Hybrid Monte-Carlo on Hilbert Spaces

Alexandros Beskos\textsuperscript{a,∗}, F.J. Pinski\textsuperscript{b}, J. M. Sanz-Serna\textsuperscript{c}, Andrew Stuart\textsuperscript{d}

\textsuperscript{a}Department of Statistical Science, University College London, Gower Street, London, WC1E 6BT, UK
\textsuperscript{b}Physics Department, University of Cincinnati, Geology-Physics Building, P. O. Box 210011, Cincinnati, OH 45221, USA
\textsuperscript{c}Departamento de Matemática Aplicada, Universidad de Valladolid, Spain
\textsuperscript{d}Mathematics Institute, University of Warwick, Coventry, CV4 7AL, UK

Abstract

The Hybrid Monte-Carlo (HMC) algorithm provides a framework for sampling from complex, high-dimensional target distributions. In contrast with standard Markov chain Monte-Carlo (MCMC) algorithms, it generates nonlocal, non-symmetric moves in the state space, alleviating random walk type behaviour for the simulated trajectories. However, similarly to algorithms based on random walk or Langevin proposals, the number of steps required to explore the target distribution typically grows with the dimension of the state space. We define a generalized HMC algorithm which overcomes this problem for target measures arising as finite-dimensional approximations of measures $\pi$ which have density with respect to a Gaussian measure on an infinite-dimensional Hilbert space. The key idea is to construct an MCMC method which is well-defined on the Hilbert space itself.

We successively address the following issues in the infinite-dimensional setting of a Hilbert space: (i) construction of a probability measure $\Pi$ in an enlarged phase space having the target $\pi$ as a marginal, together with a Hamiltonian flow that preserves $\Pi$; (ii) development of a suitable geometric numerical integrator for the Hamiltonian flow; and (iii) derivation of an accept/reject rule to ensure preservation of $\Pi$ when using the above numerical integrator instead of the actual Hamiltonian flow. Experiments are reported that compare the new algorithm with standard HMC and with a version of the Langevin MCMC method defined on a Hilbert space.

1. Introduction

Several applications of current interest give rise to the problem of sampling a probability measure $\pi$ on a separable Hilbert space $(\mathcal{H}, \langle \cdot, \cdot \rangle, |\cdot|)$ defined via

*Corresponding Author

Email addresses: alex@stats.ucl.ac.uk (Alexandros Beskos), frank.pinski@uc.edu (F.J. Pinski), sanzsern@mac.uva.es (J. M. Sanz-Serna), a.m.stuart@warwick.ac.uk (Andrew Stuart)
its density with respect to a Gaussian measure $\pi_0$:

$$\frac{d\pi}{d\pi_0}(q) \propto \exp(-\Phi(q)).$$

(1)

Measures with this form arise, for example, in the study of conditioned diffusions [11] and the Bayesian approach to inverse problems [20]. The aim of this paper is to develop a generalization of the Hybrid Monte-Carlo (HMC) method which is tailored to the sampling of measures $\pi$ defined as in (1).

Any algorithm designed to sample $\pi$ will in practice be implemented on a finite-dimensional space of dimension $N$; key to the efficiency of the algorithm will be its cost as a function of $N$. Mathematical analysis in the simplified scenario of product targets [17, 16, 1], generalizations to the non-product case in [2] and practical experience together show that these MCMC methods require $O(N^a)$ steps to explore the approximate target in $\mathbb{R}^N$, for some $a > 0$. Indeed for specific targets and proposals it is proven that for the standard Random-Walk Metropolis (RWM), Metropolis-adjusted Langevin algorithm (MALA) and HMC methods $a = 1, 1/3$ and $1/4$ respectively. The growth of the required number of steps with $N$ occurs for one or both of the following two reasons: either because the algorithms are not defined in the limit $N = \infty$ or because the proposals at $N = \infty$ are distributed according to measures which are not absolutely continuous with respect to the target measure $\pi$. Finite-dimensional approximations then require, as $N$ increases, smaller and smaller moves to control these shortcomings. On the other hand, when $\pi_0$ is Gaussian it is now understood that both the MALA and RWM algorithms can be generalized to obtain methods which require $O(1)$ steps to explore the approximate target in $\mathbb{R}^N$ [3, 4, 5]. This is achieved by designing algorithms where the proposal is absolutely continuous with respect to $\pi_0$ (thus, also with respect to $\pi$) so that the method is well-defined even on an infinite-dimensional Hilbert space $\mathcal{H}$. In this paper we show that similar ideas can be developed for the HMC method.

