arXiv: math.PR/1103.0542

OPTIMAL SCALING AND DIFFUSION LIMITS FOR THE LANGEVIN ALGORITHM IN HIGH DIMENSIONS

By Natesh S. Pillai [*]

Department of Statistics,
Harvard University

By Andrew M. Stuart [†]

Mathematics Institute,
Warwick University

By Alexandre H. Thiéry [‡]

Department of Statistics,
Warwick University

The Metropolis-adjusted Langevin (MALA) algorithm is a sampling algorithm which makes local moves by incorporating information about the gradient of the logarithm of the target density. In this paper we study the efficiency of MALA on a natural class of target measures supported on an infinite dimensional Hilbert space. These natural measures have density with respect to a Gaussian random field measure and arise in many applications such as Bayesian nonparametric statistics and the theory of conditioned diffusions. We prove that, started in stationarity, a suitably interpolated and scaled version of the Markov chain corresponding to MALA converges to an infinite dimensional diffusion process. Our results imply that, in stationarity, the MALA algorithm applied to an $N-$dimensional approximation of the target will take $\mathcal{O}(N^{\frac{1}{3}})$ steps to explore the invariant measure, comparing favourably with Random Walk Metropolis which was recently shown to require $\mathcal{O}(N)$ steps when applied to the same class of problems. As a by-product of the diffusion limit it also follows that the MALA algorithm is optimized at an average acceptance probability of 0.574. Previous results were proved only for targets which are products of one dimensional distributions, or for variants of this situation, limiting their applicability. The correlation in our target means that the rescaled MALA algorithm converges weakly to an infinite dimensional Hilbert space valued diffusion, and the limit cannot be described through analysis of scalar diffusions. The limit theorem is proved by showing that a drift-martingale decomposition of the Markov chain, suitably scaled, closely resembles a weak Euler-Maruyama discretization of the putative limit. An invariance principle is proved for the Martingale and a continuous mapping argument is used to complete the proof.

[*]Research supported by the NSF grant DMS 1107070.

[†]Research supported by EPSRC and ERC.

[‡]Research supported by (EPSRC-funded) CRISM.

AMS 2000 subject classifications: Primary 60J20 ; secondary 65C05, Markov Chain Monte Carlo, Metropolis Adjusted Langevin Algorithm, Scaling limit, Diffusion Approximation

1. Introduction.

Sampling probability distributions π^N in $\mathbb{R}^N$ for N large is of interest in numerous applications arising in applied probability and statistics. The Markov chain-Monte Carlo (MCMC) methodology [RC04] provides a framework for many algorithms which effect this sampling. It is hence of interest to quantify the computational cost of MCMC methods as a function of dimension N. This paper is part of a research program designed to develop the analysis of MCMC in high dimensions so that it may be usefully applied to understand target measures which arise in applications. The simplest class of target measures for which analysis can be carried out are perhaps target distributions π^N of the form

$$\frac{d\pi^N}{d\lambda^N}(x) = \Pi_{i=1}^N f(x_i). \tag{1.1}$$

Here $\lambda^N(dx)$ is the N-dimensional Lebesgue measure and $f(x)$ is a one-dimensional probability density function. Thus π^N has the form of an i.i.d. product. Using understanding gained in this situation, we will develop an analysis that is relevant to an important class of non-product measures which arise in a range of applications.

We start by describing the MCMC methods which are studied in this paper. Consider a π^N-invariant Metropolis Hastings Markov chain $\{x^{k,N}\}_{k\geq 1}$. From the current state x we propose y drawn from the kernel $q(x,y)$; this is then accepted with probability

$$\alpha(x,y) = 1 \wedge \frac{\pi^N(y)q(y,x)}{\pi^N(x)q(x,y)}.$$

Two widely used proposals are the random walk proposal (obtained from the discrete approximation of Brownian motion)

$$y = x + \sqrt{2\delta}Z^N, \quad Z^N \sim \mathrm{N}(0, \mathrm{I}_N), \tag{1.2}$$

and the Langevin proposal (obtained from the time discretization of the Langevin diffusion)

$$y = x + \delta\,\nabla \log \pi^N(x) + \sqrt{2\delta}\,Z^N, \qquad Z^N \sim \mathrm{N}(0, \mathrm{I}_N)\ . \tag{1.3}$$

Here 2δ is the proposal variance, a parameter quantifying the size of the discrete time increment; we will consider "local proposals" for which δ is small. The Markov chain corresponding to the proposal (1.2) is the Random Walk Metropolis (RWM) algorithm [MRTT53], and the Markov transition rule constructed from the proposal (1.3) is known as the Metropolis Adjusted Langevin Algorithm (MALA) [RC04]. This paper is aimed at analyzing the computational complexity of the MALA algorithm in high dimensions.

A fruitful way to quantify the computational cost of these Markov chains which proceed via local proposals is to determine the "optimal" size of increment δ as a function of dimension N (the precise notion of optimality is discussed below). A simple heuristic suggests the existence of such an "optimal scale" for δ: smaller values of the proposal variance lead to high acceptance rates but the chain does not move much even when accepted, and therefore may not be efficient. Larger values of the proposal variance lead to larger moves, but then the acceptance probability is tiny. The optimal scale for the proposal variance strikes a balance between making large moves and still having a reasonable acceptance probability. In order to quantify this idea it is useful to define a continuous interpolant of the Markov chain as follows:

$$z^N(t) = \Big(\frac{t}{\Delta t} - k\Big)\, x^{k+1,N} + \Big(k+1 - \frac{t}{\Delta t}\Big)\, x^{k,N}, \qquad \text{for} \qquad k\Delta t \leq t < (k+1)\Delta t. \tag{1.4}$$

We choose the proposal variance to satisfy $\delta = \ell \Delta t$, with $\Delta t = N^{-\gamma}$ setting the scale in terms of dimension and the parameter ℓ a "tuning" parameter which is indepedent of the dimension N. Key questions, then, concern the choice of γ and ℓ. If z^N converges weakly to a suitable stationary diffusion process then it is natural to deduce that the number of Markov chain steps required in stationarity is inversely proportional to the proposal variance, and hence to Δt, and so grows like N^γ. The parametric dependence of the limiting diffusion process then provides a selection mechanism for ℓ. A research program along these lines was initiated by Roberts and coworkers in the pair of papers [RGG97, RR98]. These papers concerned the RWM and MALA algorithms respectively when appplied to the target (1.1). In both cases it was shown that the projection of z^N into any single fixed coordinate direction x_i converges weakly in $C([0,T];\mathbb{R})$ to z, the scalar diffusion process

$$\frac{dz}{dt} = h(\ell)[\log f(z)]' + \sqrt{2h(\ell)}\frac{dW}{dt} \tag{1.5}$$

for $h(\ell) > 0$ a constant determined by the parameter ℓ from the proposal variance. For RWM the scaling of the proposal variance to achieve this limit is determined by the choice $\gamma = 1$ ([RGG97]) whilst for MALA $\gamma = \frac{1}{3}$ ([RR98]). The analysis shows that the number of steps required to sample the target measure grows as $\mathcal{O}(N)$ for RWM, but only as $\mathcal{O}(N^{\frac{1}{3}})$ for MALA. This quantifies the efficiency gained by use of MALA over RWM, and in particular from employing local moves informed by the gradient of the logarithm of the target density. A second important feature of the analysis is that it suggests that the optimal choice of ℓ is that which maximizes $h(\ell)$. This value of ℓ leads in both cases to a universal (independent of $f(\cdot)$) optimal average acceptance probability (to three significant figures) of 0.234 for RWM and 0.574 for MALA.

These theoretical analyses have had a huge practical impact as the optimal acceptance probabilities send a concrete message to practitioners: one should "tune" the proposal variance of the RWM and MALA algorithms so as to have acceptance probabilities of 0.234 and 0.574 respectively. However, practitioners use these tuning criteria far outside the class of targer distrubutions given by (1.1). It is natural to ask whether they are wise to do so. Extensive simulations (see [RR01, SFR10]) show that these optimality results also hold for more complex target distributions. Furthermore, a range of subsequent theoretical analyses confirmed that the optimal scaling ideas do indeed extend beyond (1.1); these papers studied slightly more complicated models such as products of one dimensional distributions with different variances and elliptically symmetric distributions ([Béd07, Béd09, BPS04, CRR05]). However the diffusion limits obtained remained essentially one dimensional in all of these extensions.[1] In this paper we study considerably more complex target distributions which are not of the product form and the limiting diffusion takes values in an infinite dimensional space.

Our perspective on these problems is motivated by applications such as Bayesian nonparametric statistics, for example in application to inverse problems [Stu10], and the theory of conditioned diffusions [HSV11]. In both these areas the target measure of interest, π, is on an infinite dimensional real separable Hilbert space $\mathcal{H}$ and, for Gaussian priors (inverse problems) or additive noise (diffusions) is absolutely continuous with respect to a Gaussian measure π_0 on $\mathcal{H}$ with mean zero and covariance operator $\mathcal{C}$. This framework for the analysis of MCMC in high dimensions was first studied in the papers [BRSV08, BRS09, BS09]. The Radon-Nikodym derivative defining the target

[1]The paper [BR00] contains an infinite dimensional diffusion limit, but we have been unable to employ the techniques of that paper.

measure is assumed to have the form

$$\frac{d\pi}{d\pi_0}(x) = M_\Psi \exp(-\Psi(x)) \tag{1.6}$$

for a real-valued functional $\Psi : \mathcal{H}^s \mapsto \mathbb{R}$ defined on a subspace $\mathcal{H}^s \subset \mathcal{H}$ that contains the support of the reference measure π_0; here M_Ψ is a normalizing constant. We are interested in studying MCMC methods applied to finite dimensional approximations of this measure found by projecting onto the first N eigenfunctions of the covariance operator $\mathcal{C}$ of the Gaussian reference measure π_0.

It is proved in [DPZ92, HAVW05, HSV07] that the measure π is invariant for $\mathcal{H}-$valued SDEs (or stochastic PDEs – SPDEs) with the form

$$\frac{dz}{dt} = -h(\ell)\big(z + \mathcal{C}\nabla\Psi(z)\big) + \sqrt{2\,h(\ell)}\,\frac{dW}{dt}, \quad z(0) = z^0 \tag{1.7}$$

where W is a Brownian motion (see [DPZ92]) in $\mathcal{H}$ with covariance operator $\mathcal{C}$. In [MPS11] the RWM algorithm is studied when applied to a sequence of finite dimensional approximations of π as in (1.6). The continuous time interpolant of the Markov chain z^N given by (1.4) is shown to converge weakly to z solving (1.7) in $C([0,T];\mathcal{H}^s)$. Furthermore, as for the i.i.d target measure the scaling of the proposal variance which achieves this scaling limit is inversely proportional to N (*i.e.* corresponds to the exponent $\gamma = 1$) and the speed of the limiting diffusion process is maximized at the same universal acceptance probability of 0.234 that was found in the i.i.d case. Thus, remarkably, the i.i.d. case has been of fundamental importance in understanding MCMC methods applied to complex infinite dimensional probability measures arising in practice. The paper [MPS11] developed an approach for deriving diffusion limits for such algorithms, using ideas from numerical analysis. These techniques can be built on to derive scaling limits for a wide range of Metropolis-Hastings algorithms with local proposals.

The purpose of this article is to develop the techniques in the context of the MALA algorithm. To the best of our knowledge, the only paper to consider the optimal scaling for the MALA algorithm for non-product targets is [BPS04], in the context of non-linear regression. In [BPS04] the target measure has a structure similar to that of the mean field models studied in statistical mechanics and hence behaves asymptotically like a product measure when the dimension goes to infinity. Thus the diffusion limit obtained in [BPS04] is finite dimensional.

The main contribution of our work is the proof of a diffusion limit for the output of the MALA algorithm, suitably interpolated, to the SPDE (1.7), when applied to $N-$dimensional approximations of the target measures (1.6) with proposal variance inversely proportional to $N^{\frac{1}{3}}$. Moreover we show that the speed $h(\ell)$ of the limiting diffusion is maximized for an average acceptance probability of 0.574, just as in the i.i.d product scenario [RR98]. Thus in this regard, our work is the first extension of the remarkable results in [RR98] for the Langevin algorithm to target measures which are not of product form. This adds theoretical weight to the results observed in computational experiments which demonstrate the robustness of the optimality creiteria developed in [RGG97, RR98]. In particular the paper [BRSV08] shows numerical results indicating the need to scale time-step as a function of dimension to obtain $\mathcal{O}(1)$ acceptance probabilities.

In section 2 we state the main theorem of the paper, having defined precisely the setting in which it holds. Section 3 contains the proof of the main theorem, postponing the proof of a number of key technical estimtes to section 4. In section 5 we conclude by summarizing and providing the outlook for further research in this area.

2. Main Theorem. This section is devoted to stating the main theorem of the article. However the setting is complex and we develop this in a step-by-step fashion, before the theorem statement. In subsection 2.1 we introduce the form of the reference, or prior, Gaussian measure π_0, followed in subsection 2.2 by the change of measure which induces a genuinely non-product structure. In subsection 2.3 we describe finite dimensional approximation of the measure, enabling us to define application of a variant MALA type algorithm in subsection 2.4. We then discuss in subsection 2.5 how the choice of scaling used in the theorem emerges from study of the acceptance probabilities. Finally, in subsection 2.6, we state the main theorem.

Throughout the paper we use the following notation in order to compare sequences and to denote conditional expectations.

- Two sequences $\{\alpha_n\}$ and $\{\beta_n\}$ satisfy $\alpha_n \lesssim \beta_n$ if there exists a constant $K > 0$ satisfying $\alpha_n \le K\beta_n$ for all $n \ge 0$. 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 $\{f_n\}$ and $\{g_n\}$ defined on the same set D satisfy $f_n \lesssim g_n$ if there exists a constant $K > 0$ satisfying $f_n(x) \le K g_n(x)$ for all $n \ge 0$ and all $x \in D$. The notations $f_n \asymp g_n$ means that $f_n \lesssim g_n$ and $g_n \lesssim f_n$.
- The notation $\mathbb{E}_x\big[f(x,\xi)\big]$ denotes expectation with respect to ξ with the variable x fixed.

2.1. *Gaussian Reference Measure.* Let $\mathcal{H}$ be a separable Hilbert space of real valued functions with scalar product denoted by $\langle \cdot,\cdot \rangle$ and associated norm $\|x\|^2 = \langle x,x \rangle$. Consider a Gaussian probability measure π_0 on $(\mathcal{H}, \|\cdot\|)$ with covariance operator $\mathcal{C}$. The general theory of Gaussian measures [DPZ92] ensures that the operator $\mathcal{C}$ is positive and trace class. Let $\{\varphi_j, \lambda_j^2\}_{j\ge 1}$ be the eigenfunctions and eigenvalues of the covariance operator $\mathcal{C}$:

$$\mathcal{C}\varphi_j = \lambda_j^2\,\varphi_j, \qquad\qquad j \ge 1.$$

We assume a normalization under which the family $\{\varphi_j\}_{j\ge 1}$ forms a complete orthonormal basis in the Hilbert space $\mathcal{H}$, which we refer to us as the Karhunen-Loève basis. Any function $x \in \mathcal{H}$ can be represented in this basis via the expansion

$$x = \sum_{j=1}^{\infty} x_j\,\varphi_j, \qquad x_j \stackrel{\text{def}}{=} \langle x, \varphi_j \rangle. \tag{2.1}$$

Throughout this paper we will often identify the function x with its coordinates $\{x_j\}_{j=1}^{\infty} \in \ell^2$ in this eigenbasis, moving freely between the two representations. The Karhunen-Loève expansion (see [DPZ92], section *White Noise expansions*), refers to the fact that a realization x from the Gaussian measure π_0 can be expressed by allowing the coordinates $\{x_j\}_{j\ge 1}$ in (2.1) to be independent random variables distributed as $x_j \sim \mathrm{N}(0, \lambda_j^2)$. Thus, in the coordinates $\{x_j\}_{j\ge 1}$, the Gaussian reference measure π_0 has a product structure.

