MAP estimators for piecewise continuous inversion

You may also be interested in:

Inverse Modeling: Functional analytic tools
G Nakamura and R Potthast

Inverse Modeling: Analysis: uniqueness, stability and convergence questions
G Nakamura and R Potthast

Inverse Modeling: Nonlinear operators
G Nakamura and R Potthast

Well-posed Bayesian geometric inverse problems arising in subsurface flow
Marco A Iglesias, Kui Lin and Andrew M Stuart

Bayesian approach to inverse problems for functions with a variable index Besov prior
Junxiong Jia, Jigen Peng and Jinghuai Gao

A TV-Gaussian prior for infinite-dimensional Bayesian inverse problems and its numerical
implementations
Zhewei Yao, Zixi Hu and Jinglai Li

Inverse problems with Poisson data: statistical regularization theory, applications and algorithms
Thorsten Hohage and Frank Werner

MAP estimators for piecewise continuous inversion

M M Dunlop[1] and A M Stuart

Mathematics Institute, University of Warwick, Coventry, CV4 7AL, UK

E-mail: matthew.dunlop@warwick.ac.uk and a.m.stuart@warwick.ac.uk

CrossMark

Abstract

We study the inverse problem of estimating a field u^a from data comprising a finite set of nonlinear functionals of u^a, subject to additive noise; we denote this observed data by y. Our interest is in the reconstruction of piecewise continuous fields u^a in which the discontinuity set is described by a finite number of geometric parameters a. Natural applications include groundwater flow and electrical impedance tomography. We take a Bayesian approach, placing a prior distribution on u^a and determining the conditional distribution on u^a given the data y. It is then natural to study maximum *a posterior* (MAP) estimators. Recently (Dashti *et al* 2013 *Inverse Problems* **29** 095017) it has been shown that MAP estimators can be characterised as minimisers of a generalised Onsager–Machlup functional, in the case where the prior measure is a Gaussian random field. We extend this theory to a more general class of prior distributions which allows for piecewise continuous fields. Specifically, the prior field is assumed to be piecewise Gaussian with random interfaces between the different Gaussians defined by a finite number of parameters. We also make connections with recent work on MAP estimators for linear problems and possibly non-Gaussian priors (Helin and Burger 2015 *Inverse Problems* **31** 085009) which employs the notion of Fomin derivative. In showing applicability of our theory we focus on the groundwater flow and EIT models, though the theory holds more generally. Numerical experiments are implemented for the groundwater flow model, demonstrating the feasibility of determining MAP estimators for these piecewise continuous models, but also that the geometric formulation can lead to multiple nearby (local) MAP estimators. We relate these MAP estimators to the behaviour of output from

---

[1] Author to whom any correspondence should be addressed.

MCMC samples of the posterior, obtained using a state-of-the-art function space Metropolis–Hastings method.

(Some figures may appear in colour only in the online journal)

1. Introduction

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 paper 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 characterised variationally, linking to deterministic methods.

There are two main approaches taken to the study of the posterior. The first is to discretise the space, and then apply finite dimensional Bayesian methodology [18]. 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 discretise to implement. This approach has been studied for linear problems in [12, 25, 27], and more recently for nonlinear problems [10, 21, 22, 33]. It is the latter approach that we focus on in this paper.

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 [24]. Three different estimates are commonly considered: the posterior mean which minimises L^2 loss, the posterior median which minimises L^1 loss, and posterior modes which minimise zero-one loss. The former two estimates are unique [28], 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 paper 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 [18]. In the infinite-dimensional setting there is no Lebesgue density, but there has been recent research aimed at characterising the mode variationally and linking to the classical regularisation techniques described in, for example, [9] in the case when Gaussian priors are adopted. Non-Gaussian priors have also been considered in the infinite dimensional setting—in [14] 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.
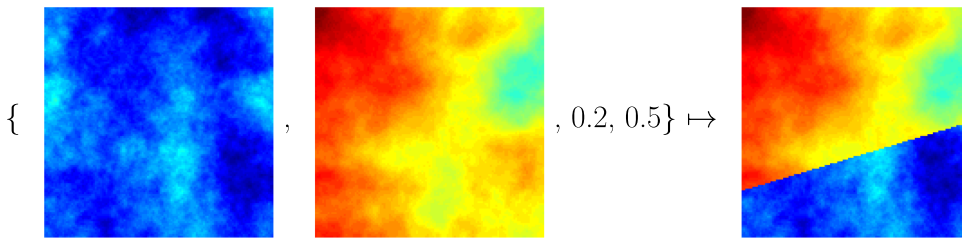
Figure 1. An example of construction of a piecewise continuous field, using two continuous fields and two scalar parameters. Here the scalar parameters determine the points where the interface meets each side of the domain. We work on the space of continuous fields and parameters, but it is pushforward of these by the construction map that represents the piecewise continuous field we aim to recover.

In this paper we make a significant extension of the work in [9] 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 [5], 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. Such piecewise continuous fields were also previously studied in a groundwater flow context in [16], where existence and well-posedness of the posterior distribution were shown. 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 [14], by characterising the MAP estimator via the Fomin derivative.

Throughout this paper 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, with applications of great economic and societal value. MAP estimation in such problems has been studied previously [2, 4, 17, 31]. 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.

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 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
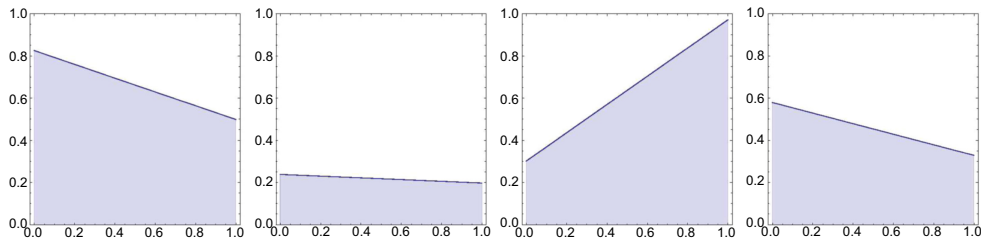
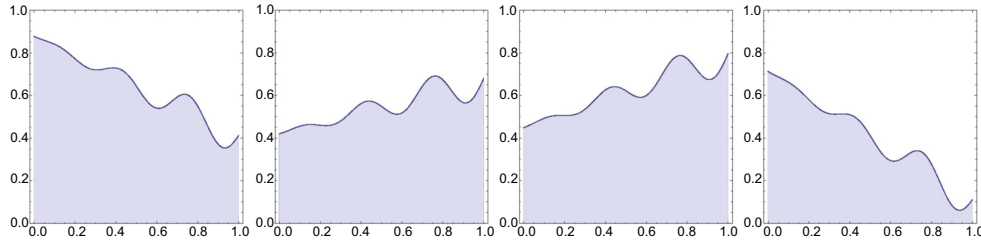**Figure 2.** Possible sets A_i corresponding to example 2.1.

Figure 3. Possible sets A_i, corresponding to example 2.2.
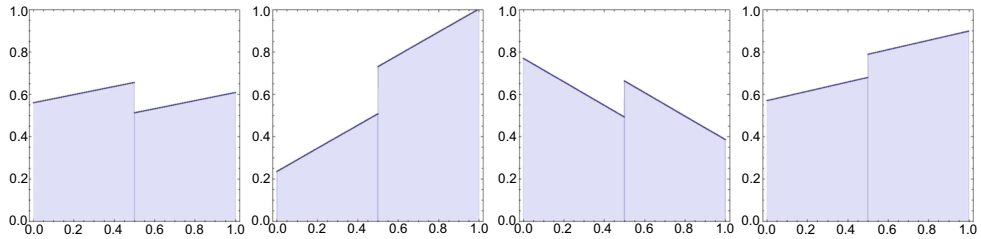
Figure 4. Possible sets A_i, corresponding to example 2.3 in the case $p = 1/2$.

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 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 [9]. Given $(u, a) \in X \times \Lambda$, denote by $B^\delta(u, a)$ the ball of radius δ centred at (u, a).

Definition 1.1 (**MAP estimator**)**.** For each $\delta > 0$, define

$$(u^\delta, a^\delta) = \underset{(u,a)\in X\times\Lambda}{\operatorname{argmax}} \mu(B^\delta(u, a)).$$
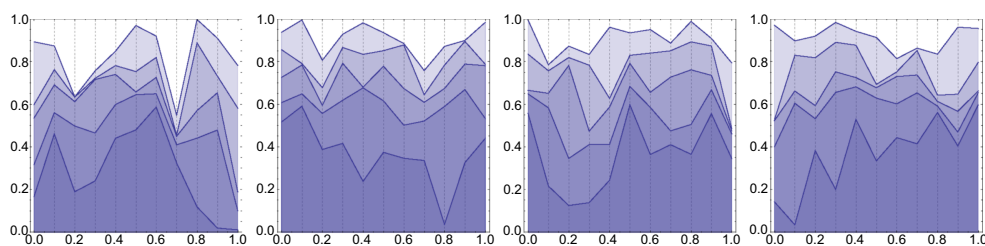
Figure 5. Possible sets A_i, corresponding to example 2.4 in the case $K = 11$, $N = 6$.

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.

1.3. Our contribution

The primary contributions of the paper 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 [9, 14], 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 minimisation problem.
(iii) We show that the theory applies to the groundwater flow model as in [16] and we show that the theory applies to the EIT problem as in [11].
(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.

1.4. Structure of the paper
   • In section 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 4–5.
   • In section 3 we describe the choice of, and assumptions upon, the prior distribution whose samples comprise piecewise Gaussian random fields with random interfaces.
   • In section 4 we show existence and uniqueness of the posterior distribution.
   • In section 5 we define MAP estimators and prove their equivalence to minimisers of an appropriate Onsager–Machlup functional.
   • In section 6 we present numerics for the groundwater flow problem. We consider three different prior models and investigate maximisers of the posterior distribution.
   • In section 7 we conclude and outline possible future work in the area.
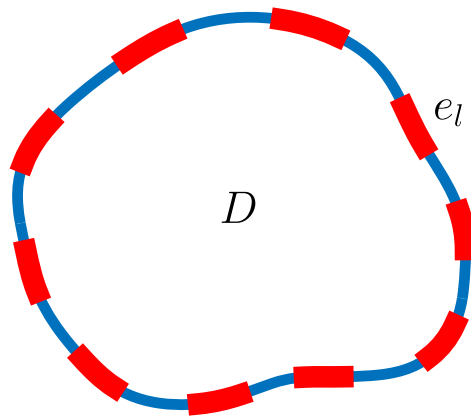
Figure 6. An example domain D, with attached electrodes $(e_l)_{l=1}^{L}$, for the EIT problem.

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 section 2.1 by defining the construction map $(u, a) \mapsto u^a$ for the piecewise continuous fields. In sections 2.2 and 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.

2.1. 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. Take a collection of set-valued maps $A_i : \Lambda \to \mathcal{B}(D)$, $i = 1, \ldots, N$ such that for each $a \in \Lambda$ we have

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

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:

$$A \Delta B := (A \backslash B) \bigcup (B \backslash A).$$

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

$$u^a = F(u, a) := \sum_{i=1}^{N} u_i \mathbb{1}_{A_i(a)}. \tag{2.1}$$

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 2.1. 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 $A_2(a, b) = D \backslash A_1(a, b)$. Example sets $A_i(a, b)$ for various parameters a, b are shown in figure 2.

Example 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 $A_2(a, b) = D \backslash A_1(a, b)$. 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

$$|A_1(a_1, b_1) \Delta A_1(a_2, b_2)| = |A_2(a_1, b_1) \Delta A_2(a_2, b_2)|$$
$$\leqslant \int_0^1 |H(a_1, b_1)(x) - H(a_2, b_2)(x)| \, dx$$
$$\leqslant \|H(a_1, b_1) - H(a_2, b_2)\|_\infty$$

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 4.

Example 2.3. We can generalise 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 3.

