MAP ESTIMATORS AND POSTERIOR CONSISTENCY IN BAYESIAN NONPARAMETRIC INVERSE PROBLEMS

M. DASHTI,* K.J.H. LAW, A.M. STUART[†] AND J. VOSS[‡]

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 μ_0. We work under a natural set of conditions on the likelihood which imply the existence of a well-posed posterior measure, μ^y. Under these conditions we show that the *maximum a posteriori* (MAP) estimator is well-defined as the minimiser 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. 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. Consider a centred (mean-zero) Gaussian measure μ_0 on a Banach space $\big(X, \|\cdot\|_X\big)$, with Cameron-Martin space $\big(E, \langle\cdot,\cdot\rangle_E, \|\cdot\|_E\big)$, and a measure μ which has a density with respect to μ_0:

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

Measures μ of this form arise naturally in a number of applications, including the theory of conditioned diffusions [16] and the Bayesian approach to inverse problems [28]. In these settings there are many applications where $\Phi\colon X \to \mathbb{R}$ 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 μ, and in particular the *maximum a posteriori* 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 is finite dimensional and the prior μ_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 μ with respect to the Lebesgue measure, given by (1.1), is maximised at minimisers of

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

where $\|\cdot\|_E = |\mathcal{C}_0^{-1/2}u|$ is the Cameron-Martin norm of the Gaussian measure μ_0. We would like to derive such a result in the infinite dimensional setting. The main

*DEPARTMENT OF MATHEMATICS, UNIVERSITY OF SUSSEX, BRIGHTON BN1 9QH, UK

[†]MATHEMATICS INSTITUTE, UNIVERSITY OF WARWICK, COVENTRY, CV4 7AL, UK

[‡]SCHOOL OF MATHEMATICS, UNIVERSITY OF LEEDS, LEEDS, LS2 9JT, UK

technical difficulty that is encountered stems from the fact that the Cameron-Martin space has measure zero with respect to the distributions μ_0 and μ.

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 δ centred at $z \in X$. If there is a functional I, defined on E, which satisfies

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

for all $z_1, z_2 \in E \subset X$, then I is termed the *Onsager-Machlup* functional [11, 19]. 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 μ and is clearly given by minimisers of the Onsager-Machlup function. In finite dimensions it is clear that I given by (1.2) is the Onsager-Machlup functional. We will generalize this result to the infinite dimensional setting and show that the MAP estimators, which we define using centres of shrinking small balls with maximal probability, are characterised by minimisers of functional I.

When the probability measure μ 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 data 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 Φ and μ_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 minimiser of this functional;
- in section 4 we study the problem of Bayesian posterior consistency by studying limits of Onsager-Machlup minimisers in the small noise and large data 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 the functional I in (1.2) resembles a Tikhonov-Phillips regularization of the minimisation problem for Φ [12], with the Cameron-Martin norm of the prior determining the regularization. In the theory of classical non-statistical inversion, formulation via 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 μ 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 [23, 24, 27] 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). 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 [29]; see [30] for an overview. Our general setting is similar to that used to study the specific applications arising in the papers [9, 29, 30]. By working with small ball properties of Gaussian measures, and assuming that Φ has natural continuity properties, we are able to derive results in considerable generality. There is a recent related definition of MAP estimators in [17], with application to density estimation in [14]. However, whilst the goal of minimising I is also identified in [17], the proof in that paper is only valid in finite dimensions since it implicitly assumes that the Cameron-Martin norm is μ_0-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 [13]. The study of posterior consistency for the Bayesian approach to inverse problems is starting to receive considerable attention. The papers [20, 1] are devoted to obtaining rates of convergence for linear inverse problems with conjugate Gaussian priors, whilst the papers [4, 25] 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].

2. Set-up. Throughout this paper we assume that $\big(X, \|\cdot\|_X\big)$ is a separable Banach space and that μ_0 is a centred Gaussian (probability) measure on X with Cameron-Martin space $\big(E, \langle\cdot,\cdot\rangle_E, \|\cdot\|_E\big)$. The measure μ of interest is given by (1.1) and we make the following assumptions concerning the *potential* Φ.

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

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

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

(ii) Φ 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

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

(iii) Φ 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

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

From [28] we know that under Assumptions 2.1 the function $\exp(-\Phi)$ is integrable with respect to μ_0 and thus μ given by (1.1) can indeed be normalized to give a probability measure μ. Finally, we define a function $I\colon X \to \mathbb{R}$ 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 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 centred Gaussian measure μ_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 μ given by (1.1). Then Theorem 3.5 and Corollary 3.10, show that the MAP estimators, defined precisely in Definition 3.1, are characterised by the minimisers of the Onsager-Machlup functional.

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

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

be the mass of the ball centred around $z \in X$. We first define the MAP estimator for μ as follows:

DEFINITION 3.1. *Let*

$$z^\delta = \arg\max_{z \in X} J^\delta(z).$$

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 μ 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 minimises 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 minimisers of I characterise all MAP estimators of μ.

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 μ, i.e. for any $z_1, z_2 \in E$ we have*

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

Proof. Note that $J^\delta(z)$ is finite and positive for any $z \in E$ by Assumptions 2.1(i),(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 [21]:

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

This is the key estimate in the proof since it transfers questions about probability, naturally asked on the space X of full measure under μ_0, into statements concerning the Cameron-Martin norm of μ_0, which is almost surely infinite under μ_0.

