[81]

S.L. Cotter, M. Dashti and A.M. Stuart,

Approximation of Bayesian Inverse Problems.

SIAM Journal of Numerical Analysis 48

(2010) 322-345.

APPROXIMATION OF BAYESIAN
INVERSE PROBLEMS FOR PDES[*]

S. L. COTTER[†], M. DASHTI[†], AND A. M. STUART[†]

Abstract. Inverse problems are often ill posed, with solutions that depend sensitively on data. In any numerical approach to the solution of such problems, regularization of some form is needed to counteract the resulting instability. This paper is based on an approach to regularization, employing a Bayesian formulation of the problem, which leads to a notion of well posedness for inverse problems, at the level of probability measures. The stability which results from this well posedness may be used as the basis for quantifying the approximation, in finite dimensional spaces, of inverse problems for functions. This paper contains a theory which utilizes this stability property to estimate the distance between the true and approximate posterior distributions, in the Hellinger metric, in terms of error estimates for approximation of the underlying forward problem. This is potentially useful as it allows for the transfer of estimates from the numerical analysis of forward problems into estimates for the solution of the related inverse problem. It is noteworthy that, when the prior is a Gaussian random field model, controlling differences in the Hellinger metric leads to control on the differences between expected values of polynomially bounded functions and operators, including the mean and covariance operator. The ideas are applied to some non-Gaussian inverse problems where the goal is determination of the initial condition for the Stokes or Navier–Stokes equation from Lagrangian and Eulerian observations, respectively.

Key words. inverse problem, Bayesian, Stokes flow, data assimilation, Markov chain–Monte Carlo

AMS subject classifications. 35K99, 65C50, 65M32

DOI. 10.1137/090770734

1. Introduction. In applications it is frequently of interest to solve *inverse problems* [15, 28, 32]: to find u, an input to a mathematical model, given y an observation of (some components of, or functions of) the solution of the model. We have an equation of the form

$$(1.1) \qquad y = \mathcal{G}(u)$$

to solve for $u \in X$, given $y \in Y$, where X and Y are Banach spaces. We refer to evaluating $\mathcal{G}$ as solving the *forward problem*.[1] We refer to y as *data* or *observations*. It is typical of inverse problems that they are *ill posed*: there may be no solution, or the solution may not be unique and may depend sensitively on y. For this reason some form of regularization is often employed [8] to stabilize computational approximations.

We adopt a Bayesian approach to regularization [4, 12] which leads to the notion of finding a *probability measure* μ on X, containing information about the relative probability of different states u, given the data y. Adopting this approach is natural in situations where an analysis of the source of data reveals that the observations y

---

[†]Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK (s.l.cotter@warwick.ac.uk, M.Dashti@warwick.ac.uk, A.M.Stuart@warwick.ac.uk).

[1]In the applications in this paper $\mathcal{G}$ is found from composition of the forward model with some form of observation operator, such as pointwise evaluation at a finite set of points. The resulting observation operator is often denoted with the letter $\mathcal{H}$ in the atmospheric sciences community [11]; because we need $\mathcal{H}$ for Hilbert space later on, we use the symbol $\mathcal{G}$.

are subject to noise. A more appropriate model equation is then often of the form

$$(1.2) \qquad y = \mathcal{G}(u) + \eta,$$

where η is a mean-zero random variable, whose statistical properties we might know, or make a reasonable mathematical model for, but whose actual value is unknown to us; we refer to η as the *observational noise*. We assume that it is possible to describe our prior knowledge about u, before acquiring data, in terms of a *prior probability measure* μ_0. It is then possible to use Bayes's formula to calculate the *posterior probability measure* μ for u given y.

In the infinite dimensional setting the most natural version of Bayes theorem is a statement that the posterior measure is absolutely continuous with respect to the prior [27] and that the Radon–Nikodým derivative (density) between them is determined by the data likelihood. This gives rise to the formula

$$(1.3) \qquad \frac{d\mu}{d\mu_0}(u) = \frac{1}{Z(y)} \exp\bigl(-\Phi(u; y)\bigr),$$

where the *normalization constant* $Z(y)$ is chosen so that μ is a probability measure:

$$(1.4) \qquad Z(y) = \int_X \exp\bigl(-\Phi(u; y)\bigr) d\mu_0(u).$$

In the case where y is finite dimensional and η has Lebesgue density ρ this is simply

$$(1.5) \qquad \frac{d\mu}{d\mu_0}(u) \propto \rho(y - \mathcal{G}(u)).$$

More generally Φ is determined by the distribution of y given u. We call $\Phi(u; y)$ the *potential* and sometimes, for brevity, refer to the evaluation of $\Phi(u; y)$ for a particular $u \in X$ as *solving the forward problem* as it is defined through $\mathcal{G}(\cdot)$. Note that the solution to the inverse problem is a probability measure μ which is defined through a combination of solution of the forward problem Φ, the data y, and a prior probability measure μ_0. Bayesian and classical regularization are linked via the fact that the minimizer of the Tikhonov-regularized nonlinear least squares problem coincides with MAP estimators, which maximize the posterior probability—see section 5.3 of [27].

In general it is hard to obtain information from a formula such as (1.3) for a probability measure. One useful approach to extracting information is to use *sampling*: generate a set of points $\{u^{(k)}\}_{k=1}^K$ distributed (perhaps only approximately) according to μ. In this context it is noteworthy that the integral $Z(y)$ appearing in formula (1.3) is not needed to enable implementation of MCMC (Markov chain–Monte Carlo) methods to sample from the desired measure. These methods incur an error which is well understood and which decays as $\sqrt{K}$ [17]. However, for inverse problems on function space there is a second source of error, arising from the need to approximate the inverse problem in a finite dimensional subspace of dimension N. The purpose of this paper is to quantify such approximation errors. The *key analytical idea* is that we transfer approximation properties of the forward problem Φ into approximation properties of the inverse problem defined by (1.3). Since the solution to the Bayesian inverse problem is a probability measure, we will need to use metrics on probability measures to quantify the effect of approximation. We will employ the Hellinger metric because this leads directly to bounds on the approximation error incurred when calculating the expectation of functions. The main general results concerning approximation properties are Theorems 2.2 and 2.4, together with Corollary 2.3. The *key*

practical implication of the theoretical developments is that it allows the practitioner to apportion computational resources so as to balance the Monte Carlo error and the error from finite dimensional approximation.

The main Theorem 2.2 is proved by application of the general "consistency plus stability implies convergence" approach. By this, we mean that convergence of the approximation of the forward problem (a form of consistency), along with well posedness of the inverse problem, implies convergence of the posterior measure. Defining well posedness is, of course, a serious obstacle in most inverse problem formulations. However, in [5] it was demonstrated that, at the level of probability measures, there is a useful notion of well posedness, and this was used to prove that the posterior measure is Lipschitz in the data. Here we develop these ideas to encompass the more complex perturbations arising from finite dimensional approximation. We also extend the framework of [5], which allows only for finite data sets, to include potential functions Φ which are not bounded below, hence allowing for data which is a function. Examples of such inverse problems, and the Lipschitz continuity of the posterior with respect to the data, are developed in [27], and we build on this framework here.

In section 2 we provide the general approximation theory, for measures μ given by (1.3), upon which the following two sections build. In section 3 we study the inverse problem of determining the initial condition for the Stokes equation, given a finite set of observations of Lagrangian trajectories defined through the time-dependent velocity field solving the Stokes equation. This section also includes numerical results showing the convergence of the posterior distribution under refinement of the finite dimensional approximation, as predicted by the theory. Section 4 is devoted to the related inverse problem of determining the initial condition for the Navier–Stokes equation, given direct observation of the time-dependent velocity field at a finite set of points at positive times.

The underlying philosophy in this paper is to formulate the Bayesian inverse problem on function space, and then approximate it. This philosophy forms the cornerstone of the review article [27] and is also central to the book [28] where linear Gaussian problems are studied from this viewpoint. Indeed, for such problems the idea appeared four decades ago in [9]. An alternative philosophy is to discretize the forward problem and then apply the Bayesian approach to the resulting inverse problems in finite dimensions [14], often with some attempt to model the affect of discretization error statistically. This approach has been very successful in practice and is readily adapted to a variety of scenarios [2, 13, 21]. There has also been some work on finite dimensional linear inverse problems, using the Bayesian approach to regularization, and considering infinite dimensional limits [10, 18]. Another motivation for discretizing the forward problem before applying the Bayesian formulation is that it easily allows for white noise priors (identity covariance operators) providing a direct link to the basic form of Tikhonov regularization used in many inverse problems [33]. In function space the Bayesian version of this setup leads to technical obstacles because the identity is not trace class in the natural Hilbert space, and hence the resulting probability measure is supported on (for example) spaces of distributions and not on the Hilbert space itself [16]. Nonetheless there are significant advantages to formulating inverse problems on function space, beyond the clarity of the mathematical formulation, including the possibility of designing sampling methods which do not degenerate under mesh refinement, as demonstrated in [6], and the ability to transfer approximation results from classical numerical analysis of forward problems into the study of inverse problems, as demonstrated in this article. Considerable practical

experience will need to be gathered in order to fully evaluate the relative merits of discretizing before or after formulation of the Bayesian inverse problem.

Overviews of inverse problems arising in fluid mechanics, such as those studied in sections 3 and 4, may be found in [19]. The connection between the Tikhonov-regularized least squares and Bayesian approaches to inverse problems in fluid mechanics is overviewed in [1].

2. General framework. In this section we establish three useful results which concern the effect of approximation on the posterior probability measure μ given by (1.3). These three results are Theorem 2.2, Corollary 2.3, and Theorem 2.4. The key point to notice about these results is that they simply require the proof of various bounds and approximation properties *for the forward problem*, and yet they yield approximation results concerning the Bayesian inverse problem. The connection to probability comes only through the choice of the space X, in which the bounds and approximation properties must be proved, which must have full measure under the prior μ_0.

The probability measure of interest (1.3) is defined through a density with respect to a prior reference measure μ_0 which, by shift of origin, we take to have mean zero. Furthermore, we assume that this reference measure is Gaussian with covariance operator $\mathcal{C}$. We write $\mu_0 = \mathcal{N}(0, \mathcal{C})$. In fact, the only consequence of the Gaussian prior assumption that we use is the Fernique Theorem A.3. Hence the results may be trivially extended to all measures which satisfy the conclusion of this theorem. The Fernique theorem holds for all Gaussian measures on a separable Banach space [3] and also for other measures with tails which decay at least as fast as a Gaussian.

