



Non-stationary phase of the MALA algorithm

Juan Kuntz¹ · Michela Ottobre²  ·
Andrew M. Stuart³

Received: 3 August 2017 / Published online: 17 April 2018
© The Author(s) 2018

Abstract The Metropolis-Adjusted Langevin Algorithm (MALA) is a Markov Chain Monte Carlo method which creates a Markov chain reversible with respect to a given target distribution, π^N , with Lebesgue density on \mathbb{R}^N ; it can hence be used to approximately sample the target distribution. When the dimension N is large a key question is to determine the computational cost of the algorithm as a function of N . The measure of efficiency that we consider in this paper is the *expected squared jumping distance* (ESJD), introduced in Roberts et al. (Ann Appl Probab 7(1):110–120, 1997). To determine how the cost of the algorithm (in terms of ESJD) increases with dimension N , we adopt the widely used approach of deriving a diffusion limit for the Markov chain produced by the MALA algorithm. We study this problem for a class of target measures which is *not* in product form and we address the situation of practical relevance in which the algorithm is started out of stationarity. We thereby significantly extend previous works which consider either measures of product form, when the Markov chain is started out of stationarity, or non-product measures (defined via a density with respect to a Gaussian), when the Markov chain is started in stationarity. In order to work in this non-stationary and non-product setting, significant new analysis is

✉ Michela Ottobre
michelaottobre@gmail.com

Juan Kuntz
juankuntz@protonmail.com

Andrew M. Stuart
astuart@caltech.edu

¹ Imperial College London, London SW7 2AZ, UK

² Mathematics Department, Heriot Watt University, Edinburgh EH14 4AS, UK

³ Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA 91125, USA

required. In particular, our diffusion limit comprises a stochastic PDE coupled to a scalar ordinary differential equation which gives a measure of how far from stationarity the process is. The family of non-product target measures that we consider in this paper are found from discretization of a measure on an infinite dimensional Hilbert space; the discretised measure is defined by its density with respect to a Gaussian random field. The results of this paper demonstrate that, in the non-stationary regime, the cost of the algorithm is of $\mathcal{O}(N^{1/2})$ in contrast to the stationary regime, where it is of $\mathcal{O}(N^{1/3})$.

Keywords Markov Chain Monte Carlo · Metropolis-Adjusted Langevin Algorithm · Diffusion limit · Optimal scaling

Mathematics Subject Classification Primary 60J22; Secondary 60J20 · 60H10

1 Introduction

1.1 Context

Metropolis–Hastings algorithms are Markov Chain Monte Carlo (MCMC) methods used to sample from a given probability measure, referred to as the target measure. The basic mechanism consists of employing a proposal transition density $q(x, y)$ in order to produce a reversible Markov chain $\{x^k\}_{k=0}^\infty$ for which the target measure π is invariant [11]. At step k of the chain, a proposal move y^k is generated by using $q(x, y)$, i.e. $y^k \sim q(x^k, \cdot)$. Then such a move is accepted with probability $\alpha(x^k, y^k)$:

$$\alpha(x^k, y^k) = \min \left\{ 1, \frac{\pi(y^k)q(y^k, x^k)}{\pi(x^k)q(x^k, y^k)} \right\}. \quad (1.1)$$

The computational cost of this algorithm when the state space has high dimension N is of practical interest in many applications. The measure of computational cost considered in this paper is the *expected squared jumping distance*, introduced in [19] and related works. Roughly speaking [we will be more precise about this in the next Sect. 1.2, see comments before (1.8)], if the size of the proposal moves is too large, i.e. if we propose moves which are too far away from the current position, then such moves tend to be frequently rejected; on the other hand, if the algorithm proposes moves which are too close to the current position, then such moves will be most likely accepted, however the chain will have not moved very far away. In either extreme cases, the chain tends to get stuck and will exhibit slow mixing, and this is more and more true as the dimension N of the state space increases. It is therefore clear that one needs to strike a balance between these two opposite scenarios; in particular, the optimal size of the proposed moves (i.e., the proposal variance) will depend on N . If the proposal variance scales with N like $N^{-\zeta}$, for some $\zeta > 0$, then we will say that the cost of the algorithm, in terms of ESJD, is of the order N^ζ .

A widely used approach to tackle this problem is to study diffusion limits for the algorithm. Indeed the scaling used to obtain a well defined diffusion limit corresponds

to the optimal scaling of the proposal variance (see Remark 1.1). This problem was first studied in [19], for the Random Walk Metropolis algorithm (RWM); in this work it is assumed that the algorithm is started in stationarity and that the target measure is in product form. In the case of the MALA algorithm, the same problem was considered in [20,21], again in the stationary regime and for product measures. In this setting, the cost of RWM has been shown to be $\mathcal{O}(N)$, while the cost of MALA is $\mathcal{O}(N^{\frac{1}{3}})$. The same $\mathcal{O}(N^{\frac{1}{3}})$ scaling for MALA, in the stationary regime, was later obtained in the setting of non-product measures defined via density with respect to a Gaussian random field [17]. In the paper [6] extensions of these results to non-stationary initializations were considered, however only for the Gaussian targets. For Gaussian targets, RWM was shown to scale the same in and out of stationarity, whilst MALA scales like $\mathcal{O}(N^{\frac{1}{2}})$ out of stationarity. In [12,13] the RWM and MALA algorithms were studied out of stationarity for quite general product measures and the RWM method shown again to scale the same in and out of stationarity. For MALA the appropriate scaling was shown to differ in and out of stationarity and, crucially, the scaling out of stationarity was shown to depend on a certain moment of the potential defining the product measure. In this paper we contribute further understanding of the MALA algorithm when initialized out of stationarity by considering non-product measures defined via density with respect to a Gaussian random field. Considering such a class of measures has proved fruitful, see e.g. [15,17]. Relevant to this strand of literature, is also the work [5].

In this paper our primary contribution is the study of diffusion limits for the MALA algorithm, out of stationarity, in the setting of general non-product measures, defined via density with respect to a Gaussian random field. Significant new analysis is needed for this problem because the work of [17] relies heavily on stationarity in analyzing the acceptance probability, whilst the work of [13] uses propagation of chaos techniques, unsuitable for non-product settings.

The challenging diffusion limit obtained in this paper is relevant both to the picture just described and, in general, due to the widespread practical use of the MALA algorithm. The understanding we obtain about the MALA algorithm when applied to realistic non-product targets is one of the main motivations for the analysis that we undertake in this paper. The diffusion limit we find is given by an SPDE coupled to a one-dimensional ODE. The evolution of such an ODE can be taken as an indicator of how close the chain is to stationarity (see Remark 1.1 for more details on this). The scaling adopted to obtain such a diffusion limit shows that the cost of the algorithm is of order $N^{1/2}$ in the non-stationary regime, as opposed to what happens in the stationary phase, where the cost is of order $N^{1/3}$. It is important to recognize that, for measures absolutely continuous with respect to a Gaussian random field, algorithms exist which require $\mathcal{O}(1)$ steps in and out of stationarity; see [7] for a review. Such methods were suggested by Radford Neal in [16], and developed by Alex Beskos for conditioned stochastic differential equations in [4], building on the general formulation of Metropolis–Hastings methods in [23]; these methods are analyzed from the point of view of diffusion limits in [18]. It thus remains open and interesting to study the MALA algorithm out of stationarity for non-product measures which are not defined via density with respect to a Gaussian random field; however the results in

[12] demonstrate the substantial technical barriers that will exist in trying to do so. An interesting starting point of such work might be the study of non i.i.d. product measures as pioneered by Bédard [2,3].

1.2 Setting and the main result

Let $(\mathcal{H}, \langle \cdot, \cdot \rangle, \| \cdot \|)$ be an infinite dimensional separable Hilbert space and consider the measure π on \mathcal{H} , defined as follows:

$$\frac{d\pi}{d\pi_0} \propto \exp(-\Psi), \quad \pi_0 := \mathcal{N}(0, \mathcal{C}). \tag{1.2}$$

That is, π is absolutely continuous with respect to a Gaussian measure π_0 with mean zero and covariance operator \mathcal{C} . Ψ is some real valued functional with domain $\tilde{\mathcal{H}} \subseteq \mathcal{H}$, $\Psi : \tilde{\mathcal{H}} \rightarrow \mathbb{R}$. Measures of the form (1.2) naturally arise in Bayesian nonparametric statistics and in the study of conditioned diffusions [10,22]. In Sect. 2 we will give the precise definition of the space $\tilde{\mathcal{H}}$ and identify it with an appropriate Sobolev-like subspace of \mathcal{H} (denoted by \mathcal{H}^s in Sect. 2). The covariance operator \mathcal{C} is a positive, self-adjoint, trace class operator on \mathcal{H} , with eigenbasis $\{\lambda_j^2, \phi_j\}$:

$$\mathcal{C}\phi_j = \lambda_j^2\phi_j, \quad \forall j \in \mathbb{N}, \tag{1.3}$$

and we assume that the set $\{\phi_j\}_{j \in \mathbb{N}}$ is an orthonormal basis for \mathcal{H} .

We will analyse the MALA algorithm designed to sample from the finite dimensional projections π^N of the measure (1.2) on the space

$$X^N := \text{span}\{\phi_j\}_{j=1}^N \subset \mathcal{H} \tag{1.4}$$

spanned by the first N eigenvectors of the covariance operator. Notice that the space X^N is isomorphic to \mathbb{R}^N . To clarify this further, we need to introduce some notation. Given a point $x \in \mathcal{H}$, $\mathcal{P}^N(x) := \sum_{j=1}^n \langle \phi_j, x \rangle \phi_j$ is the projection of x onto the space X^N and we define the approximations of functional Ψ and covariance operator \mathcal{C} :

$$\Psi^N := \Psi \circ \mathcal{P}^N \quad \text{and} \quad \mathcal{C}_N := \mathcal{P}^N \circ \mathcal{C} \circ \mathcal{P}^N. \tag{1.5}$$

With this notation in place, our target measure is the measure π^N (on $X^N \cong \mathbb{R}^N$) defined as

$$\frac{d\pi^N}{d\pi_0^N}(x) = M_{\Psi^N} e^{-\Psi^N(x)}, \quad \pi_0^N := \mathcal{N}(0, \mathcal{C}_N), \tag{1.6}$$

where M_{Ψ^N} is a normalization constant. Notice that the sequence of measures $\{\pi^N\}_{N \in \mathbb{N}}$ approximates the measure π (in particular, the sequence $\{\pi^N\}_{N \in \mathbb{N}}$ converges to π in the Hellinger metric, see [22, Section 4] and references therein). In

order to sample from the measure π^N in (1.6), we will consider the MALA algorithm with proposal

$$y^{k,N} = x^{k,N} + \delta C_N \nabla \log \pi^N(x^{k,N}) + \sqrt{2\delta} C_N^{1/2} \xi^{k,N}, \tag{1.7}$$

where

$$\xi^{k,N} = \sum_{i=1}^N \xi_i \phi_i, \quad \xi_i \stackrel{\mathcal{D}}{\sim} \mathcal{N}(0, 1) \text{ i.i.d.},$$

and $\delta > 0$ is a positive parameter. We rewrite $y^{k,N}$ as

$$y^{k,N} = x^{k,N} - \delta(x^{k,N} + C_N \nabla \Psi^N(x^{k,N})) + \sqrt{2\delta} C_N^{1/2} \xi^{k,N}.$$

The proposal defines the kernel q and enters in the accept-reject criterion α , which is added to preserve detailed balance with respect to π^N (more details on the algorithm will be given in Sect. 2.2). The proposal is a discretization of a π^N -invariant diffusion process with time step δ ; in the MCMC literature δ is often referred to as the proposal variance. The accept-reject criterion compensates for the discretization, which destroys the π^N -reversibility. A crucial parameter to be appropriately chosen in order to optimize the performance of the algorithm is δ ; such a choice will depend on the dimension N of the state space. To be more precise, set $\delta = \ell N^{-\zeta}$, where ℓ, ζ are two positive parameters, the latter being, for the time, the most relevant to this discussion. As explained when outlining the context of this paper, if ζ is too large (so that δ is too small) then the algorithm will tend to move very slowly; if ζ is too big, then the proposed moves will be very large and the algorithm will tend to reject them very often. In this paper we show that, if the algorithm is started out of stationarity then, in the non-stationary regime, the optimal choice of ζ is $\zeta = 1/2$. In particular, if

$$\delta = \ell / \sqrt{N} \tag{1.8}$$

then the acceptance probability is $\mathcal{O}(1)$. Furthermore, starting from the Metropolis–Hastings chain $\{x^{k,N}\}_{k \in \mathbb{N}}$, we define the continuous interpolant

$$\begin{aligned} x^{(N)}(t) &= (N^{1/2}t - k)x^{k+1,N} + (k + 1 - N^{1/2}t)x^{k,N}, \\ t_k \leq t < t_{k+1}, \text{ where } t_k &= \frac{k}{N^{1/2}}. \end{aligned} \tag{1.9}$$

This process converges weakly to a diffusion process. The precise statement of such a result is given in Theorem 4.2 (and Sect. 4 contains heuristic arguments which explain how such a result is obtained). In proving the result we will use the fact that $W(t)$ is a \mathcal{H}_s -valued Brownian motion with covariance C_s , with \mathcal{H}_s a (Hilbert) subspace of \mathcal{H} and C_s the covariance in this space. Details of these spaces are given in Sect. 2, see in particular (2.4) and (2.5). Below $C([0, T]; \mathcal{H}_s)$ denotes the space of \mathcal{H}_s -valued continuous functions on $[0, T]$, endowed with the uniform topology; α_ℓ, h_ℓ and b_ℓ

are real valued functions, which we will define immediately after the statement, and $x_j^{k,N}$ denotes the j th component of the vector $x^{k,N} \in X^N$ with respect to the basis $\{\phi_1, \dots, \phi_N\}$ (more details on this notation are given in Sect. 2.1.)

Main Result *Let $\{x^{k,N}\}_{k \in \mathbb{N}}$ be the Metropolis–Hastings Markov chain to sample from π^N and constructed using the MALA proposal (1.7) (i.e. the chain (2.14)) with δ chosen to satisfy (1.8). Then, for any deterministic initial datum $x^{0,N} = \mathcal{P}^N(x^0)$, where x^0 is any point in \mathcal{H}_s , the continuous interpolant $x^{(N)}$ defined in (1.9) converges weakly in $C([0, T]; \mathcal{H}_s)$ to the solution of the SDE*

$$dx(t) = -h_\ell(S(t))(x(t) + C\nabla\Psi(x(t))) dt + \sqrt{2h_\ell(S(t))} dW(t), \quad x(0) = x^0, \tag{1.10}$$

where $S(t) \in \mathbb{R}_+ := \{s \in \mathbb{R} : s \geq 0\}$ solves the ODE

$$dS(t) = b_\ell(S(t)) dt, \quad S(0) := \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{j=1}^N \frac{|x_j^{0,N}|^2}{\lambda_j^2}. \tag{1.11}$$

In the above the initial datum $S(0)$ is assumed to be finite and $W(t)$ is a \mathcal{H}_s -valued Brownian motion with covariance \mathcal{C}_s .

The functions $\alpha_\ell, h_\ell, b_\ell : \mathbb{R} \rightarrow \mathbb{R}$ in the previous statement are defined as follows:

$$\alpha_\ell(s) = 1 \wedge e^{\ell^2(s-1)/2} \tag{1.12}$$

$$h_\ell(s) = \ell\alpha_\ell(s) \tag{1.13}$$

$$b_\ell(s) = 2\ell(1 - s) \left(1 \wedge e^{\ell^2(s-1)/2}\right) = 2(1 - s)h_\ell(s). \tag{1.14}$$

Remark 1.1 We make several remarks concerning the main result.

- Since the effective time-step implied by the interpolation (1.9) is $N^{-1/2}$, the main result implies that the number of steps required by the Markov chain in its non-stationary regime is $\mathcal{O}(N^{1/2})$. A more detailed discussion on this fact can be found in Sect. 4.
- Notice that Eq. (1.11) evolves independently of Eq. (1.10). Once the MALA algorithm (2.14) is introduced and an initial state $x^0 \in \mathcal{H}$ is given such that $S(0)$ is finite, the real valued (double) sequence $S^{k,N}$,

$$S^{k,N} := \frac{1}{N} \sum_{i=1}^N \frac{|x_i^{k,N}|^2}{\lambda_i^2} \tag{1.15}$$

started at $S_0^N := \frac{1}{N} \sum_{i=1}^N \frac{|x_i^{0,N}|^2}{\lambda_i^2}$ is well defined. For fixed N , $\{S^{k,N}\}_k$ is not, in general, a Markov process (however it is Markov if e.g. $\Psi = 0$). Consider the continuous interpolant $S^{(N)}(t)$ of the sequence $S^{k,N}$, namely

$$S^{(N)}(t) = (N^{1/2}t - k)S^{k+1,N} + (k + 1 - N^{1/2}t)S^{k,N}, \quad t_k \leq t < t_{k+1}, \quad t_k = \frac{k}{N^{1/2}}. \tag{1.16}$$

In Theorem 4.1 we prove that $S^{(N)}(t)$ converges in probability in $C([0, T]; \mathbb{R})$ to the solution of the ODE (1.11) with initial condition $S_0 := \lim_{N \rightarrow \infty} S_0^N$. Once such a result is obtained, we can prove that $x^{(N)}(t)$ converges to $x(t)$. We want to stress that the convergence of $S^{(N)}(t)$ to $S(t)$ can be obtained independently of the convergence of $x^{(N)}(t)$ to $x(t)$.

- Let $S(t) : \mathbb{R} \rightarrow \mathbb{R}$ be the solution of the ODE (1.11). We will prove (see Theorem 3.1) that $S(t) \rightarrow 1$ as $t \rightarrow \infty$; this is also consistent with the fact that, in stationarity, $S^{k,N}$ converges to 1 as $N \rightarrow \infty$ (for every $k > 0$), see Remark 4.1. In view of this and the above comment, $S(t)$ (or $S^{k,N}$) can be taken as an indication of how close the chain is to stationarity. Moreover, notice that $h_\ell(1) = \ell$; heuristically one can then argue that the asymptotic behaviour of the law of $x(t)$, the solution of (1.10), is described by the law of the following infinite dimensional SDE:

$$dz(t) = -\ell(z(t) + \mathcal{C}\nabla\Psi(z(t)))dt + \sqrt{2\ell}dW(t). \tag{1.17}$$

It was proved in [9, 10] that (1.17) is ergodic with unique invariant measure given by (1.2). Our deduction concerning computational cost is made on the assumption that the law of (1.10) does indeed tend to the law of (1.17), although we will not prove this here as it would take us away from the main goal of the paper which is to establish the diffusion limit of the MALA algorithm.

- In [12, 13] the diffusion limit for the MALA algorithm started out of stationarity and applied to i.i.d. target product measures is given by a non-linear equation of McKean-Vlasov type. This is in contrast with our diffusion limit, which is an infinite-dimensional SDE. The reason why this is the case is discussed in detail in [14, Section 1.2]. The discussion in the latter paper is in the context of the Random Walk Metropolis algorithm, but it is conceptually analogous to what holds for the MALA algorithm and for this reason we do not spell it out here.
- In this paper we make stronger assumptions on Ψ than are required to prove a diffusion limit in the stationary regime [17]. In particular we assume that the first derivative of Ψ is bounded, whereas [17] requires only boundedness of the second derivative. Removing this assumption on the first derivative, or showing that it is necessary, would be of interest but would require different techniques to those employed in this paper and we do not address the issue here.

Remark 1.2 The proposal we employ in this paper is the standard MALA proposal. It can be seen as a particular case of the more general proposal introduced in [4, equation (4.2)] see also [1]; in our notation this proposal can be written as

$$y^{k+1,N} = x^{k,N} + \delta \{ -(1 - \theta)x^{k,N} - \theta y^{k+1,N} - \mathcal{C}_N \nabla \Psi^N(x^{k,N}) \} + \sqrt{2\delta} \xi^{k,N}. \tag{1.18}$$

In the above, $\theta \in [0, 1]$ is a parameter. The choice $\theta = 0$ corresponds to our proposal. When $\theta = 1/2$, the resulting algorithm is well posed in infinite dimensions; as a

consequence a diffusion limit is obtained, in and out of stationarity, without scaling δ with respect to N ; see Remark 4.3. When $\theta \neq 1/2$ the algorithms all suffer from the curse of dimensionality: it is necessary to scale δ inversely with a power of N to obtain an acceptable acceptance probability. In this paper we study how the efficiency decreases with N when $\theta = 0$; results analogous to the one we prove here will hold for any $\theta \neq 1/2$, but proving them at this level of generality would lengthen the article without adding insight. Furthermore, for non-Gaussian priors practitioners might use the algorithm with $\theta = 0$ and so our results shed light on that case; if the prior is actually Gaussian practitioners should use the algorithm with $\theta = \frac{1}{2}$. There is no reasons to use any other values of θ in practice, as far as we are aware.

