Diffusion Limit For The Random Walk Metropolis Algorithm Out Of stationarity

Juan Kuntz, Michela Ottobre and Andrew M. Stuart*

Juan Kuntz
Imperial College London,
SW7 2AZ, London, UK,
e-mail: juan.kuntz08@imperial.ac.uk

Michela Ottobre
Heriot Watt University, Mathematics Department,
Edinburgh, EH14 4AS, UK
e-mail: m.ottobre@hw.ac.uk

Andrew M. Stuart
Department of Computing and Mathematical Sciences,
California Institute of Technology, CA 91125, USA e-mail: astuart@caltech.edu

Abstract:
The Random Walk Metropolis (RWM) algorithm is a Metropolis-Hastings Markov Chain Monte Carlo algorithm designed to sample from a given target distribution π^N with Lebesgue density on $\mathbb{R}^N$. Like any other Metropolis-Hastings algorithm, RWM constructs a Markov chain by randomly proposing a new position (the "proposal move"), which is then accepted or rejected according to a rule which makes the chain reversible with respect to π^N. When the dimension N is large a key question is to determine the optimal scaling with N of the proposal variance: if the proposal variance is too large, the algorithm will reject the proposed moves too often; if it is too small, the algorithm will explore the state space too slowly. Determining the optimal scaling of the proposal variance gives a measure of the cost of the algorithm as well. One approach to tackle this issue, which we adopt here, is to derive diffusion limits for the algorithm. Such an approach has been proposed in the seminal papers [RGG97, RR98]; in particular in [RGG97] the authors derive a diffusion limit for the RWM algorithm under the two following assumptions: i) the algorithm is started in stationarity; ii) the target measure π^N is in product form. The present paper considers the situation of practical interest in which both assumptions i) and ii) are removed. That is a) we study the case (which occurs in practice) in which the algorithm is started out of stationarity and b) we consider target measures which are in non-product form. In particular, we work in the setting in which families of measures on spaces of increasing dimension are found by approximating a measure, on an infinite dimensional Hilbert space, which is defined by its density with respect to a Gaussian. The target measures that we consider arise in Bayesian nonparametric statistics and in the study of conditioned diffusions. We prove that, out of stationarity, the optimal scaling for the proposal variance is $O(N^{-1})$, as it is in stationarity. In this optimal scaling a diffusion limit is obtained and the cost of reaching and exploring the invariant measure scales as $O(N)$. Notice that the optimal scaling in and out of stationatity need not be the same in general, and indeed they differ e.g. in the case of the MALA algorithm [KOS16].

MSC 2010 subject classifications: Primary 60J22; secondary 60J20, 60H10.
Keywords and phrases: Markov Chain Monte Carlo, Random Walk Metropolis algorithm, diffusion limit, optimal scaling.

1. Introduction

1.1. Setting and Main Result

Metropolis-Hastings algorithms are popular MCMC methods used to sample from a given target measure, π^N, defined via its density with respect to Lebesgue measure on $\mathbb{R}^N$ (with abuse of notation, we often denote both the measure and the density with the same letter). The basic mechanism consists of employing a proposal transition density $q(x, y)$ in order to produce a reversible chain $\{x_k\}_{k=0}^{\infty}$ which has the target measure as invariant distribution [Tie98]. At step k of the chain, a proposal move y_{k+1} is generated by using

*Supported by ERC and EPSRC

a proposal kernel $q(x,y)$, i.e. $y_{k+1} \sim q(x_k,\cdot)$. Then such a move is accepted with probability $\alpha(x_k, y_{k+1})$, where

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

If the move is accepted then the chain is updated to the state $x_{k+1} := y_{k+1}$, otherwise $x_{k+1} := x_k$. When the proposal kernel $q(x,y)$ is symmetric in its variables, the expression for the acceptance probability simplifies to

$$\alpha(x,y) = \min\left\{1, \frac{\pi(y)}{\pi(x)}\right\}.$$

Random Walk Metropolis (RWM) belongs to the family of Metropolis-Hastings algorithms with symmetric proposal, as the proposal move is generated according to a random walk. A key question for Metropolis-Hastings methods in general, and for RWM in particular, is to determine the cost of the algorithm as a function of the dimension N. The present paper aims at studying the cost of the RWM algorithm by the use of diffusion limits. Precisely, we identify scalings of the proposal variance with resepct to the dimension N which lead to a diffiusion limit. Since the inverse proposal variance has the interpretation as a time-step in a discretization of the limiting diffusion, this scaling determines the number of steps required to reach and explore the desired target distribution. We study the situation of practical interest where the algorithm is started out of stationarity and the target measure is in non-product form.

In what follows we first introduce the class of target measures that we will be considering and we then specify the RWM algorithm for such a class of targets (more details on the algorithm and on the class of target measures can be found in Section 2 and in Section 3, respectively). We then clarify the problem that is the subject of the paper, we present our main result and, immediately after (see Remark 1.1), we explain the practical implications of such a result in terms of cost of the algorithm (in this context we will specify what we mean by "cost of the algorithm").

The class of target measures that we consider are determined by approximations of a measure on an infinite dimensional Hilbert space. In particular, let π be a probability measure defined on an infinite dimensional separable Hilbert space $(\mathcal{H}, \langle\cdot,\cdot\rangle, \|\cdot\|)$ and absolutely continuous with respect to a Gaussian measure π_0 with mean zero and covariance operator $\mathcal{C}$:

$$\frac{d\pi}{d\pi_0} \propto \exp(-\Psi), \qquad \pi_0 \overset{\mathcal{D}}{\sim} \mathcal{N}(0,\mathcal{C}), \tag{1.1}$$

where $\Psi : \tilde{\mathcal{H}} \to \mathbb{R}$ is some real valued functional with domain $\tilde{\mathcal{H}} \subseteq \mathcal{H}$ and $\pi_0(\tilde{\mathcal{H}}) = 1$. In Section 3 we will detail our assumptions on Ψ and 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 Section 3). The covariance operator $\mathcal{C}$ is a positive, self-adjoint, trace class operator on $\mathcal{H}$, with eigenbasis $\{\lambda_j^2, \phi_j\}_{j\in\mathbb{N}}$:

$$\mathcal{C}\phi_j = \lambda_j^2\phi_j, \quad \forall j \in \mathbb{N}, \tag{1.2}$$

where $\{\phi_j\}_{j\in\mathbb{N}}$ is an orthonormal basis of $\mathcal{H}$. We will analyse the RWM algorithm designed to sample from the finite dimensional projections π^N of the measure (1.1) on the space

$$\mathcal{H} \supset X^N := \mathrm{span}\{\phi_j\}_{j=1}^N \tag{1.3}$$

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 $x \in \mathcal{H}$, $x^N := \mathcal{P}^N(x)$ is the projection of x onto the space X^N; $x^{i,N}$ will be the i-th component of the vector $x^N \in \mathbb{R}^N$, i.e. $x^{i,N} = \langle\phi_i, x^N\rangle$. [1] Similar notation is also used for y, ξ and other vectors; we do not give details. We will also denote $\Psi^N(x) := \Psi(\mathcal{P}^N(x))$ and $\mathcal{C}_N$ will be, effectively, an $N \times N$ diagonal matrix with i-th diagonal component equal to λ_i^2. More formally,

$$\Psi^N := \Psi \circ \mathcal{P}^N \quad \text{and} \quad \mathcal{C}_N := \mathcal{P}^N \circ \mathcal{C} \circ \mathcal{P}^N. \tag{1.4}$$

[1] Notice that if $x^N = \mathcal{P}^N(x)$ and $1 \le i \le N$ then $x^{i,N} = \langle\phi_i, x^N\rangle = \langle\phi_i, x\rangle$.

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

$$\frac{d\pi^N}{d\pi_0^N}(x) = M_{\Psi^N} e^{-\Psi^N(x)}, \qquad \pi_0^N \sim \mathcal{N}(0, \mathcal{C}_N), \tag{1.5}$$

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 $\{\pi^N\}_{N\in\mathbb{N}}$ convereges to π in the Hellinger metric [Stu10]).

Letting $\ell > 0$ denote a positive parameter, consider the RWM algorithm with proposal

$$y = x + \sqrt{\frac{2\ell^2}{N}} \mathcal{C}_N^{1/2} \xi^N, \qquad \xi^N = \sum_{j=1}^N \xi^{j,N} \phi_j, \quad \xi^{j,N} \overset{\mathcal{D}}{\sim} \mathcal{N}(0,1) \text{ i.i.d..} \tag{1.6}$$

The current position x and the proposal y belong to $\mathcal{H}$; however, because the noise is finite dimensional, effectively only the first N components of x are modified when a proposal is accepted, namely the components belonging to X^N.

Using the proposal (1.6) we construct the RWM - Markov chain $\{x_k\}_k \subset \mathcal{H}$, through the "accept-reject" mechanism described earlier. In computational practice one uses the projected chain $x_k^N = \mathcal{P}^N(x_k)$, which samples from the measure π^N, i.e. for any fixed $N \in \mathbb{N}$, the chain $\{x_k^N\}_{k\in\mathbb{N}} \subset X^N$ can be used to sample from the measure π^N. However we often work in $\mathcal{H}$ rather than in X^N (and therefore consider the chain $\{x_k\}_{k\in\mathbb{N}}$ rather than the chain $\{x_k^N\}_{k\in\mathbb{N}}$) only because in $\mathcal{H}$ the analysis is cleaner.

To explain the problem at hand consider for a moment, instead of the proposal (1.6), the following proposal:

$$y = x + \sqrt{\frac{2\ell^2}{N^\beta}} \mathcal{C}_N^{1/2} \xi^N, \tag{1.7}$$

where $\beta > 0$ is a positive parameter to be chosen. As is well known, if β is "too large" then the proposal variance (that is, informally, the size of the jumps of the chain) is "too small", therefore the algorithm will move in state space very slowly. On the other hand, if β is "too small" then the proposal variance is too large and the algorithm will tend to reject the proposed moves too frequently (and this is more and more the case as the dimension N increases). We will show that the value of β that strikes the balance between these two opposing scenarios is $\beta = 1$.

We are now in a position to present our main result: let $x^{(N)}(t)$ be the continuous interpolant of the chain $\{x_k\}$, namely

$$x^{(N)}(t) = (Nt - k)x_{k+1} + (k + 1 - Nt)x_k, \quad t_k \le t < t_{k+1}, \text{ where } t_k = k/N. \tag{1.8}$$

The main result of this paper is the diffusion limit for the RWM algorithm started out of stationarity. We informally state such a result below, with the functions D_ℓ, Γ_ℓ and A_ℓ defined immediately after the statement. The rigorous statement of the result, with precise conditions, appears in Theorem 5.1 and Theorem 5.4. Below we denote by $C([0,T];\tilde{\mathcal{H}})$ the space of $\tilde{\mathcal{H}}$- valued continuous functions on $[0,T]$, endowed with the uniform topology.

Main Result. *Let $\{x_k\}_{k\in\mathbb{N}}$ be the Markov chain constructed using the RWM proposal (1.6) and starting from the (deterministic) initial datum $x_0 \in \tilde{\mathcal{H}}$. Assume*

$$S_0 := \lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^N \frac{\left|x_0^{j,N}\right|^2}{\lambda_j^2} < \infty. \tag{1.9}$$

Then the continuous interpolant of the chain x_k, i.e. the sequence of processes $x^{(N)}(t)$ defined in (1.8), converges weakly in $C([0,T];\tilde{\mathcal{H}})$ (as $N \to \infty$) to the solution of the SDE

$$dx(t) = [-x(t) - \mathcal{C}\nabla\Psi(x(t))] D_\ell(S(t))\, dt + \sqrt{\Gamma_\ell(S(t))}\, dW(t), \quad x(0) = x_0\,, \tag{1.10}$$

where $S(t) : \mathbb{R}_+ \to \mathbb{R}_+ := \{s \in \mathbb{R} : s \ge 0\}$ is a deterministic function which solves the ODE

$$dS(t) = A_\ell(S(t))\, dt, \qquad S(0) = S_0\,, \tag{1.11}$$

and $W(t)$ is a $\tilde{\mathcal{H}}$-valued $\tilde{\mathcal{C}}$-Brownian motion. [2]

If we denote by $\Phi(x)$ the cdf of a standard Gaussian, the functions $D_\ell, \Gamma_\ell, A_\ell : \mathbb{R}_+ \to \mathbb{R}$ that appear in the above statement are defined as follows: for $x > 0$ and $\ell > 0$ a positive parameter, we define

$$D_\ell(x) := 2\ell^2 e^{\ell^2(x-1)}\Phi\left(\frac{\ell(1-2x)}{\sqrt{2x}}\right), \tag{1.12}$$

$$\Gamma_\ell(x) := D_\ell(x) + 2\ell^2\Phi\left(-\frac{\ell}{\sqrt{2x}}\right), \tag{1.13}$$

$$A_\ell(x) := (1-2x)D_\ell(x) + 2\ell^2\Phi\left(-\frac{\ell}{\sqrt{2x}}\right) = -2xD_\ell(x) + \Gamma_\ell(x) \tag{1.14}$$

and for $x = 0$ and $\ell > 0$ we set

$$D_\ell(0) = \Gamma_\ell(0) = A_\ell(0) = 2\ell^2 e^{-\ell^2}. \tag{1.15}$$

Remark 1.1. We make several remarks concerning the main result.

- The effective time-step implied by the interpolation (1.8) is N^{-1} so, in this sense, the main result indicates that, started out of stationarity, the RWM algorithm will take $\mathcal{O}(N)$ steps to reach and explore target measures found by approximating π in $\mathbb{R}^N$. In this respect, we say that the computational cost of the algorithm is of order N. To put it differently, our result proves that the proposal variance which delivers a diffusion limit scales like N^{-1} with dimension and that, therefore, the cost of the algorithm is of order N.
- Notice that equation (1.11) evolves independently of equation (1.10). Once the RWM chain $\{x_k\}_k$ is introduced (see (2.3) for a precise description of the chain) and an initial state $x_0 \in \tilde{\mathcal{H}}$ is given such that $S(0)$ is finite, the real valued (double) sequence S_k^N,

$$S_k^N := \frac{1}{N}\sum_{i=1}^N \frac{\left|x_k^{i,N}\right|^2}{\lambda_i^2} \tag{1.16}$$

started at $S_0^N := \frac{1}{N}\sum_{i=1}^N \frac{\left|x_0^{i,N}\right|^2}{\lambda_i^2}$ is well defined. We can then consider the continuous interpolant $S^{(N)}(t)$ of the chain $\{S_k^N\} \subset \mathbb{R}_+$, namely

$$S^{(N)}(t) = (Nt-k)S_{k+1}^N + (k+1-Nt)S_k^N, \quad t_k \le t < t_{k+1}, \text{ where } t_k = k/N. \tag{1.17}$$

In Theorem 5.1 we prove that $S^{(N)}(t)$ converges in probability in $C([0,T];\mathbb{R})$ to the solution of (1.11) with initial condition $S_0 := \lim_{N\to\infty} S_0^N$. 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 $x(t)$. Moreover, notice that S_k^N is not a Markov Chain in general (unless e.g. $\Psi = 0$.)
- Let $S(t) : \mathbb{R}_+ \to \mathbb{R}_+$ be the solution of the ODE (1.11). We will prove (see Theorem 4.1) that $S(t) \to 1$ as $t \to \infty$. With this in mind, notice that $D_\ell(1) = 2\ell^2\Phi(-\ell/\sqrt{2}) =: h_\ell$ and $\Gamma_\ell(1) = 2D_\ell(1) = 2h_\ell$. Heuristically one can then argue that the asymptotic behaviour of the law of $x(t)$, solution of (1.10), is described by the law of the following infinite dimensional SDE:

$$dz(t) = -h_\ell(z + \mathcal{C}\nabla\Psi(z)) + \sqrt{2h_\ell}dW. \tag{1.18}$$

It was proved in [HSVW05, HSV07] that (1.18) is ergodic with unique invariant measure given by our target measure (1.1). Our deduction concerning computational cost is made on the assumption that the law of (1.10) does indeed tend to the law of (1.18), 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 RWM algorithm.

□

[2] The operator that here we denote generically by $\tilde{\mathcal{C}}$, to avoid getting in too much notation at this stage, will be more clearly defined in Section 3 and there denoted by $\mathcal{C}_s$. More precisely, as we will explain, $W(t)$ is a Brownian motion with covariance $\mathcal{C}_s$, see Section 3.

1.2. Relation to the Literature

As already explained, in this paper we consider target measures in non-product form, when the chain is started out of stationarity. When the target measure is in product form, a diffusion limit for the resulting Markov chain was studied in the seminal paper [RGG97], where it is assumed that

$$p(x^N) = \Pi_{i=1}^N e^{-V(x^{i,N})}, \quad x^N = (x^{1,N}, \dots, x^{N,N}) \in \mathbb{R}^N, \tag{1.19}$$

and the potential V is such that the measure p is normalized. That work assumed that the chain is started in stationarity, leading to the conclusion that, in stationarity, $\mathcal{O}(N)$ steps are required to explore the target distribution. In [CRR05] the same question was addressed in the case where p is the density of an isotropic Gaussian, when the chain is started out of stationarity. Recently the papers [JLM15, JLM14] made the significant extension of considering the product case (1.19) for quite general potentials V, again out of stationarity. The work in [CRR05, JLM15] demonstrates that the same scaling of the proposal variance is required both in and out of stationarity, in the product case, and that then $\mathcal{O}(N)$ steps are required to explore the target distribution. Again recently, diffusion limits for RWM started in stationarity have also been considered for measures in non-product form [MPS12], using families of target measures found by approximating (1.1), as we consider in this paper; once again the conclusion is that $\mathcal{O}(N)$ steps are required to explore the target distribution. In the present paper we combine the settings of [MPS12] and [JLM15] and make a significant extension of the analysis to consider measures in non-product form, when the chain is started out of stationarity, again showing that $\mathcal{O}(N)$ steps are required to explore the target distribution.

In [RGG97] the diffusion limit is for a single coordinate of the Markov chain and takes the form

$$dX(t) = -h_\ell V'(X(t))dt + \sqrt{2h_\ell}dB(t), \tag{1.20}$$

with $X_t \in \mathbb{R}$ and $B(t)$ a one dimensional Brownian motion. Each coordinate of the Markov chain has the same weak limit. In [JLM15, JLM14] a similar limit is obtained for each coordinate, but because the system is out of stationarity the coordinates are coupled together, leading to a one dimensional nonlinear (in the sense of McKean) diffusion process

$$dX(t) = -d_\ell(t)V'(X(t))dt + \sqrt{2g_\ell(t)}dB(t), \tag{1.21}$$

with $X_t \in \mathbb{R}$ and $B(t)$ a one dimensional Brownian motion and

$$d_\ell(t) = \mathcal{G}_\ell\left(\mathbb{E}\left[V'(X(t))\right]^2, \mathbb{E}\left[V''(X(t))\right]\right), \quad g_\ell(t) = \frac{1}{2}\tilde{\Gamma}_\ell\left(\mathbb{E}\left[V'(X(t))\right]^2, \mathbb{E}\left[V''(X(t))\right]\right).$$

The definition of the functions $\mathcal{G}_\ell$ and $\tilde{\Gamma}_\ell$ can be found in [JLM15, (1.7) and (1.6)], respectively. While we don't repeat the full definition here, we point out the two main facts which are relevant in the present context: i) in stationarity $d_\ell(t) = h_\ell$ and $g_\ell(t) = h_\ell$ and so (1.21) is identical to (1.20), but out of stationarity the variation of these quantities reflects what remains of the coupling between different coordinates in the limit of large N; ii) regarding the functions $D_\ell(x)$ and $\Gamma_\ell(x)$ (defined in (1.12) and (1.13), respectively), notice that $D_\ell(x) = \mathcal{G}_{\ell\sqrt{2}}(x, 1)$, $\Gamma_\ell(x) = \tilde{\Gamma}_{\ell\sqrt{2}}(x, 1)$.

In [MPS12], since the target measure is no longer of product form, the continuous interpolant of the RWM chain x_k defined in (2.3) has diffusion limit given by the solution of the infinite dimensional SDE (1.18), when the chain is started in stationarity. In contrast, in this paper where we study the same target measure as in [MPS12], but started out of stationarity, the limiting diffusion is (1.10), with $S(t)$ solving (1.11). The relationship between (1.20) and (1.21) is entirely analogous to the relationship between (1.18) and (1.10). It is natural to ask, then, why we do not obtain an infinite dimensional nonlinear (in the McKean sense) diffusion process as the limit in this paper? The reason for this is related to the fact that our underlying reference measure is Gaussian. Indeed in the case of Gaussian product measure the limiting diffusion (1.21) simplifies in the sense that the the equations for $d_\ell(t)$ and $g_\ell(t)$ depend only on the process X through the quantity $M(t) := \mathbb{E}(X_t)^2$ and it is explicitly noted in [JLM15] that $M(t)$ solves precisely the ODE (1.11). It is also relevant to observe at this point that the weak limit $S^{(N)} \xrightarrow{d} S$ (in $C([0,T], \mathbb{R}_+)$) has already been proven in [CRR05] in the Gaussian case where all the components $x_k^{i,N}$ are identically distributed.

On a technical note, we observe that in [JLM15, JLM14] the symmetry of the target measure allows the authors to employ propagation of chaos techniques so that these two papers have brought together two thus far distant worlds: MCMC and probabilistic methods for nonlinear PDEs. In our case, due to the lack of symmetry in the proposal, the propagation of chaos point of view cannot be used so we base our analysis on the more "hands on" approach used in [MPS12]. As already mentioned, the latter paper is devoted to the study of the diffusion limit for the same chain that we are analysing here and in the same infinite dimensional context as well. The difference with our paper is that the chain in [MPS12] is started in stationarity. As a consequence, the quantity that here we call $S(t)$ is, in their case, equal to 1 for every $t \geq 0$; to better phrase it, if we start the chain in stationarity, then

$$S_k^N = \frac{1}{N}\sum_{i=1}^{N} \frac{\left|x_k^{i,N}\right|^2}{\lambda_i^2} \overset{N\to\infty}{\longrightarrow} 1\,, \quad \text{almost surely, for all } k \geq 0. \tag{1.22}$$

Recalling that $S(t) \to 1$ as $t \to \infty$, this is coherent with our results. Although the approach we use here is similar to the one developed in [MPS12], significant extensions of that work are required in order to handle the technical complications introduced by the non stationarity of the chain. Throughout the paper we will flag up the main steps where our analysis differs from that in [MPS12] (see in particular Section 5.2, the comments at the end of Section 5.3 and Remark 8.7). Let us just say for the moment that if we start the chain in stationarity then $x_k^N \sim \pi^N$ for all $k \geq 0$. Because π^N is a change of measure from a Gaussian measure, all the almost sure properties of the chain only need to be shown for $x \sim \pi_0$. In the non stationary case we cannot reduce the analysis to the Gaussian case and therefore some of the estimates become more involved. The above discussion motivates our interest in the problem studied in this paper: on the one hand we want to extend the analysis of [JLM15] away from the non-practical i.i.d. product form for the target; on the other hand we drop the assumption of stationarity in [MPS12].

We mention for completeness that the non stationary case has also been considered in [PST14, OPPS16], for the pCN (preconditioned Crank-Nicolson) algorithm and for the SOL-HMC (Second Order Langevin - Hamiltonian Monte Carlo) scheme, respectively. These algorithms are well-defined in the infinite dimensional limit and hence do not require a scaling of the time-step which is inversely proportional to a power of the dimension. On a related note, we remark that when we want to sample from measures of the form (1.1), RWM is not the optimal choice. Indeed both pCN and the SOL-HMC exactly preserve the Gaussian measure π_0 and hence, in the case $\Psi \equiv 0$, such algorithms are exact; it is for this reason that they are well-defined in the infinite dimensional limit, and do not require a scaling of the time-step with dimension. However it is still of interest to study the behaviour of RWM on measures of the form (1.1) because they provide an explicit class of non-product measures for which analysis is possible and for which the scaling of cost with dimention is the same as in the product case, suggesting broader validity of the conclusions in the papers [CRR05, JLM15, JLM14].

