Advanced MCMC Methods for Sampling on Diffusion Pathspace

Alexandros Beskos

Department of Statistical Science, University College London, UK

Kostas Kalogeropoulos

Department of Statistics, London School of Economics

Erik Pazos

Department of Statistical Science, University College London, UK

Abstract

The need to calibrate increasingly complex statistical models requires a persistent effort for further advances on available, computationally intensive Monte-Carlo methods. We study here an advanced version of familiar Markov-chain Monte-Carlo (MCMC) algorithms that sample from target distributions defined as change of measures from Gaussian laws on general Hilbert spaces. Such a model structure arises in several contexts: we focus here at the important class of statistical models driven by diffusion paths whence the Wiener process constitutes the reference Gaussian law. Particular emphasis is given on advanced Hybrid Monte-Carlo (HMC) which makes large, derivative-driven steps in the state space (in contrast with local-move Random-walk-type algorithms) with analytical and experimental results. We illustrate its computational advantages in various diffusion processes and observation regimes; examples include stochastic volatility and latent survival models. In contrast with their standard MCMC counterparts, the advanced versions have *mesh-free* mixing times, as these will not deteriorate upon refinement of the approximation of the inherently infinite-dimensional diffusion paths by finite-dimensional ones used in practice when applying the algorithms on a computer.

Keywords: Gaussian measure, diffusion process, covariance operator, Hamiltonian dynamics, mixing time, stochastic volatility.

Email addresses: `a.beskos@ucl.ac.uk` (Alexandros Beskos), `K.Kalogeropoulos@lse.ac.uk` (Kostas Kalogeropoulos), `erikpazos05@gmail.com` (Erik Pazos)

1. Introduction

Markov chain Monte-Carlo (MCMC) methods provide an intuitive, powerful mechanism for sampling from complex posterior distributions arising in applications (see [21] for a review of algorithms and applications). The rapidly increasing complexity of statistical models employed by practitioners requires a parallel effort at advancing MCMC methodology to deliver algorithms that can provide fast model exploration. Ideally, suggested algorithms should be of enough flexibility to cover a wide range of model structures. This paper will move along such directions by proposing and studying advanced versions of standard MCMC algorithms to improve algorithmic performance on complex, high-dimensional models.

The advanced algorithms are relevant for target distributions defined as change of measures from Gaussian laws. Within such a structure, we will be focusing here upon the important class of statistical models driven by Stochastic Differential Equations (SDEs) posing concrete computational challenges; here, Brownian motion constitutes the reference Gaussian measure. The paper will develop and test advanced MCMC algorithms for the computationally demanding task of reproducing sample paths of SDEs under various direct or indirect observation regimes. The ability to sample the realized dynamics driving the data mechanism is of high importance for understanding the model behavior. Also, due to the typical intractability of the likelihood function in SDE contexts, the underlying diffusion path is many times treated as a latent variable within Gibbs samplers and its fast sampling is critical for the efficiency and feasibility of parametric inference procedures.

The advanced MCMC methods follow closely recent developments [9, 6] over algorithms that take advantage of the structure of target distributions Π being determined as a change of measure from a Gaussian one $\Pi_0 \equiv N(0, \mathcal{C})$, that is:

$$\frac{d\Pi}{d\Pi_0}(x) = \exp\{-\Phi(x)\} \ , \tag{1}$$

for some function Φ defined on a Hilbert space $\mathcal{H}$. The method exploits the relation with the Gaussian measure to evolve standard MCMC algorithms into advanced ones with the critical computational advantage that their convergence properties are *mesh-free*, i.e. their mixing times do not deteriorate as the dimension of the state space increases when refining relevant finite-dimensional projections (used in practice on a computer) to better approximate inherently infinite-dimensional elements of $\mathcal{H}$. In the SDE context, finite-difference methods are commonly employed to approximate the infinite-dimensional diffusion sample paths whence a discretization mesh will be specified. We will be looking on advanced versions of Random-walk Metropolis (RWM), Metropolis-adjusted Langevin algorithm (MALA) and Hybrid Monte-Carlo (HMC). MALA and HMC both use information about the derivative of the log-target to drive the algorithmic dynamics, whereas RWM uses blind proposals. Emphasis will be given on HMC, employing Hamiltonian dynamics, as its nature to perform global

designated steps on the SDE pathspace seems to provide significant computational advantages compared to the other two local-move algorithms; this will be illustrated both analytically and experimentally.

A first methodological contribution of the paper will be to define HMC on a Hilbert space, similar to Hilbert-space valued RWM and MALA derived in [9]. A similar attempt for HMC has taken place in [6]; the method there required strict, *analytical* Lipschitz-continuity assumptions on the Φ-derivative $\delta\Phi$ [1]. In this paper, a much simpler derivation will be proven under minimal *probabilistic* assumptions on Φ relevant for a much wider range of practical applications that, for instance, could involve inherently stochastic terms (e.g. stochastic integrals) at the specification of Φ. The concrete consequence of the mathematical verification of a well-defined algorithm on the infinite-dimensional Hilbert space $\mathcal{H}$ is an anticipated mesh-free mixing time for the practical algorithm that will run on some N-dimensional projection (typically, finite-difference) of the pathspace. Another methodological contribution of the paper will be that it will analytically illustrate, in the context of directly observed diffusions, that 'clever' use of information on the derivative of the log-target (within HMC, compared to MALA) can remove orders of complexity from the computational costs of the MCMC algorithm. Albeit proven in a linear scenario, the result will be later on empirically manifested in applications on realistic models. In addition to verifying the theoretical results, the empirical contributions of the paper relate with the implementation of the pathspace algorithms on several practical problems including stochastic volatility and various diffusion driven models. As demonstrated in relevant applications, HMC algorithms avoid the random-walk-type behaviour of the other algorithms and greatly outperform them, thus constituting a powerful method for tackling high-dimensional path-sampling problems.

1.1. Diffusion-Driven Models

Stochastic Differential Equations (SDEs) provide a powerful and flexible probabilistic structure for modeling phenomena in a multitude of disciplines: finance, biology, molecular dynamics, chemistry, survival analysis, epidemiology, just to name a few (see [30] for a review of applications). A large class of SDEs can be specified as follows:

$$dV_u = \mu(V_u;\theta)dt + \sigma(V_u;\theta)dB_u \ , \tag{2}$$

for drift and diffusion coefficient $\mu(\cdot;\theta) : \mathbb{R}^d \mapsto \mathbb{R}^d$ and $\sigma(\cdot;\theta) : \mathbb{R}^d \times \mathbb{R}^d \mapsto \mathbb{R}^d \times \mathbb{R}^d$ respectively (of known functional form up to some unknown parameter $\theta \in \mathbb{R}^p$), and Brownian motion B_u. Such a differential modeling structure is sometimes implied by physical laws (e.g. Langevin equation in physics) or is

[1] the δ-notation refers in the most general setting to the Fréchet generalisation of differentiation of real-valued functions defined on general Hilbert spaces; in our case we will work with the associated (to the differential operator) element of the dual space, so $\delta\Phi$ will be treated as an element of the pathspace.

selected due to its adaptiveness by the practitioner to represent time-evolving quantities.

The diffusion process defined by (2) may be observed in a number of ways. The case where data form a discrete skeleton of the diffusion path has been studied extensively (see for example [2, 7, 17, 10]), but different types of observation regimes often appear in applications. For example in PK/PD studies [34], growth curve models [14] and bioinformatics [24], the stochastic dynamical systems are observed with error. In financial applications, and in particular stochastic volatility models [20], the observation model consists of partial data or integrals of the underlying diffusion process [12]. Event time models as in [1, 35], involve a latent diffusion process that is observed through random barrier hitting times. In this paper we demonstrate how the above framework can be unified through formulations that consist of a diffusion process defined by (2) and are completed by an appropriate (typically Lebesgue) density $p(y|v, \theta)$ for the observations Y conditionally on the diffusion path V and θ. We then proceed and develop a general and efficient HMC algorithm to handle models within this framework.

Inferential procedures about θ are perplexed by the typical unavailability of the likelihood function $p(y|\theta)$ in closed form. MCMC methods could in principle overcome such an issue via *data augmentation*, whereby Gibbs sampler steps within the algorithm would switch from sampling the unobserved diffusion path V conditionally on θ and Y to sampling θ conditionally on V and Y. Our advanced MCMC methods are relevant here for the challenging high-dimensional path-update of V.

The construction of such data augmentation MCMC samplers has to address two important issues arising from inherent characteristics of diffusion processes. The first issue is the singularity between the parameters in the diffusion coefficient and the latent diffusion paths (i.e. conditionally on V, the parameters in the diffusion coefficient have a Dirac distribution) caused by the quadratic variation process. The problem was identified in [36] noting that the convergence of an MCMC Independence sampler deteriorates as $\mathcal{O}(N)$ unless appropriate reparametrisation is applied (N being the number of discrete instances of the path considered in practice when executing the method on a personal computer). Various such reparametrizations are currently available and some of them will be used in the sequel. The second issue is the construction of efficient proposals for the update of the (high-dimensional) latent paths. The efficiency and tuning of such updates are critical for the overall performance of the algorithms. Most existing approaches adopt Independence sampler steps proposing paths from easier-to-handle diffusion processes (typically, Brownian motion paths) as candidates for paths of the target $V|Y, \theta$. While these approaches may work well in some cases, they can also perform unacceptably poorly in many other ones when target paths are very different from Brownian ones.

The advanced MCMC samplers studied in this paper connect with both above issues. First, methods used for decoupling the dependence between V and θ, achieve this via the consideration of some 1-1 transform $X = \eta(V; \theta)$ so that the distribution of X has a density w.r.t. a probability measure that

will not involve θ: this reference measure is precisely a Brownian motion (or a Brownian bridge). Thus, the decoupling approaches conveniently deliver target distributions which are change of measures from Gaussian measures of the structure (1). More analytically, using the reparametrization method of [24] our advanced samplers will be, in principle, applicable under the condition:

$$v \mapsto \sigma(v;\theta) \text{ is invertible, for any } v \text{ in the state space of } V \ . \tag{3}$$

For the second issue, the advanced HMC sampler provides a powerful mechanism for delivering efficient updates for the high-dimensional latent paths as: (i) it will propose non-blind path updates, driven by the log-derivative of the target distribution; (ii) proposed updates will make large steps in the path space thus offering the potential for efficient mixing; (iii) it's mixing time will be mesh-free, i.e. will not deteriorate with increasing N.

The paper is organised as follows: Section 2 contains some background material regarding Gaussian distributions on separable Hilbert Spaces. The advanced HMC sampler is developed in Section 3 where other MCMC samplers for diffusion pathspaces are also presented. In Section 4 we demonstrate the advantages offered by the HMC algorithm through an analytical study on the Ornstein-Uhlenbeck (OU) diffusion process. The theory of Section 3 is used in Section 5 to extend the framework the HMC algorithm to a rich class family of diffusion models with general diffusion coefficients and various observation regimes. Section 6 verifies the theoretical results and illustrates the developed algorithms in various simulation experiments. Further generalizations are provided in Section 7, while Section 8 concludes with some relevant discussion and future directions.

2. Gaussian Measures on Hilbert Spaces

We collect here some background material (see e.g. [13]) on Gaussian distributions on a separable Hilbert space $\mathcal{H}$ that will assist at the presentation of the later sections. The Cameron-Martin space, $\mathcal{H}_0$, of the Gaussian law $\Pi_0 \equiv N(0,\mathcal{C})$ coincides with the image space of $\mathcal{C}^{1/2}$. Essentially, $\mathcal{H}_0$ includes all elements of the Hilbert space which preserve the absolute continuity properties of Π_0 upon translation. This is made mathematically explicit via the following proposition.

Proposition 2.1. *If $T(v) = v + \mathcal{C}^{1/2}v_0$ for a constant $v_0 \in \mathcal{H}$ then Π_0 and $\Pi_0 \circ T^{-1}$ are absolutely continuous w.r.t. each other with density:*

$$\frac{d\left\{\,\Pi_0 \circ T^{-1}\,\right\}}{d\Pi_0}(v) = \exp\left\{\langle v_0, \mathcal{C}^{-1/2}v\rangle - \tfrac{1}{2}|v_0|^2\right\} \ .$$

Proof. This is Theorem 2.21 of [13]. □

For the diffusion pathspace we focus upon in this paper, the target distribution $\Pi(dx)$ is defined on the Hilbert space of squared integrable paths $\mathcal{H} =$

$L^2([0,\ell],\mathbb{R})$ (with appropriate boundary conditions) for some length $\ell > 0$. The centered Gaussian reference measure Π_0 will correspond to a Brownian motion (thus, boundary condition $x(0) = 0$) or a Brownian Bridge ($x(0) = x(\ell) = 0$). The covariance operator is connected with the covariance function $c(u, v)$ of the Gaussian process via:

$$(\mathcal{C}f)(u) = \int_0^\ell c(u,v)f(v)dv \ , \quad f \in \mathcal{H} \ .$$

With this in mind, the covariance operators $\mathcal{C}^{bm}$, $\mathcal{C}^{bb}$ of the Brownian motion and Brownian bridge respectively are as below:

$$(\mathcal{C}^{bm}f)(u) = \int_0^\ell (u \wedge v)\, f(v)dv = u\int_0^\ell f(v)dv - \int_0^u \int_0^s f(v)dv\, ds \ ; \tag{4}$$

$$\begin{aligned}(\mathcal{C}^{bb}f)(u) &= \int_0^\ell (u \wedge v - \tfrac{uv}{\ell})\, f(v)dv \\ &= \frac{u}{\ell}\int_0^\ell \int_0^s f(v)dv\, ds - \int_0^u \int_0^s f(v)dv\, ds \ .\end{aligned} \tag{5}$$

The Cameron-Martin spaces $\mathcal{H}_0^{bm}$ and $\mathcal{H}_0^{bb}$ of a Brownian motion and Brownian bridge respectively are analytically specified as follows (see e.g. Lemma 2.3.14 of [11] for the case of Brownian motion; Brownian bridge involves the extra boundary condition $x(\ell) = 0$):

$$\begin{aligned}\mathcal{H}_0^{bm} &= \big\{x : [0,\ell] \mapsto \mathbb{R} : \exists\ f \in L^2([0,\ell],\mathbb{R}) \text{ such that } x(u) = \int_{[0,u]} f(s)ds\big\} \ ; \\ \mathcal{H}_0^{bb} &= \big\{x : [0,\ell] \mapsto \mathbb{R} : \exists\ f \in L^2([0,\ell],\mathbb{R}) \text{ such that} \\ &\qquad\qquad x(u) = \int_{[0,u]} f(s)ds,\ x(\ell) = 0\big\} \ .\end{aligned}$$

The Karhunen-Loève representation of $N(0,\mathcal{C})$ will be used later on. Analytically, considering the eigen-decomposition $\{\lambda_p, \phi_p\}_{p=1}^\infty$ of $\mathcal{C}$ so that $\mathcal{C}\,\phi_p = \lambda_p\,\phi_p$, we have that the (normalized) eigenfunctions $\{\phi_p\}_{p=1}^\infty$ constitute an orthonormal basis for the Hilbert space $\mathcal{H}$. In particular, for $x \sim N(0,\mathcal{C})$ we have the expansion:

$$x = \sum_{p=1}^\infty \langle x, \phi_p\rangle\, \phi_p = \sum_{p=1}^\infty x_p\, \phi_p = \sum_{p=1}^\infty \sqrt{\lambda_p}\, \xi_p\, \phi_p \ , \tag{6}$$

where $\{\xi_p\}_{p=1}^\infty$ are iid variables from $N(0,1)$.

3. Advanced HMC on Hilbert Space

HMC uses the derivative of the log-target density, in the form of Hamiltonian dynamics, to generate moves on the state space. An appropriate accept/reject decision will then force reversibility w.r.t. the target distribution.

Uniquely amongst alternative MCMC algorithms, HMC allows for the synthesis of a number of derivative-guided steps before the accept/reject test, thus permitting large, designated steps in the state space.

The standard version of HMC first appeared in [15]. An advanced version for infinite-dimensional targets of the form (1) appeared in [6]. One of our objectives here is to extend the relevance of the method in [6] to a much wider class of SDE-driven applications. In particular, [6] defines an HMC algorithm on a Hilbert space under strong analytical Lipschitz-continuity assumptions on Sobolev norms of $\delta\Phi$ which are difficult to verify and are, indeed, not relevant for the common scenario when Φ will involve inherently stochastic terms (like stochastic integrals). Here, we provide an alternative derivation of advanced HMC which avoids such strong assumptions of earlier works, and is relevant for many practical applications. Such a construction of the algorithm on the Hilbert space will provide a justification for order $\mathcal{O}(1)$-mixing times for the related N-dimensional projected algorithm. This justification will not be analytically proven here as it would require a level of mathematical technicalities far beyond the application-motivated scope of this paper; we note that all numerical applications shown later will empirically verify the $\mathcal{O}(1)$-mixing times of the samplers.

We will only require the following condition (recall that $\mathcal{C}$ is the covariance operator of the Gaussian measure at the definition of the target Π in (1)):

Assumption 3.1. *$\mathcal{C}\,\delta\Phi(x)$ is an element of the Cameron-Martin space of the Gaussian measure Π_0 (so $\mathcal{C}\,\delta\Phi(x) \in \operatorname{Im}\mathcal{C}^{1/2}$) for all x in a set with probability 1 under Π_0.*

We will discuss Assumption 3.1 in the context of our applications in the sequel. Note that the final algorithm suggested later in this section coincides with the one presented in [6] when the strict analytical conditions of that paper on Φ are satisfied; the new theoretical development here is to construct an alternative proof for the well-definition of the algorithm on the Hilbert space that requires only Assumption 3.1 and, as a consequence, greatly extends the relevance of the developed method for applications.

