Inversion of Magnetotelluric Data in an Anisotropic Domain

A new approach to the magnetotelluric inverse problem in an anisotropic domain driven by different geophysics data

by

ERIC MANDOLESI

A thesis submitted to
The Faculty of Science, National University of Ireland, Galway
in fulfilment of the requirements for the degree of

Doctor of Philosophy (PhD)

August 19, 2013

SUPERVISION BY:
Prof. Dr. Alan G. Jones, Dublin Institute for Advanced Studies, Dublin, Ireland
AND
Dr. Colin Brown, National University of Ireland, Galway, Ireland

School of Cosmic Physics,
Geophysics Section,
Dublin Institute for Advanced Studies,
Dublin, Ireland

Faculty of Science,
Department of Earth and Ocean Science,
National University of Ireland,
Galway, Ireland

Contents

List of Acronyms

1D	one-dimensional 8
2D	two-dimensional 2
3D	three-dimensional 14
APM	absolute plate motion 99
DFT	discrete Fourier transform 38
EM	electromagnetic 1
GA	genetic algorithm 46
ICHEC	Ireland's High-Performance Computing Center 64
JPD	joint probability distribution 71
LM	Levenberg-Marquardt 46
MI	mutual information 2
MT	magnetotelluric 1
PDE	partial differential equation 43
RHS	right hand side 119
RMS	root mean square 64
SVD	single value decomposition 44

Abstract

The aim of this study was to develop a set of tools – algorithms that would be translated into computer programs at a later stage – to invert anisotropic magnetotelluric data within a framework that integrates multi-disciplinary information relative to the subsurface.

The initial stages of research was aimed at understanding and summarizeing the available modeling strategies for the magnetotelluric method, to identify an inversion strategy that was both accurate and effective, and to explore a method to allow information from different disciplines to be used in the inversion process. Over a year was spent testing the suitability of a selected genetic algorithm to be used in solving the inverse problem. Even though this genetic algorithm was successfully used in different inverse problems with magnetotelluric data, it was not successful here because of the lack of efficiency in the framework that was used.

The final part of this study was dedicated to the development, testing and appraisal of newly-developed codes based on the concept of "mutual information", information shared between two images that can be quantified in a probabilistic sense. Preliminary studies relating to synthetic tests were performed, and the results analyzed from a numerical perspective. A real anisotropic one-dimensional dataset from the DIE magnetotelluric station deployed in Central Germany was inverted. The resulting model was consistent with the most up-to-date models from independent research, surpassing these models by effectively constraining the amount of anisotropy needed to fit the data. These results made the electrical conductivity values obtained in the studied region compatible with laboratory measurements. The anisotropic two-dimensional approach is demonstrated on a test model, and shown to be very effective at elucidating conductivity structures.

Acknowledgements

It is impossible to name here all the people that have helped and supported me during my time at DIAS as a PhD student, this thesis is the result of the influence of every one of them. I am thus persuaded to cite here the ones that had the strongest impact on my work, leaving to the others my silent gratitude.

I want to thank my supervisors Alan Jones and Colin Brown for their guidance, encouragement and especially for the freedom they gave me during my research. Besides my supervisors, I wish to express all my gratitude to Josef Pek as any chat with him was helpful and encouraging; moreover, he insisted on paying all the pub's bills during my visit at the Academy of Sciences of the Czech Republic in November 2011.

During my visit to the University of British Columbia in November 2009 and June 2010, Eldad Haber lent me his ideas regarding the use of mutual information in inverse problems, while Doug Oldenburg reminded me of both the importance of regularization and the importance of obtaining a degree. I wish to thank them both for the inspiration and the advice they provided.

All the inhabitants of 5 Merrion Square were at a certain point involved in this thesis. I want to name here Andrew Schaeffer, one of the common elements in both sets of my friends and my colleagues, as well as Sarah, Thomas and Robert who (thanks Sarah!) had the most important influence in my writing style (or lack of style).

A big, fat, thanks is for "the big guys" of the upper floors: Estelle Roux, Mark Muller, Marion Miensopust, Celine Tyrell, Sergei Lebedev, Jan Vözar and the lately arrived Florian "doc" LePape.

Finally I need to thank my family: my parents's support has been most important in crescendo and my brother Fulvio has helped me in maintaining a link, however weak, with real life.

Last but not least, thank you Marlene, simply for everything.

Science is a way of trying not to fool yourself. The first principle is that you must not fool yourself, and you are the easiest person to fool.

– Richard Feynman

1 Introduction

"Our understanding of how the Earth operates, particularly its tectonic history and secular variation of tectonic processes, is severely limited by both our lack of knowledge, and the intrinsic bias in knowledge, at depth" [*Jones*, 2007].

Despite the great improvements in both the quality and the quantity of geophysical data in the last 20 years, the resolution of these data, due to both noise and the physical limits of the geophysical methods, limits the level of detail that can be achieved by subsurface imaging techniques. Recently, several authors have attempted a fusion between different geophysical techniques, led by the idea that a specific method can improve the resolution of some aspects where others are weak or lack it. This approach makes us sure that when different imaging methods are carefully chosen, at least one of the many datasets used has the resolution appropriate to constrain the subsurface image at the depth required for the specific study undertaken. An alternative possibility is to increase the amount of measured data of a certain type and model these data whilst sacrificing some approximations and simplifications in the modelling schemes. With this strategy the maximum amount of information is extracted from the data, avoiding artifacts caused by incorrect simplifying assumptions or imprecise approximations.

In this thesis, a combination of these two strategies is attempted by the development of an inversion scheme for magnetotelluric (MT) data that is able to link the electrical conductivity distribution of the subsurface to a given reference model, computed from the most adequate method available to bound the electromagnetic (EM) retrieved model. More precisely, the MT data are inverted using the the subsurface as an anisotropic domain. The study of electrical anisotropy is a key factor to understand the physical – and tectonic – processes that take place in the upper mantle. Notwithstanding this consideration, many authors continue to use simplified modeling schemes that are unable to work with data measured in an anisotropic environment. By aiming to constrain the obtained image of

the subsurface, we test the effectiveness of a mutual information (MI) based constraint.

The goals of this thesis are multiple. Firstly it is firm the intention to develop a computer code to solve the inverse MT problem in an anisotropic domain. Despite the well-known difficulties that affect inverse problems, we attempt to keep the model domain characterization as detailed as possible. With this goal in mind, we decided to attempt the solution of the MT inverse problem in an anisotropic layered half-space and in a more complex two-dimensional (2D) anisotropic domain. In both cases we reject the commonly used simplifying assumptions relative to the electrical conductivity tensor, and treat it in its full, complete form.

Secondly, to make the constraint based on different geophysical models effective, we attempt to explore a new way to keep the MT model close to the one chosen to "drive" the inverse process. Despite its popularity in information theory, the MI is rarely – if ever – used in geophysical inverse problems, but commonly used in many others areas of science. For example, tomographic and imaging methods based on MI are broadly used in image registration, computer vision or medical imaging.

We chose to present our research in three parts. The first part presents the theories and mathematical settings that allow the new approach to the inverse problem, the object of this thesis, to be assembled. In this part we present the physical approximations to Maxwell's equations and the formal methods that allow us to model and solve the MT equations. We then present the fundamental strategies used to solve the inverse problems, with special regard to the methods we select to solve the problem treated in this thesis. Finally we present the basic definitions of information theory, using some examples to help the reader to appreciate the versatility of the distance function used.

The second part includes results of tests performed using data from synthetic models and from a real survey. Within this part, we highlight the problems and successes of the tests themselves.

The third part summarizes our conclusions and outlooks, and includes further material, such as a bibliography and a list of software used.

Part I

Theoretical background

The first step to be taken, is to study carefully the fundamental phenomenon above described, and to examine all the various circumstances under which it presents itself.

– Jean-Baptiste Biot

2

The Principles of the Magnetotelluric Method

The MT method is a frequency-domain EM technique that makes use of the natural EM fields as signal source. *Tikhonov* [1950] and independently *Cagniard* [1953] and *Rikitake* [1950, 1951a,b,c] were the first that developed the basic theory of MT. In this chapter the fundamental equations of the MT problem will be obtained from Maxwell's equations and discussed in relation to physical Earth properties.

The electric field $\vec{E}$ and magnetic field $\vec{B}$ are vector fields that describe interactions between electric charges, respectively stationary or moving with velocity $\vec{v}$. $\vec{E}$ and $\vec{B}$ values at any position $\vec{r}$ and at any time t are defined through the force $\vec{F}$ that in $(\vec{r}, t)$ is exerted on a point test charge. This is expressed through the Lorentz force

$$\vec{F} = q(\vec{E} + \vec{v} \times \vec{B}). \tag{2.1}$$

Defined in this way, $\vec{E}$ and $\vec{B}$ (and the auxiliary fields $\vec{D}$ and $\vec{H}$ used to describe interaction between electromagnetic field and matter) are linked to the Maxwell equations' sources, that are the starting point of the discussion proposed in this chapter, reported as follows:

$$\vec{\nabla} \cdot \vec{D} = \varrho; \tag{2.2}$$

$$\vec{\nabla} \cdot \vec{B} = 0; \tag{2.3}$$

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}; \tag{2.4}$$

$$\vec{\nabla} \times \vec{H} = \vec{J} + \frac{\partial \vec{D}}{\partial t}. \tag{2.5}$$

The sources of electromagnetic fields, the charge density ϱ and the current density $\vec{J}$, are not fully independent; they are subject to charge conservation formalized through the continuity equation:

$$\vec{\nabla} \cdot \vec{J} + \frac{\partial \varrho}{\partial t} = 0. \tag{2.6}$$

A number of assumptions are made in order to derive the equations forming the MT problem:

1. Maxwell's equations are obeyed. Cases in which this assumption is invalid are unknown.
2. The Earth does not generate electromagnetic energy, only dissipates or absorbs it.
3. All fields are treated as conservative and analytic away from their sources.
4. For interpretation Magnetotelluric source fields are treated as uniform, plane-polarized electromagnetic waves striking the Earth on a nearly vertical incidence.
5. Charge is conserved and the Earth is treated as an ohmic conductor, obeying the Ohm's law

$$\vec{J} = \sigma \vec{E}. \tag{2.7}$$

6. Displacement currents $\frac{\partial \vec{D}}{\partial t}$ are negligible.
7. Linear relationship is assumed between $\vec{D}$ and $\vec{E}$ and between $\vec{H}$ and $\vec{B}$ and changes in electrical permittivities ε and magnetic permeabilities μ, so

$$\vec{B} = \mu_0 \vec{H}, \tag{2.8}$$
$$\vec{D} = \varepsilon_0 \vec{E}. \tag{2.9}$$

Situation in which assumptions 1 or 3 are invalid are unknown. Assumption 4 is discussed in the Section 2.6, while in regard to the assumption 7 it is the classic quasi-static approximation. All of these assumption have been discussed in several publications (cf. for example *Cagniard* [1953], *Price* [1962]; *Price* [1973] or *Vozoff* [1991]) and summarized in *Simpson and Bahr* [2005]. Under these assumptions, Maxwell's equations are

approximated as follows

$$\vec{\nabla} \cdot \vec{E} = \frac{\varrho}{\varepsilon_0}; \tag{2.10}$$

$$\vec{\nabla} \cdot \vec{B} = 0; \tag{2.11}$$

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}; \tag{2.12}$$

$$\vec{\nabla} \times \vec{B} = \mu_0 \sigma \vec{E}. \tag{2.13}$$

Moving from the time-domain into a frequency-domain through Fourier transform, derivatives simplify to multiplications, so

$$-\frac{\partial \vec{B}(t)}{\partial t} \xrightarrow{\mathcal{F}} -i\omega \vec{B}(\omega) \tag{2.14}$$

at any angular frequency ω. Applying the divergence to equation 2.13, we obtain

$$\vec{\nabla} \cdot \left(\vec{\nabla} \times \vec{B}\right) = \vec{\nabla} \cdot \left(\mu_0 \sigma \vec{E}\right) \tag{2.15}$$

$$= \mu_0 \vec{\nabla} \cdot \left(\sigma \vec{E}\right) \tag{2.16}$$

$$= \mu_0 \left(\sigma \vec{\nabla} \cdot \vec{E} + \vec{E} \vec{\nabla} \sigma\right). \tag{2.17}$$

Using the well-known vectorial relation $\vec{\nabla} \cdot \left(\vec{\nabla} \times \vec{a}\right) = 0$ for any $\vec{a}$, it results from equation 2.17 that $\vec{\nabla} \cdot \vec{E} = -\frac{\vec{E}\vec{\nabla}\sigma}{\sigma}$. Replacing the term $\vec{\nabla} \cdot \vec{E}$ in equation 2.10, it follows that

$$\frac{\varrho}{\varepsilon_0} = -\vec{E} \cdot \frac{\vec{\nabla}\sigma}{\sigma} = -\vec{E} \cdot \vec{\nabla} \ln \sigma. \tag{2.18}$$

Maxwell's equations therefore become

$$\vec{\nabla} \cdot \vec{E} = -\vec{E} \cdot \vec{\nabla} \ln \sigma; \tag{2.19}$$

$$\vec{\nabla} \cdot \vec{B} = 0; \tag{2.20}$$

$$\vec{\nabla} \times \vec{E} = -i\omega \vec{B}; \tag{2.21}$$

$$\vec{\nabla} \times \vec{B} = \mu_0 \sigma \vec{E}. \tag{2.22}$$

If the curl of equations 2.21 and 2.22 is taken, it is possible to reduce the four Maxwell's equations into two, which describe the behavior of the electromagnetic field under the mentioned assumptions.

$$\vec{\nabla} \times \vec{\nabla} \times \vec{E} = -i\omega\vec{\nabla} \times \vec{B} \tag{2.23}$$
$$= -\nabla^2\vec{E} + \vec{\nabla}\left(\vec{\nabla} \cdot \vec{E}\right). \tag{2.24}$$

Substituting the term $\vec{\nabla} \cdot \vec{E}$ with equation 2.19

$$\nabla^2\vec{E} = i\omega\mu_0\sigma\vec{E} - \vec{\nabla}\left(\vec{E}\vec{\nabla}\ln\sigma\right). \tag{2.25}$$

Similarly, it is computed from the curl of equation 2.22

$$\nabla^2\vec{B} = i\omega\mu_0\sigma\vec{B} + \mu_0\vec{E} \times \vec{\nabla}\sigma. \tag{2.26}$$

2.1. Homogeneous half-space

In order to derive the basic concepts of MT it is useful to explore the behaviour of equations 2.25 and 2.26 in the most simple environment examinable. In a homogeneous half-space the conductivity σ inside the Earth is constant and equal to σ_0. Equations 2.25 and 2.26 become

$$\nabla^2\vec{F} = i\omega\mu_0\sigma\vec{F} \tag{2.27}$$

with $\vec{F} = \vec{E}$ or $\vec{B}$ respectively. Equation 2.27 is a diffusion equation with solutions in the form

$$\vec{F} = \vec{F}_1 e^{i\omega t - \alpha z} + \vec{F}_2 e^{i\omega t + \alpha z} \tag{2.28}$$

where $\alpha^2 = i\omega\mu_0\sigma$ and α chosen so that $\mathcal{R}(\alpha) > 0$. In a homogeneous half-space that does not contain any field source, $\vec{F}_2$ must be identically zero, because waves have to travel to deepest z and reflections are not possible. It results that the solution to the magnetotelluric problem in a homogeneous half-space has the form

$$\vec{E} = \vec{E}_0 e^{i\omega t - \alpha z}; \tag{2.29}$$
$$\vec{B} = \vec{B}_0 e^{i\omega t - \alpha z}. \tag{2.30}$$

Taking the second derivative with regard to depth of equation 2.28 yields

$$\frac{\partial^2\vec{F}}{\partial z^2} = \alpha^2\vec{F}_1 e^{i\omega t - \alpha z} = \alpha^2\vec{F}. \tag{2.31}$$

In the uniform half-space, assuming a planar wave source, $\frac{\partial^2\vec{F}}{\partial x^2} = \frac{\partial^2\vec{F}}{\partial y^2} = 0$ and it is possible

to relate equation 2.31 with equation 2.27, computing α

$$\alpha = \sqrt{i\omega\mu_0\sigma} = (i+1)\sqrt{\frac{\omega\mu_0\sigma}{2}} \tag{2.32}$$

The inverse of α, C, is called the Schmucker-Weidelt transfer function:

$$C = \frac{1}{\alpha} = \frac{p}{2} - i\frac{p}{2} \tag{2.33}$$

where $p = \frac{1}{\mathrm{Re}(\alpha)}$ is the electromagnetic skin depth or penetration depth of an electric field at the frequency $f = 2\pi\omega$ into an half-space of conductivity σ.

The transfer function C links physical properties of the Earth and the measured EM fields: combining equation 2.31 for the field $\vec{E}$ derived with respect to z

$$\frac{\partial\vec{E}}{\partial z} = -\alpha\vec{E} \tag{2.34}$$

and equation 2.21, gives:

$$\frac{\partial E_x}{\partial z} = -i\omega B_y = -\alpha E_x. \tag{2.35}$$

Thus,

$$C = \frac{1}{\alpha} = \frac{E_x}{i\omega B_y} = -\frac{E_y}{i\omega B_x}. \tag{2.36}$$

In this way it is possible to compute C from the measured fields $\vec{E}$ and $\vec{B}$ and combining equation 2.36 and equation 2.32 the conductivity σ of the half-space is finally computed through

$$\frac{1}{\sigma} = \frac{1}{|\alpha^2|}\mu_0\omega = |C^2|\,\mu_0\omega. \tag{2.37}$$

This result justifies the name of transfer function, giving the idea that the Earth works as a linear system for which the knowledge of C, computed by the measurement of the electromagnetic field, is sufficient to determine the physical properties of the Earth, namely the sole conductivity σ.

In practice, the impedance tensor $\hat{Z}$ is used in spite of C and is defined, in this one-dimensional (1D) isotropic case, as

$$\hat{Z} = \begin{pmatrix} 0 & Z_{xy} \\ Z_{yx} & 0 \end{pmatrix} = \begin{pmatrix} 0 & Z_{xy} \\ -Z_{xy} & 0 \end{pmatrix} \tag{2.38}$$

with

$$Z_{xy} = -Z_{yx} = \frac{\mu_0 E_x}{B_y} = -\frac{\mu_0 E_y}{B_x} = i\omega\mu_0 C\ [\Omega] \tag{2.39}$$

or equivalently

$$Z_{xy} = -Z_{yx} = \frac{E_x}{B_y} = -\frac{E_y}{B_x} = i\omega C\left[m \cdot s^{-1}\right] \tag{2.40}$$

in which the only difference is the unit of measure: "field-unit" in the first case, "SI-units" in the second. The SI-units will be used unless not otherwise indicated.

2.2. Layered half-space

Considering an m-layered mono-dimensional Earth model leads to a result that is different from the elegant solution found for the homogeneous half-space. In a layered Earth, waves travel both downwards and – due to reflection at inhomogeneous boundaries – upwards; it is therefore mandatory to consider the complete solution 2.28, in which both F_1 and F_2 terms are non zero in all layers except for the bottom-most . The solutions are then connected at the interfaces using the correct boundary conditions, with the exception of the lowermost half-space, where waves travel only downwards, leading to a solution formally identical to 2.31.

In the generic n-th layer

$$E_x^n = a_n e^{-\alpha_n z} + b_n e^{+\alpha_n z}, \tag{2.41}$$

and applying equation 2.21 yields

$$B_x^n = \frac{\alpha_n}{i\omega}\left(a_n e^{-\alpha_n z} - b_n e^{+\alpha_n z}\right). \tag{2.42}$$

It follows that the transfer function inside the n-th layer is given by

$$C_n(z) = \frac{E_x^n}{i\omega B_x^n} = \frac{a_n e^{-\alpha_n z} + b_n e^{+\alpha_n z}}{\alpha_n\left(a_n e^{-\alpha_n z} - b_n e^{+\alpha_n z}\right)}. \tag{2.43}$$

Specifying the solution for the top of the layer, $z = z_n$, and at the bottom, $z = z_{n+1}$, gives:

$$C_n(z_n) = \frac{a_n e^{-\alpha_n z_n} + b_n e^{+\alpha_n z_n}}{\alpha_n\left(a_n e^{-\alpha_n z_n} - b_n e^{+\alpha_n z_n}\right)}; \tag{2.44}$$

$$C_n(z_{n+1}) = \frac{a_n e^{-\alpha_n z_{n+1}} + b_n e^{+\alpha_n z_{n+1}}}{\alpha_n\left(a_n e^{-\alpha_n z_{n+1}} - b_n e^{+\alpha_n z_{n+1}}\right)}. \tag{2.45}$$

Calculating a_n from 2.45 and replacing it in 2.44, with some algebra, yields,

$$C_n(z) = \frac{1}{\alpha_n} \frac{\alpha_n C_n(z_{n+1}) + \tanh\left[\alpha_n(z_{n+1} - z_n)\right]}{1 + \alpha_n C_n(z_{n+1}) \tanh\left[\alpha_n(z_{n+1} - z_n)\right]}, \tag{2.46}$$

and by applying the boundary conditions required by C in order to be continuous at the boundaries,

$$C_n(z_{n-1}) = \lim_{z \to z_n^-} C_n(z) = \lim_{z \to z_n^+} C_{n+1}(z) = C_{n+1}(z_n) \tag{2.47}$$

it is possible to directly obtain the Wait's recursion formula (cf. for example *Wait* [1954] or *Wait* [1981]):

$$C_n(z_{n-1}) = \frac{1}{\alpha_n} \frac{\alpha_n C_{n+1}(z_n) + \tanh(\alpha_n l_n)}{1 + \alpha_n C_{n+1}(z_n) \tanh(\alpha_n l_n)}. \tag{2.48}$$

Equation 2.48 is useful to compute the response of a layered Earth to an MT source, imposing the solution of a homogeneous half-space in the lowermost layer and then propagating the solution to the top of the uppermost layer. With $C_1(z_0 = 0)$ computed, it is possible to compare the value of the transfer function computed on the Earth surface directly with field data. In this context the apparent resistivity ρ_a can be introduced as a generalization of 2.37.

$$\rho_a(\omega) = |C(\omega)|^2 \mu_0 \omega = \left|\hat{Z}(\omega)\right|^2 \mu_0 \omega. \tag{2.49}$$

As the elements of $\hat{Z} \in \mathbb{C}$ it is possible to define the impedance phase ϕ representing the phase difference between $\vec{E}$ and $\vec{B}$:

$$\phi = \tan^{-1} \frac{\mathrm{Im}(\hat{Z})}{\mathrm{Re}(\hat{Z})}. \tag{2.50}$$

Apparent resistivity and phase are often used to represent magnetotelluric data. They do not add any information to the impedance tensor, but are commonly used to display data because of their relationship with physical properties. The apparent resistivity is in fact the average electrical resistively taken on the volume of the half-space that is penetrated. The phase can be expressed [*Weidelt*, 1972] as

$$\phi(\omega) = \frac{\pi}{4} - \frac{\omega}{\pi} \int_o^{+\infty} \log \frac{\rho_a(x)}{\rho_0} \frac{dx}{x^2 - \omega^2}. \tag{2.51}$$

Equation 2.51 indicates that – in an isotropic 1D domain – the ρ_a can be predicted by ϕ if the scaling factor ρ_0 is known.

2.3. Vertical Discontinuities: the 2D Case

In a more complex environment, the subsurface is not perfectly stratified, and more detailed modeling has to be undertaken. A two-dimensions domain allows for a more accurate characterization of the subsurface. In a Cartesian reference system with the z axis directed downward, the electrical conductivity may be written as

$$\sigma = \sigma(y, z), \tag{2.52}$$

being independent of the x coordinate, which defines the "strike" direction of the subsurface structures. While the fact that the strike direction has a geological or a geodynamic meaning is a matter of regional considerations, in the modeling problem the independence of σ from x simplifies the MT equations and their solution. Bearing in mind that in this configuration the x axis is parallel to the plane defined by the vertical discontinuities, it is possible to derive the MT equations for this configuration. As the fields are invariant with[1] x, from equations 2.12 and 2.13 it results:

$$\vec{\nabla} \times \vec{E} = \hat{\imath}\left(\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z}\right) + \hat{\jmath}\left(\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x}\right) + \hat{k}\left(\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y}\right) = -\hat{\imath}\,(i\omega B_x) - \hat{\jmath}\left(i\omega B_y\right) - \hat{k}\,(i\omega B_z)$$

$$\vec{\nabla} \times \vec{B} = \hat{\imath}\left(\frac{\partial B_z}{\partial y} - \frac{\partial B_y}{\partial z}\right) + \hat{\jmath}\left(\frac{\partial B_x}{\partial z} - \frac{\partial B_z}{\partial x}\right) + \hat{k}\left(\frac{\partial B_y}{\partial x} - \frac{\partial B_x}{\partial y}\right) = \hat{\imath}\,(\mu_0\sigma E_x) + \hat{\jmath}\left(\mu_0\sigma E_y\right) + \hat{k}\,(\mu_0\sigma E_z)$$

where $\hat{\imath}, \hat{\jmath}, \hat{k}$ are the unit vectors of the Cartesian reference system.

In this environment the induced electric and magnetic fields parallel to the $y - z$ plane are decoupled, so it is possible to define the E-polarization in which the electric field along the x direction induces a magnetic field in the $y - z$ plane, and a B-polarization in which the magnetic field along x induces an electric field in $y - z$. The separated equations are:

$$\text{E-polarization} \begin{cases} \frac{\partial E_x}{\partial y} &= i\omega B_z \\ \frac{\partial E_x}{\partial z} &= -i\omega B_y \\ \frac{\partial B_z}{\partial y} - \frac{\partial B_y}{\partial z} &= \mu_0\sigma E_x \end{cases} \tag{2.53}$$

$$\text{B-polarization} \begin{cases} \frac{\partial B_x}{\partial y} &= \mu_0\sigma E_z \\ \frac{\partial B_x}{\partial z} &= -\mu_0\sigma E_y \\ \frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} &= i\omega B_x \end{cases} . \tag{2.54}$$

These two polarizations are also called *TE-mode* and *TM-mode* respectively. In this 2D

[1] that means that $\frac{\partial \vec{F}}{\partial x} = 0$ for any $\vec{F}$

scenario, it is fundamental to consider the boundary condition at the contact between the two media.

Discontinuity Conditions

Consider the flux of $\vec{D}$ exiting from a cylindrical elementary volume τ in which the bases S_1 and S_2 are parallel to the surface Σ of separation between the two media, as illustrated in Figure 2.1. Integrating the equation 2.2 on the volume τ

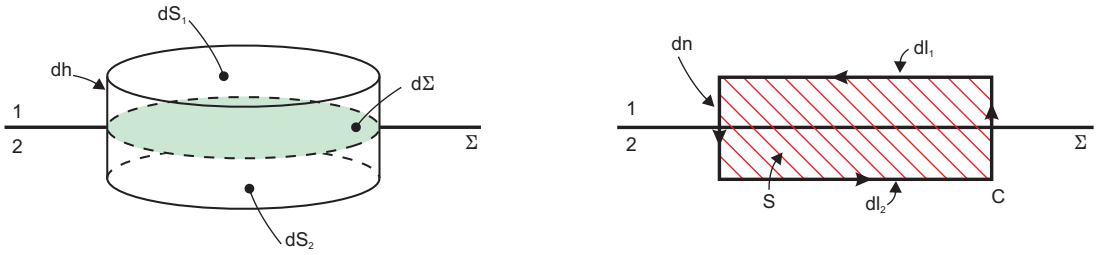


Fig. 2.1.: Left panel: elementary volume τ cutter by the boundary between medium 1 and medium 2. Right panel: its section. The $d\star$ quantities are the infinitesimal elements considered from the quantity $\star$.

$$\int_\tau \left(\vec{\nabla}\cdot\vec{D}\right) d\tau = \int_\tau \varrho d\tau \tag{2.55}$$

and using the divergence theorem on the integral in the left side returns

$$\int_\tau \left(\vec{\nabla}\cdot\vec{D}\right) d\tau = \int_S \vec{D}\cdot d\vec{S}. \tag{2.56}$$

The limit as $dh \to 0$ is:

$$(D_{1n} - D_{2n})\, d\Sigma = \sigma_\Sigma d\Sigma \tag{2.57}$$

in which D_{1n} and D_{2n} are the components on $\vec{D}$ normal to Σ and σ_Σ is the charge density per unit of surface present on Σ itself

$$(D_{1n} - D_{2n}) = \sigma_\Sigma. \tag{2.58}$$

Taking into account equation 2.4, it is possible to compute the flux of both sides through the surface S that has C as boundary:

$$\int_S \left(\vec{\nabla}\times\vec{E}\right)\cdot d\vec{S} = -\int_S \left(\frac{\partial\vec{B}}{\partial t}\right)\cdot d\vec{S}, \tag{2.59}$$

and using the Stokes theorem on the left side returns

$$\oint_C \vec{E} \cdot d\vec{l} = -\int_S \left(\frac{\partial \vec{B}}{\partial t}\right) \cdot d\vec{S}. \tag{2.60}$$

If the limit $dn \to 0$ is considered, the result is:

$$\lim_{dn \to 0} \int_S \left(\frac{\partial \vec{B}}{\partial t}\right) \cdot d\vec{S} = 0 \tag{2.61}$$

because $\left(\frac{\partial \vec{B}}{\partial t}\right)$ is finite and $S \to 0$. In the same way it is possible to compute the contact boundary conditions for the fields $\vec{H}$ and $\vec{B}$. The contact conditions are so written

$$\begin{cases} D_{1n} - D_{2n} = \sigma_\Sigma & ; \quad B_{1n} - B_{2n} = 0 \\ E_{1t} - E_{2t} = 0 & ; \quad H_{1t} - H_{2t} = J_S \end{cases} \tag{2.62}$$

in which J_S is the *surface current* measured in $A \cdot m^{-1}$.

The component of the field $\vec{D}$ – and because of the assumptions specified in the beginning of this chapter, of the field $\vec{E}$ – crossing the boundary between the two media then has to be discontinuous and the magnitude of discontinuity results in $\frac{\sigma_2}{\sigma_1}$.

Impedance Tensor

Defined as the tensor linking $\vec{E}$ and $\vec{B}$, the impedance tensor $\hat{Z}$ is commonly used in MT problems:

$$\begin{pmatrix} E_x \\ E_y \end{pmatrix} = \begin{pmatrix} Z_{xx} & Z_{xy} \\ Z_{yx} & Z_{yy} \end{pmatrix} \begin{pmatrix} \frac{B_x}{\mu_0} \\ \frac{B_y}{\mu_0} \end{pmatrix} \text{ or } \vec{E} = \frac{1}{\mu_0} \hat{Z} \vec{B}. \tag{2.63}$$

In this representation Z_{ij} is defined as the ratio $\frac{E_i}{H_j}$ and it is measured in Ω.

$\hat{Z}$ contains information about configuration of the domains in which it is computed. In a 1D domain, the tensor elements Z_{xx} and Z_{yy} are 0. In fact due to the lack of vertical gradients in conductivity, the horizontal components of $\vec{E}$ can only induce orthogonal magnetic fields and vice-versa. The relations between tensor elements are summarized in

$$\text{1D} \begin{cases} Z_{xx} = & Z_{yy} \ (\equiv 0) \\ Z_{xy} = & -Z_{yx} \end{cases} \tag{2.64}$$

$$\text{2D} \begin{cases} Z_{xx} = & -Z_{yy} \\ Z_{xy} \neq & -Z_{yx} \end{cases} \tag{2.65}$$

Note that $Z_{xx} = -Z_{yy} \equiv 0$ if x (or y) are oriented towards the geoelectric strike direction.

In a more complex domain configuration (i.e. three-dimensional (3D) structures, anisotropic media) the elements of the impedance tensor do not follow these simple conditions.

2.4. Anisotropic Forward Problem

Depending on the domain features, equations 2.25 and 2.26 can be solved to predict the electromagnetic field measured on the Earth's surface given a known electrical conductivity distribution $\sigma(x, y, z)$. In an isotropic conductor the electric current density $\vec{J}$ and the electric field $\vec{E}$ follow Ohm's law. In this configuration $\vec{E}$ and $\vec{J}$ are always parallel, whilst in a more general case the transport properties vary according to the direction. In this last anisotropic environment the parallelism between electric field and electric current density is ensured only in three mutually orthogonal directions in which $\vec{E}$ can be decomposed. Mathematically in an anisotropic environment equation 2.7 is still valid, but in this case the tensor that links $\vec{E}$ and $\vec{J}$ is a second-rank tensor $\hat{\sigma}$ that takes the general form

$$\hat{\sigma} = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{bmatrix}. \tag{2.66}$$

The conductivity tensor $\hat{\sigma}$ is symmetric and positive definite [*Onsager*, 1931; *Yin and Weidelt*, 1999]: these features imply the existence of three mutually orthogonal eigendirections – associated with the three different eigenvalues – in which $\vec{E}$ and $\vec{J}$ are parallel. It is thus easy to find three angles α, β, γ and three eigenvalues $\sigma_1, \sigma_2, \sigma_3$ such that

$$\hat{\sigma} = \mathbf{R_z}(-\gamma)\mathbf{R_x}(-\beta)\mathbf{R_z}(-\alpha) \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{bmatrix} \mathbf{R_z}(\alpha)\mathbf{R_x}(\beta)\mathbf{R_z}(\gamma), \tag{2.67}$$

where $\mathbf{R_n}$ is the elementary rotation matrix around the axis $n \in [x, y, z]$. Explicitly

$$\mathbf{R_x}(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & \sin(\theta) \\ 0 & -\sin(\theta) & \cos(\theta) \end{bmatrix}, \tag{2.68}$$

$$\mathbf{R_z}(\theta) = \begin{bmatrix} \cos(\theta) & \sin(\theta) & 0 \\ -\sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{2.69}$$

Using this representation of the electrical conductivity it is possible to derive the solution to the MT problem in an anisotropic environment. This solution is presented in the following section.

2.4.1. 1D Anisotropic Domain

In a layered 1D medium, Maxwell's equations have to be simplified considering that in such a domain, $\frac{\partial}{\partial x} = \frac{\partial}{\partial y} \equiv 0$ for both $\vec{E}$ and $\vec{B}$. In the following, the MT anisotropic 1D solution described by *Pek and Santos* [2002] is presented. In mono-dimensional anisotropic conditions, Maxwell equations become (for each frequency ω):

$$\frac{\partial E_x}{\partial z} = i\omega\mu_0 H_y \quad , \quad \frac{\partial H_y}{\partial z} = -\sigma_{xx}E_x - \sigma_{xy}E_y - \sigma_{xz}E_z, \tag{2.70}$$

$$\frac{\partial E_y}{\partial z} = -i\omega\mu_0 H_x \quad , \quad \frac{\partial H_x}{\partial z} = \sigma_{yx}E_x + \sigma_{yy}E_y + \sigma_{yz}E_z, \tag{2.71}$$

$$H_z = 0 \quad , \quad \sigma_{zx}E_x + \sigma_{zy}E_y + \sigma_{zz}E_z = 0. \tag{2.72}$$

in each layer. These six equations can be rewritten as a system of second-order differential equations as

$$\begin{cases} \frac{\partial^2 E_x}{\partial z^2} + i\omega\mu_0 \left[\overbrace{\left(\sigma_{xx} - \frac{\sigma_{xz}\sigma_{zx}}{\sigma_{zz}}\right)}^{A_{xx}} E_x + \overbrace{\left(\sigma_{xy} - \frac{\sigma_{xz}\sigma_{zy}}{\sigma_{zz}}\right)}^{A_{xy}} E_y \right] = 0 \\ \frac{\partial^2 E_y}{\partial z^2} + i\omega\mu_0 \left[\overbrace{\left(\sigma_{xx} - \frac{\sigma_{xz}\sigma_{zx}}{\sigma_{zz}}\right)}^{A_{yx}} E_x + \overbrace{\left(\sigma_{xy} - \frac{\sigma_{xz}\sigma_{zy}}{\sigma_{zz}}\right)}^{A_{yy}} E_y \right] = 0. \end{cases} \tag{2.73}$$

Due of the symmetry of $\hat{\sigma}$, it is trivial to see that $A_{xy} \equiv A_{yx}$, thus the EM field measured on the surface depends only on the three values A_{xx}, A_{xy}, A_{yy}. Moreover, if these aggregated conductivities are arranged in a 2×2 matrix $\mathbf{A}$, it can be seen as an effective conductivity tensor. In fact it is not possible to distinguish changes in $\hat{\sigma}$ elements if $\mathbf{A}$ elements are unchanged. $\mathbf{A}$ is again symmetric and positive definite, thus it is possible to write it in terms of its principal components and an effective rotation around the z axis. In matrix form:

$$\begin{pmatrix} A_{xx} & A_{xy} \equiv A_{yx} \\ A_{yx} \equiv A_{xy} & A_{yy} \end{pmatrix} = \begin{pmatrix} \cos\beta_s & -\sin\beta_s \\ \sin\beta_s & \cos\beta_s \end{pmatrix} \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix} \begin{pmatrix} \cos\beta_s & \sin\beta_s \\ -\sin\beta_s & \cos\beta_s \end{pmatrix}, \tag{2.74}$$

therefore, the MT problem in a 1D anisotropic layered medium can be simplified to an equivalent problem of azimuthal electrical anisotropy in which in each layer two principal directions are defined – the high- and low-conductive field propagation directions[2] – and are mutually orthogonal. The effective conductivity A_1 takes place along the direction indicated by β_s if the matrix $\mathbf{A}$ is arrayed so that $A_1 > A_2$.

[2] waves propagate downwards.

More information about the conductivity tensor would require knowledge of elements violating the approximations made to derive the MT fundamental equations; i.e. source, intrinsic Earth magnetism and so forth.

To compute the MT field on the Earth's surface, system 2.73 must be solved. The general solution is in the form of $e^{\pm\alpha z}$, and two modes relating to two different values of α can be found:

$$\alpha^2_{1,2} = -\frac{i\omega\mu_0}{2}\left[A_{xx} + A_{yy} \pm \sqrt{\left(A_{xx} + A_{yy}\right)^2 + \overbrace{4A_{xy}A_{yx}}^{2A^2_{xy}}}\right], \tag{2.75}$$

or, written in terms of $A_{1,2}$,

$$\alpha^2_{1,2} = -\frac{i\omega\mu_0}{2}\left(A_1 + A_2 \pm |A_1 - A_2|\right). \tag{2.76}$$

In the isotropic case, $A_{xx} = A_{yy} = \sigma$ and $A_{xy} = A_{yx} \equiv 0$, so that $\alpha^2 = i\omega\mu_0\sigma$ and the value of α is thus equivalent to the value just found for α in equation 2.32: in that case the horizontal components of electric field are completely decoupled. If the $\alpha^2_{1,2}$ values do not collapse to a single value, the horizontal components of electric field are linked by relations 2.73. By substituting $E_x = e^{\pm\alpha z}$ in one of the equations 2.73, it is easy find that $E_y = \frac{i\omega\mu_0 A_{xy}}{\alpha^2_{1,2}+i\omega\mu_0 A_{yy}}E_x$. The MT field in a homogeneous anisotropic layer of the 1D domain can be expressed , at depth z, in matrix form, after some algebra, as follows:

$$\overbrace{\begin{pmatrix} E_x \\ E_y \\ H_x \\ H_y \end{pmatrix}}^{\mathbf{F}(z,\omega)} = \overbrace{\begin{pmatrix} e^{+\alpha_1 z} & e^{-\alpha_1 z} & e^{+\alpha_2 z} & e^{-\alpha_2 z} \\ \frac{i\omega\mu_0 A_{xy}}{\alpha_1^2+i\omega\mu_0 A_{yy}}e^{+\alpha_1 z} & \frac{i\omega\mu_0 A_{xy}}{\alpha_1^2+i\omega\mu_0 A_{yy}}e^{-\alpha_1 z} & \frac{i\omega\mu_0 A_{xy}}{\alpha_2^2+i\omega\mu_0 A_{yy}}e^{+\alpha_2 z} & \frac{i\omega\mu_0 A_{xy}}{\alpha_2^2+i\omega\mu_0 A_{yy}}e^{-\alpha_2 z} \\ -\frac{\alpha_1 A_{xy}}{\alpha_1^2+i\omega\mu_0 A_{yy}}e^{+\alpha_1 z} & \frac{\alpha_1 A_{xy}}{\alpha_1^2+i\omega\mu_0 A_{yy}}e^{-\alpha_1 z} & -\frac{\alpha_2 A_{xy}}{\alpha_2^2+i\omega\mu_0 A_{yy}}e^{+\alpha_2 z} & \frac{\alpha_2 A_{xy}}{\alpha_2^2+i\omega\mu_0 A_{yy}}e^{-\alpha_2 z} \\ -\frac{i\alpha_1}{\mu_0\omega}e^{+\alpha_1 z} & \frac{i\alpha_1}{\mu_0\omega}e^{-\alpha_1 z} & -\frac{i\alpha_2}{\mu_0\omega}e^{+\alpha_2 z} & \frac{i\alpha_2}{\mu_0\omega}e^{-\alpha_2 z} \end{pmatrix}}^{\mathbf{M}(z,\omega)} \overbrace{\begin{pmatrix} D_1^+ \\ D_1^- \\ D_2^+ \\ D_2^- \end{pmatrix}}^{\mathbf{D}} \tag{2.77}$$

where the vector $\mathbf{D}$ contains constraints that scale the amplitude of the downgoing ($D^-_\star$) and the upgoing ($D^+_\star$) wave modes in the layer. The elements of $\mathbf{D}$ have to be fixed to respect the boundary conditions. In the lowermost layer, only the downgoing wave exists, thus $D^+_{1,N+1} = D^+_{2,N+1} \equiv 0$. Moreover, at the top of each layer l the longitudinal field

components must be continuous, so that:

$$\mathbf{F_l}(z_l, \omega) = \mathbf{F_{l+1}}(z_l, \omega), \tag{2.78}$$

therefore

$$\mathbf{D_l} = \mathbf{M_l^{-1}}(z_l, \omega)\mathbf{M_{l+1}}(z_l, \omega)\mathbf{D_{l+1}}, \text{ with } l = 0, 1, ..., N. \tag{2.79}$$

It is easy in this notation to propagate the field value from the the top of the basement to the *l-th* layer

$$\begin{aligned}\mathbf{F_l}(z, \omega) = \mathbf{M_l}(z, \omega)\mathbf{D_l} = \overbrace{\mathbf{M_l}(z, \omega)\mathbf{M_l^{-1}}(z_l, \omega)}^{\mathbf{S_l}(z_l - z, \omega)}\mathbf{M_{l+1}}(z_l, \omega)\mathbf{D_{l+1}} = ... = \\ = \mathbf{S_l}(z_l - z, \omega)\left[\prod_{n=l+1}^{N}\mathbf{S_n}(h_n, \omega)\right]\mathbf{M_{N+1}}(z_N, \omega)\mathbf{D_{N+1}} = \mathbf{S_l^{\Pi}}(z, \omega)\mathbf{D_{N+1}}\end{aligned} \tag{2.80}$$

in which $\mathbf{S_n}(h_n, \omega) = \mathbf{M_n}(z_{n-1}, \omega)\mathbf{M_n^{-1}}(z_n, \omega)$, $h_n = z_n - z_{n-1}$ and $z \in (z_{l-1}, z_l)$. The only two unknowns of equation 2.80 are the two elements of $\mathbf{D}$ relative to the upgoing waves $D^-_{\star,N+1}$. These values can be fixed by imposing, for example, the surface boundary condition. If, for instance, the magnetic field components are known at the surface $z_0 = 0$, the solution to the system

$$\begin{cases} H_x(0, \omega) = \mathbf{S_1^{\Pi}}(z_0, \omega)_{3,2}D^-_{N+1,1} + \mathbf{S_1^{\Pi}}(z_0, \omega)_{3,4}D^-_{N+1,2} \\ H_y(0, \omega) = \mathbf{S_1^{\Pi}}(z_0, \omega)_{4,2}D^-_{N+1,1} + \mathbf{S_1^{\Pi}}(z_0, \omega)_{4,4}D^-_{N+1,2} \end{cases} \tag{2.81}$$

allows for determination of all the elements of the vector $\mathbf{D}$ and thus to completely solve the problem.

