Optimal Scaling and Diffusion Limits for the Langevin Algorithm in High Dimensions

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

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

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

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

Abstract: The Metropolis-adjusted Langevin (MALA) algorithm is a sampling algorithm which makes local moves by incorporating information about the gradient 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 at 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. 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. Until now such results were proved only for targets which are products of one dimensional distributions, or for variants of this situation. Our result is the first derivation of scaling limits for the MALA algorithm to target measures which are not of the product form. As a consequence the rescaled MALA algorithm converges weakly to an infinite dimensional Hilbert space valued diffusion, and not to a scalar diffusion.The diffusion limit is proved by showing that a drift-martingale decomposition of the Markov chain, suitably scaled, closely resembles an 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.

1. Introduction

Sampling probability distributions 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. The simplest class of target measures for which such an analysis can be carried out are perhaps target distributions 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.

Consider a π^N-invariant Metropolis Hastings Markov chain $x^N = \{x^{k,N}\}_{k\geq 1}$ which employs local moves, *i.e.*, the chain moves from current state $x^{k,N}$ to a new state $x^{k+1,N}$ via proposing a candidate y according to a proposal kernel $q(x^{k,N}, y)$ and accepting the proposed value with probability $\alpha(x^{k,N}, y) = 1 \wedge \frac{\pi^N(y)q(y,x^{k,N})}{\pi^N(x)q(x^{k,N},y)}$. Two widely used proposals are the Random walk proposal (obtained from the discrete approximation of

imsart-generic ver. 2007/12/10 file: langevin.tex date: March 4, 2011

Brownian motion)

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

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

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

Here 2δ is the proposal variance, a small parameter representing the discrete time increment. 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].

A fruitful way to quantify the computational cost of these Markov chains which proceed via local moves is to determine the "optimal" size of increment δ as a function of dimension N (the precise notion of optimality is discussed below). To this end it is useful to define a continuous interpolant of the Markov chain as follows:

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

The proposal variance is $2\ell\Delta t$, 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. These are addressed in the following discussion.

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. 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. 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

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

Here $h(\ell)$ is 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]). These analyses show the number of steps required to sample the target measure grows as N for RWM, but only as $N^{\frac{1}{3}}$ for MALA. The MALA proposal is more sophisticated than the RWM proposal as the MALA proposal incorporates information about the target density by via the log gradient term (1.3). The optimal scaling results mentioned enable us to quantify the efficiency gained by the MALA compared to that of RWM by employing "informed" local moves using the gradient 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 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. The criteria of optimality used here is defined on the scale set by choosing the largest proposal variance, as a function of dimension, which leads to $\mathcal{O}(1)$ acceptance probability, as the dimension grows. A simple heuristic suggests the existence of such an "optimal scale". Smaller values of the proposal variance lead to high acceptance rates but the chain only moves locally and therefore may not be efficient. Larger values of the proposal variance lead to global proposal 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. Once on this optimal scale, a diffusion limit is obtained and a further optimization, with respect to ℓ, leads the optimal choice of acceptance probability.

Although the optimal acceptance probability and the asymptotic diffusive behavior derived in the original papers of Roberts and coworkers required i.i.d product measures (1.1), extensive simulations (see [RR01, SFR10]) show that these results also hold for more complex target distributions. A wide range of subsequent research [Béd07, Béd09, BPS04, BR00, CRR05] confirmed this in slightly more complicated models such as products of one dimensional distributions with different variances and elliptically symmetric distributions, but the diffusion limits obtained were essentially one dimensional. This is because if one considers target distributions which are not of the product form, the different coordinates interact with each other and therefore the limiting diffusion must take values in an infinite dimensional space.

Our perspective on these problems is motivated by applications such Bayesian nonparametric statistics, for example in application to inverse problems [Stu10], and the theory of conditioned diffusions [HSV10]. 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 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 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 π_0-measurable functional $\Psi : \mathcal{H}^s \mapsto \mathbb{R}$, for some subspace $\mathcal{H}^s$ contained in $\mathcal{H}$; 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 C.

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

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

where W is a Brownian motion (see [DPZ92]) in $\mathcal{H}$ with covariance operator C. In [MPS09] 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 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 purpose of this article is to develop such an analysis for the MALA algorithm. The above mentioned papers primarily study the RWM algorithm. To our knowledge, the only paper to consider the optimal scaling for the Langevin algorithm for non-product targets is [BPS04], in the context of non-linear regression. In [BPS04] the target measure a structure similar to that of the mean field models studied in statistical mechanics so that the target measures behave 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 scaling limit of the Metropolis adjusted Langevin algorithm to the SPDE (1.7), when applied to 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 [RGG97]. Thus in this regard, our work is the first genuine extension of the remarkable results in [RR98] for the Langevin algorithm to target measures which are not of the product form, confirming the results observed from simulation. Recently, in [MPS09], the first two authors developed an approach for deriving diffusion limits for such algorithms. This approach to obtaining invariance principles for Markov chains yields insights into the behavior of MCMC algorithms in high dimensions. In this paper we further develop this method and we believe that the techniques developed here can be built on to derive scaling limits for other popular MCMC algorithms.

In section 2 we set-up the mathematical framework that we adopt througout, and define the version of the MALA algorithm which is the object of our study. Section 3 contains statement of the main theorem, and we outline the proof strategy. Section 4 is devoted to a variety of key estimates, in particular to quantify a Gaussian approximation in the acceptance probability for MALA and, using this, estimates of the mean drift and diffusion. Then, in section 5, we combine these preliminary estimates in order to prove the main theorem. We end, in section 6, with some concluding remarks about future directions.

2. MALA Algorithm

In this section we introduce the Karhunen-Loéve representation used throughout the paper, and define the precise version of the MALA algorithm that we study in detail. 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 $\alpha_n \leq K\beta_n$ for all $n \geq 0$. The notations $\alpha_n \asymp \beta_n$ means that $\alpha_n \lesssim \beta_n$ and $\beta_n \lesssim \alpha_n$.
- Two sequences of real functions $\{f_n\}$ and $\{g_n\}$ defined on the same set D satisfy $f_n \lesssim g_n$ if there exists a constant $K > 0$ satisfying $f_n(x) \leq K g_n(x)$ for all $n \geq 0$ 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. Karhunen-Loève Basis

Let $\mathcal{H}$ be a separable Hilbert space of real valued functions with scalar product denoted by $\langle \cdot, \cdot \rangle$ and associated norm $\|x\|^2 = \langle x, x\rangle$. Let C be a positive, trace class operator on $\mathcal{H}$ and $\{\varphi_j, \lambda_j^2\}$ be the eigenfunctions and eigenvalues of C respectively, so that

$$C\varphi_j = \lambda_j^2\,\varphi_j, \ j \in \mathbb{N}.$$

We assume a normalization under which $\{\varphi_j\}$ forms a complete orthonormal basis in $\mathcal{H}$, which we refer to us as the Karhunen-Loève Basis. We assume throughout the sequel that the eigenvalues are arranged in decreasing order.

Any function $x \in \mathcal{H}$ can be represented in the orthonormal eigenbasis of C 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. Note, in particular, that C is diagonal with respect to the coordinates in this eigenbasis. By the Karhunen-Loève expansion [DPZ92], a realization x from the Gaussian measure π_0 can be expressed by allowing the x_j to be independent random variables distributed as $x_j \sim N(0, \lambda_j^2)$. Thus, in the coordinates $\{x_j\}_{j=1}^{\infty}$, the prior π_0 in (1.6) has a product structure. For the particular MALA algorithm studied in this paper we assume that the eigenpairs $\{\lambda_j^2, \varphi_j\}$ are known so that sampling from π_0 is straightforward.

The measure π is absolutely continuous with respect to π_0 and hence any almost sure property under π_0 is also true under π. For example, it is a consequence of the law of large numbers that, almost surely with respect to π_0,

$$\frac{1}{N}\sum_{j=1}^{N} \frac{x_j^2}{\lambda_j^2} \to 1 \quad \text{as} \quad N \to \infty. \tag{2.2}$$

This, then, also holds almost surely with respect to π, implying that a typical draw from the target measure π must behave like a typical draw from π_0 in the large j coordinates. It is this structure which creates the link between the original results proven for product measures, and those we prove in this paper. In particular this structure enables us to exploit the product structure of the underlying Gaussian measure, when represented in the Karhunen-Loève coordinate system.

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

$$\langle x, y\rangle_r \stackrel{\text{def}}{=} \sum_{j=1}^{\infty} j^{2r} x_j y_j, \quad \|x\|_r^2 \stackrel{\text{def}}{=} \sum_{j=1}^{\infty} j^{2r}\, x_j^2. \tag{2.3}$$

Notice that $\mathcal{H}^0 = \mathcal{H}$ and so that the $\mathcal{H}$ inner-product and norm are given by $\langle\cdot,\cdot\rangle = \langle\cdot,\cdot\rangle_0$ and $\|\cdot\| = \|\cdot\|_0$. Furthermore $\mathcal{H}^r \subset \mathcal{H} \subset \mathcal{H}^{-r}$ for any $r > 0$. The Hilbert-Schmidt norm $\|\cdot\|_C$ is defined as

$$\|x\|_C = \sum_j \lambda_j^{-2} x_j^2.$$

For $r \in \mathbb{R}$, let $B_r : \mathcal{H} \mapsto \mathcal{H}$ denote the operator which is diagonal in the basis $\{\varphi_j\}$ with diagonal entries j^{2r}, *i.e.*,

$$B_r\, \varphi_j = j^{2r} \varphi_j$$

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:

$$\langle x, y\rangle_r = \langle B_r^{\frac{1}{2}} x, B_r^{\frac{1}{2}} y\rangle \qquad \text{and} \qquad \|x\|_r^2 = \|B_r^{\frac{1}{2}} x\|^2. \tag{2.4}$$

Since $\|B_r^{-1/2}\varphi_k\|_r = \|\varphi_k\| = 1$, we deduce that $\{B_r^{-1/2}\varphi_k\}_{k\geq 0}$ form 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

$$\mathrm{Tr}_{\mathcal{H}^r}(D) \stackrel{\text{def}}{=} \sum_{j=1}^{\infty} \langle (B_r^{-\frac{1}{2}} \varphi_j), D(B_r^{-\frac{1}{2}} \varphi_j)\rangle_r. \tag{2.5}$$

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 $\otimes_{\mathcal{H}^r}$ denote the outer product operator in $\mathcal{H}^r$ defined by

$$(x \otimes_{\mathcal{H}^r} y) z \stackrel{\text{def}}{=} \langle y, z\rangle_r\, x \qquad \forall x, y, z \in \mathcal{H}^r. \tag{2.6}$$

For an operator $L : \mathcal{H}^r \mapsto \mathcal{H}^l$, we denote the operator norm by

$$\|L\|_{\mathcal{L}(\mathcal{H}^r, \mathcal{H}^l)} \stackrel{\text{def}}{=} \sup_{\|x\|_r = 1} \|Lx\|_l.$$

For self-adjoint L and $r = l = 0$ this is, of course, the spectral radius of L.

