# A function space HMC algorithm with second order Langevin diffusion limit

MICHELA OTTOBRE[1], NATESH S. PILLAI[2], FRANK J. PINSKI[3] and ANDREW M. STUART[4]

[1]*Department of Mathematics, Imperial College, 180 Queen's Gate, SW7 2AZ, London, UK.*
*E-mail: michelaottobre@gmail.com*
[2]*Department of Statistics, Harvard University, 1 Oxford Street, Cambridge, MA 02138, USA.*
*E-mail: pillai@fas.harvard.edu*
[3]*Department of Physics, University of Cincinnati, P.O. Box 210011, Cincinnati, OH 45221-0377, USA.*
*E-mail: pinskifj@ucmail.uc.edu*
[4]*Mathematics Institute, Warwick University, CV4 7AL, UK. E-mail: a.m.stuart@warwick.ac.uk*

We describe a new MCMC method optimized for the sampling of probability measures on Hilbert space which have a density with respect to a Gaussian; such measures arise in the Bayesian approach to inverse problems, and in conditioned diffusions. Our algorithm is based on two key design principles: (i) algorithms which are well defined in infinite dimensions result in methods which do not suffer from the curse of dimensionality when they are applied to approximations of the infinite dimensional target measure on $\mathbb{R}^N$; (ii) nonreversible algorithms can have better mixing properties compared to their reversible counterparts. The method we introduce is based on the hybrid Monte Carlo algorithm, tailored to incorporate these two design principles. The main result of this paper states that the new algorithm, appropriately rescaled, converges weakly to a second order Langevin diffusion on Hilbert space; as a consequence the algorithm explores the approximate target measures on $\mathbb{R}^N$ in a number of steps which is independent of $N$. We also present the underlying theory for the limiting nonreversible diffusion on Hilbert space, including characterization of the invariant measure, and we describe numerical simulations demonstrating that the proposed method has favourable mixing properties as an MCMC algorithm.

*Keywords:* diffusion limits; function space Markov chain Monte Carlo; hybrid Monte Carlo algorithm; second order Langevin diffusion

## 1. Introduction

Markov chain Monte Carlo (MCMC) algorithms for sampling from high dimensional probability distributions constitute an important part of Bayesian statistical inference. This paper is focussed on the design and analysis of such algorithms to sample a probability distribution on an infinite dimensional Hilbert space $\mathcal{H}$ defined via a density with respect to a Gaussian; such problems arise in the Bayesian approach to inverse problems (or Bayesian nonparametrics) [21] and in the theory of conditioned diffusion processes [9]. Metropolis–Hastings algorithms [10] constitute a popular class of MCMC methods for sampling an arbitrary probability measure. They proceed by constructing an irreducible, *reversible* Markov chain by first proposing a candidate move and then accepting it with a certain probability. The acceptance probability is chosen so as to preserve the detailed balance condition ensuring reversibility. In this work, we build on the generalized

Hybrid Monte Carlo (HMC) method of [11] to construct a new *nonreversible* MCMC method appropriate for sampling measures defined via density with respect to a Gaussian measure on a Hilbert space. We also demonstrate that, for a particular set of parameter values in the algorithm, there is a natural diffusion limit to the second order Langevin (SOL) equation with invariant measure given by the target. We thus name the new method the SOL-HMC algorithm. Our construction is motivated by the following two key design principles:

1. designing proposals which are well-defined on the Hilbert space results in MCMC methods which do not suffer from the curse of dimensionality when applied to sequences of approximating finite dimensional measures on $\mathbb{R}^N$;
2. nonreversible MCMC algorithms, which are hence not from the Metropolis–Hastings class, can have better sampling properties in comparison with their reversible counterparts.

The idea behind the first principle is explained in [6] which surveys a range of algorithms designed specifically to sample measures defined via a density with respect to a Gaussian; the unifying theme is that the proposal is reversible with respect to the underlying Gaussian so that the accept–reject mechanism depends only on the likelihood function and not the prior distribution. The second principle above is also well-documented: nonreversible Markov chains, often constructed by performing individual time-reversible[1] steps successively [12,13], or by building on Hamiltonian mechanics [7,11,15], may have better mixing properties.

Since the target distribution has support on an infinite dimensional space, practical implementation of MCMC involves discretizing the parameter space, resulting in a target measure on $\mathbb{R}^N$, with $N \gg 1$. It is well known that such discretization schemes can suffer from the curse of dimensionality: the efficiency of the algorithm decreases as the dimension $N$ of the discretized space grows large. One way of understanding this is through diffusion limits of the algorithm. In the context of measures defined via density with respect to Gaussian, this approach is taken in the papers [14,18] which show that the random walk Metropolis and Langevin algorithms require $\mathcal{O}(N)$ and $\mathcal{O}(N^{1/3})$ steps respectively, to sample the approximating target measure in $\mathbb{R}^N$. If, however, the algorithm is defined on Hilbert space then it is possible to explore the target in $\mathcal{O}(1)$ steps and this may also be demonstrated by means of a diffusion limit. The paper [17] uses this idea to study a Metropolis–Hastings algorithm which is defined on Hilbert space and is a small modification of the random walk Metropolis method; the diffusion limit is a first order reversible Langevin diffusion. Moreover the diffusion limits in [14,18] are derived under stationarity whereas the results in [18] hold for *any* initial condition. The above discussion has important practical consequences: as implied by the above diffusion limits, algorithms which are well defined on the function spaces show an order of magnitude improvement in the mixing time in these high dimensional sampling problems.

Here we employ similar techniques as that of [17] to study our new nonreversible MCMC method, and show that, after appropriate rescaling, it converges to a second order nonreversible Langevin diffusion. Our new algorithm is inspired by similar algorithms in finite dimensions, starting with the work of [11], who showed how the momentum updates could be correlated in the original HMC method of [8], and the more recent work [5] which made the explicit connection to second order Langevin diffusions; a helpful overview and discussion may be found in [15].

---

[1]For a definition of time-reversibility, see Section 2.3.

Diffusion limit results similar to ours are proved in [4,5] for finite dimensional problems. In those papers, an accept–reject mechanism is appended to various standard integrators for the first and second order Langevin equations, and shown not to destroy the strong pathwise convergence of the underlying methods. The reason for this is that rejections are rare when small time-steps are used. The same reasoning underlies the results we present here, although we consider an infinite dimensional setting and use only weak convergence. Another existing work underpinning that presented here is the paper [2] which generalizes the hybrid Monte Carlo method for measures defined via density with respect to a Gaussian so that it applies on Hilbert space. Indeed the algorithm we introduce in this paper includes the one from [2] as a special case and uses the split-step (non-Verlet) integrator first used there. The key idea of the splitting employed is to split according to linear and nonlinear dynamics within the numerical Hamiltonian integration step of the algorithm, rather than according to position and momentum. This allows for an algorithm which exactly preserves the underlying Gaussian reference measure, without rejections, and is key to the fact that the methods are defined on Hilbert space even in the non-Gaussian case.

We now define the class of models to which our main results are applicable. Let $\pi_0$ and $\pi$ be two measures on a Hilbert space $(\mathcal{H}, \langle \cdot, \cdot \rangle, \| \cdot \|)$ and assume that $\pi_0$ is Gaussian so that $\pi_0 = \mathrm{N}(0, C)$, with $C$ a covariance operator. The target measure $\pi$ is assumed to be absolutely continuous with respect to $\pi_0$ and given by the identity

$$\frac{\mathrm{d}\pi}{\mathrm{d}\pi_0}(x) = M_\Psi \exp\bigl(-\Psi(x)\bigr), \qquad x \in \mathcal{H} \tag{1.1}$$

for a real valued functional $\Psi$ (which denotes the negative log-likelihood in the case of Bayesian inference) and $M_\Psi$ a normalizing constant. Although the above formulation may appear quite abstract, we emphasize that this points to the wide-ranging applicability of our theory: the setting encompasses a large class of models arising in practice, including nonparametric regression using Gaussian random fields and statistical inference for diffusion processes and bridge sampling [9, 21].

In Section 2, we introduce our new algorithm. We start in a finite dimensional context and then explain parametric choices made with reference to the high or infinite dimensional setting. We demonstrate that various other algorithms defined on Hilbert space, such as the function space MALA [3] and function space HMC algorithms [2], are special cases. In Section 3, we describe the infinite dimensional setting in full and, in particular, detail the relationship between the change of measure, encapsulated in $\Psi$, and the properties of the Gaussian prior $\pi_0$. Section 4 contains the theory of the SPDE which both motivates our class of algorithms, and acts as a limiting process for a specific instance of our algorithm applied on a sequence of spaces of increasing dimension $N$. We prove existence and uniqueness of solutions to the SPDE and characterize its invariant measure. Section 5 contains statement of the key diffusion limit Theorem 5.1. Whilst the structure of the proof is outlined in some detail, various technical estimates are left for Appendices A and B. Section 6 contains some numerics illustrating the new algorithm in the context of a problem from the theory of conditioned diffusions. We make some brief concluding remarks in Section 7.

The new algorithm proposed and analyzed in this paper is of interest for two primary reasons. Firstly, it contains a number of existing function space algorithms as special cases and hence plays a useful conceptual role in unifying these methods. Secondly, numerical evidence

demonstrates that the method is comparable in efficiency to the function space HMC method introduced in [2] for a test problem arising in conditioned diffusions; until now, the function space HMC method was the clear best choice as demonstrated numerically in [2]. Furthermore, our numerical results indicate that for certain parameter choices in the SOL-HMC algorithm, and for certain target measures, we are able to improve upon the performance of the function space HMC algorithm, corroborating a similar observation made in [11] for the finite dimensional samplers that form motivation for the new family of algorithms that we propose here. From a technical point of view, the diffusion limit proved in this paper is similar to that proved for the function space MALA in [18], extending to the nonreversible case; however significant technical issues arise which are not present in the reversible case and, in particular, incorporating momentum flips into the analysis, which occur for every rejected step, requires new ideas.

## 2. The SOL-HMC algorithm

In this section, we introduce the SOL-HMC algorithm studied in this paper. We first describe the basic ideas from stochastic dynamics underlying this work, doing so in the finite dimensional setting of $\mathcal{H} = \mathbb{R}^N$, i.e., when the target measure $\pi(q)$ is a probability measure on $\mathbb{R}^N$ of the form

$$\frac{\mathrm{d}\pi}{\mathrm{d}\pi_0}(q) \propto \exp\left(-\Psi(q)\right),$$

where $\pi_0$ is a mean zero Gaussian with covariance matrix $\mathcal{C}$ and $\Psi(q)$ is a function defined on $\mathbb{R}^N$. A key idea is to work with an extended phase space in which the original variables are viewed as 'positions' and then 'momenta' are added to complement each position. We then explain the advantages of working with 'velocities' rather than 'momenta', in the large dimension limit. And then finally we introduce our proposed algorithm, which is built on the measure preserving properties of the second order Langevin equation. As already mentioned, our algorithm will build on some basic facts about Hamiltonian mechanics. For a synopsys about the Hamiltonian formalism see Appendix C.

### 2.1. Measure preserving dynamics in an extended phase space

Introduce the auxiliary variable $p$ ('momentum') and $\mathcal{M}$ a user-specified, symmetric positive definite 'mass' matrix. Let $\Pi_0'$ denote the Gaussian on $\mathbb{R}^{2N}$ defined as the independent product of Gaussians $N(0, \mathcal{C})$ and $N(0, \mathcal{M})$ on the $q$ and $p$ coordinates, respectively, and define $\Pi'$ by

$$\frac{\mathrm{d}\Pi'}{\mathrm{d}\Pi_0'}(q, p) \propto \exp\left(-\Psi(q)\right).$$

A key point to notice is that the marginal of $\Pi'(q, p)$ with respect to $q$ is the target measure $\pi(q)$. Define the Hamiltonian in $\mathsf{H}: \mathbb{R}^{2N} \to \mathbb{R}$ given by

$$\mathsf{H}(q, p) = \tfrac{1}{2}\langle p, \mathcal{M}^{-1}p\rangle + \tfrac{1}{2}\langle q, \mathcal{L}q\rangle + \Psi(q),$$

where $\mathcal{L} = \mathcal{C}^{-1}$. The corresponding canonical Hamiltonian differential equation is given by

$$\frac{dq}{dt} = \frac{\partial H}{\partial p} = \mathcal{M}^{-1}p, \qquad \frac{dp}{dt} = -\frac{\partial H}{\partial q} = -\mathcal{L}q - D\Psi(q). \qquad (2.1)$$

This equation preserves any smooth function of $H(q, p)$ and, as a consequence, the Liouville equation corresponding to (2.1) preserves the probability density of $\Pi'(q, p)$, which is proportional to $\exp(-H(q, p))$. This fact is the basis for HMC methods [8] which randomly sample momentum from the Gaussian $N(0, \mathcal{M})$ and then run the Hamiltonian flow for $T$ time units; the resulting Markov chain on $q$ is $\pi(q)$ invariant. In practice, the Hamiltonian flow must be integrated numerically, but if a suitable integrator is used (volume-preserving and time-reversible) then a simple accept–reject compensation corrects for numerical error.

Define

$$z = \begin{pmatrix} q \\ p \end{pmatrix}$$

and

$$J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}.$$

Then the Hamiltonian system can be written as

$$\frac{dz}{dt} = JDH(z), \qquad (2.2)$$

where, abusing notation, $H(z) := H(q, p)$. The equation (2.2) preserves the measure $\Pi'$.

Now define the matrix

$$\mathcal{K} = \begin{pmatrix} \mathcal{K}_1 & 0 \\ 0 & \mathcal{K}_2 \end{pmatrix},$$

where both $\mathcal{K}_1$ and $\mathcal{K}_2$ are symmetric. The following SDE also preserves the measure $\Pi'$:

$$\frac{dz}{dt} = -\mathcal{K}DH(z) + \sqrt{2\mathcal{K}}\frac{dW}{dt}.$$

Here $W = (W_1, W_2)$ denotes a standard Brownian motion on $\mathbb{R}^{2N}$. This SDE decouples into two independent equations for $q$ and $p$; the equation for $q$ is what statisticians term the Langevin equation [18], namely

$$\frac{dq}{dt} = -\mathcal{K}_1\big(\mathcal{L}q + D\Psi(q)\big) + \sqrt{2\mathcal{K}_1}\frac{dW_1}{dt},$$

whilst the equation for $p$ is simply the Ornstein–Uhlenbeck process:

$$\frac{dp}{dt} = -\mathcal{K}_2\mathcal{M}^{-1}p + \sqrt{2\mathcal{K}_2}\frac{dW_2}{dt}.$$

Discretizing the Langevin equation (resp., the random walk found by ignoring the drift) and adding an accept–reject mechanism, leads to the Metropolis–Adjusted Langevin (MALA) (resp., the Random Walk Metropolis (RWM) algorithm).

A natural idea is to try and combine benefits of the HMC algorithm, which couples the position and momentum coordinates, with the MALA and RWM methods. This thought experiment suggests considering the second order Langevin equation[2]

$$\frac{\mathrm{d}z}{\mathrm{d}t} = J D\mathsf{H}(z) - \mathcal{K} D\mathsf{H}(z) + \sqrt{2\mathcal{K}}\frac{\mathrm{d}W}{\mathrm{d}t}, \tag{2.3}$$

which also preserves $\Pi'$ as a straightforward calculation with the Fokker–Planck equation shows.

## 2.2. Velocity rather than momentum

Our paper is concerned with using the equation (2.3) to motivate proposals for MCMC. In particular, we will be interested in choices of the matrices $\mathcal{M}$, $\mathcal{K}_1$ and $\mathcal{K}_2$ which lead to well-behaved algorithms in the limit of large $N$. To this end, we write the equation (2.3) in position and momentum coordinates as

$$\frac{\mathrm{d}q}{\mathrm{d}t} = \mathcal{M}^{-1}p - \mathcal{K}_1\big(\mathcal{L}q + D\Psi(q)\big) + \sqrt{2\mathcal{K}_1}\frac{\mathrm{d}W_1}{\mathrm{d}t},$$

$$\frac{\mathrm{d}p}{\mathrm{d}t} = -\big(\mathcal{L}q + D\Psi(q)\big) - \mathcal{K}_2\mathcal{M}^{-1}p + \sqrt{2\mathcal{K}_2}\frac{\mathrm{d}W_2}{\mathrm{d}t}.$$

In our subsequent analysis, which concerns the large $N$ limit, it turns out to be useful to work with velocity rather than momentum coordinates; this is because the optimal algorithms in this limit are based on ensuring that the velocity and position coordinates all vary on the same scale. For this reason, we introduce $v = \mathcal{M}^{-1}p$ and rewrite the equations as

$$\frac{\mathrm{d}q}{\mathrm{d}t} = v - \mathcal{K}_1\big(\mathcal{L}q + D\Psi(q)\big) + \sqrt{2\mathcal{K}_1}\frac{\mathrm{d}W_1}{\mathrm{d}t},$$

$$\mathcal{M}\frac{\mathrm{d}v}{\mathrm{d}t} = -\big(\mathcal{L}q + D\Psi(q)\big) - \mathcal{K}_2 v + \sqrt{2\mathcal{K}_2}\frac{\mathrm{d}W_2}{\mathrm{d}t}.$$

In the infinite dimensional setting, i.e., when $\mathcal{H}$ is an infinite dimensional Hilbert space, this equation is still well posed (see (2.5) below and Theorem 4.1). However in this case $W_1$ and $W_2$ are cylindrical Wiener processes on $\mathcal{H}$ (see Section 3.1) and $\mathcal{L} = \mathcal{C}^{-1}$ is necessarily an unbounded operator on $\mathcal{H}$ because the covariance operator $\mathcal{C}$ is trace class on $\mathcal{H}$. The unbounded operators introduce undesirable behaviour in the large $N$ limit when we approximate them; thus we choose $\mathcal{M}$ and the $\mathcal{K}_i$ to remove the appearance of unbounded operators. To this end, we

---

[2]Physicists often refer to this as *the* Langevin equation for the choice $\mathcal{K}_1 \equiv 0$ which leads to noise only appearing in the momentum equation.

set $\mathcal{M} = \mathcal{L} = \mathcal{C}^{-1}$, $\mathcal{K}_1 = \Gamma_1 \mathcal{C}$ and $\mathcal{K}_2 = \Gamma_2 \mathcal{C}^{-1}$ and assume that $\Gamma_1$ and $\Gamma_2$ commute with $\mathcal{C}$ to obtain the equations

$$\frac{dq}{dt} = v - \Gamma_1\big(q + \mathcal{C}D\Psi(q)\big) + \sqrt{2\Gamma_1\mathcal{C}}\frac{dW_1}{dt}, \tag{2.4a}$$

$$\frac{dv}{dt} = -\big(q + \mathcal{C}D\Psi(q)\big) - \Gamma_2 v + \sqrt{2\Gamma_2\mathcal{C}}\frac{dW_2}{dt}, \tag{2.4b}$$

or simply

$$\frac{dq}{dt} = v - \Gamma_1\big(q + \mathcal{C}D\Psi(q)\big) + \sqrt{2\Gamma_1}\frac{dB_1}{dt}, \tag{2.5a}$$

$$\frac{dv}{dt} = -\big(q + \mathcal{C}D\Psi(q)\big) - \Gamma_2 v + \sqrt{2\Gamma_2}\frac{dB_2}{dt}. \tag{2.5b}$$

In the above, $B_1$ and $B_2$ are $\mathcal{H}$-valued Brownian motions with covariance operator $\mathcal{C}$. This equation is well-behaved in infinite dimensions provided that the $\Gamma_i$ are bounded operators, and under natural assumptions relating the reference measure, via its covariance $\mathcal{C}$, and the log density $\Psi$, which is a real valued functional defined on an appropriate subspace of $\mathcal{H}$. Detailed definitions and assumptions regarding (2.5) are contained in the next Section 3. Under such assumptions, the function

$$F(q) := q + \mathcal{C}D\Psi(q) \tag{2.6}$$

has desirable properties (see Lemma 3.4), making the existence theory for (2.5) straightforward. We develop such theory in Section 4 – see Theorem 4.1. Furthermore, in Theorem 4.2 we will also prove that equation (2.5) preserves the measure $\Pi(dq, dv)$ defined by

$$\frac{d\Pi}{d\Pi_0}(q, v) \propto \exp\big(-\Psi(q)\big), \tag{2.7}$$

where $\Pi_0$ is the independent product of $N(0, \mathcal{C})$ with itself. The measure $\Pi$ (resp., $\Pi_0$) is simply the measure $\Pi'$ (resp., $\Pi_0'$) in the case $\mathcal{M} = \mathcal{C}^{-1}$ and rewritten in $(q, v)$ coordinates instead of $(q, p)$. In finite dimensions, the invariance of $\Pi$ follows from the discussions concerning the invariance of $\Pi'$.