Some of the calculations carried out in the development of the algorithm in the sequel make explicit sense for $\mathcal{H} = \mathbb{R}^N$ and only formal sense when $\mathcal{H}$ is an infinite-dimensional separable Hilbert space. We can write (formally in the case of infinite-dimensional $\mathcal{H}$):

$$\Pi(x) \propto \exp\{-\Phi(x) - \tfrac{1}{2}\langle x, Lx\rangle\} \ , \tag{7}$$

where we have defined $L = \mathcal{C}^{-1}$.

3.1. Hamiltonian Dynamics

The state space is now extended via the introduction of an auxiliary 'velocity' $v \in \mathcal{H}$; the original argument $x \in \mathcal{H}$ can be thought of as 'location'. We consider the 'total energy' function:

$$H(x, v; M) = \Phi(x) + \tfrac{1}{2}\,\langle x, Lx\rangle + \tfrac{1}{2}\,\langle v, Mv\rangle \ , \quad x \in \mathcal{H} \ , \tag{8}$$

for a user-specified 'mass' operator M such that M^{-1} is a well-defined covariance operator on $\mathcal{H}$. We define the distribution on $\mathcal{H} \times \mathcal{H}$:

$$\Pi(x, v) \propto \exp\{-H(x, v; M)\} = \exp\{-\Phi(x) - \tfrac{1}{2}\langle x, Lx\rangle - \tfrac{1}{2}\langle v, Mv\rangle\} \ .$$

Note that under $\Pi(x, v)$ we have $v \sim N(0, M^{-1})$. The Hamiltonian dynamics on $\mathcal{H} = \mathbb{R}^N$ are defined as follows:

$$\frac{dx}{dt} = M^{-1}\frac{\partial H}{\partial v} \ , \quad M\,\frac{dv}{dt} = -\frac{\partial H}{\partial x} \ ,$$

or, equivalently:

$$\frac{dx}{dt} = v \ , \quad M\,\frac{dv}{dt} = -Lx - \delta\Phi(x) \ . \tag{9}$$

Hamiltonian equations preserve the total energy, and in a probabilistic context they have $\Pi(x, v)$ as their invariant distribution. The proof (for the case $\mathcal{H} = \mathbb{R}^N$) under regularity conditions (see e.g. [15]) is straightforward and based on the fact that the solution operator of (9) is volume-preserving and energy-preserving. One motivation for the selection of mass matrix M could be that directions of higher variance under the target $\Pi(dx)$ should give higher marginal velocity variances and vice versa. In particular, selecting $M = L$ seems optimal in the case when $\Phi \equiv 0$ and the target distribution is simply $N(0, \mathcal{C})$, as we effectively transform the target to $N(0, I)$ and equalize all marginal variances.

Caution is needed when $\mathcal{H}$ is an infinite-dimensional Hilbert space. Following [6], one now is forced to choose the mass matrix $M = L$ to end up with a well-defined algorithm. In this case the energy function will be:

$$H(x, v) = \Phi(x) + \tfrac{1}{2}\langle x, Lx\rangle + \tfrac{1}{2}\langle v, Lv\rangle \ , \quad x \in \mathcal{H} \ , \tag{10}$$

and the Hamiltonian equations will become:

$$\frac{dx}{dt} = v \ , \quad \frac{dv}{dt} = -x - \mathcal{C}\,\delta\Phi(x) \ . \tag{11}$$

Showing in infinite dimensions that a solution operator of (11) exists and preserves $\Pi(x, v)$ is harder: an analytical proof is given in [6], under carefully formulated assumptions on $\mathcal{C}$ and Φ. As the differential equations in (11) cannot be solved analytically, HMC solves them numerically and then uses an accept/reject step to correct for violating preservation of total energy (and invariance of $\Pi(x, v)$).

3.2. Standard HMC on $\mathcal{H} = \mathbb{R}^N$

The standard HMC algorithm developed in [15] discretises the Hamiltonian equations (9) via a leapfrog scheme (we work under the selection $M = L$):

$$\begin{aligned} v_{h/2} &= v_0 - \tfrac{h}{2}\,x_0 - \tfrac{h}{2}\,\mathcal{C}\,\delta\Phi(x_0) \ , \\ x_h &= x_0 + h\,v_{h/2} \ , \\ v_h &= v_{h/2} - \tfrac{h}{2}\,x_h - \tfrac{h}{2}\,\mathcal{C}\,\delta\Phi(x_h) \ , \end{aligned} \tag{12}$$

giving rise to an operator $(x_0, v_0) \mapsto (x_h, v_h) = \psi_h(x_0, v_0)$ which is volume-preserving, and has the symmetricity property $\psi_h(x_h, -v_h) = (x_0, -v_0)$. HMC looks at Hamiltonian dynamics up to some time horizon $T > 0$, via the synthesis of

$$I = \lfloor \tfrac{T}{h} \rfloor \tag{13}$$

leapfrog steps, so we define ψ_h^I to be the synthesis of I mappings ψ_h. The standard HMC is given in Table 1 ($\mathcal{P}_x$ denotes projection on the x-argument). Due to the properties of the leapfrog operator mentioned above, it is easy to verify ([15]) that under regulatory conditions the employed acceptance probability provides Markov dynamics with invariant distribution $\Pi(x)$ in (1).

HMC on $\mathbb{R}^N$:

(i) Start with an initial value $x^{(0)} \in \mathbb{R}^N$ and set $k = 0$.

(ii) Given $x^{(k)}$ sample $v^{(k)} \sim N(0, \mathcal{C})$ and propose

$$x^\star = \mathcal{P}_x \, \psi_h^I(x^{(k)}, v^{(k)}) \ .$$

(iii) Calculate the acceptance probability

$$a(x^{(k)}, v^{(k)}) = 1 \wedge \exp\{-\Delta H(x^{(k)}, v^{(k)})\} \tag{14}$$

for $\Delta H(x, v) = H(\psi_h^I(x, v)) - H(x, v)$.

(iv) Set $x^{(k+1)} = x^\star$ with probability a; otherwise set $x^{(k+1)} = x^{(k)}$.

(v) Set $k \to k + 1$ and go to (ii).

Table 1: HMC on $\mathbb{R}^N$, with target $\Pi(x)$ in (1).

3.3. Advanced HMC on $\mathcal{H}$

Applying the standard algorithm in Table 1 on an N-dimensional projection of the SDE pathspace would give an algorithm for which: the proposal $x^\star$ would become an increasingly inappropriate candidate for a sample from the target with increasing N ([6]); thus, the acceptance probability would vanish with increasing N, assuming parameters h, T were kept fixed. Indeed, employing the mapping ψ_h^I for elements of the pathspace would project Brownian motion paths to paths of the wrong quadratic variation which would then necessarily have had acceptance probability 0 (see [9] for an analytical illustration in the case of MALA); the results in [8] suggest that one must decrease the step-size

h as $\mathcal{O}(N^{-1/4})$ to control the acceptance probability for increasing N. The advanced HMC algorithm avoids this degeneracy by exploiting the definition of the target as a change of measure from a Gaussian law. The development below follows [6]; a similar splitting of the Hamiltonian equations as the one showed below is also used in [5], but in a different context.

Hamiltonian equations (11) are split into the following two equations:

$$\frac{dx}{dt} = 0 \ , \quad \frac{dv}{dt} = -\mathcal{C}\,\delta\Phi(x) \ ; \tag{15}$$

$$\frac{dx}{dt} = v \ , \quad \frac{dv}{dt} = -x \ . \tag{16}$$

Notice that both equations can be solved analytically. We construct a numerical integrator for (11) by synthesizing steps on (15) and (16). Analytically, we define the solution operators of (15) and (16):

$$\Xi_t(x,v) = (x,\, v - t\,\mathcal{C}\,\delta\Phi(x)) \ ; \tag{17}$$

$$\tilde{\Xi}_t(x,v) = \big(\cos(t)\,x + \sin(t)\,v,\, -\sin(t)\,x + \cos(t)\,v\big) \ . \tag{18}$$

The numerical integrator for (11) is defined as follows:

$$\Psi_h = \Xi_{h/2} \circ \tilde{\Xi}_{h^*} \circ \Xi_{h/2} \ , \tag{19}$$

for small $h > 0$ and h^* a function of h defined in (21) below. We can synthesize steps up to some time horizon T. Defining I as in (13), Ψ_h^I will correspond to the synthesis of I steps Ψ_h. Ψ_h^I will provide the proposals for the MCMC steps.

Remark 3.1. *Critically, operators $\Xi_t(x,v)$, $\tilde{\Xi}_t(x,v)$ have the property that they preserve the absolute continuity properties of an input random pair (x,v) distributed according to the Gaussian law:*

$$Q_0(x,v) \propto \exp\{-\tfrac{1}{2}\langle x, Lx\rangle - \tfrac{1}{2}\langle v, Lv\rangle\} \ , \tag{20}$$

(so, also of any other distribution absolutely continuous w.r.t. Q_0). This is obvious for $\tilde{\Xi}_t(x,v)$ as it defines a rotation, so this map is in fact invariant for Q_0. Then, as illustrated with Proposition 2.1, Assumption 3.1 guarantees precisely that also $\Xi_t(x,v)$ preserves absolute continuity of Q_0. We re-emphasize here, that the solver employed for the standard HMC algorithm would generate proposals of a distribution singular to the target distribution.

We will use h^* such that:

$$\cos(h^*) = \tfrac{1-h^2/4}{1+h^2/4} \ , \tag{21}$$

though any choice is allowed. For this choice, it can be easily checked that the integrator $(x_0, v_0) \mapsto \Psi_h(x_0, v_0) =: (x_h, v_h)$ can be equivalently expressed as:

$$\begin{aligned} v_{h/2} &= v_0 - \tfrac{h}{2}\,\frac{x_0 + x_h}{2} - \tfrac{h}{2}\,\mathcal{C}\,\delta\Phi(x_0) \ , \\ x_h &= x_0 + h\,v_{h/2} \ , \\ v_h &= v_{h/2} - \tfrac{h}{2}\,\frac{x_0 + x_h}{2} - \tfrac{h}{2}\,\mathcal{C}\,\delta\Phi(x_h) \ , \end{aligned} \tag{22}$$

which can now be interpreted as a semi-implicit-type integrator of (11). Under the interpretation (22), the justification for the choice (21) is that it delivers an integrator Ψ_h that carries out steps of similar size, h, in the x and v directions, in accordance with standard HMC.

The complete algorithm is determined in Table 2.

HMC on Hilbert space $\mathcal{H}$:

(i) Start with an initial value $x^{(0)} \sim \Pi_0 \equiv N(0, \mathcal{C})$ and set $k = 0$.

(ii) Given $x^{(k)}$ sample $v^{(k)} \sim N(0, \mathcal{C})$ and propose

$$x^\star = \mathcal{P}_x \, \Psi_h^I(x^{(k)}, v^{(k)}) \ .$$

(iii) Consider

$$a(x^{(k)}, v^{(k)}) = 1 \wedge \exp\{-\Delta H(x^{(k)}, v^{(k)})\} \tag{23}$$

for $\Delta H(x, v) = H(\Psi_h^I(x, v)) - H(x, v)$.

(iv) Set $x^{(k+1)} = x^\star$ with probability a; otherwise set $x^{(k+1)} = x^{(k)}$.

(v) Set $k \to k + 1$ and go to (ii).

Table 2: HMC on $\mathcal{H}$, with target $\Pi(x)$ in (1).

Remark 3.2. *The acceptance probability in the table is at the moment defined only formally, as $H(x, v) = \infty$ a.s.. To see that, notice that using the Karhunen-Loève expansion in (6) for $x \sim \Pi_0$ we have $\langle x, Lx \rangle \equiv \sum_{p=1}^{\infty} \xi_p^2$, for ξ_p iid $N(0, 1)$. We re-express the acceptance probability in the following section in a way that illustrates that the difference $\Delta H(x, v) = H(\Psi_h^I(x, v)) - H(x, v)$ is a.s. well-defined; from a practical point of view, for the N-dimensional projection used in practice one could still use directly the expression $\Delta H(x, v) = H(\Psi_h^I(x, v)) - H(x, v)$ as each of the two H-terms will grow as $\mathcal{O}(N)$.*

Remark 3.3. *We will not prove the existence of a solution for the Hamiltonian equations on Hilbert space (15)-(16) or that the solution would preserve $\Pi(x, v)$ as such proofs would require a level and amount of technicalities out of the scope of the paper. In Section 3.4 below we will prove the validity of the algorithm in Table 2 which uses directly the numerical integrators of these equations in (17)-(18). This seems to suffice from a practical point of view: our proof below indicates that the algorithm will not collapse as $N \to \infty$ but will converge to a limit, with N being the dimension of the vector used instead of complete infinite-dimensional diffusion paths when running the algorithms on a personal*

computer; then, for a fixed finite dimension N one can resort to properties of finite-dimensional Hamiltonian equations to justify that, under standard regulatory conditions, they will indeed preserve the N-dimensional target distribution, thus one can attain average acceptance probabilities arbitrarily close to 1 by decreasing the step-size h.

3.4. Proof of Validity for Advanced HMC

We consider the Gaussian product measure $Q_0 = N(0,\mathcal{C}) \otimes N(0,\mathcal{C})$ on $\mathcal{H}\times\mathcal{H}$ as in (20) and the bivariate distribution Q via the change of measure:

$$Q(dx, dv) = \exp\{-\Phi(x)\}\, Q_0(dx, dv) \ .$$

We also consider the sequence of probability measures on $\mathcal{H} \times \mathcal{H}$:

$$Q^{(i)} = Q \circ \Psi_h^{-i} \ , \quad 1 \leq i \leq I$$

the sequence $(x_i, v_i) = \Psi_h^i(x_0, v_0)$, and set:

$$g(x) := -\mathcal{C}^{1/2}\, \delta\Phi(x) \ , \quad x \in \mathcal{H} \ .$$

Note that under Assumption 3.1, $g(x)$ is a well-defined element of the Hilbert space $\mathcal{H}$ a.s. under Π_0. Using Proposition 2.1, we can now prove the following:

Proposition 3.1. *We have that:*

$$\frac{dQ^{(i)}}{dQ_0}(x_i, v_i) = \frac{dQ^{(i-1)}}{dQ_0}(x_{i-1}, v_{i-1}) \cdot G(x_i, v_i) \cdot G(x_{i-1}, v_{i-1} + \tfrac{h}{2}\, \mathcal{C}^{1/2} g(x_{i-1})) \quad ,$$

where we have defined:

$$G(x, v) = \exp\left\{\langle \tfrac{h}{2}\, g(x), \mathcal{C}^{-1/2} v\rangle - \tfrac{1}{2}|\tfrac{h}{2}\, g(x)|^2\right\} \ .$$

Proof. We will use the chain rule and Proposition 2.1. Recall that for any two measurable spaces $(E, \mathcal{E})$, $(E', \mathcal{E}')$, probability measures M, M_0 on $(E, \mathcal{E})$ and 1-1 mapping $F : (E, \mathcal{E}) \mapsto (E', \mathcal{E}')$, we have the following identity rule for the Radon-Nikodym derivative:

$$\frac{d\{M_1 \circ F^{-1}\}}{d\{M_0 \circ F^{-1}\}}(x) = \frac{dM_1}{dM_0}(F^{-1}(x)) \ . \tag{24}$$

We now work as follows. Following the definition of Ψ_h from (19), we have the equality of probability measures:

$$Q^{(i)} = Q^{(i-1)} \circ \Xi_{h/2}^{-1} \circ \tilde{\Xi}_{h^*}^{-1} \circ \Xi_{h/2}^{-1} \ .$$

Thus, we have that:

$$\begin{aligned}
\frac{dQ^{(i)}}{dQ_0}(x_i, v_i) &= \frac{d\left\{Q^{(i-1)} \circ \Xi_{h/2}^{-1} \circ \tilde{\Xi}_{h^*}^{-1} \circ \Xi_{h/2}^{-1}\right\}}{dQ_0}(x_i, v_i) \\
&= \frac{d\left\{Q^{(i-1)} \circ \Xi_{h/2}^{-1} \circ \tilde{\Xi}_{h^*}^{-1} \circ \Xi_{h/2}^{-1}\right\}}{d\left\{Q_0 \circ \Xi_{h/2}^{-1}\right\}}(x_i, v_i) \times \frac{d\left\{Q_0 \circ \Xi_{h/2}^{-1}\right\}}{dQ_0}(x_i, v_i) \\
&= \frac{d\left\{Q^{(i-1)} \circ \Xi_{h/2}^{-1} \circ \tilde{\Xi}_{h^*}^{-1}\right\}}{dQ_0}(\Xi_{h/2}^{-1}(x_i, v_i)) \times G(x_i, v_i) \ ;
\end{aligned} \tag{25}$$

we have used the chain rule in the second line, then (24) and Proposition 2.1 (in this case $v_0 \equiv \frac{h}{2}\, g(x)$) in the third line. Using the fact that $Q_0 \circ \tilde{\Xi}_{h^*}^{-1} \equiv Q_0$ and that also $(\tilde{\Xi}_{h^*}^{-1} \circ \Xi_{h/2}^{-1})(x_i, v_i) \equiv \Xi_{h/2}(x_{i-1}, v_{i-1})$ we have that:

$$\frac{d\,\{Q^{(i-1)} \circ \Xi_{h/2}^{-1} \circ \tilde{\Xi}_{h^*}^{-1}\}}{dQ_0}\,(\Xi_{h/2}^{-1}(x_i, v_i)) \equiv \frac{d\,\{Q^{(i-1)} \circ \Xi_{h/2}^{-1}\}}{dQ_0}\,(\Xi_{h/2}(x_{i-1}, v_{i-1}))\ .$$

