Bayesian inverse problems in measure spaces with application to Burgers and Hamilton–Jacobi equations with white noise forcing

Hoang, Viet Ha.

2012

Hoang, V. H. (2012). Bayesian inverse problems in measure spaces with application to Burgers and Hamilton–Jacobi equations with white noise forcing. *Inverse Problems*, 28(2), 025009.

<https://hdl.handle.net/10356/100482>

<https://doi.org/10.1088/0266-5611/28/2/025009>

© 2012 IOP Publishing Ltd. This paper was published in *Inverse Problems* and is made available as an electronic reprint (preprint) with permission of IOP Publishing Ltd. The paper can be found at the following official DOI: [<http://dx.doi.org/10.1088/0266-5611/28/2/025009>]. One print or electronic copy may be made for personal use only. Systematic or multiple reproduction, distribution to multiple locations via electronic or other means, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper is prohibited and is subject to penalties under law.

Downloaded on 26 Apr 2023 19:03:50 SGT

Bayesian inverse problems in measure spaces with application to Burgers and Hamilton–Jacobi equations with white noise forcing

This content has been downloaded from IOPscience. Please scroll down to see the full text.

2012 Inverse Problems 28 025009

(<http://iopscience.iop.org/0266-5611/28/2/025009>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 155.69.4.4

This content was downloaded on 10/12/2013 at 04:30

Please note that [terms and conditions apply](#).

Bayesian inverse problems in measure spaces with application to Burgers and Hamilton–Jacobi equations with white noise forcing

Viet Ha Hoang

Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 637371

E-mail: vhhoang@ntu.edu.sg

Received 1 May 2011, in final form 4 November 2011

Published 20 January 2012

Online at stacks.iop.org/IP/28/025009

Abstract

This paper formulates Bayesian inverse problems for inference in a topological measure space given noisy observations. Conditions for the validity of the Bayes' formula and the well posedness of the posterior measure are studied. The abstract theory is then applied to Burgers and Hamilton–Jacobi equations on a semi-infinite time interval with forcing functions which are white noise in time. Inference is made on the white noise forcing, assuming the Wiener measure as the prior.

1. Introduction

Bayesian inverse problems are attracting considerable attention due to their importance in many applications (see the extensive survey by Stuart [25]). The unknown quantities are to be determined from a set of data, which are subject to observational error. Given a functional $\mathcal{G} : X \rightarrow \mathbb{R}^m$ where X is a function space, the observation $y \in \mathbb{R}^m$ of $\mathcal{G}$ is subject to an unbiased noise σ :

$$y = \mathcal{G}(x) + \sigma. \quad (1.1)$$

The inverse problem determines x given the noisy observation y .

Given x , the quantity y can be found by solving the forward (or direct) problem, which are well-posed partial differential equations: for each $x \in X$, there is a unique solution y which depends smoothly on x . However, the inverse problem (1.1) may not have a unique solution; a solution may not even exist; or it can depend sensitively on x .

The solution x of the ill-posed problem (1.1) may be defined reasonably as the minimizer of the functional

$$\|y - \mathcal{G}(x)\|_Y, \quad (1.2)$$

where $\|\cdot\|_Y$ is a measuring norm, for example, the $\mathbb{R}^m$ norm in our setting. However, (1.2) may not possess a solution as minimizing sequences can be highly oscillating and diverge.

Classical approaches solve a well-posed regularized problem which approximates (1.1). The Tikhonov method looks for a minimizer of the well-posed regularized functional

$$F_\delta(x) = \|y - \mathcal{G}(x)\|_Y^2 + \delta \|x\|_E^2$$

for a chosen norm $\|\cdot\|_E$ of x . Another method is the truncated singular value decomposition which restricts the problem to an approximating finite-dimensional space in which the problem becomes well posed. Details about classical methods can be found in [9, 15, 17].

Classical methods regard the observational error σ as deterministic and the approximating solution to (1.2) is found as a single deterministic function. Ill-posedness is removed in a rather ad hoc manner [17]. The Bayesian approach for inverse problems treats σ as a random variable whose distribution is known and x as a random variable with values in the function space X , which is equipped with a predetermined prior probability. The information we seek for the unknown quantity x is encoded in the conditional probability of x given the observation y which contains an unknown observational noise σ whose statistics is known. This conditional probability is the solution of the inverse problem.

Kaipio and Somersalo [17] provide a foundation of the Bayesian approach to inverse problems. In finite-dimensional spaces, the existence of a solution is guaranteed by Bayes' theorem. The approach of Kaipio and Somersalo requires discretizing the problems first, and then applying the Bayesian framework to the discrete problems. However, for the discrete problem to represent the original accurately, the dimension of the approximating space must be very high. Further, in many problems, as discussed by Stuart [25], important insights can be obtained by deferring discretization to the last possible moments. With this viewpoint, in their pioneering work, Cotter *et al* [7] formulate a rigorous mathematical framework for Bayesian inverse problems in Banach spaces X which are equipped with a prior probability μ_0 . When $\mathcal{G}$ is μ_0 measurable, they show that Bayes' formula (2.3) holds for the conditional probability $\mu^y = \mathbb{P}(x|y)$ in this infinite-dimensional setting. When μ_0 is Gaussian, assuming that $\mathcal{G}(x)$ grows polynomially with respect to the norm $\|x\|_X$, they show that μ^y , with respect to the Hellinger distance, depends continuously on y . The problem is thus well-posed.

In some situations, inverse problems in general metric spaces need to be considered, for which the theory in [7] does not readily apply. In fluid turbulence, the behavior of solutions of stochastic partial differential equations after a long time, when initial conditions have negligible effect, is of interest. The behavior of the solution depends solely on the realization of the random forcing. The problems are thus considered over a semi-infinite time interval $(-\infty, T)$; see [22] for Navier–Stokes equations and [8] and [4] for the Burgers equation. Inferring the forcing (as considered in [7]) requires formulating a Bayesian inverse problem in $C(-\infty, T]$ which is a metric space. Another typical example is inference of a positive random field (e.g. the random conductivity) which is strictly bounded away from zero, or a random field which can be expressed as the sum of an infinite series of random variables (e.g. the Kahnen–Love expansion) which are distributed in a finite domain (see e.g. [6, 14, 23]). The unknown x belongs to a topological measure space.

In this paper, we formulate Bayesian inverse problems in general measure spaces. We demonstrate that while the validity of the Bayes' theorem for general measure spaces follow from the results by Cotter *et al* [7], the well-posedness of the posterior measure can be significantly complicated. For Banach spaces X , Cotter *et al* [7] assume that the functional Φ in (2.2) grows polynomially with respect to the norm $\|x\|_X$ and is locally Lipschitz with respect to y and the Lipschitz constant also grows polynomially with respect to $\|x\|_X$. The well-posedness of the posterior measure is deduced from the Fernique theorem. In general measure spaces, conditions for well-posedness become abstract. Unlike in Banach spaces, in metric spaces, the functions $M(r)$ and $G(r, x)$ in assumption 2.3 generally cannot be expressed in

terms of the metric. There is no general description of how to identify them; their identification is problem dependent. We illustrate these by considering inference of the white noise forcing for Burgers and Hamilton–Jacobi equations in the semi-infinite time interval $(-\infty, T]$. Semi-infinite time intervals and the use of the Wiener measure as the prior combine to necessitate the use of a metric structure.

As shown below, for problems in a spatially compact domain, in particular the periodic torus $\mathbb{T}^d$, the proof for the well-posedness of the posterior measure is not so complicated, as the function $G(r, x)$ in assumption 2.3 can be formulated in terms of a seminorm. It is the problems in noncompact domains that highlight the complexity of Bayesian problems in measure spaces. For the case where the forcing potential $F(x)$ in (3.2) has a large maximum and a small minimum in a compact set in comparison to the potential outside, $G(r, x)$ depends on the first time a fluid particle trajectory hits this compact set. Showing the required condition that the function $G(r, x)$ is square integrable with respect to μ_0 is highly nontrivial. Burgers and Hamilton–Jacobi equations in general noncompact domains are far more complicated than this particular case where the forcing potential has a large maximum and a small minimum; similar Bayesian inverse problems in general noncompact domains should be far more technically demanding. This shows that when the well-posedness conditions cannot be formulated in terms of a norm, as in the Banach space setting, Bayesian inverse problems can become far more difficult.

The Burgers equation plays an important role in turbulent theory, cosmology, acoustic theory and non-equilibrium statistical physics (see, e.g., [11, 8, 4]). It is simpler than Navier–Stokes equations due to the absence of the pressure but still keeps the essential nonlinear features. The Burgers equation with white noise forcing in the spatially periodic case is studied by E *et al* [8] in one dimension and by Iturriaga and Khanin [16] in more than one dimension. They prove that almost surely, there is a unique solution that exists for all time. They establish many properties for shocks and for the dynamics of the Lagrange minimizers of the Lax–Oleinik functional. Much less is understood in the non-compact case. Hoang and Khanin [13] study the special case where the forcing potential has a large maximum and a small minimum. They prove that there is a solution that exists for all time; it is the limit of finite-time solutions with the zero initial condition.

In the case where spatially the forcing is the gradient of a potential, assuming that the velocity is also a gradient, the velocity potential satisfies a Hamilton–Jacobi equation with random forcing. This is the Kardar–Parisi–Zhang (KPZ) equation that describes lateral growth of an interface. In the dissipative case, inverse problems to determine the dissipative coefficient for the KPZ equation was studied in Lam and Sander [19] using the classical least-squares fitting method. The forcing for the KPZ equation considered in [19] is white in both space and time. The dissipative coefficient is inferred from the expectation of the observed heights of the growing interface at some spatial and temporal points with respect to the random forcing’s distribution. This expectation is taken as the Monte Carlo average of the surface height at these locations which is obtained from a large number of independent simulations of the forcing. This approach is not feasible when the expectation is not available but instead, data are only obtained from one forcing’s realization that occurs in a practical situation. The white noise forcing may need to be inferred. This viscous (or dissipative) problem requires consideration of a stochastic control problem which ‘becomes’ the variational problem in this paper when the viscosity converges to 0 (see [12]). Though we do not assume a spatially white forcing, and we only consider the inviscid problems, our approach may well be adopted for this problem.

The Burgers equation and equations of the Burgers type are employed as models for studying data assimilation in Bardos and Pironneau [2], Lundvall *et al* [21], Apte *et al* [1] and Freitag *et al* [10].

For the Hamilton–Jacobi equation (3.3), we assume that the solution (i.e. the velocity potential ϕ) is observed at fixed spatial points at fixed moments. This assumption is reasonable as ϕ is continuous and is well defined everywhere (though it is determined within an additive constant). The inviscid Burgers equation possesses many shocks, at which the solution is not defined, but for a given time, the solution is in $L^1_{\text{loc}}(\mathbb{R}^d)$, so observations in the form of a continuous functional in $L^1_{\text{loc}}(\mathbb{R}^d)$ are considered.

In section 2, we establish Bayes’ formula for the posterior measures of Bayesian inverse problems in measure spaces, and derive conditions under which the problems are well-posed. Under this condition, we show that the Hellinger and the Kullback–Leibler distances are locally Lipschitz with respect to y . The results are generalizations of those in the Banach space setting by Cotter *et al* in [7]. When the prior measure μ_0 is a Radon Gaussian measure in a locally convex topological space (see, for example, Bogachev [5] or Lifshits [20]), we show the existence of the maximum *a posteriori* (MAP) estimator. Section 3 introduces the randomly forced Burgers and Hamilton–Jacobi equations and formulates the Bayesian inverse problems. The spatially periodic case is considered in section 4, using the results for random Burgers and Hamilton–Jacobi equations by E *et al* [8] and Iturriaga and Khanin [16]. We establish the validity of Bayes’ formula and the posterior measure’s local Lipschitzness with respect to the Hellinger and the Kullback–Leibler distances; the main results are recorded in theorems 4.6 and 4.12. We also show the existence of a MAP estimator in theorems 4.7 and 4.13. The non-periodic case is studied in section 5, using the results by Hoang and Khanin [13] where we restrict our consideration to forcing potentials that satisfy assumption 5.1, with a large maximum and a small minimum. Under this assumption, for both the Hamilton–Jacobi and Burgers equations, Bayes’ formula is valid and the posterior measure is locally Lipschitz with respect to the Hellinger and the Kullback–Leibler distances. The main results are presented in theorems 5.8 and 5.14. Solutions to the Burgers and Hamilton–Jacobi equations are shown to exist almost surely in Hoang and Khanin [13]. It is not clear whether a solution exists for the forcing function’s realizations in the Cameron–Martin space so the existence of a MAP estimator is open for this case. We discuss this in remark 5.9. We devote section 6 to a conclusion summarizing the paper. In section 7, we prove some technical results that are used in the previous two sections.

Throughout the paper, we define $\langle \cdot, \cdot \rangle_{\Sigma} = \langle \Sigma^{-1/2} \cdot, \Sigma^{-1/2} \cdot \rangle_{\mathbb{R}^m}$ for a positive definite matrix Σ ; the corresponding norm $|\Sigma^{-1/2} \cdot|_{\mathbb{R}^m}$ is denoted $|\cdot|_{\Sigma}$. The $\mathbb{R}^m$ norm is denoted $|\cdot|$. We denote by c various positive constants that do not depend on the white noise; the value of c can differ from one appearance to the next.

