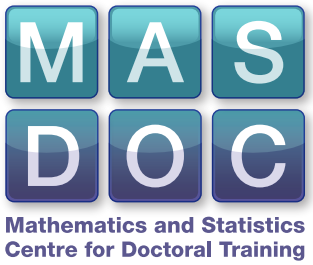


Analysis and Computation for Bayesian Inverse Problems

by

Matthew M. Dunlop

Thesis

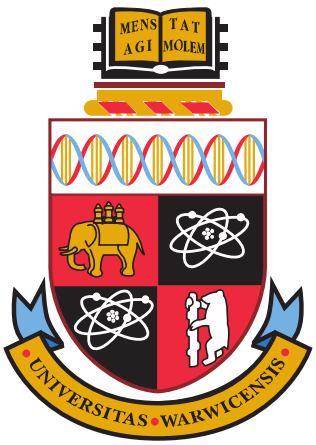
Submitted for the degree of

Doctor of Philosophy

Mathematics Institute

The University of Warwick

June 2016



Contents

Acknowledgments

I would first like to thank my supervisors Andrew Stuart and Marco Iglesias for their help, guidance and support, without which this thesis would not have been possible. They have introduced me to a fascinating, rich field of mathematics that regularly provides new and interesting problems. I'd also like to thank my examiners Carola-Bibiane Schönlieb and Marie-Therese Wolfram for taking the time to read this thesis, and for their many helpful comments. Thank you also to everybody involved with the EQUIP programme, for the support and stimulating discussions.

Next, I thank everyone I've shared an office with over the past four years, providing a welcoming environment for questions and discussion. In particular, I'd like to thank Graham Hobbs, Ben Lees and David O'Connor for the many helpful chats over lunches, tea, coffee and otherwise.

I would also like to thank the Engineering and Physical Sciences Research Council (EPSRC) for funding my position in MASDOC, and the University of Warwick Mathematics department for their hospitality.

Finally I would like to thank my friends and family for their continued support and encouragement throughout my PhD.

Declarations

The work described in this thesis is the author's own, conducted under the supervision of Andrew Stuart (University of Warwick) and Marco Iglesias (University of Nottingham), except where otherwise stated. More specifically:

(i) Chapter 1 is an introduction containing a literature review and outline of the thesis content.

(ii) Chapter 2 is work done in collaboration with Andrew Stuart, and has been submitted for publication [44]. Changes were made on the advice of two anonymous referees.

(iii) Chapter 3 is work done in collaboration with Andrew Stuart, and has been submitted for publication [45].

(iv) Chapter 4 is work done in collaboration with Andrew Stuart and Marco Iglesias, and has been submitted for publication [43]. The numerical simulations presented in section 4.4.2 are due to Marco Iglesias.

This thesis has not been submitted for a degree at any other university. It has not been submitted for award at any other institution for any other qualification.

Abstract

Many inverse problems involve the estimation of a high dimensional quantity, such as a function, from noisy indirect measurements. These problems have received much study from both classical and statistical directions, with each approach having its own advantages and disadvantages. In this thesis we focus on the Bayesian approach, in which all uncertainty is modelled probabilistically.

Recently the Bayesian approach to inversion has been developed in function space. Much of the existing work in the area has been focused on the case when the prior distribution produces samples which are continuous functions, however it is of interest, both in terms of applications and mathematically, to consider cases when these samples are discontinuous. Natural applications are those in which we wish to infer the shape and locations of interfaces between different materials, such as in tomography.

In this thesis we consider Bayesian inverse problems in which the unknown function is piecewise continuous or piecewise constant. Based on prior information, the problem is then to infer the discontinuity set, the values the function takes away from the discontinuities, or both simultaneously. These problems are considered both from analytic and computational points of view. In order to ensure numerical robustness, we formulate any algorithms directly on function space before discretizing. This requires a number of technical issues to be considered, such as the equivalence and singularity of measures on such spaces.

Chapter 1

Introduction

In section 1.1 we describe what is meant by an inverse problem, and provide some examples of challenges associated with the solution of such problems. In subsection 1.1.1 we give an overview of some classical approaches to overcome these challenges, first in the linear finite-dimensional case, and then more general cases. In subsection 1.1.2 we then describe the modern Bayesian approach to inverse problems, where the solution is now a probability measure, rather than a state or set of states. In section 1.2 an outline of the content of the thesis is given, with a brief overview of the results in each chapter.

1.1 Overview

Inverse problems arise in many scientific disciplines, from geophysics and oceanography, to non-destructive testing and machine learning. The objective of such problems is typically to recover a function or other high-dimensional quantity from a number of indirect measurements. This function could for example represent the permeability of the subsurface in geophysical applications, or the internal structure of a body in medical applications. Such problems in these areas have been of significant interest in recent years; their effective resolution is of both fundamental scientific interest, as well as having potential economic value. As an example, consider the recovery of the subsurface structure of a petroleum reservoir from measurements such as flow data in a well, water table height or seismic reflection data. The measurements are highly indirect, yet knowledge of the subsurface is crucial for making significant financial decisions, such as choosing the optimal location for an

oil well to be constructed. In a medical context, data could arise via X-ray measurements or some other tomographic method. Decisions made based on the analysis of this data could include how to or whether to intervene surgically.

A common theme in many of these applications is that the data is not complete – it cannot uniquely determine the unknown quantity. In this case, the problem could be represented in terms of an equation of the form

$$y = \mathcal{G}(u) \tag{1.1.1}$$

where $\mathcal{G} : (X, \|\cdot\|_X) \to (Y, \|\cdot\|_Y)$ is a (often non-linear, non-invertible) mapping between two Banach spaces, termed the *forward map*. In this equation $u \in X$ represents the unknown input to be determined from a number of (frequently noisy and indirect) measurements, and $y \in Y$ represents the observed data. In applications, the forward map is typically formed by composing a *forward model* with an *observation map*. The forward model could for example involve the solution of a partial differential equation related to a physical system, given some input u, and the observation map could take this solution to its values y at a discrete set of observation points.

Another theme is that there may be significant noise on measurements, meaning that even if there is enough data, it may not be consistent with both the physical model and itself. In this case the observations may, for example, take the form

$$y = \mathcal{G}(u) + \eta \tag{1.1.2}$$

where η is a realization of some random noise. Physically this could correspond to measurement instruments having only finite precision, or might be included to account for differences between the computer model and physical reality [78]. Often this noise will be assumed Gaussian, though other distributions are not uncommon depending on the context [59, 103].

Uncertainty in inverse problems hence typically exists both due to missing data and noise on measurements. These problems are usually ill-posed in the sense of Hadamard [60]: there may be no solution, there may be no *unique* solution, or the solution may not depend continuously on the input. In the next subsection we will outline some popular classical approaches that have been used to provide a notion of solution to problems of the form (1.1.1) and (1.1.2). We will then give an overview of the Bayesian approach to inversion, which permits quantification of the uncertainty in the unknown, and its propagation to uncertainty in quantities of interest.

1.1.1 Classical Approaches to Inversion

In order to remove or reduce the effects of ill-posedness from an inverse problem, one could attempt to modify the problem in an appropriate manner, or modify what is meant by a solution to the problem. Well-studied techniques for doing so include variational regularization approaches, truncation of spectral decompositions, definition of notions of quasisolutions, and contruction of dynamical systems with suitable ergodic properties. Such techniques have been considered for a large variety of problems, including inverse spectral problems [46, 110, 140], inverse obstacle scattering [32, 81, 118], tomography [29, 58, 120], image processing and deconvolution [87, 88, 129], and seismic inversion [111, 113, 127]. The structure in these problems, such as the regularity of the forward map, the dimension (or effective dimension) of the input and the data, and the amplitude of the noise, can have a strong effect on what technique is appropriate and how effectively the input can be recovered.

In order to illustrate the technologies available, we will first outline least-squares based methods for linear forward maps on finite dimensional spaces, before discussing methods for more general maps.

Let X and Y be finite dimensional Hilbert spaces, and let the forward map $A : X \to Y$ be linear so that the data y now arises via

$$y = Au.$$

Assuming that A is not invertible, an alternative way of characterizing solutions to this problem is in the least-squares sense. Denote by $\Phi : X \to \mathbb{R}$ the least-squares functional associated with this problem,

$$\Phi(v) := \frac{1}{2}\|Av - y\|_Y^2,$$

and denote by $\mathcal{U} \subseteq X$ the set of minimizers of Φ. Since Φ is smooth and convex, we can characterize $\mathcal{U}$ by for example checking first- and second-order optimality constraints. That is $v \in \mathcal{U}$ if and only if $\Phi'(v) = 0$ and $\Phi''(v)$ is non-negative definite. We calculate

$$\Phi'(v) = A^*Av - A^*y, \quad \Phi''(v) = A^*A$$

and deduce that $v \in \mathcal{U}$ if and only if it satisfies the *normal equations*:

$$A^*Av = A^*y. \tag{1.1.3}$$

However, unless A^*A is invertible, this system will admit an infinity of solutions. Consider the case when $X = \mathbb{R}^n$ and $Y = \mathbb{R}^m$ are Euclidean spaces and $A \in \mathbb{R}^{m\times n}$. Suppose first that $m \geq n$ so that the system is overdetermined, and $\operatorname{rank}(A) = n$. Then all of the singular values of A are strictly positive, hence the same is true of the eigenvalues of A^*A: there exists a unique solution to (1.1.3). If instead we have $m < n$ and $\operatorname{rank}(A) = m$, then $\operatorname{rank}(A^*A) = m < n$ and so there exist infinitely many solutions to (1.1.3). In general, we have infinitely many solutions whenever $\operatorname{rank}(A) < \min\{m, n\}$, so that $\operatorname{rank}(A^*A) = \operatorname{rank}(A) < n$. In all cases there exists at least one solution – this could also be seen by noting that Φ is continuous and coercive.

Even when we are in the case that there exists a unique solution, the behavior of this solution with respect to perturbations in the data y may be poor. If A admits the singular value decomposition $A = U\Sigma V^*$, then it can be checked that the solution is given by

$$u = V\Sigma^{-1}U^*y \tag{1.1.4}$$

where $\Sigma^{-1} \in \mathbb{R}^{n\times m}$ is given by $(\Sigma^{-1})_{ii} = 1/\Sigma_{ii}$ for $i = 1, \ldots, n$ and $(\Sigma^{-1})_{ij} = 0$ otherwise. Hence if any of the singular values Σ_{ii} of A are close to zero, small changes in y can lead to large changes in u.

Typically there will be an infinite number of solutions to (1.1.3), and so it may be preferable to pick out a particular solution u with certain properties. For example, it may be of interest to to choose the solution $u_0 \in \mathcal{U}$ whose norm in X is minimal, that is,

$$u_0 = \operatorname*{argmin}_{v\in\mathcal{U}} \frac{1}{2}\|v\|_X^2. \tag{1.1.5}$$

A relaxation of this approach is to introduce a regularization parameter $\lambda > 0$, and instead seek minimizers of the perturbed least-squares functional $J_\lambda : X \to \mathbb{R}$,

$$J_\lambda(v) := \Phi(v) + \frac{\lambda}{2}\|v\|_X^2.$$

The parameter λ allows us to balance fitting the data with the regularity of the solution. An advantage of the regularized approach is that we can immediately deduce that the functional J_λ has a unique minimizer $u_\lambda \in X$ for each $\lambda > 0$, since

the corresponding normal equations are now given by

$$(A^*A + \lambda I)v = A^*y,$$

and $(A^*A + \lambda I)$ is always positive-definite and hence invertible. Moreover, we have that $u_\lambda \to u_0 \in \mathcal{U}$ as $\lambda \to 0$, where u_0 is as defined by (1.1.5).

Returning to the Euclidean case $X = \mathbb{R}^n$, $Y = \mathbb{R}^m$ discussed earlier, it can be verified that instead of (1.1.4), the solution u_λ to the perturbed problem may be represented as

$$u_\lambda = V\Sigma_\lambda^{-1}U^*y$$

where $\Sigma_\lambda^{-1} \in \mathbb{R}^{n\times m}$ is given by

$$(\Sigma_\lambda^{-1})_{ii} = \frac{\Sigma_{ii}}{\Sigma_{ii}^2 + \lambda}$$

for $i = 1, \ldots, n$ and $(\Sigma_\lambda^{-1})_{ij} = 0$ otherwise. Hence by choosing λ sufficiently large, the effect of small singular values of A on data-sensitivity of the solution can be controlled. This is one sense in which the perturbation of the objective functional can be thought of as a regularization of the problem.

Alternatively, returning to the general case, if one takes a subspace $(E, \|\cdot\|_E)$ embedded in X, it may be of interest to penalize the norm in E rather than the norm in X in order to promote certain properties of the solution [48], an approach commonly referred to as *Tikhonov-Phillips regularization*. Additionally, the problem of making the optimal choice of regularization parameter λ has received significant interest, since it is not always appropriate or efficient to consider the limit $\lambda \to 0$ [47, 54, 125].

Such an approach can also be used when the data is noisy, that is

$$y = Au + \eta.$$

If the noise is known to be Gaussian distributed with zero mean and covariance Γ, i.e. $\eta \sim N(0, \Gamma)$, then it is natural to consider the weighted least-squares functional

$$\Phi(v) := \frac{1}{2}\|Av - y\|_\Gamma^2 := \frac{1}{2}\|\Gamma^{-\frac{1}{2}}(Av - y)\|_Y^2$$

and associated regularized functionals; this will be discussed later in the context of Bayesian inversion. Note that if Γ is diagonal so that each component of the noise is independent, more weight will be placed upon components with smaller amounts

of noise.

Other misfit functions and regularization terms have received much study [40,48,80]. In general, the spaces X, Y will be infinite-dimensional Banach spaces and $\mathcal{G} : X \to Y$ will not necessarily be linear. With data given by (1.1.1) or (1.1.2) it still makes sense to formulate the problem variationally, so that solutions can be defined to be minimizers of the functional $J : X \to \mathbb{R} \cup \{\infty\}$,

$$J(v) = \Phi(v) + \lambda R(v). \tag{1.1.6}$$

Here, $\Phi : X \to \mathbb{R}$ again represents the model-data misfit, and now $R : X \to \mathbb{R} \cup \{\infty\}$ is a general (typically convex) regularization term. These methods are often referred to as *Tikhonov-type* or *generalized Tikhonov* methods. When X and Y are function spaces, possible choices for Φ and R may include, for example,

$$\Phi(v) = \frac{1}{p}\|\mathcal{G}(v) - y\|_{L^p}^p,$$

$$R(v) = \frac{1}{q}\|Bv\|_{L^q}^q \text{ or } R(v) = \frac{1}{2}\|v\|_E^2 \text{ or } R(v) = \frac{1}{2}\|v\|_{\mathrm{TV}},$$

for some $p, q \geq 1$, some non-negative operator B, and some $(E, \|\cdot\|_E)$ compactly embedded in X. B may for example be a differential operator, in which case the regularity of minimizers can be controlled, or a multiple of the identity, in which case their amplitudes can be controlled. As in the finite-dimensional linear case, the choice of the squared E-norm as a regularization term is known as Tikhonov-Phillips regularization [48]. The choice of the total variation (TV) regularization term is often used when the unknown function is piecewise continuous; the TV norm penalizes the surface area of the discontinuity set, and so over-fitting of data can be avoided [114].

Note that due to the non-linearity in $\mathcal{G}$, the minimizers of J in the case $p = q = 2$ do not satisfy a linear system as before. This presents issues both from analytical and computational perspectives. Another issue is that in general Φ will not be convex, and so more work is needed to obtain existence of minimizers of J. In these cases there may exist multiple local minima which are not the global minimum. These local minima may however be of interest in order to get a fuller picture of the solution of the problem, instead of just looking at the global minimum.

Another class of approaches for solving (1.1.2) are known as *residual methods* [57]. In this case, with the same notation as above, one considers the problem of minimizing

the regularization term R subject to the condition that the misfit Φ is not too large. Specifically, one can look for solutions in the set

$$\operatorname*{argmin}_{v \in X} \big\{ R(v) \;\big|\; \Phi(v) \leq \delta \big\}$$

where δ is approximately the size of the data error, that is $\delta \approx \Phi(u^\dagger)$, and $u^\dagger$ is the true state that generates the data. Note that the calculation of δ does not in general require knowledge of $u^\dagger$, only of the model: if for example Φ is given by the least-squares functional, then we have that

$$\Phi(u^\dagger) = \frac{1}{2}\|\eta\|_Y^2$$

and so we may use the distributional properties of η to estimate δ.

Finally one could consider a combination of the Tikhonov-type and residual methods. Denote by u_λ the solution to (1.1.6) using regularization parameter λ. We can look for the choice of λ such that $\Phi(u_\lambda) = \Phi(u^\dagger)$. This is a root finding problem, and is known as Morozov's discrepancy principle [33].

1.1.2 The Bayesian Approach to Inversion

Probability can be used to account for incomplete data and noise on the measurements. The Bayesian approach treats all unknowns as random variables, and so inference about these random variables then allows us to perform inference on the system itself. Bayes' formula tells us quantitatively how to marry data with prior beliefs to produce the *posterior distribution*. This is a probability measure which, in the context of inverse problems, will typically be defined on an infinite-dimensional Banach space. Quantities of interest, for example posterior mean, variance and modes, can then be studied under this distribution. Since the solution to the problem is now a probability measure rather than a single state, we are able to quantify any uncertainty arising in these quantities via integration. Moreover, whilst the classical approaches discussed in the previous subsection were somewhat ad hoc, the Bayesian approach provides a constructive approach to regularization, especially through the connection with MAP estimators elaborated below.

The following are the main components in the Bayesian approach:

1. *Prior distribution.* A probability measure on X that describes our beliefs about the solution to the inverse problem before any data has been collected.

We typically denote this by μ_0.

2. *Likelihood.* A function $\mathcal{L}$ representing how likely data is to have arisen from a given state u, i.e. the conditional density of y given u. We will typically write this in the form
$$\mathcal{L}(y;u) = \exp(-\Phi(u;y)).$$

3. *Posterior distribution.* A probability measure on X that arises by combining the prior distribution and likelihood using Bayes' theorem, representing the conditional distribution of u given the data y. Denoting this measure by μ^y, it is given by
$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(u) = \frac{1}{Z(y)}\exp(-\Phi(u;y)) \tag{1.1.7}$$
where $Z(y)$ is the normalization constant.

The posterior distribution μ^y is the solution to the Bayesian inverse problem. Once we have this, many questions may arise. For example:

(i) How sensitive is the posterior distribution to perturbations of the data? How about the choice of prior distribution?

(ii) If the forward model isn't perfect, either mathematically or computationally, what effect will this have?

(iii) How can we sample the posterior numerically, and ensure that sampling efficiency does not decay as any approximation meshes are refined?

Such questions are addressed in generality in the lecture notes [39] and the references therein. The sensitivity in (i) and (ii) is typically characterized with respect to the Hellinger metric, which leads to error bounds on expectations of different quantities. Question (iii) is addressed by formulating any sampling algorithms on function space *before* any discretization, and showing such algorithms are well-defined.

The Bayesian and classical approaches to inversion may be related via the modes of the posterior distribution, termed MAP estimators. If the posterior distribution μ^y admits a Lebesgue density, so that it can be written
$$\frac{\mathrm{d}\mu^y}{\mathrm{d}u}(u) \propto \exp(-J(u)),$$
then a MAP estimator for μ^y is defined as a maximizer of this density, or equivalenty a minimizer of the functional J. Determination of MAP estimators is hence a

variational problem, and the choice of prior distribution is akin to the choice of regularization term in the classical approaches discussed earlier. In general there will not exist a Lebesgue density, however the definition of MAP estimators may be extended to this case. In [38] the authors provide a definition involving ratios of measures of balls of diminishing size, in such a way that it agrees with the standard definition when Lebesgue densities exist. Later on we will see a direct relation between MAP estimation and Tikhonov-Phillips regularized minimization, in the case of Gaussian observations and Gaussian priors, proved in [38].

An overview of the area of statistical, and in particular, Bayesian approaches to inverse problems is provided in the text [75], with a strong focus on cases where Lebesgue densities exist. This is often the case if, for example, the forward map is assumed to be defined on a finite element space rather than an infinite dimensional function space. Many computational issues are also discussed therein, related to approximation of integrals of quantities of interest against the posterior distribution. Markov Chain Monte Carlo (MCMC) methods are commonly used in practice, which produce Markov chains whose states, in stationarity, are drawn from the posterior. These states can then be used to approximate integrals via Monte Carlo approximation. MCMC techniques typically require many evaluations of the likelihood to produce accurate approximations of quantities of interest, due to strong correlations between states. One is often therefore interested in how the posterior distribution could be approximated by sampling simpler distributions, such as a Gaussians, since these may be sampled much more cheaply. Possible approximations that have been considered previously include linearization around the MAP [113], randomized maximum likelihood [112] and ensemble Kalman filtering [1]. A comparison of these approximations in the context of reservoir simulation is provided in [70].

As in the classical case, the situation when the forward map is linear and defined on finite dimensional spaces can be easier to analyze. Let X, Y be finite dimensional Hilbert spaces and let $A : X \to Y$ denote the forward map. Take a Gaussian prior $\mu_0 = N(m_0, \mathcal{C}_0)$ on u, and assume the noise has a Gaussian distribution $\eta \sim N(0, \Gamma)$. Then the posterior is again Gaussian, with mean m and covariance $\mathcal{C}$ given by

$$\begin{aligned} \mathcal{C} &= (A^*\Gamma^{-1}A + \mathcal{C}_0)^{-1}, \\ m &= \mathcal{C}(A^*\Gamma^{-1}y + \mathcal{C}_0^{-1}m_0). \end{aligned}$$

The matrix $A^*\Gamma^{-1}A$ is the Fisher information associated with the likelihood. Since this is positive semi-definite, the posterior covariance is 'smaller' in some sense

than the prior covariance. If the Fisher information is zero, then any data will be uninformative and the posterior equals the prior. On the other hand, as the size of the Fisher information increases, the posterior variance contracts to zero and the posterior mean converges to the true state that generates the data – this is an example of what is termed *posterior consistency* [30].

Note that the posterior mean is the unique minimizer of the functional

$$J(v) = \frac{1}{2}\|Av - y\|_\Gamma^2 + \frac{1}{2}\|\mathcal{C}_0^{-\frac{1}{2}}(v - m_0)\|_X^2 \tag{1.1.8}$$

and thus there exists a clear link between classical and Bayesian approaches in the linear Gaussian case. Note also that the posterior Lebesgue density is given by

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}u}(u) \propto \exp(-J(u)).$$

This is maximized whenever J is minimized, and so the posterior mean is the unique MAP estimator in this case.

The non-linear non-parametric case can be significantly more difficult to analyze. The lecture notes [39] provide an overview of many areas associated with such problems. Posterior consistency results have been obtained in, for example, [4, 5, 38]. It is this setting that is considered throughout the majority of this thesis. One of the main obstacles with the non-parametric approach is the the lack of existence of a Lebesgue measure on the space X, and so absolute continuity between measures becomes a much more central property. In particular, the absolute continuity (1.1.7) of the posterior with respect to the prior means that any almost sure properties of the prior, such as sample regularity, are also almost sure properties of the posterior.

In [38] the authors provide the following result regarding MAP estimators in the non-linear non-parametric case. Suppose that the prior μ_0 is chosen to be Gaussian[1] with associated Cameron-Martin space $(E, \|\cdot\|_E)$ compactly embedded in $(X, \|\cdot\|_X)$. Then under appropriate conditions on the forward map $\mathcal{G}$, MAP estimators are shown to be equivalent to minimizers of the functional $J : X \to \mathbb{R} \cup \{\infty\}$ given by

$$J(v) = \frac{1}{2}\|\mathcal{G}(v) - y\|_\Gamma^2 + \frac{1}{2}\|v - m_0\|_E^2$$

for $v - m_0 \in E$, and infinity otherwise. A link between Bayesian and classical inversion techniques is hence again evident. Note also that in the case that X is a

[1]The definition of a Gaussian measure on a Banach space may be found in the appendix.

separable Hilbert space and the prior covariance $\mathcal{C}_0$ is strictly positive, the Cameron-Martin space takes the form $(E, \|\cdot\|_E) = (\mathcal{C}_0^{1/2} X, \|\mathcal{C}_0^{-1/2} \cdot\|_X)$, and so J takes the same form as (1.1.8) whenever it is finite. In general, however, minimizers of J will not coincide with the posterior mean due to the non-linearity of $\mathcal{G}$.

1.2 Outline of the Thesis

The recent development of Bayesian inversion on function space described in the previous section is focused predominantly on cases where fields sampled from the prior distribution (and hence posterior distribution) are almost surely continuous. There are however many applications where the unknown field is expected, or known a priori, to be discontinuous. This includes for example applications in subsurface and medical imaging, where interfaces between different materials correspond to jump discontinuities [20, 85, 113], and deconvolution of piecewise constant signals [28, 64, 121]. In this thesis we develop the mathematically challenging case of reconstruction of piecewise continuous fields from noisy non-direct measurements. The recovery of these fields could be thought of as the joint recovery of the values the fields take in the regions where they are continuous, and the shape of the interfaces between these regions. We can consider different ways to parametrize both of these.

In what follows the forward map will often take the following form. Let X, Y, V denote three Banach spaces, and let Z denote a function space. Define the following three maps between these spaces:

- $F : X \to Z$ is a construction map, which maps a (typically infinite) set of parameters to a function. This allows for construction of functions with a certain structure, such as being piecewise continuous and/or positive.
- $K : Z \to V$ is a forward model. This is typically a non-linear map involving the solution of a partial differential equation.
- $O : V \to Y$ is an observation map, which for example maps a function to its values at a discrete set of points.

The forward map $\mathcal{G} : X \to Y$ is then defined as the composition $\mathcal{G} = O \circ K \circ F$. In such a setup, it is usually the function $F(u)$ that is of interest, rather than the parameters $u \in X$ themselves, however it is often advantageous to perform inference directly on u.

1.2.1 Chapter 2 – MAP Estimators for Piecewise Constant Inversion

In this chapter we consider the case where the interfaces can be parametrized finite dimensionally, but the fields infinite dimensionally. As such, the parameter space X takes the form $X = \Lambda \times W$ where $\Lambda \subseteq \mathbb{R}^k$ denotes the set of geometric parameters defining the interfaces, and the space $W = C(D; \mathbb{R}^N)$ contains the fields between these interfaces. Such a setup was considered using a Bayesian approach in [71]. A natural application would be in a reservoir model where we are trying to infer the subsurface structure: it may be known that there is a fault occurring in the geology, but the location and size of this fault, along with the permeabilities of the different media, may be unknown. We place a compactly supported prior with continuous Lebesgue density on the geometric parameters, and a Gaussian prior on the fields, so that the prior permeability is piecewise Gaussian with random interfaces.

We look at MAP estimation for this problem, i.e. determination of posterior modes, as has been considered in a function space context in [38]. Since X is infinite dimensional, the posterior distribution does not admit a Lebesgue density that can be maximized; we instead use the definition of mode provided in for example [38], involving maximizing ratios of measures of balls of diminishing size. Similarly to [38] the MAP estimators can be characterized as minimizers of a particular functional on the product space of fields and geometric parameters. Numerically solving this minimization problem we observe that, when a non-trivial geometric model is used, the posterior distribution arising can be highly multi-modal. Such observations are confirmed by the behavior of MCMC simulations.

The work in this chapter is in collaboration with my PhD supervisor Andrew Stuart (University of Warwick), and is contained in [44]. This work complements that of [71] and [38] by extending the existence and well-posedness theory from [71] to more general forward maps than groundwater flow, and by analyzing the problem of MAP estimation for more general priors than the Gaussians considered in [38].

1.2.2 Chapter 3 – The Bayesian Formulation of EIT

In this chapter we focus on the inverse problem associated with electrical impedance tomography (EIT). The EIT problem is closely related to a classical mathematical inverse problem, called Calderón's problem, concerning the recovery of the coefficient of a divergence form elliptic PDE given either full or partial knowledge of its

Dirichlet-to-Neumann (DtN) or Neumann-to-Dirichlet (NtD) map. More precisely, if $D \subseteq \mathbb{R}^d$ and $g \in H^{1/2}(\partial D)$ is given, let $u \in H^1(D)$ solve

$$\begin{cases} -\nabla \cdot (\sigma \nabla u) = 0 & \text{in } D \\ u = g & \text{on } \partial D. \end{cases}$$

Then does the DtN mapping $\Lambda_\sigma : H^{1/2}(\partial D) \to H^{-1/2}(\partial D)$ given by

$$g \mapsto \sigma \nabla u \cdot \nu,$$

where ν is the outward unit normal to the boundary ∂D, determine the coefficient σ in D? An associated, more ill-posed inverse problem is to ask whether noisy partial knowledge of the NtD map $(\Lambda_\sigma)^{-1} : H^{-1/2}(D) \to H^{1/2}(D)$ allows for determination of σ. Much work has been done in this area, for example [6,77,106,132]. An overview of results in the DtN case is given in [20].

A physically realistic PDE model for the problem, of interest in medical imaging, was introduced in [128]. In this chapter we provide a rigorous Bayesian formulation of this EIT problem in an infinite dimensional setting, leading to well-posedness in the Hellinger metric with respect to the data. We focus particularly on the reconstruction of piecewise constant fields where the interface between different media is the primary unknown. We consider three different prior models, corresponding to three different choices for the space X and the map F defined at the start of the subsection. Numerical simulations based on the implementation of MCMC are performed, illustrating the advantages and disadvantages of each type of prior in the reconstruction, in the case where the true conductivity is a binary field, and exhibiting the properties of the resulting posterior distribution.

The work in this chapter is in collaboration with my PhD supervisor Andrew Stuart (University of Warwick), and is contained in [45]. This work gives, as far as we are aware, the first rigorous Bayesian formulation of the EIT problem on function space, for three flexible families of priors. Such a formulation allows for the implementation of function space based algorithms, which provide attractive mesh-independent properties.

1.2.3 Chapter 4 – Hierarchical Bayesian Level Set Inversion

In this chapter we consider the case where the construction map F thresholds a continuous function at several levels in order to define a piecewise constant func-

tion. The inverse problem thus concerns the recovery of interfaces between different phases. This approach is known as the level set approach, and was introduced in the 1980's for the study of interface problems [115]. In the context of inverse problems, the level set approach gained traction after the seminal paper by Santosa [123]. One of the key advantages of the level set approach is the admission of topological changes from the initial 'guess' for the solution. Such methods typically evolve the level set function u via a Hamilton-Jacobi equation,

$$\frac{\partial u}{\partial t} = v|\nabla u|,$$

for some suitable choice of normal velocity field v related to the data misfit functional [41, 98, 123].

The Bayesian level set approach to such inverse problems was introduced in [72] and leads to well-posed posterior distributions under appropriate assumptions on the prior and forward map. However the resulting posterior distribution can admit strong sensitivity to the length and amplitude scales encoded in the prior probabilistic model. In this chapter we show that the scale-sensitivity can be circumvented by means of a hierarchical approach, using a single scalar parameter. The hierarchical approach we consider relies on the equivalence of a certain family of Gaussian measures indexed by this parameter, related to Whittle-Matérn distributions. This equivalence of measures allows for the formulation of a Metropolis-within-Gibbs algorithm on function space, namely, an algorithm that alternates between Metropolis-Hastings updates for the field and the scalar parameter.

The equivalence of measures is fundamental for performance of sampling algorithms on high-dimensional discretizations of the parameter space. As was done for the non-hierarchical approach, well-posedness of the problem is shown. We also show effectiveness of a resulting algorithm via numerical simulations, in the context of groundwater flow, EIT and linear observations – even when no 'true' lengthscale is specified in the data, a consistent value for the lengthscale is identified in simulation, regardless of the initial guess of this lengthscale.

The work in this chapter is in collaboration with my PhD supervisors Marco Iglesias (University of Nottingham) and Andrew Stuart (University of Warwick), and is contained in [43]. This work extends that of [72] to allow for hierarchical priors, leading to more accurate reconstruction when less is known a priori; it also points towards the analysis of more general hierarchical models, such as those involving anisotropic lengthscales, or deeper hierarchies.

Chapter 2

MAP Estimators for Piecewise Continuous Inversion

2.1 Introduction

2.1.1 Context and Literature Review

A common inverse problem is that of estimating an unknown function from noisy measurements of a (possibly nonlinear) map applied to the function. Statistical and deterministic approaches to this problem have been considered extensively. In this chapter we focus on the the study of MAP estimators within the Bayesian approach; these estimators provide a natural link between deterministic and statistical methods. In the Bayesian formulation, we describe the solution probabilistically and the distribution of the unknown, given the measurements and a prior model, is termed the posterior distribution. MAP estimators attempt to work with a notion of solutions of maximal probability under this posterior distribution and are typically characterized variationally, linking to deterministic methods.

There are two main approaches taken to the study of the posterior. The first is to discretize the space, and then apply finite dimensional Bayesian methodology [75]. An advantage to this approach is the availability of a Lebesgue density and a large amount of previous work which can then be built upon; but issues may arise (for example computationally) when the dimension of the discretisation space is increased. An alternative approach is to apply infinite dimensional methodology directly on the original space, to derive algorithms, and then discretize to implement.

This approach has been studied for linear problems in [50,96,101], and more recently for nonlinear problems [39, 91, 92, 131]. It is the latter approach that we focus on in this chapter.

In some situations it may be that point estimates are more desirable, or more computationally feasible, than the entire posterior distribution. A detailed study of point estimates can be found in for example [95]. Three different estimates are commonly considered: the posterior mean which minimizes L^2 loss, the posterior median which minimizes L^1 loss, and posterior modes which minimize zero-one loss. The former two estimates are unique [104], but a distribution may possess more than one mode. A consequence of this is that the posterior mean and median may be misleading in the case of a multi-modal posterior. Posterior modes are often termed maximum a posteriori (MAP) estimators in the literature.

In this chapter we focus on MAP estimation. If the posterior has Lebesgue density ρ, MAP estimators are given by the global maxima of ρ. The problem of MAP estimation in this case is hence a deterministic variational problem, and has been well-studied [75]. In the infinite-dimensional setting there is no Lebesgue density, but there has been recent research aimed at characterizing the mode variationally and linking to the classical regularisation techniques described in, for example, [38] in the case when Gaussian priors are adopted. Non-Gaussian priors have also been considered in the infinite dimensional setting – in [65] weak MAP (wMAP) estimators are defined as generalisations of MAP estimators, and a variational characterisation of them is provided in the case that the forward map is linear, using the notion of Fomin derivative.

In this chapter we make a significant extension of the work in [38] to include priors which are defined by a combination of Gaussian random fields and a finite number of geometric parameters which define the different domains in which the different random fields apply. We thereby study the reconstruction of piecewise continuous fields with interfaces defined by a finite number of parameters. Our motivation for doing so comes from the work in [27], and its predecessors. In that paper a Bayesian inverse problem for piecewise constant fields, modelling the permeability appearing in a two-phase subsurface flow model, was studied. The idea of single point estimates being misleading is discussed and the existence of multiple local MAP estimators is shown. We also link our work to that in [65], by characterizing the MAP estimator via the Fomin derivative.

Such piecewise continuous fields were previously studied in the context of ground-

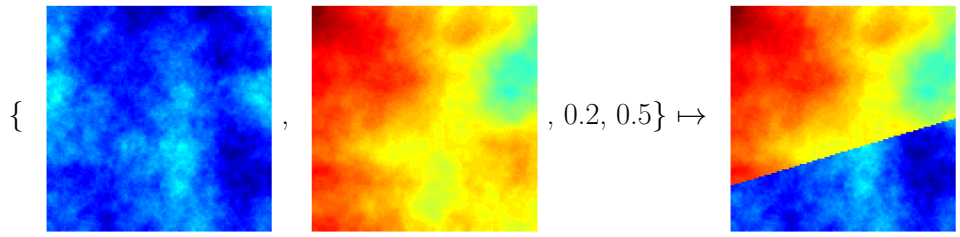


Figure 2.1: An example of construction of a piecewise continuous field, using two continuous fields and two scalar parameters. Here the scalar parameters determine the points where the interface meets each side of the domain. We work on the space of continuous fields and parameters, but it is pushforward of these by the construction map that represents the piecewise continuous field we aim to recover.

water flow [71], where existence and well-posedness of the posterior distribution were shown. Throughout this chapter we focus on two model problems: groundwater flow and electrical impedance tomography (EIT). Both of these problems are important examples of large scale inverse problems, sith applications of great economic and societal value. MAP estimation in such problems has been studied previously [14, 21, 73, 113]. However our formulation is quite general; for brevity we simply illustrate the theory for groundwater flow and EIT, and the numerics only in the case of groundwater flow.

2.1.2 Mathematical Setting

Let X be a separable Banach space and let $\Lambda \subseteq \mathbb{R}^k$. X should be thought of as a function space and Λ a space of geometric parameters. Given $(u, a) \in X \times \Lambda$, we construct another function $u^a \in Z$, say. Considering the ingredients u and a in the construction of this function u^a separately will be useful in what follows. An example of such a construction is shown in Figure 2.1.

Suppose we have a (typically nonlinear) forward operator $\mathcal{G} : X \times \Lambda \to Y$, where $Y = \mathbb{R}^J$. If (u, a) denotes the true input to our forward problem, we observe data $y \in Y$ given by

$$y = \mathcal{G}(u, a) + \eta$$

where $\eta \sim N(0, \Gamma)$, $\Gamma \in \mathbb{R}^{J \times J}$ positive definite, is some centred Gaussian noise on Y. Modelling everything probabilistically, we build up the joint distribution of (u, a, y) by specifying a prior distribution $\mu_0 \times \nu_0$ on (u, a) and an independent noise

model on η. We are then interested in the posterior μ on (u, a) given y. Denote $|\cdot|$ the Euclidean norm on $\mathbb{R}^J$, and for any positive definite $A \in \mathbb{R}^{J\times J}$ denote $|\cdot|_A := |A^{-1/2}\cdot|$ the weighted norm on $\mathbb{R}^J$. Under certain conditions, using a form of Bayes' theorem, we may write μ in the form

$$\mu(\mathrm{d}u, \mathrm{d}a) \propto \exp\left(-\frac{1}{2}|\mathcal{G}(u, a) - y|_\Gamma^2\right)\mu_0(\mathrm{d}u)\nu_0(\mathrm{d}a).$$

The modes of the posterior distribution, termed MAP (maximum a posteriori) estimators, can be considered 'best guesses' for the state (u, a) given the data y. We now state rigorously what we mean by a MAP estimator for μ, as in [38]. Given $(u, a) \in X \times \Lambda$, denote by $B^\delta(u, a)$ the ball of radius δ centred at (u, a).

Definition 2.1.1 (MAP estimator)**.** *For each $\delta > 0$, define*

$$(u^\delta, a^\delta) = \operatorname*{argmax}_{(u,a)\in X\times\Lambda} \mu(B^\delta(u, a)).$$

Any point $(\bar{u}, \bar{a}) \in X \times \Lambda$ satisfying

$$\lim_{\delta\downarrow 0} \frac{\mu(B^\delta(\bar{u}, \bar{a}))}{\mu(B^\delta(u^\delta, a^\delta))} = 1$$

is called a MAP estimator for the measure μ.

If this definition is applied to probability measures defined via a Lebesgue density, MAP estimators coincide with maxima of this density. Here we extend the notion to the study of piecewise continuous fields. Note that it is not clear from this definition when MAP estimators will exist or when they are unique; however the connection with variational problems described in section 2.5.1 makes this more apparent.

2.1.3 Our Contribution

The primary contributions of the chapter are fourfold:

(i) We develop the MAP estimator theory for infinite dimensional geometric inverse problems involving discontinuous fields, building on theory in both of the recent papers [38, 65], and opening up new avenues for the study of MAP estimators in infinite dimensional inverse problems.

(ii) We explicitly link MAP estimation for these geometric inverse problems to a variational Onsager-Machlup minimization problem.

(iii) We show that the theory applies to the groundwater flow model as in [71] and we show that the theory applies to the EIT problem as in [45].

(iv) We implement numerical experiments for the groundwater flow model and demonstrate the feasibility of computing (local) MAP estimators within the geometric formulation, but also show that they can lead to multiple nearby solutions. We relate these multiple MAP estimators to the behaviour of output from MCMC to probe the posterior.

2.1.4 Structure of the Chapter

- In section 2.2 we describe the forward maps associated with the groundwater flow and EIT problems, and show that they have the appropriate regularity needed in sections 2.4–2.5.
- In section 2.3 we describe the choice of, and assumptions upon, the prior distribution whose samples comprise piecewise Gaussian random fields with random interfaces.
- In section 2.4 we show existence and uniqueness of the posterior distribution.
- In section 2.5 we define MAP estimators and prove their equivalence to minimizers of an appropriate Onsager-Machlup functional.
- In section 2.6 we present numerics for the groundwater flow problem. We consider three different prior models and investigate maximizers of the posterior distribution.
- In section 2.7 we conclude and outline possible future work in the area.

2.2 The Forward Problem

We consider two model problems. Our first problem (groundwater flow) is that of determining the piecewise continuous permeability of a medium, given noisy measurements of water pressure (or hydraulic head) within it. The second problem (EIT) is determination of the piecewise continuous conductivity within a body from boundary voltage measurements.

In what follows, the finite dimensional space Λ will be a space of geometric parameters defining the interfaces between different media, and X will be a product of

function spaces defining the values of the permeabilities/conductivities between the interfaces.

We begin in subsection 2.2.1 by defining the construction map $(u, a) \mapsto u^a$ for the piecewise continuous fields. In subsections 2.2.2 and 2.2.3 we describe the models for groundwater flow and EIT respectively, and prove regularity properties of the resulting forward maps; these properties are required for our subsequent theory.

2.2.1 Defining the Interfaces

Let $D \subseteq \mathbb{R}^d$ be the domain of interest and let $\Lambda \subseteq \mathbb{R}^k$ be the space of geometric parameters. Let $\mathcal{B}(D)$ denote the Borel σ-algebra on D, defined in the Appendix. Take a collection of set-valued maps $A_i : \Lambda \to \mathcal{B}(D)$, $i = 1, \dots, N$ such that for each $a \in \Lambda$ we have

$$\bigcup_{i=1}^{N} A_i(a) = D, \quad A_i(a) \cap A_j(a) = \varnothing \text{ if } i \neq j.$$

We assume that each map A_i is continuous in the sense that

$$|a - b| \to 0 \Rightarrow |A_i(a) \Delta A_i(b)| \to 0$$

where Δ denotes the symmetric difference:

$$A \Delta B := (A \setminus B) \cup (B \setminus A).$$

Let $X = C^0(D; \mathbb{R}^N)$. Given $u = (u_1, \dots, u_N) \in X$ and $a \in \Lambda$ we define the function $u^a \in L^\infty(D)$ by

$$u^a = F(u, a) := \sum_{i=1}^{N} u_i \mathbb{1}_{A_i(a)}. \tag{2.2.1}$$

where $F : X \times \Lambda \to L^\infty(D)$ is the construction map.

We give four examples of the functions A_i and the sets/interfaces they define.

Example 2.2.1. *Let $D = [0, 1]^2$, $\Lambda = [0, 1]^2$ and $N = 2$. We specify points a and b on either side of the square D and join them with a straight line. We then let $A_1(a, b)$ be the region of D below this line and $A_2(a, b) = D \setminus A_1(a, b)$. Example sets $A_i(a, b)$ for various parameters a, b are shown in Figure 2.2.*

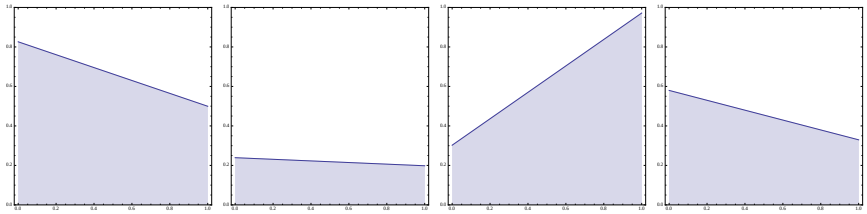


Figure 2.2: Possible sets A_i corresponding to Example 2.2.1

Example 2.2.2. *Let $D = [0,1]^2$, $\Lambda = [0,1]^2$ and $N = 2$. Choose a continuous map $H : \Lambda \to L^\infty([0,1])$ such that $H(a,b)(0) = a$ and $H(a,b)(1) = b$ for all $(a,b) \in \Lambda$. Let $A_1(a,b)$ be the region of D beneath the graph of the curve $H(a,b)$ and let $A_2(a,b) = D \setminus A_1(a,b)$. This setup includes the previous example: $H(a,b)(x) = a + (b-a)x$ defines the appropriate straight lines.*

The continuity of A_1 and A_2 can be seen by noting that

$$\begin{aligned} |A_1(a_1,b_1)\Delta A_1(a_2,b_2)| &= |A_2(a_1,b_1)\Delta A_2(a_2,b_2)| \\ &\leq \int_0^1 |H(a_1,b_1)(x) - H(a_2,b_2)(x)| \,\mathrm{d}x \\ &\leq \|H(a_1,b_1) - H(a_2,b_2)\|_\infty \end{aligned}$$

and using the continuity of H into $L^\infty([0,1])$.

For example, one may take H to be given by

$$H(a,b)(x) = a + (b-a)x + x\sin(6\pi x)/10$$

which can be seen to be continuous into $L^\infty([0,1])$. Example sets $A_i(a,b)$ for various parameters a, b, with this choice of H, are shown in Figure 2.3.

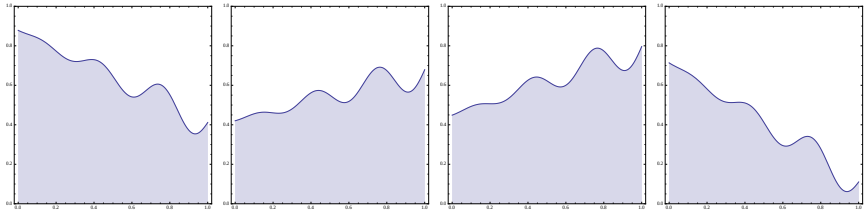


Figure 2.3: Possible sets A_i, corresponding to Example 2.2.2

Example 2.2.3. *We can generalize the previous example to allow the inclusion of a fault. Let $D = [0,1]^2$, $\Lambda = [0,1]^2 \times [-1,1]$ and $N = 2$. Let $p \in (0,1)$ denote the*

horizontal location of the fault. Given $H : [0,1]^2 \to L^\infty([0,1])$ *as in the previous example, define* $\tilde{H} : \Lambda \to L^\infty([0,1])$ *by*

$$\tilde{H}(a,b,c)(x) = \begin{cases} H(a,b)(x) & x \in [0,p] \\ c + H(a,b)(x) & x \in (p,1] \end{cases}$$

so that the parameter c *determines the (signed) magnitude of the fault. Defining the sets* $A_1(a,b,c)$ *and* $A_2(a,b,c)$ *as the regions of* D *beneath and above the curve* $\tilde{H}(a,b,c)$ *respectively, the continuity can be seen in a similar manner to the previous example. Example sets* $A_i(a,b,c)$ *for various parameters* a,b,c *are shown in Figure 2.4.*

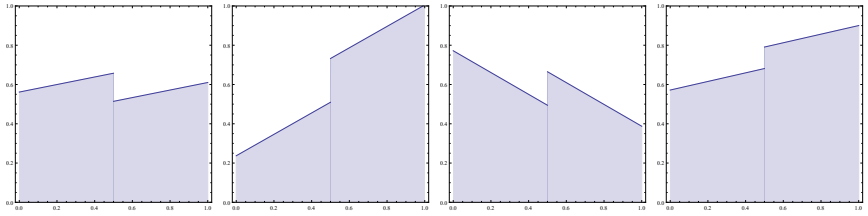


Figure 2.4: Possible sets A_i, corresponding to Example 2.2.3 in the case $p = 1/2$.

Example 2.2.4. *Again working with* $D = [0,1]^2$*, but with a much larger parameter space, one could also select points at specific* x*-coordinates and linearly interpolate between them. Fix* $K, N \in \mathbb{N}$ *and set* $\Lambda = \Xi_{N-1}^K \subseteq [0,1]^{(N-1)\times K}$*, where* Ξ_{N-1} *is the simplex*

$$\Xi_{N-1} = \{(y_1, \ldots, y_{N-1}) \in [0,1]^{N-1} \mid 0 \leq y_1 \leq \ldots \leq y_{N-1} \leq 1\}.$$

Then given $a \in \Lambda$*, define the functions* $f_i(a)$*,* $i = 1, \ldots, N-1$*, to be the linear interpolation of the points* $\left(\frac{j-1}{K-1}, a_{ij}\right)_{j=1}^K$*. The sets* $A_i(a)$*,* $i = 1, \ldots, N-1$*, are then defined to be the regions between the graphs of the functions* $f_i(a)$ *and* $f_{i-1}(a)$*, and* $A_N(a) = D \setminus \cup_{i=1}^{N-1} A_i(a)$*. Note that the sets* $(A_i(a,b))_{i=1}^N$ *are disjoint for fixed* $a, b \in \Lambda$ *by the choice of* Λ*. Example sets* $A_i(a,b)$ *for various parameters* a,b*, with this choice of* H*, are shown in Figure 2.5.*

In order to see the continuity of these maps, we first partition the domain into strips D_j*,*

$$D_j = \left\{ (x,y) \in D \;\middle|\; \frac{j-1}{K-1} \leq x \leq \frac{j}{K-1} \right\}, \quad j = 1, \ldots, K-1$$

so that we have

$$A_i(a) = \bigcup_{j=1}^{K-1} A_i(a) \cap D_j.$$

It follows from properties of the symmetric difference that

$$|A_i(a) \Delta A_i(b)| \leq \sum_{j=1}^{K-1} |(A_i(a) \cap D_j) \Delta (A_i(b) \cap D_j)|.$$

It hence suffices to show that the maps $A_i(\cdot) \cap D_j$ are continuous for all i, j. This follows from the same argument as in Example 2.2.2, for sufficiently small $|a - b|$.

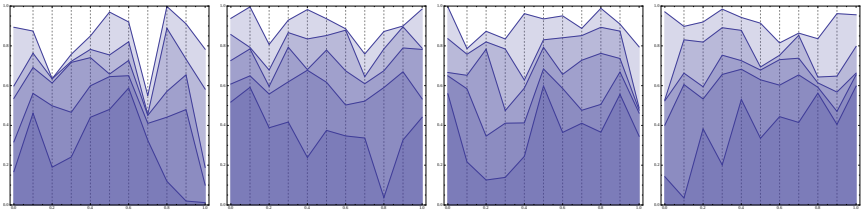


Figure 2.5: Possible sets A_i, corresponding to Example 2.2.4 in the case $K = 11$, $N = 6$

2.2.2 The Darcy Model for Groundwater Flow

We consider the Darcy model for groundwater flow on a domain $D \subseteq \mathbb{R}^d$, $d = 1, 2, 3$. Let $\kappa = (\kappa_{ij})$ denote the permeability tensor of the medium, p the pressure of the water, and assume the viscosity of the water is constant. Darcy's law [37] tells us that the velocity is proportional to the gradient of the pressure:

$$v = -\kappa \nabla p.$$

Additionally, a local form of mass conservation tells us that

$$\nabla \cdot v = f.$$

where f is a recharge term. Combining these two equations, and imposing Dirichlet boundary conditions for simplicity, results in the PDE

$$\begin{cases} -\nabla \cdot (\kappa \nabla p) = f & \text{in } D \\ p = g & \text{on } \partial D. \end{cases}$$

This is the PDE we will consider in the forward model, and it gives rise to a solution map $\kappa \mapsto p$.

For simplicity we will work in the case where κ is an isotropic (scalar) permeability, bounded above and below by positive constants, and so it can be represented as the image of some bounded function under a positive continuously differentiable map $\sigma : \mathbb{R} \to \mathbb{R}^+$.

Let $V = H^1(D)$, the Sobolev space of square integrable once weakly differentiable functions on D [55]. Then given $f \in H^{-1}(D)$, $g \in H^{1/2}(\partial D)$, $u \in X$ and $a \in \Lambda$, define $p_{u,a} \in V$ to be the solution of the weak form of the PDE

$$\begin{cases} -\nabla \cdot (\sigma(u^a) \nabla p_{u,a}) = f & \text{in } D \\ p_{u,a} = g & \text{on } \partial D. \end{cases} \tag{2.2.2}$$

We are first interested in the regularity of the map $\mathcal{R} : X \times \Lambda \to V$ given by $\mathcal{R}(u, a) = p_{u,a}$. We first recall what it means for $p_{u,a}$ to be a solution of (2.2.2). Since $g \in H^{1/2}(\partial D)$, by the trace theorem [55] there exists $G \in V$ such that $\mathrm{tr}(G) = g$. The solution $p_{u,a}$ of (2.2.2) is then given by $p_{u,a} = q_{u,a} + G$, where $q_{u,a} \in H_0^1(D)$ solves the PDE

$$\begin{cases} -\nabla \cdot \big(\sigma(u^a) \nabla q_{u,a}\big) = f + \nabla \cdot \big(\sigma(u^a) \nabla G\big) & \text{in } D \\ q_{u,a} = 0 & \text{on } \partial D. \end{cases} \tag{2.2.3}$$

The following lemma tells us that the map $\mathcal{R}$ is well defined and has certain regularity properties. Its proof is given in the appendix.

Lemma 2.2.5. *The map $\mathcal{R} : X \times \Lambda \to V$ is well-defined and satisfies:*

(i) for each $(u, a) \in X \times \Lambda$,

$$\|\mathcal{R}(u, a)\|_V \leq (\|f\|_{V^*} + \|\sigma(u^a)\|_{L^\infty} \|G\|_V)/\kappa_{\min}(u, a) + \|G\|_V$$

where $\kappa_{\min}(u,a)$ *is given by*

$$\kappa_{\min}(u,a) = \operatorname*{essinf}_{x\in D} \sigma(u^a(x)) > 0$$

and $\|f\|_{V^*}$ *denotes the dual norm of* f *as defined in the Appendix;*

(ii) for each $a \in \Lambda$, $\mathcal{R}(\cdot,a) : X \to V$ *is locally Lipschitz continuous, i.e. for every* $r > 0$ *there exists* $L(r) > 0$ *such that, for all* $u, v \in X$ *with* $\|u\|_X, \|v\|_X < r$ *and all* $a \in \Lambda$, *we have*

$$\|\mathcal{R}(u,a) - \mathcal{R}(v,a)\|_V \leq L(r)\|u - v\|_X;$$

(iii) for each $u \in X$, $\mathcal{R}(u,\cdot) : \Lambda \to V$ *is continuous.*

We now choose a continuous linear observation operator $\ell : V \to \mathbb{R}^J$. For example, writing $\ell = (\ell_1, \ldots, \ell_J)$, we could take

$$\ell_i(p) = \int_D \frac{1}{(2\pi\varepsilon)^{d/2}} e^{-|x_i - y|^2/2\varepsilon} p(y)\,\mathrm{d}x, \quad i = 1, \ldots, J \tag{2.2.4}$$

for some $\varepsilon > 0$, so that ℓ_i approximates a point observation at the point $x_i \in D$. Our forward operator $\mathcal{G} : X \times \Lambda \to \mathbb{R}^J$ is then defined by $\mathcal{G} = \ell \circ \mathcal{R}$, so that it can be written as the composition

$$(u,a) \mapsto u^a \mapsto \kappa = \sigma(u^a) \mapsto p \mapsto \ell(p)$$

From the above regularity of $\mathcal{R}$ we can deduce the following regularity properties of our forward operator $\mathcal{G}$:

Proposition 2.2.6. *Define the map* $\mathcal{G} : X \times \Lambda \to \mathbb{R}^J$ *as above. Then* $\mathcal{G}$ *satisfies*

1. *For each* $r > 0$ *and* $u, v \in X$ *with* $\|u\|_X, \|v\|_X < r$, *there exists* $C(r) > 0$ *such that for all* $a \in \Lambda$,

$$|\mathcal{G}(u,a) - \mathcal{G}(v,a)| \leq C(r)\|u - v\|_X.$$

2. *For each* $u \in X$, *the map* $\mathcal{G}(u,\cdot) : \Lambda \to \mathbb{R}^J$ *is continuous.*

Proof. 1. The map ℓ is defined to be a continuous linear functional, and so in particular is Lipschitz. Since we have $\mathcal{G} = \ell \circ \mathcal{R}$ the result follows from Lemma 2.2.5(ii).

2. This follows from the continuity of ℓ and Lemma 2.2.5(iii).

□

2.2.3 The Complete Electrode Model for EIT

Electrical Impedance Tomography (EIT) is an imaging technique that aims to make inference about the internal conductivity of a body from surface voltage measurements. Electrodes are attached to the surface of the body, current is injected, and the resulting voltages on the electrodes are measured. Applications include both medical imaging, where the aim is to non-invasively detect internal abnormalities within a human patient, and subsurface imaging, where material properties of the subsurface are differentiated via their conductivities. Early references include [66] in the context of medical imaging and [90] in the context of subsurface imaging.

The complete electrode model (CEM) is proposed for the forward model in [128], and shown to agree with experimental data up to measurement precision. In its strong form, the PDE reads

$$\begin{cases} -\nabla \cdot (\kappa(x)\nabla v(x)) = 0 & x \in D \\ \displaystyle\int_{e_l} \kappa \frac{\partial v}{\partial n} \,\mathrm{d}S = I_l & l = 1, \dots, L \\ \kappa(x)\dfrac{\partial v}{\partial n}(x) = 0 & x \in \partial D \setminus \bigcup_{l=1}^{L} e_l \\ v(x) + z_l \kappa(x)\dfrac{\partial v}{\partial n}(x) = V_l & x \in e_l, l = 1, \dots, L. \end{cases} \tag{2.2.5}$$

The domain D represents the body, and $(e_l)_{l=1}^L \subseteq \partial D$ the electrodes attached to its surface with corresponding contact impedances $(z_l)_{l=1}^L$. A current I_l is injected into each electrode e_l, and a voltage measurement V_l made. Here κ represents the conductivity of the body, and v the potential within it. Note that the solution comprises both a function $v \in H^1(D)$ and a vector $(V_l)_{l=1}^L \in \mathbb{R}^L$ of boundary voltage measurements.

A corresponding weak form exists, and is shown to have a unique solution (up to constants) given appropriate conditions on κ, $(z_l)_{l=1}^L$ and $(I_l)_{l=1}^L$ – see [128] for details. Moreover, under some additional assumptions, the mapping $\kappa \mapsto (V_l)_{l=1}^L$ is known to be Fréchet differentiable when we equip the conductivity space with the supremum norm [73].

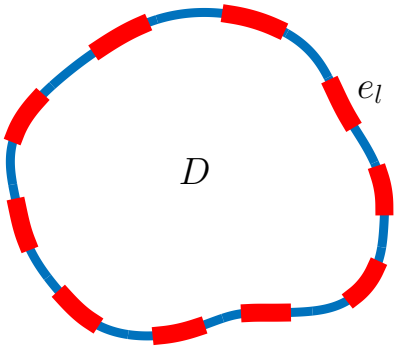


Figure 2.6: An example domain D, with attached electrodes $(e_l)_{l=1}^L$, for the EIT problem.