Finally, working as in (25) we have that:

$$\begin{aligned}\frac{d\,\{Q^{(i-1)} \circ \Xi_{h/2}^{-1}\}}{dQ_0}\,(\Xi_{h/2}(x_{i-1}, v_{i-1})) &= \\ &= \frac{dQ^{(i-1)}}{dQ_0}\,(x_{i-1}, v_{i-1}) \times \frac{d\,\{Q_0 \circ \Xi_{h/2}^{-1}\}}{dQ_0}(\Xi_{h/2}(x_{i-1}, v_{i-1})) \\ &= \frac{dQ^{(i-1)}}{dQ_0}\,(x_{i-1}, v_{i-1}) \times G(\Xi_{h/2}(x_{i-1}, v_{i-1}))\ .\end{aligned}$$

The definition of $\Xi_{h/2}$ gives $G(\Xi_{h/2}(x_{i-1}, v_{i-1})) \equiv G(x_{i-1}, v_{i-1} + \frac{h}{2}\,\mathcal{C}^{1/2} g(x_{i-1}))$. Following the calculation from (25) we have now proven the requested result. □

Thus, using Proposition 3.1 iteratively we have now obtained that:

$$\frac{dQ^{(I)}}{dQ_0}(x_I, v_I) = \frac{dQ}{dQ_0}(x_0, v_0) \times \prod_{i=1}^{I} G(x_i, v_i)\, G(x_{i-1}, v_{i-1} + \tfrac{h}{2}\,\mathcal{C}^{1/2} g(x_{i-1}))\ . \quad (26)$$

Now, following the definition of Ψ_h in (19), we set:

$$\begin{aligned} v_{i-1}^{-} &= \mathcal{P}_v\, \Xi_{h/2}(x_{i-1}, v_{i-1}) \equiv v_{i-1} + \tfrac{h}{2}\,\mathcal{C}^{1/2} g(x_{i-1})\ ; \\ v_i^{+} &= \mathcal{P}_v\, \tilde{\Xi}_{h^*}(x_{i-1}, v_{i-1}) \equiv v_i - \tfrac{h}{2}\,\mathcal{C}^{1/2} g(x_i)\ . \end{aligned}$$

Using these definitions, for any $h, h^* > 0$ we have that:

$$\begin{aligned}&\log\{\, G(x_i, v_i)\, G(x_{i-1}, v_{i-1} + \tfrac{h}{2}\,\mathcal{C}^{1/2} g(x_{i-1}))\,\} = \\ &\quad = \langle \tfrac{h}{2}\, g(x_i), \mathcal{C}^{-1/2} v_i\rangle - \tfrac{1}{2}|\tfrac{h}{2} g(x_i)|^2 + \langle \tfrac{h}{2}\, g(x_{i-1}), \mathcal{C}^{-1/2} v_{i-1}\rangle + \tfrac{1}{2}|\tfrac{h}{2} g(x_{i-1})|^2 \\ &\quad = \tfrac{1}{2}\,\langle v_i, L v_i\rangle - \tfrac{1}{2}\,\langle v_i^{+}, L v_i^{+}\rangle - \tfrac{1}{2}\,\langle v_{i-1}, L v_{i-1}\rangle + \tfrac{1}{2}\,\langle v_{i-1}^{-}, L v_{i-1}^{-}\rangle \\ &\quad = \tfrac{1}{2}\,\langle x_i, L x_i\rangle + \tfrac{1}{2}\,\langle v_i, L v_i\rangle - \tfrac{1}{2}\,\langle x_{i-1}, L x_{i-1}\rangle - \tfrac{1}{2}\,\langle v_{i-1}, L v_{i-1}\rangle\ .\end{aligned}$$

The last equation is due to the mapping $(x_{i-1}, v_{i-1}^{-}) \mapsto (x_i, v_i^{+})$ corresponding to the modulus-preserving rotation $\tilde{\Xi}_{h^*}$. Thus, we can rewrite (26) as follows:

$$\frac{dQ^{(I)}}{dQ_0}(x_I, v_I) = \exp\{\Delta H(x_0, v_0) - \Phi(x_I)\}\ . \quad (27)$$

The above expression will be used at proving the main result below.

Remark 3.4. *The operator Ψ_h (thus, also Ψ_h^I) has the following properties:*

i) Ψ_h is symmetric, that is $\Psi_h \circ S \circ \Psi_h = S$ where $S(x, v) = (x, -v)$.

ii) Ψ_h is (formally) volume-preserving, as it preserves volume when $\mathcal{H} \equiv \mathbb{R}^d$.

Theorem 3.1. *The Markov chain with transition dynamics specified in Table 2 has invariant distribution $\Pi(x)$ in (1).*

Proof. Assuming stationarity, so $(x_0, v_0) \sim Q$, we can write for the next position, x', of the Markov chain (recall that $(x_I, v_I) = \Psi_h^I(x_0, v_0)$):

$$x' = \mathrm{I}\,[\,U \leq a(\Psi_h^{-I}(x_I, v_I))\,]\,x_I + \mathrm{I}\,[\,U > a(x_0, v_0)\,]\,x_0 \ ,$$

for a uniform random variable $U \sim \mathrm{Un}\,[0, 1]$. Let $f : \mathcal{H} \mapsto \mathbb{R}$ be bounded and continuous. We need to prove that:

$$\mathrm{E}\,[\,f(x')\,] = \mathrm{E}\,[\,f(x_0)\,] \ .$$

Integrating out U from above we get:

$$\mathrm{E}\,[\,f(x')\,] = \mathrm{E}[\,f(x_I)\,a(x_0, v_0)\,] - \mathrm{E}[\,f(x_0)\,a(x_0, v_0)\,] + \mathrm{E}\,[\,f(x_0)\,] \ . \tag{28}$$

Note now that (here, we need to stress the integrators in expectations/integrals and will show them explicitly as a subscript of E):

$$\begin{aligned}
\mathrm{E}[\,f(x_I)\,a(x_0, v_0)\,] &= \mathrm{E}_{Q^{(I)}}[\,f(x_I)\,a(\Psi_h^{-I}(x_I, v_I))\,] \\
&\overset{(27)}{=} \mathrm{E}_{Q_0}[\,f(x_I)\,a(\Psi_h^{-I}(x_I, v_I))\,e^{\Delta H(\Psi_h^{-I}(x_I, v_I)) - \Phi(x_I)}\,] \\
&= \mathrm{E}_{Q_0}[\,f(x_I)\,(\,1 \wedge e^{\Delta H(\Psi_h^{-I}(x_I, v_I))}\,)\,e^{-\Phi(x_I)}\,] \\
&= \mathrm{E}_{Q}[\,f(x_I) \cdot 1 \wedge e^{\Delta H(\Psi_h^{-I}(x_I, v_I))}\,] \\
&= \mathrm{E}_{Q}[\,f(x_I) \cdot 1 \wedge e^{\Delta H(\Psi_h^{-I}(x_I, -v_I))}\,] \ .
\end{aligned} \tag{29}$$

(For the last equation, notice that (x_I, v_I) and $(x_I, -v_I)$ have the same law Q.) Now, that due to the symmetricity property $\Psi_h^I \circ S \circ \Psi_h^I = S$ of the leapfrog operator in Remark 3.4 we have that $\Psi_h^{-I} \circ S = S \circ \Psi_h^I$. Thus, we have:

$$\begin{aligned}
\Delta H(\Psi_h^{-I}(x_I, -v_I))) &= \Delta H(S \circ \Psi_h^I(x_I, v_I))) \\
&= H(S(x_I, v_I)) - H(S \circ \Psi_h^I(x_I, v_I)) \equiv -\Delta H(x_I, v_I) \ ,
\end{aligned}$$

where is the last equation we used the fact that $H \circ S = H$ due to the energy H being quadratic in the velocity v. Thus, using this in (29), we have that:

$$\mathrm{E}[\,f(x_I)\,a(x_0, v_0)\,] = \mathrm{E}_Q[\,f(x_I) a(x_I, v_I)\,] \equiv \mathrm{E}[\,f(x_0)\,a(x_0, v_0)\,] \ . \tag{30}$$

So, from (28), the proof is now complete. □

Remark 3.5. *The demonstration of validity of standard HMC [15] does not require the recursive calculation of the forward density (27) as it exploits the preservation of volume (unit Jacobian) for the mapping $(x_0, v_0) \mapsto \psi_h^I(x_0, v_0)$ to directly prove the analogue to (30). So, finding (27) overcomes the difficulty of making sense of a Jacobian for the transform Ψ_h^I on the infinite-dimensional Hilbert space.*

3.5. Advanced MCMC Samplers on Pathspace

A number of MCMC algorithms corresponding to an upgrade of standard RWM, MALA and HMC on the infinite-dimensional pathspace are now available. So far in Section 3 we have defined HMC on pathspace under Assumption 3.1. In [9] the definition of advanced MALA on the pathspace is provided; also advanced RWM on pathspace was defined in that paper. We now briefly review these two local-move algorithms.

The starting point for MALA is a Langevin SDE with drift $\frac{1}{2}\,\mathcal{C}\,\delta\log\Pi(x)$ and coefficient $\mathcal{C}^{1/2}$, that is, after a calculation on the drift:

$$\frac{dx}{dt} = -\tfrac{1}{2}\,x - \tfrac{1}{2}\,\mathcal{C}\,\delta\Phi(x) + \mathcal{C}^{1/2}\,\frac{dw}{dt}\ . \tag{31}$$

In an Euclidean setting $\{w_t\}$ denotes a standard Brownian motion, whereas in the pathspace it denotes a cylindrical Brownian motion. In both cases, the process can be easily understood via the distribution of it's increments, as $\mathcal{C}^{1/2}\,\frac{(w_{t+s}-w_t)}{\sqrt{s}} \sim N(0,\mathcal{C})$. On pathspace, the SDE (31) is shown in [9] to have invariant distribution Π under Lipschitz continuity and absolute boundedness assumptions on $\delta\Phi$. In the practically interesting case of nonlinearity, this SDE cannot be solved analytically. So, a proposal can be derived via the following Euler-type scheme on (31) for an finite increment $\Delta t > 0$:

$$x^* - x = -\Delta t\,(\theta\,\tfrac{x^*}{2} + (1-\theta)\tfrac{x}{2}) - \tfrac{\Delta t}{2}\,\mathcal{C}\,\delta\,\Phi(x) + \sqrt{\Delta t}\,N(0,\mathcal{C})\ . \tag{32}$$

Standard MALA is derived from an explicit Euler scheme with $\theta = 0$ and advanced pathspace MALA from a semi-implicit scheme with $\theta = 1/2$. Contrasting (32) with the leapfrog steps in (12) and (22), one can easily check that standard (resp. advanced) MALA is a particular case of standard (resp. advanced) HMC when choosing $h = \sqrt{\Delta t}$ and a single leapfrog step $I = 1$.

Finally, a RWM algorithm on pathspace is derived in [9] via proposal (32) for $\theta = 1/2$ but also with omitting the nonlinear term $\mathcal{C}\,\delta\Phi(x)$. That is, the proposal for advanced RWM is:

$$x^* = \rho\,x + \sqrt{1-\rho^2}\,N(0,\mathcal{C})\ ,$$

with parameter

$$\rho = \frac{1-\frac{\Delta t}{4}}{1+\frac{\Delta t}{4}}\ .$$

The Metropolis-Hastings acceptance probability for this proposal (see [9]) is reminiscent of the one for standard RWM, namely $1 \wedge \frac{\Pi(x^*)}{\Pi(x)}$, which also explains the interpretation of this algorithm as 'advanced RWM'. Table 3 summarises the three pathspace samplers looked at in this paper together with their standard versions for finite-dimensional spaces.

Algorithm	Pathspace Proposal	Standard Proposal
HMC	$x^* = \mathcal{P}_x \Psi_h^I(x,v)$	$x^* = \mathcal{P}_x \psi_h^I(x,v)$
MALA	$x^* = \rho\, x + \sqrt{1-\rho^2}\, v - \frac{\Delta t}{2}\, \mathcal{C}\, \delta\Phi(x)$	$x^* = (1-\frac{\Delta t}{2})x - \frac{\Delta t}{2}\, \mathcal{C}\, \delta\Phi(x) + \sqrt{\Delta t}\, v$
RWM	$x^* = \rho\, x + \sqrt{1-\rho^2}\, v$	$x^* = x + \sqrt{\Delta t}\, v$

Table 3: MCMC algorithms on pathspace together with their standard versions. In all cases $v \sim N(0, \mathcal{C})$. HMC for $I = 1$ and $h = \sqrt{\Delta t}$ coincides with MALA.

4. HMC Superiority in an Analytical Study

In the applications that we show in the next sections HMC appears to be much more efficient than MALA, even if they both use the same information about the target distribution in the form of the derivative $\delta \log \Pi$. So, the synthesis of deterministic steps for HMC, by avoiding random-walk-type behaviour for the Markov chain dynamics, seems to be providing significant computational advantages for HMC. We will illustrate this analytically in this section for the case of a linear target distribution corresponding to an Ornstein-Uhlenbeck (OU) diffusion process. In particular, we will show that HMC gains orders of complexity compared to MALA and RWM when MALA itself does not gain a complexity benefit over RWM. So, an important message derived here is that designated use of information on the derivative can have significant effect on the efficiency of such MCMC methods.

We will consider the OU bridge:

$$\begin{aligned} dX_u &= -\kappa X_u\, dt + dB_u \ , \\ X_0 &= X_\ell = 0 \ , \end{aligned} \tag{33}$$

for reversion parameter $\kappa > 0$ and path-length $\ell > 0$. From Girsanov's theorem ([33]) we get that the target distribution is defined on the Hilbert space $\mathcal{H} = L^2([0,\ell],\mathbb{R})$ and is expressed as in the general form (1), so that:

$$\frac{d\Pi}{d\Pi_0}(x) = \exp\{-\Phi(x)\} \ ; \quad \Pi_0 = N(0, \mathcal{C}^{bb}) \ , \ \Phi(x) = \tfrac{\kappa^2}{2} \int_0^\ell x^2(u) du + c \ , \tag{34}$$

for some constant $c \in \mathbb{R}$, with $N(0, \mathcal{C}^{bb})$ the distribution of a Brownian bridge with $x(0) = x(\ell) = 0$. We will look at the complexity of pathspace samplers as a function of the length ℓ of the bridge. Our main result summarises the mixing times as follows:

$$\begin{aligned} \text{RWM}: &\ \ \mathcal{O}(\ell^2) \ ; \\ \text{MALA}: &\ \ \mathcal{O}(\ell^2) \ ; \\ \text{HMC}: &\ \ \mathcal{O}(\ell) \ . \end{aligned} \tag{35}$$

The notion of mixing time is used here in an informal, practical manner and should not be confused with analytical definitions of various different versions of

mixing times appearing in the Markov chain literature. In particular, the results below provide appropriate scalings of the step-sizes for the relevant MCMC samplers as a function of ℓ that deliver non-vanishing acceptance probabilities as ℓ grows. Then, informal arguments will be used to connect mixing times with the inverse of such step-sizes.

4.1. Acceptance Probability for RWM, MALA, HMC

The proof in Proposition 4.1 below follows closely the derivations in the PhD thesis [39], where a similar scaling problem has been considered over the reversion parameter κ. We have decided to include an analytical proof in the Appendix for reasons of completeness, but also because we have made some modifications that make it easier for the reader to follow the derivations.

Karhunen-Loève expansion for Brownian Bridge and OU Bridge

The Karhunen-Loéve expansion (see Section 2) of the Gaussian distributions corresponding to the target OU bridge and the reference Brownian bridge will be used in this section. In particular, we will use the orthonormal basis $\{\phi_p\}_{p=1}^{\infty}$ of $\mathcal{H}$ corresponding to the eigenfunctions of $\mathcal{C}^{bb}$ and make the standard correspondence $x \mapsto \{x_p\}_{p=1}^{\infty}$ between an element $x \in \mathcal{H}$ and it's squared summable co-ordinates $x_p = \langle x, \phi_p \rangle$ w.r.t. the basis $\{\phi_p\}$. In particular, the eigen-structure $\{\lambda_p, \phi_p\}_{p=1}^{\infty}$ of $\mathcal{C}^{bb}$ is specified as follows (see e.g. [39]):

$$\lambda_p = \frac{\ell^2}{\pi^2 p^2} \; ; \quad \phi_p(u) = \sqrt{\frac{2}{\ell}} \, \sin(\frac{\pi p u}{\ell}) \; . \tag{36}$$

Then, the Karhunen-Loève expansion of the two Gaussian distributions w.r.t. the above basis of sinusoidals is as below (see e.g. [39]):

$$\text{BB:} \;\; x = \sum_{p=1}^{\infty} \frac{\ell}{\pi p} \, \xi_p \, \phi_p \; ; \quad \text{OU Bridge:} \;\; x = \sum_{p=1}^{\infty} \frac{1}{\sqrt{\frac{\pi^2 p^2}{\ell^2} + \kappa^2}} \, \xi_p \, \phi_p \; , \tag{37}$$

where $\{\xi_p\}_{p=1}^{\infty}$ are iid variables from $N(0,1)$.