Let π_0 denote a mean zero Gaussian measure on $\mathcal{H}$ with covariance operator C, *i.e.*, $\pi_0 \stackrel{\text{def}}{=} \mathrm{N}(0, C)$. If $x \stackrel{\mathcal{D}}{\sim} \pi_0$, then the x_j in (2.1) are independent $N(0, \lambda_j^2)$ Gaussians and we may write (Karhunen-Loéve)

$$x = \sum_{j=1}^{\infty} \lambda_j\, \rho_j\, \varphi_j, \ \text{ with } \rho_j \stackrel{\mathcal{D}}{\sim} \mathrm{N}(0,1) \text{ i.i.d.} \tag{2.7}$$

Because $\{B_r^{-1/2}\varphi_k\}_{k\geq 0}$ form an orthonormal basis for $\mathcal{H}^r$, we may also write (2.7) as

$$x = \sum_{j=1}^{\infty} (\lambda_j\, j^r)\, \rho_j\, (B_r^{-1/2} \varphi_j), \ \text{ with } \rho_j \stackrel{\mathcal{D}}{\sim} \mathrm{N}(0,1) \text{ i.i.d.} \tag{2.8}$$

Define

$$C_r = B_r\, C = B_r^{1/2}\, C\, B_r^{1/2}\,. \tag{2.9}$$

Let Ω denote the probability space in which the sequences $\{\rho_j\}_{j\in\mathbb{N}}$ are defined. Then the series in (2.7) may be shown to converge in $L^2(\Omega; \mathcal{H}^r)$ as long as

$$\mathrm{Tr}_{\mathcal{H}^r}(C_r) = \sum_{j=1}^{\infty} \lambda_j^2\, j^{2r} < \infty.$$

The induced distribution of π_0 on $\mathcal{H}^r$ is then identical to that of a centered Gaussian measure on $\mathcal{H}^r$ with covariance operator C_r in the sense that, if $\xi \overset{\mathcal{D}}{\sim} \pi_0$, then $\mathbb{E}\big[\langle \xi, u\rangle_r \langle \xi, v\rangle_r\big] = \langle u, C_r v\rangle_r$ for $u, v \in \mathcal{H}^r$. Thus in what follows, we freely alternate between the Gaussian measures $\mathrm{N}(0, C)$ on $\mathcal{H}$ and $\mathrm{N}(0, C_r)$ on $\mathcal{H}^r$.

Our goal is to sample from a measure π on $\mathcal{H}$, given by (1.6) with π_0 as constructed above. Frequently in applications the function Ψ may not be defined on all of $\mathcal{H}$, but only on a subspace $\mathcal{H}^r \subset \mathcal{H}$, for some exponent $r > 0$. Even though the Gaussian measure π_0 is defined on $\mathcal{H}$, depending on the decay of the eigenvalues of C, there exists an entire range of values of r such that $\mathrm{Tr}_{\mathcal{H}^r}(C_r) < \infty$ so that the measure π_0 has full support on $\mathcal{H}^r$, *i.e.*, $\pi_0(\mathcal{H}^r) = 1$. From now onwards we fix a distinguished exponent $s > 0$ and assume that $\Psi : \mathcal{H}^s \to \mathbb{R}$ and that $\mathrm{Tr}_{\mathcal{H}^s}(C_s) < \infty$. Then $\pi_0 \overset{\mathcal{D}}{\sim} \mathrm{N}(0, C)$ on $\mathcal{H}$ and $\pi(\mathcal{H}^s) = 1$. For ease of notation we define

$$\hat{\varphi}_k = B_s^{-\frac{1}{2}} \varphi_k$$

so that we may view π_0 as a Gaussian measure $\mathrm{N}(0, C_s)$ on $\big(\mathcal{H}^s, \langle\cdot,\cdot\rangle_s\big)$, and $\{\hat{\varphi}_k\}$ form an orthonormal basis of $\big(\mathcal{H}^s, \langle\cdot,\cdot\rangle_s\big)$.

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

$$W(t) = \sum_j (j^s \lambda_j)\, \beta_j(t) \hat{\varphi}_j \tag{2.10}$$

is a Brownian motion in $\mathcal{H}^s$ with covariance operator C_s; equivalently, this same process $\{W(t)\}_{t\geq 0}$ can be described as a Brownian motion in $\mathcal{H}$ with covariance operator equal to C since Equation (2.10) may also be expressed as

$$W(t) = \sum_j \lambda_j \beta_j(t) \varphi_j$$

.

To approximate π and π_0 by finite N-dimensional measures π^N and π_0^N living in

$$X^N \overset{\text{def}}{=} \mathrm{span}\{\hat{\varphi}_1, \hat{\varphi}_2, \cdots, \hat{\varphi}_N\},$$

we introduce the orthogonal projection on X^N denoted by $P^N : \mathcal{H}^s \mapsto X^N \subset \mathcal{H}^s$. The function $\Psi^N : X^N \mapsto \mathbb{R}$ is defined by $\Psi^N \overset{\text{def}}{=} \Psi \circ P^N$. The probability distribution π^N supported by X^N is defined by

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

where M_{Ψ^N} is a normalization constant and π_0^N is a Gaussian measure supported on $X^N \subset \mathcal{H}^s$ and defined by the property that $x \overset{\mathcal{D}}{\sim} \pi_0^N$ is given by

$$x = \sum_{j=1}^N \lambda_j \xi_j \varphi_j = (C^N)^{\frac{1}{2}} \xi^N$$

where ξ_j are i.i.d standard Gaussian random variables, $\xi^N = \sum_{j=1}^N \xi_j \varphi_j$ and $C^N = P^N \circ C \circ P^N$. Notice that π^N has Lebesgue density [1] on X^N equal to

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

where the Hilbert-Schmidt norm $\|\cdot\|_{C^N}$ on X^N is given by the scalar product

$$\langle u, v\rangle_{C^N} = \langle u, (C^N)^{-1} v\rangle \qquad \forall u, v \in X^N.$$

The operator C^N is invertible on X^N because the eigenvalues of C are assumed to be strictly positive.

[1] 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 $\mathbb{R}^N$

2.2. The Algorithm

Define

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

and

$$\mu^N(x) = -\Big(P^N x + C^N\nabla\Psi^N(x)\Big). \tag{2.14}$$

Setting $h(\ell) = 1$ in (1.7) we see that the measure π given by (1.6) is invariant with the respect to the diffusion process

$$\frac{dz}{dt} = \mu(z) + \sqrt{2}\,\frac{dW}{dt}, \quad z(0) = z^0$$

where W is a Brownian motion (see [DPZ92]) in $\mathcal{H}$ with covariance operator C. Similarly, the measure π^N given by (2.11) is invariant with respect to the diffusion process

$$\frac{dz}{dt} = \mu^N(z) + \sqrt{2}\,\frac{dW^N}{dt}, \quad z(0) = z^0 \tag{2.15}$$

where W^N is a Brownian motion in $\mathcal{H}$ with covariance operator C^N. The idea of the MALA algorithm is to make a proposal based on Euler-Maruyama discretization of a diffusion process which is invariant with respect to the target. To this end we consider, from state $x \in X^N$, proposals $y \in X^N$ given by

$$y - x = \delta\,\mu^N(x) + \sqrt{2\delta}\,(C^N)^{\frac{1}{2}}\xi^N \qquad \text{where} \qquad \delta = \ell N^{-\frac{1}{3}} \tag{2.16}$$

with $\xi^N = \sum_{i=1}^N \xi_i\,\varphi_i$ and $\xi_i \overset{\mathcal{D}}{\sim} \mathrm{N}(0,1)$. Thus δ is the time-step in an Euler-Maruyama discretization of (2.15). 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. 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 $\{x^{k,N}\}_{k\geq 0}$ 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 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

$$\alpha^N(x,\xi^N) = 1 \wedge \frac{\pi^N(y)T^N(y,x)}{\pi^N(x)T^N(x,y)} = 1 \wedge e^{Q^N(x,\xi^N)}. \tag{2.17}$$

where the function T^N given by

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

is the density of the Langevin proposals. The local mean acceptance probability $\alpha^N(x)$ is defined by

$$\alpha^N(x) = \mathbb{E}_x\big[\alpha^N(x,\xi^N)\big]. \tag{2.18}$$

The Markov chain for $x^N = \{x^{k,N}\}$ can also be expressed as

$$\begin{cases} y^{k,N} &= x^{k,N} + \delta\mu^N(x^{k,N}) + \sqrt{2\delta}\,(C^N)^{\frac{1}{2}}\,\xi^{k,N} \\ x^{k+1,N} &= \gamma^{k,N} y^{k,N} + (1-\gamma^{k,N})\,x^{k,N} \end{cases} \tag{2.19}$$

where $\xi^{k,N}$ are i.i.d samples distributed as ξ^N and $\gamma^{k,N} = \gamma^N(x^{k,N}, \xi^{k,N})$ is a Bernoulli random variable with 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 (2.17) may be expressed as

$$Q^N(x, \xi^N) = -\frac{1}{2}\Big(\|y\|^2_{C^N} - \|x\|^2_{C^N}\Big) - \Big(\Psi^N(y) - \Psi^N(x)\Big) \\ - \frac{1}{4\delta}\Big\{\|x - y - \delta\mu^N(y)\|^2_{C^N} - \|y - x - \delta\mu^N(x)\|^2_{C^N}\Big\}. \tag{2.20}$$

As will be seen in the next section, a key idea behind our proof 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 $\{x^{k,N}\}$ 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 $N \to \infty$.

3. Diffusion Limit and Proof Strategy

In this section we state the main theorem, set it in context, and explain the proof technique that we adopt.