For every $x \in \mathcal{H}$ we have the representation (2.1). Using this expansion, we define Sobolev-like spaces $\mathcal{H}^r, r \in \mathbb{R}$, with the inner-products and norms defined by

$$\langle x, y \rangle_r \stackrel{\text{def}}{=} \sum_{j=1}^{\infty} j^{2r} x_j y_j, \qquad \|x\|_r^2 \stackrel{\text{def}}{=} \sum_{j=1}^{\infty} j^{2r}\, x_j^2. \tag{2.2}$$

Notice that $\mathcal{H}^0 = \mathcal{H}$ and $\mathcal{H}^r \subset \mathcal{H} \subset \mathcal{H}^{-r}$ for any $r > 0$. The Hilbert-Schmidt norm $\|\cdot\|_{\mathcal{C}}$ associated to the covariance operator $\mathcal{C}$ is defined as

$$\|x\|_{\mathcal{C}}^2 = \sum_j \lambda_j^{-2} x_j^2.$$

For $x, y \in \mathcal{H}^r$, the outer product operator in $\mathcal{H}^r$ is the operator $x \otimes_{\mathcal{H}^r} y : \mathcal{H}^r \to \mathcal{H}^r$ defined by $(x \otimes_{\mathcal{H}^r} y) z \stackrel{\text{def}}{=} \langle y, z \rangle_r \, x$ for every $z \in \mathcal{H}^r$. For $r \in \mathbb{R}$, let $B_r : \mathcal{H} \mapsto \mathcal{H}$ denote the operator which is diagonal in the basis $\{\varphi_j\}_{j \geq 1}$ with diagonal entries j^{2r}. The operator B_r satisfies $B_r \, \varphi_j = j^{2r} \varphi_j$ so that $B_r^{\frac{1}{2}} \, \varphi_j = j^r \varphi_j$. The operator B_r lets us alternate between the Hilbert space $\mathcal{H}$ and the Sobolev spaces $\mathcal{H}^r$ via the identities $\langle x, y \rangle_r = \langle B_r^{\frac{1}{2}} x, B_r^{\frac{1}{2}} y \rangle$. Since $\|B_r^{-1/2} \varphi_k\|_r = \|\varphi_k\| = 1$, we deduce that $\{B_r^{-1/2} \varphi_k\}_{k \geq 0}$ forms an orthonormal basis for $\mathcal{H}^r$. For a positive, self-adjoint operator $D : \mathcal{H} \mapsto \mathcal{H}$, we define its trace in $\mathcal{H}^r$ by

$$\mathrm{Tr}_{\mathcal{H}^r}(D) \stackrel{\text{def}}{=} \sum_{j=1}^{\infty} \langle (B_r^{-\frac{1}{2}} \varphi_j), D (B_r^{-\frac{1}{2}} \varphi_j) \rangle_r. \tag{2.3}$$

Since $\mathrm{Tr}_{\mathcal{H}^r}(D)$ does not depend on the orthonormal basis, the operator D is said to be trace class in $\mathcal{H}^r$ if $\mathrm{Tr}_{\mathcal{H}^r}(D) < \infty$ for some, and hence any, orthonormal basis of $\mathcal{H}^r$. Let us define the operator $\mathcal{C}_r \stackrel{\text{def}}{=} B_r^{1/2} \, \mathcal{C} \, B_r^{1/2}$. Notice that $\mathrm{Tr}_{\mathcal{H}^r}(\mathcal{C}_r) = \sum_{j=1}^{\infty} \lambda_j^2 \, j^{2r}$. In [MPS11] it is shown that under the condition

$$\mathrm{Tr}_{\mathcal{H}^r}(\mathcal{C}_r) < \infty, \tag{2.4}$$

the support of π_0 is included in $\mathcal{H}^r$ in the sense that π_0-almost every function $x \in \mathcal{H}$ belongs to $\mathcal{H}^r$. Furthermore, the induced distribution of π_0 on $\mathcal{H}^r$ is identical to that of a centered Gaussian measure on $\mathcal{H}^r$ with covariance operator $\mathcal{C}_r$. For example, if $\xi \stackrel{\mathcal{D}}{\sim} \pi_0$, then $\mathbb{E}\big[\langle \xi, u \rangle_r \langle \xi, v \rangle_r\big] = \langle u, \mathcal{C}_r v \rangle_r$ for any functions $u, v \in \mathcal{H}^r$. Thus in what follows, we alternate between the Gaussian measures $\mathrm{N}(0, \mathcal{C})$ on $\mathcal{H}$ and $\mathrm{N}(0, \mathcal{C}_r)$ on $\mathcal{H}^r$, for those r for which (2.4) holds.

2.2. *Change of Measure.* Our goal is to sample from a measure π defined through the change of probability formula (1.6). As described in section 2.1, the condition $\mathrm{Tr}_{\mathcal{H}^r}(\mathcal{C}_r) < \infty$ implies that the measure π_0 has full support on $\mathcal{H}^r$, *i.e.*, $\pi_0(\mathcal{H}^r) = 1$. Consequently, if $\mathrm{Tr}_{\mathcal{H}^r}(\mathcal{C}_r) < \infty$, the functional $\Psi(\cdot)$ needs only to be defined on $\mathcal{H}^r$ in order for the change of probability formula (1.6) to be valid. In this section we give assumptions on the decay of the eigenvalues of the covariance operator $\mathcal{C}$ of π_0 that ensure the existence of a real number $s > 0$ such that π_0 has full support on $\mathcal{H}^s$. The functional $\Psi(\cdot)$ is assumed to be defined on $\mathcal{H}^s$ and we impose regularity assumptions on $\Psi(\cdot)$ that ensure that the probability distribution π is not too different from π_0, when projected into directions associated with φ_j for j large. 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 linear functionals on $\mathcal{H}^s$. However, we may identify $(\mathcal{H}^s)^*$ with $\mathcal{H}^{-s}$ 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 as an element of $\mathcal{L}(\mathcal{H}^s, \mathcal{H}^{-s})$. To avoid technicalities we assume that $\Psi(\cdot)$ is quadratically bounded, with first derivative linearly bounded and second derivative globally bounded. Weaker assumptions could be dealt with by use of stopping time arguments.

Assumptions 2.1. *The covariance operator $\mathcal{C}$ and functional Ψ satisfy the following:*

1. ***Decay of Eigenvalues λ_j^2 of $\mathcal{C}$:*** *there is an exponent $\kappa > \frac{1}{2}$ such that*

$$\lambda_j \asymp j^{-\kappa}. \tag{2.5}$$

2. ***Assumptions on*** Ψ*: There exist constants* $M_i \in \mathbb{R}, i \le 4$ *and* $s \in [0, \kappa - 1/2)$ *such that for all* $x \in \mathcal{H}^s$ *the functional* $\Psi : \mathcal{H}^s \to \mathbb{R}$ *satisfies*

$$M_1 \le \Psi(x) \le M_2 \left(1 + \|x\|_s^2\right) \tag{2.6}$$

$$\|\nabla\Psi(x)\|_{-s} \le M_3 \left(1 + \|x\|_s\right) \tag{2.7}$$

$$\|\partial^2\Psi(x)\|_{\mathcal{L}(\mathcal{H}^s, \mathcal{H}^{-s})} \le M_4. \tag{2.8}$$

REMARK 2.2. *The condition* $\kappa > \frac{1}{2}$ *ensures that the covariance operator* $\mathcal{C}$ *is trace class in* $\mathcal{H}$*. In fact, Equation (2.4) shows that* $\mathcal{C}_r$ *is trace-class in* $\mathcal{H}^r$ *for any* $r < \kappa - \frac{1}{2}$*. It follows that* π_0 *has full measure in* $\mathcal{H}^r$ *for any* $r \in [0, \kappa - 1/2)$*. In particular* π_0 *has full support on* $\mathcal{H}^s$*.*

REMARK 2.3. *The functional* $\Psi(x) = \frac{1}{2}\|x\|_s^2$ *satisfies Assumptions 2.1. It is defined on* $\mathcal{H}^s$ *and its derivative at* $x \in \mathcal{H}^s$ *is given by* $\nabla\Psi(x) = \sum_{j \ge 0} j^{2s} x_j \varphi_j \in \mathcal{H}^{-s}$ *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* $\sum_{j \ge 0} j^{2s} \langle u, \varphi_j \rangle \varphi_j \in \mathcal{H}^s$*: its norm satisfies* $\|\partial^2\Psi(x)\|_{\mathcal{L}(\mathcal{H}^s, \mathcal{H}^{-s})} = 1$ *for any* $x \in \mathcal{H}^s$*.*

Since the eigenvalues λ_j^2 of $\mathcal{C}$ decrease as $\lambda_j \asymp j^{-\kappa}$, the operator $\mathcal{C}$ has a smoothing effect: $\mathcal{C}^\alpha h$ gains $2\alpha\kappa$ orders of regularity in the sense that the $\mathcal{H}^\beta$-norm of $\mathcal{C}^\alpha h$ is controlled by the $\mathcal{H}^{\beta - 2\alpha\kappa}$-norm of $h \in \mathcal{H}$. Indeed, under Assumption 2.1, the following estimates holds:

$$\|h\|_{\mathcal{C}} \asymp \|h\|_\kappa \qquad \text{and} \qquad \|\mathcal{C}^\alpha h\|_\beta \asymp \|h\|_{\beta - 2\alpha\kappa}. \tag{2.9}$$

The proof follows the methodology used to prove Lemma 3.3 of [MPS11]. The reader is referred to this text for more details.

2.3. *Finite Dimensional Approximation.* We are interested in finite dimensional approximations of the probability distribution π. To this end, we introduce the vector space spanned by the first N eigenfunctions of the covariance operator,

$$X^N \stackrel{\text{def}}{=} \operatorname{span}\{\varphi_1, \varphi_2, \dots, \varphi_N\}.$$

Notice that $X^N \subset \mathcal{H}^r$ for any $r \in [0; +\infty)$. In particular, X^N is a subspace of $\mathcal{H}^s$. Next, we define N-dimensional approximations of the functional $\Psi(\cdot)$ and of the reference measure π_0. To this end, we introduce the orthogonal projection on X^N denoted by $P^N : \mathcal{H}^s \mapsto X^N \subset \mathcal{H}^s$. The functional $\Psi(\cdot)$ is approximated by the functional $\Psi^N : X^N \mapsto \mathbb{R}$ defined by

$$\Psi^N \stackrel{\text{def}}{=} \Psi \circ P^N. \tag{2.10}$$

The approximation π_0^N of the reference measure π_0 is the Gaussian measure on X^N given by the law of the random variable

$$\pi_0^N \stackrel{\mathcal{D}}{\sim} \sum_{j=1}^N \lambda_j \xi_j \varphi_j \;=\; (\mathcal{C}^N)^{\frac{1}{2}}\, \xi^N$$

where ξ_j are i.i.d standard Gaussian random variables, $\xi^N = \sum_{j=1}^N \xi_j \varphi_j$ and $\mathcal{C}^N = P^N \circ \mathcal{C} \circ P^N$. Consequently we have $\pi_0^N = \mathrm{N}(0, \mathcal{C}^N)$. Finally, one can define the approximation π^N of π by the change of probability formula

$$\frac{d\pi^N}{d\pi_0^N}(x) = M_{\Psi^N} \exp\big(-\Psi^N(x)\big) \tag{2.11}$$

where M_{Ψ^N} is a normalization constant. Notice that the probability distribution π^N is supported on X^N and has Lebesgue density[2] on X^N equal to

$$\pi^N(x) \propto \exp\Big(-\frac{1}{2}\|x\|^2_{\mathcal{C}^N} - \Psi^N(x)\Big). \tag{2.12}$$

In formula (2.12), the Hilbert-Schmidt norm $\|\cdot\|_{\mathcal{C}^N}$ on X^N is given by the scalar product $\langle u,v\rangle_{\mathcal{C}^N} = \langle u, (\mathcal{C}^N)^{-1}v\rangle$ for all $u,v\in X^N$. The operator $\mathcal{C}^N$ is invertible on X^N because the eigenvalues of $\mathcal{C}$ are assumed to be strictly positive. The quantity $\mathcal{C}^N\nabla\log\pi^N(x)$ is repeatedly used in the text and in particular appears in the function $\mu^N(x)$ given by

$$\mu^N(x) = -\Big(P^N x + \mathcal{C}^N\nabla\Psi^N(x)\Big) \tag{2.13}$$

which, upto an additive constants, is $\mathcal{C}^N\nabla\log\pi^N(x)$. This function is the drift of an ergodic Langevin diffusion that leaves π^N invariants. Similarly, one defines the function $\mu:\mathcal{H}^s\to\mathcal{H}^s$ given by

$$\mu(x) = -\Big(x + \mathcal{C}\nabla\Psi(x)\Big) \tag{2.14}$$

which can informally be seen as $\mathcal{C}\nabla\log\pi(x)$, upto an additive constant. In the sequel, Lemma 4.1 shows that, for π_0-almost every function $x\in\mathcal{H}$, we have $\lim_{N\to\infty}\mu^N(x) = \mu(x)$. This quantifies the manner in which $\mu^N(\cdot)$ is an approximation of $\mu(\cdot)$.

The next lemma gathers various regularity estimates on the functional $\Psi(\cdot)$ and $\Psi^N(\cdot)$ that are repeatedly used in the sequel. These are simple consequences of Assumptions 2.1 and proofs can be found in [MPS11].

Lemma 2.4. **(Properties of Ψ)** *Let the functional $\Psi(\cdot)$ satisfy Assumptions 2.1 and consider the functional $\Psi^N(\cdot)$ defined by Equation (2.10). The following estimates hold.*

1. *The functionals $\Psi^N:\mathcal{H}^s\to\mathbb{R}$ satisfy the same conditions imposed on Ψ given by Equations (2.6), (2.7) and (2.8) with constants that can be chosen independent of N.*
2. *The function $\mathcal{C}\nabla\Psi:\mathcal{H}^s\to\mathcal{H}^s$ is globally Lipschitz on $\mathcal{H}^s$: there exists a constant $M_5>0$ such that*

$$\|\mathcal{C}\nabla\Psi(x) - \mathcal{C}\nabla\Psi(y)\|_s \le M_5\,\|x-y\|_s \qquad \forall x,y\in\mathcal{H}^s.$$

   Moreover, the functions $\mathcal{C}^N\nabla\Psi^N:\mathcal{H}^s\to\mathcal{H}^s$ also satisfy this estimate with a constant that can be chosen independent of N.
3. *The functional $\Psi(\cdot):\mathcal{H}^s\to\mathbb{R}$ satisfies a second order Taylor formula[3]. There exists a constant $M_6>0$ such that*

$$\Psi(y) - \Big(\Psi(x) + \langle\nabla\Psi(x), y-x\rangle\Big) \le M_6\,\|x-y\|_s^2 \qquad \forall x,y\in\mathcal{H}^s. \tag{2.15}$$

   Moreover, the functionals $\Psi^N(\cdot)$ also satisfy this estimates with a constant that can be chosen independent of N.

Remark 2.5. *The regularity Lemma 2.4 shows in particular that the function $\mu:\mathcal{H}^s\to\mathcal{H}^s$ defined by (2.14) is globally Lipschitz on $\mathcal{H}^s$. Similarly, it follows that $\mathcal{C}^N\nabla\Psi^N:\mathcal{H}^s\to\mathcal{H}^s$ and $\mu^N:\mathcal{H}^s\to\mathcal{H}^s$ given by (2.13) are globally Lipschitz with Lipschitz constants that can be chosen uniformly in N.*

---

[2] For ease of notation we do not distinguish between a measure and its density, nor do we distinguish between the representation of the measure in X^N or in coordinates in $\mathbb{R}^N$

[3] We extend $\langle\cdot,\cdot\rangle$ from an inner-product on $\mathcal{H}$ to the dual pairing between $\mathcal{H}^{-s}$ and $\mathcal{H}^s$.

2.4. *The Algorithm.* The MALA algorithm is defined in this section. This method is motivated by the fact that the probability measure π^N defined by Equation (2.11) is invariant with respect to the Langevin diffusion process

$$\frac{dz}{dt} = \mu^N(z) + \sqrt{2}\,\frac{dW^N}{dt}, \tag{2.16}$$

where W^N is a Brownian motion in $\mathcal{H}$ with covariance operator $\mathcal{C}^N$. The drift function $\mu^N : \mathcal{H}^s \to \mathcal{H}^s$ is the gradient of the log-density of π^N, as described by Equation (2.13). The idea of the MALA algorithm is to make a proposal based on Euler-Maruyama discretization of the diffusion (2.16). To this end we consider, from state $x \in X^N$, proposals $y \in X^N$ given by

$$y - x = \delta\, \mu^N(x) + \sqrt{2\delta}\,(\mathcal{C}^N)^{\frac{1}{2}} \xi^N \qquad \text{where} \qquad \delta = \ell N^{-\frac{1}{3}} \tag{2.17}$$

with $\xi^N = \sum_{i=1}^N \xi_i\, \varphi_i$ and $\xi_i \overset{\mathcal{D}}{\sim} \mathrm{N}(0,1)$. Notice that $(\mathcal{C}^N)^{\frac{1}{2}} \xi^N \overset{\mathcal{D}}{\sim} \mathrm{N}(0, \mathcal{C}^N)$. The quantity δ is the time-step in an Euler-Maruyama discretization of (2.16). We introduce a related parameter

$$\Delta t := \ell^{-1}\delta = N^{-\frac{1}{3}}$$

which will be the natural time-step for the limiting diffusion process derived from the proposal above, after inclusion of an accept-reject mechanism. The scaling of Δt, and hence δ, with N will ensure that the average acceptance probability is of order 1 as N grows. This is discussed in more detail in section 2.5. The quantity $\ell > 0$ is a fixed parameter which can be chosen to maximize the speed of the limiting diffusion process: see the discussion in the introduction and after the Main Theorem below.