It is demonstrated in [27] that in many applications, including those considered here, the potential $\Phi(\cdot; y)$ satisfies certain natural bounds on a Banach space $(X, \|\cdot\|_X)$. It is then natural to choose the prior μ_0 so that $\mu_0(X) = 1$. Such bounds on the forward problem Φ are summarized in the following assumptions. We assume that the data y lay in a Banach space $(Y, \|\cdot\|_Y)$. The key point about the form of Assumption 1(i) is that it allows the use of the Fernique theorem to control integrals against μ. Assumption 1(ii) may be used to obtain lower bounds on the normalization constant $Z(y)$.

Assumption 1. For some Banach space X with $\mu_0(X) = 1$, the function $\Phi : X \times Y \to \mathbb{R}$ satisfies the following:
   (i) for every $\varepsilon > 0$ and $r > 0$ there is $M = M(\varepsilon, r) \in \mathbb{R}$ such that for all $u \in X$ and $y \in Y$ with $\|y\|_Y < r$

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

   (ii) for every $r > 0$ there is a $L = L(r) > 0$ such that for all $u \in X$ and $y \in Y$ with $\max\{\|u\|_X, \|y\|_Y\} < r$

$$\Phi(u; y) \leqslant L(r).$$

For Bayesian inverse problems in which a finite number of observations are made and the observation error η is mean zero Gaussian with covariance matrix Γ, the potential Φ has the form

$$(2.1) \qquad \Phi(u; y) = \frac{1}{2}|y - \mathcal{G}(u)|_\Gamma^2,$$

where $y \in \mathbb{R}^m$ is the data, $\mathcal{G} : X \to \mathbb{R}^m$ is the forward model, and $|\cdot|_\Gamma$ is a covariance weighted norm on $\mathbb{R}^m$ given by $|\cdot|_\Gamma = |\Gamma^{-\frac{1}{2}} \cdot|$ and $|\cdot|$ denotes the standard Euclidean norm. In this case it is natural to express conditions on the measure μ in terms of $\mathcal{G}$.

Assumption 2. For some Banach space X with $\mu_0(X) = 1$, the function $\mathcal{G} : X \to \mathbb{R}^m$ satisfies the following: for every $\varepsilon > 0$ there is $M = M(\varepsilon) \in \mathbb{R}$ such that, for all $u \in X$,

$$|\mathcal{G}(u)|_\Gamma \leqslant \exp(\varepsilon \|u\|_X^2 + M).$$

LEMMA 2.1. *Assume that* $\Phi : X \times \mathbb{R}^m \to \mathbb{R}$ *is given by* (2.1) *and let* $\mathcal{G}$ *satisfy Assumption* 2. *Assume also that* μ_0 *is a Gaussian measure satisfying* $\mu_0(X) = 1$. *Then* Φ *satisfies Assumption* 1.

Proof. Assumption 1(i) is automatic since Φ is positive; Assumption 1(ii) follows from the bound

$$\Phi(u; y) \leqslant |y|_\Gamma^2 + |\mathcal{G}(u)|_\Gamma^2$$

and the use of the exponential bound on $\mathcal{G}$. $\square$

Since the dependence on y is not relevant in this paper, we suppress it notationally and study measures μ given by

$$(2.2) \qquad \frac{d\mu}{d\mu_0}(u) = \frac{1}{Z} \exp(-\Phi(u)),$$

where the normalization constant Z is given by

$$(2.3) \qquad Z = \int_X \exp(-\Phi(u)) d\mu_0(u).$$

We approximate μ by approximating Φ over some N-dimensional subspace of X. In particular, we define μ^N by

$$(2.4) \qquad \frac{d\mu^N}{d\mu_0}(u) = \frac{1}{Z^N} \exp(-\Phi^N(u)),$$

where

$$(2.5) \qquad Z^N = \int_X \exp(-\Phi^N(u)) d\mu_0(u).$$

The potential Φ^N should be viewed as resulting from an approximation to the solution of the forward problem. Our interest is in translating approximation results for Φ into approximation results for μ. The following theorem proves such a result. Again the particular exponential dependence of the error constant for the forward approximation is required so that we may use the Fernique theorem to control certain expectations arising in the analysis.

THEOREM 2.2. *Assume that* Φ *and* Φ^N *satisfy Assumptions* 1(i) *and* 1(ii) *with constants uniform in* N. *Assume also that, for any* $\varepsilon > 0$, *there is* $K = K(\varepsilon) > 0$ *such that*

$$(2.6) \qquad |\Phi(u) - \Phi^N(u)| \leqslant K \exp(\varepsilon \|u\|_X^2) \psi(N),$$

where $\psi(N) \to 0$ *as* $N \to \infty$. *Then the measures* μ *and* μ^N *are close with respect to the Hellinger distance: there is a constant* C, *independent of* N, *such that*

$$(2.7) \qquad d_{\mathrm{Hell}}(\mu, \mu^N) \leqslant C \psi(N).$$

Consequently all moments of $\|u\|_X$ are $\mathcal{O}\big(\psi(N)\big)$ close. In particular, the mean and, in the case X is a Hilbert space, the covariance operator are $\mathcal{O}\big(\psi(N)\big)$ close.

Proof. Throughout the proof, all integrals are over X. The constant C may depend upon r and changes from occurrence to occurrence. Using Assumption 1(ii) gives

$$|Z| \geqslant \int_{\{\|u\|_X \leqslant r\}} \exp\big(-L(r)\big) d\mu_0(u) \geqslant \exp\big(-L(r)\big) \mu_0\{\|u\|_X \leqslant r\}.$$

This lower bound is positive because μ_0 has a full measure on X and is Gaussian so that all balls in X have positive probability. We have an analogous lower bound for $|Z^N|$.

From Assumptions 1(i) and (2.6), using the fact that μ_0 is a Gaussian probability measure so that the Fernique Theorem A.3 applies, we obtain

$$|Z - Z^N| \leqslant \int K\psi(N) \exp\big(\varepsilon\|u\|_X^2 - M\big) \exp\big(\varepsilon\|u\|_X^2\big) d\mu_0(u)$$
$$\leqslant C\psi(N).$$

From the definition of Hellinger distance we have

$$2d_{\mathrm{Hell}}(\mu,\mu^N)^2 = \int \left(Z^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\Phi(u)\right) - (Z^N)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\Phi^N(u)\right) \right)^2 d\mu_0(u)$$
$$\leqslant I_1 + I_2,$$

where

$$I_1 = \frac{2}{Z} \int \left(\exp\left(-\frac{1}{2}\Phi(u)\right) - \exp\left(-\frac{1}{2}\Phi^N(u)\right) \right)^2 d\mu_0(u),$$
$$I_2 = 2|Z^{-\frac{1}{2}} - (Z^N)^{-\frac{1}{2}}|^2 \int \exp(-\Phi^N(u)) d\mu_0(u).$$

Now, again using Assumption 1(i) and (2.6), together with the Fernique Theorem A.3,

$$\frac{Z}{2} I_1 \leqslant \int \frac{1}{4} K^2 \psi(N)^2 \exp\big(3\varepsilon\|u\|_X^2 - M\big) d\mu_0(u)$$
$$\leqslant C\psi(N)^2.$$

Note that the bounds on Z, Z^N from below are independent of N. Furthermore,

$$\int \exp\big(-\Phi^N(u)\big) d\mu_0(u) \leqslant \int \exp\big(\varepsilon\|u\|_X^2 - M\big) d\mu_0(u)$$

with bound independent of N, by the Fernique Theorem A.3. Thus

$$I_2 \leqslant C\big(Z^{-3} \vee (Z^N)^{-3}\big)|Z - Z^N|^2$$
$$\leqslant C\psi(N)^2.$$

Combining gives the desired continuity result in the Hellinger metric.

Finally all moments of u in X are finite under the Gaussian measure μ_0 by the Fernique Theorem A.3. It follows that all moments are finite under μ and μ^N because, for $f : X \to Z$ polynomially bounded,

$$\mathbb{E}^\mu \|f\| \leqslant \left(\mathbb{E}^{\mu_0} \|f\|^2\right)^{\frac{1}{2}} \left(\mathbb{E}^{\mu_0} \exp(-2\Phi(u;y))\right)^{\frac{1}{2}},$$

and the first term on the right-hand side is finite since all moments are finite under μ_0, while the second term may be seen to be finite by the use of Assumption 1(i) and the Fernique Theorem A.3. $\square$

For Bayesian inverse problems with finite data the potential Φ has the form given in (2.1) where $y \in \mathbb{R}^m$ is the data, $\mathcal{G} : X \to \mathbb{R}^m$ is the forward model, and $|\cdot|_\Gamma$ is a covariance weighted norm on $\mathbb{R}^m$. In this context the following corollary is useful.

COROLLARY 2.3. *Assume that Φ is given by (2.1) and that $\mathcal{G}$ is approximated by a function $\mathcal{G}^N$ with the property that, for any $\varepsilon > 0$, there is $K' = K'(\varepsilon) > 0$ such that*

$$|\mathcal{G}(u) - \mathcal{G}^N(u)| \leqslant K' \exp\left(\varepsilon \|u\|_X^2\right) \psi(N), \tag{2.8}$$

where $\psi(N) \to 0$ as $N \to \infty$. If $\mathcal{G}$ and $\mathcal{G}^N$ satisfy Assumption 2 uniformly in N, then Φ and $\Phi^N := \frac{1}{2}|y - \mathcal{G}^N(u)|_\Gamma^2$ satisfy the conditions necessary for application of Theorem 2.2 and all the conclusions of that theorem apply.

