Non-Linear Inversion for Relative Permittivity

by

Kim Lam

B.A.Sc, University of British Columbia 2000
M.A.Sc, University of British Columbia 2003

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

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate Studies

(Electrical Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA
(VANCOUVER)

April 2009

ABSTRACT

The gradient method, Levenberg-Marquardt (LM) method, L_2 cooled roughness (CRL2) method and L_1 cooled roughness (CRL1) method are applied to the problem of recovering the relative permittivity structure of a dielectric object. The CRL1 method is a novel technique for the recovery of the relative permittivity structure of a dielectric object introduced in this work. The frequencies used in this work range from 0.8GHz to 1.2 GHz. The size of the permittivity structure is approximately 1 wavelength, which is approximately 30cm at 1GHz.

The gradient method and LM method were unable to recover the relative permittivity structure unless the starting model is very close to the target. Both methods require a starting model that is close to the target model for them to be successful. The CRL2 method was able to recover a blurry approximation to the target relative permittivity structure. The blurriness is due to the L_2 norm. The CRL1 method is able to recover "blocky" structure. In the absence of noise, the CRL1 method was able to recover structure that was approximately one third wavelength in size. The recovery of structure at a fraction of a wavelength is highly sensitive to noise. Even at 0.01% noise, the CRL1 algorithm had difficulty recovering the exact structure.

TABLE OF CONTENTS

Chapter 1: Introduction

Chapter 2: Volume modeling and calculation of sensitivities

Chapter 3: Least square minimization techniques

Chapter 4: Performance of minimization algorithms and data selection

Chapter 5: Limits of the L1 cooled roughness regularization

Chapter 6: Summary and conclusion

LIST OF TABLES

Chapter 3

Chapter 4

Chapter 5

LIST OF FIGURES

Chapter 4

Chapter 5

Chapter 1

Introduction

1.1 Introduction to chapter 1

This thesis explores the recovery of a three dimensional structure of an unknown dielectric object using electromagnetic(EM) waves in the microwave frequency regime. The recovery is accomplished by illuminating the structure with EM-waves. The illuminated structure produces secondary waves, known as the scattered field. From the scattered field, the structure can be recovered.

Algorithms for recovering the permittivity structure can be separated into two broad classes. The first class are algorithms based on radar techniques. The goal of these techniques is to simply recover the structure of the object, but not the relative permittivity. Fear et al. (2002) is one author working in this area to detect breast cancer. The second broad category of techniques attempts to recover the relative permittivity and the structure. These techniques are denoted as tomographic techniques. The tomographic techniques require a forward modeling operator and a method to recover the structure based on the forward modeling operator. The current work falls into the category of tomographic techniques. Joachimowicz et al. (1991) did early work on tomographic recovery for biological tissue. In her research she attempted to recover the bone structure within a cross section of a leg.

Both branches of microwave imaging(MI) borrow heavily from work done in geophysics. Geophysics has used inverse techniques for exploration purposes. Although, the body of work in

geophysics is large, the problem of MI has distinct challenges and different requirements. The largest difference is the availability of the transmitted field. In geophysics, especially in seismology, the reflected field is used. In geophysics many different types of signal sources are employed, including acoustic waves, DC resistivity measurements just to name a few, while in this work microwaves are used. The scattering behavior of each source type is different, leading to a different underlying forward model. Lastly, the scaling of the problems are different. In MI, the size of an individual permittivity cell is smaller than a wavelength leading to highly non-linear behavior of the scattering with respect to the permittivity.

Many applications have been proposed for microwave imaging. Breast cancer detection, land mine detection, non-destructive testing and wood-grading are just a few examples. MI offers many theoretical advantages over existing systems. The microwave spectrum is non-ionizing unlike that of the x-ray spectrum. MI does not require a large magnetic field like MRI. Indeed, the original motivation for this work was to detect the interior structure of living trees, outlining the potential utility and portability of such a system.

However, all of these applications remain theoretical. The author makes no apologies for the state of the current technology. The work presented still cannot image an entire tree and such a goal remains lofty. Among one of the key problems for tomographic techniques is the computational cost. The cost is not insurmountable, but does require that either computing power becomes cheaper or an application sufficiently valuable is found that justifies the cost.

MI remains a laboratory technology, although is has over a 20 year history. Recent developments in general purpose graphics processing units (GPGPU) and massively parallel processors in the consumer space will substantially lower the computing cost, eventually making applications like MI economically feasible. Such a development history where the algorithm precedes the applications is not uncommon in research. One only has to look at the development of digital signal processing

(DSP), where early researchers were unable perform their algorithms in real time or even acquire voice quality digital samples. The processing power for DSP is now readily available and it is used in many applications.

Although the original application is not yet feasible, this work does present several advancements in the area of tomographic algorithms for MI. MI does outline the richness in electromagnetic scattering behavior and this works explores that.

1.2 Description of MI tomography

Electromagnetics has a history of modeling the scattering behavior of an object through the use of computer simulation. The process of modeling the scattering behavior is known as forward modeling. Tomographic microwave imaging tries to invert the forward modeling operator to recover the relative permittivity distribution. Because the dependence of the scattered field on relative permittivity is non-linear, an iterative scheme is typically used.
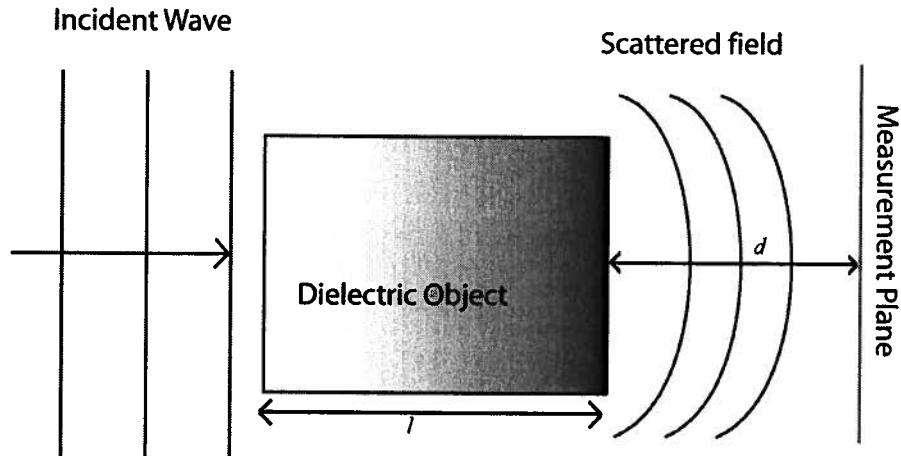
Figure 1.1: Schematic diagram of physical system. Incident waves enter the dielectric object and the dielectric object alters the incident wave to produce the scattered field. The field is measured at the measurement plane.

Figure 1.1 shows an example of a setup that can be used for MI. The size of the object, l, for this thesis is 0.3m, 1 wavelength at 1GHz. The distance, d, for this thesis is typically 0.3m, 1 wavelength at 1GHz. For inverse scattering a dielectric object is illuminated by an incident wave. The

wave interacts with the dielectric object. The interaction can include resonance behavior as the waves bounces back and forth inside the dielectric object. This resonance behavior implies that the scattered field contains information regarding the entire interior of the dielectric object. The scattered field is measured at the measurement plane to recover the relative permittivity distribution of the dielectric object. Because it is unfeasible to measure the entire scattered field, complications in data selection are introduced. The number of measurement points and the placement of the measurement plane needs to be chosen such that there is sufficient information to recover the structure via an inversion algorithm which in effect approximates the inverse operator.

```
┌──────────────────┐     ┌──────────────────┐     ┌──────────────────┐
│ Scattered Field  │     │                  │     │ Recovered        │
│ Measurements     │────▶│ Inverse Operator │────▶│ Permittivity     │
│                  │     │                  │     │                  │
└──────────────────┘     └──────────────────┘     └──────────────────┘
```
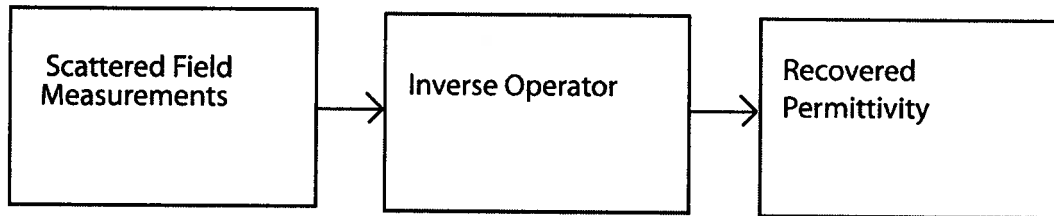
Figure 1.2: Simplified schematic view of microwave imaging. The scattered data is fed into an inversion operator to produce the recovered distribution.

For linear problems, the inverse operator can be a simple matrix inverse. For non-linear problems the inverse operator is typically approximated by an iterative algorithm. The current research falls into the the non-linear category, and therefore uses an iterative algorithm.

1.3 Previous work

Previous work has illustrated that many techniques could be used to recover the structure of a dielectric object. The structure of the dielectric object can be recovered using time domain signals or frequency domain signals. Optimization techniques ranging from genetic algorithms to nonlinear optimization algorithms have been used.

1.3.1 Radar techniques

In the area of radar-techniques, Fear et al. (2002) has made the most progress. She uses a radar based techniques to try and detect breast cancer. In this technique the radar return of a tumor from

4

several different locations is measured. Fear has successfully recovered the 3-D structure using simulated data. However, producing an ultra-wide band antenna capable of generating a pulse has proven to be difficult. The main advantage of this technique is the computational simplicity relative to that of tomographic techniques.

1.3.2 Tomographic techniques

In the area of tomographic techniques there have been many attempts at developing a MI algorithm. Most of the early works focused on recovering two dimensional relative permittivity distributions. This was due to the computational cost of MI algorithms. One common technique employed by researchers has been to construct a cost functional and then proceed to minimize the aforementioned functional. The minimization process requires many iterations, leading to long computational time.

Joachimowicz et al. (1991) experiments with local search techniques using the Newton-Kantorovich procedure. Aside from the computation cost, she points out that the starting model to be a significant problem. Because the technique are local search techniques, a starting model close to the actual distribution is required to recover the actual distribution. Franchois et al. (1998) explores the use of the Levenberg-Marquardt method for image recovery and discusses effect of receiver geometry of the recovery. She discovers many of the same problems as Joachimowicz. That is that the success of local search techniques is largely dependent on the starting model.

Because of the improvements in the performance of computers, more recently, people have begun to explore the recovery problem in 3-D. Some researchers have concentrated on producing experimental results. Geffrin et al. (2007), Yu et al. (2008), and Zaeytijd et al. (2007) have demonstrated reconstruction of dielectric objects. In particular, they have demonstrated the recovery of two dielectric cubes with real data. Others have concentrated on the theoretical aspect of creating a

three dimensional imaging algorithm. Bulyshev et al. (2001) has explored computational modeling for microwave imaging. Zhang et al. (2004) presents a three dimensional microwave imaging algorithm for biomedical imaging using the contrast source inversion method. Li et al. (2004) develops an imaging algorithm for the recovery on an object in layered media using the distorted Born method.

One artifact of the current generation of algorithms, is that the recovered relative permittivity distribution tends to be blurry. The commonality between all these techniques is that they use the L_2 norm, which strongly suggests that the blurriness is due to the L_2 norm. Lobel et al. (1996) used the edge-preserving regularization to address "blurriness" problem in the recovery of a 2-D relative permittivity distributions. In this thesis, the L_1 norm is used for 3-D microwave imaging and to address the "blurriness" problem. It recovers sharper images than previous methods. Although the method introduced in this paper is computationally more expensive than that of previous deterministic methods, it still requires less computation time than that of stochastic methods.

1.4 Novel contributions

The main novel contribution of this work is the application of the L_1 norm, through the cooled roughness L_1 method (CRL1), to the field of microwave imaging. In reaching this goal several novel concepts were developed.

A forward modeler for dielectric objects using cuboid shape functions and rooftop basis functions was developed. What is unique about this modeler is the inclusion of analytic sensitivity calculations for both the field internal and external to the dielectric object. The sensitivities are the change of the field with respect to a small perturbation of the dielectric properties of the object. These calculations are unique to each modeler.

A thorough examination of the misfit surface and how different optimization algorithms perform for this problem were presented. In examining the behavior of the CRL1 algorithm, the

concept of using a checkerboard as a test pattern was introduced to the field of microwave imaging. Although the practice is common in geophysics, such test patterns are absent from the field of microwave imaging. Through the use of different relative permittivity distributions, the limitations of the L_1, absolute value norm, and the L_2, Euclidean norm was presented.

In producing the comparison of the CRL1 method to other methods, the novel contribution of calculating the gradient for the transmitted field without the calculation of the Jacobian was introduced. This provides computational savings for researchers using the gradient method, which was not the main focus of this work.

1.5 Thesis organization

The progression of this thesis begins with a formulation of a forward modeler, followed by a description of local search techniques and ends with a broad analysis of selected numerical experiments.

Chapter 2 presents a forward modeler with roof-top basis functions and cube shape functions. In addition to simulating the internal field of a dielectric object, the calculation for the external field is presented and the sensitivities of both the external and internal fields are presented. The inclusion of the sensitivity is needed for local search techniques, because such techniques use the local sensitivity and gradient information to determine the next search location. Techniques such as genetic algorithms do not require this. The inclusion of the sensitivity is a novel contribution, because typical forward modelers are not accompanied by an analytical sensitivity calculation.

Chapter 3 presents some local search techniques. Included techniques are the gradient method, the Newton method, the Gauss-Newton Method, the Levenberg-Marquardt method, the L_2 cooled roughness method, and the L_1 cooled roughness method. The L_1 norm is achieved through the use of iterative least squares, and is based on Faquarson et al.'s (1998) work on optimization using

general norms. Different techniques are presented so that their performance can be compared. Since all these techniques are local search techniques one cannot be said to be better than any other because their performance is also dependent on the functional to be minimized. Understanding their behavior with respect to the functional surface gives great insight into the construction of the inverse operators.

Chapter 4 presents the results of using different local search techniques for MI recovery. Two test cases are analyzed so that any results are not an artifact of a given technique. Data selection is explored both in the number of illumination angles as well as frequencies to determine the influence of each selection on the performance on the algorithms. This chapter presents the case for the use of an L_1 norm to produce sharp images and the use of cooled roughness to overcome the starting model problem as described earlier.

Chapter 5 presents the the limitations of the L_1 cooled roughness method (CRL1). The CRL1 method is used to recover four different permittivity distributions. For some of these cases the CRL1 method will fail. Through these examples, the limitations can be defined.

Chapter 6 presents conclusions and future work. Although the field of MI still requires considerable effort before it becomes a practical system, this thesis makes several important contribution to understanding the problem of MI. The chapter will discuss the implications of this understanding and what it suggests for future work.

Chapter 2:

Volume modeling and calculation of sensitivities

2.1 Introduction to chapter 2

The scattering of electromagnetic waves from a dielectric object can be studied through computer simulation. The mathematics of the simulation is based on the time harmonic formulation of Maxwell's equations (Harrington, 1961) and the method of moments (Harrington, 1968). For an inversion algorithm in this thesis a volume formulation is used. For the inverse problem the permittivity parameter is assumed to be constant in a small volume, and therefore a volume modeler is most appropriate.

Volume modeling separates the dielectric object to be studied into smaller subvolumes, such that the permittivity in each subvolume is assumed to be constant and the dielectric permittivity of the object is approximated by the subvolumes. Wang (1991) uses cuboid subvolumes to model the permittivity and the electric field is assumed to be constant within a subvolume. Wang's method does not enforce the continuity of the perpendicular components of displacement field across an interface for the case where there are no free charges in the volume (Griffiths, 1989).

A second model presented by Schaubert et al. (1984) uses tetrahedral subvolumes to model the permittivity and rooftop basis functions to model the electric field. The tetrahedral shape allows for better approximation of the boundary of irregularly shaped objects. More recently Usner et al. (2006) has developed a method using curvilinear shape functions. This method has the least error in terms of

modeling the geometry of the dielectric object. The rooftop function is used to enforce the continuity of the perpendicular component of the displacement field across a charge free interface.

The modeling method presented below combines cube shape functions for modeling the permittivity and roof top basis functions for modeling the electric field. Because the structure of the object is unknown in the inverse problem, the use of tetrahedral volumes will lead to a more severe sawtooth effect than the use of a cube, because the vertices of the tetrahedron have a smaller solid angle than that of a cube.

A tetrahedron is a more appropriate choice for the shape functions when the structure of the object is known and the tetrahedrons can be used to approximate the boundary, thereby reducing the error caused by a mismatch in the shape of the simulated object and the real object. In the case of structure recovery, the shape is not known a-priori. Therefore the tetrahedral shape is no longer advantageous.

In addition to the forward model, sensitivity calculations for the internal field and scattered field are included in this thesis. The sensitivity represents the change that will occur in the electric field, when the permittivity is changed by a small amount. Sensitivities are typically not included in forward modeling formulations. The sensitivities will be used in the search methods presented in chapter 3. These sensitivities provide local information regarding how the field changes with respect to a change in the permittivity. Using this information, the structure of an object can be recovered based on the field.

The method presented is based on Schaubert et al. (1984) work. The method presented will replace the tetrahedral shape functions used by Schaubert with cube shape functions. The use of the cube shape function leads to a reduction in the computational effort needed to setup the MI algorithm due to a fewer number of cubes to approximate the volume. The use of cuboid shape functions leads to

significant difference in the formulation. In addition to the above change, the calculation of the field external to the dielectric object is included, which was not included in Shaubert's work. This allows the modeler to simulate the scattered field from an object. Lastly, the sensitivities for both the internal and external fields of the dielectric object are included. Because this thesis focuses on the problem of microwave imaging, the sensitivities are needed to facilitate the search algorithm. In summary, although this work is based on Schaubert, significant extension to that work is included in this chapter.

2.2 Mathematical theory

The solution to the scattering from a dielectric object by a time harmonic incident electromagnetic field in a homogeneous background medium is now considered. For the dielectric volume, V, the electric field as a result of its presence can be written as a summation of the known field without the volume, $\vec{E}^i(\vec{r})$, and an unknown perturbation to $\vec{E}^i(\vec{r})$, $\vec{E}^s(\vec{r})$. The expression for the electric field can be written

$$\vec{E}(\vec{r})=\vec{E}^i(\vec{r})+\vec{E}^s(\vec{r}) \ , \tag{2.1}$$

where $\vec{r}$ is the position vector. The known portion of the field $\vec{E}^i(\vec{r})$, is called the incident field, and the perturbation, $\vec{E}^s(\vec{r})$, is called the scattered field. The total electric field is denoted by $\vec{E}(\vec{r})$. The arrow above the symbols denote vectorial quantities. In this case, the fields are complex vectors in three dimensions and the position vector is a real vector in three dimensions.

The electric field is separated into two parts because the scattered field can be calculated in terms of the total electric field. The separation allows for the development of a set of linear equations to determine $\vec{E}(\vec{r})$.

The scattered field can be written as

$$\vec{E}^s(\vec{r})=-j\omega\vec{A}(\vec{r})-\nabla\Phi(\vec{r}) \ , \tag{2.2}$$

where, w is the angular frequency, $j\omega\vec{A}(\vec{r})$ is the field caused by the equivalent current elements and $\nabla\Phi(\vec{r})$ is the field caused by the charge density. The terms for the foregoing equation are given by

$$\vec{A}(\vec{r})=\frac{\mu_0}{4\pi}\int_v \vec{J}(\vec{r}')\frac{e^{-jk_0|\vec{r}-\vec{r}'|}}{|\vec{r}-\vec{r}'|}dv' \ , \text{and} \tag{2.3}$$

$$\Phi(\vec{r})=\frac{1}{4\pi\varepsilon_0}\int_v \rho(\vec{r}')\frac{e^{-jk_0|\vec{r}-\vec{r}'|}}{|\vec{r}-\vec{r}'|}dv' \ . \tag{2.4}$$

12

The wave number in free space, k_0 , can be written

$$k_0 = \omega \sqrt{\mu_0 \varepsilon_0} = 2\pi / \lambda_0 \quad,$$ (2.5)

where λ_0 is the free space wavelength.

In the above formulation, the dielectric object is replaced by an equivalent current and equivalent free charge. The equivalent current, $\vec{J}(\vec{r})$, inside a dielectric object due to the electric field is given by

$$\vec{J}(\vec{r}) = j\omega [\hat{\varepsilon}(\vec{r}) - \varepsilon_0] \vec{E}(\vec{r}) \quad,$$ (2.6)

where $\hat{\varepsilon}(\vec{r})$ is the complex permittivity as a function of position , and $\vec{E}(\vec{r})$ is the electric field. The current is proportional to the electric field and dielectric permittivity contrast. Equation (2.3) is simply the integral contribution of all the currents, and (2.4) is the integral contribution of all the equivalent free charges.

From the continuity equation, the charge density is proportional to the divergence of the current and can be written as

$$\rho(\vec{r}) = \frac{-\nabla \cdot \vec{J}(\vec{r})}{j\omega} \quad.$$ (2.7)

Therefore (2.3) and (2.4) can be written in terms of $\vec{J}(\vec{r})$ and they can be written in terms of $\vec{E}(\vec{r})$ using (2.6).

2.2.1 The tessellations

For this formulation, the arbitrary volume V is approximated by M cubes. Each cuboid subvolume is denoted as v_m . For this work, the cube v_m has sidelength s , which is the same for every subvolume.
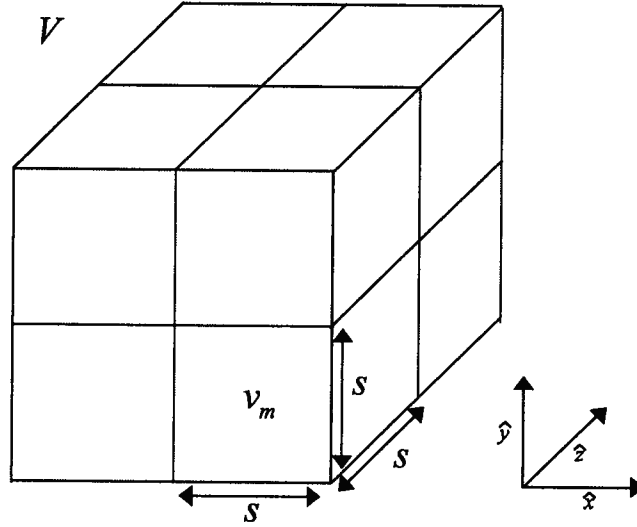
Figure 2.1. Sample tessellation for a cuboid volume, V. It consists of 4 cuboid subvolumes, each with sidelength s. Each sub-volume is denoted by V_m, where m is the index of the sub-volume which is assigned arbitrarily.

The relative permittivity is assumed to be constant in each sub-volume and is denoted ε_m. The sides of the cube are parallel to the coordinate axises. The centroid of the cube is denoted by $\vec{r}_m$.

Given the constraints, the centroid and the relative permittivity are sufficient to fully define the cube.

Such constraints can be imposed because the geometry of the object is unknown. Techniques that are employed to to reduce numerical error by matching geometry of the dielectric object are not beneficial in this case.

14

2.2.2 Basis functions: 1D example
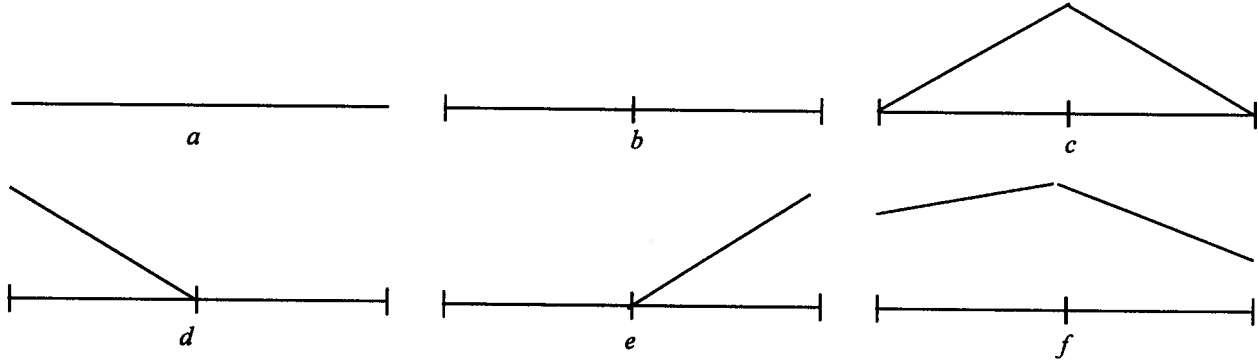
Figure 2.2: An example of tessellation and function approximation in 1D. a) Interval in which function is to be approximated. b) Interval in a separated into two sub-intervals c) Rooftop basis function spanning the entire interval. d) Rooftop basis function spanning the left line segment, defining the left side boundary. e) roof-top basis function spanning the right line segment, defining the right side boundary. f) A piecewise linear function that is a linear combination of the 3 roof-top basis functions.

Before delving into the intricacies of defining the set of basis functions, a one dimensional example is provided to illustrate some of the concepts is presented. Shown in figure 2.2a is an interval on which a function is to be approximated. Roof-top basis functions will be used to approximate the function. The interval in figure 2.2a is separated into two sub-intervals as shown in figure 2.2b. Figure 2.2c shows a basis function that spans both sub-intervals and is shaped like a roof-top. It is composed of 2 linear pieces. Roof-top basis functions typically span two intervals unless they lie on a boundary, in which case they span one interval. The value is zero at both ends and maximum at the center. Figure 2.2d shows a basis function that spans only the left sub-interval, defining the left side boundary. It is simply a linear function. Figure 2.2e shows a basis function spanning only the right sub-interval, defining the right side boundary. Figure 2.2f shows a piecewise linear function approximated by the three basis functions.

2.2.3 3D basis function for moment method modeling

This modeling algorithm will use roof-top basis functions that typically span 2 adjacent cuboid subvolumes. The basis functions that define the boundary of the object only span 1 subvolume. The

roof-top basis functions will approximate the displacement field, $\vec{D}$, to guarantee that it is continuous in the normal direction. The displacement field is related to the electric field (Griffiths, 1989) by the relationship

$$\vec{D}(\vec{r})=\varepsilon(\vec{r})\,\vec{E}(\vec{r}) \quad .$$

(2.8)

The displacement field at any given point is proportional to the electric field, weighted by the local permittivity. The displacement field, in the volume V, is represented in by a set of N basis functions and N unknown coefficients,

$$\vec{D}(\vec{r})=\sum_{n=1}^{N} D_n\,\vec{f}_n(\vec{r}) \quad .$$

(2.9)

The coefficients, D_n, are unknown, but they can be determined by the use of the method of moments (Harrington, 1968).

There is no single expression that relates the number of sub-volumes, M, to the number of basis functions for an arbitrary volume. The relationship is dependent on the volume and surface area of V. As seen in the 1-D example, the boundary must be defined by a special one sided basis function. Basis functions can occupy either 2 adjacent cubes, or 1 cube if they define the boundary of the object. The basis functions that occupy 1 cube have at least one face on the surface of V.
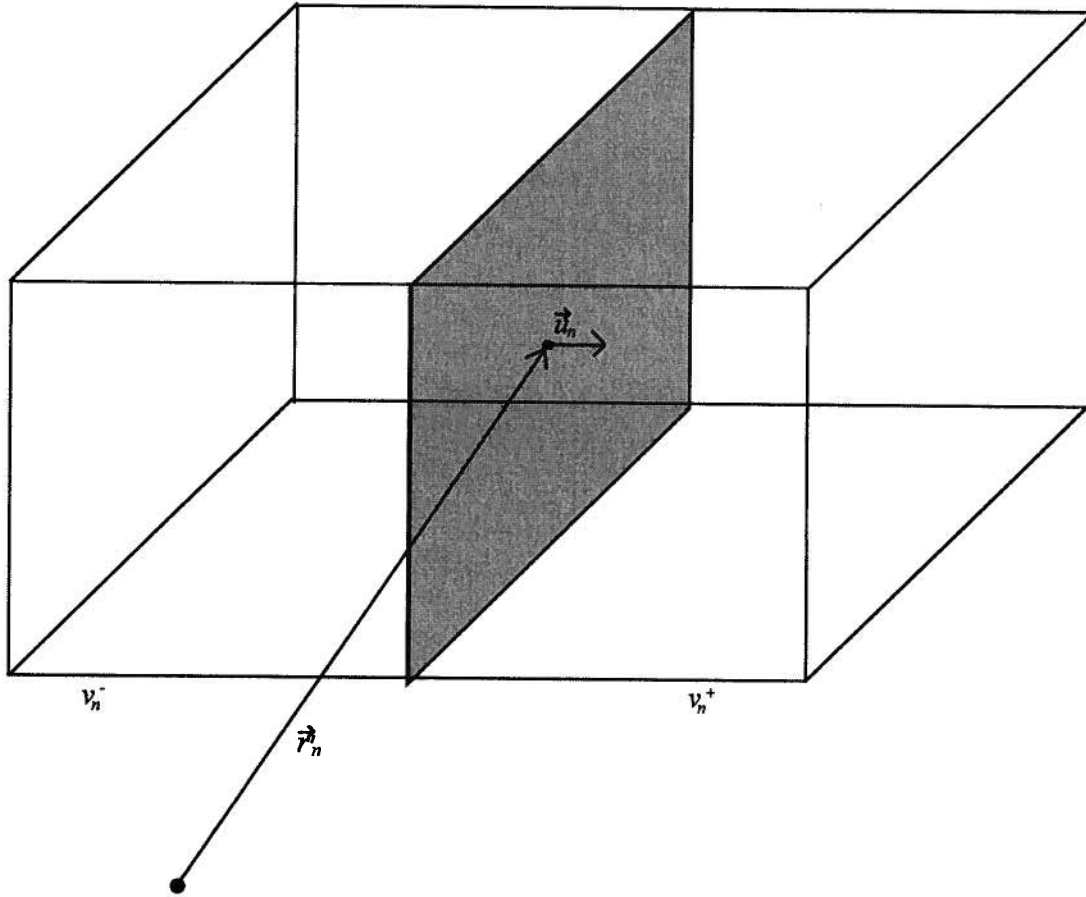
Figure 2.3: Example of the volume supports for a rooftop basis function, where the basis function does not define a boundary. It contains to subvolumes v_n^+ and v_n^-. The normal vector $\vec{u}_n$ is perpendicular to the common face of the two volumes. The position vector $\vec{r}'_n$ defines the centroid of the common face.

Figure 2.3 is a graphical representation of a basis function. The basis function has non-zero

values in two sub-volumes, unless it defines a boundary. These sub-volumes are known as the support

for the basis function. The vector $\vec{u}_n$ is normal to the common face of both cubes and is used to

define the direction of the basis function. There are three unique directions for $\vec{u}_n$, given the choice

cuboid shape functions that are all oriented in the same way. The subscript of n is used as a matter of

notational convenience such that (2.9) can be written as a single summation, rather than 3 separate

summations with one summation for each direction. For those basis functions defining the boundary,

$\vec{u}_n$ is normal to the boundary surface. The positive and negative subvolumes that support the basis

function are denoted by v_n^+ and v_n^- respectively. The positive subvolume is the subvolume that lies in the positive $\vec{u}_n$ direction of the common face and the negative subvolume is the subvolume that lies in the negative $\vec{u}_n$ direction of the common face. For notational convenience the permittivity associated with the positive and negative subvolume of the n'th basis function is denoted ε_n^+ and ε_n^- respectively. However, there are only M unique permittivity variables, one for each sub-volume.

Because the sides of the cubes are aligned with the coordinate axis, the normal vectors lie in the same direction as the coordinate axis. The position vector is given the symbol $\vec{r}'_n$, which represents the centroid of the common surface between two sub-volumes. For the case where the basis function has one sub-volume, $\vec{r}'_n$ represents the centroid of the face of the cube defining the boundary of the dielectric object.

For basis functions that only have one support, the subvolume is designated positive or negative by its relative position to the boundary surface. If the subvolume lies in the positive $\vec{u}_n$ direction of the boundary, it is designated a positive subvolume and if the subvolume lies in the negative $\vec{u}_n$ direction of the boundary, it is designated a negative subvolume.

The value of the basis function is defined by

$$\vec{f}_n(\vec{r}) = \begin{cases} \left(1 - \dfrac{(\vec{r}-\vec{r}'_n)\cdot\vec{u}_n}{s}\right)\vec{u}_n & \text{if } r \in +ve \ \ volume \\ \left(1 + \dfrac{(\vec{r}-\vec{r}'_n)\cdot\vec{u}_n}{s}\right)\vec{u}_n & \text{if } r \in -ve \ \ volume \\ 0 & otherwise \end{cases}$$

(2.10)

If the subvolume does not exist, then the value of the basis function is 0. The basis function is 0 outside its supporting subvolumes. The value of the basis function only varies along $\vec{u}_n$ and is constant along directions orthogonal to $\vec{u}_n$. The shape of the basis function when plotted along

$\vec{u}_n$ resembles a roof-top and is shown in in figure 2.4.
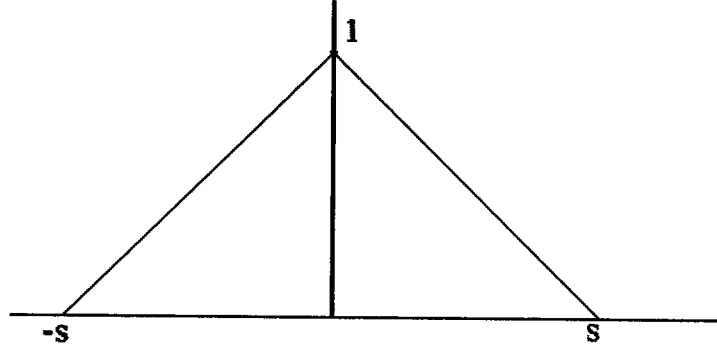
Figure 2.4. Value of a basis function plotted along the u_n direction. The value are zero at the ends and 1 and the center, creating a rooftop like shape.

The current can be written in terms of the displacement field,

$$\vec{J}(\vec{r}) = k(\vec{r}) \vec{D}(\vec{r}) \ ,$$ (2.11)

where k is given by

$$k(\vec{r}) = \frac{\varepsilon(\vec{r}) - \varepsilon_0}{\varepsilon(\vec{r})} \ .$$ (2.12)

The equivalent current is proportional to the displacement field, weighted by the dielectric contrast and should not be confused with the free space wavenumber, k_0 .

The expression for current is needed in (2.3), the expression for the magnetic vector potential.

2.3 Testing procedure

To find the unknown coefficients, D_n , in (2.9) the testing function

$$\langle \vec{g}, \vec{f} \rangle = \int_v \vec{f} \cdot \vec{g}\, dv \qquad (2.13)$$

is used to set up a linear system. This is known as the moment method[Harrington 1968].

Substituting (2.2) into (2.13) and using (2.8) and applying the testing procedure gives,

$$\langle D/\bar{\varepsilon}, \vec{f}_m \rangle + jw \langle A, \vec{f}_m \rangle + \langle \nabla \Phi, \vec{f}_m \rangle = \langle E^i, \vec{f}_m \rangle \quad . \qquad (2.14)$$

The testing functions are chosen to be the same as the basis functions as prescribed by the Galerkin

method (Volakis et al., 1998). The testing procedure is used to generate a set of linear equations, such

that the unknown field can be determined. Equation 2.14 is repeated N times, once for each basis

function, generating N equations. Within each application of (2.14) there are N unknowns. The

subscript m denotes the m'th basis function, as well as the row index when the (2.14) is converted to

matrix notation.

The right hand side of the equation represents the projection of the incident field onto the basis

functions. The first term on the left hand side represent the projection of the electric field onto the

basis function and the the last two terms on the left hand side represents the projection of the scattered

field onto the basis function.

Examining (2.3) for A and (2.4) for Φ the field at the volume of f_m , can be seen as

radiated from the other volumes. Therefore, the sub-volume where f_m is non-zero can be said to be

the destination volume, while the other sub-volumes can be said to be source volumes.

2.3.1 Individual terms in the test procedure

The individual terms are examined separately. Each term represents a contribution to the electric field

of the destination from the source cell.

2.3.2 Testing the electric field, $\langle D/\vec{\varepsilon}, f_m \rangle$

The equation

$$\langle D/\vec{\varepsilon}, f_m \rangle = \sum_{n=1}^{N} \int_V \frac{D_n}{\varepsilon(\vec{r})} \vec{f}_n \cdot \vec{f}_m \, dv \qquad (2.15)$$

will be evaluated and be written as

$$\langle D/\vec{\varepsilon}, f_m \rangle = \sum_{n=1}^{N} \left(\frac{T^1_{mn}}{\varepsilon_n^+} + \frac{T^2_{nn}}{\varepsilon_n^-} \right) D_n \quad . \qquad (2.16)$$

In (2.16), the dependence on permittivity is shown explicitly. The explicit presentation of the permittivity will aid in the calculation of the sensitivities. The coefficients T^1_{mn} and T^1_{mn} are presented below.

When $\vec{f}_n$ and $\vec{f}_m$ have non-overlapping volumes, (2.15) is zero. It is also zero when $\vec{u}_n$ and $\vec{u}_m$ are orthogonal. When n = m,

$$\int_V \frac{1}{\varepsilon(\vec{r})} \vec{f}_n \cdot \vec{f}_m \, dv = \frac{s^3}{3} \left(\frac{1}{\varepsilon_m^+} + \frac{1}{\varepsilon_n^-} \right) \quad . \qquad (2.17)$$

This is the self term, because electric field component of the n'th basis function is projected onto itself. If the positive volume of $\vec{f}_n$ overlaps the negative volume of $\vec{f}_m$,

$$\int_V \frac{1}{\varepsilon(\vec{r})} \vec{f}_n \cdot \vec{f}_m \, dv = \frac{s^3}{6} \frac{1}{\varepsilon_n^+} \quad . \qquad (2.18)$$

If the negative volume of $\vec{f}_n$ overlaps the positive volume of $\vec{f}_m$,

$$\int_V \frac{1}{\vec{\varepsilon}(\vec{r})} \vec{f}_n \cdot \vec{f}_m \, dv = \frac{s^3}{6} \frac{1}{\varepsilon_n^-} \qquad (2.19)$$

The permittivity independent term associated with the positive volume of $\vec{f}_n$ is

$$T^1_{mn} = \begin{cases} \dfrac{s^3}{3} & \text{if } m=n \\[2mm] \dfrac{s^3}{6} & \text{if } v_n^+ = v_m^- \\[2mm] 0 & \text{otherwise} \end{cases} \quad . \qquad (2.20)$$

The permittivity independent term associated with the negative volume of $\vec{f}_n$ is

$$T^2_{mn} = \begin{cases} \dfrac{s^3}{3} & \text{if } m = n \\ \dfrac{s^3}{6} & \text{if } v_n^- = v_m^+ \\ 0 & \text{otherwise} \end{cases} \qquad (2.21)$$

2.3.3 Testing the current contribution, $jw\langle \vec{A}(\vec{r}), f_m \rangle$

The next term, $jw\langle \vec{A}(\vec{r}), f_m \rangle$ represents the contribution to the electric field from the

equivalent currents inside the object. To make the problem tractable $\vec{A}(\vec{r})$ is assumed to be constant

within each subvolume of $\vec{f}_n$.

The value of $\vec{A}(\vec{r})$ is taken at the center of the cubes of $\vec{f}_m$.

Substituting (2.9) into (2.11) and (2.11) into (2.2) the expression for $\vec{A}(\vec{r})$ can be written as

$$\vec{A}(\vec{r}) = \frac{\mu_0}{4\pi} \sum_{n=1}^{N} D_n \int_V jwk(\vec{r}') \vec{f}_n(\vec{r}) \frac{e^{-jk_0|\vec{r}-\vec{r}'|}}{|\vec{r}-\vec{r}'|} dv' \quad . \qquad (2.22)$$

The testing term, $\langle \vec{A}, \vec{f}_m \rangle$, can also be written as

$$\langle \vec{A}, \vec{f}_m \rangle = \sum_{n=1}^{N} D_n (k_n^+ (V^1_{mn} + V^3_{mn}) + k_n^- (V^2_{mn} + V^4_{mn})) \quad , \qquad (2.23)$$

where the dependence on the permittivity is written with respect to the permittivity contrast. The co-

efficients in (2.23) are presented below.

The basis function $\vec{f}_m$ is separated into its constituent subvolumes and the integration is

evaluated separately for each subvolume. Because the basis function has a value of 0 outside the

subvolume, the integral only needs to be evaluated inside the subvolumes. Evaluating $\vec{A}(\vec{r})$ at a

given $\vec{r}$ results in,

$$\vec{A}(\vec{r}) = \frac{\mu_0 jw}{4\pi} \sum_{n=1}^{N} D_n \left[k_n^+ \int_{v_n^+} \vec{f}_n(\vec{r}') \frac{e^{-jk_0|\vec{r}-\vec{r}'|}}{|\vec{r}-\vec{r}'|} dv' + k_n^- \int_{v_n^-} \vec{f}_n(\vec{r}') \frac{e^{-jk_0|\vec{r}-\vec{r}'|}}{|\vec{r}-\vec{r}'|} dv' \right] . \qquad (2.24)$$

The integral in (2.23) can be evaluated numerically.

Applying the testing procedure and assuming $\vec{A}(\vec{r})$ is constant within each subvolume results in

$$\langle \vec{A}, \vec{f}_m \rangle = \sum_{n=1}^{N} \vec{A}(\vec{d}_m^+)\cdot \int_{v_m^+} f_m(\vec{r})\,dv + \vec{A}(\vec{d}_m^-)\cdot \int_{v_m^-} f_m(\vec{r})\,dv \qquad (2.25)$$

The vectors $\vec{d}_m^+$ is the center of the positive volume and $\vec{d}_m^-$ is the center of the negative volume

for the basis function $\vec{f}_m$. If the positive subvolume does not exist then $\int_{v_m^+} \vec{f}_m(\vec{r})\,dv$ is equal to

0, and if the negative subvolume does not exist then $\int_{v_m^-} \vec{f}_m(\vec{r})\,dv$ is equal to 0.

For the case where the basis function occupy two subvolumes, (2.25) can be written as

$$jw\langle \vec{A}, \vec{f}_m \rangle = jw\,\vec{A}(\vec{d}_m^+)\cdot\vec{u}_m\frac{s^3}{2} + jw\,\vec{A}(\vec{d}_m^-)\cdot\vec{u}_m\frac{s^3}{2} \ . \qquad (2.26)$$

When $\vec{u}_n$ and $\vec{u}_m$ are orthogonal, then $\langle \vec{A}, \vec{f}_m \rangle$ is 0. To reduce the amount of computation

needed, orthogonality can be determined first, before carrying out a lengthy computation.

Four permittivity independent expressions can be written for (2.26) and are listed below:

The contribution of the current from the positive volume of the source basis function, to the positive

side of the test basis function is given by

$$V_{mn}^1 = \begin{cases} \dfrac{-\mu_0 w^2 s^3}{8\pi} \displaystyle\int_{v_n^+} f_n(r')\cdot u_m \dfrac{e^{-jk_0|\vec{d}_m^+ - \vec{r}'|}}{|\vec{d}_m^+ - \vec{r}'|}\,dv' & \text{if } v_m^+ \text{ and } v_n^+ \text{ exists} \\[4pt] 0 & \text{otherwise} \end{cases} \qquad (2.27)$$

The contribution of the current from the negative volume of the source basis function, to the positive

volume of the test basis function is given by

$$V_{mn}^2 = \begin{cases} \dfrac{-\mu_0 w^2 s^3}{8\pi} \displaystyle\int_{v_n^-} f_n(r')\cdot u_m \dfrac{e^{-jk_0|\vec{d}_m^+ - \vec{r}'|}}{|\vec{d}_m^+ - \vec{r}'|}\,dv' & \text{if } v_m^+ \text{ and } v_n^- \text{ exists} \\[4pt] 0 & \text{otherwise} \end{cases} \qquad (2.28)$$

The contribution of the current from the positive volume of the source basis function, to the negative

volume of the test basis function is given by

$$V_{mn}^3 = \begin{cases} \dfrac{-\mu_0 w^2 s^3}{8\pi} \displaystyle\int_{v_n^+} f_n(r') \cdot u_m \dfrac{e^{-jk_0|\vec{d_m^-}-\vec{r}'|}}{|\vec{d_m^-}-\vec{r}'|} dv' & \text{if } d_m^- \text{ and } v_n^+ \text{ exists} \\ 0 & \text{otherwise} \end{cases} \qquad (2.29)$$

The contribution of the current from the negative volume of the source basis function, to the negative volume of the test basis function is give by

$$V_{mn}^4 = \begin{cases} \dfrac{-\mu_0 w^2 s^3}{8\pi} \displaystyle\int_{v_n^-} f_n(r') \cdot u_m \dfrac{e^{-jk_0|\vec{d_m^-}-\vec{r}'|}}{|\vec{d_m^-}-\vec{r}'|} dv' & \text{if } v_m^- \text{ and } v_n^- \text{ exists} \\ 0 & \text{otherwise} \end{cases} \qquad (2.30)$$

In summary, the contribution from the equivalent currents can be evaluated numerically. Four coefficients can be defined such that the dependence on permittivity contrast can be shown explicitly, as shown in (2.23).

2.3.4 Testing the charge contribution, $\langle \nabla \Phi, \vec{f}_m \rangle$

The $\langle \nabla \Phi, \vec{f}_m \rangle$ term is given by

$$\langle \nabla \Phi, \vec{f}_m \rangle = \oint_S \Phi \vec{f}_m \cdot \vec{n} \, ds - \int_V \Phi \nabla \cdot \vec{f}_m \, dv \qquad (2.31)$$

and it represents the contribution to the electric field by the equivalent free charges. The expression for the contribution to the field can be written as

$$\langle \nabla \Phi, \vec{f}_m \rangle = \sum_{n=1}^{N} D_n(k_n^+ (C_{mn}^1 + C_{mn}^4) + k_n^- (C_{mn}^2 + C_{mn}^5) + (k_n^- - k_n^+)(C_{mn}^3 + C_{mn}^6)) , \qquad (2.32)$$

where the dependence on the permittivity contrast is explicitly shown.

Examining the divergence of the basis function results in

$$\nabla \cdot f_m = \begin{cases} \dfrac{-1}{s} & v_m^+ \\ \dfrac{1}{s} & v_m^- \\ 0 & \text{otherwise} \end{cases} , \qquad (2.33)$$

which is simply the slope of the rooftop function.

To make the problem tractable $\Phi(\vec{r})$ is evaluated at the centroid of the cubes of the basis function.

24

An expression involving D_n will be derived from

$$\Phi(\vec{r}) = \frac{1}{4\pi\varepsilon_0} \int_V \rho(\vec{r}\,') \frac{e^{-jk_0|\vec{r}-\vec{r}\,'|}}{|\vec{r}-\vec{r}\,'|} dv' \quad .$$ (2.34)

To arrive at an expression involving D_n , (2.11) and (2.9) is substituted into (2.7) giving an expression for the charge,

$$\rho(\vec{r}) = -\sum_{n=1}^{N} D_n k(\vec{r}) \nabla \cdot f_n(\vec{r}) - \sum_{n=1}^{N} D_n f_n(\vec{r}) \cdot \nabla k(\vec{r}) \quad .$$ (2.35)

The gradient of $k(\vec{r})$ only exists on the interface between the two volumes of the basis function and is perpendicular to the surface. The divergence of the basis function is given in (2.33).

Substituting (2.35) into (2.34) results in an expression for the contribution of the charge to the electric field,

$$\Phi(\vec{r}) = \sum_{n=1}^{N} \frac{D_n}{4\pi\varepsilon_0} \left[\frac{-k_n^+}{s} \int_{v_n^+} \frac{e^{-jk_0 R}}{|\vec{r}-\vec{r}\,'|} dv' + \frac{k_n^-}{s} \int_{v_n^-} \frac{e^{-jk_0|\vec{r}-\vec{r}\,'|}}{|\vec{r}-\vec{r}\,'|} dv' + (k_n^- - k_n^+) \int_{a_n} \frac{e^{-jk_0|\vec{r}-\vec{r}\,'|}}{|\vec{r}-\vec{r}\,'|} ds' \right]$$ (2.36)

The first two integrals in (2.36) are volume integrals and is 0 if the volume does not exist. The third integral is a surface integral and is taken along the boundary between the positive and negative volume of the basis function. If the negative volume does not exist, k_n^- is 0 and if the positive volume does not exist then k_n^+ is taken to be 0, which is equivalent of substituting the background permittivity into (2.12) which in this case the background permittivity is ε_0 .

Evaluating $\Phi(\vec{r})$ at the centers of the subvolumes of the basis function , substituting (2.33) directly into the (2.31) and integrating, the following expression is obtained.

$$\langle \nabla \Phi, f_m \rangle = \begin{cases} s^2 [\Phi(\vec{d}_m^+) - \Phi(\vec{d}_m^-)] & \text{if } v_m^+ \text{ and } v_m^- \text{ exists} \\ s^2 [\pm \Phi(\vec{d}_m^\pm) \mp \Phi(\vec{d}_m^\pm \mp \frac{s}{2}\vec{u}_m)] & \text{if } v_m^\mp \text{ does not exist} \end{cases} \quad .$$ (2.37)

The vectors $\vec{d}_m^+$ and $\vec{d}_m^-$ represents the centers of the positive subvolume and negative subvolume of f_m , respectively. If the subvolume does not exist, the integral is evaluated at the boundary

surface. The surface integral in (2.31) is 0 except at the boundaries of the object.

The contribution from the positive volume of the source basis function to the positive subvolume of the test basis function is given by

$$
C_{mn}^1 = \begin{cases} \dfrac{s}{4\pi\varepsilon_0} \displaystyle\int\limits_{v_n^+} \dfrac{e^{-jk_0|\vec{d}_m^+ - \vec{r}\,'|}}{|\vec{d}_m^+ - \vec{r}\,'|}\, dv' & \text{if } v_m^+ \text{ and } v_n^+ \text{ exists} \\[2em] \dfrac{s}{4\pi\varepsilon_0} \displaystyle\int\limits_{v_n^+} \dfrac{e^{-jk_0|\vec{d}_m^+ + \frac{s}{2}\vec{u}_m - \vec{r}\,'|}}{|\vec{d}_m^+ + \frac{s}{2}\vec{u}_m - \vec{r}\,'|}\, dv' & \text{if } v_m^+ \text{ does not exist and } v_n^+ \text{ exists} \\[2em] 0 & \text{otherwise} \end{cases}
\qquad (2.38)
$$

The contribution from the negative volume of the source basis function to the positive subvolume of the test basis function is given by

$$
C_{mn}^2 = \begin{cases} \dfrac{s}{4\pi\varepsilon_0} \displaystyle\int\limits_{v_n^-} \dfrac{e^{-jk_0|\vec{d}_m^+ - \vec{r}\,'|}}{|\vec{d}_m^+ - \vec{r}\,'|}\, dv' & \text{if } v_m^+ \text{ and } v_n^- \text{ exists} \\[2em] \dfrac{s}{4\pi\varepsilon_0} \displaystyle\int\limits_{v_n^-} \dfrac{e^{-jk_0|\vec{d}_m^- + \frac{s}{2}\vec{u}_m - \vec{r}\,'|}}{|\vec{d}_m^- + \frac{s}{2}\vec{u}_m - \vec{r}\,'|}\, dv' & \text{if } v_m^+ \text{ does not exist and } v_n^- \text{ exists} \\[2em] 0 & \text{otherwise} \end{cases}
\qquad (2.39)
$$

The contribution of the surface charge to the positive subvolume of the test basis function is given by

$$
C_{mn}^3 = \begin{cases} \displaystyle\int\limits_{a_n} \dfrac{e^{-jk_0|\vec{d}_m^+ - \vec{r}\,'|}}{|\vec{d}_m^+ - \vec{r}\,'|}\, ds' & \text{if } v_{mn}^+ \text{ exists} \\[2em] \displaystyle\int\limits_{a_n} \dfrac{e^{-jk_0|\vec{d}_m^- + \frac{s}{2}\vec{u}_m - \vec{r}\,'|}}{|\vec{d}_m^- + \frac{s}{2}\vec{u}_m - \vec{r}\,'|}\, ds' & \text{otherwise} \end{cases}
\qquad (2.40)
$$

The contribution from the negative subvolume of the source basis function to the positive subvolume of the test basis function is given by

$$C^4_{mn} = \begin{cases} \dfrac{s}{4\pi\varepsilon_0} \displaystyle\int\limits_{v_n^+} \dfrac{e^{-jk_0|\vec{d}_m^- - \vec{r}\,'|}}{|\vec{d}_m^- - \vec{r}\,'|}\, dv & \text{if } v_m^- \text{ and } v_n^+ \text{ exists} \\[18pt] \dfrac{s}{4\pi\varepsilon_0} \displaystyle\int\limits_{v_n^+} \dfrac{e^{-jk_0|\vec{d}_m^+ - \frac{s}{2}\vec{u}_m - \vec{r}\,'|}}{|\vec{d}_m^+ - \frac{s}{2}\vec{u}_m - \vec{r}\,'|}\, dv & \text{if } v_m^- \text{ does not exist and } v_n^+ \text{ exists} \\[18pt] 0 & \text{otherwise} \end{cases}$$
(2.41)

The contribution from the negative subvolume of the source basis function to the negative subvolume of the test basis function is given by

$$C^5_{mn} = \begin{cases} \dfrac{s}{4\pi\varepsilon_0} \displaystyle\int\limits_{v_n^-} \dfrac{e^{-jk_0|\vec{d}_m^- - \vec{r}\,'|}}{|\vec{d}_m^- - \vec{r}\,'|}\, dv & \text{if } v_m^- \text{ and } v_n^- \text{ exists} \\[18pt] \dfrac{s}{4\pi\varepsilon_0} \displaystyle\int\limits_{v_n^-} \dfrac{e^{-jk_0|\vec{d}_m^+ - \frac{s}{2}\vec{u}_m - \vec{r}\,'|}}{|\vec{d}_m^+ - \frac{s}{2}\vec{u}_m - \vec{r}\,'|}\, dv & \text{if } v_m^- \text{ does not exist and } v_n^- \text{ exists} \\[18pt] 0 & \text{otherwise} \end{cases}$$
(2.42)

The contribution of the surface charge to the negative test volume is given by

$$C^6_{mn} = \begin{cases} \displaystyle\int\limits_{a_n} \dfrac{e^{-jk_0|\vec{d}_m^- - \vec{r}\,'|}}{|\vec{d}_m^- - \vec{r}\,'|}\, ds & \text{if } v_{mn}^- \text{ exists} \\[18pt] \displaystyle\int\limits_{a_n} \dfrac{e^{-jk_0|\vec{d}_m^+ - \vec{r}\,'|}}{|\vec{d}_m^+ - \frac{s}{2}\vec{u}_m - \vec{r}\,'|}\, ds & \text{otherwise} \end{cases}$$
(2.43)

In summary, the contribution by the induced charges can be written as 4 volume integrals and 2 surface integrals where the integrals are evaluated numerically.

2.3.5 Total expression for forward modeling

Now that the individual components of (2.14) have been analyzed, a set of linear equations can be developed to solve for the unknown coefficients D_n.

This set of linear equations can be written as

$$\langle E^i, f_m \rangle = \sum_{n=1}^{N} D_n S_{mn} \, .$$
(2.44)

The testing of the incident field $\langle E^i, f_m \rangle$, can be carried out numerically.

The coefficient, S_{mn} , can be expressed as

$$S_{mn} = \frac{T^1_{mn}}{\varepsilon_n^+} + \frac{T^2_{mn}}{\varepsilon_n^-} + k_n^+ \, T^3_{mn} + k_n^- \, T^4_{mn} \tag{2.45a}$$

where

$$T^3_{mn} = (V^1_{mn} + V^3_{mn} + C^1_{mn} + C^4_{mn} + C^3_{mn} - C^6_{mn}) \tag{2.45b}$$

is coefficient associated with the positive volume and

$$T^4_{mn} = (V^2_{mn} + V^4_{4mn} + C^2_{mn} + C^5_{mn} - C^3_{mn} + C^6_{mn}) \tag{2.45c}$$

is the coefficient associated with the negative volume.

2.3.6 Calculation of derivative for the internal field

The derivative with respect to the k'th permittivity value can be calculated from (2.44) using the formula

$$\sum_{n=1}^{N} \frac{\partial S_{mn} D_n}{\partial \varepsilon_k} = \frac{\partial \langle \vec{E}^i_m, f_m \rangle}{\partial \varepsilon_k} \tag{2.46}$$

resulting in

$$\sum_{n=1}^{N} \left(\frac{\partial S_{mn}}{\partial \varepsilon_k} D_n + S_{mn} \frac{\partial D_n}{\partial \varepsilon_k} \right) = 0 \quad , \tag{2.47}$$

because the incident field has no dependence on the permittivity.

To calculate $\dfrac{\partial D_n}{\partial \varepsilon_k}$ the set of linear equations,

$$\sum_{n=1}^{N} S_{mn} \frac{\partial D_n}{\partial \varepsilon_k} = -\sum_{n=1}^{N} \frac{\partial S_{mn}}{\partial \varepsilon_k} D_n \quad , \tag{2.48}$$

can be solved. The right hand side is known, since the solution to D_n can be determined before the calculation of the sensitivity.

The partial derivative of the individual matrix terms is given by

$$\frac{\partial S_{mn}}{\partial \varepsilon_k} = \frac{-T^1_{mn}}{\varepsilon_n^{-2}} \frac{\partial \varepsilon_n^+}{\partial \varepsilon_k} - \frac{T^2_{mn}}{\varepsilon_n^{+2}} \frac{\partial \varepsilon_n^-}{\partial \varepsilon_k} + \frac{\partial k_n^-}{\partial \varepsilon_k} T^3_{mn} + \frac{\partial k_n^+}{\partial \varepsilon_k} T^4_{mn} \quad . \tag{2.49}$$

By calculating and storing the sub-terms of S_{mn} the right hand side can be generated without the recalculation of the individual integrals. The left hand side of (2.48) is a matrix-vector product, with the same matrix as is used in (2.44).

2.4 The field external to the object

The field external to the dielectric object can be calculated directly from equation (2.2). The current contribution, $\vec{A}(\vec{r})$, can be computed using (2.23).

To calculate the contribution of the charges, $\nabla \Phi(\vec{r})$, a finite difference approximation can be used. That is ,

$$
\begin{aligned}
\nabla \Phi(\vec{r}) = & (\Phi(\vec{r}+0.5\Delta_x \vec{i}) - \Phi(\vec{r}-0.5\Delta_x \vec{i}))\frac{\vec{i}}{\Delta_x} \\
& + (\Phi(\vec{r}+0.5\Delta_y \vec{j}) - \Phi(\vec{r}-0.5\Delta_y \vec{j}))\frac{\vec{j}}{\Delta_y} \\
& + (\Phi(\vec{r}+0.5\Delta_z \vec{k}) - \Phi(\vec{r}-0.5\Delta_z \vec{k}))\frac{\vec{z}}{\Delta_k}
\end{aligned}
\qquad (2.50)
$$

can be used to approximate the gradient term.

Once again, it is useful for the calculation of the derivative with respect to the permittivity to define a permittivity independent term.

The contribution of the current from the positive volume is given by

$$
V_n^1(\vec{r}) = \begin{cases} \dfrac{j w \mu_0}{4\pi} \displaystyle\int_{v_n^+} f_n(\vec{r}')\dfrac{e^{-jk_0|\vec{r}-\vec{r}'|}}{|\vec{r}-\vec{r}'|}dv' & \text{if } v_n^+ \text{ exists} \\ 0 & \text{otherwise} \end{cases}
\qquad (2.51)
$$

The contribution of the negative volume is given by

$$
V_n^2(\vec{r}) = \begin{cases} \dfrac{j w \mu_0}{4\pi} \displaystyle\int_{v_n^-} f_n(\vec{r}')\dfrac{e^{-jk_0|\vec{r}-\vec{r}'|}}{|\vec{r}-\vec{r}'|}dv' & \text{if } v_n^- \text{ exists} \\ 0 & \text{otherwise} \end{cases}
\qquad (2.52)
$$

The contribution from the currents can be summarized as,

$$
\vec{A}(\vec{r}) = \sum_{n=1}^{N} D_n\left(k_n^+ \, V_n^1(\vec{r}) + k_n^- \, V_n^2(\vec{r})\right) .
\qquad (2.53)
$$

The contribution of the charges in the positive volume

$$C_n^1(\vec{r}) = \begin{cases} \dfrac{-1}{4\pi\varepsilon_0 s} \displaystyle\int_{v_n^+} \dfrac{e^{-jk_0|\vec{r}-\vec{r}'|}}{|\vec{r}-\vec{r}'|} \, dv' & \text{if } v_n^+ \text{ exists} \\ 0 & \text{otherwise} \end{cases}$$ (2.53)

The contribution of the charges in the negative volume

$$C_n^2(\vec{r}) = \begin{cases} \dfrac{1}{4\pi\varepsilon_0 s} \displaystyle\int_{v_n^-} \dfrac{e^{-jk_0|\vec{r}-\vec{r}'|}}{|\vec{r}-\vec{r}'|} \, dv' & \text{if } v_n^- \text{ exists} \\ 0 & \text{otherwise} \end{cases}$$ (2.54)

the contribution of the surface charges is given by

$$C_n^3(\vec{r}) = \frac{1}{4\pi\varepsilon_0} \int_{a_n} \frac{e^{-jk_0|\vec{r}-\vec{r}'|}}{|\vec{r}-\vec{r}'|} \, ds' \quad .$$ (2.55)

The total expression for $\Phi(\vec{r})$ is given by

$$\Phi(\vec{r}) = \sum_{n=1}^{N} D_n\left(k_n^+ \, C_n^1(\vec{r}) + k_n^- \, C_n^2(\vec{r}) + (k_n^- - k_n^+)C_3(\vec{r}) \right)$$ (2.56)

2.4.1 Sensitivities of the observations

The sensitivities of the scattered data can be calculated directly from (2.2) and is given by

$$\frac{\partial \vec{E}^s(\vec{r})}{\partial \varepsilon_k} = \frac{-\partial \, jw\, \vec{A}(\vec{r})}{\partial \varepsilon_k} - \frac{\partial \nabla \Phi(\vec{r})}{\partial \varepsilon_k}$$ (2.57)

The terms of the right hand side of (2.57) can be calculated.

The derivative with respect to the current contribution is given by

$$\frac{\partial \vec{A}(\vec{r})}{\partial \varepsilon_k} = \sum_{n=1}^{N} \frac{\partial D_n}{\partial \varepsilon_k}\left(k_n^+ \, V_n^1(\vec{r}) + k_n^- \, V_n^2(\vec{r}) \right) + D_n\left(\frac{\partial k_n^+}{\partial \varepsilon_k} V_n^1(\vec{r}) + \frac{\partial k_n^-}{\partial \varepsilon_k} V_n^2(\vec{r}) \right) \quad .$$ (2.58)

Because the finite difference approximation to the gradient of is $\nabla \Phi(\vec{r})$ used, the following expression

$$\begin{aligned} \frac{\partial \Phi(\vec{r})}{\partial \varepsilon_k} = &\sum_{n=1}^{N} \frac{\partial D_n}{\partial \varepsilon_k}\left(k_n^+ \, C_n^1(\vec{r}) + k_n^- \, C_n^2(\vec{r}) + (k_n^- - k_n^+)C_3(\vec{r}) \right) \\ &+ D_n\left(\frac{\partial k_n^+}{\partial \varepsilon_k} C_n^1(\vec{r}) + \frac{\partial k_n^-}{\partial \varepsilon_k} C_n^2(\vec{r}) + \frac{\partial(k_n^- - k_n^+)}{\partial \varepsilon_k} C_3(\vec{r}) \right) \end{aligned}$$ (2.59)

can be used to determine the derivative with respect to the permittivity. To determine the derivative, (2.59) is applied 6 times, because in (2.50), $\Phi(\vec{r})$ appears 6 times.

31

2.5 Verification of the formulation

To verify the formulation presented above, the scattered field generated by the pulsed based moment method (Wang, 1991), ComSol and the formulation presented above will be compared. The three forward modelers should produce similar results because they model the same physical system.