2. Bayesian inverse problems in measure spaces

We study Bayesian inverse problems in measure spaces, in particular, the conditions for the validity of Bayes’ formula, and the well-posedness of the problems. In [7], Cotter *et al* require the functional $\mathcal{G}$ in (1.1) to be measurable with respect to the prior measure for the validity of the Bayes formula. Restricting their consideration to Banach spaces, this requirement holds when $\mathcal{G}$ is continuous. However, this is true for any topological spaces X . Cotter *et al* [7] impose a polynomial growth on $\mathcal{G}$ to prove the well-posedness. Using Fernique’s theorem, they show that when the prior measure is Gaussian, with respect to the Hellinger metric, the posterior measure is locally Lipschitz with respect to y . We will generalize this result to measure spaces and introduce a general condition for the local Lipschitzness of the posterior measure, without imposing a Gaussian prior.

2.1. Problem formulation

For a topological space X with a σ -algebra $\mathcal{B}(X)$, we consider a functional $\mathcal{G} : X \rightarrow \mathbb{R}^m$. An observation y of $\mathcal{G}$ is subject to a multivariate Gaussian noise σ , i.e

$$y = \mathcal{G}(x) + \sigma, \quad (2.1)$$

where $\sigma \sim \mathcal{N}(0, \Sigma)$.

Let μ_0 be the prior probability measure on $(X, \mathcal{B}(X))$. Our purpose is to determine the conditional probability $\mathbb{P}(x|y)$. Let

$$\Phi(x, y) = \frac{1}{2}|y - \mathcal{G}(x)|_{\Sigma}^2. \quad (2.2)$$

We have the following theorem.

Theorem 2.1 (Cotter *et al* [7] theorem 2.1). *If $\mathcal{G} : X \rightarrow \mathbb{R}^m$ is μ_0 -measurable, then the posterior measure $\mu^y(dx) = \mathbb{P}(dx|y)$ is absolutely continuous with respect to the prior measure $\mu_0(dx)$ and the Radon–Nikodym derivative satisfies*

$$\frac{d\mu^y}{d\mu_0}(x) \propto \exp(-\Phi(x; y)). \quad (2.3)$$

Although Cotter *et al* [7] assume that X is a Banach space, their proof holds for any topological measure spaces X .

Following [7], corollary 2.2, we have the following corollary.

Corollary 2.2. *If $\mathcal{G} : X \rightarrow \mathbb{R}^m$ is continuous, then the measure $\mu^y(dx) = \mathbb{P}(dx|y)$ is absolutely continuous with respect to $\mu_0(x)$ and the Radon–Nikodym derivative satisfies (2.3).*

Proof. If $\mathcal{G} : X \rightarrow \mathbb{R}^m$ is continuous, then it is measurable ([18], lemma 1.5). The conclusion follows. $\square$

2.2. Well-posedness of the problem

For Banach spaces X , when the prior measure μ_0 is Gaussian, the well-posedness of the inverse problem is established by Cotter *et al* [7] by assuming that the function $\Phi(x; y)$ has a polynomial growth in $\|x\|_X$ fixing y and is Lipschitz in y with the Lipschitz coefficient being bounded by a polynomial of $\|x\|_X$. The proof uses Fernique's theorem. For a general measure space X , we make the following assumption.

Assumption 2.3. *The function $\Phi : X \times \mathbb{R}^m \rightarrow \mathbb{R}$ satisfies:*

- *For each $r > 0$, there is a constant $M(r) > 0$ and a set $X(r) \subset X$ of positive μ_0 measure such that for all $x \in X(r)$ and for all y with $|y| \leq r$*

$$0 \leq \Phi(x; y) \leq M(r).$$

- *There is a function $G : \mathbb{R} \times X \rightarrow \mathbb{R}$ such that for each $r > 0$, $G(r, \cdot) \in L^2(X, d\mu_0)$ and if $|y| \leq r$ and $|y'| \leq r$ then*

$$|\Phi(x; y) - \Phi(x; y')| \leq G(r, x)|y - y'|.$$

To study the well-posedness of the Bayesian inverse problems (2.1), following Cotter *et al* [7], we employ the Hellinger metric

$$d_{\text{Hell}}(\mu, \mu') = \sqrt{\frac{1}{2} \int_X \left(\sqrt{\frac{d\mu}{d\mu_0}} - \sqrt{\frac{d\mu'}{d\mu_0}} \right)^2 d\mu_0},$$

where μ and μ' are measures on $(X, \mathcal{B}(X))$. The probability measure μ^y is determined as

$$\frac{d\mu^y}{d\mu_0} = \frac{1}{Z(y)} \exp(-\Phi(x, y)),$$

where the normalization constant is

$$Z(y) = \int_X \exp(-\Phi(x, y)) d\mu_0(x). \quad (2.4)$$

With respect to the Hellinger metric, the measure μ^y depends continuously on y as shown in the following theorem.

Theorem 2.4. *Under assumption 2.3, the measure μ^y is locally Lipschitz in the data y with respect to the Hellinger metric: for each positive constant r there is a positive constant $c(r)$ such that if $|y| \leq r$ and $|y'| \leq r$, then*

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

Proof. First we show that for each $r > 0$, there is a positive constant $\alpha(r)$ such that $Z(y) \geq \alpha(r)$ when $|y| \leq r$. It is obvious from (2.4) and from assumption 2.3(i) that when $|y| \leq r$,

$$Z(y) \geq \mu_0(X(r)) \exp(-M(r)). \quad (2.5)$$

Using the inequality $|\exp(-a) - \exp(-b)| \leq |a - b|$ for $a > 0$ and $b > 0$, we have

$$|Z(y) - Z(y')| \leq \int_X |\Phi(x; y) - \Phi(x; y')| d\mu_0(x).$$

From assumption 2.3(ii), when $|y| \leq r$ and $|y'| \leq r$:

$$|\Phi(x; y) - \Phi(x; y')| \leq G(r, x)|y - y'|.$$

As $G(r, x)$ is μ_0 -integrable,

$$|Z(y) - Z(y')| \leq c(r)|y - y'|. \quad (2.6)$$

The Hellinger distance satisfies

$$\begin{aligned} 2d_{\text{Hell}}(\mu^y, \mu^{y'})^2 &= \int_X \left(Z(y)^{-1/2} \exp\left(-\frac{1}{2}\Phi(x, y)\right) \right. \\ &\quad \left. - Z(y')^{-1/2} \exp\left(-\frac{1}{2}\Phi(x, y')\right) \right)^2 d\mu_0(x) \\ &\leq I_1 + I_2, \end{aligned}$$

where

$$I_1 = \frac{2}{Z(y)} \int_X \left(\exp\left(-\frac{1}{2}\Phi(x, y)\right) - \exp\left(-\frac{1}{2}\Phi(x, y')\right) \right)^2 d\mu_0(x)$$

and

$$I_2 = 2|Z(y)^{-1/2} - Z(y')^{-1/2}|^2 \int_X \exp(-\Phi(x, y')) d\mu_0(x).$$

Using again the inequality $|\exp(-a) - \exp(-b)| \leq |a - b|$, we deduce from (2.5)

$$\begin{aligned} I_1 &\leq c(r) \int_X |\Phi(x, y) - \Phi(x, y')|^2 d\mu_0(x) \\ &\leq c(r) \int_X (G(r, x))^2 d\mu_0(x) |y - y'|^2 \leq C(r) |y - y'|^2. \end{aligned}$$

Furthermore,

$$|Z(y)^{-1/2} - Z(y')^{-1/2}|^2 = \frac{|Z(y) - Z(y')|^2}{(Z(y)^{1/2} + Z(y')^{1/2})Z(y)Z(y')}.$$

From (2.5) and (2.6),

$$|Z(y)^{-1/2} - Z(y')^{-1/2}|^2 \leq c(r) |Z(y) - Z(y')|^2 \leq c(r) |y - y'|^2.$$

The conclusion then follows. $\square$

The Kullback–Leibler distance

$$d_{\text{KL}}(\mu, \mu') = \int_X \log \left(\frac{d\mu}{d\mu'} \right) d\mu$$

is larger than the Hellinger distance. However, we can also show a similar result for the Kullback–Leibler distance of μ^y and $\mu^{y'}$.

Theorem 2.5. *Under assumption 2.3, the posterior measure μ^y is locally Lipschitz in y with respect to the Kullback–Leibler distance: for each positive constant r there is a positive constant $c(r)$ such that if $|y| \leq r$ and $|y'| \leq r$, then*

$$d_{\text{KL}}(\mu^y, \mu^{y'}) \leq c(r) |y - y'|.$$

Proof. We have

$$\begin{aligned} d_{\text{KL}}(\mu^y, \mu^{y'}) &= \int_X \left[\log \left(\frac{Z(y')}{Z(y)} \right) - (\Phi(x, y) - \Phi(x, y')) \right] d\mu^y \\ &\leq \int_X \left[\log \left(\frac{Z(y')}{Z(y)} \right) \right] d\mu^y + \int_X |\Phi(x, y) - \Phi(x, y')| d\mu^y. \end{aligned}$$

From assumption 2.3(ii), we have

$$\begin{aligned} \int_X |\Phi(x, y) - \Phi(x, y')| d\mu^y &\leq \left(\int_X G(r, x) \frac{1}{Z(y)} \exp(-\Phi(x, y)) d\mu_0 \right) |y - y'| \\ &\leq \frac{1}{\alpha(r)} \left(\int_X G(r, x) d\mu_0 \right) |y - y'| \leq c(r) |y - y'|. \end{aligned}$$

We note that

$$\frac{Z(y')}{Z(y)} \leq 1 + \left| \frac{Z(y')}{Z(y)} - 1 \right| \leq 1 + \frac{1}{\alpha(r)} Z(y) \left| \frac{Z(y')}{Z(y)} - 1 \right| = 1 + \frac{1}{\alpha(r)} |Z(y) - Z(y')|.$$

Therefore,

$$\log \left(\frac{Z(y')}{Z(y)} \right) \leq \log \left(1 + \frac{1}{\alpha(r)} |Z(y) - Z(y')| \right) \leq \frac{1}{\alpha(r)} |Z(y) - Z(y')| \leq c(r) |y - y'|$$

where the last inequality is deduced from (2.6). $\square$

Remark 2.6. To show theorem 2.5, it is sufficient to assume that $G(r, \cdot) \in L^1(X, d\mu_0)$ instead of the stronger condition $G(r, \cdot) \in L^2(X, d\mu_0)$ in assumption 2.3(ii).

2.3. MAP estimator

In infinite-dimensional spaces, a MAP estimator is identified as the points small balls with centers at which will have the largest probability in the asymptotic limit when the radius converges to 0. To show the existence of a MAP estimator, it is essential to employ asymptotic behavior of the probability of these small balls. For Gaussian probability measures in Hausdorff locally convex topological spaces, this has been well established. However, for general probability measures, similar small deviation results are still underdeveloped so we will restrict our consideration to the case of a Radon Gaussian prior probability μ_0 in a Hausdorff locally convex topological space X . The result in this section is a generalization of that for Banach spaces by Cotter *et al* [7].

We assume that the mean m of μ_0 is in the Cameron–Martin space E . Let A be a convex, symmetric about zero, weakly bounded Borel subset of X . Let x be a vector in the Cameron–Martin space E . We then have

$$\mu_0(x + rA) \sim \exp(-\|x - m\|_E^2/2) \mu_0(m + rA)$$

when $r \rightarrow 0$ (see Lifshits [20], p 270). We therefore define the probability maximizers to be the minimizers of the functional

$$I(x) = \Phi(x, y) + \frac{1}{2} \|x - m\|_E^2.$$

Indeed, when $r \rightarrow 0$, the probability of set $x + rA$ will be maximized when x is a minimizer of this functional. We have the following existence result.

Theorem 2.7. *Let X be a Hausdorff locally convex topological space and μ_0 be a Radon Gaussian measure on X with mean m in the Cameron–Martin space E . Assume that $\mathcal{G}(x)$ is continuous with respect to the topology of X . Then, there exists a minimizer in the Cameron–Martin space E for the functional I .*

Proof. As $\Phi(x, y)$ is nonnegative, the functional $I(x)$ has a nonnegative infimum. Let $\{x_n\}_n \subset E$ be a minimizing sequence of I . Then, $\|x_n\|_E$ is bounded. As μ_0 is a Radon Gaussian measure, the unit ball in the space E is compact in X (see Bogachev [5], corollary 3.2.4); therefore, we can extract a subsequence x_{n_i} that converges strongly in the topology of X to a limit $\bar{x}$ in E . On the other hand, from the Alaoglu theorem, we can choose the subsequence $\{x_{n_i}\}$ so that it converges weakly to $\bar{x}$ in the Hilbert space E , with $\|\bar{x} - m\|_E \leq \liminf_{n \rightarrow \infty} \|x_n - m\|_E$. As $\Phi(x, y)$ is continuous, we deduce that $I(\bar{x}) \leq \liminf_{n \rightarrow \infty} I(x_n)$ so $I(\bar{x}) = \inf_{x \in E} I(x)$. $\square$

Remark 2.8. Following the same procedure in Cotter *et al* [7], we can show that the subsequence converges strongly to $\bar{x}$.

