# MAP estimators and their consistency in Bayesian nonparametric inverse problems

**M Dashti**[1], **K J H Law**[2], **A M Stuart**[3] **and J Voss**[4]

[1] Department of Mathematics, University of Sussex, Brighton BN1 9QH, UK
[2] Division of Computer, Electrical and Mathematical Sciences and Engineering (CEMSE),
4700 King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900,
Kingdom of Saudi Arabia
[3] Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK
[4] School of Mathematics, University of Leeds, Leeds LS2 9JT, UK

E-mail: M.Dashti@sussex.ac.uk

## Abstract

We consider the inverse problem of estimating an unknown function $u$ from noisy measurements $y$ of a known, possibly nonlinear, map $\mathcal{G}$ applied to $u$. We adopt a Bayesian approach to the problem and work in a setting where the prior measure is specified as a Gaussian random field $\mu_0$. We work under a natural set of conditions on the likelihood which implies the existence of a well-posed posterior measure, $\mu^y$. Under these conditions, we show that the *maximum a posteriori* (MAP) estimator is well defined as the minimizer of an Onsager–Machlup functional defined on the Cameron–Martin space of the prior; thus, we link a problem in probability with a problem in the calculus of variations. We then consider the case where the observational noise vanishes and establish a form of Bayesian posterior consistency for the MAP estimator. We also prove a similar result for the case where the observation of $\mathcal{G}(u)$ can be repeated as many times as desired with independent identically distributed noise. The theory is illustrated with examples from an inverse problem for the Navier–Stokes equation, motivated by problems arising in weather forecasting, and from the theory of conditioned diffusions, motivated by problems arising in molecular dynamics.

## 1. Introduction

This paper considers questions from Bayesian statistics in an infinite-dimensional setting, for example in function spaces. We assume our state space to be a general separable Banach space $(X, \|\cdot\|_X)$. While in the finite-dimensional setting, the prior and posterior distribution of such statistical problems can typically be described by densities w.r.t. the Lebesgue measure, such a characterization is no longer possible in the infinite-dimensional spaces we consider here: it can be shown that no analogue of the Lebesgue measure exists in infinite-dimensional spaces. One way to work around this technical problem is to replace the Lebesgue measure with a Gaussian measure on $X$, i.e. with a Borel probability measure $\mu_0$ on $X$ such that all

finite-dimensional marginals of $\mu_0$ are (possibly degenerate) normal distributions. Using a fixed, centred (mean-zero) Gaussian measure $\mu_0 = \mathcal{N}(0, \mathcal{C}_0)$ as a reference measure, we then assume that the distribution of interest, $\mu$, has a density with respect to $\mu_0$:

$$\frac{\mathrm{d}\mu}{\mathrm{d}\mu_0}(u) \propto \exp(-\Phi(u)). \tag{1.1}$$

Measures $\mu$ of this form arise naturally in a number of applications, including the theory of conditioned diffusions [18] and the Bayesian approach to inverse problems [33]. In these settings, there are many applications where $\Phi : X \to \mathbb{R}$ is a locally Lipschitz continuous function, and it is in this setting that we work.

Our interest is in defining the concept of 'most likely' functions with respect to the measure $\mu$, and in particular the *maximum a posteriori* (MAP) estimator in the Bayesian context. We will refer to such functions as MAP estimators throughout. We will define the concept precisely and link it to a problem in the calculus of variations, study posterior consistency of the MAP estimator in the Bayesian setting and compute it for a number of illustrative applications.

To motivate the form of MAP estimators considered here, we consider the case where $X = \mathbb{R}^d$ is finite dimensional and the prior $\mu_0$ is Gaussian $\mathcal{N}(0, \mathcal{C}_0)$. This prior has density $\exp(-\frac{1}{2}|\mathcal{C}_0^{-1/2}u|^2)$ with respect to the Lebesgue measure, where $|\cdot|$ denotes the Euclidean norm. The probability density for $\mu$ with respect to the Lebesgue measure, given by (1.1), is maximized at minimizers of

$$I(u) := \Phi(u) + \tfrac{1}{2}\|u\|_E^2, \tag{1.2}$$

where $\|\cdot\|_E = |\mathcal{C}_0^{-1/2}u|$. We would like to derive such a result in the infinite-dimensional setting.

The natural way to talk about MAP estimators in the infinite-dimensional setting is to seek the centre of a small ball with maximal probability and then study the limit of this centre as the radius of the ball shrinks to zero. To this end, let $B^\delta(z) \subset X$ be the open ball of radius $\delta$ centred at $z \in X$. If there is a functional $I$, defined on $E$, which satisfies

$$\lim_{\delta \to 0} \frac{\mu(B^\delta(z_2))}{\mu(B^\delta(z_1))} = \exp(I(z_1) - I(z_2)), \tag{1.3}$$

then $I$ is termed the *Onsager–Machlup* functional [11, 21]. For any fixed $z_1$, the function $z_2$ for which the above limit is maximal is a natural candidate for the MAP estimator of $\mu$ and is clearly given by minimizers of the Onsager–Machlup function. In the finite-dimensional case, it is clear that $I$ given by (1.2) is the Onsager–Machlup functional.

From the theory of infinite-dimensional Gaussian measures [5, 25], it is known that copies of the Gaussian measure $\mu_0$ shifted by $z$ are absolutely continuous w.r.t. $\mu_0$ itself, if and only if $z$ lies in the Cameron–Martin space $(E, \langle \cdot, \cdot \rangle_E, \|\cdot\|_E)$; furthermore, if the shift direction $z$ is in $E$, then the shifted measure $\mu_z$ has density

$$\frac{\mathrm{d}\mu_z}{\mathrm{d}\mu_0} = \exp\left(\langle z, u \rangle_E - \frac{1}{2}\|z\|_E^2\right). \tag{1.4}$$

In the finite-dimensional example, above, the Cameron–Martin norm of the Gaussian measure $\mu_0$ is the norm $\|\cdot\|_E$, and it is easy to verify that (1.4) holds for all $z \in \mathbb{R}^d$. In the infinite-dimensional case, it is important to keep in mind that (1.4) only holds for $z \in E \subsetneq X$. Similarly, relation (1.3) only holds for $z_1, z_2 \in E$. In our application, the Cameron–Martin formula (1.4) is used to bound the probability of the shifted ball $B^\delta(z_2)$ from equation (1.3). (For an exposition of the standard results about small ball probabilities for Gaussian measures, we refer to [5, 25]; see also [24] for related material.) The main technical difficulty that is encountered stems from the fact that the Cameron–Martin space $E$, while being dense in $X$, has measure zero with respect to $\mu_0$. An example where this problem can be explicitly seen is

the case where $\mu_0$ is the Wiener measure on $L^2$; in this example, $E$ corresponds to a subset of the Sobolev space $H^1$, which has indeed measure zero w.r.t. the Wiener measure.

Our theoretical results assert that despite these technical complications, the situation from the finite-dimensional example, above, carries over to the infinite-dimensional case essentially without change. In theorem 3.2, we show that the Onsager–Machlup functional in the infinite-dimensional setting still has the form (1.2), where $\| \cdot \|_E$ is now the Cameron–Martin norm associated with $\mu$ (using $\|z\|_E = \infty$ for $z \in X \setminus E$), and in corollary 3.10, we show that the MAP estimators for $\mu$ lie in the Cameron–Martin space $E$ and coincide with the minimizers of the Onsager–Machlup functional $I$.

In the second part of the paper, we consider the inverse problem of estimating an unknown function $u$ in a Banach space $X$, from a given observation $y \in \mathbb{R}^J$, where

$$y = G(u) + \zeta; \tag{1.5}$$

here, $G : X \to \mathbb{R}^J$ is a possibly nonlinear operator, and $\zeta$ is a realization of an $\mathbb{R}^J$-valued centred Gaussian random variable with known covariance $\Sigma$. A prior probability measure $\mu_0(\mathrm{d}u)$ is put on $u$, and the distribution of $y|u$ is given by (1.5), with $\zeta$ assumed to be independent of $u$. Under appropriate conditions on $\mu_0$ and $G$, the Bayes theorem is interpreted as giving the following formula for the Radon–Nikodym derivative of the posterior distribution $\mu^y$ on $u|y$ with respect to $\mu_0$:

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(u) \propto \exp(-\Phi(u; y)), \tag{1.6}$$

where

$$\Phi(u; y) = \tfrac{1}{2}|\Sigma^{-\frac{1}{2}}(y - G(u))|^2. \tag{1.7}$$

The derivation of the Bayes formula (1.6) for problems with finite-dimensional data, and $\zeta$ in this form, is discussed in [7]. Clearly, then, Bayesian inverse problems with Gaussian priors fall into the class of problems studied in this paper, for potentials $\Phi$ given by (1.7) which depend on the observed data $y$. When the probability measure $\mu$ arises from the Bayesian formulation of inverse problems, it is natural to ask whether the MAP estimator is close to the truth underlying the data, in either the small noise or large sample size limits. This is a form of Bayesian posterior consistency, here defined in terms of the MAP estimator only. We will study this question for finite observations of a nonlinear forward model, subject to Gaussian additive noise.

The paper is organized as follows:

- in section 2, we detail our assumptions on $\Phi$ and $\mu_0$;
- in section 3, we give conditions for the existence of an Onsager–Machlup functional $I$ and show that the MAP estimator is well defined as the minimizer of this functional;
- in section 4, we study the problem of Bayesian posterior consistency by studying the limits of Onsager–Machlup minimizers in the small noise and large sample size limits;
- in section 5, we study applications arising from data assimilation for the Navier–Stokes equation, as a model for what is done in weather prediction;
- in section 6, we study applications arising in the theory of conditioned diffusions.

We conclude the introduction with a brief literature review. We first note that MAP estimators are widely used in practice in the infinite-dimensional context [22, 30]. We also note that the functional $I$ in (1.2) resembles a Tikhonov–Phillips regularization of the minimization problem for $\Phi$ [12], with the Cameron–Martin norm of the prior determining the regularization. In the theory of classical non-statistical inversion, formulation via the Tikhonov–Phillips regularization leads to an infinite-dimensional optimization problem and

has led to deeper understanding and improved algorithms. Our aim is to achieve the same in a probabilistic context. One way of defining a MAP estimator for $\mu$ given by (1.1) is to consider the limit of parametric MAP estimators: first discretize the function space using $n$ parameters, and then apply the finite-dimensional argument above to identify an Onsager–Machlup functional on $\mathbb{R}^n$. Passing to the limit $n \to \infty$ in the functional provides a candidate for the limiting Onsager–Machlup functional. This approach is taken in [27, 28, 32] for problems arising in conditioned diffusions. Unfortunately, however, it does not necessarily lead to the correct identification of the Onsager–Machlup functional as defined by (1.3). The reason for this is that the space on which the Onsager–Machlup functional is defined is smoother than the space on which small ball probabilities are defined. Small ball probabilities are needed to properly define the Onsager–Machlup functional in the infinite-dimensional limit. This means that discretization and use of standard numerical analysis limit theorems can, if incorrectly applied, use more regularity than is admissible in identifying the limiting Onsager–Machlup functional. We study the problem directly in the infinite-dimensional setting, without using discretization, leading, we believe, to greater clarity. Adopting the infinite-dimensional perspective for MAP estimation has been widely studied for diffusion processes [9] and related stochastic PDEs [34]; see [35] for an overview. Our general setting is similar to that used to study the specific applications arising in [9, 34, 35]. By working with small ball properties of Gaussian measures, and assuming that $\Phi$ has natural continuity properties, we are able to derive results in considerable generality. There is a recent related definition of MAP estimators in [19], with application to density estimation in [16]. However, whilst the goal of minimizing $I$ is also identified in [19], the proof in that paper is only valid in finite dimensions since it implicitly assumes that the Cameron–Martin norm is $\mu_0$ -almost surely (a.s.) finite. In our specific application to fluid mechanics, our analysis demonstrates that widely used *variational methods* [2] may be interpreted as MAP estimators for an appropriate Bayesian inverse problem and, in particular, that this interpretation, which is understood in the atmospheric sciences community in the finite-dimensional context, is well defined in the limit of infinite spatial resolution.

Posterior consistency in Bayesian nonparametric statistics has a long history [15]. The study of posterior consistency for the Bayesian approach to inverse problems is starting to receive considerable attention. The papers [1, 23] are devoted to obtaining rates of convergence for linear inverse problems with conjugate Gaussian priors, whilst the papers [4, 29] study non-conjugate priors for linear inverse problems. Our analysis of posterior consistency concerns nonlinear problems, and finite data sets, so that multiple solutions are possible. We prove an appropriate weak form of posterior consistency, without rates, building on ideas appearing in [3].

Our form of posterior consistency is weaker than the general form of Bayesian posterior consistency since it does not concern fluctuations in the posterior, simply a point (MAP) estimator. However, we note that for linear Gaussian problems there are examples where the conditions which ensure convergence of the posterior mean (which coincides with the MAP estimator in the linear Gaussian case) also ensure posterior contraction of the entire measure [1, 23].

## 2. Set-up

Throughout this paper, we assume that $(X, \|\cdot\|_X)$ is a separable Banach space and that $\mu_0$ is a centred Gaussian (probability) measure on $X$ with Cameron–Martin space $(E, \langle\cdot,\cdot\rangle_E, \|\cdot\|_E)$. The measure $\mu$ of interest is given by (1.1) and we make the following assumptions concerning the *potential* $\Phi$.