It can be proven [*Pek and Santos*, 2002; *Simpson and Bahr*, 2005; *Jones*, 2012] that the solution of the one-dimensional anisotropic problem produces an impedance tensor with the form

$$\hat{Z} = \begin{pmatrix} Z_{xx} & Z_{xy} \\ Z_{yx} & Z_{yy} \end{pmatrix} = \begin{pmatrix} Z_{xx} & Z_{xy} \\ Z_{yx} & -Z_{xx} \end{pmatrix}. \tag{2.82}$$

The tensor representation is formally identical to the 2D isotropic case presented in equation 2.65. The main difference is that while in the 2D isotropic case it is always possible in principle to find a strike angle θ such that $Z_{xx} = Z_{yy} \equiv 0$, in the 1D anisotropic case. This rotation can be found if and only if in the layers exists only one single value for the effective strike direction $\beta_{s,i}$. This condition makes the problem symmetric with respect to the plane defined by the axis z and the high-conductive (or low-conductive) direction. If more than one value exists for i.e. $\beta_{s,i} \neq \beta_{s,i+1}$ for at least a certain i, then it is not possible find any planar symmetry and no θ exists that reduces the 1D anisotropic impedance tensor to an equivalent 2D one.

2.4.2. 2D Anisotropic Domain

The strategy used to derive the equations that describes MT fields in an anisotropic two-dimensional problem have been analyzed in detail in several classic and recent papers [*Reddy and Rankin*, 1975; *Pek and Verner*, 1997; *Li*, 2002]. Maxwell's equations become, along the strike direction (x axis):

$$\begin{aligned}\frac{\partial^2 E_x}{\partial y^2} + \frac{\partial^2 E_x}{\partial z^2} + i\omega\mu_0\left(\sigma_{xx} + \frac{S_z\sigma_{xy}}{\sigma_{yy}\sigma_{zz}-\sigma_{yz}^2} + \frac{S_y\sigma_{xz}}{\sigma_{yy}\sigma_{zz}-\sigma_{yz}^2}\right)E_x + \\ + i\omega\mu_0 S_y\frac{\partial H_x}{\partial y} - i\omega\mu_0 S_z\frac{\partial H_x}{\partial z} = 0\end{aligned} \tag{2.83}$$

$$\begin{aligned}\frac{\partial}{\partial y}\left(\frac{\sigma_{yy}}{\sigma_{yy}\sigma_{zz}-\sigma_{yz}^2}\frac{\partial H_x}{\partial y}\right) + \frac{\partial}{\partial z}\left(\frac{\sigma_{zz}}{\sigma_{yy}\sigma_{zz}-\sigma_{yz}^2}\frac{\partial H_x}{\partial z}\right) + \frac{\partial}{\partial y}\left(\frac{\sigma_{yz}}{\sigma_{yy}\sigma_{zz}-\sigma_{yz}^2}\frac{\partial H_x}{\partial z}\right) + \frac{\partial}{\partial z}\left(\frac{\sigma_{yz}}{\sigma_{yy}\sigma_{zz}-\sigma_{yz}^2}\frac{\partial H_x}{\partial y}\right) + \\ -\frac{\partial(S_y E_x)}{\partial y} + \frac{\partial(S_z E_x)}{\partial z} + i\omega\mu_0 H_x = 0,\end{aligned} \tag{2.84}$$

where $S_y = (\sigma_{yz}\sigma_{yx} - \sigma_{yy}\sigma_{xz})$, $S_z = (\sigma_{xz}\sigma_{yz} - \sigma_{zz}\sigma_{yx})$.

Given the partial differential equations 2.83 and 2.84, the MT problem is completed by defining the boundary conditions at infinity and at each of the subdomain discontinuities. The classic solution follows the hypothesis that the discontinuities in electrical conductivity are bounded in the finite region that defines the model. Far from this region, the boundary conditions are defined by solutions of 1D sections at the edges of the model. Inside the domain, across a discontinuity, the relations summarized in equations 2.62 require the continuity of

- E_x
- $\frac{\partial E_x}{\partial n}$
- H_x
- $\frac{\left[\left(\sigma_{yy}n_y+\sigma_{yz}n_z\right)\frac{\partial H_x}{\partial y}+\left(\sigma_{yz}n_y+\sigma_{zz}n_z\right)\frac{\partial H_x}{\partial z}\right]}{\sigma_{yy}\sigma_{zz}-\sigma_{yz}^2} - \left(S_y n_y - S_z n_z\right)E_x$

where $n = (0, n_y, n_z)$ is the unit vector orthogonal to the discontinuity. Equations 2.83 and 2.84, jointly with the boundary conditions completely define the problem for the electromagnetic field's component along the strike direction E_x, H_x. The other components can be computed via

$$\begin{aligned} H_y &= -\frac{i}{\omega\mu_0}\frac{\partial E_x}{\partial z}, & E_y &= \frac{\sigma_{yz}}{\sigma_{yy}\sigma_{zz}-\sigma_{yz}^2}\frac{\partial H_x}{\partial y} + \frac{\sigma_{zz}}{\sigma_{yy}\sigma_{zz}-\sigma_{yz}^2}\frac{\partial H_x}{\partial z} + S_z E_x \\ H_z &= \frac{i}{\omega\mu_0}\frac{\partial E_x}{\partial z}, & E_z &= \frac{\sigma_{yy}}{\sigma_{yy}\sigma_{zz}-\sigma_{yz}^2}\frac{\partial H_x}{\partial y} - \frac{\sigma_{yz}}{\sigma_{yy}\sigma_{zz}-\sigma_{yz}^2}\frac{\partial H_x}{\partial z} + S_y E_x \end{aligned} . \tag{2.85}$$

In the two-dimensional anisotropic case the structure of the MT impedance tensor is identical to the 3D case. Given that the components of the field are coupled via the elements of $\hat{\sigma}$, it is not possible to decouple the two modes because of the lack of symmetry in the problem.

The explicit solution of the anisotropic 2D MT problem is performed numerically. In order to produce a solution, a code from Pek and Santos, described in *Pek and Verner* [1997]; *Pek et al.* [2003] has been used as an engine for the inversion scheme presented in this thesis. An important point is that, after the projection of the problem on a numerical mesh, the problem itself can be solved by solving a system of linear equation in the form

$$\mathbf{A}\mathbf{x} = \mathbf{b} \tag{2.86}$$

where $\mathbf{A}$ is a banded matrix of coefficients, $\mathbf{b}$ contains the problem boundary conditions and $\mathbf{x}$ is the vector containing the field values at each mesh point. The linear form obtained by the use of finite-difference methods is extremely useful. In fact, the standard method used to compute the parametric sensitivities with respect to the subdomains conductivities[3] is to differentiate directly the system 2.86 respect to the parameter p, obtaining

$$\mathbf{A}\frac{\partial \mathbf{x}}{\partial p} \equiv \mathbf{A}\mathbf{x}^p = -\frac{\partial \mathbf{A}}{\partial p}\mathbf{x} + \frac{\partial \mathbf{b}}{\partial p} \equiv \mathbf{r}^p, \tag{2.87}$$

a system characterized by the same coefficients matrix $\mathbf{A}$ that characterizes the finite-differences system 2.86. If a Gaussian elimination procedure is used to solve 2.86, the complete elimination procedure for the matrix $\mathbf{A}$ needs to be carried only once, storing the result to be used to solve both systems 2.86 and 2.87, only by changing the right hand side in the back-substitution process.

2.5. Conductivity in Earth Materials

The interest in knowledge of properties, and specifically in electrical properties of Earth materials, is driven from the desire to understand the composition, and more in general the dynamics, that rules the internal part of our planet. In an inverse problem solution it is important to know the limits of the estimated parameters. Due that the conductivity is a physical parameter, it is bounded by its physical nature, and in the following the limits of the electrical conductivity values are explored.

Due that the conductivity is defined by the Ohm's law

$$\mathbf{J} = \hat{\sigma}\mathbf{E} \tag{2.88}$$

where $\hat{\sigma}$ is the conductivity tensor, $\mathbf{E}$ electric field and $\mathbf{J}$ density of current. In the next sections the main ways of conduction of electrons will be presented and results from laboratories experiments on these mechanisms will be presented. It is important to bear in mind that $\mathbf{J}$ in 2.88 depends on pressure and temperature, as $\hat{\sigma}$ is a function of physical state, then environment conditions are essential to define conductivity $\hat{\sigma}$. In other terms, it

[3] which are needed to solve the 2D MT inverse problem, refer to Chapter 4

is essential to consider the depth at which we are viewing a certain material and a certain way of conduction.

These two kinds of information, material properties and conduction mechanisms, can be gathered to an effective idea on the value on conductivity $\hat{\sigma}$ in the different layers and in the different materials.

2.5.1. Charge Transport in Crustal and Mantle Materials

The mechanisms of conduction in materials derive from the materials' physical and chemical properties It is possible discriminate different mechanisms of charge conduction, each one of these mechanisms is predominant with respect to the others in determinate conditions. MT measurements show that the Earth's crust has zones of different conductivity variating only with depth coexisting with structures that have laterally heterogeneities in electrical conductivity. Some of these structures are recognizable in the mantle as well.

In the upper crustal layer, rocks are mainly characterized by pores and fractures. Under these conditions, the electrolytic process is the dominant mechanism of conduction; in fact, pores in rocks are easily filled with electrolytic fluids. In the environment of pressure p and temperature T of the upper crust, salinities of fluids can vary from few grams per liter to more than 150 g of NaCl per l. In these conditions the electrolytic conduction is the most important mechanism of conduction [*Nover*, 2005].

In special geological environments, like fault structures, high electrical conductivity is assumed to be caused by electronic conduction through interconnected graphite films. In these cases, it is debated if the high value of electrical conductivity is caused by the the presence of saline liquid inside the rocks or by the graphite films on the fault edges. Both these models could justify the high electrical conductivity in faults and the debate regarding the most important source for conductivity is open.

Increasing in depth, with the increases in p, T in the lower crust and upper mantle, semi-conduction is the dominant transport mechanism. It is strongly dependent on T and on oxygen partial pressure, by the importance of oxygen in oxidizing processes that controls the number of charges available for semi-conduction processes, for example the quantity of Fe^{3+} or Fe^{2+} ions.

In volcanic zones, it may happens that T can be high enough to cause conductivity anomalies. These can be explained either by the formation of partial melts or by the presence of saline fluids, in which conductivity is definitely higher than in solid rocks. These values of T can be anyway considered realistic just for volcanic areas or at deep depths in the mantle (cf. e.g. *Shankland and Waff* [1977]).

In the next sections details about this transport mechanisms will be discussed.

2.5.2. Electrolytic Conduction

Electrolytic conduction, also called ionic conduction, occurs when ions under the influence of an electric field $\vec{E}$ are free to move in a liquid (more rarely in a solid). Micro-

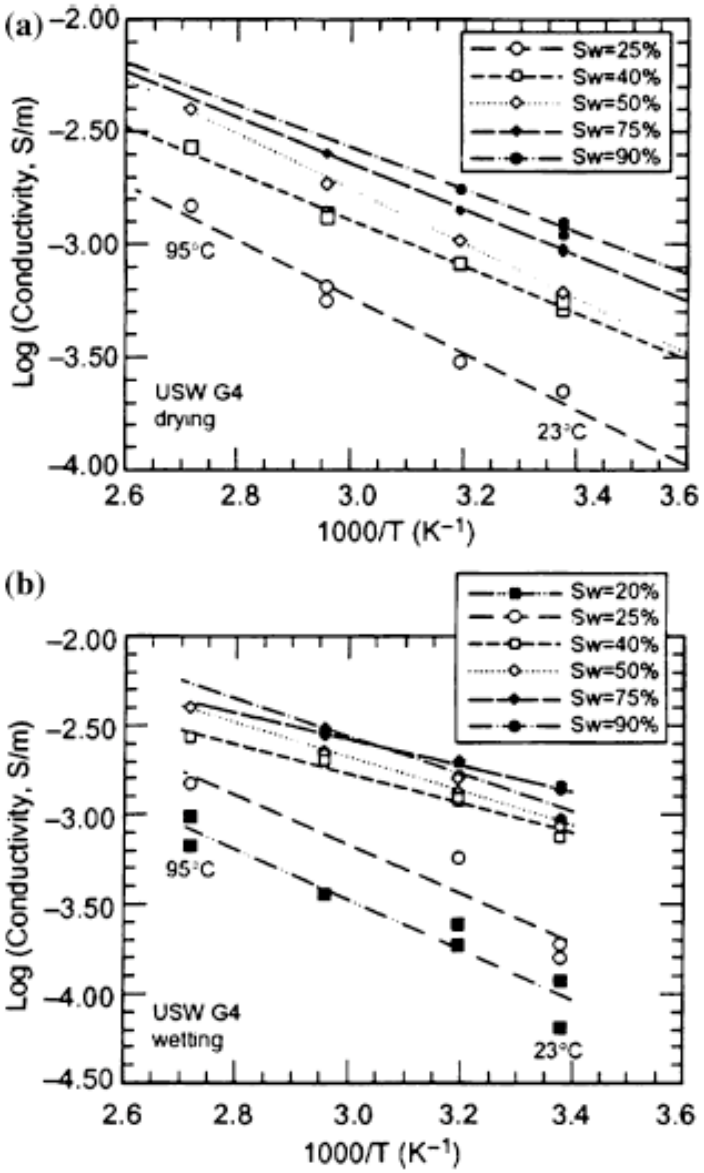


Fig. 2.2.: $\log \sigma$ versus reciprocal temperature for a tuff sample between 23 and 95^oC, for both wetting(a) and drying (b) at varying saturations. Activation energies range between 0.15 and 0.28eV[*Roberts*, 2002]

structures that exist in rocks in the upper crust are assumed to be filled by electrolytic fluids. The magnitude of the electrolytic conduction is influenced by the following factors:

concentration of ions: electrolytic conduction increases as the concentration of ions in the electrolytic solution increases;

pressure: high pressure conditions can modify the geometry of pores and fractures. In these conditions, the system of pores and electrolytic fluid may be treated as capacitors;

temperature: temperature influences the solubility of salts in the electrolytic fluid.

Electrolytic liquids can affect the value of electrical conductivity σ as shown in laboratory experiments [*Roberts*, 2002]. In these tests, the electrical resistivity $\rho = \frac{1}{\sigma}$ of densely welded tuff (Topopah Spring tuff) was measured as a function of saturation by water between 23^o and 145^oC. Measurements in the temperature window $23°C < T < 95°C$ at ambient pressure and measurements up to $145°C$ at pressures up to $5.5MPa$ were takenin an externally heated pressure vessel. Pore and confining pressures were controlled independently, allowing electrical measurements as the sample was subjected to boiling conditions. At low saturation $\leq 20\%$ an exponential dependency of resistivity is reported [*Roberts*, 2002].

$$\rho = \rho e^{CS_W} \tag{2.89}$$

where C is a constant and S_W is the saturation. Results of these measurements are reported in Figure 2.2

2.5.3. Electronic Conduction

Electronic conduction may increase the electrical conductivity by orders of magnitude if the medium in which it happens, like ores or graphite, form an interconnected network. In that cases, from 2.88, σ can be calculated directly as

$$\sigma = \frac{|\mathbf{J}|}{|\mathbf{E}|} \tag{2.90}$$

The origin of carbon at great depths is not clear. Greenschist facies p, T conditions were found to be sufficient to transform carbon into graphite, but the source of carbon is still uncertain. These conditions are found at pressure $p < 0.6\ GPa$, $T > 500°C$ [*Nover et al.*, 2005]. The electrical properties of graphite depends on the structural arrangement of the Basis Structural Units (BCU). One of the reasons supposed to explain anisotropy in electrical conductivity is caused by the bonding forces within the hexagonal carbon rings [*Nover et al.*, 2005]. Carbon gas, such as CO_2 and CH_4, are supposed to be source

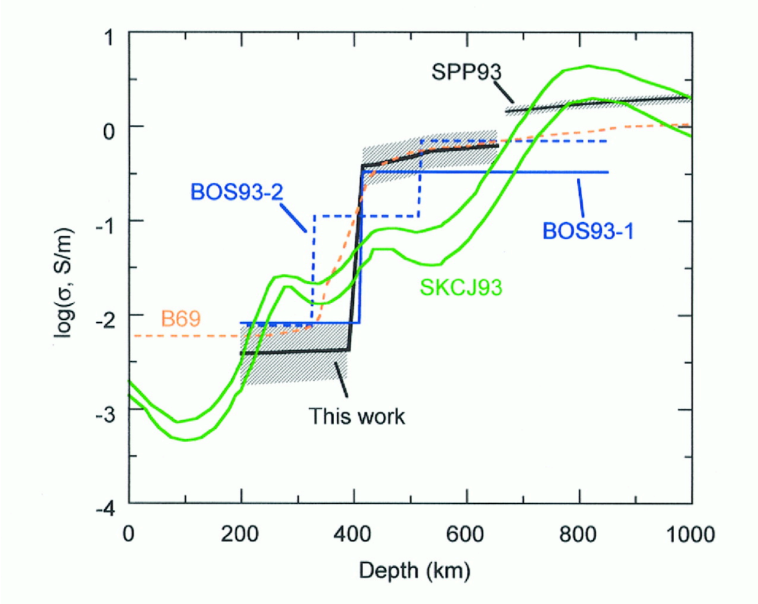


Fig. 2.3.: Conductivity model for the upper mantle calculated from laboratory data (black line) with the addiction of low mantle curve for perovskite+magnesiowustite having bulk iron fraction $\frac{Fe}{Fe+Mg} = 0.11$. Shaded areas illustrate the effect on the model of a $\pm 100^o$ temperature variation. Refer to [*Xu et al.*, 1998] for other information.

of graphite at depth, but the causes and mechanisms of their reactions are not completely clear.

2.5.4. Semi-conduction

Within the mantle, olivine – the principal constituent of the mantle itself – presents a behaviour of semi-conduction. In these areas, models obtained by MT sounding measurements present discontinuities and jumps in conductivity values up to 1.5 orders of magnitude or more[4]. In figure 2.3 these measurements are shown as reported in *Xu et al.* [1998]. At higher temperature gradient conditions, the increased thermal energy can permit some electrons in the crystalline lattice to jump from the valence band into the conduction band in the solid structure. This leaves a hole in the valence band that acts as another charge carrier. In this situation the current density $\mathbf{J}$ is given by

$$\mathbf{J} = \mathbf{J_e} + \mathbf{J_h} = (-q)n_e\mathbf{v_e} + qn_h\mathbf{v_h} \tag{2.91}$$

where q is the electron charge, n_e, n_h and $\mathbf{v_e}$, $\mathbf{v_h}$ are respectively the densities and velocities of electron and hole charge carriers. If the mobilities of holes and electrons $\mu_h = \frac{|\mathbf{v_h}|}{|\mathbf{E}|}; \mu_e =$

[4] it depends on the model

$\frac{|\mathbf{v_e}|}{|\mathbf{E}|}$ are considered, we obtain for the density of current

$$\mathbf{J} = q(n_e\mu_e + n_h\mu_h)\mathbf{E} \tag{2.92}$$

If only the charges provided by semi-conduction are considered – $n = n_h = n_e$ – the Equation 2.92 become

$$\mathbf{J} = qn(\mu_e + \mu_h)\mathbf{E} = \sigma\mathbf{E}. \tag{2.93}$$

The dependence of σ on T is then easy to show; in fact the concentration of charge carriers n depends on temperature as follow

$$n = C(k_B T)^{\frac{2}{3}} e^{\frac{-E_g}{2k_B T}} \tag{2.94}$$

in which C is a constant that depends on the materials, k_B is the Boltzmann's constant and E_g is the energy band gap.

These equations describe the condition of transport in semiconductors in a minimum conduction regime. In fact, if other charge carriers are available the value of σ may be significantly higher, depending directly on the density of charge carriers and indirectly on temperature.

Olivine-Spinel

From the early seventies, the importance of oxygen fugacity and temperature in processes of charges transport in olivine was clear [*Duba and Nicholls*, 1973]. In laboratory experiments with olivine containing $8 - 12\%$ of fayalite, as is supposed to be in typical mantle olivine, temperature is the main parameter that controls the number and mobility of charge carriers, while oxygen fugacity dominates the process of oxidation of Fe^{2+} to Fe^{3+}. This oxidation makes more charge carriers available in the form of both electrons and holes [*Duba and von der Gönna*, 1994]. In figure 2.4 results of these experiment are summarized.

To obtain these results, the presence of water in the rocks is needed to have oxidation reactions that allow the presence of holes and electrons.

With regards to the dependance of σ on frequency of the source field, it has been shown that for olivine and synthetic forsterite and fayalite, in the range between 1 to 100 kHz, the influence of the grain boundary effects may be neglected if the temperature is above 800^oC. For frequency below 100 Hz, electrodes influence the measurements [*Nover*, 2005]. Fayalite does not show any energy dispersion in range between 100Hz-20kHz and temperatures up to 1200^oC. Measurements of electrical conductivities and thermoelectric effect S[5] as functions of temperature in the range from 1000^o to 1500^oC and oxygen

[5] Seebeck effect is the phenomena of the conversion of temperature differences directly into electricity. More precisely the potential difference is given from the relation $V = \int_{T_2}^{T_1}[S_B(T) - S_A(T)]dT$ where

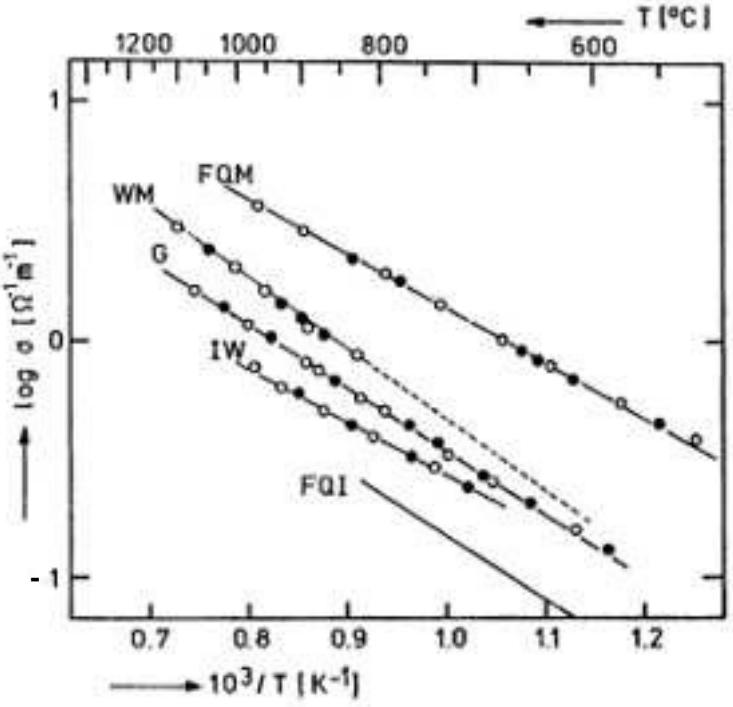


Fig. 2.4.: Arrhenius diagram of the conductivity of fayalite at various oxygen partial pressure in field of olivine. Reducing conditions are given by buffer mixture FQI-fayalite quartz iron, whereas oxidizing conditions were established using a fayalite quartz magnetite (FQM) buffer mixture. WM-wustite magnetite; G-Graphite; IW-iron wustite(von der Gonna1997)[*Nover*, 2005]

partial pressure in the range from 10^{-10} to 10^{4} Pa were done. The two most striking observations are strong conductivity anisotropy in forsterite and a sign change in S in olivine at 1390°C. These results are interpreted to show that both materials have mixed ionic and extrinsic electronic conduction under these conditions [*Schock et al.*, 1989]. Anyhow *Constable and Roberts* [1997] published results that show that at least one other conducting defect species must be invoked to explain the observed magnitude and oxygen fugacity dependence of thermopower. An electron-polaron model cannot be made to fit the conductivity and thermopower data well, but a polaron-magnesium vacancy model fits the data if a constant polaron or magnesium vacancy term is included. Consequently the transformation from polaron dominance in conduction to magnesium vacancy dominance at the threshold of about $T = 1300^{o}C$ was inferred [*Constable and Roberts*, 1997].

On the other hand, recent experiments show that, in the mantle, olivine behavior is different. In fact recent studies [*Takashi Yoshino and Katsura*, 2008] show that the main conduction mechanism in the mantle's olivine is dominated by electron conduction. In this way, the presence of water in the mantle is no longer needed. Result of this work are

S_A, S_B are the Seebeck coefficients of materials A and B and T_1, T_2 are the temperatures of the two junctions.

shown in figure 2.5.

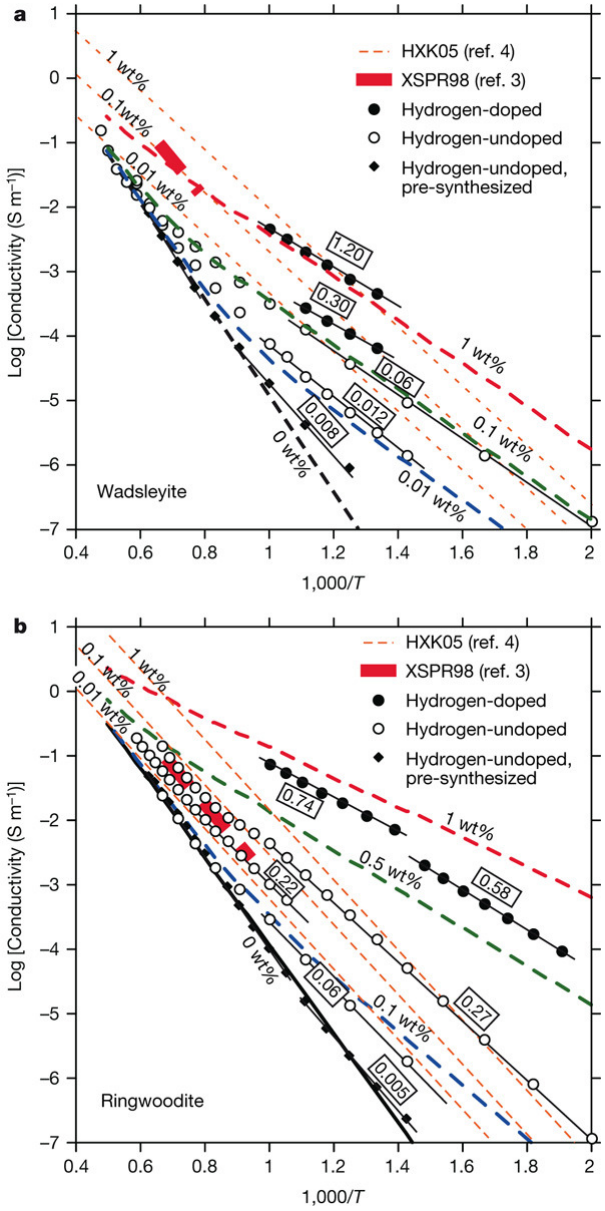


Fig. 2.5.: Electrical conductivity of waldesite and ringwoodite as a function of reciprocal temperature. a, Waldesite; b, ringwoodite. The symbols indicate raw data for each sample with different water contents. Previous results from [*Xiaoge Huang and Karato*, 2005] [*Xu et al.*, 1998] are shown as function of water content. Colored thick dashed lines indicate the electrical conductivity. Numbered boxes denote the estimate water content. [*Takashi Yoshino and Katsura*, 2008]

The presence of water is thus required to explain the oxidation reactions that allow the presence of electrons-holes pairs, but is not required to justify a conductivity value that can be caused by the electron conduction on its own.

These results lead to both a different global conductivity model, reported in figure 2.6 and to a debate between authors [*Karato and Dai*, 2009], [*Yoshino and Katsura*, 2009]

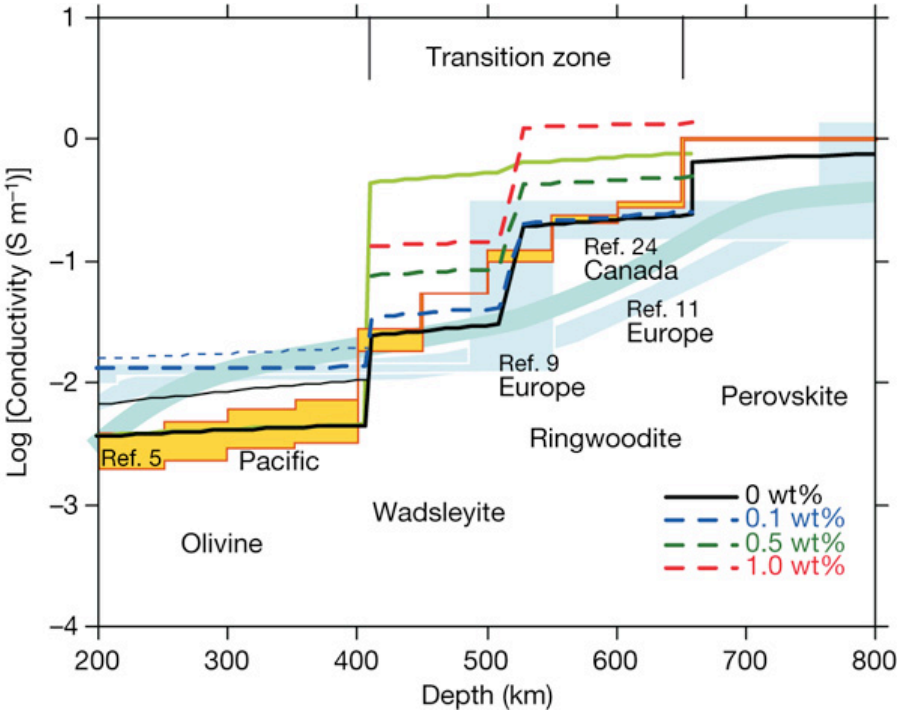


Fig. 2.6.: [*Takashi Yoshino and Katsura*, 2008] Electrical conductivity profiles beneath the Pacific, and the estimated water content in the mantle transition zone.The orange and bluish area represent geophysically observed conductivity profiles in the pacific from [*Kuvshinov et al.*, 2005] and the continental mantle from [*Olsen*, 1998], [*Tarits et al.*, 2004] and [*Neal et al.*, 2000], respectively. Light green solid line denotes previous experimental result of Xu et al.[*Xu et al.*, 1998].

2.5.5. Partial Melts

One hypothesis to explain unusual high conductivities inferred for the zone within lower crust and upper mantle - and in volcanic areas - is the existence of partially melted materials. This phenomena can occur in certain $[p, T]$ conditions, and it depends on chemical composition of partial melted minerals.

Using the effective medium theory [*Shankland and Waff*, 1974] high electrical conductivity anomalies in the upper mantle were explained as caused by partial melting [*Shankland and Waff*, 1977]. It was assumed that the melt fraction forms an interconnected network along the grain boundaries of the olivine matrix. In the same article [*Shankland and Waff*, 1977] a conductivity difference of two order of magnitude was reported if a network of this kind was present. Melt conductivities present a very low dependence on chemical composition, and they are totally independent on oxygen fugacity [*Shankland*, 1979].

The influence of melt composition was studied for olivine and a tholeiitic mid-ocean basalt [*Roberts, Jeffery J.*, 1999]. Results suggest that the larger part of variation in electrical conductivity depends on temperature rather then on composition. When the temperature is high enough to start the melting process ($> 1120^oC$) the presence of partial melts cause an increment in the measured electrical conductivity.

2.6. Field Sources

One of the most restrictive assumptions in modeling MT data is the plane wave approximation of the source. Commonly the source for MT experiments is from EM interaction that happens between ionosphere and magnetosphere and solar winds or from lightening storms.

Actually this approximation is good enough to model data only in certain zones of the Earth. Complex effects due to current flows in the ionosphere and magnetosphere may cause an appreciable distortion in data, especially if the station is close enough to latitudes close to the magnetic equator or the magnetic poles.

In the MT technique, both lightening storms and solar winds are used as primary field sources to explore the Earth's interior. In both cases, the perturbation induced in the electromagnetic field is propagated in the atmosphere. Because of this role of signal propagation media, it is essential in MT surveys to understand the complex interactions between the EM field and the atmosphere, especially in the uppermost layer of the atmosphere itself. Perturbations in the ionosphere and magnetosphere cause, of course, perturbations on the fields measured on the Earth's surface. If these effects are negligible, the plane wave approximation is convenient to model survey data.

2.6.1. Ionosphere and Magnetosphere

The ionosphere is a shell of electrons and electrically charged atoms and molecules that surrounds the Earth, stretching from a height of about 50 km to more than 1000 km. It owes its existence primarily to ultraviolet radiation from the sun.

The lowest part of the Earth's atmosphere, the troposphere, extends from the surface to about 10 km. Above 10 km is the stratosphere, followed by the mesosphere. In the stratosphere incoming solar radiation creates the ozone layer. At heights of above 80 km, in the thermosphere, the atmosphere is so thin that free electrons can exist for short periods of time before they are captured by a nearby positive ion.

Solar radiation, acting on the different compositions of the atmosphere with height, generates layers of ionization:

- The *D* region is the innermost layer that exists approximately between 50 km and 95 km above the Earth's surface. In this region the ionization is low and it is caused mainly by Lyman series hydrogen radiation ionizing nitric oxide (NO).

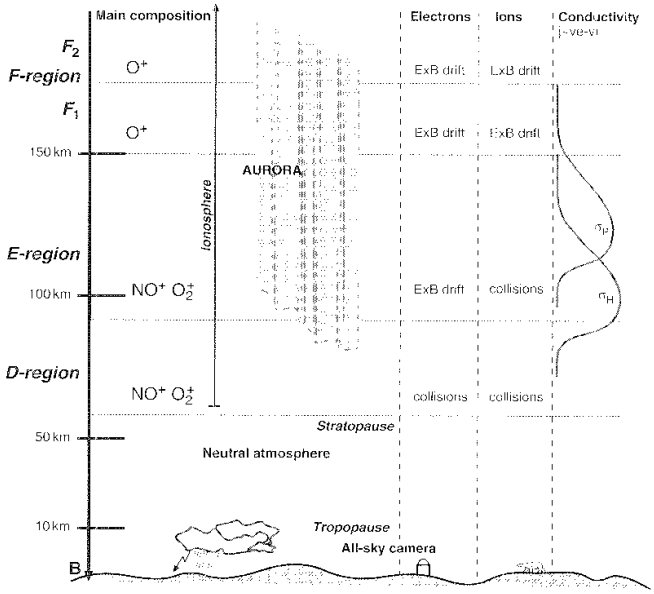


Fig. 2.7.: Regions and the main ion constituents of the ionosphere, the typical altitudes of the aurora, as well as a sketch of Hall and Pedersen conductivity profiles. Illustration by Mikko Syrjäsuo (Finnish Meteorological Institute). [*Chave and Jones*, 2012]

- The E layer is the middle layer, 90 km to 120 km above the surface of the Earth. Ionization is due to soft X-ray (1-10 nm) and far ultraviolet (UV) solar radiation ionization of molecular oxygen (O_2).

- The F layer is 120 km to 400 km above the surface of the Earth. It is the topmost layer of the ionosphere. Here extreme ultraviolet (10-100 nm) solar radiation ionizes atomic oxygen. The F layer consists of one layer at night, while in the presence of sunlight it divides into two layers, labeled F_1 and F_2. These F layers are responsible for most skywave propagation of EM signal at high radio frequencies.

Other than these layers, a rare situation of intense ionization can exist which can lead to a propagation environment close to the one caused by the E layer; it is named sporadic E layer (E_S).

The magnetosphere of Earth is a region in space whose shape is determined by the extent of Earth's internal magnetic field, the solar wind plasma, and the interplanetary magnetic field. In the magnetosphere, a mix of free ions and electrons from both the solar wind and the Earth's ionosphere is confined by electromagnetic forces that are much stronger than gravity.

In spite of its name the magnetosphere it is distinctly non-spherical. All known planetary magnetospheres in the solar system possess more of an oval tear-drop shape due to the effects of the solar wind. Planets that have an internal magnetic field, caused by core dynamics, have a protecting magnetosphere, whereas the surface of planets that do not

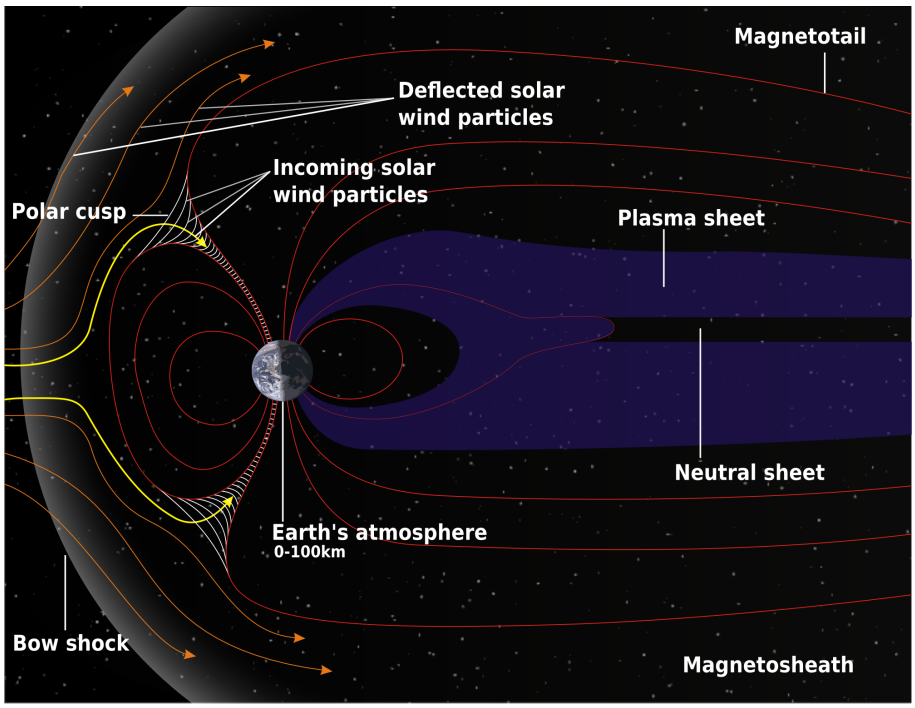


Fig. 2.8.: Typical shape of magnetosphere. It is immediately recognizable the conical symmetry around the Sun-Earth axis.

are constantly irradiated by the solar wind.

On the side facing the Sun, the distance to its boundary (which varies with solar wind intensity) is about 70,000 km (10-12 Earth radii or R_E, where 1 R_E=6371 km; unless otherwise noted, all distances here are from the Earth's centre). The boundary of the magnetosphere ("magnetopause") is roughly bullet shaped, about 15 R_E abreast of Earth and on the night side (in the "magnetotail" or "geotail") approaching a cylinder with a radius 20-25 R_E. The tail region stretches well past 200 R_E, and the way it ends is not well-know.

2.6.2. Electrojets

An electroject is an electric current which travels around the E region of the Earth's ionosphere. There are three electrojets: above the magnetic equator (the equatorial electrojet), and near the Northern and Southern Polar Circles (the Auroral Electrojets). Electrojets are currents carried primarily by electrons at altitudes from 100 to 150 km. [*Mareschal*, 1986]

These currents are particularly interesting in MT because while the MT plane wave criterion is usually satisfied, in areas near the auroral and equatorial electrojets the distortions in the fields sources (i.e. the violation of plane-wave assumption) may seriously hamper the interpretation of data [*Mareschal*, 1986].

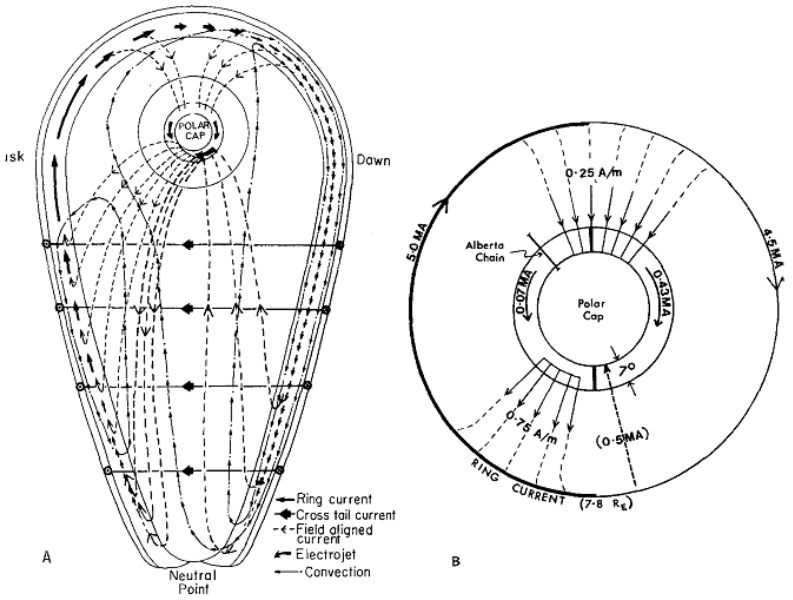


Fig. 2.9.: (A) a schematic representation of various magnetospheric currents, (B) simulation of these currents with a near-Earth model (figure from *Rostoker et al.* [1982]).

2.6.3. Modeling Strategies

The most trivial solution in modeling data affected by source distortion is selection of events to be analyzed in conjunction with robust techniques (cf. e.g. *Jones and Spratt* [2002]). While this method is acceptable in theory, there are two big consequences to face. Primarily, selection of data is done *a posteriori* and thus a lot of data may result useless [*Viljanen et al.*, 1999]. Secondly, statistically robust techniques are routinely performed, introducing a bias in an unpredictable way, so that conductivity image obtained with these techniques may return a robust result that is inconsistent with respect to the measured data.

The other strategy consists in a complex model of atmospheric electric currents, that are not laterally constant giving up the plane wave MT classical assumption [*Viljanen et al.*, 1999; *Chave and Jones*, 2012]. This technique is rarely implemented, anyway.