1.3 Structure of the paper

The paper is organized as follows. In Sect. 2 we introduce the notation and the assumptions that we use throughout this paper. In particular, Sect. 2.1 introduces the infinite dimensional setting in which we work, Sect. 2.2 discusses the MALA algorithm and the assumptions we make on the functional Ψ and on the covariance operator \mathcal{C} . Section 3 contains the proof of existence and uniqueness of solutions for the limiting Eqs. (1.10) and (1.11). With these preliminaries in place, we give, in Sect. 4, the formal statement of the main results of this paper, Theorems 4.1 and 4.2. In this section we also provide heuristic arguments outlining how the main results are obtained. The complete proof of these results builds on a continuous mapping argument presented in Sect. 5. The heuristics of Sect. 4 are made rigorous in Sects. 6–8. In particular, Sect. 6 contains some estimates of the size of the chain's jumps and the growth of its moments, as well as the study of the acceptance probability. In Sects. 7 and 8 we use these estimates and approximations to prove Theorems 4.1 and 4.2, respectively. Readers interested in the structure of the proofs of Theorems 4.1 and 4.2 but not in the technical details may wish to skip the ensuing two sections (Sects. 2 and 3) and proceed directly to the statement of these results and the relevant heuristics discussed in Sect. 4.

2 Notation, algorithm, and assumptions

In this section we detail the notation and the assumptions (Sects. 2.1 and 2.3, respectively) that we will use in the rest of the paper.

2.1 Notation

Let $(\mathcal{H}, \langle \cdot, \cdot \rangle, \|\cdot\|)$ denote a real separable infinite dimensional Hilbert space, with the canonical norm induced by the inner-product. Let π_0 be a zero-mean Gaussian measure on \mathcal{H} with covariance operator \mathcal{C} . By the general theory of Gaussian measures [8], \mathcal{C} is a positive, trace class operator. Let $\{\phi_j, \lambda_j^2\}_{j \geq 1}$ be the eigenfunctions and eigenvalues of \mathcal{C} , respectively, so that (1.3) holds. We assume a normalization under which $\{\phi_j\}_{j \geq 1}$

forms a complete orthonormal basis of \mathcal{H} . Recalling (1.4), we specify the notation that will be used throughout this paper:

- x and y are elements of the Hilbert space \mathcal{H} ;
- the letter N is reserved to denote the dimensionality of the space X^N where the target measure π^N is supported;
- x^N is an element of $X^N \cong \mathbb{R}^N$ (similarly for y^N and the noise ξ^N);
- for any fixed $N \in \mathbb{N}$, $x^{k,N}$ is the k th step of the chain $\{x^{k,N}\}_{k \in \mathbb{N}} \subseteq X^N$ constructed to sample from π^N ; $x_i^{k,N}$ is the i th component of the vector $x^{k,N}$, that is $x_i^{k,N} := \langle x^{k,N}, \phi_i \rangle$ (with abuse of notation).

For every $x \in \mathcal{H}$, we have the representation $x = \sum_{j \geq 1} x_j \phi_j$, where $x_j := \langle x, \phi_j \rangle$. Using this expansion, we define Sobolev-like spaces $\mathcal{H}^s, s \in \mathbb{R}$, with the inner-products and norms defined by

$$\langle x, y \rangle_s = \sum_{j=1}^{\infty} j^{2s} x_j y_j \quad \text{and} \quad \|x\|_s^2 = \sum_{j=1}^{\infty} j^{2s} x_j^2.$$

The space $(\mathcal{H}^s, \langle \cdot, \cdot \rangle_s)$ is also a Hilbert space. Notice that $\mathcal{H}^0 = \mathcal{H}$. Furthermore $\mathcal{H}^s \subset \mathcal{H} \subset \mathcal{H}^{-s}$ for any $s > 0$. The Hilbert–Schmidt norm $\|\cdot\|_{\mathcal{C}}$ associated with the covariance operator \mathcal{C} is defined as

$$\|x\|_{\mathcal{C}}^2 := \sum_{j=1}^{\infty} \lambda_j^{-2} x_j^2 = \sum_{j=1}^{\infty} \frac{|\langle x, \phi_j \rangle|^2}{\lambda_j^2}, \quad x \in \mathcal{H},$$

and it is the Cameron–Martin norm associated with the Gaussian measure $\mathcal{N}(0, \mathcal{C})$. Such a norm is induced by the scalar product

$$\langle x, y \rangle_{\mathcal{C}} := \langle \mathcal{C}^{-1/2} x, \mathcal{C}^{-1/2} y \rangle, \quad x, y \in \mathcal{H}.$$

Similarly, \mathcal{C}_N defines a Hilbert–Schmidt norm on X^N ,

$$\|x^N\|_{\mathcal{C}_N}^2 := \sum_{j=1}^N \frac{|\langle x^N, \phi_j \rangle|^2}{\lambda_j^2}, \quad x^N \in X^N, \tag{2.1}$$

which is induced by the scalar product

$$\langle x^N, y^N \rangle_{\mathcal{C}_N} := \langle \mathcal{C}_N^{-1/2} x^N, \mathcal{C}_N^{-1/2} y^N \rangle, \quad x^N, y^N \in X^N.$$

For $s \in \mathbb{R}$, let $L_s : \mathcal{H} \rightarrow \mathcal{H}$ denote the operator which is diagonal in the basis $\{\phi_j\}_{j \geq 1}$ with diagonal entries j^{2s} ,

$$L_s \phi_j = j^{2s} \phi_j,$$

so that $L_s^{\frac{1}{2}} \phi_j = j^s \phi_j$. The operator L_s lets us alternate between the Hilbert space \mathcal{H} and the interpolation spaces \mathcal{H}^s via the identities:

$$\langle x, y \rangle_s = \left\langle L_s^{\frac{1}{2}} x, L_s^{\frac{1}{2}} y \right\rangle \quad \text{and} \quad \|x\|_s^2 = \left\| L_s^{\frac{1}{2}} x \right\|^2.$$

Since $\left\| L_s^{-1/2} \phi_k \right\|_s = \|\phi_k\| = 1$, we deduce that $\{\hat{\phi}_k := L_s^{-1/2} \phi_k\}_{k \geq 1}$ forms an orthonormal basis of \mathcal{H}^s . An element $y \sim \mathcal{N}(0, \mathcal{C})$ can be expressed as

$$y = \sum_{j=1}^{\infty} \lambda_j \rho_j \phi_j \quad \text{with} \quad \rho_j \stackrel{\mathcal{D}}{\sim} \mathcal{N}(0, 1) \text{ i.i.d.} \tag{2.2}$$

If $\sum_j \lambda_j^2 j^{2s} < \infty$, then y can be equivalently written as

$$y = \sum_{j=1}^{\infty} (\lambda_j j^s) \rho_j (L_s^{-1/2} \phi_j) \quad \text{with} \quad \rho_j \stackrel{\mathcal{D}}{\sim} \mathcal{N}(0, 1) \text{ i.i.d.} \tag{2.3}$$

For a positive, self-adjoint operator $D : \mathcal{H} \mapsto \mathcal{H}$, its trace in \mathcal{H} is defined as

$$\text{Trace}_{\mathcal{H}}(D) := \sum_{j=1}^{\infty} \langle \phi_j, D \phi_j \rangle.$$

We stress that in the above $\{\phi_j\}_{j \in \mathbb{N}}$ is an orthonormal basis for $(\mathcal{H}, \langle \cdot, \cdot \rangle)$. Therefore, if $\tilde{D} : \mathcal{H}^s \rightarrow \mathcal{H}^s$, its trace in \mathcal{H}^s is

$$\text{Trace}_{\mathcal{H}^s}(\tilde{D}) = \sum_{j=1}^{\infty} \left\langle L_s^{-\frac{1}{2}} \phi_j, \tilde{D} L_s^{-\frac{1}{2}} \phi_j \right\rangle_s.$$

Since $\text{Trace}_{\mathcal{H}^s}(\tilde{D})$ does not depend on the orthonormal basis, the operator \tilde{D} is said to be trace class in \mathcal{H}^s if $\text{Trace}_{\mathcal{H}^s}(\tilde{D}) < \infty$ for some, and hence any, orthonormal basis of \mathcal{H}^s . Because \mathcal{C} is defined on \mathcal{H} , the covariance operator¹

$$\mathcal{C}_s = L_s^{1/2} \mathcal{C} L_s^{1/2} \tag{2.4}$$

is defined on \mathcal{H}^s . Thus, for all the values of r such that $\text{Trace}_{\mathcal{H}^s}(\mathcal{C}_s) = \sum_j \lambda_j^2 j^{2s} < \infty$, we can think of y as a mean zero Gaussian random variable with covariance operator

¹ In this paper, we commit a slight abuse of our notation by writing \mathcal{C}_s to mean the covariance operator on the Sobolev-like subspace \mathcal{H}^s and \mathcal{C}_N to mean that on the finite dimensional subspace X^N as defined in (1.5). We distinguish these two by always employing N as the subscript for the latter, and lower case letters such as s or r for the former.

\mathcal{C} in \mathcal{H} and \mathcal{C}_s in \mathcal{H}^s [see (2.2) and (2.3)]. In the same way, if $\text{Trace}_{\mathcal{H}^s}(\mathcal{C}_s) < \infty$, then

$$W(t) = \sum_{j=1}^{\infty} \lambda_j w_j(t) \phi_j = \sum_{j=1}^{\infty} \lambda_j j^r w_j(t) \hat{\phi}_j, \tag{2.5}$$

where $\{w_j(t)\}_{j \geq 1}$ a collection of i.i.d. standard Brownian motions on \mathbb{R} , can be equivalently understood as an \mathcal{H} -valued \mathcal{C} -Brownian motion or as an \mathcal{H}^s -valued \mathcal{C}_s -Brownian motion.

We will make use of the following elementary inequality,

$$|\langle x, y \rangle|^2 = \left| \sum_{j=1}^{\infty} (j^s x_j)(j^{-s} y_j) \right|^2 \leq \|x\|_s^2 \|y\|_{-s}^2, \quad \forall x \in \mathcal{H}^s, \quad y \in \mathcal{H}^{-s}. \tag{2.6}$$

Throughout this paper we study sequences of real numbers, random variables and functions, indexed by either (or both) the dimension N of the space on which the target measure is defined or the chain’s step number k . In doing so, we find the following notation convenient.

- Two (double) sequences of real numbers $\{A^{k,N}\}$ and $\{B^{k,N}\}$ satisfy $A^{k,N} \lesssim B^{k,N}$ if there exists a constant $K > 0$ (independent of N and k) such that

$$A^{k,N} \leq K B^{k,N},$$

for all N and k such that $\{A^{k,N}\}$ and $\{B^{k,N}\}$ are defined.

- If the $A^{k,N}$ s and $B^{k,N}$ s are random variables, the above inequality must hold almost surely (for some deterministic constant K).
- If the $A^{k,N}$ s and $B^{k,N}$ s are real-valued functions on \mathcal{H} or \mathcal{H}^s , $A^{k,N} = A^{k,N}(x)$ and $B^{k,N} = B^{k,N}(x)$, the same inequality must hold with K independent of x , for all x where the $A^{k,N}$ s and $B^{k,N}$ s are defined.

As is customary, $\mathbb{R}_+ := \{s \in \mathbb{R} : s \geq 0\}$ and for all $b \in \mathbb{R}_+$ we let $[b] = n$ if $n \leq b < n + 1$ for some integer n . Finally, for time dependent functions we will use both the notations $S(t)$ and S_t interchangeably.

2.2 The algorithm

A natural variant of the MALA algorithm stems from the observation that π^N is the unique stationary measure of the SDE

$$dY_t = C_N \nabla \log \pi^N(Y_t) dt + \sqrt{2} dW_t^N, \tag{2.7}$$

where W^N is an X^N -valued Brownian motion with covariance operator C_N . The algorithm consists of discretising (2.7) using the Euler-Maruyama scheme and adding

a Metropolis accept-reject step so that the invariance of π^N is preserved. The variant on MALA which we study is therefore a Metropolis–Hastings algorithm with proposal

$$y^{k,N} = x^{k,N} - \delta \left(x^{k,N} + C_N \nabla \Psi^N(x^{k,N}) \right) + \sqrt{2\delta} C_N^{1/2} \xi^{k,N}, \tag{2.8}$$

where

$$\xi^{k,N} := \sum_{j=1}^N \xi_j^{k,N} \phi_j, \quad \xi_j^{k,N} \sim \mathcal{N}(0, 1) \text{ i.i.d.}$$

We stress that the Gaussian random variables $\xi_i^{k,N}$ are independent of each other and of the current position $x^{k,N}$. Motivated by the considerations made in the introduction (and that will be made more explicit in Sect. 4.1), in this paper we fix the choice

$$\delta := \frac{\ell}{N^{1/2}}. \tag{2.9}$$

If at step k the chain is at $x^{k,N}$, the algorithm proposes a move to $y^{k,N}$ defined by Eq. (2.8). The move is then accepted with probability

$$\alpha^N(x^{k,N}, y^{k,N}) := \frac{\pi^N(y^{k,N})q^N(y^{k,N}, x^{k,N})}{\pi^N(x^{k,N})q^N(x^{k,N}, y^{k,N})}, \tag{2.10}$$

where, for any $x^N, y^N \in \mathbb{R}^N \simeq X^N$,

$$q^N(x^N, y^N) \propto e^{-\frac{1}{4\delta} \|(y^N - x^N) - \delta \nabla \log \pi^N(x^N)\|_{C_N}^2}. \tag{2.11}$$

If the move to $y^{k,N}$ is accepted then $x^{k+1,N} = y^{k,N}$, if it is rejected the chain remains where it was, i.e. $x^{k+1,N} = x^{k,N}$. In short, the MALA chain is defined as follows:

$$x^{k+1,N} := \gamma^{k,N} y^{k,N} + (1 - \gamma^{k,N}) x^{k,N}, \quad x^{0,N} := \mathcal{P}^N(x^0), \tag{2.12}$$

where in the above

$$\gamma^{k,N} \stackrel{\mathcal{D}}{\sim} \text{Bernoulli}(\alpha^N(x^{k,N}, y^{k,N})); \tag{2.13}$$

that is, conditioned on $(x^{k,N}, y^{k,N})$, $\gamma^{k,N}$ has Bernoulli law with mean $\alpha^N(x^{k,N}, y^{k,N})$. Equivalently, we can write

$$\gamma^{k,N} = \mathbf{1}_{\{U^{k,N} \leq \alpha^N(x^{k,N}, y^{k,N})\}},$$

with $U^{k,N} \stackrel{\mathcal{D}}{\sim} \text{Uniform}[0, 1]$, independent of $x^{k,N}$ and $\xi^{k,N}$.

For fixed N , the chain $\{x^{k,N}\}_{k \geq 1}$ lives in $X^N \cong \mathbb{R}^N$ and samples from π^N . However, in view of the fact that we want to study the scaling limit of such a chain as $N \rightarrow \infty$,

the analysis is cleaner if it is carried out in \mathcal{H} ; therefore, the chain that we analyse is the chain $\{x^k\}_k \subseteq \mathcal{H}$ defined as follows: the first N components of the vector $x^k \in \mathcal{H}$ coincide with $x^{k,N}$ as defined above; the remaining components are not updated and remain equal to their initial value. More precisely, using (2.8) and (2.12), the chain x^k can be written in a component-wise notation as follows:

$$x_i^{k+1} = x_i^{k+1,N} = x_i^{k,N} - \gamma^{k,N} \left[\frac{\ell}{N^{1/2}} \left(x_i^{k,N} + [C_N \nabla \Psi^N(x^{k,N})]_i \right) + \sqrt{\frac{2\ell}{N^{1/2}}} \lambda_i \xi^{k,N} \right] \quad \forall i \leq N \tag{2.14}$$

and

$$x_i^{k+1} = x_i^k = 0 \quad \forall i \geq N + 1. \tag{2.15}$$

For the sake of clarity, we specify that $[C_N \nabla \Psi^N(x^{k,N})]_i$ denotes the i th component of the vector $C_N \nabla \Psi^N(x^{k,N}) \in \mathcal{H}^s$. From the above it is clear that the update rule (2.14) only updates the first N coordinates (with respect to the eigenbasis of \mathcal{C}) of the vector x^k . Therefore the algorithm evolves in the finite-dimensional subspace X^N . From now on we will avoid using the notation $\{x^k\}_k$ for the “extended chain” defined in \mathcal{H} , as it can be confused with the notation x^N , which instead is used throughout to denote a generic element of the space X^N .

We conclude this section by remarking that, if $x^{k,N}$ is given, the proposal $y^{k,N}$ only depends on the Gaussian noise $\xi^{k,N}$. Therefore the acceptance probability will be interchangeably denoted by $\alpha^N(x^N, y^N)$ or $\alpha^N(x^N, \xi^N)$.

2.3 Assumptions

In this section, we describe the assumptions on the covariance operator \mathcal{C} of the Gaussian measure $\pi_0 \stackrel{\mathcal{D}}{\sim} \mathcal{N}(0, \mathcal{C})$ and those on the functional Ψ . We fix a distinguished exponent $s \geq 0$ and assume that $\Psi : \mathcal{H}^s \rightarrow \mathbb{R}$ and $\text{Trace}_{\mathcal{H}^s}(\mathcal{C}_s) < \infty$. In other words, \mathcal{H}^s is the space that we were denoting with $\tilde{\mathcal{H}}$ in the introduction. Since

$$\text{Trace}_{\mathcal{H}^s}(\mathcal{C}_s) = \sum_{j=1}^{\infty} \lambda_j^2 j^{2s}, \tag{2.16}$$

the condition $\text{Trace}_{\mathcal{H}^s}(\mathcal{C}_s) < \infty$ implies that $\lambda_j j^s \rightarrow 0$ as $j \rightarrow \infty$. Therefore the sequence $\{\lambda_j j^s\}_j$ is bounded:

$$\lambda_j j^s \leq C, \tag{2.17}$$

for some constant $C > 0$ independent of j .

For each $x \in \mathcal{H}^s$ the derivative $\nabla \Psi(x)$ is an element of the dual $\mathcal{L}(\mathcal{H}^s, \mathbb{R})$ of \mathcal{H}^s , comprising the linear functionals on \mathcal{H}^s . However, we may identify $\mathcal{L}(\mathcal{H}^s, \mathbb{R}) = \mathcal{H}^{-s}$

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

$$\|\nabla\Psi(x)\|_{\mathcal{L}(\mathcal{H}^s, \mathbb{R})} = \|\nabla\Psi(x)\|_{-s}. \tag{2.18}$$

To avoid technical complications we assume that the gradient of $\Psi(x)$ is bounded and globally Lipschitz. More precisely, throughout this paper we make the following assumptions.

Assumption 2.1 The functional Ψ and covariance operator \mathcal{C} satisfy the following:

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

$$j^{-\kappa} \lesssim \lambda_j \lesssim j^{-\kappa}.$$

2. **Domain of Ψ** : the functional Ψ is defined everywhere on \mathcal{H}^s .
3. **Derivatives of Ψ** : The derivative of Ψ is bounded and globally Lipschitz:

$$\|\nabla\Psi(x)\|_{-s} \lesssim 1, \quad \|\nabla\Psi(x) - \nabla\Psi(y)\|_{-s} \lesssim \|x - y\|_s. \tag{2.19}$$

Remark 2.1 The condition $\kappa > s + \frac{1}{2}$ ensures that $\text{Trace}_{\mathcal{H}^s}(\mathcal{C}_s) < \infty$. Consequently, π_0 has support in \mathcal{H}^s ($\pi_0(\mathcal{H}^s) = 1$). □

Example 2.1 The functional $\Psi(x) = \sqrt{1 + \|x\|_s^2}$ satisfies all of the above.

Remark 2.2 Our assumptions on the change of measure (that is, on Ψ) are less general than those adopted in [14, 17] and related literature (see references therein). This is for purely technical reasons. In this paper we assume that Ψ grows linearly. If Ψ was assumed to grow quadratically, which is the case in the mentioned works, finding bounds on the moments of the chain $\{x^{k,N}\}_{k \geq 1}$ (much needed in all of the analysis) would become more involved than it already is, see Remark C.1. However, under our assumptions, the measure π (or π^N) is still, generically, of non-product form. □

We now explore the consequences of Assumption 2.1. The proofs of the following lemmas can be found in Appendix A.