3. Bayesian inverse problems for equations with white noise forcing

3.1. Stochastic Burgers and Hamilton–Jacobi equations

We consider the inviscid Burgers equation

$$\frac{\partial u}{\partial t} + (u \cdot \nabla)u = f^W(x, t), \quad (3.1)$$

where $u(x, t) \in \mathbb{R}^d$ is the velocity of a fluid particle in $\mathbb{R}^d$. The forcing function $f^W(x, t) \in \mathbb{R}^d$ depends on t via a one-dimensional white noise $\dot{W}$ and is of the form

$$f^W(x, t) = f(x)\dot{W}(t).$$

We study the case where $f(x)$ is a gradient, i.e. there is a potential $F(x)$ such that

$$f(x) = -\nabla F(x). \quad (3.2)$$

Throughout this paper, $F(x)$ is three times differentiable with bounded derivatives up to the third order. We consider the case where $u(x, t)$ is also a gradient, i.e. $u(x, t) = \nabla \phi(x, t)$. The velocity potential ϕ satisfies the Hamilton–Jacobi equation

$$\frac{\partial \phi(x, t)}{\partial t} + \frac{1}{2} |\nabla \phi(x, t)|^2 + F^W(x, t) = 0, \quad (3.3)$$

where the forcing potential $F^W(x, t)$ is

$$F^W(x, t) = F(x) \dot{W}(t).$$

The viscosity solution ϕ of (3.3) that corresponds to the solution u of the Burgers equation (3.1) is determined by the Lax–Oleinik variational principle. With the initial condition ϕ_0 at time 0, $\phi(x, t)$ is given by

$$\phi(x, t) = \inf \left\{ \phi_0(\gamma(0)) + \int_0^t \left(\frac{1}{2} |\dot{\gamma}|^2 - F^W(\gamma(\tau), \tau) \right) d\tau \right\}, \quad (3.4)$$

where the infimum is taken among all the absolutely continuous curves $\gamma : [0, t] \rightarrow \mathbb{R}^m$ such that $\gamma(t) = x$.

When the forcing function is spatially periodic, problem (3.3) possesses a unique solution (up to an additive constant) that exists for all time t (see [8, 16]). The non-compact setting is far more complicated. In [13], it is proved that when $F(x)$ has a large maximum and a small minimum, (3.3) possesses a solution that exists for all time. These imply the existence of a solution to the Burgers equation that exists for all time.

We employ the framework for measure spaces developed in section 2 to study Bayesian inverse problems for Burgers equation (3.1) and Hamilton–Jacobi equation (3.3) with white noise forcing.

3.2. Bayesian inverse problems

We study the Bayesian inverse problems that make inference on the white noise forcing given the observations at a finite set of times.

For Hamilton–Jacobi equation (3.3), fixing m spatial points x_i and m times $t_i, i = 1, \dots, m$, the function ϕ is observed at (x_i, t_i) . The observations are subject to an unbiased Gaussian noise. They are

$$z_i = \phi(x_i, t_i) + \sigma'_i,$$

where σ'_i are independent Gaussian random variables.

Since ϕ is determined within an additive constant, we assume that it is measured at a further point (x_0, t_0) where, for simplicity only, $t_0 < t_i$ for all $i = 1, \dots, m$.

Let

$$y_i = z_i - z_0 = \phi(x_i, t_i) - \phi(x_0, t_0) + \sigma'_i - \sigma'_0.$$

Let $\sigma_i = \sigma'_i - \sigma'_0$. We assume that $\sigma = \{\sigma_i : i = 1, \dots, m\} \in \mathbb{R}^m$ follows a Gaussian distribution with zero mean and covariance Σ . The vector

$$\mathcal{G}_{HJ}(W) = \{\phi(x_i, t_i) - \phi(x_0, t_0) : i = 1, \dots, m\} \in \mathbb{R}^m \quad (3.5)$$

is uniquely determined by the Wiener process W . We recover W given

$$y = \mathcal{G}_{HJ}(W) + \sigma.$$

The solution $u(x, t)$ of Burgers equation (3.1) is not defined at shocks. However, in the cases studied here, $u(\cdot, t) \in L^1_{\text{loc}}(\mathbb{R}^m)$ for all time t . We assume that noisy measurements

$$y_i = l_i(u(\cdot, t_i)) + \sigma_i, \quad i = 1, \dots, m,$$

are made for $l_i(u(\cdot, t_i))$ ($i = 1, \dots, m$) where $l_i : L^1_{\text{loc}}(\mathbb{R}^m) \rightarrow \mathbb{R}$ are continuous and bounded functionals. The noise $\sigma = (\sigma_1, \dots, \sigma_m) \sim \mathcal{N}(0, \Sigma)$. Letting

$$\mathcal{G}_B(W) = (l_1(u(\cdot, t_1)), \dots, l_m(u(\cdot, t_m))) \in \mathbb{R}^m, \quad (3.6)$$

given

$$y = \mathcal{G}_B(W) + \sigma,$$

we determine W .

Let $t_{\max} = \max\{t_1, \dots, t_m\}$. Let $\mathcal{X}$ be the metric space of continuous functions W on $(-\infty, t_{\max}]$ with $W(t_{\max}) = 0$. The space is equipped with the metric

$$D(W, W') = \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{\sup_{-n \leq t \leq t_{\max}} |W(t) - W'(t)|}{1 + \sup_{-n \leq t \leq t_{\max}} |W(t) - W'(t)|} \quad (3.7)$$

(see Stroock and Varadhan [24]). Let μ_0 be the Wiener measure on $\mathcal{X}$. Let μ^y be the conditional probability defined on $\mathcal{X}$ given y . For $\mathcal{G}_{HJ}$ in (3.5), we define the function $\Phi_{HJ} : \mathcal{X} \times \mathbb{R}^m \rightarrow \mathbb{R}$ by

$$\Phi_{HJ}(W; y) = \frac{1}{2} |y - \mathcal{G}_{HJ}(W)|_{\Sigma}^2, \quad (3.8)$$

and for $\mathcal{G}_B$ in (3.6), we define

$$\Phi_B(W; y) = \frac{1}{2} |y - \mathcal{G}_B(W)|_{\Sigma}^2. \quad (3.9)$$

We will prove that the Radon–Nikodym derivative of μ^y satisfies

$$\frac{d\mu^y}{d\mu_0} \propto \exp(-\Phi_{HJ}(W; y))$$

and

$$\frac{d\mu^y}{d\mu_0} \propto \exp(-\Phi_B(W; y)),$$

respectively, and that the posterior measure μ^y continuously depends on $y \in \mathbb{R}^m$.

4. Periodic problems

We first consider the case where the problems are set in the torus $\mathbb{T}^d$. The forcing potential $F^W(x, t)$ is spatially periodic. Randomly forced Burgers and Hamilton–Jacobi equations in a torus have been studied thoroughly by E *et al* in [8] for one dimension and by Iturriaga and Khanin in [16] for multidimensions. We first review their results.

4.1. Existence and uniqueness for periodic problems

The mean value

$$b = \int_{\mathbb{T}^d} u(x, t) \, dx$$

is unchanged for all t . We denote the solution ϕ of the Hamilton–Jacobi equation (3.3) by ϕ_b^W . It is of the form

$$\phi_b^W(x, t) = b \cdot x + \psi_b^W(x, t), \quad (4.1)$$

where ψ_b^W is spatially periodic. Let $\mathcal{C}(s, t; \mathbb{T}^d)$ be the space of absolutely continuous functions from (s, t) to $\mathbb{T}^d$. For each vector $b \in \mathbb{R}^d$, we define the operator $\mathcal{A}_{s,t}^{W,b} : \mathcal{C}(s, t; \mathbb{T}^d) \rightarrow \mathbb{R}$ as

$$\mathcal{A}_{s,t}^{W,b}(\gamma) = \int_s^t \left(\frac{1}{2} |\dot{\gamma}(\tau) - b|^2 - F^W(\gamma(\tau), \tau) - \frac{b^2}{2} \right) d\tau. \quad (4.2)$$

The variational formulation for the function $\psi_b^W(x, t)$ in (4.1) is

$$\psi_b^W(x, t) = \inf_{\gamma \in \mathcal{C}(s, t; \mathbb{T}^d)} (\psi(\gamma(s), s) + \mathcal{A}_{s,t}^{W,b}(\gamma)), \quad (4.3)$$

where $s < t$. This is written in terms of the Lax operator as

$$\mathcal{K}_{s,t}^{W,b} \psi_b^W(\cdot, s) = \psi_b^W(\cdot, t).$$

Using integration by parts, we get

$$\begin{aligned} \mathcal{A}_{s,t}^{W,b}(\gamma) = \int_s^t \left(\frac{1}{2} |\dot{\gamma}(\tau) - b|^2 + \nabla F(\gamma(\tau)) \cdot \dot{\gamma}(\tau) (W(\tau) - W(t)) - \frac{b^2}{2} \right) d\tau \\ - F(\gamma(s)) (W(t) - W(s)). \end{aligned} \quad (4.4)$$

We study the functional $\mathcal{A}_{s,t}^{W,b}$ via minimizers which are defined as follows [16].

Definition 4.1. A curve $\gamma : [s, t] \rightarrow \mathbb{R}^d$ such that $\gamma(s) = x$ and $\gamma(t) = y$ is called a minimizer over $[s, t]$ if it minimizes the action $\mathcal{A}_{s,t}^{W,b}$ among all the absolutely continuous curves with endpoints at x and y at times s and t , respectively.

For a function ψ , a curve $\gamma : [s, t] \rightarrow \mathbb{R}^d$ with $\gamma(t) = x$ is called a ψ minimizer over $[s, t]$ if it minimizes the action $\psi(\gamma(s)) + \mathcal{A}_{s,t}^{W,b}(\gamma)$ among all the curves with endpoint at x at time t .

A curve $\gamma : (-\infty, t] \rightarrow \mathbb{R}^d$ with $\gamma(t) = x$ is called a one-sided minimizer if it is a minimizer over all the time intervals $[s, t]$.

Iturriaga and Khanin [16] proved the following theorem.

Theorem 4.2 ([16], theorem 1).

- (i) For almost all W , all $b \in \mathbb{R}^d$ and $T \in \mathbb{R}$, there exists a unique (up to an additive constant) function $\phi_b^W(x, t)$, $x \in \mathbb{R}^m$, $t \in (-\infty, T]$ such that

$$\phi_b^W(x, t) = b \cdot x + \psi_b^W(x, t),$$

where $\psi_b^W(\cdot, t)$ is a $\mathbb{T}^d$ periodic function and for all $-\infty < s < t < T$,

$$\psi_b^W(\cdot, t) = \mathcal{K}_{s,t}^{W,b} \psi_b^W(\cdot, s).$$

- (ii) The function $\psi_b^W(\cdot, t)$ is Lipschitz in $\mathbb{T}^d$. If x is a point of differentiability of $\psi_b^W(\cdot, t)$, then there exists a unique one-sided minimizer $\gamma_{x,t}^{W,b}$ at (x, t) and its velocity is given by the gradient of ϕ : $\dot{\gamma}_{x,t}^{W,b}(t) = \nabla \phi_b^W(x, t) = b + \nabla \psi_b^W(x, t)$. Further, any one-sided minimizer is a ψ_b^W minimizer on finite-time intervals.

- (iii) For almost all W and all $b \in \mathbb{R}^d$:

$$\lim_{s \rightarrow -\infty} \sup_{\eta \in \mathcal{C}(\mathbb{T}^d)} \min_{C \in \mathbb{R}} \max_{x \in \mathbb{T}^d} |\mathcal{K}_{s,t}^{W,b} \eta(x) - \psi_b^W(x, t) - C| = 0.$$

- (iv) The unique solution $u_b^W(x, t)$ that exists for all time of the Burgers equation (3.1) is determined by $u_b^W(\cdot, t) = \nabla \phi_b^W(\cdot, t)$.

We will employ these results to study the Bayesian inverse problems formulated in section 2.

4.2. Bayesian inverse problem for the spatially periodic Hamilton–Jacobi equation (3.3)

We first study the Bayesian inverse problem for Hamilton–Jacobi equation (3.3) in the torus $\mathbb{T}^d$. The observation $\mathcal{G}_{HJ}(W)$ in (3.5) becomes

$$\mathcal{G}_{HJ}(W) = \{\psi_b^W(x_i, t_i) - \psi_b^W(x_0, t_0) : i = 1, \dots, m\}.$$

To establish the validity of the Bayesian formula (2.3), we first show that $\mathcal{G}_{HJ}(W)$ as a map from $\mathcal{X}$ to $\mathbb{R}^m$ is continuous with respect to the metric D in (3.7).

The following proposition holds.

Proposition 4.3. *The map $\mathcal{G}_{HJ}(W) : \mathcal{X} \rightarrow \mathbb{R}^m$ is continuous with respect to the metric (3.7).*

Proof. Let W_k converge to W in the metric (3.7). There are constants C_k which do not depend on i such that

$$\lim_{k \rightarrow \infty} |\phi_b^{W_k}(x_i, t_i) - \phi_b^W(x_i, t_i) - C_k| \rightarrow 0 \quad (4.5)$$

for all $i = 0, \dots, m$. The proof of (4.5) is technical; we present it in section 7.1. From this we deduce that

$$\lim_{k \rightarrow \infty} |(\phi_b^{W_k}(x_i, t_i) - \phi_b^{W_k}(x_0, t_0)) - (\phi_b^W(x_i, t_i) - \phi_b^W(x_0, t_0))| = 0,$$

so

$$\lim_{k \rightarrow \infty} |\mathcal{G}_{HJ}(W_k) - \mathcal{G}_{HJ}(W)| = 0.$$

The conclusion then follows. $\square$

To establish the continuous dependence of μ^y on y , we verify assumption 2.3. First we prove a bound for $\mathcal{G}_{HJ}(W)$.

