Inversion of CSEM Data for Subsurface Structure Identification and Numerical Assessment of the Upstream Mobility Scheme

Svenn Tveit

Dissertation for the degree of Philosophiae Doctor (PhD)

Department of Mathematics
University of Bergen

November 2014

Preface

This dissertation is submitted as a partial fulfillment of the requirements for the degree of Philosophiae Doctor (PhD) in Applied and Computational Mathematics at the University of Bergen (UiB), Norway.

The PhD project ‘Robust inversion of controlled source electromagnetic (CSEM) data’ has been supported by VISTA – a basic research program funded by Statoil, conducted in close collaboration with The Norwegian Academy of Science and Letters.

The supervising committee for this PhD project has been Trond Mannseth (Uni Research CIPR, UiB), Martha Lien (Octio AS), and Shaaban A. Bakr (Assuit University, Egypt).

Outline

The work done in this thesis is divided into two separate parts; Part I: ‘Inversion of CSEM Data for Subsurface Structure Identification’ and Part II: ‘Numerical Assessment of the Upstream Mobility Scheme’. These parts provide the scientific backgrounds for the collection of research papers given in Part III. In Part IV, co-authored research papers associated with Part II are given. These are not considered as the main part of this thesis, but are given for completeness.

A brief outline of Part I and Part II follows.

Part I

In Chapter 1, a short introduction that provides an overview of marine exploration and controlled source electromagnetic (CSEM) inversion is given.

A presentation of different aspects of the CSEM method, together with a brief overview of some related exploration methods follows in Chapter 2.

In Chapter 3, a general discussion of the inverse problem is presented. We also present different solution methods within the classical and Bayesian approach for solving an inverse problem.

Chapter 4 includes an overview of different formulations of Maxwell’s equations and how these are typically solved numerically in the CSEM problem.

The representation of the unknown parameter function in the inverse problem is discussed in Chapter 5. The focus is on a composite parameterization based on the level-set idea. Moreover, a reparameterization of the parameter function is discussed.

In Chapter 6, the concept of kernel functions is discussed in detail. Based on the theory on kernel functions, several applications are present.

A summary of the research papers associated with Part I is given in Chapter 7.

Part II

In Chapter 8, a short introduction to different aspects of reservoir simulation with a focus on the upstream mobility scheme is given.

Chapter 9 introduces two-phase flow in heterogeneous porous media, which, mathematically, is modelled as a hyperbolic conservation law with a discontinuous flux function. A detailed overview of the theory of conservation laws with discontinuous flux is thus also presented.

In Chapter 10, numerical schemes for approximating solutions to conservation laws with discontinuous flux are introduced. Specifically, the upstream mobility scheme for approximating two-phase flow in heterogeneous porous media is presented.

A summary of the research papers associated with Part II is given in Chapter 11.

List of papers

Paper A: S. Tveit, S. A. Bakr, M. Lien, and T. Mannseth, *Identification of subsurface structures using electromagnetic data and shape priors*, submitted to Journal of Computational Physics, 2014.

Paper B: S. Tveit, S. A. Bakr, M. Lien, and T. Mannseth, *Ensemble-based, Bayesian inversion of CSEM data using structural prior information*, in proceedings of the 76th EAGE Conference & Exhibition, Amsterdam, the Netherlands, 2014.

Paper C: S. Tveit, S. A. Bakr, M. Lien, and T. Mannseth, *Ensemble-based Bayesian inversion of CSEM data for subsurface structure identification*, resubmitted to Geophysical Journal International, 2014.

Paper D: S. Tveit and I. Aavatsmark, *Errors in the upstream mobility scheme for countercurrent two-phase flow in heterogeneous porous media*, Computational Geosciences **16**(3), 809-825, 2012.

For completeness, the following co-authored papers are include. These papers are associated with Part II of the thesis.

Paper E: T. S. Mykkeltvedt, I. Aavatsmark, and S. Tveit, *Errors in the upstream mobility scheme for counter-current two-phase flow with discontinuous permeabilities*, in proceedings of the 13th European Conference on the Mathematics of Oil Recovery (ECMOR XIII), Biarritz, France, 2012.

Paper F: T. S. Mykkeltvedt, I. Aavatsmark, and S. Tveit, *On the performance of the upstream mobility scheme applied to counter-current two-phase flow in a heterogeneous porous medium*, in proceedings of the SPE Reservoir Simulation Symposium, Texas, USA, 2013.

Abstract

Part I

In this part of the thesis, two different methodologies for solving the inverse problem of mapping the subsurface electric conductivity distribution using controlled source electromagnetic (CSEM) data are presented. The two inversion methodologies are based on a classical and a Bayesian approach for solving inverse problems, respectively.

In the classical approach, we regularize the inverse problem by incorporating structural prior information available from, e.g., interpreted seismic data. In many cases, the outcome of an interpretation of seismic data cannot be well approximated by a Gaussian distribution. Hence, to incorporate non-Gaussian prior information we have applied the shape prior technique. Here, an implicit transformation of variables facilitates the incorporation of non-Gaussian prior information, at the expense of an application-dependent kernel function.

In the Bayesian approach, a combination of prior knowledge and observed data results in a solution given as a posterior probability density function (PDF). To sample from the posterior PDF, a sequential Bayesian method, the ensemble Kalman filter (EnKF), is applied. Structural prior information is naturally incorporated as a part of the Bayesian framework.

To represent large-scale subsurface structures two model-based, composite parameterizations based on the level-set representation are applied in the inversion methodologies. By using a reduced number of parameters in the representation, a regularization of the inverse problem is achieved. Moreover, it enables the use of second-order gradient-based optimization algorithms in the classical approach.

Part II

In this part of the thesis, a numerical investigation of the upstream mobility scheme for calculating fluid flow in porous media is presented. Previous studies have shown that the upstream mobility scheme experienced erroneous behaviour when approximating pure gravity segregation flow in 1D heterogeneous porous media. The errors shown, however, were small in magnitude. In this work, numerical experiments, where we include both advection and gravity segregation, are conducted. It is shown that the errors produced in this case may be larger in magnitude than for pure gravity segregation, but are only found for countercurrent flow situations.

Acknowledgements

First and foremost, I would like to thank my supervisors Trond Mannseth, Martha Lien, and Shaaban A. Bakr for their advices and guidance during my PhD. I am grateful for the time they took in sharing their vast knowledge through many interesting discussions. Furthermore, I would like to thank Ivar Aavatsmark for our collaboration on Part II of this thesis. Also, I would like to thank Trine S. Mykkeltvedt for our collaboration on the related papers and the many non-scientific discussions.

I am grateful for the financial support from VISTA. Furthermore, I would like to thank my friends and colleagues at Uni Research CIPR for providing a great research and social environment. Especially, I would like to thank Kristian Fossum for taking the time to discuss the (countless) scientific problems I encountered during my PhD; but most of all, I am grateful for our laughs and off-topic conversations, making the PhD-period a fun time.

Finally, I thank my friends for always being supportive and helping me taking my mind off work. And last but not least, I want to express my deepest gratitude to my family. Their support and encouragement have always given me strength during my studies.

Svenn Tveit
Bergen, November 2014

Contents

F On the performance of the upstream mobility scheme applied to counter-current two-phase flow in a heterogeneous porous medium

*"‘Be what you would seem to be’ – or if you’d like it put more simply –
‘Never imagine yourself not to be otherwise than what it might appear to
others that what you were or might have been was not otherwise than what
you had been would have appeared to them to be otherwise.’"*

Alice’s Adventures in Wonderland, Lewis Caroll.

Part I

Scientific background
–
Inversion of CSEM Data for Subsurface Structure Identification

Chapter 1

Introduction

Over the last decades, it has become apparent that satisfying the world's energy need is one of humanity's greatest problems. To meet the energy demands, hydrocarbon energy sources have been, and will still be in the future, a major contributor. The challenge is, of course, that most of the 'easy' reservoirs of hydrocarbon resources have been found and depleted, and thus the more difficult targets remain. In many countries, Norway included, the reservoirs are located offshore, which adds to the challenge.

The cost of drilling wells in a marine environment is expensive. To increase the probability of hitting a hydrocarbon target in a drilling operation, the area is first prospected with a geophysical exploration method. Among the most used geophysical exploration methods are seismic reflection and rarefaction, controlled source electromagnetic (CSEM), magnetotelluric, gravity, magnetics, and ground-penetrating radar. (See Chapter 2 for a description of the first three methods.) Although seismic methods have dominated the marine geophysical exploration industry the last decades, other geophysical methods have shown their usefulness, especially as the hydrocarbon targets have become increasingly harder to locate. In particular, the industry have in recent years applied several geophysical exploration methods together to lower the risk of drilling 'false positives' (i.e., non-existing reservoir targets predicted by a geophysical exploration method).

The overall principle in any geophysical exploration method is to transmit energy into the subsurface and measure the returning signals with receivers at the surface level. The measured signals will change according to the subsurface physical properties, and thus an inversion of the measured signals can provide an image of the subsurface. This image must then be interpreted by geologists and geophysicists, and turned into a geological model where subsurface formations, rock types, and, importantly, the reservoir quality are described. Based on the geological model, an informed economic evaluation of the prospected area can be made before any drilling is done.

In this thesis, we focus on the CSEM exploration method where the transmitted energy is an electromagnetic (EM) field. Most of the EM applications in geophysics attempt to use the measured EM signals to map the electric conductivity distribution in the subsurface (see Sections 2.1 and 2.3). Other physical properties that affect the EM signal propagation are the electric permittivity and magnetic permeability, but they are seldom included in an inversion process.

An important part of the CSEM inversion process is to understand how the signals propagate in the subsurface. Since we cannot directly monitor the signal propagation in the subsurface, we must rely on mathematical equations that model the physics behind the propagation of signals, which, in the case of CSEM, are the Maxwell's equations (see Chapter 4). By setting up the mathematical model in the exact same manner as the real-world exploration survey, we could, in theory, have described the signal propagation by solving the equations. Unfortunately, the mathematical equations are generally not possible to solve analytically, and thus numerical methods must be used (see Section 4.6). In the numerical approach, the subsurface is represented by a set of grid cells, where each grid cell is populated by the parameters governing the signal propagation, which in the CSEM case is electric conductivity. When the numerical model has been set up, usually on a computer, a simulation of the real-world CSEM survey can be done, where the mathematical equations are solved approximately. With today's computational resources, the signal propagation can be modelled with considerable accuracy. Even so, prospecting more difficult areas with complex geological formations requires better and less resource demanding numerical methods. Thus the numerical simulation of CSEM signal propagation is a research area within itself.

Setting up the numerical model in the exact same manner as the real-world experiment is of course impossible. However, we can try to minimize the discrepancy between the outcome of the numerical simulation and the actual measured signals from a survey. The minimum is achieved by changing the electric conductivity values in the numerical model in such a way that it simulates similar signals in the receivers as the measured signals. The numerical model will then provide the sought map of the subsurface electric conductivity distribution. This is the basics in the classical approach for solving an inverse problem (see Section 3.2). Traditionally, the adjustment of the numerical model parameters was done manually, but over the years automatic adjustment of the parameters have been developed. With the advancement of optimization methods, automatic adjustment of parameters have become a much used approach by the geophysical community.

In an inversion process, we often have to deal with uncertainties. The numerical model is not a perfect reconstruction of the real-world survey, and measured signals always contain noise. Moreover, since the signals are measured at sparse locations at the surface, the map of the electric conductivity resulting from an inversion process will necessarily contain uncertainties. It can thus be advantageous to view the outcome of the numerical simulations, the numerical model parameters, and measured signals as stochastic variables associated with a probability density function (PDF). This is the basic set up of a entirely different approach for solving the inverse problem than the classical – the Bayesian approach (see Section 3.3). The solution procedure of a Bayesian approach is to update a prior PDF of the electric conductivity distribution by conditioning to the measured signals using Bayes' rule. The result of this procedure is a posterior PDF where an updated knowledge on the uncertainty of the subsurface electric conductivity is contained.

The focus of this thesis is to develop computationally efficient and robust methodologies for the inversion of CSEM signals. Moreover, we want to incorporate information from other geophysical methods (most notably seismic methods) to increase the reliability of the CSEM inversion result. In particular, we want to use the interpreted result from an inversion of a geophysical method as prior model in an inversion of CSEM signals. To do so, we have investigated solution procedures following both the classical and the Bayesian approach. While prior models are naturally incorporated into the Bayesian framework as a prior PDF, they must be incorporated as an additional term in a classical inversion methodology. Due to the ambiguity in the interpretation procedure, prior models can exhibit non-Gaussian features, and are thus difficult to incorporate. Hence, when developing the classical inversion methodology, the prior models are incorporated using the shape prior technique (see Section 6.3). In short, the shape prior technique allows for the incorporation of non-Gaussian prior models using an implicit transformation of variables.

An important part of this thesis is to produce maps of the subsurface conductivity distribution which renders the complex geological formations known to exist in the subsurface (e.g., faults, pinchouts, anticlines, and so on). To be able to represent the conductivity distribution with such flexiblity, model-based representations based on level sets (see Chapter 5) are considered. When chosen properly, model-based representations can preserve geological knowledge about the subsurface and also reduce the ambiguity in the inversion result.

Chapter 2

Controlled source electromagnetic method and marine exploration

In this chapter, we discuss the CSEM method, together with some selected marine exploration methods. The aim of the discussion is to highlight the strengths and weaknesses of each method, thus providing a sense of what sets each method apart, and, importantly, how the different methods can work together for a richer view of the subsurface geology.

2.1 Controlled source electromagnetic

The CSEM method used today was initially developed by Cox in 1980 [48] to study seafloor geology (although similar configurations had been proposed earlier, see, e.g., [19].) From this point CSEM was further developed both technically and theoretically, but the applications were mostly academic. In 2000, the first field testing of CSEM as a tool for hydrocarbon exploration was conducted [58]. Since then CSEM has been successfully used in many field operations, both as a frontier exploration tool and in conjunction with other exploration methods. For a comprehensive history of CSEM see, e.g., [44, 47]. A detailed review on the instrumentation used in CSEM can be found in [45].

2.1.1 Acquisition – source and receivers

A generic CSEM survey consists of first deploying a set of receivers, which are sensitive to both electric and magnetic fields. These receivers are typically positioned in a straight line or in a 2D array on the sea floor. Since the receivers are released from a boat and freely fall to the sea floor, a perfect set up of the receivers is typically not possible. The receivers record EM signals using a set of electrodes and inductions coils. Unfortunately, magnetic fields are difficult to record due to high noise sensitivity.

The source is towed behind a boat over the acquisition area, transmitting low-frequency (typically 0.1–10 Hz) EM signals into the subsurface. The source is usually a 100–300 m long cable, which is towed 25–100 m over the sea floor. The reason the source has to be towed near the sea floor is to minimize the signal attenuation in conductive sea water, and also to reduce the influence of the ‘airwave’ (see Figure 2.1).

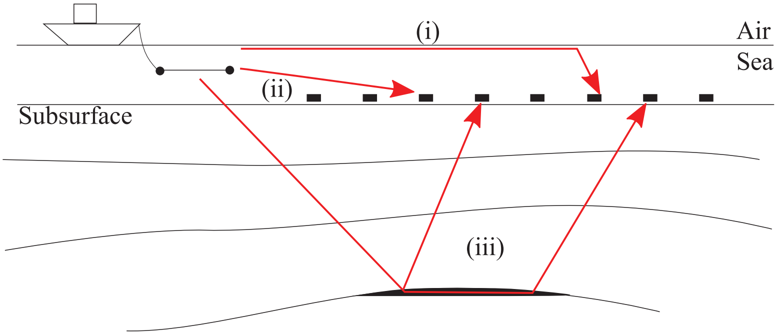

Figure 2.1: An illustration of CSEM signal propagation. (i) ‘Airwave’; (ii) direct wave; (iii) and (iv) signals interacting with the subsurface.

There are four basic source geometries associated with CSEM: horizontal electric dipole (HED), vertical electric dipole (VED), horizontal magnetic dipole (HMD), and vertical magnetic dipole (VMD) [38]. The electric dipole sources are typically current carrying bipoles (i.e., there is a length between the positive and negative pole), which for receivers far away from the source can be treated as dipoles (a point in space with positive and negative pole). The magnetic dipole sources are loops of electric wires that create a magnetic field when a current is active. Both HMD and HED sources provide horizontal and vertical current flows, while VED and VMD sources only provide vertical or horizontal currents, respectively. HED sources are by far the most widely applied, as it is much easier to generate electric currents than magnetic currents [43].

There are two main approaches in CSEM: frequency domain and time domain. The underlying physics of both approaches are the same; EM fields in the frequency domain are Fourier transformed fields from the time domain, and vice versa. However, there is a difference in their practical application. For frequency-domain CSEM, the source transmits broadband waveform signals with high amplitude in some key frequencies, improving the sensitivity to both shallow and deep structures. In time-domain CSEM, the source transmits continuously over the frequency spectrum with a step-on/step-off transmitter current, see, e.g., [91] and references therein.

Lastly, we mention that a towed streamer acquisition has also been developed [59]. Similarly to the seismic method (see Section 2.2 below), a line of receivers is towed behind the source, making the acquisition faster and cheaper, but only inline fields are recorded.

2.1.2 Physical behaviour

A simplified illustration of the CSEM signal propagation is given in Figure 2.1. In reality, the signals cannot be readily explained by simple ray paths as shown in the figure. The signals are 3D vector fields, which interact with the subsurface in a more complicated way than, e.g., seismic. In the following, we give a basic overview of different factors that impact the physical behaviour of EM signals.

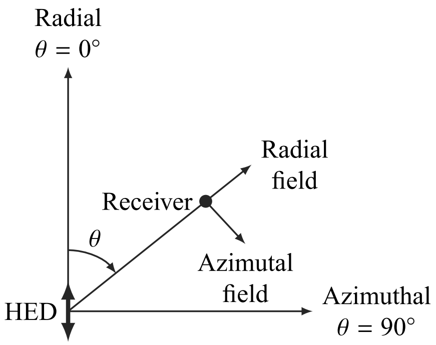

Figure 2.2: Plane view of the HED source-receiver geometry outlining the two modes: radial and azimuthal.

Modes

The source-receiver geometry is important in CSEM as it determines which type of interactions with the subsurface geology that will be recorded. Considering the HED source in Figure 2.2, the EM fields transmitted in-line with the direction of the source are called *radial* (zero azimuth, $\theta = 0°$). On the other hand, the EM signals transmitted in the broadside direction are called *azimuthal* (full azimuth, $\theta = 90°$). For a receiver in between these extremes, both radial and azimuthal fields are recorded.

Electric conductivity

The main physical attribute of CSEM signals is the sensitivity to the subsurface electric conductivity distribution, σ, measured in siemens/meter (S/m) (or, more seldom, in mho/m). Often the reciprocal of conductivity, resistivity, is used, denoted by ρ and measured in ohm/m (Ω/m). In sea water, the conductivity is almost entirely dependent on temperature and salinity, and typically ranges from 3.2–5 S/m.

In the subsurface, the range of conductivity is vast and highly dependent on the environment in which the sediments were deposited and on processes occurring after deposition. Importantly, the uppermost sediments are porous and can thus contain fluids, which contributes greatly to the conductivity distribution in the subsurface. Saline water is the dominating fluid, filling most of the porous media in the upper crust. The difference in conductivity between different brine filled sediments is mostly due to the composition of the rock matrix. Depending on the minerals which the rock consists of, the conductivity of the uppermost layers typically range from 0.1–10 S/m (assuming that the brine contains one type of salt).

Important for marine exploration is the presence of hydrocarbons in the subsurface. Compared to brine, hydrocarbons are regarded as insulators with conductivity typically ranging from 0.01–0.1 S/m. The contrast in conductivity between brine filled and hydrocarbon filled porous media makes CSEM an attractive hydrocarbon indicator. Note that there also exist highly resistive bodies other than hydrocarbons in the subsurface, e.g., tight carbonates and volcanic rock. Thus, the interpretation of CSEM signals can often be difficult.

Table 2.1: Skin depth, δ, for some given frequencies, f, and conductivity values, σ.

$f \backslash \sigma$	100	10	3.33	1	0.1	0.01
0.1	159.15	503.29	872.16	1591.55	5032.92	15915.49
1	50.33	159.15	275.80	503.29	1591.55	5032.92
5	22.51	71.18	123.34	225.08	711.76	2250.79
10	15.92	50.33	87.22	159.15	503.29	1591.55

Mechanisms

There are several mechanisms related to a changing EM field. First, we have the attenuation of a diffusive EM wave in a conductive medium, which can be described by the *skin depth*:

$$\delta = \sqrt{\frac{2}{\omega \mu_0 \sigma}}. \tag{2.1}$$

Here, $\omega = 2\pi f$, with f being the source frequency, and $\mu_0 = 4\pi \times 10^{-7}$ H/m is the magnetic permeability of free space. The skin depth is defined as the distance at which the amplitude of a EM plane wave is reduced by a factor of $1/e$ (≈ 0.37). Although it is defined for a plane wave, it gives an idea of how much a more complicated EM wave will attenuate in the subsurface. The skin depth for some frequencies and conductivity values are given in Table 2.1. From this table it is seen that there is little to no attenuation of the EM fields in very resistive targets like, e.g., hydrocarbon filled sediments. Hence, the EM signals will propagate through hydrocarbon reservoirs as 'guided waves' [163].

The second mechanism is the *galvanic* effect. This effect occurs when currents cross a boundary between two bodies. The continuity equation (see Chapter 4) requires the normal component of the current density to be continuous across a boundary, leading to a jump in the electric field, due to Ohm's law (see Chapter 4). This perturbation in the electric field is measurable at sea floor receivers. Pure radial source-receiver geometry will generate EM fields where the galvanic effect dominates when intersecting a boundary in the subsurface.

The third mechanism is the *inductive* effect. This effect occurs when the current flow is parallel to the boundary between two bodies. At the boundary, the tangential electric field is continuous. Hence, if one of the bodies is more conductive than the other, a strong electric current must flow in the high conductive body, due to Ohm's law (see Chapter 4). This will in turn generate magnetic fields according to Ampere's law (see Chapter 4), which work against the electric field. Consequently, attenuation increases according to the skin depth due to inductive effects. On the other hand, if one of the bodies is more resistive (e.g., hydrocarbon reservoirs), the attenuation will be less than the conductive surroundings, resulting in a change in EM fields measurable by the receivers. The change in amplitude in the inductive effect is much less than in the galvanic effect. Contrary to the galvanic effect, however, a change in phase occurs in the inductive effect, which can be used to determine, e.g., the geometry of a resistive target [57]. In pure azimuthal source-receiver geometry, the EM fields are largely horizontal and produce inductive effects when intersecting a boundary in the subsurface.

For a more complete description and investigation of the different mechanisms in CSEM surveys see, e.g., [44, 153].

2.2 Seismic

Seismic methods are by far the most widely used marine exploration methods. A typical seismic acquisition consists of a towed air gun transmitting elastic (or acoustic) waves through the seawater and into the subsurface. The elastic waves that have interacted with the subsurface are recorded by hydrophones, which are towed some distance behind the acquisition boat. The hydrophones are located on a single line cable, or, more commonly, on a 2D array of cables.

There are three main types of elastic waves: P-waves, S-waves, and surface waves [142]. Of the three wave types, P-waves (or primary waves) are the most important in seismic acquisition. P-waves travel by compression and rarefaction of the media, and the direction of travel is thus perpendicular to the wave fronts. Lines following the travel direction are called ray paths, and are important in first guess seismic interpretation. S-waves (or shear waves) are transverse waves, that is, the movement is perpendicular to the direction of the wave propagation (similar to EM waves). Since water does not have any shear strength, S-waves cannot be recorded by towed hydrophones. If receivers are placed on the sea floor, however, it is possible to record both S-waves and P-waves. Surface waves are waves traveling near the interface between different materials, and do not illuminate the subsurface. Hence, surface waves are mainly considered as noise, which is most severe in onshore acquisition.

The fundamental principles in seismic surveying are reflection and refraction. When an incident elastic wave meets an interface between two media with different elastic properties (primarily density), the wave will reflect or refract, or both, according to Snell's laws. Due to the complex geometry of the subsurface, the wave propagation can be quite involved. An example is the occurrence of multiples. These are waves that have reflected multiple times in the subsurface before reaching the receiver, hence, making the interpretation of such signals a difficult process.

At the receivers, the incoming seismic waves are recorded together with travel times (time from transmission to recording), and form the basis of seismic data. The seismic data are processed, before an interpreter converts them to one or more likely geological scenarios. Interpretation is a difficult process, and requires expert knowledge from different parts of geology. Most often such knowledge comes from studying outcrop analogues and well logs.

Contrary to CSEM, seismic signals do not give direct information on presence of hydrocarbons. The presence of hydrocarbons must be interpreted from typical signal reflections occurring at the interface between hydrocarbon-bearing sediments and surround sediments, so called ‘flat spots’, ‘bright spots’, and ‘dim spots’.

The resolution of seismic imaging is better than for CSEM. In CSEM, the signal propagation is diffusive in nature, and is thus limited to a large-scale description of the subsurface. Seismic waves have typically much higher frequency than CSEM waves, and have thus higher resolution. Hence, seismic methods can better resolve the structures and stratigraphy of the subsurface, while correction of the large-scale structures and detection of possible hydrocarbon formations can be done with CSEM. This means that seismic and CSEM data can give complementary information about the subsurface (see [106] for a discussion on this topic).

2.3 Magnetotelluric

A closely related method to CSEM is the magnetotelluric (MT) method. Equivalently to CSEM, the signals measured in MT are EM signals. Contrary to CSEM, natural sources are responsible for the MT signals. Typical sources are lightning activity in the ionosphere, which emits signals in the frequency range of 1 Hz–10 kHz, and disturbance in the magnetosphere, which emits signals in the frequency range below 1 Hz. The receivers in MT are equivalent to CSEM receivers, hence, it is possible to record both MT and CSEM signals in one survey.

The EM fields associated with MT are largely horizontal plane waves. Hence, the interaction with the subsurface only produces the inductive effect, and is entirely governed by the skin depth (see description in Section 2.1). Consequently, MT signals have small to no sensitivity for (thin) resistive targets such as hydrocarbon reservoirs. Thus, we can only map subsurface geological structures with a contrast in conductivity. Moreover, since the EM waves have to pass through the highly conductive sea column, only the low-frequency signals will pass through to the subsurface. Hence, MT data have low spatial resolution.

MT and CSEM data can be used together for an improved understanding of the subsurface geology. Where CSEM signals are primarily concentrated between source and receiver, the nature of MT signals leads to local and regional information about the subsurface [46]. Hence, CSEM and MT can be used together to map large-scale geological formations. Moreover, since MT has little (or no) sensitivity to hydrocarbon reservoirs, while CSEM has, an inversion process where MT data can be used to map the subsurface structures and CSEM data can be used to provide information about potential hydrocarbon reservoirs, can be made.

Chapter 3

Inverse problems

Consider a physical system described by a set of mathematical equations. To apply the mathematical equations on a specific problem, the parameters governing the equations must be determined. Direct observation of these parameters is, however, often costly, inaccurate, or impossible. We must instead rely on observations of responses from the physical system to determine the parameters.

According to [150], a study of a physical system can be divided into three steps:

1. *Parameterization*. Find a set of model parameters that fully describes the system. These parameters can either be scalars or described by a function.

2. *Forward modelling*. Find a set of mathematical equations that allows us to calculate a unique response of a physical system (to some stimulus). Typical forward models are numerical methods solving boundary-/initial-value problems of partial, or ordinary, differential equations, or integral equations.

3. *Inverse problem*. Using a set of observed responses of a physical system to determine a plausible set of model parameters, or information about these model parameters, to some stimulus. In this work, we will also refer to this as *parameter estimation*.

In the next sections, we will briefly cover subjects within the two main approaches in inverse problems: the classical approach and probabilistic (Bayesian) approach. Before we present these approaches, we will discuss inverse problems in general and outline their mathematical properties. We confine our cover of inverse problems to the discrete case and refer to, e.g., [60, 111] for theory on continuous inverse problems.

3.1 Inverse problem formulation

Let $\mathbf{m}$ denote the $N_m \times 1$ vector containing the model parameters of a system. Furthermore, let $\mathbf{d}$ be a $N_d \times 1$ vector containing the observed responses (or observed data) of a physical system, and let $\mathbf{g}(\cdot)$ denote the corresponding nonlinear forward model operator. The theoretical relationship between $\mathbf{m}$ and $\mathbf{d}$ is then given as

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

In practice, the observed data almost always contain measurement noise, due to, e.g., faulty measurements of the physical system. Hence, we can envision $\mathbf{d}$ containing noiseless observations from a 'perfect' experiment, $\mathbf{d}_{true}$, plus measurement noise, $\boldsymbol{\epsilon}_d$,

$$\begin{aligned} \mathbf{d} &= \mathbf{g}(\mathbf{m}_{true}) + \boldsymbol{\epsilon}_d, \\ &= \mathbf{d}_{true} + \boldsymbol{\epsilon}_d, \end{aligned} \tag{3.2}$$

where $\mathbf{m}_{true}$ contains the true model parameters of the physical system, and $\mathbf{d}_{true}$ contains the true observed responses assuming $\mathbf{g}(\cdot)$ is the exact forward model operator [12]. In practice, $\mathbf{g}(\cdot)$ is not exact (e.g., using numerical approximations), leading to an additional noise term $\boldsymbol{\epsilon}_m$, denoted model noise. In this case, we must replace $\boldsymbol{\epsilon}_d$ in (3.2), with $\boldsymbol{\epsilon} = \boldsymbol{\epsilon}_d + \boldsymbol{\epsilon}_m$. In practice, $\boldsymbol{\epsilon}_m$ is often neglected as it is difficult to quantify.

Solving the inverse problem is generally a very difficult task due to the fact that it is most often an *ill-posed problem*. In [77], Hadamard gave the mathematical properties of a *well-posed problem*:

Existence. A solution must exist.

Unique. A solution must be unique.

Stability. A solution must depend continuously on the data.

If any of the above properties are violated, the problem is said to be ill-posed.

The first property is almost always violated in inverse problems since no set of model parameters makes forward model responses that exactly fits the observed data, due to contamination of noise. In practice, we need to settle for solutions that fits the data within a given accuracy.

A violation of the second property frequently occurs in many inverse problems. In this case, there exist several (or possibly infinite) sets of model parameters that satisfies (3.1). In this case, one has to either decide which of the solutions is of interest, or use additional information to specify the expected nature of the solution [60].

The last property is difficult to attain. A violation of the stability property implies that small perturbations in the observed data lead to (arbitrary) large perturbations in the solution. A remedy for this is called regularization, which we come back to in Section 3.2.3.

In the following sections, we let $\tilde{\mathbf{d}} \in \mathbb{C}^{N_d/2}$ and $\tilde{\mathbf{g}}(\mathbf{m}) \in \mathbb{C}^{N_d/2}$ denote vectors of observed data and forward model outputs, respectively. Complex-valued responses are typical for physical systems governing the frequency-domain CSEM method. Instead of the complex-valued $\tilde{\mathbf{d}}$ and $\tilde{\mathbf{g}}(\mathbf{m})$, we use a real composite form (see Appendix A)

$$\mathbf{d} = \begin{bmatrix} \mathrm{Re}\{\tilde{\mathbf{d}}\} \\ \mathrm{Im}\{\tilde{\mathbf{d}}\} \end{bmatrix}, \quad \mathbf{g}(\mathbf{m}) = \begin{bmatrix} \mathrm{Re}\{\tilde{\mathbf{g}}(\mathbf{m})\} \\ \mathrm{Im}\{\tilde{\mathbf{g}}(\mathbf{m})\} \end{bmatrix}, \tag{3.3}$$

where $\mathrm{Re}\{\cdot\}$ and $\mathrm{Im}\{\cdot\}$ denote the real and imaginary part of the argument, respectively, thus $\mathbf{d}, \mathbf{g}(\mathbf{m}) \in \mathbb{R}^{N_d}$. Note that $\mathbf{m} \in \mathbb{R}^{N_m}$.

3.2 Classical approach

As we discussed in the previous section, an exact solution of (3.1) may not exist, and we must instead look for a solution that fits the data within a given accuracy. A standard approach to achieve this goal is to solve the inverse problem as an optimization problem. Define the weighted least-squares objective function as

$$J(\mathbf{m}) \;= (\mathbf{g}(\mathbf{m}) - \mathbf{d})^T \mathbf{C}_d^{-1}(\mathbf{g}(\mathbf{m}) - \mathbf{d}), \tag{3.4}$$

where T denotes matrix and vector transpose, and $\mathbf{C}_d$ is a data weighting matrix, usually a diagonal matrix containing estimates of the variance of the measurement and model noise, $\boldsymbol{\epsilon}$. $J(\mathbf{m})$, as given in (3.4), is often denoted *data misfit function*. The objective in an optimization problem is to minimize $J(\mathbf{m})$ until some termination conditions are met. Mathematically, the optimization problem can be stated as

$$\arg\min_{\mathbf{m}} \; J(\mathbf{m}). \tag{3.5}$$

Ideally, we want to find the global minimizer of (3.5), that is, we want to find a point $\mathbf{m}^*$ such that $J(\mathbf{m}^*) \leq J(\mathbf{m})$, $\forall \mathbf{m} \in \mathbb{R}^{N_m}$. However, since $\mathbf{g}(\cdot)$ is nonlinear, local minimizers of (3.5) may exist, that is, we may only be able to find a point $\mathbf{m}^*$ such that $J(\mathbf{m}^*) \leq J(\mathbf{m})$ for $\mathbf{m}$ in a neighbourhood around $\mathbf{m}^*$ [125]. Indeed, the algorithms discussed in the next sections only search for local minimizers.

3.2.1 Newton-type methods

A necessary condition for $\mathbf{m}^*$ to be a (global or local) minimizer is

$$\nabla J(\mathbf{m}) = \mathbf{0}. \tag{3.6}$$

Here, ∇ denotes the gradient with respect to $\mathbf{m}$. The most common way of solving (3.6) is by applying the Newton-Raphson algorithm. Here, an initial guess on $\mathbf{m}$, denoted $\mathbf{m}^0$, is iteratively updated by [125]

$$\mathbf{m}^{n+1} = \mathbf{m}^n + \Delta\mathbf{m}^n, \tag{3.7}$$

where $\Delta\mathbf{m}^n$ is obtained by solving the linear system

$$\mathbf{N}(\mathbf{m}^n)\Delta\mathbf{m}^n = -\nabla J(\mathbf{m}). \tag{3.8}$$

In (3.8), n is the iteration index and $\mathbf{N}(\mathbf{m})$ is the Hessian matrix

$$\mathbf{N}(\mathbf{m}) = \nabla\left((\nabla J(\mathbf{m}))^T\right). \tag{3.9}$$

The gradient and Hessian of the objective function (3.4) are given, respectively, by

$$\nabla J(\mathbf{m}) \;= 2\mathbf{S}^T \mathbf{C}_d^{-1}(\mathbf{g}(\mathbf{m}) - \mathbf{d}), \tag{3.10}$$

and

$$\mathbf{N} = 2\mathbf{S}^T \mathbf{C}_d^{-1}\mathbf{S} + \mathbf{Q}, \tag{3.11}$$

where $\mathbf{S}$ is the $N_d \times N_m$ sensitivity matrix (see Section 3.2.2) and $\mathbf{Q}$ is a matrix containing second derivatives of $\mathbf{g}(\mathbf{m})$.

It can be shown that the Newton-Raphson algorithm has quadratic convergence given that $J(\mathbf{m})$ is twice continuously differentiable in a neighbourhood around $\mathbf{m}^*$, $\mathbf{m}^0$ is sufficiently close to $\mathbf{m}^*$, and $\mathbf{N}(\mathbf{m}^*)$ is positive definite [125]. A drawback of the algorithm is the calculation of $\mathbf{N}$, which can be difficult (or even infeasible) in most applications. We will in the following briefly discuss Newton-type methods with different approximations of $\mathbf{N}$.

Gauss-Newton method

A simple approximation of $\mathbf{N}$ is to neglect $\mathbf{Q}$ in (3.11), leading to $\mathbf{N} \approx 2\mathbf{S}^T\mathbf{C}_d^{-1}\mathbf{S}$. This approximation is only valid if $2\mathbf{S}^T\mathbf{C}_d^{-1}\mathbf{S}$ is large in magnitude compare to $\mathbf{Q}$, which in many cases holds true, especially around the solution $\mathbf{m}^*$. The linear system (3.8) then becomes

$$\mathbf{S}(\mathbf{m}^n)^T\mathbf{C}_d^{-1}\mathbf{S}(\mathbf{m}^n)\Delta\mathbf{m}^n = -\mathbf{S}(\mathbf{m}^n)^T\mathbf{C}_d^{-1}(\mathbf{g}(\mathbf{m}^n) - \mathbf{d}), \tag{3.12}$$

and is denoted the Gauss-Newton (GN) method. As seen in (3.12), only the sensitivity matrix $\mathbf{S}$ is needed to solve the linear system, which saves computational time compared to the Newton-Raphson algorithm.

Often, GN is implemented with a line search algorithm. In short, a line search algorithm searches along the direction of the current iterate $\mathbf{m}^n$ for a new iterate with lower objective function value. In practical applications, an inexact line search algorithm is implemented, where a few trial searches are done and accepted when some conditions are fulfilled (often used stopping criteria are the Wolfe conditions) [125].

An interesting connection between GN and linear least-squares problems can be made by noting that (3.12) is identical in form to normal equations. That is, $\Delta\mathbf{m}^n$ is the solution of the linear least-squares problem [125]

$$\arg\min_{\Delta\mathbf{m}} \left\|\mathbf{C}_d^{-1/2}[\mathbf{S}(\mathbf{m}^n)\Delta\mathbf{m} + (\mathbf{g}(\mathbf{m}^n) - \mathbf{d})]\right\|_2^2. \tag{3.13}$$

This subproblem can be much easier to solve using, e.g., QR decomposition where $\mathbf{S}^T\mathbf{C}_d^{-1}\mathbf{S}$ does not need to be calculated explicitly [66].

In the best case, GN will have similar convergence properties as that of Netwon-Raphson. However, the convergence relies on $\mathbf{S}$ having full rank. For application of GN in CSEM inversion see, e.g., [2].

Levenberg-Marquardt method

In the Levenberg-Marquardt (LM) method, $\mathbf{N}$ is replaced by $2(\mathbf{S}^T\mathbf{C}_d^{-1}\mathbf{S}+\lambda^n\mathbf{I}_{N_m})$, where $\mathbf{I}_{N_m}$ is the $N_m \times N_m$ identity matrix, and $\lambda^n \geq 0$ is chosen at each iteration to improve the downhill search direction. The linear system (3.8) then becomes

$$\left[\mathbf{S}(\mathbf{m}^n)^T\mathbf{C}_d^{-1}\mathbf{S}(\mathbf{m}^n) + \lambda^n\mathbf{I}_{N_m}\right]\Delta\mathbf{m}^n = -\mathbf{S}(\mathbf{m}^n)^T\mathbf{C}_d^{-1}(\mathbf{g}(\mathbf{m}^n) - \mathbf{d}). \tag{3.14}$$

The obvious connection between (3.12) and (3.14) leads to LM sometimes being referred to as the damped GN method.

The value λ^n controls both the search direction and the step size. If $\lambda^n \approx 0$, then LM reduces to GN, and typically large steps are made. On the other hand, if $\mathbf{S}^T\mathbf{C}_d^{-1}\mathbf{S} + \lambda^n\mathbf{I}_{N_m} \approx \lambda^n\mathbf{I}_{N_m}$, then LM reduces to the steepest decent method,

$$\Delta\mathbf{m}^n \approx -\frac{1}{\lambda^n}\mathbf{S}(\mathbf{m}^n)^T\mathbf{C}_d^{-1}(\mathbf{g}(\mathbf{m}^n) - \mathbf{d}), \tag{3.15}$$

which ensures convergence, but is typically very slow (linear convergence). Hence, careful adjustment of λ^n must be done at each iteration to ensure good convergence properties for LM. An example of a selection procedure for λ^n can be found in [65].

A related implementation of LM is the trust region method. In short, the trust region method ensures that the length and direction of $\Delta\mathbf{m}^n$ is chosen within a trusted region around the current iterate. A popular implementation of the trust region method is given by More in [118].

The LM method has been used in CSEM inversion, see, e.g., [116]. It was also used in Paper A.