We have

$$\begin{aligned}
\frac{J^\delta(z_1)}{J^\delta(z_2)} &= \frac{\int_{B^\delta(z_1)} \exp(-\Phi(u))\, \mu_0(\mathrm{d}u)}{\int_{B^\delta(z_2)} \exp(-\Phi(v))\, \mu_0(\mathrm{d}v)} \\
&= \frac{\int_{B^\delta(z_1)} \exp(-\Phi(u)+\Phi(z_1)) \exp(-\Phi(z_1))\, \mu_0(\mathrm{d}u)}{\int_{B^\delta(z_2)} \exp(-\Phi(v)+\Phi(z_2)) \exp(-\Phi(z_2))\, \mu_0(\mathrm{d}v)}.
\end{aligned}$$

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

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

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

$$\begin{aligned}
\frac{J^\delta(z_1)}{J^\delta(z_2)} &\le \mathrm{e}^{\delta(L_1+L_2)} \frac{\int_{B^\delta(z_1)} \exp(-\Phi(z_1))\, \mu_0(\mathrm{d}u)}{\int_{B^\delta(z_2)} \exp(-\Phi(z_2))\, \mu_0(\mathrm{d}v)} \\
&= \mathrm{e}^{\delta(L_1+L_2)} \mathrm{e}^{-\Phi(z_1)+\Phi(z_2)} \frac{\int_{B^\delta(z_1)} \;\mu_0(\mathrm{d}u)}{\int_{B^\delta(z_2)} \;\mu_0(\mathrm{d}v)}.
\end{aligned}$$

Now, by (3.1), we have

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

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

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

Similarly we obtain

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

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

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

Inequalities (3.2) and (3.3) give the desired result. □

We note that similar methods of analysis show the following:

Corollary 3.3. *Let the Assumptions of Theorem 3.2 hold. Then for any $z \in E$*

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

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

Proof. Noting that we consider μ 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

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

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 $\delta \to 0$. □

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

Proof. The existence of a minimiser of I in E, under the given assumptions, is proved as Theorem 5.4 in [28] (and as Theorem 2.7 in [7] in the case that Φ is non-negative). □

The rest of this section is devoted to a proof of the result that MAP estimators can be characterised as minimisers 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 $\Phi(u) \ge M$ for any $u \in X$.*

i) Let $z^\delta = \arg\max_{z\in X} J^\delta(z)$. 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 minimiser 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 characterising the MAP estimators as minimisers of Onsager-Machlup functional.

Lemma 3.6. *Let $\delta > 0$. For any centred Gaussian measure μ_0 on a separable Banach space X we have*

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

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

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

$$J_{0,n}^\delta(z) := \int_{B^\delta(z)} \mathrm{e}^{-\frac{1}{2}(a_1x_1^2+\cdots+a_nx_n^2)}\,\mathrm{d}x, \quad \text{for any } z \in \mathbb{R}^n, \tag{3.4}$$

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

$$\begin{aligned}\frac{J^\delta_{0,n}(z)}{J^\delta_{0,n}(0)} &= \frac{\int_{B^\delta(z)} \mathrm{e}^{-\frac{1}{2}\left(a_1x_1^2+\cdots+a_nx_n^2\right)}\,\mathrm{d}x}{\int_{B^\delta(0)} \mathrm{e}^{-\frac{1}{2}\left(a_1x_1^2+\cdots+a_nx_n^2\right)}\,\mathrm{d}x}\\ &< \frac{\mathrm{e}^{-\frac{1}{2}(a_1-\varepsilon)(|z|-\delta)^2}}{\mathrm{e}^{-\frac{1}{2}(a_1-\varepsilon)\delta^2}}\frac{\int_{B^\delta(z)} \mathrm{e}^{-\frac{1}{2}\left(\varepsilon x_1^2+(a_2-a_1+\varepsilon)x_2^2+\cdots+(a_n-a_1+\varepsilon)x_n^2\right)}\,\mathrm{d}x}{\int_{B^\delta(0)} \mathrm{e}^{-\frac{1}{2}\left(\varepsilon x_1^2+(a_2-a_1+\varepsilon)x_2^2+\cdots+(a_n-a_1+\varepsilon)x_n^2\right)}\,\mathrm{d}x}\\ &< c\,\mathrm{e}^{-\frac{1}{2}(a_1-\varepsilon)(|z|-\delta)^2}\frac{\int_{B^\delta(z)}\hat{\mu}_0(\mathrm{d}x)}{\int_{B^\delta(0)}\hat{\mu}_0(\mathrm{d}x)},\end{aligned}$$

for any $\varepsilon < a_1$ and where $\hat{\mu}_0$ is a centred Gaussian measure on $\mathbb{R}^n$ with the Covariance matrix $\mathrm{diag}[1/\varepsilon, 1/(a_2-a_1+\varepsilon), \cdots, 1/(a_n-a_1+\varepsilon)]$ (noting that $a_n \geq a_{n-1} \geq \cdots \geq a_1$). By Anderson's inequality $\hat{\mu}_0(B(z,\delta)) \leq \hat{\mu}_0(B(0,\delta))$ and therefore

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

and since ε 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 μ_0, separates the points in X, therefore $T: u \to \{e_j(u)\}_{j=1}^\infty$ is an injective map from X into $\mathbb{R}^\infty$. Let $u_j = e_j(u)$ and

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

Then, since μ_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 $A_z = P_n^{-1}(P_n(B^\delta(0))$ 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]. Let $z_j = (z, e_j)$ and $z^n = (z_1, z_2, \cdots, z_n, 0, \cdots)$ and for $0 < \varepsilon_1 < \delta/2$, $n > N$ large enough so that $\|z - z^n\|_X \leq \varepsilon_1$. With $\alpha = c\,\mathrm{e}^{-\frac{a_1}{2}(\|z\|_X-\varepsilon_1-\delta)^2}$ we have

$$\begin{aligned}J^\delta_0(z) &\leq J^\delta_{0,n}(z^n) + \varepsilon_0\\ &\leq \alpha J^\delta_{0,n}(0) + \varepsilon_0\\ &\leq \alpha J^\delta_0(0) + (1+\alpha)\varepsilon_0.\end{aligned}$$

Since ε_0 and ε_1 converge to zero as $n \to \infty$, the result follows. □

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

$$\frac{J^\delta_0(z^\delta)}{J^\delta_0(0)} < \varepsilon.$$

Proof. Let $\mathcal{C}$ be the covariance operator of μ_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 μ_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 $z^\delta \rightharpoonup \bar{z}$ 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 exists N sufficiently large and $\tilde{\delta} > 0$ sufficiently small such that