Example 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

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

Then given $a \in \Lambda$, define the functions $f_i(a)$, $i = 1, ..., N - 1$, to be the linear interpolation of the points $\left(\frac{j-1}{K-1}, a_{ij} \right)_{j=1}^{K}$. $A_i(a)$, $i = 1, ..., N - 1$, is then defined to be the region between the

graphs of the functions $f_i(a)$ and $f_{i-1}(a)$, and $A_N(a) = D \setminus \bigcup_{i=1}^{N-1} A_i(a)$. Example sets $A_i(a)$ for various parameters a are shown in figure 5.

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

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

so that we have

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

It follows from properties of the symmetric difference that

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

It hence suffices to show that the maps $A_i(\cdot) \bigcap D_j$ are continuous for all i, j. This follows from the same argument as in example 2.2, for sufficiently small $|a - b|$.

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 [8] 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

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

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 once weakly differentiable functions on D [13]. 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

$$\begin{cases} -\nabla \cdot (\sigma(u^a) \nabla p_{u,a}) = f & \text{in } D \\ p_{u,a} = g & \text{on } \partial D. \end{cases} \tag{2.2}$$

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). Since $g \in H^{1/2}(\partial D)$, by the trace theorem [13] there exists $G \in V$ such that $\mathrm{tr}(G) = g$. The solution $p_{u,a}$ of (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

$$\begin{cases} -\nabla \cdot (\sigma(u^a)\nabla q_{u,a}) = f + \nabla \cdot (\sigma(u^a)\nabla G) & \text{in } D \\ q_{u,a} = 0 & \text{on } \partial D. \end{cases} \quad (2.3)$$

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

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

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

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

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

(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

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

(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 $\ell = (\ell_1, \dots, \ell_J)$, we could take

$$\ell_i(p) = \int_D \frac{1}{(2\pi\varepsilon)^{d/2}} e^{-|x_i - y|^2/2\varepsilon} p(y) \mathrm{d}x, \quad i = 1, \dots, J \quad (2.4)$$

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

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

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

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

Proof.

(i) 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 2.5(ii).
(ii) This follows from the continuity of ℓ and lemma 2.5(iii). □

2.3. The complete electrode model (CEM) for EIT

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 [15] in the context of medical imaging and [20] in the context of subsurface imaging.

The CEM is proposed for the forward model in [32], and shown to agree with experimental data up to measurement precision. In its strong form, the PDE reads

$$\begin{cases} -\nabla \cdot (\kappa(x)\nabla v(x)) = 0 & x \in D \\ \int_{e_l} \kappa \dfrac{\partial v}{\partial n}\, \mathrm{d}S = I_l & l = 1,\ldots,L \\ \kappa(x)\dfrac{\partial v}{\partial n}(x) = 0 & x \in \partial D \setminus \bigcup_{l=1}^{L} e_l \\ v(x) + z_l \kappa(x)\dfrac{\partial v}{\partial n}(x) = V_l & x \in e_l,\, l = 1,\ldots,L. \end{cases} \tag{2.5}$$

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 $(z_l)_{l=1}^{L}$. Figure 6 shows an example domain and attached electrodes. 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 [32] 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 [17].

We can apply different current stimulation patterns to the electrodes 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

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

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

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

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

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

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[2] 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.

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 Lebesgue measure is not defined. Suppose that μ is a measure equivalent to a Gaussian measure μ_0. Then the Onsager–Machlup functional for μ is defined as follows.

Definition 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 $x, y \in E$,

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

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

Remarks 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 normalised.
(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 regularised least squares functional.

---

[2] 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 $A \in \mathcal{M}$.

(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 [9] tells us that the Onsager–Machlup functional for μ is given by

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

(iv) In this paper, 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 does not apply; we provide a suitable extension to the definition in section 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 $|a| \to 0$.

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 [10] 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 [26].

Remark 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 3.4. Define the negative Laplacian with Neumann boundary conditions as follows:

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

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 [10] for details.

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 $\Lambda \backslash \partial S$ 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 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 minimisers. The definition we have chosen is appropriate: if any minimising sequence $(a_n)_{n \geqslant 1} \subseteq \text{int}(S)$ 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.

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 (2.1), then our prior permeability/conductivity distribution on $L^\infty(D)$ is given by the pushforward[3] $\mu_0^* = F^{\#}(\mu_0 \times \nu_0)$. 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 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), $(v, b) \in E \times \text{int}(S)$,

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

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

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

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

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

---

[3] Given a measurable map $F : (X, \mathcal{X}) \to (Y, \mathcal{Y})$ between two measurable spaces, the pushforward of a measure μ on X is the measure $F^{\#}\mu$ on Y defined by $(F^{\#}\mu)(A) = \mu(F^{-1}(A))$ for $A \in \mathcal{Y}$. If a random variable u on X has law μ, then the random variable $F(u)$ on Y has law $F^{\#}\mu$.

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

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

$$\Phi(u, a; y) = \frac{1}{2}|\mathcal{G}(u, a) - y|_\Gamma^2. \tag{4.1}$$

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

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

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 (4.1), and instead assert only that Φ satisfies the following assumptions.

Assumptions 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 $a \in \Lambda'$

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

(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 $|y_1|, |y_2| < r$,

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

(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 $\|u_1\|_X, \|u_2\|_X < r$,

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

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

Theorem 4.2 (**Existence and well-posedness**). *Let assumptions 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,

$$\mu^y(\mathrm{d}u, \mathrm{d}a) = \frac{1}{Z(y)} \exp(-\Phi(u, a; y)) \mu_0(\mathrm{d}u) \nu_0(\mathrm{d}a) \tag{4.2}$$

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 $|y|, |y'| < r$,

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

Remark 4.3. In this paper 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 [10, 23].

We show that for both choices of test models, the potential (4.1) satisfies assumptions 4.1:

Proposition 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. 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*

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

Then Φ satisfies assumptions 4.1, with $X' \times \Lambda' = X \times \Lambda$.

Proof.

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

(ii) Fix $u \in X'$ and $y \in Y$. Propositions 2.6 and 2.7 tell us that $\mathcal{G}(u, \cdot)$ is continuous for either choice of test model. The map $z \mapsto |z - y|_{\Gamma}^2$ is continuous, and so $\Phi(u, \cdot; y)$ is continuous too.

(iii) A consequence of propositions 2.6 and 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 2.6 and 2.7 tell us that $\mathcal{G}(\cdot, a)$ is locally Lipschitz for either choice of test model. The map $z \mapsto |z - y|_{\Gamma}^2$ 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 3, we can therefore apply theorem 4.2 in the cases where the forward map is one of the two described in section 2 and the observational noise is Gaussian. In this case, the constant $M_2(r, \|u\|_X, |a|)$ appearing in assumptions 4.1(iii) is independent of $\|u\|_X$ and $|a|$, and so the integrability condition (iii) in theorem 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.

5. MAP estimators

In section 5.1 we characterise the MAP estimators for the posterior μ in terms of the Onsager–Machlup functional for μ. In section 5.2 we relate this Onsager–Machlup functional to the Fomin derivative of μ, with reference to the work [14].

5.1. MAP estimators and the Onsager–Machlup functional

Throughout this section we assume that μ is given by (4.2). Furthermore we assume that μ_0 has mean zero for simplicity. Additionally, when we assume that assumptions 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

$$I(u, a) = \Phi(u, a) + J(u) + K(a), \tag{5.1}$$

where J, K are as defined in sections 3.2 and 3.3 respectively. In this section we attain the following three results concerning I and μ, which are proved in the appendix.

Theorem 5.1. *Let assumptions 4.1 hold. Then the function I defined by (5.1) is the Onsager–Machlup functional for μ, where the Onsager–Machlup functional is as defined in definition 3.6.*

Theorem 5.2. *Let assumptions 4.1 hold. Then there exists $(\bar{u}, \bar{a}) \in E \times S$ such that*

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

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

Theorem 5.3. *Let assumptions 4.1 hold. Assume also that there exists an $M \in \mathbb{R}$ such that $\Phi(u, a) \geqslant M$ for any $(u, a) \in X \times \Lambda$.*

(i) *Let $(u^\delta, a^\delta) = \underset{(u,a) \in X \times \Lambda}{\mathrm{argmax}}\ \mu(B^\delta(u, a))$. 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 minimiser of I.*

A consequence of theorem 5.3 is that, under its assumptions, MAP estimators and minimisers of the Onsager–Machlup functional are equivalent. The proof of this corollary is identical to that of corollary 3.10 in [9]:

Corollary 5.4. *Under the conditions of theorem 5.3 we have the following.*

(i) *Any MAP estimator minimises the Onsager–Machlup functional I.*
(ii) *Any $(u^*, a^*) \in E \times S$ which minimises the Onsager–Machlup functional I is a MAP estimator for the measure μ given by (4.2).*

5.2. The Fomin derivative approach

In recent work of Helin and Burger [14], the concept of MAP estimators was generalised to 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 minimisers of a particular functional. The assumptions do not hold in our case, since our forward map is nonlinear and our prior $\mu_0 \times \nu_0$ is not 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 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 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 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 $a \in \text{int}(S)$ 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.

(iii) Again using the Cameron–Martin and dominated convergence theorems, we see that with ν_0 and μ_0 as above, for any $(u, a) \in X \times \text{int}(S)$ 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 characterise the Fomin derivative of our posterior distribution μ, given by (4.2).

Theorem 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*

$$\beta_{(h,b)}^\mu(u, a) = -\partial_{(h,b)} \Phi(u, a) - \langle u, h \rangle_E + \partial_b \log \rho(a)$$
$$= -\partial_{(h,b)} I(u, a).$$

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

Proof. We use result (2.1.13) from [3], 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 h 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 5.6 (iii) above. □

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 section 2.2. We both compute minimisers 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.

(i) The posterior distribution can be highly multi-modal, especially when the parameterised 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 initiallised at these calculated minimisers.

(ii) When the geometry is incorrect the fields attempt to compensate, which presumably contributes to the existence of multiple local minimisers of the objective functional; this occurs in both the MAP estimation and the MCMC samples. A consequence is that many of the local minimisers lack the desired sharp interfaces. These minimisers could however be used to suggest more appropriate geometric parameters for the initialisation.

(iii) The mixing rates of MCMC chains have a strong dependence upon which local minimiser they are initialised 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.

(iv) Though often there are many local minimisers, they can be separated into classes of minimisers 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 minimiser 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.
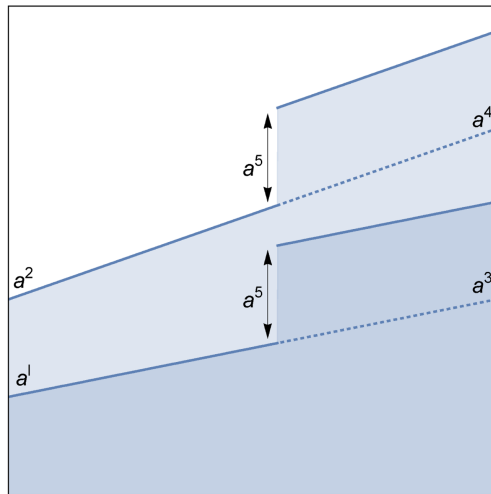
6.1. Test models

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

In what follows, as in example 3.4, we define the negative Laplacian with Neumann boundary conditions:
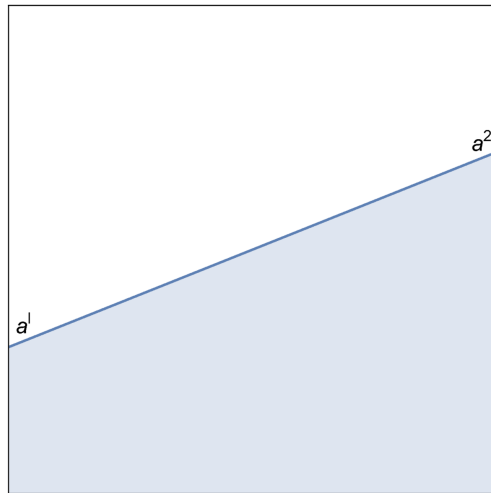
Figure 7. The definition of the geometric parameters $a = (a^1, a^2)$ in model 1.
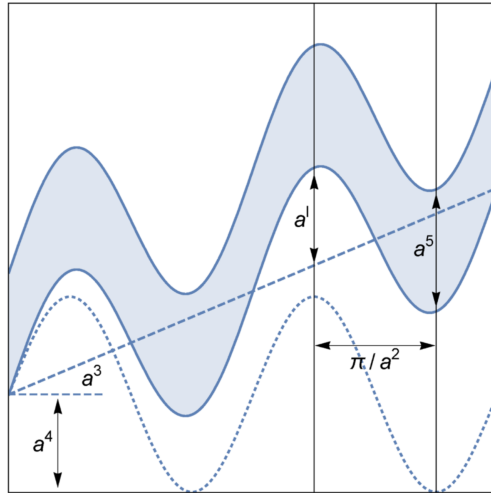

Figure 8. The definition of the geometric parameters $a = (a^1, a^2, a^3, a^4, a^5)$ in model 2.