3.1. Main Theorem

Consider the constant $\alpha(\ell) = \mathbb{E}\big[1 \wedge e^{Z_\ell}\big]$ where $Z_\ell \overset{\mathcal{D}}{\sim} \mathrm{N}(-\frac{\ell^3}{4}, \frac{\ell^3}{2})$ and define the speed function

$$h(\ell) = \ell\alpha(\ell). \tag{3.1}$$

The quantity $\alpha(\ell)$ represents the limiting expected acceptance probability of the MALA algorithm while $h(\ell)$ is the asymptotic speed function of the limiting diffusion. The following is the main result of this article (it is stated in full, with conditions, as Theorem 5.2):

Main Theorem: *Let the initial condition $x^{0,N}$ of the MALA algorithm be such that $x^{0,N} \overset{\mathcal{D}}{\sim} \pi^N$ and let $z^N(t)$ be a piecewise linear, continuous interpolant of the MALA algorithm (2.19) as defined in (1.4) with $\Delta t = N^{-\frac{1}{3}}$. Then, for any $T > 0$, $z^N(t)$ converges weakly in $C([0,T], \mathcal{H}^s)$ to the diffusion process $z(t)$ given by Equation (1.7) with $z(0) \overset{\mathcal{D}}{\sim} \pi$.*

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 tuning the average acceptance probability to 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 (2.19) 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. The second implication follows from Equation (1.7) for $z(t)$, together with the definition (3.1) of $h(\ell)$ itself. The maximum of the speed of the limiting diffusion $h(\ell)$ occurs at an average acceptance probability of $\alpha^\star = 0.574$, to three decimal places. Thus, remarkably, the optimal acceptance probability identified in [RR98] for product measures, is also optimal for the non-product measures studied in this paper.

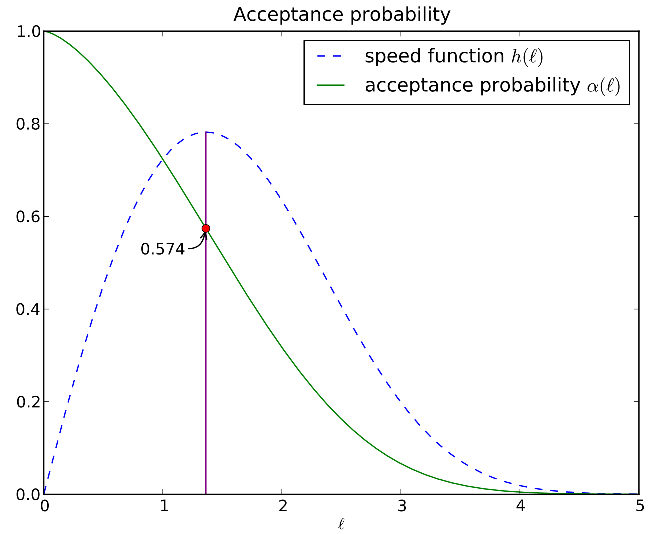

FIG 1. *Optimal acceptance probability* = 0.574

3.2. 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 Lipshitz 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 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. $\xi^k \overset{\mathcal{D}}{\sim} \mathrm{N}(0,1)$ converges weakly, when interpolated using a time-acceleration factor of $N^{\frac{1}{3}}$, to the scalar diffusion $dz(t) = \ell\mu\big(z(t)\big)\, dt + \sqrt{2\ell}\, dW(t)$ where W is a Brownian motion with variance $\mathrm{Var}\big(W(t)\big) = Ct$. 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 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

$$dz(t) = h(\ell)\mu\big(z(t)\big)\, dt + \sqrt{2h(\ell)}\, dW(t) \qquad \text{where} \qquad h(\ell) = \ell\alpha(\ell).$$

This shows that the Bernoulli random variables $\{\gamma^k\}$ have slowed down the original Markov chain by a factor $\alpha(\ell)$.

The proof of Theorem 5.2 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 (2.16) 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 (2.18) 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 (2.20): the main observation is that $Q^N(x,\xi^N)$ can be approximated by a Gaussian random variables

$$Q^N(x,\xi^N) \;\approx\; Z_\ell \tag{3.2}$$

where $Z_\ell \stackrel{\mathcal{D}}{\sim} \mathrm{N}(-\frac{\ell^3}{4}, \frac{\ell^3}{2})$. These approximations are made rigorous in Lemma 4.1 and Lemma 4.2. 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

$$\alpha(\ell) = \mathbb{E}\big[1 \wedge e^{Z_\ell}\big]. \tag{3.3}$$

Thus, the limiting acceptance probability of the MALA algorithm is as given in Equation (3.3).
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

$$d^N(x) = \big(h(\ell)\Delta t\big)^{-1}\mathbb{E}\big[x^{1,N} - x^{0,N}\,|x^{0,N} = x\big] \tag{3.4}$$

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

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

The normalization constant $h(\ell)$ defined in Equation (3.1) 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

$$x^{k+1,N} - x^{k,N} = h(\ell)\Delta t d^N(x^{k,N}) + \sqrt{2h(\ell)\Delta t}\,\Gamma^{k,N}. \tag{3.6}$$

Lemma 4.4 and Lemma 4.5 exploit the Gaussian behaviour of $Q^N(x,\xi^N)$ described in Equation (3.2) in order to give quantitative versions of the following approximations,

$$d^N(x) \;\approx\; \mu(x) \qquad \text{and} \qquad \Gamma^{k,N} \;\approx\; \mathrm{N}(0,C) \tag{3.7}$$

where $\mu(x) = -\Big(x + C\nabla\Psi(x)\Big)$. From Equation (3.6) it follows that for large N the evolution of the Markov chain ressembles the Euler discretization of the limiting diffusion (1.7). 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

$$W^N(t) = \sqrt{\Delta t}\,\sum_{j=0}^{k}\Gamma^{j,N} \;+\; \frac{t - k\Delta t}{\sqrt{\Delta t}}\,\Gamma^{k+1,N} \qquad \text{for} \qquad k\Delta t \leq t < (k+1)\Delta t. \tag{3.8}$$

The sequence of processes $\{W^N\}$ converges weakly in $C([0;T],\mathcal{H}^s)$ to a Brownian motion W in $\mathcal{H}^s$ with covariance operator equal to C_s. Indeed, Proposition 4.7 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 (1.7) 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

$$z(t) = z^0 - h(\ell) \int_0^t \mu(z)\, du + \sqrt{2h(\ell)}\, W(t) \qquad \forall t \in [0,T]. \tag{3.9}$$

Notice that $z = \Theta(z^0, W)$ solves the SPDE (1.7). The Itô map Θ is continuous, essentially because the noise in (1.7) is additive (does not depend on the state z). The piecewise constant interpolant $\bar{z}^N$ of x^N is defined by

$$\bar{z}^N(t) = x^k \qquad \text{for} \qquad k\Delta t \le t < (k+1)\Delta t. \tag{3.10}$$

The continuous piecewise linear interpolant z^N defined in Equation (1.4) satisfies

$$z^N(t) = x^{0,N} + h(\ell) \int_0^t d^N(\bar{z}^N(u))\, du + \sqrt{2h(\ell)}\, W^N(t) \qquad \forall t \in [0,T]. \tag{3.11}$$

Using the closeness of $d^N(x)$ and $\mu(x)$, 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

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

so that $z^N = \Theta(x^{0,N}, \widehat{W}^N)$. 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 and is stated precisely in Theorem 5.2.

3.3. Assumptions

Here we give assumptions on the decay of the eigenvalues of C – that they decay like $j^{-\kappa}$ for some $\kappa > \frac{1}{2}$ – and then assumptions on Ψ that are linked to the eigenvalues of C through the need for the change of measure in (2.11) to be π_0−measurable. Let $\nabla\Psi : \mathcal{H}^s \to \mathcal{R}$. Then 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 $\Psi(x)$ 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 3.1. *The operator C and functional Ψ satisfy the following:*

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

$$\lambda_j \asymp j^{-\kappa}$$

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

$$M_1 \le \Psi(x) \le M_2\Big(1 + \|x\|_s^2\Big) \;\; \forall x \in \mathcal{H}^s \tag{3.12}$$

$$\|\nabla\Psi(x)\|_{-s} \le M_3\Big(1 + \|x\|_s\Big) \;\; \forall x \in \mathcal{H}^s \tag{3.13}$$

$$\|\partial^2\Psi(x)\|_{\mathcal{L}(\mathcal{H}^s,\mathcal{H}^{-s})} \le M_4 \qquad \forall x \in \mathcal{H}^s. \tag{3.14}$$

Remark 3.2. *The condition* $\kappa > \frac{1}{2}$ *ensures that* C *is trace class in* $\mathcal{H}$*. In fact* C_r *is trace-class in* $\mathcal{H}^r$ *for any* $r < \kappa - \frac{1}{2}$*. It follows that the* $\mathcal{H}^r$ *norm of* $x \overset{\mathcal{D}}{\sim} N(0, C)$ *is* π_0*-almost surely finite for any* $r < \kappa - \frac{1}{2}$ *because* $\mathbb{E}\Big(\|x\|_r^2\Big) = \mathrm{Tr}_{\mathcal{H}^r}(C_r) < \infty$*.*

Remark 3.3. *The functional* $\Psi(x) = \frac{1}{2}\|x\|_s^2$ *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 C decrease as $\lambda_j \asymp j^{-\kappa}$, the operator C has a smoothing effect: $C^\alpha h$ gains $2\alpha\kappa$ orders of regularity.

