

Noname manuscript No. (will be inserted by the editor)

Noisy Gradient Flow from a Random Walk in Hilbert Space

Natesh S. Pillai · Andrew M. Stuart · Alexandre H. Thiéry

Received: date / Accepted: date

Abstract Consider a probability measure on a Hilbert space defined via its density with respect to a Gaussian. The purpose of this paper is to demonstrate that an appropriately defined Markov chain, which is reversible with respect to the measure in question, exhibits a diffusion limit to a noisy gradient flow, also reversible with respect to the same measure. The Markov chain is defined by applying a Metropolis-Hastings accept-reject mechanism [27] to an Ornstein-Uhlenbeck (OU) proposal which is itself reversible with respect to the underlying Gaussian measure. The resulting noisy gradient flow is a stochastic partial differential equation driven by a Wiener process with spatial correlation given by the underlying Gaussian structure.

There are two primary motivations for this work. The first concerns insight into Monte Carlo Markov Chain (MCMC) methods for sampling of measures on a Hilbert space defined via a density with respect to a Gaussian measure. These measures must be approximated on finite dimensional spaces of dimension N in order to be sampled. A conclusion of the work herein is that MCMC methods based on prior-reversible OU proposals will explore the target measure in $\mathcal{O}(1)$ steps with respect to dimension N . This is to be contrasted with standard MCMC methods based on the random walk or Langevin proposals which require $\mathcal{O}(N)$ and $\mathcal{O}(N^{1/3})$ steps respectively [23, 24]. The second motivation relates to optimization. There are many applications where it is of interest to find global or local minima of a functional defined on an infinite dimensional Hilbert space. Gradient flow or steepest descent is a natural approach to this problem, but in its basic form requires computation of a gradient which, in some applications, may be an expensive or complex task. This paper shows that a stochastic gradient descent described by a stochastic partial differential equation can emerge from certain carefully specified Markov chains. This idea is well-known in the finite state [21, 4] or finite dimensional context [13, 14, 5, 20]. The novelty of the work in this paper is that the emergence of the noisy gradient flow is developed on an infinite dimensional Hilbert space. In the context of global optimization, when the noise level is also adjusted as part of the algorithm, methods of the type studied here go by the name of simulated-annealing; see the review [2] for further references. Although we

Natesh S. Pillai
Department of Statistics,
Harvard University
1 Oxford Street, Cambridge
02138, MA, USA
E-mail: pillai@stat.harvard.edu

Andrew M. Stuart
Mathematics Institute
Warwick University
CV4 7AL, UK
E-mail: a.m.stuart@warwick.ac.uk

Alexandre H. Thiéry
Department of Statistics
Warwick University
CV4 7AL, UK
E-mail: a.h.thiery@warwick.ac.uk

arXiv:1108.1494v4 [math.ST] 18 Apr 2014

do not consider adjusting the noise-level as part of the algorithm, the noise strength is a tuneable parameter in our construction and the methods developed here could potentially be used to study simulated annealing in a Hilbert space setting.

The transferable idea behind this work is that conceiving of algorithms directly in the infinite dimensional setting leads to methods which are robust to finite dimensional approximation. We emphasize that discretizing, and then applying standard finite dimensional techniques in \mathbb{R}^N , to either sample or optimize, can lead to algorithms which degenerate as the dimension N increases.

Keywords Optimisation · Simulated annealing · Markov Chain Monte Carlo · Diffusion approximation

Mathematics Subject Classification (2000) 60-08 · 60H15 · 60J25

1 Introduction

There are many applications where it is of interest to find global or local minima of a functional

$$J(x) = \frac{1}{2} \|C^{-1/2}x\|^2 + \Psi(x) \quad (1)$$

where C is a self-adjoint, positive and trace-class linear operator on a Hilbert space $(\mathcal{H}, \langle \cdot, \cdot \rangle, \|\cdot\|)$. Gradient flow or steepest descent is a natural approach to this problem, but in its basic form requires computation of the gradient of Ψ which, in some applications, may be an expensive or complex task. The purpose of this paper is to show how a stochastic gradient descent described by a stochastic partial differential equation can emerge from certain carefully specified random walks, when combined with a Metropolis-Hastings accept-reject mechanism [27]. In the finite state [21, 4] or finite dimensional context [13, 14, 5, 20] this is a well-known idea, which goes by the name of simulated-annealing; see the review [2] for further references. The novelty of the work in this paper is that the theory is developed on an infinite dimensional Hilbert space, leading to an algorithm which is robust to finite dimensional approximation: we adopt the ‘‘optimize then discretize’’ viewpoint (see [19], Chapter 3). We emphasize that discretizing, and then applying standard finite dimensional techniques in \mathbb{R}^N to optimize, can lead to algorithms which degenerate as N increases; the diffusion limit proved in [23] provides a concrete example of this phenomenon for the standard random walk algorithm.

The algorithms we construct have two basic building blocks: (i) drawing samples from the centred Gaussian measure $N(0, C)$ and (ii) evaluating Ψ . By judiciously combining these ingredients we generate (approximately) a noisy gradient flow for J with tunable temperature parameter controlling the size of the noise. In finite dimensions the basic idea is built from Metropolis-Hastings methods which have an invariant measure with Lebesgue density proportional to $\exp(-\tau^{-1}J(x))$. The essential challenge in transferring these finite-dimensional algorithms to an infinite-dimensional setting is that there is no Lebesgue measure. This issue can be circumvented by working with measures defined via their density with respect to a Gaussian measure, and for us the natural Gaussian measure on \mathcal{H} is

$$\pi_0^\tau = N(0, \tau C). \quad (2)$$

The quadratic form $\|C^{-\frac{1}{2}}x\|^2$ is the square of the Cameron-Martin norm corresponding to the Gaussian measure π_0^τ . Given π_0^τ we may then define the (in general non-Gaussian) measure π^τ via its Radon-Nikodym derivative with respect to π_0^τ :

$$\frac{d\pi^\tau}{d\pi_0^\tau}(x) \propto \exp\left(-\frac{\Psi(x)}{\tau}\right). \quad (3)$$

We assume that $\exp(-\tau^{-1}\Psi(\cdot))$ is in $L^1_{\pi_0^\tau}$. Note that if \mathcal{H} is finite dimensional then π^τ has Lebesgue density proportional to $\exp(-\tau^{-1}J(x))$.

Our basic strategy will be to construct a Markov chain which is π^τ -invariant and to show that a piecewise linear interpolant of the Markov chain converges weakly (in the sense of probability measures) to the desired

noisy gradient flow in an appropriate parameter limit. To motivate the Markov chain we first observe that the linear SDE in \mathcal{H} given by

$$\begin{aligned} dz &= -z dt + \sqrt{2\tau} dW \\ z_0 &= x, \end{aligned} \tag{4}$$

where W is a Brownian motion in \mathcal{H} with covariance operator equal to C , is reversible and ergodic with respect to π_0^τ given by (2) [10]. If $t > 0$ then the exact solution of this equation has the form, for $\delta = \frac{1}{2}(1 - e^{-2t})$,

$$\begin{aligned} z(t) &= e^{-t}x + \sqrt{\left(\tau(1 - e^{-2t})\right)}\xi \\ &= (1 - 2\delta)^{\frac{1}{2}}x + \sqrt{2\delta\tau}\xi, \end{aligned} \tag{5}$$

where ξ is a Gaussian random variable drawn from $N(0, C)$. Given a current state x of our Markov chain we will propose to move to $z(t)$ given by this formula, for some choice of $t > 0$. We will then accept or reject this proposed move with probability found from pointwise evaluation of Ψ , resulting in a Markov chain $\{x^{k,\delta}\}_{k \in \mathbb{Z}^+}$. The resulting Markov chain corresponds to the preconditioned Crank-Nicolson, or pCN, method, also referred to as the PIA method with $(\alpha, \theta) = (0, \frac{1}{2})$ in the paper [3] where it was introduced; this is one of a family of Metropolis-Hastings methods defined on the Hilbert space \mathcal{H} and the review [8] provides further details.

From the output of the pCN Metropolis-Hastings method we construct a continuous interpolant of the Markov chain defined by

$$z^\delta(t) = \frac{1}{\delta}(t - t_k)x^{k+1,\delta} + \frac{1}{\delta}(t_{k+1} - t)x^{k,\delta} \quad \text{for} \quad t_k \leq t < t_{k+1} \tag{6}$$

with $t_k \stackrel{\text{def}}{=} k\delta$. The main result of the paper is that as $\delta \rightarrow 0$ the Hilbert-space valued function of time z^δ converges weakly to z solving the Hilbert space valued SDE, or SPDE, following the dynamics

$$dz = -\left(z + C\nabla\Psi(z)\right) dt + \sqrt{2\tau}dW \tag{7}$$

on pathspace. This equation is reversible and ergodic with respect to the measure π^τ [10, 16]. It is also known that small ball probabilities are asymptotically maximized (in the small radius limit), under π^τ , on balls centred at minimizers of J [11]. The result thus shows that the algorithm will generate sequences which concentrate near minimizers of J .

Because the SDE (7) does not possess the smoothing property, almost sure fine scale properties under its invariant measure π^τ are not necessarily reflected at any finite time. For example, if C is the covariance operator of Brownian motion or Brownian bridge then the quadratic variation of draws from the invariant measure, an almost sure quantity, is not reproduced at any finite time in (7) unless $z(0)$ has this quadratic variation; the almost sure property is approached asymptotically as $t \rightarrow \infty$. This behaviour is reflected in the underlying Metropolis-Hastings Markov chain pCN with weak limit (7), where the almost sure property is only reached asymptotically as $n \rightarrow \infty$. In a second result of this paper we will show that almost sure quantities such as the quadratic variation under pCN satisfy a limiting linear ODE with globally attractive steady state given by the value of the quantity under π^τ . This gives quantitative information about the rate at which the pCN algorithm approaches statistical equilibrium.

We have motivated the limit theorem in this paper through the goal of creating noisy gradient flow in infinite dimensions with tuneable noise level, using only draws from a Gaussian random variable and evaluation of the non-quadratic part of the objective function. A second motivation for the work comes from understanding the computational complexity of MCMC methods, and for this it suffices to consider τ fixed at 1. The paper [23] shows that discretization of the standard Random Walk Metropolis algorithm, S-RWM, will also have diffusion limit given by (7) as the dimension of the discretized space tends to infinity, whilst the time increment δ in (6), decreases at a rate inversely proportional to N . The condition on δ is a form of CFL condition, in the language of computational PDEs, and implies that $\mathcal{O}(N)$ steps will be required to sample the desired probability distribution. In contrast the pCN method analyzed here has no CFL restriction: δ may tend to zero independently of dimension; indeed in this paper we work directly in the setting of infinite

dimension. The reader interested in this computational statistics perspective on diffusion limits may also wish to consult the paper [24] which demonstrates that the Metropolis adjusted Langevin algorithm, MALA, requires a CFL condition which implies that $\mathcal{O}(N^{\frac{1}{3}})$ steps are required to sample the desired probability distribution. Furthermore, the formulation of the limit theorem that we prove in this paper is closely related to the methodologies introduced in [23] and [24]; it should be mentioned nevertheless that the analysis carried out in this article allows to prove a diffusion limit for a sequence of Markov chains evolving in a possibly non-stationary regime. This was not the case in [23] and [24].

We prove in Theorem 4 that for a fixed temperature parameter $\tau > 0$, as the time increment δ goes to 0, the pCN algorithm behaves as a stochastic gradient descent. By adapting the temperature $\tau \in (0, \infty)$ according to an appropriate cooling schedule it is possible to locate global minima of J ; standard heuristics show that the distribution π^τ concentrates on a $\tau^{1/2}$ -neighbourhood around the global minima of the functional J . We stress though that all the proofs presented in this article assume a constant temperature. The asymptotic analysis of the effect of the cooling schedule is left for future work; the study of such Hilbert space valued simulated annealing algorithms presents several challenges, one of them being that the probability distributions π^τ are mutually singular for different temperatures $\tau > 0$.

In section 2 we describe some notation used throughout the paper, discuss the required properties of Gaussian measures and Hilbert-space valued Brownian motions, and state our assumptions. Section 3 contains a precise definition of the Markov chain $\{x^{k,\delta}\}_{k \geq 0}$, together with statement and proof of the weak convergence theorem that is the main result of the paper. Section 4 contains proof of the lemmas which underly the weak convergence theorem. In section 5 we state and prove the limit theorem for almost sure quantities such as quadratic variation; such results are often termed ‘‘fluid limits’’ in the applied probability literature. An example is presented in section 6. We conclude in section 7.

2 Preliminaries

In this section we define some notational conventions, Gaussian measure and Brownian motion in Hilbert space, and state our assumptions concerning the operator C and the functional Ψ .

2.1 Notation

Let $(\mathcal{H}, \langle \cdot, \cdot \rangle, \|\cdot\|)$ denote a separable Hilbert space of real valued functions with the canonical norm derived from the inner-product. Let C be a positive symmetric trace class operator on \mathcal{H} and $\{\varphi_j, \lambda_j^2\}_{j \geq 1}$ be the eigenfunctions and eigenvalues of C respectively, so that $C\varphi_j = \lambda_j^2 \varphi_j$ for $j \in \mathbb{N}$. We assume a normalization under which $\{\varphi_j\}_{j \geq 1}$ forms a complete orthonormal basis in \mathcal{H} . For every $x \in \mathcal{H}$ we have the representation $x = \sum_j x_j \varphi_j$ where $x_j = \langle x, \varphi_j \rangle$. Using this notation, we define Sobolev-like spaces $\mathcal{H}^r, r \in \mathbb{R}$, with the inner-products and norms defined by

$$\langle x, y \rangle_r \stackrel{\text{def}}{=} \sum_{j=1}^{\infty} j^{2r} x_j y_j \quad \text{and} \quad \|x\|_r^2 \stackrel{\text{def}}{=} \sum_{j=1}^{\infty} j^{2r} x_j^2. \quad (8)$$

Notice that $\mathcal{H}^0 = \mathcal{H}$. Furthermore $\mathcal{H}^r \subset \mathcal{H} \subset \mathcal{H}^{-r}$ and $\{j^{-r} \varphi_j\}_{j \geq 1}$ is an orthonormal basis of \mathcal{H}^r for any $r > 0$. For a positive, self-adjoint operator $D : \mathcal{H}^r \mapsto \mathcal{H}^r$, its trace in \mathcal{H}^r is defined as

$$\text{Trace}_{\mathcal{H}^r}(D) \stackrel{\text{def}}{=} \sum_{j=1}^{\infty} \langle (j^{-r} \varphi_j), D(j^{-r} \varphi_j) \rangle_r.$$

Since $\text{Trace}_{\mathcal{H}^r}(D)$ does not depend on the orthonormal basis $\{\varphi_j\}_{j \geq 1}$, the operator D is said to be trace class in \mathcal{H}^r if $\text{Trace}_{\mathcal{H}^r}(D) < \infty$ for some, and hence any, orthonormal basis of \mathcal{H}^r . Let $\otimes_{\mathcal{H}^r}$ denote the outer product operator in \mathcal{H}^r defined by

$$(x \otimes_{\mathcal{H}^r} y)z \stackrel{\text{def}}{=} \langle y, z \rangle_r x \quad (9)$$

for vectors $x, y, z \in \mathcal{H}^r$. For an operator $L : \mathcal{H}^r \mapsto \mathcal{H}^l$, we denote its operator norm by $\|\cdot\|_{\mathcal{L}(\mathcal{H}^r, \mathcal{H}^l)}$ defined by $\|L\|_{\mathcal{L}(\mathcal{H}^r, \mathcal{H}^l)} \stackrel{\text{def}}{=} \sup \{\|Lx\|_l, : \|x\|_r = 1\}$. For self-adjoint L and $r = l = 0$ this is, of course, the spectral radius of L . Throughout we use the following notation.

- Two sequences $\{\alpha_n\}_{n \geq 0}$ and $\{\beta_n\}_{n \geq 0}$ satisfy $\alpha_n \lesssim \beta_n$ if there exists a constant $K > 0$ satisfying $\alpha_n \leq K\beta_n$ for all $n \geq 0$. The notations $\alpha_n \asymp \beta_n$ means that $\alpha_n \lesssim \beta_n$ and $\beta_n \lesssim \alpha_n$.
- Two sequences of real functions $\{f_n\}_{n \geq 0}$ and $\{g_n\}_{n \geq 0}$ defined on the same set Ω satisfy $f_n \lesssim g_n$ if there exists a constant $K > 0$ satisfying $f_n(x) \leq Kg_n(x)$ for all $n \geq 0$ and all $x \in \Omega$. The notations $f_n \asymp g_n$ means that $f_n \lesssim g_n$ and $g_n \lesssim f_n$.
- The notation $\mathbb{E}_x[f(x, \xi)]$ denotes expectation with variable x fixed, while the randomness present in ξ is averaged out.
- We use the notation $a \wedge b$ instead of $\min(a, b)$.