2.6.4. Natural EM Source from Ionosphere

In regions where the Earth can be treated as a horizontally stratified conductor, all induced currents, regardless of the actual geometry of the source, flow in horizontal planes and, thus, an ionospheric 'equivalent' current system is sufficient to represent the fields relevant to the induction process. An exact representation of the equivalent current system is not often easy to achieve, either in the wave number or space domain, and, at short periods (T less than a few hours) is only performed satisfactorily at mid-latitudes. Figure 2.9 illustrates the electric currents lying in the ionosphere and magnetosphere.

2.6.5. Diurnal Variations

Magnetic variations at all latitudes exhibit the effects of diurnal variations. A pure diurnal variation is, by definition, essentially due to DC magnetospheric currents fixed in inertial space, and gives rise to apparently oscillating fields at a given geographical position as the earth rotates within the magnetosphere [*Mareschal*, 1986].

2.6.6. Mid-Latitude Variations

At latitude below 50^o, variations are mainly caused by the solar quiet (*Sq*) current system. The major cause of these variation is the effects of atmospheric winds displacing charged particles across geomagnetic field lines, and thus generating electric fields and currents where the ionospheric conductivity is large enough.

Most of the current still appears to be concentrated in the ionosphere *E*-region [*Mareschal*, 1986].

2.6.7. Low-Latitude Variations

The existence of the daytime equatorial electrojet can essentially be explained by an extension of the *Sq* electric field to the equatorial regions, where **E** is basically parallel to the equator. In those region, **B** is virtually horizontal, and thus an effective Cowling channel[6] is created along the equator with its characteristic enhanced conductivity and currents.

Note that the equatorial electrojet current displays longitudinal gradients, with regional and seasonal variations. Such variations are known to be strongly influenced by the asymmetric way in which the *Sq* vortices intrude on the equatorial regions in the northern and southern hemispheres, and must therefore be reproducible by any realistic electrojet model. A good model must also be able to generate the disturbance fields associated with the electrojet [*Mareschal*, 1986].

2.6.8. High-Latitude Variations

The problem of extending the *Sq* variations to the polar regions is augmented by the ever present effect of various current flows generated by different physical processes mapping via field lines to different parts of the magnetosphere.

A simple model of the possible interaction, illustrated by *Zanetti et al.* [1983], is reproduced in Figure 2.10. Within the electrojets themselves, the ionospheric current density **J** is given by

$$\mathbf{J} = \sigma_P \mathbf{E}_\perp + \sigma_H \frac{\mathbf{B} \times \mathbf{E}_\perp}{|\mathbf{B}|}$$

[6]Cowling channel: it is a channel of highly ionized air in which electronic currents develop readily.

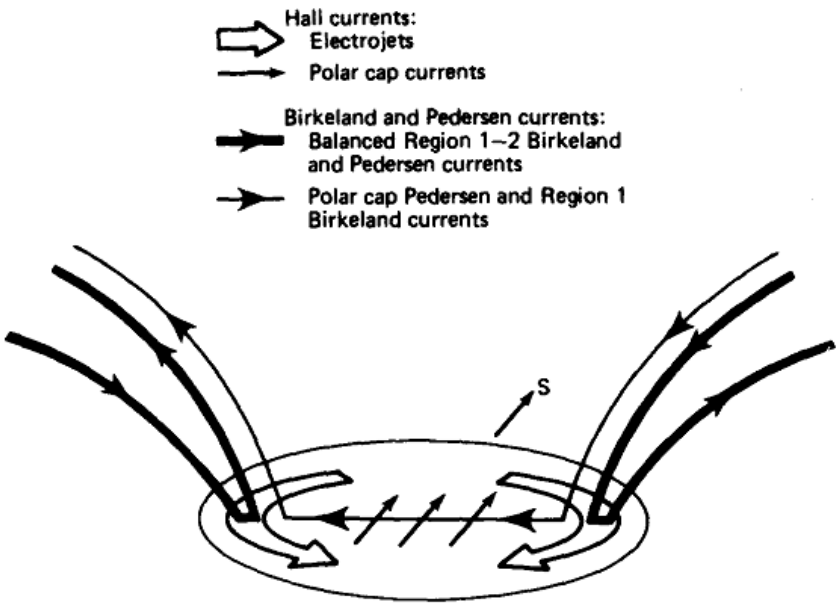


Fig. 2.10.: Possible interaction between field-aligned currents, convection electrojets and cross-polar cap current flow in the sunlit polar ionosphere [*Zanetti et al.*, 1983].

where $\mathbf{E}_\perp$ is the component of the electric field orthogonal to $\mathbf{B}$. The conductivity in the auroral zone being primarily due to precipitation of electrons energized in the magnetosphere, the conductance has a minimum in the morning sector (7-10 S) and a maximum in the midnight sector (10-20 S). The electric field pattern always reflects the large scale magnetospheric convection, $\mathbf{E}_\perp$ being primarily poleward in the region of the eastward electro jet (afternoon local time) and equator-ward in the region of the westward electrojet (morning sector).

All that probably matters in terms of modeling high-latitude localized sources for induction studies, is to know that most current systems above 55^o of geomagnetic latitude can be represented by various combinations of the two basic current loops shown in Figure 2.11 (*Mareshal*, 1984).

2.6.9. Effects on MT and GDS

In the last section lateral variations in ionospheric current are presented. All these effects cause distortion in EM field on the Earth surface. Because of this is important to keep in mind these effects using EM techniques to obtain a conductivity map.

Figure 2.12 shows an example of the differences between the response of a conductive Earth to a source of plane wave (dotted line) and a source in which westward traveling surge (WTS) is modeled (solid line).

The difference and the lack of regularity in EM response is evident [*Mareschal*, 1986].

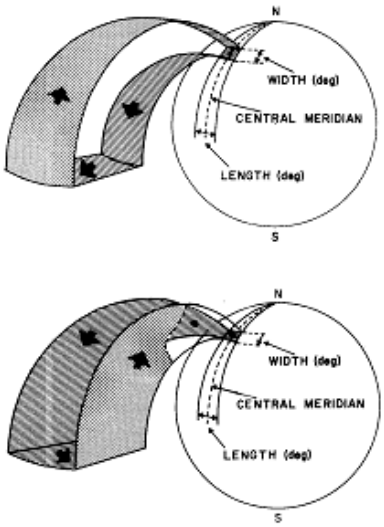


Fig. 2.11.: Models of three-dimensional current loops always present in the high-latitude regions [*Kisabeth and Rostoker*, 1977]. The E-W(N-S) loop is basically Hall (Pedersen) current flow.

2.7. Dimensionality and Distortion of Magnetotelluric Data

Notwithstanding the elegant theory developed in the past decades to solve the MT problem, the schematizations used to restrict the real world in a mathematical model force to distinguish between a "good" MT response and a "distortion" effect, the good response being the part of MT data that follows the superimposed modeling schema and the distortion the deviation of real data from the schema chosen to model the problem. Of course, the primary step to perform to obtain a reliable model for the subsurface conductivity is the selection of an appropriate domain in which to solve the problem.

The formal way used to distinguish between MT responses and distortion makes use of the formal properties of the impedance tensor highlighted in equations 2.64, 2.65 and 2.82. Far away from the goals of this thesis is an exploration of the huge production of tools developed in the years to discriminate and remove the distortion from MT data. Bearing in mind that the problem exists, we appreciate the relation of theoretical impedance tensor. A general approach to the problem of distortion recognition and removal can be found in *Simpson and Bahr* [2005]. A more detailed analysis is presented by Alan G. Jones in *Chave and Jones* [2012] and references therein. As a general rule, before proceeding to model MT data, the so-called data dimensionality has to be fixed. This process is largely empirical, due that the condition expressed by the relations 2.64, 2.65 and 2.82 are necessary to model data in respectively 1D, 2D and 1D anisotropic domain, but they are not sufficient to ensure that the chosen dimensionality is adequate to model the data themselves. In this thesis, when real data from central Germany are modeled, the distor-

tion removal procedure developed by *Jones* [2012] has been used.

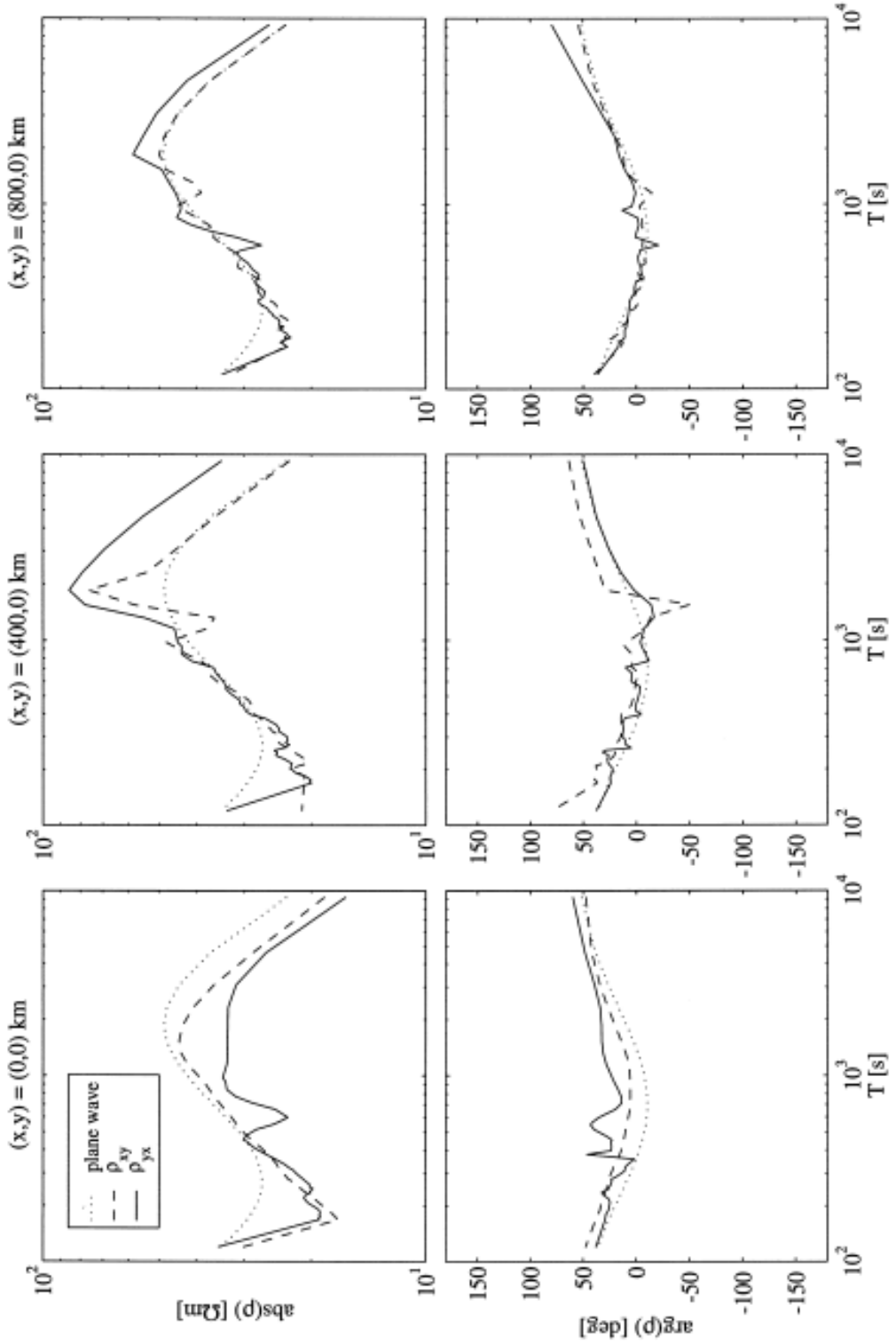


Fig. 2.12.: Absolute values and phases of the apparent resistivities ρ_{xy} and ρ_{yx} as functions of period in the case of the WTS model. The plane wave value is plotted as a dotted line.

3 From Field Measurements to Impedance Tensors

3.1. Data Processing

So far, the theory supporting MT modeling has been described without any detail relevant to field measurements. This section will describe MT data processing and data analysis. MT data collected during a typical survey consist of a time record of electromagnetic field components, particularly a record of $\mathcal{S}(t)$ is sampled, with the signal $\mathcal{S} \in [E_x; E_y; B_x; B_y; B_z]$. In an ideal situation, the treatment of these data is straightforward.

- Divide the recorded data into segments of equal lengths, the length depends on the lowest frequency needed;
- multiply each segment by an appropriate tapering function in order to minimize the spectral leakage (cf. e.g. *Harris* [1978]);
- (Fourier) transform each segment;
- compute the auto– and cross–spectrum relative to each segment;
- compute the mean and variance relative to each spectral component and
- compute the elements of the impedance tensor with relative errors.

In terms of implementation this procedure can be performed with a number of different techniques, each developed to solve a particular theoretical or practical request. In the following section the straight–forward basic theory is presented, referring to other authors' works when needed in order to present solutions to some of the well-known problems affecting MT data processing. The component relative to each required frequency ω_i is

computed via the mean of the discrete Fourier transform (DFT)s of each segment x_l

$$\mathcal{S}(\omega_i) = \sum_{k=0}^{\frac{N}{L-1}} \sum_{l=k\cdot L}^{(k+1)\cdot L-1} x_l e^{-i\omega_i l}. \tag{3.1}$$

Due to the fact that the multiplication of the signal with a boxcar function – needed to produce the segments – in the frequency domain is equivalent to a convolution of the transformed signal with a sinc[1] function that causes spectral leakage, the DFT is computed by multiplying each segment by an appropriate window function. Usually the standard window functions used in signal processing, namely the *Hann window* and the *Humming window* [*Nuttall*, 1981] are sufficient to avoid spectral leakage in the transformation of EM time-series.

With the field components expressed in the frequency domain, the estimation of the impedance tensor components is obtained by the solution of

$$E_x = Z_{xx}H_x + Z_{xy}H_y \tag{3.2}$$

$$E_y = Z_{yx}H_x + Z_{yy}H_y \tag{3.3}$$

Since every physical measurement is affected by noise we have to estimate errors for Z_{ij} ($i, j \in [x; y]$) and there will not be any couple Z_{ii}, Z_{ij} that can satisfy exactly 3.2 or 3.3. In 1971 *Sims et al.* [1971] suggested the minimization, in a least–squares sense, of all the points from the plane represented by the Equation 3.2 (or 3.3). In the same paper *Sims et al.* [1971] gave the expression of the impedance tensor elements in term of DFT of fields components, namely

$$Z_{xy} = \frac{\overline{H_xH_x^\star} \cdot \overline{E_xH_y^\star} - \overline{H_xH_y^\star} \cdot \overline{E_xH_x^\star}}{\overline{H_xH_x^\star} \cdot \overline{H_yH_y^\star} - \overline{H_xH_y^\star} \cdot \overline{H_yH_x^\star}}, \tag{3.4}$$

with similar expressions given for the other impedance tensor elements. These complicated expressions are chosen because they are not biased by random noise on the electric channel, but only by the random noise affecting the magnetic channel. This choice is made because the magnetic channels are usually less affected by random noise, and the estimations of Z_{ij} based on the expression 3.4 are considered to be more reliable.

A number of techniques have been developed over the years in order to produce the best estimation possible of the MT transfer functions [*Chave and Jones*, 2012], the following are the most–commonly used:

- In 1979 *Gamble et al.* [1979a] presented the so-called remote reference method: when more than one instrument are recording at the same time, it is possible to obtain improved quality transfer functions. The magnetic field is, in the plane–wave approximation, either equal at any point on the surface (in a 1D domain) or

[1] the sinc function is defined as $\mathrm{sinc}(x) = \frac{\sin x}{x}$

changes coherently (in a 2D domain). When the noise of the magnetic field relative to the remote site is orthogonal to the noise of the magnetic field relative to the local site it is possible to compute unbiased estimates [*Gamble et al.*, 1979b].

- *Jones and Jödicke* [1984] proposed a coherence–based method. The weight of each segment is then given from the coherence between the electric and magnetic fields, that under ideal circumstances should be $\simeq 1$.
- The reduction of the weight of outliers improves the robustness of the described procedure, i.e. it is less influenced by segments that for whatever reasons fall far from the mean [*Egbert and Booker*, 1986; *Chave and Thomson*, 1989].
- In 2003 *Chave and Thomson* [2003] introduced the bounded-influence method to identify and reduce the influence of problematic leverage points that bias the estimated transfer functions.

3.2. The Distortion Problem

As stated in Section 2.7, the term *distortion* is used to indicate any effect that distorts the behavior of a dataset with respect to the superimposed modeling scheme and that is not caused by the field sources approximation. In detail, since the MT fields are related to both the electric and the magnetic natural fields, it is possible to operate an initial distinction between galvanic (i.e. electric field) and inductive (i.e. magnetic field) distortion. Both these effects are, in principle, non-negligible and may lead to major problems in terms of MT interpretation in different depth ranges. *Jiracek* [1990] suggested that in fact while the galvanic distortion effect saturates as the frequency increases, the inductive distortion effect will increase to saturation as the frequency decreases.

3.2.1. Galvanic Distortion Effects

Let us imagine the existence of a small, finite–dimension, conductive (or resistive) heterogeneity embedded in a uniform medium. The exiting electric field – the primary field – forces a charge accumulation on the discontinuity boundary. This accumulation of charges is the source of a secondary field that distorts the primary electric field. This effect is known as galvanic distortion, and depends mainly on the electrical conductivity of the heterogeneity [*Jiracek*, 1990].

3.2.2. Inductive Distortion Effects

The currents induced in the embedded body by the primary electromagnetic field are, as described by the Equation 2.4 (the Faraday-Neumann Law), sources of a secondary magnetic field that distorts the primary magnetic field.

3.3. Data Analysis

The process that removes the removable distorting effects takes the name of *data analysis*.

In the cases in which it is correct to neglect the inductive distortion, the measured impedance tensor can be described [*Groom and Bahr*, 1992] as

$$\hat{Z}_{meas} = \hat{C}\hat{Z}_{und} \tag{3.5}$$

where $\hat{Z}_{meas}$ is the measured impedance tensor, $\hat{Z}_{und}$ the undistorted impedance tensor, and $\hat{C}$ the telluric distortion tensor. Distorting effects are thus characterized by the rank-two tensor $\hat{C}$ that does not depend on frequency [*Groom et al.*, 1993].

Because the solution of Equation 3.5 is non-unique, it is not possible to solve it autonomously. *Groom and Bailey* [1989] proposed to decompose the tensor $\hat{C}$ into determinable and indeterminable parts, using the Pauli matrices $\mathbf{\Sigma_n}$ for the spin description as the tensor base:

$$\hat{C} = g\hat{T}\hat{S}\hat{A} \tag{3.6}$$

in which

- g is a scalar, scale factor, called site *gain*
- $\hat{T} \propto [\mathbf{I} - i(t\mathbf{\Sigma_2})]$ is a rotation tensor
- $\hat{A} \propto (\mathbf{I} + s\mathbf{\Sigma_3})$ is a tensor that distorts field elements along the principal axis direction of the anisotropy
- $\hat{S} \propto (\mathbf{I} + e\mathbf{\Sigma_1})$ is a tensor that develops anisotropy on an axis that bisects the regional inductive principal axis
- e, t are tangents of characteristic angular deformation and s is the parameter that characterizes the distortion anisotropy.

In this way it is possible to consider all the different factors that influence distortion.

$\hat{T}$ performs only a rotation of the undistorted induction tensor, orientating the coordinate system of $\hat{Z}_{und}$ to coincide with the coordinate system of $\hat{Z}_{meas}$. Figure 3.1 shows the effect of anisotropy $\hat{A}$, in picture (b), over the vector field shown in picture (a). It is the same for Figure 3.2 for the $\hat{S}$ matrix.

The Groom-Bailey decomposition is characterized by twelve unknowns, namely g, t, e, s, plus the real and imaginary parts of the four impedance tensor elements, and eight knowns, the real and imaginary parts of the measured tensor elements. Because the only frequency-dependent unknowns are the ones relative to the distorted impedance tensor, in the special cases in which it is possible to reduce the amounts of frequency-dependent unknowns[2]

[2] e.g. because of the particular structure of the impedance tensor, cf. for example the Equations 2.64 or 2.65

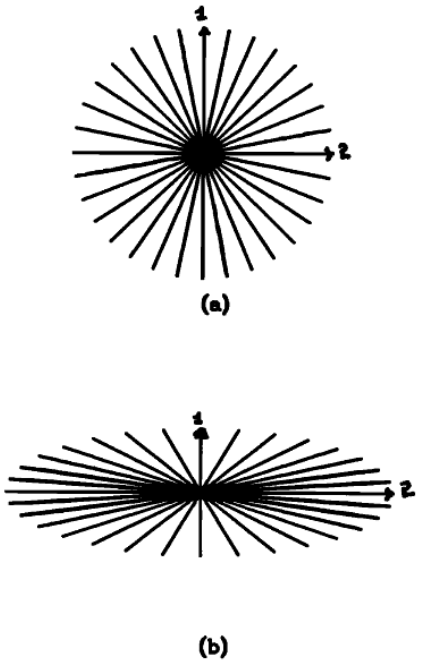


Fig. 3.1.: A effect, $s > 0$

it is possible, in principle, to solve system 3.5 and determine the undistorted MT fields except for the gain factor, g.

So far, the simple theory underlying the Groom-Bailey decomposition has been presented. For the particular problem of 2D structures (a good compromise between real Earth subsurface description and numerical problems arising in inverse problems with many parameters), the impedance tensor as presented in Equation 2.65 has been studied and discussed by several authors (cf. e.g. *Mcneice and Jones* [2001], *Becken and Burkhardt* [2004]). A complete description of the methods used to remove the distortion from data is available in *Chave and Jones* [2012] and references therein.

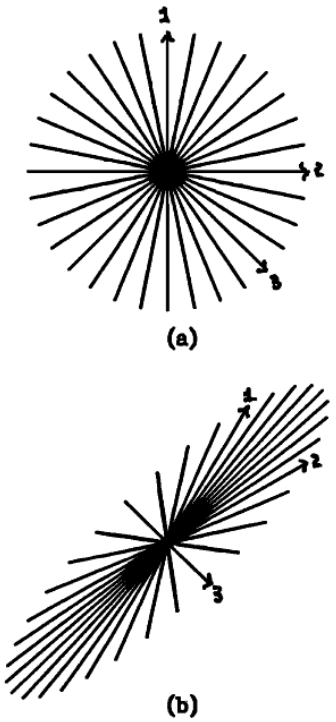


Fig. 3.2.: **S** effect, $e > 0$

4

Inverse Problem Theory in the Context of Geophysics Problems

4.1. Inverse problems

Inverse problems are a class of mathematical problems routinely solved in all physics and engineering environments, with an incredible number of different techniques. This kind of problem arises when the real physical world is schematized in a set of parameters characterizing a **model** m.

In the following, it will be assumed that a physical theory describing the studied phenomena is properly developed and understood, so, the knowledge of the model parameters m allows the prediction of a set of data d through the theory represented by the G operator[1], in the formula:

$$G(m) = d. \tag{4.1}$$

Of course, in the real world there is no "perfect experiment", so every measure of a physical quantity is affected by an amount of uncertainty, sometimes known or estimable, other times unknown or unestimable. The resulting problem can be cast in a similar way

$$G(m_{true}) = d_{true} + \eta, \tag{4.2}$$

in which m_{true} is the model that predicts the true data d_{true} affected by the noise η, assuming that the forward modeling (G) is exact.

In other terms, the problem of data prediction from a physical model is usually called the **forward problem**, whereas the **inverse problem** arises after the experiment, when data are collected and a model of the unknown true world is derived from collected data. The computation of $G(m)$ involves different strategies, depending on the physics theory that describes the studied process. In MT it usually involves the solution of a partial

[1] $G(\star)$ is known as *forward operator*.

differential equation (PDE) system derived from Maxwell's equations 2.2, while in other cases the relation between the model and data is expressed through an algorithm for which it is not possible to give an analytical[2] formula (an example of this is the solution of the MT problem in a 1D domain, that can be solved by the *Wait's recursion formula* 2.48).

In geophysics – and in tomography problems especially – the model is often a representation of the Earth itself, a region of its subsurface in the specific case examined here, and the parameters that form the model are the values that a certain physical parameter assumes in the region in which the subsurface is discretized. This subclass of inverse problems is referred to as **discrete inverse problems** and is usually cast as

$$G(\mathbf{m}) = \mathbf{d} \tag{4.3}$$

in which $\mathbf{m}$ is an n-element column vector of the model parameters, and $\mathbf{d}$ the m-element vector of the data predicted from the model $\mathbf{m}$.

The aim of the theory presented in the next sections is to invert the relation G and find the model $\mathbf{m}$.

4.1.1. Linear Inverse Problems

A special subset of discrete inverse problems is the set of inverse problems in which the Equation 4.3 can be written as

$$G\mathbf{m} = \mathbf{d}. \tag{4.4}$$

In this special case, the G operator can be expressed as a matrix and the solution of the inverse problem is simplified. If the G operator is sufficiently regular (i.e. the matrix representing it is a square matrix and $\det(G) \neq 0$) it is possible compute its inverse directly, via linear-algebra classic results, and the solution of the inverse problem become

$$\mathbf{m} = G^{-1}\mathbf{d}. \tag{4.5}$$

Otherwise it is often possible invert the matrix G via the generalized inverse and single value decomposition (SVD) [*Moore*, 1920; *Penrose*, 1955].

Due that the MT problem is highly non-linear (cf. Chapter 2) we refer to *Aster et al.* [2004] and references therein for a more detailed overview of problems and solutions arising in the linear inverse problems field.

[2] in this context the use of the word *analytical* is rigorous, meaning locally given by a convergent power series.

4.1.2. Difficulties in Inverse Problems Solution

While it could be straightforward to invert a simple relation and find the solution of an inverse problem, there are three main problems that make this process challenging

Existence. A set of parameters $\mathbf{m}$ so that $G(\mathbf{m})$ exactly fit the data $\mathbf{d}$ does not exist. This problem might be caused by noise in measurements or because the forward operator G is wrong or makes use of incorrect approximations to describe the physical process or even because the model is too simple.

Uniqueness. The sole existence of a solution is not sufficient to warrant that an inverse problem can be solved. There might be several – even infinite – solutions that fit $\mathbf{d}$ equally well. A classic example is the external electric field generated by a spherically symmetric charge distribution. The external electric field depends solely on the total charge and not on its radial distribution.

Instability. The solution of a problems as expressed in Equation 4.1 is often an unstable process, that means that a small change in the vector $\mathbf{d}$ leads to dramatic changes in the solution $\mathbf{m}$. For instance a small variation in the noise η in equation 4.2 could result in a total different model. Following the classic nomenclature [*Hadamard*, 1902], we call an unstable system **ill-posed** if the Equation 4.1 is in the form of a PDE system, or **ill-conditioned** if it represent a discrete linear system. The solution of instability problems is usually performed via a constraint that biases the solution, referred to as **regularization**. Other methods used to stabilize the solution of an unstable system (for example the procedure of preconditioning an ill-conditioned system of linear equation or trough a coarse discretization [*Engl, H. W. and A. Neubauer*, 1996]) will not be used in this work.

4.2. Optimization Problems

The most common way to handle the problem of the lack of existence of a solution in inverse problems is to cast the whole problem in another way. In spite of looking for the true model $\mathbf{m}_{true}$ that exactly predicts the data $\mathbf{d}_{true}$ and the noise η, an approximated solution $\mathbf{m}$ is built such that it minimizes a certain measure of the misfit between the measured data and the predicted data $G(\mathbf{m})$. A common measure of this misfit is the 2-norm of the **residual** vector, computed as

$$\mathbf{r} = \mathbf{d} - G(\mathbf{m}). \tag{4.6}$$

In order to find the model vector $\mathbf{m}$, the inverse problem is then cast as an **optimization problem**. So an **objective function** ϕ that has to be minimized is defined. Bearing in

mind the equation 4.6,

$$\phi = \sum_{i=1}^{m} \left[\frac{\mathbf{d}_i - G(\mathbf{m})_i}{\sigma_i} \right]^2 \tag{4.7}$$

in which σ_i is the error associated with the *i-th* measure.

This way of casting an inverse problem has several major advantages:

1. if ϕ is well-defined (i.e. $\sigma_i \neq \{0, \pm\infty\}\, \forall i$) it always has al least one minimum value, thus the associated inverse problem admit at least one solution, solving the problem of solution existence;

2. the quantity minimized is proportional to the χ^2, so it gains a statistical meaning[3];

3. being well-studied, specific algorithms to solve this problem have been developed, cf. e.g. *Levenberg* [1944] and *Marquardt* [1963] (Levenberg-Marquardt (LM) algorithm) or *Wolfe* [1959] (simplex method for quadratic programming).

Optimization problems are solved using a large number of algorithms, and an enumeration of these will lead far from the goal of this thesis. Nevertheless it is important to understand advantages and limits of the algorithm used in the solution of an inverse problem. In the following sections two different approaches are described:

- the LM algorithm, used in the developed code;
- the genetic algorithm (GA), tested but not used, because of slow convergence [*Mandolesi et al.*, 2009a].

4.3. Newton-type Optimization Methods

In this section, the equations governing the non-linear optimization algorithm used will be derived. It is important to understand the limitations and advantages of these approaches, in particular because we select an algorithm, the LM algorithm, in this class of optimization algorithms to solve the 2D MT anisotropic problem.

First of all it is important to understand that the lack of linearity of the forward operator G governing the MT problem leads to the necessity of using an algorithm specifically designed to solve nonlinear problems. The theory used to solve linear problems, in which the inverse problem can be cast as in Equation 4.4 is a classic of linear algebra theory and will not be discussed here.

[3]To appreciate the statistical aspects of this choice of objective function, refer to the Section 4.5

4.3.1. Newton's Method

Let us define $\mathbf{F}(\mathbf{x}) = \mathbf{0}$, a nonlinear system of n equations with n unknowns, the aim of Newton's method is to construct a sequence of vectors $\mathbf{x^0}, \mathbf{x^1}, ..., \mathbf{x^n}$ that converge to the solution $\mathbf{x}^*$. Assuming that $\mathbf{F}$ is C^∞, it is possible to expand $\mathbf{F}$ in a Taylor series about $\mathbf{x^0}$:

$$\mathbf{F}(\mathbf{x^0} + \mathbf{\Delta x}) \approx \mathbf{F}(\mathbf{x^0}) + \nabla\mathbf{F}(\mathbf{x^0})\Delta\mathbf{x} \tag{4.8}$$

in which $\nabla\mathbf{F}(\mathbf{x^0})$ represents the Jacobian

$$\nabla\mathbf{F}(\mathbf{x^0}) = \begin{bmatrix} \frac{\partial F_1(\mathbf{x^0})}{\partial x_1} & \frac{\partial F_1(\mathbf{x^0})}{\partial x_2} & \cdots & \frac{\partial F_1(\mathbf{x^0})}{\partial x_m} \\ \frac{\partial F_2(\mathbf{x^0})}{\partial x_1} & \frac{\partial F_2(\mathbf{x^0})}{\partial x_2} & \cdots & \frac{\partial F_2(\mathbf{x^0})}{\partial x_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial F_m(\mathbf{x^0})}{\partial x_1} & \frac{\partial F_m(\mathbf{x^0})}{\partial x_2} & \cdots & \frac{\partial F_m(\mathbf{x^0})}{\partial x_m} \end{bmatrix}. \tag{4.9}$$

Computing the difference between $\mathbf{x^0}$ and $\mathbf{x}^*$ we obtain

$$\mathbf{F}(\mathbf{x}^*) = \mathbf{0} \approx \mathbf{F}(\mathbf{x^0}) + \nabla\mathbf{F}(\mathbf{x^0}) \overbrace{\Delta\mathbf{x}}^{\mathbf{x}^*-\mathbf{x^0}}. \tag{4.10}$$

Thus

$$\nabla\mathbf{F}(\mathbf{x^0})\Delta\mathbf{x} \approx -\mathbf{F}(\mathbf{x^0}) \tag{4.11}$$

is the relation that allows us to perform *Newton's method*: given a guess initial solution $\mathbf{x^0}$, the following steps are repeated until necessary the precision is achieved

- solve $\nabla\mathbf{F}(\mathbf{x^k})\Delta\mathbf{x} = -\mathbf{F}(\mathbf{x^k})$.
- Let $\mathbf{x^{k+1}} = \mathbf{x^k} + \Delta\mathbf{x}$.
- Let $k = k + 1$.

Quadratic Convergence of Newton's Method

The main characteristic concerning Newton's method is that the sequence of the "errors"

$$h_1 = \mathbf{x}^* - \mathbf{x^1}, h_2 = \mathbf{x}^* - \mathbf{x^2}, ..., h_n = \mathbf{x}^* - \mathbf{x^n} \tag{4.12}$$

converge to zero quadratically, in the sense that $|h_{n+1}| \leq \mu h_n^2$ with a fixed constant μ (cf. e,g, *Courant and John* [1989]).

This quadratic convergence means that in the neighborhood of $\mathbf{x}^*$ the number of accurate digits in the solution doubles with each iteration. This quick convergence rate comes at the price of the request for continuously differentiability of $\mathbf{F}$. If this hypothesis is

not satisfied, the convergence rate of the Newton's method is slow and not proved [*Aster et al.*, 2004].

The *damped Newton's method* is a modified version of this algorithm in which the search direction is updated at each iteration. For $f(\mathbf{x})$, a scalar valued C^2 function, we can expand it as a Taylor series

$$f(\mathbf{x^0} + \Delta\mathbf{x}) \approx f(\mathbf{x^0}) + \nabla f(\mathbf{x^0})^T \Delta\mathbf{x} + \frac{1}{2}\Delta\mathbf{x}^T \nabla^2 f(\mathbf{x^0})\Delta\mathbf{x} \tag{4.13}$$

in which $\nabla f(\mathbf{x^0})$ and $\nabla^2 f(\mathbf{x^0})$ are respectively the gradient (Jacobian matrix) and the Hessian of f.

$\nabla f(\mathbf{x}^*) = \mathbf{0}$ is a necessary condition for $\mathbf{x}^*$ to be a minimum of f. In the neighborhood of x^0 it is possible to approximate the gradient by

$$\overbrace{\nabla f(\mathbf{x^0} + \Delta\mathbf{x})}^{=0} \approx \nabla f(\mathbf{x^0}) + \nabla^2 f(\mathbf{x^0})\Delta\mathbf{x} \tag{4.14}$$

thus

$$\nabla^2 f(\mathbf{x^0})\Delta\mathbf{x} = -\nabla f(\mathbf{x^0}). \tag{4.15}$$

Newton's method for minimizing $f(\mathbf{x})$ is performed by repeatedly solving equation 4.15: given an initial guess solution $\mathbf{x^0}$. The following steps are repeated until the necessary precision is achieved

- Solve $\nabla^2 f(\mathbf{x^n})\Delta\mathbf{x} = -\nabla f(\mathbf{x^n})$.
- Let $\mathbf{x^{n+1}} = \mathbf{x^n} + \Delta\mathbf{x}$.
- Let $n = n + 1$.

An important method on its own, Newton's method for minimizing $f(\mathbf{x})$ is needed in the derivation of the Levenberg-Marquardt method, the algorithm chosen to solve the MT problem in this thesis.

4.3.2. The Levenberg-Marquardt Algorithm

Consider the non-linear problem of fitting a data vector $\mathbf{d}$, consisting of m measures, by forward modeling of the model $\mathbf{m}$. A vector of standard deviations σ relative to the data is also known. The aim of an optimization algorithm in this context is to find a model $\mathbf{m}$ so that the 2-norm of the residuals is minimized. We also suppose that the measurements errors are normally distributed. Our problem becomes to minimize the objective function

$$\phi(\mathbf{m}) = \sum_{i=1}^{m} \left(\frac{G(\mathbf{m})_i - d_i}{\sigma_i} \right)^2 . \tag{4.16}$$

Let us write

$$\phi_i(\mathbf{m}) = \frac{G(\mathbf{m})_i - d_i}{\sigma_i}, i = 1, 2, ..., m \tag{4.17}$$

so

$$\Phi(\mathbf{m}) = \begin{bmatrix} \phi_1(\mathbf{m}) \\ \phi_2(\mathbf{m}) \\ \vdots \\ \phi_m(\mathbf{m}) \end{bmatrix}, \tag{4.18}$$

and

$$\phi(\mathbf{m}) = \sum_{i=1}^{m} \phi_i(\mathbf{m})^2. \tag{4.19}$$

The gradient $\nabla\phi$ can be written as the sum of the gradients of the individual terms

$$\nabla\phi(\mathbf{m}) = \sum_{i=1}^{m} \nabla\left(\phi_i(\mathbf{m})^2\right); \tag{4.20}$$

and the elements of the gradient are

$$\nabla\phi(\mathbf{m})_j = \sum_{i=1}^{m} 2\nabla\phi_i(\mathbf{m})_j \Phi(\mathbf{m})_j. \tag{4.21}$$

In matrix notation it is possible to write

$$\nabla\phi(\mathbf{m}) = 2\mathbf{J}(\mathbf{m})^T \Phi(\mathbf{m}) \tag{4.22}$$

in which $\mathbf{J}$ is the Jacobian matrix

$$\mathbf{J}(\mathbf{m}) = \begin{bmatrix} \frac{\partial\phi_1(\mathbf{m})}{\partial m_1} & \frac{\partial\phi_1(\mathbf{m})}{\partial m_2} & \cdots & \frac{\partial\phi_1(\mathbf{m})}{\partial m_n} \\ \frac{\partial\phi_2(\mathbf{m})}{\partial m_1} & \frac{\partial\phi_2(\mathbf{m})}{\partial m_2} & \cdots & \frac{\partial\phi_2(\mathbf{m})}{\partial m_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial\phi_m(\mathbf{m})}{\partial m_1} & \frac{\partial\phi_m(\mathbf{m})}{\partial m_2} & \cdots & \frac{\partial\phi_m(\mathbf{m})}{\partial m_n} \end{bmatrix}. \tag{4.23}$$

In the same way, the Hessian can be expressed as

$$\nabla^2\phi(\mathbf{m}) = \sum_{i=1}^{m} \nabla^2\left(\phi_i(\mathbf{m})^2\right) = \sum_{i=1}^{m} \mathbf{H}^i(\mathbf{m}) \tag{4.24}$$

where $\mathbf{H}^i(\mathbf{m})$ is the Hessian of $\phi_i(\mathbf{m})^2$. The j-th, k-th element of $\mathbf{H}^i(\mathbf{m})$ can be expressed as

$$H^i_{j,k}(\mathbf{m}) = \frac{\partial^2 \left(\phi_i(\mathbf{m})^2\right)}{\partial m_j \partial m_k} \tag{4.25}$$

$$= \frac{\partial}{\partial m_j}\left(2\phi_i(\mathbf{m})\frac{\partial \phi_i(\mathbf{m})}{\partial m_k}\right) \tag{4.26}$$

$$= 2\left(\frac{\partial \phi_i(\mathbf{m})}{\partial m_j}\frac{\partial \phi_i(\mathbf{m})}{\partial m_k} + \phi_i(\mathbf{m})\frac{\partial^2 \phi_i(\mathbf{m})}{\partial m_j \partial m_k}\right), \tag{4.27}$$

thus

$$\nabla^2\phi(\mathbf{m}) = 2\mathbf{J}(\mathbf{m})^T\mathbf{J}(\mathbf{m}) + 2\sum_{i=1}^{m}\phi_i(\mathbf{m})\nabla^2\phi_i(\mathbf{m}). \tag{4.28}$$

Ignoring the term $2\sum_{i=1}^{m}\phi_i(\mathbf{m})\nabla^2\phi_i(\mathbf{m})$ leads to the *Gauss-Newton method*; the Hessian is then approximated by

$$\nabla^2\phi(\mathbf{m}) \approx 2\mathbf{J}(\mathbf{m})^T\mathbf{J}(\mathbf{m}). \tag{4.29}$$

Substituting the gradient 4.22 and the Hessian 4.29 into equation 4.15 and dividing both sides by 2, the equation that leads to the solution of the Gauss-Newton method becomes

$$\mathbf{J}(\mathbf{m})^T\mathbf{J}(\mathbf{m})\Delta\mathbf{m} = -\mathbf{J}(\mathbf{m})^T\Phi(\mathbf{m}). \tag{4.30}$$

The Levenberg-Marquardt Algorithm is a classic modification of the Gauss-Newton method [*Levenberg*, 1944], [*Marquardt*, 1963] in which the equation to be solved is modified as follows

$$\left(\mathbf{J}(\mathbf{m^k})^T\mathbf{J}(\mathbf{m^k}) + \lambda\mathbf{I}\right)\Delta\mathbf{m} = -\mathbf{J}(\mathbf{m^k})^T\Phi(\mathbf{m^k}). \tag{4.31}$$

in which $\mathbf{I}$ is the identity matrix and λ a positive value called *damping parameter*. For extreme values of λ it is possible to recognize two limit behaviours for the algorithm

- $\mathbf{J}(\mathbf{m^k})^T\mathbf{J}(\mathbf{m^k}) + \lambda\mathbf{I} \approx \lambda\mathbf{I}$, if $\lambda \to +\infty$. In this case $\Delta\mathbf{m} \approx -\frac{1}{\lambda}\nabla\phi(\mathbf{m})$ and a *steepest-descent* method is achieved. The algorithm moves toward the direction that reduces the value of $\phi(\mathbf{m})$ most quickly. This kind of algorithm ensures the convergence to at least a local minimum, but lacks the efficiency in convergence of the Gauss-Newton method.

- $\mathbf{J}(\mathbf{m^k})^T\mathbf{J}(\mathbf{m^k}) + \lambda\mathbf{I} \approx \mathbf{J}(\mathbf{m^k})^T\mathbf{J}(\mathbf{m^k})$, if $\lambda \to 0^+$ leads to the original Gauss-Newton method, taking advantage of its potential fast convergence rate but with the possibility of failure in convergence at all [*Aster et al.*, 2004].

In our problem we have normalized the system performing the divisions

$$\frac{\left[\mathbf{J}(\mathbf{m^k})^T\mathbf{J}(\mathbf{m^k})\right]_{i,j}}{\left[\mathbf{J}(\mathbf{m^k})^T\mathbf{J}(\mathbf{m^k})\right]_{i,i}} \text{ and } \frac{\left[-\mathbf{J}(\mathbf{m^k})^T\Phi(\mathbf{m^k})\right]_i}{\left[\mathbf{J}(\mathbf{m^k})^T\mathbf{J}(\mathbf{m^k})\right]_{i,i}}, \forall i, j. \tag{4.32}$$

The resulting system has the same solution as the original one (as proved by the Rouché-Capelli theorem) but all the diagonal elements are normalized.