Quasi-Newton methods

The motivation in quasi-Newton (QN) methods is to replace the (usually) computationally costly $\mathbf{N}$ with a less computationally expensive matrix $\mathbf{B}$. The matrix $\mathbf{B}$ is a symmetric, positive definite matrix that, compared to $\mathbf{N}$, has an easily computable inverse, $\mathbf{B}^{-1}$. It is of course desirable to use information about the Hessian to build $\mathbf{B}$, without actually computing it. To this end, $\mathbf{B}$ is updated at each iteration by solving the secant equation

$$\mathbf{B}^{n+1}\Delta\mathbf{m}^n = \nabla J(\mathbf{m}^{n+1}) - \nabla J(\mathbf{m}^n) = \mathbf{w}^n. \tag{3.16}$$

(From Taylor's theorem the second-derivative of a function can be approximated by the difference in gradients, see, e.g., [125]). The secant equation is not enough to uniquely determine $\mathbf{B}$, hence, additional constraints are needed. In most QN methods, the following problem is solved [125]:

$$\begin{aligned} &\min_{\mathbf{B}} \|\mathbf{B} - \mathbf{B}^n\|_F, \\ \text{subject to} \quad &\mathbf{B} = \mathbf{B}^T, \quad \mathbf{B}\Delta\mathbf{m}^n = \mathbf{w}^n, \quad (\Delta\mathbf{m}^n)^T\mathbf{w}^n > 0, \end{aligned} \tag{3.17}$$

where $\|\cdot\|_F$ is a weighted Frobenius norm. Solving (3.17) leads to the most used QN methods: Davidson-Fletcher-Powell (DFP) method and Broyden-Fletcher-Goldfarb-Shanno (BFGS) method.

The advantage of QN methods is that only the gradient of $J(\mathbf{m})$ is needed to update $\mathbf{B}$. The convergence rate of QN methods is superlinear, which is lower than GN. However, the cost per iteration in QN is lower than GN since $\mathbf{B}^{-1}$ is easier to calculated than $\mathbf{N}^{-1}$. In the case of BFGS, $\mathbf{B}^{-1}$ is found directly by solving (3.17) where $\mathbf{B}$ is replaced with $\mathbf{B}^{-1}$.

The main disadvantage of QN methods is that $\mathbf{B}^n$ must be stored for next iteration (in order to update $\mathbf{B}^{n+1}$), which for large-scale problems can be memory consuming. To alleviate this problem, a limited memory version of BFGS was introduced where updates of $\mathbf{B}^{n+1}$ are done using vectors that implicitly stores information about $\mathbf{B}^n$ [125].

QN has been used for solving CSEM inversion problems, see, e.g., [131].

3.2.2 Sensitivity calculation

In Newton-Rapshon and the Newton-type methods, the sensitivity matrix, **S**, is required in the gradient and Hessian of $J(\mathbf{m})$. The sensitivity matrix is a $N_d \times N_m$ matrix given by

$$\mathbf{S} = \frac{\partial g_i}{\partial m_j}, \quad i = 1,\ldots,N_d, \quad j = 1,\ldots,N_m, \tag{3.18}$$

hence it provides information about how changes in the model parameters affect the output of the forward model.

There are generally four ways of calculating **S**: perturbation method, automatic differentiation, direct method, and adjoint method [109]. In the perturbation method, the derivatives in (3.18) are approximated by some finite difference approach. This method is not widely used in practice, due to its inefficiency and difficulties associated with selection of perturbation length. Automatic differentiation exploits the fact that any function is written in a computer program using elementary functions (e.g., sin, exp, etc.) and arithmetic operations (+, −, etc.). By using basic differentiation rules, the derivatives can be found automatically by the computer program to working precision (see, e.g., [121]). The two last methods are analytical methods, which we will briefly discuss in the next two sections.

Direct method

In the direct method, $\mathbf{g}(\mathbf{m})$ is differentiated directly with respect to $\mathbf{m}$. In many applications, $\mathbf{g}(\mathbf{m})$ is not readily available as an explicit expression; typically, forward model outputs are given as a numerical solution of a differential, or integral, equation. In this case, let $\mathbf{f}(\mathbf{g}(\mathbf{m}),\mathbf{m}) = \mathbf{0}$ denote the system of equations that are solved in order to get the numerical solution $\mathbf{g}(\mathbf{m})$. Total differentiation of $\mathbf{f}$ with respect to m_j gives [127]

$$\frac{\mathrm{d}\mathbf{f}}{\mathrm{d}m_j} = \frac{\partial \mathbf{f}}{\partial \mathbf{g}}\frac{\partial \mathbf{g}}{\partial m_j} + \frac{\partial \mathbf{f}}{\partial m_j} = \mathbf{0}, \quad j = 1,\ldots,N_m. \tag{3.19}$$

The first expression, $\partial\mathbf{f}/\partial\mathbf{g}$, is often simple to calculate. The numerical solutions of differential equations are often given on the form $\mathbf{f}(\mathbf{g}(\mathbf{m}),\mathbf{m}) = \mathbf{A}\mathbf{g} - \mathbf{b}$, where $\mathbf{A}$ is the coefficient matrix resulting from discretization of the differential equation and $\mathbf{b}$ is a vector containing the source term; hence, $\partial\mathbf{f}/\partial\mathbf{g} = \mathbf{A}$. In the case where $\mathbf{f}(\mathbf{g}(\mathbf{m}),\mathbf{m})$ is a nonlinear function, $\mathbf{f}(\mathbf{g}(\mathbf{m}),\mathbf{m}) = \mathbf{0}$ must be solved iteratively (using, e.g., a Netwon-type method, see Section 3.2.1) where $\mathbf{A} = \partial\mathbf{f}/\partial\mathbf{g}$ is calculated at each iteration. The second expression, $\partial\mathbf{f}/\partial m_j$, can be calculated analytically since the functional relationship between $\mathbf{f}$ and $\mathbf{m}$ will be known.

Inserting $\partial\mathbf{f}/\partial\mathbf{g} = \mathbf{A}$ into (3.19) and rearranging leads to

$$\mathbf{A}\frac{\partial \mathbf{g}}{\partial m_j} = -\frac{\partial \mathbf{f}}{\partial m_j}, \quad j = 1,\ldots,N_m. \tag{3.20}$$

This is a linear system that can be solved in a similar manner as $\mathbf{f}(\mathbf{g}(\mathbf{m}),\mathbf{m}) = \mathbf{0}$. From (3.20) we see that the right-hand side gives rise to a new linear system for each unknown model parameter, m_j; hence, N_m linear systems must be solved to get the entries in $\mathbf{S}$.

Adjoint method

To present the adjoint method, we follow the description in [127] and define a function $L = h(\mathbf{g}(\mathbf{m}), \mathbf{m})$. From the previous section we know that $\mathbf{f}(\mathbf{g}(\mathbf{m}), \mathbf{m}) = \mathbf{0}$, hence, we can define the Lagrangian

$$L = h(\mathbf{g}(\mathbf{m}), \mathbf{m}) + \boldsymbol{\mu}^T \mathbf{f}(\mathbf{g}(\mathbf{m}), \mathbf{m}), \tag{3.21}$$

where $\boldsymbol{\mu}$ is a vector containing the Lagrange multipliers. Total differentiation of L with respect to m_j leads to

$$\begin{aligned} \frac{\mathrm{d}L}{\mathrm{d}m_j} &= \frac{\partial h}{\partial \mathbf{g}} \frac{\partial \mathbf{g}}{\partial m_j} + \frac{\partial h}{\partial m_j} + \boldsymbol{\mu}^T \left[\frac{\partial \mathbf{f}}{\partial \mathbf{g}} \frac{\partial \mathbf{g}}{\partial m_j} + \frac{\partial \mathbf{f}}{\partial m_j} \right], \\ &= \frac{\partial h}{\partial m_j} + \boldsymbol{\mu}^T \frac{\partial \mathbf{f}}{\partial m_j} + \left[\frac{\partial h}{\partial \mathbf{g}} + \boldsymbol{\mu}^T \frac{\partial \mathbf{f}}{\partial \mathbf{g}} \right] \frac{\partial \mathbf{g}}{\partial m_j}. \end{aligned} \tag{3.22}$$

$\boldsymbol{\mu}$ can be freely chosen, hence, we require it to satisfy

$$\boldsymbol{\mu}^T \frac{\partial \mathbf{f}}{\partial \mathbf{g}} = -\frac{\partial h}{\partial \mathbf{g}}, \tag{3.23}$$

or

$$\left(\frac{\partial \mathbf{f}}{\partial \mathbf{g}} \right)^T \boldsymbol{\mu} = -\left(\frac{\partial h}{\partial \mathbf{g}} \right)^T. \tag{3.24}$$

This linear system is referred to as the adjoint system.

Inserting (3.23) into (3.22), we get

$$\frac{\mathrm{d}L}{\mathrm{d}m_j} = \frac{\partial h}{\partial m_j} + \boldsymbol{\mu}^T \frac{\partial \mathbf{f}}{\partial m_j}. \tag{3.25}$$

Hence, when $\boldsymbol{\mu}$ is found from (3.24), it can be inserted into (3.25) to find the sensitivity $\mathrm{d}L/\mathrm{d}m_j$.

If we let $L = h(\mathbf{g}(\mathbf{m}), \mathbf{m}) = g_i$ and recall from the previous section that $\partial \mathbf{f}/\partial \mathbf{g} = \mathbf{A}$, then (3.24) and (3.25) give

$$\mathbf{A}^T \boldsymbol{\mu} = -\left(\frac{\partial g_i}{\partial \mathbf{g}} \right)^T, \tag{3.26}$$

and

$$\frac{\mathrm{d}g_i}{\mathrm{d}m_j} = \boldsymbol{\mu}^T \frac{\partial \mathbf{f}}{\partial m_j}, \tag{3.27}$$

respectively, where $i = 1, \ldots, N_d$ and $j = 1, \ldots, N_m$. To get the sensitivity matrix $\mathbf{S}$ we need to solve (3.26) N_d times and insert into (3.27). Hence, the adjoint method is well suited for problems where $N_d < N_m$, whereas the direct method is well suited for problems where $N_m < N_d$.

3.2.3 Tikhonov regularization

To alleviate the instability issue in an ill-posed inverse problem, additional features of the solution are imposed by using regularization techniques. The additional features are *a priori* assumptions on the nature of the solution. Hence, regularization techniques will always bias the solution. A compromise between accuracy and stability is thus needed when choosing the regularization technique.

Perhaps the most widely applied regularization technique is Tikhonov regularization [152]. In the least-squares framework, a Tikhonov regularization is given by

$$J(\mathbf{m}) \ = (\mathbf{g}(\mathbf{m}) - \mathbf{d})^T \mathbf{C}_d^{-1} (\mathbf{g}(\mathbf{m}) - \mathbf{d}) + \alpha \|\mathbf{L}\mathbf{m}\|_2^2. \tag{3.28}$$

Here, $\mathbf{L}$ is called the roughening matrix, and $\alpha > 0$ is the regularization parameter. In an optimization, a solution is sought where both terms in (3.28) are minimized. In addition to the data misfit term (first term), we have a penalization term (second term). If we let $\mathbf{L} = \mathbf{I}_{N_m}$, then $\alpha\|\mathbf{m}\|_2^2$ will penalize solutions with large norms. This is often denoted zero-order Tikhonov regularization (or ridge regression in statistics). In higher-order Tikhonov regularization, $\mathbf{L}$ is given as a numerical approximation of some differential operator. This will penalize solutions that are rough in some derivative sense. Note that $\|\cdot\|_2$ in the above description is the Euclidean norm. If we instead use the l_1-norm and let $\mathbf{L} = \mathbf{I}_{N_m}$, then (3.28) reduces to the LASSO (least absolute shrinkage and selection operator) method [151].

The regularization parameter α can be difficult to adjust. Essentially α controls the degree of which the regularization should be emphasized. Some strategies for selecting α have been proposed, e.g., the L-curve strategy [96]. In practice, however, α is usually adjusted during the iterations by applying some kind of ‘cooling’ scheme, see, e.g., [124]. The basic idea with a ‘cooling’ scheme is to have a large α in the beginning, leading to (3.28) being almost quadratic, and subsequently reduce α in later iterations to put more emphasis on the data misfit term.

3.3 Bayesian approach

A different approach from the classical one is to view $\mathbf{m}$ and $\mathbf{d}$ as random variables associated with probability distributions. In this case, we are not only interested in identifying one particular solution of the inverse problem but we are also interested in quantifying the uncertainty of the solution. Towards this end, let $\mathbf{m}$ be associated with the PDF $f(\mathbf{m})$ and let $\mathbf{d}$ be associated with the likelihood function $f(\mathbf{d}|\mathbf{m})$. The conditional PDF of $\mathbf{m}$ given $\mathbf{d}$ is given by *Bayes' rule*:

$$f(\mathbf{m}|\mathbf{d}) = \frac{f(\mathbf{d}|\mathbf{m})f(\mathbf{m})}{f(\mathbf{d})}. \tag{3.29}$$

The PDF $f(\mathbf{d})$ is a normalizing constant ensuring that $f(\mathbf{m}|\mathbf{d})$ integrates to 1. In practice, exact knowledge about $f(\mathbf{d})$ is not needed, and (3.29) can be expressed as

$$f(\mathbf{m}|\mathbf{d}) \propto f(\mathbf{d}|\mathbf{m})f(\mathbf{m}). \tag{3.30}$$

In (3.30), $f(\mathbf{m})$ is the knowledge and uncertainty we have on the model parameters before conditioning to data, and is denoted *prior* PDF. Consequently, $f(\mathbf{m}|\mathbf{d})$ is denoted *posterior* PDF, and it provides the complete solution of the inverse problem. The likelihood function $f(\mathbf{d}|\mathbf{m})$ relates $\mathbf{m}$ with $\mathbf{d}$ through (3.1). (Note that the difference between $f(\mathbf{d}|\mathbf{m})$ being a likelihood function or conditional PDF is whether $\mathbf{d}$ is a fixed or a random variable, leading to $\mathbf{m}$ being a random or a fixed variable, respectively. Since we want to use observed data that are fixed, e.g., from an experiment, $f(\mathbf{d}|\mathbf{m})$ is the likelihood function.)

If we assume that measurement errors and model errors, $\boldsymbol{\epsilon}$, are Gaussian with mean zero and covariance matrix $\mathbf{C}_d$, then $f(\mathbf{d}|\mathbf{m})$ is written as

$$\begin{aligned} f(\mathbf{d}|\mathbf{m}) &= f(\boldsymbol{\epsilon} = \mathbf{g}(\mathbf{m}) - \mathbf{d}), \\ &= c \exp\left(-(\mathbf{g}(\mathbf{m}) - \mathbf{d})^T \mathbf{C}_d^{-1} (\mathbf{g}(\mathbf{m}) - \mathbf{d})\right), \end{aligned} \tag{3.31}$$

where c is a normalizing constant. Similarly, if we assume that the prior PDF is Gaussian with mean $\mathbf{m}_{prior}$ and covariance matrix $\mathbf{C}_m$, then $f(\mathbf{m})$ is given by

$$f(\mathbf{m}) = c \cdot \exp(-(\mathbf{m} - \mathbf{m}_{prior})^T \mathbf{C}_m^{-1} (\mathbf{m} - \mathbf{m}_{prior})). \tag{3.32}$$

Inserting (3.31) and (3.32) into (3.30) yields

$$f(\mathbf{m}|\mathbf{d}) = c \exp(-J(\mathbf{m})), \tag{3.33}$$

where

$$J(\mathbf{m}) = (\mathbf{g}(\mathbf{m}) - \mathbf{d})^T \mathbf{C}_d^{-1} (\mathbf{g}(\mathbf{m}) - \mathbf{d}) + (\mathbf{m} - \mathbf{m}_{prior})^T \mathbf{C}_m^{-1} (\mathbf{m} - \mathbf{m}_{prior}). \tag{3.34}$$

Note that $f(\mathbf{m}|\mathbf{d})$ is not a Gaussian PDF since $\mathbf{g}(\cdot)$ is a nonlinear forward model operator. Hence, a complete characterization of $f(\mathbf{m}|\mathbf{d})$ can be very difficult or even impossible. (Note that the above Gaussian PDFs can be written as generalized complex Gaussian PDFs using a complex augmented form of $\tilde{\mathbf{d}}$ and $\tilde{\mathbf{g}}(\mathbf{m})$, see Appendix A. However, the generalized complex Gaussian PDFs are no more general than the Gaussian PDFs based on the real composite form given above.)

Traditionally, a single 'best' estimate can be obtained from $f(\mathbf{m}|\mathbf{d})$ by finding its maximum a posteriori solution. Mathematically, this is stated as

$$\begin{aligned} \mathbf{m}_{MAP} &= \arg\max_{\mathbf{m}} \; f(\mathbf{m}|\mathbf{d}), \\ &= \arg\min_{\mathbf{m}} \; J(\mathbf{m}), \end{aligned} \tag{3.35}$$

where the second equality follows from (3.33). Note that this is an optimization problem that can be solved using the methods described in Section 3.2. Moreover, comparing $J(\mathbf{m})$ from (3.34) and (3.28), we see that the second term in (3.34) is a type of zero-order Tikhonov regularization term; it will penalize solutions that are far away from $\mathbf{m}_{prior}$ (in some weighted norm sense). In Paper A, we used a similar regularization term as in (3.34), which was derived by assuming a Gaussian prior PDF in a high-dimensional feature space $\mathcal{Y}$ (confer Section 6.3).

In the following sections, we will discuss methods for characterizing the whole posterior PDF, not only providing a 'single' best estimate. First, we present the widely used, and important, Markov Chain Monte Carlo (MCMC) methods. Subsequently, we will discuss alternative methods to MCMC, which are based on the sequential Bayesian framework.

3.3.1 Markov chain Monte Carlo

Since a complete characterization of the posterior PDF is generally infeasible when the forward model operator is nonlinear, assessment of the posterior PDF can only be done by sampling. A popular choice for sampling from the posterior PDF is MCMC algorithms, due to their simplistic implementation. The basic idea of MCMC algorithms is to construct Markov chains that converge to the posterior PDF within a finite number of steps. In most MCMC algorithms, this is an iterative procedure:

1. $j = 1 \rightarrow$ generate initial sample $\mathbf{m}^1$.
2. Generate a sample $\mathbf{m}^\star$ from a proposal PDF $q(\mathbf{m}|\mathbf{m}^j)$.
3. Generate a random number from the uniform distribution $u \sim \mathcal{U}[0, 1]$, and calculate the acceptance probability $\beta(\mathbf{m}^\star, \mathbf{m}^j)$:
$$\beta(\mathbf{m}^\star, \mathbf{m}^j) = \min\left[1, \frac{f(\mathbf{m}^\star|\mathbf{d})q(\mathbf{m}^j|\mathbf{m}^\star)}{f(\mathbf{m}^j|\mathbf{d})q(\mathbf{m}^\star|\mathbf{m}^j)}\right].$$
4. If $\beta(\mathbf{m}^\star, \mathbf{m}^j) > u \rightarrow \mathbf{m}^{j+1} = \mathbf{m}^\star$; else $\mathbf{m}^{j+1} = \mathbf{m}^j$.
5. Repeat steps 2–4 until a predetermined number of iteration steps is reached.

The proposal PDF, $q(\mathbf{m}^i|\mathbf{m}^j)$, is the probability of proposing a transition from $\mathbf{m}^i$ to $\mathbf{m}^j$. It can be arbitrarily chosen as long as it is possible to propose a transition to any $\mathbf{m}$ within a finite number of steps [127]. The acceptance probability, $\beta(\mathbf{m}^i, \mathbf{m}^j)$, given in step 3, is the widely used Metropolis-Hastings algorithm [79, 113]. Other algorithms exist, e.g., Gibbs sampling and slice sampling. (Note that to evaluate $f(\mathbf{m}|\mathbf{d})$ in step 3 we insert (3.30), or if the likelihood and prior is given from (3.31) and (3.32), respectively, then we insert (3.33).)

The convergence of MCMC algorithms is usually very slow. Depending on the proposal PDF, the number of iterations can be $O(10^4) \sim O(10^5)$, or even higher. Since we need to run the forward model as many times as we have iterations, the computational costs can be prohibitively costly for complex models .

For CSEM application of MCMC algorithms see, e.g., [39, 132]. These applications involved a fast computing 1D numerical model, which enabled the use of a MCMC algorithm.

3.3.2 Sequential Bayesian formulations

In the following sections, it is advantageous to introduce some notation. Let $\mathbf{d}_k \in \mathbb{R}^{N_{d_k}}$ denote a subset of the observed data, where $k = 1, \ldots, N_s$, with N_s denoting the total number of subsets, that is, $N_d = \sum_{k=1}^{N_s} N_{d_k}$. The corresponding forward model output is denoted $\mathbf{g}_k(\mathbf{m}_{k-1}) \in \mathbb{R}^{N_{d_k}}$, and $\mathbf{m}_k \in \mathbb{R}^{N_m}$. We introduce a subscript notation to denote a sequence, e.g., $\mathbf{d}_{k:1} = \{\mathbf{d}_k, \ldots, \mathbf{d}_1\}$. For simplicity of notation, we augment the forward model response and model parameter vector in a new vector

$$\boldsymbol{\psi}_k = \begin{bmatrix} \mathbf{g}_k(\mathbf{m}_{k-1}) \\ \mathbf{m}_k \end{bmatrix}, \tag{3.36}$$

often denoted the joint-state vector. Note that $\boldsymbol{\psi}_k \in \mathbb{R}^{N_{\psi_k}}$, where $N_{\psi_k} = N_{d_k} + N_m$.

Following [64], we use the above notation to write the posterior PDF at step k as

$$f(\boldsymbol{\psi}_{k:0}|\mathbf{d}_{k:1}) \propto f(\mathbf{d}_{k:1}|\boldsymbol{\psi}_{k:0}) f(\boldsymbol{\psi}_{k:0}). \tag{3.37}$$

If we assume that the evolution of $\boldsymbol{\psi}_k$ from one sequential step to the next is a first-order Markov process; that data vectors at different sequential steps are independent from each other; and that the data at a particular sequential step only depend on the joint-state at this step; then (3.37) can be written as

$$f(\boldsymbol{\psi}_{k:0}|\mathbf{d}_{k:1}) \propto f(\mathbf{d}_k|\boldsymbol{\psi}_k) f(\boldsymbol{\psi}_{k:0}|\mathbf{d}_{k-1:1}). \tag{3.38}$$

If we note that $f(\boldsymbol{\psi}_{k:0}|\mathbf{d}_{k-1:1})$ is just the posterior PDF from step $k-1$ evolved forward to step k, then (3.38) is just a sequential version of (3.37). Observe that at step k, the information from $\mathbf{d}_k$ is used to update all joint-state vectors from steps $0, \ldots, k$. In the literature, this is called a *smoother*; hence, (3.37) is denoted a *general smoother* and (3.38) is denoted a *sequential smoother*.

From (3.38) it is possible to derive the *general filter* by integrating over the solutions $\boldsymbol{\psi}_{k-1:0}$,

$$f(\boldsymbol{\psi}_k|\mathbf{d}_{k:1}) \propto f(\mathbf{d}_k|\boldsymbol{\psi}_k) f(\boldsymbol{\psi}_k|\mathbf{d}_{k-1:1}). \tag{3.39}$$

Similarly to (3.38), we note that $f(\boldsymbol{\psi}_k|\mathbf{d}_{k-1:1})$ is just the posterior PDF from step $k-1$ evolved forward to step k. Contrary to (3.38), $\mathbf{d}_k$ only updates $\boldsymbol{\psi}_k$. Note that at step k, (3.38) and (3.39) will provide the same estimate, but the general filter estimate at step k is suboptimal at later sequential steps since observed data for steps $k+1, \ldots, N_s$ will not be used to update $\boldsymbol{\psi}_k$.

3.3.3 Kalman filter

The posterior PDF in the general filter problem (3.39) has a closed-form solution when the forward model is linear, and the likelihood function and prior PDF are Gaussian. This was discovered by Kalman in his now famous 1960 paper [89], and is thus named Kalman filter (KF).

We follow the traditional notation for filter problems and consider the current state of the system given by

$$\boldsymbol{\psi}_k = \mathbf{F}_k \boldsymbol{\psi}_{k-1} + \boldsymbol{\tau}_k = \begin{bmatrix} \mathbf{0} & \mathbf{P}_k \\ \mathbf{0} & \mathbf{I}_{N_m} \end{bmatrix} \begin{bmatrix} \mathbf{P}_{k-1}\mathbf{m}_{k-2} \\ \mathbf{m}_{k-1} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_k^m \\ \mathbf{0} \end{bmatrix}, \tag{3.40}$$

with $\mathbf{P}$ denoting the linear forward model operator, i.e., $\mathbf{g}(\mathbf{m}) = \mathbf{P}\mathbf{m}$ in (3.36). $\boldsymbol{\epsilon}_k^m \in \mathbb{R}^{N_{d_k}}$ is a zero-mean Gaussian model error, i.e., $\boldsymbol{\epsilon}_k^m \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{m_k})$. Note that the model parameters are not changed from step $k-1$ to step k in (3.40); they are only used to produce forward model output. Furthermore, the relationship between $\mathbf{d}_k$ and $\boldsymbol{\psi}_k$ is given by

$$\mathbf{d}_k = \mathbf{H}_k \boldsymbol{\psi}_k + \boldsymbol{\epsilon}_k^d, \tag{3.41}$$

where $\mathbf{H}_k = [\mathbf{I}_{N_{d_k}}, \mathbf{0}]$, and $\boldsymbol{\epsilon}_k^d \in \mathbb{R}^{N_{d_k}}$ is a zero-mean Gaussian measurement error, i.e., $\boldsymbol{\epsilon}_k^d \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{d_k})$. It is assumed that $\boldsymbol{\psi}_0$, $\boldsymbol{\epsilon}_k^m$ and $\boldsymbol{\epsilon}_k^d$ are mutually independent.

Since all the involved PDFs in (3.39) are Gaussian and the forward model operator is linear, the posterior PDF at step $k-1$ will also be Gaussian

$$f(\boldsymbol{\psi}_{k-1}|\mathbf{d}_{k-1:1}) \sim \mathcal{N}(\boldsymbol{\psi}^a_{k-1}, \mathbf{C}^a_{\psi_{k-1}}), \tag{3.42}$$

hence, completely described by its mean $\boldsymbol{\psi}^a_{k-1}$ and covariance matrix $\mathbf{C}^a_{\psi_{k-1}}$. The prior PDF at the next step, k, is given by

$$f(\boldsymbol{\psi}_k|\mathbf{d}_{k-1:1}) \sim \mathcal{N}(\boldsymbol{\psi}^f_k, \mathbf{C}^f_{\psi_k}). \tag{3.43}$$

From (3.40) and the fact that $\boldsymbol{\epsilon}^m_k$ has zero mean, $\boldsymbol{\psi}^f_k$ and $\mathbf{C}^f_{\psi_k}$ are given by [85]

$$\boldsymbol{\psi}^f_k = \mathbf{F}_k\boldsymbol{\psi}^a_{k-1}, \tag{3.44}$$

$$\mathbf{C}^f_{\psi_k} = \mathbf{F}_k\mathbf{C}^a_{\psi_{k-1}}\mathbf{F}^T_k + \mathbf{C}_{\tau_k}, \tag{3.45}$$

where

$$\mathbf{C}_{\tau_k} = \begin{bmatrix} \mathbf{C}_{m_k} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}. \tag{3.46}$$

The equations (3.44) and (3.45) describes the first stage in KF, the *forecast step*; hence the superscript ‘f’.

The second stage of KF is the *analysis step* (hence the superscript ‘a’) where the mean and covariance matrix of the posterior PDF at step k, $\boldsymbol{\psi}^a_k$ and $\mathbf{C}^a_{\psi_k}$, respectively, is calculated. By recognizing that $f(\boldsymbol{\psi}_k|\mathbf{d}_{k:1})$ can be derived from the joint distribution $f(\boldsymbol{\psi}_k, \mathbf{d}_k|\mathbf{d}_{k-1:1})$, we can outline a simple derivation of the analysis step; see, e.g., [110] for a full derivation. From multivariate statistics, a joint Gaussian distribution,

$$\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix}\right), \tag{3.47}$$

yields a conditional PDF $f(\mathbf{x}_1|\mathbf{x}_2)$ that is also Gaussian with mean and covariance matrix, respectively, given by

$$\boldsymbol{\mu}_{1|2} = \boldsymbol{\mu}_1 + \mathbf{C}_{12}\mathbf{C}^{-1}_{22}(\mathbf{x}_2 - \boldsymbol{\mu}_2), \tag{3.48}$$

$$\mathbf{C}_{1|2} = \mathbf{C}_{11} - \mathbf{C}_{12}\mathbf{C}^{-1}_{22}\mathbf{C}_{21}. \tag{3.49}$$

We first note that $\mathbf{x}_1 = \boldsymbol{\psi}_k$, and $\boldsymbol{\mu}_1 = \boldsymbol{\psi}^f_k$ and $\mathbf{C}_{11} = \mathbf{C}^f_{\psi_k}$ from (3.44) and (3.45), respectively. Secondly, $\mathbf{x}_2 = \mathbf{d}_k$, $\boldsymbol{\mu}_2 = \mathbf{H}_k\boldsymbol{\psi}^f_k$ and $\mathbf{C}_{22} = \mathbf{H}_k\mathbf{C}^f_{\psi_k}\mathbf{H}^T_k + \mathbf{C}_{d_k}$. Lastly, the cross-covariance matrices are given as $\mathbf{C}_{12} = \mathbf{C}^T_{21} = \mathbf{C}^f_{\psi_k}\mathbf{H}^T_k$. Inserted into (3.48) and (3.49) yields

$$\boldsymbol{\psi}^a_k = \boldsymbol{\psi}^f_k + \mathbf{K}_k(\mathbf{d}_k - \mathbf{H}_k\boldsymbol{\psi}^f_k), \tag{3.50}$$

$$\mathbf{C}^a_{\psi_k} = (\mathbf{I}_{N_{\psi_k}} - \mathbf{K}_k\mathbf{H}_k)\mathbf{C}^f_{\psi_k}, \tag{3.51}$$

where the $N_{\psi_k} \times N_{d_k}$ matrix $\mathbf{K}_k$ is the *Kalman gain*

$$\mathbf{K}_k = \mathbf{C}^f_{\psi_k}\mathbf{H}^T_k(\mathbf{H}_k\mathbf{C}^f_{\psi_k}\mathbf{H}^T_k + \mathbf{C}_{d_k})^{-1}. \tag{3.52}$$

In summary, KF is a two-stage recursive method: first, (3.44) – (3.45) are calculated giving the forecast distribution; and, secondly, (3.50) – (3.52) are utilized to correct the forecast distribution by conditioning to observed data. It can be shown that KF is the optimal (minimum variance) filter in the case of Gaussian PDFs and linear forward model operator [85].

Discussion

In the above derivation of the KF equations, we considered the joint-state vector given in (3.36). Recall from (3.3) that $\mathbf{g}(\mathbf{m})$ and $\mathbf{d}$ are augmented vectors containing the real and imaginary part of $\tilde{\mathbf{g}}(\mathbf{m})$ and $\tilde{\mathbf{d}}$ (i.e., the real composite form). Thus, if we define $\tilde{\boldsymbol{\psi}} = [\tilde{\mathbf{g}}(\mathbf{m})^T, \mathbf{m}^T]^T$, then the equivalent expression to (3.36) is (with some rearrangement) $\boldsymbol{\psi} = [\mathrm{Re}\{\tilde{\boldsymbol{\psi}}\}^T, \mathrm{Im}\{\tilde{\boldsymbol{\psi}}\}^T]^T$. From Appendix A, we know that $\tilde{\boldsymbol{\psi}}$ can be expressed in a complex augmented form: $\underline{\boldsymbol{\psi}} = [\tilde{\boldsymbol{\psi}}^T, \tilde{\boldsymbol{\psi}}^H]^T$, where H is the Hermitian (conjugate transpose). Moreover, this is equivalent to $\boldsymbol{\psi}$ since we can use the real-to-complex transform (confer (A.5)) to express $\underline{\boldsymbol{\psi}}$ in terms of $\boldsymbol{\psi}$, and vice versa.

Using the complex augmented form of the state vector, $\underline{\boldsymbol{\psi}}$, Dini *et al.* [53] derived an alternative (and equally valid) form of the KF equations given above, and denoted the method augmented complex Kalman filter (ACKF). More importantly, they showed that KF methods that only considers $\tilde{\boldsymbol{\psi}}$ instead of $\underline{\boldsymbol{\psi}}$ is a special case of ACKF with restrictions on the measurement and model error (confer Appendix A for a comparison of the covariance matrix in a Gaussian distribution for complex augmented vectors and ‘standard’ complex vectors).

3.3.4 Ensemble Kalman filter

When $\mathbf{g}(\cdot)$ is nonlinear, $f(\boldsymbol{\psi}_k|\mathbf{d}_{k:1})$ will become non-Gaussian, and thus KF will not provide a closed-form solution of (3.39). As we noted above, MCMC provides an accurate characterization of $f(\boldsymbol{\psi}_k|\mathbf{d}_{k:1})$ but at an extremely high computational cost. To reduce the computational cost, an approximate sampling method must be used. An approximate sampling method that has received a lot of attention in recent years is the ensemble Kalman filter (EnKF), first introduced by Evensen [61]. The main idea in EnKF is to use a Monte Carlo, or *ensemble*, representation of the PDFs in (3.39), and update the ensemble with the KF analysis equations. In the following, we only give an overview of the method and discuss some of its features. For a more thorough discussion on EnKF see, e.g., [1, 63].

We consider the nonlinear state-space problem, where the current state of the system is given by

$$\boldsymbol{\psi}_k = \mathbf{F}_k(\boldsymbol{\psi}_{k-1}) + \boldsymbol{\tau}_k = \begin{bmatrix} \mathbf{g}_k(\mathbf{m}_{k-1}) \\ \mathbf{m}_{k-1} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_k^m \\ \mathbf{0} \end{bmatrix}. \tag{3.53}$$

Similarly to (3.41), the relationship between $\mathbf{d}_k$ and $\boldsymbol{\psi}_k$ is assumed to be linear,

$$\mathbf{d}_k = \mathbf{H}_k \boldsymbol{\psi}_k + \boldsymbol{\epsilon}_k^d. \tag{3.54}$$

Note that if the relationship between $\mathbf{d}_k$ and $\boldsymbol{\psi}_k$ is nonlinear, say $\mathbf{h}_k(\boldsymbol{\psi}_k)$, then a linear relationship can be obtained by defining a new augmented vector $\widehat{\boldsymbol{\psi}}_k = [\boldsymbol{\psi}_k^T, \mathbf{h}_k(\boldsymbol{\psi}_k)^T]^T$ and a new associated matrix $\widehat{\mathbf{H}}_k$ that picks out $\mathbf{g}_k(\mathbf{m}_{k-1})$ from $\widehat{\boldsymbol{\psi}}_k$. Note that $\boldsymbol{\epsilon}_k^m \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{m_k})$ and $\boldsymbol{\epsilon}_k^d \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{d_k})$ as for KF.

To describe the EnKF algorithm, we need to introduce some notation. Let $\mathbf{M}_k = [\mathbf{m}_k^1, \ldots, \mathbf{m}_k^{N_e}] \in \mathbb{R}^{N_m \times N_e}$ and $\mathbf{G}_k(\mathbf{M}_{k-1}) = [\mathbf{g}_k(\mathbf{m}_{k-1}^1), \ldots, \mathbf{g}_k(\mathbf{m}_{k-1}^{N_e})] \in \mathbb{R}^{N_{d_k} \times N_e}$ denote the ensemble of model parameters and forward model outputs, respectively, where

N_e denotes the number of ensemble members. With these definitions, an ensemble matrix $\mathbf{\Psi}_k \in \mathbb{R}^{N_{\psi_k} \times N_e}$ containing $\mathbf{M}_k$ and $\mathbf{G}_k(\mathbf{M}_{k-1})$ is given on a similar form as (3.36)

$$\mathbf{\Psi}_k = \begin{bmatrix} \mathbf{G}_k(\mathbf{M}_{k-1}) \\ \mathbf{M}_k \end{bmatrix}. \tag{3.55}$$

The sample mean matrix of the ensemble members is given by

$$\overline{\mathbf{\Psi}}_k = \mathbf{\Psi}_k \mathbf{1}_{N_e}, \tag{3.56}$$

where $\mathbf{1}_{N_e}$ is a $N_e \times N_e$ matrix with all entries equal $1/N_e$. With the definition of $\overline{\mathbf{\Psi}}_k$, the sample covariance matrix for $\mathbf{\Psi}_k$ is given by

$$\mathbf{C}^e_{\psi_k} = \frac{1}{N_e - 1}(\mathbf{\Psi}_k - \overline{\mathbf{\Psi}}_k)(\mathbf{\Psi}_k - \overline{\mathbf{\Psi}}_k)^T. \tag{3.57}$$

To ensure that $\mathbf{C}^e_{\psi_k}$ is updated with correct statistics, an ensemble of observed data is also needed [34]. To this end, let $\mathbf{D}_k = [\mathbf{d}^1_k, \ldots, \mathbf{d}^{N_e}_k] \in \mathbb{R}^{N_{d_k} \times N_e}$ denote the ensemble of observed data, where each entry is a realization of a Gaussian measurement distribution,

$$\mathbf{d}^j_k \sim \mathcal{N}(\mathbf{d}^{true}_k, \mathbf{C}_{d_k}), \quad j = 1, \ldots, N_e, \tag{3.58}$$

where $\mathbf{d}^{true}_k$ contains the true observed data. The relationship between $\mathbf{\Psi}_k$ and $\mathbf{D}_k$ is given in a similar manner as (3.54),

$$\mathbf{D}_k = \mathbf{H}_k \mathbf{\Psi}_k + \mathbf{E}^d_k, \tag{3.59}$$

where $\mathbf{E}^d_k = [(\boldsymbol{\epsilon}^d_k)^1, \ldots, (\boldsymbol{\epsilon}^d_k)^{N_e}]$, with $(\boldsymbol{\epsilon}^d_k)^j \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{d_k})$ for $j = 1, \ldots, N_e$.

EnKF follows the same forecast-analysis workflow as KF. In the forecast step, we are interested in calculating the forward propagation of the posterior PDF at the previous step $k-1$, $f(\boldsymbol{\psi}_{k-1}|\mathbf{d}_{k-1:1})$, to the next sequential step k; that is, we want to calculate the prior PDF $f(\boldsymbol{\psi}_k|\mathbf{d}_{k-1:1})$ in (3.39). In general, the evolution of a PDF, $f(\cdot)$, is governed by the Fokker-Planck equation (also called Kolmogorov's forward equation) [85, p. 130],

$$\frac{\partial f(\cdot)}{\partial t} + \sum_i \frac{\partial F_i f(\cdot)}{\partial \psi_i} = \frac{1}{2} \sum_{i,j} \frac{\partial^2 f(\cdot)(C_{\tau_k})_{ij}}{\partial \psi_i \partial \psi_j}. \tag{3.60}$$

Essentially, this is a diffusion equation where the 'movement' of the PDF is described by the left-hand side, and the 'flattening', or 'diffusion', of the PDF due to model errors is described by the right-hand side. Inserting $f(\boldsymbol{\psi}_k|\mathbf{d}_{k-1:1})$ into (3.60) and assuming that the forward model operator is linear (i.e., $\mathbf{F}_k(\cdot)$ is given as in (3.40)) leads to the KF forecast equations, (3.44) and (3.45) (proof can be found in [85, examples 4.19–4.21]).

In the EnKF forecast step, the Fokker-Planck equation for $f(\boldsymbol{\psi}_k|\mathbf{d}_{k-1:1})$ is solved by a Monte Carlo method. That is, the updated ensemble from step $k-1$, $\mathbf{\Psi}^a_{k-1}$, is propagated forward to step k using (3.53)

$$\mathbf{\Psi}^f_k = \begin{bmatrix} \mathbf{G}^f_k \\ \mathbf{M}^f_k \end{bmatrix} = \begin{bmatrix} \mathbf{G}_k(\mathbf{M}^a_{k-1}) \\ \mathbf{M}^a_{k-1} \end{bmatrix} + \begin{bmatrix} \mathbf{E}^m_k \\ \mathbf{0} \end{bmatrix}, \tag{3.61}$$

where $\mathbf{E}_k^m = [(\boldsymbol{\epsilon}_k^m)^1, \ldots, (\boldsymbol{\epsilon}_k^m)^{N_e}]$, with $(\boldsymbol{\epsilon}_k^m)^j \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{m_k})$ for $j = 1, \ldots, N_e$. As we noted for KF, the model parameters are not changed from step $k-1$ to k; hence $\mathbf{M}_k^f = \mathbf{M}_{k-1}^a$. By using a Monte Carlo approximation we have in the limit of $N_e \to \infty$ a consistent description of $f(\boldsymbol{\psi}_k | \mathbf{d}_{k-1:1})$. Consequently, the only approximation in the forecast step is the use of a finite ensemble size.