The standard HMC algorithm was introduced in [8]. It is based on the observation that the exponential of a separable Hamiltonian, with potential energy given by the negative logarithm of the target density, is invariant under the Hamiltonian flow. In contrast with standard Markov chain Monte-Carlo (MCMC) methods such as RWM and MALA, the HMC algorithm generates nonlocal moves in the state space, offering the potential to overcome undesirable mixing properties associated with random walk behaviour; see [14] for an overview. It is thus highly desirable to generalize HMC to the infinite-dimensional setting required by the need to sample measures of the form (1).

The paper proceeds as follows. In Section 2 we review the standard HMC method. The new Hilbert space algorithm is presented in Section 3. Therein, we successively address the following issues in the infinite-dimensional setting of a Hilbert space: (i) construction of a probability measure $\Pi$ in an enlarged

---

1 The restriction is then analogous to a Courant stability restriction in the numerical approximation of partial differential equations.
phase space having the target \( \pi \) as a marginal, together with a Hamiltonian flow that preserves \( \Pi \); (ii) development of a suitable geometric numerical integrator for the Hamiltonian flow; and (iii) derivation of an accept/reject rule to ensure preservation of \( \Pi \) when using the above numerical integrator instead of the actual Hamiltonian flow. All required proofs have been collected in Section 4. Section 5 contains numerical experiments illustrating the advantages of our generalized HMC method over both the standard HMC method [8] and the modified MALA algorithm which is defined in a Hilbert space [3]. We make some concluding remarks in Section 6 and, in the Appendix, we gather some results from the Hamiltonian formalism.

2. Standard HMC on \( \mathbb{R}^N \)

In order to facilitate the presentation of the new Hilbert-space-valued algorithm to be introduced in Section 3 below, it is convenient to review first the standard HMC method defined in [8] (for additional details see e.g. [14]).

The aim is to sample from a probability density function \( \pi \) in \( \mathbb{R}^N \) which has the form

\[
\pi(q) \propto \exp \left( -\frac{1}{2} \langle q, Lq \rangle - \Phi(q) \right),
\]

(2)

where \( L \) is a symmetric, positive semi-definite matrix (at this stage the choice \( L = 0 \) is not excluded). HMC is based on the combination of three elements: (i) a Hamiltonian flow, (ii) a numerical integrator and (iii) an accept/reject rule. Each of these is discussed in a separate subsection. The final subsection examines the choice of the mass matrix required to apply the method.

2.1. Hamiltonian Flow

Consider a Hamiltonian function (‘energy’) in \( \mathbb{R}^{2N} \) ([18]) associated with the target density (2):

\[
H(q, p) = \frac{1}{2} \langle p, M^{-1} p \rangle + \frac{1}{2} \langle q, Lq \rangle + \Phi(q).
\]

(3)

Here \( p \) is an auxiliary variable (‘momentum’) and \( M \) a user-specified, symmetric positive definite ‘mass’ matrix. Denoting by

\[
f = -\nabla \Phi
\]

the ‘force’ stemming from the ‘potential’ \( \Phi \), the corresponding canonical Hamiltonian differential equations read as (see the Appendix)

\[
\frac{dq}{dt} = \frac{\partial H}{\partial p} = M^{-1} p, \quad \frac{dp}{dt} = -\frac{\partial H}{\partial q} = -Lq + f(q).
\]

(4)