Wang's variation of the moment method uses cube shape functions to represent the dielectric object and pulse basis functions to represent the electric field. The moment method presented in this thesis uses cube shape functions to represent the dielectric object and rooftop basis functions to represent the electric field. ComSol is a commercially available multiphysics solver based on the Finite Element Method (Jin, 2002). The data will be generated using the RF module, in version 3.4 of ComSol.
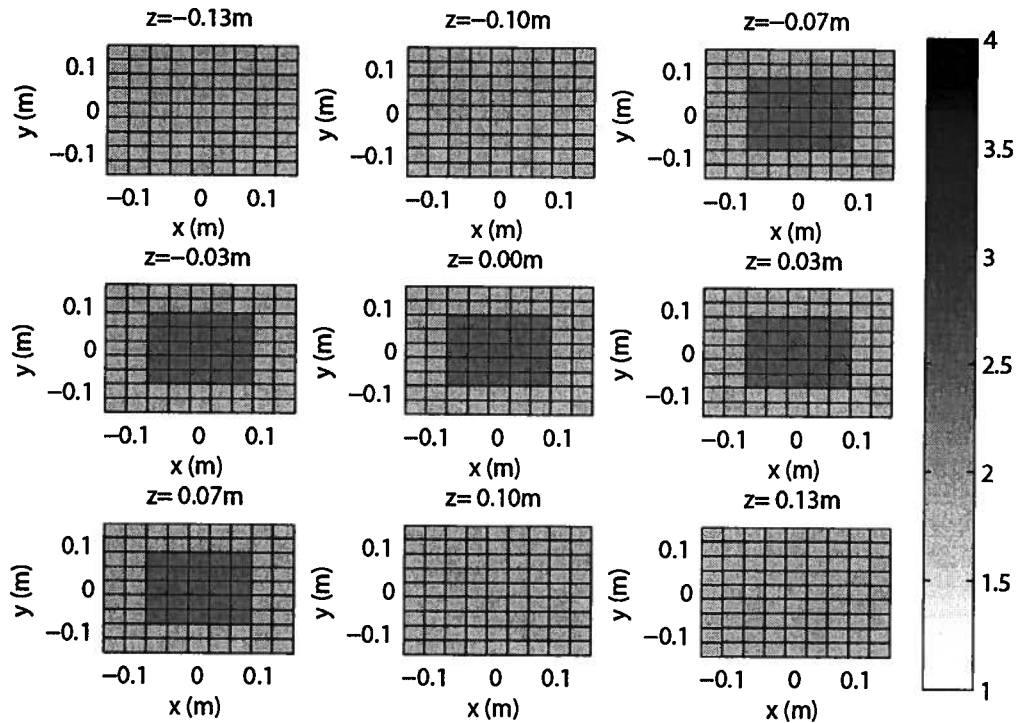
Figure 2.5: Relative permittivity distribution consisting of two cubes. The outer cube has a relative permittivity of 2.0 and a sidelength of 0.3m. The inner cube has a relative permittivity of 3.0 and a sidelength of 0.17m.

The scattered field generated with the aforementioned forward modelers for the relative

permittivity distribution in figure 2.5 will be compared. The excitation frequency is 1GHz. The

freespace wavelength at 1GHZ, λ_0 , is 0.3m. The relative permittivity distribution will be

illuminated by a planewave that is propagating in the x direction, and polarized in the z direction. The

data will be measured on the z-y plane, at x=0.30m ($1\lambda_0$). The permittivity distribution consists of

two cubes. The outer cube has sidelength of 0.30m($1\lambda_0$) and relative permittivity of 2.0. The inner

cube has of sidelength 0.17m($0.56\lambda_0$) and relative of permittivity 3.0.

Figure 2.6: A comparison between the x-component of the scattered field from 3 forward modelers. The image on the left shows the scattered field from ComSol. The image in the center shows the scattered field from the moment method using rooftop basis functions. The image on the right shows the scattered field from the moment method using pulse basis functions.

Figure 2.6 shows surface plots of the x-component of the simulated scattered field on the

measurement plane for the three forward modelers. The plot on the left shows the x-component of the

scattered field generated by ComSol. The plot in the center shows the x-component of the scattered

field generated by the rooftop moment method. The plot on the right shows the x-component of the

scattered field generated by the pulse moment method. A large dark spot appears in the upper and

lower halfs of all three images. The x-component of the scattered field generated by ComSol has a

greater magnitude than those generated by the other two forward modelers.

Figure 2.7: Surface plot of the difference between the x-component of the electric field for the three forward modelers. The surfaces plot the magnitude of the difference between the magnitude of the fields of the three forward modelers.

Figure 2.7 shows the differences between the magnitudes of the x-component of the electric fields from the three forward modelers. The plot on the left shows the difference between rooftop moment method and the pulse moment method. The plot in the center shows the difference between ComSol and the pulse moment method. The plot on the right shows the difference between ComSol and the rooftop moment method.

Figure 2.8: Comparison of the of the x-component of the scattered field between 3 forward modelers. The top graph represents the magnitude and the bottom graph represents the phase. The data was measured on the line x=0.3m, y=0m and z is the independent variable.

Figure 2.8 compares the x-component of the scattered field along the line where x=0.3m, y=0m and z is the independent variable. The top graph of figure 2.8 shows that the x-component of the three scattered fields have similarly shaped envelopes but differ in magnitude. In figure 2.8, x-component of the scattered field generated by ComSol has the greatest magnitude, while the x-component of the scattered field generated by the pulse moment method has the smallest magnitude. In the bottom graph of figure 2.8, the phase of the x-component of the scattered fields generated by ComSol and the rooftop moment method are similar.

Figure 2.9: A comparison between the y-component of the scattered field from 3 forward modelers. The image on the left shows the result from ComSol. The image in the center shows the results from the moment method using rooftop basis functions. The image on the right shows the results from the moment method using pulse basis functions.

Figure 2.9 shows surface plots of the y-component of the simulated scattered field on the measurement plane for the three forward modelers. The plot on the left shows the y-component of the scattered field generated by ComSol. The plot in the center shows the y-component of the scattered field generated by the rooftop moment method. The plot on the right shows the y-component of the scattered field generated by the pulse moment method. The the y-component of the three scattered fields look similar. Each has a white cross shaped region in the center. In figure 2.9, four dark regions, one region for each corner, are also apparent in each surface plot.

Figure 2.10: Surface plot of the difference between the y-component of the electric field for the three forward modelers. The surfaces plot the magnitude of the difference between the magnitude of the fields of the three forward modelers.

Figure 2.10 shows the differences between the magnitudes of the y-component of the electric fields from the three forward modelers. The plot on the left shows the difference between rooftop moment method and the pulse moment method. The plot in the center shows the difference between

36

ComSol and the pulse moment method. The plot on the right shows the difference between ComSol and the rooftop moment method.

Figure 2.11: Comparison of the of the y-component of the scattered field between 3 forward modelers. The top graph represents the magnitude and the bottom graph represents the phase. The data was measured on the the line x=0.3m, y=0m and z is the independent variable.

Figure 2.11 compares the y-component of the scattered field along the line where x=0.3m, y=0m and z is the independent variable. The bottom graph in figure 2.11 represents the phase and is meaningless in this case because of the small magnitude of the y-component of the scattered field. The graph at the top of figure 2.11 represents the magnitude of the y-component. From this graph the error associated with the ComSol modeler can be estimated. While the magnitude of the y-component of the scattered fields generated by the pulse moment method and the rooftop moment method have zero

magnitude, the y-component of the scattered field generated by ComSol does not. The ComSol results resemble that of a noise. If both moment methods are assumed to be correct, then the ComSol results can be used to estimate the error associated with ComSol modeler.

Figure 2.12: Comparison of the of the y-component of the scattered field between 3 forward modelers. The top graph represents the magnitude and the bottom graph represents the phase. The data was measured on the the line where x=0.3m and y=-z line.

Figure 2.12 compares the y-component of the scattered field along the line where x=0.3m and y=-z. The top graph of figure 2.12 shows the magnitude of the y-component of the scattered fields. The y-component of the scattered fields generated by the pulse moment method and ComSol have similar magnitudes. The y-component of the scattered field generated by the rooftop moment method has a slightly smaller magnitude.

38
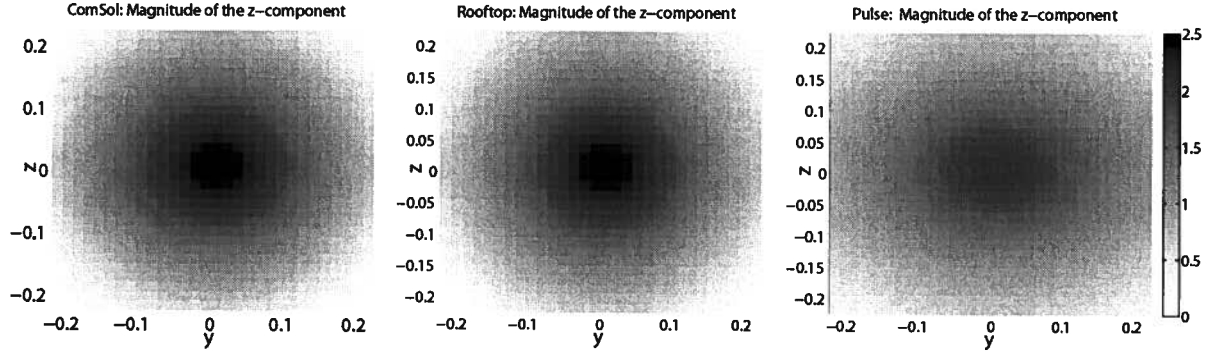
Figure 2.13: A comparison between the z-component of the scattered field from 3 forward modelers. The image on the left shows the result from ComSol. The image in the center shows the results from the moment method using rooftop basis functions. The image on the right shows the results from the moment method using pulse basis functions.

Figure 2.13 shows surface plots of the z-component of the simulated scattered field on the measurement plane for the three forward modelers. The plot on the left shows the z-component of the scattered field generated by ComSol. The plot in the center shows the z-component of the scattered field generated by the rooftop moment method. The plot on the right shows the z-component of the scattered field generated by the pulse moment method. All three plots contain a large dark spot in the center. The magnitude of the the z-component of scattered field generated by the pulse moment method is noticeably smaller.
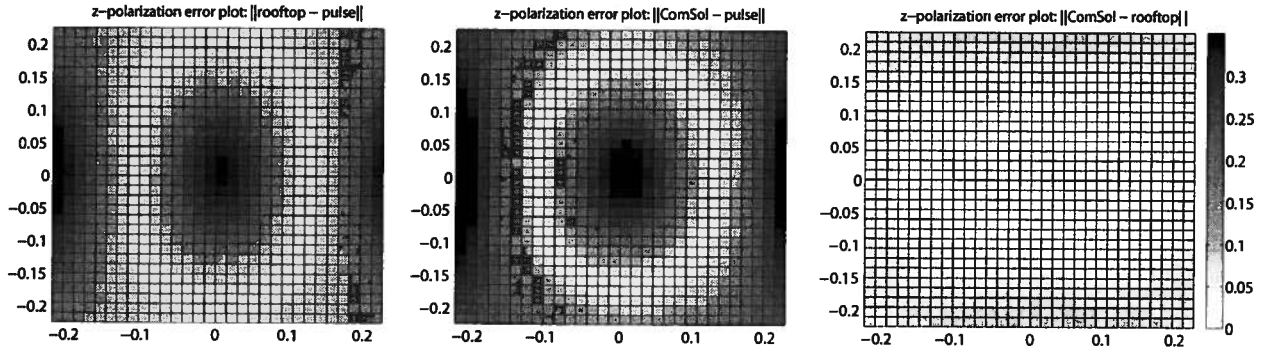
Figure 2.14: Surface plot of the difference between the z-component of the electric field for the three forward modelers. The surfaces plot the magnitude of the difference between the magnitude of the fields of the three forward modelers.

Figure 2.14 shows the differences between the magnitudes of the z-component of the electric fields from the three forward modelers. The plot on the left shows the difference between rooftop moment method and the pulse moment method. The plot in the center shows the difference between

39

ComSol and the pulse moment method. The plot on the right shows the difference between ComSol and the rooftop moment method.
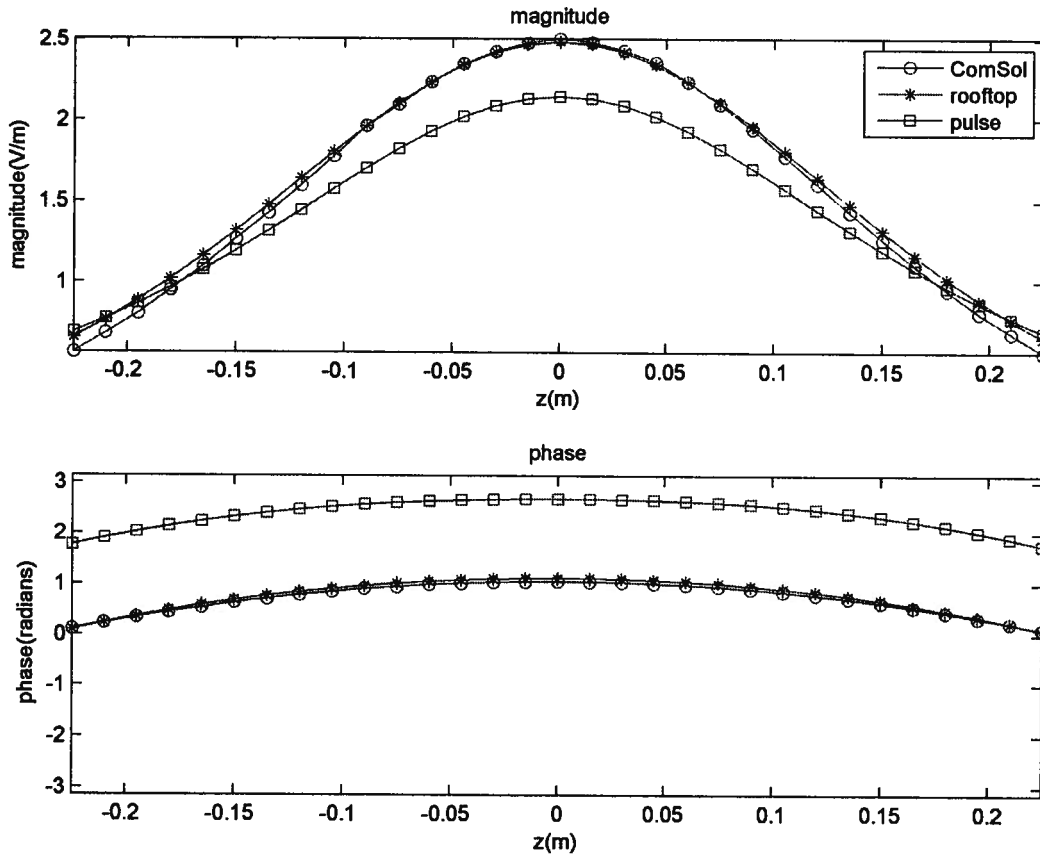
Figure 2.15: Comparison of the of the z-component of the scattered field between 3 forward modelers. The top graph represents the magnitude and the bottom graph represents the phase. The data was measured on the the line x=0.3m, y=0m and z is the independent variable.

Figure 2.15 compares the z-component of the scattered field along the line where x=0.3m, y=0m and z is the independent variable. The top graph of figure 2.15 shows the magnitude of the z-component of the scattered fields. The z-component of scattered fields generated by the rooftop moment method and ComSol have similar magnitudes. The z-component of scattered field generated by the pulse moment method has a slightly smaller magnitude. The phases of the z-component of the scattered fields generated by ComSol and the rooftop moment method show good agreement.
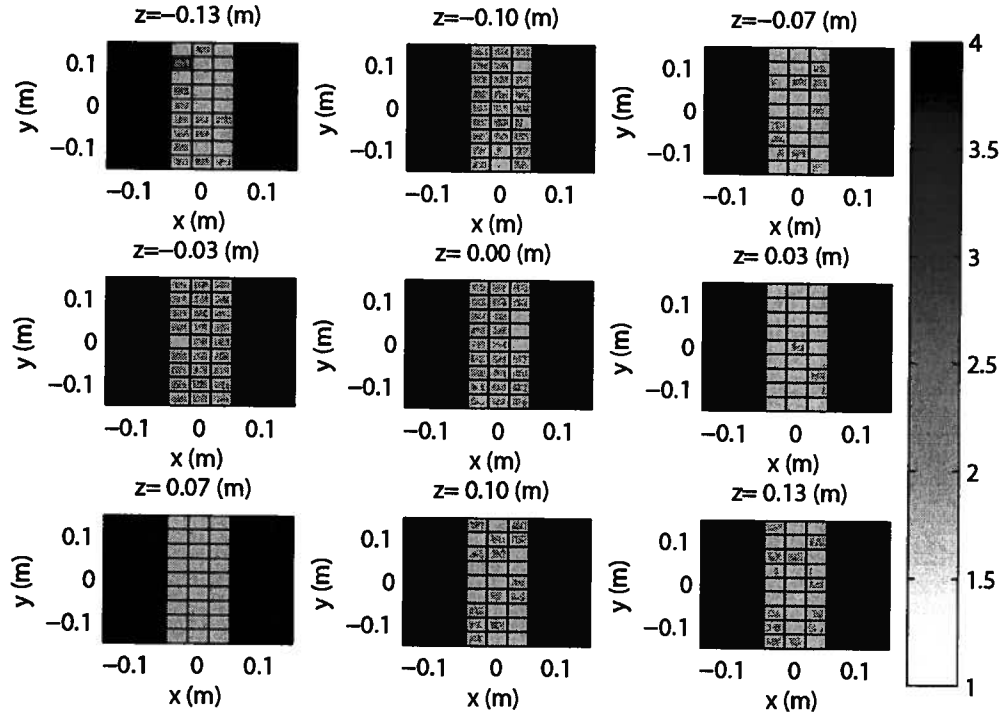
Figure 2.16: Relative permittivity distribution consisting of 3 layers. Each layer is 0.1m thick. The middle layer has a relative permittivity of 2.0. The two out layers have a relative permittivity of 4.0.

Figure 2.16 shows a second relative permittivity distribution that will be used to verify the rooftop moment method presented in this thesis. The excitation frequency is 1GHz. The freespace wavelength at 1GHZ, λ_0 , is 0.3m. The relative permittivity distribution will be illuminated by a planewave that is propagating in the x-direction and polarized in the z-direction. The data will be measured on the z-y plane at x=0.30m ($1\lambda_0$). The relative permittivity distribution consists of three layers. Each layer is 0.1m($0.33\lambda_0$) thick. The width is 0.3m($1\lambda_0$) and the height is 0.3m($1\lambda_0$). The two outer layers have a relative permittivity of 4.0. The middle layer has a relative permittivity of 2.0.

41
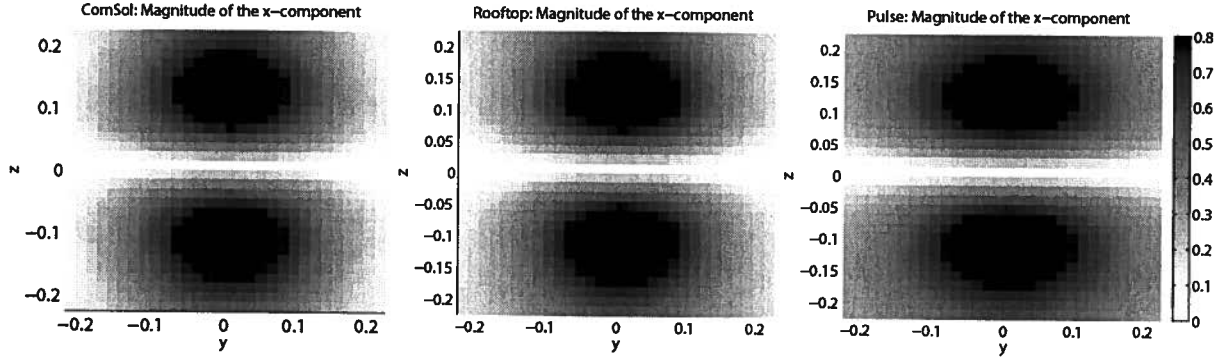
Figure 2.17: A comparison between the x-component of the scattered field from 3 forward modelers. The image on the left shows the result from ComSol. The image in the center shows the results from the moment method using rooftop basis functions. The image on the right shows the results from the moment method using pulse basis functions.

Figure 2.17 compares the magnitude of the x-component of the scattered field generated by the

three modelers. The plot on the left shows the x-component of the scattered field generated by

ComSol. The plot in the center shows the x-component of the scattered field generated by the rooftop

moment method. The plot on the right shows the x-component of the scattered field generated by the

pulse moment method. All three plots contain a large dark spot in the upper and lower halfs. The

magnitudes of the x-component of the scattered fields generated by the three forward modelers share

similar characteristics.
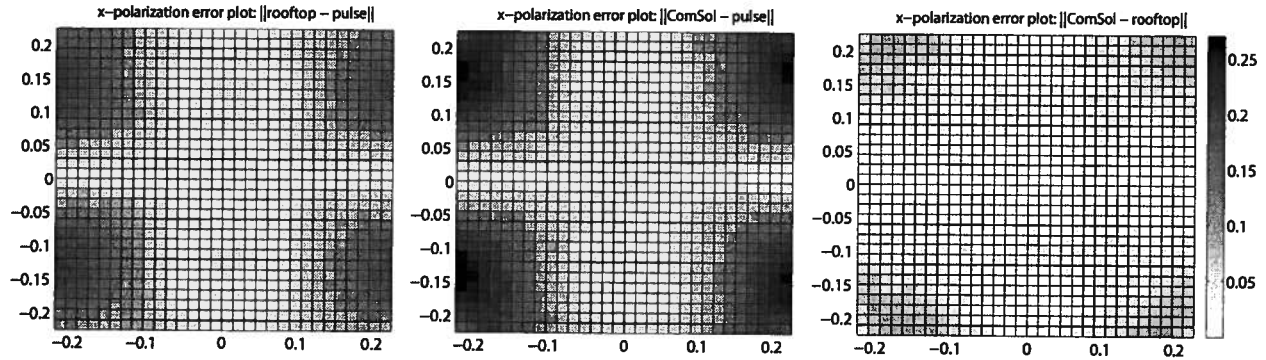
Figure 2.18: Surface plot of the difference between the x-component of the electric field for the three forward modelers. The surfaces plot the magnitude of the difference between the magnitude of the fields of the three forward modelers.

Figure 2.18 shows the differences between the magnitudes of the x-component of the electric

fields from the three forward modelers. The plot on the left shows the difference between rooftop

42

moment method and the pulse moment method. The plot in the center shows the difference between ComSol and the pulse moment method. The plot on the right shows the difference between ComSol and the rooftop moment method.
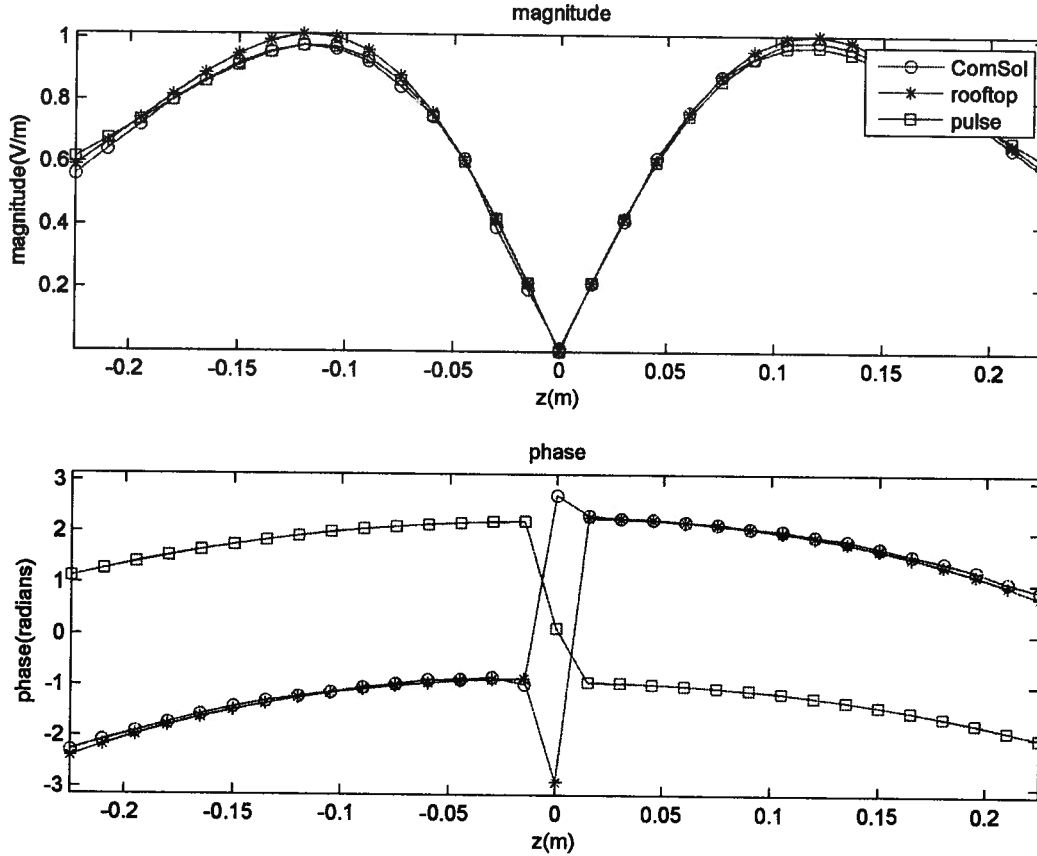
Figure 2.19: Comparison of the of the x-component of the scattered field between 3 forward modelers. The top graph represents the magnitude and the bottom graph represents the phase. The data was measured on the the line where x=0.3m, y=0m and z is the independent variable.

Figure 2.19 shows the comparison of the x-component of the scattered fields generated by the three forward modelers on the line where x=0.3m, y=0 and z is the independent variable. The x-component of the scattered field generated by all three modelers show good agreement for the magnitude plot, shown at the top of figure 2.19. The phases of the x-component of the scattered fields generated by ComSol and the rooftop moment method have good agreement.
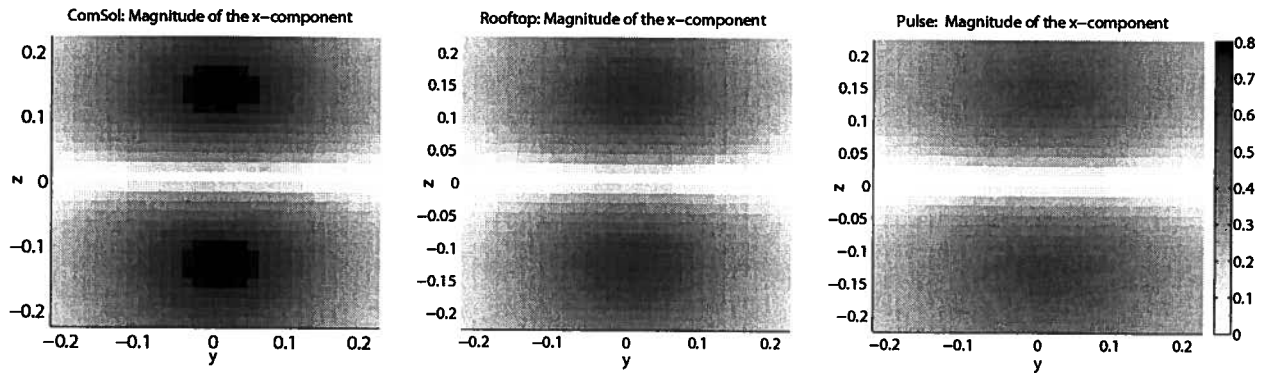
43
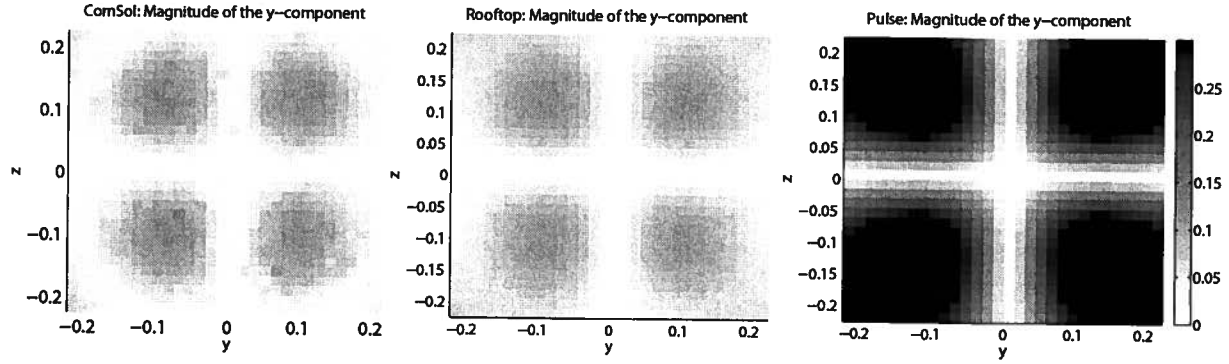
Figure 2.20: Comparison between the y-component of the scattered field from 3 forward modelers. The image on the left shows the result from ComSol. The image in the center shows the results from the moment method using rooftop basis functions. The image on the right shows the results from the moment method using pulse basis functions.

Figure 2.20 compares the magnitude of the y-component of the scattered fields generated by the three modelers. The plot on the left shows the y-component of the scattered field generated by ComSol. The plot in the center shows the y-component of the scattered field generated by the rooftop moment method. The plot on the right shows the y-component of the scattered field generated by the pulse moment method. The y-component of the scattered field generated by the pulse moment method has a much larger magnitude than those generated by the other two forward modelers.
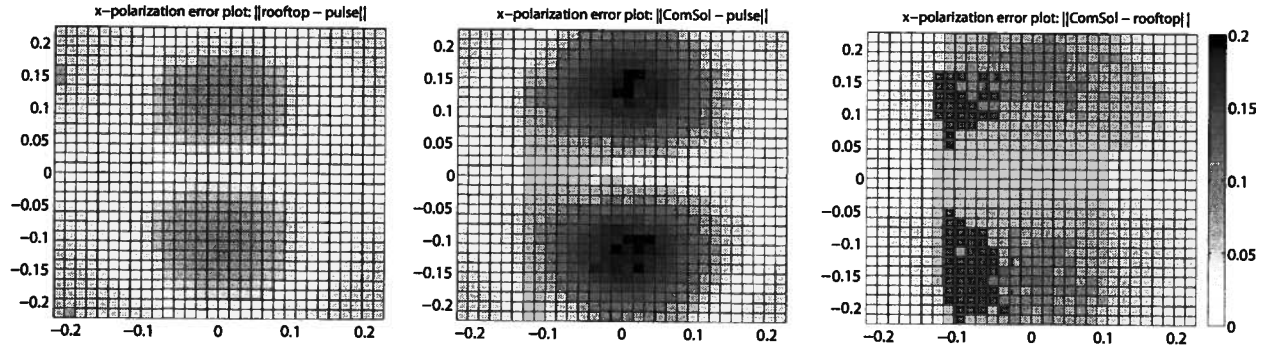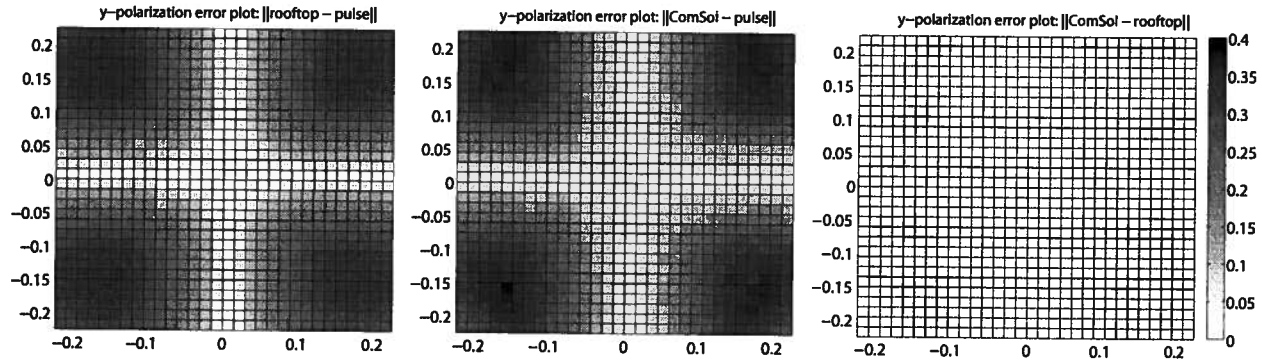

Figure 2.21: Surface plot of the difference between the y-component of the electric field for the three forward modelers. The surfaces plot the magnitude of the difference between the magnitude of the fields of the three forward modelers.

Figure 2.21 shows the differences between the magnitudes of the y-component of the electric fields from the three forward modelers. The plot on the left shows the difference between rooftop moment method and the pulse moment method. The plot in the center shows the difference between

44

ComSol and the pulse moment method. The plot on the right shows the difference between ComSol and the rooftop moment method.
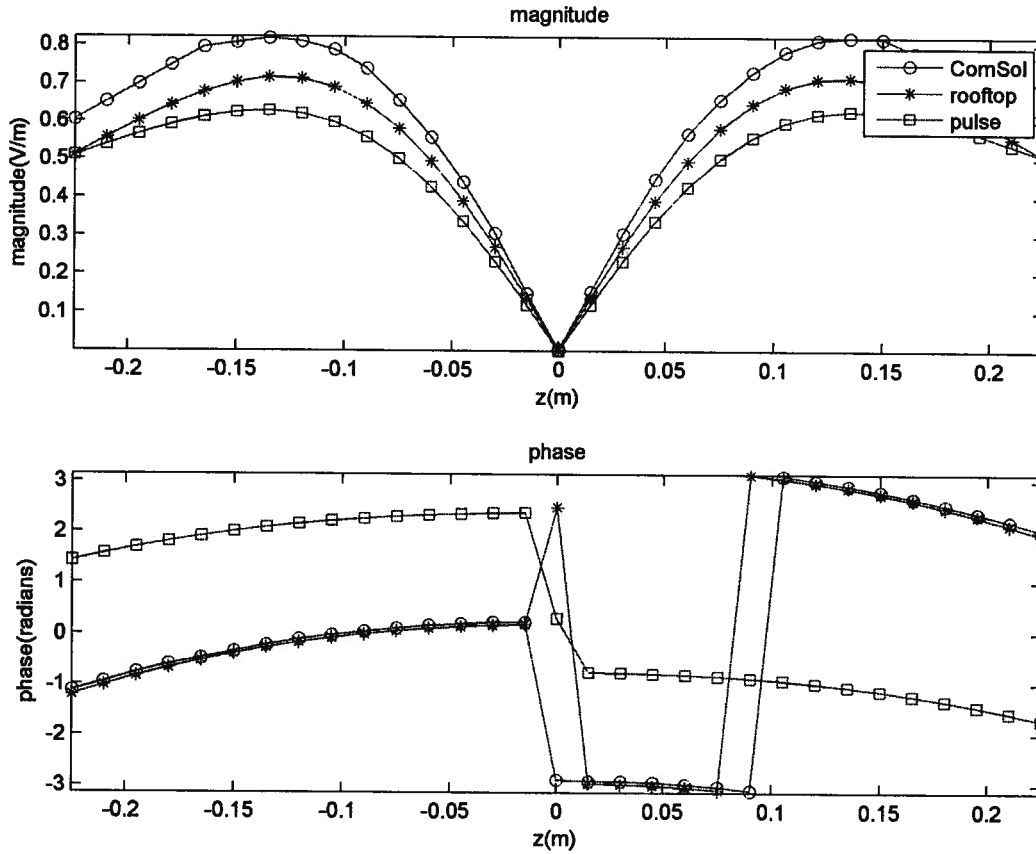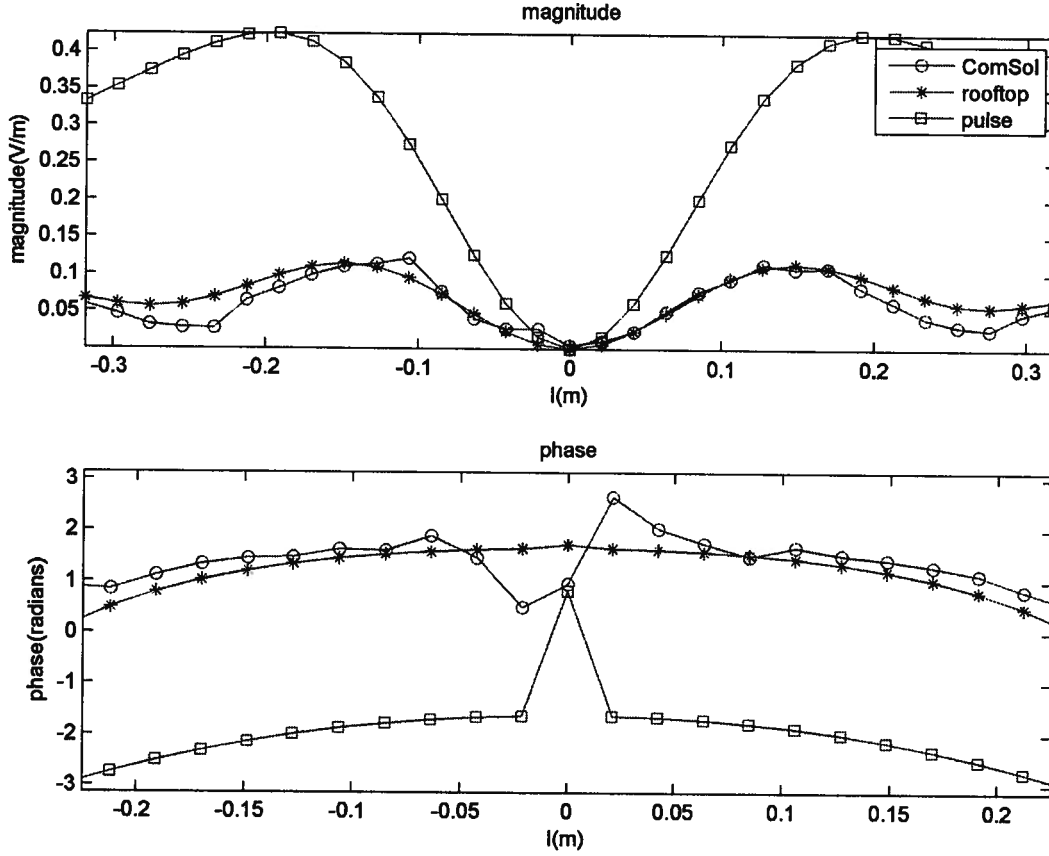

Figure 2.22: Comparison of the of the y-component of the scattered field between 3 forward modelers. The top graph represents the magnitude and the bottom graph represents the phase. The data was measured on the the line x=0.3m and y=-z line.

Figure 2.22 compares the y-component of the scattered fields generated by the three forward modelers. The data were measured on the line x=0.3m and y=-z line. The y-component of the scattered field generated by the pulse moment method has a much greater magnitude that those generated by the other two forward modelers. The y-component of the scattered fields generated by ComSol and the rooftop moment method are similar, agreeing in both magnitude and phase. Numerical error is apparent in the y-component of the scattered field generated by ComSol. At the point l=-0.1, where l is

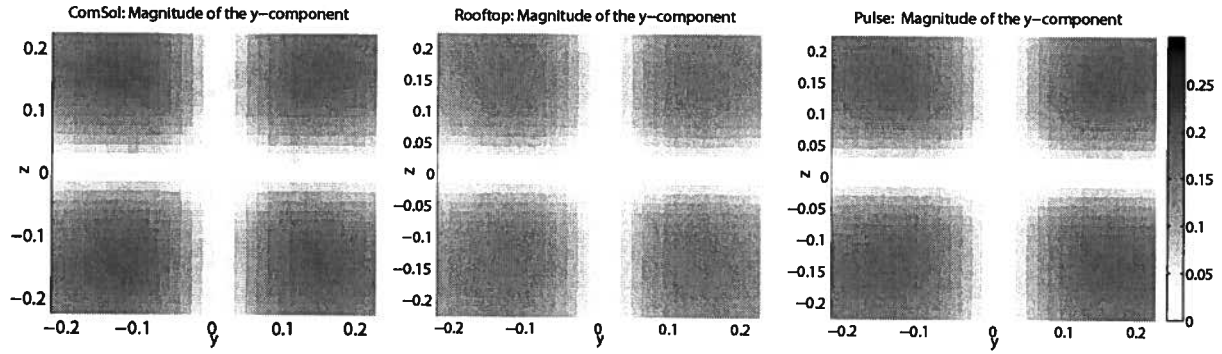the independent variable, the ComSol curve does not appear smooth.
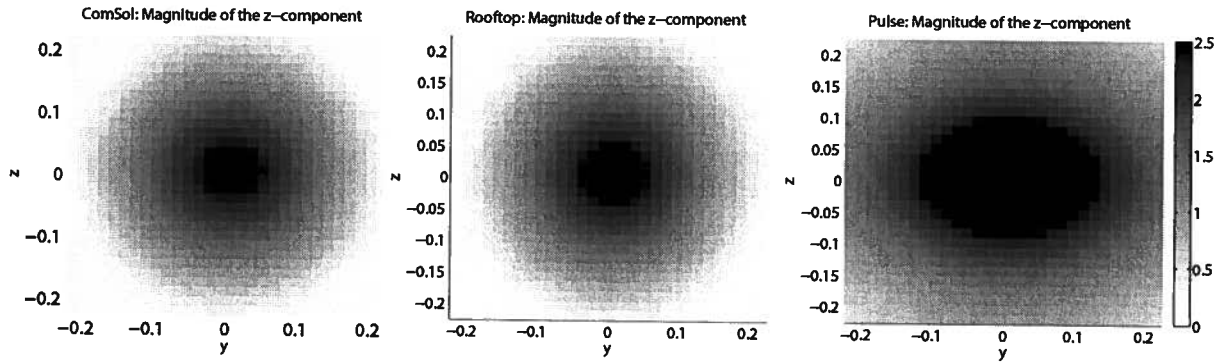
Figure 2.23: A comparison between the z-component of the scattered field from 3 forward modelers. The image on the left shows the result from ComSol. The image in the center shows the results from the moment method using rooftop basis functions. The image on the right shows the results from the moment method using pulse basis functions.

Figure 2.23 compares the magnitude of the z-component of the scattered field generated by the three modelers. The plot on the left shows the z-component of the scattered field generated by ComSol. The plot in the center shows the z-component of the scattered field generated by the rooftop moment method. The plot on the right shows the z-component of the scattered field generated by the pulse moment method. All three plots contain a large dark spot in center. The magnitudes of the z-component of the scattered fields generated by the three forward modelers share similar characteristics.
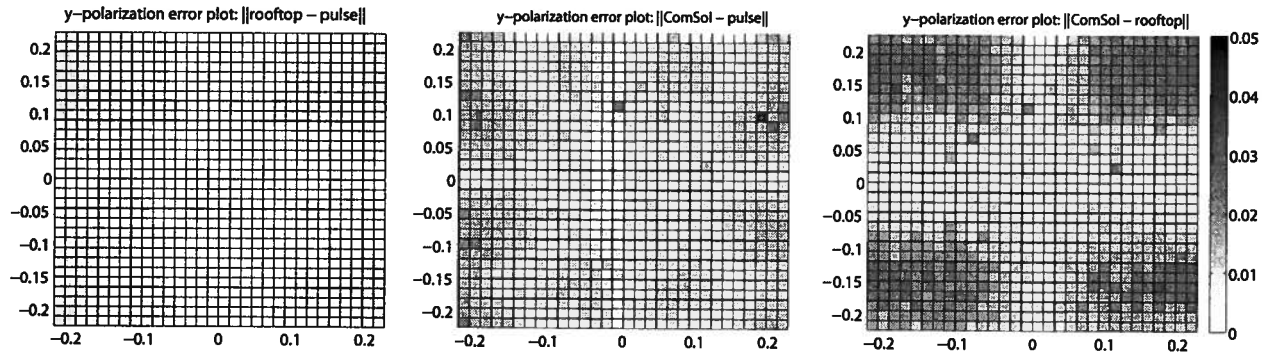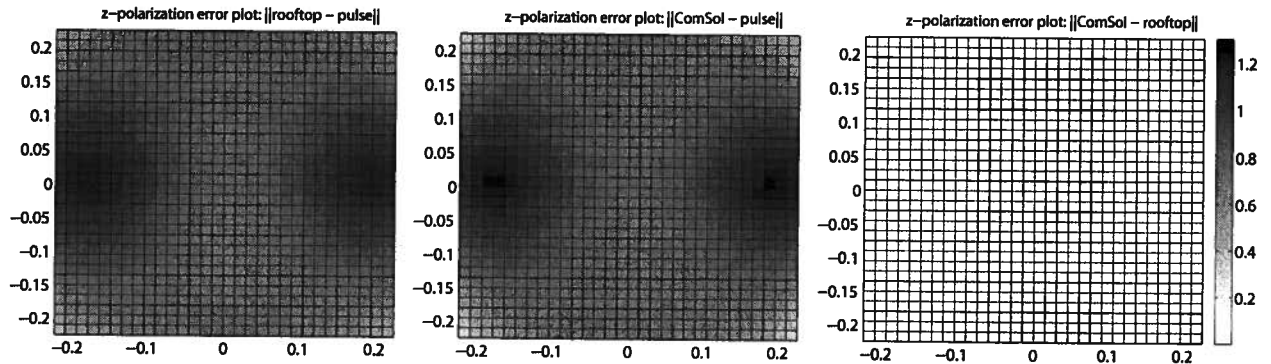

Figure 2.24: Surface plot of the difference between the z-component of the electric field for the three forward modelers. The surfaces plot the magnitude of the difference between the magnitude of the fields of the three forward modelers.

Figure 2.24 shows the differences between the magnitudes of the z-component of the electric fields from the three forward modelers. The plot on the left shows the difference between rooftop

46

moment method and the pulse moment method. The plot in the center shows the difference between

ComSol and the pulse moment method. The plot on the right shows the difference between ComSol
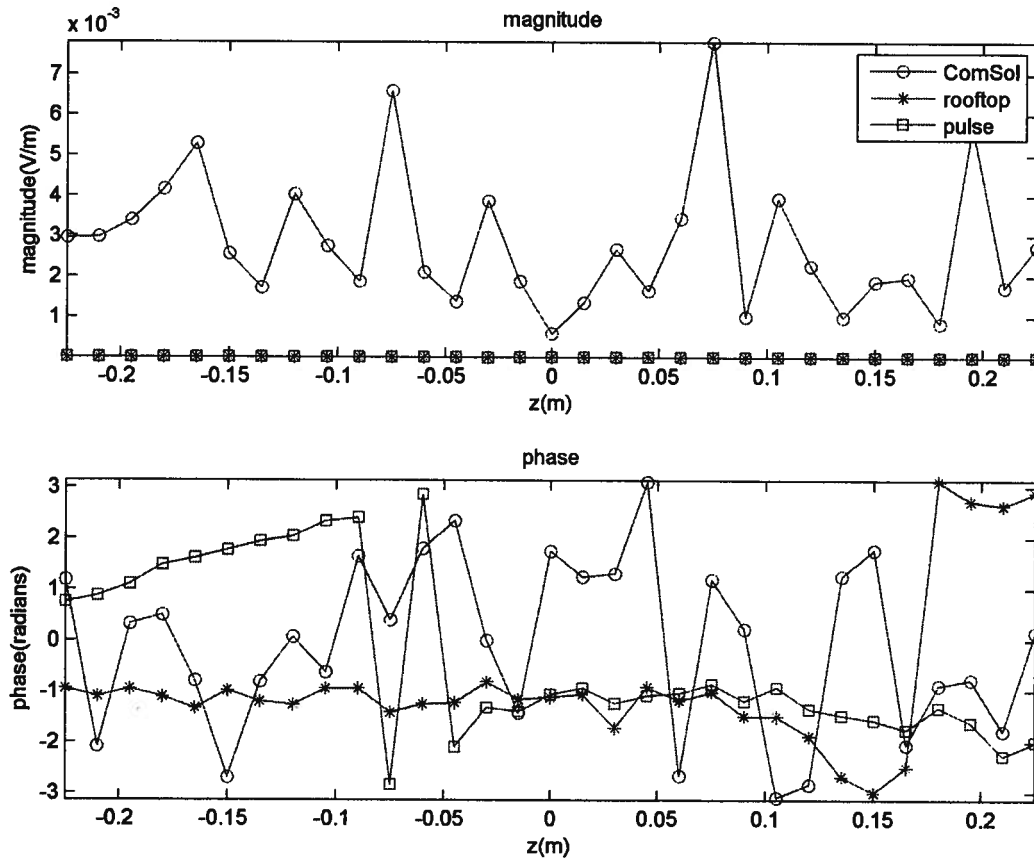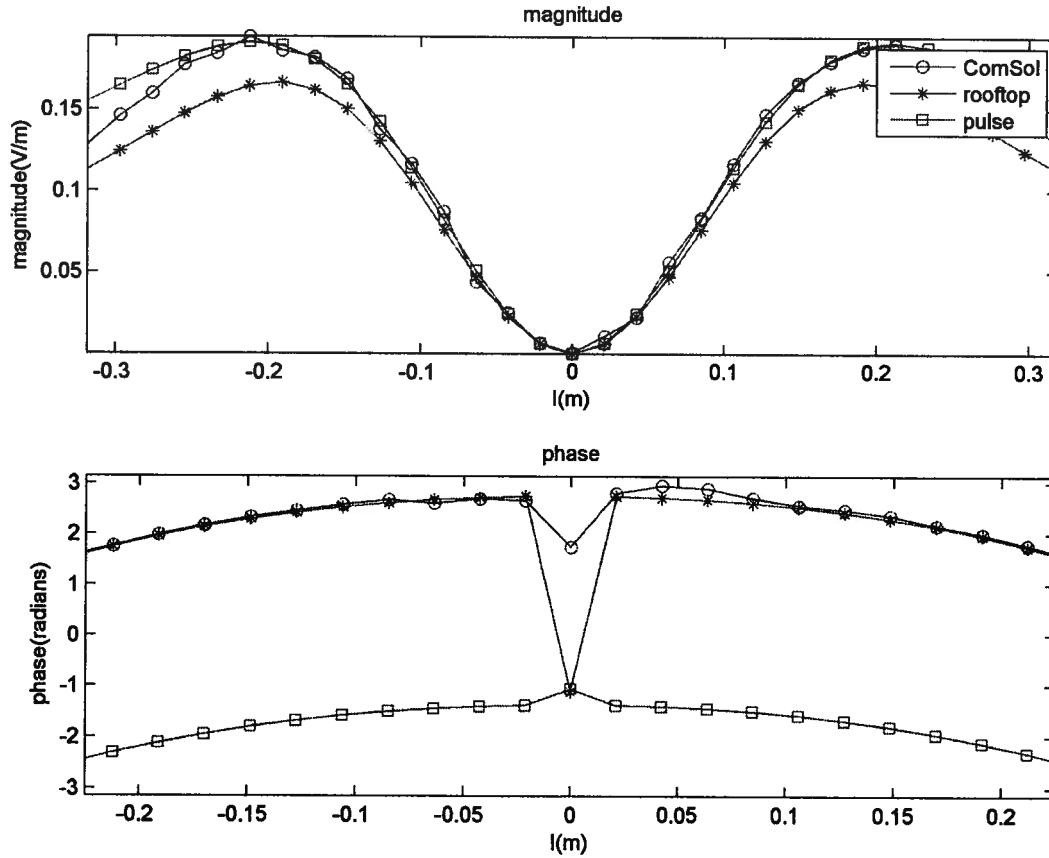
and the rooftop moment method.
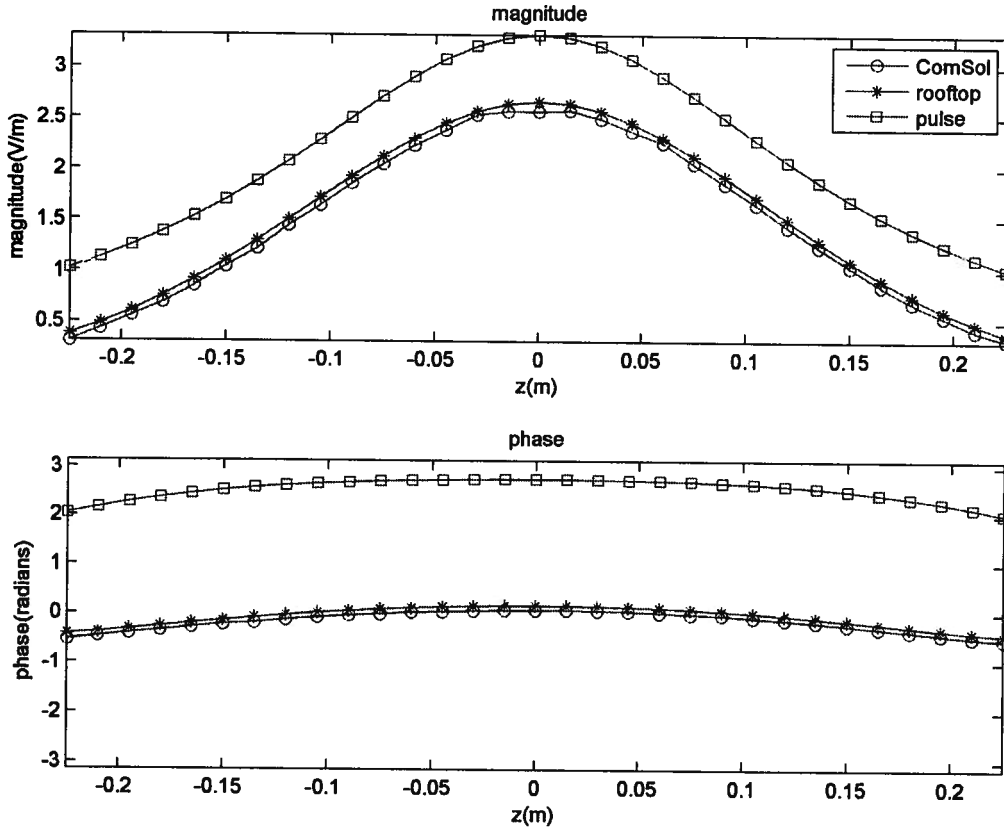
Figure 2.25: Comparison of the of the z-component of the scattered field between 3 forward modelers. The top graph represents the magnitude and the bottom graph represents the phase. The data was measured on the the line where x=0.3m, y=0m and z is the independent variable.

Figure 2.25 compares the z-component of the scattered field along the line where x=0.3m,

y=0m and z is the independent variable. The z-component of the scattered fields generated by ComSol

and the rooftop moment method show good agreement. The z-component of scattered field generated

by the pulse moment method has a larger greatest magnitude.

The scattered fields generated by ComSol and the rooftop moment method for the two relative

permittivity distributions show good agreement. For the relative permittivity distribution with 3 layers,

the scattered field generated by the pulse moment method differed from that of the scattered fields generated by the other two methods.

2.6 Conclusion

A moment method forward modeler using cube shape functions and rooftop basis function is presented. The formulation enforces the continuity of the displacement field across dielectric boundaries. In addition to the internal fields an extension to the external fields are also computed, a novel result in this thesis. The derivative calculations for the external and internal fields are also presented.

To verify that the formulation presented in the chapter, it was compared with the commercially available multiphysics solver, ComSol. The solution from the moment method using rooftop basis functions shows good agreement with ComSol and in the cases of modeling small signals, the moment method using rooftop basis function produces better results.

Chapter 3:

Least square minimization techniques

3.1 Introduction to chapter 3

The problem of recovering the permittivity structure of an unknown dielectric object can be cast as an unconstrained non-linear optimization problem. Attempts have been made to cast the problem as a linear problem using the Born approximation (Chew et al., 1994). However, the scattering from a dielectric object on the length scale of a wavelength does not satisfy the linear approximation and has yielded poor results.

To find the unknown permittivity distribution a functional is constructed based on observed data that has been measured or simulated, and predicted data, based on an estimate of the structure. The difference between the observed data and the predicted data, subject to the estimate of the structure, is minimized.

In this chapter, several least square minimization techniques will be presented and their behaviors will be examined. Other authors have explored the recovery of the permittivity structure, with the focus on the 2-D structure using non-linear optimization techniques. Franchois et al. (1998) uses the Levenberg-Marquardt method to recover high permittivity objects, submerged in a high permittivity background. Her work explores 2-D recovery, but also touches on small scale 3-D recovery.

Souvorov et al. (1998) uses a regularized Gauss-Newton method to recover a 2-D object for a submerged object. He was able to recover the permittivity distribution of a cylinder with a large

number of unknowns. Chew et al. (1995) introduces multiple frequencies in this kind of calculation using a Newton type minimization. Tanaka et al. (2001) applies the Levenberg-Marquardt to a 2-D cylinder.

With such a diversity of techniques that have demonstrated success in the past it would seem that the problem would be easy. Although these techniques have worked in 2-D, there is no clear indication that they will work in the 3-D case. The only way to ascertain a suitable method and provide a fair basis of comparison is to utilize a number of techniques on the recovery problem. The use of optimization techniques to recover structure is motivated by applications outside the field of electrical engineering. The field of geophysics has yielded many useful techniques for a slightly different problems. Jackson (1979) states that the regularization should be more than mathematical in nature, especially in the case of under-determined problems. In the under-determined case, the regularization chooses one solution out of an infinite set of possible solution. Ideally, the regularization will choose a solution which has features that are in the real target model. The L_2 norm technique arises naturally by first linearizing the mapping between the dielectric structure and the scattered field, and then squaring the difference between the measured data and the data that is predicted by the linearized mapping. This method also has the benefit of having known algorithms to determine a solution.

Although, the techniques presented have been discussed in previous texts, all of the algorithms presented have independent parameters that must be adjusted for the current inversion problem. It is necessary to define the behavior of these independent parameters before a successful inversion technique can be achieved. The importance of the selection and amount of data will be deferred to the next chapter. Even the best minimization algorithm can not recover the correct solution if insufficient data are present.

One major difficulty in inverse problems is the inability to determine whether a global

minimum has been reached, and another major difficulty is the possibility of a non-unique solution. Because of this, any solution that has a sufficiently small misfit value is an acceptable solution, although it may not represent the actual permittivity structure. Sometimes this can be caused by lack of data, while at other times it may be inherent in the problem (Lam at al., 2005).

The rest of the chapter will present the gradient method, Newton method, Gauss-Newton method, and the Levenberg-Marquardt method, L_2 cooled-roughness method, and the L_1 cooled-roughness method. To handle the L_1 norm the iteratively re-weighted least squares algorithm is presented.

3.2 The sum of squares

The techniques presented will be based on a sum of squares functional,

$$\Phi^d = \frac{1}{2}\sum_{f=1}^{N_f}\sum_{n=1}^{M_f}\|d_{fn}^{ref} - d_{fn}^{pred}(\vec{\varepsilon})\|^2 \quad . \tag{3.1}$$

Rewritten in a more compact fashion,

$$\Phi^d = \frac{1}{2}\sum_{f=1}^{N_f}\sum_{n=1}^{M_f}\|\delta d_f(\vec{\varepsilon})\|^2 \quad , \tag{3.2}$$

where

$$\delta d_{fn}(\vec{\varepsilon}) = d_{fn}^{ref} - d_{fn}^{pred}(\vec{\varepsilon}) \tag{3.3}$$

is a quantification of the difference between the field that has been measured or simulated, and the one predicted by $\vec{\varepsilon}$, where $\vec{\varepsilon}$ is a vector representing the permittivity distribution. The sum of squares has the benefit that the sign of (3.3) is disregarded, implying that only the magnitude of the error contribute to the functional. If the direction of the error is used, this may result in two or more errors canceling each other out in the functional.

Data from multiple observations can be easily included in the inversion scheme. The constant N_f refers to the number of observation sets. The subscript f denotes the particular data set. Each data set has M_f data points and the subscript n refers to the n'th data point within the data set. The superscript d on Φ^d indicate that this is the least square summation associated with the data. Other terms related to the regularization will be introduced later. The reference data is denoted by d_{fn}^{ref} and the data predicted by the permittivity distribution, $\vec{\varepsilon}$, is given by d_{fn}^{pred}. The summation represents a measure of difference between the reference data and the data predicted by

$\vec{\varepsilon}$. The values of d_{fn}^{ref} and d_{fn}^{pred} are complex. The factor $\dfrac{1}{2}$ is introduced for mathematical convenience.

The least square summation is more commonly stated as (3.2) in optimization texts. The summation associated with N_f in (3.2) omitted in other texts. However it is convenient to include it when solving inverse problems with multiple observations sets, because it illustrates the relationship between the different observation set in the minimization. Measurements in an observation set share the same incident field and frequency.

3.3 The gradient method:

The gradient method (Gill at al., 1986), also known as the method of steepest descent, is an

iterative method employed to find the minimum of a functional by searching along the negative

gradient direction.
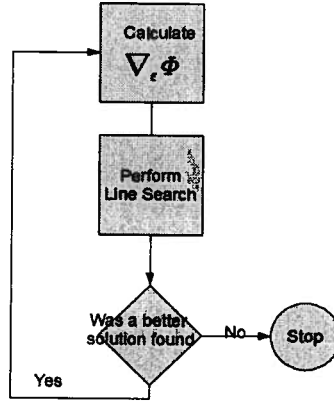
Figure 3.1: Flow chart for the gradient descent method.

Figure 3.1 is a diagrammatic representation of the gradient method. The gradient is computed

and the perturbation is determined from the gradient The gradient method is relatively simple in

concept. It uses the fact that the gradient direction is defined as the direction of greatest increase and

negative gradient direction is the direction of the steepest decrease. During each iteration, the

permittivity for the next iteration is determined by

$$\vec{\varepsilon}_{k+1} = \vec{\varepsilon}_k - \alpha_k \nabla_\varepsilon \Phi^d (\vec{\varepsilon}_k) \ .$$ (3.4)

The subscript k refers to the iteration number. The equation consists of two parts, the current

permittivity, $\vec{\varepsilon}_k$, and the perturbation to the permittivity. The perturbation has two parts, the

gradient and a constant α_k which weights the gradient. The magnitude of the gradient multiplied by

α_k is the step length, and the gradient normalized to its magnitude is the search direction for the

gradient method. Writing the perturbation in the form of (3.4) does not require the explicit

normalization of the gradient. The gradient for (3.1) is

$$\nabla_\varepsilon \Phi^d(\vec{\varepsilon}_k) = \sum_{f=1}^{N} \Re\{J_f^H(\vec{\varepsilon}_k)\delta d_f(\vec{\varepsilon}_k)\} \quad ,$$ (3.5)

where $J_f(\vec{\varepsilon}_k)$ represents the Jacobian matrix for an observation set and $J_f^H(\vec{\varepsilon}_k)$ is the Hermitian transpose of the Jacobian and $\Re$ is the real value operator. The Jacobian matrix calculation is demonstrated in chapter 2 and represents the change in the scattered field relative to a small change in permittivity. The gradient of the least square functional is simply summation of these individual changes.

3.3.1 The step length

After the gradient direction has been determined, the step-length needs to be determined.

The step length is the magnitude of the gradient multiplied by the α_k , so it's sufficient to determine α_k in this case. For the linear problem α_k can be calculated from (Pratt et al. 1998)

$$\alpha_k = \frac{\|\nabla_\varepsilon \Phi^d(\vec{\varepsilon}_k)\|^2}{\sum_{f=1}^{N_f} \|J_f^H(\vec{\varepsilon}_k)\nabla_\varepsilon \Phi^d(\vec{\varepsilon}_k)\|^2} \quad .$$ (3.6)

For non-linear problems the constant α_k must be determined by a line search, but (3.6) can be used as a starting point. The line search can be approximate or exact. For this work, an approximate line search centered around the linear solution is used because a forward problem must be performed for each search point and that represents a large computational cost.

3.3.2 Calculation of the gradient without the explicit calculation of the Jacobian

In chapter 2, the calculation of the Jacobian is shown. The gradient can be calculated using the Jacobian shown in (3.5). Pratt et al. (1998) has shown an alternative method for calculating the

gradient, without the explicit calculation of the Jacobian. His method represents a significant computational savings in cases where the sensitivities are not readily available. In order for it to be applied to the scattered field it is extended, one of the novel contributions contained in this thesis. The arguments below are an extension of Pratt's work to the case of electromagnetic scattering. Pratt demonstrated that only one right hand side needs to be solved to calculate the gradient and the Jacobian does not need to be calculated explicitly.

For the electromagnetic scattering, where the observation points are known before hand, the data for a single frequency and excitation combination can be written as the product of the transmission matrix multiplied by the field internal to the dielectric object,

$$\vec{d}_f^{pred} = T_f \vec{D}_f \ , \tag{3.7}$$

or more explicitly

$$\vec{d}_f^{pred} = T_f S_f^{-1} \vec{E}_f^i \ , \tag{3.8}$$

a function of the incident field and inverse of the forward modeling matrix. The symbol $\vec{D}_f$ represents the displacement field and $\vec{d}_f^{pred}$ represents the scattered field. The symbol $\vec{E}_f^i$ represents the incident field tested against the basis function using equation (2.13). The term T_f represents the transmission matrix that maps internal fields to a set of observations points that have been predefined. The matrix S_f is the forward modeling matrix as defined by chapter 2.

Written in summation component fashion, (3.8) becomes

$$d_{fm}^{pred} = \sum_{n=1}^{M_f} T_{fmn} D_{fn} \ , \tag{3.9}$$

where m is the row index representing the destination location and n is the column index representing the radiation source location. The constant M_f represents the number of observation points.