**Assumption 2.1.** *The function* $\Phi : X \to \mathbb{R}$ *satisfies the following conditions.*

*(i) For every* $\varepsilon > 0$, *there is an* $M \in \mathbb{R}$, *such that for all* $u \in X$,

$$\Phi(u) \geqslant M - \varepsilon \|u\|_X^2.$$

*(ii)* $\Phi$ *is locally bounded from above, i.e. for every* $r > 0$, *there exists* $K = K(r) > 0$ *such that, for all* $u \in X$ *with* $\|u\|_X < r$, *we have*

$$\Phi(u) \leqslant K.$$

*(iii)* $\Phi$ *is locally Lipschitz continuous, i.e. for every* $r > 0$, *there exists* $L = L(r) > 0$, *such that for all* $u_1, u_2 \in X$ *with* $\|u_1\|_X, \|u_2\|_X < r$, *we have*

$$|\Phi(u_1) - \Phi(u_2)| \leqslant L \|u_1 - u_2\|_X.$$

Assumption 2.1(i) ensures that expression (1.1) for the measure $\mu$ is indeed normalizable to give a probability measure; the specific form of the lower bound is designed to ensure that application of the Fernique theorem (see [5] or [25]) proves that the required normalization constant is finite. Assumption 2.1(ii) enables us to obtain explicit bounds from below on small ball probabilities and assumption 2.1(iii) allows us to use continuity to control the Onsager–Machlup functional. Numerous examples satisfying these conditions are given in [18, 33]. Finally, we define a function $I : X \to \mathbb{R}$ by

$$I(u) = \begin{cases} \Phi(u) + \frac{1}{2} \|u\|_E^2 & \text{if } u \in E, \text{ and} \\ +\infty & \text{else.} \end{cases} \tag{2.1}$$

We will see in section 3 that $I$ is the Onsager–Machlup functional.

**Remark 2.2.** We close with a brief remark concerning the definition of the Onsager–Machlup function in the case of a non-centred reference measure $\mu_0 = \mathcal{N}(m, \mathcal{C}_0)$. Shifting coordinates by $m$, it is possible to apply the theory based on the centred Gaussian measure $\mu_0$, and then undo the coordinate change. The relevant Onsager–Machlup functional can then be shown to be

$$I(u) = \begin{cases} \Phi(u) + \frac{1}{2} \|u - m\|_E^2 & \text{if } u - m \in E, \text{ and} \\ +\infty & \text{else.} \end{cases}$$

## 3. MAP estimators and the Onsager–Machlup functional

In this section, we prove two main results. The first, theorem 3.2, establishes that $I$ given by (1.2) is indeed the Onsager–Machlup functional for the measure $\mu$ given by (1.1). Then theorem 3.5 and corollary 3.10 show that the MAP estimators, defined precisely in definition 3.1, are characterized by the minimizers of the Onsager–Machlup functional.

For $z \in X$, let $B^\delta(z) \subset X$ be the open ball centred at $z \in X$ with radius $\delta$ in $X$. Let

$$J^\delta(z) = \mu\big(B^\delta(z)\big)$$

be the mass of the ball $B^\delta(z)$. We first define the MAP estimator for $\mu$ as follows.

**Definition 3.1.** *Let*

$$z^\delta = \underset{z \in X}{\mathrm{argmax}}\, J^\delta(z).$$

*Any point* $\tilde{z} \in X$ *satisfying* $\lim_{\delta \to 0}(J^\delta(\tilde{z})/J^\delta(z^\delta)) = 1$ *is a MAP estimator for the measure* $\mu$ *given by (1.1).*

We show later on (theorem 3.5) that a strongly convergent subsequence of $\{z^\delta\}_{\delta>0}$ exists, and its limit, that we prove to be in $E$, is a MAP estimator and also minimizes the Onsager–Machlup functional $I$. Corollary 3.10 then shows that any MAP estimator $\tilde{z}$ as given in definition 3.1 lives in $E$ as well, and minimizers of $I$ characterize all MAP estimators of $\mu$.

One special case where it is easy to see that the MAP estimator is unique is the case where $\Phi$ is linear, but we note that, in general, the MAP estimator cannot be expected to be unique. To achieve uniqueness, stronger conditions on $\Phi$ would be required.

We first need to show that $I$ is the Onsager–Machlup functional for our problem.

**Theorem 3.2.** *Let assumption 2.1 hold. Then the function $I$ defined by (2.1) is the Onsager–Machlup functional for $\mu$, i.e. for any $z_1, z_2 \in E$, we have*

$$\lim_{\delta \to 0} \frac{J^\delta(z_1)}{J^\delta(z_2)} = \exp(I(z_2) - I(z_1)).$$

**Proof.** Note that $J^\delta(z)$ is finite and positive for any $z \in E$ by assumptions 2.1(i) and (ii) together with the Fernique theorem and the positive mass of all balls in $X$, centred at points in $E$, under Gaussian measure [5]. The key estimate in the proof is the following consequence of proposition 3 in section 18 of [25]:

$$\lim_{\delta \to 0} \frac{\mu_0(B^\delta(z_1))}{\mu_0(B^\delta(z_2))} = \exp\left(\frac{1}{2}\|z_2\|_E^2 - \frac{1}{2}\|z_1\|_E^2\right). \tag{3.1}$$

This is the key estimate in the proof since it transfers questions about probability, naturally asked on the space $X$ of full measure under $\mu_0$, into statements concerning the Cameron–Martin norm of $\mu_0$, which is almost surely infinite under $\mu_0$.

We have

$$\frac{J^\delta(z_1)}{J^\delta(z_2)} = \frac{\int_{B^\delta(z_1)} \exp(-\Phi(u))\mu_0(du)}{\int_{B^\delta(z_2)} \exp(-\Phi(v))\mu_0(dv)}$$

$$= \frac{\int_{B^\delta(z_1)} \exp(-\Phi(u) + \Phi(z_1)) \exp(-\Phi(z_1))\mu_0(du)}{\int_{B^\delta(z_2)} \exp(-\Phi(v) + \Phi(z_2)) \exp(-\Phi(z_2))\mu_0(dv)}.$$

By assumption 2.1(iii), for any $u, v \in X$,

$$-L\|u - v\|_X \leqslant \Phi(u) - \Phi(v) \leqslant L\|u - v\|_X,$$

where $L = L(r)$ with $r > \max\{\|u\|_X, \|v\|_X\}$. Therefore, setting $L_1 = L(\|z_1\|_X + \delta)$ and $L_2 = L(\|z_2\|_X + \delta)$, we can write

$$\frac{J^\delta(z_1)}{J^\delta(z_2)} \leqslant e^{\delta(L_1+L_2)} \frac{\int_{B^\delta(z_1)} \exp(-\Phi(z_1))\mu_0(du)}{\int_{B^\delta(z_2)} \exp(-\Phi(z_2))\mu_0(dv)}$$

$$= e^{\delta(L_1+L_2)} e^{-\Phi(z_1)+\Phi(z_2)} \frac{\int_{B^\delta(z_1)} \mu_0(du)}{\int_{B^\delta(z_2)} \mu_0(dv)}.$$

Now, by (3.1), we have

$$\frac{J^\delta(z_1)}{J^\delta(z_2)} \leqslant r_1(\delta) \, e^{\delta(L_2+L_1)} \, e^{-I(z_1)+I(z_2)}$$

with $r_1(\delta) \to 1$ as $\delta \to 0$. Thus,

$$\limsup_{\delta \to 0} \frac{J^\delta(z_1)}{J^\delta(z_2)} \leqslant e^{-I(z_1)+I(z_2)}. \tag{3.2}$$

Similarly, we obtain

$$\frac{J^\delta(z_1)}{J^\delta(z_2)} \geqslant \frac{1}{r_2(\delta)} \, e^{-\delta(L_2+L_1)} \, e^{-I(z_1)+I(z_2)}$$

with $r_2(\delta) \to 1$ as $\delta \to 0$ and deduce that

$$\liminf_{\delta \to 0} \frac{J^\delta(z_1)}{J^\delta(z_2)} \geqslant e^{-I(z_1)+I(z_2)}. \tag{3.3}$$

Inequalities (3.2) and (3.3) give the desired result. $\qquad\square$

We note that similar methods of analysis show the following.

**Corollary 3.3.** *Let the assumptions of theorem 3.2 hold. Then for any $z \in E$,*

$$\lim_{\delta \to 0} \frac{J^\delta(z)}{\int_{B^\delta(0)} \mu_0(\mathrm{d}u)} = \frac{1}{Z} \, e^{-I(z)},$$

*where $Z = \int_X \exp(-\Phi(u)) \, \mu_0(\mathrm{d}u)$.*

**Proof.** Noting that we consider $\mu$ to be a probability measure and hence,

$$\frac{J^\delta(z)}{\int_{B^\delta(0)} \mu_0(\mathrm{d}u)} = \frac{\frac{1}{Z} \int_{B^\delta(z)} \exp(-\Phi(u))\mu_0(\mathrm{d}u)}{\int_{B^\delta(0)} \mu_0(\mathrm{d}u)},$$

with $Z = \int_X \exp(-\Phi(u)) \, \mu_0(\mathrm{d}u)$, arguing along the lines of the proof of the above theorem gives

$$\frac{1}{Z} \frac{1}{r(\delta)} \, e^{-\delta\hat{L}} \, e^{-I(z)} \leqslant \frac{J^\delta(z)}{\int_{B^\delta(0)} \mu_0(\mathrm{d}u)} \leqslant \frac{1}{Z} r(\delta) \, e^{\delta\hat{L}} \, e^{-I(z)}$$

with $\hat{L} = L(\|z\|_X + \delta)$ (where $L(\cdot)$ is as in definition 2.1) and $r(\delta) \to 1$ as $\delta \to 0$. The result then follows by taking lim sup and lim inf as $\delta \to 0$. $\qquad\square$

**Proposition 3.4.** *Suppose assumptions 2.1 hold. Then the minimum of $I : E \to \mathbb{R}$ is attained for some element $z^* \in E$.*

**Proof.** The existence of a minimizer of $I$ in $E$, under the given assumptions, is proved as theorem 5.4 in [33] (and as theorem 2.7 in [7] in the case where $\Phi$ is non-negative). $\qquad\square$

The rest of this section is devoted to a proof of the result that MAP estimators can be characterized as minimizers of the Onsager–Machlup functional $I$ (theorem 3.5 and corollary 3.10).

**Theorem 3.5.** *Suppose that assumptions 2.1(ii) and (iii) hold. Assume also that there exists an $M \in \mathbb{R}$ such that $\Phi(u) \geqslant M$ for any $u \in X$.*

*(i) Let $z^\delta = argmax_{z \in X} J^\delta(z)$. There is a $\bar{z} \in E$ and a subsequence of $\{z^\delta\}_{\delta>0}$ which converges to $\bar{z}$ strongly in $X$.*
*(ii) The limit $\bar{z}$ is a MAP estimator and a minimizer of $I$.*

The proof of this theorem is based on several lemmas. We state and prove these lemmas first and defer the proof of theorem 3.5 to the end of the section where we also state and prove a corollary characterizing the MAP estimators as minimizers of the Onsager–Machlup functional.

**Lemma 3.6.** *Let $\delta > 0$. For any centred Gaussian measure $\mu_0$ on a separable Banach space X, we have*

$$\frac{J_0^\delta(z)}{J_0^\delta(0)} \leqslant c\, \mathrm{e}^{-\frac{a_1}{2}(\|z\|_X - \delta)^2},$$

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

**Proof.** We first show that this is true for a centred Gaussian measure on $\mathbb{R}^n$ with the covariance matrix $C = \mathrm{diag}[\lambda_1, \ldots, \lambda_n]$ in basis $\{e_1, \ldots, e_n\}$, where $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$. Let $a_j = 1/\lambda_j$, and $|z|^2 = z_1^2 + \cdots + z_n^2$. Define

$$J_{0,n}^\delta(z) := \int_{B^\delta(z)} \mathrm{e}^{-\frac{1}{2}(a_1 x_1^2 + \cdots + a_n x_n^2)}\, \mathrm{d}x, \quad \text{for any } z \in \mathbb{R}^n, \tag{3.4}$$

and with $B^\delta(z)$ the ball of radius $\delta$ and centre $z$ in $\mathbb{R}^n$. We have

$$
\begin{aligned}
\frac{J_{0,n}^\delta(z)}{J_{0,n}^\delta(0)} &= \frac{\int_{B^\delta(z)} \mathrm{e}^{-\frac{1}{2}(a_1 x_1^2 + \cdots + a_n x_n^2)}\, \mathrm{d}x}{\int_{B^\delta(0)} \mathrm{e}^{-\frac{1}{2}(a_1 x_1^2 + \cdots + a_n x_n^2)}\, \mathrm{d}x} \\
&< \frac{\mathrm{e}^{-\frac{1}{2}(a_1-\varepsilon)(|z|-\delta)^2}}{\mathrm{e}^{-\frac{1}{2}(a_1-\varepsilon)\delta^2}} \frac{\int_{B^\delta(z)} \mathrm{e}^{-\frac{1}{2}(\varepsilon x_1^2 + (a_2-a_1+\varepsilon)x_2^2 + \cdots + (a_n-a_1+\varepsilon)x_n^2)}\, \mathrm{d}x}{\int_{B^\delta(0)} \mathrm{e}^{-\frac{1}{2}(\varepsilon x_1^2 + (a_2-a_1+\varepsilon)x_2^2 + \cdots + (a_n-a_1+\varepsilon)x_n^2)}\, \mathrm{d}x} \\
&< c\, \mathrm{e}^{-\frac{1}{2}(a_1-\varepsilon)(|z|-\delta)^2} \frac{\int_{B^\delta(z)} \hat{\mu}_0(\mathrm{d}x)}{\int_{B^\delta(0)} \hat{\mu}_0(\mathrm{d}x)},
\end{aligned}
$$

for any $\varepsilon < a_1$ and where $\hat{\mu}_0$ is a centred Gaussian measure on $\mathbb{R}^n$ with the covariance matrix $\mathrm{diag}[1/\varepsilon, 1/(a_2 - a_1 + \varepsilon), \ldots, 1/(a_n - a_1 + \varepsilon)]$ (noting that $a_n \geqslant a_{n-1} \geqslant \cdots \geqslant a_1$). By Anderson's inequality for the infinite-dimensional spaces (see theorem 2.8.10 of [5]), we have $\hat{\mu}_0(B(z, \delta)) \leqslant \hat{\mu}_0(B(0, \delta))$ and therefore

$$\frac{J_{0,n}^\delta(z)}{J_{0,n}^\delta(0)} < c\, \mathrm{e}^{-\frac{1}{2}(a_1-\varepsilon)(|z|-\delta)^2},$$

and since $\varepsilon$ is arbitrarily small, the result follows for the finite-dimensional case.