We will study the Markov chain $x^N = \{x^{k,N}\}_{k \geq 0}$ resulting from Metropolizing this proposal when it is started at stationarity: the initial position $x^{0,N}$ is distributed as π^N and thus lies in X^N. Therefore, the Markov chain evolves in X^N; as a consequence, only the first N components of an expansion in the eigenbasis of $\mathcal{C}$ are nonzero and the algorithm can be implemented in $\mathbb{R}^N$. However the analysis is cleaner when written in $X^N \subset \mathcal{H}^s$. The acceptance probability only depends on the first N coordinates of x and y and has the form

$$\alpha^N(x, \xi^N) = 1 \wedge \frac{\pi^N(y) T^N(y,x)}{\pi^N(x) T^N(x,y)} = 1 \wedge e^{Q^N(x,\xi^N)} \tag{2.18}$$

where the proposal y is given by Equation (2.17). The function $T^N(\cdot,\cdot)$ is the density of the Langevin proposals (2.17) and is given by

$$T^N(x,y) \propto \exp\Big\{ -\frac{1}{4\delta} \|y - x - \delta \mu^N(x)\|^2_{\mathcal{C}^N} \Big\}$$

The local mean acceptance probability $\alpha^N(x)$ is defined by

$$\alpha^N(x) = \mathbb{E}_x\big[\alpha^N(x, \xi^N)\big]. \tag{2.19}$$

It is the expected acceptance probability when the algorithm stands at $x \in \mathcal{H}$. The Markov chain $x^N = \{x^{k,N}\}_{k \geq 0}$ can also be expressed as

$$\begin{cases} y^{k,N} &= x^{k,N} + \delta \mu^N(x^{k,N}) + \sqrt{2\delta}\,(\mathcal{C}^N)^{\frac{1}{2}}\, \xi^{k,N} \\ x^{k+1,N} &= \gamma^{k,N} y^{k,N} + (1 - \gamma^{k,N})\, x^{k,N} \end{cases} \tag{2.20}$$

where $\xi^{k,N}$ are i.i.d samples distributed as ξ^N and $\gamma^{k,N} = \gamma^N(x^{k,N}, \xi^{k,N})$ creates a Bernoulli random sequence with k^{th} success probability $\alpha^N(x^{k,N}, \xi^{k,N})$. We may view the Bernoulli random variable as $\gamma^{k,N} = 1_{\{U^k < \alpha^N(x^{k,N}, \xi^{k,N})\}}$ where $U^k \overset{\mathcal{D}}{\sim} \text{Uniform}(0,1)$ is independent from $x^{k,N}$ and $\xi^{k,N}$. The quantity Q^N defined in Equation (2.18) may be expressed as

$$Q^N(x,\xi^N) = -\frac{1}{2}\Big(\|y\|^2_{\mathcal{C}^N} - \|x\|^2_{\mathcal{C}^N}\Big) - \Big(\Psi^N(y) - \Psi^N(x)\Big) \\ - \frac{1}{4\delta}\Big\{\|x - y - \delta\mu^N(y)\|^2_{\mathcal{C}^N} - \|y - x - \delta\mu^N(x)\|^2_{\mathcal{C}^N}\Big\}. \tag{2.21}$$

As will be seen in the next section, a key idea behind our diffusion limit is that, for large N, the quantity $Q^N(x, \xi^N)$ behaves like a Gaussian random variable independent from the current position x.

In summary, the Markov chain that we have described in $\mathcal{H}^s$ is, when projected onto X^N, equivalent to a standard MALA algorithm on $\mathbb{R}^N$ for the Lebesgue density (2.12). Recall that the target measure π in (1.6) is the invariant measure of the SPDE (1.7). Our goal is to obtain an invariance principle for the continuous interpolant (1.4) of the Markov chain $x^N = \{x^{k,N}\}_{k\geq 0}$ started in stationarity, *i.e*, to show weak convergence in $C([0,T];\mathcal{H}^s)$ of $z^N(t)$ to the solution $z(t)$ of the SPDE (1.7), as the dimension $N \to \infty$.

2.5. *Optimal scale* $\gamma = \frac{1}{3}$. In this section, we informally describe why the optimal scale for the MALA proposals (2.17) is given by the exponent $\gamma = \frac{1}{3}$. For product-form target probability described by Equation (1.1), the optimality of the exponent $\gamma = \frac{1}{3}$ was first obtained in [RR98]. For further discussion, see also [BRS09]. To keep the exposition simple in this explanatory subsection we focus on the case $\Psi(\cdot) = 0$. The analysis is similar with a non-vanishing function $\Psi(\cdot)$, because absolute continuity ensures that the effect of $\Psi(\cdot)$ is small compared to the dominant Gaussian effects described here. Inclusion of non-vanishing $\Psi(\cdot)$ is carried out in Lemma 4.4.

In the case $\Psi(\cdot) = 0$, straightforward algebra shows that the acceptance probability $\alpha^N(x, \xi^N) = 1 \wedge e^{Q^N(x,\xi^N)}$ satisfies

$$Q^N(x,\xi^N) = -\frac{\ell\Delta t}{4}\Big(\|y\|^2_{\mathcal{C}^N} - \|x\|^2_{\mathcal{C}^N}\Big).$$

For $\Psi(\cdot) = 0$ and $x \in X^N$, the proposal y is distributed as $y = (1 - \ell\Delta t)x + \sqrt{2\ell\Delta t}(\mathcal{C}^N)^{\frac{1}{2}}\,\xi^N$. It follows that

$$\|y\|^2_{\mathcal{C}^N} - \|x\|^2_{\mathcal{C}^N} = -2\ell\Delta t\left(\|x\|^2_{\mathcal{C}^N} - \|(\mathcal{C}^N)^{\frac{1}{2}}\xi^N\|^2_{\mathcal{C}^N}\right) + (\ell\Delta t)^2\,\|x\|^2_{\mathcal{C}^N} \\ + 2\sqrt{2\ell\Delta t}(1-\Delta t)\,\langle x, (\mathcal{C}^N)^{\frac{1}{2}}\xi^N\rangle_{\mathcal{C}^N}.$$

The details can be found in the proof of Lemma 4.4. Since the Markov chain $x^N = \{x^{k,N}\}_{k\geq 0}$ evolves in stationarity, for all $k \geq 0$ we have $x^{k,N} \overset{\mathcal{D}}{\sim} \pi^N = \text{N}(0, \mathcal{C}^N)$. Therefore, with $x \overset{\mathcal{D}}{\sim} \text{N}(0,\mathcal{C}^N)$ and $\xi^N \overset{\mathcal{D}}{\sim} \text{N}(0,\mathcal{C}^N)$, the Law of Large Numbers shows that both $\|x\|^2_{\mathcal{C}^N}$ and $\|(\mathcal{C}^N)^{\frac{1}{2}}\xi^N\|^2_{\mathcal{C}^N}$ are of order $\mathcal{O}(N)$, whilst the Central Limit Theorem shows that $\langle x, (C^N)^{\frac{1}{2}}\xi^N\rangle_{\mathcal{C}^N} = \mathcal{O}(N^{\frac{1}{2}})$ and $\|x\|^2_{\mathcal{C}^N} - \|(\mathcal{C}^N)^{\frac{1}{2}}\xi^N\|^2_{\mathcal{C}^N} = \mathcal{O}(N^{\frac{1}{2}})$. For $\Delta t = \ell N^{-\gamma}$ and $\gamma < \frac{1}{3}$, it follows

$$Q^N(x,\xi^N) = -\frac{(\ell\Delta t)^3}{4}\|x\|^2_{\mathcal{C}^N} + \mathcal{O}(N^{\frac{1}{2}-\frac{3\gamma}{2}}) \;\approx\; -\frac{\ell^3}{4}N^{1-3\gamma},$$

which shows that the acceptance probability is exponentially small of order $\exp(-\frac{\ell^3}{4}N^{1-3\gamma})$. The same argument shows that for $\gamma > \frac{1}{3}$ we have $Q^N(x,\xi^N) \to 0$, which shows that the average acceptance probability converges to 1. For the critical exponent $\gamma = \frac{1}{3}$ the acceptance probability is of order $\mathcal{O}(1)$. In fact Lemma 4.4 shows that for $\gamma = \frac{1}{3}$, even when $\Psi(\cdot)$ is non-zero, the following Gaussian approximation holds:

$$Q^N(x,\xi^N) \;\approx\; \mathrm{N}(-\frac{\ell^3}{4}, \frac{\ell^3}{2}).$$

This approximation is key to derivation of the diffusion limit. In summary, choosing $\gamma > \frac{1}{3}$ leads to exponentially small acceptance probabilities: almost all the proposals are rejected so that the expected squared jumping distance $\mathbb{E}_{\pi^N}[\|x^{k+1,N} - x^{k,N}\|^2]$ converges exponentially quickly to 0 as the dimension N goes to infinity. On the other hand, for any exponent $\gamma \geq \frac{1}{3}$, the acceptance probabilities are bounded away from zero: the Markov chain moves with jumps of size $\mathcal{O}(N^{-\frac{\gamma}{2}})$ and the expected squared jumping distance is of order $\mathcal{O}(N^{-\gamma})$. If we adopt the expected squared jumping distance as measure of efficiency, the optimal exponent is thus given by $\gamma = \frac{1}{3}$. This viewpoint is analyzed further in [BRS09].

2.6. *Statement of Main Theorem.* The main result of this article describes the behavior of the MALA algorithm for the optimal scale $\gamma = \frac{1}{3}$; the proposal variance is given by $\delta = 2\ell N^{-\frac{1}{3}}$. In this case, Lemma 4.4 shows that the local mean acceptance probability $\alpha^N(x,\xi^N) = 1 \wedge e^{Q^N(x,\xi^N)}$ satisfies $Q^N(x,\xi^N) \to Z_\ell \stackrel{\mathcal{D}}{\sim} \mathrm{N}(-\frac{\ell^3}{4}, \frac{\ell^3}{2})$. As a consequence, the asymptotic mean acceptance probability of the MALA algorithm can be explicitly computed as a function of the parameter $\ell > 0$,

$$\alpha(\ell) \;\stackrel{\text{def}}{=}\; \lim_{N\to\infty} \mathbb{E}^{\pi^N}\big[\alpha^N(x,\xi^N)\big] \;=\; \mathbb{E}\big[1 \wedge e^{Z_\ell}\big].$$

This result is rigorously proved as Corollary 4.6. We then define the "speed function"

$$h(\ell) = \ell\alpha(\ell). \tag{2.22}$$

Note that the time step made in the proposal is $\delta = l\Delta t$ and that, if this is accepted a fraction $\alpha(\ell)$ of the time, then a naive argument invoking independence shows that the effective time-step is reduced to $h(l)\Delta t$. This is made rigorous in the following Theorem 2.6 which shows that the quantity $h(\ell)$ is the asymptotic speed function of the limiting diffusion obtained by rescaling the Metropolis-Hastings Markov chain $x^N = \{x^{k,N}\}_{k\geq 0}$.

THEOREM 2.6. **(Main Theorem)** *Let the reference measure π_0 and the function $\Psi(\cdot)$ satisfy Assumptions 2.1. Consider the MALA algorithm (2.20) with initial condition $x^{0,N} \stackrel{\mathcal{D}}{\sim} \pi^N$. Let $z^N(t)$ be the piecewise linear, continuous interpolant of the MALA algorithm as defined in (1.4), with $\Delta t = N^{-\frac{1}{3}}$. Then $z^N(t)$ converges weakly in $C([0,T],\mathcal{H}^s)$ to the diffusion process $z(t)$ given by*

$$\frac{dz}{dt} = -h(\ell)\big(z + \mathcal{C}\nabla\Psi(z)\big) + \sqrt{2\,h(\ell)}\,\frac{dW}{dt} \tag{2.23}$$

with initial distribution $z(0) \stackrel{\mathcal{D}}{\sim} \pi$.

We now explain the following two important implications of this result.

- Since time has to be accelerated by a factor $(\Delta t)^{-1} = N^{\frac{1}{3}}$ in order to observe a diffusion limit, it follows that in stationarity the work required to explore the invariant measure scales as $\mathcal{O}(N^{\frac{1}{3}})$.
- The speed at which the invariant measure is explored, again in stationarity, is maximized by choosing ℓ so as to maximize $h(\ell)$; this is achieved at an average acceptance probability 0.574. From a practical point of view, this shows that one should "tune" the proposal variance of the MALA algorithm so as to have a mean acceptance probability of 0.574.

The first implication follows from (1.4) since this shows that $\mathcal{O}(N^{\frac{1}{3}})$ steps of the MALA Markov chain (2.20) are required for $z^N(t)$ to approximate $z(t)$ on a time interval $[0, T]$ long enough for $z(t)$ to have explored its invariant measure. To understand the second implication, note that if $Z(t)$ solves (2.23) with $h(\ell) \equiv 1$ then, in law, $z(t) = Z(h(\ell)t)$. This result suggests choosing the value of ℓ that maximizes the speed function $h(\cdot)$ since $z(t)$ will then explore the invariant measure as fast as possible. For practitioners, who often tune algorithms according to the acceptance probability, it is relevant to express the maximization principle in terms of the asymptotic mean acceptance probability $\alpha(\ell)$. Figure 2.6 shows that the speed function $h(\cdot)$ is maximized for an optimal acceptance probability of $\alpha^\star = 0.574$, to three decimal places. This is precisely the argument used in [RR98] for the case of product target measures and it is remarkable that the optimal acceptance probability identified in that context is also optimal for the non-product measures studied in this paper.

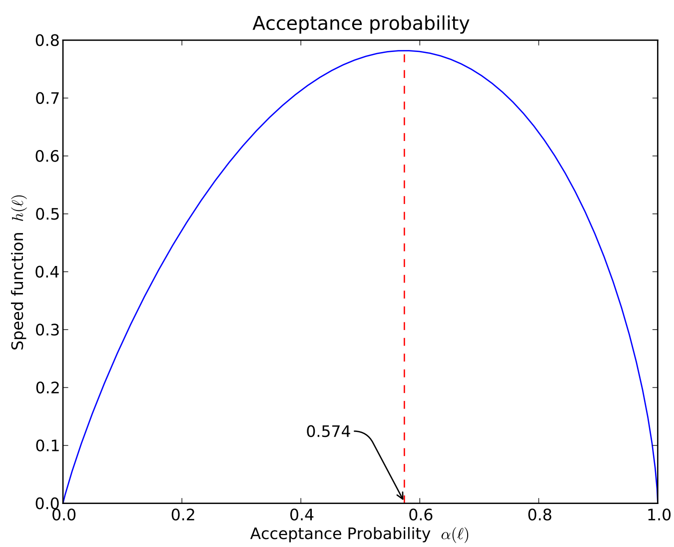

FIG 1. *Optimal acceptance probability* $= 0.574$

3. Proof of Main Theorem. In subsection 3.1 we outline the proof strategy and introduce the drift-martginale decomposition of our discrete-time Markov chain which underlies it. Subsection 3.2 contains statement and proof of a general diffusion approximation, Proposition 3.1. In subsection 3.3 we use this proposition to prove the main theorem of this paper, pointing to section 4 for the key estimates required.

3.1. *Proof Strategy.* To communicate the main ideas, we give a heuristic of the proof before proceeding to give full details in subsequent sections. Let us first examine a simpler situation: consider a scalar Lipschitz function $\mu : \mathbb{R} \to \mathbb{R}$ and two scalar constants $\ell, c > 0$. The usual theory of diffusion approximation for Markov processes [EK86] shows that the sequence $x^N = \{x^{k,N}\}$ of Markov chains

$$x^{k+1,N} - x^{k,N} = \mu(x^{k,N})\, \ell N^{-\frac{1}{3}} + \sqrt{2\ell N^{-\frac{1}{3}}}\; c^{\frac{1}{2}} \xi^k,$$

with i.i.d. $\xi^k \stackrel{\mathcal{D}}{\sim} \mathrm{N}(0,1)$ converges weakly, when interpolated using a time-acceleration factor of $N^{\frac{1}{3}}$, to the scalar diffusion $dz(t) = \ell\mu(z(t))\, dt + \sqrt{2\ell}\, dW(t)$ where W is a Brownian motion with variance $\mathrm{Var}(W(t)) = ct$. Also, if γ^k is an i.i.d. sequence of Bernoulli random variables with success rate $\alpha(\ell)$, independent from the Markov chain x^N, one can prove that the sequence $x^N = \{x^{k,N}\}$ of Markov chains given by

$$x^{k+1,N} - x^{k,N} = \gamma^k \left\{ \mu(x^{k,N}) \ell N^{-\frac{1}{3}} + \sqrt{2\ell N^{-\frac{1}{3}}}\; c^{\frac{1}{2}} \xi^k \right\}$$

converges weakly, when interpolated using a time-acceleration factor $N^{\frac{1}{3}}$, to the diffusion

$$dz(t) = h(\ell)\mu(z(t))\, dt + \sqrt{2h(\ell)}\, dW(t)$$

where the speed function is given by $h(\ell) = \ell\alpha(\ell)$. This shows that the Bernoulli random variables $\{\gamma^k\}_{k \geq 0}$ have slowed down the original Markov chain by a factor $\alpha(\ell)$. The proof of Theorem 2.6 is an application of this idea in a slightly more general setting. The following complications arise.

- Instead of working with scalar diffusions, the result holds for a Hilbert space-valued diffusion. The correlation structure between the different coordinates is not present in the preceding simple example and has to be taken into account.
- Instead of working with a single drift function μ, A sequence of approximations d^N converging to μ has to be taken into account.
- The Bernoulli random variables $\gamma^{k,N}$ are not i.i.d. and have an autocorrelation structure. On top of that, the Bernoulli random variables $\gamma^{k,N}$ are not independent from the Markov chain $x^{k,N}$. This is the main difficulty in the proof.
- It should be emphasized that the main theorem uses the fact that the MALA Markov chain is started at stationarity: this in particular implies that $x^{k,N} \stackrel{\mathcal{D}}{\sim} \pi^N$ for any $k \geq 0$, which is crucial to the proof of the invariance principle as it allows us to control the correlation between $\gamma^{k,N}$ and $x^{k,N}$.