## 2.3. Function space algorithm

We note that the choice $\Gamma_1 \equiv 0$ gives the standard (physicists) Langevin equation

$$\frac{d^2q}{dt} + \Gamma_2\frac{dq}{dt} + \big(q + \mathcal{C}D\Psi(q)\big) = \sqrt{2\Gamma_2\mathcal{C}}\frac{dW_2}{dt}. \tag{2.8}$$

In this section, we describe an MCMC method designed to sample the measure $\Pi$ given by (2.7) and hence, by marginalization, the measure $\pi$ given by (1.1). The method is based on discretization of the second order Langevin equation (2.8), written as the hypo-elliptic first order

equation (2.9) below. In the finite dimensional setting, a method closely related to the one that we introduce was proposed in [11]; however we will introduce different Hamiltonian solvers which are tuned to the specific structure of our measure, in particular to the fact that it is defined via density with respect to a Gaussian. We will be particularly interested in choices of parameters in the algorithm which ensure that the output (suitability interpolated to continuous time) behaves like (2.8) whilst, as is natural for MCMC methods, exactly preserving the invariant measure. This perspective on discretization of the (physicists) Langevin equation in finite dimensions was introduced in [4,5].

In position/velocity coordinates, and using (2.6), (2.5) becomes

$$
\begin{aligned}
\frac{\mathrm{d}q}{\mathrm{d}t} &= v, \\
\frac{\mathrm{d}v}{\mathrm{d}t} &= -F(q) - \Gamma_2 v + \sqrt{2\Gamma_2 \mathcal{C}} \frac{\mathrm{d}W_2}{\mathrm{d}t}.
\end{aligned}
\tag{2.9}
$$

The algorithm we use is based on splitting (2.9) into an Ornstein–Uhlenbeck (OU) process and a Hamiltonian ODE. The OU process is

$$
\begin{aligned}
\frac{\mathrm{d}q}{\mathrm{d}t} &= 0, \\
\frac{\mathrm{d}v}{\mathrm{d}t} &= -\Gamma_2 v + \sqrt{2\Gamma_2 \mathcal{C}} \frac{\mathrm{d}W_2}{\mathrm{d}t},
\end{aligned}
\tag{2.10}
$$

and the Hamiltonian ODE is given by

$$
\begin{aligned}
\frac{\mathrm{d}q}{\mathrm{d}t} &= v, \\
\frac{\mathrm{d}v}{\mathrm{d}t} &= -F(q).
\end{aligned}
\tag{2.11}
$$

The solution of the OU process (2.10) is denoted by $(q(t), v(t)) = \Theta_0(q(0), v(0); \xi^t)$; here $\xi^t$ is a mean zero Gaussian random variable with covariance operator $\mathcal{C}(I - \exp(-2t\Gamma_2))$. Notice that the dynamics given by both (2.10) and by (2.11) preserve the target measure $\Pi$ given in (2.7). This naturally suggests constructing an algorithm based on alternating the above two dynamics. However, note that whilst (2.10) can be solved exactly, (2.11) requires a further numerical approximation. If the numerical approximation is based on a volume-preserving and time-reversible numerical integrator, then the accept–reject criterion for the resulting MCMC algorithm can be easily expressed in terms of the energy differences in H. A flow $\varphi^t$ on $\mathbb{R}^{2N}$ is said to be time-reversible if $\varphi^t(q(0), v(0)) = (q(t), v(t))$ implies $\varphi^t(q(t), -v(t)) = (q(0), -v(0))$. Defintion of time-reversible and discussion of the roles of time-reversible and volume-preserving integrators may be found in [20].

To construct volume-preserving and time-reversible integrators, the Hamiltonian integration will be performed by a further splitting of (2.11). The usual splitting for the widely used Verlet method is via the velocity and the position coordinates [11]. Motivated by our infinite dimensional setting, we replace the Verlet integration by the splitting method proposed in [2]; this

leads to an algorithm which is exact (no rejections) in the purely Gaussian case where $\Psi \equiv 0$. The splitting method proposed in [2] is via the linear and nonlinear parts of the problem, leading us to consider the two equations

$$\frac{dq}{dt} = v, \qquad \frac{dv}{dt} = -q, \tag{2.12}$$

with solution denoted as $(q(t), v(t)) = \mathsf{R}^t(q(0), v(0))$; and

$$\frac{dq}{dt} = 0, \qquad \frac{dv}{dt} = -\mathcal{C}D\Psi(q), \tag{2.13}$$

with solution denoted as $(q(t), v(t)) = \Theta_1^t(q(0), v(0))$. We note that the map

$$\chi^t = \Theta_1^{t/2} \circ \mathsf{R}^t \circ \Theta_1^{t/2}$$

is a volume-preserving and time-reversible second order accurate approximation of the Hamiltonian ODE (2.11). We introduce the notation

$$\chi_\tau^t = \left(\chi^t \circ \cdots \circ \chi^t\right), \qquad \left\lfloor \frac{\tau}{t} \right\rfloor \text{ times}$$

to denote integration, using this method, up to time $\tau$. This integrator can be made to preserve the measure $\Pi$ if appended with a suitable accept–reject rule as detailed below. On the other hand the stochastic map $\Theta_0^t$ preserves $\Pi$ since it leaves $q$ invariant and since the OU process, which is solved exactly, preserves $\Pi_0$. We now take this idea to define our MCMC method. The infinite dimensional Hilbert space $\mathcal{H}^s \times \mathcal{H}^s$ in which the chain is constructed will be properly defined in the next section. Here we focus on the algorithm, which will be explained in more details and analyzed in Section 5.

Define the operation $'$ so that $v'$ is the velocity component of $\Theta_0^\delta(q, v)$. The preceding considerations suggest that from point $(q^0, v^0) \in \mathcal{H}^s \times \mathcal{H}^s$ we make the proposal

$$\left(q_*^1, v_*^1\right) = \chi_\tau^h \circ \Theta_0^\delta\left(q^0, v^0\right)$$

and that the acceptance probability is given by

$$\alpha\left(x^0, \xi^\delta\right) := 1 \wedge \exp\left(\mathsf{H}\left(q^0, \left(v^0\right)'\right) - \mathsf{H}\left(q_*^1, v_*^1\right)\right),$$

where

$$\mathsf{H}(q, v) = \tfrac{1}{2}\langle q, \mathcal{C}^{-1}q \rangle + \tfrac{1}{2}\langle v, \mathcal{C}^{-1}v \rangle + \Psi(q), \tag{2.14}$$

$\langle \cdot, \cdot \rangle$ denoting scalar product in $\mathcal{H}$. One step of the resulting MCMC method is then defined by setting

$$\begin{aligned}
\left(q^1, v^1\right) &= \left(q_*^1, v_*^1\right) && \text{with probability } \alpha\left(x^0, \xi^\delta\right) \\
&= \left(q^0, -\left(v^0\right)'\right) && \text{otherwise.}
\end{aligned} \tag{2.15}$$

We will make further comments on this algorithm and on the expression (2.14) for the Hamiltonian in Section 5, see Remark 5.6. Here it suffices simply to note that whilst $H$ will be almost surely infinite, the energy difference is well-defined for the algorithms we employ. We stress that when the proposal is rejected the chain does not remain in $(q^0, v^0)$ but it moves to $(q^0, -(v^0)')$; i.e., the position coordinate stays the same while the velocity coordinate is first evolved according to (2.10) and then the sign is flipped. This flipping of the sign, needed to preserve reversibility, leads to some of the main technical differences with respect to [17]; see Remark 5.13. For the finite dimensional case with Verlet integration the form of the accept–reject mechanism and, in particular, the sign-reversal in the velocity, was first derived in [11] and is discussed in further detail in Section 5.3 of [15]. The algorithm (2.15) preserves $\Pi$ and we refer to it as the SOL-HMC algorithm. Recalling that $v'$ denotes the velocity component of $\Theta_0^\delta(q, v)$, we can equivalently use the notations $\alpha(x, \xi^\delta)$ and $\alpha(q, v')$, for $x = (q, v)$ (indeed, by the definition of $\Theta_0^\delta$, $v'$ depends on $\xi^\delta$). With this in mind, the pseudo-code for the SOL-HMC is as follows.

---

SOL-HMC in $\mathcal{H}^s$:

1. Pick $(q^0, v^0) \in \mathcal{H}^s \times \mathcal{H}^s$ and set $k = 0$;
2. given $(q^k, v^k)$, define $(v^k)'$ to be the $v$-component of $\Theta_0^\delta(q^k, v^k)$ and calculate the proposal

$$\left(q_*^{k+1}, v_*^{k+1}\right) = \chi_\tau^h\left(q^k, \left(v^k\right)'\right);$$

3. define the acceptance probability $\alpha(q^k, (v^k)')$;
4. set $(q^{k+1}, v^{k+1}) = (q_*^{k+1}, v_*^{k+1})$ with probability $\alpha(q^k, (v^k)')$;
   otherwise set $(q^{k+1}, v^{k+1}) = (q^k, -(v^k)')$;
5. set $k \to k + 1$ and go to (2).

---

**Theorem 2.1.** *Let Assumption* 3.1 *hold. For any* $\delta, h, \tau > 0$, *the Markov chain defined by* (2.15) *is invariant with respect to* $\Pi$ *given by* (2.7).

**Proof.** See Appendix A. □

**Remark 2.2.** We first note that if $\delta \to \infty$ then the algorithm (2.15) is that introduced in the paper [2]. From this it follows that, if $\delta = \infty$ and $\tau = h$, then the algorithm is simply the function-space Langevin introduced in [3].

Secondly, we mention that, in the numerical experiments reported later, we will choose $\Gamma_2 = I$. The solution of the OU process (2.10) for $v$ is thus given as

$$v(\delta) = \left(1 - \iota^2\right)^{1/2} v(0) + \iota w, \tag{2.16}$$

where $w \sim N(0, C)$ and $e^{-2\delta} = (1 - \iota^2)$. The numerical experiments will be described in terms of the parameter $\iota$ rather than $\delta$.

# 3. Preliminaries

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

## 3.1. Notation

Let $(\mathcal{H}, \langle \cdot, \cdot \rangle, \| \cdot \|)$ denote a separable Hilbert space of real valued functions with the canonical norm derived from the inner-product. Let $\mathcal{C}$ be a positive, trace class operator on $\mathcal{H}$ and $\{\varphi_j, \lambda_j^2\}_{j \geq 1}$ be the eigenfunctions and eigenvalues of $\mathcal{C}$, respectively, so that

$$\mathcal{C}\varphi_j = \lambda_j^2 \varphi_j \qquad \text{for } j \in \mathbb{N}.$$

We assume a normalization under which $\{\varphi_j\}_{j \geq 1}$ forms a complete orthonormal basis in $\mathcal{H}$. For every $x \in \mathcal{H}$ we have the representation $x = \sum_j x_j \varphi_j$, where $x_j = \langle x, \varphi_j \rangle$. Using this notation, we define Sobolev-like spaces $\mathcal{H}^r$, $r \in \mathbb{R}$, with the inner products and norms defined by

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

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

$$\|x\|_{\mathcal{C}}^2 = \left\| \mathcal{C}^{-1/2} x \right\|^2 = \sum_{j=1}^{\infty} \lambda_j^{-2} x_j^2.$$

For $r \in \mathbb{R}$, let $Q_r : \mathcal{H} \mapsto \mathcal{H}$ denote the operator which is diagonal in the basis $\{\varphi_j\}_{j \geq 1}$ with diagonal entries $j^{2r}$, i.e.,

$$Q_r \varphi_j = j^{2r} \varphi_j$$

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

$$\langle x, y \rangle_r = \left\langle Q_r^{1/2} x, Q_r^{1/2} y \right\rangle \quad \text{and} \quad \|x\|_r^2 = \left\| Q_r^{1/2} x \right\|^2.$$

Since $\|Q_r^{-1/2} \varphi_k\|_r = \|\varphi_k\| = 1$, we deduce that $\{Q_r^{-1/2} \varphi_k\}_{k \geq 1}$ forms an orthonormal basis for $\mathcal{H}^r$. A function $y \sim N(0, \mathcal{C})$ can be expressed as

$$y = \sum_{j=1}^{\infty} \lambda_j \rho_j \varphi_j \qquad \text{with } \rho_j \overset{\mathcal{D}}{\sim} N(0, 1) \text{ i.i.d;} \tag{3.1}$$

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

$$y = \sum_{j=1}^{\infty} (\lambda_j j^r) \rho_j (Q_r^{-1/2} \varphi_j) \qquad \text{with } \rho_j \overset{\mathcal{D}}{\sim} N(0,1) \text{ i.i.d.} \tag{3.2}$$

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

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

We stress that in the above $\{\varphi_j\}_{j \in \mathbb{N}}$ 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

$$\text{Trace}_{\mathcal{H}^r}(\tilde{D}) \overset{\text{def}}{=} \sum_{j=1}^{\infty} \langle Q_r^{-1/2} \varphi_j, \tilde{D} Q_r^{-1/2} \varphi_j \rangle_r.$$

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

$$\mathcal{C}_r = Q_r^{1/2} \mathcal{C} Q_r^{1/2} \tag{3.3}$$

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 (3.1) and (3.2)). In the same way, if $\text{Trace}_{\mathcal{H}^r}(\mathcal{C}_r) < \infty$ then

$$B_2(t) = \sum_{j=1}^{\infty} \lambda_j \beta_j(t) \varphi_j = \sum_{j=1}^{\infty} \lambda_j j^r \beta_j(t) \hat{\varphi}_j,$$

with $\{\beta_j(t)\}_{j \in \mathcal{N}}$ 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. In the next section, we will need the cylindrical Wiener process $W(t)$ which is defined via the sum

$$W(t) := \sum_{j=1}^{\infty} \beta_j(t) \varphi_j.$$

This process is $\mathcal{H}^r$-valued for any $r < -\frac{1}{2}$. Observe now that if $\{\hat{\varphi}_j\}_{j \in \mathbb{N}}$ is an orthonormal basis of $\mathcal{H}^r$ then, denoting $\mathcal{H}^r \times \mathcal{H}^r \ni \hat{\varphi}_j^1 = (\hat{\varphi}_j, 0)$ and $\mathcal{H}^r \times \mathcal{H}^r \ni \hat{\varphi}_j^2 = (0, \hat{\varphi}_j)$, $\mathfrak{F} = \{\hat{\varphi}_j^1, \hat{\varphi}_j^2\}_{j \in \mathbb{N}}$ is an orthonormal basis for $\mathcal{H}^r \times \mathcal{H}^r$. Let $\mathfrak{C}_r : \mathcal{H}^r \times \mathcal{H}^r \to \mathcal{H}^r \times \mathcal{H}^r$ be the diagonal operator such that

$$\mathfrak{C}_r \hat{\varphi}_j^1 = (0,0), \qquad \mathfrak{C}_r \hat{\varphi}_j^2 = j^{2r} \lambda_j^2 \hat{\varphi}_j^2 = (0, \mathcal{C}_r \hat{\varphi}_j) \qquad \forall j \in \mathbb{N}$$

and $\tilde{\mathfrak{C}}_r : \mathcal{H}^r \times \mathcal{H}^r \to \mathcal{H}^r \times \mathcal{H}^r$ be the diagonal operator such that

$$\tilde{\mathfrak{C}}_r \hat{\varphi}_j^1 = j^{2r} \lambda_j^2 \hat{\varphi}_j^1 = (\mathcal{C}_r \hat{\varphi}_j, 0), \qquad \tilde{\mathfrak{C}}_r \hat{\varphi}_j^2 = j^{2r} \lambda_j^2 \hat{\varphi}_j^2 = (0, \mathcal{C}_r \hat{\varphi}_j) \qquad \forall j \in \mathbb{N}. \qquad (3.4)$$

Consistently, $B(t) := (0, B_2(t))$ will denote an $\mathcal{H}^r \times \mathcal{H}^r$ valued Brownian motion with covariance operator $\mathfrak{C}_r$ and $\tilde{B}(t) := (B_1(t), B_2(t))$ will denote a $\mathcal{H}^r \times \mathcal{H}^r$ valued Brownian motion with covariance operator $\tilde{\mathfrak{C}}_r$. In other words, $B_1(t)$ and $B_2(t)$ are independent $\mathcal{H}^r$-valued $\mathcal{C}_r$-Brownian motions.

Throughout, we use the following notation.

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

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

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

For an operator $A : \mathcal{H}^r \mapsto \mathcal{H}^l$, we denote its operator norm by $\| \cdot \|_{\mathcal{L}(\mathcal{H}^r, \mathcal{H}^l)}$ defined by

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

For self-adjoint $A$ and $r = l = 0$ this is, of course, the spectral radius of $A$. Finally, in the following we will consider the product space $\mathcal{H}^r \times \mathcal{H}^r$. The norm of $w = (w_1, w_2) \in \mathcal{H}^r \times \mathcal{H}^r$ is

$$\|w\|_{r \times r}^2 := \|w_1\|_r^2 + \|w_2\|_r^2.$$

## 3.2. Assumptions

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

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

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 and second derivative globally bounded. Weaker assumptions could be dealt with by use of stopping time arguments.

**Assumption 3.1.** *The functional $\Psi$, covariance operator $\mathcal{C}$ and the operators $\Gamma_1, \Gamma_2$ satisfy the following assumptions.*

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

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

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

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

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

4. *Derivatives of $\Psi$: The derivatives of $\Psi$ satisfy*

$$\left\| D\Psi(x) \right\|_{-s} \lesssim 1 + \|x\|_s \quad \text{and} \quad \left\| \partial^2\Psi(x) \right\|_{\mathcal{L}(\mathcal{H}^s, \mathcal{H}^{-s})} \lesssim 1. \tag{3.5}$$

5. *Properties of the $\Gamma_i$: The operators $\Gamma_1, \Gamma_2$ commute with $\mathcal{C}$ and are bounded linear operators from $\mathcal{H}^s$ into itself.*

**Remark 3.2.** The condition $\kappa > \frac{1}{2}$ ensures that $\text{Trace}_{\mathcal{H}^r}(\mathcal{C}_r) < \infty$ for any $r < \kappa - \frac{1}{2}$: this implies that $\pi_0(\mathcal{H}^r) = 1$ for any $\tau > 0$ and $r < \kappa - \frac{1}{2}$.

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

The Assumption 3.1 ensure that the functional $\Psi$ behaves well in a sense made precise in the following lemma.

**Lemma 3.4.** *Let Assumption 3.1 hold.*

1. *The function $F(x)$ given by (2.6) is globally Lipschitz on $\mathcal{H}^s$:*

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

2. *The second order remainder term in the Taylor expansion of $\Psi$ satisfies*

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

**Proof.** See [14,17]. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

## 4. SPDE theory

In this section, we study the SDE (2.5) in the infinite dimensional Hilbert space setting; we work under the assumptions specified in the previous section. Recall that our goal is to sample the measure $\pi$ in (1.1), but that we have extended our state space to obtain the measure $\Pi$ given by (2.7), with $q$ marginal given by $\pi$. Here $\Pi_0$ is the independent product of $\pi_0 = N(0, \mathcal{C})$ with itself in the $q$ and $p$ coordinates. The finite dimensional arguments in Section 2 show that the equation (2.5) preserves $\Pi_0$. The aim of this section is to show that these steps all make sense in the infinite dimensional context, under the assumptions laid out in the previous section.