Lemma 2.1 *Suppose that Assumption 2.1 holds. Then*

1. *The function $\mathcal{C}\nabla\Psi(x)$ is bounded and globally Lipschitz on \mathcal{H}^s , that is*

$$\|\mathcal{C}\nabla\Psi(x)\|_s \lesssim 1 \quad \text{and} \quad \|\mathcal{C}\nabla\Psi(x) - \mathcal{C}\nabla\Psi(y)\|_s \lesssim \|x - y\|_s. \tag{2.20}$$

Therefore, the function $F(z) := -z - \mathcal{C}\nabla\Psi(z)$ satisfies

$$\|F(x) - F(y)\|_s \lesssim \|x - y\|_s \quad \text{and} \quad \|F(x)\|_s \lesssim 1 + \|x\|_s. \tag{2.21}$$

2. The function $\Psi(x)$ is globally Lipschitz and therefore also $\Psi^N(x) := \Psi(\mathcal{P}^N(x))$ is globally Lipschitz:

$$\left| \Psi^N(y) - \Psi^N(x) \right| \lesssim \|y - x\|_s. \tag{2.22}$$

Before stating the next lemma, we observe that by definition of the projection operator \mathcal{P}^N we have that

$$\nabla \Psi^N = \mathcal{P}^N \circ \nabla \Psi \circ \mathcal{P}^N. \tag{2.23}$$

Lemma 2.2 *Suppose that Assumption 2.1 holds. Then the following holds for the function Ψ^N and for its the gradient:*

1. If the bounds (2.19) hold for Ψ , then they hold for Ψ^N as well:

$$\left\| \nabla \Psi^N(x) \right\|_{-s} \lesssim 1, \quad \left\| \nabla \Psi^N(x) - \nabla \Psi^N(y) \right\|_{-s} \lesssim \|x - y\|_s. \tag{2.24}$$

2. Moreover,

$$\left\| \mathcal{C}_N \nabla \Psi^N(x) \right\|_s \lesssim 1, \tag{2.25}$$

and

$$\left\| \mathcal{C}_N \nabla \Psi^N(x) \right\|_{\mathcal{C}_N} \lesssim 1. \tag{2.26}$$

We stress that in (2.24)–(2.26) the constant implied by the use of the notation “ \lesssim ” (see end of Sect. 2.1) is independent of N . Lastly, in what follows we will need the fact that, due assumptions on the covariance operator,

$$\mathbb{E} \left\| \mathcal{C}_N^{1/2} \xi^N \right\|_s^2 \lesssim 1, \quad \text{uniformly in } N, \tag{2.27}$$

where $\xi^N := \sum_{j=1}^N \xi_j \phi_j$ and $\xi_i \stackrel{\mathcal{D}}{\sim} \mathcal{N}(0, 1)$ i.d.d., see [15, (2.32)] or [14, first proof of Appendix A]

3 Existence and uniqueness for the limiting diffusion process

The main results of this section are Theorems 3.1, 3.2 and 3.3. Theorems 3.1 and 3.2 are concerned with establishing existence and uniqueness for Eqs. (1.10) and (1.11), respectively. Theorem 3.3 states the continuity of the Itô maps associated with Eqs. (1.10) and (1.11). The proofs of the main results of this paper (Theorems 4.1 and 4.2) rely heavily on the continuity of such maps, as we illustrate in Sect. 5. Once Lemma 3.1 below is established, the proofs of the theorems in this section are completely analogous to the proofs of those in [14, Section 4]. For this reason, we omit them and refer the reader to [14]. In what follows, recall that the definition of the functions α_ℓ , h_ℓ and b_ℓ has been given in (1.12), (1.13) and (1.14), respectively.

Lemma 3.1 *The functions $\alpha_\ell(s)$, $h_\ell(s)$ and $\sqrt{h_\ell(s)}$ are positive, globally Lipschitz continuous and bounded. The function $b_\ell(s)$ is globally Lipschitz and it is bounded above but not below. Moreover, for any $\ell > 0$, $b_\ell(s)$ is strictly positive for $s \in [0, 1)$, strictly negative for $s > 1$ and $b_\ell(1) = 0$.*

Proof of Lemma 3.1 When $s > 1$, $\alpha_\ell(s) = 1$ while for $s \leq 1$ $\alpha_\ell(s)$ has bounded derivative; therefore $\alpha_\ell(s)$ is globally Lipschitz. A similar reasoning gives the Lipschitzianity of the other functions. The further properties of b_ℓ are straightforward from the definition. □

In the case of (1.11) we have the following.

Theorem 3.1 *For any initial datum $S(0) > 0$, there exists a unique solution $S(t)$ to the ODE (1.11). The solution is strictly positive for any $t > 0$, it is bounded and has continuous first derivative for all $t \geq 0$. In particular*

$$\lim_{t \rightarrow \infty} S(t) = 1$$

and

$$0 \leq \min\{S(0), 1\} \leq S(t) \leq \max\{S(0), 1\}. \tag{3.1}$$

For (1.10) we have that:

Theorem 3.2 *Let Assumption 2.1 hold and consider Eq. (1.10), where $W(t)$ is any \mathcal{H}^s -valued C_s -Brownian motion and $S(t)$ is the solution of (1.11). Then for any initial condition $x^0 \in \mathcal{H}^s$ and any $T > 0$ there exists a unique solution of Eq. (1.10) in the space $C([0, T]; \mathcal{H}^s)$.*

Consider the deterministic equations

$$dz(t) = [-z(t) - C\nabla\Psi(z(t))]h_\ell(S(t)) dt + d\zeta(t), \quad z(0) = z^0 \tag{3.2}$$

and

$$d\mathfrak{S}(t) = b_\ell(\mathfrak{S}(t)) dt + dw(t), \quad \mathfrak{S}(0) = \mathfrak{S}^0, \tag{3.3}$$

where S is the solution of (1.11), $z^0 \in \mathcal{H}^s$, $\mathfrak{S}^0 \in \mathbb{R}$, and ζ and w are functions in $C([0, T]; \mathcal{H}^s)$ and $C([0, T]; \mathbb{R})$, respectively. Throughout the paper, we endow the spaces $C([0, T]; \mathcal{H}^s)$ and $C([0, T]; \mathbb{R})$ with the uniform topology. The following is the starting point of the continuous mapping arguments presented in Sect. 5.

Theorem 3.3 *Suppose that Assumption 2.1 is satisfied. Both (3.2) and (3.3) have unique solutions in $C([0, T]; \mathcal{H}^s)$ and $C([0, T]; \mathbb{R})$, respectively. The Itô maps*

$$\begin{aligned} \mathcal{J}_1 : \mathcal{H}^s \times C([0, T]; \mathcal{H}^s) &\longrightarrow C([0, T]; \mathcal{H}^s) \\ (z^0, \zeta) &\longrightarrow z \end{aligned}$$

and

$$\begin{aligned} \mathcal{J}_2 : \mathbb{R}_+ \times C([0, T]; \mathbb{R}) &\longrightarrow C([0, T]; \mathbb{R}) \\ (\mathfrak{S}^0, w) &\longrightarrow \mathfrak{S} \end{aligned}$$

are continuous.

4 Main theorems and heuristics of proofs

In order to state the main results, we first set

$$\mathcal{H}_\cap^s := \left\{ x \in \mathcal{H}^s : \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \frac{|x_i|^2}{\lambda_i^2} < \infty \right\}, \tag{4.1}$$

where we recall that in the above $x_i := \langle x, \phi_i \rangle$.

Theorem 4.1 *Let Assumption 2.1 hold and let $\delta = \ell/N^{\frac{1}{2}}$. Let $x^0 \in \mathcal{H}_\cap^s$ and $T > 0$. Then, as $N \rightarrow \infty$, the continuous interpolant $S^{(N)}(t)$ of the sequence $\{S^{k,N}\}_{k \in \mathbb{N}} \subseteq \mathbb{R}_+$ (defined in (1.16)) and started at $S^{0,N} = \frac{1}{N} \sum_{i=1}^N |x_i^0|^2 / \lambda_i^2$, converges in probability in $C([0, T]; \mathbb{R})$ to the solution $S(t)$ of the ODE (1.11) with initial datum $S^0 := \lim_{N \rightarrow \infty} S^{0,N}$.*

For the following theorem recall that the solution of (1.10) is interpreted precisely through Theorem 3.2 as a process driven by an \mathcal{H}^s -valued Brownian motion with covariance C_s , and solution in $C([0, T]; \mathcal{H}^s)$.

Theorem 4.2 *Let Assumption 2.1 hold let $\delta = \ell/N^{\frac{1}{2}}$. Let $x^0 \in \mathcal{H}_\cap^s$ and $T > 0$. Then, as $N \rightarrow \infty$, the continuous interpolant $x^{(N)}(t)$ of the chain $\{x^{k,N}\}_{k \in \mathbb{N}} \subseteq \mathcal{H}^s$ (defined in (1.9) and (2.14), respectively) with initial state $x^{0,N} := \mathcal{P}^N(x^0)$, converges weakly in $C([0, T]; \mathcal{H}^s)$ to the solution $x(t)$ of Eq. (1.10) with initial datum x^0 . We recall that the time-dependent function $S(t)$ appearing in (1.10) is the solution of the ODE (1.11), started at $S(0) := \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N |x_i^0|^2 / \lambda_i^2$.*

Both Theorems 4.1 and 4.2 assume that the initial datum of the chains $x^{k,N}$ is assigned deterministically. From our proofs it will be clear that the same statements also hold for random initial data, as long as (i) $x^{0,N}$ is not drawn at random from the target measure π^N or from any other measure which is a change of measure from π^N (i.e. we need to be starting out of stationarity) and (ii) $S^{0,N}$ and $x^{0,N}$ have bounded moments (bounded uniformly in N) of sufficiently high order and are independent of all the other sources of noise present in the algorithm. Notice moreover that the convergence in probability of Theorem 4.1 is equivalent to weak convergence, as the limit is deterministic.

The rigorous proof of the above results is contained in Sects. 5–8. In the remainder of this section we give heuristic arguments to justify our choice of scaling $\delta \propto N^{-1/2}$ and we explain how one can formally obtain the (fluid) ODE limit (1.11) for the

double sequence $S^{k,N}$ and the diffusion limit (1.10) for the chain $x^{k,N}$. We stress that the arguments of this section are only formal; therefore, we often use the notation “ \simeq ”, to mean “approximately equal”. That is, we write $A \simeq B$ when $A = B +$ “terms that are negligible” as N tends to infinity; we then justify these approximations, and the resulting limit theorems, in the following Sects. 5–8.

4.1 Heuristic analysis of the acceptance probability

As observed in [17, equation (2.21)], the acceptance probability (2.10) can be expressed as

$$\alpha^N(x^N, \xi^N) = 1 \wedge e^{Q^N(x^N, \xi^N)}, \tag{4.2}$$

where, using the notation (2.1), the function $Q^N(x, \xi)$ can be written as

$$Q^N(x^N, \xi^N) := -\frac{\delta}{4} \left(\|y^N\|_{C_N}^2 - \|x^N\|_{C_N}^2 \right) + r^N(x^N, \xi^N) \tag{4.3}$$

$$= \left[\frac{\delta^2}{2} \left(\|x^N\|_{C_N}^2 - \|C_N^{1/2} \xi^N\|_{C_N}^2 \right) \right] - \frac{\delta^3}{4} \|x^N\|_{C_N}^2 - \left(\frac{\delta^{3/2}}{\sqrt{2}} - \frac{\delta^{5/2}}{\sqrt{2}} \right) \langle x^N, C_N^{1/2} \xi^N \rangle_{C_N} + r_\Psi^N(x^N, \xi^N). \tag{4.4}$$

We do not give here a complete expression for the terms $r^N(x^N, \xi^N)$ and $r_\Psi^N(x^N, \xi^N)$. For the time being it is sufficient to point out that

$$\begin{aligned} r^N(x^N, \xi^N) &:= I_2^N + I_3^N \\ r_\Psi^N(x^N, \xi^N) &:= r^N(x^N, \xi^N) + \frac{(\delta^2 - \delta^3)}{2} \langle x^N, C_N \nabla \Psi^N(x^N) \rangle_{C_N} \\ &\quad - \frac{\delta^3}{4} \|C_N \nabla \Psi^N(x^N)\|_{C_N}^2 + \frac{\delta^{5/2}}{\sqrt{2}} \langle C_N \nabla \Psi^N(x^N), C_N^{1/2} \xi^N \rangle_{C_N} \end{aligned} \tag{4.5}$$

where I_2^N and I_3^N will be defined in (6.10) and (6.11), respectively. Because I_2^N and I_3^N depend on Ψ , r_Ψ^N contains all the terms where the functional Ψ appears; moreover r_Ψ^N vanishes when $\Psi = 0$. The analysis of Sect. 6 (see Lemma 6.4) will show that with our choice of scaling, $\delta = \ell/N^{1/2}$, the terms r^N and r_Ψ^N are negligible (for N large). Let us now illustrate the reason behind our choice of scaling. To this end, set $\delta = \ell/N^\zeta$ and observe the following two simple facts:

$$S^{k,N} = \frac{1}{N} \sum_{j=1}^N \frac{|x_j^{k,N}|^2}{\lambda_j^2} = \frac{1}{N} \|x^{k,N}\|_{C_N}^2 \tag{4.6}$$

and

$$\left\| \mathcal{C}_N^{1/2} \xi^N \right\|_{\mathcal{C}_N}^2 = \sum_{i=1}^N |\xi_i|^2 \simeq N, \tag{4.7}$$

the latter fact being true by the Law of Large Numbers. Neglecting the terms containing Ψ , at step k of the chain we have, formally,

$$\begin{aligned} Q^N(x^{k,N}, \xi^{k+1,N}) &\simeq \frac{\ell^2}{2} N^{1-2\zeta} \left(S^{k,N} - 1 \right) \\ &- \frac{\ell^3}{4} N^{1-3\zeta} S^{k,N} - \frac{\ell^{3/2}}{\sqrt{2}} N^{(1-3\zeta)/2} \frac{\langle x^{k,N}, \mathcal{C}_N^{1/2} \xi^{k,N} \rangle_{\mathcal{C}_N}}{\sqrt{N}} \end{aligned} \tag{4.8}$$

$$+ \frac{\ell^{5/2}}{\sqrt{2}} N^{(1-5\zeta)/2} \frac{\langle x^{k,N}, \mathcal{C}_N^{1/2} \xi^{k,N} \rangle_{\mathcal{C}_N}}{\sqrt{N}}. \tag{4.10}$$

The above approximation (which, we stress again, is only formal and will be made rigorous in subsequent sections) has been obtained from (4.4) by setting $\delta = \ell/N^\zeta$ and using (4.6) and (4.7), as follows:

$$\begin{aligned} \frac{\delta^2}{2} \left[\left\| x^N \right\|_{\mathcal{C}_N}^2 - \left\| \mathcal{C}_N^{1/2} \xi^N \right\|_{\mathcal{C}_N}^2 \right] &\simeq (4.8), \\ -\delta^3 \frac{\left\| x^N \right\|_{\mathcal{C}_N}^2}{4} - \frac{\delta^{3/2}}{\sqrt{2}} \langle x^N, \mathcal{C}_N^{1/2} \xi^N \rangle_{\mathcal{C}_N} &\simeq (4.9), \\ + \frac{\delta^{5/2}}{\sqrt{2}} \langle x^N, \mathcal{C}_N^{1/2} \xi^N \rangle_{\mathcal{C}_N} &= (4.10). \end{aligned} \tag{4.11}$$

Looking at the decomposition (4.8)–(4.10) of the function Q^N , we can now heuristically explain the reason why we are lead to choose $\zeta = 1/2$ when we start the chain out of stationarity, as opposed to the scaling $\zeta = 1/3$ when the chain is started in stationarity. This is explained in the following remark.

Remark 4.1 First notice that the expression (4.4) and the approximation (4.8)–(4.10) for Q^N are valid both in and out of stationarity, as the first is only a consequence of the definition of the Metropolis–Hastings algorithm and the latter is implied just by the properties of Ψ and by our definitions.

- If we start the chain in stationarity, i.e. $x_0^N \sim \pi^N$ (where π^N has been defined in (1.6)), then $x^{k,N} \sim \pi^N$ for every $k \geq 0$. As we have already observed, π^N is absolutely continuous with respect to the Gaussian measure $\pi_0^N \sim \mathcal{N}(0, \mathcal{C}_N)$; because all the almost sure properties are preserved under this change of measure, in the stationary regime most of the estimates of interest need to be shown only for $x^N \sim \pi_0^N$. In particular if $x^N \sim \pi_0^N$ then x^N can be represented as $x^N = \sum_{i=1}^N \lambda_i \rho_i \phi_i$, where ρ_i are i.i.d. $\mathcal{N}(0, 1)$. Therefore we can use the law of large numbers and observe that $\|x^N\|_{\mathcal{C}_N}^2 = \sum_{i=1}^N |\rho_i|^2 \simeq N$.

- Suppose we want to study the algorithm in stationarity and we therefore make the choice $\zeta = 1/3$. With the above point in mind, notice that if we start in stationarity then by the Law of Large numbers $N^{-1} \sum_{i=1}^N |\rho_i|^2 = S^{k,N} \rightarrow 1$ (as $N \rightarrow \infty$, with speed of convergence $N^{-1/2}$). Moreover, if $x^N \sim \pi_0^N$, by the Central Limit Theorem the term $\langle x^N, C_N^{1/2} \xi^N \rangle_{C_N} / \sqrt{N}$ is $O(1)$ and converges to a standard Gaussian. With these two observations in place we can then heuristically see that, with the choice $\zeta = 1/3$ the term in (4.10) are negligible as $N \rightarrow \infty$ while the terms in (4.9) are $O(1)$. The term in (4.8) can be better understood by looking at the LHS of (4.11) which, with $\zeta = 1/3$ and $x^N \sim \pi_0^N$, can be rewritten as

$$\frac{\ell^2}{2N^{2/3}} \sum_{i=1}^N (|\rho_i|^2 - |\xi_i|^2). \tag{4.12}$$

The expected value of the above expression is zero. If we apply the Central Limit Theorem to the i.i.d. sequence $\{|\rho_i|^2 - |\xi_i|^2\}_i$, (4.12) shows that (4.8) is $O(N^{1/2-2/3})$ and therefore negligible as $N \rightarrow \infty$. In conclusion, in the stationary case the only $O(1)$ terms are those in (4.9); therefore one has the heuristic approximation

$$Q^N(x, \xi) \sim \mathcal{N}\left(-\frac{\ell^3}{4}, \frac{\ell^3}{2}\right).$$

For more details on the stationary case see [17].

- If instead we start out of stationarity the choice $\zeta = 1/3$ is problematic. Indeed in [6, Lemma 3] the authors study the MALA algorithm to sample from an N -dimensional isotropic Gaussian and show that if the algorithm is started at a point x^0 such that $S(0) < 1$, then the acceptance probability degenerates to zero. Therefore, the algorithm stays stuck in its initial state and never proceeds to the next move, see [6, Figure 2] (to be more precise, as N increases the algorithm will take longer and longer to get unstuck from its initial state; in the limit, it will never move with probability 1). Therefore the choice $\zeta = 1/3$ cannot be the optimal one (at least not irrespective of the initial state of the chain) if we start out of stationarity. This is still the case in our context and one can heuristically see that the root of the problem lies in the term (4.8). Indeed if out of stationarity we still choose $\zeta = 1/3$ then, like before, (4.9) is still order one and (4.10) is still negligible. However, looking at (4.8), if x^0 is such that $S(0) < 1$ then, when $k = 0$, (4.8) tends to minus infinity; recalling (4.2), this implies that the acceptance probability of the first move tends to zero. To overcome this issue and make Q^N of order one (irrespective of the initial datum) so that the acceptance probability is of order one and does not degenerate to 0 or 1 when $N \rightarrow \infty$, we take $\zeta = 1/2$; in this way the terms in (4.8) are $O(1)$, all the others are small. Therefore, the intuition leading the analysis of the non-stationary regime hinges on the fact that, with our scaling,

$$Q^N(x^{k,N}, \xi^{k,N}) \simeq \frac{\ell^2}{2} (S^{k,N} - 1); \tag{4.13}$$

hence

$$\alpha^N(x^{k,N}, \xi^{k,N}) = (1 \wedge e^{Q^N(x^{k,N}, \xi^{k,N})}) \simeq \alpha_\ell(S^{k,N}), \tag{4.14}$$

where the function α_ℓ on the RHS of (4.14) is the one defined in (1.12). The approximation (4.13) is made rigorous in Lemma 6.4, while (4.14) is formalized in Sect. 6.1 (see in particular Proposition 6.1).

- Finally, we mention for completeness that, by arguing similarly to what we have done so far, if $\zeta < 1/2$ then the acceptance probability of the first move tends to zero when $S(0) < 1$. If $\zeta > 1/2$ then $Q^N \rightarrow 0$, so the acceptance probability tends to one; however the size of the moves is small and the algorithm explores the phase space slowly.

Remark 4.2 Notice that in stationarity the function Q^N is, to leading order, independent of ξ ; that is, Q^N and ξ are asymptotically independent (see [17, Lemma 4.5]). This can be intuitively explained because in stationarity the leading order term in the expression for Q^N is the term with $\delta^3 \|x\|^2$. We will show that also out of stationarity Q^N and ξ are asymptotically independent. In this case such an asymptotic independence can, roughly speaking, be motivated by the approximation (4.13), (as the interpolation of the chain $S^{k,N}$ converges to a deterministic limit). The asymptotic correlation of Q^N and the noise ξ is analysed in Lemma 6.5.

Remark 4.3 When one employs the more general proposal (1.18), assuming $\Psi \equiv 0$, the expression for Q^N becomes

$$Q^N(x^{k,N}, y^{k,N}) = -\frac{\delta}{4}(1 - 2\theta) \left(\|x^{k,N}\|_{\mathcal{C}^N}^2 - \|y^{k,N}\|_{\mathcal{C}^N}^2 \right).$$

So, if $\theta = 1/2$, the acceptance probability would be exactly one (for every N), i.e. the algorithm would be sampling exactly from the prior hence there is no need of rescaling δ with N .

