MODELLING AND INVERSION OF
TWO-DIMENSIONAL MAGNETOTELLURIC DATA

AI JUN ZHANG

Doctor of Philosophy

University of Edinburgh

1988
DECLARATION

I hereby declare that the work presented in this thesis is my own, unless otherwise stated in the text and that the thesis has been composed by myself.

Ai Jun Zhang
Two dimensional modelling has become an important research area for the magnetotelluric method of Geophysical exploration since many real cases are two (or even three) dimensional. These problems can only be solved numerically and require a large amount of CPU time so that not much work has been done in this area. My thesis concentrates on two dimensional modelling, mainly dealing with inversion, which contributes to this research area and could be extended to further study in three dimensional modelling.

A two dimensional inversion modelling study requires the efficient supply of Jacobian information from the forward program. For this purpose a two dimensional forward program has been modified by an iterative scheme, which can approximate the Jacobian information whilst trading-off the speed and accuracy. The convergence of this iterative scheme is fully analysed and discussed. The two dimensional forward program used here as a basis is the one made by Brewitt-Taylor and Weaver (1976).

In this thesis, two dimensional inversion modelling is made using the ridge regression method. This method's characteristics and applications are described fully. For the resolution study of the resulting model, some modifications have been made to Jackson's most square method in determining the resolution of the model parameters.

An analysis is made concerning one of the possible phenomena in two dimensional inversion modelling, namely the δ-like model formulation, arising especially when the model area has been parameterized in much detail. To solve this problem and to construct an acceptable, simple model a smoothing technique has been introduced. This smooths the model subjectively whilst
the misfit is hardly affected.

Introduction of the boundary parameterization enables two dimensional inversion modelling to be conducted in a more practical and efficient way. For the newly introduced parameterization, the parameters are in different spaces i.e. the resistivity and the horizontal boundary parameters are in logarithmic space while the vertical boundary has to be in linear space. The solution of the inversion will therefore be affected when using the ridge regression method. To overcome this problem a scaling matrix has been introduced. The function of the scaling matrix acting on the solution has been analysed in detail.

As a synthetic example of 2D inversion, one model from the COMMEMI project (labelled there as 3-2a) has been chosen for study. The data are given at 8 sites for 15 periods and are artificially contaminated by random noise to simulate the field case. During the iterations of the inversion, the misfit reduced to a local minimum. One possible way to escape from this local minimum and to approach the global minimum is to parameterize the model in more detail. When a model has been parameterized in more detail the result of the inversion often leads to an unnecessarily complex model. At this stage the model may be smoothed for simplification while the misfit is hardly affected. After 10 iterations of the above scheme, the inversion shows a satisfactory result.

Finally a field case has been studied by using the above methods. For this case, the fast two dimensional forward modelling program demonstrates its advantage in conducting 15 iterations of inversion with many parameters. After several iterations, the inversion again reached a local minimum. At this stage some of the sites already give a good fit to the data but others do not.
To escape from the local minimum and to enable those poorly fitting sites to be improved, some areas are parameterized in more detail. This does make the model converge, especially for those previously poorly fitting sites. But the resulting model is very complicated and there even exists a kind of $\delta$-model phenomenon. The smoothing technique has to be applied, which once again shows its advantage.

Two dimensional magnetotelluric inversion using the ridge regression method has been shown to be very effective. Along with it, some further techniques are necessary and useful, such as boundary parameterization in addition to resistivity parameterization and also model smoothing.
ACKNOWLEDGEMENTS

I would like to gratefully acknowledge my supervisor Dr. B.A. Hobbs for introducing me to the Magnetoelluric two dimensional modelling study, for his continuing support and guidance through the whole period, and especially for his patience in helping me in English correction for the completion of this thesis.

I also thank Dr. V.R.S. Hutton for her kind help in this field, and Dr. Hipkin for his help in some computer aspects. Special thanks go to G. Dawes for some beneficial discussions in the research area.

I would like to express my thanks to my colleagues: E.R.G. Hill for the guidance of the work in chapter 4; Sergio Luiz Fontes, T. Harinarayana and M. Meju for some helpful discussions; R. Parr for showing me some aspects of field work.

For the financial support to my study, I am indebted to the Chinese Education Department and the British Council for a Technology Co-operation Scholarship and the CVCP (London) for Oversea Research Student Award.

Finally, I would like to thank the friends I have met in Edinburgh for their warm companionship. My deep thanks for the encouragement, love and help of my family and Y. Kuang.
## TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>DECLARATION</td>
<td>1</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>2</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>5</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>6</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>9</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>12</td>
</tr>
<tr>
<td>1 Introduction</td>
<td>13</td>
</tr>
<tr>
<td>1.1 The source field</td>
<td>13</td>
</tr>
<tr>
<td>1.2 General review and brief description of the research work in this thesis</td>
<td>15</td>
</tr>
<tr>
<td>2 Theory of Magnetotelluric Sounding and field work</td>
<td>22</td>
</tr>
<tr>
<td>2.1 The fundamental theory of Magnetotellurics</td>
<td>22</td>
</tr>
<tr>
<td>2.2 Induction in one dimensional structures</td>
<td>23</td>
</tr>
<tr>
<td>2.3 Induction in two dimensional and three dimensional structures</td>
<td>24</td>
</tr>
<tr>
<td>2.4 Apparent resistivity, phase and dimensionality indicator</td>
<td>26</td>
</tr>
<tr>
<td>2.5 Field work</td>
<td>28</td>
</tr>
<tr>
<td>3 Modelling Methods</td>
<td>29</td>
</tr>
<tr>
<td>3.1 One dimensional modelling</td>
<td>29</td>
</tr>
<tr>
<td>3.2 Two and three dimensional forward modelling</td>
<td>32</td>
</tr>
<tr>
<td>3.2.1 General description</td>
<td>32</td>
</tr>
<tr>
<td>3.2.2 MT 2D finite difference programme</td>
<td>34</td>
</tr>
<tr>
<td>3.3 Two and three dimensional inversion modelling</td>
<td>36</td>
</tr>
<tr>
<td>3.3.1 General description</td>
<td>36</td>
</tr>
<tr>
<td>3.3.2 The least square method</td>
<td>37</td>
</tr>
<tr>
<td>3.3.3 Ridge regression method</td>
<td>39</td>
</tr>
<tr>
<td>3.4 Analysis of the ridge regression method and its application in MT 2D inversion</td>
<td>40</td>
</tr>
<tr>
<td>3.4.1 The general characteristics</td>
<td>40</td>
</tr>
<tr>
<td>3.4.2 The decomposition of the Jacobian matrix</td>
<td>41</td>
</tr>
<tr>
<td>3.4.3 The analysis of the ill-conditioned solution</td>
<td>43</td>
</tr>
<tr>
<td>3.5 The choice of the damping factor</td>
<td>44</td>
</tr>
<tr>
<td>3.5.1 Computation description</td>
<td>46</td>
</tr>
<tr>
<td>4 An iterative scheme for use in the MT 2D forward program — to speed up the production of the Jacobian information and to improve the accuracy of the inverse matrix</td>
<td>49</td>
</tr>
<tr>
<td>4.1 Introduction</td>
<td>49</td>
</tr>
<tr>
<td>4.2 Using an iterative scheme to approximate the inverse matrix for the Jacobian information</td>
<td>51</td>
</tr>
<tr>
<td>4.3 Case study</td>
<td>54</td>
</tr>
<tr>
<td>4.4 Using the iterative scheme to improve the accuracy of the originally calculated inverse matrix</td>
<td>56</td>
</tr>
<tr>
<td>4.5 Examples and computer procedure</td>
<td>57</td>
</tr>
<tr>
<td>4.6 Conclusion and discussion</td>
<td>63</td>
</tr>
</tbody>
</table>
5 Resolution study

5.1 The resolution formulation given by Jackson 65
5.2 The modified resolution 66
  5.2.1 Extremizing the increments in conjunction with stabilization 66
  5.2.2 Extremizing each parameter increment with others being constrained 68

6 Analysis of the delta-like model formulation and model smoothing 70

6.1 Introduction 70
6.2 An example of 2D inversion and its resultant model 72
6.3 Analysis of the formulation of the delta-like model 81
6.4 Model smoothing 87
6.5 Conclusion and discussion 96

7 Analysis of the effect of boundary parameterization in MT 2D inversion 97

7.1 An alternative parameterization for the two dimensional model inversion 97
7.2 Problems arising 104
7.3 The use of a scaling matrix projecting on the Jacobian matrix to restore the affected solution. 110
  7.3.1 Introduction of the scaling matrix 110
  7.3.2 General application 111
7.4 Analysis of the ridge regression method when the scaling matrix is applied. 112
7.5 Conclusion and discussion 115

8 A model example from the COMMENI project 118

8.1 Model description and data generation 118
8.2 The construction of the initial 2D model and its first iteration 120
8.3 Escaping from the local minimum 130
8.4 The result of further iterations 132
8.5 The resolution study for the model 144
8.6 Conclusion and discussion 145

9 Field data modelling 147

9.1 Introduction 147
9.2 The inversion procedure 152
9.3 The misfit at site L and relevant parameter adjustments 158
9.4 Further inversion procedure and the model smoothing 165
9.5 The resolution study 173
9.6 Geological interpretation and conclusion 175

10 Conclusions and suggestions 177

10.1 Conclusions 177
10.2 Suggestions 178
APPENDIX 1 180
APPENDIX 2 185
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>16</td>
</tr>
<tr>
<td>3.1</td>
<td>45</td>
</tr>
<tr>
<td>3.2</td>
<td>48</td>
</tr>
<tr>
<td>4.1</td>
<td>58</td>
</tr>
<tr>
<td>4.2</td>
<td>62</td>
</tr>
<tr>
<td>6.1</td>
<td>73</td>
</tr>
<tr>
<td>6.2</td>
<td>74</td>
</tr>
<tr>
<td>6.3</td>
<td>76</td>
</tr>
<tr>
<td>6.4</td>
<td>78</td>
</tr>
<tr>
<td>6.5</td>
<td>80</td>
</tr>
<tr>
<td>6.6</td>
<td>82</td>
</tr>
<tr>
<td>6.7</td>
<td>85</td>
</tr>
<tr>
<td>6.8</td>
<td>87</td>
</tr>
<tr>
<td>6.9</td>
<td>91</td>
</tr>
<tr>
<td>6.10</td>
<td>93</td>
</tr>
<tr>
<td>6.11</td>
<td>94</td>
</tr>
<tr>
<td>6.12</td>
<td>95</td>
</tr>
</tbody>
</table>

1.1 A. Amplitudes of natural variations in the horizontal geomagnetic field useful in induction research. B. Corresponding amplitudes in the earth's electric field, computed for a model earth of uniform resistivity $20 \ \Omega$m.

3.1 A ridge plot graph for the increment vector $x$.

3.2 The computation procedure diagram for using the ridge regression method.

4.1 An example of a two-dimensional model used for the iterative scheme studies.

4.2 The computation procedure for the modified two-dimensional forward program.

6.1 A two-dimensional model used for the inversion studies.

6.2 Locations of the 22 sites providing data for inversion relative to the true and initial models.

6.3 The initial model area as in Fig. 6.2 is divided into 28 blocks (dashed lines) for the inversion parameterization.

6.4 The model is parameterized by a parameter vector $p=[p_1, p_2, \ldots, p_n]^T$, where $p_i$ is the $i$th block's resistivity value.

6.5 The resulting models during the iterations.

6.6 The resulting models as different damping factors are chosen.

6.7 The results for the increment vector $x$ when one element $x_{13}$ is extremized.

6.8 As for Fig. 6.7 except that parameter increment $x_{25}$ is extremized.

6.9 As in Fig. 6.5, the iterations being carried out here follow Fig. 6.5 and the smoothing technique is applied.

6.10 As in Fig. 6.5, the iterations being carried out here follow Fig. 6.5 but no smoothing technique is applied.

6.11 As for Fig. 6.5 but the data are contaminated by noise.

6.12 As in Fig. 6.11 and the iteration results carried out here by using the smoothing technique.
7.1 2D inversion modelling example when the boundary parameterization has been applied.

7.2 The sequences of convergence for boundaries only of Fig. 7.1.

7.3 As for Fig. 7.1 except that the data are contaminated by noise.

7.4 The plot of \( y(b) = |10^{b-1}/|b| \) for the range \(-1 \leq b \leq 1\).

8.1 A two dimensional model from the COMMEMI project (labelled there as 3-2a).

8.2 The one dimensional inversion result at each site.

8.3 The initial two dimensional model.

8.4 Comparison of the response of the initial model in Fig. 8.3 to the given response function data.

8.5 The resulting model from the first iteration.

8.6 The resulting model after further iterations following Fig. 8.5.

8.7 The resulting model after four more iterations from the model of Fig. 8.6.

8.8 The model produced by two further iterations from the model of Fig. 8.7.

8.9 The final model.

8.10 Comparison of the response of the final model in Fig. 8.9 to the given response function data.

9.1 Geological sketch map of Sardinia.

9.2 The initial 2D model constructed from 1D inversions at each site.

9.3 Comparison of the response of the initial 2D model (Fig. 9.2) with the field data.

9.4 The resulting model after five iterations of inversion.

9.5 Comparison of the response of the model in Fig. 9.4 with the field data.

9.6 Part of the model in Fig. 9.5, showing only sites J, K and L.

9.7 The resulting resistivity values of the blocks in Fig. 9.6 after another 6 iterations.
9.8 The result after smoothing each parameter within its group with its adjacent parameters (using Fig. 9.7).

9.9 The final model after fifteen iterations of inversion.

9.10 Comparison of the response of the final model with the field data.
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>An example of using the iterative scheme to improve the accuracy of the original inverse matrix.</td>
<td>59</td>
</tr>
<tr>
<td>4.2</td>
<td>An example of using the iterative scheme to approximate the inverse matrix for the Jacobian information.</td>
<td>59</td>
</tr>
<tr>
<td>7.1</td>
<td>The inversion modelling results at some of the stages of the iterations.</td>
<td>99</td>
</tr>
<tr>
<td>7.2</td>
<td>As for Table 7.1 except that the data is contaminated by noise.</td>
<td>103</td>
</tr>
<tr>
<td>8.1</td>
<td>Resolution for the resistivity block study.</td>
<td>145</td>
</tr>
<tr>
<td>9.1</td>
<td>The grids used for the first four iterations.</td>
<td>156</td>
</tr>
<tr>
<td>9.2</td>
<td>The grids used for the rest of the iterations.</td>
<td>158</td>
</tr>
<tr>
<td>9.3</td>
<td>The resolution study for some blocks.</td>
<td>174</td>
</tr>
</tbody>
</table>
CHAPTER 1
INTRODUCTION

Geophysics is the application of principles and practices of physics and geology. It can be divided into seven subdivisions—density, gravity, magnetic, electric phenomena, deformations, heat flow, radioactivity (Howell, 1959)—which cover a region from the ionosphere and magnetosphere down to the deep structure in the earth. It contributes to earth science as a whole involving both theoretical and experimental approaches to an understanding of the planet we inhabit. The magnetotelluric method (abbreviation MT) is one tool in understanding the earth, which investigates the earth's conductivity structure by using the natural electromagnetic field as the source. It has been accepted as an important method of geophysical exploration of the earth's crust and mantle.

1.1. The source field

The earth consists of a solid inner core, a liquid outer core, a solid stony mantle and crust including the conductive ocean, an insulating atmosphere and an enveloping plasma called the magnetosphere. All of these, except the atmosphere, contribute to or influence the geomagnetic field.

The evidence for motion in a liquid core is suggested by the presence of the earth's magnetic field and its long term time variations. The complicated interaction between the earth and particle radiation from the sun is revealed by magnetic storms.

The fields measured on the surface of the earth are vector sums of a number of different constituent fields each of which originates in a different way and varies differently in time and space. The field as a whole varies in a
complicated way in both time and space.

Analyses may be used to separate each of the time varying fields into two parts. One part is due to sources within the earth, the other part to sources outside the surface of the earth (Parkinson, 1983).

It is known that the natural electromagnetic field of the earth arises from a variety of causes. For application of the magnetotelluric method, the fields of interest are described below.

At frequencies above a few Hertz, the field source may be due to man-made fields, such as power distribution systems and radio stations, and to meteorological activity. If the man-made sources are far away from the observation sites then the fields are considered uniform and useful for magnetotelluric studies (if the signals are strong enough). If the man-made sources are located very near our observation sites they can be used only if their distribution is known; otherwise they are considered as noise. The meteorological activity portion is considered rather an important source field, in particular lightning associated with thunderstorms, which provides a surprisingly uniform source for electromagnetic energy (Strangway et al., 1973, Keller and Frischknecht, 1966).

The principal interest in applying the magnetotelluric method is in exploring to great depths. Therefore the electromagnetic field at frequencies below one Hz is of most interest. At these low frequencies, the natural electromagnetic field originates from complex interactions between the earth's permanent magnetic field and the flow of plasma from the sun (Yanovskiy, 1978).

A general geomagnetic spectrum in the interval 1 cycle per year to 1 cycle
per second (Serson, 1973) is shown in Fig. 1.1. The telluric response that would be observed over a uniform earth of resistivity 20 Ωm is also shown in the figure.

The frequency of 1 cycle per year is due to the earth's rotation around the sun; the 27 days period is due to the solar rotation; the daily variation (Sq) of period 24 hours, and its harmonics, arise from ionospheric dynamo action, which is observed best during geomagnetically quiet periods (Mishin, et. al., 1975); geomagnetic bay variations of one hour period are due to the simplest type of worldwide magnetic storms; micropulsations of 1-600 seconds period originate as hydromagnetic waves due to the interaction between the solar wind and the magnetosphere. They are either continuous harmonic pulsations(Pc's) or transient phenomenon(Pi's) usually observed around local midnight (Orr, 1973).

1.2. General review and brief description to the research work in this thesis

It was Cagniard (1953) who first recommended that the resistivity structure beneath the earth's surface is determinable when the ratio and phase of the horizontal components of the electric and magnetic fields at the earth's surface have been measured using natural source fields. Following Cagniard (1953), Price (1962) modified the theory taking account of the dimensions of the inducing source field.

The propagation and attenuation of electromagnetic waves into the earth are best understood by developing the theory of electromagnetic induction, contained in Maxwell equations. Chapter 2 gives the basic description and an introduction to data collection.

The application of natural electromagnetic fields for the magnetotelluric
A. Amplitudes of natural variations in the horizontal geomagnetic field useful in induction research.

B. Corresponding amplitudes in the earth-electric field, computed for a model earth of uniform resistivity 20 $\Omega$m.

(After Serson, 1973)
method theoretically requires that the field is a vertically propagating plane wave. In reality this condition cannot be satisfied. Madden and Nelson (1964) showed numerically that, for nearly all practical purposes, the source fields could be treated as planar. Dmitriev and Berdichevsky (1979) showed that, when conductivity varied only with depth, source fields may vary linearly in the horizontal direction without effecting the result.

The signals in the field are often irregular and noise like in appearance. In order to give repeatable, geologically meaningful response curves, a good quality data set is required. Digital recording and the Fast Fourier Transform (FFT) make it possible to carry out spectral and tensor analyses routine so that good quality data can be obtained (Cantwell, 1960; Swift, 1967). Kao and Rankin (1977) introduced a concept of remote reference in MT to improve the field data's quality.

The next stage is the interpretation of the collected MT data, this is called modelling and inversion. One dimensional cases can be solved analytically, this makes inversion modelling very convenient (Parker, 1970). For general two and three dimensional cases there are no analytic solutions available and numerical solutions have to be used. Jones and Prices (1971) introduced the finite difference method for MT application, Raiche (1974) applied the integral equation method, Swift (1967) applied the transient network method and Reddy and Rankin (1973) applied the finite element method. Vozoff and Jupp (1975) used the ridge regression method to approach MT one and two dimension inversion problems.

The interpretation of the collected data requires modelling, which is a most important stage in the Magnetotelluric method. The interpretation may move a simple model to a more complex model, then back to be as simple as
possible if it fits the data within the error bounds. The procedure involves one
dimensional modelling at each site, then two dimensional modelling across the
profile, it might even require three dimension modelling of the region if it is
complex. There are many one dimensional modelling techniques available
compared with those for two and three dimensions. Chapter 3 gives an
introduction about different modelling methods and a detailed description
about the associated two dimensional modelling procedure. This will be used
throughout the thesis, both in forward modelling and in inversion such as the
finite difference method applied in the 2D forward program and the ridge
regression method used for 2D inversion. In reality, most cases are two, or
even three dimensional so that it is very necessary for geophysists to
challenge these problems.

Two and three-dimensional modelling has been set as an emphasized topic
for the next International Association of Geomagnetism and Aeronomy (IAGA)
Workshop in USSR in October, 1988. The work in this thesis will contribute to
two dimensional modelling analysis in Magnetotelluric investigation.

The numerical calculation for two dimension modelling involves much
computation. Specifically, inversion modelling requires the Jacobian
information (model response functions used to form Jacobian matrix), which
has to be supplied by the forward modelling program. This increases the
computation time and makes inversion very inefficient and inconvenient.
Chapter 4 introduces an iterative scheme to approximate the Jacobian
information, this speeds up forward modelling. In contrast to Rodi's (1976)
method, the iterative scheme enables a trade-off between the speed and the
accuracy to be obtained. Using this faster forward program, inversion
modelling can be conducted conveniently.
After the field data are interpreted by inversion modelling, the question follows as to how well the model parameters are resolved. Jackson (1976) used the most square inversion method to judge the resolution of the parameters by extremizing them while the misfit remains fixed. The ranges of the parameters are used to judge their resolution. If the parameter’s range is very large, this parameter will be considered not well resolved; alternatively it is well resolved. Chapter 5 presents a modified Jackson’s formulation, which enables us to consider different conditions, such as taking account of the ill-condition of the parameters’ Jacobian matrix or constraining the final model.

As the inversion procedure is carried out, many phenomena can occur. One of them is the occurrence of $\delta$-like models, especially when the model region has been divided into many blocks. Chapter 6 gives the analysis of this phenomenon by using a simple 2D example. A smoothing technique has been introduced to overcome this problem. It makes the model more acceptable and simple whilst the misfit, which is the difference between the response of the present model and the field data, is not affected much. Compared with Neumann’s (1987) a-priori information smoothing method, the method introduced in chapter 6 can achieve a smooth model without requiring a-priori information and extra iterations.

One of the important aspects about modelling is parameterization. Chapter
7 introduces the boundary parameterization in addition to resistivity parameterization. The problem arising from this new parameterization is that these parameters are not in the same space and this will effect the solution. Chapter 7 gives an analysis of this problem and introduces a scaling matrix for achieving a proper solution. The function of the scaling matrix acting on the solution is well discussed.

In chapter 8, a synthetic example for inversion modelling is made by using a known 2D model, which is from the COMMEMI project. The COMMEMI project, a Comparison of Modelling Methods in Electro-magnetic Induction problems was first proposed by Zhdanov (1982) at the 6th IAGA workshop in 1982. The purpose of the project was to compare programs for two dimension modelling. All the proposed conductivity models are quite complex and realistic and have certain difficulties in modelling.

Following the above synthetic example, chapter 9 presents an inversion modelling of real two dimensional data. The region between two of the sites has to be subdivided into a number of blocks to introduce more parameters to obtain better convergence. A possible δ model phenomenon occurs and the smoothing technique gives a very helpful tool for producing a simple and acceptable inversion model.

Through the modelling of chapters 8 and 9, many of the above techniques have been shown to be very useful and necessary in order to obtain a satisfactory inversion model. Chapter 10 gives conclusions and suggestions for further work.
CHAPTER 2

THEORY OF MAGNETOTELLURIC SOUNDING AND FIELD WORK

2.1. The fundamental theory of Magnetotellurics

The basic equations for electromagnetic induction are described by Maxwell equations. In S.I. system, these are given as:

\[ \nabla \times E = -\partial B / \partial t \]  
\[ (2.1) \]

\[ \nabla \times H = J + \partial D / \partial t \]  
\[ (2.2) \]

\[ \nabla \cdot B = 0 \]  
\[ (2.3) \]

\[ \nabla \cdot D = \rho' \]  
\[ (2.4) \]

where \( E \), \( H \) and \( B \) are the electric field strength, magnetic field strength and magnetic flux density field vectors respectively. \( D \) is the electric displacement vector, \( J \) is the electric current density and \( \rho' \) is the volume electric charge density.

Applying the constitutive relationships, where \( \varepsilon \), \( \mu \) and \( \sigma \) are the respective electric permittivity, magnetic permeability and the conductivity of the medium,

\[ D = \varepsilon E \]  
\[ (2.5) \]

\[ J = \sigma E \]  
\[ (2.6) \]

\[ B = \mu H \]  
\[ (2.7) \]

Taking the curl of eq.(2.1), and using eq.(2.5) and eq.(2.6), gives

\[ \nabla \times \nabla \times E = -\mu \sigma \varepsilon E / \partial t - \mu \varepsilon \sigma^2 E / \partial t^2 \]  
\[ (2.8) \]

In any conductivity medium space charge does not accumulate, as any initial density decays with time according to

\[ \rho'(t) = \rho'_0 e^{-\sigma t / \varepsilon} \]

For the rocks of the earth's crust and mantle \( \sigma / \varepsilon \gg 1 \) so that \( \rho' = 0 \) and eq.(2.4) becomes
Then eq.(2.8) becomes

\[ \nabla \cdot D = 0 \]  

(2.9)

Considering the range of the frequencies to be used and the earth's crust and mantle conductivities, the displacement current is negligible (Ingham, 1981), eq.(2.10) is thus simplified to

\[ \nabla^2 E = \mu_0 \sigma \partial E / \partial t \]  

(2.11)

Similarly taking the curl of eq.(2.2) and using eq.(2.1) and eq.(2.2) and neglecting the displacement current gives

\[
- \nabla^2 H = \sigma \nabla \times E + \nabla \sigma \times E \\
- \nabla^2 H = -\sigma \partial B / \partial t + \nabla \sigma \times (\nabla \times H) / \sigma
\]

(2.12)

where \( \mu \) has been put equal to the permeability of free space \( \mu_0 \) (\( \mu_0 = 4\pi \times 10^{-7} \) H.m\(^{-1} \)).

2.2. Induction in one dimensional structures

The definition here of a one dimensional structure is that the conductivity of the earth varies only as a function of depth, that is \( \sigma = \sigma(z) \), where \( z \) is measured vertically downwards in Cartesian co-ordinates.