The summation form is used for equation (3.8) for ease of analysis. Each data element in the scattered field, therefore consists of contributions from each internal element. The transmission coefficient represents the weighting of the internal elements, based partly on the distance to the observation point, and partly on the permittivity.

The derivative for a single data element is given by

$$\frac{\partial d^{pred}_{fm}}{\partial \varepsilon_p} = \sum_{n=1}^{M_f} \frac{\partial T_{fmn}}{\partial \varepsilon_p} D_{fn} + \sum_{n=1}^{M_f} T_{fmn} \frac{\partial D_{fn}}{\partial \varepsilon_p} \quad . \tag{3.10}$$

The derivative can be found, through the simple application of the product rule. The derivative consists of two terms, one portion consists of the known derivative of the transmission terms and the displacement field, the second term consists of the known transmission coefficient and the unknown derivative of the internal field. Pratt's work eliminates the need to determine the unknown derivative of the internal field explicitly.

The assembly of $\dfrac{\partial d^{pred}_{fm}}{\partial \varepsilon_p}$ is denoted at J_f . Rewriting (3.10) in matrix form results in

$$J_f = C_f + T_f [\frac{\partial \vec{D}_f}{\partial \varepsilon}] \quad , \tag{3.11}$$

where C_f represents the first term in (3.10) consolidated in matrix form,

$$C_f = \begin{bmatrix} \sum_{n=1}^{M_f} \frac{\partial T_{f1n}}{\partial \varepsilon_1} D_{fn} & \cdots & \sum_{n=1}^{M_f} \frac{\partial T_{fM_fn}}{\partial \varepsilon_1} D_{fn} \\ \vdots & \ddots & \vdots \\ \sum_{n=1}^{M_f} \frac{\partial T_{f1n}}{\partial \varepsilon_M} D_{fn} & \cdots & \sum_{n=1}^{M_f} \frac{\partial T_{fM_fn}}{\partial \varepsilon_M} D_{fn} \end{bmatrix} \quad . \tag{3.12}$$

The Jacobian written explicitly is given by

$$J_f = C_f + T_f S_f^{-1} F_f \quad . \tag{3.13}$$

If $S_f^{-1} F_f$ were evaluated explicitly M_f right hand sides would need to be solved.

$$F_f = \begin{bmatrix} \sum_{n=1}^{N} \dfrac{\partial S_{f1n}}{\partial \varepsilon_1} D_{fn} & \cdots & \sum_{n=1}^{N} \dfrac{\partial S_{fNn}}{\partial \varepsilon_1} D_{fn} \\ \vdots & \ddots & \vdots \\ \sum_{n=1}^{N} \dfrac{\partial S_{f1n}}{\partial \varepsilon_M} D_{fn} & \cdots & \sum_{n=1}^{N} \dfrac{\partial S_{fNn}}{\partial \varepsilon_M} D_{fn} \end{bmatrix} \tag{3.14}$$

Plugging the new expression for the Jacobian,(3.13), into the expression for the gradient, (3.5), results in

$$\nabla_\varepsilon \Phi = \frac{1}{2} \sum_{f=1}^{N_f} \Re \{ \left[\left(C_f + T_f S_f^{-1} F_f \right)^T \left(\vec{d}_f^{ref} - \vec{d}_f^{pred} \right)^* \right] \} , \tag{3.15}$$

an expression that does not involve the Jacobian explicitly.

Evaluating the transpose operator in (3.15) results in

$$\nabla_\varepsilon \Phi = \frac{1}{2} \sum_{f=1}^{N_f} \Re \{ \left[\left(C_f^T + F_f^T (S_f^{-1})^T T_f^T \right) \left(\vec{d}_f^{ref} - \vec{d}_f^{pred} \right)^* \right] \} . \tag{3.16}$$

The back propagation vector, $\vec{v}_f$, is defined

$$\vec{v}_f = (S_f^{-1})^T T_f^T \left(\vec{d}_f^{ref} - \vec{d}_f^{pred} \right)^* . \tag{3.17}$$

Using the identity $\left(S^{-1} \right)^T = S^{-1}$ for a symmetric matrix (3.17) is written

$$\vec{v}_f = S_f^{-1} T_f^T \left(\vec{d}_f^{ref} - \vec{d}_f^{pred} \right)^* \tag{3.18}$$

and can be solved as a linear system. The back propagation vector can be substituted into (3.16), an expression for the gradient, thus reducing the number of linear systems to be solved to calculate the gradient. A total of $3 N_f$ linear systems needed to be solved for the linear gradient descent case. One set for the forward problem, one set for the gradient and one set for the linear step length. Without the speed up, $M_f N_f$ linear system needs to be solved.

This provides a small amount of computational savings, because the an LU decomposition

solution is approximately $O(M_f^3)$ in terms of floating point operations and solving M_f right hand sides is also $O(M_f^3)$. Therefore, by using the speed up proposed, the computation time can be reduced by a constant factor. The exact factor depends on the efficiency of the LU solver and the upper and lower triangular solvers.

3.4 The Newton method

The Newton method is a second derivative method because, it uses second derivative information. It is an iterative method based on the second order approximation of a functional.

The second order Taylor series expansion of a general functional with respect to a small perturbation $\vec{\varepsilon}_k$ can be written as,

$$F\left(\vec{\varepsilon}_k + \Delta \vec{\varepsilon}_k\right) = F(\vec{\varepsilon}_k) + \vec{g}_k^T \Delta \vec{\varepsilon}_k + \frac{1}{2} \Delta \vec{\varepsilon}_k^T G_k \Delta \vec{\varepsilon}_k + O\left(\|\vec{\varepsilon}_k\|^3\right) \quad . \tag{3.19}$$

The expansion consists of the value of the functional at the expansion point, the gradient vector multiplied by a small perturbation to the expansion point, and the Hessian is multiplied by the equivalent of the square of the perturbation. Once again, as in the gradient method, the subscript k represents the iteration number.

The direction of the update vector is based both on the gradient and the Hessian. The solution takes advantage of the fact that the derivative at the minimum is equal to zero.

The permittivity at the next iteration is given by

$$\vec{\varepsilon}_{k+1} = \vec{\varepsilon}_k + \Delta \vec{\varepsilon}_k \quad . \tag{3.20}$$

The gradient of (3.19) can be written as,

$$\nabla_\varepsilon F\left(\Delta \vec{\varepsilon}_k\right) = g_k + G_k \Delta \vec{\varepsilon}_k \quad , \tag{3.21}$$

which is simply the first order Taylor expansion of the gradient.

Using the fact that the gradient is zero at the minimum the solution for $\Delta \vec{\varepsilon}_k$ can be found by solving

$$-g_k = G_k \Delta \vec{\varepsilon}_k \quad . \tag{3.22}$$

When applied to the least square functional, the gradient, omitting the iteration index , can be written as

$$\sum_{f=1}^{N_f} \Re\{J_f^H \delta d_f\} \quad,$$

(3.23)

or it can be written as

$$\sum_{f=1}^{N_f} \Re\{J_f^T \delta \vec{d}_f^*\} \quad.$$

(3.24)

The gradient for each observation set is simply a multiplication of the Jacobian with the data, and the gradient of each observation set is simply summed to produce the gradient for the functional.

The Hessian can be written

$$G = \sum_{f=1}^{N_f} \Re\{J_f^H J_f + Q_f\} \quad,$$

(3.25)

It is the sum of first derivative information, $\displaystyle\sum_{f=1}^{N_f} \Re\{J_f^H J_f\}$, and second derivative information Q_f . The tensor product Q_f contains the second derivative information.

Figure 3.2 illustrates an iterative process for the Newton method.

61
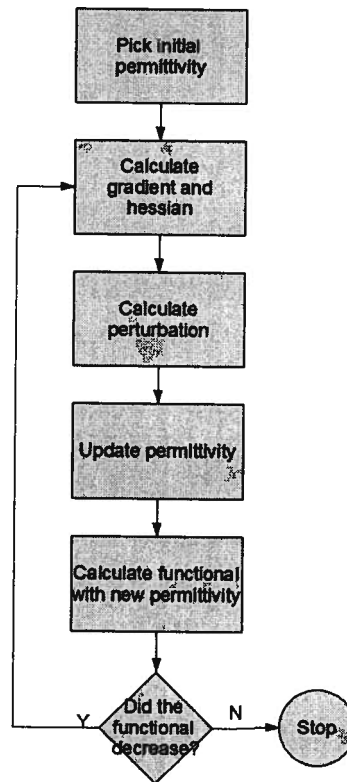
Figure 3.2: Flow chart of the Newton method

The algorithm can be stated as follows:

1) Pick an initial permittivity

2) Calculate the gradient and the Hessian

3) Calculate the perturbation based on (3.22)

4) Update the functional with the new permittivity using (3.20)

5) Calculate the functional value, (3.19)

6) If the functional value decreases go to step 2 to start the next iteration

7) stop

3.5 The Gauss-Newton method

The Gauss-Newton method is an approximation to the Newton method, for sum of square problems. The calculation of the Q_f portion of the Hessian matrix can be time consuming. The

Gauss Newton method is an approximation to Newton-Method using $\sum_{f=1}^{N_f} \Re\{J_f^H J_f\}$ to approximate

the Hessian matrix, G_k in (3.22).

This results in the equation

$$\sum_{f=1}^{N_f} \Re\{J_f^H J_f\} \Delta \vec{\varepsilon} = -\Re\{J_f^H \delta \vec{d}_f^{\,*}\} \quad , \tag{3.26}$$

which can be solved to determine the perturbation to the permittivity. The Gauss Newton method is equivalent to minimizing

$$\frac{1}{2} \sum_{f=1}^{N_f} \|J_f^H \Delta \vec{\varepsilon} - \delta \vec{d}_f\|^2 \quad . \tag{3.27}$$

The Gauss Newton may fail when $\sum_{f=1}^{N_f} \Re\{J_f^H J_f\}$ is badly conditioned or does not have full rank.
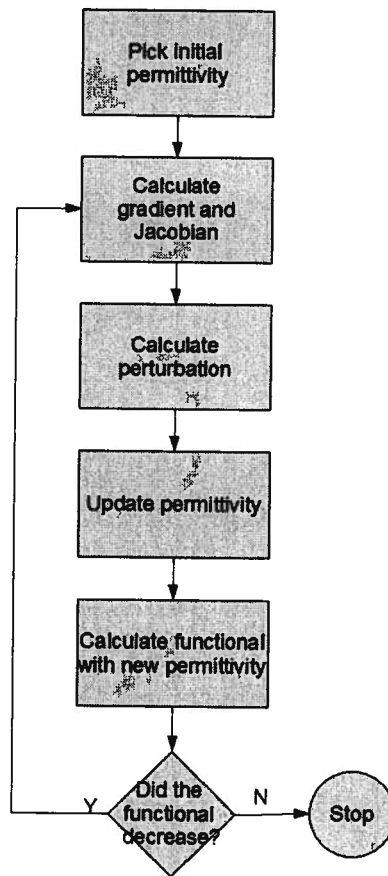
Figure 3.3: Diagrammatic representation of the Gauss-Newton method.

Figure 3.3 is a flow chart of the Gauss-Newton Method.

The algorithm can be stated as follows:

1. Pick an initial permittivity

2. Calculate the gradient and the Jacobian

3. Calculate the perturbation based on (3.26)

4. Update the functional with the new permittivity using (3.20)

5. Calculate the functional value, (3.19)

6. If the functional value decreases go to step 2 to start the next iteration

7. stop

3.6 The method of successive linearization

For the method of successive linearization, $\left(\vec{d}_f^{ref} - \vec{d}_f^{pred}\right)$ is linearized at each iteration using a first order Taylor expansion. The second derivative terms are ignored.

An additional term is added to (3.1) resulting in

$$\Phi = \frac{1}{2}\sum_{f=1}^{M_f}\|\left(\vec{d}_f^{ref} - \vec{d}_f^{pred}\right)\|^2 + \frac{1}{2}\beta\|F\left(\vec{\varepsilon}^{ref} - \vec{\varepsilon}\right)\|^2 \quad , \tag{3.28}$$

to produce a stable system for this algorithm.

The second term,

$$\Phi^r = \beta\sum_{m=1}^{M}\frac{1}{2}\|\sum_{j=1}^{M}F_{mj}\left(\varepsilon_j^{ref} - \varepsilon_j\right)\|^2 \quad , \tag{3.29}$$

is a regularization term and can be chosen with some physical meaning. The parameter β is determined later, but determines the weighting between the regularization and data. The operator F is an operator on the permittivity distribution. Several choices can be chosen for F and will be discussed in the next section. Later, in the Levenberg-Marquardt section method an operator on the perturbation will be introduced. The regularization can be seen as defining a feasible region in which the correct solution exists. In the successive linearization method the region is centered on $\vec{\varepsilon}^{ref}$.

Successive linearization is an iterative technique and the subscript k is introduced to denote this. For successive linearization two steps are taken. First, a linear approximation of the forward modeling problem is applied to (3.28) to obtain

$$\Phi_k = \frac{1}{2}\sum_{f=1}^{M_f}\|\left(\vec{d}_f^{ref} - \vec{d}_f^{pred}(\vec{\varepsilon}_k)\right) - J_f\Delta\vec{\varepsilon}_k\|^2 + \frac{1}{2}\beta_k\|F\left(\vec{\varepsilon}_k + \Delta\vec{\varepsilon}_k - \vec{\varepsilon}^{ref}\right)\|^2 \quad . \tag{3.30}$$

The equation is rewritten in terms of a perturbation to a fixed point $\vec{\varepsilon}_k$. The perturbation vector, to $\Delta\vec{\varepsilon}_k$ is determined at each iteration. The second step is to calculate the gradient of (3.30),

$$\nabla_{\varepsilon}\Phi_k = \sum_{f=1}^{M_f} \Re\{J_{fk}^H(\vec{d}_f^{ref} - \vec{d}_f^{pred}(\vec{\varepsilon}_k)) - J_{fk}^H J_{fk}\Delta\vec{\varepsilon}_k\} + \beta_k F^T F(\vec{\varepsilon}_k + \Delta\vec{\varepsilon}_k - \vec{\varepsilon}^{ref}) \quad, \tag{3.31}$$

with respect to the perturbation vector. Setting the gradient to zero, (3.31) can be rewritten as

$$\sum_{f=1}^{M_f} \Re\{J_{fk}^H J_{fk}\}\Delta\vec{\varepsilon}_k - \beta_k F^T F\Delta\vec{\varepsilon}_k = \sum_{f=1}^{M_f} \Re\{J_{fk}^H(\vec{d}_f^{ref} - \vec{d}_f^{pred}(\vec{\varepsilon}_k))\} + \beta_k F^T F(\vec{\varepsilon}_k - \vec{\varepsilon}^{ref}) \tag{3.32}$$

which can be solved to find a solution to the perturbation vector, $\Delta\vec{\varepsilon}$.

The perturbation is updated after each iteration with the equation

$$\vec{\varepsilon}_{k+1} = \vec{\varepsilon}_k + \Delta\vec{\varepsilon}_k \quad. \tag{3.33}$$

If β_k is set to zero, successive linearization and Gauss-Newton produces the same solution,

$$\sum_{f=1}^{M_f} \Re\{J_{fk}^H J_{fk}\}\Delta\vec{\varepsilon}_k = \sum_{f=1}^{M_f} \Re\{J_{fk}^H(\vec{d}_f^{ref} - \vec{d}_f^{pred}(\vec{\varepsilon}_k))\} \quad, \tag{3.34}$$

but obtained via a different approach.
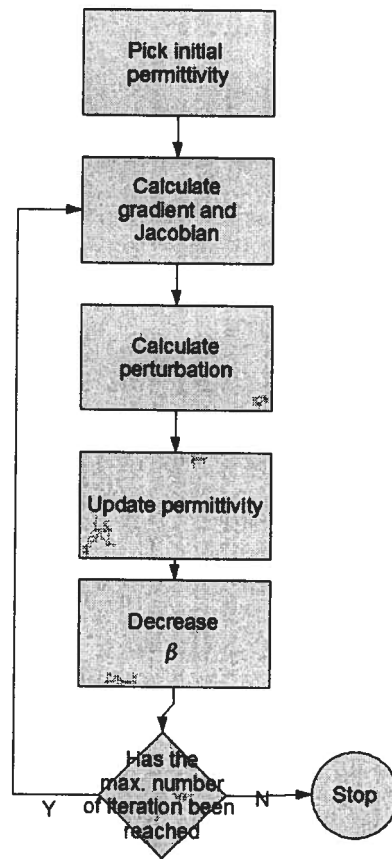
Figure 3.4: Diagrammatic representation of the successive linearization algorithm

Figure 3.3 is a flow chart of the successive linearization algorithm.

One algorithm for successive linearization can be stated as follows:

1. Pick and initial permittivity
2. Calculate the gradient and the Jacobian
3. Calculate the perturbation based on (3.32)
4. Update the functional with the new permittivity using (3.33)
5. Modify the value of β
6. If the maximum of iterations hasn't been reached goto step 2
7. stop

3.6.1 The choice of F

In (3.24), the matrix F represents an operator that constrains the overall solution.

One choice of F is the identity matrix. This choice, measures how close the recovered model is to a reference model, $\vec{\varepsilon}^{ref}$. Alternatively it can be said, if F is chosen to be the identity matrix, the solution in the region surrounding $\vec{\varepsilon}^{ref}$ is favored. The feasible region is thus, the region around $\vec{\varepsilon}^{ref}$.

The parameter β_k controls the weighting between the regularization and the data and for this case it controls how much the model can deviate from the reference model. For high values of β_k only the reference model will be recovered, because the region of feasible solution will be the point $\vec{\varepsilon}^{ref}$. The region increases in size, as β_k decreases.

A second choice for F is

$$F\vec{\varepsilon} \approx \nabla \vec{\varepsilon} , \qquad\qquad (3.35)$$

where the gradient is taken as the spatial gradient. This choice of F limits the structure in the recovered model. The spatial gradient agrees with an intuitive idea on what a feature is. That is an area with nearly homogeneous properties would be considered a feature. In a problem with 2 degrees of freedom, the region around a 45° line, passing through $\vec{\varepsilon}^{ref}$ is defined as a feasible region. If β_k is very large, a solution on this line is recovered. Solutions are allowed to deviate from this line when the values of β_k decrease.

The simplest choice for β_k is to leave it constant for the entire iterative scheme. The value β_k can be calculated for some criterion with respect to the initial linearized system. Ideally β_k is chosen as small as to not prevent features from being found.

A second choice is to perform a line search for β_k, during each iteration that yields the best Φ^d. This method, like the line search to find α_k in the gradient method, requires a linear system to be solved for each search point. A line search is necessary because the problem is non-linear. This method is also very computationally expensive. Like the line search in the gradient method, each value of β_k requires a system of linear equations to be solved.

A third choice is simply to decrease β_{k-1} by a prescribed amount after each iteration to produce β_k. Empirical evidence has shown that this to be an effective method of determining β_k, while avoiding the necessity to search for a new value of β_k after each iteration. This method has been shown to be successful in the past, but experimentation is necessary to determine the rate of decrease. The major draw back of this method, is the lack of visible progress during many of the iterations, resulting in the need for a large number of iterations for find a solution.

3.7 The Levenberg-Marquardt method
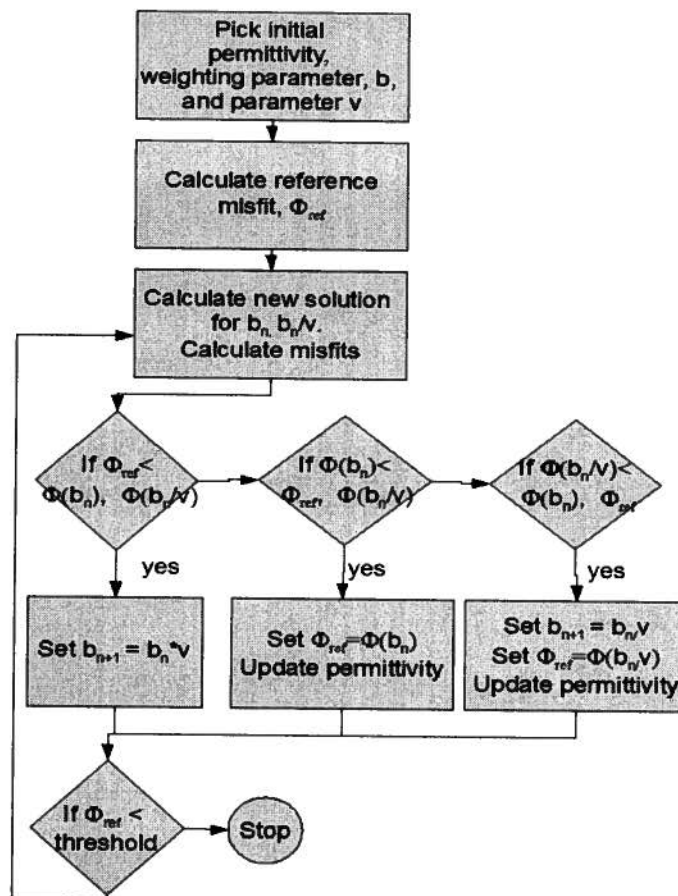
Figure 3.5: Diagrammatic representation of the Levenberg-Marquardt algorithm.

The Levenberg-Marquardt method is another popular method in minimization. It was first developed by Levenberg (1944), and then later rediscovered by Marquardt (1963). The Levenberg-Marquardt places a penalty on the magnitude of the search vector. The Levenberg-Marquardt method is introduced here and will be compared to the other methods in terms of its performance.

The Levenberg-Marquart is seen as a compromise between the Gauss-Newton method and the gradient method. The Gauss-Newton method has very quick convergence properties. However because the Jacobian matrix may be badly conditioned, the Gauss-Newton method may produce

70

unsatisfactory results.

The functional to be minimized by the Levenberg-Marquardt method the is given by,

$$\Phi_k^{LM} = \frac{1}{2} \sum_{f=1}^{M_f} \left\| \left(\vec{d}_f^{ref} - \vec{d}_f^{pred}(\vec{\varepsilon}_k) \right) - J_f \Delta \vec{\varepsilon}_k \right\|^2 + \frac{1}{2} \beta_k^{LM} \left\| F(\Delta \vec{\varepsilon}_k) \right\|^2 \quad . \tag{3.36}$$

As in the Gauss-Newton method, the functional is minimized with respect to $\Delta \vec{\varepsilon}_k$. An additional

regularization term is added, $\frac{1}{2} \beta^{LM} \left\| F(\Delta \vec{\varepsilon}_k) \right\|^2$. The equation is solved with respect to $\Delta \vec{\varepsilon}_k$ and

the permittivity for the next iteration is given by

$$\vec{\varepsilon}_{k+1} = \vec{\varepsilon}_k + \Delta \vec{\varepsilon}_k \quad . \tag{3.37}$$

The operator, F , is the same as that for successive linearization. In this case F operates on the

perturbation rather than the total permittivity distribution in this case. If, F , is chosen to have a

physical interpretation, it limits some property of $\Delta \vec{\varepsilon}_k$.

The typical choice of F while operating on the perturbation is the identity matrix which leads

to the Levenberg-Marquardt method. It is referred to as a trust region method, because the

regularization limits the size of the perturbation vector, through the second term in (3.36) which

defines a region around the expansion point as the trust region. In Levenberg-Marquardt, the feasible

region is defined around the expansion point, which changes with each iteration, while in successive

linearization the feasible region is defined around a fixed point $\vec{\varepsilon}^{ref}$.

The Levenberg-Marquardt method is an iterative method. The value of β_k^{LM} varies from

iteration to iteration. Marquardt (1963) suggest a heuristic approach for changing the value of β_k^{LM} .

Marquardt proposes to compute two solutions per an iteration. One solution for the current value of

β_k^{LM} , one for a decreased value of β_k^{LM} . The value used for the next iteration is the value which

yields the best solution.

The algorithm is outlined below:

1) pick a starting $\vec{\varepsilon}_0$, a starting β_0^{LM}, and a decrease factor v, such that $v > 1$.

2) Calculate Φ^d for $\vec{\varepsilon}_0$

3) Calculate $\Delta \vec{\varepsilon}_k$ for β_k^{LM} and β_k^{LM}/v, labeled $\Delta \vec{\varepsilon}_k^a$ and $\Delta \vec{\varepsilon}_k^b$ respectively

4) Calculate $\Phi^d(\vec{\varepsilon}_k^a)$ and $\Phi^d(\vec{\varepsilon}_k^b)$

5) if $\Phi^d(\vec{\varepsilon}_k^a) < \Phi_k^d$ and $\Phi^d(\vec{\varepsilon}_k^a) < \Phi^d(\vec{\varepsilon}_k^b)$ set $\vec{\varepsilon}_{k+1} = \vec{\varepsilon}_k + \Delta \vec{\varepsilon}_k^a$, $\beta_{k+1}^{LM} = \beta_k^{LM}$

6) if $\Phi^d(\vec{\varepsilon}_k^b) < \Phi_k^d$ and $\Phi^d(\vec{\varepsilon}_k^b) < \Phi^d(\vec{\varepsilon}_k^a)$ set $\vec{\varepsilon}_{k+1} = \vec{\varepsilon}_k + \Delta \vec{\varepsilon}_k^b$, $\beta_{k+1}^{LM} = \beta_k^{LM}/v$

7) if $\Phi_k^d < \Phi^d(\vec{\varepsilon}_k^a)$ and $\Phi_k^d < \Phi^d(\vec{\varepsilon}_k^b)$ set $\beta_{k+1}^{LM} = \beta_k^{LM} * v$

8) repeat from step 3.

3.8 Iterative re-weighted least squares

The iteratively re-weighted least square (IRLS) method can be seen as an extension of the successive linearization, that allows the minimization with respect to a general norms. IRLS is used to explore alternative norms other than L_2 in inverse problems (Farquarson at al. 1998). In this thesis, the L_1 measure is explored and its performance is compared. One motivation to use L_1 is the existence of sharp edges in the permittivity distribution. For successive linearization, the regularization is chosen to be the gradient of the permittivity distribution. The regularization penalizes sharp transitions less if the L_1 norm is used.

To generate the IRLS, an arbitrary functional can be defined as,

$$\phi = \sum_{n=1}^{N} \rho(x_n) \qquad (3.38)$$

where $\rho(x_n)$ is some general norm and in (3.1) it is defined as the square of the magnitude of x_n. The term x_n represents an error term. In the case of microwave imaging, it is the difference between the measured scattered field and predicted scattered field as shown in (3.3).

The derivative of (3.38) with respect to an arbitrary free parameter, m_k, can be expressed as

$$\frac{\partial \phi(\vec{x})}{\partial m_k} = \sum_{n=1}^{N} \rho'(x_n) \frac{\partial x_n}{\partial m_k} \qquad (3.39)$$

To find the minimum of (3.38) an iterative procedure is used. The derivative of (3.38) is rewritten such that x_n appears explicitly in the equation.

Rewriting (3.39) in matrix form, and taking into account all the free parameters at once :

$$\frac{\partial \phi(\vec{x})}{\partial \vec{m}} = B^T R \vec{x} , \qquad (3.40)$$

where R is given by

$$R = diag\{\rho'(x_1)/x_1, ..., \rho'(x_N)/x_N\} \quad , \tag{3.41}$$

for the case of the least square problem.

The terms of the matrix B are given by

$$B_{ij} = \partial x_i / \partial m_j \quad , \tag{3.42}$$

and B is the Jacobian matrix.

Equation (3.39) cannot be applied directly to the case where x_n is complex, for L_p norms

$$\rho(\vec{x}) = \sum_{n=0}^{N} \|x_n\|^p \tag{3.43}$$

because the derivative of $\rho'(x_n)$ does not exist in some cases. For the electromagnetic problem the real and imaginary parts can be interpreted as quadrature measurements. The compromise is therefore to treat the real and imaginary parts of the data separately. The derivative for the the L_1 norm is given by

$$\frac{\partial \rho(\vec{x})}{\partial x_n} = \begin{cases} 1 & \text{if } x_n > 0 \\ -1 & \text{if } x_n < 0 \end{cases} \quad . \tag{3.44}$$

Numerical difficulty arises with equation (3.42) for the L_1 norm. When the value of x_n is zero, the derivative is undefined in (3.44) and the denominator in (3.41) is zero. To overcome this difficulty maximum value for the elements of R is defined and the individual terms can be written as

74

$$R_{nn} = \begin{cases} \dfrac{1}{|x_n|} & \text{if } |x_n| > c \\[2mm] \dfrac{1}{|c|} & \text{if } |x_n| \le c \end{cases} \quad .$$

(3.45)

When the real and imaginary parts are written explicitly, (3.39) can be written as

$$\phi(\vec{x}) = \rho(\Re[\vec{x}]) + \rho(\Im[\vec{x}]) \quad .$$

(3.46)

By treating the the two complex and real parts independently, the contribution of the two parts is simply summed.

In this case, the real operator takes the real part of the elements of the vector and the imaginary operator takes the imaginary part of the elements of the vector.

The derivative of (3.46) can be written as

$$\frac{\partial \phi(\vec{x})}{\partial \vec{m}} = \frac{\partial \rho(\Re[\vec{x}])}{\partial \Re[\vec{x}]} \Re\left[\frac{\partial \vec{x}}{\partial \vec{m}}\right] + \frac{\partial \rho(\Im[\vec{x}])}{\partial \Im[\vec{x}]} \Im\left[\frac{\partial \vec{x}}{\partial \vec{m}}\right] \quad ,$$

(3.47)

and is needed to generate a solution.

Reformulating the problem in IRLS to take advantage of the ability to minimize with arbitrary norms results in

$$\Phi_k = \sum_{f=1}^{N} \left(\rho(\Re[\vec{d}_f^{ref} - \vec{d}_{fk}^{prd} - J_{fk} \Delta \vec{\varepsilon}_k]) + \rho(\Im[\vec{d}_f^{ref} - \vec{d}_{fk}^{prd} - J_{fk} \Delta \vec{\varepsilon}_k]) \right) + \beta_k \rho(F(\varepsilon_k + \Delta \varepsilon_k))$$

(3.48)

where the subscript k represents the iteration number.

We now use the formulation outlined in (3.38) to (3.42). Applying the IRLS formulation to the derivative to the first term of the summation in (3.48) produces

$$\frac{\partial \rho(\Re[\vec{d}_f^{ref} - \vec{d}_{fk}^{prd} - J_{fk} \Delta \vec{\varepsilon}_k])}{\partial \Delta \vec{\varepsilon}_k} = \Re[J_{fk}^T] R_{rk}(\Re[J_{fk} \Delta \varepsilon_k - d_f^{ref}]) \quad .$$

(3.49)

Applying the IRLS formulation to the derivative to the second term of the summation in (3.48) produces

$$\frac{\partial \rho(\Im[\vec{d}_f^{ref}-\vec{d}_{fk}^{prd}-J_{fk}\Delta\vec{\varepsilon}_k])}{\partial\Delta\vec{\varepsilon}_k}=\Im[J_{fk}^T]R_{ik}(\Im[J_{fk}\Delta\varepsilon_k-d_f^{ref}]) \quad .$$

(3.50)

Applying the IRLS formulation to the derivative to the last term of (3.48) produces

$$\frac{\partial\rho(F(\varepsilon_k+\Delta\varepsilon_k))}{\partial\Delta\vec{\varepsilon}_k}=F^T R_{sk} F(\varepsilon_k+\Delta\varepsilon_k) \quad .$$

(3.51)

By setting the derivative of (3.48) to zero and using (3.49) to (3.51) a system of linear equations is obtained:

$$\sum_{f=1}^{N}(\Re[J_k^T]R_{rk}\Re[d_f^{ref}]+\Im[J_k^T]R_{ik}\Im[d_f^{ref}])-\beta_k F^T R_{sk} F\varepsilon_k$$
$$=(\Re[J_k^T]R_{rk}\Re[J_k]+\Im[J_k^T]R_{rk}\Im[J_k]+\beta_k F^T R_{sk} F)\Delta\varepsilon_k$$

(3.52)

to solve for $\Delta\varepsilon_k$.

The minimum of (3.46) can found by solving (3.52) repeatedly in an iterative procedure. The R matrices will change after each iteration. The solution of (3.52) is the same form of the solution for the least square problem. Hence the algorithm is known as the iteratively re-weighted least squares algorithm. For the L_1 norm, there is an additional difficulty where x_n is equal to zero. The entry in the R matrix is singular. This can be mitigated with a small threshold value, which replaces x_n with a constant if x_n falls below the constant.

The problem cast as an iterative procedure will utilize the following equation,

$$\vec{\varepsilon}_{k+1}=\vec{\varepsilon}_k+\Delta\vec{\varepsilon}_k$$

(3.53)

for updating the permittivity vector at each iteration and the perturbation vector is determined by (3.52) for each iteration.

The choice of the free parameter β_k can be chosen in 2 manners. One method is to maintain β_k as constant till a minimum is reached and then decrease beta only after a minimum is reached, yielding two sets of iterations. The first set of iterations minimizes (3.48) using the L_l norm with a fixed β_k.

The second method is simply to decrease beta after each iteration. The second method does not find the exact L_1 minimum at each iteration. For the case of noiseless data, the global minimum for fitting the data for the L_l and L_2 case should be at the same location. That is to say, (3.46) and (3.1) have the same global minimum. The interpretation of the regularization term is the same as that of successive linearization. The difference is that the weighting of sharp edges in the roughness is decreased.

The algorithm is thus:

One algorithm for successive linearization can be stated as follows:
1. Pick an initial permittivity
2. Calculate the gradient, the Jacobian and R matrixes
3. Calculate the perturbation based on (3.52)
4. Update the functional with the new permittivity using (3.53)
5. Modify the value of β
6. If the maximum of iterations hasn't been reached goto step 2
7. stop

3.9 Simple comparisons

To compare the gradient method, the Gauss-Newton method, and the Levenberg-Marquardt method two examples are presented. The first example will be that of a parabola. Only the Gauss Newton method and the gradient method are compared. Typically, the Levenberg-Marquardt and the regularized version of successive linearization aren't used because for this problem because there is a unique solution and no local minima.

The second example will be the superposition of the first example and a sinusoidal function. This more detailed example will better illustrate the behavior of the algorithms. The problem is chosen such that the minimum value of the parabola has a value of zero. The sinusoidal function provides a strong local minima behavior, which helps illustrate the non-ideal behaviors of the various algorithms.

3.9.1 Example 1

For the first example the functional ,

$$\Phi_1(\vec{x}) = \frac{1}{2} \sum_{n=1}^{N_1} \| b_{1n} x_n - a_{1n} \|^2 \tag{3.54}$$

is used. It consists of an offset, a_n , from the origin and a stretch parameter, b_n . Here, N_1 is the number of observations and it is equal to 2. The subscript 1 refers to the fact that this is the first observation set. The number of parameters is also equal to 2, because two parameters can be plotted and the behavior of the algorithm can be illustrated graphically.

The gradient is given by

$$\nabla \Phi_1(\vec{x}) = \begin{bmatrix} b_{11}(b_{11} x_1 - a_{11}) \\ b_{12}(b_{12} x_2 - a_{12}) \end{bmatrix} . \tag{3.55}$$

The Jacobian is simply the diagonal matrix,

$$J_1 = \begin{bmatrix} b_{11} & 0 \\ 0 & b_{12} \end{bmatrix} , \qquad (3.56)$$

because the terms inside the norm are linear. Hence, the terms in the Jacobian are constant.

The Hessian is given by

$$G_1 = \begin{bmatrix} b_{11}^2 & 0 \\ 0 & b_{12}^2 \end{bmatrix} \qquad (3.57)$$

for this example the Hessian is equal to $J^H J$, and the Gauss-Newton method is exactly the same as the Newton method.

For this experiment the following parameters are used

$$\vec{b}_1 = \begin{bmatrix} 2.5 \\ 1 \end{bmatrix} , \qquad (3.58)$$

$$\vec{a}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} . \qquad (3.59)$$

These parameters are picked arbitrarily and the global minimum occurs at

$$x_{min} = \begin{bmatrix} 0.4 \\ 2 \end{bmatrix} \qquad (3.60)$$

with a value of 0 at the minimum.

Because the shape of the functional is a parabola, a starting point can be picked arbitrarily and the algorithms will be well behaved. For the all the iterative procedures the starting point

$$x_0 = \begin{bmatrix} 3 \\ 3 \end{bmatrix} , \qquad (3.61)$$

will be used.

The contour lines in figure 3.6 are ellipses , because the elements of $\vec{b}$ are not the same. If they were the same, the contour lines would be circles. The major and minor axis of the ellipses are aligned with the coordinate axes because there are no off diagonal terms in our Jacobian.
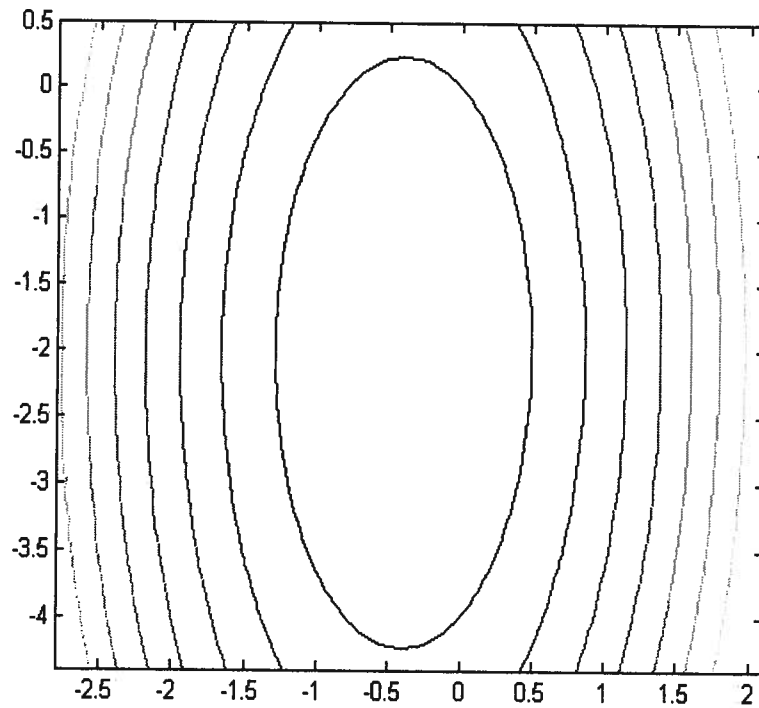
Figure 3.6: Contours of a parabolic functional.
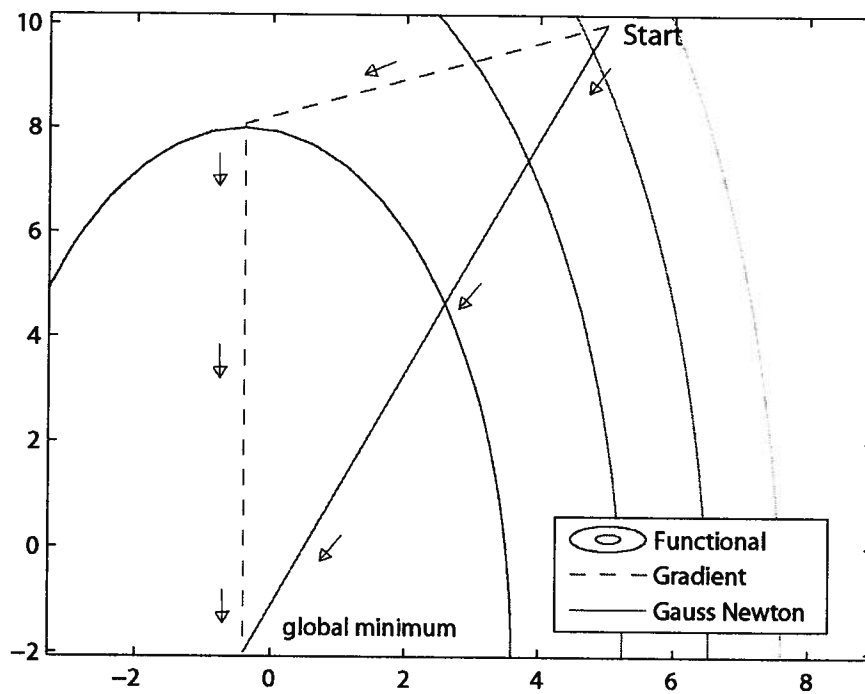
Figure 3.7: Paths taken from the initial point to the minimum. The Gauss-Newton method takes the a direct path to the minimum, while the gradient method requires two steps.

Figure 3.7 shows the path taken by the Gauss-Newton method, and the gradient method. The Gauss-Newton method goes directly from the start point to the global minimum. The gradient method takes two steps to reach the minimum.
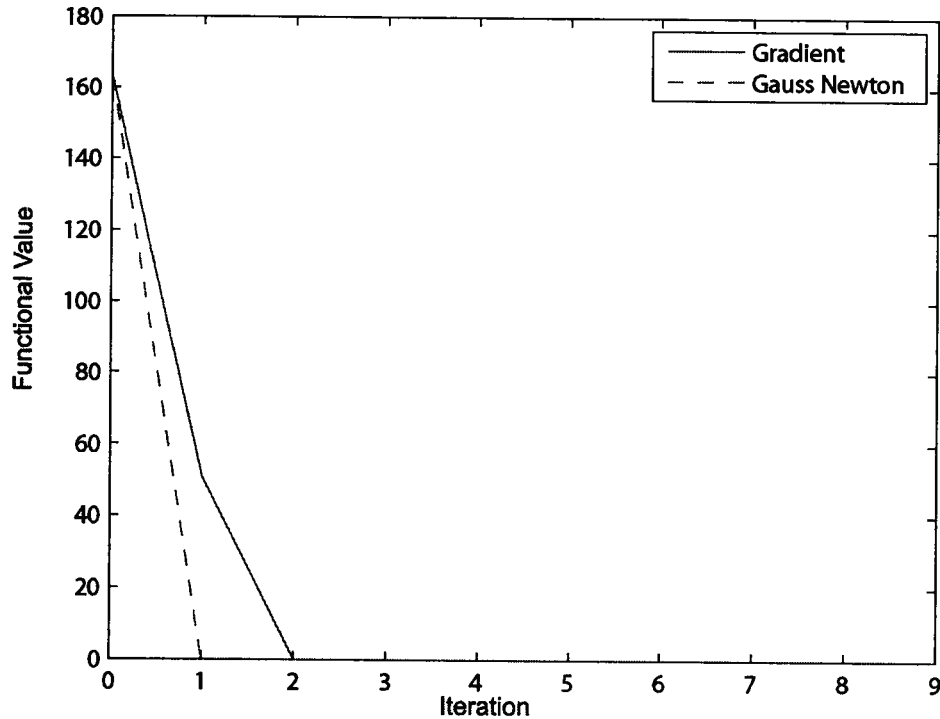
Figure 3.8: Functional values for the minimization of a parabola. For this example, the gradient method takes 2 iterations to reach the global minimum and the Gauss-Newton method takes one.

From the example, the Gauss-Newton method takes one iteration to reach the solution, and the gradient method takes two iterations as seen in figure 3.8. The Gauss-Newton method will always take the direct path to the minimum regardless of starting point, because the Gauss-Newton functional is exactly the same as the functional is defined in (3.27).

The gradient method takes more than one iteration because the initial gradient, does not point to the minimum, as can be seen in figure 3.7. Because the line search used is approximate, the first gradient search does lead to the minimum point along the gradient. The first search vector lands on the top of an ellipse, if the vector was extended slightly, the end point of the vector would be inside the ellipse, implying a smaller functional value. This results from using the gradient method with an approximate line search. While an exact line search would yield a smaller functional value after the

82

first iteration, it would also require more computational effort. In this thesis, the approximate linear search is used, because of the computational expense of calculating the field values is far too great for an exact line search to be feasible.

3.9.2 Example 2

For a second example, the behavior of the optimization techniques can be better illustrated, and the less than ideal behaviors are more apparent. The functional in example 1 and a sinusoidal functional are added together. This creates a functional with many local minima.

The sinusoidal functional is given by,

$$\phi_2 = \frac{1}{2} \sum_{n=1}^{N_2} \| \sin(b_{2n}(x_n + a_{2n})) \|^2 \quad . \tag{3.62}$$

It consists of sinusoidal functions and has an infinite number of local minima.

The gradient is given by,

$$\nabla \Phi_2 = \begin{bmatrix} b_{21} \sin(b_{21}(x_1 + a_{21})) \cos(b_{21}(x_1 + a_{21})) \\ b_{22} \sin(b_{22}(x_2 + a_{22})) \cos(b_{22}(x_2 + a_{22})) \end{bmatrix} \quad . \tag{3.63}$$

The Jacobian is given by,

$$J_2 = \begin{bmatrix} b_{21} \cos(b_{21}(x_1 + a_{21})) & 0 \\ 0 & b_{22} \cos(b_{22}(x_2 + a_{22})) \end{bmatrix} \quad . \tag{3.64}$$

The Hessian is given by

$$G_2 = \begin{bmatrix} b_{21} - 2 b_{21}^2 \sin(b_{21}(x_1 + a_{21}))^2 & 0 \\ 0 & b_{22} - 2 b_{22}^2 \sin(b_{22}(x_2 + a_{22}))^2 \end{bmatrix} \quad . \tag{3.65}$$

The Hessian is no longer well approximated by $J^H J$ when the functional is far away from a

minima, but close to the minima $J^H J$ remains a good approximation, because the sine term is near 0.

As mentioned earlier the second example will combine the first, (3.54) and second functionals, (3.62) and can be written as

$$\Phi = \frac{1}{2} \sum_{n=1}^{2} \| b_{1n} x_n - a_{1n} \|^2 + \frac{1}{2} \sum_{n=1}^{2} \| \sin(b_{2n}(x_n + a_{2n})) \|^2 \quad .$$

(3.66)

The parameters for the first functional are

$$\vec{b}_1 = \begin{bmatrix} 0.5 \\ 0.33 \end{bmatrix} \quad ,$$

(3.67)

and

$$\vec{a}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad ,$$

(3.68)

resulting in a parabola centered on (-2, 6.06).

The parameters for the second functional are

$$\vec{b}_2 = \begin{bmatrix} 2.5 \\ 1 \end{bmatrix}$$

(3.69)

and

$$\vec{a}_2 = \begin{bmatrix} 2 \\ 6.06 \end{bmatrix} \quad ,$$

(3.70)

resulting in a minimum at the same location as the first functional and several minima around that location. The parameters values are chosen such that there will be many local minima near the global minima for the total functional.

The starting point

$$x_0 = \begin{bmatrix} -2.5 \\ -4.5 \end{bmatrix} , \qquad (3.71)$$

is chosen to be near the global minimum. The contours for the combined functional are shown in figure 3.9.
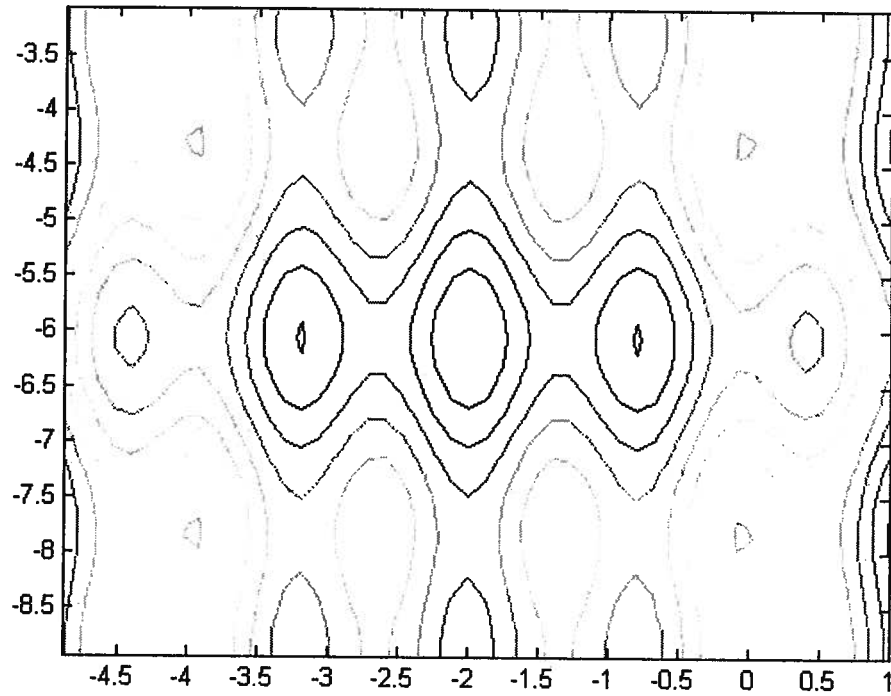
Figure 3.9: Combined functional. There are many local minima surrounding the global minima. The global minimum lies in the center of the graph.

When there are many local minima surrounding the global minimum, the search algorithms will have difficulty finding the global minimum and in many cases the global minimum will not be found. The minimum that is found will depend on the starting location.
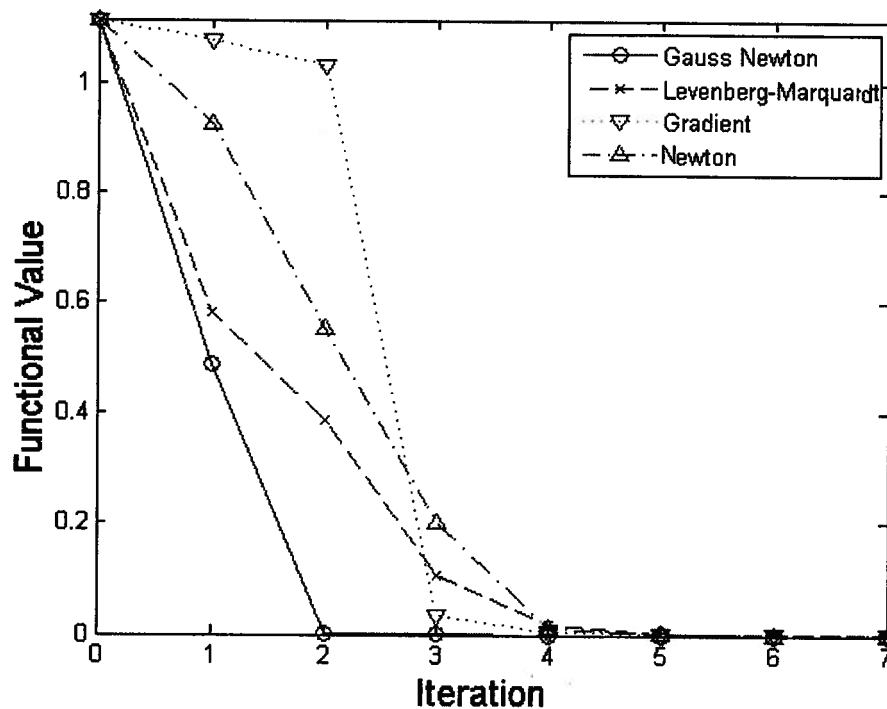
Figure 3.10: Performance chart of the gradient method, Gauss-Newton method, and the Levenberg-Marquardt method.

From figure 3.10 it can be seen that the Gauss-Newton method performs the best, because it took the least number of iterations to reach the minimum. All of the algorithms converged to a final functional value of 0. The gradient has the second best performance, it decreases slowly initially followed by a big jump, and then decrease slowly again.

The Levenberg-Marquardt method and the Newton method both decrease in even steps to the minimum. However they take the most iterations to reach the minimum. For this example, it can be said the Gauss-Newton performed the best. However this will not remain true if the starting point is changed.
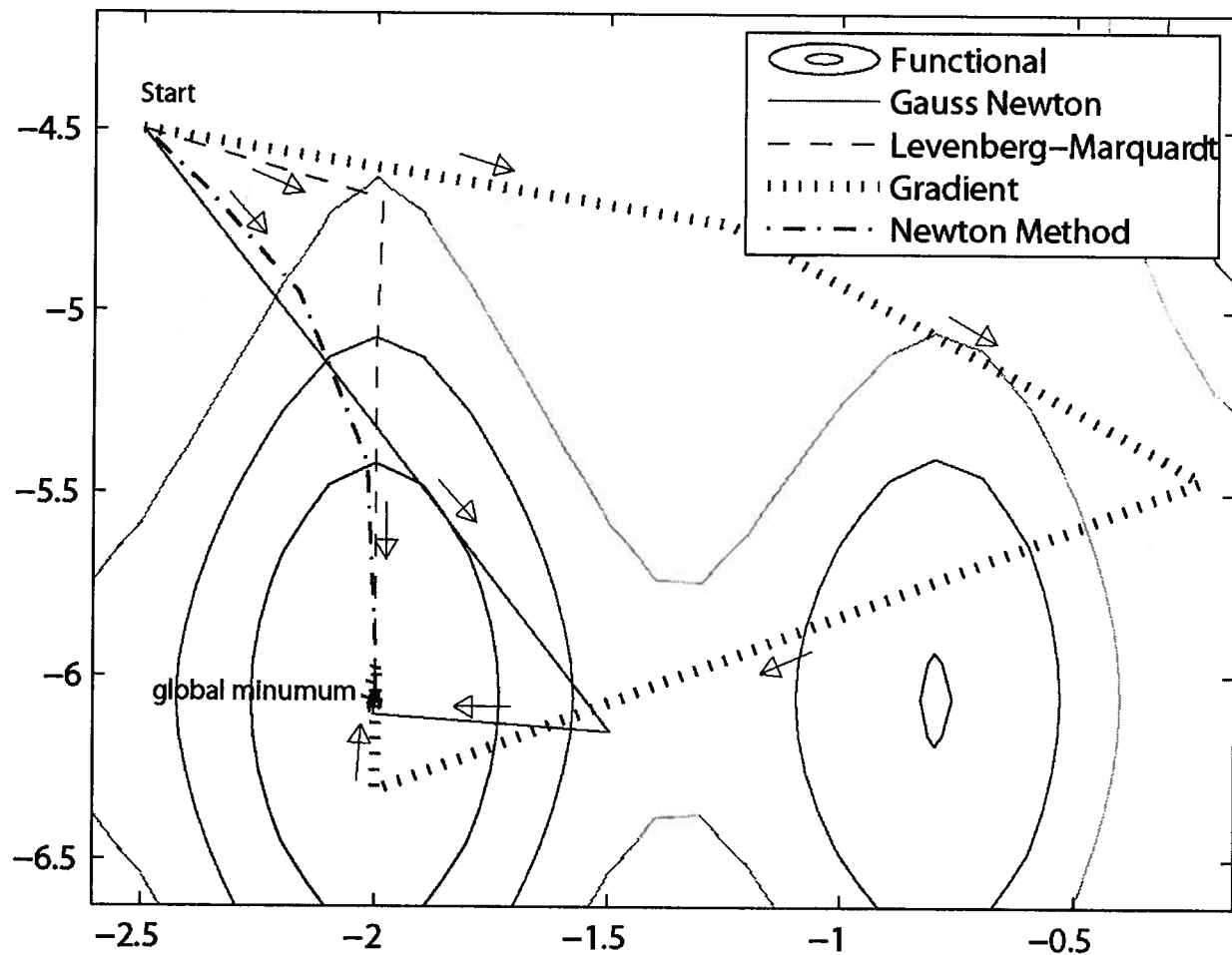
Figure 3.11: Trajectories taken by the various algorithms

In Figure 3.11 it can be seen that the trajectory of the gradient method jumps from local

minimum to minimum, which explains the sudden decrease in the functional value seen in Figure 3.11.

This is partly due to the our use of an approximate line search and partly due to the non-linear nature of

the functional. In early iterations, the gradient method gets further from the solution, by jumping over

several minima. This occurs because, far away from a minima the functional value is not proportional

to the distance to the minima. This is an undesirable behavior for inverse problems, because it implies

a solution far from actual solution can be found with a small functional value. This can be mitigated in

the inverse problem by including additional observations, but can not be completely avoided.

The Levenberg-Martquardt takes the shortest path to the minimum because the starting point was chosen to be near the minimum. This is due to the constraint on the size of the perturbation vector. The constraint on the perturbation vector has prevented the Levenberg-Marquart method, for this example, to jump to other local minima. That is to say, none of the search vectors cross over contour lines that increase in value.

The Gauss-Newton method and the Newton method take relatively uninteresting paths to the minimum. However, the Gauss-Newton method does not follow the contours as well as the Levenberg-Martquardt method, as can be seen by the trajectories passing over a several contour lines that increase in value.

To demonstrate an example where the global minimum is not found, a starting point further away from the minimum is chosen. Similar problems will occur, when applying these techniques to the microwave imaging problem. A contour plot cannot be drawn for the microwave imaging problem because of the high number of degrees of freedom. The current 2-D example can be used to build up intuition to understand the behavior of the higher dimensional minimization problem.

The starting points

$$x_0 = \begin{bmatrix} 5.0 \\ 5.0 \end{bmatrix} \text{ ,} \qquad (3.72)$$

and

$$x_0 = \begin{bmatrix} 4.0 \\ 5.0 \end{bmatrix} \text{ ,} \qquad (3.73)$$

are chosen to demonstrate that the algorithms may not find the global minimum.
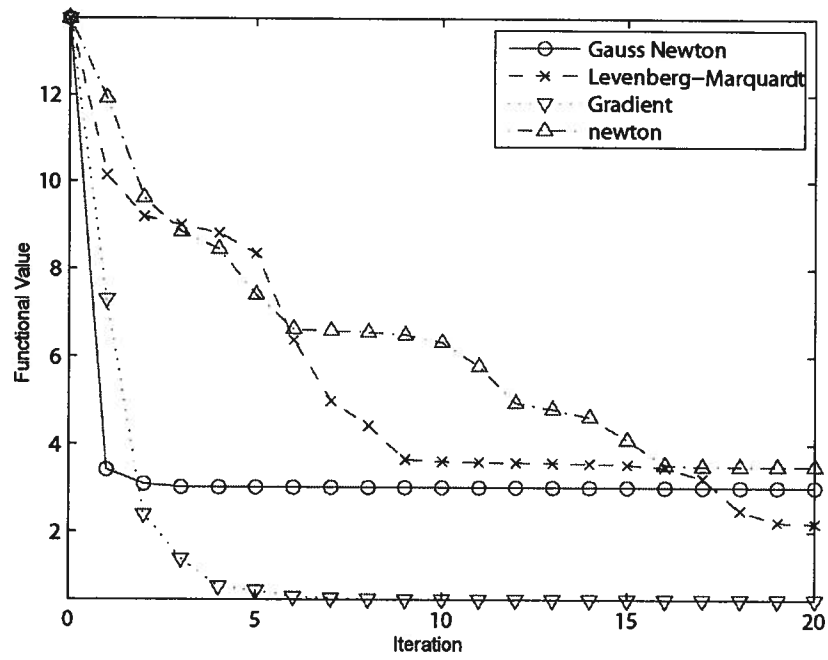
Figure 3.12: Functional value plot with the starting point $\quad x_0 = (5.0, 5.0)$
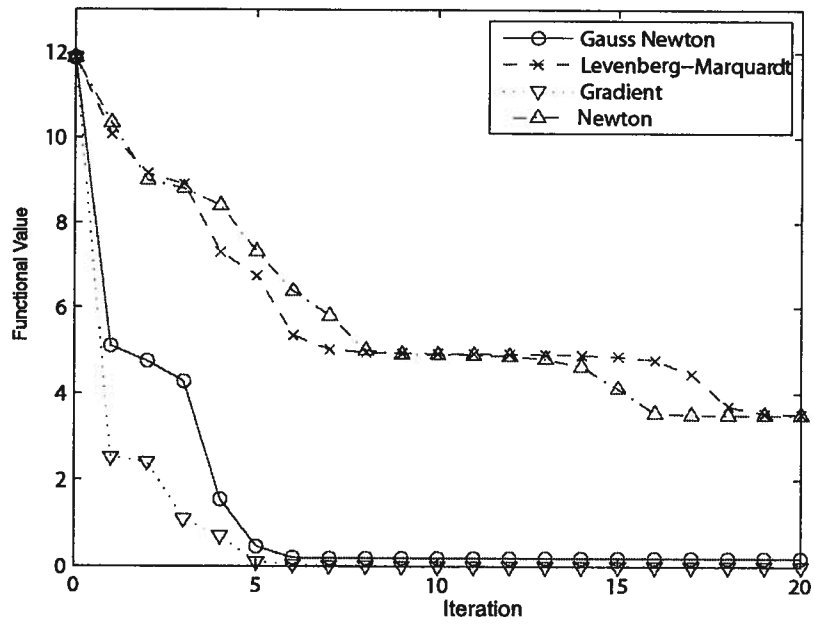
Figure 3.13: Functional value plot with starting point $\quad x_0 = (4.0, 5.0)$

Table 3.1: Comparison of algorithms for two starting points		
Method	**End value for** $x_0 = (5.0, 5.0)$	**End value for** $x_0 = (4.0, 5.0)$
Gauss-Newton	3.02308174684158	0.189764659278655
Levenberg-Marquardt	2.18768407625599	3.50601609245224
Gradient	0.482932012956024	0
Newton	3.50601375975962	3.50601375975962

For this example, as can be seen in Table 3.1, that the gradient method has performed the best because it found the smallest functional value in both cases. However, the sensitivity to starting conditions can also be seen. In the case when $x_0 = (4.0, 5.0)$, the global minimum is found, as can be seen by the functional value of 0. In the second case a local minimum is found. For the other algorithms, the convergence is different for the two starting points. It should be noted that the difference between the two starting points is small.

The Gauss-Newton method for starting point $x_0 = (5.0, 5.0)$, reaches a local minima in very few iterations. However given the second starting point, a smaller value minimum is reached, but it takes longer. This suggests that under certain contour geometries, Gauss-Newton can be easily trapped inside a local minima. It can be further said, that the convergence is highly dependent on the local properties of the functional.

The Levenberg-Marquardt method places second for the first starting point, and tied for last with the second starting point. Because the magnitude of the perturbation is penalized, the Levenberg-Marquardt takes relative small steps to reach the local minima. It should also be noted, that the Levenberg-Marquardt converged to two separate values for the two separate starting points.

The Newton method performed the worst in both cases, despite the fact that it used second derivative information. It however, is the only algorithm that converged to the same location in both cases.