Proof. That (i) and (ii) of Assumption 1 hold follows as in the proof of Lemma 2.1. Also (2.6) holds since (for some $K(\cdot)$ defined in the course of the following chain of inequalities)

$$\begin{aligned}
|\Phi(u) - \Phi^N(u)| &\leqslant \frac{1}{2}|2y - \mathcal{G}(u) - \mathcal{G}^N(u)|_\Gamma |\mathcal{G}(u) - \mathcal{G}^N(u)|_\Gamma \\
&\leqslant \left(|y| + \exp\left(\varepsilon\|u\|_X^2 + M\right)\right) \times K'(\varepsilon) \exp\left(\varepsilon\|u\|_X^2\right)\psi(N) \\
&\leqslant K(2\varepsilon) \exp(2\varepsilon\|u\|_X^2)\psi(N)
\end{aligned}$$

as required. $\square$

A notable fact concerning Theorem 2.2 is that the rate of convergence attained in the solution of the forward problem, encapsulated in the approximation of the function Φ by Φ^N, is transferred into the rate of convergence of the related inverse problem for measure μ given by (2.2) and its approximation by μ^N. Key to achieving this transfer of rates of convergence is the dependence of the constant in the forward error bound (2.6) on u. In particular, it is necessary that this constant is integrable by use of the Fernique Theorem A.3. In some applications it is not possible to obtain such dependence. Then convergence results can sometimes still be obtained, but at weaker rates. We now describe a theory for this situation.

THEOREM 2.4. *Assume that Φ and Φ^N satisfy Assumptions 1(i) and 1(ii) with constants uniform in N. Assume also that for any $R > 0$ there is $K = K(R) > 0$ such that for all u with $\|u\|_X \leqslant R$*

$$|\Phi(u) - \Phi^N(u)| \leqslant K\psi(N), \tag{2.9}$$

where $\psi(N) \to 0$ as $N \to \infty$. Then the measures μ and μ^N are close with respect to the Hellinger distance:

$$d_{\mathrm{Hell}}(\mu, \mu^N) \to 0 \tag{2.10}$$

as $N \to \infty$. Consequently all moments of $\|u\|_X$ under μ^N converge to corresponding moments under μ as $N \to \infty$. In particular the mean, and, in the case X is a Hilbert space, the covariance operator converge.

Proof. Throughout the proof, all integrals are over X unless specified otherwise. The constant C changes from occurrence to occurrence. The normalization constants Z and Z^N satisfy lower bounds which are identical to that proved for Z in the course of establishing Theorem 2.2.

From Assumption 1(i) and (2.9),

$$
\begin{aligned}
|Z - Z^N| &\leqslant \int_X |\exp\big(-\Phi(u)\big) - \exp\big(-\Phi^N(u)\big)| d\mu_0 \\
&\leqslant \int_{\{\|u\|_X \leqslant R\}} \exp\big(\varepsilon\|u\|_X^2 - M\big)|\Phi(u) - \Phi^N(u)| d\mu_0(u) \\
&\quad + \int_{\{\|u\|_X > R\}} 2\exp\big(\varepsilon\|u\|_X^2 - M\big) d\mu_0(u) \\
&\leqslant \exp\big(\varepsilon R^2 - M\big) K(R)\psi(N) + J_R \\
&:= K_1(R)\psi(N) + J_R.
\end{aligned}
$$

Here

$$
J_R = \int_{\{\|u\|_X > R\}} 2\exp\big(\varepsilon\|u\|_X^2 - M\big) d\mu_0(u).
$$

Now, again by the Fernique Theorem A.3, $J_R \to 0$ as $R \to \infty$ so, for any $\delta > 0$, we may choose $R > 0$ such that $J_R < \delta$. Now choose $N > 0$ so that $K_1(R)\psi(N) < \delta$ to deduce that $|Z - Z^N| < 2\delta$. Since $\delta > 0$ is arbitrary, this proves that $Z^N \to Z$ as $N \to \infty$.

From the definition of Hellinger distance we have

$$
2d_{\mathrm{Hell}}(\mu, \mu^N)^2 = \int \left(Z^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\Phi(u)\right) - (Z^N)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\Phi^N(u)\right) \right)^2 d\mu_0(u)
$$

$$
\leqslant I_1 + I_2,
$$

where

$$
I_1 = \frac{2}{Z} \int \left(\exp\left(-\frac{1}{2}\Phi(u)\right) - \exp\left(-\frac{1}{2}\Phi^N(u)\right) \right)^2 d\mu_0(u),
$$

$$
I_2 = 2|Z^{-\frac{1}{2}} - (Z^N)^{-\frac{1}{2}}|^2 \int \exp(-\Phi^N(u)) d\mu_0(u).
$$

Now, again using Assumption 1(i) and (2.9),

$$
\begin{aligned}
I_1 &\leqslant \frac{1}{2Z} \int_{\{\|u\|_X \leqslant R\}} K(R)^2 \psi(N)^2 \exp\big(\varepsilon\|u\|_X^2 - M\big) d\mu_0(u) \\
&\quad + \frac{4}{Z} \int_{\{\|u\|_X > R\}} 2\exp\big(\varepsilon\|u\|_X^2 - M\big) d\mu_0(u) \\
&\leqslant \frac{1}{2Z} K_2(R)\psi(N)^2 + \frac{4}{Z} J_R
\end{aligned}
$$

for suitably chosen $K_2 = K_2(R)$. An argument similar to the one above for $|Z - Z^N|$ shows that $I_1 \to 0$ as $N \to \infty$.

Note that the bounds on Z, Z^N from below are independent of N. Furthermore,

$$\int \exp\bigl(-\Phi^N(u)\bigr)d\mu_0(u) \leqslant \int \exp\bigl(\varepsilon\|u\|_X^2 - M\bigr)d\mu_0(u)$$

with bound independent of N, by the Fernique Theorem A.3. Thus

$$|Z^{-\frac{1}{2}} - (Z^N)^{-\frac{1}{2}}|^2 \leqslant C\bigl(Z^{-3} \vee (Z^N)^{-3}\bigr)|Z - Z^N|^2,$$

and so $I_2 \to 0$ as $N \to \infty$. Combining gives the desired continuity result in the Hellinger metric.

The proof may be completed by the same arguments used in Theorem 2.2. $\qquad\square$

In section 4 of [27], Theorem 2.2 is applied to the linear Gaussian problem of determining the initial condition of the heat equation from observation of the solution at a positive time. In the next two sections we show how Corollary 2.3 and Theorem 2.4 may be applied to study finite dimensional approximation of non-Gaussian inverse problems arising in the determination of the initial condition for two nonlinear, dissipative models arising in fluid mechanics.

3. Lagrangian data assimilation. In this section we study an idealized model of Lagrangian data assimilation as practiced in oceanography. The setup is exactly that in [5] except that we consider the Stokes equations in place of the Navier–Stokes equations. The velocity field v given by the incompressible Stokes ($\iota = 0$) or Navier–Stokes ($\iota = 1$) equations is given by

(3.1a) $$\frac{\partial v}{\partial t} + \iota v \cdot \nabla v = \nu \Delta v - \nabla p + f, \qquad (x, t) \in D \times [0, \infty),$$

(3.1b) $$\nabla \cdot v = 0, \qquad (x, t) \in D \times [0, \infty),$$

(3.1c) $$v = u, \qquad (x, t) \in \overline{D} \times \{0\}.$$

Here D is the unit square. We impose periodic boundary conditions on the velocity field v and the pressure p. We assume that f has zero average over D, noting that this implies the same for $v(x, t)$, provided that $u(x) = v(x, 0)$ has zero initial average.

The PDE can be formulated as a linear dynamical system on the Hilbert space

(3.2) $$\mathcal{H} = \left\{ u \in L^2_{\text{per}}(D) \,\bigg|\, \int_D u\,dx = 0, \nabla \cdot u = 0 \right\},$$

with the usual $L^2(D)$ norm and inner product on this subspace of $L^2_{\text{per}}(D)$. Throughout this article A denotes the (self-adjoint, positive) Stokes operator on $\mathbb{T}^2$ and $P : L^2_{per} \to \mathcal{H}$ the Leray projector [29, 30]. The operator A is densely defined on $\mathcal{H}$ and is the generator of an analytic semigroup. We denote by $\{(\phi_k, \lambda_k)\}_{k \in \mathbb{K}}$ a complete orthonormal set of eigenfunctions/eigenvalues for A in $\mathcal{H}$. We then define fractional powers of A by

(3.3) $$A^\alpha u = \sum_{k \in \mathbb{K}} \lambda_k^\alpha \langle u, \phi_k \rangle \phi_k.$$

For any $s \in \mathbb{R}$ we define the Hilbert spaces $\mathcal{H}^s$ by

(3.4) $$\mathcal{H}^s = \left\{ u : \sum_{k \in \mathbb{K}} \lambda_k^s |\langle u, \phi_k \rangle|^2 < \infty \right\}.$$

The norm in $\mathcal{H}^s$ is denoted by $\|\cdot\|_s$ and is given by

$$\|u\|_s^2 = \sum_{k \in \mathbb{K}} \lambda_k^s |\langle u, \phi_k \rangle|^2.$$

Of course, $\mathcal{H}^0 = \mathcal{H}$.

If we let $\psi = Pf$, then in the Stokes case $\iota = 0$ we may write (3.1) as an ODE in Hilbert space $\mathcal{H}$:

$$\frac{dv}{dt} + \nu A v = \psi, \qquad v(0) = u. \tag{3.5}$$

Our aim is to determine the initial velocity field u from Lagrangian data. To be precise we assume that we are given noisy observations of J tracers with positions z_j solving the integral equations

$$z_j(t) = z_{j,0} + \int_0^t v(z_j(s), s) ds. \tag{3.6}$$

For simplicity assume that we observe all the tracers z at the same set of positive times $\{t_k\}_{k=1}^K$ and that the initial particle tracer positions $z_{j,0}$ are known to us:

$$y_{j,k} = z_j(t_k) + \eta_{j,k}, \quad j = 1, \ldots, J, \quad k = 1, \ldots, K, \tag{3.7}$$

where the $\eta_{j,k}$'s are zero mean Gaussian random variables. Concatenating data we may write

$$y = \mathcal{G}(u) + \eta \tag{3.8}$$

with $y = (y_{1,1}^*, \ldots, y_{J,K}^*)^*$ and $\eta \sim \mathcal{N}(0, \Gamma)$ for some covariance matrix Γ capturing the correlations present in the noise. Note that $\mathcal{G}$ is a complicated function of the initial condition for the Stokes equations, describing the mapping from this initial condition into the positions of Lagrangian trajectories at positive times. We will show that the function $\mathcal{G}$ maps $\mathcal{H}$ into $\mathbb{R}^{2JK}$ and is continuous on a dense subspace of $\mathcal{H}$.