Lemma 4.4. *There is a positive constant c which only depends on the potential F such that*

$$|\mathcal{G}_{HJ}(W)| \leq c \left(1 + \sum_{i=1}^m \max_{t_0-1 \leq \tau \leq t_i} |W(\tau) - W(t_i)|^2 \right).$$

Proof. Let γ be a $\psi_b^W(\cdot, t_0 - 1)$ minimizer starting at (x_i, t_i) . Then,

$$\psi_b^W(x_i, t_i) = \psi_b^W(\gamma(t_0 - 1), t_0 - 1) + \mathcal{A}_{t_0-1, t_i}^{W, b}(\gamma).$$

From (4.4), there is a constant c which depends only on b and F such that

$$\begin{aligned} \mathcal{A}_{t_0-1, t_i}^{W, b}(\gamma) &\geq \frac{1}{2(t_i - t_0 + 1)} \left(\int_{t_0-1}^{t_i} |\dot{\gamma}(\tau)| d\tau \right)^2 \\ &\quad - c \left(1 + \max_{t_0-1 \leq \tau \leq t_i} |W(\tau) - W(t_i)| \right) \int_{t_0-1}^{t_i} |\dot{\gamma}(\tau)| d\tau \\ &\quad - c(1 + |W(t_i) - W(t_0 - 1)|). \end{aligned}$$

The right-hand side is a quadratic form of $\int_{t_0-1}^{t_i} |\dot{\gamma}(\tau)| d\tau$ so

$$\mathcal{A}_{t_0-1, t_i}^{W, b}(\gamma) \geq -c \left(1 + \max_{t_0-1 \leq \tau \leq t_i} |W(\tau) - W(t_i)|^2 \right).$$

Therefore,

$$\psi_b^W(x_i, t_i) \geq \psi_b^W(\gamma(t_0 - 1), t_0 - 1) - c \left(1 + \max_{t_0-1 \leq \tau \leq t_i} |W(\tau) - W(t_i)|^2 \right).$$

Let γ' be the linear curve connecting (x_0, t_0) and $(\gamma(t_0 - 1), t_0 - 1)$. We have

$$\begin{aligned}\psi_b^W(x_0, t_0) &\leq \psi_b^W(\gamma(t_0 - 1), t_0 - 1) + \mathcal{A}_{t_0-1, t_0}^{W, b}(\gamma') \\ &\leq \psi_b^W(\gamma(t_0 - 1), t_0 - 1) + c(1 + \max_{t_0-1 \leq \tau \leq t_0} |W(\tau) - W(t_0)|).\end{aligned}$$

Therefore,

$$\psi_b^W(x_i, t_i) - \psi_b^W(x_0, t_0) \geq -c(1 + \max_{t_0-1 \leq \tau \leq t_i} |W(\tau) - W(t_i)|^2). \quad (4.6)$$

Let $\bar{\gamma}$ be the linear curve connecting (x_0, t_0) and (x_i, t_i) . We then find that

$$\psi_b^W(x_i, t_i) \leq \psi_b^W(x_0, t_0) + \mathcal{A}_{t_0, t_i}^{W, b}(\bar{\gamma}).$$

Thus,

$$\psi_b^W(x_i, t_i) \leq \psi_b^W(x_0, t_0) + c(1 + \max_{t_0 \leq \tau \leq t_i} |W(\tau) - W(t_i)|). \quad (4.7)$$

From (4.6) and (4.7), we conclude that there exists a positive constant c such that

$$|\psi_b^W(x_i, t_i) - \psi_b^W(x_0, t_0)| \leq c(1 + \max_{t_0-1 \leq \tau \leq t_i} |W(\tau) - W(t_i)|^2). \quad (4.8)$$

Therefore,

$$|\mathcal{G}_{HJ}(W)| \leq c \left(1 + \sum_{i=1}^m \max_{t_0-1 \leq \tau \leq t_i} |W(\tau) - W(t_i)|^2 \right). \quad (4.9) \quad \square$$

Proposition 4.5. *Assumption 2.3 holds.*

Proof. Fixing a constant M , from (4.9) there is a set of positive Wiener measures so that

$$|\mathcal{G}_{HJ}(W)| \leq M.$$

Therefore, there is a constant $M(r)$ and a set of positive measures such that when $|y| \leq r$

$$|\Phi_{HJ}(W; y)| = \frac{1}{2} |y - \mathcal{G}_{HJ}(W)|_{\Sigma}^2 \leq \frac{1}{2} \|\Sigma^{-1/2}\|_{\mathbb{R}^m, \mathbb{R}^m}^2 (|y| + M)^2 \leq M(r).$$

Assumption 2.3(i) holds.

Now we prove assumption 2.3(ii). We have

$$\begin{aligned}|\Phi_{HJ}(W; y) - \Phi_{HJ}(W; y')| &= \frac{1}{2} \left| \langle \Sigma^{-1/2}(y - \mathcal{G}_{HJ}(W)), \Sigma^{-1/2}(y - y') \rangle \right. \\ &\quad \left. + \langle \Sigma^{-1/2}(y - y'), \Sigma^{-1/2}(y' - \mathcal{G}_{HJ}(W)) \rangle \right| \\ &\leq \frac{1}{2} \|\Sigma^{-1/2}\|_{\mathbb{R}^m, \mathbb{R}^m}^2 (|y| + |y'| + 2|\mathcal{G}_{HJ}(W)|) |y - y'|.\end{aligned} \quad (4.10)$$

Letting

$$G(r, W) = \|\Sigma^{-1/2}\|_{\mathbb{R}^m, \mathbb{R}^m}^2 \left[r + c \left(1 + \sum_{i=1}^m \max_{t_0-1 \leq \tau \leq t_i} |W(\tau) - W(t_i)|^2 \right) \right],$$

obviously $G(r, \cdot) \in L^2(\mathcal{X}, \mu_0)$. $\square$

From propositions 4.3, 4.5, corollary 2.2 and theorem 2.4 we have

Theorem 4.6. *For the spatially periodic randomly forced Hamilton–Jacobi equation (3.3), the measure $\mu^y(dW) = \mathbb{P}(dW|y)$ is absolutely continuous with respect to the Wiener measure $\mu_0(dW)$; the Radon–Nikodym derivative satisfies*

$$\frac{d\mu^y}{d\mu_0} \propto \exp(-\Phi_{HJ}(W; y)).$$

When $|y| \leq r$ and $|y'| \leq r$, there is a constant $c(r)$ such that

$$d_{\text{Hell}}(\mu^y, \mu^{y'}) \leq c(r)|y - y'| \quad \text{and} \quad d_{\text{KL}}(\mu^y, \mu^{y'}) \leq c(r)|y - y'|.$$

We establish the existence of a MAP estimator in the following theorem.

Theorem 4.7. *For the spatially periodic randomly forced Hamilton–Jacobi equation (3.3), there exists a MAP estimator for the posterior measure μ^y .*

Proof. Iturriaga and Khanin [16] establish the almost sure existence of a solution for the Hamilton–Jacobi equation with white noise forcing in periodic domains. A close examination of their proofs shows that for a solution to exist, it requires the velocity of finite-time minimizers to be uniformly bounded, and the existence of an ε -narrow place that satisfies the hypothesis of lemma 7.1, i.e. the existence of an interval $[s_1, s_2]$ of length T such that

$$\max_{s_1 \leq \tau \leq s_2} |W(\tau) - W(s_2)| \leq \frac{1}{T^2}$$

for T arbitrarily large.

Iturriaga and Khanin [16] show that the velocity is uniformly bounded as long as W is continuous. This condition is satisfied by all the functions W in the Cameron–Martin space of the Wiener space $\mathcal{X}$.

For ε narrow places, we note that when W is in the Cameron–Martin space, for a fixed T ,

$$\max_{s_1 \leq \tau \leq s_2} |W(\tau) - W(s_2)| \leq \int_{s_1}^{s_2} |\dot{W}(\tau)| d\tau \leq T^{1/2} \left(\int_{s_1}^{s_2} |\dot{W}(\tau)|^2 d\tau \right)^{1/2},$$

which can be arbitrarily small when s_2 is small as $\int_{-\infty}^0 |\dot{W}(\tau)|^2 d\tau$ is finite. This shows the existence of an ε -narrow place. Therefore, there is a unique solution that exists for all time for the forcing realizations in the Cameron–Martin space. The Wiener space $\mathcal{X}$ with the metric 3.7 is Frechet, so all the requirements of theorem 2.7 hold. Thus, there exists a MAP estimator. $\square$

4.3. Bayesian inverse problem for the spatially periodic Burgers equation

We consider the Bayesian inverse problem for Burgers equation (3.1) in the torus $\mathbb{T}^d$. We first establish the continuity of

$$\mathcal{G}_B(W) = (l_1(u_b^W(\cdot, t_1)), \dots, l_m(u_b^W(\cdot, t_m)))$$

as a map from $\mathcal{X}$ to $\mathbb{R}^m$ to prove the validity of the Bayesian formula (2.3).

First we prove the uniform boundedness of $|\dot{\gamma}(t)|$ for any minimizers γ on $[t-1, t]$. The result is proved in [8] and [16], using Gronwall’s inequality. We present another proof without using this inequality, and establish explicitly the dependence of the bound on W .

Lemma 4.8. *For any minimizer γ of $\mathcal{A}_{t-1,t}^{W,b}$,*

$$|\dot{\gamma}(t)| \leq c \left(1 + \max_{t-1 \leq \tau \leq t} |W(\tau) - W(t)|^2 \right).$$

Proof. $\bar{\gamma}$ be the linear curve connecting $(\gamma(t), t)$ and $(\gamma(t-1), t-1)$. As $|\gamma(t) - \gamma(t-1)| \leq d^{1/2}$,

$$\mathcal{A}_{t-1,t}^{W,b}(\bar{\gamma}) \leq c \left(1 + \max_{t-1 \leq \tau \leq t} |W(\tau) - W(t)| \right).$$

We have

$$\begin{aligned} \mathcal{A}_{t-1,t}^{W,b}(\gamma) &\geq \frac{1}{2} \left(\int_{t-1}^t |\dot{\gamma}(\tau)| d\tau \right)^2 - c \left(1 + \max_{t-1 \leq \tau \leq t} |W(\tau) - W(t)| \right) \int_{t-1}^t |\dot{\gamma}(\tau)| d\tau \\ &\quad - c \left(1 + |W(t-1) - W(t)| \right). \end{aligned}$$

Because $\mathcal{A}_{t-1,t}^{W,b}(\gamma) \leq \mathcal{A}_{t-1,t}^{W,b}(\tilde{\gamma})$, solving the quadratic equation, we get

$$\int_{t-1}^t |\dot{\gamma}(\tau)| d\tau \leq c(1 + \max_{t-1 \leq \tau \leq t} |W(\tau) - W(t)|). \quad (4.11)$$

The minimizer γ satisfies (see [8] and [16])

$$\dot{\gamma}(t) = \dot{\gamma}(s) + \int_s^t \nabla^2 F(\gamma(\tau)) \dot{\gamma}(\tau) (W(\tau) - W(t)) d\tau + \nabla F(\gamma(s)) (W(s) - W(t)) \quad (4.12)$$

for all $t-1 \leq s \leq t$. Therefore,

$$\begin{aligned} |\dot{\gamma}(s)| &\geq |\dot{\gamma}(t)| - c \max_{t-1 \leq \tau \leq t} |W(\tau) - W(t)| \left(1 + \int_{t-1}^t |\dot{\gamma}(\tau)| d\tau\right) \\ &\geq |\dot{\gamma}(t)| - c \left(1 + \max_{t-1 \leq \tau \leq t} |W(\tau) - W(t)|^2\right). \end{aligned}$$

From this and (4.11), we deduce that

$$\begin{aligned} |\dot{\gamma}(t)| &\leq c \left(1 + \max_{t-1 \leq \tau \leq t} |W(\tau) - W(t)|^2\right) + c(1 + \max_{t-1 \leq \tau \leq t} |W(\tau) - W(t)|) \\ &\leq c \left(1 + \max_{t-1 \leq \tau \leq t} |W(\tau) - W(t)|^2\right). \end{aligned}$$

□

We have the following lemma on the Lipschitzness of the solution $\phi_b^W(\cdot, t)$.

Lemma 4.9. *The solution $\phi_b^W(\cdot, t)$ satisfies*

$$|\phi_b^W(x, t) - \phi_b^W(x', t)| \leq c \left(1 + \max_{t-1 \leq \tau \leq t} |W(\tau) - W(t)|^2\right) |x - x'|.$$

The lemma is proved in [16], but to show the explicit Lipschitz constant, we briefly mention the proof here.

