KU LEUVEN

FACULTY OF SCIENCE

Wavelet-based inversion of electromagnetic induction surveys

Wouter DELEERSNYDER

Thesis presented in
fulfillment of the requirements
for the degree of Master of Science
in Physics

Supervisor: Prof. Dr. D. Dudal
Mentor: B. Maveau

Academic year 2018-2019

ACKNOWLEDGEMENTS

I use the first words of this thesis to thank a few people. First I want to thank prof. dr. David Dudal for the guidance, the suggestions and comments that have contributed to this thesis, while maintaining a healthy degree of independence. My mentor Benjamin Maveau helped me with all that I needed, especially in my first steps with (object-oriented) programming in Python. He was always willing to assist me with the implementations. Together with Benjamin and Marieke Paepen (UGent, PhD candidate in prof. dr. T. Herman's team), we carried out a measurement in De Panne, Westhoek. Thank you for also highlighting this side of the research. Through these contacts with the Geology department, I have realized the societal relevance of my research.

I would also like to thank everyone in the physics department at KULAK. It was pleasant to work near the lab's coffee corner. These conversations led to suggestions for my thesis that were worth analysing.

This thesis certainly contains influences of various professors at the KU Leuven who have inspired me in those five years. I am especially grateful to those who made me recognise the beauty of a good model.

I also thank my parents and my family for the continuous stream of support and encouragement. With equally great importance are the beautiful times that I have had with my friends, which I will remember now and later. Together with the necessary distraction, we could forget about the thesis tension. Last but not least, I thank Chelsea for the love and because you always listened with great interest to what I was doing in my thesis.

SUMMARY

In geophysics, electromagnetic induction (EMI) methods are commonly used to explore the near subsurface non-invasively (i.e. without digging). That subsurface consists of earth layers with different electrical conductivities. A geologist can interpret these conductivities via Archie's empirical law, which maps the soil characteristics (e.g. porosity, water content, the electrical conductivity of the brine) to the electrical conductivity. The goal of this thesis is to study the state-of-the-art models of the EMI problem and to develop an inversion scheme that allows determining the electrical conductivities from EMI measurements. The focus is on EMI surveys in a coastal regions, where saltwater intrusion can be monitored (salt water is more conductive than fresh water).

A typical electromagnetic induction survey is performed in the frequency domain, where an alternating magnetic dipole with constant frequency (the primary magnetic field) is placed above the soil. The dipole induces eddy currents in the earth that in return generate the measured secondary magnetic field. The response is inverted such that the conductivity profile of the subsurface is obtained. This inverse problem requires a reliable forward model that quickly computes the response of any conductivity profile. Furthermore, it requires a fast optimization technique in order to minimize the data misfit between the response of an initial guess of the conductivity profile (calculated via the forward model) and the response from EMI measurements.

In this thesis, three forward models are discussed: an exact model, the LIN approximation and the damped model. The first model is by [Wait, 1951] and solves the Maxwell equations in the quasi-stationary field regime (where displacement currents can be neglected for sufficiently low frequencies). The conductivity profile consists of horizontally stratified earth layers (and thus no lateral variation); this is a justified assumption in coastal regions, where sedimentary layers are approximately horizontally stratified. The model takes into account both the electromagnetic dampening in conductive media (skin effect) and the electromagnetic couplings between eddy currents. The relations include integration over Bessel functions and a recursive formula; therefore it has a large computational burden. Since in our optimization method, the response needs to be calculated at every iteration, other simplified forward models are studied. The first approximation [Wait, 1962] is a linear model that slices the subsurface in infinitesimal layers and sums the contributions to the secondary magnetic field from each independent sheet. The model does not take into account any couplings between the layers or any the electromagnetic dampening. This model is known as the LIN approximation and is only valid under the Low Induction Numbers (LIN) condition. This LIN assumption breaks down in conductive environments, such as coastal regions; therefore a recently developed model is examined. The damped model by [Maveau et al., 2017] is similar to the LIN approximation but embeds the independent sheets in a conductive background. This effectively models the electromagnetic dampening of the electromagnetic fields. It slightly complexifies the model; depending on the choice of the conductive background, the model is linear or non-linear. Both the LIN and damped model allow to calculate the response rather quickly, at the cost of unmodelled effects. The damped model approximates the exact model more accurately.

The inverse problem is solved by the minimization of an objective function. For a properly defined objective function, that minimum corresponds to the true conductivity profile that describes the data. Starting from an initial guess of the conductivity profile, the minimum of the objective function is obtained iteratively via the L-BFGS-B algorithm [Byrd et al., 1995], a quasi-Newton method. The objective function has two terms: the data misfit and the model misfit. The data misfit measures the discrepancy between the measured response and the response generated via the forward model and some conductivity profile. The ill-posedness of the inverse problem (ill-conditioned and non-unique), however, requires to add a model misfit term that regularizes (or stabilizes) the problem. The model misfit term in our inversion scheme is based on Occam's principle, which states that simpler solutions are more likely than complex ones. In the case of horizontally stratified earth structures, this likely solution corresponds to blocky structures. Consider a basis transformation in which likely solutions can be represented in a sparse form. Such a basis transform in combination with a sparsity promoting measure would impose that the solution is simple or blocky. Such a basis transformation is available under the discrete wavelet transform. The ℓ_1-norm is the traditional sparsity promoting measure. However, the optimization algorithm uses gradient information and the ℓ_1-norm is non-differentiable at zero. Therefore, other measures must be used (which are differentiable), such as the perturbed ℓ_p-norm measure of Ekblom [Ekblom, 1987].

A regularization parameter balances the importance of the model misfit w.r.t. the data misfit. A large regularization parameter filters out much small-scale effects. Selecting an optimal value is challenging. Common techniques such as the discrepancy principle and the L-curve criterion are examined. The combination of both principles can result in a good selection of the regularization parameter. Other regularization strategies were also developed and tested in which the regularization parameter is no longer held constant.

Simple, two or three-layered conductivity profiles have a sparse representation with Haar wavelets (block and step functions with compact support). For complex conductivity profiles (e.g. with soft boundaries), we must resort to other wavelets, such a as Daubechies wavelets with different properties (e.g. with more regularity). Choosing the right wavelet is an essential part of the inversion scheme and allows to apply the scheme on a wide variety of conductivity profiles.

We have developed an effective inversion scheme that significantly alleviates the ill-posedness of the inverse problem. The scheme can be tailored to the context of the survey and the needs of the geologist.

Summary for the General Audience

De waterspiegel in Vlaanderen staat onder druk. Recente droogtes leiden tot dalende waterspiegels. Dit kan tot verzilting in de kuststreek leiden, door zout zeewater dat in de ondergrondse lagen dringt. In De Panne wordt dit versterkt door de aanwezigheid van een inham in de duinen, die zeewater tot enkele keren per jaar tot in de duinen laat vloeien. Dit bevordert de zoutwaterfauna en -flora, maar bedreigt anderzijds het evenwicht tussen zoet- en zoutwater. Dat zoetwaterreservoir in de duinen wordt bovendien ook gebruikt voor onze drinkwatervoorziening. Er mag dus niet te veel worden opgepompt.

Er bestaat een geofysische methode om dit evenwicht te monitoren, gebaseerd op het principe van elektromagnetische inductie. Een wisselend magnetisch veld wordt boven de aarde geplaatst. Dit zal verschillende wervelstroompjes in de aardlagen opwekken, waarvan de respons terug wordt opgemeten. Dit is vergelijkbaar met de werking van een metaaldetector.

Via deze opgemeten respons trachten we kenmerkende parameters van de aardlagen te achterhalen. Als we daarin slagen, kennen we de elektrische geleidbaarheid van elke aardlaag op elke diepte. Deze combinatie van parameters noemen we een conductiviteitsprofiel. Een sterke geleidbaarheid van een aardlaag komt dan typisch overeen met een aardlaag waar veel zoutwater in aanwezig is.

Een eerste uitdaging is om een goed model te hebben die de respons van een conductiviteitsprofiel berekent. Met andere woorden, stel dat je de ondergrondse parameters kent, dan kan je met zo'n model berekenen wat je zou opmeten. Dit noemen we een voorwaarts model. Een exact voorwaarts model bestaat al sinds 1951, maar rekent op de computer heel traag. Daarom wordt in de thesis onderzocht welke vereenvoudigingen van het model voldoende accuraat zijn, maar sneller kunnen worden berekend. Dit is vooral belangrijk voor het omgekeerde probleem, waar we uit de opgemeten respons de ondergrondse parameters willen bepalen. Dit wordt het invers probleem genoemd. In dat probleem zullen we een gok doen naar het ondergrondse conductiviteitsprofiel. Daarna vergelijken we of de respons, berekend via een voorwaarts model, overeenkomt met de opgemeten respons. Bij een sterke overeenkomst komt onze gok heel waarschijnlijk overeen met de echte aardstructuur. Vermits we in de praktijk veel gokken zullen moeten doen, is een snelle computationele berekening zeker noodzakelijk.

Een tweede uitdaging is om dit invers probleem op te lossen. Er zijn namelijk verschillende conductiviteitsprofielen die ongeveer dezelfde respons genereren. Dit probleem wordt gestabiliseerd door een extra voorwaarde op te leggen voor het profiel. Deze voorwaarde komt erop neer dat we de eenvoudige profielen verkiezen boven ingewikkelde. Als een opgegeven respons zowel correspondeert met een tweelagige structuur als een veertiglagige structuur, zal de tweelagige structuur fysisch waarschijnlijker zijn. Wavelettheorie is een wiskundige theorie die toelaat om de complexiteit van zo'n profiel te kwantificeren. Dit gebruiken we om betere gokken van het conductiviteitsprofiel te kunnen doen en om een realistischere uitkomst te bekomen.

Deze masterthesis introduceert een geavanceerdere procedure om vanuit een meting de ondergrondse aardstructuur te achterhalen, zonder dat hiervoor dure boorgaten moeten gegraven worden. Deze ontwikkelingen dragen bij tot een betere monitoring van hedendaagse uitdagingen, zoals zoutwaterintrusie in de kuststreek.

List of Abbreviations

EM	electromagnetic
EMI	electromagnetic induction
DOI	depth of investigation
HCP	horizontal coplanar coil set-up
VCP	vertical coplanar coil set-up
VCA	vertical co-axial coil set-up
PERP	perpendicular coil set-up
LIN	low induction numbers
TDEM	time domain electromagnetic
FDEM	frequency domain electromagnetic
DWT	discrete wavelet transform
FWT	fast wavelet transform
FFT	fast Fourier transform
MRA	multiresolution analysis
PR	perfect reconstruction
L-BFGS-B-method	limited-memory Broyden–Fletcher–Goldfarb–Shanno method for bound constraints

List of Symbols

Table 1: List of physical quantities and units.

SYMBOL	UNIT	MEANING
σ	[Siemens/meter]	electrical conductivity
σ_a	[Siemens/meter]	apparent conductivity
s	[meter]	intercoil distance
h_0	[meter]	height above the surface of the earth
h	[meter]	depth, measured from the surface of the soil
$\mathbf{E}$	[Volt/meter]	electric field strength
$\mathbf{H}$	[Ampere/meter]	magnetic field strength
$\mathbf{H}_p$	[Ampere/meter]	primary magnetic field
$\mathbf{H}_s$	[Ampere/meter]	secondary magnetic field
$\mathbf{A}$	[Volt×seconds/meter]	magnetic vector potential
$\mathbf{J}_d$	[Ampere/(meter)^2]	displacement current density
ω	[radians/second]	angular frequency
B	[1]	low induction number
$\boldsymbol{\eta}$	[Ampere/meter]	measurement error
$\boldsymbol{\kappa}$	[Ampere/meter]	total error (measurement error + unmodelled effects)

Table 2: List of the most important symbols.

SYMBOL	MEANING
$\mathbf{m}$	vector of model parameters, the parametrization of the conductivity profile
$\mathbf{x}$	vector of model parameters in wavelet domain
$\mathbf{d}$	data vector with magnetic field ratio
K	functional that maps the conductivity profile on the magnetic field ratio
ϕ	objective function
ϕ_d	data misfit functional
ϕ_m	model misfit functional
$\mu(\mathbf{x})$	a measure (assigns non-negative real numbers to subsets)
λ	regularization parameter
W	wavelet transformation in matrix expression
φ	scaling function
v_{nk}	scaling coefficient at level of resolution n and translation k
ψ	wavelet function
w_{nk}	wavelet coefficient at level of resolution n and translation k
N	level of the discrete wavelet transform
n_m	number of model parameters in space domain, length of $\mathbf{m}$
n_x	number of model parameters in wavelet domain, length of $\mathbf{x}$

Contents

Chapter 1

Geophysics

1.1 Motivation

1.1.1 Geophysics

A geophysicist is concerned with the physical processes and properties of the Earth, such as the Earth's gravitational and magnetic fields, the Earth's internal structure and the dynamics in plate tectonics. He will utilize quantitative methods based on physical laws. For example, the laws of electricity and magnetism are used to better understand the ionosphere and magnetosphere.

When one tries to determine the geology of the subsurface, one can drill expensive boreholes in order to prospect the subsurface locally, however this technique is slow. In exploration geophysics, other non-invasive techniques will be used to determine the physical properties of the subsurface. These properties will be mapped to the geology, via mostly empirical laws.

1.1.2 Exploration methods

Geophysical exploration methods are classified into natural field and artificial source methods [Kearey et al., 2013]. The first class studies perturbations in natural fields of the Earth, for example, the gravitational field. These methods are appropriate for studying geology at a large scale and at great depths. For example, in magnetic surveying, the spatial variations in the Earth's magnetic field are recorded and used to detect local perturbations or anomalies. Early magnetic surveying at sea [Heirtzler et al., 1968] showed that the oceanic crust is characterized by a pattern of linear magnetic anomalies, the oceanic crust is alternately magnetized in a normal and reverse direction[1]. This discovery led to the theory of seafloor spreading[2] and it provided a time scale for polarity transitions in the Earth's magnetic field.

[1] In geological sciences, alternately magnetized patterns are a well-known consequence of geomagnetic reversal. This is the reversal of the Earth's magnetic north and south.

[2] Seafloor spreading is the effect of new oceanic crust that gradually moves away from a ridge, this is mostly at mid-ocean ridges with volcanic activity.

The other class of exploration techniques uses artificial (man-made) sources, for which, in contrast to natural field methods, the source can be tuned. These techniques drive to a more detailed image of the subsurface, but on the other hand, the signal does not penetrate to great depths. Seismic imaging is a common artificial source geophysical method. The density and elastic moduli of the sediments determine the propagation velocity of seismic waves and by measuring travel times of reflected seismic waves, the distribution of the subsurface parameters can be obtained. Seismic methods have the disadvantage to be logistically intensive. Alternatively, electrical methods are adopted. In resistivity methods, the electrical conductivity is obtained by measuring the subsurface's resistance. Electromagnetic methods can be used as well, in which the response to electromagnetic (EM) radiation is measured. The measurement of terrain resistivity has several shortcomings: It requires placement of electrodes all over the terrain, which involves relatively large amount of manpower[3]. Secondly, inhomogeneities, small compared to the depth of exploration, cause significant error in the measurements[4]. Additionally, resistivity techniques are sensitive to telluric currents, these natural electric currents flow through the soil and cause regional potential gradients. This led academics [Stefanesco et al., 1930] and companies such as Geonics Limited [McNeill, 1980] to examine the potential of electromagnetic techniques. These techniques have similarities with resistivity methods, although data acquisition can be executed faster, e.g. via airborne measurements. This thesis will elaborate on certain challenges related to these electromagnetic (induction) techniques. A third common electromagnetic technique is ground penetrating radar, where the penetration of radar waves (with frequencies ranging from 10MHz to 10 000MHz) are used. It is analogous to seismic imaging, because the reflection of the radar wave is used. The propagation velocity of radar waves depends mainly on the dielectric constant.

1.1.3 Hydrogeophysics

The link between physics and hydrogeology is made via Archie's law [Archie et al., 1942], an empirical law that maps the electrical conductivity to the soil characteristics. The law determines the conductivity of the soil σ from the conductivity of brine σ_w, the porosity fraction θ and the amount of water in the pores S:

$$\sigma = \sigma_w \theta^m S^n. \tag{1.1}$$

The exponents m and n need to be determined experimentally and depend on the soil.

Geophysical surveying has multiple applications and is especially valuable when a properly performing reconstruction scheme is available. It has economic value as it can be used to locate metalliferous mineral deposits. Archaeological surveys can be executed faster by imaging the underground structures non-invasively [Tang et al., 2018]. Geophysical surveying also has societal merits, for example in coastal areas, where through groundwater prospection, the salt intrusion of seawater can be mapped. In the work of [Hermans

[3]It must be noted that this is a historical sketch. Advances in instrument technology have led to the development of electrodes in the form of antennae that only need to be towed over the surface or spiked wheels that ensure continuous galvanic contact with the soil. [Loke et al., 2013]

[4]More technically, one can overcome this issue by measuring at various inter-electrode spacings, at the cost of more laborious data acquisition efforts.

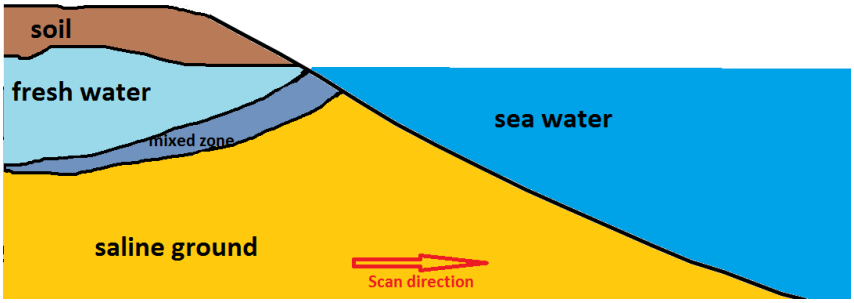


Figure 1.1: The standard seawater intrusion model exhibits an approximately horizontally stratified structure.

et al., 2012], the natural reserve of Westhoek, Belgium is prospected for salt water intrusion. In that area, drinking water is produced via the extraction of fresh water from the available aquifers. Too much water extraction puts pressure on the hydrogeological equilibrium and it is therefore vital to map the spatial extent of the body of infiltrated salt water. Indeed, concerning this context, electromagnetic techniques are appropriate, since salty water greatly influences the electrical conductivities. Now we have translated the hydrogeophysical problem back to Maxwell theory. We need to propose a model for the structure of the soil, a *conductivity profile*. In coastal areas, the subsurface layers are often flat and hence a horizontally stratified structure will be proposed in Section 1.3. Figure 1.1 illustrates a typical profile of the coastal subsurface.

1.2 Electromagnetic induction exploration

1.2.1 Fundamental principle

Non-contacting conductivity measurements are based on the principle of electromagnetic induction. By M. Faraday's experiments reported in 1831, we comprehend that a changing magnetic field induces an electric field. Later, J. Maxwell came up with a corrected version of Ampère's law, widely known as Ampère-Maxwell law, from which it follows that a changing electric field induces a magnetic field. These fundamental principles form the basis of the geophysical Electromagnetic Induction (EMI) technique. If we are able to generate a changing magnetic field above the soil, so-called *eddy currents* will appear in the subsurface, which in turn generate a response magnetic field.

1.2.2 A standard set-up

There are various conceivable experimental set-ups of the electromagnetic induction technique. In this thesis, we focus on the set-up of a mobile measuring device as shown in Figure 1.2. A *primary magnetic field* is generated in the transmitter coil. Such a transmitter coil consists of a wire that is coiled several times around a ring. When an electric current is carried through the wire, the transmitter coil can be modelled as an ideal magnetic dipole[5] (via the Biot-Savart law). However, in EMI, an alternating current

[5]The transmitter coil is not an ideal magnetic dipole (it is so when it has a vanishing diameter), but assuming an ideal dipole simplifies the calculations. The approximation is valid when the quadrupole

will be carried through the wire and the sinusoidal current will generate an alternating magnetic dipole. If the soil is conductive, the alternating magnetic dipole will generate eddy currents in the subsurface.

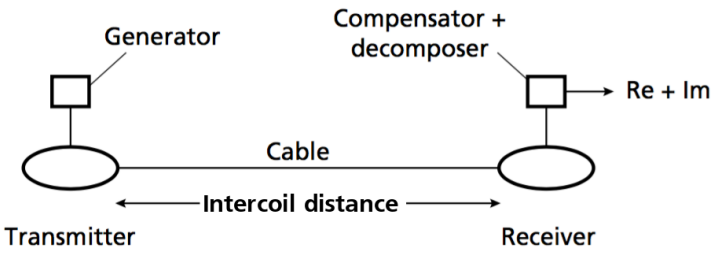


Figure 1.2: Mobile transmitter–receiver electromagnetic field equipment. Figure adapted from [Kearey et al., 2013].

These eddy currents, in turn, generate a response of the soil, called the *secondary magnetic field*, which is measured in the receiver coil. Note that in the case of a continuous alternating current in the transmitter coil, the primary and secondary fields will both be measured in the receiver coil. This coil has an identical construction as the transmitter coil, except that it is connected with a data acquisition device instead of a current generator.

The transmitter and receiver coil are usually connected by a cable, which connects the information (frequency, magnitude of the voltage) from the current generator to the logger device. The *intercoil distance* is the distance between the two coils. It turns out that this distance determines the depth of exploration (see later). There is freedom in the orientation of the coils (they can greatly alter the difficulty of the analytical modelling!). If the transmitter coil is placed horizontally above the surface, a vertical dipole will be generated, while a vertical placement generates a horizontal dipole. The orientation of the receiver coil determines which component of the secondary magnetic field that is measured. Examples of common configurations are shown in Figure 1.3. The configurations 1-3 are most common. Note that the NULL configuration is also perpendicular (as PERP), but the configuration is not responsive to the cylindrically symmetric stratified earth, when varied along the vertical plane of the transmitter.

The set-up considered in this section is deliberately quite general. More details on a realistic experimental set-up and instrumentation will be provided in Section 2.8, where the data acquisition performed in De Panne, Belgium in collaboration with prof. Dr. T. Hermans' team[6] is discussed.

As mentioned in Section 1.1, artificial source methods (as EMI) are particularly suitable for mapping the near subsurface. In general, for electromagnetic induction techniques, the

and higher order terms are negligible w.r.t. the dipole term. This is especially true at larger distances. More details can be found Chapter 3 of [Wait, 1982].

[6]I notably want to thank Benjamin Maveau and Marieke Paepen again for working together during the field work.

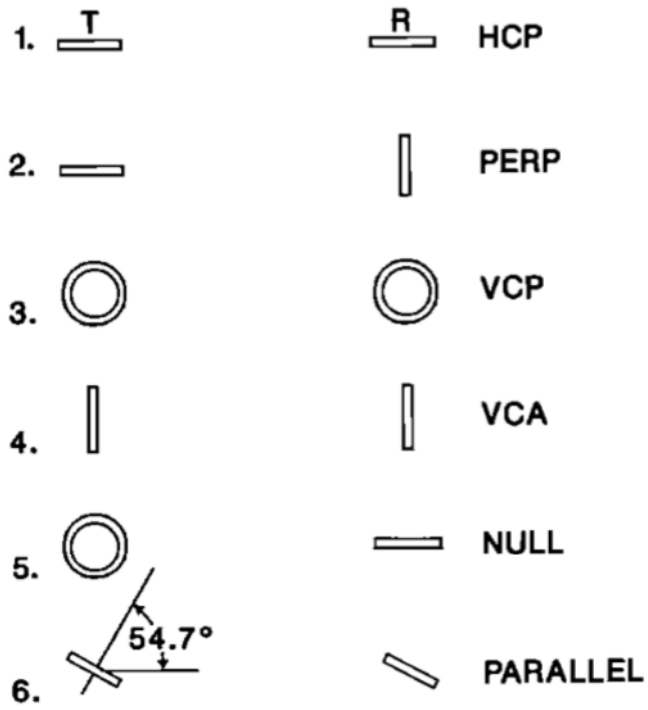


Figure 1.3: Common dipolar loop configurations (HCP = Horizontal Coplanar, VCP = Vertical Coplanar, VCA = Vertical Co-Axial). Figure retrieved from [Nabighian and Corbett, 1988].

depth of penetration, also called depth of investigation (DOI), is a measure of how far the electromagnetic fields reach in the matter and it is understood by the *skin effect*. The skin depth δ is defined as the distance it takes to reduce the amplitude of an electromagnetic wave by a factor $e^{-1} \approx 1/3$ [Griffiths, 2005]

$$\delta = \sqrt{\frac{2}{\omega\sigma\mu}}. \tag{1.2}$$

From this equation, it is difficult to determine an exact formula for the penetration depth. We can deduce that electromagnetic fields penetrate less deeply into highly conductive media and for alternating high-frequency electromagnetic fields. The frequency can thus be tuned to get the desired depth range. Note that the skin effect is derived for electromagnetic plane waves. In a general setting, the skin effect gives some maximum on the depth of penetration, but in our specific setting, it is different.

In Figure 1.2, a magnetic dipole is placed above the soil and it turns out under conditions that will be further elaborated in Section 2.4, that the depth of penetration is primarily limited by the fall-off of the magnetic dipole field. It rather depends on the intercoil distance s, than on the electrical conductivity and frequency. To give a sense of the numbers, take as a rule of thumb that for a horizontal dipole position the depth of exploration is $3s/4$ and $3s/2$ for vertical dipoles [McNeill, 1980]. Note that the amplitude of the magnetic dipole is still damped, due to of the skin effect. This effect will be examined in next chapter in the discussion of the recently proposed Damped model [Maveau et al., 2017].

1.2.3 Time domain vs frequency domain

A continuous sinusoidal electric current yields a measurement of both the primary and secondary magnetic field in the receiver coil. Knowing that the secondary magnetic field is much smaller than the primary field, we will, therefore, have poor accuracy. Time Domain EM surveying (TDEM) can resolve this issue. Instead of a continuous alternating current, current pulses are sent through the transmitter coil. The pulses are separated by an inactive primary magnetic dipole and the secondary field can be measured in the absence of the primary field. The principal physical principles are a bit different in this context. Eddy currents diffuse and gradually dissipate by resistive heat release. In higher conductive bodies, eddy currents circulate around the boundary and decay more slowly. Measurement of the decay rate of the waning eddy currents allows to estimate the conductivities. Most methods propose to solve the problem in the frequency domain and subsequently apply the inverse Fourier transform, in order to translate the solution in the time-domain [Ward and Hohmann, 1988], [Farquharson and Oldenburg, 1993], [Ralph-Uwe et al., 2008].

The analytical TDEM analysis actually exploits the work that has been done in the frequency domain, for which an exact formulation was available quite early. As already mentioned, in Frequency Domain EM surveying (FDEM), a continuous alternating current is carried through the loop and as we will see in the next chapter, the Maxwell equations can be reduced to the quasi-stationary field equations. The secondary magnetic field is generally a complex function of conductivities, intercoil distance and frequency. We will see that under some conditions, the relationship can be reduced to a linear model. These models are both by J. Wait, in [Wait, 1951] and [Wait, 1962][7] respectively. A recent model, proposed by [Maveau et al., 2017], will also be examined.

Notwithstanding the potential of TDEM analysis, the FDEM will be studied in this thesis. Sufficient analytical models are available, however solving the inverse problem remains challenging. We will propose a new inversion scheme for this problem. An *inversion scheme* is a sequence of consecutive steps that allows to reconstruct a conductivity profile from a given dataset.

1.3 Problem description and general method

In the hydrogeophysical setting of salinization in coastal areas, it is justified to present the geology of the subsurface as stratified earth with only horizontal layers (this is called an 1D model). Each layer has a thickness Δh_i, conductivity σ_i and magnetic permeability μ, which we set equal for every layer[8]. We assume a vanishing conductivity of air. This profile is shown in Figure 1.4, where a vertical dipole is depicted at height h_0. h denotes the total depth, measured from the surface of the soil, while h_i is the depth of layer i. The M^{th} layer has a semi-infinite thickness. In Chapter 2, different forward models will

[7] In the literature, the model is sometimes referred to as by J.D. McNeill [McNeill, 1980], however J. Wait notes in his book [Wait, 1982] that McNeill's interpretation approach was already described in [Wait, 1962].

[8] Permeability is the measure of the ability of a material to support the formation of a magnetic field, we presume that this does not differ significantly from sediment to sediment.

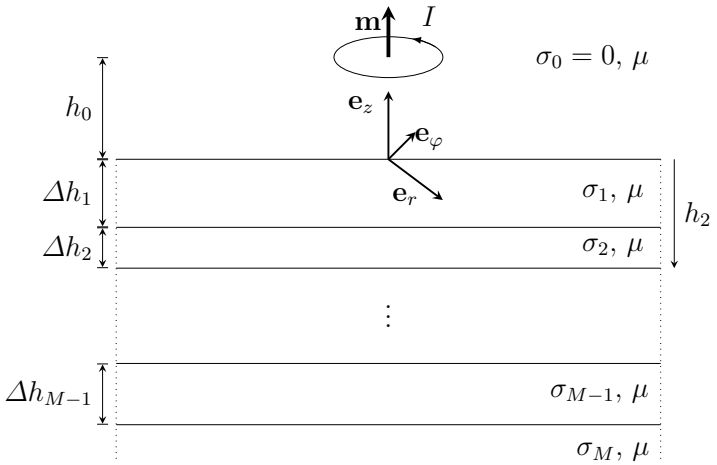


Figure 1.4: Stratified earth model with a vertical dipole with magnetic moment $\mathbf{m}$. The layers differ in depth and conductivity.

be presented for the geophysical electromagnetic induction technique. A *forward model* is a model where the response (i.e. secondary magnetic field) is expressed in terms of the primary magnetic field and conductivities for both a vertical and horizontal dipole.

After data acquisition (measuring the magnetic field), the inverse problem is considered. This problem is concerned with determining the conductivity profile from the data. In Chapter 3, we examine how the parameter distribution can be recovered and argue why the inverse problem is challenging. The optimization problem will be exactly defined. We will need to rely on iterative methods, such as the quasi-Newton limited-memory Broyden–Fletcher–Goldfarb–Shanno method for bound constraints (L-BFGS-B-method).

An inversion scheme usually consists of fitting the data, via the minimization of a least squares functional. For stable inversion, a so-called *regularization* term will be added to the minimization functional. This regularization term will impose an additional constraint. In our scheme, minimum-structure solutions will be promoted. This is an example of Occam's razor, it is the principle or belief that simpler solutions are more likely to be correct than complex ones. We will consider a basis transformation in which minimum structure solutions can be represented in a sparse form and impose sparsity on the solution in the regularization term. Wavelet theory provides us with such a basis transformation. With all the aforementioned ingredients, an inversion scheme for this inverse problem will be proposed and tested in Chapter 5 using the wavelet theory developed in Chapter 4

Chapter 2

Forward Modelling of the EMI-problem

2.1 Introduction

The electrical conductivity σ is the principal parameter that is measured, with units in Siemens per meter (S/m) (In geophysical literature, it is often referred to as ground conductivity in mS/m). The objective of this chapter is to deduce the soil's response to a magnetic dipole placed at height h_0 above the surface and a soil with a specific parameter distribution. All the models that we discuss in this thesis are related to the same stratified conductivity profile, as discussed in Section 1.3.

First, we describe the quasi-stationary regime in which we rewrite and solve the Maxwell equations. We then provide a summary of the long calculations of the *exact model* by Wait [Wait, 1951] and a brief interpretation based on what we know about transmission lines. Under the Low Induction Number (LIN) condition (see below), the complicated exact model can be simplified. It does not take into account couplings between the eddy currents and the electromagnetic dampening of fields in conductive media. This simplification is especially necessary from a computational point of view and results in the *LIN approximation*, a linear model. This model was developed in the 60s [Wait, 1962] but is still used today (e.g. [Corwin and Lesch, 2003]). We additionally look at a recently proposed model, the *damped model* [Maveau et al., 2017], which takes into account the dampening of electromagnetic fields in conductive media. Such simplified models pave the path for 2D and 3D problems, where subsurface layers do not necessarily need to be horizontal. Afterwards, we briefly look at a quantitative comparison of the models. In this thesis, we do not focus on the modelling problem, accordingly the comparison is kept concise. In view of the inverse problem, the expressions for the magnetic fields are discretized in Section 2.7. Finally, more details about data acquisition instrumentation will be provided and the results from measurements in the Westhoek will be reported.

2.2 Quasi-stationary Maxwell equations

In this section, the standard Maxwell equations in the quasi-stationary field regime are examined. The Maxwell equations are [Griffiths, 2005]

$$\nabla \cdot \mathbf{E} = \frac{1}{\epsilon_0}\rho \qquad \text{Gauss's law,} \tag{2.1}$$

$$\nabla \cdot \mathbf{B} = 0, \tag{2.2}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \qquad \text{Faraday's law,} \tag{2.3}$$

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} + \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \qquad \text{Ampère-Maxwell law,} \tag{2.4}$$

where ρ is the charge density and μ_0 and ϵ_0 are respectively the permeability and permittivity of the vacuum. Historically, Maxwell added the *displacement current*

$$\mathbf{J}_d = \epsilon_0 \frac{\partial \mathbf{E}}{\partial t}. \tag{2.5}$$

In the derivation of the forward models, a magnetic dipole with harmonic time dependence will be assumed. It is convenient to write this with the exponential $\exp(i\omega t)$, where $\omega = 2\pi f$ is the angular frequency. The actual physical quantity is the real part of the complex phasor. The Maxwell equations in the frequency domain are obtained by substituting $\mathbf{E}(\mathbf{r}, t) = \mathbf{E}(\mathbf{r}) \exp(i\omega t) \rightarrow \mathbf{E}(\mathbf{r})$ and $\frac{\partial \mathbf{E}(\mathbf{r},t)}{\partial t} = i\omega \mathbf{E}(\mathbf{r}, \mathbf{t}) \rightarrow i\omega \mathbf{E}(\mathbf{r})$ and analogously for the other vector field quantities. Using the constitutive equation in the absence of magnetic materials, $\mathbf{B} = \mu_0 \mathbf{H}$, yields

$$\nabla \cdot \mathbf{E} = \frac{1}{\epsilon_0}\rho = 0, \tag{2.6}$$

$$\nabla \cdot \mathbf{H} = 0, \tag{2.7}$$

$$\nabla \times \mathbf{E} = -i\mu_0 \omega \mathbf{H} \tag{2.8}$$

$$\nabla \times \mathbf{H} = \mathbf{J} + i\epsilon_0 \omega \mathbf{E}, \tag{2.9}$$

where the subsurface is assumed to be electrically neutral. This means that there are no net electrical charges, i.e. $\rho = 0$.

Finally, the displacement current (2.5) can be neglected when $i\epsilon_0\omega \ll \sigma$, seeing that Eq. (2.9) with Ohm's law yields

$$\nabla \times \mathbf{H} = \mathbf{J} + \mathbf{J}_d \tag{2.10}$$

$$= \sigma \mathbf{E} + i\epsilon_0 \omega \mathbf{E} \tag{2.11}$$

$$= (\sigma + i\epsilon_0 \omega)\mathbf{E}. \tag{2.12}$$

Since the direct derivation of a formula for the fields is an elaborate process, we resort to using potentials instead of fields. It is well-known [Griffiths, 2005] that the electric field $\mathbf{E}$ and the magnetic induction $\mathbf{B}$ can be written in terms of a magnetic vector potential

A and an electric scalar potential V

$$\mathbf{B} = \boldsymbol{\nabla} \times \mathbf{A}, \tag{2.13}$$

$$\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t} - \boldsymbol{\nabla} V, \tag{2.14}$$

where gauge fixing copes with the redundant degrees of freedom. [Maveau et al., 2017] use the Weyl or temporal gauge, because this brings the derivation as close as possible to the magnetostatics case. In the Weyl gauge, the electric scalar potential is zero, hence

$$\mathbf{H} = \frac{1}{\mu_0} \boldsymbol{\nabla} \times \mathbf{A}, \tag{2.15}$$

$$\mathbf{E} = -i\omega \mathbf{A}. \tag{2.16}$$

The Maxwell equations now need to be written in terms of the vector potential **A**. The following Helmholtz equation[1] is obtained

$$(\nabla^2 - k_i^2)\mathbf{A}_i = \mathbf{0} \quad k_i^2 = i\omega\mu_0\sigma_i, \tag{2.17}$$

by substituting Equations (2.15) and (2.16) in the Ampère-Maxwell equation and with Gauss' law. The wave number k_i is purely imaginary due to the neglect of the displacement current in our dynamics.

2.3 Exact solution in real space

2.3.1 Summary of the calculations

In this section, the full model is derived within the quasi-stationary regime of the previous section. This result will be referred to as the exact model. The horizontally stratified conductivity profile has already been introduced in Section 1.3, Chapter 1, including the meaning of the symbols.

The results of cumbersome calculations are summarized [Maveau et al., 2017], [Deleersnyder et al., 2017], [Wait, 1982]. First, the case of a vertical dipole is considered, because the cylindrical symmetry simplifies the calculations. The Helmholtz Equation (2.17) is solved using separation of variables, yielding exponential functions in z and cylindrical Bessel functions $J_{1,2}$ for the radial part. Omitting the non-physical (i.e. exploding) solutions and summing (integrating) over all eigenvalues, yields

$$\mathbf{A}_0 = \mathbf{e}_\phi \frac{m\mu_0}{4\pi} \int_0^{+\infty} f(\lambda) \exp(-\lambda z) J_1(\lambda s)\, \mathrm{d}\lambda + \frac{\mu_0}{4\pi} \frac{\mathbf{m} \times \mathbf{r}}{r^3}, \tag{2.18}$$

$$\mathbf{A}_i = \mathbf{e}_\phi \frac{m\mu_0}{4\pi} \int_0^{+\infty} g_i(\lambda) \exp(\gamma_i z)[1 + x_i(\lambda) \exp(-2\gamma_i z)] J_1(\lambda s)\, \mathrm{d}\lambda\,, \tag{2.19}$$

$$\mathbf{A}_N = \mathbf{e}_\phi \frac{m\mu_0}{4\pi} \int_0^{+\infty} g_N(\lambda) \exp(\gamma_N z) J_1(\lambda s)\, \mathrm{d}\lambda\,, \tag{2.20}$$

[1] It is not a surprise to find a time-independent form of the wave equation, since we have dealt with the time-dependence by considering the frequency domain.

for the components of the magnetic vector potential, where

$$\gamma_i = \sqrt{\lambda^2 + k_i^2}. \tag{2.21}$$

The second term in $\mathbf{A}_0$ is the magnetic vector potential of an ideal magnetic dipole with moment $\mathbf{m}$. The functions $f(\lambda)$, $g_i(\lambda)$ and $x_i(\lambda)$ result from solving the separated differential equations and do depend on the boundary conditions.

The next step is to apply boundary conditions to Eqns. (2.18)-(2.20). In this setting, the relevant boundary conditions at the interface of two adjacent layers 1 and 2 are [Griffiths, 2005]

$$\mu_1 \mathbf{H}_1^{\perp} - \mu_2 \mathbf{H}_2^{\perp} = 0, \tag{2.22}$$

$$\mathbf{H}_1^{\parallel} - \mathbf{H}_2^{\parallel} = \mathbf{K}_f \times \hat{\mathbf{n}}, \tag{2.23}$$

where $\mathbf{K}_f$ is surface current density. The unit vector $\hat{\mathbf{n}}$ is perpendicular to the interface and points from layer 2 toward layer 1. We do not expect surface currents, since they are only expected for highly conductive (metallic) layers. Also, we assume that all permeabilities are equal and thus all components in the magnetic field $\mathbf{H}$ are continuous at every boundary. In terms of the vector potential $\mathbf{A}$, we write

$$\boldsymbol{\nabla} \times (\mathbf{A}_i - \mathbf{A}_{i+1}) = 0. \tag{2.24}$$

After applying the boundary conditions, a recursive relation is obtained

$$f(\lambda) = \lambda \frac{\gamma_0 - Y_1}{\gamma_0 + Y_1} \exp(-2\lambda h_0), \tag{2.25}$$

$$Y_i := \gamma_i \frac{1 - x_i \exp(-2\gamma_i h_{i-1})}{1 + x_i \exp(-2\gamma_i h_{i-1})} \tag{2.26}$$

$$= \gamma_i \frac{Y_{i+1} + \gamma_i \tanh \gamma_i \Delta h_i}{\gamma_i + Y_{i+1} \tanh \gamma_i \Delta h_i}. \tag{2.27}$$

The starting point of the recursive relation is for $x_N = 0$ (this can be seen from Eq. (2.20)) or $Y_M = \gamma_M$.

For a horizontal dipole, cylindrical symmetry is lost and the calculations become more cumbersome. [Wait, 1982] circumvents this issue by solving the problem with a magnetic monopole instead of a dipole. Afterwards, the solution is transformed to a dipole-case, via

$$\frac{\mathbf{m}}{q} \cdot \nabla_{\mathbf{r}'} \Big|_{\mathbf{r}'=0}, \tag{2.28}$$

where q is the strength of the monopole and $\mathbf{m}$ the strength of the dipole. After tedious calculations, the function $f(\lambda)$ is

$$f(\lambda) = -\frac{\lambda - Y_1}{\lambda + Y_1} \exp(-2\lambda h_0), \tag{2.29}$$

while the functions Y_i remain the same.

Let us finally note that other derivations exist for this specific stratified earth problem. Zhdanov derives an expression in the Fourier domain. It can be shown (e.g. in [Deleersnyder et al., 2017]) via the Hankel transform[2] that these approaches are equivalent.

2.3.2 Analogy with transmission lines

The recursive relations (2.26) and (2.27) are complex. At first glance, they do not seem to provide much insight in what physically happens. However, a reader familiar with transmission lines will recognize the relations. In transmission lines [Johnson et al., 2003], the propagation of electromagnetic waves (or alternating voltages) are described over long distances (that is: the size of the line is typically much larger than the wavelength), where the current is not the same all the way down the line. If we instantaneously put a voltage source across a transmission line, a certain current flows to create a voltage wave that travels down the transmission line.

In direct current circuits, the resistance describes the ratio of the voltage over the current, while with alternating currents, complex impedances are used. The characteristic impedance Z_i depends on the physical characteristics of the line (as that is also the case with resistance in DC circuits) and determines the magnitude of the current. At an impedance discontinuity, the reflection coefficient describes how much of an electromagnetic wave is reflected due to the discontinuity (that is the same as with acoustic waves). The reflection coefficient of the discontinuity between characteristic impedances Z_1 and Z_2 is then

$$\Gamma = \frac{Z_1 - Z_2}{Z_1 + Z_2}. \tag{2.30}$$

The impedance measured at a given distance h (between source and discontinuity) of the discontinuity is[3]

$$Z_{in} = Z_1 \frac{1 + \Gamma e^{-2\gamma h}}{1 - \Gamma e^{-2\gamma h}}, \tag{2.31}$$

where γ is the (complex) propagation constant. It depends on the characteristic impedance Z_1 of the medium and the reflection Γ. Rearranging the above equation yields

$$Z_{\text{in}}(h) = Z_1 \frac{Z_2 + Z_1 \tanh(\gamma h)}{Z_1 + Z_2 \tanh(\gamma h)}, \tag{2.32}$$

in terms of the Z_2 instead of the reflection coefficient. Z_1 is the characteristic impedance of the sub-circuit in the transmission line while the actual impedance at a specific position

[2]The Hankel transform $F_\nu(k) = \int_0^\infty f(r) J_\nu(kr)\, r\, \mathrm{d}r$, where ν is the order of the Bessel function, is related to the Fourier transform. This can be seen by equating the 2D inverse Fourier transform of a function $f(\mathbf{r})$ with the inverse zero-order Hankel transform of $f(r)$.

[3]Without giving too much detail, Equation (2.31) is obtained from the Telegraphers' equations (see [Johnson et al., 2003]). When a forward wave reflects off at the far end, then the reflected wave returns to the near end. The length will determine the decay of the electromagnetic wave (indeed, in conductive media, electromagnetic waves decay). For the voltage at length h, it is simply the sum of both forward and reflected wave: $V_{\text{in}} = V_+ \exp(+\gamma h) + V_- \Gamma \exp(-\gamma h)$. The current can be obtained via the Telegraphers equation $\frac{\mathrm{d}V}{\mathrm{d}z} = -(j\omega L + R)I$. We write $I_{\text{in}} = I_+ \exp(+\gamma h) - I_- \Gamma \exp(-\gamma h)$. With $Z_{\text{in}} = V_{\text{in}}/I_{\text{in}}$, Equation (2.31) is obtained.

is as in Eq. (2.31). For multiple discontinuities, the relation is recursive.

An impedance discontinuity corresponds with a new subsurface layer (indeed, the different ground conductivity will yield a different electrical impedance). Every layer corresponds with a specific sub-circuit with its characteristics. The sub-circuit of the air is quite simple. The conductivity $\sigma_0 = 0 \Rightarrow k_i = 0 \Rightarrow \gamma_0 = \lambda$. The characteristic impedance of the air is $Z_0 = \gamma_0$ with reflection coefficient

$$\Gamma_0 = \frac{Z_0 - Y_1}{Z_0 + Y_1}. \tag{2.33}$$

From the equation for $f(\lambda)$ and the z dependence in the integral of $\mathbf{A}$, rewrite

$$f(\lambda)\exp(-\lambda z) = \lambda\frac{\gamma_0 - Y_1}{\gamma_0 + Y_1}\exp(-2\lambda h_0) \tag{2.34}$$

$$\Leftrightarrow f(\lambda)\exp(-\lambda(z - h_0)) = Z_0\Gamma_0\exp(-\lambda h_0). \tag{2.35}$$

$f(\lambda)\exp(-\lambda z)$ describes the reflected measured impedance at a given distance in the air.

2.4 LIN Approximation

The model in Section 2.3 is highly non-linear. In order to reduce the computational burden, a linear model was proposed. [Wait, 1962] proposed to linearise the model by slicing the subsurface earth in thin conducting sheets and to neglect the induction between these sheets and the self induction[4], i.e. 'independent sheets'. The total secondary magnetic field was obtained by summing the contributions of all these thin sheets, or integrating over infinitesimally small sheets.

2.4.1 Derivation via the exact solution

This LIN approach can be derived from the model by Wait in Section 2.3. First, an expression for the contribution to the secondary magnetic field for an infinitesimal, independent sheet is required. This expression is obtained from an infinitesimal thick layer $\mathrm{d}h$ at depth h with conductivity $\sigma(h)$ embedded in free space. In the framework of Wait's model, this yields

$$Y_1 = \gamma_1\frac{\lambda + \gamma_1\tanh(\gamma_1\mathrm{d}h)}{\gamma_1 + \lambda\tanh(\gamma_1\mathrm{d}h)}, \tag{2.36}$$

$$f(\lambda) = -\frac{k^2\mathrm{d}h}{2}\exp(-2\lambda h). \tag{2.37}$$

The next step is to calculate the secondary magnetic field response of such an infinitesimally thin sheet. Equations (2.36) and (2.37) are substituted in Eq. (2.18). Via the definition of the magnetic vector potential (2.13), the secondary magnetic field of the

[4]In his work, this approximation was effectively only retaining the first term in a series expansion. The approximation was only valid for sufficiently low frequencies. Retaining only the first term corresponded physically with neglecting self-induction in the sheet. Summing over all the independent sheets corresponds physically with neglecting the magnetic couplings (or linearisation).

sheet is obtained.

The secondary magnetic field of the complete semi-infinite subsurface is obtained by integration or superposition of the elementary secondary sources. Using the dimensionless quantity $\eta = hs^{-1}$ and putting the dipole field H_p at $\eta = 0$ yields the following magnetic field ratios

$$\frac{H_{s,z}}{H_{p,\text{vert}}} = \frac{i\omega\mu_0 s^2}{4}\sigma_{a,z} \tag{2.38}$$

and

$$\frac{H_{s,\rho}}{H_{p,\text{vert}}} = \frac{i\omega\mu_0 s^2}{4}\sigma_{a,\rho}, \tag{2.39}$$

while the φ-component is irrelevant, due to cylindrical symmetry. The expressions for the apparent conductivities are given by

$$\sigma_{a,z} = \int_0^\infty \sigma(\eta s)\frac{4\eta}{(4\eta^2+1)^{\frac{3}{2}}}\mathrm{d}\eta, \tag{2.40}$$

and

$$\sigma_{a,\rho} = \int_0^\infty \sigma(\eta s)\frac{2}{(4\eta^2+1)^{\frac{3}{2}}}\mathrm{d}\eta. \tag{2.41}$$

The z-component is measured in a horizontal coplanar system (HCP) coil setting, while the ρ-component is measured in a perpendicular (PERP) setting.

The dimensionless quantity η will be used on plots in Section 2.4.4, because in this model (under the LIN condition), there is no coupling and thus the depth of penetration only depends on the fall-off of the magnetic dipole. It is a linear model, in the sense that a function $\sigma(h)$ is linearly mapped on the magnetic field ratio. This linearity will make it possible to easily discretize this forward model (see Section 2.7 and the associated inverse problem in the next chapter).

For a horizontal dipole, the magnetic field ratio's are

$$\frac{H_{s,\rho}}{H_{p,\text{hor}}} = -\frac{i\omega\mu_0}{4}xy\int_0^{+\infty}\sigma(\eta s)2\left(2-\frac{4\eta}{\sqrt{4\eta^2+1}}-\frac{2\eta}{(4\eta^2+1)^{3/2}}\right)\mathrm{d}\eta, \tag{2.42}$$

$$\frac{H_{s,\varphi}}{H_{p,\text{hor}}} = -\frac{i\omega\mu_0}{4}s^2\int_0^{+\infty}\sigma(\eta s)\left[\frac{y^2-x^2}{s^2}\left(2-\frac{4\eta}{\sqrt{4\eta^2+1}}\right)-\frac{y^2}{s^2}\frac{2\eta}{(4\eta^2+1)^{3/2}}\right]\mathrm{d}\eta, \tag{2.43}$$

$$\frac{H_{s,z}}{H_{p,\text{hor}}} = -\frac{i\omega\mu_0}{4}ys\int_0^{+\infty}\sigma(\eta s)\frac{2}{(4\eta^2+1)^{3/2}}\mathrm{d}\eta. \tag{2.44}$$

The ρ-component of the secondary magnetic field is measured in the VCA coil orientation. The φ-component is measured in a VCP coil orientation and the z-component in the NULL configuration. Note that in the latter configuration, the response will *not* be null if the receiver coil is located outside the vertical plane of the transmitter coil.

2.4.2 Low Induction Numbers (LIN) condition

In an alternative derivation of the LIN approach, the result is obtained by a series expansion of an integral (e.g. see [Wait, 1962]). Only retaining the first term in the series expansion is only valid for $B = s/\delta$ sufficiently small. The dimensionless quantity B is called the induction number, s is the intercoil distance and δ is the skin depth. Rewriting the *low induction number condition* $B \ll 1$ yields

$$\frac{\omega\mu_0\sigma s^2}{2} \ll 1. \tag{2.45}$$

The meaning of the LIN condition is that if the skin depth is much larger than the path the electromagnetic field has to traverse, the dampening can be neglected. We have already mentioned that the path length is restricted by the fall-off of the magnetic dipole. It has the same order of the intercoil distance s.

Now, the condition can be interpreted via Eq. (2.45). For highly conductive media, the interaction between sheets is strong and cannot be neglected. Indeed, the induction number B is larger and the approximation is poor. For large intercoil distances, the lower lying layers are probed and the dampening effect is more pronounced. It thus makes sense that the approximation is poorer for larger intercoil distances. Finally, the angular frequency of the magnetic dipole has to be sufficiently low. A larger frequency yields a larger magnetic flux, also for the secondary magnetic fields and then the current loops are more strongly coupled. In Section 2.6, the models will be compared and it will be possible to recognize when the LIN condition breaks down.

2.4.3 The LIN condition and electronic circuits analogy

[McNeill, 1980] adds some insight by translating the problem to elementary circuit theory. While we have used a transmission line interpretation in the wave representation for the exact model, it can also be viewed as a series of circuits with load impedance Z_L. Consider two current loops with currents i_1 and i_2 respectively. An electromotive force $\mathcal{E}$ (voltage) energizes the current in the first loop i_1. From Faraday's law, it is well-known that an alternating electromotive force induces an alternating current. Additionally, a larger frequency induces a larger magnetic flux. In alternating current loops, there are three impedances that restrict the current to be infinite. Obviously, the electrical resistance R_1 in the loop limits the current. Secondly, the self-inductance L_1 and thirdly the mutual inductance M_{12} between both loops play a role. The complex impedance Z_L describes all effects of magnetic coupling between both loops. We have

$$Z_L = i\omega L_1 + \frac{\omega^2 M^2}{R_2 + i\omega L_2} \tag{2.46}$$

and

$$i_1 = \frac{\mathcal{E}}{R_1 + Z_L}. \tag{2.47}$$

From these equations, it can be easily seen that for sufficiently low frequencies, $|Z| \ll R_1$, the loops decouple. This is another manifestation of the independence at low frequencies. While it seems consistent with the LIN condition, the skin effect is not taken into account in this interpretation (current loops do not facilitate such behaviour).

2.4.4 Relative contribution in different coil settings

To get some more detailed intuition about the depth of exploration, the instrumental response as a function of depth is plotted in Figure 2.1. Recall that the dimensionless quantity η is used. $\xi(\eta)$ is the relative contribution (or relative response) to the secondary magnetic field from the normalized depth η and it is recognized from the expression of apparent conductivity

$$\sigma_a = \int_0^\infty \sigma(\eta s)\xi(\eta)\mathrm{d}\eta. \tag{2.48}$$

For the vertical dipole, the HCP coil orientation probes the lower lying conductivities more than those near the surface. The maximum contribution to the apparent conductivity will be from (more or less) a layer at depth $h = 0.5s$. For the PRP coil orientation, the first layer will contribute most significantly. In the HCP setting, the eddy currents in the upper layer do not contribute solely due to geometrical reasons. Eddy currents in conductive sheets behave as smoke rings[5]. The diameter of the current loop is small for the upper layers and the magnetic flux (z-component) is only large right above the eddy current.

For a horizontal dipole, we go to cylindrical coordinates $x = s\cos(\theta), y = s\sin(\theta)$. For $\theta = 0$, only the VCP coil configuration depends on the stratified earth. For $\theta = \pi/2$, only PRP,V or VCA does not depend on the stratified earth.

2.5 The damped model

In the previous section, we have discussed the LIN approximation. In this section, a hybrid model is derived in the same way as the LIN approach. The model is hybrid in the sense that it does not take into account magnetic couplings between current loops, but the effect of the dampening of the propagation of the fields is included. This model attempts to improve the LIN approach while preserving the spirit of simplicity (a closed form relation of the fields and (quasi-)linearity).

The derivation is analogous as with the LIN approach, but instead of embedding an infinitesimal thin sheet in free space, it is embedded in a half-space with a background conductivity σ_b. The layers with background conductivity are modelled to have no eddy currents, it is just there to model the dampening effect. The conductivity profile is described in three layers: the upper layer has a conductivity σ_b and reaches from the top of the soil to the thin sheet, the second layer is the infinitesimal thin sheet at depth h with conductivity $\sigma(h)$. The third layer is semi-infinite and has the background conductivity σ_b (see Figure 2.2a). Applying the exact model to this conductivity profile to first order in dh yields

$$Y_3 = \gamma_b, \qquad Y_2 \approx \gamma_b + (\gamma_h^2 - \gamma_b^2)\mathrm{d}h, \qquad Y_1 \approx \gamma_b + (\gamma_h^2 - \gamma_b^2)\exp(-2\gamma_b h)\mathrm{d}h, \tag{2.49}$$

$$f(\lambda) \approx \lambda\frac{\lambda - \gamma_b}{\lambda + \gamma_b}\left[1 + 2\lambda\frac{\sigma(h) - \sigma_b}{\sigma_b}\exp(-2\gamma_b h)\mathrm{d}h\right]\exp(-2\lambda h_0). \tag{2.50}$$

[5]There is a beautiful paper that gives some insight in the geometry of eddy currents, based on Maxwell's receding image construction [Saslow, 1992], esp. Section IV is relevant for our problem.