To show the result for an infinite-dimensional separable Banach space $X$, we first note that $\{e_j\}_{j=1}^\infty$, the orthogonal basis in the Cameron–Martin space of $X$ for $\mu_0$, separates the points in $X$; therefore, $T : u \rightarrow \{e_j(u)\}_{j=1}^\infty$ is an injective map from $X$ to $\mathbb{R}^\infty$. Let $u_j = e_j(u)$ and

$$P_n u = (u_1, u_2, \ldots, u_n, 0, 0, \ldots).$$

Then, since $\mu_0$ is a Radon measure, for the balls $B(0, \delta)$ and $B(z, \delta)$, for any $\varepsilon_0 > 0$, there exists large enough $N$ such that the cylindrical sets $A_0 = P_n^{-1}(P_n(B^\delta(0)))$ and $A_z = P_n^{-1}(P_n(B^\delta(z)))$ satisfy $\mu_0(B^\delta(0) \triangle A_0) < \varepsilon_0$ and $\mu_0(B^\delta(z) \triangle A_z) < \varepsilon_0$ for $n > N$ [5], where $\triangle$ denotes the symmetric difference. Let $z_j = (z, e_j)$ and $z^n = (z_1, z_2, \ldots, z_n, 0, \ldots)$, and for $0 < \varepsilon_1 < \delta/2$, $n > N$ large enough so that $\|z - z^n\|_X \leqslant \varepsilon_1$. With $\alpha = c\, \mathrm{e}^{-\frac{a_1}{2}(\|z\|_X - \varepsilon_1 - \delta)^2}$, we have

$$
\begin{aligned}
J_0^\delta(z) &\leqslant J_{0,n}^\delta(z^n) + \varepsilon_0 \\
&\leqslant \alpha J_{0,n}^\delta(0) + \varepsilon_0 \\
&\leqslant \alpha J_0^\delta(0) + (1 + \alpha)\varepsilon_0.
\end{aligned}
$$

Since $\varepsilon_0$ and $\varepsilon_1$ converge to zero as $n \rightarrow \infty$, the result follows. $\qquad \square$

**Lemma 3.7.** *Suppose that $\bar{z} \notin E$, $\{z^\delta\}_{\delta>0} \subset X$ and $z^\delta$ converges weakly to $z$ in X as $\delta \rightarrow 0$. Then for any $\varepsilon > 0$, there exists $\delta$ small enough such that*

$$\frac{J_0^\delta(z^\delta)}{J_0^\delta(0)} < \varepsilon.$$

**Proof.** Let $\mathcal{C}$ be the covariance operator of $\mu_0$, and $\{e_j\}_{j\in\mathbb{N}}$ the eigenfunctions of $\mathcal{C}$ scaled with respect to the inner product of $E$, the Cameron–Martin space of $\mu_0$, so that $\{e_j\}_{j\in\mathbb{N}}$ forms an orthonormal basis in $E$. Let $\{\lambda_j\}$ be the corresponding eigenvalues and $a_j = 1/\lambda_j$. Since $z^\delta$ converges weakly to $\bar{z}$ in $X$ as $\delta \to 0$,

$$e_j(z^\delta) \to e_j(\bar{z}) \quad \text{for any } j \in \mathbb{N}, \tag{3.5}$$

and as $\bar{z} \notin E$ for any $A > 0$, there exist $N$ sufficiently large and $\tilde{\delta} > 0$ sufficiently small, such that

$$\inf_{z \in B^\delta(\bar{z})} \left\{ \sum_{j=1}^{N} a_j x_j^2 \right\} > A^2,$$

where $x_j = e_j(z)$. By (3.5), for $\delta_1 < \tilde{\delta}$ small enough, we have $B^{\delta_1}(z^{\delta_1}) \subset B^{\tilde{\delta}}(\bar{z})$ and therefore

$$\inf_{z \in B^{\delta_1}(z^{\delta_1})} \left\{ \sum_{j=1}^{N} a_j x_j^2 \right\} > A^2. \tag{3.6}$$

Let $T_n : X \to \mathbb{R}^n$ map $z$ to $(e_1(z), \ldots, e_n(z))$, and consider $J_{0,n}^\delta(z)$ to be defined as in (3.4). Having (3.6), and choosing $\delta \leqslant \delta_1$ such that $\mathrm{e}^{-\frac{1}{4}(a_1+\cdots+a_N)\delta^2} > 1/2$, for any $n \geqslant N$, we can write

$$\frac{J_{0,n}^\delta(T_n z^\delta)}{J_{0,n}^\delta(0)} = \frac{\int_{B^\delta(T_n z^\delta)} \mathrm{e}^{-\frac{1}{2}(a_1 x_1^2 + \cdots + a_n x_n^2)} \, \mathrm{d}x}{\int_{B^\delta(0)} \mathrm{e}^{-\frac{1}{2}(a_1 x_1^2 + \cdots + a_n x_n^2)} \, \mathrm{d}x}$$

$$\leqslant \frac{\int_{B^\delta(T_n z^\delta)} \mathrm{e}^{-\frac{1}{4}(a_1 x_1^2 + \cdots + a_N x_N^2)} \, \mathrm{e}^{-\frac{1}{2}(\frac{a_1}{2} x_1^2 + \cdots + \frac{a_N}{2} x_N^2 + a_{N+1} x_{N+1}^2 \cdots + a_n x_n^2)} \, \mathrm{d}x}{\int_{B^\delta(0)} \mathrm{e}^{-\frac{1}{4}(a_1 x_1^2 + \cdots + a_N x_N^2)} \, \mathrm{e}^{-\frac{1}{2}(\frac{a_1}{2} x_1^2 + \cdots + \frac{a_N}{2} x_N^2 + a_{N+1} x_{N+1}^2 \cdots + a_n x_n^2)} \, \mathrm{d}x}$$

$$\leqslant \frac{\mathrm{e}^{-\frac{1}{4}A^2} \int_{B^\delta(T_n z^\delta)} \mathrm{e}^{-\frac{1}{2}(\frac{a_1}{2} x_1^2 + \cdots + \frac{a_N}{2} x_N^2 + a_{N+1} x_{N+1}^2 \cdots + a_n x_n^2)} \, \mathrm{d}x}{\frac{1}{2} \int_{B^\delta(0)} \mathrm{e}^{-\frac{1}{2}(\frac{a_1}{2} x_1^2 + \cdots + \frac{a_N}{2} x_N^2 + a_{N+1} x_{N+1}^2 \cdots + a_n x_n^2)} \, \mathrm{d}x}$$

$$\leqslant 2 \, \mathrm{e}^{-\frac{1}{4}A^2}.$$

As $A > 0$ was arbitrary, the constant in the last line of the above equation can be made arbitrarily small, by making $\delta$ sufficiently small and $n$ sufficiently large. Having this and arguing in a similar way to the final paragraph of the proof of lemma 3.6, the result follows. $\qquad\square$

**Corollary 3.8.** *Suppose that $z \notin E$. Then*

$$\lim_{\delta \to 0} \frac{J_0^\delta(z)}{J_0^\delta(0)} = 0.$$

**Lemma 3.9.** *Consider $\{z^\delta\}_{\delta>0} \subset X$ and suppose that $z^\delta$ converges weakly and not strongly to $0$ in $X$ as $\delta \to 0$. Then for any $\varepsilon > 0$, there exists $\delta$ small enough such that*

$$\frac{J_0^\delta(z^\delta)}{J_0^\delta(0)} < \varepsilon.$$

**Proof.** Since $z^\delta$ converges weakly and not strongly to $0$, we have

$$\liminf_{\delta \to 0} \|z^\delta\|_X > 0,$$

and therefore, for $\delta_1$ small enough, there exists $\alpha > 0$ such that $\|z^\delta\|_X > \alpha$ for any $\delta < \delta_1$. Let $\lambda_j$, $a_j$ and $e_j$, $j \in \mathbb{N}$, be defined as in the proof of lemma 3.7. Since $z^\delta \rightharpoonup 0$ as $\delta \to 0$,

$$e_j(z^\delta) \to 0, \quad \text{for any } j \in \mathbb{N}. \tag{3.7}$$

Also, as for $\mu_0$-almost every $x \in X$, $x = \sum_{j \in \mathbb{N}} e_j(x)\hat{e}_j$ and $\{\hat{e}_j = e_j/\sqrt{\lambda_j}\}$ is an orthonormal basis in $X^*_{\mu_0}$ (closure of $X^*$ in $L^2(\mu_0)$) [5], we have

$$\sum_{j \in \mathbb{N}} (e_j(x))^2 < \infty \quad \text{for } \mu_0\text{-almost every } x \in X. \tag{3.8}$$

Now, for any $A > 0$, let $N$ large enough such that $a_N > A^2$. Then, having (3.7) and (3.8), one can choose $\delta_2 < \delta_1$ small enough and $N_1 > N$ large enough so that for $\delta < \delta_2$ and $n > N_1$,

$$\sum_{j=1}^{N} (e_j(z^\delta))^2 < \frac{C\alpha}{2} \quad \text{and} \quad \sum_{j=N+1}^{n} (e_j(z^\delta))^2 > \frac{C\alpha}{2}.$$

Therefore, letting $J^\delta_{0,n}(z)$ and $T_n$ be defined as in the proof of lemma 3.7, we can write

$$\frac{J^\delta_{0,n}(T_n z^\delta)}{J^\delta_{0,n}(0)} = \frac{\int_{B^\delta(T_n z^\delta)} e^{-\frac{1}{2}(a_1 x_1^2 + \cdots + a_n x_n^2)} \, dx}{\int_{B^\delta(0)} e^{-\frac{1}{2}(a_1 x_1^2 + \cdots + a_n x_n^2)} \, dx}$$

$$\leqslant \frac{\int_{B^\delta(T_n z^\delta)} e^{-\frac{A^2}{2}(x_{N+1}^2 + \cdots + x_n^2)} e^{-\frac{1}{2}(a_1 x_1^2 + \cdots + a_N x_N^2 + (a_{N+1} - A^2) x_{N+1}^2 \cdots + (a_n - A^2) x_n^2)} \, dx}{\int_{B^\delta(0)} e^{-\frac{A^2}{2}(x_{N+1}^2 + \cdots + x_n^2)} e^{-\frac{1}{2}(a_1 x_1^2 + \cdots + a_N x_N^2 + (a_{N+1} - A^2) x_{N+1}^2 \cdots + (a_n - A^2) x_n^2)} \, dx}$$

$$\leqslant \frac{e^{-\frac{1}{2}A^2(\frac{C\alpha}{2} - \delta^2)} \int_{B^\delta(T_n z^\delta)} e^{-\frac{1}{2}(\frac{a_1}{2} x_1^2 + \cdots + \frac{a_N}{2} x_N^2 + a_{N+1} x_{N+1}^2 \cdots + a_n x_n^2)} \, dx}{e^{-\frac{1}{2}A^2 \delta^2} \int_{B^\delta(0)} e^{-\frac{1}{2}(\frac{a_1}{2} x_1^2 + \cdots + \frac{a_N}{2} x_N^2 + a_{N+1} x_{N+1}^2 \cdots + a_n x_n^2)} \, dx}$$

$$\leqslant 2 e^{-\frac{C\alpha}{4} A^2},$$

if $\delta < \delta_2$ is small enough so that $e^{A\delta^2} < 2$. Having this and arguing in a similar way to the final paragraph of the proof of lemma 3.6, the result follows. □

Having these preparations in place, we can give the proof of theorem 3.5.

**Proof of theorem 3.5.** (i) We first show that $\{z^\delta\}$ is bounded in $X$. By assumption 2.1(ii), for any $r > 0$, there exists $K = K(r) > 0$ such that

$$\Phi(u) \leqslant K(r)$$

for any $u$ satisfying $\|u\|_X < r$; thus, $K$ may be assumed to be a non-decreasing function of $r$. This implies that

$$\max_{z \in E} \int_{B^\delta(z)} e^{-\Phi(u)} \mu_0(du) \geqslant \int_{B^\delta(0)} e^{-\Phi(u)} \mu_0(du) \geqslant e^{-K(\delta)} \int_{B^\delta(0)} \mu_0(du).$$

We assume that $\delta \leqslant 1$, and then the inequality above shows that

$$\frac{J^\delta(z^\delta)}{\int_{B^\delta(0)} \mu_0(du)} \geqslant \frac{1}{Z} e^{-K(1)} = \varepsilon_1, \tag{3.9}$$

noting that $\varepsilon_1$ is independent of $\delta$.

We can also write

$$ZJ^\delta(z) = \int_{B^\delta(z)} e^{-\Phi(u)} \mu_0(du)$$

$$\leqslant e^{-M} \int_{B^\delta(z)} \mu_0(du)$$

$$=: e^{-M} J^\delta_0(z),$$

which implies that for any $z \in X$ and $\delta > 0$,

$$J_0^\delta(z) \geqslant Z e^M J^\delta(z). \tag{3.10}$$

Now suppose $\{z^\delta\}$ is not bounded in $X$, so that for any $R > 0$, there exists $\delta_R$ such that $\|z^{\delta_R}\|_X > R$ (with $\delta_R \to 0$ as $R \to \infty$). By (3.10), (3.9) and the definition of $z^{\delta_R}$, we have

$$J_0^{\delta_R}(z^{\delta_R}) \geqslant Z e^M J^{\delta_R}(z^{\delta_R}) \geqslant Z e^M J^{\delta_R}(0) \geqslant e^M e^{-K(1)} J_0^{\delta_R}(0),$$

implying that for any $\delta_R$ and corresponding $z^{\delta_R}$

$$\frac{J_0^{\delta_R}(z^{\delta_R})}{J_0^{\delta_R}(0)} \geqslant c = e^M e^{-K(1)}.$$

This contradicts the result of lemma 3.6 (below) for $\delta_R$ small enough. Hence, there exists $R, \delta_R > 0$ such that

$$\|z^\delta\|_X \leqslant R \quad \text{for any } \delta < \delta_R.$$

Therefore, there exist a $\bar{z} \in X$ and a subsequence of $\{z^\delta\}_{0 < \delta < \delta_R}$ which converges weakly in $X$ to $\bar{z} \in X$ as $\delta \to 0$.