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

Recall that if $u \sim N(0, A^{-\alpha})$ with $\alpha > d/2$, then u is almost surely continuous [10].

6.1.1. Model 1 (two layer). This model is described in example 2.1. The geometric parameters $a = (a^1, a^2)$ are defined as in figure 7. For simulations, we use the choice of prior

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

Figure 9. The definition of the geometric parameters $a = (a^1, a^2, a^3, a^4, a^5)$ in model 3.

Table 1. The relative error on the data, when each measurement is perturbed by an instance of $N(0, 0.01^2)$ noise.

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

6.1.2. Model 2 (*three layer with fault*). This model is described in [16], 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 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 $S = \{(x, y) \mid 0 \leqslant x \leqslant 1, x \leqslant y \leqslant 1\}$.

6.1.3. Model 3 (*channel*). This model is described in [16], 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 9. Here a^1, a^2, a^3, a^4, a^5 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 (2.4) 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 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.

6.2. MAP estimation

Based on the theory in section 5, we can calculate MAP estimators by minimising the Onsager–Machlup functional for the posterior distribution. We compute local minimisers of the Onsager–Machlup functional using the following iterative alternating method.

Algorithm 6.1.

(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 [30], 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 minimised 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 $\mathrm{TOL} = 10^{-5}$. 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 10–12.

We first comment on the minimisers 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 minimisers associated with model 2 admit much more variation, though it is possible to partition them into smaller subsets of minimisers which share similar characteristics to one another, as mentioned in point (iv) at the beginning of the section. The clustering of the different minimisers is performed by eye, classifying them according to similar geometric parameters. Additionally we have an other class, containing the minimisers 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 maximisers of the posterior probability.

The minimisers of the Onsager–Machlup functional for model 3 show even more variation than those for model 2, with the geometry in half of the minimisers 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,
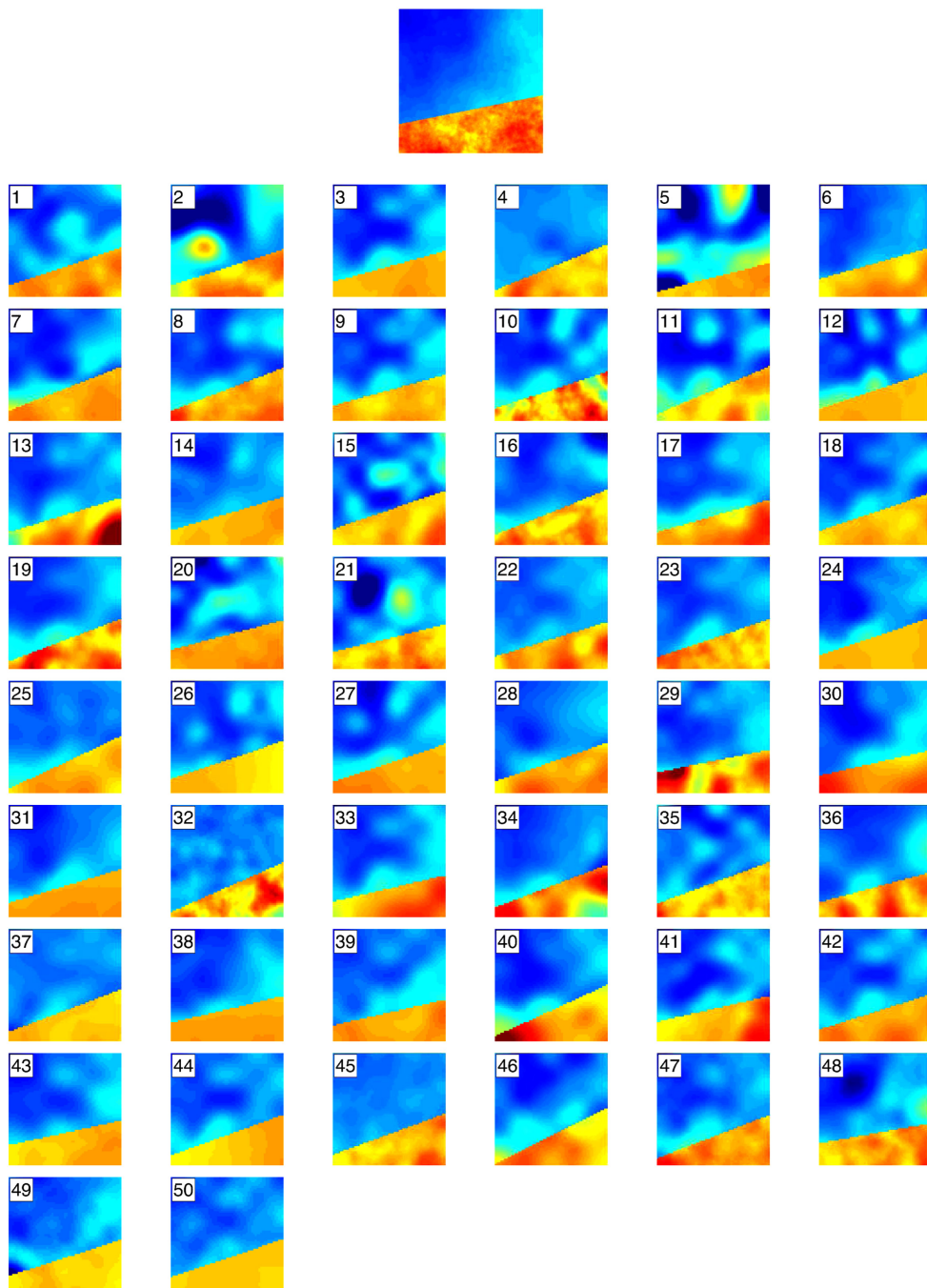
Figure 10. (Model 1) The true log-permeability field (top), and 50 local minimisers arising from minimisation initialised at draws from a smoothed prior distribution. Simulation 12 has the lowest functional value, with $I(u_{\mathrm{MAP}}^{12}, a_{\mathrm{MAP}}^{12}) = 2847$.
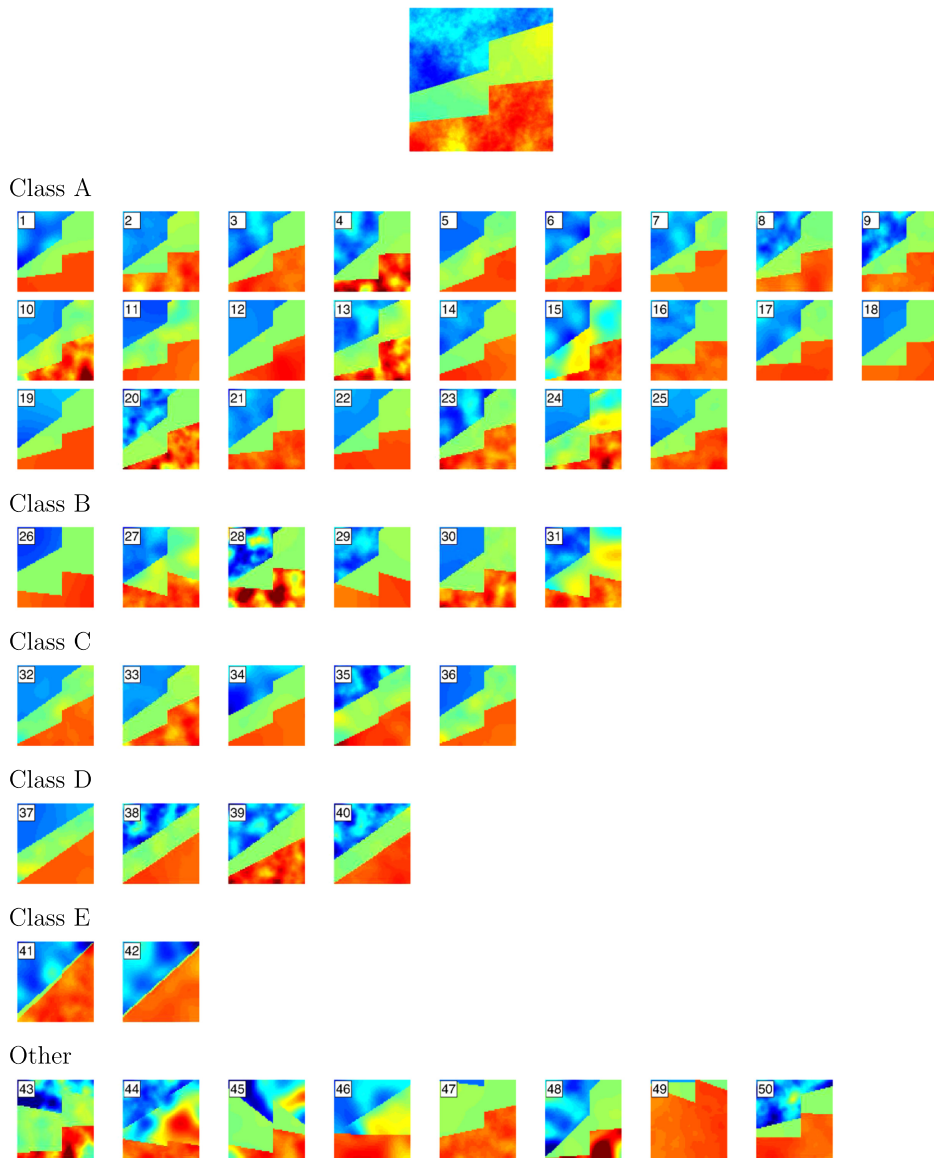
Figure 11. (Model 2) The true log-permeability field (top), and 50 local minimisers arising from minimisation initialised at draws from a smoothed prior distribution. Simulation 7 has the lowest functional value, with $I(u_{\text{MAP}}^7, a_{\text{MAP}}^7) = 2567$. The minimisers have been divided into classes based on similar characteristics.

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 maximisers of the posterior probability.