We can apply different current stimulation patterns to the electrodes, that is, different values of $(I_l)_{l=1}^L$ in (2.2.5), to yield additional information. Assume that we have M different (linearly independent) current stimulation patterns $(I^{(m)})_{m=1}^M$. This yields M different mappings $\kappa \mapsto (V_l^{(m)})_{l=1}^L$ each with the regularity above, or equivalently a mapping $\kappa \mapsto V$ where $V \in \mathbb{R}^J$ with $J = LM$.

Analogously to the Darcy model case, we will consider isotropic conductivities of the form $\kappa = \sigma(u^a)$, where $\sigma : \mathbb{R} \to \mathbb{R}^+$ is positive and continuously differentiable. Our forward operator $\mathcal{G} : X \times \Lambda \to \mathbb{R}^J$, is then given by the composition

$$(u, a) \mapsto u^a \mapsto \kappa = \sigma(u^a) \mapsto \left((v^{(1)}, V^{(1)}), \ldots, (v^{(M)}, V^{(M)})\right) \mapsto (V^{(1)}, \ldots, V^{(M)}).$$

We show in the appendix that the map defined in this way has the same regularity as the map corresponding to the Darcy model.

Proposition 2.2.7. *Define the map $\mathcal{G} : X \times \Lambda \to \mathbb{R}^J$ as above. Then $\mathcal{G}$ satisfies*

1. *For each $r > 0$ and $u, v \in X$ with $\|u\|_X, \|v\|_X < r$, there exists $C(r) > 0$ such that for all $a \in \Lambda$,*

$$|\mathcal{G}(u, a) - \mathcal{G}(v, a)| \leq C(r)\|u - v\|_X.$$

2. *For each $u \in X$, the map $\mathcal{G}(u, \cdot) : \Lambda \to \mathbb{R}^J$ is continuous.*

2.3 Onsager-Machlup Functionals and Prior Modelling

In this section we recall the definition of an Onsager-Machlup functional for a measure which is equivalent[1] to a Gaussian measure. We then introduce the prior measures that we will consider, first on the function space X, then the geometric parameter space Λ, and finally the product space $X \times \Lambda$. We conclude the section by extending the definition of Onsager-Machlup functional so that it is appropriate for the measures we consider here, supported on fields and geometric parameters which are combined to make piecewise continuous functions.

2.3.1 Onsager-Machlup Functionals

The Onsager-Machlup functional of a measure is the negative logarithm of its Lebesgue density when such a density exists, and otherwise can be thought of analogously. We start by defining it precisely for measures defined via density with respect to a Gaussian, allowing for infinite dimensional spaces on which the Lebesgue measure is not defined. Suppose that μ is a measure equivalent to a Gaussian measure μ_0; the definition of such measures is found in the Appendix. Then the Onsager-Machlup functional for μ is defined as follows.

Definition 2.3.1 (Onsager-Machlup functional I)**.** *Let μ be a measure on a Banach space Z which is equivalent to μ_0, where μ_0 is a Gaussian measure on Z with Cameron-Martin space E. Let $B^\delta(z)$ denote the ball of radius δ centred at $z \in Z$. A functional $I : Z \to \overline{\mathbb{R}}$ is called the* Onsager-Machlup functional *for μ if, for each $x, y \in E$,*

$$\lim_{\delta \downarrow 0} \frac{\mu(B^\delta(x))}{\mu(B^\delta(y))} = \exp\left(I(y) - I(x)\right)$$

and $I(x) = \infty$ for $x \notin E$.

Remarks 2.3.2. *(i) The Onsager-Machlup functional is only defined up to addition of a constant.*

(ii) If Z is finite dimensional and μ admits a positive Lebesgue density ρ, then $I(x) = -\log \rho(x)$ for all $x \in Z$. In light of the previous remark, this is true even if ρ is not normalized.

[1] Two measures ν, μ on a measurable space $(M, \mathcal{M})$ are equivalent if $\nu(A) = 0$ if and only if $\mu(A) = 0$, for $A \in \mathcal{M}$.

(iii) *Let $Z = \mathbb{R}^n$ be finite dimensional, and let $\mu_0 = N(0, \Sigma)$ be a Gaussian measure on Z. Let $\Gamma \in \mathbb{R}^{m\times m}$ be a positive-definite matrix, $A \in \mathbb{R}^{m\times n}$ and $y \in \mathbb{R}^m$. Define μ by*

$$\frac{\mathrm{d}\mu}{\mathrm{d}\mu_0}(x) \propto \exp\left(-\frac{1}{2}|Ax - y|_\Gamma^2\right)$$

so that

$$\frac{\mathrm{d}\mu}{\mathrm{d}x}(x) \propto \exp\left(-\frac{1}{2}|Ax - y|_\Gamma^2 - \frac{1}{2}|x|_\Sigma^2\right).$$

Then by the previous remark, the Onsager-Machlup functional for μ is given by

$$I(x) = \frac{1}{2}|Ax - y|_\Gamma^2 + \frac{1}{2}|x|_\Sigma^2$$

for all $x \in Z$, which is a Tikhonov regularized least squares functional.

(iv) *The preceding example (iii) may be extended to an infinite dimensional setting. Let Z be a separable Banach space, and let $\mu_0 = N(0, \mathcal{C}_0)$ be a Gaussian measure on Z with Cameron-Martin space $(E, \langle\cdot,\cdot\rangle_E, \|\cdot\|_E)$. Let $\Gamma \in \mathbb{R}^{m\times m}$ be a positive-definite matrix, $A : X \to \mathbb{R}^m$ a bounded linear operator and $y \in \mathbb{R}^m$. Define μ by*

$$\frac{\mathrm{d}\mu}{\mathrm{d}\mu_0}(x) \propto \exp\left(-\frac{1}{2}|Ax - y|_\Gamma^2\right).$$

Then Theorem 3.2 in [38] tells us that the Onsager-Machlup functional for μ is given by

$$I(x) = \frac{1}{2}|Ax - y|_\Gamma^2 + \frac{1}{2}\|x\|_E^2.$$

(v) *In this chapter, the posterior distribution will be a measure on the product space $Z = X \times \Lambda$. The prior distribution will be an independent product of a Gaussian on X and a compactly supported measure on Λ. Due to the assumption of compact support, the prior will not be equivalent to a Gaussian measure on Z and so the above definition doesn't apply; we provide a suitable extension to the definition in subsection 2.3.4.*

As we are taking a Bayesian approach to the inverse problem, we incorporate our prior beliefs about the permeability/conductivity into the model via probability

measures on X and Λ. We will combine these into a prior measure on the product space $X \times \Lambda$. We equip this space with any (complete) norm $\|(\cdot,\cdot)\|$ such that if $\|(u,a)\| \to 0$, then $\|u\|_X \to 0$ and $|a| \to 0$.

2.3.2 Priors for the Fields

We wish to put priors on the fields $u_1, \ldots, u_N \in C^0(D)$. We use independent Gaussian measures $u_i \sim \mu_0^i := N(m_i, \mathcal{C}_i)$, where the means $m_i \in C^0(D)$, and each covariance operator $\mathcal{C}_i : C^0(D) \to C^0(D)$ is trace-class and positive definite. It follows that the vector $(u_1, \ldots, u_N) \sim \mu_0^1 \times \ldots \times \mu_0^N =: \mu_0$ is Gaussian on X:

$$\mu_0 = N\left(m, \bigoplus_{i=1}^{N} \mathcal{C}_i\right)$$

where $m = (m_1, \ldots, m_N) \in X$. If E_i denotes the Cameron-Martin space [39] of μ_0^i, then that of μ_0 is given by

$$E = \bigoplus_{i=1}^{N} E_i$$

with inner product given by the sum of those of its component spaces.

The Onsager-Machlup functional of μ_0 is known to be given by

$$J(u) = \begin{cases} \frac{1}{2}\|u-m\|_E^2 & u - m \in E \\ \infty & u - m \notin E. \end{cases}$$

This can be seen, for example, as a consequence of Proposition 18.3 in [97].

Remark 2.3.3. *We may assume that the different fields are correlated under the prior, so long as μ_0 remains Gaussian on X – this does not affect any of the following theory. Allowing correlations between the fields and the geometric parameters under the prior is a more technical issue however, and so we will assume that these are independent.*

Example 2.3.4. *Define the negative Laplacian with Neumann boundary conditions as follows:*

$$A = -\Delta, \quad \mathcal{D}(A) = \left\{ u \in H^2(D) \;\middle|\; \frac{\mathrm{d}u}{\mathrm{d}\nu} = 0 \text{ on } \partial D, \int_D u(x)\,\mathrm{d}x = 0 \right\}.$$

Then A is invertible. We can define $\mathcal{C}_i = A^{-\alpha_i}$, where each $\alpha_i > d/2$. Then each $\mathcal{C}_i$ is trace-class and positive definite, and samples from each μ_0^i will be almost surely continuous and so μ_0 can be considered as a Gaussian measure on X. Moreover, regularity of the samples will increase as α_i increases, see [39] for details.

2.3.3 Priors for the Geometric Parameters

We also want to put a prior measure on the geometric parameters, i.e. we want to choose a probability measure on Λ. Since $\Lambda \subseteq \mathbb{R}^k$ the analysis is more straightforward than the infinite dimensional case. Let ν be a probability measure on Λ with compact support $S \subseteq \Lambda$. We assume ν is absolutely continuous with respect to the Lebesgue measure and that its density ρ is continuous on S. Despite being defined on a finite dimensional space, the measure ν is not necessarily equivalent to the Lebesgue measure on the whole of $\mathbb{R}^k$ and so the previous definition of Onsager-Machlup functional does not apply. We hence must formulate a new definition for this case.

Since $\rho > 0$ on $\text{int}(S)$, we can use the continuity of ρ to calculate the limits of ratios of small ball probabilities for ν on $\text{int}(S)$. Let $a_1, a_2 \in \text{int}(S)$, then

$$\begin{aligned}
\lim_{\delta \downarrow 0} \frac{\nu(B^\delta(a_1))}{\nu(B^\delta(a_2))} &= \lim_{\delta \downarrow 0} \frac{\int_{B^\delta(a_1)} \rho(a)\,\mathrm{d}a}{\int_{B^\delta(a_2)} \rho(a)\,\mathrm{d}a} \\
&= \lim_{\delta \downarrow 0} \frac{\frac{1}{|B^\delta(a_1)|}\int_{B^\delta(a_1)} \rho(a)\,\mathrm{d}a}{\frac{1}{|B^\delta(a_2)|}\int_{B^\delta(a_2)} \rho(a)\,\mathrm{d}a} \\
&= \frac{\rho(a_1)}{\rho(a_2)} \\
&= \exp\left(\log \rho(a_1) - \log \rho(a_2)\right).
\end{aligned}$$

If either a_1 or a_2 lie outside of S the limit can be seen to be 0 or ∞ respectively. It hence makes sense to define the Onsager-Machlup functional for ν on $\Lambda \setminus \partial S$ as

$$K(a) = \begin{cases} -\log \rho(a) & a \in \text{int}(S) \\ \infty & a \notin S. \end{cases}$$

For $a \in \partial S$, we define $K(a)$ to be the limit of K from the interior:

$$K(a) = -\lim_{\substack{b \to a \\ b \in \text{int}(S)}} \log \rho(b) \quad a \in \partial S$$

which is well defined due to the continuity of ρ on $\text{int}(S)$. K is then continuous on the whole of S.

Remark 2.3.5. *If we were to define K on ∂S in the same way that we defined it on $\Lambda \setminus \partial S$, K would have a positive jump at the boundary related to the geometry of S. This would mean that K was not lower semi-continuous on S which would cause problems when seeking minimizers. The definition we have chosen is appropriate: if any minimizing sequence $(a_n)_{n\geq 1} \subseteq \text{int}(S)$ of K has an accumulation point on ∂S, then ν has a mode at that point.*

If we have no prior knowledge about the interfaces and Λ is compact, we could place a uniform prior on the whole of Λ. Otherwise we could either choose a prior with smaller support, or one that weights certain areas more than others.

2.3.4 Priors on $X \times \Lambda$

We assume that the priors on the fields and the geometric parameters are independent, so that we may take the product measure $\mu_0 \times \nu_0$ as our prior on $X \times \Lambda$. Note that if $F : X \times \Lambda \to L^\infty(D)$ denotes the construction map $(u, a) \mapsto u^a$ defined earlier by (2.2.1), then our prior permeability/conductivity distribution on $L^\infty(D)$ is given by the pushforward $\mu_0^* = F^\#(\mu_0 \times \nu_0)$, where the pushforward is as defined in the Appendix. This is much more cumbersome to deal with however, since for example $L^\infty(D)$ is not separable. It is for this reason we incorporate the mapping F into the forward map $\mathcal{G}$. Assuming now that the prior $\mu_0 \times \nu_0$ is as described above, we can define the Onsager-Machlup functional for measures μ on $X \times \Lambda$ which are equivalent to $\mu_0 \times \nu_0$.

Definition 2.3.6 (Onsager-Machlup functional II)**.** *Let μ be a measure on $X \times \Lambda$ equivalent to $\mu_0 \times \nu_0$, where μ_0 and ν_0 satisfy the assumptions detailed above. Let $B^\delta(u, a)$ denote the ball of radius δ centred at $(u, a) \in X \times \Lambda$. A functional $I : X \times \Lambda \to \overline{\mathbb{R}}$ is called the* Onsager-Machlup functional *for μ if,*

(i) for each $(u, a), (v, b) \in E \times \text{int}(S)$,

$$\lim_{\delta \downarrow 0} \frac{\mu(B^\delta(u, a))}{\mu(B^\delta(v, b))} = \exp\left(I(v, b) - I(u, a)\right);$$

(ii) for each $(u, a) \in E \times \partial S$,

$$I(u,a) = \lim_{\substack{b \to a \\ b \in \mathrm{int}(S)}} I(u,b);$$

(iii) $I(u,a) = \infty$ *for* $u \notin E$ *or* $a \notin S$.

2.4 Likelihood and Posterior Distribution

We return to the abstract setting mentioned in the introduction. Let X be a separable Banach space, $\Lambda \subseteq \mathbb{R}^k$ and $Y = \mathbb{R}^J$. Suppose we have a forward operator $\mathcal{G} : X \times \Lambda \to Y$. If (u, a) denotes the true input to our forward problem, we observe data $y \in Y$ given by

$$y = \mathcal{G}(u,a) + \eta$$

where $\eta \sim \mathbb{Q}_0 := N(0, \Gamma)$, $\Gamma \in \mathbb{R}^{J \times J}$ positive definite, is Gaussian noise on Y independent of the prior.

It is clear that we have $y|(u,a) \sim \mathbb{Q}_{u,a} := N(\mathcal{G}(u,a), \Gamma)$. We can use this to formally find the distribution of $(u,a)|y$. First note that

$$\mathbb{Q}_{u,a}(\mathrm{d}y) = \exp\left(-\Phi(u,a;y) + \frac{1}{2}|y|_\Gamma^2\right)\mathbb{Q}_0(\mathrm{d}y)$$

where the potential (or negative log-likelihood) $\Phi : X \times \Lambda \times Y \to \mathbb{R}$ is given by

$$\Phi(u,a;y) = \frac{1}{2}|\mathcal{G}(u,a) - y|_\Gamma^2. \tag{2.4.1}$$

Hence under suitable regularity conditions, Bayes' theorem tells us that the distribution μ of $(u,a)|y$ satisfies

$$\mu(\mathrm{d}u, \mathrm{d}a) \propto \exp\big(-\Phi(u,a;y)\big)\mu_0(\mathrm{d}u)\nu_0(\mathrm{d}a)$$

after absorbing the $\exp\left(\frac{1}{2}|y|_\Gamma^2\right)$ term into the normalisation constant.

We now make this statement rigorous. To keep the situation general, we do not insist that Φ takes the form (2.4.1), and instead assert only that Φ satisfies the following assumptions.

Assumptions 2.4.1. *There exists $X' \times \Lambda' \subseteq X \times \Lambda$ such that*

(i) for every $\varepsilon > 0$ there is an $M_1(\varepsilon) \in \mathbb{R}$ such that for all $u \in X'$ and all $a \in \Lambda'$

$$\Phi(u, a; y) \geq M_1(\varepsilon) - \varepsilon \|u\|_X^2;$$

(ii) for each $u \in X'$ and $y \in Y$, the potential $\Phi(u, \cdot; y) : \Lambda' \to \mathbb{R}$ is continuous;

(iii) there exists a strictly positive $M_2 : \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ monotonic non-decreasing separately in each argument, such that for each $r > 0$, $u \in X'$ and $a \in \Lambda'$, and $y_1, y_2 \in Y$ with $|y_1|, |y_2| < r$,

$$|\Phi(u, a; y_1) - \Phi(u, a; y_2)| \leq M_2(r, \|u\|_X, |a|)|y_1 - y_2|;$$

(iv) there exists a strictly positive $M_3 : \mathbb{R}^+ \times \Lambda \times Y \to \mathbb{R}^+$, continuous in its second component, such that for each $r > 0$, $a \in \Lambda'$ and $y \in Y$, and $u_1, u_2 \in X'$ with $\|u_1\|_X, \|u_2\|_X < r$,

$$|\Phi(u_1, a; y) - \Phi(u_2, a; y)| \leq M_3(r, a, y)\|u_1 - u_2\|_X.$$

These assumptions are used in the proof of existence and well-posedness of the posterior distribution, which is given in the appendix:

Theorem 2.4.2 (Existence and well-posedness)**.** *Let Assumptions 2.4.1 hold. Assume that $(\mu_0 \times \nu_0)(X' \times \Lambda') = 1$, and that $(\mu_0 \times \nu_0)((X' \times \Lambda') \cap B) > 0$ for some bounded set $B \subseteq X \times \Lambda$. Then*

(i) Φ is $\mu_0 \times \nu_0 \times \mathbb{Q}_0$-measurable;

(ii) for each $y \in Y$, $Z(y)$ given by

$$Z(y) = \int_{X \times \Lambda} \exp(-\Phi(u, a; y))\, \mu_0(\mathrm{d}u)\nu_0(\mathrm{d}a)$$

is positive and finite, and so the probability measure μ^y,

$$\mu^y(\mathrm{d}u, \mathrm{d}a) = \frac{1}{Z(y)} \exp(-\Phi(u, a; y))\, \mu_0(\mathrm{d}u)\nu_0(\mathrm{d}a) \tag{2.4.2}$$

is well-defined.

(iii) Assume additionally that, for every fixed $r > 0$, there exists $\varepsilon > 0$ with

$$\exp(\varepsilon\|u\|_X^2)(1 + M_2(r, \|u\|_X, |a|)^2) \in L^1_{\mu_0 \times \nu_0}(X \times \Lambda; \mathbb{R}).$$

Then there is $C(r) > 0$ such that for all $y, y' \in Y$ with $|y|, |y'| < r$,

$$d_{\text{Hell}}(\mu^y, \mu^{y'}) \leq C|y - y'|.$$

where d_{Hell} is the Hellinger metric, defined in the Appendix.

Remark 2.4.3. *In this chapter we are focused on the case when the field prior μ_0 is taken to be Gaussian. However, the above existence and well-posedness result still holds if, for example, μ_0 is taken to be Besov rather than Gaussian, since a Fernique-type theorem holds for such priors [39, 94].*

We show that for both choices of test models, the potential (2.4.1) satisfies Assumptions 2.4.1:

Proposition 2.4.4. *Let $X = C^0(D; \mathbb{R}^N)$, and let $\mathcal{G} : X \times \Lambda \to Y$ denote the forward map corresponding to either the groundwater flow or EIT problem, as detailed in section 2.2. Let $y \in Y$ and let $\Gamma \in \mathbb{R}^{J \times J}$ be positive definite. Define the potential $\Phi : X \times \Lambda \times Y \to \mathbb{R}$ by*

$$\Phi(u, a; y) = \frac{1}{2}|\mathcal{G}(u, a) - y|^2_\Gamma.$$

Then Φ satisfies Assumptions 2.4.1, with $X' \times \Lambda' = X \times \Lambda$.

Proof. (i) $\Phi \geq 0$ so this is true with $M_1 \equiv 0$.

(ii) Fix $u \in X'$ and $y \in Y$. Propositions 2.2.6 and 2.2.7 tell us that $\mathcal{G}(u, \cdot)$ is continuous for either choice of test model. The map $z \mapsto |z - y|^2_\Gamma$ is continuous, and so $\Phi(u, \cdot; y)$ is continuous too.

(iii) A consequence of Propositions 2.2.6 and 2.2.7 is that for each $u \in X$ and $a \in \Lambda$, $\mathcal{G}(u, a)$ can be bounded in terms of $\|u\|_X$ and $|a|$. The result then follows from the local Lipschitz property of the map $y \mapsto |y|^2$.

(iv) Propositions 2.2.6 and 2.2.7 tell us that $\mathcal{G}(\cdot, a)$ is locally Lipschitz for either choice of test model. The map $z \mapsto |z - y|^2_\Gamma$ is locally Lipschitz, and hence we conclude that $\Phi(\cdot, a; y)$ is locally Lipschitz, with Lipschitz constant independent of a.

□

With a choice of prior as described in section 2.3, we can therefore apply Theorem 2.4.2 in the cases where the forward map is one of the two described in section 2.2 and the observational noise is Gaussian. In this case, the constant $M_2(r, \|u\|_X, |a|)$ appearing in Assumptions 2.4.1(iii) is independent of $\|u\|_X$ and $|a|$, and so the integrability condition (iii) in Theorem 2.4.2 always holds via Fernique's theorem. The condition on positivity of a bounded set can be seen by taking, for example, $B = B^1(0) \times S$, where S is the (compact) support of ν_0.

2.5 MAP Estimators

In subsection 2.5.1 we characterize the MAP estimators for the posterior μ in terms of the Onsager-Machlup functional for μ. In subsection 2.5.2 we relate this Onsager-Machlup functional to the Fomin derivative of μ, with reference to the work [65].

2.5.1 MAP Estimators and the Onsager-Machlup Functional

Throughout this section we assume that μ is given by (2.4.2). Furthermore we assume that μ_0 has mean zero for simplicity. Additionally, when we assume that Assumptions 2.4.1 hold, we will assume that $X' \times \Lambda' = X \times \Lambda$.

Suppressing the dependence of Φ on the data y since it is not relevant in the sequel, we define the functional $I : X \times \Lambda \to \mathbb{R}$ by

$$I(u, a) = \Phi(u, a) + J(u) + K(a) \tag{2.5.1}$$

where J, K are as defined in subsections 2.3.2, 2.3.3 respectively:

$$J(u) = \begin{cases} \frac{1}{2}\|u - m\|_E^2 & u - m \in E \\ \infty & u - m \notin E, \end{cases} \quad \text{and} \quad K(a) = \begin{cases} -\log \rho(a) & a \in \operatorname{int}(S) \\ -\lim\limits_{\substack{b \to a \\ b \in \operatorname{int}(S)}} \log \rho(b) & a \in \partial S \\ \infty & a \notin S. \end{cases}$$

In this section we attain the following three results concerning I and μ, which are proved in the appendix.

Theorem 2.5.1. *Let Assumptions 2.4.1 hold. Then the function I defined by (2.5.1) is the Onsager-Machlup functional for μ, where the Onsager-Machlup functional is as defined in Definition 2.3.6.*

Theorem 2.5.2. *Let Assumptions 2.4.1 hold. Then there exists $(\bar{u}, \bar{a}) \in E \times S$ such that*

$$I(\bar{u}, \bar{a}) = \inf\{I(u, a) \,|\, u \in E,\ a \in S\}.$$

Furthermore, if $(u_n, a_n)_{n\geq 1}$ is a minimizing sequence satisfying $I(u_n, a_n) \to I(\bar{u}, \bar{a})$, then there is a subsequence $(u_{n_k}, a_{n_k})_{k\geq 1}$ converging to $(\bar{u}, \bar{a})$ (strongly) in $E \times S$.

Theorem 2.5.3. *Let Assumptions 2.4.1 hold. Assume also that there exists an $M \in \mathbb{R}$ such that $\Phi(u, a) \geq M$ for any $(u, a) \in X \times \Lambda$.*

(i) Let $(u^\delta, a^\delta) = \underset{(u,a)\in X\times\Lambda}{\operatorname{argmax}} \mu(B^\delta(u, a))$. There is a $(\bar{u}, \bar{a}) \in E \times S$ and a subsequence of $(u^\delta, a^\delta)_{\delta>0}$ which converges to $(\bar{u}, \bar{a})$ strongly in $X \times \Lambda$.

(ii) The limit $(\bar{u}, \bar{a})$ is a MAP estimator and minimizer of I.

A consequence of Theorem 2.5.3 is that, under its assumptions, MAP estimators and minimizers of the Onsager-Machlup functional are equivalent. The proof of this corollary is identical to that of Corollary 3.10 in [38]:

Corollary 2.5.4. *Under the conditions of Theorem 2.5.3 we have the following.*

(i) Any MAP estimator minimizes the Onsager-Machlup functional I.

(ii) Any $(u^*, a^*) \in E \times S$ which minimizes the Onsager-Machlup functional I is a MAP estimator for the measure μ given by (2.4.2).

2.5.2 The Fomin Derivative Approach

In recent work of Helin and Burger [65], the concept of MAP estimators was generalized to weak MAP (wMAP) estimators using the notion of Fomin differentiability of measures. The definition of wMAP estimators is such that if $\hat{u}$ is a MAP estimator then it is a wMAP estimator, but not necessarily vice versa. Under certain assumptions, they show that wMAP estimators are equivalent to minimizers of a particular functional. The assumptions do not hold in our case, since our forward map is non-linear and our prior $\mu_0 \times \nu_0$ isn't necessarily convex, however the functional agrees with our objective functional I. Thus in what follows we provide a

link between the Fomin derivative of the posterior μ and our objective functional I.

The Fomin derivative of a measure on a Banach space X equipped with its Borel σ-algebra $\mathcal{B}(X)$ is defined as follows.

Definition 2.5.5. *A measure λ on X is called Fomin differentiable along the vector $z \in X$ if, for every set $A \in \mathcal{B}(X)$, there exists a finite limit*

$$d_z\lambda(A) = \lim_{t\to 0} \frac{\lambda(A+tz) - \lambda(A)}{t}.$$

The measure $d_z\lambda$ is called the Fomin derivative of λ in the direction z. The Radon-Nikodym density of $d_z\lambda$ with respect to λ is denoted β_z^λ, and is called the logarithmic derivative of λ along z.

Example 2.5.6. *(i) Let ν_0 be a measure on $\mathbb{R}^k$ with Lebesgue density ρ, supported and continuously differentiable on $S \subseteq \mathbb{R}^k$. Then for any $a \in \operatorname{int}(S)$ and $b \in \mathbb{R}^k$ we have*

$$\beta_b^{\nu_0}(a) = \frac{\nabla\rho(a)}{\rho(a)} \cdot b = \partial_b \log \rho(a).$$

(ii) Let μ_0 be a Gaussian measure on a Banach space X with Cameron-Martin space $(E, \langle\cdot,\cdot\rangle_E)$. Then for any $u \in X$ and $h \in E$ we have

$$\beta_h^{\mu_0}(u) = -\langle u, h\rangle_E.$$

This follows from the Cameron-Martin and dominated convergence theorems, both of which are given in the Appendix.

(iii) Again using the Cameron-Martin and dominated convergence theorems, we see that with ν_0 and μ_0 as above, for any $(u,a) \in X \times \operatorname{int}(S)$ and $(h,b) \in E \times \mathbb{R}^k$,

$$\beta_{(h,b)}^{\mu_0\times\nu_0}(u,a) = \beta_h^{\mu_0} + \beta_b^{\nu_0}.$$

We can use the above example to characterize the Fomin derivative (or equivalently the logarithmic derivatives) of our posterior distribution μ, given by (2.4.2).

Theorem 2.5.7. *Assume that $\Phi : X \times \Lambda \to \mathbb{R}$ is bounded measurable with uniformly bounded derivative, and assume that ρ is continuously differentiable*

on S. Then for each $(u,a) \in X \times \text{int}(S)$ and $(h,b) \in E \times \mathbb{R}^k$, we have

$$\begin{aligned}\beta^{\mu}_{(h,b)}(u,a) &= -\partial_{(h,b)}\Phi(u,a) - \langle u,h\rangle_E + \partial_b \log \rho(a) \\ &= -\partial_{(h,b)} I(u,a)\end{aligned}$$

Therefore, $(\hat{u},\hat{a})$ is a critical point of I if and only if $\beta^{\mu}_{(h,b)}(\hat{u},\hat{a}) = 0$ for all $(h,b) \in E \times \mathbb{R}^k$.

Proof. We use result (2.1.13) from [16], which tells us that if λ is a measure differentiable along z and f is a bounded measurable function with uniformly bounded partial derivative $\partial_z f$, then the measure $f \cdot \lambda$ is differentiable along z as well and

$$d_z(f \cdot \lambda) = \partial_z f \cdot \lambda + f \cdot d_z \lambda.$$

We apply this result with $\lambda = \mu_0 \times \nu_0$, $f = \exp(-\Phi)/Z$ and $z = (u,a)$. Note that f satisfies the assumptions of (2.1.13) due to the assumptions on Φ. The result then follows using Example 2.5.6 (iii) above. □

2.6 Numerical Experiments

In this section we perform some numerical experiments related to the theory above for a variety of geometric models, in the case of the groundwater flow forward map introduced in subsection 2.2.2. We both compute minimizers of the relevant Onsager-Machlup functional (i.e. MAP estimators), and we sample the posterior distribution using a state-of-the-art function space Metropolis-Hastings MCMC method. We then relate the samples to the MAP estimators. From these numerical experiments we observe the following behaviour of the posterior distribution.

1. The posterior distribution can be highly multi-modal, especially when the parameterized geometry is non-trivial. This is evident from the sensitivity of the minimisation of the objective functional on its initial state, and the behaviour of MCMC chains initialized at these calculated minimizers.

2. When the geometry is incorrect the fields attempt to compensate, which presumably contributes to the existence of multiple local minimizers of the objective functional; this occurs in both the MAP estimation and the MCMC samples. A consequence is that many of the local minimizers lack the desired

sharp interfaces. These minimizers could however be used to suggest more appropriate geometric parameters for the initialisation.

3. The mixing rates of MCMC chains have a strong dependence upon which local minimizer they are initialized at: acceptance rates can vary wildly when the initial state is changed but all other parameters are kept fixed. This provides some insight into the shape of the posterior distribution.

4. Though often there are many local minimizers, they can be separated into classes of minimizers sharing similar characteristics, such as close geometry. MCMC chains typically tend to stay within these classes, which can be observed by monitoring the closest local minimizer to an MCMC chain's state at each step. This suggests that the posterior can possess several clusters of nearby modes.

One conclusion we can draw from the above points is that there are often many different geometries that are consistent with the data. This is not necessarily an effect of noise on the measurements, and the effect may persist as the noise level goes to zero, since it is unknown if these geometric parameters are uniquely identifiable in general.

2.6.1 Test Models

We consider three different geometric models: a two parameter, two layer model; a five parameter, three layer model with fault; and a five parameter channelized model.

In what follows, as in Example 2.3.4, we define the negative Laplacian with Neumann boundary conditions:

$$A = -\Delta, \quad \mathcal{D}(A) = \left\{ u \in H^2(D) \;\middle|\; \frac{\mathrm{d}u}{\mathrm{d}\nu} = 0 \text{ on } \partial D, \int_D u(x)\,\mathrm{d}x = 0 \right\}.$$

Recall that if $u \sim N(0, A^{-\alpha})$ with $\alpha > d/2$, then u is almost surely continuous [39].

Additionally, if $W \subset \mathbb{R}^k$ is a subset of Euclidean space with positive and finite Lebesgue measure, $U(W)$ will denote the uniform probability distribution on W.

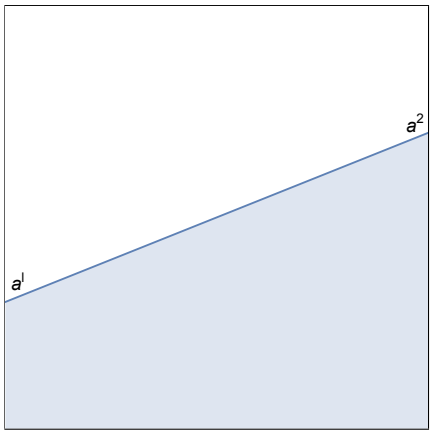


Figure 2.7: The definition of the geometric parameters $a = (a^1, a^2)$ in Model 1.

Model 1 (Two layer)

This model is described in Example 2.2.1. The geometric parameters $a = (a^1, a^2)$ are defined as in Figure 2.7. For simulations, we use the choice of prior

$$\mu_0 = N(1, A^{-1.4}) \times N(-1, A^{-1.8}),$$
$$\nu_0 = U([0,1]) \times U([0,1]).$$

Model 2 (Three layer with fault)

This model is described in [71], where it is labelled *Test Model 1.* The geometric parameters $a = (a^1, a^2, a^3, a^4, a^5)$ are defined as in Figure 2.8, with the fault occurring at $x = 0.55$. For simulations, we use the choice of prior

$$\mu_0 = N(2, 2A^{-1.4}) \times N(0, A^{-1.8}) \times N(-2, 2A^{-1.4}),$$
$$\nu_0 = U(S) \times U(S) \times U([-0.3, 0.3]),$$

where $S \subseteq [0,1]^2$ is the simplex $S = \{(x, y) \mid 0 \leq x \leq 1, x \leq y \leq 1\}$.

Model 3 (Channel)

This model is described in [71], where it is labelled *Test Model 2.* The geometric parameters $a = (a^1, a^2, a^3, a^4, a^5)$ are defined as in Figure 2.9. Here a^1, a^2, a^3, a^4, a^5

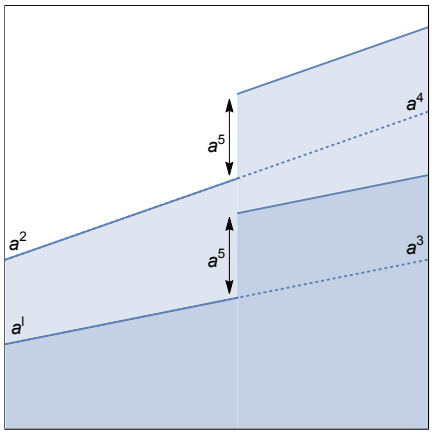


Figure 2.8: The definition of the geometric parameters $a = (a^1, a^2, a^3, a^4, a^5)$ in Model 2.

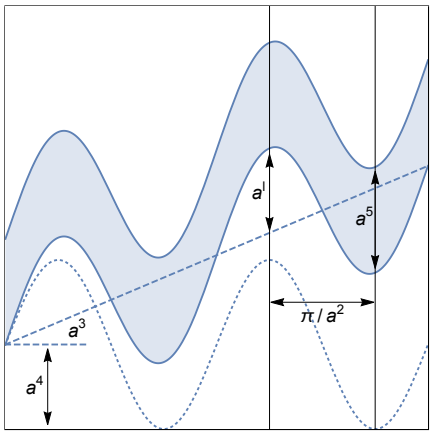


Figure 2.9: The definition of the geometric parameters $a = (a^1, a^2, a^3, a^4, a^5)$ in Model 3.

represent the channel amplitude, frequency, angle, initial point and width respectively. For simulations, we use the choice of prior

$$\mu_0 = N(1, A^{-1.4}) \times N(-1, A^{-1.8}),$$
$$\nu_0 = U([0,1]) \times U([\pi, 4\pi]) \times U([-\pi/4, \pi/4]) \times U([0,1]) \times U([0, 0.4]).$$

For each model, we fix a true permeability $(u^\dagger, a^\dagger)$ as a draw from the corresponding prior distribution, generated on a mesh of 256^2 points. For the forward model, we take the coefficient map $\sigma(\cdot) = \exp(\cdot)$. We observe the pressure on a grid $(x_i)_{i=1}^{25}$ of 25 uniformly spaced points, via the maps (2.2.4) with $\varepsilon = 0.05$. We add i.i.d. Gaussian noise $N(0, \gamma^2)$ to each observation, taking $\gamma = 0.01$. The resulting relative

errors on the data can be seen in Table 2.1. Small relative errors of this size typically make the posterior distribution hard to sample as they lead to measure concentration phenomena; MAP estimation can thus be particularly important.

Model Number	**Mean relative error (%)**	**Range of relative errors (%)**
1	0.5	$0.02 - 3.5$
2	0.9	$0.1 - 4.0$
3	0.3	$0.1 - 1.0$

Table 2.1: The relative error on the data, when each measurement is perturbed by an instance of $N(0, 0.01^2)$ noise.

2.6.2 MAP Estimation

Based on the theory in section 2.5, we can calculate MAP estimators by minimizing the Onsager-Machlup functional for the posterior distribution. We compute local minimizers of the Onsager-Machlup functional using the following iterative alternating method.

Algorithm 1 Iterative Alternating Method

1. Choose an initial state $(u_0, a_0) \in X \times \Lambda$.
2. Update the geometric parameters simultaneously using the Nelder-Mead algorithm.
3. Update each field individually using a line-search in the direction provided by the Gauss-Newton algorithm.
4. Go to 2.

The Nelder-Mead and Gauss-Newton algorithms are discussed in [108], in sections 9.5 and 10.3 respectively. Since we do not update the fields and geometric parameters simultaneously, it is possible that this algorithm will get caught in a saddle point: consider for example the function $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, $f(x, y) = xy$, at the point $(0, 0)$, being minimized alternately in the coordinate directions. Hence once the algorithm stalls, we propose a large number of random simultaneous updates in an attempt to find a lower functional value. If this is successful, we return to step 2 of the algorithm. We terminate the algorithm once the difference between successive values of Φ is below $\text{TOL} = 10^{-5}$. Calculations are performed on a mesh of 64^2 points in order to avoid an inverse crime.

To ensure that we explore the support of the posterior distribution, we choose a variety of initial states $(u_0, a_0) \in X \times \Lambda$ for the minimisation such that $I(u_0, a_0) < \infty$

in the continuum setting. To this end, we let a_0 be a draw from the prior distribution ν_0, and take u_0 to lie in the Cameron-Martin space of μ_0. Specifically, if a component of $u \in X$ has prior distribution $N(m, A^{-\alpha})$, we take the corresponding component of u_0 to be a draw from $N(m, A^{-\alpha-d/2})$. Output of the algorithm is shown in Figures 2.10-2.12.

We first comment on the minimizers of the Onsager-Machlup functional for Model 1. Generally the geometric parameters are closely recovered regardless of the initialisation state, though there is more variation in the fields. In the simulations where the geometry is inaccurate, for example simulations 7, 17 and 46, the fields can be seen to be compensating by forming a 'soft' interface where the true interface is.

The minimizers associated with Model 2 admit much more variation, though it is possible to partition them into smaller subsets of minimizers which share similar characteristics to one another, as mentioned in point (iv) at the beginning of the section. The clustering of the different minimizers is performed by eye, classifying them according to similar geometric parameters. Additionally we have an Other class, containing the minimizers which do not appear similar to one another nor appear to fit into any other class. We see later with MCMC simulations that these states do still act as local maximizers of the posterior probability.

The minimizers of the Onsager-Machlup functional for Model 3 show even more variation than those for Model 2, with the geometry in half of the minimizers not even being close to the true geometry. In the cases where the geometry is drastically wrong the fields have again attempted to compensate. This behaviour is particularly evident in the Other class, and is echoed in the MCMC simulations later. The Other class here is much larger than for Model 2, though as with Model 2 these states do appear to act as distinct local maximizers of the posterior probability.

This multi-modality of the posterior distribution is not unexpected. The paper [27] considers the history matching problem in reservoir simulation, in which inference is done jointly on both geometric and permeability parameters in the IC fault model. Though the forward map and observation maps are different in our model, we observe the same clustering of nearby local MAP estimators, and increased multi-modality as the dimension of the parameter space increases. In [27] it is observed that the global minimum often does not correspond to the truth, especially in the presence of measurement noise, and so all local minimizers of the Onsager-Machlup functional should be sought before drawing conclusions about the permeabilities – this appears to be the case in our model as well. We note that MCMC can be useful

in identifying a range of such minimizers, in view of the links established in the next subsection between MCMC and MAP estimation.

2.6.3 MCMC and Local Minimizers

We now observe the behaviour of MCMC chains initialized at these local minimizers of the Onsager-Machlup functional. We use a Metropolis-within-Gibbs algorithm for the sampling, alternating between preconditioned Crank-Nicolson (pCN) updates for the fields, see [34] for details, and Random Walk Metropolis updates for the geometric parameters. Again, simulations are performed on a mesh of 64^2 points in order to avoid an inverse crime. 10^5 samples are taken for each chain, with the initial 2×10^4 discarded as burn-in. The conditional means calculated from the samples are shown in Figures 2.13-2.15.

We monitor the value that Φ takes along the chain $(u^{(n)}, a^{(n)})$, and compare it with the value Φ takes on the local minimizers $(u^i_{\text{MAP}}, a^i_{\text{MAP}})$. This is shown in Figures 2.16-2.18, with the horizontal lines being the different values of $\Phi(u^i_{\text{MAP}}, a^i_{\text{MAP}})$. Note that it makes no sense to monitor the value that the objective functional I takes along the chain as the fields almost surely do not lie in the corresponding Cameron-Martin spaces, and so I is almost surely infinite along the chain in the continuum setting.

In addition, we monitor which minimizer the chain is nearest at each step, in the permeability space. Specifically, we look at

$$m_n := \operatorname*{argmin}_i \|F(u^{(n)}, a^{(n)}) - F(u^i_{\text{MAP}}, a^i_{\text{MAP}})\|_{L^2(D)} \tag{2.6.1}$$

where $F : X \times \Lambda \to L^\infty(D)$ is the construction map (2.2.1) from the state space to the permeability space. We make the choice of the L^2 norm over the L^∞ norm for the permeability space to avoid over-penalizing incorrect geometry. A selection of traces of m_n are shown in Figures 2.19-2.21. These illustrate that even though some of the local minimizers are very far from the true log-permeability, they do indeed act as local maximizers of the posterior probability.

We now discuss the above monitored quantities, and their relation to MAP estimators, on a model-by-model basis. Despite the slight variation in the fields of the minimizers from Model 1, the conditional means arising from the MCMC are nearly all identical. Simulation 23 stands out from the rest due to its slightly incorrect geometry – this effect can be seen in the trace plot of Φ, Figure 2.16, where the

value of Φ remains larger than the simulations started elsewhere. The traces of Φ for all other initialisations behave similarly to one another, taking similar misfit values after 2×10^4 samples. From Figure 2.19, it can be seen that the MCMC chains considered all spend a lot of time close to MAP estimator 38, despite this not being the estimator with the lowest functional value.

For Model 2, typically the conditional means within the different classes are very similar to one another. Classes A and C resemble each other, and Class B has compensated for incorrect geometry with the centre field. Faults have developed in Class D, though there is still some compensation in the field. The centre field and a small fault has appeared in Class E, but again the fields are compensating. The geometric parameters for the permeabilities in the Other class remain relatively unchanged, but the fields have more freedom to attain a lower misfit than in the Onsager-Machlup functional minimisation due to the lack of regularisation term. Figure 2.17 shows evidence for a number of local minima with a large data misfit value Φ, with some chains appearing to remain stuck in their vicinity. The four chains visible in Figure 2.17 (top) correspond to chains 49, 47, 45 and 43, from highest to lowest Φ value, all lying in the Other class – despite their significantly incorrect geometry, the corresponding MAP estimators appear to be genuine local maximizers of the posterior probability.

In the channelized model, Model 3, there is yet more variation between local minimizers. Here the compensation effect by the fields is even more apparent *in the conditional means*, especially in the Other class. From Figure 2.18 it appears that the local minima are much sharper and more sparsely distributed than the previous two models. Again the chains with the largest Φ values were initialized at minimizers in the Other class, suggesting the existence of many posterior modes with incorrect geometry.

The mixing of the MCMC chains varies heavily based on the initialisation points of the chains: with the same jump parameters for the field and geometric parameter proposals, acceptance rates vary largely based on which minimizer the chain was started from. This indicates that some of the minima are much sharper than others. This is also evident from the traces of m_n defined above, Figures 2.19-2.21, especially in Model 3. Note also from these figures that the nearest local minimum typically lies in the same class as the initialisation state, though jumps between classes are possible. Though not shown, in Model 2, whenever the initial state lies in Class A, then the nearest minimizer always lies in Class A.

2.7 Conclusions and Future Work

We have made a new contribution to the recently developed theory of MAP estimation in infinite dimensions [38, 65]. We link MAP estimation to a variational Onsager-Machlup functional. The work is focused on priors for piecewise Gaussian random fields, with random interfaces parameterized finite-dimensionally. Such fields arise naturally in applications such as groundwater flow and EIT, and these are used to illustrate the theory and numerics. The work opens up several new avenues for investigation. A major theoretical direction is to fully reconcile the approaches in [38] and [65]. That is, show equivalence of weak MAP estimators and minimizers of a functional related to the Fomin derivative of the posterior, under the more general assumptions of a nonlinear forward map and a certain class of non-convex priors. The work in this chapter suggests that this may be possible. On the applications side an important new direction would be to consider problems in which the geometric parameters are no longer independent from the fields a priori. A possible extension could be to treat the geometric parameters as hyperparameters for the fields under the prior. This would allow, for example, the fields to have specific boundary conditions at the interfaces, which may be more physically appropriate in some contexts. A related hierarchical model was considered in [107], in which prior samples were piecewise white; this could be extended to allow for spatial correlations in the continuum setting. Computationally an exciting direction is to build upon definitions of MAP estimators to develop hybrid algorithms which fully exploit local minimizer structure of the Onsager-Machlup functional within MCMC.

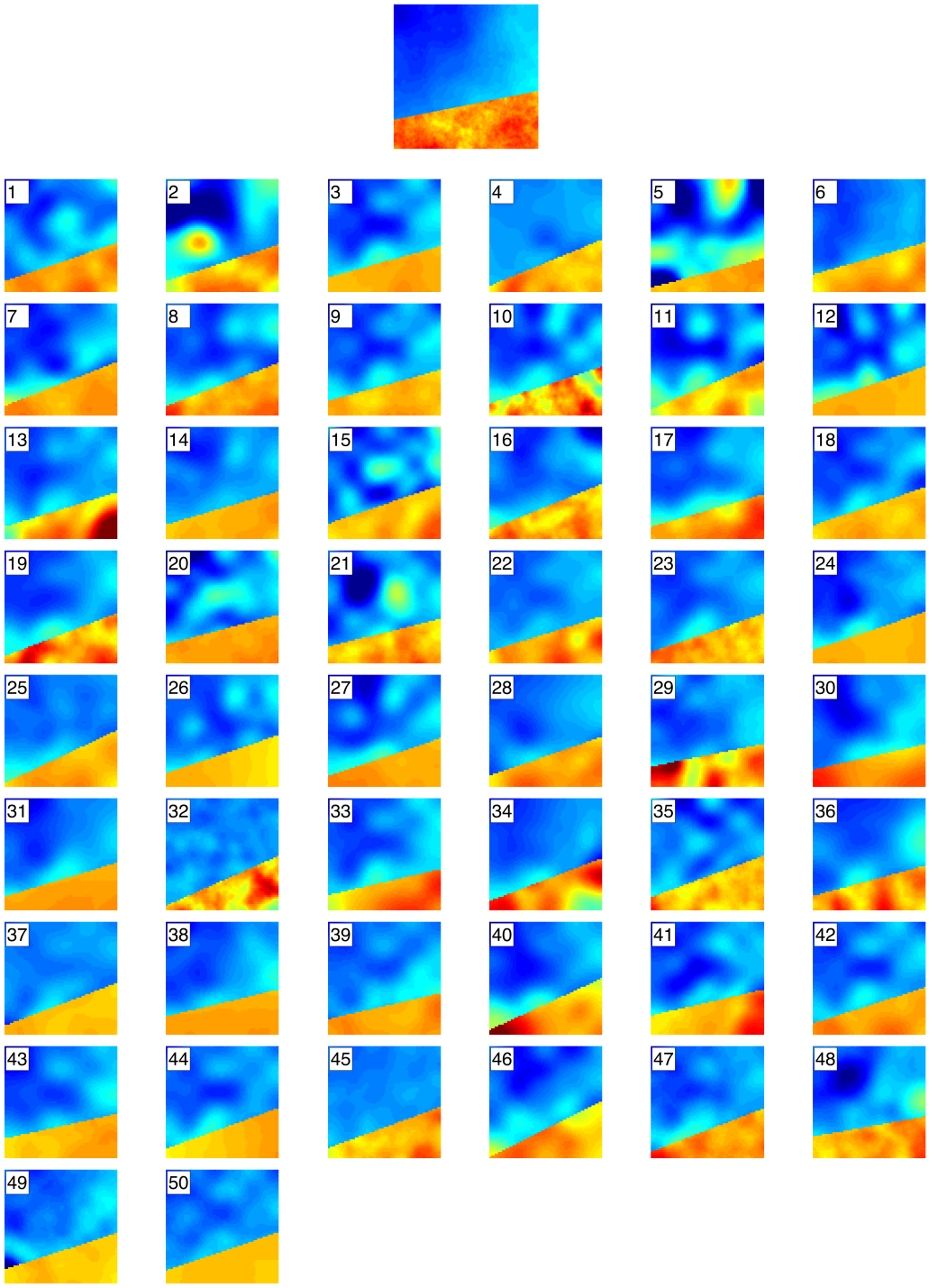


Figure 2.10: (Model 1) the true log-permeability field (top), and 50 local minimizers arising from minimisation initialized at draws from a smoothed prior distribution. Simulation 12 has the lowest functional value, with $I(u^{12}_{\mathrm{MAP}}, a^{12}_{\mathrm{MAP}}) = 2847$.

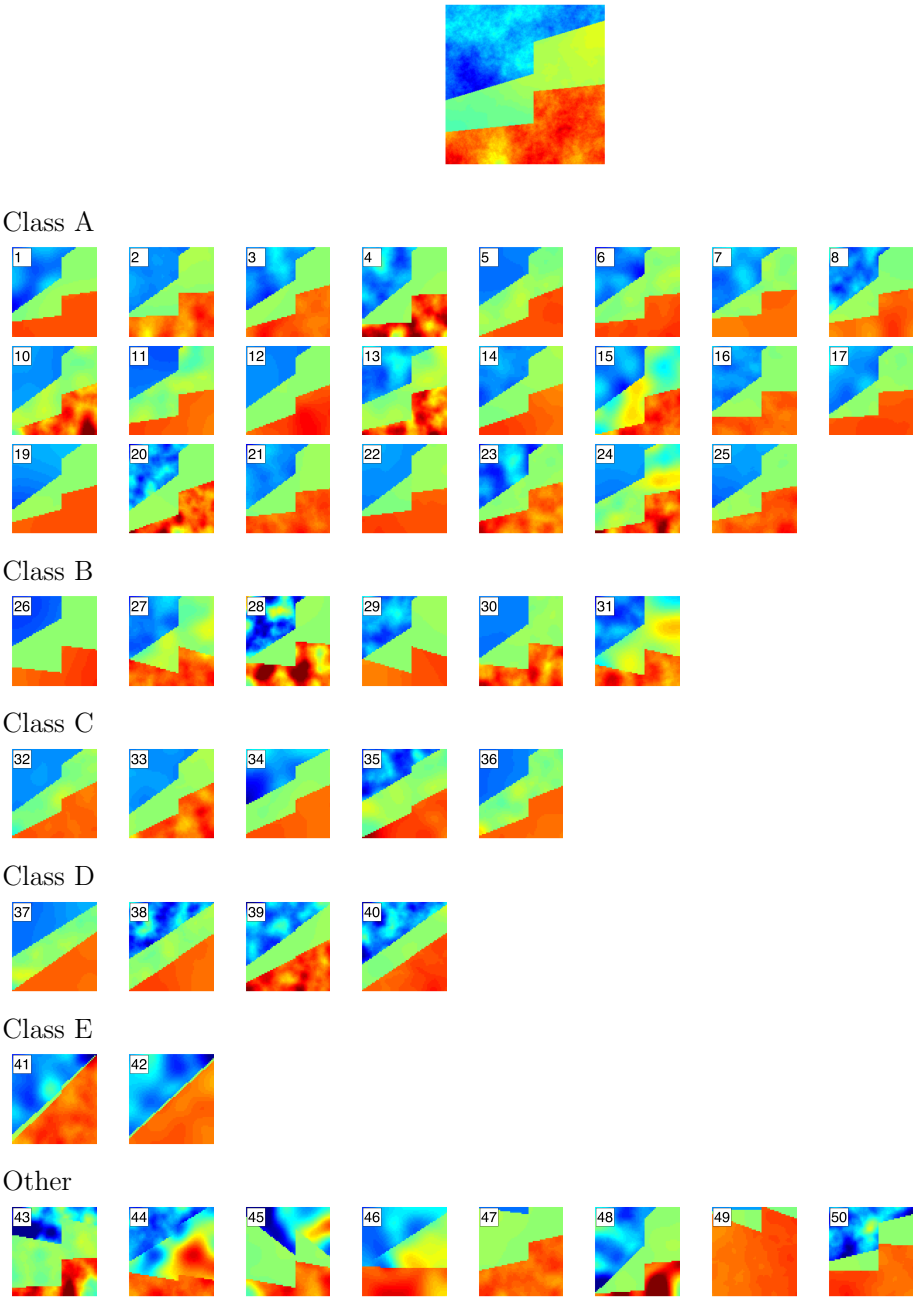


Figure 2.11: (Model 2) the true log-permeability field (top), and 50 local minimizers arising from minimisation initialized at draws from a smoothed prior distribution. Simulation 7 has the lowest functional value, with $I(u^7_{\text{MAP}}, a^7_{\text{MAP}}) = 2567$. The minimizers have been divided into classes based on similar characteristics.

Figure 2.12: (Model 3) the true log-permeability field (top), and 50 local minimizers arising from minimisation initialized at draws from a smoothed prior distribution. Simulation 20 has the lowest functional value, with $I(u^{20}_{\text{MAP}}, a^{20}_{\text{MAP}}) = 2117$. The minimizers have been divided into classes based on similar characteristics.

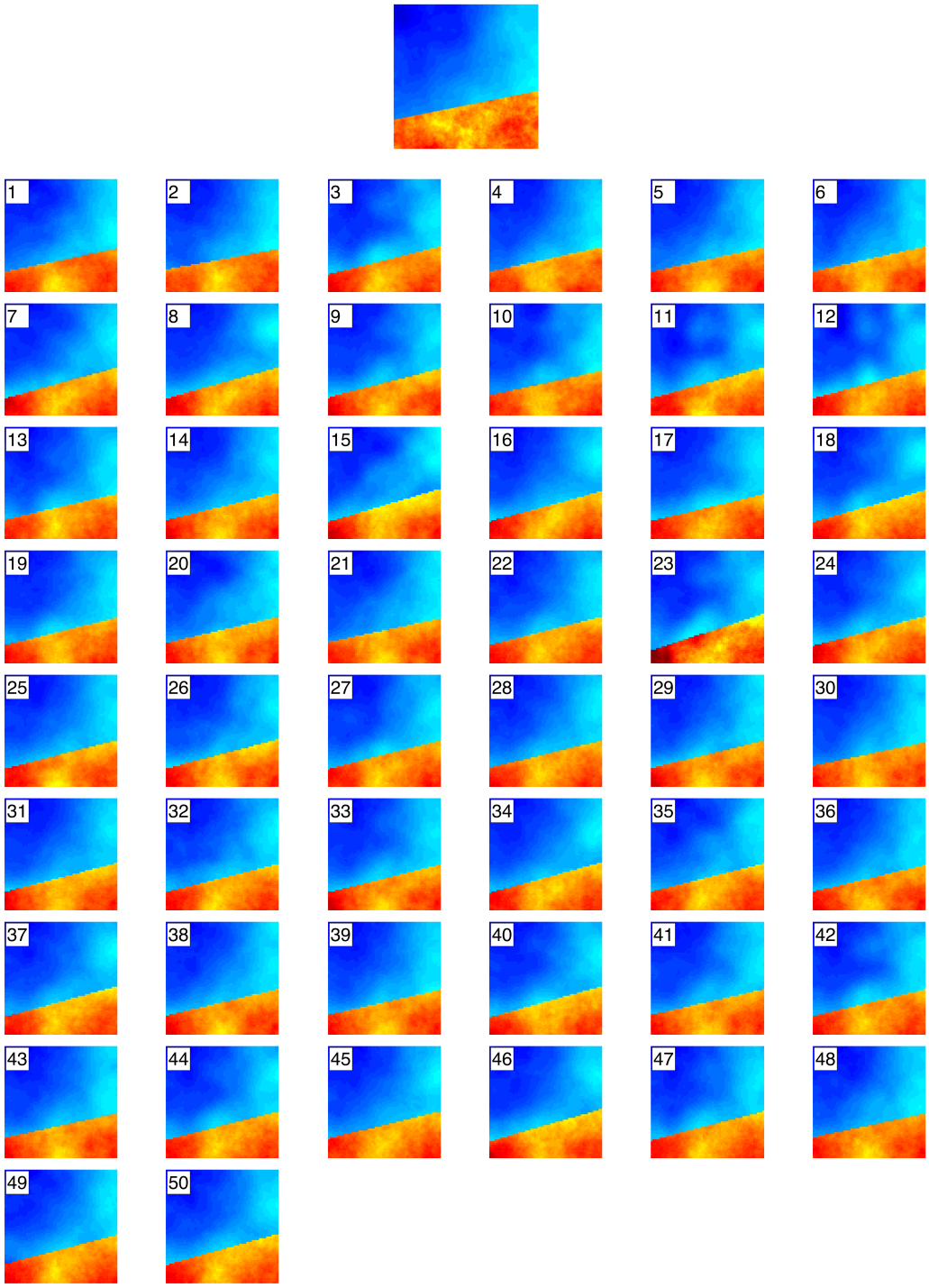


Figure 2.13: (Model 1) the true log-permeability field (top), and the conditional mean arising from MCMC chains initialized at the corresponding local minimizers above.