2.2 Gaussian Measure on Hilbert Space

The following facts concerning Gaussian measures on Hilbert space, and Brownian motion in Hilbert space, may be found in [9]. Since C is self-adjoint, positive and trace-class we may associate with it a centred Gaussian measure π_0 on \mathcal{H} with covariance operator C , *i.e.*, $\pi_0 \stackrel{\text{def}}{=} \text{N}(0, C)$. If $x \stackrel{\mathcal{D}}{\sim} \pi_0$ then we may write its Karhunen-Loève expansion,

$$x = \sum_{j=1}^{\infty} \lambda_j \rho_j \varphi_j, \quad (10)$$

with $\{\rho_j\}_{j \geq 1}$ an i.i.d sequence of standard centered Gaussian random variables; since C is trace-class, the above sum converges in L^2 . Notice that for any value of $r \in \mathbb{R}$ we have $\mathbb{E}\|X\|_r^2 = \sum_{j \geq 1} j^{2r} \langle X, \varphi_j \rangle^2 = \sum_{j \geq 1} j^{2r} \lambda_j^2$ for $X \stackrel{\mathcal{D}}{\sim} \pi_0$. For values of $r \in \mathbb{R}$ such that $\mathbb{E}\|X\|_r^2 < \infty$ we indeed have $\pi_0(\mathcal{H}^r) = 1$ and the random variable X can also be described as a Gaussian random variable in \mathcal{H}^r . One can readily check that in this case the covariance operator $C_r : \mathcal{H}^r \rightarrow \mathcal{H}^r$ of X when viewed as a \mathcal{H}^r -valued random variable is given by

$$C_r = B_r^{1/2} C B_r^{1/2}. \quad (11)$$

where $B_r : \mathcal{H} \mapsto \mathcal{H}$ denote the operator which is diagonal in the basis $\{\varphi_j\}_{j \geq 1}$ with diagonal entries j^{2r} . In other words, $B_r \varphi_j = j^{2r} \varphi_j$ so that $B_r^{1/2} \varphi_j = j^r \varphi_j$ and $\mathbb{E}[\langle X, u \rangle_r \langle X, v \rangle_r] = \langle u, C_r v \rangle_r$ for $u, v \in \mathcal{H}^r$ and $X \stackrel{\mathcal{D}}{\sim} \pi_0$. The condition $\mathbb{E}\|X\|_r^2 < \infty$ can equivalently be stated as

$$\text{Trace}_{\mathcal{H}^r}(C_r) < \infty.$$

This shows that even though the Gaussian measure π_0 is defined on \mathcal{H} , depending on the decay of the eigenvalues of C , there exists an entire range of values of r such that $\mathbb{E}\|X\|_r^2 = \text{Trace}_{\mathcal{H}^r}(C_r) < \infty$ and in that case the measure π_0 has full support on \mathcal{H}^r .

Frequently in applications the functional Ψ arising in (1) may not be defined on all of \mathcal{H} , but only on a subspace $\mathcal{H}^s \subset \mathcal{H}$, for some exponent $s > 0$. From now onwards we fix a distinguished exponent $s > 0$ and assume that $\Psi : \mathcal{H}^s \rightarrow \mathbb{R}$ and that $\text{Trace}_{\mathcal{H}^s}(C_s) < \infty$ so that $\pi(\mathcal{H}^s) = \pi_0^r(\mathcal{H}^s) = \pi^\tau(\mathcal{H}^s) = 1$; the change of measure formula (3) is well defined. For ease of notations we introduce

$$\hat{\varphi}_j = B_s^{-1/2} \varphi_j = j^{-s} \varphi_j$$

so that the family $\{\hat{\varphi}_j\}_{j \geq 1}$ forms an orthonormal basis for $(\mathcal{H}^s, \langle \cdot, \cdot \rangle_s)$. We may view the Gaussian measure $\pi_0 = \text{N}(0, C)$ on $(\mathcal{H}, \langle \cdot, \cdot \rangle)$ as a Gaussian measure $\text{N}(0, C_s)$ on $(\mathcal{H}^s, \langle \cdot, \cdot \rangle_s)$.

A Brownian motion $\{W(t)\}_{t \geq 0}$ in \mathcal{H}^s with covariance operator $C_s : \mathcal{H}^s \rightarrow \mathcal{H}^s$ is a continuous Gaussian process with stationary increments satisfying $\mathbb{E}[\langle W(t), x \rangle_s \langle W(t), y \rangle_s] = t \langle x, C_s y \rangle_s$. For example, taking $\{\beta_j(t)\}_{j \geq 1}$ independent standard real Brownian motions, the process

$$W(t) = \sum_j (j^s \lambda_j) \beta_j(t) \hat{\varphi}_j \quad (12)$$

defines a Brownian motion in \mathcal{H}^s with covariance operator C_s ; equivalently, this same process $\{W(t)\}_{t \geq 0}$ can be described as a Brownian motion in \mathcal{H} with covariance operator equal to C since Equation (12) may also be expressed as $W(t) = \sum_{j=1}^{\infty} \lambda_j \beta_j(t) \varphi_j$.

2.3 Assumptions

In this section we describe the assumptions on the covariance operator C of the Gaussian measure $\pi_0 = \mathbb{N}(0, C)$ and the functional Ψ , and the connections between them. Roughly speaking we will assume that the second-derivative of Ψ is globally bounded as an operator acting between two spaces which arise naturally from understanding the domain of the function Ψ ; furthermore the domain of Ψ must be a set of full measure with respect to the underlying Gaussian. If the eigenvalues of C decay like $j^{-2\kappa}$ and $\kappa > \frac{1}{2}$ then $\pi_0^\tau(\mathcal{H}^s) = 1$ for all $s < \kappa - \frac{1}{2}$ and so we will assume eigenvalue decay of this form and assume the domain of Ψ is defined appropriately. We now formalize these ideas.

For each $x \in \mathcal{H}^s$ the derivative $\nabla\Psi(x)$ is an element of the dual $(\mathcal{H}^s)^*$ of \mathcal{H}^s , comprising the linear functionals on \mathcal{H}^s . However, we may identify $(\mathcal{H}^s)^* = \mathcal{H}^{-s}$ and view $\nabla\Psi(x)$ as an element of \mathcal{H}^{-s} for each $x \in \mathcal{H}^s$. With this identification, the following identity holds

$$\|\nabla\Psi(x)\|_{\mathcal{L}(\mathcal{H}^s, \mathbb{R})} = \|\nabla\Psi(x)\|_{-s}$$

and the second derivative $\partial^2\Psi(x)$ can be identified with an element of $\mathcal{L}(\mathcal{H}^s, \mathcal{H}^{-s})$. To avoid technicalities we assume that $\Psi(x)$ is quadratically bounded, with first derivative linearly bounded and second derivative globally bounded. Weaker, localized assumptions could be dealt with by use of stopping time arguments.

Assumptions 1 *The functional Ψ and the covariance operator C satisfy the following assumptions.*

A1. Decay of Eigenvalues λ_j^2 of C : *there exists a constant $\kappa > \frac{1}{2}$ such that*

$$\lambda_j \asymp j^{-\kappa}. \quad (13)$$

A2. Domain of Ψ : *there exists an exponent $s \in [0, \kappa - 1/2)$ such Ψ is defined on \mathcal{H}^s .*

A3. Size of Ψ : *the functional $\Psi : \mathcal{H}^s \rightarrow \mathbb{R}$ satisfies the growth conditions*

$$0 \leq \Psi(x) \lesssim 1 + \|x\|_s^2.$$

A4. Derivatives of Ψ : *The derivatives of Ψ satisfy*

$$\|\nabla\Psi(x)\|_{-s} \lesssim 1 + \|x\|_s \quad \text{and} \quad \|\partial^2\Psi(x)\|_{\mathcal{L}(\mathcal{H}^s, \mathcal{H}^{-s})} \lesssim 1.$$

Remark 1 The condition $\kappa > \frac{1}{2}$ ensures that $\text{Trace}_{\mathcal{H}^r}(C_r) < \infty$ for any $r < \kappa - \frac{1}{2}$: this implies that $\pi_0^\tau(\mathcal{H}^r) = 1$ for any $\tau > 0$ and $r < \kappa - \frac{1}{2}$.

Remark 2 The functional $\Psi(x) = \frac{1}{2}\|x\|_s^2$ is defined on \mathcal{H}^s and satisfies Assumptions 1. Its derivative at $x \in \mathcal{H}^s$ is given by $\nabla\Psi(x) = \sum_{j \geq 0} j^{2s} x_j \varphi_j \in \mathcal{H}^{-s}$ with $\|\nabla\Psi(x)\|_{-s} = \|x\|_s$. The second derivative $\partial^2\Psi(x) \in \mathcal{L}(\mathcal{H}^s, \mathcal{H}^{-s})$ is the linear operator that maps $u \in \mathcal{H}^s$ to $\sum_{j \geq 0} j^{2s} \langle u, \varphi_j \rangle \varphi_j \in \mathcal{H}^{-s}$: its norm satisfies $\|\partial^2\Psi(x)\|_{\mathcal{L}(\mathcal{H}^s, \mathcal{H}^{-s})} = 1$ for any $x \in \mathcal{H}^s$.

The Assumptions 1 ensure that the functional Ψ behaves well in a sense made precise in the following lemma.

Lemma 2 *Let Assumptions 1 hold.*

1. *The function $d(x) \stackrel{\text{def}}{=} -(x + C\nabla\Psi(x))$ is globally Lipschitz on \mathcal{H}^s :*

$$\|d(x) - d(y)\|_s \lesssim \|x - y\|_s \quad \forall x, y \in \mathcal{H}^s. \quad (14)$$

2. *The second order remainder term in the Taylor expansion of Ψ satisfies*

$$|\Psi(y) - \Psi(x) - \langle \nabla\Psi(x), y - x \rangle| \lesssim \|y - x\|_s^2 \quad \forall x, y \in \mathcal{H}^s. \quad (15)$$

Proof See [23].

In order to provide a clean exposition, which highlights the central theoretical ideas, we have chosen to make *global* assumptions on Ψ and its derivatives. We believe that our limit theorems could be extended to localized version of these assumptions, at the cost of considerable technical complications in the proofs, by means of stopping-time arguments. The numerical example presented in section 6 corroborates this assertion. There are many applications which satisfy local versions of the assumptions given, including the Bayesian formulation of inverse problems [26] and conditioned diffusions [17].

3 Diffusion Limit Theorem

This section contains a precise statement of the algorithm, statement of the main theorem showing that piecewise linear interpolant of the output of the algorithm converges weakly to a noisy gradient flow described by a SPDE, and proof of the main theorem. The proofs of various technical lemmas are deferred to section 4.

3.1 pCN Algorithm

We now define the Markov chain in \mathcal{H}^s which is reversible with respect to the measure π^τ given by Equation (3). Let $x \in \mathcal{H}^s$ be the current position of the Markov chain. The proposal candidate y is given by (5), so that

$$y = (1 - 2\delta)^{\frac{1}{2}}x + \sqrt{2\delta\tau}\xi \quad \text{where} \quad \xi \stackrel{\mathcal{D}}{\sim} \mathcal{N}(0, C) \quad (16)$$

and $\delta \in (0, \frac{1}{2})$ is a small parameter which we will send to zero in order to obtain the noisy gradient flow. In Equation (16), the random variable ξ is chosen *independent* of x . As described in [3] (see also [7, 26]), at temperature $\tau \in (0, \infty)$ the Metropolis-Hastings acceptance probability for the proposal y is given by

$$\alpha^\delta(x, \xi) = 1 \wedge \exp\left(-\frac{1}{\tau}(\Psi(y) - \Psi(x))\right). \quad (17)$$

For future use, we define the local mean acceptance probability at the current position x via the formula

$$\alpha^\delta(x) = \mathbb{E}_x[\alpha^\delta(x, \xi)]. \quad (18)$$

The chain is then reversible with respect to π^τ . The Markov chain $x^\delta = \{x^{k,\delta}\}_{k \geq 0}$ can be written as

$$\begin{aligned} x^{k+1,\delta} &= \gamma^{k,\delta} y^{k,\delta} + (1 - \gamma^{k,\delta}) x^{k,\delta} \\ y^{k,\delta} &= (1 - 2\delta)^{\frac{1}{2}} x^{k,\delta} + \sqrt{2\delta\tau}\xi^k \end{aligned} \quad (19)$$

In the above equation, the ξ^k are i.i.d Gaussian random variables $\mathcal{N}(0, C)$ and the $\gamma^{k,\delta}$ are Bernoulli random variables which account for the accept-reject mechanism of the Metropolis-Hastings algorithm,

$$\gamma^{k,\delta} \stackrel{\text{def}}{=} \gamma^\delta(x^{k,\delta}, \xi^k, U^k) = \mathbb{1}_{\{U^k < \alpha^\delta(x^{k,\delta}, \xi^k)\}} \stackrel{\mathcal{D}}{\sim} \text{Bernoulli}\left(\alpha^\delta(x^{k,\delta}, \xi^k)\right). \quad (20)$$

for an i.i.d sequence $\{U^k\}_{k \geq 0}$ of random variables uniformly distributed on the interval $(0, 1)$ and independent from all the other sources of randomness. The next lemma will be repeatedly used in the sequel. It states that the size of the jump $y - x$ is of order $\sqrt{\delta}$.

Lemma 3 *Under Assumptions 1 and for any integer $p \geq 1$ the following inequality*

$$\mathbb{E}_x[\|y - x\|_s^p]^{\frac{1}{p}} \lesssim \delta \|x\|_s + \sqrt{\delta} \lesssim \sqrt{\delta} (1 + \|x\|_s)$$

holds for any $\delta \in (0, \frac{1}{2})$.

Proof The definition of the proposal (16) shows that $\|y - x\|_s^p \lesssim \delta^p \|x\|_s^p + \delta^{\frac{p}{2}} \mathbb{E}[\|\xi\|_s^p]$. Fernique's theorem [9] shows that ξ has exponential moments and therefore $\mathbb{E}[\|\xi\|_s^p] < \infty$. This gives the conclusion.

3.2 Diffusion Limit Theorem

Fix a time horizon $T > 0$ and a temperature $\tau \in (0, \infty)$. The piecewise linear interpolant z^δ of the Markov chain (19) is defined by Equation (6). The following is the main result of this article. Note that “weakly” refers to weak convergence of probability measures.

Theorem 4 *Let Assumptions 1 hold. Let the Markov chain x^δ start at a fixed position $x_* \in \mathcal{H}^s$. Then the sequence of processes z^δ converges weakly to z in $C([0, T], \mathcal{H}^s)$, as $\delta \rightarrow 0$, where z solves the \mathcal{H}^s -valued stochastic differential equation*

$$\begin{aligned} dz &= -\left(z + C\nabla\Psi(z)\right) dt + \sqrt{2\tau} dW \\ z_0 &= x_* \end{aligned} \quad (21)$$

and W is a Brownian motion in \mathcal{H}^s with covariance operator equal to C_s .