This multi-modality of the posterior distribution is not unexpected. The paper [5] 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
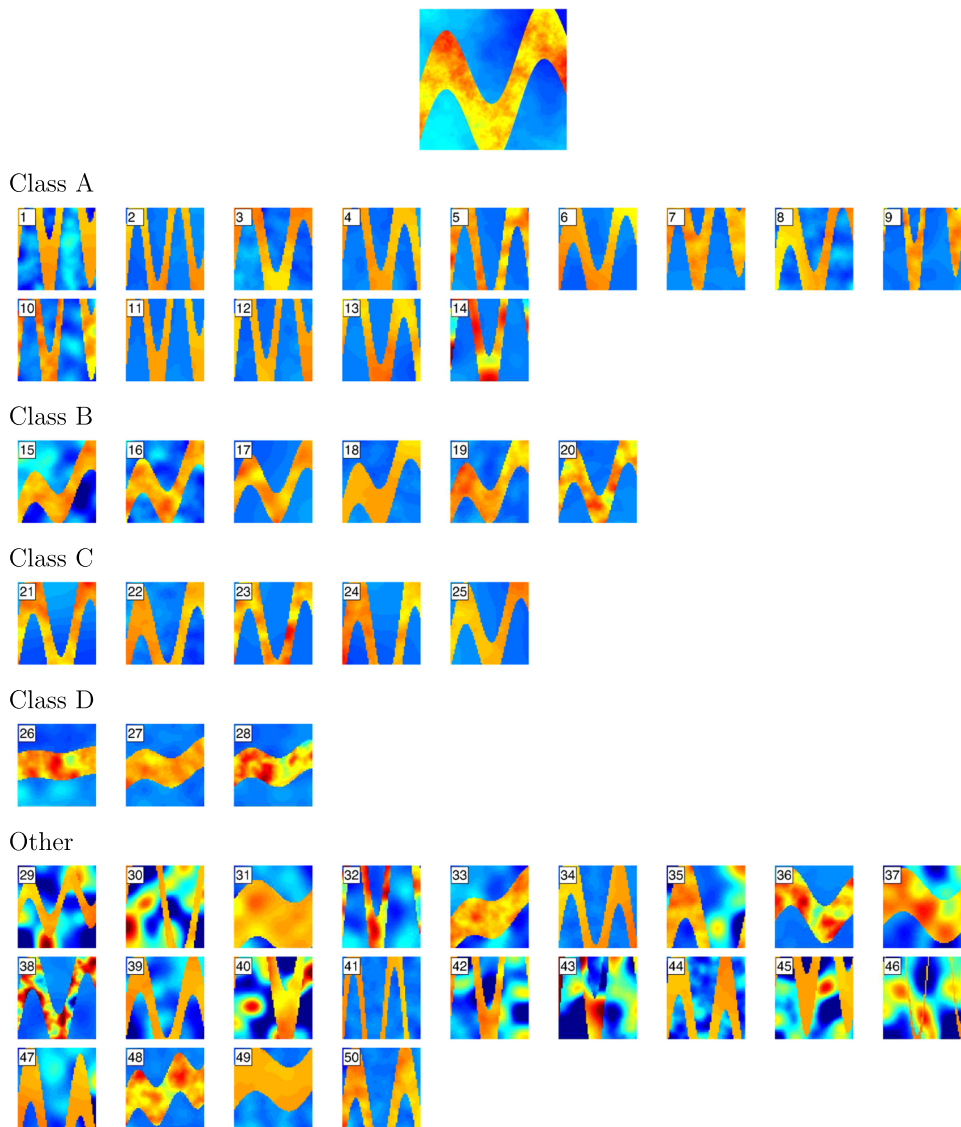
Figure 12. (Model 3) The true log-permeability field (top), and 50 local minimisers arising from minimisation initialised at draws from a smoothed prior distribution. Simulation 20 has the lowest functional value, with $I(u_{\text{MAP}}^{20}, a_{\text{MAP}}^{20}) = 2117$. The minimisers have been divided into classes based on similar characteristics.

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 [5] 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 minimisers 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 minimisers, in view of the links established in the next subsection between MCMC and MAP estimation.

6.3. MCMC and local minimisers

We now observe the behaviour of MCMC chains initialised at these local minimisers of the Onsager–Machlup functional. We use a Metropolis-within-Gibbs algorithm for the sampling, alternating between preconditioned Crank–Nicolson updates for the fields, see [6] 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 13–15.

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 minimisers $(u^i_{\mathrm{MAP}}, a^i_{\mathrm{MAP}})$. This is shown in figures 16–18, with the horizontal lines being the different values of $\Phi(u^i_{\mathrm{MAP}}, a^i_{\mathrm{MAP}})$. 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 minimiser the chain is nearest at each step, in the permeability space. Specifically, we look at

$$m_n := \underset{i}{\mathrm{argmin}} \ \|F(u^{(n)}, a^{(n)}) - F(u^i_{\mathrm{MAP}}, a^i_{\mathrm{MAP}})\|_{L^2(D)}, \tag{6.1}$$

where $F : X \times \Lambda \to L^\infty(D)$ is the construction map (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-penalising incorrect geometry. A selection of traces of m_n are shown in figures 19–21. These illustrate that even though some of the local minimisers are very far from from the true log-permeability, they do indeed act as local maximisers of the posterior probability. Moreover, they show the interaction between the different classes of minimisers in the cases of models 2 and 3. Specifically, they show that the MCMC chains can easily move within these classes, but moving between classes is more difficult.

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 minimisers 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 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 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 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
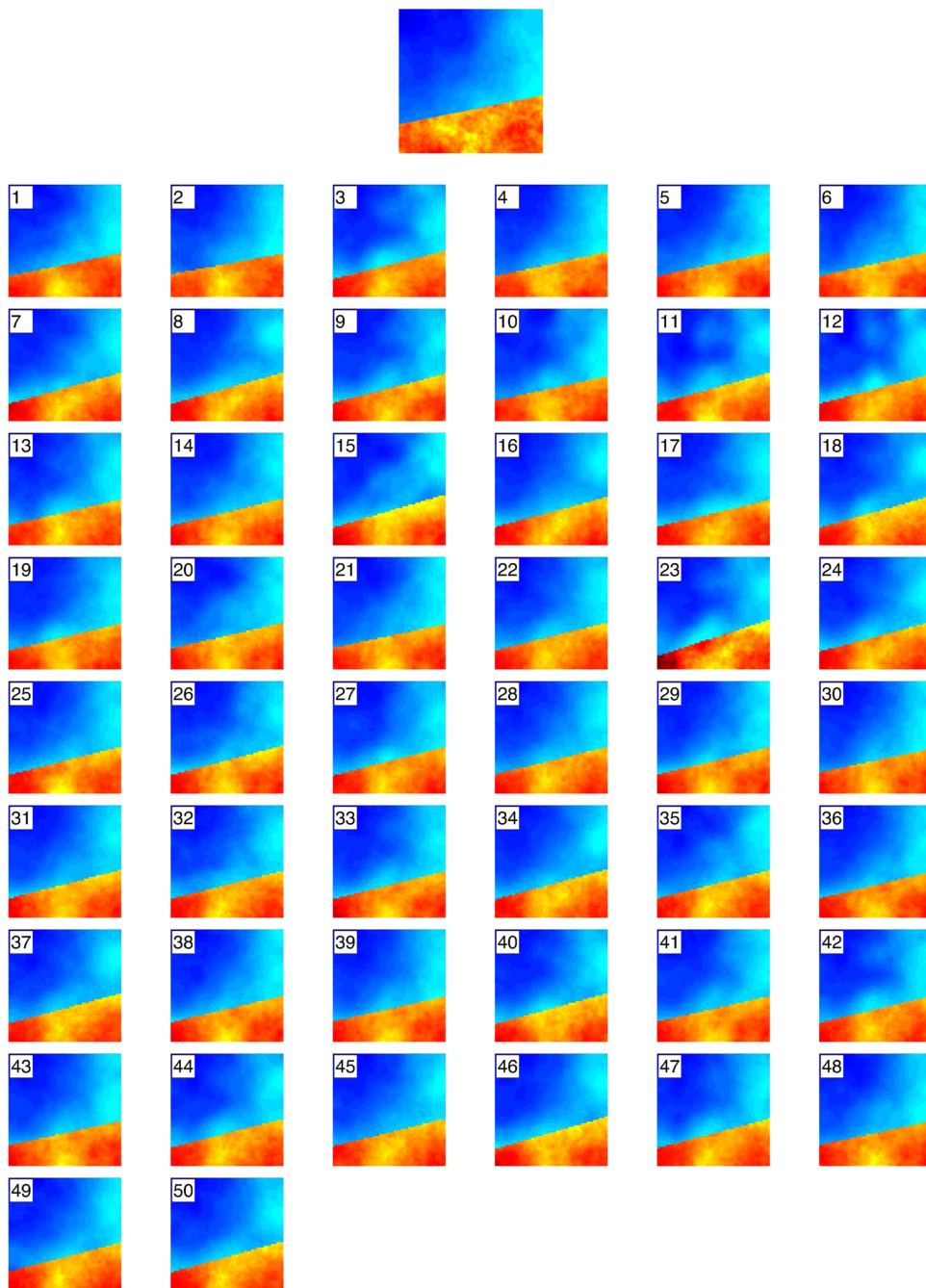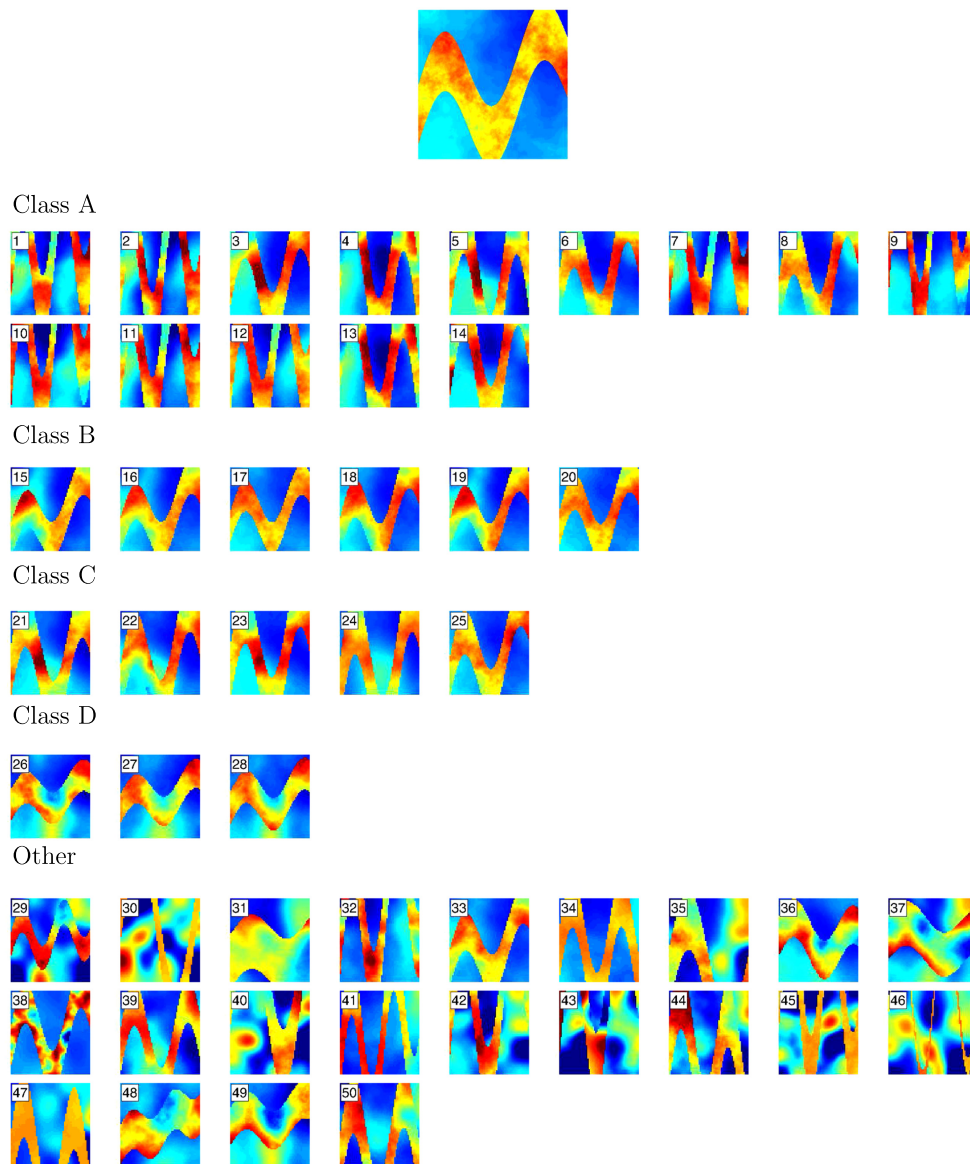
Figure 13. (Model 1) The true log-permeability field (top), and the conditional mean arising from MCMC chains initialised at the corresponding local minimisers above.
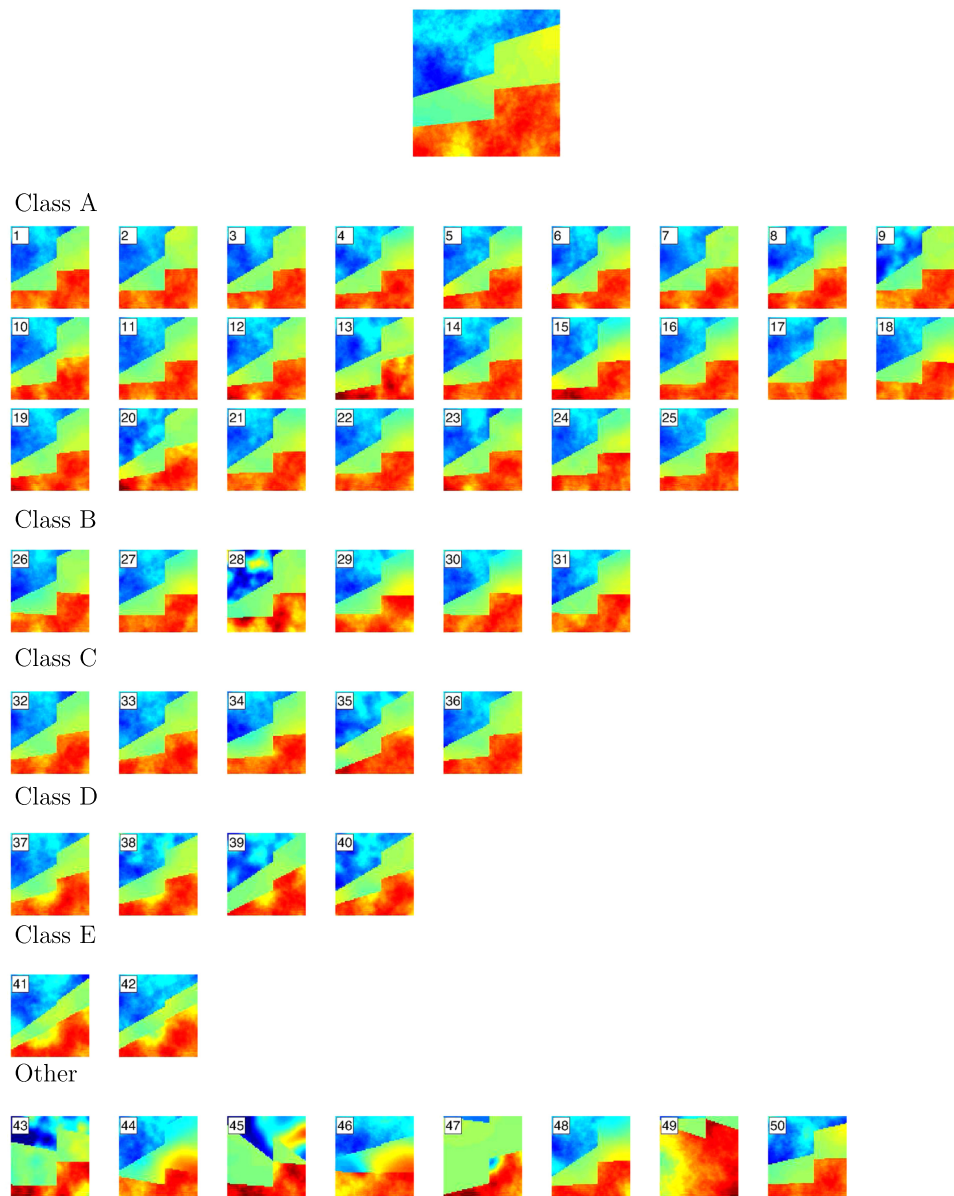
Figure 14. (Model 2) The true log-permeability field (top), and the conditional mean arising from MCMC chains initialised at the corresponding local minimisers above. We group them into the same classes as the local minimisers.