Now, suppose either

(a) there is no strongly convergent subsequence of $\{z^\delta\}$ in $X$, or
(b) if there is one, its limit $\bar{z}$ is not in $E$.

Let $U_E = \{u \in E : \|u\|_E \leqslant 1\}$. Each of the above situations implies that for any positive $A \in \mathbb{R}$, there is a $\delta^\dagger$ such that for any $\delta \leqslant \delta^\dagger$,

$$B^\delta(z^\delta) \cap \left( B^\delta(0) + A U_E \right) = \emptyset. \tag{3.11}$$

We first show that $\bar{z}$ has to be in $E$. By the definition of $z^\delta$, we have (for $\delta < 1$)

$$1 \leqslant \frac{J^\delta(z^\delta)}{J^\delta(0)} \leqslant \frac{e^M}{e^{-K(1)}} \frac{\int_{B^\delta(z^\delta)} \mu_0(\mathrm{d}u)}{\int_{B^\delta(0)} \mu_0(\mathrm{d}u)}. \tag{3.12}$$

Supposing $\bar{z} \notin E$, in lemma 3.7, we show that for any $\varepsilon > 0$, there exists $\delta$ small enough such that

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

Hence, choosing $A$ in (3.11) such that $e^{-A^2/2} < \frac{1}{2} e^{K(1)} e^{-M}$, and setting $\varepsilon = e^{-A^2/2}$, from (3.12), we obtain $1 \leqslant J^\delta(z^\delta)/J^\delta(0) < 1$, which is a contradiction. We therefore have $\bar{z} \in E$.

Now, knowing that $\bar{z} \in E$, we can show that $z^\delta$ converges strongly in $X$. Suppose it does not. Then for $z^\delta - \bar{z}$, the hypotheses of lemma 3.9 are satisfied. Again choosing $A$ in (3.11) such that $e^{-A^2/2} < \frac{1}{2} e^{K(1)} e^{-M}$, and setting $\varepsilon = e^{-A^2/2}$, from lemma 3.9 and (3.12), we obtain $1 \leqslant J^\delta(z^\delta)/J^\delta(0) < 1$, which is a contradiction. Hence, there is a subsequence of $\{z^\delta\}$ converging strongly in $X$ to $\bar{z} \in E$.

(ii) Let $z^* = \arg \min I(z) \in E$; the existence is assured by theorem 3.2. By assumption 2.1(iii), we have

$$\frac{J^\delta(z^\delta)}{J^\delta(\bar{z})} \leqslant e^{-\Phi(z^\delta) + \Phi(\bar{z})} e^{(L_1 + L_2)\delta} \frac{\int_{B^\delta(z^\delta)} \mu_0(\mathrm{d}u)}{\int_{B^\delta(\bar{z})} \mu_0(\mathrm{d}u)},$$

with $L_1 = L(\|z^\delta\|_X + \delta)$ and $L_2 = L(\|\bar{z}\|_X + \delta)$. Therefore, since $\Phi$ is continuous on $X$ and $z^\delta \to \bar{z}$ in X,

$$\limsup_{\delta \to 0} \frac{J^\delta(z^\delta)}{J^\delta(\bar{z})} \leqslant \limsup_{\delta \to 0} \frac{\int_{B^\delta(z^\delta)} \mu_0(\mathrm{d}u)}{\int_{B^\delta(\bar{z})} \mu_0(\mathrm{d}u)}.$$

11

Suppose $\{z^\delta\}$ is not bounded in $E$, or if it is, it only converges weakly (and not strongly) in $E$. Then, $\|\bar{z}\|_E < \liminf_{\delta\to 0}\|z^\delta\|_E$, and hence for small enough $\delta$, $\|\bar{z}\|_E < \|z^\delta\|_E$. Therefore, for the centered Gaussian measure $\mu_0$, since $\|z^\delta - \bar{z}\|_X \to 0$, we have

$$\limsup_{\delta\to 0} \frac{\int_{B^\delta(z^\delta)}\mu_0(\mathrm{d}u)}{\int_{B^\delta(\bar{z})}\mu_0(\mathrm{d}u)} \leqslant 1.$$

Since by the definition of $z^\delta$, $J^\delta(z^\delta) \geqslant J^\delta(\bar{z})$ and hence

$$\liminf_{\delta\to 0}\big(J^\delta(z^\delta)/J^\delta(\bar{z})\big) \geqslant 1,$$

this implies that

$$\lim_{\delta\to 0}\frac{J^\delta(z^\delta)}{J^\delta(\bar{z})} = 1. \tag{3.13}$$

In the case where $\{z^\delta\}$ converges strongly to $\bar{z}$ in $E$, by the Cameron–Martin theorem we have

$$\frac{\int_{B^\delta(z^\delta)}\mu_0(\mathrm{d}u)}{\int_{B^\delta(\bar{z})}\mu_0(\mathrm{d}u)} = \frac{\mathrm{e}^{-\frac{1}{2}\|z^\delta\|_E^2}\int_{B^\delta(0)}\mathrm{e}^{\langle z^\delta,u\rangle_E}\mu_0(\mathrm{d}u)}{\mathrm{e}^{-\frac{1}{2}\|\bar{z}\|_E^2}\int_{B^\delta(0)}\mathrm{e}^{\langle\bar{z},u\rangle_E}\mu_0(\mathrm{d}u)},$$

and then by an argument very similar to the proof of theorem 18.3 of [25], one can show that

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

and (3.13) follows again in a similar way. Therefore, $\bar{z}$ is a MAP estimator of measure $\mu$.

It remains to show that $\bar{z}$ is a minimizer of $I$. Suppose $\bar{z}$ is not a minimizer of $I$ so that $I(\bar{z}) - I(z^*) > 0$. Let $\delta_1$ be small enough so that in the equation before (3.2), $1 < r_1(\delta) < \mathrm{e}^{I(\bar{z})-I(z^*)}$ for any $\delta < \delta_1$ and therefore,

$$\frac{J^\delta(\bar{z})}{J^\delta(z^*)} \leqslant r_1(\delta)\,\mathrm{e}^{-I(\bar{z})+I(z^*)} < 1. \tag{3.14}$$

Let $\alpha = r_1(\delta)\,\mathrm{e}^{-I(\bar{z})+I(z^*)}$. We have

$$\frac{J^\delta(z^\delta)}{J^\delta(z^*)} = \frac{J^\delta(z^\delta)}{J^\delta(\bar{z})}\frac{J^\delta(\bar{z})}{J^\delta(z^*)},$$

and this by (3.14) and (3.13) implies that

$$\limsup_{\delta\to 0}\frac{J^\delta(z^\delta)}{J^\delta(z^*)} \leqslant \alpha \limsup_{\delta\to 0}\frac{J^\delta(z^\delta)}{J^\delta(\bar{z})} < 1,$$

which is a contradiction, since by the definition of $z^\delta$, $J^\delta(z^\delta) \geqslant J^\delta(z^*)$ for any $\delta > 0$.　　□

**Corollary 3.10.** *Under the conditions of theorem 3.5, we have the following.*

 (i) *Any MAP estimator, given by definition 3.1, minimizes the Onsager–Machlup functional I.*
 (ii) *Any $z^* \in E$ which minimizes the Onsager–Machlup functional I is a MAP estimator for measure $\mu$ given by (1.1).*

**Proof.**

 (i) Let $\tilde{z}$ be a MAP estimator. By theorem 3.5, we know that $\{z^\delta\}$ has a subsequence which strongly converges in $X$ to $\bar{z}$. Let $\{z^\alpha\}$ be the said subsequence. Then by (3.13) one can show that

$$\lim_{\delta\to 0}\frac{J^\delta(z^\delta)}{J^\delta(\bar{z})} = \lim_{\alpha\to 0}\frac{J^\alpha(z^\alpha)}{J^\alpha(\bar{z})} = 1.$$

By the above equation and since $\tilde{z}$ is a MAP estimator, we can write

$$\lim_{\delta \to 0} \frac{J^\delta(\tilde{z})}{J^\delta(\bar{z})} = \lim_{\delta \to 0} \frac{J^\delta(z^\delta)}{J^\delta(\bar{z})} \lim_{\delta \to 0} \frac{J^\delta(\tilde{z})}{J^\delta(z^\delta)} = 1.$$

Then corollary 3.8 implies that $\tilde{z} \in E$, and supposing that $\tilde{z}$ is not a minimizer of $I$ would result in a contradiction using an argument similar to that in the last paragraph of the proof of the above theorem.

(ii) Note that the assumptions of theorem 3.5 imply those of theorem 3.2. Since $\bar{z}$ is a minimizer of $I$ as well, by theorem 3.2 we have

$$\lim_{\delta \to 0} \frac{J^\delta(\bar{z})}{J^\delta(z^*)} = 1.$$

Then we can write

$$\lim_{\delta \to 0} \frac{J^\delta(z^*)}{J^\delta(z^\delta)} = \lim_{\delta \to 0} \frac{J^\delta(\bar{z})}{J^\delta(z^\delta)} = \lim_{\delta \to 0} \frac{J^\delta(z^*)}{J^\delta(\bar{z})} = 1.$$

The result follows by definition 3.1.                                    □

## 4. Bayesian inversion and posterior consistency

The structure (1.1), where $\mu_0$ is Gaussian, arises in the application of the Bayesian methodology to the solution of inverse problems. In that context, it is interesting to study *posterior consistency*: the idea that the posterior concentrates near the truth which gave rise to the data, in the small noise or large sample size limits; these two limits are intimately related and indeed there are theorems that quantify this connection for certain linear inverse problems [6].

In this section, we describe the Bayesian approach to nonlinear inverse problems, as outlined in the introduction. We assume that the data are found from the application of $G$ to the truth $u^\dagger$ with additional noise:

$$y = G(u^\dagger) + \zeta.$$

The posterior distribution $\mu^y$ is then of the form (1.6) and in this case, it is convenient to extend the Onsager–Machlup functional $I$ to a mapping from $X \times \mathbb{R}^J$ to $\mathbb{R}$, defined as

$$I(u; y) = \Phi(u; y) + \tfrac{1}{2}\|u\|_E^2.$$

We study the posterior consistency of MAP estimators in both the small noise and large sample size limits. The corresponding results are presented in theorems 4.4 and 4.1, respectively. Specifically, we characterize the sense in which the MAP estimators concentrate on the truth underlying the data in the small noise and large sample size limits.

### 4.1. Large sample size limit

Let us denote the exact solution by $u^\dagger$ and suppose that as data we have the following $n$ random vectors:

$$y_j = \mathcal{G}(u^\dagger) + \eta_j, \quad j = 1, \ldots, n,$$

with $y_j \in \mathbb{R}^K$ and $\eta_j \sim \mathcal{N}(0, \mathcal{C}_1)$ independent identically distributed random variables. Thus, in the general setting, we have $J = nK$, $G(\cdot) = (\mathcal{G}(\cdot), \ldots, \mathcal{G}(\cdot))$ and $\Sigma$ a block diagonal matrix with $\mathcal{C}_1$ in each block. We have $n$ independent observations, each polluted by $\mathcal{O}(1)$

noise, and we study the limit $n \to \infty$. Corresponding to this set of data and given the prior measure $\mu_0 \sim \mathcal{N}(0, \mathcal{C}_0)$, we have the following formula for the posterior measure on $u$:

$$\frac{d\mu^{y_1,\ldots,y_n}}{d\mu_0}(u) \propto \exp\left(-\frac{1}{2}\sum_{j=1}^{n}|y_j - \mathcal{G}(u)|_{\mathcal{C}_1}^2\right).$$

Here, and in the following, we use the notations $\langle \cdot, \cdot \rangle_{\mathcal{C}_1} = \langle \mathcal{C}_1^{-1/2}\cdot, \mathcal{C}_1^{-1/2}\cdot \rangle$ and $|\cdot|_{\mathcal{C}_1}^2 = \langle \cdot, \cdot \rangle_{\mathcal{C}_1}$. By corollary 3.10, MAP estimators for this problem are minimizers of

$$I_n := \|u\|_E^2 + \sum_{j=1}^{n}|y_j - \mathcal{G}(u)|_{\mathcal{C}_1}^2. \tag{4.1}$$

Our interest is in studying the properties of the limits of minimizers $u_n$ of $I_n$, namely the MAP estimators corresponding to the preceding family of posterior measures. We have the following theorem concerning the behaviour of $u_n$ when $n \to \infty$.

**Theorem 4.1.** *Assume that $\mathcal{G} : X \to \mathbb{R}^K$ is Lipschitz on bounded sets and $u^\dagger \in E$. For every $n \in \mathbb{N}$, let $u_n \in E$ be a minimizer of $I_n$ given by (4.1). Then there exist a $u^* \in E$ and a subsequence of $\{u_n\}_{n \in \mathbb{N}}$ that converges weakly to $u^*$ in $E$, almost surely. For any such $u^*$, we have $\mathcal{G}(u^*) = \mathcal{G}(u^\dagger)$.*

We describe some preliminary calculations useful in the proof of this theorem, then give lemma 4.2, also useful in the proof, and finally give the proof itself.