If the structure is presumed to be a uniform conducting half space from the earth's surface downward, that is \( \sigma = \sigma_0 \), the fields are obtained as

\[ E_x(z) = \alpha e^{-ikz} \]  

(2.13)

\[ H_y(z) = (k / \mu \omega) \alpha e^{-ikz} \]  

(2.14)

where \( \alpha \) is a constant related to the source field, \( k = (i \sigma \mu \omega)^{1/2} = (1+i)/\delta \) is the wave number, \( \delta = (2 / \sigma \mu \omega)^{1/2} \) is defined as the skin depth in which the field value has decayed by a factor 1/e of the surface field value, \( \omega = 2\pi / T \) and \( T \) is the field period. In place of \( \delta \) we use another parameter which is called the wave
The wave impedance measured at $z=0$ for the uniform medium half space is defined by

$$Z_{xy} = \frac{E_y(0)}{H_y(0)} \quad (2.15)$$

$$Z_{yx} = \frac{E_x(0)}{H_x(0)} \quad (2.16)$$

Using eq.(2.13), eq.(2.14), eq.(2.15) and eq.(2.16), shows that

$$Z = Z_{xy} = -Z_{yx} = -\frac{\omega \mu}{k} = (\omega \mu / \sigma)^{\frac{1}{2}} e^{i\pi/4} \quad (2.17)$$

Knowing $Z$, the resistivity of the half space can be obtained as

$$\rho = 1 / \sigma = ZZ^*/\omega \mu \quad (2.18)$$

where $Z^*$ is the complex conjugate of $Z$.

If the earth is presumed to consist of $n$ layers with conductivities $\sigma_i$ ($z_{i-1} < z < z_{i+1}$) ($i = 1, \ldots, n$), where $z_0 = 0$, $z_n = \infty$, the solution for the wave impedance measured at the surface is then

$$Z = \frac{E_x(0)}{H_y(0)} = -\frac{E_y(0)}{H_x(0)}$$

$$Z = (i\omega / \theta_1) \coth\left[ \theta_1 h_1 + \coth^{-1}\left( (\theta_1 / \theta_2) \coth\left[ \theta_2 h_2 + \ldots \right] \right) \right] (2.19)$$

where $h_i$ ($i = 1, \ldots, n-1$) is the $i$th layer's interface depth and $\theta_i$ ($i = 1, \ldots, n$) is defined by

$$\theta_i = (1 + i)(\mu \omega \sigma_i / 2)^{\frac{1}{2}}$$

(Ward, et. al., 1973).

2.3. Induction in two dimensional and three dimensional structures

A two dimensional structure is defined such that the conductivity is a function of two coordinates (usually $y$ and $z$), for which the criterion is there is no conductivity variation along the other horizontal axis ($x$-axis) for a distance much greater than the skin depth of the incident field. If that criterion is not
satisfied, then the structure is considered three dimensional.

In the two dimension structure, the conductivity is defined as \( \sigma = \sigma(y,z) \), and the fields satisfy eq.(2.11) eq.(2.12). A general two dimensional field can be separated into two fields which represent E and H polarizations. The E polarization field is

\[
E = E_i; \quad H = (i/\mu \omega)((\partial E/\partial z)j - (\partial E/\partial y)k)
\] (2.20)

and the H polarization field is

\[
H = H_i; \quad E = (1/\sigma)((\partial H/\partial z)j - (\partial H/\partial y)k)
\] (2.21)

where \( i, j, k \) are unit vectors along the axes of Cartesian co-ordinates \( x, y, z \) respectively.

Equations (2.11) and (2.12) cannot in general be solved analytically. Analytical solutions exist only for specialized cases such as those discussed by e.g. d'Erceville and Kunetz(1962), Rankin(1962) and Weaver, et. al.(1985, 1986).

Solutions to equations (2.11), (2.12) have to satisfy the following boundary conditions:

(i) at any boundary, the tangential components of E and H and the normal components of B and D are continuous

(ii) the electric current density perpendicular to any boundary is continuous across it and is zero across \( z=0 \); the latter implies that \( E_z = 0 \) inside the conductor at \( z=0 \)

(iii) at infinite depth in the earth, both E and H are zero

(iv) the left and right boundaries are assumed far enough from any lateral
discontinuity so that the fields can be considered as one dimensional there

(v) at the earth's surface and everywhere above it, H is a constant for the H polarization case; there is no similar simplification for E polarization.

Two dimensional numerical modelling methods will be discussed in detail in chapter 3.

For a three dimensional structure, the conductivity is defined as \( \sigma=\sigma(x,y,z) \) and Maxwell's equations (2.11), (2.12) cannot be simplified. Equations eq.(2.11), eq.(2.12) cannot be solved analytically for three dimensional structures. Difficulties in determining numerical solutions to three dimensional induction problems are not due to difficulties with the mathematical formulation but to the large computer storage and the time required for the computation. These two factors must therefore be considered in the choice of a method for three dimensional modelling.

2.4. Apparent resistivity, phase and dimensionality indicator

We only consider the horizontal component of the electromagnetic fields on the surface of the earth to determine the earth's conductivity structure. Different depths are explored by different periods. Those surface components are related by the following equations.

\[
E_x = Z_{xx} H_x + Z_{xy} H_y \quad (2.22)
\]
\[
E_y = Z_{yx} H_x + Z_{yy} H_y \quad (2.23)
\]

where \( Z_{xy} \) and \( Z_{yx} \) are called the principal impedances, \( Z_{xx} \) and \( Z_{yy} \) the additional impedances due to contributions from parallel components of the magnetic field.

For a one dimensional situation
\[ Z_{xx} = Z_{yy} = 0 \quad \text{and} \quad Z_{xy} = -Z_{yx} \]

which shows that the electric field components are only related to their respective orthogonal magnetic field components. The apparent resistivity \( \rho_a \) and phase \( \phi_a \) are thus defined by Cagniard (1953) as

\[
\rho_a = \frac{1}{\mu \omega} |Z_{xy}(vy)|^2 \quad \phi_a = \text{arg}(Z_{xy}(vy))
\]

\( \rho_a, \phi_a \) are related to the ratio of orthogonal electric and magnetic fields, which vary with period(T) of the inducing field. The variation of \( \rho_a \) and \( \phi_a \) with period(T) indicates the resistivity variation with depth.

For a two-dimensional structure, \( Z_{xx} + Z_{yy} = 0 \) but \( Z_{xx}, Z_{yy} \) are in general non-zero and vary as the measurement frame is rotated with respect to the direction in which the conductivity is invariant (the strike direction). The impedances \( Z_{xx} = Z_{yy} = 0 \) only when the axes of the measurement frame are parallel and perpendicular to the strike. Then \( Z_{xy} \) and \( Z_{yx} \) are either at a maximum or a minimum (Abramovici, 1974).

For E polarization, since the electric field is parallel to the strike (denoted by \( x \) axis), the impedance is defined by

\[ Z_{xy} = E_x / H_y \]

and the apparent resistivity and phase are given by

\[
\rho_a^E = |Z_{xy}|^2 / (\omega \mu) \quad \phi_a^E = \text{arg}(Z_{xy})
\]

For H polarization

\[ Z_{yx} = E_y / H_x \]
\[ \rho_\alpha^H = |Z_{yx}|^2 / (\omega \mu) \]

\[ \phi_\alpha^H = \text{arg}(Z_{yx}) \]

In this thesis, for two dimensional modelling, \( \rho_\alpha^E, \phi_\alpha^E, \rho_\alpha^H, \phi_\alpha^H \) are considered as the response functions of the model for E and H polarizations.

For a general three dimensional structure \( Z_{xy} \) and \( Z_{yx} \), will not have coincident maxima and minima, \( Z_{xx} \) and \( Z_{yy} \) will not be small, \( Z_{xx} + Z_{yy} \neq 0 \). There is a dimensionality indicator \( s \) (the skew) given by

\[ s = \frac{|Z_{xx} - Z_{yy}|}{|Z_{xx} + Z_{yy}|} \]

For a truly two dimensional case the skew \( s \) should equal zero. However in practice an upper limit is set to 0.4, for which a two dimensional interpretation is considered valid (Vozoff, 1972).

2.5. Field work

In the field, the electric components \( E \) are measured with the electrodes buried in the earth surface. The telluric lines joining the electrodes run north south and east west. The electric field at the surface of the earth is approximated by the potential difference measured between the electrodes divided by their separation. Directional magnetometers are used to measure the magnetic fields along the same axes as the electric fields.

The signals are amplified and filtered electronically before being recorded. The recorded time series are split into “events” and each event is decomposed by a Fourier transform to obtain its frequency power spectra. The power spectra and cross power spectra are then used to derive the apparent resistivity and phase for each frequency for each event. A typical response is obtained by stacking or averaging the responses of many events for each frequency, thus obtaining the mean and standard deviation (Kaufman and Keller, 1981).
CHAPTER 3
MODELLING METHODS

After magnetotelluric data have been collected from several sites in the field, the next step is to process the data in order to produce a model of the conductivity structure of the region. The procedure starts with one-dimension modelling at each site, and these 1D inversion results can be combined to form an initial structure for a particular profile. If the profile is two dimensional then the data will be processed by two dimension modelling. The combined inversion results along several such profiles can indicate whether the region is three dimensional. If it is, three dimension modelling is necessary. One dimensional forward calculations can be performed analytically, and many inversion techniques are available. For two and three dimensional modelling, forward modelling is done numerically since no general analytical method is available. Several forward and inverse methods have been applied to two and three dimension modelling, but the techniques are comparatively far behind 1D modelling research.

3.1. One dimensional modelling

As one dimensional forward modelling can be solved analytically eq.(2.19), there are several different ways of approaching the inversion model. One of the simplest is to retain the best fitting models out of a number of randomly chosen models by giving initial bounds for the model parameters. This is called the Monte Carlo method. The program in our department was written by Jones and Hutton (1979) and a modification has been made by Dawes. The scheme starts by selecting the best fitting 20 models out of 100 random models within the given bounds, then the parameter ranges of these best selected 20 models are expanded by a given percentage to form a new set of
bounds for the parameters.

Within the new bounds another best fitting 20 models are selected out of 100 randomly chosen models, then the bounds are again moved as above. After a number of iterations, the solution is presented by the 20 best fitting models with their parameters.

Jupp and Vozoff (1975), Fischer and Quang (1981), use methods very different from the Monte Carlo method. They start the inversion with a uniform conductivity structure, which is discretized into a number of layers, each layer being parameterized. The inversion scheme is based on local linearization using Marquardt methods to alter the model parameters iteratively to approach the final model.

Fischer et. al. (1981), Nabetani and Rankin (1969), Patella (1976) solve the inversion problem in a reciprocal way to the above two methods, by moving from the data to a structure rather than from the structure to the data. The procedure starts with the shortest period of the available data set and the upper layer's resistivity as a known parameter. Then the observed data at the shortest period can be explained by a two-layer structure for which the upper layer's depth and the second layer's resistivity can be solved. Including the next longest period in the data the structure is considered as a three-layer case so that the second layer's depth and the third layer's resistivity can be determined, and so on. The process shifts successively to the long periods so that discrete new layers are introduced at progressively greater depths. The basic assumption for this method is that at a given period the observational data on the earth's surface can only be influenced by the structure that lies above a maximum depth H. The number of layers is reduced to a minimum so as to give a simple
one dimensional model.

Following the inversion theory originated by Backus and Gilbert (1967, 1978, 1970), that for geophysically interesting data there is essentially a Fréchet Kernel function which is the derivative of the response function with respect to the earth model parameters (i.e. conductivity for MT), a number of investigations have been made in one dimensional inversion [Parker (1970, 1977, 1980, 1983), Oldenburg (1979), Weidelt (1972), Parker and Whaler (1981), Hobbs (1982, 1983) and Gomez-Trevino (1987)]. Using this method the conductivity increment vector \( \delta \sigma \) can be expressed as the sum of Fréchet-Kernel functions \( (F_j, j=1, ..., N) \).

\[
\delta \sigma = \sum_{j=1}^{N} \alpha_j F_j
\]

where the coefficients \( \alpha_j \) can be obtained from the \( N \) linear equations.

\[
\sum_{k=1}^{N} (F_j, F_k) \alpha_k = E_j(\sigma') - E_j(\sigma)
\]

This is derived from the result of minimizing the norm

\[
\| \delta \sigma \|^2 = (\delta \sigma, \delta \sigma)
\]

subject to the constraints

\[
E_j(\sigma) + (F_j, \delta \sigma) = E_j(\sigma'), (j=1, ..., N)
\]

where \( E_j(\sigma') (j=1, ..., N) \) is the field data, and \( E_j(\sigma) (j=1, ..., N) \) are the response function of the present starting model.

This method has been used for investigating the earth (or moon) structure globally by assuming the earth (or moon) has a spherically symmetric conductivity distribution. Below the greatest penetration depth the field data can provide is considered as a perfectly conducting core (Parker, 1970; Jady, 1974; Jady et. al., 1983; Hobbs, et. al., 1984; Hobbs, 1987).
3.2. Two and three dimensional forward modelling

3.2.1. General description

Only a few simple two and three dimensional cases can be solved analytically. A cylinder or a sphere buried in a homogeneous media was considered by Wait (1959). d'Erceville and Kunetz (1962) considered the case of two media of different resistivity in contact along a vertical plane overlying a resistive or conductive horizontal basement (H polarization). Analytic series approximation solution for a dyke in a homogeneous media with a resistive horizontal basement (for H polarization) was given by Rankin (1962). Some further analytical cases can be found in the review paper of Hobbs (1975). Recently Weaver, et al. (1985, 1986) considered a dyke in an inhomogeneous medium with a conductive basement for both polarizations, this can be used as a control model for two dimensional forward modelling. The model response functions can also be obtained by laboratory modelling; this experimental technique has been developed and used in some coast effects and regional studies (Dosso, 1966; Dosso et al., 1986; Dosso and Nienaber, 1986).

Due to the limitations of analytic methods and the inconvenience of laboratory modelling, numerical techniques appear to be especially important for two and three dimensional forward modelling. Several numerical methods have been developed and applied in magnetotelluric studies. The finite difference method is one of them. This discretizes the whole model space. The electromagnetic equations are then approximated by finite difference equations, and the discretized field values can be solved by a system of linear algebraic equations. This was first applied by Jones and Price (1970), modified by Jones and Pascoe (1971), Pascoe and Jones (1972), Jones and Thomson...
(1974), Williamson et al. (1974) and further improvement was made by Brewitt-Taylor and Weaver (1976), Weaver and Brewitt-Taylor (1978). Many case studies in two and three dimensional modelling have been investigated with the use of this method (Jones and Price, 1971a, 1971b; Jones and Pascoe, 1972; Brewitt-Taylor, 1975, 1976; Jones and Vozoff, 1978; Dey and Morrison, 1979; Dawson and Weaver, 1979; Hutton, et al., 1981).

The finite element method is based on the minimization of the total energy of the electromagnetic system, the field values within each triangular element being approximated linearly. The variation of the total energy must be zero within the given area where the source field and the induced field energy are included. Thus the field at each discreted point can be obtained by solving a system of linear algebraic equations. This method has been widely used in engineering. The initial application in magnetotellurics was done by Coggon (1971). Compared with the finite difference method, there is an advantage in being able to treat complex model geometries and boundary conditions, but the grid preparation consumes a large amount of CPU time. With the use of this method, some slope structure modelling studies was done by Silverster and Haslam (1972), Reddy and Rankin (1973) and Reddy, et al. (1977).

The transmission surface analogy method is based on an equivalent circuit for Maxwell's field equations (Kron, 1944). Thus the field values at each discretized point can be solved by a circuit network, consisting of resistances, inductances, and capacitances. Further work was done by Johns and Beurie (1971), Johns (1977) and Brewitt-Taylor and Johns (1980). Applications in magnetotellurics were done by Swift (1971); Vozoff (1971); Ku, Hsieh and Lim (1973).

The integral equation method is another method for two and three
dimensional forward modelling. More complicated mathematical treatment is required to solve the Maxwell equations compared with other methods. The fields at all the sites can be obtained by integration only over the anomaly (or target) area so that less CPU time is required compared with other methods. The integration is done numerically by dividing the anomalous area into many cells. The surrounding medium should be uniform or layered with electrical parameters constant in each layer. This method has been used to solve some simple two and three dimension cases (Hohmann, 1971; Parry and Ward, 1971; Raiche, 1974; Weidelt, 1975; Ting and Hohmann, 1981; Zhdanov and Varentsov, 1983; Wannamaker, Hohmann and Sanfillipo, 1984; Sanfillipo and Hohman, 1985).

The computer programs for all these methods are available and widely used by many geophysicists in various case studies. As a result of the COMMEMI project the finite difference program made by Brewitt-Taylor and Weaver was recommended as one of the best. This two dimensional forward program is used throughout our modelling. The details of the program's formulation are given below.

3.2.2. MT 2D finite difference programme

The program used for two dimensional modelling was the diakoptic formulation (Brewitt-Taylor and Johns, 1980) of the finite difference method of Brewitt-Taylor and Weaver (1976).

The upper boundary of the numerical calculation region for E polarization is set quite a distance away from the earth's surface and the magnetic field there is assumed to be constant. The lower boundary has to be three to four skin depths distance away from the earth's surface and at that boundary the electric field is assumed to be zero. The left and right boundaries are required
to be three to four skin distances away from the two dimensional model area and at these distances the model is assumed to be one dimensional.

The upper boundary for H polarization is set to be the earth's surface, where the magnetic field is constant. Other boundaries are set as for E polarization.

For E polarization, the fields can be expressed as (eq.(2.11))

\[
(\frac{\partial^2 E}{\partial y^2}) + (\frac{\partial^2 E}{\partial z^2}) = i\omega \mu_0 \sigma E
\]

(3.1)

Define \( \theta = 1/\omega \mu_0 \), for H polarization, (eq.(2.12))

\[
(\frac{\partial^2 B}{\partial y^2}) + \theta (\frac{\partial^2 B}{\partial z^2}) + (\frac{\partial \theta}{\partial y})(\frac{\partial B}{\partial y}) + (\frac{\partial \theta}{\partial z})(\frac{\partial B}{\partial z}) = iB
\]

(3.2)

Let the numerical region be divided into M grids in the y direction \((i=1, ..., M)\) (from left to right), and N grids in the z direction \((j=1, ..., N)\) (downwards). Using the finite difference numerical method to represent the differential equations then eq.(3.1) and eq.(3.2) can be represented by

\[
c_{i,j-1} f_{i,j-1} + c_{i,j} f_{i,j} + c_{i,j+1} f_{i,j+1} + c_{i+1,j} f_{i+1,j} + c_{i,j+1} f_{i,j+1} = 0
\]

(3.3)

\((1<i\leq M, 1<j\leq N)\)

For E polarization (H polarization) \( f_{ij} \) represents the electric field (magnetic field) at the node \( i,j \), where \( c_{i,j} \) are the coefficients which depend on the polarization, the grid spacing, the frequency and the conductivities. The fields in eq.(3.3) must be solved together with the boundary conditions resulting in the equation

\[
A x = y
\]

(3.4)

where \( A \) is a \( K \) by \( K \) matrix and \( K \) is the product of \( M \) and \( N \). The matrix \( A \) contains the fields' coefficients and is a banded matrix dependent on \( M \) and \( N \). Vector \( x \) contains the field value to be solved for and vector \( y \) contains the values derived from the boundary condition. Eq.(3.4) is solved using the
Bi-factorization method (Zollenkopf, 1971).

The Bi-factorization method is based on the equation below

\[ L^K L^{K-1} \ldots L^2 L^1 A R^1 R^2 \ldots R^{K-1} R^K = I \]

where \( L^j \) (\( j = 1, \ldots, K \)) are left-hand factor matrices, \( R^j \) (\( j = 1, \ldots, K \)) are the right-hand factor matrices and \( I \) is the unity matrix. So that

\[ A^{-1} = R^1 R^2 \ldots R^{K-1} R^K L^K L^{K-1} \ldots L^2 L^1 \]

\( L^j \) (\( j = 1, \ldots, K \)) are very sparse and differ from the unity matrix in only column \( j \); \( R^j \) (\( j = 1, \ldots, K \)) are also very sparse and differ from the unity matrix in only row \( j \). Since \( A^{-1} \) is obtained explicitly this method allows repeated solutions for different right-hand sides (as \( y \) in eq.(3.4)) without repeating the reduction process.

After the field value is obtained by \( x = A^{-1} y \), for E polarization, the magnetic field value can be solved by eq.(2.20), and for H polarization, the electric field value can be solved by eq.(2.21). Knowing both the electric and magnetic field values, the impedance and the apparent resistivities and phases along the earth's surface can be calculated.

3.3. Two and three dimensional inversion modelling

3.3.1. General description

The response functions are not linear functions, that the data observed at the selected sites are not linearly related to the model parameters. With local linearization to approximate the non-linear functions the inversion can be carried out by multiple linear regressions to reach the best fit model, i.e. the difference between the corresponding data of the resulting model and the field data is a minimum. This method, called the least square method, has been used in nearly every field of science. The method has several variants, mainly

In some geophysical problems, the inversion is carried out considering a-priori information. This is done by a probabilistic treatment of the a-priori information as constraints on the parameters during the inversion iterations (Backus and Gilbert, 1970a, 1970b; Jackson, 1979; Pous, Marcuello and Queralt, 1987).

3.3.2. The least square method

Let the response function estimated from the field data (or data sample points) be $g_i$ ($i=1, ..., M$), and let the model function be $f_i(p)$ ($i=1, ..., M$) where $p$ is the model parameter vector of dimension $N$, $p=[p_1, ..., p_N]$. The number of the data sample points is often far larger than the number of model parameters $N$ ($M \gg N$), in addition, $f_i(p)$ ($i=1, ..., M$) is a non-linear function. The solution of this overdetermined systems of nonlinear equations raises problems. In this case the equations are in general not consistent so that a solution must be sought in the least square sense. That is the parameter vector $p$ is obtained by the following minimization.
\[
\text{Min}\left(\sum_{i=1}^{M} (g_i - f_i(p))^2\right) \tag{3.5}
\]

If \( f_i(p) \) is a linear function \( p \) can be found straight away. If \( f_i(p) \) is non-linear then \( f_i(p) \) may be approximated by a first order expansion of its Taylor series, that is

\[
f_i(p) = f_i(p_0 + \delta p) = f_i(p_0) + \sum_{j=1}^{N} \left( \frac{\partial f_i(p_0)}{\partial p_j} \right) \Delta p_j \tag{3.6}
\]

where \( p_0 \) is denoted as the parameter vector of the present model, and \( \delta p \) is the parameter increments vector to be solved for and \( \delta p = (\Delta p_1, ..., \Delta p_N)^T \). Let \( x_j = \Delta p_j \) and \( x = (x_1, ..., x_N)^T \), \( y = g_i - f_i(p) \) and \( y = (y_1, ..., y_M)^T \). Also let \( a_{ij} = \frac{\partial f_i(p_0)}{\partial p_j} \) and \( A = (a_{ij}) \), \( (i = 1, ..., M) \), \( (j = 1, ..., N) \) so that \( A \) is a matrix of dimension \( M \) by \( N \). \( A \) is often called the Jacobian matrix. Then the above minimization (3.5) may be written as

\[
\text{Min}[\ (Ax-y)^T(Ax-y)] \tag{3.7}
\]

Define

\[
s(x) = (Ax-y)^T(Ax-y) \tag{3.8}
\]

To find the minimization, it is required to satisfy the following necessary condition for a stationary value.

\[
\frac{\partial s(x)}{\partial x_j} = 2e_j^T A^T Ax - 2e_j^T A^T y = 0, \quad (i = 1, ..., N)
\]

where \( e_j \) is a vector with the \( j \)th element equal to 1 and others zero.

Thus \( x \) must satisfy the set of linear equations (the normal equations)

\[
A^T Ax = A^T y \tag{3.9}
\]

If \( A \) has rank \( N \), \( Ax = 0 \) only if \( x = 0 \). It follows that the matrix \((A^T A)\) is positive definite and hence is nonsingular. The solution \( x \) is determined uniquely by

\[
x = (A^T A)^{-1} A^T y \tag{3.10}
\]

and this is called the least square solution.
3.3.3. Ridge regression method

The linear approximation to \( f(p) \) can often lead to an unacceptable solution \( x \), which means that the parameter increments and hence the error in the linear approximation to \( f(p) \), can be large in which case the model diverges. Considering this fact, the Euclidean length \( x^Tx \) has to be constrained to seek possible convergence for the inversion.

Let the eigenvalues of \( A^TA \) be \( \gamma_j \) (\( j=1, \ldots, N \)), then a seriously nonorthogonal (or ill-conditioned) problem is characterized by the fact that the smallest eigenvalue \( \gamma_{\text{Min}} \) can be \( << 1 \), or \( \gamma_{\text{Max}}/\gamma_{\text{Min}} \) is very large. If the data collected from the field has been contaminated by random noise \( \epsilon \), then the Expectations \( E(\epsilon)=0 \) and \( E(\epsilon\epsilon^T)=\sigma^2 I_M \). The variance for the solved parameter increments \( x \) is

\[
V(x)=\sigma^2 (A^TA)^{-1} \tag{3.11}
\]

Since \( \gamma_{\text{Min}} << 1 \), \( \sigma^2/\gamma_{\text{Min}} \) can show the dramatic inadequacy of least square solution for nonorthogonal problems. To better approximate the solution the matrix \( A^TA \) needs to be stabilized. One way to deal with this problem is to add a diagonal matrix \( \lambda I \) (\( \lambda > 0 \)) to the matrix \( A^TA \). The solution for the parameter increments will be

\[
x=(A^TA+\lambda I)^{-1}A^Ty \tag{3.12}
\]

\( \lambda \) is called the damping factor. The above solution is called the ridge regression solution. It can be represented as \( \text{Min}(Ax-y)^T(Ax-y) \) with constraints \( x^Tx=c \), where \( c \) is a given constant. The conditioned minimization leads to

\[
\text{Min}[ (Ax-y)^T(Ax-y)+\lambda(x^Tx-c) ] \tag{3.13}
\]

resulting in
\[(\Lambda^T\Lambda + \lambda I)x = \Lambda^Ty\]  

which gives the solution to \(x\) as represented by eq.(3.12).

With a proper choice of the damping factor \(\lambda\), the inverse of matrix \((\Lambda^T\Lambda)\) is stabilized so that the effect of errors in the data can be reduced, and the Euclidean length of the solution vector \(x\) can be constrained to make the linear approximation of the response function valid.

3.4. Analysis of the ridge regression method and its application in MT 2D inversion

3.4.1. The general characteristics

Comparing the solution obtained from eq.(3.12) to the least square solution in eq.(3.10), Marquardt (1970) shows that the damped solution has the following identities.

(i) Let \(\lambda(>0)\) be arbitrary, and let \(x\) satisfy (3.12). Then \(x\) minimizes the sum of the square of residuals of eq.(3.8) on the sphere centred at the origin whose squared radius is \((x^Tx)\).