1.3. Outline of Paper

The paper is organized as follows. In the next Section 2 we present in more detail the RWM algorithm. In Section 3 we introduce the notation that we will use in the rest of the paper and the assumptions we make on the nonlinearity Ψ and on the covariance operator $\mathcal{C}$. Section 4 contains the proof of existence and uniqueness for the limiting equations (1.10) and (1.11). With these preliminaries in place, we give, in Section 5, the precise statement of the main results of this paper, Theorem 5.1 and Theorem 5.4. In Section 5 we also provide heuristic arguments to explain how the main results are obtained. Such arguments are then made rigorous in Section 7 and Section 8, which contain the proof of Theorem 5.1 and Theorem 5.4, respectively. The continuous mapping argument on which these proofs rely is presented in Section 6.

2. The Algorithm

Once the current state x of the chain is given, the proposed move (1.6) depends only on the noise ξ^N. For this reason, in defining the acceptance probability for our algorithm, we can use the notations $\alpha(x^N, y^N)$ or

$\alpha(x^N, \xi^N)$ exchangeably. With this in mind, let us define the acceptance probability

$$\alpha(x^N, \xi^N) := 1 \wedge \exp\left(Q(x^N, \xi^N)\right) \tag{2.1}$$

where

$$Q(x^N, \xi^N) := \frac{1}{2}\|\mathcal{C}^{-1/2}x^N\|^2 - \frac{1}{2}\|\mathcal{C}^{-1/2}y^N\|^2 + \Psi(x^N) - \Psi(y^N). \tag{2.2}$$

Consider the Markov chain $\{x_k\}_{k=0}^{\infty} \subset \mathcal{H}$ constructed as follows

$$x_{k+1} = x_k + \gamma_{k+1}\sqrt{\frac{2\ell^2}{N}}\mathcal{C}_N^{1/2}\xi_{k+1}^N\,,^{3} \tag{2.3}$$

where

$$\gamma_{k+1} \overset{\mathcal{D}}{\sim} \text{Bernoulli}(\alpha_{k+1}) \qquad \text{with} \qquad \alpha_{k+1} = \alpha(x_k^N, \xi_{k+1}^N).$$

That is, given α_{k+1}, the random variable γ_{k+1} is independent of any other source of noise and has Bernoulli law with mean $\alpha(x_k^N, \xi_{k+1}^N)$. Therefore, (2.3) can be spelled out as follows: if the chain is currently in x_k, the proposal

$$y_{k+1} = x_k + \sqrt{\frac{2\ell^2}{N}}(\mathcal{C}_N)^{1/2}\xi_{k+1}^N$$

is accepted with probability α_{k+1} and rejected with probability $1 - \alpha_{k+1}$. We specify that in the above

$$\xi_{k+1}^N := \sum_{i=1}^{N} \xi_{k+1}^{i,N}\phi_i, \qquad \text{where} \qquad \xi_{k+1}^{i,N} \overset{\mathcal{D}}{\sim} \mathcal{N}(0,1) \text{ i.i.d.},$$

and therefore for α_k, Q and γ_k actually depend on N (we suppress the superscript N in the notation for convenience). In a less compact notation, (2.3) and (2.2) can be rewritten as

$$\begin{aligned} x_{k+1}^{i,N} &= x_k^{i,N} + \gamma_{k+1}\sqrt{\frac{2\ell^2}{N}}\,\lambda_i\,\xi_{k+1}^{i,N}, \qquad \text{for } i = 1, \dots, N \\ x_{k+1} &= x_k = x_0 \qquad \text{on } \mathcal{H} \setminus X^N \end{aligned} \tag{2.4}$$

and

$$Q_k := Q(x_k^N, \xi_{k+1}^N) = \frac{1}{2}\sum_{i=1}^{N}\frac{|x_k^{i,N}|^2}{\lambda_i^2} - \frac{1}{2}\sum_{i=1}^{N}\frac{|y_{k+1}^{i,N}|^2}{\lambda_i^2} + \Psi(x_k^N) - \Psi(y_{k+1}^N), \tag{2.5}$$

respectively. As we have already observed in the introduction, in computational practice the above algorithm is implemented in $\mathbb{R}^N$. That is, for any N fixed, in order to sample from the measure π^N (defined in (1.5)), one considers the projected chain $\{x_k^N = \mathcal{P}^N(x_k)\}_{k\in\mathbb{N}}$.

3. Preliminaries

In this section we detail the notation and the assumptions (Subsection 3.1 and Subsection 3.2 , respectively) that we will use in the rest of the paper.

3.1. Notation

Let $(\mathcal{H}, \langle\cdot,\cdot\rangle, \|\cdot\|)$ denote an infinite dimensional separable Hilbert space with the canonical norm derived from the inner-product. Let $\mathcal{C}$ be a positive, trace class operator on $\mathcal{H}$ and $\{\phi_j, \lambda_j^2\}_{j\geq 1}$ be the eigenfunctions and eigenvalues of $\mathcal{C}$ respectively, so that (1.2) holds. We assume a normalization under which $\{\phi_j\}_{j\geq 1}$ forms a complete orthonormal basis in $\mathcal{H}$. Throughout the paper we will use the following notation:

[3] Notice that also the state of the chain $\{x_k\}_{k\in\mathbb{N}} \subset \mathcal{H}$ depends on N, as only the first N components are updated. However this is not reflected in the notation to avoid confusion between the finite-dimensional chain $\{x_k^N\} \subset X^N$ and the infinite-dimensional chain $\{x_k\} \subset \mathcal{H}$.

- The letter N denotes exclusively the dimensionality of the space X^N (defined in (1.3)) where the target measure π^N is supported.
- As already stressed in the introduction, if $x \in \mathcal{H}$, then $x^N := \mathcal{P}^N(x)$ is the projection of x on the space X^N defined in (1.3). For every $x \in \mathcal{H}$ we have the representation $x = \sum_j x^j \phi_j$, where here $x^j = \langle x, \phi_j \rangle$, i.e. x^j is the j-th component of x. $x^{j,N}$ denotes the j-th component of x^N, so that $x^j = x^{j,N}$, for $1 \le j \le N$. Similar notation holds for the proposal vector y and the noise vector ξ as well.
- x_k^N denotes the k-th step of projected chain $\{\mathcal{P}^N(x_k)\} \subset X^N$, where x_k has been defined in (2.3). Accordingly, $x_k^{i,N}$ is the i-th component of the vector $x_k^N \in X^N$.

Using this notation, we define Sobolev-like spaces $\mathcal{H}^r, r \in \mathbb{R}$, with the inner-products and norms defined by

$$\langle x, y \rangle_r = \sum_{j=1}^{\infty} j^{2r} x^j y^j \qquad \text{and} \qquad \|x\|_r^2 = \sum_{j=1}^{\infty} j^{2r} \left| x^j \right|^2 .$$

$(\mathcal{H}^r, \langle \cdot, \cdot \rangle_r)$ is a Hilbert space. Notice that $\mathcal{H}^0 = \mathcal{H}$. Furthermore $\mathcal{H}^r \subset \mathcal{H} \subset \mathcal{H}^{-r}$ for any $r > 0$. The Hilbert-Schmidt norm $\| \cdot \|_{\mathcal{C}}$ is defined as

$$\|x\|_{\mathcal{C}}^2 = \|\mathcal{C}^{-\frac{1}{2}} x\|^2 = \sum_{j=1}^{\infty} \lambda_j^{-2} \left| x^j \right|^2 \tag{3.1}$$

and it is the Cameron-Martin norm associated with the Gaussian $\mathcal{N}(0, \mathcal{C})$. For $r \in \mathbb{R}$, let $L_r : \mathcal{H} \to \mathcal{H}$ denote the operator which is diagonal in the basis $\{\phi_j\}_{j \ge 1}$ with diagonal entries j^{2r}, *i.e.*,

$$L_r \, \phi_j = j^{2r} \phi_j, \qquad j \in \mathbb{N},$$

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

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

Since $\|L_r^{-1/2} \phi_k\|_r = \|\phi_k\| = 1$, we deduce that $\{\hat{\phi}_k := L_r^{-1/2} \phi_k\}_{k \ge 1}$ forms an orthonormal basis for $\mathcal{H}^r$. If $y \sim N(0, \mathcal{C})$, then y can be expressed as

$$y = \sum_{j=1}^{\infty} \lambda_j \rho_j \phi_j \qquad \text{with} \qquad \rho_j \overset{\mathcal{D}}{\sim} N(0,1) \text{ i.i.d;} \tag{3.2}$$

if $\sum_j \lambda_j^2 j^{2r} < \infty$ then y can be equivalently written as

$$y = \sum_{j=1}^{\infty} (\lambda_j j^r) \rho_j \hat{\phi}_j \qquad \text{with} \qquad \rho_j \overset{\mathcal{D}}{\sim} N(0,1) \text{ i.i.d.} \tag{3.3}$$

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 $\tilde{D} : \mathcal{H}^r \to \mathcal{H}^r$, its trace in $\mathcal{H}^r$ is

$$\text{Trace}_{\mathcal{H}^r}(\tilde{D}) := \sum_{j=1}^{\infty} \langle L_r^{-\frac{1}{2}} \phi_j, \tilde{D} L_r^{-\frac{1}{2}} \phi_j \rangle_r.$$

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

Because $\mathcal{C}$ is defined on $\mathcal{H}$, the covariance operator

$$\mathcal{C}_r = L_r^{1/2} \mathcal{C} L_r^{1/2}$$

is defined on $\mathcal{H}^r$. With this definition, for all the values of r such that $\text{Trace}_{\mathcal{H}^r}(\mathcal{C}_r) = \sum_j \lambda_j^2 j^{2r} < \infty$, we can think of y as a mean zero Gaussian random variable with covariance operator $\mathcal{C}$ in $\mathcal{H}$ and $\mathcal{C}_r$ in $\mathcal{H}^r$ (see (3.2) and (3.3)). In the same way, if $\text{Trace}_{\mathcal{H}^r}(\mathcal{C}_r) < \infty$ then

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

with $\{w_j(t)\}_{j \in \mathbb{N}}$ 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 $\mathcal{H}^r$-valued $\mathcal{C}_r$-Brownian motion.
Throughout we use the following notation.

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

As 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.

3.2. Assumptions

In this section we describe the assumptions on the covariance operator $\mathcal{C}$ of the Gaussian measure $\pi_0 \overset{\mathcal{D}}{\sim} \mathcal{N}(0, \mathcal{C})$ and the functional Ψ. We fix a distinguished exponent $s \geq 0$ and assume that $\Psi : \mathcal{H}^s \to \mathbb{R}$ and $\text{Trace}_{\mathcal{H}^s}(\mathcal{C}_s) < \infty$. In other words the space $\mathcal{H}^s$ is the one that we were denoting by $\tilde{\mathcal{H}}$ in the introduction. For each $x \in \mathcal{H}^s$ the derivative $\nabla \Psi(x)$ is an element of the dual $(\mathcal{H}^s)^*$ of $\mathcal{H}^s$, comprising the linear functionals on $\mathcal{H}^s$. However, we may identify $(\mathcal{H}^s)^* = \mathcal{H}^{-s}$ and view $\nabla \Psi(x)$ as an element of $\mathcal{H}^{-s}$ for each $x \in \mathcal{H}^s$. With this identification, the following identity holds

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

furthermore, the second derivative $\partial^2 \Psi(x)$ can be identified with an element of $\mathcal{L}(\mathcal{H}^s, \mathcal{H}^{-s})$. To avoid technicalities we assume that $\Psi(x)$ is quadratically bounded, with first derivative linearly bounded at infinity and second derivative globally bounded.

Assumptions 3.1. *The functional Ψ and covariance operator $\mathcal{C}$ satisfy the following assumptions.*

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

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

2. **Domain of Ψ:** *there exists an exponent $s \in [0, \kappa - 1/2)$ such that Ψ is defined everywhere on $\mathcal{H}^s$.*
3. **Size of Ψ:** *the functional $\Psi : \mathcal{H}^s \to \mathbb{R}$ satisfies the growth conditions*

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

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

$$\|\nabla\Psi(x)\|_{-s} \lesssim \|x\|_s^{\varsigma} \vee \|x\|_s \qquad \text{and} \qquad \|\partial^2\Psi(x)\|_{\mathcal{L}(\mathcal{H}^s;\mathcal{H}^{-s})} \lesssim 1, \tag{3.4}$$

for some $1/2 \le \varsigma < 1$.

Remark 3.2. Regarding the first of Assumptions 3.1, the condition $\kappa > \frac{1}{2}$ ensures that $\mathrm{Trace}_{\mathcal{H}^s}(\mathcal{C}_s) < \infty$ for any $0 \le s < \kappa - \frac{1}{2}$; this implies that $\pi_0(\mathcal{H}^s) = 1$ for any $0 \le s < \kappa - \frac{1}{2}$. As for the first of the requirements in (3.4), this is slightly less general than the corresponding condition imposed in [MPS12] (there it is required that $\|\nabla\Psi(x)\|_{-s} \lesssim 1 + \|x\|_s$). This is to avoid excessive technicalities (particularly in the proof of (8.10), which is the only place where this simplification is actually used, see Remark 8.12 and Remark 8.9 on this point). □

Example 3.3. The functional $\Psi(x) = \frac{1}{2}\|x\|_s^2$ is defined on $\mathcal{H}^s$ and its derivative at $x \in \mathcal{H}^s$ is given by $\nabla\Psi(x) = \sum_{j\ge 0} j^{2s} x^j \phi_j \in \mathcal{H}^{-s}$ with $\|\nabla\Psi(x)\|_{-s} = \|x\|_s$. The second derivative $\partial^2\Psi(x) \in \mathcal{L}(\mathcal{H}^s, \mathcal{H}^{-s})$ is the linear operator that maps $u \in \mathcal{H}^s$ to $\sum_{j\ge 0} j^{2s}\langle u, \phi_j\rangle\phi_j \in \mathcal{H}^{-s}$: its norm satisfies $\|\partial^2\Psi(x)\|_{\mathcal{L}(\mathcal{H}^s,\mathcal{H}^{-s})} = 1$ for any $x \in \mathcal{H}^s$. □

The Assumptions 3.1 ensure that the functional Ψ behaves well in a sense made precise in the following lemma. We set

$$F(z) = -z - \mathcal{C}\nabla\Psi(z). \tag{3.5}$$

Lemma 3.4. *Let Assumptions 3.1 hold.*

1. *The function $\mathcal{C}\nabla\Psi(z)$ is globally Lipshitz on $\mathcal{H}^s$ and hence the same holds for the function $F(z)$:*

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

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

$$\left|\Psi(y) - \Psi(x) - \langle\nabla\Psi(x), y - x\rangle\right| \lesssim \|y - x\|_s^2 \qquad \forall x, y \in \mathcal{H}^s. \tag{3.6}$$

Proof. See [MPS12]. □

We would also like to recall that because of our assumptions on the covariance operator, for all $p \ge 0$ there is a constant $c = c(p)$ such that

$$\mathbb{E}\|(\mathcal{C}_N)^{1/2}\xi^N\|_s^p \le c, \quad \text{uniformly in } N, \tag{3.7}$$

if ξ^N is the Gaussian defined in (1.6). We will prove this inequality in Appendix A. For the moment we just stress that $c > 0$ is a constant independent of N but that does depend on p.

4. Existence And Uniqueness For the Limiting SDE

The main statements of this section are Theorem 4.1, Theorem 4.3 and Theorem 4.6. In Theorem 4.1 and Theorem 4.3 we prove existence and uniqueness for the solution to equation (1.11) and equation (1.10), respectively. Theorem 4.6 is a "continuous mapping" result and it is crucial for the arguments of Section 6 (and, ultimately, it is the backbone of the proof of our main results).

Theorem 4.1. *For any initial datum $S(0) \in \mathbb{R}_+$, there exists a unique solution $S(t)$ to the ODE (1.11). Such a solution is strictly positive for every $t > 0$. Furthermore, $S(t)$ is bounded with continuous first derivative for all $t \ge 0$. In particular*

$$\lim_{t\to\infty} S(t) = 1 \tag{4.1}$$

and

$$0 \le \min\{S(0), 1\} \le S(t) \le \max\{S(0), 1\}, \qquad \text{for } t \ge 0. \tag{4.2}$$

Before proving the above theorem we state Lemma 4.2, which gathers all the properties of the real valued functions D_ℓ, Γ_ℓ and A_ℓ, defined in (1.12)-(1.15).

Lemma 4.2. *The functions $D_\ell(x)$, $\Gamma_\ell(x)$ and $\sqrt{\Gamma_\ell(x)}$ are positive, globally Lipshitz continuous and bounded, with bounded first derivative. $A_\ell(x)$ is bounded above but not below; it has continuous first derivative on the whole of $\mathbb{R}_+$ and it is globally Lipshitz. Moreover, for any $\ell > 0$, $A_\ell(x)$ is strictly positive for $x \in [0, 1)$, strictly negative for $x > 1$ and $A_\ell(1) = 0$.*

Proof of lemma 4.2. The proof of the above Lemma 4.2 follows from the same arguments used in [JLM15, Proof of Lemma 2]. We sketch the proof in Appendix A for completeness. A plot of the function $A_\ell(x)$ for various values of ℓ can be found in [CRR05, page 258]. Figure 1 contains a plot of $A_\ell(x)$ for $\ell = 1$ and $\ell = 2$. Plots of the functions D_ℓ, Γ_ℓ and of the derivative of A_ℓ can be found in Appendix A. □

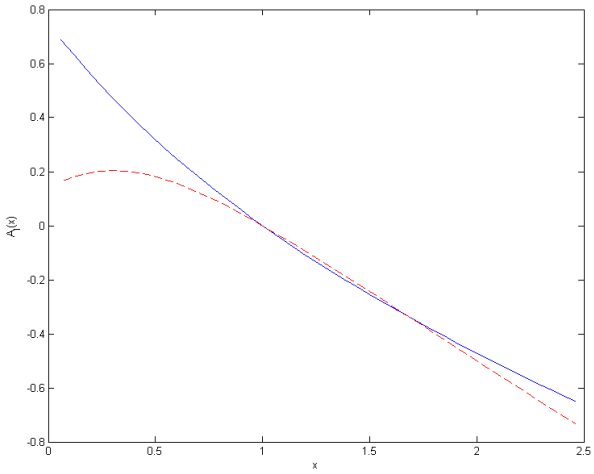

FIG 1. *Plots of the function $A_\ell(x)$ for $\ell = 1$ and $\ell = 2$ (dashed line).*

Proof of Theorem 4.1. Existence and uniqueness for (1.11) is standard, since A_ℓ is globally Lipshitz. The limit (4.1) and the bound (4.2) are a consequence of the last statement of Lemma 4.2. Indeed, if we start with an initial datum $S_0 \in [0, 1)$ then $S(t)$ will increase towards 1. If $S(0) > 1$ then $S(t)$ will decrease towards 1. □

We now come to existence and uniqueness for equation (1.10), which we rewrite as

$$dx(t) = F(x(t))D_\ell(S(t))dt + \sqrt{\Gamma_\ell(S(t))}dW(t),$$

where $W(t)$ is an $\mathcal{H}^s$ valued $\mathcal{C}_s$-Brownian Motion and the function F has been defined in (3.5). The above is a short notation for the integral form

$$x(t) = x(0) + \int_0^t F(x(v))D_\ell(S(v))dv + \int_0^t \sqrt{\Gamma_\ell(S(v))}dW(v)\,. \tag{4.3}$$

In view of the next statement we emphasize that throughout the paper the spaces $C([0, T]; \mathcal{H}^s)$ and $C([0, T]; \mathbb{R}_+)$ are endowed with the uniform topology.

Theorem 4.3. *Let Assumption 3.1 hold. Then, for any initial condition $x(0) \in \mathcal{H}^s$, any $T > 0$ and every $\mathcal{H}^s$-valued $\mathcal{C}_s$-Brownian motion $W(t)$, there exists a unique solution of equation (1.10) (with $S(t)$ given by (1.11)) in the space $C([0, T]; \mathcal{H}^s)$.*

Before proving the above theorem, let us make a remark on the statement.

Remark 4.4. In the statement of Theorem 4.3 we refer to equation (1.10) or, equivalently, to equation (4.3). With the notation introduced so far, the function $S(t)$ appearing in (4.3) is the solution of the ODE (1.11). However the proof of Theorem 4.3 is still valid if $S(t) : \mathbb{R}_+ \to \mathbb{R}_+$ is any continuous and bounded function with continuous first derivative. □

Proof of Theorem 4.3. Once we prove continuity of the map

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

where $x(t)$ is defined by (4.3), existence and uniqueness for equation (4.3) for a small enough time interval follow from a standard contraction mapping argument, see e.g. [MPS12]. To show the continuity of such a map, let $x^\sharp(t)$ and $x^\dagger(t)$ be the images through the map $\mathcal{J}$ of the pairs $(x^\sharp(0), W^\sharp(t))$ and $(x^\dagger(0), W^\dagger(t))$, respectively. For all $0 \le t \le T$, we then have

$$\begin{aligned} \|x^\sharp(t) - x^\dagger(t)\|_s \le \int_0^t \|F(x_v^\sharp)D_\ell(S_v) - F(x_v^\dagger)D_\ell(S_v)\|_s dv \\ + \|x^\sharp(0) - x^\dagger(0)\|_s + \left\| \int_0^t \sqrt{\Gamma_\ell(S_v)}(dW_v^\sharp - dW_v^\dagger) \right\|_s . \end{aligned}$$

Thanks to the Lipshitzianity of F, Lemma 3.4, and the boundedness of D_ℓ, Lemma 4.2, the drift coefficient of (4.3), i.e.

$$\Theta(x,S) := F(x)D_\ell(S), \quad (x,S) \in \mathcal{H}^s \times \mathbb{R}_+,$$

is globally Lipshitz, uniformly in time. Therefore, integrating by parts in the stochastic integral, we get

$$\begin{aligned} \|x^\sharp(t) - x^\dagger(t)\|_s \lesssim \int_0^t \|x^\sharp(v) - x^\dagger(v)\|_s + \|x^\sharp(0) - x^\dagger(0)\|_s \\ + \left\| \Gamma_\ell^{1/2}(S_t)(W_t^\sharp - W_t^\dagger) - \int_0^t \frac{d}{dv}\left(\Gamma_\ell^{1/2}(S_v)\right)(W_v^\sharp - W_v^\dagger) \right\|_s . \end{aligned}$$

We now further work on the right hand side of the above as follows:

$$\left\| \int_0^t \frac{d}{dv}\left(\Gamma_\ell^{1/2}(S_v)\right)(W_v^\sharp - W_v^\dagger) \right\|_s \le \sup_{v\in[0,t]} \|W_v^\sharp - W_v^\dagger\|_s \int_0^t \left| \frac{d}{dv}\left(\Gamma_\ell^{1/2}(S_v)\right) \right| .$$

Clearly,

$$\left| \frac{d}{dv}\left(\Gamma_\ell^{1/2}(S_v)\right) \right| = \frac{1}{2} \left| \frac{\frac{d}{dx}\Gamma_\ell(x)}{\Gamma_\ell^{1/2}(x)} \Big|_{x=S_v} (A_\ell(S_v)) \right| .$$

From the definition of Γ_ℓ (see (1.13) and (1.15)), for any $x \ge 0$, Γ_ℓ is bounded below away from zero. Moreover, Γ_ℓ has bounded derivative (see Lemma 4.2) and A_ℓ, being continuous, is bounded on compacts. These facts, together with (4.2), imply the bound

$$\int_0^t \left| \frac{d}{dv}\left(\Gamma_\ell^{1/2}(S_v)\right) \right| \lesssim t, \tag{4.4}$$

hence

$$\|x^\sharp(t) - x^\dagger(t)\|_s \lesssim \int_0^T \sup_{v\in[0,t]} \|x^\sharp(v) - x^\dagger(v)\|_s + \|x^\sharp(0) - x^\dagger(0)\|_s + T \sup_{t\in[0,T]} \|W_t^\sharp - W_t^\dagger\|_s . \tag{4.5}$$

Taking the supremum over $t \in [0,T]$ on the left hand side of the above gives the desired contractivity, thanks to the Gronwall Lemma, for a small enough time interval, say $[0,T_0]$ and hence a unique solution can be constructed for $t \in [0,T_0]$. Such a solution can then be extended to $t \ge 0$, thanks to the specific form of (4.5), which, we stress again, is a consequence of (4.2). Indeed, thanks to the fact that the drift of the equation is Lipshitz uniformly in time and to (4.4), the time dependence of the RHS of (4.5) will stay the same when we try and construct a solution starting from T_0. We will therefore be able to construct a solution over the interval $[T_0, 2T_0]$. Continuing inductively we can cover the whole real axis. This concludes the proof. □

Consider now the following equation

$$dz(t) = [-z(t) - \mathcal{C}\nabla\Psi(z(t))]D_\ell(S(t))\,dt + d\eta(t), \tag{4.6}$$

where $S(t)$ is the solution of (1.11) and $\eta(t)$ is any time-continuous function taking values in $\mathcal{H}^s$. Also, let $\mathfrak{S}(t) : \mathbb{R}_+ \to \mathbb{R}$ be the solution of

$$d\mathfrak{S}(t) = A_\ell(\mathfrak{S}(t))\,dt + a\,dw(t), \tag{4.7}$$

where $w(t)$ is a real valued standard Brownian motion and $a \in \mathbb{R}_+$ is a constant.