We first observe that, under the assumption that $\mathcal{G}$ is Lipschitz on bounded sets, assumptions 2.1 hold for $\Phi$. We note that

$$I_n = \|u\|_E^2 + \sum_{j=1}^{n}|y_j - \mathcal{G}(u)|_{\mathcal{C}_1}^2$$

$$= \|u\|_E^2 + n|\mathcal{G}(u^\dagger) - \mathcal{G}(u)|_{\mathcal{C}_1}^2 + 2\sum_{j=1}^{n}\langle \mathcal{G}(u^\dagger) - \mathcal{G}(u), \mathcal{C}_1^{-1}\eta_j\rangle.$$

Hence,

$$\arg\min_u I_n = \arg\min_u \left\{\|u\|_E^2 + n|\mathcal{G}(u^\dagger) - \mathcal{G}(u)|_{\mathcal{C}_1}^2 + 2\sum_{j=1}^{n}\langle \mathcal{G}(u^\dagger) - \mathcal{G}(u), \mathcal{C}_1^{-1}\eta_j\rangle\right\}.$$

Define

$$J_n(u) = |\mathcal{G}(u^\dagger) - \mathcal{G}(u)|_{\mathcal{C}_1}^2 + \frac{1}{n}\|u\|_E^2 + \frac{2}{n}\sum_{j=1}^{n}\langle \mathcal{G}(u^\dagger) - \mathcal{G}(u), \mathcal{C}_1^{-1}\eta_j\rangle.$$

We have

$$\arg\min_u I_n = \arg\min_u J_n.$$

**Lemma 4.2.** *Assume that $\mathcal{G} : X \to \mathbb{R}^K$ is Lipschitz on bounded sets. Then for fixed $n \in \mathbb{N}$ and almost surely, there exists $u_n \in E$ such that*

$$J_n(u_n) = \inf_{u \in E} J_n(u).$$

**Proof.** We first observe that, under the assumption that $\mathcal{G}$ is Lipschitz on bounded sets and because for a given $n$ and fixed realizations $\eta_1, \ldots, \eta_n$ there exists an $r > 0$ such that $\max\{|y_1|, \ldots, |y_n|\} < r$, assumptions 2.1 hold for $\Phi$. Since $\arg\min_u I_n = \arg\min_u J_n$, the result follows by proposition 3.4. $\square$

We may now prove the posterior consistency theorem. From (4.3) onwards, the proof is an adaptation of the proof of theorem 2 of [3]. We note that the assumptions on limiting behaviour of measurement noise in [3] are stronger: property (9) of [3] is not assumed here for our $J_n$. On the other hand, a frequentist approach is used in [3], while here, since $J_n$ is coming from a Bayesian approach, the norm in the regularization term is stronger (it is related to the Cameron–Martin space of the Gaussian prior). That is why in our case, asking what if $u^\dagger$ is not in $E$ and only in $X$ is relevant and is answered in corollary 4.3 below.

**Proof of theorem 4.1** By the definition of $u_n$, we have

$$|\mathcal{G}(u^\dagger) - \mathcal{G}(u_n)|_{\mathcal{C}_1}^2 + \frac{1}{n}\|u_n\|_E^2 + \frac{2}{n}\sum_{j=1}^{n}\langle\mathcal{G}(u^\dagger) - \mathcal{G}(u_n), \mathcal{C}_1^{-1}\eta_j\rangle \leqslant \frac{1}{n}\|u^\dagger\|_E^2.$$

Therefore,

$$|\mathcal{G}(u^\dagger) - \mathcal{G}(u_n)|_{\mathcal{C}_1}^2 + \frac{1}{n}\|u_n\|_E^2 \leqslant \frac{1}{n}\|u^\dagger\|_E^2 + \frac{2}{n}|\mathcal{G}(u^\dagger) - \mathcal{G}(u_n)|_{\mathcal{C}_1}|\sum_{j=1}^{n}\mathcal{C}_1^{-1/2}\eta_j|.$$

Using Young's inequality (see lemma 1.8 of [31], for example) for the last term on the right-hand side, we obtain

$$\frac{1}{2}|\mathcal{G}(u^\dagger) - \mathcal{G}(u_n)|_{\mathcal{C}_1}^2 + \frac{1}{n}\|u_n\|_E^2 \leqslant \frac{1}{n}\|u^\dagger\|_E^2 + \frac{2}{n^2}\left(\sum_{j=1}^{n}\mathcal{C}_1^{-1/2}\eta_j\right)^2.$$

Taking expectation and noting that the $\{\eta_j\}$ are independent, we obtain

$$\frac{1}{2}\mathbb{E}|\mathcal{G}(u^\dagger) - \mathcal{G}(u_n)|_{\mathcal{C}_1}^2 + \frac{1}{n}\mathbb{E}\|u_n\|_E^2 \leqslant \frac{1}{n}\|u^\dagger\|_E^2 + \frac{2K}{n},$$

where $K = \mathbb{E}|\mathcal{C}_1^{-1/2}\eta_1|^2$. This implies that

$$\mathbb{E}|\mathcal{G}(u^\dagger) - \mathcal{G}(u_n)|_{\mathcal{C}_1}^2 \to 0 \quad \text{as} \quad n \to \infty \tag{4.2}$$

and

$$\mathbb{E}\|u_n\|_E^2 \leqslant \|u^\dagger\|_E^2 + 2K. \tag{4.3}$$

(1) We first show using (4.3) that there exist $u^* \in E$ and a subsequence $\{u_{n_k(k)}\}_{k\in\mathbb{N}}$ of $\{u_n\}$ such that

$$\mathbb{E}\langle u_{n_k(k)}, v\rangle_E \to \mathbb{E}\langle u^*, v\rangle_E \quad \text{for any } v \in E. \tag{4.4}$$

Let $\{\phi_j\}_{j\in\mathbb{N}}$ be a complete orthonormal system for $E$. Then,

$$\mathbb{E}\langle u_n, \phi_1\rangle_E \leqslant \mathbb{E}\|u_n\|_E\|\phi_1\|_E \leqslant \|u^\dagger\|_E^2 + 2K.$$

Therefore, there exist $\xi_1 \in \mathbb{R}$ and a subsequence $\{u_{n_1(k)}\}_{k\in\mathbb{N}}$ of $\{u_n\}_{n\in\mathbb{N}}$ such that $\mathbb{E}\langle u_{n_1(k)}, \phi_1\rangle \to \xi_1$. Now considering $\mathbb{E}\langle u_{n_1(k)}, \phi_2\rangle$ and using the same argument, we conclude that there exist $\xi_2 \in \mathbb{R}$ and a subsequence $\{u_{n_2(k)}\}_{k\in\mathbb{N}}$ of $\{u_{n_1(k)}\}_{k\in\mathbb{N}}$ such that $\mathbb{E}\langle u_{n_2(k)}, \phi_2\rangle \to \xi_2$. Continuing similarly, we can show that there exist $\{\xi_j\} \in \mathbb{R}^\infty$ and $\{u_{n_1(k)}\}_{k\in\mathbb{N}} \supset \{u_{n_2(k)}\}_{k\in\mathbb{N}} \supset \ldots \supset \{u_{n_j(k)}\}_{k\in\mathbb{N}}$ such that $\mathbb{E}\langle u_{n_j(k)}, \phi_j\rangle \to \xi_j$ for any $j \in \mathbb{N}$ as $k \to \infty$. Therefore,

$$\mathbb{E}\langle u_{n_k(k)}, \phi_j\rangle_E \to \xi_j, \quad \text{as } k \to \infty \text{ for any } j \in \mathbb{N}.$$

15

We need to show that $\{\xi_j\} \in \ell^2(\mathbb{R})$. We have, for any $N \in \mathbb{N}$,

$$\sum_{j=1}^{N} \xi_j^2 \leqslant \lim_{k \to \infty} \mathbb{E} \sum_{j=1}^{N} \langle u_{n_k(k)}, \phi_j \rangle_E^2 \leqslant \limsup_{k \to \infty} \mathbb{E} \|u_{n_k(k)}\|_E^2 \leqslant \|u^\dagger\|_E^2 + 2K.$$

Therefore, $\{\xi_j\} \in \ell^2(\mathbb{R})$ and $u^* := \sum_{j=1}^{\infty} \xi_j \phi_j \in E$. We can now write for any nonzero $v \in E$,

$$\mathbb{E} \langle u_{n_k(k)} - u^*, v \rangle_E = \mathbb{E} \sum_{j=1}^{\infty} \langle v, \phi_j \rangle_E \langle u_{n_k(k)} - u^*, \phi_j \rangle_E$$

$$\leqslant N \|v\|_E \mathbb{E} \sup_{j \in \{1, \dots, N\}} |\langle u_{n_k(k)} - u^*, \phi_j \rangle_E|$$

$$+ (\|u^\dagger\|_E^2 + 2K)^{1/2} \sum_{j=N}^{\infty} |\langle v, \phi_j \rangle_E|.$$

Now for any fixed $\varepsilon > 0$, we choose $N$ large enough so that

$$(\|u^\dagger\|_E^2 + 2K)^{1/2} \sum_{j=N}^{\infty} |\langle v, \phi_j \rangle_E| < \frac{1}{2}\varepsilon,$$

and then $k$ large enough so that

$$N \|v\|_E \mathbb{E} |\langle u_{n_k(k)} - u^*, \phi_j \rangle_E| < \tfrac{1}{2}\varepsilon \quad \text{for any } 1 \leqslant j \leqslant N.$$

This demonstrates that $\mathbb{E} \langle u_{n_k(k)} - u^*, v \rangle_E \to 0$ as $k \to \infty$.

(2) Now we show the almost sure existence of a convergent subsequence of $\{u_{n_k(k)}\}$. By (4.2), we have $|\mathcal{G}(u_{n_k(k)}) - \mathcal{G}(u^\dagger)|_{\mathcal{C}_1} \to 0$ in probability as $k \to \infty$. Therefore, there exists a subsequence $\{u_{m(k)}\}$ of $\{u_{n_k(k)}\}$ such that

$$\mathcal{G}(u_{m(k)}) \to \mathcal{G}(u^\dagger) \quad \text{a.s. as } k \to \infty.$$

Now by (4.4), we have $\langle u_{m(k)} - u^*, v \rangle_E \to 0$ in probability as $k \to \infty$, and hence there exists a subsequence $\{u_{\hat{m}(k)}\}$ of $\{u_{m(k)}\}$ such that $u_{\hat{m}(k)}$ converges weakly to $u^*$ in $E$ almost surely as $k \to \infty$. Since $E$ is compactly embedded in $X$, this implies that $u_{\hat{m}(k)} \to u^*$ in $X$ almost surely as $k \to \infty$. The result now follows by continuity of $\mathcal{G}$. $\qquad\square$

In the case where $u^\dagger \in X$ (and not necessarily in $E$), we have the following weaker result.

**Corollary 4.3.** *Suppose that $\mathcal{G}$ and $u_n$ satisfy the assumptions of theorem 4.1, and that $u^\dagger \in X$. Then there exists a subsequence of $\{\mathcal{G}(u_n)\}_{n \in \mathbb{N}}$ converging to $\mathcal{G}(u^\dagger)$ almost surely.*

**Proof.** For any $\varepsilon > 0$, by the density of $E$ in $X$, there exists $v \in E$ such that $\|u^\dagger - v\|_X \leqslant \varepsilon$. Then by the definition of $u_n$, we can write

$$|\mathcal{G}(u^\dagger) - \mathcal{G}(u_n)|_{\mathcal{C}_1}^2 + \frac{1}{n}\|u_n\|_E^2 + \frac{2}{n} \sum_{j=1}^{n} \langle \mathcal{G}(u^\dagger) - \mathcal{G}(u_n), \mathcal{C}_1^{-1}\eta_j \rangle$$

$$\leqslant |\mathcal{G}(u^\dagger) - \mathcal{G}(v)|_{\mathcal{C}_1}^2 + \frac{1}{n}\|v\|_E^2 + \frac{2}{n} \sum_{j=1}^{n} \langle \mathcal{G}(u^\dagger) - \mathcal{G}(v), \mathcal{C}_1^{-1}\eta_j \rangle.$$

Therefore, dropping $\frac{1}{n}\|u_n\|_E^2$ on the left-hand side, and using Young's inequality, we obtain

$$\frac{1}{2}|\mathcal{G}(u^\dagger) - \mathcal{G}(u_n)|_{\mathcal{C}_1}^2 \leqslant 2|\mathcal{G}(u^\dagger) - \mathcal{G}(v)|_{\mathcal{C}_1}^2 + \frac{1}{n}\|v\|_E^2 + \frac{3}{n^2} \sum_{j=1}^{n} |\mathcal{C}_1^{-1/2}\eta_j|^2.$$

By the local Lipschitz continuity of $\mathcal{G}$, $|\mathcal{G}(u^\dagger) - \mathcal{G}(v)|_{\mathcal{C}_1} \leqslant C\varepsilon^2$, and therefore, taking the expectations and noting the independence of $\{\eta_j\}$, we obtain

$$\mathbb{E}|\mathcal{G}(u^\dagger) - \mathcal{G}(u_n)|^2_{\mathcal{C}_1} \leqslant 4C\varepsilon^2 + \frac{2C_\varepsilon}{n} + \frac{6K}{n},$$

implying that

$$\limsup_{n \to \infty} \mathbb{E}|\mathcal{G}(u^\dagger) - \mathcal{G}(u_n)|^2_{\mathcal{C}_1} \leqslant 4C\varepsilon^2.$$

Since the $\liminf$ is obviously positive and $\varepsilon$ was arbitrary, we have $\lim_{n\to\infty} \mathbb{E}|\mathcal{G}(u^\dagger) - \mathcal{G}(u_n)|^2_{\mathcal{C}_1} = 0$. This implies that $|\mathcal{G}(u^\dagger) - \mathcal{G}(u_n)|_{\mathcal{C}_1} \to 0$ in probability. Therefore, there exists a subsequence of $\{\mathcal{G}(u_n)\}$ which converges to $\mathcal{G}(u^\dagger)$ almost surely. $\qquad\square$

### 4.2. Small noise limit

Consider the case where, as data, we have the random vector

$$y_n = \mathcal{G}(u^\dagger) + \frac{1}{n}\eta_n \tag{4.5}$$

for $n \in \mathbb{N}$ and with $u^\dagger$ again as the true solution and $\eta_j \sim \mathcal{N}(0, \mathcal{C}_1)$, $j \in \mathbb{N}$, Gaussian random vectors in $\mathbb{R}^K$. Thus, in the preceding general setting, we have $G = \mathcal{G}$ and $J = K$. Rather than having $n$ independent observations, we have an observation noise scaled by small $\gamma = 1/n$ converging to zero. For these data and given the prior measure $\mu_0$ on $u$, we have the following formula for the posterior measure:

$$\frac{\mathrm{d}\mu^{y_n}}{\mathrm{d}\mu_0}(u) \propto \exp\left(-\frac{n^2}{2}|y_n - \mathcal{G}(u)|^2_{\mathcal{C}_1}\right).$$

By the result of the previous section, the MAP estimators for the above measure are the minimizers of

$$I_n(u) := \|u\|^2_E + n^2|y_n - \mathcal{G}(u)|^2_{\mathcal{C}_1}. \tag{4.6}$$

Our interest is in studying the properties of the limits of minimizers of $I_n$ as $n \to \infty$. We have the following almost-sure convergence result.