figure 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 maximisers of the posterior probability.

In the channelised model, model 3, there is yet more variation between local minimisers. Here the compensation effect by the fields is even more apparent *in the conditional means*, especially in the other class. From figure 18 it appears that the local minima are much sharper

Figure 15. (Model 3) The true log-permeability field (top), and the conditional mean arising from MCMC chains initialised at the corresponding local minimisers above. We group them into the same classes as the local minimisers.

and more sparsely distributed than the previous two models. Again the chains with the largest Φ values were initialised at minimisers 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 minimiser the chain was started from. This indicates that some of the minima are much sharper than others. This is also evident from the
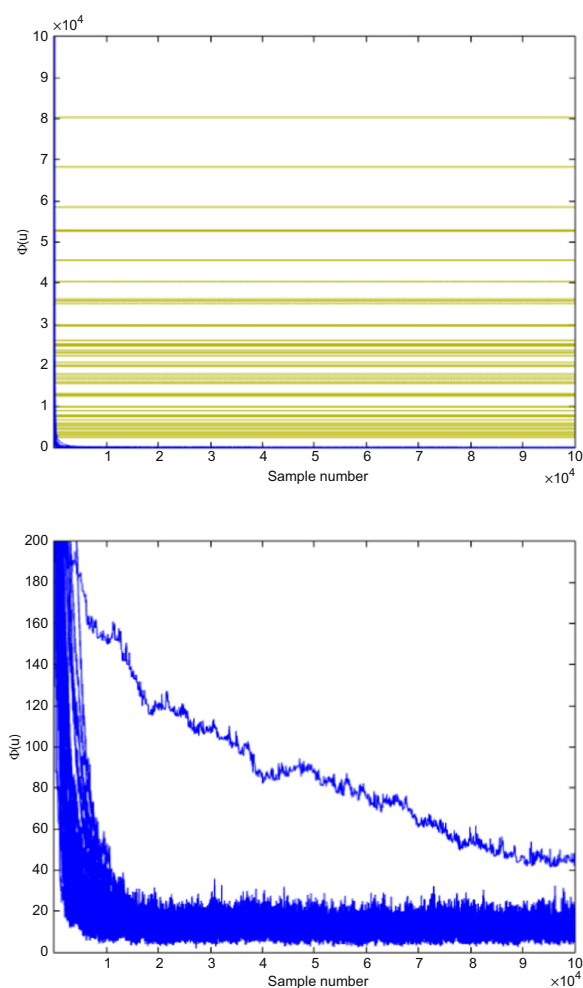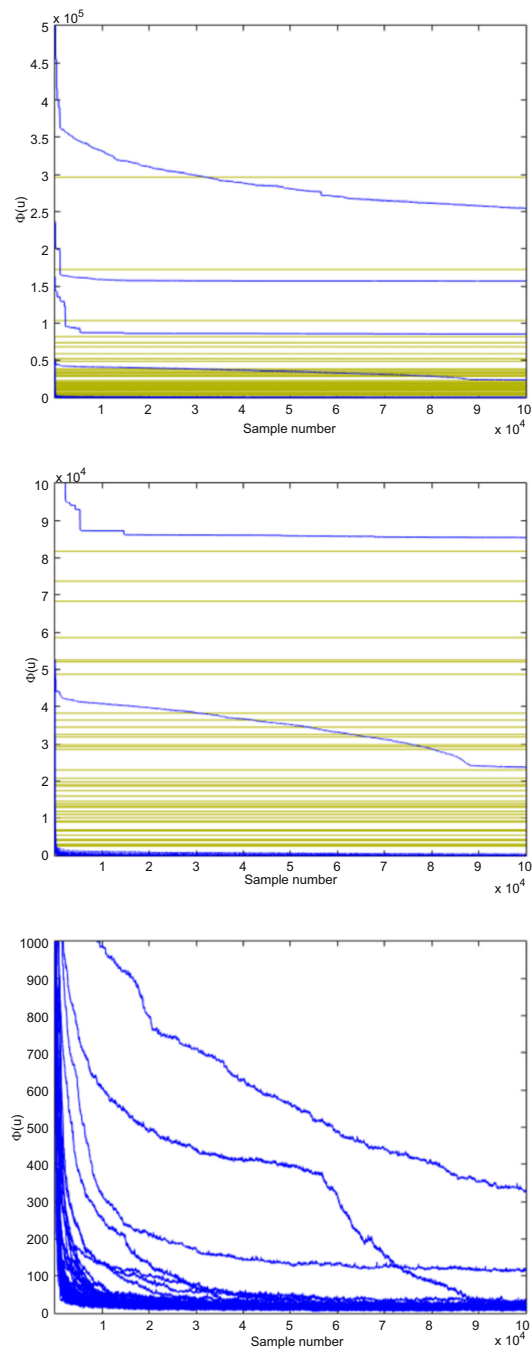
Figure 16. (Model 1) The evolution of Φ as the MCMC chains progress. The horizontal lines represent the value of each local minimiser under Φ. Nearly all of the simulations find a small value of Φ almost immediately, but simulation 23 remains caught in the local minimiser for some time before it follows.

traces of m_n defined above, figures 19–21, 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 minimiser always lies in class A.

7. Conclusions and future work

We have made a new contribution to the recently developed theory of MAP estimation in infinite dimensions [9, 14]. We link MAP estimation to a variational Onsager–Machlup functional. The work is focused on priors for piecewise Gaussian random fields, with random interfaces parameterised 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

Figure 17. (Model 2) The evolution of Φ as the MCMC chains progress. The horizontal lines represent the value of each local minimiser under Φ. 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 14, and generally correspond to those in the same class as the initial state.
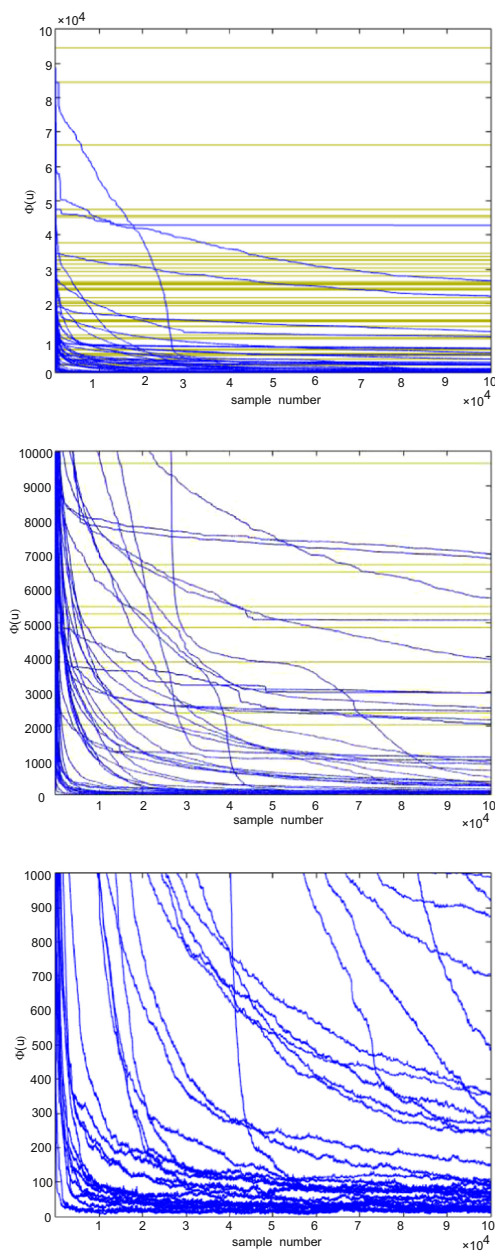
Figure 18. (Model 3) The evolution of Φ as the MCMC chains progress. The horizontal lines represent the value of each local minimiser under Φ. The majority of the simultions find a small value of Φ almost immediately, but numerous fail to reach there, settling in local minima. The shape of these minima can be seen in figure 15, and generally correspond to those in the same class as the initial state.

Figure 19. (Model 1) The trace of m_n as defined by (6.1), when the chain is initialised at a variety of minimisers—specifically numbers $1, 2, \ldots, 8$.