Remarks 4.5. Before stating the next theorem we need to be more precise about equations (4.6) and (4.7).

- We consider equation (4.7), which is (1.11) perturbed by noise, in view of the contraction mapping argument (explained in Section 6) that we will use to prove our main results. Observe that (4.7) still admits a unique solution (by the Lipshitzianity of A_ℓ). Analogous observations hold for equation (4.6), which has the same structure as equation (1.10).
- The solution to (4.7) might not stay positive if started from a positive initial datum (as opposed to the solution to (1.11), which preserves positivity). However A_ℓ is only defined for positive arguments, (see (1.14)). To make sense of the notation in (4.7), we extend A_ℓ to the negative semiaxis. In other words, the function A_ℓ appearing in (4.7) is not the same A_ℓ defined in (1.14); we should use a different notation for such a function but we refrain from doing so for simplicity. In conclusion, the function $A_\ell(s)$ in (4.7) is intended to be a strictly positive function for any $\mathbb{R} \ni s < 1$ and we fix it equal to 1 if $s \leq -1/2$; it smoothly interpolates between -1/2 and $A_\ell(0)$ if $-1/2 < s < 0$ and it coincides with $A_\ell(s)$ as defined in (1.14) if $s \geq 0$. Therefore such an A_ℓ will still be globally Lipshitz.
- We emphasize that (4.6) and (4.7) are decoupled as the function $S(t)$ appearing in (4.6) is the solution of (1.11). This fact will be particularly relevant in the remainder of this section as well as in Section 6.1 and Section 6.2.

□

The statement of the following theorem is crucial to the proof of the main results of this paper, Theorem 5.1 and Theorem 5.4, stated in the next section.

Theorem 4.6. *With the notation introduced so far (and in particular with the clarifications of Remarks 4.5) let $z(t)$ and $\mathfrak{S}(t)$ be solutions of (4.6) and (4.7), respectively. Then under Assumption 3.1 the maps*

$$\begin{aligned} \mathcal{J}_1 : \mathcal{H}^s \times C([0,T];\mathcal{H}^s) &\longrightarrow C([0,T];\mathcal{H}^s \times \mathbb{R}) \\ (z_0, \eta(t)) &\longrightarrow z(t) \end{aligned}$$

and

$$\begin{aligned} \mathcal{J}_2 : \mathbb{R}_+ \times C([0,T];\mathbb{R}) &\longrightarrow C([0,T];\mathbb{R}) \\ (\mathfrak{S}_0, w(t)) &\longrightarrow \mathfrak{S}(t) \end{aligned}$$

are continuous maps.

Proof. Continuity of the map $\mathcal{J}_1$ can be shown with a calculation in the same spirit of the one done for the map $\mathcal{J}$, so we only sketch the proof of the continuity of the map $\mathcal{J}_2$. To this end we will use (4.2) and the Lipshitzianity of A_ℓ. Let $\mathfrak{S}^\sharp(t)$ and $\mathfrak{S}^\dagger(t)$ be the images through the map $\mathcal{J}_2$ of the pairs $(\mathfrak{S}_0^\sharp, w^\sharp(t))$ and $(\mathfrak{S}_0^\dagger, w^\dagger(t))$, respectively. Then

$$\begin{aligned} \left|\mathfrak{S}_t^\sharp - \mathfrak{S}_t^\dagger\right| &\leq \left|\mathfrak{S}_0^\sharp - \mathfrak{S}_0^\dagger\right| + \int_0^t \left|A_\ell(\mathfrak{S}_v^\sharp) - A_\ell(\mathfrak{S}_v^\dagger)\right| dv + a\left|w_t^\sharp - w_t^\dagger\right| \\ &\lesssim \left|\mathfrak{S}_0^\sharp - \mathfrak{S}_0^\dagger\right| + \int_0^t \left|\mathfrak{S}_v^\sharp - \mathfrak{S}_v^\dagger\right| dv + a\left|w_t^\sharp - w_t^\dagger\right|. \end{aligned} \tag{4.8}$$

Now we can conclude by Gronwall's Lemma. □

5. Statement of Main Theorems and Heuristics of Proofs

In this section we give a precise statement of the main results of the paper, Theorem 5.1 and Theorem 5.4 below, and outline the heuristic arguments which are at the basis of the proof of such results. The rigorous proofs of Theorem 5.1 and Theorem 5.4 are detailed in Section 7 and Section 8, respectively, and they consist in quantifying the formal approximations presented in this section. The structure of such proofs relies on the continuous mapping argument which is presented in Section 6.

While describing the main intuitive ideas of the proof, we will also try and emphasize the differences with the analysis presented in [MPS12] in the stationary case. Here and throughout the paper we will use a notation analogous to the one used in [MPS12].

5.1. Statement of Main Results

Let us define the set $\mathcal{H}^s_\cap$ as follows:

$$\mathcal{H}^s_\cap := \left\{ x \in \mathcal{H}^s : \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \frac{\left|x^i\right|^2}{\lambda_i^2} < \infty \right\}. \tag{5.1}$$

Theorem 5.1. *Let Assumption 3.1 hold and let $x_0 \in \mathcal{H}^s_\cap$. Let $\{S^N_k\} \subset \mathbb{R}_+$ be the double sequence defined in (1.16) and started at $S^N_0 = \frac{1}{N}\sum_{i=1}^N \left|x_0^{i,N}\right|^2 / \lambda_i^2$. Let $S^{(N)}(t)$, defined in (1.17), be the continuous interpolant of S^N_k. Then, as $N \to \infty$, $S^{(N)}(t)$ converges weakly in $C([0,T];\mathbb{R})$ to the solution $S(t)$ of the ODE (1.11) started at $S_0 := \lim_{N\to\infty} S^N_0$.*

We will prove Theorem 5.1 in Section 7. For the time being, let us make the following observations.

Remark 5.2. Notice that the weak limit of the double sequence S^N_k is a deterministic function, therefore the above theorem also implies convergence in probability in $C([0,T];\mathbb{R})$ of $S^{(N)}(t)$ to $S(t)$. □

Let us now introduce the piecewise constant interpolant of the (double) sequence S^N_k, i.e. the (sequence of) functions $\bar{S}^{(N)}(t)$ defined as follows:

$$\bar{S}^{(N)}(t) = S^N_k, \quad \text{for } \; t_k \le t < t_{k+1}, \; t_k = k/N\,. \tag{5.2}$$

Lemma 5.3. *Under the assumptions of Theorem 5.1, for every fixed $t > 0$,*

$$S^{(N)}(t) \to S(t) \quad \textit{almost surely}$$

and

$$\bar{S}^{(N)}(t) \to S^{(N)}(t) \quad \textit{almost surely.}$$

Therefore,

$$\bar{S}^{(N)}(t) \to S(t) \quad \textit{almost surely.}$$

Proof of Lemma 5.3. The proof of this lemma can be found in Appendix B. □

Consider now the set $\mathcal{H}^s_{\cap\cap}$ defined as the set of $x \in \mathcal{H}^s_\cap$ such that

- for all $p \ge 0$,

$$\lim_{N\to\infty} \sum_{i=1}^{N} i^{2s} \lambda_i^2 \frac{\left|x^i\right|^{2p}}{\lambda_i^{2p}} < \infty, \tag{5.3}$$

- there exists some $\epsilon > 0$, such that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \frac{\left|x^i\right|^2}{\lambda_i^2} \ge \epsilon > 0. \tag{5.4}$$

Theorem 5.4. *Let Assumption 3.1 hold and $x_0 \in \mathcal{H}^s_{\cap\cap}$. Then, as $N \to \infty$ the continuous interpolant $x^{(N)}(t)$ of the chain $\{x_k\}_k \subset \mathcal{H}^s$ (defined in (1.8) and (2.4), respectively) and started at x_0, converges weakly in $C([0,T];\mathcal{H}^s)$ to the solution $x(t)$ of equation (1.10) started at 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 $S_0 := \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N \left|x_0^i\right|^2/\lambda_i^2$.*

We will prove Theorem 5.4 in Section 8. Note that in the above statement we are picking a deterministic initial condition. However it is worth noting that $x_0 \in \mathcal{H}^s_{\cap\cap}$ almost surely if x_0 is drawn at random from the stationary measure (1.1) . We will make some remarks on condition (5.4) at the end of Subsection 5.2. As for condition (5.3), strictly speaking this does not need to be satisfied for all $p \geq 0$; a finite, sufficiently large p would suffice. However we refrain from determining the optimal p, which would distract from the main goals of the paper, and we state the result as it is, based on (5.3).

5.2. Formal Analysis of the Acceptance Probability

Gaining an intuition about the behaviour of the acceptance probability $\alpha(x,\xi)$, defined in (2.1), is at the core of the proof of the main result of this paper, Theorem 5.4. We present here a formal calculation that helps impart such intuition. We stress again that the calculations of this section are purely formal and will be made rigorous from Section 7 on . In this spirit we will use the loose notation $A^N \simeq B^N$ when, for N large, A^N is "approximately equal" to B^N, and $A^N \approx B^N$ when, for N large, A^N is "approximately distributed" according to B^N.

Let us recall the notation $\Psi^N := \Psi \circ \mathcal{P}^N$ (that is, $\Psi^N(x) := \Psi(\mathcal{P}^N(x))$) and set

$$\zeta_k^N := (\mathcal{C}_N)^{-1/2} x_k^N + (\mathcal{C}_N)^{1/2} \nabla\Psi^N(x_k), \quad \text{where } \nabla\Psi^N(x_k) = \mathcal{P}^N(\nabla\Psi(x_k^N)). \tag{5.5}$$

With these definitions, we can further rewrite the expression (2.5) for $Q(x_k^N, \xi_{k+1}^N)$:

$$\begin{aligned}
Q(x_k,\xi_{k+1}) &= -\frac{\ell^2}{N}\sum_{i=1}^N \left|\xi_{k+1}^{i,N}\right|^2 - \sqrt{\frac{2\ell^2}{N}}\sum_{i=1}^N \frac{x_k^{i,N}\xi_{k+1}^{i,N}}{\lambda_i} + \Psi(x_k^N) - \Psi(y_{k+1}^N) \\
&= -\frac{\ell^2}{N}\sum_{i=1}^N \left|\xi_{k+1}^{i,N}\right|^2 - \sqrt{\frac{2\ell^2}{N}}\langle \mathcal{C}^{-1/2} x_k^N, \xi_{k+1}^N\rangle + \Psi(x_k^N) - \Psi(y_{k+1}^N) \\
&= -\frac{\ell^2}{N}\|\xi_{k+1}^N\|^2 - \sqrt{\frac{2\ell^2}{N}}\langle \zeta_k^N, \xi_{k+1}^N\rangle + \Psi(x_k^N) - \Psi(y_{k+1}^N) + \sqrt{\frac{2\ell^2}{N}}\langle \mathcal{C}_N^{1/2}\nabla\Psi^N(x_k^N), \xi_{k+1}^N\rangle.
\end{aligned}$$

Therefore setting

$$r^N(x_k,\xi_{k+1}) := \Psi(x_k^N) - \Psi(y_{k+1}^N) + \sqrt{\frac{2\ell^2}{N}}\langle \mathcal{C}_N^{1/2}\nabla\Psi^N(x_k^N), \xi_{k+1}^N\rangle$$

and

$$R(x_k^N,\xi_{k+1}^N) := -\frac{\ell^2}{N}\sum_{i=1}^N \left|\xi_{k+1}^{i,N}\right|^2 - \sqrt{\frac{2\ell^2}{N}}\langle \zeta_k^N, \xi_{k+1}^N\rangle, \tag{5.6}$$

we obtain

$$Q(x_k^N,\xi_{k+1}^N) = R(x_k^N,\xi_{k+1}^N) + r^N(x_k,\xi_{k+1})\,. \tag{5.7}$$

In [MPS12] it is shown that

$$\left|r^N(x_k,\xi)\right| \lesssim \frac{\|\mathcal{C}^{1/2}\xi\|_s^2}{N}\,; \tag{5.8}$$

therefore

$$\mathbb{E}\left|r^N(x_k,\xi_{k+1})\right| \lesssim \frac{1}{N}, \tag{5.9}$$

see [MPS12, eqn. (2.32)]. The above (5.8)-(5.9) are true whether the chain is started in stationarity or not, as they are only a consequence of the properties of Ψ (see (3.6)) and of the noise ξ_{k+1}, see (3.7). Using (5.9),

$$Q(x_k^N,\xi_{k+1}^N) \simeq R(x_k^N,\xi_{k+1}^N). \tag{5.10}$$

Looking at the definition of R, equation (5.6), and observing that by the Law of Large Numbers

$$\frac{1}{N}\sum_{j=1}^{N}\left|\xi_{k+1}^{j,N}\right|^2 \longrightarrow 1\,, \tag{5.11}$$

we deduce that $R \simeq G$ (see Lemma 8.1), where

$$G := -\ell^2 - \sqrt{\frac{2\ell^2}{N}}\sum_{j=1}^{N}\zeta_k^{j,N}\xi_{k+1}^{j,N}, \quad \text{so that, given } x_k, \quad G \sim \mathcal{N}\left(-\ell^2, \frac{2\ell^2}{N}\sum_{j=1}^{N}\left|\zeta_k^{j,N}\right|^2\right). \tag{5.12}$$

We will show

$$\frac{1}{N}\sum_{j=1}^{N}\left|\zeta_k^{j,N}\right|^2 \simeq \frac{1}{N}\sum_{j=1}^{N}\frac{\left|x_k^{j,N}\right|^2}{\lambda_j^2} = S_k^N.$$

This can be intuitively understood by observing that in (5.5) the "dominating contribution" comes from the first addend. The above approximation is formalized by (8.51) and (7.3) and it implies $G \approx Z_{\ell,k}$, where

$$Z_{\ell,k} := -\ell^2 - \sqrt{\frac{2\ell^2}{N}}\sum_{j=1}^{N}\frac{x_k^{j,N}}{\lambda_j}\xi_{k+1}^{j,N} \quad \text{so that, given } x_k^N, \quad Z_{\ell,k} \sim \mathcal{N}(-\ell^2, 2\ell^2 S_k^N). \tag{5.13}$$

In conclusion, the formal analysis presented so far suggests that we may use the approximations

$$Q(x_k^N, \xi_{k+1}^N) \simeq R \approx \mathcal{N}\left(-\ell^2, 2\ell^2\, S_k^N\right). \tag{5.14}$$

In [MPS12] it is proved that if we start from stationarity then the sequence S_k^N converges (for fixed k, as $N \to \infty$) to 1 almost surely (see (1.22)). We will show that if we start the chain out of stationarity, i.e. x_0 is any point in $\mathcal{H}^s$, then

$$S_k^N = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|x_k^{i,N}\right|^2}{\lambda_i^2} \xrightarrow{d} S(t), \quad \text{as } N \to \infty, \ \text{ for } t_k \le t < t_{k+1}, \tag{5.15}$$

where $t_k = k/N$ and $S(t)$ is the solution of the ODE (1.11). This is the main conceptual difference between our work and [MPS12], all the other differences are technical consequences of this fact.

Looking at (5.14)-(5.15), we can explain why we are assuming (5.4): roughly speaking, if the initial datum S_0 is strictly positive then the limit $S(t)$ is strictly positive for every $t \ge 0$, so the Gaussian variable on the RHS of (5.14) always has a strictly positive variance. If instead $S_0 = 0$, then at zero one would have $Q_0 = Q(x_0, \xi_1) \simeq -\ell^2$ and therefore the acceptance probability at the first step simply tends to $e^{-\ell^2}$; however this would only be true at zero as, even if $S_0 = 0$, the solution of the ODE (1.11) becomes immediately strictly positive for $t > 0$ (see Theorem 4.1). To avoid having to take into account also this further possibility (which does not add anything to the overall understanding of the algorithm), and to streamline the analysis, we make the simplifying assumption (5.4).

The approximation (5.14) dictates the behaviour of the acceptance probability. With the present algorithm the average acceptance probability does not tend to one (as $N \to \infty$, for $t_k \le t < t_{k+1}$). This is one of the disadvantages of using the method analysed in this paper, in comparison to using algorithms which are well defined in infinite dimensions.

5.3. Formal Derivation of the Drift Coefficient of Equation (1.10)

Let us first clarify the use of the notation that we will make in the following. The definition of x_{k+1} (2.3) contains two sources of randomness: the Gaussian noise ξ_{k+1} and the Bernoulli random variable γ_{k+1}. With this in mind, when we write $\mathbb{E}_k(\cdot)$ we will mean expectation with respect to ξ_{k+1} and γ_{k+1}, given x_k. In some

cases, when we want to emphasize the fact that the expectation is taken with respect to ξ_{k+1} and γ_{k+1}, we will write explicitly $\mathbb{E}_k^{\xi,\gamma}$. In the same way, if we want to stress that expectation is being taken with respect to ξ_{k+1}, we write $\mathbb{E}_k^{\xi}$. According to (2.4), the i-th component of the approximate drift is given by

$$\begin{aligned} N\mathbb{E}_k(x_{k+1}^{i,N} - x_k^{i,N}) &= N\mathbb{E}_k\left(\gamma_{k+1}\sqrt{\frac{2\ell^2}{N}}\lambda_i\,\xi_{k+1}^{i,N}\right) = \sqrt{2N\ell^2}\lambda_i\,\mathbb{E}_k^{\xi,\gamma}(\gamma_{k+1}\,\xi_{k+1}^{i,N}) \\ &= \sqrt{2N\ell^2}\lambda_i\,\mathbb{E}_k^{\xi}(\alpha_{k+1}\,\xi_{k+1}^{i,N}) = \sqrt{2N\ell^2}\lambda_i\,\mathbb{E}_k^{\xi}\left[\left(1\wedge e^{Q(x_k^N,\xi_{k+1}^N)}\right)\xi_{k+1}^{i,N}\right]. \end{aligned} \tag{5.16}$$

(We briefly explain at the end of Appendix A how the first equality in (5.16) is obtained.) For a reason that will be clear in a few lines we further split the RHS of (5.6) as follows [4]

$$\begin{aligned} R(x_k^N,\xi_{k+1}^N) &= -\frac{\ell^2}{N}\sum_{j\neq i}^{N}\left|\xi_{k+1}^{j,N}\right|^2 - \sqrt{\frac{2\ell^2}{N}}\sum_{j\neq i}\zeta_k^{j,N}\xi_{k+1}^{j,N} - \frac{\ell^2}{N}\left|\xi_{k+1}^{i,N}\right|^2 - \sqrt{\frac{2\ell^2}{N}}\zeta_k^{i,N}\xi_{k+1}^{i,N} \\ &=: R^i(x_k^N,\xi_{k+1}^N) - \frac{\ell^2}{N}\left|\xi_{k+1}^{i,N}\right|^2 - \sqrt{\frac{2\ell^2}{N}}\zeta_k^{i,N}\xi_{k+1}^{i,N}, \end{aligned} \tag{5.17}$$

hence

$$Q(x_k^N,\xi_{k+1}^N) \simeq R^i(x_k^N,\xi_{k+1}^N) - \sqrt{\frac{2\ell^2}{N}}\zeta_k^{i,N}\xi_{k+1}^{i,N}. \tag{5.18}$$

Using (5.18) we then have

$$\mathbb{E}_k^{\xi}\left[\left(1\wedge e^{Q(x_k^N,\xi_{k+1}^N)}\right)\xi_{k+1}^{i,N}\right] \simeq \mathbb{E}_k^{\xi}\left[\left(1\wedge e^{R^i(x_k^N,\xi_{k+1}^N)-\sqrt{\frac{2\ell^2}{N}}\zeta_k^{i,N}\xi_{k+1}^{i}}\right)\xi_{k+1}^{i,N}\right]. \tag{5.19}$$

We now use [MPS12, eqn. (2.36)], which we recast here for the reader's convenience.

Lemma 5.5. *Let X be a real valued r.v., $X\sim\mathcal{N}(0,1)$. Then for any $a,b\in\mathbb{R}$,*

$$\mathbb{E}\left[X\left(1\wedge e^{aX+b}\right)\right] = ae^{\frac{a^2}{2}+b}\,\Phi\left(-\frac{b}{|a|}-|a|\right). \tag{5.20}$$

Proof. See [MPS12, Lemma 2.4]. □

Now notice that, given x_k, R^i is independent of ξ_{k+1}^i as it only contains the random variables ξ_{k+1}^j for $i\neq j$. Therefore the expected value $\mathbb{E}_k^{\xi}$ can be calculated by first evaluating $\mathbb{E}_k^{\xi^i}$ and then $\mathbb{E}_k^{\xi^i_-}$, where the latter denotes expectation with respect to $\xi\backslash\xi^i$. With this observation we can use the above Lemma 5.5 with

[4] This splitting is standard in the analysis of high dimensional MCMC, see [MPS12].

$a = -\sqrt{\frac{2\ell^2}{N}}\zeta_k^{i,N}$ and $b = R^i$ to further evaluate the RHS of (5.19); we get

$$\begin{aligned}
\mathbb{E}_k^\xi\left[\left(1\wedge e^{Q(x_k^N,\xi_{k+1}^N)}\right)\xi_{k+1}^{i,N}\right] &\simeq \mathbb{E}_k^\xi\left[\left(1\wedge e^{R^i(x_k^N,\xi_{k+1}^N)-\sqrt{\frac{2\ell^2}{N}}\zeta_k^{i,N}\xi_{k+1}^i}\right)\xi_{k+1}^{i,N}\right] \\
&\overset{(5.20)}{=} -\sqrt{\frac{2\ell^2}{N}}\zeta_k^{i,N}e^{\frac{\ell^2}{N}|\zeta_k^{i,N}|^2}\mathbb{E}_k^{\xi^i_-}e^{R^i}\Phi\left(-\frac{R^i}{\sqrt{\frac{2\ell^2}{N}}\left|\zeta_k^{i,N}\right|}-\sqrt{\frac{2\ell^2}{N}}\left|\zeta_k^{i,N}\right|\right) \qquad (5.21)\\
&\simeq -\sqrt{\frac{2\ell^2}{N}}\zeta_k^{i,N}\,\mathbb{E}_k^{\xi^i_-}e^{R^i}\Phi\left(\frac{-R^i}{\sqrt{\frac{2\ell^2}{N}}\left|\zeta_k^{i,N}\right|}\right)\\
&= -\sqrt{\frac{2\ell^2}{N}}\zeta_k^{i,N}\,\mathbb{E}_k^{\xi}e^{R^i}\Phi\left(\frac{-R^i}{\sqrt{\frac{2\ell^2}{N}}\left|\zeta_k^{i,N}\right|}\right)\\
&\simeq -\sqrt{\frac{2\ell^2}{N}}\zeta_k^{i,N}\,\mathbb{E}_k^{\xi}e^{R^i}\mathbf{1}_{\{R^i<0\}} \qquad (5.22)\\
&\simeq -\sqrt{\frac{2\ell^2}{N}}\zeta_k^{i,N}\,\mathbb{E}_k^{\xi}e^{R}\mathbf{1}_{\{R<0\}}\,.
\end{aligned}$$

Therefore, using the approximation (5.14) (and the notation (5.13)),

$$\mathbb{E}_k^\xi\left[\left(1\wedge e^{Q(x_k^N,\xi_{k+1}^N)}\right)\xi_{k+1}^{i}\right] \simeq -\sqrt{\frac{2\ell^2}{N}}\zeta_k^{i,N}\,\mathbb{E}_k^{\xi}e^{Z_{\ell,k}}\mathbf{1}_{\{Z_{\ell,k}<0\}}\,. \qquad (5.23)$$

Now a straightforward calculation shows that if $X\sim\mathcal{N}(\mu,\sigma^2)$ then

$$\mathbb{E}\left(e^X\mathbf{1}_{X<0}\right) = e^{\mu+\sigma^2/2}\,\Phi\left(-\frac{\mu}{\sigma}-\sigma\right).$$

In particular this means that if $X\sim\mathcal{N}(-\ell^2,2\ell^2 a)$, for some $a>0$, then

$$\mathbb{E}\left(e^X\mathbf{1}_{X<0}\right) = e^{\ell^2(a-1)}\Phi\left(\frac{\ell(1-2a)}{\sqrt{2a}}\right) = \frac{1}{2\ell^2}D_\ell(a). \qquad (5.24)$$

From (5.24), (5.23) and (5.13), we then get

$$\mathbb{E}_k^\xi\left[\left(1\wedge e^{Q(x_k^N,\xi_{k+1}^N)}\right)\xi_{k+1}^{i,N}\right] \simeq -\sqrt{\frac{2\ell^2}{N}}\zeta_k^{i,N}\frac{1}{2\ell^2}D_\ell(S_k^N) = -\frac{1}{\sqrt{2\ell^2 N}}\zeta_k^{i,N}D_\ell(S_k^N)\,.$$

Combining the above with (5.16) gives

$$N\mathbb{E}_k^\xi(x_{k+1}^{i,N}-x_k^{i,N}) \simeq -\lambda_i\zeta_k^{i,N}\,D_\ell(S_k^N),$$

which is the desired drift, after observing that $\lambda_i\zeta_k^{i,N}$ is the i-th component of $\mathcal{C}_N^{1/2}\zeta_k^N$ and

$$\mathcal{C}_N^{1/2}\zeta_k^N = x_k^N + \mathcal{C}_N\,\nabla\Psi^N(x_k).$$

As already mentioned in the introduction, as a consequence of (1.22), if we started the chain in stationarity then the approximate drift would not be time dependent and we would have

$$N\mathbb{E}_k^\xi(x_{k+1}^{i,N}-x_k^{i,N}) \simeq -\lambda_i\zeta_k^{i,N}\,D_\ell(1),$$

which is the approximate drift of (1.18).