**Theorem 4.1.** *Let Assumption 3.1 hold. Then, for any initial condition $(q(0), v(0)) \in \mathcal{H}^s \times \mathcal{H}^s$, any $T > 0$ and almost every $\mathcal{H}^s \times \mathcal{H}^s$-valued $\tilde{\mathcal{C}}_s$-Brownian motion $\tilde{B}(t) = (B_1(t), B_2(t))$, there exists a unique solution of the SDE (2.5) in the space $C([0, T], \mathcal{H}^s \times \mathcal{H}^s)$. Furthermore, the Itô map $(B_1, B_2) \in C([0, T]; \mathcal{H}^s \times \mathcal{H}^s) \mapsto (q, v) \in C([0, T]; \mathcal{H}^s \times \mathcal{H}^s)$ is Lipschitz.*

**Proof.** If we define

$$x = \begin{pmatrix} q \\ v \end{pmatrix},$$

together with the operator

$$\Gamma = \begin{pmatrix} \Gamma_1 & 0 \\ 0 & \Gamma_2 \end{pmatrix},$$

then equation (2.5) takes the form

$$\frac{dx}{dt} = G(x) + \sqrt{2\Gamma} \frac{d\tilde{B}}{dt}, \qquad (4.1)$$

where

$$G(x) = \begin{pmatrix} v - \Gamma_1 F(q) \\ -F(q) - \Gamma_2 v \end{pmatrix}. \qquad (4.2)$$

A solution of (2.5) satisfies the integral equation

$$x(t) = x_0 + \int_0^t G\big(x(s)\big)\, ds + \sqrt{2\Gamma}\, \tilde{B}(t),$$

where $x(0) = x_0$. By virtue of Lemma 3.4, we see that $G : \mathcal{H}^s \times \mathcal{H}^s \to \mathcal{H}^s \times \mathcal{H}^s$ is globally Lipschitz. Furthermore, Remark 3.2 shows that $\tilde{B} \in C([0, T]; \mathcal{H}^s \times \mathcal{H}^s)$ almost surely. To prove existence and uniqueness of a solution, we consider the map $\Xi : C([0, T]; \mathcal{H}^s \times \mathcal{H}^s) \mapsto C([0, T]; \mathcal{H}^s \times \mathcal{H}^s)$ defined by

$$\Xi(x)(t) := x_0 + \int_0^t G\big(x(s)\big)\, ds + \sqrt{2\Gamma}\, \tilde{B}(t).$$

Since $F$ is globally Lipschitz from $\mathcal{H}^s$ into itself, it follows that $G$ is globally Lipschitz from $\mathcal{H}^s \times \mathcal{H}^s$ into itself. This in turn implies that $\Xi$ is Lipschitz and that, furthermore, the Lipschitz constant may be made less than one, by choosing $t$ sufficiently small. From this existence and uniqueness of a solution follows by the contraction mapping principle, on time-intervals sufficiently small. The argument may then be repeated on successive time-intervals to prove the result on any time-interval $[0, T]$.

Now let

$$\Upsilon : (x_0, \tilde{B}) \in \mathcal{H}^s \times \mathcal{H}^s \times C([0, T]; \mathcal{H}^s \times \mathcal{H}^s) \mapsto x \in C([0, T]; \mathcal{H}^s \times \mathcal{H}^s). \qquad (4.3)$$

The arguments used in Lemma 3.7 of [14] show that $\Upsilon$ is Lipschitz continuous and hence the desired properties of the Itô map follow. $\qquad \square$

For $N \in \mathbb{N}$, let $\mathcal{H}^N$ denote the linear span of the first $N$ eigenfunctions of $\mathcal{C}$, $P^N : \mathcal{H} \mapsto \mathcal{H}^N$ denote the projection map and $\Psi^N = \Psi \circ P^N$. Define $Q^N = I - P^N$. Recall equations (2.4). Let $\Upsilon^N$ denote the Ito map obtained by replacing $D\Psi$ by $P^N D\Psi^N$ in (2.4).

The following is the key result of this section. Our choices of measure (2.7) and dynamics (2.4) have been coordinated to ensure that the resulting stochastic dynamics preserves $\Pi$:

**Theorem 4.2.** *For any initial condition* $(q(0), v(0)) \sim \Pi$ *and any* $T > 0$, *the equation* (2.4) *preserves* $\Pi$: $(q(T), v(T)) \sim \Pi$.

**Proof.** The proof follows along the lines of Theorem 3.1 of [2]. The key idea is to exploit the fact that for finite dimensional $\mathcal{H}$, the invariance of $\Pi$ under the dynamics $(q(T), v(T))$ follows easily. From this, the invariance for an infinite dimensional $\mathcal{H}$ follows from an application of the dominated convergence theorem which we outline below.

We let $\mathbb{W}$ denote the Weiner measure on $X = C([0, T]; \mathcal{H}^s \times \mathcal{H}^s)$ induced by Brownian motions with covariance the same as that of $\Pi_0$. For any continuous, bounded function $g : \mathcal{H}^s \times \mathcal{H}^s \mapsto \mathbb{R}$ and $T > 0$, we need to show that

$$\int_{\mathcal{H} \times X} g\big(\Upsilon(q, v, W)\big) \exp\big(-\Psi(q)\big) \, d\Pi_0(q, v) \, d\mathbb{W}(W)$$
$$= \int_{\mathcal{H}} g(q, v) \exp\big(-\Psi(q)\big) \, d\Pi_0(q, v). \qquad (4.4)$$

First, we claim that for any $N \in \mathbb{N}$,

$$\int_{\mathcal{H} \times X} g\big(\Upsilon^N(q, v, W)\big) \exp\big(-\Psi^N(q)\big) \, d\Pi_0(q, v) \, d\mathbb{W}(W)$$
$$= \int_{\mathcal{H}} g(q, v) \exp\big(-\Psi^N(q)\big) \, d\Pi_0(q, v). \qquad (4.5)$$

This follows from the fact that the flow $\Upsilon^N$ preserves the invariant measure proportional to $\exp(-\Psi^N)\Pi_0$ as obtained below in Lemma 4.3.

In Lemma 4.4 below, we will show that $\Upsilon^N$ converges pointwise to $\Upsilon$. Thus by the continuity of $g$, $g(\Upsilon^N(q, v, W))$ converges pointwise to $g(\Upsilon(q, v, W))$. Clearly, $\exp(-\Psi^N(q))$ converges to $\exp(-\Psi(q))$ pointwise. Since $g$ is bounded and $\Psi$, $\Psi^N$ are positive, by the dominated convergence theorem the right, respectively, left-hand side of (4.5) converges to the right, respectively, left-hand side of (4.4) and the claim follows.                                              $\square$

**Lemma 4.3.** *Let Assumption* 3.1 *hold. The measure* $\Pi^N \propto \exp(-\Psi^N)\Pi_0$ *factors as the product of two measures on* $P^N\mathcal{H}^s$ *and* $Q^N\mathcal{H}^s$. *The measure* $\Pi^N \propto \exp(-\Psi^N)\Pi_0$ *is preserved by* $\Upsilon^N$.

**Proof.** By construction, the measure $\Pi_0$ factors as the product of two measures $\mu_0 = N(0, P^N \mathcal{C} P^N)$ and $\mu_0^\perp = N(0, Q^N \mathcal{C} Q^N)$. Since $\Psi^N$ is 0 on $Q^N$, it follows that $\Pi^N$ factors into $\mu_1 \propto \exp(-\Psi^N)\mu_0$ on $P^N\mathcal{H}^s$ and $\mu_1^\perp = \mu_0^\perp$ on $Q^N\mathcal{H}^s$.

Now, as explained in Section 2 for any $N$, $\mu_1$ is invariant for $P^N\Upsilon^N$. Also setting $\Psi = 0$ in (2.4) results in an OU flow on $\mathcal{H}^s$ for which $\Pi_0$ is invariant. Thus if $D\Psi$ is replaced by $P^N D\Psi^N$ in (2.4), the resulting flow on $Q^N$ is an Orstein–Uhlenbeck process with invariant measure $\mu_1^\perp$. Since $\Pi^N$ is a product of $\mu_1$ and $\mu_1^\perp$, the result follows.                                          $\square$

The following result shows the pointwise convergence of $\Upsilon^N$ to $\Upsilon$.

**Lemma 4.4.** *Let Assumption* 3.1 *hold. As* $N \to \infty$, $\Upsilon^N(x_0, \tilde{B})$ *converges to* $\Upsilon(x_0, \tilde{B})$ *for every* $(x_0, \tilde{B}) \in \mathcal{H}^s \times \mathcal{H}^s \times C([0, T]; \mathcal{H}^s \times \mathcal{H}^s)$.

**Proof.** Proceeding similarly as in Theorem 4.1, set

$$\frac{dx^N}{dt} = G^N(x) + \sqrt{2\Gamma}\frac{d\tilde{B}}{dt},$$

where

$$G^N(x) = \begin{pmatrix} v - \Gamma_1 F^N(q) \\ -F^N(q) - \Gamma_2 v \end{pmatrix} \tag{4.6}$$

with $F^N(q) = q + \mathcal{C}P^N D\Psi^N(q)$. Let $x(t)$ denote the solution of (2.4) and $x^N$ above satisfies

$$x^N(t) = x_0 + \int_0^t G^N(x^N(s))\, ds + \sqrt{2\Gamma}\tilde{B}(t), \tag{4.7}$$

where $x(0) = x_0$. Set $e = x - x^N$. The pointwise convergence of $\Upsilon^N$ to $\Upsilon$ is established by showing that $e \to 0$ in the path space $C([0, T]; \mathcal{H}^s \times \mathcal{H}^s)$. We first decompose:

$$G(x) - G^N(x^N) = (G(x) - G^N(x)) + (G^N(x) - G^N(x^N)). \tag{4.8}$$

Next, it can be shown that $G^N$ is globally Lipschitz with the Lipschitz constant $L$ independent of $N$ (see [18], Lemma 4.1). Thus we have $\|G^N(x(t)) - G^N(x^N(t))\|_s \le L\|e(t)\|_s$. Combining

this bound with (4.7) and (4.8),

$$\left\| e(t) \right\|_s \leq \int_0^t L \left\| e(u) \right\|_s du + \int_0^t \left\| G\big(x(u)\big) - G^N\big(x(u)\big) \right\|_s du.$$

Thus by Gronwall's inequality, it suffices to show that

$$\sup_{0 \leq t \leq T} \left\| G\big(x(t)\big) - G^N\big(x(t)\big) \right\|_s \to 0$$

as $N \to \infty$. To this end, write

$$F(x) - F^N(x) = \big(\mathcal{C} D\Psi(x) - \mathcal{C} P^N D\Psi(x)\big) + \big(\mathcal{C} P^N D\Psi(x) - \mathcal{C} P^N D\Psi^N(x)\big).$$

Since $\mathcal{C} D\Psi$ is globally Lipschitz,

$$\left\| G\big(x(t)\big) - G^N\big(x(t)\big) \right\|_s \lesssim \left\| (I - P^N)\mathcal{C} D\Psi\big(x(t)\big) \right\|_s + \left\| (I - P^N)x(t) \right\|_s. \qquad (4.9)$$

From the existence of a global solution for (2.4) as shown in Theorem 4.1, it follows that $\sup_{0 \leq t \leq T} \|x(t)\|_s < \infty$. Thus from (4.9) we infer that $\sup_{0 \leq t \leq T} \|G(x(t)) - G^N(x(t))\|_s \to 0$, and the claim follows. $\qquad \square$

# 5. Diffusion limit of algorithms

The main result of this section is the diffusion limit Theorem 5.1: using the prescription (2.15) and setting $\delta = h = \tau$, we construct a sequence of Markov chains $x^{k,\delta}$ (i.e., for every fixed delta, $\{x^{k,\delta}\}_k$ is a Markov chain) and consider the process $z^\delta(t)$ which is the continuous time interpolant of the chain $x^{k,\delta}$. Then $z^\delta(t)$ converges to the solution of the SDE (5.9), which is a specific instance of (4.1), when $\Gamma_1 = 0$. By Theorem 4.2, the flow (5.9) preserves the measure $\Pi$ defined in (2.7).

More precisely, for $q, v \in \mathcal{H}^s$, let $x \in \mathcal{H}^s \times \mathcal{H}^s$ denote the pair $x = (q, v)$; we recall that the norm of $x$ is then

$$\|x\|_{s \times s}^2 := \|q\|_s^2 + \|v\|_s^2.$$

With the algorithm described in Section 2.3, taking $\delta = h = \tau$ we construct the Markov chain $x^{k+1,\delta} := (q^{k+1,\delta}, v^{k+1,\delta})$ as follows

$$
\begin{aligned}
\big(q^{k+1,\delta}, v^{k+1,\delta}\big) &= \big(q_*^{k+1,\delta}, v_*^{k+1,\delta}\big) \qquad \text{with probability } \alpha^{k,\delta} \\
&= \big(q^{k,\delta}, -\big(v^{k,\delta}\big)'\big) \qquad \text{otherwise,}
\end{aligned}
\qquad (5.1)
$$

where

$$\alpha^{k,\delta} = \alpha\big(x^{k,\delta}, \xi^\delta\big) := 1 \wedge \exp\big(\mathsf{H}\big(q^{k,\delta}, \big(v^{k,\delta}\big)'\big) - \mathsf{H}\big(q_*^{k+1,\delta}, v_*^{k+1,\delta}\big)\big).$$

We specify that in the above

$$(q_*^{k+1,\delta}, v_*^{k+1,\delta}) = \chi^\delta \circ \Theta_0^\delta(q^{k,\delta}, v^{k,\delta}) \quad \text{and}$$

$$((q^{k,\delta})', -(v^{k,\delta})') = (q^{k,\delta}, -\mathcal{P}_v(\Theta_0^\delta(q^{k,\delta}, v^{k,\delta}))),$$

where for $x \in \mathcal{H}^s \times \mathcal{H}^s$, we denote by $\mathcal{P}_q(x)$ and $\mathcal{P}_v(x)$ the projection of $x$ on the $q$ and $v$ component, respectively. Notice that introducing $\gamma^{k,\delta} \sim \text{Bernoulli}(\alpha^{k,\delta})$, the algorithm (5.1) can be also written as

$$(q^{k+1,\delta}, v^{k+1,\delta}) = \gamma^{k,\delta}(q_*^{k+1,\delta}, v_*^{k+1,\delta}) + (q^{k,\delta}, -(v^{k,\delta})').$$

Following [17], we consider the piecewise linear and the piecewise constant interpolant of the chain $x^{k,\delta}$, $z^\delta(t)$ and $\bar{z}^\delta(t)$, respectively:

$$z^\delta(t) := \frac{1}{\delta}(t - t_k)x^{k+1,\delta} + \frac{1}{\delta}(t_{k+1} - t)x^{k,\delta}, \qquad t_k \le t < t_{k+1}, t_k = k\delta, \tag{5.2}$$

$$\bar{z}^\delta(t) := x^{k,\delta}, \qquad t_k \le t < t_{k+1}, t_k = k\delta. \tag{5.3}$$

Decompose the chain $x^{k,\delta}$ into its drift and martingale part:

$$x^{k+1,\delta} = x^{k,\delta} + \delta G^\delta(x^{k,\delta}) + \sqrt{2\delta\mathcal{S}}M^{k,\delta},$$

where

$$\mathcal{S} = \begin{bmatrix} \text{Id} & 0 \\ 0 & \Gamma_2 \end{bmatrix},$$

$$G^\delta(x) := \frac{1}{\delta}\mathbb{E}_x[x^{k+1,\delta} - x^{k,\delta}|x^{k,\delta} = x], \tag{5.4}$$

$$M^{k,\delta} := \frac{\mathcal{S}^{-1/2}}{\sqrt{2\delta}}(x^{k+1,\delta} - x^{k,\delta} - \delta G^\delta(x^{k,\delta})), \tag{5.5}$$

$$M^\delta(x) := \mathbb{E}[M^{k,\delta}|x^{k,\delta} = x]. \tag{5.6}$$

Notice that with this definition, if $\mathcal{F}^{k,\delta}$ is the filtration generated by $\{x^{j,\delta}, \gamma^{j,\delta}, \xi^\delta, j = 0, \ldots, k\}$, we have $\mathbb{E}[M^{k,\delta}|\mathcal{F}^{k,\delta}] = 0$. Also, let us introduce the rescaled noise process

$$\tilde{B}^\delta(t) := \sqrt{2\mathcal{S}\delta}\sum_{j=0}^{k-1}M^{j,\delta} + \sqrt{\frac{2\mathcal{S}}{\delta}}(t - t_k)M^{k,\delta}, \qquad t_k \le t < t_{k+1}. \tag{5.7}$$

A simple calculation, which we present in Appendix A, shows that

$$z^\delta(t) = \Upsilon(x_0, \hat{B}^\delta), \tag{5.8}$$

where $\Upsilon$ is the map defined in (4.3) and $\hat{B}^\delta$ is the rescaled noise process $\tilde{B}^\delta$ plus a term which we will show to be small:

$$\hat{B}^\delta(t) := \tilde{B}^\delta(t) + \int_0^t \left[G^\delta\big(\bar{z}^\delta(u)\big) - G\big(z^\delta(u)\big)\right] du;$$

we stress that in the above and throughout this section the map $G(x)$ is as in (4.2) with $\Gamma_1 = 0$.

Let $B_2(t)$ be an $\mathcal{H}^s$-valued $\mathcal{C}_s$-Brownian motion (we recall that the covariance operator $\mathcal{C}_s$ has been defined in (3.3)) and $\mathcal{H}^s \times \mathcal{H}^s \ni B(t) = (0, B_2(t))$. Recall the SPDE (2.5) written in the form (4.1). The main result of this section is the following diffusion limit of the Markov chain (5.1) to (4.1).

**Theorem 5.1 (Diffusion limit).** *Let Assumption* 3.1 *hold and let* $(\mathcal{H}^s, \langle \cdot, \cdot \rangle_s)$ *be a separable Hilbert space,* $x^{k,\delta}$ *be the Markov chain* (5.1) *starting at* $x^{0,\delta} = x_0 \in \mathcal{H}^s \times \mathcal{H}^s$ *and let* $z^\delta(t)$ *be the process defined by* (5.2). *If Assumption* 3.1 *holds, then* $z^\delta(t)$ *converges weakly in* $C([0, T]; \mathcal{H}^s \times \mathcal{H}^s)$ *to the solution* $z(t) \in \mathcal{H}^s \times \mathcal{H}^s$ *of the stochastic differential equation*

$$dz(t) = G(z)\,dt + \sqrt{2\Gamma}\,dB(t),$$
$$z(0) = x_0.$$

$$(5.9)$$

The diffusion limit can be proven as a consequence of [17], Lemma 3.5. Proposition 5.4 below is a slightly more general version of [17], Lemma 3.5.

**Proof of Theorem 5.1.** Theorem 5.1 follows as a consequence of Proposition 5.4 and Lemma 5.5 below. $\qquad\square$

Consider the following conditions:

**Condition 5.2.** *The Markov chain* $x^{k,\delta} \in \mathcal{H}^s \times \mathcal{H}^s$ *defined in* (5.1) *satisfies*

- Convergence of the approximate drift. *There exist a globally Lipshitz function* $G : \mathcal{H}^s \times \mathcal{H}^s \to \mathcal{H}^s \times \mathcal{H}^s$, *a real number* $a > 0$ *and an integer* $p \geq 1$ *such that*

$$\left\|G^\delta(x) - G(x)\right\|_{s \times s} \lesssim \delta^a\big(1 + \|x\|_{s \times s}^p\big).$$

$$(5.10)$$

- Size of the increments. *There exist a real number* $r > 0$ *and an integer* $n \geq 1$ *such that*

$$\mathbb{E}\big[\big\|x^{k+1,\delta} - x^{k,\delta}\big\|_{s \times s}\big|x^{k,\delta} = x\big] \lesssim \delta^r\big(1 + \|x\|_{s \times s}^n\big).$$

$$(5.11)$$

- A priori bound. *There exists a real number* $\varepsilon$ *such that* $1 - \varepsilon + (a \wedge r) > 0$ *(with* $a$ *and* $r$ *as in* (5.10) *and* (5.11), *resp.) and the following bound holds:*

$$\sup_{\delta \in (0, 1/2)} \left\{ \delta^\varepsilon \mathbb{E}\left[\sum_{k\delta \leq T} \|x^{k,\delta}\|_{s \times s}^{p \vee n}\right]\right\} < \infty.$$

$$(5.12)$$

- Invariance principle. *As $\delta$ tends to zero the sequence of processes $\tilde{B}^{\delta}$ defined in* (5.7) *converges weakly in $C([0, T]; \mathcal{H}^s \times \mathcal{H}^s)$ to the Brownian motion $\mathcal{H}^s \times \mathcal{H}^s \ni B = (0, B_2)$ where $B_2$ is a $\mathcal{H}^s$-valued, $\mathcal{C}_s$-Brownian motion.*

***Remark 5.3.*** Notice that if (5.10) holds for some $a > 0$ and $p \geq 1$, then

$$\left\| \mathbb{E}\left[ x^{k+1,\delta} - x^{k,\delta} | x^{k,\delta} \right] \right\|_{s \times s} \lesssim \delta \left( 1 + \|x\|_{s \times s}^p \right) \tag{5.13}$$

and

$$\left\| G^{\delta}(x) \right\|_{s \times s} \lesssim 1 + \|x\|_{s \times s}^p. \tag{5.14}$$

Indeed

$$\left\| \mathbb{E}\left[ x^{k+1,\delta} - x^{k,\delta} | x^{k,\delta} = x \right] \right\|_{s \times s}$$
$$= \delta \left\| G^{\delta}(x) \right\|_{s \times s} \leq \delta \left\| G^{\delta}(x) - G(x) \right\|_{s \times s} + \delta \left\| G(x) \right\|_{s \times s} \lesssim \delta \left( 1 + \|x\|_{s \times s}^p \right),$$

having used the Lipshitzianity of the map $G(x)$. Analogously one can obtain (5.14) as well.