In the analysis step, each member of the forecast ensemble, $\boldsymbol{\Psi}_k^f$, is updated using the KF analysis equation (3.50). Omitting the step index k, this is written in matrix form as

$$\boldsymbol{\Psi}^a = \boldsymbol{\Psi}^f + \mathbf{K}^e(\mathbf{D} - \mathbf{H}\boldsymbol{\Psi}^f), \tag{3.62}$$

where the $N_{\psi_k} \times N_{d_k}$ matrix $\mathbf{K}^e$ is the *approximate Kalman gain*,

$$\mathbf{K}^e = \mathbf{C}_{\psi f}^e \mathbf{H}^T (\mathbf{H}\mathbf{C}_{\psi f}^e \mathbf{H}^T + \mathbf{C}_d)^{-1}. \tag{3.63}$$

It is possible to find separate update equations for $\mathbf{G}$ and $\mathbf{M}$ from (3.62) and (3.63), using (3.57) and (3.61). Towards this end, inserting (3.61) in (3.57) yields

$$\mathbf{C}_{\psi f}^e = \begin{bmatrix} \mathbf{C}_g^e & \mathbf{C}_{gm}^e \\ \mathbf{C}_{mg}^e & \mathbf{C}_m^e, \end{bmatrix} \tag{3.64}$$

where

$$\mathbf{C}_g^e = \frac{1}{N_e - 1}(\mathbf{G}^f - \overline{\mathbf{G}}^f)(\mathbf{G}^f - \overline{\mathbf{G}}^f)^T, \quad \mathbf{C}_{gm}^e = \frac{1}{N_e - 1}(\mathbf{G}^f - \overline{\mathbf{G}}^f)(\mathbf{M}^f - \overline{\mathbf{M}}^f)^T,$$
$$\mathbf{C}_{mg}^e = \frac{1}{N_e - 1}(\mathbf{M}^f - \overline{\mathbf{M}}^f)(\mathbf{G}^f - \overline{\mathbf{G}}^f)^T, \quad \mathbf{C}_m^e = \frac{1}{N_e - 1}(\mathbf{M}^f - \overline{\mathbf{M}}^f)(\mathbf{M}^f - \overline{\mathbf{M}}^f)^T.$$

The terms involving $\mathbf{C}_{\psi f}^e$ in (3.63) are $\mathbf{C}_{\psi f}^e \mathbf{H}^T$ and $\mathbf{H}\mathbf{C}_{\psi f}^e \mathbf{H}^T$. Computing these terms using (3.64) and $\mathbf{H}_k = [\mathbf{I}_{N_{d_k}}, \mathbf{0}]$, and, subsequently, inserting (3.63) into (3.62) yields

$$\begin{bmatrix} \mathbf{G}^a \\ \mathbf{M}^a \end{bmatrix} = \begin{bmatrix} \mathbf{G}^f \\ \mathbf{M}^f \end{bmatrix} + \begin{bmatrix} \mathbf{C}_g^e \\ \mathbf{C}_{mg}^e \end{bmatrix} (\mathbf{C}_g^e + \mathbf{C}_d)^{-1}(\mathbf{D} - \mathbf{G}^f). \tag{3.65}$$

The bottom equation is identified as the update equation for the model parameters. (Note that $\mathbf{M}^f$ is actually $\mathbf{M}^a$ from the previous sequential step. However, to avoid confusing notation the superscript 'f' is used in (3.65).)

Some observations can be made on the EnKF algorithm. First, we note that in the case of Gaussian PDFs and linear forward model, and in the limit of $N_e \to \infty$, the sample mean of $\boldsymbol{\Psi}^a$ and the mean vector $\boldsymbol{\psi}^a$ given in (3.50) for KF are identical. When $\mathbf{g}(\cdot)$ is nonlinear, the EnKF analysis equation will only provide an approximate solution.

Second, we note that even though the analysis step is a linear update step, the updated ensemble will not be a resampled Gaussian distribution. Since the forecast ensemble is a Monte Carlo representation of $f(\boldsymbol{\psi}_k | \mathbf{d}_{k-1:1})$, the updated ensemble will inherent some of the non-Gaussian features from the forecast ensemble. Hence, the updated ensemble can be thought of as something in between a linear Gaussian and full Bayesian solution [63].

It can be shown that the updated ensemble for the model parameters is a linear combination of the initial ensemble, $\mathbf{M}_0$ (see, e.g., [1]). Hence, it is important that the initial ensemble for the model parameters properly spans the prior uncertainty. In practical applications of EnKF, much effort is put into the generation of the initial ensemble to obtain the best possible results.

The strength of EnKF is the ability to approximate covariance matrices using an ensemble. Hence, the computational cost associated with EnKF is mostly dependent on N_e and the computational cost of producing a forward model output. To keep the computational cost at a moderate level, N_e is often of $O(10) \sim O(100)$. Since N_e is usually much lower than N_m and/or N_{d_k}, some issues may arise. One issue is related to the inversion of the matrix $(\mathbf{C}_g^e + \mathbf{C}_d)$ in (3.65). If $\mathbf{C}_g^e$ is badly scaled and low variance is assumed in $\mathbf{C}_d$, then $(\mathbf{C}_g^e + \mathbf{C}_d)$ may be singular. In CSEM, this may be a problem if data from receivers far away from the source is not suppressed (e.g., by setting a minimum data level).

Another issue that may arise is what in the literature is denoted as 'ensemble collapse', that is, all ensemble members collapse into one vector. This arises when multiple forecast-analysis cycles are done, or if very few ensemble members are used. In either case, the internal correlation between the ensemble members increases with each analysis step. This is due to the fact that the forecast ensemble is updated using a Kalman gain that is approximated from the same forecast ensemble.

A third issue that may arise when using a relatively small ensemble size is spurious correlations. In CSEM, this can result in an update of model parameters in a part of the subsurface that is far away from the source where it is not expected that the EM signals would have sensitivity.

In the literature, ad hoc solutions to the abovementioned issues have been made, see, e.g., [1, 63].

Discussion

Alternative ensemble-based methods exist, for example, the ensemble smoother (ES) [156]. In short, ES produces a solution to the general Bayesian problem, (3.30), by following the EnKF procedure outline in the previous section, but now all observed data is conditioned simultaneously (i.e., no sequential steps). ES has shown to produce inferior results compared to EnKF in some applications, see, e.g., [62]. In [68, 69], an investigation on simultaneous and sequential methods was conducted, with the conclusion that a sequential method should outperform a simultaneous method in most cases. Some preliminary investigations for CSEM inversion studied in this work supported this conclusion. Thus, utilizing EnKF, and not ES, for inversion of CSEM data seems advantageous. Moreover, it was also suggested in [68, 69] that a particular grouping of the data based on a measure of nonlinearity of $\mathbf{g}(\cdot)$ could improve the results for a sequential method. In Paper B, no systematic grouping based on measure of nonlinearity was done.

It is also interesting to compare EnKF to a classical method like, e.g., Newton-type methods (see Section 3.2.1). A clear advantage of EnKF is the fact that no sensitivity calculation is needed to update the model parameters. EnKF updates the model parameters by relating the changes in forward model output to changes in the model parameters through correlations established in the sample covariance matrices, and not through a sensitivity matrix.

We mention that iterative ensemble-based methods exist where the sensitivity matrix is approximated from the ensemble of model parameters and forward model output, that is, $\mathbf{S} \approx (\mathbf{G} - \overline{\mathbf{G}})(\mathbf{M} - \overline{\mathbf{M}})^{\dagger}$, where '†' denotes pseudoinverse [40]. The approximation of the sensitivity matrix reduces the computational cost compared to computing the real sensitivity matrix. However, compared to EnKF, where no iterations are done, the computational cost is at best equal (if only one iteration is done), but is usually higher. Hence, using iterative ensemble-based methods is a question of how expensive the computational costs are compared to the improvement of the final results.

Lastly, we note that it is possible to also use the analysis equation from the ACKF to update the ensemble of model parameters, see Appendix B for a derivation.

Chapter 4

Electromagnetic theory and numerical modelling

In 1865, James Clerk Maxwell published the paper 'A dynamical theory of the electromagnetic field' [108], which gathered all previous theory and experimental knowledge on electricity and magnetism into a cohesive theory. He also discovered that EM fields propagate as waves, and most famously, that light were EM waves. The theory was proven experimentally over 20 years later by Heinrich Hetz. The system of equations that governs EM fields – Maxwell's equations – were gathered in their familiar form by Oliver Heaviside (he reduced the original twenty equations down to the four we use today), which greatly simplified the application of Maxwell's EM theory.

In this chapter, we formulate Maxwell's equations in several (equally valid) forms, all of which have been extensively used to model EM fields. We will also briefly summarize analytical and numerical approaches that have been widely applied in geophysical EM modelling.

4.1 Maxwell's equations in the time domain

Maxwell's equations are given in the time domain as

$$\nabla \times \mathbf{e} = -\frac{\partial \mathbf{b}}{\partial t}, \tag{4.1}$$

$$\nabla \times \mathbf{b} = \mu_0 \mathbf{j} + \mu_0 \epsilon_0 \frac{\partial \mathbf{e}}{\partial t}, \tag{4.2}$$

$$\nabla \cdot \mathbf{e} = \frac{\rho}{\epsilon_0}, \tag{4.3}$$

$$\nabla \cdot \mathbf{b} = 0, \tag{4.4}$$

where $\mathbf{e}$ is the electric field intensity (V/m), $\mathbf{b}$ is the magnetic field intensity (T), $\mathbf{j}$ is the electric current density (A/m^2), ρ is the electric charge density (C/m^3), ϵ_0 is permittivity of free space (F/m) with value $\epsilon_0 = 8.85 \times 10^{-12}$ F/m, and μ_0 is magnetic permeability of free space (H/m) with value $\mu_0 = 4\pi \times 10^{-7}$ H/m.

The first of the Maxwell's equations, (4.1), is Faraday's law. It states that a changing magnetic field induces an electric field. The second equation, (4.2), is Ampere's law (sometimes called Maxwell-Ampere's law), which states that a changing electric field together with a conducting current generates a magnetic field. Gauss' law, (4.3), states that the electric flux through any closed surface is proportional to the charge enclosed by the surface. In other words, a positive electric charge emits an electric field that diverges away from the charge. The last of Maxwell's equations, (4.4), bears no name (although sometimes called Gauss' law for magnetism), and it simple states that there are no magnetic charges analogous to electric charges.

Applying the divergence to (4.2) and using (4.3), we get

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{j} = 0, \tag{4.5}$$

which is the mathematical expression of conservation of charge, also called the continuity equation. It states that the change of charge inside a volume is equal to the current in or out through the surface.

For electromagnetic fields in matter, (4.1) – (4.4) can be expressed in a more convenient way. In matter, ρ and $\mathbf{j}$ can be separated into free and bound parts, where the bound parts are due to polarization and magnetization of the material. Since polarization and magnetization are properties of the material, it is convenient to express Maxwell's equations in terms of the free parts of ρ and $\mathbf{j}$ (which we can control). To do so, the following constitutive relations are introduced (assuming linear media)

$$\mathbf{d} = \epsilon \mathbf{e}, \tag{4.6}$$

$$\mathbf{h} = \frac{1}{\mu}\mathbf{b}, \tag{4.7}$$

where **d** is the electric displacement and **h** is an auxiliary vector (many authors denote this, and not **b**, as the 'magnetic field', in which case **b** is denoted 'magnetic induction' or 'magnetic flux density'). Furthermore, $\epsilon = \epsilon_0(1 + \chi_e)$ and $\mu = \mu_0(1 + \chi_m)$ with χ_e and χ_m being the electric and magnetic susceptibility, respectively. Using (4.6) and (4.7), the unknowns in Maxwell's equations can be any combination of **e**, **b**, **d** and **h**, if wanted. However, the common practice is to let **e** and **h** be the unknown vector fields, in which case (4.1) – (4.4) becomes

$$\nabla \times \mathbf{e} = -\mu \frac{\partial \mathbf{h}}{\partial t}, \tag{4.8}$$

$$\nabla \times \mathbf{h} = \mathbf{j}_f + \epsilon \frac{\partial \mathbf{e}}{\partial t}, \tag{4.9}$$

$$\nabla \cdot \epsilon \mathbf{e} = \rho_f, \tag{4.10}$$

$$\nabla \cdot \mu \mathbf{h} = 0, \tag{4.11}$$

where $\mathbf{j}_f$ and ρ_f are the free current and charge densities, respectively.

In CSEM, it is advantageous to split $\mathbf{j}_f$ as

$$\mathbf{j}_f = \mathbf{j}_c + \mathbf{j}_e = \sigma \mathbf{e} + \mathbf{j}_e, \tag{4.12}$$

where $\mathbf{j}_c$ is the current density induced in conducting matter by an external source current density, $\mathbf{j}_e$. The equality $\mathbf{j}_c = \sigma \mathbf{e}$ is due to Ohm's law, and it introduces the electric conductivity, σ (S/m).

Lastly in this section, we mention that the parameters ϵ, μ, and σ are tensors for anisotropic media and scalars for isotropic media. Furthermore, they are functions of position for heterogeneous media, whereas for homogeneous media they are not.

4.2 Maxwell's equations in the frequency domain

If $\mathbf{e}$, $\mathbf{h}$, and $\mathbf{j}_e$ are harmonic oscillating functions with a single frequency, they are referred to as time-harmonic [86]. Let the vector fields vary with the harmonic time convention $\exp(-i\omega t)$, where $\omega = 2\pi f$ denote the angular frequency with f being the ordinary frequency (Hz), and $i = \sqrt{-1}$. Assuming no free charges ($\rho_f = 0$), the time-domain Maxwell's equations (4.8) – (4.12) can now be expressed in the frequency domain as

$$\nabla \times \mathbf{e} = i\omega\mu\mathbf{h}, \tag{4.13}$$

$$\nabla \times \mathbf{h} = \mathbf{j}_e + \sigma\mathbf{e} - i\omega\epsilon\mathbf{e}, \tag{4.14}$$

$$\nabla \cdot \epsilon\mathbf{e} = 0, \tag{4.15}$$

$$\nabla \cdot \mu\mathbf{h} = 0. \tag{4.16}$$

The frequency-domain Maxwell's equations are widely used in CSEM, and the advantage with this approach is the ability to solve the equations only for a few frequencies of interest.

4.3 Wave equation

Maxwell's equations, as given in Section 4.1 or 4.2, are first-order coupled partial differential equations. It is possible to decouple the equations with some simple mathematical manipulations. For the time-domain Maxwell's equations, applying the curl operator to (4.8) and substituting (4.9) with (4.12) eliminates $\mathbf{h}$, resulting in

$$\nabla \times (\nabla \times \mathbf{e}) + \mu\epsilon\frac{\partial^2\mathbf{e}}{\partial t^2} + \mu\sigma\frac{\partial\mathbf{e}}{\partial t} = -\mu\frac{\partial\mathbf{j}_e}{\partial t}. \tag{4.17}$$

Equivalent manipulations can be done with the frequency-domain Maxwell's equations, leading to

$$\nabla \times (\nabla \times \mathbf{e}) - (\omega^2\mu\epsilon + i\omega\mu\sigma)\mathbf{e} = i\omega\mu\mathbf{j}_e. \tag{4.18}$$

In CSEM, it is common to neglect the contribution from the displacement current (the so-called quasi-static approximation [83]), due to the application of low-frequency signals. Removing the displacement current term in (4.17) and (4.18) leads to

$$\nabla \times (\nabla \times \mathbf{e}) + \mu\sigma\frac{\partial\mathbf{e}}{\partial t} = -\mu\frac{\partial\mathbf{j}_e}{\partial t}, \tag{4.19}$$

and

$$\nabla \times (\nabla \times \mathbf{e}) - i\omega\mu\sigma\mathbf{e} = i\omega\mu\mathbf{j}_e, \tag{4.20}$$

respectively. Informally speaking, by neglecting displacement currents, the wave-type equations (4.17) and (4.18) are changed to diffusion-type equations (4.19) and (4.20) [104]. The price paid for decoupling Maxwell's equations is that (4.19) and (4.20) are now second-order partial differential equations.

When $\mathbf{e}$ has been found from either (4.19) or (4.20), $\mathbf{h}$ can be found using Faraday's law,

$$\frac{\partial\mathbf{h}}{\partial t} = -\frac{1}{\mu}\nabla \times \mathbf{e} \tag{4.21}$$

for the time domain, or

$$\mathbf{h} = \frac{1}{i\omega\mu}\nabla \times \mathbf{e} \tag{4.22}$$

for the frequency domain.

For completeness, we mentioned that **e** can be eliminated by similar manipulations as above. Applying the curl operator to (4.9) and using (4.8) with (4.12) (neglecting the contribution from the displacement current), the resulting wave equations for **h** are given as

$$\nabla \times (\nabla \times \mathbf{h}) + \mu\sigma\frac{\partial \mathbf{h}}{\partial t} = \nabla \times \mathbf{j}_e, \tag{4.23}$$

and with similar manipulations for the frequency domain Maxwell's equations we have

$$\nabla \times (\nabla \times \mathbf{h}) - i\omega\mu\sigma\mathbf{h} = \nabla \times \mathbf{j}_e. \tag{4.24}$$

4.4 Scalar and vector potentials

An alternative approach to the wave equations for finding the unknown vector fields, **e** and **h**, is to use scalar and vector potentials. Since **h** is divergenceless, it is possible to express it in terms of a vector potential **a**,

$$\mathbf{h} = \frac{1}{\mu}\nabla \times \mathbf{a}. \tag{4.25}$$

Substituting (4.25) into (4.13) gives

$$\nabla \times \mathbf{e} = i\omega\nabla \times \mathbf{a}, \tag{4.26}$$

or

$$\nabla \times (\mathbf{e} - i\omega\mathbf{a}) = 0. \tag{4.27}$$

Since ($\mathbf{e} - i\omega\mathbf{a}$) has vanishing curl, it can be expressed in terms of a scalar potential V,

$$\mathbf{e} - i\omega\mathbf{a} = \nabla V. \tag{4.28}$$

or

$$\mathbf{e} = i\omega\mathbf{a} + \nabla V. \tag{4.29}$$

Similar considerations in the time domain yields

$$\mathbf{e} = -\frac{\partial \mathbf{a}}{\partial t} - \nabla V. \tag{4.30}$$

The decomposition of a vector field into vector and scalar potentials, as done in (4.29) and (4.30), is denoted Helmoltz decomposition.

Inserting (4.25) and (4.29) into (4.14) (neglecting contributions from displacement currents), and inserting (4.29) into (4.15), leads to the following system of equations

$$\nabla \times (\nabla \times \mathbf{a}) - \mu\sigma\,(i\omega\mathbf{a} + \nabla V) = \mu\mathbf{j}_e, \tag{4.31}$$

$$\nabla^2 V + i\omega\nabla \cdot \mathbf{a} = 0. \tag{4.32}$$

Similarly for the time domain, we have the system of equations

$$\nabla \times (\nabla \times \mathbf{a}) + \mu\sigma \left(\frac{\partial \mathbf{a}}{\partial t} + \nabla V \right) = \mu \mathbf{j}_e, \tag{4.33}$$

$$\nabla^2 V + \frac{\partial}{\partial t}(\nabla \cdot \mathbf{a}) = 0. \tag{4.34}$$

For 3D problems the system of equations (4.31) and (4.32), or (4.33) and (4.34), effectively reduces the search for the six unknowns in **e** and **h** to a search for the four unknowns in V and **a**.

The vector fields **e** and **h** are not uniquely described by the potentials. This can be seen, e.g., in (4.29) where **e** is described by both V and **a**, which means that we have an extra degree of freedom. As long as **e** and **h** do not change, extra conditions can thus be imposed on V and **a**; these are called gauge transformations. Many gauge transformations have been introduced to solve (4.31) and (4.32), or (4.33) and (4.34), more easily, e.g., Coulomb gauge and Lorentz gauge.

4.5 Analytical solutions

Analytical solutions to Maxwell's equations can be found when the earth's subsurface can be modeled as a set of horizontally stratified layers. In this case, the material properties only varies in one direction. For a typical CSEM setup with a HED source, the components of **e** and **h** are given as integrals of the form [90]

$$\int_0^\infty f(\lambda) J_i(\lambda r)\, \mathrm{d}\lambda. \tag{4.35}$$

The integral (4.35) is called the Hankel transform, and is essentially a double Fourier transform. The kernel function $f(\lambda)$ depends on the subsurface material properties, and $J_i(\lambda)$ is an i'th order Bessel function of the first kind. Derivation of **e** and **h** on the form (4.35) can be found, e.g., in [38, 161].

In most 1D EM modelling, evaluation of (4.35) is done numerically using a digital filter approach (see, e.g., [7]), which provides fast and accurate solutions. The numerical evaluation of (4.35) can also be done using quadrature methods (see, e.g., [37]), although they are less popular in geophysical EM applications (see [90] for a comparison of digital filter and quadrature approaches).

1D EM modelling has been widely used in many areas of CSEM. To mention a few applications, 1D solutions provided valuable insights on the physics of CSEM (see, e.g., [153]), and helped with experimental design [67]. Furthermore, horizontally stratified layers are often assumed as background models when calculating solutions for an anomalous conductivity distribution in the integral equation formulation (see Section 4.6.3 below) [160]. 1D solutions can also be used to generate source terms for 3D modelling applications [122]. Analytical solutions also exist when the horizontally layered subsurface consists of an anisotropic conductivity distribution (see, e.g., [105, 165]).

4.6 Numerical modelling

In many geological scenarios, the subsurface contains structures that varies in 2D or 3D. Hence, the subsurface cannot be well approximated as a set of horizontally stratified layers, and, consequently, analytical solutions of Maxwell's equations are not valid. For general subsurface models, calculations of **e** and **h** must be done using numerical methods. Roughly speaking, the numerical approaches for EM modelling can be divided into three main groups: differential equation (DE) approaches, integral equation (IE) approaches, and hybrid approaches. DE approaches are easy to implement even for complex subsurface structures, and result in a sparse matrix system. IE approaches contain more involved mathematics and result in a dense matrix system, but the vector fields are only calculated in an anomalous region. Hybrid approaches try to combine parts of DE and IE approaches. In the following sections, we briefly discuss the DE methods, specifically finite difference (FD) and finite element (FE) methods, and the IE and hybrid methods. For a comprehensive review on the these topics see, e.g., [13, 28].

4.6.1 Finite difference methods

In FD, the solution procedure is relatively simple [134]: discretize the subsurface model into a grid of nodes; approximate the governing differential equations by finite differences; and solve the differential equations subject to boundary and/or initial conditions. Most often, a structured rectangular grid is employed to discretize the subsurface model, which leads to an easy implementation of the FD method. In particular, the staggered grid approach of Yee [167] has been widely applied in EM modelling.

FD approximation of the first-order Maxwell's equations can be found, e.g., in [42, 159]. For FD approximation of the wave equation formulation of Maxwell's equations see, e.g., [119, 123]. Some authors have solved the scalar-vector potential formulation of Maxwell's equations using FD approximation, see, e.g., [11, 81]. We also mention that FD methods have been applied for anisotropic subsurface structures, see, e.g., [51, 164].

4.6.2 Finite element methods

The solution procedure for any FE method involves the following steps [134]: discretize the subsurface model into a finite number of elements; derive governing equations for an element; assemble all elements in the solution region; and solve the system of equations obtained. Typically, and contrary to most FD methods, unstructured grids are applied to discretize the subsurface model. This provides greater flexibility to conform grids to complex model features. The development of robust mesh generators have reduced the complication associated with model discretization (a popular choice for triangular mesh generation can be found in [144]).

The FE method has been applied to EM problems by many authors: for application on the first-order Maxwell's equations see, e.g., [92, 115]; for application on the wave equation see, e.g., [35, 93]; and for application on the scalar-vector potential formulation see, e.g., [16, 94]. The FE methods have also been applied on models with anisotropic conductivity distributions, see, e.g., [95, 99].

A widely applied FE formulation is the 2.5D formulation (see, e.g., [80, 148]). Here, the vector fields, **e**, **h**, and **j**, are functions in 3D, while the material properties, σ, ϵ and μ, are functions only in 2D; thus the last dimension is defined as the strike direction. The splitting is done using a 1D Fourier transform, and the 3D EM problem is effectively reduced to a sequence of 2D problems (one for each Fourier mode). Several authors have applied the 2.5D formulation for the EM modelling problem; see, e.g., [115, 129]. In Papers A, B, and C, a 2.5D FE method was used to generate forward model outputs for the inversion of CSEM data.

4.6.3 Integral equations and hybrid methods

The IE approach for modelling EM fields was introduced in a paper by Dmitriev [54]. In this approach, the first-order Maxwell's equations are reduced to Fredholm integral equations (of first or second kind). Moreover, with the IE approach, the electric conductivity is divided into a background part (in the following denoted by a superscript 'b') and anomalous part (in the following denoted by a superscript 'a'), $\sigma = \sigma^b + \sigma^a$, which in turn splits **e** and **h** into a background and anomalous part. Without providing the derivation, the expressions for the background and anomalous part of **e** and **h** are given as

$$\mathbf{f}^b(\mathbf{r}') = \int_Q \widehat{\mathbf{G}}_{\mathrm{f}}(\mathbf{r}'|\mathbf{r})\mathbf{j}_e(\mathbf{r})\,\mathrm{d}v, \tag{4.36}$$

and

$$\mathbf{f}^a(\mathbf{r}') = \int_D \widehat{\mathbf{G}}_{\mathrm{f}}(\mathbf{r}'|\mathbf{r})\sigma^a(\mathbf{r})\left[\mathbf{f}^b(\mathbf{r}) + \mathbf{f}^a(\mathbf{r})\right]\,\mathrm{d}v, \tag{4.37}$$

respectively. Here, **f** is either **e** or **h**, and $\widehat{\mathbf{G}}_{\mathrm{f}}$ is the Green's tensor calculated for the background conductivity. Furthermore, Q is the region containing the source, D is the anomalous region, and **r** and $\mathbf{r}'$ are vectors denoting position. The calculation of $\mathbf{f}^b$ is done analytically if the background model is a horizontally stratified model, hence it is only necessary to discretize D in order to solve the EM modelling problem.

In summary, the procedure for calculating **f** is: calculate $\mathbf{f}^b$ in D ($\mathbf{r}' \in D$) and in the receivers ($\mathbf{r}' \in R$, with R being the receiver positions) from (4.36); calculate $\mathbf{f}^a$ in D from (4.37) ($\mathbf{r}, \mathbf{r}' \in D$); and compute $\mathbf{f}^a$ in the receivers from (4.37) ($\mathbf{r} \in D$ and $\mathbf{r}' \in R$). Many authors have implemented the IE approach, see, e.g., [14, 82].

The computationally intensive part of IE approaches is the calculation of $\mathbf{f}^a$ in D. To alleviate the computational expenses, hybrid methods have been introduced where calculation of $\mathbf{f}^a$ in D is done using a FE or FD approach. Hybrid methods where FE and IE are combined can be found in, e.g., [27, 76]. Hybrid methods where FD and IE methods are combined can be found in, e.g., [17, 168].

4.6.4 Boundary conditions

When solving the EM modelling problem on a computational grid, Ω, suitable boundary conditions (BC) must be supplemented to account for the behaviour of the vector fields, **e** and **h**, on the computational boundary, $\partial\Omega$. Traditionally, the BCs in numerical methods are the Dirichlet, Neumann, and Cauchy BC (also called Dirichlet BC of first, second, and third order). Dirichlet BC fix the field values at the boundary; Neumann BC fix the gradient of the fields normal to the boundary; and Cauchy BC is a combination of Dirichlet and Neumann BC. A typically used Dirichlet BC is the perfect conducting surface: $\mathbf{n} \times \mathbf{e} = \mathbf{0}$ and $\mathbf{n} \cdot \mathbf{h} = 0$, where $\mathbf{n}$ is a normal vector on $\partial\Omega$. To apply the perfect conducting surface BC, $\partial\Omega$ has to be significantly far away from the region of interest.

Among other BCs are the infinite element technique [35], absorbing BC [120], perfectly matched layers [20], and asymptotic BC [162].

4.6.5 Linear system solver

Whether a DE, IE, or hybrid approach is applied, the numerical modelling of (any version of) Maxwell's equations results in a linear system,

$$\mathbf{A}\mathbf{x} = \mathbf{s}. \tag{4.38}$$

Here, $\mathbf{A}$ is a square coefficient matrix, which for FD and FE methods is a large, sparse matrix; while for IE methods $\mathbf{A}$ is a dense matrix, which is typically smaller in size than for FD and FE methods. The unknown, $\mathbf{x}$, is either $\mathbf{e}$ or $\mathbf{h}$, and $\mathbf{s}$ is a vector containing the external source, $\mathbf{j}_e$.

The algorithms solving the linear system (4.38) can be divided into two approaches: direct methods and iterative methods. Basically, all direct methods are based on Gauss elimination, and, in particular, decomposition methods are widely used. Among the most popular decomposition methods are the LU decomposition, LDL decomposition (modified Cholesky decomposition), frontal and multifrontal methods. For a detailed presentation of decomposition methods see, e.g., [70].

In some cases, decomposition methods can be slow, particular when $\mathbf{A}$ becomes large. An alternative is to use iterative methods. In short, the procedure of iterative methods is to iteratively improve an approximate solution to $\mathbf{x}$ until a termination criterion is met. Among the most used iterative methods are Krylov subspace methods, including conjugate and biconjugate gradient, generalized minimal residual, quasi-minimal residual and biconjugate gradient stabilized. For a review on Krylov subspace methods see, e.g., [145]. Lastly, we mention that the rate of convergence in iterative methods can be (greatly) improved using a preconditioner (see, e.g., [15]).

Chapter 5

Parameter representation using level sets

In Chapter 3, we listed ‘parameterization’ as one of three steps for studying a physical system. It consisted of finding a suitable set of model parameters to describe the physical system under study. In general, the model parameters are described by a parameter function of some kind. However, we are interested in solving the discrete inverse problem, thus we need to approximate the parameter function by a set of discrete quantities; this is called *reparameterization*.

In this chapter, we discuss a widely popular parameterization method – *the level-set representation* – and outline some possible reparameterizations of this method. Furthermore, we will briefly discuss different strategies for solving the inverse problem of identifying the unknown parameter function.

5.1 Parameterization and the level-set representation

Let D denote the region where we want to estimate a parameter function, q. Furthermore, let $\mathbf{r}$ be a vector denoting spatial position in D. A general representation of q is often given as a linear series expansion,

$$q(\mathbf{r}) = \sum_{j=1}^{N_c} c_j(\mathbf{r})\Phi_j(\mathbf{r}), \tag{5.1}$$

where $\mathbf{\Phi}(\mathbf{r}) = [\Phi_1(\mathbf{r}),\ldots,\Phi_{N_c}(\mathbf{r})]$ are the basis functions and $\mathbf{c}(\mathbf{r}) = [c_1(\mathbf{r}),\ldots,c_{N_c}(\mathbf{r})]^T$ are the corresponding expansion functions. The representation (5.1) divides D into N_c regions given by $\operatorname{supp}\Phi_j$, with $q(\mathbf{r}) = c_j(\mathbf{r})$ in each region. The linear series expansion (5.1) describes a *parameterization* of q, and is often referred to as a *model-based representation* of q. In the special case where the regions are given by the N_g grid cells in a numerical method (either DE or IE approach; see Section 4.6) and $\mathbf{c}$ is independent of $\mathbf{r}$, we can write (5.1) as $\mathbf{q} = \mathbf{c} = [c_1,\ldots,c_{N_g}]^T$. This describes a so-called *pixel-based representation*.

In the following, we drop the dependence of $\mathbf{c}$ on $\mathbf{r}$ and assume for simplicity that each region has a constant value $q(\mathbf{r}) = c_j$. In the seminal paper [128], Osher and Sethian introduced the level-set (LS) representation as a part of a front propagation method – the LS method. The LS representation is simple, yet powerful: the boundary between two regions are represented implicitly. That is, the region boundary is derived

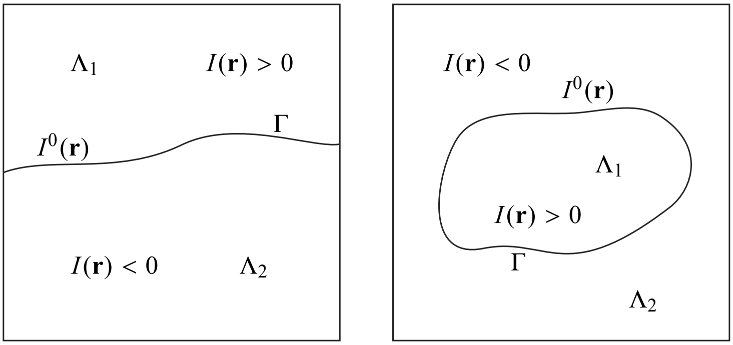

Figure 5.1: Two examples of the LS representation with two regions

from a higher-dimensional function, which again is represented explicitly. The higher-dimensional function is called the LS function, and we denote it as $I(\mathbf{r})$.

Generally, the boundary is given by the zero-contour of $I(\mathbf{r})$ (this choice is arbitrary; we could, with minor modifications, have given it as any contour of $I(\mathbf{r})$). When $I(\mathbf{r})$ is a function in 3D, the zero-contour, and hence the region boundary, is a line in 2D. Similarly, when $I(\mathbf{r})$ is a function in 4D, the zero-contour is plane in 3D. In the following, we will only consider the case when $I(\mathbf{r})$ is a 3D function, and we will denote the 2D zero-contour of $I(\mathbf{r})$ as $I^0(\mathbf{r})$.

For simplicity, consider the partition of D into two regions Λ_1 and Λ_2 with a common edge $\Gamma = \partial\Lambda_1 \cap \partial\Lambda_2$. The LS function is then given as the continuous function satisfying

$$\begin{aligned} I(\mathbf{r}) &> 0, \quad \text{for } \mathbf{r} \in \Lambda_1, \\ I(\mathbf{r}) &< 0, \quad \text{for } \mathbf{r} \in \Lambda_2, \\ I(\mathbf{r}) &= I^0(\mathbf{r}), \quad \text{for } \mathbf{r} \in \Gamma. \end{aligned} \tag{5.2}$$

An illustration is given in Figure 5.1.

In the case of two regions ($N_c = 2$), the parameter function (5.1) is given as

$$q(\mathbf{r}) = c_1\Phi_1(\mathbf{r}) + c_2\Phi_2(\mathbf{r}). \tag{5.3}$$

Define $\Phi_1(\mathbf{r}) = H(I(\mathbf{r}))$ and $\Phi_2(\mathbf{r}) = 1 - H(I(\mathbf{r}))$, where H is the Heaviside function

$$H(I) = \begin{cases} 1, & \text{if } I \geq 0, \\ 0, & \text{if } I < 0. \end{cases} \tag{5.4}$$

Inserting this into (5.3) yields the LS representation for two regions

$$q(\mathbf{r}) = c_1 H(I(\mathbf{r})) + c_2(1 - H(I(\mathbf{r}))). \tag{5.5}$$

From (5.2) and (5.4) we see that (5.5) gives $q = c_1$ for $\mathbf{r} \in \Lambda_1$ and $q = c_2$ for $\mathbf{r} \in \Lambda_2$.

For $N_c = 2$, the LS representation is uniquely given by (5.5). However, in the case when $N_c > 2$, the LS representation is not unique and several formulations have been proposed. In the following sections, we will present two LS representations important for this work. Alternative LS formulations can be found, e.g., in [56, 100, 103, 169].

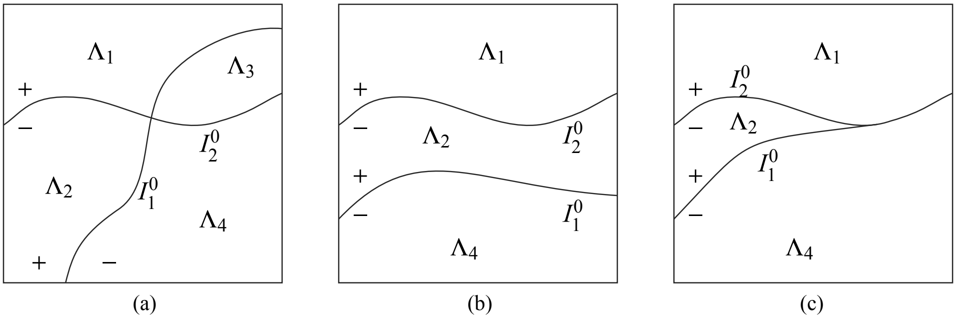

Figure 5.2: Illustrations using the Vese-Chan representation. (a) Four regions represented by I_1^0 and I_2^0 intersecting. (b) Three regions represented by I_1^0 and I_2^0 not intersecting. (c) Λ_1 and Λ_4 made adjacent by partially coinciding I_1^0 and I_2^0.

Lastly in this section, we mention that the LS method has been applied in a wide range of scientific fields; see, e.g., [31, 55] for reviews. For inverse problems, the LS method was introduced by Santosa in [136], where he suggested two approaches: the *evolution approach*, where the LS functions are adjusted with a velocity corresponding to the descent direction; and the *optimization approach*, where the update of the LS functions are determined directly by the optimization algorithm. In this work, we follow the latter approach.

5.1.1 Vese-Chan representation

The most commonly applied LS representation for $N_c > 2$ is the Vese-Chan representation [157]; see also [36, 158]. Let $I_i(\mathbf{r})$ denote the i'th LS function where $i = 1, \ldots, N_I$, with N_I being the total number of LS functions. Furthermore, let b_j^i denote element number j in the N_I-dimensional binary representation of $(j-1)$, $\text{bin}(j-1) = \left[b_j^1, \ldots, b_j^{N_I}\right]$, for $j = 1, \ldots, 2^{N_I}$. With these definitions, the Vese-Chan representation is given by

$$q(\mathbf{r}) = \sum_{j=1}^{N_c} c_j \prod_{i=1}^{N_I} E_j(I_i(\mathbf{r})), \tag{5.6}$$

where

$$E_j(I_i) = \begin{cases} H(I_i), & \text{if } b_j^i = 0, \\ 1 - H(I_i), & \text{if } b_j^i = 1. \end{cases} \tag{5.7}$$

From the definition of $\text{bin}(j-1)$ it is possible to represent $N_c = 2^{N_I}$ regions for N_I LS functions. Note, however, that 2^{N_I} is the maximum number of regions for N_I LS functions. The actual number of regions occurring in D depends on the configuration of $\{I_i^0(\mathbf{r})\}_{i=1}^{N_I}$. Let us illustrate this with an example. If we let $N_I = 2$, then there is

$2^{N_I} = 2^2 = 4$ possible combinations for $\text{bin}(j-1)$,

$$\begin{bmatrix} \text{bin}(0) \\ \text{bin}(1) \\ \text{bin}(2) \\ \text{bin}(3) \end{bmatrix} = \begin{bmatrix} b_1^1 & b_1^2 \\ b_2^1 & b_2^2 \\ b_3^1 & b_3^2 \\ b_4^1 & b_4^2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}, \tag{5.8}$$

and using (5.7) we get

$$\begin{aligned} E_1(I_1) &= E_2(I_1) = H(I_1), \\ E_1(I_2) &= E_3(I_2) = H(I_2), \\ E_3(I_1) &= E_4(I_1) = 1 - H(I_1), \\ E_2(I_2) &= E_4(I_2) = 1 - H(I_2). \end{aligned}$$

Inserted into (5.6) yields

$$\begin{aligned} q(\mathbf{r}) &= c_1 E_1(I_1)E_1(I_2) + c_2 E_2(I_1)E_2(I_2) + c_3 E_3(I_1)E_3(I_2) + c_4 E_4(I_1)E_4(I_2), \\ &= c_1 H(I_1)H(I_2) + c_2 H(I_1)(1 - H(I_2)) \\ &\quad + c_3(1 - H(I_1))H(I_2) + c_4(1 - H(I_1)(1 - H(I_2))), \end{aligned} \tag{5.9}$$

where the dependence of I_i on $\mathbf{r}$ has been suppressed for convenience. With q given as in (5.9) the maximum number of regions possible is $N_c = 2^2 = 4$, and an illustration of this fact for a configuration of I_1^0 and I_2^0 is given in Figure 5.2a. The + and - signs in the figure indicates which side of I_i^0 the LS functions are positive and negative. In Figure 5.2a, we see that in order to represent four regions, I_1^0 and I_2^0 must intersect. If we consider a configuration where they do not intersect, as in Figure 5.2b, only three regions will occur in D. In the case of the non-intersecting I_1^0 and I_2^0 in Figure 5.2b, Λ_3 has vanished since $\text{supp}\,(1 - H(I_1))H(I_2) = \emptyset$. In general, the number of regions that can occur in D for a particular number of LS functions is $p \in [N_I + 1, 2^{N_I}]$. The lower bound accounts for the case where none of the I_i^0's are intersecting, while the upper bound accounts for the case where all of the I_i^0's are intersecting.

The above example illustrates a topological constraint of the Vese-Chan representation. When $\{I_i^0\}_{i=1}^{N_I}$ are intersecting, new regions, which have no relation to the neighbouring regions, are introduced. Consider again Figure 5.2a where we have seen that Λ_3 occurs due to the intersection of I_1^0 and I_2^0. In order for Λ_4 and Λ_1 to become adjacent, I_1^0 and I_2^0 must partially coincide as illustrated in Figure 5.2c. As pointed out in [56, 107], this is a drawback when using the Vese-Chan representation in an estimation setting, since partial joining of the I_i^0's is unlikely to happen during a practical estimation.