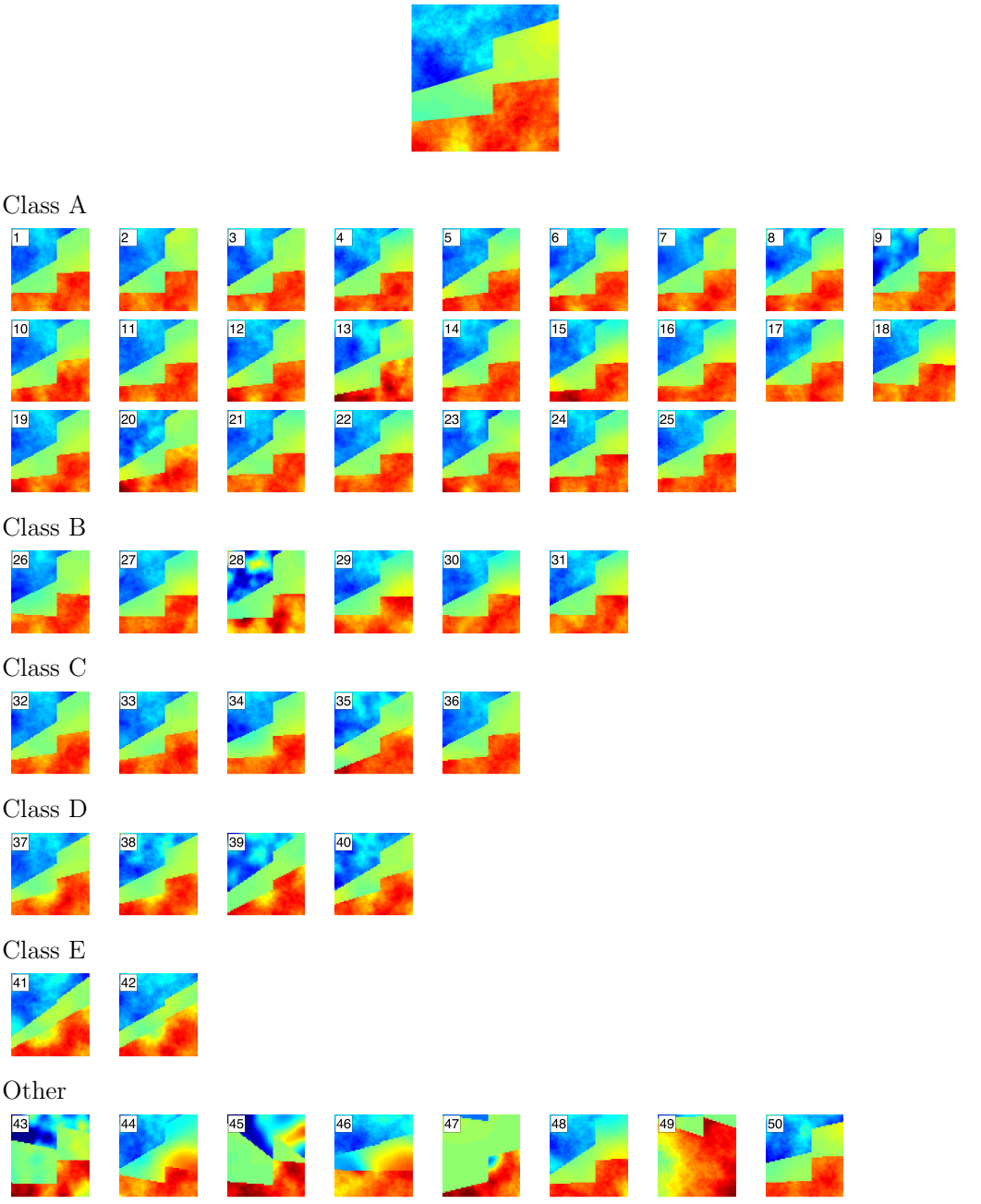


Figure 2.14: (Model 2) the true log-permeability field (top), and the conditional mean arising from MCMC chains initialized at the corresponding local minimizers above. We group them into the same classes as the local minimizers.

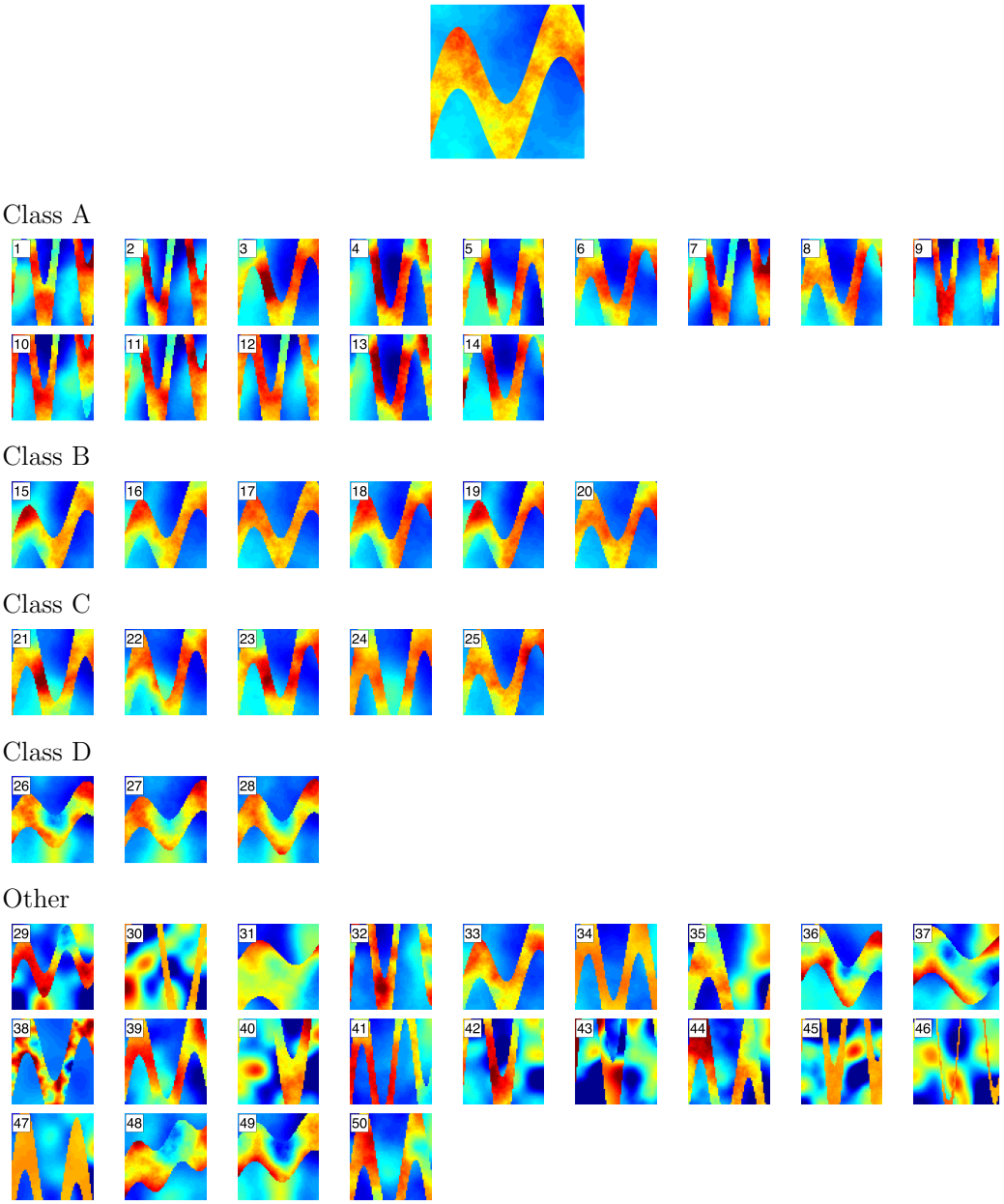


Figure 2.15: (Model 3) the true log-permeability field (top), and the conditional mean arising from MCMC chains initialized at the corresponding local minimizers above. We group them into the same classes as the local minimizers.

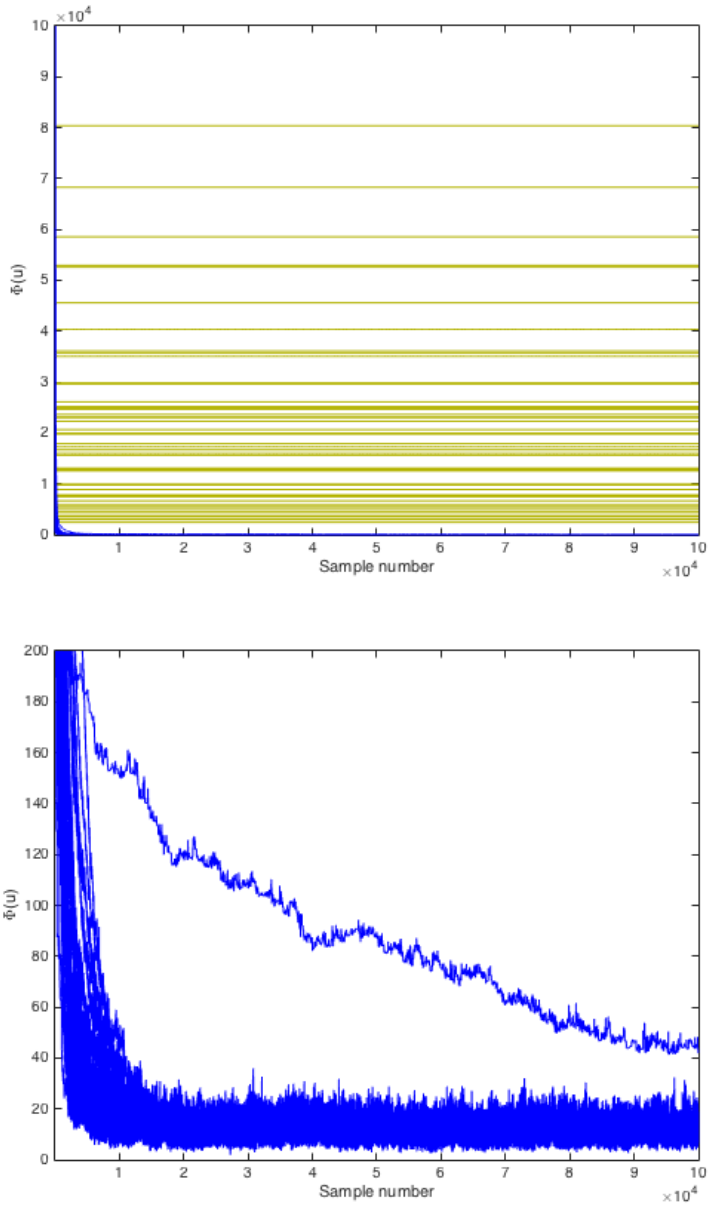


Figure 2.16: (Model 1) The evolution of Φ as the MCMC chains progress. The horizontal lines represent the value of each local minimizer under Φ. The same traces are shown in each figure with different ranges on the vertical axes. Nearly all of the simulations find a small value of Φ almost immediately, but simulation 5 remains caught in the local minimizer for some time before it follows.

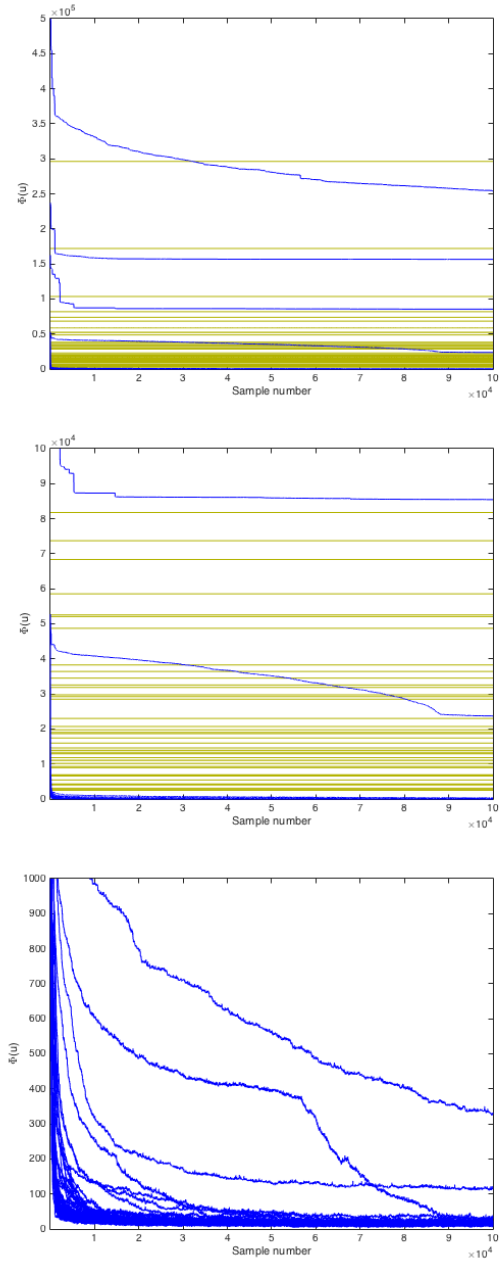


Figure 2.17: (Model 2) The evolution of Φ as the MCMC chains progress. The horizontal lines represent the value of each local minimizer under Φ. The same traces are shown in each figure with different ranges on the vertical axes. The majority of the simultions find a small value of Φ almost immediately, but numerous fail to reach there, settling in local minima. The shape of these minima can be seen in Figure 2.14, and generally correspond to those in the same class as the initial state.

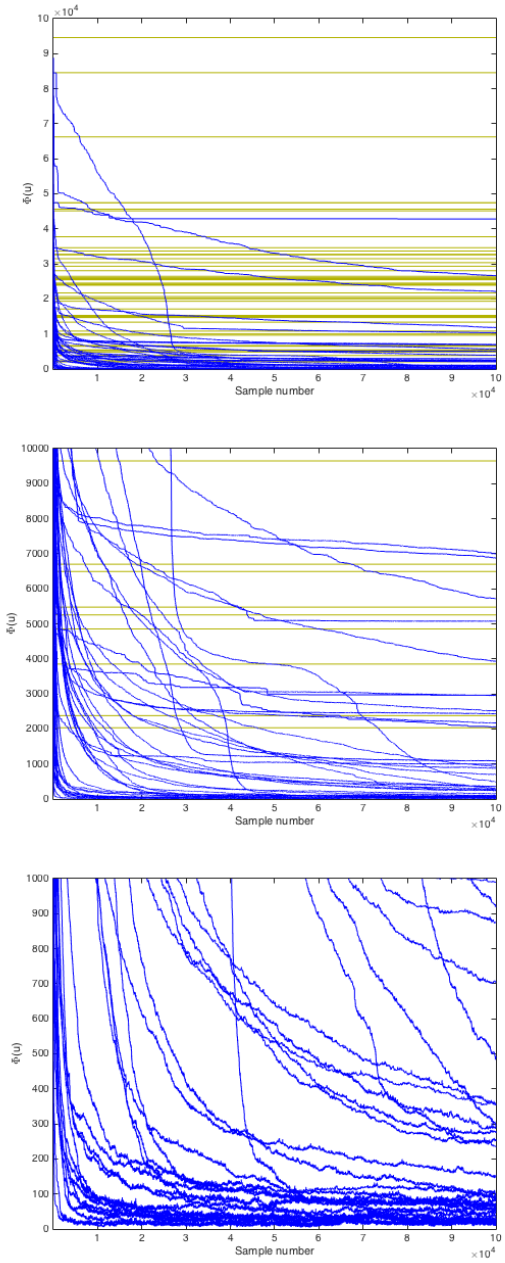


Figure 2.18: (Model 3) The evolution of Φ as the MCMC chains progress. The horizontal lines represent the value of each local minimizer under Φ. The same traces are shown in each figure with different ranges on the vertical axes. The majority of the simulations find a small value of Φ almost immediately, but numerous fail to reach there, settling in local minima. The shape of these minima can be seen in Figure 2.15, and generally correspond to those in the same class as the initial state.

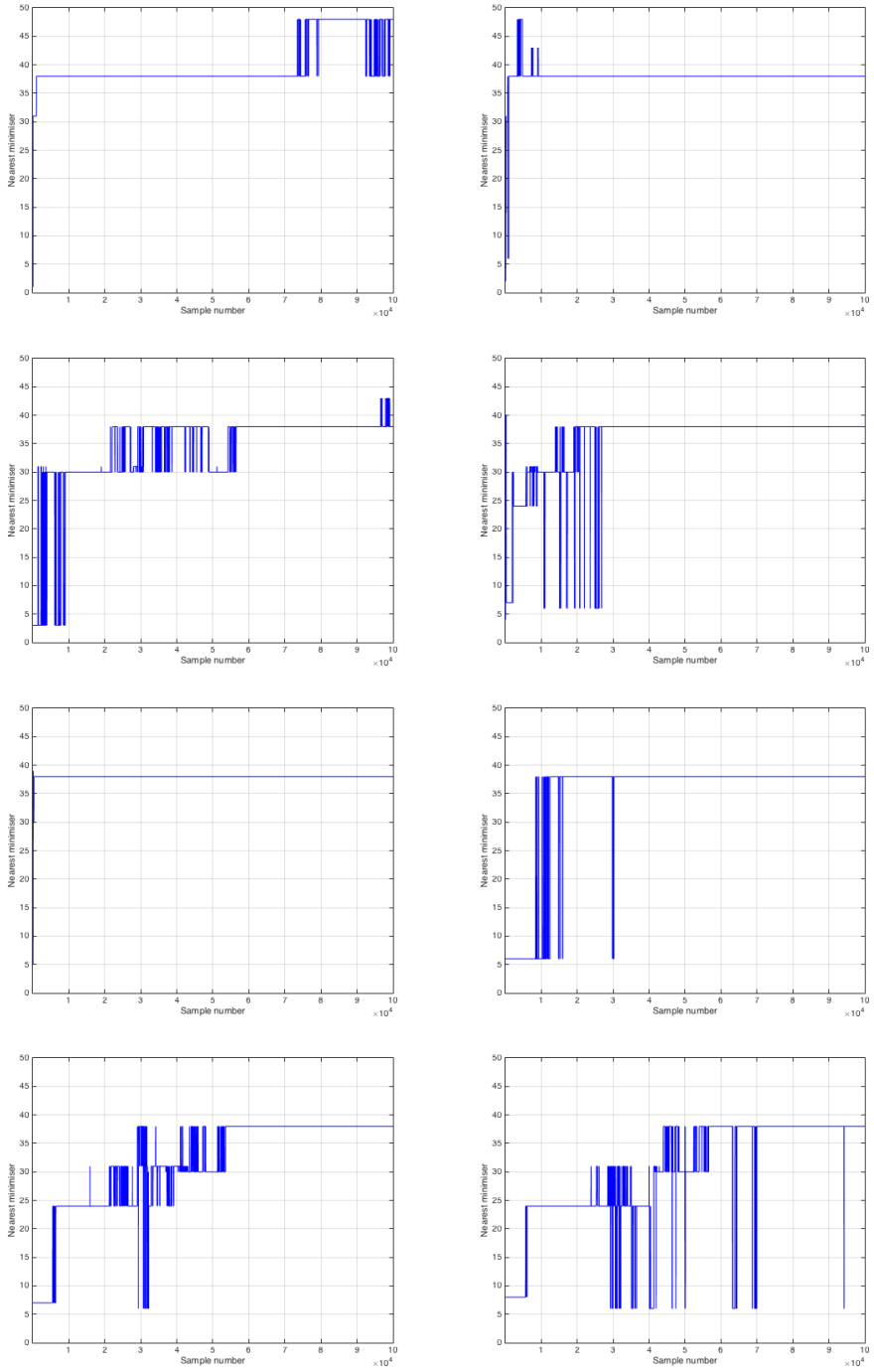


Figure 2.19: (Model 1) The trace of m_n as defined by (2.6.1), when the chain is initialized at a variety of minimizers – specifically numbers $1, 2, \ldots, 8$.

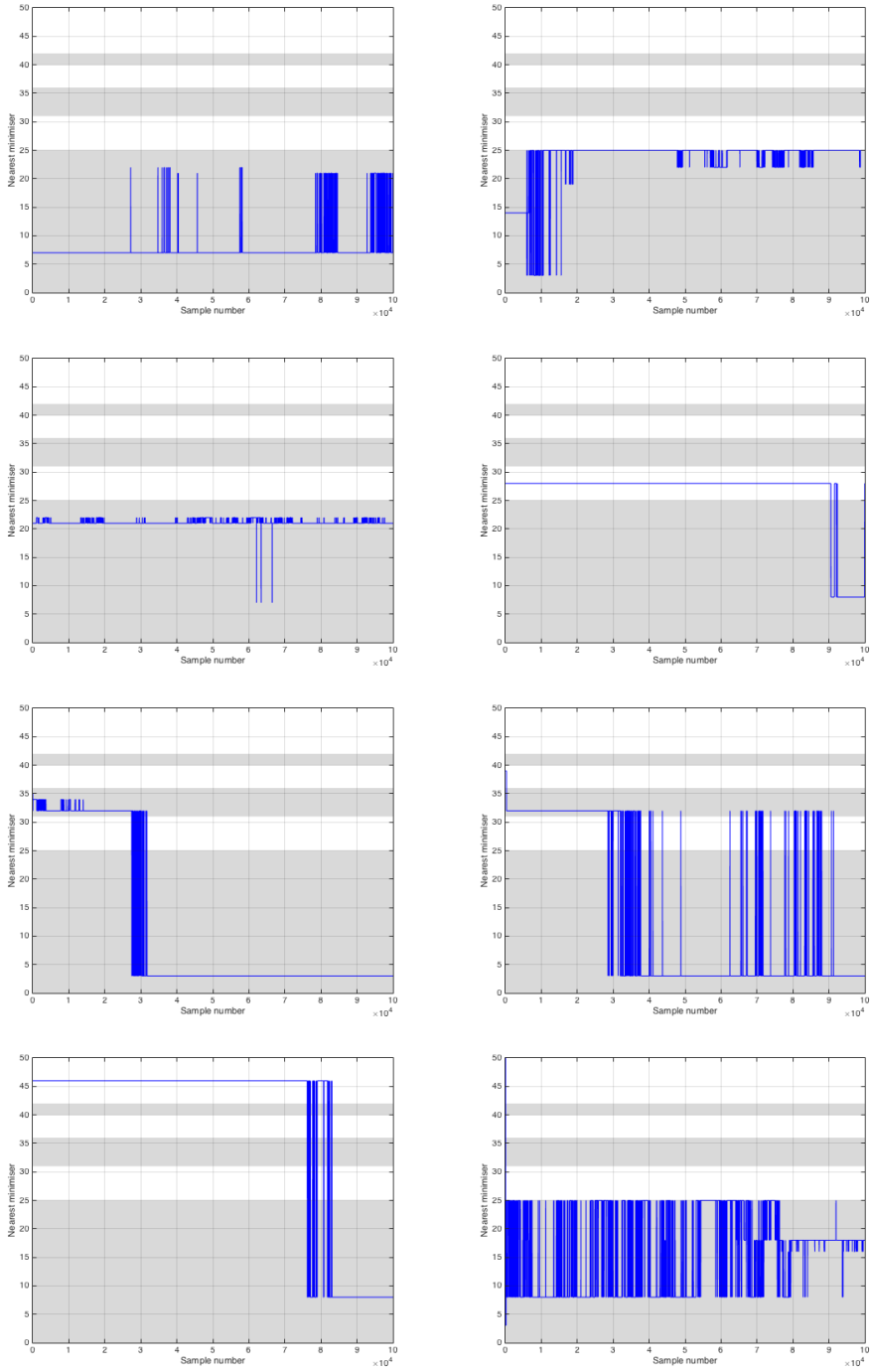


Figure 2.20: (Model 2) The trace of m_n as defined by (2.6.1), when the chain is initialized at a variety of minimizers – specifically numbers $7, 14, 21, 28, 35, 39, 46$ and 50. The different classes are alternately shaded.

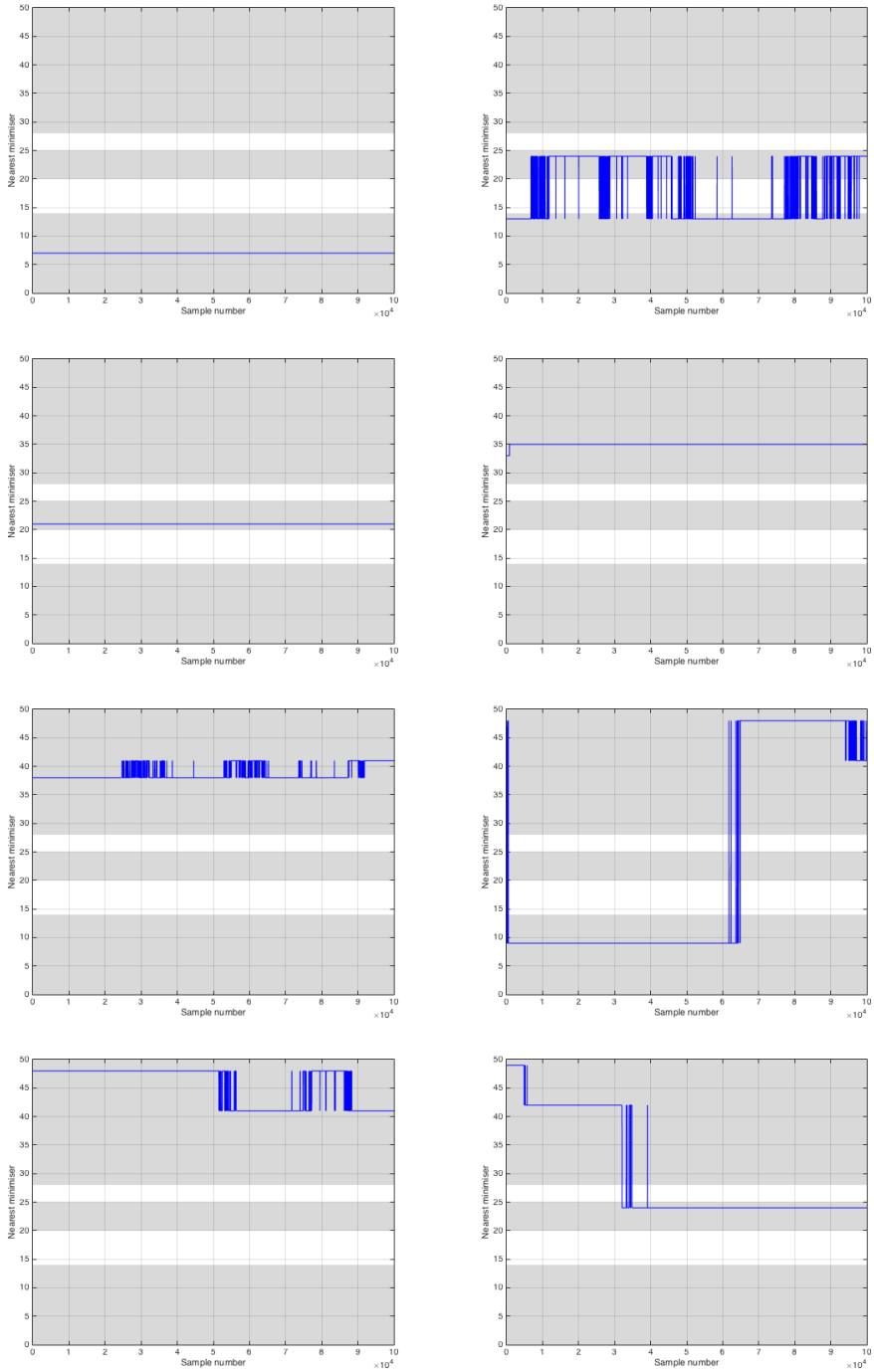


Figure 2.21: (Model 3) The trace of m_n as defined by (2.6.1), when the chain is initialized at a variety of minimizers – specifically numbers $7, 13, 21, 33, 38, 47, 48$ and 49. The different classes are alternately shaded.

2.8 Appendix

In this appendix we provide proofs of the results given in the chapter.

2.8.1 Results From Section 2.2

Before we prove Lemma 2.2.5 we require the following lemma.

Lemma 2.8.1. *Let $(X, \mathcal{F}, \mu)$ be a measure space and $f \in L^1(X, \mathcal{F}, \mu)$. Let $B_n \subseteq \mathcal{F}$ be a sequence of measurable subsets of X with $\mu(B_n) \to 0$ as $n \to \infty$. Then*

$$\int_{B_n} f(x)\, \mu(\mathrm{d}x) \to 0 \quad \text{as } n \to \infty.$$

Proof. Write $f_n(x) = f(x)\mathbb{1}_{B_n}(x)$. We have that $f_n \to 0$ in measure: for any $\delta > 0$,

$$\mu(\{x \in X \mid |f_n(x)| > \delta\}) \leq \mu(\{x \in X \mid |f_n(x)| \neq 0\}) \leq \mu(B_n) \to 0.$$

Now suppose that $\|f_n\|_{L^1}$ does not tend to zero. Then there exists $\delta > 0$ and a subsequence $(f_{n_k})_{k \geq 1}$ such that $\|f_{n_k}\|_{L^1} \geq \delta$ for all $k \geq 1$. This subsequence still converges to zero in measure, and so admits a further subsequence that converges to zero almost surely. We can bound this subsequence above uniformly by f, and so an application of the dominated convergence theorem leads to a contradiction. The result follows. □

Proof of Lemma 2.2.5. Showing that $\mathcal{R}$ is well-defined is equivalent to showing that PDE (2.2.3) has a unique solution for all $(u, a) \in X \times \Lambda$. Since $u^a \in L^\infty(D)$ it is bounded, and so by the continuity and positivity of σ there exist $\kappa_{\min}, \kappa_{\max} > 0$ with $\kappa_{\min} \leq \sigma(u^a) \leq \kappa_{\max}$. The associated bilinear form is hence bounded and coercive. The right hand side can be seen to lie in $H^{-1}(D)$ since $G \in H^1(D)$ and $\sigma(u^a) \leq \kappa_{\max}$, and so a unique solution exists by Lax-Milgram.

(i) In its weak form, the PDE (2.2.3) is given by

$$\int_D \sigma(u^a) \nabla q_{u,a} \cdot \nabla \varphi = f(\varphi) - \int_D \sigma \nabla G \cdot \nabla \varphi \quad \text{for all } \varphi \in V.$$

Taking $\varphi = q_{u,a}$ we deduce that

$$\begin{aligned}\kappa_{\min}(u,a)\|\nabla q_{u,a}\|_{L^2}^2 &\leq \int_D \sigma(u^a)\nabla q_{u,a}\cdot\nabla q_{u,a}\\ &= f(q_{u,a}) - \int_D \sigma(u^a)\nabla G\cdot\nabla q_{u,a}\\ &\leq \|f\|_{V^*}\|q_{u,a}\|_V + \|\sigma(u^a)\|_{L^\infty}\|\nabla G\|_{L^2}\|\nabla q_{u,a}\|_{L^2}\end{aligned}$$

and so we have the estimate

$$\begin{aligned}\|p_{u,a}\|_V &\leq \|q_{u,a}\|_V + \|G\|_V\\ &\leq (\|f\|_{V^*} + \|\sigma(u^a)\|_{L^\infty}\|G\|_V)/\kappa_{\min}(u,a) + \|G\|_V.\end{aligned}$$

(ii) Let $u, v \in X$ and $a \in \Lambda$. Then $p_{u,a} - p_{v,a}$ satisfies the PDE

$$\begin{cases} -\nabla\cdot\big(\sigma(u^a)\nabla(p_{u,a}-p_{v,a})\big) &= \nabla\cdot\big((\sigma(u^a)-\sigma(v^a))\nabla p_{v,a}\big) \quad \text{in } D\\ p_{u,a}-p_{v,a} &= 0 \qquad\qquad\qquad\qquad\qquad\qquad\quad \text{on } \partial D.\end{cases}$$

Setting $\kappa_*(u,v,a) = \kappa_{\min}(u,a)\wedge\kappa_{\min}(v,a)$, we see

$$\begin{aligned}\kappa_*(u,v,a)\|\nabla(p_{u,a}-p_{v,a})\|_{L^2}^2 &\leq \int_D \sigma(u^a)|\nabla(p_{u,a}-p_{v,a})|^2\\ &= \int_D (\sigma(u^a)-\sigma(v^a))\nabla(p_{u,a}-p_{v,a})\cdot\nabla p_{v,a}\\ &\leq \|\sigma(u^a)-\sigma(v^a)\|_{L^\infty}\|\nabla(p_{u,a}-p_{v,a})\|_{L^2}\|\nabla p_{v,a}\|_{L^2}\end{aligned}$$

and so by (i),

$$\begin{aligned}\|p_{u,a}-p_{v,a}\|_V &\leq \|p_{v,a}\|_V\|\sigma(u^a)-\sigma(v^a)\|_{L^\infty}/\kappa_*(u,v,a)\\ &\leq \|\sigma(u^a)-\sigma(v^a)\|_{L^\infty}\\ &\qquad\times\big((\|f\|_{V^*}+\|\sigma(u^a)\|_{L^\infty}\|G\|_V)/\kappa_*(u,a)^2 + \|G\|_V/\kappa_*(u,a)\big).\end{aligned}$$

Using that the A_i are disjoint gives that

$$\begin{aligned}\|\sigma(u^a)-\sigma(v^a)\|_{L^\infty} &= \left\|\sigma\left(\sum_{i=1}^N u_i\mathbb{1}_{A_i(a)}\right) - \sigma\left(\sum_{i=1}^N v_i\mathbb{1}_{A_i(a)}\right)\right\|_{L^\infty}\\ &= \|\sigma(u_k)-\sigma(v_k)\|_{L^\infty}\end{aligned}$$

for some $k = k(a)$. Now suppose that $\|u\|_X, \|v\|_X < r$. Then the C^1 property

of σ yields

$$\|\sigma(u_k) - \sigma(v_k)\|_{L^\infty} \leq \max_{|t|\leq r} |\sigma'(t)| \cdot \|u_k - v_k\|_{L^\infty} \leq \max_{|t|\leq r} |\sigma'(t)| \cdot \|u - v\|_X.$$

Finally we deal with the κ_*^{-j} terms:

$$\begin{aligned}
\kappa_*(u, v, a)^{-j} &= \left[\left(\operatorname*{essinf}_{x\in D} e^{u^a(x)} \right) \wedge \left(\operatorname*{essinf}_{x\in D} e^{v^a(x)} \right) \right]^{-j} \\
&\leq \left(\min_{|t|\leq r} \sigma(t) \wedge \min_{|t|\leq r} \sigma(t) \right)^{-j} \\
&= \left(\min_{|t|\leq r} \sigma(t) \right)^{-j}.
\end{aligned}$$

We bound the $\|\sigma(u^a)\|_{L^\infty}$ term similarly. Putting the above bounds together, we have

$$\begin{aligned}
\|\mathcal{R}(u, a) - \mathcal{R}(v, a)\|_V &= \|p_{u,a} - p_{v,a}\|_V \\
&\leq \max_{j=1,2} \left(\min_{|t|\leq r} \sigma(t) \right)^{-j} \left(\|f\|_{V^*} + \|G\|_V \left(\max_{|t|\leq r} \sigma(t) + 1 \right) \right) \\
&\qquad \times \max_{|t|\leq r} |\sigma'(t)| \cdot \|u - v\|_X \\
&= L(r)\|u - v\|_X.
\end{aligned}$$

Note that the constant $L(r)$ is uniform in a.

(iii) We use a similar approach to the previous part. Given $u \in X$ and $a, b \in \Lambda$, the difference $p_{u,a} - p_{u,b}$ satisfies

$$\begin{cases}
-\nabla \cdot (\sigma(u^a)\nabla(p_{u,a} - p_{u,b})) &= \nabla \cdot ((\sigma(u^a) - \sigma(u^b))\nabla p_{u,b}) \quad \text{in } D \\
p_{u,a} - p_{u,b} &= 0 \qquad\qquad\qquad\qquad\qquad\qquad\quad \text{on } \partial D
\end{cases}$$

which leads to the bound

$$\begin{aligned}
\kappa_\dagger(u, a, b)\|\nabla(p_{u,a} - p_{u,b})\|_{L^2}^2 &\leq \int_D \sigma(u^a)|\nabla(p_{u,a} - p_{u,b})|^2 \\
&= \int_D (\sigma(u^a) - \sigma(u^b))\nabla(p_{u,a} - p_{u,b}) \cdot \nabla p_{u,b} \\
&\leq \|\nabla(p_{u,a} - p_{u,b})\|_{L^2} \|(\sigma(u^a) - \sigma(u^b))\nabla p_{u,b}\|_{L^2}
\end{aligned}$$

where $\kappa_\dagger(u,a,b) = \kappa_{\min}(u,a) \wedge \kappa_{\min}(u,b)$. It follows that

$$\|p_{u,a} - p_{u,b}\|_V \leq \|(\sigma(u^a) - \sigma(u^b))\nabla p_{u,b}\|_{L^2} / \kappa_\dagger(u,a,b).$$

Again by the disjointness of the A_i and the C^1 property of σ,

$$\begin{aligned}
\|(\sigma(u^a) - \sigma(u^b))\nabla p_{u,b}\|_{L^2} &= \left\| \left(\sigma\left(\sum_{i=1}^N u_i \mathbb{1}_{A_i(a)} \right) - \sigma\left(\sum_{i=1}^N u_i \mathbb{1}_{A_i(b)} \right) \right) \nabla p_{u,b} \right\|_{L^2} \\
&= \left\| \sum_{i=1}^N \left(\sigma\left(u_i \mathbb{1}_{A_i(a)} \right) - \sigma\left(u_i \mathbb{1}_{A_i(b)} \right) \right) \nabla p_{u,b} \right\|_{L^2} \\
&\leq \sum_{i=1}^N \left\| \left(\sigma\left(u_i \mathbb{1}_{A_i(a)} \right) - \sigma\left(u_i \mathbb{1}_{A_i(b)} \right) \right) \nabla p_{u,b} \right\|_{L^2} \\
&\leq \sum_{i=1}^N \max_{|t| \leq \|u_i\|_\infty} |\sigma'(t)| \cdot \left\| \left| u_i \mathbb{1}_{A_i(a)} - u_i \mathbb{1}_{A_i(b)} \right| \nabla p_{u,b} \right\|_{L^2} \\
&\leq \sum_{i=1}^N \max_{|t| \leq \|u_i\|_\infty} |\sigma'(t)| \cdot \|u_i\|_\infty \left\| \mathbb{1}_{A_i(a) \Delta A_i(b)} \nabla p_{u,b} \right\|_{L^2}
\end{aligned}$$

since $|\mathbb{1}_A - \mathbb{1}_B| = \mathbb{1}_{A \Delta B}$. Now as before we can bound $\kappa_\dagger^{-1}$:

$$\begin{aligned}
\kappa_\dagger(u,v,a)^{-1} &= \left[\left(\operatorname*{essinf}_{x \in D} e^{u^a(x)} \right) \wedge \left(\operatorname*{essinf}_{x \in D} e^{u^b(x)} \right) \right]^{-1} \\
&\leq \left(\min_{|t| \leq \max \|u_i\|_\infty} \sigma(t) \wedge \min_{|t| \leq \max \|u_i\|_\infty} \sigma(t) \right)^{-1} \\
&\leq \left(\min_{|t| \leq \|u\|_X} \sigma(t) \right)^{-1}.
\end{aligned}$$

Putting the above bounds together, we have

$$\begin{aligned}
&\|\mathcal{R}(u,a) - \mathcal{R}(u,b)\|_V = \|p_{u,a} - p_{u,b}\|_V \\
&\quad \leq \left(\min_{|t| \leq \|u\|_X} \sigma(t) \right)^{-1} \sum_{i=1}^N \max_{|t| \leq \|u_i\|_\infty} |\sigma'(t)| \cdot \|u_i\|_\infty \left\| \mathbb{1}_{A_i(a) \Delta A_i(b)} \nabla p_{u,b} \right\|_{L^2} \\
&\quad \leq \left(\min_{|t| \leq \|u\|_X} \sigma(t) \right)^{-1} \sum_{i=1}^N \max_{|t| \leq \|u_i\|_\infty} |\sigma'(t)| \cdot \|u_i\|_\infty \left(\int_{A_i(a) \Delta A_i(b)} |\nabla p_{u,b}|^2 \right)^{1/2}.
\end{aligned}$$

The right hand goes to zero as each $|A_i(a) \Delta A_i(b)| \to 0$ by Lemma 2.8.1, since $|\nabla p_{u,b}| \in L^2(D)$, and so the continuity of $\mathcal{R}(u,\cdot)$ follows from the assumed

continuity of the maps A_i.

□

Proof of Proposition 2.2.7. 1. Theorem 2.3 in [73] tells us that the mapping from the conductivity to the weak solution of (2.2.5) is Fréchet differentiable with respect to the supremum norm, and hence locally Lipschitz. Note that the mapping from the solution to the boundary voltage measurements, $(v, V) \mapsto V$, is smooth, and the assumptions on σ imply that it is Lipschitz. It hence suffices to show that the mapping $u \mapsto F(u, a)$ is Lipschitz for each $a \in \Lambda$. Let $u, v \in X$ and $a \in \Lambda$, then

$$\|F(u,a) - F(v,a)\|_\infty \leq \sum_{i=1}^{N} \|u_i - v_i\|_\infty \mathbb{1}_{A_i(a)} \leq C\|u - v\|_X$$

and the result follows.

2. By Corollary 2.8 in [45] and the continuity of σ, it suffices to show that $a_n \to a$ in Λ implies that $F(u, a_n) \to F(u, a)$ in measure. For any $p \in (1, \infty)$ we have that

$$\begin{aligned}\int_D |F(u,a_n) - F(u,a)|^p \,\mathrm{d}x &\leq \sum_{i=1}^{N} \int_D |u_i|^p \mathbb{1}_{A_i(a_n)\Delta A_i(a)} \,\mathrm{d}x \\ &\leq \sum_{i=1}^{N} \|u_i\|_\infty^p \cdot |A_i(a_n)\Delta A_i(a)|\end{aligned}$$

From the assumed continuity of $A_i(\cdot)$ it follows that $F(u, a_n) \to F(u, a)$ in L^p for any $p \in (1, \infty)$, and hence in measure.

□

2.8.2 Results From Section 2.4

Proof of Theorem 2.4.2. (i) We first claim that the assumptions on Φ mean that $\Phi(\cdot, \cdot; y) : X' \times \Lambda' \to \mathbb{R}$ is continuous for each $y \in Y$. Fix $y \in Y$ and $(u, a) \in X' \times \Lambda'$. Choose any approximating sequence $(u_n, a_n)_{n\geq 1} \subseteq X' \times \Lambda'$ such that $(u_n, a_n) \to (u, a)$. Then the assumptions on the norm on $X \times \Lambda$ means that $\|u_n - u\|_X \to 0$ and $|a_n - a| \to 0$. Letting $r > \max\{\|u\|_X, \sup_n \|u_n\|_X\}$, we

may approximate

$$\begin{aligned}|\Phi(u_n, a_n; y) - \Phi(u, a; y)| &\leq |\Phi(u_n, a_n; y) - \Phi(u, a_n; y)| + |\Phi(u, a_n; y) - \Phi(u, a; y)| \\ &\leq M_3(r, a_n, y)\|u_n - u\|_X + |\Phi(u, a_n; y) - \Phi(u, a; y)| \\ &\leq \left(\sup_{k\in\mathbb{N}} M_3(r, a_k, y)\right) \cdot \|u_n - u\|_X + |\Phi(u, a_n; y) - \Phi(u, a; y)|\end{aligned}$$

where the supremum is finite due the continuity of M_3 in its second component. Since Φ is also continuous in its second component, we see that the right-hand side tends to zero as $(u_n, a_n) \to (u, a)$.

Now as $\Phi(\cdot, \cdot; y) : X' \times \Lambda' \to \mathbb{R}$ is continuous and $(\mu_0 \times \nu_0)(X' \times \Lambda') = 1$, $\Phi(\cdot, \cdot; y)$ is $\mu_0 \times \nu_0$-measurable. Setting $Z = X' \times \Lambda'$, we can consider $\Phi : Z \times Y \to \mathbb{R}$. This is a Caratheodory function, and it is known that these are jointly measurable, see for example [7]. We conclude that Φ is $\mu_0 \times \nu_0 \times \mathbb{Q}_0$ measurable.

(ii) We first show $Z(y)$ is finite. Since μ_0 is Gaussian, by Fernique's theorem there exists $\alpha > 0$ such that

$$\int_X \exp(\alpha\|u\|_X^2)\, \mu_0(\mathrm{d}u) < \infty.$$

Then using Assumptions 2.4.1(i), we have the lower bound

$$\Phi(u, a; y) \geq M_1(\alpha) - \alpha\|u\|_X^2$$

from which we conclude that $Z(y) < \infty$.

Now fix $r > 0$. Let $y \in Y$ and take $(u, a) \in X' \times \Lambda'$ with $\max\{\|u\|_X, |a|\} < r$. Then we have by the local Lipschitz property

$$|\Phi(u, a; y)| \leq M_3(r, y)\|u\|_X + |\Phi(0, a; y)| \leq M_3(r, a, y)r + |\Phi(0, a; y)|.$$

Using the continuity of Φ and M_3 in a, we can maximize the right hand side over $|a| < r$ to deduce that

$$|\Phi(u, a; y)| \leq K(r, y).$$

Thus $\Phi(\cdot, \cdot; y)$ is bounded on bounded sets.

Now we can proceed as in [39]. Using that $(\mu_0 \times \nu_0)(X' \times \Lambda') = 1$, we have

that

$$Z(y) = \int_{X' \times \Lambda'} \exp(-\Phi(u, a; y))\, \mu_0(\mathrm{d}u)\nu_0(\mathrm{d}a).$$

Set $B' = (X' \times \Lambda') \cap B$, and set

$$R = \sup\{\max\{\|u\|_X, |a|\} \,|\, (u, a) \in B'\}.$$

We deduce that

$$\sup_{(u,a)\in B'} \Phi(u, a; y) \leq K(R, y) < \infty$$

and so

$$Z(y) \geq \int_{B'} \exp(-K(R, y))\, \mu_0(\mathrm{d}u)\nu_0(\mathrm{d}a) = \exp(-K(R, y))(\mu_0 \times \nu_0)(B') > 0.$$

Hence the measure μ^y is well-defined.

(iii) The well-posedness of the posterior is proved in virtually the same way as Theorem 4.5 in [39].

□

2.8.3 Results From Section 2.5

Throughout this section, for $\delta > 0$ and $(u, a) \in X \times \Lambda$, we will denote $\mathcal{J}^\delta(u, a) = \mu(B^\delta(u, a))$. To prove Theorems 2.5.1 and 2.5.2, we first require two lemmas.

Lemma 2.8.2. *Let* $(u_1, a_1), (u_2, a_2) \in E \times \mathrm{int}(S)$. *Then*

$$\begin{aligned}\lim_{\delta\downarrow 0} \frac{(\mu_0 \times \nu_0)(B^\delta(u_1, a_1))}{(\mu_0 \times \nu_0)(B^\delta(u_2, a_2))} &= e^{\frac{1}{2}\|u_2\|_E^2 - \frac{1}{2}\|u_1\|_E^2} \cdot \frac{\rho(a_1)}{\rho(a_2)} \\ &= \exp\left(J(u_2) + K(a_2) - J(u_1) - K(a_1)\right).\end{aligned}$$

Proof. We adapt the proof of Proposition 18.3 in [97] to first show that

$$(\mu_0 \times \nu_0)(B^\delta(u_1, a_1)) \sim e^{-\frac{1}{2}\|u_1\|_E^2}(\mu_0 \times \nu_0)(B^\delta(0, a_1)) \quad \text{as } \delta \downarrow 0.$$

The first half of the proof is almost identical to that in [97], though some care must be taken since we cannot (a priori) separate the integrals over balls in $X \times \Lambda$ into

products of those over balls in X and Λ. Using the Cameron-Martin theorem we see that

$$(\mu_0 \times \nu_0)(B^\delta(u_1, a_1)) = e^{-\frac{1}{2}\|u_1\|_E^2} \int_{B^\delta(0,a_1)} e^{\langle u_1, u\rangle_E}\, \mu_0(\mathrm{d}u)\nu_0(\mathrm{d}a).$$

Since $\langle u_1, -u\rangle_E = -\langle u_1, u\rangle_E$ and $B^\delta(0, a_1)$ is symmetric about $0 \in X$, it follows that

$$\begin{aligned}\int_{B^\delta(0,a_1)} e^{\langle u_1, u\rangle_E}\, \mu_0(\mathrm{d}u)\nu_0(\mathrm{d}a) &= \int_{B^\delta(0,a_1)} \frac{1}{2}\left(e^{\langle u_1, u\rangle_E} + e^{-\langle u_1, u\rangle_E}\right) \mu_0(\mathrm{d}u)\nu_0(\mathrm{d}a) \\ &\geq (\mu_0 \times \nu_0)(B^\delta(0, a_1))\end{aligned}$$

which gives the inequality

$$(\mu_0 \times \nu_0)(B^\delta(u_1, a_1)) \geq e^{-\frac{1}{2}\|u_1\|_E^2}(\mu_0 \times \nu_0)(B^\delta(0, a_1)). \tag{2.8.1}$$

For the opposite bound, we write $\langle u_1, \cdot\rangle_E$ as the sum of two functionals z_c and z_s on E. We aim to choose z_c to be continuous on E, and z_s 'small' in some sense. Then we have that

$$\begin{aligned}(\mu_0 \times \nu_0)(B^\delta(u_1, a_1)) &= e^{-\frac{1}{2}\|u_1\|_E^2} \int_{B^\delta(0,a_1)} e^{z_c(u)+z_s(u)}\, \mu_0(\mathrm{d}u)\nu_0(\mathrm{d}a) \\ &\leq \exp\left(-\frac{1}{2}\|u_1\|_E^2 + \delta \cdot \sup_{(u,a)\in B^1(0,a_1)} z_c(u)\right) \cdot \\ &\qquad \left[(\mu_0 \times \nu_0)(B^\delta(0, a_1)) + \int_{B^\delta(0,a_1)} (e^{z_s(u)} - 1)\, \mu_0(\mathrm{d}u)\nu_0(\mathrm{d}a)\right]\end{aligned}$$

where we have used the linearity of z_c to extract δ from the supremum. As in [97], using a result from [138], a special case of the Gaussian correlation conjecture, it follows that for any $C \in \mathbb{R}$ and any convex set $B \subseteq X$ symmetric about 0,

$$\mu_0(B \cap \{u \in X \mid |z_s(u)| > C\}) \leq \mu_0(B)\mu_0(|z_s(\cdot)| > C).$$

Then for any increasing function $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$, one has

$$\begin{aligned}
\int_{B^\delta(0,a_1)} \varphi(|z_s(u)|)\,\mu_0(\mathrm{d}u)\nu_0(\mathrm{d}a) &= \int_{X\times\Lambda} \varphi(|z_s(u)|)\mathbb{1}_{B^\delta(0,a_1)}(u,a)\,\mu_0(\mathrm{d}u)\nu_0(\mathrm{d}a) \\
&= \int_0^\infty (\mu_0\times\nu_0)(\{(u,a)\in B^\delta(0,a_1) \mid \varphi(|z_s(u)|) > t\})\,\mathrm{d}t \\
&= \int_0^\infty (\mu_0\times\nu_0)(\{(u,a)\in B^\delta(0,a_1) \mid |z_s(u)| > \varphi^{-1}(t)\})\,\mathrm{d}t \\
&= \int_0^\infty \int_\Lambda \mu_0(\{u\in X \mid (u,a)\in B^\delta(0,a_1), |z_s(u)| > \varphi^{-1}(t)\})\,\nu_0(\mathrm{d}a)\mathrm{d}t \\
&\le \int_0^\infty \int_\Lambda \mu_0(\{u\in X \mid (u,a)\in B^\delta(0,a_1)\})\mu_0(|z_s(\cdot)| > \varphi^{-1}(t))\,\nu_0(\mathrm{d}a)\mathrm{d}t \\
&= \int_0^\infty \mu_0(|z_s(\cdot)| > \varphi^{-1}(t)) \underbrace{\left(\int_\Lambda \mu_0(\{u\in X \mid (u,a)\in B^\delta(0,a_1)\})\,\nu_0(\mathrm{d}a)\right)}_{(\mu_0\times\nu_0)(B^\delta(0,a_1))}\mathrm{d}t \\
&= (\mu_0\times\nu_0)(B^\delta(0,a_1)) \int_0^\infty \mu_0(|z_s(\cdot)| > \varphi^{-1}(t))\,\mathrm{d}t \\
&= (\mu_0\times\nu_0)(B^\delta(0,a_1)) \int_0^\infty \mu_0(\varphi(|z_s(\cdot)|) > t)\,\mathrm{d}t \\
&= (\mu_0\times\nu_0)(B^\delta(0,a_1)) \int_X \varphi(|z_s(u)|)\,\mu_0(\mathrm{d}u).
\end{aligned}$$

Choosing $\varphi(\cdot) = \exp(\cdot) - 1$ in this formula gives

$$\int_{B^\delta(0,a_1)} (e^{|z_s(u)|} - 1)\,\mu_0(\mathrm{d}u)\nu_0(\mathrm{d}a) \le (\mu_0\times\nu_0)(B^\delta(0,a_1)) \int_X (e^{|z_s(u)|} - 1)\,\mu_0(\mathrm{d}u).$$

The space of linear measurable functionals on E, which contains $\langle u_1, \cdot\rangle_E$, is the L^2 closure of E^*. Thus for any $\varepsilon > 0$, the functionals z_c, z_s can be chosen in order that the first of them is continuous and the second of them satisfies the inequality

$$\int_X (e^{|z_s(u)|} - 1)\,\mu_0(\mathrm{d}u) \le \varepsilon.$$

It follows that for each $\varepsilon > 0$ we have

$$\begin{aligned}
&(\mu_0\times\nu_0)(B^\delta(u_1,a_1)) \\
&\qquad \le \exp\left(-\frac{1}{2}\|u_1\|_E^2 + \delta\cdot \sup_{(u,a)\in B^1(0,a_1)} z_c(u)\right)(\mu_0\times\nu_0)(B^\delta(0,a_1))(1+\varepsilon).
\end{aligned} \tag{2.8.2}$$

Since balls are bounded, $\varepsilon > 0$ is arbitrary and z_c is continuous, we can combine

(2.8.1) and (2.8.2) to deduce that there exists $M > 0$ such that

$$\begin{aligned} e^{-\frac{1}{2}\|u_1\|_E^2}(\mu_0 \times \nu_0)(B^\delta(0, a_1)) &\leq (\mu_0 \times \nu_0)(B^\delta(u_1, a_1)) \\ &\leq e^{-\frac{1}{2}\|u_1\|_E^2 + M\delta}(\mu_0 \times \nu_0)(B^\delta(0, a_1)). \end{aligned}$$

Now looking at the ratio of measures we see

$$\lim_{\delta \downarrow 0} \frac{(\mu_0 \times \nu_0)(B^\delta(u_1, a_1))}{(\mu_0 \times \nu_0)(B^\delta(u_2, a_2))} = e^{\frac{1}{2}\|u_2\|_E^2 - \frac{1}{2}\|u_1\|_E^2} \cdot \lim_{\delta \downarrow 0} \frac{(\mu_0 \times \nu_0)(B^\delta(0, a_1))}{(\mu_0 \times \nu_0)(B^\delta(0, a_2))}.$$

We now deal with the geometric parameters. Let $a^* \in \mathrm{int}(S)$ so that ρ is positive in a neighbourhood of a^* (we may take $a^* = a_1$ or a_2 since we assume they lie in $\mathrm{int}(S)$). Then

$$\begin{aligned} \frac{(\mu_0 \times \nu_0)(B^\delta(0, a_1))}{(\mu_0 \times \nu_0)(B^\delta(0, a_2))} &= \frac{\int_{B^\delta(0,a_1)} \rho(a)\mu_0(\mathrm{d}u)\mathrm{d}a}{\int_{B^\delta(0,a_2)} \rho(a)\mu_0(\mathrm{d}u)\mathrm{d}a} \\ &= \frac{\int_{B^\delta(0,a^*)} \rho(a + a_1 - a^*)\, \mu_0(\mathrm{d}u)\mathrm{d}a}{\int_{B^\delta(0,a^*)} \rho(a + a_2 - a^*)\, \mu_0(\mathrm{d}u)\mathrm{d}a} \\ &= \frac{\int_{B^\delta(0,a^*)} \frac{\rho(a+a_1-a^*)}{\rho(a)}\, \mu_0(\mathrm{d}u)\nu_0(\mathrm{d}a)}{\int_{B^\delta(0,a^*)} \frac{\rho(a+a_2-a^*)}{\rho(a)}\, \mu_0(\mathrm{d}u)\nu_0(\mathrm{d}a)}. \end{aligned}$$

For sufficiently small δ both of the integrands are continuous. A mean-value property hence holds for the integrals, and so we may divide both the numerator and denominator by $(\mu_0 \times \nu_0)(B^\delta(0, a^*))$ and take limits to obtain

$$\begin{aligned} \lim_{\delta \downarrow 0} \frac{(\mu_0 \times \nu_0)(B^\delta(0, a_1))}{(\mu_0 \times \nu_0)(B^\delta(0, a_2))} &= \frac{\left.\dfrac{\rho(a + a_1 - a^*)}{\rho(a)}\right|_{a=a^*}}{\left.\dfrac{\rho(a + a_2 - a^*)}{\rho(a)}\right|_{a=a^*}} \\ &= \frac{\rho(a_1)}{\rho(a_2)}. \end{aligned}$$

We conclude that

$$\begin{aligned} \lim_{\delta \downarrow 0} \frac{(\mu_0 \times \nu_0)(B^\delta(u_1, a_1))}{(\mu_0 \times \nu_0)(B^\delta(u_2, a_2))} &= e^{\frac{1}{2}\|u_2\|_E^2 - \frac{1}{2}\|u_1\|_E^2} \cdot \frac{\rho(a_1)}{\rho(a_2)} \\ &= \exp\left(J(u_2) + K(a_2) - J(u_1) - K(a_1)\right). \end{aligned}$$

□

Lemma 2.8.3. *Let $f, g : \Lambda \to \mathbb{R}$ be continuous, and $(u_1, a_1), (u_2, a_2) \in E \times \mathrm{int}(S)$.*

Then

$$\lim_{\delta \downarrow 0} \frac{\int_{B^\delta(u_1,a_1)} f(a)\,\mu_0(\mathrm{d}u)\nu_0(\mathrm{d}a)}{\int_{B^\delta(u_2,a_2)} g(a)\,\mu_0(\mathrm{d}a)\nu_0(\mathrm{d}a)} = e^{\frac{1}{2}\|u_2\|_E^2 - \frac{1}{2}\|u_1\|_E^2} \cdot \frac{\rho(a_1)}{\rho(a_2)} \cdot \frac{f(a_1)}{g(a_2)}.$$

Proof. Let $\varepsilon > 0$. Then by the continuity of f and g, and the assumption on the norm on $X \times \Lambda$, there exists $\delta > 0$ such that

$$\begin{aligned}\frac{(f(a_1)-\varepsilon)(\mu_0\times\nu_0)(B^\delta(u_1,a_1))}{(g(a_2)+\varepsilon)(\mu_0\times\nu_0)(B^\delta(u_2,a_2))} &\leq \frac{\int_{B^\delta(u_1,a_1)} f(a)\,\mu_0(\mathrm{d}u)\nu_0(\mathrm{d}a)}{\int_{B^\delta(u_2,a_2)} g(a)\,\mu_0(\mathrm{d}u)\nu_0(\mathrm{d}a)} \\ &\leq \frac{(f(a_1)+\varepsilon)(\mu_0\times\nu_0)(B^\delta(u_1,a_1))}{(g(a_2)-\varepsilon)(\mu_0\times\nu_0)(B^\delta(u_2,a_2))}.\end{aligned}$$

The result now follows by the previous lemma. □

Proof of Theorem 2.5.1. Let $(u_1,a_1),(u_2,a_2) \in E \times \mathrm{int}(S)$. The case $\Phi \equiv 0$ is the result of Lemma 2.8.2. Now proceeding analagously to [38],

$$\begin{aligned}\frac{\mathcal{J}^\delta(u_1,a_1)}{\mathcal{J}^\delta(u_2,a_2)} &= \frac{\int_{B^\delta(u_1,a_1)} \exp(-\Phi(u,a))\,\mu_0(\mathrm{d}u)\nu_0(\mathrm{d}a)}{\int_{B^\delta(u_2,a_2)} \exp(-\Phi(u,a))\,\mu_0(\mathrm{d}u)\nu_0(\mathrm{d}a)} \\ &= \frac{\int_{B^\delta(u_1,a_1)} \exp(-\Phi(u,a)+\Phi(u_1,a_1))\exp(-\Phi(u_1,a_1))\,\mu_0(\mathrm{d}u)\nu_0(\mathrm{d}a)}{\int_{B^\delta(u_2,a_2)} \exp(-\Phi(u,a)+\Phi(u_2,a_2))\exp(-\Phi(u_2,a_2))\,\mu_0(\mathrm{d}u)\nu_0(\mathrm{d}a)}.\end{aligned}$$

Using Assumptions 2.4.1(iv), we have that for any $(u,a),(v,b) \in X \times \Lambda$,

$$|\Phi(u,a)-\Phi(v,b)| \leq M_3(r,a)\|u-v\|_X + |\Phi(v,a)-\Phi(v,b)|$$

where $r > \max\{\|u\|_X, \|v\|_X\}$. Now set

$$\begin{aligned}L_1 &= \max_{|a|<|a_1|+\delta} M_3(\|u_1\|_X+\delta,a), \\ L_2 &= \max_{|a|<|a_2|+\delta} M_3(\|u_2\|_X+\delta,a),\end{aligned}$$

which are finite due to the continuity assumption on M_3. Then

$$\begin{aligned}\frac{\mathcal{J}^\delta(u_1,a_1)}{\mathcal{J}^\delta(u_2,a_2)} &\leq e^{\delta(L_1+L_2)} e^{-\Phi(u_1,a_1)+\Phi(u_2,a_2))} \\ &\quad \times \frac{\int_{B^\delta(u_1,a_1)} \exp(|\Phi(u_1,a)-\Phi(u_1,a_1)|)\,\mu_0(\mathrm{d}u)\nu_0(\mathrm{d}a)}{\int_{B^\delta(u_2,a_2)} \exp(-|\Phi(u_2,a)-\Phi(u_2,a_2)|)\,\mu_0(\mathrm{d}u)\nu_0(\mathrm{d}a)}.\end{aligned}$$

Note that both integrands are continuous in a, and so we may use the previous

lemma. Taking $\limsup_{\delta\downarrow 0}$ of both sides gives

$$\limsup_{\delta\downarrow 0} \frac{\mathcal{J}^\delta(u_1,a_1)}{\mathcal{J}^\delta(u_2,a_2)} \leq e^{-I(u_1,a_1)+I(u_2,a_2)}.$$

A similar method gives that the $\liminf_{\delta\downarrow 0}$ is bounded below by the RHS and so we have that for any $(u_1,a_2),(u_2,a_2) \in E \times \text{int}(S)$,

$$\lim_{\delta\downarrow 0} \frac{\mathcal{J}^\delta(u_1,a_1)}{\mathcal{J}^\delta(u_2,a_2)} = e^{I(u_2,a_2)-I(u_1,a_1)}.$$

Noting that I is continuous on $E\times S$, we see that I agrees with the Onsager-Machlup functional on $E\times S$. Finally note that $I(u,a)=\infty$ on $(X\setminus E)\times\Lambda$ and $E\times(\Lambda\setminus S)$. $\square$

Remark 2.8.4. *Note that the limit above is independent of the choice of norm used on the product space $X\times\Lambda$ when referring to the balls. If we use the norm given by*

$$\|(x,a)\| = \max\{\|x\|_X, |a|\}$$

then we have that

$$B^\delta(u,a) = B^\delta(u)\times B^\delta(a)$$

and so may deduce that, for any choice of norm on $X\times\Lambda$,

$$\begin{aligned}\lim_{\delta\downarrow 0}\frac{(\mu_0\times\nu_0)(B^\delta(u_1,a_1))}{(\mu_0\times\nu_0)(B^\delta(u_2,a_2))} &= \lim_{\delta\downarrow 0}\frac{(\mu_0\times\nu_0)(B^\delta(u_1)\times B^\delta(a_1))}{(\mu_0\times\nu_0)(B^\delta(u_2)\times B^\delta(a_2))}\\ &= \lim_{\delta\downarrow 0}\frac{\mu_0(B^\delta(u_1))}{\mu_0(B^\delta(u_2))}\cdot\frac{\nu_0(B^\delta(a_1))}{\nu_0(B^\delta(a_2))}.\end{aligned}$$

This will be useful later for separating integrals.