In inverse problems solution, the trade-off between fast convergence and certain convergence is usually balanced by performing a *line-search* along the direction of the step. A line search requires the computation of solutions for several forward problems, nevertheless it provides an affordable method to find the adequate value for λ. In the implementation of the LM algorithm used in this thesis, the parameter λ has been chosen iteratively with the following procedure [*Marquardt*, 1963]:

1. Fix a value for $\lambda = v^n, n \in \mathbb{N}$ to be used in the first iteration.
2. Solve the equation 4.31.
   a) If $\Phi(m^k) < \Phi(m^{k-1})$, $\lambda_{k+1} = \frac{\lambda_k}{v}$, go to 3.
   b) else $\lambda_k = v \cdot \lambda_k$. Go to 2.
3. Store the value of λ_k and exit the line search.

The selection of the dumping parameter is largely discretional. We decid to selected an initial value for λ so that results $\lambda \gg 1$, ensuring an initial convergence toward a local minimum. Using a scaling factor of $v = 2$, we decided to fix the initial value of lambda to the first available integer two order of magnitude larger than 1, resulting in the selection of 2^7 as starting value for λ.

Defining λ equal to 128 rather then 1 ensured, in each of the tests performed, that the steps in the early iterations of the algorithm favor the steepest-descent direction. This method ensures a good convergence rate with the certainty of achieving a minimum in the objective function. Beginning the LM algorithm choosing a dumping parameter $\lambda \gg 1$ ensures that the minimum reached is really the local one, i.e. the solution found is not too far from the initial guess. Moreover, the presence of the term $\lambda\mathbf{I}$ ensures the non-singularity of the term $\left(\mathbf{J}(\mathbf{m^k})^T\mathbf{J}(\mathbf{m^k}) + \lambda\mathbf{I}\right)$ and thus the possibility of its inverse computation.

4.4. Stochastic Methods

Stochastic optimization methods are optimization methods grounded in the generation of random variables. Stochastic methods used in geophysics are in the class of *metaheuristics*: methods that solve a problem by iteratively trying to improve a candidate solution. Other terms having a similar meaning as metaheuristic are: derivative-free, direct search, black-box, or indeed just heuristic optimizer. There are several reasons why algorithms

that search for a solution with a random path were developed. In geophysics, noise in data make the misfit surface extremely rough, and linearized methods cannot escape a local minimum when they find one.

A popular stochastic method for the MT problem is the genetic algorithm (GA): several implementations of GA's [*Everett and Schultz*, 1993; *Shi et al.*, 2000; *Pérez-Flores and Schultz*, 2002; *Moorkamp et al.*, 2007; *Roux et al.*, 2011] have been used in the last years to solve different MT inverse problems. We did not use the GA to solve our problem, but it is interesting to understand its mechanics and why it is not appropriate for the 2D MT anisotropic problem.

4.4.1. Monte Carlo Methods: Common Features

The search for an optimal solution in a broad model space can be extremely difficult. With the exception of small problems, a systematic search in the model space is highly inefficient, thus several algorithms that perform a random (or pseudo-random) search were designed. The name of Monte-Carlo was given to these methods by the team of Los Alamos and is nowadays generally recognized as a synonym of random search [*Tarantola*, 2005]. The main advantage of this method is the ability to avoid local minima in the objective function, coming at the cost of an intensive computation requirements. Nevertheless several Monte-Carlo methods other than GA have been used in MT inversion, including random search [*Jones and Hutton*, 1979], simulated annealing [*Sharma and Kaikkonen*, 1998], Markov chain algorithms [*Grandis et al.*, 1999] and particle swarm [*Shaw and Srivastava*, 2007]. There are several advantages to using stochastic methods in geophysics:

- stochastic methods do not requires the objective functions to be continuous and derivable;
- stochastic methods naturally fit problems in which the model space is limited;
- stochastic methods have the opportunity to avoid local minima;
- being derivative-free, there is no need to compute or store the Jacobian or Hessian matrixes.

On the other hand the necessity of an extremely large number of forward computations makes stochastic methods often inapplicable.

4.4.2. Genetic Algorithm

Genetic algorithms are a class of global search algorithms that mimic biological evolution strategies in order to find the global minimum of the objective function. One of the appreciated properties of the GA's is their ability of exploring the model space, avoiding

the localized search strategy that characterizes linearized methods. A broad description of GA's development and their main properties can be found in *Goldberg* [1989].

The general work flow for a genetic algorithm is presented in figure 4.1, with the different GA's characterized by different implementations of the steps described in the graph. The selection of parameterization influences most of the following steps because it influences the representation of model parameters in the GA. In the oldest implementations of GA's, model parameters are coded by concatenated binary strings, while more recent representation use real number encoding. The importance of the earliest representation is in the analogy between the binary string and the genetic DNA chain, making the modification operators of crossover and mutation simple and independent from the individual modeled parameters. The technique used to obtain the binary string is straightforward: given the model vector $\mathbf{m}$, the string is completely defined for each parameter m_i by the minimum value m_i^{min}, the discretization step δ_i and the number of bits used to encode the parameter n_i. So the relation between the parameter value and its binary representation will be

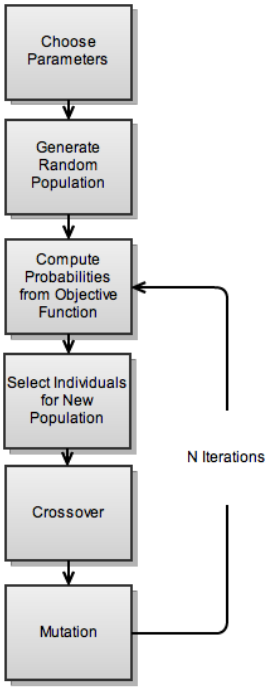


Fig. 4.1.: Program flow for genetic algorithm

$$m_i = m_i^{min} + \delta_i \sum_{j=1}^{n_i} 2^{j-1} s_j \tag{4.33}$$

in which s_j is the part of the genetic string that codes the *i-th* parameter. This representation provides both advantages and disadvantages in the solution of inverse problems. In geophysics problems it is usually possible to bound values of parameters in realistic intervals. The values of conductivities of the rocks spread in the interval[4] $\left[10^{-8}, 10^{6}\right] S \cdot m^{-1}$. Representing this range of variability is challenging and representing the interval could require a large discretization step or a high number of bits in the binary representation. Nevertheless having the possible values of the parameters bounded in the same range as the "reasonable" values, reduces the size of model space allowing the exploration of a broader part of it.

Once the encoding step provides a useful representation of the model space, a starting set of random modes – the so-called *population* – is created. The size of the population is a key factor in the algorithm, and specifically it controls the randomness of

[4] read Section 2.5 and references therein for a better understanding of these values

the starting models, completeness of model space exploration and the length of run-time of the algorithm. Once the starting population has been generated, each member of the population is unencrypted from the binary form to its real representation and its fitness is computed. Usually, in inverse problem applications, the fitness is inversely proportional to some kind of misfit measure (for example, equation 4.6). Nevertheless GA's do not require the objective function to be differentiable, so the misfit function can be modified in a huge number of ways.

From the fitness, each member of the population is assigned a probability to be chosen. Selected members' genetic strings are then saved and stored in a set called the *mating pool* in which they are randomly selected and mated to generate the next generation. This probabilistic approach helps the GA in escaping local minima. After the selection of the "parents", their genetic code is broken and assembled in pairs, producing a new population. This process is usually referred as *crossover*. Another process helps in keeping diversity and randomness in the population, the so-called *mutation*. Each element of a genetic code is associated with a small – usually less than 0.05 – probability of mutation. Mutation changes only one bit of the entire genetic string, however it helps again in escaping local minima, especially when these minima are close to the actual global minimum. In order to ensure that the fitness is not decreasing with the number of iterations, the best fitting member of the parents population is overwritten to any random member of the children population. This process is known as *elitism*.By applying the mutation process to offspring built from the crossover of fitness parents, this new population is ready to be used as starting population for a new iteration of the algorithm.

A genetic algorithm has lots of apparent advantages. It can naturally explore a discretized model space, it can avoid local minima and there is no need to know the derivatives of the objective function with respect to model parameters, saving storage space in the process. However these advantages come with some risks. The GA's "implicit" knowledge of the gradient direction does not allow searching of the model space in a direct way and moreover, while non divergence of the process is ensured by the elitism, there is not proof of convergence.

Following the steps of *Moorkamp et al.* [2007] GA has been tested as an engine for solving the optimization problem, but the approach was not followed further because of problems that will be highlighted in section 4.6.

4.5. Confidence Intervals

A classic solution of an inverse problem leads naturally to the necessity of quantifying a measure of the goodness of the model parameters estimation. *Backus and Gilbert* [1968], in their classic paper, introduce the model covariance as a tool to appraise inversion results. Both resolution and model covariance are developed to appraise the solutions of linear inverse problems, nevertheless it is possible to compute an approximation to these quantities useful for appraising the solution of a non-linear inverse problem, at least to a

first order approximation.

Analyzing the problem as in *Aster et al.* [2004], if a vector of data $\mathbf{d}$ has a multivariate normal distribution, the product $\mathbf{A}\cdot\mathbf{d}$ also has a multivariate normal distribution, assuming that $\mathbf{A}$ is a matrix of the appropriate size, and its covariance is

$$\text{Cov}(\mathbf{A}\cdot\mathbf{d}) = \mathbf{A}\cdot\text{Cov}(\mathbf{d})\cdot\mathbf{A}^T. \tag{4.34}$$

In a linear inverse problem, in which a problem has the form 4.4, if the data errors are independent and normally distributed, the maximum likelihood principle solution is the least squares solution, and the probability density for the datum d_i of $\mathbf{d}$ is

$$f_i(d_i|\mathbf{m}) := \frac{1}{\sqrt{(2\pi)}\sigma_i} e^{\frac{-(d_i-(\mathbf{Gm})_i)^2}{2\sigma_i^2}}. \tag{4.35}$$

The likelihood function L for the complete dataset is given by

$$L(\mathbf{m}|\mathbf{d}) = \frac{1}{\sqrt{(2\pi)^m}\prod_{i=1}^m \sigma_i} \prod_{i=1}^m e^{\frac{-(d_i-(\mathbf{Gm})_i)^2}{2\sigma_i^2}}. \tag{4.36}$$

Looking for the maximum of L, the maximization problem to be solved is

$$\max \prod_{i=1}^m e^{\frac{-(d_i-(\mathbf{Gm})_i)^2}{2\sigma_i^2}}, \tag{4.37}$$

treating the term $\frac{1}{\sqrt{(2\pi)^m}\prod_{i=1}^m \sigma_i}$ in 4.36 constant.

In order to solve the equation 4.37, we can apply the logarithmic function

$$\max \prod_{i=1}^m e^{\frac{-(d_i-(\mathbf{Gm})_i)^2}{2\sigma_i^2}} = \tag{4.38}$$

$$\max \log\left[\prod_{i=1}^m e^{\frac{-(d_i-(\mathbf{Gm})_i)^2}{2\sigma_i^2}}\right] = \tag{4.39}$$

$$\max -\sum_{i=1}^m \frac{-(d_i-(\mathbf{Gm})_i)^2}{2\sigma_i^2} = \tag{4.40}$$

$$\min \sum_{i=1}^m \frac{-(d_i-(\mathbf{Gm})_i)^2}{\sigma_i^2}. \tag{4.41}$$

Equation 4.41 is the solution to the problem 4.4 if each datum in the considered set is weighted by its error. In matrix notation

$$\mathbf{W} = \text{diag}(\sigma_1^{-1}, \sigma_2^{-1}, ..., \sigma_m^{-1}). \tag{4.42}$$

Equation 4.41 is the solution to the least square problem

$$\mathbf{G_w} \cdot \mathbf{m} = \mathbf{d_w} \tag{4.43}$$

where $\mathbf{G_w} = \mathbf{W} \cdot \mathbf{G}$ and $\mathbf{d_w} = \mathbf{W} \cdot \mathbf{d}$.

The least-square solution to equation 4.43 is

$$\mathbf{m}_{L_2} = \left(\mathbf{G_w}^T \mathbf{G_w}\right)^{-1} \mathbf{G_w}^T \mathbf{d_w}. \tag{4.44}$$

The covariance $\mathrm{Cov}(\mathbf{m}_{L_2})$ can therefor be computed by substituting $\mathbf{A} = \left(\mathbf{G_w}^T \mathbf{G_w}\right)^{-1} \mathbf{G_w}^T$ in equation 4.34.

The data, while weighted, have a covariance matrix equal to $\mathrm{Cov}(\mathbf{d_w}) = \mathbf{I}$ and equation 4.34 becomes

$$\mathrm{Cov}(\mathbf{m}_{L_2}) = \left(\mathbf{G_w}^T \mathbf{G_w}\right)^{-1} \mathbf{G_w}^T \mathbf{I} \left(\mathbf{G_w}^T \mathbf{G_w}\right)^{-1} = \left(\mathbf{G_w}^T \mathbf{G_w}\right)^{-1}. \tag{4.45}$$

If the errors are independent and normally distributed, the expected value of the least squares solution to 4.43 is

$$E\left[\mathbf{m}_{L_2}\right] = \left(\mathbf{G_w}^T \mathbf{G_w}\right)^{-1} \mathbf{G_w}^T E\left[\mathbf{d_w}\right]. \tag{4.46}$$

Substituting $E[\mathbf{d_w}] = \mathbf{d_{true_w}}$ and $\mathbf{G_w m_{true}} = \mathbf{d_{true}}$ in 4.46,

$$E\left[\mathbf{m}_{L_2}\right] = \left(\mathbf{G_w}^T \mathbf{G_w}\right)^{-1} \mathbf{G_w}^T \mathbf{G_w m_{true}} = \mathbf{m_{true}}. \tag{4.47}$$

It is now easy to compute any specified interval of confidence for the model parameters. Using the fact that each *i-th* element of $\mathbf{m}$, m_i, has a normal distribution with mean $\mathbf{m_{true}}$ and variance given by the $[i, i]$ element of the covariance matrix 4.45, for each specified interval of confidence C the interval of confidence is given by

$$\mathbf{m}_{L_2} \pm p_C \cdot \mathrm{diag}\left(\mathrm{Cov}(\mathbf{m}_{L_2})\right)^{\frac{1}{2}} \tag{4.48}$$

with p_C the solution of the equation

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-p_C\sigma}^{p_C\sigma} e^{-\frac{x^2}{2\sigma^2}} dx = C. \tag{4.49}$$

In classic statistical analysis, it is common to fix $C = 95\%$ that returns a value $p_C \approx 1.96$.

In the context of a non-linear problem, it is possible to consider a linear approximation as a first-order estimator of the goodness of the inverse problem result. Considering a linearization of the misfit function 4.18

$$\Phi(\hat{\mathbf{m}} + \Delta\mathbf{m}) \approx \Phi(\hat{\mathbf{m}}) + \mathbf{J}(\hat{\mathbf{m}}) \cdot \Delta\mathbf{m}, \tag{4.50}$$

a linear relation between variations in Φ and $\mathbf{m}$ can be written:

$$\Delta\Phi \approx \mathbf{J}(\hat{\mathbf{m}})\Delta\mathbf{m}. \tag{4.51}$$

If the linearization proposed is a good approximation, $\mathbf{J}(\hat{\mathbf{m}})$ may approximate $\mathbf{G}$ in an estimation of the model parameters covariance. In the derivation of the LM method the errors σ_i have been incorporated in the formula for ϕ (equation 4.16), thus $\mathrm{Cov}(\mathbf{d}) = \mathbf{I}$, leading to the approximate formal for the covariance

$$\mathrm{Cov}(\hat{\mathbf{m}}) \approx \left(\mathbf{J}(\hat{\mathbf{m}})^T\mathbf{J}(\hat{\mathbf{m}})\right)^{-1}. \tag{4.52}$$

In the case that the errors are independent and normally distributed and the standard deviations unknown, it is possible to estimate the standard deviation s as

$$s = \sqrt{\frac{\sum_{i=1}^{m} r_i^2}{m-n}} \tag{4.53}$$

where r_i is the $i-th$ residual $r_i = G(\hat{\mathbf{m}})_i - d_i$. This approximation leads to the covariance estimation for the model parameters

$$\mathrm{Cov}(\hat{\mathbf{m}}) = s^2\left(\mathbf{J}(\hat{\mathbf{m}})^T\mathbf{J}(\hat{\mathbf{m}})\right)^{-1}. \tag{4.54}$$

Knowing both $\hat{\mathbf{m}}$ and $\mathrm{Cov}(\hat{\mathbf{m}})$, it is possible to establish an interval of confidence using the same method as in equation 4.48. It is essential to bear in mind that this estimation of interval of confidence – exact in 4.48 – is only approximate in the case of non-linear inversion. It can be used as a tool to appraise the resulting model, but in the case of extremely noisy data further investigation may be required to appraise inversion results.

4.6. Linearized and Stochastic Methods Compared. Efficiency Matters.

Several tests have been performed in order to choose an optimization strategy for the problem object of this thesis. The two-dimensional anisotropic MT problem is extremely challenging and the selection of an appropriate algorithm to solve the inverse problem associated with it is a key topic of this study. In this test results from GA and LM algorithms applied to a simple non-linear problem will be compared and discussed.

4.6.1. Test 1

We need to test the efficiency of GA and LM. Dealing with an optimization problem, we define a test function ϕ to be minimized. Let be

$$\phi(x) = x + x^2 \cos(x) + 20, \tag{4.55}$$

defined on the interval $[-2\pi, 2\pi]$, our aim is to compare how many forward evaluations of the test function are needed in by the two algorithms in order to find the global minimum of the test function ϕ, presented in Figure 4.2. While ϕ maybe not realistic as a misfit function, it shares two major features with the functional minimized in a realistic inverse problem: it is positive on the interval on which is defined and it has multiple minima (two in this case, a local one located in $x^- < 0$ and a global one located in $x^+ > 0$). Being the LM method dependent on the initial guess, we expect that it is able to find the global minimum if and only if the initial guess $x_0 : \|x_0 - x^-\| \leq \varepsilon^\star$, with $\varepsilon^\star > 0$, whereas the GA performances do not depend on the initial guess selection.

Since the computing time-consuming part of a realistic problem is in the forward problem solution, we used a similar number of forward evaluations for both the algorithms.

We run a sample of 1000 LM algorithms using an uniform distributed random variable as starting guess, drawn in the interval $[-2\pi, 2\pi]$. It results that from a random starting guess, the LM method converges, averagely and in relation to this particular function, computing 52 forward evaluations[5]. An histogram of results occurrences is presented in Figure 4.3. Leaving the GA running for 52 forward evaluations – using a population of 13 individuals for 4 generations in this case – neither the global nor the local minimum were retrieved with precision. A comparison between the values retrieved is presented in Figure 4.2.

Results presented in Figure 4.2 clarify the problem relative to the quickness of convergence. While the use of the LM method could prevent us from retrieving the global minimum, the convergence to a local one is ensured, whereas the GA's results are not defined and not precise. Resolved several time with the GA the same problem, we built an histogram relative to the density of the solutions, reproduced in Figure 4.4.

A qualitatively similar results are given if the GA parameters are set to consider a smaller population and a longer evolution, as shown in Figure 4.5.

Clearly, the ability in escaping the local minimum, comes at the cost of a longer computation. Moreover, it is mandatory to remind to readers that while the LM method converges to a minimum – possibly local – the GA select a population on the basis of the misfit function. This difference is conceptually important: the LM algorithm converges, minimizing a least-square problem and finding a maximum likelihood solution, the GA retrieves a set of models that fit the problem, minimizing a function in this case, but none of these is proved to be a solution that satisfies the maximum likelihood principle.

[5]This number takes account of the forward evaluations needed to compute the misfit, to compute the Jacobian via finite-differences and to compute the line-search for λ.

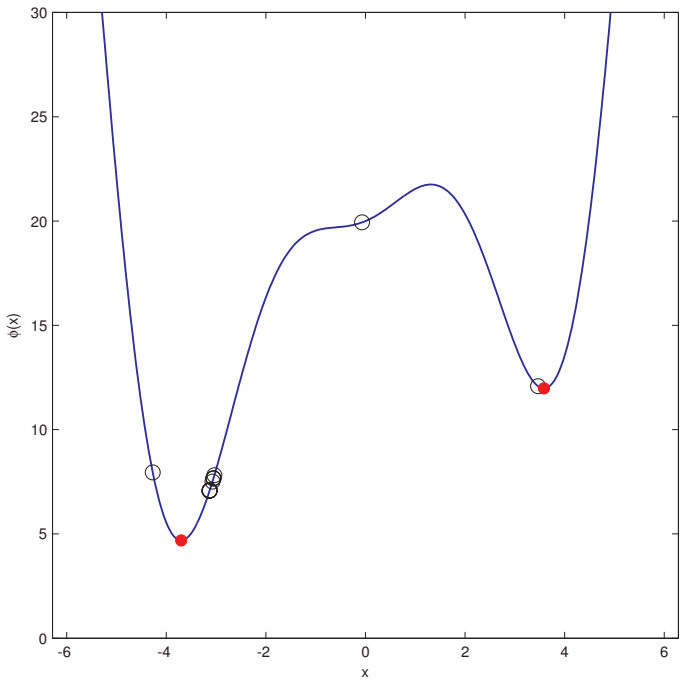


Fig. 4.2.: Results relative to minimization of function $\phi(x)$. Minima retrieved with the LM algorithm are marked with red dots, results produce making use of GA are marked by black circles. Plotted in blue: the objective function minimized in Test 1. The function is characterized by the presence of two minima, in correspondence of two value of $x = \{x^+, x^-\}$, a local one, located on the positive *x-axis* ($x^+ > 0$), and a global one located on the negative *x-axis* ($x^- < 0$)

The main result from this simple example is the evidence of a longer computing time – approximated by the number of forward problems solved – required by the GA respect to the LM method to converge.

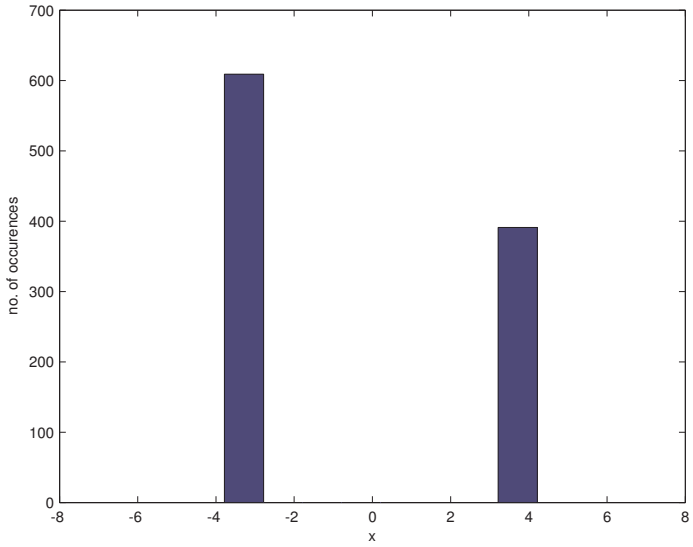


Fig. 4.3.: Distribution of the solutions to the problem proposed in the Test 1, computed with the LM algorithm using a random starting guess drawn from the interval $[-2\pi, 2\pi]$.

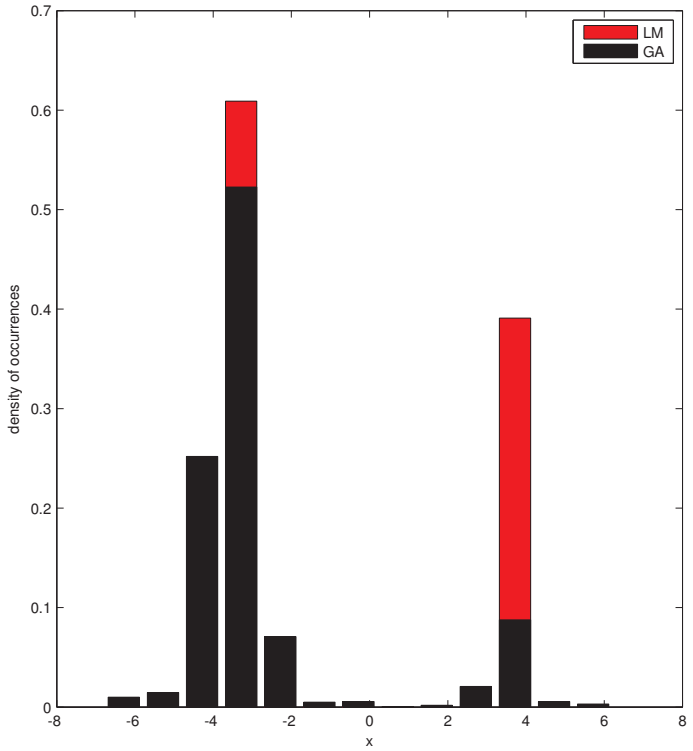


Fig. 4.4.: Density of solutions relative to the Test 1. In black the density of occurrences relative to the GA, in red the ones relative to the LM algorithm. Densities are computed on results from a sample of 1000 runs.

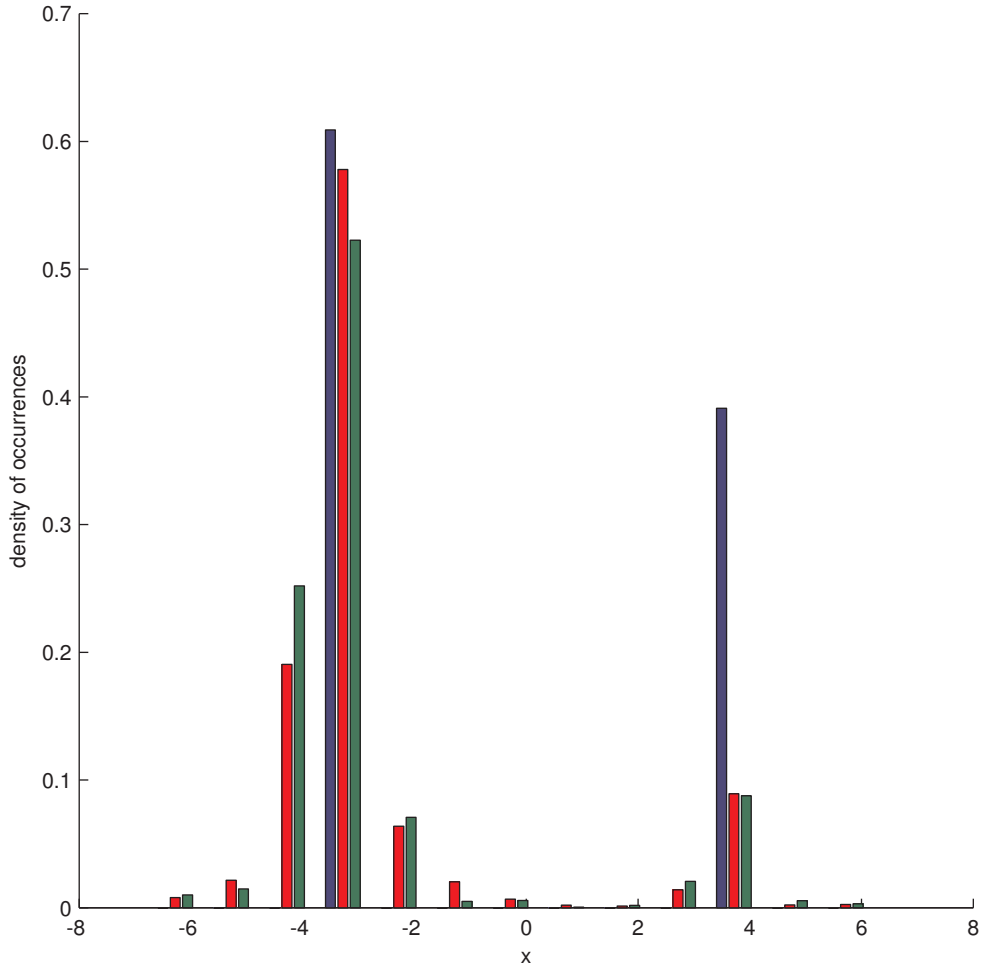


Fig. 4.5.: Density of solutions relative to the Test 1. In red and green marked the densities of occurrences relative to the GA (red bars: population size = 4, no. of generations = 13. Green bars: population size = 13, no. of generations = 4), in blue the one relative to the LM algorithm. Densities are computed on results from a sample of 1000 runs.

4.6.2. Test 2

The first test presented was able to show that an higher number of forward evaluations is required by the genetic algorithm, respect to the Levenberg-Marquardt method in order to individuate a minimum, meanwhile for the GA is not proved the convergence strictu sensu.

In this test a real RMS function is minimized from a synthetic dataset, obtained from a modified version of the function 4.55. Let us suppose that a certain problem can be described by the equation

$$d_i = m_1 x_i + m_1 x_i^2 \cos(m_2 x_i), \tag{4.56}$$

where d_i is the *i-th* datum predicted in correspondence of x_i, as shown in Figure 4.6, and $\vec{m} = [m_1, m_2]$ is the model vector and $i \in \mathbb{N} \cap [1, 41]$. Let us fix the model parameters $m_1 = 1.0$ and $m_2 = 1.0$. An array of synthetic data d^{noisy}, is then produced, introducing a gaussian-distributed noise, with $\sigma_i = 1 \forall i$. The synthetic dataset is presented in Figure 4.6. The effect of gaussian noise is reflected in the objective/misfit function values. Let us define the misfit function as:

$$\phi(\mathbf{m}) = \sqrt{\frac{1}{41} \sum_{i=1}^{41} (d_i^{noisy} - d_i)^2}. \tag{4.57}$$

The misfit surfaces are presented in Figure 4.7 and their difference $\phi(d^{true}, \mathbf{m}) - \phi(d^{noisy}, \mathbf{m})$ in Figure 4.8. In Figure 4.9 the final population of the inverse problem proposed, obtained with GA, is presented. It is evident the second limit of this approach: whereas in the final population the correct solution is present, most of the population's models are clustered around a sub-optimum point in the model space. Due that the model that build the population are interdependent, it is not even possible ti make a statistic relative to the final population, its mean and variance.

Summarizing the results of the two test presented, the GA performances are good and its ability in avoiding local minima is powerful, meanwhile there are two major setbacks:

- the requirement of a higher number of forward problem evaluations makes the GA highly computational-expensive;
- the "implicit" knowledge of the misfit surface allows to solve an optimization problem, but in the same time it is never endured the convergence to a minimum.

4.6.3. Application of genetic algorithm to magnetotelluric

As stated before, the genetic algorithm have been applied several times as the optimization engine in magnetotelluric inverse problems. The major disadvantage in using GA is the computation time required to solve numerous forward problems. Usually this problem is

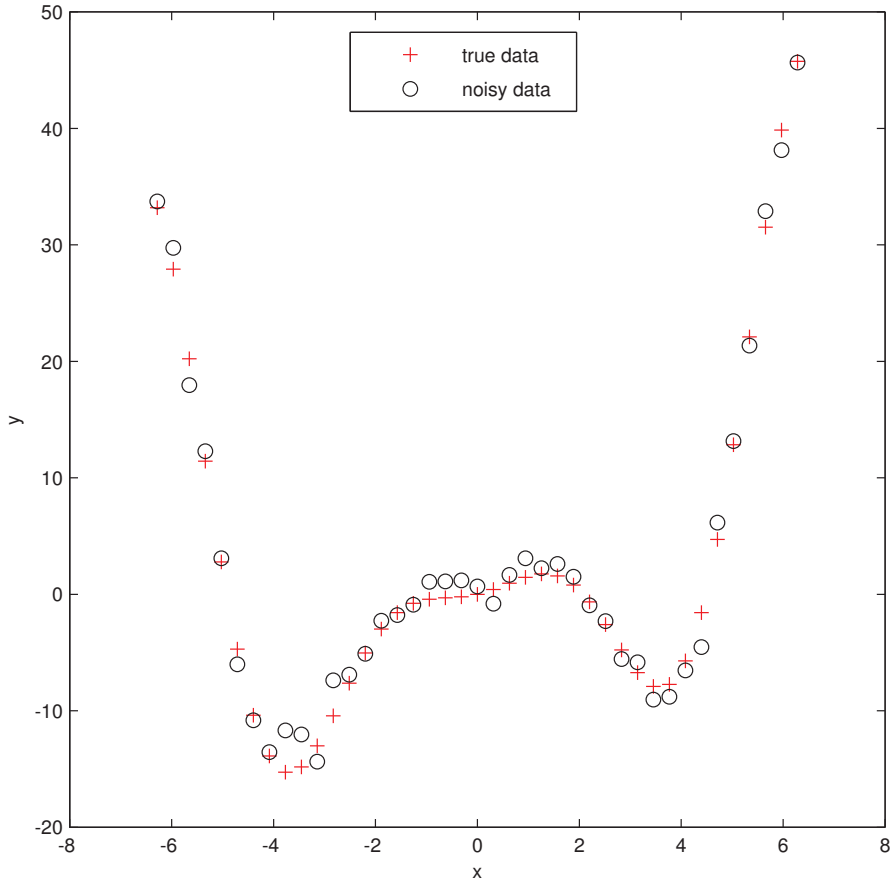


Fig. 4.6.: Synthetic data used in Test 2. Data from the true model are marked with red crosses, noisy data are marked with black circles.

outdone by restricting the model space to a coarse discretization. Seen from this angle, 2D models presented – for example – in *Pérez-Flores and Schultz* [2002] are discretized in a fixed grid of 16 parameters[6] to test their recombinant genetic algorithm with synthetic data and 48 and 72 parameters[7] to use the GA with real data. In *Moorkamp et al.* [2007] a joint inversion of teleseismic and magnetotellurics data in an isotropic 1D domain was bounded using 11 layers[8] whereas in *Roux et al.* [2011] the joint inversion surface-waves phase velocity and MT data in an anisotropic environment – 1D – was undertaken discretizing the system in 29 parameters.

As examined in the section 2.4, to represent anisotropic conductivity, a tensor with six independent parameters is needed. For example, in order to allow anisotropy in the checkerboard synthetic test performed in *Pérez-Flores and Schultz* [2002] 96 parameters in spite of 16 have to be estimated. In international conferences [*Mandolesi et al.*,

[6] arranged in 4 rows and 4 columns
[7] respectively arranged in 6 rows and 8 columns and 6 raws and 12 columns
[8] 33 parameters to describes both the conductivity and the sound velocity models

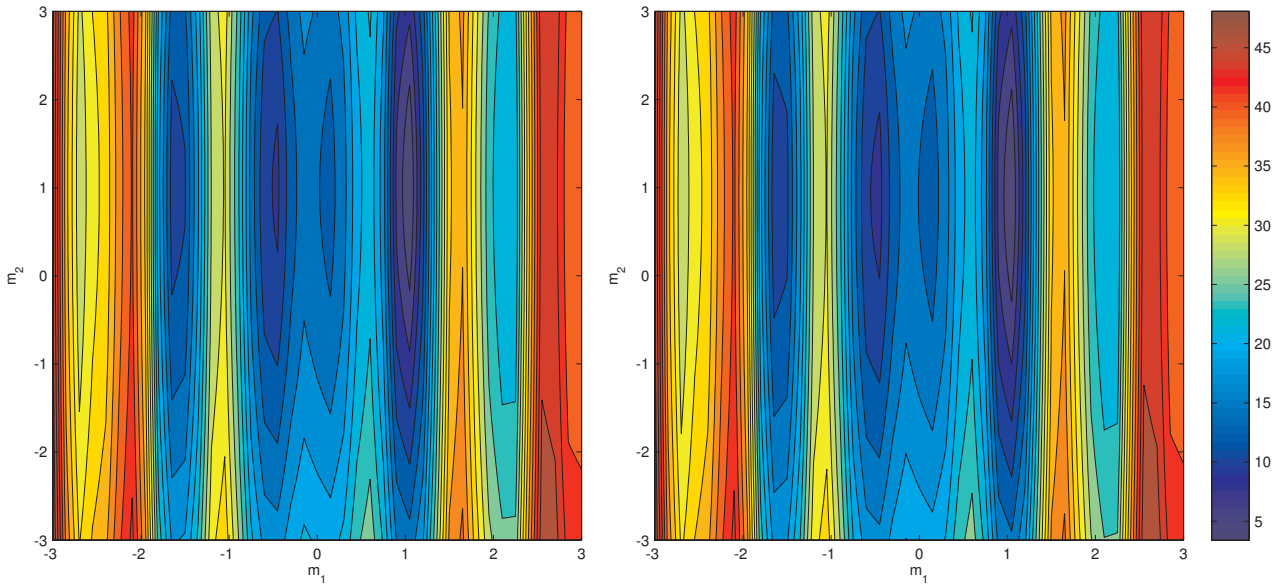


Fig. 4.7.: Misfit surfaces relative to the Test 2. On the left panel the misfit surface relative to the noise-free dataset; on the right panel the misfit surface relative to the noisy dataset. The presence of different minima is highlighted by the presence of multiple closed contour lines.

2009b,a] the low convergence rate provided by GA in MT have been highlighted. In Figure 4.10 RMS evolution from a run of GA in a 1D isotropic environment, discretized in 99 parameters, is reported. The expected value for the root mean square (RMS) is 1, and the reached value for this run was 10^4 after 12000 iterations considering a population of 1024 individuals. This run required the computation of 12288000 forward evaluations. In a complex domain, as the 2D anisotropic earth, this number of forward evaluations creates two major problems, as highlighted before:

- the computation time required by a forward evaluation is longer then the one required by the simple 1D problem, for example the computation of the forward solution of the test problem solved at page 89 is on average of 0.5 s per period on a computing node of the Stokes supercomputer[9];
- convergence properties for the GA have not been proved, thus it is not ensured that such number of forward evaluations is sufficient to converge into a misfit minimum.

These two reasons drove us to select a Newton-type algorithm over a stochastic method for minimizing the objective function.

[9] for information about the Ireland's High-Performance Computing Center (ICHEC)'s hardware visit the webpage www.ichec.ie

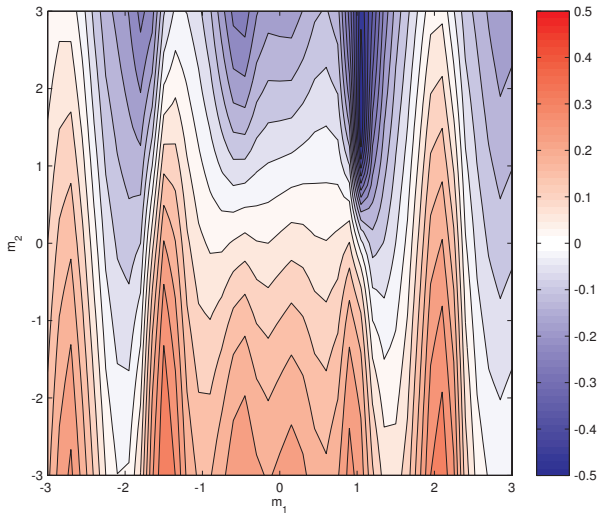


Fig. 4.8.: Misfit surface difference. The position of the global minimum is clearly biased by the presence of noise.

4.7. Tikhonov regularization

As stated before, solution of an inverse problem is often problematic because of the non-uniqueness of the solution itself. Bearing in mind that data contain noise, there is no advantage in fitting this noise exactly, thus many solutions can be regarded as adequate as soon as $\|G(\mathbf{m}) - \mathbf{d}\|$ is small enough.

The *Tikhonov regularization* (cf. *Tikhonov and Goncharsky* [1987] and references therein) is perhaps the most common technique used to reduce non-uniqueness in inverse problems [*Aster et al.*, 2004]. This method is based on the intuitive concept that the "best" model fits the data and has its norm $\|\mathbf{m}\|$ minimized. The idea is that a model built respecting these requirements will be zero everywhere except in the regions in which complexity has to be introduced in order to fit the data. There are several ways to express this concept in mathematical language, the most useful – solving a least square inverse problem – is

$$\min \|G(\mathbf{m}) - \mathbf{d}\|^2 + \lambda^2 \|\mathbf{m}\|^2 \tag{4.58}$$

where λ is a parameter that balances the trade-off between regularization and data fitting.

Equation 4.58 presents the so-called zero-order Tikhonov regularization. Of course it is possible to regularize the solution by defining different regularizers. Other functionals commonly used in as addition to the objective function are the first or the second derivative of the model, leading to the so-called first- and second-order Tikhonov regularization, which produce the flattest and the smoothest model, respectively. In the general case, a regularizer suitable for the LM method can be written as

$$\|\mathbf{L}(\mathbf{m})\|^2 \tag{4.59}$$

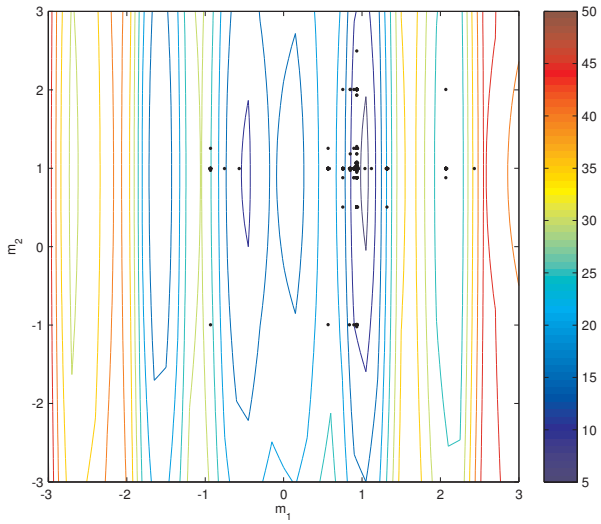


Fig. 4.9.: Final population relative to the inverse problem proposed in the Test 2 solved by GA. The clustering of the solutions around a sub-optimal point is remarkable.

thus, the objective function becomes

$$\phi(\mathbf{m}) = \|G(\mathbf{m}) - \mathbf{d}\|^2 + \lambda^2 \|\mathbf{L}(\mathbf{m})\|^2. \tag{4.60}$$

If we consider Tikhonov regularization, the regularization term can be written as $\|\mathbf{L} \cdot \mathbf{m}\|^2$, and $\mathbf{L}$ is given the name *roughening matrix*. In the 1D problem $\mathbf{L}$ will be written as the matrix that makes the product $\mathbf{L} \cdot \mathbf{m}$ proportional to the first or the second derivative of $\mathbf{m}$. Specifically,

$$\mathbf{L} = \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \dots & & \\ & & -1 & 1 & \\ & & & -1 & 1 \end{bmatrix} \tag{4.61}$$

to implement the first-order Tikhonov regularization and

$$\mathbf{L} = \begin{bmatrix} 1 & -2 & 1 & & \\ & 1 & -2 & 1 & \\ & & & \dots & \\ & & 1 & -2 & 1 \\ & & & 1 & -2 & 1 \end{bmatrix} \tag{4.62}$$

is the roughening matrix used to implement the second-order Tikhonov regularization. In an environment more complex, such as a 2D or 3D domain, the roughening matrix used is often a finite-difference approximation of the Laplacian operator.