For conceptual clarity, we derive Theorem 4 as a consequence of the general diffusion approximation Lemma 6. Consider a separable Hilbert space $(\mathcal{H}^s, \langle \cdot, \cdot \rangle_s)$ and a sequence of \mathcal{H}^s -valued Markov chains $x^\delta = \{x^{k,\delta}\}_{k \geq 0}$. The martingale-drift decomposition with time discretization δ of the Markov chain x^δ reads

$$\begin{aligned} x^{k+1,\delta} &= x^{k,\delta} + \mathbb{E}[x^{k+1,\delta} - x^{k,\delta} | x^{k,\delta}] \\ &\quad + \left(x^{k+1,\delta} - x^{k,\delta} - \mathbb{E}[x^{k+1,\delta} - x^{k,\delta} | x^{k,\delta}]\right) \\ &= x^{k,\delta} + d^\delta(x^{k,\delta})\delta + \sqrt{2\tau\delta} \Gamma^\delta(x^{k,\delta}, \xi^k) \end{aligned} \quad (22)$$

where the approximate drift d^δ and volatility term $\Gamma^\delta(x, \xi^k)$ are given by

$$\begin{aligned} d^\delta(x) &= \delta^{-1} \mathbb{E}[x^{k+1,\delta} - x^{k,\delta} | x^{k,\delta} = x] \\ \Gamma^\delta(x, \xi^k) &= (2\tau\delta)^{-1/2} \left(x^{k+1,\delta} - x^{k,\delta} - \mathbb{E}[x^{k+1,\delta} - x^{k,\delta} | x^{k,\delta} = x]\right). \end{aligned} \quad (23)$$

In Equation (22), the conditional expectation $\mathbb{E}[x^{k+1,\delta} - x^{k,\delta} | x^{k,\delta}]$ is given by $\alpha^\delta(x^{k,\delta}, \xi^k) \times (y^{k,\delta} - x^{k,\delta})$ for a proposal $y^{k,\delta}$ and noise term ξ^k as defined in Equation (22). Notice that $\{\Gamma^{k,\delta}\}_{k \geq 0}$, with $\Gamma^{k,\delta} \stackrel{\text{def}}{=} \Gamma^\delta(x^{k,\delta}, \xi^k)$, is a martingale difference array in the sense that $M^{k,\delta} = \sum_{j=0}^k \Gamma^{j,\delta}$ is a martingale adapted to the natural filtration $\mathcal{F}^\delta = \{\mathcal{F}^{k,\delta}\}_{k \geq 0}$ of the Markov chain x^δ . The parameter δ represents a time increment. We define the piecewise linear rescaled noise process by

$$W^\delta(t) = \sqrt{\delta} \sum_{j=0}^k \Gamma^{j,\delta} + \frac{t - t_k}{\sqrt{\delta}} \Gamma^{k+1,\delta} \quad \text{for } t_k \leq t < t_{k+1}. \quad (24)$$

We now show that, as $\delta \rightarrow 0$, if the sequence of approximate drift functions $d^\delta(\cdot)$ converges in the appropriate norm to a limiting drift $d(\cdot)$ and the sequence of rescaled noise process W^δ converges to a Brownian motion then the sequence of piecewise linear interpolants z^δ defined by Equation (6) converges weakly to a diffusion process in \mathcal{H}^s . In order to state the general diffusion approximation Lemma 6, we introduce the following:

Conditions 5 *There exists an integer $p \geq 1$ such that the sequence of Markov chains $x^\delta = \{x^{k,\delta}\}_{k \geq 0}$ satisfies*

1. **Convergence of the drift:** *there exists a globally Lipschitz function $d : \mathcal{H}^s \rightarrow \mathcal{H}^s$ such that*

$$\|d^\delta(x) - d(x)\|_s \lesssim \delta \cdot (1 + \|x\|_s^p) \quad (25)$$

2. **Invariance principle:** *as δ tends to zero the sequence of processes $\{W^\delta\}_{\delta \in (0, \frac{1}{2})}$ defined by Equation (24) converges weakly in $C([0, T], \mathcal{H}^s)$ to a Brownian motion W in \mathcal{H}^s with covariance operator C_s .*

3. **A priori bound:** *the following bound holds*

$$\sup_{\delta \in (0, \frac{1}{2})} \left\{ \delta \cdot \mathbb{E} \left[\sum_{k\delta \leq T} \|x^{k,\delta}\|_s^p \right] \right\} < \infty. \quad (26)$$

Remark 3 The a-priori bound (26) can equivalently be stated as

$$\sup_{\delta \in (0, \frac{1}{2})} \left\{ \mathbb{E} \left[\int_0^T \|z^\delta(u)\|_s^p du \right] \right\} < \infty.$$

It is now proved that Conditions 5 are sufficient to obtain a diffusion approximation for the sequence of rescaled processes z^δ defined by equation (6), as δ tends to zero. Contrary to more classical diffusion approximation for Markov processes results [25, 12] based on infinitesimal generators, the next Lemma exploits specific structures which arise when the limiting process has additive noise and, in particular, is based on exploiting preservation of weak convergence under continuous mappings, together with an explicit construction of the noise process. This idea has previously appeared in the literature in, for example, the articles [23, 24] in the context of MCMC and the article [22], and the references therein, in the context of the derivation of SDEs from ODEs with random data.

Lemma 6 (General Diffusion Approximation for Markov chains)

Consider a separable Hilbert space $(\mathcal{H}^s, \langle \cdot, \cdot \rangle_s)$ and a sequence of \mathcal{H}^s -valued Markov chains $x^\delta = \{x^{k,\delta}\}_{k \geq 0}$ starting at a fixed position in the sense that $x^{0,\delta} = x_$ for all $\delta \in (0, \frac{1}{2})$. Suppose that the drift-martingale decompositions (22) of x^δ satisfy Conditions 5. Then the sequence of rescaled interpolants $z^\delta \in C([0, T], \mathcal{H}^s)$ defined by equation (6) converges weakly in $C([0, T], \mathcal{H}^s)$ to $z \in C([0, T], \mathcal{H}^s)$ given by the stochastic differential equation*

$$dz = d(z) dt + \sqrt{2\tau} dW \quad (27)$$

with initial condition $z_0 = x_$ and where W is a Brownian motion in \mathcal{H}^s with covariance C_s .*

Proof For the sake of clarity, the proof of Lemma 6 is divided into several steps.

– **Integral equation representation.**

Notice that solutions of the \mathcal{H}^s -valued SDE (27) are nothing else than solutions of the following integral equation,

$$z(t) = x_* + \int_0^t d(z(u)) du + \sqrt{2\tau} W(t) \quad \forall t \in (0, T), \quad (28)$$

where W is a Brownian motion in \mathcal{H}^s with covariance operator equal to C_s . We thus introduce the Itô map $\Theta : C([0, T], \mathcal{H}^s) \rightarrow C([0, T], \mathcal{H}^s)$ that sends a function $W \in C([0, T], \mathcal{H}^s)$ to the unique solution of the integral equation (28): solution of (27) can be represented as $\Theta(W)$ where W is an \mathcal{H}^s -valued Brownian motion with covariance C_s . As is described below, the function Θ is continuous if $C([0, T], \mathcal{H}^s)$ is topologized by the uniform norm $\|w\|_{C([0, T], \mathcal{H}^s)} \stackrel{\text{def}}{=} \sup\{\|w(t)\|_s : t \in (0, T)\}$. It is crucial to notice that the rescaled process z^δ , defined in Equation (6), satisfies $z^\delta = \Theta(\widehat{W}^\delta)$ with

$$\widehat{W}^\delta(t) := W^\delta(t) + \frac{1}{\sqrt{2\tau}} \int_0^t [d^\delta(\bar{z}^\delta(u)) - d(z^\delta(u))] du. \quad (29)$$

In Equation (29), the quantity d^δ is the approximate drift defined in Equation (23) and \bar{z}^δ is the rescaled piecewise constant interpolant of $\{x^{k,\delta}\}_{k \geq 0}$ defined as

$$\bar{z}^\delta(t) = x^{k,\delta} \quad \text{for} \quad t_k \leq t < t_{k+1}. \quad (30)$$

The proof follows from a continuous mapping argument (see below) once it is proven that \widehat{W}^δ converges weakly in $C([0, T], \mathcal{H}^s)$ to W .

– **The Itô map Θ is continuous**

It can be proved that Θ is continuous as a mapping from $\left(C([0, T], \mathcal{H}^s), \|\cdot\|_{C([0, T], \mathcal{H}^s)}\right)$ to itself. The usual Picard's iteration proof of the Cauchy-Lipschitz theorem of ODEs may be employed: see [23].

– **The sequence of processes \widehat{W}^δ converges weakly to W**

The process $\widehat{W}^\delta(t)$ is defined by $\widehat{W}^\delta(t) = W^\delta(t) + \frac{1}{\sqrt{2\tau}} \int_0^t [d^\delta(\bar{z}^\delta(u)) - d(z^\delta(u))] du$ and Conditions 5 state that W^δ converges weakly to W in $C([0, T], \mathcal{H}^s)$. Consequently, to prove that $\widehat{W}^\delta(t)$ converges weakly to W in $C([0, T], \mathcal{H}^s)$, it suffices (Slutsky's lemma) to verify that the sequences of processes

$$(\omega, t) \mapsto \int_0^t [d^\delta(\bar{z}^\delta(u)) - d(z^\delta(u))] du \quad (31)$$

converges to zero in probability with respect to the supremum norm in $C([0, T], \mathcal{H}^s)$. By Markov's inequality, it is enough to check that $\mathbb{E}[\int_0^T \|d^\delta(\bar{z}^\delta(u)) - d(z^\delta(u))\|_s du]$ converges to zero as δ goes to zero. Conditions 5 states that there exists an integer $p \geq 1$ such that $\|d^\delta(x) - d(x)\|_s \lesssim \delta \cdot (1 + \|x\|_s^p)$ so that for any $t_k \leq u < t_{k+1}$ we have

$$\left\|d^\delta(\bar{z}^\delta(u)) - d(\bar{z}^\delta(u))\right\|_s \lesssim \delta (1 + \|\bar{z}^\delta(u)\|_s^p) = \delta (1 + \|x^{k, \delta}\|_s^p). \quad (32)$$

Conditions 5 states that $d(\cdot)$ is globally Lipschitz on \mathcal{H}^s . Therefore, Lemma 3 shows that

$$\mathbb{E}\|d(\bar{z}^\delta(u)) - d(z^\delta(u))\|_s \lesssim \mathbb{E}\|x^{k+1, \delta} - x^{k, \delta}\|_s \lesssim \delta^{\frac{1}{2}}(1 + \|x^{k, \delta}\|_s). \quad (33)$$

From estimates (32) and (33) it follows that $\|d^\delta(\bar{z}^\delta(u)) - d(z^\delta(u))\|_s \lesssim \delta^{\frac{1}{2}}(1 + \|x^{k, \delta}\|_s^p)$. Consequently

$$\mathbb{E}\left[\int_0^T \|d^\delta(\bar{z}^\delta(u)) - d(z^\delta(u))\|_s du\right] \lesssim \delta^{\frac{3}{2}} \sum_{k\delta < T} \mathbb{E}\left[1 + \|x^{k, \delta}\|_s^p\right]. \quad (34)$$

The a-priori bound of Conditions 5 shows that this last quantity converges to zero as δ converges to zero, which finishes the proof of Equation (31). This concludes the proof of $\widehat{W}^\delta(t) \implies W$.

– **Continuous mapping argument.**

It has been proved that Θ is continuous as a mapping from $\left(C([0, T], \mathcal{H}^s), \|\cdot\|_{C([0, T], \mathcal{H}^s)}\right)$ to itself. The solutions of the \mathcal{H}^s -valued SDE (27) can be expressed as $\Theta(W)$ while the rescaled continuous interpolate z^δ also reads $z^\delta = \Theta(\widehat{W}^\delta)$. Since \widehat{W}^δ converges weakly in $\left(C([0, T], \mathcal{H}^s), \|\cdot\|_{C([0, T], \mathcal{H}^s)}\right)$ to W as δ tends to zero, the continuous mapping theorem ensures that z^δ converges weakly in $\left(C([0, T], \mathcal{H}^s), \|\cdot\|_{C([0, T], \mathcal{H}^s)}\right)$ to the solution $\Theta(W)$ of the \mathcal{H}^s -valued SDE (27). This ends the proof of Lemma 6.

In order to establish Theorem 4 as a consequence of the general diffusion approximation Lemma 6, it suffices to verify that if Assumptions 1 hold then Conditions 5 are satisfied by the Markov chain x^δ defined in section 3.1. In section 4.2 we prove the following quantitative version of the approximation the function $d^\delta(\cdot)$ by the function $d(\cdot)$ where $d(x) = -\left(x + C\nabla\Psi(x)\right)$.

Lemma 7 (Drift estimate)

Let Assumptions 1 hold and let $p \geq 1$ be an integer. Then the following estimate is satisfied,

$$\|d^\delta(x) - d(x)\|_s^p \lesssim \delta^{\frac{p}{2}}(1 + \|x\|_s^{2p}). \quad (35)$$

Moreover, the approximate drift d^δ is linearly bounded in the sense that

$$\|d^\delta(x)\|_s \lesssim 1 + \|x\|_s. \quad (36)$$

It follows from Lemma (7) that Equation (25) of Conditions 5 is satisfied as soon as Assumptions 1 hold. The invariance principle of Conditions 5 follows from the next lemma. It is proved in section 4.5.

Lemma 8 (Invariance Principle)

Let Assumptions 1 hold. Then the rescaled noise process $W^\delta(t)$ defined in equation (24) satisfies

$$W^\delta \implies W$$

where \implies denotes weak convergence in $C([0, T], \mathcal{H}^s)$, and W is a \mathcal{H}^s -valued Brownian motion with covariance operator C_s .

In section 4.4 it is proved that the following a priori bound is satisfied,

Lemma 9 (A priori bound)

Consider a fixed time horizon $T > 0$ and an integer $p \geq 1$. Under Assumptions 1 the following bound holds,

$$\sup \left\{ \delta \cdot \mathbb{E} \left[\sum_{k\delta \leq T} \|x^{k,\delta}\|_s^p \right] : \delta \in (0, \frac{1}{2}) \right\} < \infty. \quad (37)$$

In conclusion, Lemmas 7 and 8 and 9 together show that Conditions 5 are consequences of Assumptions 1. Therefore, under Assumptions 1, the general diffusion approximation Lemma 6 can be applied: this concludes the proof of Theorem 4.

4 Key Estimates

This section assembles various results which are used in the previous section. Some of the technical proofs are deferred to the appendix.

4.1 Acceptance Probability Asymptotics

This section describes a first order expansion of the acceptance probability. The approximation

$$\alpha^\delta(x, \xi) \approx \bar{\alpha}^\delta(x, \xi) \quad \text{where} \quad \bar{\alpha}^\delta(x, \xi) = 1 - \sqrt{\frac{2\delta}{\tau}} \langle \nabla \Psi(x), \xi \rangle \mathbb{1}_{\{\langle \nabla \Psi(x), \xi \rangle > 0\}} \quad (38)$$

is valid for $\delta \ll 1$. The quantity $\bar{\alpha}^\delta$ has the advantage over α^δ of being very simple to analyse: explicit computations are available. This will be exploited in section 4.2. The quality of the approximation (38) is rigorously quantified in the next lemma.

Lemma 10 (Acceptance probability estimate)

Let Assumptions 1 hold. For any integer $p \geq 1$ the quantity $\bar{\alpha}^\delta(x, \xi)$ satisfies

$$\mathbb{E}_x \left[|\alpha^\delta(x, \xi) - \bar{\alpha}^\delta(x, \xi)|^p \right] \lesssim \delta^p (1 + \|x\|_s^{2p}). \quad (39)$$

Proof See Appendix A.

Recall the local mean acceptance $\alpha^\delta(x)$ defined in Equation (18). Define the approximate local mean acceptance probability by $\bar{\alpha}^\delta(x) \stackrel{\text{def}}{=} \mathbb{E}_x[\bar{\alpha}^\delta(x, \xi)]$. One can use Lemma 10 to approximate the local mean acceptance probability $\alpha^\delta(x)$.

Corollary 1 Let Assumptions 1 hold. For any integer $p \geq 1$ the following estimates hold,

$$|\alpha^\delta(x) - \bar{\alpha}^\delta(x)| \lesssim \delta (1 + \|x\|_s^2) \quad (40)$$

$$\mathbb{E}_x \left[|\alpha^\delta(x, \xi) - 1|^p \right] \lesssim \delta^{\frac{p}{2}} (1 + \|x\|_s^p) \quad (41)$$

Proof See Appendix A.

4.2 Drift Estimates

Explicit computations are available for the quantity $\bar{\alpha}^\delta$. We will use these results, together with quantification of the error committed in replacing α^δ by $\bar{\alpha}^\delta$, to estimate the mean drift (in this section) and the diffusion term (in the next section).