Proof of Theorem 2.5.2. We follow the idea of the proof of Theorem 5.4 in [131], which is based on [36] and [79], and first show $I=\Phi+J+K$ is weakly lower semicontinuous on $E\times S$. Let $(u_n,a_n)\rightharpoonup(\bar{u},\bar{a})$ in $E\times S$. Since $S\subseteq\mathbb{R}^k$, weak convergence of the second component is equivalent to strong convergence. Since $\mu_0(X)=1$, E is compactly embedded in X and so $u_n\to\bar{u}$ strongly in X. In the proof of existence of the posterior distribution we showed that Φ is continuous on $X\times\Lambda$, and so we deduce that $\Phi(u_n,a_n)\to\Phi(u,a)$. Hence Φ is weakly continuous on $E\times S$. The functional J is weakly lower semicontinuous on E and K is continuous on S, and so I is weakly lower semicontinuous on $E\times S$.

Now we show I is coercive on $E \times S$. Since E is compactly embedded in X there exists a $C > 0$ such that

$$\|u\|_X^2 \leq C\|u\|_E^2.$$

Therefore by Assumption 2.4.1(i) it follows that, for any $\varepsilon > 0$, there is an $M(\varepsilon) \in \mathbb{R}$ such that

$$I(u,a) \geq M(\varepsilon) + \left(\frac{1}{2} - C\varepsilon\right)\|u\|_E^2 + K(a).$$

Since K is bounded below[2] by $-\log\|\rho\|_\infty$, we may incorporate this into the constant term $M(\varepsilon)$:

$$I(u,a) \geq \widetilde{M}(\varepsilon) + \left(\frac{1}{2} - C\varepsilon\right)\|u\|_E^2.$$

By choosing $\varepsilon = 1/4C$, we see that there is an $M \in \mathbb{R}$ such that, for all $(u,a) \in E \times S$,

$$I(u,a) \geq \frac{1}{4}\|u\|_E^2 + M$$

which establishes coercivity.

Now take a minimizing sequence (u_n, a_n) such that for any $\delta > 0$ there exists an $N_1 = N_1(\delta)$ such that

$$M \leq \bar{I} \leq I(u_n, a_n) \leq \bar{I} + \delta, \quad \forall n \geq N_1.$$

From the coercivity it can be seen that the sequence (u_n, a_n) is bounded in $E \times S$. Since $E \times S$ is a closed subset of a Hilbert space, there exists $(\bar{u}, \bar{a}) \in E \times S$ such that (possibly along a subsequence) $(u_n, a_n) \rightharpoonup (\bar{u}, \bar{a})$ in $E \times S$. From the weak lower semicontinuity of I it follows that, for any $\delta > 0$,

$$\bar{I} \leq I(\bar{u}, \bar{a}) \leq \bar{I} + \delta.$$

Since δ is arbitrary the first result follows.

Now consider the subsequence $(u_n, a_n) \rightharpoonup (\bar{u}, \bar{a})$. The convergence of $a_n \to \bar{a}$ is strong, so all that needs to be checked is that $u_n \to \bar{u}$ strongly in X. This follows from exactly the same argument as in the proof of Theorem 5.4 in [131] (taking $\bar{a}$

[2]Recall in subsection 2.3.3 we assumed ρ to be continuous on the compact set S, and hence bounded.

as the second parameter in I and Φ) and so the second result follows. $\square$

Before proving Theorem 2.5.3 we first collect some results on centred Gaussian measures from [38], specifically Lemmas 3.6, 3.7, and 3.9. For $u \in X$, let

$$\mathcal{J}_0^\delta(u) = \mu_0(B^\delta(u)).$$

Proposition 2.8.5. *(i) Let $\delta > 0$ and $u \in X$. Then we have*

$$\frac{\mathcal{J}_0^\delta(u)}{\mathcal{J}_0^\delta(0)} \leq ce^{-\frac{a_1}{2}(\|u\|_X - \delta)^2}$$

where $c = \exp\left(\frac{a_1}{2}\delta^2\right)$ and a_1 is a constant independent of z and δ.

(ii) Suppose that $\bar{u} \notin E$, $(u^\delta)_{\delta>0} \subseteq X$ and u^δ converges weakly to $\bar{u} \in X$ as $\delta \downarrow 0$. Then for any $\varepsilon > 0$ there exists δ small enough such that

$$\frac{\mathcal{J}_0^\delta(u^\delta)}{\mathcal{J}_0^\delta(0)} < \varepsilon.$$

(iii) Consider $(u^\delta)_{\delta>0} \subseteq X$ and suppose that u^δ converges weakly and not strongly to 0 in X as $\delta \downarrow 0$. Then for any $\varepsilon > 0$ there exists δ small enough such that

$$\frac{\mathcal{J}_0^\delta(u^\delta)}{\mathcal{J}_0^\delta(0)} < \varepsilon.$$

Proof of Theorem 2.5.3. (i) We first show (u^δ, a^δ) is bounded in $X \times \Lambda$. The boundedness of the second component is clear since S is bounded, so it suffices to show that (u^δ) is bounded in X. This is proved in the same way as in Theorem 3.5 in [38].

In the proof of existence of the posterior measure, Theorem 2.4.2, we show that if $r > 0$ and $\|u\|_X, |a| < r$, then there exists $K(r) > 0$ such that $\Phi(u, a) \leq K(r)$. Letting $c = e^M e^{-K(1)} > 0$, it follows in the same was as [38] that, given any $a \in S$, for $\delta < 1$ we have

$$\mathcal{J}_0^\delta(u^\delta, a) \geq c\mathcal{J}_0^\delta(0, a).$$

Suppose that (u^δ) is not bounded in X so that for any $R > 0$ there exists δ_R such that $\|u^{\delta_R}\|_X > R$, with $\delta_R \to 0$ as $R \to \infty$. Then the above bound says

that

$$\frac{\mathcal{J}_0^\delta(u^\delta, a)}{\mathcal{J}_0^\delta(0, a)} = \frac{\mu_0(B^\delta(u^\delta))}{\mu_0(B^\delta(0))} \cdot \frac{\nu_0(B^\delta(a))}{\nu_0(B^\delta(a))} \geq c.$$

This contradicts Proposition 2.8.5(i) above. Therefore there exists $R, \delta_R > 0$ such that

$$\|(u^\delta, a^\delta)\|_{X\times\Lambda} \leq R \quad \text{for any } \delta < \delta_R.$$

Hence there exist $(\bar{u}, \bar{a}) \in X \times \Lambda$ and a subsequence of $(u^\delta, a^\delta)_{0<\delta<\delta_R}$ which converges weakly in $X \times \Lambda$ to $(\bar{u}, \bar{a})$ as $\delta \downarrow 0$. For simplicity of notation we still call this subsequence (u^δ, a^δ).

We now show that (u^δ, a^δ) converges strongly to an element of $E \times S$. We first show that $(\bar{u}, \bar{a}) \in X \times S$.

Note that any limit point of a^δ must lie in S. Suppose it did not, and a limit point was $a^* \notin S$. Then there exists $\delta^\dagger > 0$ such that along a subsequence converging to a^*, $\delta < \delta^\dagger$ implies $a^\delta \notin S$ since S is closed. For $\delta < \frac{1}{2}\text{dist}(a^*, S) \wedge \delta^\dagger$ we then have $B^\delta(a^\delta) \cap S = \varnothing$. In particular $\nu_0(B^\delta(a^\delta)) = 0$ for all such δ, which in turn implies $\mathcal{J}^\delta(u, a^\delta) = 0$ for any $u \in X$ contradicting the definition of a^δ. It follows that we must have $\bar{a} \in S$.

We need to show $\bar{u} \in E$. From the definition of (u^δ, a^δ) and the bounds on Φ we have for δ small enough and some[3] α close to 1,

$$\begin{aligned} 1 \leq \frac{\mathcal{J}^\delta(u^\delta, 0)}{\mathcal{J}^\delta(0, 0)} &\leq \alpha \frac{e^{-M} \int_{B^\delta(u^\delta)} \mu_0(\mathrm{d}u) \int_{B^\delta(0)} \nu_0(\mathrm{d}a)}{e^{-K(1)} \int_{B^\delta(0)} \mu_0(\mathrm{d}u) \int_{B^\delta(0)} \nu_0(\mathrm{d}a)} \\ &= \alpha e^{K(1)-M} \frac{\int_{B^\delta(u^\delta)} \mu_0(\mathrm{d}u)}{\int_{B^\delta(0)} \mu_0(\mathrm{d}u)}. \end{aligned}$$

We use Proposition 2.8.5(ii). Supposing $\bar{u} \notin E$, for any $\varepsilon > 0$ there exists δ small enough such that

$$\frac{\int_{B^\delta(u^\delta)} \mu_0(\mathrm{d}u)}{\int_{B^\delta(0)} \mu_0(\mathrm{d}u)} < \varepsilon.$$

We may choose $\varepsilon = \frac{1}{2\alpha} e^{M-K(1)}$ to deduce that there exists δ small enough

[3] Remark 2.8.4 tells us that we can separate the integrals in the limit $\delta \downarrow 0$.

such that

$$1 \leq \frac{\mathcal{J}^\delta(u^\delta, 0)}{\mathcal{J}^\delta(0,0)} < \frac{1}{2}$$

which is a contradiction, and so $\bar{u} \in E$.

Knowing that $(\bar{u}, \bar{a}) \in E \times S$ we now show that the convergence is strong. Any convergence of the second component will be strong and so we just need to show that $u^\delta \to \bar{u}$ strongly in X. Suppose the convergence is not strong, then we may use Proposition 2.8.5(iii) on the sequence $u^\delta - \bar{u}$. The same choice of ε as above leads to the same contradiction, and so we deduce that $\bar{u} \to \bar{u}$ strongly in X and the first result is proved.

(ii) We now show that $(\bar{u}, \bar{a})$ is a MAP estimator and minimizes I. As in [38], and the proof of Theorem 2.5.1, we can use Assumptions 2.4.1(iii) to see that

$$\frac{\mathcal{J}^\delta(u^\delta, a^\delta)}{\mathcal{J}^\delta(\bar{u}, \bar{a})} \leq e^{\delta(L_1+L_2)} e^{-\Phi(u^\delta, a^\delta)+\Phi(\bar{u},\bar{a}))} \times \frac{\int_{B^\delta(u^\delta, a^\delta)} \exp(|\Phi(u^\delta, a) - \Phi(u^\delta, a^\delta)|)\, \mu_0(\mathrm{d}u)\nu_0(\mathrm{d}a)}{\int_{B^\delta(\bar{u},\bar{a})} \exp(-|\Phi(\bar{u}, a) - \Phi(\bar{u}, \bar{a})|)\, \mu_0(\mathrm{d}u)\nu_0(\mathrm{d}a)}$$

where

$$L_1 = \max_{|a| \leq |a_1|+\delta} M_3(\|u^\delta\|_X + \delta, a),$$
$$L_2 = \max_{|a| \leq |a_2|+\delta} M_3(\|\bar{u}\|_X + \delta, a).$$

Therefore using the continuity of Φ, as shown in the proof of existence of the posterior distribution, and that $(u^\delta, a^\delta) \to (\bar{u}, \bar{a})$ strongly in $X \times \Lambda$,

$$\limsup_{\delta \downarrow 0} \frac{\mathcal{J}^\delta(u^\delta, a^\delta)}{\mathcal{J}^\delta(\bar{u}, \bar{a})} \leq \limsup_{\delta \downarrow 0} \frac{\int_{B^\delta(u^\delta, a^\delta)} \mu_0(\mathrm{d}u)\nu_0(\mathrm{d}a)}{\int_{B^\delta(\bar{u},\bar{a})} \mu_0(\mathrm{d}u)\nu_0(\mathrm{d}a)}.$$

Suppose u^δ is not bounded in E, or if it is, it only converges weakly (and not strongly) in E. Then $\|\bar{u}\|_E < \liminf_{\delta \downarrow 0} \|u^\delta\|_E$, and hence for small enough δ,

$\|\bar{u}\|_E < \|u^\delta\|_E$. Therefore, since μ_0 is centered and $\|u^\delta - \bar{u}\|_X \to 0$, $|a^\delta - \bar{a}| \to 0$,

$$\begin{aligned}
\limsup_{\delta\downarrow 0} \frac{\int_{B^\delta(u^\delta,a^\delta)} \mu_0(\mathrm{d}u)\nu_0(\mathrm{d}a)}{\int_{B^\delta(\bar{u},\bar{a})} \mu_0(\mathrm{d}u)\nu_0(\mathrm{d}a)} &= \limsup_{\delta\downarrow 0} \frac{\int_{B^\delta(u^\delta)} \mu_0(\mathrm{d}u) \int_{B^\delta(a^\delta)} \nu_0(\mathrm{d}a)}{\int_{B^\delta(\bar{u})} \mu_0(\mathrm{d}u) \int_{B^\delta(\bar{a})} \nu_0(\mathrm{d}a)} \\
&\leq \limsup_{\delta\downarrow 0} \frac{\int_{B^\delta(u^\delta)} \mu_0(\mathrm{d}u)}{\int_{B^\delta(\bar{u})} \mu_0(\mathrm{d}u)} \cdot \limsup_{\delta\downarrow 0} \frac{\int_{B^\delta(a^\delta)} \nu_0(\mathrm{d}a)}{\int_{B^\delta(\bar{a})} \nu_0(\mathrm{d}a)} \\
&\leq \limsup_{\delta\downarrow 0} \frac{\int_{B^\delta(a^\delta)} \nu_0(\mathrm{d}a)}{\int_{B^\delta(\bar{a})} \nu_0(\mathrm{d}a)} \\
&= \limsup_{\delta\downarrow 0} \frac{\frac{1}{|B^\delta(a^\delta)|}\int_{B^\delta(a^\delta)} \rho(a)\,\mathrm{d}a}{\frac{1}{|B^\delta(\bar{a})|}\int_{B^\delta(\bar{a})} \rho(a)\,\mathrm{d}a} \\
&= 1.
\end{aligned}$$

The final equality above follows from the continuity of the integrand and the fact that $|a^\delta - \bar{a}| \to 0$: both the numerator and the denominator tend to $\rho(\bar{a})$.

Since by definition of (u^δ, a^δ), $\mathcal{J}^\delta(u^\delta, a^\delta) \geq \mathcal{J}^\delta(\bar{u}, \bar{a})$ and hence

$$\liminf_{\delta\downarrow 0} \frac{\mathcal{J}^\delta(u^\delta, a^\delta)}{\mathcal{J}^\delta(\bar{u}, \bar{a})} \geq 1,$$

this implies that

$$\lim_{\delta\downarrow 0} \frac{\mathcal{J}^\delta(u^\delta, a^\delta)}{\mathcal{J}^\delta(\bar{u}, \bar{a})} = 1. \tag{2.8.3}$$

In the case where (u^δ) converges strongly to $\bar{u}$ in E, we see from the proof of Lemma 2.8.2 that we have

$$\begin{aligned}
e^{\frac{1}{2}\|\bar{u}\|_E^2 - \frac{1}{2}\|u^\delta\|_E^2 - M\delta} \frac{(\mu_0\times\nu_0)(B^\delta(0,a^\delta))}{(\mu_0\times\nu_0)(B^\delta(0,\bar{a}))} &\leq \frac{(\mu_0\times\nu_0)(B^\delta(u^\delta,a^\delta))}{(\mu_0\times\nu_0)(B^\delta(\bar{u},\bar{a}))} \\
&\leq e^{\frac{1}{2}\|\bar{u}\|_E^2 - \frac{1}{2}\|u^\delta\|_E^2 + M\delta} \frac{(\mu_0\times\nu_0)(B^\delta(0,a^\delta))}{(\mu_0\times\nu_0)(B^\delta(0,\bar{a}))}.
\end{aligned}$$

Since we have $u^\delta \to \bar{u}$ strongly in E we have in particular that $\|u^\delta\|_E \to \|\bar{u}\|_E$. It follows that $e^{\frac{1}{2}\|\bar{u}\|_E^2 - \frac{1}{2}\|u^\delta\|_E^2 \pm M\delta} \to 1$ as $\delta \downarrow 0$. Now using the continuity of ρ and the fact that $|a^\delta - \bar{a}| \to 0$, an argument similar to that in the proof of Lemma 2.8.2 shows that

$$\lim_{\delta\downarrow 0} \frac{(\mu_0\times\nu_0)(B^\delta(0,a^\delta)}{(\mu_0\times\nu_0)(B^\delta(0,\bar{a}))} = 1.$$

We therefore deduce that

$$\lim_{\delta \downarrow 0} \frac{\int_{B^\delta(u^\delta, a^\delta)} \mu_0(\mathrm{d}u)\nu_0(\mathrm{d}a)}{\int_{B^\delta(\bar{u},\bar{a})} \mu_0(\mathrm{d}u)\nu_0(\mathrm{d}a)} = 1$$

and (2.8.3) follows again. Therefore $(\bar{u}, \bar{a})$ is a MAP estimator of the measure μ.

The proof that $(\bar{u}, \bar{a})$ minimizes I is identical to that in the proof of Theorem 3.5 in [38].

□

Chapter 3

The Bayesian Formulation of EIT

3.1 Introduction

3.1.1 Background

Electrical Impedance Tomography (EIT) is an imaging technique in which the conductivity of a body is inferred from electrode measurements on its surface. Examples include medical imaging, where the methodology is used to non-invasively detect abnormal tissue within a patient, and subsurface imaging where material properties of the subsurface are determined from surface (or occasional interior) measurements of the electrical response; the methodology is often referred to as electrical resistance tomography – ERT – in this context and discussed in more detail below. The concept of EIT appears as early as the late 1970's [66] and ERT the 1930's [90].

A very influential mathematical formulation of the inverse problem associated with EIT dates back to 1980, due to Calderón. He formulated an abstract version of the problem, in which the objective is recovery of the coefficient of a divergence form elliptic PDE from knowledge of its Neumann-to-Dirichlet or Dirichlet-to-Neumann operator. Specifically, in the Dirichlet-to-Neumann case, if $D \subseteq \mathbb{R}^d$ and $g \in H^{1/2}(\partial D)$ is given, let $u \in H^1(D)$ solve

$$\begin{cases} -\nabla \cdot (\sigma \nabla u) = 0 & \text{in } D \\ u = g & \text{on } \partial D. \end{cases}$$

The problem of interest is then to ask does the mapping $\Lambda_\sigma : H^{1/2}(\partial D) \to H^{-1/2}(\partial D)$ given by

$$g \mapsto \sigma \frac{\partial u}{\partial \nu}$$

determine the coefficient σ? Physically, g corresponds to boundary voltage measurements, and $\Lambda_\sigma(g)$ corresponds to the current density on the boundary. Much study has been carried out on this problem – some significant results, in the case where all conductivities are in $C^2(\overline{D})$ and $d \geq 3$, concern uniqueness [132], reconstruction [105], stability [6] and partial data [77]. Details of these results are summarized in [122].

In 1996, Nachman proved global uniqueness and provided a reconstruction procedure for the case $d = 2$, involving the use of a scattering transform and solving a D-bar problem [106]. The D-bar equation involved is a differential equation of the form $\overline{\partial} q = f$, where $\overline{\partial}$ denotes the conjugate of the complex derivative and f depends on the scattering transform. A regularized D-bar approach, involving the truncation of the scattering transform, was provided in [82, 83], enabling the recovery of features of discontinuous permeabilities. The regularized D-bar approach is also used in [84], for the case when the data is not of infinite precision. Other work in the area includes joint inference of the shape of the domain and conductivity [86].

For applications, a more physically appropriate model for EIT was provided in 1992 in [128]. This model, referred to as the complete electrode model (CEM), replaces complete boundary potential measurements with measurements of constant potential along electrodes on the boundary, subject to contact impedances. The authors show that predictions from this model agree with experimental measurements up to the measurement precision of 0.1%. For this model they also prove existence and uniqueness of the associated electric potential. It is this model that we shall consider in this chapter, and it is outlined in section 3.2.

When using the CEM, there is a limitation on the number of measurements that can be taken to provide additional information due to the linear relationship between current and potential. The data is therefore finite dimensional in the inverse problem, as distinct from the Calderón problem where knowledge of an infinite dimensional operator is assumed. As a consequence, reconstruction using the CEM often makes use of Tikhonov regularisation. The paper [52] analyses numerical convergence when an H^1 or TV penalty term is used, with a finite element discretisation of the problem. We will adopt a Bayesian approach to regularisation, and this is discussed below.

A closely related problem to EIT is Electrical Resistivity Tomography (ERT), which concerns subsurface inference from surface potential measurements, see for example [90] which discussed the problem as early as 1933. Physically the main difference between EIT and ERT is that alternating current is typically used for the former, and direct current for the latter. Additionally, due to the scale of ERT, it is a reasonable approximation to model the electrodes as points, rather than using the CEM. Another difference between the two is that the relative contrast between the conductivities of different media are typically higher in subsurface applications than medical applications, which permits the approximation of the problem by a network of resistors in some cases [116]. Nonetheless, the Bayesian theory and MCMC methodology introduced here will be useful for the ERT problem as well as the EIT problem.

Statistical, in particular Bayesian, approaches to EIT inversion have previously been studied, for example in [73, 74, 93]. In [73], the authors prove certain regularity of the forward map associated with the CEM, formulate the Bayesian inverse problem in terms of the discretized model, and investigate the effect of different priors on reconstruction and behaviour of the posterior. The paper [93] focuses on Whittle-Matèrn priors, using EIT and ERT as examples for numerical simulation. The paper [74] presents a regularized version of the inverse problem, which admits a Bayesian interpretation.

The Bayesian approach to inverse problems, especially in infinite dimensions, is a relatively new technique. Two approaches are typically taken: discretize first and use finite dimensional Bayesian techniques, or apply the Bayesian methodology directly on function space prior to discretizing. The former approach is outlined in [75]. The latter approach for linear problems has been studied in [50, 96, 101]. More recently, this approach has been applied to nonlinear problems [39, 91, 92, 131]. It is this approach that we will be taking in this chapter.

3.1.2 Our Contribution

The main contributions of this chapter are as follows:

(i) This is the first rigorous Bayesian formulation of EIT given in infinite dimensions.

(ii) We employ a variety of prior models, based on the assumption that the underlying conductivity we wish to recover is binary. We initially look at log-Gaussian

priors, before focusing on priors which enforce the binary field property. These binary field priors include both single star-shaped inclusions parametrized by their centre and by a radial function [22], and arbitrary geometric interfaces between the two conductivity values defined via level set functions [72].

(iii) This setting leads to proof that the posterior measure is Lipschitz in the data, with respect to the Hellinger metric, for all three priors studied; further stability properties of the posterior with respect to perturbations, such as numerical approximation of the forward model, may be proved similarly.

(iv) Numerical results using state of the art MCMC demonstrate the importance of the prior choice for accurate reconstruction in the severely underdetermined inverse problems arising in EIT.

3.1.3 Organisation of the Chapter

In section 3.2 we describe the forward map associated with the EIT problem, and prove relevant regularity properties. In section 3.3 we formulate the inverse problem rigorously and describe our three prior models. We then prove existence and well-posedness of the posterior distribution for each of these choices of prior. In section 3.4 we present results of numerical MCMC simulations to investigate the effect of the choice of prior on the recovery of certain binary conductivity fields. We conclude in section 3.5.

3.2 The Forward Model

In subsection 3.2.1 we describe the complete electrode model for EIT as given in [128]. In subsection 3.2.2 we give the weak formulation of this model, stating assumptions required for the quoted existence and uniqueness result. Then in subsection 3.2.3 we define the forward map in terms of this model, and prove that this map is continuous with respect to both uniform convergence and convergence in measure.

3.2.1 Problem Statement

Let $D \subseteq \mathbb{R}^d$, $d \leq 3$, be a bounded open set representing a body, with conductivity $\sigma : \overline{D} \to \mathbb{R}$. A number L of electrodes are attached to the surface of the body. We

treat these as subsets $(e_l)_{l=1}^L$ of the boundary ∂D, and assume that they have contact impedances $(z_l)_{l=1}^L \in \mathbb{R}^L$. A current stimulation pattern $(I_l)_{l=1}^L \in \mathbb{R}^L$ is applied to the electrodes. Then the electric potential v within the body and boundary voltages $(V_l)_{l=1}^L \in \mathbb{R}^L$ on $(e_l)_{l=1}^L$ are modelled by the following partial differential equation.

$$\begin{cases} -\nabla \cdot (\sigma(x)\nabla v(x)) = 0 & x \in D \\ \displaystyle\int_{e_l} \sigma \frac{\partial v}{\partial n} \,\mathrm{d}S = I_l & l = 1, \ldots, L \\ \sigma(x)\dfrac{\partial v}{\partial n}(x) = 0 & x \in \partial D \setminus \bigcup_{l=1}^L e_l \\ v(x) + z_l \sigma(x) \dfrac{\partial v}{\partial n}(x) = V_l & x \in e_l, l = 1, \ldots, L \end{cases} \tag{3.2.1}$$

This model was first proposed in [128]; a derivation can be found therein. Note that the inputs to this forward model are the conductivity σ, input current $(I_l)_{l=1}^L$ and contact impedances $(z_l)_{l=1}^L$. The solution comprises the function $v : D \to \mathbb{R}$ *and* the vector $V \in \mathbb{R}^L$ of voltages. Also note that solutions to this equation are only defined up to addition of a constant: if (v, V) solves the equation, then so does $(v + c, V + c)$ for any $c \in \mathbb{R}$. This is because it is necessary to choose a reference ground voltage.

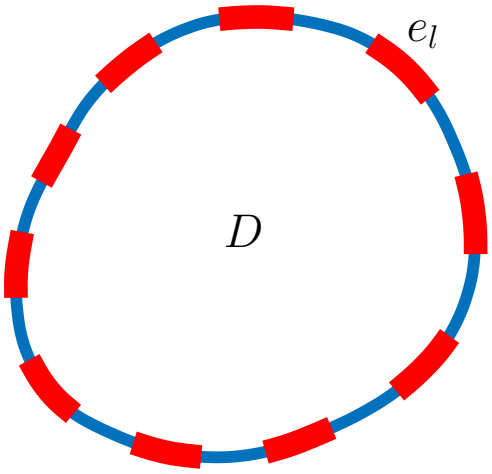


Figure 3.1: An example domain D with electrodes $(e_l)_{l=1}^L$ attached to its boundary.

3.2.2 Weak Formulation

We first define the space in which the solution to equation (3.2.1) will live. Using the notation of [128] we set

$$\begin{aligned}\mathbb{H} &= H^1(D) \oplus \mathbb{R}^L,\\ \|(v,V)\|_{\mathbb{H}}^2 &= \|v\|_{H^1(D)}^2 + \|V\|_{\ell^2}^2\\ &= \|v\|_{L^2(D)}^2 + \|\nabla v\|_{L^2(D)}^2 + \|V\|_{\ell^2}^2.\end{aligned}$$

Since solutions are only defined up to addition of a constant, we define the quotient space $(\dot{\mathbb{H}}, \|\cdot\|_{\dot{\mathbb{H}}})$ by

$$\begin{aligned}\dot{\mathbb{H}} &= \mathbb{H}/\mathbb{R},\\ \|(v,V)\|_{\dot{\mathbb{H}}} &= \inf_{c\in\mathbb{R}} \|(v-c, V-c)\|_{\mathbb{H}}.\end{aligned}$$

We will often use the notation $v' = (v, V)$ for brevity. It is more convenient to equip $\dot{\mathbb{H}}$ with an equivalent norm, as stated in the following lemma from [128]:

Lemma 3.2.1. *Define* $\|v'\|_*$ *by*

$$\|v'\|_*^2 = \|\nabla v\|_{L^2(D)}^2 + \sum_{l=1}^{L} \int_{e_l} |v(x) - V_l|^2 \,\mathrm{d}S.$$

Then $\|\cdot\|_*$ *and* $\|\cdot\|_{\dot{\mathbb{H}}}$ *are equivalent.*

We can now state the weak formulation of the problem as derived in [128]. For this let $w' = (w, W)$.

Proposition 3.2.2. *Let* $B : \dot{\mathbb{H}} \times \dot{\mathbb{H}} \to \mathbb{R}$ *and* $r : \dot{\mathbb{H}} \to \mathbb{R}$ *be defined by*

$$\begin{aligned}B(v', w'; \sigma) &= \int_D \sigma \nabla v \cdot \nabla w \,\mathrm{d}x + \sum_{l=1}^{L} \frac{1}{z_l} \int_{e_l} (v - V_l)(w - W_l) \,\mathrm{d}S,\\ r(w') &= \sum_{l=1}^{L} I_l W_l.\end{aligned}$$

Then if $v' \in \dot{\mathbb{H}}$ *is a strong solution to the problem (3.2.1), it satisfies*

$$B(v', w'; \sigma) = r(w') \quad \textit{for all } w' \in \dot{\mathbb{H}}. \tag{3.2.2}$$

We will use the weak formulation (3.2.2) to define our forward map for the complete

electrode model (3.2.1). In order to guarantee a solution to this problem, we make the following assumptions.

Assumptions 3.2.3. *The conductivity $\sigma : \overline{D} \to \mathbb{R}$, the contact impedances $(z_l)_{l=1}^L \in \mathbb{R}^L$ and the current stimulation pattern $(I_l)_{l=1}^L \in \mathbb{R}^L$ satisfy*

(i) $\sigma \in L^\infty(\overline{D}; \mathbb{R}), \quad \operatorname{essinf}_{x \in D} \sigma(x) = \sigma_- > 0;$

(ii) $0 < z_- \leq z_l \leq z_+ < \infty, \quad l = 1, \ldots, L;$

(iii) $\displaystyle\sum_{l=1}^{L} I_l = 0.$

Under these assumptions, existence of a unique solution to (3.2.2) is proved in [128] and stated here for convenience:

Proposition 3.2.4. *Let Assumptions 3.2.3 hold, then there is a unique $[(v, V)] \in \dot{\mathbb{H}}$ solving (3.2.2). We may, without loss of generality, choose the element $(v, V) \in [(v, V)]$ of the equivalence class of solutions to be that which satisfies*

$$\sum_{l=1}^{L} V_l = 0. \tag{3.2.3}$$

Remark 3.2.5. *Assumptions 3.2.3 (i) and (ii) are to ensure coercivity and boundedness of $B(\cdot, \cdot; \sigma)$. Assumption 3.2.3 (iii) is necessary for continuity of $r(\cdot)$, and physically may be thought of as a conservation of charge condition. Choosing a solution from the equivalence class corresponds to choosing a reference ground voltage.*

3.2.3 Continuity of the Forward Map

In what follows we will restrict to the set of admissible conductivities, which is defined as follows.

Definition 3.2.6. *A conductivity field $\sigma : \overline{D} \to \mathbb{R}$ is said to be admissible if*

(i) there exists $N \in \mathbb{N}$, $\{D_n\}_{n=1}^N$ open disjoint subsets of D for which $\overline{D} = \bigcup_{n=1}^N \overline{D}_j$;

(ii) $\sigma\big|_{D_j} \in C(\overline{D}_j)$; and

(iii) there exist $\sigma^-, \sigma^+ \in (0, \infty)$ such that $\sigma^- \leq \sigma(x) \leq \sigma^+$ for all $x \in \overline{D}$.

The set of all such conductivities will be denoted $\mathcal{A}(D)$.

Note that any $\sigma \in \mathcal{A}(D)$ will satisfy Assumptions 3.2.3(i). Assume that the current stimulation pattern $(I_l)_{l=1}^L \in \mathbb{R}^L$ and contact impedances $(z_l)_{l=1}^L \in \mathbb{R}^L$ are known and satisfy Assumptions 3.2.3(ii)-(iii). Then we may define the solution map $\mathcal{M} : \mathcal{A}(D) \to \mathbb{H}$ to be the unique solution to (3.2.2) satisfying (3.2.3). The above existence and uniqueness result tells us that this map is well-defined.

In [73] it is shown that $\mathcal{M} : \mathcal{A}(D) \to \mathbb{H}$ is Fréchet differentiable when we equip $\mathcal{A}(D)$ with the supremum norm. Though this is a strong result, this choice of norm is not appropriate for all of the conductivities that we will be considering. More specifically, as we are mainly interested in binary conductivities, any two conductivities with different interfaces separating the phases will be regarded as a constant distance from one another in the supremum norm, regardless of how little the interfaces differ.

We hence establish the following continuity result.

Proposition 3.2.7. *Fix a current stimulation pattern $(I_l)_{l=1}^L \in \mathbb{R}^L$ and contact impedances $(z_l)_{l=1}^L \in \mathbb{R}^L$ satisfying Assumptions 3.2.3. Define the solution map $\mathcal{M} : \mathcal{A}(D) \to \mathbb{H}$ as above. Let $\sigma \in \mathcal{A}(D)$ and let $(\sigma_\varepsilon)_{\varepsilon>0} \subseteq \mathcal{A}(D)$ be such that either*

(i) σ_ε converges to σ uniformly; or

(ii) σ_ε converges to σ in measure, and there exist $\sigma^-, \sigma^+ \in (0, \infty)$ such that for all $\varepsilon > 0$ and $x \in D$, $\sigma^- \leq \sigma_\varepsilon(x) \leq \sigma^+$.

Then $\|\mathcal{M}(\sigma_\varepsilon) - \mathcal{M}(\sigma)\|_* \to 0$.

Proof. Define the maps $B : \mathbb{H}\times\mathbb{H}\times\mathcal{A}(D) \to \mathbb{R}$ and $r : \mathbb{H} \to \mathbb{R}$ as in Lemma 3.2.2, but on $\mathbb{H}$ rather than $\dot{\mathbb{H}}$. Then denoting $u'_\varepsilon = (v_\varepsilon, V^\varepsilon) = \mathcal{M}(\sigma_\varepsilon)$ and $v' = (v, V) = \mathcal{M}(\sigma)$, we have for all $w' \in \mathbb{H}$,

$$B(v'_\varepsilon, w'; \sigma_\varepsilon) = r(w'), \quad B(v', w'; \sigma) = r(w').$$

It follows that

$$\begin{aligned} 0 &= B(v'_\varepsilon, w'; \sigma_\varepsilon) - B(v', w'; \sigma_\varepsilon) + B(v', w'; \sigma_\varepsilon) - B(v', w'; \sigma) \\ &= \int_D \sigma_\varepsilon \nabla(v_\varepsilon - v) \cdot \nabla w \, \mathrm{d}x + \sum_{l=1}^L \frac{1}{z_l} \int_{e_l} ((v_\varepsilon - v) - (V_l^\varepsilon - V_l))(w - W_l) \, \mathrm{d}S \\ &\quad + \int_D (\sigma_\varepsilon - \sigma) \nabla v \cdot \nabla w \, \mathrm{d}x. \end{aligned}$$

Letting $w' = (v_\varepsilon - v, V^\varepsilon - V)$, we see that

$$\int_D \sigma_\varepsilon |\nabla(v_\varepsilon - v)|^2 \,\mathrm{d}x + \sum_{l=1}^{L} \frac{1}{z_l} \int_{e_l} ((v_\varepsilon - v) - (V_l^\varepsilon - V_l))^2 \,\mathrm{d}S \\ \leq \int_D |\sigma_\varepsilon - \sigma||\nabla v \cdot \nabla(v_\varepsilon - v)| \,\mathrm{d}x.$$

In both cases (i) and (ii), we have that $(\sigma_\varepsilon)_{\varepsilon>0}$ is bounded uniformly below by a positive constant. Hence for small enough ε, the left hand side above may be bounded below by $C\|v'_\varepsilon - v'\|_*^2$. We then have by Cauchy-Schwarz

$$\begin{aligned} \|v'_\varepsilon - v'\|_*^2 &\leq C \int_D |\sigma_\varepsilon - \sigma||\nabla v \cdot \nabla(v_\varepsilon - v)| \,\mathrm{d}x \\ &\leq C \left(\int_D |\sigma_\varepsilon - \sigma|^2 |\nabla v|^2 \,\mathrm{d}x \right)^{1/2} \cdot \|\nabla(v_\varepsilon - v)\|_{L^2} \\ &\leq C \left(\int_D |\sigma_\varepsilon - \sigma|^2 |\nabla v|^2 \,\mathrm{d}x \right)^{1/2} \cdot \|v'_\varepsilon - v'\|_* \qquad (3.2.4) \\ &\leq C \|\sigma_\varepsilon - \sigma\|_\infty \|\nabla v\|_{L^2} \|v'_\varepsilon - v'\|_*. \qquad (3.2.5) \end{aligned}$$

If $\sigma_\varepsilon \to \sigma$ uniformly, we deduce from (3.2.5) that $\|v'_\varepsilon - v'\|_* \to 0$ and the result follows. If $|\sigma_\varepsilon - \sigma| \to 0$ in measure, then since $|D| < \infty$, it follows that the integrand in (3.2.4) tends to zero in measure, see for example Corollary 2.2.6 in [17]. Since σ_ε is assumed to be uniformly bounded, the integrand is dominated by a scalar multiple of the integrable function $|\nabla v|^2$. We claim that this implies that the integrand tends to zero in L^1. Suppose not, and denote the integrand f_ε. Then there exists $\delta > 0$ and a subsequence $(f_{\varepsilon_i})_{i\geq 1}$ such that $\|f_{\varepsilon_i}\|_{L^1} \geq \delta$ for all i. This subsequence still converges to zero in measure, and so admits a further subsequence that converges to zero almost surely. An application of the dominated convergence theorem leads to a contradiction, hence we deduce that f_ε tends to zero in L^1 and the result follows. □

Denote the projection $\Pi : \mathbb{H} \to \mathbb{R}^L$, $(v, V) \mapsto V$. The following lemma shows that the above result still holds if we replace $\mathcal{M}$ by $\Pi \circ \mathcal{M}$.

Corollary 3.2.8. *Let the assumptions of Proposition 3.2.7 hold. Then*

$$|\Pi \circ \mathcal{M}(\sigma_\varepsilon) - \Pi \circ \mathcal{M}(\sigma)|_{\ell^2} \to 0.$$

Proof. We show that there exists $C > 0$ such that for all $(v, V) \in \mathbb{H}$ with $\sum_{l=1}^{L} V_l =$

0, $\|(v,V)\|_* \geq C|V|_{\ell^2}$. By the equivalence of $\|\cdot\|_*$ and $\|\cdot\|_{\dot{\mathbb{H}}}$, Lemma 3.2.1, we have

$$\|(v,V)\|_* \geq C \inf_{c\in\mathbb{R}} (\|v-c\|_{H^1} + |V-c|_{\ell^2}) \geq C \inf_{c\in\mathbb{R}} |V-c|_{\ell^2}$$

The infimum on the right-hand side is attained at

$$c = \frac{1}{L}\sum_{l=1}^{L} V_l = 0.$$

Then by Proposition 3.2.7, we have

$$0 \leq |\Pi \circ \mathcal{M}(\sigma_\varepsilon) - \Pi \circ \mathcal{M}(\sigma)|_{\ell^2} \leq \|\mathcal{M}(\sigma_\varepsilon) - \mathcal{M}(\sigma)\|_* \to 0$$

□

3.3 The Inverse Problem

We are interested in the inverse problem of determining the conductivity field from measurements of the voltages $(V_l)_{l=1}^L$ on the boundary, for a variety of input currents $(I_l)_{l=1}^L$ on the boundary. To this end we introduce the following version of Ohm's law. Observe that the mapping $I \mapsto v'$, taking the current stimulation pattern to the solution of (3.2.2), is linear. Then given a conductivity field $\sigma \in \mathcal{A}(D)$, there exists a resistivity matrix $R(\sigma) \in \mathbb{R}^{L\times L}$ such that the boundary voltage measurements $V(\sigma)$ arising from the solution of the forward model are related to I via

$$V(\sigma) = R(\sigma)I$$

By applying several different current stimulation patterns we should be able to infer more about the conductivity σ. Note however that since the mapping $I \mapsto V$ is linear, only linearly independent stimulation patterns will provide more information[1]. Since we have the conservation of charge condition on I, there are at most $L-1$ linearly independent patterns we can use.

Assume that J linearly independent current patterns $I^{(j)} \in \mathbb{R}^L$, $j = 1,\ldots,J$, $J \leq$

[1] If there is noise on the measurements, additional linearly dependent observations can be made to effectively reduce the noise level on the original measurements. We can assume that this has been done and scale the noise appropriately.

$L-1$ are applied, and noisy measurements of $V^{(j)} = R(\sigma)I^{(j)}$ are made:

$$y_j = V^{(j)} + \eta_j, \quad \eta_j \sim N(0, \Gamma_0) \text{ iid.}$$

We have

$$y_j = \mathcal{G}_j(\sigma) + \eta_j$$

where $\mathcal{G}_j(\sigma) = R(\sigma)I^{(j)}$. Concatenating these observations, we write

$$y = \mathcal{G}(\sigma) + \eta, \quad \eta \sim N(0, \Gamma)$$

where $\Gamma = \text{diag}(\Gamma_0, \ldots, \Gamma_0)$. The inverse problem is then to recover the conductivity field σ from the data y. This problem is highly ill-posed: the data is finite dimensional, yet we wish to recover a function which, typically, lies in an infinite dimensional space. We take a Bayesian approach by placing a prior distribution on σ. The choice of prior may have significant effect on the resulting posterior distribution, and different choices of prior may be more appropriate depending upon the prior knowledge of the particular experimental set-up under consideration.

In subsection 3.3.1 we outline three different families of prior models, and show the appropriate regularity of the forward maps arising from them. In subsection 3.3.2 we describe the likelihood and posterior distribution formally, before rigorously proving that the posterior distribution exists and is Lipschitz with respect to the data in the Hellinger metric.

3.3.1 Choices of Prior

In this section we consider three priors, labelled by $i = 1, 2, 3$, defined by functions $F_i : X_i \to \mathcal{A}(D)$ which map draws from prior measures on the Banach spaces X_i to the space of conductivities $\mathcal{A}(D)$. Our prior conductivity distributions will then be the pushfoward of the prior measures by these maps F_i. We describe these maps, and establish continuity properties of them needed for the study of the posterior later.

Log-Gaussian prior

We first consider the simple case that the coefficient is given by the exponential of a continuous function. Let $F_1 : C^0(\overline{D}) \to \mathcal{A}(D)$ be defined by $F_1(u) = \exp(u)$. Then it is easily seen that F_1 does indeed map into $\mathcal{A}(D)$. Furthermore, since D is

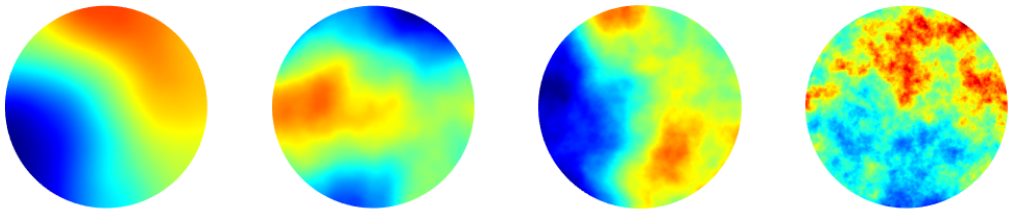


Figure 3.2: Example draws from log-Gaussian priors defined in Example 3.3.1.

bounded, if $u \in C^0(\overline{D})$ and $(u_\varepsilon)_{\varepsilon>0} \subseteq C^0(\overline{D})$ is a sequence such that $\|u_\varepsilon - u\|_\infty \to 0$, then $\|F_1(u_\varepsilon) - F_1(u)\|_\infty \to 0$.

In this case, we will take our prior measure μ_0 on u to be a Gaussian measure $N(m_0, \mathcal{C}_0)$ on $C^0(\overline{D})$. Note that the push forward of a Gaussian measure by F_1 is a log-Gaussian measure.

Example 3.3.1. *Consider the case $D = B(0,1) \subseteq \mathbb{R}^2$. Suppose that u is drawn from a Gaussian measure $\mu_0 = N(0, \mathcal{C})$. Typical samples from $F_1^{\#}(\mu_0)$ are shown in Figure 3.2*[2]*. The covariance $\mathcal{C}$ is chosen such that the samples u almost surely have regularity $u \in H^s(D)$ for all $s < t$, where from left to right $t = 2, 1.5, 1, 0.5$ respectively. Here the samples are generated on $[-1,1]^2 \supseteq D$ and then restricted to D, for computational simplicity.*

Star-shaped prior

We now consider star-shaped inclusions, that is, inclusions parametrized by their centre and a radial function. These were studied in two-dimensions in the paper [22] to parametrize domains for a Bayesian inverse shape scattering problem. In [22] the authors prove well-posedness of the inverse problem in an infinite dimensional setting through the use of shape derivatives and Riesz-Fredholm theory.

Let $D \subseteq \mathbb{R}^d$, and $R_{d-1} = (-\pi, \pi] \times [0, \pi]^{d-2} \subseteq \mathbb{R}^{d-1}$. Let $h : \mathbb{R}^d \to R_{d-1}$ be the continuous function representing the mapping from Cartesian to angular polar coordinates. Define the mapping $A : C_P^0(R_{d-1}) \times D \to \mathcal{B}(D)$ by

$$A(r, x_0) = \left\{ x \in D \;\middle|\; |x - x_0| \leq r(h(x - x_0)) \right\}$$

[2]Given a measure μ on $(X, \mathcal{X})$ and a measurable map $F : (X, \mathcal{X}) \to (Y, \mathcal{Y})$ between measurable spaces, $F^{\#}(\mu)$ denotes the pushforward of μ by F, i.e. the measure on $(Y, \mathcal{Y})$ given by $F^{\#}(\mu)(A) = \mu(F^{-1}(A))$ for all $A \in \mathcal{Y}$.

where $C_P^0(R_{d-1})$ is the space of continuous periodic functions on R_{d-1}. Then $A(r, x_0)$ describes the set of points in D which lie within the closed surface parametrized in polar coordinates centred at x_0 by

$$\Gamma(\Theta) = (\Theta, r(\Theta)), \quad \Theta \in R_{d-1}.$$

In two dimensions, we have $R_1 = (-\pi, \pi]$ and the mapping $h : \mathbb{R}^2 \to R_1$ is given by

$$h(x, y) = \operatorname{atan2}(y, x) \equiv 2 \arctan\left(\frac{y}{\sqrt{x^2 + y^2} + x}\right)$$

where atan2 is the two-parameter inverse tangent function.

In three dimensions, we have $R_2 = (-\pi, \pi] \times [0, \pi]$ and the mapping $h : \mathbb{R}^3 \to R_2$ is given by

$$h(x, y, z) = \left(\operatorname{atan2}(y, x), \operatorname{arccot}\left(\frac{z}{\sqrt{x^2 + y^2}}\right)\right).$$

Similar expressions for h exist in higher dimensions, though for applications we are only interested in the case $d = 2, 3$.

Define now the map $F_2 : C_P^0(R_{d-1}) \times D \to \mathcal{A}(D)$ by

$$\begin{aligned} F_2(r, x_0) &= u_+ \mathbb{1}_{A(r,x_0)} + u_- \mathbb{1}_{D \setminus A(r,x_0)} \\ &= (u_+ - u_-)\mathbb{1}_{A(r,x_0)} + u_-. \end{aligned}$$

Again it can easily be seen that F_2 does indeed map into $\mathcal{A}(D)$. We claim that this map is continuous in the following sense:

Proposition 3.3.2. *Define the map $F_2 : C_P^0(R_{d-1}) \times D \to \mathcal{A}(D)$ as above. Let $x_0 \in D$ and let $r \in C_P^0(R_{d-1})$ be Lipschitz continuous.*

(i) Suppose that $(r_\varepsilon)_{\varepsilon>0} \subseteq C_P^0(R_{d-1})$ is a sequence of functions such that $\|r_\varepsilon - r\|_\infty \to 0$. Then $F_2(r_\varepsilon, x_0) \to F_2(r, x_0)$ in measure[3].

(ii) Suppose that $(x_0^\varepsilon)_{\varepsilon>0} \subseteq D$ is a sequence of points such that $|x_0^\varepsilon - x_0| \to 0$. Then $F_2(r, x_0^\varepsilon) \to F_2(r, x_0)$ in measure.

(iii) Let $(r_\varepsilon)_{\varepsilon>0}$, $(x_0^\varepsilon)_{\varepsilon>0}$ be as above. Then $F_2(r_\varepsilon, x_0^\varepsilon) \to F_2(r, x_0)$ in measure.

[3]A sequence of functions $(f_\varepsilon)_{\varepsilon>0}$, $f_\varepsilon : D \to \mathbb{R}$, is said to converge in measure to a function $f : D \to \mathbb{R}$ if for all $\delta > 0$, $|\{x \in D \mid |f_\varepsilon(x) - f(x)| > \delta\}| \to 0$. Here $|B|$ denotes the Lebesgue measure of a set $B \subseteq \mathbb{R}^d$.

Proof. In order to show that a sequence of functions $f_\varepsilon : D \to \mathbb{R}$ converges to $f : D \to \mathbb{R}$ in measure, it suffices to show that there exists a sequence of sets $Z_\varepsilon \subseteq D$ with $|Z_\varepsilon| \to 0$ such that $|f_\varepsilon - f| \leq C\mathbb{1}_{Z_\varepsilon}$. Then for each $\delta > 0$ we have

$$|\{x \in D \,|\, |f_\varepsilon(x) - f(x)| > \delta\}| \leq |\{x \in D \,|\, |f_\varepsilon(x) - f(x)| \neq 0\}| \leq |Z_\varepsilon| \to 0.$$

(i) Fix the centre $x_0 \in D$. Denote $A(r) = A(r, x_0)$. Let $r \in C^0_P(R_{d-1})$ and let $(r_\varepsilon)_{\varepsilon>0} \subseteq C^0_P(R_{d-1})$ be a sequence of functions such that $\|r_\varepsilon - r\|_\infty \to 0$. Then there exists $\gamma(\varepsilon) \to 0$ such that $\|r_\varepsilon - r\|_\infty < \gamma(\varepsilon)$. By definition we then have

$$r(x) - \gamma(\varepsilon) \leq r_\varepsilon(x) \leq r(x) + \gamma(\varepsilon) \quad \text{for all } x \in D \text{ and } \varepsilon > 0.$$

It follows that we have the inclusions

$$A(r - \gamma(\varepsilon)) \subseteq A(r_\varepsilon) \subseteq A(r + \gamma(\varepsilon)),$$
$$A(r - \gamma(\varepsilon)) \subseteq A(r) \subseteq A(r + \gamma(\varepsilon)).$$

Let Δ denote the symmetric difference. We deduce that

$$A(r_\varepsilon)\Delta A(r) \subseteq A(r + \gamma(\varepsilon)) \setminus A(r - \gamma(\varepsilon)).$$

Now the right-hand side is given by

$$A(r + \gamma(\varepsilon)) \setminus A(r - \gamma(\varepsilon))$$
$$= \left\{x \in D \;\middle|\; r(h(x - x_0)) - \gamma(\varepsilon) < |x - x_0| \leq r(h(x - x_0)) + \gamma(\varepsilon)\right\}.$$

As $\varepsilon \to 0$, this set decreases to the boundary set

$$\partial A(r) = \left\{x \in D \;\middle|\; |x - x_0| = r(h(x - x_0))\right\}.$$

Since the graph of a continuous function has Lebesgue measure zero, we deduce that $|\partial A(r)| = 0$. It follows that

$$\lim_{\varepsilon \to 0} |A(r_\varepsilon)\Delta A(r)| = 0.$$

To conclude, note that

$$|F_2(r_\varepsilon, x_0) - F_2(r, x_0)| \leq |u_+ - u_-||\mathbb{1}_{A(r_\varepsilon)} - \mathbb{1}_{A(r_\varepsilon)}| = C\mathbb{1}_{A(r_\varepsilon)\Delta A(r)}.$$

(ii) Let $r \in C_P^0(R_{d-1})$ be Lipschitz continuous. Denote $A(x_0) = A(r, x_0)$. Let $(x_0^\varepsilon) \subseteq D$ be a sequence of points such that $|x_0^\varepsilon - x_0| \to 0$. Note that we may write

$$\begin{aligned} A(x_0^\varepsilon) &= \{x \in D \mid |x - x_0^\varepsilon| \leq r(h(x - x_0^\varepsilon))\} \\ &= \{x \in \mathbb{R}^d \mid |x - x_0^\varepsilon| \leq r(h(x - x_0^\varepsilon))\} \cap D \\ &= ((x_0^\varepsilon - x_0) + \{x \in \mathbb{R}^d \mid |x - x_0| \leq r(h(x - x_0))\}) \cap D \\ &=: ((x_0^\varepsilon - x_0) + A(x_0)^*) \cap D. \end{aligned}$$

By the distributivity of intersection over symmetric difference, we then have that

$$\begin{aligned} A(x_0^\varepsilon)\Delta A(x_0) &= [((x_0^\varepsilon - x_0) + A(x_0)^*) \cap D]\Delta[A(x_0)^* \cap D] \\ &= [((x_0^\varepsilon - x_0) + A(x_0)^*)\Delta A(x_0)^*] \cap D \\ &\subseteq ((x_0^\varepsilon - x_0) + A(x_0)^*)\Delta A(x_0)^*. \end{aligned}$$

Therefore, using Theorem 1 from [126], we see that

$$\begin{aligned} |A(x_0^\varepsilon)\Delta A(x_0)| &\leq |((x_0^\varepsilon - x_0) + A(x_0)^*)\Delta A(x_0)^*| \\ &\leq |x_0^\varepsilon - x_0|\mathcal{H}^{d-1}(\partial A(x_0)^*) \end{aligned}$$

where $\mathcal{H}^{d-1}$ is the $(d-1)$-dimensional Hausdorff measure. Since we assume that r is Lipschitz, the surface area $\mathcal{H}^{d-1}(\partial A(x_0)^*)$ of the boundary of $A(x_0)^*$ is finite, and so it follows that

$$\lim_{\varepsilon \to 0} |A(x_0^\varepsilon)\Delta A(x_0)| = 0.$$

As before, we conclude by noting that

$$|F_2(r, x_0^\varepsilon) - F_2(r, x_0)| \leq |u_+ - u_-||\mathbb{1}_{A(x_0^\varepsilon)} - \mathbb{1}_{A(x_0)}| = C\mathbb{1}_{A(x_0^\varepsilon)\Delta A(x_0)}.$$

(iii) We have that

$$\begin{aligned} |F_2(r_\varepsilon, x_0^\varepsilon) - F_2(r, x_0)| &\leq |F_2(r_\varepsilon, x_0^\varepsilon) - F_2(r, x_0^\varepsilon)| + |F_2(r, x_0^\varepsilon) - F_2(r, x_0)| \\ &\leq C(\mathbb{1}_{A(r_\varepsilon, x_0^\varepsilon)\Delta A(r, x_0^\varepsilon)} + \mathbb{1}_{A(r, x_0^\varepsilon)\Delta A(r, x_0)}) \\ &\leq C\mathbb{1}_{[A(r_\varepsilon, x_0^\varepsilon)\Delta A(r, x_0^\varepsilon)] \cup [A(r, x_0^\varepsilon)\Delta A(r, x_0)]}. \end{aligned}$$

Now note that

$$|A(r_\varepsilon, y_0)\Delta A(r, y_0)| \leq |A(r_\varepsilon, y_0)^* \Delta A(r, y_0)^*|.$$

The right hand-side is independent of y_0 by translation invariance of the Lebesgue measure. By the same argument as part (i) we conclude that it tends to zero. We then have that

$$\begin{aligned} |[A(r_\varepsilon, x_0^\varepsilon)\Delta A(r, x_0^\varepsilon)] \cup [A(r, x_0^\varepsilon)\Delta A(r, x_0)]| & \\ \leq |A(r_\varepsilon, x_0^\varepsilon)\Delta A(r, x_0^\varepsilon)| + |A(r, x_0^\varepsilon)\Delta A(r, x_0)| & \\ \leq \sup_{y_0 \in D} |A(r_\varepsilon, y_0)\Delta A(r, y_0)| + |A(r, x_0^\varepsilon)\Delta A(r, x_0)| & \end{aligned}$$

which tends to zero by the discussion above and part (ii).