Proof. Let γ be a one-sided minimizer starting at (x, t) . We consider the curve

$$\tilde{\gamma}(\tau) = \gamma(\tau) + (x' - x)(\tau - t + 1),$$

(modulus $\mathbb{T}^d$) connecting x' and $\gamma(t-1)$. As

$$\psi_b^W(x, t) = \psi_b^W(\gamma(t-1), t-1) + \mathcal{A}_{t-1,t}^{W,b}(\gamma)$$

and

$$\psi_b^W(x', t) \leq \phi_b^W(\gamma(t-1), t-1) + \mathcal{A}_{t-1,t}^{W,b}(\tilde{\gamma}),$$

we have

$$\begin{aligned} \psi_b^W(x', t) - \psi_b^W(x, t) &\leq \mathcal{A}_{t-1,t}^{W,b}(\tilde{\gamma}) - \mathcal{A}_{t-1,t}^{W,b}(\gamma) \\ &\leq c \left[1 + \max_{t-1 \leq s \leq t} |W(s) - W(t)| \left(1 + \int_{t-1}^t |\dot{\gamma}(\tau)| d\tau\right)\right] |x - x'|. \end{aligned}$$

Similarly, the same estimate holds for $\psi_b^W(x, t) - \psi_b^W(x', t)$. From (4.11), we have

$$|\phi_b^W(x, t) - \phi_b^W(x', t)| \leq c \left(1 + \max_{t-1 \leq \tau \leq t} |W(\tau) - W(t)|^2\right) |x - x'|.$$

□

We now establish the continuity of $\mathcal{G}_B(W)$ as a map from $\mathcal{X}$ to $\mathbb{R}^m$.

Proposition 4.10. *The function $\mathcal{G}_B(W) : \mathcal{X} \rightarrow \mathbb{R}^m$ is continuous with respect to the metric D in (3.7).*

Proof. From lemma 4.9, when $W_k \rightarrow W$ in the metric D , $\phi_b^{W_k}$ is uniformly Lipschitz in $C^{0,1}(\mathbb{T}^d)/\mathbb{R}$. We can therefore extract a subsequence that converges in $C^{0,1}(\mathbb{T}^d)/\mathbb{R}$. From (4.5), every subsequence converges to ϕ_b^W . Thus, $u_b^{W_k} \rightarrow u_b^W$ whenever $\phi_b^{W_k}$ and ϕ_b^W are differentiable ($\phi_b^{W_k}$ and ϕ_b^W are differentiable almost everywhere). From lemma 4.9, $|u_b^{W_k}(x, t)|$ are uniformly bounded. The dominated convergence theorem implies that $u_b^{W_k}(\cdot, t_i) \rightarrow u_b^W(\cdot, t_i)$ in $L^1(\mathbb{T})$ for $i = 1, \dots, m$. $\square$

For the well-posedness of the Bayesian inverse problem, we prove assumption 2.3.

Proposition 4.11. *For the spatially periodic randomly forced Burgers equation (3.1), assumption 2.3 holds.*

Proof. From lemma 4.9, we deduce that

$$|\mathcal{G}_B(W)| \leq c \left(1 + \sum_{i=1}^m \max_{t_{i-1} \leq \tau \leq t_i} |W(\tau) - W(t_i)|^2 \right). \quad (4.13)$$

Therefore, for each positive constant M , there exists a subset of $\mathcal{X}$ of positive Wiener measure such that for all W in that set: $|\mathcal{G}(W)| \leq M$. By the same argument as in the proof of proposition 4.5, assumption 2.3(i) holds.

From (4.10) and (4.13), assumption 2.3(ii) holds with

$$G(r, W) = \|\Sigma^{-1/2}\|_{\mathbb{R}^m, \mathbb{R}^m}^2 \left[r + c \left(1 + \sum_{i=1}^m \max_{t_{i-1} \leq \tau \leq t_i} |W(\tau) - W(t_i)|^2 \right) \right],$$

which is in $L^2(\mathcal{X}, d\mu_0)$. $\square$

From propositions 4.10 and 4.11, corollary 2.2 and theorem 2.4, we have

Theorem 4.12. *For the spatially periodic randomly forced Burgers equation (3.1), the posterior measure μ^y is absolutely continuous with respect to the prior measure μ_0 and satisfies*

$$\frac{d\mu^y}{d\mu_0} \propto \exp(-\Phi_B(W; y)),$$

where $\Phi_B(W; y)$ is defined in (3.9), and is locally Lipschitz in y with respect to the Hellinger and Kullback–Leibler distances, i.e.

$$d_{\text{Hell}}(\mu^y, \mu^{y'}) \leq c(r)|y - y'|, \quad \text{and} \quad d_{\text{KL}}(\mu^y, \mu^{y'}) \leq c(r)|y - y'|$$

when $|y| \leq r$ and $|y'| \leq r$; $c(r)$ depends on r .

Similarly to the Hamilton–Jacobi equation, we can show that there exists a MAP estimator for the posterior measure μ^y .

Theorem 4.13. *For the spatially periodic randomly forced Burgers equation (3.1), there is a MAP estimator for the posterior measure μ^y .*

5. Non-periodic problems

We consider Bayesian inverse problems making inference on white noise forcing for Burgers and Hamilton–Jacobi equations in non-compact domains. Burgers and Hamilton–Jacobi equations with white noise forcing are reasonably understood when the forcing potential $F(x)$ has a large maximum and a small minimum (Hoang and Khanin [13]). We first present the setting up of the problem and introduce the basic results that will be used.

5.1. Existence and uniqueness for non-periodic problems

Let $W(t)$ be a standard Brownian motion starting at 0 at the time 0. Let

$$E_1 = \mathbb{E}\{|W(1)|\},$$

where $\mathbb{E}$ denotes expectation with respect to the Wiener measure. For all $0 < t < 1$, there is a constant E_2 not depending on t such that

$$\mathbb{E}\left\{\max_{1-t \leq \tau \leq 1} |W(\tau) - W(1)|\right\} = E_2 t^{1/2}.$$

We also denote

$$E_3 = \mathbb{E}\left\{\max_{0 \leq \tau \leq 1} |W(\tau) - W(1)|^2\right\}.$$

Following Hoang and Khanin [13], we make the following assumption.

Assumption 5.1.

(i) The forcing potential F has a maximum at $x^{\max}$ and a minimum at $x^{\min}$ such that

$$F(x^{\max}) > L \text{ and } F(x^{\min}) < -L; \quad |x^{\max}| < a, \quad |x^{\min}| < a,$$

where a and L are positive constants; there is a constant $b > a$ such that

$$|\nabla F(x)| < K \text{ when } |x| < b$$

and for $x \geq b$,

$$|\nabla F(x)| < K_1 \ll K \text{ and } |F(x)| \ll L_1 \ll L.$$

(ii) The function $F(x)$ is three times differentiable; $|\nabla^2 F(x)|$ and $|\nabla^3 F(x)|$ are uniformly bounded for all x .

(iii) The constants L, K, L_1, K_1 satisfy

$$\frac{128a^4 E_2^2 K^2}{E_1^3} < L^3, \quad 8a^2 < LE_1, \quad K_1^2 < \frac{LE_1}{3E_3} \text{ and } L_1 < \frac{L}{16}.$$

We denote by $B(b)$ the ball with center 0 and radius b .

The solution of the Hamilton–Jacobi equation (3.3), denoted by ϕ^W in this case, satisfies the Lax–Oleinik formula

$$\phi^W(x, t) = \inf_{\gamma \in \mathcal{C}(s, t; \mathbb{R}^d)} \{\phi(\gamma(s), s) + \mathcal{A}_{s, t}^W(\gamma)\}, \quad (5.1)$$

where $\mathcal{C}(s, t; \mathbb{R}^d)$ is the space of absolutely continuous functions from (s, t) to $\mathbb{R}^d$; the operator $\mathcal{A}_{s, t}^W$ is defined as

$$\begin{aligned} \mathcal{A}_{s, t}^W(\gamma) = & \int_s^t \left(\frac{1}{2} |\dot{\gamma}(\tau)|^2 + \nabla F(\gamma(\tau)) \dot{\gamma}(\tau) (W(\tau) - W(t)) \right) d\tau \\ & + F(\gamma(s)) (W(s) - W(t)). \end{aligned} \quad (5.2)$$

We denote (5.1) as

$$\phi^W(x, t) = \mathcal{K}_{s, t}^W \phi^W(\cdot, s).$$

Minimizers for the operator $\mathcal{A}_{s, t}^W$ are defined as in definition 4.1.

Under assumption 5.1, the following existence and uniqueness results for the Hamilton–Jacobi equation (3.3) and Burgers equation (3.1) are established in [13].

Theorem 5.2.

- (i) With probability 1, there is a solution ϕ^W to the Hamilton–Jacobi equation (3.3) such that for all $t > 0$, $\phi^W(\cdot, t) = \mathcal{K}_{s,t}^W \phi^W(\cdot, s)$ for all $s < t$. The solution $\phi^W(\cdot, t)$ is locally Lipschitz for all t . Let $\phi_s^W(x, t)$ be the solution to the Hamilton–Jacobi equation with the zero initial condition at time s ; then, there is a constant C_s such that

$$\lim_{s \rightarrow -\infty} \sup_{|x| \leq M} |\phi^W(x, t) - \phi_s^W(x, t) - C_s| = 0$$

for all $M > 0$. The solution is unique in this sense up to an additive constant.

- (ii) For every (x, t) , there is a one-sided minimizer $\gamma_{x,t}^W$ starting at (x, t) , and there is a sequence of time s_i that converges to $-\infty$, and zero minimizers γ_{s_i} on $[s_i, t]$ starting at (x, t) such that γ_{s_i} converge uniformly to $\gamma_{x,t}^W$ on any compact time intervals when $s_i \rightarrow -\infty$; further $\gamma_{x,t}^W$ is a ϕ^W minimizer on any finite time interval. If $x \in \mathbb{R}^m$ is a point of differentiability of $\phi^W(\cdot, t)$ then such a one-sided minimizer starting at x at time t is unique and $\dot{\gamma}_{x,t}^W(t) = \nabla \phi^W(x, t)$.
- (iii) The function $u^W(\cdot, t) = \nabla \phi^W(\cdot, t)$ is a solution of the Burgers equation (3.1) that exists for all time. It is the limit of the solutions to (3.1) on finite time intervals with the zero initial condition. In this sense, the solution u^W is unique.

From now on, by one-sided minimizers, we mean those that are limits of zero minimizers on finite time intervals as in theorem 5.2(ii).

The key fact that Hoang and Khanin [13] used to prove these results is that when $t - s$ is sufficiently large, any zero minimizers starting at (x, t) over a time interval (s, t) will be inside the ball $B(b)$ at a time in an interval $(\bar{t}(W), t)$ where $\bar{t}(W)$ only depends on W , x and t .

Let

$$\alpha = \min \{1, L^2 E_1^2 / (16a^2 K^2 E_2^2)\}.$$

We define for each integer l the random variable

$$P_l = \frac{2a^2}{\alpha} + 2aK \max_{l+1-\bar{t} \leq \tau \leq l+1} |W(\tau) - W(l+1)| - L|W(l) - W(l+1)| \\ + L_1|W(l) - W(l+1)| + \frac{K_1^2}{2} \max_{l \leq \tau \leq l+1} |W(\tau) - W(l+1)|^2. \quad (5.3)$$

These random variables are independent and identically distributed. Their expectation satisfies

$$\mathbb{E}\{P_l\} < -\frac{LE_1}{2}, \quad (5.4)$$

which can be shown from assumption 5.1(iii) (see [13], p 830).

Let t'_i and t''_i be the integers that satisfy

$$t_i \leq t'_i < t_i + 1 \quad \text{and} \quad t_i - 2 < t''_i \leq t_i - 1. \quad (5.5)$$

The following result is shown in [13].

Proposition 5.3 ([13], pp 829–31). *Fixing an index i , there are constants $c(x_i, t_i, F)$ and $c(F)$ such that if $\bar{t}_i(W)$ is the largest integer for which*

$$\sum_{l=\bar{t}}^{t''_i-1} P_l > -c(x_i, t_i, F) \left(1 + \sum_{j=t''_0}^{t'_i} \max_{j \leq \tau \leq j+1} |W(\tau) - W(j+1)|^2 \right) \\ - c(F) \left(1 + \max_{\bar{t}-2 \leq \tau \leq \bar{t}} |W(\tau) - W(\bar{t})|^2 \right), \quad (5.6)$$

for all integers $\bar{t} \leq \bar{t}_i(W) + 2$, then when s is sufficiently small, every zero minimizer over (s, t_i) starting at (x_i, t_i) is inside the ball $B(b)$ at a time $s' \in [\bar{t}_i(W), t''_0]$.

Indeed, Hoang and Khanin [13] prove the result for $s' \in [\bar{t}_i(W), t_i - 1]$ but the proof for the requirement $s' \in [\bar{t}_i(W), t_0'']$ is identical. The law of large numbers guarantees the existence of such a value $\bar{t}_i(W)$. It is obvious that all one-sided minimizers in theorem 5.2(ii) are in $B(b)$ at a time $s' \in [\bar{t}_i(W), t_0'']$.

We employ these results to prove the validity of the Bayes formula and the well-posedness of the Bayesian inverse problems for the Burgers and Hamilton–Jacobi equations.