The acceptance probability of the proposal (2.17) is equal to $\alpha^N(x, \xi^N) = 1 \wedge e^{Q^N(x,\xi^N)}$ and the quantity $\alpha^N(x) = \mathbb{E}_x[\alpha^N(x, \xi^N)]$ given by (2.19) represents the mean acceptance probability when the Markov chain x^N stands at x. For our proof it is important to understand how the acceptance probability $\alpha^N(x, \xi^N)$ depends on the current position x and on the source of randomness ξ^N. Recall the quantity Q^N defined in Equation (2.21): the main observation is that $Q^N(x, \xi^N)$ can be approximated by a Gaussian random variable

$$Q^N(x, \xi^N) \;\approx\; Z_\ell \tag{3.1}$$

where $Z_\ell \stackrel{\mathcal{D}}{\sim} \mathrm{N}(-\frac{\ell^3}{4}, \frac{\ell^3}{2})$. These approximations are made rigorous in Lemma 4.4 and Lemma 4.5. Therefore, the Bernoulli random variable $\gamma^N(x, \xi^N)$ with success probability $1 \wedge e^{Q^N(x,\xi^N)}$ can be approximated by a Bernoulli random variable, independent of x, with success probability equal to

$$\alpha(\ell) = \mathbb{E}\big[1 \wedge e^{Z_\ell}\big]. \tag{3.2}$$

Thus, the limiting acceptance probability of the MALA algorithm is as given in Equation (3.2). Recall that $\Delta t = N^{-\frac{1}{3}}$. With this notation we introduce the drift function $d^N : \mathcal{H}^s \to \mathcal{H}^s$ given by

$$d^N(x) = \big(h(\ell)\Delta t\big)^{-1}\mathbb{E}\big[x^{1,N} - x^{0,N} \,|x^{0,N} = x\big] \tag{3.3}$$

and the martingale difference array $\{\Gamma^{k,N} : k \geq 0\}$ defined by $\Gamma^{k,N} = \Gamma^N(x^{k,N}, \xi^{k,N})$ with

$$\Gamma^{k,N} = \big(2h(\ell)\Delta t\big)^{-\frac{1}{2}}\Big(x^{k+1,N} - x^{k,N} - h(\ell)\Delta t\, d^N(x^{k,N})\Big). \tag{3.4}$$

The normalization constant $h(\ell)$ defined in Equation (2.22) ensures that the drift function d^N and the martingale difference array $\{\Gamma^{k,N}\}$ are asymptotically independent from the parameter ℓ. The drift-martingale decomposition of the Markov chain $\{x^{k,N}\}_k$ then reads

$$x^{k+1,N} - x^{k,N} = h(\ell)\Delta t\, d^N(x^{k,N}) + \sqrt{2h(\ell)\Delta t}\; \Gamma^{k,N}. \tag{3.5}$$

Lemma 4.7 and Lemma 4.8 exploit the Gaussian behavior of $Q^N(x, \xi^N)$ described in Equation (3.1) in order to give quantitative versions of the following approximations,

$$d^N(x) \;\approx\; \mu(x) \qquad \text{and} \qquad \Gamma^{k,N} \;\approx\; \mathrm{N}(0, \mathcal{C}) \tag{3.6}$$

where the function $\mu(\cdot)$ is defined by Equation (2.14). From Equation (3.5) it follows that for large N the evolution of the Markov chain ressembles the Euler discretization of the limiting diffusion (2.23). The next step consists of proving an invariance principle for a rescaled version of the martingale difference array $\{\Gamma^{k,N}\}$. The continuous process $W^N \in C([0;T], \mathcal{H}^s)$ is defined as

$$W^N(t) = \sqrt{\Delta t}\,\sum_{j=0}^{k} \Gamma^{j,N} \;+\; \frac{t - k\Delta t}{\sqrt{\Delta t}}\,\Gamma^{k+1,N} \qquad \text{for} \qquad k\Delta t \leq t < (k+1)\Delta t. \tag{3.7}$$

The sequence of processes $\{W^N\}_{N\geq 1}$ converges weakly as $N \to \infty$ in $C([0;T], \mathcal{H}^s)$ to a Brownian motion W in $\mathcal{H}^s$ with covariance operator equal to $\mathcal{C}_s$. Indeed, Proposition 4.10 proves the stronger result

$$(x^{0,N}, W^N) \Longrightarrow (z^0, W)$$

where $\Longrightarrow$ denotes weak convergence in $\mathcal{H}^s \times C([0;T], \mathcal{H}^s)$ and $z^0 \overset{\mathcal{D}}{\sim} \pi$ is independent of the limiting Brownian motion W. Using this invariance principle and the fact that the noise process is additive (the diffusion coefficient of the SPDE (2.23) is constant), the main theorem follows from a continuous mapping argument which we now outline. For any $W \in C([0,T]; \mathcal{H}^s)$ we define the Itô map

$$\Theta\colon \mathcal{H}^s \times C([0,T]; \mathcal{H}^s) \to C([0,T]; \mathcal{H}^s)$$

which maps (z^0, W) to the unique solution of the integral equation

$$z(t) = z^0 - h(\ell)\int_0^t \mu(z)\,du + \sqrt{2h(\ell)}\,W(t) \qquad \forall t \in [0,T]. \tag{3.8}$$

Notice that $z = \Theta(z^0, W)$ solves the SPDE (2.23). The Itô map Θ is continuous, essentially because the noise in (2.23) is additive (does not depend on the state z). The piecewise constant interpolant $\bar{z}^N$ of x^N is defined by

$$\bar{z}^N(t) = x^k \qquad \text{for} \qquad k\Delta t \leq t < (k+1)\Delta t. \tag{3.9}$$

Using this definition it follows that the continuous piecewise linear interpolant z^N defined in Equation (1.4) satisfies

$$z^N(t) = x^{0,N} - h(\ell)\int_0^t d^N(\bar{z}^N(u))\,du + \sqrt{2h(\ell)}\,W^N(t) \qquad \forall t \in [0,T]. \tag{3.10}$$

Using the closeness of $d^N(\cdot)$ and $\mu(\cdot)$, and of z^N and $\bar{z}^N$, we will see that there exists a process $\widehat{W}^N \Rightarrow W$ as $N \to \infty$ such that

$$z^N(t) = x^{0,N} - h(\ell)\int_0^t \mu(z^N(u))\,du + \sqrt{2h(\ell)}\,\widehat{W}^N(t).$$

Thus we may write $z^N = \Theta(x^{0,N}, \widehat{W}^N)$. By continuity of the Itô map Θ, it follows from the continuous mapping theorem that $z^N = \Theta(x^{0,N}, \widehat{W}^N) \Longrightarrow \Theta(z^0, W) = z$ as N goes to infinity. This weak convergence result is the principal result of this article.

3.2. *General Diffusion Approximation.* In this section we state and prove a proposition containing a general diffusion approximation result. Using this, we then prove our Main Theorem in section 3.3. To this end, consider a general sequence of Markov chains $x^N = \{x^{k,N}\}_{k \geq 0}$ evolving at stationarity in the separable Hilbert space $\mathcal{H}^s$ and introduce the drift-martingale decomposition

$$x^{k+1,N} - x^{k,N} = h(\ell)\, d^N(x_k)\,\Delta t \;+\; \sqrt{2h(\ell)\Delta t}\;\Gamma^{k,N} \tag{3.11}$$

where $h(\ell) > 0$ is a constant parameter and Δt is a time-step decreasing to 0 as N goes to infinity. Here d^N and $\Gamma^{k,N}$ are as defined above. We introduce the rescaled process $W^N(t)$ as in (3.7). The main diffusion approximation result is the following.

Proposition 3.1. **(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 $x^N = \{x^{k,N}\}_{k\geq 0}$ with invariant distribution π^N. Suppose that the Markov chains start at stationarity $x^{0,N} \overset{\mathcal{D}}{\sim} \pi^N$ and that the drift-martingale decomposition (3.11) satisfies the following assumptions.*

1. **Convergence of initial conditions:** *π^N converges in distribution to the probability measure π where π has a finite first moment, that is $\mathbb{E}^{\pi}[\|x\|_s] < \infty$.*
2. **Invariance principle:** *the sequence $(x^{0,N}, W^N)$ defined by Equation (3.7) converges weakly in $\mathcal{H}^s \times C([0,T],\mathcal{H}^s)$ to (z^0, W) where $z^0 \overset{\mathcal{D}}{\sim} \pi$ and W is a Brownian motion in $\mathcal{H}^s$, independent from z^0, with covariance operator $\mathcal{C}_s$.*
3. **Convergence of the drift:** *there exists a globally Lipschitz function $\mu : \mathcal{H}^s \to \mathcal{H}^s$ that satisfies*

$$\lim_{N\to\infty} \mathbb{E}^{\pi^N}\Big[\|d^N(x) - \mu(x)\|_s\Big] = 0.$$

Then the sequence of rescaled interpolants $z^N \in C([0,T],\mathcal{H}^s)$ defined by Equation (1.4) converges weakly in $C([0,T],\mathcal{H}^s)$ to $z \in C([0,T],\mathcal{H}^s)$ given by

$$\begin{aligned} \frac{dz}{dt} &= h(\ell)\mu(z(t)) \;+\; \sqrt{2h(\ell)}\frac{dW}{dt}, \\ z(0) &\overset{\mathcal{D}}{\sim} \pi. \end{aligned}$$

Here W is a Brownian motion in $\mathcal{H}^s$ with covariance $\mathcal{C}_s$ and initial condition $z^0 \overset{\mathcal{D}}{\sim} \pi$ independent of W.

Proof. Define $\bar{z}^N(t)$ as in (3.9). It then follows that

$$z^N(t) = x^{0,N} + h(\ell)\int_0^t d^N(\bar{z}^N(u))\,du + \sqrt{2h(\ell)}\,W^N(t) \qquad (3.12)$$
$$= z^{0,N} + h(\ell)\int_0^t \mu(z^N(u))\,du + \sqrt{2h(\ell)}\,\widehat{W}^N(t)$$

where the process $W^N \in C([0,T],\mathcal{H}^s)$ is defined by Equation (3.7) and

$$\widehat{W}^N(t) = W^N(t) + \sqrt{\frac{h(\ell)}{2}}\int_0^t [d^N(\bar{z}^N(u)) - \mu(z^N(u))]\,du.$$

Define the Itô map $\Theta\colon \mathcal{H}^s \times C([0,T];\mathcal{H}^s) \to C([0,T];\mathcal{H}^s)$ that maps (z_0, W) to the unique solution $z \in C([0,T],\mathcal{H}^s)$ of the integral equation

$$z(t) = z_0 + h(\ell)\int_0^t \mu(z(u))\,du + \sqrt{2h(\ell)}\,W(t), \qquad \forall t \in [0,T].$$

The Equation (3.12) is thus equivalent to $z^N = \Theta(x^{0,N}, \widehat{W}^N)$. The proof of the diffusion approximation is accomplished through the following steps.

- **The Itô map $\Theta : \mathcal{H}^s \times C([0,T],\mathcal{H}^s) \to C([0,T],\mathcal{H}^s)$ is continuous.**
  This is Lemma 3.7 of [MPS11].
- **The pair $(x^{0,N}, \widehat{W}^N)$ converges weakly to (z^0, W).**
  In a separable Hilbert space, if the sequence $\{a_n\}_{n\in\mathbb{N}}$ converges weakly to a and the sequence $\{b_n\}_{n\in\mathbb{N}}$ converges in probability to 0 then the sequence $\{a_n + b_n\}_{n\in\mathbb{N}}$ converges weakly to a. It is assumed that $(x^{0,N}, W^N)$ converges weakly to (z^0, W) in $\mathcal{H}^s \times C([0,T],\mathcal{H}^s)$. Consequently, to prove that $\widehat{W}^N$ converges weakly to W it suffices to prove that $\int_0^T \|d^N(\bar{z}^N(u)) - \mu(z^N(u))\|_s\,du$ converges in probability to 0. For any time $k\Delta t \le u < (k+1)\Delta t$, the stationarity of the chain shows that

  $$\|d^N(\bar{z}^N(u)) - \mu(\bar{z}^N(u))\|_s = \|d^N(x^{k,N}) - \mu(x^{k,N})\|_s \overset{\mathcal{D}}{\sim} \|d^N(x^{0,N}) - \mu(x^{0,N})\|_s$$
  $$\|\mu(\bar{z}^N(u)) - \mu(z^N(u))\|_s \le \|\mu\|_{\mathrm{Lip}} \cdot \|x^{k+1,N} - x^{k,N}\|_s \overset{\mathcal{D}}{\sim} \|\mu\|_{\mathrm{Lip}} \cdot \|x^{1,N} - x^{0,N}\|_s.$$

  where in the last step we have used the fact that $\|\bar{z}^N(u) - z^N(u)\|_s \le \|x^{k+1,N} - x^{k,N}\|_s$. Consequently,

  $$\mathbb{E}^{\pi^N}\Big[\int_0^T \|d^N(\bar{z}^N(u)) - \mu(z^N(u))\|_s\,du\Big] \le T\cdot\mathbb{E}^{\pi^N}[\|d^N(x^{0,N}) - \mu(x^{0,N})\|_s] + T\cdot\|\mu\|_{\mathrm{Lip}}\cdot\mathbb{E}^{\pi^N}[\|x^{1,N} - x^{0,N}\|_s].$$

  The first term goes to zero since it is assumed that $\lim_N \mathbb{E}^{\pi^N}\Big[\|d^N(x) - \mu(x)\|_s\Big] = 0$. Since $\mathrm{Tr}_{\mathcal{H}^s}(\mathcal{C}_s) < \infty$, the second term is of order $\mathcal{O}(\sqrt{\Delta t})$ and thus also converges to 0. Therefore $\widehat{W}^N$ converges weakly to W, hence the conclusion.
- **Continuous mapping argument.**
  We have proved that $(x^{0,N}, \widehat{W}^N)$ converges weakly in $\mathcal{H}^s \times C([0,T],\mathcal{H}^s)$ to (z^0, W) and the Itô map $\Theta\colon \mathcal{H}^s \times C([0,T],\mathcal{H}^s) \to C([0,T],\mathcal{H}^s)$ is a continuous function. The continuous mapping theorem thus shows that $z^N = \Theta(x^{0,N}, \widehat{W}^N)$ converges weakly to $z = \Theta(z^0, W)$, finishing the proof of Proposition 3.1.

□

3.3. *Proof of Main Theorem.* We now prove Theorem 2.6. The proof consists in checking that the conditions needed for Proposition 3.1 to apply are satisfied by the sequence of MALA Markov chains (2.20). The key estimates are proved later in section 4.

1. By Lemma 4.3 the sequence of probability measures π^N converges weakly in $\mathcal{H}^s$ to π.
2. Proposition 4.10 proves that $(x^{0,N}, W^N)$ converges weakly in $\mathcal{H} \times C([0,T], \mathcal{H}^s)$ to (z^0, W), where W is a Brownian motion with covariance $\mathcal{C}_s$ independent from $z^0 \overset{\mathcal{D}}{\sim} \pi$.
3. Lemma 4.7 states that $d^N(x)$ defined by Equation (3.3) satisfies $\lim_N \mathbb{E}^{\pi^N}[\|d^N(x) - \mu(x)\|_s^2] = 0$ and Proposition 2.4 shows that $\mu : \mathcal{H}^s \to \mathcal{H}^s$ is a Lipschitz function.

The three assumptions needed for Lemma 3.1 to apply are satisfied, which concludes the proof of Theorem 2.6.

4. Key Estimates.

Subsection 4.1 contains some technical lemmas of use throughout. In section 4.2 we study the large N Gaussian approximation of the acceptance probability, simultanesouly establishing asymptotic independence of the current state of the Markov chain. This approximation is then used in subsections 4.3 and 4.4 to give quantitative versions of the heuristics (3.6). The section concludes with subsection 4.5 in which we prove an invariance principle for W^N given by (3.7).

4.1. *Technical Lemmas.* The first Lemma shows that, for π_0-almost every function $x \in \mathcal{H}^s$, the approximation $\mu^N(x) \approx \mu(x)$ holds as N goes to infinity.