□

Remark 3.3.3. *Above we assumed that $r : R_{d-1} \to \mathbb{R}$ was Lipschitz continuous. This assumption is only used in the proof of part (ii) of the proposition. If the centre of the star-shaped region is known, this assumption may then be dropped to allow for rougher boundaries.*

We need to choose a prior measure μ_0 on (r, x_0). We assume that r and x_0 are independent under the prior so that we may factor $\mu_0 = \sigma_0 \otimes \tau_0$ where σ_0 is a measure on $C_P^0(R_{d-1})$ and τ_0 is a measure on D. We will assume that σ_0 is such that $\sigma_0(B) > 0$ for all balls $B \subseteq C_P^0(R_{d-1})$. We equip $C_P^0(R_{d-1}) \times D$ with the norm $\|(r, x_0)\| = \max\{\|r\|_\infty, |x_0|\}$.

Example 3.3.4. *Consider the case $D = B(0,1) \subseteq \mathbb{R}^2$. Suppose that r is drawn from a log-Gaussian measure σ_0 on $C_P^0((-\pi, \pi])$, and x_0 is drawn from $\tau_0 = U([-0.5, 0.5]^2)$. Note that $[-0.5, 0.5]^2 \subseteq B(0,1)$. Typical samples from $F_2^\#(\mu_0)$ are shown in Figure 3.3. The covariance of σ_0 is chosen such that the samples r almost surely have regularity $r \in H^s((-\pi, \pi])$ for all $s < t$, where from left to right $t = 2.5, 2, 1.5, 1$ respectively.*

Level set prior

We finally consider the case where the inclusions can be described by a single level set function, as in [72]. A discussion of the use of level set functions in geometric inversion is provided in the introduction of the next chapter. Let $n \in \mathbb{N}$ and fix

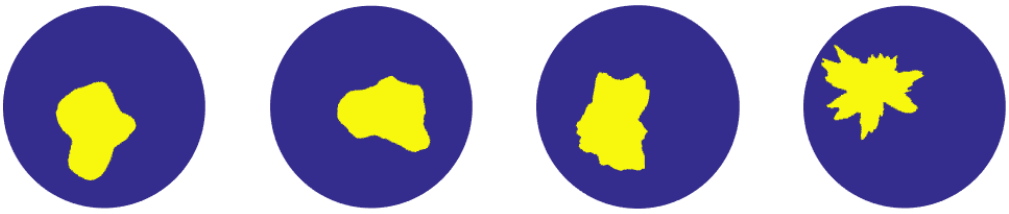


Figure 3.3: Example draws from star-shaped priors defined in Example 3.3.4.

constants $-\infty = c_0 < c_1 < \ldots < c_n = \infty$. Given $u : D \to \mathbb{R}$, define $D_i \subseteq D$ by

$$D_i = \{x \in D \,|\, c_{i-1} \leq u(x) < c_i\}, \quad i = 1, \ldots, n$$

so that $\overline{D} = \bigcup_{i=1}^{n} \overline{D}_i$ and $D_i \cap D_j = \varnothing$ for $i \neq j$, $i, j \geq 1$. Define also the level sets

$$D_i^0 = \overline{D}_i \cap \overline{D}_{i+1} = \{x \in D \,|\, u(x) = c_i\}, \quad i = 1, \ldots, n-1.$$

Now given strictly positive functions $f_1, \ldots, f_n \in C^0(\overline{D})$, we define the map $F_3 : C^0(\overline{D}) \to \mathcal{A}(D)$ by

$$F_3(u) = \sum_{i=1}^{n} f_i \mathbb{1}_{D_i}.$$

Since each f is continuous and strictly positive on a compact set $\overline{D}$, they are uniformly bounded above and below by positive constants, and so F_3 does indeed map into $\mathcal{A}(D)$.

In this chapter we are primarily concerned with the case of binary fields, $n = 2$ and f_i constant above, however the theory in proved in the general case. We have the following result regarding continuity of this map, by the same arguments as in [72].

Proposition 3.3.5. *Define the map $F_3 : C^0(\overline{D}) \to \mathcal{A}(D)$ as above. Let $u \in C^0(\overline{D})$ be such that $|D_i^0| = 0$ for $i = 1, \ldots, n-1$. Suppose that $(u_\varepsilon)_{\varepsilon > 0} \subseteq C^0(\overline{D})$ is an approximating sequence of functions so that $\|u_\varepsilon - u\|_\infty \to 0$. Then $F_3(u_\varepsilon) \to F_3(u)$ in measure.*

Proof. Denote by $D_{i,\varepsilon}$ and $D_{i,\varepsilon}^0$ the sets as defined above associated with the ap-

proximating functions u_ε. We can write

$$F_3(u_\varepsilon) - F_3(u) = \sum_{i=1}^{n}\sum_{j=1}^{n}(f_i - f_j)\mathbb{1}_{D_{i,\varepsilon}\cap D_j} = \sum_{\substack{i,j=1\\ i\neq j}}^{n}(f_i - f_j)\mathbb{1}_{D_{i,\varepsilon}\cap D_j}.$$

Since $\|u_\varepsilon - u\|_\infty \to 0$, there exists $\gamma(\varepsilon) \to 0$ with $\|u_\varepsilon - u\|_\infty < \gamma(\varepsilon)$. Then we have for all $x \in D$ and $\varepsilon > 0$

$$u(x) - \gamma(\varepsilon) < u_\varepsilon(x) < u(x) + \gamma(\varepsilon).$$

Hence for $|j - i| > 1$ and ε sufficiently small, $D_{i,\varepsilon} \cap D_j = \varnothing$. If $|j - i| = 1$, then

$$\begin{aligned} D_{i,\varepsilon} \cap D_{i+1} &\subseteq \widetilde{D}_{i,\varepsilon} := \{x \in D \,|\, c_i \leq u(x) < c_i + \gamma(\varepsilon)\} \to D_i^0, \\ D_{i,\varepsilon} \cap D_{i-1} &\subseteq \hat{D}_{i-1,\varepsilon} := \{x \in D \,|\, c_i - \gamma(\varepsilon) \leq u(x) < c_i\} \to \varnothing. \end{aligned}$$

By the uniform boundedness of the (f_i), for sufficiently small ε we can then write

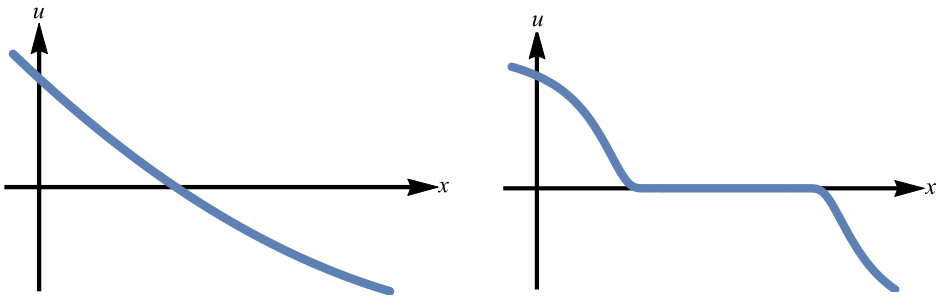
$$\begin{aligned} |F_3(u_\varepsilon) - F_3(u)| &\leq \sum_{i=1}^{n-1}|f_i - f_{i+1}|\mathbb{1}_{\widetilde{D}_{i,\varepsilon}} + \sum_{i=2}^{n}|f_i - f_{i-1}|\mathbb{1}_{\hat{D}_{i-1,\varepsilon}} \\ &\leq C\mathbb{1}_{Z_\varepsilon} \end{aligned} \tag{3.3.1}$$

where $Z_\varepsilon \subseteq D$ is given by

$$Z_\varepsilon = \left(\bigcup_{i=1}^{n-1}\widetilde{D}_{i,\varepsilon}\right) \cup \left(\bigcup_{i=2}^{n}\hat{D}_{i-1,\varepsilon}\right) \to \bigcup_{i=1}^{n-1} D_i^0.$$

By the assumption that $|D_i^0| = 0$ for all i, it follows that $|Z_\varepsilon| \to 0$, and so the result follows from the comment at the start of the proof of Proposition 3.3.2. □

Note that bound (3.3.1) actually above implies the slightly stronger result that, when the c_i-level sets of $u \in X$ have zero measure, F_3 is continuous into $L^p(D)$, $1 \leq p < \infty$, at u. The assumption that the level sets have zero measure is an important one, as illustrated by Figure 3.4: an arbitrarily small perturbation of u can lead to an order 1 change in $F_3(u)$. To see this, note that function u in the right hand figure of Figure 3.4 takes the value zero on a set of positive measure. If F_3 thresholds u at zero, then the value of $F_3(u)$ on this set changes significantly as the value of u on this set is perturbed above or below zero; such a perturbation can be arbitrarily small in L^p.



$F_3(\cdot)$ is continuous at u $F_3(\cdot)$ is discontinuous at u

Figure 3.4: The discontinuity of F_3 into $L^p(D)$.

.

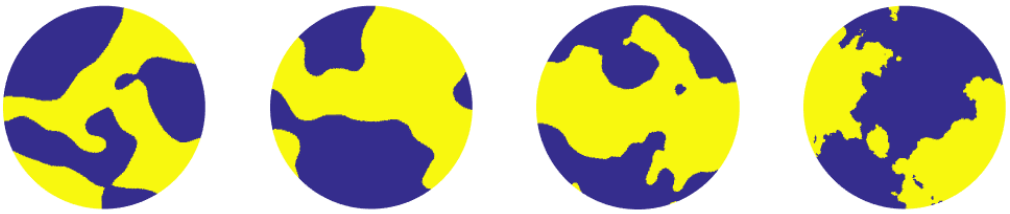


Figure 3.5: Example draws from level set priors defined in Example 3.3.6.

In the Bayesian approach we are taking to this problem, we may choose a prior measure on u such that, almost surely, the Lebesgue measure of the level sets is zero. This is shown to hold for Gaussian measures in [72]. As a result, F_3 will be almost surely continuous under the prior, and this is enough to give the measurability required in Bayes' theorem, as shown in [72].

As in the log-Gaussian case, we take our prior measure μ_0 on u to be a Gaussian measure $N(m_0, \mathcal{C}_0)$ on $C^0(\overline{D})$.

Example 3.3.6. *Consider the case $D = B(0,1) \subseteq \mathbb{R}^2$, $n = 2$, $c_1 = 0$, $f_1 \equiv 1$ and $f_2 \equiv 2$. Suppose that u is drawn from a centred Gaussian measure $\mu_0 = N(0, \mathcal{C})$ on $C^0(\overline{D})$. Typical samples from $F_3^{\#}(\mu_0)$ are show in Figure 3.5. The covariance $\mathcal{C}$ is chosen such that the samples u almost surely have regularity $u \in H^s(D)$ for all $s < t$, where from left to right $t = 4, 3, 2, 1$ respectively. As in the log-Gaussian case, here the samples are generated on $[-1,1]^2 \supseteq D$ and then restricted to D, for computational simplicity.*

3.3.2 The Likelihood and Posterior Distribution

The inverse problem was introduced at the beginning of the section. Now that we have introduced prior distributions, we may provide the Bayesian formulation of the problem.

Let X be a separable Banach space and $F : X \to \mathcal{A}(D)$ a map from the space X where the unknown parameters live to the conductivity space. Choose a set of current stimulation patterns $I^{(j)} \in \mathbb{R}^L$, $j = 1, \dots, J$ and let $\mathcal{M}_j : \mathcal{A}(D) \to \mathbb{H}$ denote the solution map when using stimulation pattern $I^{(j)}$. Recall the projection map $\Pi : \mathbb{H} \to \mathbb{R}^L$ was defined by $\Pi(v, V) = V$.

The data y_j from the jth stimulation pattern is assumed to arise from the map $\mathcal{G}_j : X \to \mathbb{R}^L$, $\mathcal{G}_j = \Pi \circ \mathcal{M}_j \circ F$, via

$$y_j = \mathcal{G}_j(u) + \eta_j, \quad \eta_j \sim N(0, \Gamma_0) \text{ iid.}$$

We concatenate these observations to get data $y \in \mathbb{R}^{JL}$ given by

$$y = \mathcal{G}(u) + \eta, \quad \eta \sim \mathbb{Q}_0 := N(0, \Gamma)$$

where $\Gamma = \operatorname{diag}(\Gamma_0, \dots, \Gamma_0)$ and $\mathcal{G} : X \to \mathbb{R}^{JL}$. This coincides with the setup at the start of the section, with $\sigma = F(u)$.

Assume that $u \sim \mu_0$, where μ_0 is independent of $\mathbb{Q}_0$. From the above, we see that $y|u \sim \mathbb{Q}_u := N(\mathcal{G}(u), \Gamma)$. We use this to find the distribution of $u|y$. First note that

$$\frac{\mathrm{d}\mathbb{Q}_u}{\mathrm{d}\mathbb{Q}_0}(y) = \exp\left(-\Phi(u; y) + \frac{1}{2}|y|_\Gamma^2\right)$$

where the potential (or negative log-likelihood) $\Phi : X \times Y \to \mathbb{R}$ is given by

$$\Phi(u; y) = \frac{1}{2}|\mathcal{G}(u) - y|_\Gamma^2. \tag{3.3.2}$$

Then under suitable regularity conditions, Bayes' theorem tells us that the distribution μ^y of $u|y$ is as given below:

Theorem 3.3.7 (Existence and Well-Posedness)**.** *Let $(X, \mathcal{F}, \mu_0)$ denote any of the probability spaces associated with any of the three priors introduced in the previous subsection, and let $\Phi : X \times Y \to \mathbb{R}$ be the potential (4.2.9) associated with the corresponding forward map. Then the posterior distribution μ^y of the state u given*

data y is well-defined. Furthermore, $\mu^y \ll \mu_0$ with Radon-Nikodym derivative

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(u) = \frac{1}{Z_\mu}\exp(-\Phi(u;y)) \tag{3.3.3}$$

where for y $\mathbb{Q}_0$-a.s.,

$$Z_\mu := \int_X \exp(-\Phi(u;y))\,\mu_0(\mathrm{d}u) > 0.$$

Additionally, the posterior measure μ^y is locally Lipschitz with respect to y, in the Hellinger distance: for all $y, y' \in Y$ with $\max\{|y|_\Gamma, |y'|_\Gamma\} < \rho$, there exists $C = C(\rho) > 0$ such that

$$d_{\mathrm{Hell}}(\mu^y, \mu^{y'}) \leq C|y - y'|_\Gamma.$$

In the proof of the above theorem we will make use of the following version of Bayes' theorem from [39].

Proposition 3.3.8 (Bayes' theorem)**.** *Define the measure $\nu_0(\mathrm{d}u, \mathrm{d}y) = \mu_0(\mathrm{d}u)\mathbb{Q}_0(\mathrm{d}y)$ on $X \times Y$. Assume that $\Phi : X \times Y \to \mathbb{R}$ is ν_0-measurable and that, for y $\mathbb{Q}$-a.s.*

$$Z_\mu = \int_X \exp(-\Phi(u;y))\,\mu_0(\mathrm{d}u) > 0.$$

Then the conditional distribution of $u|y$ exists and is denoted by μ^y. Furthermore $\mu^y \ll \mu_0$ and, for y $\mathbb{Q}_0$-a.s.,

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0} = \frac{1}{Z_\mu}\exp\left(-\Phi(u;y)\right).$$

We need to verify that the assumptions of this theorem are satisfied. To proceed we first give some regularity properties of the potential Φ:

Proposition 3.3.9. *Let $(X, \mathcal{F}, \mu_0)$ denote any of the probability spaces associated with the priors introduced in the previous subsection. Then the potential $\Phi : X \times Y \to \mathbb{R}$ associated with the corresponding forward map, given by (4.2.9), admits the following properties.*

(i) There is a continuous $K : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ such that for every $\rho > 0$, $u \in X$ and $y \in Y$ with $|y|_\Gamma < \rho$,

$$0 \leq \Phi(u;y) \leq K(\rho, \|u\|_X).$$

In the cases $F = F_2$ and $F = F_3$, K has no dependence on $\|u\|_X$.

(ii) For any fixed $y \in Y$, $\Phi(\cdot; y) : X \to \mathbb{R}$ is continuous μ_0-almost surely on the probability space $(X, \mathcal{F}, \mu_0)$.

(iii) There exists $C : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ such that for every $y_1, y_2 \in Y$ with $\max\{|y_1|_\Gamma, |y_2|_\Gamma\} < \rho$, and every $u \in X$,

$$|\Phi(u; y_1) - \Phi(u; y_2)| \leq C(\rho, \|u\|_X)|y_1 - y_2|_\Gamma.$$

Moreover, $C(\rho, \|\cdot\|_X) \in L^2_{\mu_0}(X)$ for all $\rho > 0$.

Proof. (i) From equation (3.2.5) in the proof of Proposition 3.2.7, we see that there exists $C > 0$ such that each $\mathcal{M}_j : \mathcal{A}(D) \to \mathbb{H}$ satisfies

$$\|\mathcal{M}_j(\sigma_1) - \mathcal{M}_j(\sigma_2)\|_* \leq C\|\mathcal{M}_j(\sigma_2)\|_*\|\sigma_1 - \sigma_2\|_\infty$$

for all $\sigma_1, \sigma_2 \in \mathcal{A}(D)$. Taking $\sigma_2 \equiv 1$, say, we deduce that

$$\|\mathcal{M}_j(\sigma_1)\|_* \leq C\|\mathcal{M}_j(1)\|_*\|\sigma_1 - 1\|_\infty + \|\mathcal{M}_j(1)\|_* \leq C(1 + \|\sigma_1\|_\infty).$$

Hence $\|\sigma\|_\infty < \rho$ implies that $\|\mathcal{M}_j(\sigma)\|_* < C(1 + \rho)$. By Corollary 3.2.8, it follows that $\Pi \circ \mathcal{M}_j : \mathcal{A}(D) \to \mathbb{R}^L$ is bounded on bounded sets with respect to $\|\cdot\|_\infty$ for all j.

In the case $F = F_1$, if $u \in X$ then $\|F(u)\|_\infty \leq e^{\|u\|_X}$. It follows that $|\mathcal{G}(u)|_\Gamma \leq \max_j |\mathcal{G}_j(u)|_\Gamma \leq C(1 + e^{\|u\|_X})$.

Now note that

$$\Phi(u; y) \leq |\mathcal{G}(u)|^2_\Gamma + |y|^2_\Gamma.$$

Then for any $y \in Y$ with $|y| < \rho$, we may bound

$$\Phi(u; y) \leq C(1 + e^{2\|u\|_X} + \rho^2) =: K(\rho, \|u\|_X).$$

In the cases $F = F_2$ and $F = F_3$, we have that $\|F(u)\|_\infty$ is bounded uniformly over $u \in X$ and so $|\mathcal{G}(u)|_\Gamma \leq \max_j |\mathcal{G}_j(u)|_\Gamma \leq C$. Hence we obtain the bound

$$\Phi(u; y) \leq C(1 + \rho^2) =: K(\rho).$$

(ii) Let $u \sim \mu_0$ and suppose $F : X \to \mathcal{A}(D)$ is such that $\|u_\varepsilon - u\|_X \to 0$ implies that $F(u_\varepsilon) \to F(u)$ either uniformly or in measure. Then Proposition 3.2.7

tells us that $\mathcal{M}_j \circ F : X \to \mathbb{H}$ is continuous at u for each j. The projection $\Pi : \mathbb{H} \to \mathbb{R}^L$ is continuous, and so $\mathcal{G}_j = \Pi \circ \mathcal{M}_j \circ F$ is continuous at u for each j. In §3.3.1 it is shown that this is true for $F = F_1$ and $F = F_2$ for any u. For $F = F_3$ it is only true at points u whose level sets have zero measure, however since we are assuming $u \sim \mu_0$, a Gaussian measure, it follows from Proposition 7.2 in [72] that u μ_0-almost surely has this property.

(iii) Let $u \in X$ and $y_1, y_2 \in Y$ with $\max\{|y_1|_\Gamma, |y_2|_\Gamma\} < \rho$. Then we have

$$\begin{aligned}
|\Phi(u; y_1) - \Phi(u; y_2)| &= \frac{1}{2}|\langle y_1 + y_2 - 2\mathcal{G}(u), y_1 - y_2\rangle_\Gamma| \\
&\leq \frac{1}{2}(|y_1|_\Gamma + |y_2|_\Gamma + 2|\mathcal{G}(u)|_\Gamma)|y_1 - y_2|_\Gamma \\
&\leq (\rho + |\mathcal{G}(u)|)|y_1 - y_2|_\Gamma \\
&=: C(\rho, \|u\|_X)|y_1 - y_2|_\Gamma
\end{aligned}$$

For the square-integrability we consider cases separately based on the prior. In the log-Gaussian case, we may bound

$$C(\rho, \|u\|_X) \leq C(1 + \rho^2 + e^{\|u\|_X})$$

using the bound from the proof of part (i), and so square-integrability follows since Gaussians have exponential moments.

In the star-shaped and level set prior cases, we have that $|\mathcal{G}(u)|$ is bounded uniformly by a constant. We may hence bound $C(\rho, \|u\|_X)$ above independently of u, and so again the square-integrability follows.

□

Proof of Theorem 3.3.7. Define the product measure $\nu_0(\mathrm{d}u, \mathrm{d}y) = \mu_0(\mathrm{d}u)\mathbb{Q}_0(\mathrm{d}y)$ on $X \times Y$. We showed in Proposition 3.3.9 that $\Phi(\cdot; y) : X \to \mathbb{R}$ is almost-surely continuous under the prior for all $y \in Y$, and $\Phi(u; \cdot) : Y \to \mathbb{R}$ is locally Lipschitz for all $u \in X$. Together these imply that $\Phi : X \times Y \to \mathbb{R}$ is almost-surely jointly continuous under ν_0. To see this, let $(u, y) \in X \times Y$ and let $(u_n, y_n)_{n\geq 1} \subseteq X \times Y$ be an approximating sequence so that $\|u_n - u\|_X \to 0$ and $|y_n - y|_\Gamma \to 0$. Then we have

$$|\Phi(u_n, y_n) - \Phi(u, y)| \leq |\Phi(u_n, y_n) - \Phi(u_n, y)| + |\Phi(u_n, y) - \Phi(u, y)|.$$

The second term tends to zero μ_0-almost surely by continuity. For the first term,

note that the sequences $(\|u_n\|_X)_{n\geq 1}$ and $(|y_n|_\Gamma)_{n\geq 1}$ are bounded, by K and R respectively, say. Then we can use the local Lipschitz property to deduce that

$$|\Phi(u_n, y_n) - \Phi(u_n, y)| \leq C(R, K)|y_n - y|_\Gamma$$

since $C(\cdot,\cdot) : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ is monotonically increasing in both components. Therefore this term tends to zero, and we obtain the desired continuity. It follows, see for example Lemma 6.1 in [72], that Φ is ν_0-measurable.

Now by Proposition 3.3.9(i), we may bound $\exp(-\Phi(u; y))$ by 1, and so $Z_\mu \leq 1$. For the lower bound, we consider cases separately based on the prior. First we consider the log-Gaussian and level set prior cases so that μ_0 is Gaussian. Let $B \subseteq X$ be any ball. Fix any $\rho > |y|_\Gamma$ and define

$$R = \sup_{u\in B} K(\rho, \|u\|_X)$$

where K is the upper bound from Proposition 3.3.9(i). This supremum is finite by the continuity of K. Then we have

$$\begin{aligned}\int_X \exp(-\Phi(u; y))\,\mu_0(\mathrm{d}u) &\geq \int_B \exp(-\Phi(u; y))\,\mu_0(\mathrm{d}u)\\ &\geq \int_B \exp(-K(\rho, \|u\|))\,\mu_0(\mathrm{d}u)\\ &\geq \exp(-R)\mu_0(B).\end{aligned}$$

Since μ_0 is Gaussian, $\mu_0(B) > 0$ and so $Z_\mu > 0$.

In the star-shaped prior case, proceed as above but take $B = B_1 \times D$ where $B_1 \subseteq C^0_P(R_{d-1})$ is any ball. Then we have

$$\mu_0(B) = (\sigma_0 \times \tau_0)(B_1 \times D) = \sigma_0(B_1)\tau_0(D) > 0$$

by the assumption that σ_0 assigns positive mass to balls, and so again $Z_\mu > 0$. The above hold for all $y \in Y$, and so in particular for y $\mathbb{Q}_0$-almost-surely. We may now apply Bayes' Theorem 3.3.8 to obtain the existence of μ^y.

The proof of well-posedness is almost identical to that of the analogous result Theorem 2.2 in [72] and is hence omitted. □

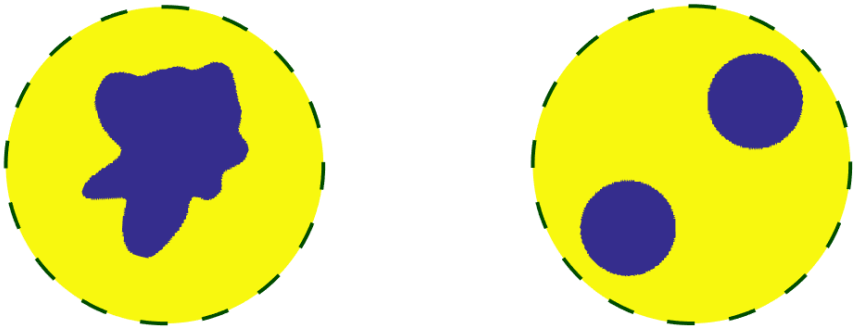


Figure 3.6: The two binary fields we attempt to recover. Conductivity A is drawn from the star-shaped prior, with $\sigma_0 = h^{\#}\left[N(0.5, 10^9 \cdot (30^2 - \mathcal{A}_D)^{-3})\right]$, $h(z) = (1 + \tanh z)/2$, and $\tau_0 = U([-0.5, 0.5]^2)$. Conductivity B is constructed explicitly, rather than being drawn from a prior.

3.4 Numerical Experiments

We investigate the effect of the choice of prior on the recovery of certain binary conductivity fields. The specific fields we consider are shown in Figure 3.6, where blue represents a conductivity of 1 and yellow a conductivity of 2. Simulations are performed using the EIDORS software [2] to solve the forward model; a mesh of 43264 elements is used to create the data and a mesh of 10816 elements is used for simulations in order to avoid an inverse crime [75].

In subsection 3.4.1 we describe the MCMC sampling algorithm that we will use. In subsection 3.4.2 we define the parameters we will use for the forward model and the MCMC simulations. We also describe how the data is created, and define our choices of prior distributions. Finally in subsection 3.4.3 we present the results of the simulation, looking at quality of reconstruction, convergence of the algorithm and some properties of the posterior distribution.

3.4.1 Sampling Algorithm

We aim to produce a sequence of samples from μ^y on X, where μ^y is given by (3.3.3). We make use of the preconditioned Crank-Nicolson Markov Chain Monte Carlo (pCN-MCMC) method. The pCN-MCMC method is a modification of the standard Random Walk Metropolis MCMC method which is well-adapted to Gaussian priors in high dimensions. It was introduced in [13], and its dimension independent prop-

erties are analysed and illustrated numerically in [62] and [34] respectively; the pCN nomenclature was introduced in [34]. The algorithm is stated in the Appendix. In the case of the star-shaped prior, we use a Metropolis-within-Gibbs algorithm [134], alternately updating the field with the pCN method above and updating the centre with the standard RWM method.

An advantage of these MCMC methods is that derivatives of the forward map are not needed, only black-box solution of the forward model. However in order to accurately compute some quantity of interest, such as the conditional mean, we may need to produce a very large number of samples and tuning the algorithm to minimize this effect is important. For this reason we compute the effective sample size from the integrated autocorrelation (neglecting a burn-in period) of a quantity of interest, as in [76].

3.4.2 Data and Parameters

We work on a circular domain of radius 1, with 16 equally spaced electrodes on its boundary providing 50% coverage. We take all contact impedances $z_l = 0.01$. We stimulate adjacent electrodes with a current of 0.1, so that the matrix of stimulation patterns $I = (I^{(j)})_{j=1}^{15} \in \mathbb{R}^{16\times 15}$ is given by

$$I = 0.1 \times \begin{pmatrix} +1 & 0 & \cdots & 0 \\ -1 & +1 & \cdots & 0 \\ 0 & -1 & \ddots & 0 \\ \vdots & \vdots & \ddots & +1 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

The conductivity is chosen such that it takes values 1 and 2. We perturb the measurements with white noise $\eta \sim N(0, \gamma^2 I)$, $\gamma = 0.0002$, so that the mean relative error on both sets of data is approximately 10%. The true conductivity fields used to generate the data, henceforth referred to as *Conductivity A* and *Conductivity B*, are shown in Figure 3.6. In all cases we generate $N = 2.5 \times 10^6$ samples with a burn-in of $k_0 = 5 \times 10^5$ samples.

Our priors on fields will make use of Gaussians with covariances of the form

$$\mathcal{C} = q(\tau^2 - \Delta)^{-\alpha}. \tag{3.4.1}$$

These are essentially rescaled Whittle-Matern covariances [93], with τ representing

the inverse length scale of the samples, α proportional to their regularity, and q proportional to their amplitude. These distributions are discussed in detail in the following chapter, section 4.2.1.

In what follows, denote by $\mathcal{A}_N$ the Laplacian with Neumann boundary conditions on $[-1,1]^2$, restricted to D, so that its domain is given by

$$\mathcal{D}(\mathcal{A}_N) = \left\{ u|_D \;\middle|\; u \in H^2([-1,1]^2), \frac{\partial u}{\partial n} = 0 \right\}.$$

Defining the Laplacian first on a square and then restricting to D will allow for efficient generation of Gaussian samples via the fast Fourier transform. Note that if we were to consider priors of the form (3.4.1) with $\tau = 0$, we should restrict $\mathcal{D}(\mathcal{A}_N)$ further to ensure the invertibility of $\mathcal{A}_N$.

Additionally, denote by $\mathcal{A}_D$ the Laplacian with Dirichlet boundary conditions on $R_1 = (-\pi, \pi]$, so that its domain is given by

$$\mathcal{D}(\mathcal{A}_D) = \big\{ u \in H^2\big((-\pi,\pi]\big) \;\big|\; u(-\pi) = u(\pi) = 0 \big\}.$$

Gaussian prior

States are defined on a grid of $2^7 \times 2^7$ points. For both simulations the pCN jump parameter β is taken to be 0.01, with choice of prior

$$\mu_0 = \exp^{\#}\left[N(0.5 \log 2, 10^{16} \cdot (40^2 - \mathcal{A}_N)^{-6}) \right].$$

Star-shaped prior

Radial states are defined on a grid of 2^8 points. For Conductivity A, we choose the pCN jump parameter $\beta = 0.03$ and the RWM jump parameter $\delta = 0.01$. For Conductivity B we choose $\beta = 0.01$ and $\delta = 0.005$. For both simulations we use the choice of prior $\mu_0 = \sigma_0 \times \tau_0$, with

$$\sigma_0 = h^{\#}\left[N(0.5, 10^9 \cdot (30^2 - \mathcal{A}_D)^{-3}) \right], \quad \tau_0 = U([-0.5, 0.5]^2),$$

where $h(z) = (1 + \tanh z)/2$. Note that we choose Dirichlet boundary conditions here to ensure that the boundaries of the star-shaped inclusions generated are closed curves.

Level set prior

States are defined on a grid of $2^7 \times 2^7$ points. For both simulations the pCN jump parameter β is taken to be 0.005, with choice of prior

$$\mu_0 = N(0, (35^2 - \mathcal{A}_N)^{-5}).$$

3.4.3 Results

Recovery

Figures 3.7 and 3.8 show conductivities arising from the MCMC chains, and Figure 3.9 shows the values of the misfit Φ at the different sample means. The sample means are calculated in the sample spaces X_i and then pushed forward to the conductivity space by the maps F_i, so that Figures 3.7-3.8 show estimates of $F_i(\mathbb{E}(u))$. This preserves the binary nature of the fields in the cases of the star-shaped and level set priors, as distinct from estimates of $\mathbb{E}(F_i(u))$.

For Conductivity A, the sample mean arising from the star-shaped prior provides a better reconstruction than the other two prior choices. This is expected, since the true conductivity was drawn from this prior. Whilst the sample mean arising from the level set prior is fairly close to the true conductivity (both visually and in terms of Φ), the boundary of the interface appears to have too large a length-scale. Appropriate choice of prior length-scale is a key issue the with the level set method; treating the length-scale hierarchically as another unknown in the problem may be beneficial. The sample mean arising from the Gaussian prior fails to recover both the sharp interface and the values of the conductivity, which is reflected in the value Φ takes.

For Conductivity B, the level set prior is most effective in the reconstruction, since a specific number of inclusions isn't fixed a priori as it is for the star-shaped prior. Again the Gaussian prior fails to recover both the sharp interface and the values of the conductivity, however it appears to do a better job than the star-shaped prior at identifying the location and shape of the two inclusions.

In both of the above cases, even though the individual samples coming from using the level set prior contain many small inclusions, these do not show up in the sample means.

Convergence

In Figure 3.10, we show the approximate effective sample size (ESS) associated with different quantities of interest. For all choices of prior, these are significantly smaller than the total 2.5×10^6 samples generated. Many more samples may hence be required to produce accurate approximations of the posterior mean.

The chain associated with the star-shaped prior results in the largest ESS, likely because we are only attempting to infer 2^8+2 parameters rather than 2^{14} parameters as in the log-Gaussian and level set cases.

In order to accelerate the convergence of the MCMC we can adjust the jump parameters β and δ. Larger choices of these parameters mean that accepted states will be less correlated with the current state, however the proposed states are less likely to be accepted. The choice $\beta = 1$ in pCN produces proposed states that are independent of the current state, but dependent upon how far the prior is from the posterior, very few or no states may be accepted so that the chain never moves. Similarly, smaller choices of these jump parameters mean that more proposals will be accepted, but the states will be more correlated. A balance hence must be achieved – in our simulations we choose the parameters such that approximately 20-30% of proposals are accepted, though in general the optimal acceptance rate is not known [11].

Alternatively, reconstruction may be accelerated by looking at an approximation of the posterior instead of the exact posterior, for example using the ensemble Kalman filter [69] or a sequential Monte Carlo method [12]. We could also initialize the MCMC chains from EnKF estimates to significantly reduce the burn-in period and hence computational cost. If the derivative of the forward map is available, Hybrid Monte Carlo (HMC) methods could be used to accelerate the convergence [42]. Emulators could also be used to reduce the computational burden of derivative calculation, allowing the use of geometric MCMC methods such as Riemannian Manifold Hamiltonian Monte Carlo (RHMC) and Lagrangian Monte Carlo (LMC) [89].

Posterior Behaviour

In Figures 3.11-3.13 we show kernel density estimates for a number of quantities associated with each posterior distribution. The most regular densities arise in the star-shaped case, with the distribution of all quantities appearing to be very close

to uni-modal. More irregularity is seen for the log-Gaussian case, especially in the joint distributions, but they are still relatively close to uni-modal.

The least regular, highly multi-modal densities come from the level set prior. One reason for this is likely the lack of identifiability of the level set function: the forward model only 'sees' the zero level set of the state, and hence cannot distinguish between infinitely many different states. The prior can however distinguish between these states, and will weight them appropriately, which can help explain the shape of the posterior densities. Another reason for the lack of regularity could be that the MCMC chain failed to converge within the burn-in period, and hence artefacts from the transient period appear in the density estimates.

3.5 Conclusions

The primary contributions of this chapter are:

- We have formulated the EIT problem rigorously in the infinite dimensional Bayesian framework.
- We have studied three different prior models, each with their own advantages and disadvantages based on prior knowledge and the nature of the field we are trying to recover.
- With each of these choices of prior we obtain well-posedness of the problem. We can obtain well-posedness using additional prior models, as long as the mapping from the state space to the conductivity space has appropriate regularity.
- The infinite dimensional formulation of the problem leads to the use of state of the art function space MCMC methods for sampling the posterior distribution.
- Simulations performed using these methods illustrate that the conditional mean provides a reasonable reconstruction of the conductivity, even with fairly significant noise on the measurements. They also illustrate the fact that the choice of prior has a significant impact on reconstruction and, in particular, that the geometric priors (star-shaped and level set) can be particularly effective for the (approximately) piecewise constant fields that arise in many applications.

Future research directions could include the following:

- Sampling the exact posterior distribution using MCMC can be computationally expensive. Methods that approximate the posterior may be as effective for calculating quantities such as the conditional mean, with much lower computational load. The relative effectiveness versus cost of different methods could be studied. This could be especially important for simulations in three dimensions, where forward model evaluations are even more expensive.
- When using the level set prior, the length scale of samples could be treated hierarchically as an additional unknown in the problem. This is the topic of the following chapter, for a more general class of forward models than just EIT.
- The star-shaped prior could be extended to describe multiple inclusions, either with the number of inclusions fixed or as an additional unknown.

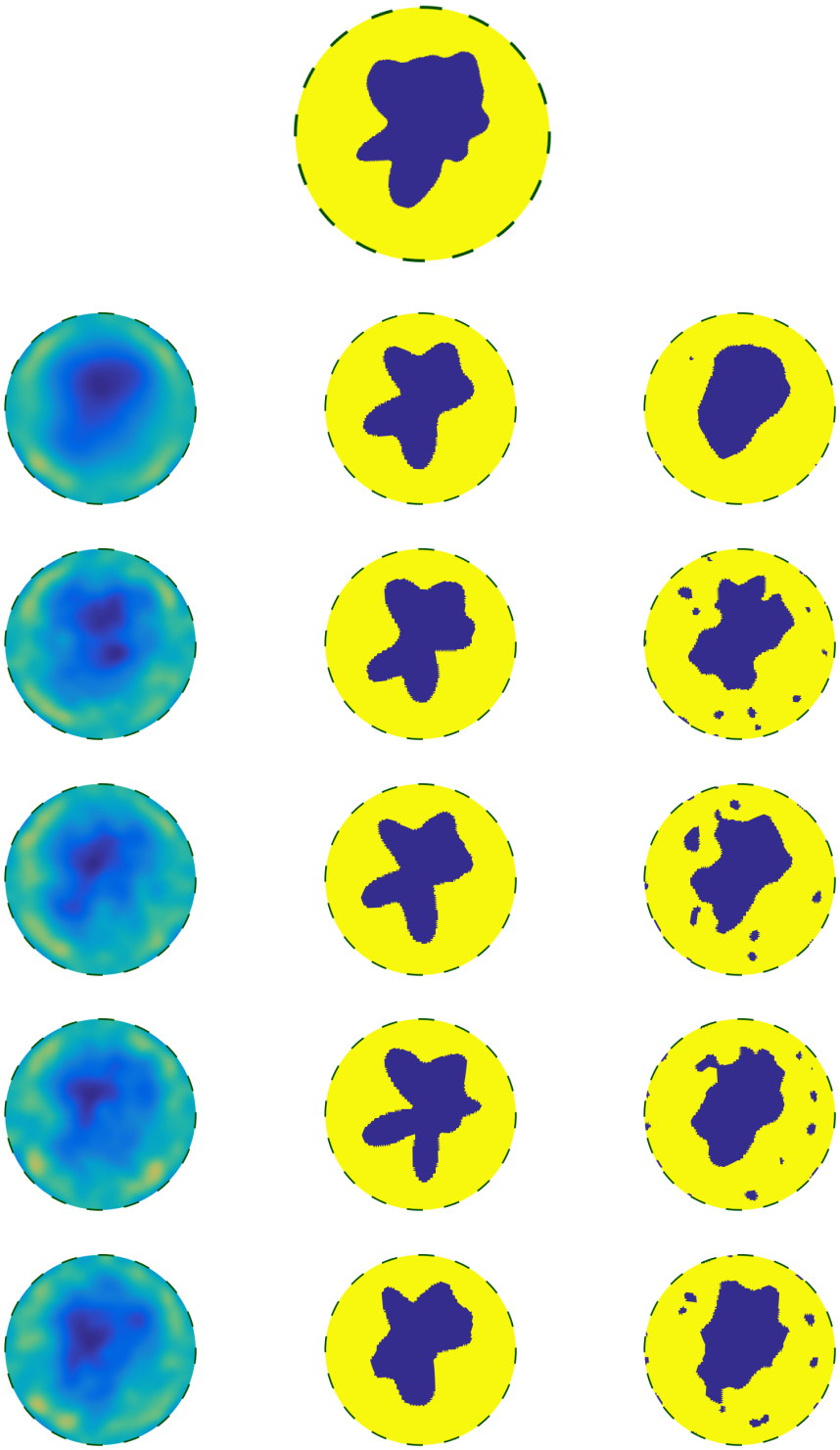


Figure 3.7: Recovery of Conductivity A. From left to right, the log-Gaussian, star-shaped and level set priors are used. (Top) True conductivity (Line 2) Posterior means (Lines 3-6) Posterior samples.

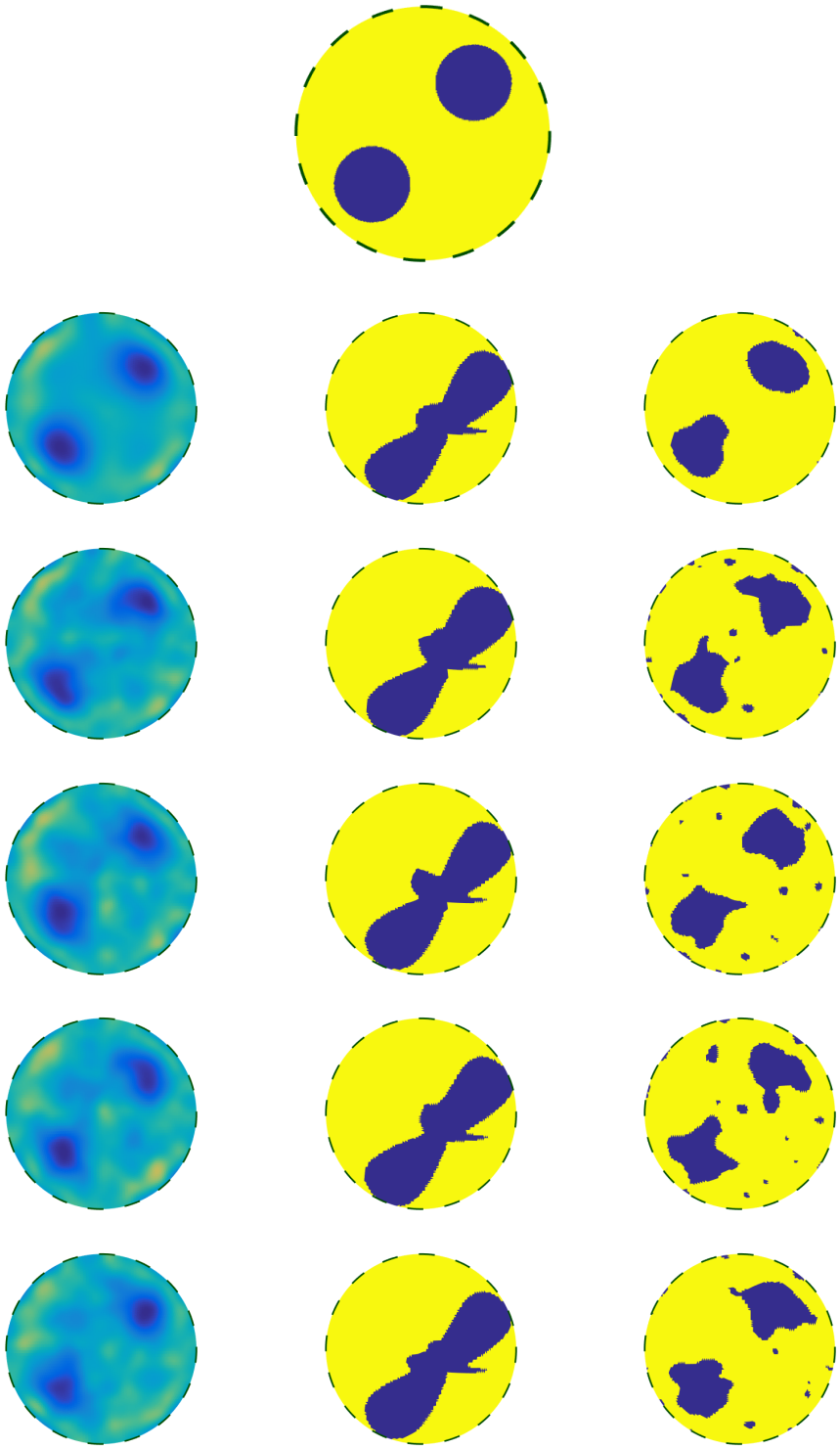


Figure 3.8: Recovery of Conductivity B. From left to right, the log-Gaussian, star-shaped and level set priors are used. (Top) True conductivity (Line 2) Posterior means (Lines 3-6) Posterior samples.

	Conductivity A	Conductivity B
Log-Gaussian prior	11596	12219
Star-shaped prior	228.84	266.22
Level set prior	284.28	193.03

Figure 3.9: The values of the misfit Φ at the sample mean, for the different conductivities and prior distributions.

Quantity	Estimated ESS	Quantity	Estimated ESS
$\hat{u}(0,1)$	40.0	$x_0^{(1)}$	241.7
$\hat{u}(0,2)$	90.7	$x_0^{(2)}$	89.6
$\hat{u}(1,1)$	35.4	$\hat{r}(1)$	101.1
$\hat{u}(1,2)$	44.5	$\hat{r}(2)$	179.4
$\hat{u}(1,3)$	36.0	$\hat{r}(3)$	277.8
$\hat{u}(2,1)$	101.9	$\hat{r}(4)$	214.8
$\hat{u}(2,2)$	37.9	$\hat{r}(5)$	146.7
$\hat{u}(2,3)$	89.7	$\hat{r}(6)$	146.7

Quantity	Estimated ESS
$\hat{u}(0,1)$	26.4
$\hat{u}(0,2)$	28.9
$\hat{u}(1,1)$	27.2
$\hat{u}(1,2)$	23.5
$\hat{u}(1,3)$	23.5
$\hat{u}(2,1)$	26.0
$\hat{u}(2,2)$	26.6
$\hat{u}(2,3)$	24.1

Figure 3.10: (Conductivity B) The estimated effective sample size (ESS) for each chain, approximated using a variety of quantities, for the different choices of prior. In all cases 2.5×10^6 total MCMC samples are produced, with the initial 5×10^5 discarded as burn-in.

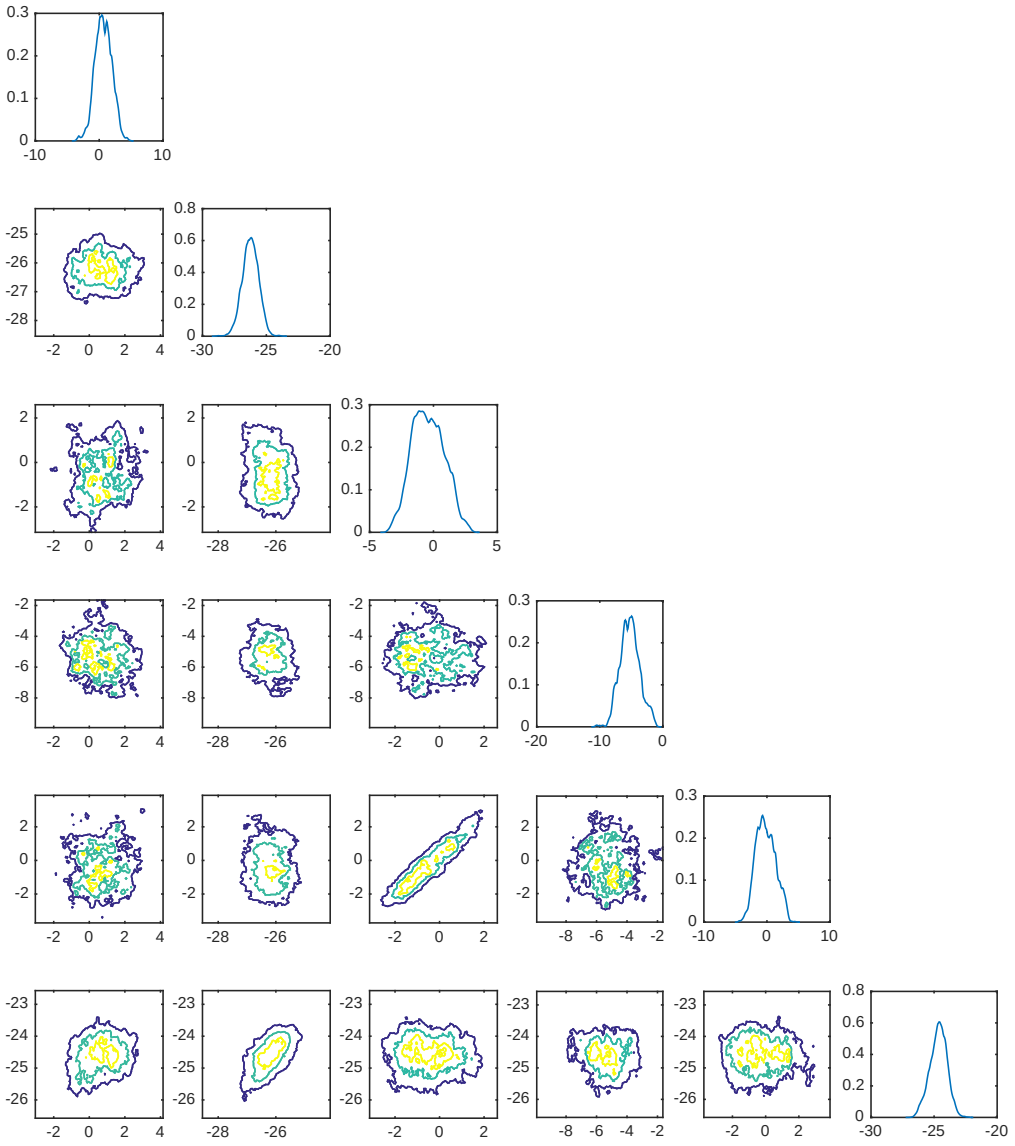


Figure 3.11: (Conductivity B, log-Gaussian prior) Kernel density estimates associated with six Fourier coefficients of u. The diagonal displays the marginal densities of each coefficient, and the off-diagonals the marginal densities of corresponding pairs of coefficients.

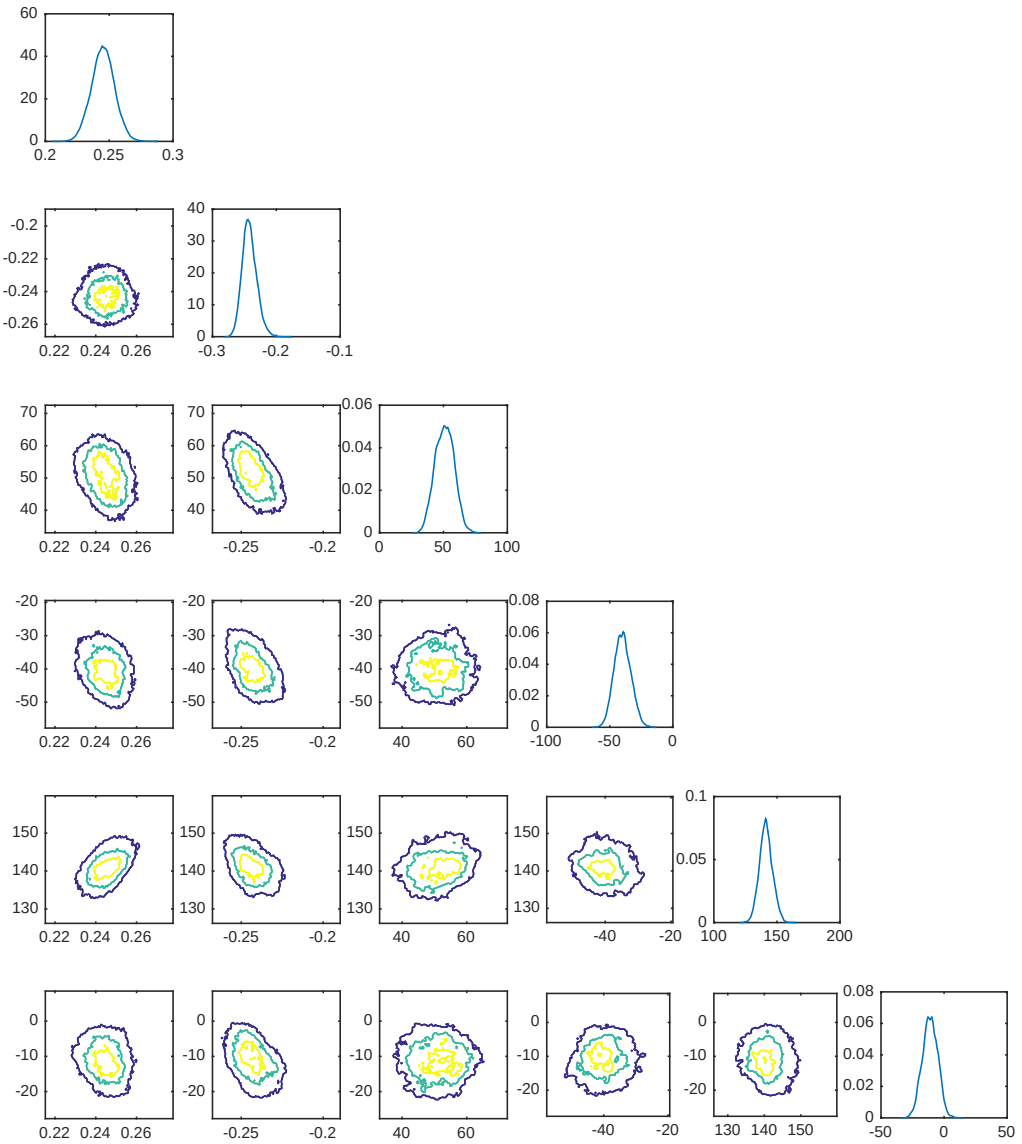


Figure 3.12: (Conductivity B, star-shaped prior) Kernel density estimates associated with the centre $(x_0^{(1)}, x_0^{(2)})$ and four Fourier coefficients of r. The diagonal displays the marginal densities of each quantity, and the off-diagonals the marginal densities of corresponding pairs of quantities.

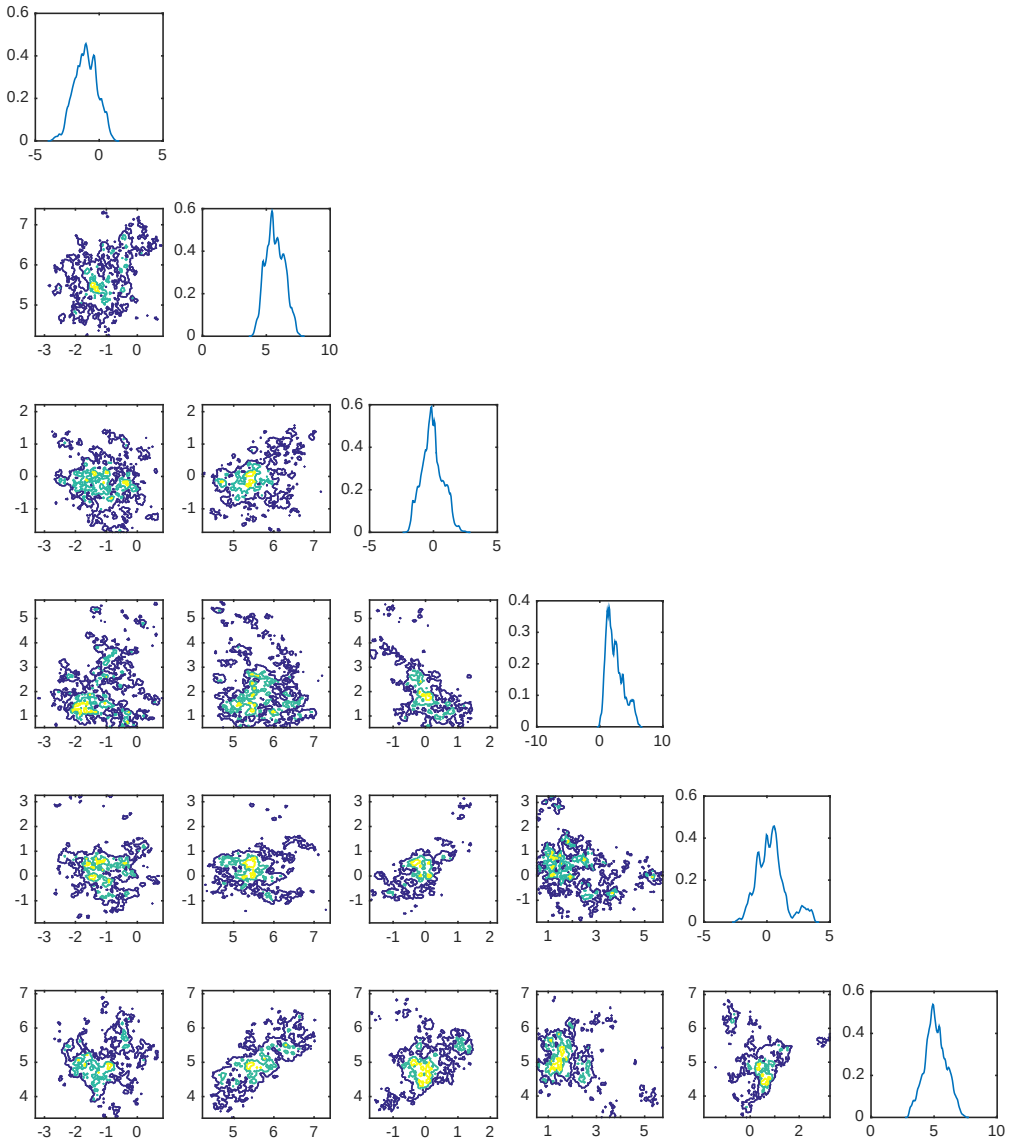


Figure 3.13: (Conductivity B, level set prior) Kernel density estimates associated with six Fourier coefficients of u. The diagonal displays the marginal density of each coefficient, and the off-diagonals the marginal densities of corresponding pairs of coefficients. Axes are rescaled by 10^6 for clarity.