5.2. Bayesian inverse problem for the Hamilton–Jacobi equation (3.3)

We consider Hamilton–Jacobi equation (3.3) with the forcing potential F satisfying assumption 5.1. To prove the validity of the Bayes formula, we show the continuity of the function

$$\mathcal{G}_{HJ}(W) = (\phi^W(x_1, t_1) - \phi^W(x_0, t_0), \dots, \phi^W(x_m, t_m) - \phi^W(x_0, t_0)).$$

Proposition 5.4. *The function $\mathcal{G}_{HJ}$ is continuous as a map from $\mathcal{X}$ to $\mathbb{R}^m$ with respect to the metric D in (3.7).*

Proof. Let W_k converge to W in the metric D of $\mathcal{X}$. As shown in section 7.2, there are constants C_k which do not depend on i such that

$$\lim_{k \rightarrow \infty} |\phi^{W_k}(x_i, t_i) - \phi^W(x_i, t_i) - C_k| = 0 \quad (5.7)$$

for all $i = 1, \dots, m$. From this,

$$\lim_{k \rightarrow \infty} |(\phi^{W_k}(x_i, t_i) - \phi^{W_k}(x_0, t_0)) - (\phi^W(x_i, t_i) - \phi^W(x_0, t_0))| = 0.$$

The conclusion then follows. $\square$

To show the continuous dependence of the posterior measure on the noisy data y , we show assumption 2.3. We first prove the following lemmas.

Lemma 5.5. *There is a constant c that depends on x_i, x_0, t_i, t_0 and F such that*

$$\phi^W(x_i, t_i) - \phi^W(x_0, t_0) \leq c \left(1 + \max_{t_0 \leq \tau \leq t_i} |W(\tau) - W(t_i)| \right).$$

Proof. Let $\bar{\gamma}$ be the linear curve that connects (x_i, t_i) with (x_0, t_0) . We have

$$\phi^W(x_i, t_i) \leq \phi^W(x_0, t_0) + \mathcal{A}_{t_0, t_i}^W(\bar{\gamma}).$$

Since $\dot{\bar{\gamma}}(\tau) = (x_i - x_0)/(t_i - t_0)$, from (5.2)

$$\begin{aligned} |\mathcal{A}_{t_0, t_i}^W| &\leq \frac{|x_i - x_0|^2}{2|t_i - t_0|} + \max_x |\nabla F(x)| \max_{t_0 \leq \tau \leq t_i} |W(\tau) - W(t_i)| |x_i - x_0| \\ &\quad + \max_x |F(x)| |W(t_i) - W(t_0)|. \end{aligned}$$

The conclusion then follows. $\square$

Lemma 5.6. *The following estimate holds*

$$\phi^W(x_i, t_i) - \phi^W(x_0, t_0) \geq -c(x_0, x_i, F) \sum_{l=\bar{t}_i(W)}^{t_i'} \left(1 + \max_{l \leq \tau \leq l+1} |W(\tau) - W(l+1)|^2 \right),$$

where the constant $c(x_0, x_i, F)$ depends on x_0, x_i and F .

Proof. We prove this lemma in section 7.3. $\square$

From these two lemmas, we have

Proposition 5.7. *Assumption 2.3 holds.*

Proof. From lemmas 5.5 and 5.6,

$$|\mathcal{G}_{HJ}(W)| \leq S(W),$$

where

$$S(W) = c + c \sum_{i=1}^m \max_{t_0 \leq \tau \leq t_i} |W(\tau) - W(t_i)| + c \sum_{i=1}^m \sum_{l=\bar{t}_i(W)}^{t'_i} \left(1 + \max_{l \leq \tau \leq l+1} |W(\tau) - W(l+1)|^2 \right),$$

where the constants do not depend on the Brownian path W .

Let I be a positive integer such that the subset $\mathcal{X}_I$ containing all $W \in \mathcal{X}$ that satisfy $\bar{t}_i(W) > I$ for all $i = 1, \dots, m$ has a positive probability. There is a constant $\bar{c}$ such that the subset containing the paths $W \in \mathcal{X}_I$ that satisfy

$$S(W) < \bar{c}$$

has a positive probability. Assumption 2.3(i) holds.

To prove assumption 2.3(ii), we get from (4.10)

$$|\Phi_{HJ}(W; y) - \Phi_{HJ}(W; y')| \leq \frac{1}{2} \|\Sigma^{-1/2}\|_{\mathbb{R}^m, \mathbb{R}^m}^2 (|y| + |y'| + 2|\mathcal{G}_{HJ}(W)|) |y - y'|.$$

Let

$$G(r, W) = \|\Sigma^{-1/2}\|_{\mathbb{R}^m, \mathbb{R}^m}^2 (r + S(W)). \quad (5.8)$$

We prove that $G(r, W)$ is in $L^2(\mathcal{X}, \mu_0)$ in 7.4. $\square$

From propositions 5.4, 5.7, corollary 2.2 and theorem 2.4, we get

Theorem 5.8. *Under assumption 5.1 for the Hamilton–Jacobi equation (3.3), the posterior probability measure $\mu^y(dW) = \mathbb{P}(dW|y)$ is absolutely continuous with respect to the Wiener measure $\mu_0(dW)$; its Radon–Nikodym derivative satisfies*

$$\frac{d\mu^y}{d\mu_0} \propto \exp(-\Phi_{HJ}(W; y))$$

where Φ_{HJ} is defined in (3.8). For $|y| \leq r$ and $|y'| \leq r$, there is a constant $c(r)$ depending on r such that

$$d_{\text{Hell}}(\mu^y, \mu^{y'}) \leq c(r)|y - y'| \text{ and } d_{KL}(\mu^y, \mu^{y'}) \leq c(r)|y - y'|.$$

We now remark on the existence of a MAP estimator for the Hamilton–Jacobi equations in the noncompact setting. This remark also holds for the Burgers equation considered below.

Remark 5.9. The existence of a solution for the Hamilton–Jacobi and Burgers equations in [13] is almost sure. Unlike in the periodic case, the proof for the noncompact setting is purely probabilistic, where the law of a large number plays a central role. It is not clear whether a solution exists when W belongs to the Cameron–Martin space. Therefore, the existence of a MAP estimator for this case is open.

5.3. Bayesian inverse problem for the Burgers equation

We consider Burgers equation (3.1) in the non-compact setting. We first prove that

$$\mathcal{G}_B(W) = (l_1(u(\cdot, t_1)), \dots, l_m(u(\cdot, t_m)))$$

is continuous with respect to W . To this end, we prove a bound for $|\dot{\gamma}(t_i)|$ where $\gamma(t_i)$ is a one-sided minimizer. Another bound is proved in [13] using Gronwall's inequality. This bound involves an exponential function of the Brownian motion, which is difficult for constructing the function $G(r, W)$ in assumption 2.3.

Lemma 5.10. *All one-sided minimizers γ that start at (x, t_0'') satisfy*

$$|\dot{\gamma}(t_i)| \leq c(t_i - \bar{t}_i(W)) \left(1 + |x|^2 + \max_{\bar{t}_i(W) \leq \tau \leq t_i} |W(\tau) - W(t_i)|^2 \right), \quad (5.9)$$

where the constant c does not depend on x , t_i and $\bar{t}_i(W)$.

Proof. Let $s \in (\bar{t}_i(W), t_i)$ such that $\gamma(s) \in B(b)$. From (5.2), we have

$$\begin{aligned} \mathcal{A}_{s,t_i}^W(\gamma) &\geq \frac{1}{2(t_i - s)} \left(\int_s^{t_i} |\dot{\gamma}(\tau)| d\tau \right)^2 - c \max_{s \leq \tau \leq t_i} |W(\tau) - W(t_i)| \\ &\quad - W(t_i) \int_s^{t_i} |\dot{\gamma}(\tau)| d\tau - c|W(s) - W(t_i)|. \end{aligned} \quad (5.10)$$

We consider the linear curve $\tilde{\gamma}$ that connects (x, t_i) and $(\gamma(s), s)$. As $\dot{\tilde{\gamma}}(\tau) \leq (|x| + |b|)/(t_i - s)$,

$$\mathcal{A}_{s,t}^W(\tilde{\gamma}) \leq \frac{(|x| + |b|)^2}{2(t_i - s)} + c(|x| + b) \max_{s \leq \tau \leq t_i} |W(\tau) - W(t_i)| + c|W(s) - W(t_i)|. \quad (5.11)$$

The constants c in (5.10) and (5.11) do not depend on x . From $\mathcal{A}_{s,t_i}^W(\gamma) \leq \mathcal{A}_{s,t_i}^W(\tilde{\gamma})$, solving the quadratic equation, we deduce that

$$\int_s^{t_i} |\dot{\gamma}(\tau)| d\tau \leq (t_i - s)c \left(1 + |x| + \max_{s \leq \tau \leq t_i} |W(\tau) - W(t_i)| \right),$$

where the constant c does not depend on x .

The rest of the proof is similar to that of lemma 4.8. $\square$

Lemma 5.11. *For any x and x'*

$$\begin{aligned} |\phi^W(x, t_i) - \phi^W(x', t_i)| &\leq c(t_i - \bar{t}_i(W))^2 [1 + |x|^2 + |x'|^2 \\ &\quad + \max_{\bar{t}_i(W) \leq \tau \leq t_i} |W(\tau) - W(t_i)|^2] |x - x'|. \end{aligned}$$

The proof of this lemma is similar to that of lemma 4.9.

Proposition 5.12. *The function $\mathcal{G}_B$ is a continuous map from $\mathcal{X}$ to $\mathbb{R}^m$ with respect to the metric (3.7).*

Proof. When $W_k \rightarrow W$ in the metric (3.7), $\bar{t}_i(W_k)$ as determined in proposition 5.3 converges to $\bar{t}_i(W)$. From lemma 5.11, $\phi^{W_k}(\cdot, t_i)$ are uniformly bounded in $C^{0,1}(D)/\mathbb{R}$ for any compact set $D \subset \mathbb{R}^m$. Therefore, from (5.7), $\phi^{W_k}(\cdot, t_i)$ converges to $\phi^W(\cdot, t_i)$ in $C^{0,1}(D)/\mathbb{R}$. Thus, $u^{W_k}(\cdot, t_i)$ are uniformly bounded in D and converges to $u^W(\cdot, t_i)$ in $L^1(D)$. Therefore, $\mathcal{G}_B(W^k) \rightarrow \mathcal{G}_B(W)$. $\square$

Now we prove the continuous dependence of μ^y on y .

Proposition 5.13. *Assumption 2.3 holds.*

Proof. From lemma 5.11, we find that at the points where ϕ^W is differentiable

$$|u(x, t_i)| \leq c(t_i - \bar{t}_i(W)) \left[1 + |x|^2 + (t_i - \bar{t}_i(W) + 1)^2 \sum_{l=\bar{t}_i(W)}^{t'_i-1} \max_{l \leq \tau \leq l+1} |W(\tau) - W(l+1)|^2 \right],$$

where t'_i is defined in (5.5). The rest of the proof is similar to that for proposition 5.7.

The function $G(r, W)$ is defined as

$$G(r, W) = \|\Sigma\|_{\mathbb{R}^m, \mathbb{R}^m}^2 \left[r + c \sum_{i=1}^m (t_i - \bar{t}_i(W)) \left(1 + |x|^2 + (t_i - \bar{t}_i(W) + 1)^2 \sum_{l=\bar{t}_i(W)}^{t'_i-1} \max_{l \leq \tau \leq l+1} |W(\tau) - W(l+1)|^2 \right) \right], \quad (5.12)$$

where c does not depend on x , t_i and $\bar{t}_i(W)$.

The proof that the function $G(r, W)$ in (5.12) is in $L^2(\mathcal{X}, d\mu_0)$ is similar to that in section 7.4 for the function $G(r, W)$ defined in (5.8). $\square$

From propositions 5.12 and 5.13, corollary 2.2 and theorem 2.4, we deduce

Theorem 5.14. *Under assumption 5.1, for the randomly forced Burgers equation (3.1) in the non-compact setting, the posterior measure μ^y is absolutely continuous with respect to the prior measure μ_0 and satisfies*

$$\frac{d\mu^y}{d\mu_0} \propto \exp(-\Phi_B(W; y))$$

where Φ_B is defined in (3.9). For $|y| \leq r$ and $|y'| \leq r$, there is a constant $c(r)$ depending on r such that

$$d_{\text{Hell}}(\mu^y, \mu^{y'}) \leq c(r)|y - y'| \quad \text{and} \quad d_{KL}(\mu^y, \mu^{y'}) \leq c(r)|y - y'|.$$

6. Conclusions

We formulated general conditions in this paper for the existence and well-posedness of a posterior probability measure for Bayesian inverse problems in general topological measure spaces. We showed that the conditions for well-posedness are abstract and are problem dependent. Even when metrics and/or seminorms are available, we may not be able to write these conditions in terms of the metric or a seminorm; identifying them can be very challenging. With the well-posedness conditions, we showed that the posterior measure is locally Lipschitz with respect to the Hellinger and the Kullback–Leibler distances. We then applied the abstract framework to Burgers and Hamilton–Jacobi equations with a white noise forcing on a semi-infinite time interval $(-\infty, T]$. This is the problem of inferring the forcing when the physical process has occurred over a long time and no longer depends on the particular initial condition it started with. Employing the available results in the literature for the Burgers and Hamilton–Jacobi equations, we consider the Bayesian inverse problems for both spatially periodic and nonperiodic domains.