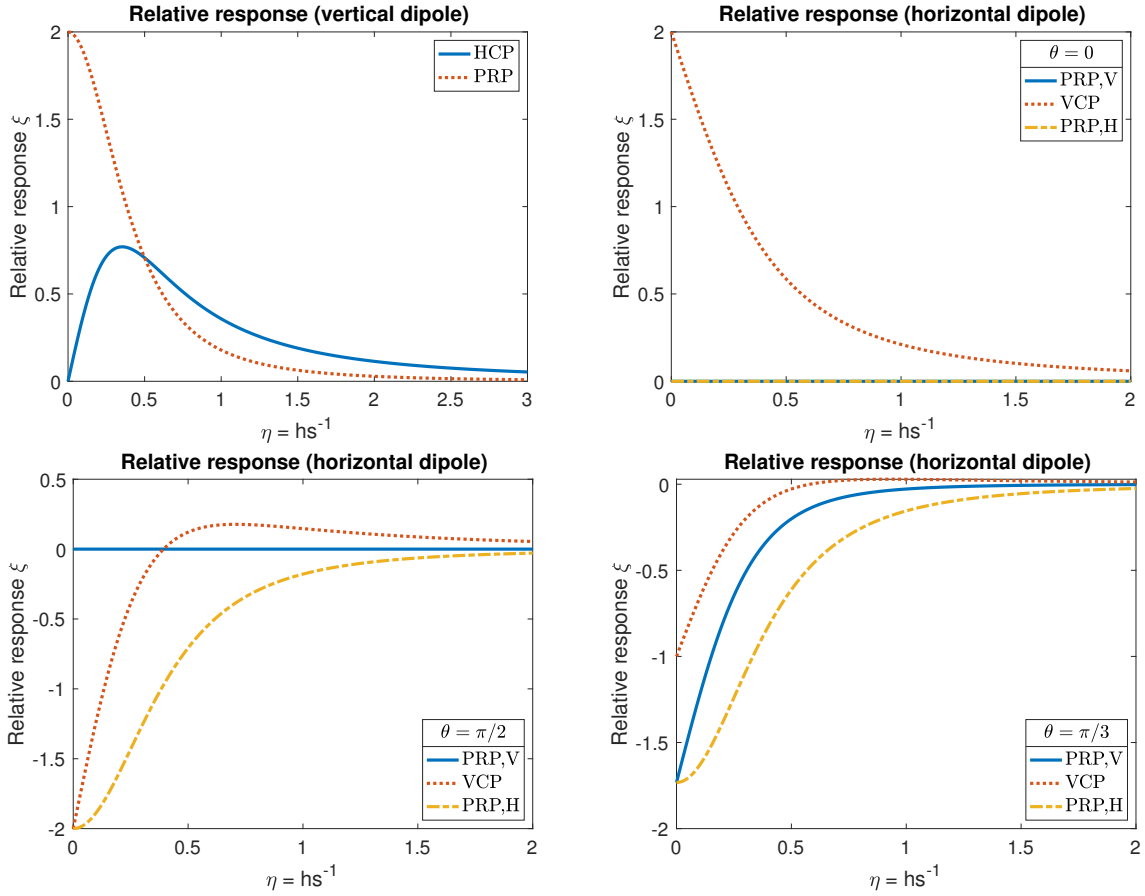
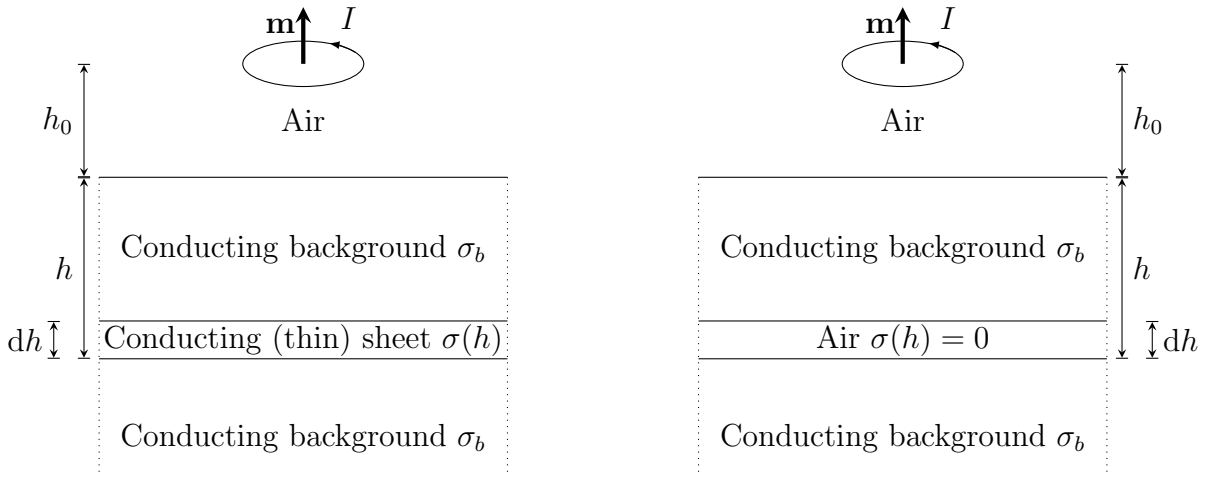


Figure 2.1: Relative response (in the LIN approach) for different coil orientations.



(a) Conductive sheet (b) Non-conductive sheet

Figure 2.2: A scheme of the rationale behind the Damped model. It consist of a dipole at height h_0 above the ground, a thin (non-)conductive sheet embedded in a conductive background.

As in the LIN approach, we need to integrate over all the sheets. The background conductivity σ_b should not directly contribute to the secondary magnetic field H_s, since no eddy currents are modelled in those layers. The unwanted contribution can easily be eliminated by subtracting the contribution of a thin sheet with vanishing conductivity with the same background σ_b (see Figure 2.2b). This development yields an integral with no known analytic solution. A further simplification (a Taylor approximation, see [Maveau et al., 2017]) allows to write the contribution of an infinitesimally small conducting sheet to the secondary magnetic field as

$$\frac{H_{\mathrm{d}h,z}}{H_{p,\mathrm{vert}}} \approx \frac{i\omega\mu_0 s^2}{4} 4\sigma(h)\mathrm{d}\eta \exp\left(-k_b s\sqrt{4z^2+1}\right) \frac{z}{4z^2+1}\left(k_b s + \frac{1}{\sqrt{4z^2+1}}\right) \tag{2.51}$$

$$\frac{H_{s,\rho}}{H_{p,\mathrm{vert}}} \approx \frac{i\omega\mu_0 s^2 s\mathrm{d}h}{4}\sigma(h) 2\frac{\partial^2 T(s,2h)}{\partial(2h)\partial s} \tag{2.52}$$

$$T(s,z) = \int_0^{+\infty} \frac{1}{\gamma}\exp(-\gamma z) J_0(\lambda\rho)\mathrm{d}\lambda = I_0\left[\frac{k}{2}(r-z)\right] K_0\left[\frac{k}{2}(r+z)\right] \tag{2.53}$$

for a vertical dipole, while for a horizontal dipole, the relevant kernel is

$$f(\lambda) \approx \frac{i\mu_0\omega\sigma(h)\mathrm{d}h}{2}\exp(-2\gamma_b h). \tag{2.54}$$

The functions I_0 and K_0 are the modified Bessel functions [Abramowitz and Stegun, 1965]. The explicit secondary magnetic fields will be given in Section 2.7, where the discretization of the magnetic field will be considered.

It was assumed that the dipoles lie on the surface. This is not an issue, because dipoles at a height h_0 can be modelled to lie on a horizontal layer with vanishing conductivity, i.e.

$$\tilde{\sigma}(h) = \begin{cases} 0 & 0 < h < h_0 \\ \sigma(h-h_0) & h_0 < h \end{cases}. \tag{2.55}$$

The second issue is to find a proper value for the background conductivity σ_b. A fixed value yields a linear model. However, [Maveau et al., 2017] argues that pinpointing the exact value is difficult. A small deviation in the optimal σ_b results in a large error in the magnetic field. We can give up the linearity, by considering a more difficult background conductivity based on knowledge of the profile (prior or acquired). Dampening is primarily caused by the layers on top of the thin conducting sheet. When the secondary magnetic field contribution of the i^{th} layer is calculated, σ_b is approximated as the weighted average of the conductivities of the the layers on top of the layer. The weights will be the thicknesses of those layers.

2.6 Behaviour of models

To gain more insight into the typical magnetic field ratios, the response is illustrated (Figure 2.3) for a conductivity profile in Liège [Hermans and Irving, 2017]. The conductivity profile is obtained via borehole logging. Both the ρ and z-component for measurements at different intercoil distances ranging from 1 to 40 metres at a height of 0.1 meters, as well as measurements at different heights from 0 to 20 meters with intercoil distance 20 meters are plotted in Figure 2.4. Note that the latter setting is not practical. In practice, measurements with different intercoil distances will occur more often. The response is slightly different for each model and the difference is most apparent for the z-component.

The LIN approximation's magnetic field ratio depends on a pre-factor that is quadratic in s and the apparent conductivity σ_a. Recall that the apparent conductivity is some weighted average of the subsurface's conductivity profile, and this also depends on s! The maximum of the relative contribution ξ thus depends on the intercoil distance. The LIN condition

$$\frac{\omega \mu_0 \sigma s^2}{2} \ll 1 \tag{2.56}$$

is quadratic in s, thus, for larger intercoil distances, the LIN assumption breaks down. This is clearly visible for the z-component of H_s/H_p. The damped model follows the exact response much more closely.

Let us now examine the measurements at different heights. If measurements are carried out at larger heights, then the secondary magnetic field drops. The strength of the magnetic dipole drops cubically with distance and is, therefore, less strong before it reaches the soil. Consequently, the secondary magnetic field is weak. It is also clear how the damped model follows the exact model much better than the LIN approximation. The LIN approximation deviates from the exact model at $h = 0$ metres with 1.37×10^{-3} A/m, which linearly drops to $h = 20$ meters with 1.14×10^{-3} A/m. The difference between the LIN approximation and the other models is so manifest because the intercoil distance s is already 20 meters (LIN condition breaks down).

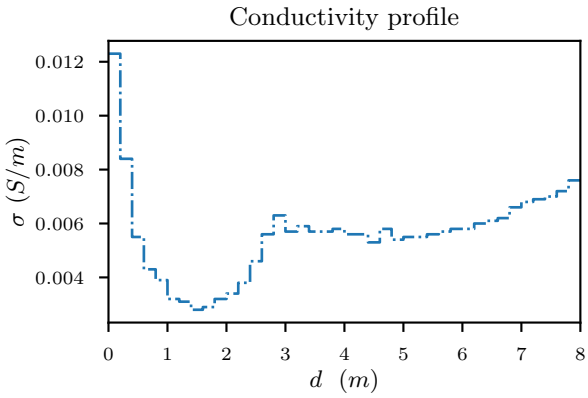


Figure 2.3: Conductivity profile from a site in Liège, obtained via borehole logging. Data retrieved from [Hermans and Irving, 2017].

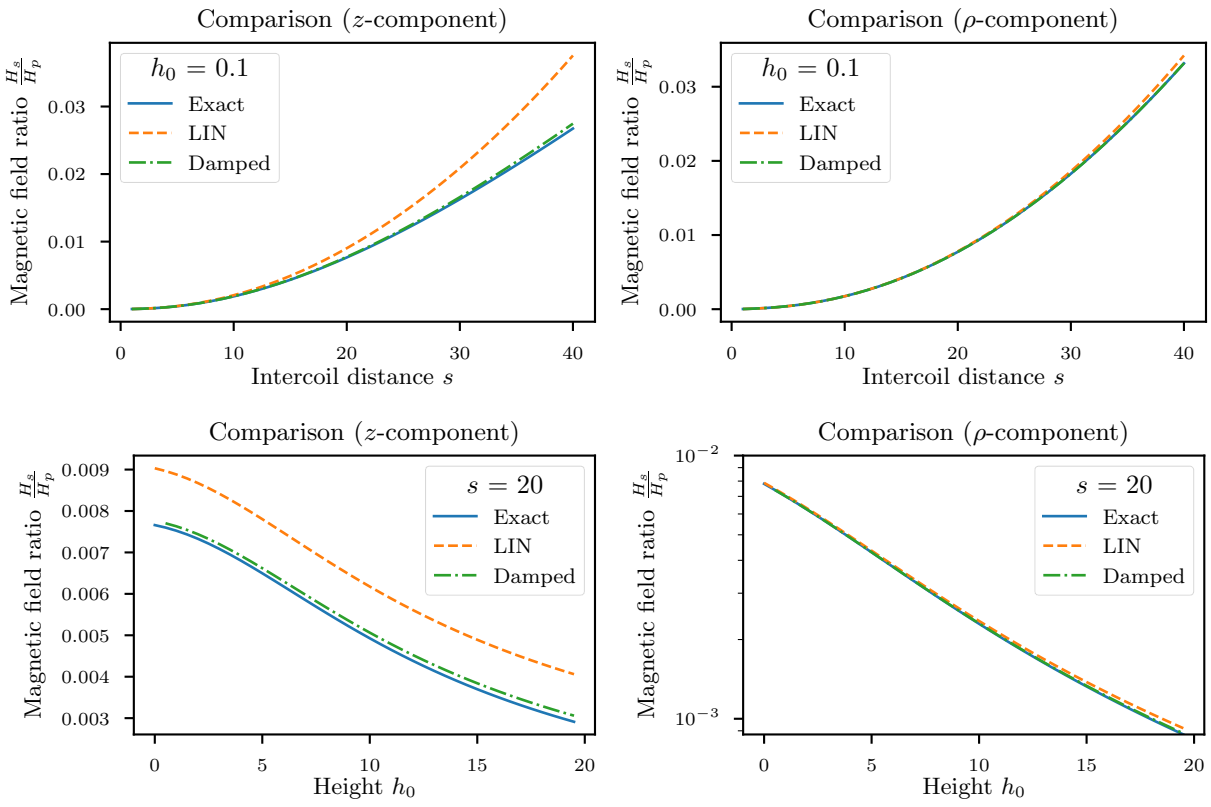


Figure 2.4: Behaviour of the response (in terms of heights and intercoil distances) for the conductivity profile from Figure 2.3.

2.7 Discretization

There is no need to have an expression that maps $\sigma(h)$ on $H_s(\rho, \phi, z)$, since our conductivity profile is assumed to be a series of step functions. This allows us to easily discretize the expressions.

2.7.1 Discretization of the LIN approximation

Discretizing the LIN approximation is basically discretizing the apparent conductivity

$$\sigma_a(h_0) = \int_0^\infty \sigma(\eta s)\xi(\eta), \tag{2.57}$$

associated to a measurement conducted at height h_0.

Consider a measurement of the field ratio conducted at heights $h_{0,1}$, ..., $h_{0,m}$ above the surface (dipole height as well as receiver height). Let $\boldsymbol{\sigma}_a$ denote the m-dimensional vector containing all of these values. As we will see below, the data from measurements with the EM34-3 are given in these apparent conductivities.

As the conductivity in a layer is a constant, Eq. (2.57) can be discretized via Fredholm integral equation of the first kind [Aster et al., 2018]. Consider an n-layered stratified earth, where the n-th layer is semi-infinite. Then the vector $\boldsymbol{\sigma}$ contains the electrical conductivities of these layers. As an example, the i-th entry of $\boldsymbol{\sigma}_a$ can be computed as follows:

$$\sigma_{a,i} = \sigma_a(h_{0,i}) = \sigma_1 \int_0^{h_1} \xi(h + h_{0,i})\mathrm{d}h + \sigma_2 \int_{h_1}^{h_2} \xi(h + h_{0,i})\mathrm{d}h + \cdots + \sigma_n \int_{h_{n-1}}^{\infty} \xi(h + h_{0,i})\mathrm{d}h.$$

Therefore the following matrix identity is obtained

$$\boldsymbol{\sigma}_A = K\boldsymbol{\sigma}, \tag{2.58}$$

where the ij-th entry of the matrix $K \in \mathbb{C}^{m\times n}$ is given by

$$K_{ij} = \int_{h_{j-1}}^{h_j} \xi(h + h_{0,i})\mathrm{d}h.$$

Then,

$$\frac{H_s}{H_p}(h_{0,i}) = \frac{i\omega\mu_0 s^2}{4} K_{ij}\sigma_j. \tag{2.59}$$

This matrix scheme is the forward implementation of the LIN approximation. The linearity of the LIN approximation allows a matrix K, independent of the conductivities $\boldsymbol{\sigma}$.

The experimenter might also want to conduct measurements at a variety of values for the intercoil distance. Similarly as above, this yields a unique matrix K, independent of the conductivities $\boldsymbol{\sigma}$.

2.7.2 Discretization of the damped model

Analogously, the discretization of the damped model is obtained. The matrix K is now dependent on the background conductivity σ_b. From [Maveau et al., 2017]:

Vertical dipole

$$\frac{H_{i,s,HCP}}{H_p} \approx -\frac{i\omega\mu_0\sigma_i s^2}{4}\left[\frac{\exp\left(-k_b s\sqrt{4\eta^2+1}\right)}{\sqrt{4\eta^2+1}}\right]_{\eta_i}^{\eta_{i+1}}, \tag{2.60}$$

$$\frac{H_{i,s,PRP}}{H_p} \approx \frac{i\omega\mu_0\sigma_i s^2}{4}\left[\frac{k_b s}{2\sqrt{4\eta^2+1}}\left(I_1(r_-)K_0(r_+)-I_0(r_-)K_1(r_+)\right)\right]_{\eta_i}^{\eta_{i+1}}, \tag{2.61}$$

Horizontal dipole

$$\frac{H_{i,s,PRP,V}}{H_p} \approx \frac{i\omega\mu_0\sigma_i}{4}xy\left[2I_{1/2}(r_-)K_{1/2}-\frac{1}{\sqrt{4\eta^2+1}}\exp\left(-k_b s\sqrt{4\eta^2+1}\right)\right]_{\eta_i}^{\eta_{i+1}} \tag{2.62}$$

$$\frac{H_{i,s,VCP}}{H_p} \approx \frac{i\omega\mu_0\sigma_i}{4}s^2\left[\frac{y^2-x^2}{s^2}2I_{1/2}(r_-)K_{1/2}-\frac{y^2}{s^2\sqrt{4\eta^2+1}}\exp\left(-k_b s\sqrt{4\eta^2+1}\right)\right]_{\eta_i}^{\eta_{i+1}}, \tag{2.63}$$

$$\frac{H_{i,s,PRP,H}}{H_p} \approx -\frac{i\omega\mu_0\sigma_i}{4}s^2\left[\frac{k_b y}{2\sqrt{4\eta^2+1}}\left(I_1(r_-)K_0(r_+)-I_0(r_-)K_1(r_+)\right)\right]_{\eta_i}^{\eta_{i+1}}, \tag{2.64}$$

where

$$r_{i,\pm} = \frac{k_b s}{2}\left(\sqrt{4\eta_i^2+1}\pm 2\eta\right) \tag{2.65}$$

and recall $\eta_i = h_i/s$.

2.8 Data acquisition at De Panne

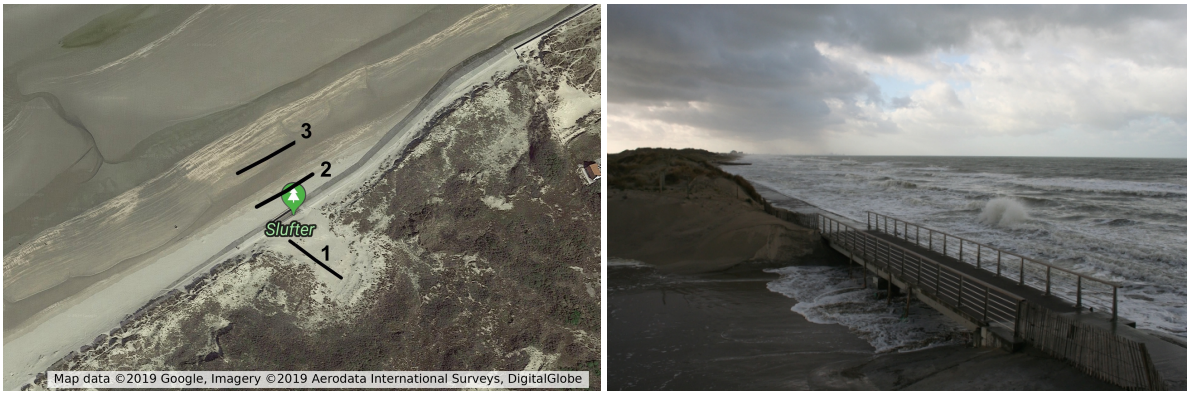
2.8.1 Location

In the context of this master's thesis, I was given the opportunity to collaborate in a measurement campaign. The data was obtained at a sea inlet[6] in the Flemish Nature Reserve "The Westhoek", situated along the French-Belgian border. The sea inlet (Figure 2.5b) was created in the fore dunes in 2004 [Hermans et al., 2012] [Verwaest et al., 2005], with the aim of promoting biodiversity, esp. salt tolerant flora. Therefore, the concrete dunefoot revetment was removed locally up to a hight such that sea water could approximately 12 times a year flood into the dunes. During a storm or high tide, the sea water would be able to access the dune slacks[7]. Nevertheless, the project is not running as planned. The sea inlet is completely silted up because water could not penetrate the sea inlet often enough. Consequently, the dune slack is also completely silted.

[6] in Dutch: "slufter", as indicated on Figure 2.5a by Google.

[7] Dune slacks are low-lying areas within the dunes that are occasionally flooded. [Grootjans et al., 1998] They can form infiltration ponds.

From a hydrogeological point of view, it remains an interesting area, because there is (only sporadically) salt water infiltration. Additional recharge of salt water can threaten the fresh water lens in the dune aquifer[8]. As already mentioned in Chapter 1, the area is used for drinking water production, whereby fresh water is pumped up from the fresh water lens. The hydrogeological equilibrium (fresh vs. salty water) should therefore be monitored. EMI measurements were carried out along the black lines in the sea inlet area, see Figure 2.5a.



(a) Map of sea inlet area. Measurements were carried ou along the black lines. Map retrieved from [Google, 2019].

(b) The sea inlet at high tide. Water floods towards the dune slack. Figure retrieved from [Copejans, 2007].

Figure 2.5: Site location at the sea inlet, situated in the Flemish Nature Reserve "The Westhoek".

2.8.2 Instrumentation

The EM34-3 by Geonics [McNeill, 1980] was used to measure the apparent conductivities (in mS/m) at different intercoil distances, according to the LIN approximation. It consists of a self-contained dipole transmitter and receiver, a reference (shielded) cable, the required power sources and consoles for the transmitter and receiver (see Figure 2.6). It is two-man portable. The instrumentation is composed of three standard intercoil spacings with a corresponding operating frequency. The alternating current will have a frequency of 6.4, 1.6 and 0.4 kHz for respective intercoil spacing of 10 m, 20 m or 40 m. The conductivity is measured in ranges of 10, 100 or 1000 mS/m with a $\pm 0.1\%$ resolution of the full scale. The measurement accuracy is $\pm 5\%$ at 20 mS/m and the noise level is 0.2 mS/m, which can be greater in regions of high power line interference. For that reason, there exists the EM34-3XL with larger transmitter and receiver coils. These larger coils improve the signal-to-noise ratio by a factor of ten at the 40 m spacing and a factor of four at the 10 m and 20 m spacings [Geonics, 2012]. This larger instrument, particularly useful in regions of high cultural noise, was not used. The maximal depth of exploration is about 60 m for an intercoil spacing of 40 m in a vertical dipole (HCP) coil setting.

[8]According to [Hubbard and Rubin, 2005]: "An aquifer is a body of rock that contains voids (such as pores, fissures or joints) and is thus capable of conducting groundwater."

Figure 2.6: Typical measurement set-up with the EM34-3. Figure retrieved from [Geonics, 2012].

The instrumentation is built to operate at three intercoil spacings. Of course, one can make measurements at numerous intercoil distances. The experimenter will need to manually write down the intercoil spacings, while the spacings in the logging device will correspond to the operating frequencies. This is how we have conducted the measurements.

2.8.3 Data

The magnetic field ratio's for measurements along line 1 in Figure 2.5a are shown in Figure 2.7. The raw data is tabulated in Appendix A, together with the data from lines 2 and 3. To get the magnetic field ratio's from the raw EM34-3 data, the following operations are required: The frequency is determined by the spacing that is set on the transmitter device. Spacing 10, 20 and 40 meters correspond respectively with 6.4, 1.3 and 0.4 kHz[9]. While the instrumentation measures the magnetic field, a reading of the apparent conductivity σ_a is given. In [McNeill, 1980], the conversion formula to the magnetic field ratio's is given. The formula is used for both HCP (vertical dipole) and VCP (horizontal dipole) setting:

$$\left(\frac{H_s}{H_p}\right)_V \approx \left(\frac{H_s}{H_p}\right)_H \approx \frac{i\omega\mu_0\sigma_a s^2}{4}. \tag{2.66}$$

[9]Indeed, a large intercoil distance is compensated by a small angular frequency, as required by the LIN condition.

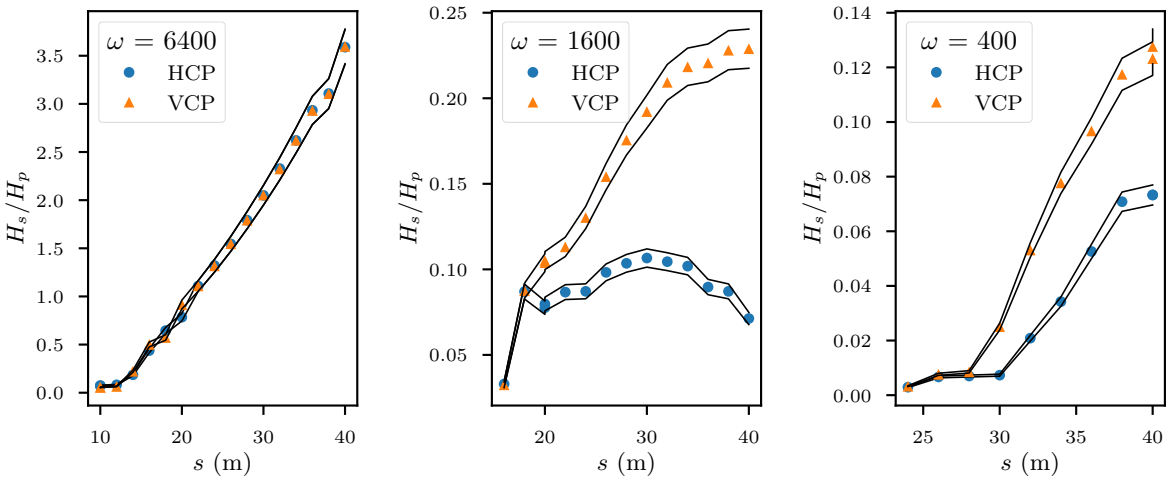


Figure 2.7: Magnetic field ratio's obtained along line 1 in Figure 2.5a. The solid lines represent the 5% accuracy level on the instrument readings (on the apparent conductivity σ_a).

Chapter 3

The Inverse Problem

3.1 Introduction

In the previous chapter, the *forward problem* was discussed in which the electromagnetic field response for a horizontally stratified earth, with known electric conductivities and thicknesses, to the field of an alternating magnetic dipole as source was examined. This problem is theoretically interesting and challenging, however, the inverse problem is of greater practical importance. An *inverse problem* is about finding the model parameters, given an input signal and its response (see Figure 3.1). Here, an alternating magnetic field is applied and a magnetic field is being measured. The question is, what is the structure of the subsurface, i.e. what are the conductivities and thicknesses of the different layers?

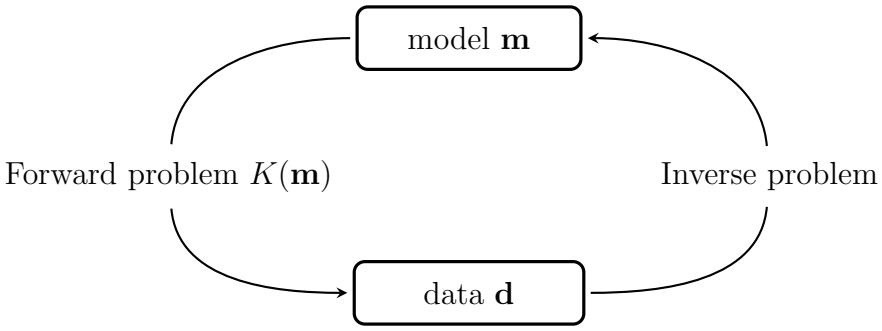


Figure 3.1: Forward vs. inverse problem.

More formally, the objective of an inverse problem is to find the model parameters $\mathbf{m} \in \mathbb{R}^{+,n}$ such that

$$\mathbf{d} = K(\mathbf{m}) \tag{3.1}$$

holds, where $\mathbf{d} \in \mathbb{C}^m$ is the vector that contains the measured data. K is the operator that describes the relation between $\mathbf{m}$ and $\mathbf{d}$. The forward models discussed in Chapter 2 are examples of such operators K. For a linear model, Eq. (3.1) reduces to

$$d = Km$$

where $K \in \mathbb{R}^{m \times n}$ is now a matrix. The LIN approximation is, after discretizing the continuous expression, an example of such a linear theory. The vectors $\mathbf{m}$ and $\mathbf{d}$ contain the

electrical conductivities and magnetic field ratios respectively.

The inverse problem is further elaborated in this chapter. In Section 3.2, we define an objective function that describes how well a vector $\mathbf{m}$ approximates the true model. The minimum of this function should corresponds to the best set of conductivities for a given data set $\mathbf{d}$. Finding such a minimum is not a simple task. We will rely on iterative methods that require gradient information from the objective function (see Section 3.3). In Section 3.4, we discuss the iterative method that will be used in our inversion scheme.

3.2 The regularized objective function

3.2.1 Least-squares fitting of an ill-posed problem

Since there is no direct technique for solving the non-linear inverse problem, we use least-squares fitting. An objective function $\phi(\mathbf{m})$ is defined, which describes how the predicted data $K(\mathbf{m})$ from the model parameters $\mathbf{m}$ deviates from the observed data $\mathbf{d}$. A standard functional for a parametrized model is

$$\phi(\mathbf{m}) = \frac{1}{2}\|\mathbf{d} - K(\mathbf{m})\|_2^2, \tag{3.2}$$

where the intuitive ℓ_2-norm measures the Euclidean distance between the predicted data $K(\mathbf{m})$ and the observed data $\mathbf{d}$.

A successful iterative inversion scheme requires fast and accurate computation of the forward model. For the purpose of fast computation, the LIN approximation was introduced, at the cost of a consistent error on the predicted data (cf. Section 2.6). Additionally, the accuracy of the computation depends on the stability of the implementation and the conditioning of the problem. Especially, if the inverse problem is ill-conditioned, then a slightly perturbed data set produces largely deviating model parameters. The general definition of the relative condition number of a function is [Trefethen and Bau III, 1997]

$$\kappa = \lim_{\delta\to 0} \sum_{||\delta x||\leq\delta} \left(\frac{||\delta f||}{||f(x)||} / \frac{||\delta x||}{||x||} \right), \tag{3.3}$$

where δx denotes a small perturbation on x and $\delta f = f(x + \delta x) - f(x)$. The condition number of a matrix is [Trefethen and Bau III, 1997]

$$\kappa(A) = ||A|| \cdot ||A^{-1}||. \tag{3.4}$$

For a rectangular matrix of full rank, the inverse A^{-1} is replaced by the pseudoinverse $A^+ = (A^*A)^{-1}A^*$. There is freedom to choose the norm, but for simplicity we take the ℓ_2-norm, because then[1] $||A||_2 = \sigma_{\max}$, the first singular value of matrix A. Here, the condition number $\kappa(A) = \sigma_{\max}/\sigma_{\min}$, where $\sigma_{\max}$ is the first singular value of matrix A and $\sigma_{\min}$ is the smallest singular value of A. A simulation of the condition number of the linear LIN matrix K reveals $\kappa \sim 10^{11}$, and consequently the inverse problem is ill conditioned. Measurement errors make the task even more demanding and solutions will differ greatly

[1] Theorem 5.3 in [Trefethen and Bau III, 1997]

for every remeasurement. We have thus three sources of noise: numerical round-off, unmodelled influences (e.g. the subsurface is not perfectly horizontally stratified) and noise due to instrument readings. The minimum of our least-squares functional (3.2) will generally not be zero:

$$\mathbf{d} = \mathbf{d^{true}} + \boldsymbol{\eta} \Rightarrow \phi(\mathbf{m}) = \|\mathbf{d^{true}} + \boldsymbol{\eta} - K(\mathbf{m}^{\text{true}}) = \|\boldsymbol{\eta}\|_2^2 > 0, \tag{3.5}$$

where $\boldsymbol{\eta}$ represents the experimental noise[2] and $\mathbf{d^{true}}$ is the exact magnetic field ratio for a given model $\mathbf{m}^{\text{true}}$.

Furthermore, the inverse problem is ill-posed. There are in our set-up usually more model parameters than measurement positions, i.e. one has more unknowns than knowns and hence the solution is non-unique. Finally, the objective function could be multimodal for non-linear least-squares. In Section 3.4, it will become clear that this exacerbates the optimization problem.

The non-uniqueness requires the problem to be reformulated for numerical treatment. The next example will demonstrate what is meant with such a reformulation. Consider two solutions $\mathbf{m_1}$ and $\mathbf{m_2}$ for which $\phi(\mathbf{m_1}) = \phi(\mathbf{m_2})$, i.e. the data misfits of both solutions are equal. For example, if $\mathbf{m_1}$ is strongly oscillatory and $\mathbf{m_2}$ is a constant solution, then the inversion scheme should prefer the constant solution, for these parameters are closer to realistic model parameters. Indeed, it is expected that there will be a set of layers with equal conductivities σ rather than spatially oscillatory behaviour of the conductivity profile. The task is thus to define, or reformulate, an objective function for which $\phi(\mathbf{m_1}) > \phi(\mathbf{m_2})$. Such an adjustment is made by imposing additional constraints on the model. Adding these terms to the objective function is a technique known as *regularization*. Various constraints can be imposed on the model, but in a general regularization problem, the structure or variation in the model parameters is minimized. That regularization can serve as a means of preventing overfitting due to noise, produced by unmodelled artefacts (e.g. a small metal object in the subsurface). Minimum structure regularization is a formalization of as Occam's razor idea that states that simpler solutions are more likely to be correct than complex ones.

In general, the objective function $\phi(\mathbf{m})$ is separated in two terms

$$\phi(\mathbf{m}) = \phi_d(\mathbf{m}) + \lambda\phi_m(\mathbf{m}), \tag{3.6}$$

where the ϕ_d is the data misfit term (i.e. the least-squares fitting) and ϕ_m is the model misfit or regularization term. In the next section, a traditional regularization method and the role of the regularization parameter λ will be discussed.

3.2.2 Tikhonov regularization

The traditional regularization method is Tikhonov regularization [Kaipio and Somersalo, 2006], where the the model misfit term contains an ℓ_2-norm and a smoothing operator. We define the objective function as

$$\phi(\mathbf{m}) = \phi_d(\mathbf{m}) + \lambda\phi_m(\mathbf{m}) = \frac{1}{2}\|\mathbf{d} - K(\mathbf{m})\|_2^2 + \lambda\|L_n\mathbf{m}\|_2^2 \tag{3.7}$$

[2]Noise due to unmodeled influences would yield a different operator K and is not considered here.

where L_n is a n-th order discrete differential operator which together with the ℓ_2-norm imposes continuity or better: smoothness. To make things more concrete, consider the following example: Consider a conductivity profile of two subsurface layers with respective conductivities $\sigma_1 = 0.1$ S/m and $\sigma_2 = 0.01$ S/m, the parametrization is such that $\mathbf{m} = (0.1, 0.1, 0.1, 0.1, 0.01, 0.01, 0.01, 0.01)^T$. If $n = 2$, then

$$||L_2\mathbf{m}|| = ||(0, 0, -0.09, 0.09, 0, 0)^T|| = 0.1273. \tag{3.8}$$

More complex structures yield larger derivatives and a larger ℓ_2-norm and hence a larger 'cost' in the objective function ϕ. By minimizing ϕ_m, a minimum structure model $\mathbf{m}$ will be a more probable outcome.

The parameter λ in Eq. (3.7) is the regularization parameter, it controls the relative contributions of both misfit functions ϕ_d and ϕ_m. There exist several strategies to choose regularization parameter λ. One can in an iterative optimization method, among other heuristic strategies, start with a relatively large regularization parameter, which strongly imposes minimum structure on the parameters $\mathbf{m}$. This yields a low resolution approximation of the parameters. By decreasing the parameter λ after several iterations, one allows the solution $\mathbf{m}$ to have more structure and details. Alternatively, more advanced techniques, such as the L-curve criterion, can, under some conditions, determine the optimal regularization parameter. In Chapter 5, several strategies will be employed and compared.

The regularization scheme in Eq. (3.7) can be useful in our inverse problem, because it improves the stability of our inversion. The reason behind this is that it smears out small-scale perturbations, which we do not expect in our model and consider as noise (measurement errors or artefacts in the subsurface). The disadvantage, however, is that for our model parameters, a certain blockiness is expected: for some neighbouring parameters in the same earth layer, the smoothness constraint is appropriate, but at the interface of two layers, an abrupt change in conductivity is expected. By imposing the smoothness constraint, high resolution is lost. In brief, Tikhonov regularization with a ℓ_2-norm in combination with a differential operator fails to simultaneously produce high resolution solutions and to be stable.

3.2.3 Sparsity based regularization

Seeing that the ℓ_2-norm has been deemed insufficient, other regularization terms need to be examined. There are many other regularization schemes that can be thought of. As discussed before, it is expected that the spatial derivative of our model $\mathbf{m}$ will have many zero entries. Vectors with many zero-entries are called *sparse* vectors. An S-sparse vector has S non-zero entries. A potentially successful regularization term would promote sparsity, because it favours sharp interfaces over smeared-out models. We will now examine which focusing functions promote sparsity.

For a regularization term with an ℓ_p-norm[3], all norms with $p \leq 1$ promote sparse solu-

[3]The ℓ_p-norm is defined as

$$||\mathbf{x}||_p = \sqrt[p]{\sum_i |x_i|^p}. \tag{3.9}$$

tions, but the ℓ_1-norm is the most attractive norm, because of its relatively easy numerical computation. A disadvantage is that the ℓ_1-norm is not differentiable in $x = 0$, which will be important in the next section, when the objective function will be minimized.

It is not obvious to understand why the ℓ_1-norm promotes sparse solutions. It is most easily illustrated with a linear underdertermined system with matrix A, a desired sparse solution $\mathbf{x}$ and a data vector $\mathbf{b}$. Consider in 2D

$$A = [0.1, 1], \quad \mathbf{x} = (x_1, x_2)^T, \quad b = 1, \tag{3.10}$$

then we have an infinite number of solutions $x_2 = 1 - 0.1x_1$, as in Figure 3.2. As an additional requirement, sparse solutions are promoted, which is formalized as the minimum ℓ_1 norm solutions. In 2D, 1-sparse solutions correspond with solutions on the axes. Indeed, $\mathbf{x^a} = (0, 1)$ and $\mathbf{x^b} = (10, 0)$ are both 1-sparse solutions of our simple 2D problem. $\mathbf{x^a} = (0, 1)$ has smaller ℓ_1-norm and our minimization problem will promote this solution[4]. We can understand graphically that no other non-sparse solutions exist with smaller norm by plotting the ℓ_1-norm of $\mathbf{x^a}$ in Figure 3.2.

It is important to note that not all 1-sparse vectors will be favoured above all other solutions. Consider both solutions $\mathbf{x^b} = (10, 0)$ and $\mathbf{x^c} = (1, 0.9)$, where $\|\mathbf{x^b}\|_1 = 10$ and $\|\mathbf{x^c}\|_1 = 1.9$, for which the non-sparse solution $\mathbf{x^c}$ will be promoted over $\mathbf{x^b}$. This could be the outcome when the data misfit terms ϕ_d and regularization parameter λ are such that

$$[\phi(\mathbf{x^a}) = \phi_d(\mathbf{x^a}) + \lambda] > \left[\phi(\mathbf{x^b}) = \phi_d(\mathbf{x^b}) + 10\lambda\right] > [\phi(\mathbf{x^c}) = \phi_d(\mathbf{x^c}) + 1.9\lambda] \,.$$

A potential problem with ℓ_1-norm minimization is that the slope of the locus of solutions of our underdetermined problem is at an angle of $\pm\pi/4$. The ℓ_1-norm of infinitely many (including non-sparse) solutions coincides with the minimum ℓ_1-norm diamond shape. The $\ell_{1/2}$ norm, shown in Figure 3.3, resolves this issue and it always promotes sparse solutions, however, the norm is computationally more expensive[5]. Luckily, this $\pm\pi/4$-outcome is less probable in higher dimensions[6] and therefore we assume that it will not occur. From Figure 3.3, it becomes clear that the ℓ_2-norm does not promote sparse solutions: it minimizes the Euclidean length of the solution $\mathbf{x}$.

The objective function is now

$$\phi(\mathbf{m}) = \phi_d(\mathbf{m}) + \lambda\phi_m(\mathbf{m}) = \frac{1}{2}\|\mathbf{d} - K(\mathbf{m})\|_2^2 + \lambda\|L_n\mathbf{m}\|_1. \tag{3.11}$$

[4]Because that corresponds with the minimum of the objective function. Recall that $\phi = \phi_m + \lambda\phi_d$ and that λ determines the trade-off. The promotion of sparse solutions does not mean that the solution is guaranteed to be sparse.

[5]The ℓ_p-norm for $0 \leq p < 1$ is non-convex. It is known that non-convex optimization is strongly NP-hard or computationally difficult to solve exactly. (see [Ge et al., 2011]) and [Ramirez et al., 2013]). We don't delve in the issues of computational complexity theory.

[6]A higher dimensional set of solutions need to have a slope of $\pm\pi/4$ along every axis, before this issue occurs.

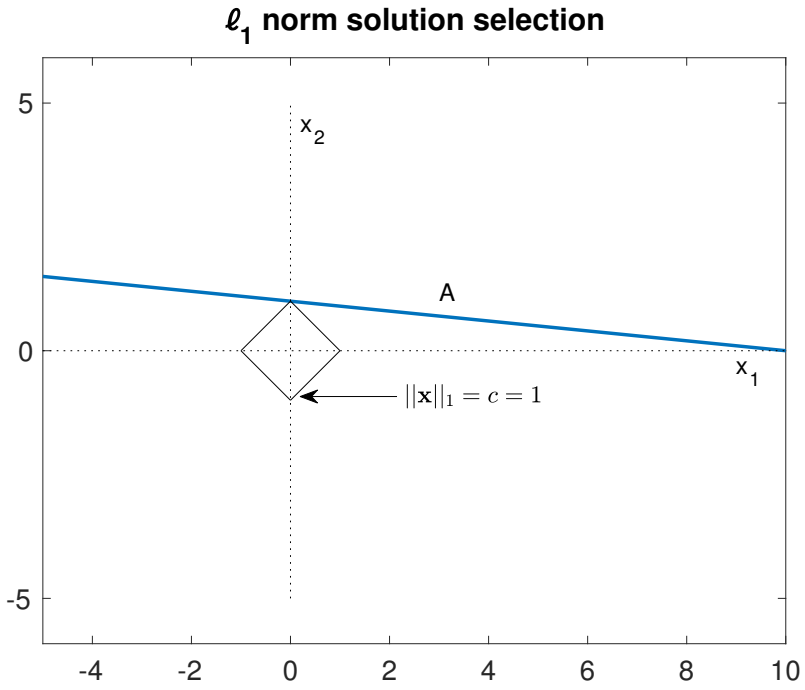


Figure 3.2: Sparse solution selection of an underdetermined problem, via ℓ_1-norm minimization. The diamond represents the set of pairs (x_1, x_2) for which the ℓ_1-norm equals the constant $c = 1$.

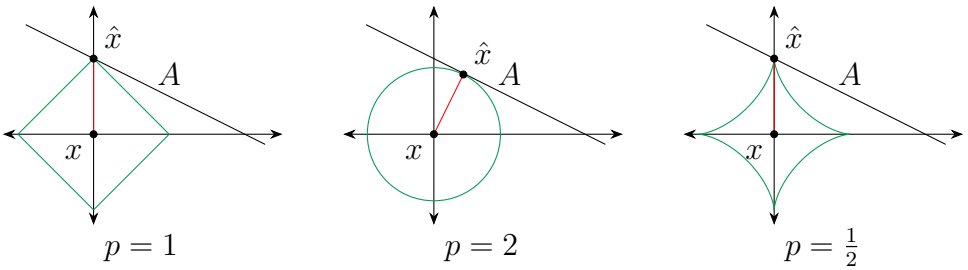


Figure 3.3: Geometric interpretations of ℓ_p-norms

Norms are not the only measures that can be considered in regularization schemes. Other measures, which will be called focusing functions $\mu(v_j)$, map weights to every entry of a vector $\mathbf{v}$. The model misfit ϕ_m is then the sum of these weights:

$$\phi_m(\mathbf{m}) = \sum_j \mu(v_j), \tag{3.12}$$

where the vector $\mathbf{v}$ is e.g. the spatial derivative of the model $\mathbf{m}$.

Focusing functions can be designed to promote sparsity. That is, in general, the case when large values are mapped on large values and zeros mapped on the minimum of the focusing function. For small values, we require $\mu(v_j) > \mu(0)$. There are innumerable measures that can be constructed with this property and therefore focusing inversion seems to provide additional freedom in finding a perfect measure for the regularization term ϕ_m. However, that freedom must be balanced against the fact that these functions often depend on additional parameters, the question then remains to find the optimal parameter.

In geophysical inverse problems, the M-measure of Huber [Huber et al., 1964] is often considered ([Farquharson, 2007], [Ha et al., 2009]). The function

$$\mu_{\delta,\text{Huber}}(x) = \begin{cases} \frac{1}{2}x^2 & \text{for } |x| \leq \delta, \\ \delta(|x| - \frac{1}{2}\delta) & \text{otherwise} \end{cases} \tag{3.13}$$

is quadratic for small values of x and linear for large values of x (see Figure 3.4a). The M-measure of Huber is differentiable, the slope at $|x| = \delta$ is equal for both sections. This focusing function corresponds to to taking a ℓ_1-norm when the residuals are large and half of the ℓ_2-norm when residuals are small. [Guitton and Symes, 2003] show that for several types of noise in seismic applications, more robust results can be obtained via the M-measure of Huber than with the ℓ_2-norm. The Huber measure is also known from statistics, esp. in robust regression [Vogel, 2002].

[Portniaguine and Zhdanov, 1999] use the following measure, the minimum support functional,

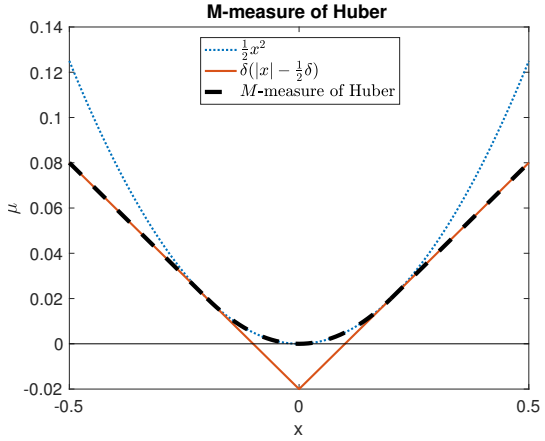
$$\mu_{\text{Zdh}}(x) = \frac{x^2}{x^2 + \epsilon^2}, \tag{3.14}$$

where ϵ is a parameter that we can choose (see Figure 3.4b). This measure, for ϵ small enough, is proportional to the number of non-zero elements in $\mathbf{x}$. [Portniaguine and Zhdanov, 1999] have used this measure in 3D inversion, where the total volume within the model $\mathbf{m}$ for which the gradient is non-zero was minimized.

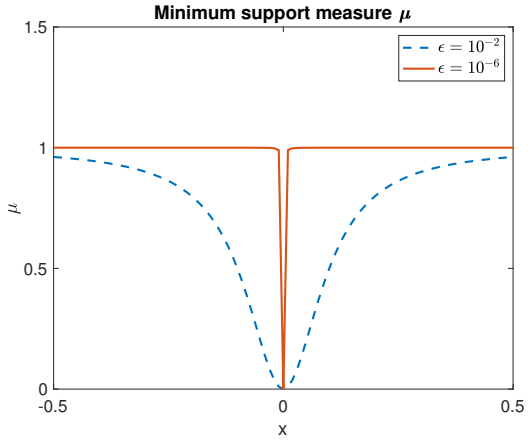
Fast optimization techniques usually use gradient information from the objective function ϕ. The ℓ_1-norm is not differentiable in $x = 0$ and therefore the perturbed ℓ_p-norm measure of Ekblom [Ekblom, 1987] for $p = 1$ often replaces the ℓ_1-norm. The measure

$$\mu_{\text{Ekblom}}(x) = \left(x^2 + \epsilon\right)^{p/2}, \tag{3.15}$$

where ϵ is a small number, is shown in Figure 3.5.



(a) The M-measure of Huber for $\delta = 0.2$ [Huber et al., 1964]



(b) The minimum support measure [Portniaguine and Zhdanov, 1999].

Figure 3.4: Focusing functions

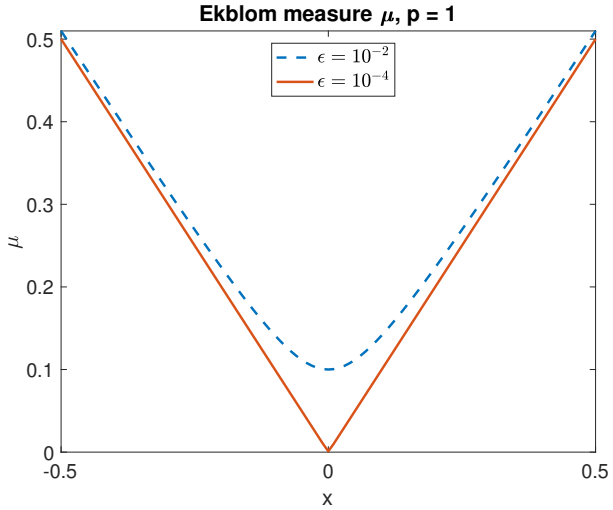


Figure 3.5: The perturbed ℓ_p-norm measure of Ekblom [Ekblom, 1987].

One can invent innumerable measures for sparsity based inversion. Even machine learning can be used to find better regularization functionals. [Haber and Tenorio, 2003] use supervised learning techniques that determine a regularization functional, based on a realistic training set.

3.2.4 Reformulation in sparse basis

There are other strategies for a minimum structure inversion than imposing sparsity on the spatial derivative on the model $\mathbf{m}$. Suppose that there exists a basis in which the true model parameters $\mathbf{m}$, known to have minimum structure, are represented in a sparse form. Then,

$$\mathbf{x} = W\mathbf{m}, \tag{3.16}$$

where W is the basis transformation. The objective function ϕ in terms of the sparse basis is

$$\boxed{\phi(\mathbf{x}) = \phi_d(\mathbf{x}) + \lambda\phi_m(\mathbf{x}) = \frac{1}{2}\|\mathbf{d} - K(W^{-1}\mathbf{x})\|_2^2 + \lambda\|\mathbf{x}\|_1,} \tag{3.17}$$

where ϕ_m is the new regularization term.

In summary, the objective function (3.17) has a data misfit term ϕ_d and a model misfit term ϕ_m. The data misfit term describes the error on a proposed model $\mathbf{x}$ in some basis, while the regularization term ϕ_m imposes minimum structure on the model $\mathbf{m}$, by imposing sparsity via the ℓ_1-norm on its transform $\mathbf{x}$. The trade off between the two terms is governed by the regularization parameter λ.

The choice of the basis, in which a minimum-structure model $\mathbf{m}$ is represented in a sparse fashion, is crucial and is the topic of next chapter. Indeed, that basis preferably represents a strongly oscillatory model $\mathbf{m}$ in a *non*-sparse fashion. It remains to find the minimum of the objective function, which we will discuss next.

3.3 The optimization problem

In the previous section, we have proposed an objective function, for which finding the minimum remains a challenging task. Fast optimization techniques exist (see Section 3.4), but they make use of the objective function's gradient. We will examine this further in this section and we will introduce the relevant optimization problem formulations.

The minimum of the objective function is at least an approximate solution of the real model parameters $\mathbf{m}$, or $\mathbf{x}$ in some sparse basis. It remains to select the best elements that minimize that function ϕ. Many such selection methods exists. In the field of mathematical optimization, the most general problem is the following:

$$\begin{aligned} \text{minimize} \quad & \phi(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) = 0, i = 1, \cdots, \ell \\ & h_i(\mathbf{x}) < 0, i = 1, \cdots, \ell \end{aligned} \tag{3.18}$$

where ϕ is the objective function, $\mathbf{x}$ the solution and h_i, g_i are equality and inequality constraints respectively. In the remainder of the text, 'subject to' will often be abbreviated as 's.t.'.

An optimization method or algorithm will iteratively find improved estimates of $\mathbf{x}$, starting from an initial guess $\mathbf{x}^{(0)}$. The iterate at iteration k is denoted with $\mathbf{x}^{(k)}$. A stop criterion will be defined such that the algorithm terminates at a (hopefully) good approximate solution. The algorithm that will used is described in Section 3.4.

3.3.1 The minimization problem

In optimization theory, our type of objective function is known as the LASSO-problem (Least Absolute Shrinkage and Selection Operator) or BPDN (Basis Pursuit Denoising). There are typically two strategies that try to resolve the differentiability issue of the ℓ_1-norm in $x = 0$. One strategy transforms the problem to a linear program and the other replaces the ℓ_1-norm with a differentiable measure. In this section, we examine how the general problem formulation Eq. (3.18) can be reformulated.

There are two special cases of Eq. (3.18): The Linear Program (LP) and the Quadratic Program (QP). In a *linear program*, the objective function and the constraints are linear

$$\begin{aligned} \text{minimize} \quad & \mathbf{a}^T\mathbf{x} \\ \text{s.t.} \quad & A\mathbf{x} = \mathbf{b} \\ & C\mathbf{x} \leq \mathbf{c}. \end{aligned} \tag{3.19}$$

The minimization problem Eq. (3.17) from Section 3.2.4 can be written as

$$\begin{aligned} \text{minimize} \quad & ||\mathbf{x}||_1 \\ \text{s.t.} \quad & KW^{-1}\mathbf{x} = \mathbf{d}, \end{aligned} \tag{3.20}$$

under the assumptions that $m < n$ (i.e. there are more model parameters $\mathbf{m}$ than measurements $\mathbf{d}$), K is linear (i.e. LIN approx.) and that $\mathbf{x}$ would exactly yield $\mathbf{d}$. In Eq. (3.5) from Section 3.2.1, we have argued that noise, originating from various sources, yields that $KW^{-1}\mathbf{x}^{\text{true}} \neq \mathbf{d}^{\text{true}}$ and esp. the restriction to use only the LIN approximation amplifies the error.

Keeping the assumptions in mind, the constrained, non-smooth minimization problem (3.20) is equivalent to the constrained, smooth linear program, by writing $\mathbf{x} = \mathbf{p} - \mathbf{w}$ via the following construction for all i:

$$\begin{cases} p_i = x_i & \text{if } x_i \geq 0 \\ w_i = -x_i & \text{if } x_i \leq 0 \end{cases}. \tag{3.21}$$

Then,

$$\|\mathbf{x}\|_1 = \sum_i |x_i| = \sum_i (p_i + w_i), \tag{3.22}$$

and hence,

$$\begin{aligned} \text{minimize} \quad & \sum_{i=1}(p_i + w_i) \\ \text{s.t.} \quad & KW^{-1}(\mathbf{p} - \mathbf{w}) = \mathbf{d} \\ & \mathbf{p}, \mathbf{w} \geq 0 \end{aligned} \tag{3.23}$$

The variables $\mathbf{p}, \mathbf{w}$ in the objective function and in the inequality constraints are called *slack variables*. Using slack variables is a standard method to deal with objective functions ϕ with an ℓ_1-norm. Linear programs can be solved easily [Nocedal and Wright, 2006], however we turn our focus to a slightly more general optimization problem that can handle noise.

In a *quadratic program* (QP), the objective function is non-linear and is written as

$$\begin{aligned} \text{minimize} \quad & \mathbf{a}^T\mathbf{x} + \frac{1}{2}\mathbf{x}^T H\mathbf{x} \\ \text{s.t.} \quad & A\mathbf{x} = \mathbf{b} \\ & C\mathbf{x} \leq \mathbf{c}. \end{aligned} \tag{3.24}$$

where $H \in \mathbb{R}^{n\times n}$ is a symmetric matrix. It turns out that QPs with a positive semidefinite matrix[7] H can be solved reliably to their global optimality [Patrinos, 2018]. As with the linear program, the minimization problem (3.17) can be transformed to a QP. While linear problems are easier to solve, it is better to use the quadratic variant. The latter allows to minimize the error on the solution and is thus able to deal with measurement error and unmodelled effects. The objective function expressed in (3.17) is equivalent to the following quadratic problem

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}||K\left(W(\mathbf{p} + \mathbf{w})\right) - \mathbf{b}||_2^2 + \lambda \sum_i (p_i + w_i) \\ \text{s.t.} \quad & \mathbf{p}, \mathbf{w} \geq 0 \end{aligned} \tag{3.25}$$

Alternatively, the non-differentiability issue of the ℓ_1-norm in $x = 0$ is resolved by replacing the ℓ_1-norm with other, differentiable measures, such as the Ekblom measure and Huber measure μ.

$$||\mathbf{x}||_1 = \sum_{i=1}^{n} \mu(x_i) \tag{3.26}$$

In conclusion, the problem with the non-differentiability of the ℓ_1-norm can be tackled by reformulating the problem using slack variables or by replacing the ℓ_1-norm with another differentiable focusing function.