Lemma 11 *For any $x \in \mathcal{H}^s$ the approximate acceptance probability $\bar{\alpha}^\delta(x, \xi)$ satisfies*

$$\sqrt{\frac{2\tau}{\delta}} \mathbb{E}_x \left[\bar{\alpha}^\delta(x, \xi) \cdot \xi \right] = -C\nabla\Psi(x).$$

Proof Let $u = \sqrt{\frac{2\tau}{\delta}} \mathbb{E}_x \left[\bar{\alpha}^\delta(x, \xi) \cdot \xi \right] \in \mathcal{H}^s$. To prove the lemma it suffices to verify that for all $v \in \mathcal{H}^{-s}$ we have $\langle u, v \rangle = -\langle C\nabla\Psi(x), v \rangle$. To this end, use the decomposition $v = \alpha\nabla\Psi(x) + w$ where $\alpha \in \mathbb{R}$ and $w \in \mathcal{H}^{-s}$ satisfies $\langle C\nabla\Psi(x), w \rangle = 0$. Since $\xi \stackrel{\mathcal{D}}{\sim} N(0, C)$ the two Gaussian random variables $Z_\Psi \stackrel{\text{def}}{=} \langle \nabla\Psi(x), \xi \rangle$ and $Z_w \stackrel{\text{def}}{=} \langle w, \xi \rangle$ are independent: indeed, (Z_Ψ, Z_w) is a Gaussian vector in \mathbb{R}^2 with $\text{Cov}(Z_\Psi, Z_w) = 0$. It thus follows that

$$\begin{aligned} \langle u, v \rangle &= -2 \langle \mathbb{E}_x \left[\langle \nabla\Psi(x), \xi \rangle 1_{\{\langle \nabla\Psi(x), \xi \rangle > 0\}} \cdot \xi \right], \alpha\nabla\Psi(x) + w \rangle \\ &= -2 \mathbb{E}_x \left[\alpha Z_\Psi^2 1_{\{Z_\Psi > 0\}} + Z_w Z_\Psi 1_{\{Z_\Psi > 0\}} \right] \\ &= -2\alpha \mathbb{E}_x \left[Z_\Psi^2 1_{\{Z_\Psi > 0\}} \right] = -\alpha \mathbb{E}_x \left[Z_\Psi^2 \right] \\ &= -\alpha \langle C\nabla\Psi(x), \nabla\Psi(x) \rangle = \langle -C\nabla\Psi(x), \alpha\nabla\Psi(x) + w \rangle \\ &= -\langle C\nabla\Psi(x), v \rangle, \end{aligned}$$

which concludes the proof of Lemma 11.

We now use this explicit computation to give a proof of the drift estimate Lemma 7.

Proof (Proof of Lemma 7) The function d^δ defined by Equation (23) can also be expressed as

$$d^\delta(x) = \left\{ \frac{(1-2\delta)^{\frac{1}{2}} - 1}{\delta} \alpha^\delta(x) x \right\} + \left\{ \sqrt{\frac{2\tau}{\delta}} \mathbb{E}_x[\alpha^\delta(x, \xi) \xi] \right\} = B_1 + B_2, \quad (42)$$

where the mean local acceptance probability $\alpha^\delta(x)$ has been defined in Equation (18) and the two terms B_1 and B_2 are studied below. To prove Equation (35), it suffices to establish that

$$\|B_1 + x\|_s^p \lesssim \delta^{\frac{p}{2}} (1 + \|x\|_s^{2p}) \quad \text{and} \quad \|B_2 + C\nabla\Psi(x)\|_s^p \lesssim \delta^{\frac{p}{2}} (1 + \|x\|_s^{2p}). \quad (43)$$

We now establish these two bounds.

– Lemma 10 and Corollary 1 show that

$$\begin{aligned} \|B_1 + x\|_s^p &= \left\{ \frac{(1-2\delta)^{\frac{1}{2}} - 1}{\delta} \alpha^\delta(x) + 1 \right\}^p \|x\|_s^p \\ &\lesssim \left\{ \left| \frac{(1-2\delta)^{\frac{1}{2}} - 1}{\delta} - 1 \right|^p + |\alpha^\delta(x) - 1|^p \right\} \|x\|_s^p \\ &\lesssim \left\{ \delta^p + \delta^{\frac{p}{2}} (1 + \|x\|_s^p) \right\} \|x\|_s^p \lesssim \delta^{\frac{p}{2}} (1 + \|x\|_s^{2p}). \end{aligned} \quad (44)$$

– Lemma 10 shows that

$$\begin{aligned} \|B_2 + C\nabla\Psi(x)\|_s^p &= \left\| \sqrt{\frac{2\tau}{\delta}} \mathbb{E}_x[\alpha^\delta(x, \xi) \xi] + C\nabla\Psi(x) \right\|_s^p \\ &\lesssim \delta^{-\frac{p}{2}} \left\| \mathbb{E}_x[\{\alpha^\delta(x, \xi) - \bar{\alpha}^\delta(x, \xi)\} \xi] \right\|_s^p \end{aligned} \quad (45)$$

$$+ \underbrace{\left\| \sqrt{\frac{2\tau}{\delta}} \mathbb{E}_x[\bar{\alpha}^\delta(x, \xi) \xi] + C\nabla\Psi(x) \right\|_s^p}_{=0}.$$

By Lemma 11, the second term on the right hand equals to zero. Consequently, the Cauchy-Schwarz inequality implies that

$$\begin{aligned} \|B_2 + C\nabla\Psi(x)\|_s^p &\lesssim \delta^{-\frac{p}{2}} \mathbb{E}_x[|\alpha^\delta(x, \xi) - \bar{\alpha}^\delta(x, \xi)|^2]^{\frac{p}{2}} \\ &\lesssim \delta^{-\frac{p}{2}} \left(\delta^2(1 + \|x\|_s^4) \right)^{\frac{p}{2}} \lesssim \delta^{\frac{p}{2}}(1 + \|x\|_s^{2p}). \end{aligned}$$

Estimates (44) and (45) give Equation (43). To complete the proof we establish the bound (36). The expression (42) shows that it suffices to verify $\delta^{-\frac{1}{2}} \mathbb{E}_x[\alpha^\delta(x, \xi) \xi] \lesssim 1 + \|x\|_s$. To this end, we use Lemma 11 and Corollary 1. By the Cauchy-Schwarz inequality,

$$\begin{aligned} \left\| \delta^{-\frac{1}{2}} \mathbb{E}_x[\alpha^\delta(x, \xi) \cdot \xi] \right\|_s &= \left\| \delta^{-\frac{1}{2}} \mathbb{E}_x[(\alpha^\delta(x, \xi) - 1) \cdot \xi] \right\|_s \\ &\lesssim \delta^{-\frac{1}{2}} \mathbb{E}_x[(\alpha^\delta(x, \xi) - 1)^2]^{\frac{1}{2}} \lesssim 1 + \|x\|_s, \end{aligned}$$

which concludes the proof of Lemma 7.

4.3 Noise Estimates

In this section we estimate the error in the approximation $\Gamma^{k,\delta} \approx \mathbf{N}(0, C_s)$. To this end, let us introduce the covariance operator $D^\delta(x) = \mathbb{E}[\Gamma^{k,\delta} \otimes_{\mathcal{H}^s} \Gamma^{k,\delta} | x^{k,\delta} = x]$ of the martingale difference Γ^δ . For any $x, u, v \in \mathcal{H}^s$ the operator $D^\delta(x)$ satisfies

$$\mathbb{E}[\langle \Gamma^{k,\delta}, u \rangle_s \langle \Gamma^{k,\delta}, v \rangle_s | x^{k,\delta} = x] = \langle u, D^\delta(x)v \rangle_s.$$

The next lemma gives a quantitative version of the approximation of $D^\delta(x)$ by the operator C_s .

Lemma 12 (Noise estimates)

Let Assumptions 1 hold. For any pair of indices $i, j \geq 1$, the martingale difference term $\Gamma^\delta(x, \xi)$ satisfies

$$|\langle \hat{\varphi}_i, D^\delta(x) \hat{\varphi}_j \rangle_s - \langle \hat{\varphi}_i, C_s \hat{\varphi}_j \rangle_s| \lesssim \delta^{\frac{1}{8}} \cdot (1 + \|x\|_s) \quad (46)$$

$$|\text{Trace}_{\mathcal{H}^s}(D^\delta(x)) - \text{Trace}_{\mathcal{H}^s}(C_s)| \lesssim \delta^{\frac{1}{8}} \cdot (1 + \|x\|_s^2). \quad (47)$$

with $\{\hat{\varphi}_j = j^{-s} \varphi_j\}_{j \geq 0}$ is an orthonormal basis of \mathcal{H}^s .

Proof See Appendix A.

4.4 A Priori Bound

Now we have all the ingredients for the proof of the a priori bound presented in Lemma 9 which states that the rescaled process z^δ given by Equation (6) does not blow up in finite time.

Proof (Proof Lemma 9) Without loss of generality, assume that $p = 2n$ for some positive integer $n \geq 1$. We now prove that there exist constants $\alpha_1, \alpha_2, \alpha_3 > 0$ satisfying

$$\mathbb{E}[\|x^{k,\delta}\|_s^{2n}] \leq (\alpha_1 + \alpha_2 k \delta) e^{\alpha_3 k \delta}. \quad (48)$$

Lemma 9 is a straightforward consequence of Equation 48 since this implies that

$$\delta \sum_{k\delta < T} \mathbb{E}[\|x^{k,\delta}\|_s^{2n}] \leq \delta \sum_{k\delta < T} (\alpha_1 + \alpha_2 k \delta) e^{\alpha_3 k \delta} \asymp \int_0^T (\alpha_1 + \alpha_2 t) e^{\alpha_3 t} dt < \infty.$$

For notational convenience, let us define $V^{k,\delta} = \mathbb{E}[\|x^{k,\delta}\|_s^{2n}]$. To prove Equation (48), it suffices to establish that

$$V^{k+1,\delta} - V^{k,\delta} \leq K \delta \cdot (1 + V^{k,\delta}), \quad (49)$$

where $K > 0$ is constant independent from $\delta \in (0, \frac{1}{2})$. Indeed, iterating inequality (49) leads to the bound (48), for some computable constants $\alpha_1, \alpha_2, \alpha_3 > 0$. The definition of V^k shows that

$$\begin{aligned} V^{k+1,\delta} - V^{k,\delta} &= \mathbb{E}[\|x^{k,\delta} + (x^{k+1,\delta} - x^{k,\delta})\|_s^{2n} - \|x^{k,\delta}\|_s^{2n}] \\ &= \mathbb{E}\left[\|x^{k,\delta}\|_s^2 + \|x^{k+1,\delta} - x^{k,\delta}\|_s^2 \right. \\ &\quad \left. + 2\langle x^{k,\delta}, x^{k+1,\delta} - x^{k,\delta} \rangle_s\right]^n - \|x^{k,\delta}\|_s^{2n} \end{aligned} \quad (50)$$

where the increment $x^{k+1,\delta} - x^{k,\delta}$ is given by

$$x^{k+1,\delta} - x^{k,\delta} = \gamma^{k,\delta} \left((1 - 2\delta)^{\frac{1}{2}} - 1 \right) x^{k,\delta} + \sqrt{2\delta} \gamma^{k,\delta} \xi^k. \quad (51)$$

To bound the right-hand-side of Equation (50), we use a binomial expansion and control each term. To this end, we establish the following estimate: for all integers $i, j, k \geq 0$ satisfying $i+j+k = n$ and $(i, j, k) \neq (n, 0, 0)$ the following inequality holds,

$$\begin{aligned} \mathbb{E}\left[\left(\|x^{k,\delta}\|_s^2\right)^i \times \left(\|x^{k+1,\delta} - x^{k,\delta}\|_s^2\right)^j \right. \\ \left. \times \left(\langle x^{k,\delta}, x^{k+1,\delta} - x^{k,\delta} \rangle_s\right)^k\right] \lesssim \delta (1 + V^{k,\delta}). \end{aligned} \quad (52)$$

To prove Equation (52), we separate two different cases.

- Let us suppose $(i, j, k) = (n-1, 0, 1)$. Lemma 7 states that the approximate drift has a linearly bounded growth so that

$$\left\| \mathbb{E}[x^{k+1,\delta} - x^{k,\delta} | x^{k,\delta}] \right\|_s = \delta \|d^\delta(x^{k,\delta})\|_s \lesssim \delta (1 + \|x^{k,\delta}\|_s).$$

Consequently, we have

$$\begin{aligned} \mathbb{E}\left[\left(\|x^{k,\delta}\|_s^2\right)^{n-1} \langle x^{k,\delta}, x^{k+1,\delta} - x^{k,\delta} \rangle_s\right] &\lesssim \mathbb{E}\left[\|x^{k,\delta}\|_s^{2(n-1)} \|x^{k,\delta}\|_s \left(\delta (1 + \|x^{k,\delta}\|_s)\right)\right] \\ &\lesssim \delta (1 + V^{k,\delta}). \end{aligned}$$

This proves Equation (52) in the case $(i, j, k) = (n-1, 0, 1)$.

- Let us suppose $(i, j, k) \notin \{(n, 0, 0), (n-1, 0, 1)\}$. Because for any integer $p \geq 1$,

$$\mathbb{E}_x \left[\|x^{k+1,\delta} - x^{k,\delta}\|_s^p \right]^{\frac{1}{p}} \lesssim \delta^{\frac{1}{2}} (1 + \|x\|_s)$$

it follows from the Cauchy-Schwarz inequality that

$$\mathbb{E}\left[\left(\|x^{k,\delta}\|_s^2\right)^i \left(\|x^{k+1,\delta} - x^{k,\delta}\|_s^2\right)^j \left(\langle x^{k,\delta}, x^{k+1,\delta} - x^{k,\delta} \rangle_s\right)^k\right] \lesssim \delta^{j+\frac{k}{2}} (1 + V^{k,\delta}).$$

Since we have supposed that $(i, j, k) \notin \{(n, 0, 0), (n-1, 0, 1)\}$ and $i+j+k = n$, it follows that $j + \frac{k}{2} \geq 1$.

This concludes the proof of Equation (52),

The binomial expansion of Equation (50) and the bound (52) show that Equation (49) holds. This concludes the proof of Lemma 9.

4.5 Invariance Principle

Combining the noise estimates of Lemma 12 and the a priori bound of Lemma 9, we show that under Assumptions 1 the sequence of rescaled noise processes defined in Equation 24 converges weakly to a Brownian motion. This is the content of Lemma 8 whose proof is now presented.

Proof (Proof of Lemma 8) As described in [1] [Proposition 5.1], in order to prove that W^δ converges weakly to W in $C([0, T], \mathcal{H}^s)$ it suffices to prove that for any $t \in [0, T]$ and any pair of indices $i, j \geq 0$ the following three limits hold in probability,

$$\lim_{\delta \rightarrow 0} \delta \sum_{k\delta < t} \mathbb{E} \left[\| \Gamma^{k, \delta} \|_s^2 | x^{k, \delta} \right] = t \cdot \text{Trace}_{\mathcal{H}^s}(C_s) \quad (53)$$

$$\lim_{\delta \rightarrow 0} \delta \sum_{k\delta < t} \mathbb{E} \left[\langle \Gamma^{k, \delta}, \hat{\varphi}_i \rangle_s \langle \Gamma^{k, \delta}, \hat{\varphi}_j \rangle_s | x^{k, \delta} \right] = t \langle \hat{\varphi}_i, C_s \hat{\varphi}_j \rangle_s \quad (54)$$

$$\lim_{\delta \rightarrow 0} \delta \sum_{k\delta < T} \mathbb{E} \left[\| \Gamma^{k, \delta} \|_s^2 \mathbb{1}_{\{ \| \Gamma^{k, \delta} \|_s^2 \geq \delta^{-1} \varepsilon \}} | x^{k, \delta} \right] = 0 \quad \forall \varepsilon > 0. \quad (55)$$

We now check that these three conditions are indeed satisfied.