***Proposition 5.4.*** *Let Assumption* 3.1 *hold and let $(\mathcal{H}^s, \langle \cdot, \cdot \rangle_s)$ be a separable Hilbert space and $x^{k,\delta}$ a sequence of $\mathcal{H}^s \times \mathcal{H}^s$ valued Markov chains with $x^{0,\delta} = x_0$. Suppose the drift martingale decomposition* (5.4)–(5.5) *of $x^{k,\delta}$ satisfies Condition* 5.2. *Then the sequence of interpolants $z^{\delta}(t)$ defined in* (5.2) *converges weakly in $C([0, T]; \mathcal{H}^s \times \mathcal{H}^s)$ to the solution $z(t) \in \mathcal{H}^s \times \mathcal{H}^s$ of the stochastic differential equation* (5.9).

**Proof.** Thanks to the Lipshitzianity of the map $\Upsilon$ in (5.8) (see Theorem 4.1), the proof is analogous to the proof of [17], Lemma 3.5. We sketch it in Appendix A. $\qquad\square$

**Lemma 5.5.** *Let Assumption* 3.1 *hold and let $x^{k,\delta}$ be the Markov chain* (5.1) *starting at $x^{0,\delta} = x_0 \in \mathcal{H}^s \times \mathcal{H}^s$. Under Assumption* 3.1 *the drift martingale decomposition of $x^{k,\delta}$,* (5.4)–(5.5), *satisfies Condition* 5.2.

The remainder of this section is devoted to proving Lemma 5.5, which is needed to prove Theorem 5.1. First, in Section 5.1 we list and explain several preliminary technical lemmata, which will be proved in Appendix B. The main one is Lemma 5.7, where we study the acceptance probability. Then, in Section 5.2 and Section 5.3, we prove Lemma 5.5; in order to prove such a lemma we need to show that if Assumption 3.1 holds, the four conditions listed in Condition 5.2 are satisfied by the chain $x^{k,\delta}$. To this end, Lemma 5.10 proves that (5.10) holds with $a = 1$ and $p = 6$; Lemma 5.11 shows that (5.11) is satisfied with $r = 1/2$ and $n = 6$; the a priori bound (5.12) is proved to hold for $\varepsilon = 1$ and for any power of $\|x^{k,\delta}\|_{s \times s}$ in Lemma 5.12; finally, Lemma 5.18 is the invariance principle.

**Proof of Lemma 5.5.** Lemma 5.5 follows as a consequence of Lemma 5.10, Lemma 5.11, Lemma 5.12 and Lemma 5.18. $\qquad\square$

## 5.1. Preliminary estimates

We first analyse the acceptance probability. Given the current state of the chain $x^{k,\delta} = x = (q, v)$, the acceptance probability of the proposal $(q_*, v_*)$ is

$$\alpha^\delta := \alpha^{0,\delta}(x, \xi^\delta) = 1 \wedge \exp(H(q, v') - H(q_*, v_*)) = 1 \wedge \exp(\Delta H(q, v')). \qquad (5.15)$$

Similarly, we denote

$$\gamma^\delta := \gamma^{0,\delta} \sim \text{Bernoulli}(\alpha^\delta).$$

For an infinite dimensional Hilbert space setting, the matter of the well-posedness of the expression for the acceptance probability is not obvious; we comment on this below.

**Remark 5.6.** Before proceeding to the analysis, let us make a few observations about the expression (5.15) for the acceptance probability.

- As we have already mentioned, the flip of the sign of the velocity in case of rejection of the proposal move guarantees time-reversibility. As a consequence, the proposal moves are symmetric and the acceptance probability can be defined only in terms of the energy difference.
- We are slightly abusing notation in going from the original $H(q, p)$ to $H(q, v)$. However notice that $H(q, v)$ is preserved by the flow (2.11).
- The relevant energy difference here is $H(q, v') - H(q_*, v_*)$ (rather than $H(q, v) - H(q_*, v_*)$); indeed the first step in the definition of the proposal $(q_*, v_*)$, namely the OU process $\Theta_0^\delta(q, v)$, is based on an exact integration and preserves the desired invariant measure. Therefore, the accept–reject mechanism (which is here only to preserve the overall reversibility of the chain by accounting for the numerical error made by the integrator $\chi_\tau^h$) doesn't need to include also the energy difference $H(q, v) - H(q, v')$.
- The Hamiltonian $H(q, v)$, defined in (2.14), is almost surely infinite in an infinite dimensional context; this can be seen by just applying a zero–one law to the series representation of the scalar product $\langle q, \mathcal{C}^{-1}q \rangle$. However, in order for the acceptance probability to be well defined, all we need is for the difference $H(q, v') - H(q_*, v_*)$ to be almost surely finite, i.e., for $\Delta H(q, v')$ to be a bounded operator. This is here the case thanks to the choice of the Verlet algorithm. Indeed from [2], page 2212, we know that

$$\Delta H(q, v') = \Psi(q) - \Psi(q_*) - \frac{\delta}{2}(\langle D\Psi(q), v' \rangle + \langle D\Psi(q_*), v_* \rangle)$$

$$+ \frac{\delta^2}{8}(\|\mathcal{C}^{1/2}D\Psi(q_*)\|^2 - \|\mathcal{C}^{1/2}D\Psi(q)\|^2).$$

More details on this fact can be found in [2], pages 2210, 2212, 2227.

**Lemma 5.7.** *Let Assumption* 3.1 *hold. Then, for any* $p \geq 1$,

$$\mathbb{E}_x |1 - \alpha^\delta|^p \lesssim \delta^{2p}(1 + \|q\|_s^{4p} + \|v\|_s^{4p}). \qquad (5.16)$$

**Proof.** The proof of Lemma 5.7 can be found in Appendix B.                                          □

The above (5.16) quantifies the intuition that the acceptance rate is very high, i.e., the proposal is rejected very rarely. Therefore, the analysis of Section 5.2 and Section 5.3 is done by bearing in mind that "everything goes as if $\alpha^\delta$ were equal to one". We now state a few technical results, gathered in Lemma 5.8 and Lemma 5.9, that will be frequently used in the following.

**Lemma 5.8.** *Let Assumption* 3.1 *hold. Then, for any* $q, \tilde{q}, v, \tilde{v} \in \mathcal{H}^s$,

$$\left\| \mathcal{C}D\Psi(q) - \mathcal{C}D\Psi(\tilde{q}) \right\|_s \lesssim \|q - \tilde{q}\|_s \quad and \quad \left\| \mathcal{C}D\Psi(q) \right\|_s \lesssim \left( 1 + \|q\|_s \right); \quad (5.17)$$

$$\left| \langle D\Psi(q), v \rangle \right| \lesssim \left( 1 + \|q\|_s \right) \|v\|_s;$$

$$\left| \langle D\Psi(q), v \rangle - \langle D\Psi(\tilde{q}), \tilde{v} \rangle \right| \lesssim \|v\|_s \|q - \tilde{q}\|_s + \left( 1 + \|\tilde{q}\|_s \right) \|v - \tilde{v}\|_s; \quad (5.18)$$

$$\left\| \mathcal{C}^{1/2}D\Psi(q) \right\| \lesssim 1 + \|q\|_s;$$

$$\left\| \mathcal{C}^{1/2}D\Psi(q) - \mathcal{C}^{1/2}D\Psi(\tilde{q}) \right\| \lesssim \|q - \tilde{q}\|_s. \quad (5.19)$$

**Proof.** See [2], Lemma 4.1.                                                                      □

Recall that $B_2(t)$ is an $\mathcal{H}^s$-valued $\mathcal{C}_s$-Brownian motion and that $\xi^\delta$ is the noise component of the OU process $\Theta_0^\delta$, i.e.,

$$v' = e^{-\delta\Gamma_2}v + \int_0^\delta e^{-(\delta-u)\Gamma_2}\sqrt{2\Gamma_2}\, dB_2(u) =: e^{-\delta\Gamma_2}v + \xi^\delta. \quad (5.20)$$

By integrating $\chi^\delta$ and $\Theta_0^\delta$, the proposal move at step $k$, $x_*^{k+1,\delta} = (q_*^{k+1,\delta}, v_*^{k+1,\delta})$, is given by

$$q_*^{k+1,\delta} = \cos\delta q^{k,\delta} + \sin\delta \left( v^{k,\delta} \right)' - \frac{\delta}{2}\sin\delta \mathcal{C}D\Psi\left( q^{k,\delta} \right), \quad (5.21)$$

$$v_*^{k+1,\delta} = -\sin\delta q^{k,\delta} + \cos\delta \left( v^{k,\delta} \right)' - \frac{\delta}{2}\cos\delta \mathcal{C}D\Psi\left( q^{k,\delta} \right) - \frac{\delta}{2}\mathcal{C}D\Psi\left( q_*^{k+1,\delta} \right). \quad (5.22)$$

If $\gamma^k := \gamma^{k,\delta} \sim \text{Bernoulli}(\alpha^{k,\delta})$, then the $(k+1)$th step of the Markov chain is

$$\begin{aligned} q^{k+1,\delta} &= \gamma^k q_*^{k+1,\delta} + \left( 1 - \gamma^k \right) q^{k,\delta}, \\ v^{k+1,\delta} &= \gamma^k v_*^{k+1,\delta} - \left( 1 - \gamma^k \right) \left( v^{k,\delta} \right)'. \end{aligned} \quad (5.23)$$

**Lemma 5.9.** *Let Assumption* 3.1 *hold. Then, for any* $p \geq 1$, *we have*

$$\mathbb{E}\left\| \xi^\delta \right\|_s^p \lesssim \delta^{p/2}; \quad (5.24)$$

$$\mathbb{E}\left\| \left( v^{k,\delta} \right)' | x^{k,\delta} = x \right\|_s^p \lesssim 1 + \|v\|_s^p; \quad (5.25)$$

$$\mathbb{E}\left[ \left\| q_*^{k+1,\delta} - q^{k,\delta} \right\|_s^p | x^{k,\delta} = x \right] \lesssim \delta^p \left( 1 + \|q\|_s^p + \|v\|_s^p \right). \quad (5.26)$$

**Proof.** See Appendix B. □

## 5.2. Analysis of the drift

Let $G(x)$ be the map in (4.2) with $\Gamma_1 = 0$, i.e.,

$$G(x) = G(q, v) = \begin{bmatrix} v \\ -q - \mathcal{C}D\Psi(q) - \Gamma_2 v \end{bmatrix},$$

and $G_i(x)$ and $G_i^\delta$, $i = 1, 2$, be the $i$th component of $G$ and $G^\delta$, respectively.

**Lemma 5.10.** *Let Assumption 3.1 hold. Then, for any $x = (q, v) \in \mathcal{H}^s \times \mathcal{H}^s$,*

$$\begin{aligned} \left\| G_1^\delta(x) - G_1(x) \right\|_s &\lesssim \delta \left( 1 + \|q\|_s^6 + \|v\|_s^6 \right), \\ \left\| G_2^\delta(x) - G_2(x) \right\|_s &\lesssim \delta \left( 1 + \|q\|_s^6 + \|v\|_s^6 \right). \end{aligned} \tag{5.27}$$

**Proof.** By (5.23),

$$\begin{aligned} q^{k+1,\delta} - q^{k,\delta} &= \gamma^k \left( q_*^{k+1,\delta} - q^{k,\delta} \right), \\ v^{k+1,\delta} - v^{k,\delta} &= \gamma^k v_*^{k+1,\delta} + \left( \gamma^k - 1 \right) \left( v^{k,\delta} \right)' - v^{k,\delta}. \end{aligned} \tag{5.28}$$

So if we define

$$\begin{aligned} A_1 &:= \frac{1}{\delta} \left\| \mathbb{E}_x \left[ \gamma^\delta (\cos \delta - 1) q \right] \right\|_s, \\ A_2 &:= \left\| \mathbb{E}_x \left( \gamma^\delta \frac{\sin \delta}{\delta} e^{-\delta \Gamma_2} v, \right) - v \right\|_s, \\ A_3 &:= \left\| \mathbb{E}_x \left[ \gamma^\delta \frac{\sin \delta}{\delta} \xi^\delta - \gamma^\delta \frac{\sin \delta}{2} \mathcal{C}D\Psi(q) \right] \right\|_s \end{aligned}$$

and

$$\begin{aligned} E_1 &:= \left\| q - \mathbb{E}_x \left( \gamma^\delta \frac{\sin \delta}{\delta} q \right) \right\|_s, \\ E_2 &:= \left\| \mathcal{C}D\Psi(q) + \mathbb{E}_x \left( -\frac{\gamma^\delta}{2} \cos \delta \mathcal{C}D\Psi(q) - \frac{\gamma^\delta}{2} \mathcal{C}D\Psi \left( q_*^{k+1,\delta} \right) \right) \right\|_s, \\ E_3 &:= \left\| \mathbb{E}_x \left( \gamma^\delta \frac{\cos \delta}{\delta} e^{-\delta \Gamma_2} v \right) - \frac{1}{\delta} v + \Gamma_2 v + \mathbb{E}_x \left[ \frac{\gamma^\delta - 1}{\delta} e^{-\delta \Gamma_2} v \right] \right\|_s, \\ E_4 &:= \frac{1}{\delta} \left\| \mathbb{E}_x \left[ \gamma^\delta \cos \delta \xi^\delta + \left( \gamma^\delta - 1 \right) \xi^\delta \right] \right\|_s, \end{aligned}$$

by the definition of $G^\delta$ (equation (5.4)) and using (5.20) and (5.21), we obtain

$$\left\| G_1^\delta(x) - G_1(x) \right\|_s \le A_1 + A_2 + A_3 \quad \text{and} \quad \left\| G_2^\delta(x) - G_2(x) \right\|_s \le E_1 + E_2 + E_3 + E_4.$$

We will bound the $A_i$'s and the $E_i$'s one by one. To this end, we will repeatedly use the following simple bounds:

$$\gamma^\delta, \gamma^k \in \{0, 1\} \quad \text{and} \quad 0 \le \alpha^\delta \le 1; \tag{5.29}$$

$$\mathbb{E}[\xi^\delta] = 0; \tag{5.30}$$

$$\left\| \mathbb{E}[(\alpha^\delta - 1)\xi^\delta] \right\|_s \le \left[ \mathbb{E}(\alpha^\delta - 1)^2 \right]^{1/2} \left[ \mathbb{E} \| \xi^\delta \|_s^2 \right]^{1/2} \lesssim \delta^{5/2}(1 + \|q\|_s^4 + \|v\|_s^4). \tag{5.31}$$

(5.31) follows from using Bochner's inequality[3] and Cauchy–Schwartz first and then (5.24) and (5.16). Using (5.29), it is straightforward to see that

$$A_1 \le \delta \|q\|_s.$$

As for $A_2$,

$$
\begin{aligned}
A_2 &= \left\| \left( I - \mathbb{E}_x(\alpha^\delta) \frac{\sin \delta}{\delta} e^{-\delta \Gamma_2} \right) v \right\|_s \\
&\le \left| 1 - \mathbb{E}_x(\alpha^\delta) \right| \|v\|_s + \left\| \mathbb{E}_x(\alpha^\delta) \left( 1 - \frac{\sin \delta}{\delta} e^{-\delta \Gamma_2} \right) v \right\|_s \\
&\le \delta^2 (1 + \|q\|_s^4 + \|v\|_s^4) \|v\|_s + \delta \|v\|_s \le \delta (1 + \|q\|_s^6 + \|v\|_s^6),
\end{aligned}
$$

having used, in the second inequality, (5.16) and (5.29). $A_3$ is bounded by using (5.30), (5.31) and (5.17):

$$
\begin{aligned}
A_3 &\le \left\| \frac{\sin \delta}{\delta} \mathbb{E}_x[(\alpha^\delta - 1)\xi^\delta + \xi^\delta] \right\|_s + \left\| \mathbb{E}_x \delta \mathcal{C} D\Psi(q) \right\|_s \\
&\lesssim \delta^{5/2}(1 + \|q\|_s^4 + \|v\|_s^4) + \delta(1 + \|q\|_s) \le \delta(1 + \|q\|_s^4 + \|v\|_s^4).
\end{aligned}
$$

Hence (5.27) has been proven. We now come to estimating the $E_i$'s. Proceeding as in the bound for $A_2$ above we obtain:

$$
\begin{aligned}
E_1 &\le \left\| q - \mathbb{E}_x(\alpha^\delta) q \right\|_s + \left\| \mathbb{E}_x(\alpha^\delta) \left( 1 - \frac{\sin \delta}{\delta} \right) q \right\|_s \\
&\le \delta^2 (1 + \|q\|_s^4 + \|v\|_s^4) \|q\|_s + \delta^2 \|q\|_s \le \delta^2 (1 + \|q\|_s^6 + \|v\|_s^6).
\end{aligned}
$$

---

[3]Let $(X, \|\cdot\|)$ be a Banach space and $f \in L^1((\Omega, \mathcal{F}, \mu); X)$. Then $\| \int f \, d\mu \| \le \int \|f\| \, d\mu$. For a proof of the Bochner's inequality see [19].

Also,

$$
\begin{aligned}
E_2 &\leq \left\| CD\Psi(q) - \mathbb{E}_x(\alpha^\delta)\cos\delta CD\Psi(q) \right\|_s \\
&\quad + \left\| \tfrac{1}{2}\mathbb{E}_x\big[\alpha^\delta \cos\delta CD\Psi(q) - \alpha^\delta CD\Psi(q_*^{k+1,\delta})\big] \right\|_s \\
&\lesssim \left\| (1 - \mathbb{E}_x(\alpha^\delta))CD\Psi(q) \right\|_s + \left\| (\cos\delta - 1)CD\Psi(q) \right\|_s \\
&\quad + \left\| \mathbb{E}_x\big(CD\Psi(q) - CD\Psi(q_*^{k+1,\delta})\big) \right\|_s \\
&\lesssim \delta^2\big(1 + \|q\|_s^6 + \|v\|_s^6\big) + \delta\mathbb{E}_x\big\|q_*^{k+1,\delta} - q\big\|_s \overset{(5.26)}{\lesssim} \delta\big(1 + \|q\|_s^6 + \|v\|_s^6\big),
\end{aligned}
$$

where the penultimate inequality is obtained by using (5.16) and (5.17).

For the last two terms:

$$
\begin{aligned}
E_3 &\leq \frac{1}{\delta}\left\| \mathbb{E}_x(\alpha^\delta)(\cos\delta - 1)e^{-\delta\Gamma_2}v \right\|_s + \frac{1}{\delta}\left\| \mathbb{E}_x(\alpha^\delta - 1)e^{-\delta\Gamma_2}v \right\|_s \\
&\quad + \frac{1}{\delta}\left\| \mathbb{E}_x(e^{-\delta\Gamma_2} - 1 + \delta\Gamma_2)v \right\|_s \\
&\overset{(5.29)}{\lesssim} \delta\|v\|_s + \frac{1}{\delta}\mathbb{E}\big|\alpha^\delta - 1\big|\|v\|_s \\
&\overset{(5.16)}{\lesssim} \delta\|v\|_s + \delta\big(1 + \|q\|_s^4 + \|v\|_s^4\big)\|v\|_s \lesssim \delta\big(1 + \|q\|_s^6 + \|v\|_s^6\big).
\end{aligned}
$$

Finally, from (5.30) and (5.31),

$$
\begin{aligned}
E_4 &\leq \frac{1}{\delta}\left\| \mathbb{E}_x(\alpha^\delta \cos\delta\xi^\delta) \right\|_s + \frac{1}{\delta}\left\| \mathbb{E}_x(\alpha^\delta - 1)\xi^\delta \right\|_s \\
&\lesssim \frac{1}{\delta}\left\| \mathbb{E}_x\big[(\alpha^\delta - 1)\cos\delta\xi^\delta + \cos\delta\xi^\delta\big] \right\|_s + \frac{1}{\delta}\left\| \mathbb{E}_x\big[(\alpha^\delta - 1)\xi^\delta\big] \right\|_s \\
&\leq \delta^{3/2}\big(1 + \|q\|_s^4 + \|v\|_s^4\big).
\end{aligned}
$$

This concludes the proof. □

Let us now show that condition (5.11) is satisfied as well.