5.1.2 Hierarchical representation

To alleviate the unwanted topological constraint seen for the Vese-Chan representation (see Section 5.1.1), the hierarchical representation was introduced in [107]. The hierarchical representation is best presented as a scale-by-scale description of q. To simplify

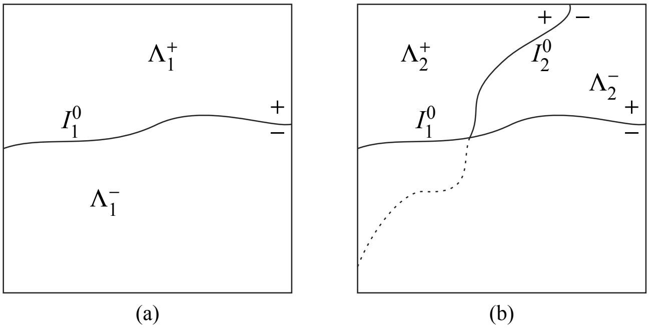

Figure 5.3: Illustrations using the hierarchical representation. (a) Example of scale 1 representation. (b) Example of scale 2 representation. I_2^0 divide Λ_1^+ from (a) into two new regions, $\Lambda_2^\pm$.

the notation, let $H_i^+ = H(I_i(\mathbf{r}))$ and $H_i^- = 1 - H(I_i(\mathbf{r}))$, and let the corresponding expansion functions be given as c_i^+ and c_i^-. Furthermore, let the region where $I_i(\mathbf{r}) > 0$ be denoted Λ_i^+ and the region where $I_i(\mathbf{r}) < 0$ be denoted Λ_i^-.

At scale 1, the domain D is divided into two regions, Λ_1^+ and Λ_1^-, using I_1^0, and q is expressed as in (5.5), which, using the above notation, is written as

$$q = c_1^+ H_1^+ + c_1^- H_1^-. \tag{5.10}$$

See Figure 5.3a for an arbitrary example of a scale 1 representation.

At scale 2, we have two possibilities: divide Λ_1^+ or divide Λ_1^-. Hence, c_1^+ and c_1^- may be expanded as

$$\begin{aligned} c_1^+ &= c_2^+ H_2^+ + c_2^- H_2^-, \quad \text{and/or} \\ c_1^- &= c_3^+ H_3^+ + c_3^- H_3^-. \end{aligned} \tag{5.11}$$

As an illustration of the scale 2 representation, consider the case where Λ_1^+ from Figure 5.3a has been divided into two new regions $\Lambda_2^\pm$ by the introduction of I_2^0, see Figure 5.3b. Since Λ_1^+ has been divided, c_1^+ must be expanded according to (5.11), leading to q being given by

$$q = (c_2^+ H_2^+ + c_2^- H_2^-) H_1^+ + c_1^- H_1^-. \tag{5.12}$$

From Figure 5.3b we note that I_2^0 has a partially solid and partially dashed part. The solid part indicates where I_2^0 has an effect on the representation of q, while the dashed part indicates where I_2^0 has no effect on q due to annihilation by H_1^+ (confer (5.12)). It is this annihilation effect that is responsible for avoiding the unwanted topological constraint seen in the Vese-Chan representation, and thus there are no restrictions on which regions that can be adjacent to each other.

The above procedure continues at subsequent scales with further expansion of the $c_j^\pm$'s as new regions are introduced. Since the introduction of a new I_i^0 leads to two new regions in an already existing region (compare Figures 5.3a and 5.3b), it is clear that $N_I = N_c - 1$ LS functions are needed to represent N_c regions.

By noticing that the regions at one scale are 'parents' to the regions at the next scale, the above procedure can be ordered in a binary tree structure. Hence, at scale 1, $c_1^\pm$ are the 'children' of the domain D and are the 'parents' of $c_2^{\pm}$ and $c_3^{\pm}$ at scale 2, which again will be parents for subsequent regions at scale 3, and so on. A general formula for expanding $c_j^{\pm}$'s at scale $(L-1)$ can then be written as

$$\begin{aligned} &c_j^{(-)^i} = c_i^+ H_i^+ + c_i^- H_i^-, \\ &j \in [2^{L-2}, 2^{L-1}-1], \quad i \in [2j, 2j+1], \end{aligned} \tag{5.13}$$

where

$$(-)^i = \begin{cases} -, & \text{if } i \text{ is odd,} \\ +, & \text{if } i \text{ is even,} \end{cases} \tag{5.14}$$

is introduced as a shorthand for determining the superscript on the $c_j^{\pm}$'s.

5.1.3 Related parameterization methods

If we let $\Phi_j(\mathbf{r}) = \chi_j(\mathbf{r})$, where χ_j is the indicator function (e.g., the Heaviside function), and let c_j be independent of $\mathbf{r}$, then (5.1) is given as

$$q(\mathbf{r}) = \sum_{j=1}^{N_c} c_j \chi_j(\mathbf{r}). \tag{5.15}$$

This corresponds to the widely used parameterization method called standard (or classical) zonation [84]. With standard zonation, D is divided into fixed regions with constant value $q(\mathbf{r}) = c_j$. Hence, during an estimation it is not possible to adjust the region boundaries; only the c_j's can be adjusted. This is different from the LS representation where it is possible to (implicitly) adjust the region boundaries by changing the value of the LS functions. Due to the connection with standard zonation, LS representations are sometimes denoted non-standard zonation methods.

There also exist parameterization methods where the region boundaries are explicitly adjusted, for example, the sharp boundary approach [52, 146]. Here, nodes with interpolated lines in between make up the region boundaries, and each region has a constant parameter value $q(\mathbf{r}) = c_j$. Only the vertical position of the nodes are adjustable. In [2], the sharp boundary method was expanded using 2D polygons for defining regions.

A drawback of using explicit representation of the region boundaries is the handling of regions merging or splitting apart, which can happen during an estimation sequence. To prevent region boundaries colliding, a minimum distance between nodes was set in [2]. In [52], region boundaries that intersected were instead given equal depth or removed completely. As seen above, due to the implicit representation of the region boundaries in the LS representation, regions merging or splitting does not pose a problem. For the Vese-Chan representation a new region is introduced if regions merge, while for the hierarchical representation the scale-by-scale description prevents new regions appearing if regions merge.

5.1.4 Smoothed level-set representation

In both the Vese-Chan representation and hierarchical representation, q was represented using the standard Heaviside function. This leads to a discontinuous transition between the regions, which again can lead to difficulties in an estimation setting. A numerical study done in [102] showed that a smooth approximation of the Heaviside function was needed to reduce the nonlinearity of $\mathbf{g}(\cdot)$ in the estimation problem. In other words, a smooth representation of q reduces the risk of finding a non-optimal solution to the estimation problem.

Another reason for replacing the standard Heaviside function with a smooth approximation is that in many classical estimation algorithms, the derivative of the Heaviside function is needed. In the sense of distributions, the derivative of the Heaviside function is the Dirac delta, $\delta(x)$, which cannot be expressed outside an integral. Hence, many authors have suggested approximations to the Dirac delta and Heaviside function, see, e.g., [157, 158]. In Papers A, B, and C, we used the approximations given in [157],

$$\tilde{H}(I) = \frac{1}{\pi}\tan^{-1}(I) + \frac{1}{2}, \qquad \tilde{\delta}(I) = \frac{1}{\pi(1+I^2)}. \tag{5.16}$$

Note that the $\tilde{\delta}(I)$ is the derivative of $\tilde{H}(I)$.

5.2 Reparameterization

To adjust the functions $\mathbf{\Phi}(\mathbf{r})$ and $\mathbf{c}(\mathbf{r})$ in an estimation sequence we need to introduce some control coefficients in (5.1):

$$q(\mathbf{r};\mathbf{w},\mathbf{a}) = \sum_{j=1}^{N_c} c_j(\mathbf{r};\mathbf{w})\Phi_j(\mathbf{r};\mathbf{a}). \tag{5.17}$$

$\mathbf{\Phi}(\mathbf{r};\mathbf{a})$ and $\mathbf{c}(\mathbf{r};\mathbf{w})$ are thus determined by the coefficients $\mathbf{a} = [a_1,\dots,a_{N_a}]^T$ and $\mathbf{w} = [w_1,\dots,w_{N_w}]^T$, respectively. The $\mathbf{a}$-coefficients are introduced to be able to change the shape and position of the regions, while the $\mathbf{w}$-coefficients are introduced to be able to change $c_j(\mathbf{r};\mathbf{w})$ within the regions. When $\mathbf{w}$ and $\mathbf{a}$ are determined, $q(\mathbf{r};\mathbf{w},\mathbf{a})$ can be found for any $\mathbf{r} \in D$; hence, $q(\mathbf{r};\mathbf{w},\mathbf{a})$ is *reparameterized* by the coefficients $\mathbf{w}$ and $\mathbf{a}$.

In the next two sections, we look at how the LS function, $I_i(\mathbf{r})$, and the expansion functions, $\mathbf{c}(\mathbf{r})$, in can be represented using $\mathbf{a}$ and $\mathbf{w}$, respectively.

5.2.1 Representation of level-set functions

If we consider the case where D is divided into a grid with N_g cells, e.g., associated with a numerical method for computing forward model outputs, then a straightforward representation of the LS functions is given by

$$I_i(\mathbf{r};\mathbf{a}_i) = \sum_{k=1}^{N_g} a_k^i \chi_k(\mathbf{r}), \tag{5.18}$$

where χ_k is the indicator function for grid cell number k (see Figure 5.4a). Note that $\mathbf{a}_i$ is a subset of the $\mathbf{a}$-coefficients that are associated with I_i; hence $\mathbf{a} = [\mathbf{a}_1^T,\dots,\mathbf{a}_{N_I}^T]^T$. Although flexible, it leads to $N_a = N_I \times N_g$, which can become a prohibitively large number when $N_I > 1$.

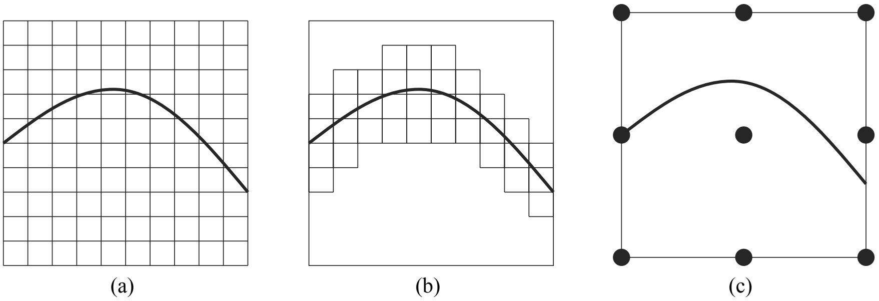

Figure 5.4: Conceptual illustrations of (a) the straightforward, (b) the narrow-band, and (c) the interpolation representation of an arbitrary I. The solid line indicates I^0.

An alternative to the straightforward representation is the narrow-band representation, first introduced in [41] and analyzed extensively in [3]. Here, only the parameters, a^i_k, associated with grid cells in the vicinity of I^0_i are allowed to change (see Figure 5.4b, where only the narrow-band has been illustrated). Some difficulties are associated with the narrow-band representation. First, selecting the number of grid cells in the vicinity of I^0_i is not trivial. Second, re-initialization of the bands of grid cells around the I^0_i's must be done each time the I^0_i's reach the edge the bands, which can be a cumbersome procedure.

Instead of representing the LS functions on the numerical model grid, interpolation techniques can be used. The LS functions are then given on the form

$$I_i(\mathbf{r};\mathbf{a}_i) = \sum_{k=1}^{N^i_a} a^i_k \theta^i_k(\mathbf{r}). \tag{5.19}$$

Here, $\left\{\theta^i_k\right\}_{k=1}^{N^i_a}$ denotes a set of basis functions, which are given on a parameter grid (typically detached from the numerical grid), and $\mathbf{a}_i = [a^i_1, \ldots, a^i_{N^i_a}]^T$ is the associated coefficient vector (see Figure 5.4c). Note that the total number of associated coefficients is $N_a = \sum_{i=1}^{N_I} N^i_a$.

In Papers A, B, and C, we followed [25, 26] and used the simplest interpolation technique – bilinear interpolation. In bilinear interpolation, the parameter grid consists of non-overlapping rectangles, where, in finite-element fashion, the cell area are denoted elements and cell corners are denoted nodes. The basis function θ^i_k is given as a normalized piecewise bilinear function with support on the parameter elements adjacent to node no. k. A normalized bilinear function with support on an arbitrary element, $[x_1, x_2] \times [z_1, z_2]$, is given by

$$((x_2 - x_1)(z_2 - z_1))^{-1} \begin{cases} (x_2 - x)(z_2 - z), \\ (x - x_1)(z_2 - z), \\ (x - x_1)(z - z_1), \\ (x_2 - x)(z - z_1). \end{cases} \tag{5.20}$$

With the bilinear representation a minimum of four nodes are needed to represent I_i, and increasing the number of nodes allows for a more detailed representation. In [25] it was shown (for a reservoir estimation problem) that using too many nodes can become problematic in an estimation procedure. With a high number of nodes, we can have a scenario where I_i has oscillating perturbations, which can lead to difficulties in the estimation problem.

For other representations of LS functions using interpolation techniques see [23] where θ^i_k is given as a B-spline, and [166] where θ^i_k is given as a T-spline (a generalization of non-uniform rational B-splines).

5.2.2 Representation of expansion functions

In this section, we consider the representation of the expansion functions $\mathbf{c}(\mathbf{r})$. The expansion functions determines if $q(\mathbf{r};\mathbf{w},\mathbf{a})$ in a particular region should be homogeneous or heterogeneous. In the homogeneous case, $\mathbf{c}$ is a vector consisting of one constant per region, hence, $\mathbf{c}$ is independent of $\mathbf{r}$, and $\mathbf{w}$ can be neglected (this was assumed in Paper A). In the heterogeneous case, the representation of $\mathbf{c}(\mathbf{r};\mathbf{w})$ can be given in the same manner as (5.19) for the LS functions,

$$c_j(\mathbf{r};\mathbf{w}_j) = \sum_{k=1}^{N_w^j} w_k^j \vartheta_k^j(\mathbf{r}). \tag{5.21}$$

Here, $\left\{\vartheta_k^j\right\}_{k=1}^{N_w^j}$ are interpolation basis functions with $\mathbf{w}_j = [w_1^j, \ldots, w^j_{N_w^j}]^T$ being the associated coefficient vector. The associated coefficient vectors for all regions are gathered in the vector $\mathbf{w} = [\mathbf{w}_1^T, \ldots, \mathbf{w}_{N_w}^T]^T$. The total number of associated coefficients is $N_w = \sum_{j=1}^{N_c} N_w^j$.

In Paper C, $\mathbf{c}(\mathbf{r};\mathbf{w})$ was given as a bilinear interpolation function, i.e., ϑ_k^j was a normalized piecewise bilinear function (see (5.20)). Hence, $q(\mathbf{r};\mathbf{w},\mathbf{a})$ was a smoothly varying function within each region.

5.3 Parameter estimation and multiscale approaches

Recall from Chapter 3 that the model parameters were gathered in a vector denoted $\mathbf{m}$. With the reparameterization of q by the coefficients $\mathbf{w}$ and $\mathbf{a}$, the model parameter vector is given as $\mathbf{m} = [\mathbf{w}^T, \mathbf{a}^T]^T$. Thus, the inverse problem can in this case be stated as: use a set of observed responses from a physical system to identify both the structure of the regions and parameter value within each region. If the inverse problem is solved by a classical approach, the optimization problem can be stated (confer (3.5))

$$\underset{\mathbf{w},\mathbf{a}}{\arg\min}\ (\mathbf{g}(\mathbf{w},\mathbf{a}) - \mathbf{d})^T \mathbf{C}_d^{-1}(\mathbf{g}(\mathbf{w},\mathbf{a}) - \mathbf{d}). \tag{5.22}$$

If some type of regularization term is added, the optimization problem is given by

$$\underset{\mathbf{w},\mathbf{a}}{\arg\min}\ (\mathbf{g}(\mathbf{w},\mathbf{a}) - \mathbf{d})^T \mathbf{C}_d^{-1}(\mathbf{g}(\mathbf{w},\mathbf{a}) - \mathbf{d}) + J_{reg}(\mathbf{w},\mathbf{a}), \tag{5.23}$$

where $J_{reg}(\mathbf{w},\mathbf{a})$ can, e.g., be a Tikhonov regularization term (see Section 3.2.3) or shape prior regularization term (see Section 6.3). Note that $\mathbf{g}(\mathbf{w},\mathbf{a})$ is a shorthand notation for $\mathbf{g}(q(\mathbf{r};\mathbf{w},\mathbf{a}))$.

If, on the other hand, the Bayesian approach is used, the parameter estimation problem is stated as (confer (3.30))

$$f(\mathbf{w},\mathbf{a}|\mathbf{d}) \propto f(\mathbf{d}|\mathbf{w},\mathbf{a})f(\mathbf{w},\mathbf{a}). \tag{5.24}$$

The parameter estimation problem can also be stated as sequential Bayesian problem following the description given in Section 3.3.2.

In the last two solution procedures, (5.23) and (5.24), regularization was applied by adding terms to penalize solutions in some sense. However, the choice of reparameterization can also regularize an inverse problem. The idea is to identify a low-dimensional representation of the 'true' parameter function (denoted q_{true}) using q given by (5.17). This regularization method is called reduced parameterization (alternatively, regularization by projection or regularization by discretization, see, e.g., [60]). By decreasing the number of degrees of freedom in the sought parameter function, a stable estimate may be obtained. The effect of reducing the number of degrees of freedom was studied in, e.g., [74].

Choosing the dimensionality of the reduced representation requires some kind of *a priori* knowledge, e.g., of the resolution power of the observed data, or about q_{true} itself. Without such *a priori* knowledge, different strategies can be employed to solve the parameter estimation problem. A well-known strategy is the multiscale approach. With q given by (5.17), two multiscale approaches can be employed, either in conjuction, or separately.

The first strategy is related to the problem of *a priori* choosing the number of regions. To solve this problem, we start with a few number of regions (typically one region) and, subsequently, refine the number of regions during the estimation process. An example of an application of this approach can be found in [24, 102]. Here, an adaptive refinement strategy based on [75] was combined with a LS corrector.

The second strategy is related to the problem of determining the fine-scale variations of the model, that is, determining the resolution of $\mathbf{c}(\mathbf{r};\mathbf{w})$ and $\mathbf{\Phi}(\mathbf{r};\mathbf{a})$. The multiscale strategy is to start with a coarse grid for $\mathbf{w}$ and $\mathbf{a}$, and, subsequently, refine to get fine-scale variations in both the parameter value within each region, given by $\mathbf{c}(\mathbf{r};\mathbf{w})$, and the region structure, given by $\mathbf{\Phi}(\mathbf{r};\mathbf{a})$. An example of an application of this strategy for $\mathbf{\Phi}(\mathbf{r};\mathbf{a})$ can be found in [26, 101]. Here, an adaptive refinement procedure for $\mathbf{a}$ using a LS representation was applied, again based on [75].

Chapter 6

Kernel methods

Extracting the most of the information contained in a set of data is an important part of science. In this section, we will refer to data in a wide sense, that is, it can be the output of any observation, measurement, or recording device [143]. In many applications, extraction of information, or *features*, from the data cannot be well described by linear algorithms. However, in the middle of the 1990s it was discovered that nonlinear features could be extracted using theory on *kernel functions*. Kernel functions had already been widely applied in functional analysis, dating back to the early 1900s. The term nonlinear features can (loosely speaking) be translated to features that can only be related through some nonlinear function. In statistics, data with nonlinear features follow a non-Gaussian probability distribution. In any case, kernel methods allow for extraction of nonlinear features by transforming the data to a *feature space* and using linear algorithms in this new space.

In this chapter, we will start by introducing kernels and present some of the fundamental theory. Most of the theory will follow the textbooks [140, 143]. Subsequently, we will discuss some applications that take advantage of the theory on kernel functions.

6.1 Kernel functions and their properties

Throughout this chapter, we will consider the mapping

$$\phi : \mathbf{x} \in \mathbb{R}^n \mapsto \phi(\mathbf{x}) \in \mathcal{Y}, \tag{6.1}$$

where $\mathcal{Y}$ is a $N_{\mathcal{Y}}$-dimensional space, denoted feature space, and ϕ is a nonlinear function, denoted feature map. Using the mapping (6.1), we aim to convert nonlinear features in the data to linear ones; hence, enabling the use of linear algorithms to extract features in $\mathcal{Y}$. This, however, poses two major difficulties: first, to convert nonlinear features to linear features, $\mathcal{Y}$ must usually be a very high-dimensional space (in some cases, it must be infinite dimensional). Second, an explicit expression for ϕ is generally unknown.

Instead of relying on a direct measure in $\mathcal{Y}$ via ϕ, we can instead use some indirect measure to extract the information we need from $\mathcal{Y}$. A simple, yet powerful indirect measure is the inner product (sometimes also denoted scalar product or dot product),

$$\langle \cdot, \cdot \rangle : \mathcal{Y} \times \mathcal{Y} \mapsto \mathbb{R}. \tag{6.2}$$

The inner product is an operation yielding information on how similar two vectors are in $\mathcal{Y}$. Moreover, the inner product yields geometrical information on the length of a vector, and the distance between two vectors (which we will exploit later when discussing dissimilarity).

The inner product still requires the computation of ϕ. To circumvent the explicit dependence on ϕ, we introduce a function $k : \mathbb{R}^n \times \mathbb{R}^n \mapsto \mathbb{R}$,

$$k(\mathbf{x},\mathbf{y}) = \langle \phi(\mathbf{x}), \phi(\mathbf{y}) \rangle. \tag{6.3}$$

The function k is denoted kernel function or just kernel. What remains now is to see what kind of kernel functions admits a representation of the form (6.3); that is, given a kernel, can we construct a $\mathcal{Y}$ with an associated ϕ such that (6.3) holds? In the following sections, we will present two theories from functional analysis that answers this question. First, we define some properties of k.

6.1.1 Positive definite kernels

Let a set of data be given as $\mathbf{t}^1, \ldots, \mathbf{t}^m \in \mathcal{X}$. Note that $\mathcal{X}$ is a general set where $\mathbb{R}^n$ is a special case. The $m \times m$ matrix $\mathbf{K}$ given as

$$K_{ij} = k(\mathbf{t}^i, \mathbf{t}^j) \tag{6.4}$$

is denoted kernel matrix. A kernel function is said to be positive (semi-) definite (PD) iff $\mathbf{K}$ is PD, i.e., $\mathbf{s}^T\mathbf{K}\mathbf{s} \geq 0, \forall \mathbf{s} \in \mathbb{R}^m$. Furthermore, k is a symmetric function iff $\mathbf{K}$ is symmetric, i.e., $\mathbf{K} = \mathbf{K}^T$. Note that, PD and symmetry for k must hold for all choices of $\mathbf{t}^i$. When there is no conflicting notation, we will refer to symmetric PD kernels as just kernels. Note that we restrict our study to real-valued kernels.

In the following section, we will need the Cauchy-Schwarz inequality for kernels, which is given by

$$|k(\mathbf{x},\mathbf{y})|^2 \leq k(\mathbf{x},\mathbf{x})k(\mathbf{y},\mathbf{y}). \tag{6.5}$$

Proof can be found in [140].

6.1.2 Reproducing kernel Hilbert space

The goal of this section is to construct a reproducing kernel Hilbert space (RKHS) and prove that (6.3) holds true in this space. To do so, we define a pre-Hilbert space, or strict inner-product space, that, when completed, will define a RKHS. For a more rigorous derivation of RKHS see, e.g., [22].

We start by defining the vector space

$$\mathcal{Y}_0 = \left\{ \sum_{i=1}^{m} \xi_i k(\mathbf{t}^i, \cdot) \ : \ \xi_i \in \mathbb{R}, m \in \mathbb{N}, \mathbf{t}^i \in \mathcal{X} \right\}.$$

Note that $\mathbf{t}^1, \ldots, \mathbf{t}^m$ are arbitrary. Let $f, g \in \mathcal{Y}_0$ be given as

$$f(\cdot) = \sum_{i=1}^{m} \xi_i k(\mathbf{t}^i, \cdot), \quad g(\cdot) = \sum_{j=1}^{m'} \eta_j k(\mathbf{u}^j, \cdot). \tag{6.6}$$

Now, define the operation $\langle\cdot,\cdot\rangle$ on $\mathcal{H}_0 \times \mathcal{H}_0$ as

$$\langle f,g\rangle = \sum_{i=1}^{m}\sum_{j=1}^{m'} \xi_i\eta_j k(\mathbf{t}^i,\mathbf{u}^j) = \sum_{j=1}^{m'} \eta_j f(\mathbf{u}^j) = \sum_{i=1}^{m} \xi_i g(\mathbf{t}^i). \tag{6.7}$$

The two last equalities follow from (6.6) and the fact that k is symmetric. Hence, by definition, $\langle\cdot,\cdot\rangle$ is also symmetric, i.e., $\langle f,g\rangle = \langle g,f\rangle$. Moreover, the two last equalities show that $\langle f,g\rangle$ does not depend on the particular expansion of f and g, and also that $\langle\cdot,\cdot\rangle$ is bilinear (e.g., $\langle f+h,g\rangle = \langle f,g\rangle + \langle h,g\rangle$, which follows from the second equality and the fact that $\mathcal{H}_0$ is a vector space.) Since k is PD, then

$$\langle f,f\rangle = \sum_{i,j=1}^{m} \xi_i\xi_j k(\mathbf{t}^i,\mathbf{t}^j) \geq 0, \tag{6.8}$$

thus $\langle\cdot,\cdot\rangle$ is also PD.

Hence, we only need to show that $\langle f,f\rangle = 0$ implies $f = 0$ to show that $\langle\cdot,\cdot\rangle$ is a strict inner product. To this end, we make two observations. The first observation follows directly from (6.6) and (6.7),

$$\langle f,k(\mathbf{x},\cdot)\rangle = f(\mathbf{x}). \tag{6.9}$$

A kernel fulfilling (6.9) is called a reproducing kernel. In particular, we have

$$\langle k(\mathbf{x},\cdot),k(\mathbf{y},\cdot)\rangle = k(\mathbf{x},\mathbf{y}). \tag{6.10}$$

The second observation is that $\langle\cdot,\cdot\rangle$ is actually itself a kernel. This follows from $\langle\cdot,\cdot\rangle$ being bilinear and PD. To see this, note that for any functions $f_1,\ldots,f_n$ and coefficients $c_1,\ldots,c_n \in \mathbb{R}$ we have

$$\sum_{i,j=1}^{n} c_ic_j\langle f_i,f_j\rangle = \left\langle \sum_{i=1}^{n} c_if_i, \sum_{i=j}^{n} c_jf_j \right\rangle \geq 0, \tag{6.11}$$

where the equality follows from bilinearity of $\langle\cdot,\cdot\rangle$, and the inequality follows from (6.8). From the definition of a kernel in Section 6.1.1, we thus see that $\langle\cdot,\cdot\rangle$ is a kernel.

Since $\langle\cdot,\cdot\rangle$ is a kernel, we can use the Cauchy-Schwartz inequality (6.5) and the reproducing property (6.9) to get

$$|f(\mathbf{x})|^2 = |\langle f,k(\mathbf{x},\cdot)\rangle|^2 \leq |\langle f,f\rangle| k(\mathbf{x},\mathbf{x}), \tag{6.12}$$

from which $\langle f,f\rangle = 0$ directly implies $f = 0$. Since $\langle\cdot,\cdot\rangle$ is a strict inner product on $\mathcal{H}_0$, then $\mathcal{H}_0$ is a pre-Hilbert space. By standard approaches, $\mathcal{H}_0$ can be turned into a Hilbert space $\mathcal{H}$ (with the norm $||f|| = \sqrt{\langle f,f\rangle}$). A Hilbert space with the reproducing property is denoted RKHS. Note that the RKHS uniquely determines a kernel.

Letting $\phi : \mathcal{X} \mapsto \mathbb{R}^{\mathcal{X}}$, where $\mathbb{R}^{\mathcal{X}}$ is the space of functions mapping $\mathcal{X}$ to $\mathbb{R}$, be given as

$$\phi(\mathbf{x}) = k(\mathbf{x},\cdot), \tag{6.13}$$

we see that the relation (6.3) follows directly from (6.10). Hence, we have shown that a kernel infers a RKHS. The converse also holds, that is, if ϕ is a feature map to a RKHS, then $k(\mathbf{x},\mathbf{y})$ given by (6.3) is PD. To this end, $\forall s_i \in \mathbb{R}$ and $\mathbf{t}^1,\ldots,\mathbf{t}^m \in \mathcal{X}$ we have

$$\sum_{i,j=1}^{m} s_i s_j k(\mathbf{t}^i,\mathbf{t}^j) = \sum_{i,j=1}^{m} s_i s_j \langle \phi(\mathbf{t}^i), \phi(\mathbf{t}^j)\rangle = \left\langle \sum_{i=1}^{m} s_i \phi(\mathbf{t}^i), \sum_{j=1}^{m} s_j \phi(\mathbf{t}^j) \right\rangle$$
$$= \left\| \sum_{i=1}^{m} s_i \phi(\mathbf{t}^i) \right\|^2 \geq 0, \tag{6.14}$$

where the inequality follows from the nonnegativity of the norm.

In conclusion, we have shown that kernels are functions with the property that there exist a feature map ϕ to a $\mathcal{Y}$ such that (6.3) holds.

6.1.3 The Mercer kernel map

In this section, we construct an alternative RKHS from the one presented in the previous section, which is commonly the way (6.3) is justified in various applications. Assume now that k is a continuous kernel, and that $\mathcal{X}$ is a compact metric space with a finite measure ν. We can then define an integral operator $T_k : L_2(\mathcal{X}) \mapsto L_2(\mathcal{X})$:

$$(T_k f)(\mathbf{x}) = \int_{\mathcal{X}} k(\mathbf{x},\mathbf{y}) f(\mathbf{y}) \, \mathrm{d}\nu(\mathbf{y}), \tag{6.15}$$

where $L_2(\mathcal{X})$ is the the space of square integrable functions. Since k is symmetric, T_k is self-adjoint, i.e., $\langle f, T_k g\rangle = \langle T_k f, g\rangle$. Moreover, since k is PD, T_k is a positive operator, i.e., $\langle f, T_k f\rangle \geq 0$, and since k is continuous, T_k is compact. Since T_k is a self-adjoint positive compact operator, it can, by the spectral theorem be decomposed into a sum of eigenvalues $\zeta_j > 0$ and orthonormal eigenfunctions $e_j \in L_2(\mathcal{X})$, where $j = 1,\ldots,N_{\mathcal{Y}}$ with $N_{\mathcal{Y}} = \mathbb{N}$ or $N_{\mathcal{Y}} = \infty$.

We are now ready to state Mercer' theorem (without proof) [112].

Theorem 1 *Let k be a continuous kernel on $\mathcal{X}$. Then $\forall \mathbf{x},\mathbf{y} \in \mathcal{X}$*

$$k(\mathbf{x},\mathbf{y}) = \sum_{j=1}^{N_{\mathcal{Y}}} \zeta_j e_j(\mathbf{x}) e_j(\mathbf{y}), \tag{6.16}$$

which converges absolutely and uniformly for $N_{\mathcal{Y}} = \infty$.

The relation (6.3) follows from Mercer's theorem by letting $\phi : \mathcal{X} \mapsto l_2^{N_{\mathcal{Y}}}$ be given as

$$\phi(\mathbf{x}) = \left\{ \sqrt{\zeta_j} e_j(\mathbf{x}) \right\}_{j=1}^{N_{\mathcal{Y}}} \tag{6.17}$$

where $l_2^{N_{\mathcal{Y}}}$ is a $N_{\mathcal{Y}}$-dimensional space of square summable sequences.

It is now possible to construct a RKHS in the same manner as in Section 6.1.2. Omitting the details, the RKHS is given by

$$\mathcal{Y} = \left\{ \sum_{j=1}^{N_{\mathcal{Y}}} \xi_j e_j \;:\; \frac{\xi_j}{\sqrt{\zeta_j}} \in l_2^{N_{\mathcal{Y}}} \right\}, \tag{6.18}$$

with an inner product

$$\left\langle \sum_{j=1}^{N_{\mathcal{Y}}} \xi_j e_j, \sum_{j=1}^{N_{\mathcal{Y}}} \eta_j e_j \right\rangle = \sum_{j=1}^{N_{\mathcal{Y}}} \frac{\xi_j \eta_j}{\zeta_j}. \tag{6.19}$$

(The inner product is chosen such that $\langle e_i, e_j \rangle = \delta_{ij}/\zeta_j$, where δ_{ij} is the Kroeneker delta.)

The reproducing property is easily checked by letting $f = \sum_{j=1}^{N_{\mathcal{Y}}} \xi_j e_j$ and noticing from Mercer's theorem that $k(\mathbf{x}, \cdot) = \sum_{j=1}^{N_{\mathcal{Y}}} \zeta_j e_j(\mathbf{x}) e_j$,

$$\langle f, k(\mathbf{x}, \cdot) \rangle = \left\langle \sum_{j=1}^{N_{\mathcal{Y}}} \xi_j e_j, \sum_{j=1}^{N_{\mathcal{Y}}} \zeta_j e_j(\mathbf{x}) e_j \right\rangle = \sum_{j=1}^{N_{\mathcal{Y}}} \frac{\xi_j \zeta_j e_j(\mathbf{x})}{\zeta_j} = f(\mathbf{x}). \tag{6.20}$$

In conclusion, kernels made by Mercer's theorem are also reproducing kernels, and, hence, are associated with a unique RKHS. Moreover, we showed above that the relation (6.3) is ensured by Mercer's theorem.

6.1.4 Kernel trick

The results in the preceding sections justified the relation (6.3) provided that the kernel was real-valued, symmetric, and PD. We can now proceed to the most important consequence of (6.3), the kernel trick: *Any algorithm expressed in terms of inner products can instead be expressed in terms of kernels.* Consequently, we can transform our data from $\mathbb{R}^n$ to $\mathcal{Y}$ and use standard linear algebra algorithms involving inner products to extract nonlinear features.

The kernel trick does not tell us what kernel to use for each application. Other than a kernel needing to be symmetric and PD, there is usually no restriction on which kernel to use for an application. However, in some applications, there exist theoretical proofs that narrows down the choice of valid kernels. In other applications, a kernel can be constructed using simpler kernels as building blocks, see, e.g., [143, section 3.4]. Often, however, a kernel is chosen by trail-and-error, or based on user experience.

It is important to stress that the choice of kernel implicitly defines the feature space $\mathcal{Y}$, or more specific, it defines how the data is related in $\mathcal{Y}$. Hence, to extract the most information from the data, it is important to choose the kernel wisely.

The above kernel trick relies on the notion of similarity (through the inner product in (6.3)). It turns out that there exist an equivalent kernel trick which is based on dissimilarity, that is, there exists a relation between a kernel and a distance measure. These types of kernels will be important in the applications described in the sections below, thus we introduce their theory in the next section.

6.1.5 Conditionally positive definite kernels

The purpose of this section is to establish a similar relation to (6.3), but now between a distance measure and a kernel. The kernels we will be discussing belongs to a large class than PD kernels. They are conditionally positive definite (CPD) kernels. Since we will encounter both PD and CPD kernels in this section, we denote a PD kernel with k_{PD} and CPD kernel with k_{CPD}.

Let **1** be a m-vector where each element is equal 1. A kernel function is said to be CPD iff $\mathbf{K}$ is CPD, i.e., $\mathbf{s}^T\mathbf{K}\mathbf{s} \geq 0$, $\forall \mathbf{s} \in \mathbb{R}^m$ with $\mathbf{s}^T\mathbf{1} = 0$. Note that a PD kernel is also CPD.

Two formulas will be important for relating CPD kernels and distances in $\mathcal{Y}$. First, a squared distance in $\mathcal{Y}$ for $\mathbf{x}, \mathbf{y} \in \mathcal{X}$ is given by

$$\begin{aligned} \|\phi(\mathbf{x}) - \phi(\mathbf{y})\|^2 &= \langle\phi(\mathbf{x}),\phi(\mathbf{x})\rangle - 2\langle\phi(\mathbf{x}),\phi(\mathbf{y})\rangle + \langle\phi(\mathbf{y}),\phi(\mathbf{y})\rangle, \\ &= k_{PD}(\mathbf{x},\mathbf{x}) - 2k_{PD}(\mathbf{x},\mathbf{y}) + k_{PD}(\mathbf{y},\mathbf{y}), \end{aligned} \tag{6.21}$$

where the last equality follows from (6.3). This tells us that there is a connection between squared distances in $\mathcal{Y}$ and PD kernels. A connection between PD and CPD kernels is given by the next formula [21]:

$$k_{PD}(\mathbf{x},\mathbf{y}) = \frac{1}{2}(k_{CPD}(\mathbf{x},\mathbf{y}) - k_{CPD}(\mathbf{x},\mathbf{x}_0) - k_{CPD}(\mathbf{x}_0,\mathbf{y}) + k_{CPD}(\mathbf{x}_0,\mathbf{x}_0)), \tag{6.22}$$

where $\mathbf{x}_0 \in \mathcal{X}$ is arbitrary.

Combining (6.21) and (6.22) yields the Hilbert space representation of CPD kernels: If $k : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$ is a CPD kernel, then there exists a Hilbert space with mapping $\phi : \mathcal{X} \mapsto \mathcal{Y}$ such that

$$k_{CPD}(\mathbf{x},\mathbf{y}) - \frac{1}{2}(k_{CPD}(\mathbf{x},\mathbf{x}) + k_{CPD}(\mathbf{y},\mathbf{y})) = -\|\phi(\mathbf{x}) - \phi(\mathbf{y})\|^2. \tag{6.23}$$

In particular, if $k_{CPD}(\mathbf{x},\mathbf{x}) = k_{CPD}(\mathbf{y},\mathbf{y}) = 0$ and $k_{CPD}(\mathbf{x},\mathbf{y}) \neq 0$ for $\mathbf{x} \neq \mathbf{y}$, then

$$k_{CPD}(\mathbf{x},\mathbf{y}) = -\|\phi(\mathbf{x}) - \phi(\mathbf{y})\|^2. \tag{6.24}$$

Thus, we have showed that k_{CPD} is the negative of a distance measure in $\mathcal{Y}$ (see, e.g., [21] for a rigorous proof).

For the applications discussed later, a generalization of (6.22) will be useful. First, we note that (6.22) is itself a generalization. Consider $\mathbf{x}, \mathbf{y}, \mathbf{x}_0 \in \mathcal{X}$, and the inner product in $\mathcal{X}$. The effect of a translation of the data $\mathbf{x}$ and $\mathbf{y}$ by $\mathbf{x}_0$ for the inner product can be expressed as

$$\langle\mathbf{x} - \mathbf{x}_0, \mathbf{y} - \mathbf{x}_0\rangle = \frac{1}{2}(-\|\mathbf{x} - \mathbf{y}\|^2 + \|\mathbf{x} - \mathbf{x}_0\|^2 + \|\mathbf{x}_0 - \mathbf{x}\|^2). \tag{6.25}$$

Noticing that $-\|\mathbf{x} - \mathbf{y}\|^2$ is a CPD kernel and $\langle\mathbf{x} - \mathbf{x}_0, \mathbf{y} - \mathbf{x}_0\rangle$ is a PD kernel, we see that (6.25) follows from (6.22). Moreover, in light of (6.24) (or (6.23)) we can infer that (6.22) also must hold true in the general case. This means that k_{PD} in (6.22) essentially describes a translation with respect to a point $\phi(\mathbf{x}_0)$ in $\mathcal{Y}$. If we want to translate with respect to other types of orgins, e.g., the mean of some mapped data set, $\phi(\mathbf{t}^1), \ldots, \phi(\mathbf{t}^m) \in \mathcal{Y}$, then the following generalization of (6.22) is advantageous. Let $\mathbf{t}^1, \ldots, \mathbf{t}^m \in \mathcal{X}$; $\mathbf{x}, \mathbf{y} \in \mathcal{X}$; and $\gamma_i \in \mathbb{R}$ where $\sum_{i=1}^m \gamma_i = 1$. Then we have the following relation [140]

$$\begin{aligned} k_{PD}(\mathbf{x},\mathbf{y}) = k_{CPD}(\mathbf{x},\mathbf{y}) &- \sum_{i=1}^m \gamma_i k_{CPD}(\mathbf{x},\mathbf{t}^i) - \sum_{i=1}^m \gamma_i k_{CPD}(\mathbf{t}^i,\mathbf{y}) \\ &+ \sum_{i,j=1}^m \gamma_i\gamma_j k_{CPD}(\mathbf{t}^i,\mathbf{t}^j). \end{aligned} \tag{6.26}$$

6.2 Kernel principal component analysis

Principal component analysis (PCA) is a powerful and widely used method for reducing the dimensionality of a data set. The method is an orthogonal linear transform of coordinates where the data set is projected onto a new coordinate system. The projected data form a new set of uncorrelated variables, denoted principal components (PC), where each PC reflects the variability of the data. Hence, by ordering the PC in decreasing order, the first few PCs will represent the most of the variability in the data.