- Condition (53): since $\mathbb{E} \left[\| \Gamma^{k, \delta} \|_s^2 | x^{k, \delta} \right] = \text{Trace}_{\mathcal{H}^s}(D^\delta(x^{k, \delta}))$, Lemma 12 shows that

$$\mathbb{E} \left[\| \Gamma^{k, \delta} \|_s^2 | x^{k, \delta} \right] = \text{Trace}_{\mathcal{H}^s}(C_s) + \mathbf{e}_1^\delta(x^{k, \delta})$$

where the error term \mathbf{e}_1^δ satisfies $|\mathbf{e}_1^\delta(x)| \lesssim \delta^{\frac{1}{8}} (1 + \|x\|_s^2)$. Consequently, to prove condition (53) it suffices to establish that

$$\lim_{\delta \rightarrow 0} \mathbb{E} \left[\left| \delta \sum_{k\delta < T} \mathbf{e}_1^\delta(x^{k, \delta}) \right| \right] = 0.$$

We have $\mathbb{E} \left[\left| \delta \sum_{k\delta < T} \mathbf{e}_1^\delta(x^{k, \delta}) \right| \right] \lesssim \delta^{\frac{1}{8}} \left\{ \delta \cdot \mathbb{E} \left[\sum_{k\delta < T} (1 + \|x^{k, \delta}\|_s^2) \right] \right\}$ and the a priori bound presented in Lemma 9 shows that

$$\sup_{\delta \in (0, \frac{1}{2})} \left\{ \delta \cdot \mathbb{E} \left[\sum_{k\delta < T} (1 + \|x^{k, \delta}\|_s^2) \right] \right\} < \infty.$$

Consequently $\lim_{\delta \rightarrow 0} \mathbb{E} \left[\left| \delta \sum_{k\delta < T} \mathbf{e}_1^\delta(x^{k, \delta}) \right| \right] = 0$, and the conclusion follows.

- Condition (54): Lemma 12 states that

$$\mathbb{E}_k \left[\langle \Gamma^{k, \delta}, \hat{\varphi}_i \rangle_s \langle \Gamma^{k, \delta}, \hat{\varphi}_j \rangle_s \right] = \langle \hat{\varphi}_i, C_s \hat{\varphi}_j \rangle_s + \mathbf{e}_2^\delta(x^{k, \delta})$$

where the error term \mathbf{e}_2^δ satisfies $|\mathbf{e}_2^\delta(x)| \lesssim \delta^{\frac{1}{8}} (1 + \|x\|_s)$. The exact same approach as the proof of Condition (53) gives the conclusion.

- Condition (55): from the Cauchy-Schwarz and Markov's inequalities it follows that

$$\begin{aligned} \mathbb{E} \left[\| \Gamma^{k, \delta} \|_s^2 \mathbb{1}_{\{ \| \Gamma^{k, \delta} \|_s^2 \geq \delta^{-1} \varepsilon \}} \right] &\leq \mathbb{E} \left[\| \Gamma^{k, \delta} \|_s^4 \right]^{\frac{1}{2}} \cdot \mathbb{P} \left[\| \Gamma^{k, \delta} \|_s^2 \geq \delta^{-1} \varepsilon \right]^{\frac{1}{2}} \\ &\leq \mathbb{E} \left[\| \Gamma^{k, \delta} \|_s^4 \right]^{\frac{1}{2}} \cdot \left\{ \frac{\mathbb{E} \left[\| \Gamma^{k, \delta} \|_s^4 \right]}{(\delta^{-1} \varepsilon)^2} \right\}^{\frac{1}{2}} \\ &\leq \frac{\delta}{\varepsilon} \cdot \mathbb{E} \left[\| \Gamma^{k, \delta} \|_s^4 \right]. \end{aligned}$$

Lemma 7 readily shows that $\mathbb{E} \| \Gamma^{k, \delta} \|_s^4 \lesssim 1 + \|x\|_s^4$. Consequently we have

$$\mathbb{E} \left[\left| \delta \sum_{k\delta < T} \mathbb{E} \left[\| \Gamma^{k, \delta} \|_s^2 \mathbb{1}_{\{ \| \Gamma^{k, \delta} \|_s^2 \geq \delta^{-1} \varepsilon \}} | x^{k, \delta} \right] \right| \right] \leq \frac{\delta}{\varepsilon} \times \left\{ \delta \cdot \mathbb{E} \left[\sum_{k\delta < T} (1 + \|x^{k, \delta}\|_s^4) \right] \right\}$$

and the conclusion again follows from the a priori bound Lemma 9.

5 Quadratic Variation

As discussed in the introduction, the SPDE (7), and the Metropolis-Hastings algorithm pCN which approximates it for small δ , do not satisfy the smoothing property and so almost sure properties of the limit measure π^τ are not necessarily seen at finite time. To illustrate this point, we introduce in this section a functional $V : \mathcal{H} \rightarrow \mathbb{R}$ that is well defined on a dense subset of \mathcal{H} and such that $V(X)$ is π^τ -almost surely well defined and satisfies $\mathbb{P}(V(X) = 1) = \tau$ for $X \stackrel{\mathcal{D}}{\sim} \pi^\tau$. The quantity V corresponds to the usual quadratic variation if π_0 is the Wiener measure. We show that the quadratic variation like quantity $V(x^{k,\tau})$ of a pCN Markov chain converges as $k \rightarrow \infty$ to the almost sure quantity τ . We then prove that piecewise linear interpolation of this quantity solves, in the small δ limit, a linear ODE (the ‘‘fluid limit’’) whose globally attractive stable state is the almost sure quantity τ . This quantifies the manner in which the pCN method approaches statistical equilibrium.

5.1 Definition and Properties

Under Assumptions 1, the Karhunen-Lo eve expansion shows that π_0 -almost every $x \in \mathcal{H}$ satisfies

$$\lim_{N \rightarrow \infty} N^{-1} \sum_{j=1}^N \frac{\langle x, \varphi_j \rangle^2}{\lambda_j^2} = 1.$$

This motivates the definition of the quadratic variation like quantities

$$V_-(x) \stackrel{\text{def}}{=} \liminf_{N \rightarrow \infty} N^{-1} \sum_{j=1}^N \frac{\langle x, \varphi_j \rangle^2}{\lambda_j^2} \quad \text{and} \quad V_+(x) \stackrel{\text{def}}{=} \limsup_{N \rightarrow \infty} N^{-1} \sum_{j=1}^N \frac{\langle x, \varphi_j \rangle^2}{\lambda_j^2}.$$

When these two quantities are equal the vector $x \in \mathcal{H}$ is said to possess a *quadratic variation* $V(x)$ defined as $V(x) = V_-(x) = V_+(x)$. Consequently, π_0 -almost every $x \in \mathcal{H}$ possesses a quadratic variation $V(x) = 1$. It is a straightforward consequence that π_0^τ -almost every and π^τ -almost every $x \in \mathcal{H}$ possesses a quadratic variation $V(x) = \tau$. Strictly speaking this only coincides with quadratic variation when C is the covariance of a (possibly conditioned) Brownian motion; however we use the terminology more generally in this section. The next lemma proves that the quadratic variation $V(\cdot)$ behaves as it should do with respect to additivity.

Lemma 13 (Quadratic Variation Additivity)

Consider a vector $x \in \mathcal{H}$ and a Gaussian random variable $\xi \stackrel{\mathcal{D}}{\sim} \pi_0$ and a real number $\alpha \in \mathbb{R}$. Suppose that the vector $x \in \mathcal{H}$ possesses a finite quadratic variation $V(x) < +\infty$. Then almost surely the vector $x + \alpha\xi \in \mathcal{H}$ possesses a quadratic variation that is equal to

$$V(x + \alpha\xi) = V(x) + \alpha^2.$$

Proof Let us define $V_N \stackrel{\text{def}}{=} N^{-1} \sum_1^N \frac{\langle x, \varphi_j \rangle \cdot \langle \xi, \varphi_j \rangle}{\lambda_j^2}$. To prove Lemma 13 it suffices to prove that almost surely the following limit holds

$$\lim_{N \rightarrow \infty} V_N = 0.$$

Borel-Cantelli Lemma shows that it suffices to prove that for every fixed $\varepsilon > 0$ we have $\sum_{N \geq 1} \mathbb{P}[|V_N| > \varepsilon] < \infty$. Notice then that V_N is a centred Gaussian random variables with variance

$$\text{Var}(V_N) = \frac{1}{N} \left(N^{-1} \sum_1^N \frac{\langle x, \varphi_j \rangle^2}{\lambda_j^2} \right) \asymp \frac{V(x)}{N}.$$

It readily follows that $\sum_{N \geq 1} \mathbb{P}[|V_N| > \varepsilon] < \infty$, finishing the proof of the Lemma.

5.2 Large k Behaviour of Quadratic Variation for pCN

The pCN algorithm at temperature $\tau > 0$ and discretization parameter $\delta > 0$ proposes a move from x to y according to the dynamics

$$y = (1 - 2\delta)^{\frac{1}{2}}x + (2\delta\tau)^{\frac{1}{2}}\xi \quad \text{with} \quad \xi \stackrel{\mathcal{D}}{\sim} \pi_0.$$

This move is accepted with probability $\alpha^\delta(x, y)$. In this case, Lemma 13 shows that if the quadratic variation $V(x)$ exists then the quadratic variation of the proposed move $y \in \mathcal{H}$ exists and satisfies

$$\frac{V(y) - V(x)}{\delta} = -2(V(x) - \tau). \quad (56)$$

Consequently, one can prove that for any finite time step $\delta > 0$ and temperature $\tau > 0$ the quadratic variation of the MCMC algorithm converges to τ .

Proposition 14 (Limiting Quadratic Variation) *Let Assumptions 1 hold and $\{x^{k,\delta}\}_{k \geq 0}$ be the Markov chain of section 3.1. Then almost surely the quadratic variation of the Markov chain converges to τ ,*

$$\lim_{k \rightarrow \infty} V(x^{k,\delta}) = \tau.$$

Proof Let us first show that the number of accepted moves is infinite. If this were not the case, the Markov chain would eventually reach a position $x^{k,\delta} = x \in \mathcal{H}$ such that all subsequent proposals $y^{k+l} = (1 - 2\delta)^{\frac{1}{2}}x^k + (2\delta\tau)^{\frac{1}{2}}\xi^{k+l}$ would be refused. This means that the i.i.d. Bernoulli random variables $\gamma^{k+l} = \text{Bernoulli}(\alpha^\delta(x^k, y^{k+l}))$ satisfy $\gamma^{k+l} = 0$ for all $l \geq 0$. This can only happen with probability zero. Indeed, since $\mathbb{P}[\gamma^{k+l} = 1] > 0$, one can use Borel-Cantelli Lemma to show that almost surely there exists $l \geq 0$ such that $\gamma^{k+l} = 1$. To conclude the proof of the Proposition, notice then that the sequence $\{u_k\}_{k \geq 0}$ defined by $u_{k+1} - u_k = -2\delta(u_k - \tau)$ converges to τ .

5.3 Fluid Limit for Quadratic Variation of pCN

To gain further insight into the rate at which the limiting behaviour of the quadratic variation is observed for pCN we derive an ODE ‘‘fluid limit’’ for the Metropolis-Hastings algorithm. We introduce the continuous time process $t \mapsto v^\delta(t)$ defined as continuous piecewise linear interpolation of the process $k \mapsto V(x^{k,\delta})$,

$$v^\delta(t) = \frac{1}{\delta}(t - t_k)V(x^{k+1,\delta}) + \frac{1}{\delta}(t_{k+1} - t)V(x^{k,\delta}) \quad \text{for } t_k \leq t < t_{k+1}. \quad (57)$$

Since the acceptance probability of pCN approaches one as $\delta \rightarrow 0$ (see Corollary 1) Equation (56) shows heuristically that the trajectories of the process $t \mapsto v^\delta(t)$ should be well approximated by the solution of the (non-stochastic) differential equation

$$\dot{v} = -2(v - \tau). \quad (58)$$

We prove such a result, in the sense of convergence in probability in $C([0, T], \mathbb{R})$:

Theorem 15 (Fluid Limit For Quadratic Variation) *Let Assumptions 1 hold. Let the Markov chain x^δ start at fixed position $x_* \in \mathcal{H}^s$. Assume that $x_* \in \mathcal{H}$ possesses a finite quadratic variation, $V(x_*) < \infty$. Then the function $v^\delta(t)$ converges in probability in $C([0, T], \mathbb{R})$, as δ goes to zero, to the solution of the differential equation (58) with initial condition $v_0 = V(x_*)$.*

As already indicated, the heart of the proof consists in showing that the acceptance probability of the algorithm converges to one as δ goes to zero. We prove such a result as Lemma 16 below, and then proceed to prove Theorem 15. To this end we introduce $t^\delta(k)$, the number of accepted moves,

$$t^\delta(k) \stackrel{\text{def}}{=} \sum_{l \leq k} \gamma^{l,\delta},$$

where $\gamma^{j,\delta} = \text{Bernoulli}(\alpha^\delta(x, y))$ is the Bernoulli random variable defined in Equation (20). Since the acceptance probability of the algorithm converges to 1 as $\delta \rightarrow 0$, the approximation $t^\delta(k) \approx k$ holds. In order to prove a fluid limit result on the interval $[0, T]$ one needs to prove that the quantity $|t^\delta(k) - k|$ is small when compared to δ^{-1} . The next Lemma shows that such a bounds holds uniformly on the interval $[0, T]$.

Lemma 16 (Number of Accepted Moves) *Let Assumptions 1 hold. The number of accepted moves $t^\delta(\cdot)$ verifies*

$$\lim_{\delta \rightarrow 0} \sup \{ \delta \cdot |t^\delta(k) - k| : 0 \leq k \leq T\delta^{-1} \} = 0$$

where the convergence holds in probability.

Proof The proof is given in Appendix B.

We now complete the proof of Theorem 15 using the key Lemma 16.

Proof (of Theorem 15) The proof consists in showing that the trajectory of the quadratic variation process behaves as if all the move were accepted. The main ingredient is the uniform lower bound on the acceptance probability given by Lemma 16.

Recall that $v^\delta(k\delta) = V(x^{k,\delta})$. Consider the piecewise linear function $\hat{v}^\delta(\cdot) \in C([0, T], \mathbb{R})$ defined by linear interpolation of the values $\hat{v}^\delta(k\delta) = u^\delta(k)$ and where the sequence $\{u^\delta(k)\}_{k \geq 0}$ satisfies $u^\delta(0) = V(x_*)$ and

$$u^\delta(k+1) - u^\delta(k) = -2\delta(u^\delta(k) - \tau).$$

The value $u^\delta(k) \in \mathbb{R}$ represents the quadratic variation of $x^{k,\delta}$ if the k first moves of the MCMC algorithm had been accepted. One can readily check that as δ goes to zero the sequence of continuous functions $\hat{v}^\delta(\cdot)$ converges in $C([0, T], \mathbb{R})$ to the solution $v(\cdot)$ of the differential equation (58). Consequently, to prove Theorem 15 it suffices to show that for any $\varepsilon > 0$ we have

$$\lim_{\delta \rightarrow 0} \mathbb{P} \left[\sup \left\{ |V(x^{k,\delta}) - u^\delta(k)| : k \leq \delta^{-1}T \right\} > \varepsilon \right] = 0. \quad (59)$$

The definition of the number of accepted moves $t^\delta(k)$ is such that $V(x^{k,\delta}) = u^\delta(t^\delta(k))$. Note that

$$u^\delta(k) = (1 - 2\delta)^k u_0 + (1 - (1 - 2\delta)^k) \tau. \quad (60)$$

Hence, for any integers $t_1, t_2 \geq 0$, we have $|u^\delta(t_2) - u^\delta(t_1)| \leq |u^\delta(|t_2 - t_1|) - u^\delta(0)|$ so that

$$|V(x^{k,\delta}) - u^\delta(k)| = |u^\delta(t^\delta(k)) - u^\delta(k)| \leq |u^\delta(k - t^\delta(k)) - u^\delta(0)|.$$

Equation (60) shows that $|u^\delta(k) - u^\delta(0)| \lesssim (1 - (1 - 2\delta)^k)$. This implies that

$$|V(x^{k,\delta}) - u^\delta(k)| \lesssim 1 - (1 - 2\delta)^{k - t^\delta(k)} \lesssim 1 - (1 - 2\delta)^{\delta^{-1}S}$$

where $S = \sup \{ \delta \cdot |t^\delta(k) - k| : 0 \leq k \leq T\delta^{-1} \}$. Since for any $a > 0$ we have $1 - (1 - 2\delta)^{a\delta^{-1}} \rightarrow 1 - e^{-2a}$, Equation (59) follows if one can prove that as δ goes to zero the supremum S converges to zero in probability: this is precisely the content of Lemma 16. This concludes the proof of Theorem 15.