The objective of the inverse problem is thus to find the initial velocity field u, given y. We adopt a Bayesian approach, place a prior $\mu_0(du)$ on u, and identify the posterior $\mu(du) = \mathbb{P}(u|y)du$. We now spend some time developing the Bayesian framework, culminating in Theorem 3.3 which shows that μ is well defined. The reader interested purely in the approximation of μ can skip straight to Theorem 3.4.

The following result shows that the tracer equations (3.6) have a solution, under mild regularity assumptions on the initial data. An analogous result is proved in [7] for the case where the velocity field is governed by the Navier–Stokes equation, and the proof may be easily extended to the case of the Stokes equations.

THEOREM 3.1. *Let $\psi \in L^2(0, T; \mathcal{H})$ and let $v \in C([0, T]; \mathcal{H})$ denote the solution of* (3.5) *with initial data $u \in \mathcal{H}$. Then the integral equation* (3.6) *has a unique solution $z \in C([0, T], \mathbb{R}^2)$.*

We assume throughout that ψ is sufficiently regular that this theorem applies. To determine a formula for the probability of u given y, we apply the Bayesian approach described in [5] for the Navier–Stokes equations and easily generalized to the Stokes equations. For the prior measure we take $\mu_0 = \mathcal{N}(0, \beta A^{-\alpha})$ for some $\beta > 0, \alpha > 1$, with the condition on α chosen to ensure that draws from the prior are in $\mathcal{H}$, by

Lemma A.4. We condition the prior on the observations, to find the *posterior* measure on u. The likelihood of y given u is

$$\mathbb{P}\left(y \mid u\right) \propto \exp\left(-\frac{1}{2}|y - \mathcal{G}(u)|_{\Gamma}^2\right).$$

This suggests the formula

$$\text{(3.9)} \qquad \frac{d\mu}{d\mu_0}(u) \propto \exp(-\Phi(u; y)),$$

where

$$\text{(3.10)} \qquad \Phi(u; y) := \frac{1}{2}|y - \mathcal{G}(u)|_{\Gamma}^2$$

and μ_0 is the prior Gaussian measure. We now make this assertion rigorous. The first step is to study the properties of the forward model $\mathcal{G}$. Proof of the following lemma is given after statement and proof of the main approximation result, Theorem 3.4.

LEMMA 3.2. *Assume that* $\psi \in C([0,T]; \mathcal{H}^\gamma)$ *for some* $\gamma \geqslant 0$. *Consider the forward model* $\mathcal{G} : \mathcal{H} \to \mathbb{R}^{2JK}$ *defined by* (3.7) *and* (3.8).

- *If* $\gamma \geqslant 0$, *then for any* $\ell \geqslant 0$ *there is* $C > 0$ *such that for all* $u \in \mathcal{H}^\ell$

$$|\mathcal{G}(u)| \leqslant C\big(1 + \|u\|_\ell\big).$$

- *If* $\gamma > 0$, *then for any* $\ell > 0$ *and* $R > 0$ *and for all* u_1, u_2 *with* $\|u_1\|_\ell \vee \|u_2\|_\ell < R$, *there is* $L = L(R) > 0$ *such that*

$$|\mathcal{G}(u_1) - \mathcal{G}(u_2)| \leqslant L\|u_1 - u_2\|_\ell.$$

Furthermore, for any $\varepsilon > 0$, *there is* $M > 0$ *such that* $L(R) \leqslant M\exp(\varepsilon R^2)$. *Thus* $\mathcal{G}$ *satisfies Assumption 2 with* $X = \mathcal{H}^s$ *and any* $s \geqslant 0$.

Since $\mathcal{G}$ is continuous on $\mathcal{H}^\ell$ for $\ell > 0$ and since, by Lemma A.4, draws from μ_0 are almost surely in $\mathcal{H}^s$ for any $s < \alpha - 1$, use of the techniques in [5], employing the Stokes equation in place of the Navier–Stokes equation, shows the following.

THEOREM 3.3. *Assume that* $\psi \in C([0,T]; \mathcal{H}^\gamma)$, *for some* $\gamma > 0$, *and that the prior measure* $\mu_0 = \mathcal{N}(0, \beta\mathcal{A}^{-\alpha})$ *is chosen with* $\beta > 0$ *and* $\alpha > 1$. *Then the measure* $\mu(du) = \mathbb{P}(du|y)$ *is absolutely continuous with respect to the prior* $\mu_0(du)$, *with the Radon–Nikodým derivative given by* (3.9).

In fact, the theory in [5] may be used to show that the measure μ is Lipschitz in the data y, in the Hellinger metric. This well posedness underlies the following study of the approximation of μ in a finite dimensional space. We define P^N to be the orthogonal projection in $\mathcal{H}$ onto the subspace spanned by $\{\phi_k\}_{|k|\leqslant N}$; recall that $k \in \mathbb{K} := \mathbb{Z}^2\backslash\{0\}$. Since P^N is an orthogonal projection in any $\mathcal{H}^a$, we have $\|P^N u\|_X \leqslant \|u\|_X$. Define

$$\mathcal{G}^N(u) := \mathcal{G}(P^N u).$$

The approximate posterior measure μ^N is given by (3.9) with $\mathcal{G}$ replaced by $\mathcal{G}^N$. This approximate measure coincides with the prior on the orthogonal complement of $P^N\mathcal{H}$. On $P^N\mathcal{H}$ itself the measure is finite dimensional and amenable to sampling techniques as demonstrated in [6]. We now quantify the error arising from the approximation of $\mathcal{G}$ in the finite dimensional subspace $P^N X$.

THEOREM 3.4. *Let the assumptions of Theorem 3.3 hold. Then, for any $q < \alpha - 1$, there is a constant $c > 0$, independent of N, such that $d_{\mathrm{Hell}}(\mu, \mu^N) \leqslant cN^{-q}$. Consequently the mean and covariance operator of μ and μ^N are $\mathcal{O}(N^{-q})$ close in the $\mathcal{H}$ and $\mathcal{H}$-operator norms, respectively.*

Proof. We set $X = \mathcal{H}^s$ for any $s \in (0, \alpha - 1)$. We employ Corollary 2.3. Clearly, since $\mathcal{G}$ satisfies Assumption 2 by Lemma 3.2, so too does $\mathcal{G}^N$, with constants uniform in N. It remains to establish (2.8). Write $u \in \mathcal{H}^s$ as

$$u = \sum_{k \in \mathbb{K}} u_k \phi_k$$

and note that

$$\sum_{k \in \mathbb{K}} |k|^{2s} |u_k|^2 < \infty.$$

We have, for any $\ell \in (0, s)$,

$$\begin{aligned}
\|u - P^N u\|_\ell^2 &= \sum_{|k| > N} |k|^{2\ell} |u_k|^2 \\
&= \sum_{|k| > N} |k|^{2(\ell - s)} |k|^{2s} |u_k|^2 \\
&\leqslant N^{-2(s-\ell)} \sum_{|k| > N} |k|^{2s} |u_k|^2 \\
&\leqslant C \|u\|_s^2 N^{-2(s-\ell)}.
\end{aligned}$$

By the Lipschitz properties of $\mathcal{G}$ from Lemma 3.2 we deduce that, for any $\ell \in (0, s)$,

$$\begin{aligned}
|\mathcal{G}(u) - \mathcal{G}(P^N u)| &\leqslant M \exp\left(\varepsilon \|u\|_\ell^2\right) \|u - P^N u\|_\ell \\
&\leqslant C^{\frac{1}{2}} M \exp\left(\varepsilon \|u\|_s^2\right) \|u\|_s N^{-(s-\ell)}.
\end{aligned}$$

This establishes the desired error bound (2.8). It follows from Corollary 2.3 that μ^N is $\mathcal{O}(N^{-(s-\ell)})$ close to μ in the Hellinger distance. Choosing s arbitrarily close to its upper bound, and ℓ arbitrarily close to zero, yields the optimal exponent q as appears in the theorem statement. $\square$

Proof of Lemma 3.2. Throughout the proof, the constant C may change from instance to instance, but is always independent of the u_i. It suffices to consider a single observation so that $J = K = 1$. Let $z^{(i)}(t)$ solve

$$z^{(i)}(t) = z_0^{(i)} + \int_0^t v^{(i)}(z^{(i)}(\tau), \tau) d\tau,$$

where $v^{(i)}(x, t)$ solves (3.1) with $u = u_i$.

Let $\ell \in [0, 2 + \gamma)$. By (A.3),

$$(3.11) \qquad \|v^{(i)}(t)\|_s \leqslant C \left(\frac{1}{t^{(s-\ell)/2}} \|u_i\|_\ell + \|\psi\|_{C([0,T];\mathcal{H}^\gamma)} \right)$$

for $s \in [\ell, 2 + \gamma)$. Also, by linearity and (A.2),

$$(3.12) \qquad \|v^{(1)}(t) - v^{(2)}(t)\|_s \leqslant \frac{C}{t^{(s-\ell)/2}} \|u_1 - u_2\|_\ell.$$

To prove the first part of the lemma note that, by the Sobolev embedding theorem, for any $s > 1$,

$$|z^{(i)}(t)| \leqslant |z_0^{(i)}| + \int_0^t \|v^{(i)}(\cdot, \tau)\|_{L^\infty} d\tau$$

$$\leqslant C\Big(1 + \int_0^t \|v^{(i)}(\cdot, \tau)\|_s d\tau\Big)$$

$$\leqslant C\Big(1 + \int_0^t \frac{1}{\tau^{(s-\ell)/2}} \|u_i\|_\ell d\tau\Big).$$

For any $\gamma \geqslant 0$ and $\ell \in [0, 2 + \gamma)$ we may choose s such that $s \in [\ell, 2 + \gamma) \bigcap (1, \ell + 2)$. Thus the singularity is integrable and we have, for any $t \geqslant 0$,

$$|z^{(i)}(t)| \leqslant C(1 + \|u_i\|_\ell)$$

as required.