Mathematically, the objective in PCA can be stated as follows: given a set of data $\mathbf{t}^1, \ldots, \mathbf{t}^m \in \mathbb{R}^n$, find a $n \times n$ matrix $\mathbf{V}$ of unit vectors $\mathbf{v}^1, \ldots, \mathbf{v}^n$ that determines the change of variables from $\mathbf{x}$ to a new set of variables

$$\mathbf{y} = \mathbf{V}^T\mathbf{x}, \tag{6.27}$$

such that $\mathbf{y} = [y_1, \ldots, y_n]^T$ are ordered by decreasing variance and are uncorrelated. It turns out that $\mathbf{v}^1, \ldots, \mathbf{v}^n$ are easily found from the sample covariance matrix,

$$\mathbf{C}_t^e = \frac{1}{m}\sum_{i=1}^{m}(\mathbf{t}^i - \bar{\mathbf{t}})(\mathbf{t}^i - \bar{\mathbf{t}})^T, \tag{6.28}$$

where $\bar{\mathbf{t}}$ denotes the sample mean,

$$\bar{\mathbf{t}} = \frac{1}{m}\sum_{i=1}^{m}\mathbf{t}^i. \tag{6.29}$$

By performing an eigenvalue decomposition of $\mathbf{C}_t^e$, we obtain the unit vectors $\mathbf{v}^1, \ldots, \mathbf{v}^n$:

$$\lambda_k\mathbf{v}^k = \mathbf{C}_t^e\mathbf{v}^k. \tag{6.30}$$

The eigenvalues, $\{\lambda_k\}_{k=1}^n \geq 0$, give the variance of the PCs, and are thus ordered in decreasing order. The PCs are uncorrelated since the eigenvectors $\{\mathbf{v}^k\}_{k=1}^n$ are orthogonal. A full derivation of PCA can be found, e.g., in [87].

The PCs defined by (6.27) give a linear projection of the data set onto $\mathbf{V}$. In some applications, a linear projection will not reveal all the features in the data. For example, if the data is given by multiple clusters in $\mathbb{R}^n$, the standard PCA method will not provide PCs that accounts for such nonlinear features. In this case, a generalization of the PCA method is needed. In [139], the authors provided a nonlinear version of PCA using the kernel relation (6.3); they referred to the new method as kernel PCA. Below we provide the derivation of kernel PCA, which follows directly from the standard PCA (from now denoted linear PCA).

Consider the feature map $\phi : \mathbb{R}^n \mapsto \mathcal{Y}$, and let $\phi(\mathbf{t}^1), \ldots, \phi(\mathbf{t}^m) \in \mathcal{Y}$ denote the mapped data set. The sample mean and covariance matrix is given by

$$\bar{\boldsymbol{\phi}} = \frac{1}{m}\sum_{i=1}^{m}\phi(\mathbf{t}^i), \tag{6.31}$$

and

$$\mathbf{C}_\phi^e = \frac{1}{m}\sum_{i=1}^{m}(\phi(\mathbf{t}^i) - \bar{\boldsymbol{\phi}})(\phi(\mathbf{t}^i) - \bar{\boldsymbol{\phi}})^T, \tag{6.32}$$

respectively. To shorten the notation we let $\tilde{\phi}(\mathbf{x}) = \phi(\mathbf{x}) - \bar{\boldsymbol{\phi}}$ in the following. Recall that $N_{\mathcal{Y}}$ is large (possibly infinite) and ϕ is generally unknown, hence, a direct eigenvalue decomposition of $\mathbf{C}^e_\phi$ is not possible. To proceed, we inserting (6.32) into the eigenvalue decomposition of $\mathbf{C}^e_\phi$,

$$\lambda_k \mathbf{v}^k = \mathbf{C}^e_\phi \mathbf{v}^k = \frac{1}{m} \sum_{i=1}^{m} \langle \tilde{\phi}(\mathbf{t}^i), \mathbf{v}^k \rangle \tilde{\phi}(\mathbf{t}^i). \tag{6.33}$$

Rearranging gives

$$\mathbf{v}^k = \sum_{i=1}^{m} \frac{\langle \tilde{\phi}(\mathbf{t}^i), \mathbf{v}^k \rangle}{m\lambda_k} \tilde{\phi}(\mathbf{t}^i) = \sum_{i=1}^{m} \alpha_i^k \tilde{\phi}(\mathbf{t}^i), \tag{6.34}$$

where

$$\alpha_i^k = \frac{\langle \tilde{\phi}(\mathbf{t}^i), \mathbf{v}^k \rangle}{m\lambda_k}. \tag{6.35}$$

Consequently, $\mathbf{v}^k$ lies in the span of $\{\tilde{\phi}(\mathbf{t}^1), \ldots, \tilde{\phi}(\mathbf{t}^m)\}$. Hence, we can consider the equivalent eigenvalue decomposition of $\mathbf{C}^e_\phi$ given by

$$\langle \tilde{\phi}(\mathbf{t}^l), \lambda_k \mathbf{v}^k \rangle = \langle \tilde{\phi}(\mathbf{t}^l), \mathbf{C}^e_\phi \mathbf{v}^k \rangle \tag{6.36}$$

Inserting (6.32) and (6.34) into (6.36) leads to

$$\lambda_k \sum_{i=1}^{m} \alpha_i^k \langle \tilde{\phi}(\mathbf{t}^l), \tilde{\phi}(\mathbf{t}^i) \rangle = \frac{1}{m} \sum_{j=1}^{m} \sum_{i=1}^{m} \alpha_i^k \langle \tilde{\phi}(\mathbf{t}^l), \tilde{\phi}(\mathbf{t}^j) \rangle \langle \tilde{\phi}(\mathbf{t}^j), \tilde{\phi}(\mathbf{t}^i) \rangle. \tag{6.37}$$

From (6.3) and (6.4) we can define $\widetilde{K}_{ij} = \tilde{k}(\mathbf{t}^i, \mathbf{t}^j) = \langle \tilde{\phi}(\mathbf{t}^i), \tilde{\phi}(\mathbf{t}^j) \rangle$. Moreover, we let $\boldsymbol{\alpha}^k = [\alpha_1^k, \ldots, \alpha_m^k]^T$ and $l = 1, \ldots, m$. Inserted into (6.37) yields

$$m\lambda_k \widetilde{\mathbf{K}} \boldsymbol{\alpha}^k = \widetilde{\mathbf{K}}^2 \boldsymbol{\alpha}^k, \tag{6.38}$$

or, since $\widetilde{\mathbf{K}}$ is a symmetric PD matrix

$$\tilde{\lambda}_k \boldsymbol{\alpha}^k = \widetilde{\mathbf{K}} \boldsymbol{\alpha}^k, \tag{6.39}$$

where $\tilde{\lambda}_k = m\lambda_k$. This is just the expression for an eigenvalue decomposition of $\widetilde{\mathbf{K}}$.

In summary, we have reduced the eigenvalue decomposition of $\mathbf{C}^e_\phi$ to the eigenvalue decomposition of $\widetilde{\mathbf{K}}$. Note that since $\mathbf{v}^k$ is a unit vector, $\boldsymbol{\alpha}^k$ needs to be normalized [139],

$$\langle \mathbf{v}^k, \mathbf{v}^k \rangle = 1 \quad \Rightarrow \quad \tilde{\lambda}_k \langle \boldsymbol{\alpha}^k, \boldsymbol{\alpha}^k \rangle = 1. \tag{6.40}$$

Finally, the PCs in kernel PCA are given in a similar manner as in linear PCA (see (6.27)):

$$y_k = \langle \mathbf{v}^k, \tilde{\phi}(\mathbf{x}) \rangle = \sum_{i=1}^{m} \alpha_i^k \langle \tilde{\phi}(\mathbf{x}), \tilde{\phi}(\mathbf{t}^i) \rangle = \sum_{i=1}^{m} \alpha_i^k \tilde{k}(\mathbf{x}, \mathbf{t}^i), \tag{6.41}$$

for $k = 1, \ldots, m$. Since the PCs are given in terms of a kernel, which generally is a nonlinear function, we denote y_k as nonlinear PCs.

A comment on $\widetilde{\mathbf{K}}$ is needed. $\widetilde{\mathbf{K}}$ was given in terms of a kernel function $\tilde{k}$, which we will denote the centered kernel. Since we do not know $\tilde{\phi}(\mathbf{x})$, we cannot compute $\widetilde{\mathbf{K}}$ directly. However, from the definition of $\tilde{k}$ and $\tilde{\phi}$, and using (6.3) it follows that [139]

$$\begin{aligned}
\widetilde{K}_{ij} &= \langle \tilde{\phi}(\mathbf{t}^i), \tilde{\phi}(\mathbf{t}^j) \rangle = \langle \phi(\mathbf{t}^i) - \frac{1}{m}\sum_{k=1}^{m} \phi(\mathbf{t}^k), \phi(\mathbf{t}^j) - \frac{1}{m}\sum_{l=1}^{m} \phi(\mathbf{t}^l) \rangle, \\
&= \langle \phi(\mathbf{t}^i), \phi(\mathbf{t}^j) \rangle - \frac{1}{m}\left[\sum_{k=1}^{m} \langle \phi(\mathbf{t}^k), \phi(\mathbf{t}^j) \rangle + \sum_{l=1}^{m} \langle \phi(\mathbf{t}^i), \phi(\mathbf{t}^l) \rangle \right] \\
&\quad + \frac{1}{m^2} \sum_{k,l=1}^{m} \langle \phi(\mathbf{t}^k), \phi(\mathbf{t}^l) \rangle, \\
&= k(\mathbf{t}^i, \mathbf{t}^j) - \frac{1}{m}\left[\sum_{k=1}^{m} k(\mathbf{t}^k, \mathbf{t}^j) + \sum_{l=1}^{m} k(\mathbf{t}^i, \mathbf{t}^l) \right] + \frac{1}{m^2} \sum_{k,l=1}^{m} k(\mathbf{t}^k, \mathbf{t}^l). \qquad (6.42)
\end{aligned}$$

Hence, we have expressed $\widetilde{\mathbf{K}}$ in terms of the familiar PD kernel $k(\mathbf{x}, \mathbf{y})$. Note that we can also get the same expression for CPD kernels. Recall from the discussion made at the end of Section 6.1.5, where we noted that data translated to a general origin in $\mathcal{Y}$ could be expressed in kernel form by (6.26). Above, we have used $\tilde{\phi}$ to center the mapped data set to the sample mean $\bar{\boldsymbol{\phi}}$ in $\mathcal{Y}$. Hence, it is readily seen that letting $\gamma_i = 1/m$, $\mathbf{x} = \mathbf{t}^i$, and $\mathbf{y} = \mathbf{t}^j$ in (6.26) leads to (6.42). Thus, CPD kernels are a valid choice in kernel PCA. Using (6.42), $\widetilde{\mathbf{K}}$ can be calculated via $\mathbf{K}$ as

$$\widetilde{\mathbf{K}} = \mathbf{K} - \frac{1}{m}[\mathbf{K}\mathbf{1}\mathbf{1}^T + \mathbf{1}\mathbf{1}^T\mathbf{K}] + \frac{1}{m^2}(\mathbf{1}^T\mathbf{K}\mathbf{1})\mathbf{1}\mathbf{1}^T, \qquad (6.43)$$

where $\mathbf{1}$ is a m-dimensional vector with all entries equal 1.

We note that in the derivation given above for kernel PCA we did not do anything we could not have done for linear PCA. In fact, letting $k(\mathbf{x}, \mathbf{y}) = \langle \mathbf{x}, \mathbf{y} \rangle$ in kernel PCA we get the linear PCA algorithm. Consequently, all mathematical and statistical properties of linear PCA directly applies to kernel PCA. In particular, the first nonlinear PCs correspond to the most of the (nonlinear) variability in the data.

In the next section, we will discuss a method related to kernel PCA; in fact, the method can be viewed as a probabilistic extension of kernel PCA.

6.3 Shape prior regularization

In Section 3.2.3, we discussed Tikhonov regularization, which was introduced to remedy the instability issues in ill-posed inverse problems. Moreover, in Section 3.3, we pointed out that the Gaussian prior model had strong connections with Tikhonov regularization in the sense that deviations from a predetermined prior model was penalized during a MAP solution procedure. In some applications, assuming a Gaussian prior model in the input space $\mathbb{R}^n$ may not be appropriate, that is, a set of data may not be well described by a Gaussian PDF. For example, data that are divided into separate clusters will necessarily lead to a multimodal PDF (see, example given below).

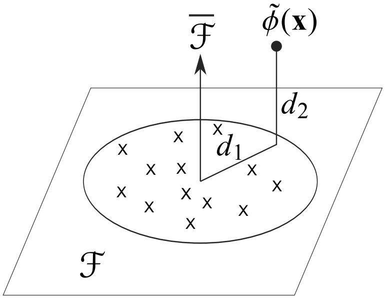

Figure 6.1: Illustration of the decomposition of $\mathcal{Y}$ to $\mathcal{F}$ (kernel PCA space) and $\overline{\mathcal{F}}$ (the orthogonal complement to kernel PCA space). The (squared) distances (6.48) and (6.49) are indicated, together with the mapped data set (×).

The limitation of assuming a Gaussian prior PDF in $\mathbb{R}^n$ was noticed in [50] for the image segmentation problem. They suggested an extension of the standard approach for incorporating prior information by instead assuming a Gaussian PDF in $\mathcal{Y}$. The method was denoted shape prior. In the following, we will give a brief derivation of the method, before we discuss some important properties.

Let $\mathbf{t}^1, \ldots, \mathbf{t}^m \in \mathbb{R}^n$ be a set of data that comprises some prior knowledge about a model under consideration in an inverse problem (in some applications, e.g., machine learning, the $\mathbf{t}^i$'s are called training data). The sample mean, $\overline{\boldsymbol{\phi}}$, and covariance matrix, $\mathbf{C}^e_\phi$, of the mapped data set $\phi(\mathbf{t}^1), \ldots, \phi(\mathbf{t}^m) \in \mathcal{Y}$ is given by (6.31) and (6.32), respectively. Furthermore, recall that $\tilde{\phi}(\mathbf{x}) = \phi(\mathbf{x}) - \overline{\boldsymbol{\phi}}$. The shape prior regularization term is given as the corresponding energy of the Gaussian PDF in $\mathcal{Y}$ estimated from the mapped data set. That is, it is defined as

$$J_{prior}(\mathbf{x}) = \tilde{\phi}(\mathbf{x})^T (\widehat{\mathbf{C}}^e_\phi)^{-1} \tilde{\phi}(\mathbf{x}). \tag{6.44}$$

Note that $\widehat{\mathbf{C}}^e_\phi$ is a regularized covariance matrix [50]

$$\begin{aligned} \widehat{\mathbf{C}}^e_\phi &= \mathbf{C}^e_\phi + \lambda_\perp (\mathbf{I} - \mathbf{V}\mathbf{V}^T), \\ &= \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T + \lambda_\perp (\mathbf{I} - \mathbf{V}\mathbf{V}^T) \end{aligned} \tag{6.45}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix containing the eigenvalues, λ_k, and $\mathbf{V}$ is matrix containing the eigenvectors, $\mathbf{v}^k$, of $\mathbf{C}^e_\phi$, with $k = 1, \ldots, r \leq m$ (that is, there are at most $r \leq m$ positive eigenvalues of $\mathbf{C}^e_\phi$). In most cases $r << N_\mathcal{Y}$, which results in $\mathbf{C}^e_\phi$ not having full rank and thus is not invertible. Hence, we add the term $\lambda_\perp(\mathbf{I} - \mathbf{V}\mathbf{V}^T)$ to replace the $r + 1, \ldots, N_\mathcal{Y}$ zero eigenvalues of $\mathbf{C}^e_\phi$ by $\lambda_\perp \in (0, \lambda_r)$. To see that this is valid, note that if $r \geq N_\mathcal{Y}$ then $\mathbf{V}\mathbf{V}^T = \mathbf{I}$ and $\widehat{\mathbf{C}}^e_\phi = \mathbf{C}^e_\phi$. For $r < N_\mathcal{Y}$, we have

$$\mathbf{V}\mathbf{V}^T + \mathbf{V}_0\mathbf{V}_0^T = \mathbf{I}, \tag{6.46}$$

where $\mathbf{V}_0\mathbf{V}_0^T$ denote the eigenvectors corresponding to the $r + 1, \ldots, N_\mathcal{Y}$ zero eigenvalues. Rearranging (6.46) and multiplying by $\lambda_\perp$ yields the second term in (6.45).

Inserting (6.45) into (6.44), and writing it as a series expansion, yields

$$J_{prior}(\mathbf{x}) = \sum_{k=1}^{r} \lambda_k^{-1} \langle \mathbf{v}^k, \tilde{\phi}(\mathbf{x}) \rangle^2 + \lambda_\perp^{-1} \left(\langle \tilde{\phi}(\mathbf{x}), \tilde{\phi}(\mathbf{x}) \rangle - \sum_{k=1}^{r} \langle \mathbf{v}^k, \tilde{\phi}(\mathbf{x}) \rangle^2 \right). \qquad (6.47)$$

From this expression it is possible to make some insightful observations. First, we immediately recognize $\langle \mathbf{v}^k, \tilde{\phi}(\mathbf{x}) \rangle$ as the nonlinear PC, y_k, from kernel PCA (confer (6.41)). Moreover, $\langle \tilde{\phi}(\mathbf{x}), \tilde{\phi}(\mathbf{x}) \rangle = ||\tilde{\phi}(\mathbf{x})||^2$. Hence, we can split J_{prior} in two (squared) distances [50, 117]:

$$d_1 = \sum_{k=1}^{r} \lambda_k^{-1} \langle \mathbf{v}^k, \tilde{\phi}(\mathbf{x}) \rangle^2 = \sum_{k=1}^{r} \frac{y_k^2}{\lambda_k}, \qquad (6.48)$$

and

$$\begin{aligned} d_2 &= \lambda_\perp^{-1} \left(\langle \tilde{\phi}(\mathbf{x}), \tilde{\phi}(\mathbf{x}) \rangle - \sum_{k=1}^{r} \langle \mathbf{v}^k, \tilde{\phi}(\mathbf{x}) \rangle^2 \right), \\ &= \lambda_\perp^{-1} \left(||\tilde{\phi}(\mathbf{x})||^2 - \sum_{k=1}^{r} y_k^2 \right). \qquad (6.49) \end{aligned}$$

Let the subspace spanned by $\mathbf{v}^k$ be denoted $\mathcal{F} \subseteq \mathcal{Y}$ (kernel PCA space), and its orthogonal compliment be denoted $\overline{\mathcal{F}}$. Then d_1 is a (squared) distance in $\mathcal{F}$ while d_2 is a (squared) distance in $\overline{\mathcal{F}}$; see Figure 6.1. Hence, by regularizing $\mathbf{C}_\phi^e$ we have the possibility to search for solutions that are outside the span of the mapped data set, but since $\lambda_\perp < \lambda_r$ they are less probable than all the solutions within the span of the mapped data set.

The expression (6.47) is not calculable in its current form. To finalize the derivation, we insert for the PCs and eigenvalues from Section 6.2, and let $\tilde{k}(\mathbf{x}, \mathbf{y}) = \langle \tilde{\phi}(\mathbf{x}), \tilde{\phi}(\mathbf{y}) \rangle$. Omitting the details, the final expression becomes [50]

$$J_{prior}(\mathbf{x}) = \lambda_\perp^{-1} \tilde{k}(\mathbf{x}, \mathbf{x}) + \sum_{k=1}^{r} \left(\sum_{i=1}^{m} \alpha_i^k \tilde{k}(\mathbf{t}^i, \mathbf{x}) \right)^2 (\lambda_k^{-1} - \lambda_\perp^{-1}). \qquad (6.50)$$

From (6.42) we know that the centered kernel, $\tilde{k}$, can be evaluated from an 'uncentered' kernel, k, which can be CPD or PD. Hence, (6.50) is eaily calculated when an appropriate k has been chosen.

It is important to note that although J_{prior} is quadratic in $\mathcal{Y}$, it is not necessarily convex in $\mathbb{R}^n$. This is illustrated in the 1D example given in Figure 6.2a and 2D example given in Figure 6.2b. For the 1D example in Figure 6.2a we have also drawn the two terms in (6.50) separately: the dashed line denotes the first term (i.e., the term equivalent to the Parzen estimator, confer Appendix C), while the dotted line denotes the second term in (6.50). It is seen that the first term favors area with more data points (as a kernel density estimator tends to do), while the second term tends to compensate this such that J_{prior} (solid line) is impacted by all input data. This can also be seen for the 2D example in Figure 6.2b where only J_{prior} is plotted. The three clusters of data are encapsulated separately by J_{prior}, but also individual points outside the main clusters have an impact.

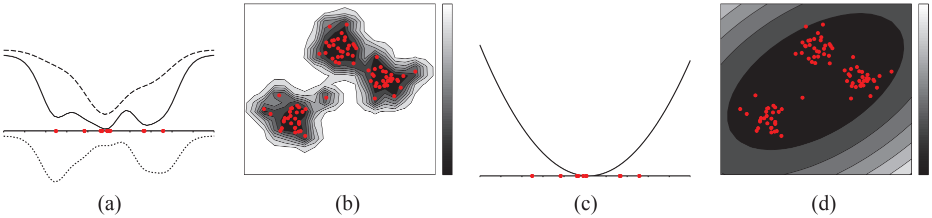

Figure 6.2: Examples of J_{prior} for a set of data (red dots). (a) 1D example with J_{prior} (solid), first (dashed) and second (dotted) term in (6.50). (b) 2D example of J_{prior}. The correspond examples calculated with a regularization term Gaussian in $\mathbb{R}^n$, J_{gauss}, in (c) 1D and (d) 2D.

The examples have also been calculated using a regularization term which is Gaussian in $\mathbb{R}^n$ (confer (3.32)); we denote this by J_{gauss}. Figure 6.2c shows J_{gauss} for the 1D example, and Figure 6.2d shows J_{gauss} for the 2D example. The main advantage of the shape prior regularization term is clearly seen by comparing Figure 6.2b and Figure 6.2d. While the information from the three clusters are smoothed out by J_{gauss}, they are separately encapsulated by J_{prior}. If, for example, the three clusters were data from three separate objects, say three types of subsurface structures in the CSEM case, then the information from the three objects would be smoothed out with J_{gauss}, but not with J_{prior}.

In Paper A, we chose the CPD kernel called power kernel, as it was shown to have a computational advantage in the CSEM inversion compared to the widely applied PD kernel, the Gaussian kernel. In [49, Appendix C] and [50], it was argued that the choice of Gaussian kernel lead to an interpretation of J_{prior} as a generalization of the Parzen kernel density estimator [130]. In Appendix C, we show that same argument can be made for the power kernel.

6.4 Kernel ensemble Kalman filter

In this section, we discuss a nonlinear generalization of EnKF (confer Section 3.3.4) based on the kernel relation (6.3) – the kernel-based EnKF (KEnKF). The main idea in KEnKF is to do the analysis step (confer (3.62) or (3.65)) in a high-dimensional feature space, $\mathcal{Y}$, instead of the original space, $\mathbb{R}^{N_{\psi_k}}$ [137]. By doing the analysis step in $\mathcal{Y}$, one hopes to capture more of the statistical properties of the posterior PDF than with EnKF. In the following, we only briefly outline the KEnKF algorithm, expanding on the notation and quantities introduced in Section 3.3.4. We refer to [137] for a full derivation of KEnKF.

For simplicity we assume that only the model parameters, $\mathbf{m}$, are mapped to the feature space; hence, the measurement and model noise are assumed Gaussian (an extension to non-Gaussian noise terms is given in [137]). We thus define the feature map as

$$\phi : \mathbf{m} \in \mathbb{R}^{N_m} \mapsto \phi(\mathbf{m}) \in \mathcal{Y}. \tag{6.51}$$

The forecast ensemble matrix is now given by (omitting the sequential step index)

$$\mathbf{Y}^f = \begin{bmatrix} \mathbf{G}^f \\ \mathbf{\Phi}_m^f \end{bmatrix}, \tag{6.52}$$

where $\mathbf{\Phi}_m = [\phi(\mathbf{m}^1), \ldots, \phi(\mathbf{m}^{N_e})]$. Note that $\mathbf{Y}^f$ is a $N_y \times N_e$ matrix, with $N_y = N_d + N_{\mathcal{Y}}$. The sample covariance matrix of the forecast ensemble, $\mathbf{C}^e_{y^f}$, is given in a similar manner as (6.32),

$$\mathbf{C}^e_{y^f} = \frac{1}{N_e} \widetilde{\mathbf{Y}}^f (\widetilde{\mathbf{Y}}^f)^T, \tag{6.53}$$

where $\widetilde{\mathbf{Y}}^f = \mathbf{Y}^f - \overline{\mathbf{Y}}^f$ with $\overline{\mathbf{Y}}^f = \mathbf{Y}^f \mathbf{1}_{N_e}$.

With the above definitions, the approximate Kalman gain (confer (3.63)) is given by

$$\mathbf{K}^e_y = \mathbf{C}^e_{y^f} \mathbf{H}^T (\mathbf{H}\mathbf{C}^e_{y^f}\mathbf{H}^T + \mathbf{C}_d)^{-1}, \tag{6.54}$$

where $\mathbf{H} = [\mathbf{I}_{N_d}, \mathbf{0}]$. Note that $\mathbf{K}^e_y$ is of size $N_y \times N_d$. Due the high dimensionality of ϕ, $\mathbf{K}^e_y$ will be, at most, of rank N_e. Moreover, the columns in $\mathbf{K}^e_y$ is a linear combination of the columns in $\widetilde{\mathbf{Y}}^f$ (this follows from (6.34) for $\mathbf{C}^e_{y^f}$, which translates directly to $\mathbf{K}^e_y$). Hence (6.54) can be written as

$$\mathbf{K}^e_y = \widetilde{\mathbf{Y}}^f \mathbf{A}. \tag{6.55}$$

With some manipulations we get

$$\mathbf{A} = \frac{1}{N_e} (\mathbf{G} - \overline{\mathbf{G}})^T (\mathbf{C}^e_g + \mathbf{C}_d)^{-1}. \tag{6.56}$$

The analysis step in KEnKF is given by

$$\begin{aligned} \mathbf{Y}^a &= \mathbf{Y}^f + \mathbf{K}^e_y(\mathbf{D} - \mathbf{H}\mathbf{Y}^f), \\ &= \mathbf{Y}^f + (\mathbf{Y}^f - \mathbf{Y}^f \mathbf{1}_{N_e})\mathbf{A}(\mathbf{D} - \mathbf{G}^f), \\ &= \mathbf{Y}^f \mathbf{B}, \end{aligned} \tag{6.57}$$

where $\mathbf{B} = (\mathbf{I}_{N_e} - \mathbf{1}_{N_e})\mathbf{A}(\mathbf{D} - \mathbf{G}^f)$ is a $N_e \times N_e$ matrix. From (6.57) the updated ensemble of mapped model parameters is given by

$$\mathbf{\Phi}^a_m = \mathbf{\Phi}^f_m \mathbf{B} \tag{6.58}$$

We are, however, not interested in the ensemble of updated model parameters in $\mathcal{Y}$, but rather the ensemble of model parameters in $\mathbb{R}^{N_m}$, $\mathbf{M}^a = [(\mathbf{m}^a)^1, \ldots, (\mathbf{m}^a)^{N_e}]$. To get $\mathbf{M}^a$, the inverse feature map, ϕ^{-1}, is required. This is denoted the exact pre-image problem [140]. Due to the large dimensionality of $\mathcal{Y}$ the exact pre-image is generally not possible to calculate. Instead we calculate an approximate pre-image, which is given by the following optimization problem

$$\begin{aligned} (\mathbf{m}^a)^j &= \arg\min_{\mathbf{m}} \; ||\phi(\mathbf{m}) - \phi((\mathbf{m}^a)^j)||^2, \\ &= \arg\min_{\mathbf{m}} \langle \phi(\mathbf{m}), \phi(\mathbf{m}) \rangle - 2\langle \phi((\mathbf{m}^a)^j), \phi(\mathbf{m}) \rangle + \langle \phi((\mathbf{m}^a)^j), \phi((\mathbf{m}^a)^j) \rangle, \\ &= \arg\min_{\mathbf{m}} \; k(\mathbf{m}, \mathbf{m}) - \sum_{i=1}^{N_e} B_{ij} k((\mathbf{m}^f)^j, \mathbf{m}) + \text{const.}, \end{aligned} \tag{6.59}$$

where we have used (6.3) and (6.58), with $j = 1,\dots,N_e$. The optimization problem (6.59) can be solved using, e.g., fixed-point iterations.

In [137], the authors argued for the use of a polynomial kernel given as

$$k(\mathbf{x},\mathbf{y}) = \sum_{l=1}^{q} \langle \mathbf{x},\mathbf{y} \rangle^{l}, \tag{6.60}$$

which made it possible to honour up to $2q$-order statistical moments (see also [138]); in particular, $q = 1$ reduces KEnKF to the standard EnKF algorithm.

6.4.1 Discussion

In [137], it was noted that using high-order polynomial in (6.60) could lead to ensemble collapse (i.e., all ensemble member collapsing to one vector), which is clearly an unwanted behaviour. The optimization problem (6.59) can also be difficult to solve in some cases, leading to larger computational expenses in KEnKF than in EnKF. In this work, preliminary tests using EnKF and KEnKF on CSEM inversion lead to little or no difference between the methods. Hence, it was concluded that for our CSEM inversion problem there was no clear advantage of using KEnKF above EnKF.

The goal in KEnKF was to extract more statistical information about the posterior PDF. As we note in Section 3.3.3 it is only in the case of Gaussian PDFs and linear forward model that a complete description of the posterior PDF is possible. In recent years, however, a vast theory has been developed that aims to describe a general probability distribution by mapping it to a RKHS; such mapping is called kernel embedding (see, e.g., [147] and references therein). In kernel embedding, the *whole* probability distribution is described by *one* point in $\mathcal{Y}$. The mapping for a univariate distribution is defined as

$$\boldsymbol{\mu}_X = E_X[\phi(\mathbf{X})] = \int_{\mathcal{X}} \phi(\mathbf{x})\,\mathrm{d}P(\mathbf{x}), \tag{6.61}$$

and for a general joint distribution, it is defined as

$$\mathbf{C}_{XY} = E_{XY}[\phi(\mathbf{X}) \otimes \phi(\mathbf{Y})] = \int_{\mathcal{X}\times\mathcal{X}} \phi(\mathbf{x}) \otimes \phi(\mathbf{y})\,\mathrm{d}P(\mathbf{x},\mathbf{y}). \tag{6.62}$$

Here, $\mathbf{X}$ and $\mathbf{Y}$ are random variables in $\mathcal{X}$ with instantiations $\mathbf{x}$ and $\mathbf{y}$, and probability distribution $P(\mathbf{x})$ and $P(\mathbf{y})$, respectively. Based on these mappings, basic statistical operations, such as Bayes' rule, can easily be generalized. To actually calculate the embeddings, sample versions of (6.61) and (6.62) must be defined. Using (6.3), the embeddings are calculable by choosing an appropriate kernel. It turns out that the kernels must be characteristic. Informally speaking, characteristic kernels are associated with a feature space large enough such that all the statistical moments of $P(\mathbf{x})$ are mapped to $\mathcal{Y}$; an example is the Gaussian kernel. Although some applications have used embedded versions of Bayes' rule, preliminary study on this topic did not reveal how an implementation for the CSEM inversion problem should look like.

Chapter 7

Summary of the papers and future work

In this chapter, the main results of Paper A – C associated with Part I of this thesis are given. The summaries will be based on the scientific background presented in Part I. Comments on potential future work will also be given.

7.1 Summary of Paper A

Title:	*Identification of subsurface structures using electromagnetic data and shape priors*
Authors:	S. Tveit, S. A. Bakr, M. Lien, and T. Mannseth

In Paper A, we presented a methodology for inversion of controlled source electromagnetic (CSEM) data using a reduced representation (see Chapter 5) and shape prior regularization (see Section 6.3). The inversion methodology was applied to identify large-scale subsurface structures where the contrast in electric conductivity can be small.

To represent the structure boundaries, the Vese-Chan level-set representation discussed in Section 5.1.1 was applied. The bilinear interpolation technique was used to represent the level-set functions (see Section 5.2.1), and smoothing of the level-set functions was done according to the description in Section 5.1.4. This allowed for a flexible representation of geological formations with a minimum number of parameters.

To obtain accurate placement of the structure boundaries, structural prior information were incorporated. Such type of prior information most often comes from seismic interpretation (see Section 2.2). The shape prior regularization technique was used to incorporate the structural information. Using the shape prior technique we were able to incorporate different types of subsurface structures. That is, two or more equally probable types of geological models could be incorporated, since each type of geological model will be captured by the shape prior technique (see Figure 6.2b for a toy example).

The shape prior regularization term was incorporated using a different kernel function than the one suggested in the original paper [50]. For the application of CSEM inversion, a CPD kernel – the power kernel (see Appendix C) – was shown to be more useful. As shown in Appendix C, choosing the power kernel did not result in a loss of theoretical properties for the shape prior regularization term.

The reduced representation enable the use of Newton-type methods where Hessian information could be employed. Specifically, the Levenberg-Marquardt method (see Section 3.2.1) was used. The shape prior regularization term was incorporated as a Tikhonov regularization term (see Sections 3.2.3 and 6.3) where the regularization parameter was chosen such that both the data misfit and regularization term was equal initially, and subsequently reduced to put more emphasis on the observed data in later iterations.

The methodology was tested on several different test cases, where the impact of the shape prior regularization term was the main objective. Only the shape and position of the structures were estimated; hence, $\mathbf{c}(\mathbf{r})$ was assumed known and constant within each structure. In some test cases, inversion without applying shape prior regularization could be approximately identified, but the use of shape prior regularization clearly improved the results, and in some cases was crucial to get even an approximate identification.

7.2 Summary of Papers B and C

Paper B:

Title:	*Ensemble-based, Bayesian inversion of CSEM data using structural prior information*
Authors:	S. Tveit, S. A. Bakr, M. Lien, and T. Mannseth

Paper C:

Title:	*Ensemble-based Bayesian inversion of CSEM data for subsurface structure identification*
Authors:	S. Tveit, S. A. Bakr, M. Lien, and T. Mannseth

In Papers B and C, a Bayesian inversion methodology for the identification of large-scale subsurface structures from CSEM data was presented. The inversion methodology applied a reduced representation and solved the Bayesian inverse problem using the ensemble Kalman filter (EnKF), see Section 3.3.4. In Paper B, we only considered the identification of shape and position of the structures, while in Paper C we also considered the identification of the conductivity value within each structure.

The reduced representation utilized in the papers was the hierarchical level-set representation presented in Section 5.1.2. With the hierarchical representation, it was possible to construct complex geological formations like, e.g., faults and pinchouts. Bilinear interpolation was used for the representation of level-set functions in both papers.

In Paper B, the structures were assumed to have constant conductivity value. In Paper C, however, we considered heterogeneous conductivity distribution, hence, the expansion function, $\mathbf{c}(\mathbf{r})$, was represented using the interpolation technique discussed in Section 5.2.2. Specifically, we used bilinear interpolation to get a smooth conductivity distribution within each structure.

The update of model parameters ($\mathbf{m} = \mathbf{a}$ for Paper B and $\mathbf{m} = [\mathbf{w}^T, \mathbf{a}^T]^T$ for Paper C) was done using the EnKF algorithm (confer (3.65)). To make the initial ensemble for EnKF, samples must be generated from an initial prior PDF. It was assumed that the prior PDF consisted of structural prior information from, e.g., a seismic interpretation (as for Paper A). Moreover, for the conductivity distribution within each structure it was assumed that prior information was available from, e.g., well logs.

The initial prior PDF was assumed to be Gaussian, hence, a mean prior model together with a covariance matrix had to be generated. While a mean prior model is easily generated, much effort can be put into the generation of the covariance matrix. It was assumed that the **w**- and **a**-coefficients were not correlated, and moreover, the $\mathbf{w}_j$'s were not correlated with each other, and the same for the $\mathbf{a}_i$'s. Consequently, separate covariance matrices were made for the conductivity distribution within each structure and for each structure boundary. By assuming that the correlation between two coefficients was only dependent on their normalized spatial distance (not their physical distance), an analytical covariance model (the spherical covariance model) could be utilized. The flexibility of the chosen covariance model made it possible to make a wide variety of prior models in the numerical experiments.

The numerical test cases made in Paper B showed that the methodology was able to recover fairly complex subsurface structures (included was, e.g., a fault). It was also able to identify the correct shape and position of included hydrocarbon reservoirs. Moreover, the methodology was able to completely remove a non-existing reservoir present in the prior models, and identify the remaining structures. By plotting the structure boundaries of each ensemble member, it was seen that the uncertainty in the initial ensemble was significantly reduced in the final ensemble.

In the numerical experiments in Paper C, the inversion methodology was applied on subsurface models where the structures had a heterogeneous conductivity distribution. In each test case, the prior model was far away from the reference model (the conductivity distribution was homogeneous in each structure, and the shape and position of the structures was far away from the reference solution). The numerical results showed that the methodology was able to recover the reference model reasonably well in each test case. A similar test case as we did in Paper B, where a non-existing reservoir was present in the prior models, was conducted in Paper C also. Now, the reservoir was almost completely removed, and the remaining subsurface structures were fairly close to the reference model. Similarly to what was seen in Paper B, by plotting the structure boundaries of the individual ensemble members, it was seen that the uncertainty was reduced from initial ensemble to final ensemble. However, since the conductivity distribution within each structure was also estimated, a more quantitative assessment of the quality of the final ensemble compared to the initial ensemble was done by computing the data misfit for both. A significant reduction was also seen for the data misfit, indicating that the uncertainty present in the initial ensemble was reduced in the final ensemble.

Comparing the results from Paper B and C, it was concluded that introducing the **w**-coefficients in the inversion process, which allowed for the estimation of a smoothly varying conductivity distribution, made the identification of region boundaries (**a**-coefficients) more difficult.

7.3 Future work

In a recent paper [18], a novel numerical method for simulating 3D CSEM measurements has been developed. Here, one part of the computational domain is modelled using the 2.5D approach discussed in Section 4.6.2, while the remaining parts are discretized utilizing a 3D FE method. Hence, targets in the subsurface where it is important to know the spatial extent of the body in all three directions, like, e.g., hydrocarbon reservoirs, can be modelled with smaller computational effort than a full 3D FE method. In this context, it would be interesting to extend the model-based representation used in this thesis to be able to represent the subsurface structures outside a potential reservoir in 2D, and the target reservoir in 3D. The obvious difficulty is how to couple the 2D and 3D part with the implicit representation presented in Chapter 5, without creating unnecessary overhead computations.

In the numerical experiments shown in Papers A, B, and C, the number of structures in the initial models were the same as the reference models. In reality, the number of structures may not be known *a priori*. As discussed in Section 5.3, it is possible to apply a multiscale approach where new structures can be introduced as a part of the estimation sequence. Moreover, a multiscale refinement procedure for the **w**- and **a**-coefficients can be beneficial in the estimation. An alternative to refining **w** and **a** is to increase the polynomial order of the interpolation function in the representation of the level-set functions and expansion functions (confer (5.19) and (5.21), respectively).

An implementation of the hierarchical representation (see Section 5.1.2) in a multiscale approach would be advantageous. The scale-by-scale description of the hierarchical representation naturally allows for the introduction of new region boundaries that are confined within the area of interest.

As noted in Section 5.3, multiscale strategies have been implemented using an adaptive approach [24, 25, 26, 101, 102]. Here, a linearized data misfit function (confer (3.4)) based on one Gauss-Newton step (confer (3.12)) was computed to decide which part of the model domain that needed refinement. An extension of this decision procedure to include a prior term, in particular the shape prior regularization term (see Section 6.3), would be interesting.

The classical inversion methodology presented in Paper A, used shape prior regularization together with a Netwon-type method (see Section 3.2.1) where the sensitivity matrix was computed using the direct method (see Section 3.2.2). When complex subsurface models are considered, more parameters are needed in the representation, and thus the sensitivity matrix becomes more expensive to calculate. An alternative way of calculating sensitivity was outlined in the Discussion after Section 3.3.4, where an ensemble of model parameters and forward model outputs can be used to calculate the sensitivity matrix in an iterative ensemble-based method. Incorporating the shape prior technique in an iterative ensemble-based framework could be both computationally efficient and robust. In general, it would be interesting to further investigate the use of kernel methods (see Section 6) in a Bayesian inversion of CSEM, for example, based on the kernel embedding theory briefly mentioned in Section 6.4.1.

The results given in Papers B and C, showed that there is potential in using EnKF as method for inversion of CSEM. As mentioned in the Discussion after Section 3.3.4, how we define the subset of observed data in EnKF can have an affect on the inversion result. A thorough investigation on the impact of different groupings of the observed data could reveal an even greater potential for using EnKF in CSEM inversion.