HMC is based on the fact that, for any fixed \( t \), the \( t \)-flow \( \Xi^t \) of (4), i.e. the map \( \Xi^t : \mathbb{R}^{2N} \rightarrow \mathbb{R}^{2N} \) such that

\[
(q(t), p(t)) = \Xi^t(q(0), p(0)),
\]
preserves both the volume element $dq \, dp$ and the value of $H$. As a result, $\Xi^t$ also preserves the measure in the phase space $\mathbb{R}^{2N}$ with density

$$\Pi(q, p) \propto \exp \left( -\frac{1}{2} \langle p, M^{-1} p \rangle \right) \exp \left( -\frac{1}{2} \langle q, L q \rangle - \Phi(q) \right), \quad (5)$$

whose $q$ and $p$ marginals are respectively the target (2) and a centred Gaussian with $M$ as a covariance matrix. It follows that if we assume that the initial value $q(0)$ is distributed according to (2) and we draw $p(0) \sim N(0, M)$, then $q(t)$ will also follow the law (2). This shows that the implied Markov transition kernel $q(0) \mapsto q(T)$, with a user-defined fixed $T > 0$, defines a Markov chain in $\mathbb{R}^N$ that has (2) as an invariant density and makes nonlocal moves in that state space [19]. Furthermore, the chain is reversible, in view of the symmetry of the distribution $N(0, M)$ and of the time-reversibility of the dynamics of (4): that is, if $\Xi^T(q, p) = (q', p')$, then $\Xi^T(q', -p') = (q, -p)$.

2.2. Numerical Integrator

In general the analytic expression of the flow $\Xi^t$ is not available and it is necessary to resort to numerical approximations to compute the transitions. The integrator of choice, the Verlet/leap-frog method, is best presented as a splitting algorithm, see e.g. [18]. The Hamiltonian (3) is written in the form

$$H = H_1 + H_2, \quad H_1 = \frac{1}{2} \langle q, L q \rangle + \Phi(q), \quad H_2 = \frac{1}{2} \langle p, M^{-1} p \rangle,$$

where the key point is that the flows $\Xi_1^t, \Xi_2^t$ of the split Hamiltonian systems

$$\frac{dq}{dt} = \frac{\partial H_1}{\partial p} = 0 , \quad \frac{dp}{dt} = -\frac{\partial H_1}{\partial q} = -L q + f(q)$$

and

$$\frac{dq}{dt} = \frac{\partial H_2}{\partial p} = M^{-1} p , \quad \frac{dp}{dt} = -\frac{\partial H_2}{\partial q} = 0$$

may be explicitly computed:

$$\Xi_1^t(q, p) = (q, p - tL q + tf(q)) , \quad \Xi_2^t(q, p) = (q + tM^{-1} p, p) .$$

Then a time-step of length $h > 0$ of the Verlet algorithm is, by definition, carried out by composing three substeps:

$$\Psi_h = \Xi_{1}^{h/2} \circ \Xi_{2}^{h} \circ \Xi_{1}^{h/2} ; \quad (6)$$

and the exact flow $\Xi^T$ of (4) is approximated by the transformation $\Psi_h^{(T)}$ obtained by concatenating $\left\lfloor \frac{T}{h} \right\rfloor$ Verlet steps:

$$\Psi_h^{(T)} = \Psi_h^{\left\lfloor \frac{T}{h} \right\rfloor} .$$

Since the mappings $\Xi_1^t$ and $\Xi_2^t$ are exact flows of Hamiltonian systems, the transformation $\Psi_h^{(T)}$ itself is symplectic and preserves the volume element $dq \, dp$
(see the Appendix). Also the symmetry in the right hand-side of (6) (Strang’s splitting) results in $\Psi_h^{(T)}$ being time-reversible:

$$\Psi_h^{(T)}(q,p) = (q',p') \Leftrightarrow \Psi_h^{(T)}(q',-p') = (q,-p).$$

The map $\Psi_h^{(T)}$ is an example of a geometric integrator [10] as it preserves these various geometric properties of the flow $\Xi_T$. However $\Psi_h^{(T)}$ preserves the value of $H$ only up to an $O(h^2)$ error and, accordingly, it does not preserve exactly the measure with density (5).