Chapter 4

Hierarchical Bayesian Level Set Inversion

4.1 Introduction

4.1.1 Background

The level set method has been pervasive as a tool for the study of interface problems since its introduction in the 1980s [115]. In a seminal paper in the 1990s, Santosa demonstrated the power of the approach for the study of inverse problems with unknown interfaces [123]. The key benefit of adopting the level set parametrization of interfaces is that topological changes are permitted. In particular for inverse problems the number of connected components of the field does not need to be known a priori. The idea is illustrated in Figure 4.1. The type of unknown functions that we might wish to reconstruct are piecewise continuous functions, illustrated in the bottom row by piecewise constant ternary functions. However in the inversion we work with a smooth function, shown in the top row and known as the *level-set function*, which is thresholded to create the desired unknown function in the bottom row. This allows the inversion to be performed on smooth functions, and allows for topological changes to be detected during the course of algorithms. After Santosa's paper there were many subsequent papers employing the level set representation for classical inversion, and examples include [23, 31, 41, 133], and the references therein.

In many inverse problems arising in modern day science and engineering, the data is noisy and prior regularizing information is naturally expressed probabilistically

since it contains uncertainties. In this context, Bayesian inversion is a very attractive conceptual approach [75]. Early adoption of the Bayesian approach within level set inversion, especially in the context of history matching for reservoir simulation, includes the papers [99,100,117,139]. In a recent paper [72] the mathematical foundations of Bayesian level set inversion were developed, and a well-posedness theorem established, using the infinite dimensional Bayesian framework developed in [39, 91, 92, 131]. An ensemble Kalman filter method has also been applied in the Bayesian level set setting [67] to produce estimates of piecewise constant permeabilities/conductivities in groundwater flow/electrical impedance tomography (EIT) models.

For linear Bayesian inverse problems, the adoption of Gaussian priors leads to Gaussian posteriors, formulae for which can be explicitly computed [50,96,101]. However the *level set map*, which takes the smooth underlying level set function (top row, Figure 4.1) into the physical unknown function (bottom row, Figure 4.1) is nonlinear; indeed it is discontinuous. As a consequence, Bayesian level set inversion, even for inverse problems which are classically-speaking 'linear', does not typically admit closed form solutions for the posterior distribution on the level set function. Thus, in order to produce samples from the posterior arising in the Bayesian approach, MCMC methods are often used. Since the posterior is typically defined on an infinite-dimensional space in the context of inverse problems, it is important that the MCMC algorithms used are well-defined on such spaces. A formulation of the Metropolis-Hastings algorithm on general state spaces is given in [135]. A particular case of this algorithm, well-suited to posterior distributions on function spaces and Gaussian priors, is the preconditioned Crank-Nicolson (pCN) method introduced (although not named this way) in [13]. As the method is defined directly on a function space, it has desirable properties related to discretization – in particular the method is robust with respect to mesh refinement (discretization invariance) – see [34] and the references therein. On the other hand, the need for hierarchical models in Bayesian statistics, and in particular in the context of non-parametric (i.e. function space) methods in machine learning, is well-established [15]. However, care is needed when using hierarchical methods in order to ensure that discretization invariance is not lost [3]. In this chapter we demonstrate how hierarchical methods can be employed in the context of discretization-invariant MCMC methods for Bayesian level set inversion.

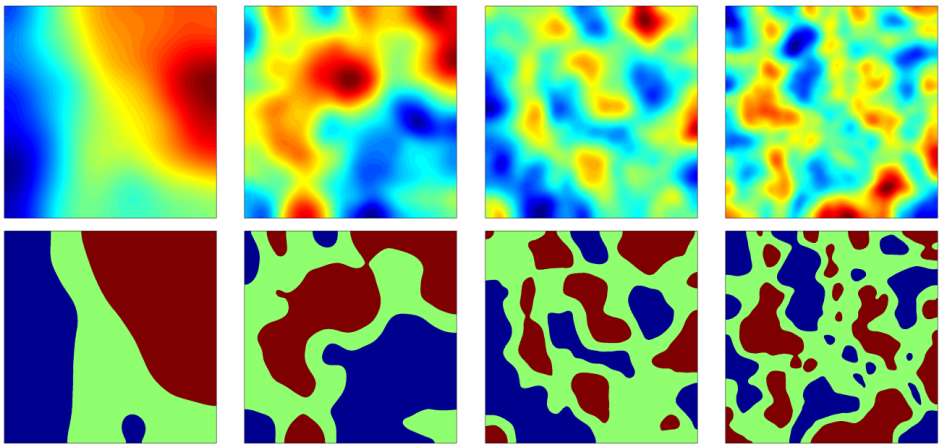


Figure 4.1: Four continuous scalar fields (top) and the corresponding ternary fields formed by thresholding these fields at two levels (bottom). The smooth function in the top row is known as the *level-set function* and is used in the inversion procedure. The discontinuous function in the bottom row is the physical unknown.

4.1.2 Key Contributions of the Chapter

Study of Figure 4.1 suggests that the ability of the level set representation to accurately reconstruct piecewise continuous fields depends on two important scale parameters:

- the length-scale of the level set function, and its relation to the typical separation between discontinuities;
- the amplitude-scale of the level set function, and its relation to the levels used for thresholding.

If these two scale parameters are not set correctly then MCMC methods to determine the level set function from data can perform poorly; this was illustrated numerically in the previous chapter in the context of EIT. This immediately suggests the idea of using hierarchical Bayesian methods in which these parameters are learned from the data. However there is a second consideration which interacts with this discussion. From the work of Tierney [135] it is known that absolute continuity of certain measures arising in the definition of Metropolis-Hastings methods is central for their well-definedness, and hence to discretization invariant MCMC methods [34]. The key contribution of this chapter is to show how enforcing absolute continuity links the two scale parameters, and hence leads to the construction of a hierarchical

Bayesian level set method with a *single scalar* hierarchical parameter which deals with the scale and absolute continuity issues simultaneously, resulting in effective sampling algorithms.

The hierarchical parameter is an inverse length-scale within a Gaussian random field prior for the level set function. In order to preserve absolute continuity of different priors on the level set function as the length-scale parameter varies, and relatedly to make well-defined MCMC methods, the mean square amplitude of this Gaussian random field must decay proportionally to a power of the inverse length-scale. It is thus natural that the level values used for thresholding should obey this power law relationship with respect to the hierarchical parameter. As a consequence the likelihood depends on the hierarchical parameter, leading to a novel form of posterior distribution.

We construct this posterior distribution and demonstrate how to sample from it using a Metropolis-within-Gibbs algorithm which alternates between updating the level set function and the inverse length scale. As a second contribution of the chapter, we demonstrate the applicability of the algorithm on three inverse problems, by means of simulation studies. The first concerns reconstruction of a ternary piecewise constant field from a finite noisy set of point measurements. The other two concern reconstruction of the coefficient of a divergence form elliptic PDE from measurements of its solution; in particular, groundwater flow (in which measurements are made in the interior of the domain) and EIT (in which measurements are made on the boundary).

4.1.3 Structure of the Chapter

In section 4.2 we describe a family of prior distributions on the level set function, indexed by an inverse length scale parameter, which remain absolutely continuous with respect to one another when we vary this parameter; we then place a hyper-prior on this parameter. We describe an appropriate level set map, dependent on the length-scale parameter because length and amplitude scales are intimately connected through absolute continuity of measures, to transform these fields into piecewise constant ones, and use this level set map in the construction of the likelihood. We end by showing existence and well-posedness of the posterior distribution on the level set function and the inverse length scale parameter. In section 4.3 we describe a Metropolis-within-Gibbs MCMC algorithm for sampling the posterior distribution, taking advantage of existing state-of-the-art function space MCMC,

and the absolute continuity of our prior distributions with respect to changes in the inverse length scale parameter, established in the previous section. Section 4.4 contains numerical experiments for three different forward models: a linear map comprising pointwise observations, groundwater flow and EIT; these illustrate the behavior of the algorithm and, in particular, demonstrate significant improvement with respect to non-hierarchical Bayesian level set inversion.

4.2 Construction of the Posterior

In subsection 4.2.1 we recall the definition of the Whittle-Matérn covariance functions, and define a related family of covariances parametrized by an inverse length scale parameter τ. We use these covariances to define our prior on the level set function u, and also place a hyperprior on the parameter τ, yielding a prior $\mathbb{P}(u, \tau)$ on a product space. In subsection 4.2.2 we construct the level set map, taking into account the amplitude scaling of prior samples with τ, and incorporate this into the forward map. The inverse problem is formulated, and the resulting likelihood $\mathbb{P}(y|u, \tau)$ is defined. Finally in subsection 4.2.3 we construct the posterior $\mathbb{P}(u, \tau|y)$ by combining the prior $\mathbb{P}(u, \tau)$ and likelihood $\mathbb{P}(y|u, \tau)$ using Bayes' formula. Well-posedness of this posterior is established.

4.2.1 Prior

As discussed in the introduction it can be important, within the context of Bayesian level set inversion, to attempt to learn the length-scale of the level set function whose level sets determine interfaces in piecewise continuous reconstructions. This is because we typically do not know a-priori the typical separation of interfaces. It is also computationally expedient to work with Gaussian random field priors for the level set function, as demonstrated in [45,72]. A family of covariances parameterized by length scale is hence required.

A widely used family of distributions, allowing for control over sample regularity, amplitude and length scale, are Whittle-Matérn distributions. These are a family of stationary Gaussian distributions with covariance function

$$c_{\sigma,\nu,\ell}(x, y) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{|x-y|}{\ell} \right)^{\nu} K_\nu \left(\frac{|x-y|}{\ell} \right)$$

where K_ν is the modified Bessel function of the second kind of order ν [102, 130].

These covariances interpolate between exponential covariance, for $\nu = 1/2$, and Gaussian covariance, for $\nu \to \infty$. As a consequence, the regularity of samples increases as the parameter ν increases. The parameter $\ell > 0$ acts as a characteristic length scale (sometimes referred to as the spatial range) and σ as an amplitude scale (σ^2 is sometimes referred to as the marginal variance). On $\mathbb{R}^d$, samples from a Gaussian distribution with covariance function $c_{\sigma,\nu,\ell}$ correspond to the solution of a particular stochastic partial differential equation (SPDE). This SPDE can be derived using the Fourier transform and the spectral representation of covariance functions – the paper [93] derives the appropriate SPDE for the covariance function above:

$$\frac{1}{\sqrt{\beta \ell^d}}(I - \ell^2 \Delta)^{(\nu+d/2)/2} v = W \tag{4.2.1}$$

where W is a white noise on $\mathbb{R}^d$, and

$$\beta = \sigma^2 \frac{2^d \pi^{d/2} \Gamma(\nu + d/2)}{\Gamma(\nu)}.$$

Computationally, implementation of this SPDE approach requires restriction to a bounded subset $D \subseteq \mathbb{R}^d$, and hence the provision of boundary conditions for the SPDE in order to obtain a unique solution. Choice of these boundary conditions may significantly affect the autocorrelations near the boundary. The effects for different boundary conditions are discussed in [93]. Nonetheless, the computational expediency of the SPDE formulation makes the approach very attractive for applications and, if necessary, boundary effects can be ameliorated by generating the random fields on larger domains which are a superset of the domain of interest.

From (4.2.1) it can be seen that the covariance operator corresponding to the covariance function $c_{\sigma,\nu,\ell}$ is given by

$$\mathcal{D}_{\sigma,\nu,\ell} = \beta \ell^d (I - \ell^2 \Delta)^{-\nu - d/2}. \tag{4.2.2}$$

The fact that the scalar multiplier in front of the covariance operator $\mathcal{D}_{\sigma,\nu,\ell}$ changes with the length-scale means that the family of measures $\{N(0, \mathcal{D}_{\sigma,\nu,\ell})\}_\ell$, for fixed σ and ν, are mutually singular. This leads to problems when trying to design hierarchical methods based around these priors, even for latent Gaussian random field models [141]. We hence work instead with the modified covariances

$$\mathcal{C}_{\alpha,\tau} = (\tau^2 I - \Delta)^{-\alpha}$$

where $\tau = 1/\ell > 0$ now represents an *inverse* length scale, and $\alpha = \nu + d/2$ still controls the sample regularity. The inconsistency problems arising from varying σ now disappear for latent Gaussian random field models [136]. To be concrete we will always assume that the domain of the Laplacian is chosen so that $\mathcal{C}_{\alpha,\tau}$ is well-defined for all $\tau \geq 0$; for example we may choose a periodic box, with domain restricted to functions which integrate to zero over the box, Neumann boundary conditions on a box, again with domain restricted to functions which integrate to zero over the box, or Dirichlet boundary conditions. We have the following theorem concerning the family of Gaussians $\{N(m, \mathcal{C}_{\alpha,\tau})\}_{\tau \geq 0}$, proved in Appendix.

Theorem 4.2.1. *Let $D = \mathbb{T}^d$ be the d-dimensional torus, and fix $\alpha > 0$. Define the family of Gaussian measures $\mu_0^\tau = N(m, \mathcal{C}_{\alpha,\tau})$, $\tau \geq 0$. Then*

(i) for $d \leq 3$, the $(\mu_0^\tau)_{\tau \geq 0}$ are mutually equivalent;

(ii) if $u \sim \mu_0^\tau$, then μ_0^τ-a.s. we have $u \in H^s(\overline{D})$ and $u \in C^{\lfloor s \rfloor, s - \lfloor s \rfloor}(\overline{D})$ for all $s < \alpha - d/2$. [1]

(iii) $\mathbb{E}\|u - m\|^2 \propto \tau^{d-2\alpha}$ with constant of proportionality independent of τ.

Remark 4.2.2. *Proof of this theorem is driven by the smoothness of the eigenfunctions of the Laplacian subject to periodic boundary conditions, together with the growth of the eigenvalues, which is like like $j^{2/d}$. These properties extend to Laplacians on more general domains and with more general boundary conditions, and to Laplacians with lower order perturbations, and so the above result still holds in these cases. For discussion of this in relation to (ii) see [39]; for parts (i) and (iii) the reader can readily extend the proof given in the Appendix.*

Let $X = C(\overline{D})$ denote the space of continuous real-valued functions on domain D. In what follows we will always assume that $\alpha - d/2 > 0$ in order that the measures have samples in X almost-surely. Additionally we shall write $\mathcal{C}_\tau$ in place of $\mathcal{C}_{\alpha,\tau}$ when the parameter α is not of interest.

In subsection 4.2.2, we pass the inverse length scale parameter τ to the forward map and treat it as an additional unknown in the inverse problem. We therefore require a joint prior $\mathbb{P}(u, \tau)$ on both the field and on τ. We will treat τ as a hyper-parameter, so that $\mathbb{P}(u, \tau)$ takes the form $\mathbb{P}(u, \tau) = \mathbb{P}(u|\tau)\mathbb{P}(\tau)$. Specifically, we will take the conditional distribution $\mathbb{P}(u|\tau)$ to be given by $\mu_0^\tau = N(m_0, \mathcal{C}_\tau)$, where $m_0 \in X$ is constant, and the hyper-prior $\mathbb{P}(\tau)$ to be any probability measure π_0 on $\mathbb{R}^+$, the set

[1] i.e. the functions has s weak (possibly fractional) derivatives in the Sobolev sense, and the $\lfloor s \rfloor^{th}$ classical derivative is Hölder with exponent $s - \lfloor s \rfloor$;

of positive reals; in practice it will always have a Lebesgue density on $\mathbb{R}^+$. The joint prior μ_0 on $X \times \mathbb{R}^+$ is therefore assumed to be given by

$$\mu_0(\mathrm{d}u, \mathrm{d}\tau) = \mu_0^\tau(\mathrm{d}u)\pi_0(\mathrm{d}\tau). \tag{4.2.3}$$

Discussion of prior choice for the hierarchical parameters in latent Gaussian models may be found in [51].

4.2.2 Likelihood

In the previous subsection we defined a prior distribution μ_0 on $X \times \mathbb{R}^+$. We now define a way of constructing a piecewise constant field from a sample (u, τ). In [72], where the Bayesian level set method was introduced, the piecewise constant field was constructed purely as a function of u as follows. Let $n \in \mathbb{N}$ and fix constants $-\infty = c_0 < c_1 < \ldots < c_n = \infty$. Given $u \in X$, define $D_i(u) \subseteq D$ by

$$D_i(u) = \{x \in D \,|\, c_{i-1} \leq u(x) < c_i\}, \quad i = 1, \ldots, n$$

so that $\overline{D} = \bigcup_{i=1}^n \overline{D}_i(u)$ and $D_i(u) \cap D_j(u) = \varnothing$ for $i \neq j$, $i, j \geq 1$. Then given $\kappa_1, \ldots, \kappa_n \in \mathbb{R}$, define the map $F : X \to Z$ by

$$F(u) = \sum_{i=1}^n \kappa_i \mathbb{1}_{D_i(u)}.$$

We may take $Z = L^p(\overline{D})$, the space of p-integrable functions on D, for any $1 \leq p \leq \infty$. This construction is effective for a fixed value of τ, but in light of Theorem 4.2.1(iii), the amplitude of samples from $N(m_0, \mathcal{C}_{\alpha,\tau})$, varies with τ. More specifically, since $d - 2\alpha < 0$ by assumption, samples will decay toward the mean as τ increases. For this reason, employing fixed levels $(c_i)_{i=0}^n$ and then changing the value of τ during a sampling method may render the levels out of reach. We can compensate for this by allowing the levels to change with τ, so that they decay towards the mean at the same rate as the samples.

From Theorem 4.2.1(iii) we deduce that samples u from $N(m_0, \mathcal{C}_{\alpha,\tau})$ decay towards m_0 at a rate $\tau^{d/2-\alpha}$ with respect to τ. This suggests allowing for the following dependence of the levels on the constant mean m_0 and length scale parameter τ:

$$c_i(\tau) = m_0 + \tau^{d/2-\alpha}(c_i - m_0), \quad i = 1, \ldots, n. \tag{4.2.4}$$

In order to update these levels, we must pass the parameter τ to the level set map F. We therefore define a new level set map $F : X \times \mathbb{R}^+ \to Z$ as follows. Let $n \in \mathbb{N}$, fix initial levels $-\infty = c_0 < c_1 < \ldots < c_n = \infty$ and define $c_i(\tau)$ by (4.2.4) for $\tau > 0$. Given $u \in X$ and $\tau > 0$, define $D_i(u, \tau) \subseteq D$ by

$$D_i(u, \tau) = \{x \in D \mid c_{i-1}(\tau) \leq u(x) < c_i(\tau)\}, \quad i = 1, \ldots, n, \tag{4.2.5}$$

so that $\overline{D} = \bigcup_{i=1}^n \overline{D}_i(u, \tau)$ and $D_i(u, \tau) \cap D_j(u, \tau) = \varnothing$ for $i \neq j$, $i, j \geq 1$. Now given $\kappa_1, \ldots, \kappa_n \in \mathbb{R}$, we define the map $F : X \times \mathbb{R}^+ \to Z$ by

$$F(u, \tau) = \sum_{i=1}^n \kappa_i \mathbb{1}_{D_i(u,\tau)}. \tag{4.2.6}$$

We can now formulate the inverse problem. Let $Y = \mathbb{R}^J$ be the data space, and let $S : Z \to Y$ be a forward operator. Define $\mathcal{G} : X \times \mathbb{R}^+ \to Y$ by $\mathcal{G} = S \circ F$. Assume we have data $y \in Y$ arising from observations of some $(u, \tau) \in X \times \mathbb{R}^+$ under $\mathcal{G}$, corrupted by Gaussian noise $\eta \sim \mathbb{Q}_0 := N(0, \Gamma)$ on Y:

$$y = \mathcal{G}(u, \tau) + \eta. \tag{4.2.7}$$

We now construct the likelihood $\mathbb{P}(y|u, \tau)$. In the Bayesian formulation, we place a prior μ_0 of the form (4.2.3) on the pair (u, τ). Assuming $\mathbb{Q}_0$ is independent of μ_0, the conditional distribution $\mathbb{Q}_{u,\tau}$ of y given (u, τ) is given by

$$\frac{\mathrm{d}\mathbb{Q}_{u,\tau}}{\mathrm{d}\mathbb{Q}_0}(y) = \exp\left(-\Phi(u, \tau; y) + \frac{1}{2}|y|_\Gamma^2 \right) \tag{4.2.8}$$

where the potential (or negative log-likelihood) $\Phi : X \times \mathbb{R}^+ \to \mathbb{R}$ is defined by

$$\Phi(u, \tau; y) = \frac{1}{2}|y - \mathcal{G}(u, \tau)|_\Gamma^2. \tag{4.2.9}$$

and $|\cdot|_\Gamma := |\Gamma^{-1/2} \cdot|$.

Denote $\mathrm{Im}(F) \subseteq Z$ the image of $F : X \times \mathbb{R}^+ \to Z$. In what follows we make the following assumptions on $S : Z \to Y$.

Assumptions 4.2.3. *(i) S is continuous on* $\mathrm{Im}(F)$.

(ii) For any $r > 0$ there exists $C(r) > 0$ such that for any $z \in \mathrm{Im}(F)$ with $\|z\|_{L^\infty} \leq r$, $|S(z)| \leq C(r)$.

In the next subsection we show that, under the above assumptions, the posterior distribution μ^y of (u, τ) given y exists, and study its properties.

4.2.3 Posterior

Bayes' theorem provides a way to construct the posterior distribution $\mathbb{P}(u, \tau|y)$ using the ingredients of the prior $\mathbb{P}(u, \tau)$ and the likelihood $\mathbb{P}(y|u, \tau)$ from the previous two subsections. Informally we have

$$\begin{aligned}\mathbb{P}(u, \tau|y) &\propto \mathbb{P}(y|u, \tau)\mathbb{P}(u, \tau) \\ &\propto \exp\left(-\Phi(u, \tau; y)\right) \mu_0^\tau(u)\pi_0(\tau)\end{aligned}$$

after absorbing $y-$dependent constants from the likelihood into the normalization constant. In order to make this formula rigorous some care must be taken, since μ_0^τ does not admit a Lebesgue density. The following is proved in the Appendix.

Theorem 4.2.4. *Let μ_0 be given by (4.2.3), y by* (4.2.7) *and Φ be given by (4.2.9). Let Assumptions 4.2.3 hold. If $\mu^y(du, d\tau)$ is the regular conditional probability measure on $(u, \tau)|y$, then $\mu^y \ll \mu_0$ with Radon-Nikodym derivative*

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(u, \tau) = \frac{1}{Z} \exp\big(-\Phi(u, \tau; y)\big)$$

where, for y almost surely,

$$Z := \int_{X\times\mathbb{R}^+} \exp\big(-\Phi(u, \tau; y)\big)\, \mu_0(\mathrm{d}u, \mathrm{d}\tau) > 0.$$

Furthermore μ^y is locally Lipschitz with respect to y, in the Hellinger distance: for all y, y' with $\max\{|y|_\Gamma, |y'|_\Gamma\} < r$, there exists a $C = C(r) > 0$ such that

$$d_{\mathrm{Hell}}(\mu^y, \mu^{y'}) \le C|y - y'|_\Gamma.$$

This implies that, for all $f \in L^2_{\mu_0}(X \times \mathbb{R}^+; E)$ for separable Banach space E,

$$\|\mathbb{E}^{\mu^y} f(u, \tau) - \mathbb{E}^{\mu^{y'}} f(u, \tau)\|_E \le C|y - y'|.$$

To the best of our knowledge this form of Bayesian inverse problem, in which the prior hyper-parameter appears in the likelihood because it is natural to scale a thresholding function with that parameter, is novel. A different form of thresholding

is studied in the paper [19] where boundaries defining regions in which certain events occur with a specified (typically close to 1) probability is studied.

4.3 MCMC Algorithm for Posterior Sampling

Having constructed the posterior distribution on $(u, \tau)|y$ we are now faced with the task of sampling this probability distribution. We will use the Metropolis-within-Gibbs formalism, as described in for example [119], section 10.3. This algorithm constructs the Markov chain $(u^{(k)}, \tau^{(k)})$ with the structure

- $u^{(k+1)} \sim \mathbb{K}^{\tau^{(k)},y}(u^{(k)}, \cdot)$,
- $\tau^{(k+1)} \sim \mathbb{L}^{u^{(k+1)},y}(\tau^{(k)}, \cdot)$,

where $\mathbb{K}^{\tau,y}$ is a Metropolis-Hastings Markov kernel reversible with respect to $u|(\tau, y)$ and $\mathbb{L}^{u,y}$ is a Metropolis-Hastings Markov kernel reversible with respect to $\tau|(u, y)$. The Metropolis-Hastings method is outlined in chapter 7 of [119]. See [53] for related blocking methodologies for Gibbs samplers in the context of latent Gaussian models.

In defining the conditional distributions, and the Metropolis methods to sample from them, a key design principle is to ensure that all measures and algorithms are well-defined in the infinite-dimensional setting, so that the resulting algorithms are robust to mesh-refinement [34]. This thinking has been behind the form of the prior and posterior distributions developed in the previous section, as we now demonstrate.

In subsection 4.3.1 we define the kernel $\mathbb{K}^{\tau,y}$ and in subsection 4.3.2 we define the kernel $\mathbb{L}^{u,y}$. Then in the final subsection 4.3.3 we put all these building blocks together to specify the complete algorithm used.

4.3.1 Proposal and Acceptance Probability for $u|(\tau, y)$

Samples from the distribution of $u|(\tau, y)$ can be produced using a pCN Metropolis Hastings method [34], with proposal and acceptance probability as follows:

1. Given u, propose

$$v = m_0 + (1 - \beta^2)^{1/2}(u - m_0) + \beta\xi, \quad \xi \sim N(0, \mathcal{C}_\tau).$$

2. Accept with probability

$$\alpha(u,v) = \min\left\{1, \exp\left(\Phi(u,\tau;y) - \Phi(v,\tau;y)\right)\right\}$$

or else stay at u.

4.3.2 Proposal and Acceptance Probability for $\tau|(u,y)$

Producing samples of $\tau|(u,y)$ is more involved, since we must first make sense of this conditional distribution. To do this, define the three measures η_0, ν_0, and ν on $X \times \mathbb{R}^+ \times Y$ by

$$\begin{aligned}
\eta_0(\mathrm{d}u, \mathrm{d}\tau, \mathrm{d}y) &= \mu_0^0(\mathrm{d}u)\pi_0(\mathrm{d}\tau)\mathbb{Q}_0(\mathrm{d}y), \\
\nu_0(\mathrm{d}u, \mathrm{d}\tau, \mathrm{d}y) &= \mu_0^\tau(\mathrm{d}u)\pi_0(\mathrm{d}\tau)\mathbb{Q}_0(\mathrm{d}y), \\
\nu(\mathrm{d}u, \mathrm{d}\tau, \mathrm{d}y) &= \mu_0^\tau(\mathrm{d}u)\pi_0(\mathrm{d}\tau)\mathbb{Q}_{u,\tau}(\mathrm{d}y).
\end{aligned}$$

Here $\mathbb{Q}_0 = N(0,\Gamma)$ is the distribution of the noise, and $\mathbb{Q}_{u,\tau}$ is as defined in (4.2.8). Then we have the chain of absolute continuities $\nu \ll \nu_0 \ll \eta_0$, with

$$\begin{aligned}
\frac{\mathrm{d}\nu_0}{\mathrm{d}\eta_0}(u,\tau,y) &= \frac{\mathrm{d}\mu_0^\tau}{\mathrm{d}\mu_0^0}(u) =: L(u,\tau), \\
\frac{\mathrm{d}\nu}{\mathrm{d}\nu_0}(u,\tau,y) &= \frac{\mathrm{d}\mathbb{Q}_{u,\tau}}{\mathrm{d}\mathbb{Q}_0}(y) = \exp\left(-\Phi(u,\tau;y) + \frac{1}{2}|y|_\Gamma^2\right),
\end{aligned}$$

and so by the chain rule we have $\nu \ll \eta_0$ and

$$\frac{\mathrm{d}\nu}{\mathrm{d}\eta_0}(u,\tau,y) = \frac{\mathrm{d}\mathbb{Q}_{u,\tau}}{\mathrm{d}\mathbb{Q}_0}(y) \cdot \frac{\mathrm{d}\mu_0^\tau}{\mathrm{d}\mu_0^0}(u) =: \varphi(u,\tau,y).$$

We use the conditioning lemma, Theorem 3.1 in [39], to prove the existence of the desired conditional distribution.

Theorem 4.3.1. *Assume that $\Phi : X \times Y \to \mathbb{R}$ is $\mu_0^0 \times \mathbb{Q}_0$ measurable and $\mu_0^0 \times \mathbb{Q}_0$-a.s. finite. Assume also that, for (u,y) $\mu_0^0 \times \mathbb{Q}_0$-a.s.,*

$$Z_\pi := \int_{\mathbb{R}^+} \exp\big(-\Phi(u,\tau;y)\big) L(u,\tau)\, \pi_0(\mathrm{d}\tau) > 0.$$

Then the regular conditional distribution of $\tau|(u,y)$ exists under ν, and is denoted

by $\pi^{u,y}$. Furthermore, $\pi^{u,y} \ll \pi_0$ and, for (u,y) ν-a.s,

$$\frac{\mathrm{d}\pi^{u,y}}{\mathrm{d}\pi_0}(\tau) = \frac{1}{Z_\pi}\exp\big(-\Phi(u,\tau;y)\big)L(u,\tau).$$

Proof. The conditional random variable $\tau|(u,y)$ exists under η_0, and its distribution is just π_0 since η_0 is a product measure. Theorem 3.1 in [39] then tells us that the the conditional random variable $\tau|(u,y)$ exists under ν. We denote its distribution $\pi^{u,y}$. Define

$$\begin{aligned} c(u,y) &= \int_{\mathbb{R}^+} \varphi(u,\tau,y)\pi_0(\mathrm{d}\tau) \\ &= \exp\left(\frac{1}{2}|y|_\Gamma^2\right)\int_{\mathbb{R}^+}\exp\big(-\Phi(u,\tau;y)\big)L(u,\tau)\pi_0(\mathrm{d}\tau). \end{aligned}$$

Now since $\exp\left(\frac{1}{2}|y|_\Gamma^2\right) \in (0,\infty)$ $\mu_0^0 \times \mathbb{Q}_0$-a.s., we deduce that $c(u,y) > 0$ $\mu_0^0 \times \mathbb{Q}_0$-a.s. by the μ_0^0-a.s. positivity of Z_π. By the absolute continuity $\nu \ll \eta_0$, we deduce that $c(u,y) > 0$ ν-a.s. Therefore, again by Theorem 3.1 in [39], we have $\pi^{u,y} \ll \pi_0$ and, for (u,y) ν-a.s.,

$$\begin{aligned} \frac{\mathrm{d}\pi^{u,y}}{\mathrm{d}\pi_0}(\tau) &= \frac{1}{c(u,y)}\varphi(u,\tau,y) \\ &= \frac{1}{Z_\pi}\exp\big(-\Phi(u,\tau;y)\big)L(u,\tau). \end{aligned}$$

□

Remark 4.3.2. *Above we have used μ_0^0 as a reference measure, and the function $L(u,\tau)$ enters our expression for the posterior. But any μ_0^λ will suffice since the entire family of measures $\{\mu_0^\tau\}_{\tau\geq 0}$ are equivalent to one another. A straightforward calculation with the chain rule gives*

$$\begin{aligned} \frac{\mathrm{d}\pi^{u,y}}{\mathrm{d}\pi_0}(\tau) &= \frac{1}{Z_{\pi,\lambda}}\frac{\mathrm{d}\mu_0^\tau}{\mathrm{d}\mu_0^\lambda}(u)\exp\big(-\Phi(u,\tau;y)\big) \\ &:= \frac{1}{Z_{\pi,\lambda}}L_\lambda(u,\tau)\exp\big(-\Phi(u,\tau;y)\big). \end{aligned}$$

We now wish to sample from $\pi^{u,y}$ using a Metropolis-Hastings algorithm. We assume from now on that π_0 admits a Lebesgue density, so that $\pi^{u,y}$ also admits a Lebesgue density. Abusing notation and using $\pi^{u,y}, \pi_0$ to denote Lebesgue densities as well as the corresponding measures we have

$$\pi^{u,y}(\tau) \propto \exp\big(-\Phi(u,\tau;y)\big)L(u,\tau)\pi_0(\tau).$$

Take a proposal kernel $Q(\tau, \mathrm{d}\gamma) = q(\tau, \gamma)\,\mathrm{d}\gamma$. Define the two measures ρ, ρ^T on $(\mathbb{R} \times \mathbb{R}, \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}))$ by

$$\begin{aligned}\rho(\mathrm{d}\tau, \mathrm{d}\gamma) &= \pi^{u,y}(\mathrm{d}\tau)Q(\tau, \mathrm{d}\gamma) \\ &\propto \exp\big(-\Phi(u, \tau; y)\big)L(u, \tau)\pi_0(\tau)q(\tau, \gamma)\,\mathrm{d}\tau\mathrm{d}\gamma, \\ \rho^T(\mathrm{d}\tau, \mathrm{d}\gamma) &= \mu(\mathrm{d}\gamma, \mathrm{d}\tau).\end{aligned}$$

Then under appropriate conditions on π_0 and q, these two measures are equivalent. Define $r(\tau, \gamma)$ to be the Radon-Nikodym derivative

$$\begin{aligned}r(\tau, \gamma) &:= \frac{\mathrm{d}\rho^T}{\mathrm{d}\rho}(\tau, \gamma) \\ &= \exp\big(\Phi(u, \tau; y) - \Phi(u, \gamma; y)\big) \cdot \frac{\mathrm{d}\mu_0^\gamma}{\mathrm{d}\mu_0^\tau}(u) \cdot \frac{\pi_0(\gamma)q(\gamma, \tau)}{\pi_0(\tau)q(\tau, \gamma)}.\end{aligned}$$

The general form of the Metropolis-Hastings algorithm, as for example given in [135], says that we produce samples from $\pi^{u,y}$ by iterating the follow two steps:

1. Given τ, propose $\gamma \sim Q(\tau, \mathrm{d}\gamma)$.
2. Accept with probability $\alpha(\tau, \gamma) = \min\big\{1, r(\tau, \gamma)\big\}$, or else stay at τ.

In order to implement this algorithm, we need an expression for the Radon-Nikodym derivative $\frac{\mathrm{d}\mu_0^\gamma}{\mathrm{d}\mu_0^\tau}(u)$. Using Proposition 4.6.3, we see that

$$\begin{aligned}\frac{\mathrm{d}\mu_0^\gamma}{\mathrm{d}\mu_0^\tau}(u) &= \prod_{j=1}^{\infty} \frac{\lambda_j(\tau)^{1/2}}{\lambda_j(\gamma)^{1/2}} \exp\left(\frac{1}{2}\sum_{j=1}^{\infty}\left(\frac{1}{\lambda_j(\tau)} - \frac{1}{\lambda_j(\gamma)}\right)\langle u - m_0, \varphi_j\rangle^2\right) \qquad (4.3.1) \\ &= \exp\left(\frac{1}{2}\sum_{j=1}^{\infty}\left(\frac{1}{\lambda_j(\tau)} - \frac{1}{\lambda_j(\gamma)}\right)\langle u - m_0, \varphi_j\rangle^2 + \log\left(\frac{\lambda_j(\tau)}{\lambda_j(\gamma)}\right)\right)\end{aligned}$$

where the $\lambda_j(\cdot)$ are the eigenvalues of $\mathcal{C}_\cdot$ as defined earlier, and φ_j are the corresponding eigenvectors.

From Theorem 4.2.1 we know that μ_0^τ and μ_0^γ are equivalent, and so it must be the case that the expressions for the derivative above are almost-surely finite. However this is not immediately clear from inspection of the expression; thus we provide some intuition about why it is so in the following theorem. The proof is given in the Appendix.

Theorem 4.3.3. *Assume that $u \sim N(m_0, \mathcal{C}_0)$. Then for each $\tau > 0$,*

(i) $\sum_{j=1}^{\infty}\left(\frac{1}{\lambda_j(\tau)}-\frac{1}{\lambda_j(0)}\right)\langle u-m_0,\varphi_j\rangle^2$ *is almost-surely finite if and only if* $d=1$*; and*

(ii) $\sum_{j=1}^{\infty}\left[\left(\frac{1}{\lambda_j(\tau)}-\frac{1}{\lambda_j(0)}\right)\langle u-m_0,\varphi_j\rangle^2+\log\left(\frac{\lambda_j(\tau)}{\lambda_j(0)}\right)\right]$ *is almost-surely finite if* $d\leq 3$.

A consequence of part (i) of this result is that in dimensions 2 and 3, both the product and the sum in (4.3.1) diverge, despite the whole expression being finite. This means that care is required when numerically implementing the Gibbs update of τ.

4.3.3 The Algorithm

Putting the theory above together, we can write down a Metropolis-within-Gibbs algorithm for sampling the posterior distribution. Recall that we assumed the proposal kernel Q admitted a Lebesgue density q: $Q(\tau,\mathrm{d}\gamma)=q(\tau,\gamma)\mathrm{d}\gamma$.

Let $(\lambda_j(\tau),\varphi_j)_{j\geq 1}$ denote the eigenbasis associated with $\mathcal{C}_\tau$. Define

$$w(\tau,\gamma)=\exp\left(\frac{1}{2}\sum_{j=1}^{\infty}\left(\frac{1}{\lambda_j(\tau)}-\frac{1}{\lambda_j(\gamma)}\right)\langle u-m_0,\varphi_j\rangle^2+\log\left(\frac{\lambda_j(\tau)}{\lambda_j(\gamma)}\right)\right)$$

and set

$$\alpha^{\tau}(u,v)=\min\Big\{1,\exp\big(\Phi(u,\tau;y)-\Phi(v,\tau;y)\big)\Big\},$$
$$\alpha^{u}(\tau,\gamma)=\min\left\{1,\exp\big(\Phi(u,\tau;y)-\Phi(u,\gamma;y)\big)\cdot w(\tau,\gamma)\cdot\frac{\pi_0(\tau)q(\tau,\gamma)}{\pi_0(\gamma)q(\gamma,\tau)}\right\}.$$

Fix jump parameter $\beta\in(0,1]$, and generate $(u^{(k)},\tau^{(k)})_{k\geq 0}$ as follows:

Then $(u^{(k)},\tau^{(k)})_{k\geq 0}$ is a Markov chain which is invariant with respect to $\mu^y(du,d\tau)$.

4.4 Numerical Results

We perform a variety of numerical experiments to illustrate the performance of the hierarchical algorithm described in section 4.3. We focus on three different forward models. The first is pointwise observations composed with the identity – the simplicity of this model allows us to probe the behavior of the algorithm at low

Algorithm 2 Metropolis-within-Gibbs

1. Set $k = 0$ and pick initial state $(u^{(0)}, \tau^{(0)}) \in X \times \mathbb{R}^+$.
2. Propose $v^{(k)} = m_0 + (1 - \beta^2)^{1/2}(u^{(k)} - m_0) + \beta\xi^{(k)}$, where $\xi^{(k)} \sim N(0, \mathcal{C}_\tau)$.
3. Set $u^{(k+1)} = v^{(k)}$ with probability $\alpha^{\tau^{(k)}}(u^{(k)}, v^{(k)})$, independently of $(u^{(k)}, v^{(k)})$.
4. Set $u^{(k+1)} = u^{(k)}$ otherwise.
5. Propose $\gamma^{(k)} \sim Q(\tau^{(k)}, \cdot)$.
6. Set $\tau^{(k+1)} = \gamma^{(k)}$ with probability $\alpha^{u^{(k+1)}}(\tau^{(k)}, \gamma^{(k)})$, independently of $(\tau^{(k)}, \gamma^{(k)})$.
7. Set $\tau^{(k+1)} = \tau^{(k)}$ otherwise.
8. $k \to k + 1$ and return to 2.

computational cost, and such models are also of interest in applications such as image reconstruction – see for example [8, 124] and the references therein. The other two, groundwater flow and EIT, are physical models which have previously been studied extensively, including study of non-hierarchical Bayesian level set methods [45, 72]. A review of studies on inverse problems associated with EIT is given in [20].

4.4.1 Identity Map

The first inverse problem is based on reconstruction of a piecewise constant field from noisy pointwise observations.

The forward model

Let $D = [0, 1]^2$ and define a grid of observation points $\{x_j\}_{j=1}^J \subseteq D$. Let $Z = L^p(D)$ for some $1 \leq p < \infty$ and let $Y = \mathbb{R}^J$. The forward operator $S : Z \to Y$ is defined by

$$S(z) = (z(x_1), \ldots, z(x_J)).$$

We are then interested in finding z, given the prior information that it is piecewise constant, and taking a number of known prescribed values. Let $\mathcal{G} = S \circ F : X \times \mathbb{R}^+ \to Y$. We reconstruct (u, τ) and hence $z = F(u, \tau)$. The map S is not continuous, and so Assumptions 4.2.3 do not hold. However Proposition 4.6.2 shows that the map $\mathcal{G}$ is uniformly bounded, and almost-surely continuous under the priors considered. From this the conclusions of Proposition 4.6.1 follow, and it is possible to deduce the conclusions of Theorem 4.2.4.

Simulations and results

We study the effect of different length scales, for both hierarchical and non-hierarchical methods, demonstrating the advantages of the former over the latter. To this end we define $\tau_i^\dagger = 5i$, $i = 1, \ldots, 10$, and generate 10 different true level set fields $u_i^\dagger \sim \mu_0^{\tau_i^\dagger}$ on a mesh of $2^{10} \times 2^{10}$ points. This leads to 10 sets of data y_i, given by

$$y_i = \mathcal{G}(u_i^\dagger, \tau_i^\dagger) + \eta_i, \quad \eta_i \sim N(0, \Gamma) \text{ i.i.d.}$$

where we take the noise covariance $\Gamma = \text{diag}(0.2^2)$. The level set map F is defined such that there are 3 phases, taking the constant values $1, 3$ and 5. The mean relative error on the generated data sets ranges from 6% to 9%.

One of the motivations for developing a hierarchical method is that little knowledge may be known a priori about the length scale associated with the unknown field. We therefore sample from each hierarchical posterior distribution associated with each y_i using a variety of initial values for the length scale parameter. This allows us to check that, computationally, we can recover a good approximation to the true length scale even if our initial guess is poor. Specifically, for each set of data we run 10 hierarchical MCMC simulations started at the different values of $\tau = \tau_k^\dagger$, giving a total of 100 hierarchical MCMC chains. For all chains we place a relatively flat prior of $N(20, 10^2)$ on τ, and fix the smoothness parameter $\alpha = 5$ on the prior for u.

We also wish to compare how the hierarchical method compares with the non-hierarchical method. We therefore look at the 10 different posterior distributions that arise from each set of data y_i when using each of 10 fixed prior inverse length scale $\tau_k^\dagger$, which gives another 100 MCMC chains.

We perform all sampling on a mesh of $2^7 \times 2^7$ points to avoid an inverse crime, and the observation grid $\{x_i\}_{i=1}^{100}$ is taken to be a uniformly spaced grid of 100 points. We produce 5×10^6 samples for each chain, and discard the first 10^6 samples as burn-in when calculating quantities of interest.

In Figure 4.2 we look at the recovery of the true value of τ with the hierarchical method. For large enough τ_0, the mean of τ after the burn-in period is roughly constant with respect to varying the initialization point, for each posterior. This makes sense from a theoretical point of view since these means arise from the same posterior distribution, for a fixed truth, but it is also reassuring from a computational point of view since the output is close to independent of the initial guess for the

length scale. There does however appear to be an issue with initializing the value of τ at too low a value, with the value τ tending to get stuck far from the truth when initialized at $= 5$. This effect has been detected in several other experiments and models – initializing the value of τ much lower than the true inverse length can cause the parameter to become stuck in a local minimum. Such an effect has not been observed however when the parameter is initialized significantly larger than the true value. Table 4.1 shows that recovery of the true value of τ is very good for $\tau^\dagger \leq 35$, though becomes slightly worse for larger values of $\tau^\dagger$. The means here are calculated without the $\tau_0 = 5$ sample means since they are clearly outliers for most of the posteriors. One possible explanation for the lack of recovery in the cases $\tau^\dagger = 40$, 45 and 50 is to do with the structure of the observation map S. The observation grid has a length scale associated with it, related to distances between observation points, and so issues could arise when trying to detect the length scale of the field that is significantly shorter than this. Additionally, the length scales $1/\tau$ are closer for larger τ and so it may be more difficult to distinguish between particular values.

For brevity we now focus on the case where $\tau^\dagger = 15$. The traces of the values of τ along the hierarchical chains corresponding to this truth is shown in Figure 4.3. After approximately 10^6 samples, all chains have become centred around the true length scale. This convergence appears to be roughly linear for each chain.

Figure 4.4 shows the push forwards of the sample means from the different chains under the level set map, that is, approximations of $F(\mathbb{E}(u), \mathbb{E}(\tau))$. This figure also shows approximations of $\mathbb{E}(F(u, \tau))$ and typical samples of $F(u, \tau)$ coming from the different chains. We see that these conditional means for the hierarchical method appear to agree with one other. This is reassuring for the reason mentioned above – they are all estimates of the mean of the same distribution. The figures for the non-hierarchical posteriors admit greater variation, especially near the boundary for higher values of τ. Moreover, not all inclusions are detected when the length scale parameter is taken to be $\tau = 5$. Note that the mean from the hierarchical posterior agrees closely with that from the non-hierarchical posterior using the fixed true length-scale $\tau = 15$. Additionally, even though the means are reasonable approximations to the truth in most cases, the typical samples are much worse when using the non-hierarchical method with an incorrect length scale parameter.

We can also consider the sample variance of the pushforward of the samples by the level set map, i.e. approximations of the quantity $\mathrm{Var}(F(u, \tau))$. In Figure 4.5 we show this quantity for both the hierarchical and non-hierarchical priors. Note

Table 4.1: (Identity model) The value of τ used to create the data y_i, and the mean value of τ across the MCMC chains and the different initial values of τ.

$\tau^\dagger$	Mean sample mean of τ
5	6.10
10	10.0
15	15.5
20	21.8
25	24.8
30	30.0
35	35.4
40	44.6
45	50.8
50	40.6

that for the non-hierarchical priors, the variance increases both at the boundary and away from the observation points for larger values of τ. Variance is also higher along the interfaces and within the central phase, since points in these locations are more likely to switch between all three phases. The hierarchical approximations all appear to agree. Whilst the hierarchical means are very similar to the non-hierarchical means using the true length scale, as seen in Figure 4.4, the hierarchical variances are smaller away from the observation points.

Additionally, we look at the level set function u itself in Figure 4.6. In these plots we rescale the level set function by $\tau^{\alpha-d/2} = \tau^4$ so that they are all of approximately the same amplitude. The means for both the hierarchical and non-hierarchical methods are again quite similar to one another, though the difference between the typical samples is much more stark.

Finally, in Figure 4.7, we look at the joint densities of the inverse length scale parameter τ and first five Karhunen-Loève (KL) modes of the level set function u.[2] Non-trivial correlations are evident between τ and each of these modes, with the support of the densities appearing non-convex. This is likely related to the non-linear scaling between the length-scale and the amplitude of the level-set function under the prior. Conversely the KL modes, whilst still correlated with one-another other, have simpler joint densities. Note, also, that the posterior on the length scale is centred close to the true value of the inverse length scale parameter τ.

[2]KL modes are the eigenfunctions of the covariance operator, here ordered by decreasing eigenvalue.

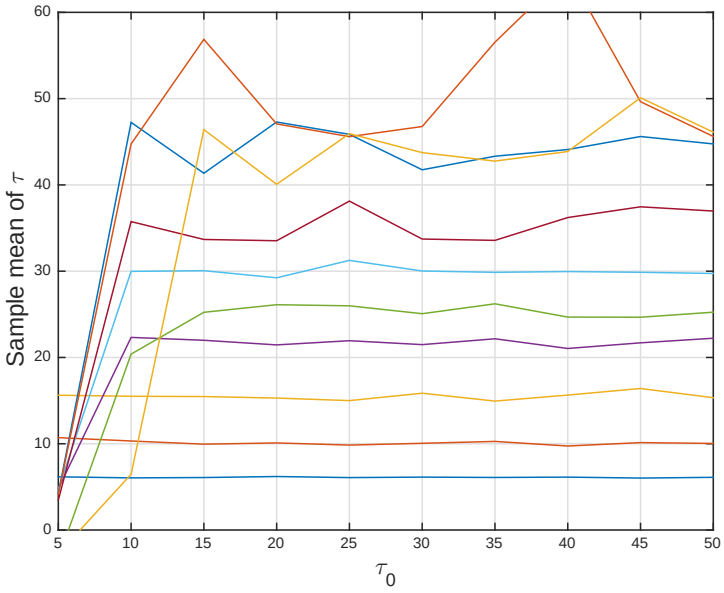


Figure 4.2: (Identity model) The sample mean of τ along each hierarchical MCMC chain, against the initial value of τ. The different curves arise from using different data y_i.

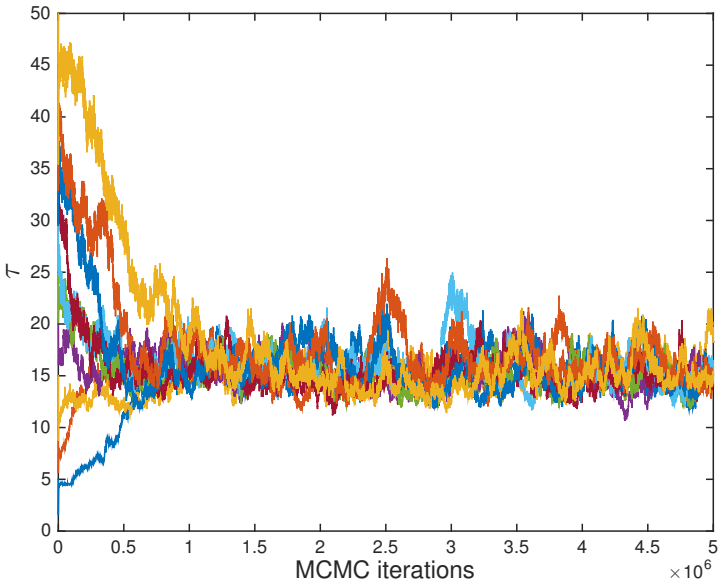
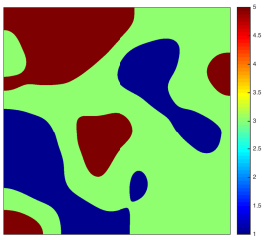
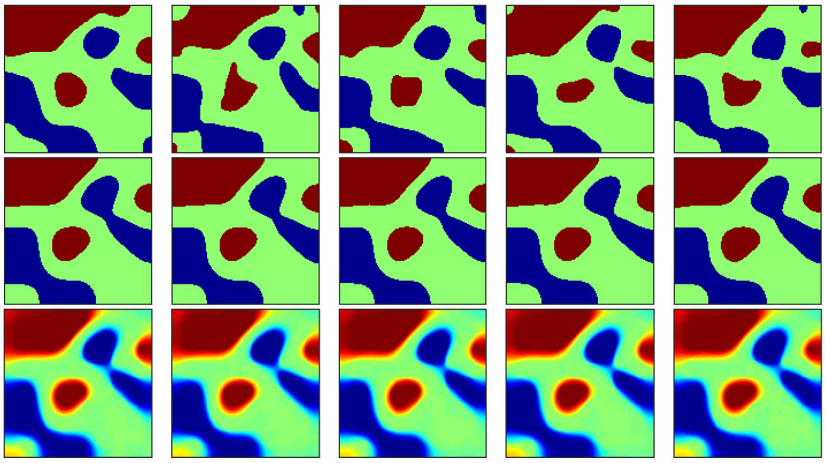


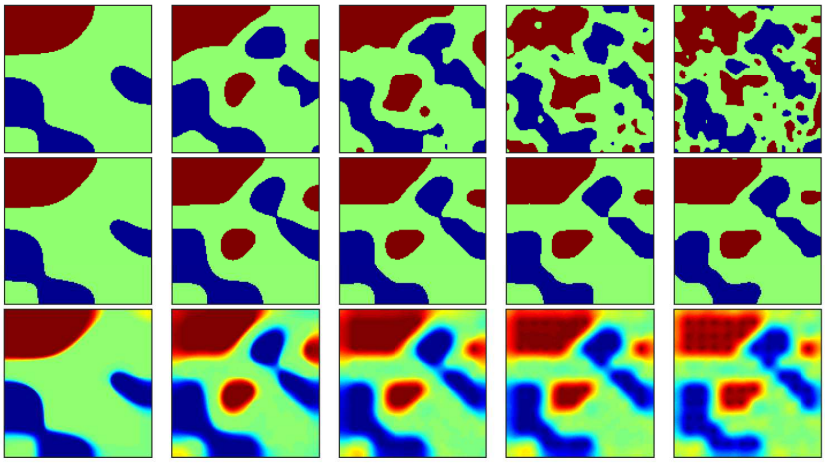
Figure 4.3: (Identity model) The trace of τ along the MCMC chain, when initialized at the 10 different initial values. True inverse length scale is $\tau = 15$.



(a) The true geometric field used to generate the data y, with true inverse length scale $\tau = 15$.



(b) (Top) Representative samples of $F(u, \tau)$ under the hierarchical posterior. (Middle) Approximations of $F(\mathbb{E}(u), \mathbb{E}(\tau))$. (Bottom) Approximations of $\mathbb{E}(F(u, \tau))$. From left-to-right, τ is initialized at $\tau = 5, 15, 25, 35, 45$.



(c) As in (b), using the non-hierarchical method. From left-to-right, τ is fixed at $\tau = 5, 15, 25, 35, 45$.

Figure 4.4: Simulations for the identity model.

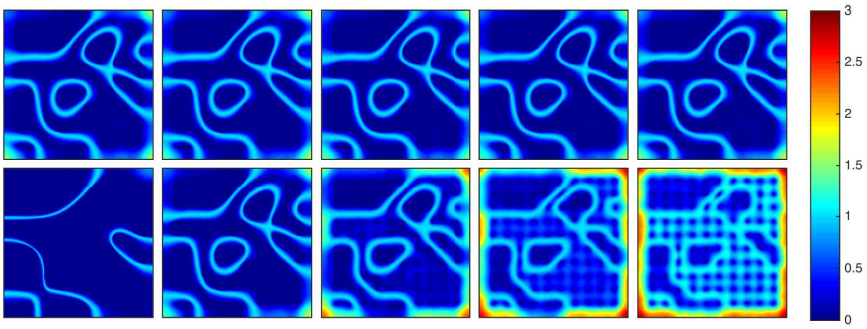
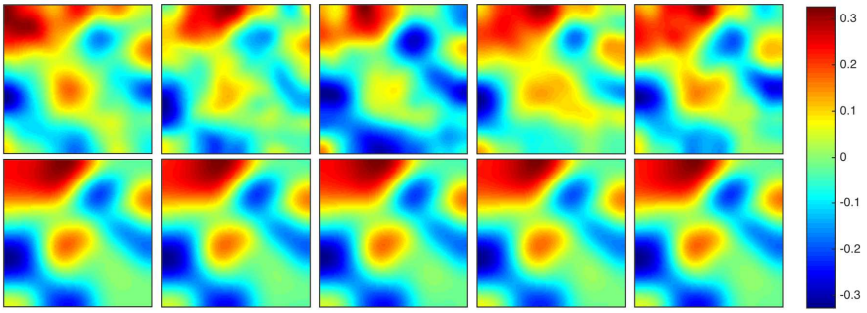
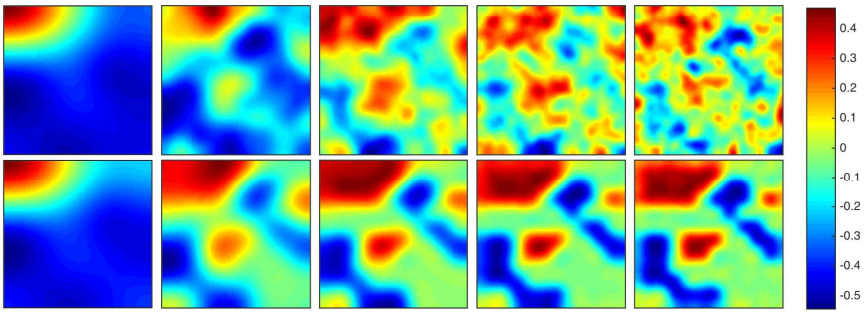


Figure 4.5: (Identity model) Approximations of $\mathrm{Var}(F(u,\tau))$ using the hierarchical (top) and fixed (bottom) priors, initialized or fixed at $\tau = 5, 15, 25, 35, 45$, from left-to-right. True inverse length scale is $\tau = 15$.



(a) (Top) Representative samples of the rescaled level-set function $\tau^4 \cdot u$ and (bottom) approximations of $\mathbb{E}(\tau^4 \cdot u)$ using the hierarchical method. From left-to-right, τ is initialized at $\tau = 5, 15, 25, 35, 45$.



(b) As in (a), using the non-hierarchical method. From left-to-right, τ is fixed at $\tau = 5, 15, 25, 35, 45$.

Figure 4.6: (Identity model) Respresentative samples and sample means of the level set function. The rescaling τ^4 means that the above quantities have the same approximate amplitude. True inverse length scale is $\tau = 15$.