In addition to the abovementioned recommendations for future work, the following suggestions are also of interest:

- Applying the inversion methodology on numerical experiments were anisotropic electric conductivity distribution is considered.
- Use the developed methodologies on an inversion of MT and CSEM data as suggested in Section 2.3.
- Straightforward extension of the inversion methodologies to 3D subsurface models.
- Application of the developed inversion methodologies on real-world observed data.

Part II

Scientific background
–
Numerical Assessment of the Upstream Mobility Scheme

Chapter 8

Introduction

The understanding of fluid flow in the subsurface is important in many areas of science. For example, in the petroleum industry, an understanding of subsurface fluid flow is required to increase the recovery of oil and gas from a reservoir. This has become increasingly important over the last decades as more complicated recovery strategies have to be implemented in order to fully profit from mature petroleum reservoirs. Another example is the sequestration of CO_2. To lower the CO_2 emission into the atmosphere, CO_2 can be stored in deep subsurface reservoirs. It is then important to understand short and long term effects of injecting CO_2 in a geological formation.

The dynamics of fluid flow in the subsurface is modelled by mathematical equations (see Chapter 9). The complex structure of the reservoir and involved interactions between the fluids make it difficult to solve these mathematical equations analytically. Hence, we are dependent on numerical methods (see Chapter 10) to make fluid flow predictions. In the petroleum industry, large investments are made to develop commercial reservoir simulators, which can be used in planning of well placements, testing different production scenarios, investigation of new enhanced oil recovery methods, etc. Reservoir simulators are also important in history matching. Here, the parameters that govern the fluid flow in a reservoir (e.g., permeability, porosity, fault transmissibility, etc.) are adjusted in such a way that the outcome of the reservoir simulator matches real observed data (to some stimulus). With the updated parameters, the reservoir simulator can make more reliable predictions, e.g., on future oil and gas recovery rates.

From the above discussion it is clear that a fundamental understanding of the properties and limitations of the numerical methods behind the reservoir simulators is crucial. Over the years, numerical methods for fluid flow have been studied extensively both theoretically and numerically, and in most cases, the applicability of the methods has been established. However, the applicability of even the most well-established methods is, in some cases, not fully understood. In this thesis, one such method is under consideration, namely the upstream mobility scheme (see Section 10.5).

The upstream mobility scheme is a widely applied numerical method for calculating the fluid flux through a grid cell. Many reservoir simulators have implemented the upstream mobility scheme due to its simplistic and intuitive nature. Even so, a general convergence proof of the scheme has not been established. Among practitioners, however, it has been known for some time to produce erroneous results, and some experiments in the literature have also indicated minor errors for the scheme for some simple flow situations [114].

Influenced by previous results and experiences, the main objective of this thesis is to investigate the upstream mobility scheme numerically for flow situations commonly associated with the fluid flow in the subsurface, as exemplified in the beginning of this chapter. The investigations are done in 1D where it is possible to compare the numerical results to analytical solutions (see Section 9.3.3).

Chapter 9

Mathematical model

In this chapter, we discuss the mathematical model behind two-phase fluid flow in a 1D heterogeneous porous medium. Starting from the basic equations, we derive the model most often associated with two-phase flow, namely the Buckley-Leverett equation. The Buckley-Leverett equation is a hyperbolic conservation law, and for heterogeneous porous media, the associated flux function will be discontinuous. Hence, we will present the theory on hyperbolic conservation laws with discontinuous flux functions, and discuss the correct physical solution to the Buckley-Leverett equation in heterogeneous porous media.

9.1 Reservoir properties

To describe the physical properties of fluid flow in a reservoir, the porous rock saturated with fluids is regarded as a *continuum*; that is, all the involved components fill the entire space under consideration. Roughly speaking, the involved properties can be divided into three groups: rock properties, fluid properties, and properties that result from the interaction between rock and fluid. In the following, we will only give a brief overview of the three groups, and refer to standard textbooks on the topic, e.g., [170], for a full description.

9.1.1 Rock properties

A typical reservoir consists of rock with void spaces, pores, where fluid can accumulate or flow freely. The solid part of the rock is denoted matrix. On the continuum scale, a porous rock is described by averaged parameters, which are only valid if the scale of the reservoir is (much) larger compared to effects of micro-scale phenomena.

Porosity

There are two definitions of porosity: absolute and effective porosity. The first considers the total void space in the rock, while the latter only considers the void space that are interconnected. Since fluid can only flow through interconnected pore volumes, we will only consider effective porosity. It is defined as

$$\phi = \frac{V_p}{V_t}, \tag{9.1}$$

where V_p is the volume of interconnected pores, and V_t is the total volume of the rock. Note that ϕ is a dimensionless quantity.

The effective porosity can be regarded as the storage capacity of the reservoir, and is determined by several factors, such as grain size, cementation, and rock type (sandstone, limestone, etc.).

Absolute permeability

The absolute permeability (K) is the porous medium's capability of transmitting fluids through the network of interconnected pores. It depends on many factors, most importantly, effective porosity. Although there is no simple relationship between absolute permeability and effective porosity, they are correlated (high effective porosity usually leads to high absolute permeability). Absolute permeability is, in general, a symmetric, positive definite tensor, that is, it is direction dependent. Usually, the horizontal permeability is much larger than the vertical permeability. The SI unit of absolute permeability is m^2, but conventionally the unit of measure is Darcy (D).

9.1.2 Fluid properties

A description of the fluid properties is necessary as they affect the fluid's inherent ability to flow in a porous medium. The properties of the fluid is highly dependent on temperature and pressure. A change in either can drastically change fluid flow properties, and in extreme cases a change of state can occur, e.g., a liquid can become gas or solid. In this work, we will not consider any phase transitions.

Viscosity and density

Viscosity (μ) is the fluid's internal resistance to shear stress. The viscosity is highly dependent on the temperature, where it typically decreases for a liquid, while it increases for a gas. It is also dependent on the pressure, and shape and size of fluid particles. Due to the friction introduced by viscosity, the flow velocity of a fluid in porous media is typically highest in the middle of the pores and decreasing towards the wall. The SI unit of viscosity is pascal-seconds (Pa·s), however, it is often given in poise (P).

Density (ρ) is the mass per unit volume of a fluid. Changes in density of a fluid is most often due to change in pressure. In some cases, the composition of the fluid changes with temperature which, consequently, changes its density. The latter is typical for hydrocarbon fluids.

9.1.3 Petrophysics

When two or more fluids are present in the rock, interactions internally between the fluids, and between fluids and rock can occur. In this work, we assume that the reservoir is occupied by maximum two fluids, which are both in the liquid phase. In addition, the two liquids are immiscible, and can thus be classified according to wettability, i.e., the attraction to solid rock. The notation $\alpha = n, nw$ is introduced to denote either the wetting (w) or nonwetting (nw) phase.

Saturation

The fraction of pore volume occupied by a particular fluid is defined by the dimensionless quantity saturation,

$$S_\alpha = \frac{V_\alpha}{V_p}, \tag{9.2}$$

where V_α is the volume occupied by phase α. Due to the cohesive and adhesive forces, a phase may become immobile under certain circumstances. In other words, there is a minimum saturation at which each phase will become immobile, denoted by $S_{\alpha,r}$. Hence, it is often convenient to normalize the saturation according to

$$u_\alpha = \frac{S_\alpha - S_{\alpha,r}}{1 - S_{nw,r} - S_{w,r}}. \tag{9.3}$$

u will thus vary between zero and one. Note that the total volume of the pores must be filled by all phases, hence

$$u_w + u_{nw} = 1. \tag{9.4}$$

Although u is a normalized saturation, it will in the following be referred to as just 'saturation'.

Capillary pressure

The capillary pressure (P_c) is defined as the (molecular) pressure difference across the interface of two immiscible fluids,

$$P_c = p_{nw} - p_w, \tag{9.5}$$

where p_α is the internal pressure of phase α. The capillary pressure depends on many factors, most notably saturation. The unit is the same as pressure, pascal (P).

At large scales, capillary pressures are often neglected, but, as we will discuss below (see Section 9.3.4), it can have an important impact on the solution of the flow equations at the interface between two different homogeneous porous media.

Relative permeability

When more than one fluid occupies the pore volume, the fluid flow of one phase is limited, not only by the rock properties, but also the fluid properties. A common assumption is that fluid flow of a phase can be modeled as a single phase flow through a reduced pore volume due to the presence of the other phase. This is called effective permeability, and is denoted K_α^e. For practical purposes, relative permeability has been introduced to describe the relationship between effective and absolute permeability. It is a dimensionless quantity defined as

$$k_\alpha = \frac{K_\alpha^e}{K}. \tag{9.6}$$

The relative permeability is generally a nonlinear function of saturation, and the sum of k_α is always less than one.

9.2 Two-phase flow in heterogeneous porous media

We consider the flow of two phases in a 1D heterogeneous reservoir. The heterogeneous porous medium is given by two or more adjacent homogenous porous media with disparate permeability values. The change in permeability can occur in absolute or relative permeability, or both. For simplicity we assume that porosity does not change between the different porous media (in any case, it is only a scaling factor for the fluid flow). In this work, we consider the simplest case of two adjacent homogeneous media with an interface at $x = x_h$. In the following, a superscript 'L' and 'R' is adopted to indicated quantities that are different on the left and right side of x_h, respectively. We also follow the commonly used subscript notation $\cdot_t = \frac{\partial\cdot}{\partial t}$ and $\cdot_x = \frac{\partial\cdot}{\partial x}$.

The velocity of each phase in the porous medium is assumed to be well approximated by Darcy's law

$$q_\alpha^{L,R} = -\lambda_\alpha^{L,R}\left((p_\alpha)_x - \rho_\alpha g \cos\beta\right), \tag{9.7}$$

where we have introduced the effective mobility of phase α

$$\lambda_\alpha^{L,R}(u) = \frac{K^{L,R}k_\alpha^{L,R}(u)}{\mu_\alpha}, \tag{9.8}$$

and $g\cos\beta$ is the influence of gravity at the angle β from the vertical axis. Note that Darcy's law is an empirical law giving the bulk velocity of phase α, $q_\alpha^{L,R}$.

Conservation of mass in an arbitrary interval $[x_1,x_2]$ without a source is given as

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{x_1}^{x_2} \phi\rho_\alpha u_\alpha \,\mathrm{d}x + \left[\rho_\alpha q_\alpha^{L,R}\right]_{x_1}^{x_2} = 0. \tag{9.9}$$

The first term describes the rate of change of total mass in $[x_1,x_2]$, while the second term describes the total flux into or out of $[x_1,x_2]$. With some modifications and recognizing that $[x_1,x_2]$ was arbitrary, (9.9) can be written on differential form,

$$\phi(u_\alpha)_t + (q_\alpha^{L,R})_x = 0, \tag{9.10}$$

where we have assumed incompressible rock and fluids.

From (9.4) and (9.10), we get

$$(q_w^{L,R} + q_{nw}^{L,R})_x = 0, \quad \Rightarrow \quad q = q_w^{L,R} + q_{nw}^{L,R}, \tag{9.11}$$

where q is the total bulk velocity, which is independent of x. Subtracting (9.7) for each phase, and using (9.5) and (9.11) leads to

$$q_w^{L,R} = \frac{\lambda_w^{L,R}}{\lambda_w^{L,R} + \lambda_{nw}^{L,R}}[q + (\gamma_w - \gamma_{nw} + (P_c)_x)\lambda_{nw}^{L,R}]. \tag{9.12}$$

where $\gamma_\alpha = \rho_\alpha g\cos\beta$. For ease of notation, we define

$$f(u_w) = \frac{\lambda_w^R}{\lambda_w^R + \lambda_{nw}^R}[q + (\gamma_w - \gamma_{nw})\lambda_{nw}^R], \tag{9.13}$$

$$g(u_w) = \frac{\lambda_w^L}{\lambda_w^L + \lambda_{nw}^L}[q + (\gamma_w - \gamma_{nw})\lambda_{nw}^L], \tag{9.14}$$

$$F(u_w,x) = \begin{cases} f(u_w), & \text{if } x \geq x_h, \\ g(u_w), & \text{if } x < x_h. \end{cases} \tag{9.15}$$

Inserted into (9.10) for the wetting phase yields the parabolic equation

$$\phi(u_w)_t + F(u_w,x)_x = \left(\frac{\lambda_w^{L,R} \lambda_{nw}^{L,R}}{\lambda_w^{L,R} + \lambda_{nw}^{L,R}} (P_c)_x \right)_x . \tag{9.16}$$

In the case of negligible capillary pressure, we get the hyperbolic conservation law for the wetting phase

$$\phi(u_w)_t + F(u_w,x)_x = 0, \tag{9.17}$$

also known as the heterogeneous Buckley-Leverett equation. When u_w is found from (9.17), u_{nw} is easily calculated from (9.4). The functions $F(u_w,x)$, $f(u_w)$, and $g(u_w)$ are denoted flux functions. In the case of negligible capillary pressure, $F(u_w,x)$ is equal to the bulk velocity of the wetting phase. Solutions of (9.17) are discussed in the next section.

9.3 Conservation laws with discontinuous flux

In the following, we will present the general theory on conservation laws with a discontinuous flux function, in which two-phase flow in heterogeneous media can be seen as a special application. To this end, consider the following Cauchy problem

$$\begin{aligned} u_t + F(u,x)_x &= 0, \\ u(x,0) &= u_0(x), \end{aligned} \tag{9.18}$$

where $u = u(x,t) \in [0,1]$, $x \in \mathbb{R}$, and $t = [0,T)$. The flux function $F(u,x)$ is given as in (9.15), written in an equivalent form as

$$F(u,x) = H(x - x_h) f(u) + (1 - H(x - x_h)) g(u), \tag{9.19}$$

with H being the Heaviside function. For now, we let $f(u)$ and $g(u)$ be general continuous nonlinear functions (further specifications are made in Section 9.3.2). With (9.19), (9.18) can be separated into two autonomous conservation laws

$$u_t + f(u)_x = 0, \quad \text{for } x \geq x_h, \tag{9.20}$$

$$u_t + g(u)_x = 0, \quad \text{for } x < x_h. \tag{9.21}$$

Hence, away from the interface x_h the problem is determined by either (9.20) or (9.21), which are just classical nonlinear conservation laws. We will discuss solutions of (9.20) or (9.21) in the next section. At the interface x_h, special care must be taken to ensure that the correct solution is found. This will be more discussed in Sections 9.3.2 and 9.3.3.

An important special case of (9.18) arise when the initial value is discontinuous,

$$u_0(x) = \begin{cases} u_l^0, & \text{if } x < 0, \\ u_r^0, & \text{if } x > 0, \end{cases} \tag{9.22}$$

which is called the Riemann problem. The solution to a Riemann problem is a similarity solution, that is, u is a function of x/t alone and is self-similar at all times. Indeed, the two-phase flow problem given in (9.17) is often posed as a Riemann problem. Riemann problems are also important for the numerical schemes discussed in the next chapter. The solution procedure below, however, is not restricted to a particular instance of $u_0(x)$, and we will thus keep the presentation general.

9.3.1 Away from the interface

The solution of (9.20) and (9.21) is given as [98]

$$u(x,t) = u_0(x - f'(u)t), \quad \text{for } x > x_h, \tag{9.23}$$
$$u(x,t) = u_0(x - g'(u)t), \quad \text{for } x < x_h. \tag{9.24}$$

That is, the solution is constant along the curves $x = x_0 + f'(u)t$ and $x = x_0 + g'(u)t$ in the $x - t$ plane, denoted characteristics. In the case where the characteristics do not cross, the solution will be continuous at all times. Such solutions are called rarefaction waves. Due to $f(u)$ and $g(u)$ being general nonlinear functions, characteristics on either side of x_h may cross at some finite time. Hence, even if $u_0(x)$ is smooth, the solution may become multivalued, which is not physically acceptable. Consequently, a weaker form of the solution must be sought. To this end, let $u \in L^\infty(\mathbb{R} \times [0,T))$ be a solution of (9.18) satisfying

$$\int_0^\infty \int_{-\infty}^\infty [u\varphi_t + F(u,x)\varphi_x] \,\mathrm{d}x\mathrm{d}t + \int_{-\infty}^\infty \varphi(x,0)u(x,0) \,\mathrm{d}x = 0, \tag{9.25}$$

for all test functions $\varphi(x,t) \in C_c^\infty(\mathbb{R} \times [0,T))$. Note that (9.25) is found from the integral form of the conservation law and multiplying in the test function.

With (9.25), we can allow u to develop discontinuities, or shocks, whenever the characteristics cross, and consequently, circumvent the problem of the unphysical multivalued solution. The speed of a shock is determined by the Rankine-Hugoniot condition. Let $\sigma_{L,R}$ denote the speed of a shock to the left or right of x_h, and let u_l and u_r be the solution immediately to the left and right of the shock discontinuity. Then the Rankine-Hugoniot condition is given by

$$\sigma_R = \frac{f(u_r) - f(u_l)}{u_r - u_l}, \quad \text{for } x > x_h, \tag{9.26}$$
$$\sigma_L = \frac{g(u_r) - g(u_l)}{u_r - u_l}, \quad \text{for } x < x_h. \tag{9.27}$$

Even though u is a weak solution, it may not necessarily be unique. Hence, we need to evoke further conditions to pick out the unique weak solution, which, moreover, must be physically correct. To this end, so-called *entropy conditions* have been developed to check if a solution is admissible. Entropy conditions have strong connections to the second law of thermodynamics, which states that the total physical entropy of a system must be nondecreasing in time. This can be state in mathematical form by defining a convex entropy function $\eta(u)$ with $\eta''(u) > 0$, and an entropy flux $\psi(u)$ satisfying $\psi'(u) = \eta'(u)f'(u)$ for $x > x_h$ and $\psi'(u) = \eta'(u)g'(u)$ $x < x_h$. A weak solution u is said to be an entropy weak solution if, for any $\varphi(x,t) \in C_c^\infty(\mathbb{R} \times [0,T))$ it satisfies [98]

$$\int_0^\infty \int_{-\infty}^\infty [\eta(u)\varphi_t + \psi(u)\varphi_x] \,\mathrm{d}x\mathrm{d}t + \int_{-\infty}^\infty \varphi(x,0)\eta(u(x,0)) \,\mathrm{d}x \geq 0. \tag{9.28}$$

A particularly useful entropy pair is the Kružkov entropies given as $\eta(u) = |u - c|$ and $\psi(u) = \operatorname{sign}(u - c)(f(u) - f(c))$ for $x > x_h$ and $\psi(u) = \operatorname{sign}(u - c)(g(u) - g(c))$ for $x < x_h$, where $c \in \mathbb{R}$ is arbitrary. Inserting the Kružkov entropies into (9.28) and

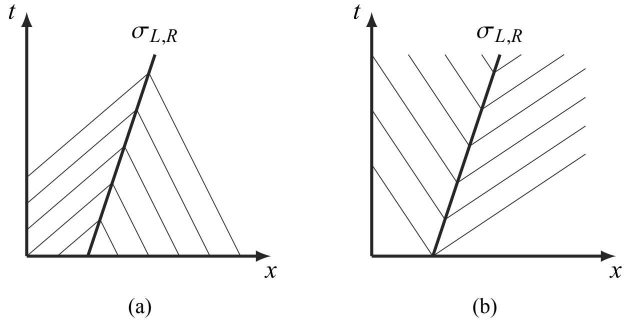

Figure 9.1: Illustrations of characteristics (on an arbitrary side of x_h) near a shock. (a) Valid and (b) invalid characteristics according to the Lax entropy conditions.

studying the behaviour of an entropy weak solution around a shock discontinuity leads to the famous Oleinik entropy conditions [126]

$$\frac{f(u)-f(u_l)}{u-u_l} \geq \sigma_R \geq \frac{f(u)-f(u_r)}{u-u_r}, \quad \text{for } x > x_h, \tag{9.29}$$

$$\frac{g(u)-g(u_l)}{u-u_l} \geq \sigma_L \geq \frac{g(u)-g(u_r)}{u-u_r}, \quad \text{for } x < x_h. \tag{9.30}$$

Hence, if a solution to (9.18) develops a shock away from x_h, it must satisfy the Oleinik entropy conditions to be a physically valid solution.

Letting $u \to u_{l,r}$ in (9.29) and (9.30) leads to the Lax entropy conditions

$$f'(u_l) \geq \sigma_R \geq f'(u_r), \quad \text{for } x > x_h, \tag{9.31}$$

$$g'(u_l) \geq \sigma_L \geq g'(u_r), \quad \text{for } x < x_h. \tag{9.32}$$

These are weaker conditions than (9.29) and (9.30) as they do not provide information on intermediate solutions. However, they provide insight on the behaviour of the characteristics near a shock. That is, characteristics may never come out of a shock, only go into one, as time advances; see Figure 9.1.

9.3.2 At the interface

The entropy conditions given in (9.29) and (9.30) only ensures a unique solution away from x_h. To get a fully unique solution of (9.18), an entropy condition for the solution at x_h is necessary. It turns out, however, that no unique entropy condition at the interface exists for conservation laws with a discontinuous flux function. In fact, many authors have suggested different admissibility criteria for a solution at x_h (see, e.g., [10] and references therein). In [6], the authors gather previously existing admissible solutions to (9.18) into one unified theory, which was further justified in [33]. (More recently, the theory was further summarized, and admissible solutions were given as a family of 'elementary solutions', see [10].) In the following, we will give an overview of the entropy condition for the interface x_h presented in [6, 33].

We start with some restrictions on the flux functions. Let $f, g \in \text{Lip}([0,1])$, and let $\theta_f \in [0,1]$ and $\theta_g \in [0,1]$ denote the global extrema of $f(u)$ and $g(u)$, respectively.

Furthermore, $f(0) = g(0)$ and $f(1) = g(1)$. Then the flux functions satisfy either of the following properties:

(i) $f(u)$ and $g(u)$ have one global maximum and no local minimum in $(0,1)$. Such flux functions are denoted $CC([0,1])$ (concave type).

(ii) $f(u)$ and $g(u)$ have one global minimum and no local maximum in $(0,1)$. Such flux functions are denoted $CV([0,1])$ (convex type).

Let $u^+(t) = \lim_{x\to x_h^+} u(x,t)$ and $u^-(t) = \lim_{x\to x_h^-} u(x,t)$ denote the right and left traces, respectively. The flux must necessarily be equal on both sides of x_h, which is given by the following Rankine-Hugoniot condition

$$f(u^+(t)) = g(u^-(t)). \tag{9.33}$$

To derive the entropy condition at the interface the following definition is crucial.

Definition 1 *The pair* $(A,B) \in [0,1]$ *is called a* connection *if*

1. $f(A) = g(B)$.

2. For $f,g \in CC([0,1])$ *we have* $\theta_g \leq A \leq 1$ *and* $0 \leq B \leq \theta_f$.

3. For $f,g \in CV([0,1])$ *we have* $0 \leq A \leq \theta_g$ *and* $\theta_f \leq B \leq 1$.

Note that the second and third entry essentially says that A lies in the region where $g'(u) \leq 0$ and B lies in the region where $f'(u) \geq 0$.

With Definition 1, we can define the following function [33]

$$c^{AB}(x) = H(x-x_h)B + (1 - H(x-x_h))A = \begin{cases} A, & \text{for } x \leq x_h, \\ B, & \text{for } x > x_h. \end{cases} \tag{9.34}$$

Simliarly to the Kružkov entropies we defined in the previous section, we now let $\eta = |u - c^{AB}(x)|$ and $\psi = \operatorname{sign}(u - c^{AB}(x))(F(u,x) - F(c^{AB}(x),x))$. Inserted into (9.28), leads to the following entropy condition [33]

$$\operatorname{sign}(u^+(t) - B)(f(u^+(t)) - f(B)) - \operatorname{sign}(u^-(t) - A)(g(u^-(t)) - g(A)) \leq 0, \tag{9.35}$$

which is understood in the weak sense. This is called the interface entropy condition. In summary, the interface entropy condition says that any transition of u across x_h given by $u^-(t)$ and $u^+(t)$ must fulfill (9.35) for any choice of (A,B) that satisfy the conditions in Definition 1. Note that $u^-(t) = A$ and $u^+(t) = B$ is only a special case in which (9.35) is fulfilled with equality.

A similar characteristic condition as the Lax entropy conditions, (9.31) and (9.32), can be derived from (9.35) [33]

$$\min\{0, g'(u^-(t))\} \max\{0, f'(u^+(t))\} = 0, \quad \text{if } (u^-(t), u^+(t)) \neq (A,B). \tag{9.36}$$

This says that the characteristics must lead back to the x-axis at least on one side of x_h, unless $(u^-(t), u^+(t)) = (A,B)$, in which case there are no restrictions on the characteristics.

In the last two sections, we have presented the entropy conditions that are valid both away from and at the interface x_h. The following definition summaries valid entropy solutions to conservation laws with a discontinuous flux function [6].

Definition 2 *$u \in L^\infty(\mathbb{R} \times [0,T))$ is a unique entropy solution of* (9.18) *if the following holds*

1. *u is a weak solution, i.e.,* (9.25) *holds.*

2. *u satisfies* (9.29) *for $x > x_h$, and* (9.30) *for $x < x_h$.*

3. *u satisfies* (9.35) *for $x = x_h$.*

The proof of uniqueness, existence, and L_1-stability of the entropy solution u satisfying the above definition is found in [6, 33].

Remarkably, the definition of a unique entropy solution is valid for any choice of (A,B), which means that, in theory, there exist an infinite number of entropy solutions to (9.18). Hence, we need to choose (A,B) carefully such that the correct entropy solution for each application is found. This is discussed more in the next section.

9.3.3 The physically meaningful entropy solution

Most hyperbolic conservation laws are derived by letting a higher-order dissipation term in a parabolic equation approach zero. In the limit, we expect the behaviour of the parabolic solution to be the same as the hyperbolic solution. In fact, the entropy conditions discussed in the previous sections are derived by studying the behaviour of the entropy functions and fluxes in the limit between parabolic and hyperbolic solutions.

When $f = g$, there is only one entropy condition (i.e., the Oleinik entropy condition), which is derived from the viscous equation [98]

$$u_t^\epsilon + h(u^\epsilon)_x = \epsilon u_{xx}^\epsilon, \tag{9.37}$$

where h is the flux function for the continuous conservation law. In the case of $f \neq g$, as we have in (9.18), a unique entropy solution must be derived by considering each physical model by itself, that is, we need to consider from which parabolic differential equation the hyperbolic conservation law is derived from. In the following sections, we present two entropy conditions most often encountered when discussing two-phase flow in heterogeneous porous media.

Optimal entropy condition

In [6], a functional measuring the total variation of the solution for each choice of (A,B) was defined and, subsequently, minimized to find the optimal connection (A_o,B_o).

Definition 3 *The optimal connection (A_o,B_o).*

- *For $f,g \in CC([0,1])$:*
  If $f(\theta_f) \leq g(\theta_g)$, then let $\overline{\theta}_f \geq \theta_g$ such that $f(\theta_f) = g(\overline{\theta}_f)$. Moreover, if $f(\theta_f) > g(\theta_g)$, then let $\overline{\theta}_g \leq \theta_f$ such that $f(\overline{\theta}_g) = g(\theta_g)$. Then

$$(A_o,B_o) = \begin{cases} (\theta_g,\overline{\theta}_g), & \text{if } f(\theta_f) \geq g(\theta_g), \\ (\overline{\theta}_f,\theta_f), & \text{if } f(\theta_f) \leq g(\theta_g). \end{cases} \tag{9.38}$$

- *For $f,g \in CV([0,1])$:*
  If $f(\theta_f) \leq g(\theta_g)$, then let $\overline{\theta}_g \geq \theta_f$ such that $f(\overline{\theta}_g) = g(\theta_g)$. Moreover, if $f(\theta_f) > g(\theta_g)$, then let $\overline{\theta}_f \leq \theta_g$ such that $f(\theta_f) = g(\overline{\theta}_f)$. Then

$$(A_o,B_o) = \begin{cases} (\theta_g,\overline{\theta}_g), & \text{if } f(\theta_f) \leq g(\theta_g), \\ (\overline{\theta}_f,\theta_f), & \text{if } f(\theta_f) \geq g(\theta_g). \end{cases} \tag{9.39}$$

In summary, A_o and B_o are chosen by simple consideration on $g(\theta_g)$ and $f(\theta_f)$.

With (A_o,B_o) given in the definition, the characteristic condition (9.36) reduces to [33]

$$\min\{0,g'(u^-(t))\}\max\{0,f'(u^+(t))\} = 0, \tag{9.40}$$

that is, the characteristics must always be traced back to the x-axis on at least one side of x_h.

The optimal entropy connection was derived from a mathematical point of view. However, it was noted that it corresponded exactly with the entropy condition derived by Kaasschieter in [88] where the heterogeneous Buckley-Leverett equation was considered. The entropy condition given in [88] was derived considering the vanishing capillary limit, that is, by studying the solution when the capillary pressure term (right-hand side of (9.16)) approaches zero. This was considered the physically relevant solution for two-phase fluid flow in a heterogeneous porous medium in [6, 114] and also in Papers D, E, and F.

Minimal jump condition

The minimal jump condition was introduced in [71]. Although the conditions on the flux functions given in [71] are more relaxed than for $f(u)$ and $g(u)$ given above, the entropy condition can still be adapted to the (A,B) framework.

Definition 4 *Let either $f,g \in CC([0,1])$ or $f,g \in CV([0,1])$. Then the minimal jump connection (A_m,B_m) is chosen such that*

$$|A_m - B_m| \tag{9.41}$$

is minimized.

With (A_m,B_m) given in the definition, the characteristic condition (9.36) reduces to [33]

$$\min\{0,g'(u^-(t))\}\max\{0,f'(u^+(t))\} = 0 \quad \text{if } u^-(t) \neq u^+(t). \tag{9.42}$$

Hence, the characteristics must lead back to the x-axis at least on one side of x_h, unless $u^-(t) = u^+(t)$, in which case there is no restriction on the characteristics.

In [71], the minimal jump connection was associated with the vanishing viscosity limit, that is, in the limit of a parabolic equation of type (9.37). It was recognized in [32] that the minimal jump connection was the correct entropy solution for the clarifier-thickener model, that is, a model where a source is present at x_h. It was also associated with two-phase fluid flow in a heterogeneous porous medium in, e.g., [72].

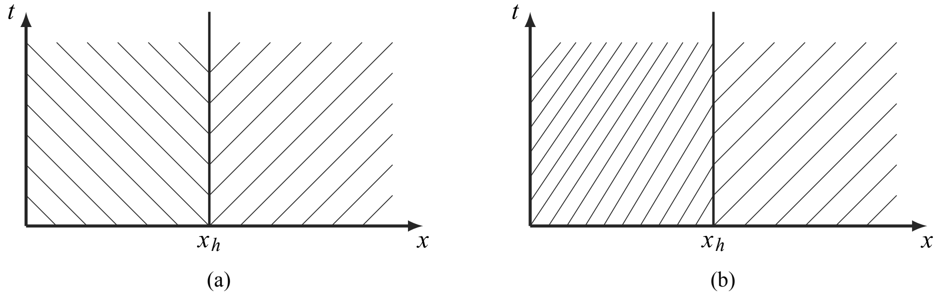

Figure 9.2: Illustration of (a) the MJ solution and (b) the OE solution around x_h for flux functions crossing in an undercompressive manner.

9.3.4 Discussion

In the previous section, we introduced two entropy conditions, the optimal entropy (OE) condition and minimal jump (MJ) condition, and indicated that both had been associated with two-phase flow in heterogeneous porous media. Hence, it is useful to discuss the difference between the two entropy solutions.

It can easily be shown that for $f,g \in CC([0,1])$ and $f,g \in CV([0,1])$ the OE and MJ condition produces different result only when flux functions cross in an *undercompressive manner*, that is, $f'(u_\chi) > 0$ and $g'(u_\chi) < 0$ for the intersection value u_χ. In this case, Definition 4 gives $A_m = B_m = u_\chi$, and, depending on the initial condition, the solution can thus be $u^-(t) = u^+(t) = u_\chi$. From (9.42) we see that the MJ condition produces characteristics that go out of x_h when $u^-(t) = u^+(t) = u_\chi$ (see Figure 9.2a). Consequently, the MJ solution is continuous across the interface x_h. From Definition 3, we see that $(A_o,B_o) \neq (u_\chi,u_\chi)$, and thus $(u^-(t),u^+(t)) \neq (u_\chi,u_\chi)$. The characteristic condition (9.40) also tells us that that the characteristics of an OE solution always lead back to the x-axis on at least one side of x_h (see Figure 9.2b).

As mentioned in Section 9.3.3, the OE condition has been considered the physically correct entropy condition for two-phase flow in heterogeneous porous media in many studies. However, in recent papers [8, 9], a new entropy condition for the heterogeneous Buckley-Leverett equation was suggested where one has to allow the interface to 'generate information' (similarly as seen for the MJ condition above). In short, the entropy condition presented in [8, 9] requires the knowledge about the capillary forces, which are neglected in the hetereogeneous Buckley-Leverett equation (i.e., in the limit of (9.16) to (9.17)), to compute the solution to (9.17). This entropy condition is different from the one presented in [88] (i.e., the OE condition), where knowledge of the capillary forces is not needed. Specifically, in [8, 9] the capillary pressure function must be continuous at the interface together with the flux function (c.f. (9.33)). The continuity of capillary pressure was regarded in [88] as only 'merely coincidental'.

Chapter 10

Numerical schemes

In most practical applications, conservation laws must be solved with a numerical method. It is then important that the numerical method honours the same principles underlying the conservation law. To wit, the conservation principle must be followed to allow for discontinuous solutions (shocks), and, moreover, the correct entropy solution must be captured. In the following, we present the basic concepts for a finite volume approximation of the conservation law with discontinuous flux function we studied in the previous chapter (confer (9.18)). Furthermore, we present two schemes, the Godunov and Engquist-Osher scheme, that provide solutions with respect to the (A,B)-connection defined in Section 9.3.2, and two schemes, the local Lax-Friedrichs and upstream mobility scheme, that are just extensions of their classical counterparts.

10.1 Finite volume methods

For simplicity, we discretize the spatial domain $\mathbb{R}$ into grid cells (or 'finite volumes') with equidistant width $\Delta x > 0$. The cell edges are given by

$$x_{j+1/2} = j\Delta x + x_h, \quad \text{for } j \geq 0, \qquad x_{j-1/2} = j\Delta x + x_h, \quad \text{for } j \leq 0,$$

where $x_{-1/2} = x_{1/2} = x_h$ is the interface (confer Chapter 9), and $j \in \mathbb{Z}$. The cell centers are given by

$$x_j = (j-1/2)\Delta x + x_h, \quad \text{for } j \geq 1, \qquad x_j = (j+1/2)\Delta x + x_h, \quad \text{for } j \leq -1.$$

The time domain $[0,T)$ is discretized into $n = 0,\dots,N$ equidistant time steps, i.e., $t^n = n\Delta t$ for $\Delta t > 0$.

To get a numerical method on conservation form, we integrate (9.18) over an arbitrary cell, $[x_{j-1/2}, x_{j+1/2}] \times [t^n, t^{n+1}]$, in the $x - t$ plane

$$\int_{x_{j-1/2}}^{x_{j+1/2}} u(x,t^{n+1})\,\mathrm{d}x - \int_{x_{j-1/2}}^{x_{j+1/2}} u(x,t^n)\,\mathrm{d}x + \int_{t^n}^{t^{n+1}} F(u(x_{j+1/2},t),x)\,\mathrm{d}t - \int_{t^n}^{t^{n+1}} F(u(x_{j-1/2},t),x)\,\mathrm{d}t = 0. \quad (10.1)$$

Since the first term describes the change of u in a time step and the second term describes the change of total flux in or out of $[x_{j-1/2}, x_{j+1/2}]$, (10.1) is a conservative

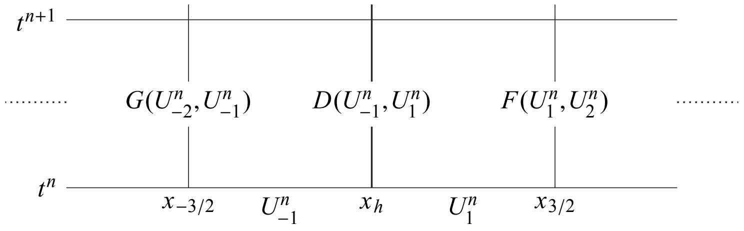

Figure 10.1: Illustration of the finite volume scheme around the interface x_h. Shown in the $x - t$ plane.

scheme. In general, we cannot evaluate the above integrals since we do not know the exact solution. However, by approximating the integrals, the numerical method will be based on the necessary conservation principles. A standard approach is to use the following averages [98]:

$$U^n_j \approx \frac{1}{\Delta x}\int_{x_{j-1/2}}^{x_{j+1/2}} u(x,t^n)\,\mathrm{d}x, \tag{10.2}$$

$$\mathcal{F}^n_{j+1/2} \approx \frac{1}{\Delta t}\int_{t^n}^{t^{n+1}} F(u(x_{j+1/2},t),x)\,\mathrm{d}t. \tag{10.3}$$

Using (10.2) and (10.3), (10.1) can be written as

$$U^{n+1}_j = U^n_j - \frac{\Delta t}{\Delta x}\left(\mathcal{F}^n_{j+1/2} - \mathcal{F}^n_{j-1/2}\right), \tag{10.4}$$

which is the basis for all (explicit) finite volume methods.

To get a fully discrete method, we need $\mathcal{F}^n$ to depend on U^n, and not on u. For hyperbolic problems, the information propagates with finite speed, hence, it is reasonable to approximate $\mathcal{F}^n_{j+1/2}$ based on the neighbouring values U^n_j and U^n_{j+1} [98]. To this end, we define the numerical flux function as

$$\mathcal{F}^n_{j+1/2} = \mathcal{F}(U^n_j,U^n_{j+1}) = \begin{cases} F(U^n_j,U^n_{j+1}), & \text{for } j \geq 1 \\ D(U^n_{-1},U^n_1), & \\ G(U^n_{j-1},U^n_j), & \text{for } j \leq -1. \end{cases} \tag{10.5}$$

Hence, $F(U^n_j,U^n_{j+1})$ approximates $f(u)$ for $x > x_h$, and $G(U^n_{j-1},U^n_j)$ approximates $g(u)$ for $x < x_h$. At $x = x_h$, however, special considerations must be made to ensure that the correct solution is approximated, which is taken care of by the function $D(U^n_{-1},U^n_1)$. See Figure 10.1 for a schematic of the finite volume scheme around the interface x_h.

From (10.4) and (10.5), we see that a finite volume method depends on how F, G, and D are defined. Hence, when we present the different finite volume schemes below, we will only describe their numerical flux functions. First, however, we briefly discuss the CFL-condition and give an overview of convergence criteria for (10.4).

10.1.1 CFL condition

In the previous section, we indicated that for hyperbolic problems the information propagates with finite speed and, hence, we could approximate the flux over a cell edge as a function of the neighbouring U^n values. However, this approximation is highly dependent on the speed of propagation, which is determined by the derivative of the flux functions, i.e., $f'(u)$ and $g'(u)$. Since the distance that a solution of (9.18) propagates in one time step is $f'(u)\Delta t$ for $x > x_h$ and $g'(u)\Delta t$ for $x < x_h$, the following restrictions must be met

$$\Delta t \max_{u\in[0,1]} |f'(u)| \leq \Delta x, \quad \Delta t \max_{u\in[0,1]} |g'(u)| \leq \Delta x. \tag{10.6}$$

This is called the Courant-Friedrich-Lewy (CFL) condition, and it is a necessary (but not sufficient) condition of a finite volume method to converge.

10.1.2 Convergence

When discussing the theory on conservation laws in Chapter 9, two aspects were important: a solution had to be on a weak form and it had to fulfill an entropy condition. It is clear that U^n also has to fulfill these conditions to converge to the true physically relevant solution. A useful theorem in this context is the Lax-Wendroff theorem [97], which states that if a solution U^n of a conservative scheme converges to a solution u in the limit of $\Delta x, \Delta t \to 0$, then u is a weak solution. Although an important theorem, it does not guarantee convergence, and if convergence is guaranteed, it does not say if the solution fulfills an entropy condition or not. Hence, more notions on convergence are needed.

The first notion needed is *consistency*, that is, the finite volume method needs to be consistent with the conservation law it approximates. For (10.4) this is satisfied if

$$F(\overline{u},\overline{u}) = f(\overline{u}), \quad G(\overline{u},\overline{u}) = g(\overline{u}). \tag{10.7}$$

For D, the notion of consistency is difficult since the flux through x_h is highly problem dependent. As we will note below, convergence can still be achieved, even though consistency is not fulfilled for D.