2.3. Accept/Reject Rule

The invariance of (5) in the presence of integration errors is ensured through an accept/reject Metropolis-Hastings mechanism; the right recipe is given in steps (iii) and (iv) of Table 1, that summarizes the standard HMC algorithm [8, 14].

---

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

(i) Pick $q^{(0)} \in \mathbb{R}^N$ and set $n = 0$.

(ii) Given $q^{(n)}$, compute

$$(q^*, p^*) = \Psi_h^{(T)}(q^{(n)}, p^{(n)})$$

where $p^{(n)} \sim N(0, M)$ and propose $q^*$.

(iii) Calculate

$$a = \min \left(1, \exp \left(H(q^{(n)}, p^{(n)}) - H(q^*, p^*)\right) \right).$$

(iv) Set $q^{(n+1)} = q^*$ with probability $a$; otherwise set $q^{(n+1)} = q^{(n)}$.

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

---

Table 1: Standard HMC algorithm on $\mathbb{R}^N$. It generates a Markov chain $q^{(0)} \mapsto q^{(1)} \mapsto \ldots$ reversible with respect to the target probability density function (2).

2.4. Choice of Mass Matrix

As pointed out above, the mass-matrix $M$ is a ‘parameter’ to be selected by the user; the particular choice of $M$ will have great impact on the efficiency of the algorithm ([9]). A rule of thumb ([12]) is that directions where the target (2) possesses larger variance should be given smaller mass so that the Hamiltonian flow can make faster progress along them.

In order to gain understanding concerning the role of $M$ and motivate the material in Section 3, we consider in the remainder of this section the case where
in (2) the matrix $L$ is positive-definite and $\Phi(q)$ is small with respect to $\langle q, Lq \rangle$, i.e. the case where the target is a perturbation of the distribution $N(0, L^{-1})$.

In agreement with the rule of thumb above, we set $M = L$ so that (4) reads

$$\frac{dq}{dt} = L^{-1}p, \quad \frac{dp}{dt} = -Lq + f(q).$$

(7)

Let us now examine the limit situation where the perturbation vanishes, i.e. $\Phi \equiv 0$. From (5), at stationarity, $q \sim N(0, L^{-1})$, $p \sim N(0, L)$. Furthermore in (7), $f \equiv 0$ so that, after eliminating $p$,

$$\frac{d^2q}{dt^2} = -q.$$

Thus, $q(t)$ undergoes oscillations with angular frequency 1 regardless of the size of the eigenvalues of $L/(\text{co})$-variances of the target. From a probabilistic point of view, this implies that, if we think of $q$ as decomposed in independent Gaussian scalar components, the algorithm (as intended with the choice of mass matrix $M = L$) automatically adjusts itself to the fact that different components may possess widely different variances. From a numerical analysis point of view, we see that the Verlet algorithm will operate in a setting where it will not be necessary to reduce the value of $h$ to avoid stability problems originating from the presence of fast frequencies.\textsuperscript{2}

Remark 1. Let us still keep the choice $M = L$ but drop the assumption $\Phi \equiv 0$ and suppose that $L$ has some very large eigenvalues (i.e. the target distribution presents components of very small variance). As we have just discussed, we do not expect such large eigenvalues to affect negatively the dynamics of $q$. However we see from the second equation in (7) that $p$ (which, recall, is only an auxiliary variable in HMC) will in general be large. In order to avoid variables of large size, it is natural to rewrite the algorithm in Table 1 using throughout the scaled variable

$$v = L^{-1}p$$

rather than $p$. Since $v = M^{-1}p = dq/dt$, the scaled variable possesses a clear meaning: it is the ‘velocity’ of $q$.