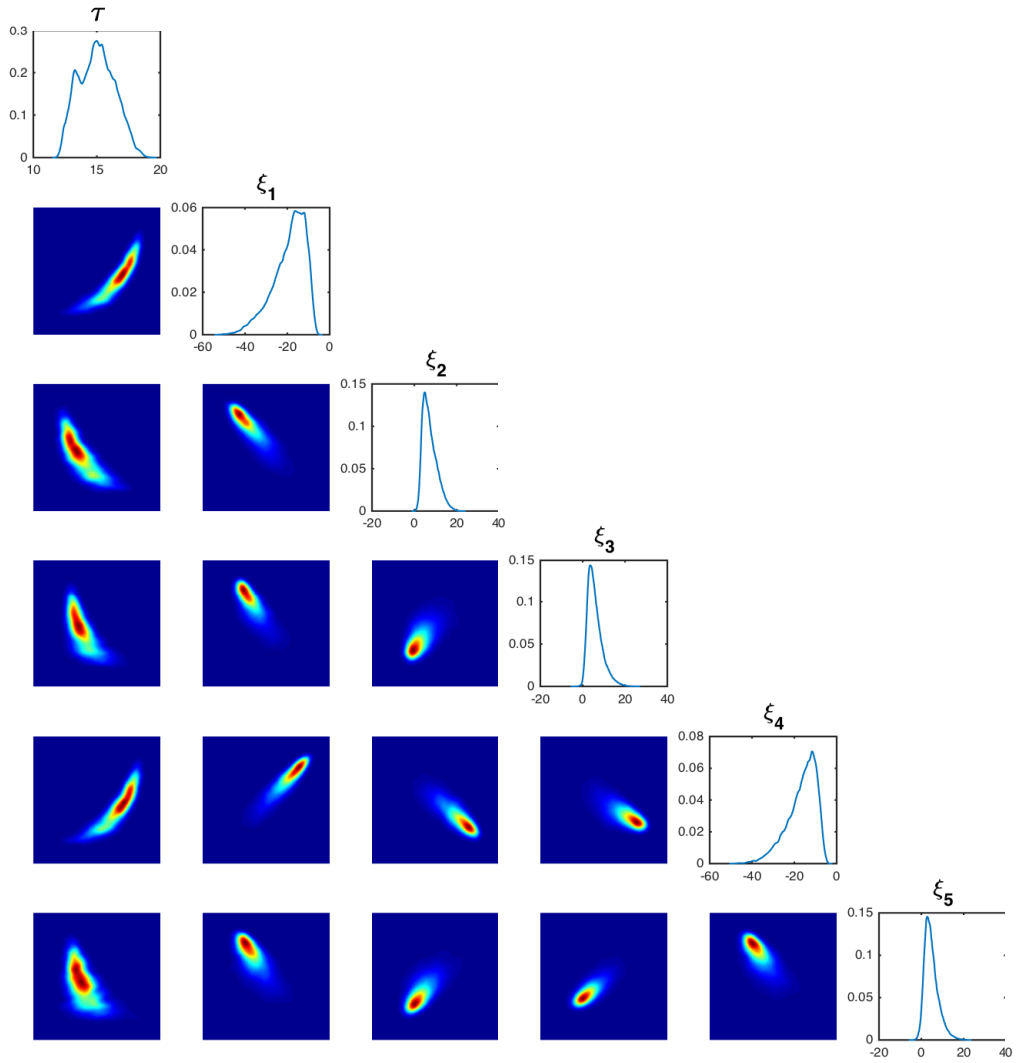


Figure 4.7: (Identity model) (diagonal) Empirical densities of τ and the first five KL modes of u. (off-diagonal) Empirical joint densities. True inverse length scale is $\tau = 15$.

4.4.2 Identification of Geologic Facies in Groundwater Flow

The identification of geologic facies in subsurface flow applications is a common example of a large scale inverse problem that involves the recovery of unknown interfaces. In the case of groundwater flow, for example, the inverse problem concerns the recovery of the interface between regions with different hydraulic conductivity given measurements of hydraulic head. Geometric inverse problems of this type have recently received a lot of attention by the research community [99,100,117,139]. Indeed, it has been recognized that the geometry determined by the aforementioned interfaces constitutes one of the main sources of uncertainty that must be quantified and reduced by means of Bayesian inversion.

In the context of groundwater flow, the identification of interfaces between regions associated with different types of geological properties can be posed as the recovery of a piecewise constant conductivity field parameterized with a level set function. A fully Bayesian level set framework for the solution of the aforementioned type of inverse problems has been recently developed in [72]. The MCMC method applied in [72] performs well when the prior of the level set function properly encodes the intrinsic length-scales of the unknown interfaces. Clearly, in practical applications such length-scales are most likely unknown and their incorrect specification may result in inaccurate and uncertain estimates of the unknown interfaces. The purpose of this section is to show that the proposed hierarchical Bayesian framework enables us to determine an optimal length-scale in the prior of the level set function which, in turn, captures more accurately the intrinsic length-scale of the unknown interface.

The forward model

We are interested in the identification of a piecewise constant hydraulic conductivity, denoted by κ, of a two-dimensional confined aquifer whose physical domain is $D = [0,6] \times [0,6]$. We assume single-phase steady-state Darcy flow. The piezometric head, denoted by $h(x)$ ($x \in D$), which describes the flow within the aquifer can be modeled by the solution of [10]

$$-\nabla \cdot \kappa \nabla h \;\; = f \qquad \text{in} \;\; D \tag{4.4.1}$$

where f represents sources/sinks and where boundary conditions need to be specified. For the present work we consider the setup from the Benchmark used in

[26,63,67–69,72]. In concrete, we assume that f is a recharge term of the form

$$f(x_1, x_2) = \begin{cases} 0 & \text{if} \quad 0 < x_2 \leq 4, \\ 137 & \text{if} \quad 4 < x_2 < 5, \\ 274 & \text{if} \quad 5 \leq x_2 < 6. \end{cases} \tag{4.4.2}$$

and we consider the following boundary conditions

$$\begin{aligned} h(x_1, 0) = 100, \qquad & \frac{\partial h}{\partial x_1}(6, x_2) = 0, \\ -\kappa \frac{\partial h}{\partial x_1}(0, x_2) = 500, \qquad & \frac{\partial h}{\partial x_2}(x_1, 6) = 0. \end{aligned} \tag{4.4.3}$$

We consider the inverse problem of recovering κ from observations $\{\ell_j(h)\}_{j=1}^{64}$ of h given by (4.4.1)-(4.4.3). We assume we have smoothed point observations given by

$$\ell_j(h) = \int_D \frac{1}{2\pi\varepsilon^2} e^{-\frac{1}{2\varepsilon^2}(x-q_i)^2} h(x)\, \mathrm{d}x$$

where $\varepsilon > 0$ and $\{q_i\}_{i=1}^{64} \subseteq D$ is a grid of 64 observation points equally distributed on D. Let $Z = L^p(D)$ for some $1 \leq p < \infty$ and $Y = \mathbb{R}^{64}$. Then the forward map $S : Z \to Y$ is given by

$$h \mapsto (\ell_1(h), \ldots, \ell_{64}(h)).$$

We assume that each κ_i in the definition of the level set map F is strictly positive. The image of F is contained in the set of bounded fields on D bounded below by $\min_i \kappa_i > 0$. In [72] the map S is shown to be continuous and uniformly bounded on such fields, with respect to $\|\cdot\|_{L^p(D)}$ for some p, and so Assumptions 4.2.3 hold. As a consequence Theorem 4.2.4 applies directly.

Simulations and results

In the previous example we illustrate, with a simple model, the capabilities of the proposed framework to recover a specified true length-scale and a true level set function that defines a true discontinuous field from which synthetic data are generated. However, we must reiterate that, in practice, we wish to recover the true discontinuous field; the level set function is merely an artifact that we use for the parameterization of such a field. In practical applications the aim of the proposed hierarchical Bayesian level set framework is to infer a length-scale alongside with a level set function which, by means of expression (4.2.6), produces a discontinuous

field that captures the desired piecewise constant field as accurately as possible and, in particular, the intrinsic length-scale separation of the interfaces determined by the discontinuities of the true field. Therefore, in order to test our methodology in the applied setting of groundwater flow, rather than a true level set function, in this subsection we consider the true hydraulic conductivity $\kappa^\dagger$ whose logarithm is displayed in Figure 4.9 (top). This $\kappa^\dagger$ is defined such that it takes the constant values $e^{1.5}$, e^4 and $e^{6.5}$. This is channelized conductivity typical of fluvial environments and often used as Benchmarks for subsurface flow inversion [72, 100, 117, 139]. Note that the values that the conductivity can take on the three different regions differ by at least one of order of magnitude, due to the logarithmic transformation. While there is indeed an intrinsic length-scale in the channelized structure, this true conductivity field does not come from a specified level set prior.

Synthetic data are generated by means of

$$y = (\ell_1(h^\dagger), \dots, \ell_{64}(h^\dagger)) + \eta, \quad \eta \sim N(0, \Gamma) \text{ i.i.d.}$$

where $h^\dagger$ is the solution to (4.4.1)-(4.4.3) for $\kappa = \kappa^\dagger$. Equations (4.4.1)-(4.4.3) have been solved with cell-centered finite differences [9]. In order to avoid inverse crimes, synthetic data are generated on a grid finer (160x160 cells) than the one used for the inversion (80x80 cells). In addition, Γ is a diagonal matrix given by $\Gamma_{i,i} = 0.0175\ell_i(h^\dagger)$. In other words, we add noise that corresponds to 1.75% of the size of the noise-free observations.

We consider a Gaussian prior $N(35, 10^2)$ for τ. We then apply the hierarchical MCMC method from subsection 4.3.3 initialized with the following six different choices of $\tau = 1, 10, 30, 50, 70, 90$ and a sample of the prior (with that given τ) of the level set function u. We thus produce six MCMC chains of length 4×10^6 and discard the first 10^6 as burn-in for the computation of quantities of interest. The trace plots of τ are displayed in Figure 4.8 from which we clearly observe that all chains, regardless of their initial point, seem to stabilize and produce samples around $\tau = 18$. In the middle-top of Figure 4.9 we display the logarithm of some representatives samples of $F(u, \tau)$ under the hierarchical posterior. The middle-middle panel of Figure 4.9 shows the logarithm of $F(\mathbb{E}(u), \mathbb{E}(\tau))$, i.e., the pushforward of the posterior means obtained using the hierarchical method. The middle-bottom of Figure 4.9 displays the logarithm of the approximations of $\mathbb{E}(F(u, \tau))$. That is, the expected value of the pushforward samples under the posterior. The aforementioned results corresponds to five MCMC chains with τ initialized $\tau = 10, 30, 50, 70, 90$ (the results for $\tau = 1$ have been omitted). Similarly, Figure 4.10 (top) shows the approxima-

tions of the variance of the pushforward samples of the posterior, i.e. $\mathrm{Var}\big(F(u,\tau)\big)$. Clearly, both $\mathbb{E}(F(u,\tau))$ and $F(\mathbb{E}(u),\mathbb{E}(\tau))$ result in fields that provide a reasonable approximation of the true field. Note that, as expected, the largest uncertainty in the distribution of the pushforward samples is around the interface between the regions with different conductivity. In Figure 4.11 we show representative samples of u (top-top) and approximations to $\mathbb{E}(u)$ (top-bottom). In these plots, as before, we rescale the level set function by $\tau^{\alpha-d/2} = \tau^4$ so that they are all of approximately the same amplitude. In Figure 4.12 we display the empirical densities of τ and the first five KL modes of u. A key observation is that, although the true hydraulic conductivity is not generated by thresholding a Gaussian random field, and hence there is no "true" length scale, the posterior nonetheless settles on a narrow range of values of τ which are consistent with the data.

From the aforementioned results we can also clearly see that the hierarchical MCMC algorithm produces similar outcomes regardless of the initialization of the inverse of the length-scale τ, reflecting ergodicity of the Markov chain. The results from $\tau = 1$ are not shown but they are very similar to the ones from other chains. As with the results from the previous subsection, the similarity in outcomes between all six chains is not surprising as these are aimed at sampling from the same posterior distribution; but the fact that this posterior distribution on τ concentrate near to a single value is of particular interest because it shows that the true unknown field has an intrinsic length-scale, even though it was no constructed via the map $F(u,\tau)$. Furthermore, this similarity of outcomes between chains showcases the main advantage of the proposed framework with respect to the non-hierarchical one. Indeed, as stated earlier, the proposed method has the ability to recover a distribution for the intrinsic length-scale which gives rise to reasonably accurate estimates (i.e. $F(\mathbb{E}(u),\mathbb{E}(\tau))$ and $\mathbb{E}(F(u,\tau))$) of the true unknown field. We now present the numerical results from applying a non-hierarchical MCMC algorithm in which the inverse of length-scale τ is fixed. We consider again six MCMC chains as before with the (now fixed) values of $\tau = 1, 10, 30, 50, 70, 90$ that we used to initialized the hierarchical chains used before. Analogous results to the ones presented for the hierarchical method can be found in the bottom panels of Figure 4.9 as well as the bottom of Figures 4.10 and 4.11. Clearly, the lack of properly prescribing the intrinsic length-scale in the non-hierarchical method results in inaccurate estimates of the true field. We clearly observe that for $\tau \geq 30$ the estimates of the truth given by $F(\mathbb{E}(u),\mathbb{E}(\tau))$ and $\mathbb{E}(F(u,\tau))$ are substantially inaccurate and the uncertainty measured by $\mathrm{Var}\big(F(u,\tau)\big)$ is large. The non-hierarchical MCMC for $\tau = 1$ did not converge; the results are not shown. The non-hierarchical MCMC only provides rea-

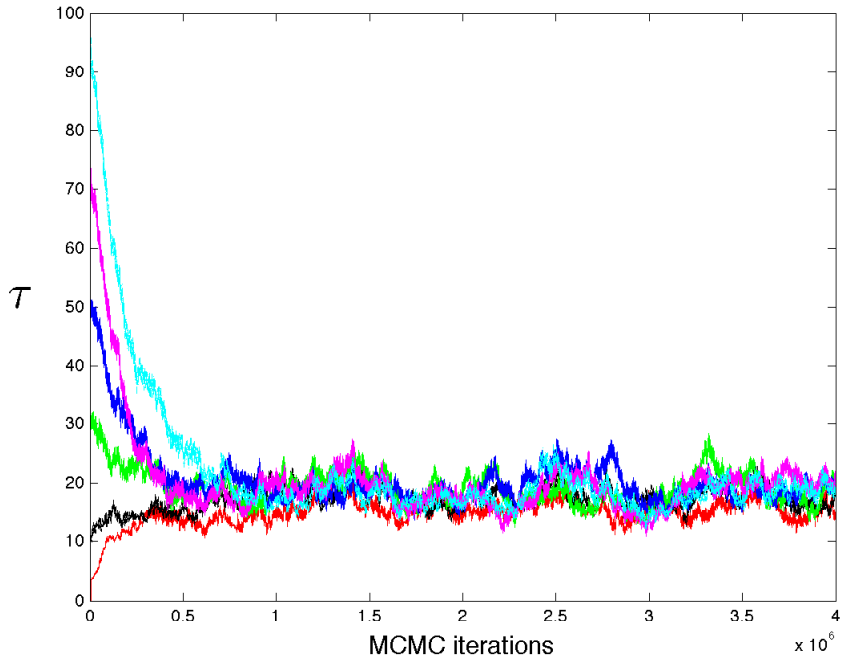


Figure 4.8: (Groundwater flow model) Trace plots of τ obtained from six hierarchical MCMC chains.

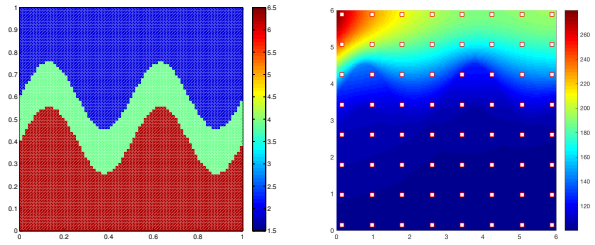
sonable estimates for $\tau = 10$ and $\tau = 30$. However, we can visually appreciate that these results are still suboptimal when compared to the results from the hierarchical framework.

4.4.3 Electrical Impedance Tomography

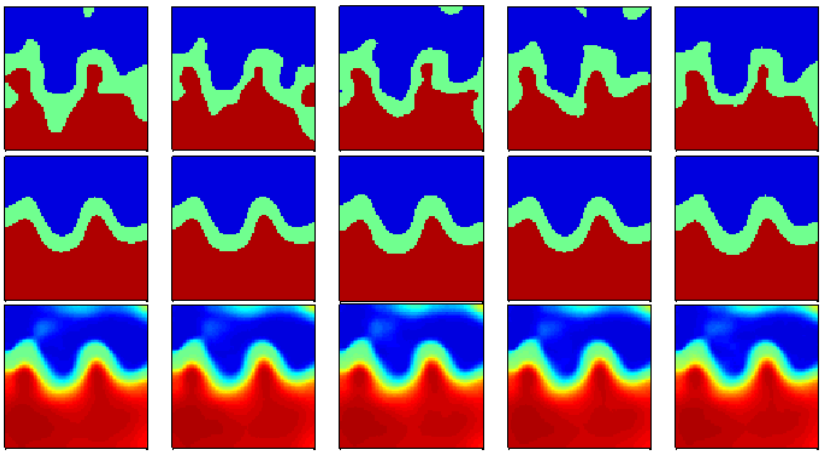
Finally we consider the electrical impedance tomography (EIT) problem. This problem has previously been approached with a non-hierarchical Bayesian level set method [45]. In this subsection we show that the hierarchical approach outperforms the non-hierarchical approach in the case where the true conductivity is a binary field, given the same number of forward model evaluations.

The forward model

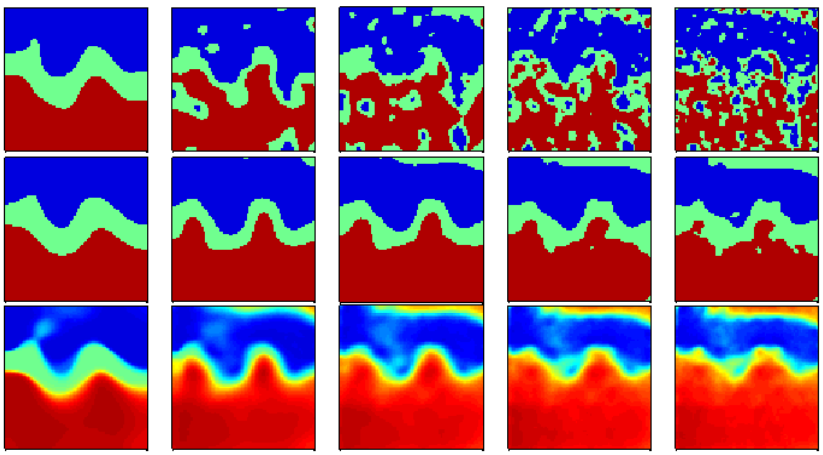
EIT is an imaging technique which attempts to infer the internal conductivity of a body from boundary voltage measurements. Typical applications include medical imaging, as well as subsurface imaging where it is known as electrical resistivity tomography (ERT). We utilize the complete electrode model (CEM), proposed in [128].



(a) (Left) Logarithm of the true hydraulic conductivity field used to generate the data y. (Right) True pressure field, and the grid of observation points.



(b) (Top) Logarithm of representative samples of $F(u, \tau)$ under the hierarchical posterior. (Middle) Logarithm of the approximations of $F(\mathbb{E}(u), \mathbb{E}(\tau))$. (Bottom) Logarithm of the approximations of $\mathbb{E}(F(u, \tau))$. From left-to-right, τ is initialized at $\tau = 10, 30, 50, 70, 90$.



(c) As in (b), using the non-hierarchical method. From left-to-right, τ is fixed at $\tau = 10, 30, 50, 70, 90$.

Figure 4.9: Simulations for the groundwater flow model.

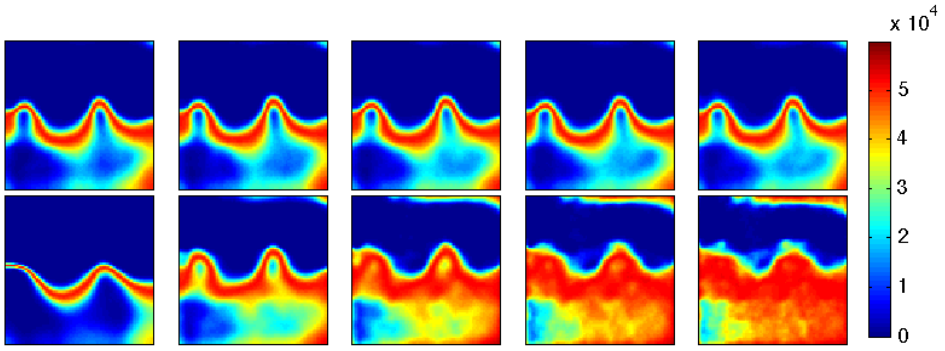
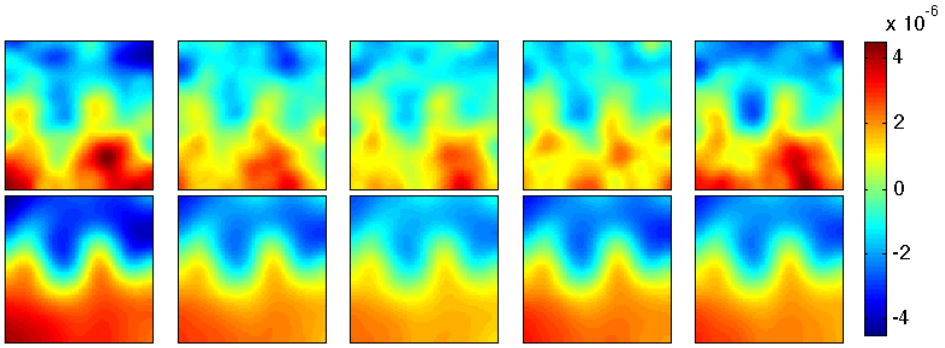
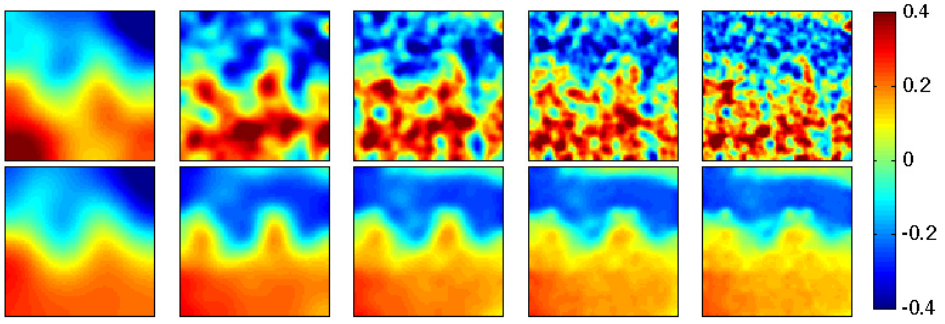


Figure 4.10: (Groundwater flow model) Approximations of $\mathrm{Var}\big(F(u,\tau)\big)$ using the hierarchical (top) and the non-hierarchical (bottom) MCMC.



(a) (Top) Representative samples of the rescaled level-set function $\tau^4 \cdot u$ and (bottom) approximations of $\mathbb{E}(\tau^4 \cdot u)$ using the hierarchical method. From left-to-right, τ is initialized at $\tau = 10, 30, 50, 70, 90$.



(b) As in (a), using the non-hierarchical method. From left-to-right, τ is fixed at $\tau = 10, 30, 50, 70, 90$.

Figure 4.11: (Groundwater flow model) Respresentative samples and sample means of the level set function. The rescaling τ^4 means that the above quantities have the same approximate amplitude. True inverse length scale is $\tau = 15$.

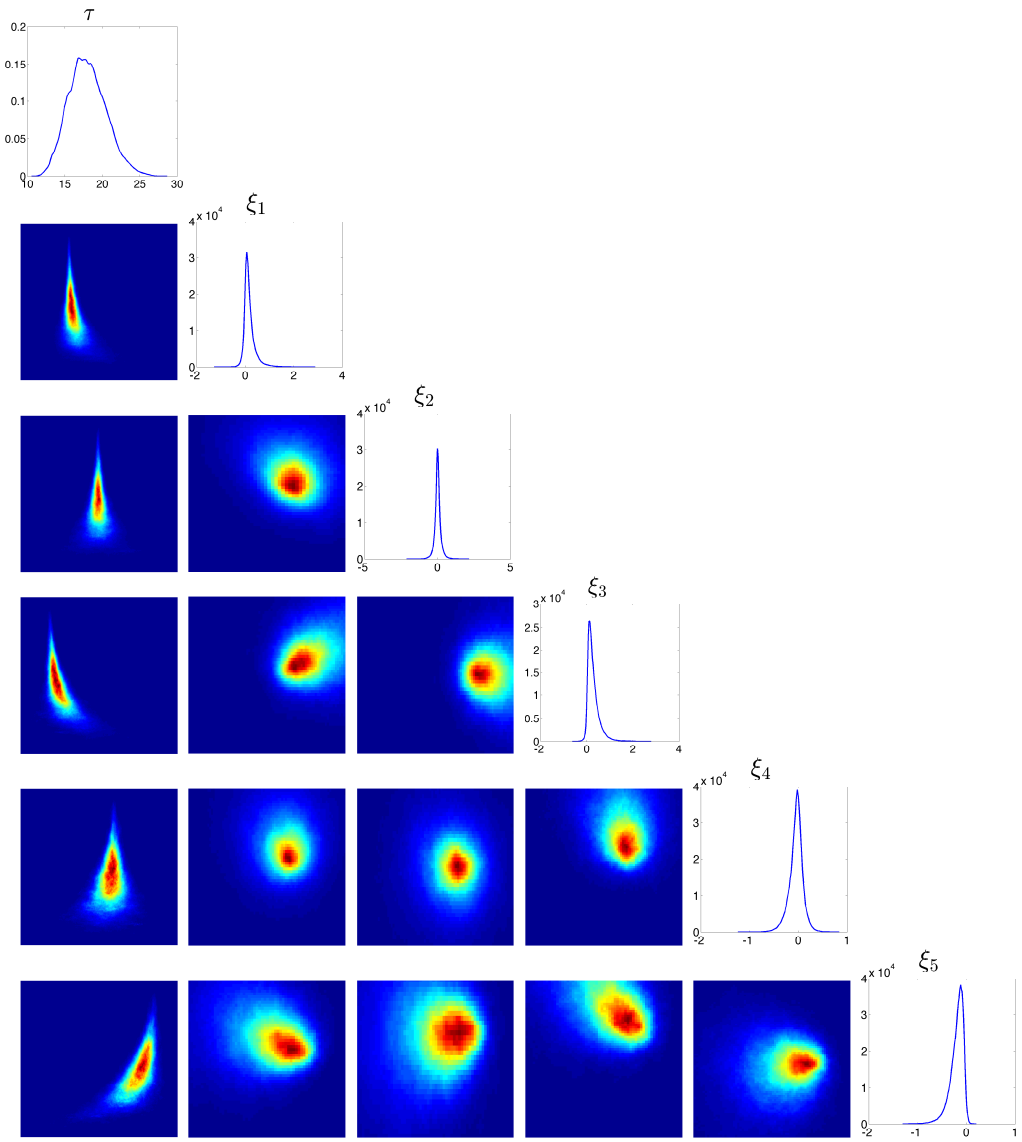


Figure 4.12: (Groundwater flow model) (diagonal) Empirical densities of τ and the first five KL modes of u. (off-diagonal) Empirical joint densities.

This is a physically accurate model which has been shown to agree with experimental data up to measurement precision. The strong form of the PDE governing the model is given by

$$\begin{cases} -\nabla \cdot (\sigma(x)\nabla v(x)) = 0 & x \in D \\ \int_{e_l} \sigma \dfrac{\partial v}{\partial n} \, \mathrm{d}S = I_l & l = 1, \ldots, L \\ \sigma(x)\dfrac{\partial v}{\partial n}(x) = 0 & x \in \partial D \setminus \bigcup_{l=1}^{L} e_l \\ v(x) + z_l \sigma(x) \dfrac{\partial v}{\partial n}(x) = V_l & x \in e_l, l = 1, \ldots, L. \end{cases}$$

Here $D \subseteq \mathbb{R}^2$ is the domain and $\{e_l\}_{l=1}^L \subseteq \partial D$ are electrodes on the boundary upon which currents $\{I_l\}_{l=1}^L$ are injected and voltages $\{V_l\}_{l=1}^L$ are read. The numbers $\{z_l\}_{l=1}^L$ represent the contact impedances of the electrodes. The field σ represents the conductivity of the body and v represents the potential within the body. It should be noted that the solution of this PDE comprises both a potential $v \in H^1(D)$ and a vector $\{V_l\}_{l=1}^L$ of boundary voltage measurements.

The inverse problem we consider is the recovery of σ from a sequence of boundary voltage measurements. A number of (linearly independent) current stimulation patterns $\{I_l\}_{l=1}^L$ may be performed to provide more information; we assume that we perform the maximum $M = L - 1$ measurements. Let $Z = L^p(D)$ for some $1 \leq p < \infty$ and $Y = \mathbb{R}^J$ where $J = LM$. We can concatenate the boundary voltage measurements arising from different stimulation patterns to yield a map $S : Z \to Y$,

$$\sigma \mapsto (V^{(1)}, V^{(2)}, \ldots, V^{(M)})$$

where $V^{(m)} = \{V_l^{(m)}\}_{l=1}^L \in \mathbb{R}^L$, $m = 1, \ldots, M$.

For the experiments we work on a circular domain $D = \{x \in \mathbb{R}^2 \mid |x| < 1\}$. 16 electrodes are spaced equally around the boundary providing 50% coverage. All contact impedances are taken to be $z_l = 0.01$. Adjacent electrodes are stimulated with a current of 0.1, so that the matrix of stimulation patterns $I = (I^{(j)})_{j=1}^{15} \in$

$\mathbb{R}^{16\times 15}$ is given by

$$I = 0.1 \times \begin{pmatrix} +1 & 0 & \cdots & 0 \\ -1 & +1 & \cdots & 0 \\ 0 & -1 & \ddots & 0 \\ \vdots & \vdots & \ddots & +1 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

We define our forward map $\mathcal{G} : X \times \mathbb{R}^+ \to \mathbb{R}^J$ by $\mathcal{G} = S \circ F$. As in the groundwater flow example, assume that each κ_i in the definition of the level set map is strictly positive. We do not have a continuity result for the map S on L^p for any $1 \leq p < \infty$. However the almost-sure continuity of the map $\mathcal{G}$ can be seen via a modification of the proof of Proposition 3.5 in [45] to include the parameter τ; this modification is almost identical to the proof of Proposition 4.6.1 given in the appendix. The uniform boundedness of $\mathcal{G}$ follows from a result in [45] similarly. Hence as was the case with the identity map example, the conclusions of Proposition 4.6.1 follow, and we can deduce the conclusions of Theorem 4.2.4.

Simulations and results

We fix a true conductivity $\sigma^\dagger$, shown in Figure 4.14. As with the groundwater flow experiments, this is constructed explicitly and does not have a true value of τ associated with it. We generate data y as

$$y = S(\sigma^\dagger) + \eta, \quad \eta \sim N(0, \Gamma)$$

where we take the noise covariance $\Gamma = \text{diag}(0.0002^2)$. The mean relative error on the generated data is approximately 12%. The data is generated using a mesh of 43264 elements and simulations are performed used a mesh of 10816 elements, in order to avoid an inverse crime. All field samples are defined on the square $[-1, 1]^2$ and restricted to the domain D. This has the advantage of allowing for efficient sampling via the Fast Fourier Transform, though has the drawback of introducing possibly non-trivial boundary effects on the domain; no such effects are observed in our problem, however.

The level set map F is defined such that there are 2 phases, taking the constant values 1 and 10. We take the prior field mean to be zero, so that in this case F (and hence Φ) becomes independent of τ. Thus a forward model evaluation is not required for the Gibbs update of τ, and each sample of (u, τ) using the hierarchical method

costs virtually the same as one of u using the non-hierarchical method.

Similarly to the previous experiments, we initialize the hierarchical sampling from $\tau = 10, 30, 50, 70, 90$ to check for robustness of the method. We use a sharper prior on τ than was used previously, $\pi_0 = N(10, 3^2)$, in order to reduce the length of the burn-in period, given the expense of forward model evaluations. We fix the smoothness parameter $\alpha = 5$ in the prior for u. We again wish to compare how the hierarchical method compares with the non-hierarchical method. We therefore also look at the 5 different posterior distributions that arise when using each of 5 fixed prior inverse length scales $\tau = 10, 30, 50, 70, 90$, which gives another 5 MCMC chains. For both the methods we produce 4×10^6 samples for each chain, and discard the first 2×10^6 samples as burn-in when calculating quantities of interest.

The traces of the values of τ along the hierarchical chains are shown in Figure 4.13. With the exception of the chain initialized at $\tau = 10$, the chains converge to the sample approximate value of τ. Unlike in previous experiments, the traces have a relatively flat period before the approximate linear convergence to the common length scale. Initializing $\tau = 90$ requires an additional 10^6 samples to converge, over the other converging chains.

Figure 4.14 shows the push forwards of the sample means from different chains under the level set map, along with approximations of $\mathbb{E}(F(u, \tau))$ and typical samples of $F(u, \tau)$ coming from the different posteriors. In both the hierarchical and non-hierarchical methods, the chains initialized/fixed at $\tau = 10$ fail to recover the true conductivity, similarly to what was observed with the identity map experiments when initialising at $\tau = 5$. The other chains for the hierarchical method produce very similar results to one another, whilst the effect of fixing the length scale to be too short is apparent in the figures for the non-hierarchical method.

In Figure 4.15 we see approximations to $\mathrm{Var}(F(u, \tau))$ under the different posteriors. In both cases, variance is highest around the boundaries of the two inclusions. The difference between the hierarchical and non-hierarchical methods is more apparent here, with higher variance between the two inclusions when the length scale is fixed to be too short.

Again, we look at the level set function u itself in Figure 4.16. In these plots, as before, we rescale the level set function by $\tau^{\alpha - d/2} = \tau^4$ so that they are all of approximately the same amplitude. As in the previous experiments, there is noticeable contrast between the means for the hierarchical and non-hierarchical methods, and yet more contrast between the typical samples.

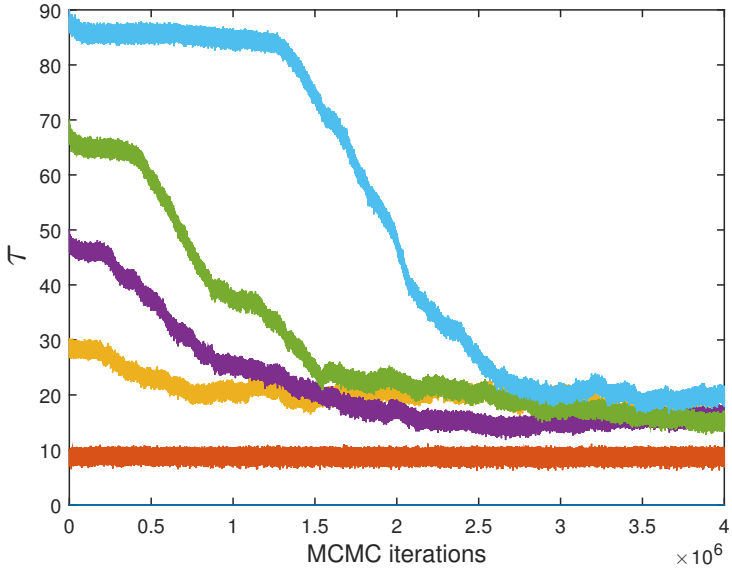


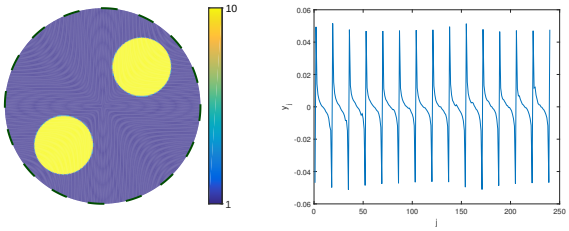
Figure 4.13: (EIT model) The trace of τ along the MCMC chain, when initialized at the 5 different values $\tau = 10, 30, 50, 70, 90$.

Finally, in Figure 4.17, we show the posterior densities on the inverse length scale and the first five KL modes, as well as correlations between them. As with the grounwater flow example, although there is no "true" inverse length scale, the data is sufficiently informative to define a small range of values for this parameter under the posterior.

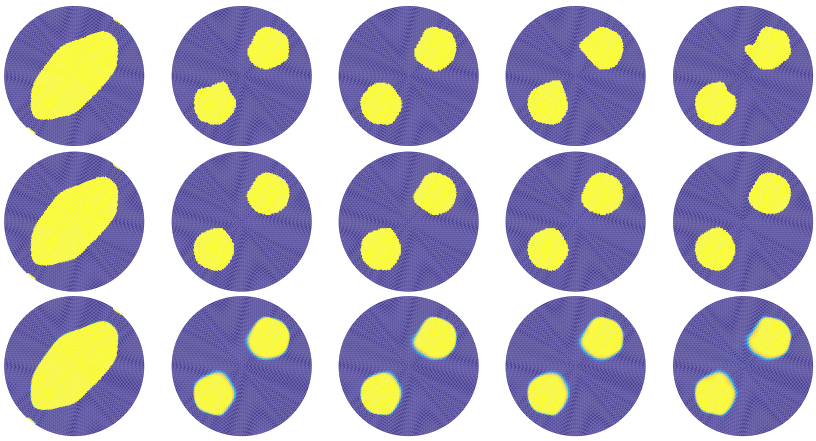
4.5 Conclusions

The level set method is an attractive approach to inverse problems for the detection of interfaces. Furthermore the Bayesian approach is particularly desirable when there is a need to quantify uncertainty. In this chapter we have shown that Bayesian level set inversion is considerably enhanced by a hierarchical approach in which the length scale of the underlying level set function is inferred from the data. We have demonstrated this by means of three examples of interest arising in, respectively, the information, physical and medical sciences; however many potential applications remain to be explored and this provides an interesting avenue for future work.

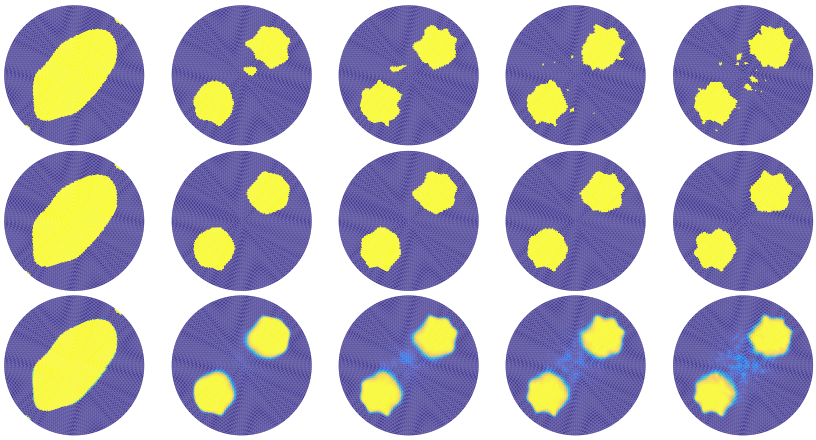
We also developed the theoretical underpinnings for our hierarchical method. Our work is based on a Metropolis-within-Gibbs approach which alternates between up-



(a) (Left) True conductivity field used to generate the data y. (Right) The entries y_i of the data vector y, plotted against i.



(b) (Top) Representative samples of $F(u, \tau)$ under the hierarchical posterior. (Middle) Approximations of $F(\mathbb{E}(u), \mathbb{E}(\tau))$. (Bottom) Approximations of $\mathbb{E}(F(u, \tau))$. From left-to-right, τ is initialized at $\tau = 10, 30, 50, 70, 90$.



(c) As in (b), using the non-hierarchical method. From left-to-right, τ is fixed at $\tau = 10, 30, 50, 70, 90$.

Figure 4.14: Simulations for the groundwater flow model.

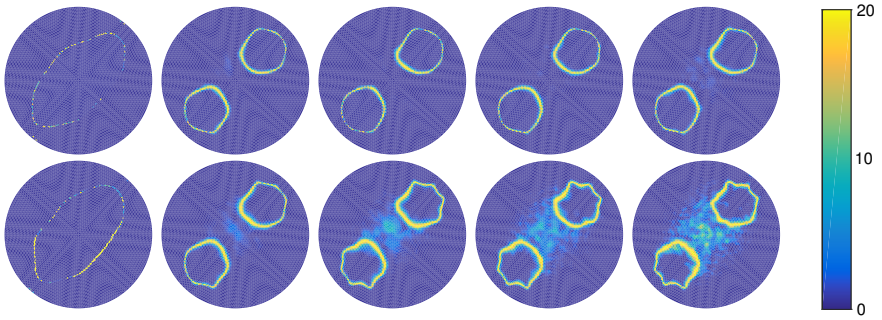
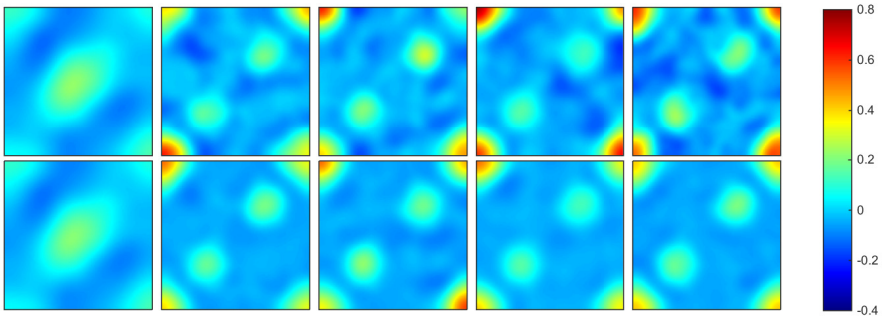
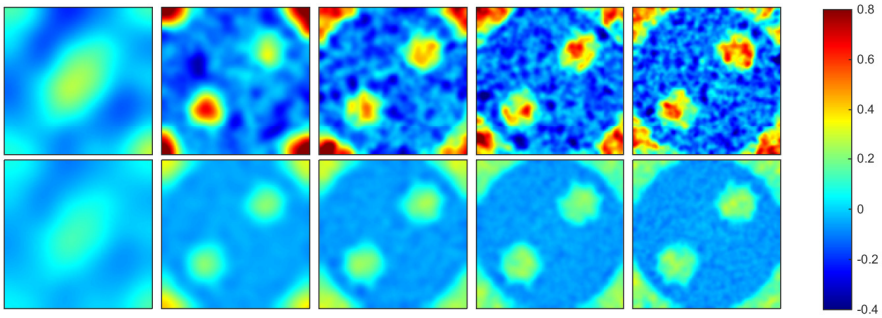


Figure 4.15: (EIT model) Approximations of $\mathrm{Var}(F(u,\tau))$ using the hierarchical (top) and fixed (bottom) priors, with τ initialized or fixed at $\tau = 10, 30, 50, 70, 90$, from left-to-right.



(a) (Top) Representative samples of the rescaled level-set function $\tau^4 \cdot u$ and (bottom) approximations of $\mathbb{E}(\tau^4 \cdot u)$ using the hierarchical method. From left-to-right, τ is initialized at $\tau = 10, 30, 50, 70, 90$.



(b) As in (a), using the non-hierarchical method. From left-to-right, τ is fixed at $\tau = 10, 30, 50, 70, 90$.

Figure 4.16: (EIT model) Representative samples and sample means of the level set function. The rescaling τ^4 means that the above quantities have the same approximate amplitude. True inverse length scale is $\tau = 15$.

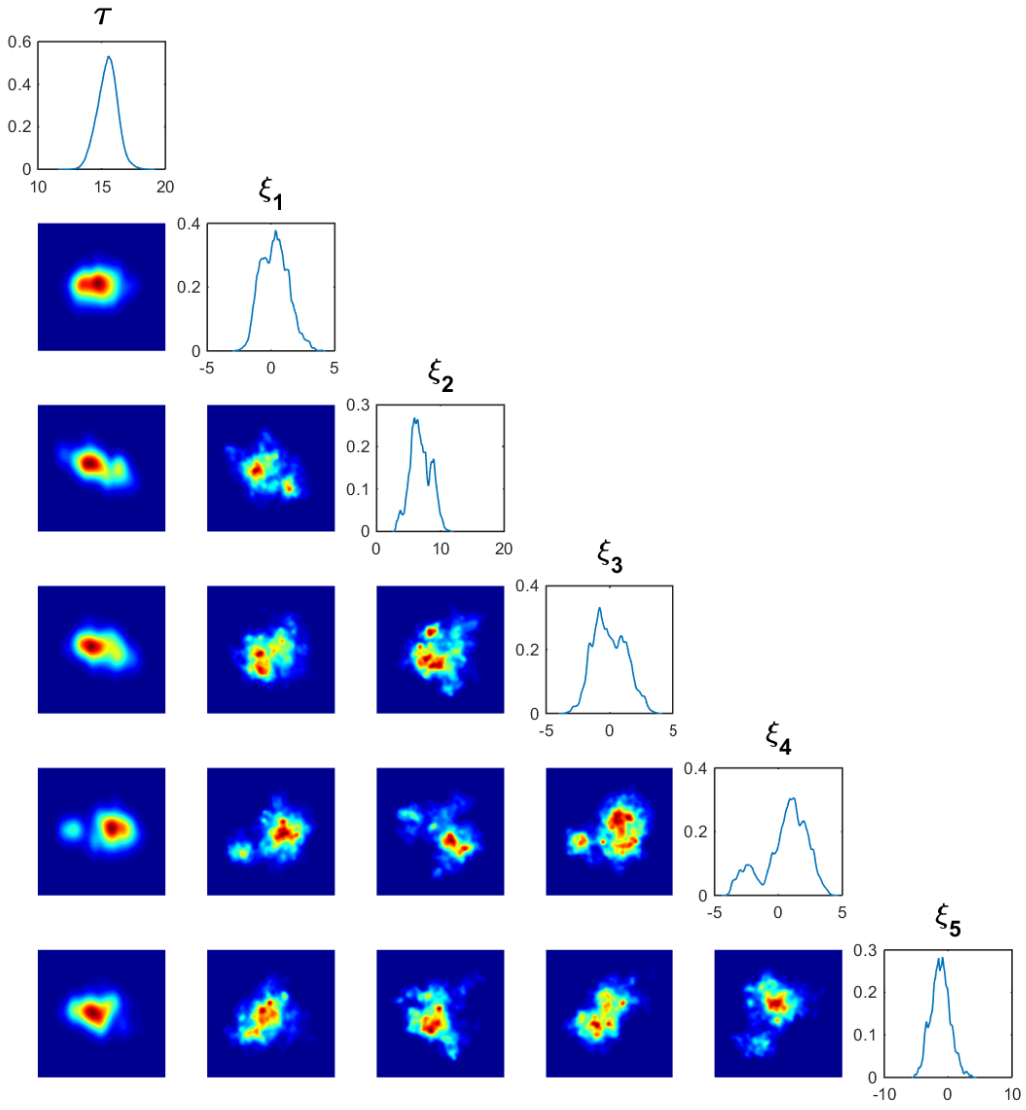


Figure 4.17: (EIT model) (diagonal) Empirical densities of τ and the first five KL modes of u. (off-diagonal) Empirical joint densities.

dating the level set function and the length-scale. The Metropolis method we use for the field update does not use derivatives of the log-likelihood, and could be improved by doing so, using the infinite dimensional variants on MALA and HMC (which use first derivative information, see the citations in [34]) or the manifold MALA and HMC methods, which use higher order derivatives [56]. Another interesting direction for future work is the design of methods with more informed proposals which exploit correlations in the level set function and its length-scale. And finally it would be interesting to consider pseudo-marginal methods to sample the hierarchical parameter alone, as in [49].

Finally we mention that the use of a single length-scale within an isotropic prior is a simple example of more sophisticated hierarchical approaches which attempt to learn non-stationary and non-isotropic [24, 25] features of the level set function from the data. This provides an interesting opportunity for future work and for ideas from machine learning to play a role in the solution of inverse problems for interfaces.

4.6 Appendix

4.6.1 Proof of Theorems

Proof of Theorem 4.2.1. (i) Note that it suffices to show that $\mu_0^\tau \sim \mu_0^0$ for all $\tau > 0$. (Here $\sim$ denotes "equivalent as measures"). It is known that the eigenvalues of $-\Delta$ on $\mathbb{T}^d$ grow like $j^{2/d}$, and hence the eigenvalues $\lambda_j(\tau)$ of $\mathcal{C}_{\alpha,\tau}$ decay like

$$\lambda_j(\tau) \asymp (\tau^2 + j^{2/d})^{-\alpha}, \quad j \geq 1.$$

Using Proposition 4.6.3, we see that $\mu_0^\tau \sim \mu_0^0$ if

$$\sum_{j=1}^{\infty} \left(\frac{\lambda_j(\tau)}{\lambda_j(0)} - 1 \right)^2 < \infty.$$

Now we have

$$\left|\frac{\lambda_j(\tau)}{\lambda_j(0)} - 1\right| \asymp \left|\left(1 + \frac{\tau^2}{j^{2/d}}\right)^{-\alpha} - 1\right|$$
$$\leq \left|\exp\left(\frac{\alpha\tau^2}{j^{2/d}}\right) - 1\right|$$
$$\leq C\frac{\alpha\tau^2}{j^{2/d}}.$$

Here we have used that $(1+x)^{-\alpha} - 1 \leq \exp(\alpha x) - 1$ for all $x \geq 0$ to move from the first to the second line, and that $\exp(x) - 1 \leq Cx$ for all $x \in [0, x_0]$ to move from the second to third line. Now note that when $d \leq 3$, $j^{-4/d}$ is summable, and so it follows that $\mu_0^\tau \sim \mu_0^0$.

(ii) The case $\tau = 0$ is Theorem 2.18 in [39]; the general result follows from the equivalence above.

(iii) Let $v \sim N(m, \mathcal{D}_{\nu,\ell})$ where $\mathcal{D}_{\nu,\ell}$ is as given by (4.2.2). These samples satisfy $\mathbb{E}\|v - m\|^2 = \sigma^2$ independent of ℓ. Now note that we have

$$\begin{aligned}\mathcal{D}_{\nu,\ell} &= \beta\ell^d(I - \ell^2\Delta)^{-\nu-d/2} \\ &= \beta\ell^d\ell^{-2\nu-d}(\ell^{-2}I - \Delta)^{-\nu-d/2} \\ &= \beta\tau^{2\alpha-d}(\tau^2 I - \Delta)^{-\alpha} \\ &= \beta\tau^{2\alpha-d}\mathcal{C}_{\alpha,\tau}.\end{aligned}$$

and so letting $u \sim N(m, \mathcal{C}_{\alpha,\tau})$, we see that

$$\mathbb{E}\|u - m\|^2 = \mathrm{tr}(\mathcal{C}_{\alpha,\tau}) = \frac{\sigma^2}{\beta}\tau^{d-2\alpha}$$

□

Proof of Theorem 4.2.4. Proposition 4.6.1 which follows shows that μ_0 and Φ satisfy Assumptions 2.1 in [72], with $U = X \times \mathbb{R}^+$. Theorem 2.2 in [72] then tells us that the posterior exists and is Lipschitz with respect to the data. □

Proposition 4.6.1. *Let μ_0 be given by (4.2.3) and $\Phi : X \times \mathbb{R}^+ \to \mathbb{R}$ be given by (4.2.9). Let Assumptions 4.2.3 hold. Then*

(i) for every $r > 0$ there is a $K = K(r)$ such that, for all $(u, \tau) \in X \times \mathbb{R}^+$ and all

$y \in Y$ *with* $|y|_\Gamma < r$,

$$0 \leq \Phi(u, \tau; y) \leq K;$$

(ii) for any fixed $y \in Y$, $\Phi(\cdot, \cdot; y) : X \times \mathbb{R}^+ \to \mathbb{R}$ *is continuous* μ_0*-almost surely on the complete probability space* $(X \times \mathbb{R}^+, \mathcal{X} \otimes \mathcal{R}, \mu_0)$*;*

(iii) for $y_1, y_2 \in Y$ *with* $\max\{|y_1|_\Gamma, |y_2|_\Gamma\} < r$*, there exists a* $C = C(r)$ *such that for all* $(u, \tau) \in X \times \mathbb{R}^+$,

$$|\Phi(u, \tau; y_1) - \Phi(u, \tau; y_2)| \leq C|y_1 - y_2|_\Gamma.$$

Proof. (i) Recall the level set map F defined by (4.2.6) defined via the finite constant values κ_i taken on each subset D_i of $\overline{D}$. We may bound F uniformly:

$$|F(u, \tau)| \leq \max\{|\kappa_1|, \dots |\kappa_n|\} =: F_{\max}$$

for all $(u, \tau) \in X \times \mathbb{R}^+$. Combining this with Assumption 4.2.3(ii) it follows that $\mathcal{G}$ is uniformly bounded on $X \times \mathbb{R}^+$. The result then follows from the continuity of $y \mapsto \frac{1}{2}|y - \mathcal{G}(u, \tau)|^2_\Gamma$.

(ii) This holds, with very little additional work, in the more general case where the prior mean m_0 has continuous spatial dependence; we prove it in this case which may be of independent interest. Note that this means that the rescaled levels $c_i(\tau)$ will also gain a spatial dependence and so we write $c_i(\tau) = c_i(\tau, x)$.

Let $(u, \tau) \in X \times \mathbb{R}^+$ and let $D_i(u, \tau)$ be as defined by (4.2.5), and define $D_i^0(u, \tau)$ by

$$\begin{aligned} D_i^0(u, \tau) &= \overline{D}_i(u, \tau) \cap \overline{D}_{i+1}(u, \tau) \\ &= \{x \in D \,|\, u(x) = c_i(\tau, x)\}, \quad i = 1, \dots, n-1. \end{aligned}$$

We first show that $\mathcal{G}$ is continuous at (u, τ) whenever $|D_i^0(u, \tau)| = 0$ for $i = 1, \dots, n-1$.

Choose an approximating sequence $(u_\varepsilon, \tau_\varepsilon)_{\varepsilon > 0}$ of (u, τ) such that $\|u_\varepsilon - u\|_\infty + |\tau_\varepsilon - \tau| < \varepsilon$ for all $\varepsilon > 0$. We will first show that $\|F(u_\varepsilon, \tau_\varepsilon) - F(u, \tau)\|_{L^p(D)} \to 0$

for any $p \in [1, \infty)$. As in [72] Proposition 2.4, we can write

$$\begin{aligned} F(u_\varepsilon, \tau_\varepsilon) - F(u, \tau) &= \sum_{i=1}^{n} \sum_{j=1}^{n} (\kappa_i - \kappa_j) \mathbb{1}_{D_i(u_\varepsilon, \tau_\varepsilon) \cap D_j(u, \tau)} \\ &= \sum_{\substack{i,j=1 \\ i \neq j}}^{n} (\kappa_i - \kappa_j) \mathbb{1}_{D_i(u_\varepsilon, \tau_\varepsilon) \cap D_j(u, \tau)}. \end{aligned}$$

From the definition of $(u_\varepsilon, \tau_\varepsilon)$,

$$u(x) - \varepsilon < u_\varepsilon(x) < u(x) + \varepsilon, \quad \tau - \varepsilon < \tau_\varepsilon < \tau + \varepsilon$$

for all $x \in D$ and $\varepsilon > 0$. Hence for $|i - j| > 1$ and ε sufficiently small, $D_i(u_\varepsilon, \tau_\varepsilon) \cap D_i(u, \tau) = \varnothing$. We hence look at the cases $|i - j| = 1$. We have that

$$\begin{aligned} D_i(u_\varepsilon, \tau_\varepsilon) = \big\{ x \in D \;\big|\; & m_0(x) + \tau_\varepsilon^{d/2-\alpha} \big(c_{i-1} - m_0(x) \big) \\ & \leq u_\varepsilon(x) \leq m_0(x) + \tau_\varepsilon^{d/2-\alpha} \big(c_i - m_0(x) \big) \big\} \\ = \big\{ x \in D \;\big|\; & c_{i-1} \leq m_0(x) + \tau_\varepsilon^{\alpha-d/2} \big(u_\varepsilon(x) - m_0(x) \big) < c_i \big\}. \end{aligned}$$

For small ε we may approximate

$$(\tau \pm \varepsilon)^{\alpha-d/2} = \tau^{\alpha-d/2} \pm \varepsilon(\alpha - d/2)\tau^{\alpha-d/2-1} + \mathcal{O}(\varepsilon^2)$$

and so, since $\alpha - d/2 > 0$,

$$\begin{aligned} & \tau_\varepsilon^{\alpha-d/2} \big(u_\varepsilon(x) - m_0(x) \big) \\ & \quad < (\tau + \varepsilon)^{\alpha-d/2} \big(u(x) - m_0(x) + \varepsilon \big) \\ & \quad = \tau^{\alpha-d/2} \big(u(x) - m_0(x) \big) \\ & \qquad + \varepsilon \big(\tau^{\alpha-d/2} + \big(u(x) - m_0(x) \big)(\alpha - d/2)\tau^{\alpha-d/2-1} \big) + \mathcal{O}(\varepsilon^2) \\ & \quad \leq \tau^{\alpha-d/2} \big(u(x) - m_0(x) \big) + \varepsilon\gamma + \mathcal{O}(\varepsilon^2). \end{aligned}$$

The above $\gamma \in \mathbb{R}$ exists since u and m_0 are bounded. Similarly we may obtain the lower bound

$$\tau_\varepsilon^{\alpha-d/2} \big(u_\varepsilon(x) - m_0(x) \big) > \tau^{\alpha-d/2} \big(u(x) - m_0(x) \big) - \varepsilon\gamma + \mathcal{O}(\varepsilon^2).$$

We therefore have that, for small ε,

$$\begin{aligned} D_i(u_\varepsilon,\tau_\varepsilon) \subseteq \big\{ x \in D \;\big|\; & c_{i-1} - \gamma\varepsilon + \mathcal{O}(\varepsilon^2) \\ & < m_0(x) + \tau^{\alpha-d/2}\big(u(x) - m_0(x)\big) < c_i + \varepsilon\gamma + \mathcal{O}(\varepsilon^2) \big\} \end{aligned}$$

from which it follows that

$$\begin{aligned} D_i(u_\varepsilon,\tau_\varepsilon) \cap D_{i+1}(u,\tau) &\subseteq \big\{ x \in D \;\big|\; c_i \leq m_0(x) + \tau^{\alpha-d/2}\big(u(x) - m_0(x)\big) \\ &\qquad < c_i + \varepsilon\gamma + \mathcal{O}(\varepsilon^2) \big\} \\ &\to \big\{ x \in D \;\big|\; m_0(x) + \tau^{\alpha-d/2}\big(u(x) - m_0(x)\big) = c_i \big\} \\ &= \big\{ x \in D \;\big|\; u(x) = m_0(x) + \tau^{d/2-\alpha}\big(c_i - m_0(x)\big) \big\} \\ &= D_i^0(u,\tau) \end{aligned}$$

and also

$$\begin{aligned} D_i(u_\varepsilon,\tau_\varepsilon) \cap D_{i-1}(u,\tau) &\subseteq \big\{ x \in D \;\big|\; c_{i-1} - \varepsilon\gamma + \mathcal{O}(\varepsilon^2) \\ &\qquad < m_0(x) + \tau^{\alpha-d/2}\big(u(x) - m_0(x)\big) < c_{i-1} \big\} \\ &\to \varnothing. \end{aligned}$$

Assume that each $|D_i^0(u,\tau)| = 0$, then it follows that $|D_i(u_\varepsilon,\tau_\varepsilon) \cap D_j(u,\tau)| \to 0$ whenever $i \neq j$. Therefore we have that

$$\begin{aligned} \|F(u_\varepsilon,\tau_\varepsilon) - F(u,\tau)\|_{L^p(D)}^p &= \sum_{\substack{i,j=1 \\ i\neq j}}^{n} \int_{D_i(u_\varepsilon,\tau_\varepsilon)\cap D_j(u,\tau)} |\kappa_i - \kappa_j|^p \,\mathrm{d}x \\ &\leq (2F_{\max})^p \sum_{\substack{i,j=1 \\ i\neq j}}^{n} |D_i(u_\varepsilon,\tau_\varepsilon) \cap D_j(u,\tau)| \\ &\to 0. \end{aligned}$$

Thus F is continuous at (u,τ). By Assumption 4.2.3(i) it follows that $\mathcal{G}$ is continuous at (u,τ).