Further \(|x|\) is a continuous monotonic decreasing function of \(\lambda\) such that as \(\lambda \to \infty\), \(|x| \to 0\), while \(s(x)\) is a monotonic increasing function of \(\lambda\). That is

\[
\lim_{\lambda \to \infty} |x| = 0
\]

\[
\lim_{\lambda \to \infty} s(x) = y^Ty
\]

Thus by reducing the length of the regression coefficient vector, the ridge estimates have the potential to overcome the most serious deficiency of the least square solution. But the ridge solution requires some increase of the residual sum of squares above the minimum value when \(x\) is obtained by eq.(3.10).
(ii) Define $b=\mathbf{A}^\top \mathbf{y}$. Let $\phi_\lambda$ be the angle between $\mathbf{x}$ and $\mathbf{b}$. Then $\phi_\lambda$ is a continuous monotone decreasing function of $\lambda$, such that as $\lambda \to \infty$, $\phi_\lambda \to 0$. Since $\mathbf{b}$ is independent of $\lambda$, it follows that $\mathbf{x}$ rotates toward $\mathbf{b}$ as $\lambda \to \infty$.

(iii) Denote the solution obtained from the least square method as $\mathbf{x}_L$ (as in eq.(3.10)), then the ridge solution $\mathbf{x}$ (as in eq.(3.12)) is a linear transform of $\mathbf{x}_L$. That is

$$\mathbf{x} = (\mathbf{A}^\top \mathbf{A} + \lambda \mathbf{I})^{-1}(\mathbf{A}^\top \mathbf{A})$$

$$\mathbf{x} = Z_\lambda \mathbf{x}_L$$

where

$$Z_\lambda = (\mathbf{A}^\top \mathbf{A} + \lambda \mathbf{I})^{-1}(\mathbf{A}^\top \mathbf{A})$$

It follows immediately that $\mathbf{x}$ is a biased estimator of $\mathbf{x}_L$.

$$\mathbb{E}(\mathbf{x}) = Z_\lambda \mathbf{x}_L$$

The variance of $\mathbf{x}$ is

$$\mathbb{V}(\mathbf{x}) = \sigma^2 (\mathbf{A}^\top \mathbf{A} + \lambda \mathbf{I})^{-1} \mathbf{A}^\top \mathbf{A} (\mathbf{A}^\top \mathbf{A} + \lambda \mathbf{I})^{-1}$$

The above gives a detailed analysis of the ridge regression method, which is going to guide us in the application to MT 2D inversion modelling.

3.4.2. The decomposition of the Jacobian matrix

Lanczos (1958) introduced a decomposition which provides a spectral expansion for an arbitrary matrix.

Let $\mathbf{A}$ be a real $M$ by $N$ matrix $M>N$ with rank $N$, then

$$\mathbf{A} = \mathbf{U} \mathbf{Q} \mathbf{V}^\top$$

where $\mathbf{U}$ is a matrix of $M$ by $N$, $\mathbf{Q}$ is a diagonal matrix
The diagonal elements of $Q$ are the non-negative square roots of the eigenvalues of $A^T A$; they are called singular values. We shall assume that

$$q_1 \geq q_2 \geq \ldots \geq q_N \geq 0 \quad (3.23)$$

For $U$ and $V$, they have the following relation

$$U^T U = V^T V = V V^T = I_N \quad (3.24)$$

where $I_N$ is the $N$ by $N$ identity matrix.

With this decomposition, $A^T A$ is then expressed as

$$A^T A = V Q^2 V^T \quad (3.25)$$

thus

$$A^T A + \lambda I = V (Q^2 + \lambda I) V^T \quad (3.26)$$

and

$$(A^T A + \lambda I)^{-1} A^T = V (Q^2 + \lambda I)^{-1} V^T \quad (3.27)$$

Eq.(3.12) becomes

$$x = V (Q^2 + \lambda I)^{-1} Q U^T y \quad (3.28)$$

If the rank of matrix $A$ is less than $N$, i.e. rank($A$) = $p$, $p < N$, then $q_{p+1} = \ldots = q_N = 0$.

The decomposition in eq.(3.22) is called the singular value decomposition (SVD).

Then eq.(3.22) becomes

$$A = U_p Q_p V_p^T$$

where
and

\[ Q_p = \text{diag}(q_1, \ldots, q_p) \]  \hspace{1cm} (3.30)

The computational method used for this decomposition is given by Golub (1970). In the application to MT 2D inversion modelling each parameter is chosen to represent aspects of the model so that often rank(A) = N. The cases we are going to study are mostly

\[ q_i > 0 \quad (i = 1, \ldots, N) \]  \hspace{1cm} (3.31)

But cases like

\[ q_1 / q_N >> 1 \]  \hspace{1cm} (3.32)

often exist, where \( q_1, q_N \) are the maximum and minimum eigenvalues respectively. The equations are then called ill-conditioned. In this case the use of the damping factor helps us to obtain a better solution.

3.4.3. The analysis of the ill-conditioned solution

To analyse the problem, let us consider a special case. Suppose the columns in matrix A are orthogonal then \( A^T A \) is a diagonal matrix. If some of the parameters of the model are far less sensitive than others then the corresponding diagonal elements in \( A^T A \) will be far smaller than others.

\[ A^T A = \text{diag}(\alpha_1, \ldots, \alpha_N) \]  \hspace{1cm} (3.33)

Those smaller elements can be assumed as \( \alpha_{p+1}, \ldots, \alpha_N \) (1 < p < N). The least square solution from (3.10) is then

\[ x = \text{diag}(1/\alpha_1, \ldots, 1/\alpha_N) A^T y \]  \hspace{1cm} (3.34)

It is seen the the most insensitive parameter (or so-called unimportant parameter) will get the biggest increment step for the solution. These unimportant parameters, which contribute little to the variation of the model, can undergo very large changes for a small change in the data. It is not what
we expect the model to do.

In addition, the large increment step for the solution may distort the model, and cause the step to exceed the limit in which the linear approximation is valid. A model example in MT is that some highly conductive layers or blocks make parameters for the deeper region "unimportant".

If the above is damped then

$$\mathbf{x} = \text{diag}(1/((\alpha_1+\lambda), \ldots, 1/((\alpha_N+\lambda))A^T \mathbf{y}$$  \hspace{1cm} (3.35)

The use of a proper damping factor can therefore overcome this problem.

3.5. The choice of the damping factor

To choose the proper damping factor $\lambda$, its ridge plot is required. This is a graph of parameter increments $\mathbf{x}$ against misfit. From this plot the damping factor can be selected around the point beyond which each element in $\mathbf{x}$ turns out to be stable (Hoerl and Kennard, 1970) (e.g. Fig. 3.1). It is also possible in some well conditioned cases that $\lambda$ can be taken as zero. It is very subjective and experimental to deal with the selection of the damping factor.

3.5.1. Computation description

Before the MT 2D inversion modelling is carried out, the Jacobian matrix $A$ has to be obtained by running the 2D forward program. Then the matrix $A$ and the vector $y$ can be set up. The decomposition of matrix $A$ enables the inverse matrix $(A^T A)^{-1}$ or $(A^T A+\lambda I)^{-1}$ to be calculated conveniently. For any given damping factor $\lambda$ the solution $\mathbf{x}$ may be solved using eq.(3.28). The ridge plot has to be produced in order for a proper damping factor to be chosen. The damping factor is given in the form of
Figure 3.1 Each line is a parameter increment $x_i$ versus the damping factor $\lambda$. The graph indicates that even after rather heavy damping with $\lambda=32$ or $\lambda=53.4$, some of the elements in $x$ are still unstable. The proper damping factor is chosen as $\lambda=515$ where $x$ becomes quite stable. These three values of $\lambda$ are shown on the graph as broken lines.

This example of a ridge plot refers to the model used in chapter 6 and is derived from the first iteration for the 28 resistivity parameter increments shown in Fig. 6.4.
\[ \lambda = \alpha^2 (q_1 - q_N)^2 \]

$q_1, q_N$ correspond to the maximum and minimum eigen values in the diagonal matrix $Q$. $0 < \alpha < 1$. Many sample points will be taken for $\lambda$. As $\alpha$ increases from 0 to 1, the density of the sample points will decrease.

Using the ridge plot a proper damping factor $\lambda$ can be chosen and the corresponding parameter increment solution $x$ is obtained (Fig. 3.1). The computation diagram description is given in Fig. 3.2.
input the Jacobian information

set the Jacobian matrix $A$ and vector $Y$

decompose $A$ using SVD method

For any given damping factor $\lambda$, $x$ can be obtained by eq. (3.28)

produce the ridge plot graph

choose the proper damping factor and obtain the solutions $x$, $p$

stop

Figure 3.2 The computation procedure diagram for using the ridge regression method.
CHAPTER 4

AN ITERATIVE SCHEME OF USE IN THE MT 2D FORWARD PROGRAM

TO SPEED UP THE PRODUCTION OF THE JACOBIAN INFORMATION

AND TO IMPROVE THE ACCURACY OF THE INVERSE MATRIX

An important stage in two dimensional magnetotelluric modelling is the calculation of the earth's response functions for an assumed conductivity model and the calculation of the associated Jacobian information relating those response functions to the model parameters. The efficiency of the calculation of the Jacobian information will affect the efficiency for inversion modelling. Rodi (1976) produced all the Jacobian elements using a single inverse matrix in an approximate first order algorithm. Since only one inverse matrix required calculation then the procedure speeded up the inversion. Following the suggestion of Hill (1987), this chapter presents an iterative scheme to improve the approximation to the Jacobian. It takes a little longer than Rodi's algorithm. However sacrificing a little time enables the solution to be obtained more accurately. It was found that the Jacobian elements could be produced in 10% of the time required to calculate an inverse matrix or to calculate a 2D starting model. A modification of the algorithm could further be used to improve the original inverse matrix calculated in the 2D finite difference program and hence the solution this program produced. The convergence of the iteration scheme is given in this chapter and is discussed in detail. By using this modified forward program together with an inversion routine, 2D inversion can be carried out conveniently.

4.1. Introduction

The finite difference program (Brewitt-Taylor and Weaver, 1976) is not efficient when the Jacobian information is needed for a magnetotelluric (or
MT) 2D inversion when many parameters are involved. At each stage of the inversion a new model is produced, and for this new model another Jacobian matrix is required for the next inversion, and so on. This routine continues until a satisfactory resolving model is obtained. At each stage a new starting model may be described by a vector parameter \( p \). That is

\[ p = (p_1, p_2, \ldots, p_N)^T \]  

where \( p_j \) \((j=1,N)\) are the model parameters (resistivities, depths, etc.)

For a model with parameter \( p \), the response function may be defined as

\[ f(p) = (f_1(p), f_2(p), \ldots, f_M(p))^T \]  

where \( M \) is the number of data sample points (e.g. \( M \) is the product of the number of frequencies, number of sites, number of polarisation, etc.), and \( f_i(p) \) is the response for the corresponding \( i \)th data sample point.

When \( p \) changes to \( p_1 \) by \( \delta p \), its Jacobian matrix will be

\[
J = \begin{bmatrix}
\frac{\partial f_1}{\partial p_1} \\
\vdots \\
\frac{\partial f_M}{\partial p_1}
\end{bmatrix}_{j=1,N}
\]  

\( J \) is an \( M \) by \( N \) matrix.

The inversion method used for MT 2D modelling is the ridge regression method (or Marquardt method) (presented in section 3.3.3.). Because of the non-linearity of the response function \( f(p) \), it is necessary to calculate its Jacobian information for each new starting model. The Jacobian information has to be calculated by the 2D forward program.

When a new model is established by using the finite difference method the corresponding system in the program is \( Ax = y \) which can be solved as \( x = A^{-1}y \), where \( A \) is an \( N \) by \( N \) matrix, \( y \) is a vector of dimension \( N \) and \( x \) is a vector of dimension \( N \) to be solved for which contains the field variables. As \( p \) changes
to $p_1$ by $\delta_p$ then the system changes to $A_1x'=y_1$. In the inversion, in addition to $x$ ($x=A^{-1}y$), either $x'$ or $\partial x/\partial p$ is required to obtain the Jacobian matrix since the relationship between $f$, $x$ and $p$ is $f(p)$. Rodi (1976) used the first order approximation of the partial derivatives to approximate the field response, that is he differentiated $Ax=y$ with respect to $p$, which gives

$$\frac{\partial \lambda}{\partial p} x + \lambda \frac{\partial x}{\partial p} = \frac{\partial y}{\partial p}$$

then

$$\frac{\partial x}{\partial p} = \lambda^{-1} \frac{\partial y}{\partial p} - \lambda^{-1} \frac{\partial \lambda}{\partial p} x$$

The above expression for $\partial x/\partial p$ is then used by Rodi to calculate the Jacobian information. This procedure is also used elsewhere (Jupp and Vozoff, 1977, Oristaglio and Worthington, 1980). To obtain a better approximation to $x'$ and hence the Jacobian information, the program is modified using the method below.

4.2. Using an iterative scheme to approximate the inverse matrix for the Jacobian information

For the system we have set if a model parameter $p$ is changed by $\delta_p$, then $A$ changes to $A_1$, $y$ to $y_1$, and $x'$ should be obtained by

$$A_1x'=y_1$$

i.e.

$$x'=A_1^{-1}y_1$$

Since the calculation $A_1^{-1}$ would take about 90% of the time of the entire run of the forward program, it is better to use the inverse matrix $A^{-1}$ to approximate $A_1^{-1}$, since $A^{-1}$ has already been calculated for each new start model. This is done by the algebra below.

Let $x_0=x=A^{-1}y$, the initial difference $\Delta_0$ is found by
\[ A_1 x_0 - y_1 = \Delta_0 \]

and this is then used to calculate \( x_1 \) via

\[ x_1 = -A^{-1} \Delta_0 + x_0 \quad (4.7) \]

\( x_1 \) is used in the calculation of \( \Delta_1 \):

\[ A_1 x_1 - y_1 = \Delta_1 \quad (4.8) \]

\[ x_2 = A^{-1} \Delta_1 + x_1 \quad (4.9) \]

... ...

and so on

\[ A_1 x_{n-1} - y_1 = \Delta_{n-1} \quad (4.10) \]

\[ x_n = -A^{-1} \Delta_{n-1} + x_{n-1} \quad (4.11) \]

Define

\[ \delta_A = A_1 - A \]

\[ \delta_y = y_1 - y \]

The relation between the \( n \)th and the \((n-1)\)th estimate of \( x \) is

\[ A x_n + \delta_A x_{n-1} = y_1 \quad (4.12) \]

\( x_n \) may be expressed as

\[ x_n = A^{-1} y_1 - A^{-1} \delta_A x_{n-1} \]

Iteratively it is

\[ x_n = (-1)^n (A^{-1} \delta_A)^n x_0 + \left[ (-1)^{n-1}(A^{-1} \delta_A)^{n-1} + \ldots + (-1)A^{-1} \delta_A + 1 \right] A^{-1} y_1 \]

\[ = \left[ 1 + (-1)A^{-1} \delta_A + (-1)^2 (A^{-1} \delta_A)^2 + \ldots + (-1)^n (A^{-1} \delta_A)^n \right] A^{-1} y_1 \quad (4.13) \]

If the norm of \( A^{-1} \delta_A \) is less than 1 (\( \| A^{-1} \delta_A \| < 1 \)), the above series is convergent (Appendix 1(b)).
When $n \to \infty$, by eq. (4.20) in Appendix 1(b) and eq. (4.22) in Appendix 1(c)

$$x_n \to (1 + A^{-1} \delta_A)^{-1} \Lambda^{-1} y_1 = \Lambda^{-1} y_1$$

showing that $x_n (n \to \infty)$ converges to $x'$.

Comparing eq. (4.12) with eq. (4.6), when $x_n = x_{n-1}$ we have $x_n = x' = A^{-1} y_1$.

In the program certain criteria were used to determine when convergence was reached, namely:

\[ c1 \]
\[ \|x_{n+1} - x_n\|^2 \leq \varepsilon \]
\[ c2 \]
\[ \xi = \frac{\|A_1 x_n - y_1\|^2}{\|A_1 x_0 - y_1\|^2} \leq \varepsilon \]

for some small $\varepsilon$.

Alternatively the iterations were stopped after a given number of cycles. An example is given in table 4.2.

Following the norm definition in Appendix 1(a), then the norm $\|A^{-1} \delta_A\|$ is equal to the square root of the largest eigenvalue of the positive semidefinite Hermitian matrix $(A^{-1} \delta_A)^* (A^{-1} \delta_A)$ where $(A^{-1} \delta_A)^*$ is the complex conjugate of the transpose of the matrix $(A^{-1} \delta_A)$, that is

$$\|A^{-1} \delta_A\| = \left( \lambda_{\max} \left[ (A^{-1} \delta_A)^* (A^{-1} \delta_A) \right] \right)^{\frac{1}{2}}$$

(Franklin, 1968)

An upper bound of the norm $\|A^{-1} \delta_A\|$ is the square root of the sum of the squared elements of the matrix $A^{-1} \delta_A$ (eq. (4.25), Appendix 1(d)).

It often appears that three or four iterations are adequate. The iterative
procedure used to approximate \((A+\delta A)^{-1}\) was found to be about ten times faster than that used to calculate \(A_1^{-1}\) directly. Since each parameter would require an inverse \(A_1^{-1}\); the above method speeds up the program efficiently for a many parameter problem compared to using the entire program to calculate a number of inverses \(A_1^{-1}\).

For the case \(n=1\), the above algorithm includes Rodi's. From eq.(4.7)

\[
x_1 = -A_1^{-1}A_0 + x_0 = -A_1^{-1}(A_1x_0 - y_1) + x_0
\]

move \(x_0\) to the left and let \(A_1 = (A_1 - A) + A\)

\[
x_1 = A_1^{-1}(y_1 - y) - A_1^{-1}(A_1 - A)x
\]

divide both sides by the change made in the parameter vector \(p\)

\[
\frac{\partial x}{\partial p} = A_1^{-1} \frac{\partial y}{\partial p} - A_1^{-1} \frac{\partial \Lambda}{\partial p} x
\]

This is identical to eq.(4.4)

4.3. Case study

Using an ICL 2988 computer with 8 Mb main memory, a 77 by 56 mesh requires 251 CPU seconds to calculate the inverse matrix \(A_1^{-1}\) and 7 CPU seconds to get \(x\) from \(x = A_1^{-1}y\). 11 CPU seconds are required to set up and calculate the apparent resistivity and phase. In this case the proportion of time taken to calculate \(A_1^{-1}\) is 93% CPU time of the entire run of the program.

When the iterative method was used to approximate \((A+\delta A)^{-1}\), on average three or four iterations were adequate. Each iteration took 7 CPU seconds, and hence the program is speeded up by about a factor of ten. If \(n\) is the number of iterations, then the speed-up factors \(f\) can be expressed as below. \(f\) is the time ratio between calculating \(A_1^{-1}\) directly to resolve \(x_1\) and using the iterations to approximate it.
n=1  \( f = \frac{251}{7n} = 36 \) (times)
n=3  \( f = \frac{251}{7n} = 12 \) (times)
n=4  \( f = \frac{251}{7n} = 9 \) (times)

This is very significant, especially if there are many frequencies and parameters to deal with; the saving in time is then very important.

If the computer system is AMDAHL 470V/8 with 16 Mb as the main memory, a 77 by 56 mesh requires 31.24 CPU seconds to calculate the inverse matrix \( A^{-1} \) directly. If the iteration scheme is used to approximate the inverse matrix, it requires 0.87 CPU seconds to finish one iteration; then the factor \( f \) as defined above is

\[
n=1 \quad f = \frac{31.24}{0.87n} = 36 \text{ (times)}
n=3 \quad f = \frac{31.24}{0.87n} = 12 \text{ (times)}
n=4 \quad f = \frac{31.24}{0.87n} = 9 \text{ (times)}
\]

This shows for a given grid size the speed factor \( \alpha \) does not change even though the computer system changes.

Taking the above particular example, for a computer system and mesh size 77 by 56, a formula for the speed-up time for the total problem can be given. The times quoted are proportional to the number of nodes in the finite element mesh. \( M \) is number of model parameters, \( n \) is the chosen number of iterations. \( r \) is the time ratio between the unmodified and the modified iteration method presented here.

\[
r = \frac{11(M+1)+(251+7)(M+1)}{251+11(M+1)+(7n+7)(M+1)}
\]

For \( M=0 \), no Jacobian information is required hence \( n=0 \) and \( r=1 \).
For $M \to \infty$, 
\[ r = \frac{(1+251+7)}{(7n+7+11)} = \frac{269}{7n+7+11} \]

The maximum number of iterations normally required for sufficient accuracy is 4, giving

\[ n=4, \quad r = 6 \text{ (times)} \]

The minimum number of operations is when $n=1$, this will give the maximum value to the time ratio $r$ which is what Rodi's first order approximation obtains, namely

\[ n=1, \quad r = 11 \text{ (times)} \]

Hence the range for the time ratio (or the speed up) is approximately 6 to 11.

4.4. Using the iterative scheme to improve the accuracy of the originally calculated inverse matrix

The above shows the algebra for approximating the inverse matrix when the Jacobian information is required. Before that, another calculation should be considered, that is the accuracy of the calculation of $A^{-1}$ which can be improved. If $B^{-1}$ is the result of the actual calculation for the inverse of matrix $A$ in the program using Bi-factorization (Zollenkopf, 1971), the fact is often that $AB^{-1} \neq I$. Then the solution for $x$ is $x = B^{-1}y_0$ but $Ax - y_0 \neq 0$.

Let $A = B + \delta_B$, and use the iterative scheme in section 4.2, where $A$ is equivalent to $A_1$, $B$ to $A$, $y_0$ to $y$ and $\delta_B$ to $\delta_A$ in that section. If the norm of $B^{-1}\delta_B$ is less than 1, then as $n \to \infty$

\[ x_n \to (I + B^{-1}\delta_B)^{-1}B^{-1}y_0 = Ax_0 \]

The criteria for stopping the iterations are the same as given by $c_1$ or $c_2$ (in section 4.2). If the iteration stops at $n=r$, and $x = x_r$, then
\[ Ax - y_0 = e \]

where \( e \) is a vector with dimension of \( N \) and contains the difference between \( Ax \) and \( y_0 \). So before calculating the Jacobian information, the original system has to be set by

\[ Ax - y = 0 \]

where \( y = y_0 + e \). Then any change in the vector of parameters \( p \) causing a change in \( A \) and \( y \) determines the change in \( x \), and this is obtained by the iterative scheme presented in section 4.2.

4.5. Examples and computer procedure

In the model of Fig. 4.1, \( R_2 = 600 \Omega \), \( R_1 = 5 \Omega \), \( Y_1 = -3 \text{.0 km} \), \( Y_2 = 3 \text{.0 km} \), \( Z_1 = 1.55 \text{.0 km} \), \( Z_2 = 6 \text{.0 km} \). For a period of \( T = 0 \text{.1 (sec.)} \) Table 4.1 shows an example of the iterative scheme to improve the accuracy of the originally calculated inverse matrix \( A^{-1} \). For both polarizations the first one or two iterations improve the accuracy of the original calculated inverse matrix significantly but there is slow convergence for further iterations. The E polarization scheme diverged at the sixth iteration, while the H polarization scheme diverged at the third iteration. The convergence criterion is given as \( c_2 \) (in section 4.2).

Following above, \( R_2 \) was changed from 600\( \Omega \) to 500\( \Omega \), i.e. by 17\%. For a period of \( T = 0 \text{.1 (sec.)} \) Table 4.2 shows an example of approximating the inverse matrix for the Jacobian information. For both polarizations, the first two or three iterations show significant convergence rate and are very necessary in approximating the inverse matrix for the Jacobian information compared with when only one iteration is taken (\( n = 1 \), which is equivalent to Rodi's first order approximation). The iterations still converged very well for some further iterations. However, the E polarisation scheme diverged at the eighth iteration and the H polarisation scheme diverged at the ninth iteration but
Figure 4.1 A block with resistivity $R_2$ is surrounded by resistivity $R_1$. $Z_1, Z_2$ are the upper and lower depths of the block, $Y_1, Y_2$ are the left and right boundaries of the block, relative to the origin $O$. 
Table 4.1

<table>
<thead>
<tr>
<th>No. n</th>
<th>n=0</th>
<th>n=1</th>
<th>n=2</th>
<th>n=3</th>
<th>n=4</th>
<th>n=5</th>
<th>n=6</th>
</tr>
</thead>
<tbody>
<tr>
<td>E_P, ξ</td>
<td>1</td>
<td>0.0963</td>
<td>0.0881</td>
<td>0.0856</td>
<td>0.0845</td>
<td>0.0757</td>
<td>0.08</td>
</tr>
<tr>
<td>H_P, ξ</td>
<td>1</td>
<td>0.0268</td>
<td>0.0228</td>
<td>0.246</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Table 4.1* E_P is the abbreviation for E polarization and H_P is for H polarization. ξ is the convergence criterion given as c2 (in section 4.2) and n is the iteration number.

Table 4.2

<table>
<thead>
<tr>
<th>No. n</th>
<th>E_P, ξ</th>
<th>H_P, ξ</th>
</tr>
</thead>
<tbody>
<tr>
<td>n=1</td>
<td>0.0565</td>
<td>0.028</td>
</tr>
<tr>
<td>n=2</td>
<td>0.0029</td>
<td>0.0007</td>
</tr>
<tr>
<td>n=3</td>
<td>0.0001</td>
<td>0.0002</td>
</tr>
<tr>
<td>n=4</td>
<td>0.7x10^{-5}</td>
<td>0.16x10^{-5}</td>
</tr>
<tr>
<td>n=5</td>
<td>0.2x10^{-5}</td>
<td>0.1x10^{-5}</td>
</tr>
<tr>
<td>n=6</td>
<td>0.15x10^{-5}</td>
<td>0.87x10^{-6}</td>
</tr>
<tr>
<td>n=7</td>
<td>0.14x10^{-5}</td>
<td>0.85x10^{-6}</td>
</tr>
<tr>
<td>n=8</td>
<td>0.15x10^{-5}</td>
<td>0.8x10^{-6}</td>
</tr>
<tr>
<td>n=9</td>
<td>0.85x10^{-6}</td>
<td></td>
</tr>
</tbody>
</table>

*Table 4.2* E_P is the abbreviation for E polarization and H_P is for H polarization. ξ is the convergence criterion given as c2 (in section 4.2) and n is the iteration number.
the first two or three iterations provide sufficient accuracy. The convergence
criterion is given as $c_2$ (in section 4.2).

The computation details are described below.

For each polarization at each period, the forward program has a separate
stage which is called file preparation. Before running the main program a file
has to be prepared containing information such as the grids, the boundary
conditions and the structure of the model. Let file model$_0$ be the file prepared
as the current starting model. File model$_1$ is then the file in which parameter
$p_1$ changes by $\delta$ to obtain the Jacobian information for $p_1$. Similarly files
model$_2$, ..., model$_N$ are prepared for obtaining Jacobian information for
parameters $p_2$, ..., $p_N$.

Let $A$ and $x$ be the corresponding matrix and solution for file model$_0$ and
$A_1'$, ..., $A_N'$ and $x_1'$, ..., $x_N'$ the corresponding files for model$_1$, ..., model$_N$ (they
are equivalent to $A_1$ and $x'$ in section 4.2 when the iteration scheme is used
for approximating the relevant inverse matrix). The computation procedure for
running the main part of the program can be described as follows (Fig. 4.2).