Lemma 3.4. *For any vector* $h \in \mathcal{H}$ *and exponent* $\beta \in \mathbb{R}$*,* $\|h\|_C \asymp \|h\|_\kappa$ *and* $\|C^\alpha h\|_\beta \asymp \|h\|_{\beta - 2\alpha\kappa}$*.*

Proof. Under Assumption 3.1 the eigenvalues satisfy $\lambda_j \asymp j^{-\kappa}$. Hence

$$\|h\|_C^2 = \sum_{j\ge 1} \lambda_j^{-2} h_j^2 \asymp \sum_{j\ge 1} j^{2\kappa} h_i^2 = \|h\|_\kappa^2.$$

Also,

$$\|C^\alpha h\|_\beta^2 = \sum_{j\ge 1} j^{2\beta}\langle C^\alpha h, \varphi_j\rangle^2 = \sum_{j\ge 1} j^{2\beta}\big(\lambda_j^{2\alpha} h_j\big)^2 \asymp \sum_{j\ge 1} j^{2\beta} j^{-4\kappa\alpha} h_j^2 = \|h\|_{\beta-2\alpha\kappa}^2,$$

which concludes the proof of the lemma. □

For simplicity we assume throughout this paper that $\Psi^N(\cdot) = \Psi(P^N \cdot)$. From this definition it follows that $\nabla\Psi^N(x) = P^N \nabla\Psi(P^N x)$ and $\partial^2\Psi^N(x) = P^N \partial^2\Psi(P^N x) P^N$. Other approximations could be handled similarly. The function Ψ^N may be shown to satisfy the following:

Assumptions 3.5. *The functions* $\Psi^N : \mathcal{H}^s \to \mathbb{R}$ *satisfy the same conditions imposed on* Ψ *given by Equations* (3.12),(3.12) *and* (3.14) *with the same constants uniformly in* N *.*

Notice also that the above assumptions on Ψ and Ψ^N imply [2] that for all $x, y \in \mathcal{H}^s$,

$$\Psi^N(y) = \Psi^N(x) + \langle\nabla\Psi^N(x), y - x\rangle + \mathrm{rem}(x, y) \qquad \text{with} \qquad \mathrm{rem}(x, y) \le M_5\|x - y\|_s^2 \tag{3.15}$$

for some constant $M_5 > 0$. The following result will be used repeatedly in the sequel:

Lemma 3.6. ($C\nabla\Psi : \mathcal{H}^s \to \mathcal{H}^s$ is globally Lipschitz)
Under Assumptions 3.1 the functional $C\Psi$ *is globally Lipschitz on* $\mathcal{H}^s$*: there exists* $M_6 > 0$ *satisfying*

$$\|C\nabla\Psi(x) - C\nabla\Psi(y)\|_s \le M_6\,\|x - y\|_s \qquad \forall x, y \in \mathcal{H}^s.$$

Proof. Because $s - 2\kappa < -s$ we have $\|C\nabla\Psi(x) - C\nabla\Psi(y)\|_s \asymp \|\nabla\Psi(x) - \nabla\Psi(y)\|_{s-2\kappa} \lesssim \|\nabla\Psi(x) - \nabla\Psi(y)\|_{-s}$ so that it suffices to prove that $\|\nabla\Psi(y) - \nabla\Psi(y)\|_{-s} \lesssim \|y - x\|_s$. Assumption 3.1 states that $\partial^2\Psi$ is uniformly bounded in $\mathcal{L}(\mathcal{H}^s, \mathcal{H}^{-s})$ so that

$$\|\nabla\Psi(y) - \nabla\Psi(y)\|_{-s} = \Big\|\int_0^1 \partial^2\Psi\big(x + t(y - x)\big) \cdot (y - x)\, dt\Big\|_{-s} \tag{3.16}$$

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

$$\leq \int_0^1 \|\partial^2\Psi\big(x+t(y-x)\big)\cdot(y-x)\|_{-s}\,dt$$
$$\leq M_4\int_0^1 \|y-x\|_s\,dt,$$

which finishes the proof of Lemma 3.6. □

Remark 3.7. *Lemma 3.6 shows in particular that the function $\mu:\mathcal{H}^s\to\mathcal{H}^s$ defined by (2.13) is globally Lipschitz on $\mathcal{H}^s$. The same proof shows that $C^N\nabla\Psi^N:\mathcal{H}^s\to\mathcal{H}^s$ and $\mu^N:\mathcal{H}^s\to\mathcal{H}^s$ given by (2.14) are globally Lipschitz and that the Lipschitz constants can be chosen uniformly in N.*

The globally Lipschitz property of μ ensures that the solution of the Langevin diffusion (1.7) is defined for all time. The global Lipschitzness of μ is also used to show the continuity of the Itô map $\Theta\colon \mathcal{H}^s\times C([0,T],\mathcal{H}^s)\to C([0,T],\mathcal{H}^s)$ which maps $(Z^0,W)\in\mathcal{H}^s\times C([0,T],\mathcal{H}^s)$ to the unique solution of the integral equation (3.9) below.
We now show that the sequence of functions $\mu^N:\mathcal{H}^s\to\mathcal{H}^s$ defined by

$$\mu^N(x)\ \stackrel{\text{def}}{=}\ -\Big(P^Nx+C^N\nabla\Psi^N(x)\Big)=-\Big(P^Nx+P^NCP^N\nabla\Psi(P^Nx)\Big)$$

converge to μ in an appropriate sense.

Lemma 3.8. **(μ^N converges π_0-almost surely to μ)**
Let Assumption 3.1 hold. The sequences of functions μ^N 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

$$\lim_{N\to\infty}\|P^Nx-x\|_s=0 \tag{3.17}$$
$$\lim_{N\to\infty}\|CP^N\nabla\Psi(P^Nx)-C\nabla\Psi(x)\|_s=0. \tag{3.18}$$

- Let us prove Equation (3.17). For $x\in\mathcal{H}^s$ we have $\sum_{j\geq1}j^{2s}x_j^2<\infty$ so that

$$\lim_{N\to\infty}\|P^Nx-x\|_s^2\ =\ \lim_{N\to\infty}\sum_{j=N+1}^{\infty}j^{2s}x_j^2=0. \tag{3.19}$$

- Let us prove (3.18). The triangle inequality shows that

$$\|CP^N\nabla\Psi(P^Nx)-C\nabla\Psi(x)\|_s\leq\|CP^N\nabla\Psi(P^Nx)-CP^N\nabla\Psi(x)\|_s+\|CP^N\nabla\Psi(x)-C\nabla\Psi(x)\|_s$$

  The same proof as Lemma 3.6 reveals that $CP^N\nabla\Psi:\mathcal{H}^s\to\mathcal{H}^s$ is globally Lipschitz, with a Lipschitz constant that can be chosen independent from N. Consequenly, Equation (3.19) shows that

$$\|CP^N\nabla\Psi(P^Nx)-CP^N\nabla\Psi(x)\|_s\lesssim\|P^Nx-x\|_s\to0.$$

  Also, $z=\nabla\Psi(x)\in\mathcal{H}^{-s}$ so that $\|\nabla\Psi(x)\|_{-s}^2=\sum_{j\geq1}j^{-2s}z_j^2<\infty$. The eigenvalues of C satisfy $\lambda_j^2\asymp j^{-2\kappa}$ with $s<\kappa-\frac{1}{2}$. Consequently,

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

□

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

Lemma 3.9. *Consider y given by (2.16). Under Assumptions 3.5, 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 3.5 the functional μ^N is globally Lipschitz on $\mathcal{H}^s$, with Lipschitz constant that can be chosen independent from N. Thus

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

We have $\mathbb{E}^{\pi^0}\Big[\|C^{\frac{1}{2}}\xi^N\|_s^p\Big] \leq \mathbb{E}^{\pi^0}\Big[\|\zeta\|_s^p\Big] < \infty$, where $\zeta \overset{\mathcal{D}}{\sim} \mathrm{N}(0,C)$. Fernique's theorem [DPZ92] shows that $\mathbb{E}^{\pi^0}\Big[\|\zeta\|_s^p\Big] < \infty$. Consequently, $\mathbb{E}^{\pi^0}\Big[\|C^{\frac{1}{2}}\xi^N\|_s^p\Big]$ 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 3.10. (Finite dimensional approximation π^N of π)
Under the Assumptions 3.5 on Ψ^N 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. By definition, $M_{\Psi^N}^{-1} = \int_{\mathcal{H}} \exp\{-\Psi^N(x)\}\pi_0(dx) \geq \int_{\mathcal{H}} \exp\{-M_2(1+\|x\|_s^2)\}\pi_0(dx) \geq e^{-2M_2}\mathbb{P}(\|x\|_s \leq 1)$ and therefore $\inf\{M_{\Psi^N}^{-1} : N \in \mathbb{N}\} > 0$, which shows that the normalization constants M_{Ψ^N} are uniformly bounded. Because Ψ^N is uniformly lower bounded by a constant M_1, for any $f:\mathcal{H}\mapsto\mathbb{R}$ we have $\mathbb{E}^{\pi^N}|f(x)| \leq \sup_N\big\{M_{\Psi^N}\big\}\,\mathbb{E}^{\pi_0}\Big[e^{-\Psi^N(x)}|f(x)|\Big] \leq e^{-M_1}\sup_N\big\{M_{\Psi^N}\big\}\,\mathbb{E}^{\pi_0}|f(x)|$.

Let us now 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

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

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 $g(P^Nx)M_{\Psi^N}e^{-\Psi(P^Nx)}$ 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 (3.13) we have

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

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 $\mathbb{E}^{\pi^0}\big[\|x\|_s^p\big] < \infty$. It thus follows from Lemma 3.10 that for any $p \geq 0$

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

This estimate is repeatedly used in the sequel.

4. Key Estimates

In this section we describe a Gaussian approximation within the acceptance probability for the MALA algorithm and, using this, quantify the mean one-step drift and diffusion.

4.1. Gaussian approximation of Q^N

Recall the quantity Q^N defined in Equation (2.20). This section proves that Q^N has a Gaussian behaviour in the sense that

$$Q^N(x,\xi^N) = Z^N(x,\xi^N) \ + \ i^N(x,\xi^N) \ + \ \mathbf{e}^N(x,\xi^N) \tag{4.1}$$

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

$$Z^N(x,\xi^N) = -\frac{\ell^3}{4} - \frac{\ell^{\frac{3}{2}}}{\sqrt{2}} N^{-\frac{1}{2}} \sum_{j=1}^N \lambda_j^{-1}\xi_j x_j \tag{4.2}$$

$$i^N(x,\xi^N) = \frac{1}{2}(\ell\Delta t)^2\Big(\|x\|^2_{C^N} - \|(C^N)^{\frac{1}{2}}\xi^N\|^2_{C^N}\Big) \tag{4.3}$$

with i^N and e^N small. Thus the principal contribution 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, by virtue of (2.2), we have that almost surely with respect to x, the Gaussian distribution approaches that 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 4.2 we then quantify the convergence of $Z^N(x,\xi^N)$ to Z_ℓ.