Proposition 4.1. *Consider the advanced MALA and RWM algorithms described in Table 3 with target distribution Π in (34). If $a = a(x,v)$ is the acceptance probability of the proposal for current position x and $v \sim N(0, \mathcal{C}^{bb})$, then in stationarity ($x \sim \Pi$) we have the following:*

- *If $\Delta t = c/\ell^2$ for some constant $c > 0$, then $\limsup_{\ell} \mathrm{E}\,[\,a\,] > 0$.*
- *If $\Delta t = c/\ell^{\epsilon}$ for $\epsilon \in (0,2)$ and a constant $c > 0$, then $\lim_{\ell \to \infty} \mathrm{E}\,[\,a\,] = 0$.*

Proof. See the Appendix. □

We will now derive a corresponding result for HMC. Recall that the step-size for HMC is denoted by h instead of Δt. We will show that the scaling $h = c/\ell$ for some constant $c > 0$ will control the average acceptance probability. We work as follows (under the choice $h = c/\ell$). Each leapfrog step Ψ_h, defined via (19), (21), can in this case be written as a linear operator:

$$\Psi_h = \begin{pmatrix} \rho - (1-\rho)\kappa^2 \mathcal{C}^{bb} & \sqrt{1-\rho^2} \\ -\frac{1-\left(\rho-(1-\rho)\kappa^2\mathcal{C}^{bb}\right)^2}{\sqrt{1-\rho^2}} & \rho - (1-\rho)\kappa^2\mathcal{C}^{bb} \end{pmatrix}$$

where we have set here:

$$\rho = \frac{1 - \frac{h^2}{4}}{1 + \frac{h^2}{4}} \ .$$

So, the mapping Ψ_h above, can be equivalently expressed in terms of it's effect to the pth co-ordinates x_p, v_p of x, v respectively w.r.t. the orthonormal basis $\{\phi_p\}$ in (36) as follows:

$$\Psi_{h,p} = \begin{pmatrix} \rho - (1-\rho)\kappa^2 \lambda_p & \sqrt{1-\rho^2} \\ -\frac{1-\left(\rho-(1-\rho)\kappa^2\lambda_p\right)^2}{\sqrt{1-\rho^2}} & \rho - (1-\rho)\kappa^2\lambda_p \end{pmatrix} \ .$$

Powers of the above matrix will be determined via it's eigenstructure. We will only consider the case when there will be complex eigenvalues, i.e. when:

$$|\rho - (1-\rho)\,\kappa^2\lambda_p| < 1 \ ,$$

as in the alternative scenario there will be an eigenvalue of modulus greater than one (since the Jacobian of the above matrix is unit) whose powers will explode rendering the algorithm unstable. The above is equivalent to requiring that $(4 - \frac{c^2}{\ell^2} - \frac{2\,c^2\,\kappa^2}{p^2\,\pi^2})/(4 + \frac{c^2}{\ell^2})$ lies in $(-1, 1)$, which can be easily seen to be guaranteed, for any $\ell \geq \ell_0 > 0$ and for all $p \geq 1$, under the condition:

$$c\,\kappa < 2\pi \ . \tag{38}$$

This condition specifies the region of stability (see e.g. [32]) for the discretisation scheme of the Hamiltonian dynamics in our context. Under (38), we can conveniently write:

$$\Psi_{h,p}^I = \begin{pmatrix} \cos(\theta_p) & a_p\,\sin(\theta_p) \\ -\frac{1}{a_p}\,\sin(\theta_p) & \cos(\theta_p) \end{pmatrix}^I \equiv \begin{pmatrix} \cos(I\theta_p) & a_p\,\sin(I\theta_p) \\ -\frac{1}{a_p}\,\sin(I\theta_p) & \cos(I\theta_p) \end{pmatrix} \tag{39}$$

where we have set:

$$\cos(\theta_p) = \rho - (1-\rho)\,\kappa^2\lambda_p \ ; \quad \sin(\theta_p) = \sqrt{1-\cos^2(\theta_p)} \ ; \quad a_p = \frac{\sqrt{1-\rho^2}}{\sin(\theta_p)} \ . \tag{40}$$

We can now derive the following result.

Proposition 4.2. *Consider the advanced HMC algorithm in Table 2 with target distribution* Π *in (34). If* $a = a(x, v)$ *is the acceptance probability of the proposal for current position* x *and* $v \sim N(0, \mathcal{C}^{bb})$*, then in stationarity (*$x \sim \Pi$*) we have the following:*

- *If* $h = c/\ell$ *with* $c\,\kappa < 2\pi$ *then* $\liminf_\ell \mathrm{E}\,[\,a\,] > 0$.

Proof. See the Appendix. □

So, under the selection $h = c/\ell$, the average acceptance probability is lower bounded by a constant for arbitrarily long bridges.

Remark 4.1. *We could now informally connect the above step-sizes that control the average acceptance probability for the advanced RWM, MALA and HMC algorithms with their mixing times, which will involve their inverses, as stated in (35). One could think of the effect of the proposal of each algorithm for increasing* ℓ *on a fixed time-window for a path, say on* $[0, \ell_0]$ *for some* $\ell_0 > 0$*. For HMC, the synthesis of* $I = \lfloor \frac{T}{h} \rfloor$ *leapfrog steps will give a proposal moving the whole sub-path on* $[0, \ell_0]$ *an* $\mathcal{O}(1)$*-distance within it's state space. To show that, we ignore for a moment the effect of the nonlinear map* $\Xi_{h/2}$ *at the the leapfrog update in (19) and focus on the synthesis of* I *linear maps* $\tilde{\Xi}_{h^*}$*. This gives:*

$$\tilde{\Xi}^I_{h^*} = \begin{pmatrix} \cos(Ih^*) & \sin(Ih^*) \\ -\sin(Ih^*) & \cos(Ih^*) \end{pmatrix} \longrightarrow \begin{pmatrix} \cos(T) & \sin(T) \\ -\sin(T) & \cos(T) \end{pmatrix} , \quad \text{as } \ell \to \infty \ .$$

The effect of the nonlinear operator $\Xi_{h/2}$ *does not have a similarly simple interpretation, but should not offset the main effect of proposals making* $\mathcal{O}(1)$*-steps from a current position, for arbitrarily large* ℓ*. Thus, as a function of* ℓ*, the mixing time for advanced HMC only corresponds to the order of the number of leapfrog steps,* $\mathcal{O}(\ell)$*. For advanced RWM, shown in Table 3, for* $\Delta t = c/\ell^2$ *we can express the proposal as:*

$$x^* = (1 + \mathcal{O}(\ell^{-2}))\, x + \tfrac{\sqrt{c}}{\ell}\,(1 + \mathcal{O}(\ell^{-2}))\, \xi \ . \tag{41}$$

Here, due to the random-walk nature of the proposal, the algorithm will have to synthesize $\mathcal{O}(\ell^2)$*-steps to move* $\mathcal{O}(1)$*-distance from a current position for a fixed point of the sub-path in* $(0, \ell_0]$*, thus the* $\mathcal{O}(\ell^2)$*-mixing time. Finally, for MALA, one has to refer to the interpretation of the algorithm as a discretisation of an SDE on the pathspace, expressed in (31): without being rigorous here, advanced MALA essentially carries out steps of size* $\Delta t = \mathcal{O}(\ell^{-2})$ *along the continuous-time dynamics, thus will require* $1/\Delta t = \mathcal{O}(\ell^2)$ *steps to propagate a point of the sub-path on* $[0, \ell_0]$ *an* $\mathcal{O}(1)$*-distance from it's current position.*

Of course, a rigorous analysis of mixing times would involve characterising the eigenvalues of the Markov chains, but this is beyond the scope of this paper.

5. Advanced MCMC for Diffusion-Driven Models

In this section we return to the general framework of Section 1.1 that can accommodate a wide range of applications. Recall that the statistical model here is driven by the solution V of the SDE (2) involving some unknown parameter θ; then, conditionally on V and θ there are some observations Y via the Lebesgue density $p(y|v,\theta)$. We will now apply the previously defined advanced MCMC samplers in the context of such diffusion-driven models to efficiently sample from the path-valued distribution $p(v|y,\theta)$. Such models provide a generic framework where target distributions defined as a change of measure from Gaussian laws as in (1) will naturally rise due to the probabilistic dynamics being driven by Brownian motion. We will first give a brief description of inferential issues related with such models, before we show details over the applicability of the advanced MCMC methods in such a context.

The method in [36] succeeds in breaking down the singularity between V and the parameters in the diffusion coefficient $\sigma(\cdot;\theta)$ through appropriate use of the Lamperti transform; see [27] and [28] for extensions. This framework covers the case of *reducible* diffusions, i.e. SDEs that can be transformed via an 1-1 map to ones of unit diffusion coefficient. Now, [29] offers some extensions to irreducible diffusions utilizing time change transformations; nevertheless the most general framework for irreducible diffusions is offered from [12, 24] requiring only invertibility of the diffusion coefficient. For ease of exposition we will first consider the case of univariate ($d=1$) reducible diffusions and work with the reparametrisation recipe of [36]. However, advanced HMC algorithms can also be defined under (3) with the approach of [24, 12]; details are postponed until Section 7.

For reducible diffusions we can apply the Lamperti transform on the original SDE model (2), i.e. $V_u \to \eta(V_u;\theta) =: X_u$ where

$$\eta(v;\theta) = \int_{V_0}^{v} \frac{1}{\sigma(z,\theta)} dz$$

is the antiderivative of $\sigma^{-1}(\cdot;\theta)$. Depending on the context, V_0 can be treated as fixed or as an additional unknown parameter. Assuming that $\sigma(\cdot;\theta)$ is continuously differentiable, an application of Ito's lemma provides the SDE of the transformed diffusion X as:

$$dX_u = \nu(X_u;\theta)du + dB_u \ , \quad X_0 = 0 \ , \tag{42}$$

where

$$\nu(x;\theta) = \frac{\mu\big(\eta^{-1}(x;\theta);\theta\big)}{\sigma\big(\eta^{-1}(x;\theta);\theta\big)} - \tfrac{1}{2}\sigma'\big(\eta^{-1}(x;\theta);\theta\big) \ .$$

Girsanov's theorem gives that (for Π_0 being the standard Brownian motion on $[0,\ell]$):

$$\frac{d\Pi}{d\Pi_0}(x|\theta) = \exp\Big\{ \int_0^{\ell} \nu\big(x(u);\theta\big)dx(u) - \tfrac{1}{2}\int_0^{\ell} |\nu\big(x(u);\theta\big)|^2\, du \Big\} \ .$$

Thus, for the distribution of X conditionally on the data Y we have that:

$$\begin{aligned}\frac{d\Pi}{d\Pi_0}(x|y,\theta) &\propto p(y|\eta^{-1}(x;\theta),\theta)\,\frac{d\Pi}{d\Pi_0}(x|\theta)\\ &\propto \exp\{-\Phi(x;y,\theta)\}\ ,\end{aligned} \tag{43}$$

where we have defined:

$$\begin{aligned}\Phi(x;y,\theta) = -\int_0^\ell \nu(x(u);\theta)dx(u) + \tfrac{1}{2}\int_0^\ell \nu^2(x(u);\theta)du\\ - \log p(y|\eta^{-1}(x;\theta),\theta)\ .\end{aligned} \tag{44}$$

Next, assuming a generic structure for $p(y|v,\theta)$ relevant for practical applications, we will verify Assumption 3.1 and therefore demonstrate that our advanced HMC method is well-defined for such models.

5.1. Calculation of $\mathcal{C}\,\delta\Phi(x)$

Motivated by the expression in (44) for $\Phi(x|y,\theta)$ and the structure of the data density $p(Y|V,\theta)$ arising in applications, we will carry out calculations assuming the following general form:

$$\Phi(x) = \alpha(x(u_1),x(u_2),\ldots,x(u_M)) + \beta(I_1,I_2,\ldots,I_L) + \gamma(S_1,S_2,\ldots,S_J) \tag{45}$$

where we have set:

$$I_l = \int_0^\ell z_l(u,x(u))du\ \ ,\ 1\le l\le L\ ;\quad S_j = \int_0^\ell r_j(u,x(u))dx(u)\ \ ,\ 1\le j\le J\ ,$$

for positive integers M, L, J, times $u_1 < u_2 < \cdots < u_M$ in $[0,\ell]$ that could be determined by the data Y and functions $\alpha,\beta,\gamma,z_l,r_j$ determined via the particular model. Explicit instances of this structure will be provided in the example applications in the sequel. Here, the target posterior distribution $\Pi(dx)$ is defined on the Hilbert space of squared integrable paths $\mathcal{H} = L^2([0,\ell],\mathbb{R})$ (with appropriate boundary conditions). The centered Gaussian reference measure Π_0 will correspond to a Brownian motion (thus, boundary condition $x(0)=0$) or a Brownian Bridge ($x(0)=x(\ell)=0$). Recall here the specification of the covariance operators $\mathcal{C}^{bm}$, $\mathcal{C}^{bb}$ and Cameron-Martin spaces $\mathcal{H}_0^{bm}$, $\mathcal{H}_0^{bb}$ of a Brownian motion and Brownian bridge respectively in Section 2. We make the following definitions, for the relevant range of subscripts:

$$\begin{aligned}\alpha_m = \frac{\partial\alpha}{\partial x(u_m)}(x(u_1),x(u_2),\ldots,x(u_M))\ ;\quad \beta_l = \frac{\partial\beta}{\partial I_l}(I_1,I_2,\ldots,I_L)\ ;\\ \gamma_j = \frac{\partial\gamma}{\partial S_j}(S_1,S_2,\ldots,S_J)\ ;\quad z_l' = \frac{\partial z_l}{\partial x}\ ;\quad r_j' = \frac{\partial r_j}{\partial x}\ .\end{aligned}$$

Remark 5.1. *With a somewhat abuse of notation, path-elements $\{\mathcal{C}^{bm}\delta\Phi(x)\}$, $\{\mathcal{C}^{bb}\delta\Phi(x)\}$ found in Proposition 5.1 below are obtained (at least for the terms*

in $\Phi(x)$ involving stochastic integrals) by recognising that the finite-difference N-dimensional algorithm used in practice corresponds to applying the finite-difference scheme on the Hilbert-space-valued algorithm employing precisely the shown paths $\{\mathcal{C}^{bm}\delta\Phi(x)\}$, $\{\mathcal{C}^{bb}\delta\Phi(x)\}$ within it's specification. (So, here $\delta\Phi(x)$ corresponds to a variational derivative only formally.) This remark applies also to a similar result shown in Proposition 7.1 in a subsequent section.

Proposition 5.1. *For the functional $\Phi(x)$ given in (45), for any $x \in \mathcal{H}$:*

$$\begin{aligned}
\big(\mathcal{C}^{bm}\delta\Phi(x)\big)(u) &= \sum_{m=1}^{M} \alpha_m \cdot \big(u\,\mathbb{I}\,[\,u < u_m\,] + u_m\,\mathbb{I}\,[\,u \geq u_m\,]\,\big) \\
&+ \sum_{l=1}^{L} \beta_l \cdot \Big(u \int_0^{\ell} z_l'(v, x(v))dv - \int_0^u \int_0^s z_l'(v, x(v))dv\,ds \Big) \\
&+ \sum_{j=1}^{J} \gamma_j \cdot \Big(u\,\big(r_j(\ell, x(\ell)) + \int_0^{\ell} dq_j(v)\,\big) - \int_0^u \int_0^s dq_j(v)\,ds \Big)\ , \quad u \in [0, \ell]\ ,
\end{aligned}$$

for the integrator

$$dq_j(v) = r_j'(v, x(v))dx(v) - dr_j(v, x(v))\ .$$

Also:

$$\begin{aligned}
\big(\mathcal{C}^{bb}\delta\Phi(x)\big)(u) &= \sum_{m=1}^{M} \alpha_m \cdot \big(u\,\mathbb{I}\,[\,u < u_m\,] + u_m\,\mathbb{I}\,[\,u \geq u_m\,] - u\,u_m/\ell\,\big) \\
&+ \sum_{l=1}^{L} \beta_l \cdot \Big(\frac{u}{\ell} \int_0^{\ell} \int_0^s z_l'(v, x(v))dv\,ds - \int_0^u \int_0^s z_l'(v, x(v))dv\,ds \Big) \\
&+ \sum_{j=1}^{J} \gamma_j \cdot \Big(\frac{u}{\ell} \int_0^{\ell} \int_0^s dq_j(v) - \int_0^u \int_0^s dq_j(v)\,ds \Big)\ , \qquad u \in [0, \ell]\ .
\end{aligned}$$

Proof. See the Appendix. □

Thus, in both cases the first terms appearing in the specification of $\{\mathcal{C}^{bm}\delta\Phi(x)\}$, $\{\mathcal{C}^{bb}\delta\Phi(x)\}$ in the proposition, are continuous and piece-wise linear in u (there is a turn at the time instances of the observations) so still lie within the Cameron-Martin spaces $\mathcal{H}_0^{bm}$, $\mathcal{H}_0^{bb}$ respectively (even if the variational derivative $\delta\alpha$ itself will not necessarily lie within the Hilbert space, as shown in the proof). The second terms are clearly a.s. elements of the corresponding spaces $\mathcal{H}_0^{bm}$, $\mathcal{H}_0^{bb}$ under weak continuity conditions on z_l'. Finally, for the third terms, again weak regulatory conditions on r_j and r_j' guarantee that the corresponding paths in u are elements of the appropriate Cameron-Martin spaces.