Prepare files model$_0$, ..., model$_N$, then input each file individually. Most of
the computation is done when file model$_0$ is input. First of all matrix $A$ and
vector $y_0$ are set, then $A^{-1}$ has to be calculated by using Bi-factorization
(Zollenkopf, 1971). Thereafter the iterative scheme is used to improve the
accuracy of this inverse matrix (as stated in section 4.4). The solution $x$ can be
obtained by $x = A^{-1}y_0$ and the corresponding output for the present model can
be given (denoted by response$_0$). Even though some improvement has been
made to the calculation of $A^{-1}$ by using the iterative scheme, a residue may
still occur as $e = Ax - y_0$. Set $y = y_0 + e$, which enables $Ax - y = 0$. $A^{-1}$, $x$ and $y$ are
stored for later use.

Then input file model_1, A is replaced by $A_1'$ and $y$ is replaced by $y_1$, $x_1'$ can be obtained by using the iterative scheme presented in section 4.2. The corresponding output can be obtained (denoted by response_1). Similar procedures for the other files result in corresponding outputs(response_2, ..., response_N). The Jacobian information for the present model can then be calculated from these outputs (response_0, response_1, ..., response_N) for the given polarization at one period. If the field data contain both polarizations and a number of periods the whole procedure above has to be repeated many times to obtain the whole Jacobian information in order to carry out one iteration of the inversion.

4.6. Conclusion and discussion

When many parameters are involved in 2D inversion, the iterative scheme for approximating the Jacobian information can be very useful. Compared with the original calculation from the forward program and Rodi's first order approximation, the iterative scheme is a trade off between accuracy and speed in producing the Jacobian information. The method can also be used to improve the accuracy of the original calculation of the inverse matrix $A^{-1}$. This technique in generating Jacobian information has general applicabilities.

Some users measure $\frac{\|A_1x_n-y_1\|}{\|A_1x_{n-1}-y_1\|}$ to judge convergence of the iterations. In this case the phenomenon shown below may be noticed.

Even if the norm $\|A_1x_n-y_1\|$ starts to diverge at the nth iteration, that is

$$\|A_1x_n-y_1\| > \|A_1x_{n-1}-y_1\|$$

the iterative result $x_n$ may still converge to $x'$. That is
Figure 4.2 The computation procedure for the modified two dimension forward program in which the iterative scheme is applied to improve the accuracy of the original inverse matrix and to approximate the inverse matrix for the Jacobian information.
input file model

set matrix $A$, vector $y_0$

calculate the inverse of matrix $A$

use the iterative scheme to improve the accuracy of the inverse matrix $x$ is obtained simultaneously

Find the residue $e = Ax - y_0$
set $y = y_0 + e$
then $Ax - y = 0$

input model_1, ..., model_N individually

set matrix $A_i$, vector $y_i$

use the iterative scheme to produce $x'$ where $A_i^{-1}$ is approximated by $A^{-1}$

output files response_1, ..., response_N individually

stop
\[ \|x_n - x'\| < \|x_{n-1} - x'\| \]
as long as the following conditions hold:

\[ \|A^{-1}\delta_A\| < 1 \quad \text{and} \quad \|A\delta_A^{-1}\| < 1. \]

If we require both norms \(\|A_1 x_n - y_1\|\) and \(\|x_n - x'\|\) convergent, that is

\[ \|A_1 x_n - y_1\| < \|A_1 x_{n-1} - y_1\| \quad \text{and} \quad \|x_n - x'\| < \|x_{n-1} - x'\| \]
then the following conditions need to be satisfied:

\[ \|A^{-1}\delta_A\| < 1 \quad \text{and} \quad \|\delta_A A^{-1}\| < 1 \]

The proof of the above statement is in Appendix 1(e) ((i), (ii), (iii)).
5.1. The resolution formulation given by Jackson

After the field data have been interpreted by a model, the next question is how well the model parameters have been resolved. Jackson's (1976) method is to maximize and minimize each parameter increment \( x_k, k=1, \ldots, N \) with the sum of squares of the misfit fixed.

If we consider \( e = Ax - y \)

Then the sum of squares of the misfit is defined by

\[
S = e^T e = (Ax - y)^T (Ax - y)
\]

For the least square solution, \( x_L = (A^T A)^{-1} A^T y \)

Thus for this solution, the sum of the squares of the misfit is

\[
S_L = (Ax_L - y)^T (Ax_L - y)
\]

In this method the value of a particular parameter is maximized (or minimized) subject to the constraint that the sum of the squares of the misfit (or residual) is equal to the threshold \( s_0 \). This problem is a specific version of the problem of extremizing the linear objective function \( x^T b \) subject to the quadratic constraint \( s = s_0 \) (\( b \) is a vector of dimension \( N \)). The inner product \( x^T b \) presents a projection of the unknown vector \( x \) along a known vector \( b \).

To maximize or minimize the \( k \)th component \( x_k \), one would set \( b_i = 0, i \neq k \) and \( b_i = 1, i = k \). As long as the matrix \( A^T A \) is positive definite, the desired extremum will always take place when the constraint is exactly satisfied, i.e., \( s = s_0 \). The calculation requires differentiation of

\[
x^T b + (1/2 \mu) [(Ax - y)^T (Ax - y) - s_0]
\]

with respect to \( x \). The factor \( 1/2 \mu \) is a Lagrangian multiplier. Setting the derivative to zero results in
\[ x = (A^T A)^{-1} (A^T y - \mu b) \quad (5.2) \]

Using the constraint
\[ (A x - y)^T (A x - y) = s_0 \]
the solution is
\[
\mu = \pm \left( \frac{s_0 - y^T y + y^T A (A^T A)^{-1} A^T y}{b^T (A^T A)^{-1} b} \right)^{\frac{1}{2}}
\]
\[
= \pm \left( \frac{s_0 - s_L}{b^T (A^T A)^{-1} b} \right)^{\frac{1}{2}}
\]

By eq.(5.2), \( x \) is then given as
\[
x = (A^T A)^{-1} A^T y \pm \frac{(A^T A)^{-1} (s_0 - s_L)^{\frac{1}{2}} b}{(b^T (A^T A)^{-1} b)^{\frac{1}{2}}}
\]
\[
x = x_L \pm \frac{(A^T A)^{-1} (s_0 - s_L)^{\frac{1}{2}} b}{(b^T (A^T A)^{-1} b)^{\frac{1}{2}}} \quad (5.3)
\]

These are the maximum and minimum values of \( x \) for its extremum when subject to a given sum of the squares of the misfit (or a certain percentage perturbation of the sum of the squares of the misfit of the present model). These form a range for each \( x_i \) or the model parameter \( p_i \) (\( i = 1, ..., N \)). The bigger the range is, the worse the resolution is for that parameter.

5.2. The modified resolution

5.2.1. Extremizing the increments in conjunction with stabilization

Since \( (A^T A)^{-1} \) can be rather ill-conditioned then a damping factor is applied to stabilize the solution. Following the Marquardt (ridge regression) method, which has been fully described in section 3.3.3, the stabilized solution for \( x \) is found by minimizing the sum \( s \), where
\[
s = (A x - y)^T (A x - y) + \lambda x^T x \quad (5.4)
\]
which gives
\[ x = (A^T A + \lambda I)^{-1} A^T y \] (5.5)

This result replaces the minimization of \((Ax-y)^T (Ax-y)\) to obtain \(x = (A^T A)^{-1} A^T y\), which is the least square solution.

Using the same idea as above, Jackson's formulation for extremization can be modified by using the constraint \(s = s_0\), where \(s\) is represented by eq. (5.4). Then \(x\) is extremized along a known vector \(b\) (\(b\) is defined as before). Now differentiating

\[ x^T b + \frac{1}{2} \mu [ (Ax-y)^T (Ax-y) + \lambda x^T x - s_0 ] \] (5.6)

with respect to \(x\), and setting this derivative to zero results in

\[ x = [ A^T A + \lambda I ]^{-1} (A^{-1} y - \mu b) \] (5.7)

\(\mu\) can be obtained by substituting this \(x\) into (5.4) with the constraint \(s = s_0\).

The result is

\[ \mu = \pm \frac{s_{0} - s_{\lambda}}{b^T (A^T A + \lambda I)^{-1} b} \] (5.8)

where \(s_{\lambda} = y^T y - y^T A (A^T A + \lambda I)^{-1} A^T y\), which is the sum of the squares of the misfit from eq. (5.4) when \(x\) is solved by eq. (5.5).

After this modification, the extremization for each parameter gives a better description of the resolution since the ill-condition from \((A^T A)^{-1}\) is reduced. It is perhaps more mathematically correct to consider resolution subject to the constraint \(s_{0} = (Ax-y)^T (Ax-y)\) than to \(s_{0} = (Ax-y)^T (Ax-y) + \lambda x^T x\). However, the latter produces the same benefit as replacing the least square method by the ridge regression method where \(\lambda x^T x\) has to be constrained. Moreover, as stated in section 3.3.3, the ill-conditioned matrix \((A^T A)^{-1}\) can give unacceptably false information for the resolution if the field data have been contaminated by noise. Also the response function has been linearized, which can only be satisfied for a small increment range of the parameter \(p\). If the
step x is rather large then the condition for the linearization can no longer hold thus false information may be given.

5.2.2. Extremizing each parameter increment with others being constrained

If the matrix \((A^TA)^{-1}\) is well conditioned, the formula presented by eq.(5.3) extremizes \(x_i\), but at the same time other parameters will be affected. If the effect is considerably large compared with the extremized parameters then the assumption of resolution in respect of the present model is no longer valid. In order to give each parameter's resolution based on the present model, bounds are placed on other parameters \(p_j\) (or effectively their increments \(x_j\)) \((j \neq i)\) so that they are only allowed to change within a given percentage \(\delta\) of the present model parameter values. The procedure is described as below.

After \(x\) is solved by eq.(5.3), if \(x_i\) is being extremized, the other parameter increments \(x_j\) (effectively the parameters \(p_j\)) \((j \neq i)\) are examined to see whether or not they are within the given percentage bound \(\delta\) with respect to the present model parameter values. If a parameter \(p_j\) is within the set bounds then it is retained as the solution from eq.(5.3), if not then the parameter \(p_j\) (effectively its increment \(x_j\)) is set to its upper bound. These bounded parameter increments \(x_j\) \((j \neq i)\) then form a vector called \(x_0\). Let \(a_i\) be the ith column of \(A\), which is the corresponding column of \(x_i\) in \(A\), \(A_0\) be the remaining part in \(A\), \(y_1 = y - A_0 x_0\) and \(b_i\) be the given constant (preferably as 1). Then the extremization for \(x_i\) is once again calculated by extremizing \(x_i b_i\) subject to eq.(5.1). Since \(x_0\) is known then eq.(5.1) can be expressed as

\[
s_0 = (a_i x_i - y_1)^T (a_i x_i - y_1)
\]

Using the same algebra as in section 5.1, the solution is
After this modification, the resolution of a parameter of the present model can be given relative to other parameters changing their limits within a percentage δ when the sum of the squares of the misfit is set to $s_0$. It is found that the parameters which exceed their bounds are often those parameters which are poorly resolved. Consequently their restriction has little effect on the extremized values under consideration. In our application, $s_0$ is often perturbed by 10% with respect to the sum of the squares of the misfit the present model, thus the model parameter constraint percentage δ is also set to 10%. These two modifications for the resolution study will be used in later chapters.
CHAPTER 6
ANALYSIS OF THE DELTA-LIKE MODEL FORMULATION AND MODEL SMOOTHING

The two-dimensional inversion of magnetotelluric (or MT) data with the use of the ridge regression method can result in an unacceptable model when the area is divided into many blocks in the search for a 2D region and its conductivities. One of the unexpected models is a δ-like model, in which some resistivity blocks with large resistivity values are surrounded by blocks with lower resistivity. The result can be worse if the data are contaminated by random noise. This chapter analyses the causes of this unreasonable model with the use of the present inversion method. It also provides a smoothing technique enabling a better model with the same acceptance criteria and initial conditions to be achieved.

6.1. Introduction

If the model area is over parameterized or parameterized in detail for various reasons (i.e. to investigate a detailed or precise structure), the resulting model becomes rather complicated using the ridge regression method for inversion. One of the phenomena is the δ-like model. Neumann (1986) used a model weighting matrix to constrain the non-unique solution space to conductivity functions which are smoothly varying in space. Thus the tendency to degenerate into discrete delta functions is suppressed. Also, Neumann's later (1987) smoothing method uses a-priori smoothness and resolution constraints on the model to carry out the 2D inversion with a densely parameterized model. The a-priori information constraint to the solution can be described as below.

Jackson suggested one way of solving non-uniqueness in linear inversion when a-priori data are used (Jackson, 1979). This enables the inversion
procedure to be carried out considering the a-priori information. The parameter increments vector is solved as
\[
x = (A^T C_m^{-1} A + C_n^{-1})^{-1} (A^T C_m^{-1} Y + C_n^{-1} (\mu - p_{n-1}))
\]
(6.1)
where \(A, y, x\) are defined as before (chapter 5), \(p_{n-1}\) is the formal solution of the parameters or the initial set up for the parameters, \(C_m\) is the covariance matrix of the field data, \(C_n\) is the covariance matrix of the parameters and \(\mu\) is a vector which contains the expected values of the parameters.

If we let \(C_m=\sigma_m^{-2} I\), \(C_n=\sigma_n^{-2} I\) and \(\mu=p_{n-1}\), eq.(6.1) becomes
\[
x = (\sigma_m^{-2} A^T A + \sigma_n^{-2} I)^{-1} \sigma_m^{-2} A^T y
\]
(6.2)
This is identical to the ridge regression method, where \(\sigma_m\) and \(\sigma_n\) are the data variance and parameter variance respectively.

This method requires a-priori information which might be assigned artificially in considering an acceptable model. In addition, those parameters for which a-priori knowledge is available are not expected to move much during the iterations of inversion from the initial values. This is achieved by making the corresponding variances \(\sigma_n\) very small. The method is equivalent to using heavy damping factors for those parameters in which case other parameters may also be affected. Thus many iterations are required to complete the inversion modelling. (i.e. Pous, Marchello and Queralt, 1987).

The δ-model phenomenon is unlikely to appear if the model is not parameterized in detail (Cerv and Pek, 1981).

This chapter will analyse the δ-model formulation when an area is parameterized in detail, and another way to smooth the model. This has the advantage of not affecting the misfit and hardly requiring extra iterations to achieve a smoothed model.
6.2. An example of 2D inversion and its resultant model

Here a simple 2D case is presented. A realistic structure is assumed which allows accurate two dimensional forward computations of amplitude and phase to be made. We set resistivity equal to 1000 $\Omega$ m at depths greater than 60 Kilometers (Roberts, 1983). The proposed model is shown in Fig. 6.1, with parameter values $R_1=100\Omega$m, $R_2=5\Omega$m, $R_3=1000\Omega$m, $Z_1=1.55$km, $Z_2=4.3$km, $Z_3=60$km, $Y_1=0.0$km, $Y_2=5.0$km. In order to provide sufficient data for inversion 22 sites and 17 frequencies were set (Fig. 6.2). The 22 sites are located at $Y(km)=-16, -13, -9, -7, -6, -5, -4, -3, -2, -1.1, 0, 1.1, 2, 3, 4, 5, 6, 7, 8, 9, 13, 16$. The 17 periods are $T(\text{seconds})=0.02, 0.1, 0.5, 1.0, 3.0, 7.0, 10.0, 20.0, 60.0, 100.0, 200.0, 350.0, 500.0, 750.0, 1000.0, 2000.0, 5000.0$.

One dimensional inversion of the data was used to determine the outer limits and the average resistivity of a conductive region which was used as an initial two dimension model (Fig. 6.2), where $R_1=100\Omega$m, $R_2=20.0\Omega$m, $R_3=1000\Omega$m, $Z_1=0.75$km, $Z_2=7.2$km, $Z_3=60.0$km, $Y_1=-4.0$km, $Y_2=9.0$km. In order to make the 2D model well parameterized, the conductive region was divided into 28 blocks each with a resistivity of 20$\Omega$m or the average value obtained from the 1D inversion (Fig. 6.3). In the 2D inversion $R_1, R_3, Z_3$ were kept unchanged since the sites cover a large area where the 1D inversion already gives a satisfactory solution and they hardly change even if they are set as free parameters for the 2D inversion. All the data for the 2D model was generated by the finite difference method (Brewitt-Taylor and Weaver, 1976). The inversion part was done by a ridge regression program. The sites and the frequencies were chosen so that over each column of blocks there was at least one site, and the frequencies covered the region of the blocks in depth. The blocks were hence well defined for the inversion. In the process of the inversion, only the resistivity parameter in each block was considered.
Figure 6.1 A block with resistivity $R_2$ is surrounded by resistivity $R_1$. $Z_1, Z_2$ are the upper and lower depths of the block and $Y_1, Y_2$ are the left and right boundaries of the block relative to the origin $O$. $R_3, Z_3$ are the resistivity and depth of the half space below.
Figure 6.2 Locations of the 22 sites providing data for inversion relative to the true and initial models. The solid line is the real model with $R_2=5\Omega m, Z_1=1.55km, Z_2=4.3km, Y_1=0.0km, Y_2=5.0km$. The dashed line is the initial model set for the 2D inversion with the parameter values $R_2=20\Omega m, Z_1=0.75, Z_2=7.2km, Y_1=-4.0km, Y_2=9.0km$, where the parameters have been explained in Fig. 6.1. Since the parameters $R_1, R_3, Z_3$ are kept unchanged in the inversion, these parameters are not drawn in this figure.
\text{Sites (km)}
The initial model area as in Fig. 6.2 is divided into 28 blocks (dashed lines). The real model area is outlined with a thick solid line.
The mean sum of the squares of the misfit amplitude was of the form:

\[ M_{\rho}\text{ho} = 2\left[ \sum_{i=0}^{M/2} \left( \frac{RHO_{in} - RHO_{da}}{err_{RHO}} \right)^2 \right] / M \]  
(6.3)

The mean sum of the squares of the misfit for phase was of the form:

\[ M_{\rho\text{ha}} = 2\left[ \sum_{i=1}^{M/2} \left( \frac{PHA_{in} - PHA_{da}}{err_{PHA}} \right)^2 \right] / M \]  
(6.4)

The total mean sum of the squares of the misfit was given as

\[ M = \frac{(M_{\rho}\text{ho} + M_{\rho\text{ha}})}{2} \]  
(6.5)

M is the number of data sample points (e.g., M is the product of number of the frequencies, number of the sites, number of polarizations and number of the representations of the data (i.e., amplitude, phase)). \( err_{RHO} \) was the error in the amplitude in log space, while \( err_{PHA} \) was the error in the phase in linear space. \( RHO_{da}, PHA_{da} \) were the real data of amplitude and phase respectively, while \( RHO_{in}, PHA_{in} \) were computed from the solution of the parameters \( p \). \( p \) is a vector of dimension N, it is

\[ p = [p_1, p_2, ..., p_N]^T \]

where \( p_i \) are the model resistivity parameters as used here (Fig. 6.4). The amplitude was in log space, the phase was in linear space. After the nth iteration the solution for parameters is denoted by \( p_n \), that is

\[ p_n = p_{n-1} + x \]  
(6.6)

where the parameter and its increment are in the same space.

If \( M_s \leq 1 \), by using an approximate application of the chi-squared \( \chi^2 \) test, then the data generated from the model belonged to the data set at the 95% confidence level, i.e., were acceptable.

The data generated from the program were noise free. Hence we set the error in amplitude \( \rho \), \( err_{RHO} = 0.04 \), which was 10% of the amplitude and the error...
Figure 6.4 The parameter vector is \( \mathbf{p} = [p_1, p_2, \ldots, p_n]^T \), where \( p_i \) is the \( i \)th block's resistivity value. The real model area is that parameterized by \( p_{11}, p_{12}, p_{13}, p_{14}, p_{19}, p_{20}, p_{21}, p_{22} \).
After two iterations, the resistivity contrast between adjacent blocks became larger with $p_{18}$, $p_{23}$, $p_{26}$, $p_{27}$ assuming high values and the surrounding parameters low values (Fig. 6.5 and Fig. 6.4). We shall refer to this as a $\delta$ model. Following that, another three iterations were conducted. The resultant model was more like a $\delta$ model than previously even though the mean sum of the squares of the misfit was quite acceptable (Fig. 6.5).

6.3. Analysis of the formulation of the delta-like model

First of all we must notice that a nonlinear problem is being solved by a linear approximation. This may be ill-conditioned. Even after quite heavy damping ill-condition effects could not be removed completely (Fig. 6.6). We also notice that with 28 parameters the freedom of choosing parameter values could be rather large and hence there would be a considerable possibility of the formulation of a $\delta$ model.

The next reason we could give is that the initial model was not sufficiently near to the solution. This allowed the generation to a local minimum. In Fig. 6.5 we noticed that the model proceeded to a $\delta$ model. However in Fig. 6.9 the 8th iteration model which was in the proximity of the solution and the 9th iteration model were well resolved with lowered resistivity contrasts in those areas where the resistivity contrasts were high for the $\delta$ models.

One further reason is that the large number of parameters enabled some parameters with large values to be surrounded by parameters with low values. Such a combination produced an averaged response. It is worthwhile studying the resolution by using the formulation in section 5.5.1 (chapter 5). The extremes were given by eq.(5.7). Each element's maximum and minimum
Figure 6.5 In the initial model each value in the block is the initial resistivity value (Ωm) assigned to that block. When the iterations are taken then the value in each block is its resolved resistivity value (Ωm) after that iteration of the inversion. Each block is in the same relative position as in Fig. 6.4. $M_M$ is the misfit which is expressed as eq. (6.5).
The initial model, $M_s = 28.3$

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
</tbody>
</table>

The second iteration, $M_s = 0.383$

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>147</td>
<td>72</td>
<td>42</td>
<td>69</td>
<td>56</td>
<td>47</td>
<td>78</td>
<td>150</td>
</tr>
<tr>
<td>99</td>
<td>113</td>
<td>17</td>
<td>5</td>
<td>1.6</td>
<td>19</td>
<td>99</td>
<td>111</td>
</tr>
<tr>
<td>26</td>
<td>199</td>
<td>26</td>
<td>12</td>
<td>7</td>
<td>32</td>
<td>134</td>
<td>27</td>
</tr>
<tr>
<td>20</td>
<td>182</td>
<td>191</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>21</td>
</tr>
</tbody>
</table>

The fifth iteration, $M_s = 0.025$

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>87</td>
<td>102</td>
<td>91</td>
<td>107</td>
<td>110</td>
<td>97</td>
<td>100</td>
<td>87</td>
</tr>
<tr>
<td>142</td>
<td>64</td>
<td>9</td>
<td>4</td>
<td>3</td>
<td>8</td>
<td>69</td>
<td>160</td>
</tr>
<tr>
<td>64</td>
<td>316</td>
<td>21</td>
<td>4</td>
<td>3.6</td>
<td>15</td>
<td>242</td>
<td>54</td>
</tr>
<tr>
<td>43</td>
<td>374</td>
<td>358</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>46</td>
</tr>
</tbody>
</table>
Figure 6.6 The value in each block and the misfit $M_s$ is explained in the same way as in Fig. 6.5. Two different results of the first iteration following the initial model in Fig. 6.5 are produced when two different damping factors are chosen. The lower one $\lambda=32$ shows that some parameters reached extremely high values and were surrounded by some very low values due to ill conditioning in the present inversion method. The upper one $\lambda=515$, which is the proper damping factor value being choosing in that iteration, shows a much more stable resolution, and that extremely high values are reduced enormously and that surrounding lower values are increased to some extent compared with $\lambda=32$. It is noticed that even after such heavy damping $\lambda=515$ the combinations between high and low values which formed in $\lambda=32$ still exist while the contrast between them is reduced significantly. Fig. 3.1 is the associated ridge plot.
The first iteration (damping factor=515), $M_s=4.72$

<table>
<thead>
<tr>
<th>149</th>
<th>99</th>
<th>39</th>
<th>33</th>
<th>22</th>
<th>47</th>
<th>100</th>
<th>141</th>
</tr>
</thead>
<tbody>
<tr>
<td>42</td>
<td>75</td>
<td>17</td>
<td>4.6</td>
<td>2.8</td>
<td>14</td>
<td>63</td>
<td>48</td>
</tr>
<tr>
<td>13</td>
<td>51</td>
<td>20</td>
<td>13</td>
<td>11</td>
<td>21</td>
<td>42</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>65</td>
<td>84</td>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The first iteration (damping factor=32), $M_s=3.1$

<table>
<thead>
<tr>
<th>161</th>
<th>77</th>
<th>66</th>
<th>46</th>
<th>36</th>
<th>61</th>
<th>80</th>
<th>146</th>
</tr>
</thead>
<tbody>
<tr>
<td>46</td>
<td>80</td>
<td>4.6</td>
<td>4.1</td>
<td>1.5</td>
<td>8.1</td>
<td>63.7</td>
<td>71.4</td>
</tr>
<tr>
<td>4.6</td>
<td>2135</td>
<td>6.2</td>
<td>3.5</td>
<td>1.2</td>
<td>19.8</td>
<td>1317</td>
<td>3.36</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.3</td>
<td>191</td>
<td>341</td>
<td>66</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
results after extremization are compared with the proper resolution for the same sum of the squares of the misfit defined by eq.(5.4) (Fig. 6.7, Fig. 6.8 taken from the fifth run of the inversion). Let \( x_i \) be the parameter increment associated with the parameter \( p_i \) \((i=1, ..., N)\) as in Fig. 6.4. In Fig. 6.7 we extremized parameter increment \( x_{13} \), in Fig. 6.8 we extremized parameter \( x_{25} \). Fig. 6.7 shows that when \( x_{13} \) is maximized (the long horizontal line in the enclosed box), the value is greater than the proper solution, that is for extremization which is the proper solution (the middle horizontal line in that box). Meanwhile the parameter to the left \( x_{12} \), the one to the right \( x_{14} \) and the one below \( x_{21} \) go down in order to compensate the change in \( x_{13} \). Hence the misfit defined by eq.(5.4) is kept unchanged. When \( x_{13} \) is minimized (the short horizontal line in the enclosed box), the reverse occurs. This shows how the \( \delta \)-like model can be formed by self adjustment. Fig. 6.8 is another example of this kind of formulation.

6.4. Model smoothing

We shall now consider how to obtain an acceptable model from the inversion. We did find one way to smooth the model and the result turned out to be very successful.

The model parameters were divided into groups and the parameters within the groups were smoothed. After several iterations the groups were characterized by distinct resistivities. In our case we smoothed the model after five iterations by which time the two groups could be clearly identified. We have \( p_{11}, p_{12}, p_{13}, p_{14}, p_{19}, p_{20}, p_{21}, p_{22} \) as one group (Fig. 6.4, Fig. 6.5) which have considerably lower resistivity, while others as another group have relatively high resistivity.