Figure 20. (Model 2) The trace of m_n as defined by (6.1), when the chain is initialised at a variety of minimisers—specifically numbers 7, 14, 21, 28, 35, 39, 46 and 50. The different classes are alternately shaded.
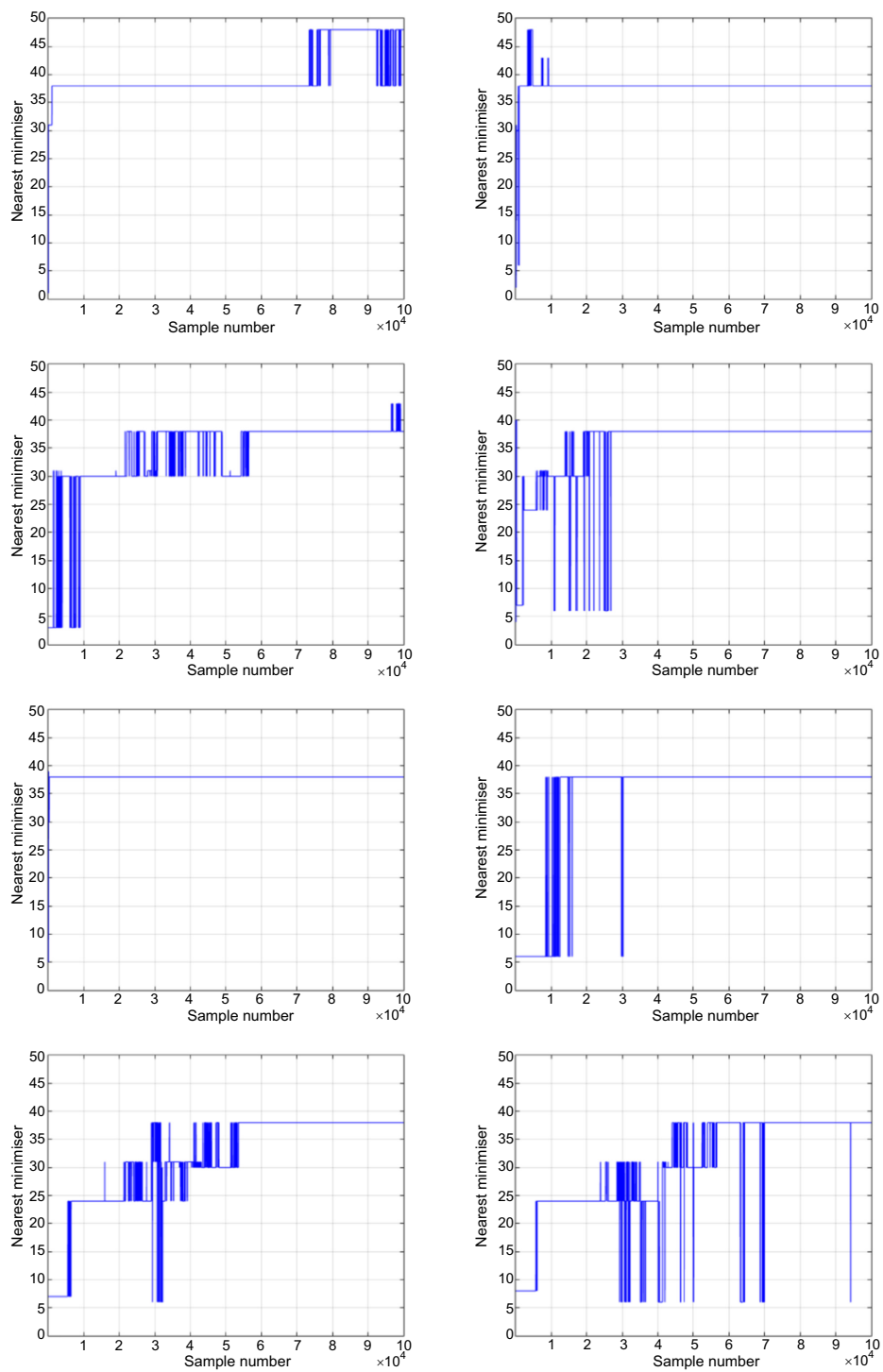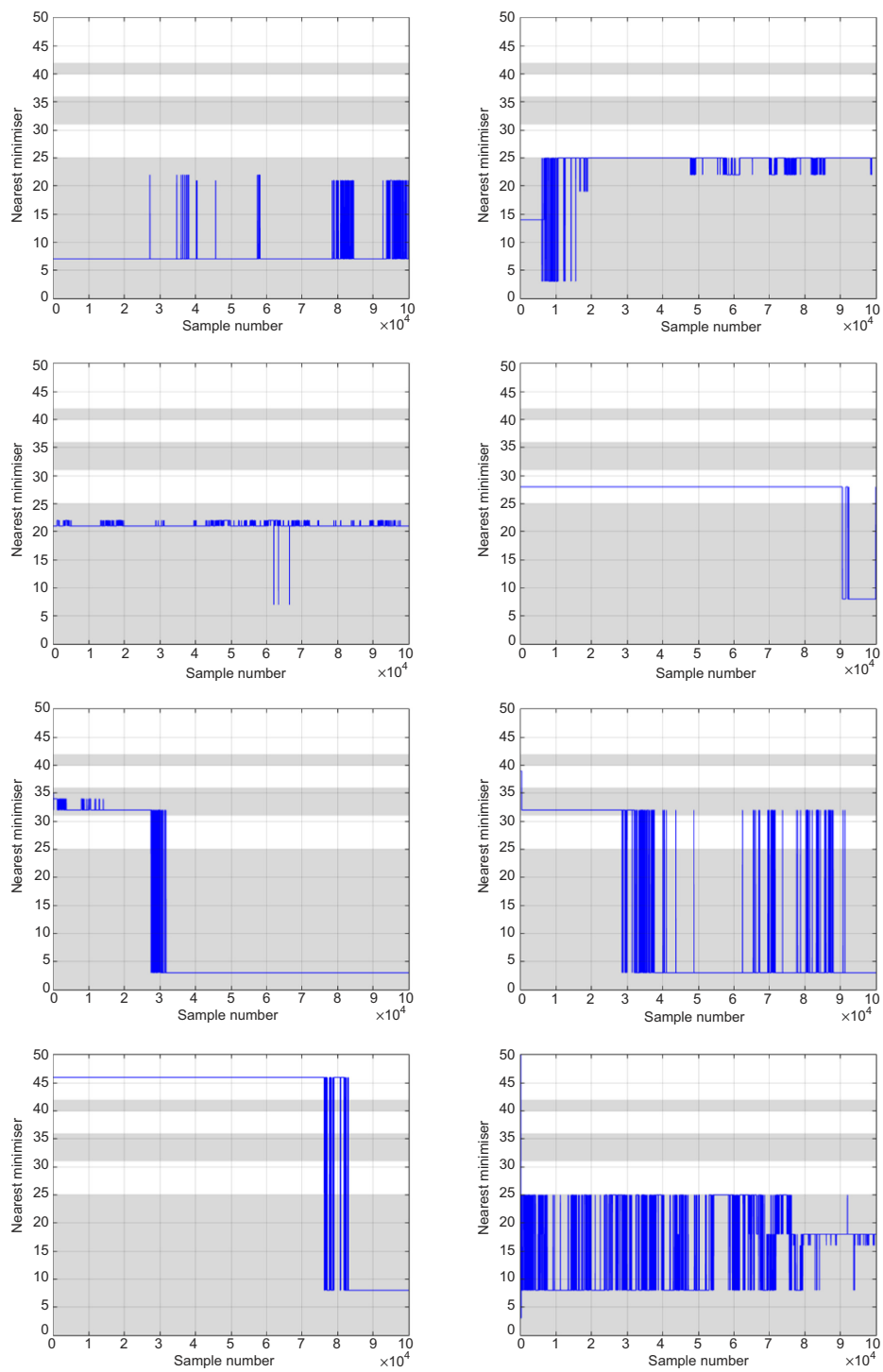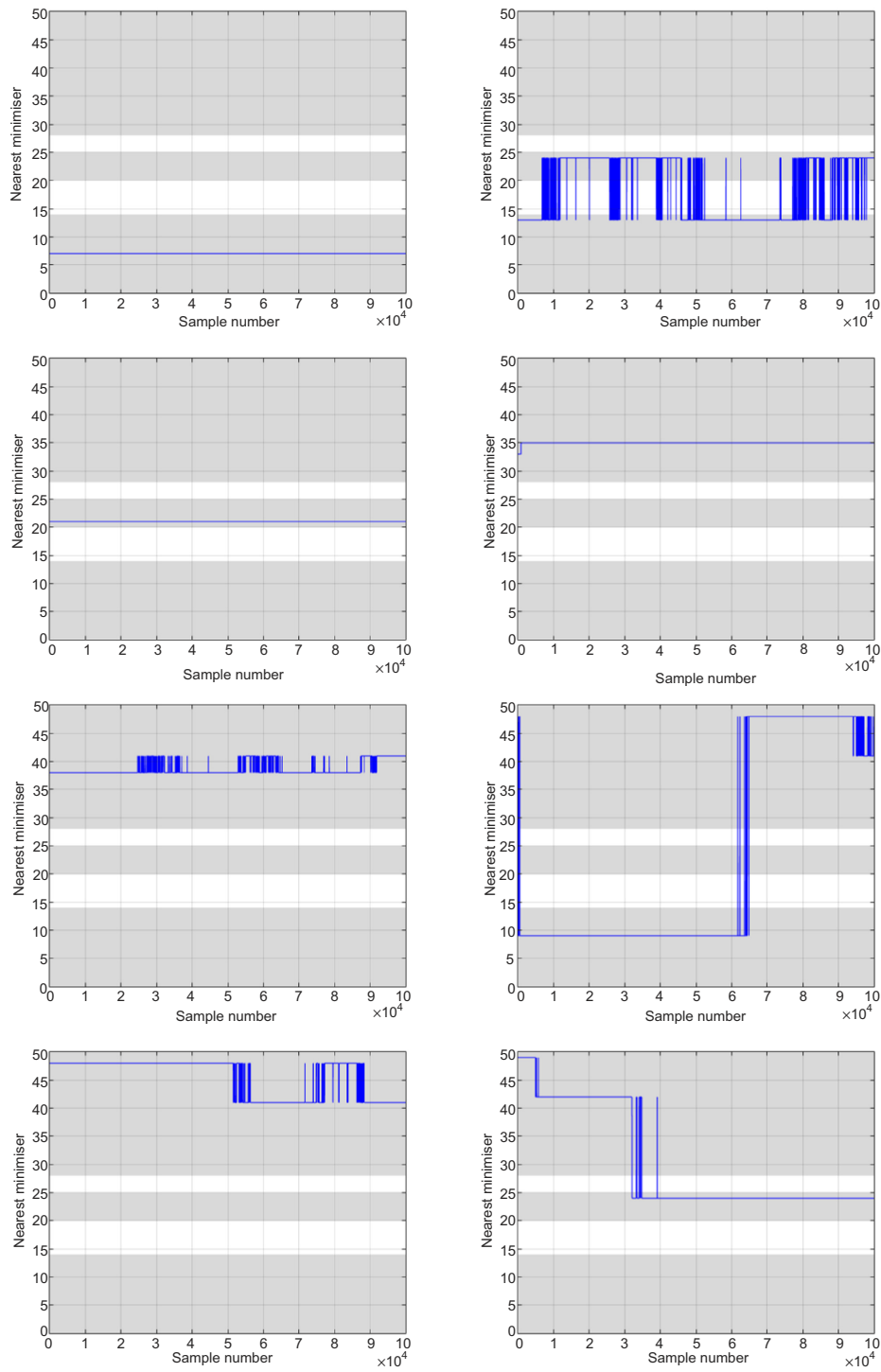
Figure 21. (Model 3) The trace of m_n as defined by (6.1), when the chain is initialised at a variety of minimisers—specifically numbers 7, 13, 21, 33, 38, 47, 48 and 49. The different classes are alternately shaded.

work opens up several new avenues for investigation. A major theoretical direction is to fully reconcile the approaches in [9, 14]; the work in this paper 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 [29], 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 minimiser structure of the Onsager–Machlup functional within MCMC.

Acknowledgments

Appendix

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

A.1. Results from section 2

Before we prove lemma 2.5 we require the following lemma.