3.3.2 Convex optimization

Convexity plays an important role in optimization, because convex optimization problems exhibit great properties (see below). Definitions [Nocedal and Wright, 2006] related to sets, functions and optimization problems are listed. First, we give a definition af a positive semidefinite matrix [Horn et al., 1990]:

[7] See Definition 1 in Section 3.3.2.

Definition 1 *(positive (semi)definite matrices) A symmetric matrix* $A \in \mathbb{R}^{n\times n}$ *is* positive definite *(*$A \succ 0$*) if*

$$\mathbf{x}^T A\mathbf{x} > 0, \quad \forall \mathbf{x} \in \mathbb{C}^n \backslash \{0\}$$

and positive semidefinite *(*$A \succeq 0$*) if*

$$\mathbf{x}^T A\mathbf{x} \geq 0, \quad \forall \mathbf{x} \in \mathbb{C}^n \backslash \{0\}.$$

The following theorem[8] serves as a practical method to check if a matrix is positive semidefinite or not.

Theorem 1 *(Symmetric positive semidefinite matrices) Let* A *be a symmetric matrix. The following statements are equivalent*

1. A *is positive semidefinite (*$A \succeq 0$*)*
2. *All eigenvalues of* A *are non-negative*
3. *There exists a factorization* $A = B^T B$*, where* $A \in \mathbb{R}^{n\times r}$ *and* r *is the rank*[9] *of* A*.*

Definition 2 *(convex set) A set* $S \in \mathbb{R}^n$ *is convex if a straight line segment connecting any two points in* S *lies entirely inside* S*.*

An example of a convex set is a unit ball.

Definition 3 *(convex function) The function* f *is a convex function if*

1. *its domain* S *is a convex set;*
2. *for any two points* x *and* y *in* S*.*

For convex functions, the following property is satisfied:

$$f(ax + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y), \quad \text{for all} \quad \alpha \in [0,1]. \tag{3.27}$$

A linear function $f(\mathbf{x}) = \mathbf{c}^T\mathbf{x} + \alpha$ is an example of a convex function. A quadratic function $f(\mathbf{x}) = \mathbf{x}^T H\mathbf{x}$, where H is symmetric and positive definite is also a convex function. The convexity of a function means that the linear interpolation between two arbitrary points on that function is never below the graph.

Definition 4 *An optimization problem of general form*

$$\begin{aligned} &\textit{minimize} \quad \phi(\mathbf{x}) \\ &\textit{s.t.} \quad \mathbf{x} \in X \end{aligned} \tag{3.28}$$

is convex when ϕ *is a convex function and* $X \subseteq \mathbf{R}^n$ *is a convex set.*

The following theorem explains why convex problems are relatively easy to solve, the proof can be found in Chapter 2 of [Nocedal and Wright, 2006].

[8] Observation 7.1.4 and Theorem 7.2.7 in [Horn et al., 1990]

[9] The **rank** of a matrix A is the dimension of the range of A, a space spanned by the columns of A.

Theorem 2 *(Local minima of convex optimization) When f is convex, any local minimum x^* is a global minimum of f. If in addition f is differentiable, then any stationary point x^* is a global minimum of f.*

However, for many optimization problems the objective function is not convex. Applying convex optimization algorithms yield local minima. For even more (exotic) functions, no local minimum will be found. The quadratic problem (3.25) is a convex problem, only if $H = (K(W\mathbf{x}))^T K(W\mathbf{x})$ is positive semidefinite. The more general problem ϕ, where the ℓ_1-norm is replaced by a measure μ, can be non-convex if H is not positive semidefinite *or* if μ is non-convex. Indeed, the ℓ_1-norm is convex, but e.g. the minimum support measure μ_{Zhd} (Eq. (3.14)) is not.

3.3.3 The sensitivity matrix

In geophysics, the *sensitivity matrix*[10] is basically the derivative of a model response with regard to the subsurface conductivities [Christensen, 2014]. The sensitivity matrix coincides with the Jacobian, as will become clear in this section. This jargon forces us to think more about changes in the parameters that propagate through the system and yield a different output. The sensitivities are useful when one tries to understand the importance of a parameter in a simulation. When a sensitivity of some parameter is large, compared to the other parameters, then this implies that changes in the parameter significantly change the output.

First consider the non-linear forward problem with a slightly different notation

$$\mathbf{d} = \mathcal{K}(\mathbf{m}) = \begin{bmatrix} K_1(\mathbf{m}) \\ K_2(\mathbf{m}) \\ \vdots \\ K_n(\mathbf{m}) \end{bmatrix}, \tag{3.29}$$

where $\mathcal{K}$ is a multivariate vector function that yields the discrete data set $\mathbf{d}$ from its discrete model parameter $\mathbf{m}$. The non-linear relation can be written via Taylor series as

$$\mathcal{K}(\mathbf{m} + \epsilon\mathbf{p}) = \mathcal{K}(\mathbf{m}) + \epsilon \sum_{i,j} \underbrace{\frac{\partial K_i(m_j)}{\partial m_j}}_{J_{ij}} p_j + \mathcal{O}(\epsilon^2) \tag{3.30}$$

$$= \mathcal{K}(\mathbf{m}) + \epsilon J\mathbf{p} + \mathcal{O}(\epsilon^2). \tag{3.31}$$

For small ϵ, the non-linear relation can be approximated by a linear relation and it provides insight in the local behaviour in the forward problem. This linear relation is used to define the sensitivity matrix (or Jacobian)

$$J = \frac{\partial \mathbf{d}}{\partial \mathbf{m}} \quad \text{or} \quad J_{ij} = \frac{\partial d_i}{\partial m_j} \quad \text{or} \quad J = \nabla K(\mathbf{m}). \tag{3.32}$$

[10] In the language of linear operators and functionals (esp. in the literature by [Zhdanov, 2015]), 'Fréchet derivative' is used instead of sensitivity matrix. When $\mathcal{K}$ is a differentiable operator, $\delta\mathcal{K}(\mathbf{m}) = F_m \delta\mathbf{m}$, where the linear operator F_m is the Fréchet derivative.

More generally, the forward problem can be written as follows:

$$\mathcal{I}(\mathbf{d}, \mathbf{m}) = 0 \tag{3.33}$$

and taking the derivative w.r.t. $\mathbf{m}$ yields a following expression:

$$\nabla_m \mathcal{I}(\mathbf{d}, \mathbf{m}) + \nabla_d \mathcal{I}(\mathbf{d}, \mathbf{m}) \frac{\delta \mathbf{d}}{\delta \mathbf{m}} = 0 \tag{3.34}$$

$$\Rightarrow \quad \boxed{J = -(\nabla_d \mathcal{I}(\mathbf{d}, \mathbf{m}))^{-1} \nabla_m \mathcal{I}(\mathbf{d}, \mathbf{m})} \tag{3.35}$$

where the $\nabla_d \mathcal{I}(\mathbf{d}, \mathbf{m})$ is assumed to be invertible. In geophysics, this formula is known as the fundamental sensitivity equation [Haber, 2014]. Let us see if we can recover the sensitivity matrix that we have obtained via the Taylor expansion in Eq. (3.32):

$$\mathcal{I}(\mathbf{d}, \mathbf{m}) = K(\mathbf{m}) - \mathbf{d} = 0, \tag{3.36}$$

$$\nabla_d \mathcal{I}(\mathbf{d}, \mathbf{m}) = -\mathbb{I}, \tag{3.37}$$

$$\nabla_m \mathcal{I}(\mathbf{d}, \mathbf{m}) = \nabla K(\mathbf{m}), \tag{3.38}$$

$$\Rightarrow J = -(-\mathbb{I})^{-1} \nabla K(\mathbf{m}) = \nabla K(\mathbf{m}). \tag{3.39}$$

The sensitivities are not only relevant for solving inverse problems. They also provide us insight in the forward problem [Haber, 2014]. The singular value decomposition (SVD) allows us to analyse the sensitivities. The sensitivity matrix $J \in \mathbb{R}^{m \times n}$ is decomposed in

$$J = U \Sigma V^T = \sum_{i=1}^{m} \sigma_i \mathbf{u}_i \mathbf{v}_i^T, \tag{3.40}$$

where $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ are both orthogonal matrices. $\Sigma = \text{diag}(\sigma_1, \cdots, \sigma_m)$ and σ_i is a singular value for which $\sigma_1 \leq \sigma_2 \cdots \leq \sigma_m$. When the model is perturbed $\mathbf{m} + \mathbf{w}$, we write

$$\mathbf{w} = \sum_{i=1}^{m} \alpha_i \mathbf{v}_i + \mathbf{w}^{\text{orth}} = V \boldsymbol{\alpha} + \mathbf{w}^{\text{orth}}. \tag{3.41}$$

Indeed,

$$J \mathbf{w}^{\text{orth}} = U \Sigma V^T \mathbf{w}^{\text{orth}} = 0, \tag{3.42}$$

from which one concludes that a perturbation of $\mathbf{m}$ in the direction of $\mathbf{w}^{\text{orth}}$ does not change the the data vector $\mathbf{d}$. On the other hand, a perturbation with a singular vector that corresponds with a large singular value, yields a significant change in the data. Indeed, we can now identify the parameters in the forward problem that are important.

Consider again the conductivity profile of Chapter 2 (Figure 2.3) from borehole logging in Liège. Let us illustrate the linearity of the LIN approximation and the non-linearity of the damped and exact model. The non-linearity is apparent from the equations in Chapter 2, but a visual inspection will help us to understand that a local linearisation (see next section) of the forward models will be justified. The parametrization is such that there are 40 layers with thicknesses h_i of 0.2 meters. For simplicity, the response is shown for measurements at an intercoil distance s of 20 meters and at a height h_0 of 0.1 meters, along the $\mathbf{m}_{40}$ axis of the model space. This semi-infinite layer is at a depth of 7.8 meters

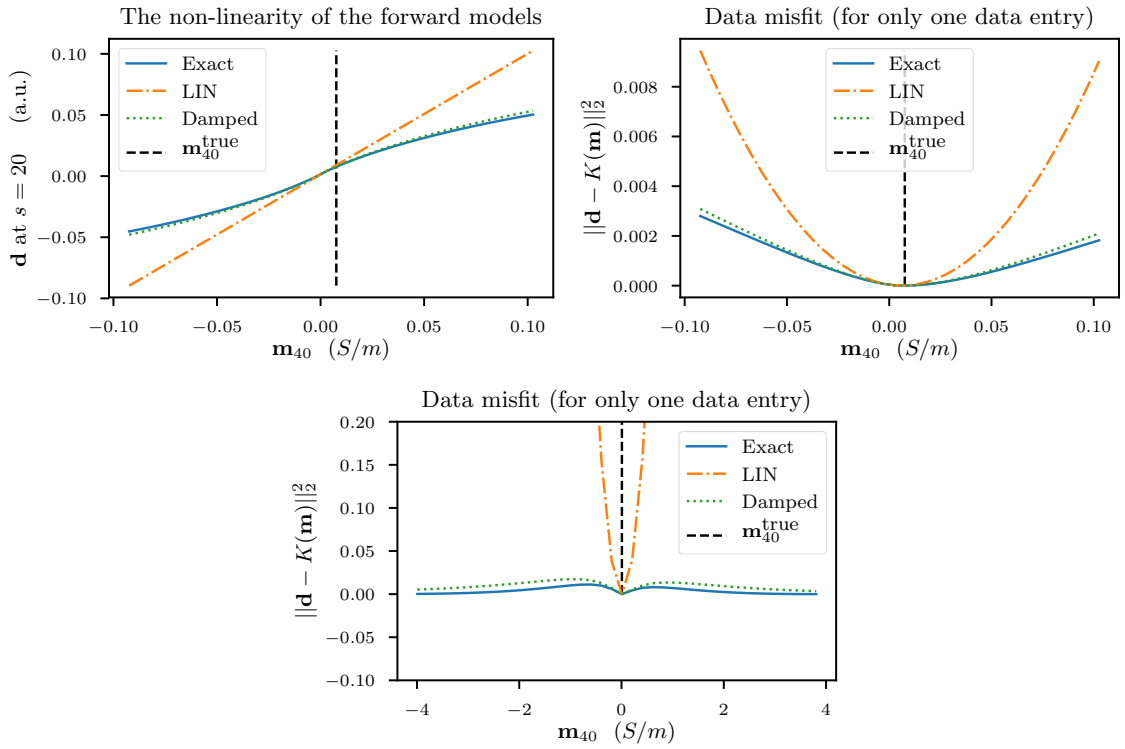


Figure 3.6: Behaviour of the response around the true conductivity of the 40-th layer.

and therefore it contributes significantly to the response. This is also seen in the singular value decomposition of the sensitivity matrix. Let us consider the SVD-decomposition of the sensitivity matrix of the damped model. The singular vector, corresponding with the largest singular value has the in absolute value largest value for the 40-th entry. This is indeed in agreement with what physically is expected. The 40-the entry is -0.989 and much larger than the other entries with values around -0.025. In Figure 3.6, the behaviour of the the response around the true electrical conductivity $\sigma_{40} = 0,0076$ S/m is shown. This confirms that the LIN approximation is linear in the conductivity $\sigma(h)$ and thus the data misfit functional ϕ_d will be quadratic and convex. The behaviour of the other models is non-linear and the data misfit functional deviates from the quadratic behaviour. Within the physical ranges[11], the data misfit functional seems to have a unique minimum. For in absolute value larger (unphysical) conductivities, a clear non-convex behaviour is observed. We will need to keep this in mind. In choosing the optimization algorithm, we opt an algorithm that allows to set bounds in the model space. By doing so, we would be able (if necessary) to constrain the model parameters to the range where the the data misfit exhibits one minimum. There is no guarantee that there is only one minimum within such a range. Let us stress that this was only an illustration, where only the $\mathbf{m}_{40}$ -dimension was considered.

3.3.4 (Non-)linear least squares

In this section, the gradient of the data misfit functional ϕ_d is obtained for both the linear and non-linear least squares term. The residual vector $\mathbf{r} = K(W^{-1}\mathbf{x}) - \mathbf{d} \in \mathbb{C}^m$ is usually

[11] positive electrical conductivities and smaller than 1 S/m

introduced in this context.

$$\nabla\phi_d(\mathbf{x}) = \nabla\frac{1}{2}\|K(W^{-1}\mathbf{x}) - \mathbf{d}\|_2^2 \tag{3.43}$$
$$= \frac{1}{2}\nabla\left(\mathbf{r}^T\mathbf{r}\right) \tag{3.44}$$
$$= \mathbf{r}^T\nabla\mathbf{r} \tag{3.45}$$
$$= \mathbf{r}^T\nabla K(W^{-1}\mathbf{x}) \tag{3.46}$$

or componentwise

$$\nabla\phi_d(\mathbf{x}) = \sum_{j=1}^{n}\frac{\partial\phi_d(\mathbf{x})}{\partial x_j} = \sum_{j=1}^{n}\frac{\partial}{\partial x_j}\frac{1}{2}\left(\sum_{i=1}^{m} r_i(\mathbf{x})r_i(\mathbf{x})\right) = \sum_{j=1}^{n}\sum_{i=1}^{m} r_i(\mathbf{x})\frac{\partial r_i(\mathbf{x})}{\partial x_j}. \tag{3.47}$$

Analogously, the Hessian[12] of ϕ_d can be obtained via

$$\nabla^2\phi_d(\mathbf{x}) = \sum_{k=1}^{n}\sum_{j=1}^{n}\sum_{i=1}^{m}\frac{\partial r_i(\mathbf{x})}{\partial x_k}\frac{\partial r_i(\mathbf{x})}{\partial x_j} + \sum_{k=1}^{n}\sum_{j=1}^{n}\sum_{i=1}^{m} r_i(\mathbf{x})\frac{\partial^2 r_i(\mathbf{x})}{\partial x_j\partial x_k} \tag{3.48}$$
$$= \sum_{i=1}^{m}\nabla r_i(\mathbf{x})\nabla r_i(\mathbf{x})^T + \sum_{i=1}^{m} r_i(\mathbf{x})\nabla^2 r_i(\mathbf{x}). \tag{3.49}$$

Expressing the gradient and Hessian in terms of the sensitivity matrix,

$$J = \begin{bmatrix} \nabla r_1(\mathbf{x})^T \\ \nabla r_2(\mathbf{x})^T \\ \vdots \\ \nabla r_m(\mathbf{x})^T \end{bmatrix}, \tag{3.50}$$

yields

$$\nabla\phi_d(\mathbf{x}) = J^T\mathbf{r} \tag{3.51}$$
$$\nabla^2\phi_d(\mathbf{x}) = J^TJ + \sum_{i=1}^{m} r_i(\mathbf{x})\nabla^2 r_i(\mathbf{x}) \tag{3.52}$$

Linear Case
In the case of a linear-least squares functional[13] (i.e. $K(W^{-1}\mathbf{x}) = KW^{-1}\mathbf{x}$), the gradient can simply be obtained from Eq. (3.46) via

$$\nabla\phi_d(\mathbf{x}) = \mathbf{r}^T\nabla\mathbf{r} \tag{3.53}$$
$$= \mathbf{r}^T\nabla KW^{-1}\mathbf{x} \tag{3.54}$$
$$= \mathbf{r}^TKW^{-1} \tag{3.55}$$
$$= (KW^{-1})^T\mathbf{r}, \tag{3.56}$$

[12]In optimization literature (such as [Nocedal and Wright, 2006]), the Hessian refers to the derivative of the gradient.

[13]Linear least squares means that the residual vector $\mathbf{r}$ is linear. Indeed, K and W are linear matrix. They do not depend on $\mathbf{x}$.

where we recognize the sensitivity matrix $J = (KW^{-1})$.

Writing componentwise makes the linearity more manifest (it will clearly demonstrate the meaning of a non-linearity). First note,

$$r_i(\mathbf{x}) = (KW^{-1}\mathbf{x}-\mathbf{d})_i = \sum_k K_{ik}(W^{-1}\mathbf{x})_k - d_i = \sum_k \sum_l K_{ik}W^{-1}_{kl}x_l - d_i \equiv \mathbf{c}^T\mathbf{x} - d_i, \quad (3.57)$$

a manifestly linear function, where $\mathbf{c}$ is simply a vector with scalars.

Non-linear Case
In the non-linear case, the Hessian in Eq. (3.52) has two terms. The first term is a symmetric and positive semidefinite matrix. The second term depends on the curvature of the objective function ϕ, indeed, it can make the Hessian negative definite. However, the second term will be small for problems with a small residual at the solution or for problems that are not very non-linear [Haber, 2014]. It is common to apply the Gauss-Newton approximation to such a non-linear problem, which linearises the residual vector $\mathbf{r}$ and thus linearises the forward model. The Gauss-Newton approximation drops the second (possible negative definite) term in the Hessian.

We can linearise the forward model via Taylor

$$\mathbf{r}(\mathbf{x} + \Delta\mathbf{x}) \approx \mathbf{r}(\mathbf{x}) + \sum_j \frac{\partial \mathbf{r}(\mathbf{x})}{\partial x_j}\Delta\mathbf{x}, \quad (3.58)$$

and thus we approximate our residual $\mathbf{r}$ in the neighbourhood of of some model $\mathbf{x}$ by a linear function. Note that K will be linear, but it still depends on $\mathbf{x}$. The sensitivity matrix is again recognized

$$\frac{\partial \mathbf{r}(\mathbf{x})}{\partial x_j} = \nabla K(W^{-1}\mathbf{x}) = J(W^{-1}\mathbf{x})W^{-1}. \quad (3.59)$$

3.3.5 The gradient of the objective function

Recall the objective function

$$\phi(\mathbf{x}) = \phi_d(\mathbf{x}) + \lambda\phi_m(\mathbf{x}) = \frac{1}{2}\|K(W^{-1}\mathbf{x}) - \mathbf{d}\|_2^2 + \lambda\|\mathbf{x}\|_1, \quad (3.60)$$

where we opt to replace the ℓ_1-norm with another measure or focusing function μ. Then,

$$\phi(\mathbf{x}) = \phi_d(\mathbf{x}) + \lambda\phi_m(\mathbf{x}) = \frac{1}{2}\|K(W^{-1}\mathbf{x}) - \mathbf{d}\|_2^2 + \lambda\sum_{i=1}^n \mu(x_i). \quad (3.61)$$

The gradient is

$$\nabla\phi_d(\mathbf{x}) = \left(J(W^{-1}\mathbf{x})W^{-1}\right)^T\mathbf{r} + \sum_{i=1}^n \frac{\mathrm{d}\mu(x_i)}{\mathrm{d}x_i} \quad (3.62)$$

where the transpose of the sensitivity matrix J^T is recognized.

The derivatives of the measures introduced in Section 3.2.3 are straightforward to calculate:

$$\frac{\mathrm{d}\mu_{\mathrm{Ekblom}}(x)}{\mathrm{d}x} = \frac{p}{2}\frac{1}{\sqrt{x^2+\epsilon}} \cdot 2x = \frac{xp}{\sqrt{x^2+\epsilon}} \tag{3.63}$$

$$\frac{\mathrm{d}\mu_{\mathrm{Huber}}(x)}{\mathrm{d}x} = \begin{cases} x & \text{for } |x| \leq \delta, \\ \\ \pm\delta, & \text{otherwise.} \end{cases} \tag{3.64}$$

and the derivative of the minimum support functional is

$$\frac{\mathrm{d}\mu_{\mathrm{Zdh}}(x)}{\mathrm{d}x} = \frac{2\epsilon^2 x}{(\epsilon^2+x^2)^2}. \tag{3.65}$$

In this section, we have discussed how the objective function can be reformulated in a proper optimization problem. There are some choices to be made. We choose the most 'flexible' optimization problem without the slack variables and with the Ekblom measure, because this measure resembles best the ℓ_1-norm for $p = 1$.

$$\min_{\mathbf{x}} \phi(\mathbf{x}) = \min_{\mathbf{x}} \left(\phi_d(\mathbf{x}) + \lambda\phi_m(\mathbf{x})\right) = \min_{\mathbf{x}} \left(\frac{1}{2}\|K(W^{-1}\mathbf{x}) - \mathbf{b}\|_2^2 + \lambda \sum_i \sqrt{x_i^2+\epsilon}\right), \tag{3.66}$$

Note that the summation can be avoided with the Hadamard product or pointwise multiplication $\star$ as

$$\nabla\phi_m(\mathbf{x}) = \mathbf{x} \star \mathbf{w} \text{ where } w_i = \frac{1}{\sqrt{x_i^2+\epsilon}}. \tag{3.67}$$

This is also the closest notation for the vectorized implementation in Python.

3.4 Optimization methods

In this section, several fundamental ingredients for powerful optimization algorithms are summarized. The line search method and Broyden-Fletcher-Goldfard-Shanno (BFGS)-method are required to understand the Limited memory (L)-BFGS-B-method, which will be used in our inversion scheme. It is important to understand these ingredients because blindly adopting the algorithm will obstruct the interpretation of our inversion scheme and our implementation.

There are a plethora of optimization algorithms available. Conversations with colleague students and members of the physics department have led us to examine several algorithms, but the BFGS-method was generally considered as a good 'standard' algorithm. A few words on the other suggestions and why they are not used: A suggestion was the use of Particle Swarm Optimization[14] (with implementation in Python: PySwarm [Lee, 2014]). This technique does not use gradient information and will be much slower than

[14]This technique utilizes a population of candidates (bees) that move around in the model space, with a given velocity. Each particle (or bee) finds its best positions (with lowest objective function), but when other bees find another better solution, they will converge towards the better position. Like bee colonies that search for flowers.

the BFGS-method. In Geophysical literature [Bunks et al., 1995], multiscale approaches are also popular. These approaches are generally slower than the BFGS-method. We will first assume that the objective function does not exhibits a strongly multimodal behaviour. (It turns out that this is true for simple, two-layered conductivity profiles, but not that this is not generally true for complex models. We will elaborate on this issue in Chapter 5.) The main idea behind multiscale approaches are summarized in Appendix C.

3.4.1 Line search methods

In general, a line search method produces a sequence of iterates that meets the requirement $\phi(\mathbf{x}^{(k+1)}) < \phi(\mathbf{x}^{(k)})$, by finding an optimal step length for a given descent direction at every iteration. The following notation for the iterates is adopted

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha^{(k)}\mathbf{d}^{(k)}, \tag{3.68}$$

where $\alpha^{(k)}$ is the step length at iteration k and $\mathbf{d}^{(k)}$ is a direction in which the objective function decreases.

The first step of the algorithm is to find the direction $\mathbf{d}^{(k)}$. We have some freedom in the choice of the direction $\mathbf{d}^{(k)}$, but it should satisfy the descent requirement

$$\nabla\phi(\mathbf{x}^{(k)})^T\mathbf{d}^{(k)} < 0. \tag{3.69}$$

A symmetric matrix $B^{(k)} \in \mathbb{R}^{n\times n}$ defines the direction from the objective function's gradient,

$$\mathbf{d}^{(k)} = -(B^{(k)})^{-1}\nabla\phi(\mathbf{x}^{(k)}), \tag{3.70}$$

for which an additional requirement can be derived. From the descent requirement Eq. (3.69) and Eq. (3.70)

$$\nabla\phi(\mathbf{x}^{(k)})^T\mathbf{d}^{(k)} = -\nabla\phi(\mathbf{x}^{(k)})^T(B^k)^{-1}\nabla\phi(\mathbf{x}^{(k)}) < 0 \tag{3.71}$$

it is clear that if $\nabla\phi(\mathbf{x}^{(k)}) \neq 0$, $B^{(k)}$ should be positive definite matrix. The negative gradient $\mathbf{d} = -\nabla\phi(\mathbf{x})$ is the simplest example, where $B = I$.

The second step is to find an optimal step length $\alpha^{(k)}$. It is a trade-off between computational efficiency and the desire for a substantial reduction. The requirement (3.69) is not sufficient. Consider the following example (from [Nocedal and Wright, 2006]) in Figure 3.7 where $f(x^{(k)}) = 5/k$. Every iteration satisfies $f(x^{(k+1)}) < f(x^{(k)})$, but the function converges to zero and not $f^* = -1$. We need stronger requirements.

Armijo condition

The Armijo condition, also known as the sufficient decrease condition, requires:

$$\phi(\mathbf{x}^{(k)} + \alpha^{(k)}\mathbf{d}^{(k)}) \leq \phi(\mathbf{x}^{(k)}) + \sigma\alpha^{(k)}\nabla\phi(\mathbf{x}^{(k)})^T\mathbf{d}^{(k)}, \tag{3.72}$$

where $\sigma \in (0, 1)$. The difference with the descent requirement is that a step length $\alpha^{(k)}$ has to be chosen such that the objective function is smaller than a first order function, rather than the constant $\phi(\mathbf{x}^{(k)})$. The larger $\alpha^{(k)}$, the larger the decrease of the objective

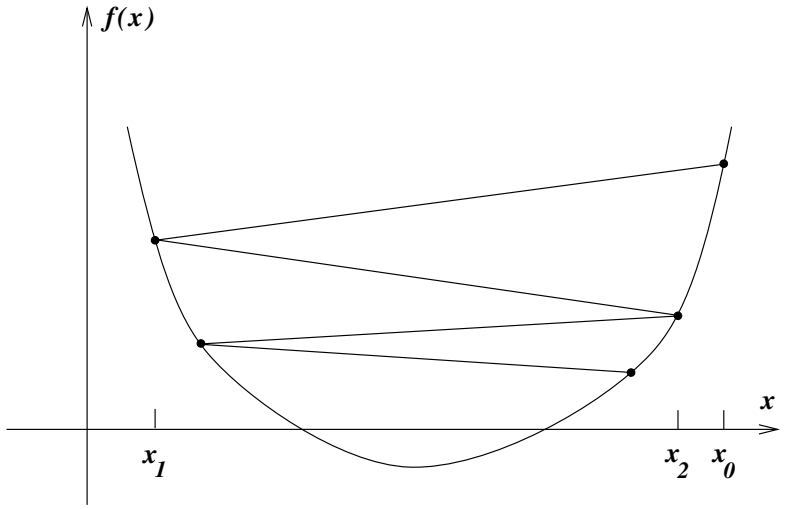


Figure 3.7: Insufficient decrease [Nocedal and Wright, 2006].

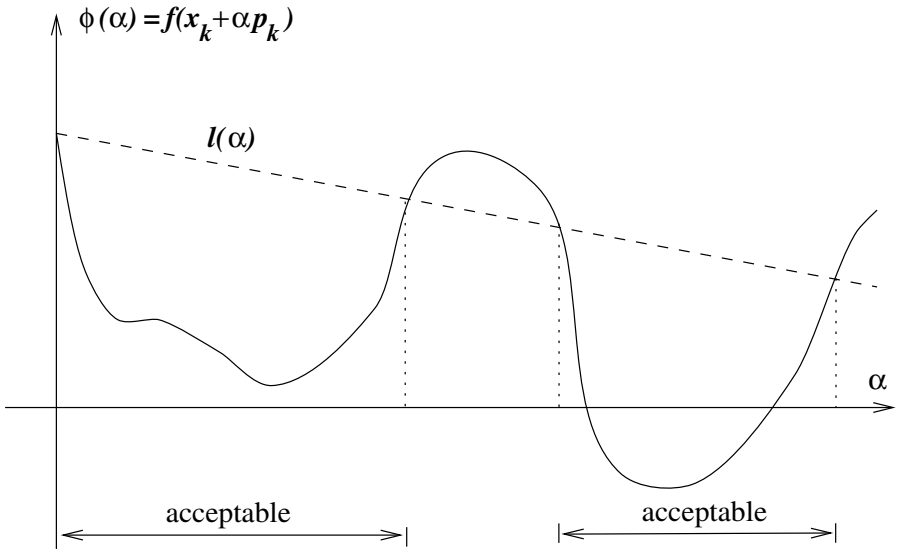


Figure 3.8: Armijo condition. Figure retrieved from [Nocedal and Wright, 2006].

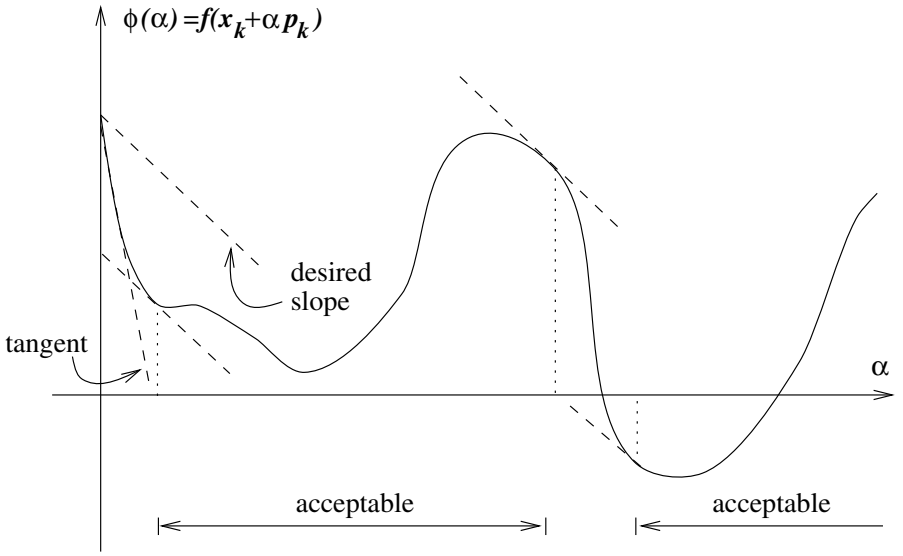


Figure 3.9: Curvature condition. Figure retrieved from [Nocedal and Wright, 2006].

function should be. The parameter σ is typically 10^{-4} [Patrinos, 2018], such that our condition is not too stringent.

There are several strategies for implementing line searches for which the Armijo condition holds. Usually, an initial step length $\alpha^{(0)}$ is chosen and subsequently, the step length is iteratively halved until the Armijo condition is satisfied. This yields the largest allowed step (smaller than the maximal step length $\alpha^{(0)}$) that sufficiently decreases the objective function. This method is referred to as the Backtracking Line Search Algorithm, presented in Algorithm 1 in Appendix B, where β is the factor by which the step length $\alpha^{(k)}$ is decreased when the Armijo condition is not satisfied.

Curvature condition

With the Armijo condition, sufficient decrease of the objective function is guaranteed. A potential failure with the Armijo condition is that the step length is too small, this is possible when the parameter σ is too large. To rule out these unacceptably small steps, an additional *curvature condition* is imposed:

$$\nabla\phi(\mathbf{x}^{(k)} + \alpha^{(k)}\mathbf{d}^{(k)})^T\mathbf{d}^{(k)} \geq \eta\nabla\phi(\mathbf{x}^{(k)})^T\mathbf{d}^{(k)}. \tag{3.73}$$

This condition ensures that the slope of $\phi(\mathbf{x}^{(k)} + \alpha^{(k)}\mathbf{d}^{(k)})$ is greater than η times the initial slope. Indeed, for $\eta < 1$, the smaller slope yields that small α's are excluded, as is shown in Figure 3.9. A strongly negative slope of $\phi(\mathbf{x}^{(k)} + \alpha^{(k)}\mathbf{d}^{(k)})$ suggest that the objective function can be significantly reduced for larger α, while a slightly negative slope is a sign that the objective function cannot be decreased much further in that direction.

The Armijo condition and curvature condition are collectively known as the Wolfe conditions. There exists a software implementation of the line search algorithm that satisfies the Wolfe conditions. The implementation by [Moré and Thuente, 1994] will be used in this

thesis. The strength of this implementation is in the sweep strategy (e.g. Backtracking) that yields minimal iterations.

3.4.2 The BFGS-method

A Newton's method requires Hessian information in order to find the minima of an objective function. The Hessian has to be computed at each iteration. When the Hessian computation is too expensive, the Broyden-Fletcher-Goldfard-Shanno (BFGS)-method can be used. It is a quasi-Newton method, which means that it uses an approximation of the true Hessian. The approximate Hessian will be obtained solely by first-order information (i.e. gradient information).

Finding minima of the objective function ϕ is equivalent to finding the roots of $\nabla\phi$. The roots of a function g are found via Newton's method. At every iteration, the function g is linearised and the root of the linear function is the next iterate. In one dimension,

$$\ell^{(k)}(x) = g(x^{(k)}) + \frac{\mathrm{d}g(x^{(k)})}{\mathrm{d}x}(x - x^{(k)}) = 0 \Rightarrow x = x^{(k+1)} = x^{(k)} - \frac{g(x^{(k)})}{g'(x^{(k)})} \tag{3.74}$$

and when the derivative g' is unknown, a finite difference approximation of g' is used:

$$g'(x^{(k)}) \approx B^{(k)} = \frac{g(x^{(k)}) - g(x^{k-1})}{x^{(k)} - x^{k-1}}. \tag{3.75}$$

This method is known as the secant method. This reasoning can be applied to $\nabla\phi$ in multiple dimensions. The series of iterates are

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - (B^{(k)})^{-1}\nabla\phi(\mathbf{x}^{(k)}), \tag{3.76}$$

where $B^{(k)}$ satisfies the *secant condition*

$$B^{(k)}(\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}) = \nabla\phi(\mathbf{x}^{(k)}) - \nabla\phi(\mathbf{x}^{k-1}). \tag{3.77}$$

In shorter notation, the secant condition is written as

$$B^{(k)}\mathbf{s}^{(k-1)} = \mathbf{y}^{(k-1)}, \text{ where } \mathbf{s}^{(k-1)} = \mathbf{x}^{(k)} - \mathbf{x}^{(k-1)} \text{ and } \mathbf{y}^{(k-1)} = \nabla\phi(\mathbf{x}^{(k)}) - \nabla\phi(\mathbf{x}^{(k-1)}). \tag{3.78}$$

It remains to find a positive definite matrix $B^{(k)}$ which satisfies the linear equations, imposed by the secant condition, and has an inverse. We have some freedom on $B^{(k)}$: there are n equations and $\frac{n(n+1)}{2}$ unknowns[15]. The BFGS method attempts to iteratively find minimum Frobenius norm[16] corrections to the Hessian, i.e. $B^{(k+1)}$ is not too far from $B^{(k)}$. Broyden, Fletcher, Goldfarb and Shanno have derived [Nocedal and Wright, 2006] the following famous formula:

$$B^{(k+1)} = B^{(k)} + \frac{\mathbf{y}^{(k)}\mathbf{y}^{(k)\mathrm{T}}}{\mathbf{y}^{(k)\mathrm{T}}\mathbf{s}^{(k)}} - \frac{B^{(k)}\mathbf{s}^{(k)}\mathbf{s}^{(k)\mathrm{T}}B^{(k)\mathrm{T}}}{\mathbf{s}^{(k)\mathrm{T}}B^{(k)}\mathbf{s}^{(k)}}, \tag{3.79}$$

which determines the descent direction $\mathbf{d}^{(k+1)}$. The complete algorithm is presented in Algorithm 2 in Appendix B.

[15]The matrix $B^{(k)}$ has n^2 elements, but the Hessian is symmetric.

[16]The Frobenius norm is defined as $\|A\|_{\mathrm{F}} = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2} = \sqrt{\mathrm{trace}\,(A^{\mathsf{T}}A)}$, according to [Golub and Van Loan, 1996]. It is a matrix norm.

3.4.3 L-BFGS-B algorithm

The actual algorithm that will be used in the inversion scheme is the L-BFGS-B method, implemented in the Python SciPy solver [Jones et al., 2001]. It uses the algorithm from [Byrd et al., 1995] and [Zhu et al., 1997]. It is different from the BFGS-method in two ways.

The first difference is that it can be used for bound constrained minimization, via a gradient projection method [Nocedal and Wright, 2006]. Each iteration can be divided in two stages. The first stage starts as a normal line search along the gradient. When a variable encounters its constraint, then this variable is held fixed, such the line search stays in the feasible model space. Along this piecewise-linear path, the first local minimum is defined as the *Cauchy point*. From that point, an active set is defined. It is the set of points (or models) for which the components of the Cauchy points are at their bounds. The second stage is then to solve the optimization problem for the free variables (those that are not at their bounds). The operator P projects the gradient $\nabla\phi$ on the free variables of the model space. $P\nabla\phi$ is thus the projected gradient.

The second difference is that limited memory matrices are used to iteratively approximate the Hessian of the objective function, via the BFGS formula (3.79). This is especially useful for large scale optimization problems.

Chapter 4

Wavelet Theory

4.1 Introduction

Wavelet theory has been used in exploration geophysics for many decades [Kumar and Foufoula-Georgiou, 1997]. While it is common in seismic applications [Kearey et al., 2013], the context is now different. Here, it is introduced as a regularization tool in our inversion scheme, which will be further discussed in Chapter 5. The goal of this chapter is to build up the basics of wavelet theory, which will provide us the right vocabulary. By the end of this chapter, one is familiar with the different properties of wavelets and understands why it is crucial to choose the right wavelet. Relations between vanishing moments, compact support and regularity of the wavelet will play an important role. While the continuous wavelet transform is widely used in seismic geophysics, this chapter is dedicated to the discrete wavelet transform for which the signals are discretely sampled. The wavelet transform has both frequency and temporal resolution and we expect that the inherent multiscale nature can be exploited in our inversion application. Understanding the discrete wavelet transform (DWT) and its fast wavelet transform (FWT) requires understanding of filter banks, for which the elementary basics are summarized in this chapter.

4.2 Setting the scene

In Fourier series, a 2π periodic signal in the continuous time domain can be represented in terms of basis functions $\{e^{ikt}\}_{k\in\mathbb{Z}}$ with Fourier coefficients $f_k = \langle e^{ikt}, f\rangle$ [Mallat, 1999]. Indeed, we write

$$f(t) = \sum_{k\in\mathbb{Z}} f_k e^{ikt}. \tag{4.1}$$

Analogously to Fourier series, wavelet theory builds up a set of basis functions, in which a signal can be represented. There are two types of such basis functions: *scaling functions* φ and *wavelet functions* ψ. Finding such bases and determining the properties are the major concern of wavelet theory.

Consider the wavelet function $\psi(t)$ and the following translations and dilatations

$$\psi_{n,k}(t) = 2^{n/2}\psi(2^n t - k) \tag{4.2}$$

where n is the dilatation parameter that makes the wavelet function's compact support wider or smaller. Dilating the wavelet function $\psi(t)$ corresponds to translations on the frequency axis. The k-parameter describes translations along the t-axis. These functions $\{\psi_{n,k}\}_{k,n\in\mathbb{Z}}$ form a basis for a function space. Assume, for now, that the basis is orthonormal and that the wavelet function $\psi(t)$ is compactly supported[1]. Similarly as with Fourier series, the representation of a signal in the wavelet basis is of the form

$$f(t) = \sum_{n,k\in\mathbb{Z}} w_{nk}\psi_{nk}(t), \tag{4.3}$$

where $w_{nk} = \langle \psi_{nk}, f\rangle$ are the wavelet coefficients.

The dimension of the wavelet basis set is infinite, which for practical reasons is not always desirable or computationally possible. One can approximate a function f by projecting it on a subspace $f_n^W \in W_n$, the space of resolution n. The basis functions $\{\psi_{n,k}\}_{k\in\mathbb{Z}}$ generate that subspace W_n. A function f at resolution n is represented by

$$f_n^W(t) = \sum_{k\in\mathbb{Z}} w_{nk}\psi_{nk}(t). \tag{4.4}$$

Exactly the same reasoning can be applied to scaling functions $\varphi(t)$. Then,

$$\varphi_{n,k}(t) = 2^{n/2}\varphi(2^n t - k) \tag{4.5}$$

and we can project f on a subspace V_n, which is generated by the basis functions $\{\varphi_{n,k}\}_{k\in\mathbb{Z}}$. Then,

$$f_n^V(t) = \sum_{k\in\mathbb{Z}} v_{nk}\varphi_{nk}(t), \tag{4.6}$$

where $v_{nk} = \langle \varphi_{nk}, f\rangle$ are the scaling coefficients. In what will become clear later, the following relations between subspaces hold:

$$V_n = V_{n-1} \oplus W_{n-1}. \tag{4.7}$$

A one level wavelet transform is the decomposition of a function $f_n \in V_n$ in its components $f_{n-1} \in V_{n-1}$ and $g_{n-1} \in W_{n-1}$.

$$f_n = \sum_k v_{nk}\varphi_{nk} = \underbrace{\sum_k v_{n-1,k}\varphi_{n-1,k}}_{f_{n-1}} + \underbrace{\sum_k w_{n-1,k}\psi_{n-1,k}}_{g_{n-1}} \tag{4.8}$$

where $v_{n-1,k}$ are the scaling coefficients and $w_{n-1,k}$ the wavelet coefficients of scale of resolution n. g_{n-1} describes the details that were present on scale n, but have disappeared from the coarser $(n-1)$-scale.

[1] [Loeb, 2016] defines in Definition 9.9.5.: "A continuous function with *compact support* on a topological space $(X, \mathcal{I})$ is a continuous function that is identically equal to 0 outside of some compact subset of X". Ignoring the technicalities of a topological space that facilitate the definition of a compact set, we paraphrase that a function has compact support if it is zero outside a compact set. The function $f : x \to x^2$ in its entire domain is not compactly supported, while a block function does have compact support.

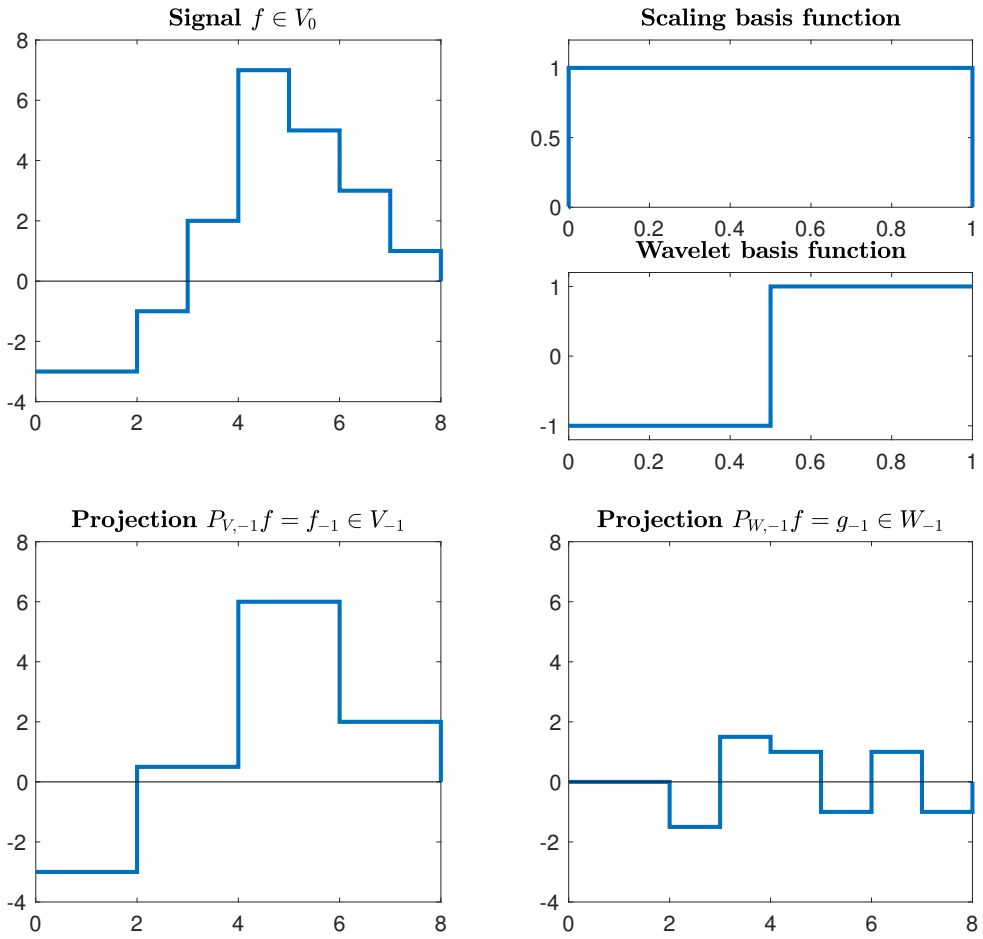


Figure 4.1: A function $f_0 \in V_0$ and its one level wavelet transform.

To make things more concrete, a classic example of such a one level transform is presented with Haar wavelets. A blocky signal f is considered in Figure 4.1. It is piecewise constant and has a domain width of 8 and each constant piece has width 1. The scaling function $\varphi(t)$ is a block function and has width 1 in V_0, whereas the wavelet function is a linear combination $\psi(t) = \varphi(2t) - \varphi(2t-1)$. At each resolution level, the basis of the subspaces are $V_n = \{\varphi_{n,k}\}_{k\in\mathbb{Z}}$ and $W_n = \{\psi_{n,k}\}_{k\in\mathbb{Z}}$, where dilatations and translations are considered as in Eq. (4.2). The scaling and wavelet function are shown in Figure 4.1. The signal f is constructed such that $f \in V_0$, so it can be decomposed as

$$f(t) = \sum_{k=0}^{7} v_{0k}\varphi_{0k}(t). \tag{4.9}$$

Applying a one level wavelet transform

$$f(t) = \sum_{k=0}^{7} v_{0k}\varphi_{0k} = \sum_{k=0}^{3} v_{-1,k}\varphi_{-1,k} + \sum_{k=0}^{3} w_{-1,k}\psi_{-1,k} \tag{4.10}$$

splits the signal in the two subspaces V_{-1} and W_{-1}, where the scaling and wavelet functions now have width 2. Only four functions span each subspace. The scaling and wavelets coefficients are the amplitudes from the block functions and wavelet functions shown in Figure 4.1.

Having illustrated what is meant by the wavelet transform, we will now move on to the motivation behind wavelet theory. As previously mentioned, wavelet theory has some similarities with Fourier analysis. The Fourier transform maps a function from time or space domain to frequency domain. By examining Fourier coefficients, valuable information about the frequency content of a signal is obtained. This is useful in e.g. signal denoising [Walker, 1997]. The Fourier representation of periodic functions are sparse and are therefore easier to work with. The cosine is an example of a perfectly localized function in frequency domain, it contains one frequency. In the time domain, however, the cosine has infinite support and thus has no localization. The wavelet transform is a decomposition that captures both frequency and space/time information, i.e. it has both frequency and temporal resolution.

Temporal resolution can be advantageous, because then only few coefficients are needed to represent local transient behaviour. Indeed, our conductivity profile is expected to exhibit no periodic behaviour, however some low frequency content will be present. Neighbouring model parameters probably belong to the same subsurface layer and have the same conductivity. Only model parameters at the boundaries will differ with the neighbouring parameters. The wavelet transform is promising for our purpose, for it has both time and frequency localization. The Fourier transform has periodical basis functions with infinite support and no time localization, it would be strange to consider such decomposition in a minimum structure inversion scheme.

Note that it is impossible to have perfect localization in both domains. [Mallat, 1999] calls it the Heisenberg Uncertainty principle: It is impossible to have a signal with finite support on the time axis which is at the same time band limited[2]. The Heisenberg principle is well known in the Fourier analysis and is best illustrated with the help of the Gaussian distribution. The Gaussian distribution does not have compact support in any domain, but entertains a notion of localization expressed in its variance. The Fourier transform of a Gaussian with variance σ^2 is again a Gaussian, yet with reciprocal variance $1/\sigma^2$. A Gaussian with small variance corresponds with strong localization and large variance with less localization. If a Gaussian has strong localization (or small variance) in the time domain, it is spread out in the frequency domain (large variance). In the limit, a Dirac pulse in the time domain is a constant function in frequency domain and vice versa. Wavelet theory has a multiscale nature by the choice of basis, expressed in Eq. (4.2). Dilating the wavelet function corresponds with translations in the frequency domain. [Mallat, 1999] uses rectangles in the time-frequency plane to explain the multiscale nature of the wavelet basis. A wavelet $\psi_{n,k}(t)$ has a specific time support centred at k and proportional to n. Smaller time support or higher resolution scales n correspond with high temporal resolution, at the cost of a more spread out Fourier transform. Figure 4.2 illustrates that relation, note that we identify $u = k$ and $s = n$. The right figure illustrates that every

[2]More formally, one needs to define a measure for expressing the width of the support of a function and of its Fourier transform.

time value is covered by many wavelets at different scales (and resolution).

The discrete Fourier transform can numerically be obtained in a reliable fashion via the fast Fourier transform (FFT), which has a complexity of only $\mathcal{O}(N \log N)$. The discrete wavelet transform also has such a fast algorithm with complexity of $\mathcal{O}(N)$. The existence of that algorithm makes wavelet theory popular in applications such as image denoising, compressed sensing and optimization.

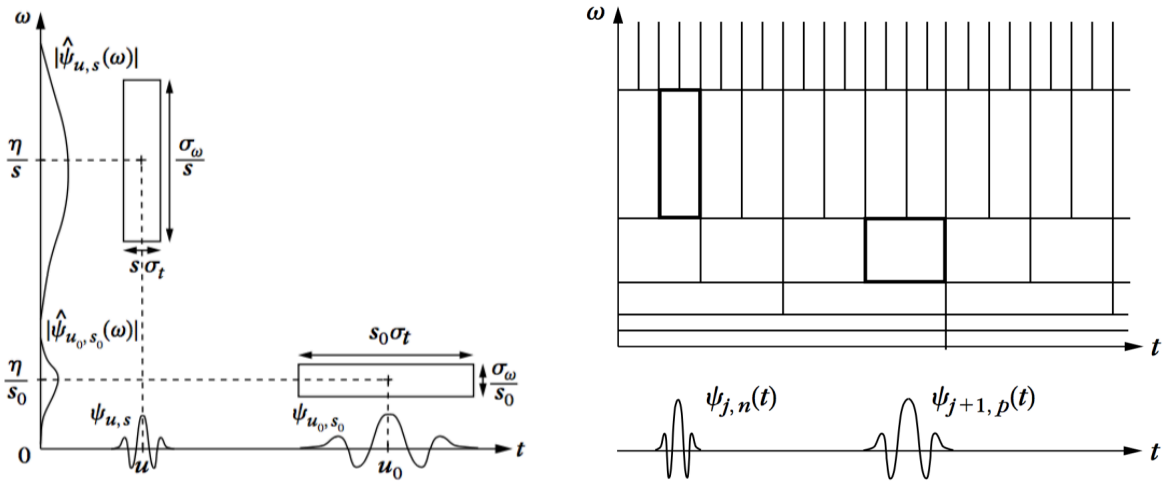


Figure 4.2: **(left)** Heisenberg time-frequency boxes of two wavelets. **(right)** The wavelet basis tiles the time-frequency plane. Figures retrieved from [Mallat, 1999].

4.3 A filter bank Interpretation

There are two main approaches to introduce wavelet theory. One approach exploits the similarity to Fourier transformations in which compactly supported basis functions $\Phi(t)$ decompose some function $f(t)$. This approach is well described in Daubechies' book [Daubechies, 1992]. The second approach is based on the deep link with filter banks and is key to the fast wavelet transform. Since we are more concerned about the discrete wavelet transform, the latter method is described here. In this text, the concept of a filter bank is introduced, which decomposes a discrete signal into two signals via high and low pass filters. The two resulting signals have half the size after a subsampling procedure. A key result is that the original signal can be reconstructed from the two half-sized output signals, when necessary and sufficient conditions are satisfied.

4.3.1 Filters

Definition 5 (Filter) *A filter $\mathcal{H}$ is an operator which maps a signal into another signal.*

The *impulse response* h_n is the effect of the filter applied to a Dirac impulse δ, $h_n = \mathcal{H}\delta$ In what follows, the z-transform is often considered. The unilateral z-transform converts

a discrete-time signal $\mathbf{s}$ into the z-domain, a complex frequency-domain:

$$\mathcal{Z}\{\mathbf{s}\} = \sum_{k=0}^{\infty} s_k z^{-k} = S(z). \tag{4.11}$$

Setting $z = e^{i\omega}$, reduces the z-transform into the discrete Fourier transform of the signal. When linear filters are considered, a filtering operation can be interpreted as a multiplication with a Toeplitz matrix[3]: $\mathbf{g} = T\mathbf{f}$ with

$$T = \begin{bmatrix} \ddots & \ddots & \ddots & \ddots & & & \\ \cdots & h_2 & h_1 & h_0 & h_{-1} & \cdots & \\ & \cdots & h_2 & h_1 & h_0 & h_{-1} & \cdots \\ & & & \ddots & \ddots & \ddots & \ddots \end{bmatrix}.$$

An example of such filter is an ideal bandpass filter, whose Fourier transform is 1 for $|\omega| \leq \pi/2$ and 0 for $|\omega| > \pi/2$. The elements of the Toeplitz matrix or the impulse response can be easily derived

$$h_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} H(\omega)e^{ik\omega}d\omega = \frac{1}{2\pi}\int_{-\pi/2}^{\pi/2} e^{ik\omega}d\omega = \frac{2}{k\pi}\sin\frac{k\pi}{2} = \frac{1}{2}\text{sinc}\frac{k\pi}{2} \tag{4.12}$$

for $k \neq 0$ and $h_0 = 1/2$. Such a filter cannot be constructed in practise, because the filter is not causal. And computationally, it would require an infinite dimensional Toeplitz matrix.

We conclude this section with a definition and some notation

Definition 6 (Transfer function) *A transfer function of a filter $\mathcal{H}$ is the z-domain representation of that filter $H(z) = \sum_k \bar{h}_k z^k$.*

The complex conjugate and time reverse of a signal $\mathbf{s} = (s_k)$ is denoted by a substar conjugate $\mathbf{s}_* = (\bar{s}_{-k})$. A linear filter $\mathcal{H}_*$ has impulse response h_* and has a Toeplitz matrix representation H^*, the Hermitian adjoint of H.