Looking at these results, it is difficult to say which algorithm is superior. The performance is largely

dependent on the local properties of the functional at the starting point.
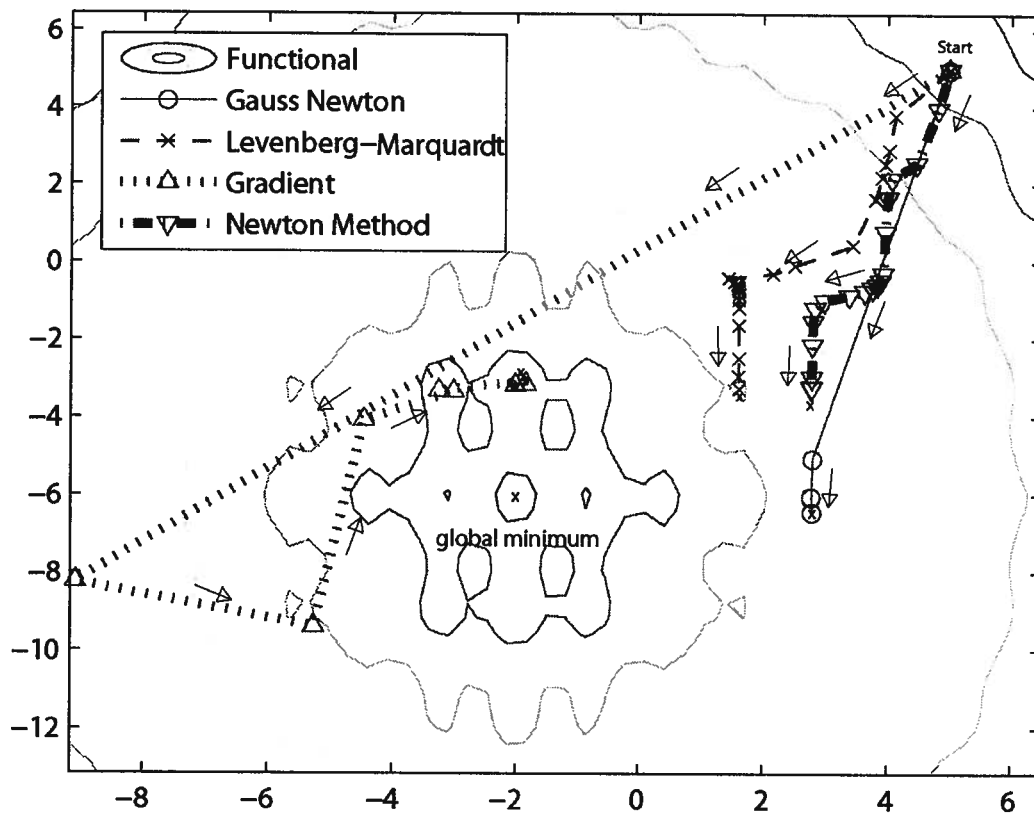
Figure 3.14: Trajectories for starting point $x_0 = (5.0, 5.0)$
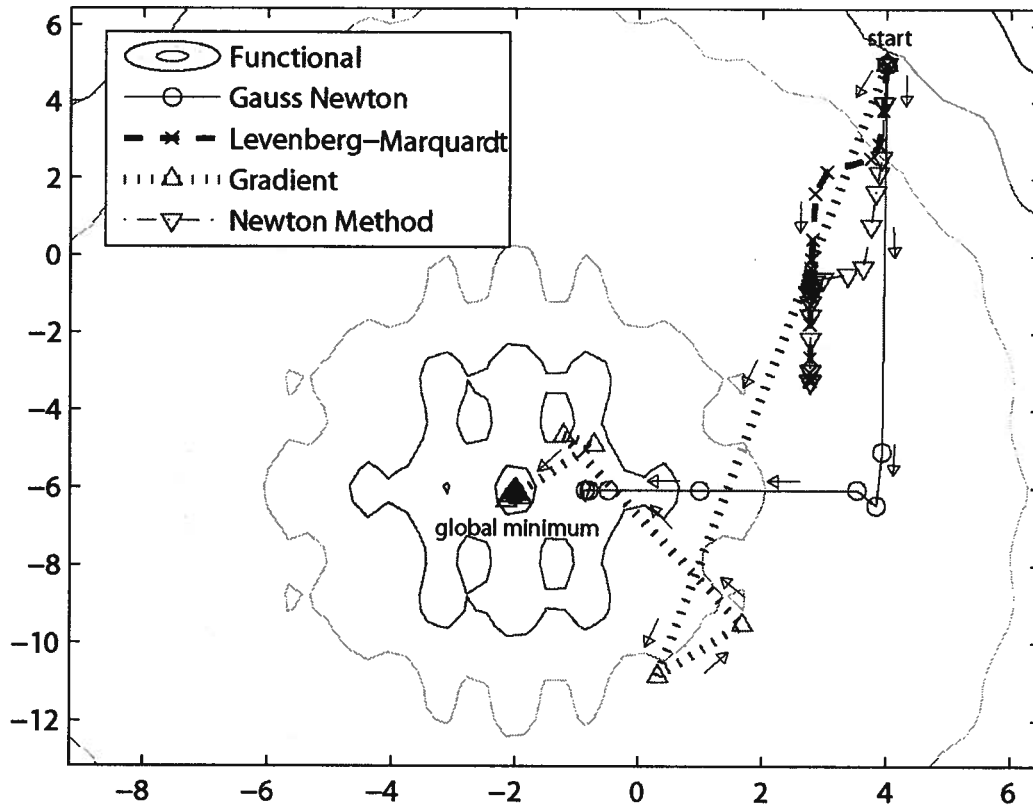
Figure 3.15: Trajectories for starting point $x_0 = (4.0, 5.0)$

From figure 3.14 and 3.15, corresponding to $x_0 = (5.0, 5.0)$ and $x_0 = (4.0, 5.0)$, the trajectories of the algorithms can be seen.

For the gradient method, comparing trajectory in figure 3.14 and 3.15, it can be seen that a completely different path is taken in both cases. In figure 3.15, the global minimum is reached, while a local minimum is reached in figure 3.14. Both trajectories share a common behavior, in that that there is a large initial jump that crosses over many contours, some increasing in value and some decreasing. For these two examples, it is precisely this behavior that leads the algorithm to find the small functional value when compared to the other algorithms. However, this behavior cannot guarantee that the smallest value is found.

The trajectory of the Gauss-Newton Method is very different for both starting points. For the starting point $(4.0, 5.0)$ a smaller minima is found, but in both cases the Gauss-Newton Method is trapped in a local minima.

The Newton Method behaved in the same manner for both starting points. As can be seen in Figure 3.14 and 3.15, the Newton Method followed the contours of the functional and converged to the same minimum for both cases.

The Levenberg-Marquardt method found different minima for both cases. The Levenberg-Marquardt method followed the contours closely due to the penalty on the size of the perturbation.

From the contour graphs it can be expected that the gradient follows the contour the least, while the Newton method follows the contours the most. This behavior however, does not lead to a better solution and in fact may lead to a higher functional value. All the algorithms get trapped in local minima, because they use local information to attempt to find the minimum. This poses a difficulty for inverse problems, as the algorithms can get stuck in local minima, and fail to find an adequate solution.

3.10 Conclusions

In this chapter several techniques for minimization of a sum of squares function are presented. Each algorithm performed differently. However none of the algorithms can be said to have superior performance because performance is dependent on the starting point.

The gradient does not follow the contours of the functional closely. In the example this led to it finding the best solutions. The Newton method followed the contour the best. However it had the poorest performance. The relative performance of the different algorithms is an artifact of the functional presented, and should not be generalized to other functionals. The ability of the algorithms to follow the contours can be generalized. The Gauss-Newton and the Levenberg-Marquardt, in terms of their ability to follow contours, is somewhere between that of the gradient method and the Newton method, with the Levenbeg-Marquardt having a slightly better contour following ability than the Gauss-Newton Method.

In terms of computation effort, the gradient method takes the least effort per iteration and the Newton takes the most. The Levenberg-Marquardt method, Successive Linearization, Gauss-Newton are equivalent in computational effort. These examples have shown that many algorithms can be used to solve the problem of minimizing a nonlinear functional. However, none can be said to have superior performance.

Chapter 4

Performance of minimization algorithms and data selection

4.1 Introduction to chapter 4

In this chapter, the behavior of the different algorithms with respect to different amounts of data is explored. Unlike linear systems, there are no clear guidelines on the amount of data or the characteristics of the data that are required to find a solution. The solution is dependent on the starting point and the misfit surface. The misfit surface is dependent on the data and the forward model.

For a linear system, if there are N unknowns then N linearly independent (Lipschutz, 1991) equations are needed to generate an unique solution. The location and frequency associated with the data points determines the system of equations and their linear independence. If there are less than N equations, then there are multiple solutions. If there are more than N linear equations, then the existence of a solution is not guaranteed.

For a non-linear system such criteria may not be possible. One possible way to understand the influence of the data on the recovery is to examine the relationship between the data and the misfit surface. This at best is an intuitive understanding of the problem. Because of the large number of unknowns it is impossible to examine the surface directly. Various combinations of algorithms and data points are presented. By examining how each experiment behaves and understanding the behavior of the algorithm, an intuition of the behavior of the surface can be developed.

For the experiments in this thesis plane wave illumination is used. Data generated using different number of frequencies and different number of angle of incidences will be presented. First the

experimental setup is presented, followed by the results for the gradient method, Levenberg-

Marquardt(LM) method, L_2 cooled roughness method, and L_1 cooled roughness method. The

Newton method is omitted due to the computational cost of the Hessian. The Gauss-Newton method is

omitted because it produces unstable solutions without regularization. The LM method and L_2

cooled roughness are variations of the Gauss-Newton method with regularization. Any information

that can be obtained by the Gauss-Newton method can also be obtained through the L_2 cooled

roughness method and the LM method.

4.2 Experimental setup

For the simulations in this chapter, the algorithms will attempt to recover two relative

permittivity distributions. Two distributions are used to verify that any observed behavior exhibited

during a recovery is not an artifact of the algorithm. The two distributions are shown in figure 4.1 and

4.2. The relative permittivity distribution shown in figure 4.1 will be known as the large anomaly case

and the relative permittivity distribution shown in figure 4.2 will be known as the small anomaly case.
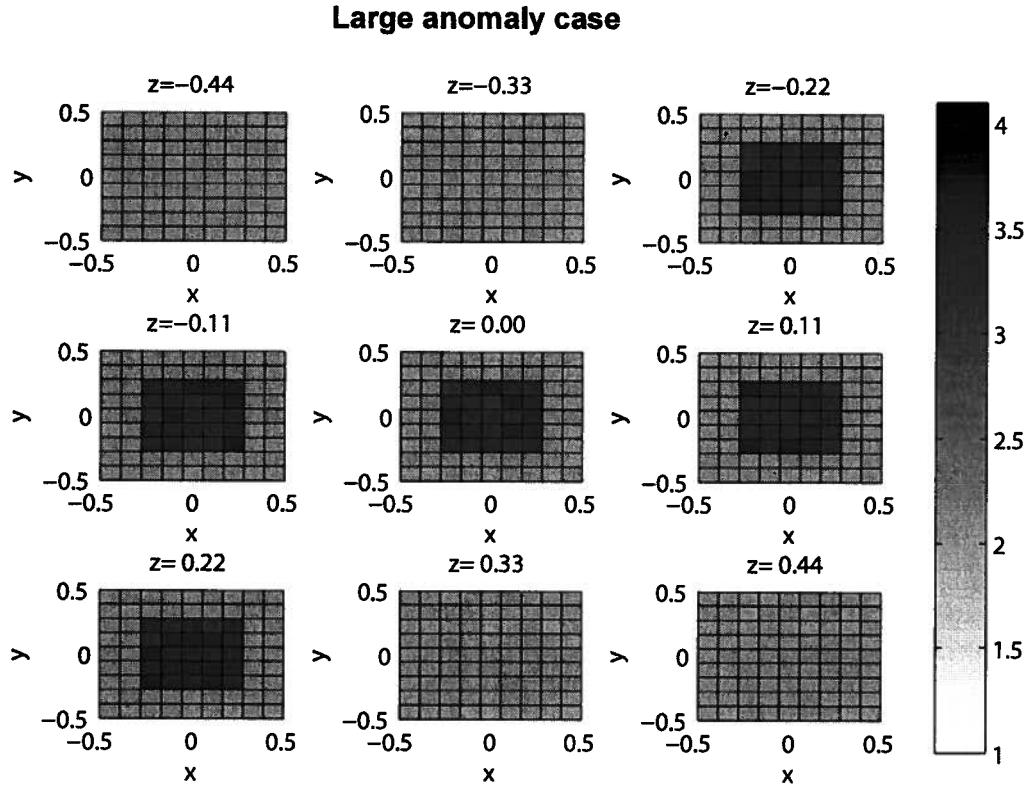
Large anomaly case

Figure 4.1: A relative permittivity distribution with a large anomaly in the center. Slices down the z axis are shown. Each slice lies on a x-y plane. The relative permittivity of the cube in the center is a uniform random distribution from 2.95 to 3.05. The relative permittivity of the outer shell is an uniform random distribution ranging from 1.95 to 2.05. The cube is 1 free space wavelength by 1 free space wavelength at 1GHz. The center of both cubes are at the origin.

Figure 4.1 represents a large cube with a smaller cube inside. Both cubes are centered on the

origin. The outer cube has a sidelength of 1 free space wavelength, relative to 1GHz. The inner cube

has a sidelength of $\frac{1}{2}$ free space wavelength relative to 1GHz. The relative permittivity of the inner

cube is taken from an uniform random distribution ranging from 2.95 to 3.05. The relative permittivity

of the outer shell is taken from an uniform random distribution ranging from 1.95 to 2.05. The random

distribution is used to introduce an unfavorable condition for the roughness regularization. The

roughness regularization penalizes non-homogeneous regions and a random distribution is non-

homogeneous.
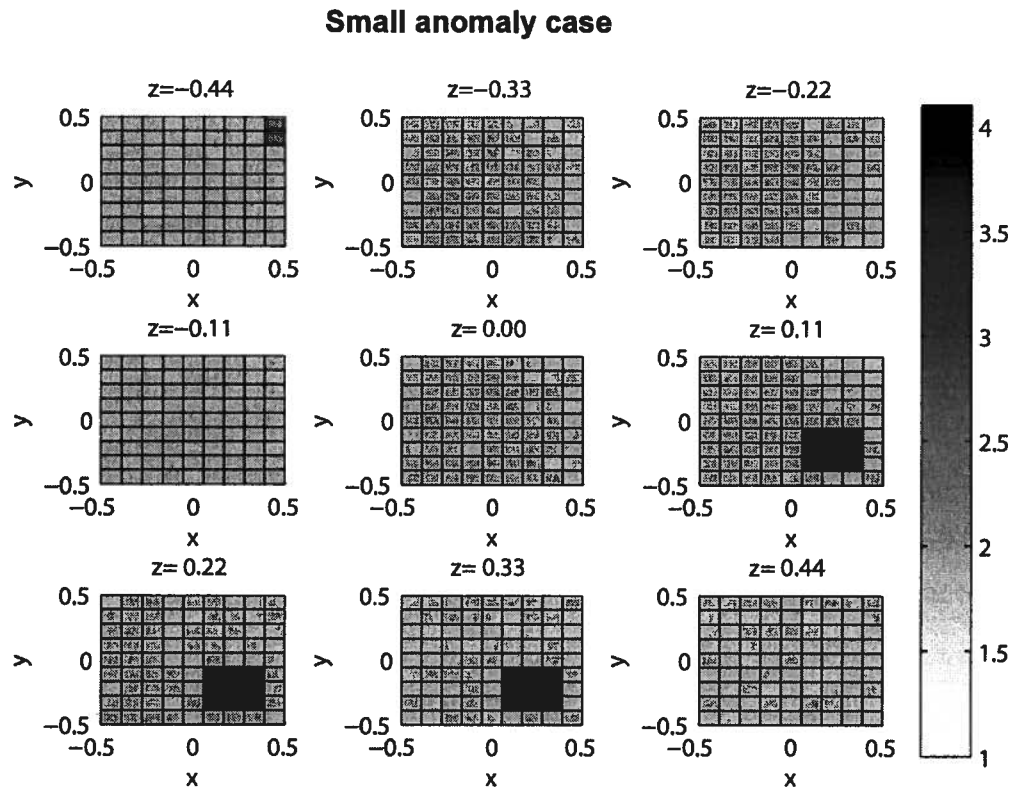
Small anomaly case

Figure 4.2: An relative permittivity distribution with a small anomaly. Slices down the z axis are shown. Each slice lies on a x-y plane. The relative permittivity of the small cube is an uniform random distribution ranging from 3.95 to 4.05. The relative permittivity of the large cube is an uniform random distribution ranging from 1.95 to 2.05. The cube is 1 free space wavelength by 1 free space wavelength long at 1GHz. The center of the large cube is at the origin. The small cube is one third wavelength by one third wavelength.

Figure 4.2 shows a large cube containing a small off center anomaly. The small anomaly is a

cube with a sidelength of $\frac{1}{3}$ free space wavelength relative to 1GHz. The large cube has a

sidelength of 1 free space wavelength. The large cube is centered around the origin and the small

anomaly is centered around x=0.22 λ, y=-022 λ, and z =0.22 λ. The relative permittivity of the small

anomaly is taken from an uniform random distribution ranging from 3.95 to 4.05. The relative

permittivity of the large cube is taken from an uniform random distribution ranging from 1.95 to 2.05.

Illumination and measurement geometry
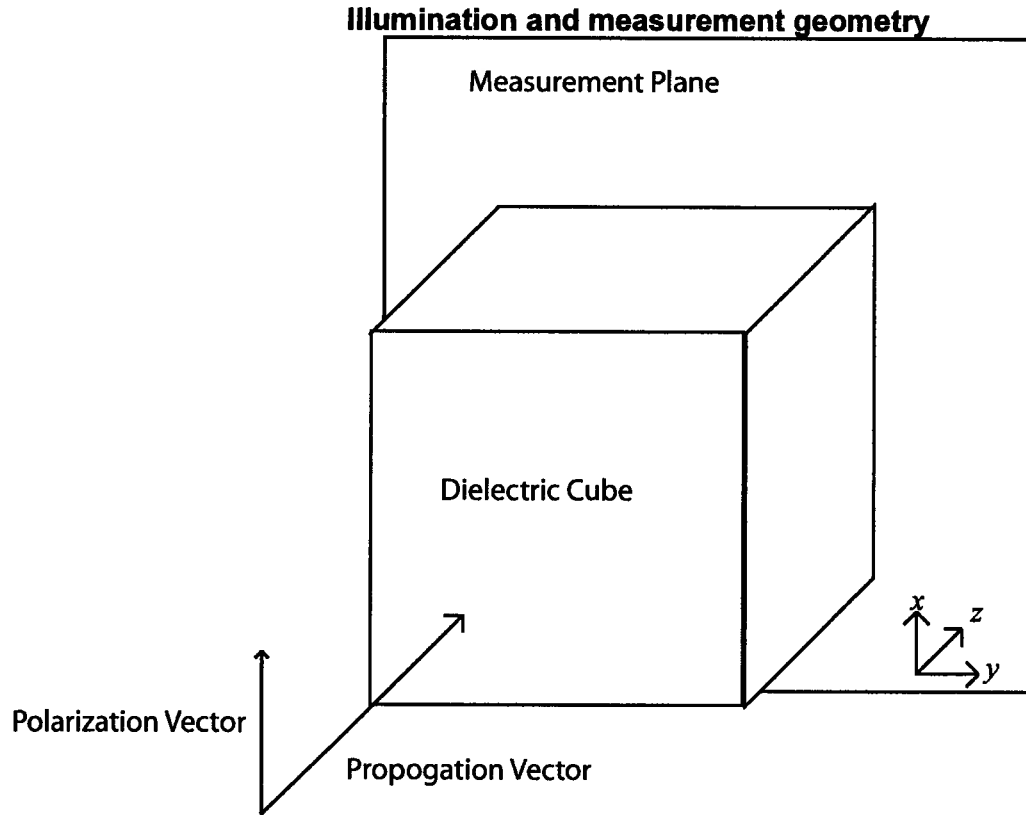
Figure 4.3: Measurement geometry with the polarization in the x direction, and the propagation vector is in the z direction. The measurement plane is an x-y plane and sits behind the dielectric cube.

Six different illumination frequencies and six different measurement geometries are used in

these simulations. The details of the geometry are given in table 4.1. An example of one of the

measurement geometries is shown in figure 4.3. All the geometries are rotations of the one shown in

figure 4.3. The transmitted field is used in the inversion. Therefore the measurement plane is placed

behind the dielectric object at a distance of 1.5 λ (45cm) from the center of the cube. The scattered

field is measured on a square aperture with a sidelength of 1 λ (30cm). The incident field's

propagation direction is normal to one face of the cube.

The dielectric cube is illuminated by a planewave. The illumination geometries are chosen such that there is an illumination geometry that corresponds to each face of the cube. The rationale for increasing the number of illumination geometries is to eliminate shadow regions; that is regions that are not directly illuminated by the planewave. Table 4.1 summarizes the combination of illumination and measurement planes used.

Table 4.1: illumination and measurement parameters				
Label	Illumination propagation direction (x,y,z)	Polarization direction (x,y,z)	Measurement plane	Measurement points in wavelengths at 1 GHz
I1	(1, 0, 0)	(0,0,1)	z-y plane	All combinations of x=1.5 y=(-0.5, -0.25, 0, 0.25, 0.5) z=(-0.5, -0.25, 0, 0.25, 0.5)
I2	(0, 1, 0)	(0,0,1)	z-x plane	All combinations of x=(-0.5, -0.25, 0, 0.25, 0.5) y=1.5 z=(-0.5, -0.25, 0, 0.25, 0.5)
I3	(-1, 0, 0)	(0,0,1)	z-y plane	All combinations of x=-1.5 y=(-0.5, -0.25, 0, 0.25, 0.5) z=(-0.5, -0.25, 0, 0.25, 0.5)
I4	(0, -1,0)	(0,0,1)	z- x plane	All combinations of x=(-0.5, -0.25, 0, 0.25, 0.5) y=-1.5 z=(-0.5, -0.25, 0, 0.25, 0.5)
I5	(0 0, 1)	(1,0,0)	x-y plane	All combinations of x=(-0.5, -0.25, 0, 0.25, 0.5) y=(-0.5, -0.25, 0, 0.25, 0.5) z=1.5
I6	(0, 0,-1)	(1,0,0)	x-y plane	All combinations of x = (-0.5, -0.25, 0, 0.25, 0.5) y=(-0.5, -0.25, 0, 0.25, 0.5) z=-1.5

By introducing multiple illumination geometries, the non-uniqueness due to shadows is reduced because the shadow regions in one illumination geometry will not be the same as that of a

second illumination geometry. If the planewave does not illuminate a region, then information regarding the region is not encoded in any of the measured data. A distinction between a shadow region and a null region needs to be defined, because they both manifest themselves as regions with a weak or zero electric field. Shadow regions are areas the illumination cannot reach, null regions are caused by standing waves. The difference between the two is in their sensitivities. Changing the permittivity of a null region will influence the measured data, because it will change the standing wave behavior compared to changing the permittivity of a shadow region which will not influence the data. The latter leads to non-uniqueness in the solution.

The frequencies used in the experiments are listed in table 4.2. The frequencies range from 0.8GHz to 1.2 GHz. Multiple frequencies are introduced to reduce non-uniqueness in the solution due to the intrinsic behavior of the scatterer (Lam, 2005). Lam has shown that for a homogeneous dielectric cylinder, there are multiple non-continuous permittivity values that lead to similar scattered fields. The non-uniqueness behavior is different for different frequencies. So, although two very different distributions may have similar fields at 1 frequency, it is unlikely that they will have similar fields at a second frequency. By including multiple frequencies, a disambiguation affect will occur. That is the more frequencies that are included, the greater the likelihood of an unique solution.

Table 4.2: The frequencies used for the experiments in this chapter.		
Label	Frequency	free space wavelength
F1	0.8GHz	37.5cm
F2	0.9GHz	33.3cm
F3	1.0GHz	30.0cm
F4	1.1GHz	27.3cm
F5	1.2GHz	25.0cm
F6	1.3GHz	23.1cm

These experimental parameters will be used in varying combinations in the rest of the chapter. Given that each method of increasing data is used to resolve different issues of non-uniqueness, the

ideal situation is to include multi-frequency data with multiple illumination geometries. To study the behavior of each however, only multiple illumination geometries or multiple frequencies are used in this chapter, so that the effect of each can be isolated.

4.3 The misfit value and the model error

The misfit value used will be the one defined in (3.1) and is presented below as a reminder to the reader.

$$\Phi^d = \frac{1}{2} \sum_{f=1}^{N_f} \sum_{n=1}^{M_f} \| d_{fn}^{ref} - d_{fn}^{pred}(\bar{\varepsilon}) \|^2 \quad . \tag{3.1}$$

This misfit is the sum of the squares of the difference between the measured data and the simulated data. Because the misfit value is dependent on the measured data, only misfit values from experiments using the same frequencies and measurement geometries can be compared directly. The contribution from different illumination geometries or frequencies is additive and corresponds to the outer summation of (3.1).

For explanation purposes let us assume two recoveries are performed, one with an experiment geometry labeled A and one with an experiment geometry labeled B. The misfit value for A will be denoted Φ_A^d and the misfit for B will be denoted Φ_B^d.

The model error is defined as the difference between the recovered relative permittivity distribution and the actual relative permittivity distribution and is written as

$$\Phi^m = \sum_{i=1}^{M} \| \varepsilon_i^{rec} - \varepsilon_i^{act} \|^2 \quad . \tag{4.1}$$

The constant M is the number of permittivity parameters, as defined in chapter 2. The symbol ε_i^{rec} denotes the i'th recovered permittivity parameter. The symbol ε_i^{act} denotes the i'th reference permittivity parameter. In practice ε_i^{act} is unknown and therefore (4.1) cannot be calculated, however in this chapter ε_i^{act} is known.

The model error for A shall be denoted Φ_A^m and the model error for B shall be denoted Φ_B^m.

Ultimately, the goal is to minimize the model error. This can not be done directly because ε_i^{act} is typically unknown. Therefore Φ^d is minimized to indirectly to find the minimum of Φ^m . However, for different observation sets $\Phi_A^d > \Phi_B^d$ does not imply that $\Phi_A^m > \Phi_B^m$, therefore misfits from different observation sets cannot be compared directly. This is due to the fact that observation sets may utilize different frequencies, data points and even number of data points. All these factors influences the magnitude of the misfit value making a direct comparison of misfits from different observation sets is meaningless because it does not directly imply a relationship between the model errors and misfits. The model errors from different experiments can be compared directly because it is independent of the aforementioned factors. It measures how close a recovered distribution is to the target distribution.

4.4 The gradient method

In this section the gradient method, as presented in chapter 3, is used to recover the relative permittivity distributions shown in figure 4.1 and 4.2. The gradient method uses the gradient of the misfit surface as a search direction to find the minimum of the misfit surface.

For the following results, a starting model of a homogeneous cube with relative permittivity of 3.0 is used. The homogeneous model is chosen due to its simplicity. The value 3.0 was chosen to give the gradient method the best chance to find the actual permittivity distribution for the large anomaly case. In chapter 3, it was shown that the gradient method works best when the starting model was near the target distribution.

Table 4.3 Misfit and model error values for the gradient method.					
Illumination geometries	Observation frequencies	Misfit/ N_f small anomaly	Model Error small anomaly	Misfit/ N_f large anomaly	Model Error large anomaly
I1	F1	1.6737	325.27	0.02255	189.01
I1 to I2	F1	0.2543	227.83	0.3958	146.03
I1 to I3	F1	0.7161	219.78	0.6442	123.25
I1 to I4	F1	0.7432	217.49	0.9642	120.59
I1 to I5	F1	0.6814	213.67	5.8278	103.70
I1 to I6	F1	0.5072	206.98	2.8943	105.79
I1	F1 to F2	1.9233	344.01	0.1979	280.27
I1	F1 to F3	0.4080	279.64	0.2148	224.20
I1	F1 to F4	3.5020	291.96	0.2411	174.96
I1	F1 to F5	0.8766	267.73	1.1656	226.30
I1	F1 to F6	1.7487	270.29	0.4871	190.65

Table 4.3 summarizes the final results for the gradient method as applied to the two anomaly models. The misfit values presented are normalized with respect to the number of data sets. For this particular example, the selection of more illumination geometries at a single frequency yielded better results than the inclusion of more frequencies because the model error is smaller for the cases with more illumination geometries. None of the results produced a good recovered model in terms of the

model error. However, the inclusion of more data leads to smaller model errors in general, with some exceptions. The gradient method's convergence to a local minimum was expected, since, as mentioned in the chapter 3, the gradient method tend to seek out local minima. Neither the inclusion of more illumination angles or more illumination frequencies can be said to be better given how close the results are. Neither case solved the local minima problem. If there were no local minima or if the local minima only surrounded the global minimum, there would have been smaller misfit values.

Figures 4.4 to 4.7 plots the data misfit as a function of iteration for the recoveries performed with the gradient method.
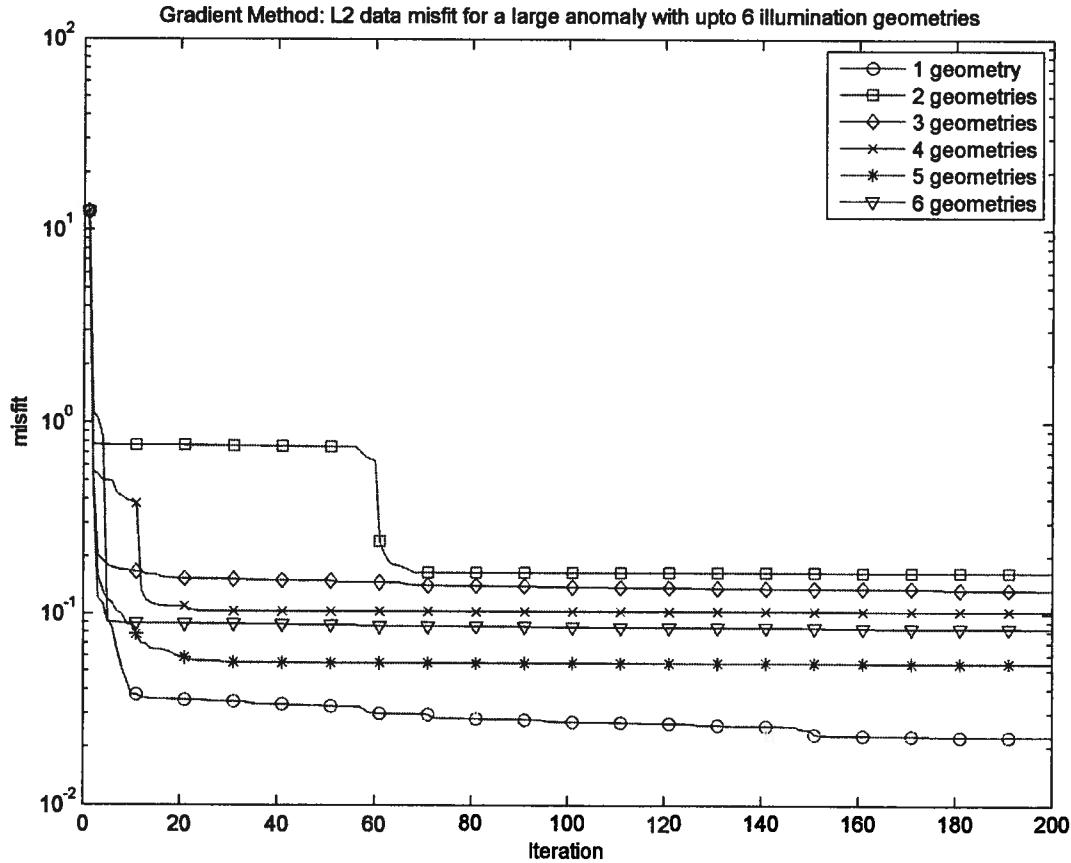
Figure 4.4: Misfit curves for the recovery of the large anomaly case using the gradient method with up to 6 illumination geometries and 1 illumination frequency.

Figures 4.4 shows the misfit curves for the recovery of the large anomaly case using up to 6

illumination geometries and 1 illumination frequency. The misfit curves seen in figure 4.4 converge to local minima. The convergence to a local minimum is characterized by a constant misfit value as a function of iteration. This can be seen in the 2 illumination geometries misfit curve in figure 4.4. From iterations 70 to 200 the misfit value is constant. The other misfit curves in figure 4.4 also exhibit the same behavior. The 2 illumination geometries misfit curve has two plateau regions. The first plateau seen in Figure 4.4 looks like it is constant, but it changes by a small amount after each iteration. With enough of these small changes, the algorithm was able to find a point on the surface that led to a local minimum. This suggests a richness in the topography of the misfit surface. All the misfit curves in figure 4.4 exhibit a large initial drop, which is characteristic of the gradient method.
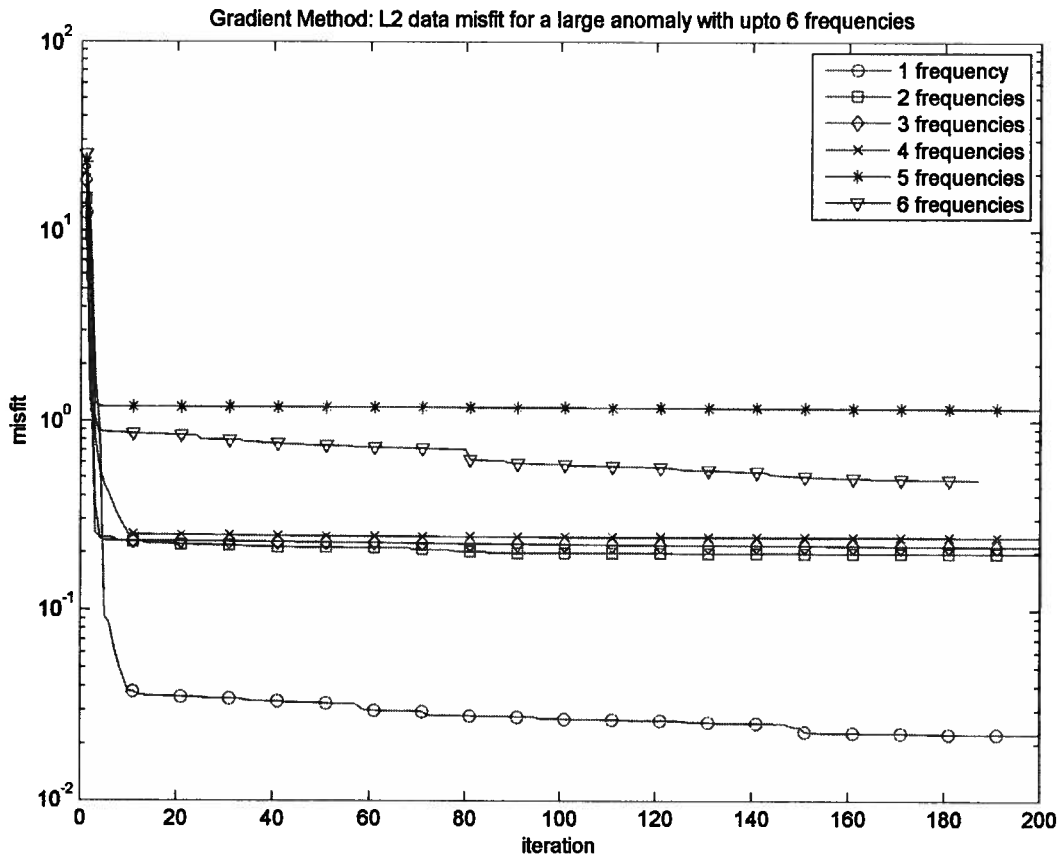
Figure 4.5: Misfit curve for the recovery of the large anomaly case using the gradient method with up to 6 illumination frequencies and 1 illumination geometry.

Figure 4.5 shows the misfit curves for the recovery of the large anomaly case using up to 6

illumination frequencies. The misfit curves in figure 4.5 exhibit the same behavior as those seen in

figure 4.4. There is a large initial drop in misfit, followed by a long plateau region. The ordering of the

final misfit value is different between that of figure 4.4 and 4.5. The solution found by the gradient

method is largely dominated by the locations of the local minima. The addition of more data does not

eliminate these local minima. The ordering of the final misfit values is then a result of the locations of

the local minima and not of the amount of data because the gradient method seeks out local minima.
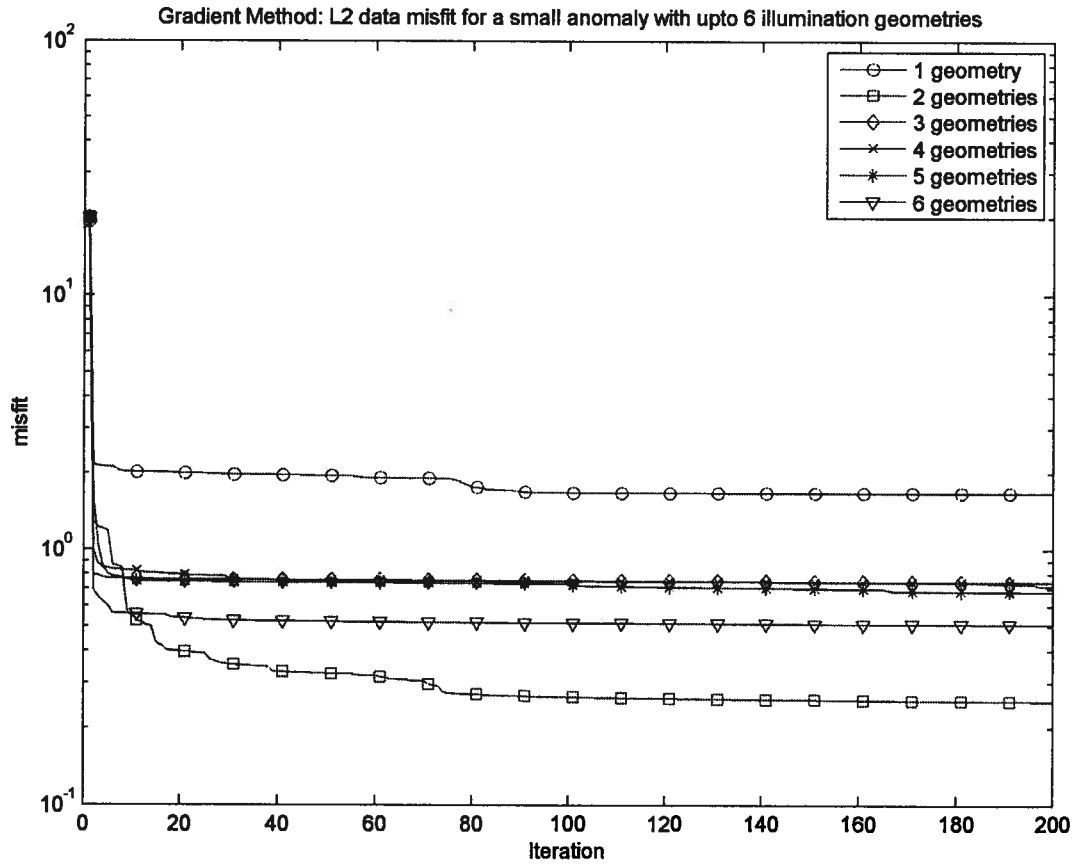
Figure 4.6:Misfit curves for the recovery of the small anomaly case with the gradient method using up to 6 illumination geometries and 1 illumination frequency.

Figure 4.6 shows the misfit curves for the recovery of the small anomaly case using up to 6

illumination geometries and 1 illumination frequency. The misfit curves in figure 4.4 exhibit the same

behavior as the misfit curves in figure 4.4. The misfit curves has their greatest drop during the first

iteration and eventually reach a local minimum. The case of 2 illumination geometries and 1

illumination frequency demonstrates a staircase behavior between iterations 1 and 80. This staircase

behavior is likely due to the gradient algorithm jumping from a region associated with one local

minimum to a region associated with another local minimum.
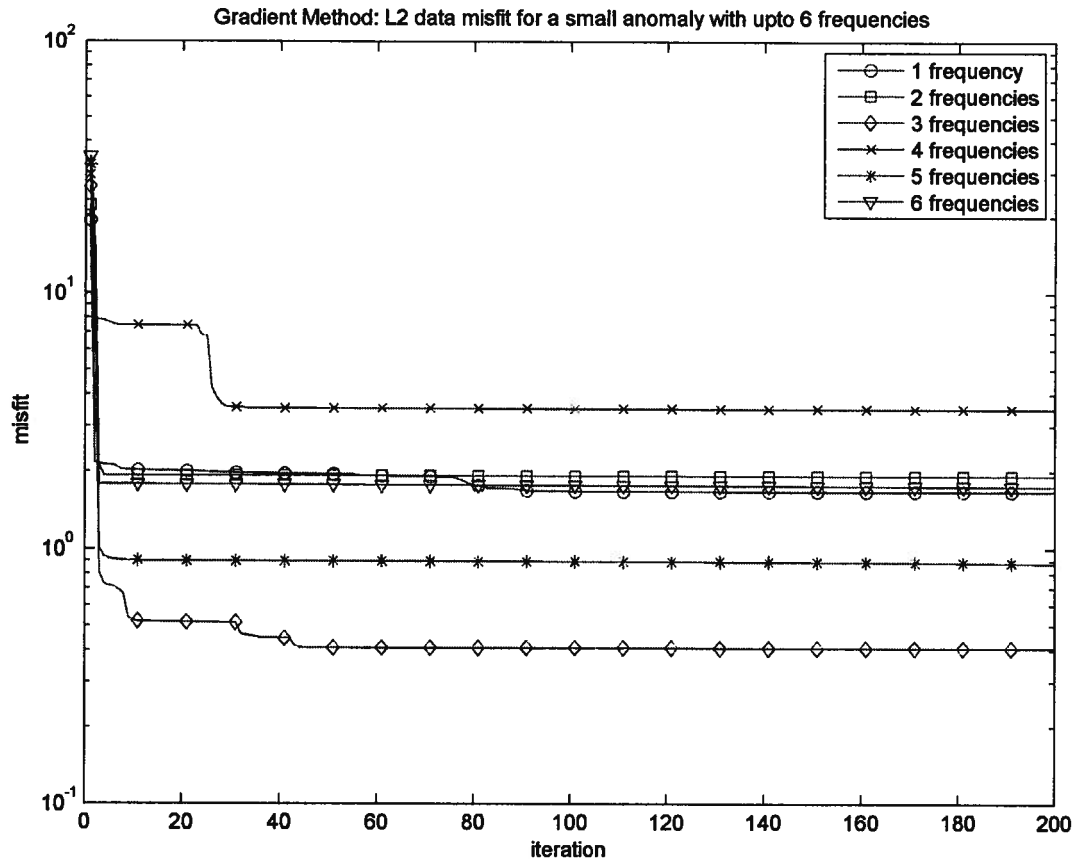
Figure 4.7:Misfit curves for the gradient method recovering the small anomaly case using up to 6 illumination frequencies and 1 illumination geometry.

Figure 4.7 shows the misfit curves for the recovery of the small anomaly case using up to 6

illumination geometries and 1 illumination frequency. The misfit curves in figure 4.7 exhibit the same

behavior as seen in previous figures. The case of 2 illumination frequencies and 1 illumination

geometry has the same double plateau behavior similar to the case of 2 illumination geometries and 1

illumination frequency case for the recovery of large anomaly as seen in figure 4.4. The misfit curves

in figure 4.7 have a large initial drop in misfit and eventually converge to a local minimum.

Figure 4.8: Model error curves for the gradient method recovering the large anomaly case using up to 6 illumination geometries and 1 illumination frequency.

Figure 4.8 shows the model error curves for the recovery of the large anomaly case using up to

6 illumination geometries. The model error curves exhibit the same behavior as that of the misfit

curves in figure 4.4. The model error curves have a steep initial drop and eventually reach a constant

value. For the case of the two illumination geometries, the double plateau is visible in both the model

error curve and the misfit curve. The difference in the scaling between model error and misfit is also

clearly visible. For the case of 1 illumination geometry, the model error decreases by less than one

order of magnitude, however the misfit in figure 4.4 decreases by approximately 3 orders of magnitude.
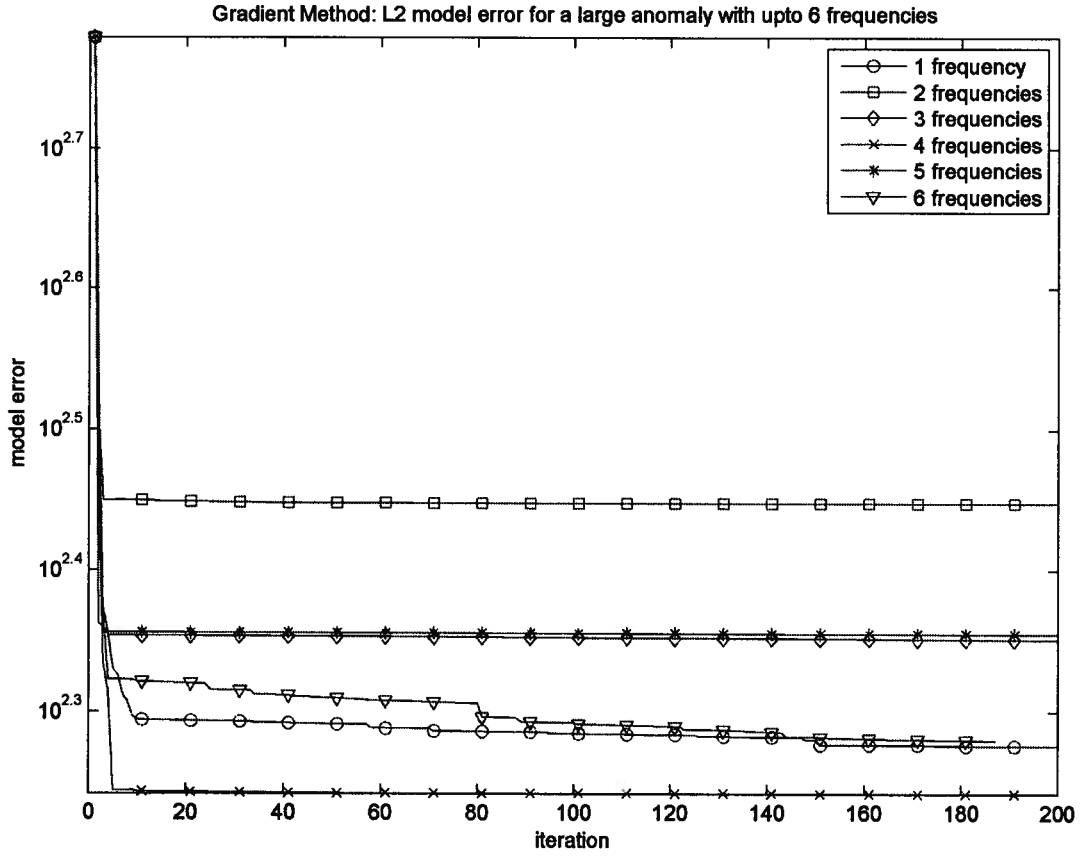
Figure 4.9: Model error curves for the gradient method recovering the large anomaly case using up to 6 illumination frequencies and 1 illumination geometry.

Figure 4.9 shows the model error curves for the recovery of the large anomaly case using up to

6 illumination frequencies and 1 illumination geometry. All the curves have a large initial drop in

misfit and eventually reach a constant value. The ordering of the final values of the model error curves

in figure 4.9 is different than that of the ordering of the misfit curves in figure 4.5. In figure 4.9 the

case using 4 illumination frequencies has the smallest model error, however in figure 4.5 the case using

1 illumination frequencies has the smallest misfit. This reinforces the fact that the misfit values

between different experiments cannot be compared directly, as discussed in section 4.2. This same

phenomenon is also visible when comparing figure 4.8 and figure 4.4, which plots the misfit and model

error curves for the recovery using up to 6 illumination geometries, respectively. Therefore the

behavior is not an artifact of using more frequencies or using more illumination geometries.
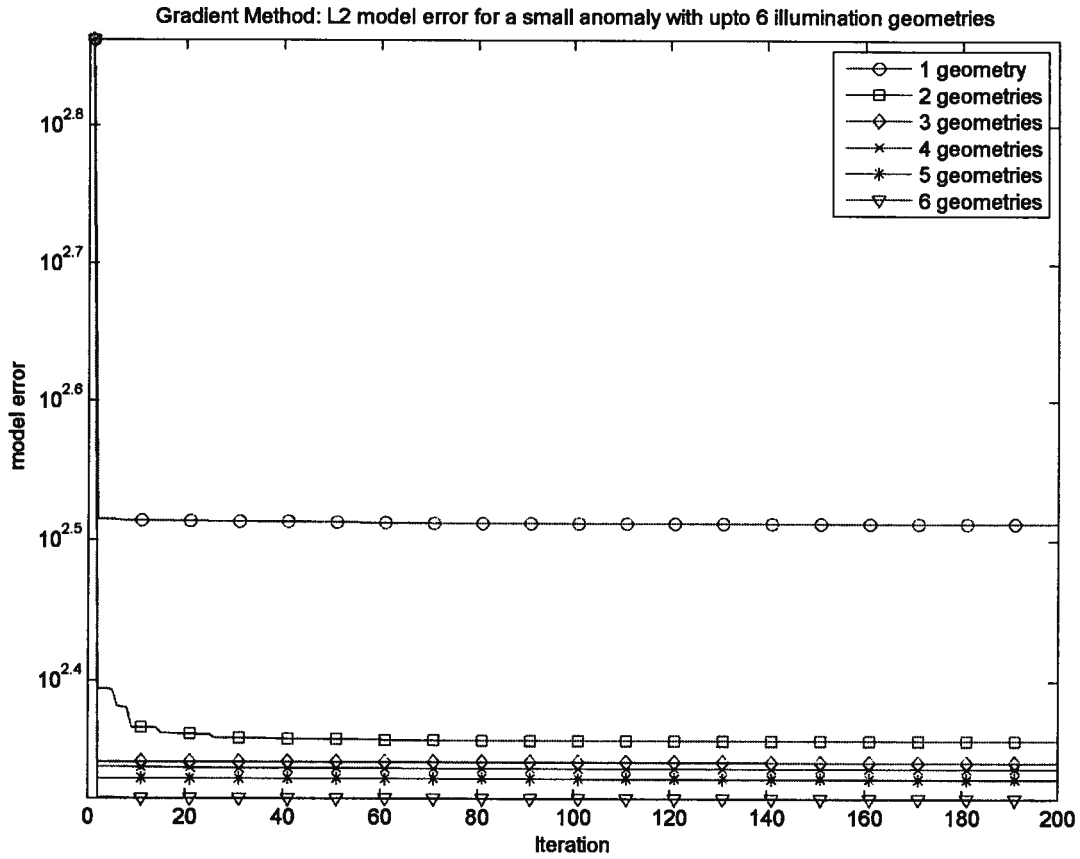
Figure 4.10: Model error curves for the gradient method recovering the small anomaly case using up to 6 illumination geometries and 1 illumination frequency.

Figure 4.10 shows the model error curves for the recovery of the small anomaly case using up to 6 illumination geometries and 1 illumination frequency. The same behavior observed for the misfit curves in figure 4.6 is also present in the model error curve in figure 4.10. All the curves exhibit a large initial decrease and converge to local minima. The staircase behavior visible in the misfit curve for the case of 2 illumination geometries in figure 4.6 is also visible in the corresponding model error curve in figure 4.10.
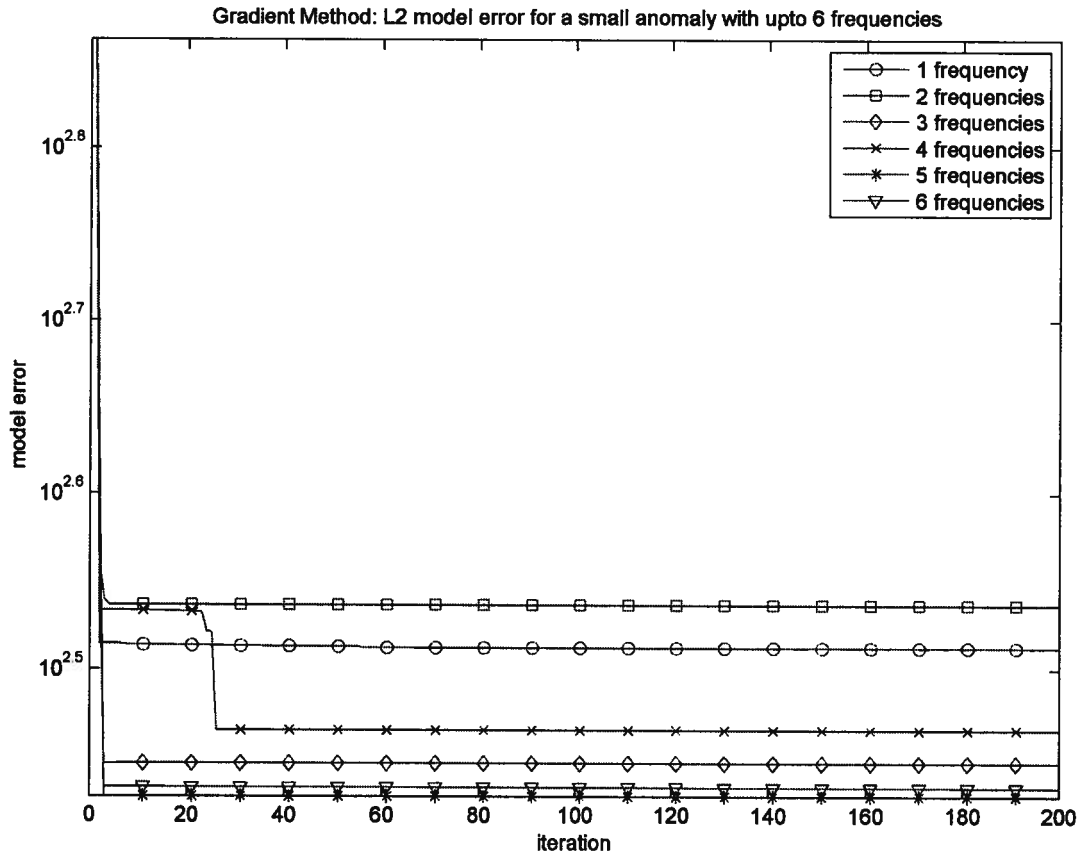
Figure 4.11: Model error curves for the gradient method recovering the small anomaly case using up to 6 illumination frequencies and 1 illumination geometry.

Figure 4.11 shows the model error curves for the recovery of the small anomaly case using up to

6 illumination frequencies and 1 illumination angle. The same behavior for the misfit curve in figure

4.7 is also observed in the model error curve in figure 4.11. All the curves exhibit a large initial

decrease and eventually converge to a constant value.

Gradient Method: Relative permittivity distribution after 1 iteration for the recovery of the large anomaly case
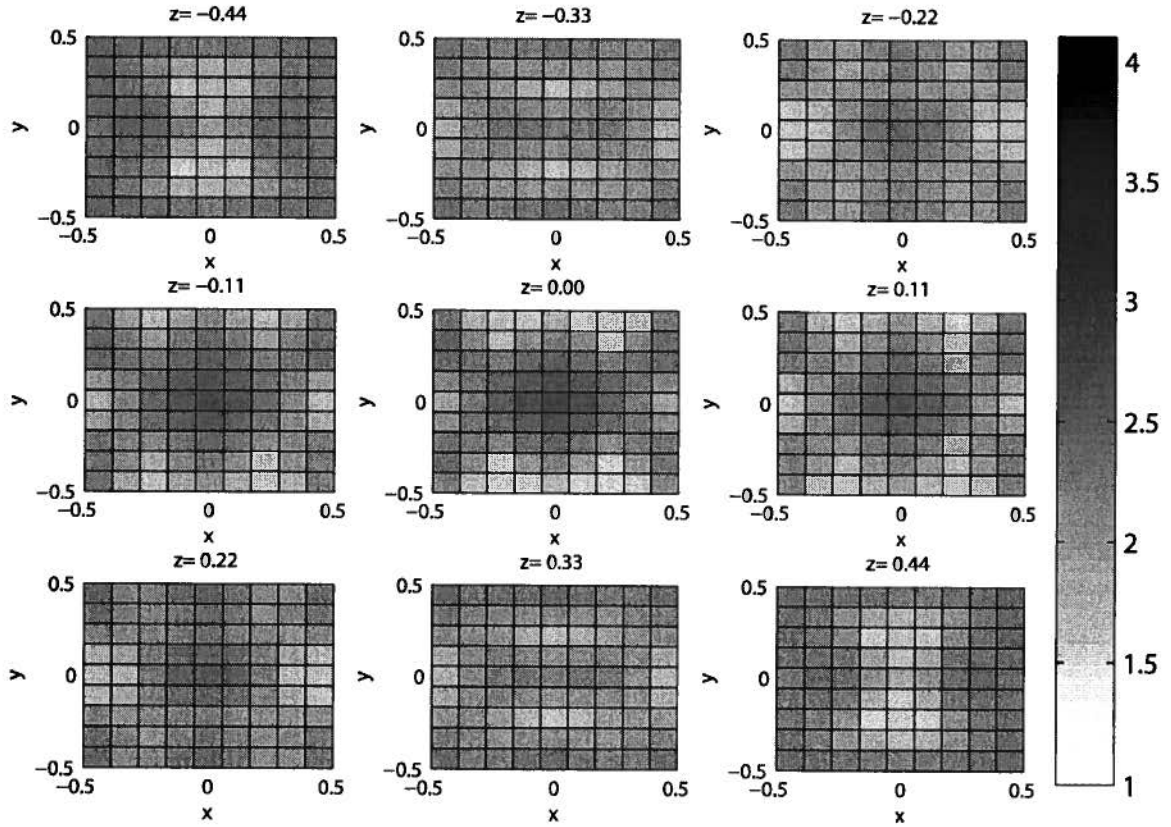
Figure 4.12: The relative permittivity distribution after 1 iteration for the gradient method while recovering the large anomaly case using 6 illumination geometries and 1 illumination frequency.

Figure 4.12 shows the relative permittivity distribution for the recovery of the large anomaly case, using 6 illumination geometries and 1 illumination frequency after the first iteration. The recovery started with a homogeneous relative permittivity distribution with a relative permittivity value of 3.0. After one iteration many features have already appeared in the permittivity distribution. This corresponds to the large decrease in misfit seen in figure 4.4.

114

Gradient Method: Relative permittivity distribution after 200 iterations for the recovery of the large anomaly case

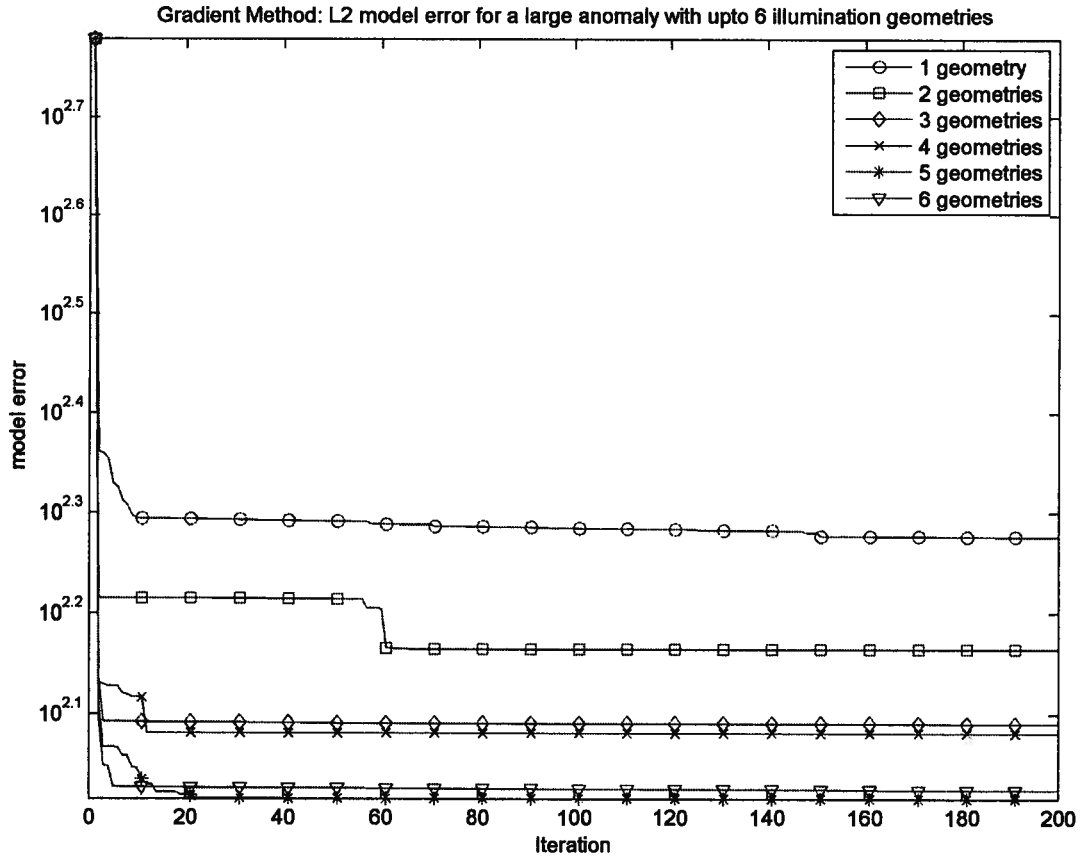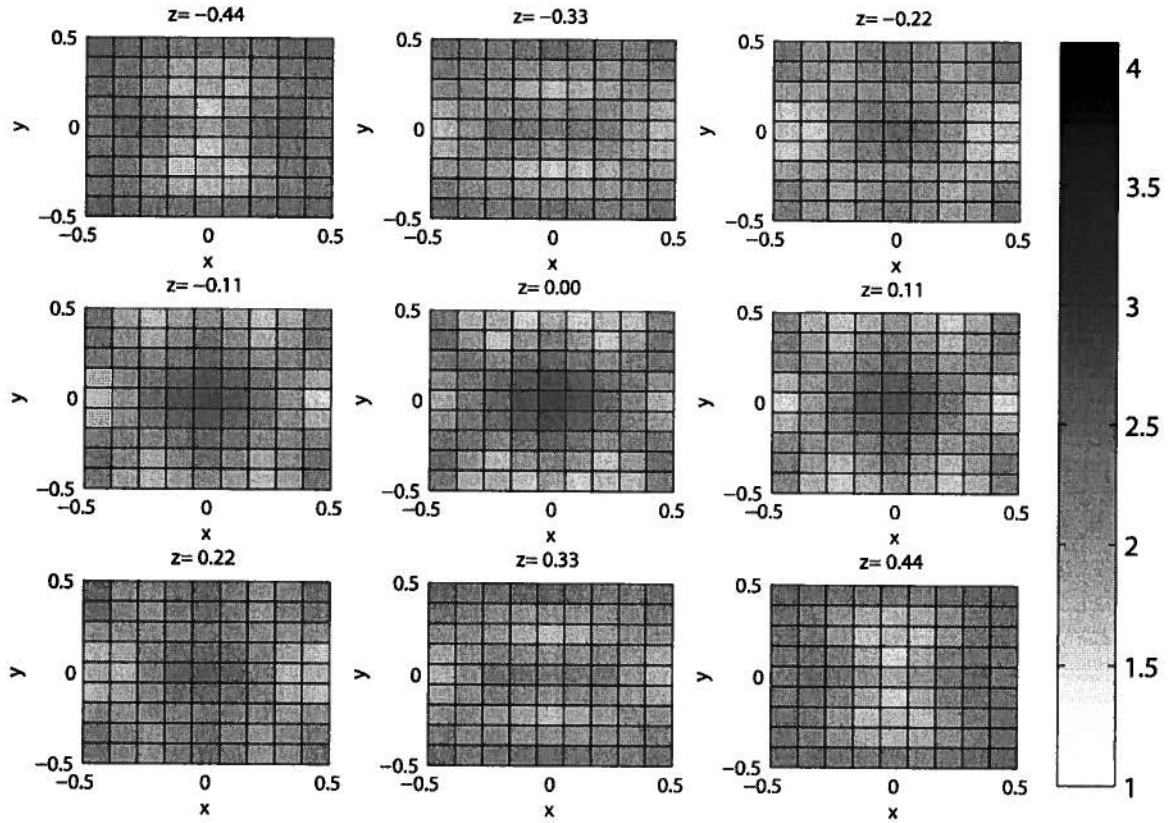*Figure 4.13: The relative permittivity distribution after 200 iterations for the gradient method while recovering the large anomaly case using 6 illumination geometries and 1 illumination frequency.*

Figure 4.13 shows the relative permittivity distribution for the recovery of the large anomaly case, using 6 illumination geometries and 1 illumination frequency after 200 iterations. The recovered relative permittivity distribution at the 200th iteration is very similar to the relative permittivity distribution after the first iteration. The result agrees with the fact that misfit has its largest decrease after the first iteration, as seen in figure 4.4. The recovered relative permittivity distribution has a large anomaly in the center which resembles the one seen in the target distribution, figure 4.1.

Gradient Method: Relative permittivity distribution after 200 iterations for the recovery of the small anomaly case
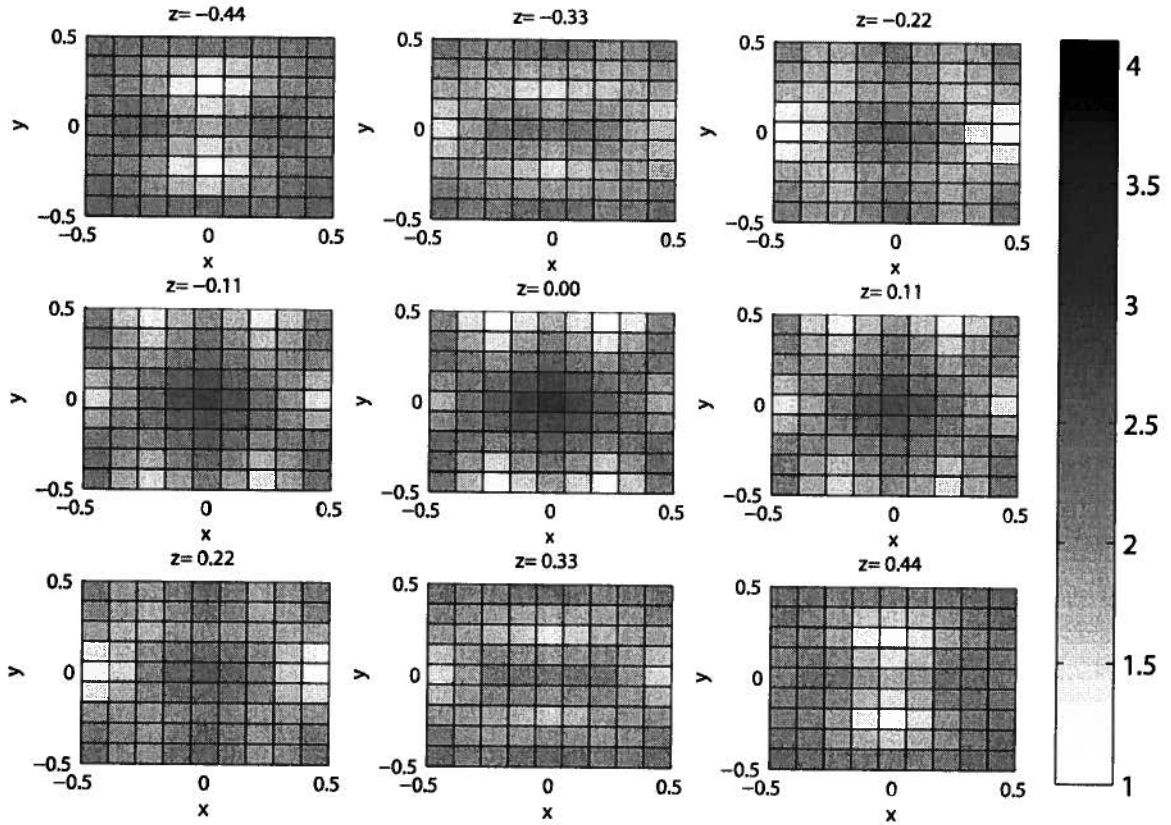
Figure 4.14: The relative permittivity distribution after 200 iterations for the gradient method while recovering the small anomaly case using 6 illumination geometries and 1 illumination frequency.