In terms of $v$ the system (7) that provides the required flow reads

$$\frac{dq}{dt} = v, \quad \frac{dv}{dt} = -q + L^{-1}f(q);$$

(8)

the value of the Hamiltonian (3) to be used in the accept/reject step is given by

$$H = \frac{1}{2}\langle v, Lv \rangle + \frac{1}{2}\langle q, Lq \rangle + \Phi(q),$$

(9)

\textsuperscript{2}Note that the Verlet algorithm becomes unstable whenever $h\omega \geq 2$, where $\omega$ is any of the angular frequencies present in the dynamics. While the choice of mass matrix $M = L$ precludes the occurrence of stability problems in the integration, the standard HMC algorithm in the present setting ($M = L$, $\Phi \equiv 0$) still suffers from the restriction $h = O(N^{1/4})$ discussed in [1] (the restriction stems from accuracy—rather than stability—limitations in the Verlet integrator). The new integrator to be introduced later in the paper is exact in the setting $\Phi \equiv 0$, and hence eliminates this problem.
and the invariant density (5) in $\mathbb{R}^{2N}$ becomes

$$\Pi(q,v) \propto \exp\left(-\frac{1}{2}(v,Lv)\right) \exp\left(-\frac{1}{2}(q,Lq) - \Phi(q)\right).$$

Note that the marginal for $v$ is

$$v \sim N(0, L^{-1});$$

the initial value $v^{(n)}$ at step (ii) of Table 1 should be drawn accordingly.

We finish the section with two comments concerning introduction of the variable $v$ in place of $p$. The algorithm expressed in terms of $v$ may be found either by first replacing $p$ by $v$ in the differential equations (7) to get (8) and then applying the Verlet algorithm; or by first applying the Verlet algorithm to the system (7) and then replacing $p$ by $v$ in the equations of the integrator: the Verlet discretization commutes with the scaling $p \mapsto v = M^{-1}p$. In addition, it is important to note that (8) is also a Hamiltonian system, albeit of a non-canonical form, see the Appendix.

3. The Algorithm

In this section we define the new algorithm on a Hilbert space $\mathcal{H}$ and outline its main mathematical properties. After introducing the required assumptions on the distribution to be sampled, we discuss successively the flow, the numerical integrator and the accept/reject strategy.

3.1. Assumptions on $\pi_0$ and $\Phi$

Throughout we assume that $\pi_0$ in (1) is a non-degenerate (non-Dirac) centred Gaussian measure with covariance operator $C$. Thus, $C$ is a positive, self-adjoint, nuclear operator (i.e. its eigenvalues are summable) whose eigenfunctions span $\mathcal{H}$. For details on properties of Gaussian measures on a Hilbert space see [6, 7].

Let $\{\phi_j\}_{j \geq 1}$ be the (normalised) eigenfunctions of $C$ and $\lambda^2_j$ its eigenvalues, so that

$$C\phi_j = \lambda^2_j \phi_j, \quad j \geq 1.$$ 

The expansion

$$q = \sum_{j=1}^{\infty} q_j \phi_j$$

establishes an isomorphism between $\mathcal{H}$ and the space

$$\ell_2 = \left\{\{q_j\}_{j=1}^{\infty} \in \mathbb{R}^\infty : \sum q_j^2 < \infty\right\}$$

that maps each element $q$ into the corresponding sequence of coefficients $\{q_j\}_{j \geq 1}$. This isomorphism gives rise to subspaces ($s > 0$) and superspaces ($s < 0$) of $\mathcal{H}$:

$$\mathcal{H}^s := \left\{\{q_j\}_{j=1}^{\infty} \in \mathbb{R}^\infty : |q|_s < \infty\right\},$$
where \(|\cdot|_s\) denotes the following Sobolev-like norm:

\[
|q|_s := \left( \sum_{j=1}^{\infty} j^{2s} q_j^2 \right)^{1/2}, \quad s \in \mathbb{R}.
\]

If \(q \sim N(0, \mathcal{C})\), then

\[
q_j \sim N(0, \lambda_j^2)
\]

independently over \(j\). Thus, \(\lambda_j\) is the standard deviation, under the reference measure \(\pi_0\), of the \(j^{th}\) coordinate. We shall impose the following condition, with the bound \(\kappa > 1/2\) being required to ensure that \(\mathcal{C}\) is nuclear, so that the Gaussian distribution is well-defined:

**Condition 3.1.** The standard deviations \(\{\lambda_j\}_{j \geq 1}\) decay at a polynomial rate \(\kappa > 1/2\), that is

\[
\lambda_j = \Theta(j^{-\kappa}).
\]

From this condition and (14), it is easy to check that

\[
|q|_s < \infty, \quad \pi_0 - \text{a.s., for any } s \in [0, \kappa - 1/2),
\]

therefore, we have the following.