Lemma A.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 $\delta > 0$,

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

Now suppose that $\|f_n\|_{L^1}$ does not tend to zero. Then there exists $\delta > 0$ and a subsequence $(f_{n_k})_{k \geqslant 1}$ such that $\|f_{n_k}\|_{L^1} \geqslant \delta$ for all $k \geqslant 1$. 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 2.5. Showing that $\mathcal{R}$ is well-defined is equivalent to showing that PDE (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 $\kappa_{\min} \leqslant \sigma(u^a) \leqslant \kappa_{\max}$. 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 $\sigma(u^a) \leqslant \kappa_{\max}$, and so a unique solution exists by Lax–Milgram.

(i) In its weak form, the PDE (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

$$\kappa_{\min}(u, a) \|\nabla q_{u,a}\|_{L^2}^2 \leqslant \int_D \sigma(u^a) \nabla q_{u,a} \cdot \nabla q_{u,a}$$

$$= f(q_{u,a}) - \int_D \sigma(u^a) \nabla G \cdot \nabla q_{u,a}$$

$$\leqslant \|f\|_{V^*} \|q_{u,a}\|_V + \|\sigma(u^a)\|_{L^\infty} \|\nabla G\|_{L^2} \|\nabla q_{u,a}\|_{L^2}$$

and so we have the estimate

$$\|p_{u,a}\|_V \leqslant \|q_{u,a}\|_V + \|G\|_V$$

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

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

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

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

$$\kappa_*(u, v, a) \|\nabla(p_{u,a} - p_{v,a})\|_{L^2}^2 \leqslant \int_D \sigma(u^a) |\nabla(p_{u,a} - p_{v,a})|^2$$

$$= \int_D (\sigma(u^a) - \sigma(v^a)) \nabla(p_{u,a} - p_{v,a}) \cdot \nabla p_{v,a}$$

$$\leqslant \|\sigma(u^a) - \sigma(v^a)\|_{L^\infty} \|\nabla(p_{u,a} - p_{v,a})\|_{L^2} \|\nabla p_{v,a}\|_{L^2}$$

and so by (i),

$$\|p_{u,a} - p_{v,a}\|_V \leqslant \|p_{v,a}\|_V \|\sigma(u^a) - \sigma(v^a)\|_{L^\infty}/\kappa_*(u, v, a)$$

$$\leqslant \|\sigma(u^a) - \sigma(v^a)\|_{L^\infty}$$

$$\times ((\|f\|_{V^*} + \|\sigma(u^a)\|_{L^\infty} \|G\|_V)/\kappa_*(u, a)^2 + \|G\|_V/\kappa_*(u, a)).$$

Using that the A_i are disjoint gives that

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

$$= \|\sigma(u_k) - \sigma(v_k)\|_{L^\infty}$$

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

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

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

$$\kappa_*(u,\,v,\,a)^{-j} = [(\operatorname*{essinf}_{x \in D} e^{u^a(x)}) \,\wedge\, (\operatorname*{essinf}_{x \in D} e^{v^a(x)})]^{-j}$$

$$\leqslant (\min_{|t| \leqslant r} \sigma(t) \,\wedge\, \min_{|t| \leqslant r} \sigma(t))^{-j}$$

$$= (\min_{|t| \leqslant r} \sigma(t))^{-j}.$$

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

$$\|\mathcal{R}(u,\,a) - \mathcal{R}(v,\,a)\|_V = \|p_{u,a} - p_{v,a}\|_V$$

$$\leqslant \max_{j=1,2} (\min_{|t| \leqslant r} \sigma(t))^{-j} (\|f\|_{V^*} + \|G\|_V (\max_{|t| \leqslant r} \sigma(t) + 1))$$

$$\times \max_{|t| \leqslant r} |\sigma'(t)| \cdot \|u - v\|_X$$

$$= L(r) \|u - v\|_X.$$

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

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

which leads to the bound

$$\kappa_\dagger(u,\,a,\,b)\|\nabla(p_{u,a} - p_{u,b})\|_{L^2}^2 \leqslant \int_D \sigma(u^a)|\nabla(p_{u,a} - p_{u,b})|^2$$

$$= \int_D (\sigma(u^a) - \sigma(u^b))\nabla(p_{u,a} - p_{u,b}) \cdot \nabla p_{u,b}$$

$$\leqslant \|\nabla(p_{u,a} - p_{u,b})\|_{L^2}\|(\sigma(u^a) - \sigma(u^b))\nabla p_{u,b}\|_{L^2},$$

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

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

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

$$\|(\sigma(u^a) - \sigma(u^b))\nabla p_{u,b}\|_{L^2} = \left\|\left(\sigma\left(\sum_{i=1}^N u_i \mathbb{1}_{A_i(a)}\right) - \sigma\left(\sum_{i=1}^N u_i \mathbb{1}_{A_i(b)}\right)\right)\nabla p_{u,b}\right\|_{L^2}$$

$$= \left\|\sum_{i=1}^N (\sigma(u_i \mathbb{1}_{A_i(a)}) - \sigma(u_i \mathbb{1}_{A_i(b)}))\nabla p_{u,b}\right\|_{L^2}$$

$$\leqslant \sum_{i=1}^N \|(\sigma(u_i \mathbb{1}_{A_i(a)}) - \sigma(u_i \mathbb{1}_{A_i(b)}))\nabla p_{u,b}\|_{L^2}$$

$$\leqslant \sum_{i=1}^N \max_{|t| \leqslant \|u_i\|_\infty} |\sigma'(t)| \cdot \||u_i \mathbb{1}_{A_i(a)} - u_i \mathbb{1}_{A_i(b)}| \nabla p_{u,b}\|_{L^2}$$

$$\leqslant \sum_{i=1}^N \max_{|t| \leqslant \|u_i\|_\infty} |\sigma'(t)| \cdot \|u_i\|_\infty \|\mathbb{1}_{A_i(a)\Delta A_i(b)}\nabla p_{u,b}\|_{L^2}$$

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

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

Putting the above bounds together, we have

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

The right hand goes to zero as each $|A_i(a) \triangle A_i(b)| \to 0$ by lemma A.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. $\qquad \square$

Proof of proposition 2.7.

(i) Theorem 2.3 in [17] tells us that the mapping from the conductivity to the weak solution of (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

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

and the result follows.

(ii) By corollary 2.8 in [11] 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

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

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. $\qquad \square$

A.2. Results from section 4

Proof of theorem 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 $(u_n, a_n)_{n \geqslant 1} \subseteq X' \times \Lambda'$ 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

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

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 [1]. 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 4.1(i), we have the lower bound

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

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

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

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

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

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

Now we can proceed as in [10]. 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

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

and so

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

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 [10]. □

A.3. Results from section 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 5.1 and 5.2, we first require two lemmas.

Lemma A.2. *Let* $(u_1, a_1), (u_2, a_2) \in E \times \text{int}(S)$. *Then*

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

$$= \exp\left(J(u_2) + K(a_2) - J(u_1) - K(a_1)\right).$$

Proof. We adapt the proof of proposition 18.3 in [26] 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 [26], 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

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

$$\geqslant (\mu_0 \times \nu_0)(B^\delta(0, a_1))$$

which gives the inequality

$$(\mu_0 \times \nu_0)(B^\delta(u_1, a_1)) \geqslant e^{-\frac{1}{2}\|u_1\|_E^2}(\mu_0 \times \nu_0)(B^\delta(0, a_1)). \tag{8.1}$$

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

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

$$\leqslant \exp\left(-\frac{1}{2}\|u_1\|_E^2 + \delta \cdot \sup_{(u, a) \in B^1(0, a_1)} z_c(u)\right)$$

$$\cdot \left[(\mu_0 \times \nu_0)(B^\delta(0, a_1)) + \int_{B^\delta(0, a_1)} (e^{z_s(u)} - 1)\mu_0(\mathrm{d}u)\nu_0(\mathrm{d}a)\right],$$

where we have used the linearity of z_c to extract δ from the supremum. As in [26], using a result from [34], 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,

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

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

$$\int_{B^\delta(0, a_1)} \varphi(|z_s(u)|) \mu_0(\mathrm{d}u) \nu_0(\mathrm{d}a) = \int_{X \times \Lambda} \varphi(|z_s(u)|) \mathbb{1}_{B^\delta(0, a_1)}(u, a) \mu_0(\mathrm{d}u) \nu_0(\mathrm{d}a)$$

$$= \int_0^\infty (\mu_0 \times \nu_0)(\{(u, a) \in B^\delta(0, a_1) \mid \varphi(|z_s(u)|) > t\}) \mathrm{d}t$$

$$= \int_0^\infty (\mu_0 \times \nu_0)(\{(u, a) \in B^\delta(0, a_1) \mid |z_s(u)| > \varphi^{-1}(t)\}) \mathrm{d}t$$

$$= \int_0^\infty \int_\Lambda \mu_0(\{u \in X \mid (u, a) \in B^\delta(0, a_1), |z_s(u)| > \varphi^{-1}(t)\}) \nu_0(\mathrm{d}a) \mathrm{d}t$$

$$\leqslant \int_0^\infty \int_\Lambda \mu_0(\{u \in X \mid (u, a) \in B^\delta(0, a_1)\}) \mu_0(|z_s(\cdot)| > \varphi^{-1}(t)) \nu_0(\mathrm{d}a) \mathrm{d}t$$

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

$$= (\mu_0 \times \nu_0)(B^\delta(0, a_1)) \int_0^\infty \mu_0(|z_s(\cdot)| > \varphi^{-1}(t)) \mathrm{d}t$$

$$= (\mu_0 \times \nu_0)(B^\delta(0, a_1)) \int_0^\infty \mu_0(\varphi(|z_s(\cdot)|) > t) \mathrm{d}t$$

$$= (\mu_0 \times \nu_0)(B^\delta(0, a_1)) \int_X \varphi(|z_s(u)|) \mu_0(\mathrm{d}u).$$

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

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

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

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

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

$$(\mu_0 \times \nu_0)(B^\delta(u_1, a_1))$$

$$\leqslant \exp\left(-\frac{1}{2} \|u_1\|_E^2 + \delta \cdot \sup_{(u, a) \in B^1(0, a_1)} z_c(u) \right) (\mu_0 \times \nu_0)(B^\delta(0, a_1))(1 + \varepsilon). \tag{8.2}$$

Since balls are bounded, $\varepsilon > 0$ is arbitrary and z_c is continuous, we can combine (8.1) and (8.2) to deduce that there exists $M > 0$ such that

$$\mathrm{e}^{-\frac{1}{2}\|u_1\|_E^2}(\mu_0 \times \nu_0)(B^\delta(0, a_1)) \leqslant (\mu_0 \times \nu_0)(B^\delta(u_1, a_1))$$

$$\leqslant \mathrm{e}^{-\frac{1}{2}\|u_1\|_E^2 + M\delta}(\mu_0 \times \nu_0)(B^\delta(0, a_1)).$$

Now looking at the ratio of measures we see

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

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

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

We conclude that

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

$\square$

Lemma A.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*

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

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

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

The result now follows by the previous lemma. $\square$

Proof of theorem 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 A.2. Now proceeding analagously to [9],

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

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

Using assumptions 4.1(iv), we have that for any $(u, a), (v, b) \in X \times \Lambda$,

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

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

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

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

$$
\frac{\mathcal{J}^\delta(u_1, a_1)}{\mathcal{J}^\delta(u_2, a_2)} \leqslant \mathrm{e}^{\delta(L_1+L_2)}\mathrm{e}^{-\Phi(u_1,a_1)+\Phi(u_2,a_2))}
$$

$$
\times \frac{\displaystyle\int_{B^\delta(u_1,a_1)} \exp(|\Phi(u_1, a) - \Phi(u_1, a_1)|)\mu_0(\mathrm{d}u)\nu_0(\mathrm{d}a)}{\displaystyle\int_{B^\delta(u_2,a_2)} \exp(-|\Phi(u_2, a) - \Phi(u_2, a_2)|)\mu_0(\mathrm{d}u)\nu_0(\mathrm{d}a)}.
$$

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

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

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

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

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 $(X\backslash E) \times \Lambda$ and $E \times (\Lambda\backslash S)$. $\qquad\square$

Remark A.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 $X \times \Lambda$,

$$
\lim_{\delta \downarrow 0} \frac{(\mu_0 \times \nu_0)(B^\delta(u_1, a_1))}{(\mu_0 \times \nu_0)(B^\delta(u_2, a_2))} = \lim_{\delta \downarrow 0} \frac{(\mu_0 \times \nu_0)(B^\delta(u_1) \times B^\delta(a_1))}{(\mu_0 \times \nu_0)(B^\delta(u_2) \times B^\delta(a_2))}
$$

$$
= \lim_{\delta \downarrow 0} \frac{\mu_0(B^\delta(u_1))}{\mu_0(B^\delta(u_2))} \cdot \frac{\nu_0(B^\delta(a_1))}{\nu_0(B^\delta(a_2))}.
$$

This will be useful later for separating integrals.