**Lemma 5.11.** *Under Assumption* 3.1, *the chain* $x^{k,\delta} \in \mathcal{H}^s \times \mathcal{H}^s$ *defined in* (5.1) *satisfies*

$$
\mathbb{E}\big[\big\|q^{k+1,\delta} - q^{k,\delta}\big\|_s \,\big|\, x^{k,\delta} = x\big] \lesssim \delta\big(1 + \|q\|_s + \|v\|_s\big), \tag{5.32}
$$

$$
\mathbb{E}\big[\big\|v^{k+1,\delta} - v^{k,\delta}\big\|_s \,\big|\, x^{k,\delta} = x\big] \lesssim \delta^{1/2}\big(1 + \|q\|_s^6 + \|v\|_s^6\big). \tag{5.33}
$$

*In particular,* (5.11) *holds with* $r = 1/2$ *and* $n = 6$.

**Proof.** (5.32) is a straightforward consequence of (5.28), (5.29) and (5.26). In order to prove (5.33) we start from (5.28) and we write

$$\mathbb{E}\big[\big\|v^{k+1,\delta} - v^{k,\delta}\big\|_s |x^{k,\delta} = x\big] = \mathbb{E}\big\|\gamma^\delta v_* + \big(\gamma^\delta - 1\big)v' - v\big\|_s \tag{5.34}$$

$$\lesssim \mathbb{E}\big\|\gamma^\delta\big(\sin\delta q + \delta\mathcal{C}D\Psi(q) + \delta\mathcal{C}D\Psi(q_*)\big)\big\|_s \tag{5.35}$$

$$+ \mathbb{E}\big\|\gamma^\delta\cos\delta v' - \big(1 - \gamma^\delta\big)v' - v\big\|_s. \tag{5.36}$$

By using (5.29), (5.17) and (5.26), we get

$$\mathbb{E}\big\|\gamma^\delta\big(\sin\delta q + \delta\mathcal{C}D\Psi(q) + \delta\mathcal{C}D\Psi(q_*)\big)\big\|_s \lesssim \delta\big(1 + \|q\|_s\big). \tag{5.37}$$

Notice that

$$\mathbb{E}\big|\gamma^\delta - 1\big|^\ell = 1 - \mathbb{E}\big(\alpha^\delta\big), \qquad \forall\ell \geq 1. \tag{5.38}$$

Therefore by (5.20) and (5.29) and repeatedly using (5.38),

$$\mathbb{E}\big\|\gamma^\delta\cos\delta v' - \big(1 - \gamma^\delta\big)v' - v\big\|_s \lesssim \mathbb{E}\big\|\big[\gamma^\delta\cos\delta - \big(1 - \gamma^\delta\big)\big]\xi^\delta\big\|_s + \mathbb{E}\big\|\big(1 - \gamma^\delta\big)e^{-\delta\Gamma_2}v\big\|_s$$

$$+ \mathbb{E}\big\|\gamma^\delta\cos\delta e^{-\delta\Gamma_2}v - v\big\|_s$$

$$\overset{(5.24)}{\lesssim} \delta^{1/2} + \mathbb{E}\big|1 - \alpha^\delta\big|\|v\|_s + \mathbb{E}\big|1 - \gamma^\delta\cos\delta\big|\big\|e^{-\delta\Gamma_2}v\big\|_s$$

$$+ \big\|e^{-\delta\Gamma_2}v - v\big\|_s \tag{5.39}$$

$$\overset{(5.16)}{\lesssim} \delta^{1/2} + \delta^2\big(1 + \|q\|_s^4 + \|v\|_s^4\big)\|v\|_s + \mathbb{E}\big|\gamma^\delta - 1\big|\big\|e^{-\delta\Gamma_2}v\big\|_s$$

$$+ \mathbb{E}\big|\gamma^\delta(\cos\delta - 1)\big|\big\|e^{-\delta\Gamma_2}v\big\|_s + \delta\|v\|_s$$

$$\lesssim \delta^{1/2}\big(1 + \|q\|_s^6 + \|v\|_s^6\big).$$

Now (5.34), (5.37) and (5.39) imply (5.33). □

Finally, the a priori bound (5.12) holds.

**Lemma 5.12.** *Let Assumption 3.1 hold. Then the chain (5.1) satisfies*

$$\sup_{\delta\in(0,1/2)}\bigg\{\delta\mathbb{E}\bigg[\sum_{k\delta<T}\big\|x^{k,\delta}\big\|_{s\times s}^\ell\bigg]\bigg\} < \infty \qquad \text{for any integer } \ell \geq 1. \tag{5.40}$$

*In particular, the bound (5.12) holds (with $\varepsilon = 1$ and for any moment of $\|x^{k,\delta}\|_{s\times s}$).*

***Remark 5.13.*** Before proving the above lemma, let us make some comments. First of all, the estimate of condition (5.12) is needed mainly because the process has not been started in stationarity and hence it is not stationary. For the same reason an analogous estimate was needed in

[17], as well. However there the approximate drift grows linearly in $x$ (see [17], equation (25)) and this is sufficient to prove an estimate of the type (5.12). Here, because in case of rejection of the proposed move the sign of the velocity is flipped, the approximate drift grows faster than linearly (see Lemma 5.10 and (5.14)). To deal with the change of sign of the velocity, we will observe that such a change of sign doesn't matter if we look at even powers of $x^{k,\delta}$ – what matters is that in moving from $x^{k,\delta}$ to $x^{k+1,\delta}$ we always "move a short distance" – and we will exploit the independence of $v^{k,\delta}$ and $\xi^\delta$, once $x^{k,\delta}$ is given.

**Proof of Lemma 5.12.** If we show that (5.40) is true for every even $\ell$, then it is true for every $\ell \geq 1$. Indeed

$$\|x\|_{s\times s}^\ell \leq \|x\|_{s\times s}^{2\ell} + 1 \quad \text{so} \quad \delta \sum_{k\delta < T} \|x\|_{s\times s}^\ell \lesssim \delta \sum_{k\delta < T} \|x\|_{s\times s}^{2\ell} + 1 < \infty.$$

Throughout this proof, $c$ will be a generic positive constant. We begin by recalling the definition of the map $\Theta_1^{\delta/2}$:

$$\Theta_1^{\delta/2}(q, p) = \left( q, v - \frac{\delta}{2}CD\Psi(q) \right),$$

hence

$$\begin{aligned}
\left\| \Theta_1^{\delta/2}(x) \right\|_{s\times s}^2 &= \|q\|_s^2 + \|v\|_s^2 + \frac{\delta^2}{4}\|CD\Psi(q)\|_s^2 - \delta\langle v, CD\Psi(q)\rangle \\
&\overset{(5.17)}{\leq} \|q\|_s^2 + \|v\|_s^2 + c\delta^2\left(1 + \|q\|_s^2\right) + c\delta\|v\|_s\left(1 + \|q\|_s\right) \\
&\leq (1 + c\delta)\|x\|_{s\times s}^2 + c\delta.
\end{aligned}$$

Because $R^\delta$ is a rotation, it preserves the norm, so also

$$\left\| \chi^\delta(x) \right\|_{s\times s}^2 = \left\| \Theta_1^{\delta/2} \circ R^\delta \circ \Theta_1^{\delta/2}(x) \right\|_{s\times s}^2 \leq (1 + c\delta)\|x\|_{s\times s}^2 + c\delta. \tag{5.41}$$

Now notice that by definition $x^{k+1,\delta} = (q^{k+1,\delta}, v^{k+1,\delta})$ is either equal to $\chi^\delta(q^{k,\delta}, (v^{k,\delta})') = (q_*^{k+1,\delta}, v_*^{k+1,\delta})$ (if the proposal is accepted) or to $(q^{k,\delta}, -(v^{k,\delta})')$ (if the proposal is rejected). Thanks to (5.41), in any of these two cases we have

$$\left\| x^{k+1,\delta} \right\|_{s\times s}^2 \leq (1 + c\delta)\left\| (x^{k,\delta})' \right\|_{s\times s}^2 + c\delta,$$

where $(x^{k,\delta})' = ((q^{k,\delta})', (v^{k,\delta})') = (q^{k,\delta}, (v^{k,\delta})')$. By (5.20),

$$\begin{aligned}
\left\| (x^{k,\delta})' \right\|_{s\times s}^2 &\leq \left\| q^{k,\delta} \right\|_s^2 + \left\| v^{k,\delta} \right\|_s^2 + \left\| \xi^\delta \right\|_s^2 + 2\langle e^{-\delta\Gamma_2}v^{k,\delta}, \xi^\delta\rangle \\
&= \left\| x^{k,\delta} \right\|_{s\times s}^2 + \left\| \xi^\delta \right\|_s^2 + 2\langle e^{-\delta\Gamma_2}v^{k,\delta}, \xi^\delta\rangle.
\end{aligned}$$

Therefore

$$\mathbb{E}\|x^{k+1,\delta}\|^2_{s\times s} = \mathbb{E}\{\mathbb{E}[\|x^{k+1,\delta}\|^2_{s\times s}|x^{k,\delta}]\}$$

$$\leq (1+c\delta)\mathbb{E}\|x^{k,\delta}\|^2_s + (1+c\delta)\mathbb{E}\|\xi^\delta\|^2_s$$

$$+ (1+c\delta)\mathbb{E}\{\mathbb{E}[\langle e^{-\delta\Gamma_2}v^{k,\delta},\xi^\delta\rangle|x^{k,\delta}]\} + c\delta.$$

By the conditional independence of $v^{k,\delta}$ and $\xi^\delta$ together with (5.30)

$$\mathbb{E}\{\mathbb{E}[\langle e^{-\delta\Gamma_2}v^{k,\delta},\xi^\delta\rangle|x^{k,\delta}]\} = 0;$$

hence, using (5.24), we obtain

$$\mathbb{E}\|x^{k+1,\delta}\|^2_{s\times s} \leq (1+c\delta)\mathbb{E}\|x^{k,\delta}\|^2_{s\times s} + c\delta.$$

Iterating the above inequality leads to

$$\mathbb{E}\|x^{k+1,\delta}\|^2_{s\times s} \leq (1+c\delta)^{[T/\delta]}\mathbb{E}\|x^0\|^2_{s\times s} + c\delta(1+c\delta)^{[T/\delta]} + c\delta,$$

which implies

$$\delta\sum_{k\delta<T}\mathbb{E}\|x^{k,\delta}\|^2_{s\times s} < \infty.$$

We now need to show that for any $j > 1$,

$$\delta\sum_{k\delta<T}\mathbb{E}\|x^{k,\delta}\|^{2j}_{s\times s} < \infty.$$

By the same reasoning as before we start with observing that

$$\|\chi^\delta(x)\|^{2j}_{s\times s} \leq (1+c\delta)\|x\|^{2j}_{s\times s} + c\delta + 2\sum_{l=1}^{j-1}(1+c\delta)^l\|x\|^{2l}_{s\times s}\delta^{j-l}$$

$$\leq (1+c\delta)\|x\|^{2j}_{s\times s} + c\delta$$

(notice that in the above $j - l \geq 1$ because $1 \leq l \leq j - 1$). Hence

$$\mathbb{E}\|x^{k+1,\delta}\|^{2j}_{s\times s} \leq (1+c\delta)\mathbb{E}\|(x^{k,\delta})'\|^{2j}_{s\times s} + c\delta.$$

From (5.20), we have

$$\|(x^{k,\delta})'\|^{2j}_{s\times s} \leq \|x^{k,\delta}\|^{2j}_{s\times s} + \|\xi^\delta\|^{2j}_{s\times s} + c(\langle e^{-\delta\Gamma_2}v^{k,\delta},\xi^\delta\rangle)^j$$

$$+ c\sum_{l=1}^{j-1}\|x^{k,\delta}\|^{2l}_{s\times s}\|\xi^\delta\|^{2(j-l)}_s$$

$$+ c \sum_{l=1}^{j-1} \|x^{k,\delta}\|_{s\times s}^{2l} \big(\big(e^{-\delta\Gamma_2} v^{k,\delta}, \xi^\delta\big)\big)^{j-l}$$

$$+ c \sum_{l=1}^{j-1} \|\xi^\delta\|_s^{2l} \big(\big(e^{-\delta\Gamma_2} v^{k,\delta}, \xi^\delta\big)\big)^{j-l}.$$

Using again the conditional independence of $v^{k,\delta}$ and $\xi^\delta$, for any $l > 1$,

$$\mathbb{E}\big\{\mathbb{E}\big[\big(\big(e^{-\delta\Gamma_2} v^{k,\delta}, \xi^\delta\big)\big)^l | x^{k,\delta} = x\big]\big\} \leq \mathbb{E}\big\{\mathbb{E}\big[\|v^{k,\delta}\|_s^l \|\xi^\delta\|_s^l | x^{k,\delta} = x\big]\big\}$$
$$\leq c\delta^{l/2} \mathbb{E}\|v^{k,\delta}\|_s^l \leq c\delta \mathbb{E}\|v^{k,\delta}\|_s^l.$$

Therefore,

$$\mathbb{E}\big\|\big(x^{k,\delta}\big)'\big\|_{s\times s}^{2j} \leq \mathbb{E}\big\|x^{k,\delta}\big\|_{s\times s}^{2j} + \delta^j + \delta\mathbb{E}\big(1 + \big\|x^{k,\delta}\big\|_{s\times s}^{2j}\big)$$

hence

$$\mathbb{E}\big\|x^{k+1,\delta}\big\|_{s\times s}^{2j} \leq (1 + c\delta)\mathbb{E}\big\|x^{k,\delta}\big\|_{s\times s}^{2j} + c\delta$$

and we can conclude as before. $\qquad\square$

## 5.3. Analysis of the noise

Let us start with defining

$$D^\delta(x) := \mathbb{E}\big[M^{k,\delta} \otimes M^{k,\delta} | x^{k,\delta} = x\big]. \tag{5.42}$$

This section is devoted to proving the invariance principle Lemma 5.18 below, as a consequence of the following Lemma 5.16 and Lemma 5.17, which we prove in Appendix B. In order to state such lemmata, consider the following set of conditions:

**Condition 5.14.** *The Markov chain* $x^{k,\delta} \in \mathcal{H}^s \times \mathcal{H}^s$ *defined in* (5.1) *satisfies*:

(i) *There exist two integers* $d_1, d_2 \geq 1$ *and two real numbers* $b_1, b_2 > 0$ *such that*

$$\big|\langle\hat{\varphi}_j^\ell, D^\delta(x)\hat{\varphi}_i^{\bar\ell}\rangle_{s\times s} - \langle\hat{\varphi}_j^\ell, \mathfrak{C}_s\hat{\varphi}_i^{\bar\ell}\rangle_{s\times s}\big| \lesssim \delta^{b_1}\big(1 + \|x\|_{s\times s}^{d_1}\big)$$
$$\forall i, j \in \mathbb{N} \text{ and } \ell, \bar\ell \in 1, 2; \tag{5.43}$$

$$\big|\mathrm{Trace}_{\mathcal{H}^s\times\mathcal{H}^s}\big(D^\delta(x)\big) - \mathrm{Trace}_{\mathcal{H}^s\times\mathcal{H}^s}\big(\mathfrak{C}_s\big)\big| \lesssim \delta^{b_2}\big(1 + \|x\|_{s\times s}^{d_2}\big), \tag{5.44}$$

*where* $\mathfrak{C}_s$ *is the covariance operator defined in* (3.4).

(ii) *There exist four real numbers* $\eta_1, \eta_2, \eta_3, \eta_4$ *such that*

$$b_1 + 1 - \eta_1 > 0, \qquad b_2 + 1 - \eta_2 > 0, \qquad 4r - \eta_3 > 0, \qquad 4 - \eta_4 > 0.$$

*Moreover, the bound*

$$\sup_{\delta \in (0,1/2)} \left\{ \delta^\eta \mathbb{E} \left[ \sum_{k\delta \leq T} \left\| x^{k,\delta} \right\|_{s\times s}^m \right] \right\} < \infty \tag{5.45}$$

*holds for $\eta = \min_{i=1,\dots,4}\{\eta_i\}$ and $m = \max\{d_1, d_2, 4n, 4p\}$. In the above $n$ and $r$ are as in Condition 5.2.*

**Remark 5.15.** Because $\|x\|_{s\times s}^b \lesssim \|x\|_{s\times s}^d + 1$ for all $d \geq b$, if (5.45) holds with $m = d$, then it also hold for any $m \leq d$.

**Lemma 5.16.** *If (5.11) is satisfied with $r > 1/4$, then the estimates (5.10) and (5.11) together with Conditions 5.14 imply the invariance principle Lemma 5.18.*

**Proof.** See Appendix B. □

**Lemma 5.17.** *Under Assumption 3.1, the estimates (5.43) and (5.44) hold with $b_1 = b_2 = 1/6$ and $d_1 = d_2 = 10$.*

**Proof.** See Appendix B. □

**Lemma 5.18.** *Under Assumption 3.1, the rescaled noise process $\mathcal{H}^s \times \mathcal{H}^s \ni \tilde{B}^\delta$ defined in (5.7), converges weakly in $C([0,T]; \mathcal{H}^s \times \mathcal{H}^s)$ to $\mathcal{H}^s \times \mathcal{H}^s \ni B = (0, B_2)$ where $B_2$ is a $\mathcal{H}_s$-valued, mean zero $\mathcal{C}_s$ Brownian motion.*

**Proof.** We use Lemma 5.16. Thanks to Lemma 5.10, (5.10) is satisfied with $a = 1$ and $p = 6$. From Lemma 5.11, (5.11) holds with $r = 1/2$. As for Conditions 5.14, Lemma 5.17 proves that Condition 5.14(i) holds. In view of Lemma 5.12, Condition 5.14(ii) is satisfied with $\eta_1 = \eta_2 = \eta_3 = \eta_4 = 1$. □

# 6. Numerics

Before describing the numerical results, we highlight the fact that the function space MALA algorithm of [3] is a special case of the function space HMC algorithm of [2] which, in turn, is a special case of the SOL-HMC algorithm introduced in this paper. All of these algorithms are designed to have dimension-independent mixing times, and are indistinguishable from this point of view. However, we expect to see different performance in practice and our numerical experiments are aimed at demonstrating this. In the paper [2], it was shown that HMC is a significant improvement on MALA for bridge diffusions [2]. It is natural to try and show that the SOL-HMC algorithm can be more efficient than HMC. To do this, we choose a target measure $\pi$ defined with respect to a reference measure $\pi_0$ which is a standard Brownian bridge on $[0, 100]$,

starting and ending at 0, and with

$$\Psi(q) = \frac{1}{2} \int_0^{100} V\big(q(\tau)\big) \, d\tau$$

and $V(u) = (u^2 - 1)^2$. Thus we may take as $\mathcal{H}$ the space $L^2((0, 100), \mathbb{R})$. The properties of measures of this type are studied in some detail in [16]. For our purposes, it is relevant to note that $\mathbb{E}^\pi q(\tau) = 0$ for $0 \le \tau \le 100$. This follows from the fact that the function $V$ is even and zero boundary conditions are imposed by the Brownian bridge meaning that the measure is invariant under $q \mapsto -q$. The precision (inverse covariance) operator for unit Brownian bridge on an interval $[0, T]$ is simply given by the negative of the Laplacian with homogeneous Dirichlet boundary conditions. Furthermore samples may be drawn simply by drawing a Brownian motion $B(t)$ and subtracting $t B(T)/T$.