5.2. Example Models

Diffusions Observed at a Discrete Skeleton

We first consider the case of discretely observed diffusions. In other words, suppose that the diffusion V is observed at times $\{u_i,\ i = 0, 1, \ldots, M\}$, with $u_0 = 0$ and $u_M = \ell$, and the observations are denoted with:

$$Y = \{Y_i = V_{u_i},\ i = 0, 1, \ldots, m\}\ .$$

The observations on the transformed unit-volatility diffusion X, defined in (42), are then:

$$\dot{Y} = \{\dot{Y}_i = X_{u_i} = \eta(Y_i; \theta),\ i = 0, 1, \ldots, M\}\ .$$

The reparametrisation of [36] involves another transformation that operates on each of the m independent constituent bridges. Between observations u_i and u_{i+1} we define $\tilde{X}_u$ as

$$\tilde{X}_u = X_u - \frac{(u_{i+1} - u)\dot{Y}_i + (u - u_i)\dot{Y}_{i+1}}{u_{i+1} - u_i}, \quad u_i \leq u \leq u_{i+1}\ ,$$

so that $X_u = b(\tilde{X}_u; \theta)$ with $b(\cdot; \theta)$ defined as the inverse of the above linear mapping. For ease of exposition, we further assume that an antiderivative of $\nu(\cdot; \theta)$ exists. Then, we can write the target log-density $\log((d\Pi/d\Pi_0)(\tilde{x}|\dot{y}_i, \dot{y}_{i+1}, \theta))$ for the path $\tilde{X}$ between u_i and u_{i+1} as (up to an additive normalising constant):

$$-\tfrac{1}{2}\int_{u_i}^{u_{i+1}} \left[\nu'\big(b(\tilde{x}(s); \theta); \theta\big) + \nu^2\big(b(\tilde{x}(s); \theta); \theta\big)\right] ds\ , \tag{46}$$

where $\nu'(x; \theta)$ denotes the derivative of $\nu(x; \theta)$ with respect to x and Π_0 is the distribution of a standard Brownian bridge (note we used Itô's Lemma in (46) to obtain an equavalent expression for the log-density that does not involve stochastic integral, though this is not a requirement for our samplers).

Diffusions observed with error

In this observation regime the data form a discrete skeleton of the diffusion path in the presence of measurement error. V is observed at the time instances $\{u_i,\ i = 1, \ldots, M\}$, with $u_M = \ell$, with observations $Y = \{Y_i,\ i = 1, \ldots, M\}$. We consider the case of independent observations conditional on the diffusion process and work with the unit volatility diffusion X, such that $V_u = \eta^{-1}(X_u; \theta)$. The probability density of the data given the latent path X is:

$$p\big(y|\eta^{-1}(x; \theta); \theta\big) = \prod_{i=1}^{M} f\left(y_i|\eta^{-1}(x(u_i); \theta); \theta\right)\ , \tag{47}$$

where f denotes the likelihood of the data given V, θ. From, (47), (43) the target log-density $\log((d\Pi/d\Pi_0)(x|y, \theta))$ is (up to an additive normalising constant):

$$\sum_{i=1}^{M}\Big\{\log f\big(y_i|\eta^{-1}(x(u_i); \theta); \theta\big) + \int_{u_i}^{u_{i+1}} \nu(x(s); \theta)dx(s) - \frac{1}{2}\int_{u_i}^{u_{i+1}} \nu^2(x(s); \theta)ds\Big\}$$

where now Π_0 denotes the distribution of a standard Brownian motion.

Stochastic volatility models

Stochastic volatility models are popular in modelling financial and econometric time series and instruments such as asset prices, interest rates and financial derivatives. In standard stochastic volatility models for asset prices, such as those in [26] and [25], the log-price S is a diffusion whose volatility is driven by another diffusion. The SDE of the bivariate diffusion model has the following form

$$\begin{cases} dS_u = \mu_s(V_u;\theta)du + \rho\sigma_s(V_u;\theta)dW_u + \sqrt{1-\rho^2}\,\sigma_s(V_u;\theta)dB_u \ ; \\ dV_u = \mu_v(V_u;\theta)du + \sigma_v(V_u;\theta)dW_u \ , \end{cases} \tag{48}$$

where B and W are independent Brownian motions and ρ reflects the leverage effect. For ease of exposition, we set $\rho = 0$. Stochastic volatility models with leverage effect are still into the framework of this paper and can be handled with the methodology described in Section 7. The Lamperti transform can be applied on V to obtain the unit volatility diffusion X as in (42).Consider a pair of observations (y_0, y_1) denoting the value of S at times 0, u_1 respectively. The distribution of the data conditional on the path of X from 0 to u_1 has a known closed form and we can write:

$$\begin{cases} y_1|y_0, x \ \sim \ N\left(M_y(x;\theta), V_y(x;\theta)\right) \ ; \\ M_y(x;\theta) = y_0 + \int_0^{u_1} \mu_s(v(u);\theta)du \ ; \\ V_y(x;\theta) = \int_0^{u_1} \sigma_s(v(u);\theta)^2 du \ ; \\ x(u) = \eta^{-1}(v(u);\theta) \ ; \\ dX_u = \nu(X_u;\theta)du + dW_u \ , \quad 0 \le u \le u_1 \ . \end{cases} \tag{49}$$

Cases with more observations are handled by splitting the data into pairs of consecutive points as above and then, using the Markov property for the bivariate diffusion (S, X), multiplying the corresponding densities calculated according to (49). The log-density $\log((d\Pi/d\Pi_0)(x|y,\theta))$ for the latent diffusion X that drives the volatility for $0 < u \le u_1$ is (up to an additive normalised constant):

$$-\frac{\log V_y(x;\theta)}{2} - \frac{[y_1 - M_y(x;\theta)]^2}{2V_y(x;\theta)} + \int_0^{u_1} \nu(x(s);\theta)dx(s) - \frac{1}{2}\int_0^{u_1} \nu^2(x(s);\theta)ds \ ,$$

where Π_0 denotes here standard Brownian motion.

Latent Diffusion Survival models

Survival models target the probability of an individual i surviving up to time u or else $P(Y > u)$, where Y denotes the event time. The aim is to model the hazard function $h(u)$ that reflects the probability that an event will occur in the infinitesimal period $[u, u + du)$ retrieving information from available data in the form of event times. Latent diffusion survival models [1, 35] provide parametric

formulations for $h(u)$, which is assumed to be a positive function $h(\cdot)$ of a diffusion process. The motivation in such models is to consider an underlying process that results in the occurrence of each event [1]. The distribution function for a single observation y_i is given by

$$F\big(y_i|x,\theta\big) = 1 - \exp\big(-\int_0^{y_i} h(x(s);\theta)ds\big) \ , \quad 0 < y_i \leq \ell \ ,$$

with density

$$f\big(y_i|x,\theta\big) = h(x(y_i);\theta)\exp\Big(-\int_0^{y_i} h(x(s);\theta)ds\Big) \ , \quad 0 < y_i \leq \ell \ ,$$

where $x(u)$ obeys a diffusion process. For ease of exposition we assign the unit diffusion defined by (42) although other choices are possible. The likelihood for the observed event times $y = (y_1, \ldots, y_n)$, with $\max_i y_i \leq \ell$, can be written as:

$$f\big(y|x,\theta\big) = \Big[\prod_{i=1}^{n} h(x(y_i);\theta)\Big]\exp\Big(-\sum_{i=1}^{n}\int_0^{y_i} h(x(s);\theta)ds\Big) \ . \tag{50}$$

Hence, the log-density $\log((d\Pi/d\Pi_0)(x|y,\theta))$ for the latent diffusion X becomes (up to an additive normalising constant):

$$\sum_{i=1}^{n}\Big\{\log h\left(x(y_i);\theta\right) - \int_0^{y_i} h(x(s);\theta)ds\Big\} + \int_0^{\ell}\nu(x(s);\theta)dx(s) - \frac{1}{2}\int_0^{\ell}\nu^2(x(s);\theta)ds$$

with Π_0 denoting the distribution of a standard Brownian motion as before. For more information about such models, including cases of censored data, the reader is referred to [35].

6. Numerical Applications

In this section, we employ the algorithms of Table 3 to perform various simulation experiments involving diffusion bridges, stochastic volatility and latent diffusion survival models. In all these experiments, we treat the parameter vector θ as known and focus on the update of the latent diffusion path. The aim is to assess and compare the performance of the algorithms on various aspects including efficiency of the MCMC output and central processor unit (CPU) time. In order to measure CPU time in two different computing environments, the simulations for diffusion bridges and stochastic volatility models were carried out in MATLAB, whereas for the latent diffusion survival models the `C` programming language was used.

The draws of each algorithm are given by a Markov chain with the target posterior as equilibrium marginal distribution. After a suitable burn-in period,

inference is based on these sequences that exhibit an amount serial correlation and are therefore not as efficient as a posterior sampler with i.i.d. draws. This dependence can be measured by the so-called inefficiency factor, or else autocorrelation time

$$INF(K) = 1 + 2\sum_{k=1}^{K} \gamma(k),$$

where $\gamma(k)$, is the autocorrelation at lag k and K is a suitably chosen truncation point. Loosely speaking the MCMC sampler must be run INF(K) times as many iterations, after the burn-in period, to match the variance of the posterior estimate obtained from a hypothetical independent posterior sampler. In this paper we focus on the inverse of INF(K), known as the effective sample size (ESS), to measure the relative efficiency of the proposed MCMC samplers. The use of INF(K) or ESS(K) is frequent in the MCMC literature; see e.g. [19] and, more specific to the context of this paper, [12] and [22] for applications on stochastic volatility models and Hybrid Monte Carlo respectively. In [22], in order to summarize the inefficiencies for all the model parameters, the minimum ESS(K) was used that was taken over a number of the univariate MCMC trajectories. In our context, the MCMC performance is assessed by monitoring the posterior draws of the diffusion, recorded at a fine partition of its path, and reporting the minimum ESS(K) over these points. By repeated applications of the algorithm, we set both K and the number of iterations of the MCMC algorithm to high enough values so that the minimum ESS(K) stabilizes for all algorithms. We therefore suppress the notation to ESS for the remainder of the paper. The value of ESS was multiplied by a factor of 100 to reflect the percentage of the total MCMC iterations that can be considered as independent draws from the posterior.

The employed MCMC algorithms consist of an independence sampler proposing from the reference Brownian path Π_0 and the advanced algorithms in Table 3. The algorithms were tuned to achieve certain acceptance probability levels that, according to our experience and previous literature, are associated with better performance. Specifically we aimed in attaining an acceptance probability around (15% - 30%) for RWM, (50%-70%) for MALA and (65%-85%) for HMC. To explore the performance of HMC we first fixed the number of leapfrog steps (e.g. to 5 or 10) and then recorded the minimum ESS for various levels of acceptance probability. We then considered cases with additional leapfrog steps. For each of these algorithms, we monitor the values of the minimum ESS, CPU times and their ratio in absolute and relative terms. The presented results contain the best version of these algorithms.

6.1. Diffusions Observed at a Discrete Skeleton

Consider the diffusion discussed in Section 4, i.e. an OU process with SDE:

$$dX_u = -\kappa X_u du + dB_u \ , \quad 0 \leq u \leq \ell \ ,$$

with $X_0 = 0$ and an observation at time $\ell = 1$. We set $X_1 = 0$ and consider 3 different values for κ, i.e. $12, 20, 30$ in our investigation of the MCMC performance.

The algorithms were constructed using the Euler-Maruyama approximation of the target density (46) from a time-discretised diffusion path. The MCMC components comprise of the equidistant points from a discrete skeleton of the diffusion. The discretisation step was set to $\delta = 0.02$. Table 4 provides the results, i.e. the values of the minimum ESS, CPU times and their absolute and relative ratio. The HMC algorithm consisted of 5 leapfrog steps with the parameter h set to values $(0.43, 0.26, 0.17)$ for values of κ $(12, 20, 30)$ respectively. For advanced MALA, that can be thought as HMC with a single leapfrog step, the corresponding values of $h = \sqrt{\Delta t}$ were very similar $(0.45, 0.26, 0.18)$ indicating much smaller total steps. Overall, advanced HMC consistently over-performs, in terms of ESS, the remaining algorithms. In particular for $\kappa = 30$, HMC is faster than the independence sampler by a factor of over 30. Its performance remains at high levels as we increase κ and does not deteriorate as δ becomes smaller, as indicated by the results obtained for $\delta = 0.01$ and $\delta = 0.005$. In line with the results of [39] and Section 4, we note a substantial improvement over advanced MALA suggesting a more efficient use of the gradient within HMC. MALA offers some improvement over RWM and independence sampler, but at a heavy additional computational cost. The independence sampler performs reasonably well for $\kappa = 12$ (acceptance rate of 16%) but its performance drops substantially as κ increases and the acceptance rate becomes smaller; 8% for $\kappa = 20$ and 1.2% for $\kappa = 30$.

6.2. Stochastic Volatility Models

The following stochastic volatility model was used to simulate data:

$$\begin{cases} dS_u = \exp(V_u/2)dB_u \ , \quad 0 \leq u \leq \ell \ ; \\ dV_u = \kappa(\mu - V_u)du + \sigma dW_u \ . \end{cases}$$

The parameters were set according to previous analyses based on similar models for the S&P 500 dataset [12]. Specifically, we set $\kappa = 0.03$, $\mu = 0.07$, $\sigma^2 = 0.03$ and $V_0 = 0$. We consider about a year measured in days ($\ell = 250$) and recorded observations at a daily frequency (250 data points). The transformation of V_u to a unit volatility diffusion was used to write the target density and construct the HMC algorithms. The model for a pair of consecutive observations, (y_{i-1}, y_i) can be written as:

$$\begin{cases} y_i | y_{i-1} \ \sim \ N\big(y_{i-1}, \ \int_{u_{i-1}}^{u_i} \exp(\sigma x(s))ds\big) \ ; \\ x(u) = v(u)/\sigma \ ; \\ dX_u = \kappa \left(\frac{\mu}{\sigma} - X_u\right) du + dW_u \ , \quad u_{i-1} \leq u \leq u_i \ . \end{cases}$$

The results are shown in Table 5. Independence sampler performs very poorly in this case, with an acceptance rate below 10^{-4}, and is omitted from the table. MALA provides a small improvement over RWM which is nevertheless not enough to cover the associated increase in the corresponding computations.

$\kappa = 12$	min(ESS)	time	$\frac{\text{min(ESS)}}{\text{time}}$	relative $\frac{\text{min(ESS)}}{\text{time}}$
IS	3.9173	4.8811	0.8025	2.1733
RWM	3.9584	5.8925	0.6718	1.8192
MALA	4.0112	10.8626	0.3693	1
HMC	35.7274	20.8695	1.7119	4.6361
HMC ($\delta = 0.01$)	35.8903	32.5594	N/A	N/A
HMC ($\delta = 0.005$)	35.5875	51.6085	N/A	N/A
$\kappa = 20$	min(ESS)	time	$\frac{\text{min(ESS)}}{\text{time}}$	relative $\frac{\text{min(ESS)}}{\text{time}}$
IS	0.5013	4.4977	0.1115	1
RWM	1.0086	5.4445	0.1853	1.6621
MALA	1.6202	10.0588	0.1611	1.4452
HMC	26.6214	20.8841	1.2747	11.4369
$\kappa = 30$	min(ESS)	time	$\frac{\text{min(ESS)}}{\text{time}}$	relative $\frac{\text{min(ESS)}}{\text{time}}$
IS	0.1012	4.7043	0.0215	1
RWM	0.4343	5.7229	0.0759	3.5277
MALA	0.5372	10.0438	0.0535	2.4863
HMC	13.3350	20.4831	0.6510	30.2631

Table 4: Relative efficiency via the minimum ESS (%) and CPU times (seconds), for the advanced pathspace algorithms - Case of OU bridges. IS denotes the Independence Sampler.

However, this is not the case for HMC that reaches its optimal performance roughly at 10 leapfrog steps. Advanced HMC offers considerable improvement, being nearly 8 times faster than RWM and 11 times faster than MALA. Parameter h that corresponded to the desired acceptance probability levels was 0.085 for MALA algorithm and 0.075 for all the versions of HMC.