We now claim that $|D_i^0(u,\tau)| = 0$ μ_0-almost surely for each i. By Tonelli's

theorem, we have that

$$\begin{aligned}
\mathbb{E}|D_i^0(u,\tau)| &= \int_{X\times\mathbb{R}^+} |D_i^0(u,\tau)|\,\mu_0(\mathrm{d}u,\mathrm{d}\tau) \\
&= \int_{X\times\mathbb{R}^+} \left(\int_{\mathbb{R}} \mathbb{1}_{D_i^0(u,\tau)}(x)\,\mathrm{d}x\right)\mu_0(\mathrm{d}u,\mathrm{d}\tau) \\
&= \int_{\mathbb{R}^d} \left(\int_{X\times\mathbb{R}^+} \mathbb{1}_{D_i^0(u,\tau)}(x)\,\mu_0(\mathrm{d}u,\mathrm{d}\tau)\right)\mathrm{d}x \\
&= \int_{\mathbb{R}^d} \left(\int_0^\infty \left(\int_X \mathbb{1}_{D_i^0(u,\tau)}(x)\,\mu_0^\tau(\mathrm{d}u)\right)\,\pi_0(\mathrm{d}\tau)\right)\mathrm{d}x \\
&= \int_{\mathbb{R}^d} \left(\int_0^\infty \mu_0^\tau(\{u\in X \mid u(x)=c_i(\tau,x)\})\,\pi_0(\mathrm{d}\tau)\right)\mathrm{d}x.
\end{aligned}$$

For each $\tau \geq 0$ and $x \in D$, $u(x)$ is a real-valued Gaussian random variable under μ_0^τ. It follows that $\mu_0^\tau(\{u \in X \mid u(x) = c_i(\tau,x)\}) = 0$, and so $\mathbb{E}|D_i^0(u,\tau)| = 0$. Since $|D_i^0(u,\tau)| \geq 0$ we have that $|D_i^0(u,\tau)| = 0$ μ_0-almost surely. The result now follows.

(iii) For fixed $(u,\tau) \in X \times \mathbb{R}^+$, the map $y \mapsto \frac{1}{2}|y - \mathcal{G}(u,\tau)|_\Gamma^2$ is smooth and hence locally Lipschitz.

□

Proof of Theorem 4.3.3. Recall that the eigenvalues of $\mathcal{C}_{\alpha,\tau}$ satisfy $\lambda_j(\tau) \asymp (\tau^2 + j^{2/d})^{-\alpha}$. Then we have that

$$\left(\frac{\lambda_j(0)}{\lambda_j(\tau)} - 1\right) \asymp (1+\tau^2 j^{-2/d})^\alpha - 1 = \mathcal{O}(j^{-2/d}).$$

It follows that

$$\sum_{j=1}^\infty \left(\frac{\lambda_j(0)}{\lambda_j(\tau)} - 1\right)^p < \infty \quad \text{if and only if } d < 2p. \tag{4.6.1}$$

(i) We first prove the 'if' part of the statement. We have $u \sim N(m_0, \mathcal{C}_0)$, and so $\mathbb{E}\langle u - m_0, \varphi_j\rangle^2 = \lambda_j(0)$. Since the terms within the sum are non-negative, by Tonelli's theorem we can bring the expectation inside the sum to see that that

$$\mathbb{E}\sum_{j=1}^\infty \left(\frac{1}{\lambda_j(\tau)} - \frac{1}{\lambda_j(0)}\right)\langle u - m_0, \varphi_j\rangle^2 = \sum_{j=1}^\infty \left(\frac{\lambda_j(0)}{\lambda_j(\tau)} - 1\right)$$

which is finite if and only if $d < 2$, i.e. $d = 1$. It follows that the sum is finite

almost surely.

For the converse, suppose that $d \geq 2$ so that the series in (4.6.1) diverges when $p = 1$. Let $\{\xi_j\}_{j\geq 1}$ be a sequence of i.i.d. $N(0,1)$ random variables so that $\langle u - m_0, \varphi_j\rangle^2$ has the same distribution as $\lambda_j(0)\xi^2$. Define the sequence $\{Z_n\}_{n\geq 1}$ by

$$\begin{aligned} Z_n &= \sum_{j=1}^{n} \left(\frac{\lambda_j(0)}{\lambda_j(\tau)} - 1\right) \xi_j^2 \\ &= \sum_{j=1}^{n} \left(\frac{\lambda_j(0)}{\lambda_j(\tau)} - 1\right) + \sum_{j=1}^{n} \left(\frac{\lambda_j(0)}{\lambda_j(\tau)} - 1\right) (\xi_j^2 - 1) \\ &=: X_n + Y_n. \end{aligned}$$

Then the result follows if Z_n diverges with positive probability. By assumption we have that X_n diverges. In order to show that Z_n diverges with positive probability it hence suffices to show that Y_n converges with positive probability. Define the sequence of random variables $\{W_j\}_{j\geq 1}$ by

$$W_j = \left(\frac{\lambda_j(0)}{\lambda_j(\tau)} - 1\right) (\xi_j^2 - 1).$$

It can be checked that

$$\mathbb{E}(W_j) = 0, \quad \mathrm{Var}(W_j) = 2\left(\frac{\lambda_j(0)}{\lambda_j(\tau)} - 1\right)^2.$$

The series of variances converges if and only if $d \leq 3$, using (4.6.1) with $p = 2$. We use Kolmogorov's two series theorem, Theorem 3.11 in [137], to conclude that $Y_n = \sum_{j=1}^{n} W_j$ converges almost surely and the result follows.

(ii) Now we have

$$\begin{aligned} \log\left(\frac{\lambda_j(\tau)}{\lambda_j(0)}\right) &= -\log\left(1 - \left(1 - \frac{\lambda_j(0)}{\lambda_j(\tau)}\right)\right) \\ &= \left(1 - \frac{\lambda_j(0)}{\lambda_j(\tau)}\right) + \frac{1}{2}\left(1 - \frac{\lambda_j(0)}{\lambda_j(\tau)}\right)^2 + \text{h.o.t.} \end{aligned}$$

Let $\{\xi_j\}_{j\geq 1}$ be a sequence of i.i.d. $N(0,1)$ random variables, so that again we have that $\langle u - m_0, \varphi_j\rangle^2$ has the same distribution as $\lambda_j(0)\xi^2$. Then it is

sufficient to show that the series

$$I = \sum_{j=1}^{\infty} \left[\left(\frac{\lambda_j(0)}{\lambda_j(\tau)} - 1 \right) \xi_j^2 + \log \left(\frac{\lambda_j(\tau)}{\lambda_j(0)} \right) \right]$$

is finite almost surely. We use the above approximation for the logarithm to write

$$I = \sum_{j=1}^{\infty} \left(\frac{\lambda_j(0)}{\lambda_j(\tau)} - 1 \right) (\xi_j^2 - 1) + \sum_{j=1}^{\infty} \left[\frac{1}{2} \left(1 - \frac{\lambda_j(0)}{\lambda_j(\tau)} \right)^2 + \text{h.o.t.} \right].$$

The second sum converges if and only if $d < 4$, i.e. $d \leq 3$. The almost sure convergence of the first term is shown in the proof of part (i).

□

Proposition 4.6.2. *Let $D \subseteq \mathbb{R}^d$. Define the construction map $F : X \times \mathbb{R}^+ \to \mathbb{R}^D$ by (4.2.6). Given $x_0 \in D$ define $\mathcal{G} : X \times \mathbb{R}^+ \to \mathbb{R}$ by $\mathcal{G}(u, \tau) = F(u, \tau)|_{x_0}$. Then $\mathcal{G}$ is continuous at any $(u, \tau) \in X \times \mathbb{R}^+$ with $u(x_0) \neq c_i(\tau)$ for each $i = 0, \ldots, n$. In particular, $\mathcal{G}$ is continuous μ_0-almost surely when μ_0 is given by (4.2.3). Additionally, $\mathcal{G}$ is uniformly bounded.*

Proof. The uniform boundedness is clear. For the continuity, let $(u, \tau) \in X \times \mathbb{R}^+$ with $u(x_0) \neq c_i(\tau)$ for each $i = 0, \ldots, n$. Then there exists a unique j such that

$$c_{j-1}(\tau) < u(x_0) < c_j(\tau). \tag{4.6.2}$$

Given $\delta > 0$, let $(u_\delta, \tau_\delta) \in X \times \mathbb{R}^+$ be any pair such that

$$\|u_\delta - u\|_\infty + |\tau_\delta - \tau| < \delta.$$

Then it is sufficient to show that for all δ sufficiently small, $x_0 \in D_j(u_\delta, \tau_\delta)$, i.e. that

$$c_{j-1}(\tau_\delta) \leq u_\delta(x_0) < c_j(\tau_\delta).$$

From this it follows that $G(u_\delta, \tau_\delta) = G(u, \tau)$.

Since the inequalities in (4.6.2) are strict, we can find $\alpha > 0$ such that

$$c_{j-1} + \alpha < u(x_0) < c_j(\tau) - \alpha. \tag{4.6.3}$$

Now c_j is continuous at $\tau > 0$, and so there exists a $\gamma > 0$ such that for any $\lambda > 0$

with $|\lambda - \tau| < \gamma$ we have

$$c_j(\lambda) - \alpha/2 < c_j(\tau) < c_j(\lambda) + \alpha/2. \tag{4.6.4}$$

We have that $\|u_\delta - u\|_\infty < \delta$, and so in particular,

$$u(x_0) - \delta < u_\delta(x_0) < u(x_0) + \delta. \tag{4.6.5}$$

We can combine (4.6.3)-(4.6.5) to see that, for $\delta < \gamma$,

$$c_{j-1}(\tau_\delta) - \delta + \alpha/2 < u_\delta(x_0) < c_j(\tau_\delta) + \delta - \alpha/2$$

and so in particular, for $\delta < \min\{\gamma, \alpha/2\}$,

$$c_{j-1}(\tau_\delta) < u_\delta(x_0) < c_j(\tau_\delta).$$

□

4.6.2 Radon-Nikodym Derivatives in Hilbert Spaces

The following proposition gives an explicit formula for the density of one Gaussian with respect to another and is used in defining the acceptance probability for the length-scale updates in our algorithm. Although we only use the proposition in the case where H is a function space and the mean m is constant, we provide a proof in the more general case where m is an arbitrary element of a separable Hilbert space H as this setting may be of independent interest.

Proposition 4.6.3. *Let $(H, \langle\cdot,\cdot\rangle, \|\cdot\|)$ be a separable Hilbert space, and let A, B be positive trace-class operators on H. Assume that A and B share a common complete set of orthonormal eigenvectors $(\varphi_j)_{j\geq 1}$, with the eigenvalues $(\lambda_j)_{j\geq 1}$, $(\gamma_j)_{j\geq 1}$ defined by*

$$A\varphi_j = \lambda_j \varphi_j, \quad B\varphi_j = \gamma_j \varphi_j$$

for all $j \geq 1$. Assume further that the eigenvalues satisfy

$$\sum_{j=1}^{\infty} \left(\frac{\lambda_j}{\gamma_j} - 1 \right)^2 < \infty.$$

Let $m \in H$ and define the measures $\mu = N(m, A)$ and $\nu = N(m, B)$. Then μ and ν

are equivalent, and their Radon-Nikodym derivative is given by

$$\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(u) = \prod_{j=1}^{\infty} \frac{\gamma_j}{\lambda_j} \cdot \exp\left(\frac{1}{2}\sum_{j=1}^{\infty}\left(\frac{1}{\gamma_j} - \frac{1}{\lambda_j}\right)\langle u - m, \varphi_j\rangle^2\right).$$

Proof. The assumption on summability of the eigenvalues means that the Feldman-Hájek theorem applies, and so we know that μ and ν are equivalent. We show that the Radon-Nikodym derivative is as given above.

Define the product measures $\hat{\mu}, \hat{\nu}$ on $\mathbb{R}^\infty$ by

$$\hat{\mu} = \prod_{j=1}^{\infty} \hat{\mu}_j, \quad \hat{\nu} = \prod_{j=1}^{\infty} \hat{\nu}_j$$

where $\hat{\mu}_j = N(0, \lambda_j)$, $\hat{\nu}_j = N(0, \gamma_j)$. As a consequence of a result of Kakutani, see [35] Proposition 1.3.5, we have that $\hat{\mu} \sim \hat{\nu}$ with

$$\begin{aligned}\frac{\mathrm{d}\hat{\mu}}{\mathrm{d}\hat{\nu}}(x) &= \prod_{j=1}^{\infty} \frac{\mathrm{d}\hat{\mu}_j}{\mathrm{d}\hat{\nu}_j}(x_j) \\ &= \prod_{j=1}^{\infty} \frac{\gamma_j}{\lambda_j} \cdot \exp\left(\frac{1}{2}\sum_{j=1}^{\infty}\left(\frac{1}{\gamma_j} - \frac{1}{\lambda_j}\right)x_j^2\right).\end{aligned}$$

We associate H with $\mathbb{R}^\infty$ via the map $G : H \to \mathbb{R}^\infty$, given by

$$G_j u = \langle u, \varphi_j\rangle, \quad j \geq 1.$$

Note that the image of G is $\ell^2 \subseteq \mathbb{R}^\infty$, and $G : H \to \ell^2$ is an isomorphism. Since A and B are trace-class, samples from $\hat{\mu}$ and $\hat{\nu}$ almost surely take values in ℓ^2. G^{-1} is hence almost surely defined on samples from $\hat{\mu}$ and $\hat{\nu}$. Define the translation map $T_m : H \to H$ by $T_m u = u + m$. Then by the Karhunen-Loève theorem, the measures μ and ν can be expressed as the push-forwards

$$\mu = T_m^{\#}(G^{-1})^{\#}\hat{\mu}, \quad \nu = T_m^{\#}(G^{-1})^{\#}\hat{\nu}.$$

Now let $f : H \to \mathbb{R}$ be bounded measurable, then we have

$$\begin{aligned}
\int_H f(u)\,\mu(\mathrm{d}u) &= \int_H f(u)\left[T_m^{\#}(G^{-1})^{\#}\hat{\mu}\right](\mathrm{d}u) \\
&= \int_{\mathbb{R}^\infty} f(G^{-1}x + m)\,\hat{\mu}(\mathrm{d}x) \\
&= \int_{\mathbb{R}^\infty} f(G^{-1}x + m)\frac{\mathrm{d}\hat{\mu}}{\mathrm{d}\hat{\nu}}(x)\,\hat{\nu}(\mathrm{d}x) \\
&= \int_H f(u)\frac{\mathrm{d}\hat{\mu}}{\mathrm{d}\hat{\nu}}(G(u-m))\left[T_m^{\#}(G^{-1})^{\#}\hat{\nu}\right](\mathrm{d}u) \\
&= \int_H f(u)\frac{\mathrm{d}\hat{\mu}}{\mathrm{d}\hat{\nu}}(G(u-m))\,\nu(\mathrm{d}u).
\end{aligned}$$

From this is follows that we have

$$\begin{aligned}
\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(u) &= \frac{\mathrm{d}\hat{\mu}}{\mathrm{d}\hat{\nu}}(G(u-m)) \\
&= \prod_{j=1}^{\infty}\frac{\gamma_j}{\lambda_j}\cdot\exp\left(\frac{1}{2}\sum_{j=1}^{\infty}\left(\frac{1}{\gamma_j}-\frac{1}{\lambda_j}\right)\langle u-m,\varphi_j\rangle^2\right).
\end{aligned}$$

□

Remark 4.6.4. *The proposition above, in the case $m = 0$, is given as Theorem 1.3.7 in [35] except that, there, the factor before the exponential is omitted. This is because it does not depend on u, and all measures involved are probability measures and hence normalized. We retain the factor as we are interested in the precise value of the derivative for the MCMC algorithm; in particular its dependence on the length-scale.*

Appendix A

Background and Preliminaries

A.1 Measure Theory

In this section we overview the notations and results in measure theory that are relevant to this thesis. We then define general Gaussian measures, and their relevant properties.

A.1.1 General Measure Theory

The definitions and results in this subsection can be found in any standard book on measure theory or measure theoretic probability, for example [17].

Definition A.1.1 (σ-algebra)**.** *Let X be a set and $\mathcal{X}$ a collection of subsets of X. $\mathcal{X}$ is called a σ-algebra on X if*

(i) $\varnothing \in \mathcal{X}$*;*

(ii) for all $A \in \mathcal{X}$*,* $X \setminus A \in \mathcal{X}$*; and*

(iii) for all countable subcollections $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{X}$*,* $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{X}$*.*

Definition A.1.2 (Borel σ-algebra)**.** *If X is a topological space, then the smallest σ-algebra containing all open sets is called the Borel σ-algebra. It is denoted $\mathcal{B}(X)$.*

Remark A.1.3. *If no σ-algebra is explicitly specified on a space, it is assumed to be equipped with its Borel σ-algebra.*

Definition A.1.4 (Measurable space)**.** *An ordered pair $(X, \mathcal{X})$ of a set X and a σ-algebra $\mathcal{X}$ on X is called a measurable space.*

Definition A.1.5 (Measure)**.** *Let $(X, \mathcal{X})$ be a measurable space. A function $\mu : \mathcal{X} \to [0, \infty]$ is called a measure on $(X, \mathcal{X})$ if*

(i) $\mu(\varnothing) = 0$; and

(ii) for all countable subcollections $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{X}$ of pairwise disjoint sets,

$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{k=1}^{\infty} \mu(A_n).$$

Definition A.1.6 (Probability measure)**.** *A measure μ on a measurable space $(X, \mathcal{X})$ is called a probability measure if $\mu(X) = 1$.*

Definition A.1.7 (Measurable function, random variable)**.** *Let $(X, \mathcal{X}, \mu)$ be a measure space and $(Y, \mathcal{Y})$ be a measurable space. A function $f : X \to Y$ is called measurable if $f^{-1}(A) \in \mathcal{X}$ for all $A \in \mathcal{Y}$. If μ is a probability measure then f is called a (Y-valued) random variable.*

Definition A.1.8. *Let $(X, \mathcal{X}, \mu)$ be a measure space, and let $A \in \mathcal{X}$. The indicator function of A is the measurable function $\mathbb{1}_A : X \to \mathbb{R}$ defined by*

$$\mathbb{1}_A(u) = \begin{cases} 1 & u \in A \\ 0 & u \notin A \end{cases}.$$

Definition A.1.9 (Integration, integrability)**.** *Let $(X, \mathcal{X}, \mu)$ be a measure space.*

(i) Let $A \in \mathcal{X}$. The integral of $\mathbb{1}_A$ with respect to μ is defined by

$$\int_X \mathbb{1}_A(u)\, \mu(\mathrm{d}u) := \mu(A).$$

This definition extends linearly to linear combinations of indicator functions, referred to as simple functions.

(ii) Let $f : X \to [0, \infty)$ be measurable. Then there exists an increasing sequence of simple functions $\{f_n\}_{n\in\mathbb{N}}$ with $f_n \to f$ pointwise. The integral of f with

respect to μ is defined by

$$\int_X f(u)\,\mu(\mathrm{d}u) := \lim_{n\to\infty} \int_X f_n(u)\,\mu(\mathrm{d}u).$$

This limit can be shown to exist in $[0,\infty]$ and to be independent of the choice of approximating sequence $\{f_n\}_{n\in\mathbb{N}}$.

(iii) Let $f : X \to \mathbb{R}$ be measurable. Then there exist measurable $f_+, f_- : X \to [0,\infty)$ with $f = f_+ - f_-$. The integral of f with respect to μ is defined by

$$\int_X f(u)\,\mu(\mathrm{d}u) := \int_X f_+(u)\,\mu(\mathrm{d}u) - \int_X f_-(u)\,\mu(\mathrm{d}u)$$

whenever at least one of the integrals on the right hand side is finite.

A measurable function for which the integral exists and is finite is called integrable.

In the above definition (*ii*) we asserted that the limit existed and was independent of the choice of increasing sequence; this is a special case of what is called the monotone convergence theorem. A related result, used in this thesis, is the dominated convergence theorem:

Theorem A.1.10 (Dominated convergence theorem)**.** *Let $(X,\mathcal{X},\mu)$ be a measure space. Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence of real valued measurable functions on $(X,\mathcal{X},\mu)$ such that $\{f_n\}_{n\in\mathbb{N}}$ converges pointwise to a function f. Suppose that there exists a real valued integrable function g on $(X,\mathcal{X},\mu)$ such that $|f_n(u)| \le Cg(u)$ for all $n \in \mathbb{N}$ and all $x \in X$. Then f is integrable, and*

$$\int_X f_n(u)\,\mu(\mathrm{d}u) \to \int_X f(u)\,\mu(\mathrm{d}u).$$

Absolute continuity of measures plays a key role throughout the thesis. It is defined as follows.

Definition A.1.11. *Let $(X,\mathcal{X})$ be a measurable space and let μ,ν be two measures on $(X,\mathcal{X})$. ν is said to be absolutely continuous with respect to μ, denoted $\nu \ll \mu$, if $\mu(A) = 0$ implies that $\nu(A) = 0$. μ and ν are said to be equivalent, denoted $\mu \sim \nu$, if $\nu \ll \mu$ and $\mu \ll \nu$.*

Theorem A.1.12 (Radon-Nikodym)**.** *Let $(X,\mathcal{X})$ be measurable space and let μ,ν be two probability measures on X such that $\nu \ll \mu$. Then there exists an integrable*

function $f : X \to [0, \infty)$ *such that for any* $A \in \mathcal{X}$,

$$\nu(A) = \int_A f(u)\, \mu(\mathrm{d}u). \tag{A.1.1}$$

The function f *is called the Radon-Nikodym derivative of* μ *with respect to* ν*, and denoted* $\frac{\mathrm{d}\mu}{\mathrm{d}\nu}$.

Remark A.1.13. *The relation (A.1.1) is often written as* $\mu(\mathrm{d}u) = f(u)\, \mu(\mathrm{d}u)$.

The notion of absolute continuity allows for the definition of certain metrics on the space of probability measures on X [39]. Note first that if μ, μ' are two probability measures, then there exists a probability measure ν such that $\mu, \mu' \ll \nu$, for example $\nu = \frac{1}{2}(\mu + \mu')$.

Definition A.1.14 (Total variation metric)**.** *Let* μ, μ' *be two probability measures absolutely continuous with respect to a probability measure* ν*. The total variation distance between* μ *and* μ' *is defined as*

$$d_{\mathrm{TV}}(\mu, \mu') = \frac{1}{2} \int_X \left| \frac{\mathrm{d}\mu}{\mathrm{d}\nu}(u) - \frac{\mathrm{d}\mu'}{\mathrm{d}\nu}(u) \right| \nu(\mathrm{d}u).$$

In particular, if μ' *is absolutely continuous with respect to* μ*, then*

$$d_{\mathrm{TV}}(\mu, \mu') = \frac{1}{2} \int_X \left| 1 - \frac{\mathrm{d}\mu'}{\mathrm{d}\mu}(u) \right| \mu(\mathrm{d}u).$$

Definition A.1.15 (Hellinger metric)**.** *Let* μ, μ' *be two probability measures absolutely continuous with respect to a probability measure* ν*. The Hellinger distance between* μ *and* μ' *is defined as*

$$d_{\mathrm{Hell}}(\mu, \mu') = \sqrt{\frac{1}{2} \int_X \left| \sqrt{\frac{\mathrm{d}\mu}{\mathrm{d}\nu}}(u) - \sqrt{\frac{\mathrm{d}\mu'}{\mathrm{d}\nu}}(u) \right|^2 \nu(\mathrm{d}u)}.$$

In particular, if μ' *is absolutely continuous with respect to* μ*, then*

$$d_{\mathrm{Hell}}(\mu, \mu') = \sqrt{\frac{1}{2} \int_X \left| 1 - \sqrt{\frac{\mathrm{d}\mu'}{\mathrm{d}\mu}}(u) \right|^2 \mu(\mathrm{d}u)}.$$

The following definition is central to the definition of Gaussian measures in the next subsection; given a measure on a space X and a map $T : X \to Y$, it allows for the construction of a particular measure on Y.

Definition A.1.16 (Pushforward measure)**.** *Let $(X, \mathcal{X})$ and $(Y, \mathcal{Y})$ be measurable spaces, and let $T : (X, \mathcal{X}) \to (Y, \mathcal{Y})$ be a measurable map. Let μ be a measure on $(X, \mathcal{X})$. Then the pushforward of μ by T is the measure $T^{\#}\mu$ on $(Y, \mathcal{Y})$ given by*

$$T^{\#}\mu(A) = \mu(T^{-1}(A))$$

for each $A \in \mathcal{Y}$.

Definition A.1.17 (Law of a random variable)**.** *Let $(X, \mathcal{X}, \mu)$ be a probability space and $(Y, \mathcal{Y})$ a measurable space. The law of a random variable $f : X \to Y$ is the probability measure on $(Y, \mathcal{Y})$ defined by the pushforward $f^{\#}(\mu)$.*

A.1.2 Gaussian Measure Theory

We provide a brief overview of the definitions and result associated with Gaussian measures used in this thesis. More detailed expositions can be found in, for example, [18, 61].

Definition A.1.18 (Gaussian measure on $\mathbb{R}$)**.** *A Borel measure μ on $\mathbb{R}$ is called a non-degenerate Gaussian measure if there exists $m \in \mathbb{R}$ and $\sigma^2 > 0$ such that*

$$\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-m)^2\right)$$

where λ denotes the Lebesgue measure on $\mathbb{R}$. It is called a degenerate Gaussian measure if there exists $m \in \mathbb{R}$ such that $\mu = \delta_m$.

Definition A.1.19 (Dual space)**.** *Let $(X, \|\cdot\|)$ be a Banach space. The dual space of X, denoted X^*, is the set of all continuous linear functionals $\ell : X \to \mathbb{R}$. It is itself a Banach space, equipped with the norm*

$$\|\ell\|_{X^*} := \sup_{\|u\|_X = 1} |\ell(u)|.$$

Definition A.1.20 (Gaussian measure on Banach space)**.** *Let X be a separable Banach space and μ be a Borel measure on X. Then μ is said to be a (non-degenerate) Gaussian measure if $\ell^{\#}\mu$ is a (non-degenerate) Gaussian measure on $\mathbb{R}$ for all $\ell \in X^*$. It is said to be centered if $\ell^{\#}\mu$ has mean zero for all $\ell \in X^*$.*

Definition A.1.21 (Covariance operator)**.** *Let μ be a centered Gaussian measure*

on a separable Banach space X. The operator $C_\mu : X^* \times X^* \to \mathbb{R}$ defined by

$$C_\mu(\ell, \ell') = \int_X \ell(u)\ell'(u)\,\mu(\mathrm{d}u)$$

is called the covariance operator of μ. If X is Hilbert, then after identification of X with its dual space, we have

$$C_\mu = \int_X (u \otimes u)\,\mu(\mathrm{d}u).$$

The covariance of a centered Gaussian measure completely determines it. We hence often write $\mu = N(0, C_\mu)$, since this defines a unique Gaussian measure.

Associated with each Gaussian measure μ on X is a Hilbert subspace $H_\mu \subset X$ called the Cameron-Martin space:

Definition A.1.22 (Cameron-Martin space)**.** *Let μ be a centered Gaussian measure on a separable Banach space X. The Cameron-Martin space $H_\mu \subset X$ of μ is defined as the closure of the space*

$$\mathring{H}_\mu = \{h \in X \,|\, \exists h^* \in X^* \text{ with } C_\mu(h^*, \ell) = \ell(h)\ \forall \ell \in X^*\}$$

under the inner product $\langle h_1, h_2 \rangle_\mu = C_\mu(h_1^*, h_2^*)$.

Remark A.1.23. *If $(X, \|\cdot\|)$ is Hilbert, then H_μ is the space $C_\mu^{1/2}X$ equipped with the inner product $\langle h_1, h_2 \rangle_\mu = \langle C_\mu^{-1/2}h_1, C_\mu^{-1/2}h_2 \rangle$.*

The space is important since it contains precisely the set of elements under which translation gives an equivalent Gaussian measure. In what follows we will write $\|\cdot\|_\mu$ for the norm arising from the inner product $\langle \cdot, \cdot \rangle_\mu$.

Theorem A.1.24 (Cameron-Martin)**.** *For $h \in X$ defined the map $T_h : X \to X$ by $T_h(u) = u + h$. Then the measure $T_h^\# \mu$ is absolutely continuous with respect to μ if and only if $h \in H_\mu$. In this case, we have*

$$\frac{\mathrm{d}(T_h^\# \mu)}{\mathrm{d}\mu}(u) = \exp\left(h^*(u) - \frac{1}{2}\|h\|_\mu^2\right).$$

Two important properties of the Cameron-Martin space H_μ are that $\mu(H_\mu) = 0$ and H_μ is dense in X. This means, for example, that $\|u\|_\mu = \infty$ μ-almost surely. This is significant when defining sampling algorithms or MAP estimation for measures associated with μ.

We conclude this overview of Gaussian measures by stating the following proposition, given as an exercise in [61].

Proposition A.1.25. *Let μ be a measure on a Banach space X. Suppose $\tilde{X} \subset X$ is a continuously embedded Banach space with $\mu(\tilde{X}) = 1$. Then the restriction of μ to $\tilde{X}$ is again a Gaussian measure.*

This proposition can be extremely useful, since Gaussian measures on Hilbert spaces are typically easier to analyze than those on Banach spaces. Thus if we have a Gaussian measure on a Hilbert space X and can show that there exists a Banach subspace $\tilde{X}$ continuously embedded in X with $\mu(\tilde{X}) = 1$, then properties of the restriction to $\tilde{X}$ can be more easily deduced.

A.2 Markov Chain Monte Carlo

In this section we outline Markov Chain Monte Carlo techniques for sampling probability measures. The book [109] provides a background on the theory of Markov chains.

Definition A.2.1 (Markov chain)**.** *Let $(X, \mathcal{X}, \mathbb{P})$ be a probability space. Let $\{u_n\}_{n\in\mathbb{N}}$ be a sequence of random variables on X. $\{u_n\}_{n\in\mathbb{N}}$ is said to be a Markov chain on X if, for each $n \in \mathbb{N}$,*

$$\mathbb{P}(u_{n+1} \in \cdot \,|\, u_1, \ldots, u_n) = \mathbb{P}(u_{n+1} \in \cdot \,|\, u_n).$$

Definition A.2.2 (Transition kernel)**.** *Let $\{u_n\}_{n\in\mathbb{N}}$ be a Markov chain on a probability space $(X, \mathcal{X}, \mathbb{P})$. A function $P : X \times \mathcal{X} \to [0, 1]$ is called the transition kernel for $\{u_n\}_{n\in\mathbb{N}}$ if*

(i) $u \mapsto P(u, A)$ is measurable for each $A \in \mathcal{X}$;

(ii) $P(u, \cdot)$ is a probability measure for each $u \in X$; and

(iii) $\mathbb{P}(u_{n+1} \in A \,|\, u_n = u) = P(u, A)$ for each $u \in X$, $A \in \mathcal{X}$ and $n \in \mathbb{N}$.

Remark A.2.3. *When such a transition kernel exists the Markov chain is said to be time-homogeneous, since the transition probabilities do not depend on the time n. In this appendix we consider only time-homogeneous Markov chains.*

Definition A.2.4 (Invariant distribution)**.** *Let $\{u_n\}_{n\in\mathbb{N}}$ be a Markov chain on a probability space $(X, \mathcal{X}, \mathbb{P})$ with transition kernel P. P (or $\{u_n\}_{n\in\mathbb{N}}$) is said to be*

invariant with respect to a probability measure μ on X if it satisfies

$$\mu(A) = \int_X \mu(\mathrm{d}v)P(v, A) \quad \textit{for all } A \in \mathcal{X}.$$

Definition A.2.5 (Reversible Markov chain)**.** *Let $\{u_n\}_{n\in\mathbb{N}}$ be a Markov chain on a probability space $(X, \mathcal{X}, \mathbb{P})$ with transition kernel P. P (or $\{u_n\}_{n\in\mathbb{N}}$) is said to be reversible with respect to a probability measure μ on X if it satisfies the detailed balance equation:*

$$\mu(\mathrm{d}u)P(u, \mathrm{d}v) = \mu(\mathrm{d}v)P(v, \mathrm{d}u).$$

Remark A.2.6. *Note that if P is reversible with respect to μ, then it is invariant with respect to μ.*

The idea of Markov Chain Monte Carlo (MCMC) is to, given a probability distribution μ on X, construct a Markov chain with transition kernel P such that μ is reversible with respect to P. The Markov chain corresponding to P can then be simulated numerically in order to produce a sequence of (correlated) samples from μ. If μ admits a Lebesgue density, then a standard construction of such a Markov chain is given by the Metropolis-Hastings algorithm:

Proposition A.2.7 (Metropolis-Hastings)**.** *Let μ be a measure on $\mathbb{R}^d$ with Lebesgue density π. Let $Q(u, \mathrm{d}v) = q(u, v)\mathrm{d}v$ be a probability distribution which proposes a new state given a state u. Construct a Markov chain $\{u_k\}_{n\in\mathbb{N}}$ on $\mathbb{R}^d$ as follows:*

1. *Set $n = 0$. Choose an initial state $u_0 \in \mathbb{R}^d$.*
2. *Propose $v_n \sim Q(u_n, \mathrm{d}v)$.*
3. *Set $u_{n+1} = v_n$ with probability*

$$\alpha(u_n, v_n) = \min\left\{1, \frac{\pi(v_n)q(v_n, u_n)}{\pi(u_n)q(u_n, v_n)}\right\}.$$

   independently of (u_n, v_n).
4. *Set $u_{n+1} = u_n$ otherwise.*
5. *Set $n \to n + 1$ and go to 2.*

Then $\{u_n\}_{n\in\mathbb{N}}$ is invariant with respect to μ.

A typical example for the proposal kernel $Q(u, \mathrm{d}v)$ is the Gaussian random walk proposal, given by $Q(u, \cdot) = N(u, \varepsilon^2 C)$ for some covariance matrix C and scalar jump parameter $\varepsilon > 0$.

Often it is advantageous to partition the state space into $\mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$. Then given a state $(u^1, u^2) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$, we can alternate updates that target the distributions of the the random variables $u^1|u^2$ and $u^2|u^1$. These are referred to as Metropolis-within-Gibbs methods [134].

Above we had to assume the existence of a Lebesgue density, though this will not be possible when the state space is infinite dimensional since there does not exist an infinite dimensional Lebesgue measure. Metropolis methods may however be generalized to these cases, as described in [135]. One particular MCMC method defined on function space is the preconditioned Crank-Nicolson method [34, 62]:

Proposition A.2.8 (pCN Method)**.** *Let X be a Hilbert space, and let μ be a measure on X given by*

$$\frac{\mathrm{d}\mu}{\mathrm{d}\mu_0}(u) \propto \exp(-\Phi(u))$$

where $\mu_0 = N(0, C)$ is Gaussian. Let $\beta \in (0, 1]$ be a scalar jump parameter. Construct a Markov chain $\{u_k\}_{n\in\mathbb{N}}$ on X as follows:

1. *Set $n = 0$. Choose an initial state $u_0 \in X$.*
2. *Propose $v_n = (1 - \beta^2)^{1/2} u_n + \beta\xi_n, \quad \xi_n \sim N(0, C)$.*
3. *Set $u_{n+1} = v_n$ with probability*

$$\alpha(u_n, v_n) = \min\left\{1, \exp(\Phi(u_n) - \Phi(v_n))\right\}.$$

   independently of (u_n, v_n).

4. *Set $u_{n+1} = u_n$ otherwise.*
5. *Set $n \to n + 1$ and go to 2.*

Then $\{u_n\}_{n\in\mathbb{N}}$ is invariant with respect to μ.

Bibliography

[1] S. I. Aanonsen, G. Nævdal, D. S. Oliver, A. C. Reynolds, B. Vallès, et al. The ensemble Kalman filter in reservoir engineering–a review. *SPE Journal*, 14(03):393–412, 2009.

[2] A. Adler and W. R. B. Lionheart. Uses and abuses of EIDORS: an extensible software base for EIT. *Physiological measurement*, 27(5):S25–S42, 2006.

[3] S. Agapiou, J. M. Bardsley, O. Papaspiliopoulos, and A. M. Stuart. Analysis of the Gibbs sampler for hierarchical inverse problems. *SIAM/ASA Journal on Uncertainty Quantification*, 2(1):511–544, 2014.

[4] S. Agapiou, S. Larsson, and A. M. Stuart. Posterior contraction rates for the Bayesian approach to linear ill-posed inverse problems. *Stochastic Processes and their Applications*, 123(10):3828–3860, 2013.

[5] S. Agapiou, A. M. Stuart, and Y.-X. Zhang. Bayesian posterior contraction rates for linear severely ill-posed inverse problems. *Journal of Inverse and Ill-posed Problems*, 22(3):297–321, 2014.

[6] G. Alessandrini. Stable determination of an inclusion by boundary measurements. *Applicable Analysis: An International Journal*, 27:153–172, 1988.

[7] C. D. Aliprantis and K. Border. *Infinite Dimensional Analysis: A Hitchhiker's Guide*. 2007.

[8] L. Alvarez and J. M. Morel. Formalization and computational aspects of image analysis. *Acta Numerica*, 3:1–59, 1994.

[9] T. Arbogast, M. F. Wheeler, and I. Yotov. Mixed finite elements for elliptic problems with tensor coefficients as cell-centered finite differences. *SIAM J. Numer. Anal.*, 34:828–852, 1997.

[10] J. Bear. *Dynamics of Fluids in Porous Media*. Dover Pulications, New York, 1972.

[11] M. Bédard. Optimal acceptance rates for Metropolis algorithms: Moving beyond 0.234. *Stochastic Processes and their Applications*, 118(12):2198–2222, 2008.

[12] A. Beskos, A. Jasra, E. A. Muzaffer, and A. M. Stuart. Sequential Monte Carlo methods for Bayesian elliptic inverse problems. *Statistics and Computing*, 25(4):727–737, 2015.

[13] A. Beskos, G. O. Roberts, A. M. Stuart, and J. Voss. MCMC methods for diffusion bridges. *Stochastics and Dynamics*, 8:319–350, 2008.

[14] L. Biegler, G. Biros, O. Ghattas, M. Heinkenschloss, D. Keyes, B. Mallick, L. Tenorio, B. van Bloemen Waanders, K. Willcox, and Y. Marzouk. *Large-scale inverse problems and quantification of uncertainty*, volume 712. John Wiley & Sons, 2011.

[15] C. M. Bishop. *Pattern recognition and machine learning*. Springer, 2006.

[16] V. I. Bogachev. Differentiable measures and the Malliavin calculus. *Journal of Mathematical Sciences*, 87(4):3577–3731, 1997.

[17] V. I. Bogachev. *Measure Theory Volume I*. Springer, 2007.

[18] V. I. Bogachev and V. I. Bogachev. *Gaussian measures*, volume 62. American Mathematical Society Providence, 1998.

[19] D. Bolin and F. Lindgren. Excursion and contour uncertainty regions for latent Gaussian models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(1):85–106, 2015.

[20] L. Borcea. Electrical impedance tomography. *Inverse Problems*, 18:R99–R136, 2002.

[21] T. Bui-Thanh and O. Ghattas. A scalable MAP solver for Bayesian inverse problems with Besov priors. 9(November):27–53, 2012.

[22] T. Bui-Thanh and O. Ghattas. An analysis of infinite dimensional Bayesian inverse shape acoustic scattering and its numerical approximation. *SIAM/ASA Journal on Uncertainty Quantification*, 2(1):203–222, 2014.

[23] M. Burger. A level set method for inverse problems. *Inverse Problems*, 17(5):1327–1355, 2001.

[24] D. Calvetti and E. Somersalo. A Gaussian hypermodel to recover blocky objects. *Inverse Problems*, 23(2):733–754, 2007.

[25] D. Calvetti and E. Somersalo. Hypermodels in the Bayesian imaging framework. *Inverse Problems*, 24(3):34013, 2008.

[26] J. Carrera and S. P. Neuman. Estimation of aquifer parameters under transient and steady state conditions: 3. application to synthetic and field data. *Water Resources Research*, 22(2):228–242, 1986.

[27] J. Carter and D. White. History matching on the imperial college fault model using parallel tempering. *Computational Geosciences*, 17(1):43–65, 2013.

[28] T. F. Chan, S. Esedoglu, and M. Nikolova. Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM journal on applied mathematics*, 66(5):1632–1648, 2006.

[29] M. Cheney, D. Isaacson, and J. C. Newell. Electrical impedance tomography. *SIAM review*, 41(1):85–101, 1999.

[30] T. Choi, R. Ramamoorthi, et al. Remarks on consistency of posterior distributions. In *Pushing the limits of contemporary statistics: contributions in honor of Jayanta K. Ghosh*, pages 170–186. Institute of Mathematical Statistics, 2008.

[31] E. T. Chung, T. F. Chan, and X.-C. Tai. Electrical impedance tomography using level set representation and total variational regularization. *Journal of Computational Physics*, 205(1):357–372, 2005.

[32] D. Colton and R. Kress. *Inverse acoustic and electromagnetic scattering theory*, volume 93. Springer Science & Business Media, 2012.

[33] D. Colton, M. Piana, and R. Potthast. A simple method using Morozov ' s discrepancy principle for solving inverse scattering problems. *Inverse Problems*, 13(6):1477–1493, 1997.

[34] S. L. Cotter, G. O. Roberts, A. M. Stuart, and D. White. MCMC methods for functions modifying old algorithms to make them faster. *Statistical Science*, 28(3):424–446, 2013.

[35] G. Da Prato and J. Zabczyk. *Second order partial differential equations in Hilbert spaces*, volume 293. Cambridge University Press, 2002.

[36] B. Dacorogna. *Direct methods in the calculus of variations*. Springer, 2008.

[37] H. Darcy. *Les fontaines publiques de la ville de Dijon.* 1856.

[38] M. Dashti, K. J. H. Law, A. M. Stuart, and J. Voss. MAP estimators and their consistency in Bayesian nonparametric inverse problems. *Inverse Problems*, 29(9):095017, 2013.

[39] M. Dashti and A. M. Stuart. The Bayesian approach to inverse problems. In *Handbook of Uncertainty Quantification*. Springer, 2016 (to appear).

[40] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics*, 57(11):1413–1457, 2004.

[41] O. Dorn and D. Lesselier. Level set methods for inverse scattering. *Inverse Problems*, 22(4):R67–R131, 2006.

[42] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth. Hybrid Monte Carlo. *Physics Letters B*, 195(2):216–222, 1987.

[43] M. M. Dunlop, M. A. Iglesias, and A. M. Stuart. Hierarchical Bayesian level set inversion. *arXiv:1601.03605*, 2016.

[44] M. M. Dunlop and A. M. Stuart. MAP estimators for piecewise continuous inversion. *arXiv:1509.03136*, 2015.

[45] M. M. Dunlop and A. M. Stuart. The Bayesian Formulation of EIT: Analysis and Algorithms. *arXiv:1508.04106*, 2015.

[46] H. Dym and H. P. McKean. *Gaussian processes, function theory, and the inverse spectral problem*. Courier Corporation, 2008.

[47] H. W. Engl and H. Gfrerer. A posteriori parameter choice for general regularization methods for solving linear ill-posed problems. *Applied numerical mathematics*, 4(5):395–417, 1988.

[48] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*, volume 375. Springer Science & Business Media, 1996.

[49] M. Filippone and M. Girolami. Pseudo-marginal Bayesian inference for Gaussian processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(11):2214–2226, 2014.

[50] J. N. Franklin. Well posed stochastic extensions of ill posed linear problems. *Journal of Mathematical Analysis and Applications*, 31(3):682–716, 1970.

[51] G.-A. Fuglstad, D. Simpson, F. Lindgren, and H. Rue. Interpretable priors for hyperparameters for Gaussian random fields. *arXiv:1503.00256*, 2015.

[52] M. Gehre, B. Jin, and X. Lu. An analysis of finite element approximation in electrical impedance tomography. *Inverse Problems*, 30(4):045013, 2014.

[53] Ó. P. Geirsson, B. Hrafnkelsson, D. Simpson, and H. Siguroarson. The MCMC split sampler: A block Gibbs sampling scheme for latent Gaussian models. *arXiv:1506.06285*, 2015.

[54] H. Gfrerer. An a posteriori parameter choice for ordinary and iterated Tikhonov regularization of ill-posed problems leading to optimal convergence rates. *Mathematics of Computation*, 49(180):507–522, 1987.

[55] D. Gilbarg and N. S. Trudinger. *Elliptic partial differential equations of second order*. Springer, 2015.

[56] M. Girolami and B. Calderhead. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 73(2):123–214, 2011.

[57] M. Grasmair, M. Haltmeier, and O. Scherzer. The residual method for regularizing ill-posed problems. *Applied Mathematics and Computation*, 218(6):2693–2710, 2011.

[58] C. W. Groetsch and C. Groetsch. *Inverse problems in the mathematical sciences*, volume 52. Springer, 1993.

[59] D. Guo, S. Shamai, and S. Verdu. Additive non-Gaussian noise channels: mutual information and conditional mean estimation. *Proceedings International Symposium on Information Theory 2005*, (1):719–723, 2005.

[60] J. Hadamard. Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton University Bulletin*, 13:49–52, 1902.

[61] M. Hairer. An introduction to stochastic pdes. *arXiv preprint arXiv:0907.4178*, 2009.

[62] M. Hairer, A. M. Stuart, and S. J. Vollmer. Spectral gaps for a Metropolis-Hastings algorithm in infinite dimensions. *arXiv*, 24(November):1–31, 2011.

[63] M. Hanke. A regularizing Levenberg-Marquardt scheme, with applications to inverse groundwater filtration problems. *Inverse Problems*, 13:79–95, 1997.

[64] R. M. Haralick, S. R. Sternberg, and X. Zhuang. Image analysis using mathematical morphology. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (4):532–550, 1987.

[65] T. Helin and M. Burger. Maximum a posteriori probability estimates in infinite-dimensional Bayesian inverse problems. *Inverse Problems*, 31(8):085009, 2015.

[66] R. P. Henderson and J. G. Webster. An impedance camera for spatially specific measurements of the thorax. *IEEE transactions on bio-medical engineering*, 25(3):250–254, 1978.

[67] M. A. Iglesias. A regularizing iterative ensemble Kalman method for PDE-constrained inverse problems. *Inverse Problems*, 32(2), 2016.

[68] M. A. Iglesias and C. Dawson. The representer method for state and parameter estimation in single-phase Darcy flow. *Computer Methods in Applied Mechanics and Engineering*, 196(1):4577–4596, 2007.

[69] M. A. Iglesias, K. J. H. Law, and A. M. Stuart. The ensemble Kalman filter for inverse problems. *Inverse Problems*, 29(4):045001, 2013.

[70] M. A. Iglesias, K. J. H. Law, and A. M. Stuart. Evaluation of Gaussian approximations for data assimilation in reservoir models. *Computational Geosciences*, 17:851–885, 2013.

[71] M. A. Iglesias, K. Lin, and A. M. Stuart. Well-posed Bayesian geometric inverse problems arising in subsurface flow. *Inverse Problems*, 30, 2014.

[72] M. A. Iglesias, Y. Lu, and A. M. Stuart. A Bayesian level set method for geometric inverse problems. *Interfaces and Free Boundary Problems*, 2016 (to appear).

[73] J. P. Kaipio, V. Kolehmainen, E. Somersalo, and M. Vauhkonen. Statistical inversion and Monte Carlo sampling methods in electrical impedance tomography. *Inverse Problems*, 16(5):1487–1522, 2000.

[74] J. P. Kaipio, V. Kolehmainen, M. Vauhkonen, and E. Somersalo. Inverse problems with structural prior information. *Inverse Problems*, 15:713–729, 1999.

[75] J. P. Kaipio and E. Somersalo. *Statistical and Computational Inverse Problems*. Springer, 2005.

[76] R. E. Kass. Markov Chain Monte Carlo in practice: A roundtable discussion. *The American Statistician*, 52(2):93–100, 1998.

[77] C. E. Kenig, J. Sjöstrand, and G. Uhlmann. The Calderón problem with partial data on manifolds and applications. *Analysis and PDE*, 6(8):2003–2048, 2013.

[78] M. C. Kennedy and A. O'Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464, 2001.

[79] D. Kinderlehrer and G. Stampacchia. *An introduction to variational inequalities and their applications*, volume 31. 1980.

[80] A. Kirsch. *An introduction to the mathematical theory of inverse problems*, volume 120. Springer Science & Business Media, 2011.

[81] A. Kirsch and R. Kress. Uniqueness in inverse obstacle scattering (acoustics). *Inverse Problems*, 9(2):285, 1993.

[82] K. Knudsen, M. Lassas, J. L. Mueller, and S. Siltanen. D-bar method for electrical impedance tomography with discontinuous conductivities. *SIAM Journal on Applied Mathematics*, 67(3):893–913, 2007.

[83] K. Knudsen, M. Lassas, J. L. Mueller, and S. Siltanen. Reconstructions of piecewise constant conductivities by the D-bar method for electrical impedance tomography. *Journal of Physics: Conference Series*, 124(1):012029, 2008.

[84] K. Knudsen, M. Lassas, J. L. Mueller, and S. Siltanen. Regularized D-bar method for the inverse conductivity problem. *Inverse Problems and Imaging*, 3(4):599–624, 2009.

[85] V. Kolehmainen, S. Arridge, W. Lionheart, M. Vauhkonen, and J. Kaipio. Recovery of region boundaries of piecewise constant coefficients of an elliptic pde from boundary data. *Inverse Problems*, 15(5):1375, 1999.

[86] V. Kolehmainen, M. Lassas, P. Ola, and S. Siltanen. Recovering boundary shape and conductivity in electrical impedance tomography. *Inverse Problems and Imaging*, 7(1):217–242, 2013.

[87] D. Krishnan and R. Fergus. Fast image deconvolution using hyper-laplacian priors. In *Advances in Neural Information Processing Systems*, pages 1033–1041, 2009.

[88] D. Kundur and D. Hatzinakos. Blind image deconvolution. *Signal Processing Magazine, IEEE*, 13(3):43–64, 1996.

[89] S. Lan, T. Bui-Thanh, M. Christie, and M. Girolami. Emulation of higher-order tensors in manifold Monte Carlo methods for Bayesian inverse problems. 2015.

[90] R. E. Langer. An inverse problem in differential equations. *Bulletin of the American Mathematical Society*, 39(10):814–821, 1933.

[91] S. Lasanen. Non-Gaussian statistical inverse problems. Part I: Posterior distributions. *Inverse Problems & Imaging*, 6(2), 2012.

[92] S. Lasanen. Non-Gaussian statistical inverse problems. Part II: Posterior convergence for approximated unknowns. *Inverse Problems & Imaging*, 6(2), 2012.

[93] S. Lasanen, J. M. J. Huttunen, and L. Roininen. Whittle-Matérn priors for Bayesian statistical inversion with applications in electrical impedance tomography. *Inverse Problems and Imaging*, 8(2):561–586, 2014.

[94] M. Lassas, E. Saksman, and S. Siltanen. Discretization-invariant Bayesian inversion and Besov space priors. *Inverse Problems and Imaging*, 3, 2009.

[95] E. L. Lehmann and G. Casella. *Theory of Point Estimation*. 1998.

[96] M. S. Lehtinen, L. Paivarinta, and E. Somersalo. Linear inverse problems for generalised random variables. *Inverse Problems*, 5(4):599–612, 1999.

[97] M. A. Lifshits. *Gaussian random functions*, volume 322. Springer, 1995.

[98] A. Litman, D. Lesselier, and F. Santosa. Reconstruction of a two-dimensional binary obstacle by controlled evolution of a level-set. *Inverse Problems*, 14(3):685–706, 1998.

[99] R. J. Lorentzen, K. M. Flornes, and G. Naevdal. History matching channelized reservoirs using the ensemble Kalman filter. *Society of Petroleum Engineers Journal*, 17, 2012.

[100] R. J. Lorentzen, G. Nævdal, and A. Shafieirad. Estimating facies fields by use of the ensemble Kalman filter and distance functions–applied to shallow-marine environments. *Society of Petroleum Engineers Journal*, 3, 2012.

[101] A. Mandelbaum. Linear estimators and measurable linear transformations on a Hilbert space. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 65(3):385–397, 1984.

[102] B. Matérn. *Spatial variation*, volume 36. Springer Science & Business Media, 2013.

[103] D. Middleton. Non-Gaussian noise models in signal processing for telecommunications : New methods and results for Class A and Class B noise models. *IEEE Transactions on Information Theory*, 45(4):1129–1149, 1999.

[104] P. Milasevic and G. R. Ducharme. Uniqueness of the spatial median. *Annals of Statistics*, 15(3):1332–1333, 1987.

[105] A. I. Nachman. Reconstructions from boundary measurements. *Annals of Mathematics*, 128(3):531–576, 1988.

[106] A. I. Nachman. Global uniqueness for a two-dimensional inverse boundary value problem. *Annals of Mathematics*, 143(1):71–96, 1996.

[107] G. K. Nicholls and C. Fox. Prior modelling and posterior sampling in impedance imaging. In A. Mohammad-Djafari, editor, *Bayesian Inference for Inverse Problems*, volume 3459, pages 116–127. SPIE, 1998.

[108] J. Nocedal and S. Wright. *Numerical optimization*. Springer Science & Business Media, 2006.

[109] J. R. Norris. *Markov chains*. Number 2. Cambridge university press, 1998.

[110] R. G. Novikov. Multidimensional inverse spectral problem for the equation $-\Delta\psi + (v(x) - Eu(x))\psi = 0$. *Functional Analysis and Its Applications*, 22(4):263–272, 1988.

[111] D. S. Oliver and Y. Chen. Recent progress on reservoir history matching: a review. *Computational Geosciences*, 15(1):185–221, 2011.

[112] D. S. Oliver, L. B. Cunha, and A. C. Reynolds. Markov chain monte carlo methods for conditioning a permeability field to pressure data. *Mathematical Geology*, 29(1):61–91, 1997.

[113] D. S. Oliver, A. C. Reynolds, and N. Liu. *Inverse theory for petroleum reservoir characterization and history matching*. Cambridge University Press, 2008.

[114] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. An iterative regularization method for total variation-based image restoration. *Multiscale Modeling & Simulation*, 4(2):460–489, 2005.

[115] S. Osher and J. A. Sethian. Fronts propagating with curvature dependent speed: algorithms based on Hamilton-Jacobi formulations. *Journal of Computational Physics*, 79:12–49, 1988.

[116] G. C. Papanicolaou and L. Borcea. Network approximation for transport properties of high contrast materials. *SIAM Journal on Applied Mathematics*, 58(2):501–539, 1998.

[117] J. Ping and D. Zhang. History matching of channelized reservoirs with vector-based level-set parameterization. *Society of Petroleum Engineers Journal*, 19, 2014.

[118] A. G. Ramm. *Scattering by an Obstacle*. Springer, 1986.

[119] C. Robert and G. Casella. *Monte Carlo statistical methods*. Springer Science & Business Media, 2013.

[120] V. G. Romanov. *Inverse problems of mathematical physics*. BRILL, 1987.

[121] B. M. Sadler and A. Swami. Analysis of multiscale products for step detection and estimation. *Information Theory, IEEE Transactions on*, 45(3):1043–1051, 1999.

[122] M. Salo. Calderón problem. *Lecture Notes*, 2008.

[123] F. Santosa. A level-set approach for inverse problems involving obstacles. *ESAIM*, 1(January):17–33, 1996.

[124] G. Sapiro. *Geometric partial differential equations and image analysis*. Cambridge University Press, 2006.

[125] O. Scherzer, H. W. Engl, and K. Kunisch. Optimal a posteriori parameter choice for Tikhonov regularization for solving nonlinear ill-posed problems. *SIAM journal on numerical analysis*, 30(6):1796–1838, 1993.

[126] D. Schymura. An upper bound on the volume of the symmetric difference of a body and a congruent copy. *Advances in Geometry*, 4(2):12, 2014.

[127] M. K. Sen and P. L. Stoffa. Nonlinear one-dimensional seismic waveform inversion using simulated annealing. *Geophysics*, 56(10):1624–1638, 1991.

[128] E. Somersalo, M. Cheney, and D. Isaacson. Existence and uniqueness for electrode models for electric current computed tomography. *SIAM Journal on Applied Mathematics*, 52(4):1023–1040, 1992.

[129] J.-L. Starck, M. K. Nguyen, and F. Murtagh. Wavelets and curvelets for image deconvolution: a combined approach. *Signal Processing*, 83(10):2279–2283, 2003.

[130] M. L. Stein. *Interpolation of spatial data: some theory for kriging*. Springer Science & Business Media, 2012.

[131] A. M. Stuart. Inverse problems : a Bayesian perspective. *Acta Numerica*, 19(May 2010):451–559, 2010.

[132] J. Sylvester and G. Uhlmann. A global uniqueness theorem for an inverse boundary value problem. *Annals of Mathematics*, 125(1):153–169, 1987.

[133] X.-C. Tai and T. F. Chan. A survey on multiple level set methods with applications for identifying piecewise constant functions. *Int. J. Numer. Anal. Model*, 1(1):25–48, 2004.

[134] L. Tierney. Markov chains for exploring posterior distributions. *Annals of Statistics*, 22(4):1701–1728, 1994.

[135] L. Tierney. A note on Metropolis-Hastings kernels for general state spaces. *Annals of Applied Probability*, 8(1):1–9, 1998.

[136] A. W. van der Vaart and J. H. van Zanten. Adaptive Bayesian estimation using a Gaussian random field with inverse gamma bandwidth. *The Annals of Statistics*, pages 2655–2675, 2009.

[137] S. R. S. Varadhan. *Probability Theory*. Courant Lecture Notes. Courant Institute of Mathematical Sciences, 2001.

[138] Z. Šidák. On multivariate normal probabilities of rectangles: their dependence on correlations. *The Annals of Mathematical Statistics*, pages 1425–1434, 1968.

[139] J. Xie, Y. Efendiev, and A. Datta-Gupta. Uncertainty quantification in history matching of channelized reservoirs using Markov chain level set approaches. *Society of Petroleum Engineers*, 2011.

[140] S. Zelditch. Survey of the inverse spectral problem. *arXiv:math/0402356*, 2004.

[141] H. Zhang. Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association*, 99(465):250–261, 2004.