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

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 $\delta \leq \delta_1$ such that $\mathrm{e}^{-\frac{1}{4}(a_1 + \cdots + a_N)\delta^2} > 1/2$, for any $n \geq N$ we can write

$$\begin{aligned}
\frac{J_{0,n}^{\delta}(T_n z^{\delta})}{J_{0,n}^{\delta}(0)} &= \frac{\int_{B^{\delta}(T_n z^{\delta})} \mathrm{e}^{-\frac{1}{2}(a_1 x_1^2 + \cdots + a_n x_n^2)}\, \mathrm{d}x}{\int_{B^{\delta}(0)} \mathrm{e}^{-\frac{1}{2}\left(a_1 x_1^2 + \cdots + a_n x_n^2\right)}\, \mathrm{d}x} \\
&\leq \frac{\int_{B^{\delta}(T_n z^{\delta})} \mathrm{e}^{-\frac{1}{4}(a_1 x_1^2 + \cdots + a_N x_N^2)} \mathrm{e}^{-\frac{1}{2}(\frac{a_1}{2} x_1^2 + \cdots + \frac{a_N}{2} x_N^2 + a_{N+1} x_{N+1}^2 \cdots + a_n x_n^2)}\, \mathrm{d}x}{\int_{B^{\delta}(0)} \mathrm{e}^{-\frac{1}{4}(a_1 x_1^2 + \cdots + a_N x_N^2)} \mathrm{e}^{-\frac{1}{2}(\frac{a_1}{2} x_1^2 + \cdots + \frac{a_N}{2} x_N^2 + a_{N+1} x_{N+1}^2 \cdots + a_n x_n^2)}\, \mathrm{d}x} \\
&\leq \frac{\mathrm{e}^{-\frac{1}{4}A^2} \int_{B^{\delta}(T_n z^{\delta})} \mathrm{e}^{-\frac{1}{2}(\frac{a_1}{2} x_1^2 + \cdots + \frac{a_N}{2} x_N^2 + a_{N+1} x_{N+1}^2 \cdots + a_n x_n^2)}\, \mathrm{d}x}{\frac{1}{2} \int_{B^{\delta}(0)} \mathrm{e}^{-\frac{1}{2}(\frac{a_1}{2} x_1^2 + \cdots + \frac{a_N}{2} x_N^2 + a_{N+1} x_{N+1}^2 \cdots + a_n x_n^2)}\, \mathrm{d}x} \\
&\leq 2\mathrm{e}^{-\frac{1}{4}A^2}.
\end{aligned}$$

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

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^{δ} converges weakly and not strongly to 0 in X as $\delta \to 0$. Then for any $\varepsilon > 0$ there exists δ small enough such that*

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

Proof. Since z^{δ} converges weakly and not strongly to 0, we have

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

and therefore for δ_1 small enough there exists $\alpha > 0$ such that $\|z^\delta\|_X > \alpha$ for any $\delta < \delta_1$. Let λ_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 μ_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

$$\begin{aligned}
&\frac{J^\delta_{0,n}(T_n z^\delta)}{J^\delta_{0,n}(0)} \\
&= \frac{\int_{B^\delta(T_n z^\delta)} \mathrm{e}^{-\frac{1}{2}(a_1 x_1^2+\cdots+a_n x_n^2)}\,\mathrm{d}x}{\int_{B^\delta(0)} \mathrm{e}^{-\frac{1}{2}\left(a_1 x_1^2+\cdots+a_n x_n^2\right)}\,\mathrm{d}x} \\
&\le \frac{\int_{B^\delta(T_n z^\delta)} \mathrm{e}^{-\frac{A^2}{2}(x_{N+1}^2+\cdots+x_n^2)}\mathrm{e}^{-\frac{1}{2}(a_1 x_1^2+\cdots+a_N x_N^2+(a_{N+1}-A^2)x_{N+1}^2\cdots+(a_n-A^2)x_n^2)}\,\mathrm{d}x}{\int_{B^\delta(0)} \mathrm{e}^{-\frac{A^2}{2}(x_{N+1}^2+\cdots+x_n^2)}\mathrm{e}^{-\frac{1}{2}(a_1 x_1^2+\cdots+a_N x_N^2+(a_{N+1}-A^2)x_{N+1}^2\cdots+(a_n-A^2)x_n^2)}\,\mathrm{d}x} \\
&\le \frac{\mathrm{e}^{-\frac{1}{2}A^2(\frac{C\alpha}{2}-\delta^2)} \int_{B^\delta(T_n z^\delta)} \mathrm{e}^{-\frac{1}{2}(\frac{a_1}{2}x_1^2+\cdots+\frac{a_N}{2}x_N^2+a_{N+1}x_{N+1}^2\cdots+a_n x_n^2)}\,\mathrm{d}x}{\mathrm{e}^{-\frac{1}{2}A^2\delta^2} \int_{B^\delta(0)} \mathrm{e}^{-\frac{1}{2}(\frac{a_1}{2}x_1^2+\cdots+\frac{a_N}{2}x_N^2+a_{N+1}x_{N+1}^2\cdots+a_n x_n^2)}\,\mathrm{d}x} \\
&\le 2\mathrm{e}^{-\frac{C\alpha}{4}A^2},
\end{aligned}$$

if $\delta < \delta_2$ is small enough so that $\mathrm{e}^{A\delta^2} < 2$. Having this and arguing in a similar way to the final paragraph of 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 $\{z^\delta\}$ is bounded in X. By Assumption 2.1.(ii) for any $r > 0$ there exists $K = K(r) > 0$ such that

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

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

$$\max_{z\in E} \int_{B^\delta(z)} \mathrm{e}^{-\Phi(u)}\,\mu_0(\mathrm{d}u) \ge \int_{B^\delta(0)} \mathrm{e}^{-\Phi(u)}\,\mu_0(\mathrm{d}u) \ge \mathrm{e}^{-K(\delta)} \int_{B^\delta(0)} \mu_0(\mathrm{d}u).$$

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

$$\frac{J^\delta(z^\delta)}{\int_{B^\delta(0)} \mu_0(\mathrm{d}u)} \ge \frac{1}{Z}\mathrm{e}^{-K(1)} = \varepsilon_1 \tag{3.9}$$

noting that ε_1 is independent of δ.