Proof of theorem 5.2. We follow the idea of the proof of theorem 5.4 in [33], which is based on [7, 19], 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

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

Therefore by assumption 4.1(i) it follows that, for any $\varepsilon > 0$, there is an $M(\varepsilon) \in \mathbb{R}$ such that

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

Since K is bounded below[4] by $-\log \|\rho\|_\infty$, we may incorporate this into the constant term $M(\varepsilon)$:

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

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

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

which establishes coercivity.

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

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

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 $\delta > 0$,

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

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

---

[4] Recall in section 3.3 we assumed ρ to be continuous on the compact set S, and hence bounded.

same argument as in the proof of theorem 5.4 in [33] (taking $\bar{a}$ as the second parameter in I and Φ) and so the second result follows. $\qquad\square$

Before proving theorem 5.3 we first collect some results on centred Gaussian measures from [9], 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 A.5.

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

$$\frac{\mathcal{J}_0^\delta(u)}{\mathcal{J}_0^\delta(0)} \leqslant c\mathrm{e}^{-\frac{a_1}{2}(\|u\|_X - \delta)^2},$$

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 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 [9].

In the proof of existence of the posterior measure, theorem 4.2, we show that if $r > 0$ and $\|u\|_X, |a| < r$, then there exists $K(r) > 0$ such that $\Phi(u, a) \leqslant K(r)$. Letting $c = \mathrm{e}^M \mathrm{e}^{-K(1)} > 0$, it follows in the same was as [9] that, given any $a \in S$, for $\delta < 1$ we have

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

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

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

This contradicts proposition A.5(i) above. Therefore there exists $R, \delta_R > 0$ such that

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

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[5] α close to 1,

$$1 \leqslant \frac{\mathcal{J}^\delta(u^\delta, 0)}{\mathcal{J}^\delta(0, 0)} \leqslant \alpha \frac{e^{-M} \int_{B^\delta(u^\delta)} \mu_0(\mathrm{d}u) \int_{B^\delta(0)} \nu_0(\mathrm{d}a)}{e^{-K(1)} \int_{B^\delta(0)} \mu_0(\mathrm{d}u) \int_{B^\delta(0)} \nu_0(\mathrm{d}a)}$$

$$= \alpha e^{K(1)-M} \frac{\int_{B^\delta(u^\delta)} \mu_0(\mathrm{d}u)}{\int_{B^\delta(0)} \mu_0(\mathrm{d}u)}.$$

We use proposition A.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 such that

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

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 A.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 minimises I. As in [9], and the proof of theorem 5.1, we can use assumptions 4.1(iii) to see that

$$\frac{\mathcal{J}^\delta(u^\delta, a^\delta)}{\mathcal{J}^\delta(\bar{u}, \bar{a})} \leqslant e^{\delta(L_1+L_2)} e^{-\Phi(u^\delta, a^\delta) + \Phi(\bar{u}, \bar{a}))}$$

$$\times \frac{\int_{B^\delta(u^\delta, a^\delta)} \exp(|\Phi(u^\delta, a) - \Phi(u^\delta, a^\delta)|) \mu_0(\mathrm{d}u) \nu_0(\mathrm{d}a)}{\int_{B^\delta(\bar{u}, \bar{a})} \exp(-|\Phi(\bar{u}, a) - \Phi(\bar{u}, \bar{a})|) \mu_0(\mathrm{d}u) \nu_0(\mathrm{d}a)},$$

---

[5] Remark A.4 tells us that we can separate the integrals in the limit $\delta \downarrow 0$.

where

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

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 $X \times \Lambda$,

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

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 centred and $\|u^\delta - \bar{u}\|_X \to 0$, $|a^\delta - \bar{a}| \to 0$,

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

$$\leqslant \limsup_{\delta \downarrow 0} \frac{\int_{B^\delta(u^\delta)} \mu_0(\mathrm{d}u)}{\int_{B^\delta(\bar{u})} \mu_0(\mathrm{d}u)} \cdot \limsup_{\delta \downarrow 0} \frac{\int_{B^\delta(a^\delta)} \nu_0(\mathrm{d}a)}{\int_{B^\delta(\bar{a})} \nu_0(\mathrm{d}a)}$$

$$\leqslant \limsup_{\delta \downarrow 0} \frac{\int_{B^\delta(a^\delta)} \nu_0(\mathrm{d}a)}{\int_{B^\delta(\bar{a})} \nu_0(\mathrm{d}a)}$$

$$= \limsup_{\delta \downarrow 0} \frac{\frac{1}{|B^\delta(a^\delta)|} \int_{B^\delta(a^\delta)} \rho(a)\mathrm{d}a}{\frac{1}{|B^\delta(\bar{a})|} \int_{B^\delta(\bar{a})} \rho(a)\mathrm{d}a}$$

$$= 1.$$

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), $\mathcal{J}^\delta(u^\delta, a^\delta) \geqslant \mathcal{J}^\delta(\bar{u}, \bar{a})$ and hence

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

this implies that

$$\lim_{\delta \downarrow 0} \frac{\mathcal{J}^\delta(u^\delta, a^\delta)}{\mathcal{J}^\delta(\bar{u}, \bar{a})} = 1. \tag{8.3}$$

In the case where (u^δ) converges strongly to $\bar{u}$ in E, we see from the proof of lemma A.2 that we have

$$e^{\frac{1}{2}\|\bar{u}\|_E^2 - \frac{1}{2}\|u^\delta\|_E^2 - M\delta} \frac{(\mu_0 \times \nu_0)(B^\delta(0, a^\delta))}{(\mu_0 \times \nu_0)(B^\delta(0, \bar{a}))} \leqslant \frac{(\mu_0 \times \nu_0)(B^\delta(u^\delta, a^\delta))}{(\mu_0 \times \nu_0)(B^\delta(\bar{u}, \bar{a}))}$$

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

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 A.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 (8.3) follows again. Therefore $(\bar{u}, \bar{a})$ is a MAP estimator of the measure μ.

The proof that $(\bar{u}, \bar{a})$ minimises I is identical to that in the proof of theorem 3.5 in [9]. $\square$

References

[1] Aliprantis C D and Border K 2007 *Infinite Dimensional Analysis: a Hitchhiker's Guide* (Berlin: Springer-Verlag) pp 1–705
[2] Biegler L, Biros G, Ghattas O, Heinkenschloss M, Keyes D, Mallick B, Tenorio L, van Bloemen Waanders B, Willcox K and Marzouk Y (ed) 2011 *Large-Scale Inverse Problems and Quantification of Uncertainty* vol 712 (New York: Wiley) pp 1–372
[3] Bogachev V I 1997 Differentiable measures and the Malliavin calculus *J. Math. Sci.* **87** 3577–731
[4] Bui-Thanh T and Ghattas O 2012 A scalable MAP solver for Bayesian inverse problems with Besov priors *Inverse Problems Imaging* **9** 27–53
[5] Carter J and White D 2013 History matching on the imperial college fault model using parallel tempering *Comput. Geosci.* **17** 43–65
[6] Cotter S L, Roberts G O, Stuart A M and White D 2013 MCMC methods for functions modifying old algorithms to make them faster *Stat. Sci.* **28** 424–46
[7] Dacorogna B 2008 *Direct Methods in the Calculus of Variations* (Berlin: Springer)
[8] Darcy H 1856 *Les fontaines publiques de la ville de Dijon* ed V Dalmont (Paris)
[9] Dashti M, Law K J H, Stuart A M and Voss J 2013 MAP estimators and their consistency in Bayesian nonparametric inverse problems *Inverse Problems* **29** 095017
[10] Dashti M and Stuart A M 2013 The Bayesian approach to inverse problems *Handbook of Uncertainty Quantification* ed R Ghanem, D Higdon and H Owhadi (Berlin: Springer)
[11] Dunlop M M and Stuart A M 2015 The Bayesian formulation of EIT: analysis and algorithms arXiv:1508.04106
[12] Franklin J N 1970 Well posed stochastic extensions of ill posed linear problems *J. Math. Anal. Appl.* **31** 682–716
[13] Gilbarg D and Trudinger N S 2001 *Elliptic Partial Differential Equations of Second Order* (Berlin: Springer) pp 1–523
[14] Helin T and Burger M 2015 Maximum *a posteriori* probability estimates in infinite-dimensional Bayesian inverse problems *Inverse Problems* **31** 085009
[15] Henderson R P and Webster J G 1978 An impedance camera for spatially specific measurements of the thorax *IEEE Trans. Biomed. Eng.* **25** 250–4
[16] Iglesias M A, Lin K and Stuart A M 2014 Well-posed Bayesian geometric inverse problems arising in subsurface flow *Inverse Problems* **30** 114001
[17] Kaipio J P, Kolehmainen V, Somersalo E and Vauhkonen M 2000 Statistical inversion and Monte Carlo sampling methods in electrical impedance tomography *Inverse Problems* **16** 1487–522

[18] Kaipio J P and Somersalo E 2005 *Statistical and Computational Inverse Problems* (Berlin: Springer) pp 1–339

[19] Kinderlehrer D and Stampacchia G 1980 *An Introduction to Variational Inequalities and Their Applications* (New York: Academic)

[20] Langer R E 1933 An inverse problem in differential equations *Bull. Am. Math. Soc.* **39** 814–21

[21] Lasanen S 2012 Non-gaussian statistical inverse problems: I. Posterior distributions *Inverse Problems Imaging* **6** 215–66

[22] Lasanen S 2012 Non-gaussian statistical inverse problems: II. Posterior convergence for approximated unknowns *Inverse Problems Imaging* **6** 267–87

[23] Lassas M, Saksman E and Siltanen S 2009 Discretization-invariant Bayesian inversion and Besov space priors *Inverse Problems Imaging* **3** 87–122

[24] Lehmann E L and Casella G 1998 *Theory of Point Estimation* (New York: Springer) pp 1–590

[25] Lehtinen M S, Paivarinta L and Somersalo E 1999 Linear inverse problems for generalised random variables *Inverse Problems* **5** 599–612

[26] Lifshits M A 1995 *Gaussian Random Functions* vol 322 (Berlin: Springer) pp 1–337

[27] Mandelbaum A 1984 Linear estimators and measurable linear transformations on a Hilbert space *Z. Wahrscheinlichkeitstheor. Verwandte Geb.* **65** 385–97

[28] Milasevic P and Ducharme G R 1987 Uniqueness of the spatial median *Ann. Stat.* **15** 1332–3

[29] Nicholls G K and Fox C 1998 Prior modelling and posterior sampling in impedance imaging *Proc. SPIE* **3459** 116

[30] Nocedal J and Wright S 2006 *Numerical Optimization* (New York: Springer-Verlag) pp 1–664

[31] Oliver D, Reynolds A and Liu N 2008 *Inverse Theory for Petroleum Reservoir Characterization* (Cambridge: Cambridge University Press) pp 1–380

[32] Somersalo E, Cheney M and Isaacson D 1992 Existence and uniqueness for electrode models for electric current computed tomography *SIAM J. Appl. Math.* **52** 1023–40

[33] Stuart A M 2010 Inverse problems: a Bayesian perspective *Acta Numer.* **19** 451–559

[34] Šidák Z 1968 On multivariate normal probabilities of rectangles: their dependence on correlations *Ann. Math. Stat.* **39** 1425–34