The second notion needed for convergence is *stability*. This is an important notion, as convergence cannot be ensured for an unstable finite volume method. There exists a wide variety of stability criteria that a finite volume method may fulfill, e.g., bounded variation, total variation, L_1-contraction, etc. For conservation laws with a continuous flux function (i.e., a classical conservation law), a particularly interesting property arise for monotone methods. Essentially, if a consistent finite volume method is a monotone method, then it converges to the correct entropy solution (i.e., it fulfills Oleiniks entropy condition) in the limit of $\Delta x, \Delta t \to 0$ [78]. Monotone methods are, however, at most first-order methods.

Unfortunately, for finite volume methods approximating a conservation law with a discontinuous flux function, as (10.4), monotonicity does not ensure that the method converges to the correct entropy solution. It can, however, be an important ingredient in a convergence proof. In any case, stability must always be fulfilled, and special care must be taken to ensure that U^n converges to the correct entropy solution, especially at the interface.

10.2 The Godunov scheme

From (10.3), it is seen that the numerical flux $\mathcal{F}^n_{j+1/2}$ depends on the solution $u(x_{j+1/2},t)$, which we cannot compute. However, it is possible to compute an approximation of $u(x_{j+1/2},t)$ by solving the Riemann problem (see Section 9.3) at $x_{j+1/2}$ with $U_L = U_j$ and $U_R = U_{j+1}$ as initial conditions. In most cases, the solution to the Riemann problem is either a shock or a rarefaction wave moving entirely to the right or left of $x_{j+1/2}$, hence, the solution is U_L or U_R, respectively. The exception is when U_L and U_R is on either side of an extrema, and the solution is a rarefaction wave. In this case, the solution will be the value of the extrema; θ_f for $f(u)$, or θ_g for $g(u)$, see Section 9.3.2. If the solution is a shock with speed equal to zero, then the solution is just given by the initial condition.

By inserting the above solutions into (10.3), the flux approximations for cell edges away from x_h can be written compactly as [73]

$$H(U_L,U_R) = \begin{cases} \min_{u\in[U_L,U_R]} h(u), & \text{for } U_L \leq U_R, \\ \max_{u\in[U_R,U_L]} h(u), & \text{for } U_R \leq U_L, \end{cases} \tag{10.8}$$

where H is either F for $x > x_h$ or G for $x < x_h$, and $h(u)$ is either $f(u)$ for $x > x_h$ or $g(u)$ for $x < x_h$.

At x_h, we can do similar consideration for the Riemann problem as we did above, but now $U_L = U_{-1}$ is associated with G and $U_R = U_1$ is associated with F. In addition, we need to consider the entropy solutions of type (A,B) introduced in Section 9.3.2. To capture the correct flux function value at the interface the authors in [6] gave the following expressions

$$D(U_L,U_R) = \min\{G(U_L,A),F(U_R,B)\}, \quad \text{for } f,g \in CC([0,1]), \tag{10.9}$$
$$D(U_L,U_R) = \max\{G(U_L,A),F(U_R,B)\}, \quad \text{for } f,g \in CV([0,1]), \tag{10.10}$$

where the definitions of $CC([0,1])$ and $CV([0,1])$ can be found in Section 9.3.2. Since the Godunov scheme approximated the flux functions using the Riemann solutions at each cell edge, it is often referred to as the exact Riemann solver.

It is easily seen from (10.8) that the scheme is consistent for $x \neq x_h$. However, from (10.9) and (10.10) it is seen that there are some cases where $D(\overline{u},\overline{u})$ is not equal to either $f(\overline{u})$ or $g(\overline{u})$ (in that case, $D(\overline{u},\overline{u}) = g(A) = f(B)$). Hence, $D(U_L,U_R)$ is not consistent. In spite of the non-consitency, convergence of the Godunov scheme to entropy solutions of type (A,B) was proven in [6].

Note that for $f(u)$ and $g(u)$ of type $CC([0,1])$ or $CV([0,1])$, the expressions in (10.8), (10.9), and (10.10) can be simplified for implementation purposes, see, e.g., [5] or Papers D, E, and F.

10.3 The Engquist-Osher scheme

The basic idea behind the Engquist-Osher scheme is to generalize the standard upwind method for conservation laws with linear flux functions by assuming that the solution is always a rarefaction wave. This can by done by defining the numerical flux functions as

$$H(U_L,U_R) = \frac{1}{2}\left[h(U_L) + h(U_R) - \int_{U_L}^{U_R} |h'(w)|\,\mathrm{d}w\right], \tag{10.11}$$

where H is either F for $x > x_h$ or G for $x < x_h$, and $h(u)$ is either $f(u)$ for $x > x_h$ or $g(u)$ for $x < x_h$. Furthermore, we have introduced the notation $U_L = U_j$ and $U_R = U_{j+1}$. In most cases, (10.11) produces the same result as (10.8). It is only in the case of a transonic shock, i.e., when $h'(U_L) > 0 > h'(U_R)$, that (10.11) produces a result different than (10.8). In this case,

$$H(U_L,U_R) = h(U_L) + h(U_R) - h(\theta_h),$$

where θ_h is either θ_f or $\theta_g$depending on if $x > x_h$ or $x < x_h$, respectively. The advantage with (10.11) compared to (10.8) is that no solution of the Riemann problem is needed.

To capture the entropy solutions of type (A,B) given in Section 9.3.2, Bürger *et al.* [33] proposed that the numerical flux at x_h should be given as

$$H(U_L,U_R) = \frac{1}{2}\left[\tilde{g}(U_L) + \tilde{f}(U_R) - \int_{B}^{U_R} |\tilde{f}(w)|\,\mathrm{d}w + \int_{A}^{U_L} |\tilde{g}(w)|\,\mathrm{d}w\right], \tag{10.12}$$

where

$$\tilde{f}(u) = \min\{f(u),f(B)\},\ \tilde{g}(u) = \min\{g(u),f(A)\}, \quad \text{for } f,g \in CC([0,1]), \tag{10.13}$$

$$\tilde{f}(u) = \max\{f(u),f(B)\},\ \tilde{g}(u) = \max\{g(u),f(A)\}, \quad \text{for } f,g \in CV([0,1]). \tag{10.14}$$

The definitions of $CC([0,1])$ and $CV([0,1])$ can be found in Section 9.3.2.

Similarly as for the Godunov scheme, the Engquist-Osher scheme is consistent away from x_h, and not consistent for $x = x_h$ (for some configurations, $D(\overline{u},\overline{u}) = g(A) = f(B)$). Nevertheless, the scheme converges to an entropy solution of type (A,B) as shown in [33].

Note that the expressions (10.11) and (10.12) can be simplified for flux functions satisfying $CC([0,1])$ and $CV([0,1])$; see Paper D.

10.4 The local Lax-Friedrichs scheme

The simplest flux approximation possible is the Lax-Friedrichs scheme. Let $U_L = U_j$ and $U_R = U_{j+1}$ be on either side of an arbitrary cell edge $x_{j+1/2}$. The numerical flux in the Lax-Friedrichs scheme is then given by

$$H(U_L,U_R) = \frac{1}{2}[h(U_L) + h(U_R) - \frac{\Delta x}{\Delta t}(U_R - U_L)], \tag{10.15}$$

where H is either F for $x > x_h$ or G for $x < x_h$, and $h(u)$ is either $f(u)$ for $x > x_h$ or $g(u)$ for $x < x_h$. From [149] it seen that the Lax-Friedrichs scheme produces the most allowable numerical viscosity (i.e., smearing of the numerical solution) of any convergent method.

An improvement of the Lax-Friedrichs scheme is obtained by replacing $\Delta x/\Delta t$ in (10.15) with a locally determined value, a_h, [98]

$$H(U_L,U_R) = \frac{1}{2}\left[h(U_L) + h(U_R) - a_h(U_R - U_L)\right], \tag{10.16}$$

where

$$a_h = \max\{|h'(u)|\}, \qquad \forall u \in [U_L,U_R]. \tag{10.17}$$

This scheme is often called the local Lax-Friedrichs scheme, although it was introduced by Rusanov [133] and is thus also known as the Rusanov method. Comparing (10.15) and (10.16), it can be shown that the numerical viscosity produced by the local Lax-Friedrichs scheme is at worst equal to the Lax-Friedrichs scheme.

At the interface x_h, (10.16) can be applied in a straightforward manner

$$D(U_L,U_R) = \frac{1}{2}\left[g(U_L) + f(U_R) - \overline{a}(U_R - U_L)\right], \tag{10.18}$$

where

$$\overline{a} = \max\{|f'(u)|,|g'(u)|\}, \qquad \forall u \in [U_L,U_R]. \tag{10.19}$$

It is easily seen from (10.16) that the local Lax-Friedrichs scheme is consistent for $x \neq x_h$. At x_h, the scheme is not consistent ($D(\overline{u},\overline{u}) = \frac{1}{2}[g(\overline{u}) + f(\overline{u})]$), similarly to the observations seen for the Godunov and the Engquist-Osher scheme. For conservation laws with a continuous flux function (i.e., classical conservation laws) the local Lax-Friedrichs scheme is monotone and thus converges to the correct entropy solution [98]. For conservation laws with a discontinuous flux function, convergence has not yet been proven for the interface flux (10.18). However, results in Paper D indicated that the scheme converges to the minimal jump solution (confer Section 9.3.3).

10.4.1 Discussion

Recently, a modified version of the Lax-Friedrichs and local Lax-Friedrichs scheme was shown to converge to entropy solutions of type (A,B) [4]. The modified schemes relied on the construction of an interface flux function in which a numerical approximation could be given. Furthermore, the numerical flux functions immediately to the right and left of the interface x_h, $F(U_1,U_2)$ and $G(U_{-1},U_{-2})$, respectively, also had to be modified. As of yet, the connection between the local Lax-Friedrichs scheme given in the previous section and the scheme presented in [4] is unclear.

10.5 The upstream mobility scheme

To predict fluid flow in porous media, an ad hoc flux approximation was invented by petroleum engineers from simple physical considerations. Recall the notation given in Sections 9.1 and 9.2 for the two-phase flow problem in heterogeneous porous media. The quantity U will in this section thus be a numerical approximation of the saturation u. The numerical flux functions away from the interface x_h can be written as

$$H(U_L,U_R) = \frac{\lambda_w^{L,R*}}{\lambda_w^{L,R*} + \lambda_{nw}^{L,R*}}[q + (\gamma_w - \gamma_{nw})\lambda_{nw}^{L,R*}]. \tag{10.20}$$

The superscript ‘R’ denotes $x > x_h$ (where H is equal to F), and ‘L’ denotes $x < x_h$ (where H is equal G). The mobility functions $\lambda_\alpha^{L,R*}(U)$ must be evaluated with the correct upstream saturation, hence, the superscript ‘*’. The direction of flow can be found from a discretization of Darcy’s law (confer (9.7)), thus $\lambda^{L,R*}$ is evaluated using the upstream saturation U in cell $j \pm 1$ if

$$p_{j\pm1} - p_j \mp \gamma_\alpha \Delta x > 0, \tag{10.21}$$

and in cell j otherwise. We can remove pressure from the equation with similar manipulations as in Section 9.2. Omitting the details, we get the following equivalence

$$p_{j\pm1} - p_j \mp \gamma_l \Delta x \quad \Longleftrightarrow \quad q + (\gamma_l - \gamma_k)\lambda_k, \quad \text{for } k \neq l. \tag{10.22}$$

The direction of flow is now determined by advection, q, and the buoyancy effects, $(\gamma_l - \gamma_k)\lambda_k$. Using the above equivalence, the mobility functions $\lambda_\alpha^{L,R*}(U)$ can thus be evaluated as

$$\lambda_l^{L,R*} = \begin{cases} \lambda_l^{L,R}(U_L), & \text{if } q + (\gamma_l - \gamma_k)\lambda_k^{L,R*} \geq 0, \\ \lambda_l^{L,R}(U_R), & \text{if } q + (\gamma_l - \gamma_k)\lambda_k^{L,R*} \leq 0, \end{cases} \tag{10.23}$$

for $k \neq l$. Note that this is an implicit formula ($\lambda_l^{L,R*}$ is dependent on $\lambda_k^{L,R*}$, and vice versa). For implementation purposes an explicit formula was given in [29]. After the mobility functions are evaluated, they are inserted into (10.20) to produce the numerical flux functions away from x_h.

To calculate the numerical flux function at the interface, (10.20) and (10.23) are applied in a straightforward manner

$$D(U_L, U_R) = \frac{\lambda_w^*}{\lambda_w^* + \lambda_{nw}^*}[q + (\gamma_w - \gamma_{nw})\lambda_{nw}^*], \tag{10.24}$$

where

$$\lambda_l^* = \begin{cases} \lambda_l^L(U_L), & \text{if } q + (\gamma_l - \gamma_k)\lambda_k^* \geq 0, \\ \lambda_l^R(U_R), & \text{if } q + (\gamma_l - \gamma_k)\lambda_k^* \leq 0. \end{cases} \tag{10.25}$$

Similarly as for (10.23), the above formula for calculating the mobility functions can be made explicit, see [114].

To check the consistency, recall that H and D are numerical approximations of the flux functions given for two-phase flow in porous media. Indeed, away from the interface, the flux approximation (10.20) is consistent with f and g given by (9.13) and (9.14), respectively. At the interface, we see from (10.25) that the mobility functions on either side of x_h might be used depending on the upstream direction of flow for each phase. Hence, in some cases D might reduce to F or G, but in the case where the upstream direction of flow for each phase is opposite (i.e. countercurrent flow), λ^L will be given for one phase and λ^R for the other. Thus, the upstream mobility scheme is not consistent at x_h, which was also the case for the Godunov, Engquist-Osher, and local Lax-Friedrichs scheme.

It can be shown that the upstream mobility scheme converges to the true entropy solution in the case of two-phase flow in *homogeneous* porous media, see, e.g., [135]. In the case of two-phase flow in *heterogeneous* porous media, convergence of the upstream mobility scheme has not yet been proven. In [114] it was shown that the scheme is stable, and converges to a weak solution (by the Lax-Wendroff theorem, confer Section 10.1.2).

Chapter 11

Summary of the papers

In this chapter, the main results of Paper D – F associated with Part II of this thesis are given. The summaries will be based on the scientific background presented in Part II.

11.1 Summary of Paper D

Title:	*Errors in the upstream mobility scheme for countercurrent two-phase flow in heterogeneous porous media*
Authors:	S. Tveit and I. Aavatsmark

In Paper D, a numerical investigation of the widely applied upstream mobility scheme (see Section 10.5) was conducted for two-phase fluid flow in a 1D heterogeneous porous medium (see Section 9.2). The scheme had been shown to fail for a pure gravity segregation problem [114], however, the errors observed in Paper D were larger in magnitude, when advection and gravity segregation were included. A secondary objective of the paper was to investigate the behaviour of the local Lax-Friedrichs scheme introduced in Section 10.4.

From Section 9.2 we know that fluid flow in a heterogeneous porous medium is modeled as conservation law with a discontinuous flux function (see Section 9.3). In Section 9.3.3, two physically relevant entropy solutions were presented for the two-phase flow problem, the optimal entropy connection and minimal jump connection. In Paper D, we followed consensus (see, e.g., [5, 114]) and considered the optimal entropy solution as the physically relevant solution, which has later been disputed (confer the discussion in Section 9.3.4 and [9]).

To evaluate the performance of the upstream mobility scheme, it was compared to the Godunov and Engquist-Osher scheme presented in Section 10.2 and Section 10.3, respectively. Both schemes were set to approximate the optimal entropy condition. Since the Godunov scheme is considered an exact Riemann solver, it acted as the true solution in the numerical experiments.

The numerical experiments were set up to model a countercurrent flow situation, that is, the two phases flow in opposite direction. With this flow situation it was shown that the upstream mobility scheme produced three types of errors: ‘spike’ solutions; solutions following the minimal jump connection; and solutions which followed neither the optimal entropy solution nor the minimal jump solution.

The ‘spike’ solution produced by the upstream mobility scheme was a ‘spike’ at the interface, x_h, that persisted even when the grid cell size was reduced, indicating that it would still be present as the grid cell size approached zero.

From the discussion made in Section 9.3.4, we know that the minimal jump solution only differs from the optimal entropy condition (for our type of flux functions, see Section 9.3.2) in the case where the flux functions cross in an undercompressive manner. Moreover, it can then produce a continuous solution across the interface x_h. This erroneous solution was observed for the upstream mobility scheme when the flux functions were made such that $\lambda_w^L(u_\chi) = \lambda_w^R(u_\chi)$ and $\lambda_{nw}^L(u_\chi) = \lambda_{nw}^R(u_\chi)$ (where u_χ is the intersection value for the flux functions, see Section 9.3.4). Although it is easy to see why the upstream mobility scheme must produce the minimal jump solution in the case of pure gravity segregation, the same conclusions could not be made for the flow situation in the Paper D experiment.

By changing the flux functions slightly from the previous experiment, the upstream mobility scheme produced a solution which did not follow the optimal entropy solution nor the minimal jump solution. The deviation of the upstream mobility solution from the optimal entropy solution was large. Moreover, the solution produced by the upstream mobility scheme looked reasonable with no apparent unusual behaviour when looked upon isolated. We also observed that the mobility functions fulfilled: $\lambda_w^L(U_{-1}) = \lambda_w^R(U_1)$ and $\lambda_{nw}^L(U_{-1}) = \lambda_{nw}^R(U_1)$, where U_{-1} and U_1 is the last saturation value to the left and right of x_h, respectively (see Figure 10.1). Note that this was only an observation and not a condition for the upstream mobility scheme.

The behaviour of the local Lax-Friedrichs scheme was studied in the same numerical experiments. It seemed to follow the minimal jump condition in all the experiments, with some numerical diffusion.

11.2 Summary of Paper E and F

Paper E:
Title: *Errors in the upstream mobility scheme for counter-current two-phase flow with discontinuous permeabilities*
Authors: T. S. Mykkeltvedt, I. Aavatsmark, and S. Tveit

Paper F:
Title: *On the performance of the upstream mobility scheme applied to counter-current two-phase flow in a heterogeneous porous medium*
Authors: T. S. Mykkeltvedt, I. Aavatsmark, and S. Tveit

In Papers E and F, the upstream mobility scheme was applied on numerical experiments involving flow of CO_2 and brine in a 1D hetereogeneous porous medium. The numerical experiments were conducted using realistic parameters for the rock and fluids under consideration. This involved relative permeability modeled by cubic spline interpolation, based on the classical Brooks-Corey [30] and van-Genuchten relations [155]. Similarly to Paper D, the upstream mobility scheme was compared to a Godunov scheme that approximated the optimal entropy condition.

The numerical experiments were set up to model the flow of CO_2 and brine in a vertical column. Since the non-wetting CO_2 phase is less dense than wetting brine phase, it will flow upwards, and thus a countercurrent flow situation occurs. Two scenarios were considered: one where a single interface was regarded (as in Paper D) and one where two interfaces were regarded. In the case where a single interface was present, similar erroneous behaviour as seen in Paper D was observed. The upstream mobility scheme produced solutions following the minimal jump condition, and solutions which followed neither the optimal entropy nor the minimal jump condition. In particular, it was observed that a small perturbation in the relative permeability produced noticeably different solutions, which indicates that the upstream mobility scheme is ill-conditioned.

When two interfaces were included in the experiments, it was clearly seen that upstream mobility scheme could produce solutions which, when looked upon isolated, appeared reasonable but was erroneous compared to the true solution.

Contribution as co-author: Experience and suggestions for the research and paper.

Appendix A

Complex Gaussian distribution

The following appendix gives a brief overview of the theory on complex random vectors with a focus on complex Gaussian distributions. For the most part we will follow the textbook [141].

Let **u** and **v** be two random vectors in $\mathbb{R}^n$, and let **z** be a *real composite* random vector in $\mathbb{R}^{2n}$ given by

$$\mathbf{z} = \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix}. \tag{A.1}$$

Furthermore, let **x** be a *complex* random vector in $\mathbb{C}^n$ given by

$$\mathbf{x} = \mathbf{u} + i\mathbf{v}, \tag{A.2}$$

where $i = \sqrt{-1}$. The complex conjugate of **x** is denoted $\mathbf{x}^* = \mathbf{u} - i\mathbf{v}$. With **x** and $\mathbf{x}^*$ we can define the *complex augmented* random vector, $\underline{\mathbf{x}} \in \mathbb{C}_*^{2n}$, as

$$\underline{\mathbf{x}} = \begin{bmatrix} \mathbf{x} \\ \mathbf{x}^* \end{bmatrix}. \tag{A.3}$$

A transformation from **z** to $\underline{\mathbf{x}}$ can easily be made with the *real-to-complex* transform matrix

$$\mathbf{T}_n = \begin{bmatrix} \mathbf{I}_n & i\mathbf{I}_n \\ \mathbf{I}_n & -i\mathbf{I}_n \end{bmatrix}, \tag{A.4}$$

which has the property $\mathbf{T}_n^{-1} = \frac{1}{2}\mathbf{T}_n^H$ where H is the Hermetian. The real-to-complex transformation thus becomes

$$\underline{\mathbf{x}} = \mathbf{T}_n\mathbf{z} \quad \Leftrightarrow \quad \mathbf{z} = \frac{1}{2}\mathbf{T}_n^H\underline{\mathbf{x}}. \tag{A.5}$$

$\underline{\mathbf{x}}$ may seem as a redundant description of **z**, but, as we will see below, it provides a valuable description of what a general complex Gaussian distribution looks like.

A Gaussian distribution is fully described by its mean and covariance matrix, i.e, by second-order statistics. Second-order statistical description of **z** and $\underline{\mathbf{x}}$ thus follows. The mean of **z** is given by

$$\boldsymbol{\mu}_z = E[\mathbf{z}] = \begin{bmatrix} E[\mathbf{u}] \\ E[\mathbf{v}] \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_u \\ \boldsymbol{\mu}_v \end{bmatrix}, \tag{A.6}$$

and its covariance matrix is

$$\mathbf{C}_{zz} = E[(\mathbf{z} - \boldsymbol{\mu}_z)(\mathbf{z} - \boldsymbol{\mu}_z)^T] = \begin{bmatrix} \mathbf{C}_{uu} & \mathbf{C}_{uv} \\ \mathbf{C}_{vu} & \mathbf{C}_{vv} \end{bmatrix}, \tag{A.7}$$

where $\mathbf{C}_{uu} = E[(\mathbf{u} - \boldsymbol{\mu}_u)(\mathbf{u} - \boldsymbol{\mu}_u)^T]$, $\mathbf{C}_{uv} = E[(\mathbf{u} - \boldsymbol{\mu}_u)(\mathbf{v} - \boldsymbol{\mu}_v)^T] = \mathbf{C}_{vu}^T$, and $\mathbf{C}_{vv} = E[(\mathbf{v} - \boldsymbol{\mu}_v)(\mathbf{v} - \boldsymbol{\mu}_v)^T]$.

The mean vector of $\underline{\mathbf{x}}$ is

$$\underline{\boldsymbol{\mu}}_x = E[\underline{\mathbf{x}}] = \begin{bmatrix} E[\mathbf{x}] \\ E[\mathbf{x}^*] \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_x^* \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_u + i\boldsymbol{\mu}_v \\ \boldsymbol{\mu}_u - i\boldsymbol{\mu}_v \end{bmatrix}, \tag{A.8}$$

where we note that $\underline{\boldsymbol{\mu}}_x = \mathbf{T}_n\boldsymbol{\mu}_z$. The covariance matrix of $\underline{\mathbf{x}}$ is given by

$$\underline{\mathbf{C}}_{xx} = E[(\underline{\mathbf{x}} - \underline{\boldsymbol{\mu}}_x)(\underline{\mathbf{x}} - \underline{\boldsymbol{\mu}}_x)^H] = \begin{bmatrix} \mathbf{C}_{xx} & \widetilde{\mathbf{C}}_{xx} \\ \widetilde{\mathbf{C}}_{xx}^* & \mathbf{C}_{xx}^* \end{bmatrix}, \tag{A.9}$$

where we note that $\underline{\mathbf{C}}_{xx} = \mathbf{T}_n\mathbf{C}_{zz}\mathbf{T}_n^H$ and $\underline{\mathbf{C}}_{xx}^H = \underline{\mathbf{C}}_{xx}$. The two covariance matrices in (A.9) are the *complex covariance matrix*,

$$\mathbf{C}_{xx} = E[(\mathbf{x} - \boldsymbol{\mu}_x)(\mathbf{x} - \boldsymbol{\mu}_x)^H] = \mathbf{C}_{uu} + \mathbf{C}_{vv} + i(\mathbf{C}_{vu} - \mathbf{C}_{uv}) \tag{A.10}$$

and the *complementary complex covariance matrix* (also called *pseudo-covariance matrix*, *conjugate covariance matrix*, and *relation matrix*),

$$\widetilde{\mathbf{C}}_{xx} = E[(\mathbf{x} - \boldsymbol{\mu}_x)(\mathbf{x} - \boldsymbol{\mu}_x)^T] = \mathbf{C}_{uu} - \mathbf{C}_{vv} + i(\mathbf{C}_{vu} + \mathbf{C}_{uv}), \tag{A.11}$$

respectively. The last equalities in (A.10) and (A.11) follows from $\mathbf{T}_n\mathbf{C}_{zz}\mathbf{T}_n^H$.

We now claim that a complete second-order statistical characterization of a complex random vector $\mathbf{x}$ is given in terms of $\underline{\boldsymbol{\mu}}_x$ and $\underline{\mathbf{C}}_{xx}$. We will not provide a rigorous proof but rather sketch an idea of why it must be true. Consider the question of uncorrelated random vectors. Two real vectors $\mathbf{z}$ and $\mathbf{w}$ are uncorrelated if $\mathbf{C}_{zw} = \mathbf{0}$. If we let $\mathbf{z} = [\mathbf{u}^T, \mathbf{v}^T]^T$ and $\mathbf{w} = [\mathbf{a}^T, \mathbf{b}^T]^T$, $\mathbf{C}_{zw} = \mathbf{0}$ implies that $\mathbf{C}_{ua} = \mathbf{C}_{ub} = \mathbf{C}_{va} = \mathbf{C}_{ub} = \mathbf{0}$. Now, if we let $\mathbf{x} = \mathbf{u} + i\mathbf{v}$ and $\mathbf{y} = \mathbf{a} + i\mathbf{b}$, the complex cross-covariance matrix between $\mathbf{x}$ and $\mathbf{y}$ is given by

$$\mathbf{C}_{xy} = \mathbf{C}_{ua} + \mathbf{C}_{vb} + i(\mathbf{C}_{va} - \mathbf{C}_{ub}). \tag{A.12}$$

Hence, $\mathbf{C}_{xy} = \mathbf{0}$ only implies $\mathbf{C}_{ua} = -\mathbf{C}_{vb}$ and $\mathbf{C}_{va} = \mathbf{C}_{ub}$. This does not lead to $\mathbf{x}$ and $\mathbf{y}$ being uncorrelated since there can be some dependence between their real and complex parts. For $\mathbf{x}$ and $\mathbf{y}$ to be completely uncorrelated the complementary cross-covariance matrix, $\widetilde{\mathbf{C}}_{xy}$, must also be $\mathbf{0}$, since

$$\widetilde{\mathbf{C}}_{xy} = \mathbf{C}_{ua} - \mathbf{C}_{vb} + i(\mathbf{C}_{va} + \mathbf{C}_{ub}) = \mathbf{0} \tag{A.13}$$

implies $\mathbf{C}_{ua} = \mathbf{C}_{vb}$ and $\mathbf{C}_{va} = -\mathbf{C}_{ub}$. Together with the result for $\mathbf{C}_{xy} = \mathbf{0}$ above we have $\mathbf{C}_{ua} = \mathbf{C}_{ub} = \mathbf{C}_{va} = \mathbf{C}_{ub} = \mathbf{0}$. Since the augmented cross-covariance matrix between $\mathbf{x}$ and $\mathbf{y}$ is

$$\underline{\mathbf{C}}_{xy} = \begin{bmatrix} \mathbf{C}_{xy} & \widetilde{\mathbf{C}}_{xy} \\ \widetilde{\mathbf{C}}_{xy}^* & \mathbf{C}_{xy}^* \end{bmatrix}, \tag{A.14}$$

$\mathbf{C}_{xy} = \mathbf{0}$ and $\widetilde{\mathbf{C}}_{xy} = \mathbf{0}$ is automatically fulfilled if $\underline{\mathbf{C}}_{xy} = \mathbf{0}$. Hence, the augmented covariance matrix provides second-order statistical information of complex random vectors.

We can now derive the general complex Gaussian probability density function (PDF). Towards this end, let again $\mathbf{z} = [\mathbf{u}^T, \mathbf{v}^T]^T \in \mathbb{R}^{2n}$ such that $\mathbf{x} = \mathbf{u} + i\mathbf{v} \in \mathbb{C}^n$ and $\underline{\mathbf{x}} = \mathbf{T}_n\mathbf{z} \in \mathbb{C}_*^{2n}$. The multivariate Gaussian PDF for $\mathbf{z}$ is given by

$$f(\mathbf{z}) = \frac{1}{(2\pi)^n\sqrt{\det \mathbf{C}_{zz}}} \exp\left(-\frac{1}{2}(\mathbf{z} - \boldsymbol{\mu}_z)^T \mathbf{C}_{zz}^{-1}(\mathbf{z} - \boldsymbol{\mu}_z)\right). \tag{A.15}$$

(Note that this is equal to the joint PDF between $\mathbf{u}$ and $\mathbf{v}$.) From [154] it can be shown that $\mathbf{C}_{zz}^{-1} = \mathbf{T}^H\underline{\mathbf{C}}_{xx}^{-1}\mathbf{T}$ and $\det \underline{\mathbf{C}}_{xx} = 2^{2n} \det \mathbf{C}_{zz}$. Inserting these identities into (A.15) yields

$$f(\mathbf{x}) = \frac{1}{\pi^n\sqrt{\det \underline{\mathbf{C}}_{xx}}} \exp\left(-\frac{1}{2}(\underline{\mathbf{x}} - \underline{\boldsymbol{\mu}}_x)^H \underline{\mathbf{C}}_{xx}^{-1}(\underline{\mathbf{x}} - \underline{\boldsymbol{\mu}}_x)\right). \tag{A.16}$$

This is often called the *generalized complex Gaussian PDF* (it is, however, no more general than (A.15)).

An important special case of (A.16) can be found when $\widetilde{\mathbf{C}}_{xx} = \mathbf{0}$. From (A.11), $\widetilde{\mathbf{C}}_{xx} = \mathbf{0}$ leads to

$$\mathbf{C}_{uu} = \mathbf{C}_{vv}, \tag{A.17}$$

$$\mathbf{C}_{uv} = -\mathbf{C}_{vu}. \tag{A.18}$$

The first equation implies that $\mathbf{u}$ and $\mathbf{v}$ have equal variance, and the second equation requires $\mathbf{C}_{uv}$ to have zero diagonal elements while the off-diagonal may be nonzero (i.e., u_i and v_i are uncorrelated, but u_i and v_j can be correlated for $i \neq j$). Furthermore, inserting (A.17) and (A.18) into (A.10) leads to

$$\mathbf{C}_{xx} = 2[\mathbf{C}_{uu} - i\mathbf{C}_{uv}] = 2[\mathbf{C}_{vv} + i\mathbf{C}_{vu}]. \tag{A.19}$$

From (A.9) we see that the augmented covariance matrix, $\underline{\mathbf{C}}_{xx}$, is now a block diagonal matrix containing $\mathbf{C}_{xx}$ and $\mathbf{C}_{xx}^*$. Inserting this into (A.16), and after some manipulations (see [154]) yields

$$f(\mathbf{x}) = \frac{1}{\pi^n \det \mathbf{C}_{xx}} \exp\left(-(\mathbf{x} - \boldsymbol{\mu}_x)^H \mathbf{C}_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x)\right). \tag{A.20}$$

This Gaussian PDF only gives a complete characterization of complex random vectors *in the special case when* $\widetilde{\mathbf{C}}_{xx} = \mathbf{0}$.

Objective functions in a number of different classical inversion approaches involving complex-valued vectors (CSEM in particular) are derived from (A.20) and are on the form

$$(\mathbf{x} - \boldsymbol{\mu}_x)^H \mathbf{C}_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x),$$

where $\mathbf{C}_{xx}$ is often a diagonal matrix containing, e.g., estimates of the variance of measurement noise. From the discussion above, a diagonal $\mathbf{C}_{xx}$ implies that the real ($\mathbf{u}$) and imaginary parts ($\mathbf{v}$) of $\mathbf{x}$ must have equal variance, which, in some cases, may not be desirable.

Appendix B

Model parameter update in an ensemble version of ACKF

In this chapter, we will discuss an ensemble version of ACKF presented in [53], and in particular derive an update equation for the ensemble of model parameters in a similar manner as we did in Section 3.3.4. From Appendix A we know that there is an equivalence between the real-composite vector and complex augmented vector, and based on these two vector forms it is possible to express a generalized complex Gaussian PDF in two, equally valid, ways. As we pointed out in the Discussion after Section 3.3.3 and shown formally in [53], the ACKF and real-valued KF are also equally valid forms of the same problem. Hence, the update equation for the ensemble of model parameters given in this appendix will be an equally valid form of the one given in Section 3.3.4. The following presentation relies on notation introduced in Appendix A and Chapter 3 (Section 3.3.3 and Section 3.3.4 in particular).

For ease of reading we give the analysis equation of the ACKF (without derivation) [53]

$$\underline{\boldsymbol{\psi}}_k^a = \underline{\boldsymbol{\psi}}_k^f + \underline{\mathbf{K}}_k(\underline{\mathbf{d}}_k - \underline{\mathbf{H}}_k\underline{\boldsymbol{\psi}}_k^f), \tag{B.1}$$

where

$$\underline{\mathbf{K}}_k = \underline{\mathbf{C}}_{\psi_k}^f \underline{\mathbf{H}}_k^H (\underline{\mathbf{H}}_k \underline{\mathbf{C}}_{\psi_k}^f \underline{\mathbf{H}}_k^H + \underline{\mathbf{C}}_{d_k})^{-1}, \tag{B.2}$$

In our application, $\underline{\boldsymbol{\psi}}_k$ is a joint-state vector defined as

$$\underline{\boldsymbol{\psi}}_k = \begin{bmatrix} \tilde{\mathbf{g}}_k(\mathbf{m}_{k-1}) \\ \mathbf{m}_k \\ \tilde{\mathbf{g}}_k^*(\mathbf{m}_{k-1}) \\ \mathbf{m}_k \end{bmatrix} \tag{B.3}$$

with a corresponding matrix

$$\underline{\mathbf{H}}_k = \begin{bmatrix} \mathbf{I}_{N_{d_k}/2} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{N_{d_k}/2} & \mathbf{0} \end{bmatrix} \tag{B.4}$$

Note that $\mathbf{m}$ is a real-valued vector, thus $\mathbf{m}^* = \mathbf{m}$.

Let the ensemble of complex forward model outputs be given as $\widetilde{\mathbf{G}}_k(\mathbf{M}_{k-1}) = [\tilde{\mathbf{g}}_k(\mathbf{m}_{k-1}^1), \ldots, \tilde{\mathbf{g}}_k(\mathbf{m}_{k-1}^{N_e})] \in \mathbb{C}^{(N_{d_k}/2)\times N_e}$. Let the complex augmented ensemble matrix, $\underline{\boldsymbol{\Psi}}_k \in \mathbb{C}_*^{N_{\underline{\psi}_k}\times N_e}$ be given by

$$\underline{\boldsymbol{\Psi}}_k = \begin{bmatrix} \widetilde{\mathbf{G}}_k(\mathbf{M}_{k-1}) \\ \mathbf{M}_k \\ \widetilde{\mathbf{G}}_k^*(\mathbf{M}_{k-1}) \\ \mathbf{M}_k \end{bmatrix}. \tag{B.5}$$

where $N_{\underline{\psi}_k} = N_{d_k} + 2N_m$. Furthermore, let the ensemble of augmented observed data be given as $\underline{\mathbf{D}}_k = [\underline{\mathbf{d}}_k^1,\dots,\underline{\mathbf{d}}_k^{N_e}] \in \mathbb{C}_*^{N_{d_k}\times N_e}$, where $\underline{\mathbf{d}}_k^j \sim \mathcal{N}(\underline{\mathbf{d}}_k^{true},\underline{\mathbf{C}}_{d_k})$. The relationship between $\underline{\boldsymbol{\Psi}}_k$ and $\underline{\mathbf{D}}_k$ is given in a similar manner as (3.59),

$$\underline{\mathbf{D}}_k = \underline{\mathbf{H}}_k\underline{\boldsymbol{\Psi}}_k + \underline{\mathbf{E}}_k^d, \tag{B.6}$$

where $\underline{\mathbf{E}}_k^d = [(\underline{\boldsymbol{\epsilon}}_k^d)^1,\dots,(\underline{\boldsymbol{\epsilon}}_k^d)^{N_e}]$ with $(\underline{\boldsymbol{\epsilon}}_k^d)^j \sim \mathcal{N}(\mathbf{0},\underline{\mathbf{C}}_{d_k})$ for $j = 1,\dots,N_e$. We denote the forecast ensemble as

$$\underline{\boldsymbol{\Psi}}_k^f = \begin{bmatrix} \widetilde{\mathbf{G}}_k^f \\ \mathbf{M}_k^f \\ (\widetilde{\mathbf{G}}_k^f)^* \\ \mathbf{M}_k^f \end{bmatrix}. \tag{B.7}$$

Based on (B.1) and (B.2), the analysis equation for $\underline{\boldsymbol{\Psi}}_k^a$ is given by (omitting the sequential step index)

$$\underline{\boldsymbol{\Psi}}^a = \underline{\boldsymbol{\Psi}}^f + \underline{\mathbf{K}}^e(\underline{\mathbf{D}} - \underline{\mathbf{H}}\,\underline{\boldsymbol{\Psi}}^f), \tag{B.8}$$

where $\underline{\mathbf{K}}^e$ is the approximate Kalman gain

$$\underline{\mathbf{K}}^e = \underline{\mathbf{C}}^e_{\underline{\psi}^f}\underline{\mathbf{H}}^H(\underline{\mathbf{H}}\,\underline{\mathbf{C}}^e_{\underline{\psi}^f}\underline{\mathbf{H}}^H + \underline{\mathbf{C}}_d)^{-1}. \tag{B.9}$$

To get the model update equation, we proceed in the same manner as for EnKF. In the following, let $\Delta\mathbf{X} = \mathbf{X} - \overline{\mathbf{X}}$, where $\mathbf{X}$ is an arbitrary ensemble matrix. Based on (A.8), (A.9), and (3.57) the sample covariance matrix, $\underline{\mathbf{C}}^e_{\underline{\psi}^f}$, omitting the superscript 'f', is given by

$$\begin{aligned}\underline{\mathbf{C}}^e_{\underline{\psi}} &= \frac{1}{N_e-1}\Delta\underline{\boldsymbol{\Psi}}\Delta\underline{\boldsymbol{\Psi}}^H = \frac{1}{N_e-1}\begin{bmatrix} \Delta\widetilde{\mathbf{G}} \\ \Delta\mathbf{M} \\ \Delta\widetilde{\mathbf{G}}^* \\ \Delta\mathbf{M} \end{bmatrix}\begin{bmatrix} \Delta\widetilde{\mathbf{G}}^H & \Delta\mathbf{M}^T & \Delta\widetilde{\mathbf{G}}^T & \Delta\mathbf{M}^T \end{bmatrix}, \\ &= \frac{1}{N_e-1}\begin{bmatrix} \Delta\widetilde{\mathbf{G}}\Delta\widetilde{\mathbf{G}}^H & \Delta\widetilde{\mathbf{G}}\Delta\mathbf{M}^T & \Delta\widetilde{\mathbf{G}}\Delta\widetilde{\mathbf{G}}^T & \Delta\widetilde{\mathbf{G}}\Delta\mathbf{M}^T \\ \Delta\mathbf{M}\Delta\widetilde{\mathbf{G}}^H & \Delta\mathbf{M}\Delta\mathbf{M}^T & \Delta\mathbf{M}\Delta\widetilde{\mathbf{G}}^T & \Delta\mathbf{M}\Delta\mathbf{M}^T \\ \Delta\widetilde{\mathbf{G}}^*\Delta\widetilde{\mathbf{G}}^H & \Delta\widetilde{\mathbf{G}}^*\Delta\mathbf{M}^T & \Delta\widetilde{\mathbf{G}}^*\Delta\widetilde{\mathbf{G}}^T & \Delta\widetilde{\mathbf{G}}^*\Delta\mathbf{M}^T \\ \Delta\mathbf{M}\Delta\widetilde{\mathbf{G}}^H & \Delta\mathbf{M}\Delta\mathbf{M}^T & \Delta\mathbf{M}\Delta\widetilde{\mathbf{G}}^T & \Delta\mathbf{M}\Delta\mathbf{M}^T \end{bmatrix}, \end{aligned} \tag{B.10}$$

Using (B.4), the terms involving $\underline{\mathbf{C}}^e_{\underline{\psi}^f}$ in (B.9) are given by

$$\underline{\mathbf{C}}^e_{\underline{\psi}}\underline{\mathbf{H}}^H = \frac{1}{N_e-1}\begin{bmatrix} \Delta\widetilde{\mathbf{G}}\Delta\widetilde{\mathbf{G}}^H & \Delta\widetilde{\mathbf{G}}\Delta\widetilde{\mathbf{G}}^T \\ \Delta\mathbf{M}\Delta\widetilde{\mathbf{G}}^H & \Delta\mathbf{M}\Delta\widetilde{\mathbf{G}}^T \\ \Delta\widetilde{\mathbf{G}}^*\Delta\widetilde{\mathbf{G}}^H & \Delta\widetilde{\mathbf{G}}^*\Delta\widetilde{\mathbf{G}}^T \\ \Delta\mathbf{M}\Delta\widetilde{\mathbf{G}}^H & \Delta\mathbf{M}\Delta\widetilde{\mathbf{G}}^T \end{bmatrix}, \tag{B.11}$$

and

$$\underline{\mathbf{H}}\,\underline{\mathbf{C}}^e_\psi\,\underline{\mathbf{H}}^H = \frac{1}{N_e - 1}\begin{bmatrix} \Delta\widetilde{\mathbf{G}}\Delta\widetilde{\mathbf{G}}^H & \Delta\widetilde{\mathbf{G}}\Delta\widetilde{\mathbf{G}}^T \\ \Delta\widetilde{\mathbf{G}}^*\Delta\widetilde{\mathbf{G}}^H & \Delta\widetilde{\mathbf{G}}^*\Delta\widetilde{\mathbf{G}}^T \end{bmatrix}. \tag{B.12}$$

Inserting (B.9) into (B.8), and using (B.11) and (B.12) yields

$$\begin{bmatrix} \tilde{\mathbf{G}}^a \\ \mathbf{M}^a \\ (\tilde{\mathbf{G}}^a)^* \\ \mathbf{M}^a \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{G}}^f \\ \mathbf{M}^f \\ (\tilde{\mathbf{G}}^f)^* \\ \mathbf{M}^f \end{bmatrix} + \frac{1}{N_e - 1}\begin{bmatrix} \Delta\widetilde{\mathbf{G}}\Delta\widetilde{\mathbf{G}}^H & \Delta\widetilde{\mathbf{G}}\Delta\widetilde{\mathbf{G}}^T \\ \Delta\mathbf{M}\Delta\widetilde{\mathbf{G}}^H & \Delta\mathbf{M}\Delta\widetilde{\mathbf{G}}^T \\ \Delta\widetilde{\mathbf{G}}^*\Delta\widetilde{\mathbf{G}}^H & \Delta\widetilde{\mathbf{G}}^*\Delta\widetilde{\mathbf{G}}^T \\ \Delta\mathbf{M}\Delta\widetilde{\mathbf{G}}^H & \Delta\mathbf{M}\Delta\widetilde{\mathbf{G}}^T \end{bmatrix}$$
$$\times\left(\frac{1}{N_e - 1}\begin{bmatrix} \Delta\widetilde{\mathbf{G}}\Delta\widetilde{\mathbf{G}}^H & \Delta\widetilde{\mathbf{G}}\Delta\widetilde{\mathbf{G}}^T \\ \Delta\widetilde{\mathbf{G}}^*\Delta\widetilde{\mathbf{G}}^H & \Delta\widetilde{\mathbf{G}}^*\Delta\widetilde{\mathbf{G}}^T \end{bmatrix} + \underline{\mathbf{C}}_d\right)^{-1}(\underline{\mathbf{D}} - \underline{\mathbf{H}}\,\underline{\boldsymbol{\Psi}}^f). \tag{B.13}$$

The update equation for the ensemble of model parameters is thus given by

$$\mathbf{M}^a = \mathbf{M}^f + \frac{1}{N_e - 1}\begin{bmatrix} \Delta\mathbf{M}\Delta\widetilde{\mathbf{G}}^H & \Delta\mathbf{M}\Delta\widetilde{\mathbf{G}}^T \end{bmatrix}$$
$$\times\left(\frac{1}{N_e - 1}\begin{bmatrix} \Delta\widetilde{\mathbf{G}}\Delta\widetilde{\mathbf{G}}^H & \Delta\widetilde{\mathbf{G}}\Delta\widetilde{\mathbf{G}}^T \\ \Delta\widetilde{\mathbf{G}}^*\Delta\widetilde{\mathbf{G}}^H & \Delta\widetilde{\mathbf{G}}^*\Delta\widetilde{\mathbf{G}}^T \end{bmatrix} + \underline{\mathbf{C}}_d\right)^{-1}(\underline{\mathbf{D}} - \underline{\mathbf{H}}\,\underline{\boldsymbol{\Psi}}^f). \tag{B.14}$$

If we define

$$\underline{\mathbf{C}}^e_{mg} = \frac{1}{N_e - 1}\Delta\mathbf{M}\Delta\underline{\mathbf{G}}^H \qquad \underline{\mathbf{C}}^e_g = \frac{1}{N_e - 1}\Delta\underline{\mathbf{G}}\Delta\underline{\mathbf{G}}^H, \tag{B.15}$$

with

$$\underline{\mathbf{G}} = \begin{bmatrix} \widetilde{\mathbf{G}} \\ \widetilde{\mathbf{G}}^* \end{bmatrix}, \tag{B.16}$$

then (B.14) can be given on similar form as (3.65).