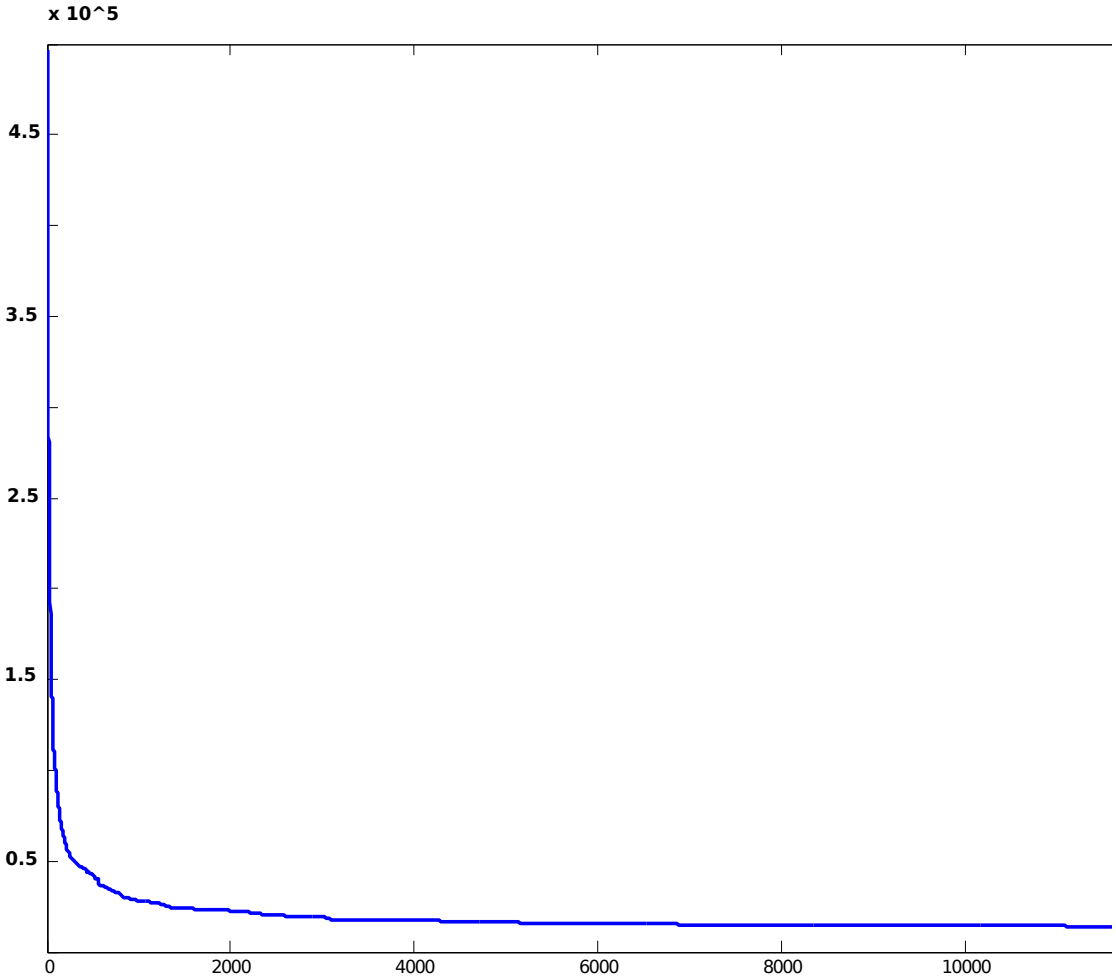


Fig. 4.10.: Convergence curve for a run of GA on a 1D isotropic problem, discretized in 99 layers. Figure form *Mandolesi et al.* [2009a].

In this work, when needed, the Tikhonov regularization have been used because of its easy implementation and in particular because of the ease in computing the derivatives of the term $\lambda^2||\mathbf{L}\cdot\mathbf{m}||^2$ needed in the LM algorithm.

5 Mutual Information Mathematical Setting

5.1. Information Theory Definitions

In 1948 Claude Elwood Shannon published two papers that are regarded as the seed of the so-called "information theory" [*Shannon*, 1948a,b]. Notwithstanding the importance of this science in its own right, in this chapter we will introduce some basic concepts from information theory and the features we used in the MT inverse problem. The goal of this chapter is to make clear the reasons why MI is a good choice if used as a constraint in a geophysics tomographic problem. In the following the symbol ":=" will denote a definition.

The *self-information* $I(A)$ that an event A may happen with probability p_A is defined as

$$I(A) := A \rightarrow \log \frac{1}{p_A} = -\log p_A. \tag{5.1}$$

$I(A)$ is measured in *bin of information* if the base of logarithmic function is 2. Other units of measurement can be introduced by changing the base of the logarithmic function, but they will be not considered here.

As considered by information theory, the *entropy* of a random variable X, $H(X)$, is the average of the self-information $I(x_i)$ of its possible values $(x_1, x_2, ..., x_n)$:

$$H(X) := E[I(x_i)] = \sum_{i=1}^{n} p(x_i) \log_2 \frac{1}{p(x_i)}. \tag{5.2}$$

The relation $H(X) \geq 0 \; \forall \; X$ derives directly by the Definition 5.2.

The relationship between entropy and information can be intuitively understood by considering that the more uncertain a random variable is, the more information about it is learned from its measurement.

Dealing with two random variables X and Y, there are other two interesting quantities

to be considered. The *conditional entropy* $H(X|Y)$ is the entropy of the variable Y conditioned on X taking a certain value x, and is the result of the average $H(Y|X = x)$ over each value X can assume.

$$\begin{aligned} H(X|Y) &:= \sum_{i=1}^{n} p(x_i) H(Y|X = x_i) && (5.3)\\ &= \sum_{i=1}^{n} p(x_i) \sum_{j=1}^{m} p(y_j|x_i) \log_2 \frac{1}{p(y_j|x_i)}, && (5.4)\\ &= -\sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i, y_j) \log_2 p(y_j|x_i), && (5.5)\\ &= \sum_{i,j} p(x_i, y_j) \log_2 \frac{p(x_i)}{p(x_i, y_j)}, && (5.6) \end{aligned}$$

where $p(x, y)$ indicates the joint probability distribution and $p(x|y)$ the conditional probability distribution of X and Y. The joint entropy is defined as

$$\begin{gathered} H(X, Y) = \sum_{x} \sum_{y} P(x, y) \log_2 P(x, y), \qquad (5.7)\\ \lim_{P(x,y)\to 0} P(x, y) \log_2 P(x, y) = 0 \end{gathered}$$

being x and y values of X and Y respectively, and $P(x, y)$ the probability that jointly $X = x$ and $Y = y$.

It is possible to define the mutual information $I(X; Y)$ as

$$I(X; Y) := \sum_{y\in Y} \sum_{x\in X} p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)}, \qquad (5.8)$$

where $p(x, y)$ is the joint probability distribution function of X and Y, and $p(x)$ and $p(y)$ are the marginal probability distribution functions of X and Y respectively. MI can be expressed equivalently as

$$\begin{aligned} I(X; Y) &= H(X) - H(X|Y), && (5.9)\\ &= H(Y) - H(Y|X), && (5.10)\\ &= H(X) + H(Y) - H(X, Y), && (5.11)\\ &= H(X, Y) - H(X|Y) - H(Y|X). && (5.12) \end{aligned}$$

Intuitively, MI measures the information that X and Y share. It measures how much the knowledge of one variable gives information about the other. For example, if X and Y are independent ,then $p(x, y) = p(x)p(y)$ and MI = 0, while if X=Y then $I(X; X) = H(X)$ and the information about X given from the knowledge of X itself is exactly its entropy.

5.2. MI in Tomography Context

Our interest in information theory is led by the opportunity that this theory offers in the definition of a metric between images. While in several geophysics inversion (and joint inversion) schemes the concept of similarity is defined by the shape of a model [*Haber and Oldenburg*, 1997; *Gallardo and Meju*, 2003], the definition of a metric in image space allows an accurate comparison of the model considered in the inversion process. For example the model considered step-by-step in equation 4.31 with a reference model.

The metric distance between X and Y defined by MI, called *variation of information*, is defined as:

$$\begin{aligned} \mathcal{D}(X,Y) &:= H(X,Y) - I(X;Y), && (5.13)\\ &= H(X) + H(Y) - 2I(X;Y). && (5.14) \end{aligned}$$

Variation of information respects all the properties a distance metric is supposed to respect. In particular

- $\mathcal{D}(X,Y) \leq \mathcal{D}(X,Z) + \mathcal{D}(Z,Y)$;
- $\mathcal{D}(X,Y) \geq |\mathcal{D}(X,Z) - \mathcal{D}(Z,Y)|$;
- $\mathcal{D}(X,Y) \geq 0 \; \forall X, Y$;
- $\mathcal{D}(X,Y) = \mathcal{D}(Y,X)$.

Bearing these relations in mind, it is possible to define a geophysical model as a picture. In fact if the model is discretized in cells, each cell can be seen as a pixel of an image (or voxel for an n-dimensional image).

From the definitions given before, the requirement to minimize the distance $\mathcal{D}(\mathbf{m},\mathbf{R})$ is achieved when equation 5.15 is maximized

$$I(\mathbf{m};\mathbf{R}) = H\left[\rho\left(\mathbf{m}\right)\right] + H\left[\rho\left(\mathbf{R}\right)\right] - H\left[\rho\left(\mathbf{m},\mathbf{R}\right)\right]. \tag{5.15}$$

The entropies computed on a continuous random variable are defined as an integral extension of the sum 5.2

$$H(\rho) := -\int \rho(t)\log_2\rho(t)dt, \tag{5.16}$$

where the probability density function $\rho(t)$ takes the place of the discrete probability distribution $P(X)$.

The great advantage in using a method developed to measure a distance between images is that it does not depend on the relations between the two images. If two images are similar, then they are close, and this fact does not depend on what the images represent or, in other words, which is the nature of the parameters that determine the model.

5.3. Estimation of Joint Probability Distribution, Mutual Information and their Derivatives

It is possible to estimate – and this estimation will be used in this thesis – the joint probability distribution (JPD) of two random variables $\rho(\mathbf{m}, \mathbf{R})$ using $\mathbf{m}$ and $\mathbf{R}$ values. Typically the estimated JPD depends on the values assumed by the components of the arrays $\mathbf{m}$ and $\mathbf{R}$, other than on the other parameters used in the estimation method. The estimated JPD is denoted in this text by the symbol ρ_Δ.

Following *Silverman* [1986] it is easy to estimate the density functions ρ_Δ. Between several methods described in *Silverman* [1986] the kernel estimator is selected, as this method is more reliable and having the major advantage that the density functions estimated are differentiable.

If a random variable X has density distribution f, it follows that

$$f(x) = \lim_{h \to 0} \frac{1}{2h} P(x - h < X < x + h). \tag{5.17}$$

Given h, it is trivial to estimate $f_h(x)$ with the proportion of the sample falling in the interval $(x - h, x + h)$, in formula

$$f(x) \approx f_h(x) = \frac{1}{2h} \frac{n_h}{n}, \tag{5.18}$$

where n is the sample size and n_h the number of extractions that fall in $(x - h, x + h)$. This concept can be expressed better by introducing a weight function $w(x)$

$$f_h(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} w\left(\frac{x - X_i}{h}\right), \tag{5.19}$$

with

$$w(x) = \begin{cases} \frac{1}{2} & \text{if } |x| < 1 \\ 0 & \text{otherwise.} \end{cases} \tag{5.20}$$

This "naive" estimator is easy to generalize, in fact, substituting the weight function $w(x)$ with a kernel function $K(x)$ so that

$$f_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \tag{5.21}$$

in which $\int_{-\infty}^{+\infty} K(x)dx = 1$. In this thesis, we chose to follow *Haber* [2004] and evaluate the joint density distribution ρ_Δ using the kernel function

$$K(\Delta, t) = \begin{cases} \frac{4}{3\Delta}\cos^4\left(\frac{\pi t}{2\Delta}\right), & \text{if } -\Delta < t < \Delta \\ 0, & \text{otherwise.} \end{cases} \tag{5.22}$$

In this function Δ takes the same role h has in equation 5.19 and controls the width of the kernel function.

This kernel function was chosen because of its properties, in particular:

- $\int_{-\infty}^{+\infty} K(\Delta, t)dt = 1 \ \forall\ \Delta > 0$;
- $\lim_{t\to\pm\Delta^{\mp}} K(\Delta, t) = \lim_{t\to\mp\Delta^{\pm}} K(\Delta, t) = 0$;
- $\lim_{t\to\pm\Delta^{\mp}} K'(\Delta, t) = \lim_{t\to\mp\Delta^{\pm}} K'(\Delta, t) = 0$;
- $K(\Delta, t) = K(\Delta, -t)$;
- $K(\Delta, t) \geq 0 \ \forall\ t$;
- $\lim_{\Delta\to 0} K(\Delta, t) \approx \delta(t)$;

where $\delta(t)$ is the Dirac delta function. Looking for a joint probability distribution that generates the two images $\mathbf{m}$ and $\mathbf{R}$ the joint probability density estimator is a linear combination of kernel functions, explicitly:

$$\rho_\Delta(\mathbf{m}, \mathbf{R}; g_m, g_R) = \sum_{j\in J} K(\Delta, \mathbf{R}(x_j) - g_R)K(\Delta, \mathbf{m}(x_j) - g_m) \tag{5.23}$$

in which $g_\star$ is the so-called gray value relative to the image $\star$, or, in density estimation terminology, represents the axis on which the kernel estimator approximates the histogram produced via Equation 5.19 and $(x_j : j \in J)$ is a discretization of the image domain [*Haber and Gazit*, 2012]. The derivative of entropy, needed by the LM algorithm described in 4.3.2, is then easily computed, in fact, on a picture composed of n pixels/voxels

$$\rho_\Delta(\mathbf{m}, \mathbf{R}; g_m, g_R) = \sum_{j=1}^{n} K(\Delta, \mathbf{R}_j - g_R)K(\Delta, \mathbf{m}_j - g_m), \tag{5.24}$$

$$\frac{\partial\rho_\Delta}{\partial\mathbf{m}_i}(\mathbf{m}, \mathbf{R}; g_m, g_R) = K(\Delta, \mathbf{R}_i - g_R)K'(\Delta, \mathbf{m}_i - g_m), \tag{5.25}$$

with

$$K'(\Delta, t) = \begin{cases} \frac{8\pi}{3\Delta^2}\cos^3\left(\frac{t\pi}{2\Delta}\right)\sin\left(\frac{t\pi}{2\Delta}\right) & \text{if } -\Delta < t < \Delta, \\ 0, & \text{otherwise.} \end{cases} \tag{5.26}$$

This method is referred to as the histogram method. In order to clarify the concept of distance measure between images, the following examples are presented.

5.3.1. MI Distance Measure Examples

Let us consider the images in figures 5.1 and 5.2, respectively figure **R** and **m**.

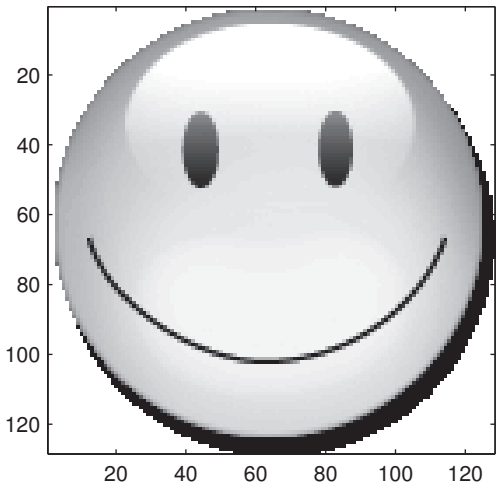


Fig. 5.1.: A black and white picture consist of a map of pixels in which in each cell a gray value g_R in the interval $[0, 1]$ is given, with 0 = black and 1 = white

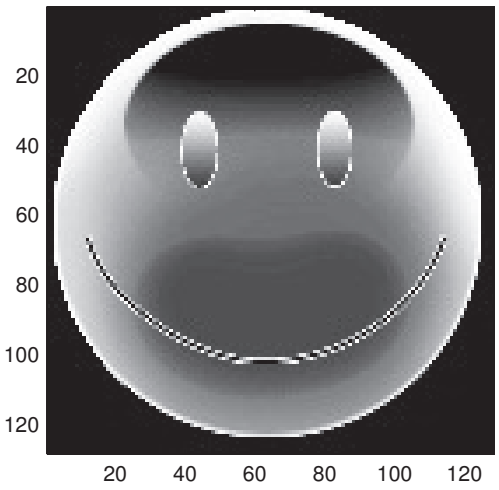


Fig. 5.2.: This image is obtained by the image displayed in figure 5.1 applying to the value of each cell the non-linear transformation $g_m = \sin(g_R \cdot \pi)$

The transformation that maps 5.2 in 5.1 must be regular enough to maintain the shape, but is non linear. It is evident that each white pixel of figure 5.1 becomes a black pixel in figure 5.2, while the inverse relation does not hold. For example, the black "half moon" shape in the lower right corner of figure 5.1 is black even in figure 5.2.

The estimated joint density function is shown in figure 5.3. The sinusoidal relation that links the grey values of the two pictures is recovered, and its integral value over the whole definition space equals 1, and $I(\mathbf{R}; \mathbf{m}) = 0.94$.

Let us consider another two pictures, presented in figures 5.4, and 5.5, respectively $\mathbf{m_1}$ and $\mathbf{m_2}$. $\mathbf{m_1}$ is the "negative" of **R** whereas $\mathbf{m_2}$ has been derived from **R** using a complex non-linear transformation. Intuition suggests that $\mathbf{m_1}$ "informs" us about **R** more than $\mathbf{m_2}$. The estimated JPD are presented in figures 5.6 and 5.7. Compared to the reference image **R**, the closest image, $\mathbf{m_1}$, presents a linear JPD. On the contrary, the quasi-random relation that links **R** and $\mathbf{m_2}$ is not retrieved. Notwithstanding the inability in finding the correct relation between **R** and $\mathbf{m_2}$, the MI value, reported in table 5.1, continues to suggest a certain relation between the two images.

As a final example, an image that has nothing to do with **R** is presented in figure 5.8: $\mathbf{m_a}$. In this case the estimated joint probability distribution $\rho_\Delta(\mathbf{R}, \mathbf{m_a})$ shows a sparse, chaotic pattern (figure 5.9). The similar shape of the apple and the smile is in itself not sufficient to communicate a significant amount of information about the smile. This evidence is translate in the MI value, $I(\mathbf{R}; \mathbf{m_a}) = 0.48$.

The error in comparing images **R** and $\mathbf{m_a}$ is in the assumption that both come from the same JPD. It is then nonsense to look for a joint probability distribution that cannot

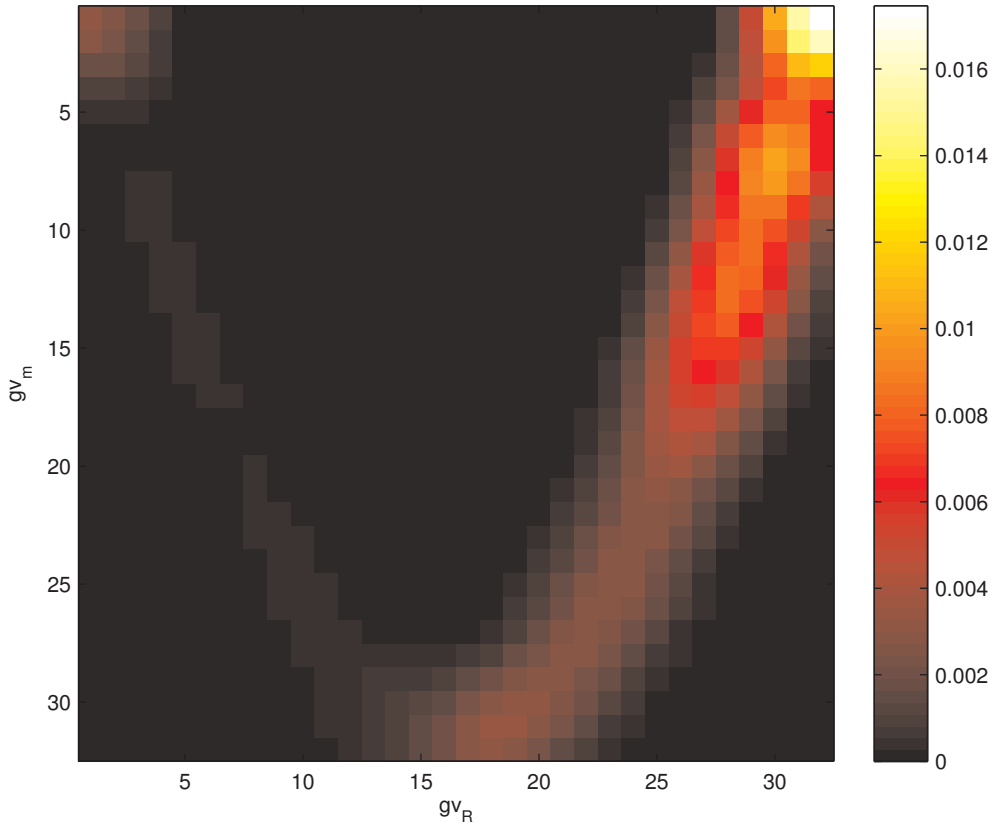


Fig. 5.3.: Joint density function $\rho_\Delta(\mathbf{m}, \mathbf{R}, g_m, g_R)$. This function has been estimated fixing $\Delta = 0.2$. The values on the axis correspond to the discretization –in this case in 32 steps – of the values that compose figures 5.1and 5.2. Thus the probability that a value $R^\star : \frac{n}{32} \leq R^\star \leq \frac{n+1}{32}$ in figure **R** correspond to a certain value $m^\star : \frac{i}{32} \leq m^\star \leq \frac{i+1}{32}$ is given by the value of $\rho_\Delta(n, i)$

exist. What lasts is the distance definition, the proof that notwithstanding the similarity in shape, the distance between the two pictures is bigger than the ones relative to all the precedent couples considered.

Table 5.1 summarizes the computed MI values for the different pictures. In agreement with intuition, the more two images are similar, the greater their MI. This means that the measure of distance between similar images is small, according with the variation of information distance definition. This metric concept allows the exact comparison of "similarity" between images – and geophysical models if seen as images – and hopefully a correct estimation of their joint probability distribution.

Summarizing: in this scheme the subsurface itself is viewed as the source of signals regarding its structure. These signals can be detected with different methods, but the results are supposed to inform us about the structure of the source and thus about the model detected with all the different methods used to model the subsurface.

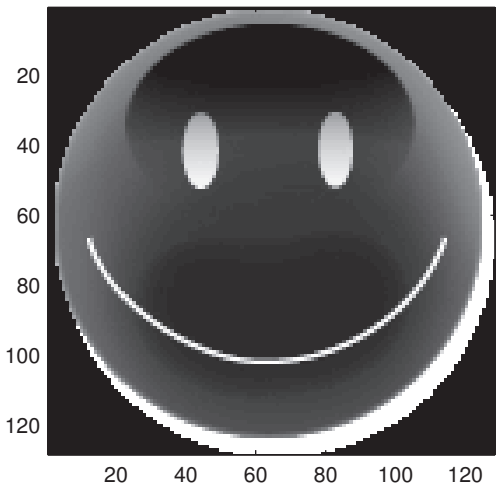


Fig. 5.4.: $\mathbf{m_1}$. Image obtained by applying to $\mathbf{R}$ the transformation $\mathbf{m_1}(i,j) = 1 - \mathbf{R}(i,j) \forall i,j$

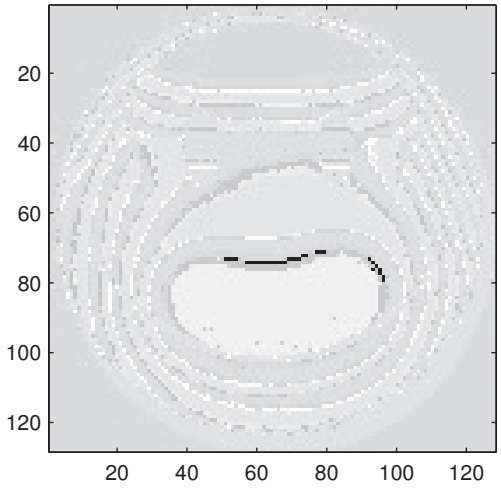


Fig. 5.5.: $\mathbf{m_2}$. Image obtained by applying to $\mathbf{R}$ the transformation $\mathbf{m_2}(i,j) = \frac{\sin[\mathbf{R}(i,j)]}{\sin[100\mathbf{R}(i,j)]} \forall i,j$

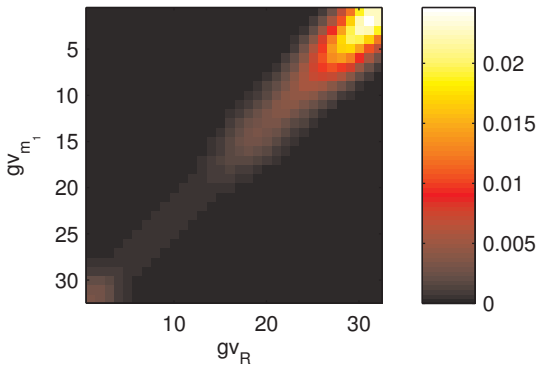


Fig. 5.6.: Joint density function $\rho_\Delta(\mathbf{m_1}, \mathbf{R}, g_{m_1}, g_R)$. This function has been estimated fixing $\Delta = 0.2$.

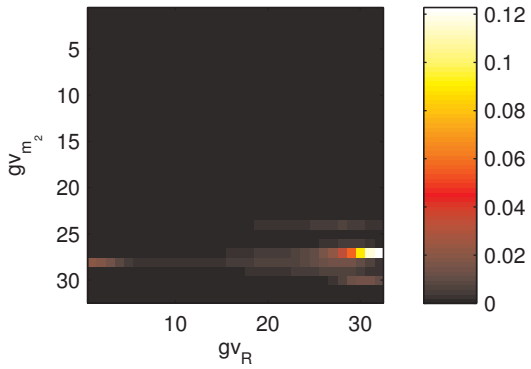


Fig. 5.7.: Joint density function $\rho_\Delta(\mathbf{m_2}, \mathbf{R}, g_{m_2}, g_R)$. This function has been estimated fixing $\Delta = 0.2$.

	MI(bits)
$I(\mathbf{R};\mathbf{R})$	1.0
$I(\mathbf{R};\mathbf{m})$	0.94
$I(\mathbf{R};\mathbf{m_1})$	0.97
$I(\mathbf{R};\mathbf{m_2})$	0.64
$I(\mathbf{R};\mathbf{m_a})$	0.48

Tab. 5.1.: MI values for the different examples considered. According with the definition 5.13, the more distant the pictures, the smaller their MI.

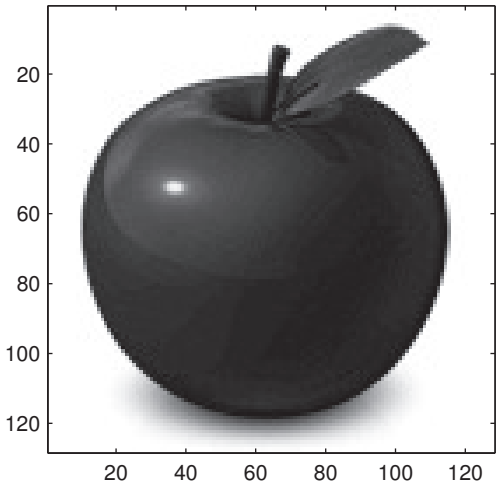


Fig. 5.8.: Joint density function $\rho_\Delta(\mathbf{m_1}, \mathbf{R}, g_{m_1}, g_R)$. This function has been estimated fixing $\Delta = 0.2$.

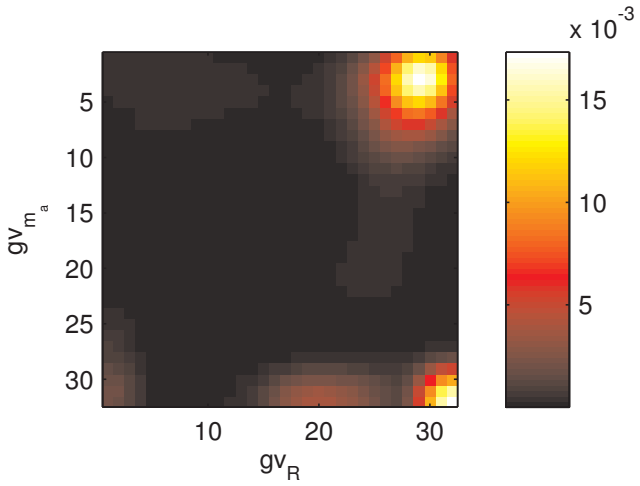


Fig. 5.9.: Joint density function $\rho_\Delta(\mathbf{m_a}, \mathbf{R}, g_{m_a}, g_R)$. This function has been estimated fixing $\Delta = 0.2$.

5.3.2. MI Robustness Examples

A discussion relative to the robustness of MI with respect to the errors in the pixels – or in the models parameters, in inverse problem terminology – is needed in order to appreciate the stability of the MI constraint with respect to errors in the reference model in an inverse problem. Robustness of maximization of MI as a distance measure is appreciated in *Unser and Thevenaz* [2000]. Without the pretension to prove this stability, in Figure 5.10 the variation of MI with noise in a picture's pixels – relative to two of the examples examined in the previous section, specifically $\mathbf{m}$ and $\mathbf{m_a}$ – is presented. It is evident that

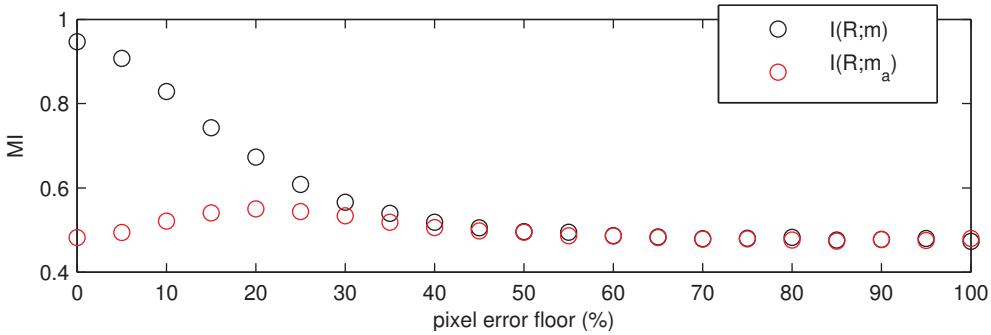


Fig. 5.10.: Variation of mutual information with noise in pixels. To each of the pixel building figures $\mathbf{m}$ and $\mathbf{m_a}$ a random number drawn from a gaussian distribution characterized by mean zero and variance equal to the error have been added.

random noise improves the similarity in two different pictures and erases the similarities in two close ones up to a certain threshold, after that threshold value, the noise completely destroys the information carried by the signal, making the two different pictures equidistant from the reference $\mathbf{R}$. In Figure 5.11 are presented some of the pictures used to compute data presented in Figure 5.10.

A better approach to consider numerically the stability of the mutual information with "errors" in the model parameter m_i is to consider the ratio $\frac{I^i(\mathbf{R};\mathbf{m})}{I(\mathbf{R};\mathbf{m})}$ with $I^i(\mathbf{R};\mathbf{m}) = \frac{\partial I(\mathbf{R};\mathbf{m})}{\partial m_i}$. This ratio represents the relative variation of the MI value with the *i-th* model parameter, and is a good candidate to represent the stability of the MI. The smaller the ratio, the more stable MI is with respect to errors in model parameters. In Figure 5.12 the variation relative to the model parameters is reported relative to the example model $\mathbf{m}$. For most of the model parameters – the white pixels in Figure 5.12 – the MI value results are extremely stable in the example examined.

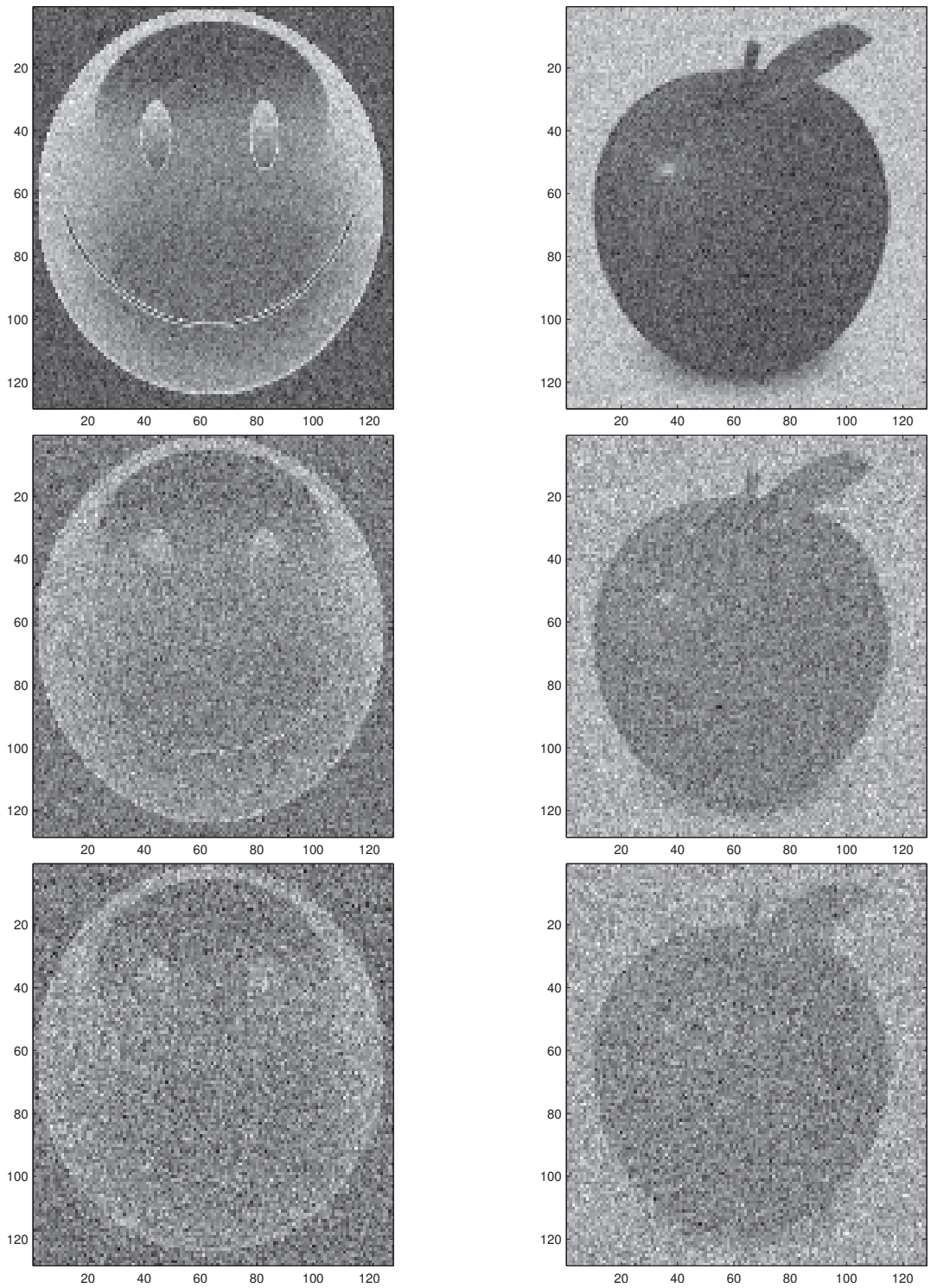


Fig. 5.11.: Some of the pictures used to compute mutual information values presented in Figure 5.10. Top to bottom: $\mathbf{m}$ (left side) and $\mathbf{m_a}$ (right side) subject to 15%, 40% and 65% random noise.

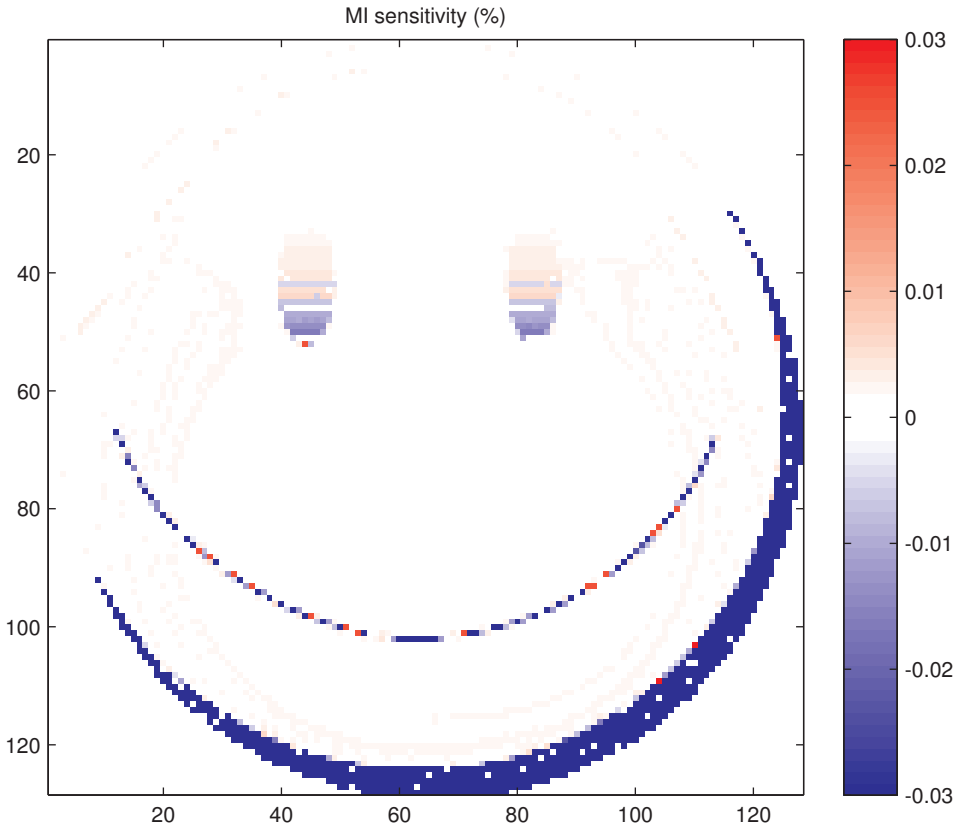


Fig. 5.12.: Relative sensitivity of MI with the variation of model parameters value, relative to the two images **m**, **R**.

Part II

A new approach to MT anisotropic inverse problem

Fundamental ideas play the most essential role in forming a physical theory. Books on physics are full of complicated mathematical formulae. But thought and ideas, not formulae, are the beginning of every physical theory. The ideas must later take the mathematical form of a quantitative theory, to make possible the comparison with experiment.

– Albert Einstein

6 A New Approach

As there is no general analytical solution available for the inversion of anisotropic MT data, over the years several computer codes have been developed to solve the problem numerically and to produce electrical conductivity maps of the subsurface from anisotropic MT datasets.

This chapter presents a new approach to working with anisotropic MT data along with results from simple synthetic case studies.

6.1. Figure Driven Inversion

Building upon the theoretical background previously discussed, it is possible to construct the new approach to the inversion of anisotropic MT data that is the subject of this thesis

In Chapter 2 we explored the theory that facilitates the so-called "forward modeling". The methods described in the Chapter 2 assume the role of the forward operator $G(\star)$, widely used in the inverse problems theory, as presented and discussed in Chapter 4.

The objective function to be minimized is defined as

$$\begin{aligned} \phi(\mathbf{m},\mathbf{R}) &:= \sum_{i=1}^{m}\left[\frac{G(\mathbf{m})_i - d_i}{\sigma_i}\right]^2 + \lambda_L^2\|\mathbf{L}(\mathbf{m})\|^2 + \\ &+ \lambda_a^2\mathbf{P}(\mathbf{m}) + \lambda_{MI}^2\left[I(\mathbf{R};\mathbf{R}) - I(\mathbf{m};\mathbf{R})\right]^2, \end{aligned} \tag{6.1}$$

where

m is the n-dimensional model vector in which the model parameters are stored;

R is the reference model. It may be a geological profile, the result of a seismic inverse problem or any kind of trusted subsurface map of the studied region;

$G(\star)_i$ is the i^{th} component of the forward operator applied to the $\star$ vector. $G(\star)$ operates on an n-dimensional model vector predicting m data;

d_i is the i^{th} component of the m-dimensional data vector;

σ_i is the i^{th} component of the m-dimensional error vector;

λ_L is a parameter that controls the trade-off between data-fitting and Tikonhov regularization;

$\mathbf{L}$ is the roughening matrix as described in the Section 4.7;

λ_a is a parameter that controls the trade-off between data-fitting and anisotropy;

$\mathbf{P}(\star)$ is an operator that penalizes high anisotropy. It takes different shapes in one and two dimensions, and it will be specified when needed;

λ_{MI} is a parameter that controls the trade-off between data-fitting and the MI influence;

$I(\mathbf{m};\mathbf{R})$ is the mutual information between $\mathbf{m}$ and $\mathbf{R}$ as described in Chapter 5;

$\lambda_\star \geq 0$.

With the minimization of ϕ, we aim to achieve the following different objectives simultaneously:

- to fit the MT data;
- to retrieve a regular[1] model of the subsurface electrical conductivity;
- to retrieve a model that is as close as possible to the reference model selected, in the MI sense.

The method chosen to minimize ϕ is the Levenberg-Marquardt algorithm, described in section 4.3.2. After the definition of an initial guess model $\mathbf{m^0}$, the equation 4.31 is iteratively solved until a certain condition ensures that a minimum – possibly a local one – has been reached.

In this context the reference model, $\mathbf{R}$, is interpreted as an image coming from the same joint probability distribution that generates $\mathbf{m}$. The requirement that the two must share information about the source (in this instance, the subsurface) is thus explained. The power of this method is evident; a direct link between the two images of the subsurface is not explicitly requested. Moreover, it is possible to use, without modification, any kind of geophysical results available from the same area to drive the inverse problem.

A short discussion about the term $\mathbf{P}(\star)$ is necessary. Since in most cases we try to fit the data within the minimum possible anisotropy, we define

$$\mathbf{P}(\star) := \int_\star (\ln \varrho_1, \ln \varrho_2, \ln \varrho_3) \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix} \begin{pmatrix} \ln \varrho_1 \\ \ln \varrho_2 \\ \ln \varrho_3 \end{pmatrix}, \tag{6.2}$$

[1] regular meaning both in a Tikhonov sense and to present the minimum possible amount of anisotropy.

where ϱ_i is the *i-th* principal conductivity, and $\star$ represents the whole model domain. The Equation 6.2 represents an "elastic force[2]" that keeps the principal resistivity values joined together: as soon as the three principal resistivity values become similar, the conductivity tensor becomes singular and can be treated as a scalar. This is translated in the physical domain to an isotropic behavior (isotropic conductivity can be defined as this special case of degenerate anisotropy).