**Theorem 4.4.** *Assume that $\mathcal{G} : X \to \mathbb{R}^K$ is Lipschitz on bounded sets, and $u^\dagger \in E$. For every $n \in \mathbb{N}$, let $u_n \in E$ be a minimizer of $I_n(u)$ given by (4.6). Then, there exist a $u^* \in E$ and a subsequence of $\{u_n\}_{n \in \mathbb{N}}$ that converges weakly to $u^*$ in $E$, almost surely. For any such $u^*$, we have $\mathcal{G}(u^*) = \mathcal{G}(u^\dagger)$.*

**Proof.** The proof is very similar to that of theorem 4.1 and so we only sketch differences. We have

$$\begin{aligned}
I_n &= \|u\|^2_E + n^2|y_n - \mathcal{G}(u)|^2_{\mathcal{C}_1} \\
&= \|u\|^2_E + n^2\left|\mathcal{G}(u^\dagger) + \frac{1}{n}\eta_n - \mathcal{G}(u)\right|^2_{\mathcal{C}_1} \\
&= \|u\|^2_E + n^2|\mathcal{G}(u^\dagger) - \mathcal{G}(u)|^2_{\mathcal{C}_1} + |\eta_n|^2_{\mathcal{C}_1} + 2n\langle\mathcal{G}(u^\dagger) - \mathcal{G}(u), \eta_n\rangle_{\mathcal{C}_1}.
\end{aligned}$$

Letting

$$J_n(u) = \frac{1}{n^2}\|u\|^2_E + |\mathcal{G}(u^\dagger) - \mathcal{G}(u)|^2_{\mathcal{C}_1} + \frac{2}{n}\langle\mathcal{G}(u^\dagger) - \mathcal{G}(u), \eta_n\rangle_{\mathcal{C}_1},$$

we hence have arg min$_u I_n = $ arg min$_u J_n$. For this $J_n$, the result of lemma 4.2 holds true, using an argument similar to the large sample size case. The result of theorem 4.4 carries over as well. Indeed, by the definition of $u_n$, we have

$$|\mathcal{G}(u^\dagger) - \mathcal{G}(u_n)|^2_{\mathcal{C}_1} + \frac{1}{n^2}\|u_n\|^2_E + \frac{2}{n}\langle\mathcal{G}(u^\dagger) - \mathcal{G}(u_n), \mathcal{C}_1^{-1}\eta_n\rangle \leqslant \frac{1}{n^2}\|u^\dagger\|^2_E.$$

Therefore,

$$|\mathcal{G}(u^\dagger) - \mathcal{G}(u_n)|^2_{\mathcal{C}_1} + \frac{1}{n^2}\|u_n\|^2_E \leqslant \frac{1}{n^2}\|u^\dagger\|^2_E + \frac{2}{n}|\mathcal{G}(u^\dagger) - \mathcal{G}(u_n)|_{\mathcal{C}_1}|\mathcal{C}_1^{-1/2}\eta_n|.$$

Using Young's inequality for the last term on the right-hand side, we obtain

$$\frac{1}{2}|\mathcal{G}(u^\dagger) - \mathcal{G}(u_n)|^2_{\mathcal{C}_1} + \frac{1}{n^2}\|u_n\|^2_E \leqslant \frac{1}{n^2}\|u^\dagger\|^2_E + \frac{2}{n^2}|\mathcal{C}_1^{-1/2}\eta_n|^2.$$

Taking expectation, we obtain

$$\mathbb{E}|\mathcal{G}(u^\dagger) - \mathcal{G}(u_n)|^2_{\mathcal{C}_1} + \frac{1}{n^2}\mathbb{E}\|u_n\|^2_E \leqslant \frac{1}{n^2}\|u^\dagger\|^2_E + \frac{2K}{n^2}.$$

This implies that

$$\mathbb{E}|\mathcal{G}(u^\dagger) - \mathcal{G}(u_n)|^2_{\mathcal{C}_1} \to 0 \quad \text{as} \quad n \to \infty \tag{4.7}$$

and

$$\mathbb{E}\|u_n\|^2_E \leqslant \|u^\dagger\|^2_E + 2K. \tag{4.8}$$

Having (4.7) and (4.8), and with the same argument as in the proof of theorem 4.1, it follows that there exist a $u^* \in E$ and a subsequence of $\{u_n\}$ that converges weakly to $u^*$ in $E$ almost surely, and for any such $u^*$, we have $\mathcal{G}(u^*) = \mathcal{G}(u^\dagger)$. □

As in the large sample size case, here also, if we have $u^\dagger \in X$ and we do not restrict the true solution to be in the Cameron–Martin space $E$, one can prove, in a similar way to the argument of the proof of corollary 4.3, the following weaker convergence result.

**Corollary 4.5.** *Suppose that $\mathcal{G}$ and $u_n$ satisfy the assumptions of theorem 4.4, and that $u^\dagger \in X$. Then there exists a subsequence of $\{\mathcal{G}(u_n)\}_{n\in\mathbb{N}}$ converging to $\mathcal{G}(u^\dagger)$ almost surely.*

## 5. Applications in fluid mechanics

In this section, we present an application of the methods presented above to filtering and smoothing in fluid dynamics, which is relevant to data assimilation applications in oceanography and meteorology. We link the MAP estimators introduced in this paper to the variational methods used in applications [2], and we demonstrate posterior consistency in this context.

We consider the 2D Navier–Stokes equation on the torus $\mathbb{T}^2 := [-1, 1) \times [-1, 1)$ with the following periodic boundary conditions:

$$\begin{array}{ll} \partial_t v - \nu\Delta v + v \cdot \nabla v + \nabla p = f & \text{for all } (x, t) \in \mathbb{T}^2 \times (0, \infty), \\ \nabla \cdot v = 0 & \text{for all } (x, t) \in \mathbb{T}^2 \times (0, \infty), \\ v = u & \text{for all } (x, t) \in \mathbb{T}^2 \times \{0\}. \end{array}$$

Here, $v : \mathbb{T}^2 \times (0, \infty) \to \mathbb{R}^2$ is a time-dependent vector field representing the velocity, $p : \mathbb{T}^2 \times (0, \infty) \to \mathbb{R}$ is a time-dependent scalar field representing the pressure, $f : \mathbb{T}^2 \to \mathbb{R}^2$ is a vector field representing the forcing (which we assume to be time independent for simplicity) and $\nu$ is the viscosity. We are interested in the inverse problem of determining the initial velocity field $u$ from pointwise measurements of the velocity field at later times. This is a model for the

situation in weather forecasting where observations of the atmosphere are used to improve the initial condition used for forecasting. For simplicity, we assume that the initial velocity field is divergence free and integrates to zero over $\mathbb{T}^2$, noting that this property will be preserved in time.

Define

$$\mathcal{H} := \left\{ \text{trigonometric polynomials } u : \mathbb{T}^2 \to \mathbb{R}^2 \middle| \nabla \cdot u = 0, \int_{\mathbb{T}^2} u(x) \, dx = 0 \right\}$$

and $H$ as the closure of $\mathcal{H}$ with respect to the $(L^2(\mathbb{T}^2))^2$ norm. We define the map $P : (L^2(\mathbb{T}^2))^2 \to H$ to be the Leray–Helmholtz orthogonal projector (see [31]). Given $k = (k_1, k_2)^T$, define $k^\perp := (k_2, -k_1)^T$. Then an orthonormal basis for $H$ is given by $\psi_k : \mathbb{R}^2 \to \mathbb{R}^2$, where

$$\psi_k(x) := \frac{k^\perp}{|k|} \exp(\pi i k \cdot x)$$

for $k \in \mathbb{Z}^2 \setminus \{0\}$. Thus, for $u \in H$, we may write

$$u = \sum_{k \in \mathbb{Z}^2 \setminus \{0\}} u_k(t) \psi_k(x),$$

where, since $u$ is a real-valued function, we have the reality constraint $u_{-k} = -\bar{u}_k$. Using the Fourier decomposition of $u$, we define the fractional Sobolev spaces

$$H^s := \left\{ u \in H \middle| \sum_{k \in \mathbb{Z}^2 \setminus \{0\}} (\pi^2 |k|^2)^s |u_k|^2 < \infty \right\}$$

with the norm $\|u\|_s := \left( \sum_k (\pi^2 |k|^2)^s |u_k|^2 \right)^{1/2}$, where $s \in \mathbb{R}$. If $A = -P\Delta$, the Stokes operator, then $H^s = D(A^{s/2})$. We assume that $f \in H^s$ for some $s > 0$.

Let $t_\ell = \ell h$, for $\ell = 0, \ldots, L$, and define $v_\ell \in \mathbb{R}^M$ to be the set of pointwise values of the velocity field given by $\{v(x_m, t_\ell)\}_{m \in \mathbb{M}}$, where $\mathbb{M}$ is some finite set of points in $\mathbb{T}^2$ with cardinality $M/2$. Note that each $v_\ell$ depends on $u$ and we may define $\mathcal{G}_\ell : H \to \mathbb{R}^M$ by $\mathcal{G}_\ell(u) = v_\ell$. We let $\{\eta_\ell\}_{\ell \in \{1, \ldots, L\}}$ be a set of random variables in $\mathbb{R}^M$ which perturbs the points $\{v_\ell\}_{\ell \in \{1, \ldots, L\}}$ to generate the observations $\{y_\ell\}_{\ell \in \{1, \ldots, L\}}$ in $\mathbb{R}^M$ given by

$$y_\ell := v_\ell + \gamma \eta_\ell, \quad \ell \in \{1, \ldots, L\}.$$

We let $y = \{y_\ell\}_{\ell=1}^L$, the accumulated data up to time $T = Lh$, with similar notation for $\eta$, and define $\mathcal{G} : H \to \mathbb{R}^{ML}$ by $\mathcal{G}(u) = (\mathcal{G}_1(u), \ldots, \mathcal{G}_L(u))$. We now solve the inverse problem of finding $u$ from $y = \mathcal{G}(u) + \gamma \eta$. We assume that the prior distribution on $u$ is a Gaussian $\mu_0 = N(0, \mathcal{C}_0)$, with the property that $\mu_0(H) = 1$ and that the observational noise $\{\eta_\ell\}_{\ell \in \{1, \ldots, L\}}$ is i.i.d. in $\mathbb{R}^M$, independent of $u$, with $\eta_1$ distributed according to a Gaussian measure $N(0, I)$. If we define

$$\Phi(u) = \frac{1}{2\gamma^2} \sum_{j=1}^L |y_j - \mathcal{G}_j(u)|^2,$$

then under the preceding assumptions, the Bayesian inverse problem for the posterior measure $\mu^y$ for $u|y$ is well defined and is Lipschitz in $y$ with respect to the Hellinger metric (see [7]). The Onsager–Machlup functional in this case is given by

$$I_{NS}(u) = \tfrac{1}{2} \|u\|_{\mathcal{C}_0}^2 + \Phi(u).$$

We are in the setting of subsection 4.2, with $\gamma = 1/n$ and $K = ML$. In the applied literature, approaches to assimilating data into mathematical models based on minimizing $I_{NS}$ are known as *variational methods*, and sometimes as 4DVAR [2].

19
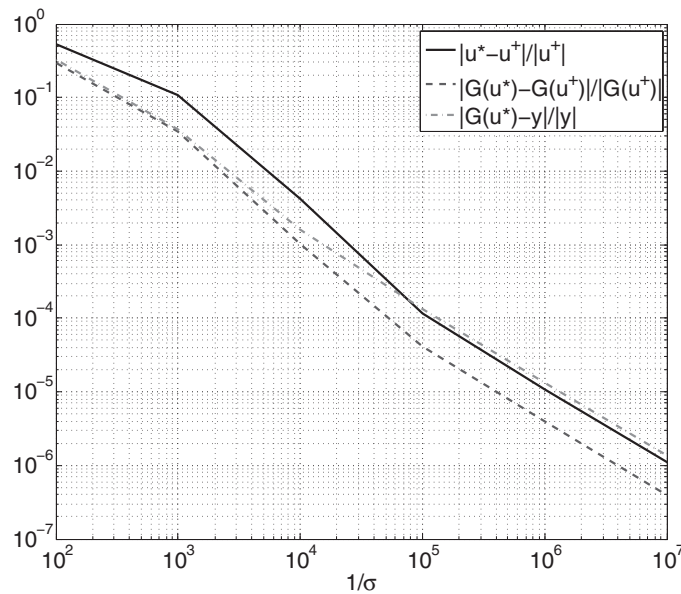
**Figure 1.** Illustration of posterior consistency in the fluid mechanics application. The three curves given are the relative error of the MAP estimator $u^*$ in reproducing the truth, $u^\dagger$ (solid), the relative error of the map $\mathcal{G}(u^*)$ in reproducing $\mathcal{G}(u^\dagger)$ (dashed) and the relative error of $\mathcal{G}(u^*)$ with respect to the observations $y$ (dash–dotted).