5.4. *Formal Derivation of the Diffusion Coefficient of Equation* (1.10)

$$N\mathbb{E}_k(x_{k+1}^{i,N}-x_k^{i,N})(x_{k+1}^{j,N}-x_k^{j,N}) = N\mathbb{E}_k^{\xi,\gamma}\left(\gamma_{k+1}\sqrt{\frac{2\ell^2}{N}}\lambda_i\xi_{k+1}^{i,N}\right)\left(\gamma_{k+1}\sqrt{\frac{2\ell^2}{N}}\lambda_j\xi_{k+1}^{j,N}\right)$$
$$= 2\ell^2\lambda_i\,\lambda_j\mathbb{E}_k^{\xi}\left(\xi_{k+1}^i\,\xi_{k+1}^j\,\left(1\wedge e^{Q(x_k,\xi_{k+1})}\right)\right), \tag{5.25}$$

where the last equality follows analogously to (5.16). We consider (5.6) as before, but this time we split

$$R(x_k^N,\xi_{k+1}^N) = R^{ij}(x_k^N,\xi_{k+1}^N) - \frac{\ell^2}{N}\left(\left|\xi_{k+1}^{i,N}\right|^2+\left|\xi_{k+1}^{j,N}\right|^2\right) - \sqrt{\frac{2\ell^2}{N}}\left(\zeta_k^{i,N}\xi_{k+1}^{i,N}+\zeta_k^{j,N}\xi_{k+1}^{j,N}\right),$$

where

$$R^{ij}(x_k^N,\xi_{k+1}^N) = -\frac{\ell^2}{N}\sum_{h\neq i,j}^{N}\left|\xi_{k+1}^{h,N}\right|^2 - \sqrt{\frac{2\ell^2}{N}}\sum_{h\neq i,j}\zeta_k^{h,N}\xi_{k+1}^{h,N}.$$

As before, $Q(x_k^N,\xi_{k+1}^N)\simeq R^{ij}(x_k^N,\xi_{k+1}^N)$, so that

$$\mathbb{E}_k^{\xi}\left(\xi_{k+1}^{i,N}\,\xi_{k+1}^{j,N}\,\left(1\wedge e^{Q(x_k^N,\xi_{k+1}^N)}\right)\right) \simeq \mathbb{E}_k^{\xi}\left(\xi_{k+1}^{i,N}\,\xi_{k+1}^{j,N}\,\left(1\wedge e^{R^{ij}(x_k^N,\xi_{k+1}^N)}\right)\right)$$
$$= \delta_{ij}\mathbb{E}_k^{\xi^{ij}}\left(1\wedge e^{R^{ij}(x_k^N,\xi_{k+1}^N)}\right)$$
$$= \delta_{ij}\mathbb{E}_k^{\xi}\left(1\wedge e^{R^{ij}(x_k^N,\xi_{k+1}^N)}\right). \tag{5.26}$$

With the same reasoning as in (5.14), we have

$$Q(x_k^N,\xi_{k+1}^N)\simeq R^{ij}\approx\mathcal{N}\left(-\ell^2,2\ell^2\,S_k^N\right).$$

(Again, if we were to consider the stationary regime, then we would have $Q(x_k^N,\xi_{k+1}^N)\approx\mathcal{N}\left(-\ell^2,2\ell^2\right)$.) Now a simple calculation shows that if $X\sim\mathcal{N}(\mu,\sigma^2)$ then

$$\mathbb{E}\left(1\wedge e^X\right) = e^{\mu+\sigma^2/2}\,\Phi\left(-\frac{\mu}{\sigma}-\sigma\right)+\Phi\left(\frac{\mu}{\sigma}\right) \tag{5.27}$$

and in particular if $X\sim\mathcal{N}(-\ell^2,2\ell^2a)$ for some $a>0$,

$$\mathbb{E}(1\wedge e^X) = \frac{1}{2\ell^2}\Gamma_\ell(a). \tag{5.28}$$

Hence

$$\mathbb{E}_k^{\xi}\left(1\wedge e^{R^{ij}(x_k^N,\xi_{k+1}^N)}\right)\simeq\frac{1}{2\ell^2}\Gamma_\ell(S_k^N). \tag{5.29}$$

Putting together (5.25), (5.26) and (5.29) we get

$$N\mathbb{E}_k(x_{k+1}^{i,N}-x_k^{i,N})(x_{k+1}^{j,N}-x_k^{j,N}) = \lambda_i\lambda_j\,\delta_{ij}\Gamma_\ell(S_k^N).$$

5.5. *Formal Derivation of Equation* (1.11)

We now want to describe the heuristic derivation of the limit (5.15). Let us start with the drift:

$$N\mathbb{E}_k(S_{k+1}^N-S_k^N) = \mathbb{E}_k\sum_{i=1}^{N}\left[\frac{\left|x_{k+1}^{i,N}\right|^2}{\lambda_i^2}-\frac{\left|x_k^{i,N}\right|^2}{\lambda_i^2}\right]$$
$$= \mathbb{E}_k\left[\gamma_{k+1}\left(\frac{2\ell^2}{N}\sum_{i=1}^{N}\left|\xi_{k+1}^{i,N}\right|^2+2\sqrt{\frac{2\ell^2}{N}}\sum_{i=1}^{N}\frac{x_k^{i,N}\xi_{k+1}^{i,N}}{\lambda_i}\right)\right] \tag{5.30}$$
$$= \mathbb{E}_k\left[\left(1\wedge e^{Q(x_k,\xi_{k+1})}\right)\left(-2R(x_k,\xi_{k+1})\right)\right]+\mathbb{E}_k\hat{r}^N \tag{5.31}$$

where

$$\hat{r}^N := -2\sqrt{\frac{2\ell^2}{N}}\left[\gamma_{k+1}\langle(\mathcal{C}_N)^{1/2}\nabla\Psi^N(x_k^N),\xi_{k+1}^N\rangle\right]. \tag{5.32}$$

We will show (as a consequence of (7.13) and (7.3)) that $\hat{r}^N$ is negligible for large N. So, by (5.10) and (5.31),

$$N\mathbb{E}_k(S_{k+1}^N - S_k^N) \simeq \mathbb{E}_k\left[\left(1\wedge e^{R(x_k^N,\xi_{k+1}^N)}\right)\left(-2R(x_k^N,\xi_{k+1}^N)\right)\right]. \tag{5.33}$$

Now observe that if $X\sim\mathcal{N}(\mu,\sigma^2)$ then,

$$\mathbb{E}\left[-2X\left(1\wedge e^X\right)\right] \overset{(5.20),(5.27)}{=} e^{\mu+\sigma^2/2}\Phi\left(-\frac{\mu}{\sigma}-\sigma\right)(-2\mu-2\sigma^2) - 2\mu\,\Phi\left(\frac{\mu}{\sigma}\right),$$

so that, if $X\sim\mathcal{N}(-\ell^2,2\ell^2 a)$ for some $a>0$, we have

$$\mathbb{E}(-2X(1\wedge e^X)) = A_\ell(a). \tag{5.34}$$

Therefore, by (5.14), (5.33) and the above, we conclude

$$N\mathbb{E}_k(S_{k+1}^N - S_k^N) \simeq A_\ell(S_k^N).$$

Showing that the diffusion coefficient for S_k^N vanishes is a consequence of the calculation that we have just done, indeed

$$\begin{aligned} N\mathbb{E}_k(S_{k+1}^N - S_k^N)^2 &= \frac{1}{N}\mathbb{E}_k\left[\sum_{i=1}^N \frac{\left|x_{k+1}^{i,N}\right|^2}{\lambda_i^2} - \frac{\left|x_k^{i,N}\right|^2}{\lambda_i^2}\right]^2 \\ &\simeq \frac{1}{N}\mathbb{E}_k\left[(1\wedge e^{R(x_k^N,\xi_{k+1}^N)})^2 R^2(x_k^N,\xi_{k+1}^N)\right] \\ &\le \frac{1}{N}\mathbb{E}_k\left[R^2(x_k^N,\xi_{k+1}^N)\right] \simeq \frac{1}{N}\mathbb{E}_k\left|Z_{\ell,k}\right|^2 \simeq \frac{2\ell^2\,S_k^N}{N}. \end{aligned}$$

We will prove that S_k^N's are uniformly bounded in N and k (in the sense of Lemma 7.4), hence $(2\ell^2\,S_k^N)/N\to 0$.

5.6. Suboptimal Scalings for the Proposal Variance

Consider the Random Walk algorithm with proposal (1.7), for $\beta\neq 1$. In this case the acceptance probability becomes

$$\alpha^\beta(x,\xi) := 1\wedge\exp Q^\beta(x,\xi),$$

where, with the same reasoning leading to (5.14),

$$Q^\beta(x_k,\xi_{k+1}) =: Q_k^\beta \simeq R_k^\beta \sim \mathcal{N}(-\ell^2 N^{1-\beta}, 2\ell^2 N^{1-\beta} S_k^N). \tag{5.35}$$

Assuming that S_0 is finite, one can show that S_k^N remains bounded (uniformly in k and N). Therefore, if we look at the average acceptance probability, we have

$$\begin{aligned} \mathbb{E}(1\wedge e^{Q^\beta(x_k,\xi_{k+1})}) &\overset{(5.27)}{=} \Phi\left(\frac{-\ell^2 N^{(1-\beta)/2}}{\sqrt{2\ell^2 S_k^N}}\right) \\ &\quad + e^{\ell^2 N^{1-\beta}(S_k^N-1)}\Phi\left(\frac{\ell^2 N^{(1-\beta)/2}(1-2S_k^N)}{\sqrt{2\ell^2 S_k^N}}\right). \end{aligned}$$

Therefore, if $\beta>1$ the acceptance probability tends to one as $N\to\infty$, if $0\le\beta<1$ it tends to zero.

6. Continuous Mapping Argument

In this section we explain the continuous mapping arguments that the proofs of Theorem 5.1 and Theorem 5.4 rely on. The continuous mapping argument that we use here is analogous to the one used in [MPS12, PST12]. The only difference is that the drift and diffusion coefficient of (1.10) are time dependent.

Section 6.1 and Section 6.2 contain the outline of the mapping argument that we will use in the proof of Theorem 5.1 and Theorem 5.4, respectively.

6.1. Continuous Mapping Argument for (1.11) (used in the Proof of Theorem 5.1)

Consider the chain S_k^N, defined in (1.16) and let $S^{(N)}(t)$ and $\bar{S}^{(N)}(t)$ be the continuous and piecewise constant interpolants of such a chain, respectively; we recall that $S^{(N)}(t)$ and $\bar{S}^{(N)}(t)$ have been defined in (1.17) and (5.2), respectively. Decompose the chain S_k^N into its drift and martingale part:

$$S_{k+1}^N = S_k^N + \frac{1}{N} A_\ell^N(x_k^N) + \frac{1}{\sqrt{N}} M_k^{2,N},$$

where

$$A_\ell^N(x_k^N) := N\mathbb{E}_k\left[S_{k+1}^N - S_k^N\right] \tag{6.1}$$

and

$$M_k^{2,N} := \sqrt{N}\left[S_{k+1}^N - S_k^N - \frac{1}{N} A_\ell^N(S_k^N)\right]. \tag{6.2}$$

We will show in Lemma 7.2 and Lemma 7.3 that $A_\ell^N(x_k^N)$ converges to $A_\ell(S(t))$. [5] Now a straightforward calculation (completely analogous to the one in [OPPS16, Appendix A]) shows that

$$S^{(N)}(t) = S_k^N + \int_{t_k}^t A_\ell^N(\bar{x}^{(N)}(v))dv + \sqrt{N}(t-t_k)M_k^{2,N}, \quad \text{when } t_k \le t < t_{k+1},$$

where $\bar{x}^{(N)}$, the piecewise constant interpolant of the chain $\{x_k\}_k$, is defined in (6.6) below. Therefore

$$S^{(N)}(t) = S_0^N + \int_0^t A_\ell^N(\bar{x}^{(N)}(v))dv + \frac{1}{\sqrt{N}}\sum_{j=0}^{k-1} M_j^{2,N} + \sqrt{N}(t-t_k)M_k^{2,N}, \quad \text{for any } t \in [0,T].$$

Setting

$$w^N(t) := \frac{1}{\sqrt{N}}\sum_{j=0}^{k-1} M_j^{2,N} + \sqrt{N}(t-t_k)M_k^{2,N}, \tag{6.3}$$

we can rewrite the above as

$$\begin{aligned} S^{(N)}(t) &= S_0^N + \int_0^t A_\ell^N(\bar{x}^{(N)}(v))dv + w^N(t) \\ &= S_0^N + \int_0^t A_\ell(S^{(N)}(v))dv + \hat{w}^N(t), \end{aligned} \tag{6.4}$$

where, for all $t \in [0,T]$,

$$\begin{aligned} \hat{w}^N(t) &:= \int_0^t \left[A_\ell^N(\bar{x}^{(N)}(v)) - A_\ell(S^{(N)}(v))\right] dv + w^N(t) \\ &= \int_0^t \left[A_\ell^N(\bar{x}^{(N)}(v)) - A_\ell(\bar{S}^{(N)}(v))\right] dv + \int_0^t \left[A_\ell(\bar{S}^{(N)}(v)) - A_\ell(S^{(N)}(v))\right] dv + w^N(t). \end{aligned} \tag{6.5}$$

[5] While the approximate drift $A_\ell^N(x_k^N)$ of the chain S_k^N depends only on x_k^N, the limiting drift A_ℓ depends only on $S(t)$. This is coherent with the fact that S_k^N depends only on x_k^N: in the limit, the dependence of the drift on S_k^N appears explicitly.

Equation (6.4) shows that $S^{(N)}(t) = \mathcal{J}_2(S_0^N, \hat{w}^N)$, where $\mathcal{J}_2$ is the map defined in Theorem 4.6. By the continuity of the map $\mathcal{J}_2$, if we show that $\hat{w}^N$ converges weakly to zero in $C([0,T];\mathbb{R})$, then $S^{(N)}(t)$ converges weakly to the solution of the ODE (1.11). The weak convergence of $\hat{w}^N$ to zero will be proved in Section 7.

Now we outline the continuous mapping argument for the chain x_k^N and in doing so we shall fix some more notation.

6.2. Continuous Mapping Argument for (1.10) (used in the Proof of Theorem 5.4)

We now consider the chain that we are actually interested in, i.e. the chain $\{x_k\}_k \subset \mathcal{H}^s$, defined in (2.4). We act analogously to what we have done for the chain S_k^N. So we start by recalling the definition of the continuous interpolant $x^{(N)}(t)$, equation (1.8), and we define the piecewise constant interpolant of the chain to be

$$\bar{x}^{(N)}(t) = x_k^N \quad \text{for} \quad t_k \le t < t_{k+1}. \tag{6.6}$$

We also recall the notation $\Theta(x,S)$ for the drift of equation (1.10), i.e.

$$\Theta(x,S) = F(x)D_\ell(S). \quad (x,S) \in \mathcal{H}^s \times \mathbb{R}_+. \tag{6.7}$$

The drift-martingale decomposition of the chain x_k^N is as follows:

$$x_{k+1}^N = x_k^N + \frac{1}{N}\Theta^N(x_k^N) + \frac{1}{\sqrt{N}}M_k^{1,N}. \tag{6.8}$$

where $\Theta^N(x)$ is

$$\Theta^N(x_k^N) := N\mathbb{E}_k\left[x_{k+1}^N - x_k^N\right] \tag{6.9}$$

and

$$M_k^{1,N} := \sqrt{N}\left[x_{k+1}^N - x_k^N - \frac{1}{N}\Theta^N(x_k^N)\right]. \tag{6.10}$$

Notice that $\Theta^N(x)$ is just a function of x; we will show (see Lemma 8.3 and (6.13)) that the approximate drift $\Theta^N(x)$ converges to $\Theta(x,S)$, the drift of the SDE (1.10); that is, in the limit the dependence on S becomes explicit (this should not surprie since, as already remarked, S_k^N depends only on x_k^N). Using again [OPPS16, Appendix A] we obtain

$$x^{(N)}(t) = x_k^N + \int_{t_k}^t \Theta^N(\bar{x}^{(N)}(v))dv + \sqrt{N}(t-t_k)M_k^{1,N}, \quad \text{when } t_k \le t < t_{k+1}$$

and therefore, for all $t \in [0,T]$,

$$x^{(N)}(t) = x_0^N + \int_0^t \Theta^N(\bar{x}^{(N)}(v))dv + \frac{1}{\sqrt{N}}\sum_{j=1}^{k-1} M_j^{1,N} + \sqrt{N}(t-t_k)M_k^{1,N}.$$

Setting

$$\eta^N(t) := \frac{1}{\sqrt{N}}\sum_{j=0}^{k-1} M_j^{1,N} + \sqrt{N}(t-t_k)M_k^{1,N}, \quad \text{when } t_k \le t < t_{k+1}, \tag{6.11}$$

we can rewrite the above as

$$\begin{aligned} x^{(N)}(t) &= x_0^N + \int_0^t \Theta^N(\bar{x}^{(N)}(v))dv + \eta^N(t) \\ &= x_0^N + \int_0^t \Theta(x^{(N)}(v), S(v))dv + \hat{\eta}^N(t)\,, \end{aligned} \tag{6.12}$$

where, for all $t \in [0,T]$,

$$\begin{aligned}\hat{\eta}^N(t) &:= \int_0^t \Big[\Theta^N(\bar{x}^{(N)}(v)) - \Theta(x^{(N)}(v), S(v))\Big]\, dv + \eta^N(t)\\ &= \int_0^t \Big[\Theta^N(\bar{x}^{(N)}(v)) - \Theta(\bar{x}^{(N)}(v), \bar{S}^{(N)}(v))\Big]\, dv\\ &+ \int_0^t \Big[\Theta(\bar{x}^{(N)}(v), \bar{S}^{(N)}(v)) - \Theta(x^{(N)}(v), S^{(N)}(v))\Big]\, dv\\ &+ \int_0^t \Big[\Theta(x^{(N)}(v), S^{(N)}(v)) - \Theta(x^{(N)}(v), S(v))\Big]\, dv + \eta^N(t). \end{aligned} \tag{6.13}$$

If we can prove that $\hat{\eta}^N(t)$ converges weakly in $C([0,T];\mathcal{H}^s)$ to

$$\eta(t) := \int_0^t \Gamma_\ell^{1/2}(S_v) dW_v, \tag{6.14}$$

where W_v is a $\mathcal{H}^s$-valued $\mathcal{C}_s$-Brownian motion, then (6.12) and the continuity of the map $\mathcal{J}_1$ allow to conclude that $x^{(N)}(t)$ converges weakly in $C([0,T];\mathcal{H}^s)$ to $x(t)$, solution of (4.3). Such an argument is the backbone of the proof of Theorem 5.4. The proof of Theorem 5.4 can be found in Section 8.

7. Proof of Theorem 5.1

Proof of Theorem 5.1. Recall the definition of the map $\mathcal{J}_2$ given in Theorem 4.6 and observe that thanks to (6.4),

$$S^{(N)}(t) = \mathcal{J}_2(S_0^N, \hat{w}^N(t)).$$

Therefore proving the statement of Theorem 5.1 amounts to proving that $\hat{w}^N(t)$ converges weakly to zero in $C([0,T];\mathbb{R})$. This is a consequence of the decomposition (6.5) together with Lemma 7.1, Lemma 7.2 and Lemma 7.3 below. □

In the following $\mathbb{E}_{x_0}$ denotes the expected value given $x_0 \in \mathcal{H}^s_\cap$, the initial value of the chain. We recall once again that the initial value of the chain x_k^N determines the initial value of the chain S_k^N.

Lemma 7.1. *Under the assumptions of Theorem 5.1, the martingale difference array $w^N(t)$ defined in (6.3) converges weakly to zero in $C([0,T];\mathbb{R})$.*

Lemma 7.2. *Under the assumptions of Theorem 5.1,*

$$\mathbb{E}_{x_0} \int_0^T \Big| A_\ell^N(\bar{x}^{(N)}(v)) - A_\ell(\bar{S}^{(N)}(v))\Big|^2\, dv \longrightarrow 0 \quad as \quad N \to \infty. \tag{7.1}$$

Lemma 7.3. *Under the assumptions of Theorem 5.1, for every fixed $T > 0$,*

$$\mathbb{E}_{x_0} \int_0^T \Big| A_\ell(\bar{S}^{(N)}(v)) - A_\ell(S^{(N)}(v))\Big|^2\, dv \longrightarrow 0 \quad as \quad N \to \infty. \tag{7.2}$$

Before proving the above lemmata, we state Lemma 7.4, which we will repeatedly use throughout this section and the next. The proof of Lemma 7.1 can be found in Section 7.2, the proof of Lemma 7.2 and Lemma 7.3 is the content of Section 7.1.

Lemma 7.4. *Let the assumptions of Theorem 5.1 hold. Then for every $m \geq 0$ there exists a constant $\bar{c} = \bar{c}(m)$ such that*

$$\mathbb{E}_{x_0} \|x_k^N\|_s^m < \bar{c}\,, \tag{7.3}$$

$$\mathbb{E}_{x_0} (S_k^N)^m < \bar{c} \tag{7.4}$$

and

$$\mathbb{E}_{x_0} e^{\frac{c}{N}\|\zeta_k^N\|^2} < \bar{c} \quad \textit{for all } c > 0. \tag{7.5}$$

We recall that ζ_k^N has been defined in (5.5). The constant $\bar{c} = \bar{c}(m)$ in the above bounds is independent of $N \in \mathbb{N}$ and of $0 \le k \le [TN]+1$ (but it depends on m).

Proof. See Appendix B. □

It is not trivial to prove Lemma 7.4 in the non-stationary regime that we are interested in. We make some more detailed remarks on this point in Remark 8.7.