The condition chosen to stop the iterative process is $\phi(\mathbf{m^i}) - \phi(\mathbf{m^{i-1}}) < \varepsilon$, meaning we consider we have reached a minimum when the i^{th} iteration changes the value of ϕ less then a given threshold value ε. From a numerical perspective, ε represents the limit that allows us to discriminate a "flat" area in the objective function hyper-surface. When a flat area has been reached, the sensitivity of the data with respect to the model parameters is numerically negligible, thus there is no need to proceed with the next iteration.

6.2. Synthetic Tests

The purpose of this section is to test and analyze results returned from the new approach discussed above. In order to test the algorithm we compute the MT response for some hypothetical synthetic models. This response will be used as MT data in the inversion scheme. Notwithstanding the innovation introduced by the variation of information in the Levenberg-Marquardt algorithm, the main goal of an inverse problem remains the finding of a model that statistically fits the data. While the inherent non-uniqueness of the solution may motivate the rejection of a model that fits data, a model that does not fit the data is to be rejected for the simple reason that the predictions made on the basis of that model are wrong.

Thus, the results are supposed to:

- fit the data;
- reproduce the synthetic model;
- be robust to changes in the values of trade-off parameters.

Whilst the data fitting is self-explanatory and the request to retrieve the correct model is the primary goal of any inverse problem, the concept of robustness with respect to the changes in the values of trade-off parameters deserves a short discussion.

As the solution of the MT inverse problem is non-unique, features in a model that appear to be minimizing a particular objective function may be the result of the path followed by the algorithm to reach the specific minimum retrieved. Adding a regularization term to the objective function (such as adding a term to keep the anisotropy rate low, or a term that maximizes the mutual information) can be interpreted in the multi-dimensional space,

[2] the penalty value relative to two different principal resistivity values is proportional to the square of their difference, as in the classic Hooke's law.

in which ϕ is defined, as an extension of the space in an extra dimension that allows the descent of the algorithm towards a local minimum to follow a different path. To illustrate this concept it is useful to look again at the expression 4.18. It is obvious that any term added to the objective function results in a new component – a new dimension in the geometrical interpretation – in the vector 4.18. Thus, any feature in a model can be an artifact due to the specific path followed because of the major influence of one of the added dimensions. This influence is controlled by the specific trade-off parameter used. Therefore, it is important when attempting to ascertain whether a certain physical feature is present in the subsurface that the results are robust to the variation of trade-off parameters.

6.2.1. Selection of inversion parameters

Equation 6.1, the estimation of mutual information and the LM algorithm present a number of parameters that need to be set to run the inversion code effectively. The trade-off parameters, the initial value for the LM damping parameter, the numerical precision and the maximum number of forward evaluations are all important factors that affect the results of the inverse problem. As described in Chapter 4, the Levenberg-Marquardt initial damping parameter was set to $\lambda = 128$. This value represented a good compromise between convergence speed and accuracy in all tests performed. The maximum number of forward evaluations, $\mathcal{N}_{max}$, is a numerical limit which prevents the algorithm from running without an effective change in the results. It only affects the inverse problem result if the local model corresponds to a very flat region of the objective function. $\mathcal{N}_{max}$ was set to $400 \cdot n$ for the 1D tests, whereas the maximum number of iteration $\mathcal{N}_{iter}$ was set to 40 for the 2D tests. The other parameter values were adjusted on a trial-and-error basis; this is described in detail in the test descriptions below.

6.2.2. 1D Preliminary Test

Beginning with noise-free data and understanding the data source, we wish to test whether the newly developed algorithm can retrieve the input model and find the original JPD. The influence of MI in the inverse scheme can thus be evaluated. The starting guess $\mathbf{m^0}$ was set as an isotropic half-space with a conductivity value of $0.03\ S \cdot m^{-1}$. The model was discretized into n_l layers with each layer completely described by its electrical conductivity tensor – as per the theory presented in the Section 2.4 – and by its fixed thickness. The half-space was discretized using layer boundaries that were logarithmically spaced to accommodate the natural exponential decay of the impedance with period. This discretization corresponds to the grid design suggestions presented in *Simpson and Bahr* [2005], and allows numbers of a similar magnitude to be used in the sensitivity matrix. The first synthetic model tested is shown in figure 6.1. The anisotropic azimuthal strike angle, β, can be seen in the left panel and the two aggregated conductivity values, A_1 and

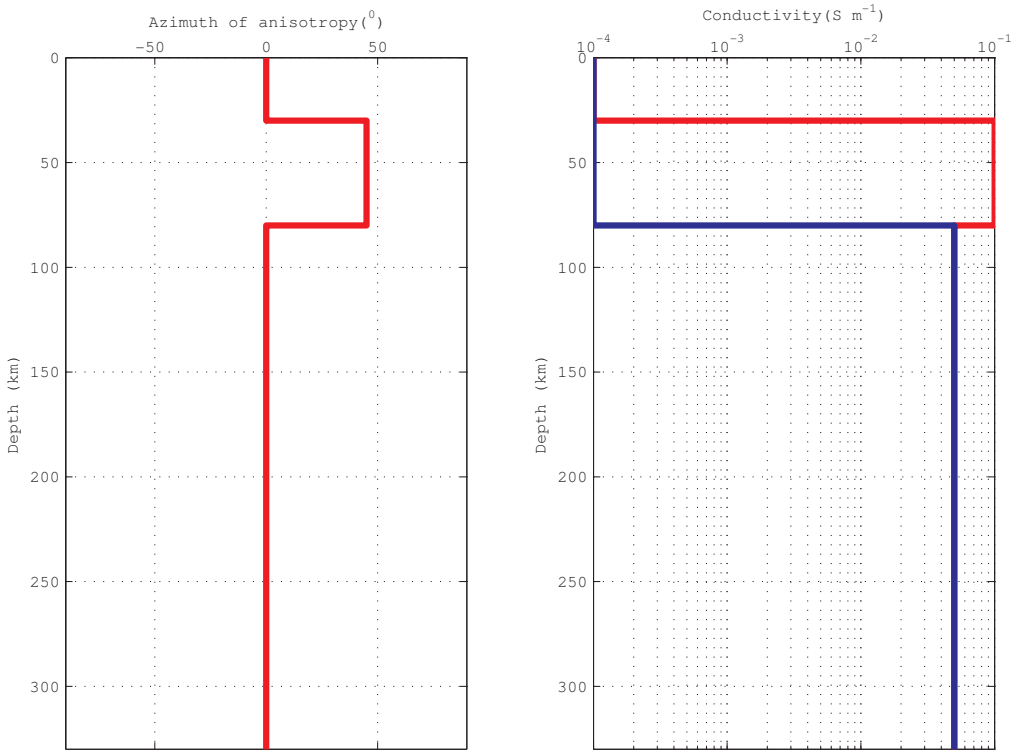


Fig. 6.1.: Synthetic model used to test the 1D algorithm. Because of the non unicity in the representation, there is a 90° ambiguity in the anisotropic strike direction, thus whenever the conductivity value in the conductive direction is smaller then the one in the resistive direction, the twos can exchange their roles with a simple 90° rotation (in the specific layer). **Left panel:** azimuthal anisotropy strike direction. **Right panel:** aggregated conductivity values A_1 (in *red*) and A_2 (in *blue*)

A_2, can be seen in the right panel. These parameters are all defined in equation 2.74. In the test we assumed an error floor of 5%.

Avoiding the black-box syndrome

It is, of course, possible to set up the inverse problem using a public computer library that implements the LM algorithm, setting a random number of layers and checking the results produced. We tested the MINPACK library, broadly used by the scientific community for decades. The MINPACK implementation of the LM algorithm was fast, but was unable to produce acceptable results even for the simple model used in this test. We have identified the problem. In the line-search procedure to find lambda, MINPACK searches for the damping parameter that minimizes the objective function, keeping the LM algorithm as close as possible to the Gauss-Newton method to increase the rate of convergence. The problem, in this case, was relative to the line-search procedure that selects a step direction that leads the model obtained by the first iteration too far from the guess model, $\mathbf{m^0}$, and thus ends up in a completely different region of the model domain where the algorithm se-

lects a different, unsuitable, local minimum of the objective function, which corresponds to an unsuitable retrieved model.

Following this, we decided to write our own implementation of the Levenberg-Marquardt method to be able to handle all the parameters we needed to control, considering that the problem to be solved is relative to a specific forward-solution and minimizes an objective function specifically tuned to the physics of this problem.

1D Preliminary Test solved

With the synthetic dataset ready, the framework used to invert the data was defined. Since a comparison of the model parameter values was required (the grey tones in the images) the number of layers[3] required to represent the model had to be fixed. It is important to note that with few parameters the link with a reference figure results unstable, as in that case there are very few voxel available to provide a "good" estimation of the JPD. Too many parameters, directly translated in layers in this scenario will over fit the data. In order to select the number of layers to adequately solve the problem, the standard statistical approach and performed an F-test to examine results obtained from an isotropic inversion have been used. The these was performed in order to checked that, if a model characterized by a certain number of layers n_l fit the data, a second model constituted by $2 \cdot n_l$ layers fits the data *significantly* better. The null hypothesis tested was that the model characterized by $2 \cdot n_l$ layers does not provide a significantly better fit than the model characterized by n_l layers (layers are logarithmically spaced as described above). The χ^2 was used as estimator, and a significance of 95% was fixed. The results of the test indicate that an isotropic inversion in a domain discretized in 32 layers improves the data fit by less than 5% with respect to an isotropic inversion in a domain discretized in 16 layers. The number of layers – logarithmically spaced – was thus fixed to 16[4]. The model domain was set so that the most penetrating wavelength's magnitude becomes negligible inside the model (cf. Chapter 2).

Once the guess model had been selected, a trial-and-error procedure was used to determine the optimal trade-off parameters for this inversion. Figure 6.2 shows the L-curve that was constructed to select the lambda parameter; a zero-order Tikhonov regularization was used for this. This L-curve method is a commonly-used, empirical method for defining a trade-off parameter [*Hansen and O'Leary*, 1993]; the "best" parameter is considered to be the one situated on the edge of the L-shaped curve of RMS plotted against λ_L.

With a final RMS $\approx$ 5.3, it is evident that the data are not fitted, and we have to reject the model relative to the local minimum retreived.

In order to introduce the MI penalty function and measure the variation of the information it is necessary to define a reference model, **R**. In this test, two different reference

[3] or equivalently the number of 1D voxels, in the MI vocabulary

[4] in this scenario that means 48 estimated parameters, 16 high conductivities, 16 low conductivities and 16 anisotropic strike directions

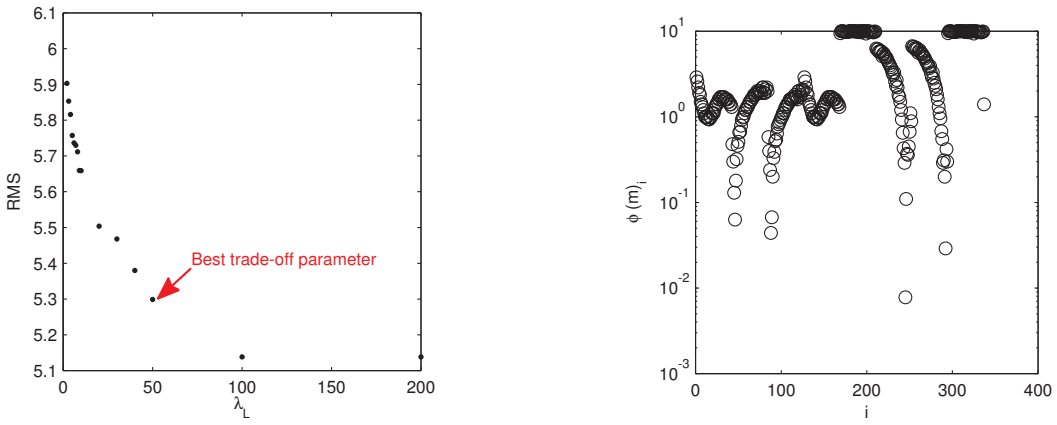


Fig. 6.2.: Left panel: L-curve computed to estimate the trade-off parameter λ_L. **Right panel**: Component of the objective function $\phi(\mathbf{m}^*)$, the trade-off parameters were$\lambda_L = 50$, $\lambda_{MI} = 0$, $\lambda_a = 0$.

models are adopted; one that is "compatible" with the synthetic model (Figure 6.3) and a second one that that is "incompatible" with the synthetic model (Figure 6.4). "Compatible" refers to a model that has a relationship with the synthetic model, and "incompatible" refers to a model that does not share information with the synthetic model. As discussed in Chapter 5, the numerical values that define the voxels do not influence the MI distance definition.

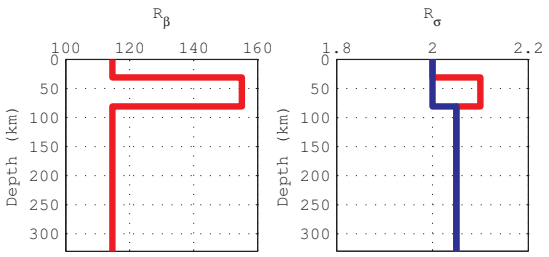


Fig. 6.3.: Compatible **R** model.

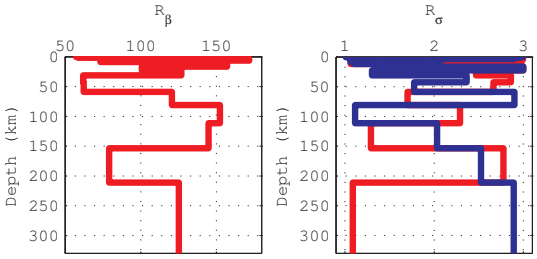


Fig. 6.4.: Incompatible **R** model.

The reference model, **R**, can be defined in a synthetic test so it is obvious that the relationship with the synthetic model, **m**, will be imposed or negated. Great care must be taken when selecting a convenient reference model. This effort will influence the selection of a reference model relative to a parameter that shares information with electrical conductivity. The retrieved model in Figure 6.5 was produced by using the compatible reference model in the computation of the mutual information penalty function. Figure 6.6 shows the data fit relative to the inversion performed in Figure 6.5. The algorithm was able to fit all the data within one error bar with none of the residuals exceeding a value of 1. Figure 6.7 shows the estimated JPD relative to the resulting model and the compatible reference model. The result from the inversion relative to the incompatible reference model produces an RMS equal to 5.9. Figure 6.8 shows the estimated joint probability

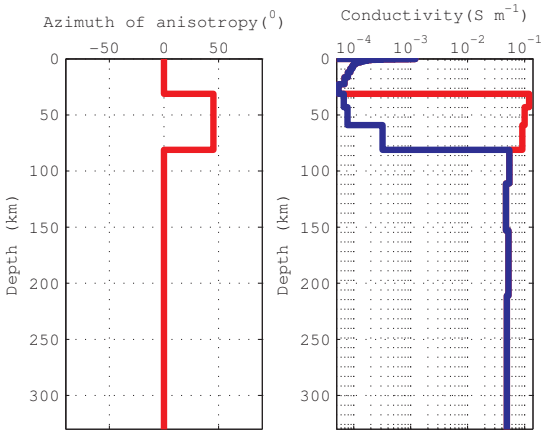


Fig. 6.5.: Solution of the inverse problem presented. To obtain this solution the compatible reference model have been used in the computation of the MI penalty function.

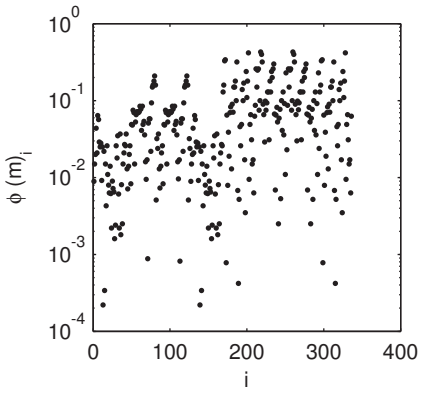


Fig. 6.6.: Components of the objective function computed in correspondence with the model presented in Figure 6.5. We highlight that all data have been fit within one error bar.

distribution relative to this inversion. It is evidently not possible from this distribution to distinguish a relationship between the two models. It is also evident that the incompatible model is not advantageous in the inversion process, but in this case it does not lead to a different minimum value to the one retrieved without the use of MI.

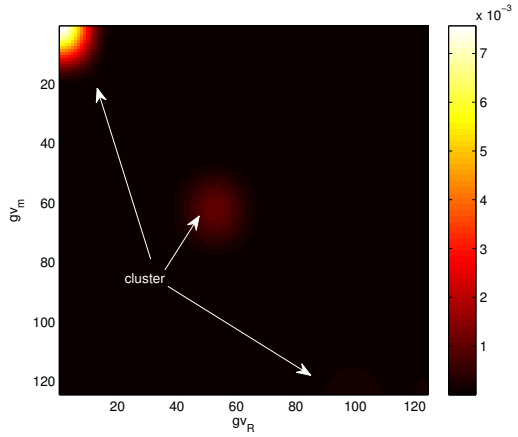


Fig. 6.7.: JPD, relative to the compatible **R**.

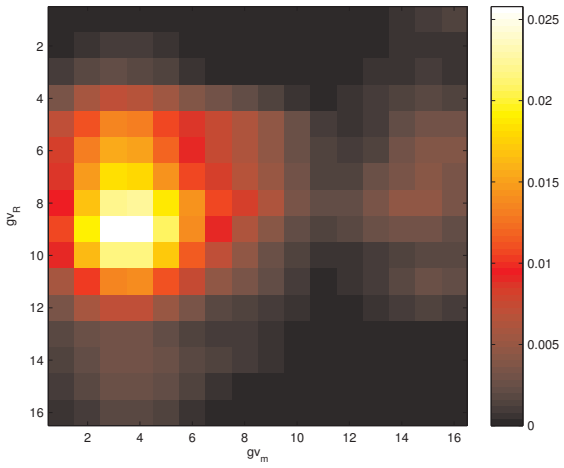


Fig. 6.8.: JPD distribution estimated in the inversion with the incompatible reference model. It is evident that there is no relation between the gray values of **R** and the ones that compose **m**.

To explore the influence of the MI penalty function, several inversions were tried. Figure 6.9 shows a plot of the variation of ϕ with respect to λ_{MI}. It is evident that the mutual information term becomes dominant with respect to other terms in the objective function after a certain threshold; thus the absolute value of ϕ increases, as expected, with λ_{MI}.

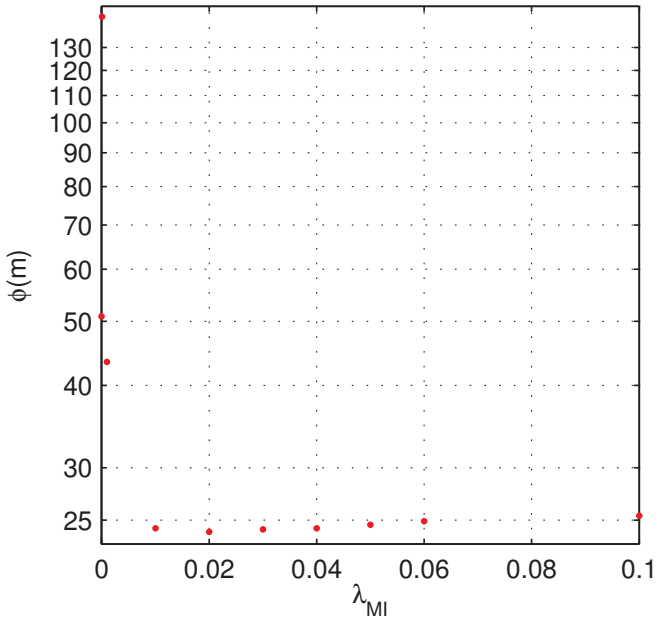


Fig. 6.9.: Variation of objective function with λ_{MI}.

6.2.3. 2D Preliminary Test

As seen by the inverse problem description the extension of the physical problem domain, from 1D to 2D, is merely translated into the introduction of a different forward solver in the Equation 6.1, thus in the accuracy of the approximation 4.29.

An example of a 2D anisotropic inversion is presented here. This was thought of as a "school test" to check the accuracy of the code and the simplicity of the design of the test has been exaggerated. The guess model, $\mathbf{m^0}$, was defined as a 1000 $\Omega \cdot m$ isotropic half-space. The conductivity tensor for the 2D model was parameterized in each homogeneous region of the structure by the following anisotropy parameters:

- the three principal electrical conductivity values;
- the three Euler's anisotropy angles, illustrated in Figure 6.10.

In this simple case, the model domain was characterized into three regions using the same boundaries that characterize the synthetic model. Two different grids were used to solve the inverse problem: an inversion grid that has the same boundaries as the synthetic model, and a finer grid used to compute the forward solutions.

The synthetic model is presented in Figure 6.11 and the discretization used to compute the forward problem is presented in Figure 6.12. Data relative to 13 periods were used – $T \in [1.0; 2.0; 5.0; 10.0; 20.0; 50.0; 100.0; 200.0; 500.0; 1000.0; 2000.0; 5000.0]$ s – along a line of 53 stations located on the vertical grid points, as shown in Figure 6.11.

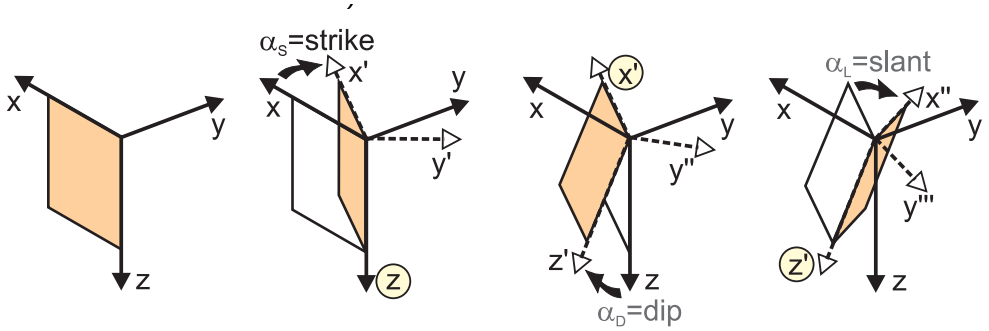


Fig. 6.10.: Euler's rotation used to describe the conductivity tensor. Due to the formal characteristics of the tensor itself, it can be represented by three principal conductivity – equivalents to the eigenvalues of the tensor – and three consecutive rotations (cf. Equation 2.67 for details; figure courtesy of Dr. Josef Pek).

The problem characteristic dimensions are, therefore, 18 (three regions defined by six parameters each) parameters and 5512 data. It was not possible to discern any difference in the inversion results produced with or without the use of the MI penalty function; the RMS was zero within the numeric precision used.

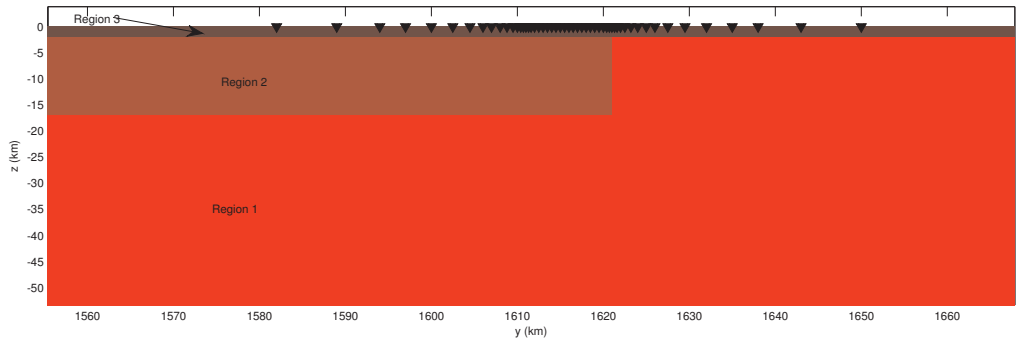


Fig. 6.11.: Synthetic model regions and position of stations. the only vertical discontinuity is reported in the figure. The conductivity values relative to the regions are reported in Table 6.1 (left value).

Table 6.1 contains the resulting model parameters. The results are presented in a physical coordinate system, with the x-axis aligned with the north-south direction and pointing north. This method of representation of the tensor was chosen in order to overcome the problems caused by the non-unique tensor representation in the principal conductivities coordinate system.

Despite the simplicity of the test, it was successful overall. Of the three aims of the test, two were obviously satisfied:

- with a final RMS equal to zero within the machine's precision, the data-fit is perfect;

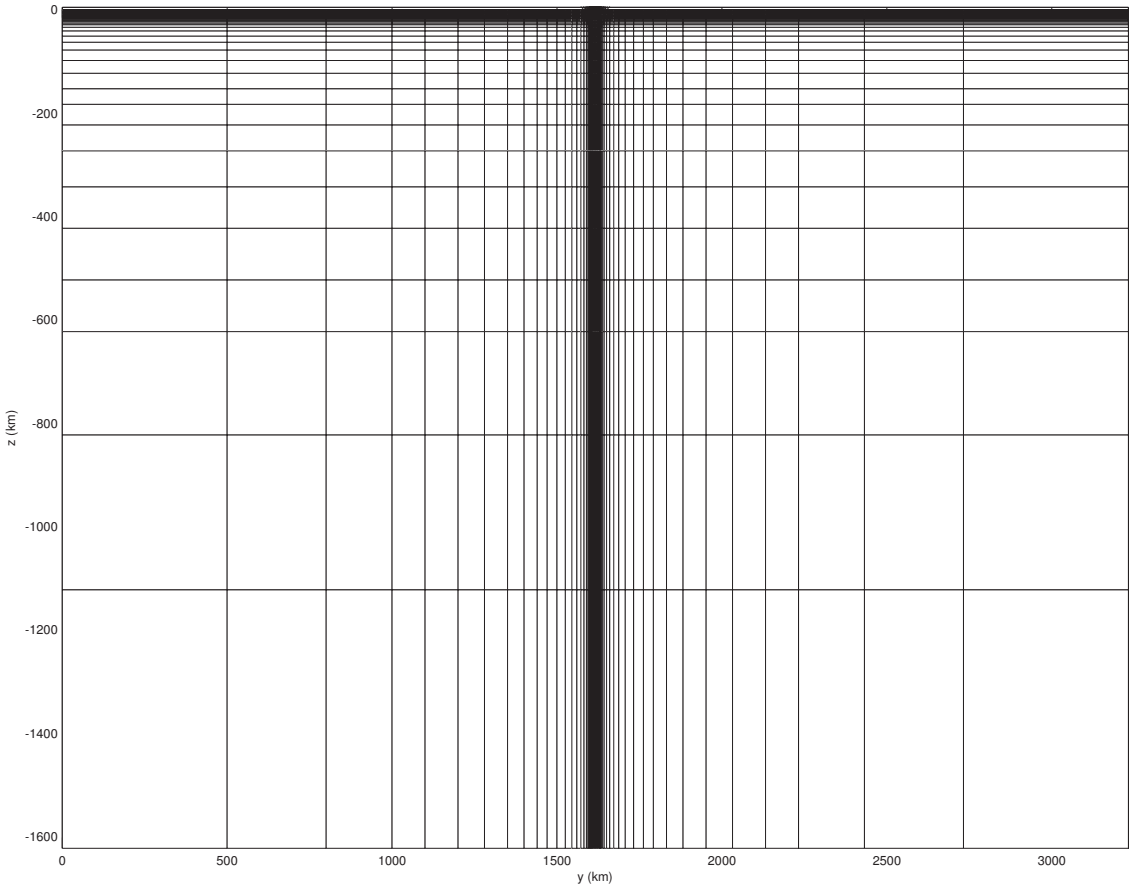


Fig. 6.12.: Discretization of the 2D anisotropic domain used to compute the forward response. The area of interest is located approximatively at the center of the y axis. The domain is much larger than the problem to reduce the influence of the boundary condition in the response.

Region	σ_{xx}	σ_{xy}	σ_{xz}	σ_{yy}	σ_{yz}	σ_{zz}
1	0.001 / 0.001	0.000 / 0.000	0.000 / 0.000	0.001 / 0.001	0.000 / 0.000	0.001 / 0.001
2	0.033 / 0.033	0.000 / 0.000	0.000 / 0.000	0.033 / 0.033	0.000 / 0.000	0.033 / 0.033
3	0.032 / 0.032	0.006 / 0.006	0.000 / 0.000	0.012 / 0.012	0.000 / 0.000	0.323 / 0.324

Tab. 6.1.: True (left value) and retrieved (right value) electrical conductivity values relative to the synthetic test described in Section 6.2.3. The RMS results zero within the machine's precision. All the conductivities are measured in $S \cdot m^{-1}$.

- if the real conductivity values and retrieved presented in Table 6.1 are compared, it is evident that we retrieved the synthetic model.

Evidently, these two results are not achieved easily; the discretization of the inversion domain, compatible with the true boundaries of the synthetic model, assumes an exact knowledge of the structure of the subsurface. Furthermore, the dimensions of the problem

($n \ll m$) is a hypothetical, ideal situation, often unattainable during a "real-world" survey. The complete insensitivity of this problem to λ_{MI} must be noted, and is attributed to the almost perfect solution retrieved without the MI penalty function. The relatively small value of n discouraged us from making use of both the anisotropy penalty function and the Tikhonov regularization.

6.3. Preliminary test conclusions

In this chapter, the objective function has been defined. From a mathematical perspective, this objective function, in conjunction with the inversion algorithm and the forward solver, completely defines the inverse problem we aim to solve. We have tested the ability of our implementation of the Levenberg-Marquardt algorithm to solve the inverse problem in both 1D and 2D anisotropic domains.

In the 1D case, we have been able to explore some of the properties of the mutual information penalty function, in particular its ability to drive the inversion towards a suitable solution and its affect on the stability of the solution with respect to the trade-off parameters.

7 Application of the new approach

7.1. 1D Anisotropic Inversion of Data from Central Germany

A more realistic application of the developed method requires the study of a real dataset. In order to further test the efficiency and the advantages of the MI penalty term, we used the objective function defined in Equation 6.1 to invert a well-known dataset related to the Rhenish Shield region (Central Germany).

In this case, a notable improvement in the retrieved model has been achieved, constraining the model's most-conductive direction with an available seismic model relative to the same region.

In the following section the data used, the reference model and the solution retrieved will be presented and discussed.

7.1.1. Data

The MT data used in this study were acquired in the Rhenish Shield region, Central Germany, between June and September 1997 and October to December 1998 [*Leibecker et al.*, 2002]. From these surveys we analyzed the data from station DIE. The location of the station and the topography of the surrounding region is shown in Fig. 7.1. Data from other stations, deployed in the studied area, present the same characteristics, but are of inferior data quality [*Roux et al.*, 2011]. We considered all the data from the station DIE, ranging in period from 10 s to 4100 s. MT data are affected by distortion caused by small heterogeneities and local structure [*Groom and Bailey*, 1989]. Based on prior studies of this area (e.g. *Leibecker et al.* [2002], *Roux et al.* [2011]) we expect a 1D anisotropic environment, both electrically and seismically. We preprocessed the data, as suggested in *Jones* [2012], to remove the determinable effects of galvanic charges on the electric field from local, small-scale, distorters. After removing distortion, we obtained a set of 18

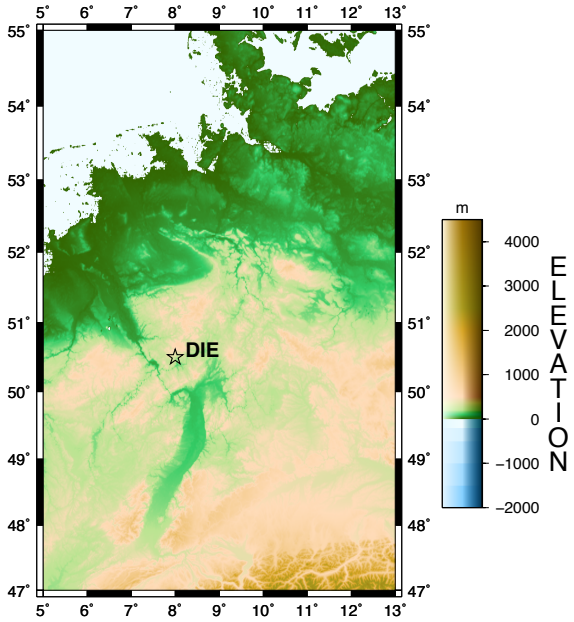


Fig. 7.1.: Topographic map of the studied area. The black star indicate the location of DIE station.

complex impedance tensors $\hat{Z}(T)$, one per period, all with $\mathrm{tr}(\hat{Z}) = 0$; this formal property is required in 1D anisotropic environments (cf. Section 2.4).

7.1.2. The Reference Model

Approximate agreement between geo-electric strike and seismic fast-axis direction in several regions, e.g., southern part of the Slave province [*Eaton et al.*, 2004], central Australia [*Simpson*, 2001], central Germany [*Gatzemeier and Moorkamp*, 2005], the Kaapvaal craton [*Hamilton et al.*, 2006] and the São Francisco craton [*Padilha et al.*, 2006], suggests that a common origin may contribute to both electrical and seismic anisotropy. Usually, lattice-preferred orientation of olivine crystals is considered to be the source of seismic anisotropy in the upper mantle [*Savage*, 1999], but laboratory measurements prove that electrical anisotropy caused by the orientation of olivine crystals cannot be the only source of the anisotropy observed in the field [*Gatzemeier and Tommasi*, 2006]. While the source of upper mantle electric anisotropy is still debated, its effect is clear in MT data. The qualitative agreement in azimuthal anisotropic directions, as detailed above, suggests that the seismic azimuthal anisotropy can effectively constrain the anisotropic strike direction of a concurrent MT model, eventually reducing the non-uniqueness inherent the solution of a one-dimensional anisotropic MT inverse problem..

The selection of a reference model, especially in the proposed inversion scheme, is not a trivial choice. In this case study, the MI distance of the direction of the most electrically

conductive direction was minimized with respect to the seismic fast-axis direction[1]. For the reference model, we chose one of the seismic models obtained by *Roux et al.* [2011], presented in Figure 7.2. In that work, *Roux et al.* [2011] implemented a genetic algorithm (NAGA-II) as their optimization engine, and inverted for a set of suitable seismic and resistivity models. Our reference model is the model A indicated with the continuous black line in the figure. Data fitting relative to the selected seismic model is shown in Figure 7.3. For information regarding the inversion technique used for the seismic data, see *Roux et al.* [2011] and references therein.

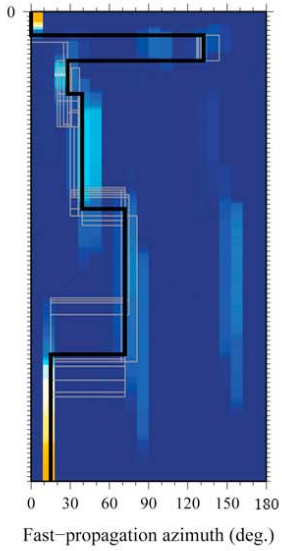


Fig. 7.2.: Final population relative to the seismic fast-axis direction obtained with GA based inversion in [*Roux et al.*, 2011]. The black continuous line is relative to the model that best fit the seismic data (model "A" in *Roux et al.* [2011].

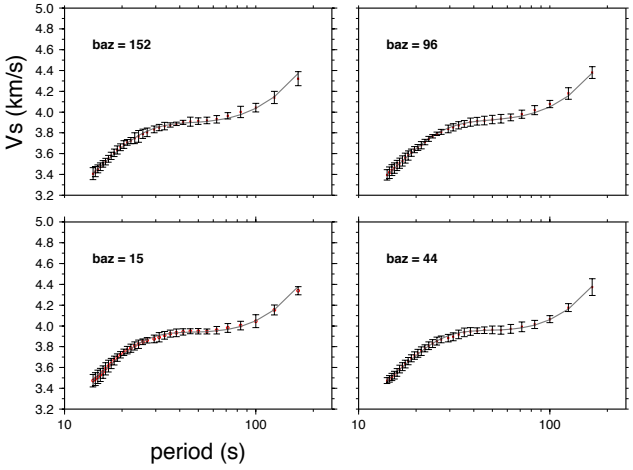


Fig. 7.3.: Seismic data fit from the reference model [*Roux et al.*, 2011].

7.1.3. Inversion details

The objective function minimized was formally identical to Equation 6.1:

$$\phi(\mathbf{m},\mathbf{R}) = \lambda_R\phi_R(\mathbf{m}) + \lambda_A\phi_A(\mathbf{m}) + \lambda_{MI}\phi_{MI}(\mathbf{m},\mathbf{R}) + \sum_{i=1}^{n}\left(\frac{G(\mathbf{m})_i - d_i}{\sigma_i}\right)^2, \qquad (7.1)$$

where $\mathbf{m}$ is the model vector, in which the values of aggregated electrical conductivities (σ_{HI_i} and σ_{LO_i} and the strike directions (β_{s_i}) are stored, $\mathbf{R}$ is the reference model vector, that stores the seismic fast-axis directions, $\phi_\star$ is the penalty function associated to the relative quantity, G is the 1D anisotropic forward operator, d_i is the component of data

[1]The azimuthal direction along which the pressure waves travel with higher velocity

vector relative to the *i-th* component, and σ_i is its statistical error. Non-uniqueness of solutions in MT inverse problems is well-known and documented (e.g. *Parker* [1980]). In 1D, this non-uniqueness is not inherent in the MT method, as there exists a uniqueness theorem [*Bailey*, 1970], but is a consequence of inaccurate, insufficient and inconsistent data. Assessing this non-uniqueness requires the use of a Tikhonov regularization term in the objective function, in order to penalize models that are too rough (e.g. *Constable et al.* [1987]). We added a quadratic term[2], that penalizes the differences between the high and low conductivities, $\sum_i(\log\sigma_{HIi} - \log\sigma_{LOi})^2$, to penalize models that present high anisotropy, and a term to introduce a first-order Tikhonov regularization between the layer-to-layer conductivities, $\lambda_R\phi_R(\mathbf{m}) = \lambda_R L^2\mathbf{m}^2$ where L is the roughening matrix

$$L = \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \cdots & & \\ & & -1 & 1 & \\ & & & -1 & 1 \end{bmatrix}, \tag{7.2}$$

and the product $L \cdot \mathbf{m}$ is a finite-difference approximation that is proportional to the first derivative of $\mathbf{m}$ (this is an example of first-order Tikhonov regularization, as presented in Section 4.7). Trade-off between data-fitting, regularization and distance from the reference model is controlled via the appropriate coefficients, selected ad-hoc during the inversion process. Specifically, any time we changed the trade-off parameters, we either divided them by 2 to reduce them, or multiplied them by 2 to increase them. The influence of anisotropy and regularization are controlled in the inversion process by the trade-off parameters λ_A and λ_R respectively. The influence of the MI in our inversion is controlled in the same way by the trade-off coefficient λ_{MI}. The term relative to the MI distance measure is $\phi_{MI}(\mathbf{m},\mathbf{R}) = I(\mathbf{R};\mathbf{R}) - I(\mathbf{m};\mathbf{R})$. $\sum_{i=1}^{n}\left(\frac{G(\mathbf{m})_i - d_i}{\sigma_i}\right)^2$ is the term that minimizes the misfit measure.

The meaning of the three $\lambda_\star$ values is not absolute, as remarked in Chapter 6. Regularization functions serve to smooth the RMS hyper-surface. This process introduces a bias in the solution, while simultaneously improving the efficiency of the LM algorithm that is not able otherwise to avoid local minima. Values used in the inversion have been selected via a trial-and-error procedure.

In detail, we followed a stepwise strategy in order to introduce complexity in the model progressively. This ensures we begin from simplicity, and introduce more complexity only when required.

1. Inversion of the MT data only starting with an initial guess m_0 to recover an isotropic model m_{ISO}. In this step the coefficient that controls anisotropy, λ_A, is set to 10^{12} (ideally $+\infty$) to prevent any anisotropy in the model. The coefficient that controls

[2] despite the formal way in which this term is cast, it is identical to the $\mathbf{P}(\star)$ operator described by Equation 6.2.

the MI, λ_{MI}, is set to 0, i.e., there is no influence from the reference model. $\lambda_R = 2$ was selected from a set of initial inversions via a simple L-curve method [*Hansen and O'Leary*, 1993].

2. Using m_{ISO} as starting model, we progressively introduced anisotropy into the model by reducing λ_A. For the data herein, in this step we reduced λ_R to the range [1.0 – 2.0]. Again, λ_{MI} is 0.

3. We progressively reduced the λ_R value whilst simultaneously increasing the value of λ_{MI} up to 8. We tested higher values for λ_{MI} but we did not notice further improvements: neither a reduction in RMS value nor an increase in the MI value.

In order to select the number of layers, we performed an F-test on results from the first step of our scheme to validate the hypothesis that a certain number, n layers, is adequate to fit the data. Starting from $n = 2$ logarithmically-spaced layers, we try to fit data progressively better by doubling the number of layers at each iteration. We accepted the hypothesis that $n = 32$ layers are adequate to fit DIE data with a significance level of 0.05. This test ensures that a finer discretization of the subsurface domain, using 64 logarithmically spaced layers instead of 32, does not improve the χ^2 statistic more than 5%.

The errors relative to the model parameters were estimated via the linearized formulæ presented in Section 4.5.

7.1.4. Results

In the previous studies of this area, taken without the constraint of MI, problems relating to constrain the minimum value of the horizontal electrical conductivity ratio $\frac{\sigma_{HI}}{\sigma_{LO}}$ needed to fit the data have been highlighted. In particular, it is important to define the minimum acceptable electrical anisotropy in the asthenosphere to compare with petrological and seismological studies. Constrained by the reference model, our resulting model, shown in Figure 7.4, presents three anisotropic regions, similar to the reference model itself. Inside the 95% confidence interval[3], a difference of more than one order of magnitude between the high conductivity and low conductivity is never required.

Interpreting the anisotropy as an indicator of physical and chemical property changes, it is possible to guess physical or chemical boundaries in the areas in which it arises. With this in mind, we can set the Moho interface between 24 km and 29 km depth, within the layer that present the highest anisotropy, and the LAB, the next important boundary, is set between 85 km and 99 km in depth, again in a highly anisotropic region. Both of these are in good agreement with the values found in independent seismic and MT studies of the area (cf. *Bischoff* [2006] and *Gatzemeier and Moorkamp* [2005]). In Figure 7.5 the

[3] computed as shown in section 4.5

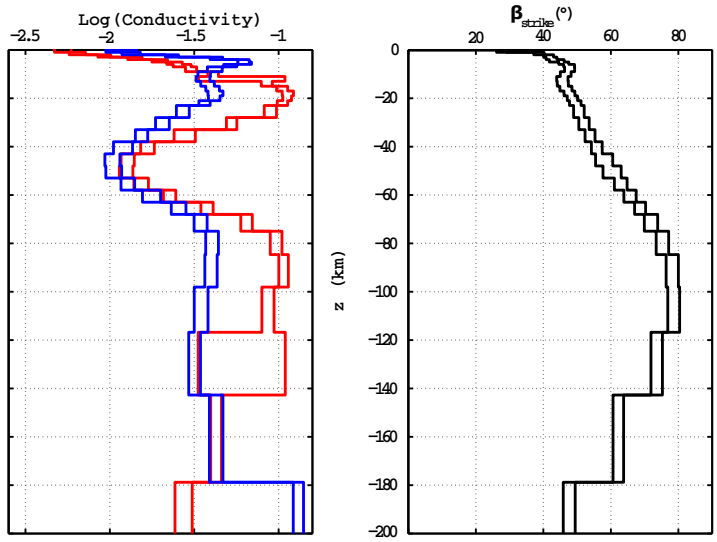


Fig. 7.4.: Electrical conductivity model from DIE station in Central Germany, relative to the complete discretization of the domain. The left panel present the logarithm of conductivities ± the computed error. Blue and red are relative to the two different aggregate conductivities. The right panel present the anisotropic strike direction ± the direction error. A detail of the upper 50 km of this model is presented in fugure 7.5.