To prove the second part of the lemma, choose $\ell \in (0, 2 + \gamma)$ and then choose $s \in [\ell - 1, 1 + \gamma) \cap (1, \ell + 1)$; this requires $\gamma > 0$ to ensure a nonempty intersection. Then

$$(3.13) \qquad \|v^{(i)}(t)\|_{1+s} \leqslant C\Big(\frac{1}{t^{(1+s-\ell)/2}} \|u_i\|_\ell + \|\psi\|_{C([0,T];\mathcal{H}^\gamma)}\Big).$$

Now we have

$$|z^{(1)}(t) - z^{(2)}(t)| \leqslant |z^{(1)}(0) - z^{(2)}(0)| + \int_0^t |v^{(1)}(z^{(1)}(\tau), \tau) - v^{(2)}(z^{(2)}(\tau), \tau)| d\tau$$

$$\leqslant \int_0^t \|Dv^{(1)}(\cdot, \tau)\|_{L^\infty} |z^{(1)}(\tau) - z^{(2)}(\tau)| d\tau$$

$$+ \int_0^t \|v^{(1)}(\cdot, \tau) - v^{(2)}(\cdot, \tau)\|_{L^\infty} d\tau$$

$$\leqslant \int_0^t \|v^{(1)}(\cdot, \tau)\|_{1+s} |z^{(1)}(\tau) - z^{(2)}(\tau)| d\tau$$

$$+ \int_0^t \|v^{(1)}(\cdot, \tau) - v^{(2)}(\cdot, \tau)\|_s d\tau$$

$$\leqslant \int_0^t C\Big(\frac{1}{\tau^{(1+s-\ell)/2}} \|u_1\|_\ell + \|\psi\|_{C([0,T];\mathcal{H}^\gamma)}\Big) |z^{(1)}(\tau) - z^{(2)}(\tau)| d\tau$$

$$+ \int_0^t \frac{C}{\tau^{(s-\ell)/2}} \|u_1 - u_2\|_\ell d\tau.$$

Both time singularities are integrable and application of the Gronwall inequality from Lemma A.1 gives, for some C depending on $\|u_1\|_\ell$ and $\|\psi\|_{C([0,T];\mathcal{H}^\gamma)}$,

$$\|z^{(1)} - z^{(2)}\|_{L^\infty((0,T);\mathbb{R}^2)} \leqslant C\|u_1 - u_2\|_\ell.$$

The desired Lipschitz bound on $\mathcal{G}$ follows. In particular, the desired dependence of the Lipschitz constant follows from the fact that for any $\varepsilon > 0$ there is $M > 0$ with the property that, for all $\theta \geqslant 0$,

$$1 + \theta \exp(\theta) \leqslant M \exp(\varepsilon \theta^2). \qquad \square$$

We conclude this section with the results of numerical experiments illustrating the theory. We consider the Stokes equations with viscosity $\nu = 0.05$. We compute the posterior distribution on the initial condition from observation of J Lagrangian trajectories at one time $t = 0.1$. The prior measure is taken to be $\mathcal{N}(0, 400 \times A^{-2})$. The initial condition used to generate the data is found by making a single draw from the prior measure and the observational noise on the Lagrangian data is independently and identically distributed (i.i.d.) $\mathcal{N}(0, \gamma^2)$ with $\gamma = 0.01$.

Note that, in the periodic geometry assumed here, the Stokes equations can be solved exactly by Fourier analysis [30]. Thus there are four sources of approximation when attempting to sample the posterior measure on u. These are

(i) the effect of generating approximate samples from the posterior measure by use of MCMC methods;

(ii) the effect of approximating u in a finite subspace found by truncating the eigenbasis of the Stokes operator;

(iii) the effect of interpolating a velocity field on a grid, found from use of the FFT, into values at the arbitrary locations of Lagrangian tracers;

(iv) the effect of the time-step in an Euler integration of the Lagrangian trajectory equations.

The MCMC method that we use is a generalization of the random walk Metropolis method and is detailed in [6]. The method is appropriate for sampling measures absolutely continuous with respect to a Gaussian in the situation where it is straightforward to sample directly from the Gaussian itself. We control the error (i) simply by running the MCMC method until time averages of various test statistics have converged; the reader interested in the effect of this Monte Carlo error should consult [6]. The error in (ii) is precisely the error which we quantify in Theorem 3.4; for the particular experiments used here we predict an error of order N^{-q} for any $q \in (0, 1)$. In this paper we have not analyzed the errors resulting from (iii) and (iv): these approximations are not included in the analysis leading to Theorem 3.4. However, we anticipate that Theorem 2.2 or Theorem 2.4 could be used to study such approximations, and the numerical evidence which follows below is consistent with this conjecture.

In the following three numerical experiments (each illustrated by a figure) we study the effect of one or more of the approximations (ii), (iii), and (iv) on the empirical distribution ("histogram") found from marginalizing data from the MCMC method onto the real part of the Fourier mode with wave vector $k = (0, 1)$. Similar results are found for other Fourier modes although it is important to note that at high values of $|k|$ the data are uninformative and the posterior is very close to the prior (see [6] for details). The first two figures use $J = 9$ Lagrangian trajectories, while the third uses $J = 400$. Figure 3.1 shows the effect of increasing the number of Fourier modes[2] used from 16, through 100 and 196, to a total of 400 modes and illustrates Theorem 3.4 in that convergence to a limit is observed as the number of Fourier modes increases. However, this experiment is conducted by using bilinear interpolation of the velocity field on the grid, in order to obtain off-grid velocities required for particle trajectories. At the cost of quadrupling the number of FFTs it is possible to implement bicubic interpolation.[3] Conducting the same refinement of the number of Fourier modes then yields Figure 3.2. Comparison of Figures 3.1

---

[2]Here by number of Fourier modes, we mean the dimension of the Fourier space approximation, i.e., the number of grid points.

[3]Bicubic interpolation with no added FFTs is also possible by using finite difference methods to find the partial derivatives, but at a lower order of accuracy
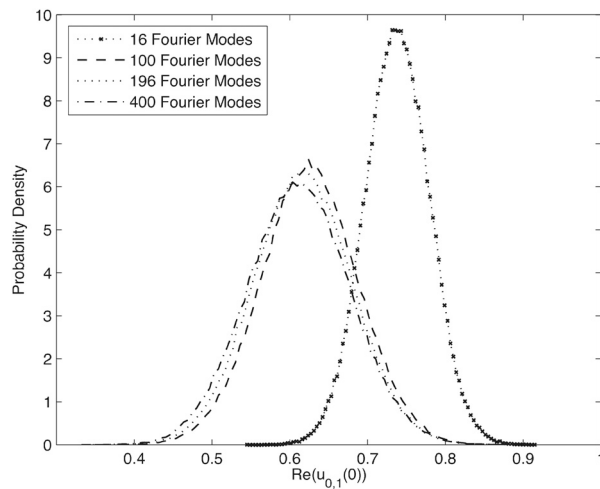
FIG. 3.1. *Marginal distributions on* $\mathrm{Re}(u_{0,1}(0))$ *with differing numbers of Fourier modes.*
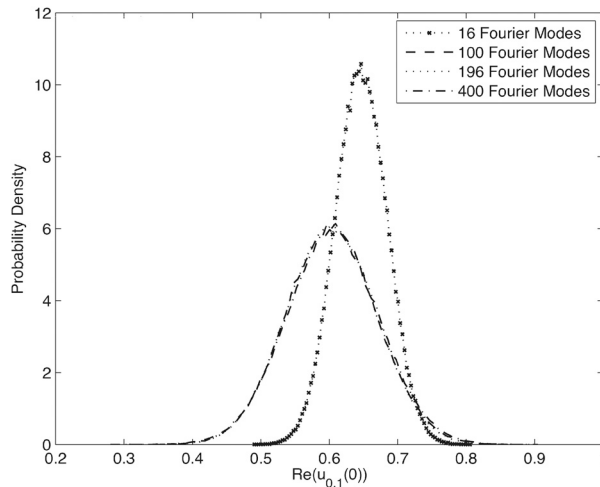

FIG. 3.2. *Marginal distributions on* $\mathrm{Re}(u_{0,1}(0))$ *with differing numbers of Fourier modes, bicubic interpolation used.*

and 3.2 shows that the approximation (iii) by increased order of interpolation leads to improved approximation of the posterior distribution, and Figure 3.2 alone again illustrates Theorem 3.4. Figure 3.3 shows the effect (iv) of reducing the time-step used in the integration of the Lagrangian trajectories. Note that many more (400) particles were used to generate the observations leading to this figure than were used in the preceding two figures. This explains the quantitatively different posterior distribution; in particular, the variance in the posterior distribution is considerably smaller because more data are present. The result shows clearly that reducing the time-step leads to convergence in the posterior distribution.

4. Eulerian data assimilation. In this section we consider a data assimilation problem that is related to weather forecasting applications. In this problem, direct observations are made of the velocity field of an incompressible viscous flow at some
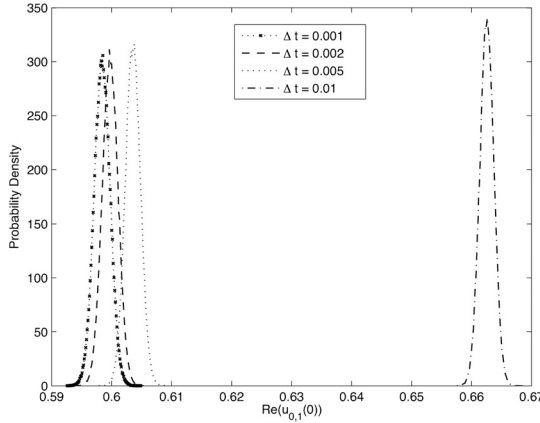
FIG. 3.3. *Marginal distributions on* $\mathrm{Re}(u_{0,1}(0))$ *with differing time-step, Lagrangian data.*

fixed points in space-time, the mathematical model is the two-dimensional Navier–Stokes equations on a torus, and the objective is to obtain an estimate of the initial velocity field. The spaces $\mathcal{H}$ and $\mathcal{H}^s$ are as defined in section 3, with $\|\cdot\|_s$ the norm in $\mathcal{H}^s$ and $\|\cdot\| = \|\cdot\|_0$. The definitions of A, the Stokes operator, and P, the Leray projector, are also as in the previous section.