Lemma 7.5. *Under the assumptions of Theorem 5.1,*

$$\bar{S}^{(N)}(t) \to S(t) \quad \textit{in } L^p(\Omega), \textit{ for every fixed } t > 0 \textit{ and any } p > 0.$$

Moreover,

$$\mathbb{E}_{x_0} \int_0^T \left| \bar{S}^{(N)}(t) - S(t) \right|^p dt \longrightarrow 0 \quad \textit{as } N \to \infty, \textit{ for all } p > 0$$

and

$$\mathbb{E}_{x_0} \int_0^T \left| S^{(N)}(t) - S(t) \right|^p dt \longrightarrow 0 \quad \textit{as } N \to \infty, \textit{ for all } p > 0\,.$$

Proof. Using Vitali's convergence theorem, the first statement is a corollary of (7.4) and Lemma 5.3 (indeed $\bar{S}^{(N)}(t) \le S_k^N + S_{k+1}^N$ and the right hand side has bounded moments of any order, so the sequence $\bar{S}^{(N)}(t)$ is uniformly integrable). As for the second statement, it can be obtained from the first by using again the bounded convergence theorem applied to the (deterministic) sequence $\mathbb{E}_{x_0} \left| \bar{S}^{(N)}(t) - S(t) \right|^p$. Indeed such a sequence tends to zero and is bounded by a multiple of the function $\mathbb{E}_{x_0} \left[\left| \bar{S}^{(N)}(t) \right|^p + \left| S(t) \right|^p \right]$, which is bounded again thanks to (7.4). The last statement is obtained similarly and we don't detail the argument. This concludes the proof of the lemma. □

7.1. Analysis of the Drift

Before starting the proof of Lemma 7.2 we observe that because $\sum_{j=1}^N (\xi_{k+1}^{j,N})^2$ has a Chi-squared distribution with N degrees of freedom, the following bound holds:

$$\mathbb{E}\left[\sum_{j=1}^N (\xi_{k+1}^{j,N})^2 \right]^m = 2^m \frac{\Gamma(m + N/2)}{\Gamma(N/2)} \lesssim N^m, \tag{7.6}$$

by Stirling's formula for the Gamma function Γ.

Proof of Lemma 7.2. Set

$$E_k^N := A_\ell^N(x_k^N) - A_\ell(S_k^N)\,. \tag{7.7}$$

Then, recalling that for any $b \in \mathbb{R}_+$ we set $[b] = n$ if $n \le b < n+1$ for some integer n,

$$\mathbb{E}_{x_0} \int_0^T \left| A_\ell^N(\bar{x}^{(N)}(v)) - A_\ell(\bar{S}^{(N)}(v)) \right|^2 dv = \mathbb{E}_{x_0} \frac{1}{N} \sum_{k=0}^{[TN]} \left| E_k^N \right|^2 \tag{7.8}$$

$$+ \left(T - \frac{[TN]}{N} \right) \mathbb{E}_{x_0} \left| E_{[TN]}^N \right|^2. \tag{7.9}$$

From the above equality and observing that $\left| T - \frac{[TN]}{N} \right| < 1/N$, it is clear that in order to show the limit (7.1) it is sufficient to prove that

$$\mathbb{E}_{x_0} \left| E_k^N \right|^2 \overset{N\to\infty}{\longrightarrow} 0\,, \quad \text{uniformly over } \; 0 \le k \le [NT].$$

To this end, we write $A_\ell(S_k^N) = \mathbb{E}_k[(1 \wedge e^{Z_{\ell,k}})(-2Z_{\ell,k})]$ (which follows from (5.34)) and use (5.31) and (6.1), obtaining

$$E_k^N = \mathbb{E}_k\left[(1 \wedge e^Q)(-2R)\right] - A_\ell(S_k^N) + \mathbb{E}_k \hat{r}^N = E_{1,k}^N + E_{2,k}^N + \mathbb{E}_k \hat{r}^N, \tag{7.10}$$

where

$$E_{1,k}^N := \mathbb{E}_k\left[((1 \wedge e^Q) - (1 \wedge e^R))(-2R)\right],$$

$$E_{2,k}^N := \mathbb{E}_k\left[(1 \wedge e^R)(-2R) - (1 \wedge e^{Z_{\ell,k}})(-2Z_{\ell,k})\right],$$

and $\hat{r}^N$ is defined in (5.32). Observe that from (5.6) and (7.6) we have

$$\mathbb{E}_k |R|^{2p} \lesssim 1 + \frac{\|\zeta_k^N\|^{2p}}{N^p}, \qquad p \geq 1, \tag{7.11}$$

as, given x_k, the sum $\sum_{i=1}^N \zeta_k^{i,N} \xi_k^{i,N}$ is Gaussian with mean zero and variance $\|\zeta_k^N\|^2$. From the definition of ζ_k^N, equation (5.5), we have

$$\frac{\|\zeta_k^N\|^2}{N} \lesssim \frac{1}{N}\sum_{i=1}^N \frac{\left|x_k^{i,N}\right|^2}{\lambda_i^2} + \frac{1}{N}\|\mathcal{C}^{1/2}\nabla\Psi^N(x_k^N)\|^2 = S_k^N + \frac{1}{N}\|\mathcal{C}^{1/2}\nabla\Psi^N(x_k^N)\|^2.$$

By acting as in [MPS12, page 915] we obtain

$$\|\mathcal{C}^{1/2}\nabla\Psi^N(x)\| \lesssim \|x\|_s \vee \|x\|_s^\varsigma \tag{7.12}$$

$$\lesssim (1 + \|x\|_s), \tag{7.13}$$

hence

$$\frac{\|\zeta_k^N\|^{2p}}{N^p} \lesssim (S_k^N)^p + \frac{1}{N^p}(1 + \|x_k^N\|_s^{2p}), \quad p \geq 1. \tag{7.14}$$

Combining (7.11) and (7.14) then gives

$$\mathbb{E}_k |R|^{2p} \lesssim 1 + (S_k^N)^p + \frac{1}{N^p}(1 + \|x_k^N\|_s^{2p}), \quad p \geq 1. \tag{7.15}$$

Therefore, using (7.15) (with $p = 1$), (7.3) and (7.4), we obtain

$$\mathbb{E}_{x_0} |R|^2 = \mathbb{E}_{x_0}\mathbb{E}_k |R|^2 \lesssim 1 + \frac{\mathbb{E}_{x_0}\|\zeta_k^N\|^2}{N} \lesssim 1 + \mathbb{E}_{x_0} S_k^N + \frac{\mathbb{E}_{x_0}\|x_k^N\|_s^2}{N} < \infty. \tag{7.16}$$

Using the Lipshitzianity of the function $1 \wedge e^x$ and (5.7), we have

$$\left|E_{1,k}^N\right| \lesssim \mathbb{E}_k \left|r^N R\right| \leq \left(\mathbb{E}_k \left|r^N\right|^2\right)^{1/2} \left(\mathbb{E}_k (R)^2\right)^{1/2}. \tag{7.17}$$

By (7.16) and (5.9) we then conclude

$$\mathbb{E}_{x_0} \left|E_{1,k}^N\right|^2 \leq \frac{1}{N^2}\mathbb{E}_{x_0}\left(S_k^N + \frac{\|x_k^N\|_s^2}{N}\right) \longrightarrow 0, \tag{7.18}$$

thanks to Lemma 7.4. As for the term $E_{2,k}^N$, we use the lipshitzianity of the function $(1 \wedge e^x)(-2x)$ to conclude

$$\mathbb{E}_{x_0} \left|E_{2,k}^N\right|^2 \lesssim \mathbb{E}_{x_0}\mathbb{E}_k |R - Z_{\ell,k}|^2 \overset{(8.72)}{\lesssim} \frac{1 + \mathbb{E}_{x_0}\|x_k\|_s^2}{N} \longrightarrow 0. \tag{7.19}$$

Finally, to estimate $\hat{r}^N$ (defined in (5.32)), we use the independence, given x_k, of $\Psi(x_k)$ from ξ_{k+1}:

$$\begin{aligned}\left|\mathbb{E}_k \hat{r}^N\right|^p &\leq \left(\mathbb{E}_k \left|\hat{r}^N\right|^2\right)^{p/2} \\ &\leq \frac{1}{N^{p/2}} \left(\mathbb{E}_k \left| \sum_{j=1}^N \left[\mathcal{C}_N^{1/2} \nabla \Psi\right]^j \xi_{k+1}^{j,N} \right|^2 \right)^{p/2} \\ &= \frac{1}{N^{p/2}} \left(\sum_{j=1}^N \mathbb{E}_k \left| \left[\mathcal{C}_N^{1/2} \nabla \Psi\right]^j \right|^2 \left| \xi_{k+1}^{j,N} \right|^2 \right)^{p/2} \\ &= \frac{1}{N^{p/2}} \|\mathcal{C}^{1/2} \nabla \Psi\|^p, \end{aligned} \tag{7.20}$$

where in the above $\left[\mathcal{C}_N^{1/2} \nabla \Psi\right]^j$ denotes the j-th component of $\mathcal{C}_N^{1/2} \nabla \Psi$. Using (7.13) we have

$$\left|\mathbb{E}_k \hat{r}^N\right|^p \leq \left(\mathbb{E}_k \left|\hat{r}^N\right|^2\right)^{p/2} \leq \frac{1 + \|x_k^N\|_s^p}{N^{p/2}} \quad \text{for all } p \geq 1. \tag{7.21}$$

Hence, (7.3) gives

$$\mathbb{E}_{x_0} \left|\mathbb{E}_k \hat{r}^N\right|^p \leq \mathbb{E}_{x_0} \left(\mathbb{E}_k \left|\hat{r}^N\right|^2\right)^{p/2} \leq \frac{1}{N^{p/2}} \quad \text{for all } p \geq 1. \tag{7.22}$$

This concludes the proof. □

Proof of Lemma 7.3. By the Lipshitzianity of A_ℓ,

$$\left|A_\ell(\bar{S}^{(N)}(v)) - A_\ell(S^{(N)}(v))\right|^2 \lesssim \left|\bar{S}^{(N)}(v) - S^{(N)}(v)\right|^2.$$

The statement is now a consequence of (8.67).

□

7.2. *Analysis of the Noise*

Proof of Lemma 7.1. By the martingale central limit theorem, all we need to prove is that

$$\frac{1}{N} \sum_{j=1}^{[TN]} \mathbb{E}_{x_0} \left|M_j^{2,N}\right|^2 \to 0 \quad \text{as } N \to \infty.$$

From the definition of $M_j^{2,N}$, equation (6.2),

$$\begin{aligned}\mathbb{E}_{x_0} \frac{\left|M_j^{2,N}\right|^2}{N} &= \mathbb{E}_{x_0} \left|S_{k+1}^N - S_k^N - \mathbb{E}_k[S_{k+1}^N - S_k^N]\right|^2 \\ &\lesssim \mathbb{E}_{x_0} \left|S_{k+1}^N - S_k^N\right|^2. \end{aligned}$$

With the same calculation as in (5.31),

$$S_{k+1}^N - S_k^N = \frac{\gamma_{k+1}(-2R)}{N} + \frac{1}{N} \hat{r}^N. \tag{7.23}$$

Therefore, using (7.16), (7.22) and $\gamma_{k+1} \leq 1$,

$$\mathbb{E}_{x_0} \left|S_{k+1}^N - S_k^N\right|^2 \lesssim \frac{1}{N^2} \mathbb{E}_{x_0} \mathbb{E}_k |R|^2 + \frac{1}{N^3} \lesssim \frac{1}{N^2}. \tag{7.24}$$

The above implies the bound

$$\mathbb{E}_{x_0}\frac{\left|M_j^{2,N}\right|^2}{N} \lesssim \frac{1}{N^2}.$$

We can therefore conclude

$$\frac{1}{N}\sum_{j=1}^{[TN]}\mathbb{E}_{x_0}\left|M_j^{2,N}\right|^2 \lesssim \frac{T}{N} \longrightarrow 0.$$

□

8. Proof of Theorem 5.4

Before starting the proof of Theorem 5.4, we state Lemma 8.1 below. We recall the definition of Wasserstain distance between two random variables X and Y:

$$Wass(X,Y) := \sup_{f\in \mathrm{Lip}_1} \mathbb{E}(f(X)-f(Y)), \tag{8.1}$$

where Lip_1 denotes the class of Lipshitz functions with Lipshitz constant equal to one. Notice that from the definition,

$$Wass(X,Y) \le \mathbb{E}\left|X-Y\right|. \tag{8.2}$$

In the next Lemma (8.1) we refer to the Wasserstein distance relative to the marginal $\mathbb{E}_k$.

Lemma 8.1. *Let the assumptions of Theorem 5.1 hold. Recalling the definitions of R, R^i, G and $Z_{\ell,k}$, (5.6), (5.17), (5.12) and (5.13) respectively, we have*

$$Wass(R,R^i) \le \mathbb{E}_k\left|R-R^i\right| \lesssim \frac{1+\left|\zeta_k^{i,N}\right|}{\sqrt{N}}, \tag{8.3}$$

$$Wass(R,G) \le \mathbb{E}_k\left|R-G\right| \lesssim \frac{1}{\sqrt{N}} \tag{8.4}$$

and

$$Wass(G,Z_{\ell,k}) \le \mathbb{E}_k\left|G-Z_{\ell,k}\right| \lesssim \frac{1+\|x_k\|_s}{\sqrt{N}}. \tag{8.5}$$

Therefore,

$$Wass(R,Z_{\ell,k}) \le \mathbb{E}_k\left|R-Z_{\ell,k}\right| \lesssim \frac{1+\|x_k\|_s}{\sqrt{N}}. \tag{8.6}$$

Proof. See Appendix B. □

Proof of Theorem 5.4. If $\mathcal{J}_1$ is the map defined in Theorem 4.6, then (6.12) means that

$$x^{(N)}(t) = \mathcal{J}_1(x_0^N, \hat{\eta}^N(t)).$$

From the continuity of $\mathcal{J}_1$, in order to prove that $x^{(N)}(t) \xrightarrow{d} x(t)$, we just need to prove that $\hat{\eta}^N(t) \xrightarrow{d} \eta(t)$, where $\eta(t)$ is the stochastic integral defined in (6.14). The weak convergence $\hat{\eta}^N(t) \xrightarrow{d} \eta(t)$ follows from Lemma 8.2, Lemma 8.3, Lemma 8.4, Lemma 8.5 and the decomposition (6.13). □

Lemma 8.2. *Let the assumptions of Theorem 5.4 hold. Then the interpolated martingale difference array $\eta^N(t)$ defined in (6.11) converges weakly in $C([0,T];\mathcal{H}^s)$ to the stochastic integral $\eta(t)$, equation (6.14).*

Lemma 8.3. *Let the assumptions of Theorem 5.4 hold. Then for every fixed $T>0$,*

$$\mathbb{E}_{x_0}\int_0^T \|\Theta^N(\bar{x}^{(N)}(v)) - \Theta(\bar{x}^{(N)}(v), \bar{S}^{(N)}(v))\|_s^2\, dv \longrightarrow 0 \quad as \quad N\to\infty. \tag{8.7}$$

Lemma 8.4. *Let the assumptions of Theorem 5.4 hold. Then for any fixed $T > 0$*

$$\mathbb{E}_{x_0} \int_0^T \|\Theta(\bar{x}^{(N)}(v), \bar{S}^{(N)}(v)) - \Theta(x^{(N)}(v), S^{(N)}(v))\|_s^2 \, dv \longrightarrow 0 \quad as \quad N \to \infty. \tag{8.8}$$

Lemma 8.5. *Let the assumptions of Theorem 5.4 hold. Then for any fixed $T > 0$*

$$\mathbb{E}_{x_0} \int_0^T \|\Theta(x^{(N)}(v), S^{(N)}(v)) - \Theta(x^{(N)}(v), S(v))\|_s^2 dv \longrightarrow 0 \quad as \quad N \to \infty.$$

We will prove Lemma 8.2 in Section 8.2 and Lemma 8.3, Lemma 8.4, Lemma 8.5 in Section 8.1.

8.1. Analysis of the Drift

In what follows we will need some preliminary estimates, which we list in Lemma 8.6 below.

Lemma 8.6. *Under the assumptions of Theorem 5.4, the following holds:*

i) *Let Y be a positive random variable such that $\mathbb{E}_{x_0} |Y|^q < \infty$ for all $q \geq 1$ (should Y depend on k and N, all the moments are assumed to be bounded independently of k and N). Then, uniformly over $0 \leq k \leq [TN] + 1$,*

$$\limsup_{N \to \infty} \mathbb{E}_{x_0} \left[Y \sum_{i=1}^N i^{2s} \lambda_i^2 \left| \zeta_k^{i,N} \right|^p \right] < \infty, \qquad for \ all \ p \geq 0. \tag{8.9}$$

ii) *Moreover,*

$$\frac{1}{N} \sum_{k=0}^{[TN]} \mathbb{E}_{x_0} \sum_{i=1}^N i^{2s} \lambda_i^2 \frac{\left| \zeta_k^{i,N} \right|^p}{(N S_k^N)^\alpha} \overset{N \to \infty}{\longrightarrow} 0, \quad for \ all \ p \geq \frac{2\alpha}{\varsigma} > 0, \tag{8.10}$$

where we recall that the constant $1/2 \leq \varsigma < 1$ is the one appearing in Assumption 3.1.

iii) *Finally,*

$$\mathbb{E}_k \frac{1}{(1 + |R| \sqrt{N})^2} \lesssim \left(1 + \|x_k^N\|_s^2\right) \frac{1}{(N S_k^N)^{1/4}} + \frac{1}{\sqrt{N S_k^N}}. \tag{8.11}$$

Proof. See Appendix B. □

Remark 8.7 (On Lemma 8.6 and Lemma 7.4)**.** The proofs of Lemma 8.6 and Lemma 7.4 bring up some of the main differences between the stationary and the non-stationary case, so it is worth making some comments.

- If we start the chain in stationarity, i.e. $x_0^N \sim \pi^N$, where π^N has been defined in (1.5), then $x_k^N \sim \pi^N$ for every $k \geq 0$. As already observed in the introduction, π^N is absolutely continuous with respect to a Gaussian measure; 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 $x \sim \pi_0$. If $x \sim \pi_0$ then $x^N = \sum_{i=1}^N \lambda_i \rho^i \phi_i$, where ρ^i are i.i.d. $\mathcal{N}(0,1)$. Therefore, recalling (5.5) (see also (8.51)), one gets
$$\left|\zeta^{i,N}\right|^p \lesssim \left|\rho^i\right|^p + \|x\|_s^p. \tag{8.12}$$
With this observation it is then clear that in stationarity the bounds (7.3) and (7.4) are trivially true, and (8.9) follows easily from (8.12) and (7.3). For the same reason, in the stationary case the estimate (7.5) is a consequence of Fernique's Theorem, see [MPS12, page 916]. With a similar reasoning and by (1.22) one can see that also (8.10) holds in stationarity.
- In our case, i.e. out of stationarity, proving the bounds of Lemma 8.6 and Lemma 7.4 requires a bit of an argument. In particular, the reason why the limit (8.10) holds can be understood at least heuristically observing that S_k^N converges to $S(t)$ (i.e. to a finite number, which is strictly positive under our assumptions and it converges to 1 if we work in stationarity, see (1.22)). Combining this observation with (8.9) gives, heuristically, (8.10).

- On a minor note, we point out that the limit (8.10) might not hold for $k = 0$ if we were to allow $S_0 = 0$. Indeed, suppose again for simplicity that $\Psi = 0$. If $S_0 = 0$ and the sequence of partial sums $\sum_{j=1}^{N} \frac{|x_0^{j,N}|^2}{\lambda_j^2}$ is convergent then the quantity on the LHS of (8.10) is in general only bounded. (However if we were to extend the proof to the case $S_0 = 0$ we would not need (8.10) to hold at $k = 0$).

□

Proof of Lemma 8.3. Set

$$e_k^N := \Theta^N(x_k^N) - \Theta(x_k^N, S_k^N) = N\mathbb{E}_k[x_{k+1}^N - x_k^N] - \Theta(x_k^N, S_k^N) \tag{8.13}$$

Then

$$\mathbb{E}_{x_0} \int_0^T \|\Theta^N(\bar{x}^{(N)}(v)) - \Theta(\bar{x}^{(N)}(v), \bar{S}^{(N)}(v))\|_s^2 \, dv = \mathbb{E}_{x_0} \frac{1}{N} \sum_{k=0}^{[TN]} \|e_k^N\|_s^2 \tag{8.14}$$

$$+ \left(T - \frac{[TN]}{N}\right) \mathbb{E}_{x_0} \|e_{[TN]}^N\|_s^2. \tag{8.15}$$

If $e_k^{i,N}$ is the $i-$th component of e_k^N, the sum on the RHS of (8.14) may be rewritten as

$$\mathbb{E}_{x_0} \frac{1}{N} \sum_{k=0}^{[TN]} \sum_{i=1}^{N} i^{2s} \left| e_k^{i,N} \right|^2.$$

The statement now follows from Lemma 8.8 below. □

Proof of Lemma 8.4. From (6.7) we have

$$\begin{aligned}
\|\Theta(\bar{x}^{(N)}(v), \bar{S}^{(N)}(v)) - \Theta(x^{(N)}(v), S^{(N)}(v))\|_s^2 &\lesssim \left| D_\ell(\bar{S}^{(N)}(v)) \right|^2 \|F(\bar{x}^{(N)}(v)) - F(x^{(N)}(v))\|_s^2 \\
&+ \|F(x^{(N)}(v))\|_s^2 \left| D_\ell(\bar{S}^{(N)}) - D_\ell(S^{(N)}) \right|^2 \\
&\lesssim \|\bar{x}^{(N)}(v) - x^{(N)}(v)\|_s^2 \\
&+ (1 + \|x^{(N)}(v)\|_s^2) \left| \bar{S}^{(N)}(v) - S^{(N)}(v) \right|^2,
\end{aligned} \tag{8.16}$$

having used the boundedness and Lipshitzianity of D_ℓ, the Lipshitzianity of F (Lemma 3.4 and Lemma 4.2, respectively) and the bound

$$\|F(z)\|_s^2 \lesssim 1 + \|z\|_s^2. \tag{8.17}$$

The above bound is a consequence of Assumption 3.1 and

$$\|\mathcal{C}\nabla\Psi(z)\|_s^2 = \sum_{i=1}^{\infty} i^{2s} \lambda_i^2 \left| (\nabla\Psi(z))^i \right|^2 \leq \|\nabla\Psi(z)\|_{-s}^2.$$

Moreover, if $t_k \leq v \leq t_{k+1}$ then from the definition (1.8), we have

$$\begin{aligned}
\mathbb{E}_{x_0} \|\bar{x}^{(N)}(v) - x^{(N)}(v)\|_s^2 &= \mathbb{E}_{x_0} \|(Nv - k)(x_{k+1} - x_k)\|_s^2 \\
&\overset{(2.4)}{\lesssim} \frac{1}{N} \mathbb{E} \|\mathcal{C}_N^{1/2} \xi_{k+1}^N\|_s^2 \overset{(3.7)}{\lesssim} \frac{1}{N}.
\end{aligned} \tag{8.18}$$

The statement of the lemma is a consequence of (8.16), (8.18), (7.3), (8.66) and (8.53). □

Proof of Lemma 8.5. Analogous to the proof of Lemma 8.4, so we only sketch it.

$$\mathbb{E}_{x_0} \|\Theta(x^{(N)}(t), S^{(N)}(t)) - \Theta(x^{(N)}(t), S(t))\|_s^2 \lesssim \mathbb{E}_{x_0} \|F(x^{(N)}(t))\|_s^2 \left| D_\ell(S^{(N)}(t)) - D_\ell(S(t)) \right|^2.$$

Now the RHS goes to zero thanks to the Lipshitzianity of D_ℓ,(8.17), Lemma 7.4 and Lemma 7.5. □

Lemma 8.8. *Let the assumptions of Theorem 5.4 hold and recall that $e_k^{i,N}$ is the i-th component of e_k^N, defined in (8.13). Then,*

$$\mathbb{E}_{x_0}\frac{1}{N}\sum_{k=0}^{[TN]}\|e_k^N\|_s^2 = \mathbb{E}_{x_0}\frac{1}{N}\sum_{k=0}^{[TN]}\sum_{i=1}^{N} i^{2s}\left|e_k^{i,N}\right|^2 \longrightarrow 0 \quad \text{as } N\to\infty.$$

Proof of Lemma 8.8. This proof is partly analogous to the proofs of [MPS12, Lemma 5.5-Lemma 5.11]. The main difference is that here we deal with time dependent coefficients. The proof will only be detailed when it differs from [MPS12]; where it does not we will provide fewer details.