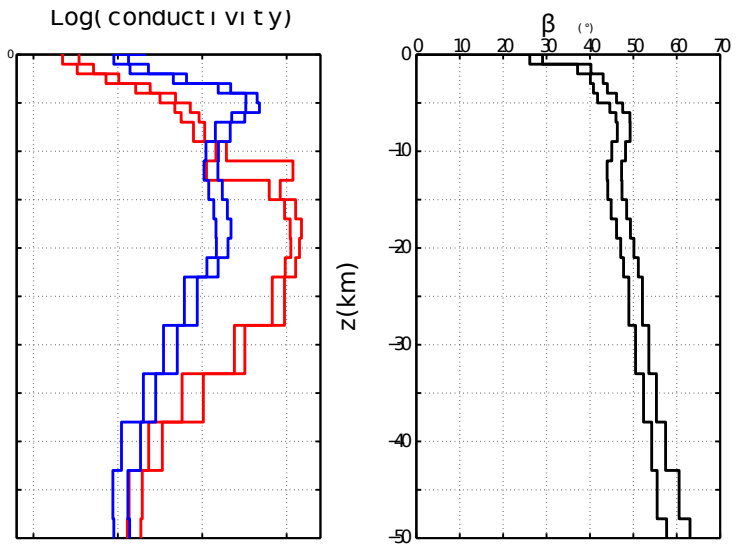


Fig. 7.5.: Electrical conductivity model from DIE station in Central Germany, relative to the top 50 km. The left panel present the logarithm of conductivities ± the computed error. Blue and red are relative to the two different aggregate conductivities.The right panel present the anisotropic strike direction ± the direction error.

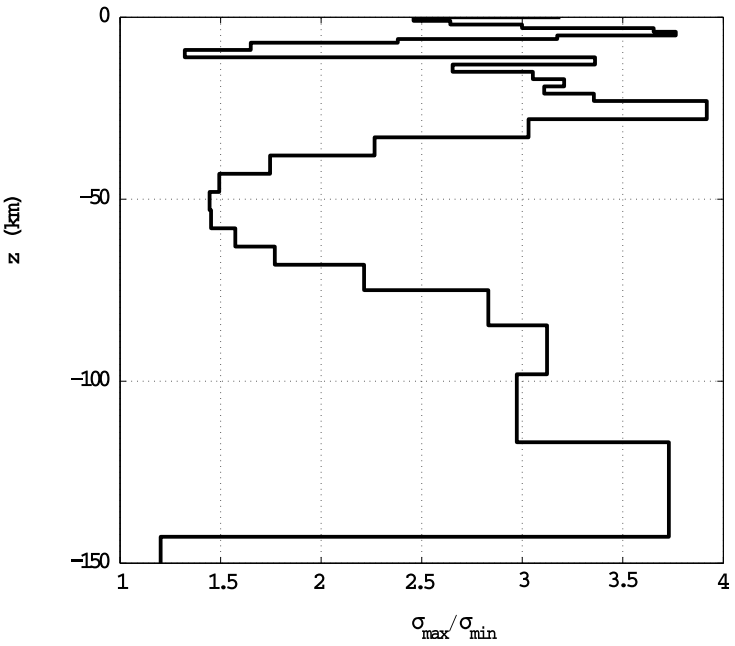


Fig. 7.6.: Maximum anisotropy rate $\frac{\sigma_{HI}+\mathrm{var}(\sigma_{HI})}{\sigma_{LO}-\mathrm{var}(\sigma_{HI})}$ computed from the model in figure 7.4.

retrieved model for to the top 50 km, where the Mohorovičić discontinuity is located, is reported. The penetration depth computed from the final model is approximate by

$$\delta \approx 0.5\sqrt{\rho_a T}\ \mathrm{km}\ \approx 120\ \mathrm{km}, \tag{7.3}$$

with $\rho_a = \mu_0\omega|Z|^2$ (cf. *Jones* [1999] or Chapter 2 for details).

Anisotropy direction is resolved in the interval of $(50 \pm 2)^\circ$ in the crust – in good agreement with the absolute plate motion direction of $[50^\circ - 55^\circ]$ for Central Europe determined by *Gripp* [1990] – and $(78 \pm 2)^\circ$ at the depth of about 100 km, which is oddly at a high angle to absolute plate motion (APM). These anisotropy directions roughly correspond to direction of NE-SW in the crust and E-W in the upper mantle. Interpretation of these directions is beyond the goal of this thesis.

The strongest anisotropy in the model is approximately at a depth of $70\ km < z < 150\ km$, and the maximum anisotropy rate, estimated by ratio $\frac{\sigma_{HI}+\mathrm{var}\sigma_{HI}}{\sigma_{LO}-\mathrm{var}\sigma_{LO}}$, in which var is the variance of the considered model parameters, reported in Figure 7.6. Overall, our results from station DIE are in good agreement with models obtained by past studies in this area (e.g. *Leibecker et al.* [2002], *Roux et al.* [2011]), with the significant improvement that our modeling vastly improves the constraint on acceptable maximum electrical anisotropy. In detail, the maximum electrical anisotropy predicted by our model – in the 95% confidence interval – is never higher then 4. This value is in agreement with the values derived from laboratory data for olivine [*Gatzemeier and Tommasi*, 2006]. In Figure 7.7

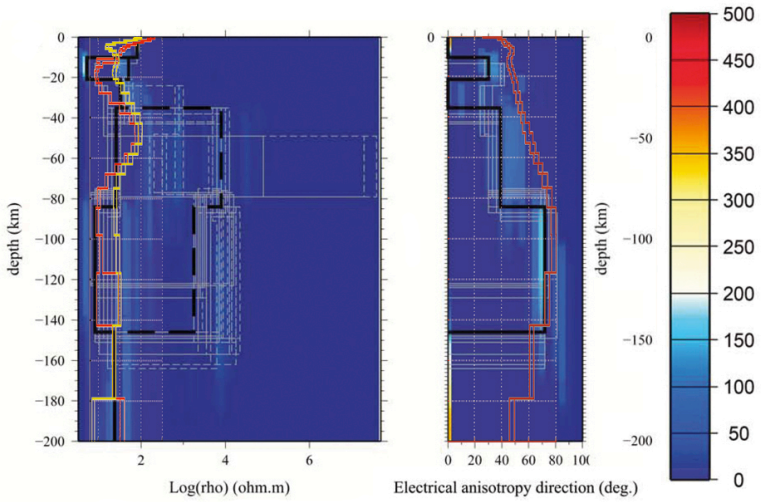


Fig. 7.7.: Retrieved model (yellow and red) compared with results from *Roux et al.* [2011] model A (black solid and dashed line). It is evident the constraint on resistivity in resistive direction.

the retrieved model is compared with results from *Roux et al.* [2011], with the constraint on resistivity in resistive direction results evident.

The data fit is illustrated in Figure 7.8. We highlight the optimum agreement between measured and predicted data along all the periods considered and relative to all the elements of the impedance tensor.

The chosen discretization of the domain prevents optimum fitting of high frequency impedances, where data relative to the longest periods are precisely predicted by the model. The process of removing distortion from the raw data produces non-negligible errors in the resulting impedances [*Jones*, 2012]. Therefore, propagation of errors via the usual linearized formula is not allowed for these impedances, preventing us from modeling apparent resistivities and phases, more commonly displayed in the MT community. The achieved RMS = 0.24 in combination with its expected value E[RMS] = 1 highlights that the parametric errors relative to the impedances are probably overestimated (see, e.g., *Chave and Jones* [1997] for a comparison of parametric and jackknife error estimates).

Lastly, we examine the estimated joint probability distribution of seismic V_s fast-axis azimuth angle and high electrical conductivity anisotropic strike angle shown in Figure 7.9. The estimated joint probability distribution shows a pattern that suggests that overall similar values of seismic fast-axis direction correspond to the high conductivity directions. In more detail, the three high probability areas are approximately aligned, suggesting a constant offset between the considered directions.

These considerations reinforce the hypothesis that the most conductive electric direction and the fast-axis seismic direction have causes that are, at least partially, in common. Given that the pattern appears non-continuous, we cannot confirm the hypothesis of a direct relation between seismic and electrical anisotropy, while the clear trend suggests a possible relation between the sources of the anisotropy. Clustering in joint probability

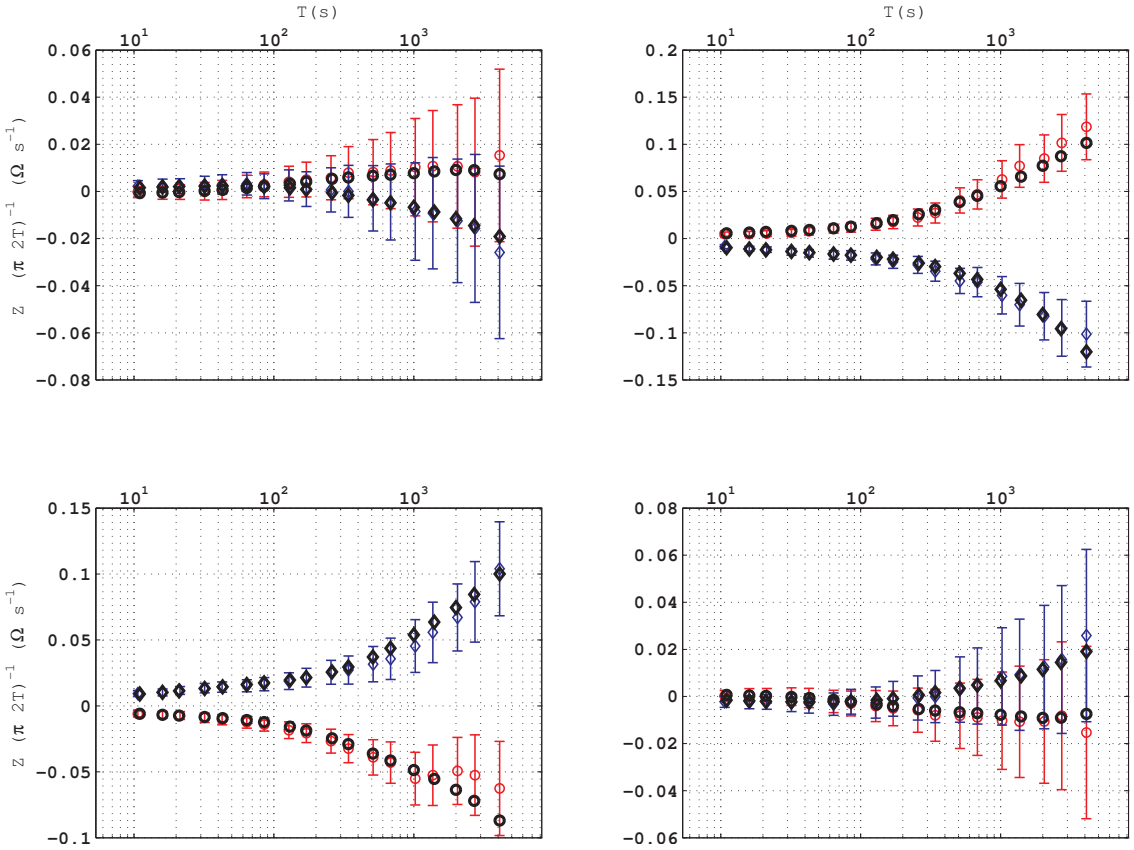


Fig. 7.8.: Measured and predicted data for DIE station. Red circles represent the real parts, blue diamonds represent imaginary parts, black circles and diamonds are the relative predicted data. The top left panel is relative to Z_{xx}, the top right to Z_{xy}, the bottom left to Z_{yx} and the bottom right to Z_{yy}. Y axis are scaled by the factor $\frac{1}{2\cdot\pi\cdot T}$ to highlight the long period data fit.

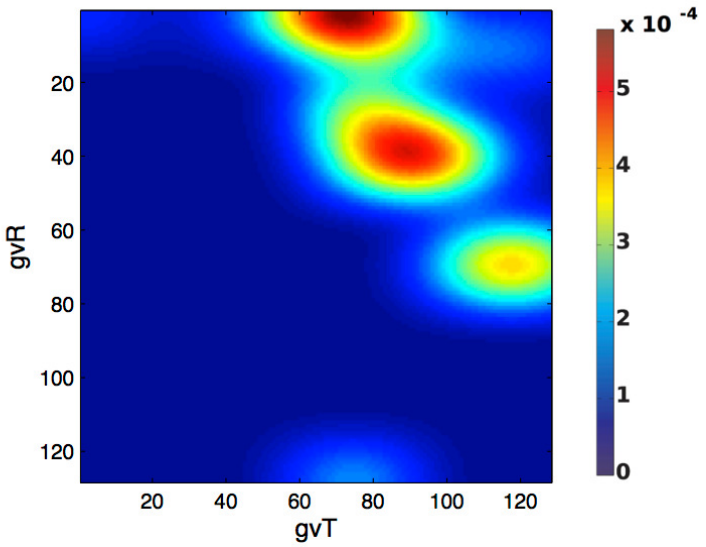


Fig. 7.9.: Computed joint probability distribution of anisotropic electric strike direction (x-axis) and azimuthal seismic fast-axis direction (y-axis).This function have been estimated fixing $s = 1.0$. The values on the axis correspond to the discretization – in this case in 128 values – of the values that define the MT strike directions and the ones that depicts the reference model **R** shown in Figure 7.3 . Thus the probability that a value $R^\star : \frac{\pi \cdot n}{128} \leq R^\star \leq \frac{\pi \cdot (n+1)}{128}$ in figure **R** correspond to a certain value $m^\star : \frac{\pi \cdot i}{128} \leq m^\star \leq \frac{\pi \cdot (i+1)}{128}$ is given by the value of $p_s(n, i)$

distribution is due to the use of a layered reference model.

7.2. Complex 2D Anisotropic Synthetic Test

The forward solver used to predict data from a two-dimensional domain allows us to build the model using up to 20 different homogeneous anisotropic subdomains. In this test we build a complex domain, thus we mixed together these subdomains, drawing the model geometry presented in Figure 7.10. We used the preview discretization – the one presented in Figure 6.12 – to compute an accurate forward response. The synthetic model is the same examined in Section 6.2.3, and the synthetic data are relative to the same stations and the same periods. The number of data is thus, as in the preview test, $m = 5512$, with which we estimate $n = 60$ parameters that constitute the model $\mathbf{m}$.

7.2.1. Results

As in the preview test, we begin this test by defining an initial guess model $\mathbf{m^0}$, and we define it as an homogeneous isotropic half-space characterized by its electrical resistivity $\rho = 1000\ \Omega \cdot m$.

We performed this test twice: once using noise free synthetic data and once adding 7% random noise. In both cases we retrieved the same final model.

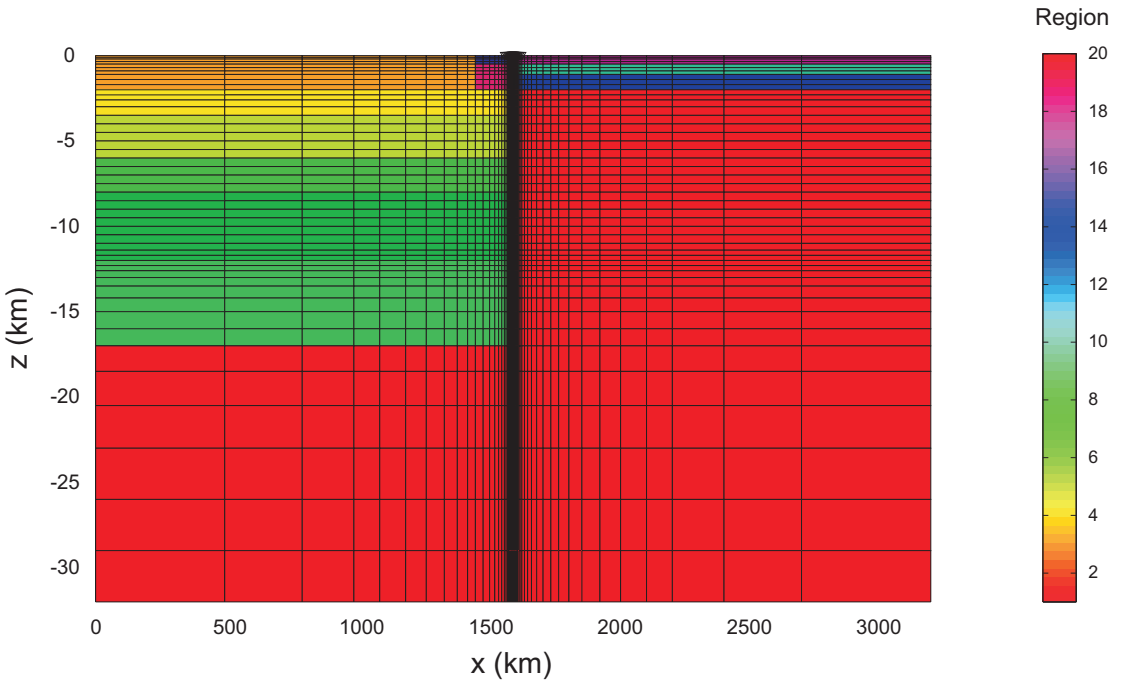


Fig. 7.10.: Discretization of the inversion grid used in the second 2D synthetic test.

The final model parameters are presented in Table 7.3.

The computed model is characterized by a good agreement with the synthetic one relative to all the conductivity tensor elements but σ_{zz}. This in a known result, in fact the MT method is not particularly sensitive to vertical conductivity (in 1D, as seen, MT is completely insensitive to σ_{zz}). This is confirmed by the value of the misfit term of the objective function computed in correspondence to the retrieved model, value reported in Table 7.2.

Notwithstanding the incorrect values of σ_{zz} computed, the data fit results were excellent in the noise free test, and acceptable – $RMS = 1.17$; $E[RMS] = 1.00$ – for the inverse problem relative to the noise dataset.

7.2.2. Results obtained via mutual information constraint

Due to the scarce sensitivity of the 2D anisotropic inverse problem with respect to the σ_{zz} parameter, we try to constrain σ_{zz} with a compatible reference model, obtained by

$$\mathbf{R}_i = \sin(\sigma^{true}_{zz\ i}). \tag{7.4}$$

The calculated model parameters relative to the inversion with the MI constraint are reported in Table 7.3. The final misfit value, relative to both the noise free dataset and the noise dataset, was slightly improved, as reported in Table 7.4,.

Region	σ_{xx}	σ_{xy}	σ_{xz}	σ_{yy}	σ_{yz}	σ_{zz}
1	0.001 / 0.001	0.000 / 0.000	0.000 / 0.000	0.001 / 0.001	0.000 / 0.000	0.001 / 0.001
2	0.033 / 0.033	0.000 / 0.000	0.000 / 0.000	0.033 / 0.033	0.000 / 0.000	0.033 / 0.033
3	0.032 / 0.032	0.006 / 0.006	0.000 / 0.000	0.012 / 0.012	0.000 / 0.000	0.323 / 0.001
4	0.033 / 0.033	0.000 / 0.000	0.000 / 0.000	0.033 / 0.033	0.000 / 0.000	0.033 / 0.035
5	0.033 / 0.033	0.000 / 0.000	0.000 / 0.000	0.033 / 0.033	0.000 / 0.000	0.033 / 0.037
6	0.033 / 0.033	0.000 / 0.000	0.000 / 0.000	0.033 / 0.033	0.000 / 0.000	0.033 / 0.027
7	0.033 / 0.033	0.000 / 0.000	0.000 / 0.000	0.033 / 0.033	0.000 / 0.000	0.033 / 0.033
8	0.033 / 0.033	0.000 / 0.000	0.000 / 0.000	0.033 / 0.033	0.000 / 0.000	0.033 / 0.012
9	0.001 / 0.001	0.000 / 0.000	0.000 / 0.000	0.001 / 0.001	0.000 / 0.000	0.001 / 0.001
10	0.032 / 0.032	0.006 / 0.006	0.000 / 0.000	0.012 / 0.016	0.000 / 0.000	0.323 / 0.067
11	0.032 / 0.032	0.006 / 0.006	0.000 / 0.000	0.012 / 0.012	0.000 / 0.000	0.323 / 0.454
12	0.032 / 0.032	0.006 / 0.006	0.000 / 0.000	0.012 / 0.012	0.000 / 0.000	0.323 / 0.333
13	0.032 / 0.032	0.006 / 0.006	0.000 / 0.000	0.012 / 0.012	0.000 / 0.000	0.323 / 0.097
14	0.032 / 0.032	0.006 / 0.006	0.000 / 0.000	0.012 / 0.012	0.000 / 0.000	0.323 / 0.027
15	0.032 / 0.032	0.006 / 0.006	0.000 / 0.000	0.012 / 0.012	0.000 / 0.000	0.323 / 0.121
16	0.032 / 0.032	0.006 / 0.006	0.000 / 0.000	0.012 / 0.012	0.000 / 0.000	0.323 / 0.324
17	0.032 / 0.032	0.006 / 0.006	0.000 / 0.000	0.012 / 0.012	0.000 / 0.000	0.323 / 0.029
18	0.032 / 0.032	0.006 / 0.006	0.000 / 0.000	0.012 / 0.012	0.000 / 0.000	0.323 / 0.294
19	0.032 / 0.032	0.006 / 0.006	0.000 / 0.000	0.012 / 0.012	0.000 / 0.000	0.323 / 0.335
20	0.032 / 0.032	0.006 / 0.006	0.000 / 0.000	0.012 / 0.012	0.000 / 0.000	0.323 / 0.337

Tab. 7.1.: True (left value) and computed (right value) electrical conductivity values relative to the synthetic test described in Section 7.2.1.

7.2.3. Discussion

With this synthetic test, we used the developed code to invert 2D anisotropic data. Whereas the synthetic model represents a simple configuration of the subsurface, the model parameters retrieved were substantially similar to the synthetic ones.

In Figure 7.12, 7.13 and 7.14 the data fit relative to some sample stations is reported. The data fit relative to the stations 2, 22 and 42 is computed by the model retrieved in Section 7.2.1, while the data fit relative to the stations 12, 32 and 52 is computed by the model retrieved in Section 7.2.2. In both cases we present the noise dataset. The position of the stations is reported in Figure 7.11.

According to the misfit value, we do not find important differences in the quality of the data fit. This agreement in the data fit quality suggests that the refinement in the model parameters is due to the introduction in the objective function of the term $\lambda_{MI}^2 [I(\mathbf{m_{zz}}; \mathbf{R})]$, in which $\mathbf{m_{zz}}$ represents the model elements storing the parameters relative to σ_{zz}.

This means that, despite the lack of sensitivity of the solution with respect to the σ_{zz} parameters, the compatible reference model was able to "drive" the inversion closer to the real solution.

random noise %	$\sum_{i=1}^{m}\left[\frac{G(\mathbf{m})_i - d_i}{\sigma_i}\right]^2$
0	$1.41 \cdot 10^{-5}$
7	1.17

Tab. 7.2.: Misfit term in the objective function relative to the test presented in Section 7.2.1.

Region	σ_{xx}	σ_{xy}	σ_{xz}	σ_{yy}	σ_{yz}	σ_{zz}
1	0.001 / 0.001	0.000 / 0.000	0.000 / 0.000	0.001 / 0.001	0.000 / 0.000	0.001 / 0.001
2	0.033 / 0.033	0.000 / 0.000	0.000 / 0.000	0.033 / 0.033	0.000 / 0.000	0.033 / 0.033
3	0.032 / 0.032	0.006 / 0.006	0.000 / 0.000	0.012 / 0.012	0.000 / 0.000	0.323 / 0.330
4	0.033 / 0.033	0.000 / 0.000	0.000 / 0.000	0.033 / 0.033	0.000 / 0.000	0.033 / 0.033
5	0.033 / 0.033	0.000 / 0.000	0.000 / 0.000	0.033 / 0.033	0.000 / 0.000	0.033 / 0.033
6	0.033 / 0.033	0.000 / 0.000	0.000 / 0.000	0.033 / 0.033	0.000 / 0.000	0.033 / 0.033
7	0.033 / 0.033	0.000 / 0.000	0.000 / 0.000	0.033 / 0.033	0.000 / 0.000	0.033 / 0.033
8	0.033 / 0.033	0.000 / 0.000	0.000 / 0.000	0.033 / 0.033	0.000 / 0.000	0.033 / 0.033
9	0.001 / 0.001	0.000 / 0.000	0.000 / 0.000	0.001 / 0.001	0.000 / 0.000	0.001 / 0.001
10	0.032 / 0.032	0.006 / 0.006	0.000 / 0.000	0.012 / 0.016	0.000 / 0.000	0.323 / 0.291
11	0.032 / 0.032	0.006 / 0.006	0.000 / 0.000	0.012 / 0.012	0.000 / 0.000	0.323 / 0.287
12	0.032 / 0.032	0.006 / 0.006	0.000 / 0.000	0.012 / 0.012	0.000 / 0.000	0.323 / 0.341
13	0.032 / 0.032	0.006 / 0.006	0.000 / 0.000	0.012 / 0.012	0.000 / 0.000	0.323 / 0.321
14	0.032 / 0.032	0.006 / 0.006	0.000 / 0.000	0.012 / 0.012	0.000 / 0.000	0.323 / 0.330
15	0.032 / 0.032	0.006 / 0.006	0.000 / 0.000	0.012 / 0.012	0.000 / 0.000	0.323 / 0.327
16	0.032 / 0.032	0.006 / 0.006	0.000 / 0.000	0.012 / 0.012	0.000 / 0.000	0.323 / 0.324
17	0.032 / 0.032	0.006 / 0.006	0.000 / 0.000	0.012 / 0.012	0.000 / 0.000	0.323 / 0.333
18	0.032 / 0.032	0.006 / 0.006	0.000 / 0.000	0.012 / 0.012	0.000 / 0.000	0.323 / 0.324
19	0.032 / 0.032	0.006 / 0.006	0.000 / 0.000	0.012 / 0.012	0.000 / 0.000	0.323 / 0.324
20	0.032 / 0.032	0.006 / 0.006	0.000 / 0.000	0.012 / 0.012	0.000 / 0.000	0.323 / 0.324

Tab. 7.3.: True (left value) and computed (right value) electrical conductivity values relative to the synthetic test described in Section 7.2.2.

This effect is due to the non-uniqueness of the solution, summarized by *Parker* [1983] in the remark "for practical data the matter of uniqueness of solution is trivial: infinitely many profiles can fit the data if one can". The improvements in the model retrieved in Section 7.2.2, with respect to the one calculated in Section 7.2.1, are thus obtained via the MI based penalty function.

As the last step of this test we present – in Figure 7.15 – the JPD relative to both the retrieved models with the reference model. The comparison of the two JPD is relatively simple: if the MI penalty function is minimized in the inversion, the joint probability distribution, estimated from values of the final result and the reference model, is characterized by a regular clustering, whereas if the MI penalty function is not taken in account, the es-

random noise %	$\sum_{i=1}^{m}\left[\frac{G(\mathbf{m})_i - d_i}{\sigma_i}\right]^2$
0	$2.21 \cdot 10^{-7}$
7	1.03

Tab. 7.4.: Misfit term in the objective function relative to the test presented in Section 7.2.2.

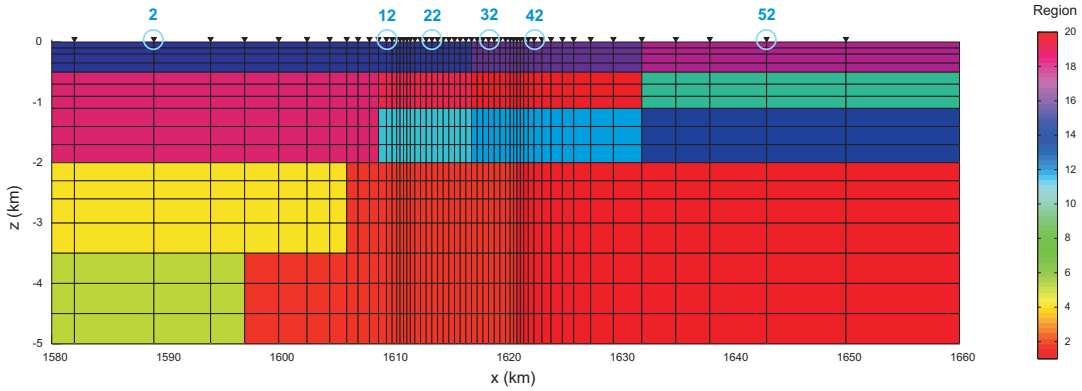
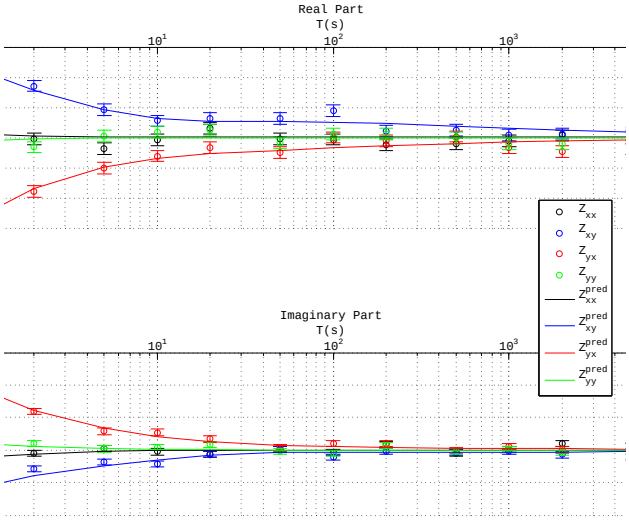


Fig. 7.11.: Position, along the synthetic profile, of the sample stations.

timated JPD clustering results more sparse. This sparsity is a clear index of an incoherent shared information between the final model and the reference one.



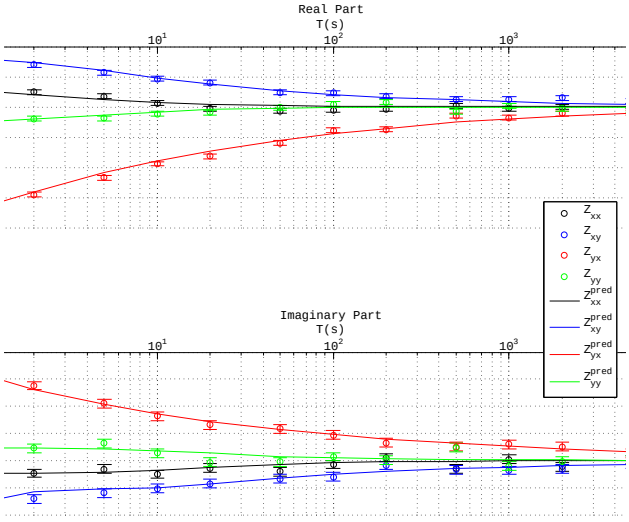
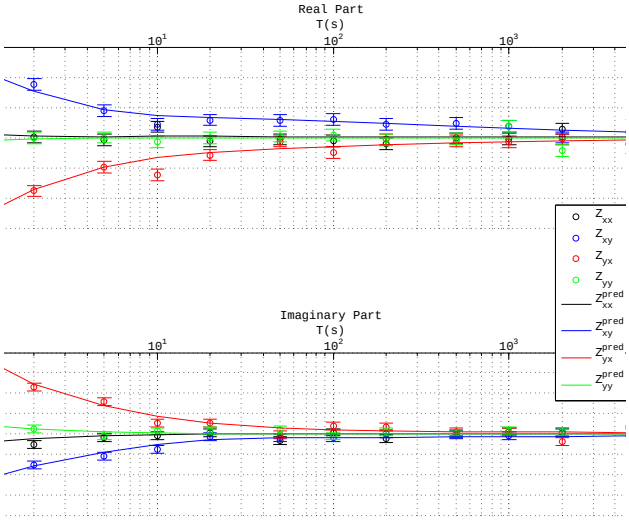


Fig. 7.12.: Upper panel: data fit relative to the station 2. This data fit was obtained without the MI penalty function. **Bottom panel:** data fit relative to the station 52. This data fit was obtained with the MI penalty function.



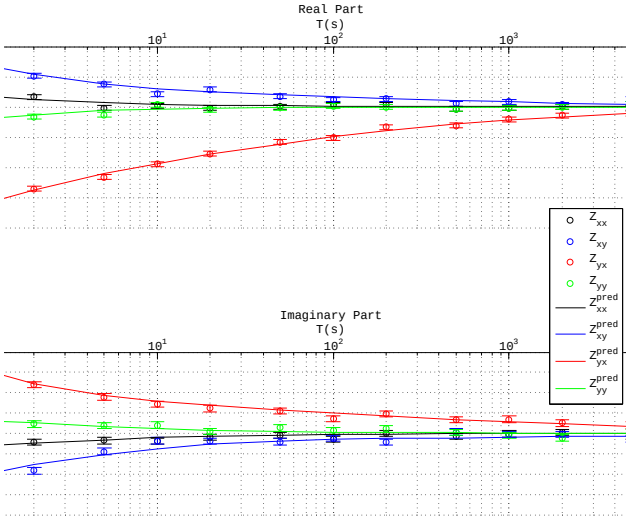
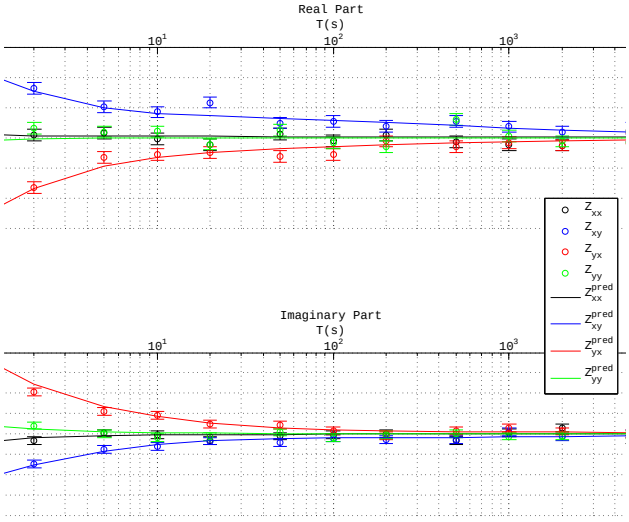


Fig. 7.13.: Upper panel: data fit relative to the station 12. This data fit was obtained without the MI penalty function. **Bottom panel:** data fit relative to the station 42. This data fit was obtained with the MI penalty function.



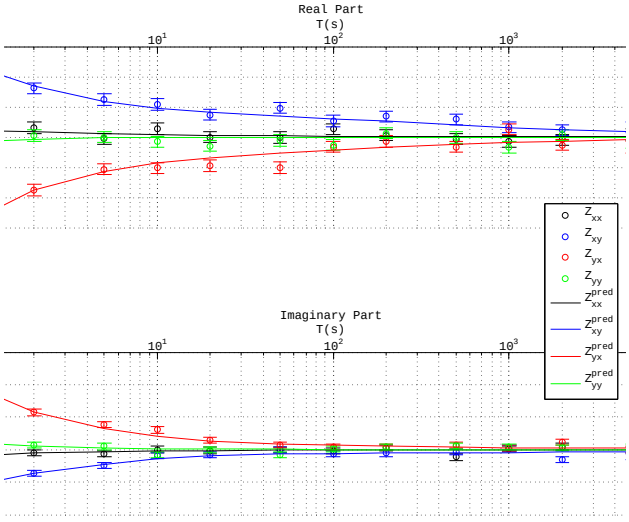


Fig. 7.14.: Upper panel: data fit relative to the station 22. This data fit was obtained without the MI penalty function. **Bottom panel:** data fit relative to the station 32. This data fit was obtained with the MI penalty function.

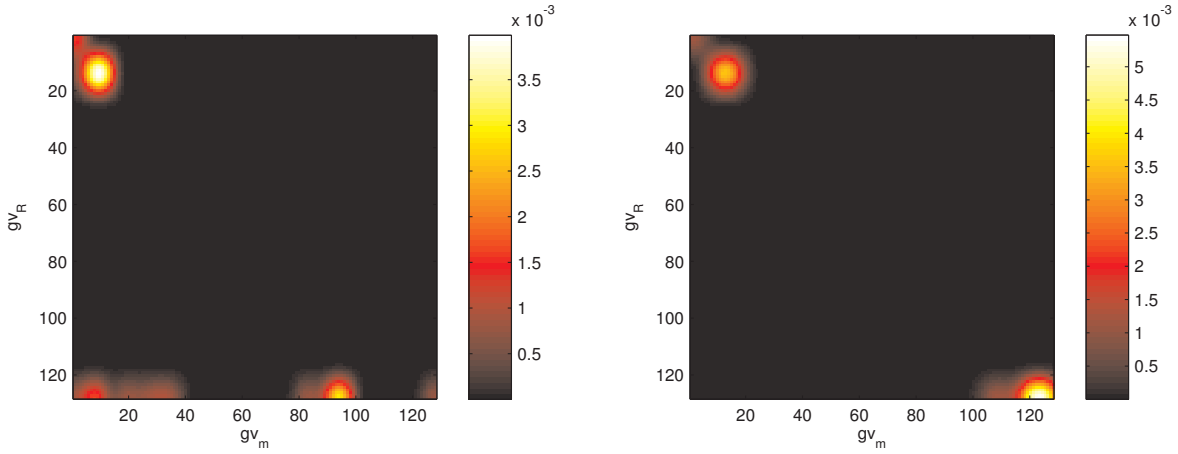


Fig. 7.15.: Left panel: JPD estimated from the MI non-constrained model and **R** . **Right panel**: JPD estimated from the MI constrained model and **R**

7.2.4. Synthetic Test with Vertical Anisotropy

As specified before, one of the main problems with general[4] anisotropic direction is the sensitivity to the vertical anisotropy. To test possible improvements achievable with the use of MI constraint term, a test that considers the same underlying geometry as before, with different anisotropic directions, was tried. In table 7.5 results obtained without the use of MI are presented.

Region	σ_{xx}	σ_{xy}	σ_{xz}	σ_{yy}	σ_{yz}	σ_{zz}
1	0.001 / 0.001	0.000 / 0.000	0.000 / 0.000	0.001 / 0.001	0.000 / 0.000	0.001 / 0.001
2	0.003 / 0.003	0.000 / 0.000	0.000 / 0.000	0.003 / 0.004	0.000 / 0.000	0.003 / 0.004
3	0.098 / 0.137	0.008 / 0.039	-0.012 / 0.008	0.070 / 0.018	0.045 / 0.003	0.034 / 0.001
4	0.003 / 0.002	0.000 / 0.000	0.001 / 0.000	0.003 / 0.002	0.000 /-0.001	0.003 / 0.002
5	0.003 / 0.002	0.000 / 0.000	0.000 / 0.000	0.003 / 0.004	0.000 / 0.000	0.003 / 0.002
6	0.003 / 0.012	0.000 / 0.000	0.000 / 0.002	0.003 / 0.001	0.000 / 0.001	0.003 / 0.004
7	0.003 / 0.003	0.000 /-0.001	0.000 /-0.006	0.003 / 0.008	0.000 / 0.005	0.003 / 0.012
8	0.003 / 0.003	0.000 / 0.004	0.000 / 0.002	0.003 / 0.007	0.000 / 0.002	0.003 / 0.001
9	0.001 / 0.011	0.000 /-0.009	0.000 / 0.000	0.001 / 0.002	0.000 /-0.017	0.001 / 0.034
10	0.098 / 0.109	0.008 / 0.096	-0.012 /-0.028	0.070 / 0.323	0.045 /-0.124	0.034 / 0.050
11	0.098 / 0.095	0.008 / 0.017	-0.012 / 0.014	0.070 / 0.032	0.045 /-0.064	0.034 / 0.156
12	0.098 / 0.097	0.008 / 0.016	-0.012 /-0.002	0.070 / 0.039	0.045 / 0.008	0.034 / 0.002
13	0.098 / 0.095	0.008 / 0.017	-0.012 /-0.002	0.070 / 0.024	0.045 / 0.005	0.034 / 0.001
14	0.098 / 0.095	0.008 / 0.025	-0.012 / 0.000	0.070 / 0.007	0.045 / 0.000	0.034 / 0.000
15	0.098 / 0.097	0.008 / 0.025	-0.012 /-0.001	0.070 /0.008	0.045 / 0.000	0.034 / 0.000
16	0.098 / 0.097	0.008 / 0.025	-0.012 /-0.001	0.070 / 0.007	0.045 / 0.000	0.034 / 0.000
17	0.098 / 0.097	0.008 / 0.025	-0.012 /-0.001	0.070 / 0.007	0.045 / 0.000	0.034 / 0.000
18	0.098 / 0.094	0.008 / 0.025	-0.012 / 0.000	0.070 / 0.017	0.045 /-0.004	0.034 / 0.003
19	0.098 / 0.095	0.008 / 0.032	-0.012 /-0.004	0.070 / 0.039	0.045 /-0.014	0.034 / 0.009
20	0.098 / 0.096	0.008 / 0.067	-0.012 /-0.019	0.070 / 0.377	0.045 /-0.162	0.034 / 0.072

Tab. 7.5.: True (left value) and computed (right value) electrical conductivity values relative to the synthetic test described in Section 7.2.4. These results are obtained without the use of the MI based constraint.

Clearly, the original "true" model was not retrieved. The initial RMS was $1.3 \cdot 10^{31}$ and the final – obtained after 115 iterations – was 9.7. The final RMS itself allows to state that the retrieved model is incorrect. The examination of retrieved values for conductivity confirms that the poor resolution of vertical conductivity prevent the algorithm to retrieve the correct model parameters.

In table 7.7 results obtained using the MI constraint are reported. Apparently the introduction of MI based term in the objective function leads, in this test as in the one presented

[4] meaning using the full electrical conductivity tensor

in Section 7.2, toward a result closer to the real one. In Table 7.6 the final RMS relative to the test presented is reported.

	RMS
without MI term	9.7
with MI term	2.9

Tab. 7.6.: RMS relative to the test described in Section 7.2.4. The test was done assuming an error floor equal to 7%.