4.2 Heuristic derivation of the weak limit of $S^{k,N}$

Let Y be any function of the random variables $\xi^{k,N}$ and $U^{k,N}$ (introduced in Sect. 2.2), for example the chain $x^{k,N}$ itself. Here and throughout the paper we use $\mathbb{E}_{x_0} [Y]$ to denote the expected value of Y with respect to the law of the variables $\xi^{k,N}$'s and $U^{k,N}$'s, with the initial state x_0 of the chain given deterministically; in other words, $\mathbb{E}_{x_0}(Y)$ denotes expectation with respect to all the sources of randomness present in Y . We will use the notation $\mathbb{E}_k [Y]$ for the conditional expectation of Y given $x^{k,N}$, $\mathbb{E}_k [Y] := \mathbb{E}_{x_0} [Y | x^{k,N}]$ (we should really be writing \mathbb{E}_k^N in place of \mathbb{E}_k , but to improve readability we will omit the further index N). Let us now decompose the chain $S^{k,N}$ into its drift and martingale parts:

$$S^{k+1,N} = S^{k,N} + \frac{1}{\sqrt{N}} b_\ell^{k,N} + \frac{1}{N^{1/4}} D^{k,N}, \tag{4.15}$$

where

$$b_\ell^{k,N} := \sqrt{N} \mathbb{E}_k[S^{k+1,N} - S^{k,N}] \tag{4.16}$$

and

$$D^{k,N} := N^{1/4} \left[S^{k+1,N} - S^{k,N} - \frac{1}{\sqrt{N}} b_\ell^{k,N}(x^{k,N}) \right]. \tag{4.17}$$

In this subsection we give the heuristics which underly the proof, given in subsequent sections, that the approximate drift $b_\ell^{k,N} = b_\ell^{k,N}(x^{k,N})$ converges to $b_\ell(S^{k,N})$,² where b_ℓ is the drift of (1.11), while the approximate diffusion $D^{k,N}$ tends to zero. This formally gives the result of Theorem 4.1. Let us formally argue such a convergence result. By (4.6) and (2.12),

$$S^{k+1,N} = \frac{1}{N} \sum_{j=1}^N \frac{|x_j^{k+1,N}|^2}{\lambda_j^2} = \frac{1}{N} \left(\gamma^{k,N} \|y^{k,N}\|_{\mathcal{C}_N}^2 + (1 - \gamma^{k,N}) \|x^{k,N}\|_{\mathcal{C}_N}^2 \right). \tag{4.18}$$

Therefore, again by (4.6),

$$\begin{aligned} b_\ell^{k,N} &= \sqrt{N} \mathbb{E}_k[S^{k+1,N} - S^{k,N}] = \frac{1}{\sqrt{N}} \mathbb{E}_k \left[\gamma^{k,N} \left(\|y^{k,N}\|_{\mathcal{C}_N}^2 - \|x^{k,N}\|_{\mathcal{C}_N}^2 \right) \right] \\ &= \frac{1}{\sqrt{N}} \mathbb{E}_k \left[(1 \wedge e^{Q^N(x^{k,N}, y^{k,N})}) \left(\|y^{k,N}\|_{\mathcal{C}_N}^2 - \|x^{k,N}\|_{\mathcal{C}_N}^2 \right) \right], \end{aligned} \tag{4.19}$$

where the second equality is a consequence of the definition of $\gamma^{k,N}$ (with a reasoning, completely analogous to the one in [14, last proof of Appendix A], see also (4.24). Using (4.3) (with $\delta = \ell/\sqrt{N}$), the fact that r^N is negligible and the approximation (4.13), the above gives

$$b_\ell^{k,N} = \sqrt{N} \mathbb{E}_k[S^{k+1,N} - S^{k,N}] \simeq -\frac{4}{\ell} \left(1 \wedge e^{\ell^2(S^{k,N}-1)/2} \right) \frac{\ell^2}{2} (S^{k,N} - 1) = b_\ell(S^{k,N}).$$

The above approximation is made rigorous in Lemma 7.5. As for the diffusion coefficient, it is easy to check (see proof of Lemma 7.2) that

$$N \mathbb{E}_k[S^{k+1,N} - S^{k,N}]^2 < \infty.$$

Hence the approximate diffusion tends to zero and one can formally deduce that (the interpolant of) $S^{k,N}$ converges to the ODE limit (1.11).

² Notice that $S^{k,N}$ is only a function of $x^{k,N}$.

4.3 Heuristic analysis of the limit of the chain $x^{k,N}$.

The drift-martingale decomposition of the chain $x^{k,N}$ is as follows:

$$x^{k+1,N} = x^{k,N} + \frac{1}{N^{1/2}} \Theta^{k,N} + \frac{1}{N^{1/4}} L^{k,N} \tag{4.20}$$

where $\Theta^{k,N} = \Theta^{k,N}(x^{k,N})$ is the approximate drift

$$\Theta^{k,N} := \sqrt{N} \mathbb{E}_k \left[x^{k+1,N} - x^{k,N} \right] \tag{4.21}$$

and

$$L^{k,N} := N^{1/4} \left[x^{k+1,N} - x^{k,N} - \frac{1}{\sqrt{N}} \Theta^{k,N}(x^{k,N}) \right] \tag{4.22}$$

is the approximate diffusion. In what follows we will use the notation $\Theta(x, S)$ for the drift of Eq. (1.10), i.e.

$$\Theta(x, S) = F(x)h_\ell(S), \quad (x, S) \in \mathcal{H}^S \times \mathbb{R}, \tag{4.23}$$

with $F(x)$ defined in Lemma 2.1. Again, we want to formally argue that the approximate drift $\Theta^{k,N}(x^{k,N})$ tends to $\Theta(x^{k,N}, S^{k,N})^3$ and the approximate diffusion $L^{k,N}$ tends to the diffusion coefficient of Eq. (1.10).

4.3.1 Approximate drift

As a preliminary consideration, observe that

$$\mathbb{E}_k \left(\gamma^{k,N} C_N^{1/2} \xi^{k,N} \right) = \mathbb{E}_k \left(\left(1 \wedge e^{Q^N(x^{k,N}, \xi^{k,N})} \right) C_N^{1/2} \xi^{k,N} \right), \tag{4.24}$$

see [14, equation (5.14)]. This fact will be used throughout the paper, often without mention. Coming to the chain $x^{k,N}$, a direct calculation based on (2.8) and on (2.12) gives

$$x^{k+1,N} - x^{k,N} = -\gamma^{k,N} \delta (x^{k,N} + C_N \nabla \Psi^N(x^{k,N})) + \gamma^{k,N} \sqrt{2\delta} C_N^{1/2} \xi^{k,N}. \tag{4.25}$$

Therefore, with the choice $\delta = \ell/\sqrt{N}$, we have

$$\begin{aligned} \Theta^{k,N} &= \sqrt{N} \mathbb{E}_k [x^{k+1,N} - x^{k,N}] \\ &= -\ell \mathbb{E}_k \left[\left(1 \wedge e^{Q^N(x^{k,N}, \xi^{k,N})} \right) (x^{k,N} + C_N \nabla \Psi^N(x^{k,N})) \right] \\ &\quad + N^{1/4} \sqrt{2\ell} \mathbb{E}_k \left[\left(1 \wedge e^{Q^N(x^{k,N}, \xi^{k,N})} \right) C_N^{1/2} \xi^{k,N} \right] \end{aligned} \tag{4.26}$$

³ Note that in the limit the dependence of the drift on $S^{k,N}$ becomes explicit.

The addend in (4.26) is asymptotically small (see Lemma 6.5 and notice that this addend would just be zero if Q^N and $\xi^{k,N}$ were uncorrelated); hence, using the heuristic approximations (4.13) and (4.14),

$$\begin{aligned} \Theta^{k,N} &= \sqrt{N} \mathbb{E}_k [x^{k+1,N} - x^{k,N}] \simeq -\ell \alpha_\ell (S^{k,N})(x^{k,N} + C_N \nabla \Psi^N(x^{k,N})) \\ &\stackrel{(1.13)}{=} -h_\ell (S^{k,N})(x^{k,N} + C_N \nabla \Psi^N(x^{k,N})); \end{aligned} \tag{4.27}$$

the right hand side of the above is precisely the limiting drift $\Theta(x^{k,N}, S^{k,N})$.

4.3.2 Approximate diffusion

We now look at the approximate diffusion of the chain $x^{k,N}$:

$$L^{k,N} := N^{1/4} (x^{k+1,N} - x^{k,N} - \mathbb{E}_k(x^{k+1,N} - x^{k,N})).$$

By definition,

$$\mathbb{E}_k \left\| L^{k,N} \right\|_s^2 = \sqrt{N} \mathbb{E}_k \left\| x^{k+1,N} - x^{k,N} \right\|_s^2 - \sqrt{N} \left\| \mathbb{E}_k \left(x^{k+1,N} - x^{k,N} \right) \right\|_s^2. \tag{4.28}$$

By (4.27) the second addend in the above is asymptotically small. Therefore

$$\begin{aligned} \mathbb{E}_k \left\| L^{k,N} \right\|_s^2 &\simeq \sqrt{N} \mathbb{E}_k \left\| x^{k+1,N} - x^{k,N} \right\|_s^2 \\ &\stackrel{(2.12),(4.25)}{\simeq} 2\ell \mathbb{E}_k \left\| \gamma^{k,N} C_N^{1/2} \xi^{k,N} \right\|_s^2 \\ &= 2\ell \mathbb{E}_k \sum_{j=1}^N j^{2s} \lambda_j^2 \left(1 \wedge e^{Q^N(x^{k,N}, \xi^{k,N})} \right) \left| \xi_j^{k,N} \right|^2. \end{aligned}$$

The above quantity is carefully studied in Lemma 6.6. However, intuitively, the heuristic approximation (4.14) (and the asymptotic independence of Q^N and ξ that (4.14) is a manifestation of) suffices to formally derive the limiting diffusion coefficient [i.e. the diffusion coefficient of (1.10)]:

$$\begin{aligned} \mathbb{E}_k \left\| L^{k,N} \right\|_s^2 &\simeq 2\ell \sum_{j=1}^N j^{2s} \lambda_j^2 \mathbb{E}_k \left[\left(1 \wedge e^{Q^N(x^{k,N}, y^{k,N})} \right) \left| \xi_j^{k,N} \right|^2 \right] \\ &\simeq 2\ell \sum_{j=1}^N j^{2s} \lambda_j^2 \mathbb{E}_k \left[\left(1 \wedge e^{\ell^2 (S^{k,N} - 1)/2} \right) \left| \xi_j^{k,N} \right|^2 \right] \end{aligned}$$

$$\begin{aligned} &\simeq 2\ell \sum_{j=1}^N j^{2s} \lambda_j^2 (1 \wedge e^{\ell^2 (S^{k,N}-1)/2}) \\ &\simeq 2\ell \operatorname{Trace}(C_s) \alpha_\ell (S^{k,N}) \stackrel{(1.13)}{=} 2\operatorname{Trace}(C_s) h_\ell (S^{k,N}). \end{aligned}$$

5 Continuous mapping argument

In this section we outline the argument which underlies the proofs of our main results. In particular, the proofs of Theorems 4.1 and 4.2 hinge on the continuous mapping arguments that we illustrate in the following Sects. 5.1 and 5.2, respectively. The details of the proofs are deferred to the next three sections: Sect. 6 contains some preliminary results that we employ in both proofs, in Sect. 7 contains the the proof of Theorem 4.1 and Sect. 8 that of Theorem 4.2.

5.1 Continuous mapping argument for (3.3)

Let us recall the definition of the chain $\{S^{k,N}\}_{k \in \mathbb{N}}$ and of its continuous interpolant $S^{(N)}$, introduced in (1.15) and (1.16), respectively. From the definition (1.16) of the interpolated process and the drift-martingale decomposition (4.15) of the chain $\{S^{k,N}\}_{k \in \mathbb{N}}$ we have that for any $t \in [t_k, t_{k+1})$,

$$\begin{aligned} S^{(N)}(t) &= (N^{1/2}t - k) \left[S^{k,N} + \frac{1}{\sqrt{N}} b_\ell^{k,N} + \frac{1}{N^{1/4}} D^{k,N} \right] + (k + 1 - tN^{1/2}) S^{k,N} \\ &= S^{k,N} + (t - t_k) b_\ell^{k,N} + N^{1/4} (t - t_k) D^{k,N}. \end{aligned}$$

Iterating the above we obtain

$$S^{(N)}(t) = S^{0,N} + (t - t_k) b_\ell^{k,N} + \frac{1}{\sqrt{N}} \sum_{j=0}^{k-1} b_\ell^{j,N} + w^N(t),$$

where

$$w^N(t) := \frac{1}{N^{1/4}} \sum_{j=0}^{k-1} D^{j,N} + N^{1/4} (t - t_k) D^{k,N} \quad t_k \leq t < t_{k+1}. \tag{5.1}$$

The expression for $S^{(N)}(t)$ can then be rewritten as

$$S^{(N)}(t) = S^{0,N} + \int_0^t b_\ell(S^{(N)}(v)) dv + \hat{w}^N(t), \tag{5.2}$$

having set

$$\hat{w}^N(t) := e^N(t) + w^N(t), \tag{5.3}$$

with

$$e^N(t) := (t - t_k)b_\ell^{k,N} + \frac{1}{\sqrt{N}} \sum_{j=0}^{k-1} b_\ell^{j,N} - \int_0^t b_\ell(S^{(N)}(v))dv. \tag{5.4}$$

Equation (5.2) shows that

$$S^{(N)} = \mathcal{J}_2(S^{0,N}, \hat{w}^N),$$

where \mathcal{J}_2 is the Itô map defined in the statement of Theorem 3.3. By the continuity of the map \mathcal{J}_2 , if we show that \hat{w}^N converges in probability in $C([0, T]; \mathbb{R})$ to zero, then $S^{(N)}(t)$ converges in probability to the solution of the ODE (1.11). We prove convergence of \hat{w}^N to zero in Sect. 7. In view of (5.3), we show the convergence in probability of \hat{w}^N to zero by proving that both e^N (Lemma 7.1) and w^N (Lemma 7.2) converge in $L_2(\Omega; C([0, T]; \mathbb{R}))$ to zero. Because $\{S^{0,N}\}_{N \in \mathbb{N}}$ is a deterministic sequence that converges to S^0 , we then have that $(S^{0,N}, \hat{w}^N)$ converges in probability to $(S^0, 0)$.

5.2 Continuous mapping argument for (3.2)

We now consider the chain $\{x^{k,N}\}_{k \in \mathbb{N}} \subseteq \mathcal{H}^s$, defined in (2.14). We act analogously to what we have done for the chain $\{S^{k,N}\}_{k \in \mathbb{N}}$. So we start by recalling the definition of the continuous interpolant $x^{(N)}$, Eq. (1.9) and the notation introduced at the beginning of Sect. 4.3. An argument analogous to the one used to derive (5.2) shows that for any $t \in [t_k, t_{k+1})$

$$\begin{aligned} x^{(N)}(t) &= x^{0,N} + (t - t_k)\Theta^{k,N} + \frac{1}{\sqrt{N}} \sum_{j=0}^k \Theta^{j,N} + \eta^N(t) \\ &= x^{0,N} + \int_0^t \Theta(x^{(N)}(v), S(v))dv + \hat{\eta}^N(t), \end{aligned} \tag{5.5}$$

where

$$\hat{\eta}^N(t) := d^N(t) + v^N(t) + \eta^N(t), \tag{5.6}$$

$$\eta^N(t) := N^{1/4}(t - t_k)L^{k,N} + \frac{1}{N^{1/4}} \sum_{j=1}^{k-1} L^{j,N}, \tag{5.7}$$

and

$$d^N(t) := (t - t_k)\Theta^{k,N} + \frac{1}{\sqrt{N}} \sum_{j=0}^{k-1} \Theta^{j,N} - \int_0^t \Theta(x^{(N)}(v), S^{(N)}(v))dv, \tag{5.8}$$

$$v^N(t) := \int_0^t \left[\Theta(x^{(N)}(v), S^{(N)}(v)) - \Theta(x^{(N)}(v), S(v)) \right] dv. \tag{5.9}$$

Equation (5.5) implies that

$$x^{(N)} = \mathcal{J}_1(x^{0,N}, \hat{\eta}^N), \tag{5.10}$$

where \mathcal{J}_1 is Itô map defined in the statement of Theorem 3.3. In Sect. 8 we prove that $\hat{\eta}^N$ converges weakly in $C([0, T]; \mathcal{H}^s)$ to the process η , where the process η is the diffusion part of Eq. (1.10), i.e.

$$\eta(t) := \int_0^t \sqrt{2h_\ell(S(v))} dW_v, \tag{5.11}$$

with W_v a \mathcal{H}^s -valued C_s -Brownian motion. Looking at (5.6), we prove the weak convergence of $\hat{\eta}^N$ to η by the following steps:

1. We prove that d^N converges in $L_2(\Omega; C([0, T]; \mathcal{H}^s))$ to zero (Lemma 8.1);
2. using the convergence in probability (in $C([0, T]; \mathbb{R})$) of $S^{(N)}$ to S , we show convergence in probability (in $C([0, T]; \mathcal{H}^s)$) of v^N to zero (Lemma 8.2);
3. we show that η^N converges weakly in $C([0, T]; \mathcal{H}^s)$ to the process η , defined in (5.11) (Lemma 8.3).

Because $\{x^{0,N}\}_{N \in \mathbb{N}}$ is a deterministic sequence that converges to x^0 , the above three steps (and Slutsky’s Theorem) imply that $(x^{0,N}, \hat{\eta}^N)$ converges weakly to (x^0, η) . Now observe that $x(t) = \mathcal{J}_1(x^0, \eta(t))$, where $x(t)$ is the solution of the SDE (1.10). The continuity of the map \mathcal{J}_1 (Theorem 3.3), (5.10) and the Continuous Mapping Theorem then imply that the sequence $\{x^{(N)}\}_{N \in \mathbb{N}}$ converges weakly to the solution of the SDE (1.10), thus establishing Theorem 4.2.

6 Preliminary estimates and analysis of the acceptance probability

This section gathers several technical results. In Lemma 6.1 we study the size of the jumps of the chain. Lemma 6.2 contains uniform bounds on the moments of the chains $\{x^{k,N}\}_{k \in \mathbb{N}}$ and $\{S^{k,N}\}_{k \in \mathbb{N}}$, much needed in Sects. 7 and 8. In Section 6.1 we detail the analysis of the acceptance probability. This allows us to quantify the correlations between $y^{k,N}$ and the noise $\xi^{k,N}$, Sect. 6.2. Throughout the paper, when referring to the function Q^N defined in (4.3), we use interchangeably the notation $Q^N(x^{k,N}, y^{k,N})$ and $Q^N(x^{k,N}, \xi^{k,N})$ (as we have already remarked, given $x^{k,N}$, the proposal $y^{k,N}$ is only a function of $\xi^{k,N}$).