From the definition of Θ, equation (6.7), the i-th component of Θ calculated at (x_k^N, S_k^N) is

$$\Theta^i(x_k^N, S_k^N) = -\lambda_i\zeta_k^{i,N}D_\ell(S_k^N) = -2\ell^2\lambda_i\zeta_k^{i,N}\mathbb{E}e^{Z_{\ell,k}}\mathbf{1}_{\{Z_{\ell,k}<0\}}\,. \tag{8.19}$$

where the second equality is a consequence of (5.24) and (5.13) . Therefore the i-th component of e_k^N is

$$\begin{aligned} e_k^{i,N} &= \sqrt{2N\ell^2}\,\lambda_i\mathbb{E}_k^\xi\left[\left(1\wedge e^{Q_k}\right)\xi_{k+1}^{i,N}\right] - \Theta^i(x_k^N, S_k^N)\\ &= \sqrt{2N\ell^2}\,\lambda_i\mathbb{E}_k^\xi\left[\left(1\wedge e^{Q_k}\right)\xi_{k+1}^{i,N}\right] + \lambda_i\zeta_k^{i,N}D_\ell(S_k^N)\,. \end{aligned}$$

Following the reasoning of Section 5.3, we decompose $e_k^{i,N}$ as follows:

$$e_k^{i,N} = \sqrt{2N\ell^2}\,\lambda_i\mathbb{E}_k^\xi\left[\left(1\wedge e^{R^i-\sqrt{\frac{2\ell^2}{N}}\zeta_k^{i,N}\xi_{k+1}^{i,N}}\right)\xi_{k+1}^{i,N}\right] - \Theta^i(x_k^N, S_k^N) + e_{1,k}^{i,N} + e_{2,k}^{i,N}\,, \quad \text{where}$$

$$e_{1,k}^{i,N} := \sqrt{2N\ell^2}\,\lambda_i\mathbb{E}_k^\xi\left[\left(\left(1\wedge e^{Q_k}\right) - \left(1\wedge e^{R}\right)\right)\xi_{k+1}^{i,N}\right] \tag{8.20}$$

$$e_{2,k}^{i,N} := \sqrt{2N\ell^2}\,\lambda_i\mathbb{E}_k^\xi\left[\left(\left(1\wedge e^{R}\right) - \left(1\wedge e^{R^i-\sqrt{\frac{2\ell^2}{N}}\zeta_k^{i,N}\xi_{k+1}^{i,N}}\right)\right)\xi_{k+1}^{i,N}\right]. \tag{8.21}$$

We now use equality (5.21), leading to:

$$e_k^{i,N} = -2\ell^2\lambda_i\zeta_k^{i,N}\mathbb{E}_k^{\xi^i_-}e^{R^i}\Phi\left(\frac{-R^i}{\sqrt{\frac{2\ell^2}{N}}\left|\zeta_k^{i,N}\right|}\right) - \Theta^i(x_k^N, S_k^N) + e_{1,k}^{i,N} + e_{2,k}^{i,N} + e_{3,k}^{i,N},$$

where

$$\begin{aligned} e_{3,k}^{i,N} := {}& -2\ell^2\lambda_i\zeta_k^{i,N}\left(e^{\frac{\ell^2}{N}|\zeta_k^{i,N}|^2}-1\right)\mathbb{E}_k^{\xi^i_-}e^{R^i}\Phi\left(\frac{-R^i}{\sqrt{\frac{2\ell^2}{N}}\left|\zeta_k^{i,N}\right|} - \sqrt{\frac{2\ell^2}{N}}\left|\zeta_k^{i,N}\right|\right)\\ & - 2\ell^2\lambda_i\zeta_k^{i,N}\mathbb{E}_k^{\xi^i_-}e^{R^i}\left[\Phi\left(\frac{-R^i}{\sqrt{\frac{2\ell^2}{N}}\left|\zeta_k^{i,N}\right|} - \sqrt{\frac{2\ell^2}{N}}\left|\zeta_k^{i,N}\right|\right) - \Phi\left(\frac{-R^i}{\sqrt{\frac{2\ell^2}{N}}\left|\zeta_k^{i,N}\right|}\right)\right]. \end{aligned} \tag{8.22}$$

Finally, by setting

$$e_{4,k}^{i,N} := -2\ell^2\lambda_i\zeta_k^{i,N}\mathbb{E}_k^\xi\left[e^{R^i}\Phi\left(\frac{-R^i}{\sqrt{\frac{2\ell^2}{N}}\left|\zeta_k^{i,N}\right|}\right) - e^{R^i}\mathbf{1}_{\{R^i<0\}}\right], \tag{8.23}$$

$$e_{5,k}^{i,N} := -2\ell^2\lambda_i\zeta_k^{i,N}\mathbb{E}_k^\xi\left[e^{R^i}\mathbf{1}_{\{R^i<0\}} - e^{Z_{\ell,k}}\mathbf{1}_{\{Z_{\ell,k}<0\}}\right], \tag{8.24}$$

and using (8.19), we obtain

$$e_k^{i,N} = \sum_{h=1}^{5} e_{h,k}^{i,N} \,. \tag{8.25}$$

Now that we have the above decomposition, we need to find bounds on each of the $e_{h,k}^{i,N}$'s, $h = 1, \dots, 5$, which is what we shall do next.

- $e_{1,k}^{i,N}$ **and** $e_{2,k}^{i,N}$**:** The bounds on $e_{1,k}^{i,N}$ and $e_{2,k}^{i,N}$ are straightforward:

$$\left|e_{1,k}^{i,N}\right| \lesssim \frac{\lambda_i}{\sqrt{N}} \quad \text{and} \quad \left|e_{2,k}^{i,N}\right| \lesssim \frac{\lambda_i}{\sqrt{N}} \,, \tag{8.26}$$

  The first estimate is a consequence of (5.6), (5.7), (5.9) and the Lipshitzianity of the function $f(x) = 1 \wedge e^x$; for the second we used definition (5.17).

- $e_{3,k}^{i,N}$**:** To study $e_{3,k}^{i,N}$, we set

$$e_{3,k}^{i,N} := \bar{e}_{3,k}^{i,N} + \tilde{e}_{3,k}^{i,N} \,, \tag{8.27}$$

  with

$$\bar{e}_{3,k}^{i,N} := -2\ell^2 \lambda_i \zeta_k^{i,N} \left(e^{\frac{\ell^2}{N}|\zeta_k^{i,N}|^2} - 1 \right) \mathbb{E}_k^{\xi^i} e^{R^i} \Phi\left(\frac{-R^i}{\sqrt{\frac{2\ell^2}{N}}\left|\zeta_k^{i,N}\right|} - \sqrt{\frac{2\ell^2}{N}}\left|\zeta_k^{i,N}\right| \right)$$

$$\tilde{e}_{3,k}^{i,N} := -2\ell^2 \lambda_i \zeta_k^{i,N} \mathbb{E}_k^{\xi^i} e^{R^i} \left[\Phi\left(\frac{-R^i}{\sqrt{\frac{2\ell^2}{N}}\left|\zeta_k^{i,N}\right|} - \sqrt{\frac{2\ell^2}{N}}\left|\zeta_k^{i,N}\right| \right) - \Phi\left(\frac{-R^i}{\sqrt{\frac{2\ell^2}{N}}\left|\zeta_k^{i,N}\right|} \right) \right].$$

  To estimate $\tilde{e}_{3,k}^{i,N}$, we use the boundedness and Lipshitzianity of Φ together with

$$\mathbb{E}_k^{\xi} e^{R^i} \lesssim e^{\frac{\ell^2}{N}\|\zeta_k^N\|^2}, \tag{8.28}$$

  see [MPS12, (5.20)]. We therefore obtain

$$\left|\tilde{e}_{3,k}^{i,N}\right| \lesssim \lambda_i \frac{\left|\zeta_k^{i,N}\right|^2}{\sqrt{N}} e^{\frac{\ell^2}{N}\|\zeta_k^N\|^2} \,. \tag{8.29}$$

  The term $\bar{e}_{3,k}^{i,N}$ will be studied separately later.

- $e_{4,k}^{i,N}$**:** We act as in the proof of [MPS12, Lemma 5.7-Lemma 5.9] and obtain

$$\left|e_{4,k}^{i,N}\right| \lesssim \lambda_i \left|\zeta_k^{i,N}\right| e^{\|\zeta_k^N\|^2/N} \left(1 + \left|\zeta_k^{i,N}\right|\right) \left(\mathbb{E}_k \frac{1}{(1 + |R|\sqrt{N})^2} \right)^{1/4} . \tag{8.30}$$

- $e_{5,k}^{i,N}$**:** Let $g(x) := e^x \mathbf{1}_{\{x<0\}}$; using the same argument as in [MPS12, page 923], if X and Y are two random variables such that one of them has a density with respect to the Lebesgue measure and such a density is bounded by M, then

$$|\mathbb{E}g(X) - \mathbb{E}g(Y)| \lesssim \sqrt{M\, Wass(X,Y)}. \tag{8.31}$$

  Such a result is applicable to R^i and $Z_{\ell,k}$ as $Z_{\ell,k}$ is (conditionally) Gaussian with variance S_k^N. Therefore using (8.3), (8.6) and (8.31) with $M = 1/\sqrt{2\pi\, S_k^N}$, we have

$$\begin{aligned} \left|\mathbb{E}_k^{\xi} e^{R^i} \mathbf{1}_{\{R^i<0\}} - \mathbb{E} e^{Z_{\ell,k}} \mathbf{1}_{\{Z_{\ell,k}<0\}}\right| &\lesssim \frac{1}{\left(S_k^N\right)^{1/4}} \sqrt{Wass(R^i, Z_{\ell,k})} \\ &\lesssim \frac{1}{N^{1/4}} \frac{1}{\left(S_k^N\right)^{1/4}} \left[\left(1 + \|x_k^N\|_s\right)^{1/2} + \sqrt{1 + \left|\zeta_k^{i,N}\right|} \right] \end{aligned}$$

The above, together with (8.24), implies

$$\left|e_{5,k}^{i,N}\right| \lesssim \frac{\lambda_i \left|\zeta_k^{i,N}\right|}{\left(S_k^N\right)^{1/4}} \frac{1}{N^{1/4}} \left[\left(1+\|x_k^N\|_s\right)^{1/2} + \sqrt{1+\left|\zeta_k^{i,N}\right|}\right] \tag{8.32}$$

From the bounds (8.26), (8.29), (8.30) and (8.32), we get

$$\begin{aligned}
\mathbb{E}_{x_0} \sum_{h=1,h\neq 3}^{5} \sum_{i=1}^{N} i^{2s} \left|e_{h,k}^{i,N}\right|^2 + \sum_{i=1}^{N} i^{2s} \left|\tilde{e}_{3,k}^{i,N}\right|^2 \lesssim\ & \mathbb{E}_{x_0} \sum_{i=1}^{N} \frac{i^{2s}\lambda_i^2}{N} \\
& + \mathbb{E}_{x_0} \sum_{i=1}^{N} i^{2s}\lambda_i^2 \frac{\left|\zeta_k^{i,N}\right|^4}{N} e^{\frac{2\ell^2}{N}\|\zeta_k^N\|^2} \\
& + \mathbb{E}_{x_0} e^{\frac{2}{N}\|\zeta_k^N\|^2} \left(\mathbb{E}_k \frac{1}{(1+|R|\sqrt{N})^2}\right)^{1/2} \sum_{i=1}^{N} i^{2s}\lambda_i^2 \left(\left|\zeta_k^{i,N}\right|^2 + \left|\zeta_k^{i,N}\right|^4\right) \\
& + \mathbb{E}_{x_0} \sum_{i=1}^{N} i^{2s}\lambda_i^2 \frac{\left|\zeta_k^{i,N}\right|^2 + \left|\zeta_k^{i,N}\right|^3}{\sqrt{N}} \frac{1}{\sqrt{S_k^N}} \\
& + \mathbb{E}_{x_0} \sum_{i=1}^{N} i^{2s}\lambda_i^2 \left|\zeta_k^{i,N}\right|^2 \frac{(1+\|x_k\|_s)}{\sqrt{N}} \frac{1}{\sqrt{S_k^N}}.
\end{aligned}$$

After simple manipulations and using Lemma 7.4 and Lemma 8.6, we have

$$\mathbb{E}_{x_0} \frac{1}{N} \sum_{k=0}^{[TN]} \left(\sum_{h=1,h\neq 3}^{5} \sum_{i=1}^{N} i^{2s} \left|e_{h,k}^{i,N}\right|^2 + \sum_{i=1}^{N} i^{2s} \left|\tilde{e}_{3,k}^{i,N}\right|^2 \right) \longrightarrow 0,$$

as $N \to \infty$.

Remark 8.9. When we apply (8.10) of Lemma 8.6 to the above, we need to enforce the condition $p \geq 2\alpha/\varsigma$, under which (8.10) holds. Rewriting such a condition as $\varsigma \geq 2\alpha/p$ and observing that this condition is always applied in the above with $p \geq 2$ and $\alpha \leq 1$, we get the constraint $\varsigma \geq 1/2$ appearing in Assumptions 3.1.

Returning to the proof, if we prove the limit,

$$\mathbb{E}_{x_0} \frac{1}{N} \sum_{k=0}^{[TN]} \sum_{i=1}^{N} i^{2s} \left|\bar{e}_{3,k}^{i,N}\right|^2 \to 0,$$

we are done. To study $\bar{e}_{3,k}^{i,N}$, we use again (8.28) and the bound $\Phi \leq 1$, obtaining

$$\left|\bar{e}_{3,k}^{i,N}\right| \lesssim \lambda_i \left|\zeta_k^{i,N}\right| \left(e^{\frac{\ell^2}{N}\left|\zeta_k^{i,N}\right|^2} - 1\right) e^{\frac{\ell^2}{N}\|\zeta_k^N\|^2}.$$

Therefore, by the weighted Jentzen inequality and (7.5),

$$\begin{aligned}
\mathbb{E}_{x_0} \sum_{i=1}^{N} i^{2s} \left|\bar{e}_{3,k}^{i,N}\right|^2 &\lesssim \mathbb{E}_{x_0} \left[e^{\frac{2\ell^2}{N}\|\zeta_k^N\|^2} \sum_{i=1}^{N} i^{2s}\lambda_i^2 \left|\zeta_k^{i,N}\right|^2 \left(e^{\frac{\ell^2}{N}\left|\zeta_k^{i,N}\right|^2} - 1\right)^2 \right] \\
&\lesssim \left(\mathbb{E}_{x_0} \sum_{i=1}^{N} i^{2s}\lambda_i^2 \left|\zeta_k^{i,N}\right|^4 \left(e^{\frac{\ell^2}{N}\left|\zeta_k^{i,N}\right|^2} - 1\right)^4 \right)^{1/2}.
\end{aligned}$$

Using the local Lipshitz property of the function e^x, we have

$$\begin{aligned}
&\mathbb{E}_{x_0}\sum_{i=1}^{N} i^{2s}\lambda_i^2\left|\zeta_k^{i,N}\right|^4\left(e^{\frac{\ell^2}{N}|\zeta_k^{i,N}|^2}-1\right)^4\\
&=\mathbb{E}_{x_0}\sum_{i=1}^{N} i^{2s}\lambda_i^2\left|\zeta_k^{i,N}\right|^4\left(e^{\frac{\ell^2}{N}|\zeta_k^{i,N}|^2}-1\right)^4\mathbf{1}_{\{\frac{\ell^2}{N}|\zeta_k^{i,N}|^2<\log\sqrt{N}\}}\\
&+\mathbb{E}_{x_0}\sum_{i=1}^{N} i^{2s}\lambda_i^2\left|\zeta_k^{i,N}\right|^4\left(e^{\frac{\ell^2}{N}|\zeta_k^{i,N}|^2}-1\right)^4\mathbf{1}_{\{\frac{\ell^2}{N}|\zeta_k^{i,N}|^2\geq\log\sqrt{N}\}}\\
&\lesssim e^{\log\sqrt{N}}\mathbb{E}_{x_0}\sum_{i=1}^{N} i^{2s}\lambda_i^2\frac{\left|\zeta_k^{i,N}\right|^6}{N}+\mathbb{E}_{x_0}\sum_{i=1}^{N} i^{2s}\lambda_i^2\left|\zeta_k^{i,N}\right|^4\left(e^{\frac{\ell^2\|\zeta_k^N\|^2}{N}}-1\right)^4\mathbf{1}_{\{\frac{\ell^2}{N}\|\zeta_k^N\|^2\geq\log\sqrt{N}\}}.
\end{aligned}$$

We now use Markov Inequality, (8.9) and (7.5) to estimate the second addend, obtaining

$$\begin{aligned}
&\mathbb{E}_{x_0}\sum_{i=1}^{N} i^{2s}\lambda_i^2\left|\zeta_k^{i,N}\right|^4\left(e^{\frac{\ell^2}{N}|\zeta_k^{i,N}|^2}-1\right)^4\\
&\leq\mathbb{E}_{x_0}\sum_{i=1}^{N} i^{2s}\lambda_i^2\frac{\left|\zeta_k^{i,N}\right|^6}{\sqrt{N}}+\left(\mathbb{P}\left\{\frac{\ell^2}{N}\|\zeta_k^N\|^2\geq\log\sqrt{N}\right\}\right)^{1/2}\\
&\lesssim\frac{1}{\sqrt{N}}+\left(\mathbb{E}_{x_0}e^{\frac{\ell^2}{N}\|\zeta_k^N\|^2}e^{-\log\sqrt{N}}\right)^{1/2}\overset{(7.5)}{\longrightarrow}0.
\end{aligned}$$

This concludes the proof. □

8.2. Analysis of the Noise

The proof of Lemma 8.2 is based on Lemma 8.10 below. In order to state such a lemma let us introduce the following notation and definitions. Let $k_N:[0,T]\to\mathbb{Z}_+$ 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 $\mathcal{F}^{k-1,N}\subset\mathcal{F}_k^N$. 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 $k_N(t)>\lim_{v\to 0+}k_N(t-v)$ and by linear interpolation otherwise. With this set up we state the following result.

Lemma 8.10. *Let $\mathfrak{T}:\mathcal{H}\to\mathcal{H}$ be a self-adjoint positive definite trace class operator on a separable Hilbert space $(\mathcal{H},\|\cdot\|)$. Suppose*

i) *there exists a continuous and positive function $f(t)$ defined on $[0,T]$ such that*

$$\lim_{N\to\infty}\sum_{k=1}^{k_N(T)}\mathbb{E}(\|X_k^N\|^2|\mathcal{F}_{k-1}^N)=\mathrm{Trace}(\mathfrak{T})\int_0^T f(t)dt\qquad\text{in probability;}$$

ii) *if $\{\phi_j\}$ is an orthonormal basis of $\mathcal{H}$ then*

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

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

$$\lim_{N\to\infty} \sum_{k=1}^{k_N(T)} \mathbb{E}(\|X_k^N\|^2)\mathbf{1}_{\{\|X_k^N\|^2\geq\epsilon\}}|\mathcal{F}_{k-1}^N) = 0, \qquad \textit{in probability.}$$

Then the sequence $\mathcal{X}^N$ *converges weakly in* $C([0,T];\mathcal{H}^s)$ *to the stochastic integral* $\int_0^T \sqrt{f(t)}dW_t$, *where* W_t *is a* $\mathcal{H}$*-valued* $\mathfrak{T}$*-Brownian motion.*

Proof. This lemma is in the same spirit as [MPS12, Proposition 4.1 and Remark 4.2]. As observed in [Ber86, Proof of Theorem 5.1], the statement just needs to be proved for a finite dimensional Hilbert space, i.e. in finite dimensions. The first two conditions are needed to ensure the weak convergence of the finite dimensional distributions of $\mathcal{X}^N$, the last condition guarantees tightness of the sequence, see [Hel82, Theorem 3.2] and [MPS12, Remark 4.2]. One may also consult the more compact [OPPS16, Section 5.5]. □

Proof of Lemma 8.2. We apply Lemma 8.10 with $k_N(t) = [tN]$, $X_k^N = M_k^{1,N}/\sqrt{N}$ and $\mathcal{F}_k^N$ the sigma-algebra generated by $\{\gamma_{h+1}^N, \xi_{h+1}^N,\, 0 \leq h \leq k\}$ to study the sequence $\eta^N(t)$, defined in (6.11), in the Hilbert space $\mathcal{H}^s$. We now check that the three conditions of Lemma 8.10 hold in the present case.

i) We need to show that

$$\frac{1}{N}\mathbb{E}_{x_0}\sum_{k=0}^{[TN]} \mathbb{E}_k\|M_k^{1,N}\|_s^2 \longrightarrow \mathrm{Trace}(\mathcal{C}_s)\int_0^T \Gamma_\ell(S(u))du\,. \tag{8.33}$$

From the definition of $M_k^{1,N}$, equation (6.10), we have

$$\frac{1}{N}\|M_k^{1,N}\|_s^2 = \|x_{k+1}^N - x_k^N - \mathbb{E}_k(x_{k+1}^N - x_k^N)\|_s^2, \tag{8.34}$$

hence

$$\begin{aligned}\frac{1}{N}\mathbb{E}_k\|M_k^{1,N}\|_s^2 &= \mathbb{E}_k\|x_{k+1}^N - x_k^N\|_s^2 - \|\mathbb{E}_k(x_{k+1}^N - x_k^N)\|_s^2 \\ &= \frac{2\ell^2}{N}\mathbb{E}_k\|\gamma_{k+1}\mathcal{C}^{1/2}\xi_{k+1}^N\|_s^2 - \|\mathbb{E}_k(x_{k+1}^N - x_k^N)\|_s^2,\end{aligned} \tag{8.35}$$

where the above equality holds thanks to (2.3). We will show that

$$\mathbb{E}_{x_0}\sum_{k=1}^{[TN]} \|\mathbb{E}_k(x_{k+1}^N - x_k^N)\|_s^2 \longrightarrow 0 \quad \text{as } N\to\infty. \tag{8.36}$$

Assuming the above for the moment, let us focus on the first addend in (8.35):

$$\begin{aligned}\frac{2\ell^2}{N}\mathbb{E}_k\|\gamma_{k+1}\mathcal{C}_N^{1/2}\xi_{k+1}^N\|_s^2 &= \frac{2\ell^2}{N}\sum_{j=1}^N \lambda_j^2 j^{2s}\mathbb{E}_k\left|\gamma_{k+1}\xi_{k+1}^{j,N}\right|^2 \\ &= \frac{2\ell^2}{N}\sum_{j=1}^N \lambda_j^2 j^{2s}\mathbb{E}_k\left[(1\wedge e^Q)\left|\xi_{k+1}^{j,N}\right|^2\right] \\ &= \frac{2\ell^2}{N}\sum_{j=1}^N \lambda_j^2 j^{2s}\mathbb{E}_k^\xi\left[(1\wedge e^{R_j})\left|\xi_{k+1}^{j,N}\right|^2\right] + \frac{a_{1,k}^N}{N},\end{aligned}$$

where

$$a_{1,k}^N := 2\ell^2\sum_{j=1}^N \lambda_j^2 j^{2s}\mathbb{E}_k\left[\left((1\wedge e^Q) - (1\wedge e^{R_j})\right)\left|\xi_{k+1}^{j,N}\right|^2\right]. \tag{8.37}$$

We now use the same technique that we used for the drift coefficient (that is, we first take expectation with respect to ξ^i and then with respect to $\xi \setminus \xi^i$), obtaining

$$\begin{aligned}\frac{2\ell^2}{N}\mathbb{E}_k\|\gamma_{k+1}\mathcal{C}^{1/2}\xi^N_{k+1}\|_s^2 &= \frac{2\ell^2}{N}\sum_{j=1}^N \lambda_j^2 j^{2s}\mathbb{E}_k^{\xi_j^-}(1\wedge e^{R_j}) + \frac{a^N_{1,k}}{N}\\ &= \frac{2\ell^2}{N}\sum_{j=1}^N \lambda_j^2 j^{2s}\mathbb{E}_k^{\xi}(1\wedge e^{Z_{\ell,k}}) + \frac{a^N_{1,k}}{N} + \frac{a^N_{2,k}}{N}\\ &= \frac{1}{N}\sum_{j=1}^N \lambda_j^2 j^{2s}\Gamma(S_k^N) + \frac{a^N_{1,k}}{N} + \frac{a^N_{2,k}}{N}, \end{aligned} \tag{8.38}$$

having used (5.13) and (5.28) and having set

$$a^N_{2,k} := 2\ell^2\sum_{j=1}^N \lambda_j^2 j^{2s}\mathbb{E}_k^{\xi}\left((1\wedge e^{R_j}) - (1\wedge e^{Z_{\ell,k}})\right). \tag{8.39}$$

Therefore

$$\frac{2\ell^2}{N}\sum_{k=0}^{[TN]}\mathbb{E}_k\|\gamma_{k+1}\mathcal{C}^{1/2}\xi^N_{k+1}\|_s^2 = \frac{1}{N}\sum_{k=0}^{[TN]}\sum_{j=1}^N \lambda_j^2 j^{2s}\Gamma(S_k^N) + \sum_{k=0}^{[TN]}\left(\frac{a^N_{1,k}}{N} + \frac{a^N_{2,k}}{N}\right). \tag{8.40}$$

If we prove that

$$\frac{1}{N}\mathbb{E}_{x_0}\sum_{k=0}^{[TN]}|a^N_{1,k}| \to 0 \quad \text{and} \quad \frac{1}{N}\mathbb{E}_{x_0}\sum_{k=0}^{[TN]}|a^N_{2,k}| \to 0, \tag{8.41}$$

then (8.33) follows from (8.35), (8.36), (8.40) and the above two limits. We therefore move on to proving the limits in (8.41). Let us start from the latter:

$$\begin{aligned}\mathbb{E}_{x_0}|a^N_{1,k}| &\lesssim \mathbb{E}_{x_0}\sum_{j=1}^N \lambda_j^2 j^{2s}\left(\mathbb{E}_k\left|(1\wedge e^{Q}) - (1\wedge e^{R_j})\right|^2\right)^{1/2}\\ &\lesssim \mathbb{E}_{x_0}\sum_{j=1}^N \lambda_j^2 j^{2s}\left(\mathbb{E}_k|Q-R_j|^2\right)^{1/2}\\ &\lesssim \mathbb{E}_{x_0}\sum_{j=1}^N \lambda_j^2 j^{2s}\left(\mathbb{E}_k|Q-R|^2\right)^{1/2} \qquad (8.42)\\ &+ \mathbb{E}_{x_0}\sum_{j=1}^N \lambda_j^2 j^{2s}\left(\mathbb{E}_k|R-R_i|^2\right)^{1/2}. \qquad (8.43)\end{aligned}$$

The addend (8.42) tends to zero as $N\to\infty$ by using (5.9) and (5.7). For (8.43) instead we have, by (5.17),

$$\begin{aligned}(8.43) &\lesssim \sum_{j=1}^N \lambda_j^2 j^{2s}\mathbb{E}_{x_0}\left(\mathbb{E}_k|R-R_i|^2\right)\\ &\lesssim \sum_{j=1}^N \lambda_j^2 j^{2s}\mathbb{E}_{x_0}\left(\frac{1}{N} + \frac{\left|\zeta_k^{j,N}\right|^2}{N}\right).\end{aligned}$$

The first limit in (8.41) now follows from (8.9). The second limit in (8.41) can be shown analogously, using this time the bounds (8.3) and (8.6).