Lemma 4.1. (Gaussian Approximation) *Let $p \ge 1$ be an integer. Under Assumptions 3.1 and 3.5 the error terms i^N and $\boldsymbol{e}^N$ in the Gaussian approximation (4.1) satisfy*

$$\Big(\mathbb{E}^{\pi^N}\big[|i^N(x,\xi^N)|^p\big]\Big)^{\frac{1}{p}} = \mathcal{O}(N^{-\frac{1}{6}}) \qquad \textit{and} \qquad \Big(\mathbb{E}^{\pi^N}\big[|\boldsymbol{e}^N(x,\xi^N)|^p\big]\Big)^{\frac{1}{p}} = \mathcal{O}(N^{-\frac{1}{3}}). \tag{4.4}$$

Proof. For notational clarity, without loss of generality, we suppose $p = 2q$. The quantity Q^N is defined in Equation (2.20) 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

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

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.

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

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

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 $\mathcal{O}(N^{-\frac{1}{3}})$:

$$\Big(\mathbb{E}^{\pi^N}[I_2^{2q}]\Big)^{\frac{1}{2q}} = \mathcal{O}(N^{-\frac{1}{3}}) \qquad \text{and} \qquad \Big(\mathbb{E}^{\pi^N}[I_3^{2q}]\Big)^{\frac{1}{2q}} = \mathcal{O}(N^{-\frac{1}{3}}). \tag{4.5}$$

- We expand I_2 and use the bound on the remainder of the Taylor expansion of Ψ described in Equation (3.15),

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

Equation (3.15) and Lemma 3.9 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 $\mathbb{E}^{\pi^N}[\|x\|_s^{4q}] \lesssim \mathbb{E}^{\pi_0}[\|x\|_s^{4q}] < \infty$. Equation (3.16) proves that $\|\nabla\Psi^N(y) - \nabla\Psi^N(x)\|_{-s} \lesssim \|y-x\|_s$. Consequently, Lemma 3.9 shows that

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

Assumption 3.5 states 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 3.6 states $C^N\nabla\Psi^N : \mathcal{H}^s \to \mathcal{H}^s$ is globally Lipschitz, with a Lipschitz constant that can be chosen uniformly in N. Therefore,

$$\|C^N\nabla\Psi^N(z)\|_s \lesssim 1 + \|z\|_s. \tag{4.6}$$

Since $\|C^N\nabla\Psi^N(z)\|_{C^N}^2 = \langle\nabla\Psi^N(z), C^N\nabla\Psi^N(z)\rangle$, the bound (3.13) gives

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

which concludes the proof of Equation (4.5).

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 (2.16) to expand I_1. For $x \in X^N$ we have $P^N x = x$. Therefore, for $x \in X^N$,

$$I_1 = -\frac{\ell\Delta t}{4}\Big(\|(1-\ell\Delta t)x - \ell\Delta t\, C^N\nabla\Psi^N(x) + \sqrt{2\ell\Delta t}\,(C^N)^{\frac{1}{2}}\xi^N\|_{C^N}^2 - \|x\|_{C^N}^2\Big)$$

$$= Z^N(x,\xi^N) \;+\; i^N(x,\xi^N) \;+\; B_1 \;+\; B_2 \;+\; B_3 \;+\; B_4.$$

with $Z^N(x,\xi^N)$ and $i^N(x,\xi^N)$ given by Equation (4.2) and (4.3) and

$$B_1 = \tfrac{\ell^3}{4}\big(1 - \tfrac{\|x\|^2_{C^N}}{N}\big) \qquad\qquad B_2 = -\tfrac{\ell^3}{4}N^{-1}\Big\{\|C^N\nabla\Psi^N(x)\|^2_{C^N} + 2\langle x, \nabla\Psi^N(x)\rangle\Big\}$$
$$B_3 = \tfrac{\ell^{\frac{5}{2}}}{\sqrt{2}}N^{-\frac{5}{6}}\langle x + C^N\nabla\Psi^N(x), (C^N)^{\frac{1}{2}}\xi^N\rangle_{C^N} \qquad\qquad B_4 = \tfrac{\ell^2}{2}N^{-\frac{2}{3}}\langle x, \nabla\Psi^N(x)\rangle.$$

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 $\mathcal{O}(N^{-\frac{1}{3}})$,

$$\Big(\mathbb{E}^{\pi^N}\big[B_j^{2q}\big]\Big)^{\frac{1}{2q}} \;=\; \mathcal{O}(N^{-\frac{1}{3}}) \qquad\qquad j = 1,\ldots,4. \tag{4.7}$$

- Lemma 3.10 shows that $\mathbb{E}^{\pi^N}[(1 - \frac{\|x\|^2_{C^N}}{N})^{2q}] \lesssim \mathbb{E}^{\pi_0}[(1 - \frac{\|x\|^2_{C^N}}{N})^{2q}]$. Under π_0,

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

  where $\rho_1,\ldots,\rho_N$ are i.i.d $\mathrm{N}(0,1)$ Gaussian random variables. Consequently, $\mathbb{E}^{\pi^N}[B_1^{2q}]^{\frac{1}{2q}} = \mathcal{O}(N^{-\frac{1}{2}})$.
- The term $\|C^N\nabla\Psi^N(x)\|^{2q}_{C^N}$ has already been bounded while proving $\mathbb{E}^{\pi^N}[I_3^{2q}] \lesssim \Big(N^{-\frac{1}{3}}\Big)^{2q}$. Equation (3.13) gives the bound $\|\nabla\Psi^N(x)\|_{-s} \lesssim 1 + \|x\|_s$ and shows that $\mathbb{E}^{\pi^N}\big[\langle x, \nabla\Psi^N(x)\rangle^{2q}\big]$ is uniformly bounded as a function of N. Consequently,

  $$\mathbb{E}^{\pi^N}\big[B_2^{2q}\big]^{\frac{1}{2q}} \;=\; \mathcal{O}(N^{-1}).$$
- We have $\langle C^N\nabla\Psi^N(x), (C^N)^{\frac{1}{2}}\xi^N\rangle_{C^N} = \langle\nabla\Psi^N(x), (C^N)^{\frac{1}{2}}\xi^N\rangle$ so that

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

  By Lemma 3.10, one can suppose $x \overset{\mathcal{D}}{\sim} \pi_0$,

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

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

  $$\Big(\mathbb{E}^{\pi^N}\big[B_3^{2q}\big]\Big)^{\frac{1}{2q}} \;=\; \mathcal{O}(N^{-\frac{5}{6}+\frac{1}{2}}) \;=\; \mathcal{O}(N^{-\frac{1}{3}}).$$
- The bound $\|\nabla\Psi^N(x)\|_{-s} \lesssim 1 + \|x\|_s$ ensures that $\Big(\mathbb{E}^{\pi^N}\big[B_4^{2q}\big]\Big)^{\frac{1}{2q}} = \mathcal{O}(N^{-\frac{2}{3}})$.

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 (4.5) and (4.7) show that $\mathbf{e}^N$ satisfies

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

We now prove that i^N is $\mathcal{O}(N^{-\frac{1}{6}})$. By Lemma 3.10, $\mathbb{E}^{\pi^N}[i^N(x,\xi^N)^{2q}] \lesssim \mathbb{E}^{\pi_0}[i^N(x,\xi^N)^{2q}]$. If $x \overset{\mathcal{D}}{\sim} \pi_0$ we have

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

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

$$\Big(\mathbb{E}^{\pi^N}\big[i^N(x,\xi^N)^{2q}\big]\Big)^{\frac{1}{2q}} \;=\; \mathcal{O}(N^{-\frac{2}{3}+\frac{1}{2}}) = \mathcal{O}(N^{-\frac{1}{6}}), \tag{4.8}$$

which ends the proof of Lemma 4.1 □

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

Lemma 4.2. (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 $\boldsymbol{e}_\star^N(x,\xi)$ satisfying*

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

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

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

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

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

The function F satisfies

$$\big|F(\mu_1,\sigma_1) \;-\; F(\mu_2,\sigma_2)\big| \;\lesssim\; |\mu_2-\mu_1| \;+\; |\sigma_2-\sigma_1|. \tag{4.9}$$

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

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

Therefore, using Lemma 3.10,

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

$$= \mathbb{E}^{\pi^N}\Big[\Big|F(-\frac{\ell^3}{4}, \mathrm{Var}\big[Z^N(x,\xi^N)\big]^{\frac12}) - F(-\frac{\ell^3}{4}, \mathrm{Var}\big[Z_\ell\big]^{\frac12})\Big|^p\Big] \;+\; \mathbb{E}^{\pi^N}\big[|\mathbf{e}^N_\star(x,\xi^N)|^p\big]$$

$$\lesssim \mathbb{E}^{\pi^N}\Big[\Big|\mathrm{Var}\big[Z^N(x,\xi^N)\big]^{\frac12} - \mathrm{Var}\big[Z_\ell\big]^{\frac12}\Big|^p\Big] \;+\; \mathbb{E}^{\pi^N}\big[|\mathbf{e}^N_\star(x,\xi^N)|^p\big]$$

$$\lesssim \mathbb{E}^{\pi_0}\Big|\Big\{\frac{\|x\|^2_{C^N}}{N}\Big\}^{\frac12} - 1\Big|^p \;+\; \mathbb{E}^{\pi^N}\big[|\mathbf{e}^N_\star(x,\xi^N)|^p\big] \;\to\; 0$$

In the last step we have used the fact that if $x \overset{\mathcal{D}}{\sim} \pi_0$ then $\frac{\|x\|^2_{C^N}}{N} \overset{\mathcal{D}}{\sim} \frac{\rho_1^2+\ldots+\rho_N^2}{N}$ where $\rho_1,\ldots,\rho_N$ are i.i.d Gaussian random variables $\mathrm{N}(0,1)$ so that $\mathbb{E}^{\pi_0}\Big|\Big\{\frac{\|x\|^2_{C^N}}{N}\Big\}^{\frac12} - 1\Big|^p \to 0$. □

Corollary 4.3. *Let $p \geq 1$ be a positive. The local mean acceptance probability $\alpha^N(x)$ defined in Equation (2.18) satisfies*

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

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

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

with $\mathbf{e}^N_\star(x,\xi^N) = i^N(x,\xi^N) + \mathbf{e}^N(x,\xi^N)$. Lemma 4.1 shows that $\lim_{N\to\infty} \mathbb{E}^{\pi^N}[\mathbf{e}^N_\star(x,\xi)^p] = 0$ and therefore Lemma 4.2 gives the conclusion. □

4.2. Drift approximation

This section proves that the approximate drift function $d^N : \mathcal{H}^s \to \mathcal{H}^s$ defined in Equation (3.4) converges to the drift function $\mu : \mathcal{H}^s \to \mathcal{H}^s$ of the limiting diffusion (1.7).