4.3.2 Analysis and synthesis

The discrete wavelet transform has a filter bank interpretation that provides more insight. A filter bank is a series of band-pass filters that separates the input in multiple frequency sub-bands, this is called the analysis side of the filter bank. This section describes the two-channel filter bank, which simplifies the description. The two-channel filter bank consists of a high pass filter $\tilde{G}_*$ and a low pass filter $\tilde{H}_*$ that act on a discrete signal $\mathbf{s}$. After the filtering operation, an input signal of length L is transformed into two output signals of length L. The low pass filter's output contains the low frequencies and thus the low resolution content of the signal. The other signal contains the high frequency content or details of the input signal. For ideal low and high pass filters, each of the two filters

[3]A Toeplitz matrix is often referred to as diagonal-constant matrix, since $T_{ij} = T_{i+1,j+1}\ \forall i \in \mathbb{Z}$

take exactly half of the frequency band. After downsampling the output signals, the two half bands are rescaled to the original full bandwidth. This can be seen in the z-domain:

$$\frac{S(z)+S(-z)}{2} = \frac{1}{2}\sum_{k=0}^{\infty} s_k \left(z^{-k} + (-z)^{-k}\right) \qquad \text{(unilateral } z\text{-transform)} \tag{4.13}$$

$$= \frac{1}{2}\sum_{k=0}^{\infty} s_{2k} \left(z^{-2k} + (-z)^{-2k}\right) \tag{4.14}$$

$$= \sum_{k=0}^{\infty} s_{2k} \left((z^2)^{-k}\right) \tag{4.15}$$

$$= \sum_{k=0}^{\infty} s'_k (z^2)^{-k} \qquad \text{(downsampling)} \tag{4.16}$$

$$= S'(z^2) \tag{4.17}$$

In the frequency domain $z = e^{i\omega}$, yielding

$$S'(e^{i\omega}) = \frac{S(e^{i\omega/2}) + S(-e^{i\omega/2})}{2} \tag{4.18}$$

and thus when the bandwidth of S is B, then the bandwidth of S' is $2B$. If the original signal $x = (x_n)$ has length L, then the two output signals have total length $2L$. By downsampling, the original length of the signal is restored.

The synthesis side of the filter bank is concerned with reconstructing the original signal from the subsampled high and low frequency parts. The synthesis is done in two steps: First, the signals are upsampled, that is adding a zero between every two samples. Secondly, filters are applied in such a way that after addition, the output signal $\tilde{\mathbf{s}}$ equals the original signal $\mathbf{s}$. The analysis and synthesis side of such a filter bank is schematically shown in Figure 4.3.

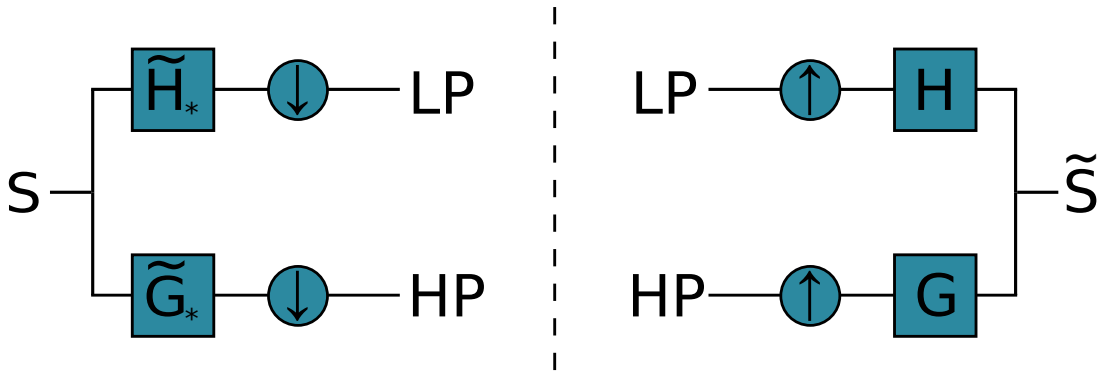


Figure 4.3: The analysis and synthesis side of a two-channel filter bank. $\tilde{H}_*$ and $\tilde{G}_*$ are respectively the analysis low pass (LP) and high pass filters (HP). H and G are respectively the reconstruction filters on the low pass and high pass filter. $\downarrow$ denotes downsampling, while $\uparrow$ denotes the upsampling operation.

A multiple-channel filter bank can be constructed by repeatedly applying a two-channel filter bank on the signal outputs. The discrete wavelet transform is such a multiple-

channel filter bank, where every level of the transform corresponds with a two-channel filter bank applied on solely the low pass output. Since the two-channel filter bank with ideal filters divides the bandwidth in equal widths, the end channels of the recursive filter bank will have different bandwidths. A schematic representation of such a filter bank is set out in Figure 4.4, together with the bandwidths of the output signals with ideal filters.

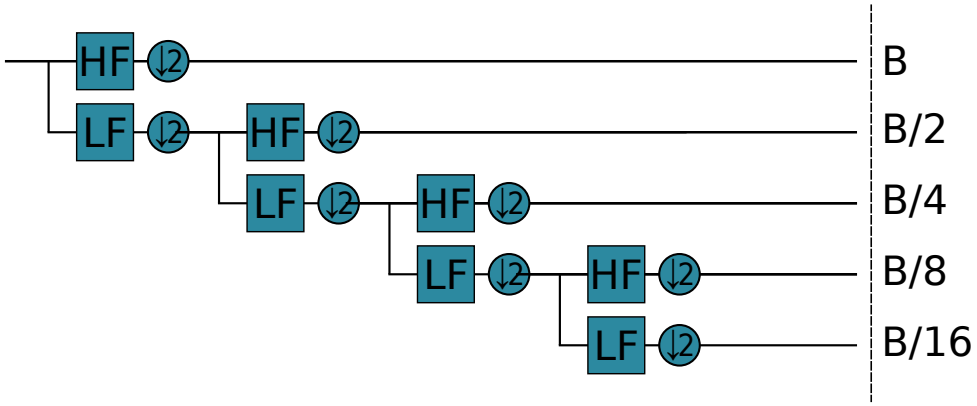


Figure 4.4: The recursive two-channel filter bank (four levels). The two-channel filter bank is applied recursively to the low pass filter. On the right side, bandwidths are given for ideal filters. $\downarrow 2$ denotes downsamling.

4.3.3 Perfect Reconstruction

Designing a two-channel filter bank entails designing good filters $\tilde{G}_*, \tilde{H}_*, G$ and H such that the output signal after synthesis equals the original signal. A filter bank with this property is called a Perfect Reconstruction (PR) filter bank. In terms of matrices, a condition that guarantees PR is easily derived:

$$\begin{bmatrix} \tilde{H}^* \\ \tilde{G}^* \end{bmatrix} \mathbf{s} \equiv \tilde{K}^* \mathbf{s} = \begin{bmatrix} \mathbf{p} \\ \mathbf{q} \end{bmatrix}. \tag{4.19}$$

On the synthesis side, we find (using Figure 4.3 as a reminder)

$$\tilde{\mathbf{s}} = H\mathbf{p} + G\mathbf{q} = \begin{bmatrix} H & G \end{bmatrix} \begin{bmatrix} \mathbf{p} \\ \mathbf{q} \end{bmatrix} \equiv K \begin{bmatrix} \mathbf{p} \\ \mathbf{q} \end{bmatrix}. \tag{4.20}$$

The PR condition requires $\tilde{\mathbf{s}} = \mathbf{s}$, then

$$K\tilde{K}^* = 1. \tag{4.21}$$

However, from that condition, it is challenging to derive conditions for the filters. Without giving any details, a lossless filter bank can be constructed by choosing $\tilde{H}(z) = H(z)$ and $\tilde{G}(z) = G(z)$ and requiring

$$H_*(z)H(z) + H_*(-z)H(-z) = 2, \tag{4.22}$$
$$G_*(z)G(z) + G_*(-z)G(-z) = 2, \tag{4.23}$$
$$H_*(z)G(z) + H_*(-z)G(-z) = 0, \tag{4.24}$$

leads to relations on the filter coefficients (they are called the double shift orthogonality relation and alternating flip relation) and $\tilde{K} = K$.

A lossless filter bank is a more restrictive type of PR filter bank and will lead to orthogonal wavelets. It forces the matrix K to be unitary, however for perfect reconstruction only $K\tilde{K}^* = 1$ is required. Two different matrices K for synthesis and $\tilde{K}$ for analysis that satisfy $K\tilde{K}^* = 1$ will lead to biorthogonal wavelets, which will be further examined in Section 4.5.

4.4 A multiresolution analysis

4.4.1 Formal definition

The example with Haar wavelets in Section 4.2 illustrates how the DWT decomposes a signal in a coarser approximation and detail vector. When applied recursively, the signal's true multiscale nature is exposed. A more formal approach is developed with the definition of a multiresolution analysis:

Definition 7 (Multiresolution)
A multiresolution analysis (MRA) of the Lebesgue space[4] L^2 *is a nested sequence of subspaces* $\cdots V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \cdots$ *such that*

1. $\bigcup_{n\in\mathbb{Z}} V_n = L^2$ *and* $\bigcap_{n\in\mathbb{Z}} V_n = \{0\}$
2. $f(t) \in V_j \Leftrightarrow f(2t) \in V_{j+1}, j \in \mathbb{Z}$
3. $f(t) \in V_0 \Leftrightarrow f(t-k) \in V_0, k \in \mathbb{Z}$
4. $\{\varphi(t-k)\}_{k\in\mathbb{Z}}$ *form a Riesz basis for* V_0.

The first requirement describes that the nested subspaces fill the whole function space. Technically, closure of $\bigcup V_n$ is required, which makes the union dense in L^2. Any function, an element from L^2, can be approximated (arbitrarily close) by elements of that union and its limit points. This is the *completeness* requirement. The intersection ensures that the nested subspaces are not redundant. The second property is that of scale invariance of our sequence. Shift or time invariance is expressed in the third property. The last part of the definition shows that one generating function, the scaling function $\varphi(t)$, together with its integer shifts form a basis for V_0. That basis is complete and non-redundant and not necessarily orthogonal (so far, we only discussed orthogonal scaling functions), but it is at least a Riesz basis[5]. This weaker requirement allows to more freedom and yields a wider variety of wavelets, more specifically biorthogonal wavelets, to be discussed in Section 4.5.

[4] L^p: A function space equipped with a p-norm.
[5] A formal definition of a Riesz basis can be found in [Christensen, 2001].

4.4.2 Bases

The formal definition of a multiresolution analysis enables us to find requirements for $\varphi(t)$. The 4th property of the MRA determines that $\varphi(t) \in V_0$ and from the 2nd property it is clear that $\varphi(t/2) \in V_{-1} \subset V_0$, thus there should exist a relation

$$\varphi(t/2) = \sum_k c_k \varphi(t-k), \tag{4.25}$$

with coefficients $c_k \in \mathbb{R}$. Eq. (4.25) is called the *two-scale relation*. This equation is hard to solve, but several solution strategies exist [Bultheel, 2003]. We are not going to derive new wavelets, therefore these strategies are not discussed. The average value of the scaling functions $\int \varphi dt$ is non-zero, this is called the partition of unity[6]. Indeed, the scaling function considered in Section 4.2 was de box function and has non-zero energy.

From these scaling functions, wavelet functions can be derived via the following wavelet relation:

$$\psi(t) = \sum_n d_n \varphi(2t-n) \in V_1, \quad d_n = (-1)^n \bar{c}_{1-n} \tag{4.26}$$

One can prove[7] that the wavelet basis at a given resolution n forms an orthonormal basis for W_n, where W_n is the orthogonal complement of V_n in V_{n+1} and thus

$$V_{n+1} = V_n \oplus W_n. \tag{4.27}$$

Note that in the literature, the following normalization is often utilized

$$h_k = \frac{c_k}{\sqrt{2}} \quad \text{and} \quad g_k = \frac{d_k}{\sqrt{2}}. \tag{4.28}$$

4.4.3 The link with filter banks

The link between scaling functions and discrete filters is based on the two-scale relation Eq. (4.25). The sequence c_k will be interpreted as a discrete filter. This filter (or its transfer function $C(z)$) describes all the properties of the scaling function $\varphi(t)$, because those (filter) coefficients completely define the function. These coefficients are indeed the only freedom in the two-scale relation.

When the DWT is applied on a function $f_n \in V_n$, then its components $f_{n-1} \in V_{n-1}$ and $g_{n-1} \in W_{n-1}$ are obtained

$$f_n = \sum_k v_{nk} \varphi_{nk} = \sum_k v_{n-1,k} \varphi_{n-1,k} + \sum_k w_{n-1,k} \psi_{n-1,k}, \tag{4.29}$$

where $v_{n-1,k}$ are the scaling coefficients and $w_{n-1,k}$ the wavelet coefficients.

[6] Theorem 6.6 + corollary in [Strang and Nguyen, 1996].
[7] Theorem 7.3, page 278 in [Mallat, 1999].

We want an efficient decomposition algorithm, where v_{nk} can be expressed in terms of $v_{n+1,k}$ and $w_{n+1,k}$:

$$\begin{aligned}
\varphi_{nk}(t) &= 2^{n/2}\varphi(2^n t - k) && \text{(Notation)} && (4.30)\\
&= 2^{n/2}\sum_i c_i \varphi(2^{n+1}t - 2k - i) && \text{(Two-scale relation)} && (4.31)\\
&= \frac{1}{\sqrt{2}}\sum_i c_i 2^{(n+1)/2}\varphi(2^{n+1}t - (2k+i)) && && (4.32)\\
&= \frac{1}{\sqrt{2}}\sum_l c_{l-2k}\varphi_{n+1,l}(t) && (l = 2k+i) && (4.33)\\
&= \sum_l h_{l-2k}\varphi_{n+1,l}(t) && \text{(Normalization, Eq. (4.28))} && (4.34)
\end{aligned}$$

and consequently

$$v_{nk} = \langle \varphi_{nk}, f\rangle = \int \bar{\varphi}_{nk}(t) f(t) dt = \sum_l \bar{h}_{l-2k}\langle \varphi_{n+1,l}, f\rangle = \sum_l \bar{h}_{l-2k} v_{n+1,l}. \tag{4.35}$$

Together with

$$\begin{aligned}
\psi_{nk}(t) &= 2^{n/2}\psi(2^n t - k) && \text{(Notation)} && (4.36)\\
&= 2^{n/2}\sum_j d_j \varphi(2^{n+1}t - 2k - j) && \text{(cf. Eq. (4.26))} && (4.37)\\
&= \frac{1}{\sqrt{2}}\sum_j d_j 2^{(n+1)/2}\varphi(2^{n+1}t - (2k+j)) && && (4.38)\\
&= \frac{1}{\sqrt{2}}\sum_l d_{l-2k}\varphi_{n+1,l}(t) && && (4.39)\\
&= \sum_l g_{l-2k}\varphi_{n+1,l}(t) && && (4.40)
\end{aligned}$$

and thus

$$w_{nk} = \sum_l \bar{g}_{l-2k} v_{n+1,l}. \tag{4.41}$$

Equations (4.35) and (4.41) are crucial results and reveal the filter bank interpretation of the discrete wavelet transform. Relation (4.35) is a convolution[8] with a downsampled and time reversed impulse response. A convolution in time domain is a multiplication in the z-domain and thus the scaling coefficient vector $\mathbf{v_n}$ is obtained by the multiplication of a suitable (Toeplitz-)matrix with the scaling coefficients at a higher scale of resolution $\mathbf{v_{n+1}}$. This will be the downsampled and time reversed filter $\mathcal{H}^*$ from Section 4.3. We thus have

$$\begin{bmatrix}\mathcal{H}^*\\ \mathcal{G}^*\end{bmatrix} v_{n+1} = \begin{bmatrix} v_n \\ w_n\end{bmatrix}, \tag{4.42}$$

[8] An ordinary discrete convolution: $v_k \star h_k = \sum_l v_l h_{k-l}$. A convolution with time reversed and downsampled filter coefficients has the following form: $v_k \star h_{-2k} = \sum_l v_l h_{l-2k}$

where the scaling coefficients v_k contain the low frequency content and the wavelet coefficients w_k the high frequency content. In linear algebra notation

$$\begin{bmatrix} H^* \\ G^* \end{bmatrix} v_{n+1} = \begin{bmatrix} v_n \\ w_n \end{bmatrix}, \tag{4.43}$$

which is shown in Figure 4.5 for the one level DWT with Haar (or Daubechies 1) wavelets.

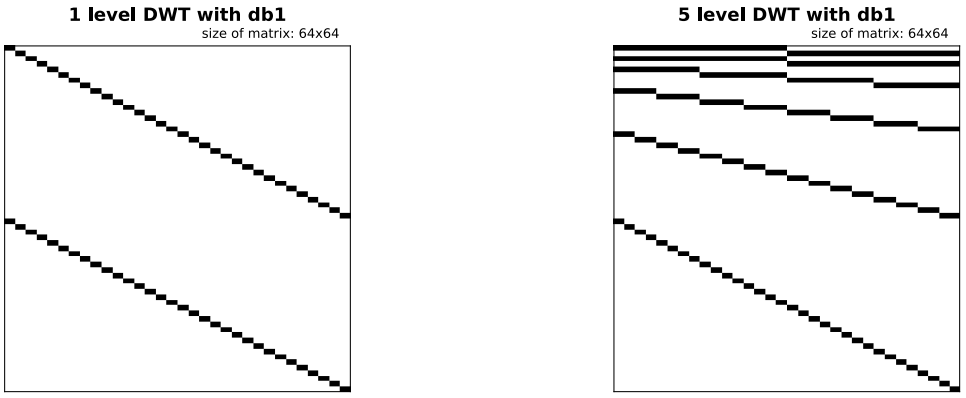


Figure 4.5: The linear algebra representation of a **(left)** one level **(right)** five level discrete wavelet transform with Haar (or Daubechies 1) wavelets. This wavelet has two filter coefficients. A black box represents a non-zero element.

Similarly to the Fast Fourier Transform for the Discrete Fourier Transform, there exists a Fast Wavelet Transform (FWT) for the Discrete Wavelet Transform. The computation cost is low due to the sparsity of the transformation matrix. Since that matrix is orthogonal, the inverse matrix is easily calculated. Finite dimensional data will often be periodically or symmetrically extended until it has a length 2^l. When the wavelets have a large number of filter coefficients, the matrices will be truncated to match the length of the signal. This will yield edge effects and orthogonality will be lost. For this purpose, the matrix has periodical boundary conditions, as is shown in Figure 4.6.

There are many implementations available for the FWT. We have used the Wavelet Toolbox [MathWorks, 2019] in Matlab and the PyWavelets package [Lee et al., 2006] in Python. These implementations do not use the linear algebra matrix, which is required in explicit form for minimizing the objective function (cf. Chapter 3). I have therefore implemented the FWT matrix in Python, which was used in Figures 4.5 and 4.6.

4.4.4 Signal extension

Because an analysis will be made with wavelets with more than two filter coefficients, we are forced to consider 'signal extension'. In this case, the matrix from the previous section no longer has periodic boundary conditions. Depending on the wavelet, the signal will be extended *at every level of the wavelet decomposition.* Indeed, it will be necessary to represent the signal with more vector entries than in space or time domain. This is a

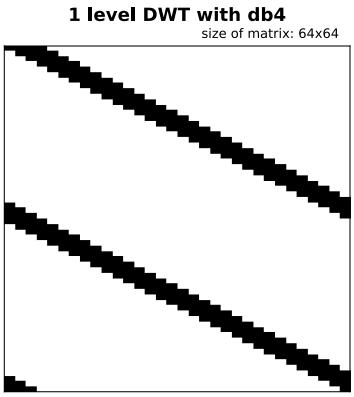


Figure 4.6: The linear algebra representation of a one level discrete wavelet transform with Daubechies 4 wavelets. This wavelet has 8 filter coefficients and in order to restore orthogonality, the matrix is periodically extended. A black box represents a non-zero element.

requirement for the perfect reconstruction property.

The need to represent the signal with more entries in the wavelet domain than in the space domain can be recognized in two ways. Let us first consider the linear algebra representation. It is impossible to fit all filter coefficients (if there are more than two) in a square matrix if we do not want to apply boundary conditions or truncate the Toeplitz-matrix. Indeed, for a one level DWT, it becomes a $n_m \times n_x$ matrix. n_m is the length of the signal in the space domain and n_x is the length in the wavelet domain. For example, if there are 5 filter coefficients, then $n_x = n_m + 3$. For each additional DWT level, the signal in the wavelet domain becomes longer.

An analogous situation occurs with convolutions[9]. With a finite signal, one must think about how the boundaries should be handled. It is virtually impossible to correlate a basis function with a part of a signal, when that basis function has wider support than the length of that part of the signal. The signal is therefore usually extended.

There are several ways to extend a signal. The most popular approach is symmetrization (e.g., this is Matlab's default [Misiti et al., 2009]), because this usually unabruptly extends the behaviour of the signal. Consequently, there will be less boundary distortion. With symmetrization, a copy of the signal is flipped and fixed to the ends of the signal. See Figure 4.7. Other strategies include periodic and antisymmetric signal extension. Reflection is the same as symmetric extension, with the difference that the first and last entry of the signal are not copied. Furthermore, the signal can be extended by only copying the first and last entry of that vector over the complete extension interval. In smooth signal extension, a linear fit is produced at each side, based on the two boundary entries. These types of signal extension are all illustrated in Figure 4.7.

The choice of signal extension will determine the boundary distortion. Boundary distor-

[9]We have already argued that the wavelet transform is some type of convolution.

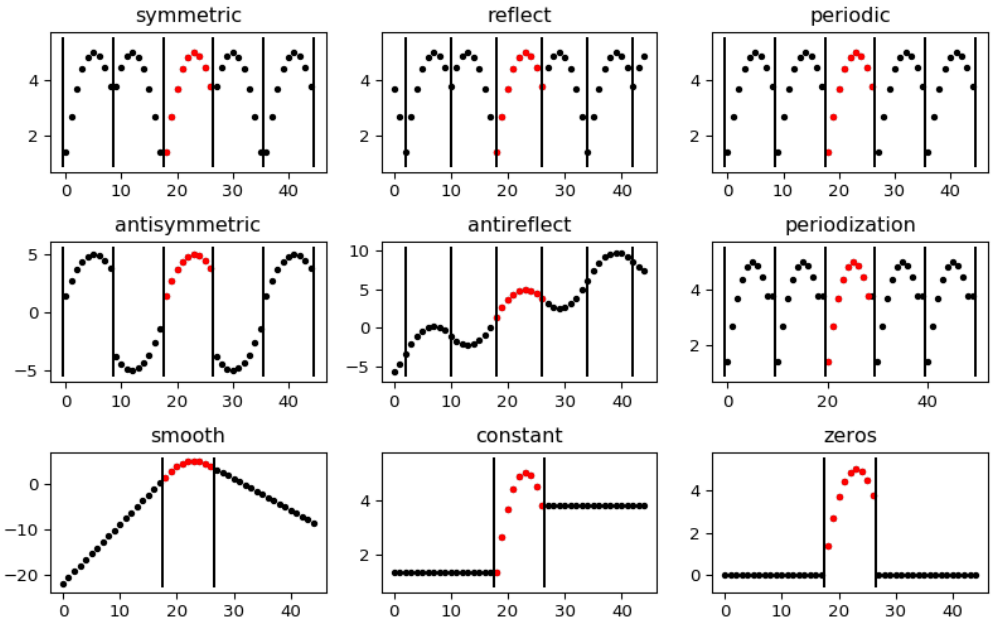


Figure 4.7: Types of signal extension for the discrete wavelet transform.

tions generally result in large values in the wavelet representation. Therefore, the wavelet representation will lose its approximating abilities (see Section 4.6). We will have to examine what works satisfactorily in our inversion scheme.

The linear algebra representation that was implemented in the previous section is no longer sufficient. A signal after a one level DWT will not necessarily have a length that is a power of two. In addition, orthogonality will be lost, so the inverse transformation matrix must also be implemented (It is now no longer possible to consider its transpose). Another approach will have to be developed that generates the matrix in explicit form.

The alternative implementation of the transformation matrix W is actually very simple, provided that you have access to a toolbox that has implemented the wavelet transform. As mentioned earlier, we use the PyWavelet toolbox in Python [Lee et al., 2006]. Consider δ_i the 'Dirac train' of length n_m with spike at the i^{th} entry. In this context, it is a column-vector. The idea is the following:

$$W = W\mathbb{I} \tag{4.44}$$
$$= W\left[\delta_1\ \delta_2 \cdots \delta_{n_m}\right] \tag{4.45}$$
$$= \left[W\delta_1\ W\delta_2 \cdots W\delta_{n_m}\right]. \tag{4.46}$$

The wavelet transform from the package PyWavelet can be used to transform the Dirac trains to the wavelet domain $W\delta_i$, specified with the correct level of DWT and extension mode. Concatenating the resulting wavelet representations yields an explicit representation of the $n_x \times n_m$ wavelet transformation matrix W.

For completeness, we note that there exist different wavelet transforms that avoid this signal extension issue, the so-called "wavelets on the interval" methods [Cohen et al., 1993], [Černá and Finěk, 2011]. They are theoretically interesting, but the concepts are hardly used in practise, due to practical limitations [Misiti et al., 2009]. Additionally, the intuition behind the transform is less clear.

4.5 Biorthogonality

Regarding the perfect reconstruction property, the matrix K was assumed to be unitary and yielded orthogonal wavelets. This orthogonality condition is too restrictive and more wavelet families can be found when more degrees of freedom become available by imposing a weaker requirement. The definition of a multiresolution analysis deliberately required only a Riesz basis and not an orthogonal basis.

There exists a weaker condition on the PR property, where K does not need to be unitary. The requirement $K\tilde{K}^* = 1$ can be fulfilled by considering two distinct matrices K and $\tilde{K}$ which are biorthogonal. The formalism slightly alters to a function space with a primal and dual basis. The functions ψ and φ generate the primal bases $\{\psi_{nk}\}$ and $\{\varphi_{nk}\}$, while the dual basis is generated by $\tilde{\psi}$ and $\tilde{\varphi}$. For these bases, the relations

$$\langle \varphi_{nk}, \tilde{\varphi}_{nl} \rangle = \delta_{k-l}, \tag{4.47}$$

$$\langle \psi_{nk}, \tilde{\psi}_{ml} \rangle = \delta_{k-l}\delta_{m-n}, \tag{4.48}$$

$$\langle \varphi_{nk}, \tilde{\psi}_{nl} \rangle = 0, \tag{4.49}$$

$$\langle \tilde{\varphi}_{nk}, \psi_{nl} \rangle = 0 \tag{4.50}$$

hold and thus a function can be expanded in both bases. We follow the convention that ψ and φ will be used on the synthesis side and that $\tilde{\psi}$ and $\tilde{\varphi}$ will be used on the analysis side. The relations in Eqns. (4.47)-(4.50) show that $V_n \perp \tilde{W}_n$ and $W_n \perp \tilde{V}_n$. It is still true that $V_n \oplus W_n = V_{n+1}$, but they are not orthogonal complements (this contrasts the orthogonal wavelet formalism).

4.6 Wavelet families and properties

The performance of a wavelet application strongly depends on the choice of the wavelet. It is often advantageous to choose a wavelet in which a function is represented in a sparse fashion, for that yields faster calculations. Other properties are regularity, vanishing moments, size of the (compact) support and the convergence rate.

Vanishing moments The number of vanishing moments is the most decisive property of a wavelet. A wavelet has p vanishing moments when

$$\int t^k \psi dt = 0 \quad \text{for } k = 0, 1 \cdots, p-1. \tag{4.51}$$

It is equivalent of saying that a multiresolution analysis is of order p, it means $t^k \in V_0$ for $k = 0, 1 \cdots, p-1$. Recall the Haar scaling function from Section 4.2, which is a box

function with support width 1. Indeed, the Haar scaling function has $t^0 \in V_0$ and thus has $p = 1$ vanishing moment. In general, p vanishing moments means that ψ is orthogonal to any polynomial of degree $p - 1$.

The number of vanishing moments of ψ is related to the number of zeros of the low pass filter $H(\omega)$. p vanishing moments of the wavelet function is equivalent to the low pas filter having p zeros at $\omega = \pi$. To see this, consider the two-scale relation for the wavelet function $\tilde{\psi}$ in the Fourier domain[10]

$$\tilde{\Psi}(2\omega) = \frac{1}{\sqrt{2}}\tilde{G}(\omega)\tilde{\Phi}(\omega), \tag{4.52}$$

where

$$\tilde{G}(\omega) = \sum_k \tilde{g}e^{-ik\omega}. \tag{4.53}$$

We have that $\tilde{\Phi}(0) \neq 0$, since[11]

$$\tilde{\Phi}(\omega = 0) = \frac{1}{\sqrt{2\pi}} \int \tilde{\varphi}(t)e^{-i0t}dt = \frac{1}{\sqrt{2\pi}} \int \tilde{\varphi}(t)dt \neq 0. \tag{4.54}$$

Eq. (4.52) for $\omega = 0$ yields

$$\tilde{\Psi}(0) = \frac{1}{\sqrt{2}}\tilde{G}(0)\tilde{\Phi}(0) \Rightarrow \tilde{G}(0) = 0. \tag{4.55}$$

For $\tilde{\psi}$ with $k - 1$ vanishing moments, we find its derivatives

$$\tilde{\Psi}^{(k)}(\omega) = \frac{1}{\sqrt{2\pi}} \int \tilde{\psi}(t)(-it)^k e^{-i\omega t}dt = (-i)^k \frac{1}{\sqrt{2\pi}} \int \tilde{\psi}(t)t^k e^{-i\omega t}dt \tag{4.56}$$

$$\Rightarrow \tilde{\Psi}^{(k)}(\omega = 0) = (-i)^k \frac{1}{\sqrt{2\pi}} \int \tilde{\psi}(t)t^k dt = 0, \tag{4.57}$$

which is used to find higher order zeros of $\tilde{G}(0)$. Differentiating the Fourier space two-scale relation for $\tilde{\psi}$ and using the condition from (4.57): $\tilde{\Psi}^{(k)}(0) = 0$ for $k = 1$ yields

$$2\tilde{\Psi}^{(1)}(2\omega) = \frac{1}{\sqrt{2}}\tilde{G}^{(1)}(\omega)\tilde{\Phi}(\omega) + \frac{1}{\sqrt{2}}\tilde{G}(\omega)\tilde{\Phi}^{(1)}(\omega) \tag{4.58}$$

$$\Rightarrow 2\tilde{\Psi}^{(1)}(0) = \frac{1}{\sqrt{2}}\tilde{G}^{(1)}(0)\tilde{\Phi}(0) + \frac{1}{\sqrt{2}} \underbrace{\tilde{G}(0)}_{\text{cf. Eq. (4.55)}} \tilde{\Phi}^{(1)}(0) \tag{4.59}$$

$$\Rightarrow \tilde{G}^{(1)}(0) = 0. \tag{4.60}$$

Differentiating $k = p - 1$ times proves that $\tilde{G}$ has $p - 1$ zeros in $\omega = 0$. By differentiating the relation[12] $\tilde{G}(\omega) = e^{-i\omega}\overline{H(\omega + \pi)}$, the p zeros of $H(\omega)$ at $\omega = \pi$ are obtained. This

[10] Obtained by multiplying the two-scale relation with $e^{-i\omega t}$ and integrating with respect to t. Alternatively, $\tilde{G}$ and $\tilde{H}$ are obtained by the z-transform, where only the unit circle $z = e^{\omega t}$ is considered.

[11] Recall the partition of unity from Section 4.4.2.

[12] See [Bultheel, 2003] Section 5.9.

property is often referred to as '*flatness*' of the wavelet.

At first sight, the compact support of a wavelet seems unrelated to the number of vanishing moments. A theorem by Daubechies [Daubechies, 1988] states that an orthogonal wavelet with p vanishing moments has at least a support of size $2p-1$. In the same paper, Daubechies shows that the order of regularity increases linearly with the support width. Regularity is related to smoothness, which is a potential desirable property of our inversion scheme.

Convergence When dealing with piecewise smooth functions, a wavelet basis is better than e.g. a Fourier basis. We touched upon this idea earlier in this chapter, but by stating a theorem[13], we make it more formal:

Theorem 3 (Convergence of wavelet approximation) *When $H(\omega)$ has p zeros at π, any p-times differentiable function $f(t)$ is approximated to order $(\Delta t)^p = 2^{-jp}$ by its projection $f_j(t)$ in V_j:*

$$||f(t) - f_j(t)|| \leq C(\Delta t)^p ||f^p(t)|| \tag{4.61}$$

Thus, the number of vanishing moments p determines the rate at which the projections onto V_j converge to f. Note the explicit condition that f should be p-times differentiable. Since wavelets are localized in time, it is sufficient that functions are only locally smooth (in the support width of the basis function). Another important theorem[14] is the following:

Theorem 4 (Decay of the Wavelet) *If $f(t)$ is p times differentiable, its wavelet coefficients decay like 2^{-jp}:*

$$|w_{j,k}| \leq C 2^{-jp} ||f^p(t)|| \tag{4.62}$$

Thus, a wavelet coefficient decays with increasing j and thus for smaller scales. Higher vanishing moments induce faster decay and thus sparser representations. This theorem can be inverted: when large wavelet coefficients are observed on all scales, then this is a result from discontinuities or singularities of f. This theorem therefore reveals an important trade off: higher vanishing moments improve the approximating abilities of a wavelet (which is interesting in applications), but also induce a larger compact support. This higher compact support width makes it more difficult to meet the conditions of the theorem: the function f must then be piecewise smooth on a larger interval.

Examples The most famous wavelets are the Daubechies (`db`) wavelets. The Haar wavelet was already used in earlier sections (the corresponding scaling function was the block function) and is an example of a Daubechies wavelet, more precisely it is referred to as the Daubechies 1 wavelet. In general, Daubechies wavelets are orthogonal and compactly supported. For a given support, they attain the highest possible number of vanishing moments. We denote `dbN` as a Daubechies wavelet of order p, i.e. with $p = N$ vanishing moments. It has a support size of $2p-1$ and a regularity (for large p) of 0.2p. The `db2-db4` wavelets are shown in Figure 4.8. Indeed, with increasing p we observe smoother wavelets and larger support widths.

[13] Theorem 7.5 p.230 in [Strang and Nguyen, 1996]
[14] Theorem 7.6 p.231 in [Strang and Nguyen, 1996]

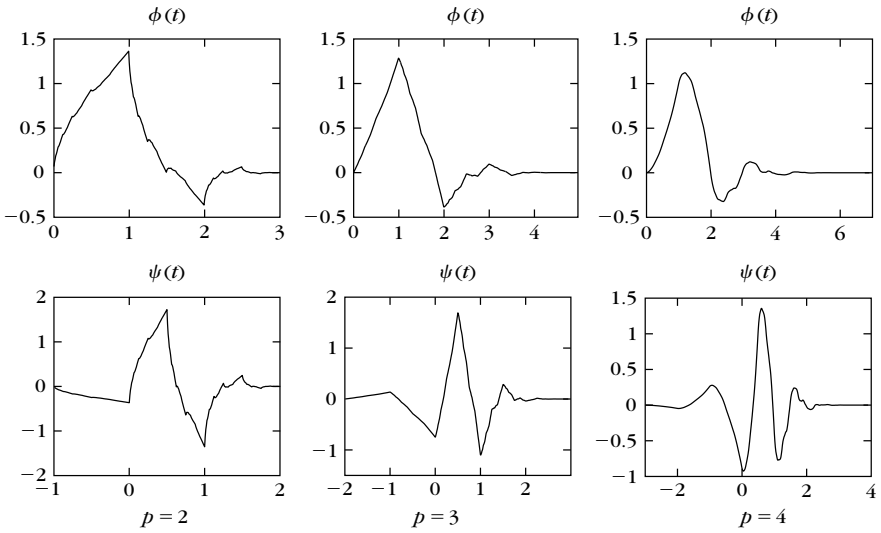


Figure 4.8: Daubechies scaling and wavelet functions with vanishing moments p. Figure retrieved from [Mallat, 1999].

In summary, wavelet theory provides a framework for finding compactly supported basis functions which are localized both in time and frequency. It has a multiscale nature, present in the wavelet basis. The wavelet transform has a fast implementation that can be viewed as a matrix operation acting on the signal. Choosing an appropriate wavelet is a crucial part in designing applications. In the next chapter, we put together all the ingredients and propose our inversion scheme.

Chapter 5

Analysis and Discussion

In this chapter, we discuss the most important results in building up a robust inversion scheme. Robust in the sense that it can handle noisy measurements and unmodelled artefacts. First, a simple scheme is proposed based on the optimization methods in Chapter 3 and wavelet theory from Chapter 4. Subsequently, the results will be analysed in order to improve the scheme. We will do this by adding or adjusting features in the inversion scheme, step by step.

5.1 The inversion scheme

5.1.1 The outline of our scheme

In Chapter 3, we have discussed a framework for solving inverse problems and have argued why a minimum structure regularization is required. This resulted in an objective function ϕ that we must minimize with iterative methods. Recall the objective function from Section 3.2.4

$$\phi(\mathbf{x}) = \phi_d(\mathbf{x}) + \lambda\phi_m(\mathbf{x}) = \frac{1}{2}\|\mathbf{d} - K(W^{-1}\mathbf{x})\|_2^2 + \lambda\|\mathbf{x}\|_1, \tag{5.1}$$

where W is the basis transformation

$$\mathbf{x} = W\mathbf{m} \tag{5.2}$$

in which the model parameters $\mathbf{m}$ are represented in a sparse fashion.

We can construct such a transformation with the concepts from Chapter 4 about wavelet theory. Then $\mathbf{x}$ becomes a vector with model parameters in the wavelet domain. There is a lot of freedom: In the first place, one can choose to implement the transformation at various levels of resolution N. Secondly, there is the choice of wavelet. We know from Section 4.6, that the wavelets in the Daubechies family are suitable candidates, because they are by construction the best wavelet in the following trade-off: more vanishing moments improve the approximation abilities of a wavelet, that is, you can make a good approximation of a vector with few non-zero wavelet coefficients. On the other hand, more vanishing moments induce larger compact support widths, which on their turn make it more difficult to meet the conditions of a proper decay of the wavelet coefficients (i.e. close to zero). Indeed, Daubechies wavelets are those orthogonal wavelets that have the

smallest support width for a given vanishing moment.

The wavelet transform has its limitations though. Recall its filter bank interpretation that on the analysis side splits the vector into two parts: a high-frequency part (detail or wavelet coefficients) and a low-frequency part (approximation or scaling coefficients). In view of facilitating a potential full discrete wavelet transform, we require the conductivity profile parameterization $\mathbf{m} \in \mathbb{R}^{+,n}$ to have a length of power two. This specification is rather a programming inconvenience/issue, than something that causes fundamental problems. Suppose that in some context, a parameterization is needed with predetermined layer thickness and that the variability in the magnetic field data is largely explained with a parameterization up to a certain depth d' (remember: this depends on the maximum intercoil distance of the measurement). If this parameterization has n' model parameters, it can be extended to a dimension of power two without any problem. Indeed, these parameters will not make a significant contribution to the variability in the data.

In Section 4.4.4, we explained that the full DWT requires a signal extension $\mathbf{x} \in \mathbb{R}^{n_x}$ in order to preserve all the properties of the wavelet and to avoid unpleasant boundary effects. As a matter of fact, we must then solve the objective function for n_x parameters for a model parameterization of only $n_m \leq n_x$ parameters, this makes the minimization problem more cumbersome. We can avoid this issue by not considering the full DWT, i.e. a $N < \log_2(n_m)$ level DWT. Consequently, we have introduced an additional difficulty: can the true model parameterization still be sufficiently sparsely represented in the wavelet domain? Fortunately, the Haar wavelets, also called Daubechies 1 (`db1`) wavelets, do not exhibit this difficulty (they only have two filter coefficients per filter). We first limit ourselves to the `db1` wavelet, which is furthermore the most intuitive. In Section 5.4, we will loosen the power of two requirement and examine the effect of a $N < \log_2(n_m)$ level DWT and its effect on boundary distortions.

The DWT with `db1` wavelet is now examined on a simple conductivity profile. This conductivity profile comprises a 5 meters thick upper layer with an electrical conductivity of 0.5 S/m and a second semi-infinite layer with a conductivity of 0.05 S/m. This profile is modelled by 8 layers ($n = 8$) with a thickness of 2.5 meters. Figure 5.1a presents this profile in the space domain, together with all the representations in the wavelet domain with the Haar wavelet. In the figure with the one level discrete wavelet transform (with subspaces of $\mathbf{m} \in V_0 = V_{-1} \oplus W_{-1}$), the pattern of the original model is recognized in the scaling coefficients, these coefficients are the first four components of $\mathbf{x}$. Note that a normalization has resized all values in $\mathbf{x}$. This ditto pattern in the scaling coefficients (the first four coefficients) is not a surprise, because the model in the space domain could exactly be represented with a parameterization $\mathbf{m}$' with only four components on a different basis, more specifically with layers (or block functions) with thicknesses of 5 meters, for which $\mathbf{m}' \in V_0' = V_{-1}$. In the two-level DWT, the first two components are still scaling coefficients. The interpretation for Haar wavelets is that the first component gives the average value of the first four components of $\mathbf{m}$ and that the second component describes the average of the last four components, except for a rescaling. The wavelet coefficients provide a 'correction' on the too simplified coarse scale scaling coefficients and ensures the perfect reconstruction property. The full DWT (here: three-level DWT) usually describes the given model in the most sparse fashion. However, we note that both the two

and three-level DWT generate a 3-sparse vector, this is not generally valid. In conclusion, we can represent a non-sparse vector, the model $\mathbf{m}$, by an S-sparse vector in the wavelet domain. In this example, we get a 3-sparse vector.

The first component of $\mathbf{x}$ in the full DWT is never zero, as in the example above. Suppose that in the full DWT, the scaling coefficient is zero, by consequence the space domain vector has zero energy (the sum of all elements in $\mathbf{m}$ is zero), something that is not expected in our physical model $\mathbf{m} \in \mathbb{R}^{+,n}$ because electrical conductivities are positive. Therefore, we can mitigate the sparsity constraint in our objective function. Note that we cannot simply repeat this reasoning on other components since we do not know a priori which components will be non-zero. The objective function with the mitigated sparsity constraint is now

$$\phi(\mathbf{x}) = \frac{1}{2}\|\mathbf{d} - K(W^{-1}\mathbf{x})\|_2^2 + \lambda \left(\sum_{i=2}^{n_x} |x_i| \right), \tag{5.3}$$

where n_x is the total number of components in $\mathbf{x}$. Since we are going to use iterative methods that use gradient information, the perturbed ℓ_p-norm measure of Ekblom (Eq.(3.15)

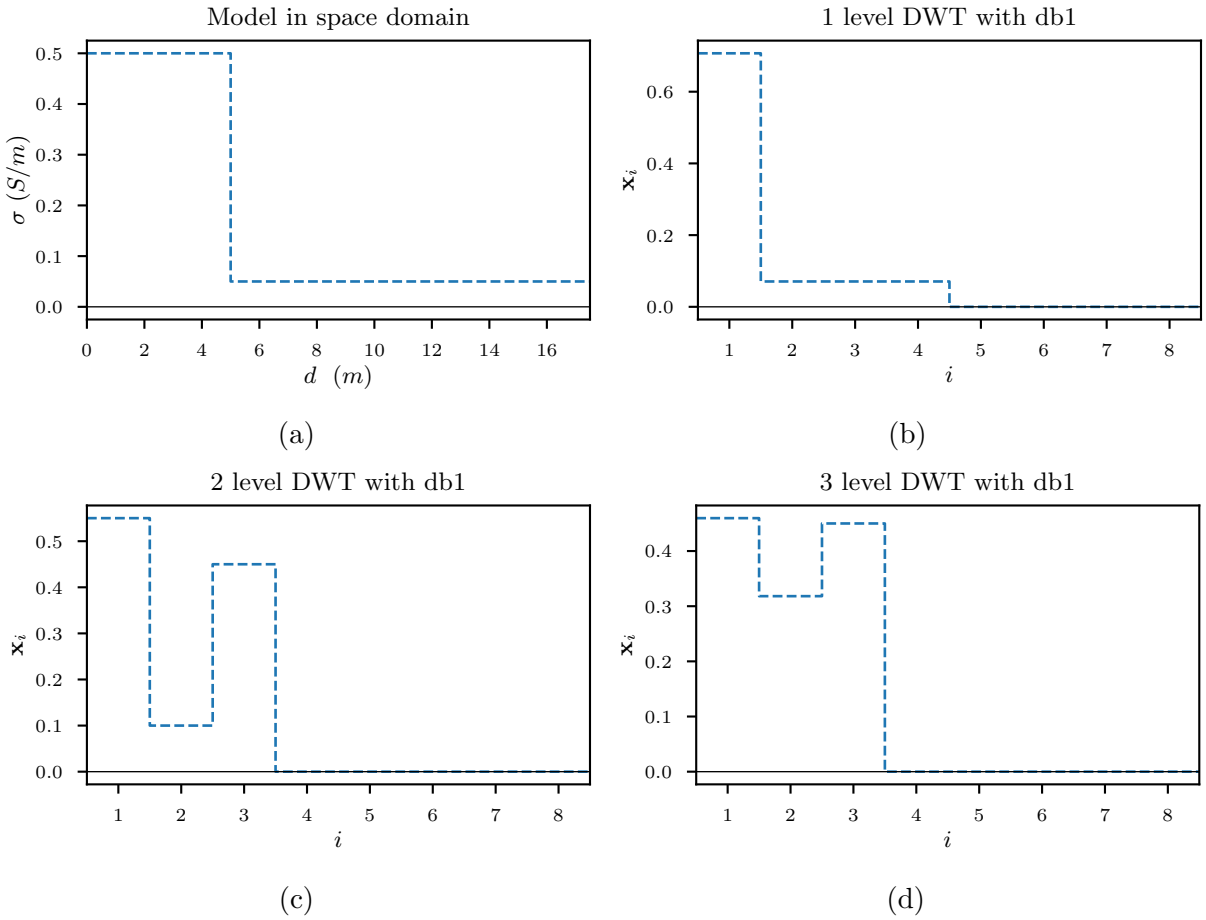


Figure 5.1: (a) The conductivity profile in the space domain. (b)-(d) level of resolution N discrete wavelet transform with the Haar or Daubechies 1 wavelet.

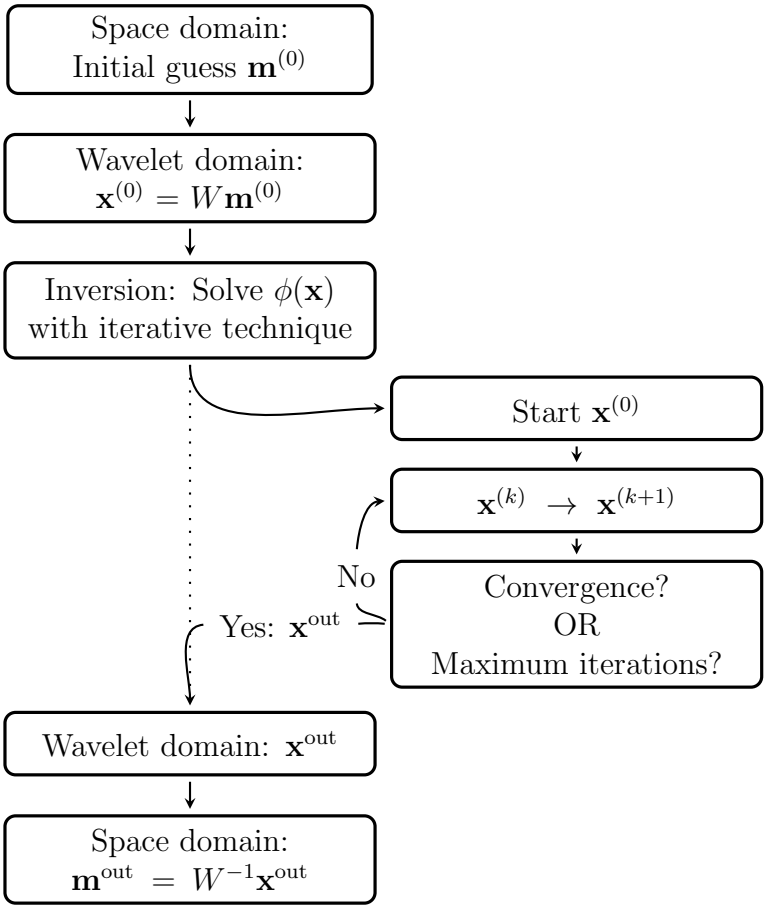


Figure 5.2: The inversion scheme

in Section 3.2.3) for $p = 1$ is used instead of the ℓ_1-norm

$$\phi(\mathbf{x}) = \frac{1}{2}\|\mathbf{d} - K(W^{-1}\mathbf{x})\|_2^2 + \lambda \sum_{i=2}^{n_x} \left(\sqrt{x_i^2 + \epsilon} \right), \tag{5.4}$$

where $\epsilon = 10^{-4}$ unless otherwise stated[1].

The synthetic data $\mathbf{d} \in \mathbb{R}^m$ is obtained from 20 measurements with intercoil distances $s = 1, 2, \cdots, 20$ meters in both horizontal coplanar (HCP) and perpendicular with vertical dipole configurations ($m = 40$) at a height of 0.1 meter. It will be further specified which forward model is used to generate the data and what noise is added in order to make the synthetic data more realistic.

An iterative inversion procedure starts with an initial guess or trial model. In a true geological inversion process, the geophysicist tries to choose the initial guess based on the available geological knowledge. Here, we try to create a robust inversion scheme and pretend that no prior knowledge is available. The simplest initial guess is a constant model, although it could just as well be a randomly generated positive model vector with values

[1]This is a good choice for ϵ, according to [Farquharson and Oldenburg, 1998].

lower than 1 S/m^2. In our inversion experiments, the initial guess is a constant, unless otherwise stated.

After determining the initial guess, it is transformed into the wavelet domain (we have the freedom to choose the level of the DWT and the wavelet). Subsequently, the objective function ϕ is minimized through the iterative method (see Section 3.4). The objective function's gradient information allows a line search algorithm to search for a minimum of the objective function. After each iteration, it is checked whether the iterate has converged. We set the following stop criteria:

$$\frac{\phi^{(k)} - \phi^{(k+1)}}{\max(1, |\phi^{(k)}|, |\phi^{(k+1)}|)} \leq 10^{-10}, \tag{5.5}$$

$$\max(|P\nabla\phi_i|, i = 1, \cdots, n) \leq 10^{-7}, \tag{5.6}$$

$$\text{maximum 1500 iterations}, \tag{5.7}$$

where $P\nabla\phi$ is the projected gradient (see Section 3.4.3).

After conversion, we (hopefully) get a minimum of the objective function $\mathbf{x}^{\text{out}}$. Eventually, the outcome is translated into the space domain through the appropriate inverse discrete wavelet transform. This inversion scheme is schematically summarized in Figure 5.2.

5.1.2 Test of the inversion scheme

The inversion scheme, as proposed in the previous section, is tested on the conductivity profile from Figure 5.1. To keep things simple, again only eight model parameters are considered. For simplicity, the data is generated and inverted via the LIN approximation and no noise was added to the synthetic data. The wavelet representation of $\mathbf{m}$ is obtained via the full DWT with Haar wavelets. The regularization parameter λ is held constant during the iterations and tuned manually. The L-BFGS-B-method [Zhu et al., 1997] is used in the standard Python SciPy [Jones et al., 2001] solver. Note that we do not yet impose bounds on our parameters and thus $P\nabla\phi = \nabla\phi$.

The inversion was performed for several regularization parameters λ. A too large regularization parameter yields the null solution (let's say a 0-sparse 'solution'). When the regularization parameter was too small, the solution exhibits complex structure, i.e. it is not sparse and is slightly overfitted. We can summarize the typical outcome after conversion for what seems a slightly too small regularization parameter and a slightly too large parameter in Figure 5.3 (note that the lowest subsurface layer is semi-infinite, and thus the solid line $\mathbf{m}^{\text{true}}$ must be thought of to be semi-infinite). The left figure shows the inversion with $\lambda = 0.001$, the solution does not show the expected minimum structure. The exact model in wavelet space is represented by a 3-sparse vector, where the first three elements are (0.459, 0.318 ,0.45), while the solution after conversion, rounded to 3

[2]This is a realistic upper bound, based based on experimental data obtained from [Hermans et al., 2012].

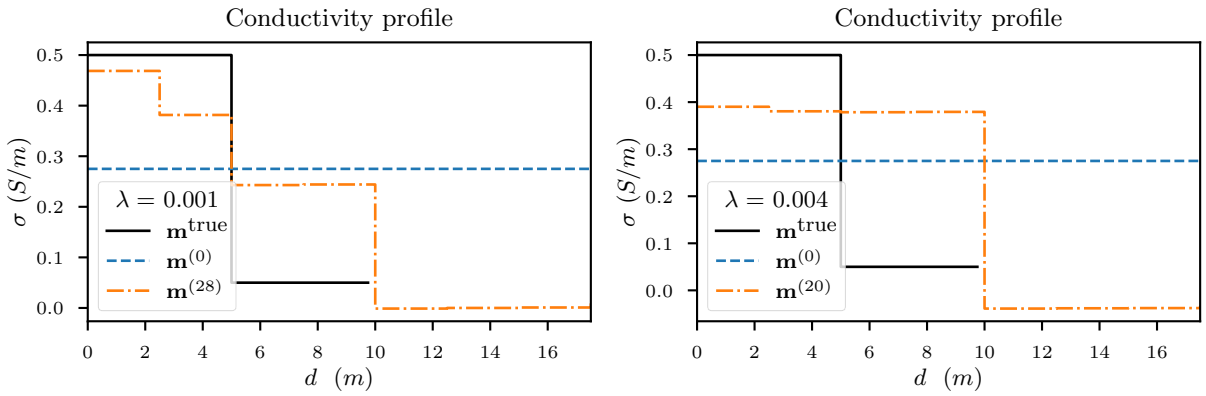


Figure 5.3: Model $\mathbf{m} = W^{-1}\mathbf{x}$ after conversion (orange -·- line) for $\lambda = 0.001$ (left) and 0.004 (right). The solid line describes the true model, which was used to generate the data. No noise was added. The initial guess $\mathbf{m}^{(0)}$ (blue, dashed line) was a constant model in space domain.

significant figures, yields

$$\mathbf{x}^{\lambda = 0.001} = \begin{bmatrix} 0.475 \\ 0.470 \\ 0.181 \\ -4.96 \times 10^{-3} \\ 0.0615 \\ -8.26 \times 10^{-4} \\ -8.68 \times 10^{-4} \\ -4.86 \times 10^{-3} \end{bmatrix}, \tag{5.8}$$

which is not sparse. One could do thresholding[3] to clean up the vector $\mathbf{x}$, although this introduces an extra issue: which threshold δ should we choose? Additionally, it somewhat obscures the effect of the sparsity constraint in the objective function which we want to assess now, therefore it is not a good approach to consider this in an initial inversion scheme.

When the regularization parameter is increased to $\lambda = 0.004$, the outcome is different and has more minimal structure (see Figure 5.3, right). Note that the conductivities at large depths have negative electrical conductivities, which is unphysical! This issue will be resolved in a later version of the inversion scheme. We cannot fully rely on the visual inspection of the outcomes. In the discussion above, we presented outcomes that seemed to be a result of too large or too small regularization parameters. However, in the light of our objective function's model misfit, both outcomes seem to have too little structure, compared to the true model. This minimum structure can be examined by considering the model misfit ϕ_m of the outcomes and to compare it with the exact solution, rounded

[3] [Liu et al., 2017] suggest that one can do thresholding after every iteration. That is removing the coefficients that are smaller than some threshold δ. However, they only mention it and do not include it in their inversion scheme.

to 3 significant figures

$$\phi_m(\mathbf{x}^{\text{true}}) = \sqrt{x_3 + \epsilon} + 5\sqrt{\epsilon} = 0.500, \tag{5.9}$$
$$\phi_m(\mathbf{x}^{\lambda\,=\,0.001}) = 0.286, \tag{5.10}$$
$$\phi_m(\mathbf{x}^{\lambda\,=\,0.004}) = 0.0645, \tag{5.11}$$

from which we observe that the regularization was in fact too strong. In the next section, we will develop a slightly different regularization term that better agrees what we visually assume a complex structure solution.