We also can write

$$\begin{aligned} ZJ^\delta(z) &= \int_{B^\delta(z)} \mathrm{e}^{-\Phi(u)}\mu_0(\mathrm{d}u) \\ &\le \mathrm{e}^{-M}\int_{B^\delta(z)} \mu_0(\mathrm{d}u) \\ &=: \mathrm{e}^{-M} J_0^\delta(z), \end{aligned}$$

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

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

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

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

implying that for any δ_R and corresponding z^{δ_R}

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

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

$$\|z^\delta\|_X \le R, \qquad \text{for any } \delta < \delta_R.$$

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

$$B^\delta(z^\delta) \cap \big(B^\delta(0) + AU_E\big) = \emptyset. \tag{3.11}$$

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

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

Supposing $\bar{z} \notin E$, in Lemma 3.7 we show that for any $\varepsilon > 0$ there exists δ 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 $\mathrm{e}^{-A^2/2} < \frac{1}{2}\,\mathrm{e}^{K(1)}\mathrm{e}^{-M}$, and setting $\varepsilon = \mathrm{e}^{-A^2/2}$, from (3.12), we get $1 \le J^\delta(z^\delta)/J^\delta(0) < 1$ which is a contradiction. We therefore have $\bar{z} \in E$.

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

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

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

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

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

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 δ, $\|\bar{z}\|_E < \|z^\delta\|_E$. Therefore for the centered Gaussian measure μ_0, since $\|z^\delta - \bar{z}\|_X \to 0$ we have

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

This, since by definition of z^δ, $J^\delta(z^\delta) \ge J^\delta(\bar{z})$ and hence

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

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 [21] 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 μ.

It remains to show that $\bar z$ is a minimiser of I. Suppose $\bar z$ is not a minimiser of I so that $I(\bar z)-I(z^*)>0$. Let δ_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

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

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

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

which is a contradiction, since by definition of z^δ, $J^\delta(z^\delta)\ge J^\delta(z^*)$ 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, minimises the Onsager-Machlup functional I.

ii) Any $z^*\in E$ which minimises the Onsager-Machlup functional I, is a MAP estimator for measure μ 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 minimiser of I would result in a contradiction using an argument similar to 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 minimiser 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 μ_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 data 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, and then study posterior consistency of MAP estimators in both the small noise and large data limits, as described 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 data limits.

4.1. Inverse Problems. Consider the problem of estimating a function u in a Banach space X, from a given vector $y \in \mathbb{R}^J$, where

$$y = G(u) + \zeta; \tag{4.1}$$

here $G\colon X \to \mathbb{R}^J$ is a possibly nonlinear operator, and ζ is a realization of an $\mathbb{R}^J$-valued centred Gaussian random variable with known covariance Σ. A prior probability measure $\mu_0(\mathrm{d}u)$ is put on u, and the distribution of $y|u$ is given by (4.1), with ζ assumed independent of u. In this paper we restrict ourselves to centred Gaussian prior measures, and denote the covariance operator of μ_0 by $\mathcal{C}_0$: $\mu_0 = \mathcal{N}(0, \mathcal{C}_0)$. Then, under appropriate conditions on μ_0 and G, Bayes theorem is interpreted as giving the following formula for the Radon-Nikodym derivative of the posterior distribution μ^y on $u|y$ with respect to μ_0:

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(u) \propto \exp\bigl(-\Phi(u;y)\bigr), \tag{4.2}$$

where

$$\Phi(u;y) = \frac{1}{2}\Bigl|\Sigma^{-\frac{1}{2}}\bigl(y - G(u)\bigr)\Bigr|^2. \tag{4.3}$$

Derivation of Bayes formula (4.2) for problems with finite dimensional data, and ζ 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 Φ given by (4.3) which depend on the observed data y. In this section we study MAP estimators for problems of the form (4.2). It is convenient to extend the Onsager-Machlup functional I to a mapping from $X \times \mathbb{R}^J$ to $\mathbb{R}$, defined as

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

We study properties of minimisers of the functional, in both the small noise and in the large data limits. We assume that the data is found from application of G to the truth $u^\dagger$ with additional noise:

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

Theorems 4.1 and 4.4 show that weak limits of MAP estimators agree, when mapped by G, with the image of the true function under G.

4.2. Large Data Limit. Let us denote the exact solution by $u^\dagger$ and suppose that as data we have the following n random vectors

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

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 $J = nK$, $G(\cdot) = \big(\mathcal{G}(\cdot), \cdots, \mathcal{G}(\cdot)\big)$ and Σ 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:

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

Here, and in the following, we use the notation $\langle \cdot, \cdot \rangle_{\mathcal{C}_1} = \left\langle \mathcal{C}_1^{-1/2} \cdot, \mathcal{C}_1^{-1/2} \cdot \right\rangle$, and $|\cdot|^2_{\mathcal{C}_1} = \langle \cdot, \cdot \rangle_{\mathcal{C}_1}$: By Corollary 3.10 MAP estimators for this problem are minimisers of

$$I_n := \|u\|_E^2 + \sum_{j=1}^{n} |y_j - \mathcal{G}(u)|^2_{\mathcal{C}_1}. \tag{4.4}$$

Our interest is in studying properties of the limits of minimisers 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 $\mathcal{G}\colon X \to \mathbb{R}^K$ is Lipschitz on bounded sets and $u^\dagger \in E$. For every $n \in \mathbb{N}$, let $u_n \in E$ be a minimiser of I_n given by (4.4). Then there exists 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 Φ. We note that

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

Hence

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

Define

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

We have

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

LEMMA 4.2. *Assume that $\mathcal{G}\colon X \to \mathbb{R}^K$ 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 realisations $\eta_1, \dots, \eta_n$ there exists an $r > 0$ such that $\max\{|y_1|, \dots, |y_n|\} < r$, Assumptions 2.1 hold for Φ. Since $\arg\min_u I_n = \arg\min_u J_n$ the result follows by Proposition 3.4. ☐

We may now prove the posterior consistency theorem. From (4.6) onwards the proof is an adaptation of the proof of Theorem 2 of [3].