Lemma 6.1 *Let $q \geq 1/2$ be a real number. Under Assumption 2.1 the following holds:*

$$\mathbb{E}_k \left\| y^{k,N} - x^{k,N} \right\|_s^{2q} \lesssim \frac{1}{N^{q/2}} \left(1 + \left\| x^{k,N} \right\|_s^{2q} \right) \tag{6.1}$$

and

$$\mathbb{E}_k \left\| y^{k,N} - x^{k,N} \right\|_{C_N}^{2q} \lesssim (S^{k,N})^q + N^{q/2}. \tag{6.2}$$

Therefore,

$$\mathbb{E}_k \left\| x^{k+1,N} - x^{k,N} \right\|_s^{2q} \lesssim \frac{1}{N^{q/2}} \left(1 + \left\| x^{k,N} \right\|_s^{2q} \right), \tag{6.3}$$

and

$$\mathbb{E}_k \left\| x^{k+1,N} - x^{k,N} \right\|_{\mathcal{C}_N}^{2q} \lesssim (S^{k,N})^q + N^{q/2}. \tag{6.4}$$

Proof of Lemma 6.1 By definition of the proposal $y^{k,N}$, Eq. (2.8),

$$\begin{aligned} \left\| y^{k,N} - x^{k,N} \right\|_s^{2q} &= \left\| \delta(x^{k,N} + C_N \nabla \Psi^N(x^{k,N})) + \sqrt{2\delta} C_N^{1/2} \xi^{k,N} \right\|_s^{2q} \\ &\lesssim \frac{1}{N^q} \left(\left\| x^{k,N} \right\|_s^{2q} + \left\| C_N \nabla \Psi^N(x^{k,N}) \right\|_s^{2q} \right) \\ &\quad + \frac{1}{N^{q/2}} \left\| C_N^{1/2} \xi^{k,N} \right\|_s^{2q}. \end{aligned}$$

Thus, using (2.25) and (2.27), we have

$$\begin{aligned} \mathbb{E}_k \left\| y^{k,N} - x^{k,N} \right\|_s^{2q} &\lesssim \frac{1}{N^q} \left(1 + \left\| x^{k,N} \right\|_s^{2q} \right) + \frac{1}{N^{q/2}} \\ &\lesssim \frac{1}{N^{q/2}} \left(1 + \left\| x^{k,N} \right\|_s^{2q} \right), \end{aligned}$$

which proves (6.1). Equation (6.2) follows similarly:

$$\begin{aligned} \mathbb{E}_k \left\| y^{k,N} - x^{k,N} \right\|_{\mathcal{C}_N}^{2q} &\lesssim \frac{1}{N^q} \left(\left\| x^{k,N} \right\|_{\mathcal{C}_N}^{2q} + \left\| C_N \nabla \Psi^N(x^{k,N}) \right\|_{\mathcal{C}_N}^{2q} \right) \\ &\quad + \frac{1}{N^{q/2}} \mathbb{E}_k \left\| C_N^{1/2} \xi^{k,N} \right\|_{\mathcal{C}_N}^{2q}. \end{aligned}$$

Since $\left\| C_N^{1/2} \xi^{k,N} \right\|_{\mathcal{C}_N}^2 = \sum_{j=1}^N (\xi_j^{k,N})^2$ has chi-squared law, applying Stirling’s formula for the Gamma function $\Gamma : \mathbb{R} \rightarrow \mathbb{R}$ we obtain

$$\mathbb{E}_k \left\| C_N^{1/2} \xi^{k,N} \right\|_{\mathcal{C}_N}^{2q} \lesssim \frac{\Gamma(q + N/2)}{\Gamma(N/2)} \lesssim N^q. \tag{6.5}$$

Hence, using (2.26), the desired bound follows. Finally, recalling the definition of the chain, Eq. (2.12), the bounds (6.3) and (6.4) are clearly a consequence of (6.1) and (6.2), respectively, since either $x^{k+1,N} = y^{k,N}$ (if the proposed move is accepted) or $x^{k+1,N} = x^{k,N}$ (if the move is rejected). □

Lemma 6.2 *If Assumption 2.1 holds, then, for every $q \geq 1$, we have*

$$\mathbb{E}_{x^0} (S^{k,N})^q \lesssim 1 \tag{6.6}$$

$$\mathbb{E}_{x^0} \left\| x^{k,N} \right\|_s^q \lesssim 1, \tag{6.7}$$

uniformly over $N \in \mathbb{N}$ and $k \in \{0, 1, \dots, [T\sqrt{N}]\}$.

Proof of Lemma 6.2 The proof of this lemma can be found in Appendix C. □

6.1 Acceptance probability

The main result of this section is Proposition 6.1, which we obtain as a consequence of Lemma 6.3 (below) and Lemma 6.2. Proposition 6.1 formalizes the heuristic approximation (4.14).

Lemma 6.3 (Acceptance probability) *Let Assumption 2.1 hold and recall the Definitions (4.2) and (1.12). Then the following holds:*

$$\mathbb{E}_k \left| \alpha^N(x^{k,N}, \xi^{k,N}) - \alpha_\ell(S^{k,N}) \right|^2 \lesssim \frac{1 + (S^{k,N})^2 + \|x^{k,N}\|_s^2}{\sqrt{N}}.$$

Before proving Lemma 6.3, we state Proposition 6.1.

Proposition 6.1 *If Assumption 2.1 holds then*

$$\lim_{N \rightarrow \infty} \mathbb{E}_{x^0} \left| \alpha^N(x^{k,N}, y^{k,N}) - \alpha_\ell(S^{k,N}) \right|^2 = 0.$$

Proof This is a corollary of Lemmas 6.3 and 6.2. □

Proof of Lemma 6.3 The function $z \mapsto 1 \wedge e^z$ on \mathbb{R} is globally Lipschitz with Lipschitz constant 1. Therefore, by (1.12) and (4.2),

$$\mathbb{E}_k \left| \alpha^N(x^{k,N}, y^{k,N}) - \alpha_\ell(S^{k,N}) \right|^2 \leq \mathbb{E}_k \left| Q^N(x^{k,N}, y^{k,N}) - \frac{\ell^2(S^{k,N} - 1)}{2} \right|^2.$$

The result is now a consequence of (6.15) below. □

To analyse the acceptance probability it is convenient to decompose Q^N as follows:

$$Q^N(x^N, y^N) = I_1^N(x^N, y^N) + I_2^N(x^N, y^N) + I_3^N(x^N, y^N) \tag{6.8}$$

where

$$\begin{aligned}
 I_1^N(x^N, y^N) &:= -\frac{1}{2} \left[\left\| y^N \right\|_{\mathcal{C}_N}^2 - \left\| x^N \right\|_{\mathcal{C}_N}^2 \right] \\
 &\quad - \frac{1}{4\delta} \left[\left\| x^N - (1 - \delta)y^N \right\|_{\mathcal{C}_N}^2 - \left\| y^N - (1 - \delta)x^N \right\|_{\mathcal{C}_N}^2 \right] \\
 &= -\frac{\delta}{4} \left(\left\| y^N \right\|_{\mathcal{C}_N}^2 - \left\| x^N \right\|_{\mathcal{C}_N}^2 \right), \tag{6.9}
 \end{aligned}$$

$$\begin{aligned}
 I_2^N(x^N, y^N) &:= -\frac{1}{2} \left[\left\langle x^N - (1 - \delta)y^N, \mathcal{C}_N \nabla \Psi^N(y^N) \right\rangle_{\mathcal{C}_N} \right. \\
 &\quad \left. - \left\langle y^N - (1 - \delta)x^N, \mathcal{C}_N \nabla \Psi^N(x^N) \right\rangle_{\mathcal{C}_N} \right] \\
 &\quad - (\Psi^N(y^N) - \Psi^N(x^N)), \tag{6.10}
 \end{aligned}$$

$$I_3^N(x^N, y^N) := -\frac{\delta}{4} \left[\left\| \mathcal{C}_N \nabla \Psi^N(y^N) \right\|_{\mathcal{C}_N}^2 - \left\| \mathcal{C}_N \nabla \Psi^N(x^N) \right\|_{\mathcal{C}_N}^2 \right]. \tag{6.11}$$

Lemma 6.4 *Let Assumption 2.1 hold. With the notation introduced above, we have:*

$$\mathbb{E}_k \left| I_1^N(x^{k,N}, y^{k,N}) - \frac{\ell^2(S^{k,N} - 1)}{2} \right|^2 \lesssim \frac{\|x^{k,N}\|_s^2}{N^2} + \frac{(S^{k,N})^2}{\sqrt{N}} + \frac{1}{N} \tag{6.12}$$

$$\mathbb{E}_k \left| I_2^N(x^{k,N}, y^{k,N}) \right|^2 \lesssim \frac{1 + \|x^{k,N}\|_s^2}{\sqrt{N}} \tag{6.13}$$

$$\mathbb{E}_k \left| I_3^N(x^{k,N}, y^{k,N}) \right|^2 \lesssim \frac{1}{N}. \tag{6.14}$$

Therefore,

$$\mathbb{E}_k \left| Q^N(x^{k,N}, y^{k,N}) - \frac{\ell^2(S^{k,N} - 1)}{2} \right|^2 \lesssim \frac{1 + (S^{k,N})^2 + \|x^{k,N}\|_s^2}{\sqrt{N}}. \tag{6.15}$$

Proof of Lemma 6.4 We consecutively prove the three bounds in the statement.

- Proof of (6.12). Using (2.8), we rewrite I_1^N as

$$\begin{aligned}
 &I_1^N(x^{k,N}, y^{k,N}) \\
 &= -\frac{\delta}{4} \left(\left\| (1 - \delta)x^{k,N} - \delta \mathcal{C}_N \nabla \Psi^N(x^{k,N}) + \sqrt{2\delta} \mathcal{C}_N^{1/2} \xi^{k,N} \right\|_{\mathcal{C}_N}^2 - \left\| x^{k,N} \right\|_{\mathcal{C}_N}^2 \right).
 \end{aligned}$$

Expanding the above we obtain:

$$I_1^N(x^{k,N}, y^{k,N}) - \frac{\ell^2(S^{k,N} - 1)}{2} = -\left(\frac{\delta^2}{2} \left\| \mathcal{C}_N^{1/2} \xi^{k,N} \right\|_{\mathcal{C}_N}^2 - \frac{\ell^2}{2} \right)$$

$$+ (r_\psi^N - r^N) + r_\xi^N + r_x^N, \tag{6.16}$$

where the difference $(r_\psi^N - r^N)$ is defined in (4.5) and we set

$$r_\xi^N := -\frac{(\delta^{3/2} - \delta^{5/2})}{\sqrt{2}} \left\langle x^{k,N}, C_N^{1/2} \xi^{k,N} \right\rangle_{C_N}, \tag{6.17}$$

$$r_x^N := -\frac{\delta^3}{4} \left\| x^{k,N} \right\|_{C_N}^2. \tag{6.18}$$

For the reader’s convenience we rearrange (4.5) below:

$$\begin{aligned} r_\psi^N - r^N &= \frac{\delta^2 - \delta^3}{2} \left\langle x^{k,N}, C_N \nabla \Psi^N(x^{k,N}) \right\rangle_{C_N} \\ &\quad - \frac{\delta^3}{4} \left\| C_N \nabla \Psi^N(x^{k,N}) \right\|_{C_N}^2 + \frac{\delta^{5/2}}{\sqrt{2}} \left\langle C_N \nabla \Psi^N(x^{k,N}), C_N^{1/2} \xi^{k,N} \right\rangle_{C_N}. \end{aligned} \tag{6.19}$$

We come to bound all of the above terms, starting from (6.19). To this end, let us observe the following:

$$\left| \left\langle x^{k,N}, C_N \nabla \Psi^N(x^{k,N}) \right\rangle_{C_N} \right|^2 = \left| \sum_{i=1}^N x_i^{k,N} [\nabla \Psi^N(x^{k,N})]_i \right|^2 \tag{6.20}$$

$$\stackrel{(2.6)}{\leq} \left\| x^{k,N} \right\|_s^2 \left\| \nabla \Psi^N(x^{k,N}) \right\|_{-s}^2 \stackrel{(2.24)}{\lesssim} \left\| x^{k,N} \right\|_s^2. \tag{6.21}$$

Moreover,

$$\mathbb{E}_k \left\| C_N^{1/2} \xi^{k,N} \right\|_{C_N}^2 = \mathbb{E}_k \sum_{j=1}^N |\xi_j|^2 = N,$$

hence

$$\left| \left\langle C_N \nabla \Psi^N(x^{k,N}), C_N^{1/2} \xi^{k,N} \right\rangle_{C_N} \right|^2 \leq \left\| C_N \nabla \Psi^N(x^{k,N}) \right\|_{C_N}^2 \left\| C_N^{1/2} \xi^{k,N} \right\|_{C_N}^2 \stackrel{(2.26)}{\lesssim} N.$$

From (6.19), (6.20), (2.26) and the above,

$$\mathbb{E}_k \left| r_\psi^N - r^N \right|^2 \lesssim \frac{\left\| x^{k,N} \right\|_s^2}{N^2} + \frac{1}{N^{3/2}}. \tag{6.22}$$

By (6.17),

$$\begin{aligned} \mathbb{E}_k \left| r_\xi^N \right|^2 &\lesssim \frac{1}{N^{3/2}} \mathbb{E}_k \left| \left\langle x^{k,N}, C_N^{1/2} \xi^{k,N} \right\rangle_{C_N} \right|^2 \\ &= \frac{1}{N^{3/2}} \mathbb{E}_k \left(\sum_{i=1}^N \frac{x_i^{k,N} \xi_i^{k,N}}{\lambda_i} \right)^2 = \frac{1}{\sqrt{N}} S^{k,N}, \end{aligned} \tag{6.23}$$

where in the last equality we have used the fact that $\{\xi_i^{k,N} : i = 1, \dots, N\}$ are independent, zero mean, unit variance normal random variables (independent of $x^{k,N}$) and (4.6). As for r_x^N ,

$$\mathbb{E}_k \left| r_x^N \right|^2 \lesssim \frac{1}{N^3} \left\| x^{k,N} \right\|_{C_N}^4 \stackrel{(4.6)}{=} \frac{(S^{k,N})^2}{N}.$$

Lastly,

$$\tilde{r}^N := \frac{\delta^2}{2} \left\| C_N^{1/2} \xi^{k,N} \right\|_{C_N}^2 - \frac{\ell^2}{2} = \frac{\ell^2}{2} \left(\frac{1}{N} \sum_{j=1}^N \xi_j^2 - 1 \right).$$

Since $\sum_{j=1}^N \xi_j^2$ has chi-squared law, $\mathbb{E}_k |\tilde{r}^N|^2 \lesssim \text{Var} \left(N^{-1} \sum_{j=1}^N \xi_j^2 \right) \lesssim N^{-1}$, by (6.5). Combining all of the above, we obtain the desired bound.

- Proof of (6.13) From (6.10),

$$\begin{aligned} I_2^N(x^{k,N}, y^{k,N}) &= - \left[\Psi^N(y^{k,N}) - \Psi^N(x^{k,N}) - \left\langle y^{k,N} - x^{k,N}, \nabla \Psi^N(x^{k,N}) \right\rangle \right] \\ &\quad + \frac{1}{2} \left\langle y^{k,N} - x^{k,N}, \nabla \Psi^N(y^{k,N}) - \nabla \Psi^N(x^{k,N}) \right\rangle \\ &\quad + \frac{\delta}{2} \left(\left\langle x^{k,N}, \nabla \Psi^N(x^{k,N}) \right\rangle - \left\langle y^{k,N}, \nabla \Psi^N(y^{k,N}) \right\rangle \right) =: \sum_{j=1}^3 d_j, \end{aligned}$$

where d_j is the addend on line j of the above array. Using (2.22), (2.24), (2.6) and Lemma 6.1, we have

$$\mathbb{E}_k |d_1|^2 \lesssim \mathbb{E}_k \left\| y^{k,N} - x^{k,N} \right\|_s^2 \lesssim \frac{1 + \left\| x^{k,N} \right\|_s^2}{\sqrt{N}}.$$

By the first inequality in (2.24),

$$\left\| \nabla \Psi^N(y^{k,N}) - \nabla \Psi^N(x^{k,N}) \right\|_{-s} \lesssim 1.$$

Consequently, again by (2.6) and Lemma 6.1,

$$\mathbb{E}_k |d_2|^2 \lesssim \mathbb{E}_k \left\| y^{k,N} - x^{k,N} \right\|_s^2 \lesssim \frac{1 + \|x^{k,N}\|_s^2}{\sqrt{N}}.$$

Next, applying (2.6) and (2.24) gives

$$\begin{aligned} |d_3| &\leq \frac{\|x^{k,N}\|_s \|\nabla \Psi^N(x^{k,N})\|_{-s} + \|y^{k,N}\|_s \|\nabla \Psi^N(y^{k,N})\|_{-s}}{\sqrt{N}} \\ &\lesssim \frac{\|x^{k,N}\|_s + \|y^{k,N}\|_s}{\sqrt{N}} \lesssim \frac{\|x^{k,N}\|_s + \|y^{k,N} - x^{k,N}\|_s}{\sqrt{N}}. \end{aligned}$$

Thus, applying Lemma 6.1 then gives the desired bound.

- Proof of (6.14) This follows directly from (2.25). □

6.2 Correlations between acceptance probability and noise $\xi^{k,N}$

Recall the definition of $\gamma^{k,N}$, Eq. (2.13), and let

$$\varepsilon^{k,N} := \gamma^{k,N} C_N^{1/2} \xi^{k,N}. \tag{6.24}$$

The study of the properties of $\varepsilon^{k,N}$ is the object of the next two lemmata, which have a central role in the analysis: Lemma 6.5 (and Lemma 6.2) establishes the decay of correlations between the acceptance probability and the noise $\xi^{k,N}$. Lemma 6.6 formalizes the heuristic arguments presented in Sect. 4.3.2.

Lemma 6.5 *If Assumption 2.1 holds, then*

$$\left\| \mathbb{E}_k \varepsilon^{k,N} \right\|_s^2 \lesssim \frac{1 + \|x^{k,N}\|_s^2}{\sqrt{N}}. \tag{6.25}$$

Therefore,

$$\left\langle \mathbb{E}_k \varepsilon^{k,N}, x^{k,N} \right\rangle_s = \mathbb{E}_k \left\langle \gamma^{k,N} C_N^{1/2} \xi^{k,N}, x^{k,N} \right\rangle_s \lesssim \frac{1}{N^{1/4}} \left(1 + \|x^{k,N}\|_s^2 \right). \tag{6.26}$$

Lemma 6.6 *Let Assumption 2.1 hold. Then, with the notation introduced so far,*

$$\left| \mathbb{E}_k \left\| \varepsilon^{k,N} \right\|_s^2 - \text{Trace}_{\mathcal{H}^s}(C_s) \alpha_\ell(S^{k,N}) \right| \lesssim \frac{1 + S^{k,N} + \|x^{k,N}\|_s}{N^{1/4}}.$$

The proofs of the above lemmata can be found in Appendix B. Notice that if $\xi^{k,N}$ and $\gamma^{k,N}$ (equivalently $\xi^{k,N}$ and Q^N) were uncorrelated, the statements of Lemmas 6.5 and 6.6 would be trivially true.

7 Proof of Theorem 4.1

As explained in Sect. 5.1, due to the continuity of the map \mathcal{J}_2 (defined in Theorem 3.3), in order to prove Theorem 4.1 all we need to show is convergence in probability of $\hat{w}^N(t)$ to zero. Looking at the definition of $\hat{w}^N(t)$, Eq. (5.3), the convergence in probability (in $C([0, T]; \mathbb{R})$) of $\hat{w}^N(t)$ to zero is consequence of Lemmas 7.1 and 7.2 below. We prove Lemma 7.1 in Sect. 7.1 and Lemma 7.2 in Sect. 7.2.

Lemma 7.1 *Let Assumption 2.1 hold and recall the definition (5.4) of the process $e^N(t)$; then*

$$\lim_{N \rightarrow \infty} \mathbb{E}_{x^0} \left(\sup_{t \in [0, T]} |e^N(t)| \right)^2 = 0.$$

Lemma 7.2 *Let Assumption 2.1 hold and recall the definition (5.1) of the process $w^N(t)$; then*

$$\lim_{N \rightarrow \infty} \mathbb{E}_{x^0} \left(\sup_{t \in [0, T]} |w^N(t)| \right)^2 = 0.$$

7.1 Analysis of the drift

In view of what follows, it is convenient to introduce the piecewise constant interpolant of the chain $\{S^{k,N}\}_{k \in \mathbb{N}}$:

$$\bar{S}^{(N)}(t) := S^{k,N}, \quad t_k \leq t < t_{k+1}, \tag{7.1}$$

where $t_k = k/\sqrt{N}$.