6 Numerical Results

In this section, we present some numerical simulations demonstrating our results. We consider the minimization of a functional $J(\cdot)$ defined on the Sobolev space $H_0^1(\mathbb{R})$. Note that functions $x \in H_0^1([0, 1])$ are continuous and satisfy $x(0) = x(1) = 0$; thus $H_0^1(\mathbb{R}) \subset C^0([0, 1]) \subset L^2(0, 1)$. For a given real parameter $\lambda > 0$, the functional $J : H_0^1([0, 1]) \rightarrow \mathbb{R}$ is composed of two competitive terms, as follows:

$$J(x) = \frac{1}{2} \int_0^1 |\dot{x}(s)|^2 ds + \frac{\lambda}{4} \int_0^1 (x(s)^2 - 1)^2 ds. \quad (61)$$

The first term penalizes functions that deviate from being flat, whilst the second term penalizes functions that deviate from one in absolute value. Critical points of the functional $J(\cdot)$ solve the following Euler-Lagrange equation:

$$\begin{aligned} \ddot{x} + \lambda x(1 - x^2) &= 0 \\ x(0) = x(1) &= 0. \end{aligned} \quad (62)$$

Clearly $x \equiv 0$ is a solution for all $\lambda \in \mathbb{R}^+$. If $\lambda \in (0, \pi^2)$ then this is the unique solution of the Euler-Lagrange equation and is the global minimizer of J . For each integer k there is a supercritical bifurcation at parameter value $\lambda = k^2\pi^2$. For $\lambda > \pi^2$ there are two minimizers, both of one sign and one being minus the other. The three different solutions of (62) which exist for $\lambda = 2\pi^2$ are displayed in Figure 1, at which value the zero (blue dotted) solution is a saddle point, and the two green solutions are the global minimizers of J . These properties of J are overviewed in, for example, [18]. We will show how these global minimizers can emerge from an algorithm whose only ingredients are an ability to evaluate Ψ and to sample from the Gaussian measure with Cameron-Martin norm $\int_0^1 |\dot{x}(s)|^2 ds$. We emphasize that we are not advocating this as the optimal method for solving the Euler-Lagrange equations (62). We have chosen this example for its simplicity, in order to illustrate the key ingredients of the theory developed in this paper.

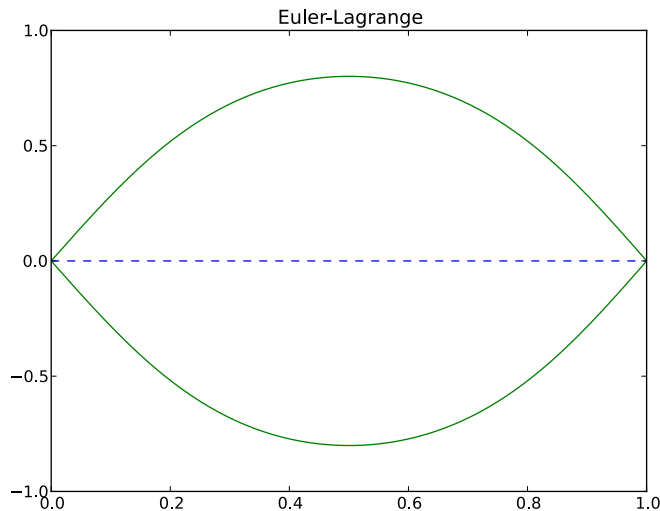


Fig. 1 The three solutions of the Euler-Lagrange Equation (62) for $\lambda = 2\pi^2$. Only the two non-zero solutions are global minimum of the functional $J(\cdot)$. The dotted solution is a local maximum of $J(\cdot)$.

The pCN algorithm to minimize J given by (61) is implemented on $L^2([0, 1])$. Recall from [9] that the Gaussian measure $N(0, C)$ may be identified by finding the covariance operator for which the $H_0^1([0, 1])$ norm $\|x\|_C^2 \stackrel{\text{def}}{=} \int_0^1 |\dot{x}(s)|^2 ds$ is the Cameron-Martin norm. In [15] it is shown that the Wiener bridge measure $\mathbb{W}_{0 \rightarrow 0}$ on $L^2([0, 1])$ has precisely this Cameron-Martin norm; indeed it is demonstrated that C^{-1} is the densely

defined operator $-\frac{d^2}{ds^2}$ with $D(C^{-1}) = H^2([0, 1]) \cap H_0^1([0, 1])$. In this regard it is also instructive to adopt the physicists viewpoint that

$$\mathbb{W}_{0 \rightarrow 0}(dx) \propto \exp\left(-\frac{1}{2} \int_0^1 |\dot{x}(s)|^2 ds\right) dx$$

although, of course, there is no Lebesgue measure in infinite dimensions. Using an integration by parts, together with the boundary conditions on $H_0^1([0, 1])$, then gives

$$\mathbb{W}_{0 \rightarrow 0}(dx) \propto \exp\left(\frac{1}{2} \int_0^1 x(s) \frac{d^2 x}{ds^2}(s) ds\right) dx$$

and the inverse of C is clearly identified as the differential operator above. See [6] for basic discussion of the physicists viewpoint on Wiener measure. For a given temperature parameter τ the Wiener bridge measure $\mathbb{W}_{0 \rightarrow 0}^\tau$ on $L^2([0, 1])$ is defined as the law of $\{\sqrt{\tau} W(t)\}_{t \in [0, 1]}$ where $\{W(t)\}_{t \in [0, 1]}$ is a standard Brownian bridge on $[0, 1]$ drawn from $\mathbb{W}_{0 \rightarrow 0}$.

The posterior distribution $\pi^\tau(dx)$ is defined by the change of probability formula

$$\frac{d\pi^\tau}{d\mathbb{W}_{0 \rightarrow 0}^\tau}(x) \propto e^{-\Psi(x)} \quad \text{with} \quad \Psi(x) = \frac{\lambda}{4} \int_0^1 (x(s)^2 - 1)^2 ds.$$

Notice that $\pi_0^\tau(H_0^1([0, 1])) = \pi^\tau(H_0^1([0, 1])) = 0$ since a Brownian bridge is almost surely not differentiable anywhere on $[0, 1]$. For this reason, the algorithm is implemented on $L^2([0, 1])$ even though the functional $J(\cdot)$ is defined on the Sobolev space $H_0^1([0, 1])$. In terms of Assumptions 1(1) we have $\kappa = 1$ and the measure π_0^τ is supported on \mathcal{H}^r if and only if $r < \frac{1}{2}$, see Remark 1; note also that $H_0^1([0, 1]) = \mathcal{H}^1$. Assumption 1(2) is satisfied for any choice $s \in [\frac{1}{4}, \frac{1}{2})$ because \mathcal{H}^s is embedded into $L^4([0, 1])$ for $s \geq \frac{1}{4}$. We add here that Assumptions 1(3-4) do not hold globally, but only locally on bounded sets, but the numerical results below will indicate that the theory developed in this paper is still relevant and could be extended to nonlocal versions of Assumptions 1(3-4), with considerable further work.

Following section 3.1, the pCN Markov chain at temperature $\tau > 0$ and time discretization $\delta > 0$ proposes moves from x to y according to

$$y = (1 - 2\delta)^{\frac{1}{2}} x + (2\delta\tau)^{\frac{1}{2}} \xi$$

where $\xi \in C([0, 1], \mathbb{R})$ is a standard Brownian bridge on $[0, 1]$. The move $x \rightarrow y$ is accepted with probability $\alpha^\delta(x, \xi) = 1 \wedge \exp(-\tau^{-1}[\Psi(y) - \Psi(x)])$. Figure 2 displays the convergence of the Markov chain $\{x^{k, \delta}\}_{k \geq 0}$ to a minimizer of the functional $J(\cdot)$. Note that this convergence is not shown with respect to the space $H_0^1([0, 1])$ on which J is defined, but rather in $L^2([0, 1])$; indeed $J(\cdot)$ is almost surely infinite when evaluated at samples of the pCN algorithm, precisely because $\pi_0^\tau(H_0^1([0, 1])) = 0$, as discussed above.

Of course the algorithm does not converge *exactly* to a minimizer of $J(\cdot)$, but fluctuates in a neighborhood of it. As described in the introduction of this article, in a finite dimensional setting the target probability distribution π^τ has Lebesgue density proportional to $\exp(-\tau^{-1} J(x))$. This intuitively shows that the size of the fluctuations around the minimum of the functional $J(\cdot)$ are of size proportional to $\sqrt{\tau}$. Figure 3 shows this phenomenon on log-log scales: the asymptotic mean error $\mathbb{E}[\|x - (\text{minimizer})\|_2]$ is displayed as a function of the temperature τ . Figure 4 illustrates Theorem 15. One can observe the path $\{v^\delta(t)\}_{t \in [0, T]}$ for a finite time step discretization parameter δ as well as the limiting path $\{v(t)\}_{t \in [0, T]}$ that is solution of the differential equation (58).

7 Conclusion

There are two useful perspectives on the material contained in this paper, one concerning optimization and one concerning statistics. We now detail these perspectives.

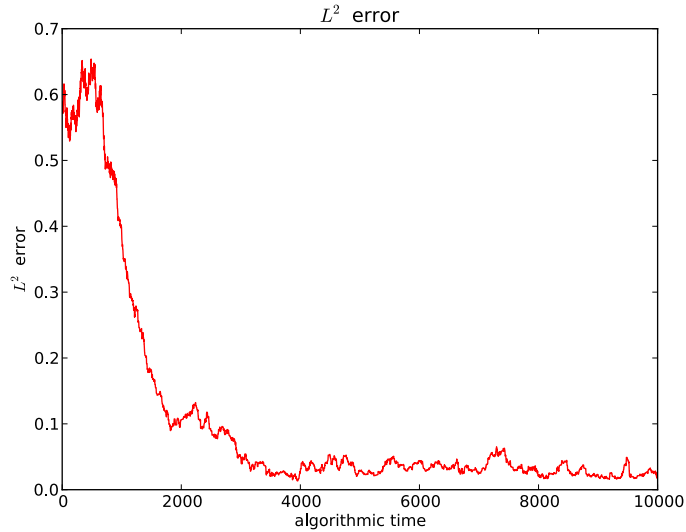


Fig. 2 pCN parameters: $\lambda = 2\pi^2$, $\delta = 1.10^{-2}$, $\tau = 1.10^{-2}$. The algorithm is started at the zero function, $x^{0,\delta}(t) = 0$ for $t \in [0, 1]$. After a transient phase, the algorithm fluctuates around a global minimizer of functional $J(\cdot)$. The L^2 error $\|x^{k,\delta} - (\text{minimizer})\|_{L^2}$ is plotted as a function of the algorithmic time k .

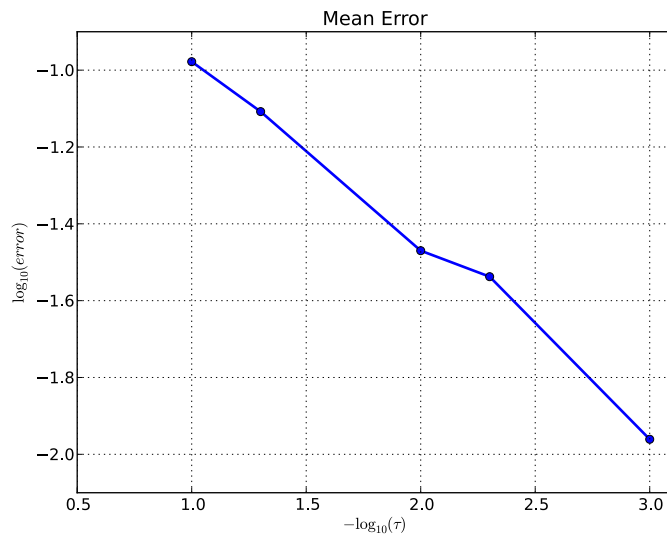


Fig. 3

Mean error $\mathbb{E}[\|x - (\text{minimizer})\|_2]$ as a function of the temperature τ .

- **Optimization** We have demonstrated a class of algorithms to minimize the functional J given by (1). The Assumptions 1 encode the intuition that the quadratic part of J dominates. Under these assumptions we study the properties of an algorithm which requires only the evaluation of Ψ and the ability to draw samples from Gaussian measures with Cameron-Martin norm given by the quadratic part of J . We demonstrate that, in a certain parameter limit, the algorithm behaves like a noisy gradient flow for the functional J and that, furthermore, the size of the noise can be controlled systematically. The advantage of constructing algorithms on Hilbert space is that they are robust to finite dimensional approximation. We turn to this point in the next bullet.

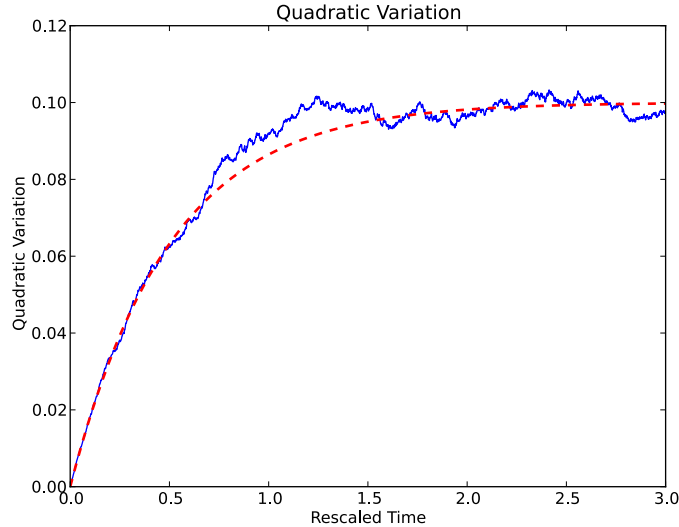


Fig. 4 pCN parameters: $\lambda = 2\pi^2$, $\tau = 1.10^{-1}$, $\delta = 1.10^{-3}$ and the algorithm starts at $x^{k,\delta} = 0$. The rescaled quadratic variation process (full line) behaves as the solution of the differential equation (dotted line), as predicted by Theorem 15. The quadratic variation converges to τ , as described by Proposition 14.

- **Statistics** The algorithm that we use is a Metropolis-Hastings method with an Ornstein-Uhlenbeck proposal which we refer to here as pCN, as in [8]. The proposal takes the form for $\xi \sim N(0, C)$,

$$y = (1 - 2\delta)^{\frac{1}{2}} x + \sqrt{2\delta\tau}\xi$$

given in (5). The proposal is constructed in such a way that the algorithm is defined on infinite dimensional Hilbert space and may be viewed as a natural analogue of a random walk Metropolis-Hastings method for measures defined via density with respect to a Gaussian. It is instructive to contrast this with the standard random walk method S-RWM with proposal

$$y = x + \sqrt{2\delta\tau}\xi.$$

Although the proposal for S-RWM differs only through a multiplicative factor in the systematic component, and thus implementation of either is practically identical, the S-RWM method is not defined on infinite dimensional Hilbert space. This turns out to matter if we compare both methods when applied in \mathbb{R}^N for $N \gg 1$, as would occur if approximating a problem in infinite dimensional Hilbert space: in this setting the S-RWM method requires the choice $\delta = \mathcal{O}(N^{-1})$ to see the diffusion (SDE) limit [23] and so requires $\mathcal{O}(N)$ steps to see $\mathcal{O}(1)$ decrease in the objective function, or to draw independent samples from the target measure; in contrast the pCN produces a diffusion limit for $\delta \rightarrow 0$ independently of N and so requires $\mathcal{O}(1)$ steps to see $\mathcal{O}(1)$ decrease in the objective function, or to draw independent samples from the target measure. Mathematically this last point is manifest in the fact that we may take the limit $N \rightarrow \infty$ (and thus work on the infinite dimensional Hilbert space) followed by the limit $\delta \rightarrow 0$.

The methods that we employ for the derivation of the diffusion (SDE) limit use a combination of ideas from numerical analysis and the weak convergence of probability measures. This approach is encapsulated in Lemma 6 which is structured in such a way that it, or variants of it, may be used to prove diffusion limits for a variety of problems other than the one considered here.