Finally, to show (8.36), observe that from (8.13),

$$\|\mathbb{E}_k(x^N_{k+1}-x^N_k)\|_s^2 \lesssim \frac{\|\Theta(x^N_k,S^N_k)\|_s^2}{N^2}+\frac{\|e^N_k\|_s^2}{N^2}.$$

The desired result now follows from Lemma 8.8, (7.3) and the bound

$$\|\Theta(x,S)\|_s \lesssim 1+\|x\|_s,$$

(which is a consequence of the definition (6.7) and (8.17)).

ii) Condition **ii)** of Lemma 8.10 can be shown to hold with similar calculations, so we will not show the details.

iii) It will suffice to show that

$$\lim_{N\to\infty}\frac{1}{N}\mathbb{E}_{x_0}\sum_{k=0}^{[TN]}\mathbb{E}_k(\|M^{1,N}_k\|_s^2\mathbf{1}_{\{\|M^{1,N}_k\|_s^2>\epsilon N\}})=0\,.$$

Using the Markov inequality,

$$\begin{aligned}\mathbb{E}_k(\|M^{1,N}_k\|_s^2\mathbf{1}_{\{\|M^{1,N}_k\|_s^2>\epsilon N\}}) &\le \left(\mathbb{E}_k\|M^{1,N}_k\|_s^4\right)^{1/2}\left(\mathbb{P}\left\{\|M^{1,N}_k\|_s^2>\epsilon N\right\}\right)^{1/2}\\ &\le \frac{1}{N\epsilon}\mathbb{E}_k\|M^{1,N}_k\|_s^4\,.\end{aligned}$$

By (8.34), (2.3) and (3.7),

$$\frac{1}{N}\mathbb{E}_k\|M^{1,N}_k\|_s^4\lesssim\mathbb{E}_k\|(x^N_{k+1}-x^N_k)\|_s^4\lesssim\frac{1}{N^2}\mathbb{E}_k\|\gamma_{k+1}\mathcal{C}^{1/2}\xi^N_{k+1}\|_s^4\lesssim\frac{1}{N^2}.$$

Therefore

$$\frac{1}{N}\mathbb{E}_{x_0}\sum_{k=0}^{[TN]}\mathbb{E}_k(\|M^{1,N}_k\|_s^2\mathbf{1}_{\{\|M^{1,N}_k\|_s^2>\epsilon N\}})\lesssim\frac{1}{N^2\epsilon}\mathbb{E}_{x_0}\sum_{k=0}^{[TN]}\mathbb{E}_k(\|M^{1,N}_k\|_s^4)\lesssim\frac{1}{N^2}\,. \tag{8.44}$$

□

Appendix A

Proof of (3.7). We just need to prove it for p even. So let $q\ge 1$; then, by the weighted Jensen's inequality

$$\mathbb{E}\|\mathcal{C}^{1/2}\xi^N\|_s^{2q}\le\mathbb{E}\left(\sum_{i=1}^\infty i^{2s}\lambda_i^2\left|\xi^{i,N}\right|^2\right)^q\le(Trace(\mathcal{C}_s))^{q-1}\sum_{i=1}^\infty i^{2s}\lambda_i^2\mathbb{E}\left|\xi^{i,N}\right|^{2q}<\infty.$$

Alternatively, one can observe that (3.7) is a consequence of Fernique's theorem. □

Before proving Lemma 4.2, we recall the following fact, which has already been pointed out in Section 1.2.

Remark 8.11. We recall that for $X\in\mathbb{R}_+$ and $b\in\mathbb{R}$, $D_\ell(X)=\mathcal{G}_{\ell\sqrt{2}}(X,1)$, where $\mathcal{G}_\ell(X,b)$ is the drift function defined in [JLM15, (1.7)]. Analogously, our $\Gamma_\ell(X)$ is $\Gamma_{\ell\sqrt{2}}(X,1)$, where $\Gamma_\ell(X,b)$ is defined in [JLM15, (1.6)]. □

Proof of Lemma 4.2. The boundedness of D_ℓ and Γ_ℓ follows from Remark 8.11 and [JLM15, Lemma 2]. Lipshitzianity follows simply observing that both functions have bounded derivative, indeed

$$\begin{aligned}\frac{d}{dx}D_\ell(x)&=\ell^2D_\ell(x)-\frac{\ell^3}{\sqrt{\pi}}\left(\frac{1}{\sqrt{x}}+\frac{1}{2\,x^{3/2}}\right)e^{-\frac{\ell^2}{4x}}\\ \frac{d}{dx}\Gamma_\ell(x)&=\ell^2D_\ell(x)-\frac{\ell^3}{\sqrt{\pi x}}e^{-\frac{\ell^2}{4x}}.\end{aligned} \tag{8.45}$$

Global Lipshitzianity of $\sqrt{\Gamma_\ell}$ then follows after observing that Γ_ℓ is bounded below away from zero (see (1.13)).

We now want to show that the derivative of $A_\ell(x)$ is bounded. From the definition of A_ℓ (equation (1.14)) we have

$$\partial_x A_\ell(x) = -2D_\ell(x) - 2x\,\partial_x D_\ell(x) + \partial_x \Gamma_\ell(x)\,. \tag{8.46}$$

We will prove that

$$\lim_{x\to+\infty} \partial_x A_\ell(x) = 0\,. \tag{8.47}$$

Because $\partial_x A_\ell$ is a continuous function on $[0,+\infty)$, (8.47) implies the boundedness of $\partial_x A_\ell(x)$. In order to prove (8.47) we will prove that all the addends on the RHS of (8.46) tend to zero (see also Figure 2 below).

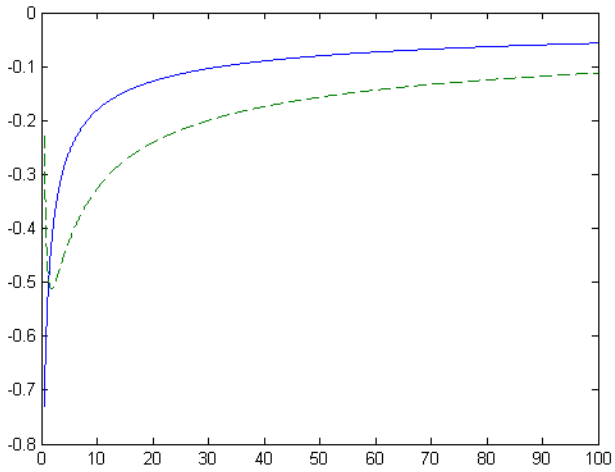

FIG 2. *Plots of the function $\partial_x A_\ell(x)$ for $\ell = 1$ and $\ell = 2$ (dashed line).*

- First of all, let us prove

$$\lim_{x\to+\infty} D_\ell(x) = 0. \tag{8.48}$$

The above limit follows from the definition of D_ℓ (1.12) by simply applying de l'Hopital's rule:

$$\begin{aligned}\lim_{x\to+\infty} \frac{\Phi\left(\frac{\ell(1-2x)}{\sqrt{2x}}\right)}{e^{-\ell^2(x-1)}} &= \lim_{x\to+\infty} \frac{e^{-\ell^2(1-2x)^2/4x}}{\sqrt{2\pi}\,\ell^2 e^{-\ell^2(x-1)}}\left(\frac{\ell}{2\sqrt{2}x^{3/2}} + \frac{\ell}{\sqrt{2x}}\right) \\ &= \lim_{x\to+\infty} e^{-\ell^2/4x}\frac{1}{\ell\sqrt{\pi}}\left(\frac{1}{4x^{3/2}} + \frac{1}{2\sqrt{x}}\right) = 0.\end{aligned} \tag{8.49}$$

- From (8.45) and (8.48), also $\partial_x\Gamma_\ell(x) \to 0$ as $x \to +\infty$.
- Now the second addend:

$$\lim_{x\to+\infty} -2x\,\partial_x D_\ell(x) = 0.$$

Indeed,

$$-2x\partial_x D_\ell = -2x\ell^2 D_\ell(x) + 2\frac{\ell^3}{\sqrt{\pi}}\left(\sqrt{x} + \frac{1}{2\sqrt{x}}\right)e^{-\ell^2/4x}\,,$$

therefore

$$\begin{aligned}\lim_{x\to+\infty} -2x\partial_x D_\ell &= \lim_{x\to+\infty} -4\ell^4 \frac{x\,\Phi\left(\frac{\ell(1-2x)}{\sqrt{2x}}\right)}{e^{-\ell^2(x-1)}} + 2\frac{\ell^3}{\sqrt{\pi}}\sqrt{x}e^{-\ell^2/4x} \\ &= \lim_{x\to+\infty} -4\ell^4 \left[\frac{\Phi\left(\frac{\ell(1-2x)}{\sqrt{2x}}\right) - e^{-\ell^2(1-2x)^2/4x}\left(\frac{\ell}{4\sqrt{\pi x}} + \frac{\ell\sqrt{x}}{2\sqrt{\pi}}\right)}{-\ell^2 e^{-\ell^2(x-1)}}\right] \\ &+ \lim_{x\to+\infty} 2\frac{\ell^3}{\sqrt{\pi}}\sqrt{x}e^{-\ell^2/4x} \\ &\overset{(8.49)}{=} \lim_{x\to+\infty} -2\frac{\ell^3}{\sqrt{\pi}}\sqrt{x}e^{-\ell^2/4x} + 2\frac{\ell^3}{\sqrt{\pi}}\sqrt{x}e^{-\ell^2/4x} = 0\end{aligned}$$

Finally, the sign of $A_\ell(x)$ is studied in [CRR05, page 258]. □

We include here plots of the functions $D_\ell(x)$ and $\Gamma_\ell(x)$, Figure 3 and Figure 4 below.

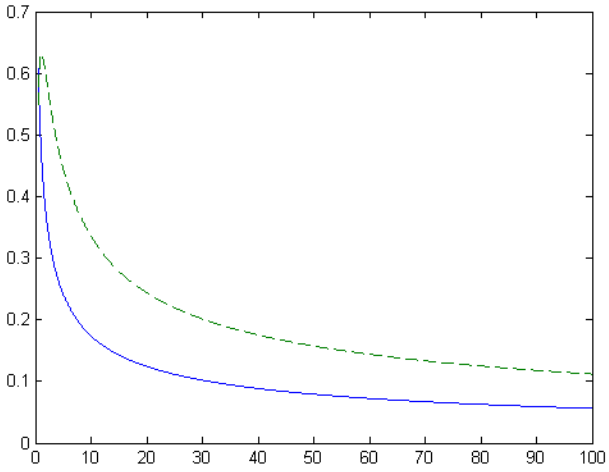

FIG 3. *Plots of the function $D_\ell(x)$ for $\ell = 1$ and $\ell = 2$ (dashed line).*

Proof of first equality in (5.16). We want to prove

$$\mathbb{E}_k^{\xi,\gamma}(\gamma_{k+1}\,\xi_{k+1}^{i,N}) = \mathbb{E}_k^{\xi}(\alpha_{k+1}\,\xi_{k+1}^{i,N}).$$

Let $f_{\gamma_{k+1},\xi_{k+1}}(\gamma,\xi)$ be the joint distribution (given x_k) of γ_{k+1} and ξ_{k+1}. Then

$$\begin{aligned}\mathbb{E}_k^{\xi,\gamma}(\gamma_{k+1}\,\xi_{k+1}^{i,N}) &= \iint \gamma\,\xi^i\, f_{\gamma_{k+1},\xi_{k+1}}(\gamma,\xi) = \int \xi^i \int \gamma f_{\xi_{k+1}}(\xi) f_{\gamma_{k+1}|\xi_{k+1}}(\gamma|\xi) \\ &= \int \xi^i f_{\xi_{k+1}}(\xi)\alpha_{k+1}(\xi) = \mathbb{E}_k^{\xi}(\alpha_{k+1}\,\xi_{k+1}^{i,N}).\end{aligned}$$

□

Appendix B

Before starting the proofs of the various lemmata, we derive equation (8.51) below, which will be repeatedly used throughout this appendix.

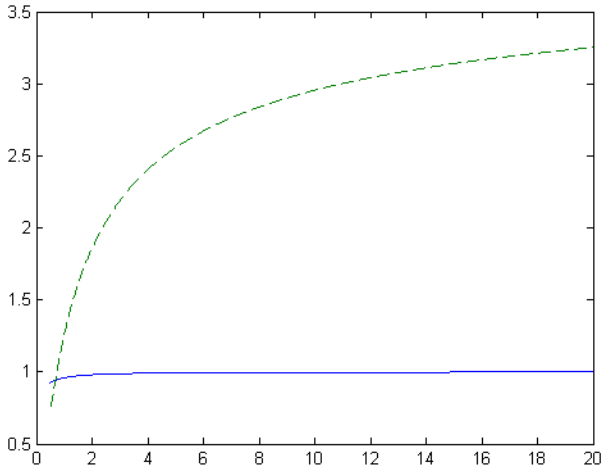

FIG 4. *Plots of the function $\Gamma_\ell(x)$ for $\ell = 1$ and $\ell = 2$ (dashed line).*

From (5.5) and recalling that $\left[\nabla\Psi^N(x_k)\right]^i$ denotes the i-th component of $\nabla\Psi^N(x_k)$,

$$\zeta_k^{i,N} = \frac{x_k^{i,N}}{\lambda_i} + \lambda_i \left[\nabla\Psi^N(x_k)\right]^i . \tag{8.50}$$

Using the bound (7.12) we have

$$\left|\lambda_i \left[\nabla\Psi^N(x_k)\right]^i\right|^2 \leq \sum_{i=1}^{\infty} \left|\lambda_i \left[\nabla\Psi^N(x_k)\right]^i\right|^2 \lesssim \|x\|_s^{2\varsigma} + \|x\|_s^2,$$

hence

$$\left|\lambda_i \left[\nabla\Psi^N(x_k)\right]^i\right|^p \lesssim \|x\|_s^{\varsigma p} + \|x\|_s^p.$$

Therefore for every $p \geq 0$

$$\left|\zeta_k^{i,N}\right|^p \lesssim \frac{\left|x_k^{i,N}\right|^p}{\lambda_i^p} + (\|x_k\|_s^{\varsigma p} + \|x_k\|_s^p), \tag{8.51}$$

Proof of Lemma 7.4. We will prove, in order, the bounds (7.4), (7.3) and (7.5).

- **Proof of** (7.4). We can act as in [PST14, Proof of Lemma 9] (in comparing our proof with [PST14, Proof of Lemma 9] set $\delta = N^{-1}$ in [PST14]). Looking at [PST14, Proof of Lemma 9], all we need to show is

$$\mathbb{E}_{x_0}(S_{k+1}^N)^{2m} - \mathbb{E}_{x_0}(S_k^N)^{2m} \lesssim \frac{1}{N}(1 + \mathbb{E}_{x_0}(S_k^N)^{2m}).$$

A close inspection of the method of proof used in [PST14] reveals that showing the above boils down to proving the following two bounds:

$$\left|\mathbb{E}_k\left[S_{k+1}^N - S_k^N\right]\right| \lesssim \frac{1}{N}\left(1 + S_k^N\right), \quad p \geq 1. \tag{8.52}$$

and

$$\left(\mathbb{E}_k\left|S_{k+1}^N - S_k^N\right|^p\right)^{1/p} \lesssim \frac{1}{\sqrt{N}}(1 + S_k^N)\,. \tag{8.53}$$

Let us start with proving (8.52). To this end let us observe that by (7.7) and (6.1), one has

$$\mathbb{E}_k\left[S_{k+1}^N - S_k^N\right] = \frac{E_k^N}{N} + \frac{A_\ell(S_k^N)}{N}\,. \tag{8.54}$$

Now notice that

$$\frac{\|x_k^N\|_s^2}{N} = \frac{1}{N}\sum_{i=1}^N i^{2s}\left|x_k^{i,N}\right|^2 = \frac{1}{N}\sum_{i=1}^N i^{2s}\lambda_i^2\frac{\left|x_k^{i,N}\right|^2}{\lambda_i^2} \lesssim \frac{1}{N}\sum_{i=1}^N \frac{\left|x_k^{i,N}\right|^2}{\lambda_i^2} = S_k^N\,, \tag{8.55}$$

where in the last inequality we have used the fact that $\sum_i \lambda_i^2 i^{2s}$ is convergent and therefore $\lambda_i^2 i^{2s}$ is bounded. To bound the RHS of (8.54), we recall that from the proof of Lemma 7.2 one has $E_k^N = E_{1,k}^N + E_{2,k}^N + \mathbb{E}_k \hat{r}^N$ (see (7.10)). Acting like we did to obtain (7.17), one has

$$\left|E_{1,k}^N\right| \lesssim \frac{1}{\sqrt{N}}\left(S_k^N + \frac{\|x_k^N\|_s^2}{N}\right)^{1/2} \overset{(8.55)}{\lesssim} \frac{1}{\sqrt{N}}(S_k^N + 1). \tag{8.56}$$

With steps analogous to those used to obtain (7.19), one also has

$$\left|E_{2,k}^N\right| \lesssim \frac{1}{\sqrt{N}}\left(1 + \|x_k^N\|_s^2\right)^{1/2} \lesssim (1 + S_k^N). \tag{8.57}$$

Now (8.52) follows from (7.10), (8.54), (8.56), (8.57), (8.55), (7.21) and

$$A_\ell(a) \lesssim (1+a), \qquad a \geq 0\,.$$

To prove (8.53) one can instead just use (7.23), (7.15) (together with $\gamma_{k+1} \leq 1$) and calculations analogous to those leading to (7.21). This concludes the proof of (7.4).

- **Proof of** (7.3). For this bound we will use the same strategy of proof that we used to show (7.4). So we only need to prove

$$\|\mathbb{E}_k(x_{k+1}^N - x_k^N)\|_s \lesssim \frac{1}{N}(1 + \|x_k^N\|_s) \tag{8.58}$$

and

$$\left(\mathbb{E}_k\|x_{k+1}^N - x_k^N\|_s^p\right)^{1/p} \lesssim \frac{1}{\sqrt{N}}(1 + \|x_k^N\|_s)\,. \tag{8.59}$$

Let us start with (8.58):

$$\begin{aligned}\|\mathbb{E}_k(x_{k+1}^N - x_k^N)\|_s &= \sqrt{\frac{2\ell^2}{N}}\|\mathbb{E}_k(\gamma_{k+1}\mathcal{C}^{1/2}\xi_{k+1}^N)\|_s \\ &= \sqrt{\frac{2\ell^2}{N}}\left(\sum_{i=1}^N i^{2s}\left|\mathbb{E}_k(\gamma_{k+1}\lambda_i\xi_{k+1}^{i,N})\right|^2\right)^{1/2}\end{aligned}$$

We therefore need to estimate $\left|\mathbb{E}_k(\gamma_{k+1}\lambda_i\xi_{k+1}^{i,N})\right|^2$. In order to do so, we make the following preliminary observation: from (5.7) and (5.17) we have

$$Q(x_k, \xi_{k+1}) = R^i(x_k, \xi_{k+1}) - \frac{\ell^2}{N}\left|\xi_{k+1}^{i,N}\right|^2 - \sqrt{\frac{2\ell^2}{N}}\zeta_k^{i,N}\xi_{k+1}^{i,N} + r^N(x_k, \xi_{k+1}).$$

As we have already said, R^i contains only terms that do not depend on the noise $\xi_{k+1}^{i,N}$, therefore we can write

$$\begin{aligned}\left|\mathbb{E}_k(\gamma_{k+1}\lambda_i\xi_{k+1}^{i,N})\right|^2 &= \left|\mathbb{E}_k\left[\left(1\wedge e^Q\right)\lambda_i\xi_{k+1}^{i,N}\right]\right|^2 \\ &= \left|\mathbb{E}_k\left[\left(\left(1\wedge e^Q\right) - \left(1\wedge e^{R^i}\right)\right)\lambda_i\xi_{k+1}^{i,N}\right]\right|^2 \\ &\leq \lambda_i^2\left|\mathbb{E}_k\left(\left|Q - R^i\right|\left|\xi_{k+1}^{i,N}\right|\right)\right|^2 \\ &\lesssim \lambda_i^2\left|\mathbb{E}_k\frac{\left|\xi_{k+1}^{i,N}\right|^3}{N} + \mathbb{E}_k\frac{\left|\zeta_k^{i,N}\right|\left|\xi_{k+1}^{i,N}\right|^2}{\sqrt{N}} + \mathbb{E}_k\left|\xi_{k+1}^{i,N}r^N\right|\right|^2.\end{aligned}$$

Using (5.8) and the fact that ζ_k depends only on x_k, we have

$$\left|\mathbb{E}_k(\gamma_{k+1}\lambda_i\xi_{k+1}^{i,N})\right|^2 \lesssim \lambda_i^2\left(\frac{1}{N^2}+\frac{\left|\zeta_k^{i,N}\right|^2}{N}\right) \overset{(8.51)}{\lesssim} \lambda_i^2\left(\frac{1}{N^2}+\frac{\left|x_k^{i,N}\right|^2}{\lambda_i^2 N}+\frac{\|x_k\|_s^2+1}{N}\right). \tag{8.60}$$

(8.58) is now a simple consequence of the above bound. For (8.59), instead, we just use $\gamma_{k+1}\leq 1$ and

$$\left(\mathbb{E}_k\|x_{k+1}^N-x_k^N\|_s^p\right)^{1/p} \lesssim \left(\frac{1}{N^{p/2}}\mathbb{E}_k\|\mathcal{C}^{1/2}\xi_{k+1}\|_s^p\right)^{1/p} \overset{(3.7)}{\lesssim} \frac{1}{\sqrt{N}}.$$

- **Proof of** (7.5). By acting as we do to obtain (8.51) (with $p=2$), it is clear that we only need to show

$$\mathbb{E}_{x_0}e^{cS_k^N}<\infty \quad \text{and} \quad \mathbb{E}_{x_0}e^{\frac{c}{N}\|x_k^N\|_s^2}<\infty, \quad \text{uniformly over } 0\leq k\leq [TN]+1, \tag{8.61}$$

for all $c>0$. However, by (8.55), proving the second of the above bounds boils down to proving the first, which is therefore the only one we need to concentrate on. Such a bound is a simple consequence of (7.3). Indeed, on inspection of the proof of (7.3), one finds that the constants $\bar{c}$ appearing on the RHS of (7.3) grows at most like d^m, where $d>0$ is some positive constant independent of m, N and k. Therefore

$$\mathbb{E}_{x_0}e^{\frac{c}{N}\|x_k^N\|_s^2}=\sum_{m=0}^{\infty}c^m\mathbb{E}_{x_0}\frac{\|x_k^N\|_s^{2m}}{N^m m!}=e^{c\,d^2/N}\lesssim 1.$$

This concludes the proof of the lemma. □

Proof of Lemma 5.3. This proof is in 2 steps. The first step proves the first part of the statement, the second step proves the second part.