Lemma 4.4. (Drift Approximation): *Let Assumptions 3.1 and 3.5 hold. The drift function $d^N : \mathcal{H}^s \to \mathcal{H}^s$ converges to μ in the sense that*

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

Proof. The approximate drift d^N is given by Equation (3.4). The definition of the local mean acceptance probability $\alpha^N(x)$ given by Equation (2.18) show that d^N can also be expressed as

$$d^N(x) = \Big(\alpha^N(x)\alpha(\ell)^{-1}\Big)\mu^N(x) + \sqrt{2\ell}h(\ell)^{-1}(\Delta t)^{-\frac12}\varepsilon^N(x)$$

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

$$\varepsilon^N(x) \;=\; \mathbb{E}_x\big[\gamma^N(x,\xi^N)\, C^{\frac12}\xi^N\big] \;=\; \mathbb{E}_x\big[\big(1 \wedge e^{Q^N(x,\xi^N)}\big)\, C^{\frac12}\xi^N\big].$$

To prove Lemma 4.4 it suffices to verify that

$$\lim_{N\to\infty} \mathbb{E}^{\pi^N}\Big[\Big\|\Big(\alpha^N(x)\alpha(\ell)^{-1}\Big)\mu^N(x) - \mu(x)\Big\|^2_s\Big] = 0 \tag{4.10}$$

$$\lim_{N\to\infty} (\Delta t)^{-1}\, \mathbb{E}^{\pi^N}\Big[\|\varepsilon^N(x)\|^2_s\Big] = 0. \tag{4.11}$$

- Let us first prove Equation (4.10). The triangle inequality and Cauchy-Schwarz inequality show that

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

By Remark 3.7 $\mu^N : \mathcal{H}^s \to \mathcal{H}^s$ 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 4.2 and Corollary 4.3 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 3.8 proves that the sequence of functions $\{\mu^N\}$ converges π_0-almost surely to $\mu(x)$ in $\mathcal{H}^s$ and Lemma 3.10 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 (4.10).

- Let us prove Equation (4.11). If the Bernoulli random variable $\gamma^N(x, \xi^N)$ were independent from the noise term $(C^N)^{\frac{1}{2}}\xi^N$, it would follow that $\varepsilon^N(x) = 0$. In general $\gamma^N(x, \xi^N)$ is not independent from $(C^N)^{\frac{1}{2}}\xi^N$ so that $\varepsilon^N(x)$ is not equal to zero. Nevertheless, as quantified by Lemma 4.2, the Bernoulli random variable $\gamma^N(x, \xi^N)$ is asymptotically independent from the current position x and from the noise term $(C^N)^{\frac{1}{2}}\xi^N$. Consequently, we can prove in Equation (4.13) 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

$$\mathbb{E}^{\pi^N}\big[\langle \varepsilon^N(x), \hat{\varphi}_j \rangle_s^2\big] \quad \lesssim \quad N^{-1}\mathbb{E}^{\pi^N}[\langle x, \hat{\varphi}_j \rangle_s^2] + N^{-\frac{2}{3}}(j^s\lambda_j)^2. \tag{4.12}$$

Summation of Equation (4.12) over $j = 1, \ldots, N$ leads to

$$\mathbb{E}^{\pi^N}\Big[\|\varepsilon^N(x)\|_s^2\Big] \quad \lesssim \quad N^{-1}\mathbb{E}^{\pi^N}\big[\|x\|_s^2\big] + N^{-\frac{2}{3}}\,\mathrm{Tr}_{\mathcal{H}^s}(C_s) \quad \lesssim \quad N^{-\frac{2}{3}}, \tag{4.13}$$

which gives the proof of Equation (4.11). To prove Equation (4.12) 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

$$\begin{cases} Q^N(x,\xi^N) &= Q^N_j(x,\xi^N) + Q^N_{j,\perp}(x,\xi^N) \\ Q^N_j(x,\xi^N) &= -\frac{1}{\sqrt{2}}\ell^{\frac{3}{2}}N^{-\frac{1}{2}}\lambda_j^{-1}x_j\xi_j - \frac{1}{2}\ell^2 N^{-\frac{2}{3}}\lambda_j^2\xi_j^2 + \mathbf{e}^N(x,\xi^N). \end{cases} \tag{4.14}$$

The definitions of $Z^N(x, \xi^N)$ and $i^N(x, \xi^N)$ in Equation (4.2) and (4.3) readily show that $Q^N_{j,\perp}(x, \xi^N)$ is independent from ξ_j. The noise term satisfies $C^{\frac{1}{2}}\xi^N = \sum_{j=1}^N (j^s\lambda_j)\xi_j\hat{\varphi}_j$. Since $Q^N_{j,\perp}(x, \xi^N)$ and ξ_j are independent and $z \mapsto 1 \wedge e^z$ is 1-Lipschitz, it follows that

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

By Lemma 4.1 $\mathbb{E}^{\pi^N}\big[\mathbf{e}^N(x,\xi^N)^2\big] \lesssim N^{-\frac{2}{3}}$. Therefore,

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

which finishes the proof of Equation (4.12).

□

4.3. Noise approximation

Recall the definition (3.5) of the martingale difference $\Gamma^{k,N}$. In this section we estimate the error in the approximation $\Gamma^{k,N} \approx \mathrm{N}(0, C_s)$. 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 $D^N(x) \approx C_s$.

Lemma 4.5. *Let Assumptions 3.1 and 3.5 hold. For any pair of indices $i, j \geq 0$ the operator $D^N(x) : \mathcal{H}^s \to \mathcal{H}^s$ satisfies*

$$\lim_{N\to\infty} \quad \mathbb{E}^{\pi^N} \big|\langle \hat{\varphi}_i, D^N(x) \hat{\varphi}_j\rangle_s - \langle \hat{\varphi}_i, C_s \hat{\varphi}_j\rangle_s\big| \;=\; 0 \tag{4.15}$$

and, furthermore,

$$\lim_{N\to\infty} \quad \mathbb{E}^{\pi^N} \big|\mathrm{Tr}_{\mathcal{H}^s}(D^N(x)) - \mathrm{Tr}_{\mathcal{H}^s}(C_s)\big| \;=\; 0. \tag{4.16}$$

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

$$\Gamma^N(x,\xi) = \alpha(\ell)^{-\frac{1}{2}} \gamma^N(x,\xi)\, C^{\frac{1}{2}} \xi \;+\; \frac{1}{\sqrt{2}} \alpha(\ell)^{-\frac{1}{2}} (\ell \Delta t)^{\frac{1}{2}} \Big\{\gamma^N(x,\xi) \mu^N(x) - \alpha(\ell) d^N(x)\Big\}. \tag{4.17}$$

We only prove Equation (4.16); the proof of Equation (4.15) is essentially identical but easier. Remark 3.7 shows that the functions $\mu, \mu^N : \mathcal{H}^s \to \mathcal{H}^s$ are globally Lipschitz and Lemma 4.4 shows that $\mathbb{E}^{\pi^N}\big[\|d^N(x) - \mu(x)\|_s^2\big] \to 0$. Therefore

$$\mathbb{E}^{\pi^N} \Big[\|\gamma^N(x,\xi) \mu^N(x) - \alpha(\ell) d^N(x)\|_s^2\Big] \;\lesssim\; 1, \tag{4.18}$$

which implies that the second term on the right-hand-side of Equation (4.17) is $\mathcal{O}\big(\sqrt{\Delta t}\big)$. Since $\mathrm{Tr}_{\mathcal{H}^s}(D^N(x)) = \mathbb{E}_x\big[\|\Gamma^N(x,\xi)\|_s^2\big]$, Equation (4.18) implies that

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

Consequently, to prove Equation (4.16) it suffices to verify that

$$\lim_{N\to\infty} \; \mathbb{E}^{\pi^N}\Big[\big|\mathbb{E}_x\big[\|\gamma^N(x,\xi)\, C^{\frac{1}{2}}\xi\|_s^2\big] - \alpha(\ell) \mathrm{Tr}_{\mathcal{H}^s}(C_s)\big|\Big] \;=\; 0. \tag{4.19}$$

We have $\mathbb{E}_x\big[\|\gamma^N(x,\xi)\, C^{\frac{1}{2}}\xi\|_s^2\big] \;=\; \sum_{j=1}^N (j^s \lambda_j)^2 \mathbb{E}_x\big[\big(1 \wedge e^{Q^N(x,\xi)}\big)\xi_j^2\big]$. Therefore, to prove Equation (4.19) it suffices to establish

$$\lim_{N\to\infty} \sum_{j=1}^N (j^s \lambda_j)^2 \mathbb{E}^{\pi^N}\Big[\big|\mathbb{E}_x\big[\big(1 \wedge e^{Q^N(x,\xi)}\big)\xi_j^2\big] - \alpha(\ell)\big|\Big] = 0. \tag{4.20}$$

Since $\sum_{j=1}^\infty (j^s \lambda_j)^2 < \infty$ and $\big|1 \wedge e^{Q^N(x,\xi)}\big| \leq 1$, the dominated convergence theorem shows that (4.20) follows from

$$\lim_{N\to\infty} \; \mathbb{E}^{\pi^N}\Big[\big|\mathbb{E}_x\big[\big(1 \wedge e^{Q^N(x,\xi)}\big)\xi_j^2\big] - \alpha(\ell)\big|\Big] = 0 \qquad \forall j \geq 0. \tag{4.21}$$

We now prove Equation (4.21). As in the proof of Lemma 4.4, we use the decomposition $Q^N(x,\xi) = Q^N_j(x,\xi) + Q^N_{j,\perp}(x,\xi)$ where $Q^N_{j,\perp}(x,\xi)$ is independent from ξ_j. Therefore, since $\mathrm{Lip}(f) = 1$,

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

Lemma 4.2 ensures that, for $f(\cdot) = 1 \wedge \exp(\cdot)$,

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

and the definition of $Q^N_i(x,\xi)$ readily shows that $\lim_{N\to\infty} \mathbb{E}^{\pi^N}\big[Q^N_j(x,\xi)^2\big] = 0$. This concludes the proof of Equation (4.21) and thus ends the proof of Lemma 4.5. □

Corollary 4.6. *More generally, for any fixed vector $h \in \mathcal{H}^s$, the following limit holds,*

$$\lim_{N\to\infty} \quad \mathbb{E}^{\pi^N}\big|\langle h, D^N(x)h\rangle_s - \langle h, C_s h\rangle_s\big| \;=\; 0. \tag{4.22}$$