Appendix C

Relation between power kernel and Parzen kernel density estimator

The objective of this appendix is to show that the power kernel used in Paper A leads to an interpretation of the shape prior regularization term (see Section 6.3), J_{prior}, as a generalization of the well known Parzen kernel density estimator (in the following denoted KDE). The discussion in this appendix mostly follows [49, Appendix C].

We first define the power kernel. Let $\mathbf{x}$, $\mathbf{y} \in \mathbb{R}^n$, then the power kernel is given by

$$k(\mathbf{x},\mathbf{y}) = \frac{1}{h^n} \begin{cases} \rho(\tau,n) - \|\frac{\mathbf{x}-\mathbf{y}}{h}\|^\tau, & \text{if } \|\frac{\mathbf{x}-\mathbf{y}}{h}\| \leq \rho(\tau,n)^{\frac{1}{\tau}}, \\ 0, & \text{otherwise,} \end{cases} \tag{C.1}$$

where $0 \leq \tau \leq 2$, and

$$\rho(\tau,n) = \left(\frac{\tau+n}{\tau V_n}\right)^{\frac{\tau}{\tau+n}}, \tag{C.2}$$

with V_n being the unit n-dimensional sphere, that is, $V_1 = 2$, $V_2 = \pi$, $V_3 = 4\pi/3$, etc. The power kernel has close connections to the CPD kernel defined as $k(\mathbf{x},\mathbf{y}) = -\|\mathbf{x}-\mathbf{y}\|^\tau$ [140]. It can be shown that any kernel of the form $k = k_{CPD} + b$, where b is a constant, is also CPD [140]. Since $\rho(\tau,n)$ is just a constant and h in (C.1) just scales the output of $k(\mathbf{x},\mathbf{y}) = -\|\mathbf{x}-\mathbf{y}\|^\tau$, the power kernel is CPD.

Let $\mathbf{t}^1,\ldots,\mathbf{t}^m \in \mathbb{R}^n$ be a set of data. A simple estimator of a PDF, $f(\mathbf{x})$, is given by the KDE [130]

$$\hat{f}(\mathbf{x}) = \frac{1}{mh^n}\sum_{i=1}^{m} K\left(\frac{\mathbf{x}-\mathbf{t}^i}{h}\right), \tag{C.3}$$

where K is a Borel measurable function, which is nonnegative and integrates to one (i.e., it fulfills the same properties as a PDF). It is easy to show a connection between KDE and J_{prior}, if we require the kernel to be on the form

$$k(\mathbf{x},\mathbf{y}) = \frac{1}{h^n}K\left(\frac{\mathbf{x}-\mathbf{y}}{h}\right). \tag{C.4}$$

From (6.50) it is seen that J_{prior} consists of two terms where one involves $\tilde{k}(\mathbf{x},\mathbf{x}) = \langle\tilde{\phi}(\mathbf{x}),\tilde{\phi}(\mathbf{x})\rangle = \|\tilde{\phi}(\mathbf{x})\|^2 = \|\phi(\mathbf{x}) - \overline{\boldsymbol{\phi}}\|^2$. Using (6.42) this term can be written

$$\|\phi(\mathbf{x}) - \overline{\boldsymbol{\phi}}\|^2 = \tilde{k}(\mathbf{x},\mathbf{x}) = k(\mathbf{x},\mathbf{x}) - \frac{2}{m}\sum_{i=1}^{m} k(\mathbf{x},\mathbf{t}^i) + \frac{1}{m^2}\sum_{i=1}^{m} k(\mathbf{t}^i,\mathbf{t}^j). \tag{C.5}$$

Inserting (C.4) leads to

$$\|\phi(\mathbf{x}) - \bar{\boldsymbol{\phi}}\|^2 = \text{const.} - \frac{2}{mh^n}\sum_{i=1}^{m} K\left(\frac{\mathbf{x}-\mathbf{t}^i}{h}\right). \tag{C.6}$$

We see that (C.6) is equivalent to (C.3) up to a scaling and a constant. Hence, J_{prior} can be seen as a generalization of KDE.

Now, we need to show that the power kernel (C.1) fulfills (C.4), that is, check if it is nonnegative and integrates to one. First, we identify $K(\mathbf{u})$ for the power kernel as

$$K(\mathbf{u}) = \begin{cases} \rho(\tau,n) - \|\mathbf{u}\|^\tau, & \text{if } \|\mathbf{u}\| \le \rho(\tau,n)^{\frac{1}{\tau}}, \\ 0, & \text{otherwise,} \end{cases} \tag{C.7}$$

where $\mathbf{u} = (\mathbf{x} - \mathbf{t}^i)/h$. From (C.7) it is easily seen that $K(\mathbf{u})$ is always nonnegative. Next, we show that $K(\mathbf{u})$ integrates to one only for $n = 2$, as it gives the basic idea for the proof for a general n. Hence, we are interested to show that

$$\int_{\infty}^{\infty}\int_{\infty}^{\infty} K(u_1,u_2)\, \mathrm{d}u_1 \mathrm{d}u_2 = 1. \tag{C.8}$$

Using polar coordinates, (r,θ), and noticing the $K(\mathbf{u})$ is nonzero only in a radius $\rho^{\frac{1}{\tau}} = \rho(\tau,2)^{\frac{1}{\tau}}$ from the origin, the integral becomes

$$\int_0^{2\pi}\int_0^{\rho^{\frac{1}{\tau}}} (\rho - r^\tau) r\, \mathrm{d}r\mathrm{d}\theta = \rho^{\frac{\tau+2}{\tau}}\left[\frac{\pi\tau}{\tau+2}\right] = 1, \tag{C.9}$$

where the last equality follows from (C.2). Using hyperspherical coordinates it is possible to show that $K(\mathbf{u})$ also integrates to one for a general n.

In summary, we have shown that J_{prior} consists partly of a term similar to the KDE if the kernel fulfills (C.4), and can thus be seen as a generalization of KDE. Moreover, we have shown that the power kernel given in (C.1) fulfills (C.4).

Bibliography

[1] Aanonsen, S. I., Nævdal, G., Oliver, D. S., Reynolds, A. C., and Vallès, B. The ensemble Kalman filter in reservoir engineering – a review. *SPE J. 14*, 3 (2009), 393–412.

[2] Abubakar, A., Habashy, T. M., Li, M., and Liu, J. Inversion algorithms for large-scale geophysical electromagnetic measurements. *Inverse Probl. 25*, 12 (2009), 123012.

[3] Adalsteinsson, D., and Sethian, J. A. A fast level set method for propagating interfaces. *J. Comput. Phys. 118*, 2 (1995), 269–277.

[4] Adimurthi, Dutta, R., Veerappa Gowda, G. D., and Jaffré, J. Monotone (A,B) entropy stable numerical scheme for scalar conservation laws with discontinuous flux. *ESAIM-Math. Model. Num.* (2014).

[5] Adimurthi, Jaffré, J., and Veerappa Gowda, G. D. Godunov-type methods for conservation laws with a flux function discontinuous in space. *SIAM J. Numer. Anal. 42*, 1 (2004), 179–208.

[6] Adimurthi, Mishra, S., and Veerappa Gowda, G. D. Optimal entropy solutions for conservation laws with discontinuous flux-functions. *J. Hyperbol. Differ. Eq. 2*, 4 (2005), 783–837.

[7] Anderson, W. L. A hybrid fast Hankel transform algorithm for electromagnetic modeling. *Geophysics 54*, 2 (1989), 263–266.

[8] Andreianov, B., and Cancès, C. Vanishing capillarity solutions of Buckley–Leverett equation with gravity in two-rocks' medium. *Comput. Geosci. 17*, 3 (2013), 551–572.

[9] Andreianov, B., and Cancès, C. A phase-by-phase upstream scheme that converges to the vanishing capillarity solution for countercurrent two-phase flow in two-rock media. *Comput. Geosci. 18*, 2 (2014), 211–226.

[10] Andreianov, B., Karlsen, K. H., and Risebro, N. H. A theory of L^1-dissipative solvers for scalar conservation laws with discontinuous flux. *Arch Ration Mech. An. 201*, 1 (2011), 27–86.

[11] Aruliah, D. A., Ascher, U. M., Haber, E., and Oldenburg, D. A method for the forward modelling of 3-D electromagnetic quasi-static problems. *Math Mod. Meth. Appl. S. 11*, 01 (2001), 1–21.

[12] Aster, R. C., Borchers, B., and Thurber, C. H. *Parameter Estimation and Inverse Problems*. Academic Press, 2005.

[13] Avdeev, D. B. Three-dimensional electromagnetic modelling and inversion from theory to application. *Surv. Geophys. 26*, 6 (2005), 767–799.

[14] Avdeev, D. B., Kuvshinov, A. V., Pankratov, O. V., and Newman, G. A. High-performance three-dimensional electromagnetic modelling using modified Neumann series. Wide-band numerical solution and examples. *J. Geomagn. Geoelectr. 49*, 11 (1997), 1519–1539.

[15] Axelsson, O. *Iterative Solution Methods*. Cambridge University Press, 1994.

[16] Badea, E. A., Everett, M. E., Newman, G. A., and Biro, O. Finite element analysis of controlled source electromagnetic induction using Coulomb gauged potentials. *Geophysics 66*, 3 (2001), 786–799.

[17] Bakr, S. A., and Mannseth, T. Feasibility of simplified integral equation modeling of low-frequency marine CSEM with a resistive target. *Geophysics 74*, 5 (2009), F107–F117.

[18] Bakr, S. A., Pardo, D., and Mannseth, T. Domain decomposition Fourier finite element method for the simulation of 3D marine CSEM measurements. *J. Comput. Phys. 255* (2013), 456–470.

[19] Bannister, P. R. Determination of the electrical conductivity of the sea bed in shallow waters. *Geophysics 33*, 6 (1968), 995–1003.

[20] Berenger, J.-P. A perfectly matched layer for the absorption of electromagnetic waves. *J. Comput. Phys. 114*, 2 (1994), 185–200.

[21] Berg, C., Christensen, J. P. R., and Ressel, P. *Harmonic Analysis on Semigroups: Theory of Positive Definite and Related Functions*. Springer, Berlin, 1984.

[22] Berlinet, A., and Thomas-Agnan, C. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Kluwer Academic Publishers, 2004.

[23] Bernard, O., Friboulet, D., Thévenaz, P., and Unser, M. Variational B-spline level-set: a linear filtering approach for fast deformable model evolution. *IEEE T. Image Process. 18*, 6 (2009), 1179–1191.

[24] Berre, I., Lien, M., and Mannseth, T. A level-set corrector to an adaptive multiscale permeability prediction. *Comput. Geosci. 11*, 1 (2007), 27–42.

[25] Berre, I., Lien, M., and Mannseth, T. Multi-level parameter structure identification for two-phase porous-media flow problems using flexible representations. *Adv. Water Resour. 32*, 12 (2009), 1777–1788.

[26] Berre, I., Lien, M., and Mannseth, T. Identification of three-dimensional electric conductivity changes from time-lapse electromagnetic observations. *J. Comput. Phys. 230*, 10 (2011), 3915–3928.

[27] Best, M. E., Duncan, P., Jacobs, F. J., and Scheen, W. L. Numerical modeling of the electromagnetic response of three-dimensional conductors in a layered earth. *Geophysics 50*, 4 (1985), 665–676.

[28] Börner, R.-U. Numerical modelling in geo-electromagnetics: advances and challenges. *Surv. Geophys. 31*, 2 (2009), 225–245.

[29] Brenier, Y., and Jaffré, J. Upstream differencing for multiphase flow in reservoir simulation. *SIAM J. Numer. Anal. 28*, 3 (1991), 685–696.

[30] Brooks, R. H., and Corey, A. T. Hydraulic properties of porous media. Tech. Rep. March, Hydrology Papers, Colorado State University, 1964.

[31] Burger, M., and Osher, S. J. A survey on level set methods for inverse problems and optimal design. *Eur. J. Appl. Math. 16*, 2 (2005), 263–301.

[32] Bürger, R., Karlsen, K. H., Mishra, S., and Towers, J. D. On Conservation Laws with Discontinuous Flux. In *Trends Appl. Math. to Mech.*, Y. Wang and K. Hutter, Eds. Shaker Verlag, 2005, pp. 75–84.

[33] Bürger, R., Karlsen, K. H., and Towers, J. D. An Engquist–Osher-type scheme for conservation laws with discontinuous flux adapted to flux connections. *SIAM J. Numer. Anal. 47*, 3 (2009), 1684–1712.

[34] Burgers, G., van Leeuwen, P. J., and Evensen, G. Analysis scheme in the ensemble Kalman filter. *Mon. Weather Rev. 126*, 6 (1998), 1719–1724.

[35] Cecot, W., Rachowicz, W., and Demkowicz, L. An hp-adaptive finite element method for electromagnetics. Part 3: a three-dimensional infinite element for Maxwell's equations. *Int J. Numer. Meth. Eng. 57*, 7 (2003), 899–921.

[36] Chan, T. F., and Tai, X.-C. Level set and total variation regularization for elliptic inverse problems with discontinuous coefficients. *J. Comput. Phys. 193*, 1 (2003), 40–66.

[37] Chave, A. D. Numerical integration of related Hankel transforms by quadrature and continued fraction expansion. *Geophysics 48*, 12 (1983), 1671–1686.

[38] Chave, A. D. On the electromagnetic fields produced by marine frequency domain controlled sources. *Geophys. J. Int. 179*, 3 (2009), 1429–1457.

[39] Chen, J., Hoversten, G. M., Vasco, D., Rubin, Y., and Hou, Z. A Bayesian model for gas saturation estimation using marine seismic AVA and CSEM data. *Geophysics 72*, 2 (2007), WA85–WA95.

[40] Chen, Y., and Oliver, D. S. Ensemble randomized maximum likelihood method as an iterative ensemble smoother. *Math. Geosci. 44*, 1 (2012), 1–26.

[41] Chopp, D. L. Computing minimal surfaces via level set curvature flow. *J. Comput. Phys. 106*, 1 (1993), 77–91.

[42] Commer, M., and Newman, G. A parallel finite-difference approach for 3D transient electromagnetic modeling with galvanic sources. *Geophysics 69*, 5 (2004), 1192–1202.

[43] Constable, S. Marine electromagnetic methods – A new tool for offshore exploration. *Lead. Edge 25*, 4 (2006), 438–444.

[44] Constable, S. Ten years of marine CSEM for hydrocarbon exploration. *Geophysics 75*, 5 (2010), 75A67–75A81.

[45] Constable, S. Review paper: Instrumentation for marine magnetotelluric and controlled source electromagnetic sounding. *Geophys. Prospect. 61* (2013), 505–532.

[46] Constable, S., Key, K., and Lewis, L. Mapping offshore sedimentary structure using electromagnetic methods and terrain effects in marine magnetotelluric data. *Geophys. J. Int. 176*, 2 (2009), 431–442.

[47] Constable, S., and Srnka, L. J. An introduction to marine controlled-source electromagnetic methods for hydrocarbon exploration. *Geophysics 72*, 2 (2007), WA3–WA12.

[48] Cox, C. Electromagnetic induction in the oceans and inferences on the constitution of the earth. *Geophys. Surv. 4*, 1-2 (1980), 137–156.

[49] Cremers, D. *Statistical Shape Knowledge in Variational Image Segmentation.* PhD thesis, Department of Mathematics and Computer Science, University of Mannheim, Germany, 2002.

[50] Cremers, D., Kohlberger, T., and Schnörr, C. Shape statistics in kernel space for variational image segmentation. *Pattern Recogn. 36*, 9 (2003), 1929–1943.

[51] Davydycheva, S., Druskin, V., and Habashy, T. An efficient finite-difference scheme for electromagnetic logging in 3D anisotropic inhomogeneous media. *Geophysics 68*, 5 (2003), 1525–1536.

[52] de Groot-Hedlin, C., and Constable, S. Inversion of magnetotelluric data for 2D structure with sharp resistivity contrasts. *Geophysics 69*, 1 (2004), 78–86.

[53] Dini, D. H., and Mandic, D. P. Class of widely linear complex Kalman filters. *IEEE T. Neural Networ. 23*, 5 (2012), 775–786.

[54] Dmitriev, V. I. Electromagnetic fields in inhomogeneous media. In *Proc. Comput. Cent.* (1969), Moscow State University (in Russian).

[55] Dorn, O., and Lesselier, D. Level set methods for inverse scattering – some recent developments. *Inverse Probl. 25*, 12 (2009), 125001.

[56] Dorn, O., and Villegas, R. History matching of petroleum reservoirs using a level set technique. *Inverse Probl. 24*, 3 (2008), 035015.

[57] Eidesmo, T., Ellingsrud, S., MacGregor, L. M., Constable, S., Sinha, M. C., Johansen, S., Kong, F. N., and Westerdahl, H. Sea Bed Logging (SBL), a new method for remote and direct identification of hydrocarbon filled layers in deepwater areas. *First Break 20*, 3 (2002), 144–152.

[58] Ellingsrud, S., Eidesmo, T., Johansen, S., Sinha, M. C., MacGregor, L. M., and Constable, S. Remote sensing of hydrocarbon layers by seabed logging (SBL): Results from a cruise offshore Angola. *Lead. Edge 21*, 10 (2002), 972–982.

[59] Engelmark, F., Mattsson, J., McKay, A., and Du, Z. Towed streamer EM comes of age. *First Break 32*, 4 (2014), 75–78.

[60] Engl, H. W., Hanke, M., and Neubauer, A. *Regularization of Inverse Problems*. Kluwer Academic Publishers, 2000.

[61] Evensen, G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res. 99*, C5 (1994), 10143–10162.

[62] Evensen, G. Advanced data assimilation for strongly nonlinear dynamics. *Mon. Weather Rev. 125*, 6 (1997), 1342–1354.

[63] Evensen, G. *Data Assimilation: The Ensemble Kalman Filter*. Springer, 2009.

[64] Evensen, G., and van Leeuwen, P. J. An ensemble Kalman smoother for nonlinear dynamics. *Mon. Weather Rev. 128*, 6 (1999), 1852–1867.

[65] Fletcher, R. A modifed Marquardt subroutine for non-linear least squares. Tech. rep., Atomic Energy Research Establishment, Harwell, England, 1971.

[66] Fletcher, R. *Practical Methods of Optimization*. Wiley, 1987.

[67] Flosadóttir, Á. H., and Constable, S. Marine controlled-source electromagnetic sounding 1. Modeling and experimental design. *J. Geophys. Res. 101*, B3 (1996), 5507–5517.

[68] Fossum, K., and Mannseth, T. Parameter sampling capabilities of sequential and simultaneous data assimilation: I. Analytical comparison. *Inverse Probl. 30*, 11 (2014), 114002.

[69] Fossum, K., and Mannseth, T. Parameter sampling capabilities of sequential and simultaneous data assimilation: II. Statistical analysis of numerical results. *Inverse Probl. 30*, 11 (2014), 114003.

[70] George, A., and Liu, J. *Computer Solution of Large Sparse Positive Definite Systems*. Prentice-Hall, 1981.

[71] Gimse, T., and Risebro, N. H. Riemann problems with a discontinuous flux function. In *Proc. Third Int. Conf. Hyperbolic Probl.* (1990).

[72] Gimse, T., and Risebro, N. H. Solution of the Cauchy problem for a conservation law with a discontinuous flux function. *SIAM J. Math. Anal. 23*, 3 (1992), 635–648.

[73] Godunov, S. K. A finite difference method for the numerical computation of discontinuous solutions of the equations of fluid dynamics. *Math. Sb. 47* (1959), 217–306.

[74] Grimstad, A.-A., and Mannseth, T. Nonlinearity, scale, and sensitivity for parameter estimation problems. *SIAM J. Sci. Comput. 21*, 6 (2000), 2096–2113.

[75] Grimstad, A.-A., Mannseth, T., Nævdal, G., and Urkedal, H. Adaptive multiscale permeability estimation. *Comput. Geosci. 7*, 1 (2003), 1–25.

[76] Gupta, P. K., Bennett, L. A., and Raiche, A. P. Hybrid calculations of the three-dimensional electromagnetic response of buried conductors. *Geophysics 52*, 3 (1987), 301–306.

[77] Hadamard, J. Sur les problèmes aux dérivées partielles et leur signification physique. *Princet. Univ. Bull.* (1903), 49–52.

[78] Harten, A., Hyman, J. M., Lax, P. D., and Keyfitz, B. On finite-difference approximations and entropy conditions for shocks. *Commun. Pure Appl. Math. 29*, 3 (1976), 297–322.

[79] Hastings, W. K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika 57*, 1 (1970), 97–109.

[80] Hohmann, G. W. Numerical Modeling for Electromagnetic Methods of Geophysics. In *Electromagn. Methods Appl. Geophys. Voume 1, Theory*. Society of Exploration Geophysicists, 1987.

[81] Hou, J., Mallan, R. K., and Torres-Verdín, C. Finite-difference simulation of borehole EM measurements in 3D anisotropic media using coupled scalar-vector potentials. *Geophysics 71*, 5 (2006), G225–G233.

[82] Hursán, G., and Zhdanov, M. S. Contraction integral equation method in three-dimensional electromagnetic modeling. *Radio Sci. 37*, 6 (2002), 1–13.

[83] Jackson, J. D. *Classical electrodynamics*. Wiley, 1998.

[84] Jacquard, P., and Jain, C. Permeability distribution from field pressure data. *Soc. Pet. Eng. J. 5*, 04 (1965), 281–294.

[85] Jazwinski, A. H. *Stochastic Processes and Filtering Theory*. Academic Press, 1970.

[86] Jin, J. *The finite element method in electromagnetics*. Wiley, 2002.

[87] Jolliffe, I. T. *Principal Component Analysis*. Springer, 2002.

[88] Kaasschieter, E. F. Solving the Buckley-Leverett equation with gravity in a heterogeneous porous medium. *Comput. Geosci. 3*, 1 (1999), 23–48.

[89] Kalman, R. E. A new approach to linear filtering and prediction problems. *J. Basic Eng. 82*, 1 (1960), 35–45.

[90] Key, K. Is the fast Hankel transform faster than quadrature? *Geophysics 77*, 3 (2012), F21–F30.

[91] Key, K. Marine electromagnetic studies of seafloor resources and tectonics. *Surv. Geophys. 33*, 1 (2012), 135–167.

[92] Key, K., and Ovall, J. A parallel goal-oriented adaptive finite element method for 2.5-D electromagnetic modelling. *Geophys. J. Int. 186*, 1 (2011), 137–154.

[93] Key, K., and Weiss, C. Adaptive finite-element modeling using unstructured grids: The 2D magnetotelluric example. *Geophysics 71*, 6 (2006), G291–G299.

[94] Koldan, J., Puzyrev, V., de la Puente, J., Houzeaux, G., and Cela, J. M. Algebraic multigrid preconditioning within parallel finite-element solvers for 3-D electromagnetic modelling problems in geophysics. *Geophys. J. Int. 197*, 3 (2014), 1442–1458.

[95] Kong, F. N., Johnstad, S. E., Røsten, T., and Westerdahl, H. A 2.5D finite-element-modeling difference method for marine CSEM modeling in stratified anisotropic media. *Geophysics 73*, 1 (2008), F9–F19.

[96] Lawson, C. L., and Hanson, R. J. *Solving Least Squares Problems*. Prentice-Hall, Jan. 1974.

[97] Lax, P., and Wendroff, B. Systems of conservation laws. *Commun. Pure Appl. Math. 13* (1960), 217–237.

[98] Leveque, R. J. *Finite Volume Methods for Hyperbolic Problems*. Cambridge University Press, 2002.

[99] Li, Y., Luo, M., and Pei, J. Adaptive finite element modeling of marine controlled-source electromagnetic fields in two-dimensional general anisotropic media. *J. Ocean Univ. China 12*, 1 (2013), 1–5.

[100] Lie, J., Lysaker, M., and Tai, X.-C. A binary level set model and some applications to Mumford-Shah image segmentation. *IEEE T. Image Process. 15*, 5 (2006), 1171–81.

[101] Lien, M. Simultaneous joint inversion of amplitude-versus-offset and controlled-source electromagnetic data by implicit representation of common parameter structure. *Geophysics 78*, 4 (2013), ID15–ID27.

[102] Lien, M., Berre, I., and Mannseth, T. Combined adaptive multiscale and level-set parameter estimation. *Multiscale Model. Sim. 4*, 4 (2005), 1349–1372.

[103] Litman, A. Reconstruction by level sets of *n*-ary scattering obstacles. *Inverse Probl. 21*, 6 (2005), S131–S152.

[104] Løseth, L. O., Pedersen, H. M., Ursin, B., Amundsen, L., and Ellingsrud, S. Low-frequency electromagnetic fields in applied geophysics: Waves or diffusion? *Geophysics 71*, 4 (2006), W29–W40.

[105] Løseth, L. O., and Ursin, B. Electromagnetic fields in planarly layered anisotropic media. *Geophys. J. Int. 170*, 1 (2007), 44–80.

[106] MacGregor, L., and Cooper, R. Unlocking the value of CSEM. *First Break 28*, 5 (2010), 49–52.

[107] Mannseth, T. Relation between level set and truncated pluri-Gaussian methodologies for facies representation. *Math. Geosci. 46*, 6 (2014), 711–731.

[108] Maxwell, J. C. A Dynamical Theory of the Electromagnetic Field. *Philos. Trans. R. Soc. London 155* (1865), 459–512.

[109] McGillivray, P. R., and Oldenburg, D. W. Methods for calculating Frèchet derivatives and sensitivities for the non-linear inverse problem: A comparative study. *Geophys. Prospect. 38*, 5 (1990), 499–524.

[110] Meinhold, R. J., and Singpurwalla, N. D. Understanding the Kalman Filter. *Am. Stat. Assoc. 37*, 2 (1983), 123–127.

[111] Menke, W. *Geophysical Data Analysis: Discrete Inverse Theory*. Elsevier, 2012.

[112] Mercer, J. Functions of positive and negative type, and their connection with the theory of integral equations. *P. Roy. Soc. A-Math. Phy. 209* (1909), 415–446.

[113] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys. 21*, 6 (1953), 1087–1092.

[114] Mishra, S., and Jaffré, J. On the upstream mobility scheme for two-phase flow in porous media. *Comput. Geosci. 14*, 1 (2009), 105–124.

[115] Mitsuhata, Y. 2-D electromagnetic modeling by finite element method with a dipole source and topography. *Geophysics 65*, 2 (2000), 465–475.

[116] Mittet, R., Brauti, K., Maulana, H., and Wicklund, T. A. CMP inversion and post-inversion modelling for marine CSEM data. *First Break 26*, 8 (2008), 59–67.

[117] Moghaddam, B., and Pentland, A. Probabilistic visual learning for object representation. *IEEE T. Pattern Anal. 19*, 7 (1997), 696–710.

[118] More, J. J. The Levenberg-Marquardt Algorithm: Implementation and Theory. In *Lecture Notes Mathematics No. 630 – Numerical Analysis*. Springer, 1978, pp. 105–116.

[119] Mulder, W. A. Geophysical modelling of 3D electromagnetic diffusion with multigrid. *Comput. Vis. Sci. 11*, 3 (2007), 129–138.

[120] Mur, G. Absorbing boundary conditions for the finite-difference approximation of the time-domain electromagnetic-field equations. *IEEE T. Electromagn. C. EMC-23*, 4 (1981), 377–382.

[121] Neidinger, R. D. Introduction to automatic differentiation and MATLAB object-oriented programming. *SIAM Rev. 52*, 3 (2010), 545–563.

[122] Newman, G. A., and Alumbaugh, D. L. Frequency-domain modelling of airborne electromagnetic responses using staggered finite differences. *Geophys. Prospect. 43*, 8 (1995), 1021–1042.

[123] Newman, G. A., and Alumbaugh, D. L. Three-dimensional induction logging problems, Part 2: A finite-difference solution. *Geophysics 67*, 2 (2002), 484–491.

[124] Newman, G. A., and Hoversten, G. M. Solution strategies for two- and three-dimensional electromagnetic inverse problems. *Inverse Probl. 16*, 5 (2000), 1357–1375.

[125] Nocedal, J., and Wright, S. J. *Numerical Optimization*. Springer, 2006.

[126] Oleinik, O. Uniqueness and stability of the generalized solution of the Cauchy problem for a quasilinear equation. *Amer. Math. Soc. Trans. 33* (1964), 285–290.

[127] Oliver, D. S., Reynolds, A. C., and Liu, N. *Inverse Theory for Petroleum Reservoir Characterization and History Matching*. Cambridge University Press, 2008.

[128] Osher, S. J., and Sethian, J. A. Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations. *J. Comput. Phys. 79*, 1 (1988), 12–49.

[129] Pardo, D., Nam, M. J., Torres-Verdín, C., Hoversten, M. G., and Garay, I. Simulation of marine controlled source electromagnetic measurements using a parallel fourier hp-finite element method. *Comput. Geosci. 15*, 1 (2010), 53–67.

[130] Parzen, E. On estimation of a probability density function and mode. *Ann. Math. Stat. 33*, 3 (1962), 1065–1076.

[131] Plessix, R.-É., and Mulder, W. A. Resistivity imaging with controlled-source electromagnetic data: depth and data weighting. *Inverse Probl. 24*, 3 (2008), 034012.

[132] Ray, A., and Key, K. Bayesian inversion of marine CSEM data with a trans-dimensional self parametrizing algorithm. *Geophys. J. Int. 191*, 3 (2012), 1135–1151.

[133] Rusanov, V. Calculation of interaction of non-steady shock waves with obstacles. *J. Comput. Math. Phys. USSR 1* (1961), 267–279.

[134] Sadiku, M. N. O. *Numerical Techniques in Electromagnetics*. CRC press LLC, 2001.

[135] Sammon, P. H. An analysis of upstream differencing. *SPE Reserv. Eng. 3*, 3 (1988), 1053–1056.

[136] Santosa, F. A level-set approach for inverse problems involving obstacles. *ESAIM Control. Optim. Calc. Var. 1* (1996), 17–33.

[137] Sarma, P., and Chen, W. H. Generalization of the ensemble Kalman filter using kernels for non-gaussian random fields. In *Proc. SPE Reserv. Simul. Symp.* (2009), SPE.

[138] Sarma, P., Durlofsky, L. J., and Aziz, K. Kernel principal component analysis for efficient, differentiable parameterization of multipoint geostatistics. *Math.Geosci. 40*, 1 (2008), 3–32.

[139] Schölkopf, B., Smola, A., and Müller, K.-R. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput. 10*, 5 (1998), 1299–1319.

[140] Schölkopf, B., and Smola, A. J. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, 2001.

[141] Schreier, P. J., and Scharf, L. L. *Statistical Signal Processing of Complex-Valued Data : The Theory of Improper and Noncircular Signals*. Cambridge University Press, 2010.

[142] Selley, R. C. *Elements of Petroleum Geology*. Academic Press, 1998.

[143] Shawe-Taylor, J., and Cristianini, N. *Kernel Methods for Pattern Analysis*. Cambridge University Press, 2004.

[144] Shewchuk, J. R. Triangle: Engineering a 2D quality mesh generator and Delaunay triangulator. In *Applied Computational Geometry Towards Geometric Engineering*. Springer, 1996, pp. 203–222.

[145] Simoncini, V., and Szyld, D. B. Recent computational developments in Krylov subspace methods for linear systems. *Numer. Linear Algebr. 14*, 1 (2007), 1–59.

[146] Smith, T., Hoversten, M., Gasperikova, E., and Morrison, F. Sharp boundary inversion of 2D magnetotelluric data. *Geophys. Prospect. 47*, 4 (1999), 469–486.

[147] Song, L., Fukumizu, K., and Gretton, A. Kernel embeddings of conditional distributions: A unified kernel framework for nonparametric inference in graphical models. *IEEE Signal Proc. Mag. 30*, 4 (2013), 98–111.

[148] Tabarovsky, L. A., Goldman, M. M., Rabinovich, M. B., and Strack, K.-M. 2.5-D modeling in electromagnetic methods of geophysics. *J. Appl. Geophys. 35*, 4 (1996), 261–284.

[149] Tadmor, E. Numerical viscosity and the entropy condition for conservative difference schemes. *Math. Comput. 43*, 168 (1984), 369.

[150] Tarantola, A. *Inverse Problem Theory and Model Parameter Estimation*. SIAM Publications, 2005.

[151] Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B 58*, 1 (1996), 267–288.

[152] Tikhonov, A. N., and Arsenin, V. Y. *Solutions of Ill-Posed Problems*. Wiley, 1977.

[153] Um, E. S., and Alumbaugh, D. L. On the physics of the marine controlled-source electromagnetic method. *Geophysics 72*, 2 (2007), WA13–WA26.

[154] van den Bos, A. The multivariate complex normal distribution – A generalization. *IEEE T. Inform. Theory 41*, 2 (1995), 537–539.

[155] van Genuchten, M. T. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Sci. Soc. Am. J. 44*, 5 (1980), 892–898.

[156] van Leeuwen, P. J., and Evensen, G. Data assimilation and inverse methods in terms of a probabilistic formulation. *Mon. Weather Rev. 124*, 12 (1996), 2898–2913.

[157] Vese, L. A., and Chan, T. F. A multiphase level set framework for image segmentation using the Mumford and Shah model. *Int. J. Comput. Vis. 50*, 3 (2002), 271–293.

[158] Wang, M. Y., and Wang, X. "Color" level sets: a multi-phase method for structural topology optimization with multiple materials. *Comput. Methods Appl. Mech. Eng. 193*, 6-8 (2004), 469–496.

[159] Wang, T., and Hohmann, G. W. A finite-difference, time-domain solution for three-dimensional electromagnetic modeling. *Geophysics 58*, 6 (1993), 797–809.

[160] Wannamaker, P. E., Hohmann, G. W., and SanFilipo, W. A. Electromagnetic modeling of three dimensional bodies in layered earths using integral equations. *Geophysics 49*, 1 (1984), 60–74.

[161] Ward, S. H., and Hohmann, G. W. Electromagnetic Theory for Geophysical Applications. In *Electromagn. Methods Appl. Geophys. Voume 1, Theory*. Society of Exploration Geophysicists, 1988, pp. 131–311.

[162] Weaver, J. T., and Brewitt-Taylor, C. R. Improved boundary conditions for the numerical solution of E-polarization problems in geomagnetic induction. *Geophys. J. R. Astron. Soc. 54*, 2 (1978), 309–317.

[163] Weidelt, P. Guided waves in marine CSEM. *Geophys. J. Int. 171*, 1 (2007), 153–176.

[164] Weiss, C. J., and Newman, G. A. Electromagnetic induction in a generalized 3D anisotropic earth, Part 2: The LIN preconditioner. *Geophysics 68*, 3 (2003), 922–930.

[165] Xiong, Z. Electromagnetic fields of electric dipoles embedded in a stratified anisotropic earth. *Geophysics 54*, 12 (1989), 1643–1646.

[166] Yang, H., and Jüttler, B. Evolution of T-spline level sets for meshing non-uniformly sampled and incomplete data. *Vis. Comput. 24*, 6 (2008), 435–448.

[167] Yee, K. S. Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media. *IEEE T. Antenn. Propag. 14*, 3 (1966), 302–307.

[168] Zaslavsky, M., Druskin, V., Davydycheva, S., Knizhnerman, L., Abubakar, A., and Habashy, T. Hybrid finite-difference integral equation solver for 3D frequency domain anisotropic electromagnetic problems. *Geophysics 76*, 2 (2011), F123–F137.

[169] Zhao, H.-K., Chan, T., Merriman, B., and Osher, S. A variational level set approach to multiphase motion. *J. Comput. Phys. 127*, 1 (1996), 179–195.

[170] Zolotukhin, A. B., and Ursin, J.-R. *Introduction to Petroleum Reservoir Engineering*. Norwegien Academic Press (Høyskoleforlaget), 2000.