Proof of Lemma 7.1 From (7.1), for any $t_k \leq t < t_{k+1}$ we have

$$\begin{aligned} \int_0^t b_\ell(\bar{S}_v^{(N)}) dv &= \int_{t_k}^t b_\ell(\bar{S}_v^{(N)}) dv + \sum_{j=1}^{k-1} \int_{t_{j-1}}^{t_j} b_\ell(\bar{S}_v^{(N)}) dv \\ &= (t - t_k) b_\ell(S^{k,N}) + \frac{1}{\sqrt{N}} \sum_{j=1}^{k-1} b_\ell(S^{j,N}). \end{aligned}$$

With this observation, we can then decompose $e^N(t)$ as

$$e^N(t) = e_1^N(t) - e_2^N(t),$$

where

$$e_1^N(t) := (t - t_k)(b_\ell^{k,N} - b_\ell(S^{k,N})) + \frac{1}{\sqrt{N}} \sum_{j=0}^{k-1} [b_\ell^{j,N} - b_\ell(S^{j,N})] \tag{7.2}$$

$$e_2^N(t) := \int_0^t [b_\ell(S_v^{(N)}) - b_\ell(\bar{S}_v^{(N)})] dv. \tag{7.3}$$

The result is now a consequence of Lemmas 7.3 and 7.4 below, which we first state and then consecutively prove. \square

Lemma 7.3 *If Assumption 2.1 holds, then*

$$\lim_{N \rightarrow \infty} \mathbb{E}_{x^0} \left(\sup_{t \in [0, T]} |e_1^N(t)| \right)^2 = 0.$$

Lemma 7.4 *If Assumption 2.1 holds, then*

$$\lim_{N \rightarrow \infty} \mathbb{E}_{x^0} \left(\sup_{t \in [0, T]} |e_2^N(t)| \right)^2 = 0.$$

Proof of Lemma 7.3 Denoting $E^{k,N} := b_\ell^{k,N} - b_\ell(S^{k,N})$, by (discrete) Jensen’s inequality we have

$$\begin{aligned} \sup_{t \in [0, T]} |e_1^N(t)|^2 &= \sup_{t \in [0, T]} \left| (t - t_k)E^{k,N} + \frac{1}{\sqrt{N}} \sum_{j=0}^{k-1} E^{j,N} \right|^2 \\ &\lesssim \frac{1}{\sqrt{N}} \sum_{j=0}^{[T\sqrt{N}]-1} |E^{j,N}|^2. \end{aligned}$$

Using Lemma 7.5 below, we obtain

$$\frac{1}{\sqrt{N}} \sum_{j=0}^{[T\sqrt{N}]-1} |E^{j,N}|^2 \lesssim \frac{1}{\sqrt{N}} \sum_{k=0}^{[T\sqrt{N}]-1} \frac{1 + (S^{k,N})^4 + \|x^{k,N}\|_s^4}{\sqrt{N}}.$$

Taking expectations on both sides and applying Lemma 6.2 completes the proof. \square

Lemma 7.5 *Let Assumption 2.1 hold. Then, for any $N \in \mathbb{N}$ and $k \in \{0, 1, \dots, [T\sqrt{N}]\}$,*

$$|E^{k,N}|^2 = |b_\ell^{k,N} - b_\ell(S^{k,N})|^2 \lesssim \frac{1 + (S^{k,N})^4 + \|x^{k,N}\|_s^4}{\sqrt{N}}.$$

Proof of Lemma 7.5 Define

$$Y_k^N := \frac{\|y^{k,N}\|_{C_N}^2 - \|x^{k,N}\|_{C_N}^2}{\sqrt{N}}, \quad \tilde{Y}_k^N := 2\ell(1 - S^{k,N}).$$

Then, from (4.19), (4.2), (1.12) and (1.14), we obtain

$$\begin{aligned} |b_\ell^{k,N} - b_\ell(S^{k,N})|^2 &= \left| \mathbb{E}_k \left(\alpha^N(x^{k,N}, y^{k,N}) Y_k^N \right) - \alpha_\ell(S^{k,N}) \tilde{Y}_k^N \right|^2 \\ &\leq \mathbb{E}_k \left| \alpha^N(x^{k,N}, y^{k,N}) Y_k^N - \alpha_\ell(S^{k,N}) \tilde{Y}_k^N \right|^2 \\ &\lesssim \mathbb{E}_k \left[\left| \alpha^N(x^{k,N}, y^{k,N}) \right|^2 \left| Y_k^N - \tilde{Y}_k^N \right|^2 \right] \\ &\quad + \mathbb{E}_k \left[\left| \tilde{Y}_k^N \right|^2 \left| \alpha^N(x^{k,N}, y^{k,N}) - \alpha_\ell(S^{k,N}) \right|^2 \right]. \end{aligned}$$

Since $|\alpha^N(x^{k,N}, y^{k,N})| \leq 1$ and \tilde{Y}_k^N is a function of $x^{k,N}$ only, we can further estimate the above as follows:

$$\left| b_\ell^{k,N} - b_\ell(S^{k,N}) \right|^2 \lesssim \mathbb{E}_k \left| Y_k^N - \tilde{Y}_k^N \right|^2 + \left| \tilde{Y}_k^N \right|^2 \mathbb{E}_k \left| \alpha^N(x^{k,N}, y^{k,N}) - \alpha_\ell(S^{k,N}) \right|^2. \tag{7.4}$$

From the definition of I_1^N , Eq. (6.9), we have

$$Y_k^N = -\frac{4}{\ell} I_1^N(x^{k,N}, y^{k,N}). \tag{7.5}$$

Therefore,

$$Y_k^N - \tilde{Y}_k^N = -\frac{4}{\ell} \left[I_1^N - \frac{\ell^2}{2} (S^{k,N} - 1) \right],$$

which implies

$$\begin{aligned} \mathbb{E}_k (Y_k^N - \tilde{Y}_k^N)^2 &\lesssim \mathbb{E}_k \left(I_1^N(x^{k,N}, y^{k,N}) - \ell^2 (S^{k,N} - 1)/2 \right)^2 \\ &\stackrel{(6.12)}{\lesssim} \frac{\|x^{k,N}\|_s^2}{N^2} + \frac{(S^{k,N})^2}{\sqrt{N}} + \frac{1}{N}. \end{aligned}$$

As for the second addend in (7.4), Lemma 6.3 gives

$$\begin{aligned} \left| \tilde{Y}_k^N \right|^2 \mathbb{E}_k \left| \alpha^N(x^{k,N}, y^{k,N}) - \alpha_\ell(S^{k,N}) \right|^2 &\lesssim (1 + (S^{k,N})^2) \left(\frac{1 + (S^{k,N})^2 + \|x^{k,N}\|_s^2}{\sqrt{N}} \right) \\ &\lesssim \frac{1 + (S^{k,N})^4 + \|x^{k,N}\|_s^4}{\sqrt{N}}. \end{aligned}$$

Combining the above two bounds and (7.4) gives the desired result. □

Proof of Lemma 7.4 By Jensen’s inequality,

$$\left(\sup_{t \in [0, T]} \left| \int_0^t b_\ell(S_v^{(N)}) - b_\ell(\bar{S}_v^{(N)}) dv \right| \right)^2 \lesssim \int_0^T \left| b_\ell(S_v^{(N)}) - b_\ell(\bar{S}_v^{(N)}) \right|^2 dv.$$

Since b_ℓ is globally Lipschitz,

$$\begin{aligned} \int_0^T \left| b_\ell(\bar{S}^N(v)) - b_\ell(S^N(v)) \right|^2 dv &\lesssim \int_0^T \left| \bar{S}^N(v) - S^N(v) \right|^2 dv \\ &= \sum_{k=0}^{[T\sqrt{N}]-1} \int_{t_k}^{t_{k+1}} \left| \bar{S}^N(v) - S^N(v) \right|^2 dv \\ &\quad + \int_{[T\sqrt{N}]}^T \left| \bar{S}^N(v) - S^N(v) \right|^2 dv \\ &\lesssim \frac{1}{\sqrt{N}} \sum_{k=0}^{[T\sqrt{N}]-1} (S^{k+1, N} - S^{k, N})^2. \end{aligned}$$

From (4.18) and (4.6),

$$\begin{aligned} \left| S^{k+1, N} - S^{k, N} \right| &\lesssim \frac{1}{N} \left(\|y^{k, N}\|_{\mathcal{C}^N}^2 - \|x^{k, N}\|_{\mathcal{C}^N}^2 \right) \\ &\stackrel{(7.5)}{\lesssim} \frac{1}{\sqrt{N}} I_1^N(x^{k, N}, y^{k, N}) \\ &= \frac{1}{\sqrt{N}} \left(I_1^N(x^{k, N}, y^{k, N}) - \frac{\ell^2(S^{k, N} - 1)}{2} \right) + \frac{1}{\sqrt{N}} \frac{\ell^2(S^{k, N} - 1)}{2}. \end{aligned}$$

Combining the above with (6.12) we obtain

$$\mathbb{E}_k(S^{k+1, N} - S^{k, N})^2 \lesssim \frac{1 + (S^{k, N})^2 + \|x^{k, N}\|_S^2}{N}. \tag{7.6}$$

Taking expectations and applying Lemma 6.2 concludes the proof. □

7.2 Analysis of noise

Proof of Lemma 7.2 Notice that we can write w^N as the linear interpolation

$$w^N(t) = (N^{1/2}t - k)M^{k, N} + (k + 1 - N^{1/2}t)M^{k-1, N} \quad \forall t_k \leq t < t_{k+1},$$

of the array

$$M^{k,N} := \frac{1}{N^{1/4}} \sum_{j=0}^{k-1} D^{j,N}, \quad \forall k = 1, \dots, [T\sqrt{N}] + 1.$$

It follows from the definition of $D^{k,N}$ in (4.17) and Lemma 6.2 that $\{M^{k,N}\}_{k \geq 1}$ is a discrete-time \mathbb{P}_{x^0} -martingale with respect to the filtration generated by $\{x^{k,N}\}_{k \geq 1}$. Since,

$$\sup_{t \in [0, T]} |w^N(t)| = \sup_{k \in \{1, \dots, [T\sqrt{N}] + 1\}} |M^{k,N}|,$$

Doob’s L^p inequality implies that

$$\begin{aligned} \mathbb{E}_{x^0} \left(\sup_{t \in [0, T]} |w^N(t)| \right)^2 &\lesssim \mathbb{E}_{x^0} \left(\sup_{k \in \{1, \dots, [T\sqrt{N}] + 1\}} |M^{k,N}|^2 \right) \\ &= \frac{1}{\sqrt{N}} \sum_{k=0}^{[T\sqrt{N}]} \mathbb{E}_{x^0} |D^{k,N}|^2, \end{aligned}$$

where the equality follows from the independence of the increments of $\{M^{k,N}\}_{k \geq 1}$. From the definition of $D^{k,N}$, Eq. (4.17), we have that

$$\begin{aligned} \frac{\mathbb{E}_{x^0} |D^{k,N}|^2}{\sqrt{N}} &= \mathbb{E}_{x^0} \left[S^{k+1,N} - S^{k,N} - \mathbb{E}_k \left(S^{k+1,N} - S^{k,N} \right) \right]^2 \\ &\lesssim \mathbb{E}_{x^0} |S^{k+1,N} - S^{k,N}|^2 \lesssim \frac{1}{N}, \end{aligned}$$

where the last inequality is a consequence of (7.6) and Lemma 6.2. The result follows immediately. □

8 Proof of Theorem 4.2

The idea behind the proof is the same as in the previous Sect. 7. First we introduce the piecewise constant interpolant of the chain $\{x^{k,N}\}_{k \in \mathbb{N}}$

$$\bar{x}^{(N)}(t) = x^{k,N} \quad \text{for } t_k \leq t < t_{k+1}. \tag{8.1}$$

Due to the continuity of the map \mathcal{J}_1 (Theorem 3.3), all we need to prove is the weak convergence of $\hat{\eta}^N(t)$ to zero (see Sect. 5.2). Looking at the definition of $\hat{\eta}^N(t)$, Eq. (5.6), this follows from Lemmas 8.1, 8.2 and 8.3 below. We prove Lemmas 8.1 and 8.2 in Sect. 8.1 and Lemma 8.3 in Sect. 8.2.

Lemma 8.1 *Let Assumption 2.1 hold and recall the definition (5.8) of the process $d^N(t)$; then*

$$\lim_{N \rightarrow \infty} \mathbb{E}_{x^0} \left(\sup_{t \in [0, T]} \left| d^N(t) \right| \right)^2 = 0.$$

Lemma 8.2 *If Assumption 2.1 holds, then v^N (defined in (5.9)) converges in probability in $C([0, T]; \mathcal{H}^s)$ to zero.*

Lemma 8.3 *Let Assumption 2.1 hold. Then the interpolated martingale difference array $\eta^N(t)$ defined in (5.7) converges weakly in $C([0, T]; \mathcal{H}^s)$ to the stochastic integral $\eta(t)$, defined in Eq. (5.11).*

8.1 Analysis of drift

Proof (Lemma 8.1) For all $t \in [t_k, t_{k+1})$, we can write

$$(t - t_k)\Theta(x^{k,N}, S^{k,N}) + \frac{1}{\sqrt{N}} \sum_{j=0}^{k-1} \Theta(x^{j,N}, S^{j,N}) = \int_0^t \Theta(\bar{x}^{(N)}(v), \bar{S}^{(N)}(v)) dv.$$

Therefore, we can decompose $d^N(t)$ as

$$d^N(t) = d_1^N(t) + d_2^N(t),$$

where

$$d_1^N(t) := (t - t_k) \left[\Theta^{k,N} - \Theta(x^{k,N}, S^{k,N}) \right] + \frac{1}{\sqrt{N}} \sum_{j=0}^{k-1} \left[\Theta^{j,N} - \Theta(x^{j,N}, S^{j,N}) \right]$$

and

$$d_2^N(t) := \int_0^t \left[\Theta(\bar{x}^N(v), \bar{S}^N(v)) - \Theta(x^{(N)}(v), S^{(N)}(v)) \right] dv.$$

The statement is now a consequence of Lemmas 8.4 and 8.5. □

Lemma 8.4 *If Assumption 2.1 holds, then*

$$\lim_{N \rightarrow \infty} \mathbb{E}_{x^0} \left(\sup_{t \in [0, T]} \left\| d_1^N(t) \right\|_s \right)^2 = 0.$$

Lemma 8.5 *If Assumption 2.1 holds, then*

$$\lim_{N \rightarrow \infty} \mathbb{E}_{x^0} \left(\sup_{t \in [0, T]} \left\| d_2^N(t) \right\|_s \right)^2 = 0.$$

Before proving Lemma 8.4, we state and prove the following Lemma 8.6. We then consecutively prove Lemmas 8.4, 8.5 and 8.2. Recall the definitions of Θ and $\Theta^{k, N}$, equations (4.23) and (4.21), respectively.

Lemma 8.6 *Let Assumption 2.1 hold and set*

$$p^{k, N} := \Theta^{k, N} - \Theta(x^{k, N}, S^{k, N}). \tag{8.2}$$

Then

$$\mathbb{E}_{x^0} \left\| p^{k, N} \right\|_s^2 \lesssim \sum_{j=N+1}^{\infty} (\lambda_j j^s)^4 + \frac{1}{\sqrt{N}}.$$

Proof of Lemma 8.6 Recalling (4.26) and (6.24), we have

$$\left\| p^{k, N} \right\|_s^2 \lesssim \sqrt{N} \left\| \mathbb{E}_k \varepsilon_k^N(x^{k, N}) \right\|_s^2 \tag{8.3}$$

$$+ \left\| \alpha_\ell(S^{k, N}) F(x^{k, N}) - \left[\mathbb{E}_k \alpha^N(x^{k, N}, y^{k, N}) \right] (x^{k, N} + C_N \nabla \Psi^N(x^{k, N})) \right\|_s^2, \tag{8.4}$$

where the function F that appears in the above has been defined in Lemma 2.1. The term on the RHS of (8.3) has been studied in Lemma 6.5. To estimate the addend in (8.4) we use (2.25), the boundedness of α_ℓ and Lemma 6.3. A straightforward calculation then gives

$$\begin{aligned} (8.4) &\lesssim \left[\alpha_\ell(S^{k, N}) - \mathbb{E}_k \alpha^N(x^{k, N}, y^{k, N}) \right]^2 \left\| x^{k, N} + C_N \nabla \Psi^N(x^{k, N}) \right\|_s^2 \\ &\quad + \left\| \alpha_\ell(S^{k, N}) \left[F(x^{k, N}) - (x^{k, N} + C_N \nabla \Psi^N(x^{k, N})) \right] \right\|_s^2 \\ &\lesssim \frac{1 + (S^{k, N})^4 + \left\| x^{k, N} \right\|_s^4}{\sqrt{N}} + \left\| C \nabla \Psi(x^{k, N}) - C_N \nabla \Psi^N(x^{k, N}) \right\|_s^2. \end{aligned}$$

From the definition of Ψ^N and $\nabla \Psi^N$, Eqs. (1.5) and (2.23), respectively,

$$\begin{aligned} \left\| C \nabla \Psi(x^{k, N}) - C_N \nabla \Psi^N(x^{k, N}) \right\|_s^2 &= \left\| C \nabla \Psi(x^{k, N}) - C_N \mathcal{P}^N(\nabla \Psi(x^{k, N})) \right\|_s^2 \\ &= \sum_{j=N+1}^{\infty} (\lambda_j j^s)^4 \mathbb{E} \left[j^{-2s} (\nabla \Psi(x^{k, N}))_j^2 \right] \end{aligned}$$

$$\lesssim \sum_{j=N+1}^{\infty} (\lambda_j j^s)^4,$$

having used (2.24) in the last inequality. The statement is now a consequence of Lemma 6.2. \square

Proof of Lemma 8.4 Following the analogous steps to those taken in the proof of Lemma 7.3, the proof is a direct consequence of Lemma 8.6, after observing that the summation $\sum_{j=N+1}^{\infty} (\lambda_j j^s)^4$ is the tail of a convergent series hence it tends to zero as $N \rightarrow \infty$. \square

Proof of Lemma 8.5 By the definition of Θ , Eq. (4.23), we have

$$\begin{aligned} & \left\| \Theta(\bar{x}^N(t), \bar{S}^N(t)) - \Theta(x^N(t), S^N(t)) \right\|_s \\ &= \left\| F(\bar{x}^N(t))h_\ell(\bar{S}^N(t)) - F(x^N(t))h_\ell(S^N(t)) \right\|_s. \end{aligned}$$

Applying (2.20) and (2.25) and using the fact h_ℓ is globally Lipschitz and bounded, we get

$$\begin{aligned} & \left\| \Theta(\bar{x}^N(t), \bar{S}^N(t)) - \Theta(x^N(t), S^N(t)) \right\|_s \\ & \lesssim \left\| \bar{x}^N(t) - x^N(t) \right\|_s + (1 + \left\| \bar{x}^N(t) \right\|_s) \left| \bar{S}^N(t) - S^N(t) \right|. \end{aligned}$$

Thus, from the definitions (1.16), (7.1), (1.9) and (8.1), if $t_k \leq t < t_{k+1}$, we have

$$\begin{aligned} & \left\| \Theta(\bar{x}^N(t), \bar{S}^N(t)) - \Theta(x^N(t), S^N(t)) \right\|_s \\ & \lesssim (t - k\sqrt{N}) \left\| x^{k+1,N} - x^{k,N} \right\|_s \\ & \quad + (t - k\sqrt{N}) \left(1 + \left\| x^{k,N} \right\|_s \right) \left| S^{k+1,N} - S^{k,N} \right|. \end{aligned}$$

Applying (6.3) and (7.6) one then concludes

$$\begin{aligned} & \mathbb{E}_k \left\| \Theta(\bar{x}^N(t), \bar{S}^N(t)) - \Theta(x^N(t), S^N(t)) \right\|_s^2 \\ & \lesssim (t - k\sqrt{N})^2 \left(\frac{1 + \left\| x^{k,N} \right\|_s^2}{\sqrt{N}} + \frac{\left\| x^{k,N} \right\|_s^4 + (S^{k,N})^4}{N} \right) \end{aligned}$$

The remainder of the proof is analogous to the proof of Lemma 7.4. \square

Proof of Lemma 8.2 For any arbitrary but fixed $\varepsilon > 0$, we need to argue that

$$\lim_{N \rightarrow \infty} \mathbb{P} \left[\sup_{t \in [0, T]} \left\| v^N(t) \right\|_s \geq \varepsilon \right] = 0.$$

From the definition of v^N we have

$$\sup_{t \in [0, T]} \left\| v^N(t) \right\|_s \leq \int_0^T \left\| F(x^{(N)}(v)) \right\|_s \left| S^{(N)}(v) - S(v) \right| dv.$$

Using (2.21) and the fact that $\|x^{(N)}(t)\|_s \leq \|x^{k, N}\|_s + \|x^{k+1, N}\|_s$ (which is a simple consequence of (1.9)), for any $t \in [t_k, t_{k+1})$

$$\begin{aligned} \sup_{t \in [0, T]} \left\| v^N(t) \right\|_s &\leq \left(\sup_{t \in [0, T]} \left| S^{(N)}(t) - S(t) \right| \right) \int_0^T \left\| F(x^{(N)}(v)) \right\|_s dv \\ &\lesssim \underbrace{\left(\sup_{t \in [0, T]} \left| S^{(N)}(t) - S(t) \right| \right)}_{=: a^N} \underbrace{\left(1 + \frac{1}{\sqrt{N}} \sum_{j=0}^{[T\sqrt{N}]-1} \left\| x^{j, N} \right\|_s \right)}_{=: u^N}. \end{aligned}$$

Using Markov’s inequality and Lemma 6.2, given any $\delta > 0$, it is straightforward to find constant M such that $\mathbb{P}[u^N > M] \leq \delta$ for every $N \in \mathbb{N}$. Thus

$$\begin{aligned} \mathbb{P} \left[\sup_{t \in [0, T]} \left\| v^N(t) \right\|_s \geq \varepsilon \right] &\leq \mathbb{P} \left[a^N u^N \geq \varepsilon \right] \\ &= \mathbb{P} \left[a^N u^N \geq \varepsilon, u^N \leq M \right] + \mathbb{P} \left[a^N u^N \geq \varepsilon, u^N > M \right] \\ &\leq \mathbb{P} \left[a^N \geq \varepsilon / M \right] + \mathbb{P} \left[u^N > M \right] \leq \mathbb{P} \left[a^N \geq \varepsilon / M \right] + \delta. \end{aligned}$$

Given that the δ was arbitrary, the result then follows from the fact that $S^{(N)}$ converges in probability to S (Theorem 4.1). □

8.2 Analysis of noise

The proof of Lemma 8.3 is based on [14, Lemma 8.9]. For the reader’s convenience, we restate [14, Lemma 8.9] below as Lemma 8.7. In order to state such a lemma let us introduce the following notation and definitions. Let $k_N : [0, T] \rightarrow \mathbb{Z}_+$ be a sequence of nondecreasing, right continuous functions indexed by N , with $k_N(0) = 0$ and $k_N(T) \geq 1$. Let \mathcal{H} be any Hilbert space and $\{X^{k, N}, \mathcal{F}^{k, N}\}_{0 \leq k \leq k_N(T)}$ be a \mathcal{H} -valued martingale difference array (MDA), i.e. a double sequence of random variables such that $\mathbb{E}[X^{k, N} | \mathcal{F}_{k-1}^N] = 0$, $\mathbb{E}[\|X^{k, N}\|^2 | \mathcal{F}_{k-1}^N] < \infty$ almost surely and sigma-algebras $\mathcal{F}^{k-1, N} \subseteq \mathcal{F}^{k, N}$. Consider the process $\mathcal{X}^N(t)$ defined by

$$\mathcal{X}^N(t) := \sum_{k=1}^{k_N(t)} X^{k, N},$$

if $k_N(t) \geq 1$ and $k_N(t) > \lim_{v \rightarrow 0+} k_N(t - v)$ and by linear interpolation otherwise. With this set up we recall the following result.