5.1.3 Scale-dependent regularization

The problem with the regularization term in the previous section is that it gives each component in $\mathbf{x}$ the same weight. We have already argued that the first component of $\mathbf{x}$ cannot be zero. We also expect that the second component will be non-zero. Wavelet coefficients at smaller scales are expected to be zero, since these correspond with neighbouring model parameters having equal conductivities. Our wavelet basis, however, exhibits better minimum structure when the high resolution wavelet coefficients $x_i \in W_{-1}$ are sparse. Theorem 4 from Chapter 4 makes this more formal for any type of wavelet with p vanishing moments, recall that if f is locally smooth and p times differentiable, the theorem states that at scale $j+1$ the wavelet coefficients, localized where f is smooth and p times differentiable, are approximately smaller than those on scale j by a factor of 2^p. This result can be used to define a new regularization term (where we assume no signal extension):

$$\phi_m(\mathbf{x}) = \mu(x_2) + 2^p \sum_{i=3}^{4} \mu(x_i) + 2^{2p} \sum_{i=5}^{8} \mu(x_i) + 2^{3p} \sum_{i=9}^{16} \mu(x_i) + \cdots, \tag{5.12}$$

where we have made the sparsity constraint for high resolution coefficients more stringent and μ is the Ekblom measure. In shorter notation, where $N = \log_2(n_m)$ and j is the scale of resolution of the subspace W_j and $\mathbf{m} \in V_0$, then[4] the relation between the i-th component of $\mathbf{x}$ and j is for $i > 1$

$$j = \lfloor \log_2(i-1) \rfloor - N \tag{5.13}$$

and thus

$$\phi_m(\mathbf{x}) = \sum_{i=2}^{n_m} 2^{\lfloor \log_2(i-1) \rfloor p} \sqrt{x_i^2 + \epsilon}. \tag{5.14}$$

In order to be able to better compare between different parameterizations, we normalize with the Euclidean norm

$$\phi_m(\mathbf{x}) = \frac{\sum_{i=2}^{n_m} 2^{\lfloor \log_2(i-1) \rfloor p} \sqrt{x_i^2 + \epsilon}}{\left(\sum_{i=2}^{n_m} (2^{\lfloor \log_2(i-1) \rfloor p})^2\right)^{\frac{1}{2}}}. \tag{5.15}$$

[4]Recall that a three level DWT yields the following decomposition of subspaces $V_0 = V_{-3} \oplus W_{-3} \oplus W_{-2} \oplus W_{-1}$, where W_j are the subspaces which correspond to with wavelet coefficients at scale of resolution j and the subspace V_{-3} corresponds to the scaling coefficient.

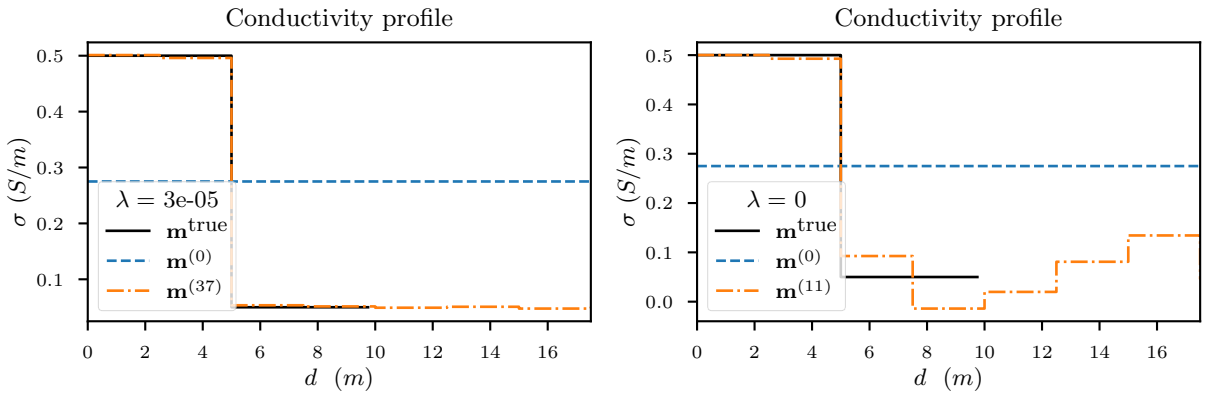


Figure 5.4: **(Left):** With scale-dependent regularization term. **(Right):** Without regularization.

The inversion was applied to the same conductivity profile with the new stabilizing term ϕ_m for a manually tuned regularization parameter and compared with the inversion without regularization ($\lambda = 0$). Visual inspection from Figure 5.4 shows that a regularized objective function results in a more stable result. The model misfits confirm this statement:

$$\phi_m(\mathbf{x}^{\text{true}}) = 0.126 \tag{5.16}$$

$$\phi_m(\mathbf{x}^{\lambda = 0}) = 0.207 \tag{5.17}$$

$$\phi_m(\mathbf{x}^{\lambda = 3 \times 10^{-5}}) = 0.125 \tag{5.18}$$

Hitherto, no noise was added to the synthetic data. Let us conclude that the regularization term with appropriate regularization parameter improves the stability of the inversion scheme. Additionally, the scale-dependent regularization significantly improves the outcome, when compared to equal regularization between different scales.

Other conductivity profiles

So far, only one simple conductivity profile was considered. In Figure 5.5, we examine two other profiles with the scale-dependent regularization strategy: Profile II is similar to the profile that was used earlier, only with switched electrical conductivities. Profile III is a somewhat more difficult conductivity profile with layers of 2.5 meters, 5 meters and a semi-infinite layer. The conductivities are 0.05, 0.5 and 0.05 S/m respectively.

The inversion scheme performs satisfactorily for Profile II, which is apparent in the model in Figure 5.5 and the model misfits are very alike:

$$\phi_m(\mathbf{x}^{\text{true}}) = 0.1636 \tag{5.19}$$

$$\phi_m(\mathbf{x}^{\text{Profile II, } \lambda = 2 \times 10^{-5}}) = 0.1633. \tag{5.20}$$

For profile III, a suitable regularization parameter was not immediately found through manual tuning. Therefore, different initial guesses were taken (randomly generated as described in Section 5.1.1) and in the various outcomes, negative conductivities were often noted, as in Figure 5.5 with $\lambda = 2 \times 10^{-6}$. In the next section, we will adjust the inversion scheme in order to avoid negative conductivities.

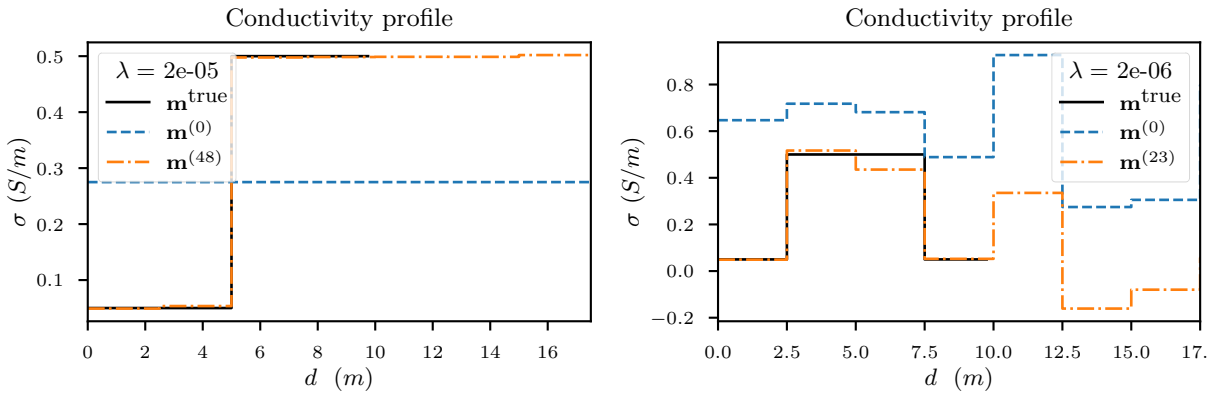


Figure 5.5: Outcomes with the scale-dependent regularization term. **(Left:)** Profile II **(Right)**: Profile III

5.1.4 Imposing positive conductivities in log-space

By applying a logarithmic transformation on the model parameters, an outcome with negative electrical conductivities is prevented. This is because the outcome is back-transformed to the normal space domain with an exponential transformation. The adjusted inversion scheme is presented in Figure 5.6. In the wavelet domain, we get

$$\mathbf{x} = W \log_{10}(\mathbf{m}) \tag{5.21}$$

and hence the objective function is

$$\min_{\mathbf{x}} \phi(\mathbf{x}) = \frac{1}{2}\|\mathbf{d} - K\left(10^{W^{-1}\mathbf{x}}\right)\|_2^2 + \lambda\phi_m(\mathbf{x}), \tag{5.22}$$

where $\phi_m(\mathbf{x})$ in Eq. (5.15) remains unaltered.

The data misfit gradient is different. The derivative of the residual vector

$$\mathbf{r} = \mathbf{d} - K\left(10^{W^{-1}\mathbf{x}}\right) \tag{5.23}$$

with respect to $\mathbf{x}$ is obtained componentwise

$$\frac{\partial r_i(\mathbf{x})}{\partial x_j} = \frac{\partial}{\partial x_j}\left(\sum_k K_{ik}(\mathbf{m})10^{y_k} - d_i\right) \qquad \text{where } y_k = \sum_l W^{-1}_{kl}x_l \tag{5.24}$$

$$= \sum_k K_{ik}10^{y_k}\log(10)\frac{\partial y_k}{x_j} \tag{5.25}$$

$$= \sum_k K_{ik}10^{y_k}\log(10)W^{-1}_{kj}, \tag{5.26}$$

where log is the natural logarithm and with the Hadamard product $\star$,

$$\frac{\partial \mathbf{r}(\mathbf{x})}{\partial \mathbf{x}} = \log(10)\left(K(\mathbf{m}) \star Y\right)W^{-1} \quad \text{where } Y_{ik} = 10^{y_k} \quad \text{for } i = 1, \cdots, m \tag{5.27}$$

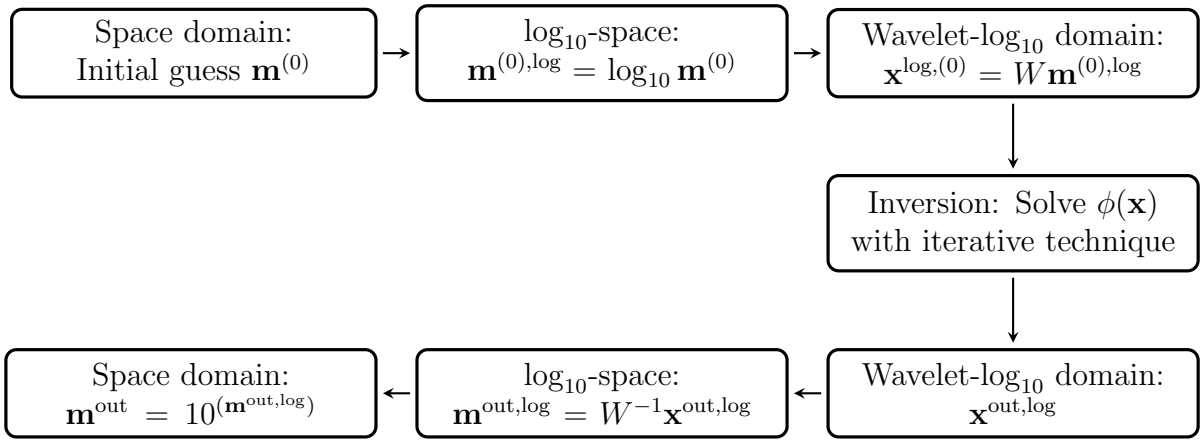


Figure 5.6: The inversion scheme in $\log_{10}$-space domain

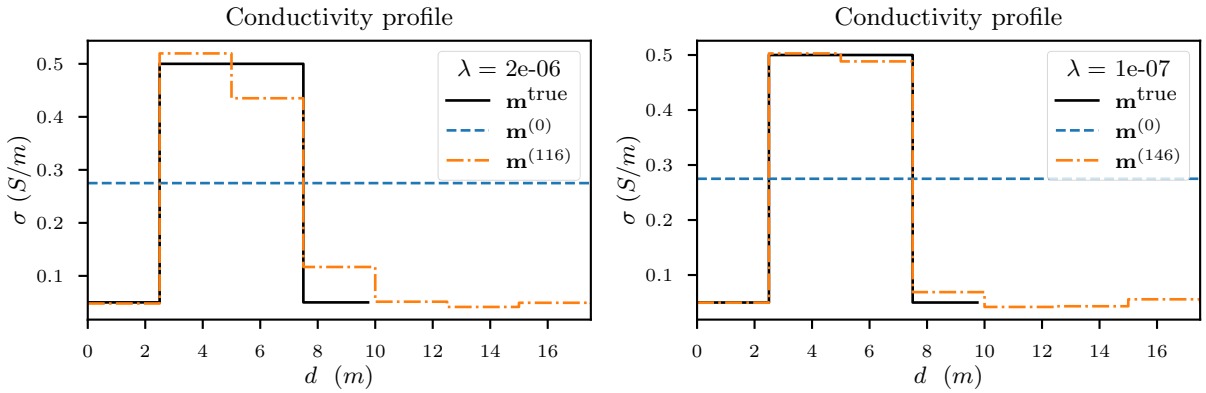


Figure 5.7: Outcomes with the inversion scheme with logarithmic transformation and scale-dependent regularization strategy. **(Left:)** Regularization parameter $\lambda = 2 \times 10^{-6}$, as in Figure 5.5. **(Right)**: Manually tuned 'optimal' regularization parameter $\lambda = 10^{-7}$.

This new inversion scheme is employed to the 3-layered conductivity profile (profile III), for which negative conductivities were observed in the previous scheme. Figure 5.7 shows the results with the same regularization parameter as in the previous section. The outcome indeed follows the real model better and the negative electrical conductivities have disappeared. Tuning the parameter to $\lambda = 10^{-7}$ produces an outcome that follows the outcome well. With the other conductivity profiles that were considered earlier, no significant differences were noticed.

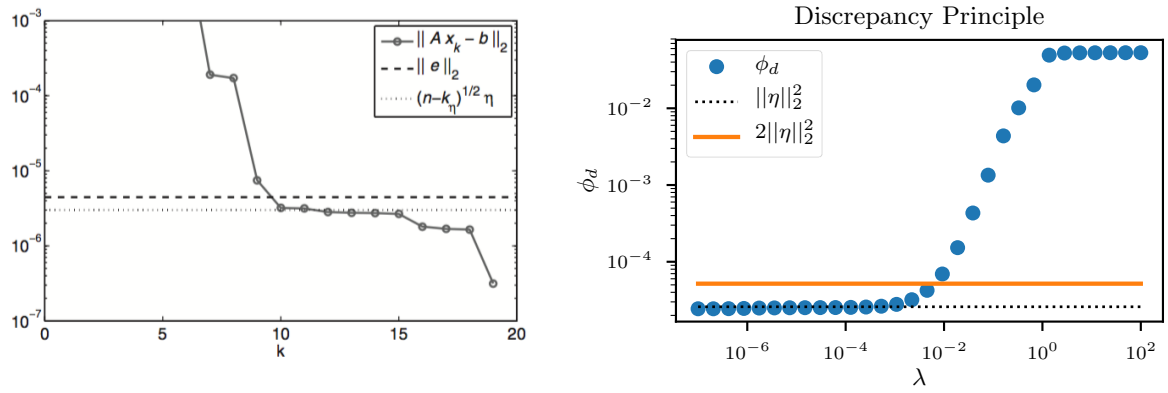
So far, the regularization parameter was tuned manually. In the next section, several automatic procedures for selecting an optimal regularization parameter will be examined.

5.2 Choosing the regularization parameter

The regularization parameter balances the importance of the data and model misfit terms in the objective function. In the inversion schemes considered above, the regularization parameter was manually tuned. This was possible, because the exact conductivity profile was known. In a real-world scenario, this is not the case and we wish to choose an optimal regularization parameter *without* prior knowledge of the exact solution. We discuss three common methods for choosing the regularization parameter. The first method, the discrepancy principle, is known in inversion literature [Hansen, 2010], however [Farquharson and Oldenburg, 2004] suggest that the 'automatic' L-curve criterion procedure performs equally well in iterative, minimum-structure, underdetermined inversion algorithms. The third method is more pragmatic. The regularization parameter will be decreased throughout the iterations, based on our own defined criteria. We will refer to this method as the λ-descent method. These methods are discussed in Section 5.2.1, 5.2.2 and 5.2.3 respectively.

5.2.1 The discrepancy principle

Recall that in Eq. (3.5) in Chapter 3, we have defined a measurement error vector $\boldsymbol{\eta}$. We know that the data misfit cannot (and should not) be lower than $1/2||\boldsymbol{\eta}||_2^2$. We can exploit this fact and find a lower bound on λ if an upper bound of $1/2||\boldsymbol{\eta}||_2^2$ is known[5].



(a) Discrepancy principle in truncated SVD. Figure retrieved from [Hansen, 2010].

(b) Discrepancy principle for Profile I with 16 parameters and constant λ.

Figure 5.8: The discrepancy principle

Morozov's *discrepancy principle* [Tikhonov et al., 2013], [Hansen, 2010] states that one should choose the regularization parameter λ such that the data misfit equals the discrepancy in the data, i.e. $\phi_d = ||\boldsymbol{\eta}||_2^2$. Since $\boldsymbol{\eta}$ is unknown and can only be estimated, it is safer to choose a slightly larger regularization parameter, such that $\phi_d = \alpha||\boldsymbol{\eta}||_2^2$ (this will become very clear in the next section about the L-curve criterion). In theoretical work, this method is popular, because it is possible to show for Tikhonov regularization with smoothness functional that the regularized solution converges to the exact solution as

[5]For the ease of notation, we will drop the 1/2 in $1/2||\boldsymbol{\eta}||_2^2$ in the remainder of this text.

$||\boldsymbol{\eta}||_2$ vanishes[6]. To our knowledge, no such proofs exist for our type of inversion scheme. Additionally, the major drawback (in both Tikhonov and our inversion scheme) is the difficulty to estimate[7] $||\boldsymbol{\eta}||_2$.

In a general setting [Hansen, 2010], the discrepancy principle is expected to behave as in Figure 5.8. In Figure 5.8a, a linear problem is considered which can be inverted via a direct method, based on truncation of the singular value coefficients. A truncation parameter k implies that only the k largest singular values are retained (the idea is that the small singular values correspond to the uncorrelated noise). Strong regularization corresponds with a small k and yields a large discrepancy on the data, while a large truncation parameter allows overfitting to happen. Overfitting means that the data is better fitted than what the quality of the data allows. The discrepancy principle plot in our inverse problem wil have the typical appearance as in Figure 5.8b, where the data misfit ϕ_d is plotted against the regularization parameter λ. The LIN approximation was used in both the data generation and inversion. Noise was added to the synthetic data (see below) and thus the exact error $||\boldsymbol{\eta}||_2^2$ is known. The solid line is the 'safe estimation' of that noise, with $\alpha = 2$.

Morozov's discrepancy principle is still used in geophysical inversion literature today ([Grasmair et al., 2015], [Oldenburg and Li, 2005]). In Section 5.2.4, we will examine if the principle can be included in our inversion scheme.

5.2.2 L-curve criterion

The L-curve criterion is a very intuitive tool for finding the 'optimal' regularization parameter, due to its graphical interpretation. The idea is to solve the inverse problem for a set of ℓ regularization parameters $0 < \lambda_1 < \lambda_2 < \cdots < \lambda_\ell$ and to log the data misfit and model misfit after conversion. Then, these misfits are plotted in log-log scale and are expected to appear as an L-shape. In Figure 5.9, a generic L-curve of an 'ideal' case is shown. The x and y-axis correspond respectively with the data misfit ϕ_d and model misfit ϕ_m.

The reason for the expected shape is intuitive. Consider the misfits $\phi_{d,k}$ and $\phi_{m,k}$ obtained after conversion with regularization parameter λ_k, where $1 < k < \ell$. Similarly, $\phi_{d,k+1}$ and $\phi_{m,k+1}$ are obtained with $\lambda_{k+1} > \lambda_k$. The larger regularization parameter λ_{k+1} gives larger weight to the model misfit in the objective function ϕ and the model misfit will, in general, be minimized more. It is expected that $\phi_{m,k+1} \leq \phi_{m,k}$, while $\phi_{d,k+1} \geq \phi_{d,k}$. Indeed, this λ_{k+1} generally yields an outcome with an ordered pair of misfits, located underneath and/or right with respect to $(\phi_{d,k}; \phi_{m,k})$. The vertical 'stick' of the L corresponds to outcomes where the model misfit is very sensitive to changes in the regularization parameter, while the horizontal part corresponds to outcomes where the data misfit is sensitive. The vertical part of the L generally corresponds with $\phi_d = ||\boldsymbol{\eta}||_2^2$. For vanishing regularization, outcomes with $\phi_d < ||\boldsymbol{\eta}||_2^2$ are obtained (overfitting). We recognize the importance of

[6] See Chapter 1 in [Tikhonov et al., 2013].

[7] For measurements with the EM34-3, the a measurement accuracy is 5%. However, there are also unmodelled effects (e.g. if one uses the LIN approximation) that will lead to an additional error. Additionally, coil misalignment also leads to a measurement error. We will denote $\boldsymbol{\kappa}$ for the total error. Especially for large intercoil distances, it is hard to estimate the error introduced by e.g. misalignment.

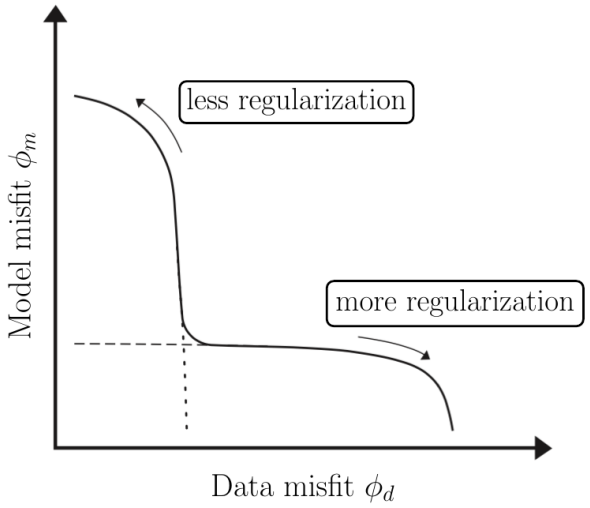


Figure 5.9: The generic form of the L-curve in log-log scale. Figure adapted from [Hansen, 1994].

taking $\phi_d = \alpha||\boldsymbol{\eta}||_2^2$ in Morozov's discrepancy principle. If the error estimation $||\boldsymbol{\eta}^{\text{est.}}||_2^2$ was such that $||\boldsymbol{\eta}^{\text{est.}}||_2^2 < ||\boldsymbol{\eta}^{\text{true}}||_2^2$, then one finds very different outcomes with very large model misfit (for vanishing regularization parameters). It is therefore safer to take $\alpha > 1$.

The L-curve divides the plane in two regions. It is impossible to construct any outcome that corresponds to a point below the L-curve. Note that the log-log scale emphasizes the L-shape. It is not guaranteed that the curve will exhibit the L-shape. However, in theoretical work [Hansen and O'Leary, 1993], there exists conditions that yield a guaranteed L-shape[8]. To our knowledge, no such theoretical work exists that is related to our inversion scheme. We will examine if we can use the L-curve criterion without the guarantee of its expected shape.

Corner selection

The optimal regularization parameter is the parameter that corresponds with the *corner* of the L-curve. There are two main approaches to determine the corner. The first approach is to connect the ordered pairs $(\phi_{d,k}, \phi_{m,k})$, for $k = 1, \cdots, \ell$ with spline curves. The corner is the point with maximal curvature. A method to map a discrete λ to the corner of the L-curve is provided in [Hansen and O'Leary, 1993]. The algorithm may sometimes mistake a local 'corner' (with larger curvature) for the global corner, as that can be the result of spline interpolation. When this issue occurs, a different spline (with e.g. different degree or order) can resolve this issue.

Nonetheless, we will rely on a more robust method: The Adaptive Pruning algorithm by [Hansen et al., 2007], as presented in Algorithm 3 in Appendix B. It should be able to

[8] If the following criteria are met with *Tikhonov regularization* with the ℓ_2-norms, then the L-curve is guaranteed to exhibit an L-shape. The conditions are (1) The error on the data $\mathbf{d}$ is 'white noise', i.e. the sequence is statistically uncorrelated; (2) The signal-to-nose ratio is reasonably large and; (3) The discrete Picard condition is satisfied. Without giving a formal treatment, it means that the data vector $\mathbf{d}$ is sufficiently smooth and not too noisy.

adaptively filter all the small local corners. Indeed, without a guarantee of the typical L-shape, we expect some local corners to be present in the curve. Therefore this robust corner method will be adopted. The idea is that the global corner can be easily found, by removing the right amount of points. The process must be adaptive, because the set of points that need to be removed so that the L-curve only has one global corner is unknown. Therefore, the algorithm will work in two stages. The first stage generates a sequence of pruned L-curves for which it locates the corner. The second stage is concerned with finding the best corner from the candidates generated in the first stage. In the belief that minimizing the data misfit is more important than minimizing the model misfit, all the candidates on the vertical part of the curve are removed. These candidates are found by calculating the slope. If the slope is greater than $\pi/4$, the candidate is removed. Secondly, only the candidates with a negative angle are retained. Finally, the smallest remaining regularization parameter λ is considered as the 'optimal' regularization parameter, that is the last candidate before reaching the vertical part.

5.2.3 Pragmatic method: decreasing the regularization parameter

In geophysical literature, the regularization parameter λ is often varied throughout the iterative minimization of ϕ. [Liu et al., 2017] uses this approach in combination with a wavelet stabilizer. In his paper, an initial value for λ of 1000 is chosen. The parameter λ is decreased with 10% if the data is misfit decreases slowly. If the regularization parameter λ equals 0.1, it will be kept constant throughout the subsequent iterations. There is no further specification as to what the precise criterion is for 'if the data misfit decreases very slowly'. Additionally, taking the lowest regularization parameter equal to 0.1 is an arbitrary decision. One can choose such a value to make it work on a synthetic data set, however, when no prior knowledge is available, there is no guarantee that this value is adequate. We will therefore not rely on such an arbitrary stop criterion.

Finding a suitable stop criterion for decreasing lambda will prove difficult. A too small λ will allow overfitting (see below), something that is difficult to detect, based on knowledge of the behaviour of ϕ_m and ϕ_d. A justified way to determine the lower limit for λ is on the basis of the discrepancy principle. However, it would still be beneficial if we find a strategy that does not use that information, because as we will see later, the discrepancy principle is not always that easy to interpret.

Decreasing the regularization parameter λ could be an interesting strategy, because it could lead to a computationally faster inversion scheme. A too low regularization parameter will allow excessive structure to build up in the model during the first iterations. Additional iterations will then be required to eliminate this excessive structure. It was therefore suggested (as in [Farquharson and Oldenburg, 2004]) that it may be more efficient to choose a fairly large *initial* regularization parameter λ_0 for the first iterations and then to decrease λ via a so-called 'cooling schedule'. In Section 5.2.6, we will develop and compare such λ-descent strategies.

5.2.4 The discrepancy principle in use

In this section, the same two-layered conductivity profile as in Section 5.1.1 is considered. The number of parameters is doubled to 16 and thus the thickness of the layers with unknown electrical conductivities is halved. To make the inversion more realistic, noise is added to the synthetic magnetic field data. A Gaussian distribution is created with a mean of one and a standard deviation of 10^{-2}, from which randomly drawn elements are multiplied with the synthetic data. We do not use additive noise, because the measured magnetic secondary field for an arrangement with intercoil distance $s = 1$ meters is only about 1.2×10^{-3} A/m, while it is 2.8×10^{-1} A/m for intercoil distance $s = 20$ meters for perpendicular with vertical dipole configuration (PERP) coil setting. Additionally, the measurement accuracy of the instrumentation is usually a percentage (e.g. in our experimental setting, this is 5% [Geonics, 2012]). First, for simplicity, data is generated and inverted via the LIN approximation, because in this way the discrepancy principle can be best understood. Afterwards, the discrepancy principle is employed on synthetic data from the exact model and inverted via the damped model. This latter setting contains noise due to unmodelled effects and it will appear that this considerably complicates the inversion problem.

For the case with the synthetic LIN data, the data misfit functional of the outcome after conversion $\phi_d(\mathbf{x}^{\text{out}})$ is plotted in terms of the regularization parameter λ. In this ersatz

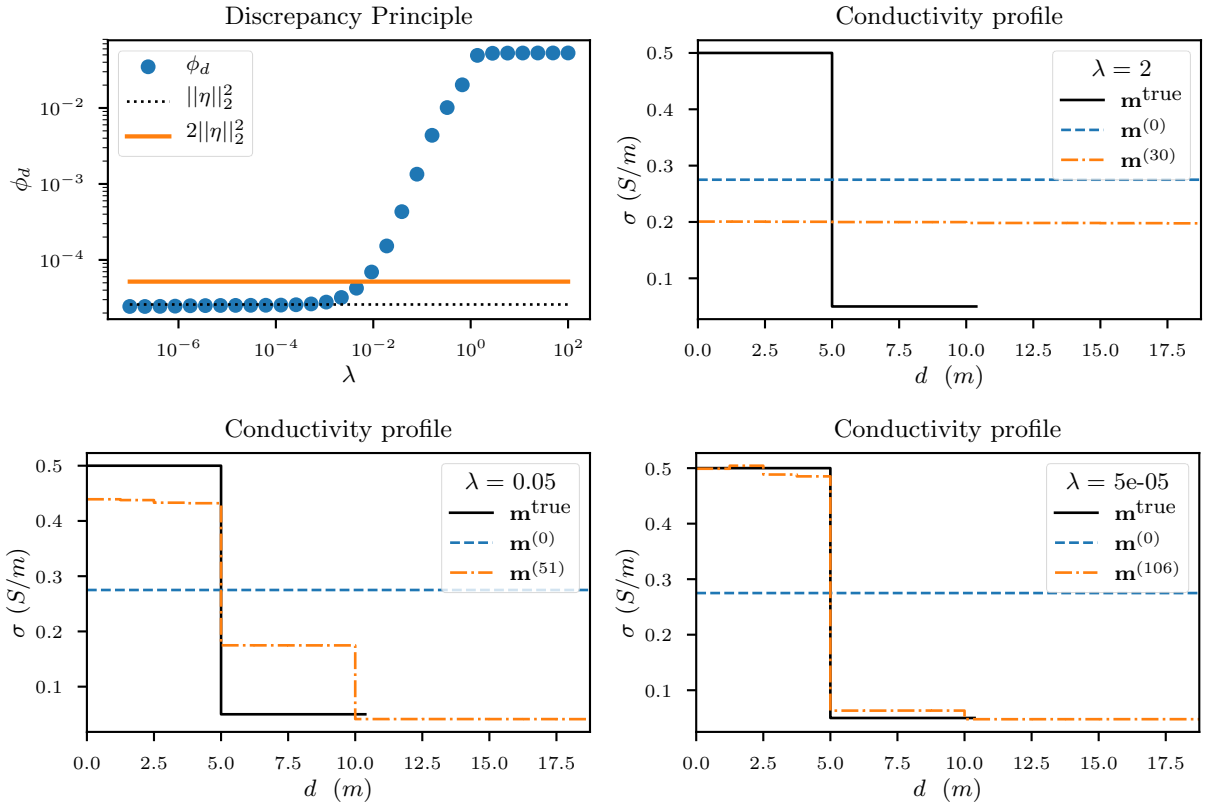
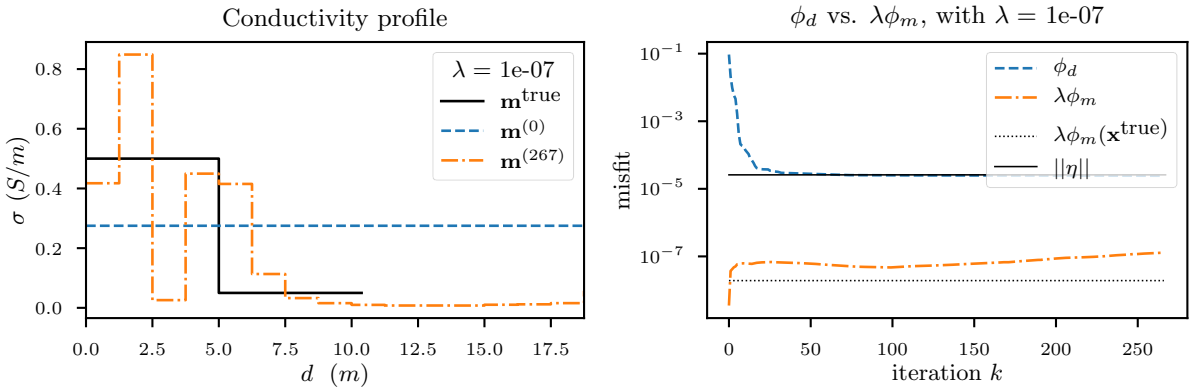


Figure 5.10: The discrepancy principle on synthetic data (with 1% noise) and inversion via LIN approximation. Three outcomes are shown for notably different regions of the graph.

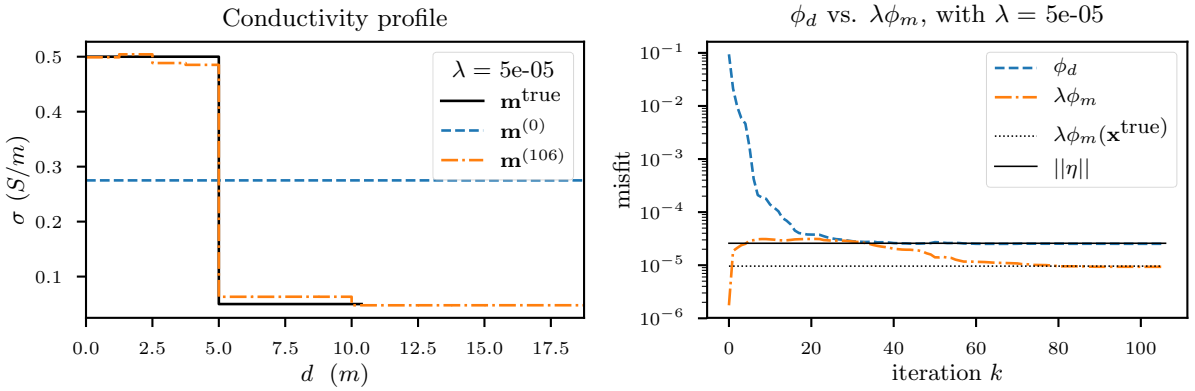
case, the noise on the data $||\boldsymbol{\eta}||_2^2$ is known and plotted (dashed line), together with $\alpha||\boldsymbol{\eta}||_2^2$ (solid line) where $\alpha = 2$. This choice for α is a suggestion in [Hansen, 2010]. The plot of the discrepancy principle is shown in Figure 5.10, accompanying three other outcomes after inversion for $\lambda = 2, 0.05$ and 5×10^{-5}. These regularization parameters correspond to three regions in the graph. The first profile ($\lambda = 2$) corresponds to the domain on the right, where the graph tends to be constant and the data misfit large. This coincides with constant solutions. Indeed, the regularization is strong and the constant conductivity profile has the absolute minimally possible structure. In the domain between 10^{-2} and 1, a lot of variation in the data misfit is observed. This seems like a domain where the interesting stuff is happening, although the outcomes are very similar. Such an outcome is shown in Figure 5.10 for $\lambda = 0.05$. The interesting domain (in terms of outcomes) is for $\lambda < 10^{-2}$, where the data misfit of the outcome is in of the order of the true noise. An example of an outcome in this region ($\lambda = 5 \times 10^{-5}$) is also shown. There is a good agreement between the true conductivity profile and the outcome after conversion.

At first glance, it seems good that for specific regularization parameters λ, the data misfit appears close to the real noise on the data. Note that there is no guarantee that this will always be the case. In fact, not all outcomes in the $\lambda < 10^{-2}$ domain are desirable. It is instructive to see how the data misfit and the model misfit evolve in terms of the number of iterations k. This is examined for 3 outcomes $\lambda = 10^{-7}, 5 \times 10^{-5}$ and 10^{-3} for which the data misfit is approximately equal to the noise: $\phi_d \approx ||\boldsymbol{\eta}||_2^2$. The results are shown in Figure 5.11. The conductivity profile of the first outcome (Figure 5.11a, $\lambda = 10^{-7}$) exhibits a complex structure solution. This is an indication of overfitting. On the plot of the misfits in terms of the iteration, we observe that the data misfit after conversion is actually slightly below the real noise. Observe that the model misfit is larger than the true complexity of the true solution (dotted line). The contribution of the minimum structure condition (i.e. regularization term) is too low and not significant in the minimization of the objective function ϕ. We conclude that the situation where $\lambda\phi_m \ll \phi_d$ yields poor outcomes. The second outcome (Figure 5.11b, $\lambda = 5 \times 10^{-5}$) is an acceptable result, the outcome follows the true conductivity profile closely. The misfits are much closer to each other (at the end *and* during the minimization process) than in the previous outcome. This is a desirable feature because both misfits have a significant contribution in the objective function ϕ during most of the iterative process. Note that for iterations $k > 60$, $\lambda\phi_m < \phi_d$ and $\phi_d(\mathbf{x}^{\text{out}}) < ||\boldsymbol{\eta}||_2^2$. In the final outcome (Figure 5.11c, $\lambda = 10^{-3}$), $\lambda\phi_m > \phi_d$ and we can conclude that the case of underfitting should likewise be rejected.

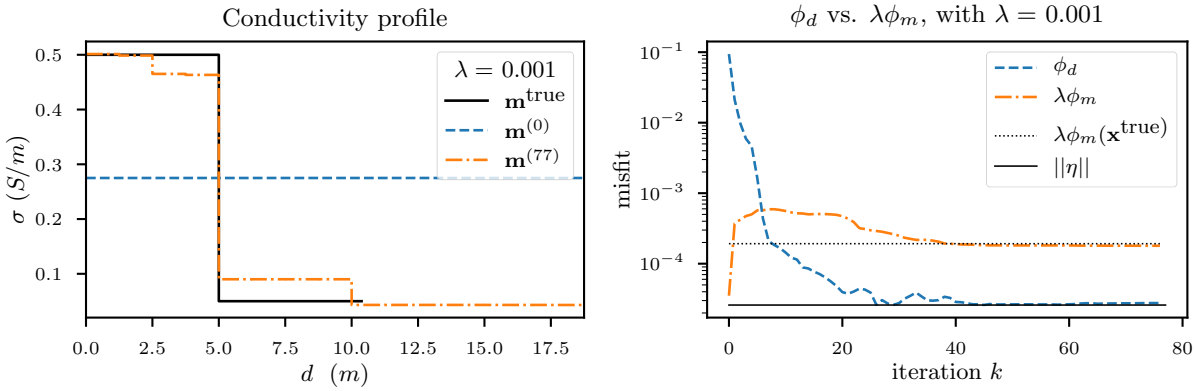
Inspired by the discrepancy principle, the same type of plot is generated for the model misfit instead of the data misfit. Instead of an increasing curve, a decreasing curve is obtained, see Figure 5.12 (left). The model misfit of the true conductivity profile $\phi_m(\mathbf{m}^{\text{true}})$ is known and plotted on the graph (dotted line). The overfitting for small regularization parameters λ is apparent. We can now put these graphs together, after normalization. Of course, this is not useful in practice, since it is impossible to normalise the misfits with the true conductivity profile. At least, something can be learned from the plot. We notice that the domain where the curves overlap is fairly wide. Therefore, the optimal regularization parameter cannot be determined, based on these types of plots. Moreover, we know from Figure 5.11c (inversion with $\lambda = 10^{-3}$) that the parameter λ for which ϕ_d crosses $\alpha||\boldsymbol{\eta}||_2^2$ is not optimal (Recall that this was the actual, original basic assumption



(a) Overfitting of the data.



(b) 'Perfect' balance between the misfits.



(c) Too much weight on the model misfit.

Figure 5.11: Further inspection of the discrepancy principle of Figure 5.10.

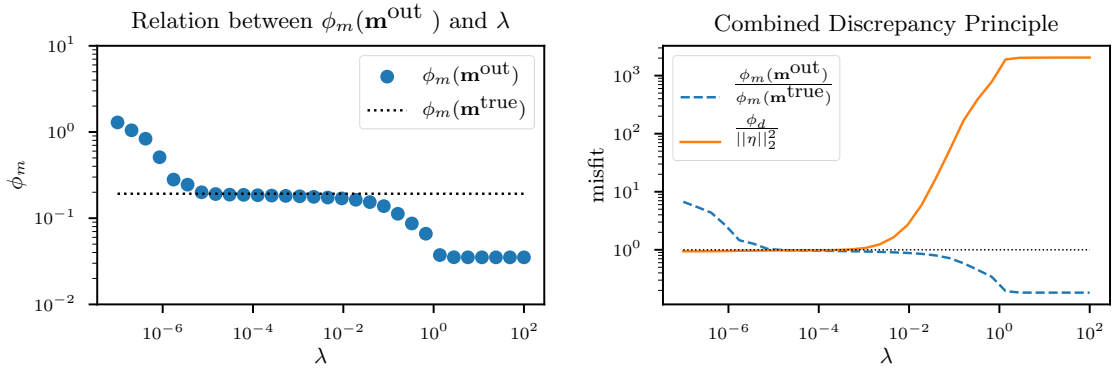


Figure 5.12: Misfits in terms of regularization parameter λ for synthetic data (1% noise) and inversion via the LIN approximation.

of Morozov's discrepancy principle). On the other hand, the discrepancy principle is very intuitive. If this behaviour occurs more generally, a parameter range (where the curves overlap) can be determined and used as input in other automatic regularization parameter selection algorithms, such as the L-curve criterion (see below). We now have to further investigate whether such behaviour also occurs in other settings.

We now turn to a more realistic case. The synthetic data is generated via the exact model and noise is added in the same way as in the previous setting. Inversion is now done using the damped model. The problem is more complex because the noise has two origins: measurement noise and noise from unmodelled interactions. The total noise is denoted with $\boldsymbol{\kappa}$, while measurement noise is still written with $\boldsymbol{\eta}$. Note that the $\boldsymbol{\kappa}$ is more relevant than $\boldsymbol{\eta}$ The behaviour of the discrepancy principle is different (see Figure 5.13), yet there are similarities. There is a domain for large lambda that corresponds with constant solutions. Secondly, for small λ, some lower bound on ϕ_d is reached. This lower bound is different from the bound in the previous case and lies in between $||\boldsymbol{\eta}||_2^2$ and $||\boldsymbol{\kappa}||_2^2$. In the combined discrepancy principle plot, we illustrate that the principle can still determine some good range for the regularization parameter λ, without prior knowledge of the errors. Within that range, the outcome for $\lambda \approx 2.5 \times 10^{-4}$ was selected for which $\lambda\phi_m(\mathbf{m}^{\text{true}}) \approx \phi_d(\mathbf{m}^{\text{true}})$, more specifically (up to three significant figures)

$$\phi_d(\mathbf{m}^{\text{true}}) = 7.09 \times 10^{-5}, \tag{5.28}$$
$$\phi_m(\mathbf{m}^{\text{true}}) = 0.429. \tag{5.29}$$

There is a good agreement up to 10 meters (see Figure 5.13), but a small discrepancy is noticeable for the semi-infinite layer. We cannot report a regularization parameter that yields a significantly better agreement with the true conductivity profile. Following Morozov's discrepancy principle, the regularization parameter that crosses the total noise $||\boldsymbol{\kappa}||_2^2$ should be considered. Here, this is for $\lambda = 9.2 \times 10^{-3}$. This outcome has the required minimal structure, but the electrical conductivities deviate largely from the true values. This outcome is not reported.

The outcomes for which $\phi_d < ||\boldsymbol{\kappa}||_2^2$ should not be viewed as cases with overfitting. In

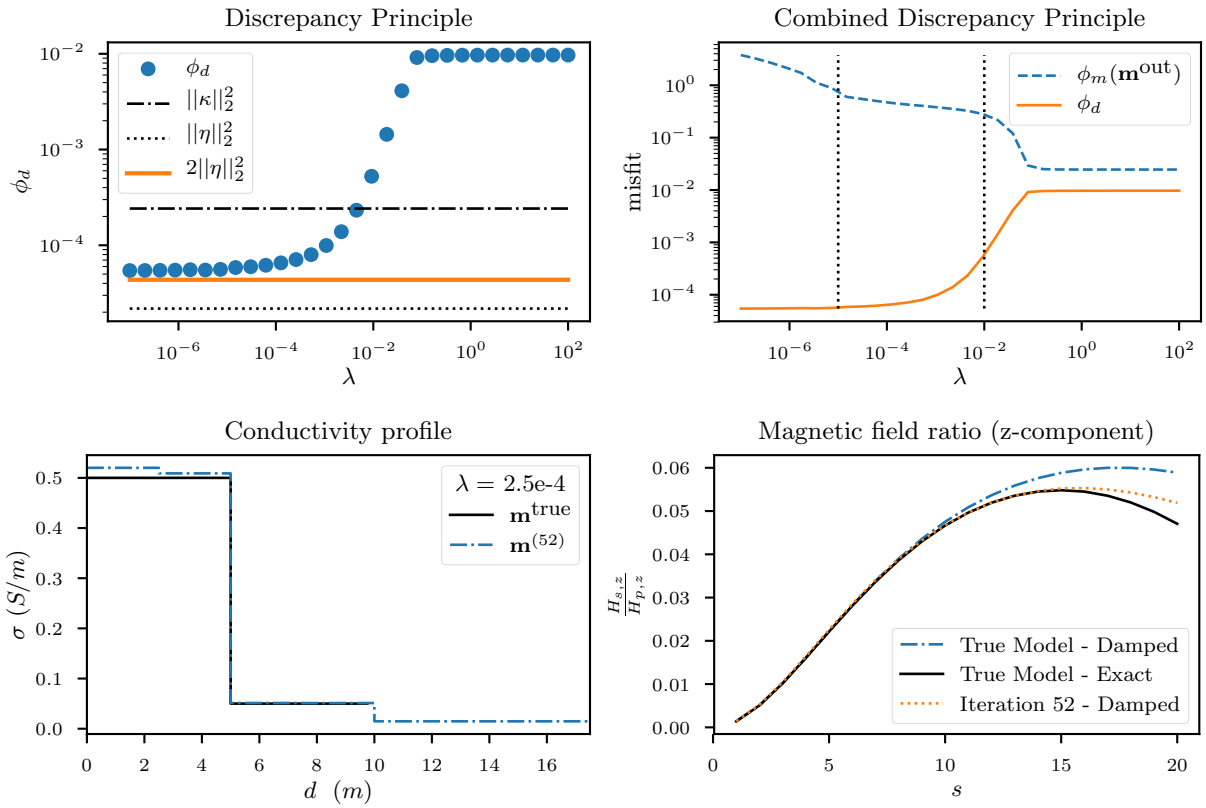


Figure 5.13: The discrepancy principle on synthetic data (with 1% noise) from the exact model and inversion via the damped model for the conductivity profile $\mathbf{m}^{\text{true}}$, together with the outcome for $\lambda = 2.5 \times 10^{-4}$.

Figure 5.13, the magnetic field ratio for the z-component is plotted[9]. This shows the response of the true model for the exact and the damped model. It also shows the response of the damped model of the outcome, obtained after 52 iterations. This illustrates the 'proof of concept': Our inversion scheme pushes the response of the outcome from the true response towards the response of the true conductivity profile, generated via the model used for inversion (i.e. damped model). This plot clearly illustrates the regularization's raison d'être.

5.2.5 The L-curve criterion in use

The L-curve criterion is applied to exactly the same inverse problem as in the previous section. First, it is examined whether the method works for the simple case with LIN data. Then, the method is applied on the more intricate problem, where the synthetic data from the exact model is inverted with the damped model.

First, the L-curve method is applied to a large range or regularization parameters $\lambda \in [10^{-7}, 10^{2}]$, see Figure 5.14. The L-curve is very similar to the generic shape of the L-curve. The Adaptive Pruning algorithm selects $\lambda = 3 \times 10^{-4}$ as the corner. The outcome is very sensitive to the regularization parameter. Therefore, a smaller range on λ is considered.

[9]Recall, also the ρ-component is measured, but this does not illustrate the point so clearly.

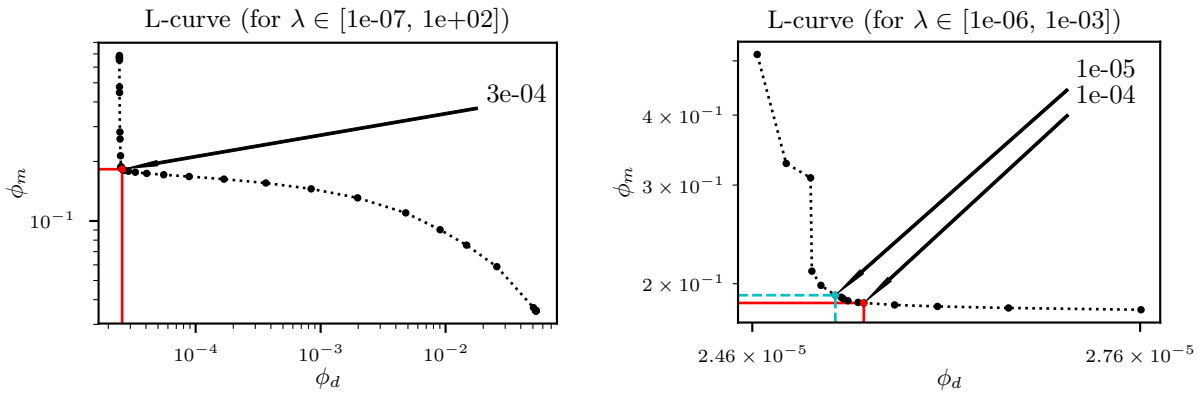


Figure 5.14: L-curve for data generation and inversion via the LIN approximation. Solid lines are corner selection by the Adaptive Pruning algorithm, while the dashed line is a corner selected by eye.

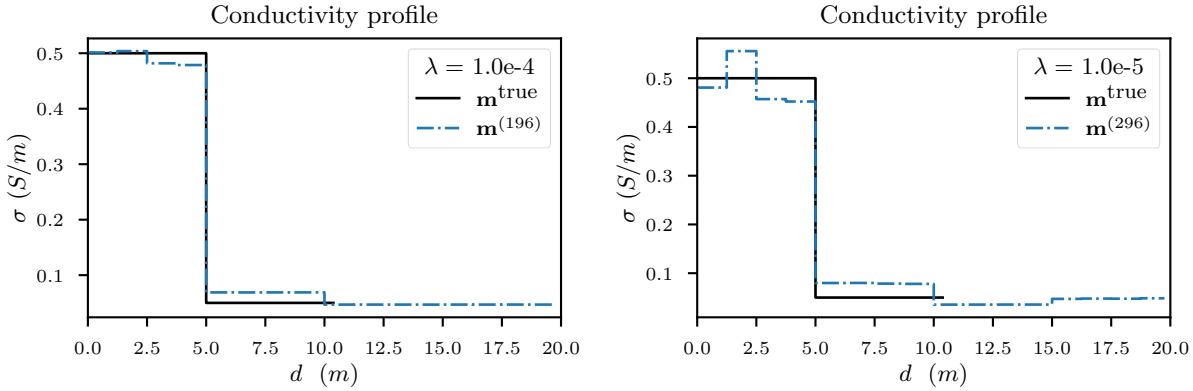


Figure 5.15: Conductivity profiles for LIN data, λ selected via L-curve criterion with Adaptive Pruning algorithm.

The choice of the bounds of that range is inspired by the combined discrepancy principle. The 'zoom' on the corner of the L-curve is as well presented in Figure 5.14. The corner selection method picked $\lambda = 1 \times 10^{-4}$. This is not really what is visually considered as the true corner. This choice for the rightmost candidate is due to the following philosophy: The rightmost corner candidate is selected for which going to the next candidate yields a larger increase in the model misfit than decrease in the data misfit, provided that the L-curve is convex for that candidate. When no such point exists, the leftmost corner is selected. We will compare this rightmost candidate with a $\lambda = 1 \times 10^{-5}$ (dashed line), which we selected by eye. The conductivity profiles are shown in Figure 5.15. In this case, that philosophy proves useful. Our manually picked $\lambda = 10^{-5}$ exhibits excessive structure at small depths d.

The L-curve for the synthetic data generated by the exact model is more difficult to interpret. Consequently, the corner selection is also more careful. In a straightforward fashion, the L-curve method is applied to the range $\lambda \in [10^{-7}; 10^{2}]$, as with the LIN data. The resulting L-curve and conductivity profile corresponding to the corner ($\lambda = 3 \times 10^{-5}$) are shown in Figure 5.16. At first sight, the L-curve resembles the curve as with the

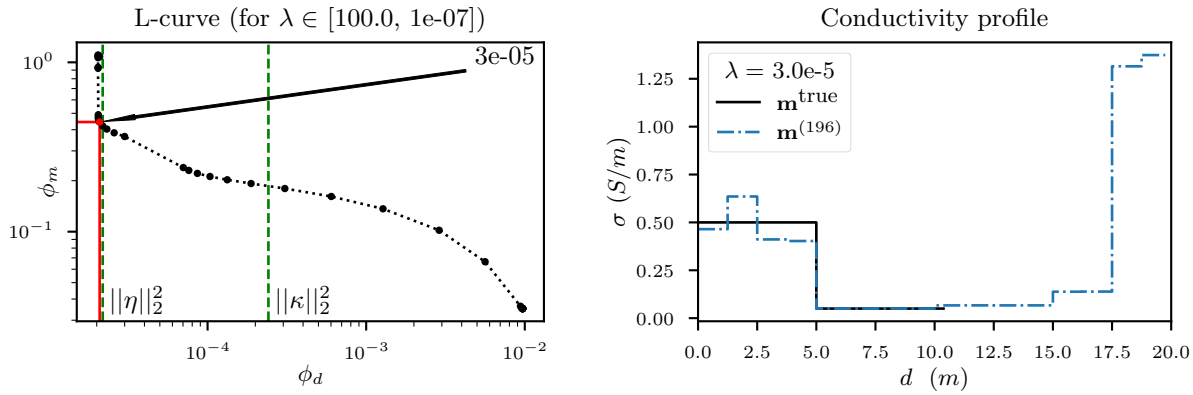


Figure 5.16: L-curve for data generation via Wait and inversion via Damped.

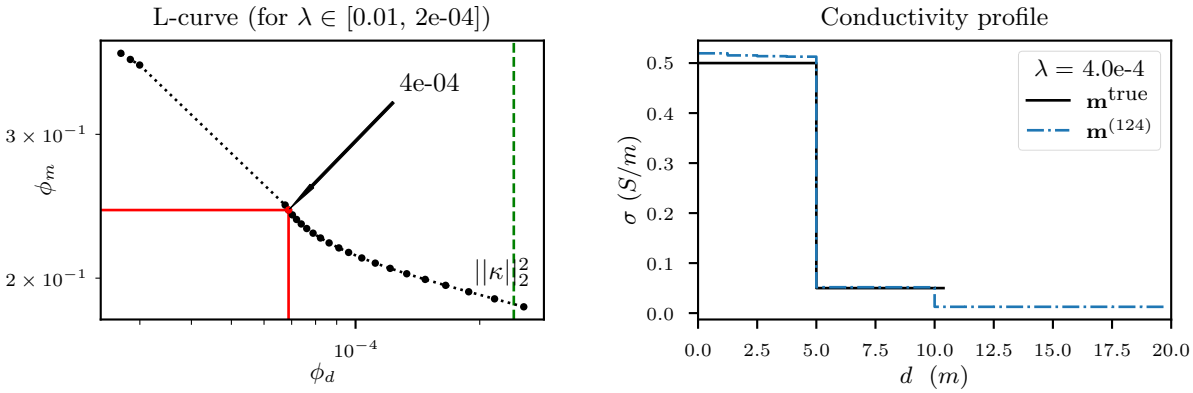


Figure 5.17: L-curve for data generation via the exact model and inversion via the damped model for a reduced range of λ.

LIN data, yet a small cusp near $\phi_d = 10^{-4}$ is observed, together with a higher density of misfit pairs. Accordingly, something interesting happens there. The conductivity profile associated with the corner selection is failing, because the regularization parameter is too small. Note that the typical vertical 'stick' of the L can be recognized and coincides with $||\boldsymbol{\eta}||_2^2$, as mentioned earlier.

A straightforward application of the L-curve criterion does not lead to the optimal outcome for this conductivity profile. As a next step, the range of the L-curve is reduced, based on the combined discrepancy principle. In this case (see previous section), an optimal λ should be between 10^{-5} and 10^{-2}. However, this will not lead to an acceptable outcome either. One really has to remove the leftmost corner, before the adaptive pruning algorithm can detect a second (local) corner. Indeed, adjusting the ranges will lead the Adaptive Pruning algorithm to select a corner in the cusp. We throw away all the data with $\phi_d < \alpha||\boldsymbol{\eta}||_2^2$ with $\alpha = 1.2$ (we cannot use $\alpha = 2$, because too much data will be thrown away). The L-curve for this limited range, with corner selection, and the corresponding result, is given in Figure 5.17. This outcome is equivalent to the best outcome we have already observed for this setting. There is again that small discrepancy in the second layer, staring at 10 meters.