6.3. Latent Diffusion Survival models

This section provides a numerical illustration in simulated data from a latent diffusion survival model appearing in [35]. Specifically, the likelihood for event times $y = \{y_1, \ldots, y_n\}$ is given by:

$$p\big(y|\eta^{-1}(x;\theta);\theta\big) = \Big[\prod_{i=1}^{n} x^2(y_i)\Big] \exp\Big(-\sum_{i=1}^{n}\int_0^{y_i} x^2(s)ds\Big)$$

where we have considered the model:

$$dX_u = -(1.4\sin(X_u)+1)du + dB_u \ , \quad X_0 = 2 \ . \tag{51}$$

We simulated a trajectory driving the hazard function from (51) that can be seen in Figure 1 (solid red line). Despite various fluctuations, the simulated

Sampler	min(ESS)	time	$100\times \frac{\text{min(ESS)}}{\text{time}}$	relative $\frac{\text{min(ESS)}}{\text{time}}$
RWM	0.1400	161.8298	0.0865	1.3561
MALA	0.2181	341.8737	0.0638	1.0000
HMC (5 steps)	2.5695	689.6767	0.3726	5.8400
HMC (10 steps)	8.1655	1188.1201	0.6873	10.7729
HMC (20 steps)	8.3216	2200.1311	0.3782	5.9288

Table 5: Relative efficiency, via the minimum ESS (%) and CPU times (seconds) for the diffusion pathspace algorithms - Case of stochastic volatility paths.

Sampler	min(ESS)	time	$100\times \frac{\text{min(ESS)}}{\text{time}}$	relative $\frac{\text{min(ESS)}}{\text{time}}$
RWM	0.1039	55.2342	0.1881	1
MALA	0.6466	87.5021	0.7389	3.9284
HMC	25.2985	248.0301	10.1997	54.2229

Table 6: Relative efficiency, via the minimum ESS (%) and CPU times (seconds) for the advanced pathspace algorithms - Case of latent diffusion survival model.

hazard is decreasing and could be used, for example, in cases of lifetime data from patients following a surgical operation. The trajectory of Figure 1 was then used, through the distribution function it defines, to draw 200 i.i.d. event times. In the simulation exercise of this section, we treat the parameters of the diffusion as known and apply the various MCMC algorithms of Table 3 to sample from the posterior of the latent diffusion process. The SDE in (51) can also be thought as a prior for the hazard function trajectory. Our focus is the behaviour and efficiency of the proposed MCMC algorithms, while keeping an eye on the ability of the model to capture the shape of the unobserved hazard.

Table 6 provides the measures of performance of the algorithms in Table 3. The calculations in this section were obtained using `C` programming language, unlike the previous two applications where MATLAB was used. Similarly with the stochastic volatility simulation experiment, the independence sampler is associated with extremely low acceptance rate rendering it infeasible in practice. RWM also performs poorly. A very small step is required to achieve the desired acceptance rate, thus resulting in very small moves around the diffusion pathspace. MALA with $h = 0.2$ performs better in this case, but massively better performance is offered by advanced HMC. Specifically, HMC with 10 leapfrog steps and $h = 0.15$ is about 54 times faster than RWM. As already mentioned, Figure 1 depicts the trajectory of the latent diffusion process that determines the hazard function and was used to generate the data. It also displays 95% pointwise credible intervals, obtained from the HMC algorithm, which indicate that the shape of the hazard function is captured reasonably well.

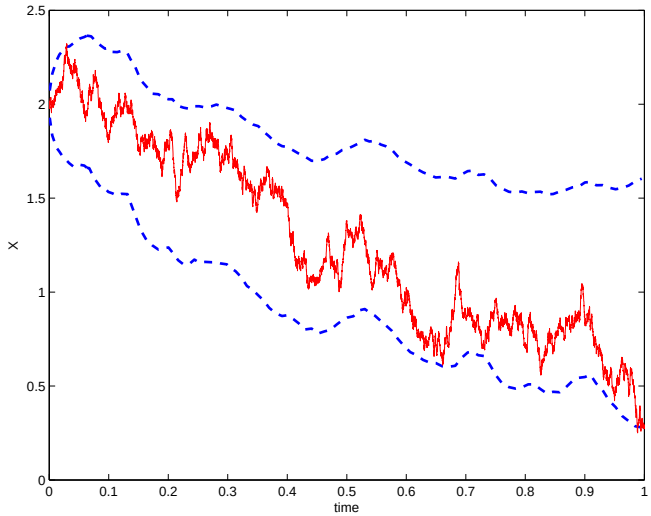

Figure 1: 95% Pointwise credible intervals (blue dashed lines) overlayed on the true path of X (red solid line).

7. Robustness in Dimension in Further Scenarios

We will briefly illustrate the well-definition of advanced HMC algorithms for cases when the target distribution of the SDE-driven models can have a different structure from the ones considered so far. In what follows, we omit reference to the parameter θ.

7.1. Beyond the Lamperti Transform

In the context of a non-scalar diffusion V_u, defined via the equation (2), it is not guaranteed at all that V can be transformed into an SDE of unit diffusion coefficient. Indeed, the Lamperti transform in such a context would require the existence of a mapping $V_u \mapsto \eta(X_u)$ (with $\eta = (\eta_1, \eta_2, \ldots, \eta_d)^\top : \mathbb{R}^d \mapsto \mathbb{R}^d$) such that, for all v in the state space of V_t:

$$D\eta(v) \cdot \sigma(v) = I_d \tag{52}$$

where $D\eta = (\partial\eta_i/\partial v_j)$. This follows directly by Itô's formula, see e.g. [2]. The work in [2] also shows that existence of a mapping η with the property (52) is equivalent to the diffusion coefficient matrix satisfying $\partial\sigma_{ij}^{-1}/\partial v_k = \partial\sigma_{ik}/\partial v_j$ for all i, j, k with $j < k$. This restricts considerably the applicability of the Lamperti transform for non-scalar diffusions. There are however other methods suggested in the literature of a wider scope. Our advanced MCMC algorithms require posterior distributions on pathspace which are absolutely continuous w.r.t. Brownian motion-related distributions, so it is of interest to briefly verify their well-definition on the pathspace when using such alternative transforms.

The method in [12] maps V_u onto the driving Wiener noise $X_u = B_u$ of the SDE. Assuming some data Y with likelihood $p(y|v)$, and since the prior on X is simply the Wiener measure $\Pi_0 = N(0, \mathcal{C}^{bm})$, we can write the posterior on X as:

$$\frac{d\Pi}{d\Pi_0}(x) \propto p(y|v) \ .$$

(Note, that we use here the mapping $x \mapsto v$ between driving noise x and process v implied by the expression of the SDE for the process V.) It remains to calculate $\mathcal{C}^{bm}\,\delta\Phi(x)$ is this context. Differentiation of $\Phi(x)$ will involve finding derivatives of v w.r.t. the driving noise x. So it is not a surprise that the dynamics of the Malliavin derivative $D_s V_u$ (see e.g. [18]) will appear in the calculations; this is defined as below:

$$\tfrac{dY_u}{Y_u} = \mu'(V_u)du + \sigma'(V_u)\,dB_u \ , \quad Y_0 = 1 \ ;$$
$$D_s V_u = \tfrac{Y_u}{Y_s}\,\sigma(V_s)\,\mathbb{I}\,[\,s \le u\,] \ .$$

We will assume here the general structure for $\Phi(x) = -\log(p(y|v))$ (compared with the structure assumed earlier in (45) we do not include here stochastic integral terms to avoid overly cumbersome expressions):

$$\begin{aligned}\Phi(x) =& \alpha(v(u_1), v(u_2), \ldots, v(u_M)) \\ &+ \beta\Big(\int_0^\ell z_1(s, v(s))ds, \int_0^\ell z_2(s, v(s))ds, \int_0^\ell z_L(s, v(s))ds\Big) \ . \end{aligned} \tag{53}$$

The terms α_m, β_l appearing below correspond to partial derivatives of the functionals α, β as in the case of Proposition 5.1.

Proposition 7.1. *For the functional $\Phi(x)$ given in (53), for any $x \in \mathcal{H}$:*

$$\begin{aligned}\big(\mathcal{C}^{bm}\delta\Phi(x)\big)(u) =& \sum_{m=1}^{M} \alpha_m \cdot \Big((u \wedge u_m)\,(F_{m,u_m} + \sigma(v(u_m))) - \int_0^{u\wedge u_m} F_{m,r}dr \Big) \\ &+ \sum_{l=1}^{L} \beta_l \cdot \Big(u\,(G_{l,\ell} + J_{l,\ell}) + \int_0^u (G_{l,r} + J_{l,r})dr \Big) \ , \quad u \in [0, \ell] \ ,\end{aligned}$$

for the processes, for $m = 1, 2, \ldots, M$ and $l = 1, 2, \ldots, L$:

$$F_{m,r} = \int_0^r e^{\int_s^{u_m} \left(\mu'(v(u))du + \sigma'(v(u))dx(u)\right)}\, dQ_s \ ;$$
$$G_{l,r} = \int_0^r \int_s^\ell z_l'(t, v(t))\, e^{\int_s^t \left(\mu'(v(u))du + \sigma'(v(u))dx(u)\right)} dt\, dQ_s \ ;$$
$$J_{l,r} = \int_0^r z_l'(s, v(s))\, \sigma(v(s))ds \ ,$$

with integrator:

$$dQ_s = \sigma(v(s))(\mu'(v(s))ds + \sigma'(v(s))dx(s)) - d\sigma(v(s)) \ .$$

Focusing on the properties of the calculated path $\mathcal{C}^{bm}\delta\Phi(x)$ over its domain $u \in [0, \ell]$, it is easy to check the following: a.s. in x, the first terms are continuous, piecewise linear with a turn at the data instances $u_1, u_2, \ldots, u_M$; then, under the weak assumption that the processes $r \mapsto G_{l,r}$, $r \mapsto J_{l,r}$ are a.s. continuous, we have that the second terms in the calculation in Proposition 7.1 are a.s. differentiable. Thus, under weak conditions Assumption 3.1 is satisfied and advanced HMC is well-defined on the pathspace.

8. Discussion - Further Directions

In this paper we developed and applied a general class of derivative driven algorithms, suitable for Hilbert spaces defined by diffusion driven models. The validity of the algorithm was established using different theory than [6], which allowed us to relax involved assumptions and generalize to widely used diffusion models across various scientific fields as mentioned in Section 1.1. The modelling framework contains diffusions observed with error, partial data, as well as observation on functionals of the diffusion process such as hitting times and integrals thereof. The extended framework also includes diffusions processes with general diffusion coefficients that can be handled with reparametrisation to the driving Wiener noise as in [12] and [24]. A number of research directions related with the methods presented in this paper have remained open:

- The algorithm can be used in contexts where a Gibbs data augmentation scheme is adopted to facilitate the step of updating the diffusion path given the parameters. The entire diffusion path can either be updated in a single step, but can also be integrated to updates based on overlapping and potentially random-sized blocks. Note that while this approach will boost the conditional step of the diffusion path updates, another important aspect for the overall performance of the MCMC sample is the posterior correlations between diffusion paths and parameters. The role of reparametrisation of diffusion paths is critical for this task. We are currently investigating extensions of the advanced HMC algorithm on the joint density of diffusion and parameters. In the numerical applications of this paper the high dimensional diffusion path was successfully updated in a single block, a fact that provides supporting evidence that the algorithm could still be very efficient when also incorporating the low dimension parameter vector θ in it. Such an approach will provide an alternative to pseudo marginal [4] or particle MCMC [3] formulations for diffusion driven models; see e.g. [37] and [16] respectively.

- In this paper the choice of the mass matrix for the Hamiltonian dynamics is guided by the prior, i.e. by the reference Gaussian law $N(0, \mathcal{C})$. Recent work in the literature ([23]) has looked at exploiting the geometric structure of state-space to consider location-specific mass matrices. It is therefore worth investigating the choice of location-specific mass matrices in the context of infinite-dimensional pathspaces to further boost the efficiency of MCMC methods.

- An important direction of further research is related with the calibration of long-memory diffusion models: an important class covers the case of processes driven by fractional Brownian motion (see e.g. [38, 31]). Applying standard Independence samplers over smaller blocks of the complete latent path will be overly expensive as calculations have to be done every time over the whole path due to the lack of Markovianity. In this case it could be very beneficial to update the whole path together using HMC and control the acceptance probability by trying different leapfrog steps.

Acknowledgments

We are grateful to two referees for many comments that greatly improved the appearance of the paper. We also thank Dr Yvo Pokern for a lot of useful advices during the development of the paper. EP acknowledges support from EPSRC for his PhD studies.

Appendix A. Proofs

Proof of Proposition 4.1. We use the notation $|\cdot|_{L_1} \equiv \mathrm{E}\,|\cdot|$. For both algorithms (RWM and MALA) the acceptance probability can be expressed as:

$$a(x,v) = 1 \wedge e^{R(x,v)}$$

for the relevant variable $R(x,v)$. Working as in [39], for the first result it suffices to prove that $\sup_\ell |R(x,v)|_{L_1} < \infty$ since we have the inequality:

$$\mathrm{E}\,[\,1 \wedge e^R\,] \geq \mathrm{E}\,[\,e^{-|R|}\,] \geq e^{-|R|_{L_1}} \ .$$

(We thank a referee for pointing at the above inequality, which replaced a slightly more complex expression we originally had.) For the second result, we will be identifying a term in R with the highest order L_1-norm, which we will denote by J. Crucially, this term will lie in $(-\infty, 0)$, thus will be growing to $-\infty$ faster than the growth of $|R-J|_{L_1}$. This provides some intuition for the zero limit of the average acceptance probability. Analytically, as in [39], we will be using the following inequality, for any $\gamma > 0$:

$$\begin{aligned}
\mathrm{E}[\,1 \wedge e^R\,] &\leq \mathrm{P}[\,R \geq -\gamma\,] + e^{-\gamma} \\
&= \mathrm{P}\,[\,\{R \geq -\gamma\} \cap \{|R-J| \leq \gamma\}\,] + \mathrm{P}\,[\,\{R \geq -\gamma\} \cap \{|R-J| > \gamma\}\,] + e^{-\gamma} \\
&\leq \mathrm{P}\,[\,J \geq -2\gamma\,] + \mathrm{P}\,[\,|R-J| > \gamma\,] + e^{-\gamma} \\
&\leq \mathrm{P}\,[\,J \geq -2\gamma\,] + \tfrac{|R-J|_{L_1}}{\gamma} + e^{-\gamma}\ , \qquad \text{(A.1)}
\end{aligned}$$

where we used the Markov inequality in the last line. So, the idea here will be to choose some γ growing to ∞ faster that the rate of growth of $|R-J|_{L_1}$ and slower than the rate of growth of $|J|_{L_1}$.

From [39], we retrieve the following collection of results:

$$\mathrm{E}\,|x|^2 = \tfrac{\ell}{2\kappa\tanh(\kappa\ell)} - \tfrac{1}{2\kappa^2}\ ; \quad \mathrm{E}\,|\mathcal{C}x|^2 = \tfrac{\ell^4}{90\kappa^2} - \tfrac{1}{2\kappa^6} - \tfrac{\ell^2}{6\kappa^4} + \tfrac{\ell}{2\kappa^5\,\tanh(\kappa\ell)}\ ;$$
$$\mathrm{E}\,[\,\langle x,\mathcal{C}x\rangle\,] = \tfrac{3+\kappa^2\ell^2}{6\kappa^4} - \tfrac{\ell}{2\kappa^3\tanh(\kappa\ell)}\ ; \tag{A.2}$$
$$\mathrm{E}\,[\,\langle x,\mathcal{C}^3x\rangle\,] = \tfrac{945+315\kappa^2\ell^2-21\kappa^4\ell^4+2\kappa^6\ell^6}{1890\kappa^8} - \tfrac{\ell}{2\kappa^7\tanh(\kappa\ell)}\ ;$$
$$\mathrm{E}\,|v|^2 = \tfrac{\ell^2}{6}\ ; \quad \mathrm{E}\,[\,\langle v,\mathcal{C}v\rangle\,] = \tfrac{\ell^4}{90}\ ;$$
$$\mathrm{E}\,[\,\langle x,v\rangle^2\,] = \tfrac{3+\kappa^2\ell^2}{6\kappa^4} - \tfrac{\ell}{2\kappa^3\tanh(\kappa\ell)}\ ; \quad \mathrm{E}\,[\,\langle \mathcal{C}x,v\rangle^2\,] = \tfrac{945+315\kappa^2\ell^2-21\kappa^4\ell^4+2\kappa^6\ell^6}{1890\kappa^8}\ ;$$
$$\mathrm{E}\,[\,\langle \mathcal{C}^2x,v\rangle^2\,] = \tfrac{467775+155925\kappa^2\ell^2-10395\kappa^4\ell^4+990\kappa^6\ell^6-99\kappa^8\ell^8+10\kappa^{10}\ell^{10}}{467775\kappa^{12}}\ . \tag{A.3}$$

For the case of RWM, analytical calculations will give:

$$R(x,v) = \tfrac{\kappa^2}{2(1+\frac{\Delta t}{4})^2}\ \Delta t\,\langle x,x\rangle - \tfrac{\kappa^2}{2(1+\frac{\Delta t}{4})^2}\ \Delta t\,\langle v,v\rangle\ - \tfrac{\kappa^2(1-\frac{\Delta t}{4})}{(1+\frac{\Delta t}{4})^2}\ \sqrt{\Delta t}\,\langle x,v\rangle\ .$$

Now, when $\Delta t = c/\ell^2$, from the results in (A.3) we find that the L_1-norm of each of the three above summands is $\mathcal{O}(1)$, so $\sup_l |R(x,v)|_{L_1} < \infty$. When $\Delta t = c/\ell^\epsilon$ with $\epsilon \in (0,2)$, a careful inspection of (A.3) gives that the term with the highest order L_1-norm is:

$$J = -\tfrac{\kappa^2}{2(1+\frac{\Delta t}{4})^2}\ \Delta t\,\langle v,v\rangle\ .$$