LEMMA 4.1. **(μ^N converges π_0-almost surely to μ)** *Let Assumption 2.1 hold. The sequences of functions $\mu^N : \mathcal{H}^s \to \mathcal{H}^s$ satisfies*

$$\pi_0\Big(\Big\{x \in \mathcal{H}^s : \lim_{N\to\infty} \|\mu^N(x) - \mu(x)\|_s = 0\Big\}\Big) = 1.$$

PROOF. It is enough to verify that for $x \in \mathcal{H}^s$ we have

$$\lim_{N\to\infty} \|P^N x - x\|_s = 0 \tag{4.1}$$

$$\lim_{N\to\infty} \|\mathcal{C}P^N \nabla\Psi(P^N x) - \mathcal{C}\nabla\Psi(x)\|_s = 0. \tag{4.2}$$

- Let us prove Equation (4.1). For $x \in \mathcal{H}^s$ we have $\sum_{j\geq 1} j^{2s} x_j^2 < \infty$ so that

$$\lim_{N\to\infty} \|P^N x - x\|_s^2 \;=\; \lim_{N\to\infty} \sum_{j=N+1}^{\infty} j^{2s} x_j^2 = 0. \tag{4.3}$$

- Let us prove (4.2). The triangle inequality shows that

$$\|\mathcal{C}P^N\nabla\Psi(P^N x) - \mathcal{C}\nabla\Psi(x)\|_s \leq \|\mathcal{C}P^N\nabla\Psi(P^N x) - \mathcal{C}P^N\nabla\Psi(x)\|_s + \|\mathcal{C}P^N\nabla\Psi(x) - \mathcal{C}\nabla\Psi(x)\|_s$$

  The same proof as Lemma 2.4 reveals that $\mathcal{C}P^N\nabla\Psi : \mathcal{H}^s \to \mathcal{H}^s$ is globally Lipschitz, with a Lipschitz constant that can be chosen independent from N. Consequenly, Equation (4.3) shows that

$$\|\mathcal{C}P^N\nabla\Psi(P^N x) - \mathcal{C}P^N\nabla\Psi(x)\|_s \lesssim \|P^N x - x\|_s \to 0.$$

Also, $z = \nabla\Psi(x) \in \mathcal{H}^{-s}$ so that $\|\nabla\Psi(x)\|_{-s}^2 = \sum_{j\geq 1} j^{-2s} z_j^2 < \infty$. The eigenvalues of $\mathcal{C}$ satisfy $\lambda_j^2 \asymp j^{-2\kappa}$ with $s < \kappa - \frac{1}{2}$. Consequently,

$$\begin{aligned}\|\mathcal{C}P^N\nabla\Psi(x) - \mathcal{C}\nabla\Psi(x)\|_s^2 &= \sum_{j=N+1}^{\infty} j^{2s}(\lambda_j^2 z_j)^2 \lesssim \sum_{j=N+1}^{\infty} j^{2s-4\kappa} z_j^2 \\ &= \sum_{j=N+1}^{\infty} j^{4(s-\kappa)} j^{-2s} z_j^2 \leq \frac{1}{(N+1)^{4(\kappa-s)}} \|\nabla\Psi(x)\|_{-s}^2 \;\to\; 0.\end{aligned}$$

□

Next lemma shows that the size of the jump $y - x$ is of order $\sqrt{\Delta t}$.

LEMMA 4.2. *Consider y given by (2.17). Under Assumptions 2.1, for any $p \geq 1$ we have*

$$\mathbb{E}_x^{\pi^N}\big[\|y - x\|_s^p\big] \lesssim (\Delta t)^{\frac{p}{2}} \cdot (1 + \|x\|_s^p).$$

PROOF. Under Assumption 2.1 the function μ^N is globally Lipschitz on $\mathcal{H}^s$, with Lipschitz constant that can be chosen independent from N. Thus

$$\|y - x\|_s \lesssim \Delta t(1 + \|x\|_s) + \sqrt{\Delta t}\, \|\mathcal{C}^{\frac{1}{2}}\xi^N\|_s.$$

We have $\mathbb{E}^{\pi^0}\Big[\|\mathcal{C}^{\frac{1}{2}}\xi^N\|_s^p\Big] \leq \mathbb{E}^{\pi^0}\Big[\|\zeta\|_s^p\Big] < \infty$, where $\zeta \overset{\mathcal{D}}{\sim} \mathrm{N}(0, \mathcal{C})$. From Fernique's theorem [DPZ92] it follows that $\mathbb{E}^{\pi^0}\Big[\|\zeta\|_s^p\Big] < \infty$. Consequently, $\mathbb{E}^{\pi^0}\Big[\|\mathcal{C}^{\frac{1}{2}}\xi^N\|_s^p\Big]$ is uniformly bounded as a function of N, proving the lemma. □

The normalizing constants M_{Ψ^N} are uniformly bounded and we use this fact to obtain uniform bounds on moments of functionals in $\mathcal{H}$ under π^N. Moreover, we prove that the sequence of probability measures π^N on $\mathcal{H}^s$ converges weakly in $\mathcal{H}^s$ to π.

LEMMA 4.3. **(Finite dimensional approximation π^N of π)** *Under the Assumptions 2.1 the normalization constants M_{Ψ^N} are uniformly bounded so that for any measurable functional $f : \mathcal{H} \mapsto \mathbb{R}$, we have*

$$\mathbb{E}^{\pi^N}\big[|f(x)|\big] \lesssim \mathbb{E}^{\pi_0}\big[|f(x)|\big].$$

Moreover, the sequence of probability measure π^N satisfies

$$\pi^N \implies \pi$$

where $\implies$ denotes weak convergence in $\mathcal{H}^s$.

PROOF. The first part is contained in Lemma 3.5 of [MPS11]. Let us prove that $\pi^N \implies \pi$. We need to show that for any bounded continuous function $g : \mathcal{H}^s \to \mathbb{R}$ we have $\lim_{N\to\infty} \mathbb{E}^{\pi^N}[g(x)] = \mathbb{E}^{\pi}[g(x)]$ where

$$\begin{aligned}\mathbb{E}^{\pi^N}[g(x)] &= \mathbb{E}^{\pi_0^N}[g(x) M_{\Psi^N} e^{-\Psi^N(x)}] \\ &= \mathbb{E}^{\pi_0}[g(P^N x) M_{\Psi^N} e^{-\Psi(P^N x)}].\end{aligned}$$

Since g is bounded, Ψ is lower bounded and since the normalization constants are uniformly bounded, the dominated convergence theorem shows that it suffices to show that $g(P^N x) M_{\Psi^N} e^{-\Psi(P^N x)}$ converges π_0-almost surely to $g(x) M_\Psi e^{-\Psi(x)}$. For this in turn it suffices to show that $\Psi(P^N\, x)$ converges π_0-almost surely to $\Psi(x)$ as this also proves almost sure convergence of the normalization constants. By (2.7) we have

$$|\Psi(P^N\, x) - \Psi(x)| \lesssim (1 + \|x\|_s + \|P^N x\|_s) \|P^N\, x - x\|_s.$$

But $\lim_{N \to \infty}\ \|P^N\ x - x\|_s \to 0$ for any $x \in \mathcal{H}^s$, by dominated convergence, and the result follows. □

Fernique's theorem [DPZ92] states that for any exponent $p \geq 0$ we have $\mathbb{E}^{\pi^0}[\|x\|_s^p] < \infty$. It thus follows from Lemma 4.3 that for any $p \geq 0$

$$\sup_N \left\{ \mathbb{E}^{\pi^N}[\|x\|_s^p] : N \in \mathbb{N} \right\} \ < \ \infty.$$

This estimate is repeatedly used in the sequel.

4.2. *Gaussian approximation of Q^N.* Recall the quantity Q^N defined in Equation (2.21). This section proves that Q^N has a Gaussian behavior in the sense that

$$Q^N(x, \xi^N) = Z^N(x, \xi^N) \ + \ i^N(x, \xi^N) \ + \ \mathbf{e}^N(x, \xi^N) \tag{4.4}$$

where the quantities Z^N and i^N are equal to

$$Z^N(x, \xi^N) = -\frac{\ell^3}{4} - \frac{\ell^{\frac{3}{2}}}{\sqrt{2}} N^{-\frac{1}{2}} \sum_{j=1}^N \lambda_j^{-1} \xi_j x_j \tag{4.5}$$

$$i^N(x, \xi^N) = \frac{1}{2} (\ell \Delta t)^2 \Big(\|x\|^2_{\mathcal{C}^N} - \|(\mathcal{C}^N)^{\frac{1}{2}} \xi^N\|^2_{\mathcal{C}^N} \Big) \tag{4.6}$$

with i^N and e^N small. Thus the principal contributions to Q^N comes from the random variable $Z^N(x, \xi^N)$. Notice that, for each fixed $x \in \mathcal{H}^s$, the random variable $Z^N(x, \xi^N)$ is Gaussian. Furthermore, the Karhunen-Loève expansion of π_0 shows that for π_0-almost every choice of function $x \in \mathcal{H}$ the sequence $\{Z^N(x, \xi^N)\}_{N \geq 1}$ converges in law to the distribution of $Z_\ell \overset{\mathcal{D}}{\sim} \mathrm{N}(-\frac{\ell^3}{4}, \frac{\ell^3}{2})$. The next lemma rigorously bounds the error terms $\mathbf{e}^N(x, \xi^N)$ and $i^N(x, \xi^N)$: we show that i^N is an error term of order $\mathcal{O}(N^{-\frac{1}{6}})$ and $\mathbf{e}^N(x, \xi)$ is an error term of order $\mathcal{O}(N^{-\frac{1}{3}})$. In Lemma 4.5 we then quantify the convergence of $Z^N(x, \xi^N)$ to Z_ℓ.

LEMMA 4.4. **(Gaussian Approximation)** *Let $p \geq 1$ be an integer. Under Assumptions 2.1 the error terms i^N and $\mathbf{e}^N$ in the Gaussian approximation (4.4) satisfy*

$$\Big(\mathbb{E}^{\pi^N}[|i^N(x, \xi^N)|^p] \Big)^{\frac{1}{p}} = \mathcal{O}(N^{-\frac{1}{6}}) \qquad \textit{and} \qquad \Big(\mathbb{E}^{\pi^N}[|\mathbf{e}^N(x, \xi^N)|^p] \Big)^{\frac{1}{p}} = \mathcal{O}(N^{-\frac{1}{3}}). \tag{4.7}$$

PROOF. For notational clarity, without loss of generality, we suppose $p = 2q$. The quantity Q^N is defined in Equation (2.21) and expanding terms leads to

$$Q^N(x, \xi^N) \ = \ I_1 \ + \ I_2 \ + \ I_3$$

where the quantities I_1, I_2 and I_3 are given by

$$I_1 = -\frac{1}{2}(\|y\|_{\mathcal{C}^N}^2 - \|x\|_{\mathcal{C}^N}^2) - \frac{1}{4\ell\Delta t}(\|x - y(1-\ell\Delta t)\|_{\mathcal{C}^N}^2 - \|y - x(1-\ell\Delta t)\|_{\mathcal{C}^N}^2)$$
$$I_2 = -\Big(\Psi^N(y) - \Psi^N(x)\Big) - \frac{1}{2}\Big(\langle x - y(1-\ell\Delta t), \mathcal{C}^N\nabla\Psi^N(y)\rangle_{\mathcal{C}^N} - \langle y - x(1-\ell\Delta t), \mathcal{C}^N\nabla\Psi^N(x)\rangle_{\mathcal{C}^N}\Big)$$
$$I_3 = -\frac{\ell\Delta t}{4}\Big\{\|\mathcal{C}^N\nabla\Psi^N(y)\|_{\mathcal{C}^N}^2 - \|\mathcal{C}^N\nabla\Psi^N(x)\|_{\mathcal{C}^N}^2\Big\}.$$

The term I_1 arises purely from the Gaussian part of the target measure π^N and from the Gaussian part of the proposal. The two other terms I_2 and I_3 come from the change of probability involving the functional Ψ^N. We start by simplifyng the expression for I_1, and then return to estimate the terms I_2 and I_3.

$$\begin{aligned} I_1 &= -\frac{1}{2}(\|y\|_{\mathcal{C}^N}^2 - \|x\|_{\mathcal{C}^N}^2) - \frac{1}{4\ell\Delta t}(\|(x-y) + \ell\Delta t\, y)\|_{\mathcal{C}^N}^2 - \|(y-x) + \ell\Delta t\, x)\|_{\mathcal{C}^N}^2) \\ &= -\frac{1}{2}(\|y\|_{\mathcal{C}^N}^2 - \|x\|_{\mathcal{C}^N}^2) - \frac{1}{4\ell\Delta t}(2\ell\Delta t[\|x\|_{\mathcal{C}^N}^2 - \|y\|_{\mathcal{C}^N}^2] + (\ell\Delta t)^2[\|y\|_{\mathcal{C}^N}^2 - \|x\|_{\mathcal{C}^N}^2]) \\ &= -\frac{\ell\Delta t}{4}\Big(\|y\|_{\mathcal{C}^N}^2 - \|x\|_{\mathcal{C}^N}^2\Big). \end{aligned}$$

The term I_1 is $\mathcal{O}(1)$ and constitutes the main contribution to Q^N. Before analyzing I_1 in more detail, we show that I_2 and I_3 are $\mathcal{O}(N^{-\frac{1}{3}})$:

$$\Big(\mathbb{E}^{\pi^N}[I_2^{2q}]\Big)^{\frac{1}{2q}} = \mathcal{O}(N^{-\frac{1}{3}}) \qquad \text{and} \qquad \Big(\mathbb{E}^{\pi^N}[I_3^{2q}]\Big)^{\frac{1}{2q}} = \mathcal{O}(N^{-\frac{1}{3}}). \tag{4.8}$$

- We expand I_2 and use the bound on the remainder of the Taylor expansion of Ψ described in Equation (2.15),

$$\begin{aligned} I_2 &= -\Big\{\Psi^N(y) - [\Psi^N(x) + \langle\nabla\Psi^N(x), y-x\rangle]\Big\} + \frac{1}{2}\langle y-x, \nabla\Psi^N(y) - \nabla\Psi^N(x)\rangle \\ &\quad + \frac{\ell\Delta t}{2}\Big\{\langle x, \nabla\Psi^N(x)\rangle - \langle y, \nabla\Psi^N(y)\rangle\Big\} \\ &= A_1 + A_2 + A_3. \end{aligned}$$

Equation (2.15) and Lemma 4.2 show that

$$\mathbb{E}^{\pi^N}[A_1^{2q}] \lesssim \mathbb{E}^{\pi^N}[\|y-x\|_s^{4q}] \lesssim (\Delta t)^{2q}\,\mathbb{E}^{\pi^N}[1 + \|x\|_s^{4q}] \lesssim (\Delta t)^{2q} = \Big(N^{-\frac{1}{3}}\Big)^{2q},$$

where we have used the fact that $\mathbb{E}^{\pi^N}[\|x\|_s^{4q}] \lesssim \mathbb{E}^{\pi_0}[\|x\|_s^{4q}] < \infty$. Assumption 2.1 states that $\partial^2\Psi$ is uniformly bounded in $\mathcal{L}(\mathcal{H}^s, \mathcal{H}^{-s})$ so that

$$\begin{aligned} \|\nabla\Psi(y) - \nabla\Psi(y)\|_{-s} &= \Big\|\int_0^1 \partial^2\Psi(x + t(y-x))\cdot(y-x)\,dt\Big\|_{-s} \\ &\le \int_0^1 \|\partial^2\Psi(x + t(y-x))\cdot(y-x)\|_{-s}\,dt \\ &\le M_4\int_0^1 \|y-x\|_s\,dt. \end{aligned} \tag{4.9}$$

This proves that $\|\nabla\Psi^N(y) - \nabla\Psi^N(x)\|_{-s} \lesssim \|y-x\|_s$. Consequently, Lemma 4.2 shows that

$$\begin{aligned}
\mathbb{E}^{\pi^N}[A_2^{2q}] &\lesssim \mathbb{E}^{\pi^N}\Big[\|y-x\|_s^{2q} \cdot \|\nabla\Psi^N(y) - \nabla\Psi^N(x)\|_{-s}^{2q}\Big] \\
&\lesssim \mathbb{E}^{\pi^N}\Big[\|y-x\|_s^{4q}\Big] \\
&\lesssim (\Delta t)^{2q}\, \mathbb{E}^{\pi^N}\Big[1 + \|x\|_s^{4q}\Big] \\
&\lesssim (\Delta t)^2 \;=\; \Big(N^{-\frac{1}{3}}\Big)^{2q}.
\end{aligned}$$

Under Assumptions 2.1, for any $z \in \mathcal{H}^s$ we have $\|\nabla\Psi^N(z)\|_{-s} \lesssim 1 + \|z\|_s$. Therefore $\mathbb{E}^{\pi^N}[A_3^{2q}] \lesssim (\Delta t)^{2q}$. Putting these estimates together,

$$\Big(\mathbb{E}^{\pi^N}[I_2^{2q}]\Big)^{\frac{1}{2q}} \;\lesssim\; \Big(\mathbb{E}^{\pi^N}[A_1^{2q} + A_2^{2q} + A_3^{2q}]\Big)^{\frac{1}{2q}} \;=\; \mathcal{O}(N^{-\frac{1}{3}}).$$

- Lemma 2.4 states $\mathcal{C}^N\nabla\Psi^N : \mathcal{H}^s \to \mathcal{H}^s$ is globally Lipschitz, with a Lipschitz constant that can be chosen uniformly in N. Therefore,

$$\|\mathcal{C}^N\nabla\Psi^N(z)\|_s \lesssim 1 + \|z\|_s. \tag{4.10}$$

Since $\|\mathcal{C}^N\nabla\Psi^N(z)\|_{\mathcal{C}^N}^2 = \langle \nabla\Psi^N(z), \mathcal{C}^N\nabla\Psi^N(z)\rangle$, the bound (2.7) gives

$$\begin{aligned}
\mathbb{E}^{\pi^N}[I_3^{2q}] &\lesssim \Delta t^{2q}\, \mathbb{E}\Big[\langle \nabla\Psi^N(x), \mathcal{C}^N\nabla\Psi^N(x)\rangle^q + \langle \nabla\Psi^N(y), \mathcal{C}^N\nabla\Psi^N(y)\rangle^q\Big] \\
&\lesssim \Delta t^{2q}\, \mathbb{E}^{\pi^N}\Big[(1+\|x\|_s)^{2q} + (1+\|y\|_s)^{2q}\Big] \\
&\lesssim \Delta t^{2q}\, \mathbb{E}^{\pi^N}\Big[1 + \|x\|_s^{2q} + \|y\|_s^{2q}\Big] \;\lesssim\; \Delta t^{2q} \;=\; \Big(N^{-\frac{1}{3}}\Big)^{2q},
\end{aligned}$$

which concludes the proof of Equation (4.8).