The smoothing technique utilized the indeterminate parts of the general
Figure 6.7 Graph of parameter increments \((x_1, x_2, \ldots, x_N)\) after the 5th iteration (middle size horizontal lines). Parameter increment \(x_{13}\) is outlined. The longest line is the maximized increment value, the shortest line is the minimized increment value. When \(x_{13}\) is so extremized, the other parameter increments that result are then shown as the longest horizontal lines (when \(x_{13}\) is maximized) and the shortest horizontal lines (when \(x_{13}\) is minimized). The positions of these parameter increments corresponding to the positions of the parameters shown in Fig. 6.4. The vertical length represents the parameter increment value normalized so that 0.5 unit = 0.205 \(\Omega m\) (in log space).
Figure 6.8 As for Fig. 6.7 except that parameter increment $x_{25}$ is extremized.
solution for \( x \) in any given iteration. In the following discussion we shall use the notation of Lanczos (Lanczos, 1958), where

\[
A = UQV^T
\]

\( V \) is a matrix of \( N \) by \( N \), which is orthogonal, \( Q \) is a diagonal matrix of \( N \) by \( N \) and contains the eigenvalues of matrix \( A \), \( U \) is an \( M \) by \( N \) matrix, \( U^TU = I \), \( UU^T = I \), which is semi-orthogonal. If the rank of matrix \( A \) is \( p \) then \( Q = [Q_p, Q_0] \), where \( Q_p \) contains the \( p \) non-zero eigenvalues of the matrix \( A \) and \( Q_0 \) contains the \( N-p \) zero eigenvalues of the matrix \( A \). Corresponding to \( Q_p, Q_0 \), \( U, V \) can be decomposed as \( U = [U_p, U_0] \), \( V = [V_p, V_0] \), where \( U_p \) is a matrix of \( M \) by \( p \), \( U_0 \) is a matrix of \( M \) by \( N-p \), and \( V_p \) is a matrix of \( N \) by \( p \), \( V_0 \) is a matrix of \( N \) by \( N-p \). Thus matrix \( A \) is expressed as

\[
A = UQV^T = [U_p, U_0] [Q_p, Q_0] [V_p, V_0]^T = U_p Q_p V_p^T
\] (6.7)

Eq.(3.12) may hence be expressed as

\[
x = (V_p Q_p^2 V_p^T + \lambda I)^{-1} V_p Q_p U_p^T y
\] (6.8)

In reality for computational purpose we may assume that the relatively small eigenvalues are equal to zero. Let \( p_n \) be the model parameter solution vector of the \( n \)th iteration of inversion and \( p_{n-1} \) be the former resolution of inversion, then by eq.(6.6) and eq.(6.8)

\[
p_n = p_{n-1} + x = p_{n-1} + (V_p Q_p^2 V_p^T + \lambda I)^{-1} V_p Q_p U_p^T y
\] (6.9)

where \( p_n \) can be expressed as

\[
p_n = (p_1, p_2, \ldots, p_N)^T
\] (6.10)

Let \( k \) be an unknown vector of dimension \( M-p \) by 1 and \( p_s \) be the desired smooth parameter vector of dimension \( N \). To obtain each element \( p_s^i \) in \( p_s \) the resolved \( p_i \) is smoothed within its adjacent parameters which belong to the same parameter group. For example referring to Fig. 6.4 since \( p_1, p_3 \), and \( p_{10} \) are adjacent to \( p_2 \) and they also belong to the same group (as indicated
after the fifth iteration (Fig. 6.5), the smoothed value of $p_2$ is given by

$$p_s^2 = \frac{(p_1 + p_2 + p_3 + p_{10})}{4} \quad (6.11)$$

We proceeded to produce a smoothed parameter resolution $p_n'$ at the $n$th iteration of the inversion by minimizing with respect to $k$ where

$$\min[(V_0k-p_n+V_0k)^T(V_0k-p_n+V_0k)] \quad (6.12)$$

The resulted smooth model parameter vector $p_n'$ is thus obtained as below.

$$p_n' = p_n + V_0k \quad (6.13)$$

where $p_n'$ is expressed as

$$p_n' = (p_1', p_2', \ldots, p_n')^T \quad (6.14)$$

Alternatively

$$p_n' = p_{n-1} + x + V_0k \quad (6.15)$$

let $x' = x + V_0k$, then

$$p_n' = p_{n-1} + x' \quad (6.16)$$

so that $x'$ is the resultant increment of the parameters after the $n$th iteration of the inversion. The projection of $V_0k$ on $A$ is zero. This means that $Ax' = Ax + AV_0k = Ax$, so that the misfit $M_s$ is not affected by adding $V_0k$ to the solution $x$. It can be shown by

$$AV_0k = U_p Q_p V_p^TV_0k \quad (6.17)$$

$V$ is an orthogonal matrix so that $V_p^TV_0$ results in a zero matrix, that is $V_p^TV_0 = 0$. Then $Ax' = Ax$ holds. Menke (1984) suggested that the $V_0$ matrix can be used for the constraint of the model in our interest.

$(N-p)<N$ so that equation $V_0k = p_0-p_n$ is indeterminate. Thus some parameter values may not be well defined. The remedy is as follows:

Set $q = p_i'/p_s^i \quad (6.18)$

$p_i'$, $p_s^i$ are the $i$th elements in vectors $p_n'$, $p_s$ respectively. If $q \geq 1.5$, $p_i'$ is altered as
which tries to restore $p_i$ back to $p_i$.

For the given model one inversion had the fifth and the eighth iterations smoothed while for another inversion of the same model none of the iterations were smoothed. The former inversion produced an acceptable model (Fig. 6.5, Fig. 6.9), while the latter inversion produced a $\delta$-like model (Fig. 6.5, Fig. 6.10).

Fig. 6.5 shows that a $\delta$-like model was produced after five iterations of the inversion. The smoothing technique was then applied at the fifth run and the resultant model is shown in Fig. 6.9, the misfit being little affected. Following this run, two more iterations were carried out, the model was then smoothed once more. At the eighth iteration this was found necessary. The inversion was completed after the ninth iteration, a satisfactory result having been obtained (Fig. 6.9).

Fig. 6.10 followed Fig. 6.5, where none of the iterations was done using the smoothing technique, and it still showed a $\delta$-like model even after nine iterations with the misfit satisfied already (Fig. 6.10).

In the field, data are usually contaminated by random noise. To approximate the field case, we added random noise (about 7.5%) to the data, then carried out the 2D model inversion (with the same starting model as above). This time the $\delta$ model structure was much stronger than in the case when no noise was added to the data. That is to say the contrast in parameter values, where the $\delta$-like model occurred, was much larger than before. We smoothed it in the same way as before at the fifth and seventh iterations and a very acceptable model was produced (Fig. 6.11, Fig. 6.12).
Figure 6.9 As in Fig. 6.5, the iterations being carried out here follow Fig. 6.5 and the smoothing technique is applied at some stages of the inversion. Where stated, smoothed means that the smoothing technique is applied during that iteration of the inversion stage.
### The Fifth iteration (smoothed), $Ms=0.07$

<table>
<thead>
<tr>
<th></th>
<th>100</th>
<th>72</th>
<th>74</th>
<th>138</th>
<th>121</th>
<th>96</th>
<th>85</th>
<th>111</th>
</tr>
</thead>
<tbody>
<tr>
<td>104</td>
<td>135</td>
<td>8.2</td>
<td>3.4</td>
<td>3.1</td>
<td>7.3</td>
<td>125</td>
<td>91</td>
<td></td>
</tr>
<tr>
<td>113</td>
<td>97</td>
<td>9.4</td>
<td>4.2</td>
<td>3.3</td>
<td>7.4</td>
<td>82</td>
<td>103</td>
<td></td>
</tr>
<tr>
<td></td>
<td>84</td>
<td>144</td>
<td>143</td>
<td>102</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### The Eighth iteration (smoothed), $Ms=0.01$

<table>
<thead>
<tr>
<th></th>
<th>97</th>
<th>99</th>
<th>102</th>
<th>106</th>
<th>106</th>
<th>101</th>
<th>101</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>108</td>
<td>100</td>
<td>5.8</td>
<td>4.5</td>
<td>4.8</td>
<td>5.5</td>
<td>100</td>
<td>105</td>
<td></td>
</tr>
<tr>
<td>92</td>
<td>96</td>
<td>5.7</td>
<td>4.7</td>
<td>4.5</td>
<td>5</td>
<td>95</td>
<td>89</td>
<td></td>
</tr>
<tr>
<td></td>
<td>96</td>
<td>134</td>
<td>135</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### The Ninth iteration, $Ms=0.0088$

<table>
<thead>
<tr>
<th></th>
<th>98</th>
<th>101</th>
<th>99</th>
<th>102</th>
<th>102</th>
<th>101</th>
<th>101</th>
<th>99</th>
</tr>
</thead>
<tbody>
<tr>
<td>104</td>
<td>97</td>
<td>5.3</td>
<td>4.9</td>
<td>4.9</td>
<td>5.2</td>
<td>100</td>
<td>106</td>
<td></td>
</tr>
<tr>
<td>93</td>
<td>99</td>
<td>5.3</td>
<td>5.1</td>
<td>4.4</td>
<td>5.4</td>
<td>92</td>
<td>95</td>
<td></td>
</tr>
<tr>
<td></td>
<td>96</td>
<td>125</td>
<td>114</td>
<td>98</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 6.10 As in Fig. 6.5, the iterations being carried out here follow Fig. 6.5 but no smoothing technique is applied.
The initial model, $\text{Ms}=29.5$

<table>
<thead>
<tr>
<th>20</th>
<th>20</th>
<th>20</th>
<th>20</th>
<th>20</th>
<th>20</th>
<th>20</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The second iteration, $\text{Ms}=1.23$

<table>
<thead>
<tr>
<th>155</th>
<th>70</th>
<th>29</th>
<th>49</th>
<th>51</th>
<th>58</th>
<th>93</th>
<th>130</th>
</tr>
</thead>
<tbody>
<tr>
<td>101</td>
<td>120</td>
<td>16</td>
<td>8</td>
<td>1</td>
<td>19</td>
<td>74</td>
<td>156</td>
</tr>
<tr>
<td>17</td>
<td>195</td>
<td>25</td>
<td>16</td>
<td>9.2</td>
<td>29</td>
<td>128</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>187</td>
<td>255</td>
<td></td>
<td></td>
<td></td>
<td>18</td>
</tr>
</tbody>
</table>

The fifth iteration, $\text{Ms}=0.6$

<table>
<thead>
<tr>
<th>80</th>
<th>85</th>
<th>52</th>
<th>82</th>
<th>94</th>
<th>101</th>
<th>126</th>
<th>85</th>
</tr>
</thead>
<tbody>
<tr>
<td>175</td>
<td>78</td>
<td>15</td>
<td>6</td>
<td>1.8</td>
<td>14</td>
<td>48</td>
<td>172</td>
</tr>
<tr>
<td>22</td>
<td>346</td>
<td>25</td>
<td>8</td>
<td>4</td>
<td>16</td>
<td>234</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>298</td>
<td>249</td>
<td></td>
<td></td>
<td></td>
<td>43</td>
</tr>
</tbody>
</table>

*Figure 6.11* As for Fig. 6.5 but the data is contaminated by noise.
### The Fifth Iteration (smoothed), $M_s=0.96$

<table>
<thead>
<tr>
<th>152</th>
<th>77</th>
<th>75</th>
<th>79</th>
<th>99</th>
<th>135</th>
<th>80</th>
<th>118</th>
</tr>
</thead>
<tbody>
<tr>
<td>86</td>
<td>183</td>
<td>14</td>
<td>5.8</td>
<td>1.3</td>
<td>89</td>
<td>132</td>
<td>81</td>
</tr>
<tr>
<td>84</td>
<td>71</td>
<td>14</td>
<td>8.1</td>
<td>4.5</td>
<td>10</td>
<td>76</td>
<td>102</td>
</tr>
<tr>
<td>73</td>
<td>145</td>
<td>151</td>
<td></td>
<td></td>
<td>89</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### The Seventh Iteration (smoothed), $M_s=0.75$

<table>
<thead>
<tr>
<th>105</th>
<th>96</th>
<th>97</th>
<th>107</th>
<th>119</th>
<th>130</th>
<th>100</th>
<th>102</th>
</tr>
</thead>
<tbody>
<tr>
<td>110</td>
<td>103</td>
<td>7.6</td>
<td>4.5</td>
<td>3.0</td>
<td>5.8</td>
<td>99</td>
<td>98</td>
</tr>
<tr>
<td>81</td>
<td>89</td>
<td>7.5</td>
<td>5.2</td>
<td>3.8</td>
<td>6.6</td>
<td>91</td>
<td>97</td>
</tr>
<tr>
<td>95</td>
<td>128</td>
<td>129</td>
<td>103</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 6.12** As for Fig. 6.9 and the iteration results from Fig. 6.11 are carried out here by using the smoothing technique at some stages of inversion. Where stated, smoothed means that the smooth technique is applied during that iteration of the inversion stage.
6.5. Conclusion and discussion

We conclude that the detailed parameterization of the model with 28 blocks had a large set of solutions because of the large number of parameters. Although fewer iterations were required for convergence and the anomaly area could be quite well defined by the parameters, δ-like models could be produced which were not acceptable. This difficulty could be overcome by the smoothing technique presented in this chapter. The use of the vector \( V_0 \) gives us a method for constraining the model in an acceptable way whilst the misfit is not effected. The fact is that with the same misfit (when \( \lambda \neq 0 \)) there are many kinds of models possible at each stage of the inversion.

We saw that noisy data produced even stronger δ-like models. Hence we should be very careful when we deal with field data using detailed model parameterization. Compared with using a-priori information to constrain the model (Neumann, 1987; Pous, et. al., 1987), the method given in this chapter achieves a smooth model without requiring either a-priori information or extra iterations. This is so since the model is smoothed according to the groups indicated after a number of iterations of inversion and this smoothing hardly affects the current misfit.
There are many ways to parameterize a 2D-model for the inversion of electromagnetic induction data. For such two dimensional models, there are often many parameters involved. The parameterization should be balanced between an adequate representation of the model's characteristics and the simplification which reduces the number of significant parameters to a minimum, because of the cost and the inefficiency of carrying out the inversion iterations with many parameters. In addition to the usual resistivity parameterization this chapter is intended to analyse and extend a boundary parameterization introduced by Hill (1987).

For the newly introduced parameterization, the parameters are in different spaces i.e. the resistivity and the horizontal boundary parameters are in logarithmic space while the vertical boundary has to be in linear space. The solution of the inversion will therefore be affected when using the ridge regression method. In particular it is found that changes in the vertical boundary parameters are suppressed. In order to reduce this effect and bring back the sensitivity of the vertical boundary parameters, a scaling matrix is introduced and its properties are analysed. For practical application, a detailed account of how to choose the proper scaling value in the scaling matrix will be presented.

7.1. An alternative parameterization for the two dimensional model inversion

Taking the same example as in chapter 6, another way to deal with the model is to consider alternative parameters. If it is required to find both the location and the resistivities of blocks within the 2D region, an initial model based on the one dimensional inversion results of each site is constructed with the minimum number of expected resistivity structure divisions. The
block resistivities and block boundaries are then considered as parameters (Hill, personal communication). In our case, we have only one block with five parameters ($Y_1, Y_2, Z_1, Z_2, R_2$) as below.

- $Y_1$ the left boundary of the block.
- $Y_2$ the right boundary of the block.
- $Z_1$ the upper boundary of the block.
- $Z_2$ the lower boundary of the block.
- $R_2$ is the block's resistivity.

We took the initial model to be that of Fig. 6.2, which was the same initial model as above and $R_1, R_3, Z_3$ are kept unchanged as in chapter 6. After eleven iterations were conducted, the model produced a very good fit and the resulting model was acceptable (Table 7.1 and Fig. 7.1). The initial model covered a much larger area than the real model. As each iteration of the inversion is carried out, the boundary parameters move and the resistivities outlined by the boundaries change — they converge to fit the field data. Fig. 7.2 shows the sequences of the approach of the boundary parameters to the real model when these eleven iterations of inversion are made, where it can be seen that before the iterations reached the eleventh step the parameter $Z_1$ moved up and down across the real model's boundary compared with other parameters. One of the reasons for this phenomenon may be that parameter $Z_1$ is a quite sensitive parameter since it is near the surface of the ground where the sites for the real model's data were taken, so it has the opportunity to move around in order to reduce the misfit. The same procedure was conducted with noise added to the data, and the result was very successful
Table 7.1

<table>
<thead>
<tr>
<th>Start model</th>
<th>Z₁</th>
<th>Z₂</th>
<th>Y₁</th>
<th>Y₂</th>
<th>R₂</th>
<th>Mₛ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.75</td>
<td>7.2</td>
<td>-4.0</td>
<td>9.0</td>
<td>20</td>
<td>28.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>second iteration</th>
<th>Z₁</th>
<th>Z₂</th>
<th>Y₁</th>
<th>Y₂</th>
<th>R₂</th>
<th>Mₛ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.3</td>
<td>7.3</td>
<td>-3.5</td>
<td>8.5</td>
<td>31.3</td>
<td>11.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ninth iteration</th>
<th>Z₁</th>
<th>Z₂</th>
<th>Y₁</th>
<th>Y₂</th>
<th>R₂</th>
<th>Mₛ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.8</td>
<td>4.86</td>
<td>-0.7</td>
<td>6.6</td>
<td>7.46</td>
<td>3.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>eleventh iteration</th>
<th>Z₁</th>
<th>Z₂</th>
<th>Y₁</th>
<th>Y₂</th>
<th>R₂</th>
<th>Mₛ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.7</td>
<td>4.5</td>
<td>-0.26</td>
<td>5.2</td>
<td>6.0</td>
<td>0.0324</td>
</tr>
</tbody>
</table>

Table 7.1 Computation details of inversion for noise free data when the model is parameterized in a different way. R₂, Z₁, Z₂, Y₁, Y₂ are the parameters explained in Fig. 6.1 Mₛ is the misfit expressed as eq. (6.5). The table presents the initial model parameter values and the misfit with the data, the resolved parameter values and the misfit with the data after that stated iteration.
Figure 7.1 The solid line is the real model with parameter values $R_2=5\Omega m$, $Z_1=1.55km$, $Z_2=4.3km$, $Y_1=0.0km$, $Y_2=5.0km$, and $R_1=100\Omega m$, $R_3=1000\Omega m$, $Z_3=60km$. Since $R_1$, $R_3$, $Z_3$ are kept unchanged during the 2D inversion they are not drawn here. The parameters' explanation is as in Fig. 6.1. The long dashed line is the initial model with parameter values $R_2=20\Omega m$, $Z_1=0.75km$, $Z_2=7.2km$, $Y_1=-4.0km$, $Y_2=9.0km$. The initial misfit is $M_s=28.7$, $M_s$ is expressed as eq.(6.5). The short dashed line is the resulting model with parameter values $R_2=5.08\Omega m$, $Z_1=1.56km$, $Z_2=4.2km$, $Y_1=0.008km$, $Y_2=4.97km$. The misfit $M_s=0.073$. 
Figure 7.2 The sequences of convergence from the initial model (long dashed line) to the real model (thick solid line). The resulting model after eleven iterations is drawn as a thick broken line, results from the ten intermediate iterations are drawn as thin dashed lines.
Compared to what happened in chapter 6, this new parameterization for the 2D inversion, considers the initial model's or the minimum divided blocks' resistivities and boundaries as the inversion parameters instead of dividing the initial model into many blocks and each block's resistivity as the inversion parameter. The new parameterization has many fewer parameters than the one before; thus much less computation is required for the 2D inversion scheme; the combinations in solution are limited hence some unexpected models (i.e. the δ-like models) may be avoided.

7.2. Problems arising

For this new parameterization, the problem to be overcome is how to scale the Jacobian matrix properly to achieve an acceptable solution since the vertical boundary parameters are in linear space while depth and resistivity parameters are in log space. For depths and resistivities there exist start positions or values (i.e. the ground is set as depth equal to zero). For the vertical boundary parameters there are no appropriate start positions available with reference to logarithmic space, so that they have to be set in linear space. When parameters are not set in the same space the solution cannot be obtained properly. Examples are analysed below.

(a) Following the definition led by eq.(3.14), which gives the solution of the parameter increments by 

\[ x=(A^T A+\lambda I)^{-1} A^T y, \]

the Jacobian matrix \( A=[a_1, a_2, \ldots, a_N] \).

Let \( a_1=a_2=\ldots=a_N=a \), where \( a=[\alpha_1, \alpha_2, \ldots, \alpha_N]^T \), then the result for \( x \) (from eq.(3.12)) is that \( x_1=x_2=\ldots=x_N \).

Proof:

As A is defined above, then
<table>
<thead>
<tr>
<th>Table 7.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start</td>
</tr>
<tr>
<td>model</td>
</tr>
<tr>
<td>Second</td>
</tr>
<tr>
<td>iteration</td>
</tr>
<tr>
<td>Ninth</td>
</tr>
<tr>
<td>iteration</td>
</tr>
<tr>
<td>eleventh</td>
</tr>
<tr>
<td>iteration</td>
</tr>
</tbody>
</table>

Table 7.2 As for Table 7.1 except that the data is contaminated by noise.
Figure 7.3 As for Fig. 7.1 except that the data is contaminated by noise and thus the resulting model is $R_2=4.7\Omega m$, $Z_1=1.57km$, $Z_2=4.1km$, $Y_1=-0.14km$, $Y_2=4.7km$. The misfit $M_2=0.7615$. 
\[ \mathbf{A}^T \mathbf{A} = \mathbf{a}^T \mathbf{a} \mathbf{E} = \xi \mathbf{E} \quad (7.1) \]

and

\[ \mathbf{A}^T \mathbf{y} = \eta \mathbf{e} \quad (7.2) \]

where \( \mathbf{E} \) is a matrix of \( N \) by \( N \) with each element equal to 1,

\[ \mathbf{\xi} = \mathbf{a}^T \mathbf{a} = \sum_{i=1}^{M} \alpha_i^2, \quad \eta = \sum_{i=1}^{M} \alpha_i y_i \]

and \( \mathbf{e} \) is a vector of dimension \( N \) in which each element is equal to 1, that is

\[ \mathbf{e} = [1,1,\ldots,1]^T \]

Using equations (7.1) and (7.2) then equation 3.14 \((\mathbf{A}^T \mathbf{A} + \mathbf{X} \mathbf{O} \mathbf{X}^T) \mathbf{y} = \mathbf{A}^T \mathbf{y} \) may be written as

\[ (\mathbf{\xi} \mathbf{E} + \lambda \mathbf{I}) \mathbf{x} = \eta \mathbf{e} \quad (7.3) \]

Let \( \lambda = k \xi \) and \( \mathbf{G} = \mathbf{E} + k \mathbf{I} \), then.

\[ \mathbf{\xi} \mathbf{G} \mathbf{x} = \eta \mathbf{e} \quad (7.4) \]

where \( \mathbf{G} \) is a matrix of \( N \) by \( N \) with the diagonal element equal to \( 1+k \) and all other elements equal to 1. Eq.(7.4) implies that

\[ k x_i + \xi = \eta / \xi \]

\[ \xi = \sum_{i=1}^{N} x_i \]

and then

\[ x_i = \left( \frac{\eta - \xi}{\xi} \right) / k \quad (i=1,\ldots,N) \quad (7.5) \]

The above expression showing that

\[ x_1 = x_2 = \ldots = x_N \]

which completes the proof.

Further, after some algebraic manipulation,
\[ x_1 = x_2 = \ldots = x_N = n/(\xi (k + N)) \]  \hspace{1cm} (7.6)

a result which will be of use later.

(b) As for (a), \( A = [a_1, a_2, \ldots, a_N] \), but \( a_i = \tau_i \alpha \), where \( \alpha = [\alpha_1, \alpha_2, \ldots, \alpha_M]^T \), and \( \tau_i \) \((i = 1, \ldots, N)\) is the \( i \)th constant associated with \( a_i \). Then the solution for \( x \) (eq.(3.12)) is that

\[ x_i = \tau_i \left( \frac{n}{\xi (k + C)} \right) \]  \hspace{1cm} (7.7)

where

\[ n = \sum_{i=1}^{M} \alpha_i y_i, \quad \xi = \sum_{i=1}^{M} \alpha_i^2, \quad \zeta = \sum_{i=1}^{N} \tau_i^2, \quad k = \lambda / \xi \]

Alternatively the solution \( x \) can be written as

\[ \frac{x_1}{\tau_1} = \frac{x_2}{\tau_2} = \ldots = \frac{x_N}{\tau_N} = \frac{n}{\xi (k + C)} \]  \hspace{1cm} (7.8)

this statement can be proved easily in a similar way as in (a).

(c) It is presumed that case (a) holds provided that all the parameter’s increments are solved in log space, that is \( A = [a_1, a_2, \ldots, a_N] \), \( a_1 = a_2 = \ldots = a_N = a \). The result is then \( x_1 = x_2 = \ldots = x_N = n/(\xi (k + N)) \). If instead the vertical boundary parameters \( p_r, p_{r+2}, \ldots, p_{r+q} \) are in linear space (where \( r \geq 1 \) and \( r+q \leq N \)), and all the other parameters (i.e. resistivity, depth) are in log space then the Jacobian matrix is changed and the result turns out differently. For a model with parameter \( p \), the response function may be defined as

\[ f(p) = (f_1(p), f_2(p), \ldots, f_M(p))^T \]  \hspace{1cm} (7.9)

Thus its Jacobian matrix \( A \) is given by

\[ A = \left( \frac{\partial f_i}{\partial p_j} \right)_{i=1, \ldots, M} \quad \text{\( A \) is a matrix of \( M \) by \( N \), with \( a_{ij} = \partial f_i / \partial p_j \)} \]  \hspace{1cm} (7.10)

The space for the model function \( f_i \) \((i = 1, \ldots, M)\) is independent of the model.
parameters’ space, if the function represents the amplitude it is given in log space, and it is in linear space if it represents the phase. The ratio between increments in these two spaces is not equal to 1 for corresponding changes’ \( \Delta p_j \). That is

\[
\tau_j = \frac{\log p_j^2 - \log p_j^1}{p_j^2 - p_j^1}
\]

is not equal to unity, where \( p_j^2 - p_j^1 \) is that \( \Delta p_j \) in linear space and \( \log p_j^2 - \log p_j^1 \) is that \( \Delta p_j \) in log space. The parameters being converted from the log space to linear space are the vertical boundary parameters. In SI units, the absolute value of the vertical boundary parameters \( |p_j| \) are considered to be larger than 1 meter for general cases.