Because of these properties of $\pi$ we expect that sufficiently long runs of MCMC methods to sample from $\pi$ should exhibit, approximately, this zero mean property. We may use the rate at which this occurs as a way of discriminating between the different algorithms. To this end, we define a quantity $E(n)$, with $n$ defined in what follows as $n = N_d N_M$. Here $N_d$ denotes the number of steps used in the deterministic integration; in MALA, $N_d = 1$ and for our implementation of HMC, $N_d > 1$ is chosen so that such that $\tau = N_d h \approx 1$. The integer $N_M$ is the number of MCMC steps taken. The quantity $n$ is thus a measure of the total number of numerical integration steps used and thus of the overall work required by the algorithm, noting that all the algorithms involve more or less the same calculations per step, and that accept/reject overheads are minimal compared with the cost arising from accumulation of numerical integration steps. We define the running average

$$\bar{q}^n(\tau) = \frac{1}{N_M} \sum_{i=1}^{N_M} q^i(\tau),$$

where the index $i$ runs over the realizations $N_M$ of the path $q$. We then define the quantity $E(n)$ as

$$E(n) = \frac{1}{100} \int_0^{100} \big|\bar{q}^n(\tau)\big| \, d\tau.$$

When viewed as a function of $n$ the rate at which $E(n)$ approaches zero determines the efficiency of the sampling. The faster $E$ decreases, the more efficient the sampling. All our numerical experiments are conducted with the form of the SOL-HMC algorithm described in Remarks 2.2. Thus $\Gamma_2 = I$ and we use the parameter $\iota$ to implicitly define $\delta$.

For the first set of numerical experiments, we use the SOL-HMC algorithm in the form which gives rise to the diffusion limit, namely with $\delta = h$ so that we make only one step in the deterministic integration $N_d = 1$. The key parameter is thus $\iota$ (and implicitly $\delta$) given in (2.16). We consider the values $\iota = 0.9, 0.99$ and $0.999$. The results are summarized in Figure 1. They demonstrate that SOL-HMC is indeed considerably better than MALA, but is not better than HMC for this problem and this choice of parameters. For $\iota = 0.999$, we see a plateauing of the value for $E$ for $n$ between 500 and 1000. It seems that such behavior is due to the sign-flip step
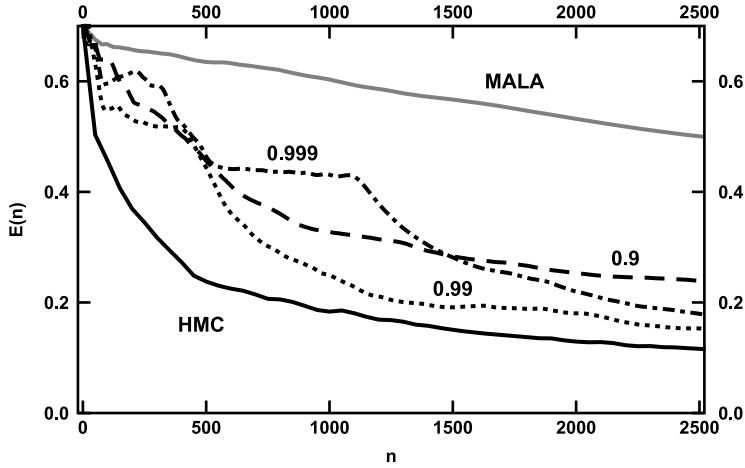
**Figure 1.** The quantity $E$ plotted as a function of number of iterations. The HMC algorithm has the best performance; MALA has the worst. For the values of $\iota$ used (0.9, 0.99 and 0.999) SOL-HMC is considerably better than MALA, but not better than HMC.

when the proposal is rejected. As Horowitz [11] noted for the standard finite dimensional algorithm L2MC that we have generalized, "If it were not for this momenta reversal, it would be a near certain conclusion that L2MC is more efficient than HMC". Our findings are consistent with this remark.

The first set of experiments, in which HMC appeared more efficient than SOL-HMC, employed a value of $\delta$ which corresponds to making small changes in the momentum in each proposal. For the second set of numerical experiments, we relax this constraint and take $\iota = 2^{-1/2}$ in all our SOL-HMC simulations. This corresponds to taking an equal mixture of the current momentum and an independent draw from its equilibrium value. Furthermore, these experiments use more than one step in the deterministic integration, $N_d > 1$. For the HMC integration, we use $N_d = 50$, $\tau = 1$ and of course, $\iota = 0$. The results are summarized in Figure 2 where we show the behaviour of the HMC algorithm in comparison with four choices of parameters in the SOL-HMC algorithm: (i) $N_d = 10$; (ii) $N_d = 25$; (iii) $N_d = 50$; and in (iv) $N_d$ is a random value, uniformly distributed between 25 and 75, in each step of the MCMC algorithm. We see that if $N_d \geq 25$ the SOL-HMC algorithm shows improved behaviour in comparison with the HMC algorithm.

## 7. Conclusions

We have introduced a parametric family of MCMC methods, the SOL-HMC algorithms, suitable for sampling measures defined via density with respect to a Gaussian. The parametric family includes a range of existing methods as special parameter choices including the function space MALA and HMC methods introduced in [2,3]. Whilst both these algorithms are reversible with respect to the target, generic parameter choices in the SOL-HMC algorithm lead to irreversible
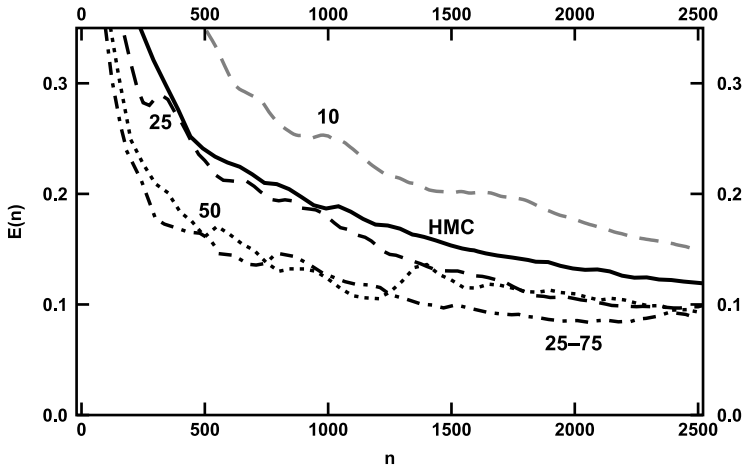
**Figure 2.** Figure showing that SOL-HMC can exhibit faster mixing than HMC when run in the regime with $\iota = 2^{-1/2}$, for appropriate choice of $\tau$ (reflected in the integer $N_d$ which labels the graphs – 25, 50, 75 and 25–75 for the random case). Note that for HMC, by definition, $\iota = 0$ and we fix $\tau \approx 1$. Note also that the vertical scale is half that of the previous graph.

algorithms which preserve the target. With one particular parameter choice, we show that the algorithm has a diffusion limit to the second order Langevin equation; this latter limit makes clear the role of irreversibility in the algorithm. Numerical results indicate that the method is comparable with the function space HMC method of [2] which, in turn, is superior to the function space MALA method of [3]. Indeed, in the example studied, we are able to exhibit situations for which the SOL-HMC algorithm outperforms the HMC method. Further application of the method is thus suggested.

We make an important observation about the diffusion limits proved in this paper, Theorem 5.1, and in [17], both of which concern algorithms that have been specifically designed to deal with target measures defined via density with respect to a Gaussian; indeed both methods would suffer no rejections in the pure Gaussian case. The limit theorems demonstrate that the number of steps required to sample may be chosen independently of the dimension of the approximating space $N$. However, in contrast to the diffusion limits identified in [14,18] the theory does not reveal an optimal choice for the time-step, or an optimal acceptance probability. The fact that an optimal acceptance probability, and implicitly an optimal time-step, can be identified in [14,18] is precisely because the proposal does not exactly preserve the underlying Gaussian reference measure and the universal optimal acceptance probability is determined purely by the Gaussian properties of the problem; the change of measure, and hence function $\Psi$, play no role. Once improved methods are used, such as SOL-HMC and the pCN method analyzed in [17], which exactly preserve the Gaussian structure, no such universal behaviour can be expected and optimality must be understood on a case by case basis.

# Appendix A

This Appendix contains the proof of Theorem 2.1, of the identity (5.8) and a sketch of the proof of Proposition 5.4.

**Proof of Theorem 2.1.** By following the arguments given in Section 5.3 of [15], it suffices to show $\chi_\tau^h$ preserves $\Pi$ when appended with the suitable accept–reject mechanism. We show this using the fact that the finite dimensional version of the above algorithm preserves the corresponding invariant measure. Since the proof of this is very similar to that of Theorem 4.2 we only sketch the details. For $N \in \mathbb{N}$ and $t > 0$ define the map

$$\chi_N^t = \Theta_{N,1}^{t/2} \circ \mathsf{R}_N^t \circ \Theta_{N,1}^{t/2},$$

where $R_N^t$ and $\Theta_N^t$ are obtained by restricting $R^t$ and $\Theta_1^t$ respectively on the first $N$ components of $(q, v)$ and with $\Psi$ replaced by $\Psi^N$, as in the proof of Theorem 4.2. Also following the ideas used in the proof of Theorem 4.2, it may be shown that $\lim_{N\to\infty} \chi_N^t(q, v) = \chi^t(q, v)$ for any $t$ and $(q, v) \in \mathcal{H}^s \times \mathcal{H}^s$. Now the proof can be completed via a dominated convergence argument similar to that used in the proof of Theorem 4.2. $\qquad\square$

**Proof of (5.8).** We recall that for any integer $j \geq 0$

$$x^{j+1,\delta} - x^{j,\delta} = \delta G^\delta\left(x^{j,\delta}\right) + \sqrt{2\delta\mathcal{S}}M^{j,\delta}. \tag{A.1}$$

With this in mind (5.8) is a straightforward consequence of the definition of $z^\delta(t)$ and $\bar{z}^\delta(t)$, given in (5.2) and (5.3), respectively; indeed if $t_k \leq t < t_{k+1}$ then

$$
\begin{aligned}
z^\delta(t) &= \frac{1}{\delta}(t - t_k)x^{k+1,\delta} + \frac{1}{\delta}(t_{k+1} - t)x^{k,\delta} \\
&\stackrel{\text{(A.1)}}{=} \frac{1}{\delta}(t - t_k)\left[x^{k,\delta} + G^\delta\left(x^{k,\delta}\right)\delta + \sqrt{2\delta\mathcal{S}}M^{k,\delta}\right] + \frac{1}{\delta}(t_{k+1} - t)x^{k,\delta} \\
&= \frac{1}{\delta}(t_{k+1} - t_k)x^{k,\delta} + (t - t_k)G^\delta\left(\bar{z}^\delta(t)\right) + \sqrt{\frac{2\mathcal{S}}{\delta}}(t - t_k)M^{k,\delta} \\
&= x^{k,\delta} + (t - t_k)G^\delta\left(\bar{z}^\delta(t)\right) + \sqrt{\frac{2\mathcal{S}}{\delta}}(t - t_k)M^{k,\delta} \\
&= x^{k,\delta} + \int_{t_k}^t G^\delta\left(\bar{z}^\delta(u)\right)\mathrm{d}u + \sqrt{\frac{2\mathcal{S}}{\delta}}(t - t_k)M^{k,\delta}.
\end{aligned}
\tag{A.2}
$$

Equation (5.8) comes from extending the above equality to the whole interval $[0, t]$. More precisely, taking the sum for $j = 0, \ldots, k - 1$ on both sides of (A.1), we obtain

$$x^{k,\delta} = x_0 + \int_0^{t_k} G^\delta\left(\bar{z}^\delta(u)\right)\mathrm{d}u + \sqrt{2\mathcal{S}\delta}\sum_{j=0}^{k-1} M^{j,\delta}. \tag{A.3}$$

Substituting (A.3) into (A.2), we obtain

$$
\begin{aligned}
z^\delta(t) &= x^0 + \int_0^t G^\delta\big(\bar{z}^\delta(u)\big)\,\mathrm{d}u + \sqrt{2\mathcal{S}\delta}\sum_{j=0}^{k-1} M^{j,\delta} + \sqrt{\frac{2\mathcal{S}}{\delta}}(t-t_k)M^{k,\delta} \\
&= x^0 + \int_0^t G\big(z^\delta(u)\big)\,\mathrm{d}u + \tilde{B}^\delta(t) + \int_0^t \big[G^\delta\big(\bar{z}^\delta(u)\big) - G\big(z^\delta(u)\big)\big]\,\mathrm{d}u \\
&= x^0 + \int_0^t G\big(z^\delta(u)\big)\,\mathrm{d}u + \hat{B}^\delta(t) = \Upsilon\big(x^0, \hat{B}^\delta\big).
\end{aligned}
$$
$\square$

**Proof of Proposition 5.4.** Once (5.8) has been established, this proof is completely analogous to the proof of [17], Lemma 3.5, which is divided in four steps. The only step that we need to specify here is the third, the rest remains unchanged. Such a step consists in showing that

$$
\lim_{\delta\to 0} \mathbb{E}\left[\int_0^T \big\|G^\delta\big(\bar{z}^\delta(u)\big) - G\big(z^\delta(u)\big)\big\|_{s\times s}\,\mathrm{d}u\right] = 0,
$$

where we recall that $z^\delta(t)$ and $\bar{z}^\delta(t)$ have been defined in (5.2) and (5.3), respectively. To this end notice first that if $t_k \le u < t_{k+1}$ then

$$
z^\delta(u) - \bar{z}^\delta(u) = \frac{1}{\delta}(u-t_k)x^{k+1,\delta} + \frac{1}{\delta}(t_{k+1}-u-\delta)x^{k,\delta} = \frac{1}{\delta}(u-t_k)\big(x^{k+1,\delta} - x^{k,\delta}\big).
$$

Hence, $\|z^\delta(u) - \bar{z}^\delta(u)\|_{s\times s} \le \|x^{k+1,\delta} - x^{k,\delta}\|_{s\times s}$. By using this inequality together with (5.10) and (5.11), we have, for $t_k \le u < t_{k+1}$,

$$
\begin{aligned}
&\mathbb{E}\big\|G^\delta\big(\bar{z}^\delta(u)\big) - G\big(z^\delta(u)\big)\big\|_{s\times s} \\
&\quad \le \mathbb{E}\big\|G^\delta\big(\bar{z}^\delta(u)\big) - G\big(\bar{z}^\delta(u)\big)\big\|_{s\times s} + \mathbb{E}\big\|G\big(\bar{z}^\delta(u)\big) - G\big(z^\delta(u)\big)\big\|_{s\times s} \\
&\quad \lesssim \delta^a\big(1 + \mathbb{E}\big\|\bar{z}^\delta(u)\big\|_{s\times s}^p\big) + \mathbb{E}\big\|z^\delta(u) - \bar{z}^\delta(u)\big\|_{s\times s} \\
&\quad \lesssim \delta^a\big(1 + \mathbb{E}\big\|x^{k,\delta}\big\|_{s\times s}^p\big) + \delta^r\big(1 + \mathbb{E}\big\|x^{k,\delta}\big\|_{s\times s}^n\big) \le \delta^{a\wedge r}\big(1 + \mathbb{E}\big\|x^{k,\delta}\big\|_{s\times s}^{p\vee n}\big).
\end{aligned}
$$

Therefore, using (5.12),

$$
\begin{aligned}
&\mathbb{E}\left[\int_0^T \big\|G^\delta\big(\bar{z}^\delta(u)\big) - G\big(z^\delta(u)\big)\big\|_{s\times s}\,\mathrm{d}u\right] \\
&\qquad \lesssim \delta^{1+(a\wedge r)}\sum_{k\delta<T}\big(1 + \mathbb{E}\big\|x^{k,\delta}\big\|_{s\times s}^{p\vee n}\big) \\
&\qquad \lesssim \delta^{a\wedge r} + \delta^{1-\varepsilon+(a\wedge r)}\left(\delta^\varepsilon \sum_{k\delta<T}\mathbb{E}\big\|x^{k,\delta}\big\|_{s\times s}^{p\vee n}\right) \to 0.
\end{aligned}
$$
$\square$

# Appendix B

This Appendix contains the proof of several technical estimates contained in the paper. In particular, it contains the proof of Lemma 5.7, Lemma 5.9, Lemma 5.16 and Lemma 5.17. We start with an analysis of the acceptance probability. We recall that from [2], page 2212, we know that

$$\Delta H(q, v') = \Psi(q) - \Psi(q_*) - \frac{\delta}{2}\big(\langle D\Psi(q), v'\rangle + \langle D\Psi(q_*), v_*\rangle\big)$$

$$+ \frac{\delta^2}{8}\big(\|\mathcal{C}^{1/2}D\Psi(q_*)\|^2 - \|\mathcal{C}^{1/2}D\Psi(q)\|^2\big)$$

$$= \mathcal{F}_\Psi + \frac{\delta^2}{8}\big(\|\mathcal{C}^{1/2}D\Psi(q_*)\|^2 - \|\mathcal{C}^{1/2}D\Psi(q)\|^2\big),$$

having set

$$\mathcal{F}_\Psi := \mathcal{F}_\Psi\big(x, \xi^\delta\big) = \Psi(q) - \Psi(q_*) - \frac{\delta}{2}\big(\langle D\Psi(q), v'\rangle + \langle D\Psi(q_*), v_*\rangle\big).$$

**Proof of Lemma 5.7.** Let

$$\tilde{\alpha}^\delta\big(x, \xi^\delta\big) = \tilde{\alpha}^\delta := 1 \wedge \exp(\mathcal{F}_\Psi) \quad \text{and} \quad \bar{\alpha}^\delta\big(x, \xi^\delta\big) = \bar{\alpha}^\delta := 1 + \mathcal{F}_\Psi \mathbf{1}_{\{\mathcal{F}_\Psi \le 0\}}.$$

Introducing the functions $h, \bar{h} : \mathbb{R} \to \mathbb{R}$ defined as

$$h(y) = 1 \wedge e^y \quad \text{and} \quad \bar{h}(y) = 1 + y\mathbf{1}_{\{y \le 0\}}$$

we have

$$\alpha^\delta = h\left(\mathcal{F}_\Psi + \frac{\delta^2}{8}\big(\|\mathcal{C}^{1/2}D\Psi(q_*)\|^2 - \|\mathcal{C}^{1/2}D\Psi(q)\|^2\big)\right), \tag{B.1}$$

$$\tilde{\alpha}^\delta = h(\mathcal{F}_\Psi) \quad \text{and} \quad \bar{\alpha}^\delta = \bar{h}(\mathcal{F}_\Psi). \tag{B.2}$$

Clearly,

$$\mathbb{E}_x\big|1 - \alpha^\delta\big|^p \lesssim \mathbb{E}_x\big|\alpha^\delta - \tilde{\alpha}^\delta\big|^p + \mathbb{E}_x\big|\tilde{\alpha}^\delta - \bar{\alpha}^\delta\big|^p + \mathbb{E}_x\big|\bar{\alpha}^\delta - 1\big|^p.$$

We will show that

$$\mathbb{E}_x\big|\alpha^\delta - \tilde{\alpha}^\delta\big|^p \lesssim \delta^{3p}\big(1 + \|q\|_s^{2p} + \|v\|_s^{2p}\big), \tag{B.3}$$

$$\mathbb{E}_x\big|\tilde{\alpha}^\delta - \bar{\alpha}^\delta\big|^p \lesssim \mathbb{E}_x|\mathcal{F}_\Psi|^{2p}, \tag{B.4}$$

$$\mathbb{E}_x\big|\bar{\alpha}^\delta - 1\big|^p \lesssim \mathbb{E}_x|\mathcal{F}_\Psi|^p. \tag{B.5}$$