We write the Navier–Stokes equations (3.1) with $\iota = 1$ as an ordinary differential equation in $\mathcal{H}$

$$\text{(4.1)} \qquad \frac{dv}{dt} + \nu Av + B(v,v) = \psi, \qquad v(0) = u,$$

which is the same as the Stokes equations (3.5), up to the addition of the bilinear form $B(v,v)$. This term arises from projection of the nonlinear term under P. Our aim is to determine u from noisy observations of the velocity field v at time $t > 0$ and at points $x_1, \ldots, x_K \in D$:

$$y_k = v(x_k, t) + \eta_k, \qquad k = 1, \ldots, K.$$

We assume that the noise is Gaussian and the η_k form an i.i.d. sequence with $\eta_1 \sim \mathcal{N}(0, \gamma^2)$. It is known (see Chapter 3 of [29], for example) that for $u \in \mathcal{H}$ and $f \in L^2(0,T;\mathcal{H}^s)$ with $s > 0$ a unique solution to (4.1) exists which satisfies $u \in L^\infty(0,T;\mathcal{H}^{1+s}) \subset L^\infty(0,T;L^\infty(D))$. Therefore for such initial condition and forcing function the value of v at any $x \in D$ and any $t > 0$ can be written as a function of u. Hence, we can write

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

where $y = (y_1^*, \cdots, y_K^*)^* \in \mathbb{R}^{2K}$ and $\eta = (\eta_1^*, \ldots, \eta_k^*)^* \in \mathbb{R}^{2K}$ is distributed as $\mathcal{N}(0, \gamma^2 I)$ and

$$\text{(4.2)} \qquad \mathcal{G}(u) = (v(x_1, t)^*, \cdots, v(x_K, t)^*)^*.$$

Now consider a Gaussian prior measure $\mu_0 \sim \mathcal{N}(u_b, \beta A^{-\alpha})$ with $\beta > 0$ and $\alpha > 1$; the second condition ensures that functions drawn from the prior are in $\mathcal{H}$, by Lemma A.4. In Theorem 3.4 of [5] it is shown that with such prior measure, the posterior measure of the above inverse problem is well defined.

THEOREM 4.1. *Assume that $f \in L^2(0, T, \mathcal{H}^s)$ with $s > 0$. Consider the Eulerian data assimilation problem described above. Define a Gaussian measure μ_0 on $\mathcal{H}$, with mean u_b and covariance operator $\beta A^{-\alpha}$ for any $\beta > 0$ and $\alpha > 1$. If $u_b \in \mathcal{H}^\alpha$, then the probability measure $\mu(du) = \mathbb{P}(du|y)$ is absolutely continuous with respect to μ_0 with Radon–Nikodým derivative*

$$(4.3) \qquad \frac{d\mu}{d\mu_0}(u) \propto \exp\left(-\frac{1}{2\gamma^2}|y - \mathcal{G}(u)|_\Sigma^2\right).$$

We now define an approximation μ^N to μ given by (4.3). The approximation is made by employing the Galerkin approximations of v to define an approximate $\mathcal{G}$. The Galerkin approximation of v, v^N, is the solution of

$$(4.4) \qquad \frac{dv^N}{dt} + \nu A v^N + P^N B(v^N, v^N) = P^N \psi, \qquad v^N(0) = P^N u,$$

with P^N as defined in the previous section. Let

$$\mathcal{G}^N(u) = \left(v^N(x_1, t), \ldots, v^N(x_K, t)\right)^T,$$

and then consider the approximate prior measure μ^N defined via its Radon–Nikodým derivative with respect to μ_0:

$$(4.5) \qquad \frac{d\mu^N}{d\mu_0} \propto \exp\left(-\frac{1}{2\gamma^2}|y - \mathcal{G}^N(u)|_\Sigma^2\right).$$

Our aim is to show that μ^N converges to μ in the Hellinger metric. Unlike the examples in the previous section, we are unable to obtain sufficient control on the dependence of the error constant on u in the forward error bound to enable application of Theorem 2.2; hence we employ Theorem 2.4. In the following lemma we obtain a bound on $\|v(t) - v^N(t)\|_{L^\infty(D)}$ and therefore on $|\mathcal{G}(u) - \mathcal{G}^N(u)|$. Following the statement of the lemma, we state and prove the basic approximation theorem for this section. The proof of the lemma itself is given after the statement and proof of the approximation theorem for the posterior probability measure.

LEMMA 4.2. *Let v^N be the solution of the Galerkin system (4.4). For any $t > t_0$*

$$\|v(t) - v^N(t)\|_{L^\infty(D)} \leqslant C(\|u\|, t_0)\,\psi(N),$$

where $\psi(N) \to 0$ as $N \to \infty$.

The above lemma leads us to the following convergence result for μ^N.

THEOREM 4.3. *Let μ^N be defined according to (4.5) and let the assumptions of Theorem 4.1 hold. Then*

$$d_{\mathrm{Hell}}(\mu, \mu^N) \to 0$$

as $N \to \infty$.

Proof. We apply Theorem 2.4 with $X = \mathcal{H}$. Assumption 2 (and hence Assumption 1) is established in Lemma 3.1 of [5]. By Lemma 4.2

$$|\mathcal{G}(u) - \mathcal{G}^N(u)| \leqslant K\psi(N)$$

with $K = K(\|u\|)$ and $\psi(N) \to 0$ as $N \to 0$. Therefore the result follows by Theorem 2.4. $\quad\square$

Proof of Lemma 4.2. Let $e_1 = v - P^N v$ and $e_2 = P^N v - v^N$. Applying P^N to (4.1) yields

$$\frac{dP^N v}{dt} + \nu A P^N v + P^N B(v, v) = P^N \psi.$$

Therefore $e_2 = P^N v - v^N$ satisfies

(4.6) $$\frac{de_2}{dt} + \nu A e_2 = P^N B(e_1 + e_2, v) + P^N B(v^N, e_1 + e_2), \qquad e_2(0) = 0.$$

Since for any $l \geqslant 0$ and for $m > l$

(4.7) $$\|e_1\|_l^2 \leqslant \frac{1}{N^{2(m-l)}} \|v\|_m^2,$$

we will obtain an upper bound for $\|e_2\|_{1+l}$, $l > 0$, in terms of the Sobolev norms of e_1 and then use the embedding $\mathcal{H}^{1+l} \subset L^\infty$ to conclude the result of the lemma.

Taking the inner product of (4.6) with e_2, and noting that P^N is self-adjoint, $P^N e_2 = e_2$, and $(B(v, w), w) = 0$, we obtain

$$\frac{1}{2} \frac{d}{dt} \|e_2\|^2 + \nu \|De_2\|^2 = (B(e_1 + e_2, v), e_2) + (B(v^N, e_1), e_2)$$

$$\leqslant c\|e_1\|^{1/2}\|e_1\|_1^{1/2}\|v\|_1\|e_2\|^{1/2}\|e_2\|_1^{1/2} + c\|e_2\|\,\|v\|_1\,\|e_2\|_1$$

$$+ c\|v^N\|^{1/2}\|v^N\|_1^{1/2}\|e_1\|_1\|e_2\|^{1/2}\|e_2\|_1^{1/2}$$

$$\leqslant c\|e_1\|^2\,\|e_1\|_1^2 + c\|v\|_1^2\,\|e_2\| + c\|e_2\|^2\,\|v\|_1^2$$

$$+ c\|v^N\|\,\|v^N\|_1\,\|e_1\|_1 + c\|e_1\|_1\,\|e_2\| + \frac{\nu}{2}\|e_2\|_1^2.$$

Therefore

$$\frac{d}{dt}(1 + \|e_2\|^2) + \nu\,\|De_2\|^2$$

$$\leqslant c\,(1 + \|v\|_1^2)\,(1 + \|e_2\|^2) + c(1 + \|e_1\|^2)\,\|e_1\|_1^2 + c\,\|v^N\|\,\|v^N\|_1\,\|e_1\|_1,$$

which gives

$$\|e_2(t)\|^2 + \nu \int_0^t \|De_2\|^2 \leqslant c\,\beta(t) \left(1 + \int_0^t \|v^N\|^2\,\|v^N\|_1^2\,d\tau\right) \int_0^t \|e_1\|_1^2\,d\tau$$

$$+ c\,\beta(t) \int_0^t (1 + \|e_1\|^2)\,\|e_1\|_1^2\,d\tau$$

with

$$\beta(t) = \exp\left(c \int_0^t 1 + \|v\|_1^2\,d\tau\right).$$

Hence

(4.8) $$\|e_2(t)\|^2 + \nu \int_0^t \|De_2\|^2 \leqslant c(1 + \|u\|^4)\,e^{c + c\|u\|^2} \int_0^t (1 + \|e_1\|^2)\,\|e_1\|_1^2\,d\tau.$$

To estimate $\|e_2(t)\|_s$ for $s < 1$, we take the inner product of (4.6) with $A^s e_2$, $0 < s < 1$, and write

$$\frac{1}{2}\frac{d}{dt}\|e_2\|_s^2 + \nu\|e_2\|_{1+s}^2 \leqslant |(((e_1 + e_2) \cdot \nabla)v, A^s e_2)| + |((v^N \cdot \nabla)(e_1 + e_2), A^s e_2)|.$$

Using

$$|((u \cdot \nabla)v, A^s w)| \leqslant c\|u\|_s \|v\|_1 \|w\|_{1+s}$$

and Young's inequality we obtain

$$\frac{d}{dt}\|e_2\|_s^2 + \nu\|e_2\|_{1+s}^2 \leqslant c\left(\|e_1\|_s^2 + \|e_2\|_s^2\right)\|v\|_1^2 + c\|v^N\|_s^2\left(\|e_1\|_1^2 + \|e_2\|_1^2\right).$$

Now integrating with respect to t over (t_0, t) with $0 < t_0 < t$ we can write

$$\|e_2(t)\|_s^2 + \nu\int_{t_0}^t \|e_2\|_{1+s}^2\, d\tau \leqslant \|e_2(t_0)\|_s^2 + c\sup_{\tau \geqslant t_0}\|v(\tau)\|_1^2 \int_0^t \|e_1\|_s^2 + \|e_2\|_s^2 d\tau$$

$$+ c\sup_{\tau \geqslant t_0}\|v^N(\tau)\|_s^2 \int_0^t \|e_1\|_1^2 + \|e_2\|_1^2\, d\tau.$$

Therefore since for $s \leqslant 1$ and $t \geqslant t_0$

$$\|v(t)\|_s^2 \leqslant \frac{c(1 + \|u\|^2)}{t_0^s},$$

and noting that the same kind of decay bounds that hold for v can be shown similarly for v^N as well, we have

$$\|e_2(t)\|_s^2 + \nu\int_{t_0}^t \|e_2\|_{1+s}^2\, d\tau \leqslant \|e_2(t_0)\|_s^2 + \frac{c}{t_0}(1 + \|u\|^6)e^{c+c\|u\|^2} \int_0^t (1 + \|e_1\|^2)\|e_1\|_1^2\, d\tau.$$

Integrating the above inequality with respect to t_0 in $(0, t)$ we obtain

$$(4.9) \quad \|e_2(t)\|_s^2 + \nu\int_{t_0}^t \|e_2\|_{1+s}^2\, d\tau \leqslant \frac{c}{t_0}(1 + \|u\|^6)e^{c+c\|u\|^2} \int_0^t (1 + \|e_1\|^2)\|e_1\|_1^2\, d\tau$$

for $t > t_0$.