Indeed, we have that:

$$|J|_{L_1} = \Theta(\Delta t\,\ell^2)\ ; \quad |R-J|_{L_1} = \mathcal{O}(\sqrt{\Delta t}\,\ell)\ .$$

We will apply (A.1) for the selection $\gamma = (\Delta t\,\ell^2)^{2/3}$. Clearly, the last two terms in the R.H.S. of (A.1) vanish as $\ell \to \infty$. For the first term we need to use the analytical definition of J. Using the rescaling properties of the Brownian bridge, we can write $v_{u\ell} = \sqrt{\ell}\,\tilde{v}_u$ for a standard Brownian bridge $\tilde{v}$ on $[0,1]$. Thus, we can re-write:

$$\langle v,v\rangle = \int_0^\ell v_u^2 du = \ell\int_0^1 v_{u\ell}^2 du \equiv \ell^2\int_0^1 \tilde{v}_u^2\,du = \ell^2\,|\tilde{v}|^2\ .$$

We now have that:

$$\mathrm{P}\,[\,J \geq -2\gamma\,] = \mathrm{P}\,[\,|\tilde{v}|^2 \leq \tfrac{4(1+\frac{\Delta t}{4})^2}{\kappa^2}\ \tfrac{\gamma}{\Delta t\,\ell^2}\,] \to \mathrm{P}\,[\,|\tilde{v}|^2 = 0\,] = 0\ .$$

For the case of MALA, some tedious calculations detailed in [39] will give that:

$$\begin{aligned}
R(x,v) = {} & \tfrac{\kappa^2}{8(1+\frac{\Delta t}{4})^2}\,\Delta t^2\,\langle x,x\rangle - \tfrac{\kappa^6}{32(1+\frac{\Delta t}{4})^2}\,\Delta t^3\,\langle \mathcal{C}x,\mathcal{C}x\rangle - \tfrac{\kappa^4(1-\frac{\Delta t}{4})}{8(1+\frac{\Delta t}{4})^2}\,\Delta t^2\,\langle x,\mathcal{C}x\rangle \\
& + \tfrac{\kappa^2}{8(1+\frac{\Delta t}{4})^2}\,\Delta t^2\langle x,\mathcal{C}x\rangle - \tfrac{\kappa^6}{32(1+\frac{\Delta t}{4})^2}\,\Delta t^3\,\langle \mathcal{C}^2x,\mathcal{C}x\rangle + \tfrac{\kappa^4(1-\frac{\Delta t}{4})}{8(1+\frac{\Delta t}{4})^2}\,\Delta t^2\,\langle \mathcal{C}x,\mathcal{C}x\rangle \\
& - \tfrac{\kappa^2}{8(1+\frac{\Delta t}{4})^2}\,\Delta t^2\,\langle v,v\rangle - \tfrac{\kappa^2}{8(1+\frac{\Delta t}{4})^2}\,\Delta t^2\,\langle v,\mathcal{C}v\rangle - \tfrac{\kappa^2(1-\frac{\Delta t}{4})}{4(1+\frac{\Delta t}{4})^2}\,\Delta t^{3/2}\,\langle x,v\rangle \\
& + \tfrac{\kappa^4}{8(1+\frac{\Delta t}{4})^2}\,\Delta t^{5/2}\,\langle \mathcal{C}x,v\rangle - \tfrac{\kappa^2(1-\frac{\Delta t}{4})}{4(1+\frac{\Delta t}{4})^2}\,\Delta t^{3/2}\,\langle \mathcal{C}x,v\rangle + \tfrac{\kappa^4}{8(1+\frac{\Delta t}{4})^2}\,\Delta t^{5/2}\,\langle \mathcal{C}^2x,v\rangle\ .
\end{aligned}$$

Now, when $\Delta t = c/\ell^2$, from the results in (A.3) we find that the L_1-norm of each of the twelve above summands is indeed $\mathcal{O}(1)$, so $\sup_l |R(x,v)|_{L_1} < \infty$. When $\Delta t = c/\ell^\epsilon$ with $\epsilon \in (0,2)$, using again the results in (A.3), we identify, among the twelve terms in the definition of $R(x,v)$, the one with the highest L_1-norm:

$$J = -\tfrac{\kappa^6}{32(1+\frac{\Delta t}{4})^2}\, \Delta t^3 \, \langle \mathcal{C}^2 x, \mathcal{C} x \rangle \ .$$

Indeed we have that:

$$|J|_{L_1} = \Theta(\Delta t^3 \ell^6) \ ; \quad |R - J|_{L_1} = \mathcal{O}(\Delta t^{5/2} \ell^5) \ .$$

We note that the second result above comes from the bound on the L_1-norm of the last summand in the definition of $R(x,v)$. In this case we apply (A.1) under the choice $\gamma = (\Delta t\, \ell^2)^{11/4}$. As before, under this choice the last two terms on the R.H.S. of (A.1) vanish as $\ell \to \infty$. For, the first term we must use the analytical definition of J from above. To handle J we can use the Karhunen-Loève expansion of Gaussian measures. In particular, using the representation for an OU-bridge from (37) we get that:

$$\langle x, \mathcal{C}^3 x \rangle = \sum_{p=1}^{\infty} \frac{1}{\frac{\pi^2 p^2}{\ell^2} + \kappa^2} \, \frac{\ell^6}{\pi^6\, p^6} \, \xi_p^2 \ .$$

Thus, we can now obtain:

$$\begin{aligned} \mathrm{P}\,[\, J \geq -2\gamma \,] &= \mathrm{P}\, \Big[\sum_{p=1}^{\infty} \tfrac{1}{\frac{\pi^2 p^2}{\ell^2} + \kappa^2} \, \tfrac{1}{\pi^6 p^6} \, \xi_p^2 \leq \tfrac{32(1+\frac{\Delta t}{4})^2}{\kappa^6} \, \tfrac{1}{(\Delta t \ell^2)^{1/4}} \Big] \\ &\leq \mathrm{P}\, \big[\tfrac{1}{\pi^2 + \kappa^2} \, \tfrac{1}{\pi^6} \, \xi_1^2 \leq \tfrac{32(1+\frac{\Delta t}{4})^2}{\kappa^6} \, \tfrac{1}{(\Delta t\, \ell^2)^{1/4}} \big] \to \mathrm{P}\,[\, \xi_1^2 \leq 0 \,] = 0 \ . \end{aligned}$$

The proof is now complete. □

Proof of Proposition 4.2. We exploit the representation of Ψ_h^I in (39). Recall that we denote $(x_i, v_i) = \Psi_h^i(x_0, v_0)$, for number of leapfrog steps $0 \leq i \leq I$. After tedious calculations, one can obtain:

$$\Delta H(x_0, v_0) = H(x_I, v_I) - H(x_0, v_0) \equiv \langle \mathcal{A} x_0, x_0 \rangle + \langle \mathcal{B} v_0, v_0 \rangle + \langle \mathcal{G} x_0, v_0 \rangle \quad \text{(A.4)}$$

for the operators:

$$\begin{aligned} \mathcal{A} = -\tfrac{1}{2}\, \sin^2(I\theta)\, \mathcal{P} \ ; \quad \mathcal{B} = \tfrac{1}{2}\, \sin^2(I\theta)\, a^2\, \mathcal{P} \ ; \quad \mathcal{G} = \tfrac{1}{2}\, \sin(2I\theta)\, a\, \mathcal{P} \ ; \\ \mathcal{P} = \kappa^2\, I + (1 - \tfrac{1}{a^2})\, (\mathcal{C}^{bb})^{-1} \ . \end{aligned} \quad \text{(A.5)}$$

We have used operator's notation: $\sin(I\theta)$ is the operator such that we have $\sin(I\theta)x = \{\sin(I\theta_p)\, x_p\}_{p=1}^{\infty}$ where $\{x_p\}$ are the co-ordinates of $x \in \mathcal{H}$ w.r.t. to the orthonormal basis corresponding to the eigen-functions of $\mathcal{C}^{bb}$; a similar interpretation stands for the operator $\sin(2I\theta)$. Also $a^2 x = \{a_p^2\, x_p\}_{p=1}^{\infty}$. The sequences $\{\theta_p\}$ and $\{a_p\}$ are defined in (40).

We denote here by $\mathcal{C}^{OU}$ the covariance matrix of the target OU bridge; the Karhunen-Loève expansion (37) implies the eigenstructure $\{\lambda_{p,OU}, \phi_p\}_{p=1}^{\infty}$ for $\mathcal{C}^{OU}$ with eigen-values:

$$\lambda_{p,OU} = \frac{1}{\frac{\pi^2\, p^2}{\ell^2} + \kappa^2} \ .$$

Examining the eigenvalues $\{P_p\}_{p=1}^{\infty}$ of $\mathcal{P}$, after some calculations one can verify that:

$$\mathcal{P} \equiv \mathcal{C}_{OU}^{-1}\, \mathcal{C}\, \frac{\kappa^2\, h^2}{(2+\frac{h^2}{2})(1+\rho)} \ ,$$

so that:

$$0 \leq P_p \leq M\, \lambda_{p,OU}^{-1}\, \lambda_p\, \frac{1}{\ell^2} \ , \tag{A.6}$$

for some constant $M > 0$. Again, after some tedious calculations we get that:

$$a_p = \lambda_p^{-1/2} \lambda_{p,OU}^{1/2}\, c_p \ ; \quad c_p^2 := \big(\frac{1}{c^2\, \kappa^2} - \frac{1}{4\, p^2 \pi^2} \big)^{-1}. \tag{A.7}$$

Note that the term on the R.H.S. of this last definition of c_p^2 is guaranteed to be positive due to condition (38). In particular we have that:

$$c_p^2 \leq M \ , \tag{A.8}$$

for a constant $M > 0$. Now, starting from (A.4) we have that:

$$\begin{aligned} \mathrm{E}\,[\,(\Delta H)^2\,] = \mathrm{E}\,[\,\langle \mathcal{A}x_0, x_0\rangle^2\,] + \mathrm{E}\,[\,\langle \mathcal{B}v_0, v_0\rangle^2\,] + \mathrm{E}\,[\,\langle \mathcal{G}x_0, v_0\rangle^2\,] \\ + 2\,\mathrm{E}\,[\,\langle \mathcal{A}x_0, x_0\rangle\,]\,\mathrm{E}\,[\,\langle \mathcal{B}v_0, v_0\rangle\,] \end{aligned} \tag{A.9}$$

as the rest of the expectations will be zero (this follows from the independency of the zero-mean variables x_0, v_0). Henceforth, $\{A_p\}_{p=1}^{\infty}$, $\{B_p\}_{p=1}^{\infty}$, $\{G_p\}_{p=1}^{\infty}$ denote the eigenvalues of the operators $\mathcal{A}$, $\mathcal{B}$, $\mathcal{G}$ respectively. Recalling that $\langle \mathcal{A}x_0, x_0\rangle = \sum_{p=1}^{\infty} A_p x_{0,p}^2$, we have:

$$\begin{aligned} \mathrm{E}\,[\,\langle \mathcal{A}x_0, x_0\rangle^2\,] &= \mathrm{Var}\,[\,\langle \mathcal{A}x_0, x_0\rangle\,] + \mathrm{E}^2\,[\,\langle \mathcal{A}x_0, x_0\rangle\,] \\ &= 2\sum_{p=1}^{\infty} A_p^2\, \lambda_{p,OU}^2 + \big(\sum_{p=1}^{\infty} A_p\, \lambda_{p,OU} \big)^2 \ . \end{aligned}$$

Doing similar calculations also for the rest of the terms on the R.H.S. of (A.9), we have:

$$\begin{aligned} \mathrm{E}\,[\,(\Delta H)^2\,] = 2\sum_{p=1}^{\infty} A_p^2\, \lambda_{p,OU}^2 + 2\sum_{p=1}^{\infty} B_p^2\, \lambda_p^2 + \Big(\sum_{p=1}^{\infty} \big(A_p\, \lambda_{p,OU} + B_p\, \lambda_p \big) \Big)^2 \\ + \sum_{p=1}^{\infty} G_p^2\, \lambda_{p,OU}\, \lambda_p \ . \end{aligned}$$

Thus, using the representations and bounds in (A.5), (A.6), (A.7) and (A.8), we have that:

$$\sum_{p=1}^{\infty} A_p^2\, \lambda_{p,OU}^2 \leq M \sum_{p=1}^{\infty} \lambda_{p,OU}^{-2}\, \lambda_p^2\, \frac{1}{\ell^4}\, \lambda_{p,OU}^2 \;= M \sum_{p=1}^{\infty} \frac{1}{p^4\, \pi^4} < \infty \;.$$

Similar calculations will give:

$$\sum_{p=1}^{\infty} B_p^2\, \lambda_p^2 \leq M \sum_{p=1}^{\infty} \lambda_p^{-2}\, \lambda_{p,OU}^2\, c_p^4\, \lambda_{p,OU}^{-2}\, \lambda_p^2\, \frac{1}{\ell^4}\, \lambda_p^2 \leq M \sum_{p=1}^{\infty} \frac{1}{p^4\, \pi^4} < \infty \;.$$

and:

$$\sum_{p=1}^{\infty} G_p^2\, \lambda_{p,OU}\, \lambda_p \leq M \sum_{p=1}^{\infty} \lambda_p^{-1}\, \lambda_{p,OU}\, c_p^2\, \lambda_{p,OU}^{-2}\, \lambda_p^2\, \frac{1}{\ell^4}\, \lambda_{p,OU}\, \lambda_p \leq M \sum_{p=1}^{\infty} \frac{1}{p^4\, \pi^4} < \infty \;.$$

Finally, we have that:

$$\begin{aligned} \Big|\sum_{p=1}^{\infty} \big(A_p\, \lambda_{p,OU} + B_p\, \lambda_p \big) \Big| &= \tfrac{1}{2} \Big| \sum_{p=1}^{\infty} \sin^2(I\theta_p)\, P_p\, (-\lambda_{p,OU} + a_p^2\, \lambda_p) \Big| \\ &= \tfrac{1}{2} \Big| \sum_{p=1}^{\infty} \sin^2(I\theta_p)\, P_p\, \lambda_{p,OU}\, (c_p^2 - 1) \Big| \\ &\leq M \sum_{p=1}^{\infty} \lambda_p\, \frac{1}{\ell^2} = M \sum_{p=1}^{\infty} \frac{1}{p^2 \pi^2} < \infty \;. \end{aligned}$$

Thus, we have proven that $\sup_l \mathrm{E}\,[\,\Delta H^2\,] < \infty$, which, as illustrated in the proof of Proposition 4.1, is sufficient for our proof.

□

Proof of Proposition 5.1. We will do our calculations using the right-most expressions in (4), (5), where $\mathcal{C}^{bm}$, $\mathcal{C}^{bb}$ are respectively specified. For the first term in the expression for Φ, namely $\alpha = \alpha(x(u_1), x(u_2), \ldots, x(u_M))$, we can formally write (with a slight abuse of notation, in the equation that follows α denotes the mapping $x \mapsto \alpha(x(u_1), x(u_2), \ldots, x(u_M))$, as it is the derivative of this path-mapping that we need to calculate):

$$(\delta\alpha)(s) = \sum_{m=1}^{M} \alpha_m \cdot \delta_{u_m}(s) \;,$$

where δ_{u_i} is the Dirac delta function centered at u_i. Applying $\mathcal{C}^{bm}$ and $\mathcal{C}^{bb}$ will give immediately the terms in the first lines of the expression for $\mathcal{C}^{bm}\, \delta\Phi(x)$ and $\mathcal{C}^{bb}\, \delta\Phi(x)$ in the statement of the proposition. For the second term $\beta = \beta(I_1, I_2, \ldots, I_L)$:

$$(\delta\beta)(s) = \sum_{l=1}^{L} \beta_l \cdot z_l'(s, x(s)) \;.$$

Again, applying $\mathcal{C}^{bm}$ and $\mathcal{C}^{bb}$ will give the terms in the second lines of the expression for $\mathcal{C}^{bm}\,\delta\Phi(x)$ and $\mathcal{C}^{bm}\,\delta\Phi(x)$ in the statement of the proposition.