Proof. If $h = \hat{\varphi}_i$, this is precisely the content of Proposition 5.1. More generally, by linearity, Proposition 5.1 shows that this is true for $h = \sum_{i\le N} \alpha_i \hat{\varphi}_i$, 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 $h^* = \sum_{j\le N} \langle h, \hat{\varphi}_j\rangle_s \,\hat{\varphi}_j$ and $e^* = h - h^*$. It follows that

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

where we have used the fact that for an non-negative self-adjoint operator $D : \mathcal{H}^s \to \mathcal{H}^s$ we have $\langle u, Dv\rangle_s \le \|u\|_s \cdot \|v\|_s \cdot \mathrm{Tr}_{\mathcal{H}^s}(D)$. Proposition 5.1 shows that $\mathbb{E}^{\pi^N}[\mathrm{Tr}_{\mathcal{H}^s}(D^N(x))] < \infty$ and Assumption 3.1 ensures that $\mathrm{Tr}_{\mathcal{H}^s}(C_s) < \infty$. Consequently,

$$\begin{aligned}\lim_{N\to\infty} \mathbb{E}^{\pi^N}\Big|\langle h, D^N(x)\; h\rangle - \langle h, C_s h\rangle\Big| &\lesssim \lim_{N\to\infty} \mathbb{E}^{\pi^N}\Big|\langle h^*, D^N(x)\; h^*\rangle - \langle h^*, C_s\; h^*\rangle\Big| \;+\; \|h - h^*\|_s\\ &= \|h - h^*\|_s.\end{aligned}$$

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

4.4. Martingale Invariance Principle

This section proves that the process W^N defined in Equation (3.8) converges to a Brownian motion.

Proposition 4.7. *Let Assumptions 3.1 and 3.5 hold. Let $z^0 \sim \pi$ and $W^N(t)$ the process defined in equation (3.8) and $x^{0,N} \overset{\mathcal{D}}{\sim} \pi^N$ the starting position of the Markov chain x^N. Then*

$$(x^{0,N}, W^N) \Longrightarrow (z^0, W), \tag{4.23}$$

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 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 [MPS09], 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 $\epsilon > 0$,

$$\lim_{N\to\infty} \quad \Delta t \sum_{k=1}^{k_N(T)} \mathbb{E}\Big[\|\Gamma^{k,N}\|_s^2 \,|\mathcal{F}^{k,N}\Big] \;=\; T\,\mathrm{Tr}_{\mathcal{H}^s}(C_s) \tag{4.24}$$

$$\lim_{N\to\infty} \quad \Delta t \sum_{k=1}^{k_N(t)} \mathbb{E}\Big[\langle\Gamma^{k,N},\hat{\varphi}_i\rangle_s\langle\Gamma^{k,N},\hat{\varphi}_j\rangle_s \,|\mathcal{F}^{k,N}\Big] \;=\; t\,\langle\hat{\varphi}_i, C_s\hat{\varphi}_j\rangle_s \tag{4.25}$$

$$\lim_{N\to\infty} \quad \Delta t \sum_{k=1}^{k_N(T)} \mathbb{E}\Big[\|\Gamma^{k,N}\|_s^2 1_{\{\|\Gamma^{k,N}\|_s^2\geq\Delta t\,\epsilon\}} \,|\mathcal{F}^{k,N}\Big] \;=\; 0 \tag{4.26}$$

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 4.5 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** (4.24) It is enough to prove that $\lim\ \mathbb{E}\Big|\Big\{\frac{1}{\lfloor N^{\frac{1}{3}}\rfloor}\sum_{k=1}^{\lfloor N^{\frac{1}{3}}\rfloor}\mathbb{E}_k\big[\|\Gamma^{k,N}\|_s^2\big]\Big\} - \mathrm{Tr}_{\mathcal{H}^s}(C_s)\Big| = 0$ 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

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

  where the last limit follows from Lemma 4.5.
- **Condition** (4.25) It is enough to prove that

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

  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 4.5.
- **Condition** (4.26) 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 (4.17) shows that for any power $p \geq 0$ we have $\sup_N\ \mathbb{E}^{\pi^N}\big[\|\Gamma^{0,N}\|_s^p\big] < \infty$. 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 C_s. In order to prove that the sequence $(x^{0,N}, W^N)$ weakly converges to (z^0, W) in $\mathcal{H} \times C([0,T], \mathcal{H}^s)$, it suffices to prove that the sequence is tight (and thus weak relatively compact in $\mathcal{H}^s \times C([0,T], \mathcal{H}^s)$) and that (z^0, W) is the unique limit point of this relatively weak compact sequence.

- Since W^N converges weakly to W in the Polish space $C([0,T], \mathcal{H}^s)$, the sequence W^N is tight in $C([0,T], \mathcal{H}^s)$. The same argument shows that $\{x^{0,N}\}$ is tight in $\mathcal{H}^s$. This is enough to conclude that the sequence $(x^{0,N}, W^N)$ is tight in $\mathcal{H}^s \times C([0,T], \mathcal{H}^s)$, and thus weak relatively compact.
- Since the finite dimensional distributions of a random variable in $C([0,T], \mathcal{H}^s)$ completely characterizes its law [Dur96] and the Fourier transform completely characterizes the law of a probability distributions living in a Hilbert space [DPZ92], in order to show that (z^0, W) is the only limit point of the sequence of $\big\{(x^{0,N}, W^N)\big\}_N$ it suffices to show that for any choice of instants $0 = t_0 < t_1 < \ldots < t_m \le T$ and vectors $h_0, h_1, \ldots, h_m \in \mathcal{H}^s$ we have

$$\begin{aligned} L = \lim_{N\to\infty} \mathbb{E}\Big[e^{i\langle h_0, x^{0,N}\rangle_s} \cdot e^{i\sum_{j=1}^m \langle h_j, W^N(t_j) - W^N(t_{j-1})\rangle_s}\Big] \\ = \mathbb{E}\Big[e^{i\langle h_0, z^0\rangle_s}\Big] \cdot \mathbb{E}\Big[e^{i\sum_{j=1}^m \langle h_j, W(t_j) - W(t_{j-1})\rangle_s}\Big]. \end{aligned} \tag{4.27}$$

Moreover, it suffices to prove that for any $t_1 \in (0, \infty)$,

$$\begin{aligned} \lim_{N\to\infty} \mathbb{E}\Big[e^{i\langle h_0, x^{0,N}\rangle_s + i\langle h_1, W^N(t_1)\rangle_s}\Big] &= \mathbb{E}\Big[e^{i\langle h_0, z^0\rangle_s}\Big] \cdot \mathbb{E}\Big[e^{i\langle h_1, W(t_1)\rangle_s}\Big] \\ &= \mathbb{E}\Big[e^{i\langle h_0, z^0\rangle_s}\Big] \cdot \mathbb{E}\Big[e^{-\frac{1}{2}t_1 \, \langle h_1, C_s \, h_1\rangle_s}\Big]. \end{aligned} \tag{4.28}$$

To see this note that, assuming (4.28), the dominated convergence theorem shows that

$$\begin{aligned} L &= \lim_{N\to\infty} \mathbb{E}\Big[e^{i\langle h_0, x^{0,N}\rangle_s} \cdot e^{i\sum_{j=1}^{m-1} \langle h_j, W^N(t_j) - W^N(t_{j-1})\rangle_s} \cdot \mathbb{E}\big[e^{i\langle h_m, W^N(t_m) - W^N(t_{m-1})\rangle_s} \,|W^N(t_{m-1})\big]\Big] \\ &= \mathbb{E}\Big[e^{i\langle h_m, W(t_m) - W(t_{m-1})\rangle_s}\Big] \cdot \lim_{N\to\infty} \mathbb{E}\Big[e^{i\langle h_0, x^{0,N}\rangle_s} \cdot e^{i\sum_{j=1}^{m-1} \langle h_j, W^N(t_j) - W^N(t_{j-1})\rangle_s}\Big] \\ &= \mathbb{E}\Big[e^{i\langle h_0, x^{0,N}\rangle_s + i\langle h_1, W^N(t_1)\rangle_s}\Big] \cdot \prod_{j=2}^m \mathbb{E}\Big[e^{i\langle h_j, W(t_j) - W(t_{j-1})\rangle_s}\Big] \\ &= \mathbb{E}\Big[e^{i\langle h_0, z^0\rangle_s}\Big] \cdot \mathbb{E}\Big[e^{i\sum_{j=1}^m \langle h_j, W(t_j) - W(t_{j-1})\rangle_s}\Big], \end{aligned}$$

which is precisely Equation (4.27). Let us now prove Equation (4.28). For notational convenience we introduce the martingale increment $Y^{k,N} = \langle h_1, \Gamma^{k,N}\rangle_s$ and its asymptotic variance $\sigma^2 = \langle h_1, C_s h_1\rangle_s$. Proving Equation (4.28) is equivalent to verifying that

$$\lim_{N\to\infty} \mathbb{E}\Big[e^{i\langle h_0, x^{0,N}\rangle_s} \prod_{1 \le k \le \lfloor N^{\frac{1}{3}} \rfloor} e^{i\sqrt{\Delta t} Y^{k,N}}\Big] = \mathbb{E}\Big[e^{i\langle h_0, z^0\rangle_s}\Big] \cdot e^{-\frac{1}{2}\sigma^2}. \tag{4.29}$$

We have $|e^{i\sqrt{\Delta t} Y^{k,N}} - (1 + i\sqrt{\Delta t} Y^{k,N} - \frac{1}{2}\Delta t \, (Y^{k,N})^2| \lesssim (\Delta t)^{\frac{3}{2}} \, |Y^{k,N}|^3$. Moreover, Proposition 4.5 and Corollary 4.6 show that

$$\mathbb{E}\Big[(Y^{k,N})^2 \,|\mathcal{F}^{k,N}\Big] \;=\; \sigma^2 + \theta^{k,N}$$

where $\{\theta^{k,N}\}_{k\ge 0}$ is a stationary sequence with $\lim_{N\to\infty} \mathbb{E}^{\pi^N}\big[|\theta^{1,N}|\big] = 0$. Since $|Y^{k,N}|^3 \lesssim \|\Gamma^{k,N}\|_s^3$ it follows that $\sup\Big\{\mathbb{E}^{\pi^N}\big[|Y^{k,N}|^3\big] \;: k, N \ge 0\Big\} < \infty$. Also, $Y^{k,N}$ is a martingale increment so that $\mathbb{E}[Y^{k,N} \,|\mathcal{F}^{k,N}] = 0$. The triangle inequality thus shows that

$$\mathbb{E}\big[e^{i\sqrt{\Delta t} Y^{k,N}} \,|\mathcal{F}^{k,N}\big] \;=\; (1 - \frac{1}{2}\Delta t \, \sigma^2) \;+\; S^{k,N} \tag{4.30}$$

where $\{S^{k,N}\}_{k\geq 0}$ is a stationary sequence with

$$\lim_{N\to\infty} (\Delta t)^{-1}\, \mathbb{E}^{\pi^N}\big[|S^{1,N}|\big] = 0. \tag{4.31}$$