Proof. *(of Theorem 4.1)* By definition of u_n we have

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

Therefore

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

Using Young's inequality for the last term in the right-hand side we get

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

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

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

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

$$\mathbb{E}|\mathcal{G}(u^\dagger) - \mathcal{G}(u_n)|^2_{\mathcal{C}_1} \to 0 \quad \text{as} \quad n \to \infty \tag{4.5}$$

and

$$\mathbb{E}\|u_n\|^2_E \le \|u^\dagger\|^2_E + 2K. \tag{4.6}$$

1) We first show using (4.6) that there exist $u^* \in E$ and a subsequence $\{u_{n_k(k)}\}_{k\in\mathbb{N}}$ of $\{u_n\}$ such that

$$\mathbb{E}\langle u_{n_k(k)}, v\rangle_E \to \mathbb{E}\langle u^*, v\rangle_E, \quad \text{for any } v \in E. \tag{4.7}$$

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

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

Therefore there exists $\xi_1 \in \mathbb{R}$ 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 exists $\xi_2 \in \mathbb{R}$ 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 $\{u_{n_1(k)}\}_{k\in\mathbb{N}} \supset \{u_{n_2(k)}\}_{k\in\mathbb{N}} \supset \cdots \supset \{u_{n_j(k)}\}_{k\in\mathbb{N}}$ such that $\mathbb{E}\langle u_{n_j(k)}, \phi_j\rangle \to \xi_j$ for any $j \in \mathbb{N}$ and as $k \to \infty$. Therefore

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

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

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

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 $v \in E$

$$\begin{aligned}\mathbb{E}\langle u_{n_k(k)} - u^*, v\rangle_E &= \mathbb{E}\sum_{j=1}^\infty \langle v, \phi_j\rangle_E \langle u_{n_k(k)} - u^*, \phi_j\rangle_E \\ &\le N\|v\|_E\; \mathbb{E}\sup_{j\in\{1,\dots,N\}} |\langle u_{n_k(k)} - u^*, \phi_j\rangle_E| + (\|u^\dagger\|_E^2 + 2K)^{1/2}\sum_{j=N}^\infty |\langle v, \phi_j\rangle_E|\end{aligned}$$

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

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

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

2) Now we show almost sure existence of a convergent subsequence of $\{u_{n_k(k)}\}$. By (4.5) 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

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

Now by (4.7) 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

$$u_{\hat{m}(k)} \rightharpoonup u^* \quad \text{in } \; E \quad \text{a.s.} \quad \text{as } k \to \infty.$$

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 $\mathcal{G}$. □

In the case that $u^\dagger \in X$ (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 density of E in X, there exists $v \in E$ such that $\|u^\dagger - v\|_X \le \varepsilon$. Then by definition of u_n we can write

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

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

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

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

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

implying that

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

Since the lim inf is obviously positive and ε 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. □

4.3. Small Noise Limit. Consider the case where as data we have the random vector

$$
y_n = \mathcal{G}(u^\dagger) + \frac{1}{n}\eta_n, \tag{4.8}
$$

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 this data and given the prior measure μ_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 minimisers of

$$I_n(u) := \|u\|_E^2 + n^2 |y_n - \mathcal{G}(u)|_{\mathcal{C}_1}^2. \tag{4.9}$$

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

THEOREM 4.4. *Assume that $\mathcal{G}\colon X \to \mathbb{R}^K$ is Lipschitz on bounded sets, and $u^\dagger \in E$. For every $n \in \mathbb{N}$, let $u_n \in E$ be a minimiser of $I_n(u)$ given by (4.9). Then there exists 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

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

Letting

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

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 data case. The result of Theorem 4.4 carries over as well. Indeed, by definition of u_n, we have

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

Therefore

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

Using Young's inequality for the last term in the right-hand side we get

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

Taking expectation we obtain

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

This implies that

$$\mathbb{E}|\mathcal{G}(u^\dagger) - \mathcal{G}(u_n)|_{\mathcal{C}_1}^2 \to 0 \quad \text{as} \quad n \to \infty \tag{4.10}$$

and

$$\mathbb{E}\|u_n\|_E^2 \le \|u^\dagger\|_E^2 + 2K. \tag{4.11}$$

Having (4.10) and (4.11), and with the same argument as the proof of Theorem 4.1, it follows that there exists 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 data 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 periodic boundary conditions:

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

Here $v\colon \mathbb{T}^2 \times (0,\infty) \to \mathbb{R}^2$ is a time-dependent vector field representing the velocity, $p\colon \mathbb{T}^2 \times (0,\infty) \to \mathbb{R}$ is a time-dependent scalar field representing the pressure, $f\colon \mathbb{T}^2 \to \mathbb{R}^2$ is a vector field representing the forcing (which we assume to be time-independent for simplicity), and ν 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

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

and H as the closure of $\mathcal{H}$ with respect to the $(L^2(\mathbb{T}^2))^2$ norm. We define the map $P\colon (L^2(\mathbb{T}^2))^2 \to H$ to be the Leray-Helmholtz orthogonal projector (see [26]). Given $k = (k_1,k_2)^{\mathrm{T}}$, define $k^\perp := (k_2, -k_1)^{\mathrm{T}}$. Then an orthonormal basis for H is given by $\psi_k\colon \mathbb{R}^2 \to \mathbb{R}^2$, where

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

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

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

with the norm $\|u\|_s := \big(\sum_k (\pi^2|k|^2)^s |u_k|^2\big)^{1/2}$, 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 $\ell = 0, \dots, L$, and define $v_\ell \in \mathbb{R}^M$ 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 point in $\mathbb{T}^2$ with cardinality $M/2$. Note that each v_ℓ depends on u and we may define $\mathcal{G}_\ell \colon H \to \mathbb{R}^M$ by $\mathcal{G}_\ell(u) = v_\ell$. We let $\{\eta_\ell\}_{\ell\in\{1,\dots,L\}}$ be a set of random variables in $\mathbb{R}^M$ which perturbs the points $\{v_\ell\}_{\ell\in\{1,\dots,L\}}$ to generate the observations $\{y_\ell\}_{\ell\in\{1,\dots,L\}}$ in $\mathbb{R}^M$ given by

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

We let $y = \{y_\ell\}_{\ell=1}^L$, the accumulated data up to time $T = Lh$, with similar notation for η, and define $\mathcal{G} \colon H \to \mathbb{R}^{ML}$ by $\mathcal{G}(u) = \big(\mathcal{G}_1(u), \dots, \mathcal{G}_L(u)\big)$. 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 $\{\eta_\ell\}_{\ell\in\{1,\dots,L\}}$ is i.i.d. in $\mathbb{R}^M$, independent of u, with η_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 μ^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

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

We are in the setting of subsection 4.3, with $\gamma = 1/n$ and $K = ML$. In the applied literature approaches to assimilating data into mathematical models based on minimising I_{NS} are known as *variational methods*, and sometimes as 4DVAR [2].