7. Proofs of technical results

7.1. Proof of (4.5)

We use the concept of ε -narrow places defined in [16], in particular, the following result.

Lemma 7.1. *For a sufficiently large T , if $s_2 - s_1 = T$ and*

$$\max_{s_1 \leq \tau \leq s_2} |W(\tau) - W(s_2)| < \frac{1}{T^2},$$

then for any minimizers γ on $[s_1, s_2]$,

$$\left| \mathcal{A}_{s_1, s_2}^{W, b}(\gamma) - \frac{b^2 T}{2} \right| < \varepsilon.$$

For each $T > 0$, for almost all W , ergodicity implies that there are infinitely many time intervals that satisfy the assumption of lemma 7.1. The proof of lemma 7.1 can be found in Iturriaga and Khanin [16] lemma 12. We call such an interval $[s_1, s_2]$ an ε -narrow place. Indeed, the definition of ε -narrow places in [16] is more general, but lemma 7.1 is sufficient for our purpose.

To prove (4.5), we need a further result from [16].

Lemma 7.2 ([16], equation (42)). *Let $[s_1, s_2]$ be an ε -narrow place. Let γ be a ψ minimizer on $[s_1, t]$ where $t > s_2$ for a function ψ . Then, for all $p \in \mathbb{T}^d$,*

$$\psi(p) \geq \psi(\gamma(s_1)) - 2\varepsilon.$$

Proof of (4.5). Let γ^{W_k} be a one-sided minimizer starting at (x_i, t_i) for the action $\mathcal{A}^{W_k, b}$. As $D(W_k, W) \rightarrow 0$ when $k \rightarrow \infty$, from lemma 4.8, $\dot{\gamma}^{W_k}(t_i)$ is uniformly bounded.

Let γ^W be a one-sided minimizer for $\mathcal{A}^{W, b}$. Let $[s_1, s_2]$ be an ε -narrow place of the action $\mathcal{A}^{W, b}$ as defined in lemma 7.1. When $D(W, W_k)$ is sufficiently small, $[s_1, s_2]$ is also an ε -narrow place for $\mathcal{A}^{W_k, b}$.

Let γ_1^W be an arbitrary $\psi_b^W(\cdot, s_1)$ minimizer on $[s_1, t_i]$. From theorem 4.2(ii), γ^W is a $\psi_b^W(\cdot, s_1)$ minimizer on $[s_1, t_i]$, so we have from lemma 7.2,

$$\psi_b^W(\gamma_1^W(s_1), s_1) \geq \psi_b^W(\gamma^W(s_1), s_1) - 2\varepsilon.$$

On the other hand, also from lemma 7.2:

$$\psi_b^W(\gamma^W(s_1), s_1) \geq \psi_b^W(\gamma_1^W(s_1), s_1) - 2\varepsilon.$$

Therefore,

$$|\psi_b^W(\gamma^W(s_1), s_1) - \psi_b^W(\gamma_1^W(s_1), s_1)| < 2\varepsilon.$$

Similarly, let $\gamma_1^{W_k}$ be an arbitrary $\psi_b^{W_k}(\cdot, s_1)$ minimizer on $[s_1, t_i]$. We then have

$$|\psi_b^{W_k}(\gamma^{W_k}(s_1), s_1) - \psi_b^{W_k}(\gamma_1^{W_k}(s_1), s_1)| < 2\varepsilon.$$

Fix γ_1^W and $\gamma_1^{W_k}$. Letting $C_k = \psi_b^W(\gamma_1^W(s_1), s_1) - \psi_b^{W_k}(\gamma_1^{W_k}(s_1), s_1)$, we deduce

$$|\psi_b^W(\gamma^W(s_1), s_1) - \psi_b^{W_k}(\gamma^{W_k}(s_1), s_1) - C_k| < 4\varepsilon. \quad (7.1)$$

We have

$$\psi_b^{W_k}(x_i, t_i) = \psi_b^{W_k}(\gamma^{W_k}(s_1), s_1) + \mathcal{A}_{s_1, s_2}^{W_k, b}(\gamma^{W_k}) + \mathcal{A}_{s_2, t_i}^{W_k, b}(\gamma^{W_k}).$$

Let $\tilde{\gamma}$ be the curve constructed as follows. For $\tau \in [s_2, t_i]$, $\tilde{\gamma}(\tau) = \gamma^{W_k}(\tau)$. On $[s_1, s_2]$, $\tilde{\gamma}$ is a minimizing curve for $\mathcal{A}_{s_1, s_2}^{W, b}$ that connects $(\gamma^{W_k}(s_2), s_2)$ and $(\gamma^W(s_1), s_1)$. We then have

$$\begin{aligned} \psi_b^W(x_i, t_i) &\leq \psi_b^W(\gamma^W(s_1), s_1) + \mathcal{A}_{s_1, t_i}^{W, b}(\tilde{\gamma}) \\ &= \psi_b^W(\gamma^W(s_1), s_1) + \mathcal{A}_{s_1, s_2}^{W, b}(\tilde{\gamma}) + \mathcal{A}_{s_2, t_i}^{W, b}(\gamma^{W_k}). \end{aligned}$$

Therefore,

$$\begin{aligned} \psi_b^W(x_i, t_i) - \psi_b^{W_k}(x_i, t_i) &\leq (\psi_b^W(\gamma^W(s_1), s_1) - \psi_b^{W_k}(\gamma^{W_k}(s_1), s_1)) \\ &\quad + (\mathcal{A}_{s_2, t_i}^{W, b}(\gamma^{W_k}) - \mathcal{A}_{s_2, t_i}^{W_k, b}(\gamma^{W_k})) + (\mathcal{A}_{s_1, s_2}^{W, b}(\tilde{\gamma}) - \mathcal{A}_{s_1, s_2}^{W_k, b}(\gamma^{W_k})). \end{aligned} \quad (7.2)$$

When $D(W_k, W) \rightarrow 0$, $|\dot{\gamma}^{W_k}(t_i)|$ are uniformly bounded so from (4.12) we deduce that $|\dot{\gamma}^{W_k}(\tau)|$ and $|\gamma^{W_k}(\tau)|$ are uniformly bounded for all $s_2 \leq \tau \leq t_i$. Therefore,

$$\lim_{k \rightarrow \infty} \mathcal{A}_{s_2, t_i}^{W, b}(\gamma^{W_k}) - \mathcal{A}_{s_2, t_i}^{W_k, b}(\gamma^{W_k}) = 0. \quad (7.3)$$

From lemma 7.1, as $[s_1, s_2]$ is an ε -narrow place for both $\mathcal{A}^{W, b}$ and $\mathcal{A}^{W_k, b}$,

$$|\mathcal{A}_{s_1, s_2}^{W, b}(\bar{\gamma}) - \mathcal{A}_{s_1, s_2}^{W_k, b}(\gamma^{W_k})| < 2\varepsilon. \quad (7.4)$$

From (7.1), (7.2), (7.3), (7.4) we have

$$\psi_b^W(x_i, t_i) - \psi_b^{W_k}(x_i, t_i) - C_k \leq 7\varepsilon,$$

when k is sufficiently large. Similarly,

$$\psi_b^{W_k}(x_i, t_i) - \psi_b^W(x_i, t_i) + C_k \leq 7\varepsilon,$$

when k is large. From these, we deduce

$$|\psi_b^W(x_i, t_i) - \psi_b^{W_k}(x_i, t_i) - C_k| \leq 7\varepsilon,$$

when k is sufficiently large. Hence, for this particular subsequence of $\{W_k\}$ (not renumbered)

$$\lim_{k \rightarrow \infty} |\phi_b^{W_k}(x_i, t_i) - \phi_b^W(x_i, t_i) - C_k| = 0 \quad (7.5)$$

for appropriate constants C_k which do not depend on x_i and t_i . From any subsequences of $\{W_k\}$, we can choose such a subsequence. We then get the conclusion. $\square$

7.2. Proof of (5.7)

The proof uses ε -narrow places for the nonperiodic setting which are defined as follows.

Definition 7.3. An interval $[s_1, s_2]$ is called an ε -narrow place for the action $\mathcal{A}^W$ defined in (5.2) if there are two compact sets S_1 and S_2 such that for all minimizing curves $\gamma(\tau)$ on an interval $[s'_1, s'_2] \supset [s_1, s_2]$ with $|\gamma(s'_1)| \leq b$ and $|\gamma(s'_2)| \leq b$, $\gamma(s_1) \in S_1$ and $\gamma(s_2) \in S_2$. Furthermore, if γ is a minimizing curve over $[s_1, s_2]$ connecting a point in S_1 to a point in S_2 , then $|\mathcal{A}_{s_1, s_2}^W(\gamma)| < \varepsilon$.

In [13] theorem 7, such an ε -narrow place is constructed as follows. Assume that $s_2 - s_1 = T$ and

$$\max_{s_1 \leq \tau \leq s_2} |W(\tau) - W(s_2)| < \frac{1}{2T}.$$

Without loss of generality, we assume that s_1 and s_2 are integers. Fixing a positive constant c that depends only on the function F and b , the law of large numbers implies that if two integers $s'_1 < s_1$ and $s'_2 > s_2$ satisfy

$$\sum_{l=s_2}^{s'_2-1} P_l + \sum_{l=s'_1}^{s_1-1} P_l \geq -c \left[1 + \max_{s'_1-2 \leq \tau \leq s'_1} |W(\tau) - W(s'_1)|^2 \right] \\ - c \left[1 + \max_{s'_2 \leq \tau \leq s'_2+2} |W(\tau) - W(s'_2)|^2 \right],$$

then $s'_1 \geq s_1^*$ and $s'_2 \leq s_2^*$ where s_1^* and s_2^* depend on the path W and the constant c only (P_l is defined in (5.3)). Let

$$M_1 = \sup_{s_1^*-2 \leq \tau \leq s_1} |W(\tau) - W(s_1)|, \quad M_2 = \sup_{s_2 \leq \tau \leq s_2^*+2} |W(\tau) - W(s_2)|.$$

Hoang and Khanin ([13], pp 833–5) show that there is a continuous function $R(M_1, M_2)$ and a constant c that depends on F such that if

$$\frac{c(R(M_1, M_2)^2 + 1)}{T} < \varepsilon,$$

then $[s_1, s_2]$ is an ε -narrow place of $\mathcal{A}^W$. By using ergodicity, they show that with probability 1, an ε -narrow place constructed this way always exists and s_2 can be chosen arbitrarily small. When $D(W, W')$ is sufficiently small, then it is also an ε -narrow place for $\mathcal{A}^{W'}$.

We then have the following lemma (indeed Hoang and Khanin [13], in the proof of their lemma 3, show a similar result for the kicked forcing case, but the proof for white noise forcing follows the same line).

Lemma 7.4. *If γ is a $\phi(\cdot, s_1)$ minimizer on $[s_1, t]$ such that $\gamma(s_1) \in S_1$ and $\gamma(s_2) \in S_2$, then for all $p \in S_1$*

$$\phi(p, s_1) > \phi(\gamma(s_1), s_1) - 2\varepsilon.$$

Proof of (5.7). The proof is similar to that of (4.5). Assume that $[s_1, s_2]$ is an ε -narrow place for both W and W_k . Any one-sided minimizers starting at (x, t) where x is in a compact set, when s_2 is sufficiently smaller than t , must intersect the ball $B(b)$ at a time larger than s_2 and at a time smaller than s_1 (see the proof of proposition 2 in [13]). Therefore, $\gamma(s_1) \in S_1$ and $\gamma(s_2) \in S_2$. We then proceed as in the proof of (4.5) in section 7.1. $\square$

7.3. Proof of lemma 5.6

In this section, we prove lemma 5.6.