We now simplify further the expression for I_1 and demonstrate that it has a Gaussian behaviour. We use the definition of the proposal y given in Equation (2.17) to expand I_1. For $x \in X^N$ we have $P^N x = x$. Therefore, for $x \in X^N$,

$$\begin{aligned}
I_1 &= -\frac{\ell\Delta t}{4}\Big(\|(1-\ell\Delta t)x - \ell\Delta t\, \mathcal{C}^N\nabla\Psi^N(x) + \sqrt{2\ell\Delta t}\,(\mathcal{C}^N)^{\frac{1}{2}}\xi^N\|_{\mathcal{C}^N}^2 - \|x\|_{\mathcal{C}^N}^2\Big) \\
&= Z^N(x,\xi^N) \;+\; i^N(x,\xi^N) \;+\; B_1 \;+\; B_2 \;+\; B_3 \;+\; B_4.
\end{aligned}$$

with $Z^N(x,\xi^N)$ and $i^N(x,\xi^N)$ given by Equation (4.5) and (4.6) and

$$B_1 = \tfrac{\ell^3}{4}\big(1 - \tfrac{\|x\|_{\mathcal{C}^N}^2}{N}\big) \qquad B_2 = -\tfrac{\ell^3}{4}N^{-1}\Big\{\|\mathcal{C}^N\nabla\Psi^N(x)\|_{\mathcal{C}^N}^2 + 2\langle x, \nabla\Psi^N(x)\rangle\Big\}$$

$$B_3 = \tfrac{\ell^{\frac{5}{2}}}{\sqrt{2}}N^{-\frac{5}{6}}\langle x + \mathcal{C}^N\nabla\Psi^N(x), (\mathcal{C}^N)^{\frac{1}{2}}\xi^N\rangle_{\mathcal{C}^N} \qquad B_4 = \tfrac{\ell^2}{2}N^{-\frac{2}{3}}\langle x, \nabla\Psi^N(x)\rangle.$$

The quantity Z^N is the leading term. For each fixed value of $x \in \mathcal{H}^s$ the term $Z^N(x,\xi^N)$ is Gaussian. Below, we prove that quantity i^N is $\mathcal{O}(N^{-\frac{1}{6}})$. We now establish that each B_j is $\mathcal{O}(N^{-\frac{1}{3}})$,

$$\Big(\mathbb{E}^{\pi^N}[B_j^{2q}]\Big)^{\frac{1}{2q}} \;=\; \mathcal{O}(N^{-\frac{1}{3}}) \qquad j = 1,\ldots,4. \tag{4.11}$$

- Lemma 4.3 shows that $\mathbb{E}^{\pi^N}[(1-\frac{\|x\|^2_{\mathcal{C}^N}}{N})^{2q}] \lesssim \mathbb{E}^{\pi_0}[(1-\frac{\|x\|^2_{\mathcal{C}^N}}{N})^{2q}]$. Under π_0,

$$\frac{\|x\|^2_{\mathcal{C}^N}}{N} \quad \overset{\mathcal{D}}{\sim} \quad \frac{\rho_1^2+\ldots+\rho_N^2}{N}.$$

  where $\rho_1,\ldots,\rho_N$ are i.i.d N(0,1) Gaussian random variables. Consequently, $\mathbb{E}^{\pi^N}[B_1^{2q}]^{\frac{1}{2q}} = \mathcal{O}(N^{-\frac{1}{2}})$.

- The term $\|\mathcal{C}^N\nabla\Psi^N(x)\|^{2q}_{\mathcal{C}^N}$ has already been bounded while proving $\mathbb{E}^{\pi^N}[I_3^{2q}] \lesssim \Big(N^{-\frac{1}{3}}\Big)^{2q}$. Equation (2.7) gives the bound $\|\nabla\Psi^N(x)\|_{-s} \lesssim 1+\|x\|_s$ and shows that $\mathbb{E}^{\pi^N}[\langle x,\nabla\Psi^N(x)\rangle^{2q}]$ is uniformly bounded as a function of N. Consequently,

$$\mathbb{E}^{\pi^N}[B_2^{2q}]^{\frac{1}{2q}} \;=\; \mathcal{O}(N^{-1}).$$

- We have $\langle \mathcal{C}^N\nabla\Psi^N(x),(\mathcal{C}^N)^{\frac{1}{2}}\xi^N\rangle_{\mathcal{C}^N} = \langle\nabla\Psi^N(x),(\mathcal{C}^N)^{\frac{1}{2}}\xi^N\rangle$ so that

$$\mathbb{E}^{\pi^N}[\langle \mathcal{C}^N\nabla\Psi^N(x),(\mathcal{C}^N)^{\frac{1}{2}}\xi^N\rangle^{2q}_{\mathcal{C}^N}] \;\lesssim\; \mathbb{E}^{\pi^N}[\|\nabla\Psi^N(x)\|^{2q}_{-s}\cdot\|(\mathcal{C}^N)^{\frac{1}{2}}\xi^N\|^{2q}_s] \;\lesssim\; 1.$$

  By Lemma 4.3, one can suppose $x \overset{\mathcal{D}}{\sim} \pi_0$,

$$\langle x,(\mathcal{C}^N)^{\frac{1}{2}}\xi^N\rangle_{\mathcal{C}^N} \;\overset{\mathcal{D}}{\sim}\; \sum_{j=1}^N \rho_j\xi_j.$$

  where $\rho_1,\ldots,\rho_N$ are i.i.d N(0,1) Gaussian random variables. Consequently $\Big(\mathbb{E}^{\pi^N}[\langle x,(\mathcal{C}^N)^{\frac{1}{2}}\xi^N\rangle^{2q}_{\mathcal{C}^N}]\Big)^{\frac{1}{2q}} = \mathcal{O}(N^{\frac{1}{2}})$, which proves that

$$\Big(\mathbb{E}^{\pi^N}[B_3^{2q}]\Big)^{\frac{1}{2q}} \;=\; \mathcal{O}(N^{-\frac{5}{6}+\frac{1}{2}}) \;=\; \mathcal{O}(N^{-\frac{1}{3}}).$$

- The bound $\|\nabla\Psi^N(x)\|_{-s} \lesssim 1+\|x\|_s$ ensures that $\Big(\mathbb{E}^{\pi^N}[B_4^{2q}]\Big)^{\frac{1}{2q}} = \mathcal{O}(N^{-\frac{2}{3}})$.

Define the quantity $\mathbf{e}^N(x,\xi^N) = I_2+I_3+B_1+B_2+B_3+B_4$ so that Q^N can also be expressed as

$$Q^N(x,\xi^N) \;=\; Z^N(x,\xi^N)+i^N(x,\xi^N)+\mathbf{e}^N(x,\xi^N).$$

Equations (4.8) and (4.11) show that $\mathbf{e}^N$ satisfies

$$\Big(\mathbb{E}^{\pi^N}[\mathbf{e}^N(x,\xi^N)^{2q}]\Big)^{\frac{1}{2q}} \;=\; \mathcal{O}(N^{-\frac{1}{3}}).$$

We now prove that i^N is $\mathcal{O}(N^{-\frac{1}{6}})$. By Lemma 4.3, $\mathbb{E}^{\pi^N}[i^N(x,\xi^N)^{2q}] \lesssim \mathbb{E}^{\pi_0}[i^N(x,\xi^N)^{2q}]$. If $x \overset{\mathcal{D}}{\sim} \pi_0$ we have

$$\begin{aligned} i^N(x,\xi^N) &= \frac{\ell^2}{2}N^{-\frac{2}{3}}\Big\{\|x\|^2_{\mathcal{C}^N} - \|(\mathcal{C}^N)^{\frac{1}{2}}\xi^N\|^2_{\mathcal{C}^N}\Big\} \\ &= \frac{\ell^2}{2}N^{-\frac{2}{3}}\,\sum_{j=1}^N(\rho_j^2-\xi_j^2). \end{aligned}$$

where $\rho_1, \ldots, \rho_N$ are i.i.d $\mathrm{N}(0,1)$ Gaussian random variables. Since $\mathbb{E}\big[\{ \sum_{j=1}^N (\rho_j^2 - \xi_j^2)\}^{2q}\big] \lesssim N^q$ it follows that

$$\Big(\mathbb{E}^{\pi^N}\big[i^N(x,\xi^N)^{2q}\big]\Big)^{\frac{1}{2q}} \;=\; \mathcal{O}(N^{-\frac{2}{3}+\frac{1}{2}}) = \mathcal{O}(N^{-\frac{1}{6}}), \tag{4.12}$$

which ends the proof of Lemma 4.4 □

The next Lemma quantifies the fact that $Z^N(x,\xi^N)$ is asymptotically independent from the current position x.

Lemma 4.5. **(Asymptotic independence)** *Let $p \geq 1$ be a positive integer and $f : \mathbb{R} \to \mathbb{R}$ be a 1-Lipschitz function. Consider error terms $\mathbf{e}_\star^N(x,\xi)$ satisfying*

$$\lim_{N\to\infty} \mathbb{E}^{\pi^N}[\mathbf{e}_\star^N(x,\xi^N)^p] = 0.$$

Define the functions $\bar{f}^N : \mathbb{R} \to \mathbb{R}$ and the constant $\bar{f} \in \mathbb{R}$ by

$$\bar{f}^N(x) = \mathbb{E}_x\Big[f\big(Z^N(x,\xi^N) + \mathbf{e}_\star^N(x,\xi^N)\big)\Big] \qquad \text{and} \qquad \bar{f} = \mathbb{E}[f(Z_\ell)].$$

Then the function f^N is highly concentrated around its mean in the sense that

$$\lim_{N\to\infty} \mathbb{E}^{\pi^N}\Big[|\bar{f}^N(x) - \bar{f}|^p\Big] = 0.$$

Proof. Let f be a 1-Lipschitz function. Define the function $F : \mathbb{R} \times [0;\infty) \to \mathbb{R}$ by

$$F(\mu,\sigma) = \mathbb{E}\big[f(\rho_{\mu,\sigma})\big] \qquad \text{where} \qquad \rho_{\mu,\sigma} \stackrel{\mathcal{D}}{\sim} \mathrm{N}(\mu,\sigma^2).$$

The function F satisfies

$$\big|F(\mu_1,\sigma_1) \;-\; F(\mu_2,\sigma_2)\big| \;\lesssim\; |\mu_2 - \mu_1| \;+\; |\sigma_2 - \sigma_1|. \tag{4.13}$$

for any choice $\mu_1, \mu_2 \in \mathbb{R}$ and $\sigma_1, \sigma_2 \geq 0$. Indeed,

$$\begin{aligned}
\big|F(\mu_1,\sigma_1) \;-\; F(\mu_2,\sigma_2)\big| &= \big|\mathbb{E}\big[f(\mu_1 + \sigma_1\,\rho_{0,1}) \;-\; f(\mu_2 + \sigma_2\,\rho_{0,1})\big]\big| \;\leq\; \mathbb{E}\Big[|\mu_2 - \mu_1| \;+\; |\sigma_2 - \sigma_1|\cdot|\rho_{0,1}|\Big] \\
&\lesssim |\mu_2 - \mu_1| \;+\; |\sigma_2 - \sigma_1|.
\end{aligned}$$

We have $\mathbb{E}_x[Z^N(x,\xi^N)] = \mathbb{E}[Z_\ell] = -\frac{\ell^3}{4}$ while the variances are given by

$$\mathrm{Var}\big[Z^N(x,\xi^N)\big] = \frac{\ell^3}{2}\frac{\|x\|^2_{\mathcal{C}^N}}{N} \qquad \text{and} \qquad \mathrm{Var}\big[Z_\ell\big] = \frac{\ell^3}{2}.$$

Therefore, using Lemma 4.3,

$$\begin{aligned}
\mathbb{E}^{\pi^N}\Big[|\bar{f}^N(x) - \bar{f}|^p\Big] &= \mathbb{E}^{\pi^N}\Big[\big|\mathbb{E}_x\big[f\big(Z^N(x,\xi^N) + \mathbf{e}_\star^N(x,\xi^N)\big) - f(Z_\ell)\big]\big|^p\Big] \\
&\lesssim \mathbb{E}^{\pi^N}\Big[\big|\mathbb{E}_x\big[f\big(Z^N(x,\xi^N)\big) - f(Z_\ell)\big]\big|^p\Big] \;+\; \mathbb{E}^{\pi^N}\big[|\mathbf{e}_\star^N(x,\xi^N)|^p\big] \\
&= \mathbb{E}^{\pi^N}\Big[\big|F(-\frac{\ell^3}{4}, \mathrm{Var}\big[Z^N(x,\xi^N)\big]^{\frac{1}{2}}) - F(-\frac{\ell^3}{4}, \mathrm{Var}\big[Z_\ell\big]^{\frac{1}{2}})\big|^p\Big] \;+\; \mathbb{E}^{\pi^N}\big[|\mathbf{e}_\star^N(x,\xi^N)|^p\big]
\end{aligned}$$

$$\lesssim \mathbb{E}^{\pi^N}\Big[\big|\mathrm{Var}[Z^N(x,\xi^N)]^{\frac{1}{2}} - \mathrm{Var}[Z_\ell]^{\frac{1}{2}}\big|^p\Big] \;+\; \mathbb{E}^{\pi^N}[|\mathbf{e}_\star^N(x,\xi^N)|^p]$$
$$\lesssim \mathbb{E}^{\pi_0}\Big|\Big\{\frac{\|x\|^2_{\mathcal{C}^N}}{N}\Big\}^{\frac{1}{2}} - 1\Big|^p \;+\; \mathbb{E}^{\pi^N}[|\mathbf{e}_\star^N(x,\xi^N)|^p] \;\to\; 0$$

In the last step we have used the fact that if $x \overset{\mathcal{D}}{\sim} \pi_0$ then $\frac{\|x\|^2_{\mathcal{C}^N}}{N} \overset{\mathcal{D}}{\sim} \frac{\rho_1^2+\ldots+\rho_N^2}{N}$ where $\rho_1,\ldots,\rho_N$ are i.i.d Gaussian random variables $\mathrm{N}(0,1)$ so that $\mathbb{E}^{\pi_0}\Big|\Big\{\frac{\|x\|^2_{\mathcal{C}^N}}{N}\Big\}^{\frac{1}{2}} - 1\Big|^p \to 0$. ☐

COROLLARY 4.6. *Let $p \geq 1$ be a positive. The local mean acceptance probability $\alpha^N(x)$ defined in Equation (2.19) satisfies*

$$\lim_{N\to\infty} \mathbb{E}^{\pi^N}[|\alpha^N(x) - \alpha(\ell)|^p] = 0.$$

PROOF. The function $f(z) = 1 \wedge e^z$ is 1-Lipschitz and $\alpha(\ell) = \mathbb{E}[f(Z_\ell)]$. Also,

$$\alpha^N(x) = \mathbb{E}_x\Big[f(Q^N(x,\xi^N))\Big] = \mathbb{E}_x\big[f(Z^N(x,\xi^N) + \mathbf{e}_\star^N(x,\xi^N)\Big]$$

with $\mathbf{e}_\star^N(x,\xi^N) = i^N(x,\xi^N) + \mathbf{e}^N(x,\xi^N)$. Lemma 4.4 shows that $\lim_{N\to\infty} \mathbb{E}^{\pi^N}[\mathbf{e}_\star^N(x,\xi)^p] = 0$ and therefore Lemma 4.5 gives the conclusion. ☐

4.3. *Drift approximation.* This section proves that the approximate drift function $d^N : \mathcal{H}^s \to \mathcal{H}^s$ defined in Equation (3.3) converges to the drift function $\mu : \mathcal{H}^s \to \mathcal{H}^s$ of the limiting diffusion (2.23).