The above three bounds, together with

$$\mathbb{E}_x|\mathcal{F}_\Psi|^p \lesssim \delta^{2p}\big(1 + \|q\|_s^{2p} + \|v\|_s^{2p}\big), \tag{B.6}$$

imply (5.16). Let us start with (B.3): from (B.1) and (B.2) and using the Lipshitzianity of the function $h$ we have

$$
\begin{aligned}
\mathbb{E}_x \left| \alpha^\delta - \tilde{\alpha}^\delta \right|^p &= \mathbb{E}_x \left| h \left( \mathcal{F}_\Psi + \frac{\delta^2}{8} \left( \left\| \mathcal{C}^{1/2} D\Psi(q_*) \right\|^2 - \left\| \mathcal{C}^{1/2} D\Psi(q) \right\|^2 \right) \right) - h(\mathcal{F}_\Psi) \right|^p \\
&\lesssim \mathbb{E}_x \left| \delta^2 \left( \left\| \mathcal{C}^{1/2} D\Psi(q_*) \right\|^2 - \left\| \mathcal{C}^{1/2} D\Psi(q) \right\|^2 \right) \right|^p \\
&\lesssim \delta^{2p} \mathbb{E}_x \left[ \|q_* - q\|_s^{2p} + \|q\|_s^p \|q_* - q\|_s^p \right] \\
&\lesssim \delta^{3p} \left( 1 + \|q\|_s^{2p} + \|v\|_s^{2p} \right),
\end{aligned}
$$

having used the elementary inequality $\|a\|^2 - \|b\|^2 \leq \|a-b\|^2 + 2\|b\| \|a-b\|$, (5.19) and (5.26). (B.4) follows after observing that $|h(y) - \bar{h}(y)| \leq \frac{y^2}{2}$. (B.5) is a consequence of $|\bar{\alpha}^\delta - 1| = |\bar{h}(\mathcal{F}_\Psi) - \bar{h}(0)|$, together with the Lipshitzianity of $\bar{h}$. We are left with showing (B.6). To this end notice first that from (5.21)

$$
\begin{aligned}
\mathbb{E}_x \left\| q_* - q - \delta v' \right\|_s^p &= \mathbb{E}_x \left\| (\cos \delta - 1)q + (\sin \delta - \delta)v' - \frac{\delta}{2} \sin \delta \mathcal{C} D\Psi(q) \right\|_s^p \\
&\lesssim \delta^{2p} \left( 1 + \|q\|_s^p + \|v\|_s^p \right),
\end{aligned}
\tag{B.7}
$$

having used (5.25) and (5.17). Analogously, from the definition of $v_*$ (5.21),

$$
\mathbb{E}_x \left\| v_* - v' \right\|_s^p \lesssim \delta^p \left( 1 + \|q\|_s^p + \|v\|_s^p \right).
\tag{B.8}
$$

Therefore

$$
\begin{aligned}
\mathbb{E}_x |\mathcal{F}_\Psi|^p &= \mathbb{E}_x \left| \Psi(q) - \Psi(q_*) - \frac{\delta}{2} \left( \langle D\Psi(q), v' \rangle + \langle D\Psi(q_*), v_* \rangle \right) \right|^p \\
&\leq \mathbb{E}_x \left| \Psi(q_*) - \Psi(q) - \langle D\Psi(q), (q_* - q) \rangle \right|^p + \mathbb{E}_x \left| \langle D\Psi(q), q_* - q - \delta v' \rangle \right|^p \\
&\quad + \mathbb{E}_x \left| \frac{\delta}{2} \langle D\Psi(q), v' - v_* \rangle \right|^p + \mathbb{E}_x \left| \frac{\delta}{2} \langle D\Psi(q) - D\Psi(q_*), v_* \rangle \right|^p \\
&\lesssim \mathbb{E}_x \|q_* - q\|_s^{2p} + \left\| D\Psi(q) \right\|_{-s}^p \mathbb{E}_x \|q_* - q - \delta v'\|_s^p \\
&\quad + \delta^p \left\| D\Psi(q) \right\|_{-s}^p \mathbb{E}_x \|v_* - v'\|_s^p + \delta^p \mathbb{E}_x \left( \|q_* - q\|_s^p \|v_*\|_s^p \right) \\
&\lesssim \delta^{2p} \left( 1 + \|q\|_s^{2p} + \|v\|_s^{2p} \right),
\end{aligned}
$$

where in the second inequality we have used (3.6) and (5.18), in the third (3.5) together with (B.7), (B.8), Lemma 5.8 and Lemma 5.9. This concludes the proof. $\qquad \square$

Before starting the proof of Lemma 5.9, recall the notation (5.20)–(5.21).

**Proof of Lemma 5.9.** $\xi^\delta$ is an $\mathcal{H}^s$ valued Gaussian random variable with mean zero and co-variance operator $\mathcal{C}_s(I - \mathrm{e}^{-2\delta\Gamma_2})$. Indeed $\Gamma_2$ is a bounded operator from $\mathcal{H}^s$ into itself and $B_2$ is an $\mathcal{H}^s$ valued $\mathcal{C}_s$-Brownian motion hence $\mathcal{C}_s(I - \mathrm{e}^{-2\delta\Gamma_2})$ is the product of a trace class operator times a bounded operator and therefore it is a trace class operator itself. So

$$\mathbb{E}\|\xi^\delta\|_s^p \lesssim \left(\mathbb{E}\|\xi^\delta\|_s^2\right)^{p/2} = \left[\mathrm{Trace}\!\left(\mathcal{C}_s\!\left(I - \mathrm{e}^{-2\delta\Gamma_2}\right)\right)\right]^{p/2}$$
$$\leq \left(\mathrm{Trace}(\mathcal{C}_s)\|I - \mathrm{e}^{-2\delta\Gamma_2}\|_{\mathcal{L}(\mathcal{H}^s,\mathcal{H}^s)}\right)^{p/2} \lesssim \delta^{p/2}.$$

This proves (5.24). (5.25) is a simple consequence of (5.24) together with (5.20). (5.26) follows from (5.21), using (5.17) and (5.25):

$$\mathbb{E}\big[\|q_*^{k+1,\delta} - q^{k,\delta}\|_s^p | x^{k,\delta} = x\big] \lesssim \mathbb{E}_x \|q(\cos\delta - 1)\|_s^p$$
$$+ \mathbb{E}_x \|\sin\delta v'\|_s^p + \mathbb{E}_x \|\delta \sin\delta \mathcal{C}D\Psi(q)\|_s^p$$
$$\lesssim \delta^{2p}\|q\|_s^p + \delta^p \|v'\|_s^p \lesssim \delta^p\big(1 + \|q\|_s^p + \|v\|_s^p\big). \qquad \square$$

We now turn to prove Lemma 5.16. To this end, we follow [17], Proof of Lemma 3.7, highlighting the slight modifications needed in our context.

**Proof of Lemma 5.16.** In [1], Theorem 5.1, it is shown that proving the weak convergence of $\tilde{B}^\delta$ to $B$ boils down to showing that the following three limits hold in probability:

$$\lim_{\delta \to 0} \delta \sum_{k\delta < t} \mathbb{E}\big[\|M^{k,\delta}\|_{s\times s}^2 | \mathcal{F}^{k,\delta}\big] = t \, \mathrm{Trace}_{\mathcal{H}^s\times\mathcal{H}^s}(\mathfrak{C}_s); \qquad (\mathrm{B}.9)$$

$$\lim_{\delta \to 0} \delta \sum_{k\delta < t} \mathbb{E}\big[\langle M^{k,\delta}, \hat{\varphi}_j^\ell\rangle_{s\times s}\langle M^{k,\delta}, \hat{\varphi}_i^{\bar{\ell}}\rangle_{s\times s} | \mathcal{F}^{k,\delta}\big] = t\langle\hat{\varphi}_j^\ell, \mathfrak{C}_s\hat{\varphi}_i^{\bar{\ell}}\rangle_{s\times s},$$
$$\forall i, j \in \mathbb{N}, \ell, \bar{\ell} \in \{1, 2\}; \qquad (\mathrm{B}.10)$$

$$\lim_{\delta \to 0} \delta \sum_{k\delta < t} \mathbb{E}\big[\|M^{k,\delta}\|_{s\times s}^2 \mathbf{1}_{\{\|M^{k,\delta}\|_s^2 \geq \delta^{-1}\zeta\}} | \mathcal{F}^{k,\delta}\big] = 0, \qquad \forall \zeta > 0. \qquad (\mathrm{B}.11)$$

Here $M^{k,\delta}$ and $D^\delta(x)$ have been defined in (5.5) and (5.42), respectively, and $\mathcal{F}^{k,\delta}$ is the filtration generated by $\{x^{j,\delta}, \gamma^{j,\delta}, \xi^\delta, j = 0, \ldots, k\}$.

Limit (B.9) condition (5.44) implies that

$$\mathbb{E}\big[\|M^{k,\delta}\|_{s\times s}^2 | \mathcal{F}^{k,\delta}\big] = \mathrm{Trace}_{\mathcal{H}^s\times\mathcal{H}^s}\mathfrak{C}_s + \mathbf{e}_1\big(x^{k,\delta}\big),$$

where $|\mathbf{e}_1(x^{k,\delta})| \lesssim \delta^{b_2}(1 + \|x^{k,\delta}\|_{s\times s}^{d_2})$. Therefore

$$\delta \sum_{k\delta < t} \mathbb{E}\big[\|M^{k,\delta}\|_{s\times s}^2 | \mathcal{F}^{k,\delta}\big] = t \, \mathrm{Trace}_{\mathcal{H}^s\times\mathcal{H}^s}\mathfrak{C}_s + \delta \sum_{k\delta < t} \mathbb{E}\big[\mathbf{e}_1\big(x^{k,\delta}\big)\big].$$

Thanks to (5.45), we have

$$\delta \sum_{k\delta < t} \mathbb{E}|\mathbf{e}_1(x^{k,\delta})| \lesssim \delta^{b_2+1} \sum_{k\delta < t} \left(1 + \mathbb{E}\|x^{k,\delta}\|_{s\times s}^{d_1}\right)$$

$$\leq \delta^{b_2} + \delta^{b_2+1-\eta_2}\left(\delta^{\eta_2} \sum_{k\delta < t} \mathbb{E}\|x^{k,\delta}\|_{s\times s}^{d_2}\right) \xrightarrow{\delta \to 0} 0.$$

Limit (B.10) can be proved as a consequence of (5.43) and (5.45), acting as we did to show (B.9).

Limit (B.11) the Cauchy–Schwartz and Markov inequalities give

$$\mathbb{E}\left[\|M^{k,\delta}\|_{s\times s}^2 \mathbf{1}_{\{\|M^{k,\delta}\|_s^2 \geq \delta^{-1}\zeta\}}|\mathcal{F}^{k,\delta}\right] \leq \left(\mathbb{E}\left[\|M^{k,\delta}\|_{s\times s}^4|\mathcal{F}^{k,\delta}\right]\right)^{1/2}\left(\mathbb{P}\left[\|M^{k,\delta}\|_{s\times s}^2 > \delta^{-1}\zeta\right]\right)^{1/2}$$

$$\leq \frac{\delta}{\zeta}\mathbb{E}\left[\|M^{k,\delta}\|_{s\times s}^4|\mathcal{F}^{k,\delta}\right],$$

hence we need to estimate $\mathbb{E}[\|M^{k,\delta}\|_{s\times s}^4|\mathcal{F}^{k,\delta}]$. To this end, we use (5.14) (which, we recall, is a consequence of (5.10)) and (5.11):

$$\mathbb{E}\left[\|M^{k,\delta}\|_{s\times s}^4|\mathcal{F}^{k,\delta}\right] = \frac{1}{\delta^2}\mathbb{E}\left[\|x^{k+1,\delta} - x^{k,\delta} - \delta G^\delta(x^{k,\delta})\|_{s\times s}^4|\mathcal{F}^{k,\delta}\right]$$

$$\lesssim \frac{1}{\delta^2}\mathbb{E}\left[\|x^{k+1,\delta} - x^{k,\delta}\|_{s\times s}^4|\mathcal{F}^{k,\delta}\right] + \delta^2\mathbb{E}\|G^\delta(x^{k,\delta})\|_s^4$$

$$\lesssim \delta^{4r-2}\mathbb{E}\left(1 + \|x^{k,\delta}\|_{s\times s}^{4n}\right) + \delta^2\mathbb{E}\left(1 + \|x^{k,\delta}\|_{s\times s}^{4p}\right).$$

Therefore

$$\delta \sum_{k\delta < t} \mathbb{E}\left[\|M^{k,\delta}\|_{s\times s}^2 \mathbf{1}_{\{\|M^{k,\delta}\|_s^2 \geq \delta^{-1}\zeta\}}|\mathcal{F}^{k,\delta}\right]$$

$$\lesssim \delta^2 \sum_{k\delta < t} \mathbb{E}\left[\|M^{k,\delta}\|_{s\times s}^4|\mathcal{F}^{k,\delta}\right]$$

$$\lesssim \delta^{4r} \sum_{k\delta < t} \mathbb{E}\left(1 + \|x^{k,\delta}\|_{s\times s}^{4n}\right) + \delta^4 \sum_{k\delta < t} \mathbb{E}\left(1 + \|x^{k,\delta}\|_{s\times s}^{4p}\right)$$

$$\lesssim \delta^{4r-1} + \delta^{4r-\eta_3}\left(\delta^{\eta_3} \sum_{k\delta < t} \mathbb{E}\|x^{k,\delta}\|_{s\times s}^{4n}\right) + \delta^3 + \delta^{4-\eta_4}\left(\delta^{\eta_4} \sum_{k\delta < t} \mathbb{E}\|x^{k,\delta}\|_{s\times s}^{4p}\right) \to 0. \qquad \square$$

**Proof of Lemma 5.17.** We will show the following two bounds:

$$\left|\langle\hat{\varphi}_j^\ell, D^\delta(x)\hat{\varphi}_i^{\bar{\ell}}\rangle_{s\times s} - \langle\hat{\varphi}_j^\ell, \mathfrak{C}_s\hat{\varphi}_i^{\bar{\ell}}\rangle_{s\times s}\right| \lesssim \delta^{1/6}\left(1 + \|x\|_{s\times s}^{10}\right) \tag{B.12}$$

$$\forall i, j \in \mathbb{N} \text{ and } \ell, \bar{\ell} \in 1, 2;$$

$$\left|\text{Trace}_{\mathcal{H}^s \times \mathcal{H}^s}\left(D^\delta(x)\right) - \text{Trace}_{\mathcal{H}^s \times \mathcal{H}^s}\left(\mathfrak{C}_s\right)\right| \lesssim \delta^{1/6}\left(1 + \|x\|_{s\times s}^{10}\right). \tag{B.13}$$

Denote $\gamma_2 := \|\Gamma_2\|_{\mathcal{L}(\mathcal{H}^s)}$, $M_1^{k,\delta} := \mathcal{P}_q(M^{k,\delta})$, $M_2^{k,\delta} := \mathcal{P}_v(M^{k,\delta})$ and recall from (5.6) that $M^\delta(x) = [M^{k,\delta}|x^{k,\delta} = x]$. Then, from (5.5),

$$M_1^\delta(x) = \frac{1}{\sqrt{2\delta}}\big[q^{1,\delta} - q - \mathbb{E}[q^{1,\delta} - q]\big], \tag{B.14}$$

$$M_2^\delta(x) = \frac{1}{\sqrt{2\delta\Gamma_2}}\big[v^{1,\delta} - v - \mathbb{E}[v^{1,\delta} - v]\big]. \tag{B.15}$$

In order to obtain (B.12) and (B.13), we start with studying $M_1^\delta(x)$ and $M_2^\delta(x)$. More precisely, we proceed as follows:

- We first show the bound

$$\mathbb{E}\big[\big\|M_1^{k,\delta}\big\|_s^2|x^{k,\delta} = x\big] = \mathbb{E}\big\|M_1^\delta(x)\big\|_s^2 \lesssim \delta\big(1 + \|x\|_{s\times s}^2\big) \tag{B.16}$$

  and the decomposition

$$M_2^\delta(x) = \frac{1}{\sqrt{2\delta\Gamma_2}}\big[R^\delta(x) + \xi^\delta\big], \tag{B.17}$$

  where $R^\delta(x)$, defined in (B.21), is such that

$$\frac{1}{\delta}\mathbb{E}\big\|R^\delta(x)\big\|_s^2 \lesssim \delta^{1/3}\big(1 + \|x\|_{s\times s}^{10}\big). \tag{B.18}$$

- We then prove that (B.12) and (B.13) are a consequence of (B.16) and (B.17)–(B.18), together with

$$\left|\frac{1}{2\delta\gamma_2}\mathbb{E}\big\|\xi^\delta\big\|_s^2 - \mathrm{Trace}_{\mathcal{H}^s}\,\mathcal{C}_s\right| \lesssim \delta^2,$$

  which is easily seen to hold true. Indeed by definition

$$\left|\frac{1}{2\delta\gamma_2}\mathbb{E}\big\|\xi^\delta\big\|_s^2 - \mathrm{Trace}_{\mathcal{H}^s}\,\mathcal{C}_s\right| = \left|\mathrm{Trace}_{\mathcal{H}^s}\left[\frac{\mathcal{C}_s - \mathcal{C}_s e^{-2\delta\Gamma_2}}{2\delta\gamma_2} - \mathcal{C}_s\right]\right|$$
$$\leq \mathrm{Trace}_{\mathcal{H}^s}(\mathcal{C}_s)\left|\frac{1 - e^{-2\delta\gamma_2}}{2\delta\gamma_2} - 1\right| \lesssim \delta^2. \tag{B.19}$$

(B.16) is a straightforward consequence of (B.14), using (5.28), (5.29) and (5.26):

$$\mathbb{E}\big[\big\|M_1^{k,\delta}\big\|_s^2|x^{k,\delta} = x\big] \lesssim \frac{1}{\delta}\mathbb{E}\|q_* - q\|_s^2 \lesssim \delta\big(1 + \|x\|_s^2\big).$$

Recalling that $\gamma^\delta \sim \mathrm{Bernoulli}(\alpha^\delta)$ (with $\alpha^\delta$ defined by equation (5.15)), to decompose $M_2^\delta(x)$ we start from (B.15) and use (5.28):

$$M_2^\delta(x) = \frac{1}{\sqrt{2\delta\Gamma_2}}\big[\gamma^\delta\big(v_* + v'\big) - \mathbb{E}\big(\gamma^\delta\big(v_* + v'\big)\big) - v' - v + \mathbb{E}\big(v' + v\big)\big].$$

By (5.20) and (5.30),

$$-v' - v + \mathbb{E}(v' + v) = -v' + \mathbb{E}v' = -\xi^\delta; \qquad (B.20)$$

so (5.21) yields:

$$\sqrt{2\delta\Gamma_2}M_2^\delta(x) = (\gamma^\delta - \mathbb{E}(\gamma^\delta))\left[-q\sin\delta - \frac{\delta}{2}\cos\delta\mathcal{C}D\Psi(q)\right] - \xi^\delta$$
$$+ \gamma^\delta\left[(\cos\delta + 1)v' - \frac{\delta}{2}\mathcal{C}D\Psi(q_*)\right] - \mathbb{E}\left[\gamma^\delta\left((\cos\delta + 1)v' - \frac{\delta}{2}\mathcal{C}D\Psi(q_*)\right)\right].$$

Let $f(x) = f^\delta(x) + \bar{f}(x)$, with

$$f^\delta(x) := -\frac{\delta}{2}\mathcal{C}D\Psi(q_*) \quad \text{and} \quad \bar{f}(x) := (\cos\delta + 1)v'.$$

Then

$$\sqrt{2\delta\Gamma_2}M_2^\delta(x) = (\gamma^\delta - \mathbb{E}(\gamma^\delta))\left(-q\sin\delta - \frac{\delta}{2}\cos\delta\mathcal{C}D\Psi(q)\right)$$
$$+ \gamma^\delta f - \mathbb{E}(\gamma^\delta f) - \xi^\delta$$
$$= (\gamma^\delta - \mathbb{E}(\gamma^\delta))\left(-q\sin\delta - \frac{\delta}{2}\cos\delta\mathcal{C}D\Psi(q)\right)$$
$$+ \gamma^\delta f - \mathbb{E}(\gamma^\delta)f + [\mathbb{E}(\gamma^\delta) - 1]f - \mathbb{E}[(\gamma^\delta - 1)f] + f - \mathbb{E}f - \xi^\delta.$$

However $\bar{f} - \mathbb{E}\bar{f} = (\cos\delta + 1)(v' - \mathbb{E}v')$, so using (B.20)

$$f - \mathbb{E}f - \xi^\delta = \bar{f} - \mathbb{E}\bar{f} + f^\delta - \mathbb{E}f^\delta - \xi^\delta = (\cos\delta)\xi^\delta + f^\delta - \mathbb{E}f^\delta;$$

therefore, setting

$$R_1^\delta(x) := (\mathbb{E}(\gamma^\delta) - \gamma^\delta)\left(q\sin\delta + \frac{\delta}{2}\cos\delta\mathcal{C}D\Psi(q)\right),$$
$$R_2^\delta(x) := (\gamma^\delta - \mathbb{E}\gamma^\delta)f = (\gamma^\delta - \mathbb{E}\gamma^\delta)\left[(\cos\delta + 1)v' - \frac{\delta}{2}\mathcal{C}D\Psi(q_*)\right],$$
$$R_3^\delta(x) := -\mathbb{E}[(\gamma^\delta - 1)f] + [(\mathbb{E}\gamma^\delta) - 1]f, \qquad (B.21)$$
$$R_4^\delta(x) := f^\delta - \mathbb{E}(f^\delta) + (\cos\delta - 1)\xi^\delta \quad \text{and}$$
$$R^\delta(x) := \sum_{i=1}^4 R_i^\delta,$$

we obtain (B.17). From now on, to streamline the notation, we will not keep track of the $x$-dependence in $R_i^\delta(x)$ and $R^\delta(x)$. In other words, we will simply denote $R_i^\delta := R_i^\delta(x)$ and $R^\delta := R^\delta(x)$.