- *Step 1.* For every fixed $t>0$ and for every $\epsilon>0$,

$$\sum_{N=1}^{\infty}\mathbb{P}\left(\left|\hat{w}^N(t)\right|>\epsilon\right)<\frac{1}{\epsilon^4}\sum_{N=1}^{\infty}\mathbb{E}_{x_0}\left(\hat{w}^N(t)\right)^4<\infty, \tag{8.62}$$

where $\hat{w}^N$ has been defined in (6.5). Assuming for the moment that (8.62) holds, by the Borel-Cantelli Lemma (8.62) implies that $\hat{w}^N(t)$ converges to zero almost surely. Because almost sure convergence is preserved under continuous transformations, this means that $S^{(N)}(t)$ converges almost surely to $S(t)$. We only sketch the proof of (8.62), as the calculations are completely analogous to those contained in the proof Theorem 5.1. From (6.5), we have

$$\begin{aligned}\mathbb{E}_{x_0}\left(\hat{w}^N(t)\right)^4 \lesssim& \int_0^t\mathbb{E}_{x_0}\left[A_\ell^N(\bar{x}^{(N)}(v))-A_\ell(\bar{S}^{(N)}(v))\right]^4dv\\ &+\int_0^t\mathbb{E}_{x_0}\left[A_\ell(\bar{S}^{(N)}(v))-A_\ell(S^{(N)}(v))\right]^4dv+\mathbb{E}_{x_0}\left|w^N(t)\right|^4.\end{aligned} \tag{8.63}$$

The estimate of the first and third addend on the right hand side of the above is done by proceeding analogously to what we have done for the proof of Lemma 7.2 and Lemma 7.1, respectively. The second addend can be studied with similar calculations (indeed, with calculations analogous to those in Step 2 of this proof). Therefore we only show how to estimate the first addend, the others can be done with a similar procedure, and we leave it to the reader. With the notation introduced in the proof of Lemma 7.2, we have

$$\mathbb{E}_{x_0}\int_0^t\left[A_\ell^N(\bar{x}^{(N)}(v))-A_\ell(\bar{S}^{(N)}(v))\right]^4dv=\mathbb{E}_{x_0}\frac{1}{N}\sum_{k=0}^{[TN]}\left|E_k^N\right|^4$$

and $\left|E_k^N\right|^4 \lesssim \left|E_{1,k}^N\right|^4 + \left|E_{2,k}^N\right|^4 + \left|\mathbb{E}_k \hat{r}^N\right|^4$ (see (7.10)). Acting as we did to obtain (7.18) and (7.19), we find

$$\mathbb{E}_{x_0} \left|E_{1,k}^N\right|^4 \lesssim \frac{1}{N^2} \qquad \text{and} \qquad \mathbb{E}_{x_0} \left|E_{2,k}^N\right|^4 \lesssim \frac{1}{N^2}. \tag{8.64}$$

Using (7.21), one finds that, overall,

$$\mathbb{E}_{x_0} \int_0^t \left[A_\ell^N(\bar{x}^{(N)}(v)) - A_\ell(\bar{S}^{(N)}(v))\right]^4 dv \lesssim \frac{1}{N^2},$$

and the sequence $a_N = N^{-2}$ is summable. Similar estimates can be obtained for the addends in (8.63). This concludes the proof of the almost sure convergence of $\hat{w}^N$ to zero.

- *Step 2.* For every $\epsilon > 0$,

$$\sum_{N=1}^{\infty} \mathbb{P}\left(\left|\bar{S}^{(N)}(t) - S^{(N)}(t)\right| > \epsilon\right) < \frac{1}{\epsilon^2} \sum_{N=1}^{\infty} \mathbb{E}_{x_0} \left|\bar{S}^{(N)}(t) - S^{(N)}(t)\right|^2 < \infty\,. \tag{8.65}$$

Again, if we prove the above, by the B-C Lemma we have almost sure convergence of $\bar{S}^{(N)}(t)$ to $S^{(N)}(t)$ and, by Step 1, to $S(t)$. From the definitions of $\bar{S}^{(N)}(t)$ and $S^{(N)}(t)$, equation (1.17) and (5.2), respectively, for $(k/N) \le t < (k+1)/N$, we have

$$\bar{S}^{(N)}(t) - S^{(N)}(t) = (Nt - k)(S_{k+1}^N - S_k^N), \tag{8.66}$$

so that

$$\mathbb{E}_{x_0} \left|\bar{S}^{(N)}(t) - S^{(N)}(t)\right|^2 = \mathbb{E}_{x_0} \left|(Nt-k)(S_{k+1}^N - S_k^N)\right|^2 \le \mathbb{E}_{x_0} \left|S_{k+1}^N - S_k^N\right|^2 \overset{(7.24)}{\lesssim} \frac{1}{N^2}. \tag{8.67}$$

This concludes the proof.

□

Proof of Lemma 8.1 . Using (8.2), the bound (8.4) is a simple consequence of the definitions of R and G, indeed

$$|R - G| \lesssim \left|\frac{1}{N}\sum_{i=1}^N \left|\xi^i\right|^2 - 1\right|,$$

hence

$$\begin{aligned}\mathbb{E}_k |R - G| \le \left(\mathbb{E}_k |R-G|^2\right)^{1/2} &\lesssim \left(\mathbb{E}\left|\frac{1}{N}\sum_{i=1}^N \left|\xi^i\right|^2 - 1\right|^2\right)^{1/2} \\ &= \left(Var\left(\frac{1}{N}\sum_{i=1}^N \left|\xi^i\right|^2\right)\right)^{1/2} \le \frac{1}{\sqrt{N}}.\end{aligned} \tag{8.68}$$

We observe now (although we will use this only later) that a similar explicit calculation also shows

$$\mathbb{E}_k |R - G|^4 \lesssim \frac{1}{N^2}. \tag{8.69}$$

Going back to the proof of the lemma, the bound (8.3) is a direct consequence of the definitions of R and R^i. The inequality (8.5) follows from (5.12), (5.13), (5.5) and the bound (7.13). Indeed,

$$G - Z_{\ell,k} = \sqrt{\frac{2\ell^2}{N}} \sum_{j=1}^N \zeta_k^{j,N} \xi_{k+1}^{j,N} - \sqrt{\frac{2\ell^2}{N}} \sum_{j=1}^N \frac{x_k^{j,N}}{\lambda_i} \xi_{k+1}^{j,N}.$$

Therefore, by (5.5), given x_k we have

$$G - Z_{\ell,k} \sim \mathcal{N}\left(0, \frac{2\ell^2}{N}\|\mathcal{C}_N^{1/2}\nabla\Psi^N(x_k^N)\|^2\right). \tag{8.70}$$

Using (7.13) one then has

$$\mathbb{E}_k\,|G - Z_{\ell,k}|^p \lesssim \frac{1+\|x_k^N\|_s^p}{N^{p/2}}, \tag{8.71}$$

hence (8.5) follows. Notice that from the above calculations we have

$$\mathbb{E}_k\,|R - Z_{\ell,k}|^p \lesssim \frac{1+\|x_k^N\|_s^p}{N^{p/2}}, \quad p \in \{2,4\}. \tag{8.72}$$

□

Proof of Lemma 8.6. We prove the three statements in the order in which they are presented.

- Proof of the bound (8.9). From (8.51),

$$\mathbb{E}_{x_0}\sum_{i=1}^N i^{2s}\lambda_i^2\left|\zeta_k^{i,N}\right|^p Y \lesssim \mathbb{E}_{x_0}\sum_{i=1}^N i^{2s}\lambda_i^2\frac{\left|x_k^{i,N}\right|^p}{\lambda_i^p}Y + \sum_{i=1}^N i^{2s}\lambda_i^2\mathbb{E}_{x_0}[(1+\|x_k\|_s^p)Y]\,. \tag{8.73}$$

The second addend is bounded thanks to the assumption on Y and (7.3). As for the first addend, (by the weighted Jensen's inequality) this is bounded (for any $p \geq 0$) as soon as we can prove that

$$v_k^N(p) := \sum_{i=1}^N i^{2s}\lambda_i^2\frac{\left|x_k^{i,N}\right|^{2p}}{\lambda_i^{2p}}$$

has bounded first moment (for every p), i.e. we want to prove $\mathbb{E}_{x_0}v_k^N(p) < c$ where $c > 0$ is a constant independent of N and $k \in \{0,1,\dots,[TN]\}$ (but possibly dependent on p). Observe that if $p = 1$ then $v_k^N(1) = \|x_k^N\|_s^2$, so the statement is a consequence of (7.3). So we can restrict to $p \geq 2$. Denoting by d a generic constant (that does not depend on N), the value of which will change from line to line, we write

$$\begin{aligned}\mathbb{E}_{x_0}v_{k+1}^N(p) &= \mathbb{E}_{x_0}\sum_{i=1}^N i^{2s}\frac{\lambda_i^2}{\lambda_i^{2p}}\left(x_k^{i,N} + \sqrt{\frac{2\ell^2}{N}}\lambda_i\gamma_{k+1}\xi_{k+1}^{i,N}\right)^{2p}\\ &\leq \mathbb{E}_{x_0}v_k^N(p) + \mathbb{E}_{x_0}d\sum_{i=1}^N i^{2s}\frac{\lambda_i^2}{\lambda_i^{2p}}\sum_{m=0}^{2p-1}(x_k^{i,N})^m\mathbb{E}_k\left(\sqrt{\frac{2\ell^2}{N}}\lambda_i\gamma_{k+1}\xi_{k+1}^{i,N}\right)^{2p-m}.\end{aligned}$$

If, in the above summation, the index m is smaller than $2p-2$, i.e. $0 \leq m \leq 2p-2$, then $p - m/2 \geq 1$ and we have the estimate

$$\begin{aligned}\mathbb{E}_{x_0}i^{2s}\frac{\lambda_i^2}{\lambda_i^{2p}}(x_k^{i,N})^m\mathbb{E}_k\left(\sqrt{\frac{2\ell^2}{N}}\lambda_i\gamma_{k+1}\xi_{k+1}^{i,N}\right)^{2p-m} &\lesssim \mathbb{E}_{x_0}i^{2s}\frac{\lambda_i^2}{N}\frac{\left|x_k^{i,N}\right|^m}{\lambda_i^m}\mathbb{E}_k\left|\gamma_{k+1}\xi_{k+1}^{i,N}\right|^{2p-m}\\ &\lesssim \mathbb{E}_{x_0}i^{2s}\frac{\lambda_i^2}{N}\frac{\left|x_k^{i,N}\right|^m}{\lambda_i^m} \qquad (8.74)\\ &\lesssim \frac{1}{N}i^{2s}\lambda_i^2\mathbb{E}_{x_0}\left(\frac{\left|x_k^{i,N}\right|^{2p}}{\lambda_i^{2p}} + 1\right). \qquad (8.75)\end{aligned}$$

If $m = 2p-1$ we instead use (8.60) and obtain

$$\mathbb{E}_{x_0} i^{2s} \frac{\lambda_i^2}{\lambda_i^{2p}} \frac{1}{\sqrt{N}} \left| x_k^{i,N} \right|^{2p-1} \left| \mathbb{E}_k (\lambda_i \gamma_k \xi_{k+1}^{i,N}) \right| \lesssim \mathbb{E}_{x_0} i^{2s} \lambda_i^2 \frac{\left| x_k^{i,N} \right|^{2p-1}}{\lambda_i^{2p-1}} \frac{1}{\sqrt{N}} \left(\frac{\|x_k^N\|_s}{\sqrt{N}} + \frac{\left| x_k^{i,N} \right|}{\lambda_i} \frac{1}{\sqrt{N}} \right)$$
$$\lesssim \mathbb{E}_{x_0} \frac{i^{2s} \lambda_i^2}{N} \left(\frac{\left| x_k^{i,N} \right|^{2p}}{\lambda_i^{2p}} + \|x\|_s^{2p} \right). \tag{8.76}$$

To obtain the last inequality we used Young's inequality with exponents $2p/(2p-1)$ and $2p$, as follows:

$$\frac{\left| x_k^{i,N} \right|^{2p-1}}{\lambda_i^{2p-1}} \cdot \|x_k^N\|_s \leq \frac{\left| x_k^{i,N} \right|^{2p}}{\lambda_i^{2p}} + \|x_k^N\|_s^{2p}.$$

From (8.75) and (8.76) (and using (7.3)) we then have

$$\mathbb{E}_{x_0} v_{k+1}^N(p) \leq \left(1 + \frac{d}{N}\right) \mathbb{E}_{x_0} v_k^N(p) + \frac{d}{N}.$$

Iterating the above $[TN]$ times we get

$$\mathbb{E}_{x_0} v_{k+1}^N(p) \leq \left(1 + \frac{d}{N}\right)^{[TN]} v_0(p) + d < \infty,$$

having denoted $v_0(p) := \sum_{i=1}^\infty i^{2s} \lambda_i^2 \frac{\left| x_0^i \right|^{2p}}{\lambda_i^{2p}}$. Notice that if $x_0 \in \mathcal{H}_{\cap\cap}^s$ then the series $v_0(p)$ is convergent for every $p \geq 0$.

- Proof of the bound (8.10). Set

$$\tilde{\zeta}_k^{i,N} := \frac{\left| x_k^{i,N} \right|}{\lambda_i}, \tag{8.77}$$

so that

$$S_k^N = \frac{1}{N} \sum_{i=1}^N \frac{\left| x_k^{i,N} \right|^2}{\lambda_i^2} = \frac{1}{N} \sum_{i=1}^N \left| \tilde{\zeta}_k^{i,N} \right|^2.$$

From (8.51) we then have

$$J_k^N := \sum_{i=1}^N i^{2s} \lambda_i^2 \frac{\left| \zeta_k^{i,N} \right|^p}{(N S_k^N)^\alpha}$$
$$\lesssim \sum_{i=1}^N i^{2s} \lambda_i^2 \frac{\left| \tilde{\zeta}_k^{i,N} \right|^p}{(N S_k^N)^\alpha} + \sum_{i=1}^N i^{2s} \lambda_i^2 \frac{\|x_k^N\|_s^{p\varsigma} + \|x_k^N\|_s^p}{(N S_k^N)^\alpha}. \tag{8.78}$$

For the first addend in (8.78) we have

$$\mathbb{E}_{x_0} \sum_{i=1}^N i^{2s} \lambda_i^2 \frac{\left| \tilde{\zeta}_k^{i,N} \right|^p}{(N S_k^N)^\alpha} = \mathbb{E}_{x_0} \sum_{i=1}^N i^{2s} \lambda_i^2 \left| \tilde{\zeta}_k^{i,N} \right|^{p-2\alpha} \left(\frac{\left| \tilde{\zeta}_k^{i,N} \right|^2}{\sum_{i=1}^N \left| \tilde{\zeta}_k^{i,N} \right|^2} \right)^\alpha$$
$$\lesssim \mathbb{E}_{x_0} v_k^N((p-2\alpha)/2),$$

thanks to the obvious estimate $\left(\left|\tilde{\zeta}_k^{i,N}\right|^2 / \sum_{i=1}^N \left|\tilde{\zeta}_k^{i,N}\right|^2\right) \leq 1$. Using (8.55), one can easily see that also the expected value of the second addend is bounded if $\varsigma p \geq 2\alpha$, as

$$\mathbb{E}_{x_0} \frac{\|x_k^N\|_s^{p\varsigma}}{(NS_k^N)^\alpha} = \mathbb{E}_{x_0} \frac{\|x_k^N\|_s^{2\alpha}}{(NS_k^N)^\alpha} \|x_k^N\|_s^{\varsigma p - 2\alpha} \lesssim \mathbb{E}_{x_0} \|x_k^N\|_s^{\varsigma p - 2\alpha} < \infty.$$

Therefore, $\mathbb{E}_{x_0} J_k^N < \infty$, uniformly over k and N. From the weighted Jensen inequality one can also see that $\mathbb{E}_{x_0} \left|J_k^N\right|^q < \infty$ for all $q \geq 1$. To conclude, for $t_k \leq t < t_{k+1}$ we write

$$\mathbb{E}_{x_0} J_k^N = \mathbb{E}_{x_0} J_k^N \mathbf{1}_{\{S_k^N \geq (S(t)/2)\}} + \mathbb{E}_{x_0} J_k^N \mathbf{1}_{\{S_k^N < (S(t)/2)\}}.$$

Now the first addend:

$$\begin{aligned}
\mathbb{E}_{x_0} J_k^N \mathbf{1}_{\{S_k^N \geq (S(t)/2)\}} &= \mathbb{E}_{x_0} \frac{1}{N^\alpha} \sum_{i=1}^N i^{2s} \lambda_i^2 \frac{\left|\zeta_k^{i,N}\right|^p}{(S_k^N)^\alpha} \mathbf{1}_{\{S_k^N \geq (S(t)/2)\}} \\
&\lesssim \frac{1}{(S(t))^\alpha N^\alpha} \mathbb{E}_{x_0} \sum_{i=1}^N i^{2s} \lambda_i^2 \left|\zeta_k^{i,N}\right|^p \overset{(8.9)}{\longrightarrow} 0.
\end{aligned}$$

The above limit follows from the assumption $S_0 \geq \epsilon$ and (4.2) (which, combined, guarantee $\min\{\epsilon, 1\} \leq S(t)$). The second addend:

$$\begin{aligned}
\mathbb{E}_{x_0} J_k^N \mathbf{1}_{\{S_k^N < (S(t)/2)\}} &\leq (\mathbb{E}_{x_0} \left|J_k^N\right|^2)^{1/2} \left(\mathbb{P}\left((S_k^N - S(t)) < -\frac{S(t)}{2}\right)\right)^{1/2} \\
&\lesssim \left(\mathbb{P}\left(\left|S_k^N - S(t)\right| > \frac{S(t)}{2}\right)\right)^{1/2} \lesssim \frac{1}{S(t)} \left(\mathbb{E}_{x_0} \left|S_k^N - S(t)\right|^2\right)^{1/2}.
\end{aligned}$$

The statement now follows from Lemma 7.5, (5.2), the assumption $S_0 \geq \epsilon$ and (4.2).

- Finally, we turn to the proof of (8.11).

$$\frac{1}{1 + |R|\sqrt{N}} = \eta + \frac{1}{1 + |Z_{\ell,k}|\sqrt{N}}, \tag{8.79}$$

where

$$\begin{aligned}
|\eta| &= \left|\frac{1}{1 + |R|\sqrt{N}} - \frac{1}{1 + |Z_{\ell,k}|\sqrt{N}}\right| \\
&\leq \frac{\sqrt{N}\,||R| - |Z_{\ell,k}||}{1 + |Z_{\ell,k}|\sqrt{N}} \leq \frac{\sqrt{N}\,|R - Z_{\ell,k}|}{1 + |Z_{\ell,k}|\sqrt{N}},
\end{aligned}$$

having used $||a| - |b|| \leq |a - b|$. Consequently,

$$\mathbb{E}_k |\eta|^2 \leq N \left(\mathbb{E}_k |R - Z_{\ell,k}|^4\right)^{1/2} \left(\mathbb{E}_k \frac{1}{(1 + |Z_{\ell,k}|\sqrt{N})^4}\right)^{1/2} \tag{8.80}$$

$$\overset{(8.72)}{\lesssim} (1 + \|x_k^N\|_s^2) \left(\mathbb{E}_k \frac{1}{(1 + |Z_{\ell,k}|\sqrt{N})^4}\right)^{1/2}. \tag{8.81}$$

Also, from (8.79),

$$\mathbb{E}_k \frac{1}{(1 + |R|\sqrt{N})^2} \lesssim \mathbb{E}_k \eta^2 + \mathbb{E}_k \frac{1}{(1 + |Z_{\ell,k}|\sqrt{N})^2}. \tag{8.82}$$

Now notice that, given x_k, $Z_{\ell,k}$ is Gaussian with variance $2\ell^2 S_k^N$ and mean $-\ell^2$. Therefore, for every $p > 1$,

$$\begin{aligned}\mathbb{E}_k \frac{1}{(1+|Z_{\ell,k}|\sqrt{N})^p} &= \int_{\mathbb{R}} \frac{1}{(1+|x|\sqrt{N})^p} \frac{1}{\sqrt{4\pi\ell^2 S_k^N}} e^{-(x+\ell^2)^2/(2\ell^2 S_k^N)} dx \\ &\lesssim \int_{\mathbb{R}} \frac{1}{(1+|x|\sqrt{N})^p} \frac{1}{\sqrt{S_k^N}} dx \\ &= \frac{1}{\sqrt{N}} \int_{\mathbb{R}} \frac{1}{(1+|y|)^p} \frac{1}{\sqrt{S_k^N}} dy \lesssim \frac{1}{\sqrt{N S_k^N}}.\end{aligned}$$

The proof can now be concluded by combining (8.82), (8.81) and the above. □

Remark 8.12. Notice that the proof of (8.10) is the only place in which we actually use (3.4) instead of the slightly more general assumption $\|\mathcal{C}^{1/2}\nabla\Psi\| \le \|\nabla\Psi\|_{-s} \lesssim 1 + \|x\|_s$. This is to avoid technicalities and streamline the proof. □

Acknowledgments We are extremely grateful to an anonimous referee for his/her careful reading, for spotting mistakes in an earlier version and for comments that helped improving the paper.

References

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

[CRR05] O. F. Christensen, G. O. Roberts, and J. S. Rosenthal. Scaling limits for the transient phase of local Metropolis-Hastings algorithms. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 67(2):253–268, 2005.

[Hel82] I. S. Helland. Central limit theorems for martingales with discrete or continuous time. *Scand. J. Statist.*, 9(2):79–94, 1982.

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

[HSVW05] M. Hairer, A.M. Stuart, J. Voss, and P. Wiberg. Analysis of SPDEs Arising in path sampling. Part I: the Gaussian case. *Comm. Math. Sci.*, 3:587–603, 2005.

[JLM14] B. Jourdain, T. Lelièvre, and B. Miasojedow. Optimal scaling for the transient phase of Metropolis Hastings algorithms: The longtime behavior. *Bernoulli*, 20(4):1930–1978, 2014.

[JLM15] B. Jourdain, T. Lelièvre, and B. Miasojedow. Optimal scaling for the transient phase of the random walk Metropolis algorithm: The mean-field limit. *The Annals of Applied Probability*, 25(4):2263–2300, 2015.

[KOS16] J. Kuntz, M. Ottobre, and A. M. Stuart. Non-stationary phase of the MALA algorithm. *ArXiv preprint*, 2016.

[MPS12] J. C. Mattingly, N. S. Pillai, and A. M. Stuart. Diffusion limits of the random walk Metropolis algorithm in high dimensions. *The Annals of Applied Probability*, 22(3):881–930, 2012.

[OPPS16] M. Ottobre, N.S. Pillai, F.J. Pinski, and A.M. Stuart. A function space HMC algorithm with second order Langevin diffusion limit. *Bernoulli*, 22(1):60–106, 2016.

[PST12] N. S. Pillai, A. M. Stuart, and A. H. Thiéry. Optimal scaling and diffusion limits for the Langevin algorithm in high dimensions. *Ann. Appl. Probab.*, 22(6):2320–2356, 2012.

[PST14] N. S. Pillai, A. M. Stuart, and A. H. Thiéry. Noisy gradient flow from a random walk in Hilbert space. *Stochastic Partial Differential Equations: Analysis and Computations*, 2(2):196–232, 2014.

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

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

[Stu10] A.M. Stuart. Inverse problems: a Bayesian perspective. *Acta Numerica*, 19:451–559, 2010.

[Tie98] L. Tierney. A note on Metropolis-Hastings kernels for general state spaces. *Ann. Appl. Probab.*, 8(1):1–9, 1998.