From the definition (eq.(7.11)) \( \tau_j > 0 \) is obvious, \( \tau_j < 1 \) is proved in appendix 2(a), thus \( 0 < \tau_j < 1 \). Let all the conditions hold as in (a), if only parameters \( p_j \) \( (j=r, \ldots, r+q) \) are in linear space then matrix \( A \) changes to \( A = [a, \ldots, \tau_i a, \ldots, \tau_{r+q} a, a, \ldots, a] \). The solution for the parameter increments solved in linear space is

\[
x_j = \tau_j \frac{n}{\xi (k+\zeta)} \quad (j=r, \ldots, r+q)
\]

and

\[
x_j = \frac{n}{\xi (k+\zeta)} \quad (j=1, \ldots, r-1, r+q+1, \ldots, N)
\]

From eq.(7.12) and eq.(7.13), The solution \( x_j \) with its parameter in linear space is decreased by a factor \( |\tau_j| \) compared to the solution with its parameter in log space.

(d) In (c) it was shown that the log spaced parameter increments in the solution of \( x_j \) decrease by a factor of \( |\tau_j| \). Now it will be shown that when the solution \( x_j \) returns back to the parameters \( p_j \) in linear space there will be another decrease for the parameters whose
Figure 7.4 It is the plot of $y(b)=|10^b-1|/|b|$ for the range $-1 \leq b \leq 1$. The horizontal axis is for $b$ and the vertical axis is for $y(b)$. 
increments are solved in log space.

Let \( x_i = b \), if it has been solved in log space then the step being added to \( p_i \) is \( p_i 10^b - p_i \), if it has been solved in linear space then the step is \( p_i + b - p_i \), the absolute ratio between them will be

\[
\frac{|p_i 10^b - p_i|}{|p_i + b - p_i|} = \frac{|p_i (10^b - 1)|}{|b|}
\]

(7.14)

(i) If \( b > 0 \), and \( p_i \geq 1 \), then \( f > 1 \). This is shown as follows

\[
f = \frac{|p_i (10^b - 1)|}{|b|} \geq \frac{|10^b - 1|}{|b|}
\]

(7.15)

Since the Taylor series for \( 10^b \) is

\[
10^b = 1 + (\ln 10) b + \sum_{n=2}^{\infty} \frac{(\ln 10)^n}{n!} b^n
\]

(7.16)

and \( b > 0 \) so that \( 10^b > 1 + (\ln 10) b \), it means that

\[
f > \frac{(\ln 10) b}{b} = \ln 10 \approx 2.3
\]

(7.17)

(ii) If \( p_i \geq 1 \), \( b < 0 \) but \( b > -0.8 \) (it implies \( |b| < 0.8 \), which has already reached the maximum increment step for the solution obtained from the logarithmic space), then \( f > 1 \). This can be shown as follows

\[
f \geq \frac{|10^b - 1|}{|b|} = \frac{|10^b - 1|}{|b|}
\]

(7.18)

Denote \( f = (10^b - 1)^2 / b^2 \), which is an increasing function of \( b \), that is

\[
\frac{df}{db} = \frac{2(10^b - 1)}{b^3} (1 - 10^b + 10^b b (\ln 10))
\]

(7.19)

Let \( g(b) = 2(10^b - 1) / b^3 \), \( l(b) = 1 - 10^b + 10^b b (\ln 10) \).

\[
\frac{dl}{db} = 10^b b (\ln 10)^2
\]

(7.20)

\( \frac{dl}{db} < 0 \) for \( b < 0 \) and \( \frac{dl}{db} > 0 \) for \( b > 0 \). Since \( l(0) = 0 \) then \( l(b) > 0 \) (\( b < 0 \), \( b > 0 \)).
b>0). For g(b), g(b)>0 for b<0 and b>0. \( \frac{\partial f}{\partial b} = g(b)l(b) > 0 \), thus f'(or f) is an increasing function of b. As \( p_i \geq 1, b \geq -0.8 \), then \( f \geq 1.05 \), it follows that \( f > 1 \) for \(-0.8 < b < 0\). (ii) is then completed.

Moreover, if b comes from the solution obtained in the linear space then the maximum step \( |b| \) can often be larger than 0.8, but since \( |b| < |p_i| \), that is \( |p_i|/|b| > 1 \), then even if \( b < -0.8 \), \( f > 1 \) can still hold.

As \(|b| \to 0, \frac{|10^b - 1|}{|b|} \to \ln 10 \approx 2.3\). The graph of \( \frac{|10^b - 1|}{|b|} \) is shown in Fig. 7.4 for \(-1 \leq b \leq 1\).

Thus from above, either the sensitivity of parameters set in linear space will be reduced, or their increments to the solution are supressed.

7.3. The use of a scaling matrix projecting on the Jacobian matrix to restore the affected solution.

7.3.1. Introduction of the scaling matrix

It is shown above that the linear spaced parameters' solution \( x_j \) has a multiplier \( \tau_j (\tau_j < 1) \) compared to the log spaced solution. In addition, the linear spaced parameter has another extra decrease when \( x_j \) is returned back to \( p_j \) compared with it remaining in log space.

In order to reduce these effects a scaling matrix \( Q \) is introduced. \( Q \) is a diagonal matrix and its ith diagonal value is \( 1/(\tau_j)^{\frac{1}{2}} \). \( Q \) projects on the Jacobian matrix \( A \), hence the ridge regression method which tries to minimize \( (Ax - y)^T(Ax - y) \) now seeks

\[
\text{Min}[ (AQQ^{-1}x - y)^T(AQQ^{-1}x - y)]
\]

and with \( \lambda \) for the stabilization, the solution is
Let all the conditions hold as given in (c). If only $p_j (j=r, \ldots, r+q)$ are in linear space then matrix $A$ in (a) is changed to $A=[a, \ldots, a, \tau_r a, \ldots, \tau_{r+q} a, a, \ldots, a]$, and 
$AQ=[a, \ldots, a, \tau_r \frac{1}{2} a, \ldots, \tau_{r+q} \frac{1}{2} a, a, \ldots, a]$. so that from (b), the solution for $x$ is:

for the parameter increments in linear space

$$x_j=(1/\tau_j)^{\frac{1}{2}}[\frac{\eta}{\xi(k+\zeta)}] \quad (j=r, \ldots, r+q)$$

for the parameter increments in log space

$$x_i=(1/\tau_i)^{\frac{1}{2}}[\frac{\eta}{\xi(k+\zeta)}] \quad (i=r, \ldots, r+q)$$

For the parameters being solved in log space, $\tau_i$ is let equal to $1$ in the diagonal scaling matrix $Q$ and $\eta$, $\xi$ and $k$ are the same as defined in (b) and $\zeta=\sum \tau_i (i=1, \ldots, N)$. With the scaling matrix the parameters in linear space regain their sensitivity. That is

$$x_i=x_j=\frac{\eta}{\xi(k+\zeta)}$$

7.3.2. General application

The special case above is solved by choosing proper elements in the scaling matrix $Q$ which changes the structure of the Jacobian matrix $A$. In reality, the value $\tau_i (i=1)$ is not a constant, and the Jacobian matrix is rather complicated to analyse. As it is mainly required to restore a parameters’ sensitivity when its increments are solved in linear space for the solution, one method for application is given below.

Each increment goes to zero as $\lambda$ goes to $\infty$. Let $L$ be a large number. When $\lambda=L$ each vertical parameter increment value $x_i$ (in linear space) is compared with the average value $\beta_i$ of its neighbouring logarithmically spaced horizontal parameter increments after being converted from log space to linear
space. Let \( r_i \) be the number of such neighbouring horizontal parameters, \( \beta_i = \left[ \sum \Delta p^j \right]/r_i \), for the appropriate \( r_i \) values of \( j \). The scaling value for \( \tau_i \) is increased from a lower bound until
\[
x_i \leq \beta_i \quad (i=r, \ldots, r+q)
\]
is reached. Therefore
\[
p^i_n - p^i_{n-1} \leq \left[ \sum (p^i_n - p^i_{n-1}) \right]/r_i
\]
where the subscript for \( p \) means the present \( n \)th or the former one \((n-1)\)th solution of the inversion, the superscript is the \( i \)th or \( j \)th the element in \( p \). All of these parameters are compared in the same space which is chosen as linear space. If the vertical parameter extends to the earth's surface or to infinity, the divisor of \( \sum (p^i_n - p^i_{n-1}) \) is then reduced to \( r_i - 1 \) (or \( r_i/2 \) if \( r_i = 1 \)), in order to better approximate that vertical boundary parameter's sensitivity for the solution.

As each parameter which is in linear space (vertical boundary) is associated with neighbouring parameters which are in log space, these parameters' sensitivity could represent the vertical boundary's sensitivity to some extent. As depth is a boundary parameter its dimension is suitable for comparison with the vertical boundary parameters, and in addition its sensitivity is comparable since they are neighbouring parameters. Thus its scaled value \( \tau_i \) is chosen where \( x_i \) is just less than or equal to that average value. The reason for choosing \( L \) sufficiently large is that the relationship for general cases becomes much more simple between the scaled and unscaled solution. This is going to be shown in the following section.

### 7.4. Analysis of the ridge regression method when the scaling matrix is applied.

Matrices \( A \), \( Q \), and vector \( y \) are defined the same as above, the solution
vector is denoted by $x'$ if the scaling matrix is applied, and from eq.(7.21)

$$x'=Q\left[ (AQ)^T AQ + \lambda I \right]^{-1} (AQ)^T y = (x_1', x_2', \ldots, x_N')$$  \hspace{1cm} (7.27)

If no scaling matrix is applied the solution vector is denoted by $x$

$$x = (AA^T + \lambda I)^{-1} A^T y = (x_1, x_2, \ldots, x_N)^T$$ \hspace{1cm} (7.28)

(i) If the damping factor $\lambda$ goes to $\infty$, then

$$x' = Q^2 x$$ \hspace{1cm} (7.29)

and so

$$x_i' = \frac{1}{\tau_i} x_i \hspace{1cm} (i=1, \ldots, N)$$ \hspace{1cm} (7.30)

(Appendix 2(b))

This means that if the damping factor is sufficiently large the scaled solution and the unscaled solution are related by a constant multiplier. For $\lambda$ sufficiently large it is clear that the effect produced by parameter increments solved in different spaces can be overcome by choosing proper scaling values $\tau_i$ ($i=1, \ldots, N$).

(ii) If $\tau_i$ goes to 0 which means the scaling matrix effecting the solution is maximum, where $i=r, \ldots, r+q$, then the limit for $x_i'$ is

$$x_i' = x_i + \lambda e_i^T \Theta x$$ \hspace{1cm} (7.31)

and the solution for $x_j$ ($j \neq i$) is

$$x_j' = x_j + \lambda e_j^T \Theta x$$ \hspace{1cm} (7.32)

where $e_i^T, e_j^T$ are vectors of dimension $N$ with the $i$th or $j$th element equal to 1 respectively while others are equal to zero, $\Theta = (A^T A + \lambda Q')^{-1} (I - Q')$ and $Q'$ is a diagonal matrix of dimension $N$ by $N$ with zero values from $r$th to $(r+q)$th diagonal position while other diagonal values are $\tau_j$ ($j=1, \ldots, r-1, r+1, \ldots, N$).

(Appendix 2(c)).

If $\tau_i$ goes to 0, for all $i=1, \ldots, N$, then the limit for $x_i$ is
\[ x'_i = x_i + \lambda e_i^T \Theta x \quad (i=1, \ldots, N) \]  

(7.33)

where \( e_i \) is defined the same as before but

\[ \Theta = \lim_{\tau_i \to 0} \left( (A^T A + \lambda Q^{-2}) \right) \quad (i=1, \ldots, N) \]  

(7.34)

If \((A^T A)^{-1}\) exists then \(\Theta = (A^T A)^{-1}\). (Appendix 2(c)).

This shows that the solution is bounded if all or just some of the \( \tau_i \to 0 \) \((i=1, \ldots, N)\).

(iii) Suppose only one parameter's corresponding column of the Jacobian matrix is scaled or only one parameter is in linear space, namely \( \tau_r \neq 1 \), and \( \tau_i = 1 \) \((i \neq r)\), then the following holds. For the non-scaled parameters' solution \( x'_i \) \((i \neq r)\), they are

\[ x'_i = x_i + \beta \quad (i \neq r) \]  

(7.35)

\( \beta \) is the \( i \)th row and \( r \)th column positioned element in \( B \), where

\[ B = \gamma_r Q(A^T AQ + \lambda Q^{-1})^{-1}(\lambda I + A^T A) x_r \]  

(Appendix 2(d))

Comparing the solution \( x'_r \) (obtained by applying the scaling matrix) with \( x_r \) (no scaling applied), it is found that \( x'_r \) is in the same direction as \( x_r \) but has a larger modulus. That is

\[ \text{if you define } r = \frac{x'_r}{x_r}, \text{ then } r > 0, r > 1 \]

The equality holds if and only if \( \lambda = 0 \). Also \( x'_r = 0 \) if and only if \( x_r = 0 \). (Appendix 2(d)).

This shows that the scaled solution is always larger than the unscaled one in absolute value, that is

\[ |x'_r| > |x_r| \]  

(7.36)

and they stay in the same direction, that is

\[ x'_r x_r \geq 0 \]  

(7.37)

The derivatives of the ratio \( r = \frac{x'_r}{x_r} \) with respect to \( \tau_r, \lambda \) are
The above equality holds if and only if \( \lambda = 0 \) (Appendix 2(e)).

This shows that the ratio between the scaled solution and the unscaled one \( r \) is a decreasing function of \( \tau_r \) and an increasing function of \( \lambda \).

It is very difficult to analyse the same character as in (iii) if the parameter increments being solved for in linear space are more than one. This analysis is left for further work when necessary.

7.5. Conclusion and discussion

Taking the boundaries and resistivities of the minimum blocks as model parameters for 2D inversion enables us to conduct the procedure in a much more convenient and efficient way. This new model parameterization brings a new problem, that is the effect on parameters not in the same space. This effect has been presented by analysing a particular case. As the scaling matrix is applied, the solution is modified so that the effect is controllable.

From eq.(7.11) in section 7.2 (c), as \( \tau_j < 1 \) then those Jacobian columns which correspond to the parameters in linear space (the vertical parameters) may have very low Euclidean lengths compared with others in the same matrix \( A \). Thus they may cause ill-conditioning for the solution and appear to be "unimportant" parameters as being discussed in section 3.4.3., their increment steps obtained by eq.(3.12) can drop dramatically as the damping factor \( \lambda \) increases.

After the scaling matrix is applied, if \((AT^TA)^{-1}\) exists (or \(AT^TA\) is non-singular), eq.(7.21) shows that for \( \lambda = 0 \) (which is the least square solution) the scaling matrix has no effect on the solution. That is \( x' = x = (AT^TA)^{-1}AT^TY \). The scaling matrix is effective on the solution when the ridge regression method is used.
(\lambda>0 \text{ in eq.(3.21)}) \text{ (section 7.4 (i))}. The effect becomes very clear especially when the damping factor is sufficiently large, in which case the relationship between the solution when the scaled matrix has been applied and when it has not is only a division by the scaling value \( \tau_i \) (\( i=1, \ldots, N \)) and for those unscaled parameters, which means their scaling values are equal to 1, their parameter increment solution would not be affected. Since for those vertical parameters the scaling values \( \tau_i \) (\( i=r, \ldots, r+q \)) are given as less than 1 then the solution obtained by applying the scaling matrix is increased in length, thus the parameters' sensitivity is increased. Moreover, behaviour like that of "unimportant" parameters can be removed.

In section 7.4., (i) and (ii) show that the solutions are bounded when either the damping factor goes to infinity or the scaling values go to zero. In section 7.4. (iii), it is shown that if only one parameter is scaled, the solution for this parameter is a decreasing function of the scaling value \( \tau_r \) (as only the \( r \)th parameter is scaled, \( \tau_{i}=1, i\neq r \)) and an increasing function of the damping factor \( \lambda \), but it is bounded in both cases (section 7.4. (i), (ii)).

One way of choosing the scaling values \( \{\tau_j, j=r, r+q\} \) of the scaling matrix \( Q=\text{diag}(1, \ldots, 1/\tau_r^\frac{1}{2}, \ldots, 1/\tau_{r+q}^\frac{1}{2}, 1, \ldots, 1) \) has been introduced, which improves the vertical boundaries' sensitivity reasonably. That is the vertical parameter is restored to its proper sensitivity by comparing it with its neighboring horizontal parameters in order to choose the proper scaling values in the applied scaling matrix. The example presented in section 7.1. used the scaling matrix and the scaling values are chosen in the way presented by section 7.3. A better way may be found to deal with this problem after more research. This discussion and analysis about the affect on the solution of the scaling matrix outlines the case. The advantage of using the scaling matrix
enable the parameter's sensitivity.
8.1. Model description and data generation

To illustrate the modelling method and the present inversion techniques, a difficult two dimensional case is to be modelled, which is taken from the COMMEMI project models (as has been explained in chapter 1). Fig. 8.1 is the model (labelled as 3-2a in the COMMEMI project), the surface layer has a low resistivity along both sides and a comparative high resistivity in between, which is connected through the whole area from the depth 10 Km to 30 Km, and is then underlain by a conductive half space.

Our two dimensional forward modelling program originally came from a copy of Brewitt-Talor and Weaver's finite difference program. Since then some alterations have been made (e.g. as detailed in chapter 4). It still agrees, in the main, with the one in use by Weaver which has been recommended as one of the best programs for 2-dimension modelling in respect of the accuracy (Zhdanov and Varentsov, 1984; Weaver, 1986)

In order to use this model as an inversion example, data at eight sites for 15 periods have been generated. The sites are located at (Fig. 8.1) -35km, -25km, -15km, 5km, 15km, 25km, 35km. The periods are selected as T (seconds)= 3.0, 5.6, 10.0, 18.0, 30.0, 56.0, 100.0, 180.0, 300.0, 560.0, 1000.0, 3000, 18000, 56000, 100000. The spacing between each two sites is 10 kilometers, and the spacing between each two periods is about one third of a decade.

The field data are calculated artificially simply by running the two dimension forward modelling program. After that 10% random noise was added to the apparent resistivity and 5 degree random noise to the phase.
Figure 8.1 A two dimensional model from the COMMEMI project (labelled there as 3-2a). R indicates the resistivity(Ωm), Y indicates the vertical boundary(Km) and Z indicates the depth(Km). The location of sites 1, ..., 8 are indicated by asterisks above the model.
The data possess both E and H polarizations, and both apparent resistivity and phase. These form the response functions for inversion.

8.2. The construction of the initial 2D model and its first iteration

The inversion procedure starts with a one dimensional inversion at each site for both E and H polarization separately. If the resulting one dimension inversion models derived from these two polarizations do not differ too much then a combination of the models is taken. If they do, then their invariant one dimension inversion result is taken. The data sets used in their invariant inversion is the averaged E and H polarization data sets. It can be given as (Kaufman and Keller, 1981)

\[ \rho_{a}^{av} = (\rho_{a}^{E} \rho_{a}^{H})^{\frac{1}{2}} \]
\[ \phi_{a}^{av} = (\phi_{a}^{E} \phi_{a}^{H})/2 \]

where \( \rho_{a}^{E} \), \( \phi_{a}^{E} \) are the apparent resistivity and phase response functions of the field model for E polarization, \( \rho_{a}^{H} \) and \( \phi_{a}^{H} \) are for H polarization, \( \rho_{a}^{av} \) and \( \phi_{a}^{av} \) are for the average of the two polarizations.

The one dimension inversion result at each site is shown in Fig. 8.2. Combining each site’s one dimension model gives the initial two dimension structure of the region (Fig. 8.3). Intermediate vertical lines are set between sites with substantially different resistivities, and neighboring blocks whose resistivities are very close are averaged and connected. The initial fit to the response function is shown graphically in Fig. 8.4.

Following this initial two dimensional model, the two dimension inversion proceeds by using the ridge regression method which has been described in chapter 3. The initial misfit (the mean sum of the square of the misfit \( M_s \) defined by eq.(6.5)) which considers the fit of the response of the initial model (Fig. 8.3) to the field data is \( M_s = 4.13 \). The inversion is carried out with the
<table>
<thead>
<tr>
<th>Site</th>
<th>Depth (km)</th>
<th>Horizontal Dimension (km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.00</td>
<td>-42.00 -34.00 -26.00 -18.00 -10.00 -2.00</td>
</tr>
<tr>
<td>2</td>
<td>0.07</td>
<td>-42.00 -34.00 -26.00 -18.00 -10.00 -2.00</td>
</tr>
<tr>
<td>3</td>
<td>0.151</td>
<td>-42.00 -34.00 -26.00 -18.00 -10.00 -2.00</td>
</tr>
<tr>
<td>4</td>
<td>0.09</td>
<td>-42.00 -34.00 -26.00 -18.00 -10.00 -2.00</td>
</tr>
<tr>
<td>5</td>
<td>0.084</td>
<td>-42.00 -34.00 -26.00 -18.00 -10.00 -2.00</td>
</tr>
<tr>
<td>6</td>
<td>0.088</td>
<td>-42.00 -34.00 -26.00 -18.00 -10.00 -2.00</td>
</tr>
<tr>
<td>7</td>
<td>0.036</td>
<td>-42.00 -34.00 -26.00 -18.00 -10.00 -2.00</td>
</tr>
<tr>
<td>8</td>
<td>0.278</td>
<td>-42.00 -34.00 -26.00 -18.00 -10.00 -2.00</td>
</tr>
</tbody>
</table>

Figure 8.2 The one dimensional inversion result at each site. The line beneath each site is the depth boundary between layers, the number is the layer's resistivity in $\Omega m$. 
Figure 8.3 The initial two dimensional model. R, Y, Z are defined as in caption 8.1. The thick lines are those vertical parameters being taken as free parameters. The tick marks in these vertical lines show where the line has been broken into separate free parameters.
Figure 8.4 Comparison of the response of the initial model in Fig. 8.3 to the given response function data. The former is represented by a cross and the latter by a circle. The vertical bar at each point is the assigned error bar of that data sample point. $E_P$ is the abbreviation for E polarization and $H_P$ is for H polarization.
MODEL AMPLITUDE RESPONSE

SITE: 4, HLP

MODEL PHASE RESPONSE

SITE: 4, HLP

MODEL AMPLITUDE RESPONSE

SITE: 5, HLP

MODEL PHASE RESPONSE

SITE: 5, HLP

MODEL AMPLITUDE RESPONSE

SITE: 6, HLP

MODEL PHASE RESPONSE

SITE: 6, HLP
MODEL AMPLITUDE RESPONSE

SITE: 7, H.P.

MODEL PHASE RESPONSE

SITE: 7, H.P.

MODEL AMPLITUDE RESPONSE

SITE: 8, H.P.

MODEL PHASE RESPONSE

SITE: 8, H.P.
vertical boundaries which connect the surface to the second layer fixed, since these parameters are extremely non-linear for H-polarization (Ingham, 1981). All the resistivity and depth parameters, and most of the vertical parameters (thick lines, Fig. 8.3) are considered as free parameters. After the first iteration, the misfit has been reduced to \( M_s = 2.95 \), but further iteration yields no improvement in the misfit if the present parameterization is retained.

8.3. Escaping from the local minimum

When no further improvement occurs, the question will be whether the inversion has stopped at a local minimum or a global minimum. It will be difficult to judge for the field case if no other information is available. But in our case since it is a test for two dimension inversion modelling based on a known structure, it is clear that the inversion has sunk to a local minimum.

The next question is whether this phenomenon is due to noise in the data or to the parameterization. As a test, the same procedure is carried out by using noise free data, and the same phenomenon occurs. Thus the problem might come from the model structure itself with the present parameterization.

As one possible way of moving the inversion from its local minimum, the model is parameterized with more resistivity parameters simply by dividing some big blocks into two or four small blocks (Fig. 8.5). The boundaries of all blocks are not free parameters but remain fixed at this stage. The reasoning is that this will provide an opportunity for a more complicated two dimension model to occur giving great flexibility for fitting the data.

With this new parameterization, the next iteration does converge (noisy data being used from here on) and the misfit has been reduced further to
Figure 8.5 The solid lines represent the model resulting from the first iteration. R, Y, Z are defined as in caption 8.1. The dashed lines indicate where blocks have been divided into some smaller blocks for additional parameterisation.
$M_s=1.38$.

8.4. The result of further iterations

Following the above, another three iterations were conducted with the same parameterization, the model slowly converged and the misfit reduced to $M_s=1.24$. The model has changed to a more complicated structure. Some newly divided blocks became more resistant, others became more conductive (Fig. 8.6).

After the above iteration, the new resistivity values indicate further boundaries that may be used as parameters. The next four iterations include these extra boundaries as parameters in addition to other resistivity and boundary parameters (see Fig. 8.6). The model derived from these four iterations is shown in Fig. 8.7, the misfit being $M_s=1$. This shows the model has converged satisfactorily. Comparing Fig. 8.6 to Fig. 8.7 it is seen that the model boundaries have been moved substantially in adjusting the model. This example shows the necessity to consider boundary parameters as free parameters in the inversion. Because in Fig. 8.7, some neighboring blocks have very close values in resistivity, then those blocks (grouped by the thick lines) can be combined using the smoothing method introduced in chapter 6. Following this smoothed and simplified model, another two iterations were carried out and the model is given by Fig. 8.8, where $M_s=1$.

Fig. 8.8 shows satisfactory convergence, but further simplification can be done to the model by applying the smoothing technique once more, which gives a more simple model as shown in Fig. 8.9, whilst the misfit has been hardly increased ($M_s=1$). Graphs showing the fit of the model in Fig. 8.9 to the field data are given in Fig. 8.10.
as free parameters which have distinctive neighbouring resistivity values.

Figure 8.6 R, Y, Z are defined as in caption 8.1. The dashed lines are those

Horiztonal Dimension (km)
Figure 8.7 Four iterations on from the model of Fig. 8.6. $R, Y, Z$ are defined as in caption 8.1. The thick lines outline the grouped blocks.
Figure 8.8 The model produced by two further iterations. R, Y, Z are defined as in caption 8.1. Again the thick lines outline grouped blocks.
Figure 8.9 The final model. R, Y, Z are defined as in caption 8.1.
Figure 8.10 Comparison of the response from the final model in Fig. 8.9 to the given response function data. The former is represented by cross and the latter by circle. The vertical bar at each point is the error bar of that data sample point. E_P is the abbreviation for E polarization and H_P is for H polarization.
MODEL AMPLITUDE RESPONSE
SITE: 4, E_P

MODEL PHASE RESPONSE
SITE: 4, E_P

MODEL AMPLITUDE RESPONSE
SITE: 5, E_P

MODEL PHASE RESPONSE
SITE: 5, E_P

MODEL AMPLITUDE RESPONSE
SITE: 6, E_P

MODEL PHASE RESPONSE
SITE: 6, E_P

139
MODEL AMPLITUDE RESPONSE

SITE: 8, E_P

AMPLITUDE (GM/H)

LOG PERIOD (SEC.)

MODEL PHASE RESPONSE

SITE: 8, E_P

PHASE (DEG.)

LOG PERIOD (SEC.)