In conclusion, the L-curve criterion performs successfully in the simplest case with LIN data. It can be used as an automatic selection method. When unmodelled effects are considered ($||\boldsymbol{\eta}|| \neq ||\boldsymbol{\kappa}||$), then there is no guarantee that the Adaptive Pruning algorithm will select a λ from the right corner. After some adjustments, it is capable of selecting the 'best you can get' outcome with our current inversion scheme.

5.2.6 λ-descent regularization in use

In this section, we summarize three types of λ-descent strategies. First, a simple regularization strategy is examined. It is the combination of the following criteria: If the relative reduction of the objective function ϕ is smaller than 10^{-5}, then the regularization parameter λ is decreased with 90%. The critical part is to define a stopping rule on the descent of λ. The above results inspired us to set the following criterion: The regularization parameter is only lowered *if* $\phi_d \leq 0.1\lambda\phi_m$. This is because we have understood that the best regularization parameter was for outcomes where ϕ_d was in the same order of ϕ_m (and often when ϕ_m was slightly larger than ϕ_d).

In order to keep things brief, we immediately jump to the case where data is generated via the exact model and inverted via the damped model, because in the case of the LIN approximation, the L-curve criterion seems to provide us with a good algorithm. As we have seen, the challenge really lies in finding a regularization scheme that can handle the unmodelled effects. We again work with the same profile and parametrization, in order to make a good comparison with the earlier methods. In our simulation, we have performed a sweep over 15 different initial regularization parameters ranging from 10^{-3} up to 10^{2}. The data reveals that for every initial parameter λ_0, there corresponds a very different trajectory in the model space. With this strategy, all regularization parameters at the final iteration λ_{out} are different. If the several final regularization parameters λ_{out} would be similar, then this could be a sign that no other automatic regularization parameter selection method would be required.

With the descent strategy outlined above, various different outcomes are obtained. Some outcomes are acceptable and the L-curve criterion would be able to determine the best outcome. However, some unexpected behaviour was observed that leads us to improve the decent strategy. In Figure 5.18, an outcome after conversion is shown for $\lambda_0 = 0.31$, together with the evolution of both misfits. Note that we have introduced $\tilde{\phi}_m$ as shorter notation for $\phi_m(\mathbf{x}^{\text{true}})$. The regularization parameter did not decrease throughout the iterative process and we have the typical situation where both misfits converge to a value of the same order. It is the case where the regularization parameter is too strong. The reason is that the total objective function ϕ always decreased with at least a relative reduction of 10^{-5}, before the regularization parameter could be decreased with 90%. In a better descent strategy, the relative reduction of the data misfit ϕ_d should be considered instead of the reduction of the objective function ϕ.

In a second strategy, we examine the effect of setting the regularization parameter (after the second iteration) equal to

$$\lambda_k = \frac{\phi_d(\mathbf{x}^{k-1})}{\phi_m(\mathbf{x}^{k-1})}. \tag{5.30}$$

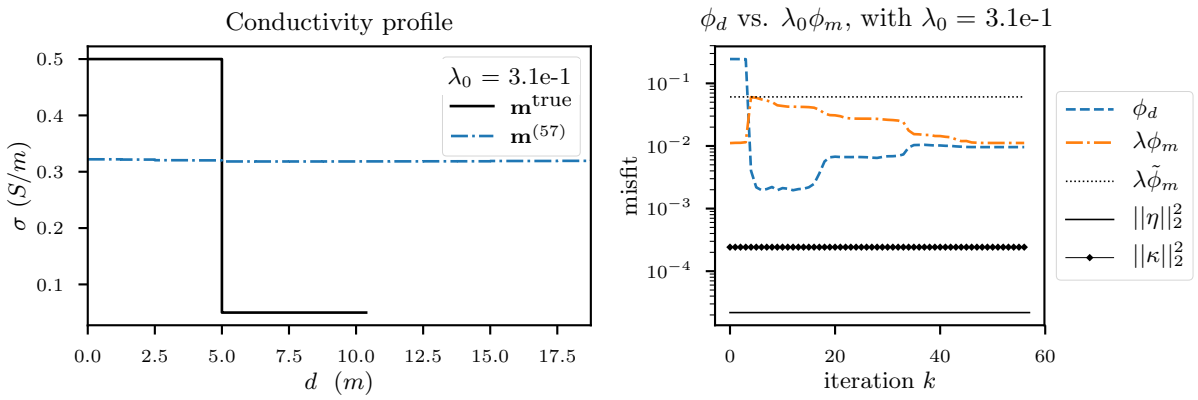


Figure 5.18: Example of an outcome where the descent strategy with criterion on the objective function ϕ was considered. $\tilde{\phi}_m$ is shorter notation for $\phi_m(\mathbf{x}^{\text{true}})$.

By doing so, the regularization parameter 'lags' the ideal balance where $\phi_d \leq \lambda\phi_m$. In Figure 5.19 (left), the evolution of λ throughout the iterative process is shown. We can divide the range of λ in two regions. For strong regularization, λ first evolves towards the value 6.9. These cases generally yield a constant outcome. Indeed, the regularization is too strong. The second range also quickly evolves towards a 'metastable' value for λ around 3×10^{-4}. Note that conversion was obtained more quickly than in other settings. The maximal number of iterations was less than 30. An example of such an outcome for $\lambda_0 = 0.028$ is shown in Figure 5.20. The outcome follows the true conductivity profile closely, except for the conductivity at a depth starting 18 meters, which translates in a relatively large model misfit. The model misfit is approximately 10 times bigger than the model misfit of the true conductivity profile. The regularization parameter at the final iteration was $\lambda_{\text{out}} = 2.0 \times 10^{-5}$. The problems with this strategy are the following: (1) it is hard to find a stopping rule that stops the λ from decreasing, (2) because both misfits are always from the same order (for k large enough), excessive structure can build up in the model. The noise in Figure 5.19 is a remnant of that excessive structure. This confirms what we have observed earlier, it is better to have ϕ_m slightly larger than ϕ_d. Indeed, the following choice

$$\lambda_k = \alpha \frac{\phi_d(\mathbf{x}^{k-1})}{\phi_m(\mathbf{x}^{k-1})}, \quad \alpha > 1 \tag{5.31}$$

could resolve this issue. However, tests did not produce significantly better results, which are therefore not reported.

The third type of strategy combines knowledge obtained from the discrepancy principle as stopping rule and the minimal iterations required for conversion from the lagging strategy. Thus the λ-descent is the same as in the previous strategy, but it is stopped when it reaches some value based on the discrepancy principle. Indeed, we have discussed that pinpointing this value is difficult and it may be somewhat arbitrary. Here, $\lambda_{\text{min}} = 10^{-3}$ is taken. The behaviour of λ in terms of iterations k is shown in Figure 5.19 (right). The range of λ can again be divided into two domains. The interesting domain is for $\lambda \leq 0.01$. From the plot, it is immediately apparent that the number of iterations is much larger

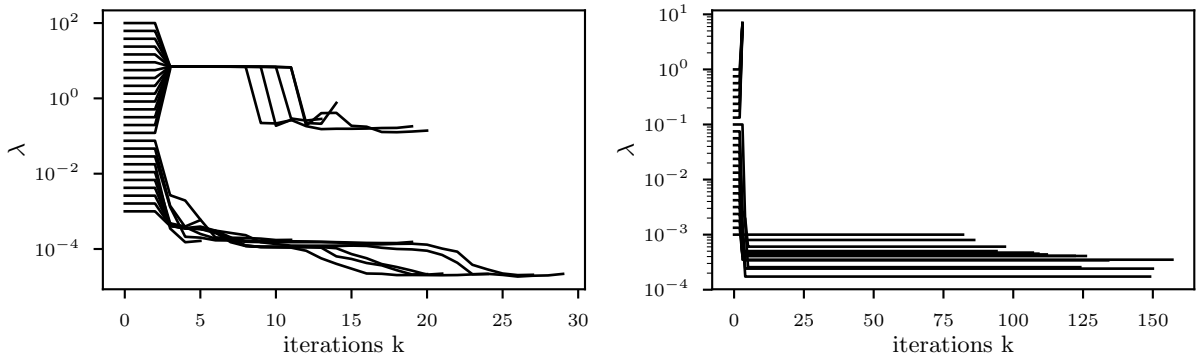


Figure 5.19: Evolution of λ throughout the iterative process (**Left**:) for the lagging strategy, (**Right**:) for the stop criterion based on the discrepancy principle.

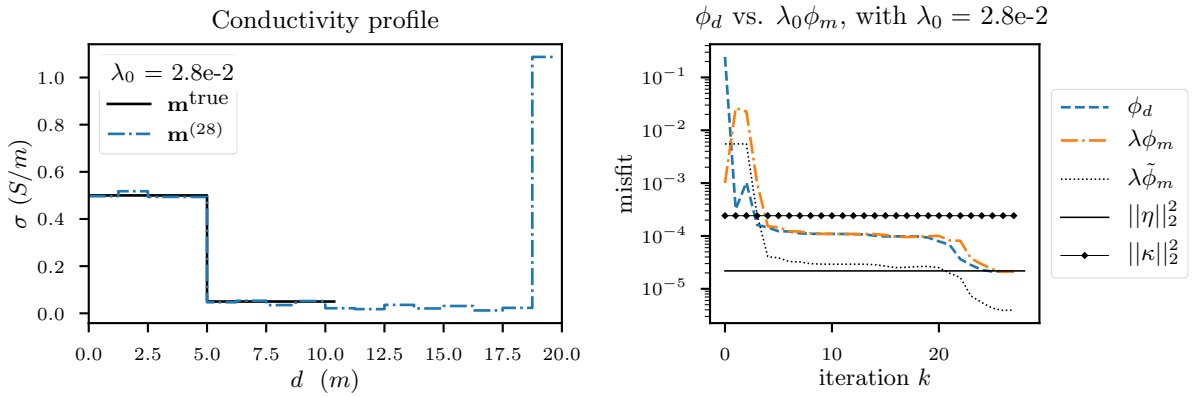


Figure 5.20: Example of an outcome with the lagging strategy. $\tilde{\phi}_m$ is shorter notation for $\phi_m(\mathbf{x}^{\text{true}})$.

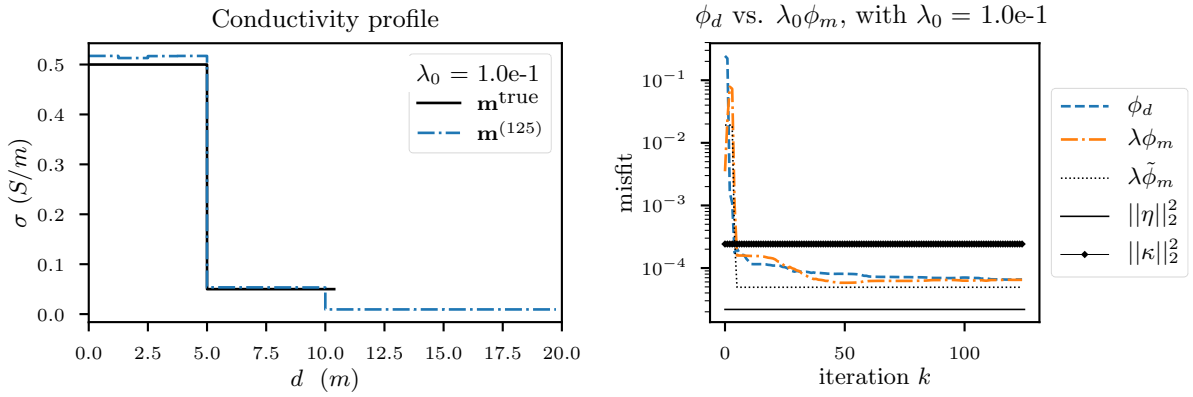


Figure 5.21: Example of an outcome with information from the discrepancy principle in the stop criterion. $\tilde{\phi}_m$ is shorter notation for $\phi_m(\mathbf{x}^{\text{true}})$.

than in the previous lagging strategy. That beneficial characteristic is thus lost. Note that there are final regularization parameters λ_{out} that are smaller than 10^{-3}. This is

because $\lambda_{\min} = 10^{-3}$ is only a stopping rule, this means that if a current $\lambda > \lambda_{\min}$, that it can still be decreased.

In conclusion, 'stand-alone' descent strategies do not allow to find the optimal outcome, because it depends on the initial regularization parameter λ_0, however it can be used in combination with the L-curve criterion or discrepancy principle. In particular, the lagging strategy is potentially useful in practise, because it reduces the number of iterations and thus the computation time. The combination with the discrepancy principle is quite stable, but the number of iterations is again large.

5.3 Improving the stabilizer

We have seen positive effects from our stabilizer so far. In this section, the stabilizer's impact is examined in more detail. For simplicity, we return to an 8 dimensional model space, because we will now look at the behaviour of each parameter.

5.3.1 Evaluating an outcome

As before, the optimal regularization parameter is selected via the L-curve criterion for this analogue problem with 8 model parameters, where the same conductivity profile was considered again (for a better comparison). The outcome after 74 iterations (Figure 5.22) is similar to the outcome from the previous sections, which showed a minor discrepancy for the lowest layer. Figure 5.22 also shows the wavelet transform of the true conductivity profile and the outcome after conversion. The sparse representation in the wavelet domain immediately becomes apparent. From this wavelet representation, it is clear that the scaling and first wavelet coefficient deviate from the true profile. Remember from wavelet theory, that the first two coefficients have the greatest effect on the outcome, for example, the first wavelet coefficient determines the total energy of the outcome. We must then question the inversion scheme and examine whether the solution *can* converge to the true profile. It will not be obvious anyway, because the scaling coefficient and the coarsest scale wavelet coefficient have an effect on all other coefficients.

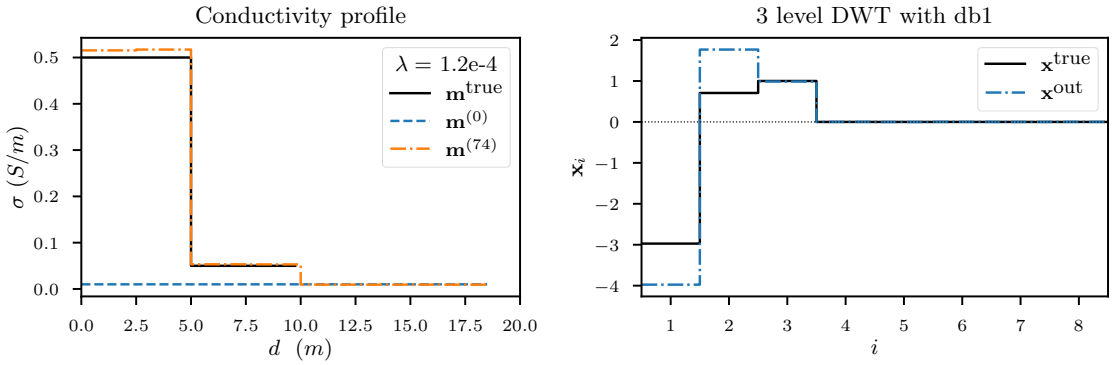


Figure 5.22: Outcome after conversion in both space and wavelet domain. Data generated by the exact model (+ 1% noise) and inverted via the damped model.

A visual inspection does not allow us to assess the inversion scheme. Let us examine the different terms in the objective function:

$\phi(\mathbf{x}^{\text{out}}) = 1.23 \times 10^{-4}$	$\phi_d(\mathbf{x}^{\text{out}}) = 6.55 \times 10^{-5}$	$\phi_m(\mathbf{x}^{\text{out}}) = 0.459$
$\phi(\mathbf{x}^{\text{true}}) = 2.85 \times 10^{-4}$	$\phi_d(\mathbf{x}^{\text{true}}) = 2.42 \times 10^{-4}$	$\phi_m(\mathbf{x}^{\text{true}}) = 0.338$

The objective function of the true profile is not zero. This is due to noise, but also because the stabilizer assigns a non-zero value to this conductivity profile. Indeed, a zero model misfit would mean the null solution, where no structure is present. However, note that $\phi_d(\mathbf{x}^{\text{out}}) < \phi_d(\mathbf{x}^{\text{true}})$ and $\phi_m(\mathbf{x}^{\text{out}}) > \phi_m(\mathbf{x}^{\text{true}})$, from where it can be deduced that the regularization parameter is too small. There is over-fitting and too complex structure. Simply increasing the regularization parameter is not the solution. How large should this parameter be? The fact that we visually opt for this outcome (also after manual tuning), indicates that the profile for the 'optimal' regularization parameter λ does not generate a visually good agreement with the true conductivity profile. If that is the case, can the stabilizer still be improved?

Figure 5.23 shows the z-component of the magnetic field ratio of the outcome after conversion (the ρ-component is not reported, but was considered in the analysis). We have already explained that a perfect inversion scheme obtains a conductivity profile where the data generated from that profile overlaps with the data generated with the forward model on the true profile that is used in the inversion (in this scheme: the damped model). With this reasoning, the regularization parameter λ is tuned so that the outcome after conversion overlaps more or less with the data from the damped model (Figure 5.23).

It is crucial to review this outcome and to understand why the scheme does not yield the outcome with the best agreement, given the optimal λ in terms of data misfit. In Figure 5.24 the solution for $\lambda = 4.5 \times 10^{-3}$ is shown in both space and wavelet domain. A typical stairlike behaviour is recognised in the space domain. This behaviour was also observed in many other inversions for various regularization parameters and other settings. It is a recurring phenomenon. Would the stabilizer be accountable for this behaviour? Note that the stairlike behaviour is small enough and can at least be considered as a relatively good agreement.

In the wavelet domain, the scaling coefficient and the second and fourth wavelet coefficient deviate from the true conductivity profile (coefficients 1, 3 and 5 in $\mathbf{x}$), with values -2,769, 0,8067 and 0,007736 respectively, instead of -2,927, 1 and 0 for the true conductivity profile. The second wavelet coefficient of the outcome is smaller than what it should be. This may be a consequence of the scale-dependent regularization, since the regularization strategy adds 2^p times more weight to the second coefficient than on the first coefficient (p is the number of vanishing moments of the wavelet). If this scale-dependent regularization scheme is responsible for this stairlike behaviour in the model space, the stabilizer should be adjusted in order to achieve a proper result.

It can further be noted that the model misfit for this stairlike outcome has a smaller

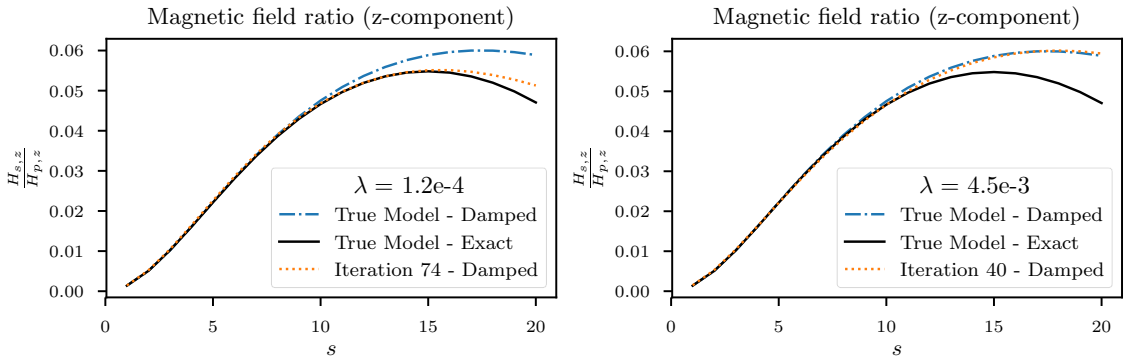


Figure 5.23: Response for a tuned regularization parameter λ, **(Left:)** based on visual overlap with true conductivity profile, **(Right:)** based on visual overlap with the response.

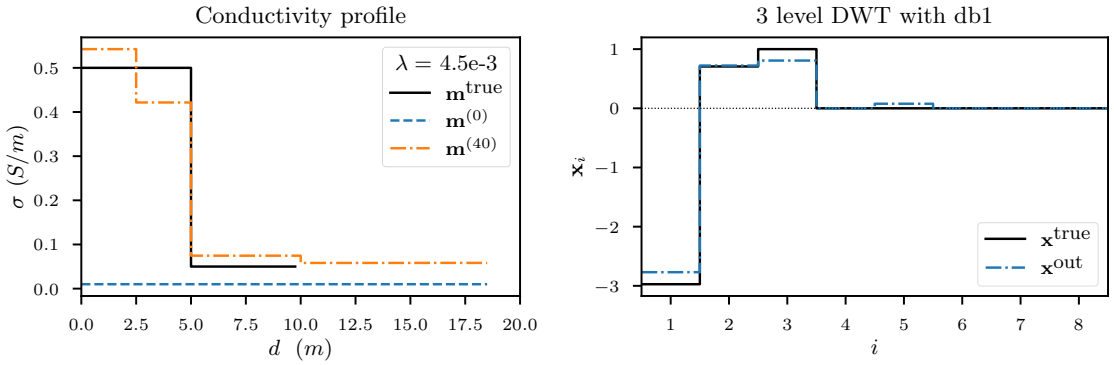


Figure 5.24: Outcome in space and wavelet domain for regularization parameter λ such that the data is close to the data from the true conductivity profile via the damped model.

model misfit than the true outcome, although that difference is small:

$$\begin{aligned}\phi(\mathbf{x}^{\lambda=4.5\times10^{-3}}) &= 1.707 \times 10^{-3}\\ \phi_d(\mathbf{x}^{\lambda=4.5\times10^{-3}}) &= 2.324 \times 10^{-4}\\ \phi_m(\mathbf{x}^{\lambda=4.5\times10^{-3}}) &= 0.326\end{aligned} \tag{5.32}$$

This means that our stabilizer promotes this outcome more than the structure in the true profile.

Some alternatives will be set out in the following sections. In Section 5.3.2, the inversion scheme will be extended with a second optimization problem. Firstly, the outcome of the current inversion scheme will serve as an input to choose a new regularization vector or scale dependent regularization approach. Secondly, in Section 5.3.3, the output of the current optimization problem will be used to threshold the outcome and to set bounds in a second optimization problem. Thirdly, in Section 5.3.4, the effect of other measures or focusing functions on the performance of the scheme will be examined.

5.3.2 Adaptive regularization vector

The inversion scheme is easily modified by simply extending the current scheme with a second optimization problem (Figure 5.25). The first optimization problem is the minimization of the objective function with the scale-dependent regularization in the misfit functional. Then, it is examined if the outcome of that optimization problem has stairlike or close to inappropriate behaviour. When that it the case, a new model misfit functional ϕ_m is proposed with a new scale-dependent regularization vector.

The scale-dependent regularization in the second optimization problem is initially the same as in the first stabilizer as described in Equation (5.15). However, if a wavelet coefficient in the outcome $\mathbf{x}^{\text{out1, log}}$ is significantly larger than ϵ (ϵ is 10^{-4} here, so let us put $x_i > 10^{-3}$ as a threshold) *and* the wavelet coefficient of level of resolution $n+1$ is less than twice a wavelet coefficient of level of resolution n (thus at the coarser scale), then the weight of the regularization vector for this wavelet coefficient of level $n+1$ will be adjusted to the same weight of level n. In this form, the over-pushing to zero for a coefficient at a smaller scale will be compensated. It will be pushed to zero as much as the coefficients on a coarser scale.

To make the idea more clear, assume that the third component of the following regularization vector $\mathbf{s}^{\text{scale-dependent}}$ forces the second wavelet coefficient (i.e. the third coefficient in $\mathbf{x}$) to zero too much, then the adaptive regularization vector is $\mathbf{s}^{\text{apdative}}$:

$$\mathbf{s}^{\text{scale-dependent}} = \begin{bmatrix} 0 \\ 2^p \\ \boxed{2^{2p}} \\ 2^{2p} \\ 2^{3p} \\ 2^{3p} \\ 2^{3p} \\ 2^{3p} \end{bmatrix} \quad \rightarrow \quad \mathbf{s}^{\text{adaptive}} = \begin{bmatrix} 0 \\ 2^p \\ \boxed{2^{p}} \\ 2^{2p} \\ 2^{3p} \\ 2^{3p} \\ 2^{3p} \\ 2^{3p} \end{bmatrix} \tag{5.33}$$

Note that we have normalised the regularization vector $\mathbf{s}/||\mathbf{s}||_2$ in Equation (5.15) and in the implementation.

This modified inversion scheme is now applied to the same problem as described in the previous section. The regularization parameter λ_1 of the first optimization problem $\min_x \phi_1$ is retained, while the L-curve criterion is used for the decision of λ_2. This results in $\lambda_2 = 3.3 \times 10^{-4}$. The L-curve with the corresponding conductivity profile in space and wavelet domain are presented in Figure 5.26. The profile displays the correct structure. Note that the fifth coefficient $\mathbf{x}_5$ in wavelet domain has disappeared. This indicates that it came into play to compensate for a too small coefficient x_3. The x_3 coefficient is now indeed larger than x_2, as it should be.

In future work, this process could be made more adaptive. In the iterative process, a criterion could be implemented that adjusts the regularization vector $\mathbf{s}$ for this issue 'online'. Note that an optimal regularization parameter must be determined for the first optimization problem $\min_{\mathbf{x}} \phi_1(\mathbf{x})$ as well as the second problem $\min_{\mathbf{x}} \phi_2(\mathbf{x})$.

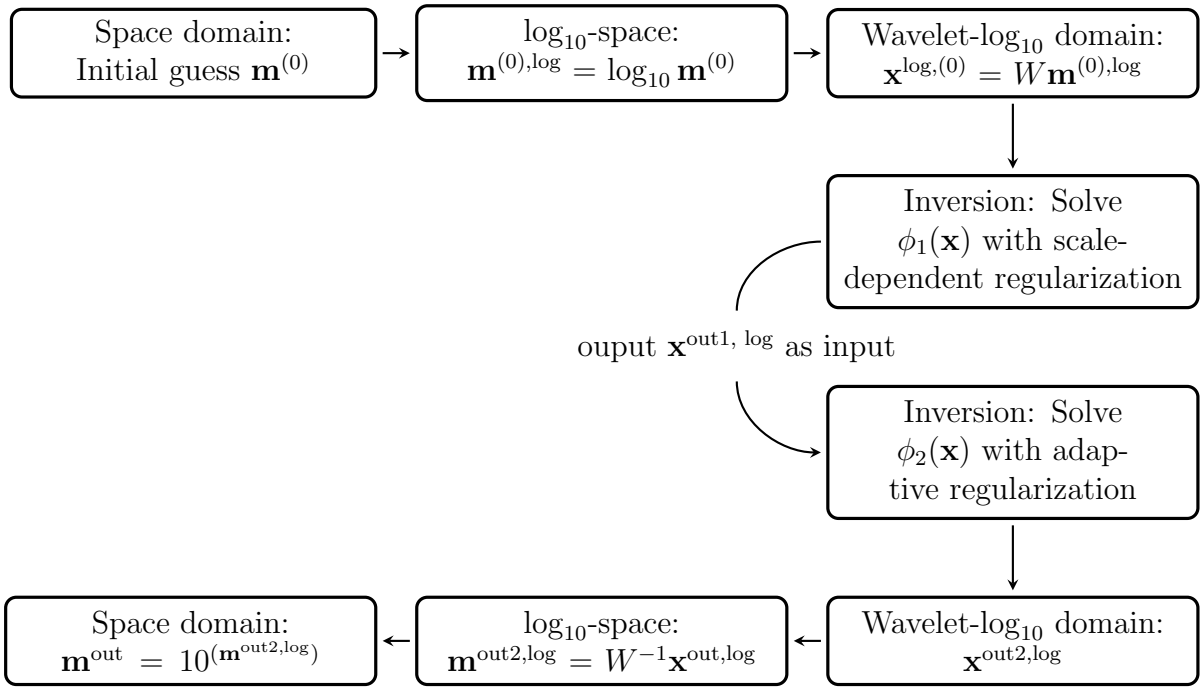


Figure 5.25: The inversion scheme in $\log_{10}$-space domain with a second optimization problem.

5.3.3 Sparsity probing

The modification described in this section is very similar to the approach from the previous section. A second minimization problem for ϕ_2 is added to the inversion scheme after the $\min_x \phi_1(\mathbf{x})$ with scale-dependent regularization. Instead of a new scale-dependent regularization vector, it is now checked which are the non-zero elements in the wavelet representation vector $\mathbf{x}^{\text{out1, log}}$. We assume that significantly large values in the outcome $\mathbf{x}^{\text{out1,log}}$ describe the real structure of the model. These large values are larger than a threshold δ. It will shortly be described how a good threshold δ can be pinpointed. Thresholding means that S coefficients, larger than δ, are considered as the only possible non-zero entries of S-sparse vector $\mathbf{x}$. As it were, by solving the minimization problem $\min_x \phi_1(\mathbf{x})$, the sparsity of the wavelet representation is probed. This is equivalent to probing the structure of the conductivity profile and not yet caring too much about the values of the true electrical conductivities. Finding accurate values for the conductivities is then the main issue for the second minimization problem $\min_x \phi_2(\mathbf{x})$. This view is why we refer to this approach as *'sparsity probing'*. Subsequently, after minimizing ϕ_1 and determining the S non-zero coefficients, the other coefficients in $\mathbf{x}$ are set to zero. This is maintained throughout the minimization of ϕ_2, by setting an upper and lower bound equal to zero in our optimization method. Indeed, the L-BFGS-B-method (see Section 3.4.3) allows us to easily impose bounds on the model. Hence, a projected gradient $P\nabla\phi$ will be considered, which will eliminate the zero-entries from the set of free variables. This is currently the easiest implementation of this approach and allows us to test and evaluate the concept. The regularization vector $\mathbf{s}$ in ϕ_2 is put equal to a vector with ones and is not normalised (hence, there is no scale-dependent regularization).

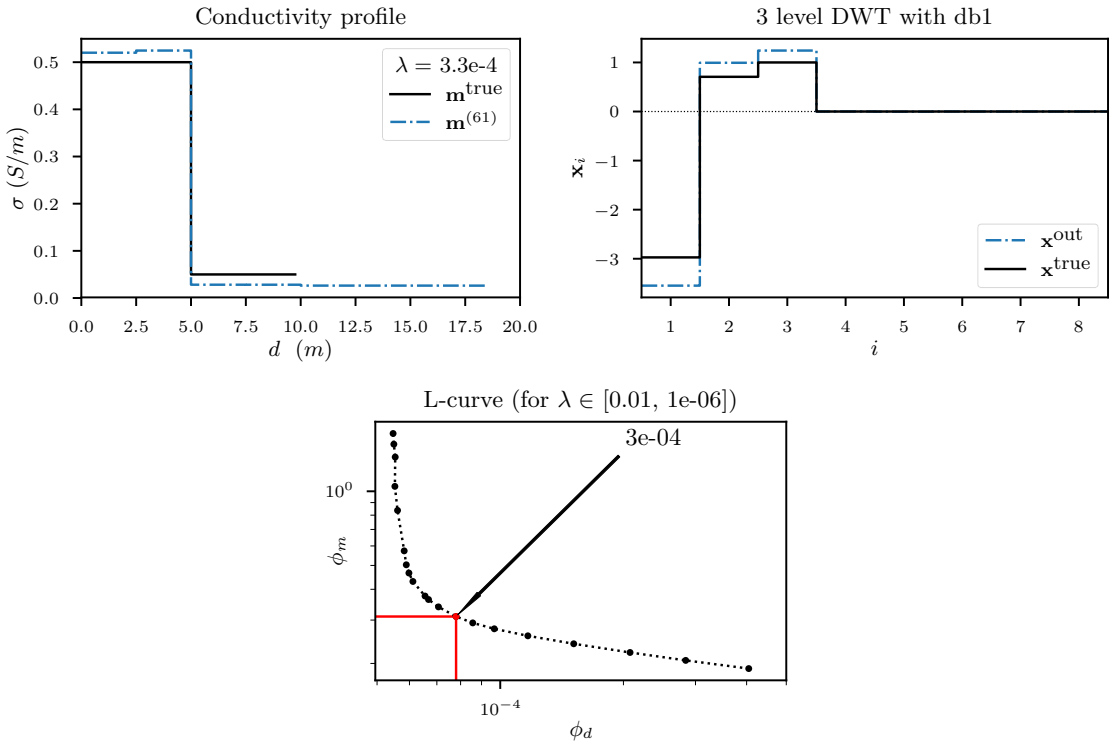


Figure 5.26: Outcome with adaptive regularization vector

There are some choices to be made, concerning the thresholding strategy. Especially in image denoising and compression, a field where wavelet theory is widely used, several thresholding algorithms exist [Chang et al., 2000], [Luisier et al., 2007]. For example, one can retain the $\chi\%$ in absolute value largest values in the vector or one can retain the values that are in absolute value larger than 10% of the largest coefficient in the vector. If the proposed approach works well, one could certainly think about more 'advanced' thresholding algorithms (as in the suggested literature). However, this would needlessly complexify the currently proposed approach and probably obstruct a simple assessment of the scheme.

One may wonder if it would not be better to take asymmetric threshold parameters $\delta_{\pm}$, since the inversion is carried out in log-space[10], such that $|10^{\delta-} - 1| = |10^{\delta+} - 1|$. The catch is that the inversion is carried out in wavelet log space. Wavelet coefficients do not contribute to the overall energy of the conductivity profile. Indeed, consider the domain where a wavelet coefficient x_i $(1 < i \leq n_x)$ is localized in space. The net effect on the total model of x_i and $-x_i$ is the same, except for its mirror image w.r.t. the centre of the domain of the wavelet.

Setting the threshold δ requires careful considerations. It depends on the maximum value of that wavelet function[11], the level of the discrete wavelet transform N *and* the value of the scaling coefficient. Let us illustrate the idea that will give us a good estimation for a

[10] Asymmetric parameters on the log-transform, such that the back-transform exhibits symmetric thresholds, i.e. one threshold parameter δ.

[11] The `db1`-wavelet has one as maximum value.

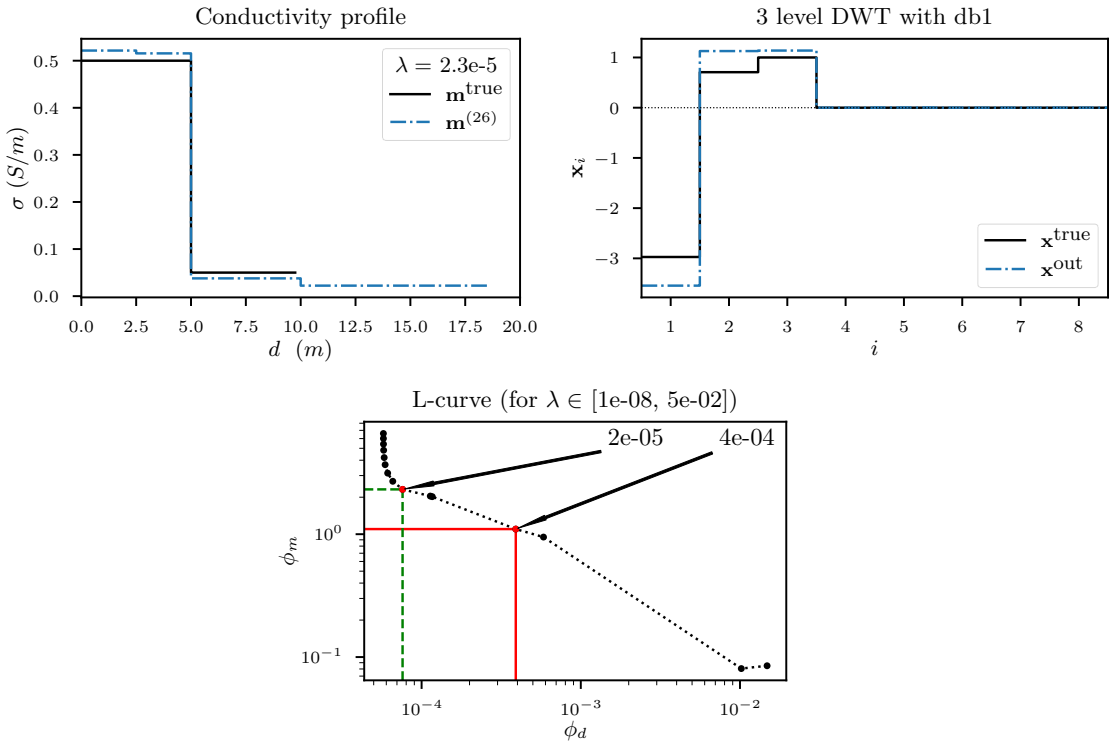


Figure 5.27: Sparsity probing approach where the outcome of a manually selected corner is shown (Dashed, green line in the L-curve plot).

good threshold δ. Consider the wavelet coefficient x_i of the smallest scale (or wavelets in the function space with the largest scale of resolution n, or $n_x/2 < i \leq n_x$). Its net effect on the model in the space domain Δm, where the wavelet coefficient is localized in space, will be

$$\Delta m \approx 10^{\frac{1}{\sqrt{2}^N} x_1 + \frac{1}{\sqrt{2}}\delta} - 10^{\frac{1}{\sqrt{2}^N} x_1}. \tag{5.34}$$

x_1 is the scaling coefficient. The factor $(1/\sqrt{2})^N$ comes into play due to the normalisation at every level of the DWT (See e.g. Eqns (4.30)-(4.34)). In practise, this threshold could be automatically determined such that it eliminates structure that is in the order of $\Delta m =$ 1 mS/m by calculating

$$\delta = \sqrt{2}\log_{10}\left(\Delta m + 10^{x_1/(2)^{N/2}}\right) - \frac{|x_1|}{2^{(N-1)/2}}. \tag{5.35}$$

In this case,

$$\delta = \sqrt{2}\log_{10}\left(10^{-3} + 10^{-3/(2)^{3/2}}\right) - (-3/2) \approx 10^{-2} \tag{5.36}$$

is the threshold that will be used.

As in the previous section, $\lambda_1 = 4.5 \times 10^{-3}$ is fixed and the L-curve criterion is applied on $\min_x \phi_2$. The Adaptive Pruning algorithm did not perform that well[12]. The corner was then manually tuned (dashed line) and the outcome is shown in Figure 5.27. After the first

[12]This is probably due to the density in the region for small λ. Many pruned L-curves will have the wrong shape and will produce wrong candidates.

optimization problem, the first, second, third and fifth component of $\mathbf{x}$ were retained after thresholding. There is again a significant improvement w.r.t. the stairlike solution from Section 5.3.2 from the scheme with only one optimization problem and scale-dependent regularization. The fifth component vanishes (as in the true profile), while it was a free variable. This is a promising result.

Let us note that there is a significant decrease in number of iterations. This is due the fact that there were only four free variables instead of eight. Especially for parametrizations with a large number of model parameters, this approach could notably reduce the computational burden.

5.3.4 Other focusing functions

The simplest modification to the scheme would be to find an alternative measure that solves the above problem. The ℓ_1-norm and the perturbed ℓ_1-norm measure of Ekblom will proportionally give a weight or 'cost' to any value in x:

$$\mu_{\text{Ekblom}}(x) > \mu_{\text{Ekblom}}(x') \text{ for } |x| > |x'|. \tag{5.37}$$

Indeed, $x = 1$ will add more weight to the objective function ϕ than $x' = 0.95$, ceteris paribus, the outcome with $x' = 0.95$ is favoured. This is an undesirable effect on the components in $\mathbf{x}$ which should be non-zero. The proportionally larger weight will favour outcomes where the non-zero coefficients are pushed to smaller values. In our context, $x \approx 1$ is definitely a dense entry in the wavelet representation. We can opt for a more binary scale for the measure.

Such a more binary measure was already introduced in Section 3.2.3: The minimum support measure μ_{Zhd}. For convenience, the curve of the measure is again shown in Figure 5.28. This measure is non-convex and the outcomes depend critically on the parameter ϵ. For small ϵ, the measure exhibits a more the binary scale. We believe that for small ϵ, the strongly non-convexity of the focusing function poses problems for the line search algorithm. Even for larger parameters, the results were not satisfying.

As a last attempt, we resort to another measure that is often used in geophysical inversion literature: The Cauchy measure (e.g. used in [Guitton, 2012]):.

$$\mu_{\text{Cauchy}}(x) = \frac{1}{2}\log\left(1 + \left(\frac{x}{\gamma}\right)^2\right), \tag{5.38}$$

where γ is a hyperparameter that needs to be selected. The effect of the hyperparameter on μ_{Cauchy} is shown in Figure 5.28. This parameter will play a more signficant role than with the Ekblom measure. Its derivative is

$$\frac{\mathrm{d}\mu_{\text{Cauchy}}(x)}{\mathrm{d}x} = \frac{1}{2}\frac{1}{\left(1 + \left(\frac{x}{\gamma}\right)^2\right)} \cdot 2\frac{x}{\gamma^2} = \frac{x}{\gamma^2 + x^2}. \tag{5.39}$$

The derivative for $x \to \pm\infty$ vanishes much more slowly than the minimum support measure. This could be advantageous for the line search algorithm: The gradient will guide

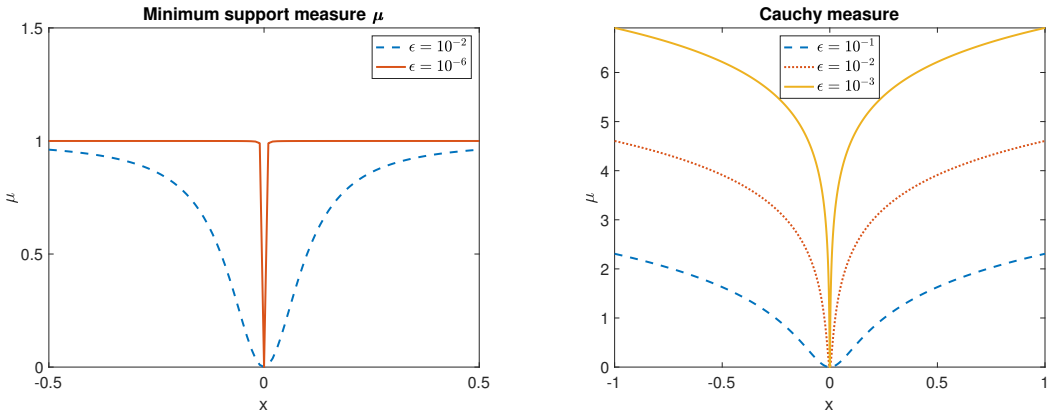


Figure 5.28: Binary-like focusing functions that replace the non-differentiable ℓ_1-norm.

the coefficients more reliably towards the minimum at $x = 0$.

Simulations were performed for $\gamma = 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}$ and 10^{-5}. The results can be summarized in Figure 5.29, where the result is shown for $\lambda = 2.9\times10^{-3}$ and $\gamma = 10^{-3}$. The L-curve is non-convex, thus the corner form the Adaptive Pruning algorithm should not be trusted (solid line). The solutions around $\phi_d = 10^{-2}$ correspond with constant solutions with no structure. The true L-curve is situated around for data misfits smaller than 10^{-4}. This corresponds with the corner around the measurement noise $||\boldsymbol{\eta}||_2^2$. We are, however, interested in the 'cusp' around the error around $||\kappa||_2^2$, where both the unmodelled and measurement noise is included. This cusp is absolutely absent in the L-curve. Taking into account that we have worked with equidistant regularization parameter distribution for the generation the L-curve, it is remarkable that there is not such a density around this point. There is thus not a broad regularization parameter range in which the outcomes are similar. Let us have a look at the misfit functionals

$$\begin{array}{lll} \phi(\mathbf{x}^{\lambda=2.9\times10^{-3}}) & = 2.976 \times 10^{-3} & \phi(\mathbf{x}^{\text{true}}) \quad = 7.34023 \times 10^{-3} \\ \phi_d(\mathbf{x}^{\lambda=2.9\times10^{-3}}) & = 1.7112 \times 10^{-4} & \phi_d(\mathbf{x}^{\text{true}}) \quad = 2.4 \times 10^{-4} \\ \phi_m(\mathbf{x}^{\lambda=2.9\times10^{-3}}) & = 2.4085 & \phi_m(\mathbf{x}^{\text{true}}) \quad = 2.3849 \end{array}$$

from which we learn that the presented outcome is indeed too complex, but it cannot be lowered, because a larger regularization parameter yields a constant outcome. The lower data misfit is also obvious from the plot from the data in Figure 5.28. The outcome in the wavelet domain has the right structure, but the scale-dependent regularization still forces the solution too much to zero. The Cauchy measure could not resolve this issue. The difference of the weight that is assigned to the true values $x_2 = 0.707$ and $x_3 = 1$ with $\gamma = 10^{-3}$ is still significant: $\mu_{\text{Cauchy}}(x_2) = 2,849$ vs. $\mu_{\text{Cauchy}}(x_3) = 3$. Parameter γ does not have much influence in this difference.

Let us finally mention that the Huber-measure, introduced in Section 3.2.3, has a quite similar behaviour as the Ekblom measure. Consequently, no significant difference in outcomes is expected. Without reporting the results, our simulations confirm our expectations.

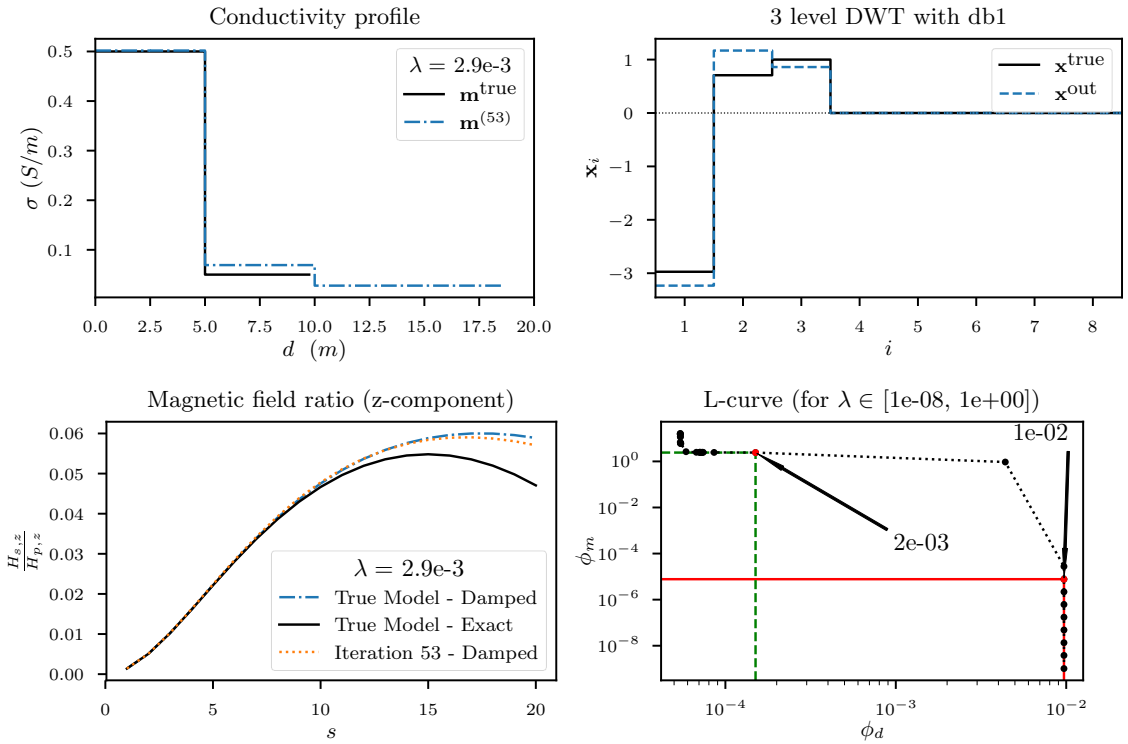


Figure 5.29: Scale-dependent regularization with the Cauchy measure μ_{Cauchy}, $\gamma = 10^{-3}$.

5.4 n-layered conductivity profiles

In this section, we turn our attention to realistic conductivity profiles, obtained via borehole logging. Two profiles from the Liège region are considered [Hermans and Irving, 2017], where a rich structure is observed with low electrical conductivities (maxima are in the order of 10^{-2} S/m). Other profiles remain in line with the setting with salt water intrusion, where conductivity profiles from De Panne are considered [Hermans et al., 2012]. These conductivity profiles have higher electrical conductivities, the maxima are in the order of 10^{-1} S/m.

The higher variability in the model may be accompanied by a need for a higher data density. This can be achieved by conducting measurements at more intercoil spacings, heights and frequencies. We will briefly illustrate the effect on the outcome in Section 5.4.5. Additionally, the number of parameters has an effect: A more precise parametrisation increases the computational burden and leads to a more ill-posed problem, i.e. there are more possibilities to approximately describe the same data. On the other hand, it increases the resolution; for example, the location of a boundary between two layers can be more adequately described. We will not focus on this trade-off, since the regularization was introduced to deal with this issue.

Recall that the lowest layer is semi-infinite. Consider a model $\mathbf{m}$ which is known to a depth of h meters. We will then generate the synthetic data with maximum intercoil distance $s = h$ meters. We parametrize the model for inversion to a depth of $2h$ meters. The part between h and $2h$ meters will only contribute 30% to the magnetic field data for

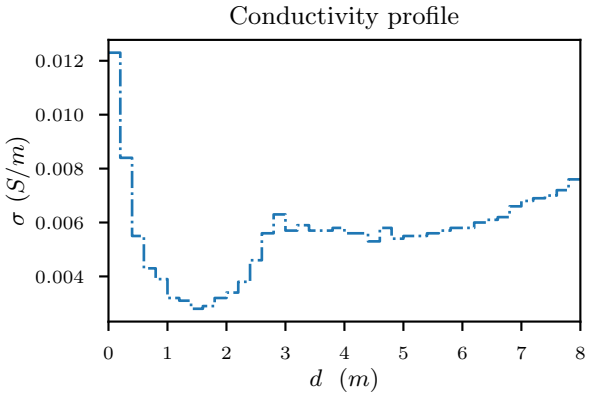


Figure 5.30: Conductivity profile from a site in Liège, obtained via borehole logging. Data retrieved from [Hermans and Irving, 2017].

the z-component, while for the ρ component this is even less. The data coverage of that range is low, but we do this because the semi-infinite layer is often not properly recovered at a depth of h meters. In this way, we push this issue to a less relevant region (see for example Figure 5.34b). However, we should keep in mind that the range between h and $2h$ meters should not be geologically interpreted.

In the subsequent section, we argue why wavelets other than Haar wavelets are more suitable for n-layered conductivity profiles.

5.4.1 Sparse representation

The main reason for the wavelet representation is that simple models have sparse representations. In Sections 5.1-5.3, simple two- or three-layered conductivity profiles were considered. More complex conductivity profiles, such as the one introduced in Chapter 2, see also Figure 5.30, are also considered as simple models in terms of the wavelet representation. An example of a model that cannot be represented in a sparse form is a model of uncorrelated noise. This model will contain just as much complexity in the wavelet domain: one will not observe the decay of the wavelet, as we explained in Section 4.6.

The difference between the blocky, two-layered conductivity profiles of Section 5.1 and the conductivity profile as in Figure 5.30 is that the two-layered structure could be *exactly* represented with only 3 coefficients in the `db1` domain, while the more complex conductivity profile can not. For more complex models, the wavelet coefficient will contain few zeros, but many very small coefficients. The influence of those small coefficients is small on the model in the space domain. The meaning of sparse has changed from having few non-zero coefficients to having few large coefficients. Wavelets other than the `db1` wavelet will be more suited.

This is easily illustrated with the conductivity profile presented in Figure 5.30 and its wavelet transformation into `db1` space and `db6` space in Figure 5.31, together with the scaling function and wavelet function. All values smaller than 10^{-3} are replaced by zeros, causing us to lose the perfect reconstruction property, but enabling us to approximate

the model with fewer entries. After thresholding, the sparse representation is back-transformed into the space domain. Observe the following characteristics: First, the shape of the wavelet is usually recognized in the back-transformed model. This is clearly visible for the `db1` wavelet. Second, higher vanishing moments p yield larger compact support and thus require more signal extension. In the case of `db1` wavelets, no signal extension was required ($n_x = n_m = 42$) while for the `db6` wavelet, the number of parameters in wavelet space was more than doubled ($n_x = 91$). The model with `db1` is represented with $S/n_x = 9/42 \approx 21\%$ of the coefficients and `db6` with $23/91 \approx 25\%$ of the coefficients. However, it is manifest that `db6` with 25% of the coefficients can be better represented than the model with `db1` and 21% of the coefficients (see Figure 5.31). Third, in the `db6` wavelet domain, the peaks at the 18-th, 30-th, 40-th ... coefficient are the result of boundary distortions, due to the signal extension. Indeed, symmetrization usually leads to the least distortions and still the effect is significant. This issue will be addressed in more detail in the next section.

Multiple Daubechies wavelets are used in this section, it is instructive to know how they look like. For an increasing number of vanishing moments, the compact support, the number of elements in wavelet domain n_x, the regularity and the number of oscillations increase. Figures such as Figure 5.31 are listed in Appendix D for all Daubechies wavelets for $p = 1, \cdots, 9$.

5.4.2 Effect of boundary distortions on sparsity

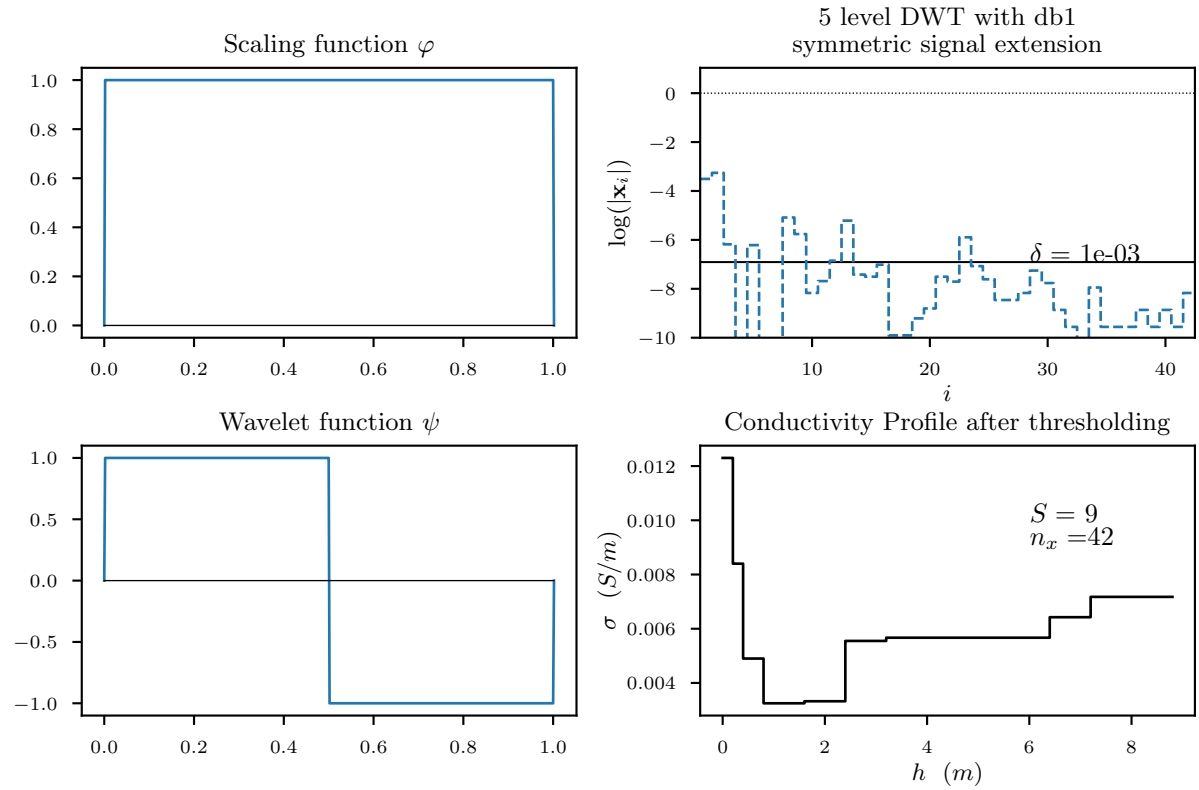
Boundary distortions emerge as a result of signal extension and affect the sparsity of the representation. We describe two aspects that are relevant in order to avoid too much border effects.