A Proofs of Lemmas; Section 4

Proof (Proof of Lemma 10) Let us introduce the two 1-Lipschitz functions $h, h_* : \mathbb{R} \rightarrow \mathbb{R}$ defined by

$$h(x) = 1 \wedge e^x \quad \text{and} \quad h_*(x) = 1 + x 1_{\{x < 0\}}. \quad (63)$$

The function h_* is a first order approximation of h in a neighborhood of zero and we have

$$\alpha^\delta(x, \xi) = h\left(-\frac{1}{\tau}\{\Psi(y) - \Psi(x)\}\right) \quad \text{and} \quad \bar{\alpha}^\delta(x, \xi) = h_*\left(-\sqrt{\frac{2\delta}{\tau}} \langle \nabla \Psi(x), \xi \rangle\right)$$

where the proposal y is a function of x and ξ , as described in Equation (16). Since $h_*(\cdot)$ is close to $h(\cdot)$ in a neighborhood of zero, the proof is finished once it is proved that $-\frac{1}{\tau}\{\Psi(y) - \Psi(x)\}$ is close to $-\sqrt{\frac{2\delta}{\tau}} \langle \nabla \Psi(x), \xi \rangle$. We have $\mathbb{E}_x[|\alpha^\delta(x, \xi) - \bar{\alpha}^\delta(x, \xi)|^p] \lesssim A_1 + A_2$ where the quantities A_1 and A_2 are given by

$$\begin{aligned} A_1 &= \mathbb{E}_x \left[\left| h\left(-\frac{1}{\tau}\{\Psi(y) - \Psi(x)\}\right) - h\left(-\sqrt{\frac{2\delta}{\tau}} \langle \nabla \Psi(x), \xi \rangle\right) \right|^p \right] \\ A_2 &= \mathbb{E}_x \left[\left| h\left(-\sqrt{\frac{2\delta}{\tau}} \langle \nabla \Psi(x), \xi \rangle\right) - h_*\left(-\sqrt{\frac{2\delta}{\tau}} \langle \nabla \Psi(x), \xi \rangle\right) \right|^p \right]. \end{aligned}$$

By Lemma 2, the first order Taylor approximation of Ψ is controlled, $|\Psi(y) - \Psi(x) - \langle \nabla \Psi(x), y - x \rangle| \lesssim \|y - x\|_s^2$. The definition of the proposal y given in Equation (16) shows that $\|(y - x) - \sqrt{2\delta\tau}\xi\|_s \lesssim \delta\|x\|_s$. Assumptions 1 state that for $z \in \mathcal{H}^s$ we have $\langle \nabla \Psi(x), z \rangle \lesssim (1 + \|x\|_s) \cdot \|z\|_s$. Since the function $h(\cdot)$ is 1-Lipschitz it follows that

$$\begin{aligned} A_1 &= \mathbb{E}_x \left[\left| h\left(-\frac{1}{\tau}\{\Psi(y) - \Psi(x)\}\right) - h\left(-\sqrt{\frac{2\delta}{\tau}} \langle \nabla \Psi(x), \xi \rangle\right) \right|^p \right] \\ &\lesssim \mathbb{E}_x \left[\left| \Psi(y) - \Psi(x) - \langle \nabla \Psi(x), y - x \rangle \right|^p + \left| \langle \nabla \Psi(x), y - x - \sqrt{2\delta\tau}\xi \rangle \right|^p \right] \\ &\lesssim \mathbb{E}_x \left[\|y - x\|_s^{2p} + (1 + \|x\|_s^p) \cdot (\delta \|x\|_s)^p \right] \lesssim \delta^p (1 + \|x\|_s^{2p}). \end{aligned} \quad (64)$$

Lemma 3 has been used to control the size of $\mathbb{E}_x[\|y - x\|^p]$. To bound A_2 , notice that for $z \in \mathbb{R}$ we have $|h(z) - h_*(z)| \leq \frac{1}{2} z^2$. Therefore the quantity A_2 can be bounded by

$$A_2 \lesssim \mathbb{E}_x \left[|\sqrt{\delta} \langle \nabla \Psi(x), \xi \rangle|^{2p} \right] \lesssim \delta^p \mathbb{E}_x \left[(1 + \|x\|_s^{2p}) \|\xi\|_s^{2p} \right] \lesssim \delta^p (1 + \|x\|_s^{2p}). \quad (65)$$

Estimates (64) and (65) together give Equation (39).

Proof (Proof of Corollary 1) Let us prove Equations (40) and (41).

- Lemma 10 and Jensen's inequality give Equation (40).
- To prove (41), one can suppose $\delta^{\frac{p}{2}} \|x\|_s^p \leq 1$. Indeed, if $\delta^{\frac{p}{2}} \|x\|_s^p \geq 1$, we have

$$\mathbb{E}_x \left[|\alpha^\delta(x, \xi) - 1|^p \right] \lesssim 1 \leq \delta^{\frac{p}{2}} \|x\|_s^p \leq \delta^{\frac{p}{2}} (1 + \|x\|_s^p),$$

which gives the result. We thus suppose from now on that $\delta^{\frac{p}{2}} \|x\|_s^p \leq 1$. Under Assumptions 1 we have $\|\nabla \Psi(x)\|_{-s} \lesssim 1 + \|x\|_s$. Lemma 2 shows that for all $x, y \in \mathcal{H}^s$ we have $|\Psi(y) - \Psi(x) - \langle \nabla \Psi(x), y - x \rangle| \lesssim \|y - x\|_s^2$. The function $h(x) = 1 \wedge e^x$ is 1-Lipschitz, $\alpha^\delta(x, \xi) = h\left(-\frac{1}{\tau}\{\Psi(y) - \Psi(x)\}\right)$ and $h(0) = 1$. Consequently,

$$\begin{aligned} \mathbb{E}_x \left[|\alpha^\delta(x, \xi) - 1|^p \right] &= \mathbb{E}_x \left[\left| h\left(-\frac{1}{\tau}\{\Psi(y) - \Psi(x)\}\right) - h(0) \right|^p \right] \\ &\lesssim \mathbb{E}_x \left[|\Psi(y) - \Psi(x)|^p \right] \lesssim \mathbb{E}_x \left[|\langle \nabla \Psi(x), y - x \rangle|^p + \|y - x\|_s^{2p} \right] \\ &\lesssim (1 + \|x\|_s^p) \cdot \mathbb{E}_x \left[\|y - x\|_s^p \right] + \mathbb{E}_x \left[\|y - x\|_s^{2p} \right]. \end{aligned}$$

By Lemma 3, for any integer $\beta \geq 1$ we have $\mathbb{E}_x[\|y - x\|_s^\beta] \lesssim \delta^\beta \|x\|_s^\beta + \delta^{\frac{\beta}{2}}$ so that the assumption $\delta^{\frac{p}{2}} \|x\|_s^p \leq 1$ leads to

$$\begin{aligned} \mathbb{E}_x \left[|\alpha^\delta(x, \xi) - 1|^p \right] &\lesssim (1 + \|x\|_s^p) \cdot (\delta^p \|x\|_s^p + \delta^{\frac{p}{2}}) + (\delta^{2p} \|x\|_s^{2p} + \delta^p) \\ &\lesssim (1 + \|x\|_s^p) \cdot (\delta^{\frac{p}{2}} + \delta^{\frac{p}{2}}) + (\delta^p + \delta^p) \\ &\lesssim \delta^{\frac{p}{2}} (1 + \|x\|_s^p). \end{aligned}$$

This finishes the proof of Corollary 1.

Proof (Proof of Lemma 12) The martingale difference $\Gamma^\delta(x, \xi)$ defined in Equation (23) can also be expressed as

$$\Gamma^\delta(x, \xi) = \xi + F(x, \xi)$$

where the error term $F(x, \xi) = F_1(x, \xi) + F_2(x, \xi)$ is given by

$$\begin{aligned} F_1(x, \xi) &= (2\tau\delta)^{-\frac{1}{2}} \left((1 - 2\delta)^{\frac{1}{2}} - 1 \right) (\gamma^\delta(x, \xi) - \mathbb{E}_x[\gamma^\delta(x, \xi)])x \\ F_2(x, \xi) &= (\gamma^\delta(x, \xi) - 1) \cdot \xi - \mathbb{E}_x[\gamma^\delta(x, \xi) \cdot \xi]. \end{aligned}$$

We now prove that the quantity $F(x, \xi)$ satisfies

$$\mathbb{E}_x \left[\|F(x, \xi)\|_s^2 \right] \lesssim \delta^{\frac{1}{4}} (1 + \|x\|_s^2) \quad (66)$$

- We have $\delta^{-\frac{1}{2}}((1-2\delta)^{\frac{1}{2}}-1) \lesssim \delta^{\frac{1}{2}}$ and $|\gamma^\delta(x, \xi)| \leq 1$. Consequently,

$$\mathbb{E}_x \left[\|F_1(x, \xi)\|_s^2 \right] \lesssim \delta \|x\|_s^2 \quad (67)$$

- Let us now prove that F_2 satisfies

$$\mathbb{E}_x \left[\|F_2(x, \xi)\|_s^2 \right] \lesssim \delta^{\frac{1}{4}} (1 + \|x\|_s^{\frac{1}{2}}). \quad (68)$$

To this end, use the decomposition

$$\begin{aligned} \mathbb{E}_x \left[\|F_2(x, \xi)\|_s^2 \right] &\lesssim \mathbb{E}_x \left[|\gamma^\delta(x, \xi) - 1|^2 \cdot \|\xi\|_s^2 \right] + \|\mathbb{E}_x [\gamma^\delta(x, \xi) \cdot \xi]\|_s^2 \\ &= I_1 + I_2. \end{aligned}$$

The Cauchy-Schwarz inequality shows that $I_1 \lesssim \mathbb{E}_x \left[|\gamma^\delta(x, \xi) - 1|^4 \right]^{\frac{1}{2}}$ where the Bernoulli random variable $\gamma^\delta(x, \xi)$ can be expressed as $\gamma^\delta(x, \xi) = \mathbb{1}_{\{U < \alpha^\delta(x, \xi)\}}$ where $U \stackrel{\mathcal{D}}{\sim} \text{Uniform}(0, 1)$ is independent from any other source of randomness. Consequently

$$\mathbb{E}_x \left[|\gamma^\delta(x, \xi) - 1|^4 \right] = \mathbb{E}_x \left[\mathbb{1}_{\{\gamma^\delta(x, \xi) = 0\}} \right] = 1 - \alpha^\delta(x)$$

where the mean local acceptance probability $\alpha^\delta(x)$ is defined by $\alpha^\delta(x) = \mathbb{E}_x[\alpha^\delta(x, \xi)] \in [0, 1]$. The convexity of the function $x \rightarrow |1 - x|$ ensures that

$$|1 - \alpha^\delta(x)| = |1 - \mathbb{E}_x[\alpha^\delta(x, \xi)]| \leq \mathbb{E}_x[|1 - \alpha^\delta(x, \xi)|] \lesssim \delta^{\frac{1}{2}} (1 + \|x\|_s)$$

where the last inequality follows from Corollary 1. This proves that $I_1 \lesssim \delta^{\frac{1}{4}} (1 + \|x\|_s^{\frac{1}{2}})$. To bound I_2 , it suffices to notice

$$\begin{aligned} I_2 &= \|\mathbb{E}_x[\gamma^\delta(x, \xi) \cdot \xi]\|_s^2 = \|\mathbb{E}_x[(\gamma^\delta(x, \xi) - 1) \cdot \xi]\|_s^2 \\ &\lesssim \mathbb{E}_x \left[|\gamma^\delta(x, \xi) - 1|^2 \cdot \|\xi\|_s^2 \right] = I_1 \end{aligned}$$

so that $I_2 \lesssim I_1 \lesssim \delta^{\frac{1}{4}} (1 + \|x\|_s^{\frac{1}{2}})$ and $\mathbb{E}_x \left[\|F_2(x, \xi)\|_s^2 \right] \lesssim \delta^{\frac{1}{4}} (1 + \|x\|_s^{\frac{1}{2}})$.

Combining Equation (67) and (68) gives Equation (66).

Let us now describe how Equations (??) and (??) follow from the estimate (66).

- We have $\mathbb{E}[\langle \hat{\varphi}_i, \xi \rangle_s \langle \hat{\varphi}_j, \xi \rangle_s] = \langle \hat{\varphi}_i, C_s \hat{\varphi}_j \rangle_s$ and $\mathbb{E}_x[\langle \hat{\varphi}_i, \Gamma^\delta(x, \xi) \rangle_s \langle \hat{\varphi}_j, \Gamma^\delta(x, \xi) \rangle_s] = \langle \hat{\varphi}_i, D^\delta(x) \hat{\varphi}_j \rangle_s$ with $\Gamma^\delta(x, \xi) = \xi + F(x, \xi)$. Consequently,

$$\begin{aligned} \langle \hat{\varphi}_i, D^\delta(x) \hat{\varphi}_j \rangle_s - \langle \hat{\varphi}_i, C_s \hat{\varphi}_j \rangle_s &= \mathbb{E}_x[\langle \hat{\varphi}_i, F(x, \xi) \rangle_s \langle \hat{\varphi}_j, F(x, \xi) \rangle_s] \\ &\quad + \mathbb{E}_x[\langle \hat{\varphi}_i, \xi \rangle_s \langle \hat{\varphi}_j, F(x, \xi) \rangle_s] \\ &\quad + \mathbb{E}_x[\langle \hat{\varphi}_i, F(x, \xi) \rangle_s \langle \hat{\varphi}_j, \xi \rangle_s]. \end{aligned}$$

We have $|\langle \hat{\varphi}_i, F(x, \xi) \rangle_s| \leq \|F(x, \xi)\|_s$ and the Cauchy-Schwarz inequality proves that

$$\begin{aligned} \mathbb{E}_x[\langle \hat{\varphi}_i, F(x, \xi) \rangle_s \langle \hat{\varphi}_j, \xi \rangle_s]^2 &\leq \mathbb{E}_x[\|F(x, \xi)\|_s \|\xi\|_s]^2 \\ &\lesssim \mathbb{E}_x[\|F(x, \xi)\|_s^2]. \end{aligned}$$

It thus follows from Equation (66) that

$$\begin{aligned} |\langle \hat{\varphi}_i, D^\delta(x) \hat{\varphi}_j \rangle_s - \langle \hat{\varphi}_i, C_s \hat{\varphi}_j \rangle_s| &\lesssim \mathbb{E}_x[\|F(x, \xi)\|_s^2] + \mathbb{E}_x[\|F(x, \xi)\|_s^2]^{\frac{1}{2}} \\ &\lesssim \delta^{\frac{1}{8}} (1 + \|x\|_s), \end{aligned}$$

finishing the proof of (46).

- We have $\text{Trace}_{\mathcal{H}^s}(C_s) = \mathbb{E}[\|\xi\|_s^2]$ and $\text{Trace}_{\mathcal{H}^s}(D^\delta(x)) = \mathbb{E}[\|\Gamma^\delta(x, \xi)\|_s^2]$. Estimate (66) thus shows that

$$\begin{aligned} |\text{Trace}_{\mathcal{H}^s}(D^\delta(x)) - \text{Trace}_{\mathcal{H}^s}(C_s)| &= |\mathbb{E}[\|\Gamma^\delta(x, \xi)\|_s^2 - \|\xi\|_s^2]| \\ &= |\mathbb{E}[\|\xi + F(x, \xi)\|_s^2 - \|\xi\|_s^2]| \\ &\lesssim |\mathbb{E}[\langle 2\xi + F(x, \xi), F(x, \xi) \rangle_s]| \lesssim \mathbb{E}[\|2\xi + F(x, \xi)\|_s \|F(x, \xi)\|_s] \\ &\lesssim \mathbb{E}[4\|\xi\|_s^2 + \|F(x, \xi)\|_s^2]^{\frac{1}{2}} \cdot \mathbb{E}[\|F(x, \xi)\|_s^2]^{\frac{1}{2}} \\ &\lesssim \left(1 + \delta^{\frac{1}{4}} (1 + \|x\|_s^{\frac{1}{2}})\right)^{\frac{1}{2}} \cdot \left(\delta^{\frac{1}{8}} (1 + \|x\|_s)\right) \lesssim \delta^{\frac{1}{8}} (1 + \|x\|_s^{\frac{1}{2}}), \end{aligned}$$

finishing the proof of (??).