Lemma 8.7 (Lemma 8.9 [14]) Let $D : \mathcal{H} \rightarrow \mathcal{H}$ be a self-adjoint positive definite trace class operator on $(\mathcal{H}, \|\cdot\|)$. Suppose the following limits hold in probability

(i) there exists a continuous and positive function $f : [0, T] \rightarrow \mathbb{R}_+$ such that

$$\lim_{N \rightarrow \infty} \sum_{k=1}^{k_N(T)} \mathbb{E} \left(\left\| X^{k,N} \right\|^2 \middle| \mathcal{F}_{k-1}^N \right) = \text{Trace}_{\mathcal{H}}(D) \int_0^T f(t) dt ;$$

(ii) if $\{\phi_j\}_{j \in \mathbb{N}}$ is an orthonormal basis of \mathcal{H} then

$$\lim_{N \rightarrow \infty} \sum_{k=1}^{k_N(T)} \mathbb{E} \left(\langle X^{k,N}, \phi_j \rangle \langle X^{k,N}, \phi_i \rangle \middle| \mathcal{F}_{k-1}^N \right) = 0 \quad \text{for all } i \neq j ;$$

(iii) for every fixed $\epsilon > 0$,

$$\lim_{N \rightarrow \infty} \sum_{k=1}^{k_N(T)} \mathbb{E} \left(\left\| X^{k,N} \right\|^2 \mathbf{1}_{\left\{ \left\| X^{k,N} \right\|^2 \geq \epsilon \right\}} \middle| \mathcal{F}_{k-1}^N \right) = 0, \quad \text{in probability,}$$

where $\mathbf{1}_A$ denotes the indicator function of the set A . Then the sequence \mathcal{X}^N converges weakly in $C([0, T]; \mathcal{H}^s)$ to the stochastic integral $t \mapsto \int_0^t \sqrt{f(v)} dW_v$, where W_t is a \mathcal{H} -valued D -Brownian motion.

Proof of Lemma 8.3 We apply Lemma 8.7 in the Hilbert space \mathcal{H}^s , with $k_N(t) = [t\sqrt{N}]$, $X^{k,N} = L^{k,N} / N^{1/4}$ [$L^{k,N}$ is defined in (4.22)] and \mathcal{F}_k^N the sigma-algebra generated by $\{\gamma^{h,N}, \xi^{h,N}, 0 \leq h \leq k\}$ to study the sequence $\eta^N(t)$, defined in (5.7). We now check that the three conditions of Lemma 8.7 hold in the present case.

(i) Note that by the definition of $L^{k,N}$, $\mathbb{E}[L^{k,N} | \mathcal{F}_{k-1}^N] = \mathbb{E}_k[L^{k,N}]$ almost surely. We need to show that the limit

$$\lim_{N \rightarrow \infty} \frac{1}{\sqrt{N}} \sum_{k=0}^{[T\sqrt{N}]} \mathbb{E}_k \left\| L^{k,N} \right\|_s^2 = 2 \text{Trace}_{\mathcal{H}^s}(C_s) \int_0^T h_\ell(S(u)) du, \quad (8.5)$$

holds in probability. By (4.28),

$$\frac{1}{\sqrt{N}} \mathbb{E}_k \left\| L^{k,N} \right\|_s^2 = \mathbb{E}_k \left\| x^{k+1,N} - x^{k,N} \right\|_s^2 - \left\| \mathbb{E}_k \left(x^{k+1,N} - x^{k,N} \right) \right\|_s^2.$$

From the above, if we prove

$$\mathbb{E}_{x^0} \sum_{k=0}^{[T\sqrt{N}]} \left\| \mathbb{E}_k \left(x^{k+1,N} - x^{k,N} \right) \right\|_s^2 \rightarrow 0 \quad \text{as } N \rightarrow \infty, \quad (8.6)$$

and that

$$\begin{aligned} & \lim_{N \rightarrow \infty} \sum_{k=0}^{[T\sqrt{N}]} \mathbb{E}_k \left\| x^{k+1,N} - x^{k,N} \right\|_s^2 \\ &= 2 \operatorname{Trace}_{\mathcal{H}^s} (C_s) \int_0^T h_\ell(S(u)) du, \quad \text{in probability,} \end{aligned} \tag{8.7}$$

then (8.5) follows. We start by proving (8.6):

$$\begin{aligned} \left\| \mathbb{E}_k \left(x^{k+1,N} - x^{k,N} \right) \right\|_s^2 &\stackrel{(2.14)}{\lesssim} \frac{1}{N} \left\| x^{k,N} + C_N \nabla \Psi^N(x^{k,N}) \right\|_s^2 \\ &\quad + \frac{1}{\sqrt{N}} \left\| \mathbb{E}_k \left(\gamma^{k,N} C_N^{1/2} \xi^{k,N} \right) \right\|_s^2 \\ &\lesssim \frac{1}{N} \left(1 + \left\| x^{k,N} \right\|_s^2 \right), \end{aligned}$$

where the last inequality follows from (2.25) and (6.25). The above and (6.7) prove (8.6). We now come to (8.7):

$$\begin{aligned} & \left| \sum_{k=0}^{[T\sqrt{N}]} \mathbb{E}_k \left\| x^{k+1,N} - x^{k,N} \right\|_s^2 - 2 \operatorname{Trace}_{\mathcal{H}^s} (C_s) \int_0^T h_\ell(S(u)) du \right| \\ &\stackrel{(2.14)}{\lesssim} \frac{1}{N} \sum_{k=0}^{[T\sqrt{N}]} \mathbb{E}_k \left\| x^{k,N} + C_N \nabla \Psi^N(x^{k,N}) \right\|_s^2 \\ &\quad + \frac{1}{N^{3/4}} \sum_{k=0}^{[T\sqrt{N}]} \mathbb{E}_k \left| \langle x^{k,N} + C_N \nabla \Psi^N(x^{k,N}), C_N^{1/2} \xi^{k,N} \rangle_s \right| \\ &\quad + \left| \frac{2\ell}{\sqrt{N}} \sum_{k=0}^{[T\sqrt{N}]} \mathbb{E}_k \left\| \gamma^{k,N} C_N^{1/2} \xi^{k,N} \right\|_s^2 - 2 \operatorname{Trace}_{\mathcal{H}^s} (C_s) \int_0^T h_\ell(S(u)) du \right|. \end{aligned}$$

The first two addends tend to zero in L_1 as N tends to infinity due to (2.25), (2.27) and Lemma 6.2. As for the third addend, we decompose it as follows

$$\begin{aligned} & \left| \frac{2\ell}{\sqrt{N}} \sum_{k=0}^{[T\sqrt{N}]} \mathbb{E}_k \left\| \gamma^{k,N} C_N^{1/2} \xi^{k,N} \right\|_s^2 - 2 \operatorname{Trace}_{\mathcal{H}^s} (C_s) \int_0^T h_\ell(S(u)) du \right| \\ &\stackrel{(1.13), (6.24)}{\lesssim} \left| \frac{\ell}{\sqrt{N}} \sum_{k=0}^{[T\sqrt{N}]} \mathbb{E}_k \left\| \varepsilon^{k,N} \right\|_s^2 - \frac{\ell}{\sqrt{N}} \sum_{k=0}^{[T\sqrt{N}]} \operatorname{Trace}_{\mathcal{H}^s} (C_s) \alpha_\ell(S^k) \right| \end{aligned}$$

$$+ \left| \frac{1}{\sqrt{N}} \sum_{k=0}^{[T\sqrt{N}]} \text{Trace}_{\mathcal{H}^s}(\mathcal{C}_s) h_\ell(S^{k,N}) - \text{Trace}_{\mathcal{H}^s}(\mathcal{C}_s) \int_0^T h_\ell(S(u)) du \right|. \tag{8.8}$$

Convergence to zero in L^1 of the first term in the above follows from Lemmas 6.2 and 6.6. As for the term in (8.8), we use the identity

$$\int_0^T h_\ell(\bar{S}^{(N)}(u)) du = \left(T - \frac{[T\sqrt{N}]}{\sqrt{N}} \right) h_\ell(S^{[T\sqrt{N}],N}) + \frac{1}{\sqrt{N}} \sum_{k=0}^{[T\sqrt{N}]} h_\ell(S^{k,N}),$$

to further split it, obtaining:

$$(8.8) \lesssim \left| \int_0^T h_\ell(\bar{S}^{(N)}(u)) - h_\ell(S^{(N)}(u)) du \right| \tag{8.9}$$

$$+ \left| \int_0^T h_\ell(S^{(N)}(u)) - h_\ell(S(u)) du \right| \tag{8.10}$$

$$+ \left(T - \frac{[T\sqrt{N}]}{\sqrt{N}} \right) h_\ell(S^{[T\sqrt{N}],N}). \tag{8.11}$$

Convergence (in L_1) of (8.9) to zero follows with the same calculations leading to (7.6), the global Lipschitz property of h_ℓ , and Lemma 6.2. The addend in (8.10) tends to zero in probability since $S^{(N)}$ tends to S in probability in $C([0, T]; \mathbb{R})$ (Theorem 4.1) and the third addend is clearly small. The limit (8.7) then follows.

- (ii) Condition (ii) of Lemma 8.7 can be shown to hold with similar calculations, so we will not show the details.
- (iii) Using (6.3), the last bound follows a calculation completely analogous to the one in [14, Section 8.2]. We omit the details here. □

Acknowledgements A.M. Stuart acknowledges support from AMS, DARPA, EPSRC, ONR. J. Kuntz gratefully acknowledges support from the BBSRC in the form of the Ph.D. studentship BB/F017510/1. M. Ottobre and J. Kuntz gratefully acknowledge financial support from the Edinburgh Mathematical Society.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

A Appendix: Proofs of the results in Sect. 2

Proof of Lemma 2.1 The bounds (2.20) are a consequence of (2.19). We show how to obtain the second bound in (2.20):

$$\begin{aligned} \|\mathcal{C}\nabla\Psi(x) - \mathcal{C}\nabla\Psi(y)\|_s^2 &= \sum_{j=1}^{\infty} \lambda_j^4 j^{2s} [(\nabla\Psi(x) - \nabla\Psi(y))_j]^2 \\ &= \sum_{j=1}^{\infty} (\lambda_j j^s)^4 j^{-2s} [(\nabla\Psi(x) - \nabla\Psi(y))_j]^2 \\ &\lesssim \|\nabla\Psi(x) - \nabla\Psi(y)\|_{-s}^2 \stackrel{(2.19)}{\lesssim} \|x - y\|_s^2, \end{aligned}$$

where in the above we have used (2.17) and $(\nabla\Psi(x) - \nabla\Psi(y))_j$ denotes the j th component of the vector $\nabla\Psi(x) - \nabla\Psi(y)$. With analogous calculations one can obtain the first bound in (2.20). As for the second equation in (2.21):

$$\|F(z)\|_s \lesssim \|z\|_s + \|\mathcal{C}\nabla\Psi(z)\|_s \stackrel{(2.20)}{\lesssim} 1 + \|z\|_s.$$

Similarly for the first bound in (2.21). The proof of Eq. (2.22) is standard, so we only sketch it: consider a line joining points x and y , $\gamma(t) = x + t(y - x)$, $t \in [0, 1]$. Then

$$\begin{aligned} \Psi(\gamma(1)) - \Psi(\gamma(0)) &= \Psi(y) - \Psi(x) \\ &= \int_0^1 dt \langle \nabla\Psi(\gamma(t)), y - x \rangle \lesssim \|y - x\|_s, \end{aligned}$$

having used (2.19) and (2.6) in the last inequality. An analogous calculation to the above can be done for Ψ^N , after proving (2.24) below. \square

Proof of Lemma 2.2 The bounds (2.24) and (2.25) are just consequences of the definition of Ψ^N and $\nabla\Psi^N$ and the analogous properties of Ψ . For the sake of clarity we just spell out how to obtain (2.25):

$$\begin{aligned} \left\| \mathcal{C}_N \nabla\Psi^N(x) \right\|_s^2 &\stackrel{(2.23)}{=} \left\| \mathcal{C}_N \mathcal{P}^N \nabla\Psi(\mathcal{P}^N(x)) \right\|_s^2 = \sum_{j=1}^N j^{2s} \lambda_j^4 \left[\nabla\Psi(\mathcal{P}^N(x)) \right]_j^2 \\ &\leq \sum_{j=1}^{\infty} j^{2s} \lambda_j^4 \left[\nabla\Psi(\mathcal{P}^N(x)) \right]_j^2 \leq \left\| \mathcal{C}\nabla\Psi(\mathcal{P}^N(x)) \right\|_s^2 \stackrel{(2.20)}{\lesssim} 1. \end{aligned}$$

As for (2.26), using (2.17):

$$\begin{aligned} \|\mathcal{C}_N \nabla\Psi^N(x)\|_{\mathcal{C}_N}^2 &= \sum_{j=1}^N \lambda_j^2 \left[(\nabla\Psi^N(x))_j \right]^2 \lesssim \sum_{j=1}^{\infty} j^{-2s} \left[(\nabla\Psi^N(x))_j \right]^2 \\ &= \|\nabla\Psi^N(x)\|_{-s}^2 \lesssim 1. \end{aligned}$$

\square

B Appendix: Proofs of Lemmas 6.5 and 6.6

To prove Lemmas 6.5 and 6.6 we decompose $Q^N(x^{k,N}, \xi^{k,N})$ into the sum of a term that depends on $\xi_j^{k,N}$ (the j th component of $\xi^{k,N}$), Q_j^N and one that is independent of $\xi_j, Q_{j,\perp}^N$:

$$Q^N = Q_j^N + Q_{j,\perp}^N,$$

where

$$Q_j^N := \left(\frac{\ell^{5/2}}{\sqrt{2}N^{5/4}} - \frac{\ell^{3/2}}{\sqrt{2}N^{3/4}} \right) \frac{x_j^{k,N} \xi_j^{k,N}}{\lambda_j} + \frac{\ell^{5/2}}{\sqrt{2}N^{5/4}} \lambda_j \xi_j^{k,N} (\nabla \Psi^N(x^{k,N}))_j - \frac{\ell^2}{2N} (\xi_j^{k,N})^2 + I_2^N(x^{k,N}, y^{k,N}) + I_3^N(x^{k,N}, y^{k,N}). \tag{B.1}$$

We recall that I_2^N and I_3^N have been defined in Sect. 6. Therefore, using (6.8),

$$Q_{j,\perp}^N = Q^N - Q_j^N = I_1^N + \tilde{Q}_j^N, \tag{B.2}$$

having set

$$\tilde{Q}_j^N := - \left(\frac{\ell^{5/2}}{\sqrt{2}N^{5/4}} - \frac{\ell^{3/2}}{\sqrt{2}N^{3/4}} \right) \frac{x_j^{k,N} \xi_j^{k,N}}{\lambda_j} - \frac{\ell^{5/2}}{\sqrt{2}N^{5/4}} \lambda_j \xi_j^{k,N} (\nabla \Psi^N(x^{k,N}))_j + \frac{\ell^2}{2N} (\xi_j^{k,N})^2. \tag{B.3}$$

Proof of Lemma 6.5 (6.26) is a consequence of the definition (6.24) and the estimate (6.25). Thus, all we have to do is establish the latter. Recalling that $\{\hat{\phi}_j\}_{j \in \mathbb{N}} := \{j^{-s} \phi_j\}_{j \in \mathbb{N}}$ is an orthonormal basis for \mathcal{H}^s , we act as in the proof of [17, Lemma 4.7] and obtain

$$\left| \left\langle \mathbb{E}_k \varepsilon^{k,N}, \hat{\phi}_j \right\rangle_s \right|^2 \lesssim j^{2s} \lambda_j^2 \mathbb{E}_k \left[Q_j^N(x^{k,N}, \xi^{k,N}) \right]^2$$

where Q_j^N has been defined in (B.1). Thus

$$\begin{aligned} \left| \left\langle \mathbb{E}_k \varepsilon^{k,N}, \hat{\phi}_j \right\rangle_s \right|^2 &\lesssim j^{2s} \lambda_j^2 \left(N^{-3/2} (x_j^{k,N})^2 \mathbb{E}_k \xi_j^2 \lambda_j^{-2} + N^{-5/2} \lambda_j^2 \mathbb{E}_k \left[\xi_j^2 (\nabla \Psi^N(x^{k,N}))_j^2 \right] \right) \\ &\quad + j^{2s} \lambda_j^2 \mathbb{E}_k \left(\left| I_2^N \right|^2 + \left| I_3^N \right|^2 \right) + j^{2s} \lambda_j^2 N^{-2} \\ &\lesssim N^{-3/2} \mathbb{E}_k (j^s x_j^{k,N})^2 + N^{-5/2} j^{-2s} (\nabla \Psi^N(x^{k,N}))_j^2 \\ &\quad + j^{2s} \lambda_j^2 N^{-2} + j^{2s} \lambda_j^2 \frac{1 + \|x^{k,N}\|_s^2}{\sqrt{N}}, \end{aligned}$$

where the second inequality follows from the boundedness of the sequence $\{\lambda_j\}$, (6.13) and (6.14). Summing over j and applying (2.24) we obtain (6.25). \square

Proof of Lemma 6.6 By definition of $\varepsilon^{k,N}$, and because $\gamma^{k,N} = [\gamma^{k,N}]^2$ (as $\gamma^{k,N}$ can only take values 0 or 1)

$$\begin{aligned} \mathbb{E}_k \left\| \varepsilon^{k,N} \right\|_s^2 &= \sum_{j=1}^N j^{2s} \lambda_j^2 \mathbb{E}_k \left[\gamma^{k,N} \left| \xi_j^{k,N} \right|^2 \right] \\ &= \sum_{j=1}^N j^{2s} \lambda_j^2 \mathbb{E}_k \left[\left(1 \wedge e^{Q^N(x^{k,N}, y^{k,N})} \right) \left| \xi_j^{k,N} \right|^2 \right]. \end{aligned}$$

Using the above, the Lipschitzianity of the function $s \mapsto 1 \wedge e^s$, (B.2) and the independence of $Q_{j,\perp}^N$ and $\xi_j^{k,N}$, we write

$$\begin{aligned} &\left| \mathbb{E}_k \left\| \varepsilon^{k,N} \right\|_s^2 - \text{Trace}(C_s) \alpha_\ell(S^{k,N}) \right| \\ &= \left| \mathbb{E}_k \sum_{j=1}^N j^{2s} \lambda_j^2 \left(1 \wedge e^{Q^N} \right) \left| \xi_j \right|^2 - \text{Trace}(C_s) \alpha_\ell(S^{k,N}) \right| \\ &\leq \left| \mathbb{E}_k \sum_{j=1}^N j^{2s} \lambda_j^2 \left(1 \wedge e^{Q_{j,\perp}^N} \right) \left| \xi_j \right|^2 - \text{Trace}(C_s) \alpha_\ell(S^{k,N}) \right| \\ &\quad + \left| \mathbb{E}_k \sum_{j=1}^N j^{2s} \lambda_j^2 \left[\left(1 \wedge e^{Q^N} \right) - \left(1 \wedge e^{Q_{j,\perp}^N} \right) \right] \left| \xi_j \right|^2 \right| \\ &\lesssim \left| \sum_{j=1}^N j^{2s} \lambda_j^2 \mathbb{E}_k \left(1 \wedge e^{Q_{j,\perp}^N} \right) - \text{Trace}(C_s) \alpha_\ell(S^{k,N}) \right| \tag{B.4} \end{aligned}$$

$$+ \left| \mathbb{E}_k \sum_{j=1}^N j^{2s} \lambda_j^2 \left| Q_j^N \right| \left| \xi_j \right|^2 \right| \tag{B.5}$$