Technically, an $N = \lfloor \log_2(n_m) \rfloor$ level DWT can be computed from a model with a parametrization of n_m entries. However, there is a common rule [Mallat, 1989] that sets a maximum on the level of the DWT that should be considered, in order to avoid too many boundary distortions:

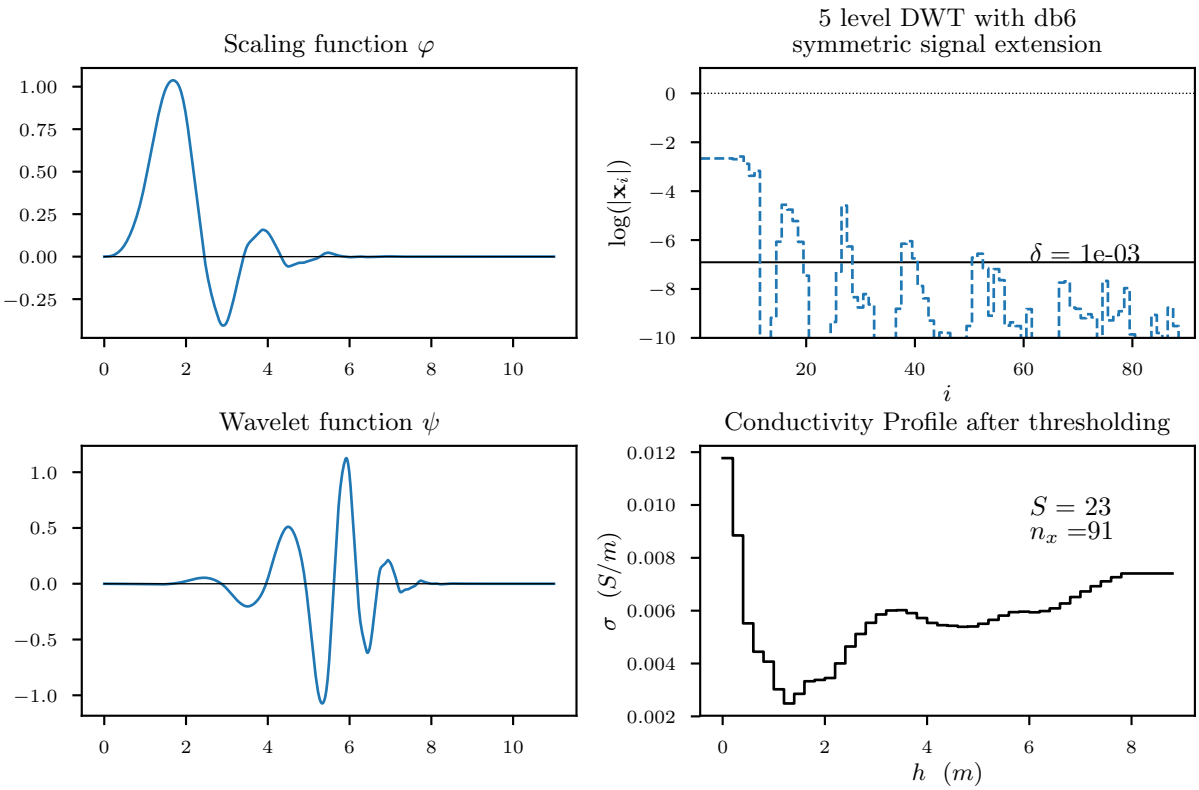
$$N = \left\lfloor \log_2 \left(\frac{n_m}{n_c - 1} \right) \right\rfloor, \tag{5.40}$$

where n_c are the number of filter coefficients in the two-scale relation (4.25). The rationale is that the wavelet decomposition stops when a signal becomes shorter than the filter length for a given wavelet. Indeed, more filter coefficients correspond with wavelets with larger compact support. In terms of that compact support, the maximum level is determined such that at least one scaling coefficient corresponds with a scaling function that is not affected by signal extension and thus border distortion.

A second aspect that impacts boundary distortions is the way by which signal extension is carried out (see Section 4.4.4). In general, periodization and symmetrization yield the most natural extension of the signal. For a two-layered structure, it is also conceivable to consider a smooth or constant extension. Different types of the extension were tested for the conductivity profile in Liège (Figure 5.30). The back-transformed conductivity profiles after thresholding were different for every type of signal extension, however, the level of sparsity was not significantly altered. At least, not to such an extent that we can



(a) Haar wavelet.



(b) Daubechies wavelet with six vanishing moments.

Figure 5.31: Full wavelet decomposition of the conductivity profile in Figure 5.30. Approximated model for threshold δ with S non-zero entries.

conclude that a certain type of extension is always better. For the remaining part of this thesis, symmetrical signal extension is used.

5.4.3 Inversion with Daubechies wavelets

In the previous sections, the considered conductivity profiles presented exactly two or three layers with sharp boundaries. The reality is often different, e.g. when there is a certain soft transition between the layers and thus no sharp boundary. This is observed in a conductivity profile from the Westhoek (Figure 5.32), where one strongly conductive layer with soft edges is present, embedded in less conductive layers. We cannot expect much more from the inversion with Haar wavelets than the rough blocky structure shown in Figure 5.32, because soft boundaries do not have a sparse representation with Haar wavelets. The L-curve exhibits the typical L-shape, the corner was selected for a regularization parameter just before the vertical part of the L-shape (as described in Section 5.2.2) and that regularization parameter lies in between the measurement error $||\boldsymbol{\eta}||_2^2$ and total error $||\boldsymbol{\kappa}||_2^2$. The result describes the coarsest structure of the layer well, except for the soft boundary. Small scale effects are filtered out. We must resort to other wavelets to recover the soft boundaries in a sparse fashion.

The `db2` wavelet can accurately represent linear functions, indeed, this wavelet has two vanishing moments and hence, each wavelet coefficient is orthogonal to any linear function. The wavelet has 4 filter coefficients and with a parameterization of 80 parameters, a 4 level DWT is considered in the inversion scheme. Via the Adaptive Pruning algorithm, the corner $\lambda = 0.3$ of the L-curve (Figure 5.33) is determined. Note that this is less regular than for `db1` wavelets, although the data misfit ϕ_d for the selected regularization parameter still lies between $||\boldsymbol{\eta}||_2^2$ and $||\boldsymbol{\kappa}||_2^2$. The result is presented in Figure 5.34a. At first glance the outcome looks relatively correct, the value of the maximum conductivity is close to the actual maximum of the electrical conductivity. The localization of the peak is not entirely correct. Note that the wavelet representation is indeed sparse, with the exception of the scaling coefficients (the first seven coefficients in the wavelet domain), where no regularization is also applied. Note, however, that all wavelet coefficients are vanishing, this suggests that the outcome is too sparse and thus that the regularization

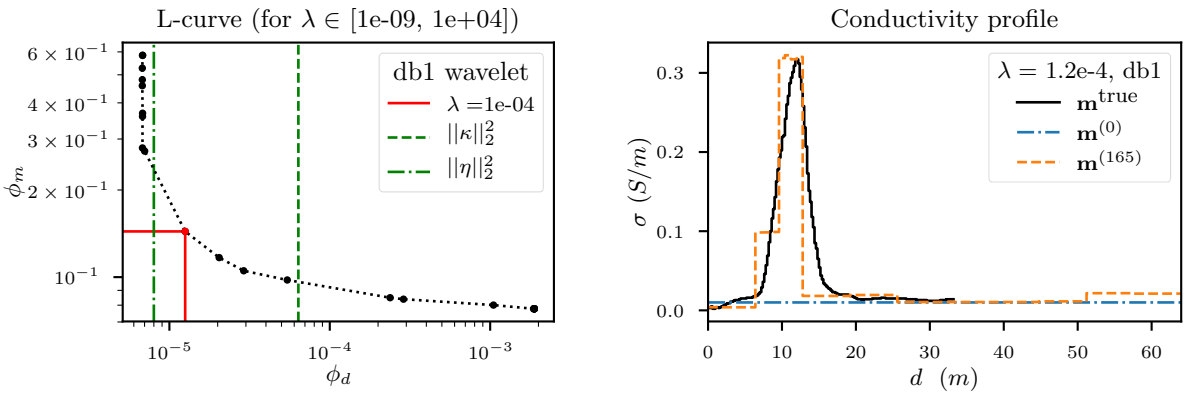


Figure 5.32: Conductivity profile from the Westhoek [Hermans et al., 2012] and L-curve with Haar wavelets.

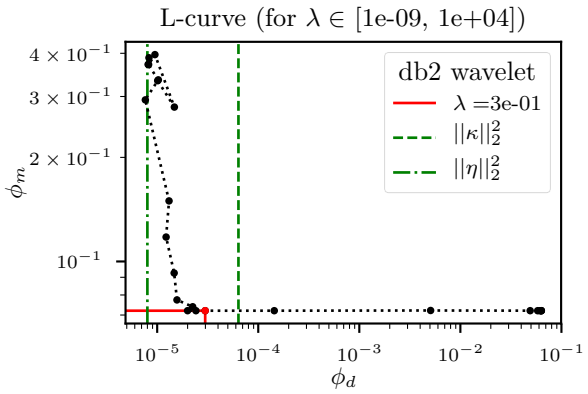
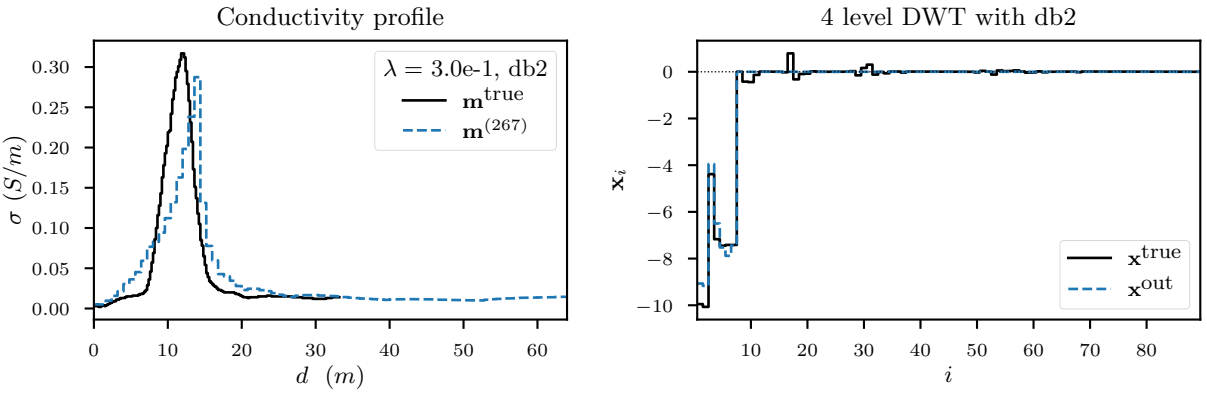
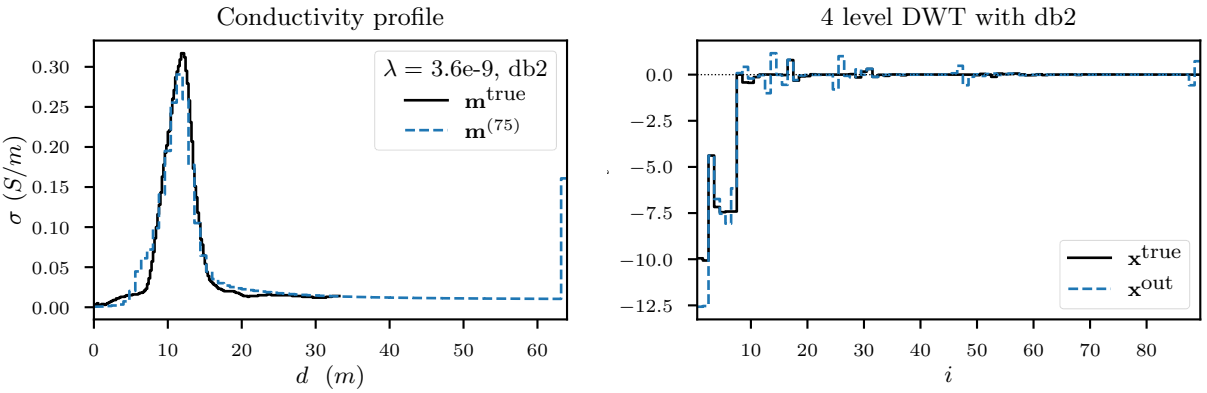


Figure 5.33: L-curve for inversion of conductivity profile from the Westhoek [Hermans et al., 2012] .



(a) The regularization parameter λ is determined from the L-curve criterion.



(b) Manual tuning of regularization parameter λ.

Figure 5.34: Conductivity profile from the Westhoek [Hermans et al., 2012] and L-curve with Haar wavelets.

was too strong. The model misfit of the outcome is equal to the model misfit of the null-solution (with no structure):

$$\phi_m(\mathbf{x}^{\text{out}}) = \phi_m(0) = 0.0724. \tag{5.41}$$

The model misfit is non-zero, due to the $\epsilon = 10^{-4}$ in the Ekblom measure. Additionally, $\lambda\phi_m = 0.0216 \gg \phi_d = 2.98 \times 10^{-5}$ and we have argued before that in an optimal case $\lambda\phi_m \approx \phi_d$.

The result can be further improved by choosing a regularization parameter $\lambda = 3.6 \times 10^{-9}$. The conductivity profiles in space and wavelet domain are shown in Figure 5.34b. The result follows the conductivity profile much more closely, especially in the space domain. The wavelet representation is less sparse and more details are now present. Note the spike at the depth of more or less 60 meters. We have warned for the effect of the lowest layer and we can ignore this for now, by our parametrization we effectively avoid the effect of the lowest layer in the interesting region of the true conductivity profile $\mathbf{m}^{\text{true}}$. A more important issue is how to select this regularization parameter. All the techniques discussed in Section 5.2 do not yield a good selection of this regularization parameter.

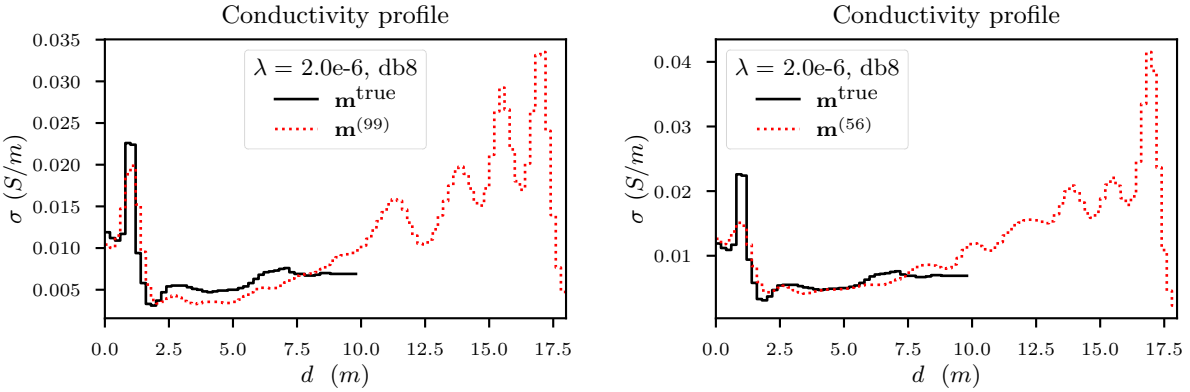
All Daubechies wavelets up to vanishing moment $p = 9$ were tested on the conductivity profile. This yielded similar outcomes, which are not reported. For an increasing number of vanishing moments p, the L-curve became very disordered: what the vertical part of the L-shape should be, is a hodgepodge of unordered points. As with the `db2` wavelet, the results with small regularization parameters showed the best agreement with the true conductivity profile.

Only Daubechies wavelets were tested, because these wavelets are the most popular wavelets due to their good characteristic: they have the largest number of vanishing moments (and thus approximating abilities) while maintaining the smallest possible compact support. Wide compact support produces more boundary distortions and consequently, the wavelet representation is less sparse. Daubechies wavelets are therefore a good choice. In Section 5.4.7 we will look more closely at biorthogonal wavelets.

The best outcomes were recognised for very small regularization parameters. This suggests to verify whether the inversion scheme in the wavelet domain with non-`db1` wavelets in itself has a stabilizing effect and whether the regularization parameter λ can be set to zero. This will turn out to be useful result in the next section.

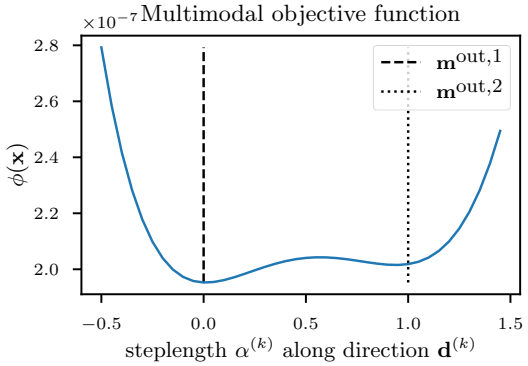
5.4.4 Multimodal objective function

In the previous section, a constant model was taken as the initial guess, with an electrical conductivity of 0.01 S/m. In that case, small regularization parameters were suitable and the L-curve was more unordered for larger p. For random initial guesses, a larger regularization parameter is required to filter out small-scale effects. The most important observation is that the outcome differs for each initial guess. The objective function exhibits *multimodal behaviour*. In Figure 5.35a-5.35b, two results with different random initial guesses are shown. These outcomes should correspond to minima in the model space (in log wavelet space). This is illustrated in Figure 5.35c, where the objective



(a) Outcome 1

(b) Outcome 2



(c) Objective function

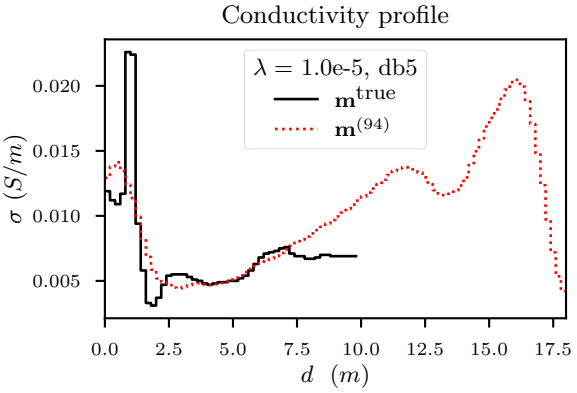
Figure 5.35: Inversion of the conductivity profile from Liège [Hermans and Irving, 2017] with the Daubechies 8 wavelet for different random initial guesses.

function is plotted along the straight line, connecting the two outcomes, by varying $\alpha^{(k)}$ in

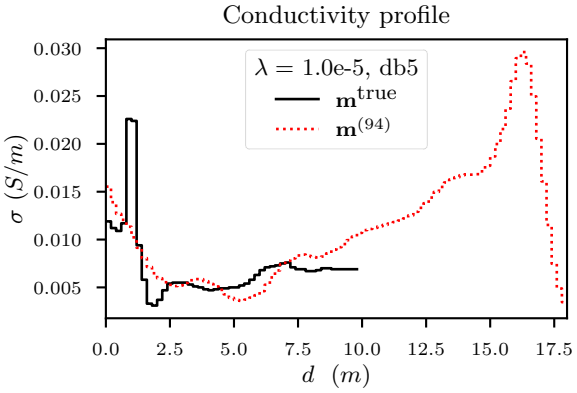
$$\phi(\mathbf{x}^{\text{out},1} + \alpha^{(k)}\mathbf{d}^{(k)}), \qquad \text{where } \mathbf{d}^{(k)} = \mathbf{x}^{\text{out},2} - \mathbf{x}^{\text{out},1}. \tag{5.42}$$

There are two approaches that we propose to handle the multimodality of the objective function that we will briefly illustrate. The first approach is a kind of statistical analysis, in which we start with many different initial guesses. In this way, we try to assign a kind of probability to an outcome. Ideally, the absolute minimum of the objective function (if it exists) is the most probable outcome and the minimum of the objective function has a good agreement with the true conductivity profile. There are more systematic and advanced multiscale techniques, which are described in Appendix C. We will return to this in the Future work in Chapter 6. The second approach starts from the constant initial guess, this is described in the Section 5.4.6, after the observations made in next section.

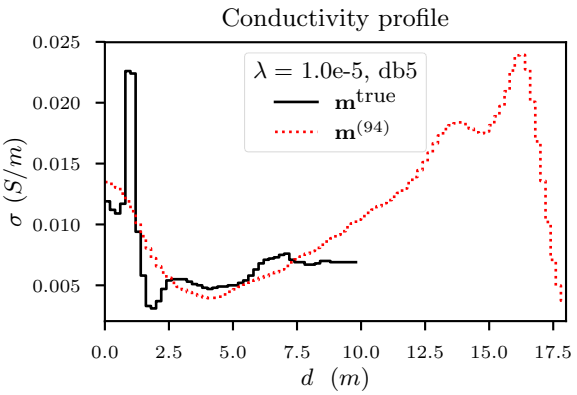
For the statistical analysis we only compare the outcomes for the parameters for depths between 0 and h meters. The lower layers also have an influence on the upper layers (Due to the wide compact support of the wavelets there is always an overlap), we therefore state that outcomes are equal to each other as the ℓ_2-norm of the difference between the



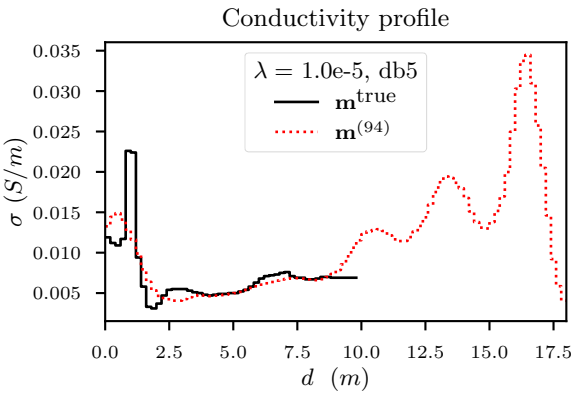
(a) Most frequent outcome: 23/40



(b) Outcome with relative frequency 8/40



(c) Outcome with relative frequency 3/40



(d) Outcome that occurred only once.

Figure 5.36: Statistical analysis of the outcomes with different initial guesses.

models in space domain differ by 0.005, i.e.

$$||\mathbf{m}^{\text{out},1} - \mathbf{m}^{\text{out},1}|| < 0.005. \tag{5.43}$$

Of the 40 different initial conditions, we find five different solutions with relative frequencies 23/40, 8/40, 3/40 and 6 outcomes that were unique. The three most frequent outcomes and fourth outcome are shown in Figure 5.36. This way of working is quite cumbersome and computationally demanding because the inversion procedure needs to be repeated numerous times. Based on this analysis, the model in Figure 5.36a would be favoured. This outcome broadly follows the conductivity profile and has the lowest value of the objective function ϕ.

5.4.5 Inversion in wavelet domain and no regularization

An inversion scheme in the wavelet domain has a clear advantage. In Figure 5.37 two outcomes are presented from minimization in the space domain vs minimization in the wavelet domain with the `db2` wavelet and without regularization (i.e. $\lambda = 0$). In the space domain, 405 iterations were needed, while only 77 in the wavelet domain. Both outcomes have a data misfit that is approximately equal to the measurement noise $||\boldsymbol{\eta}||_2^2$. The outcome in the space domain has too much structure, while the solution in the wavelet

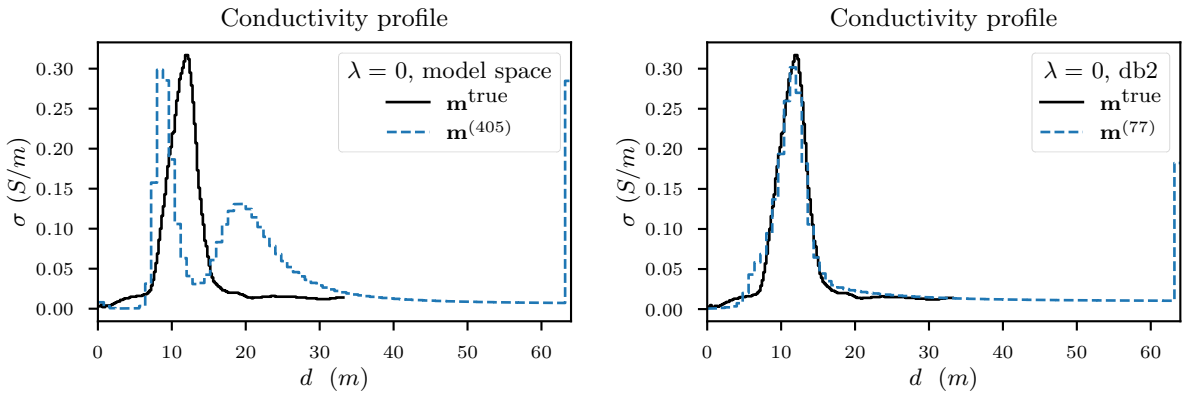


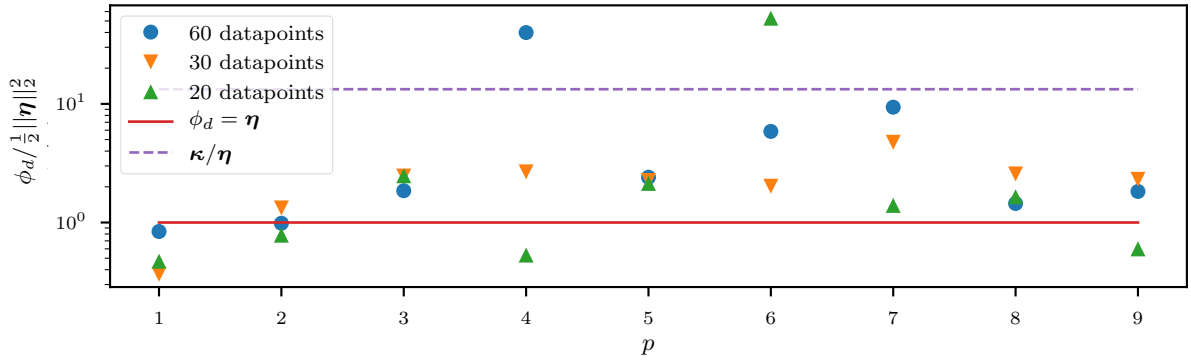
Figure 5.37: Comparison between space domain inversion and wavelet domain inversion with a Daubechies 2 wavelet.

domain follows the behaviour of the true model very well.

The remaining task is to make a distinction between the solutions with different wavelets and to determine which outcome is the best. The discrepancy principle is most relevant, although not in the usual usage. The data misfit is not plotted in terms of the regularization parameter λ but in terms of the number of vanishing moments p. Let us also immediately vary the number of data points and see how critically the outcome depends on the gradation of ill-posedness (fewer data points leads to a more ill-posed problem). The inversion is reproduced three times with 60, 30 and 20 data points. In Figure 5.38a, the ratios of the data misfits ϕ_d by the measurement noise $||\boldsymbol{\eta}||_2^2$ are plotted (in this way the data misfits remain comparable across the different simulations) in terms of the number of vanishing moments p. A wide spectrum of outcomes is observed, ranging from outcomes with overfitting and outcomes with too large data misfit. Based on the plot, it is clear that `db1` wavelets are not suitable for an inversion scheme *without* regularization. In Figure 5.38b, such an outcome for 60 data points is shown. It clearly does not allow for a good interpretation of the conductivity profile (the location of the peaks and values for electrical conductivities are too different from the true geology). For `db4` wavelets, it is clear that the data misfit depends critically on the number of data points. The data misfits of the `db5` wavelet are very close to each other and are not greater than the total noise $||\boldsymbol{\kappa}||_2^2$. This could be a hint for a potential good wavelet *for this specific conductivity profile*. The results of the inversions are shown in Figures 5.38c-5.38e. The results are of similar quality, they describe the peak and fuzzy edges fairly well, but the true maximum value of the electrical conductivity cannot be recovered. The case with only 20 data points succeeds in locating that peak exactly, but the edges are recovered less sharply.

5.4.6 Perturbation on the outcome

The analysis from the previous section was carried out in the previous section on a complex model from the Liège borehole data. All outcomes for different wavelets were quite similar, except for the unstable behaviour of the Haar wavelet. Further analysis of the results in the wavelet domain led to the observation that all outcomes are rather sparse. This is a remarkable result, given that no sparsity constraint was imposed ($\lambda = 0$). It



(a) Ratio $\phi_d/(1/2||\eta||_2^2)$ in terms of number of vanishing moments p.

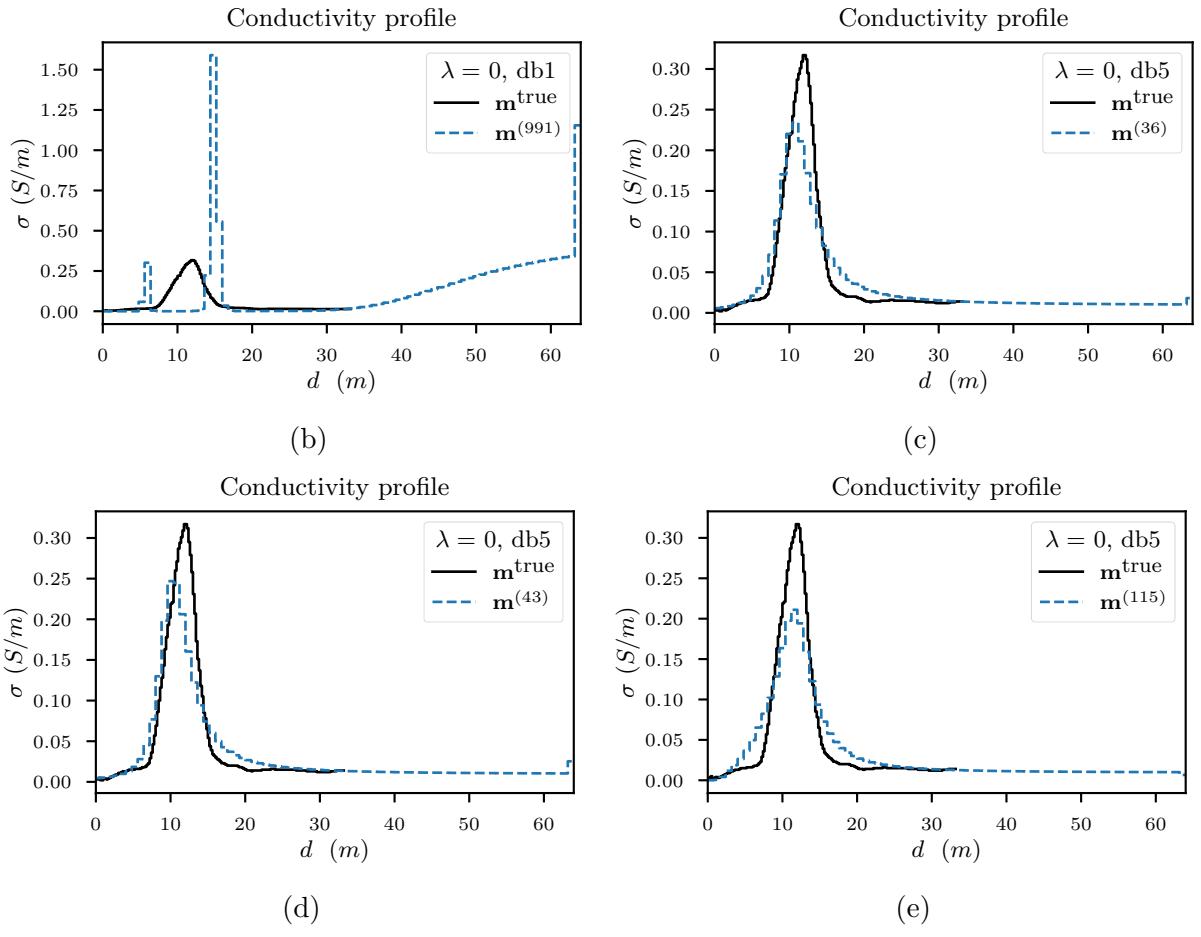


Figure 5.38: Inversion in wavelet domain without regularization and constant initial guess. Simulation was performed for several numbers of data points.

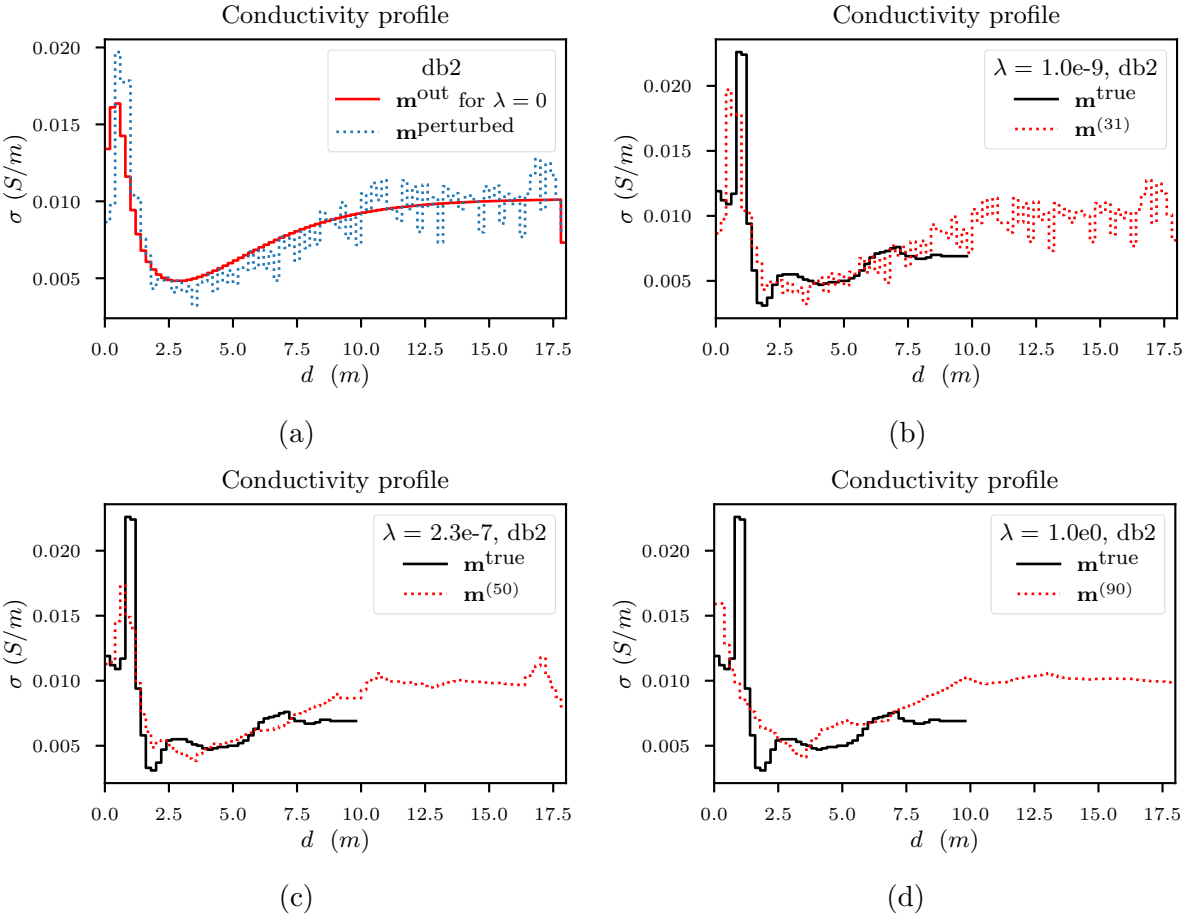
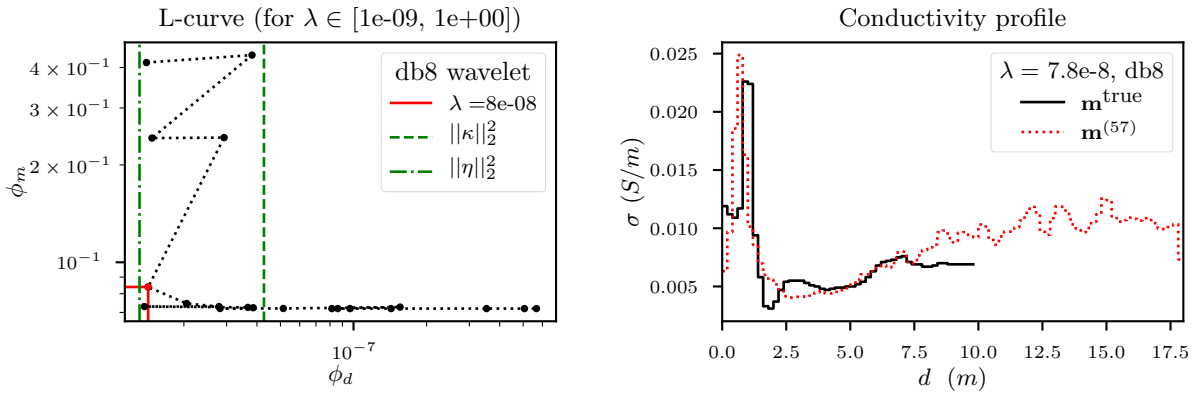


Figure 5.39: Results of the second optimization problem in the perturbative approach.

appears that the initial guess of the model has a major influence on the outcome, as we have emphasized in Section 5.4.4. The global structure is recognized from the constant initial guess, but small-scale effects do not emerge. This is different than the first approach, where a random initial guess was taken. However, if the inversion of a constant solution succeeds in finding the coarse structure in a small number of iterations (often less than 100 iterations), then we might better consider an alternative method. The inversion scheme is again extended with a second optimization problem. The first problem is an optimization problem with a Daubechies wavelet with vanishing moment $p > 1$ and with $\lambda = 0$. Subsequently, this result is perturbed, the sparse representation is made dense by adding a random model in wavelet space to the result of the first optimization problem. This random model vector has entries of a maximum of 0.01. As it were, the extra structure is added to the outcome of the first optimization method, and a second optimization problem will minimize this structure, while it will also hopefully describe the as yet undiscovered structure. In terms of the model space, the outcome is perturbed will search for another minimum in the neighbourhood.

The results of this approach with `db2` wavelets are presented in Figures 5.39 and 5.40. In Figure 5.39a, the outcome of the first optimization problem is shown (solid line), together with its perturbation (dotted line). In Figures 5.39b-5.39d, the outcomes of the second optimization problem are shown for different regularization parameters. The



(a) (b)

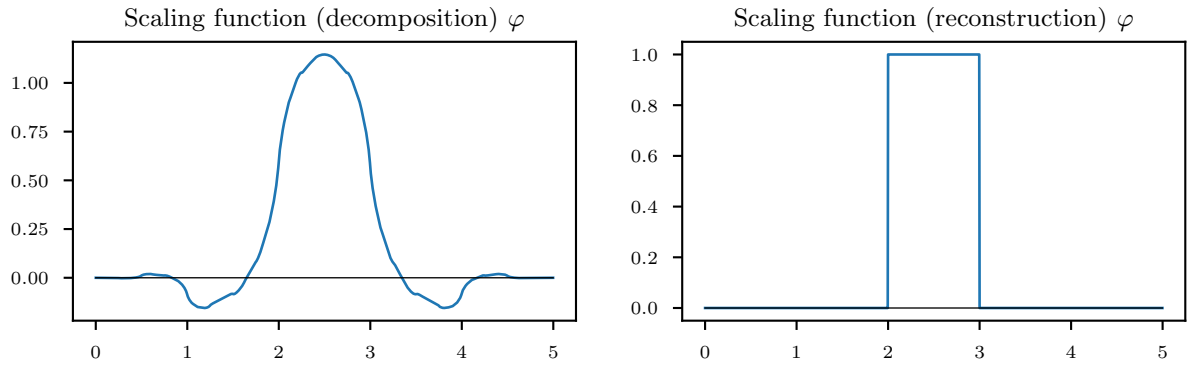
Figure 5.40: L-curve for inversion of conductivity profile from Liège [Hermans and Irving, 2017] from the perturbed outcome, see Figure 5.39a.

typical features are recognized: a too low regularization parameter leads to too much structure, while $\lambda = 1$ generates a too sparse outcome. The L-curve in Figure 5.40 does exhibit the desired L-curve. The robust Adaptive Pruning method still manages to obtain a parameter near the corner, $\lambda = 7.8 \times 10^{-7}$. The outcome has a more complex structure than the outcome of the first optimization problem. The sharp peak is recovered pretty well (the localization and value of electrical conductivity are notably good, given that this is a difficult conductivity profile with many high resolution features). Near a depth of 2.5 meters, the outcome does not follow the real profile very closely, but from 4 to 7.5 meters it does. Remember that the lower layers have very low data coverage; we should not attach too much importance to it.

5.4.7 Biorthogonal wavelets

We end the analysis of the inversion scheme with a remark about the choice of the wavelet. The best choice of the wavelet is not an exact science, but depends on the application.

Suppose that geologists are more interested in sharp boundaries and yet they expect a complex conductivity profile. The complex profile excludes the Haar wavelets from being the best choice. One can resort to biorthogonal wavelets, these allow much more freedom. We consider the case of a complex conductivity profile from the Liège region. Suppose that the geologist desires a blocky outcome. One can now look into the wavelet library; the `bior1.3` wavelet seems to be a suitable candidate for this setting. The `bior1.3` wavelet has three vanishing moments on the analysis side and one vanishing moment on the synthesis side of the filter bank. The vanishing moments at the analysis side of the DWT allow for a sparse representation, while the blocky synthesis wavelets allow a blocky reconstruction. The scaling functions of both sides are shown in Figure 5.41. The result after inversion is presented in Figure 5.42. There is, of course, no perfect agreement, but the desired blockiness is present in the outcome.



(a) (b)

Figure 5.41: The scaling functions corresponding with the Biorthogonal 1.3 wavelet.

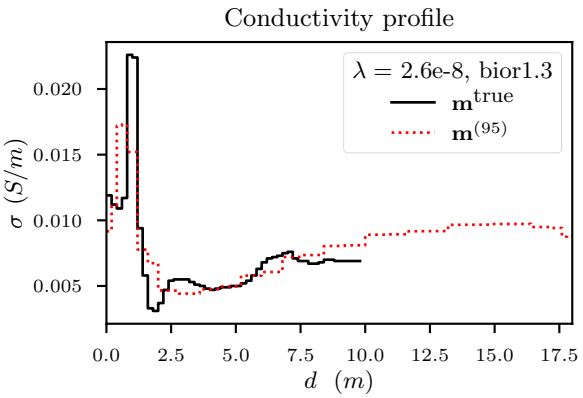


Figure 5.42: Inversion with biorthogonal wavelet.

Chapter 6

Conclusions and Future Work

Conclusions

The goal of this thesis was to examine the working principle of the Electromagnetic Induction exploration technique, its forward modelling in a conductive environment and the use of wavelet theory in the inversion scheme. The motivation and context of this thesis are to monitor the hydrogeological balance between fresh and salt water, in a typically conductive environment. Due to the higher electrical conductivity of salt water, the exploration technique based on electromagnetic induction is very suitable for prospecting salt water intrusion.

The typically horizontally stratified sedimentary structure in the Belgian coastal region justifies the use of the 1D model proposed by [Wait, 1951], which we referred to as the exact model. The computational burden was too large for the inverse problem and consequently, alternative models were examined. A first simplification was available under the LIN approximation [Wait, 1962], this model did not take into account the electromagnetic dampening and couplings between the eddy currents. This is a valid approximation under the LIN condition, but the LIN assumption breaks down for large intercoil distances and electrical conductivities, hence a third model was introduced. This recently developed model by [Maveau et al., 2017] is comparable to the LIN approximation but effectively models the electromagnetic dampening by the introduction of a conductive background.

The inverse problem is tackled by the minimization of a data misfit functional ϕ_d, where the magnetic field response of a guess of the conductivity profile (a trial model) is calculated via the forward model and compared with the response from experiments or synthetic data. The inverse problem is ill-conditioned and ill-posed. Given the presence of measurement errors and unmodelled errors, the inverse problem is reformulated for numerical treatment. Wavelet theory is suited to be incorporated in the inversion scheme, through the approximating abilities of the discrete wavelet transform. Basically, wavelet coefficients decay for higher levels of resolution, provided that the model in space domain is sufficiently differentiable (this degree depends on the number of vanishing moments of the wavelet). Put differently, an uncorrelated model (e.g. a randomly generated vector) cannot be represented in the wavelet domain in a sparse fashion. This guarantees the sparse nature of the wavelet representation of a realistic, geological model. The sparsity constraint is imposed on the conductivity profile, by adding a model misfit functional ϕ_m

to the objective function ϕ, via a sparsity promoting measure or focusing function. The ℓ_1-norm is the most well-known sparsity promoting measure, though it is not differentiable in $x = 0$. The L-BFGS-B method is used to minimize the objective function and this method requires gradient information. The perturbed ℓ_1-norm measure of Ekblom approximates the ℓ_1-norm and is differentiable on the whole domain. Our results indicate that the Ekblom measure is a suitable alternative to the ℓ_1-norm as a sparsity promoting measure.

For a two-layered subsurface, the Haar wavelets are particularly well suited, because the true model can be represented in an exactly sparse form (for other wavelets, many coefficients are almost zero). A comparison between inversion with and without regularization shows that regularization does indeed filter out small-scale effects from the model and therefore it prevents overfitting. For this purpose, a scale-dependent regularization scheme had to be implemented. In this scheme, every scale of resolution n was weighted by a factor of 2^{np}, where p is the number of vanishing moments of the wavelet. There is no regularization imposed on the scaling coefficients because it describes the (non-zero) energy of the conductivity profile. This factor p was inspired by the better approximating abilities of wavelets with a higher number of vanishing moments.

Electrical conductivities are always positive, however, the results of the inversions occasionally contained negative conductivities. This problem could easily be resolved by applying a logarithmic transform. The sequence of transforms in our inversion scheme is the following: A guess of the conductivity profile in the space domain is transformed into the logarithmic domain with base 10. Subsequently, it is transformed into the wavelet domain of a specific wavelet family, where the objective function is minimized and the outcome is back-transformed to the space domain.

We have argued in Section 5.3 that in certain cases the scale-dependent regularization generates wrong outcomes. We have proposed three solutions to this problem: adaptive regularization, sparsity probing and the use of other focusing functions. The first two methods extend the inversion scheme with a second optimization problem, where the objective function has a different regularization strategy. The adaptive strategy detects when a wavelet coefficient on a higher scale of resolution $n + 1$ is suppressed and sets the weight to 2^{np} instead of $2^{(n+1)p}$. This strategy improves the performance of the scheme. The sparsity probing strategy uses the outcome of the first optimization problem to find the true non-zero elements of the wavelet representation, via thresholding. A second optimization problem is then performed, only on the non-zero coefficients with equal regularization (i.e. not scale-dependent). This strategy also yields better results.

Other focusing functions were also tested, namely the minimum support measure, the Cauchy measure and the Huber measure. The latter is quite similar to the Ekblom measure and consequently, it does not yield very different results. We have argued that the minimum support measure is an approximate binary measure; for every pair of different values x_1, x_2 larger than some threshold (i.e. far enough from $x = 0$), the weights are similar, $\mu(x_1) \approx \mu(x_2)$. This measure does not generate good results, probably due to the strong non-convexity of the measure. The Cauchy measure is a hybrid between the behaviour of the minimum support and the Ekblom measure, but it cannot prevent that

large values on a higher scale of resolution are suppressed.

The balance between the data misfit and the model misfit is determined by the regularization parameter λ. A regularization parameter too small allows overfitting, while one that is too large yields outcomes with too little structure. Three common methods were described and tested in Section 5.2: The discrepancy principle, which uses information about the noise level; The L-curve criterion, where the model misfit is plotted in terms of the data misfit in log-log scale and for which the corner corresponds with the optimal outcome; Descent strategies, several strategies were developed and tested. The latter method is not always robust, it strongly depends on the choice of the initial regularization parameter and the stopping criterion. The lagging strategy, where the regularization parameter is updated in every iteration until the model misfit and data misfit are equal, does decrease the number of iterations and thus the computation time. A 'stand-alone' decent strategy is not sufficient, yet in combination with the discrepancy principle, the outcome is stable.

The L-curve exhibits a typical L-shape for data that is generated and inverted via the LIN approximation, though it is an unrealistic dataset. A more realistic synthetic dataset is generated by the exact model. When the inversion is conducted via the damped model, the L-curve criterion is more challenging to interpret. In the latter case, the L-curve does not entirely display the typical L-shape. The left vertical part of the L is easily recognized, it corresponds with the data misfit that is approximately equal to the measurement noise $||\boldsymbol{\eta}||_2^2$. However, near the total noise $||\boldsymbol{\kappa}||_2^2$ (i.e. measurement and unmodelled errors) a small cusp is usually noticeable; this is an atypical local corner in the L-curve. In combination with the discrepancy principle, the L-curve can be truncated. This allows the robust Adaptive Pruning Algorithm to select a regularization parameter λ which corresponds to an outcome that follows the true conductivity profile closely.

For complex conductivity profiles, i.e. profiles with soft boundaries or more than 3 layers, it is better to use other wavelets than Haar wavelets, because Haar wavelets cannot represent a complex conductivity profile in such a sparse fashion. Other wavelets are often able to represent complex profiles with fewer large coefficients. The use of other wavelets generally introduces boundary distortions, due to the inevitable fact that signal extension is required. Boundary distortions cannot be represented in a sparse form. Tests with different types of extension on true conductivity profiles are not very conclusive. For that reason, the standard symmetrical signal extension type is best used. In general settings, they are known to produce the least distortions. Moreover, it is better not to use the full wavelet transform, but only the number of levels that still allow one entire scaling function to fit in the profile completely (i.e. the number of filter coefficients is smaller than or equal to the number or model parameters in space domain).

For a heavily ill-posed problem (e.g. the number of parameters is much larger than the number of data points), the objective function is multimodal. In Section 5.4, where we have inverted complex conductivity profiles with many Daubechies wavelets, we have encountered such cases of multimodality. One can consider a statistical analysis, where many random initial guesses are generated and the inversion is conducted on every initial guess. The outcome with the lowest value of the objective function is then selected. There

exist more systematic approaches, which can be considered in future work. An alternative approach is to start from a constant initial guess, for which the global structure of the conductivity profile emerges. Even when the model misfit term is dropped from the objective function ($\lambda = 0$), the outcome is sparse. Additional details from the conductivity profile are obtained by perturbing the outcome and solving a second optimization problem. This problem will require a non-zero regularization parameter. Both approaches are promising but should be further investigated, before drawing a definite conclusion. Finally, we have illustrated that it is important to choose the right wavelet. For simple two-layered conductivity profiles, Haar wavelets are most suitable. For complex conductivity profiles with sharp boundaries, wavelets with low regularity should be used (such as Daubechies wavelets with two or three vanishing moments). With very smooth conductivity profiles, wavelets with a high degree of regularity are most fitted (such as Daubechies wavelets with six or eight vanishing moments). The choice of wavelets is not only limited to Daubechies wavelets; biorthogonal wavelets allow to combine different properties by considering a different wavelet on the analysis and on the synthesis side of the wavelet transform.

In conclusion, we have developed an effective inversion scheme that significantly alleviates the ill-posedness of the inverse problem. The scheme can be tailored to the context of the survey and the needs of the geologist.

Future work

In Chapter 2, we have described the EMI measurements in De Panne. A first future work would be to invert this data. To make this work properly, (1) the implementation has to be further developed to work with data of different frequencies; (2) the implementation of weighted least squares is most probably beneficial. The latter addition to the scheme will allow to incorporate the reciprocals of the estimates of the standard deviations of the noise. Writing the estimates in a diagonal matrix W_d yields a modified data misfit

$$\phi_d = \frac{1}{2}\|W_d\left(\mathbf{d} - K(W^{-1}\mathbf{x})\right)\|_2^2. \tag{6.1}$$

The multimodality of the objective function requires the use of more systematic approaches. In Appendix C, the issue was already addressed and the listed methods are worth implementing. Nowadays, machine learning is appearing in many fields of science, recall that it was even mentioned as a method to find better focusing functions in Section 3.2.3. Machine learning is capable of further refining the ideas from the statistical analysis and could be used in the inversion of multimodal objective functions. The papers of [Day-Lewis, 2018] and [Oware et al., 2019] can serve as a source of inspiration.

In Chapter 1, the difference between time domain electromagnetic (TDEM) surveying and frequency domain electromagnetic (FDEM) surveying was described. The accuracy of TDEM surveys is larger due to the absence of the primary magnetic field during the receiver interval. The interpretation of the analysis will be slightly different: the decay rate of the secondary magnetic field will be related to the conductivity, as we have outlined in Chapter 1. In future work, the analysis of the models in time domain could be examined. Typically, the frequency domain model is back-transformed to the time

domain. Time-domain electromagnetics is commonly used today: The TOPSOIL project [Jørgensen, 2018] is an EU cooperation supported by the Interreg VB North Sea Region program, which uses SkyTEM's airborne electromagnetics products [SkyTEM, 2019].

Finally, it would be a significant improvement to consider 2D models, where the electrical conductivity varies vertically and laterally. This can be achieved by approximating the dipping layers through a polynomial description along the measurement direction. Recent work by [Dierckx et al., 2018] could serve as a good starting point.

Appendix A

EMI-measurements

EMI-data around the sea inlet, as described in Chapter 2.

Data along line 2 in Figure 2.5a
A = intercoil distance s [m]
B = receiver orientation
C = transmitter orientation
D = apparent conductivity σ_a [mS/m]
E = range of instrument

A	B	C	D	E
10	HD	HD	175,9	100
10	HD (perp)	HD	179,4	100
10	VD	HD	180,4	100
10	VD	VD	-86	100
10	HD (perp)	VD	181,6	100
20	HD	HD	158,2	100
20	HD (perp)	HD	66,5	100
20	VD	HD	-31,7	100
20	VD	VD	-52,2	100
20	HD (perp)	VD	186,9	100
40	HD	HD	180,8	100
40	HD (perp)	HD	106,6	100
40	VD	HD	42,4	100
40	VD	VD	-41	100
40	HD (perp)	VD	186,9	100

Concluded

Data along line 3 in Figure 2.5a
A = intercoil distance s [m]
B = receiver orientation
C = transmitter orientation
D = apparent conductivity σ_a [mS/m]
E = range of instrument

A	B	C	D	E
10	HD	HD	183,3	100
10	HD (perp)	HD	183,5	100
10	VD	HD	183	100
10	VD	VD	-19,5	100
10	HD (perp)	VD	183,3	100
20	HD	HD	181,3	100
20	HD (perp)	HD	98,6	100
20	VD	HD	53,1	100
20	VD	VD	-18,1	100
20	HD (perp)	VD	186,1	100
40	HD	HD	166	100
40	HD (perp)	HD	118,1	100
40	VD	HD	101,4	100
40	VD	VD	-19,8	100
40	HD (perp)	VD	189	100

Concluded

Data along line 1 in Figure 2.5a

A = intercoil distance s [m]
B = spacing instrument in meters
C = coil orientation
D = apparent conductivity σ_a [mS/m]
E = range of instrument

A	B	C	D	E
10	10	HD	44,2	100
10	10	VD	55,7	100
10	10	VD	59,2	100
10	10	HD	44,6	100
20	20	VD	61,4	100
20	20	HD	82,2	100
40	40	HD	101	100
40	40	VD	58	100
20	20	HD	83,2	100
20	20	VD	63,1	100
40	10	VD	177,5	100
40	10	HD	177,9	100
40	20	HD	45,3	100
40	20	VD	14,1	100
40	40	VD	58	100
40	40	HD	97,5	100
38	40	HD	103	100
38	20	HD	50	100
38	10	HD	170,2	100
38	10	VD	170,3	100
38	20	VD	19,1	100
38	40	VD	62,1	100
36	40	VD	51,4	100
36	20	VD	21,9	100
36	10	VD	179,1	100
36	10	HD	179	100
36	20	HD	53,9	100
36	40	HD	94,5	100
34	10	HD	179,5	100
34	20	HD	59,8	100
34	40	HD	85,1	100
34	40	VD	37,5	100
34	20	VD	27,9	100
34	10	VD	179,4	100
32	40	VD	25,8	100

Continued

A = intercoil distance s [m]
B = spacing instrument in metres
C = coil orientation
D = apparent conductivity σ_a [mS/m]
E = range of instrument

A	B	C	D	E
32	20	VD	32,3	100
32	10	VD	179,9	100
32	10	HD	179,6	100
32	20	HD	64,7	100
32	40	HD	65,7	100
30	40	HD	35,3	100
30	20	HD	67,6	100
30	10	HD	180,5	100
30	10	VD	180,2	100
30	20	VD	37,5	100
30	40	VD	103	1000
28	40	VD	113	1000
28	20	VD	41,8	100
28	10	VD	180,9	100
28	10	HD	180,9	100
28	20	HD	70,9	100
28	40	HD	138	1000
26	40	HD	143	1000
26	20	HD	72,2	100
26	10	HD	181,2	100
26	10	VD	180,9	100
26	20	VD	46	100
26	40	VD	125	1000
24	40	VD	64	1000
24	20	VD	47,9	100
24	10	VD	181,3	100
24	10	HD	181	100
24	20	HD	71,6	100
24	40	HD	72	1000
22	20	HD	74	100
22	10	HD	181,4	100
22	10	VD	181,2	100
22	20	VD	56,7	100
20	10	VD	155,2	100
20	10	HD	181,2	100
18	10	HD	140	100
18	20	HD	85,5	100
18	20	VD	85	100
18	10	VD	157,5	100
16	20	VD	40,8	100
16	10	VD	135,3	100
16	10	HD	154,9	100

Continued

A = intercoil distance s [m]
B = spacing instrument in metres
C = coil orientation
D = apparent conductivity σ_a [mS/m]
E = range of instrument

A	B	C	D	E
16	20	HD	40,2	100
14	10	VD	75,3	100
14	10	HD	90	100
12	10	VD	45	100
12	10	HD	35,7	100

Concluded

Appendix B

Algorithms

Algorithm 1 Backtracking Line Search

$$
\begin{aligned}
&\text{Choose } \bar{\alpha}, \beta \in (0,1), \sigma \in (0,1); \\
&\alpha^k = \bar{\alpha} \\
&\text{while } \phi(\mathbf{x}^{(k)} + \alpha \mathbf{d}^{(k)}) \leq \phi(\mathbf{x}^{(k)}) + \sigma \alpha^{(k)} \nabla \phi(\mathbf{x}^{(k)})^T \mathbf{d}^{(k)} \\
&\qquad \alpha^{(k)} = \beta \alpha^{(k)}
\end{aligned}
\tag{B.1}
$$

Algorithm 2 The BFGS-method with Line Search algorithm [Vogel, 2002].

$$
\begin{aligned}
&\nu = 0; \\
&\mathbf{x}^0 = \text{initial guess for minimum;} \\
&H^0 = \text{initial guess for Hessian} \\
&\mathbf{g}^0 = \text{grad } \phi(\mathbf{x}^{(k)}) \\
&\text{begin quasi-Newton iterations} \\
&\qquad \mathbf{p}^{\nu+1} := -(H^\nu)^{-1} \mathbf{g}^\nu; \,\%\, \textit{Compute quasi-Newton step} \\
&\qquad \tau^{\nu+1} := \arg\min_{\tau>0} \phi(\mathbf{x}^\nu + \tau \mathbf{p}^\nu); \,\%\, \textit{Line search, cf. Section 3.4.1} \\
&\qquad \mathbf{x}^{\nu+1} := \mathbf{x}^\nu + \tau^\nu \mathbf{p}^\nu; \,\%\, \textit{Update approximate solution} \\
&\qquad \mathbf{g}^{\nu+1} = \text{ grad } \phi(\mathbf{x}^{\nu+1}); \\
&\qquad \mathbf{s}^\nu = \mathbf{x}^{\nu+1} - \mathbf{x}^\nu \\
&\qquad \mathbf{y}^\nu = \mathbf{g}^{\nu+1} + \mathbf{g}^\nu \\
&\qquad H_{\nu+1} = H^\nu + \frac{\mathbf{y}^\nu \mathbf{y}^{\nu\mathrm{T}}}{\mathbf{y}^{\nu\mathrm{T}} \mathbf{s}^\nu} - \frac{H^\nu \mathbf{s}^\nu \mathbf{s}^{\nu\mathrm{T}} H^{\nu\mathrm{T}}}{\mathbf{s}^{\nu\mathrm{T}} H^\nu \mathbf{s}^\nu}; \,\%\, \textit{BFGS-formula} \\
&\qquad \nu := \nu + 1; \\
&\text{end quasi-Newton iterations}
\end{aligned}
\tag{B.2}
$$

Algorithm 3 Adaptive Pruning Algorithm, from [Hansen et al., 2007].