LEMMA 4.7. **(Drift Approximation)** *Let Assumptions 2.1 hold. The drift function $d^N : \mathcal{H}^s \to \mathcal{H}^s$ converges to μ in the sense that*

$$\lim_{N\to\infty} \mathbb{E}^{\pi^N}\Big[\|d^N(x) - \mu(x)\|_s^2\Big] = 0.$$

PROOF. The approximate drift d^N is given by Equation (3.3). The definition of the local mean acceptance probability $\alpha^N(x)$ given by Equation (2.19) show that d^N can also be expressed as

$$d^N(x) = \Big(\alpha^N(x)\alpha(\ell)^{-1}\Big)\mu^N(x) + \sqrt{2\ell}h(\ell)^{-1}(\Delta t)^{-\frac{1}{2}}\varepsilon^N(x)$$

where $\mu^N(x) = -\Big(P^N x + \mathcal{C}^N \nabla\Psi^N(x)\Big)$ and the term $\varepsilon^N(x)$ is defined by

$$\varepsilon^N(x) \;=\; \mathbb{E}_x\big[\gamma^N(x,\xi^N)\,\mathcal{C}^{\frac{1}{2}}\xi^N\big] \;=\; \mathbb{E}_x\big[(1 \wedge e^{Q^N(x,\xi^N)})\,\mathcal{C}^{\frac{1}{2}}\xi^N\big].$$

To prove Lemma 4.7 it suffices to verify that

$$\lim_{N\to\infty} \mathbb{E}^{\pi^N}\Big[\|(\alpha^N(x)\alpha(\ell)^{-1})\mu^N(x) - \mu(x)\|_s^2\Big] = 0 \tag{4.14}$$
$$\lim_{N\to\infty} (\Delta t)^{-1}\, \mathbb{E}^{\pi^N}\Big[\|\varepsilon^N(x)\|_s^2\Big] = 0. \tag{4.15}$$

- Let us first prove Equation (4.14). The triangle inequality and Cauchy-Schwarz inequality show that

$$\Big(\mathbb{E}^{\pi^N}\Big[\|(\alpha^N(x)\alpha(\ell)^{-1})\mu^N(x) - \mu(x)\|_s^2\Big]\Big)^2 \lesssim \mathbb{E}[|\alpha^N(x) - \alpha(\ell)|^4] \cdot \mathbb{E}^{\pi^N}[\|\mu^N(x)\|_s^4]$$

$$+ \mathbb{E}^{\pi^N}[\|\mu^N(x) - \mu(x)\|_s^4].$$

By Remark 2.5 $\mu^N : \mathcal{H}^s \to \mathcal{H}^s$ is Lipschitz, with a Lipschitz constant that can be chosen independent of N. It follows that $\sup_N \mathbb{E}^{\pi^N}[\|\mu^N(x)\|_s^4] < \infty$. Lemma 4.5 and Corollary 4.6 show that $\mathbb{E}[|\alpha^N(x) - \alpha(\ell)|^4] \to 0$. Therefore,

$$\lim_{N\to\infty} \mathbb{E}[|\alpha^N(x) - \alpha(\ell)|^4] \cdot \mathbb{E}^{\pi^N}[\|\mu^N(x)\|_s^4] = 0.$$

The functions μ^N and μ are globally Lipschitz on $\mathcal{H}^s$, with a Lipschitz constant that can be chosen independent from N, so that $\|\mu^N(x) - \mu(x)\|_s^4 \lesssim (1 + \|x\|_s^4)$. Lemma 4.1 proves that the sequence of functions $\{\mu^N\}$ converges π_0-almost surely to $\mu(x)$ in $\mathcal{H}^s$ and Lemma 4.3 show that $\mathbb{E}^{\pi^N}[\|\mu^N(x) - \mu(x)\|_s^4] \lesssim \mathbb{E}^{\pi_0}[\|\mu^N(x) - \mu(x)\|_s^4]$. It thus follows from the dominated convergence theorem that

$$\lim_{N\to\infty} \mathbb{E}^{\pi^N}[\|\mu^N(x) - \mu(x)\|_s^4] = 0.$$

This concludes the proof of the Equation (4.14).

- Let us prove Equation (4.15). If the Bernoulli random variable $\gamma^N(x,\xi^N)$ were independent from the noise term $(\mathcal{C}^N)^{\frac{1}{2}}\xi^N$, it would follow that $\varepsilon^N(x) = 0$. In general $\gamma^N(x,\xi^N)$ is not independent from $(\mathcal{C}^N)^{\frac{1}{2}}\xi^N$ so that $\varepsilon^N(x)$ is not equal to zero. Nevertheless, as quantified by Lemma 4.5, the Bernoulli random variable $\gamma^N(x,\xi^N)$ is asymptotically independent from the current position x and from the noise term $(\mathcal{C}^N)^{\frac{1}{2}}\xi^N$. Consequently, we can prove in Equation (4.17) that the quantity $\varepsilon^N(x)$ is small. To this end, we establish that each component $\langle \varepsilon(x), \hat{\varphi}_j \rangle_s^2$ satisfies

$$\mathbb{E}^{\pi^N}[\langle \varepsilon^N(x), \hat{\varphi}_j \rangle_s^2] \quad \lesssim \quad N^{-1}\mathbb{E}^{\pi^N}[\langle x, \hat{\varphi}_j \rangle_s^2] + N^{-\frac{2}{3}}(j^s\lambda_j)^2. \tag{4.16}$$

Summation of Equation (4.16) over $j = 1, \ldots, N$ leads to

$$\mathbb{E}^{\pi^N}\Big[\|\varepsilon^N(x)\|_s^2\Big] \quad \lesssim \quad N^{-1}\mathbb{E}^{\pi^N}[\|x\|_s^2] + N^{-\frac{2}{3}}\,\mathrm{Tr}_{\mathcal{H}^s}(\mathcal{C}_s) \quad \lesssim \quad N^{-\frac{2}{3}}, \tag{4.17}$$

which gives the proof of Equation (4.15). To prove Equation (4.16) for a fixed index $j \in \mathbb{N}$, the quantity $Q^N(x,\xi)$ is decomposed as a sum of a term independent from ξ_j and another remaining term of small magnitude. To this end we introduce

$$\begin{cases} Q^N(x,\xi^N) &= Q_j^N(x,\xi^N) + Q_{j,\perp}^N(x,\xi^N) \\ Q_j^N(x,\xi^N) &= -\frac{1}{\sqrt{2}}\ell^{\frac{3}{2}} N^{-\frac{1}{2}} \lambda_j^{-1} x_j \xi_j - \frac{1}{2}\ell^2 N^{-\frac{2}{3}} \lambda_j^2 \xi_j^2 + \mathbf{e}^N(x,\xi^N). \end{cases} \tag{4.18}$$

The definitions of $Z^N(x,\xi^N)$ and $i^N(x,\xi^N)$ in Equation (4.5) and (4.6) readily show that $Q_{j,\perp}^N(x,\xi^N)$ is independent from ξ_j. The noise term satisfies $\mathcal{C}^{\frac{1}{2}}\xi^N = \sum_{j=1}^N (j^s\lambda_j)\xi_j\hat{\varphi}_j$. Since $Q_{j,\perp}^N(x,\xi^N)$ and ξ_j are independent and $z \mapsto 1 \wedge e^z$ is 1-Lipschitz, it follows that

$$\begin{aligned} \langle \varepsilon^N(x), \hat{\varphi}_j \rangle_s^2 &= (j^s\lambda_j)^2 \Big(\mathbb{E}_x[(1 \wedge e^{Q^N(x,\xi^N)})\ \xi_j]\Big)^2 \\ &= (j^s\lambda_j)^2 \Big(\mathbb{E}_x[[(1 \wedge e^{Q^N(x,\xi^N)}) - (1 \wedge e^{Q_{j,\perp}^N(x,\xi^N)})]\ \xi_j]\Big)^2 \\ &\lesssim (j^s\lambda_j)^2 \mathbb{E}_x[|Q^N(x,\xi^N)) - Q_{j,\perp}^N(x,\xi^N)|^2] \end{aligned}$$

$$= (j^s\lambda_j)^2 \mathbb{E}_x[Q_j^N(x,\xi^N)^2].$$

By Lemma 4.4 $\mathbb{E}^{\pi^N}[\mathbf{e}^N(x,\xi^N)^2] \lesssim N^{-\frac{2}{3}}$. Therefore,

$$\begin{aligned}(j^s\lambda_j)^2 \mathbb{E}^{\pi^N}[Q_j^N(x,\xi^N)^2] &\lesssim (j^s\lambda_j)^2 \Big\{ N^{-1}\lambda_j^{-2}\mathbb{E}^{\pi^N}[x_j^2\xi_j^2] + N^{-\frac{4}{3}}\mathbb{E}^{\pi^N}[\lambda_j^4\xi_j^4] + \mathbb{E}^{\pi^N}[\mathbf{e}^N(x,\xi)^2]\Big\} \\ &\lesssim N^{-1}\,\mathbb{E}^{\pi^N}[(j^s x_j)^2\xi_j^2] + (j^s\lambda_j)^2(N^{-\frac{4}{3}} + N^{-\frac{2}{3}}) \\ &\lesssim N^{-1}\,\mathbb{E}^{\pi^N}[\langle x,\hat{\varphi}_j\rangle_s^2] + (j^s\lambda_j)^2 N^{-\frac{2}{3}} \\ &\lesssim N^{-1}\,\mathbb{E}^{\pi^N}[\langle x,\hat{\varphi}_j\rangle_s^2] + (j^s\lambda_j)^2 N^{-\frac{2}{3}},\end{aligned}$$

which finishes the proof of Equation (4.16).

□

4.4. *Noise approximation.* Recall the definition (3.4) of the martingale difference $\Gamma^{k,N}$. In this section we estimate the error in the approximation $\Gamma^{k,N} \approx \mathrm{N}(0,\mathcal{C}_s)$. To this end we introduce the covariance operator

$$D^N(x) = \mathbb{E}_x\Big[\Gamma^{k,N} \otimes_{\mathcal{H}^s} \Gamma^{k,N} \,|x^{k,N} = x\Big].$$

For any $x,u,v \in \mathcal{H}^s$ the operator $D^N(x)$ satisfies

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

The next lemma gives a quantitative version of the approximation $D^N(x) \approx \mathcal{C}_s$.

Lemma 4.8. *Let Assumptions 2.1 hold. For any pair of indices $i,j \geq 0$ the operator $D^N(x) : \mathcal{H}^s \to \mathcal{H}^s$ satisfies*

$$\lim_{N\to\infty} \quad \mathbb{E}^{\pi^N} \big|\langle \hat{\varphi}_i, D^N(x)\hat{\varphi}_j\rangle_s - \langle \hat{\varphi}_i, \mathcal{C}_s \hat{\varphi}_j\rangle_s\big| \;=\; 0 \tag{4.19}$$

and, furthermore,

$$\lim_{N\to\infty} \quad \mathbb{E}^{\pi^N}\big|\mathrm{Tr}_{\mathcal{H}^s}(D^N(x)) - \mathrm{Tr}_{\mathcal{H}^s}(\mathcal{C}_s)\big| \;=\; 0. \tag{4.20}$$

Proof. The martingale difference $\Gamma^N(x,\xi)$ is given by

$$\Gamma^N(x,\xi) = \alpha(\ell)^{-\frac{1}{2}}\gamma^N(x,\xi)\,\mathcal{C}^{\frac{1}{2}}\xi \;+\; \frac{1}{\sqrt{2}}\alpha(\ell)^{-\frac{1}{2}}(\ell\Delta t)^{\frac{1}{2}}\Big\{\gamma^N(x,\xi)\mu^N(x) - \alpha(\ell)d^N(x)\Big\}. \tag{4.21}$$

We only prove Equation (4.20); the proof of Equation (4.19) is essentially identical but easier. Remark 2.5 shows that the functions $\mu, \mu^N : \mathcal{H}^s \to \mathcal{H}^s$ are globally Lipschitz and Lemma 4.7 shows that $\mathbb{E}^{\pi^N}[\|d^N(x) - \mu(x)\|_s^2] \to 0$. Therefore

$$\mathbb{E}^{\pi^N}\Big[\|\gamma^N(x,\xi)\mu^N(x) - \alpha(\ell)d^N(x)\|_s^2\Big] \;\lesssim\; 1, \tag{4.22}$$

which implies that the second term on the right-hand-side of Equation (4.21) is $\mathcal{O}(\sqrt{\Delta t})$. Since $\mathrm{Tr}_{\mathcal{H}^s}(D^N(x)) = \mathbb{E}_x[\|\Gamma^N(x,\xi)\|_s^2]$, Equation (4.22) implies that

$$\mathbb{E}^{\pi^N}\Big[\big|\alpha(\ell)\;\mathrm{Tr}_{\mathcal{H}^s}(D^N(x)) - \mathbb{E}_x[\|\gamma^N(x,\xi)\,\mathcal{C}^{\frac{1}{2}}\xi\|_s^2]\big|\Big] \lesssim (\Delta t)^{\frac{1}{2}}.$$

Consequently, to prove Equation (4.20) it suffices to verify that

$$\lim_{N\to\infty} \mathbb{E}^{\pi^N}\Big[\big|\mathbb{E}_x[\|\gamma^N(x,\xi)\,\mathcal{C}^{\frac12}\xi\|_s^2] - \alpha(\ell)\mathrm{Tr}_{\mathcal{H}^s}(\mathcal{C}_s)\big|\Big] \;=\; 0. \tag{4.23}$$

We have $\mathbb{E}_x[\|\gamma^N(x,\xi)\,\mathcal{C}^{\frac12}\xi\|_s^2] = \sum_{j=1}^N (j^s\lambda_j)^2\mathbb{E}_x[(1\wedge e^{Q^N(x,\xi)})\xi_j^2]$. Therefore, to prove Equation (4.23) it suffices to establish

$$\lim_{N\to\infty} \sum_{j=1}^N (j^s\lambda_j)^2\mathbb{E}^{\pi^N}\Big[\big|\mathbb{E}_x[(1\wedge e^{Q^N(x,\xi)})\xi_j^2] - \alpha(\ell)\big|\Big] = 0. \tag{4.24}$$

Since $\sum_{j=1}^\infty (j^s\lambda_j)^2 < \infty$ and $|1\wedge e^{Q^N(x,\xi)}| \leq 1$, the dominated convergence theorem shows that (4.24) follows from

$$\lim_{N\to\infty} \mathbb{E}^{\pi^N}\Big[\big|\mathbb{E}_x[(1\wedge e^{Q^N(x,\xi)})\xi_j^2] - \alpha(\ell)\big|\Big] = 0 \qquad \forall j\geq 0. \tag{4.25}$$

We now prove Equation (4.25). As in the proof of Lemma 4.7, we use the decomposition $Q^N(x,\xi) = Q_j^N(x,\xi) + Q_{j,\perp}^N(x,\xi)$ where $Q_{j,\perp}^N(x,\xi)$ is independent from ξ_j. Therefore, since $\mathrm{Lip}(f)=1$,

$$\begin{aligned}
\mathbb{E}_x[(1\wedge e^{Q^N(x,\xi)})\xi_j^2] &= \mathbb{E}_x[(1\wedge e^{Q_{j,\perp}^N(x,\xi)})\xi_j^2] + \mathbb{E}_x[[(1\wedge e^{Q^N(x,\xi)}) - (1\wedge e^{Q_{j,\perp}^N(x,\xi)})]\xi_j^2] \\
&= \mathbb{E}_x[1\wedge e^{Q_{j,\perp}^N(x,\xi)}] + \mathcal{O}\Big(\{\mathbb{E}_x[|Q^N(x,\xi)) - Q_{j,\perp}^N(x,\xi)|^2]\}^{\frac12}\Big) \\
&= \mathbb{E}_x[1\wedge e^{Q_{j,\perp}^N(x,\xi)}] + \mathcal{O}\Big(\{\mathbb{E}_x[Q_j^N(x,\xi)^2]\}^{\frac12}\Big).
\end{aligned}$$

Lemma 4.5 ensures that, for $f(\cdot) = 1\wedge\exp(\cdot)$,

$$\lim_{N\to\infty} \mathbb{E}^{\pi^N}\Big[\big|\mathbb{E}_x[f(Q_{j,\perp}^N(x,\xi))] - \alpha(\ell)\big|\Big] = 0$$

and the definition of $Q_i^N(x,\xi)$ readily shows that $\lim_{N\to\infty}\mathbb{E}^{\pi^N}[Q_j^N(x,\xi)^2] = 0$. This concludes the proof of Equation (4.25) and thus ends the proof of Lemma 4.8. □

Corollary 4.9. *More generally, for any fixed vector $h\in\mathcal{H}^s$, the following limit holds,*

$$\lim_{N\to\infty} \mathbb{E}^{\pi^N}\big|\langle h, D^N(x)h\rangle_s - \langle h,\mathcal{C}_s h\rangle_s\big| \;=\; 0. \tag{4.26}$$

Proof. If $h=\hat\varphi_i$, this is precisely the content of Proposition 3.1. More generally, by linearity, Proposition 3.1 shows that this is true for $h=\sum_{i\leq N}\alpha_i\hat\varphi_i$, where $N\in\mathbb{N}$ is a fixed integer. For a general vector $h\in\mathcal{H}^s$, we can use the decomposition $h = h^* + e^*$ where $h^* = \sum_{j\leq N}\langle h,\hat\varphi_j\rangle_s\,\hat\varphi_j$ and $e^* = h - h^*$. It follows that

$$\begin{aligned}
&\Big|\Big(\langle h, D^N(x)\,h\rangle_s - \langle h,\mathcal{C}_s h\rangle_s\Big) - \Big(\langle h^*, D^N(x)\,h^*\rangle_s - \langle h^*,\mathcal{C}_s h^*\rangle_s\Big)\Big| \\
&\qquad\qquad \leq \Big|\langle h+h^*, D^N(x)\,(h-h^*)\rangle_s - \langle h+h^*,\mathcal{C}_s\,(h-h^*)\rangle_s\Big| \\
&\qquad\qquad \leq 2\|h\|_s\cdot\|h-h^*\|_s\cdot\Big(\mathrm{Tr}_{\mathcal{H}^s}(D^N(x)) + \mathrm{Tr}_{\mathcal{H}_s}(\mathcal{C}_s)\Big),
\end{aligned}$$

where we have used the fact that for an non-negative self-adjoint operator $D:\mathcal{H}^s\to\mathcal{H}^s$ we have $\langle u, Dv\rangle_s \leq \|u\|_s\cdot\|v\|_s\cdot\mathrm{Tr}_{\mathcal{H}^s}(D)$. Proposition 3.1 shows that $\mathbb{E}^{\pi^N}[\mathrm{Tr}_{\mathcal{H}^s}(D^N(x))] < \infty$ and Assumption 2.1 ensures that $\mathrm{Tr}_{\mathcal{H}^s}(\mathcal{C}_s) < \infty$. Consequently,

$$\lim_{N\to\infty}\mathbb{E}^{\pi^N}\Big|\langle h, D^N(x)\,h\rangle - \langle h,\mathcal{C}_s h\rangle\Big| \lesssim \lim_{N\to\infty}\mathbb{E}^{\pi^N}\Big|\langle h^*, D^N(x)\,h^*\rangle - \langle h^*,\mathcal{C}_s\,h^*\rangle\Big| \;+\; \|h-h^*\|_s$$

$$= \|h - h^*\|_s.$$

Since $\|h-h^*\|_s$ can be chosen arbitrarily small, the conclusion follows. ☐

4.5. *Martingale Invariance Principle.* This section proves that the process W^N defined in Equation (3.7) converges to a Brownian motion.