B Proof of Lemma 16

Before proceeding to give the proof, let us give a brief proof sketch. The proof of Lemma 16 consists in showing first that for any $\varepsilon > 0$ one can find a ball of radius $R(\varepsilon)$ around 0 in \mathcal{H}^s ,

$$B_0(R(\varepsilon)) = \{x \in \mathcal{H}_s : \|x\|_s \leq R(\varepsilon)\},$$

such that with probability $1 - 2\varepsilon$ we have $x^{k,\delta} \in B_0(R(\varepsilon))$ and $y^{k,\delta} \in B_0(R(\varepsilon))$ for all $0 \leq k \leq T\delta^{-1}$. As is described below, the existence of such a ball follows from the bound

$$\mathbb{E}[\sup_{t \in [0, T]} \|x(t)\|_s] < +\infty \quad (69)$$

where $t \mapsto x(t)$ is the solution of the stochastic differential equation (21). For the sake of completeness, we include a proof of Equation (69). The solution $t \mapsto x(t)$ of the stochastic differential equation (21) satisfies $x(t) = \int_0^t d(x(u)) du + \sqrt{2\tau} W(t)$ for all $t \in [0, T]$ where the drift function $d(x) = -(x + C\nabla\Psi(x))$ is globally Lipschitz on \mathcal{H}^s , as described in Lemma 2. Consequently $\|d(x)\|_s \leq A(1 + \|x\|_s)$ for some positive constant $A > 0$. The triangle inequality then shows that

$$\|x(t)\|_s \leq A \int_0^t (1 + \|x(u)\|_s) du + \sqrt{2\tau} \|W(t)\|_s.$$

By Gronwall's inequality we obtain

$$\sup_{[0, T]} \|x(t)\|_s \leq (AT + \sup_{[0, T]} \|W(t)\|_s) [1 + ATe^{AT}].$$

Since $\mathbb{E}[\sup_{[0, T]} \|W(t)\|_s] < \infty$, the bound (69) is proved.

Proof (Proof of Lemma 16) The proof consists in showing that the acceptance probability of the algorithm is sufficiently close to 1 so that approximation $t^\delta(k) \approx k$ holds. The argument can be divided into 3 main steps. In the first part, we show that we can find a finite ball $B(0, R(\varepsilon))$ such that the trajectory of the Markov chain $\{x^{k,\delta}\}_{k \leq T\delta^{-1}}$ remains in this ball with probability at least $1 - 2\varepsilon$. This observation is useful since the function Ψ is Lipschitz on any ball of finite radius in \mathcal{H}^s . In the second part, using the fact that Ψ is Lipschitz on $B(0, R(\varepsilon))$, we find a lower bound for the acceptance probability α^δ . Then, in the last step, we use a moment estimate to prove that one can make the lower bound uniform on the interval $0 \leq k \leq T\delta^{-1}$.

– Restriction to a Ball of Finite Radius

First, we show that with high probability the trajectory of the MCMC algorithm stays in a ball of finite radius. The functional $x \mapsto \sup_{t \in [0, T]} \|x(t)\|_s$ is continuous on $C([0, T], \mathcal{H}_s)$ and $\mathbb{E}[\sup_{t \in [0, T]} \|x(t)\|_s] < \infty$ for $t \mapsto x(t)$ following the stochastic differential equation (21), as proved in Equation (69). Consequently, the weak convergence of z^δ to the solution of (21) encapsulated in Theorem 4 shows that $\mathbb{E}[\sup_{k < T\delta^{-1}} \|x^{k,\delta}\|_s]$ can be bounded by a finite universal constant independent from δ . Given $\varepsilon > 0$, Markov inequality thus shows that one can find a radius $R_1 = R_1(\varepsilon)$ large enough so that the inequality

$$\mathbb{P}[\|x^{k,\delta}\|_s < R_1 \quad \text{for all } 0 \leq k \leq T\delta^{-1}] > 1 - \varepsilon \quad (70)$$

for any $\delta \in (0, \frac{1}{2})$. By Fernique's Theorem there exists $\alpha > 0$ such that $\mathbb{E}[e^{\alpha\|\xi\|_s^2}] < \infty$. This implies that $\mathbb{P}[\|\xi\|_s > r] \lesssim e^{-\alpha r^2}$. Therefore, if $\{\xi_k\}_{k \geq 0}$ are i.i.d. Gaussian random variables distributed as $\xi \stackrel{\mathcal{D}}{\sim} \pi_0$, the union bound shows that

$$\mathbb{P}[\|\sqrt{\delta}\xi_k\|_s \leq r \quad \text{for all } 0 \leq k \leq T\delta^{-1}] \gtrsim 1 - T\delta^{-1} \exp(-\alpha\delta^{-1}r^2).$$

This proves that one can choose $R_2 = R_2(\varepsilon)$ large enough in such a manner that

$$\mathbb{P}[\|\sqrt{\delta}\xi_k\|_s < R_2 \quad \text{for all } 0 \leq k \leq T\delta^{-1}] > 1 - \varepsilon \quad (71)$$

for any $\delta \in (0, \frac{1}{2})$. At temperature $\tau > 0$ the MCMC proposals are given by $y^{k,\delta} = (1 - 2\delta)^{\frac{1}{2}} x^{k,\delta} + (2\delta\tau)^{\frac{1}{2}} \xi_k$. It thus follows from the bounds (70) and (71) that with probability at least $(1 - 2\varepsilon)$ the vectors $x^{k,\delta}$ and $y^{k,\delta}$ belong to the ball $B_0(R(\varepsilon)) = \{x \in \mathcal{H}_s : \|x\|_s < R(\varepsilon)\}$ for $0 \leq k \leq T\delta^{-1}$ where radius $R(\varepsilon)$ is given by $R(\varepsilon) = R_1(\varepsilon) + R_2(\varepsilon)$.

– Lower Bound for Acceptance Probability

We now give a lower bound for the acceptance probability $\alpha^\delta(x^{k,\delta}, \xi^k)$ that the move $x^{k,\delta} \rightarrow y^{k,\delta}$ is accepted. Assumptions 1 state that $\|\nabla\Psi(x)\|_{-s} \lesssim 1 + \|x\|_s$. Therefore, the function $\Psi : \mathcal{H}^s \rightarrow \mathbb{R}$ is Lipschitz on $B_0(R(\varepsilon))$,

$$\|\Psi\|_{\text{lip}, \varepsilon} \stackrel{\text{def}}{=} \sup \left\{ \frac{|\Psi(y) - \Psi(x)|}{\|y - x\|_s} : x, y \in B_0(R(\varepsilon)) \right\} < \infty.$$

One can thus bound the acceptance probability $\alpha^\delta(x^{k,\delta}, \xi^k) = 1 \wedge \exp(-\tau^{-1}[\Psi(y^{k,\delta}) - \Psi(x^{k,\delta})])$ for $x^{k,\delta}, y^{k,\delta} \in B_0(R(\varepsilon))$.

Since the function $z \mapsto 1 \wedge e^{-\tau^{-1}z}$ is Lipschitz with constant τ^{-1} , the definition of $\|\Psi\|_{\text{lip}, \varepsilon}$ shows that the bound

$$1 - \alpha^\delta(x^{k,\delta}, \xi^k) \leq \tau^{-1} \|\Psi\|_{\text{lip}, \varepsilon} \|y^{k,\delta} - x^{k,\delta}\|_s$$

$$\begin{aligned} &\leq \tau^{-1} \|\Psi\|_{\text{lip},\varepsilon} \left\{ [(1-2\delta)^{\frac{1}{2}} - 1] \|x^{k,\delta}\|_s + (2\delta\tau)^{\frac{1}{2}} \|\xi^k\| \right\} \\ &\lesssim \sqrt{\delta} (1 + \|\xi^k\|_s) \end{aligned}$$

holds for every $x^{k,\delta}, y^{k,\delta} \in B_0(R(\varepsilon))$. Hence, there exists a constant $K = K(\varepsilon)$ such that $\widehat{\alpha}^\delta(\xi^k) = 1 - K\sqrt{\delta}(1 + \|\xi^k\|_s)$ satisfies $\alpha^\delta(x^{k,\delta}, \xi^k) > \widehat{\alpha}^\delta(\xi^k)$ for every $x^{k,\delta}, y^{k,\delta} \in B_0(R(\varepsilon))$. Since the trajectory of the MCMC algorithm stays in the ball $B_0(R(\varepsilon))$ with probability at least $1 - 2\varepsilon$ the inequality

$$\mathbb{P}[\alpha^\delta(x^{k,\delta}, \xi^k) > \widehat{\alpha}^\delta(\xi^k) \quad \text{for all } 0 \leq k \leq T\delta^{-1}] > 1 - 2\varepsilon.$$

holds for every $\delta \in (0, \frac{1}{2})$.

– Second Moment Method

To prove that $t^\delta(k)$ does not deviate too much from k , we show that its expectation satisfies $\mathbb{E}[t^\delta(k)] \approx k$ and we then control the error by bounding the variance. Since the Bernoulli random variable $\gamma^{k,\delta} = \text{Bernoulli}(\alpha^\delta(x^{k,\delta}, \xi^k))$ are not independent, the variance of $t^\delta(k) = \sum_{l \leq k} \gamma^{l,\delta}$ is not easily computable. We thus introduce i.i.d. auxiliary random variables $\widehat{\gamma}^{k,\delta}$ such that

$$\sum_{l \leq k} \widehat{\gamma}^{l,\delta} = \widehat{t}^\delta(k) \quad \approx \quad t^\delta(k) = \sum_{l \leq k} \gamma^{l,\delta}.$$

As described below, the behaviour of $\widehat{t}^\delta(k)$ is readily controlled since it is a sum of i.i.d. random variables. The proof then exploits the fact that $\widehat{t}^\delta(k)$ is a good approximation of $t^\delta(k)$.

The Bernoulli random variables $\gamma^{k,\delta}$ can be described as $\gamma^{k,\delta} = \mathbf{1}(U_k < \alpha^\delta(x^{k,\delta}, \xi^k))$ where $\{U_k\}_{k \geq 0}$ are i.i.d. random variables uniformly distributed on $(0, 1)$. As a consequence, with probability at least $1 - 2\varepsilon$, the random variables $\widehat{\gamma}^{k,\delta} = \mathbf{1}(U_k < \widehat{\alpha}^\delta)$ satisfy $\gamma^{k,\delta} \geq \widehat{\gamma}^{k,\delta}$ for all $0 \leq k \leq T\delta^{-1}$. Therefore, with probability at least $1 - 2\varepsilon$, we have $t^\delta(k) \geq \widehat{t}^\delta(k)$ for all $0 \leq k \leq T\delta^{-1}$ where $\widehat{t}^\delta(k) = \sum_{l \leq k} \widehat{\gamma}^{l,\delta}$. Consequently, since $t^\delta(k) \leq k$, to prove Lemma 16 it suffices to show instead that the following limit in probability holds,

$$\lim_{\delta \rightarrow 0} \sup \{ \delta \cdot |\widehat{t}^\delta(k) - k| : 0 \leq k \leq T\delta^{-1} \} = 0. \quad (72)$$

Contrary to the random variables $\{\gamma^{k,\delta}\}_{k \geq 0}$, the random variables $\{\widehat{\gamma}^{k,\delta}\}_{k \geq 0}$ are i.i.d. and are thus easily controlled. By Doob's inequality we have

$$\mathbb{P} \left[\sup \{ \delta \cdot |\widehat{t}^\delta(k) - \mathbb{E}[\widehat{t}^\delta(k)]| : 0 \leq k \leq T\delta^{-1} \} > \eta \right] \leq 2 \frac{\text{Var}(\widehat{t}^\delta(T\delta^{-1}))}{(\delta^{-1}\eta)^2} \leq 2 \frac{\delta T}{\eta^2}.$$

Since $\mathbb{E}[\widehat{t}^\delta(k)] = k \cdot \{1 - K\sqrt{\delta}(1 + \mathbb{E}[\|\xi^k\|_s])\}$, Equation (72) follows. This finishes the proof of Lemma 16.

Acknowledgements The authors thank an anonymous referee for constructive comments. We are grateful to David Dunson for his comments on the implications of theory, Frank Pinski for helpful discussions concerning the behaviour of the quadratic variation; these discussions crystallized the need to prove Theorem 15. NSP gratefully acknowledges the NSF grant DMS 1107070. AMS is grateful to EPSRC and ERC for financial support. Parts of this work was done when AHT was visiting the department of Statistics at Harvard university. The authors thank the department of statistics, Harvard University for its hospitality.

References

1. Berger, E.: Asymptotic behaviour of a class of stochastic approximation procedures. *Probab. Theory Relat. Fields* **71**(4), 517–552 (1986)
2. Bertsimas, D., Tsitsiklis, J.: Simulated annealing. *Statistical Science* **8**(1), 10–15 (1993)
3. Beskos, A., Roberts, G., Stuart, A., Voss, J.: An MCMC method for diffusion bridges. *Stochastics and Dynamics* **8**(3), 319–350 (2008)
4. Černý, V.: Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm. *Journal of Optimization Theory and Applications* **45**(1), 41–51 (1985)
5. Chiang, T., Hwang, C., Sheu, S.: Diffusion for global optimization in \mathbb{R}^n . *SIAM Journal on Control and Optimization* **25**(3), 737–753 (1987)
6. Chorin, A., Hald, O.: *Stochastic Tools in Mathematics and Science*. Springer Verlag (2006)
7. Cotter, S., Dashti, M., Stuart, A.: Variational data assimilation using targetted random walks. *International Journal for Numerical Methods in Fluids* (2011). DOI 10.1002/fld.2510
8. Cotter, S., G.O.Roberts, Stuart, A., White, D.: MCMC methods for functions: modifying old algorithms to make them faster. *Statistical Science*, To Appear
9. Da Prato, G., Zabczyk, J.: *Stochastic Equations in Infinite Dimensions*, *Encyclopedia of Mathematics and its Applications*, vol. 44. Cambridge University Press, Cambridge (1992)
10. Da Prato, G., Zabczyk, J.: *Ergodicity for Infinite Dimensional Systems*. Cambridge Univ Pr (1996)

11. Dashti, M., Law, K., Stuart, A., Voss, J.: MAP estimators and their consistency in bayesian nonparametric inverse problems. *Inverse Problems*, To Appear
12. Ethier, S.N., Kurtz, T.G.: *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York (1986). Characterization and convergence
13. Geman, D.: Bayesian image analysis by adaptive annealing. *IEEE Transactions on Geoscience and Remote Sensing* **1**, 269–276 (1985)
14. Geman, S., Hwang, C.: Diffusions for global optimization. *SIAM Journal on Control and Optimization* **24**, 1031 (1986)
15. Hairer, M., A.M.Stuart, Voss, J., Wiberg, P.: Analysis of SPDEs arising in path sampling. Part 1: the gaussian case. *Comm. Math. Sci.* **3**, 587–603 (2005)
16. Hairer, M., Stuart, A.M., Voss, J.: Analysis of SPDEs arising in path sampling. PartII: the nonlinear case. *Ann. Appl. Probab.* **17**(5-6), 1657–1706 (2007)
17. Hairer, M., Stuart, A.M., Voss, J.: Signal processing problems on function space: Bayesian formulation, stochastic pdes and effective mcmc methods. *The Oxford Handbook of Nonlinear Filtering*, Editors D. Crisan and B. Rozovsky (2010). To Appear
18. Henry, D.: *Geometric Theory of Semilinear Parabolic Equations*, vol. 61. Springer-Verlag (1981)
19. Hinze, M., Pinnau, R., Ulbrich, M., Ulbrich, S.: *Optimization with PDE constraints*. Springer Verlag (2008)
20. Holley, R., Kusuoka, S., Stroock, D.: Asymptotics of the spectral gap with applications to the theory of simulated annealing. *Journal of Functional Analysis* **83**(2), 333–347 (1989)
21. Kirkpatrick, S., Jr., D., Vecchi, M.: Optimization by simulated annealing. *Science* **220**(4598), 671–680 (1983)
22. Kupferman, R., Stuart, A., Terry, J., Tupper, P.: Long-term behaviour of large mechanical systems with random initial data. *Stochastics and Dynamics* **2**(04), 533–562 (2002)
23. Mattingly, J., Pillai, N., Stuart, A.: SPDE Limits of the Random Walk Metropolis Algorithm in High Dimensions. *Ann. Appl. Probab.* (2011)
24. Pillai, N.S., Stuart, A.M., Thiery, A.H.: Optimal scaling and diffusion limits for the langevin algorithm in high dimensions. *Annals of Applied Probability* **22**(6), 2320–2356 (2012). DOI 10.1214/11-AAP828
25. Stroock, D.W., Varadhan, S.S.: *Multidimensional diffusion processes*, vol. 233. Springer (1979)
26. Stuart, A.: Inverse problems: a Bayesian perspective. *Acta Numerica* **19**(-1), 451–559 (2010)
27. Tierney, L.: A note on Metropolis-Hastings kernels for general state spaces. *Ann. Appl. Probab.* **8**(1), 1–9 (1998)