We now describe numerical experiments concerned with studying posterior consistency in the case $\gamma \to 0$. We let $\mathcal{C}_0 = A^{-2}$ noting that if $u \sim \mu_0$, then $u \in H^s$ almost surely for all $s < 1$; in particular, $u \in H$. Thus $\mu_0(H) = 1$ as required. The forcing in $f$ is taken to be $f = \nabla^\perp \Psi$, where $\Psi = \cos(\pi k \cdot x)$ and $\nabla^\perp = J\nabla$ with $J$ the canonical skew-symmetric matrix and $k = (5, 5)$. The dimension of the attractor is determined by the viscosity parameter $\nu$. For the particular forcing used, there is an explicit steady state for all $\nu > 0$, and for $\nu \geqslant 0.035$, this solution is stable (see [26], chapter 2, for details). As $\nu$ decreases, the flow becomes increasingly complex, and we focus subsequent studies of the inverse problem on the mildly chaotic regime which arises for $\nu = 0.01$. We use a time-step of $\delta t = 0.005$. The data are generated by computing a true signal solving the Navier–Stokes equation at the desired value of $\nu$, and then adding Gaussian random noise to it at each observation time. Furthermore, we let $h = 4\delta t = 0.02$ and take $L = 10$, so that $T = 0.2$. We take $M = 32^2$ spatial observations at each observation time. The observations are made at the gridpoints; thus the observations include all numerically resolved, and hence observable, wavenumbers in the system. Since the noise is added in spectral space in practice, for convenience we define $\sigma = \gamma/\sqrt{M}$ and present the results in terms of $\sigma$. The same grid is used for computing the reference solution and the MAP estimator.

Figure 1 illustrates the posterior consistency which arises as the observational noise strength $\gamma \to 0$. The three curves shown quantify (i) the relative error of the MAP estimator $u^*$ compared with the truth, $u^\dagger$, (ii) the relative error of $\mathcal{G}(u^*)$ compared with $\mathcal{G}(u^\dagger)$ and (iii) the relative error of $\mathcal{G}(u^*)$ with respect to the observations $y$. The figure clearly illustrates theorem 4.4, via the dashed curve for (ii), and indeed shows that the map estimator itself is converging to the true initial condition, via the blue curve, as $\gamma \to 0$. Recall that the observations approach the true value of the initial condition, mapped forward under $\mathcal{G}$, as

$\gamma \to 0$, and note that the dashed and dashed–dotted curves shows that the image of the MAP estimator under the forward operator $\mathcal{G}$, $\mathcal{G}(u^*)$, is closer to $\mathcal{G}(u^\dagger)$ than $y$, asymptotically as $\gamma \to 0$.

## 6. Applications in conditioned diffusions

In this section, we consider the MAP estimator for conditioned diffusions, including bridge diffusions and an application to filtering/smoothing. We identify the Onsager–Machlup functional governing the MAP estimator in three different cases. We demonstrate numerically that this functional may have more than one minimizer. Furthermore, we illustrate the results of the consistency theory in section 4 using numerical experiments. Subsection 6.1 concerns the unconditioned case, and includes the assumptions made throughout. Subsections 6.2 and 6.3 describe bridge diffusions and the filtering/smoothing problem, respectively. Finally, subsection 6.4 is devoted to numerical experiments for an example in filtering/smoothing.

### 6.1. Unconditioned case

For simplicity, we restrict ourselves to scalar processes with additive noise, taking the form

$$du = f(u)\,dt + \sigma\,dW, \quad u(0) = u^-. \tag{6.1}$$

If we let $\nu$ denote the measure on $X := C([0, T]; \mathbb{R})$ generated by the stochastic differential equation (SDE) given in (6.1), and $\nu_0$ the same measure obtained in the case $f \equiv 0$, then the Girsanov theorem states that $\nu \ll \nu_0$ with density

$$\frac{d\nu}{d\nu_0}(u) = \exp\left(-\frac{1}{2\sigma^2}\int_0^T |f(u(t))|^2\,dt + \frac{1}{\sigma^2}\int_0^T f(u(t))du(t)\right).$$

If we choose an $F : \mathbb{R} \to \mathbb{R}$ with $F'(u) = f(u)$, then an application of Itô's formula gives

$$dF(u(t)) = f(u(t))\,du(t) + \frac{\sigma^2}{2}f'(u(t))\,dt,$$

and using this expression to remove the stochastic integral, we obtain

$$\frac{d\nu}{d\nu_0}(u) \propto \exp\left(-\frac{1}{2\sigma^2}\int_0^T (|f(u(t))|^2 + \sigma^2 f'(u(t)))\,dt + \frac{1}{\sigma^2}F(u(T))\right). \tag{6.2}$$

Thus, the measure $\nu$ has a density with respect to the Gaussian measure $\nu_0$ and (6.2) takes the form (1.1) with $\mu = \nu$ and $\mu_0 = \nu_0$; we have

$$\frac{d\nu}{d\nu_0}(u) \propto \exp(-\Phi_1(u)),$$

where $\Phi_1 : X \to \mathbb{R}$ is defined by

$$\Phi_1(u) = \int_0^T \Psi(u(t))\,dt - \frac{1}{\sigma^2}F(u(T)) \tag{6.3}$$

and

$$\Psi(u) = \frac{1}{2\sigma^2}(|f(u)|^2 + \sigma^2 f'(u)).$$

We make the following assumption concerning the vector field $f$ driving the SDE.

**Assumption 6.1.** *The function $f = F'$ in (6.1) satisfies the following conditions.*

*(1) $F \in C^2(\mathbb{R}, \mathbb{R})$ for all $u \in \mathbb{R}$.*
*(2) There is $M \in \mathbb{R}$ such that $\Psi(u) \geqslant M$ for all $u \in \mathbb{R}$ and $F(u) \leqslant M$ for all $u \in \mathbb{R}$.*

Under these assumptions, we see that $\Phi_1$ given by (6.3) satisfies assumptions 2.1 and, indeed, the slightly stronger assumptions made in theorem 3.5. Let $H^1[0, T]$ denote the space of absolutely continuous functions on $[0, T]$. Then the Cameron–Martin space $E_1$ for $\nu_0$ is

$$E_1 = \left\{ v \in H^1[0, T] \Big| \int_0^T |v'(s)|^2 \, ds < \infty \text{ and } v(0) = 0 \right\}$$

and the Cameron–Martin norm is given by

$$\|v\|_{E_1} = \sigma^{-1} \|v\|_{H^1},$$

where

$$\|v\|_{H^1} = \left( \int_0^T \left| v'(s) \right|^2 ds \right)^{\frac{1}{2}}.$$

The mean of $\nu_0$ is the constant function $m \equiv u^-$ and so, using remark 2.2, we see that the Onsager–Machlup functional for the unconditioned diffusion (6.1) is thus $I_1 : E_1 \to \mathbb{R}$ given by

$$I_1(u) = \Phi_1(u) + \frac{1}{2\sigma^2} \|u - u^-\|_{H^1}^2 = \Phi_1(u) + \frac{1}{2\sigma^2} \|u\|_{H^1}^2.$$

Together, theorems 3.2 and 3.5 tell us that this functional attains its minimum over $E_1'$ defined by

$$E_1' = \left\{ v \in H^1[0, T] \Big| \int_0^T |v'(s)|^2 ds < \infty \quad \text{and } v(0) = u^- \right\}.$$

Furthermore, such minimizers define MAP estimators for the unconditioned diffusion (6.1), i.e. the most likely paths of the diffusion.

We note that the regularity of minimizers for $I_1$ implies that the MAP estimator is $C^2$, whilst sample paths of the SDE (6.1) are not even differentiable. This is because the MAP estimator defines the centre of a tube in $X$ which contains the most likely paths. The centre itself is a smoother function than the paths. This is a generic feature of MAP estimators for measures defined via density with respect to a Gaussian in infinite dimensions.

## 6.2. Bridge diffusions

In this subsection, we study the probability measure generated by solutions of (6.1), conditioned to hit $u^+$ at time 1 so that $u(T) = u^+$, and denote this measure $\mu$. Let $\mu_0$ denote the Brownian bridge measure obtained in the case $f \equiv 0$. By applying the approach to determining bridge diffusion measures in [17], we obtain, from (6.2), the expression

$$\frac{d\mu}{d\mu_0}(u) \propto \exp\left( -\int_0^T \Psi(u(t)) \, dt + \frac{1}{\sigma^2} F(u^+) \right). \tag{6.4}$$

Since $u^+$ is fixed, we now define $\Phi_2 : X \to \mathbb{R}$ by

$$\Phi_2(u) = \int_0^T \Psi(u(t)) \, dt$$

and then (6.4) takes again the form (1.1). The Cameron–Martin space for the (zero mean) Brownian bridge is

$$E_2 = \left\{ v \in H^1[0, T] \Big| \int_0^T |v'(s)|^2 \, ds < \infty \text{ and } v(0) = v(T) = 0 \right\},$$

and the Cameron–Martin norm is again $\sigma^{-1}\|\cdot\|_{H^1}$. The Onsager–Machlup function for the unconditioned diffusion (6.1) is thus $I_2 : E'_2 \to \mathbb{R}$, given by

$$I_2(u) = \Phi_2(u) + \frac{1}{2\sigma^2}\|u - m\|_{H^1}^2,$$

where $m$, given by $m(t) = \frac{T-t}{T}u^- + \frac{t}{T}u^+$ for all $t \in [0, T]$, is the mean of $\mu_0$ and

$$E'_2 = \left\{ v \in H^1[0, T] \Big| \int_0^T |v'(s)|^2 \, \mathrm{d}s < \infty \text{ and } v(0) = u^-, u(T) = u^+ \right\}.$$

The MAP estimators for $\mu$ are found by minimizing $I_2$ over $E'_2$.

### 6.3. Filtering and smoothing

We now consider conditioning the measure $\nu$ on observations of the process $u$ at discrete time points. Assume that we observe $y \in \mathbb{R}^J$ given by

$$y_j = u(t_j) + \eta_j, \tag{6.5}$$

where $0 < t_1 < \cdots < t_J < T$ and the $\eta_j$ are independent identically distributed random variables with $\eta_j \sim N(0, \gamma^2)$. Let $\mathbb{Q}_0(\mathrm{d}y)$ denote the $\mathbb{R}^J$-valued Gaussian measure $N(0, \gamma^2 I)$ and let $\mathbb{Q}(\mathrm{d}y|u)$ denote the $\mathbb{R}^J$-valued Gaussian measure $N(\mathcal{G}u, \gamma^2 I)$, where $\mathcal{G} : X \to \mathbb{R}^J$ is defined by

$$\mathcal{G}u = (u(t_1), \ldots, u(t_J)).$$

Recall $\nu_0$ and $\nu$ from the unconditioned case and define the measures $\mathbb{P}_0$ and $\mathbb{P}$ on $X \times \mathbb{R}^J$ as follows. The measure $\mathbb{P}_0(\mathrm{d}u, \mathrm{d}y) = \nu_0(\mathrm{d}u)\mathbb{Q}_0(\mathrm{d}y)$ is defined to be an independent product of $\nu_0$ and $\mathbb{Q}_0$, whilst $\mathbb{P}(\mathrm{d}u, \mathrm{d}y) = \nu(\mathrm{d}u)\mathbb{Q}(\mathrm{d}y|u)$. Then,

$$\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{P}_0}(u, y) \propto \exp\left( -\int_0^T \Psi(u(t)) \, \mathrm{d}t + \frac{1}{\sigma^2}F(u(T)) - \frac{1}{2\gamma^2}\sum_{j=1}^J |y_j - u(t_j)|^2 \right)$$

with the constant of proportionality depending only on $y$. Clearly, by continuity,

$$\inf_{\|u\|_X \leqslant 1} \exp\left( -\int_0^T \Psi(u(t)) \, \mathrm{d}t + \frac{1}{\sigma^2}F(u(T)) - \frac{1}{2\gamma^2}\sum_{j=1}^J |y_j - u(t_j)|^2 \right) > 0,$$

and hence,

$$\int_{\|u\|_X \leqslant 1} \exp\left( -\int_0^T \Psi(u(t)) \, \mathrm{d}t + \frac{1}{\sigma^2}F(u(T)) - \frac{1}{2\gamma^2}\sum_{j=1}^J |y_j - u(t_j)|^2 \right) \nu_0(\mathrm{d}u) > 0.$$

Applying the conditioning lemma 5.3 in [17] then gives

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\nu_0}(u) \propto \exp\left( -\int_0^T \Psi(u(t)) \, \mathrm{d}t + \frac{1}{\sigma^2}F(u(T)) - \frac{1}{2\gamma^2}\sum_{j=1}^J |y_j - u(t_j)|^2 \right).$$
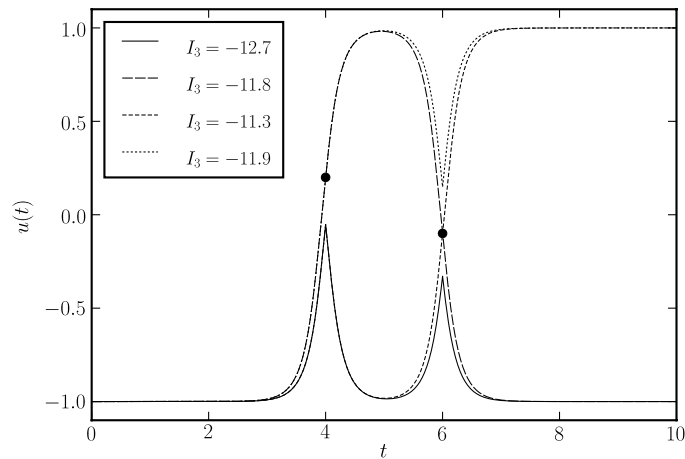
Thus, we define

$$\Phi_3(u) = \int_0^T \Psi(u(t)) \, \mathrm{d}t - \frac{1}{\sigma^2}F(u(T)) + \frac{1}{2\gamma^2}\sum_{j=1}^J |y_j - u(t_j)|^2.$$

The Cameron–Martin space is again $E_1$ and the Onsager–Machlup functional is thus $I_3 : E'_1 \to \mathbb{R}$, given by

$$I_3(u) = \Phi_3(u) + \frac{1}{2\sigma^2}\|u\|_{H^1}^2. \tag{6.6}$$

**Figure 2.** Illustration of the problem of local minima of $I$ for the smoothing problem with a small number of observations. The process $u(t)$ starts at $u(0) = -1$ and moves in a double-well potential with stable equilibrium points at $-1$ and $+1$. Two observations of the process are indicated by the two black circles. The curves correspond to four different local minima of the functional $I_3$ for this situation.