Figure 4.14 shows the relative permittivity distribution for the recovery of the small anomaly case, using 6 illumination geometries and 1 illumination frequency after the 200th iteration. Figure 4.14 is similar to figure 4.13, the large anomaly case. Both contain a blurry anomaly in the center of the cube. This indicates that the recoveries have failed because they are the recovered relative permittivity distributions for dissimilar targets. The similarity in figure 4.14 and figure 4.13 can be explained by the fact that both cases uses the same starting model, a homogeneous relative permittivity of 3.0, and both cases have a background of 2.0, with a higher permittivity object embedded inside. Because they share such similarity, it is not surprising that the initial gradient is similar. Since most of

116

the structure is recovered in the first iteration, two similar distributions were recovered that does not represent either target distribution very well.

The performance of the gradient algorithm is poor and the resulting images have some resemblance to the actual distribution for the recovery of the large anomaly but not enough to warrant this as a good recovery algorithm. The recovery of the small anomaly has clearly failed. The recovered permittivity distributions do not match the target distributions because of the local minima around the starting model.

The performance of the gradient method is dependent on the local properties of the misfit surface. The model error of the recovery of the large anomaly case using 4 illumination frequencies with the value of 174.96 is smaller than the model error for the case using 5 illumination frequencies with the value of 226.30. Typically, when more data are added the performance is expected to improve. Because the performance of the gradient algorithm is dependent on the local properties of the misfit surface, the performance of the gradient method is not solely dependent on the amount of data.

The gradient method can provide an estimate of the topography of the surface, because of it's simplicity. Interpreting the trajectory of the gradient method using the misfit curve provides an intuitive sense of the behavior of the misfit surface.

4.5 Levenberg-Marquardt

In this section the Levenberg-Marquardt(LM) method, as presented in chapter 3, is used to recover the relative permittivity distributions shown in figure 4.1 and 4.2. The LM method uses the perturbation term as a regularization term.

Table 4.4: Parameters used for the LM algorithm	
β_0^{LM}	3.0
ν	0.95

There are 2 parameters required for the LM method, they are shown in Table 4.4. One parameter is the initial weighting of the perturbation. The second parameter controls how the weighting changes. Differing parameters may lead to a convergence to a different relative permittivity distribution because a combined functional is minimized and the parameters influence the combined functional. The starting model for the LM method is the same as the one used for the gradient method. A homogeneous distribution with a relative permittivity of 3.0 is used. This provides a fair comparison between the two algorithms.

Table 4.5: Results for the LM method recovering a small and large anomaly					
Illumination geometries	Observation frequencies	Misfit/ N_f small anomaly	Model Error small anomaly	Misfit/ N_f large anomaly	Model Error large anomaly
I1	F1	2.86E-004	297.73	1.93E-003	204.05
I1 to I2	F1	1.03E-004	436.11	7.95E-005	224.48
I1 to I3	F1	5.81E-005	155.25	3.48E-005	332.18
I1 to I4	F1	5.45E-005	150.69	1.50E-008	43.12
I1 to I5	F1	2.17E-004	309.28	2.39E-006	131.99
I1 to I6	F1	1.38E-005	425.43	1.19E-005	237.85
I1	F1 to F2	1.91E-004	442.84	4.74E-004	301.5
I1	F1 to F3	3.32E-002	359.06	2.13	1039.0
I1	F1 to F4	6.78E-003	893.50	6.90E-003	368.5
I1	F1 to F5	1.85	1056.6	2.62E-003	512.73
I1	F1 to F6	5.42E-002	1744.7	1.68E-001	1744.7

Table 4.5 summarizes the results for the LM method. Once again, similar to the gradient

118

method, the misfit value is normalized with respect to the number of observation sets. The LM method produces smaller misfit values than that of the gradient method. However, when comparing the model error, the LM method does not always produce a smaller value. For the case of 1 illumination frequency and 1 illumination geometry for the recovery of a small anomaly the LM method produced a smaller model error and a smaller misfit than that of the gradient method. The LM method produced a model error of 297.7 and the gradient method produced a model error of 325.2. There are other experiments where the LM method produced a smaller model error than that of the gradient method. However, there are examples where the LM method has a larger model error than the gradient method, even though the misfit recovered by the LM method was smaller. One such case is the experiment using 1 illumination geometry and 6 illumination frequencies for the recovery of the small anomaly case. The misfit value recovered by the LM method was 0.054 compared to the gradient method's 1.75. The model error for the LM method however is much larger than that of gradient method. The model error for the LM method is 1744.7 while the model error for the gradient method is 270.3. The performance of the LM method does reinforce the idea that the misfit and the model error are loosely connected. Clearly, the goal should not be solely focused on minimizing the misfit but on the recovery of a reasonable relative permittivity distribution.

Figure 4.15: Misfit curves for the recovery of the large anomaly case with the Levenberg-Marquardt method using up to 6 illumination geometries and 1 illumination frequency.

Figure 4.15 shows the misfit curves for the large anomaly case with the LM method using up to

6 illumination geometries and 1 illumination geometry. The curves are converging to a local minimum.

Like the gradient method, the misfit has the greatest decrease at the early iterations.

For the 2 illumination geometries case, notice the sudden change in slope at approximately

iteration 120. This change of slope occurs in the other curves as well. One possible cause of this

behavior is a saddle point on the misfit surface.

The saddle point behavior will be illustrated with an example with two independent parameters.
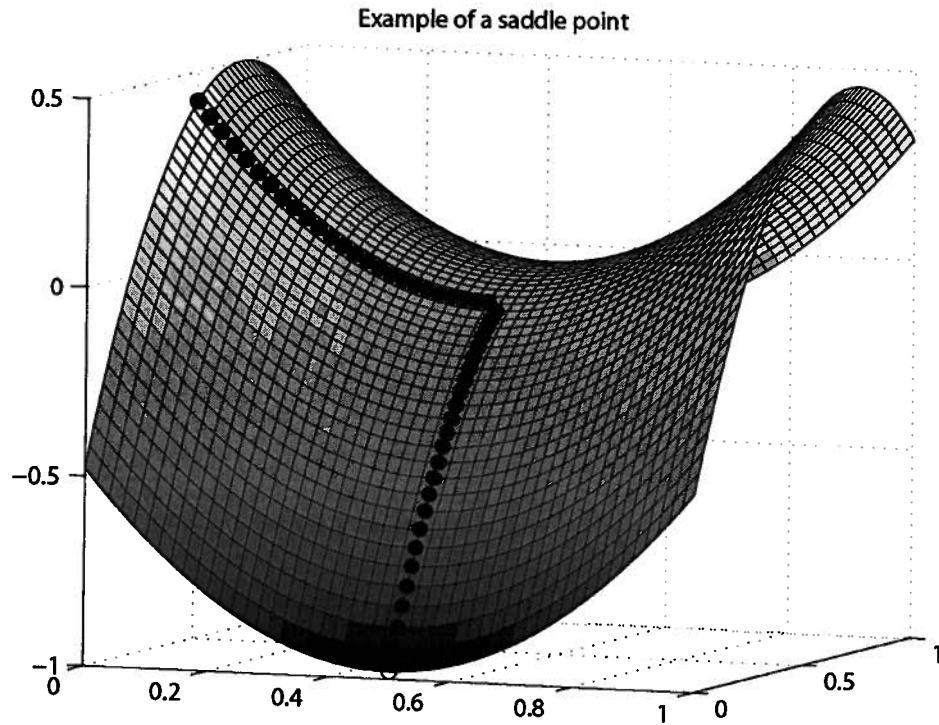
Figure 4.16: Example of a saddle point with two independent variables. Note the third dimension is used to plot the functional value. The black dotted path, represents a possible path in which the LM may take on the saddle shape Note the sharp change in direction.

Figure 4.16 shows a saddle shape with 2 independent parameters. The black dotted line indicate

a possible path the LM algorithm may take on the saddle's surface. The line changes direction when it

nears the saddle point.
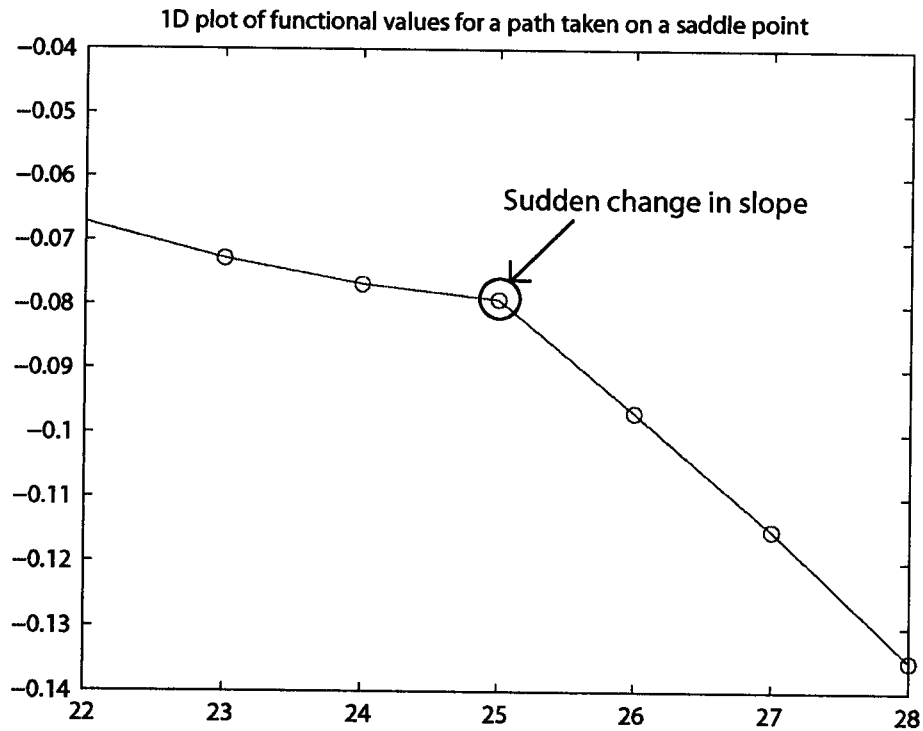
Figure 4.17: A close up one dimensional plot of the misfit values for a possible path taken by the LM method near a saddle point. Note the sudden change in slope at x=25. The x-axis represents a linear parametrization of the line.

Figure 4.17 shows a close up of the black dotted line shown in figure 4.16 plotted in one

dimension. The most apparent feature of the line is the sudden change in slope. This feature on a

misfit curve indicates the existence of a saddle point on the misfit surface. The same transition in slope

is also expected to exist in the equivalent of a saddle point in higher dimensions.

The curves in figure 4.15 never reach a minimum. The most apparent example occurs in the

case of 4 illumination geometries and 1 illumination frequency. The slope at the 400th iteration is not 0.

If the LM method had reached the local minimum, the slope would be 0.

The reason for this behavior is the addition of the regularization. Because of the definition of

the L2 misfit functional in (3.13), the shape of the functional near a minimum is a parabola. The

regularization surface for the LM method is also a parabola, centered around the expansion point.

When two parabolas are added, a third parabola is produced where location of the minimum is

122

between the location of the minima of the two original parabolas. This is shown in figure 4.18. Notice that the value of the minimum of the combined functional is greater than that of the constituent functionals.
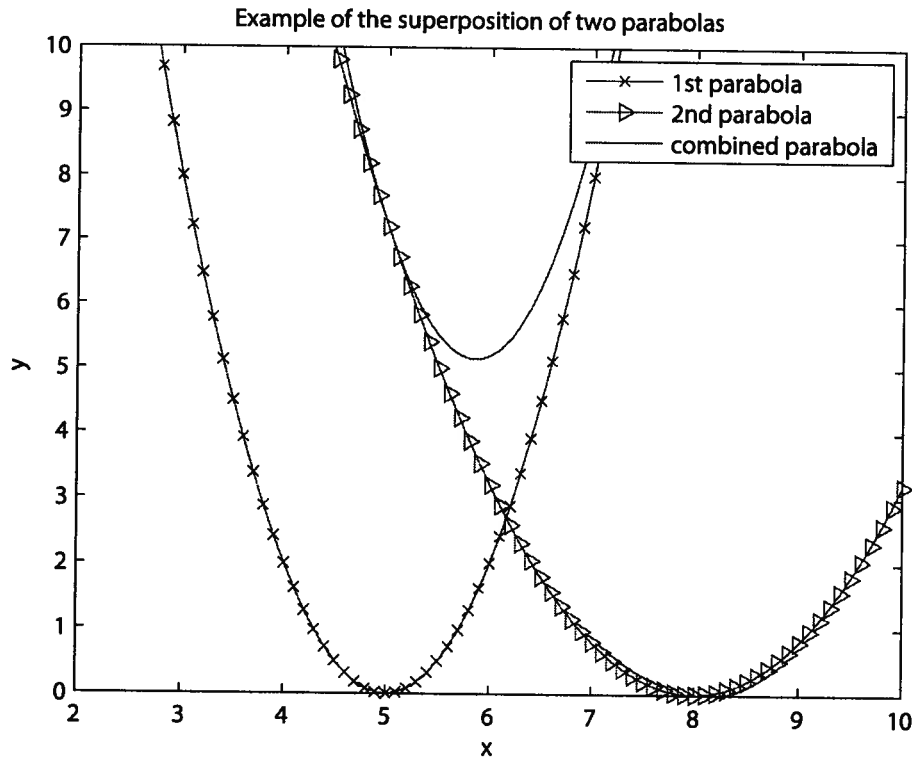
Figure 4.18: Example of the superposition of two parabolas. The local minimum of the combined functional lies between the local minima of the two constituent functions. Also note that the local minimum's value is higher than that of the constituent functions.

Theoretically the LM method can take an infinite number of iterations before the minimum is reached because the LM method operates on the combined functional. The LM misfit curves are expected to constantly decrease. However the size of the decrease shrinks with each iteration. Therefore the stopping condition for detecting a local minimum would perform poorly on the LM algorithm because the slope of the misfit curve is never 0.
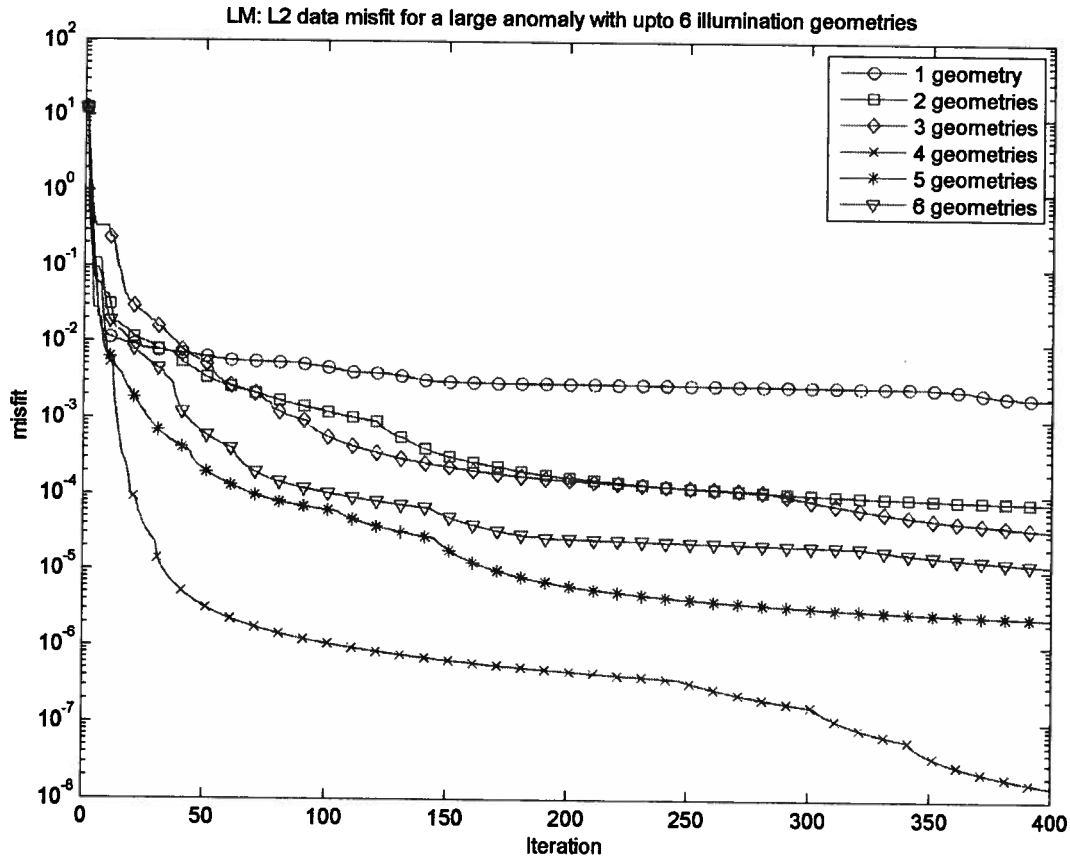
Figure 4.19: Misfit curves for the recovery of the large anomaly case with the Levenberg-Marquardt method using up to 6 illumination frequencies and 1 illumination geometry.

Figure 4.19 shows the misfit curves for the LM method recovering the relative permittivity distribution for the large anomaly case, using up to 6 illumination frequencies and 1 illumination geometry. The misfit curves in figure 4.19 exhibit the the same behavior as the curves seen in figure 4.15. The curves decrease as a function of iteration, as is guaranteed by the algorithm. The sudden change in slope can be seen in the 2 illumination frequencies case at approximately iteration 70. The fact that the curves never reaches the minimum can be seen in the 4 illumination frequencies case at the 400th iteration where the slope is not 0. Like the gradient method, the misfit has the greatest decrease at the early iterations.

Figure 4.20: Misfit curves for the recovery of the small anomaly case with the Levenberg-Marquardt method using up to 6 illumination geometries and 1 illumination frequency.

Figure 4.20 shows the misfit curves for the LM method recovering the relative permittivity

distribution for the small anomaly case, using up to 6 illumination geometries and 1 illumination

frequency. The misfit curves in figure 4.19 exhibit the the same behavior as the curves seen in figure

4.15 and 4.19. Like the gradient method, the misfit has the greatest decrease at the early iterations.

The curves decrease as a function of iteration, as is guaranteed by the algorithm. The sudden change in

slope can be seen in the 4 illumination geometries case at approximately iteration 170. The fact that the

curves never reaches the minimum can be seen in the 4 illumination geometries case at the 400th

iteration where the slope is not 0. In fact, the slopes of all the curves are non-zero at the 400th

iteration, implying that none of the experiments reached a minimum.

Figure 4.21: Misfit curves for the recovery of the small anomaly using the Levenberg-Marquardt method with up to 6 illumination frequencies and 1 illumination geometry.

Figure 4.21 shows the misfit curves for the LM method recovering the relative permittivity

distribution for the small anomaly case, using up to 6 illumination frequencies and 1 illumination

geometry. The misfit curves in figure 4.21 exhibit the the same behavior as the curves seen in figures

4.15, 4.19 and 4.20. The curves decrease as a function of iteration. The sudden change in slope can be

seen in the 6 illumination frequencies case at approximately iteration 210. The fact that the curves

never reaches the minimum can be seen in all the cases at the 400th iteration.

Figure 4.22: Model error curves for the recovery of the large anomaly cause using the Levenberg-Marquardt method with up to 6 illumination geometries and 1 illumination frequency.

Figure 4.22 shows the model error curves for the LM method recovering the relative

permittivity distribution for the large anomaly case, using up to 6 illumination geometries and 1

illumination frequency. The model error curves has its largest decrease in the early iterations and

eventually converges to a final value, which mirrors the behavior in the misfit curves seen figure 4.15.
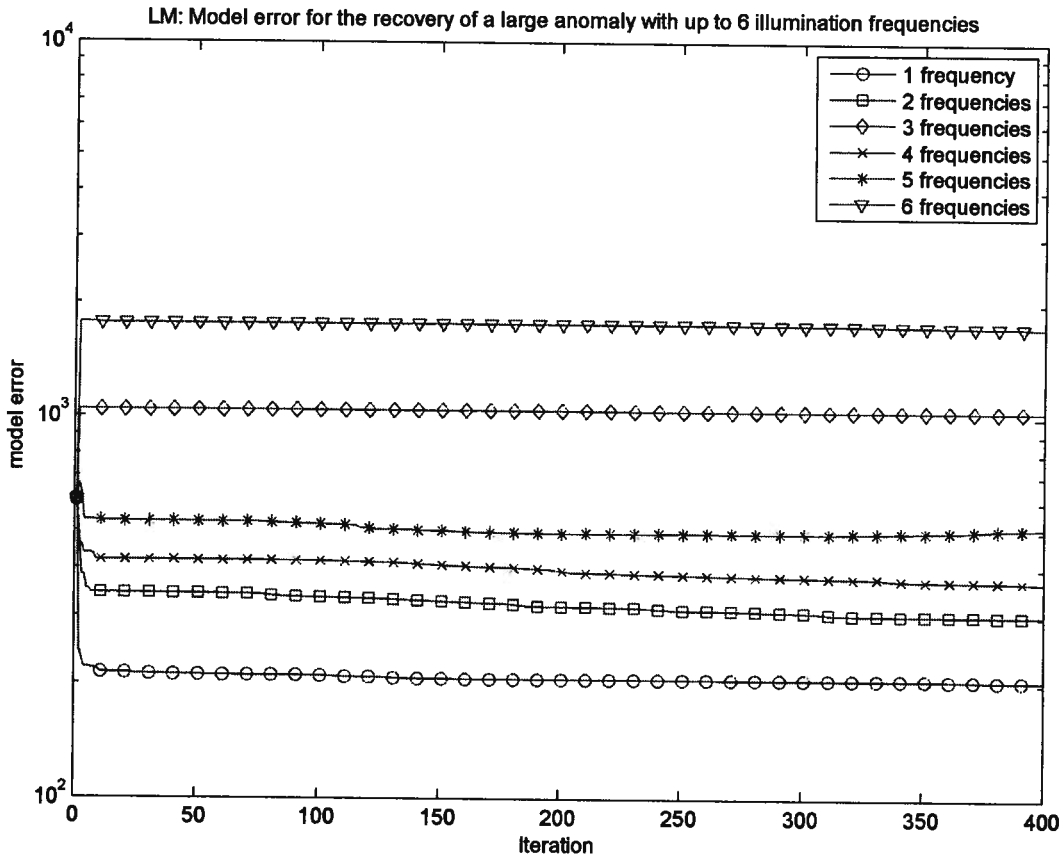
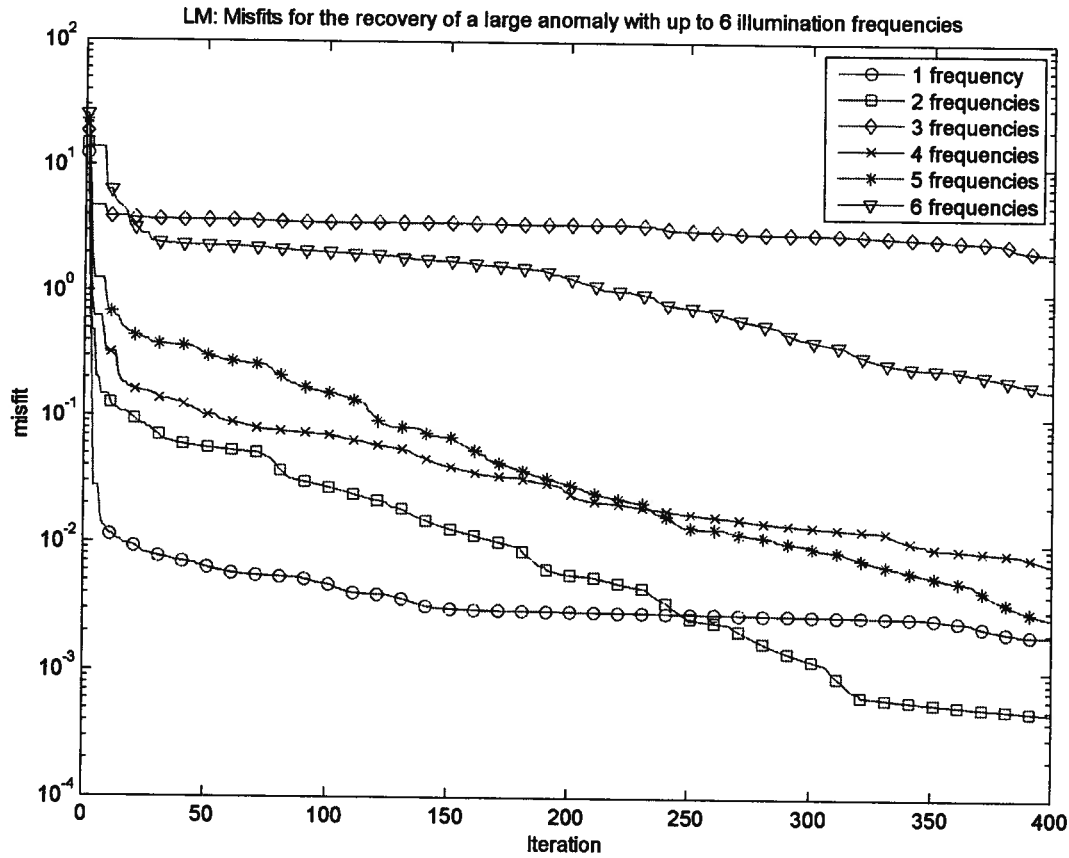*Figure 4.23: Model error curves for the recovery of the large anomaly case using the Levenberg-Marquardt method with up to 6 illumination frequencies and 1 illumination geometry.*

Figure 4.23 shows the model error curves for the LM method recovering the relative

permittivity distribution for the small anomaly case, using up to 6 illumination frequencies and 1

illumination geometry. For the 3 and 6 illumination frequencies cases, there is a sharp initial increase in

model error. There is no corresponding increase in the misfit curves seen in figure 4.19. The increase

in model error without an increase in misfit can be understood by examining the following one

dimensional example with 2 minima.
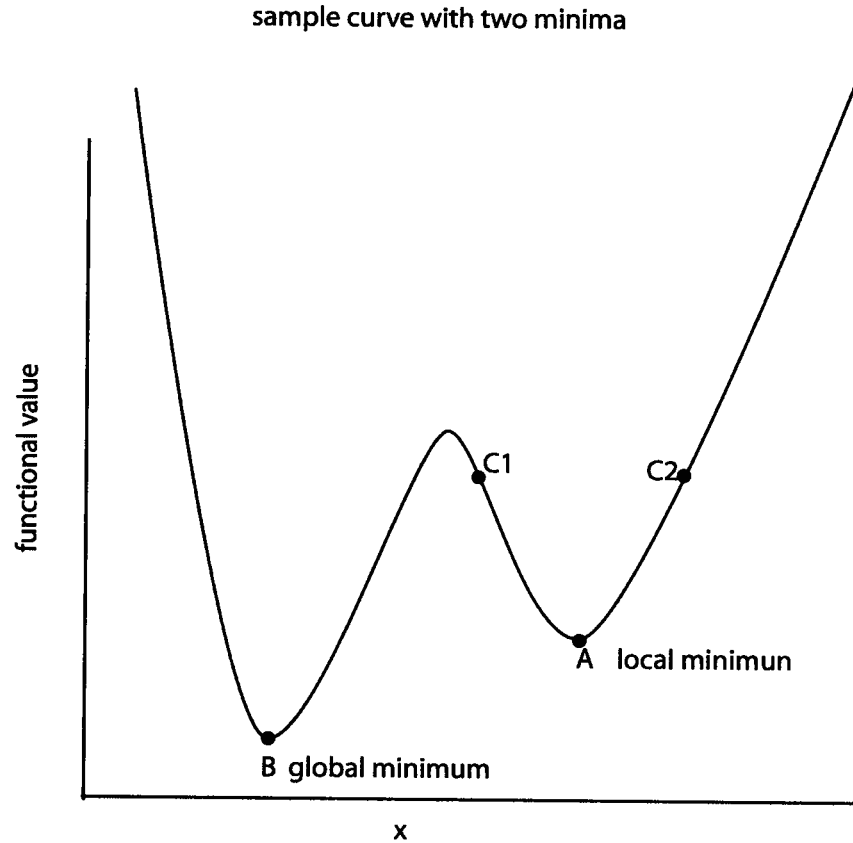
Figure 4.24: Example of a curve with two minima, one local and one global. The local minimum denoted as A. *The global minimum denoted as* B.

For the example shown in figure 4.24 the model error can be defined as

$$\Phi^m = \|B - x\|^2 .$$ (4.2)

Point B is the global minimum and point A is the local minimum. The goal is to reach B. However if the LM method is used and the starting point of C1 or C2 is used, then the algorithm will converge to A, because the LM method only uses local information and tend to seek out a local minimum. If A is approached from C2 both the functional value and the model error decreases. If A is approached from C1, the model error increases but the functional value decreases. The model error increases because the algorithm is getting further away from B. This type of behavior implies then that near a local minimum the condition that a smaller functional value corresponds to a smaller model error is not always true. The condition, that a smaller functional value corresponds to a smaller model error, is sufficiently true on a global scale, that it can be used to develop a recovery algorithm.
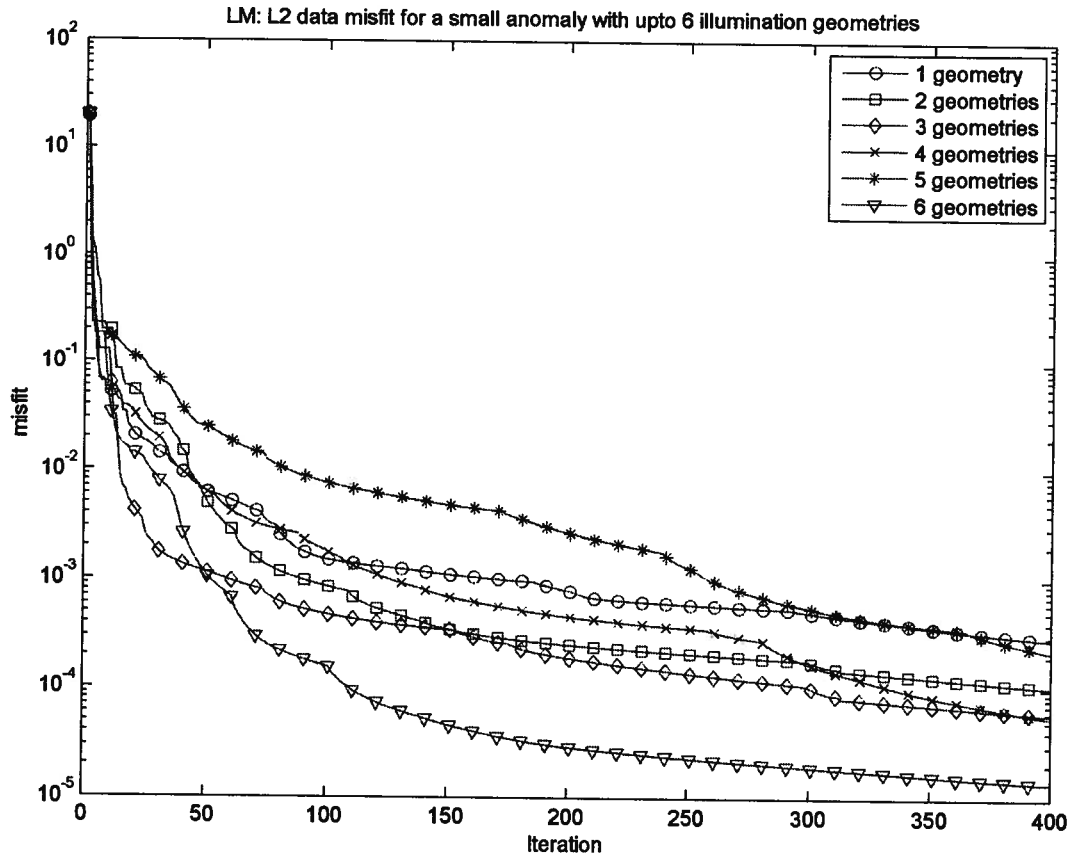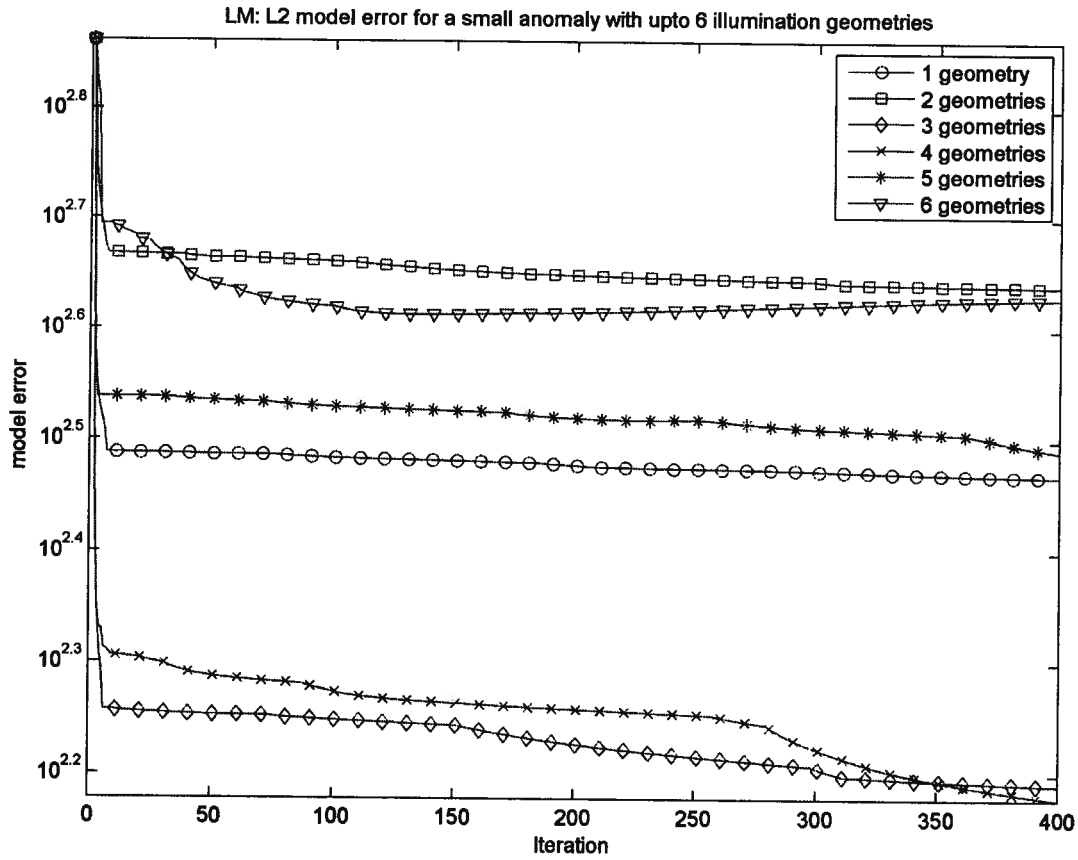
Figure 4.25: Model error curves for the recovery of the small anomaly case with the Levenberg-Marquardt method using up to 6 illumination geometries and 1 illumination frequency.

Figure 4.25 shows the model error curves for the recovery of the small anomaly case using the

LM method with up to 6 illumination geometries and 1 illumination frequency. The rise in model error

without the corresponding rise in misfit can be seen in the 6 illumination geometries case. From

iteration 150 to 400 the model error rises, though not as dramatic as the example seen in figure 4.23.

The model error curves exhibit a large initial decrease and eventually converges to a final value.

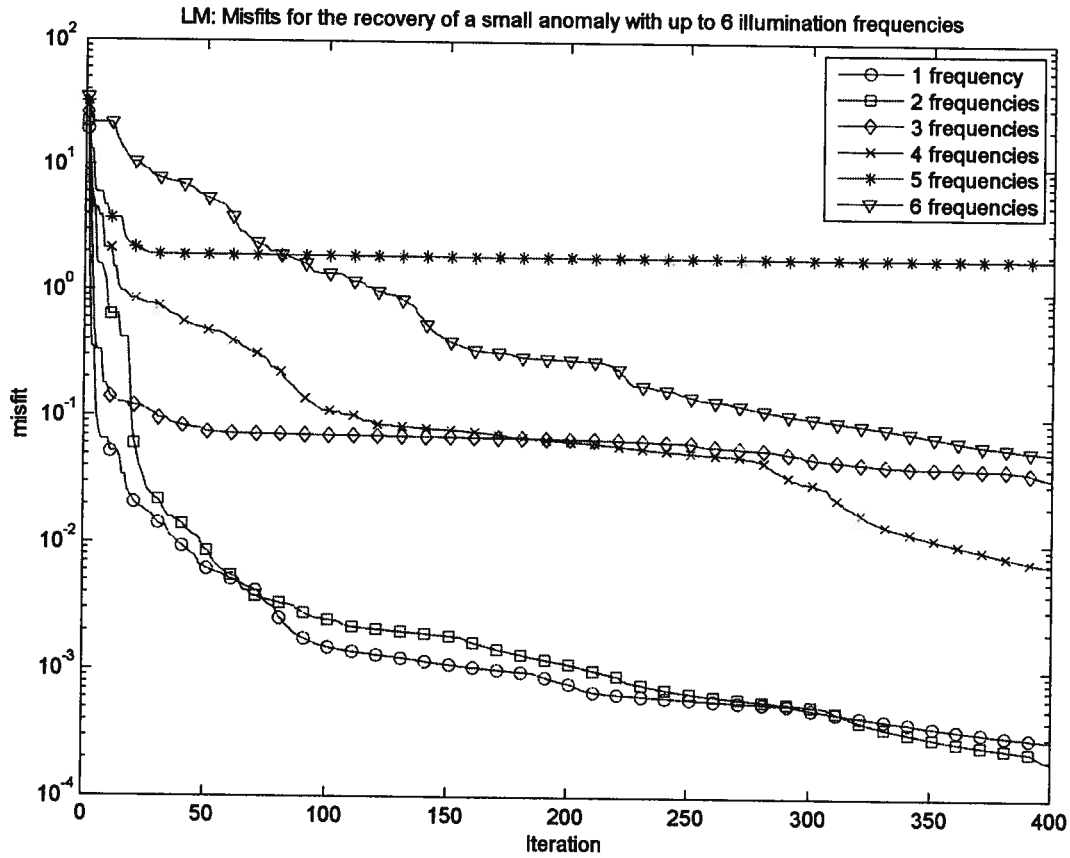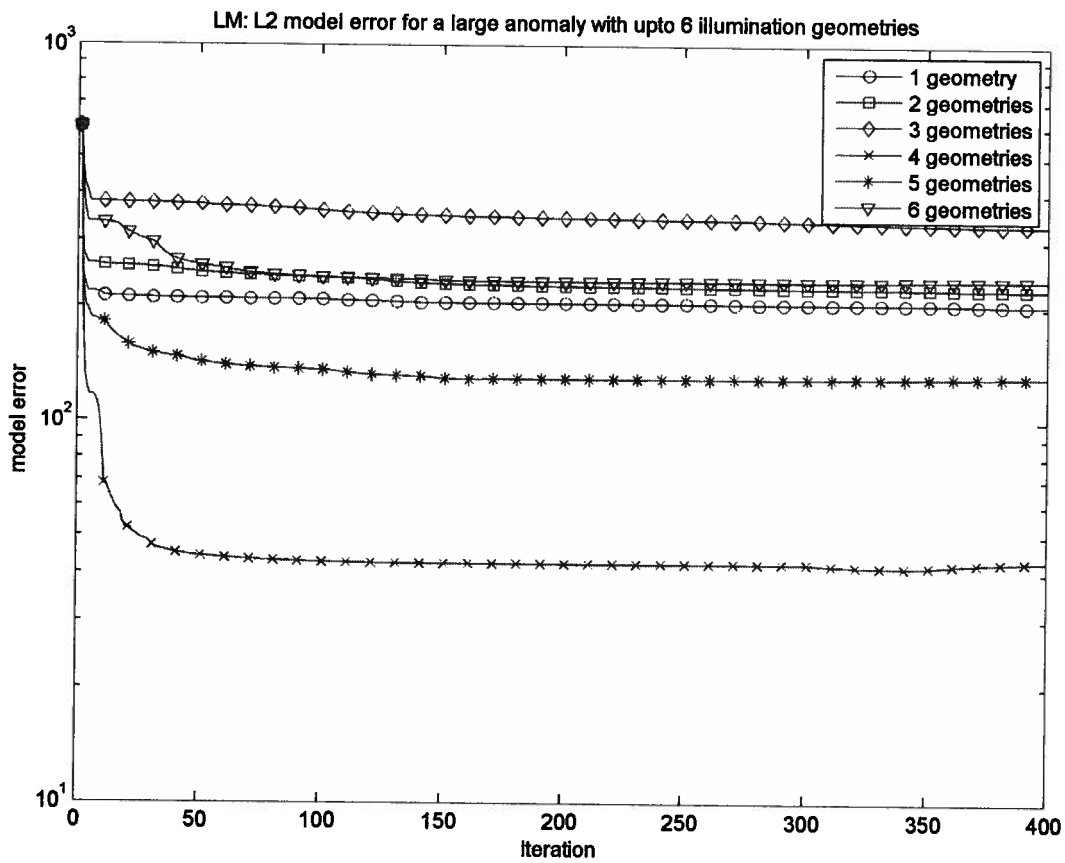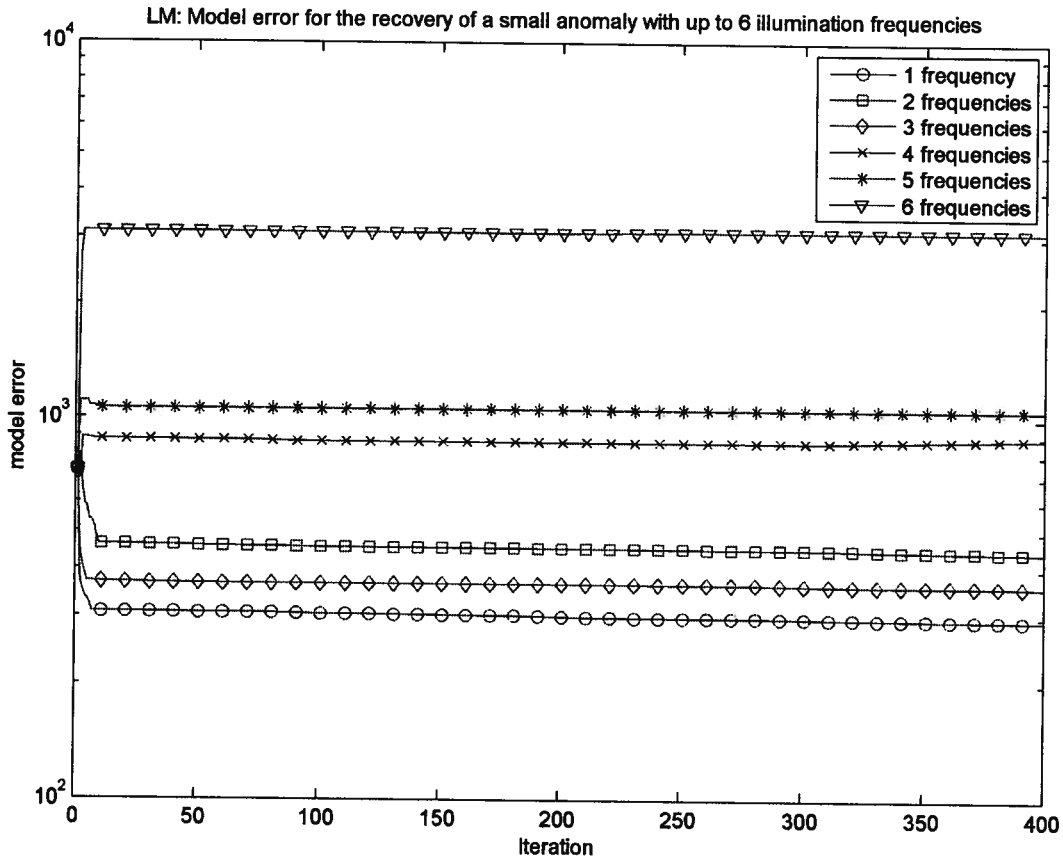*Figure 4.26: Model error curves for the recovery of the small anomaly case with the Levenberg-Marquardt method using up to 6 illumination frequencies and 1 illumination geometry.*

Figure 4.26 shows the model error curves for the recovery of the small anomaly case using the

LM method with up to 6 illumination frequencies and 1 illumination geometry. In this case, there is a

large initial rise in model error in the 6, 5, and 4 illumination frequencies cases. This does not

correspond with cases in the recovery of a large anomaly. This implies that the large initial increase is

not an artifact the choice of illumination frequencies. The initial increase is a result of the starting

model and the misfit surface as illustrated in figure 4.24. The remaining 3 model error curves in figure

4.26 have a large initial decrease in model error and converge to a final value.

LM: Relative permittivity distribution on the z=0 plane after 1 iteration for the recovery of the small anomaly case using 6 illumination frequencies and 1 illumination geometry.
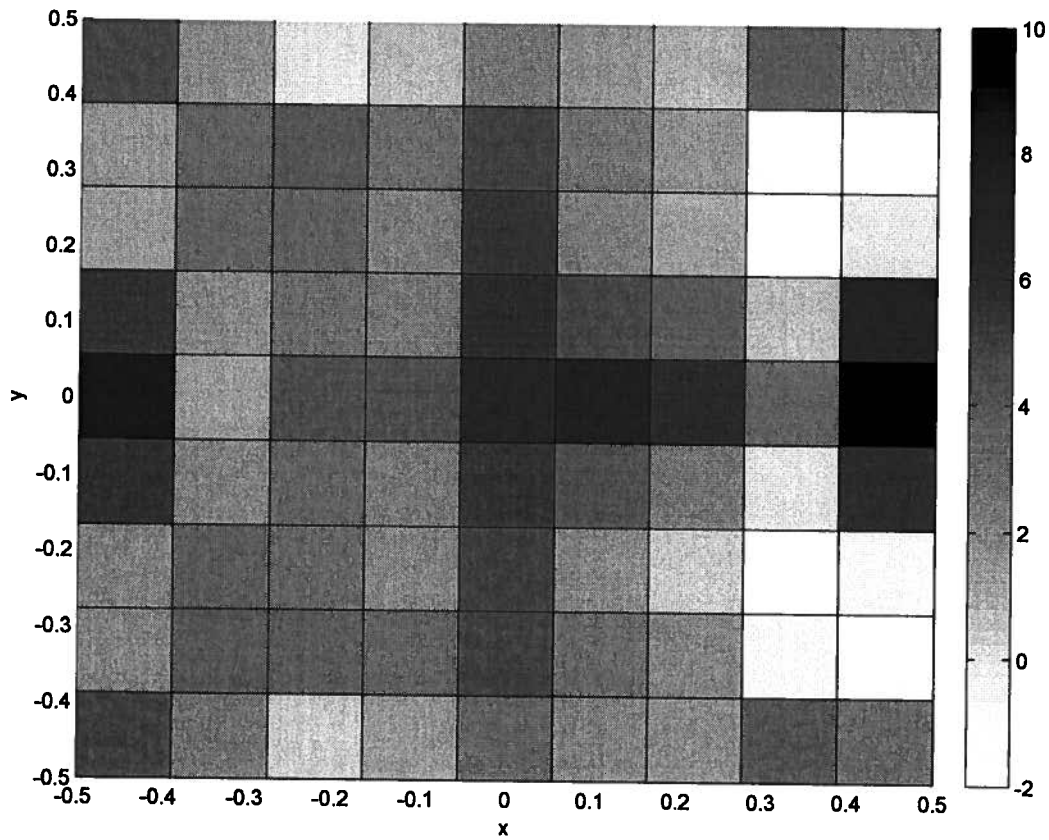
Figure 4.27: Relative permittivity distribution for the recovery of the small anomaly case using the LM method with 6 illumination frequencies and 1 illumination geometry at the 1st iteration. A cross section on the z=0 plane is shown.

Figure 4.27 shows the recovered permittivity distribution for the LM method recovering the small anomaly with 6 illumination frequencies and 1 illumination geometry at the 1st iteration. This is one of the cases where the model error increased. Negative relative permittivity values can be seen around the point y= -0.22 and x=-0.33. Comparing figures 4.27 and 4.2, there is little resemblance between the target and recovered relative permittivity distributions.

132

LM: Relative permittivity distribution on the z=0 plane after 400 iterations for the recovery of the small anomaly case using 6 illumination frequencies and 1 illumination geometry.
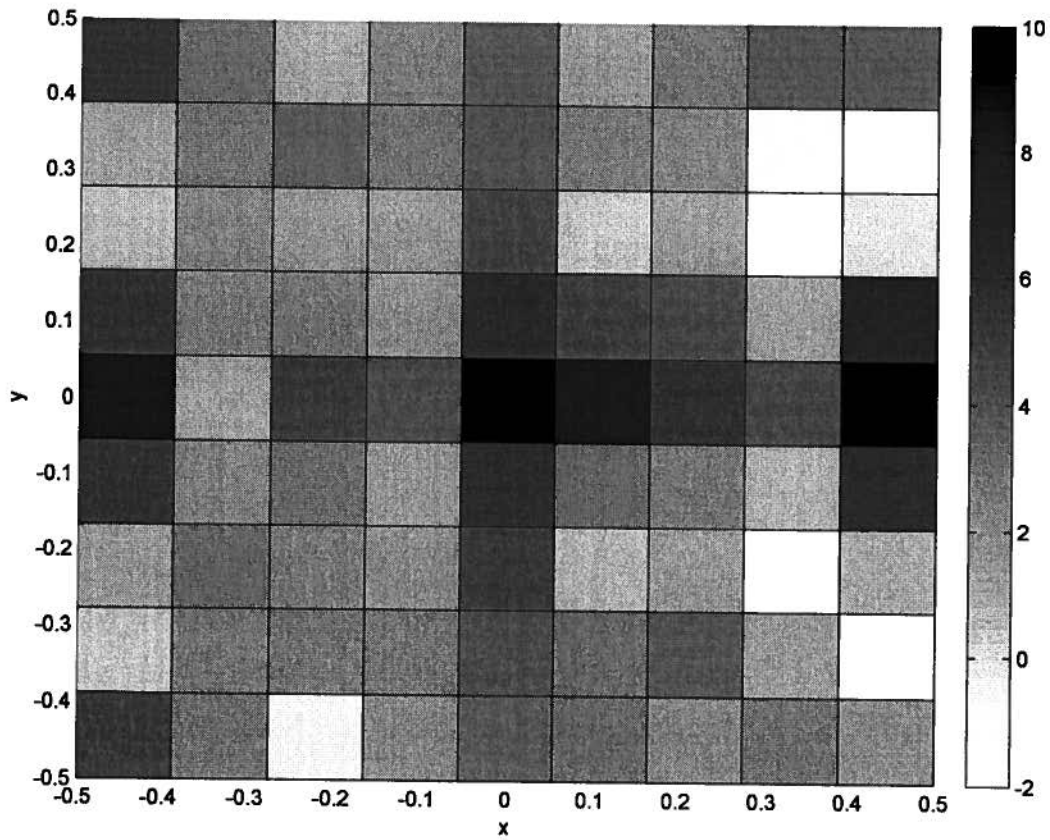
Figure 4.28: Relative permittivity distribution for the recovery of the small anomaly case using the LM method with 6 illumination frequencies and 1 illumination geometry at the 400th iteration. A cross section on the z=0 plane is shown.

Figure 4.28 shows the recovered permittivity distribution for the LM method recovering the

small anomaly with 6 illumination frequencies and 1 illumination geometry at the 400th iteration.

Comparing figure 4.27 and 4.28, the negative relative permittivity value is seen in both. The two

relative permittivity distributions looks similar implying that the LM method has reached the vicinity of

a local minimum after the first iteration.

133

LM: Relative permittivity distribution after 400 iterations for the recovery of the small anomaly case using 4 illumination geometries and 1 illumination frequency.
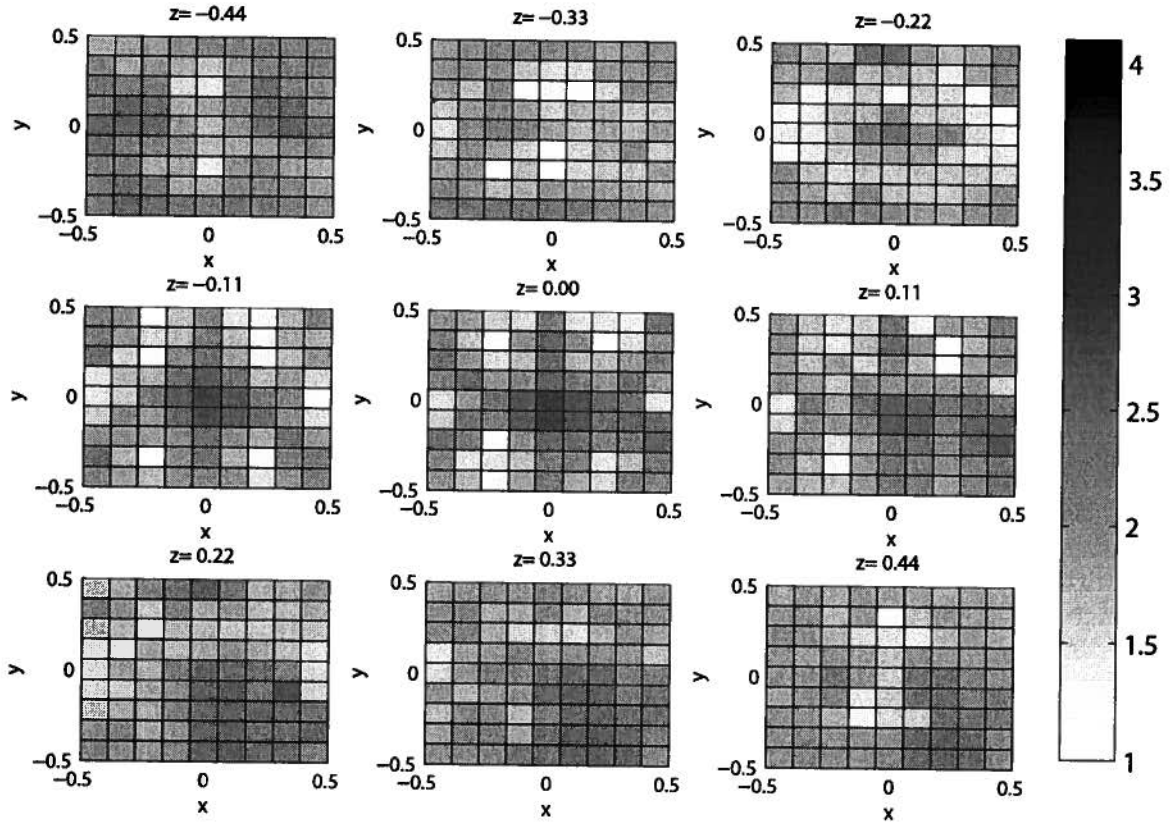
Figure 4.29: Relative permittivity distribution for the recovery of the small anomaly case using the LM method with 4 illumination geometries and 1 illumination frequency at the 400th iteration.

Figures 4.29 shows the recovered relative permittivity distribution for the small anomaly case by the LM method using 4 illumination geometries and 1 illumination frequency at the 400th iteration. This is the experiment which produced the smallest model error for the LM method. Figure 4.29 has little resemblance to figure 4.2, which is the target relative permittivity distribution. The LM method has clearly failed to recover the anomaly, because the LM method converges to a local minimum if the starting model is not sufficiently close to the target relative permittivity distribution.

When compared to the gradient method the LM method was able to find a smaller misfit value in all the examples presented. However, there are examples where the LM method produced a greater

model error value than that of the gradient method. This is due to the existence of local minima on the misfit surfaces.

For the LM method adding more data did not consistently produce a smaller model error. The final model error values are dependent on the locations of the local minima and their locations cannot be predicted apriori. Adding more data improves the local properties around the global minimum. However the effect of more data on the locations and value of the local minima is unknown. Therefore, for an algorithm which finds local minima, it is not possible to predict if adding more data will improve the model error.

4.6 L2 cooled roughness

In this section the L_2 cooled roughness(CRL2) method is used to recover the relative

permittivity distributions shown in figure 4.1 and 4.2.

The CRL2 method does not guarantee that the misfit will decrease. This condition is built into

the LM method and the gradient method. There will be a number of examples where the misfit will

increase.

The CRL2 method uses the roughness regularization which penalizes the spatial derivative of

the relative permittivity distribution. The roughness regularization is chosen based on

Jackson's[Jackson 1979] rationale for choosing a regularization. Jacksons states that if the solution is

non-unique a regularization based on physical properties should be used. In the non-linear problem

the local minima causes algorithms that only use local information to not find the global minimum,

yielding solutions that bear little resemblance to actual solution . There can be an infinite number of

local minima where the search algorithms may be mistaken as a correct solution. The roughness

regularization filters out many of these local minima, hence improving the likelihood of finding a

correct solution. By choosing a regularization that is based on some physical characteristic that is

believed to be in the actual distribution, many of the unrealistic distributions are filtered out. In this

way the regularization helps find the correct solution. The roughness regularization is chosen based on

the concept that a feature is a region which is nearly homogeneous. Relative permittivity distributions

with many sharp transitions are penalized. The problem with this regularization is that the distribution

that the algorithm is trying to recover also contains sharp transition as seen in Figure 4.1 and 4.2. That

is what distinguishes one region from another. To overcome this difficulty, the concept of cooling is

introduced, so that these transitions are allowed to slowly enter the relative permittivity distribution

because the regularization is relaxed after each iteration.

Therefore, roughness regularization controls the region in which a feasible solution may be found. The cooling allows this region to expand after each iteration. Like the LM method, the cooled roughness regularization finds a minimum between the roughness regularization and that of the paraboloid approximation based on the local information. The minimum for the LM regularization is a point, centered at the expansion point. The minimum of the roughness regularization is a line, where the line defines the set of homogeneous permittivity distribution. As the weighting of the roughness regularization decreases after each iteration, the solution can get further away from this line.

For the cooled roughness the misfit curves will the shown first because there are instabilities in the recoveries and a criteria based on the misfit for selecting what is considered to be the recovered relative permittivity distribution is needed. The recoveries for the large anomaly case stopped at 600 iterations. The recoveries for the small anomaly case stopped at 500 iterations.

Figure 4.30: Misfit curves for the L_2 cooled roughness method recovering the large anomaly case with up to 6 illumination geometries and 1 illumination frequency.

Figure 4.30 shows the misfit curves for the recovery of the large anomaly case using CRL2 with

up to 6 illumination geometries and 1 illumination frequency. All the curves in figure 4.30 have a

region of instability. For the CRL2 method the weighting of the regularization is decreased with each

iteration. When the weighting becomes too small, the algorithm becomes unstable. The point where

the algorithm becomes unstable is dependent on the features of the misfit surface. The relative

permittivity distribution at the point where the misfit curves initially increase will be considered the

recovered distribution. This point is different for each experiment.

From the misfit curves, 4 features are apparent. The first is the large initial decrease in misfit

because the algorithm starts with a large weighting of the regularization; this corresponds to the

138

algorithm finding the best homogeneous match. The second portion of the curve is a large flat region. This second region represents the slow introduction of very small scale features into the surface. The third region is where the misfit noticeably decreases. This represents the slow introduction of visible features into the the recovered relative permittivity distribution. As the weighting of the regularization decreases the surface is able to deviate from that of a homogeneous surface, representing a gradual introduction of features into the relative permittivity distribution. By gradually allowing features to enter the surface, the algorithm is able to avoid many of the local minima seen in the gradient algorithm and the LM algorithm. From the previous two methods it can be seen that there are many local minima on the surface, but the ones found by the LM method and the gradient method tends to be feature rich. The last portion CRL2 method's misfit curves displays instabilities. This portion of the curve is relatively unstable, because paraboloid approximation is inaccurate and the weighting of the regularization is too small to mitigate these inaccuracies.
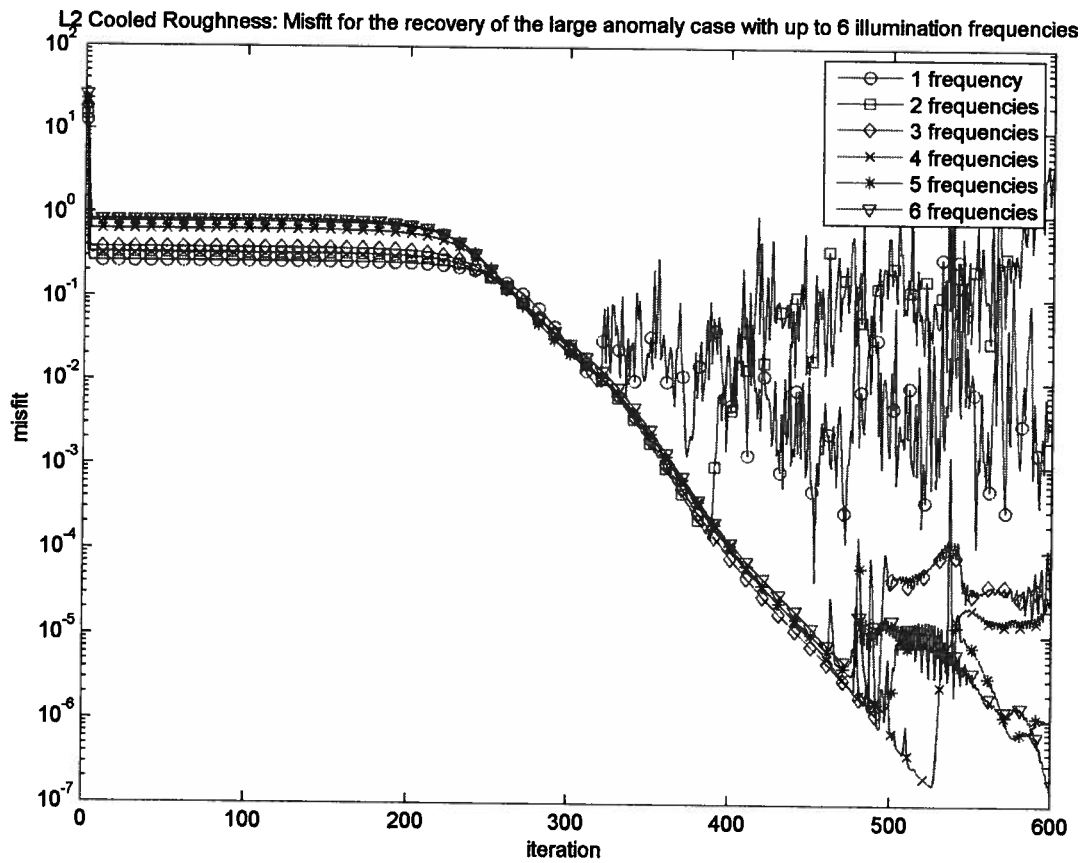
Figure 4.31: Misfit curves for the L_2 cooled roughness method recovering the large anomaly case with up to 6 illumination frequencies and 1 illumination geometry.

Figure 4.31 shows the misfit curves for the recovery of the large anomaly case using CRL2 with

up to 6 illumination frequencies and 1 illumination geometry. The misfit curves in figure 4.31 exhibit

the same behavior as the those seen in figure 4.30.