Proposition 4.10. *Let Assumptions 2.1 hold. Let $z^0 \sim \pi$ and $W^N(t)$ the process defined in equation (3.7) and $x^{0,N} \overset{\mathcal{D}}{\sim} \pi^N$ the starting position of the Markov chain x^N. Then*

$$(x^{0,N}, W^N) \Longrightarrow (z^0, W), \tag{4.27}$$

where $\Longrightarrow$ denotes weak convergence in $\mathcal{H}^s \times C([0,T];\mathcal{H}^s)$, and W is a $\mathcal{H}^s$-valued Brownian motion with covariance operator $\mathcal{C}_s$. Furthermore the limiting Brownian motion W is independent of the initial condition z^0.

Proof. As a first step, we show that W^N converges weakly to W. As described in [MPS11], a consequence of Proposition 5.1 of [Ber86] shows that in order to prove that W^N 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, the third for any $\epsilon > 0$,

$$\lim_{N\to\infty} \quad \Delta t \sum_{k=1}^{k_N(T)} \mathbb{E}\Big[\|\Gamma^{k,N}\|_s^2 \,|\mathcal{F}^{k,N}\Big] \;=\; T\,\mathrm{Tr}_{\mathcal{H}^s}(\mathcal{C}_s) \tag{4.28}$$

$$\lim_{N\to\infty} \quad \Delta t \sum_{k=1}^{k_N(t)} \mathbb{E}\Big[\langle \Gamma^{k,N}, \hat{\varphi}_i\rangle_s \langle \Gamma^{k,N}, \hat{\varphi}_j\rangle_s \,|\mathcal{F}^{k,N}\Big] \;=\; t\,\langle \hat{\varphi}_i, \mathcal{C}_s \hat{\varphi}_j\rangle_s \tag{4.29}$$

$$\lim_{N\to\infty} \quad \Delta t \sum_{k=1}^{k_N(T)} \mathbb{E}\Big[\|\Gamma^{k,N}\|_s^2 1_{\{\|\Gamma^{k,N}\|_s^2 \geq \Delta t\, \epsilon\}} \,|\mathcal{F}^{k,N}\Big] \;=\; 0 \tag{4.30}$$

where $k_N(t) = \lfloor \frac{t}{\Delta t} \rfloor$, $\{\hat{\varphi}_j\}$ is an orthonormal basis of $\mathcal{H}^s$ and $\mathcal{F}^{k,N}$ is the natural filtration of the Markov chain $\{x^{k,N}\}$. The proof follows from the estimate on $D^N(x) = \mathbb{E}\Big[\Gamma^{0,N} \otimes \Gamma^{0,N} \,|x^{0,N} = x\Big]$ presented in Lemma 4.8 For the sake of simplicity, we will write $\mathbb{E}_k[\,\cdot\,]$ instead of $\mathbb{E}[\,\cdot\,|\mathcal{F}^{k,N}]$. We now prove that the three conditions are satisfied.

- **Condition** (4.28) It is enough to prove that $\lim\, \mathbb{E}\Big|\Big\{\frac{1}{\lfloor N^{\frac{1}{3}} \rfloor} \sum_{k=1}^{\lfloor N^{\frac{1}{3}} \rfloor} \mathbb{E}_k\big[\|\Gamma^{k,N}\|_s^2\big]\Big\} - \mathrm{Tr}_{\mathcal{H}^s}(\mathcal{C}_s)\Big| = 0$ where

$$\mathbb{E}_k\Big[\|\Gamma^{k,N}\|_s^2\Big] \;=\; \mathbb{E}_k \sum_{j=1}^{N} \Big[\langle \hat{\varphi}_j, D^N(x^{k,N}) \hat{\varphi}_j\rangle_s\Big] \;=\; \mathbb{E}_k \mathrm{Tr}_{\mathcal{H}^s}(D^N(x^{k,N})).$$

  Because the Metropolis-Hastings algorithm preserves stationarity and $x^{0,N} \overset{\mathcal{D}}{\sim} \pi^N$ it follows that $x^{k,N} \overset{\mathcal{D}}{\sim} \pi^N$ for any $k \geq 0$. Therefore, for all $k \geq 0$ we have $\mathrm{Tr}_{\mathcal{H}^s}(D^N(x^{k,N})) \overset{\mathcal{D}}{\sim} \mathrm{Tr}_{\mathcal{H}^s}(D^N(x))$ where $x \overset{\mathcal{D}}{\sim} \pi^N$. Consequently, the triangle inequality shows that

$$\mathbb{E}\Big|\Big\{\frac{1}{\lfloor N^{\frac{1}{3}} \rfloor} \sum_{k=1}^{\lfloor N^{\frac{1}{3}} \rfloor} \mathbb{E}_k \|\Gamma^{k,N}\|^2\Big\} - \mathrm{Tr}_{\mathcal{H}^s}(\mathcal{C}_s)\Big| \quad \leq \quad \mathbb{E}^{\pi^N}\Big|\mathrm{Tr}_{\mathcal{H}^s}(D^N(x)) - \mathrm{Tr}_{\mathcal{H}^s}(\mathcal{C}_s)\Big| \to 0$$

  where the last limit follows from Lemma 4.8.

- **Condition** (4.29) It is enough to prove that

$$\lim \ \mathbb{E}^{\pi^N}\Big|\Big\{\frac{1}{\lfloor N^{\frac{1}{3}}\rfloor}\sum_{k=1}^{\lfloor N^{\frac{1}{3}}\rfloor}\mathbb{E}_k\Big[\langle\Gamma^{k,N},\hat{\varphi}_i\rangle_s\langle\Gamma^{k,N},\hat{\varphi}_j\rangle_s\Big]\Big\} - \langle\hat{\varphi}_i,\mathcal{C}_s\hat{\varphi}_j\rangle_s\Big| = 0$$

  where $\mathbb{E}_k\Big[\langle\Gamma^{k,N},\hat{\varphi}_i\rangle_s\langle\Gamma^{k,N},\hat{\varphi}_j\rangle_s\Big] = \langle\hat{\varphi}_i, D^N(x^{k,N})\hat{\varphi}_j\rangle_s$. Because $x^{k,N} \overset{\mathcal{D}}{\sim} \pi^N$ the conclusion again follows from Lemma 4.8.
- **Condition** (4.30) For all $k \geq 1$ we have $x^{k,N} \overset{\mathcal{D}}{\sim} \pi^N$ so that

$$\mathbb{E}^{\pi^N}\Big|\frac{1}{\lfloor N^{\frac{1}{3}}\rfloor}\sum_{k=1}^{\lfloor N^{\frac{1}{3}}\rfloor}\mathbb{E}_k[\|\Gamma^{k,N}\|_s^2 1_{\|\Gamma^{k,N}\|_s^2\geq N^{\frac{1}{3}}\varepsilon}]\Big| \ \leq \ \mathbb{E}^{\pi^N}\|\Gamma^{0,N}\|_s^2 1_{\{\|\Gamma^{0,N}\|_s^2\geq N^{\frac{1}{3}}\varepsilon\}}.$$

  Equation (4.21) shows that for any power $p \geq 0$ we have $\sup_N \ \mathbb{E}^{\pi^N}[\|\Gamma^{0,N}\|_s^p] < \infty$. Therefore the sequence $\{\|\Gamma^{0,N}\|_s^2\}$ is uniformly integrable, which shows that

$$\lim_{N\to\infty} \ \mathbb{E}^{\pi^N}\|\Gamma^{0,N}\|_s^2 1_{\{\|\Gamma^{0,N}\|_s^2\geq N^{\frac{1}{3}}\varepsilon\}} = 0.$$

The three hypothesis are satisfied, proving that W^N converges weakly in $C([0,T];\mathcal{H}^s)$ to a Brownian motion W in $\mathcal{H}^s$ with covariance $\mathcal{C}_s$. Therefore, Corollary 4.4 of [MPS11] shows that the sequence $\{(x^{0,N},W^N)\}_{N\geq 1}$ converges weakly to (z^0,W) in $\mathcal{H}\times C([0,T],\mathcal{H}^s)$. This finishes the proof of Proposition 4.10. □

5. Conclusion. We have studied the application of the MALA algorithm to sample from measures defined via density with respect to a Gaussian measure on Hilbert space. We prove that a suitably interpolated and scaled version of the Markov chain has a diffusion limit in infinite dimensions. There are two main conclusions which follow from this theory: firstly this work shows that, in stationarity, the MALA algorithm applied to an $N-$dimensional approximation of the target will take $\mathcal{O}(N^{\frac{1}{3}})$ steps to explore the invariant measure; secondly the MALA algorithm will be optimized at an average acceptance probability of 0.574. We have thus significantly extended the work [RR98] which reaches similar conclusions in the case of i.i.d. product targets. In contrast we have considered target measures with significant correlation, with structure motivated by a range of applications. As a consequence our limit theorems are in an infinite dimensional Hilbert space and we have employed an approach to the derivation of the diffusion limit which differs significantly from that used in [RR98]. This approach was developed in [MPS11] to study diffusion limits for the RWM algorithm.

There are many possible developments of this work. We list several of these.

- In [BPR+11] it is shown that the Hybrid Monte Carlo algorithm (HMC) requires, for target measures of the form (1.1), $\mathcal{O}(N^{\frac{1}{4}})$ steps to explore the invariant measure. However there is no diffusion limit in this case. Identifying an appropriate limit, and extending analysis to the case of target measures (2.11) provides a challenging avenue for exploration.
- In the i.i.d product case it is known that, if the Markov chain is started "far" from stationarity, a fluid limit (ODE) is observed [CRR05]. It would be interesting to study such limits in the present context.
- Combining the analysis of MCMC methods for hierarchical target measures [Béd09] with the analysis herein provides a challenging set of theoretical questions, as well as having direct applicability.

- It should also be noted that, for measures absolutely continuous with respect to a Gaussian, there exist new non-standard versions of RWM [BS09], MALA [BRSV08] and HMC [BPSSA11] for which the acceptance probability does not degenerate to zero as dimension N increases. These methods may be expensive to implement when the Karhunen-Loève basis is not known explicitly, and comparing their overall efficiency with that of standard RWM, MALA and HMC is an interesting area for further study.
- It is natural to ask whether analysis similar to that undertaken here could be developed for Metropolis-Hastings methods applied to other reference measures with a non-Gaussian product structure. In particular the Besov priors of [LSS09] provide an interesting class of such reference measures, and the paper [DHS11] provides a machinery for analyzing change of measure from the Besov prior, analogous to that used here in the Gaussian case. Another intertesting class of reference measures are those used in the study of uncertainty quantification for elliptic PDEs: these have the form of an infinite product of compactly supported uniform distributions; see [SS].

Acknowledgements Part of this work was done when AT was visiting the Department of Statistics at Harvard University, and we thank this Institution for its hospitality.

References.

[Béd07] M. Bédard. Weak convergence of Metropolis algorithms for non-i.i.d. target distributions. *Ann. Appl. Probab.*, 17(4):1222–1244, 2007.

[Béd09] M. Bédard. On the optimal scaling problem of metropolis algorithms for hierarchical target distributions. 2009. preprint.

[Ber86] E. Berger. Asymptotic behaviour of a class of stochastic approximation procedures. *Probab. Theory Relat. Fields*, 71(4):517–552, 1986.

[BPR+11] A. Beskos, N. Pillai, G.O. Roberts, J.-M. Sanz-Serna, and A.M. Stuart. Optimal tuning of hybrid monte-carlo. 2011.

[BPS04] L.A. Breyer, M. Piccioni, and S. Scarlatti. Optimal scaling of MaLa for nonlinear regression. *Ann. Appl. Probab.*, 14(3):1479–1505, 2004.

[BPSSA11] A. Beskos, F. Pinski, J.M. Sanz-Serna, and A.M.Stuart. Hybrid Monte-Carlo on hilbert spaces. *Stoch. Proc. Applics.*, .:., 2011.

[BR00] L. A. Breyer and G. O. Roberts. From Metropolis to diffusions: Gibbs states and optimal scaling. *Stochastic Process. Appl.*, 90(2):181–206, 2000.

[BRS09] A. Beskos, G.O. Roberts, and A.M. Stuart. Optimal scalings of metropolis-hastings algorithms for non-product targets in high dimensions. *Annals of Applied Probability*, 19:863–898, 2009.

[BRSV08] A. Beskos, G.O. Roberts, A.M. Stuart, and J. Voss. An MCMC method for diffusion bridges. *Stochastics and Dynamics*, 8(3):319–350, 2008.

[BS09] A. Beskos and A.M. Stuart. MCMC methods for sampling function space. In *Invited Lectures, Sixth International Congress on Industrial and Applied Mathematics, ICIAM07, Editors Rolf Jeltsch and Gerhard Wanner*, pages 337–364. European Mathematical Society, 2009.

[CRR05] O.F. Christensen, G.O. Roberts, and J.S. Rosenthal. Scaling limits for the transient phase of local Metropolis–Hastings algorithms. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):253–268, 2005.

[DHS11] M. Dashti, S. Harris, and A.M. Stuart. Besov priors for bayesian inverse problems. *Inverse Problems and Imaging*, http://arxiv.org/abs/1105.0889, 2011.

[DPZ92] G. Da Prato and J. Zabczyk. *Stochastic equations in infinite dimensions*, volume 44 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1992.

[EK86] S. N. Ethier and T. G. Kurtz. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986. Characterization and convergence.

[HAVW05] M. Hairer, A.M.Stuart, J. Voss, and P. Wiberg. Analysis of SPDEs arising in path sampling. Part I: the gaussian case. *Comm. Math. Sci.*, 3:587–603, 2005.

[HSV07] M. Hairer, A. M. Stuart, and J. Voss. Analysis of SPDEs arising in path sampling. PartII: the nonlinear case. *Ann. Appl. Probab.*, 17(5-6):1657–1706, 2007.

[HSV11] M. Hairer, A. M. Stuart, and J. Voss. Signal processing problems on function space: Bayesian formulation, stochastic pdes and effective mcmc methods. *The Oxford Handbook of Nonlinear Filtering, Editors D. Crisan and B. Rozovsky*, 2011.

[LSS09] M. Lassas, E. Saksman, and S. Siltanen. Discretization-invariant Bayesian inversion and Besov space priors. *Inverse Problems and Imaging*, 3:87–122, 2009.

[MPS11] J.C. Mattingly, N.S. Pillai, and A.M. Stuart. SPDE Limits of the Random Walk Metropolis Algorithm in High Dimensions. *Ann. Appl. Prob.*, 2011.

[MRTT53] N. Metropolis, A.W. Rosenbluth, M.N. Teller, and E. Teller. Equations of state calculations by fast computing machines. *J. Chem. Phys.*, 21:1087–1092, 1953.

[RC04] C. P. Robert and G. Casella. *Monte Carlo statistical methods*. Springer Texts in Statistics. Springer-Verlag, New York, second edition, 2004.

[RGG97] G. O. Roberts, A. Gelman, and W. R. Gilks. Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.*, 7(1):110–120, 1997.

[RR98] G. O. Roberts and J. S. Rosenthal. Optimal scaling of discrete approximations to Langevin diffusions. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 60(1):255–268, 1998.

[RR01] G. O. Roberts and J. S. Rosenthal. Optimal scaling for various Metropolis-Hastings algorithms. *Statist. Sci.*, 16(4):351–367, 2001.

[SFR10] C. Sherlock, P. Fearnhead, and G.O. Roberts. The random walk metropolis : linking theory and practice through a case study. *Statistical Science*, 25(2):172–190, 2010.

[SS] Ch. Schwab and A.M. Stuart. Sparse deterministic approximation of bayesian inverse problems. *Inverse Problems, (under revision)*, http://arxiv.org/abs/1103.4522.

[Stu10] A.M. Stuart. Inverse problems: a Bayesian perspective. *Acta Numerica*, 19, 2010.

Department of Statistics,
Harvard University
MA, USA, 02138-2901
E-mail: pillai@fas.harvard.edu

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

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