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 ν. For the particular forcing used there is an explicit steady state for all $\nu > 0$ and for $\nu \geq 0.035$ this solution is stable (see [22], Chapter 2 for details). As ν 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 is generated by computing a true signal solving the Navier-Stokes equation at the desired value of ν, and then adding Gaussian random noise to it at each observation time. Furthermore, we

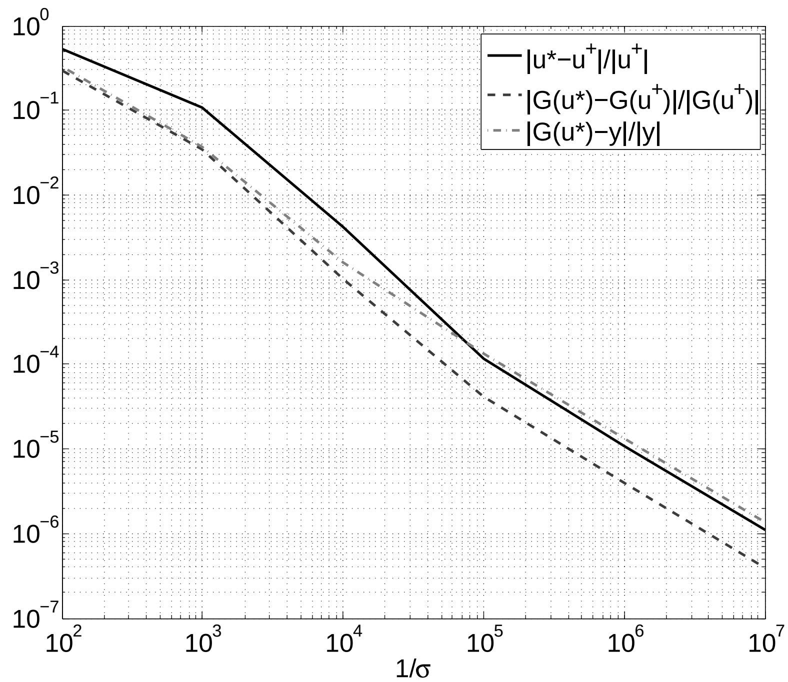

FIG. 5.1. *Illustration of posterior consistency in the fluid mechanics application. The three curves given are the relative error of the MAP estimator u^* in reproducing the truth, $u^\dagger$ (solid), the relative error of the map $\mathcal{G}(u^*)$ in reproducing $\mathcal{G}(u^\dagger)$ (dashed), and the relative error of $\mathcal{G}(u^*)$ with respect to the observations y (dash-dotted).*

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 results in terms of σ.

Figure 5.1 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 red 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 pink curve 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 minimiser. 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 ν denote the measure on $X := C\big([0,T];\mathbb{R}\big)$ generated by the stochastic differential equation (SDE) given in (6.1), and ν_0 the same measure obtained in the case $f \equiv 0$, then the Girsanov theorem states that $\nu \ll \nu_0$ with density

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

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

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

and using this expression to remove the stochastic integral we obtain

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

Thus, the measure ν has a density with respect to the Gaussian measure ν_0 and (6.2) takes the form (1.1) with $\mu = \nu$ and $\mu_0 = \nu_0$: we have

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

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

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

and

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

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 $\Psi(u) \geq M$ for all $u \in \mathbb{R}$ and $F(u) \leq M$ for all $u \in \mathbb{R}$.*

Under these assumptions, we see that Φ_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 ν_0 is

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

and the Cameron-Martin norm is given by

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

where

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

The mean of ν_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 $I_1 \colon E_1 \to \mathbb{R}$ 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

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

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

We note that the regularity of minimisers 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 μ. Let μ_0 denote the Brownian bridge measure obtained in the case $f \equiv 0$. By applying the approach to determining bridge diffusion measures in [15] we obtain, from (6.2), the expression

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

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

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

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

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

and the Cameron-Martin norm is again $\sigma^{-1}\|\cdot\|_{H^1}$. The Onsager-Machlup function for the unconditioned diffusion (6.1) is thus $I_2 \colon E_2' \to \mathbb{R}$ 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 μ_0 and

$$E_2' = \Big\{ v \in H^1[0,T] \;\Big|\; \int_0^T \big|v'(s)\big|^2 \, ds < \infty \text{ and } v(0) = u^-, u(T) = u^+ \Big\}.$$

The MAP estimators for μ are found by minimising I_2 over E_2'.

6.3. Filtering and Smoothing. We now consider conditioning the measure ν 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 η_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 $\mathcal{G} \colon X \to \mathbb{R}^J$ is defined by

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

Recall ν_0 and ν 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 ν_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

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

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

$$\inf_{\|u\|_X \le 1} \exp\Big(-\int_0^T \Psi\big(u(t)\big)\,dt + \frac{1}{\sigma^2}F\big(u(T)\big) - \frac{1}{2\gamma^2}\sum_{j=1}^J |y_j - u(t_j)|^2\Big) > 0$$

and hence

$$\int_{\|u\|_X \le 1} \exp\Big(-\int_0^T \Psi\big(u(t)\big)\,dt + \frac{1}{\sigma^2}F\big(u(T)\big) - \frac{1}{2\gamma^2}\sum_{j=1}^J |y_j - u(t_j)|^2\Big)\,\nu_0(du) > 0.$$