MODEL AMPLITUDE RESPONSE

SITE: 7, E_P

AMPLITUDE (GM/H)

LOG PERIOD (SEC.)

MODEL PHASE RESPONSE

SITE: 7, E_P

PHASE (DEG.)

LOG PERIOD (SEC.)
8.5. The resolution study for the model

The resolution study for this model uses the modified formulation described in section 5.2.2 (chapter 5), which gives the permissible range of each declared parameter while all other parameters are fixed to a given bound for the given misfit (the sum of the squares of the misfit defined as eq.(5.1)).

Since the resolution study requires a linear approximation for the studied parameters, only the resistivity parameters are used for the resolution study (Table 8.1). In this study, the sum of the squares of the misfit (eq.(5.1)) is perturbed by 10% with respect to that of the final model and the changes for the model parameters are bounded by 10% with respect to their values in the final model. From Table 8.1, it is shown that all the these resistivity parameters are well resolved.

8.6. Conclusion and discussion

The COMMEMI model has a conductive surface along both sides and an even more conductive basement. They are separated by a layer of thickness 20 km of intermediate conductivity which is thus difficult to detect from surface measurements. Indeed the COMMEMI model was chosen with this difficulty in mind.

The one dimension inversion at each site using both E and H polarization data separately gives information that enables an initial two dimension model to be constructed. The subsequent two dimension modelling illustrates a case in which the inversion reaches a local minimum due to improper parameterization. An alternative parameterization which gives the model inversion more flexibility is one way of escaping from the local minimum enabling the inversion iterations to be continued.
Table 8.1

<table>
<thead>
<tr>
<th></th>
<th>R_1</th>
<th>R_2</th>
<th>R_3</th>
<th>R_4</th>
<th>R_5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max.</td>
<td>11.47</td>
<td>107.74</td>
<td>0.13</td>
<td>1.10</td>
<td>11.62</td>
</tr>
<tr>
<td>Min.</td>
<td>9.51</td>
<td>90.19</td>
<td>0.07</td>
<td>0.93</td>
<td>9.71</td>
</tr>
</tbody>
</table>

Table 8.1 Resolution for the resistivity block study. R_1, R_2, ..., R_5 are the blocks presented at fig. 8.9.
This model study shows the importance and necessity of considering boundaries as free parameters, which enables the inversion to be conducted in a much more comprehensive and efficient way. The problem of parameters in different spaces within the inversion is treated by scaling the vertical boundary parameters properly, this has been solved and fully discussed in chapter 7. This inversion modelling shows the advantage of applying the smoothing technique in simplifying the resulting model.

From the initial two dimension model (Fig. 8.2) to the final model (Fig. 8.9) only 10 iterations are carried out and the misfit has been reduced from $M_s=4.13$ to $M_s=1$. Following the 10th iteration, the inversion hardly shows any sign in improving the misfit thus the inversion stopped at the 10th iteration as $M_s=1$ already gives a satisfactory convergence.

Fig. 8.2 shows that the intermediate layers (where $R=53.3, R=61.5$) are sandwiched between quite conductive zones, so that their resistivity values may be artificially depressed. The block ($R=299.9$) may be recognized as a kind of $\delta$-model formulation. Considering these facts the model is then smoothed by grouping all horizontal neighbouring blocks in this region (from left to right $R=53.3, R=299.6, R=101.7, R=61.5$). A simplified model as Fig. 8.3 is then obtained. However, it is still quite subjective, if some other information about the structure is available then a preferable model may be determined from Fig. 8.3 and Fig. 8.9.
9.1. Introduction

As a test of the modelling methods developed in former chapters, an application to field data is made. The data were obtained from northern Sardinia and are part of a larger scale magnetotelluric investigation. As described by Schnegg et al. (1987), in Sardinia (Fig. 9.1), one of the most striking geological features is the Campidano Graben. There is yet no consensus as to the origin of the formation of this feature. Some see it as the result of middle Pliocene to Quaternary rifting (e.g., Cherchi and Montadert, 1985), others as the remains of a subduction trench where the NE block plunged under the SW part. This suggestion stems from the geochemical character of the volcanism observed, composed of andesitic tuffs and pyroclastic lavas interpersed with ignimbrites (Cocozza and Jacobacci, 1975).

Several MT profiles have already been studied by the Observatoire Cantonal (Schnegg, Fischer, Quang and Ranieri, 1987). As a further part of an extensive campaign over Sardinia, another MT profile has been taken recently at locations Budduso, Monti, Calangianus, Riu Piatu and Luogosanto, which are presented as sites H, I, J, K and L in Fig. 9.1. These five sites are not perpendicular to the assumed geological strike in the north, but were provided from the Sardinia data only as example of an approximate two-dimensional profile. Detailed geological conclusions contributing to the tectonics of the region would require the analysis of further data.

Data from the five stations were collected by Fischer (1988, personal communication). As he suggested the profile is two dimensional. Combining this with geological information, the strike is taken in the direction of N 50° E. E-polarization data
Figure 9.1 Geological sketch map of Sardinia. (1) Quaternary sediments, (2) Miocene deposits, (3) Miocene delta sand, (4) Tertiary basalt, andeite, thyodacite, ignimbrite and tuff, (5) Mesozoic limestone and dolomite, (6) Paleozoic graniticschistose basement, (7) fault, (8) borehole, (9) thermal spring (redrawn by Schnegg, etr., 1987, after Pala et al., 1977). Circled points are the MT stations for the present profile, they are denoted by H, I, J, K, L from south to north.
cannot be used, since they are affected by a 3-D environment caused by the surrounding sea which is a relatively good conductor. Stations I to L are particularly affected. E-polarization apparent resistivity values are extremely large, because of the large current which wants to link the two parts of the sea and must do so by bridging over the land (a poor conductor). As a consequence it is the H-polarization data that should and can yield the correct structure upon 2-D modelling. Therefore only H-polarization data (both apparent resistivity and phase) will be used for our modelling.

The profile consists of five sites (Fig. 9.1, Fig. 9.2) located at 2.0km, 26.7km, 40.03km, 48.68km, 61.96km from Fischer's origin (personal communication). It has 11 periods, they are: (sec.) 0.003, 0.01, 0.03, 0.1, 0.3, 1.0, 3.0, 10.0, 30.0, 100.0, 300.0.

9.2. The inversion procedure

The inversion process starts with a determination of each site's 1D inversion model. This has already been provided by G. Fischer (personal communication). Following his work, the initial 2D model is set up by combining these 1D inversion models at each site. Between sites an intermediate vertical boundary is set (Fig. 9.2) and the mean sum of squares of the misfit ($M_s$ defined by eq.(6.5)) for this initial 2D model is $M_s=49$. The formula for the misfit $M_s$ is given as eq.(6.5) in chapter 6 and Fig. 9.3 shows the fit of the response of this initial model to the field data. The grids for the forward modelling are given in Table 9.1.

The first iteration of the inversion considered each resistivity, vertical boundary and horizontal boundary as free parameters. This iteration reduced the misfit to $M_s=2.74$. The vertical boundary parameters hardly move at all, which might be due to the large distance between any two sites and the use
Figure 9.2 The initial 2D model constructed from 1D inversions at each site. The sites are H, I, J, K, L located at 2.0km, 26.7km, 40.03km, 48.68km, 61.96km (marked by asterisks above the earth surface). The earth surface is taken as Z=0; the line has not been drawn since the surface layer is very thin. The misfit for this model is $M_s=49$. In the model, Z indicates depth(km), and R indicates resistivity(Ωm). The vertical lines are located at 14km, 33km, 44km and 56km, and these separate the 5 sites.
**Figure 9.3** Comparison of the response from the initial 2D model(Fig. 9.2) with the field data, which are represented by amplitude and phase at each site (H,I,J,K,L) with eleven periods. Circles represent the field data crosses represents the initial model responses and vertical bars indicate standard errors.
Table 9.1

The grids been used for the first four iterations, Ygrid gives the horizontal grids, and Zgrid gives the vertical grids as it is a 2D model.

Table 9.1

<table>
<thead>
<tr>
<th>YGRID(km)</th>
<th>-500.0</th>
<th>-200.0</th>
<th>-50.0</th>
<th>-10.0</th>
<th>-5.0</th>
<th>-2.0</th>
<th>0.0</th>
<th>1.2</th>
<th>1.8</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2</td>
<td>2.5</td>
<td>4.0</td>
<td>6.0</td>
<td>8.0</td>
<td>10.0</td>
<td>12.0</td>
<td>13.5</td>
<td>13.95</td>
<td>14.0</td>
<td></td>
</tr>
<tr>
<td>14.05</td>
<td>14.5</td>
<td>16.0</td>
<td>18.0</td>
<td>21.0</td>
<td>23.0</td>
<td>24.0</td>
<td>24.5</td>
<td>26.0</td>
<td>26.4</td>
<td></td>
</tr>
<tr>
<td>26.7</td>
<td>27.0</td>
<td>27.5</td>
<td>28.5</td>
<td>30.0</td>
<td>32.0</td>
<td>32.5</td>
<td>32.95</td>
<td>33.0</td>
<td>33.05</td>
<td></td>
</tr>
<tr>
<td>33.5</td>
<td>34.0</td>
<td>36.5</td>
<td>38.0</td>
<td>39.0</td>
<td>39.7</td>
<td>40.03</td>
<td>40.3</td>
<td>41.0</td>
<td>42.2</td>
<td></td>
</tr>
<tr>
<td>43.5</td>
<td>43.95</td>
<td>44.0</td>
<td>44.05</td>
<td>44.5</td>
<td>45.0</td>
<td>46.0</td>
<td>47.0</td>
<td>48.0</td>
<td>48.68</td>
<td></td>
</tr>
<tr>
<td>49.5</td>
<td>50.5</td>
<td>51.5</td>
<td>53.0</td>
<td>55.5</td>
<td>55.95</td>
<td>56.0</td>
<td>56.05</td>
<td>56.5</td>
<td>58.0</td>
<td></td>
</tr>
<tr>
<td>60.0</td>
<td>61.3</td>
<td>61.7</td>
<td>61.96</td>
<td>62.3</td>
<td>62.7</td>
<td>63.3</td>
<td>70.0</td>
<td>71.0</td>
<td>74.0</td>
<td></td>
</tr>
<tr>
<td>80.0</td>
<td>130.0</td>
<td>300.0</td>
<td>600.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ZGRID(km)</th>
<th>-0.1</th>
<th>0.0</th>
<th>0.0625</th>
<th>0.1</th>
<th>0.14</th>
<th>0.19</th>
<th>0.22</th>
<th>0.46</th>
<th>0.51</th>
<th>0.63</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.86</td>
<td>0.95</td>
<td>1.5</td>
<td>2.7</td>
<td>3.3</td>
<td>4.1</td>
<td>5.5</td>
<td>8.0</td>
<td>9.5</td>
<td>10.67</td>
<td></td>
</tr>
<tr>
<td>11.46</td>
<td>12.85</td>
<td>12.88</td>
<td>13.7</td>
<td>17.0</td>
<td>27.0</td>
<td>30.63</td>
<td>33.27</td>
<td>36.0</td>
<td>40.0</td>
<td></td>
</tr>
<tr>
<td>42.17</td>
<td>43.63</td>
<td>47.0</td>
<td>52.0</td>
<td>60.0</td>
<td>70.0</td>
<td>85.0</td>
<td>110.0</td>
<td>140.0</td>
<td>220.0</td>
<td></td>
</tr>
<tr>
<td>350.0</td>
<td>500.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
of only one polarization (H polarization). Thus the following iterations of inversion are carried out with the vertical boundaries fixed. The second iteration of the inversion reduces the misfit to $M = 1.46$, and the third iteration gives $M = 1.267$. The resulting model then shows that the third layer at site J becomes negligibly thin, and the bottom layer at sites H and J can be linked as one parameter since their resistivity values becomes very close.

The fourth iteration reduces the misfit slightly to $M = 1.256$. Before the next iteration starts the grids are refined (Table 9.2) and adjusted to suit the present model.

In addition to that, the number of parameters is reduced by combining some of the neighbouring parameters together as their values become very close. This adjustment to the model increased the misfit slightly to $M = 1.38$. The next iteration (the fifth) reduces it to $M = 1.32$.

### 9.3. The misfit at site L and relevant parameter adjustments

Further iterations do not improve the misfit with the above parameterization. The resulting model and the fit of its response to the field data after five iterations of inversion are shown in figures 9.4 and 9.5. Fig. 9.5 shows that the fit of both amplitude and phase at sites H, I, J are quite acceptable, the fit at site K needs some improvement and the fit of the phase curve at site L needs considerable improvement.

One possible way to overcome this problem is to allow a more complicated model to occur through a more detailed parameterization. This is done by dividing those resistivity blocks except the surface layers (which are already very thin) under sites K or L into regions 5–10km thick. The inversion is then carried out with all other parameters fixed and only the resistivities of
Table 9.2

<table>
<thead>
<tr>
<th>YGRID(km)</th>
<th>-500.0</th>
<th>-200.0</th>
<th>-50.0</th>
<th>-10.0</th>
<th>-5.0</th>
<th>-2.0</th>
<th>0.0</th>
<th>1.8</th>
<th>2.0</th>
<th>2.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.0</td>
<td>6.0</td>
<td>8.0</td>
<td>10.0</td>
<td>12.0</td>
<td>13.0</td>
<td>13.95</td>
<td>14.0</td>
<td>14.05</td>
<td>15.0</td>
<td></td>
</tr>
<tr>
<td>16.0</td>
<td>18.0</td>
<td>21.0</td>
<td>23.0</td>
<td>25.0</td>
<td>26.0</td>
<td>26.7</td>
<td>27.5</td>
<td>29.0</td>
<td>30.0</td>
<td></td>
</tr>
<tr>
<td>32.0</td>
<td>32.95</td>
<td>33.0</td>
<td>33.05</td>
<td>34.0</td>
<td>35.5</td>
<td>37.5</td>
<td>39.0</td>
<td>40.03</td>
<td>41.5</td>
<td></td>
</tr>
<tr>
<td>43.0</td>
<td>43.95</td>
<td>44.0</td>
<td>44.1</td>
<td>48.68</td>
<td>52.5</td>
<td>55.9</td>
<td>56.0</td>
<td>56.1</td>
<td>57.0</td>
<td></td>
</tr>
<tr>
<td>58.0</td>
<td>59.0</td>
<td>60.0</td>
<td>61.0</td>
<td>61.96</td>
<td>63.0</td>
<td>64.0</td>
<td>65.0</td>
<td>66.0</td>
<td>68.0</td>
<td></td>
</tr>
<tr>
<td>70.5</td>
<td>74.0</td>
<td>80.0</td>
<td>100.0</td>
<td>150.0</td>
<td>300.0</td>
<td>600.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ZGRID(km)</th>
<th>-0.1</th>
<th>0.0</th>
<th>0.0625</th>
<th>0.1</th>
<th>0.14</th>
<th>0.207</th>
<th>0.24</th>
<th>0.478</th>
<th>0.626</th>
<th>0.72</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0</td>
<td>6.0</td>
<td>10.27</td>
<td>11.644</td>
<td>12.5</td>
<td>15.5</td>
<td>16.5</td>
<td>17.5</td>
<td>20.5</td>
<td>24.0</td>
<td></td>
</tr>
<tr>
<td>25.2</td>
<td>26.0</td>
<td>27.0</td>
<td>30.5</td>
<td>31.2</td>
<td>32.0</td>
<td>36.0</td>
<td>38.0</td>
<td>40.0</td>
<td>43.0</td>
<td></td>
</tr>
<tr>
<td>46.0</td>
<td>50.0</td>
<td>55.0</td>
<td>60.0</td>
<td>65.0</td>
<td>70.0</td>
<td>85.0</td>
<td>90.0</td>
<td>95.0</td>
<td>100.0</td>
<td></td>
</tr>
<tr>
<td>120.0</td>
<td>140.0</td>
<td>160.0</td>
<td>180.0</td>
<td>200.0</td>
<td>220.0</td>
<td>240.0</td>
<td>260.0</td>
<td>280.0</td>
<td>300.0</td>
<td></td>
</tr>
<tr>
<td>320.0</td>
<td>340.0</td>
<td>370.0</td>
<td>400.0</td>
<td>600.0</td>
<td>1000</td>
<td>3000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 9.2 As for table 9.1, it is the grids used for the rest of the iterations.
Figure 9.4 The resulting model after five iterations of inversion. The misfit has been reduced to $M_s=1.32$. Symbols as in caption to Fig. 9.2.
Figure 9.5 Comparison of the response of the model in Fig. 9.4 with the field data. Symbols as in caption to Fig. 9.3.
these layers as free parameters. This hardly shows any effect in improving the misfit at either sites K and L. This indicates that an even more complicated structure may exist between sites K and L. To allow for this more free parameters are provided between these two sites. Between sites K and L more parameters are provided by adding vertical boundaries half way between the original intermediate vertical boundary and sites K, L. The inversion is carried out by taking the resistivity of the newly appearing blocks as free parameters while all other parameters are fixed. These free parameters are denoted by B1, B2, ..., S3, S4 (Fig. 9.6) where each letter refers to the original blocks of Fig. 9.4. Each block's initial resistivity value is given as the value resulting from the fifth iteration at the corresponding location (refering to Fig. 9.4, Fig. 9.6, B1=B2=B3=B4=20988.5, 01=02=03=04=05=06=07=08=165.3, X1=X2=X3=X4=3510.9, E1=E2=E3=E4=E5=E6=4237.6, G1=G2=392.4, S1=S2=S3=S4=0.102). The following iterations are carried out with these 28 block resistivity parameters as free parameters only, the misfit shows an improvement as the iterations are carried out, with an obvious improvement in the fit at sites K and L which is our main concern.

9.4. Further inversion procedure and the model smoothing

After another six iterations the misfit is reduced to $M_5=0.56$, and similar conductivity blocks can be joined into 9 groups with distinctive conductivity values (Fig. 9.7). They are (refering to Fig. 9.6): group1(B1, B2, B4, O1), group2(E1, E2, O5, O6, B3), group3(O2, O3, O4, O8), group4(X1, X2, X4), group5(X3), group6(O7, E3, E4, E6, G1), group7(E5), group8(G2), group9(S1,S2,S3,S4). Each group is outlined in Fig. 9.7 by a thick line. Within these groups, there are still some differences between the resistivity values. Also there exist delta-like combinations between some of the small blocks (e.g. blocks B3, O1, O7, X3, E3, E6, G1 are surrounded wholly or partially by
Figure 9.6 Part of the model in Fig. 9.5, showing only sites J, K and L. Some layers below sites K and L are divided into smaller blocks by the dashed lines.
Figure 9.7 The resulting resistivity values of the blocks in Fig. 9.6 after another 6 iterations. These can be considered as forming several groups (outlined by thick lines).
comparatively low resistivity values). The contrast is not as strong as the example presented in chapter 6 since the blocks given here are much bigger. To achieve a simple 2D model, the smoothing technique presented in chapter 6 is applied. The result is given by Fig. 9.8 while the misfit is hardly affected (Ms increases from 0.56 to 0.58). After the model is smoothed if in the same group some blocks' resistivity values become close to each other then they are combined as one parameter by taking their average value, if not they remain as free parameters in the following iterations. This smoothing technique and combinations of parameters to simplify the model is used whenever necessary as the following iterations are carried out. After several more iterations some of the resistivity values show a close connection with their neighbours namely the surface layers at site L, the bottom layer at site J. These connected parameters are taken as additional free parameters. After another 4 iterations of inversion, the final model is showed in Fig. 9.9, and the fit of the response of the final model to the field data is shown in Fig. 9.10. The misfit has been reduced to $M_s = 0.44$. Both the misfit and the graph in Fig. 9.10 show satisfactory model convergence.

9.5. The resolution study

The area beneath the sites K and L appears to be rather complicated. The resistivity blocks named as $R_1, \ldots, R_7$ are used for studying resolution with respect to the final model using the formulation presented by section 5.2.2 in chapter 5. The sum of the squares of the misfit (eq.(5.1)) (or the given bound for the misfit) is perturbed by 10% with respect to the one of the final model and the change for those present non- extremizing model parameters are bounded by 10% with respect to their values in the final model. The result of this study is given in table 9.3. With reference to Fig. 9.9, Table 9.3 shows that block $R_8$ is poorly resolved. The reason is that this block is underlying a
Figure 9.8 The result after smoothing each parameter within its group with its adjacent parameters (using Fig. 9.7).
Figure 9.9 The final model after fifteen iterations of inversion. The misfit has been reduced to $M_s = 0.44$. $Z$ indicates depth(km), and $R$ indicates resistivity(Ωm). The vertical lines are located (from left to right) at 14km, 33km, 44km, 52.5km, 56km and 59km. $R_1$ to $R_7$ are the blocks being considered in their resolution study.
Figure 9.10 Comparison of the response of the final model with the field data.

Symbols as in caption to Fig. 9.3.
Table 9.3

<table>
<thead>
<tr>
<th></th>
<th>R₁</th>
<th>R₂</th>
<th>R₃</th>
<th>R₄</th>
<th>R₅</th>
<th>R₆</th>
<th>R₇</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max.</td>
<td>2506.09</td>
<td>176.23</td>
<td>19512.03</td>
<td>4.93</td>
<td>599.30</td>
<td>6609.86</td>
<td>2663.29</td>
</tr>
<tr>
<td>Min.</td>
<td>1657.20</td>
<td>137.15</td>
<td>11648.58</td>
<td>0.20</td>
<td>237.87</td>
<td>386.22</td>
<td>1755.27</td>
</tr>
</tbody>
</table>

Table 9.3 The resolution study for the blocks named as R₁, ..., R₇ in Fig. 9.9. Max., Min. are the maximum and minimum extremes of the parameters with respect to the misfit fixed to certain value, which is 10% perturbation to the misfit of the resulting model. Each parameter is extremed while others are bounded by 10% maximum change with respect to their already existing solution.
very conductive area (i.e. it is covered by block R₄). Blocks R₁, R₂, R₄, R₇ seem to be well resolved, blocks R₃ and R₅ are reasonably well resolved.

9.6. Geological interpretation and conclusion

In conjunction with Fig. 9.1, the resulting model (Fig. 9.9) shows that in the south, where sites H, I, J are located, the resistant upper crust is covered by a thin conductive surface layer and underlain by a conductive lower crust. In the north, where site K is located there is a thin resistant surface layer which is underlain by a highly resistant upper crust and beneath site L there is a thin comparative conductive surface layer which is underlain by a resistant upper crust. A more detailed geological discussion requires the analysis of further data.

The forward program and the inversion program used here are the ones described in the former chapters. The fast forward 2D program gives a tremendous advantage here for generating the Jacobian information for the inversion. Since we have not found an effective way of measuring $\|A^T \delta A\|$ computationally yet, then it is more practical to simply run the original program in those cases when the fast program causes divergences. This occurs in less than 10% of the cases.

Using the present computer system (AMDAHL 470V/8 with 16 Mb main memory) in the above forward modelling, each period takes $T₁=40$ CPU sec.. If the program has not been speeded up the required CPU times for generating the Jacobian information for each iteration of the inversion would be $T₂=T₁$ times number of periods times (number of parameters + 1). Let the number of parameters be 28, then for 11 periods, $T₂=12760$ CPU sec.. If the number of
parameters is 9, then T2=4400 CPU sec. About 10 iterations are carried out with an average number of 28 parameters and 5 iterations with 9 parameters among the total fifteen iterations. The total CPU time for using the original forward program to produce the Jacobian information will be 10 x 12760 + 5 x 4400 = 149600 (CPU sec.).

When using the fast forward program then only 200 CPU sec. is needed to produce the Jacobian information for 28 parameters for each period. Thus 11 periods require 2200 CPU sec. About 9% are reruns which need an additional 1120 CPU sec. The total time requiring for producing the Jacobian information for 11 periods with 28 parameters is 3320 CPU sec. The same analysis for the 9 parameters, which need about 90 CPU for each period, shows that 11 periods take 990 CPU sec. 9% reruns take an additional 400 CPU sec. In total 1390 CPU sec. are required for 11 periods with 9 parameters. For all the fifteen iterations, the CPU time required to generate the Jacobian information will be 10 x 3320 + 5 x 1390 = 40150 when using the modified forward program described by chapter 4. Approximately 70% CPU time has been saved by using the speeded up program.

In this field case modelling, the ridge regression method used for MT 2D inversion study has been shown to be successful and the smoothing technique introduced in chapter 6 was found to be both necessary and successful.
CHAPTER 10

CONCLUSIONS AND SUGGESTIONS

10.1. Conclusions

From the former chapters' analysis and modelling we can draw the following conclusions.

1. The modified fast 2D forward program is a very efficient way to supply the required Jacobian information. In contrast to Rodi's first order approximation and the original calculation for the Jacobian information the iterative scheme gives us a trade off between the speed gained and the accuracy obtained. The analysis and discussion of the convergence of this iterative scheme can be a very useful guide in understanding this method.

2. The ridge regression method used for inversion has been shown to be very successful. The proper choice of the damping factor is one way to constrain the parameter increments so that the model converges.

3. The initial two dimension model constructed from the one dimension inversion result at each site seems to be a very effective way to start modelling a two dimensional region. The inversion procedure described below can be a very useful way to conduct the two dimensional inversion. First of all the model starts as simply as possible to fit the data. If this is not satisfactory then a more complicated model has to be introduced. One of the possible ways to deal with this is to subdivide some blocks where turns out to be necessary. As the model becomes very complicated, then the smoothing technique gives a useful tool for simplification. Analysis of the δ-like model gives more information about two dimensional inversion. Compared with Neuman's (1987) smoothing method, the method introduced
in chapter 6 for model smoothing can achieve the same aim without requiring a-priori information and extra iterations.

4. For resolution study, a modified resolution formulation from Jackson's most square method enables us to do a resolution study in a more practical and flexible way.

5. The introduction of the boundary parameterization in addition to the existing resistivity parameterization enables us to conduct two dimension inversion modelling more efficiently. The detailed analysis of the scaling matrix helps us to understand and carry out this new parameterization effectively. The proper choice of the scaling values can control the parameter's sensitivity successfully.

10.2. Suggestions

For the forward program, in the production of the Jacobian information efficiently using the iterative scheme, if a quick way could be found for computing the norm of $\|A\delta_A^{-1}\|$ then the application would be perfect.

Further work may be carried out to analyse the formulation of $\delta$-like models. Various other phenomena such as local minim may be analysed through different two dimensional inversion modelling cases.

To enable the boundary parameterization to work effectively, other methods may be found to control the parameter's sensitivity. Following the introduced scaling matrix, some other ways of choosing the scaling values may be given. Further analyses about the function of the scaling matrix can be investigated. This may enable us to better understand this new parameterization.
Apart from using the ridge regression method, some better methods might be found to conduct the two dimensional inversion modelling. The choice of the damping factors can be related to the parameter's sensitivity.