We proceed to the term $\gamma = \gamma(S_1, S_2, \ldots, S_J)$ with the stochastic integrals. The algorithm applied in practice will involve a finite-difference approximation of the stochastic integrals $\{S_j\}$. Below we will sacrifice accuracy at the notation to avoid taking too much space for what involves otherwise straightforward derivative calculations. Consider the discretised time instances $0 = s_0 < s_1 < \cdots s_{N-1} < s_N = \ell$. Denoting three consecutive discrete time instances among the above by $s_- < s < s_+$, the finite-difference approximation, say S_j^N, of S_j can be written as follows:

$$S_j^N = \sum_{s\in\{s_1,\ldots,s_N\}} r_j(s_-, x(s_-))(x(s) - x(s_-)) \ .$$

We can now calculate the partial derivative of S_j^N w.r.t. to the one of the N variables, $x(s)$, making up the discretised path. Notice that $x(s)$ will appear in two terms in the summation, unless it is the last point $x(s_N)$ of the x-vector when it will only appear once. This explains the following calculation of the partial derivatives:

$$\frac{\partial S_j^N}{\partial x(s)} = \Delta q_j(s) \ , \ \ s \in \{s_1, \ldots, s_{N-1}\} \ ; \ \ \frac{\partial S_j^N}{\partial x(s_N)} = r_j(s_{N-1}, x(s_{N-1})) \ , \quad \text{(A.10)}$$

where we have defined:

$$\Delta q_j(s) = r_j'(s, x(s))(x(s_+) - x(s)) - (r_j(s, x(s)) - r_j(s_-, x(s_-))) \ .$$

Overall, we have that:

$$\frac{\partial \gamma}{\partial x(s)} = \sum_{j=1}^{J} \gamma_j \cdot \frac{\partial S_j^N}{\partial x(s)} \ . \quad \text{(A.11)}$$

Then, the $N \times N$ discretised covariance operator $C^{bm} = (\min\{s_i, s_k\})_{i,k}$, corresponding to the covariance matrix of a standard Brownian motion at the time instances $s_1, s_2, \ldots, s_N$ (this is the discretised version of $\mathcal{C}^{bm}$), can be easily shown to apply as follows to a finite-dimensional vector $f \in \mathbb{R}^N$:

$$(C^{bm} f)_u = s_u \cdot \big(\sum_{i=1}^{N} f_i\big) - \sum_{k=1}^{u-1} \big(\sum_{i=1}^{k} f_i\big)\, \Delta s_{k+1} \ , \quad u = 1, 2, \ldots, N \ , \quad \text{(A.12)}$$

where $\Delta s_{k+1} = s_{k+1} - s_k$. Combining (A.10), (A.11), (A.12) will give (with ∇ denoting the gradient):

$$(C^{bm}\nabla\gamma)_u =$$
$$\sum_{j=1}^{J} \gamma_j \cdot \Big(s_u \big(r_j(s_{N-1}, x(s_{N-1})) + \sum_{i=1}^{N-1} \Delta q_j(s_i) \big) - \sum_{k=1}^{u-1} \big(\sum_{i=1}^{k} \Delta q_j(s_i)\big) \Delta s_{k+1} \Big) \ .$$

It is easy to see now that a finite-difference approximation of the term appearing in the third line of the expression for $\mathcal{C}^{bm}\delta\Phi(x)$ in the statement of the Proposition 5.1 would coincide precicely with the above expression. A similar approach for the Brownian bridge case, considering the discrete time instances $0 = s_0 < s_1 < \cdots s_{N-1} < s_N < s_{N+1} = \ell$, and the corresponding N-dimensional matrix C^{bb} represented as below:

$$(C^{bb}f)_u = \tfrac{s_u}{\ell} \cdot \sum_{k=1}^{N} \big(\sum_{i=1}^{k} f_i \big) \, \Delta s_{k+1} - \sum_{k=1}^{u-1} \big(\sum_{i=1}^{k} f_i \big) \, \Delta s_{k+1} \ , \ u = 1, \ldots, N \ , \quad \text{(A.13)}$$

where now:

$$S_j^N = \sum_{s \in \{s_1, \ldots, s_{N+1}\}} r_j(s_-, x(s_-))(x(s) - x(s_-)) \ ; \quad \frac{\partial S_j^N}{\partial x(s_i)} = \Delta q_j(s_i) \ ,$$

for $\Delta q_j(s)$ as defined earlier and $1 \leq i \leq N$. This will give the calculation:

$$(C^{bb}\,\nabla\gamma)_u = \sum_{j=1}^{J} \gamma_j \cdot \Big(\frac{s_u}{\ell} \big(\sum_{k=1}^{N} \big(\sum_{i=1}^{k} \Delta q_j(s_i) \big) \Delta s_{k+1} - \sum_{k=1}^{u-1} \big(\sum_{i=1}^{k} \Delta q_j(s_i) \big) \Delta s_{k+1} \Big) \ ,$$

immediately recognised as the finite-difference discretisation of the term appearing in the third line of the expression for $\mathcal{C}^{bb}\delta\Phi(x)$ in the statement of the proposition.

□

Proof of Proposition 7.1. Consider a collection of discrete time instances $0 < s_1 < s_2 < \cdots < s_N$ with $s_0 = 0$ and $s_N = \ell$ that include the data instances:

$$\{u_1, u_2, \ldots, u_M\} \subset \{s_1, s_2, \ldots, s_N\} \ .$$

Let $\Delta s_i = s_i - s_{i-1}$. We will consider the following finite-difference approximation $\Phi(x) = \Phi(x_1, x_2, \ldots, x_N)$ of the negative log-density:

$$\begin{aligned} \Phi(x) = \Phi_1(x) + \Phi_2(x) &= \alpha(v_{i_1}, v_{i_2}, \ldots, v_{i_M}) \qquad \text{(A.14)} \\ &+ \beta\Big(\sum_{i=1}^{N} z_1(s_{i-1}, v_{i-1})\Delta s_i, \sum_{i=1}^{N} z_2(s_{i-1}, v_{i-1})\Delta s_i, \ldots, \sum_{i=1}^{N} z_L(s_{i-1}, v_{i-1})\Delta s_i \Big) \end{aligned}$$

for indices $i_1, i_2, \ldots i_M$ such that $s_{i_m} = u_m$, for $m = 1, 2, \ldots M$, and vector v constructed via the finite-difference approximation:

$$v_i = v_{i-1} + \mu(v_{i-1})\Delta s_i + \sigma(v_{i-1})\Delta x_i \ ,$$

for $i = 1, 2, \ldots, N$ with v_0 equal to a specified fixed initial condition. We will be using the obtained expression in (A.12) for the $N \times N$ covariance matrix $C = C^{bm} = (\min\{s_i, s_j\})_{i,j}$ of a standard brownian motion at the time instances $s_1, s_2, \ldots, s_N$. The function $\Phi : \mathbb{R}^d \mapsto \mathbb{R}$ in (A.14) and the matrix C fully specify

a finite-difference approximation of the original target defined on the Hilbert space.

Now, we have the following recursion for the derivatives

$$Y_{i,j} = \frac{\partial v_i}{\partial x_j} \ ,$$

for any $j \geq 1$:

$$\begin{aligned}
Y_{i,j} &= Y_{i-1,j} + \mu'(v_{i-1})\, Y_{i-1,j} \Delta s_i + \sigma'(v_{i-1})\, Y_{i-1,j}\, \Delta x_i \ ; \quad i > j+1 \\
Y_{j+1,j} &= Y_{j,j} + \mu'(v_j)\, Y_{j,j} \Delta s_{j+1} + \sigma'(v_{i-1})\, Y_{j,j}\, \Delta x_{j+1} - \sigma(v_j) \ ; \\
Y_{j,j} &= \sigma(v_{j-1}) \\
Y_{i,j} &= 0 \ , \quad i < j \ .
\end{aligned}$$

So, we can obtain that, for $i > j+1$:

$$\log(Y_{i,j}) = \log(Y_{i-1,j}) + \log\left(1 + \mu'(v_{i-1})\Delta s_i + \sigma'(v_{i-1})\Delta x_i\right) \ ,$$

and using this recursion we get that:

$$Y_{i,j} = \Delta Q_j \times e^{\sum_{k=j+2}^{i} \log\left(1+\mu'(v_{k-1})\Delta s_k + \sigma'(v_{k-1})\Delta x_k\right)} \ , \quad i \geq j+1 \ , \tag{A.15}$$

$$Y_{j,j} = \sigma(v_{j-1}) \tag{A.16}$$

$$Y_{i,j} = 0 \ , \quad i < j \ , \tag{A.17}$$

where we have set:

$$\Delta Q_j \equiv Y_{j+1,j} = \sigma(v_{j-1})(\mu'(v_j)\Delta s_{j+1} + \sigma'(v_j)\Delta x_{j+1}) - \Delta\sigma(v_j) \ ,$$

and $\Delta\sigma(v_j) = \sigma(v_j) - \sigma(v_{j-1})$. We will also define for $1 \leq m \leq M$ and $1 \leq l \leq L$:

$$\begin{aligned}
F_{m,r} &= \sum_{j=1}^{r} Y_{i_m,j} \ , \quad r < i_m \ ; \\
G_{l,r} &= \sum_{j=1}^{r} \Big(\sum_{i \geq j+1} z_l'(s_i, v_i)\Delta s_{i+1} \Big)\, Y_{i,j} \ ; \\
J_{l,r} &= \sum_{j=1}^{r} z_l'(s_j, v_j) Y_{j,j}\, \Delta s_{j+1} \ .
\end{aligned}$$

The above sequences will appear at the calculation of partial derivatives of $\Phi(x)$. It is important to notice here that these sequences indeed constitute a finite-difference approximation of their continuous-time counterparts appearing at the statement of the proposition: to see that one only need to look at the analytical definition of $Y_{i,j}$ in equations (A.15)-(A.16), and realise that the sum $\sum_k \log\left(1+\mu'(v_{k-1})\Delta s_k + \sigma'(v_{k-1})\Delta x_k\right)$ is essentially a finite-difference approximation of $\int(\mu'(v(u))du + \sigma'(v(u))dx(u))$ as for $\epsilon \approx 0$ we have that $\log(1+\epsilon) \approx \epsilon$.

We can now proceed to the calculation of the partial derivatives of Φ. Clearly:

$$\frac{\partial \Phi}{\partial x_j} = \frac{\partial \Phi_1}{\partial x_j} + \frac{\partial \Phi_2}{\partial x_j} = \sum_{i \geq j} \big(\frac{\partial \Phi_1}{\partial v_i} \cdot Y_{i,j} + \frac{\partial \Phi_2}{\partial v_i} \cdot Y_{i,j} \big) \ .$$

We can easily get:

$$\sum_{i \geq j} \frac{\partial \Phi_1}{\partial v_i} \cdot Y_{i,j} = \sum_{m=1}^{M} \alpha_m Y_{i_m,j} \ .$$

Using (A.12), tedious, but otherwise straightforward calculation will give that, for vector index $1 \leq u \leq N$:

$$(C \, \nabla \Phi_1(x))_u = \sum_{m=1}^{M} \alpha_m \left(s_{u \wedge i_m} \left(F_{m,i_m-1} + Y_{i_m,i_m} \right) + \sum_{k=1}^{u \wedge i_m - 1} F_{m,k} \, \Delta s_{k+1} \right) \ . \tag{A.18}$$

Proceeding to the second term, $\Phi_2(x)$, we have that:

$$\sum_{i \geq j} \frac{\partial \Phi_1}{\partial v_i} \cdot Y_{i,j} = \sum_{l=1}^{L} \beta_l \sum_{i \geq j} z_l(s_i, v_i) \Delta s_{i+1} \, Y_{i,j} \ .$$

We now now multiply with the covariance matrix C to obtain, after some calculations, for $1 \leq u \leq N$:

$$(C \, \nabla \Phi_2(x))_u = \sum_{l=1}^{L} \beta_l \left(s_u \left(G_{l,N} + J_{l,N} \right) - \sum_{k=1}^{u-1} \left(G_{l,k} + J_{l,k} \right) \Delta s_{k+1} \right) \ . \tag{A.19}$$

Upon inspection, (A.18)-(A.19) provide the proof of the statement of the proposition.

□

References

[1] Odd O. Aalen and Håkon K. Gjessing. Survival models based on the Ornstein-Uhlenbeck process. *Lifetime Data Analysis*, 10:407–423, 2004.

[2] Yacine Aït-Sahalia. Closed-form likelihood expansions for multivariate diffusions. *Ann. Statist.*, 36(2):906–937, 2008.

[3] C. Andrieu, A. Doucet, and R. Holenstein. Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B*, 72(3):269–342, 2010.

[4] C. Andrieu and G.O. Roberts. The pseudo-marginal approach for efficient Monte Carlo computations. *Annals of Statistics*, 37(2):697–725, 2009.

[5] Shahbaba B., Lan S., Johnson W., and Neal R. Split Hamiltonian Monte Carlo. Technical report, University of California, 2012.

[6] A. Beskos, Stuart A., and Sanz-Serna J.M. Hybrid Monte-Carlo on Hilbert spaces. *Stochastic Process. Appl.*, 121:2201–2230, 2011.

[7] A. Beskos, O. Papaspiliopoulos, G.O. Roberts, and P. Fearnhead. Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes (with discussion). *Journal of the Royal Statistical Society: Series B*, 68(3):333–382, 2006.

[8] A. Beskos, N. Pillai, G. Roberts, A. Stuart, and J.M. Sanz-Serna. Optimal tuning of the Hybrid Monte-Carlo algorithm. Technical report, University of Warwick, 2012. To appear in Bernoulli.

[9] Alexandros Beskos, Gareth Roberts, Andrew Stuart, and Jochen Voss. MCMC methods for diffusion bridges. *Stoch. Dyn.*, 8(3):319–350, 2008.

[10] B.M. Bibby, M. Jacobsen, and M. Sørensen. Estimating functions for discretely sampled diffusion-type models. *In Aït-Sahalia, Y. and Hansen, L.P. (eds.): Handbook of Financial Econometrics*, pages 203 – 268, 2010. North Holland, Oxford.

[11] Vladimir I. Bogachev. *Gaussian measures*, volume 62 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 1998.

[12] S. Chib, M. K. Pitt, and N. Shephard. Likelihood based inference for diffusion driven state space models. 2006. Working paper.

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

[14] S. Donnet, J.L. Foulley, and A. Samson. Bayesian analysis of growth curves using mixed models defined by stochastic differential equations. *Biometrics*, 66(3):733–741, 2010.

[15] S. Duane, A.D. Kennedy, B. Pendleton, and D. Roweth. Hybrid Monte Carlo. *Phys. Lett. B*, 195(2):216–222, 1987.

[16] J. Dureau, K. Kalogeropoulos, and M. Baguelin. Capturing the time varying drivers of an epidemic using stochastic dynamical systems. 2012. Working paper.

[17] G. B. Durham and A. R. Gallant. Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *J. Bus. Econom. Statist.*, 20(3):297–316, 2002.

[18] Eric Fournié, Jean-Michel Lasry, Jérôme Lebuchoux, Pierre-Louis Lions, and Nizar Touzi. Applications of Malliavin calculus to Monte Carlo methods in finance. *Finance Stoch.*, 3(4):391–412, 1999.

[19] C.J. Geyer. Practical markov chain monte carlo. *Statist. Sci.*, 7:473–483, 1992.

[20] E. Ghysels, A. Harvey, and E. Renault. Stochastic volatily, in. *Handbook of Statistics 14, Statistical Methods in Finance*, 1996. G.S. Maddala and C.R. Rao (eds), North Holland, Amsterdam.

[21] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, editors. *Markov chain Monte Carlo in practice*. Interdisciplinary Statistics. Chapman & Hall, London, 1996.

[22] M. Girolami and B. Calderhead. Riemann manifold langevin and hamiltonian monte carlo methods (with discussion). *Journal of the Royal Statistical Society: Series B*, 73(2):1–37, 2011.

[23] Mark Girolami and Ben Calderhead. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 73(2):123–214, 2011. With discussion and a reply by the authors.

[24] A. Golightly and D. J. Wilkinson. Bayesian inference for nonlinear multivariate diffusion models observed with error. *Comput. Statist. Data Anal.*, 52(3):1674–1693, 2008.

[25] S. Heston. A closed-form solution for options with stochastic volatility. with applications to bonds and currency options. *Review of Financial Studies*, 6:327–343, 1993.

[26] John C Hull and Alan D White. The pricing of options on assets with stochastic volatilities. *Journal of Finance*, 42(2):281–300, 1987.

[27] K. Kalogeropoulos. Likelihood based inference for a class of multidimensional diffusions with unobserved paths. *Journal of Statistical Planning and Inference*, 137:3092–3102, 2007.

[28] K. Kalogeropoulos, P. Dellaportas, and G.O. Roberts. Likelihood-based inference for correllated diffusions. *Canadian Journal of Statistics*, 39(1):52–72, 2011.

[29] K. Kalogeropoulos, G.O. Roberts, and P. Dellaportas. Inference for stochastic volatility models using time change transformations. *Annals of Statistics*, 38(2):784–807, 2010.

[30] Peter E. Kloeden and Eckhard Platen. *Numerical solution of stochastic differential equations*, volume 23 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1992.

[31] S. C. Kou. Stochastic modeling in nanoscale biophysics: subdiffusion within proteins. *Ann. Appl. Stat.*, 2(2):501–535, 2008.

[32] Benedict Leimkuhler and Sebastian Reich. *Simulating Hamiltonian dynamics*, volume 14 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, Cambridge, 2004.

[33] Bernt Øksendal. *Stochastic differential equations*. Universitext. Springer-Verlag, Berlin, sixth edition, 2003. An introduction with applications.

[34] R. Overgaard, N. Jonsson, C. Tornøe, and H. Madsen. Non-linear mixed-effects models with stochastic differential equations: Implementation of an estimation algorithm. *Journal of Pharmacokinetics and Pharmacodynamics*, 32:85–107, 2005.

[35] G.O. Roberts and L. M. Sangalli. Latent diffusion models for survival analysis. *Bernoulli*, 16(2):435–458, 2010.

[36] G.O. Roberts and O. Stramer. On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm. *Biometrika*, 88(3):603–621, 2001.

[37] O. Stramer and M. Bognar. Bayesian inference for irreducible diffusion processes using the pseudo-marginal approach. *Bayesian Analysis*, 6(2):231–258, 2011.

[38] Ciprian A. Tudor and Frederi G. Viens. Statistical aspects of the fractional stochastic calculus. *Ann. Statist.*, 35(3):1183–1212, 2007.

[39] David White. *A Comparison of MCMC Methods for Conditioned Diffusions*. PhD thesis, Mathematics Institute, University of Warwick, 2009.