Applying the conditioning Lemma 5.3 in [15] then gives

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

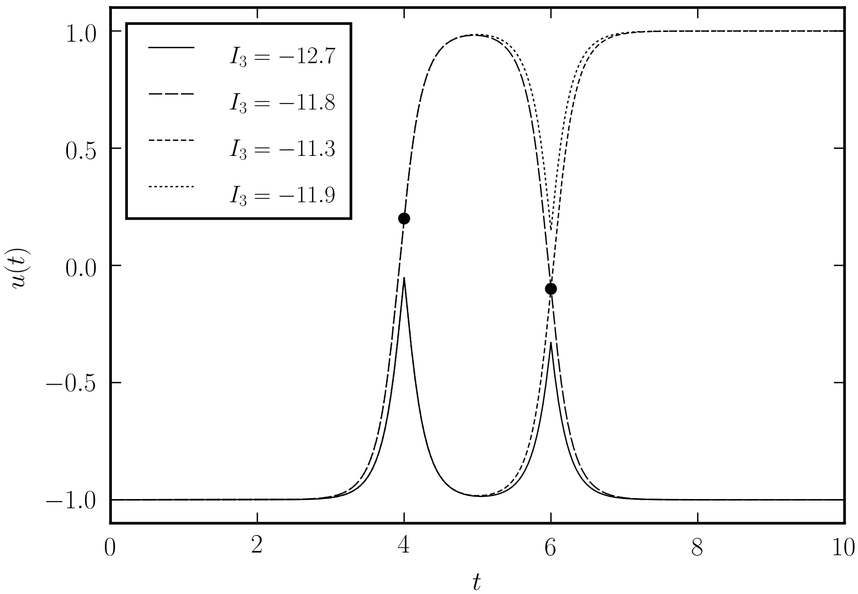

FIG. 6.1. *Illustration of the problem of local minima of I for the smoothing problem with a small number of observations. The process $u(t)$ starts at $u(0) = -1$ and moves in a double-well potential with stable equilibrium points at -1 and $+1$. Two observations of the process are indicated by the two black circles. The curves correspond to four different local minima of the functional I_3 for this situation.*

Thus we define

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

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

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

The MAP estimator for this setup is, again, found by minimising the Onsager-Machlup functional I_3.

The only difference between the potentials Φ_1 and Φ_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, minimisers of I_3 have jumps in their first derivates at $t_1, \ldots, t_J$. This effect can be clearly seen in the local minima of I_3 shown in figure 6.1.

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.

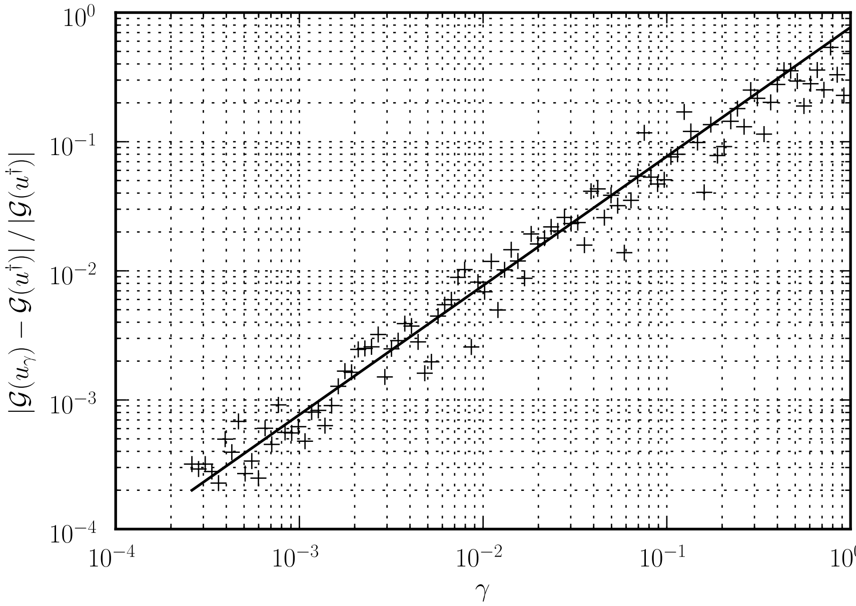

FIG. 6.2. *Illustration of posterior consistency for the smoothing problem in the small-noise limit. The marked points correspond the maximum-norm distance between the true signal $u^\dagger$ and the MAP estimator u_γ with $J = 5$ evenly spaced observations. The map $\mathcal{G}(u) = \big(u(t_1), \dots, u(t_J)\big)$ is the projection of the path onto the observation points. The solid line is a fitted curve of the form $c\gamma$.*

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 $y_1, \dots, y_J$ using (6.5). Then we implement the Onsager-Machlup functional I_3 from equation (6.6) and use numerical minimisation, employing the Broyden-Fletcher-Goldfarb-Shanno method, to find the minima of I_3.

The first experiment concerns the problem of local minima of I_3. For small number of observations we find multiple local minima; the minimisation procedure can converge to different *local* minima, depending on the starting point of the optimisation. This effect makes it difficult to find the MAP estimator, which is the *global* minimum of I_3, numerically. The problem is illustrated in figure 6.1, which shows four different local minima for the case of $J = 2$ observations. 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 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 γ^2 of the observational noise η_j converge to 0. Noting that the exact path of the SDE, denoted by $u^\dagger$ in (4.8), has the regularity of a

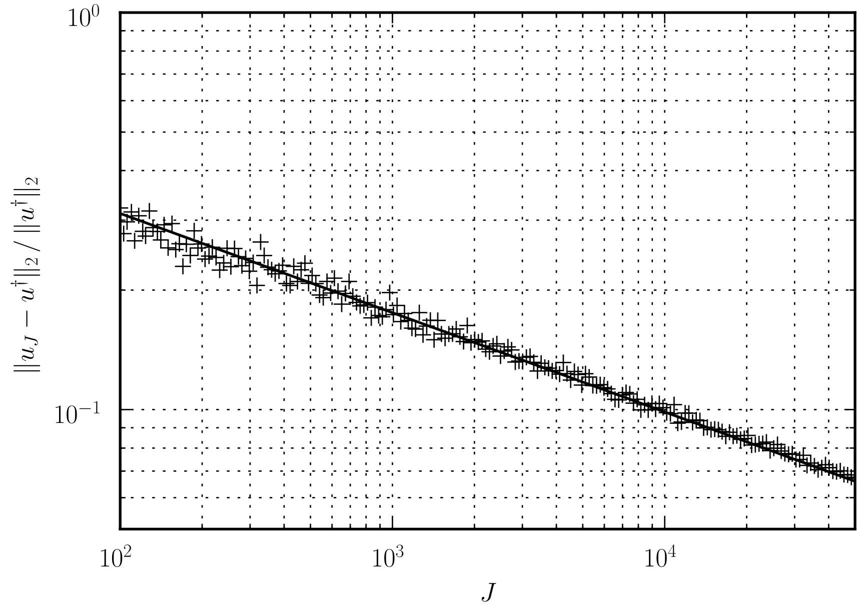

FIG. 6.3. *Illustration of posterior consistency for the smoothing problem in the large-data limit. The marked points correspond the supremum-norm distance between the true signal u^* and the MAP estimator $u_J^\dagger$ with J evenly spaced observations. The solid line give a fitted curve of the form $cJ^{-\alpha}$; the exponent $\alpha = -1/4$ was found numerically.*

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_γ denotes the MAP estimator corresponding to observational variance γ^2, confirming the result of Corollary 4.5. The result of a simulation with $J = 5$ is shown in figure 6.2.