To prove (B.18), we bound $\mathbb{E}\|R_i^\delta\|_s^2$, $i = 1, \ldots, 4$. Observe first that

$$\mathbb{E}\big(\gamma^\delta - \mathbb{E}\gamma^\delta\big)^2 = \mathbb{E}\big(\alpha^\delta\big)\big(1 - \mathbb{E}\alpha^\delta\big) \lesssim \delta^2\big(1 + \|x\|_{s\times s}^4\big),$$

which is a consequence of (5.16) and (5.29). Therefore, by (5.17),

$$\mathbb{E}\big\|R_1^\delta\big\|_s^2 \lesssim \mathbb{E}\big(\gamma^\delta - \mathbb{E}\gamma^\delta\big)^2\big\|\delta q + \delta\mathcal{C}D\Psi(q)\big\|_s^2 \lesssim \delta^3\big(1 + \|x\|_{s\times s}^6\big). \tag{B.22}$$

Now notice that the Bochner's inequality together with Jensen's inequality give

$$\|\mathbb{E}g\|_s^2 \leq \mathbb{E}\|g\|_s^2, \qquad \text{for every } \mathcal{H}^s\text{-valued, integrable } g.$$

To bound $R_2^\delta$ we split it into two terms, namely

$$R_2^\delta := \big(\gamma^\delta - \mathbb{E}\gamma^\delta\big)(\cos\delta + 1)v' - \big(\gamma^\delta - \mathbb{E}\gamma^\delta\big)\frac{\delta}{2}\mathcal{C}D\psi(q_*) =: R_{21}^\delta + R_{22}^\delta.$$

To estimate $R_{22}^\delta$ we use (5.29), (5.17) and (5.26):

$$\mathbb{E}\big\|R_{22}^\delta\big\|_s^2 \lesssim \delta^2\big[\mathbb{E}\big\|\mathcal{C}D\Psi(q_*) - \mathcal{C}D\Psi(q)\big\|_s^2 + \big\|\mathcal{C}D\Psi(q)\big\|_s^2\big] \lesssim \delta^2\big(1 + \|q\|_s^2\big).$$

To study $R_{21}^\delta$ instead, we write $\gamma^\delta - \mathbb{E}\gamma^\delta = \gamma^\delta - 1 + \mathbb{E}(1 - \gamma^\delta)$ and we repeatedly use (5.38), obtaining:

$$\begin{aligned}
\mathbb{E}\big\|R_{21}^\delta\big\|_s^2 &\lesssim \mathbb{E}\big[\big(\gamma^\delta - \mathbb{E}\gamma^\delta\big)^2\|v'\|_s^2\big] \\
&\lesssim \mathbb{E}\big[\big(\gamma^\delta - 1\big)^2\|v'\|_s^2\big] + \mathbb{E}\big(1 - \gamma^\delta\big)^2\mathbb{E}\|v'\|_s^2 \\
&\overset{(5.25)}{\lesssim} \big(\mathbb{E}\big|\gamma^\delta - 1\big|^3\big)^{2/3}\big(\mathbb{E}\|v'\|_s^6\big)^{1/3} + \mathbb{E}\big|1 - \alpha^\delta\big|\big(1 + \|v\|_s^2\big) \\
&\overset{(5.16)}{\lesssim} \delta^{4/3}\big(1 + \|x\|_{s\times s}^8\big) + \delta^2\big(1 + \|x\|_{s\times s}^6\big) \lesssim \delta^{4/3}\big(1 + \|x\|_{s\times s}^8\big).
\end{aligned}$$

Combining the estimates of $R_{21}^\delta$ and $R_{22}^\delta$ we get

$$\mathbb{E}\big\|R_2^\delta\big\|_s^2 \lesssim \delta^{4/3}\big(1 + \|x\|_{s\times s}^8\big). \tag{B.23}$$

As for $R_3^\delta$, using $\mathbb{E}\|f\|_s^2 \lesssim 1 + \|x\|_{s\times s}^2$ (which is a consequence of (5.17) and (5.25)),

$$\begin{aligned}
\mathbb{E}\big\|R_3^\delta\big\|_s^2 &\leq \mathbb{E}\big\|\big(\mathbb{E}(\gamma^\delta) - 1\big)f\big\|_s^2 + \big\|\mathbb{E}\big[(\gamma^\delta - 1)f\big]\big\|_s^2 \\
&\leq \big(\mathbb{E}(\alpha^\delta) - 1\big)^2\mathbb{E}\|f\|_s^2 + \mathbb{E}\big\|(\gamma^\delta - 1)f\big\|_s^2 \\
&\overset{(5.16)}{\lesssim} \delta^4\big(1 + \|x\|_{s\times s}^8\big)\big(1 + \|x\|_{s\times s}^2\big) + \big(\mathbb{E}\big|\gamma^\delta - 1\big|^3\big)^{2/3}\big(\mathbb{E}\|f\|_s^6\big)^{1/3} \\
&\overset{(5.38)}{\lesssim} \delta^4\big(1 + \|x\|_{s\times s}^{10}\big) + \delta^{4/3}\big(1 + \|x\|_{s\times s}^4\big)^{2/3}\big(1 + \|x\|_{s\times s}^2\big) \leq \delta^{4/3}\big(1 + \|x\|_{s\times s}^{10}\big).
\end{aligned} \tag{B.24}$$

Now the last term: from (5.17), $\mathbb{E}\|f^\delta\|_s^2 \lesssim \delta^2(1 + \|q\|_s^2)$; therefore

$$\mathbb{E}\|R_4^\delta\|_s^2 \lesssim \mathbb{E}\|f^\delta\|_s^2 + \delta^4 \mathbb{E}\|\xi^\delta\|_s^2 \lesssim \delta^2(1 + \|q\|_s^2). \tag{B.25}$$

It is now clear that (B.18) follows from (B.22), (B.23), (B.24) and (B.25).

Let us now show that (B.12) and (B.13) follow from (B.16) and (B.17)–(B.18). We start with (B.13). By definition,

$$\begin{aligned}
\text{Trace}_{\mathcal{H}^s \times \mathcal{H}^s}\left(D^\delta(x)\right) &= \mathbb{E}\left[\|M^{k,\delta}\|_s^2 | x^{k,\delta} = x\right] \\
&= \mathbb{E}\left[\|M_1^{k,\delta}\|_s^2 | x^{k,\delta} = x\right] + \mathbb{E}\left[\|M_2^{k,\delta}\|_s^2 | x^{k,\delta} = x\right]
\end{aligned} \tag{B.26}$$

and

$$\text{Trace}_{\mathcal{H}^s \times \mathcal{H}^s}(\mathfrak{C}_s) = \text{Trace}_{\mathcal{H}^s}(\mathcal{C}_s).$$

Also,

$$\begin{aligned}
\left|\mathbb{E}\left[\|M_2^{k,\delta}\|_s^2 | x^{k,\delta} = x\right] - \text{Trace}_{\mathcal{H}^s}(\mathcal{C}_s)\right| &\overset{\text{(B.17)}}{=} \left|\frac{1}{2\delta\gamma_2}\mathbb{E}\|R^\delta + \xi^\delta\|_s^2 - \text{Trace}_{\mathcal{H}^s}(\mathcal{C}_s)\right| \\
&\leq \frac{1}{2\delta\gamma_2}\mathbb{E}\|R^\delta\|_s^2 + \left|\frac{1}{2\delta\gamma_2}\mathbb{E}\|\xi^\delta\|_s^2 - \text{Trace}_{\mathcal{H}^s}(\mathcal{C}_s)\right| \\
&\quad + \frac{1}{\delta\gamma_2}\mathbb{E}\langle R^\delta, \xi^\delta\rangle_s \\
&\lesssim \frac{1}{2\delta\gamma_2}\mathbb{E}\|R^\delta\|_s^2 \\
&\quad + \left|\frac{1}{2\delta\gamma_2}\mathbb{E}\|\xi^\delta\|_s^2 - \text{Trace}_{\mathcal{H}^s}(\mathcal{C}_s)\right| \\
&\quad + \frac{1}{\delta}\left(\mathbb{E}\|R^\delta\|_s^2\right)^{1/2}\left(\|\xi^\delta\|_s^2\right)^{1/2} \\
&\overset{\text{(B.19)}}{\lesssim} \delta^{1/3}\left(1 + \|x\|_{s\times s}^{10}\right) + \delta^2 + \delta^{1/6}\left(1 + \|x\|_{s\times s}^5\right) \\
&\lesssim \delta^{1/6}\left(1 + \|x\|_{s\times s}^{10}\right).
\end{aligned} \tag{B.27}$$

(B.26) and the above (B.27) imply (B.13). (B.12) can be obtained similarly. Due to the symmetry of $D^\delta(x)$, all we need to show is that if at least one index between $\ell$ and $\bar{\ell}$ is equal to 1 then

$$\langle \hat{\varphi}_i^\ell, D^\delta(x)\hat{\varphi}_j^{\bar{\ell}}\rangle_{s\times s} \lesssim \delta^{1/2}\left(1 + \|x\|_{s\times s}^5\right). \tag{B.28}$$

If instead $\ell = \bar{\ell} = 2$ we will prove that

$$\left|\langle \varphi_i, D_{22}^\delta(x)\varphi_j\rangle_s - \langle \varphi_i, \mathcal{C}_s\varphi_j\rangle_s\right| \leq \delta^{1/6}\left(1 + \|x\|_{s\times s}^{10}\right), \tag{B.29}$$

where $D_{22}^{\delta}(x) = \mathbb{E}[M_2^{k,\delta} \otimes M_2^{k,\delta} | x^{k,\delta} = x]$. (B.28) and (B.29) imply (B.12). To prove the bound (B.28) observe first that

$$\mathbb{E}\|M_2^{\delta}(x)\|_s^2 \lesssim 1 + \|x\|_{s \times s}^{10},$$

which follows from (B.17), (B.18) and (5.24). To show (B.28) suppose, without loss of generality, that $\ell = 1, \bar{\ell} = 2$. Then

$$\left|\langle \hat{\varphi}_i^1, D^{\delta}(x)\hat{\varphi}_j^2 \rangle_{s \times s}\right| \leq \mathbb{E}\left|\langle M^{\delta}(x), \hat{\varphi}_i^1 \rangle_{s \times s}\langle M^{\delta}(x), \hat{\varphi}_i^2 \rangle_{s \times s}\right|$$

$$\leq \left(\mathbb{E}\|M_1^{\delta}(x)\|_s^2\right)^{1/2}\left(\mathbb{E}\|M_2^{\delta}(x)\|_s^2\right)^{1/2} \overset{(B.16)}{\lesssim} \delta^{1/2}\left(1 + \|x\|_{s \times s}^5\right).$$

As for (B.29), let $\xi := \xi^0$, i.e., let $\xi$ be a mean zero Gaussian random variable with covariance operator $\mathcal{C}_s$ in $\mathcal{H}^s$. Then

$$\left|\langle \hat{\varphi}_i^2, D^{\delta}(x)\hat{\varphi}_j^2 \rangle_{s \times s} - \langle \hat{\varphi}_i^2, \mathcal{C}_s\hat{\varphi}_j^2 \rangle_{s \times s}\right|$$

$$= \left|\langle \varphi_i, D_{22}^{\delta}(x)\varphi_j \rangle_s - \langle \varphi_i, \mathcal{C}_s\varphi_j \rangle_s\right|$$

$$= \left|\mathbb{E}\left(\langle M_2^{\delta}(x), \varphi_i \rangle_s\langle M_2^{\delta}(x), \varphi_j \rangle_s\right) - \mathbb{E}\left(\langle \xi, \varphi_i \rangle_s\langle \xi, \varphi_j \rangle_s\right)\right|$$

$$\overset{(B.17)}{=} \left|\mathbb{E}\left\langle \frac{R^{\delta} + \xi^{\delta}}{\sqrt{2\delta\Gamma_2}}, \varphi_i \right\rangle_s\left\langle \frac{R^{\delta} + \xi^{\delta}}{\sqrt{2\delta\Gamma_2}}, \varphi_j \right\rangle_s - \mathbb{E}\left(\langle \xi, \varphi_i \rangle_s\langle \xi, \varphi_j \rangle_s\right)\right|$$

$$\lesssim \frac{1}{\delta}\mathbb{E}\|R^{\delta}\|_s^2 + \frac{1}{\delta}\left(\mathbb{E}\|R^{\delta}\|_s^2\right)^{1/2}\left(\mathbb{E}\|\xi^{\delta}\|_s^2\right)^{1/2} + \left|\frac{1}{2\delta\gamma_2}\mathbb{E}\left(\langle \xi^{\delta}, \varphi_i \rangle_s\right)^2 - \mathbb{E}\left(\langle \xi, \varphi_i \rangle_s\right)\right|;$$

so, by using again (B.18) and (5.24) and with a reasoning analogous to that contained in (B.19) and (B.27), we obtain (B.29). $\qquad\square$

# Appendix C

We gather here some basic facts about Hamiltonian mechanics. For a more thorough discussion, the reader may consult [15,20].

Let us start from the Hamiltonian formalism in a finite dimensional setting. To a given real valued and smooth function $\mathsf{H}(q, p): \mathbb{R}^{2N} \to \mathbb{R}$, we can associate in a canonical way a system of differential equations, the so called *canonical Hamiltonian system associated to* $\mathsf{H}$, namely

$$\frac{dq}{dt} = D_p\mathsf{H}(q, p),$$

$$\frac{dp}{dt} = -D_q\mathsf{H}(q, p).$$

Using the symplectic matrix

$$J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix},$$

and denoting $z = (q, p) \in \mathbb{R}^{2N}$, the canonical Hamiltonian system can be rewritten as

$$\frac{\mathrm{d}z}{\mathrm{d}t} = J D_z \mathsf{H}(q, p). \tag{C.1}$$

The two properties of the Hamiltonian flow (C.1) that are relevant to our purposes are: (i) smooth functions of the Hamiltonian $\mathsf{H}$ remain constant along the solutions of (C.1); (ii) the flow preserves the volume element $\mathrm{d}z$. As a consequence, the Hamiltonian dynamics preserves any measure with density $\mathrm{e}^{-\mathsf{H}(z)}$ with respect to Lebesgue measure. Clearly, an analogous discussion holds for any systems obtained by making the noncanonical choice

$$\hat{J} = \begin{pmatrix} 0 & L \\ -L & 0 \end{pmatrix}, \qquad L \text{ any symmetric matrix,}$$

with corresponding dynamics

$$\frac{\mathrm{d}z}{\mathrm{d}t} = \hat{J} D_z \mathsf{H}(q, p).$$

This reasoning can be repeated in our infinite dimensional context (however in this case one cannot talk about conservation of volume element $\mathrm{d}z$). The Hamiltonian part of the equations considered in Section 2.2 is built precisely in this spirit. The change of variable which allows us to swap from momentum to velocity variable corresponds to going from the canonical to the noncanonical choice. In particular, once we fix $\mathcal{M} = \mathcal{L} = \mathcal{C}^{-1}$, our noncanonical symplectic matrix is

$$\hat{J} = \begin{pmatrix} 0 & \mathcal{C} \\ -\mathcal{C} & 0 \end{pmatrix}.$$

For more comments about the particular form of the Hamiltonian function in our infinite dimensional setting, see Remark 5.6 and [2].

## Acknowledgements

## References

[1] Berger, E. (1986). Asymptotic behaviour of a class of stochastic approximation procedures. *Probab. Theory Relat. Fields* **71** 517–552. MR0833268

[2] Beskos, A., Pinski, F.J., Sanz-Serna, J.M. and Stuart, A.M. (2011). Hybrid Monte Carlo on Hilbert spaces. *Stochastic Process. Appl.* **121** 2201–2230. MR2822774

[3] Beskos, A., Roberts, G., Stuart, A. and Voss, J. (2008). MCMC methods for diffusion bridges. *Stoch. Dyn.* **8** 319–350. MR2444507

[4] Bou-Rabee, N. and Vanden-Eijnden, E. (2010). Pathwise accuracy and ergodicity of metropolized integrators for SDEs. *Comm. Pure Appl. Math.* **63** 655–696. MR2583309

[5] Bou-Rabee, N. and Vanden-Eijnden, E. (2012). A patch that imparts unconditional stability to explicit integrators for Langevin-like equations. *J. Comput. Phys*. **231** 2565–2580. MR2881033

[6] Cotter, S.L., Roberts, G.O., Stuart, A.M. and White, D. (2013). MCMC methods for functions: Modifying old algorithms to make them faster. *Statist. Sci*. **28** 424–446. MR3135540

[7] Diaconis, P., Holmes, S. and Neal, R.M. (2000). Analysis of a nonreversible Markov chain sampler. *Ann. Appl. Probab.* **10** 726–752. MR1789978

[8] Duane, S., Kennedy, A.D., Pendleton, B.J. and Roweth, D. (1987). Hybrid Monte Carlo. *Physics Letters B* **195** 216–222.

[9] Hairer, M., Stuart, A. and Voss, J. (2011). Signal processing problems on function space: Bayesian formulation, stochastic PDEs and effective MCMC methods. In *The Oxford Handbook of Nonlinear Filtering* (D. Crisan and B. Rozovsky, eds.). 833–873. Oxford Univ. Press, Oxford. MR2884617

[10] Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57** 97–109.

[11] Horowitz, A.M. (1991). A generalized guided Monte Carlo algorithm. *Physics Letters B* **268** 247–252.

[12] Hwang, C.-R., Hwang-Ma, S.-Y. and Sheu, S.J. (1993). Accelerating Gaussian diffusions. *Ann. Appl. Probab.* **3** 897–913. MR1233633

[13] Hwang, C.-R., Hwang-Ma, S.-Y. and Sheu, S.-J. (2005). Accelerating diffusions. *Ann. Appl. Probab.* **15** 1433–1444. MR2134109

[14] Mattingly, J.C., Pillai, N.S. and Stuart, A.M. (2012). Diffusion limits of the random walk Metropolis algorithm in high dimensions. *Ann. Appl. Probab.* **22** 881–930. MR2977981

[15] Neal, R.M. (2011). MCMC using Hamiltonian dynamics. In *Handbook of Markov Chain Monte Carlo*. *Chapman & Hall/CRC Handb. Mod. Stat. Methods* 113–162. CRC Press, Boca Raton, FL. MR2858447

[16] Otto, F., Weber, H. and Westdickenberg, M. (2013). Invariant measure of the stochastic Allen–Cahn equation: The regime of small noise and large system size. Available at arXiv:1301.0408.

[17] Pillai, N.S., Stuart, A.M. and Thiery, A.H. (2014). Noisy gradient flow from a random walk in Hilbert space. *Stoch. Partial Difer. Equ. Anal. Comput.* **2** 196–232. MR3249584

[18] Pillai, N.S., Stuart, A.M. and Thiéry, A.H. (2012). Optimal scaling and diffusion limits for the Langevin algorithm in high dimensions. *Ann. Appl. Probab.* **22** 2320–2356. MR3024970

[19] Prévôt, C. and Röckner, M. (2007). *A Concise Course on Stochastic Partial Differential Equations*. *Lecture Notes in Math.* **1905**. Berlin: Springer. MR2329435

[20] Sanz-Serna, J.M. and Calvo, M.P. (1994). *Numerical Hamiltonian Problems*. *Applied Mathematics and Mathematical Computation* **7**. London: Chapman & Hall. MR1270017

[21] Stuart, A.M. (2010). Inverse problems: A Bayesian perspective. *Acta Numer.* **19** 451–559. MR2652785