L2 Cooled Roughness: Misfit for the recovery of the small anomaly case with up to 6 illumination geometries
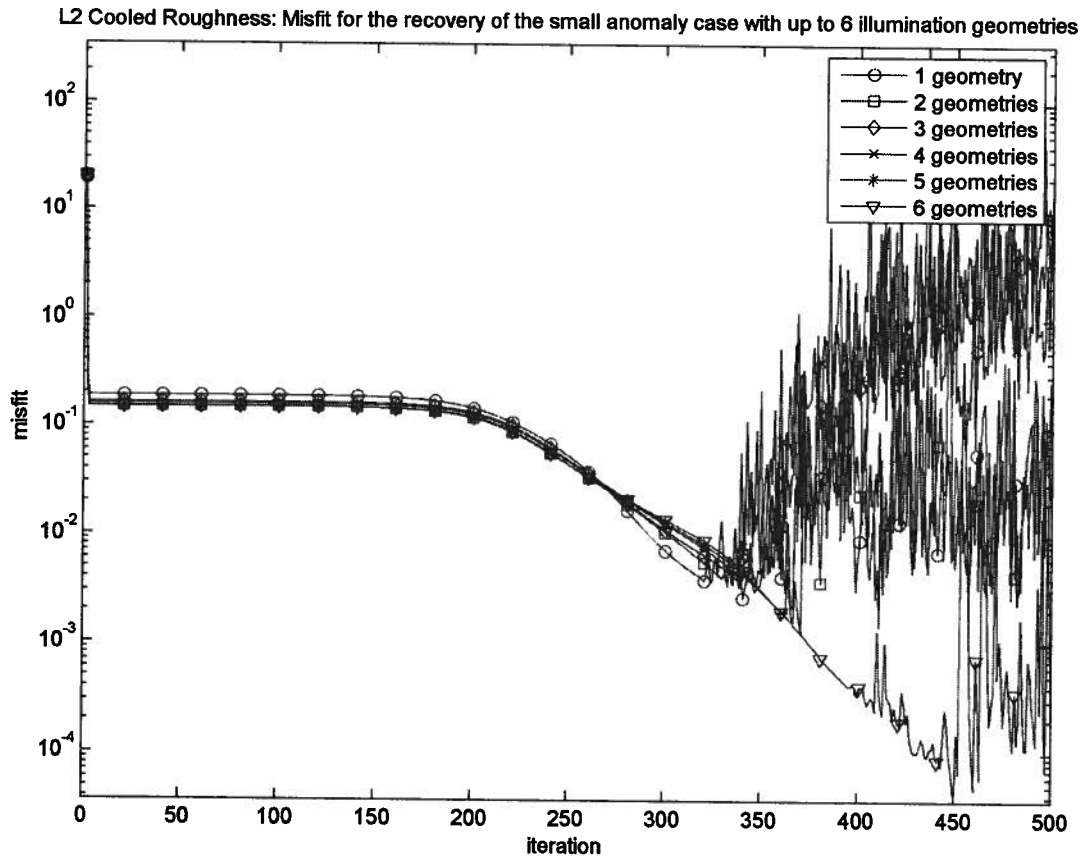
Figure 4.32: Misfit curves for the L₂ cooled roughness method recovering the small anomaly case with up to 6 illumination geometries and 1 illumination frequency.

Figure 4.32 shows the misfit curves for the recovery of the large anomaly case using CRL2 with up to 6 illumination geometries and 1 illumination frequency. The misfit curves in figure 4.32 exhibit the same behavior as the those seen in figure 4.30.

L2 Cooled Roughness: Misfit for the recovery of the small anomaly case with up to 6 illumination frequencies
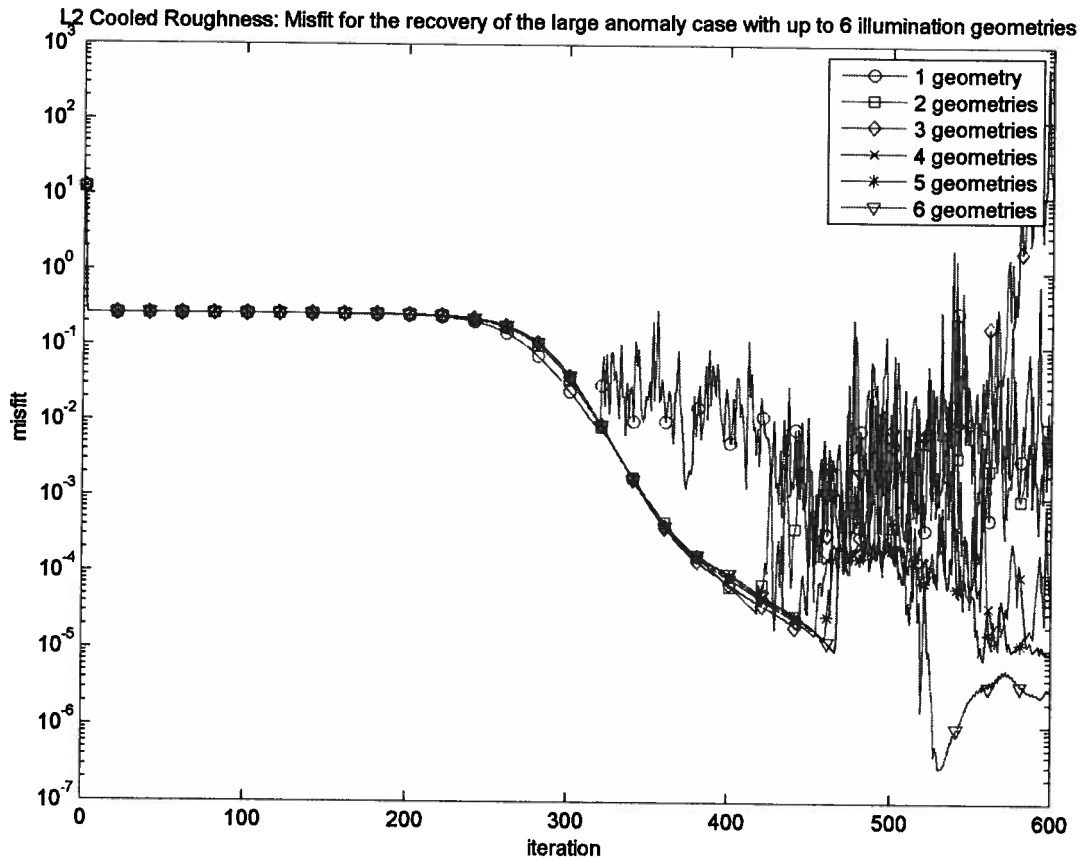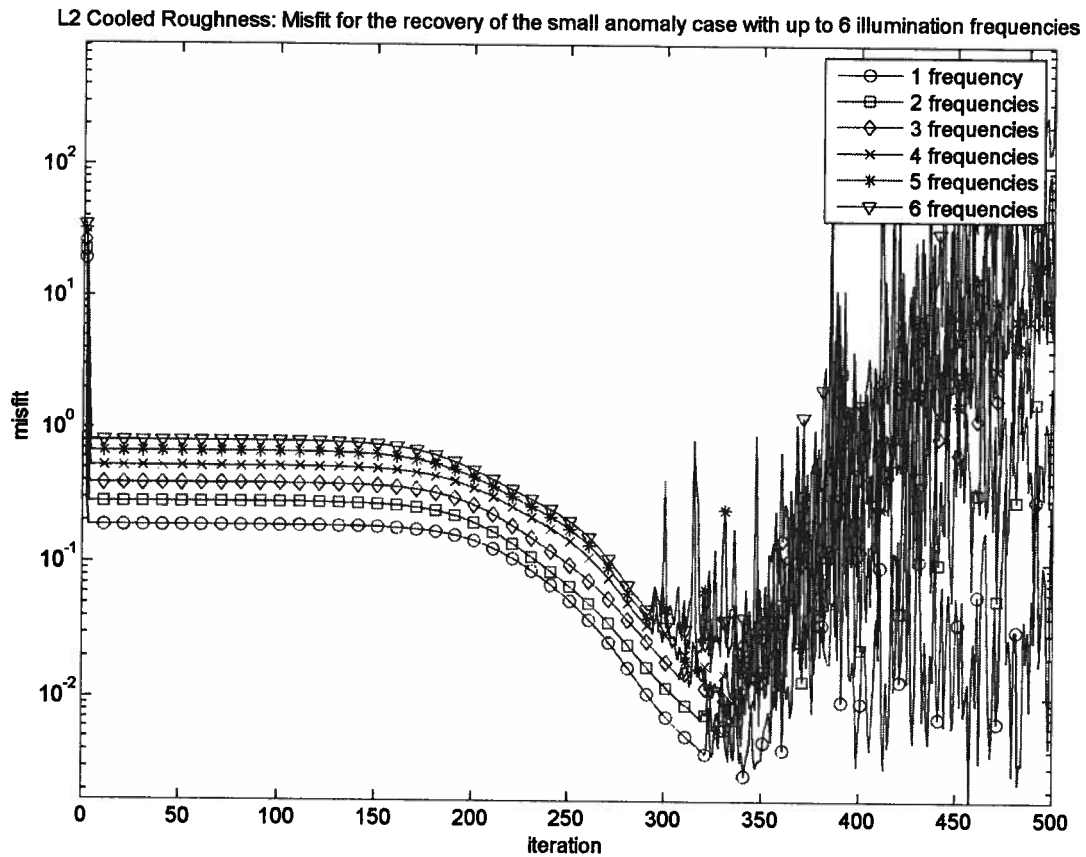
Figure 4.33: Misfit curves for the L_2 cooled roughness method recovering the small anomaly case with up to 6 illumination frequencies and 1 illumination geometry.

Figure 4.33 shows the misfit curves for the recovery of the small anomaly case using CRL2 with up to 6 illumination frequencies and 1 illumination geometry. The misfit curves in figure 4.33 exhibit the same behavior as the those seen in figure 4.30.

Table 4.6 summarizes the results for the CRL2 method. The misfit value and model error is taken at the point where the misfit curve initially increases. This criterion was chosen because the onset of the instability is indicated by an increase in misfit value. There is no increase misfit value in the other portions of the misfit curves. A homogeneous starting model with a relative permittivity value of 3.0 was used to generate the results.

Table 4.6: Results for the CRL2 method					
Illumination geometries	Observation frequencies	Misfit/ N_f small anomaly	Model Error small anomaly	Misfit/ N_f large anomaly	Model Error large anomaly
I1	F1	3.83e-03	81.84	8.73e-03	57.58
I1 to I2	F1	4.01e-03	76.42	3.04e-05	26.35
I1 to I3	F1	3.93e-03	72.44	2.18e-05	22.02
I1 to I4	F1	4.10e-03	69.82	1.57e-05	17.57
I1 to I5	F1	1.80e-03	56.50	1.36e-05	15.15
I1 to I6	F1	4.12e-04	43.46	9.66e-06	13.16
I1	F1 to F2	6.74e-03	77.56	1.62e-04	29.89
I1	F1 to F3	1.42e-02	75.37	1.64e-06	19.18
I1	F1 to F4	3.45e-02	75.88	2.21e-06	17.22
I1	F1 to F5	4.32e-02	74.53	6.14e-06	16.91
I1	F1 to F6	4.56e-02	72.33	4.56e-06	11.45

Comparing the misfit values between the CRL2 method and the LM method, it can be seen that occasionally the CRL2 method has a smaller misfit value and occasionally the LM method has a smaller misfit value. For the case of 4 illumination geometries and 1 illumination frequency for the recovery of the large anomaly, the LM method produced a misfit value of 1.50e-08 and CRL2 produced a misfit value 1.57e-05. The LM method has clearly produced a smaller misfit value, however the LM method has higher model error. The LM method's model error is 43.12 and the CRL2 method's model error is 17.57. This illustrates one of the challenges with tomographic recovery. On the misfit surface, there are many local minima with small misfit values with large model errors. This reinforces the concept that it is not sufficient to only find a small misfit value to perform a recovery.

The CRL2 method consistently found a smaller model error when compared to the LM method. The smaller model error implies that the CRL2 method has come closer to the global minimum than the LM method. The cooled roughness algorithm is able to find smaller model error values than the previous methods because it is able to avoid many of the local minima between the starting point and the global minimum because of the filtering effect of the roughness regularization. The cooled

roughness algorithm allows structure to enter the recovered permittivity distribution slowly and it is this behavior that allows the algorithm to avoid many of the local minima. It is not able to avoid all of them because the final misfit value is not zero. It is unclear whether the CRL2 method reached a minimum, global or local, because the algorithm becomes unstable. The instability occurs because the weighting on the regularization becomes too small to mitigate the bad conditioning of the Jacobian matrix.

CRL2: Recovered relative permittivity distribution for the large anomaly case at iteration 474 using 6 illumination frequencies and 1 illumination geometry.
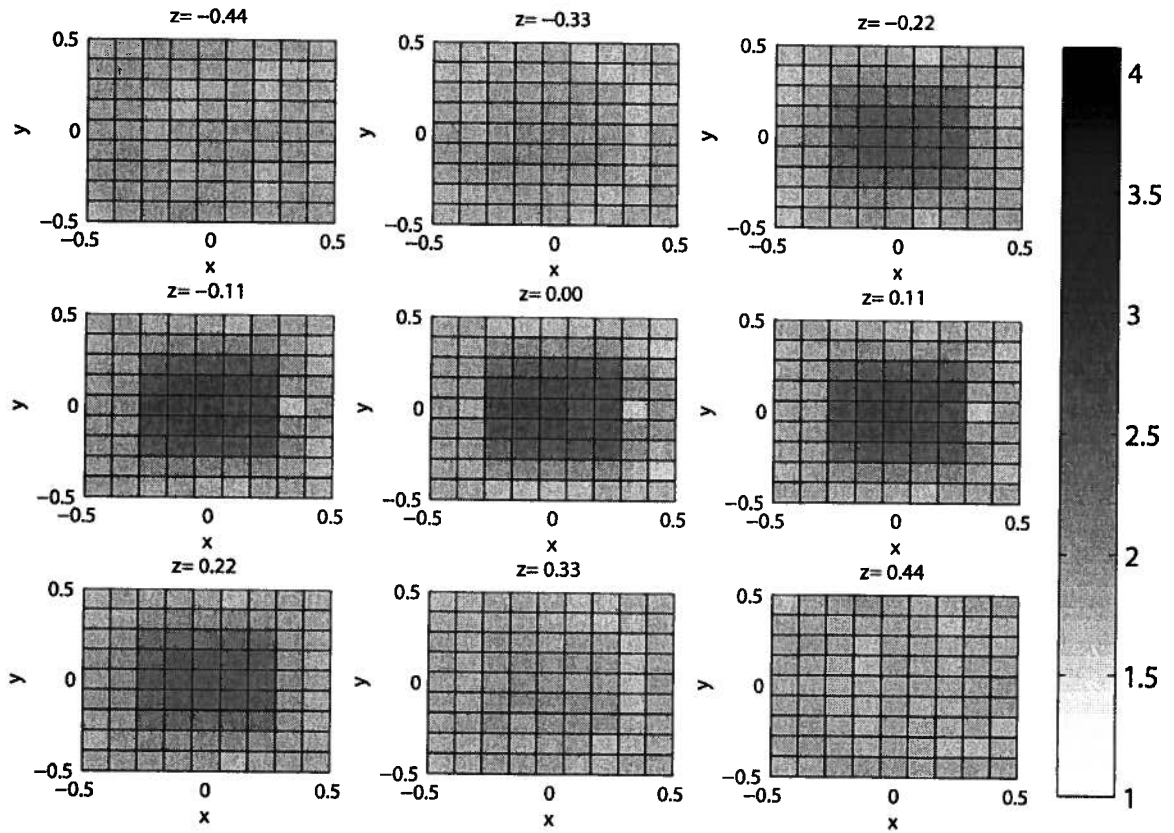
Figure 4.34: A recovered relative permittivity distribution for the large anomaly case, using L_2 cooled roughness method with 1 illumination geometry and 6 illumination frequencies at iteration 474.

Figure 4.34 shows the recovered relative permittivity distribution for the large anomaly case using the CRL2 method with 6 illumination frequencies and 1 illumination geometry. Comparing

144

figure 4.34 to figure 4.1, it can be seen that both contain the large anomaly in the center. On the z = 0.22 plane and z = -0.22 plane the relative permittivity values for the anomaly on the recovered relative permittivity distribution is smaller than those seen in figure 4.1. The anomaly appears blurry in the recovered relative permittivity distribution because of the lack of sharp transition from one region to the next.

CRL2: Recovered relative permittivity distribution for the small anomaly case at iteration 396 using 1 illumination frequency and 6 illumination geometries.

Figure 4.35:A recovered relative permittivity distribution for the small anomaly case using L_2 cooled roughness with 6 illumination geometries and 1 illumination frequency at iteration 396.

Figure 4.35 shows the recovered relative permittivity distribution for the small anomaly case using the CRL2 method with 6 illumination geometries and 1 illumination frequency. Comparing figure 4.35 to figure 4.2, it can be seen that both contain the small anomaly in the same location.

However it does not appear homogeneous in the recovered relative permittivity distribution. The anomaly in the recovered relative permittivity distribution appears blurry.

From the recoveries shown in 4.34 and 4.35 a couple of features stand out. First, the images are blurry. The blurriness contribution from the regularization is due to the penalization of sharp transitions in the recovered permittivity distribution. The same blurriness is also apparent in the research of Greffrin et al. (2007) and Yu et al. (2008). Both authors uses the L2 norm, but different regularizations, suggesting that the blurriness is due to the L2 norm on the data misfit. The second notable feature is the appearance of artifacts in the recovered distribution. These artifacts are also found in similar research. The artifacts are caused by the algorithm attempting to fit the data, but due to the blurriness caused by the L2 norm, the algorithm must introduce artifacts to fit the data. What follows is an analysis of the misfit and model error as a function of iteration number.
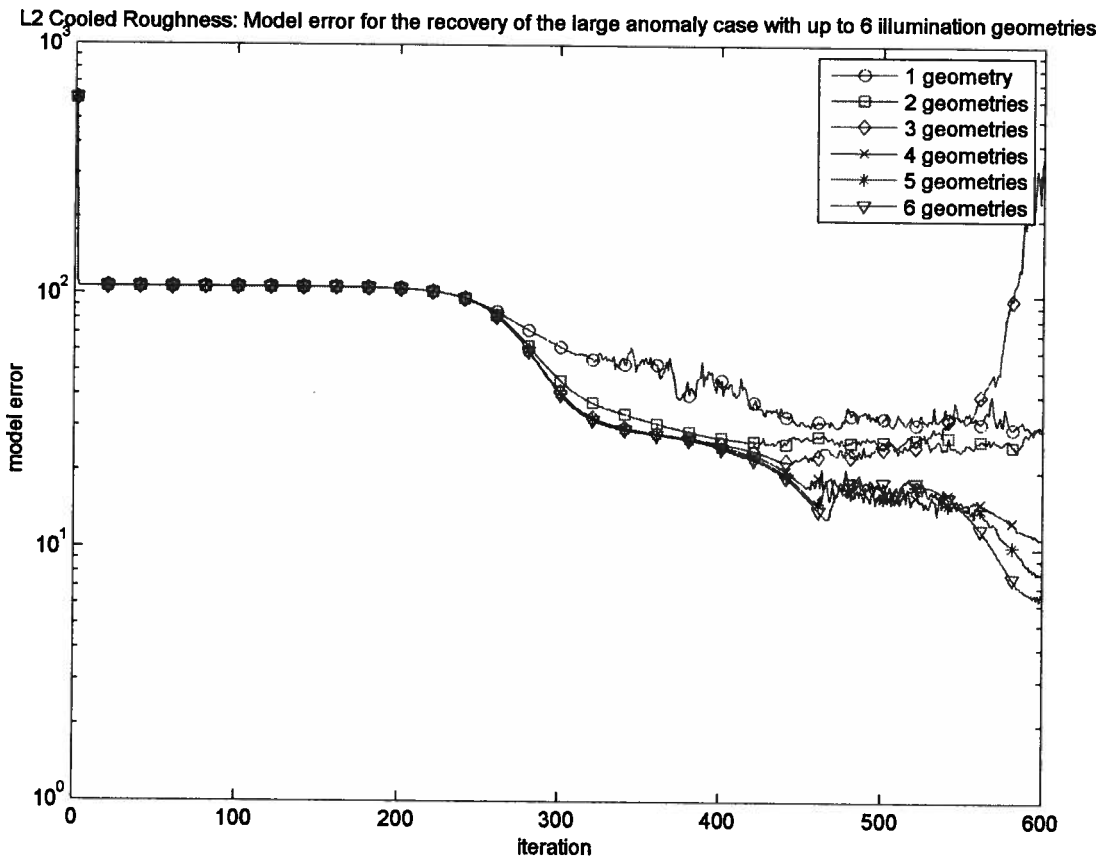
Figure 4.36: Model error curves for recovery of the large anomaly case using the L_2 cooled roughness method with up to 6 illumination geometries and 1 illumination frequency.

Figure 4.36 shows the model error curves for the recovery of the large anomaly case with the

CRL2 method using up to 6 illumination geometries and 1 illumination frequency. The instabilities

seen in the misfit curves in figure 4.30 are not as dramatic in figure 4.36. In several cases the algorithm

was able to recover from the unstable behavior and find a smaller model error.
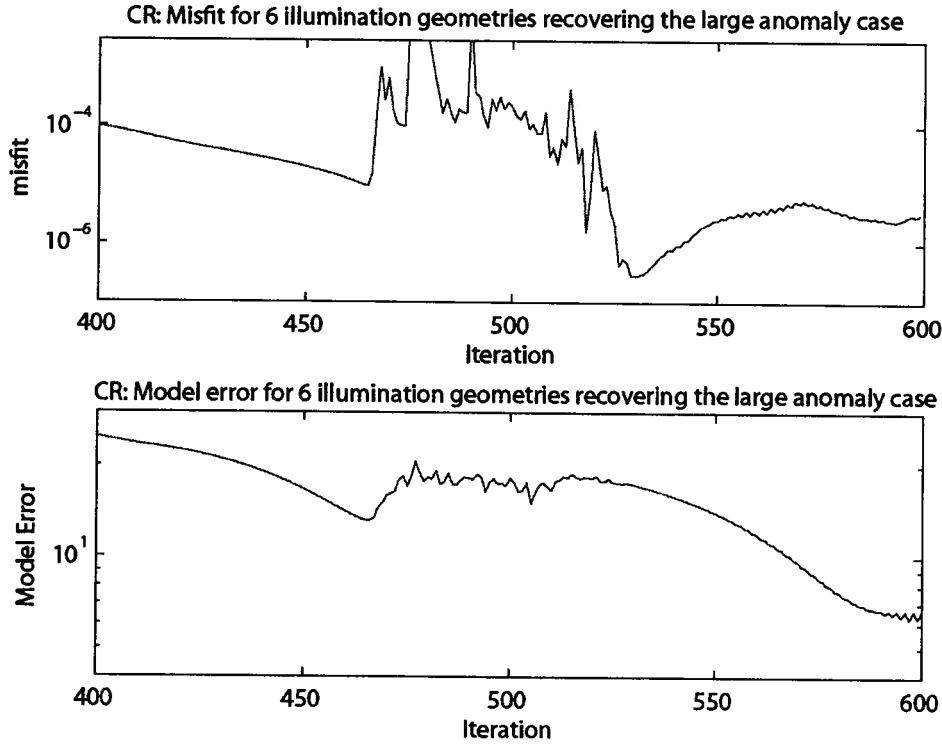
Figure 4.37: Misfit and model error for the recovery of a large anomaly using 6 illumination geometries and the L_2 cooled roughness method.

Figure 4.37 is a close up of the model error and misfit curves for the case of 6 illumination geometries and 1 illumination frequency. This is one of the cases where the CRL2 method was able to recover from the instability. At approximately iteration 460, the increase in both model error and misfit can be seen. Between iterations 520 and 540, the misfit curve increases, while the model error decreases, once again reinforcing the concept that a decrease in misfit does not always imply a decrease in model error. At the 600th iteration, the model error is lower than at the the point where the misfit initially increases. On possible explanation for the behavior of misfit curve and model error curve is the existence of many local minima near the global minimum. The misfit value at the 600th iteration is 3.27e-06 and the model error value is 6.72.

L2 cooled roughness: Recovered relative permittivity distribution for the large anomaly case at the 600th iteration using 6 illumination geometries.
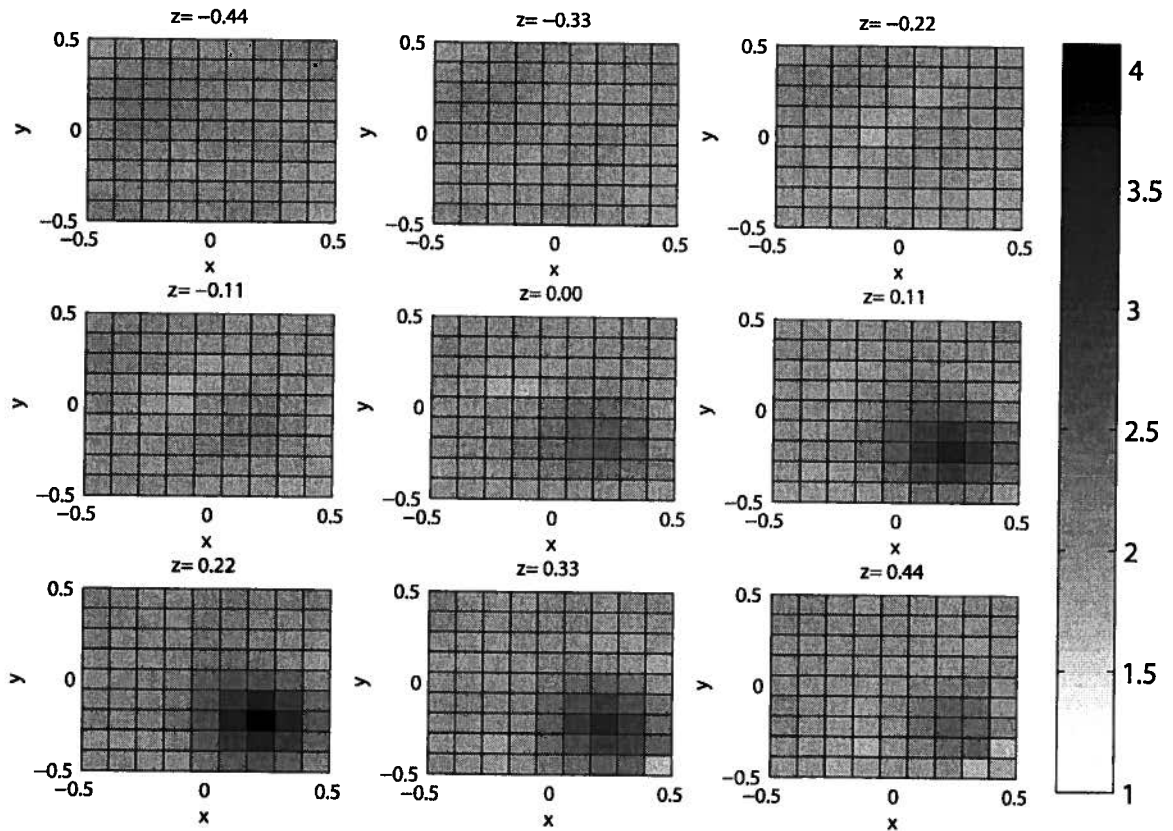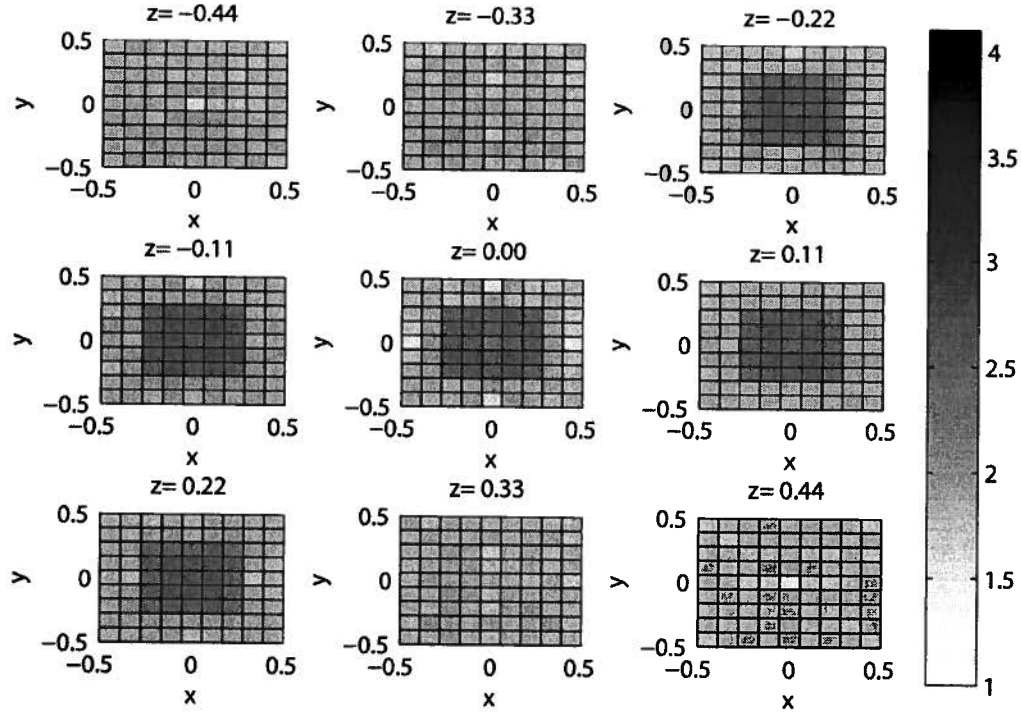
Figure 4.38: Recovered relative permittivity distribution for the large anomaly cause using the L_2 cooled roughness method with 6 illumination geometries and 1 illumination frequency.

Figure 4.38 shows the recovered relative permittivity distribution for the large anomaly case using 6 illumination geometries and 1 illumination frequency at the 600th iteration. Figure 4.38 resembles figure 4.1 indicating that the CRL2 is able to successfully recover the structure of the large anomaly case.

149
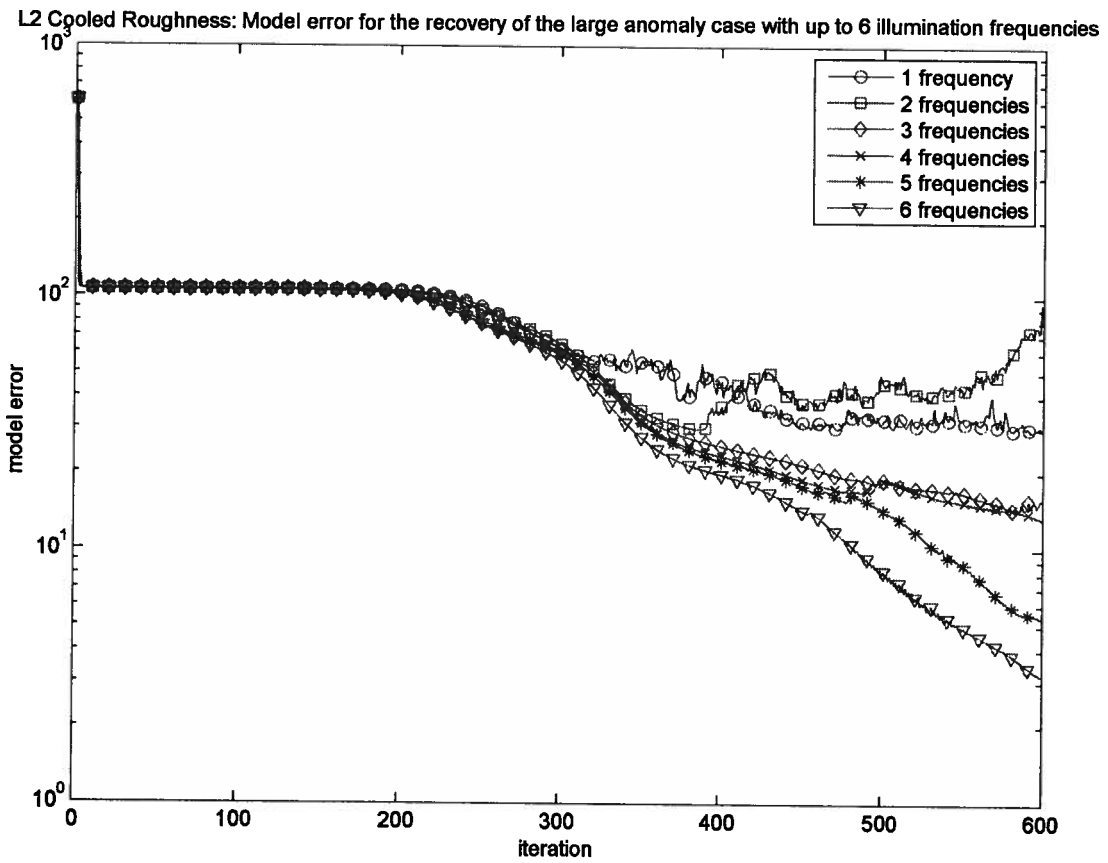
Figure 4.39: Model error results for the L_2 cooled roughness method recovering the relative permittivity for the large anomaly case, using up to 6 illumination frequencies and 1 illumination geometry.

Figure 4.39 shows the model error curves for the recovery of the large anomaly case using the

CRL2 method with 6 illumination frequencies and 1 illumination geometry. The model error curves in

figure 4.39 exhibit the same behavior as the model error curves in figure 4.36.

150
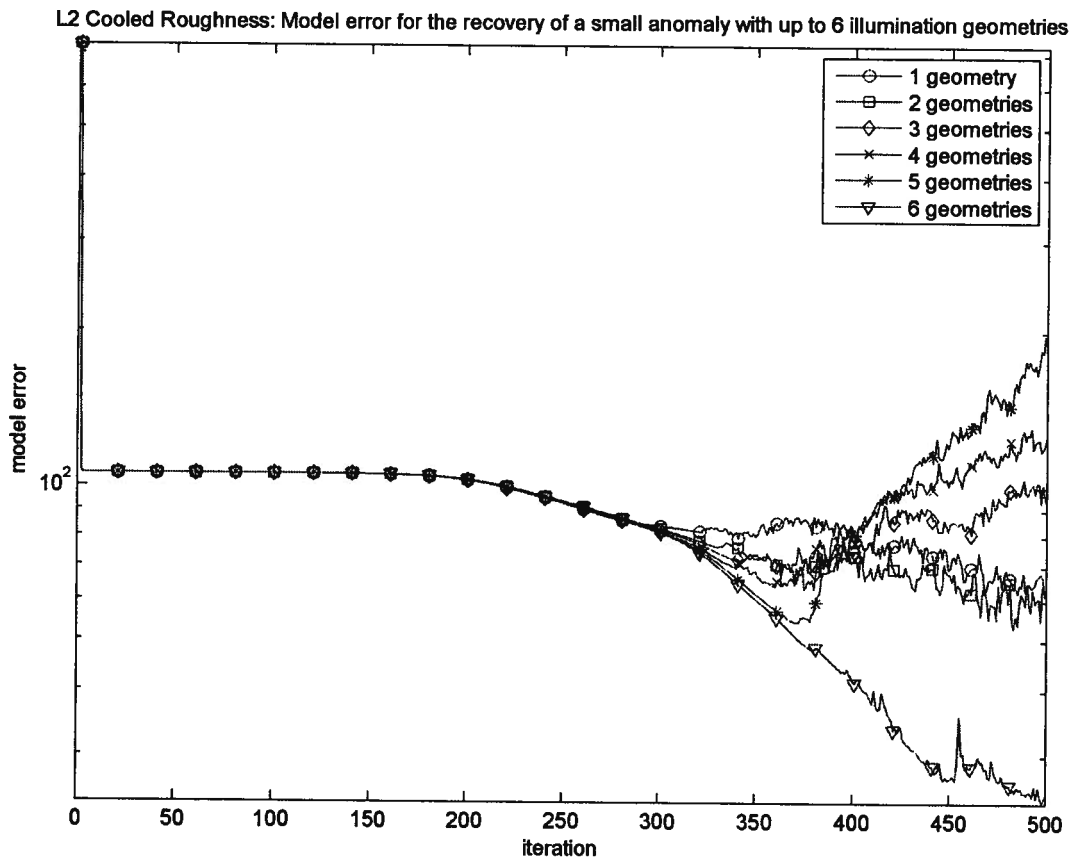
Figure 4.40: Model error results for the L_2 cooled roughness method recovering the relative permittivity distribution for the small anomaly case, using up to 6 illumination geometries and 1 illumination frequency.

Figure 4.40 shows the model error curves for the recovery of the small anomaly case using the

CRL2 method with 6 illumination geometries and 1 illumination frequency. The model error curves in

figure 4.39 exhibit the same behavior as the model error curves in figure 4.36.
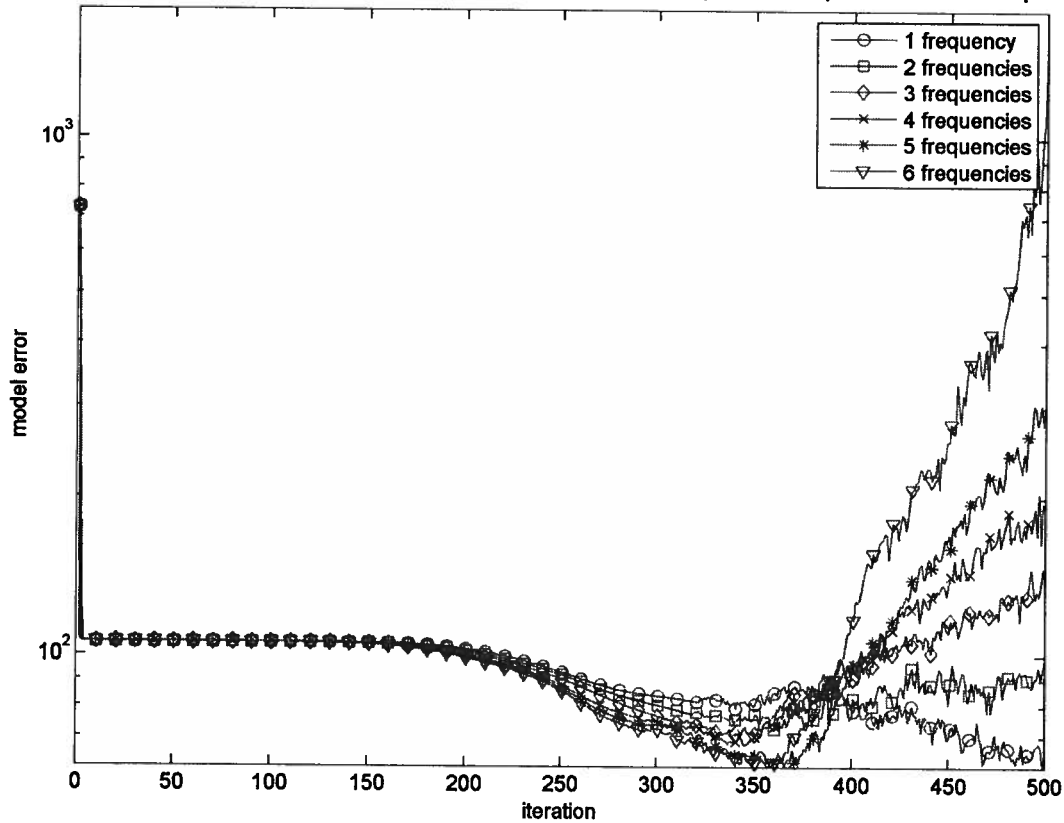
Figure 4.41: Model error results for the L_2 cooled roughness method recovering the relative permittivity distribution for the small anomaly case, using up to 6 illumination frequencies and 1 illumination geometry.

Figure 4.41 shows the model error curves for the recovery of the small anomaly case using the CRL2 method with 6 illumination frequencies and 1 illumination geometry. The model error curves in figure 4.41 exhibit the same behavior as the model error curves in figure 4.36.

The CRL2 consistently found smaller model errors values than the LM method. However, in some cases the LM method produced smaller misfit values. This behavior illustrates the need to filter out non-physical relative permittivity distributions because they can produce small misfit values. Algorithms, like the LM method, that only use local information will be trapped inside these local minima. By using the roughness regularization, the CRL2 method was able to avoid many of these local minima.

4.7 L1 cooled roughness

In this section the L_1 cooled toughness(CRL1) method, as presented in chapter 3, is used to recover the relative permittivity distributions shown in figure 4.1 and 4.2.

The CRL1 uses the iterative re-weighted least square(IRLS) approximation (Farquarson, 1998) to the L_1 norm, as presented in chapter 3. The motivation to use the L_1 norm stems from 2 factors. For the regularization, the L_2 norm penalizes sharp transitions much greater than that of a small transition. However it is these sharp transitions that define the boundary of a feature. By moving to L_1, these transition are penalized much less severely. The second motivation is the fact that blurriness is inherent in the L_2 recovery. Using the L_1 norm will reduce this blurriness.

Table 4.6 summarizes the results for the CRL1 method. A homogeneous starting model with a relative permittivity value of 3.0 was used to generate the results.

Table 4.7: Results for the L1 cooled roughness algorithm					
Illumination geometries	Observation frequencies	Misfit/ N_f small anomaly	Model Error small anomaly	Misfit/ N_f large anomaly	Model Error large anomaly
I1	F1	1.64e-11	0.746	1.93e-11	0.647
I1 to I2	F1	3.56e-11	0.443	3.68e-11	0.463
I1 to I3	F1	1.20e-10	0.386	7.51e-11	0.364
I1 to I4	F1	9.21e-11	0.316	7.96e-11	0.293
I1 to I5	F1	1.02e-10	0.248	1.04e-10	0.270
I1 to I6	F1	1.12e-10	0.218	1.08e-10	0.217
I1	F1 to F2	4.48e-11	0.453	3.89e-11	0.496
I1	F1 to F3	6.95e-11	0.408	8.32e-11	0.401
I1	F1 to F4	9.37e-11	0.336	8.89e-11	0.353
I1	F1 to F5	1.09e-10	0.275	1.01e-10	0.301
I1	F1 to F6	1.01e-10	0.241	1.04e-10	0.227

From table 4.7, it can be seen that the L_1 cooled roughness (CRL1) algorithm has achieved a smaller model error than gradient, LM, and CRL2 methods. The model error and misfit presented in

table 4.7 is the L_2 error to provide a fair comparison with the previous algorithms. The results for table 4.7 were taken at the 400th iteration. For the CRL1 method, the experiments with more data had the smallest errors. The smallest model error was 0.217 for the recovery of a large anomaly and 0.218 for the recovery of a small anomaly. Including more illumination geometries provided marginally better performance than including more illumination frequencies.

Table 4.8: Comparison of the model error for the recovery of the large anomaly				
Illumination geometries	Observation frequencies	CRL1: Model Error large anomaly	CRL2: Model Error large anomaly	LM:Model Error large anomaly
I1	F1	0.647	57.58	204.05
I1 to I2	F1	0.463	26.35	224.48
I1 to I3	F1	0.364	22.02	332.18
I1 to I4	F1	0.293	17.57	43.12
I1 to I5	F1	0.270	15.15	131.99
I1 to I6	F1	0.217	13.16	237.85
I1	F1 to F2	0.496	29.89	301.5
I1	F1 to F3	0.401	19.18	1039.0
I1	F1 to F4	0.353	17.22	368.5
I1	F1 to F5	0.301	16.91	512.73
I1	F1 to F6	0.227	11.45	1744.7

It is apparent from table 4.8 that the CRL1 consistently produced the smallest model error values and CRL2 method consistently produced the second smallest model error. The fact that the CRL2 and CRL1 methods were able to find relative permittivity distributions with smaller model errors than that of the LM method, confirms that the cooled roughness regularization led the algorithm to a relative permittivity distribution near the global minimum.

Figure 4.42: Misfit curves for the recovery of the large anomaly case using the L_1 cooled roughness method with up to 6 illumination geometries and 1 illumination frequency.

Figure 4.42 shows the misfit curves for the recovery of the large with the CRL1 method using

up to 6 illumination geometries and 1 illumination frequency. The curves exhibit a large initial drop in

misfit, which corresponds to finding the best initial homogeneous match. There is a plateau region

where small scale features enter the relative permittivity distribution. Finally, there is a decrease in

misfit which indicate the region where visible features enter the relative permittivity distribution.

Figure 4.43: Misfit curves for the recovery of the large anomaly case using the L_1 cooled roughness method with up to 6 illumination frequencies and 1 illumination geometry.

Figure 4.43 shows the misfit curves for the recovery of the large anomaly case with the CRL1

method using up to 6 illumination frequencies and 1 illumination geometry. The curves in figure 4.43

exhibit the same behavior as the curves seen in figure 4.42.

L1 cooled roughness: Misfit for the recovery of the small anomaly case with up to 6 illumination geometries.

Figure 4.44: Misfit curves for the recovery of the small anomaly case using the L_1 cooled roughness method with up to 6 illumination geometries and 1 illumination frequency.

Figure 4.44 shows the misfit curves for the recovery of the small anomaly case with the CRL1

method using up to 6 illumination geometries and 1 illumination frequency. The case of 1 illumination

geometry and 1 illumination frequency, exhibits the unstable behavior seen the CRL2 method.

However, CRL1 is able to recover from the instability for this case. Between iterations 200 and 240 a

second plateau region can be seen. This implies that the CRL1 method is not immune from the local

minima behavior seen in other algorithms. However in this case, the CRL1 method was not trapped in

the local minimum, because the regularization which imposes a global condition on the recovery

matches that of the target relative permittivity distribution. Therefore the regularization leads the

CRL1 method out of the local minimum.

157

L1 cooled roughness: Misfit for the recovery of the small anomaly case using up to 6 illumination frequencies.
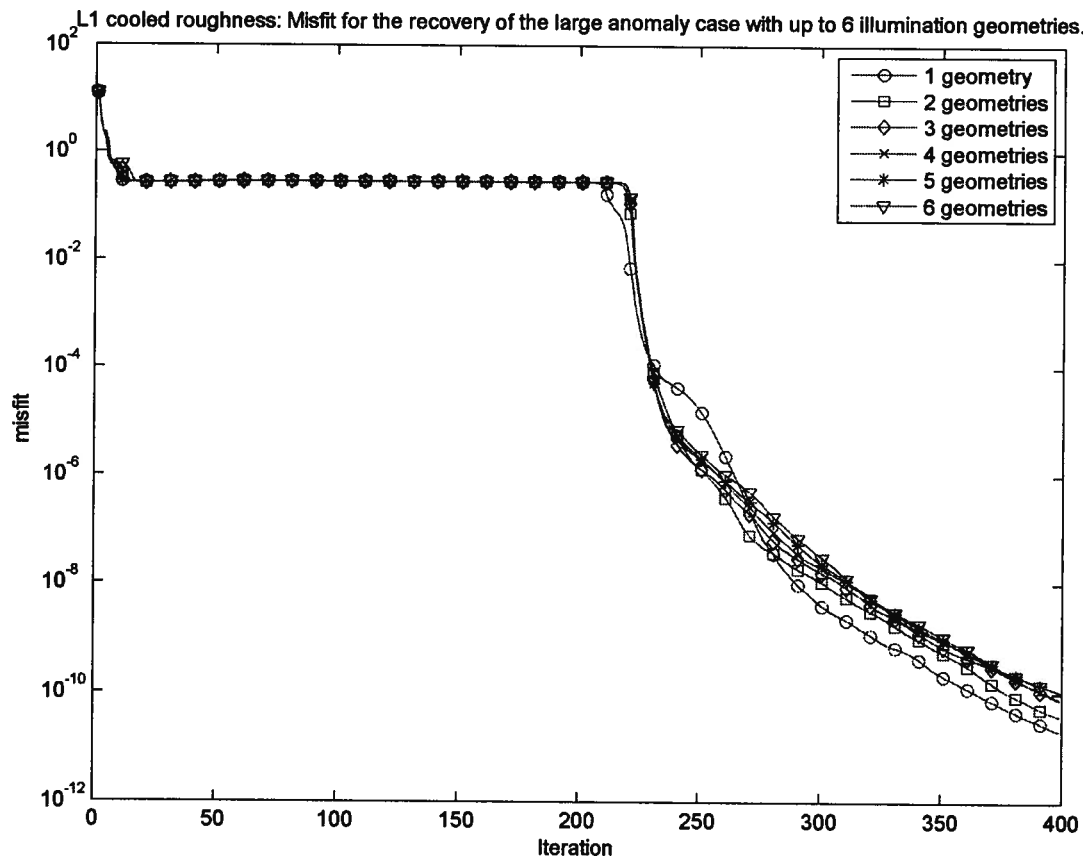
Figure 4.45: Misfit curves for the recovery of the small anomaly case using the L_1 cooled roughness method with up to 6 illumination frequencies and 1 illumination geometry.

Figure 4.45 shows the misfit curves for the recovery of the small anomaly case with the CRL1

method using up to 6 illumination frequencies and 1 illumination geometry. The behavior of of the

curves in figure 4.45 exhibits the same behavior as those in Figure 4.44.

L1 cooled roughness: Model error for the recovery of the large anomaly case with up to 6 illumination geometries.

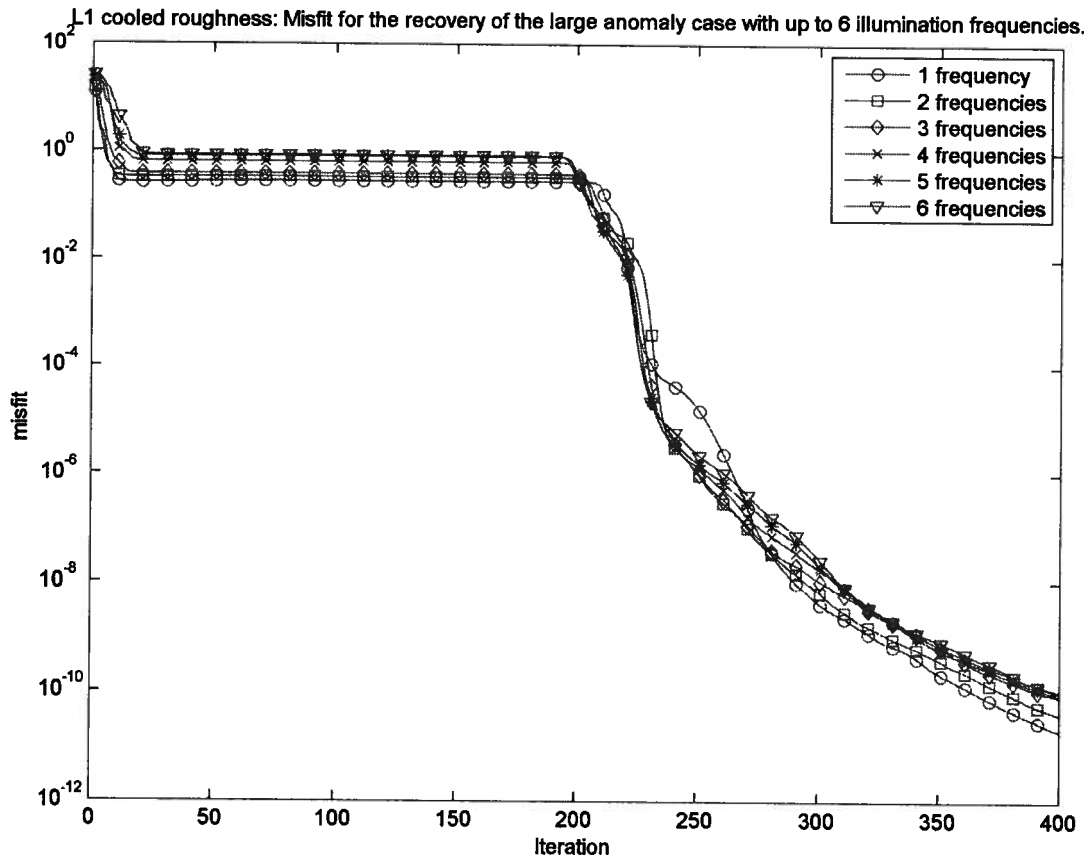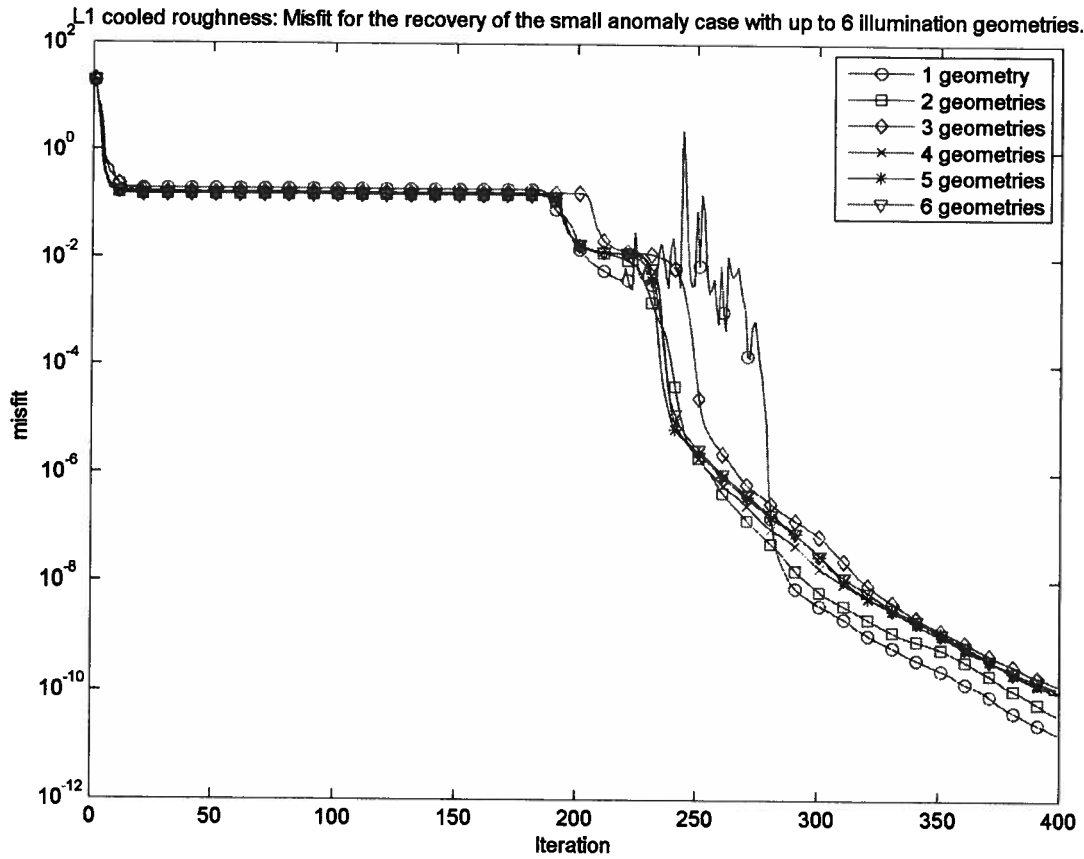*Figure 4.46: Model error curves for the recovery of the large anomaly case using the L_1 cooled roughness method with up to 6 illumination geometries and 1 illumination frequency.*

Figure 4.46 shows the model error curves for the recovery of the large with the CRL1 method

using up to 6 illumination geometries and 1 illumination frequency. For the 1 illumination frequency

and 1 illumination geometry case, the model errors rises at iteration 230. There is no corresponding rise

in the corresponding misfit curve in figure 4.42. What is seen is a dramatic change in slope at iteration

230. The likely explanation for this change in slope is that the CRL1 method was approaching a saddle

point for this case. This once again reinforces the idea that a decrease in misfit does not always imply a

decrease in model error. The remaining model error curves has a large initial decrease, which

corresponds to finding the best homogeneous match. A plateau region, where only small features are

permitted on the relative permitted distribution. A region of sleep decline, where visible features enter

159

the permittivity distribution. The region between iterations 240 and 400, where the the slope of the line approaches zero, indicates that the CRL1 algorithm is converging to a local minimum. This last region is not seen in the corresponding misfit curves, indicating that the local minima around the global minimum has small misfit values.
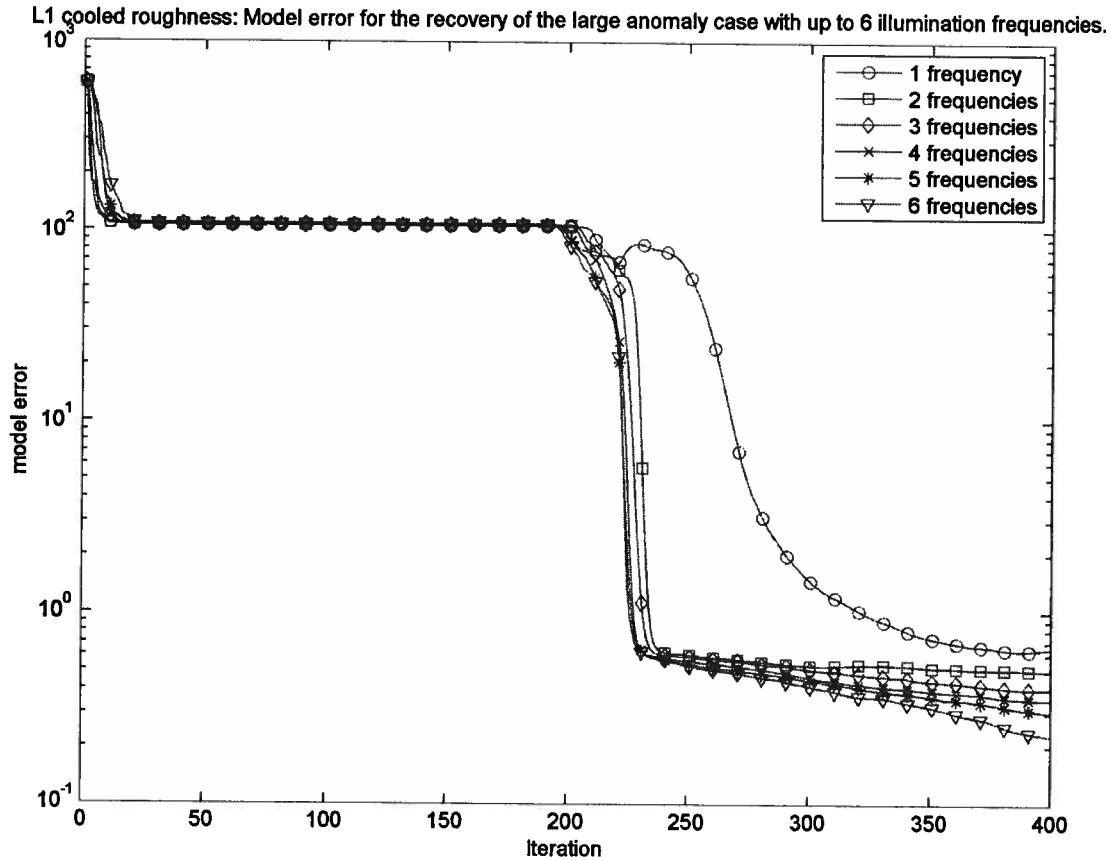
Figure 4.47: Model error curves for the recovery of the large anomaly case using the L_1 cooled roughness method with up to 6 illumination frequencies and 1 illumination geometry.

Figure 4.47 shows the model error curves for the recovery of the large anomaly case with the CRL1 method using up to 6 illumination frequencies and 1 illumination geometry. The curves in figure 4.47 exhibit the same behavior as the curves in 4.46.
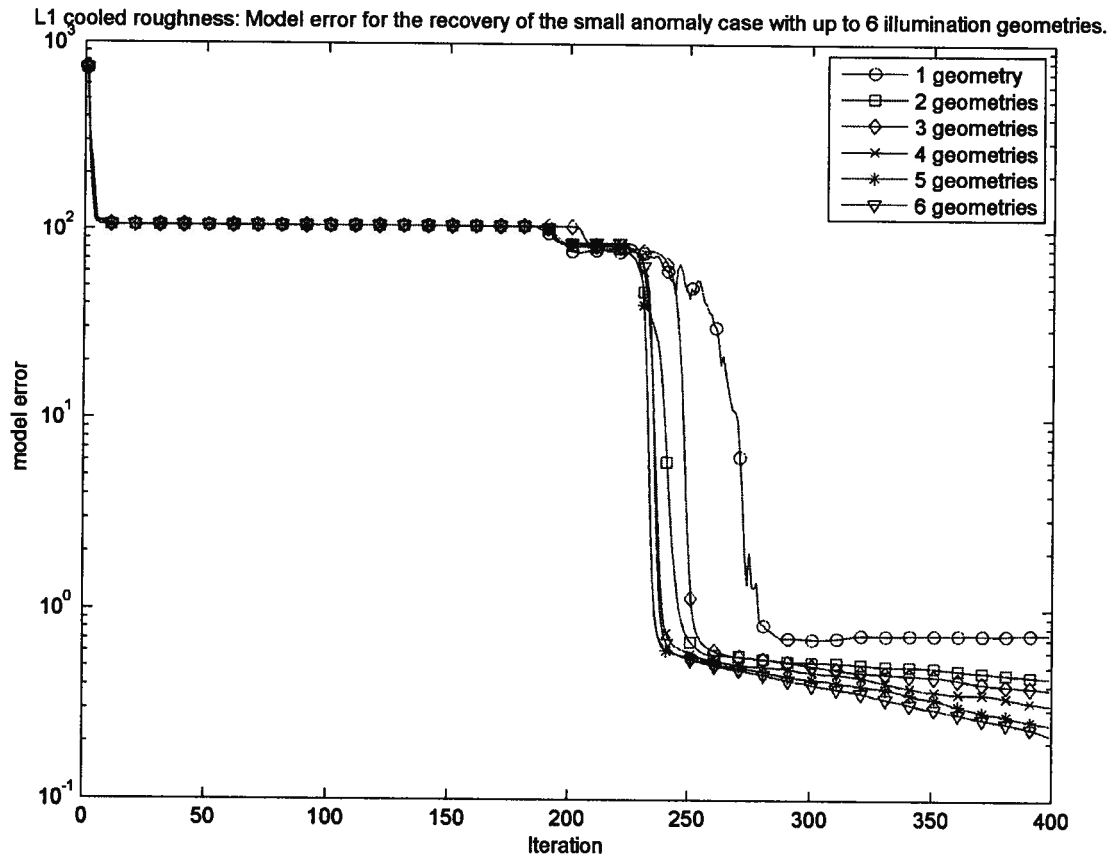
Figure 4.48: Model error curves for the recovery of the small anomaly case using the L_1 cooled roughness method with up to 6 illumination geometries and 1 illumination frequency.

Figure 4.48 shows the model error curves for the recovery of the small anomaly case with the CRL1 method using up to 6 illumination geometries and 1 illumination frequency. The instability in the misfit curve seen in figure 4.44 for 1 illumination geometry and 1 illumination frequency case did not lead to a substantial increase in model error. However, small increases in model error between iterations 230 and 280, the same interval where the misfit curve exhibits instability, can be seen.