$\hat{p} = \min(5, p-1)$;
Stage one: : while $\hat{p} < 2(p-1)$
$\hat{p} = \min(\hat{p}, p-1)$
Create a pruned L-curve consisting of the $\hat{p}$ largest line segments.
For each corner location routine
Locate the corner $\mathcal{P}_k$ using the pruned L-curve
Add the corner to the list: $\mathcal{L} = \mathcal{L} \cup \{\mathcal{P}_k\}$
$\hat{p} = 2\hat{p}$
Stage two: : if $\#\mathcal{L} = 1$ then $k = k_1$; return.
Otherwise for $i = 1, \cdots, \#\mathcal{L} - 1$
Compute the slope ϕ_i associated with point $\mathcal{P}_k$ in $\mathcal{L}$.
If $\max\{\phi_i\} < \frac{\pi}{4}$ then $k = \max\{k_i\}$; return
Otherwise let $k = \min\{k_i : \phi_i > \frac{\pi}{4} \wedge \phi(k_{i-1}, k_i, k_{i+1}) < 0\}$

Appendix C

Multiscale Optimization Approaches

[Bunks et al., 1995] proposes a *multigrid method* as an alternative to iterative inversion methods for inverting seismic data from complicated earth models. They argue that the primary difficulty is the presence of numerous local minima in the objective function. The multigrid method decomposes a problem by scale, which is illustrated for an objective function in Figure C.1(a). First, the problem is Nyquist sampled on a smaller grid. This accelerates the compuational cost and the long scale component is more manifest. For the largest length scale (see Figure C.1(e)), gradient methods easily find the global maximum (the objective function shown in the figure is convex). That global minimum serves as a initial guess for the next optimization problem: the true objective function sampled at a slightly more dense grid. This the so-called 'fining-up' procedure. This approach may succeed in finding the global minimum for the original objective function in Figure C.1(a).

There exist also probabilistic multiscale approaches, such as *simulated annealing* [Kirkpatrick et al., 1983]. This algorithm follows a probabilistic walk throughout the model space. At every step from model $\mathbf{m}^{(\mu)}$ to $\mathbf{m}^{(\nu)}$, there is a probability $P(\mu \to \nu)$ that the step is accepted. The stochastic dynamics is often split as

$$P(\mu \to \nu) = g(\mu \to \nu)A(\mu \to \nu) \tag{C.1}$$

where μ, ν are two states or models, $g(\mu \to \nu)$ is called the selection probability and $A(\mu \to \nu)$ is the acceptance ratio. The selection probability tells us which states can be generated by the algorithm from a given initial state. The acceptance ratio determines the fraction of times that an actual transition takes place. In simulated annealing, the acceptance ratio will depend on the value of the objective function $\phi(\mathbf{x}^{\mu})$ and a parameter T, the temperature. An example of such an acceptance ratio is found in the Metropolis-Hastings algorithm[1] [Chib and Greenberg, 1995], a well known Monte Carlo type of random walk. More precisely, the random walk is a Markov chain that satisfies the ergodicity[2] and detailed balance condition[3].

[1] Of course, other acceptance ratios can be considered [Khachaturyan et al., 1981].

[2] Ergodicity means that every state, or model in the model space, is reachable. For all states, there is a nonzero probability that the random walker reaches the state.

[3] The detailed balance condition is used to generate a stochastic dynamics which yields a specific stationary state. The condition requires that at an equilibrium state, each elementary process should be equilibrated by its reverse process.

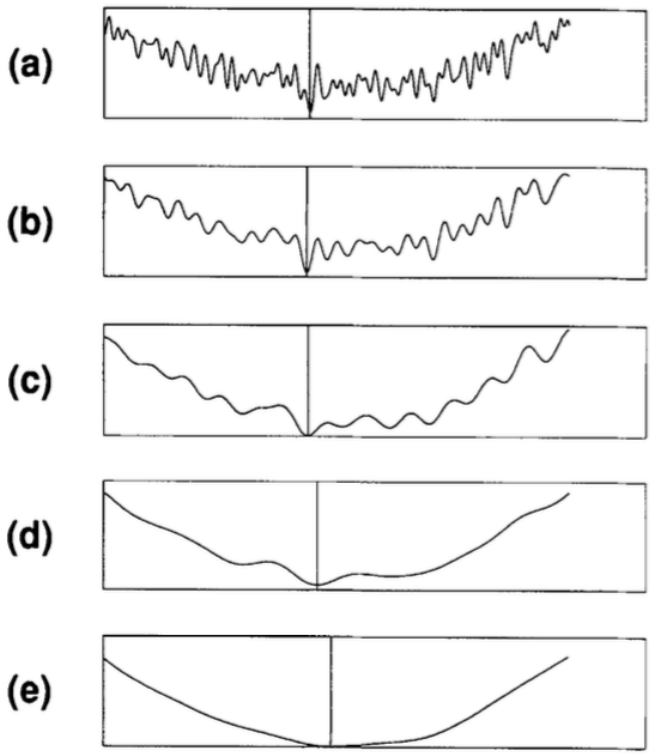


Figure C.1: Illustration of the multigrid method for non-linear problems. Figure retrieved from [Bunks et al., 1995].

In the Metropolis algorithm, the acceptance ratio is based on the Boltzmann distribution:

$$A(\mu \to \nu) = \begin{cases} e^{-\beta(E_\nu - E_\mu)} & \text{if } E_\nu > E_\mu \\ 1 & \text{otherwise} \end{cases}, \tag{C.2}$$

where $\beta = 1/(kT)$ is the thermodynamic beta. We can interpret the energy as the value of the objective function ϕ: If the energy or value of the objective function of a step $\mu \to \nu$ decreases, then that step is always accepted (i.e. $A(\mu \to \nu = 1)$). In the other case, when a step is proposed that increases the energy or the objective function, the probability A is smaller than 1 and decreases for larger energy differences. The temperature plays an important role. When the temperature is high, almost every step is accepted, since $e^{-\beta(E_\nu - E_\mu)} \approx 1$, whereas with low temperatures almost only steps are accepted that move towards a minimum of the objective function.

The algorithm slowly decreases its temperature and thus slowly decreases the probability that worse solutions (i.e. $\phi(\mathbf{m}^{(k+1)}) > \phi(\mathbf{m}^{(k)})$) are accepted. The advantage of these large temperatures is that the model space is well explored: It allows an extensive search for a global minimum.

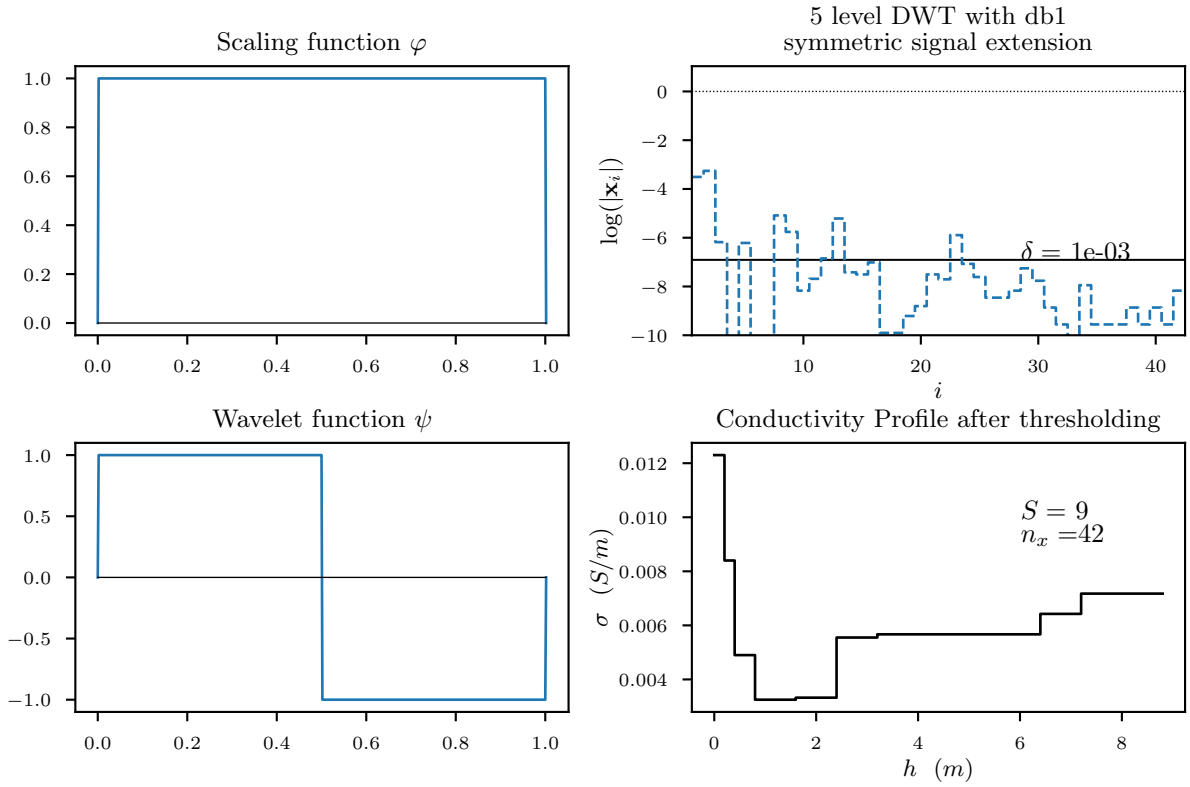
Thus in general, the algorithm selects a random model at each time step and then calculates the probability of accepting the step (depending on the change in the objective function and energy). It then generates a random number p_r that serves to decide if the step is accepted or not. During the search, the temperature is decreased slowly. There is not only the Metropolis algorithm. Simulated annealing can be designed such that it performs its optimization at different scales, by making the Markov chain step size dependent on the temperature. High temperatures allows a sampling of the objective function at long scales, while low temperatures correspond with smaller steps. Simulated annealing slowly cools down the system, that is a slow decrease in the probability of accepting a worse solution ($\phi(\mathbf{m}^{(k+1)}) > \phi(\mathbf{m}^{(k)})$).

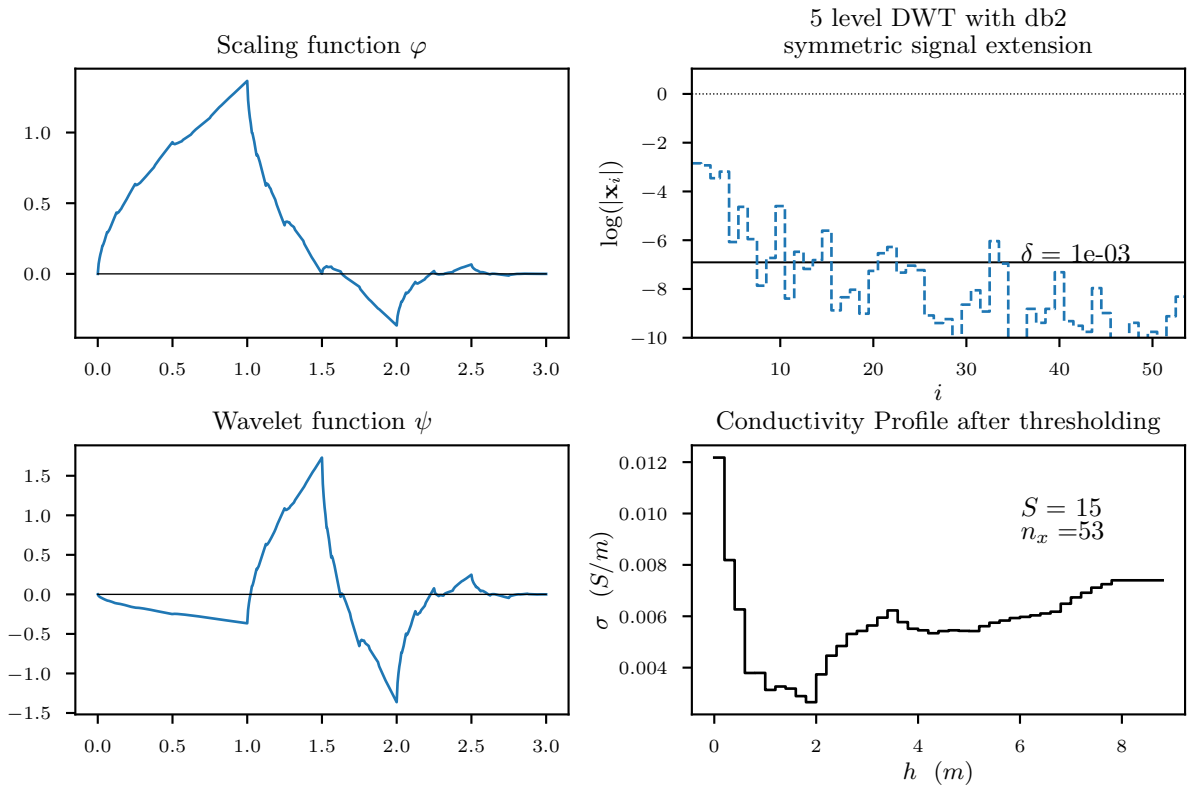
Simulated annealing does not make use of gradient information of the objective function. Consequently, the method does not take into account in which direction there is a descent. Also, it will take many steps before the algorithm gets a precise minimum.

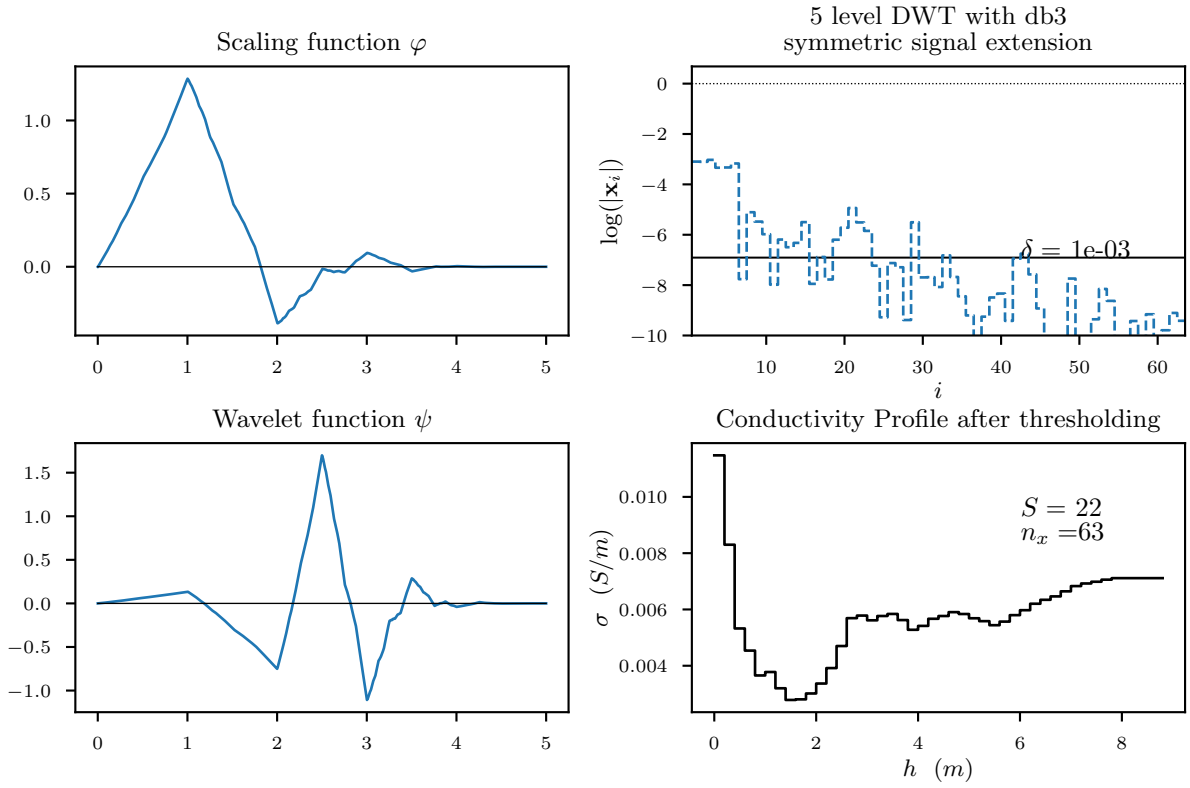
Finally, wavelet theory is another domain of research that decomposes problems by scale [Mallat, 1999]. We cite [Bunks et al., 1995]: “The method of the multigrid method is based on a frequency decomposition; however, wavelet decomposition might be more efficient and further reduce the computational burden.”.

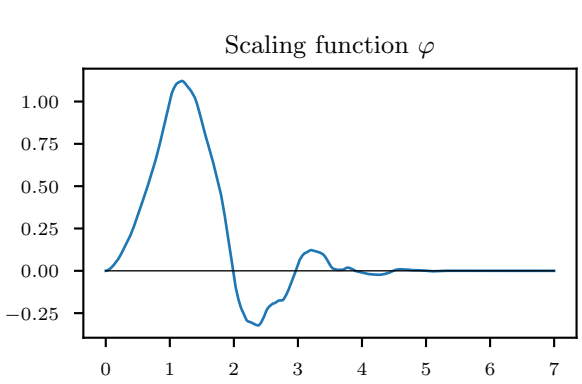
Appendix D

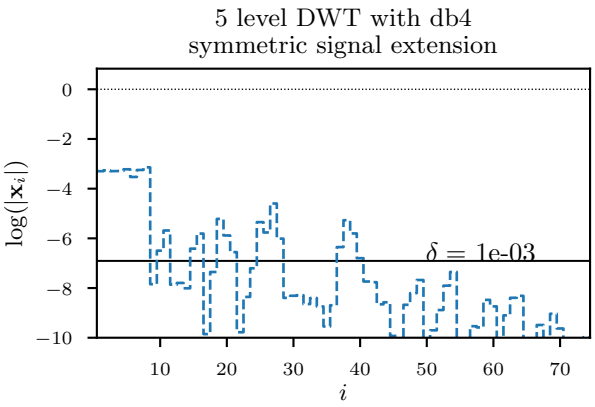
Wavelet Representation with Daubechies Wavelets

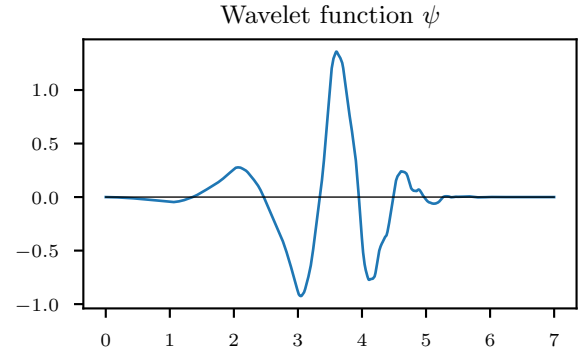


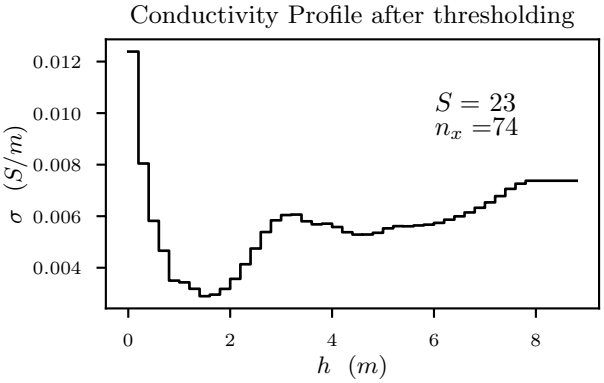

Scaling function φ
5 level DWT with db2
symmetric signal extension
$\log(|\mathbf{x}_i|)$
δ = 1e-03
i
Wavelet function ψ
Conductivity Profile after thresholding
$S = 15$
n_x =53
σ (S/m)
h (m)

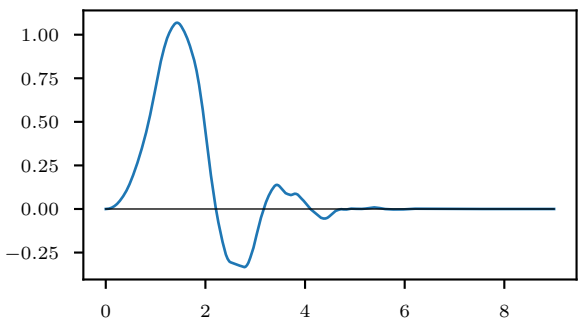

Scaling function φ
5 level DWT with db3
symmetric signal extension
$\log(|\mathbf{x}_i|)$
δ = 1e-03
i
Wavelet function ψ
Conductivity Profile after thresholding
$S = 22$
n_x =63
σ (S/m)
h (m)

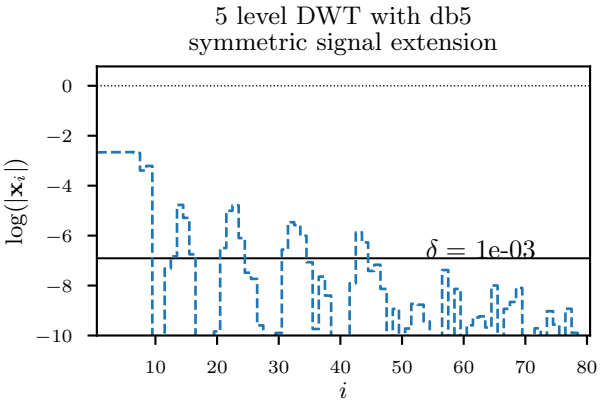

Scaling function φ

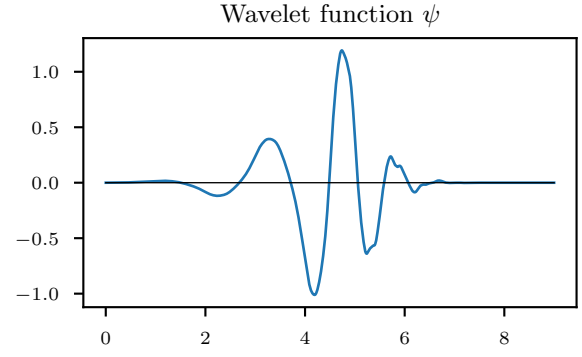

5 level DWT with db4
symmetric signal extension
$\log(|\mathbf{x}_i|)$
δ = 1e-03
i

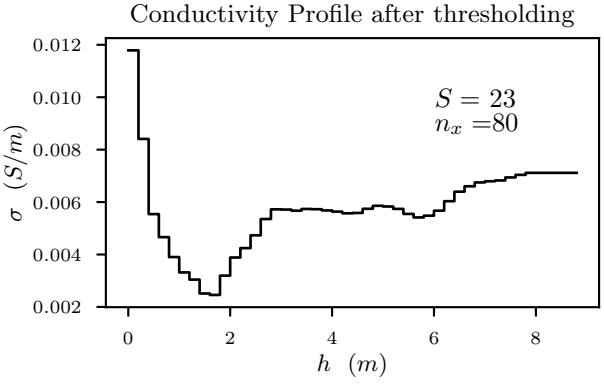

Wavelet function ψ

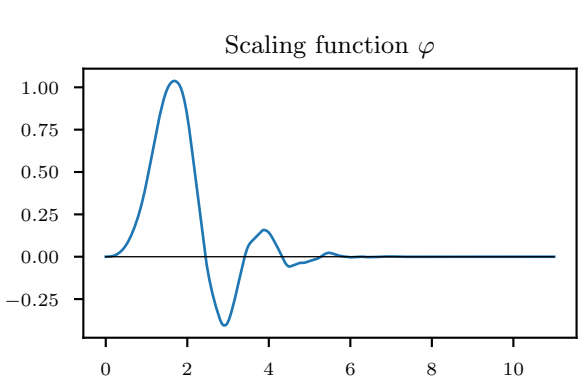

Conductivity Profile after thresholding
S = 23
n_x =74
σ (S/m)
h (m)

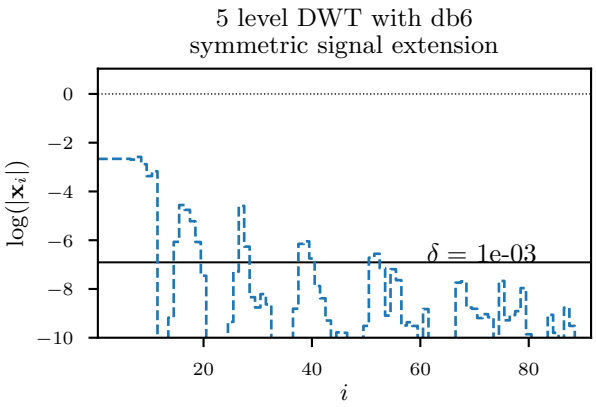

Scaling function φ

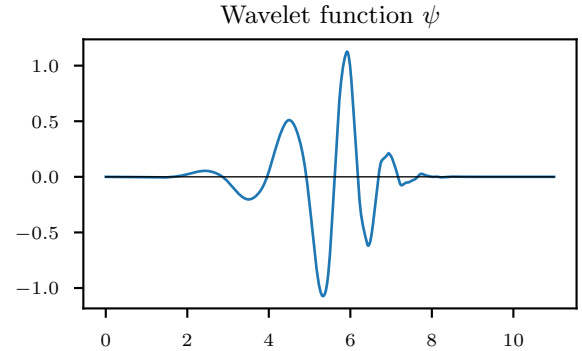

5 level DWT with db5
symmetric signal extension
$\log(|\mathbf{x}_i|)$
δ = 1e-03
i

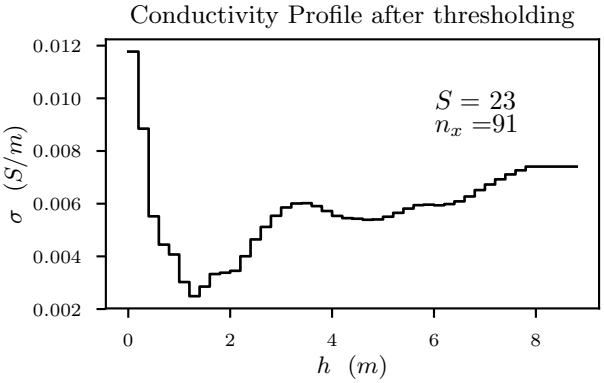

Wavelet function ψ

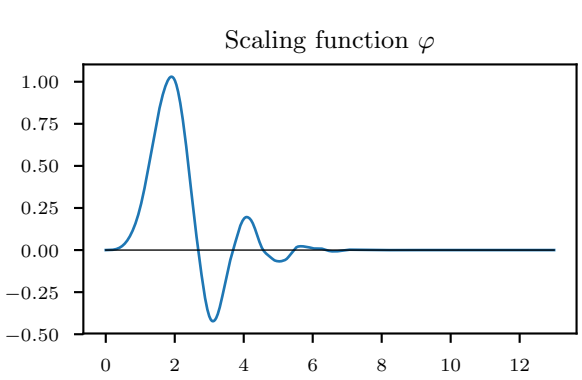

Conductivity Profile after thresholding
S = 23
n_x =80
σ (S/m)
h (m)

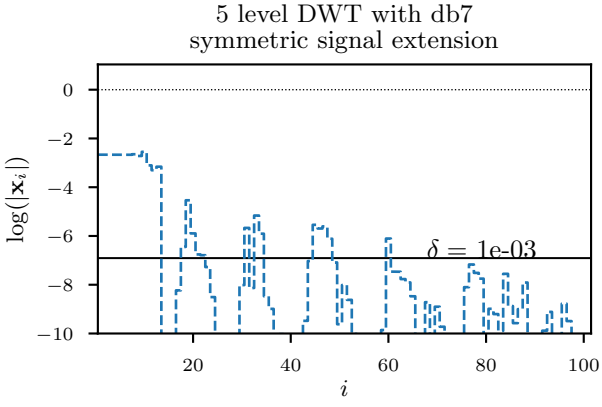

Scaling function φ

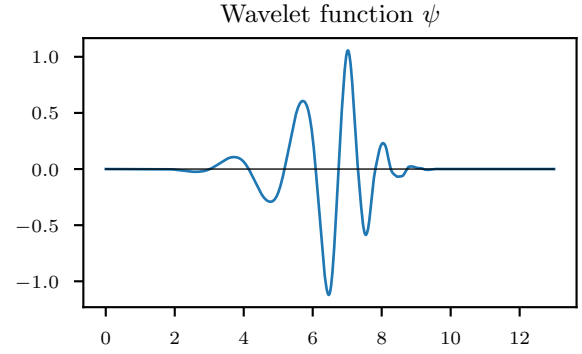

5 level DWT with db6
symmetric signal extension
$\log(|\mathbf{x}_i|)$
δ = 1e-03
i

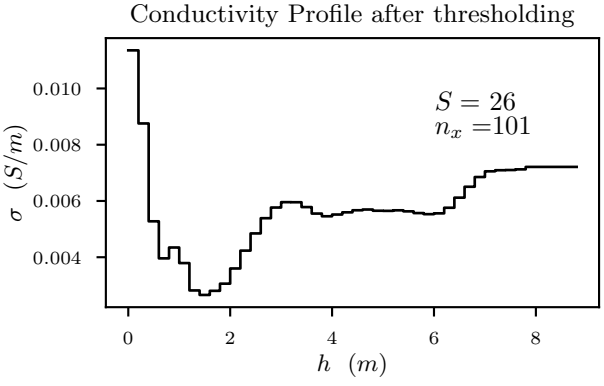

Wavelet function ψ

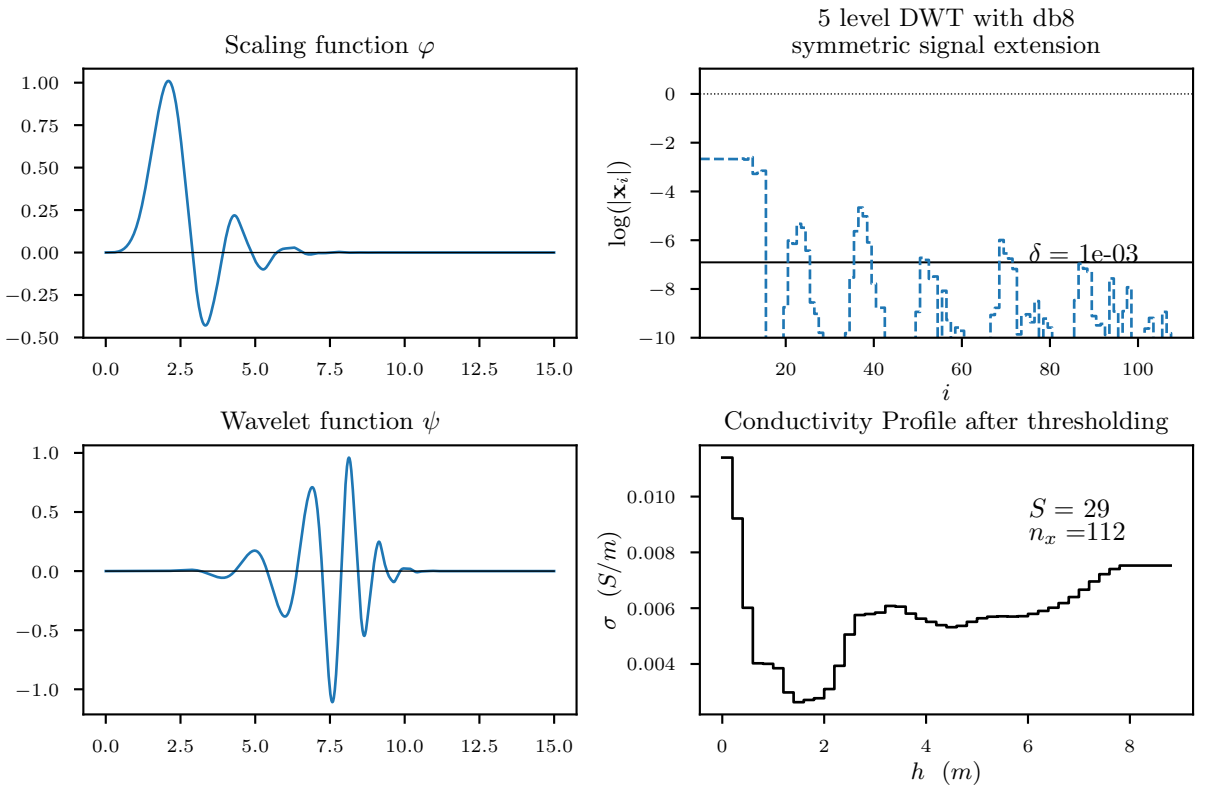

Conductivity Profile after thresholding
S = 23
n_x =91
σ (S/m)
h (m)

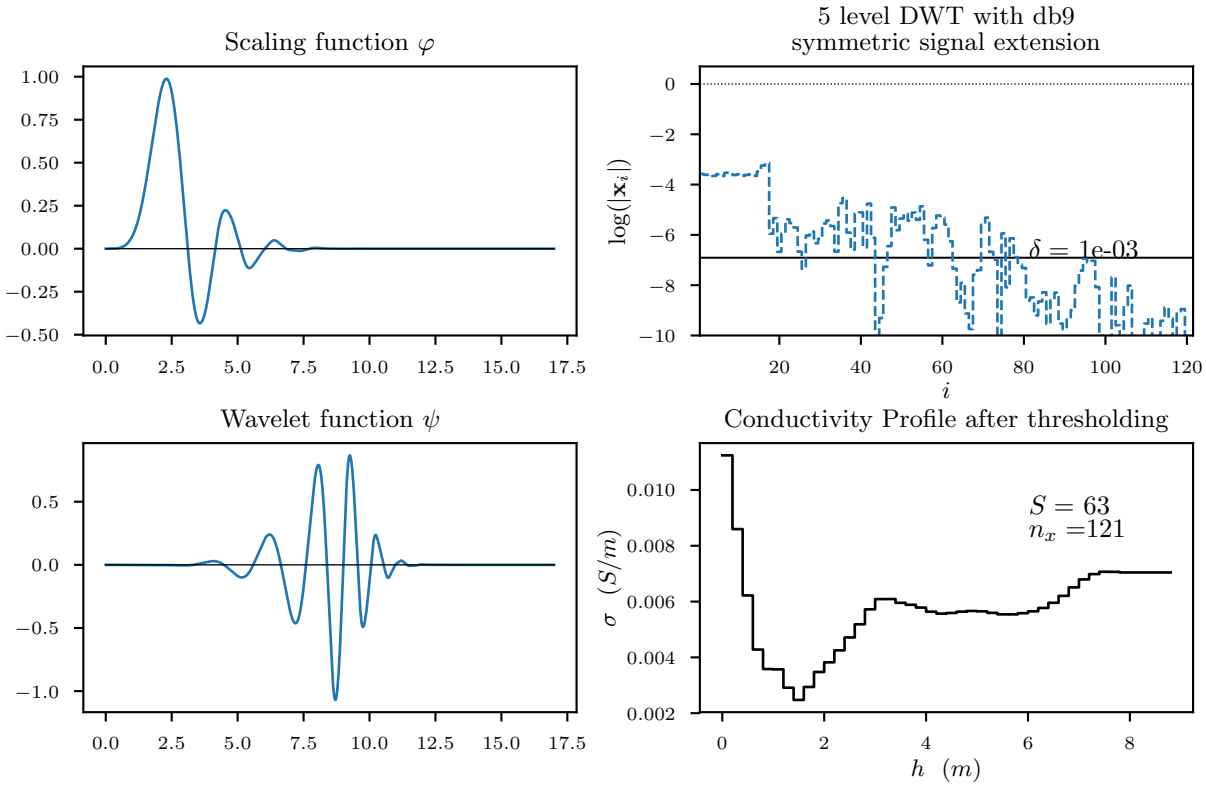

Scaling function φ


5 level DWT with db7
symmetric signal extension
$\log(|\mathbf{x}_i|)$
δ = 1e-03
i


Wavelet function ψ


Conductivity Profile after thresholding
S = 26
n_x =101
σ (S/m)
h (m)


Scaling function φ
5 level DWT with db8
symmetric signal extension
$\log(|\mathbf{x}_i|)$
δ = 1e-03
i
Wavelet function ψ
Conductivity Profile after thresholding
S = 29
n_x =112
σ (S/m)
h (m)


Scaling function φ
5 level DWT with db9
symmetric signal extension
$\log(|\mathbf{x}_i|)$
δ = 1e-03
i
Wavelet function ψ
Conductivity Profile after thresholding
S = 63
n_x =121
σ (S/m)
h (m)

Bibliography

[Abramowitz and Stegun, 1965] Abramowitz, M. and Stegun, I. A. (1965). *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, volume 55. Courier Corporation.

[Archie et al., 1942] Archie, G. E. et al. (1942). The electrical resistivity log as an aid in determining some reservoir characteristics. *Transactions of the AIME*, 146(01):54–62.

[Aster et al., 2018] Aster, R. C., Borchers, B., and Thurber, C. H. (2018). *Parameter estimation and inverse problems*. Elsevier.

[Bultheel, 2003] Bultheel, A. (2003). Wavelets with applications in signal and image processing. *Course material University of Leuven, Belgium*.

[Bunks et al., 1995] Bunks, C., Saleck, F. M., Zaleski, S., and Chavent, G. (1995). Multiscale seismic waveform inversion. *Geophysics*, 60(5):1457–1473.

[Byrd et al., 1995] Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208.

[Černá and Finěk, 2011] Černá, D. and Finěk, V. (2011). Construction of optimally conditioned cubic spline wavelets on the interval. *Advances in Computational Mathematics*, 34(2):219–252.

[Chang et al., 2000] Chang, S. G., Yu, B., and Vetterli, M. (2000). Adaptive wavelet thresholding for image denoising and compression. *IEEE transactions on image processing*, 9(9):1532–1546.

[Chib and Greenberg, 1995] Chib, S. and Greenberg, E. (1995). Understanding the metropolis-hastings algorithm. *The american statistician*, 49(4):327–335.

[Christensen, 2014] Christensen, N. B. (2014). Sensitivity functions of transient electromagnetic methods. *Geophysics*, 79(4):E167–E182.

[Christensen, 2001] Christensen, O. (2001). Frames, riesz bases, and discrete gabor/wavelet expansions. *Bulletin of the American Mathematical Society*, 38(3):273–291.

[Cohen et al., 1993] Cohen, A., Daubechies, I., and Vial, P. (1993). Wavelets on the interval and fast wavelet transforms. *Applied and computational harmonic analysis*.

[Copejans, 2007] Copejans, E. (2007). Slufters de panne — crest project. http://www.crestproject.be/en/node/30?album=214&pic=9824. (Accessed on 05/02/2019).

[Corwin and Lesch, 2003] Corwin, D. and Lesch, S. (2003). Application of soil electrical conductivity to precision agriculture. *Agronomy journal*, 95(3):455–471.

[Daubechies, 1988] Daubechies, I. (1988). Orthonormal bases of compactly supported wavelets. *Communications on pure and applied mathematics*, 41(7):909–996.

[Daubechies, 1992] Daubechies, I. (1992). *Ten lectures on wavelets*, volume 61. Siam.

[Day-Lewis, 2018] Day-Lewis, F. D. (2018). Geophysical tomography: The current state of research, challenges, and path forward. In *SEG Technical Program Expanded Abstracts 2018*, pages 5477–5481. Society of Exploration Geophysicists.

[Deleersnyder et al., 2017] Deleersnyder, W., Spenninck, J., and Vantomme, M. (2017). Electromagnetic induction scanning of stratified media. Bachelor thesis.

[Dierckx et al., 2018] Dierckx, H., De Blauwe, K., Van Meirvenne, M., and Verschelde, H. (2018). Inversion of electromagnetic induction data using a 2d geophysical response function. *arXiv preprint arXiv:1805.06003*.

[Ekblom, 1987] Ekblom, H. (1987). The l1-estimate as limiting case of an lp-or huber-estimate. In *Statistical data analysis based on the L1-norm and related methods: 31/08/1987-04/09/1987*, pages 109–116. Elsevier.

[Farquharson and Oldenburg, 1993] Farquharson, C. and Oldenburg, D. (1993). Inversion of time-domain electromagnetic data for a horizontally layered earth. *Geophysical Journal International*, 114(3):433–442.

[Farquharson, 2007] Farquharson, C. G. (2007). Constructing piecewise-constant models in multidimensional minimum-structure inversions. *Geophysics*, 73(1):K1–K9.

[Farquharson and Oldenburg, 1998] Farquharson, C. G. and Oldenburg, D. W. (1998). Non-linear inversion using general measures of data misfit and model structure. *Geophysical Journal International*, 134(1):213–227.

[Farquharson and Oldenburg, 2004] Farquharson, C. G. and Oldenburg, D. W. (2004). A comparison of automatic techniques for estimating the regularization parameter in non-linear inverse problems. *Geophysical Journal International*, 156(3):411–425.

[Ge et al., 2011] Ge, D., Jiang, X., and Ye, Y. (2011). A note on the complexity of l p minimization. *Mathematical programming*, 129(2):285–299.

[Geonics, 2012] Geonics (2012). Geonics product catalogue. http://www.geonics.com. (Accessed on 04/16/2019).

[Golub and Van Loan, 1996] Golub, H. and Van Loan, C. F. (1996). Matrix computations, johns hopkins uni. *Press, London*.

[Google, 2019] Google (2019). Google maps. https://www.google.be/maps. (Accessed on 03/20/2019).

[Grasmair et al., 2015] Grasmair, M., Haltmeier, M., and Scherzer, O. (2015). Sparsity in inverse geophysical problems. *Handbook of Geomathematics*, pages 1659–1687.

[Griffiths, 2005] Griffiths, D. J. (2005). *Introduction to electrodynamics*. AAPT.

[Grootjans et al., 1998] Grootjans, A., Ernst, W., and Stuyfzand, P. (1998). European dune slacks: strong interactions of biology, pedogenesis and hydrology. *Trends in Ecology & Evolution*, 13(3):96–100.

[Guitton, 2012] Guitton, A. (2012). Blocky regularization schemes for full-waveform inversion. *Geophysical Prospecting*, 60(5):870–884.

[Guitton and Symes, 2003] Guitton, A. and Symes, W. W. (2003). Robust inversion of seismic data using the huber norm. *Geophysics*, 68(4):1310–1319.

[Ha et al., 2009] Ha, T., Chung, W., and Shin, C. (2009). Waveform inversion using a back-propagation algorithm and a huber function norm. *Geophysics*, 74(3):R15–R24.

[Haber, 2014] Haber, E. (2014). *Computational methods in geophysical electromagnetics*, volume 1. SIAM.

[Haber and Tenorio, 2003] Haber, E. and Tenorio, L. (2003). Learning regularization functionals—a supervised training approach. *Inverse Problems*, 19(3):611.

[Hansen, 1994] Hansen, P. C. (1994). Regularization tools: a matlab package for analysis and solution of discrete ill-posed problems. *Numerical algorithms*, 6(1):1–35.

[Hansen, 2010] Hansen, P. C. (2010). *Discrete inverse problems: insight and algorithms*, volume 7. Siam.

[Hansen et al., 2007] Hansen, P. C., Jensen, T. K., and Rodriguez, G. (2007). An adaptive pruning algorithm for the discrete l-curve criterion. *Journal of computational and applied mathematics*, 198(2):483–492.

[Hansen and O'Leary, 1993] Hansen, P. C. and O'Leary, D. P. (1993). The use of the l-curve in the regularization of discrete ill-posed problems. *SIAM Journal on Scientific Computing*, 14(6):1487–1503.

[Heirtzler et al., 1968] Heirtzler, J., Dickson, G., Herron, E., Pitman, W., and Le Pichon, X. (1968). Marine magnetic anomalies, geomagnetic field reversals, and motions of the ocean floor and continents. *Journal of Geophysical Research*, 73(6):2119–2136.

[Hermans and Irving, 2017] Hermans, T. and Irving, J. (2017). Facies discrimination with electrical resistivity tomography using a probabilistic methodology: effect of sensitivity and regularisation. *Near Surface Geophysics*, 15(1):13–25.

[Hermans et al., 2012] Hermans, T., Vandenbohede, A., Lebbe, L., Martin, R., Kemna, A., Beaujean, J., and Nguyen, F. (2012). Imaging artificial salt water infiltration using electrical resistivity tomography constrained by geostatistical data. *Journal of Hydrology*, 438:168–180.

[Horn et al., 1990] Horn, R. A., Horn, R. A., and Johnson, C. R. (1990). *Matrix analysis*, chapter 7. Cambridge university press.

[Hubbard and Rubin, 2005] Hubbard, S. S. and Rubin, Y. (2005). Introduction to hydrogeophysics. In *Hydrogeophysics*, pages 3–21. Springer.

[Huber et al., 1964] Huber, P. J. et al. (1964). Robust estimation of a location parameter. *The annals of mathematical statistics*, 35(1):73–101.

[Johnson et al., 2003] Johnson, H., Johnson, H. W., and Graham, M. (2003). *High-speed signal propagation: advanced black magic*. Prentice Hall Professional.

[Jones et al., 2001] Jones, E., Oliphant, T., Peterson, P., et al. (2001). SciPy: Open source scientific tools for Python. [Online; accessed 12-01-2019].

[Jørgensen, 2018] Jørgensen, F. (2018). Topsoil - resilient soil and water resources, understanding the water beneath your feet. In *Midterm Catalogue*. Region Midtjylland.

[Kaipio and Somersalo, 2006] Kaipio, J. and Somersalo, E. (2006). *Statistical and computational inverse problems*, volume 160, chapter 2. Springer Science & Business Media.

[Kearey et al., 2013] Kearey, P., Brooks, M., and Hill, I. (2013). *An introduction to geophysical exploration*. John Wiley & Sons.

[Khachaturyan et al., 1981] Khachaturyan, A., Semenovsovskaya, S., and Vainshtein, B. (1981). The thermodynamic approach to the structure analysis of crystals. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*, 37(5):742–754.

[Kirkpatrick et al., 1983] Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983). Optimization by simulated annealing. *science*, 220(4598):671–680.

[Kumar and Foufoula-Georgiou, 1997] Kumar, P. and Foufoula-Georgiou, E. (1997). Wavelet analysis for geophysical applications. *Reviews of geophysics*, 35(4):385–412.

[Lee, 2014] Lee, A. (2014). pyswarm: Particle swarm optimization (pso) with constraint support.

[Lee et al., 2006] Lee, G., Wasilewski, F., Gommers, R., Wohlfahrt, K., O'Leary, A., and Nahrstaedt, H. (2006). Pywavelets–wavelet transforms in python.

[Liu et al., 2017] Liu, Y., Farquharson, C. G., Yin, C., and Baranwal, V. C. (2017). Wavelet-based 3-d inversion for frequency-domain airborne em data. *Geophysical Journal International*, 213(1):1–15.

[Loeb, 2016] Loeb, P. A. (2016). *Real Analysis*, chapter 9. Springer International Publishing, Cham.

[Loke et al., 2013] Loke, M., Chambers, J., Rucker, D., Kuras, O., and Wilkinson, P. (2013). Recent developments in the direct-current geoelectrical imaging method. *Journal of Applied Geophysics*, 95:135–156.

[Luisier et al., 2007] Luisier, F., Blu, T., and Unser, M. (2007). A new sure approach to image denoising: Interscale orthonormal wavelet thresholding. *IEEE Transactions on image processing*, 16(3):593–606.

[Mallat, 1999] Mallat, S. (1999). *A wavelet tour of signal processing*. Elsevier.

[Mallat, 1989] Mallat, S. G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (7):674–693.

[MathWorks, 2019] MathWorks (2019). Wavelet toolbox documentation. `https://nl.mathworks.com/help/wavelet/`. (Accessed on 02/18/2019).

[Maveau et al., 2017] Maveau, B., Delrue, S., and Dudal, D. (2017). A damped forward emi model for a horizontally stratified earth.

[McNeill, 1980] McNeill, J. (1980). Electromagnetic terrain conductivity measurement at low induction numbers.

[Misiti et al., 2009] Misiti, M., Misiti, Y., Oppenheim, G., and Poggi, J. (2009). Matlab wavelet toolbox tm 4 user's guide. *The MathWorks, Inc. Natick, Massachusetts. 153p.*

[Moré and Thuente, 1994] Moré, J. J. and Thuente, D. J. (1994). Line search algorithms with guaranteed sufficient decrease. *ACM Transactions on Mathematical Software (TOMS)*, 20(3):286–307.

[Nabighian and Corbett, 1988] Nabighian, M. N. and Corbett, J. D. (1988). *Electromagnetic methods in applied geophysics: Application*, volume 2. Society of Exploration Geophysicists Tulsa.

[Nocedal and Wright, 2006] Nocedal, J. and Wright, S. (2006). *Numerical optimization*. Springer Science & Business Media.

[Oldenburg and Li, 2005] Oldenburg, D. W. and Li, Y. (2005). Inversion for applied geophysics: A tutorial. *Investigations in geophysics*, 13:89–150.

[Oware et al., 2019] Oware, E. K., Irving, J., and Hermans, T. (2019). Basis-constrained bayesian mcmc difference inversion for geoelectrical monitoring of hydrogeological processes. *Geophysics*, 84(4):1–21.

[Patrinos, 2018] Patrinos, P. (2018). Optimization. *Course material University of Leuven, Belgium*.

[Portniaguine and Zhdanov, 1999] Portniaguine, O. and Zhdanov, M. S. (1999). Focusing geophysical inversion images. *Geophysics*, 64(3):874–887.

[Ralph-Uwe et al., 2008] Ralph-Uwe, B., Ernst, O. G., and Spitzer, K. (2008). Fast 3-d simulation of transient electromagnetic fields by model reduction in the frequency domain using krylov subspace projection. *Geophysical Journal International*, 173(3):766–780.

[Ramirez et al., 2013] Ramirez, C., Kreinovich, V., and Argaez, M. (2013). Why l1 is a good approximation to l0: A geometric explanation. *Journal of Uncertain Systems*, 7(3):203–207.

[Saslow, 1992] Saslow, W. (1992). Maxwell's theory of eddy currents in thin conducting sheets, and applications to electromagnetic shielding and maglev. *American journal of physics*, 60(8):693–711.

[SkyTEM, 2019] SkyTEM (2019). Skytem-306-hp-1.pdf. `https://skytem.com/wp-content/uploads/SkyTEM-306-HP-1.pdf`. (Accessed on 05/30/2019).

[Stefanesco et al., 1930] Stefanesco, S., Schlumberger, C., and Schlumberger, M. (1930). Sur la distribution électrique potentielle autour d'une prise de terre ponctuelle dans un terrain à couches horizontales, homogènes et isotropes. *Journal de Physique et le Radium*, 1(4):132–140.

[Strang and Nguyen, 1996] Strang, G. and Nguyen, T. (1996). *Wavelets and filter banks*. SIAM.

[Tang et al., 2018] Tang, P., Chen, F., Jiang, A., Zhou, W., Wang, H., Leucci, G., de Giorgi, L., Sileo, M., Luo, R., Lasaponara, R., and Masini, N. (2018). Multi-frequency electromagnetic induction survey for archaeological prospection: Approach and results in han hangu pass and xishan yang in china. *Surv. Geophys.*

[Tikhonov et al., 2013] Tikhonov, A. N., Goncharsky, A., Stepanov, V., and Yagola, A. G. (2013). *Numerical methods for the solution of ill-posed problems*, volume 328. Springer Science & Business Media.

[Trefethen and Bau III, 1997] Trefethen, L. N. and Bau III, D. (1997). *Numerical linear algebra*, volume 50. Siam.

[Verwaest et al., 2005] Verwaest, T., De Wolf, P., Herrier, J.-L., and Leten, M. (2005). Windows in the dunes–the creation of sea inlets in the nature reserve de westhoek in de panne. *Proceedings 'Dunes and Estuaries*, pages 33–439.

[Vogel, 2002] Vogel, C. R. (2002). *Computational methods for inverse problems*, volume 23. Siam.

[Wait, 1982] Wait, J. (1982). *Geo-electromagnetism*. Elsevier.

[Wait, 1951] Wait, J. R. (1951). The magnetic dipole over the horizontally stratified earth. *Canadian Journal of Physics*, 29(6):577–592.

[Wait, 1962] Wait, J. R. (1962). A note on the electromagnetic response of a stratified earth. *Geophysics*, 27(3):382–385.

[Walker, 1997] Walker, J. S. (1997). Fourier analysis and wavelet analysis. *Notices of the AMS*, 44(6):658–670.

[Ward and Hohmann, 1988] Ward, S. H. and Hohmann, G. W. (1988). Electromagnetic theory for geophysical applications. In *Electromagnetic Methods in Applied Geophysics: Voume 1, Theory*, pages 130–311. Society of Exploration Geophysicists.

[Zhdanov, 2015] Zhdanov, M. S. (2015). *Inverse theory and applications in geophysics*, volume 36. Elsevier.

[Zhu et al., 1997] Zhu, C., Byrd, R. H., Lu, P., and Nocedal, J. (1997). Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on Mathematical Software (TOMS)*, 23(4):550–560.

Department of
Physics and Astronomy
Celestijnenlaan 200 D bus 2412
3001 HEVERLEE, BELGIË
tel. + 32 16 32 71 24
www.fys.kuleuven.be