Region	σ_{xx}	σ_{xy}	σ_{xz}	σ_{yy}	σ_{yz}	σ_{zz}
1	0.001 / 0.001	0.000 / 0.000	0.000 / 0.000	0.001 / 0.001	0.000 / 0.000	0.001 / 0.001
2	0.003 / 0.003	0.000 / 0.000	0.001 / 0.001	0.003 / 0.003	0.000 /0.000	0.003 / 0.003
3	0.098 / 0.098	0.008 / 0.008	-0.012 /-0.010	0.070 / 0.072	0.045 / 0.041	0.034 / 0.032
4	0.003 / 0.003	0.000 / 0.000	0.001 / 0.001	0.003 / 0.003	0.000 /0.000	0.003 / 0.004
5	0.003 / 0.003	0.000 / 0.000	0.001 / 0.001	0.003 / 0.003	0.000 /0.000	0.003 / 0.003
6	0.003 / 0.003	0.000 / 0.000	0.001 / 0.001	0.003 / 0.003	0.000 /-0.001	0.003 / 0.002
7	0.003 / 0.003	0.000 / 0.000	0.001 / 0.001	0.003 / 0.003	0.000 /0.000	0.003 / 0.003
8	0.003 / 0.003	0.000 / 0.000	0.001 / 0.001	0.003 / 0.003	0.000 /0.000	0.003 / 0.005
9	0.001 / 0.001	0.000 /0.000	0.000 / 0.000	0.000 / 0.000	0.000 /0.000	0.001 / 0.001
10	0.098 / 0.097	0.008 / 0.007	-0.012 /-0.011	0.070 / 0.071	0.045 / 0.048	0.034 / 0.037
11	0.098 / 0.099	0.008 / 0.012	-0.012 /-0.019	0.070 / 0.074	0.045 / 0.045	0.034 / 0.036
12	0.098 / 0.099	0.008 / 0.011	-0.012 /-0.016	0.070 / 0.072	0.045 / 0.044	0.034 / 0.039
13	0.098 / 0.093	0.008 / 0.004	-0.012 /-0.021	0.070 / 0.055	0.045 / 0.054	0.034 / 0.041
14	0.098 / 0.095	0.008 / 0.016	-0.012 /-0.009	0.070 / 0.058	0.045 / 0.036	0.034 / 0.022
15	0.098 / 0.095	0.008 / 0.006	-0.012 /-0.012	0.070 / 0.074	0.045 / 0.034	0.034 / 0.019
16	0.098 / 0.097	0.008 / 0.010	-0.012 /-0.013	0.070 / 0.070	0.045 / 0.037	0.034 / 0.011
17	0.098 / 0.096	0.008 / 0.009	-0.012 /-0.016	0.070 / 0.072	0.045 / 0.055	0.034 / 0.012
18	0.098 / 0.096	0.008 / 0.006	-0.012 /-0.015	0.070 / 0.076	0.045 / 0.042	0.034 / 0.044
19	0.098 / 0.096	0.008 / 0.005	-0.012 /-0.022	0.070 / 0.066	0.045 / 0.042	0.034 / 0.045
20	0.098 / 0.096	0.008 / 0.014	-0.012 /-0.015	0.070 / 0.069	0.045 / 0.044	0.034 / 0.033

Tab. 7.7.: True (left value) and computed (right value) electrical conductivity values relative to the synthetic test described in Section 7.2.4. These results are obtained using the MI based constraint.

Finally, the JPD relative to both inversions are reported in Figure 7.16. Again, the JPD with minimum entropy presents a more regular pattern. Even in this case the improvements achieved by the use of MI penalty function come at a cost: in fact, without the knowledge of the underlying geometry and the correct reference model, the term introduced in the objective function does not carry itself particular meanings (while e.g. a regularization term does).

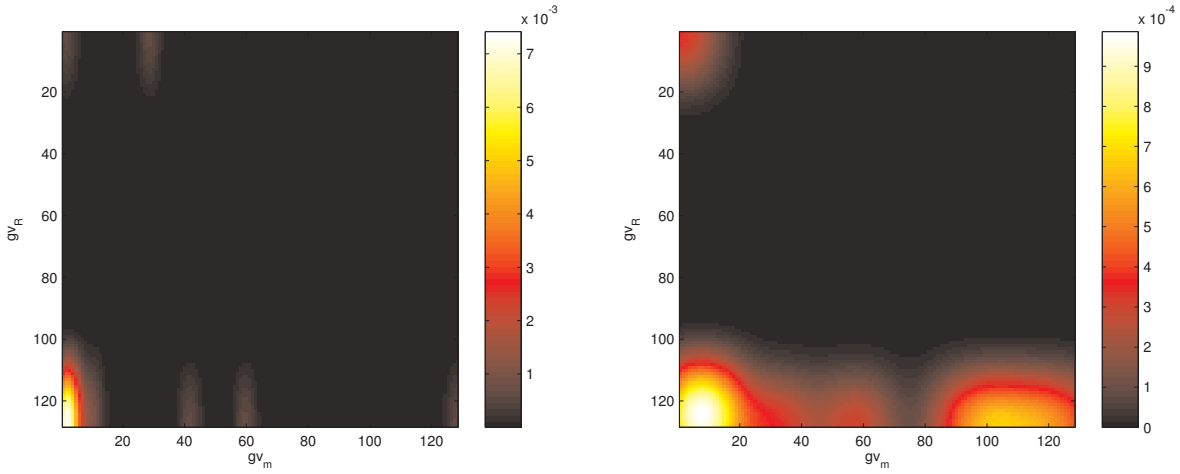


Fig. 7.16.: Left panel: JPD estimated from the MI non-constrained model and **R** . **Right panel**: JPD estimated from the MI constrained model and **R** Both these figures are relative to the synthetic test presented in Section 7.2.4

Part III

Summary, Appendix, and Bibliography

Per Aspera sic itur
ad Asthenosphæram

–

8 Summary and conclusions

The main object of this thesis has been the development of a set of tools that, at least numerically, solve the inverse problem relative to application of the MT method in an anisotropic environment. This goal has been achieved using a classic approach, the Levenberg-Marquardt algorithm, co-operated by the penalty function dependent on mutual information introduced in this work.

Prior work documented the effectiveness of MI as distance measure, and MI has been used extensively in inverse problems, especially in medical tomography (cf. for example, *Collignon et al.* [1995] or *Viola and W.M. Wells* [1995] to name but two examples of many tens). Similarly the LM algorithm is a classic tool in the inverse problem solution, and the forward 1D and 2D anisotropic MT problem can be numerically solved using the algorithms developed by *Pek and Santos* [2002] and *Pek and Verner* [1997] respectively.

The very existence of these methods does not though ensure their mutual compatibility. For example, as examined, the genetic algorithm can be an efficient tool in certain environments (i.e. *Moorkamp et al.* [2007]), while it is computationally too expensive if used in a 2D anisotropic domain.

We have partially examined the causes that could make the electrical conductivityparameter compatible with temperature, pressure or bulk composition. Notwithstanding the highlighted relations, we did not attempt to develop a joint inversion scheme because of two principal motivations: first of all a joint inversion algorithm requires the development of a suitable forward solver, a topic for a separate thesis on its own. Secondly the possibility of "switching off" the MI penalty function and using a classic LM algorithm to solve the 1D or 2D anisotropic MT problem allows tests not performed yet in the EM community, possibly highlighting unknown properties of these particular problems.

With some preliminary results ready, it is possible to design tests and experiments relative to different parameters that can be "close" to the electrical conductivity in a MI sense. Probably the 1D anisotropic domain should be preferred under this optic, for computing-cheapness and the possibility of computing the response of a whole range of geophysics problems in a 1D domain.

A remaining open problem is the selection of an appropriate reference model. Again, the existence of a trusted model ensures the minimization of the MI distance from that particular set of parameters. Interesting would be the inclusion of geological profiles in the developed scheme, in order to take advantage of geological informations available about a region in the EM data inversion process.

A further interesting topic of discussion, not treated in this thesis, is the effect of boundary conditions on the two-dimensional MT forward solutions computed. In *Pek and Verner* [1997], 1D anisotropic models are created and solved on both the left and right boundary of the domain. Of course different conditions can be imposed on the domain's boundaries, changing the forward operator and, with it, the solution of the inverse problem.

The natural extension of this work is the extension of the model domain to the third dimension. With an adequate forward solver, the same MI based penalty function can be applied to 3D isotropic or anisotropic EM inverse problems.

The complexity of these domains is naturally transferred to the forward solvers, and the updated 3D solvers make extensive use of parallel clusters, whereas we are unaware of a general 3D anisotropic forward solver. Notwithstanding the considerations relative to the efficiency of a computer program, the study of anisotropic electrical conductivities in the Earth's materials may highlight tectonic structures and geodynamic processes ignored by other geophysical methods. Therefore, electrical anisotropy measurements and recognition requires the development of tools able to handle anisotropic data.

As shown from the case-study relative to the Rhenish Shield region, the constraint introduced is capable of refining a model sharply, allowing both a statistical comparison with the reference model and a strong constraint on the maximum anisotropy rate. The computation of a minimum entropy JPD, that can be interpreted as a statistical relation that maps the retrieved model in the reference one, allows, in fact, the visualization of possible mutual relations between parameters. In the Central Germany dataset, we highlighted the alignment of clusters, in the JPD, relative to the electrical anisotropic strike direction and the seismical azimuthal fast-axis direction, suggesting a trend of constant declination between the two.

The constraint on the maximum anisotropy rate, proves the effectiveness of the developed method in finding a final model compatible with the reference one. This compatibility ensured, in the studied case, a good constraint relative to the resistivity in the resistive direction, a parameter unconstrained by the MT measurements on their own.

In the 2D synthetic tests examined, we observe, similarly, that the introduction of MI distance function is capable to refine the retrieved model and to constrain effectively an unconstrained parameter, namely σ_{zz}.

The computer code developed relative to the 2D problem, in its current form, is able to handle a two-dimensional fully anisotropic domain, discretized in up to twenty blocks of arbitrary shape on a continental flat[1] crust. Some efforts were taken to extend the limit of twenty blocks to a broader diversity of the domain, but errors in the program

[1] not considering the regional topography.

output highlight a mistake in the arrangement of the elements of the Jacobian, computed to implement this "advanced" version of the code.

As an extra effort to make the code more efficient, a computer parallelization strategy was developed, taking advantage of the structure of the Stokes supercomputer, available at the ICHEC facilities. This strategy – presented in Appendix A – was not implemented because the forward solver is presently written making use of the old FORTRAN standard, but its implementation is straightforward as soon as a "modern" implementation of the solver will be available.

A

Parallelization Issues: Efficiency Strikes Back

In the early stages of the development of this thesis, the GA was enthusiastically tested. At that stage, due to the ease of development of a parallel strategy to implement the GA, some efforts were taken to explore the available tools allowing the implementation of a code that makes use of the parallel supercomputer available at the ICHEC.

The rejection of the GA was not a sufficient reason to avoid studying of a strategy directed to the parallelization of the computer code developed relative to the 2D problem[1]. In Figure A.1, a simple scheme of the developed strategy is presented.

The Levenberg-Marquardt method requires the computation of several quantities, in order to set up the system

$$\left[\mathbf{J}(\mathbf{m})^T\mathbf{J}(\mathbf{m}) + \lambda\mathbf{I}\right]\Delta\mathbf{m} = -\mathbf{J}(\mathbf{m})^T\mathbf{F}(\mathbf{m}). \tag{A.1}$$

A.1. Computation of $\mathbf{F}(\mathbf{m})$

As stated in Chapter 2, the solution of the finite-differences system has to be performed once per each considered period (or frequency), thus the most natural parallelization strategy begins sending a forward problem relative to a certain frequency i to the computing node $node_i$. In each computing node there are eight CPUs available sharing the same RAM, and the particular structure of the cluster allows two threads per CPU.

The problem relative to the *i-th* frequency is then solved, in parallel[2], and the $\mathbf{A}$ matrix is stored in the node memory in its eliminated form. The solution of the forward problem $\mathbf{F}(\mathbf{m})$ is now available, distributed among the different nodes used.

[1] the 1D problem can be promptly solved by a serial machine

[2] this parallelization does not comport a substantial save in time computation, because the the Gaussian elimination algorithm used by the solver cannot be efficiently parallelized. For more information relative to the parallelization of the Gaussian elimination algorithm cf. for example *McGinn and Shaw* [2002].

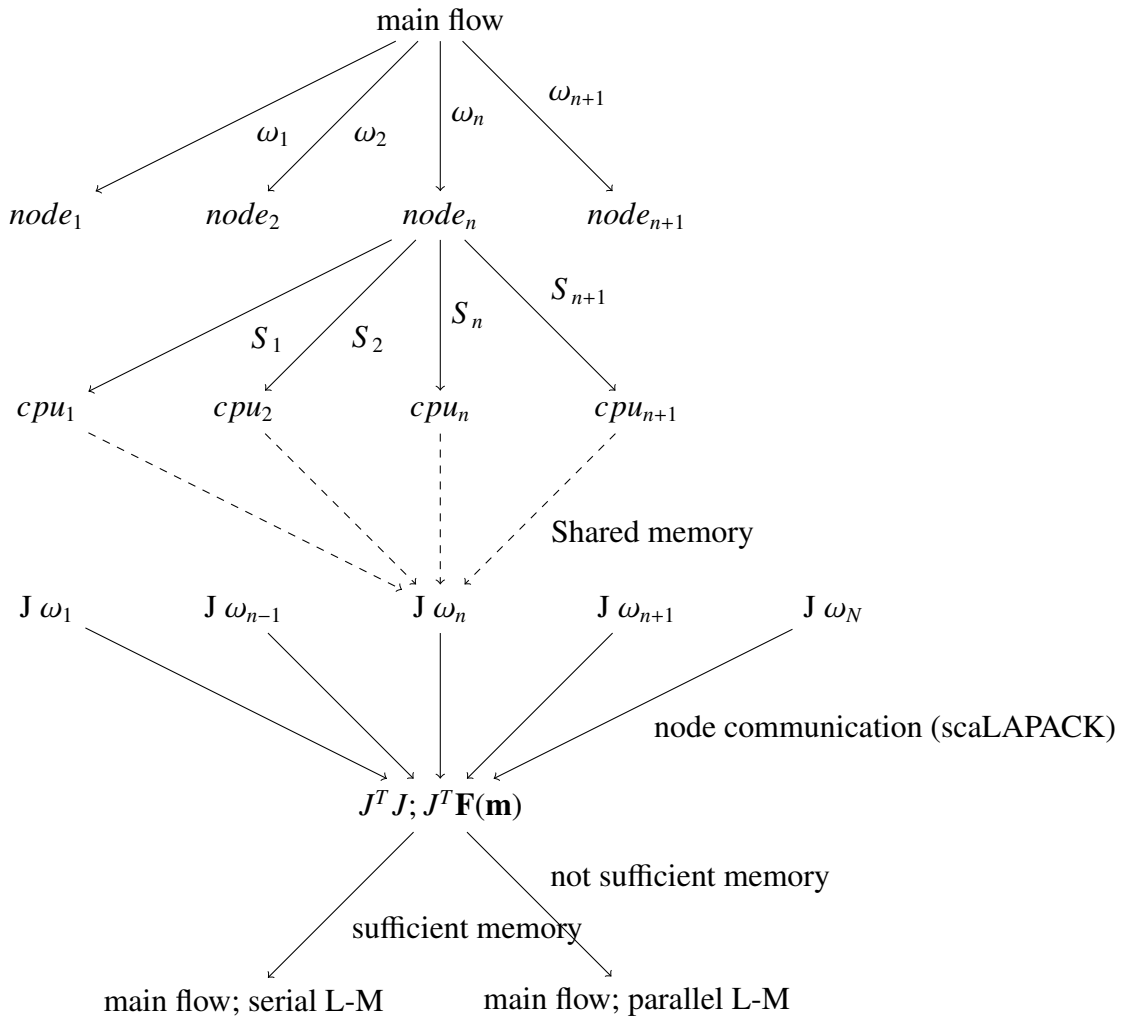


Fig. A.1.: Scheme relative to the computer parallelization strategy developed.

A.2. Computation of $\mathbf{J}(\mathbf{m})$

The Jacobian matrix can be computed by making use of the reciprocity theorem [*Pek et al.*, 2003]. Each one of pseudo-forward problems required to compute $\mathbf{J}(\mathbf{m})$ is characterized by the same $\mathbf{A}$ matrix, that is available in its eliminated form in the node RAM.

The right hand side (RHS) relative to each station $S_\star$ considered is then computed and the relative systems are solved simultaneously by the up-to-sixteen threads using the OPENMP application programming interface. In this way, in the memory of the *i-th* node, the forward solution relative to the *i-th* frequency, and the derivatives of this solution with respect to all the model parameters are stored.

A.3. Levenberg-Marquardt system solution

The remaining step of the LM algorithm involve the computation of $\mathbf{J}(\mathbf{m})^T\mathbf{J}(\mathbf{m})$ and $-\mathbf{J}(\mathbf{m})^T\mathbf{F}(\mathbf{m})$. The data needed are stored in different nodes: this is the first stage in which nodes communication – potentially time-consuming – is required.

The communication between the nodes at this stage is performed via the MPI, and the matrix-vector and matrix-matrix multiplications are handled by the appropriate scaLAPACK library subroutine.

Depending on the system dimensions (i.e., on the number of model parameters) the parallel solution of the system A.1, may be or may be not more efficient than the serial one.

In our experience the serial solution is more efficient, especially because it saves communication time in the line-search procedure undertaken to determine λ. Anyway, if the dimension of $\mathbf{m}$ does not allow the line-search using only one node, it is possible to proceed in the algorithm solving the system in a parallel scheme.

B

Software Used

"*Reddite quae sunt Caesaris, Caesari*". Following this Biblical motto, we wish to give credit to every piece of software we have used more than sparsely. With this we want to thanks those authors that releasing the products of their efforts let us to benefit from their work.

aquamacs: we enjoyed emacs, why stop using it?
http://aquamacs.org

beamer: we do not like this presentation package for its simplicity, but for the simplicity of translate through it a sparse heap of notes into a nice looking presentation.
https://bitbucket.org/rivanvx/beamer/wiki/Home

GMT: despite the theoretical nature of this thesis, we used the Generic Mapping Tool to produce Figure 7.1. Excellent documentation and examples can be found everywhere around the web, but in particular at http://gmt.soest.hawaii.edu.

GNU Emacs: "the" text editor, and much more. It can be extended to perform any action performable by a computer. Available in source or binary at
http://www.gnu.org/software/emacs/

GNU Make: the tool that makes a Makefile essential. https://www.gnu.org/software/make/

GNU Octave: if you are addicted to MATLAB and cannot find your license file, this is the methadone you can have for free from the usual pusher at
http://www.gnu.org/software/octave/

gnuplot: no one of the figures of this thesis were produced using gnuplot, but most of the ones decorating the posters from our work were.
http://www.gnuplot.info/download.html

ifort: the Intel Fortran compiler. Computing time is expensive, and this compiler helps us to save it. http://software.intel.com/en-us/intel-sdp-products

iWork Keynote: if you think that you can build any presentation easily with beamer, you need to have a look on this piece of software. It struggles with formulæ, but for everything else is quick and easy to use. http://www.apple.com/iwork/keynote/

LAPACK: without LAPACK the solution of any of the many linear equations systems solved in this thesis would have taken ages. http://www.netlib.org/lapack/

latex: "LaTeX is a document preparation system for high-quality typesetting. It is most often used for medium-to-large technical or scientific documents but it can be used for almost any form of publishing." And it saves you from using MSW***.
http://www.latex-project.org

MATLAB: yes, eventually I bought a license. Easy to use, easy to install, highly addicting. http://www.mathworks.co.uk/products/matlab/

Maxima: a computer algebra system. It make our algebraic chewing an enjoyable experience. http://maxima.sourceforge.net

Minpack: we used this numerical library to test the numerical solutions of out own implementation of the Levenberg-Marquardt algorithm.
Available at http://www.math.utah.edu/software/minpack.html

wxMaxima: we enjoyed the Maxima symbolic computation capabilities, but we are not freak enough to use it without a GUI. It really helps, and is free at
http://andrejv.github.com/wxmaxima/

Bibliography

Aster, R. C., B. Borchers, and C. Thurber (2004), *Parameter Estimation and Inverse Problems*, Academic Press.

Backus, G., and F. Gilbert (1968), The Resolving Power of Gross Earth Data, *Geophysical Journal International*, *16*, 169–205, doi:10.1111/j.1365-246X.1968.tb00216.x.

Bailey, R. C. (1970), Inversion of the Geomagnetic Induction Problem, *Royal Society of London Proceedings Series A*, *315*, 185–194, doi:10.1098/rspa.1970.0036.

Becken, M., and H. Burkhardt (2004), An ellipticity criterion in magnetotelluric tensor analysis, *Geophysical Journal International*, *159*(1), 69–82.

Bischoff, M. (2006), Lower crustal anisotropy in central europe deducedfrom dispersion analysis of love and rayleigh waves, *Geophysical Research Abstracts*, *8*.

Cagniard, L. (1953), Basic theory of the magneto-telluric method of geophysical prospecting, *Geophysics*, *18*(3), 605–635, doi:10.1190/1.1437915.

Chave, A., and A. Jones (2012), *The Magnetotelluric Method: Theory and Practice*, Cambridge University Press.

Chave, A. D., and A. G. Jones (1997), Electric and magnetic field galvanic distortion decomposition of bc87 data, *Journal of Geomagnetism and Geoelectricity*, *49*(6), 767–789.

Chave, A. D., and D. J. Thomson (1989), Some comments on magnetotelluric response function estimation, *Journal of Geophysical Research: Solid Earth*, *94*(B10), 14,215–14,225, doi:10.1029/JB094iB10p14215.

Chave, A. D., and D. J. Thomson (2003), A bounded influence regression estimator based on the statistics of the hat matrix, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *52*(3), 307–322, doi:10.1111/1467-9876.00406.

Collignon, A., F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal (1995), Automated multi-modality image registration based on information theory, *Information Processing in Medical Imaging*, pp. 263–274.

Constable, S., and J. J. Roberts (1997), Simultaneous modeling of thermopower and electrical conduction in olivine, *Physics and Chemistry of Minerals*, *24*, 319–325, doi: 10.1007/s002690050044.

Constable, S. C., R. L. Parker, and G. C. Constable (1987), Occam's inversion: A practical algorithm for generating smooth models from electromagnetic sounding data, *Geophysics*, *52*, 289, doi:10.1190/1.1442303.

Courant, R., and F. John (1989), *Introduction to Calculus and Analysis*, *Classics in Mathematics*, vol. 1, Springer.

Duba, A., and I. Nicholls (1973), The influence of oxidation state on the electrical conductivity of olivine, *Earth and Planetary Science Letters*, *18*(1), 59 – 64, doi: DOI:10.1016/0012-821X(73)90034-4.

Duba, A., and J. von der Gönna (1994), Comment on change of electrical conductivity of olivine associated with the olivine-spinel transition, *Physics of The Earth and Planetary Interiors*, *82*(1), 75 – 77, doi:DOI:10.1016/0031-9201(94)90104-X.

Eaton, D. W., A. G. Jones, and I. J. Ferguson (2004), Lithospheric anisotropy structure inferred from collocated teleseismic and magnetotelluric observations: Great Slave Lake shear zone, northern Canada, *Geophys. Res. Lett.*, *31*, L19614, doi: 10.1029/2004GL020939.

Egbert, G. D., and J. R. Booker (1986), Robust estimation of geomagnetic transfer functions, *Geophysical Journal*, *87*, 173–194.

Engl, H. W., M., and A. Neubauer (1996), *Regularization of Inverse Problems*, Kluwer Academic Publishers.

Everett, M. E., and A. Schultz (1993), Two-dimensional nonlinear magnetotelluric inversion using a genetic algorithm, *Journal of Geomagnetism and Geoelectricity*, *45*(9), 1013–1026.

Gallardo, L. A., and M. A. Meju (2003), Characterization of heterogeneous near-surface materials by joint 2D inversion of dc resistivity and seismic data, *Geophysical Research Letters*, *30*(13), 1658, doi:10.1029/2003GL017370.

Gamble, T., W. Goubau, and J. Clarke (1979a), Magnetotellurics with a remote magnetic reference, *GEOPHYSICS*, *44*(1), 53–68, doi:10.1190/1.1440923.

Gamble, T., W. Goubau, and J. Clarke (1979b), Error analysis for remote reference magnetotellurics, *GEOPHYSICS*, *44*(5), 959–968, doi:10.1190/1.1440988.

Gatzemeier, A., and M. Moorkamp (2005), 3d modelling of electrical anisotropy from electromagnetic array data: hypothesis testing for different upper mantle conduction mechanisms, *Physics of the Earth and Planetary Interiors*, *149*(3-4), 225 – 242, doi: 10.1016/j.pepi.2004.10.004.

Gatzemeier, A., and A. Tommasi (2006), Flow and electrical anisotropy in the upper mantle: Finite-element models constraints on the effects of olivine crystal preferred orientation and microstructure, *Physics of the Earth and Planetary Interiors*, *158*(2-4), 92 – 106, doi:10.1016/j.pepi.2006.01.009.

Goldberg, D. E. (1989), *Genetic Algorithms in Search, Optimization, and Machine Learning*, 1 ed., Addison-Wesley Professional.

Grandis, H., M. Menvielle, and M. Roussignol (1999), Bayesian inversion with Markov chains-I. The magnetotelluricone-dimensional case, *Geophysical Journal International*, *138*, 757–768, doi:10.1046/j.1365-246x.1999.00904.x.

Gripp, A. E. (1990), Current plate velocities relative to the hotspots incorporating the nuvel-1 global plate motion model, *Geophys. Res. Lett.*, *17*, 1109–1112.

Groom, R. W., and K. Bahr (1992), Corrections for near surface effects: Decomposition of the magnetotelluric impedance tensor and scaling corrections for regional resistivities: A tutorial, *Surveys in Geophysics*, *13*, 341–379, doi:10.1007/BF01903483.

Groom, R. W., and R. C. Bailey (1989), Decomposition of magnetotelluric impedance tensors in the presence of local three-dimensional galvanic distortion, *J. Geophys. Res*, *94*, 1913–1925, doi:10.1029/JB094iB02p01913.

Groom, R. W., R. D. Kurtz, A. G. Jones, and D. E. Boerner (1993), A quantitative methodology to extract regional magnetotelluric impedances and determine the dimension of the conductivity structure, *Geophysical Journal International*, *115*, 1095–1118, doi: 10.1111/j.1365-246X.1993.tb01512.x.

Haber, E. (2004), Multilevel optimization methods for mutual information registrations, personal communication.

Haber, E., and M. H. Gazit (2012), Model fusion and joint inversion, submitted paper.

Haber, E., and D. Oldenburg (1997), Joint inversion: a structural approach, *Inverse Problems*, *13*, 63–77, doi:10.1088/0266-5611/13/1/006.

Hadamard, J. (1902), Sur les problèmes aux dérivés partielles et leur signification physique, *Princeton University Bulletin*, *13*, 49–52.

Hamilton, M. P., A. G. Jones, R. L. Evans, S. Evans, C. J. S. Fourie, X. Garcia, A. Mountford, J. E. Spratt, and SAMTEX MT Team (2006), Electrical anisotropy of South African lithosphere compared with seismic anisotropy from shear-wave splitting analyses, *Physics of the Earth and Planetary Interiors*, *158*, 226–239, doi: 10.1016/j.pepi.2006.03.027.

Hansen, P. C., and D. P. O'Leary (1993), The use of the l-curve in the regularization of discrete ill-posed problems, *SIAM Journal on Scientific Computing*, *14*(6), 1487–1503.

Harris, F. (1978), On the use of windows for harmonic analysis with the discrete fourier transform, *Proceedings of the IEEE*, *66*(1), 51–83, doi:10.1109/PROC.1978.10837.

Jiracek, G. R. (1990), Near-surface and topographic distortions in electromagnetic induction, *Surveys in Geophysics*, *11*, 163–203, doi:10.1007/BF01901659.

Jones, A. (1999), Imaging the continental upper mantle using electromagnetic methods, *Lithos*, *48*, 57–80, doi:10.1016/S0024-4937(99)00022-5.

Jones, A. G. (2007), Geof759 rfp2007 proposal.

Jones, A. G. (2012), Distortion decomposition of the magnetotelluric impedance tensors from a one-dimensional anisotropic Earth, *Geophysical Journal International*, *189*, 268–284, doi:10.1111/j.1365-246X.2012.05362.x.

Jones, A. G., and R. Hutton (1979), A multi-station magnetotelluric study in southern Scotland ? II. Monte-Carlo inversion of the data and its geophysical and tectonic implications, *Geophysical Journal International*, *56*, 351–368, doi:10.1111/j.1365-246X.1979.tb00169.x.

Jones, A. G., and H. G. Jödicke (1984), Magnetotelluric transfer function estimation improvement by a coherence-based rejection technique, in *1984 SEG Annual Meeting*.

Jones, A. G., and J. Spratt (2002), A simple method for deriving the uniform field MT responses in auroral zones, *Earth, Planets, and Space*, *54*, 443–450.

Karato, S., and L. Dai (2009), Comments on "electrical conductivity of wadsleyite as a function of temperature and water content" by manthilake et al., *Physics of the Earth and Planetary Interiors*, *In Press, Corrected Proof*, –, doi:DOI:10.1016/j.pepi.2009.01.011.

Kisabeth, J. L., and G. Rostoker (1977), Modelling of three-dimensional current systems associated with magnetospheric substorms, *Geophysical Journal of the Royal Astronomical Society*, *49*(3), 655–683, doi:10.1111/j.1365-246X.1977.tb01310.x.

Kuvshinov, A., H. Utada, D. Avdeev, and T. Koyama (2005), 3-D modelling and analysis of Dst C-responses in the North Pacific Ocean region, revisited, *Geophysical Journal International*, *160*, 505–526, doi:10.1111/j.1365-246X.2005.02477.x.

Leibecker, J., A. Gatzemeier, M. Honig, O. Kuras, and W. Soyer (2002), Evidence of electrical anisotropic structures in the lower crust and the upper mantle beneath the rhenish shield, *Earth and Planetary Science Letters*, *202*(2), 289–302, doi:doi:10.1016/S0012-821X(02)00525-3.

Levenberg, K. (1944), A method for the solution of certain problems in least squares, *Quarterly of Applied Mathematics*, *2*, 164–168.

Li, Y. (2002), A finite-element algorithm for electromagnetic induction in two-dimensional anisotropic conductivity structures, *Geophysical Journal International*, *148*, 389–401, doi:10.1046/j.1365-246X.2002.01570.x.

Mandolesi, E., A. G. Jones, E. Roux, and S. Lebedev (2009a), Common Structure in Different Physical Properties: Electrical Conductivity and Surface Waves Phase Velocity, *AGU Fall Meeting*.

Mandolesi, E., A. G. Jones, and E. Roux (2009b), Inversion of magnetotelluric data with seismic constraints using a structural approach, *BGA Postgraduate Research in Progress Meeting*.

Mareschal, M. (1986), Modelling of natural sources of magnetospheric origin in the interpretation of regional induction studies: A review, *Surveys in Geophysics*, *8*, 261–300, doi:10.1007/BF01904062.

Marquardt, D. W. (1963), An Algorithm for Least-Squares Estimation of Nonlinear Parameters, *SIAM Journal on Applied Mathematics*, *11*(2), 431–441.

McGinn, S., and R. Shaw (2002), Parallel gaussian elimination using openmp and mpi, *High Performance Computing Systems and Applications, Annual International Symposium on*, *0*, 169, doi:http://doi.ieeecomputersociety.org/10.1109/HPCSA.2002.1019151.

Mcneice, G. W., and A. G. Jones (2001), Multi-site, multi-frequency tensor decomposition of magnetotelluric data, *Geophysics*, *66*, 158–173.

Moore, E. H. (1920), On the reciprocal of the general algebraic matrix, *Bulletin of the American Mathematical Society*, *26*, 394–395.

Moorkamp, M., A. G. Jones, and D. W. Eaton (2007), Joint inversion of teleseismic receiver functions and magnetotelluric data using a genetic algorithm: Are seismic velocities and electrical conductivities compatible?, *Geophysical Research Letters*, *341*, doi:10.1029/2007GL030519.

Neal, S. L., R. L. Mackie, J. C. Larsen, and A. Schultz (2000), Variations in the electrical conductivity of the upper mantle beneath North America and the Pacific Ocean, *J. Geophys. Res*, *105*, 8229–8242, doi:10.1029/1999JB900447.

Nover, G. (2005), Electrical properties of crustal and mantle rocks a review of laboratory measurements and their explanation, *Surveys in Geophysics*, *26*(5), 593–651, doi:http://dx.doi.org/10.1007/s10712-005-1759-6.

Nover, G., Stoll, Johannes B., and von der Gonna, Jutta (2005), Promotion of graphite formation by tectonic stress - a laboratory experiment, *Geophysical Journal International*, *160*, 1059–1067(9).

Nuttall, A. H. (1981), Some windows with very good sidelobe behavior, *Acoustics, Speech and Signal Processing, IEEE Transactions on*, *29*(1), 84–91, doi:10.1109/TASSP.1981.1163506.

Olsen, N. (1998), The electrical conductivity of the mantle beneath Europe derived from C-responses from 3 to 720 hr, *Geophysical Journal International*, *133*, 298–308, doi:10.1046/j.1365-246X.1998.00503.x.

Onsager, L. (1931), Reciprocal relations in irreversible processes. i., *Phys. Rev.*, *37*, 405–426, doi:10.1103/PhysRev.37.405.

Padilha, A. L., ´Í. Vitorello, M. B. Pádua, and M. S. Bologna (2006), Lithospheric and sublithospheric anisotropy beneath central-southeastern Brazil constrained by long period magnetotelluric data, *Physics of the Earth and Planetary Interiors*, *158*, 190–209, doi:10.1016/j.pepi.2006.05.006.

Parker, R. L. (1980), The inverse problem of electromagnetic induction - Existence and construction of solutions based on incomplete data, *Journal of Geophysical Research*, *85*, 4421–4428, doi:10.1029/JB085iB08p04421.

Parker, R. L. (1983), The magnetotelluric inverse problem, *Geophysical Surveys*, *6*, 5–25, doi:10.1007/BF01453993.

Pek, J., and F. A. M. Santos (2002), Magnetotelluric impedances and parametric sensitivities for 1-D anisotropic layered media, *Computers & Geosciences*, *28*, 939–950.

Pek, J., and T. Verner (1997), Finite-difference modelling of magnetotelluric fields in two-dimensional anisotropic media, *Geophysical Journal International*, *128*(3), 505–521, doi:10.1111/j.1365-246X.1997.tb05314.x.

Pek, J., A. M. Santos, and Y. Li (2003), Parametric Sensitivities for 2-D Anisotropic Magnetotelluric Models, *20. Kolloquium Elektromagnetische Tiefenforschung, KÃ¶nigstein, 29.09.-3.10.2003, Hrsg.: A. HÃ¶rdt und J. B. Stoll*, pp. 240–249.

Penrose, R. (1955), A generalized inverse for matrices, *Proceedings of the Cambridge Philosophical Society*, *51*, 406–413, doi:10.1017/S0305004100030401.

Pérez-Flores, M. A., and A. Schultz (2002), Application of 2-D inversion with genetic algorithms to magnetotelluric data from geothermal areas, *Earth, Planets, and Space*, *54*, 607–616.

Price, A. T. (1962), The Theory of Magnetotelluric Methods When the Source Field Is Considered, *Journal of Geophysical Research*, *67*, 1907–1918, doi:10.1029/JZ067i005p01907.

Price, A. T. (1973), The theory of geomagnetic induction, *Physics of the Earth and Planetary Interiors*, *7*(3), 227–233.

Reddy, I. K., and D. Rankin (1975), Magnetotelluric response of laterally inhomogeneous and anisotropic media, *Geophysics*, *40*(6), 1035–1045, doi:10.1190/1.1440579.

Rikitake, T. (1950), Electromagnetic induction within the earth and its relation to the electrical state of the earth's interior. part i (1), *Bull. Earthq. Res. Inst.*, *28*, 45–100.

Rikitake, T. (1951a), Electromagnetic induction within the earth and its relation to the electrical state of the earth's interior. part i (2), *Bull. Earthq. Res. Inst.*, *28*, 219–262.

Rikitake, T. (1951b), Electromagnetic induction within the earth and its relation to the electrical state of the earth's interior. part ii, *Bull. Earthq. Res. Inst.*, *28*, 263–283.

Rikitake, T. (1951c), Electromagnetic induction within the earth and its relation to the electrical state of the earth's interior. part iii, *Bull. Earthq. Res. Inst.*, *29*, 61–69.

Roberts, J. J. (2002), Electrical properties of microporous rock as a function of saturation and temperature, *Journal of Applied Physics*, *91*(3), 1687–1694, doi:10.1063/1.1430544.

Roberts, Jeffery J., T. (1999), Partial-melt electrical conductivity: influence of melt composition, *Journal of Geophysical Research*, *104*(B4), 7055–7065.

Rostoker, G., M. Mareschal, and J. C. Samson (1982), Response of dayside net downward field-aligned current to changes in the interplanetary magnetic field and to substorm perturbations, *Journal of Geophysical Research: Space Physics*, *87*(A5), 3489–3510, doi:10.1029/JA087iA05p03489.

Roux, E., M. Moorkamp, A. G. Jones, M. Bischoff, B. Endrun, S. Lebedev, and T. Meier (2011), Joint inversion of long-period magnetotelluric data and surface-wave dispersion curves for anisotropic structure: Application to data from central germany, *Geophysical Research Letters*, *38*(5), L05,304.

Savage, M. K. (1999), Seismic anisotropy and mantle deformation: what have we learned from shear wave splitting?, *Reviews of Geophysics*, *37*(1), 65–106, doi: 10.1029/98RG02075.

Schock, R. N., A. G. Duba, and T. J. Shankland (1989), Electrical conduction in olivine, *J. Geophys. Res*, *94*, 5829–5839, doi:10.1029/JB094iB05p05829.

Shankland, T. J. (1979), Physical Properties of Minerals and Melts, *Reviews of Geophysics*, *17*, 792–802, doi:10.1029/RG017i004p00792.

Shankland, T. J., and H. S. Waff (1974), Conductivity in Fluid-Bearing Rocks, *J. Geophys. Res*, *79*, 4863–4868, doi:10.1029/JB079i032p04863.

Shankland, T. J., and H. S. Waff (1977), Partlial melting and electrical conductivity anomalies in the upper mantle, *J. Geophys. Res*, *82*(33).

Shannon, C. E. (1948a,b), A Mathematical Theory of Communication, *The Bell System Technical Journal*, *27*, 379–423, 623–635.

Sharma, S. P., and P. Kaikkonen (1998), Two-dimensional non-linear inversion of VLF-R data using simulated annealing, *Geophysical Journal International*, *133*, 649–668, doi:10.1046/j.1365-246X.1998.00523.x.

Shaw, R., and S. Srivastava (2007), Particle swarm optimization: A new tool to invert geophysical data, *Geophysics*, *72*, F75, doi:10.1190/1.2432481.

Shi, X. M., J. Y. Wang, S. Y. Zhang, and X. Y. Hu (2000), Multiscale genetic algorithm and its application in magnetotelluric sounding data inversion, *Chinese Journal of GeophysicsChinese Edition*, *43*(1), 122–130.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.

Simpson, F. (2001), Resistance to mantle flow inferred from the electromagnetic strike of the Australian upper mantle, *Nature*, *412*, 632–635.

Simpson, F., and K. Bahr (2005), *Practical magnetotellurics*, Cambridge University Press.

Sims, W., F. Bostick, and H. Smith (1971), The estimation of magnetotelluric impedance tensor elements from measured data, *GEOPHYSICS*, *36*(5), 938–942, doi:10.1190/1.1440225.

Takashi Yoshino, T. M., Geeth Manthilake, and T. Katsura (2008), Dry mantle transition zone inferred from the conductivity of wadsleyite and ringwoodite, *Nature*, *451*, 326–329.

Tarantola, A. (2005), *Inverse Problem Theory and Methods for Model Parameter Estimation*, vol. 1, 352 pp., SIAM.

Tarits, P., S. Hautot, and F. Perrier (2004), Water in the mantle: Results from electrical conductivity beneath the French Alps, *Geophys. Res. Lett.*, *31*, 6612–+, doi:10.1029/2003GL019277.

Tikhonov, A. N. (1950), The determination of the electrical properties of the deep layers of the Earth's crust., *Dokl. Acad. Nauk. USSR*, *73*, 295–297.

Tikhonov, A. N., and A. V. Goncharsky (1987), *Ill-posed problems in the natural sciences*, Advances in science and technology in the USSR. Mathematics and mechanics, MIR Publishers.

Unser, M., and P. Thevenaz (2000), Optimization of mutual information for multiresolution image registration, *IEEE Transactions on Image Processing*, *9*, 2083–2099, doi:10.1109/83.887976.

Viljanen, A., R. Pirjola, and O. Amm (1999), Magnetotelluric source effect due to 3D ionospheric current system using the complex image method for 1D conductivity structures, *Earth Planets Space*, *51*, 933–945.

Viola, P., and I. W.M. Wells (1995), Alignment by maximization of mutual information, *Computer Vision, IEEE International Conference on*, *0*, 16, doi:http://doi.ieeecomputersociety.org/10.1109/ICCV.1995.466930.

Vozoff, K. (1991), The magnetotelluric method, *Electromagnetic Method in Applied Geophysics, Society of Exploration Geophysicists*, pp. 641–712.

Wait, J. R. (1954), On the Relation Between Telluric Currents and the Earth's Magnetic Field, *Geophysics*, *19*, 281, doi:10.1190/1.1437994.

Wait, J. R. (1981), *Wave Propagation Theory*, pp. 52–53, Pergamon Press.

Weidelt, P. (1972), The inverse problem of geomagnetic induction, *J. Geophys.*, *38*, 257–289.

Wolfe, P. (1959), The simplex method for quadratic programming, *Econometrica*, *27*, 382–398.

Xiaoge Huang, Y. X., and S. Karato (2005), Water content in the transition zone from electrical conductivity of wadsleyite and ringwoodite, *Nature*, *434*, 746–749.

Xu, Y., B. T. Poe, T. J. Shankland, and D. C. Rubie (1998), Electrical Conductivity of Olivine, Wadsleyite, and Ringwoodite Under Upper-Mantle Conditions, *Science*, *280*(5368), 1415–1418, doi:10.1126/science.280.5368.1415.

Yin, C., and P. Weidelt (1999), Geoelectrical fields in a layered earth with arbitrary anisotropy, *Geophysics*, *64*, 426, doi:10.1190/1.1444547.

Yoshino, T., and T. Katsura (2009), Reply to comments on "electrical conductivity of wadsleyite as a function of temperature and water content" by manthilake et al., *Physics of the Earth and Planetary Interiors*, *In Press, Corrected Proof*, –, doi:DOI:10.1016/j.pepi.2009.01.012.

Zanetti, L. J., T. A. Potemra, and W. Baumjohann (1983), Ionospheric and Birkeland current distributions inferred from the MAGSAT magnetometer data, *Journal of Geophysical Research*, *88*, 4875–4884, doi:10.1029/JA088iA06p04875.