The MAP estimator for this set-up is, again, found by minimizing the Onsager–Machlup functional $I_3$.

The only difference between the potentials $\Phi_1$ and $\Phi_3$, and thus between the functionals $I_1$ for the unconditioned case and $I_3$ for the case with discrete observations, is the presence of the term $\frac{1}{2\gamma^2} \sum_{j=1}^{J} |y_j - u(t_j)|^2$. In the Euler–Lagrange equations describing the minima of $I_3$, this term leads to Dirac distributions at the observation points $t_1, \ldots, t_J$, and it transpires that, as a consequence, minimizers of $I_3$ have jumps in their first derivatives at $t_1, \ldots, t_J$. This effect can be clearly seen in the local minima of $I_3$ shown in figure 2.

### 6.4. Numerical experiments

In this section, we perform three numerical experiments related to the MAP estimator for the filtering/smoothing problem presented in section 6.3.

For the experiments, we generate a random 'signal' by numerically solving the SDE (6.1), using the Euler–Maruyama method, for a double-well potential $F$ given by

$$F(u) = -\frac{(1 - u)^2(1 + u)^2}{1 + u^2},$$

with diffusion constant $\sigma = 1$ and initial value $u^- = -1$. From the resulting solution $u(t)$, we generate random observations $y_1, \ldots, y_J$ using (6.5). Then we implement the Onsager–Machlup functional $I_3$ from equation (6.6) and use numerical minimization, employing the Broyden–Fletcher–Goldfarb–Shanno method (see [13]; we use the implementation found in the GNU scientific library [14]), to find the minima of $I_3$. The same grid is used for numerically solving the SDE and for approximating the values of $I_3$.

The first experiment concerns the problem of local minima of $I_3$. For a small number of observations, we find multiple local minima; the minimization procedure can converge to different *local* minima, depending on the starting point of the optimization. This effect makes it difficult to find the MAP estimator, which is the *global* minimum of $I_3$, numerically. The
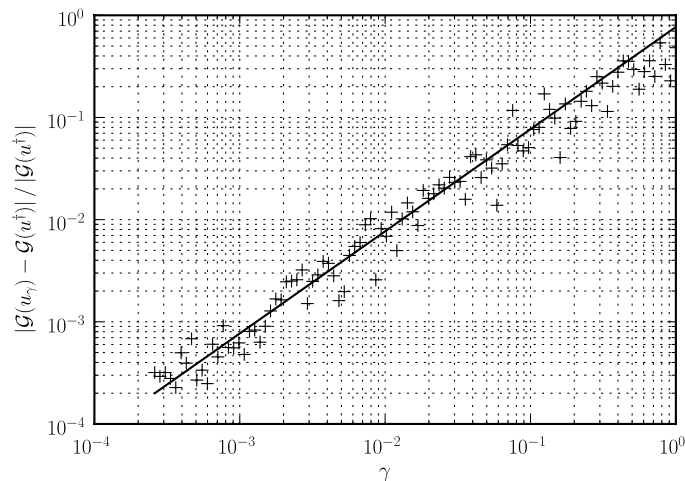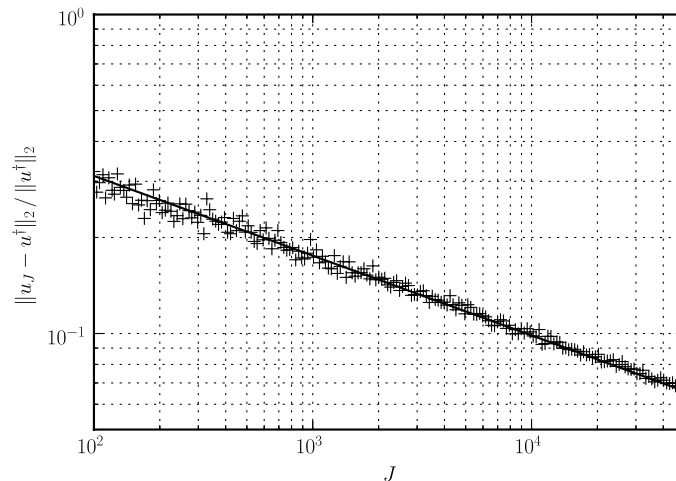
**Figure 3.** Illustration of posterior consistency for the smoothing problem in the small-noise limit. The marked points correspond the maximum-norm distance between the true signal $u^\dagger$ and the MAP estimator $u_\gamma$ with $J = 5$ evenly spaced observations. The map $\mathcal{G}(u) = (u(t_1), \ldots, u(t_J))$ is the projection of the path onto the observation points. The solid line is a fitted curve of the form $c\gamma$.

problem is illustrated in figure 2, which shows four different local minima for the case of $J = 2$ observations. In the presence of local minima, some care is needed when numerically computing the MAP estimator. For example, one could start the minimization procedure with a collection of different starting points and take the best of the resulting local minima as the result. One would expect this problem to become less pronounced as the number of observations increases, since the observations will 'pull' the MAP estimator towards the correct solution, thus reducing the number of local minima. This effect is confirmed by experiments: for larger numbers of observations, our experiments found only one local minimum.

The second experiment concerns the posterior consistency of the MAP estimator in the small noise limit. Here, we use a fixed number $J$ of observations of a fixed path of (6.1), but let the variance $\gamma^2$ of the observational noise $\eta_j$ converge to 0. Noting that the exact path of the SDE, denoted by $u^\dagger$ in (4.5), has the regularity of a Brownian motion and therefore the observed path is not contained in the Cameron–Martin space $E_3$, we are in the situation described in corollary 4.5. Our experiments indicate that we have $\mathcal{G}(u_\gamma) \to \mathcal{G}(u^\dagger)$ as $\gamma \downarrow 0$, where $u_\gamma$ denotes the MAP estimator corresponding to observational variance $\gamma^2$, confirming the result of corollary 4.5. As discussed above, for small values of $\gamma$, one would expect the minimum of $I_3$ to be unique, and indeed, experiments where different starting points of the optimization procedure were tried did not find different minima for small $\delta$. The result of a simulation with $J = 5$ is shown in figure 3.

Finally, we perform an experiment to illustrate posterior consistency in the large sample size limit: for this experiment, we still use one fixed path $u^\dagger$ of the SDE (6.1). Then, for different values of $J$, we generate observations $y_1, \ldots, y_J$ using (6.5) at equidistantly spaced times $t_1, \ldots, t_J$, for fixed $\gamma = 1$, and then determine the $L^2$ distance of the resulting MAP estimate $u_J$ to the exact path $u^\dagger$. As discussed above, for large values of $J$, one would expect the minimum of $I_3$ to be unique, and indeed, experiments where different starting points of the optimization procedure were tried did not find different minima for large $J$. The situation

25

**Figure 4.** Illustration of posterior consistency for the smoothing problem in the large sample size limit. The marked points correspond the supremum-norm distance between the true signal $u^*$ and the MAP estimator $u_J^\dagger$ with $J$ evenly spaced observations. The solid line is a fitted curve of the form $cJ^{-\alpha}$; the exponent $\alpha = -1/4$ was found numerically.

considered here is not covered by the theoretical results from section 4, but the results of the numerical experiment, shown in figure 4, indicate that posterior consistency still holds.

## References

[1] Agapiou S, Larsson S and Stuart A M 2013 Posterior consistency of the Bayesian approach to linear ill-posed inverse problems *Stoc. Proc. App.* pp 3828–60

[2] Bennett A F 2002 *Inverse Modeling of the Ocean and Atmosphere* (Cambridge: Cambridge University Press)

[3] Bissantz N, Hohage T and Munk A 2004 Consistency and rates of convergence of nonlinear Tikhonov regularization with random noise *Inverse Problems* **20** 1773–89

[4] Bochkina N A 2012 Consistency of the posterior distribution in generalised linear inverse problems arXiv:1211.3382

[5] Bogachev V I 1998 *Gaussian Measures* (*Mathematical Surveys and Monographs* vol 62) (Providence, RI: American Mathematical Society)

[6] Brown L D and Low M G 1996 Asymptotic equivalence of nonparametric regression and white noise *Ann. Stat.* **24** 2384–98

[7] Cotter S L, Dashti M, Robinson J C and Stuart A M 2009 Bayesian inverse problems for functions and applications to fluid mechanics *Inverse Problems* **25** 115008

[8] Cox S M and Matthews P C 2002 Exponential time differencing for stiff systems *J. Comput. Phys.* **176** 430–55

[9] Dembo A and Zeitouni O 1991 Onsager–Machlup functionals and maximum a posteriori estimation for a class of non-Gaussian random fields *J. Mutivariate Anal.* **36** 243–62

[10] Da Prato G and Zabczyk J 1992 *Stochastic Equations in Infinite Dimensions* (*Encyclopedia of Mathematics and Its Applications* vol 44) (Cambridge: Cambridge University Press)

[11] Dürr D and Bach A 1978 The Onsager–Machlup function as Lagrangian for the most probable path of a diffusion process *Commun. Math. Phys.* **160** 153–70

[12] Engl H W, Hanke M and Neubauer A 1996 *Regularization of Inverse Problems* (*Mathematics and Its Applications* vol 375) (Dordrecht: Kluwer)

[13] Fletcher R 2000 *Practical Methods of Optimization* 2nd edn (New York: Wiley-Blackwell)

[14] Galassi M, Davies J, Theiler J, Gough B, Jungman G, Alken P, Booth M and Rossi F 2009 *GNU Scientific Library—Reference Manual* 3rd edn (Godalming: Network Theory)

[15] Ghosal S, Ghosh J K and Ramamoorthi R V 1999 Consistency issues in Bayesian nonparametrics *Asymptotics, Nonparameters and Time Series* vol 158 (New York: Marcel Dekker Inc) pp 639–68

[16] Griebel M and Hegland M A 2010 Finite element method for density estimation with Gaussian process priors *SIAM J. Numer. Anal.* **47** 4759–92

[17] Hairer M, Stuart A M and Voss J 2007 Analysis of SPDEs arising in path sampling: II. The nonlinear case *Ann. Appl. Probab.* **17** 1657–706

[18] Hairer M, Stuart A M and Voss J 2011 Signal processing problems on function space: Bayesian formulation, stochastic PDEs and effective MCMC methods *The Oxford Handbook of Nonlinear Filtering* ed D Crisan and B Rozovsky (Oxford: Oxford University Press)

[19] Hegland M 2007 Approximate maximum a posteriori with Gaussian process priors *Constructive Approximation* **26** 205–24

[20] Hesthaven J S, Gottlieb S and Gottlieb D 2007 *Spectral Methods for Time-Dependent Problems (Cambridge Monographs on Applied and Computational Mathematics* vol 21*)* (Cambridge: Cambridge University Press)

[21] Ikeda N and Watanabe S 1989 *Stochastic Differential Equations and Diffusion Processes* 2nd edn (Amsterdam: North-Holland)

[22] Kaipio J P and Somersalo E 2005 *Statistical and Computational Inverse Problems* (Berlin: Springer)

[23] Knapik B T, van der Vaart A W and van Zanten J H 2011 Bayesian inverse problems with Gaussian priors *Ann. Stat.* **39** 2626–57

[24] Ledoux M and Talagrand M 1991 *Probability in Banach Spaces: Isoperimetry and Processes* (Berlin: Springer)

[25] Lifshits M A 1995 *Gaussian Random Functions* (Dordrecht: Kluwer)

[26] Majda A and Wang X 2006 *Non-Linear Dynamics and Statistical Theories for Basic Geophysical Flows* (Cambridge: Cambridge University Press)

[27] Markussen B 2009 Laplace approximation of transition densities posed as Brownian expectations *Stochastic Process. Appl.* **119** 208–31

[28] Mortensen R E 1968 Maximum-likelihood recursive filtering *J. Optim. Theory Appl.* **2** 386–93

[29] Ray K 2012 Bayesian inverse problems with non-conjugate priors arXiv:1209.6156

[30] Rasmussen C E and Williams C K I 2006 *Gaussian Processes for Machine Learning* (Cambridge, MA: MIT Press)

[31] Robinson J C 2001 *Infinite-Dimensional Dynamical Systems (Cambridge Texts in Applied Mathematics)* (Cambridge: Cambridge University Press)

[32] Rogers C 2010 Least-action filtering http://www.statslab.cam.ac.uk/~chris/laf.pdf

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

[34] Zeitouni O 1989 On the Onsager–Machlup functional of diffusion processes around non $C^2$ curves *Ann. Probab.* **17** 1037–54

[35] Zeitouni O 2000 MAP estimation of diffusions: an updated account *Systems Theory: Modeling, Analysis and Control (Kluwer International Series in Engineering and Computer Science* vol 518) p 145