We now proceed to bound the addends in (B.4) and (B.5), starting with the latter. Using (B.1) and (B.3), we write

$$\begin{aligned} \mathbb{E}_k \sum_{j=1}^N j^{2s} \lambda_j^2 \left| Q_j^N \right| \left| \xi_j \right|^2 &\leq \mathbb{E}_k \sum_{j=1}^N j^{2s} \lambda_j^2 \left| I_2^N \right| \left| \xi_j \right|^2 + \mathbb{E}_k \sum_{j=1}^N j^{2s} \lambda_j^2 \left| I_3^N \right| \left| \xi_j \right|^2 \\ &\quad + \mathbb{E}_k \sum_{j=1}^N j^{2s} \lambda_j^2 \left| \tilde{Q}_j^N \right| \left| \xi_j \right|^2 \end{aligned}$$

$$\begin{aligned}
 &\lesssim \sum_{j=1}^N j^{2s} \lambda_j^2 \sqrt{\mathbb{E}_k |I_2^N|^2} + \mathbb{E}_k \sum_{j=1}^N j^{2s} \lambda_j^2 \sqrt{\mathbb{E}_k |I_3^N|^2} \\
 &\quad + \sum_{j=1}^N j^{2s} \lambda_j^2 \mathbb{E}_k \left(|\tilde{Q}_j^N| |\xi_j|^2 \right) \\
 &\lesssim \frac{1 + \|x^{k,N}\|_s}{N^{1/4}} + \sum_{j=1}^N j^{2s} \lambda_j^2 \mathbb{E}_k \left(|\tilde{Q}_j^N| |\xi_j|^2 \right), \tag{B.6}
 \end{aligned}$$

where the last inequality follows from Lemma 6.4 and (2.16). As for the last addend, using (B.3):

$$\begin{aligned}
 \sum_{j=1}^N j^{2s} \lambda_j^2 \mathbb{E}_k \left[|\tilde{Q}_j^N| |\xi_j|^2 \right] &\lesssim \frac{1}{N^{3/4}} \sum_{j=1}^N j^{2s} \lambda_j |x_j^{k,N}| \mathbb{E}_k |\xi_j^{k,N}|^3 \\
 &\quad + \frac{1}{N^{5/4}} \sum_{j=1}^N j^{2s} \lambda_j^3 |(C_N \nabla \Psi^N(x^{k,N}))_j| \mathbb{E}_k |\xi_j^{k,N}|^3 \\
 &\quad + \frac{1}{N} \sum_{j=1}^N j^{2s} \lambda_j^2 \mathbb{E}_k |\xi_j^{k,N}|^4 \\
 &\lesssim \frac{1 + \|x^{k,N}\|_s^2}{N^{3/4}}, \tag{B.7}
 \end{aligned}$$

where the last inequality follows from (2.25), (2.16), the boundedness of the sequence $\{\lambda_j\}_{j \in \mathbb{N}}$ and by using Young’s inequality (more precisely, the so-called Young’s inequality “with ϵ ”), as follows:

$$\lambda_j |x_j^{k,N}| \mathbb{E}_k |\xi_j^{k,N}|^3 \leq |x_j^{k,N}|^2 + \lambda_j^2 \left(\mathbb{E}_k |\xi_j^{k,N}|^3 \right)^2.$$

This concludes the analysis of the term (B.5). As for the term (B.4), by definition of α_ℓ , Eq. (1.12),

$$\begin{aligned}
 1 \wedge e^{\mathcal{Q}_{j,\perp}^N} - \alpha_\ell(S^{k,N}) &= \left(1 \wedge e^{\mathcal{Q}_{j,\perp}^N} - 1 \wedge e^{I_1^N(x^{k,N}, y^{k,N})} \right) \\
 &\quad + \left(1 \wedge e^{I_1^N(x^{k,N}, y^{k,N})} - 1 \wedge e^{\ell^2(S^{k,N}-1)/2} \right).
 \end{aligned}$$

Exploiting the fact that $s \mapsto 1 \wedge e^s$ is globally Lipschitz, using Lemma 6.4 and manipulations of the same type as in (B.7), it follows that

$$\left| 1 \wedge e^{Q_{j,\perp}^N} - \alpha_\ell(S^{k,N}) \right| \lesssim \frac{1 + S^{k,N} + \|x^{k,N}\|_S^2}{\sqrt{N}}.$$

Putting (B.6)–(B.7) and the above together completes the proof. □

C Appendix: Uniform bounds on the moments of $S^{k,N}$ and $x^{k,N}$

Proof of Lemma 6.2 To prove both bounds, we use a strategy analogous to the one used in [18, Proof of Lemma 9]. Let $\{A_k : k \in \mathbb{N}\}$ be any sequence of real numbers. Suppose that there exists a constant $C \geq 0$ (independent of k) such that

$$A_{k+1} - A_k \leq \frac{C}{\sqrt{N}} (1 + A_k). \tag{C.1}$$

We start by showing that if the above holds then $A_k \leq e^{CT}(A_0 + CT)$, uniformly over $k = 0, \dots, [T\sqrt{N}]$. Indeed, from (C.1),

$$A_k \leq \left(1 + \frac{C}{\sqrt{N}}\right)^k A_0 + \frac{C}{\sqrt{N}} \sum_{j=0}^{k-1} \left(1 + \frac{C}{\sqrt{N}}\right)^j \leq \left(1 + \frac{C}{\sqrt{N}}\right)^k \left(A_0 + k \frac{C}{\sqrt{N}}\right).$$

Thus, for all $k = 0, \dots, [T\sqrt{N}]$,

$$A_k \leq \left(1 + \frac{C}{\sqrt{N}}\right)^{[T\sqrt{N}]} \left(A_0 + [T\sqrt{N}] \frac{C}{\sqrt{N}}\right) \leq \left(1 + \frac{C}{\sqrt{N}}\right)^{T\sqrt{N}} (A_0 + CT).$$

Since $1 + z \leq e^z$ for any $z \in \mathbb{R}$,

$$\left(1 + \frac{C}{\sqrt{N}}\right)^{\sqrt{N}} \leq \left(e^{C/\sqrt{N}}\right)^{\sqrt{N}} = e^C.$$

With this preliminary observation, we can now prove (6.6)–(6.7).

(i) Proof of (6.6). To prove (6.6) we only need to show that (C.1) holds (for some constant $C > 0$ independent of N and k) for the sequence $A_k = \mathbb{E}_{x^0}(S^{k,N})^q$. By the definition of $S^{k,N}$, we have

$$S^{k+1,N} = S^{k,N} + \frac{\|x^{k+1,N} - x^{k,N}\|_{\mathcal{C}_N}^2}{N} + \frac{2\langle x^{k+1,N} - x^{k,N}, x^{k,N} \rangle_{\mathcal{C}_N}}{N}.$$

Therefore,

$$\begin{aligned} & \mathbb{E}_{x^0}(S^{k+1,N})^q - \mathbb{E}_{x^0}(S^{k,N})^q \\ &= \sum_{\substack{n+m+l=q \\ (n,m,l) \neq (q,0,0)}} \binom{q}{n,m,l} \mathbb{E}_{x^0} \left[(S^{k,N})^n \left(\frac{\|x^{k+1,N} - x^{k,N}\|_{\mathcal{C}_N}^2}{N} \right)^m \right. \\ & \quad \left. \times \left(\frac{2\langle x^{k+1,N} - x^{k,N}, x^{k,N} \rangle_{\mathcal{C}_N}}{N} \right)^l \right]. \end{aligned} \tag{C.2}$$

Thus, to establish (C.1) it is enough to argue that each of the terms in the right-hand side of the above is bounded by $(C/\sqrt{N})(1 + \mathbb{E}(S^{k,N})^q)$. To this end, set

$$\begin{aligned} J^{k,N} &:= \mathbb{E}_{x^0} \left[(S^{k,N})^n \left(\frac{\|x^{k+1,N} - x^{k,N}\|_{\mathcal{C}_N}^2}{N} \right)^m \left(\frac{2\langle x^{k+1,N} - x^{k,N}, x^{k,N} \rangle_{\mathcal{C}_N}}{N} \right)^l \right] \\ &= \mathbb{E}_{x^0} \mathbb{E}_k \left[(S^{k,N})^n \left(\frac{\|x^{k+1,N} - x^{k,N}\|_{\mathcal{C}_N}^2}{N} \right)^m \left(\frac{2\langle x^{k+1,N} - x^{k,N}, x^{k,N} \rangle_{\mathcal{C}_N}}{N} \right)^l \right]. \end{aligned}$$

By the Cauchy–Schwartz inequality for the scalar product $\langle \cdot, \cdot \rangle_{\mathcal{C}_N}$,

$$\begin{aligned} \frac{\langle x^{k+1,N} - x^{k,N}, x^{k,N} \rangle_{\mathcal{C}_N}^l}{N^l} &\leq \frac{\|x^{k,N}\|_{\mathcal{C}_N}^l \|x^{k+1,N} - x^{k,N}\|_{\mathcal{C}_N}^l}{N^l} \\ &= (S^{k,N})^{l/2} \frac{\|x^{k+1,N} - x^{k,N}\|_{\mathcal{C}_N}^l}{N^{l/2}}, \end{aligned}$$

which gives

$$J_k^N \lesssim \mathbb{E}_{x^0} \left[(S^{k,N})^{n+l/2} \frac{\|x^{k+1,N} - x^{k,N}\|_{\mathcal{C}_N}^{2m+l}}{N^{m+l/2}} \right].$$

Using the bound (6.4) of Lemma 6.1, we also have

$$\mathbb{E}_k \frac{\|x^{k+1,N} - x^{k,N}\|_{\mathcal{C}_N}^{2m+l}}{N^{m+l/2}} \lesssim \frac{(S^{k,N})^{m+l/2}}{N^{m+l/2}} + \frac{1}{N^{(m+l/2)/2}}.$$

Putting all of the above together (and using Young’s inequality) we obtain

$$J_k^N \lesssim \frac{\mathbb{E}_{x^0}[(S^{k,N})^q]}{N^{m+l/2}} + \frac{1}{N^{(m+l/2)/2}}.$$

Now observe that $(m + l/2)/2 \geq 1/2$ except when $(n, m, l) = (q, 0, 0)$ or $(n, m, l) = (q - 1, 0, 1)$. Therefore we have shown the desired bound for all

the terms in the expansion (C.2), except the one with $(n, m, l) = (q - 1, 0, 1)$. To study the latter term, we recall that $\gamma^{k,N} \in \{0, 1\}$, and use the definition of the chain [Eqs. (2.8) and (2.12)] to obtain

$$\begin{aligned} \left| \left\langle x^{k+1,N} - x^{k,N}, x^{k,N} \right\rangle_{\mathcal{C}_N} \right| &\lesssim \delta \left\| x^{k,N} \right\|_{\mathcal{C}_N}^2 + \delta \left| \left\langle \mathcal{C}_N \nabla \Psi^N(x^{k,N}), x^{k,N} \right\rangle_{\mathcal{C}_N} \right| \\ &\quad + \sqrt{\delta} \left| \left\langle x^{k,N}, (\mathcal{C}_N)^{1/2} \xi^{k,N} \right\rangle_{\mathcal{C}_N} \right|. \end{aligned}$$

Combining (2.26) with the Cauchy–Schwartz inequality we have

$$\delta \left| \left\langle \mathcal{C}_N \nabla \Psi^N(x^{k,N}), x^{k,N} \right\rangle_{\mathcal{C}_N} \right| \lesssim N^{-1/2} \left(1 + \left\| x^{k,N} \right\|_{\mathcal{C}_N}^2 \right) \lesssim N^{-1/2} + N^{1/2} S^{k,N},$$

where in the last inequality we used the following observation

$$\left\| x^{k,N} \right\|_s^2 = \sum_{j=1}^{\infty} (x^{k,N})_j^2 j^{2s} = \sum_{j=1}^{\infty} \frac{(x^{k,N})_j^2}{\lambda_j^2} (\lambda_j^2 j^{2s}) \lesssim \sum_{j=1}^{\infty} \frac{(x^{k,N})_j^2}{\lambda_j^2} = N S^{k,N}.$$

Recalling that $\left\langle x^{k,N}, (\mathcal{C}_N)^{1/2} \xi^{k,N} \right\rangle_{\mathcal{C}_N}$, conditioned on $x^{k,N}$, is a linear combination of zero-mean Gaussian random variables, we have

$$\begin{aligned} \mathbb{E}_k \sqrt{\delta} \left| \left\langle x^{k,N}, (\mathcal{C}_N)^{1/2} \xi^{k,N} \right\rangle_{\mathcal{C}_N} \right| &\lesssim 1 + N^{-1/2} \mathbb{E}_k \left| \left\langle x^{k,N}, (\mathcal{C}_N)^{1/2} \xi^{k,N} \right\rangle_{\mathcal{C}_N} \right|^2 \\ &\lesssim 1 + \sqrt{N} S^{k,N}. \end{aligned}$$

Putting the above together and taking expectations we can then conclude

$$\begin{aligned} \mathbb{E} \left[\frac{(S^{k,N})^{q-1} \left\langle x^{k+1,N} - x^{k,N}, x^{k,N} \right\rangle_{\mathcal{C}_N}}{N} \right] &\lesssim \frac{\mathbb{E} \left[(S^{k,N})^{q-1} \right]}{N} + \frac{\mathbb{E} \left[(S^{k,N})^q \right]}{\sqrt{N}} \\ &\lesssim (1/\sqrt{N}) (1 + \mathbb{E} \left[(S^{k,N})^q \right]), \end{aligned}$$

and (6.6) follows.

- (ii) Proof of (6.7). This is very similar to the proof of (6.6), so we only sketch it. Just as before, it is enough to establish the following bound

$$\begin{aligned} &\mathbb{E} \left[\left\| x^{k,N} \right\|_s^{2n} \left\| x^{k+1,N} - x^{k,N} \right\|_s^{2m} \left\langle x^{k+1,N} - x^{k,N}, x^{k,N} \right\rangle_s^l \right] \\ &\lesssim \frac{1}{\sqrt{N}} \left(1 + \mathbb{E} \left[\left\| x^{k,N} \right\|_s^{2q} \right] \right) \end{aligned}$$

for each (n, m, l) such that $n+m+l = q$ with the exception of the triple $(n, m, l) = (q, 0, 0)$. Applying the Cauchy–Schwartz inequality for $\langle \cdot, \cdot \rangle_s$ we have

$$\left\langle x^{k+1,N} - x^{k,N}, x^{k,N} \right\rangle_s^l \leq \left\| x^{k,N} \right\|_s^l \left\| x^{k+1,N} - x^{k,N} \right\|_s^l.$$

Thus, Lemma 6.1 implies

$$\begin{aligned} & \mathbb{E}_k \left\| x^{k,N} \right\|_s^{2n} \left\| x^{k+1,N} - x^{k,N} \right\|_s^{2m} \left\langle x^{k+1,N} - x^{k,N}, x^{k,N} \right\rangle_s^l \\ & \leq \left\| x^{k,N} \right\|_s^{2n+l} \mathbb{E}_k \left\| x^{k+1,N} - x^{k,N} \right\|_s^{2m+l} \\ & \lesssim \frac{\left\| x^{k,N} \right\|_s^{2n+l} (1 + \left\| x^{k,N} \right\|_s^{2m+l})}{N^{(m+l/2)/2}}. \end{aligned}$$

The above gives us the desired bound for all (n, m, l) except for $(n, m, l) = (q - 1, 0, 1)$. Like before, to study the latter case we observe

$$\begin{aligned} & \left\langle x^{k+1,N} - x^{k,N}, x^{k,N} \right\rangle_s \\ & = \gamma^{k,N} \left(-\frac{\ell}{\sqrt{N}} \left(\left\| x^{k,N} \right\|_s^2 + \left\langle \mathcal{C}_N \nabla \Psi^N(x^{k,N}), x^{k,N} \right\rangle_s \right) \right. \\ & \quad \left. + \frac{\sqrt{2\ell}}{N^{1/4}} \left\langle (\mathcal{C}_N)^{1/2} \xi^{k,N}, x^{k,N} \right\rangle_s \right) \\ & \lesssim \frac{1}{\sqrt{N}} \left(1 + \left\| x^{k,N} \right\|_s^2 \right) + \frac{1}{N^{1/4}} \gamma^{k,N} \left\langle (\mathcal{C}_N)^{1/2} \xi^{k,N}, x^{k,N} \right\rangle_s \\ & \lesssim \frac{1}{\sqrt{N}} \left(1 + \left\| x^{k,N} \right\|_s^2 \right), \end{aligned}$$

where penultimate inequality follows from the Cauchy–Schwartz inequality, (2.25), and the fact that $\gamma^{k,N} \in \{0, 1\}$, and the last inequality follows from Lemma 6.5. This concludes the proof. \square

Remark C.1 In [17] the authors derived the diffusion limit for the chain under weaker assumptions on the potential Ψ than those we use in this paper. Essentially, they assume that Ψ is quadratically bounded, while we assume that it is linearly bounded. If Ψ was quadratically bounded the proof of Lemma 6.5 would become considerably more involved. We observe explicitly that the statement of Lemma 6.5 is of paramount importance in order to establish the uniform bound on the moments of the chain x^k contained in Lemma 6.2. In [17] obtaining such bounds is not an issue, since the authors study the chain in its stationary regime. In other words, in [17] the law of $x^{k,N}$ is independent of k , and thus the uniform bounds on the moments of $x^{k,N}$ and $S^{k,N}$ are automatically true for target measures of the form considered there (see also the first bullet point of Remark 4.1). \square

References

1. Beskos, A., Girolami, M., Lan, S., Farrell, P., Stuart, A.: Geometric MCMC for infinite-dimensional inverse problems. *J. Comput. Phys.* **335**, 327–351 (2017)
2. Bédard, M.: Weak convergence of Metropolis algorithms for non-i.i.d. target distributions. *Ann. Appl. Probab.* **17**(4), 1222–1244 (2007)
3. Bédard, M., Rosenthal, J.: Optimal scaling of Metropolis algorithms: heading toward general target distributions. *Can. J. Stat.* **36**(4), 483–503 (2008)
4. Beskos, A., Roberts, G., Stuart, A., Voss, J.: An MCMC method for diffusion bridges. *Stochast. Dyn.* **8**(3), 319–350 (2008)
5. Breyer, L., Piccioni, M., Scarlatti, S.: Optimal scaling of MALA for nonlinear regression. *Ann. Appl. Probab.* **14**(3), 1479–1505 (2004)
6. Christensen, O., Roberts, G., Rosenthal, J.: Scaling limits for the transient phase of local Metropolis–Hastings algorithms. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **67**(2), 253–268 (2005)
7. Cotter, S., Roberts, G., Stuart, A., White, D., et al.: MCMC methods for functions: modifying old algorithms to make them faster. *Stat. Sci.* **28**(3), 424–446 (2013)
8. Da Prato, G., Zabczyk, J.: *Stochastic Equations in Infinite Dimensions*. Encyclopedia of Mathematics and Its Applications. Cambridge University Press, Cambridge (1992)
9. Hairer, M., Stuart, A., Voss, J.: Analysis of SPDEs arising in path sampling. Part II: the nonlinear case. *Ann. Appl. Probab.* **17**(5–6), 1657–1706 (2007)
10. Hairer, M., Stuart, A., Voss, J., Wiberg, P.: Analysis of SPDEs arising in path sampling. Part I: the Gaussian case. *Commun. Math. Sci.* **3**, 587–603 (2005)
11. Hastings, W.: Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109 (1970)
12. Jourdain, B., Lelièvre, T., Miasojedow, B.: Optimal scaling for the transient phase of Metropolis–Hastings algorithms: the longtime behavior. *Bernoulli* **20**(4), 1930–1978 (2014)
13. Jourdain, B., Lelièvre, T., Miasojedow, B.: Optimal scaling for the transient phase of the random walk Metropolis algorithm: the mean-field limit. *Ann. Appl. Probab.* **25**(4), 2263–2300 (2015)
14. Kuntz, J., Ottobre, M., Stuart, A.: Diffusion limit for the Random Walk Metropolis algorithm out of stationarity. *Arxiv preprint* (2016)
15. Mattingly, J., Pillai, N., Stuart, A.: Diffusion limits of the random walk Metropolis algorithm in high dimensions. *Ann. Appl. Probab.* **22**(3), 881–930 (2012)
16. Neal, R.M.: Regression and classification using Gaussian process priors (with discussion). In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (eds.) *Bayesian statistics 6*. Oxford University Press (1998). <https://www.cs.toronto.edu/~radford/ftp/val6gp.pdf>
17. Pillai, N., Stuart, A., Thiéry, A.: Optimal scaling and diffusion limits for the Langevin algorithm in high dimensions. *Ann. Appl. Probab.* **22**(6), 2320–2356 (2012)
18. Pillai, N., Stuart, A., Thiéry, A.: Noisy gradient flow from a random walk in Hilbert space. *Stoch. Partial Differ. Equ. Anal. Comput.* **2**(2), 196–232 (2014)
19. Roberts, G., Gelman, A., Gilks, W.: Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.* **7**(1), 110–120 (1997)
20. Roberts, G., Rosenthal, J.: Optimal scaling of discrete approximations to Langevin diffusions. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **60**(1), 255–268 (1998)
21. Roberts, G., Tweedie, R.: Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli* **2**(4), 341–363 (1996)
22. Stuart, A.: Inverse problems: a Bayesian perspective. *Acta Numerica* **19**, 451–559 (2010)
23. Tierney, L.: A note on Metropolis–Hastings kernels for general state spaces. *Ann. Appl. Probab.* **8**(1), 1–9 (1998)