The remaining model error curves exhibit the same behavior as the the model error curves in 4.46. The double plateau seen in the misfit curves in figure 4.44 is also visible in the model error curves.
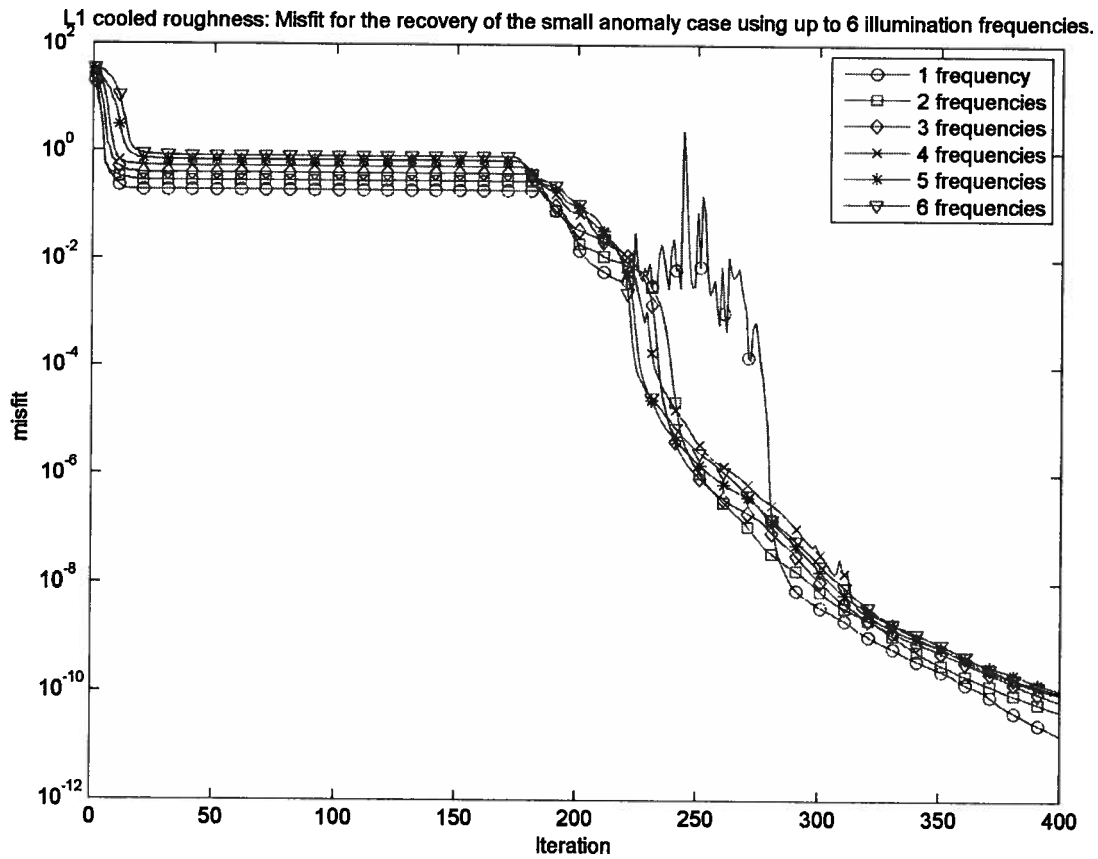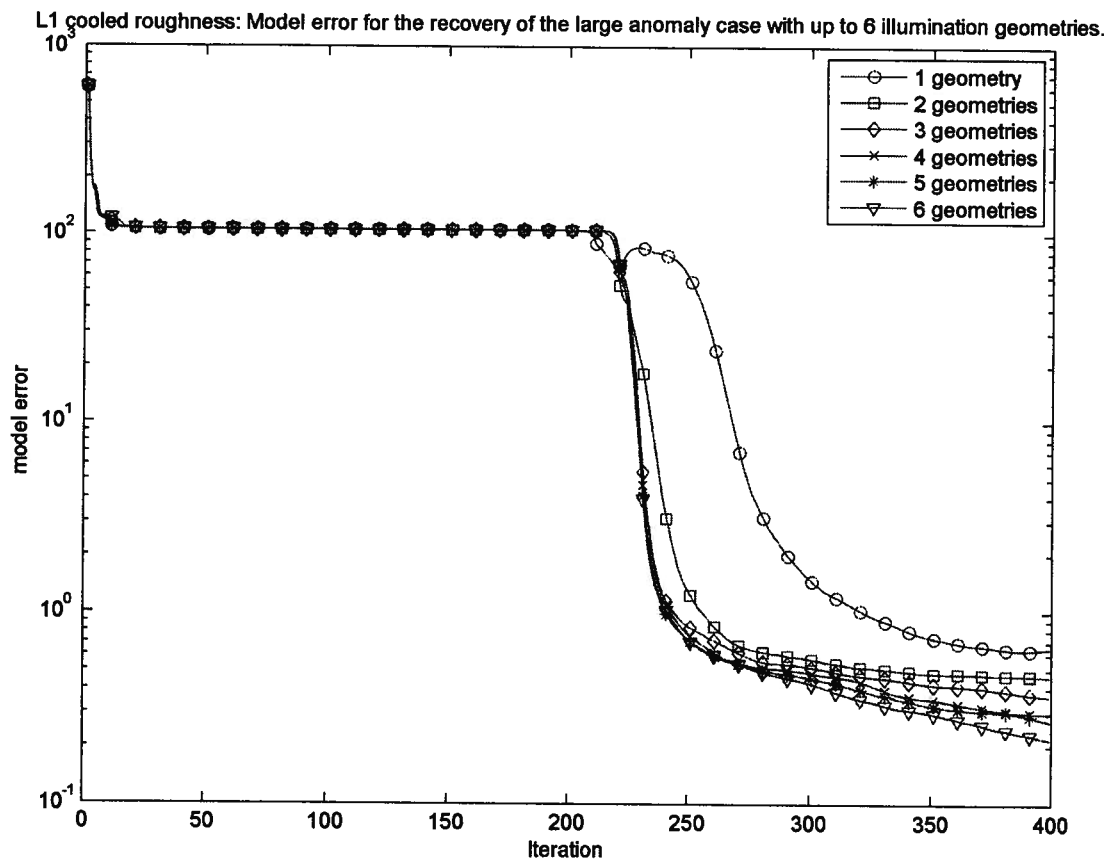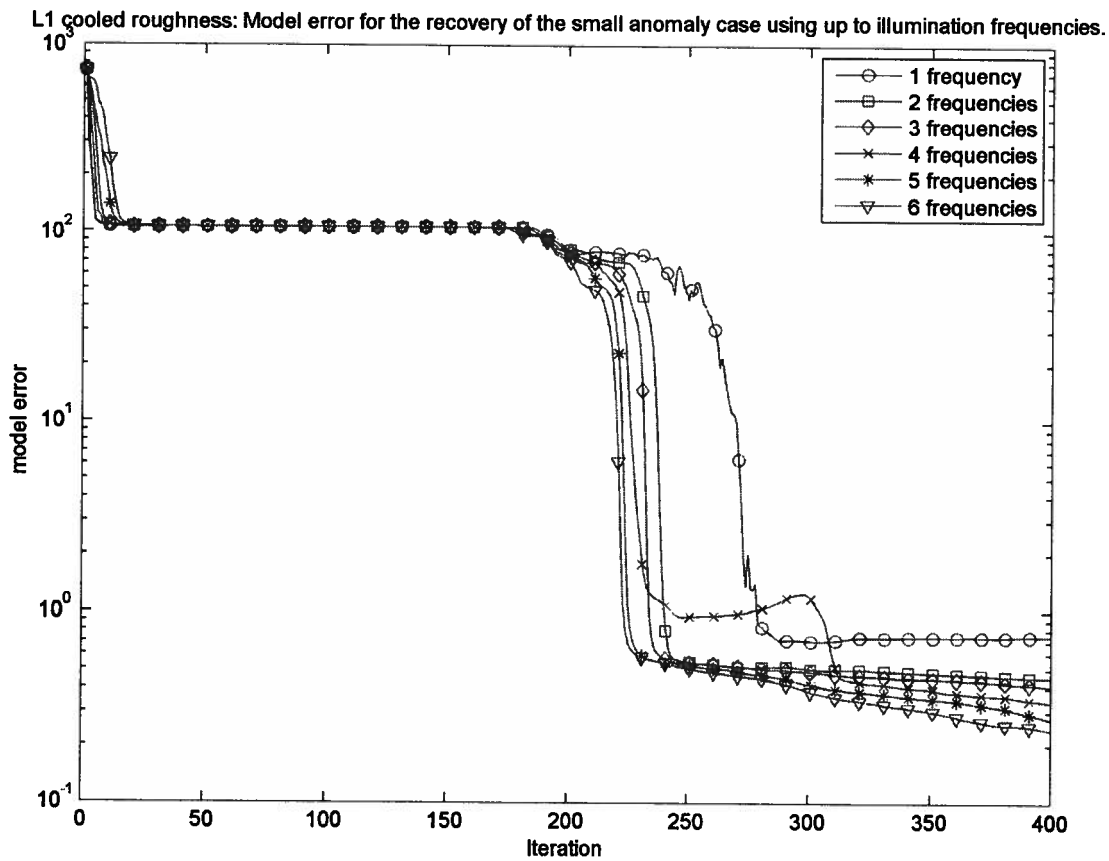
Figure 4.49: Model error curves for the recovery of the small anomaly case using the L1 cooled roughness method with up to 6 illumination frequencies and 1 illumination geometry.

Figure 4.49 shows the model error curves for the recovery of the large anomaly case with the

CRL1 method using up to 6 illumination frequencies and 1 illumination geometry. The case of 2

illumination frequencies and 1 illumination geometry shows a rise in model error between iterations

240 and 300. There is no corresponding feature in the misfit curve.

L1 cooled roughness: Relative permittivity distribution at the 220th iteration using 6 illumination geometries.

Figure 4.50: The recovered relative permittivity distribution at the 220th iteration for the recovery of the large anomaly case using the L1 cooled roughness method with 6 illumination geometries and 1 illumination frequency.

Figure 4.50 shows the recovered relative permittivity distribution at the 220th iteration for the recovery of the large anomaly case using the CRL1 method with 6 illumination geometries and 1 illumination frequency. The anomaly in figure 4.50 is in the correct position but is the wrong size, when compared to figure 4.1. The misfit value is 0.22 and the model error is 87.26. The misfit value is larger for the CRL1 method at the 220th iteration than the misfit value for the LM method for the comparable experiment. For the LM method, the misfit value is 1.19e-05. However, the model error for CRL1 at the 220th iteration is smaller than that of LM method. The model error value for the LM method is 237.85.

163

L1 cooled roughness: Relative permittivity distribution at the 400th iteration using 6 illumination geometries for the large anomaly case.

Figure 4.51: The recovered relative permittivity distribution at the 400th iteration for the recovery of the large anomaly case using the L_1 cooled roughness method with 6 illumination geometries and 1 illumination frequency.

Figure 4.51 shows the recovered relative permittivity distribution at the 400th iteration for the

recovery of the large anomaly case using the CRL1 method with 6 illumination geometries and 1

illumination frequency. The anomaly in figure 4.51 is in the correct position and of the correct size and

relative permittivity value.

Figure 4.52: Comparison of the cooled roughness methods at the y = 0 and z = 0 line, for the case of 6 illumination geometries and 1 illumination frequency. The CRL1 curve overlaps the reference curve from x=-0.33 to x=0.33.

Figure 4.52 compares the the cooled roughness methods at the y=0 and z =0 line, for the case of

6 illumination geometries and illumination frequency. For the reference relative permittivity value

curve, the values of the between x=-0.22 and x =0.22 is not constant. As mentioned in section 4.1 the

values are an uniform random distribution between 2.95 and 3.05.

The relative permittivity curve for the CRL1 at the 400th iteration matches the reference relative

permittivity curve the best. Between x=-0.22 and x =0.22 the recovered relative permittivity curve at

iteration 400 appears flatter than that of the reference relative permittivity curve. This is due to the

roughness regularization. The relative permittivity curve for CRL1 at th 220th iteration has the correct

general shape, but its value does not match the values of the reference curve.

The CRL2 curve has lower relative permittivity values than that of the reference curve and there

is dip in relative permittivity value at x =0. The dip is a result of the algorithm compensating for its

large penalization of sharp transitions.

L1 cooled roughness: Relative permittivity distribution at the 400th iteration using 6 illumination geometries for the small anomaly case.
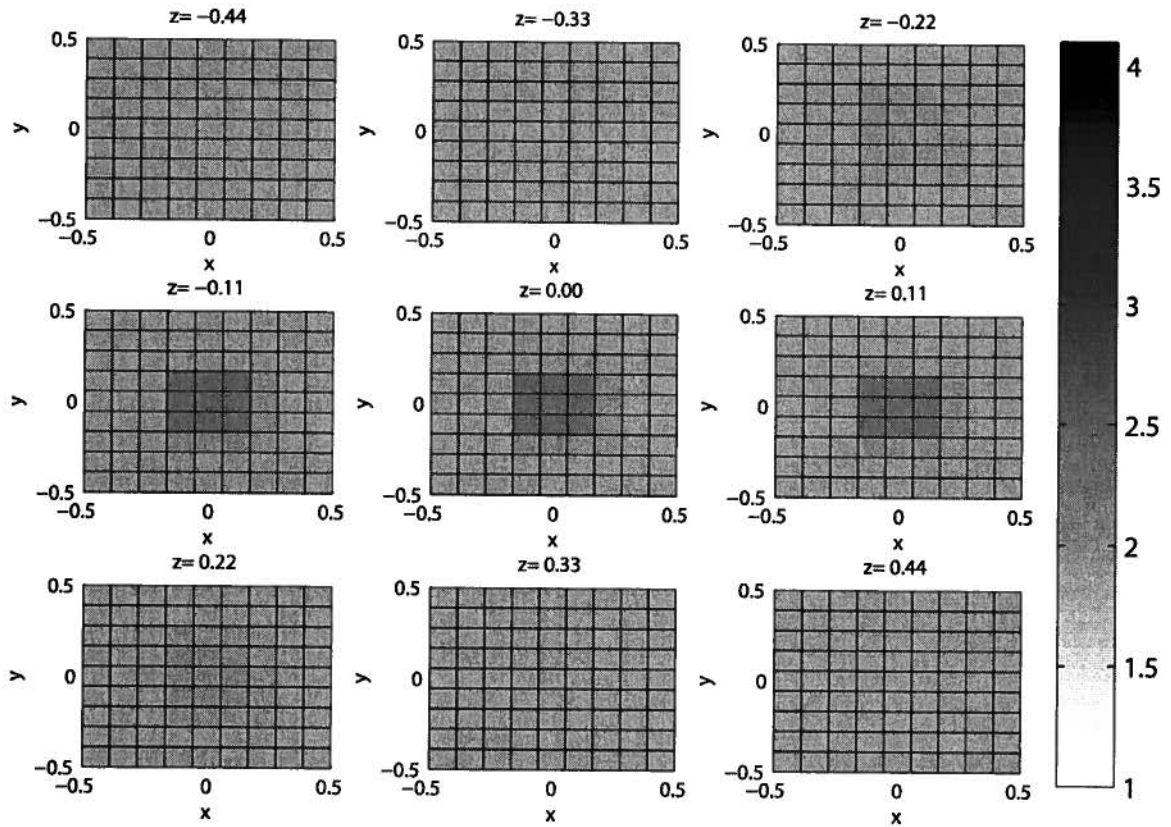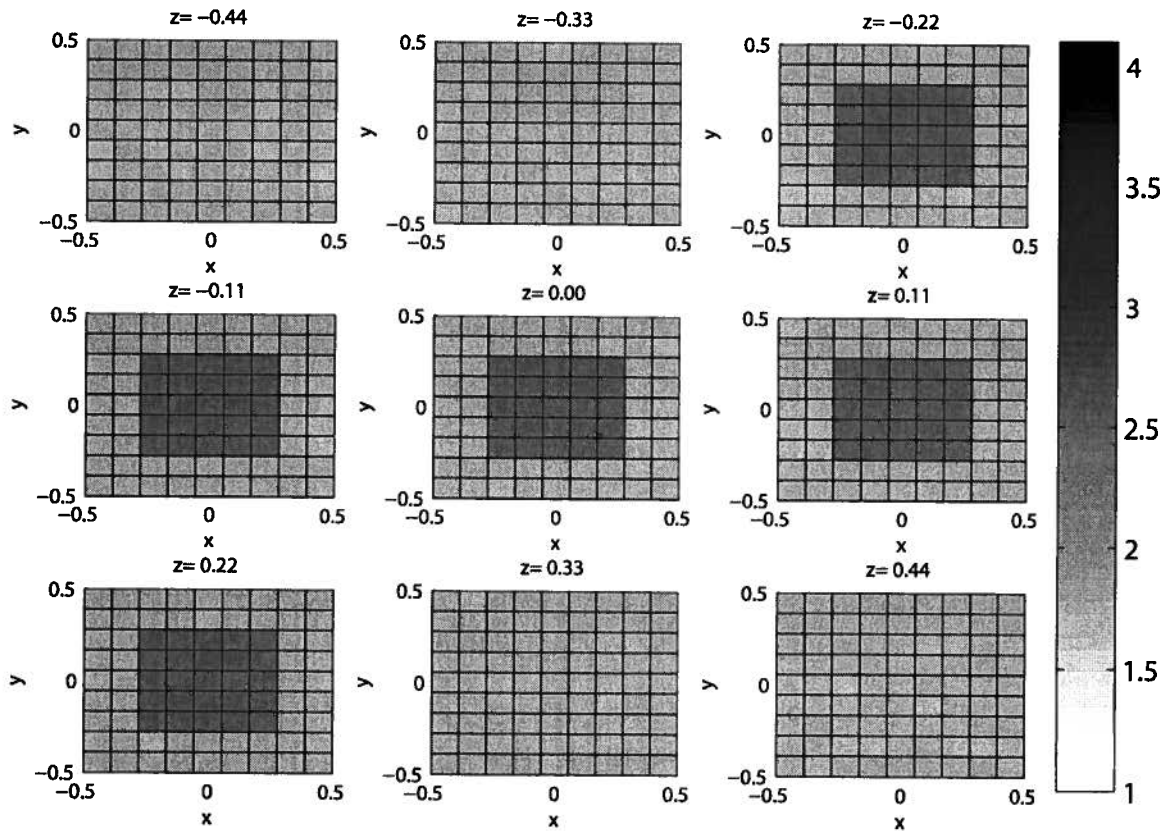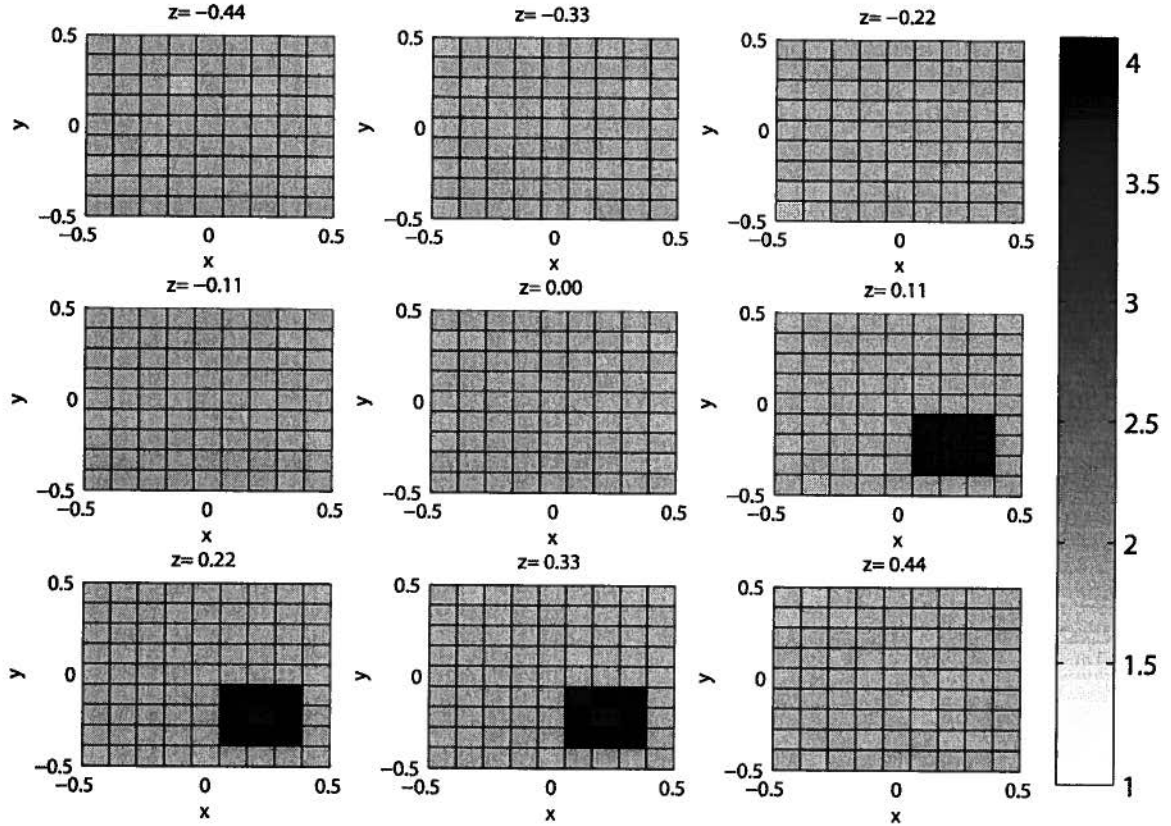

Figure 4.53: The recovered relative permittivity distribution at the 400th iteration for the recovery of the small anomaly case using the L1 cooled roughness method with 6 illumination geometries and 1 illumination frequency.

Figure 4.53 shows the recovered relative permittivity distribution for the small anomaly case using the CRL1 method with 6 illumination geometries and 1 illumination frequency at the 400th iteration. The recovered relative permittivity distribution resembles that of figure 4.2. In figure 4.43 the anomaly is centered around x = 0.22, y=-0.22, and z = 0.22, the same location as the anomaly in Figure 4.2.
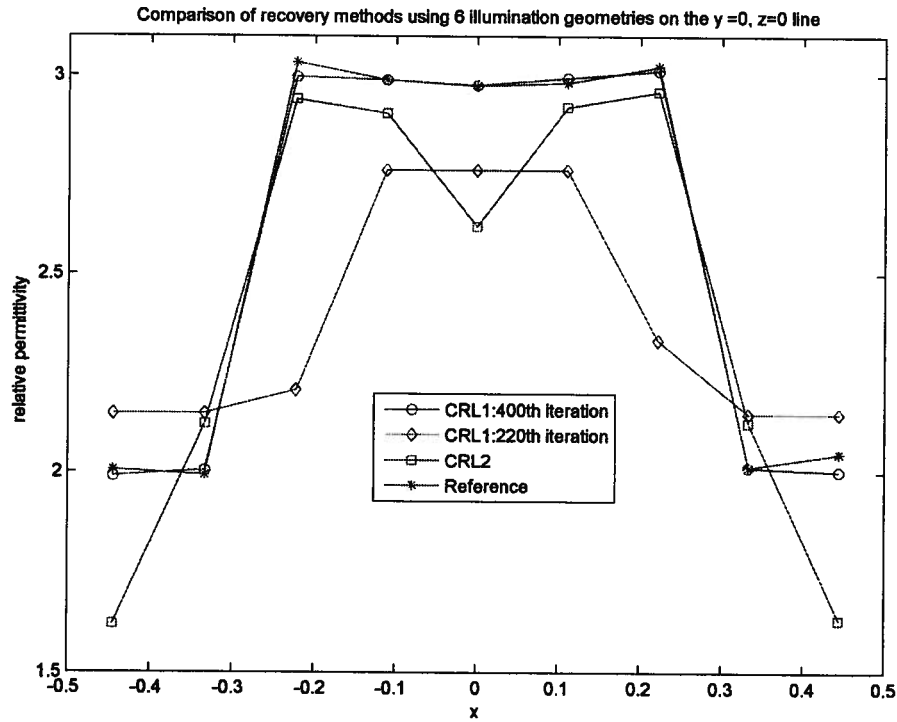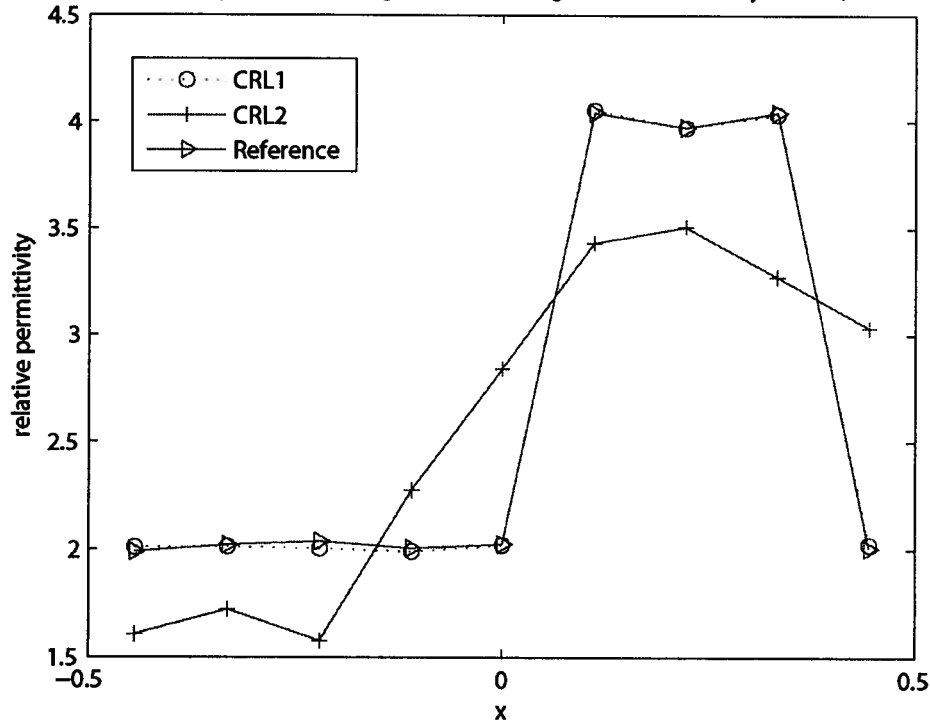
Figure 4.54: Comparison of the cooled roughness methods at the $y = -0.22$ and $z = 0.22$ line, for the case of 6 illumination geometries and 1 illumination frequency.

Figure 4.54 compares the recovered relative permittivity distribution for the CRL1 method,

CR2 method using 6 illumination geometries and 1 illumination frequency on the y=0-22 and z=0.22

line. The recovered relative permittivity distribution for the CRL1 method matches the actual relative

permittivity distribution closely. The recovered relative permittivity curve for the CRL1 method

appears flatter between x=-0.44 and x=0 than the reference relative permittivity curve.

The recovered relative permittivity distribution for the CRL2 method has features in the same

location as the actual relative permittivity distribution. However the relative permittivity values do not

match and between x = -0.22 and x = 0.11 the recovered CRL2 relative permittivity does not have a

sharp transition. The lack of sharp transitions is the cause of the blurriness seen in the CRL2 recovery

in figure 4.35.

4.8 The stopping condition

Up to now a fixed number of iterations was used as a stopping condition so that the behaviors of the algorithms could be explored. In a practical setting, a stopping condition is beneficial to prevent excess computation. From the results in this chapter, more than one stopping condition is needed. The stopping conditions are:

1) Stop after a predefined number of iterations

2) Stop when the misfit value fall belows a threshold

3) Stop at a fixed number of iterations after the misfit value begins rising.

The first stopping condition guarantees that the algorithm will stop, because there is no guarantee that the other 2 conditions will be met. If the algorithm reaches a local minimum with a high misfit value, then the misfit will not fall below the threshold required for condition 2.

The second stopping condition requires that a threshold be chosen. One method of choosing this threshold is based an estimate of the noise within the measured signal. Since noise is a random variable, the minimum misfit value for the threshold is defined as a expected value. Below this misfit value, the algorithm is fitting the noise. When fitting the noise, there is no guarantee that the model error will improve when the misfit improves. In the interests of saving computational effort, when the algorithm reaches the minimum misfit value defined by the noise, it is convenient to stop.

Let $\vec{n}$ be a noise vector representing the noise signal which is added to the noiseless data

$$\vec{d}^{\,noise} = \vec{d}^{\,pred} + \vec{n} \tag{4.3}$$

to give a new noisy data set. Plugging (4.3) into (3.1) and assuming that the reference field can be matched exactly results in

$$\Phi^{noise} = \sum_n \|n_n\|^2 \qquad (4.4)$$

for the minimum value for the misfit, which is simply the sum squares of the noise vector. Because the

noise is a random variable, the minimum threshold value can then be defined as the expectation

[Haykin 2001] of (4.4),

$$\Phi^{min} = E\left(\sum_n \|n_n\|^2\right) \qquad (4.5)$$

In (4.5), $E(x)$ is the expected value operator. In practice a higher value will be needed, due to the local

minima on the misfit surface.

The third stopping condition, where the algorithm is stopped before the misfit value rises, stops

an algorithm before it becomes unstable. From the CRL2 examples, it can be seen that a rise in the

misfit is an early indicator of the unstable behavior.

4.9 Conclusions

In this chapter the misfit surface has been thoroughly examined and many of the behaviors of the surface have been revealed.

From the gradient method, it was shown that the model error can sometime increase even when the misfit value decreases. This is most apparent when an algorithm converges around a local minimum. Adding more data does not help the gradient method find the global minimum because adding more data does not eliminate the local minima. The addition of more data however, does increase the misfit values of the local minima.

From the LM method, it was shown that an algorithm may never reach the bottom of a minimum unless the regularization is turned off. With regularization, the point that is found will always be a point between that of the minimum of the regularization and a minimum of the misfit surface. The LM method also showed that the misfit surface can be badly behaved and occasionally even find negative permittivity values. The addition of more data did not change this behavior. Therefore adding more data does not necessarily help an algorithm find the correct solution, as can be seen from the results of the LM algorithm and gradient algorithm. However, as shown in the cooled roughness method and L_1 cooled roughness method, the addition of data helps stabilize the algorithm when it is near the global minimum.

The CRL2 method demonstrated that filtering out undesirable relative permittivity distributions through the use of regularization aids in the recovery of the target distribution. The recovered relative permittivity distribution from the CRL2 method resembled the target relative permittivity distribution better than the recovered relative permittivity distribution from the LM method.

In addition to producing results with smaller misfits, CRL1 and CRL2 also solved the starting model problem. As seen in the LM method and the gradient method, an algorithm that relies solely on

local information will get trapped in a local minimum quickly. The only way for such an algorithm to find the solution is if it starts at an initial point close to the global minimum. By introducing the the cooled roughness regularization, the system was able to reach a point close to the global minimum by slowly introducing structure into the relative permittivity distribution. The starting model then, should be a homogeneous surface since it has the least features. The cooling algorithm will find the best homogeneous match at the early iterations, implying that the cooled roughness methods are insensitive to the starting relative permittivity value.

Lastly, the recovery with a L_1 norm produced better results than that of a L_2 norm for the cases presented . This due to the fact that the L_1 regularization matches the target permittivity distribution better than the L_2 regularization. The recovered relative permittivity distribution with CRL1 has less artifacts and is less blurry. Even the intermediate permittivity distribution was less blurry and had less artifacts, reinforcing the idea that the blurriness in the CRL2 recovery is due to the L2 norm.

Chapter 5

Limits of the L1 cooled roughness regularization

5.1 Introduction to chapter 5

In chapter 4, the cooled roughness regularization has been shown to be able to recover different relative permittivity distributions. However, the the permittivity distributions in chapter 4 were suited for the cooled roughness regularization. In this chapter the CRL1 method will be used to recover more complicated relative permittivity distributions. Four examples will demonstrate the limits of the CRL1 method. Four illumination geometries and 1 illumination frequency will be used.

The first example presented is an extension of the cases in chapter 4. It contains 3 anomalies and illustrates that the CRL1 method can recover relative permittivity distributions with more than 1 anomaly. The second example is a random relative permittivity distribution. This example represents a case that is filtered by the roughness regularization. The third example is a smooth anomaly. It is an anomaly that is unfavorable for the CRL1 method because the anomaly is not homogeneous The forth example is a checkered relative permittivity distribution. The fourth example gives a sense of the resolution the algorithm is able to recover. The effect of noise will also be explored with this example.

In short, the CRL1 will have difficulties with some of the examples, highlighting the difficulties with this type of recovery.

5.2 A case with 3 anomalies

The 3 anomaly case
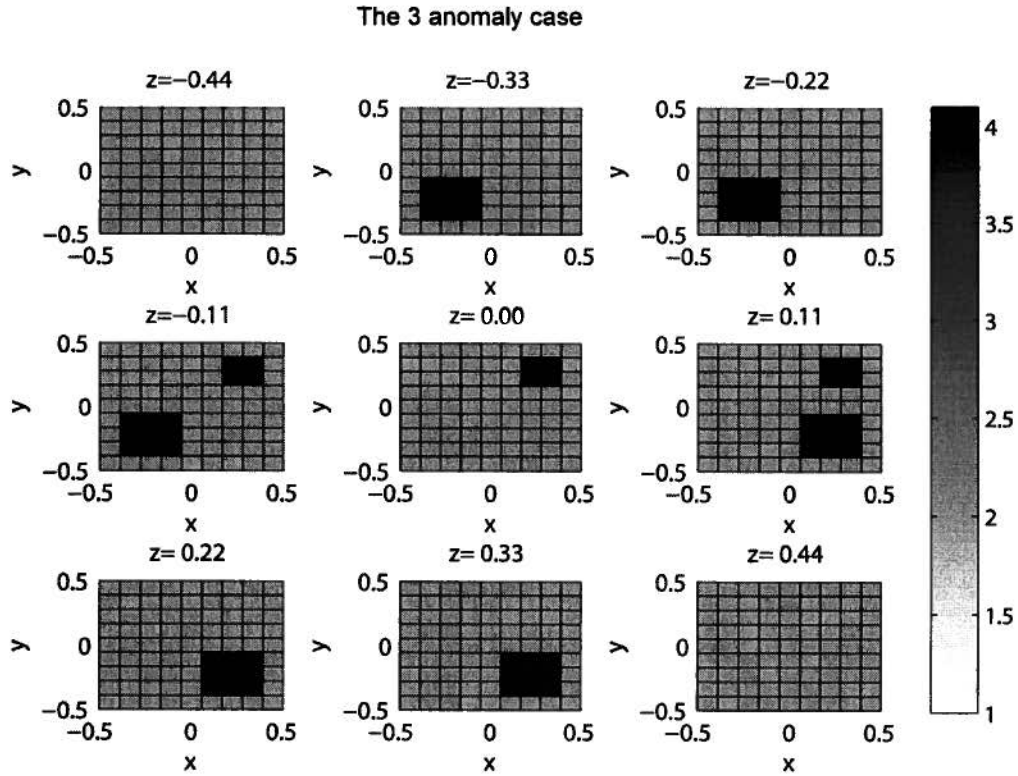
Figure 5.1: Relative permittivity distribution with 3 cube shaped anomalies

In this section of the recovery of the relative permittivity distribution shown in figure 5.1 with the CRL1 method using 4 illumination geometries and 1 illumination frequency will be examined. The illumination geometries are I1 to I4 seen in table 4.1 and the illumination frequency is 0.8GHz. Like chapter 4, a starting model of a homogeneous relative permittivity distribution with a relative permittivity of 3.0 is used.

The permittivity distribution in figure 5.1 has 3 anomalies inside a large cube. The cube has sidelength of 1 free space wavelength at 1 GHz. Their values are a uniform random distribution between 3.95 and 4.05. The relative permittivity distribution seen in figure 5.1 shall be known as the 3 anomaly case. The relative permittivity value of the anomalies is a uniform random distribution between 3.95 and 4.05. This represents an extension of the examples presented in Chapter 4.

173

Figure 5.2 shows the misfit curve for the recovery of the 3 anomaly case using the CRL1

method with 4 illumination geometries and 1 illumination frequency. The final misfit value is

3.38e-11. The misfit curves resemble those seen for CRL1 experiments in chapter 4. The curve has an

initial drop, corresponding to the CRL1 method finding the best homogeneous match. There is a

plateau region, where small scale changes enter the relative permittivity distribution. There is a region

of steep decline where visible features enter the relative permittivity distribution and a region where the

value converges to a local minimum.

Figure 5.2: Misfit curve for the recovery of the 3 anomaly case using the CRL1 method with 4 illumination geometries and 1 illumination frequency.

174

Figure 5.3: Model error curve for the recovery of the 3 anomaly case using the CRL1 method with 4 illumination geometries and 1 illumination frequency at 0.8GHz

Figure 5.3 shows the model error curve for the recovery of the 3 anomaly case using the CRL1 method with 4 illumination geometries and 1 illumination frequency. The final misfit value is 0.5865. The model error is comparable to those seen in table 4.7. The behavior of the model error resembles that of the misfit curve seen in figure 5.2. The small value of model error implies that the CRL1 has successfully recovered the target relative permittivity distribution.

175

Recovered Relative Permittivity Distribution
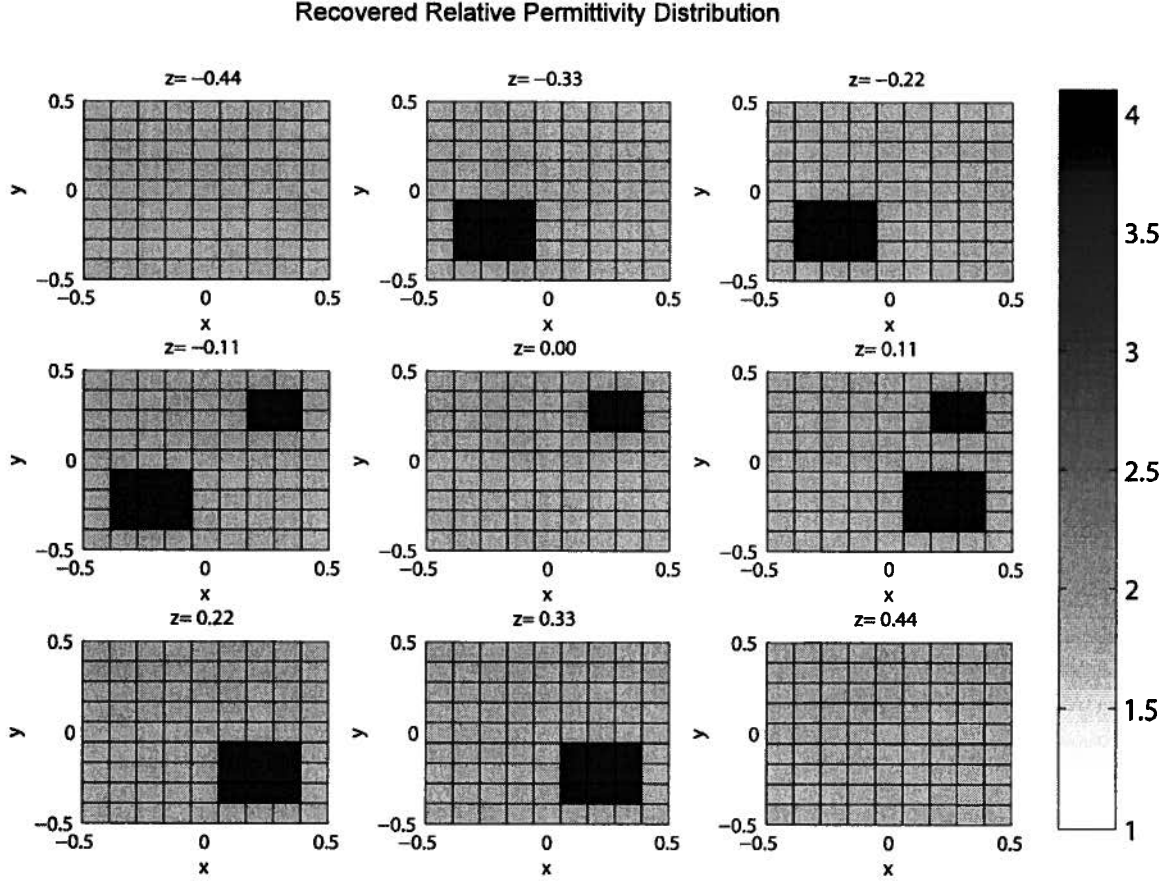
Figure 5.4: Recovered relative permittivity distribution for the recovery of the 3 anomaly case using the CRL1 method with 4 illumination geometries and 1 illumination frequency.

Figure 5.4 shows the recovered relative permittivity distribution for the recovery of the 3

anomaly case using the CRL1 method with 4 illumination geometries and 1 illumination frequency.

Figure 5.4 resembles figure 5.1. The 3 anomalies have been successfully recovered with the correct

permittivity values at the correct location. This example shows that the CRL1 method is able to

recovered more complicated relative permittivity distributions than those presented in chapter 4. The

CRL1 method was able to recover the above relative permittivity distribution because the structure of

the target relative permittivity distribution, matches that of the regularization.

176

5.3 The recovery of an unfavorable case for CRL1: The random anomaly
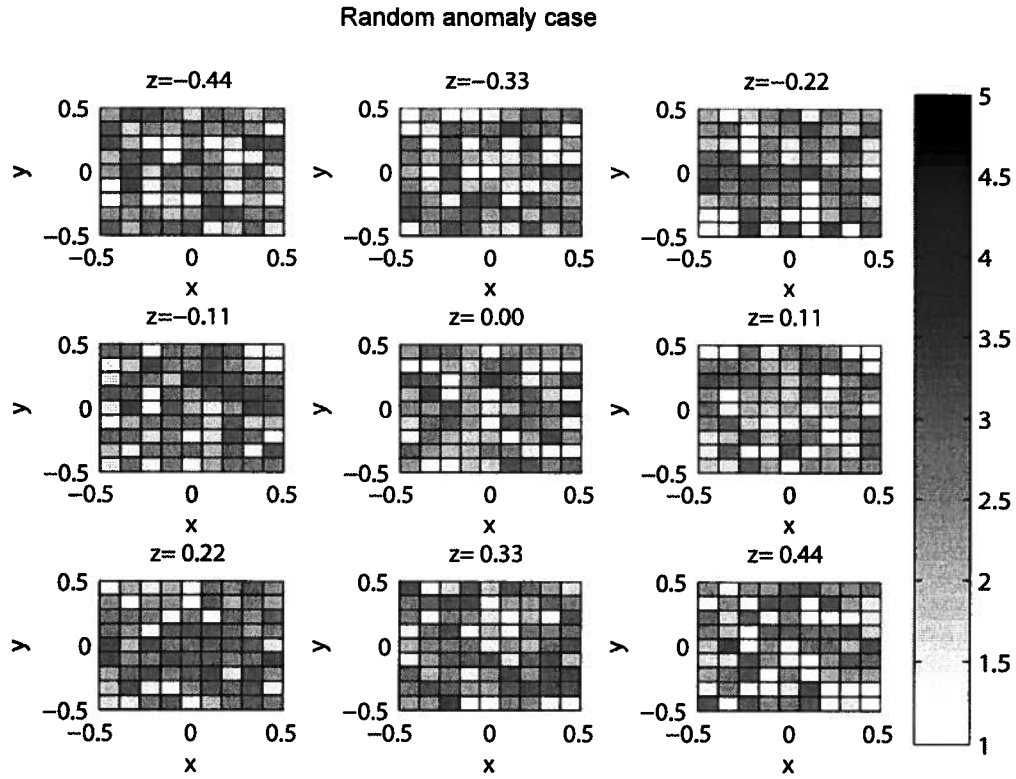
Figure 5.5: Random relative permittivity distribution.

In this section, the recovery of a cube with a sidelength of 1 free space wavelength at 1 GHz where the relative permittivity value is an uniform random distribution from 0.95 to 5.05 using the CRL1 method is explored. This represents a highly unfavorable case for the roughness regularization, because the relative permittivity distribution. shown in figure 5.5 has many sharp transitions and no homogeneous regions. Recall that the roughness regularization acts as a filter. This example explores the behavior of the CRL1 method, when the target relative permittivity distribution is one that is filtered by the regularization.

For this example the starting model is a homogeneous cube with relative permittivity of 3.0. The illumination geometries are I1 to I4 seen in table 4.1 and the illumination frequency is 0.8GHz.

CRL1: Misfit curve for the recovery of the random anomaly case
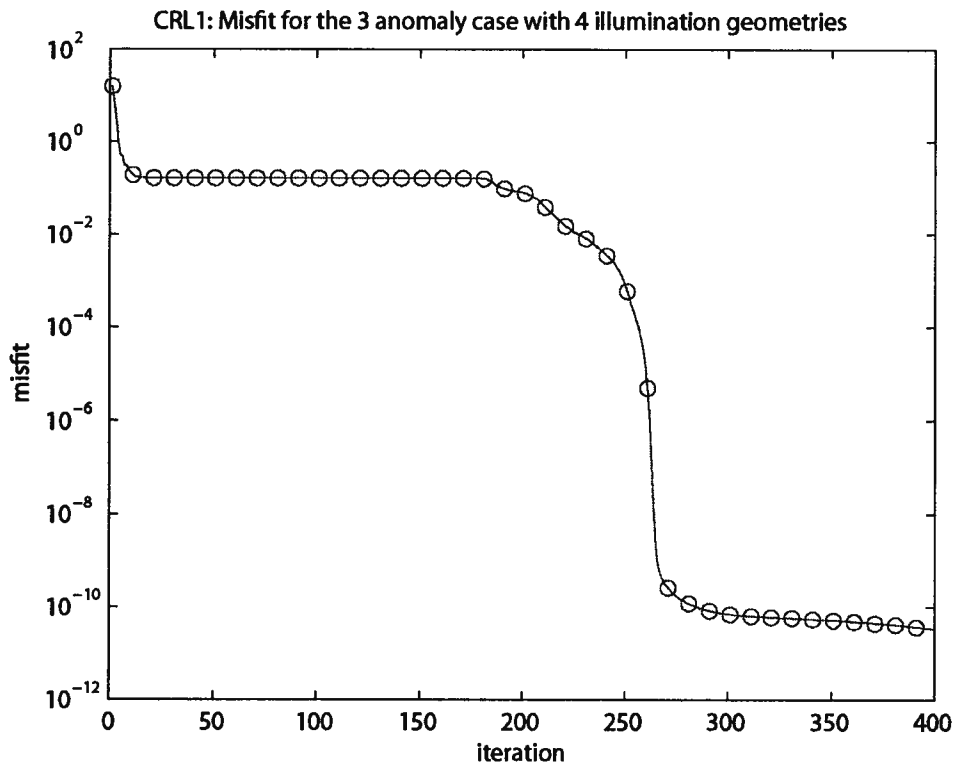with 4 illumination geometries and 1 illumination frequency.

Figure 5.6: Misfit curve for the recovery of the random anomaly case using the CRL1 method with 4 illumination geometries and 1 illumination frequency at 0.8GHz.
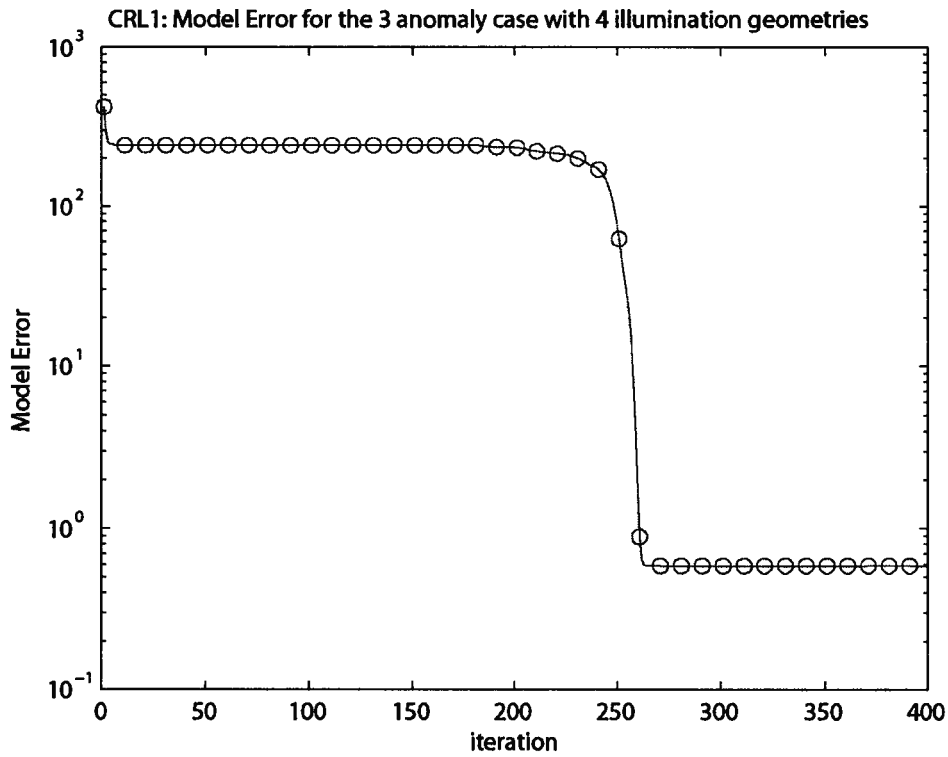
Figure 5.6 shows the misfit curve for the recovery of the random anomaly case using the CRL1 method with 4 illumination geometries and 1 illumination frequency. The recovered relative permittivity distribution will be taken at iteration 234 where the misfit curve initially increases. The misfit value at this iteration is 2.11. The misfit curve has an initial decrease, followed by a plateau region. The misfit then decreases sharply and at iteration 234, begins to oscillate. The algorithm becomes unstable after iteration 234. The inability for the CRL1 method to reach a small misfit value was expected, because the target relative permittivity distribution did not have any homogeneous regions and had many sharp transitions. Both those features are penalized by the regularization.

CRL1: Model error curve for the recovery of the random anomaly case
with 4 illumination geometries and 1 illumination frequency.

Figure 5.7: Model error curve for the recovery of the random anomaly case using the CRL1 method with 4 illumination geometries and 1 illumination frequency at 0.8GHz.

Figure 5.7 shows the model error curve for the recovery of the random anomaly case using the

CRL1 method with 4 illumination geometries and 1 illumination frequency. The value of the model

error at iteration 234 is 4340. The model error curve rises initially, followed by a plateau region, and

then a begin to sharply rise at about iteration 210. The initial rise in misfit curve is likely due to the

local minimum behavior discussed in chapter 4. The rise in misfit at approximately iteration 210 is due

to the mismatch between the regularization and target distribution. The regularization filters relative

permittivity distributions like the target relative permittivity distribution. As the algorithm progresses

the recovered relative permittivity distribution becomes more dissimilar to the target relative

permittivity distribution, leading to a rise in model error.

CRL1: Recovered relative permittivity for the recovery of the random anomaly case
with 4 illumination geometries and 1 illumination frequency.
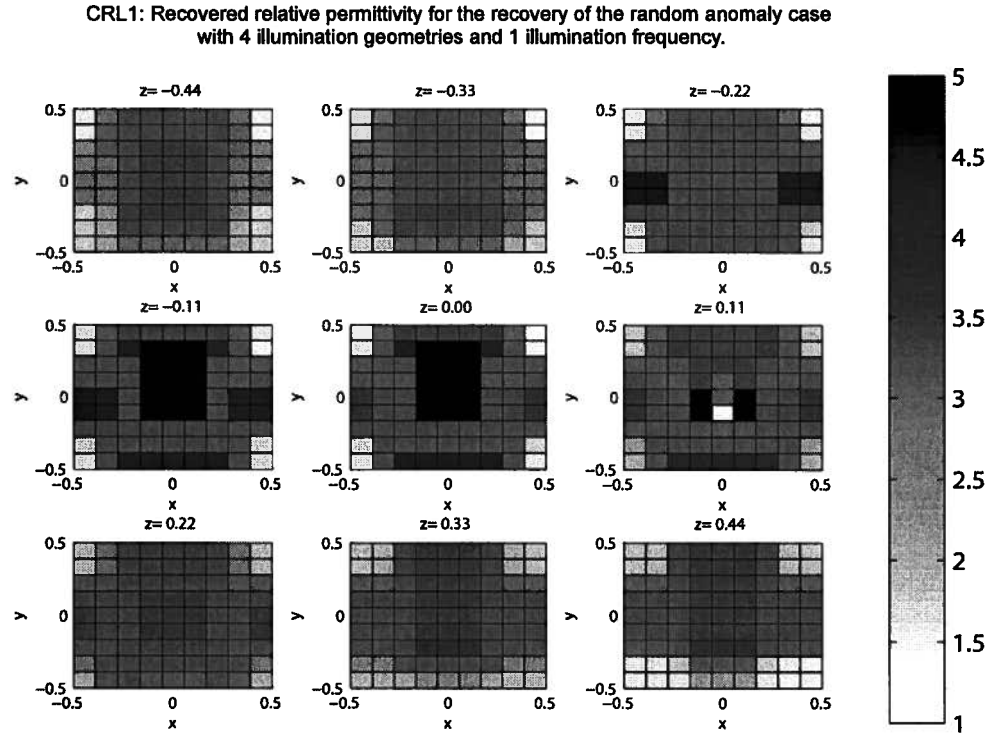
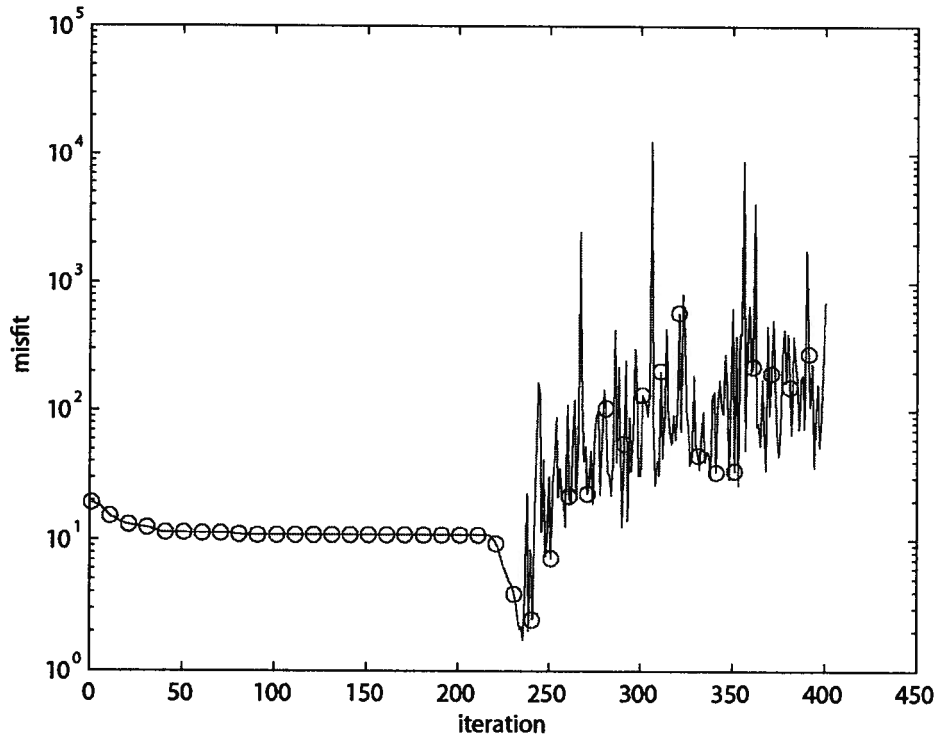*Figure 5.8: Recovered relative permittivity distribution for the recovery of the random anomaly case using the CRL1 method with 4 illumination geometries and 1 illumination frequency at 0.8GHz.*

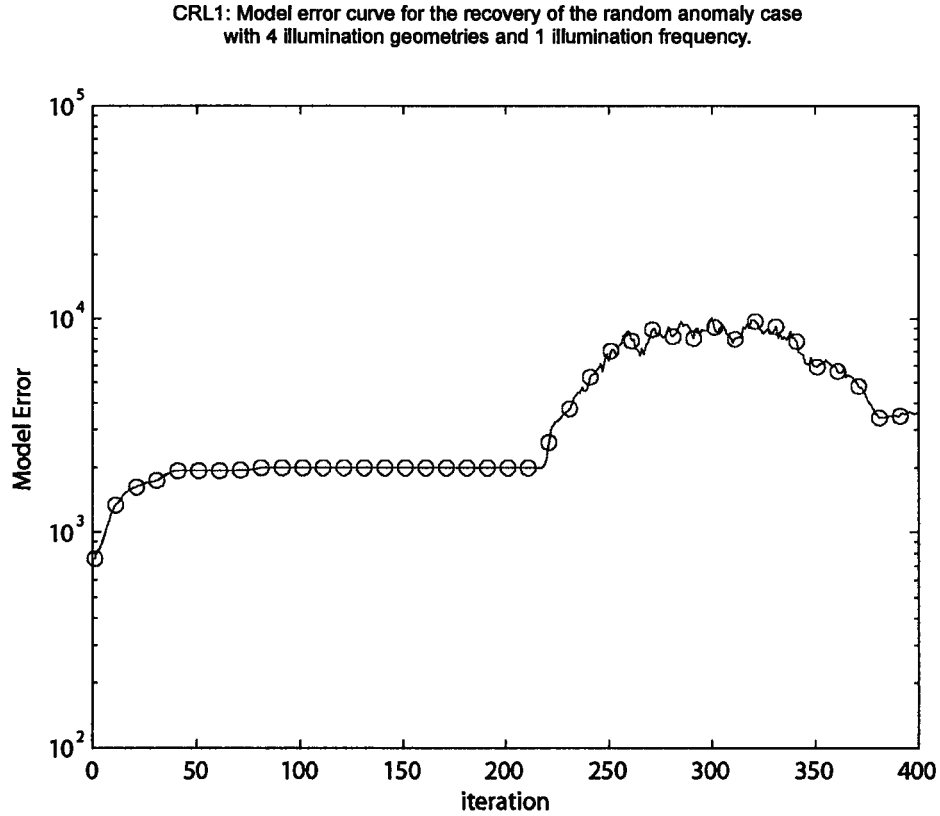Figure 5.8 shows the model recovered relative permittivity distribution for the recovery of the random anomaly case using the CRL1 method with 4 illumination geometries and 1 illumination frequency. The recovered relative permittivity distribution does not resemble the target relative permittivity distribution seen in figure 5.5. Large homogeneous regions are visible. These types of regions are favored by the CRL1 method and their appearance in the recovered relative permittivity distribution is not surprising. This illustrates that the CRL1 method suffers from the local minima behavior as seen in the other methods. For cases that are heavily penalized, the CRL1 method may not be able to recover the target distribution. It also demonstrates that the CRL1 method has a preference for recovering relative permittivity distributions with homogeneous areas, since the recovered structure in this case is not supported by the data but contains predominantly homogeneous areas.

This illustrates one of the limits of the CRL1 method. It cannot recover random relative

permittivity distributions. Further, it is expected the CRL1 method will have difficulty with any

relative distribution with many sharp transitions and lack homogeneous regions. A second possibility

for the failed recovery is due to the small feature size of the targets, relative to the wavelength.

5.4 Recovery of a parabola shaped anomaly with CRL1 and CRL2

In this section, the recovery of a cube of sidelength 1 free space wavelength at 1 GHz. The

cube contains a blurry anomaly in the center. From the results of chapter 4, the recovered relative

permittivity from the CRL2 methods are blurry. This examples tests the converse relationship, of

whether the blurry anomalies are easily recovered by the CRL2 method.

Paraboloid anomaly case
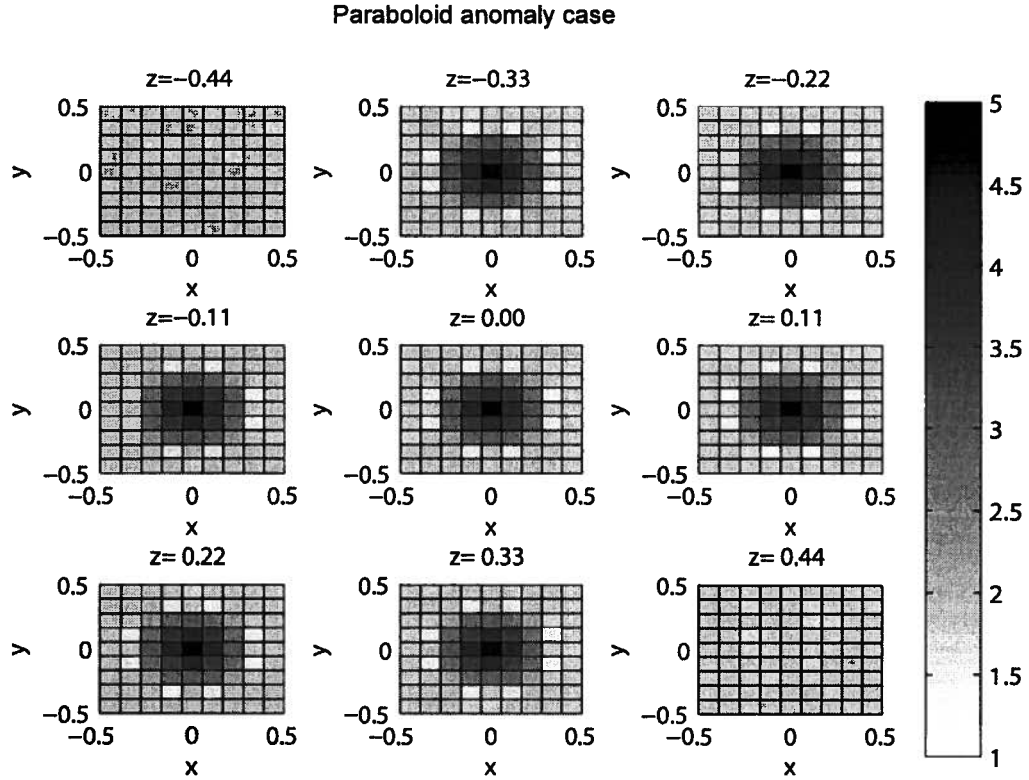
Figure 5.9: A cube of sidelength 1 free space wavelength at 1 GHz with a paraboloid shaped anomaly in the center.

Figures 5.9 shows a the target relative permittivity distribution, which is a cube with sidelength

1 free space wavelength at 1GHz and an anomaly at its center. The anomaly is blocky in the z direction

and blurry in the x and y directions. The background relative permittivity has a value of 2.0. The

anomaly is constant along the z direction between z = -0.4 and z =0.4. The misfit value within the

cylinder defined by $x^2 + y^2 < 0.16$, follows the equation

$$\varepsilon(x, y) = 2 + 27\left(\frac{1}{9} - x^2 - y^2\right)$$

which has the shape of an inverted parabola, when plotted as a function of the distance from the center of the circle.

Relative permittivity cross section on the y=0. z=0 line
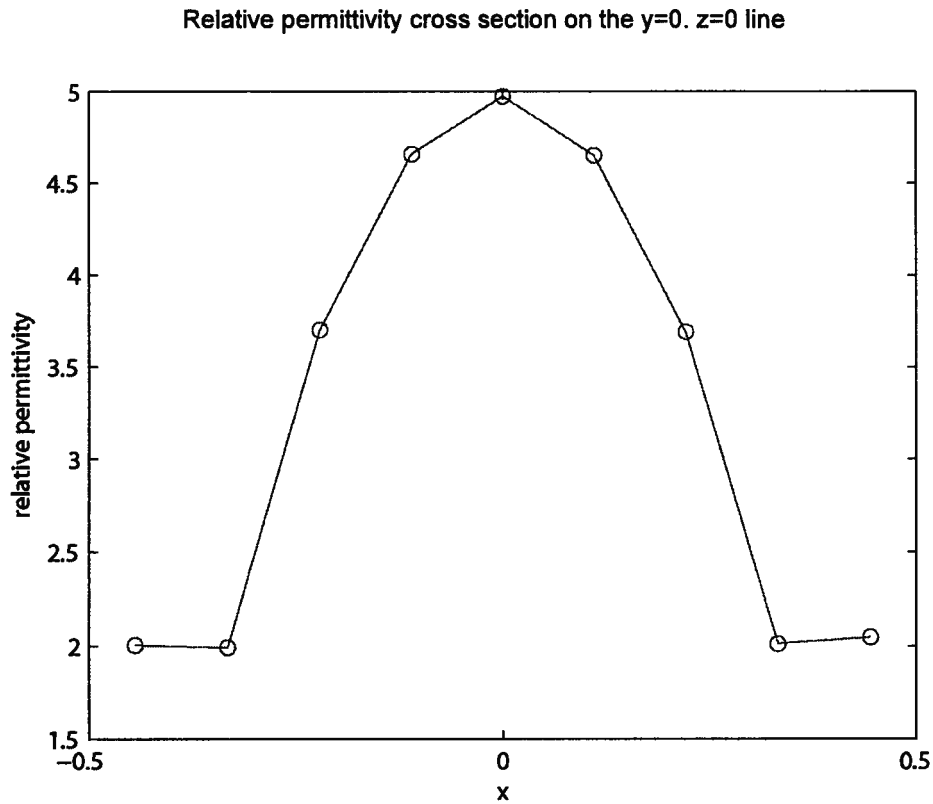
Figure 5.10: Parabola shaped relative permittivity cross section on the y=0 and z=0 line

Figure 5.10 shows the relative permittivity along the y=0 and z=0 line. The relative permittivity is shaped like an inverted parabola, with a maximum value for 5.0 and a minimum value of 2.0.

Misfit curves for the recovery of the paraboloid anomaly case
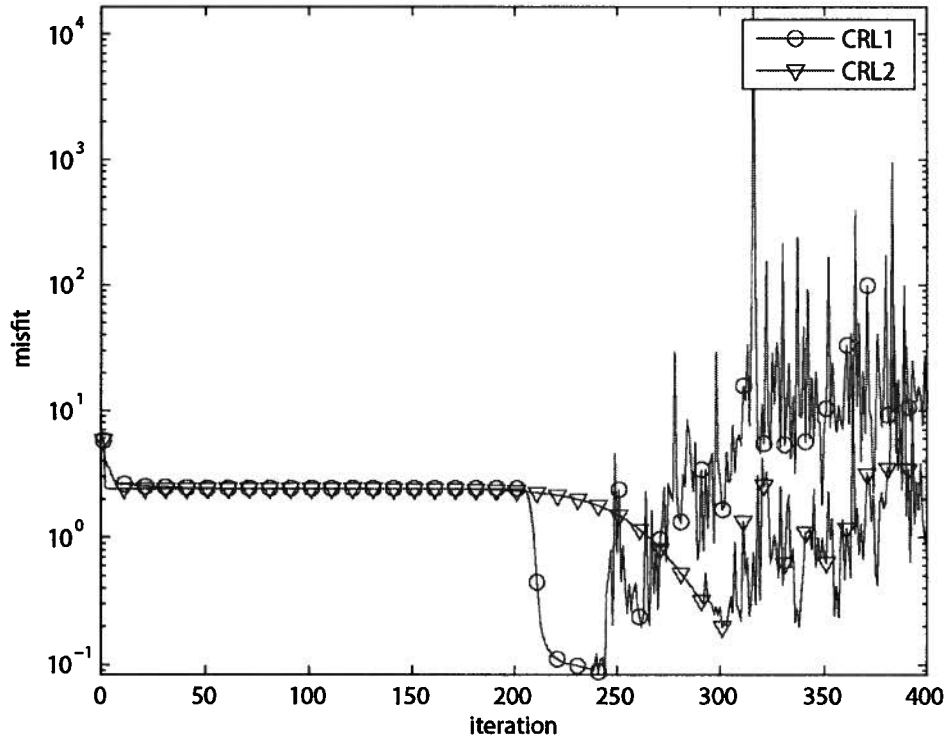
Figure 5.11: Misfit curves for the recovery of the paraboloid anomaly case using the CRL1 and CRL2 method with 4 illumination geometries and 1 illumination frequency at 0.8GHz.