Now we estimate $\|e_2(t)\|_s$ for $s > 1$. Taking the inner product of (4.6) with $A^{1+l}e_2$, $0 < l < 1$, we obtain

$$\frac{1}{2}\frac{d}{dt}\|e_2\|_{1+l}^2 + \nu\|e_2\|_{2+l}^2 \leqslant |(((e_1 + e_2) \cdot \nabla)v, A^{1+l}e_2)|$$
$$+ |((v^N \cdot \nabla)(e_1 + e_2), A^{1+l}e_2)|.$$

Since (see [5])

$$((u \cdot \nabla)v, A^{1+l}w) \leqslant c\|u\|_{1+l}\|v\|_1\|w\|_{2+l} + c\|u\|_l\|v\|_2\|w\|_{2+l}$$

and using Young's inequality, we can write

$$\frac{d}{dt}\|e_2\|_{1+l}^2 + \nu\|e_2\|_{2+l}^2 \leqslant c\|e_1\|_{1+l}^2\|v\|_1^2 + c\|e_1\|_l^2\|v\|_2^2$$
$$+ c\|e_2\|_{1+l}^2\|v\|_1^2 + c\|e_2\|_l^2\|v\|_2^2$$
$$+ c\|v^N\|_{1+l}^2\|e_1\|_1^2 + c\|v^N\|_l^2\|e_1\|_2^2$$
$$+ c\|v^N\|_{1+l}^2\|e_2\|_1^2 + c\|v^N\|_l^{2/l}\|e_2\|_{1+l}^2.$$

Now we integrate the above inequality with respect to t and over $(t_0/2 + \sigma, t)$ with $0 < t_0 < t$ and $0 < \sigma < t - t_0/2$ and obtain (noting that $\|v^N\|_s \leqslant \|v\|_s$ for any $s > 0$)

$$\|e_2(t)\|_{1+l}^2 \leqslant \|e_2(t_0/2 + \sigma)\|_{1+l}^2 + \sup_{\tau \geqslant t_0/2} \|v(\tau)\|_1^2 \int_{t_0/2+\sigma}^t \|e_1\|_{1+l}^2 + \|e_2\|_{1+l}^2 \, d\tau$$

$$+ \sup_{\tau \geqslant t_0/2} (\|e_1(\tau)\|_l^2 + \|e_2(\tau)\|_l^2) \int_{t_0/2+\sigma}^t \|v\|_2^2 \, d\tau$$

$$+ \sup_{\tau \geqslant t_0/2} (\|e_1(\tau)\|_1^2 + \|e_2(\tau)\|_1^2) \int_{t_0/2+\sigma}^t \|v^N\|_{1+l}^2 \, d\tau$$

$$+ \sup_{\tau \geqslant t_0/2} (1 + \|v^N(\tau)\|_l^{2/l}) \int_{t_0/2+\sigma}^t \|e_1\|_2^2 + \|e_2\|_{1+l}^2 \, d\tau.$$

We have, for $s > 1$ and $t > t_0$ [5],

$$\|v(t)\|_s^2 \leqslant \frac{c(1 + \|u\|^4)}{t_0^s}.$$

Therefore using (4.9) and (4.7) we conclude that

$$\|e_2(t)\|_{1+l}^2 \leqslant \|e_2(t_0/2 + \sigma)\|_{1+l}^2$$

$$+ C_p(\|u\|) \left(\frac{1}{N^{2(m-l)} \, t_0^{1+m}} + \frac{1}{t_0^{1+l}} \int_0^t (1 + \|e_1\|^2) \, \|e_1\|_1^2 \, d\tau \right.$$

$$+ \left. \frac{1}{N^{2(r-1)} \, t_0^{1+r}} \right)$$

with $r > 1$ and where $C_p(\|u\|)$ is a constant depending on polynomials of $\|u\|$. Integrating the above inequality with respect to σ over $(0, t - t_0/2)$ we obtain

$$\|e_2(t)\|_{1+l}^2 \leqslant C_p(\|u\|) \left(\frac{1}{t_0^{1+l}} + \frac{1}{t_0^{2+l}} \right) \int_0^t (1 + \|e_1\|^2) \, \|e_1\|_1^2 \, d\tau$$

$$+ C_p(\|u\|) \left(\frac{1}{N^{2(m-l)} \, t_0^{2+m}} + \frac{1}{N^{2(r-1)} \, t_0^{2+r}} \right).$$

Now to show that $\|e_1\|^2 + \int_0^t \|e_1\|_1^2 \, d\tau \to 0$ as $N \to \infty$, we note that e_1 satisfies

$$\frac{1}{2} \frac{d}{dt} \|e_1\|^2 + \nu \|De_1\|^2 \leqslant \|(I - \mathbb{P}^N)f\| \, \|e_1\| + \|(B(v, v), e_1)\|$$

$$\leqslant \|(I - \mathbb{P}^N)f\| \, \|e_1\| + \|v\|^{1/2} \, \|Dv\|^{3/2} \, \|e_1\|^{1/2} \, \|De_1\|^{1/2}$$

$$\leqslant \|(I - \mathbb{P}^N)f\| \, \|e_1\| + c \, \|v\|^{2/3} \, \|Dv\|^2 \, \|e_1\|^2 + \frac{\nu}{2} \|De_1\|^2.$$

Therefore

$$\frac{d}{dt} \|e_1\|^2 + \nu \|De_1\|^2 \leqslant c \, \|(I - \mathbb{P}^N)f\|^2 + c \, (1 + \|v\|^{2/3} \, \|Dv\|^2) \, \|e_1\|^2,$$

and after integrating, we get

$$\|e_1\|^2 + \int_0^T \|e_1\|_1^2 \, d\tau \leqslant \exp(1 + C_p(\|u\|)) \left(\|e_1(0)\|^2 + \int_0^T \|(I - \mathbb{P}^N)f\|^2 \right) \, d\tau.$$

Since $f \in L^2(0, T; \mathcal{H})$, the above integral tends to zero as $N \to \infty$ and the result follows. $\square$

5. Conclusions. In this paper we have studied the approximation of inverse problems which have been regularized by means of a Bayesian formulation. We have developed a general approximation theory which allows for the transfer of approximation results for the forward problem into approximation results for the inverse problem. The theory clearly separates analysis of the forward problem, in which no probabilistic methods are required, and the probabilistic framework for the inverse problem itself: it is simply necessary that the requisite bounds and approximation properties for the forward problem hold in a space with full measure under the prior. Indeed, the spaces in which the forward approximation theory is developed may be seen to place constraints on the prior, in order to ensure the desired robustness of the inverse problem to finite dimensional approximation.

In applications there are two sources of error when calculating expectations of functions of infinite dimensional random variables: the error for which we provide an analysis in this paper, namely, the approximation of the measure itself in a finite dimensional subspace, together with the error incurred through calculation of expectations. The two sources of error must be balanced in order to optimize computational cost and thus the analysis in this paper is of potential practical value to practitioners in estimating one source of error. The expectations themselves can be computed by MCMC methods [17], by quasi–Monte Carlo methods [24], or by the more recently analyzed polynomial chaos methods [25, 26, 23, 31]. In all these cases there are quantitative error estimates associated with the calculation of an expectation.

We have studied two specific applications, both concerned with determining the initial condition of a dissipative PDE, from observations of various kinds, at positive times. However, the general approach is applicable to a range of inverse problems for functions when formulated in a Bayesian fashion. The article [27] overviews many applications from this point of view. Furthermore we have limited our approximation of the underlying forward problem to spectral methods. However, we anticipate that the general approach will be useful for the analysis of other spatial approximations based on finite element methods, for example, and to approximation errors resulting from time discretization; indeed, it would be interesting to carry out analyses for such approximations.

Appendix A. Here we collect together some basic facts in analysis and probability. First we state the well-known Gronwall inequality in the form in which we will use it.[4]

LEMMA A.1. *Let $I = [c, d)$ with $d \in (c, \infty]$. Assume that $\alpha, u \in C(I; \mathbb{R}^+)$ and that there is $\lambda < \infty$ such that, for all intervals $J \subseteq I$, $\int_J \beta(s)ds < \lambda$. If*

$$u(t) \leqslant \alpha(t) + \int_c^t \beta(s)u(s)ds, \qquad t \in I,$$

then

$$u(t) \leqslant \alpha(t) + \int_c^t \alpha(s)\beta(s) \exp\left(\int_s^t \beta(r)dr\right) ds, \qquad t \in I.$$

In particular, if $\alpha(t) = u + 2at$ is positive in I and $\beta(t) = 2b$, then

$$u(t) \leqslant \exp(2bt)u + \frac{a}{b}(\exp(2bt) - 1), \qquad t \in I.$$

---

[4]See `http://en.wikipedia.org/wiki/Gronwall's_inequality`.