The modelling studies in this thesis show that detailed parameterization is one of the possible ways of ensuring the model converges. The question will be whether or not the following statement holds: for any given set of two dimensional data, a corresponding 2D model can be found as long as the area is parameterized in sufficient detail.

It may also be worthwhile to consider other models in the COMMEMI project as these may provide further realistic investigations and associated difficulties.

The modelling programs are available and enquires as to their use should be made to the department.
(a) One of the norms of an N by N matrix $A$ is the number defined by

$$\|A\| = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|}$$

(4.14)

where $x$ is a vector of dimension N and the norm of $x$ is defined by

$$\|x\| = (\sum_{i=1}^{N} x_i \bar{x}_i)^{\frac{1}{2}}$$

(4.15)

$x_i^*$ is the complex conjugate of $x_i$. (Strang, 1980)

(b) To give the convergence condition for the series in eq.(4.13), it is necessary to first prove the statement: if the norm (refering to (a)) $\|A\|<1$, then its eigenvalues $\lambda_i$ satisfy

$$|\lambda_i|<1, \, (i=1, \ldots, N)$$

(4.16)

Proof:

Let $x_i$ be a eigenvector of the matrix $A$.

$$Ax_i = \lambda_i x_i$$

(4.17)

So

$$\frac{\|Ax_i\|}{\|x_i\|} = |\lambda_i|$$

(4.18)

When $\|A\|<1$, by the definition of $\|A\|$, it is obvious that $|\lambda_i|<1$.

So the statement follows.

Since the following expansion is valid (Gantmacher, 1960)

$$(1-A)^{-1} = \sum_{i=0}^{\infty} A^i \quad \text{if} \quad (|\lambda_i|<1; \, i=1, 2, \ldots, N)$$

(4.19)
then if \( \| A^{-1} \delta_A \| < 1 \), then

\[
(1 + A^{-1} \delta_A)^{-1} = [1 - (-A^{-1} \delta_A)]^{-1}
\]

\[
= \sum_{i=0}^{\infty} (-1)^i (A^{-1} \delta_A)^i
\]

(4.20)

This shows that the series in eq.(4.13) is convergent if the condition 
\( \| A^{-1} \delta_A \| < 1 \) holds.

(c) The following expressions will be of use later:

\[
A_1 = A + \delta_A = A(1 + A^{-1} \delta_A)
\]

(4.21)

\[
A_1^{-1} = (A + \delta_A)^{-1} = (1 + A^{-1} \delta_A)^{-1} A^{-1}
\]

(4.22)

\[
(1 + A^{-1} \delta_A)^{-1} = 1 - A^{-1} \delta_A (I + A^{-1} \delta_A)\]

(4.23)

\[
(1 + A^{-1} \delta_A)^{-1} = 1 - A^{-1} \delta_A (I + A^{-1} \delta_A)^{-1}
\]

(4.24)

(d) To prove the norm of matrix A is less than the square root of the sum

of the squared elements in A:

\[
\| A \| \leq \left( \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}^2 \right)^{\frac{1}{2}}
\]

(4.25)

Proof:

Since

\[
\| Ax \| = \left[ \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_j^2 \right]^{\frac{1}{2}}
\]

(4.26)

and

\[
(\sum_{j=1}^{n} a_{ij} x_j^2)^2 \leq \sum_{j=1}^{n} a_{ij}^2 \sum_{k=1}^{n} x_k^2
\]

(4.27)

then we have

\[
\| Ax \| \leq \left[ \sum_{i=1}^{n} \left( \sum_{j=1}^{n} a_{ij}^2 \right) \left( \sum_{k=1}^{n} x_k^2 \right) \right]^{\frac{1}{2}}
\]

181
\[
\begin{align*}
&= \left( \sum_{k=1}^{n} x_k^2 \right)^{\frac{1}{2}} \left[ \sum_{i=1}^{n} \left( \sum_{j=1}^{n} a_{ij}^2 \right) \right]^{\frac{1}{2}} \\
&= \|\mathbf{x}\| \left[ \sum_{i=1}^{n} \left( \sum_{j=1}^{n} a_{ij}^2 \right) \right]^{\frac{1}{2}} \\
\end{align*}
\]

So that

\[
\frac{\|\Lambda \mathbf{x}\|}{\|\mathbf{x}\|} \leq \left( \sum_{i=1}^{n} \left( \sum_{j=1}^{n} a_{ij}^2 \right) \right)^{\frac{1}{2}}
\]

(4.29)

By the definition of \( \|\Lambda\| \), we have

\[
\|\Lambda\| \leq \left( \sum_{i=1}^{n} \left( \sum_{j=1}^{n} a_{ij}^2 \right) \right)^{\frac{1}{2}}
\]

(4.30)

The statement follows.

(e) To prove the following statements hold with regard to the norms \( \|\mathbf{x}_n - \mathbf{x}'\| \) and \( \|\Lambda_1 \mathbf{x}_n - \mathbf{y}_1\| \), namely

(i) \[
\|\mathbf{x}_n - \mathbf{x}'\| < \|\mathbf{x}_{n-1} - \mathbf{x}'\| \quad \text{when} \quad \|\Lambda^{-1} \delta_\Lambda\| < 1
\]

(ii) \[
\|\Lambda_1 \mathbf{x}_n - \mathbf{y}_1\| > \|\Lambda_1 \mathbf{x}_{n-1} - \mathbf{y}_1\| \quad \text{when} \quad \|\Lambda \delta_\Lambda^{-1}\| < 1
\]

(iii) \[
\|\Lambda_1 \mathbf{x}_n - \mathbf{y}_1\| < \|\Lambda_1 \mathbf{x}_{n-1} - \mathbf{y}_1\| \quad \text{when} \quad \|\delta_\Lambda \Lambda^{-1}\| < 1
\]

Proof:

By eq.(4.6) and eq.(4.11), we have

\[
\begin{align*}
\mathbf{x}_n - \mathbf{x}' &= -\Lambda^{-1} \Delta_{n-1} + \mathbf{x}_{n-1} - \Lambda_1^{-1} \mathbf{y}_1 \\
&= \Lambda^{-1} (\mathbf{y}_1 - \Lambda_1 \mathbf{x}_{n-1}) + \mathbf{x}_{n-1} - \Lambda_1^{-1} \mathbf{y}_1
\end{align*}
\]

(4.31)

Substituting \( \Lambda_1 \) and \( \Lambda_1^{-1} \) from eq.(4.21) and eq.(4.22),
\[ x_n - x' = \Lambda^{-1}(y_1 - (\Lambda + \delta_A)x_{n-1}) + x_{n-1} - (I + \Lambda^{-1}\delta_A)^{-1}\Lambda^{-1}y_1 \] (4.32)

Using eq.(4.23), and after some reduction,

\[ x_n - x' = -\Lambda^{-1}\delta_A x_{n-1} + \Lambda^{-1}\delta_A \Lambda^{-1}y_1 - (\Lambda^{-1}\delta_A)^2(1 + \Lambda^{-1}\delta_A)^{-1}\Lambda^{-1}y_1 \] (4.33)

that is

\[ x_n - x' = -\Lambda^{-1}\delta_A(x_{n-1} - \Lambda^{-1}y_1 + \Lambda^{-1}\delta_A(I + \Lambda^{-1}\delta_A)^{-1}\Lambda^{-1}y_1) \] (4.34)

For \( x_{n-1} \), using eq.(4.6),

\[ x_{n-1} - x' = x_{n-1} - \Lambda_1^{-1}y \] (4.35)

Substituting from eq.(4.22) gives

\[ x_{n-1} - x' = x_{n-1} - (I + \Lambda^{-1}\delta_A)^{-1}\Lambda^{-1}y_1 \] (4.36)

A further substitution from eq.(4.24) shows

\[ x_{n-1} - x' = x_{n-1} - \Lambda^{-1}y_1 + \Lambda^{-1}\delta_A(I + \Lambda^{-1}\delta_A)^{-1}\Lambda^{-1}y_1 \] (4.37)

Then combine eq.(4.34) with eq.(4.37), the result shows that

\[ x_n - x' = -\Lambda^{-1}\delta_A(x_{n-1} - x') \] (4.38)

Therefore

\[ \|x_n - x'\| = \|\Lambda^{-1}\delta_A(x_{n-1} - x')\| \] (4.39)

\[ \leq \|\Lambda^{-1}\delta_A\|\|x_{n-1} - x'\| \] (4.40)

Thus when \( \|\Lambda^{-1}\delta_A\| < 1 \), then \( \|x_n - x'\| < \|x_{n-1} - x'\| \) (statement (i) completes).

Using eq.(4.11), \( A_1x_n - y_1 \) can be expressed as

\[ A_1x_n - y_1 = -\Lambda_1^1 \Lambda_n^{-1}\Delta_n + A_1x_{n-1} - y_1 \] (4.41)

Using the definition of \( \Delta_n = A_1x_{n-1} - y_1 \), then

\[ A_1x_n - y_1 = (I - A_1^\Lambda)\Lambda_n^{-1} \] (4.42)

Define \( A_n = A_1x_n - y_1 \) and substitute \( A_1 \) by eq.(4.21) and then
\[ \Delta_n = (I - (\Lambda + \delta \Lambda) A^{-1}) \Delta_{n-1} \] (4.43)

\[ \Delta_n = (-1) \delta \Lambda A^{-1} \Delta_{n-1} \] (4.44)

Eq.(4.44) implies

\[ (-1) \Lambda \delta \Lambda^{-1} \Delta_n = \Delta_{n-1} \] (4.45)

Therefore

\[ \| \Lambda \delta \Lambda^{-1} \| \| \Delta_n \| \geq \| \Delta_{n-1} \| \] (4.46)

If \[ \| \Lambda \delta \Lambda^{-1} \| < 1 \] then \[ \| \Delta_n \| > \| \Delta_{n-1} \| \], that is \[ \| \Lambda_1 x_n - y_1 \| > \| \Lambda_1 x_{n-1} - y_1 \| \] (statement (ii) completes).

From eq.(4.44), the following inequality can be derived.

\[ \| \Delta_n \| \leq \| \Lambda \delta \Lambda^{-1} \| \| \Delta_{n-1} \| \] (4.47)

If \[ \| \Lambda \delta \Lambda^{-1} \| < 1 \] then \[ \| \Delta_n \| < \| \Delta_{n-1} \| \], that is \[ \| \Lambda_1 x_n - y_1 \| < \| \Lambda_1 x_{n-1} - y_1 \| \] (statement (iii) completes).

Thus statements (i), (ii), (iii) hold.
APPENDIX 2

(a) We are going to prove that $|p_i^2 - p_i^1| > |\log p_i^2 - \log p_i^1|$, where $p_i^2 = p_i^1 + \varepsilon$, $|p_i^1| > 1$, $|\varepsilon/p_i^1| < 1$.

Proof:

$$\log p_i^2 - \log p_i^1 = \log (p_i^2/p_i^1) = \log [(p_i^1 + \varepsilon)/p_i^1] = \log [1 + \varepsilon/p_i^1]$$  \hspace{1cm} (7.38)

Expanding the above in Taylor series, which is

$$= \frac{\varepsilon}{p_i^1} - \frac{\varepsilon^2}{2(p_i^1)^2} + \frac{\varepsilon^3}{3(p_i^1)^3} - \frac{\varepsilon^4}{4(p_i^1)^4} + \ldots + (-1)^{n+1} \frac{\varepsilon^n}{n!(p_i^1)^n}\hspace{1cm} (7.39)$$

when $|\varepsilon| < 1$, the above series is convergent (Hirschman, 1962). The absolute value of the above series is less than the absolute value of its first term in the series (Knopp, 1928), which means that

$$||\log p_i^2 - \log p_i^1|| < ||\varepsilon||$$  \hspace{1cm} (7.40)

but $|p_i^2 - p_i^1| = ||\varepsilon||$

so that $|p_i^2 - p_i^1| > ||\log p_i^2 - \log p_i^1||$

The statement follows. Thus

$$\frac{\log p_i^2 - \log p_i^1}{p_i^2 - p_i^1} = \tau_j, \hspace{1cm} \tau_j < 1$$  \hspace{1cm} (7.41)

As in eq.(7.11).

(b) To show the relationship between the scaled solution and the unscaled one as $\lambda \to \infty$. From eq.(7.21) the scaled solution is expressed as

$$x' = Q[(AQ)^TAQ + \lambda I]^{-1}(AQ)^Ty$$  \hspace{1cm} (7.42)

Since

$$[(AQ)^TAQ + \lambda I]^{-1}(AQ)^T = (A^TAQ + \lambda Q^{-1})^{-1}A^T$$  \hspace{1cm} (7.43)
\[ x' = Q(\lambda A^T A + \lambda I)^{-1} A^T y \]  

Since the non-scaled solution \( x = (\lambda A^T A + \lambda I)^{-1} A^T y \), the above can be written as

\[ x' = Q(\lambda A^T A + \lambda I)^{-1} A^T y (7.45) \]

\[ x' = Q(\lambda A^T A + \lambda I)^{-1} A^T A + \lambda I)x (7.46) \]

If \( \lambda \) tends to \( \infty \), then

\[ (\lambda A^T A + \lambda I)^{-1} \rightarrow Q/\lambda, \quad (\lambda A + \lambda I) \rightarrow \lambda \]

so that

\[ x' \rightarrow Q^2 x (7.47) \]

Thus

\[ x_i' = (1/\tau_i)x_i \quad (i=1, ..., N) (7.48) \]

which shows that the scaled solution \( x_i' \) is linearly related to unscaled solution \( x_i \) by a factor \( 1/\tau_i \), if the damping factor \( \lambda \) is sufficiently large.

(c) The following shows that the scaled solution \( x' \) is bounded if part (or all) of the scaling values \( \tau_i \rightarrow 0 \). From eq.(7.46) the scaled solution can be expressed as

\[ x' = Q(\lambda A^T A + \lambda I)^{-1} (A^T A + \lambda I)x \]

Since

\[ Q(\lambda A^T A + \lambda I)^{-1} = [A^T A + \lambda I]^{-1} = (A^T A + \lambda I^{-1} \lambda I - \lambda I^{-2})^{-1} (A^T A + \lambda I^{-2})^{-1} (7.49) \]

then

\[ x' = (A^T A + \lambda I)^{-1} (A^T A + \lambda I^{-1} \lambda I - \lambda I^{-2})x \]

\[ x' = x + \lambda (A^T A + \lambda I^{-2})(I - I^{-2})x \]

\[ x' = x + \lambda (A^T A + \lambda I^{-2})(I - I^{-2})x \]

Denote \( e_i \) as a vector of dimension \( N \) with the \( i \)-th value equal to 1 while others equal 0, i.e.

\[ e_i^T = (0, 0, ... 1, ..., 0, 0) \quad (7.51) \]

and \( e_i^T x = x_i \), so that
\[ x_i' = e_i^T x' = e_i^T x + e_i^T \lambda (A^T A + \lambda Q^{-2})^{-1} (I - Q^{-2}) x \]  \hspace{1cm} (7.52)

That is

\[ x_i' = x_i + \lambda e_i^T (A^T A + \lambda Q^{-2})^{-1} (I - Q^{-2}) x \]  \hspace{1cm} (7.53)

As \( \tau_i \rightarrow 0 \), for only \( i = r, \ldots, r+q \), then

\[ x_i' \rightarrow x_i + \lambda e_i^T \Theta x \]  \hspace{1cm} (7.54)

where

\[ \Theta = \lim_{\tau_i \rightarrow 0^+} \left[ (A^T A + \lambda Q^{-2})^{-1} (I - Q^{-2}) \right] = (A^T A + \lambda Q')^{-1} (I - Q') \]  \hspace{1cm} (7.55)

\( Q' \) is a diagonal matrix of dimension \( N \) by \( N \) with zero value from \( r \)th to \( (r+q) \)th diagonal position while other diagonal values are \( \tau_i \) (\( i = 1, \ldots, r-1, r+1, \ldots, N \)).

If \( \tau_i \rightarrow 0 \) for all \( i = 1, \ldots, N \), then

\[ x_i' \rightarrow x_i + \lambda e_i^T \Theta x \]  \hspace{1cm} (7.56)

where

\[ \Theta = \lim_{\tau_i \rightarrow 0^+} \left[ (A^T A + \lambda Q^{-2})^{-1} (I - Q^{-2}) \right] = (A^T A + \lambda Q')^{-1} \]  \hspace{1cm} (7.57)

If \( (A^T A)^{-1} \) exists then \( \Theta = (A^T A)^{-1} \).

This completes the statement.

(d) Suppose only one parameter is scaled, namely \( \tau_r \pm 1 \) (i.e. \( 1/\tau_r^{1/2} \) in the \( r \)th diagonal position of the scaling matrix \( Q \)) and \( \tau_i = 1 \) (\( i \neq r \)) in the diagonal scaling matrix \( Q \). From eq.(7.46), the scaled solution is

\[ x' = Q (A^T A + \lambda Q^{-1})^{-1} (A^T A + \lambda I) x \]  \hspace{1cm} (7.58)

Let \( E_{rr} \) be a \( N \) by \( N \) matrix with the element located at \( r \)th column and \( r \)th row equal to 1 and all other elements equal to zero. Define

\[ \nu = 1/\left(\tau_r \right)^{1/2} \]  \hspace{1cm} (7.59)

and \( \nu \) is let to be \( 1+\gamma_r \), that is

187
\( v = 1 + \gamma_r \)  
then \( v > 1 \) implies \( \gamma_r > 0 \). Thus

\[
Q = I + \gamma_r E_{rr}, \quad Q^{-1} = I - \gamma_r E_{rr} / v
\]

Substituting \( Q \) and \( Q^{-1} \) into eq.(7.58)

\[
\Lambda^T A Q + \lambda Q^{-1} = \Lambda^T A + \gamma_r \Lambda^T A E_{rr} + \lambda I - \lambda \gamma_r E_{rr} / v
\]  \( (7.61) \)

which implies

\[
\Lambda^T A + \lambda I = \Lambda^T A Q + \lambda Q^{-1} - \gamma_r \Lambda^T A E_{rr} + \lambda \gamma_r E_{rr} / v
\]

Hence eq.(7.58) is expressed as

\[
x' = Q[1 + \gamma_r (\Lambda^T A Q + \lambda Q^{-1})^{-1} (\lambda I / v - \Lambda^T A) E_{rr}] x
\]

\[
= Q x + \gamma_r Q (\Lambda^T A Q + \lambda Q^{-1})^{-1} (\lambda I / v - \Lambda^T A) E_{rr} x
\]  \( (7.62) \)

Since \( E_{rr} x = x_r e_r \), the rth value of the scaled \( x_r' \) in \( x' \) can be expressed as

\[
x_r' = e_r^T x' = v x_r + \gamma_r e_r^T Q (\Lambda^T A Q + \lambda Q^{-1})^{-1} (\lambda I / v - \Lambda^T A) x_r e_r
\]  \( (7.63) \)

where \( e_r \) is a vector of dimension \( N \) with the rth element equal to one and all others equal to zero.

For the non-scaled solution \( x_i \) (i=r),

\[
x_i' = e_i^T x' = x_i + \gamma_r e_i^T Q (\Lambda^T A Q + \lambda Q^{-1})^{-1} (\lambda I / v - \Lambda^T A) x_r e_r
\]  \( (7.64) \)

The last term of (7.64) is of the form \( e_i^T B e_r \) (with \( B \) an \( N \) by \( N \) matrix), let \( B_{r,i} \) be the element in the \( i \)th row and \( r \)th column of \( B \), \( e_i^T B e_r = B_{r,i} \). Then

\[
x_i' = x_i + B_{r,i}
\]  \( (7.65) \)

Using eq.(7.49), eq.(7.63) can be presented as

\[
x_r' = x_r [v + \gamma_r \lambda e_r^T (\Lambda^T A + \lambda Q^{-2})^{-1} e_r / v - \gamma_r e_r^T (\Lambda^T A + \lambda Q^{-2})^{-1} \Lambda^T A e_r]
\]  \( (7.66) \)

Substituting equation (7.60) and the following
\[ e_r^T (A^T A + \lambda Q_2^{-2})^{-1} A^T A e_r = e_r^T (A^T A + \lambda Q_2^{-2})^{-1} (A^T A + \lambda Q_2^{-2} - \lambda Q_2^{-2}) e_r \]
\[ = 1 - \lambda e_r^T (A^T A + \lambda Q_2^{-2})^{-1} Q_2^{-2} e_r \]
\[ = 1 - \lambda e_r^T (A^T A + \lambda Q_2^{-2})^{-1} e_r / \nu^2 \] (7.67)

into eq.(7.66) leads to

\[ x_r' = x_r [1 + \lambda \gamma_r (1 + \nu) e_r^T (A^T A + \lambda Q_2^{-2})^{-1} e_r / \nu^2] \] (7.68)

From eq.(7.68), the ratio between \( x_r' \) and \( x_r \) is

\[ r = \frac{x_r'}{x_r} = [1 + \lambda \gamma_r (1 + \nu) e_r^T (A^T A + \lambda Q_2^{-2})^{-1} e_r / \nu^2] \] (7.69)

Substituting eq.(7.59), gives

\[ r = \frac{x_r'}{x_r} = [1 + \lambda \gamma_r (2 + \gamma_r) e_r^T (A^T A + \lambda Q_2^{-2})^{-1} e_r / (1 + \gamma_r)^2] \]

Matrix \( A^T A + \lambda Q_2^{-2} \) is positive definite when \( \lambda \neq 0 \), and since \( \gamma_r \) is larger than zero then

\[ r > 0 \] and \[ |r| > 1 \]

The above expressions are equivalent to

\[ X_r' X_r \geq 0 \] and \[ |X_r'| \geq |X_r| \]

and from (7.69) the equality holds if and only if \( \lambda = 0 \). Also from (7.68) \( x_r' = 0 \) if and only if \( x_r = 0 \).

(e) To show that the ratio between the scaled and unscaled solution is a decreasing function of \( \tau_r \), it is necessary to prove \( \partial r / \partial \tau_r \leq 0 \), where \( r = x_r' / x_r \).

Following (d) the derivative of eq.(7.69) with respect to \( \lambda \) is

\[ \partial r / \partial \lambda = \gamma_r (1 + \nu) [e_r^T (A^T A + \lambda Q_2^{-2})^{-1} e_r + \lambda e_r^T (\partial (A^T A + \lambda Q_2^{-2})^{-1} / \partial \lambda) e_r / \nu^2] \] (7.70)

Since

\[ \partial (A^T A + \lambda Q_2^{-2})^{-1} / \partial \lambda = -(A^T A + \lambda Q_2^{-2})^{-1} Q_2^{-2} (A^T A + \lambda Q_2^{-2})^{-1} \] (7.71)

and

189
\[
\lambda \partial (A^T A + \lambda Q^{-2})^{-1} / \partial \lambda = -(A^T A + \lambda Q^{-2})^{-1} \lambda Q^{-2} (A^T A + \lambda Q^{-2})^{-1}
\]
\[
= -(A^T A + \lambda Q^{-2})^{-1} (A^T A + \lambda Q^{-2} - A^T A) (A^T A + \lambda Q^{-2})^{-1}
\]
\[
= -(A^T A + \lambda Q^{-2})^{-1} + (A^T A + \lambda Q^{-2})^{-1} A^T A (A^T A + \lambda Q^{-2})^{-1}
\]

Then eq.(7.70) can be presented as
\[
\partial r / \partial \lambda = \gamma_r (1 + \nu) e_r^T (A^T A + \lambda Q^{-2})^{-1} A^T A (A^T A + \lambda Q^{-2})^{-1} e_r / \nu^2
\] (7.72)

Let \( V = e_r^T (A^T A + \lambda Q^{-2})^{-1} A^T \), which is a vector of \( N \). Then eq.(7.72) is
\[
\partial r / \partial \lambda = \gamma_r (1 + \nu) V^T V \geq 0
\] (7.73)

This means \( r \) is an increasing function of \( \lambda \) for \( \lambda \neq 0 \).

The derivative of eq.(7.69) with respect to \( \nu \) is
\[
\partial r / \partial \nu = \partial r / \partial \gamma_r
\] (7.74)

Let \( U = (A^T A + \lambda Q^{-2})^{-1} \) then
\[
\partial r / \partial \nu = \partial r / \partial \gamma_r = [2 e_r^T U e_r / \nu + \gamma_r (1 + \nu) e_r^T (\partial U / \partial \gamma_r) e_r] \lambda / \nu^2
\] (7.75)

Since \( U^{-1} = A^T A + \lambda Q^{-2} \), which can be expressed as
\[
U^{-1} = A^T A + \lambda I + \lambda (1 / \nu^2 - 1) E_{rr},
\]
and
\[
\partial U^{-1} / \partial \gamma_r = -2 E_{rr} / (1 + \gamma_r)^3
\] (7.76)

then the derivative of \( U \) with respect to \( \gamma_r \) can be obtained from
\[
\partial U / \partial \gamma_r = -U (\partial U^{-1} / \partial \gamma_r) U
\] (7.77)

Substituting eq.(7.76), eq.(7.77) becomes
\[
\partial U / \partial \gamma_r = 2 \lambda U E_{rr} U / \nu^3
\] (7.78)

so that eq.(7.75) can be presented as
\[
\partial r / \partial \nu = 2 \lambda e_r^T U e_r / \nu^3 + 2 \lambda \gamma_r (1 + \nu) U E_{rr} U / \nu^5
\] (7.79)

\( U \) is positive definite if \( \lambda \neq 0 \), which implies that
$e_r^T U e_r > 0, \quad e_r^T U e_{rr} U e_r > 0$

so that

$\partial r / \partial \nu \geq 0$

Since $\nu$ is defined as $\nu = 1/\tau_r^{\frac{1}{2}}$, and

$\partial r / \partial \tau_r = \left[ \partial r / \partial \nu \right] = \left[ \tau_r^3 / 2 \right] \left[ \partial r / \partial \nu \right]$ \hspace{1cm} (7.80)

then

$\partial r / \partial \tau_r \leq 0$

The above equality holds if and only if $\lambda = 0$. This shows that the ratio $r$ is a decreasing function of $\tau_r$ when $\lambda \neq 0$. 

191
REFERENCES


Cherchi, A., Montadert, L., 1985. The major events responsible for the


Fischer, G. and Le Quang, B.V., 1981b. Topography and minimization of the standard deviation in one-dimensional magnetotelluric modelling.,


Lee, S., McMechan, G. and Aiken, C.L., 1987. Phase-field imaging: The electromagnetic equivalent of seismic migration., Geophysics, 52,


Rankin, D., 1962. The magnetotelluric effect on a dike., Geophysics, XXVII,


Sanfilipo, W.A. and Hohmann, G.W., Integral equation solution for the transient electromagnetic response of a three-dimensional body in a conductive half-space., Geophysics, 50, 798-809.


Swift, C.M., Jr., 1986. A magnetotelluric investigation of an electrical


Weaver, J.T., 1986. Finite difference calculations for COMMEMI., Department report, Dept. of Physics, Univ. of Victoria, Canada.