Figure 5.11 shows the misfit curves for the recovery of the paraboloid anomaly case using the

CRL1 and CRL2 method with 4 illumination geometries and 1 illumination frequency. Both curves

exhibit similar behavior. They have an initial decrease in misfit, followed by a plateau. After the

plateau there is a decrease in misfit fallowed by a region of oscillating misfit values. The CRL1

method achieved a lower misfit value. The misfit value for the CRL1, taken at iteration 243, is 0.0840.

The misfit value for the CRL2 method, taken at iteration 300, is 0.1986.

Model error curves for the recovery of the paraboloid anomaly case
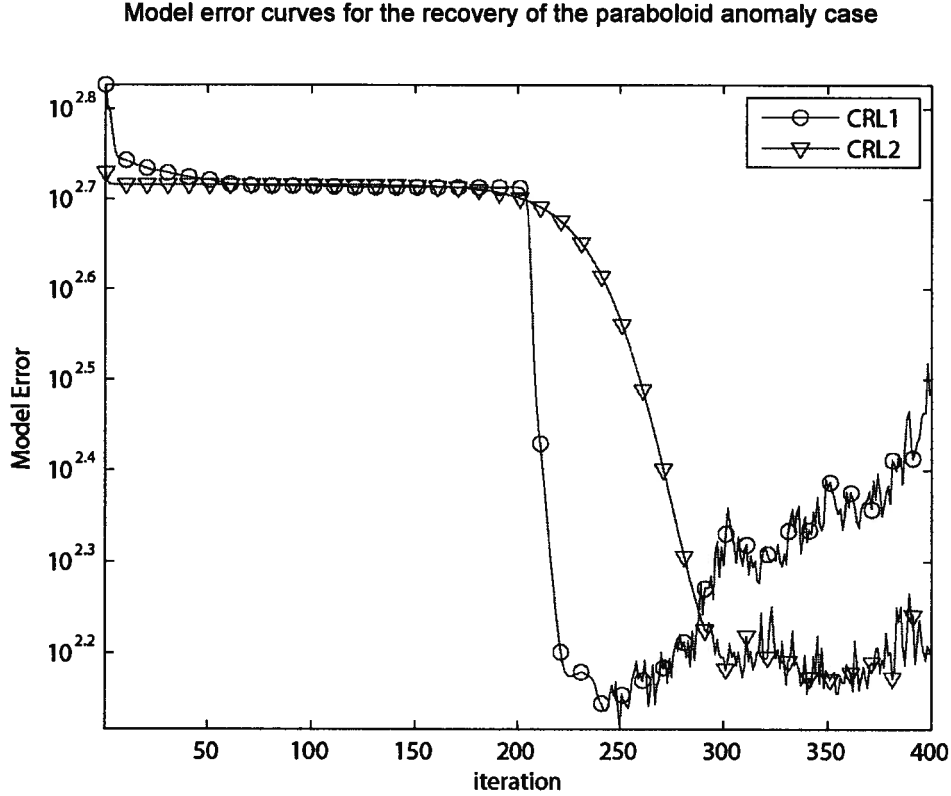
Figure 5.12: Model error curves for the recovery of the paraboloid anomaly case using the CRL1 and CRL2 method with 4 illumination geometries and 1 illumination frequency at 0.8GHz.

Figure 5.12 shows the misfit curves for the recovery of the paraboloid anomaly case using the

CRL1 and CRL2 method with 4 illumination geometries and 1 illumination frequency. The value of the

model error for the CRL1 method at iteration 243 is 136.6. The value of the model error for the CRL2

method at iteration 300 is 152.2. The CRL1 produced a marginally better result than the CRL2

method, in terms of model error. This implies that the CRL2 method has approximately the same

performance as CRL1 method. For this case the anomaly is blocky in one direction, and blurry in two.

This mix of feature type is the reason why the two algorithms have similar performance. The target

relative permittivity distribution is favored by neither algorithm.

CRL1: Recovered relative permittivity distribution at iteration 233 for the
recovery of the paraboloid anomaly case with 4 illumination geometries.

Figure 5.13: Recovered relative permittivity distribution at iteration 233 for the recovery of the paraboloid anomaly case with the CRL1 using 4 illumination geometries and 1 illumination frequency.

Figure 5.13 shows the recovered relative permittivity distribution at iteration 233 for the

recovery of the paraboloid anomaly case with the CRL1 using 4 illumination geometries and 1

illumination frequency. There are some similarities between recovered relative permittivity distribution,

figures 5.13, and the target relative permittivity distribution, figure 5.9, however they do not look

exactly the same. Both share a central anomaly at the point x=0, y=0, z=0. The anomaly in the

recovered relative permittivity distribution looks like a homogeneous block, but in the target relative

permittivity distribution it looks blurry. Generally, the target looks more blurry than that of the

recovered relative permittivity distribution.

CRL2: Recovered relative permittivity distribution at iteration 300 for the recovery of the paraboloid anomaly case with 4 illumination geometries.
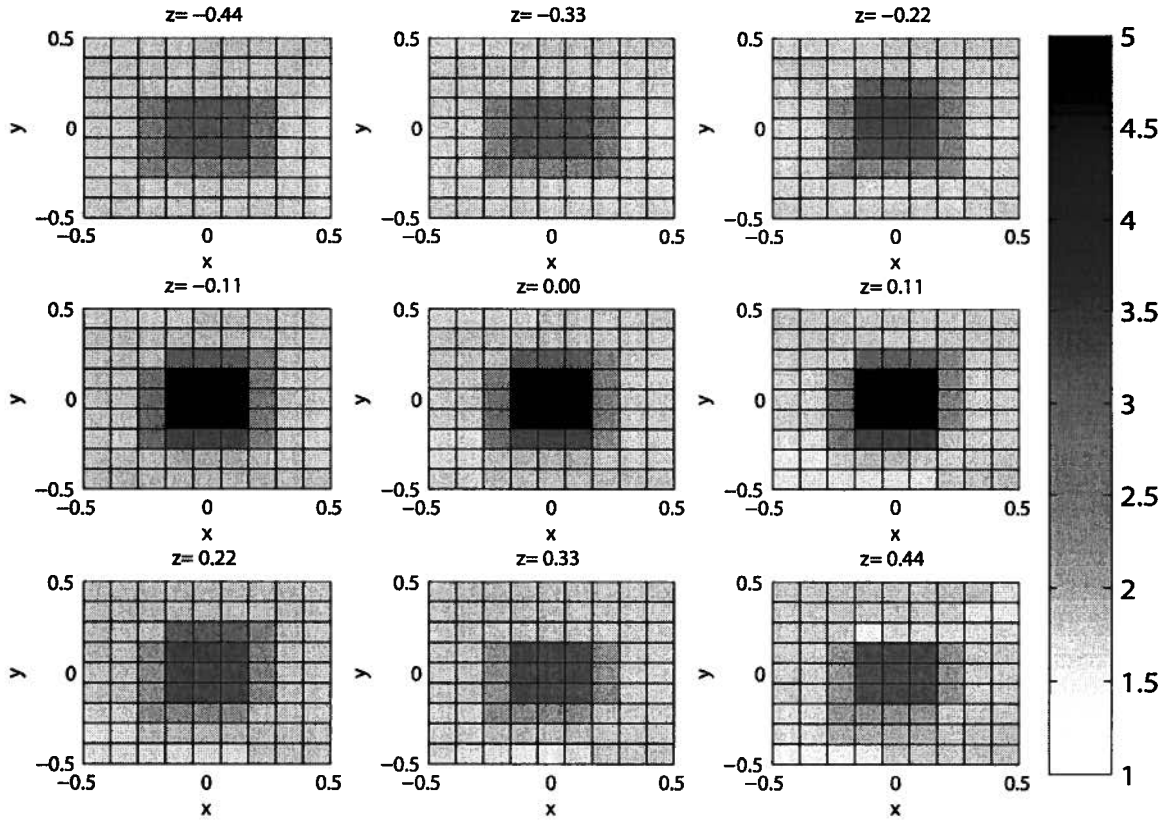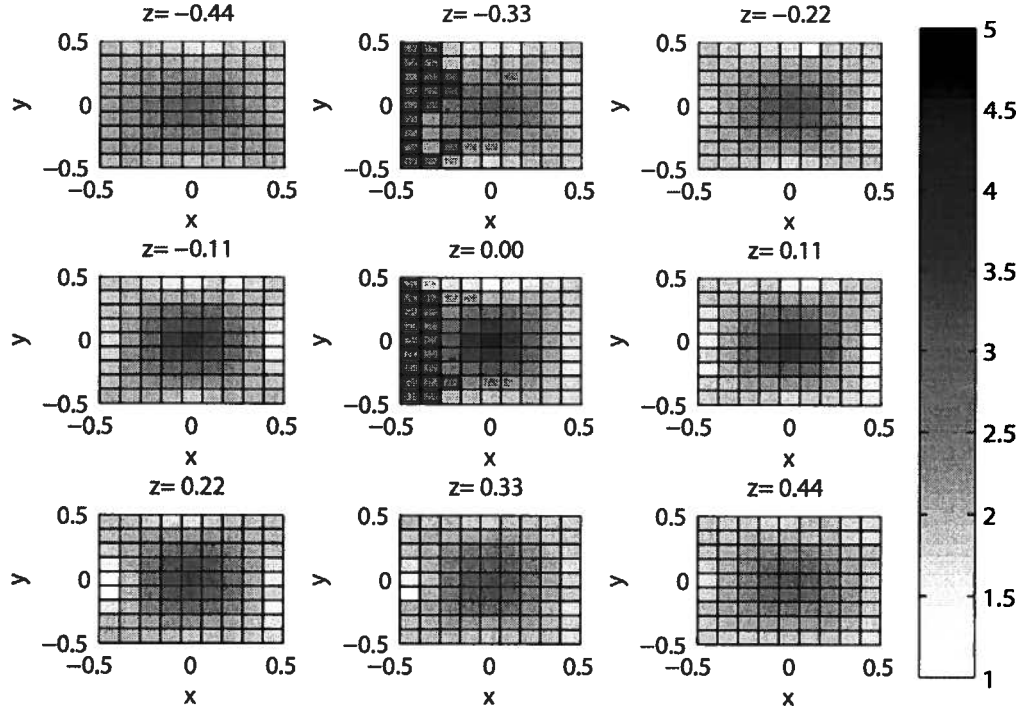
Figure 5.14: Recovered relative permittivity distribution at iteration 300 for the recovery of the paraboloid anomaly case with the CRL2 using 4 illumination geometries and 1 illumination frequency.

Figure 5.14 shows the recovered relative permittivity distribution at iteration 300 for the recovery of the paraboloid anomaly case with the CRL2 using 4 illumination geometries and 1 illumination frequency. There are some similarities between recovered relative permittivity distribution, figures 5.13, and the target relative permittivity distribution, figure 5.9. However they do not look exactly the same. Both share a central anomaly at the point x=0, y=0, z=0. The anomaly in figure 5.9 is constant along the z direction, where as the anomaly in figure 5.14 is not.

Comparison of the recovered relative permittivity distributions for the CRL1 and CRL2 along the y=0, z=0 line.

Figure 5.15: A comparison of the recovered relative permittivity distribution for the CRL1 and CRL2 methods along the y=0, z=0 line.

Figures 5.15 compares the recovered relative permittivity distribution for the CRL1 and CRL2 method with the reference relative permittivity distribution along the y=0 and z=0 line. The shape of the CRL2 curve resembles that of the shape of the reference curve. The CRL1 has the same maximum and minimum value as the reference curve. The CRL1 curved allows for several large jumps in misfit value, most notably between x = -0.22 and x=-0.11. The relative permittivity value of the CRL1 method is also constant between x=-0.11 and x=0.11. The CRL1 allows for the large jumps in relative permittivity, but requires a homogeneous region to compensate for the large change in relative permittivity.

The change in relative permittivity from 2.0 to 5.0 is more gradual with the CRL2 method, because large changes are penalized more than small changes.

Comparison of the recovered relative permittivity distributions
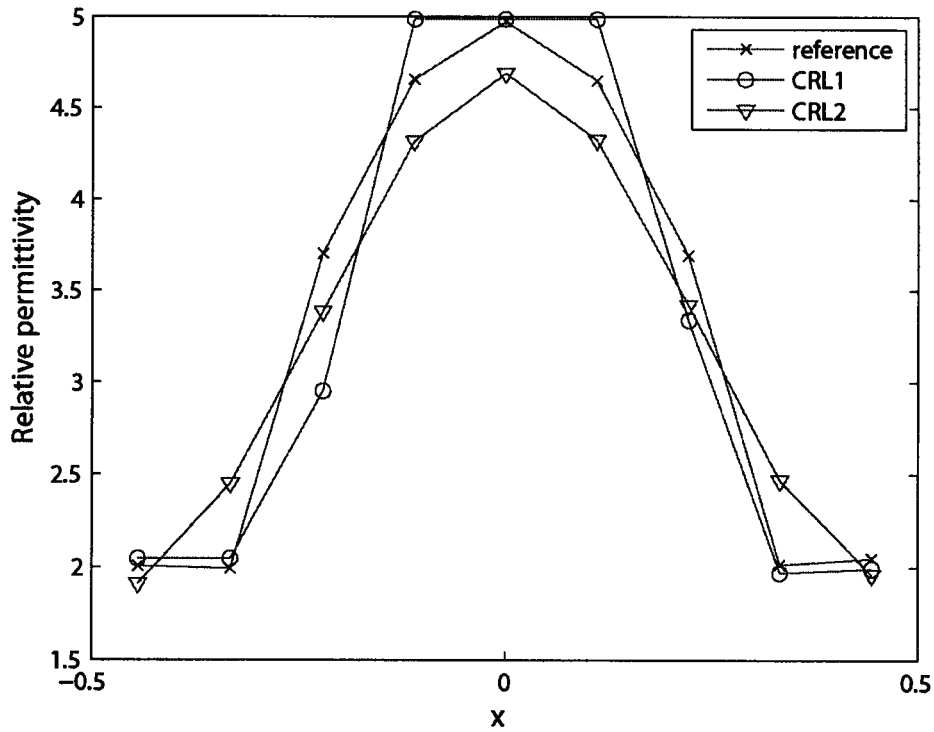for the CRL1 and CRL2 along the x=0, y=0 line.
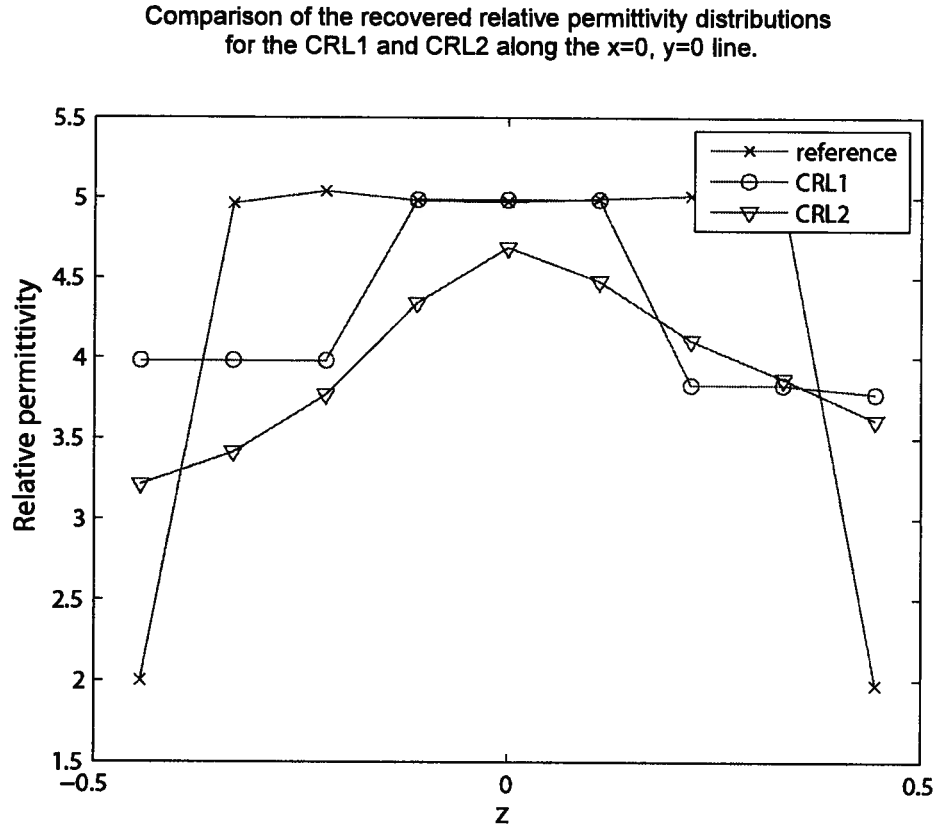
Figure 5.16: A comparison of the recovered relative permittivity distribution for the CRL1 and CRL2 methods along the x=0, y=0 line.

Figures 5.16 compares the recovered relative permittivity distribution for the CRL1 and CRL2

method with the reference relative permittivity distribution along the x=0 and y=0 line. The recovered

relative permittivity by the CRL1 method has 3 flat regions. The region between x=-0.11 and x = 0.11

has the same value as the reference relative permittivity distribution. The shape is similar to that seen

in the figure 4.52 for the case of using the CRL1 method at the 220th iteration, reinforcing the idea that

the CRL1 method has a tendency to find homogeneous regions.

The relative permittivity curve for the CRL2 method lacks sharp transitions. This is due to the

L2 norm. The shape of the recovered relative permittivity resembles a smoothed version of the

reference permittivity.

5.5 Checkered relative permittivity distribution

The checkered pattern case
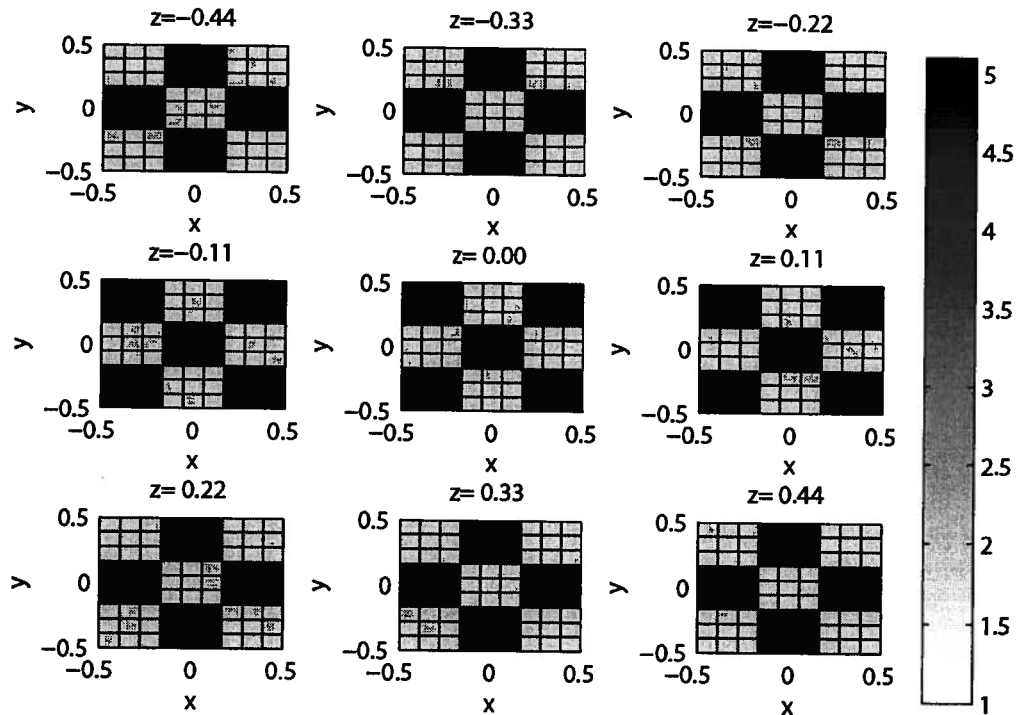
Figure 5.17: Checkered pattern relative permittivity distribution. The pattern consist of alternating cubes of relative permittivity value of 2.0 and relative value of 5.0. The cubes have a sidelength of 1/3 free space wavelength at 1GHz.

In this section the recovery of a checkered patterned relative permittivity distribution is examined. The checkered patterned is used in geophysics to ascertain the resolution of an algorithm. Leveque (1993) advises that the care must be taken when interpreting the results of the checkered pattern case. In some circumstances small sized structure can be recovered well but larger structures may be recovered poorly. This can also be seen from this work. One just has to compare the recovery of the blurry anomaly case, where there is a large blurry anomaly which wasn't recovered well to the small anomaly case in chapter 4, where the target was recovered almost exactly.

The relative permittivity distribution is shown in figure 5.17 and will be known as the checkered pattern case. It is composed of alternating high and low relative permittivity cubes with a

190

sidelength of 1/3 free space wavelength at 1GHz. One set of cubes has a relative permittivity value of 2.0 and the second set of cubes has a relative permittivity value of 5.0.

For the recovery 4 illumination geometries are used and 1 illumination frequency is used. The illumination geometries are I1 to I4 from table 4.1. The illumination frequency is 0.8GHz. The starting model is a homogeneous relative permittivity distribution with a relative permittivity value of 3.0.
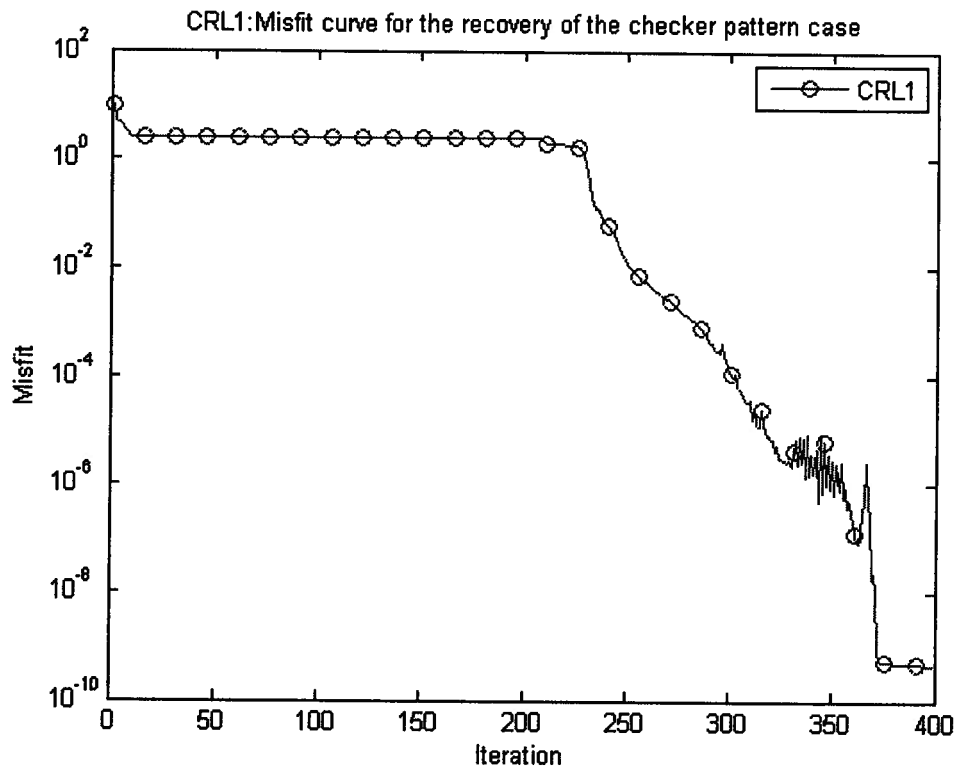
Figure 5.18: Misfit curve for the recovery of the checkered pattern case using 4 illumination geometries and 1 illumination frequency.

Figure 5.18 shows the misfit curve for the recovery of the checkered pattern case with 4 illumination geometries and 1 illumination frequency. The misfit curve exhibits the same behavior as the misfit curve seen in chapter 4. The misfit value initially decreases, followed by a plateau region. From iteration 206 to 330, the misfit value decreases. From iteration 330 to iteration 365, the misfit value oscillates. The final value of the misfit is 4.41e-10.
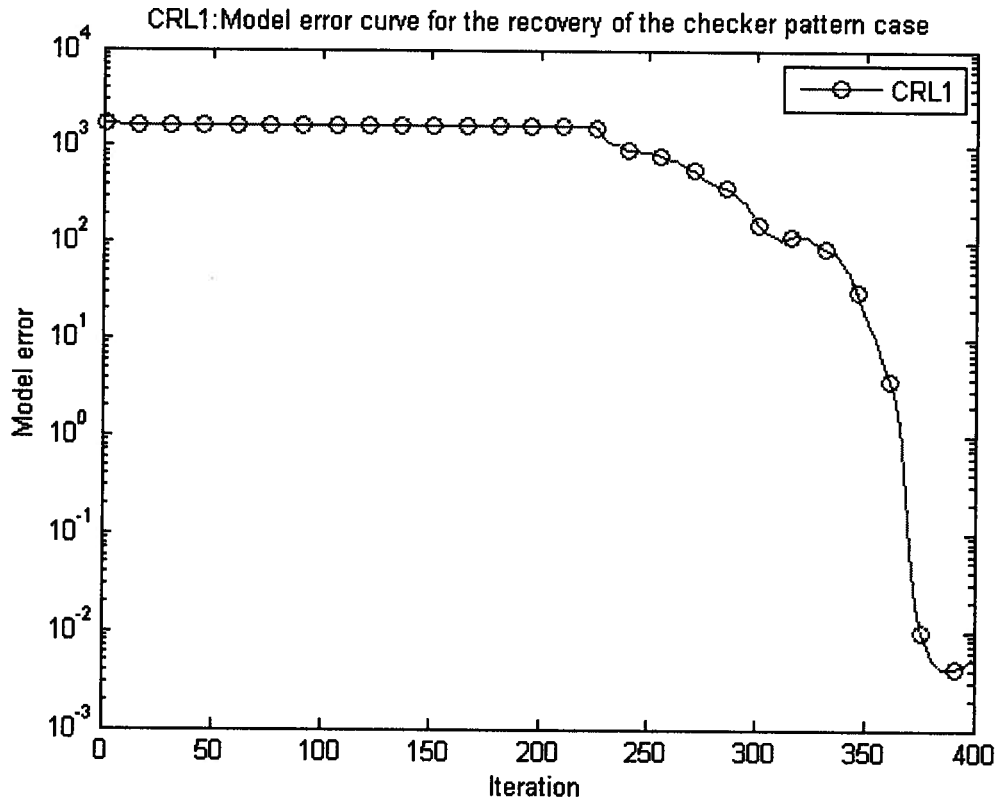
Figure 5.19: Model error curve for the recovery of the checkered pattern case using 4 illumination geometries and 1 illumination angle.

Figure 5.19 shows the model error curve for the recovery of the checkered pattern case with 4 illumination geometries and 1 illumination frequency. The model error curve initially decreases, followed by a plateau region. The model curve, then continues to decrease. The oscillatory behavior seen in the misfit curve in figure 5.18 is not visible in the model error curve. The final model error value is 0.0057, the smallest model error value is 0.0041. The rise in model error from iteration 389 to 400 indicates that the algorithm is in the vicinity of a local minimum. The small value of model error suggests that the target relative permittivity distribution has been successfully recovered.

CRL1: Recovered relative permittivity distribution
at iteration 400 for the checkered pattern case.
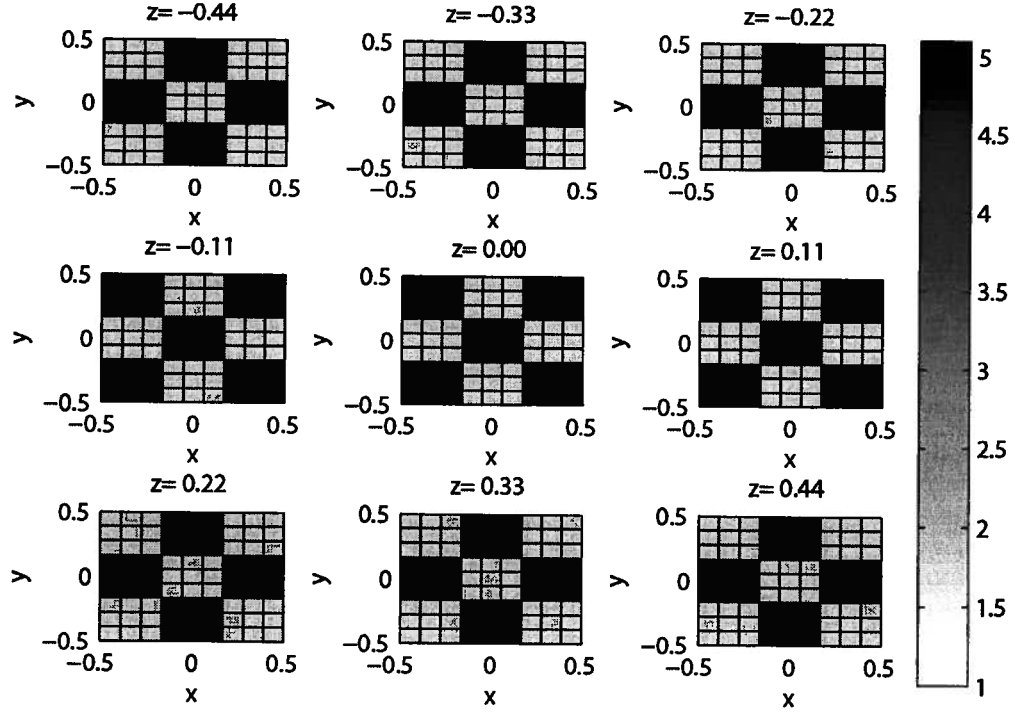
Figure 5.20: Recovered relative permittivity distribution for the CRL1 algorithm using 4 illumination geometries for the checkered pattern case.

Figures 5.20 shows the recovered relative permittivity distribution for the checkered pattern case using the CRL1 method. The recovered relative permittivity distribution seen in figure 5.20 resembles the target relative permittivity distribution see in figure 5.17. From the similarity between the two relative permittivity distribution and the small model error value, the CRL1 algorithm has successfully recovered the target distribution.

Next, the degradation of the recovered relative permittivity distribution is examined when noise is added to the data is examined. A new signal is defined as,

$$x = d + n \quad , \tag{5.2}$$

which is composed of the noiseless data, d, and a noise vector, n.

The signal to noise ratio (Haykin, 2000) is defined as,

$$SNR = \frac{P_{signal}}{P_{noise}} \qquad (5.3)$$

where P_{noise} is the noise power, and P_{power} is the signal power. For the noiseless case the SNR is infinite. The SNR is also commonly written in the decibel scale as

$$SNR_{dB} = 10 \log_{10}\left(\frac{P_{signal}}{P_{noise}}\right) . \qquad (5.4)$$

For the following experiment, Gaussian noise is used. For Gaussian noise

$$P_{Noise} = E\left(\sum_{l=0}^{L} n_l^2\right) = L\sigma^2 \qquad (5.5)$$

the expected noise energy is the number of data points multiplied by the variance of the noise. In (5.5), L represents the total number of data points, E is the expectation operator, n_l is the l'th data point associated with the noise and σ^2 is the variance of the Gaussian noise. The expected noise energy is also the expected misfit value. If the algorithm fall belows this value, it is fitting noise resulting in spurious structure entering the recovered relative permittivity distribution.

The signal to noise ratios in table 5.1 will be examined.

Table 5.1: Signal to noise ratios and their expected misfit values					
Label	Noise power as percent of signal power	SNR_{dB}	Expected misfit value	σ
N1	1.00%	20dB	2.00	0.0577
N1	0.10%	30dB	0.20	0.0183
N3	0.01%	40dB	0.02	0.0058
N4	0.00%	∞	0	0

Three cases with noise will be examined ranging from 1% noise to 0.01% noise. The noiseless case is the case presented above.

Figure 5.21: Misfit curves for the recovery of the checkered pattern case using the CRL1 algorithm with 4 illumination geometries and up to 1% noise.

Figure 5.21 shows the misfit curves for the recovery of the checkered pattern case using the CRL1 algorithm with 4 illumination geometries and up to 1% noise. The cases with 0.1% and 0.01% was able to reach a misfit value below that of the misfit value defined by the noise. The case of 1% noise, did not fall below the expected value. All the cases with noise exhibit the same behavior. There is a initial drop in misfit value, followed by a plateau region. After the plateau region, the misfit decreases again. Lastly, the misfit curve exhibit an oscillatory behavior.

CRL1: Model error curves for the recovery of the checker pattern case with upto 1% noise
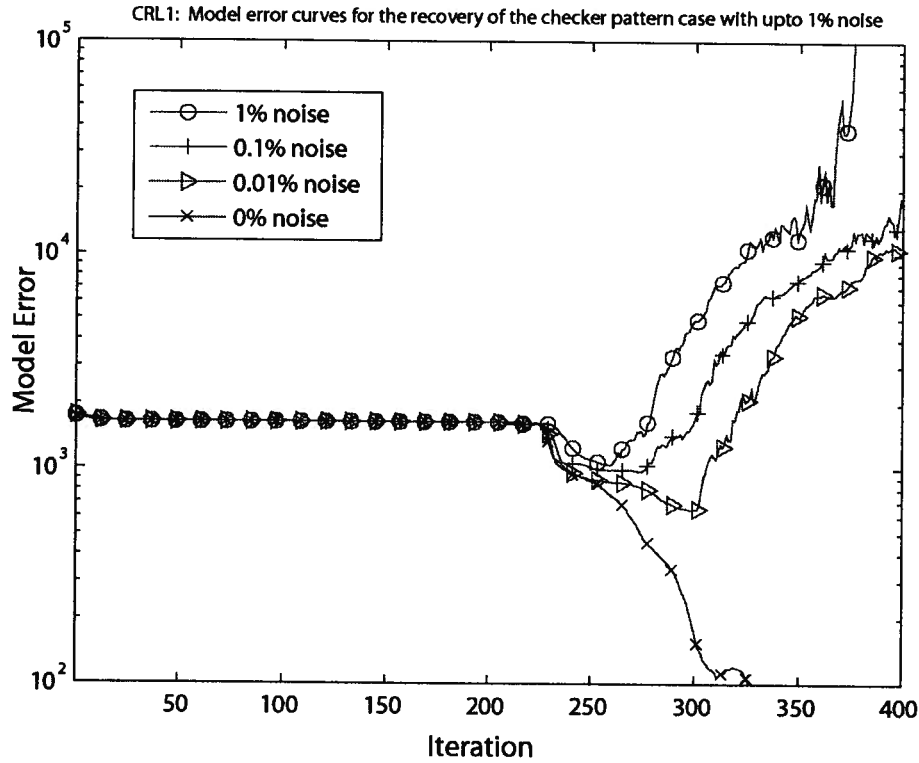
Figure 5.22: Model error curves for the recovery of the checkered pattern case using the CRL1 algorithm with 4 illumination geometries and up to 1% noise.
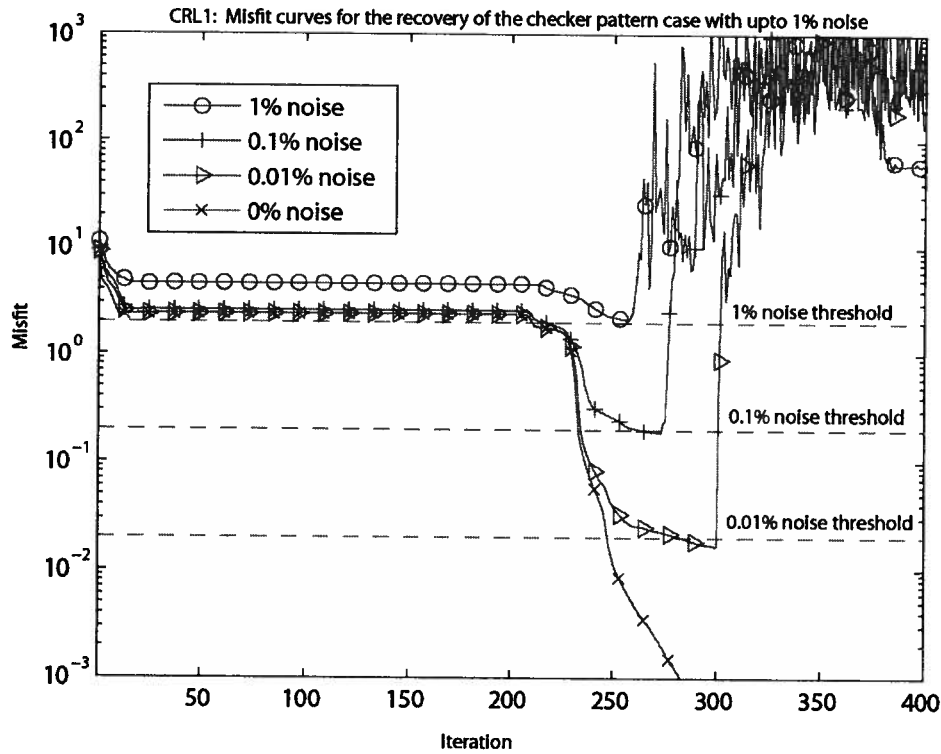
Figure 5.22 shows the model error curves for the recovery of the checkered pattern case using the CRL1 algorithm with 4 illumination geometries and up to 1% noise. All the curves exhibit an initial decrease in model error, followed by a plateau region. After the plateau region the model error decreases. For the curves where noise has been added, the model error curves eventually increases. This happens because the regularization is unable to mitigate the ill-conditioning of the Jacobian matrix.

Table 5.2 summarizes the results in Figure 5.22 and 5.21. The model error in table 5.2 is taken at the iteration before the misfit curve crosses the expected misfit value.

196

Table 5.2: Summary of results for the recovery of a checkered pattern with noise added to the data			
Noise Level	Model Error	Expected misfit value	Minimum misfit value
1% noise	1054	2.00	2.1
0.1% noise	986	0.20	0.19
0.01% noise	731	0.02	0.017
0% noise	5.70e-003	0	4.4e-10

The model error and misfit values in table 5.2 are ordered according to the amount of noise. That is, the less noise in the signal, the better the algorithm performs. The high value of model error, suggests that the recovery of structure at the current length scale is highly sensitive to noise.

The expected value of the noise signal provides a good estimate to the minimum value for the CRL1 method. The minimum value for all the cases is close to the expected misfit value. In the cases of 0.1% and 0.01% noise, the algorithm reached a minimum below that of the expected value and in the case of 1% noise the algorithm did not reach the expected value, but came very close.

CRL1: Recovered relative permittivity distribution for
the checkered pattern case with 1% noise at iteration 257

Figure 5.23: The recovered relativity permittivity distribution for the checkered pattern case with the CRL1 algorithm using 4 illumination geometries with 1% noise at iteration 257.

Figure 5.23 shows the recovered relativity permittivity distribution for the checkered pattern case with the CRL1 algorithm using 4 illumination geometries with 1% noise at iteration 257. There are similarities between the recovered relative permittivity distribution, figure 5.23 and the target relative permittivity distribution, figure 5.17. In the layers from $z=-0.44$ to $z=-0.22$ and $z=0.22$ to $z=0.44$ a cross shape is clearly visible in both the recovered relative permittivity distribution and the target relative distribution. However, the low permittivity square is missing from the recovered relative permittivity distribution. In the layers from $z=-0.11$ to $z=0.11$, the recovered relative permittivity distribution does not resemble the target relative permittivity distribution.

CRL1: Recovered relative permittivity distribution for
the checkered pattern case with 0.1% noise at iteration 263

Figure 5.24: The recovered relativity permittivity distribution for the checkered pattern case with the CRL1 algorithm using 4 illumination geometries with 0.1% noise at iteration 263.

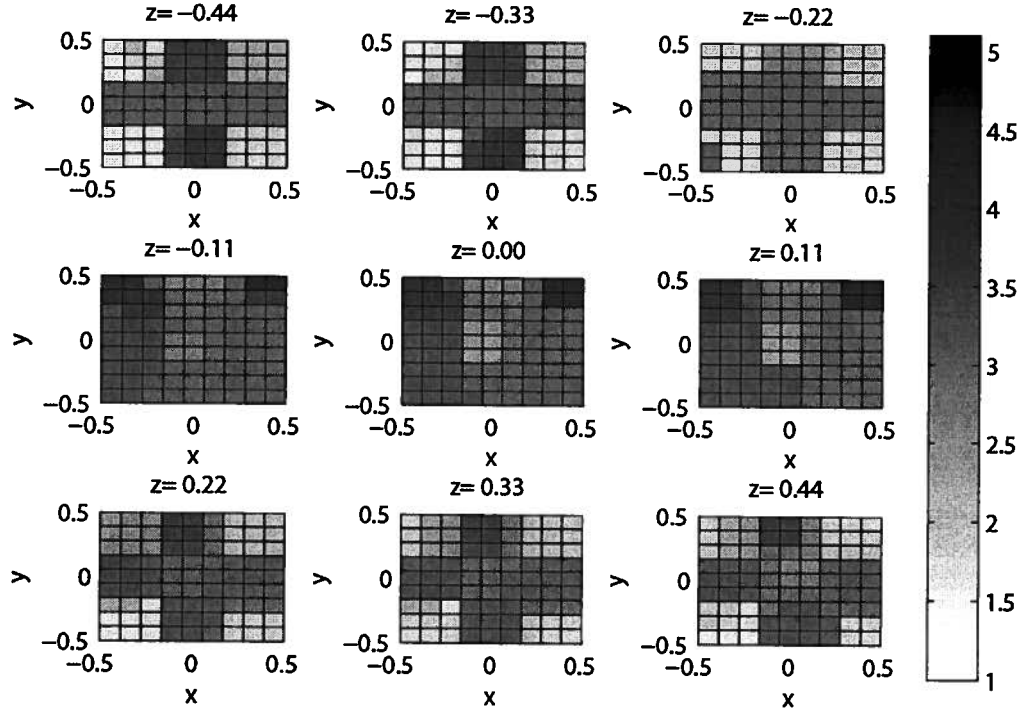Figure 5.24 shows the recovered relativity permittivity distribution for the checkered pattern case with the CRL1 algorithm using 4 illumination geometries with 0.1% noise at iteration 263. There are similarities between the recovered relative permittivity distribution, figure 5.24 and the target relative permittivity distribution, figure 5.17. In the layers from z=-0.44 to z=-0.22 and z=0.22 to z=0.44 a cross shape is clearly visible in both the recovered relative permittivity distribution and the target relative distribution. However, the low permittivity square is missing from the recovered relative permittivity distribution. In the layers from z=-0.11 to z=0.11 the structure at the perimeter of the target relative permittivity distribution is beginning to emerge at the perimeter of the recovered relative permittivity distribution.

CRL1: Recovered relative permittivity distribution for
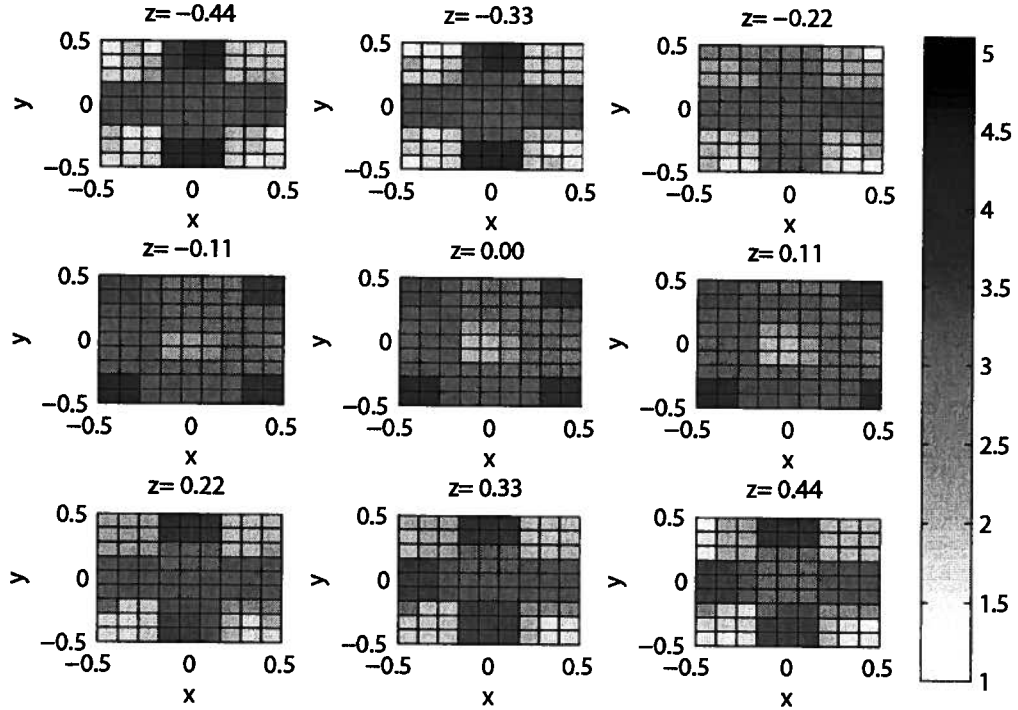the checkered pattern case with 0.01% noise at iteration 283
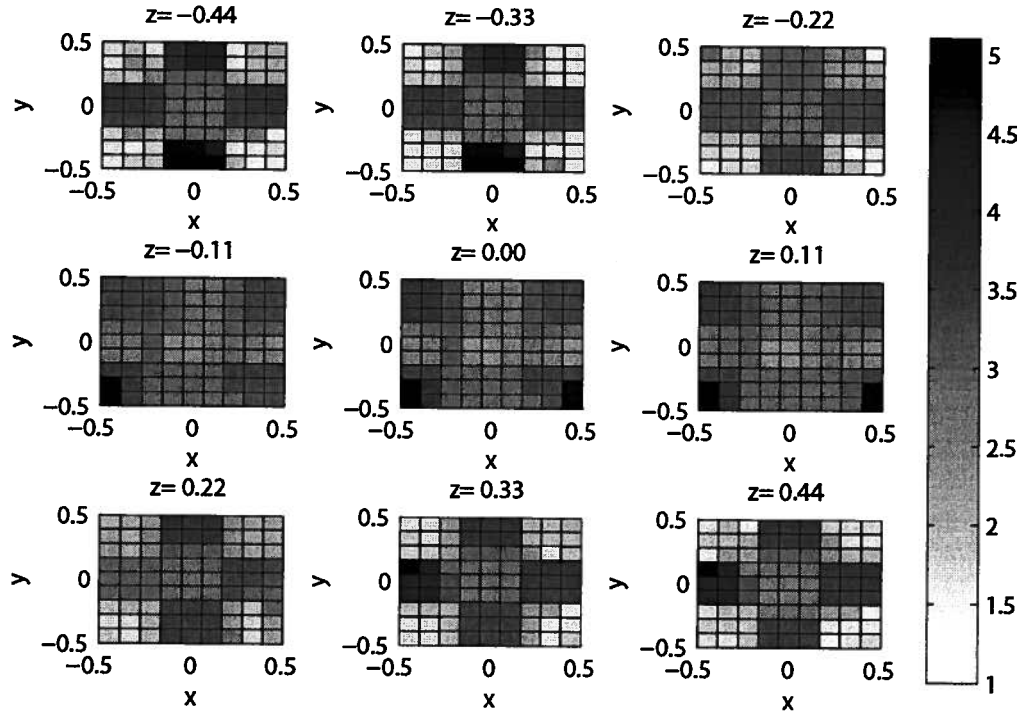
Figure 5.25: The recovered relativity permittivity distribution for the checkered pattern case with the CRL1 algorithm using 4 illumination geometries with 0.01% noise at iteration 283.

Figure 5.25 shows the recovered relativity permittivity distribution for the checkered pattern case with the CRL1 algorithm using 4 illumination geometries with 0.01% noise at iteration 283. There are similarities between the recovered relative permittivity distribution, figure 5.25 and the target relative permittivity distribution, figure 5.17. In the layers from z=-0.44 to z=-0.22 and z=0.22 to z=0.44 a cross shape is clearly visible in both the recovered relative permittivity distribution and the target relative distribution. The low permittivity cube that is visible in the layers of the target distribution is beginning to emerge in the recovered relative permittivity distribution. In the layers from z=-0.11 to z=0.11 the structure at the perimeter of the target relative permittivity distribution is becoming visible.

CRL1:*Comparison of recovered relative permittivity distributions on the y=-0.44 and z=-0.44 line*

Figure 5.26: Comparison of recovered relative permittivity distributions using the CRL1 method on the y=-0.44 and z=-0.44 line.

Figure 5.26 compares the recovered relative permittivity distribution for the recovery of the checkered pattern case, on the y=-0.44 and z=-0.44 line. This line lies on the surface of the cube. On this line the recovered distribution resembles of the target distribution. As more noise is added, the performance of the recovery decreases. The case of 0% noise matches the target exactly. The case of 0.01% noise is very close target permittivity distribution. The error in the peek height from x=-0.11 to x=0.11 decreases with decreasing noise.

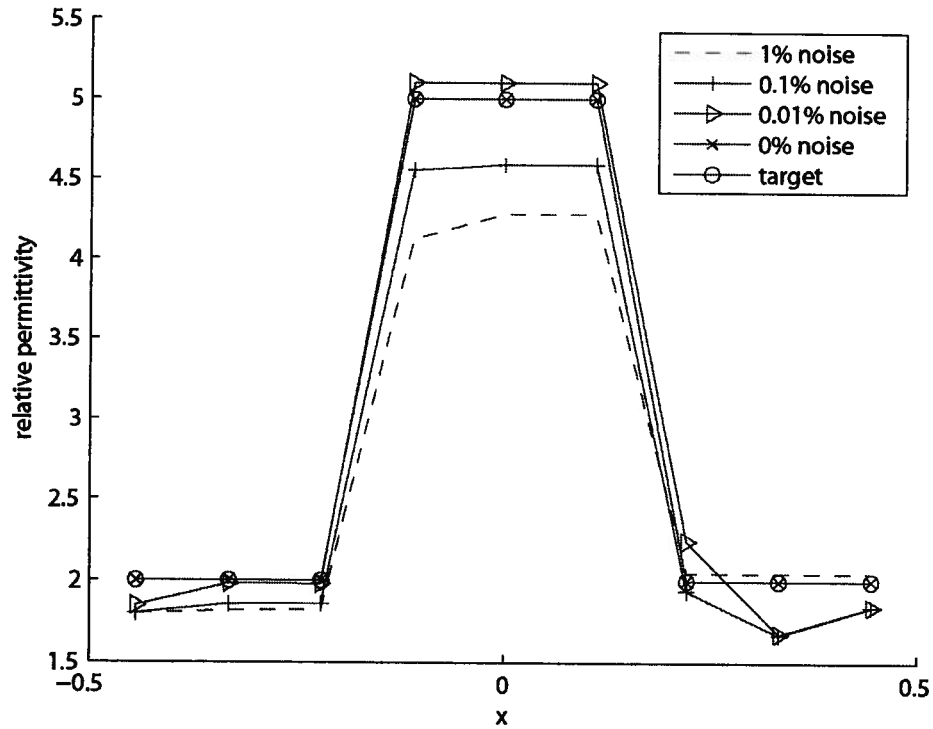CRL1:Comparison of recovered relative permittivity distributions on the y=0 and z=0 line
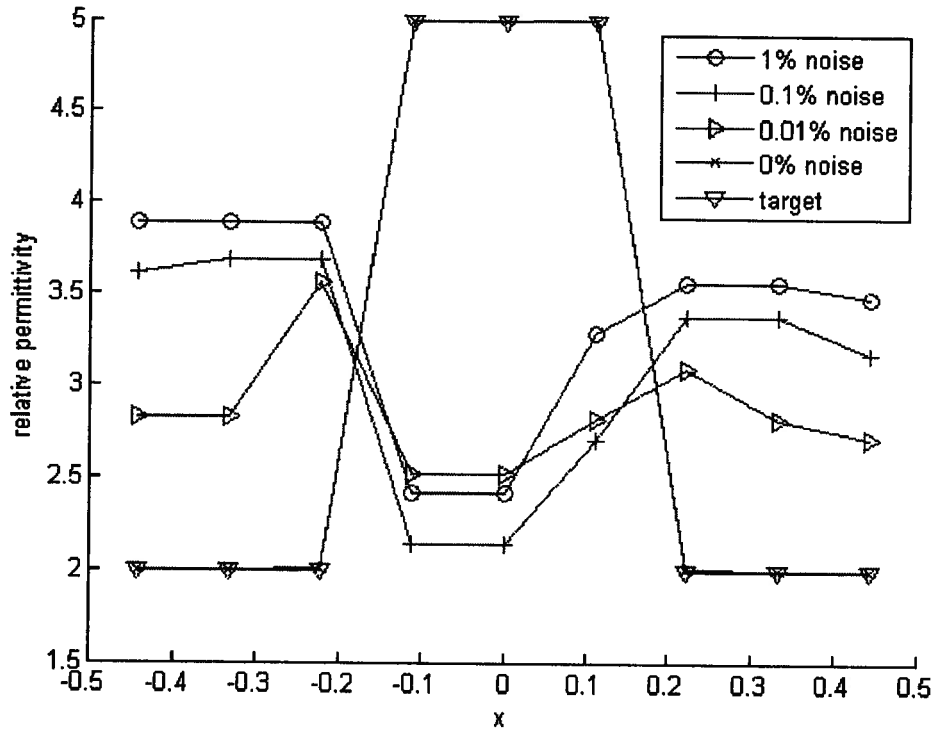
Figure 5.27: Comparison of recovered relative permittivity distributions using the CRL1 method on the y=0 and z=0 line.

Figure 5.27 compares the recovered relative permittivity distribution for the recovery of the

checkered pattern case, on the y=0 and z=0 line. In this case, the recovered relative permittivity curves

for the cases with noise looks different than the target relative permittivity curve. The relative

permittivity curve for the case with no noise matches the target relative permittivity curve exactly.

Clearly the CRL1 method has difficulty recovering the structure at the center of the target

distribution. In this case the inability of CRL1 to recover the target distribution is due to physics, as

opposed to mathematical causes like that of the random anomaly cause. This is due the illumination

wave having difficulty penetrating center dielectric object. When the illumination wave has difficulty

penetrating the center of the dielectric object, the center of the object will have little impact on the data,

manifesting itself as small features in the data. When noise is added, these features are effectively

masked by the noise, leading to the algorithms inability to recover that portion of the structure exactly.

The aforementioned problem can be addressed by altering the illumination. One the solution is to add more illumination geometries. For the above example, only illumination from 4 sides of the cube are used. This can be increased to 6. This approach may still suffer from signal to noise issues. A second solution is to change the shape of the illumination field. Up till now, only planewaves have been used. A more focused beam can be used to guarantee that a good quality signal will encode information from the center of the cube. Ultimately, the goal is to increase the signal strength from the dielectric components at the center of the anomaly.

5.6 Conclusion

In this chapter the CRL1 was used to recover 3 examples to demonstrate the limits of the CRL1 regularization.

The first example is the case with 3 anomalies. The CRL1 method was able to recover the 3 anomalies with very little error. The target relative permittivity contained homogeneous areas, so the CRL1 had no difficulty recovering the target relative permittivity distribution.

The second example is the case where the target relative permittivity distribution was a random variable. The CRL1 method was not able to recover this target relative permittivity distribution because it was made up almost exclusively of sharp transitions, which the regularization penalizes. The starting model in this case has the smallest model error and as the algorithm progressed the model error increased.

The third example is the case with a paraboloid anomaly. Neither the CRL1 nor the CRL2 method was able to recover the target relative permittivity distribution. However both algorithms were able to recover a relative permittivity distribution that resembled that of the target.

From the results of chapter 4 and chapter 5, the CRL1 method can be said to have good performance in cases where there are homogeneous regions. Lacking these regions, the CRL1 method may have difficulty.

The fourth example explored the recovery using the CRL1 method when there is noise in the data. For recovery of this length scale, that is a feature size of approximate 1/3 free space wavelength at 1GHz, the recovery is very sensitive to noise. In particular, the algorithm has has trouble recovering the permittivity at the center of the cube. This limitation was caused by the physics of the problem, rather than the properties of the misfit surface.

Chapter 6

Summary and conclusions

In this thesis algorithms for recovering the permittivity structure of an object using electromagnetic scattering data and unconstrained non-linear optimization were presented. The development of the algorithm was done in 4 steps. A time harmonic electromagnetic wave modeler was first presented. Next, optimization algorithms for the unconstrained non-linear optimization problem were presented followed by a systematic comparison of different algorithms . Lastly, the limitation of the roughness regularization was presented.

There are a number of novel contributions in this thesis. In chapter 2, a forward modeler tailored towards microwave imaging was presented. The Jacobian calculation, in addition to the modeler was included. In chapter 3, the calculation for the gradient without the explicit calculation of Jacobian for scattered data was presented. The cooled roughness method using the L1 norm was also presented, a completely novel microwave imaging algorithm. In chapter 5, some limitations of the cooled roughness regularization was presented. This thesis has added several novel ideas to the field of microwave imaging.

The problem of microwave imaging is cast as an unconstrained non-linear optimization problem. The algorithm attempts to fit scattering data iteratively by altering a relative permittivity distribution after each iteration. For this problem, the misfit surface dominates the performance of the algorithms. The gradient method and the Levenberg-Marquardt method performed poorly when applied to the problem because they used only local information to find a solution. The misfit surface is rich in local minima. To overcome the local minima problem, the cooled roughness regularization

was introduced. The cooled roughness regularization adds global information to prevent the local search algorithms from being trapped by local minima. The recovered permittivity distributions by the L2 cooled roughness method were blurry. The L1 cooled roughness method was shown to perform well when recovering blocky structure. The blocky structured is preferred for the L1 norm. When given random permittivity distribution, a blocky structure was still recovered.

Although CRL1 has the best performance in the examples of chapter 4, it has limitations. When the target structure is not blocky, the CRL1 method will have difficulty recovering the permittivity structure. When noise is added, the noise will mask portions of the signal making it difficult for the CRL1 method to recover the structure. The algorithm was able to recover a checkered pattern with cubes of 1/3 wavelength, in the absence of noise. When noise was added, parts of the pattern were not recoverable, because their contribution to the data was masked by the noise.

Future work in this area would require the acceleration of the algorithms. The computation time limits the extent the problem can be explored. By reducing the time needed to solve the problem, larger problems can be explored. For this thesis, objects of approximately 1 wavelength in size were examined. When sufficient computational resources are available, global search techniques should be explored. Global search techniques requires more computational effort, but they are less sensitive to the local minima behavior present in the misfit surface. From the examination of the noise, it is clear that the different illumination patterns may be needed to generate data that encode every aspect of the relative permittivity distribution with a strong signal.

References

Bulyshev, A., S.Y. Semenov, A.E. Souvorov, R.H. Svenson, A.G. Nazarov, Y.E. Sizov, and G.P. Tatsis. "Computational Modeling of Three-Dimensional Microwave Tomography of Breast Cancer." IEEE Transactions on Biomedical Engineering 48, no. 9 (September 2001):1053-1056.

Chew, W. C. and Q. H. Liu. "Inversion of induction tool measurements using the distorted Born iterative method and CG-FFT." IEEE Transactions on Geoscience and Remote Sensing 32, no.4 (July 1994):878–884.

Chew, W. C. and J. H. Lin. " A Frequency-Hopping Approach for microwave imaging of Large Inhomogeneous Bodies." IEEE Microwave and Guided Wave Letters 5, no. 12 (December 1995):439-441.

Farquarson, C. G. and D. W. Oldenburg "Nonlinear inversion using general measures of data misfit and model structure." Geophysical Journal International 134, no. 1 (July 1998): 213-227.

Fear, E.C., X. Li, S.C. Hagness and M.A. Stuchly."Confocal microwave imaging for breast cancer detection: localization of tumors in three dimensions." IEEE transactions on Biomedical Engineering 49, no. 8 (August 2002):812-822.

Franchois, A., A. Joisel, C. Pichot, and J. Bolomey. "Quantitative microwave imaging with a 2.45-GHz planar microwave camera." IEEE transactions on Medical Imaging 17, no. 4 (August 1998):550-561.

Geffrin, J., P. Chaumet, C. Eyraud, K. Belkebir, and P. Sabouroux. "Three dimensional permittivity reconstructions from free space experimental data." International conference on Electromagnetic in Advanced Applications, 2007, (September 2007):149 – 152, .

Gill, Philip E., Walter Murray and Margaret H. Wright. *Practical Optimization.* London, UK: Elsevier Academic Press, 1986.

Griffiths, David J. *Introduction to Electrodynamics second edition.* Upper Saddle River, NJ: Prentice Hall, 1989.

Harrington, Roger F. *Field Computation by Moment Methods.* Malabar,FL: R.E. Krieger, 1968.

Harrington, Roger F. *Time-Harmonic Electromagnetic Fields.* New York, NY: John Wiley and Son, 1961.

Haykin, Simon. *Communication Systems 4th edition.* New York, NY: John Wiley and Sons, 2001.

Jackson, D. D. "The use of a priori data to resolve non-uniqueness in linear inversion", Geophysical Journal International 57, no.1 (1979):137-157.

Joachimowicz, N., C. Pichot, and J.P. Hugonin. "Inverse Scattering: An Iterative Numerical Method for Electromagnetic Imaging." IEEE transactions on Antenna and Propagation 39, no. 12(December 1991): 1742-1752.

Jin, Jianming. *The finite Element Method in Electromagnetics.* New York, NY: John Wiley & Son. 2002.

Lam, K. and M. Yedlin. "Monochromatic nonuniqueness in plane-wave inverse scattering." IEEE Antennas and Propagation Society International Symposium 2A, (July 2005):139-142.

Levenberg, K. "A Method for the Solution of Certain Non-Linear Problems in Least Squares", The Quarterly of Applied Mathematics 2, (1944):164-168, 1944.

Leveque, J. J., L. Rivera, and G. Wittlinger. "On the use of the checker-board test to assess the resolution of tomographic inversions", Geophysical Journal International 115, no. 1 (February 1993):313-318.

Li, Fenghua, Qing Huo Liu, and Lin-Ping Song. "Three-Dimensional Reconstruction of Objects Buried in Layered Media Using Born and Distorted Born Iterative Methods." IEEE Geoscience and Remote Sensing Letters 1, no.2 (April 2004):107-111.

Lipschutz, Schultz. *Linear Algebra Second Edition.* New York, NY: McGraw and Hill 1991.

Lobel, P., L. Blanc-Feraud, C. Pichot and M. Barland. "Technical Development for an Edge-Preserving Regularization Method in Inverse Scattering." Research Note 95-73, Laboratoire Informatique Signaux et Systemes de Sophia Antipolis, 1996.

Marquardt, Donald. "An Algorithm for Least-Squares Estimation of Nonlinear Parameters." Journal of the society of Industrial Applied Mathematics 11, no. 2 (June 1963):431-441.

Pratt, R., C. Shin and G. Hicks. "Gauss-Newton and full Newton Methods in frequency-space Seismic waveform Inversion." Geophysics Journal International 133,no. 2 (1998):341-362.

Schaubert, D.H., D.R. Wilton and A. W. Glisson. "A Tetrahedral Modeling Method for Electromagnetic Scattering by Arbitrarily Shaped Inhomogeneous Dielectric Bodies." IEEE transactions on Antennas and Propagation 32, no. 1 (January 1984):77-85.

Souvorov, A., E. Bulyshev, S. Y. Semenov, et al. "Microwave Tomography: A Two Dimensional Newton Itertive Scheme." IEEE transactions on Microwave Theory and techniques 46, no. 11 (Novermber 1998):1654-1659.

Tanaka, M. and Ogata, K. "Fast Inversion Method for Electromagnetic Imaging of Cylindrical Dielectric Objects with Optimal Regularization Parameter" IEEE transactions on Communications 84, no.9 (2001): 2560-2565.

Usner, B.C., K. Sertel, M.A. Carr and J.L. Volakis. "Generalized volume-surface integral equation for modeling inhomogeneities within high contrast composite structures." IEEE transactions on Antennas and Propagation 54, no. 1 (January 2006):68-75.

Volakis, John L., Arindam Chatterjee and Leo C. Kempel. *Finite Element Method for Electromagnetics*. New York, NY: IEEE Press. 1998.

Wang, J.H., *Generalized Moment Methods in Electromagnetics*. New York, NY: John Wiley and Son, 1991.

Yu, C., M. Yaun, J. Stang, E. Bresslour, R.T. George et al. "Active Microwave Imaging II: 3-D System Prototype and Image Reconstruction From experimental Data." IEEE Transaction on Microwave Theory and Techniques 56, no. 4(April 2008): pp. 991-1000.

Zaeytijd, J.D., A. Franchois, C. Eyraud and J. Geffrin. "Full-Wave Three-Dimensional Microwave Imaging with a Regularized Gauss-Newton Method: Theory and Experiment." IEEE Transaction on Antennas and Propagation 55, no. 11 (November 2007):3279-3292.

Zhang, Z. Q. and Q. H. Liu. "Three-Dimensional Nonlinear Image Reconstruction for Microwave Biomedical Imaging." IEEE transactions on Biomedical Imaging 51, no. 3 (March 2004):544-548.