Finally, if $c = 0$, and $0 < \alpha(t) \leqslant K$ in I, then

$$u(t) \leqslant K + K\lambda \exp(\lambda), \qquad t \in I.$$

Now consider the Hilbert-space valued ODE

(A.1)
$$\frac{dv}{dt} + Av = f, \qquad v(0) = u.$$

We state some basic results in this area, provable by use of the techniques in [20], for example, or by direct calculation using the eigenbasis for A. For $f = 0$ the solution $v \in C([0, \infty), \mathcal{H}) \cap C^1((0, \infty), D(A))$ and

(A.2)
$$\|v\|_s^2 \leqslant Ct^{-(s-l)}\|u\|_l^2 \quad \forall t \in (0, T] \, s \geqslant l \geqslant 0.$$

If $f \in C([0, T], \mathcal{H}^\gamma)$ for some $\gamma \geqslant 0$, then (A.1) has a unique mild solution $u \in C([0, T]; \mathcal{H})$ and, for $0 \leqslant \ell < \gamma + 2$,

(A.3)
$$\|v(t)\|_s \leqslant C\Big(\frac{\|u\|_l}{t^{(s-l)/2}} + \|f\|_{C([0,T],\mathcal{H}^\gamma)}\Big)$$

for $s \in [\ell, 2 + \gamma)$.

It is central to this paper to estimate the distance between two probability measures. Assume that we have two probability measures μ and μ', both absolutely continuous with respect to the same reference measure ν. The *Hellinger distance* between μ and μ' is

$$d_{\mathrm{Hell}}(\mu, \mu') = \sqrt{\Big(\frac{1}{2} \int \Big(\sqrt{\frac{d\mu}{d\nu}} - \sqrt{\frac{d\mu'}{d\nu}}\Big)^2 d\nu\Big)}.$$

The Hellinger distance is particularly useful for estimating the difference between expectation values of functions of random variables under different measures. This is illustrated in the following lemma with straightforward proof.

LEMMA A.2. *Assume that two measures μ and μ' on a Banach space $(X, \|\cdot\|_X)$ are both absolutely continuous with respect to a measure ν. Assume also that $f : X \to Z$, where $(Z, \|\cdot\|)$ is a Banach space, has second moments with respect to both μ and μ'. Then*

$$\|\mathbb{E}^\mu f - \mathbb{E}^{\mu'} f\| \leqslant 2\Big(\mathbb{E}^\mu \|f\|^2 + \mathbb{E}^{\mu'}\|f\|^2\Big)^{\frac{1}{2}} d_{\mathrm{Hell}}(\mu, \mu').$$

Furthermore, if $(Z, \langle \cdot, \cdot \rangle)$ is a Hilbert space and $f : X \to Z$ has fourth moments, then

$$\|\mathbb{E}^\mu f \otimes f - \mathbb{E}^{\mu'} f \otimes f\| \leqslant 2\Big(\mathbb{E}^\mu \|f\|^4 + \mathbb{E}^{\mu'}\|f\|^4\Big)^{\frac{1}{2}} d_{\mathrm{Hell}}(\mu, \mu').$$

Note, in particular, that choosing $X = Z$, and with f chosen to be the identity mapping, we deduce that the differences in mean and covariance operators under two measures are bounded above by the Hellinger distance between the two measures.

The following Fernique theorem (see [22], Theorem 2.6) will be used repeatedly.

THEOREM A.3. *Let $x \sim \mu = \mathcal{N}(0, \mathcal{C})$ where μ is a Gaussian measure on Hilbert space H. Assume that $\mu_0(X) = 1$ for some Banach space $(X, \|\cdot\|_X)$ with $X \subseteq H$. Then there exists $\alpha > 0$ such that*

$$\int_{\mathcal{H}} \exp\big(\alpha\|x\|_X^2\big)\mu(dx) < \infty.$$

The following regularity properties of Gaussian random fields will be useful to us; the results may be proved by use of the Kolmogorov continuity criterion, together with the Karhunen–Loeve expansion (see [22], section 3.2).

LEMMA A.4. *Consider a Gaussian measure $\mu = \mathcal{N}(0, \mathcal{C})$ with $\mathcal{C} = \beta A^{-\alpha}$ where A is as defined earlier in this Appendix A. Then $u \sim \mu$ is almost surely s-Hölder continuous for any exponent $s < \min\{1, \alpha - \frac{d}{2}\}$ and $u \in \mathcal{H}^s$, μ—almost surely, for any $s < \alpha - \frac{d}{2}$.*

REFERENCES

[1] A. APTE, C. K. R. T. JONES, A. M. STUART, AND J. VOSS, *Data assimilation: Mathematical and statistical perspectives*, Internat. J. Numer. Methods Fluids, 56 (2008), pp. 1033–1046.

[2] S. R. ARRIDGE, J. P. KAIPIO, V. KOLEHMAINEN, M. SCHWEIGER, E. SOMERSALO, AND T. TARVAINEN, *Approximation errors and model reduction with an application to optical diffusion tomography*, Inverse Problems, 22 (2006), pp. 175–195.

[3] V. I. BOGACHEV, *Gaussian Measures*, American Mathematical Society, Providence, RI, 1998.

[4] D. CALVETTI AND E. SOMERSALO, *Introduction to bayesian scientific computing*, Surveys and Tutorials in the Applied Mathematical Sciences 2, Springer, New York, 2007.

[5] 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 (2010), p. 115008.

[6] S. L. COTTER, M. DASHTI, J. C. ROBINSON, AND A. M. STUART, *MCMC Methods on Function Space and Applications to Fluid Mechanics*, in preparation, 2010.

[7] M. DASHTI AND J. C. ROBINSON, *A simple proof of uniqueness of the particle trajectories for solutions of the Navier-Stokes equations*, Nonlinearity, 22 (2009), pp. 735–746.

[8] H. K. ENGL, M. HANKE, AND A. NEUBAUER, *Regularization of inverse problems*, Kluwer, Dordrecht, the Netherlands, 1996.

[9] J. N. FRANKLIN, *Well-posed stochastic extensions of ill posed linear problems*, J. Math. Anal. Appl., 31 (1970), pp. 682–716.

[10] A. HOFINGER AND H. K. PIKKARAINEN, *Convergence rates for the Bayesian approach to linear inverse problems*, Inverse Problems, 23 (2007), pp. 2469–2484.

[11] K. IDE, M. COURIER, M. GHIL, AND A. LORENC, *Unified notation for assimilation: Operational, sequential and variational*, J. Met. Soc. Japan, 75 (1997), pp. 181–189.

[12] J. KAIPIO AND E. SOMERSALO, *Statistical and computational inverse problems*, Applied Mathematical Sciences 160, Springer, New York, 2005.

[13] J. P. KAIPIO AND E. SOMERSALO, *Approximation errors in nonstationary inverse problems*, Inverse Prob. Imag., 1 (2007), pp. 77–93.

[14] J. P. KAIPIO AND E. SOMERSALO, *Statistical inverse problems: Discretization, model reduction and inverse crimes*, J. Comput. Appl. Math., 198 (2007), pp. 493–504.

[15] A. KIRSCH, *An Introduction to the Mathematical Theory of Inverse Problems*, Springer, New York, 1996.

[16] M. S. LEHTINEN, L. PAIVARINTA, AND E. SOMERSALO, *Linear inverse problems for generalized random variables*, Inverse Problems, 5 (1989), pp. 599–612.

[17] J. LIU, *Monte Carlo strategies in scientific computing*, Springer Texts in Statistics, Springer, New York, 2001.

[18] A. NEUBAUER AND H. K. PIKKARAINEN, *Convergence results for the Bayesian inversion theory*, J. Inverse Ill-Posed Probl., 16 (2008), pp. 601–613.

[19] N. K. NICHOLS, *Data assimilation: Aims and basic concepts*, Data Assimilation for the Earth System, R. Swinbank, V. Shutyaev, and W. A. Lahoz, eds., Kluwer, Dordrecht, the Netherlands, 2003, pp. 9–20.

[20] A. PAZY, *Semigroups of Linear Operators and Applications to Partial Differential Equations*, Springer, New York, 1983.

[21] H. K. PIKKARAINEN, *State estimation approach to nonstationary inverse problems: Discretization error and filtering problem*, Inverse Problems, 22 (2006), pp. 365–379.

[22] G. DA PRATO AND J. ZABCZYK, *Stochastic equations in infinite dimensions*, Encyclopedia of Mathematics and its Applications 44, Cambridge University Press, Cambridge, 1992.

[23] C. SCHWAB AND R. A. TODOR, *Karhunen-Loeve approximation of random fields in domains by generalized fast multipole methods*, J. Comput. Phys., 217 (2006), pp. 100–122.

[24] I. SLOAN AND H. WOZNIAKOWSKI, *When are quasi-Monte Carlo algorithms efficient for high-dimensional integrals?* J. Complexity, 14 (1998), pp. 1–33.

[25] P. D. SPANOS AND R. GHANEM, *Stochastic finite element expansion for random media*, J. Eng. Mech., 115 (1989), pp. 1035–1053.

[26] P. D. SPANOS AND R. GHANEM, *Stochastic Finite Elements: A Spectral Approach*, Dover, New York, 2003.

[27] A. M. STUART, *Inverse problems: A Bayesian approach*, Acta Numer., 19 (2010), to appear.

[28] A. TARANTOLA, *Inverse Problem Theory*. SIAM, Philadelphia, PA, 2005.

[29] R. TEMAM, *Navier-Stokes equations and nonlinear functional analysis*, CBMS-NSF Regional Conference Series in Applied Mathematics 66, SIAM, Philadelphia, PA, 1995.

[30] R. TEMAM, *Navier-Stokes equations*. AMS Chelsea, Providence, RI, 2001.

[31] R. A. TODOR AND C. SCHWAB, *Convergence rates for sparse chaos approximations of elliptic problems with stochastic coefficients*, IMA J. Numer. Anal., 27 (2007), pp. 232–261.

[32] C. R. VOGEL, *Computational Methods for Inverse Problems*, SIAM, Philadelphia, PA, 2002.

[33] L. W. WHITE, *A study of uniqueness for the initialiazation problem for burgers' equation*, J. Math. Anal. Appl., 172 (1993), pp. 412–431.