Consider a one-sided minimizer γ that starts at (x, t_i) . There is a time $s \in [\bar{t}_i(W), t_0'']$ ($\bar{t}_i(W)$ is defined in proposition 5.3) such that $\gamma(s) \in B(b)$. We then have

$$\phi^W(x_i, t_i) = \phi^W(\gamma(s), s) + \mathcal{A}_{s, t_i}^W(\gamma).$$

Let s' be the integer such that $1 \leq s' - s < 2$. We then have

$$\mathcal{A}_{s, t_i}^W(\gamma) = \mathcal{A}_{s, s'}^W(\gamma) + \sum_{l=s'}^{t_i''-1} \mathcal{A}_{l, l+1}^W(\gamma) + \mathcal{A}_{t_i'', t_i}^W(\gamma), \quad (7.6)$$

where t_i'' is defined in (5.5). From (5.2), we have

$$\begin{aligned} \mathcal{A}_{s, s'}^W(\gamma) &\geq \frac{1}{2|s' - s|} \left(\int_s^{s'} |\dot{\gamma}(\tau)| d\tau \right)^2 \\ &\quad - \max_x |\nabla F(x)| \max_{s \leq \tau \leq s'} |W(\tau) - W(s')| \int_s^{s'} |\dot{\gamma}(\tau)| d\tau \\ &\quad - \max_x |F(x)| |W(s) - W(s')|, \end{aligned}$$

which is a quadratic form for $\int_s^{s'} |\dot{\gamma}(\tau)| d\tau$. Therefore, there is a constant c that depends on F such that

$$\mathcal{A}_{s, s'}^W(\gamma) \geq -c(1 + \max_{s \leq \tau \leq s'} (1 + |W(\tau) - W(s')|^2)).$$

Performing similarly for other terms in (7.6), we find that there is a constant c which does not depend on x, x_0, t, t_0 such that

$$\mathcal{A}_{s, t_i}^W(\gamma) \geq -c \sum_{l=s'}^{t_i'} \left(1 + \max_{l \leq \tau \leq l+1} |W(\tau) - W(l+1)|^2 \right),$$

where t'_i is defined in (5.5), and s'' is the integer such that $0 \leq s - s'' < 1$. Thus,

$$\phi^W(x_i, t_i) \geq \phi^W(\gamma(s), s) - c \sum_{l=\bar{t}_i(W)}^{t'_i} \left(1 + \max_{l \leq \tau \leq l+1} |W(\tau) - W(l+1)|^2\right). \quad (7.7)$$

Now we consider the straight line γ_1 that connects (x_0, t_0) with $(\gamma(s), s)$. We have

$$\phi^W(x_0, t_0) \leq \phi^W(\gamma_1(s), s) + \mathcal{A}_{s,t_0}^W(\gamma_1).$$

Let t''_0 be an integer such that $1 \leq t_0 - t''_0 < 2$. We write

$$\mathcal{A}_{s,t_0}^W(\gamma_1) = \mathcal{A}_{s,s'}^W(\gamma_1) + \sum_{l=s'}^{t''_0-1} \mathcal{A}_{l,l+1}^W(\gamma_1) + \mathcal{A}_{t''_0,t_0}^W(\gamma_1).$$

The velocity $|\dot{\gamma}_1(\tau)| \leq |b - x_0|$ so that there is a constant $c(x_0)$ such that

$$\mathcal{A}_{s,t_0}^W(\gamma_1) \leq c(x_0) \sum_{l=\bar{t}_i(W)}^{t'_0} \left(1 + \max_{l \leq \tau \leq l+1} |W(\tau) - W(l+1)|\right).$$

Therefore,

$$\phi^W(x_0, t_0) \leq \phi^W(\gamma(s), s) + c(x_0) \sum_{l=\bar{t}_i(W)}^{t'_0} \left(1 + \max_{l \leq \tau \leq l+1} |W(\tau) - W(l+1)|\right). \quad (7.8)$$

From equations (7.7) and (7.8), we have

$$\phi^W(x_i, t_i) - \phi^W(x_0, t_0) \geq -c \sum_{l=\bar{t}_i(W)}^{t'_i} \left(1 + \max_{l \leq \tau \leq l+1} |W(\tau) - W(l+1)|^2\right).$$

7.4. A proof that the function $G(r, W)$ in (5.8) is in $L^2(\mathcal{X}, \mu_0)$

We now show that the function $G(r, W)$ defined in (5.8) is in $L^2(\mathcal{X}, \mu_0)$. We indeed show that

$$\sum_{l=\bar{t}_i(W)}^{t'_i} \left(1 + \max_{l \leq \tau \leq l+1} |W(\tau) - W(l+1)|^2\right)$$

is in $L^2(\mathcal{X}, \mu_0)$.

Let ε be a small positive constant. Fixing an integer k , we consider the following events:

(A):

$$\left| \sum_{l=t''_i-m}^{t'_i} \max_{l \leq \tau \leq l+1} |W(\tau) - W(l+1)|^2 - (t'_i - t''_i + m + 1)E_3 \right| < (t'_i - t''_i + m + 1)\varepsilon$$

for all $m \geq k$,

(B):

$$\left| \sum_{l=t''_i-m}^{t''_0-1} \max_{l \leq \tau \leq l+1} |W(\tau) - W(l+1)|^2 - mE_3 \right| < m\varepsilon$$

for all $m \geq k$,

(C):

$$\left| \sum_{l=t_i''-m}^{t_i''-1} P_l - m\mathbb{E}\{P_l\} \right| < m\varepsilon$$

for all $m \geq k$.Assume that all of these events occur. Then, from (A) and (B) for $m = k$, we get

$$\sum_{l=t_i''}^{t_i'} \max_{l \leq \tau \leq l+1} |W(\tau) - W(j+1)|^2 \leq (t_i' - t_i'' + 1)E_3 + (t_i' - t_i'' + 2k + 1)\varepsilon;$$

using (B) for $m = k$ and $m = k + 1$, we have

$$\max_{t_i''-(k+1) \leq \tau \leq t_i''-k} |W(\tau) - W(t_i'' - k)|^2 \leq (2k + 1)\varepsilon + E_3;$$

using (B) for $m = k + 1$ and $m = k + 2$ we have

$$\max_{t_i''-(k+2) \leq \tau \leq t_i''-(k+1)} |W(\tau) - W(t_i'' - (k + 1))|^2 \leq (2k + 3)\varepsilon + E_3;$$

and using (C) for $m = k$ we have

$$\sum_{l=t_i''-k}^{t_i''-1} P_l \leq k(\mathbb{E}\{P_l\} + \varepsilon) \leq k \left(-\frac{LE_1}{8} + \varepsilon \right)$$

(note from (5.4) that $\mathbb{E}\{P_l\} \leq -LE_1/8$).Now suppose that all the events (A), (B) and (C) hold for $k = t_i'' - \bar{t}_i(W)$; then, from (5.6) for $\bar{t} = \bar{t}_i(W)$, we get

$$(t_i'' - \bar{t}_i(W)) \left(-\frac{LE_1}{8} + \varepsilon \right) \geq -c(x_i, t_i, F) \{1 + (t_i' - t_i'' + 1)E_3 + [t_i' - t_i'' + 2(t_i'' - \bar{t}_i(W)) + 1]\varepsilon\} - c(F) \{1 + [4(t_i'' - \bar{t}_i(W)) + 4]\varepsilon + 2E_3\}.$$

Thus,

$$(t_i'' - \bar{t}_i(W)) \left[-\frac{LE_1}{8} + \varepsilon \left(1 + 2c(x_i, t_i, F) + 4c(F) \right) \right] \geq -c(x_i, t_i, F) \left[1 + (t_i' - t_i'' + 1)E_3 + (t_i' - t_i'' + 1)\varepsilon \right] - c(F)(1 + 4\varepsilon + 2E_3).$$

When ε is sufficiently small, this only holds when $t_i'' - \bar{t}_i(W)$ is smaller than an integer N which does not depend on the white noise, i.e. when $\bar{t}_i(W) \geq t_i'' - N$. If (5.6) holds but $\bar{t}_i(W) < t_i'' - N$, then at least one of the events (A), (B) and (C) must not hold. Now we denote by $\mathcal{X}_k$ the subset of $\mathcal{X}$ that contains all the Brownian paths W such that at least one of the events (A), (B) and (C) does not hold. For each number r , there is a constant $C(r, \varepsilon)$ such that

$$\mathbb{P}\{\mathcal{X}_k\} \leq \frac{C(r, \varepsilon)}{k^r} \quad (7.9)$$

(see Baum and Katz [3], theorem 1). Therefore,

$$\begin{aligned} & \mathbb{E} \left\{ \left[\sum_{l=\bar{t}_i(W)}^{t_i'} \left(1 + \max_{l \leq \tau \leq l+1} |W(\tau) - W(l+1)|^2 \right) \right]^2 \right\} \\ & \leq \mathbb{E} \left\{ \left[\sum_{l=t_i''-N}^{t_i'} \left(1 + \max_{l \leq \tau \leq l+1} |W(\tau) - W(l+1)|^2 \right) \right]^2 \right\} \end{aligned}$$

$$+ \sum_{k=1}^{\infty} \mathbb{E} \left\{ \left[\sum_{l=t_i''-k}^{t_i'} \left(1 + \max_{l \leq \tau \leq l+1} |W(\tau) - W(l+1)|^2 \right) \right]^2 \mathbf{1}_{\mathcal{X}_k} \right\}.$$

From (7.9), we have

$$\begin{aligned} & \mathbb{E} \left\{ \left[\sum_{l=t_i''-k}^{t_i'} \left(1 + \max_{l \leq \tau \leq l+1} |W(\tau) - W(l+1)|^2 \right) \right]^2 \mathbf{1}_{\mathcal{X}_k} \right\} \\ & \leq \frac{C(r, \varepsilon)}{k^{r/2}} \left[\mathbb{E} \left\{ \left[\sum_{l=t_i''-k}^{t_i'} \left(1 + \max_{l \leq \tau \leq l+1} |W(\tau) - W(l+1)|^2 \right) \right]^4 \right\} \right]^{1/2} \\ & \leq \frac{C(r, \varepsilon)}{k^{r/2}} (t_i' - t_i'' + k + 1)^{3/2} \left[\mathbb{E} \left\{ \sum_{l=t_i''-k}^{t_i'} \left(1 + \max_{l \leq \tau \leq l+1} |W(\tau) - W(l+1)|^2 \right)^4 \right\} \right]^{1/2} \\ & \leq C(r, \varepsilon) k^{-r/2} (t_i' - t_i'' + k + 1)^2. \end{aligned}$$

When r is sufficiently large, $\sum_{k=1}^{\infty} k^{-r/2} (t_i' - t_i'' + k + 1)^2$ is finite. The assertion follows.

Acknowledgments

The author thanks Andrew Stuart for suggesting the metric space framework and for helpful discussions. He also thanks Kostya Khanin for introducing him to Burgers turbulence. Financial support from the startup research grant from Nanyang Technological University is gratefully acknowledged.

References

- [1] Apte A, Auroux D and Ramaswamy M 2010 Variational data assimilation for discrete Burgers equation *Electron. J. Diff. Eqns* **19** 15–30
- [2] Bardos C and Pironneau O 2005 Data assimilation for conservation laws *Methods Appl. Anal.* **12** 103–34
- [3] Baum L E and Katz M 1963 Convergence rates in the law of large numbers *Bull. Am. Math. Soc.* **69** 771–2
- [4] Bec J and Khanin K 2007 Burgers turbulence *Phys. Rep.* **447** 1–66
- [5] Bogachev V I 1998 *Gaussian Measures (Mathematical Surveys and Monographs vol 62)* (Providence, RI: American Mathematical Society)
- [6] Cohen A, DeVore R and Schwab C 2010 Convergence rates of best N -term Galerkin approximations for a class of elliptic sPDEs *Found. Comput. Math.* **10** 615–46
- [7] Cotter S L, Dashti M, Robinson J C and Stuart A M 2009 Bayesian inverse problems for functions and applications to fluid mechanics *Inverse Problems* **25** 115008
- [8] E W, Khanin K, Mazel A and Sinai Y 2000 Invariant measures for Burgers equation with stochastic forcing *Ann. Math. (2)* **151** 877–960
- [9] Engl H W, Hanke M and Neubauer A 1996 *Regularization of Inverse Problems (Mathematics and Its Applications vol 375)* (Dordrecht: Kluwer)
- [10] Freitag M, Nichols N K and Budd C J 2010 L_1 -regularisation for ill-posed problems in variational data assimilation *Proc. Appl. Math. Mech.* **10** 665–8
- [11] Frisch U and Bec J 2001 ‘Burgulence’ *Turbulence: Nouveaux Aspects/New Trends in Turbulence (Les Houches, 2000)* (Les Ulis: EDP Sciences) pp 341–83
- [12] Gomes D, Iturriaga R, Khanin K and Padilla P 2005 Viscosity limit of stationary distributions for the random forced Burgers equation *Moscow Math. J.* **5** 613–31, 743

- [13] Hoang V H and Khanin K 2003 Random Burgers equation and Lagrangian systems in non-compact domains *Nonlinearity* **16** 819–42
- [14] Hoang V H and Schwab C 2012 Analytic regularity and polynomial approximation of stochastic, parametric elliptic multiscale PDEs *Analysis Appl.* at press
- [15] Isakov V 2006 *Inverse Problems for Partial Differential Equations (Applied Mathematical Sciences vol 127)* 2nd edn (New York: Springer)
- [16] Iturriaga R and Khanin K 2003 Burgers turbulence and random Lagrangian systems *Commun. Math. Phys.* **232** 377–428
- [17] Kaipio J and Somersalo E 2005 *Statistical and Computational Inverse Problems (Applied Mathematical Sciences vol 160)* (New York: Springer)
- [18] Kallenberg O 2002 *Foundations of Modern Probability (Probability and Its Applications)* 2nd edn (New York: Springer)
- [19] Lam C H and Sander L M 1993 Inverse methods for interface problems *Phys. Rev. Lett.* **71** 561–4
- [20] Lifshits M A 1995 *Gaussian Random Functions (Mathematics and Its Applications vol 322)* (Dordrecht: Kluwer)
- [21] Lundvall J, Kozlov V and Weinerfelt P 2006 Iterative methods for data assimilation for Burgers' equation *J. Inverse Ill-Posed Problem* **14** 505–35
- [22] Mattingly J C 1999 Ergodicity of 2 D Navier–Stokes equations with random forcing and large viscosity *Commun. Math. Phys.* **206** 273–88
- [23] Schwab C and Stuart A M 2011 Sparse deterministic approximation of Bayesian inverse problems *Technical Report SAM ETH Zurich*
- [24] Stroock D W and Srinivasa Varadhan S R 2006 *Multidimensional Diffusion Processes (Classics in Mathematics)* (Berlin: Springer) (reprint of 1997 edn)
- [25] Stuart A M 2010 Inverse problems: a Bayesian perspective *Acta Numer.* **19** 451–559