Finally, we perform an experiment to illustrate posterior consistency in the large-data 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 $u^\dagger$. The situation considered here is not covered by the theoretical results from section 4, but the results of the numerical experiment, shown in figure 6.3 indicate that posterior consistency still holds.

REFERENCES

[1] S. Agapiou, S. Larsson and A.M. Stuart. *Posterior Consistency of the Bayesian Approach to Linear Ill-Posed Inverse Problems.* `arxiv.org/abs/1210.1563`

[2] A.F. Bennett. *Inverse Modeling of the Ocean and Atmosphere.* Cambridge University Press, 2002.

[3] Bissantz N., Hohage T. and Munk A. 2004, *Consistency and rates of convergence of nonlinear Tikhonov regularization with random noise. Inverse Problems* 20, no. 6, 1773–1789.

[4] Bochkina, N.A. *Consistency of the posterior distribution in generalised linear inverse problems.* `arxiv.org/abs/1211.3382`, 2012.

[5] Bogachev, V. I. *Gaussian measures.* Mathematical Surveys and Monographs, 62. American Mathematical Society, Providence, RI, 1998.

[6] Brown, L.D. and Low, M.G. *Asymptotic equivalence of nonparametric regression and white noise.* The Annals of Statistics **24**(1996), 2384–2398.

[7] S. L. Cotter, M. Dashti, J. C. Robinson, and A. M. Stuart. Bayesian inverse problems for functions and applications to fluid mechanics. *Inverse Problems*, 25(11):115008, 43, 2009.

[8] SM Cox and PC Matthews. Exponential time differencing for stiff systems. *Journal of Computational Physics*, 176(2):430–455, 2002.

[9] Dembo A. and Zeitouni O., *Onsager-Machlup functionals and maximum a posteriori estimation for a class of non-Gaussian random fields.* J. Mutivariate Analysis **36**(1991), 243–262.

[10] Da Prato G. & Zabczyk J. 1992, *Stochastic equations in infinite dimensions.* Encyclopedia of Mathematics and its Applications, 44. Cambridge University Press, Cambridge.

[11] D. Dürr and A. Bach. The Onsager-Machlup function as lagrangian for the most probable path of a diffusion process. *Communications in Mathematical Physics*, 160:153–170, 1978.

[12] Engl, H. W., Hanke, M. and Neubauer, A. 1996, *Regularization of inverse problems.* Mathematics and its Applications, 375. Kluwer Academic Publishers Group, Dordrecht.

[13] Ghosal, S. and Ghosh, J.K. and Ramamoorthi, RV, *Consistency issues in Bayesian nonparametrics. STATISTICS TEXTBOOKS AND MONOGRAPHS*, **158**(1999), 639–668.

[14] Griebel, M. and Hegland, M. A Finite Element Method for Density Estimation with Gaussian Process Priors. *SIAM Journal on Numerical Analysis*, 47:4759–4792, 2010.

[15] M. Hairer, A. M. Stuart, and J. Voss. Analysis of SPDEs arising in path sampling, part II: The nonlinear case. *Annals of Applied Probability*, 17:1657–1706, 2007.

[16] M. Hairer, A. M. Stuart, and J. Voss. Signal processing problems on function space: Bayesian formulation, stochastic pdes and effective mcmc methods. *The Oxford Handbook of Nonlinear Filtering, Editors D. Crisan and B. Rozovsky.* Oxford University Press, 2011.

[17] M. Hegland. Approximate maximum a posteriori with Gaussian process priors. *Constructive Approximation*, 26:205–224, 2007.

[18] J.S. Hesthaven, S. Gottlieb, and D. Gottlieb. *Spectral Methods for Time-Dependent Problems*, volume 21. Cambridge Univ Pr, 2007.

[19] Nobuyuki Ikeda and Shinzo Watanabe. *Stochastic differential equations and diffusion processes.* North-Holland Publishing Co., Amsterdam, second edition, 1989.

[20] Knapik, BT and van der Vaart, AW and van Zanten, JH, *Bayesian inverse problems with Gaussian priors*, Ann. Stat. **39**(2011), 2626–2657.

[21] Lifshits M. A. 1995 *Gaussian Random Functions*, Kluwer Academic Pub.

[22] A. Majda and X. Wang. *Non-linear dynamics and statistical theories for basic geophysical flows.* Cambridge Univ Pr, 2006.

[23] Markussen, B. *Laplace approximation of transition densities posed as Brownian expectations.* Stochastic Processes and their Applications **119**(2009), p 208–231.

[24] Mortensen, R.E. *Maximum-likelihood recursive filtering.* Journal of Optimization Theory and Applications **2**(1968), 386–393.

[25] K. Ray. *Bayesian inverse problems with non-conjugate priors.* `arXiv:1209.6156`, 2012.

[26] J. C. Robinson, Infinite-Dimensional Dynamical Systems, Cambridge Texts in Applied Mathematics, Cambridge University Press, Cambridge, 2001.

[27] Rogers, C. *Least-action filtering.* `http://www.statslab.cam.ac.uk/ chris/papers/LAF.pdf`

[28] Stuart A. M. 2010, Inverse Problems: A Bayesian Approach. *Acta Numerica* **19**.

[29] Zeitouni O. 1989, *On the Onsager-Machlup Functional of Diffusion Processes Around Non C2 Curves.* Annals of Probability **17**(1989), pp. 1037–1054.

[30] Zeitouni O. 2000, *MAP estimators of diffusions: an updated account.* Systems Theory: Modeling, Analysis and Control, **518**(2000), page 145.