From (4.30) and the bound $|e^{i\langle h_0, x^{0,N}\rangle_s} \prod_{1\leq k\leq n} e^{i\sqrt{\Delta t}Y^{k,N}}| = 1$ it follows that

$$\begin{aligned}
\mathbb{E}\Big[e^{i\langle h_0, x^{0,N}\rangle_s} \prod_{1\leq k\leq \lfloor N^{\frac{1}{3}}\rfloor} e^{i\sqrt{\Delta t}Y^{k,N}}\Big] &= \mathbb{E}\Big[e^{i\langle h_0, x^{0,N}\rangle_s} \prod_{1\leq k\leq \lfloor N^{\frac{1}{3}}\rfloor -1} e^{i\sqrt{\Delta t}Y^{k,N}}\,(1-\frac{1}{2}\Delta t\,\sigma^2 + S^{\lfloor N^{\frac{1}{3}}\rfloor, N})\Big] \\
&= \mathbb{E}\Big[e^{i\langle h_0, x^{0,N}\rangle_s} \prod_{1\leq k\leq \lfloor N^{\frac{1}{3}}\rfloor -1} e^{i\sqrt{\Delta t}Y^{k,N}}\Big]\cdot(1-\frac{1}{2}\Delta t\,\sigma^2) \\
&\quad + \mathbb{E}^{\pi^N}\big[\hat{S}^{\lfloor N^{\frac{1}{3}}\rfloor, N}\big]
\end{aligned}$$

where $\{\hat{S}^{k,N}\}$ are error terms satisfying $|\hat{S}^{k,N}| = |S^{k,N}|$. Proceeding recursively we obtain

$$\mathbb{E}\Big[e^{i\langle h_0, x^{0,N}\rangle_s} \prod_{1\leq k\leq \lfloor N^{\frac{1}{3}}\rfloor} e^{i\sqrt{\Delta t}Y^{k,N}}\Big] = \mathbb{E}\Big[e^{i\langle h_0, x^{0,N}\rangle_s}\Big]\cdot(1-\frac{1}{2}\Delta t\,\sigma^2)^{\lfloor N^{\frac{1}{3}}\rfloor} + \sum_{1\leq k\leq \lfloor N^{\frac{1}{3}}\rfloor} \mathbb{E}^{\pi^N}\big[\hat{S}^{k,N}\big].$$

The Markov chain is started at stationarity so that $x^{0,N} \stackrel{\mathcal{D}}{\sim} \pi^N$. Lemma 3.10 states that π^N converges weakly in $\mathcal{H}^s$ to π: consequently, since $z^0 \stackrel{\mathcal{D}}{\sim} \pi$, we have $\lim_{N\to\infty} \mathbb{E}\big[e^{i\langle h_0, x^{0,N}\rangle_s}\big] = \mathbb{E}\big[e^{i\langle h_0, z^{0}\rangle_s}\big]$. Also, since $\Delta t = N^{-\frac{1}{3}}$, we have $\lim_{N\to\infty}(1-\frac{1}{2}\Delta t\,\sigma^2)^{\lfloor N^{\frac{1}{3}}\rfloor} = e^{-\frac{1}{2}\sigma^2}$. Finally, the stationarity of $S^{k,N}$ and the estimate (4.31) ensure that

$$\lim_{N\to\infty} \Big|\sum_{k=1}^{\lfloor N^{\frac{1}{3}}\rfloor} \mathbb{E}^{\pi^N}\big[\hat{S}^{k,N}\big]\Big| \leq \lim_{N\to\infty} \sum_{1\leq k\leq \lfloor N^{\frac{1}{3}}\rfloor} \mathbb{E}^{\pi^N}\Big[\big|S^{k,N}\big|\Big] \leq \lim_{N\to\infty} (\Delta t)^{-1}\mathbb{E}^{\pi^N}\big[|S^N|\big] \;=\; 0.$$

This finishes the proof of Equation (4.29) and thus ends the proof of Proposition 4.7.

□

5. Proof of Main Theorem

5.1. General diffusion approximation lemma

In this section we state and prove a proposition containing a general diffusion approximation result. Using this, we state precisely and then prove our Main Theorem from section 2: this is Theorem 5.2 below. 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

$$x^{k+1,N} - x^{k,N} = h(\ell)\, d^N(x_k)\,\Delta t \;+\; \sqrt{2h(\ell)\Delta t}\;\Gamma^{k,N} \tag{5.1}$$

where $h(\ell) > 0$ is a constant parameter and Δt is a time-step decreasing to 0 as N goes to infinity. We introduce the rescaled process $W^N(t)$ as in (3.8).

Proposition 5.1. (General Diffusion Approximation for Markov chains) *Consider a separable Hilbert space $\big(\mathcal{H}^s, \langle\cdot,\cdot\rangle_s\big)$ and a sequence of $\mathcal{H}^s$-valued Markov chains $x^N = \{x^{k,N}\}_{k\geq 0}$ with invariant distribution π^N and $x^{0,N} \stackrel{\mathcal{D}}{\sim} \pi^N$. Suppose that the drift-martingale decomposition (5.1) of x^N 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* (3.8) *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* C_s*.*
3. **Convergence of the drift:** *there exists a globaly Lipschitz function* $\mu : \mathcal{H}^s \to \mathcal{H}^s$ *that satisfies*

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

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*

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

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

Proof. Define $\bar{z}^N(t)$ as in (3.10). It then follows that

$$\begin{aligned} z^N(t) &= x^{0,N} + h(\ell)\int_0^t d^N(\bar{z}^N(u))\,du + \sqrt{2h(\ell)}\,W^N(t) \\ &= z^{0,N} + h(\ell)\int_0^t \mu(z^N(u))\,du + \sqrt{2h(\ell)}\,\widehat{W}^N(t) \end{aligned} \tag{5.2}$$

where

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

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 (5.2) 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.**
  Let $z_i = \Theta(z_i(0), W_i)$ for $i = 1, 2$. The definition of z_i and the fact that $\mu : \mathcal{H}^s \to \mathcal{H}^s$ is globally Lipschitz on $\mathcal{H}^s$ shows that

$$\begin{aligned} \|z_1(t) - z_2(t)\|_s \le \|z_1(0) - z_2(0)\|_s + h(\ell)\|\mu\|_{\mathrm{Lip}} \cdot \int_0^t \|z_1(u) - z_2(u)\|_s\,du \\ + \sqrt{2h(\ell)}\,\|W_1(t) - W_2(t)\|_s. \end{aligned}$$

  Therefore, we have

$$\begin{aligned} \sup_{0\le\tau\le t} \|z_1(\tau) - z_2(\tau)\|_s \le \|z_1(0) - z_2(0)\|_s \;+\; h(\ell)\|\mu\|_{\mathrm{Lip}} \cdot \int_0^t \sup_{0\le\tau\le u} \|z_1(\tau) - z_2(\tau)\|_s\,du \\ + \sqrt{2h(\ell)} \sup_{0\le\tau\le t} \|W_1(\tau) - W_2(\tau)\|_s. \end{aligned}$$

  Gronwall's lemma shows that Θ is well defined and is a continuous function.

- **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,

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

  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 $\mathrm{Tr}_{\mathcal{H}^s}(C_s) < \infty$, 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 $z = \Theta(z^0, W)$, and Proposition 5.1 is proved.

□

5.2. *Statement and Proof of The Main Theorem*

We now state and prove the main result of this paper.

Theorem 5.2. *Let the Assumptions 3.1 and 3.5 hold. Consider the MALA algorithm (2.19) with initial condition $x^{0,N} \overset{\mathcal{D}}{\sim} \pi^N$. 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 equation (1.7) with $z(0) \overset{\mathcal{D}}{\sim} \pi$.*

Proof. The proof of Theorem 5.2 consists in checking that the conditions needed for Proposition 5.1 to apply are satisfied by the sequence of MALA Markov chains (2.19).

1. By Lemma 3.10 the sequence of probability measures π^N converges weakly in $\mathcal{H}^s$ to π.
2. Proposition 4.7 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 C_s independent from $z^0 \overset{\mathcal{D}}{\sim} \pi$.
3. Lemma 4.4 states that $d^N(x)$ defined by Equation (3.4) satisfies $\lim_N \mathbb{E}^{\pi^N}\big[\|d^N(x) - \mu(x)\|_s^2\big] = 0$ and Proposition 3.6 shows that $\mu : \mathcal{H}^s \to \mathcal{H}^s$ is a Lipshitz function.

The three assumptions needed for Lemma 5.1 to apply are satisfied, which concludes the proof of Theorem 5.2.

□

6. Conclusions

We have studied the application of the MALA algorithm to approximate 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 developed an approach to the derivation of the diffusion limit which differs significantly from that used in [RR98].

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 [BPSSA] 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.

Acknowledgements AMS is grateful to EPSRC and ERC for financial support. AT is grateful to the (EPSRC-funded) CRISM for financial support. 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. Submitted, 2011.

[BPS04] Laird Arnault Breyer, Mauro Piccioni, and Sergio Scarlatti. Optimal scaling of MaLa for non-linear regression. *Ann. Appl. Probab.*, 14(3):1479–1505, 2004.

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

[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 for local Metropolis-Hastings chains on non-product targets in high dimensions. *Ann. Appl. Probab.*, 19(3):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.

[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.

[Dur96] Richard Durrett. *Stochastic Calculus: A Practical Introduction*. CRC Press; Subsequent edition (21 Jun 1996), 1996.

[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.

[HSV10] 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*, 2010. To Appear.

[MPS09] J.C. Mattingly, N.S. Pillai, and A.M. Stuart. SPDE Limits of the Random Walk Metropolis Algorithm in High Dimensions. 2009.

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

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