**Proposition 3.1.** Under Condition 3.1, the probability measure \(\pi_0\) is supported on \(\mathcal{H}_s\) for any \(s < \kappa - 1/2\).
with respect to \( \Pi_0 \). The global Lipschitz condition on \( f \) is made to simplify the analysis, and could be replaced by a local Lipschitz assumption; indeed we will give numerical results in Subsection 5.2 for a measure arising from conditioned diffusions where \( f \) is only locally Lipschitz.

We shall always assume hereafter that Conditions 3.1–3.3 are satisfied and use the symbols \( \kappa \) and \( \ell \) to refer to the relevant constants.

3.2. Flow

There is a clear analogy between the problem of sampling from \( \pi \) given by (1) in \( \mathcal{H} \) and the problem, considered in Subsection 2.4, of sampling from the density (2) in \( \mathbb{R}^N \) with \( L \) positive definite and \( \Phi(q) \) small with respect to \( \langle q, Lq \rangle \). In this analogy, \( \pi_0(dq) \) corresponds to the measure \( \exp(-\frac{1}{2}\langle q, Lq \rangle) \) and therefore the covariance operator \( \mathcal{C} \) corresponds to the matrix \( L^{-1} \): \( L \) is the precision operator. Many of the considerations that follow are built on this parallelism.

The key idea in HMC methods is to double the size of the state space by adding an auxiliary variable related to the ‘position’ \( q \). We saw in Remark 1 in Subsection 2.4, that, in the setting considered there, large eigenvalues of \( L \) lead to large values of the momentum \( p \) but do not affect the size of \( v \). In the Hilbert space setting, the role of \( L \) is played by \( \mathcal{C}^{-1} \) which has eigenvalues \( 1/\lambda_j^2 \) of arbitrarily large size. This suggests working with the velocity \( v = dq/dt \) as an auxiliary variable and not the momentum. Equation (11) prompts us to use \( \pi_0 \) as the marginal distribution of \( v \) and introduce the following Gaussian measure \( \Pi_0 \) on \( \mathcal{H} \times \mathcal{H} \)

\[
\Pi_0(dq, dv) = \pi_0(dq) \otimes \pi_0(dv).
\]

We define accordingly (cf. (10)):

\[
\frac{d\Pi}{d\Pi_0}(q, v) \propto \exp(-\Phi(q)),
\]

so that the marginal on \( q \) of \( \Pi \) is simply the target distribution \( \pi \). Furthermore (8) suggests to chose

\[
\frac{dq}{dt} = v, \quad \frac{dv}{dt} = -q + \mathcal{C} f(q)
\]

as the equations to determine the underlying dynamics that will provide (when solved numerically) proposals for the HMC algorithm with target distribution \( \pi \).

Our first result shows that (16) defines a well-behaved flow \( \Xi^t \) in the subspace \( \mathcal{H}^\ell \times \mathcal{H}^\ell \) of \( \mathcal{H} \times \mathcal{H} \) which, according to Proposition 3.1, has full \( \Pi_0 \) (or \( \Pi \)) measure. The space \( \mathcal{H}^\ell \times \mathcal{H}^\ell \) is assumed to have the product topology of the factor spaces \( (\mathcal{H}^\ell, | \cdot |_\ell) \).

Proposition 3.2.

(i) For any initial condition \( (q(0), v(0)) \in \mathcal{H}^\ell \times \mathcal{H}^\ell \) and any \( T > 0 \) there exists a unique solution of (16) in the space \( C^1([-T, T], \mathcal{H}^\ell \times \mathcal{H}^\ell) \).
(ii) Let $\Xi^t : \mathcal{H}^t \times \mathcal{H}^t \to \mathcal{H}^t \times \mathcal{H}^t$, $t \in \mathbb{R}$ denote the group flow of (16), so that
\[(q(t), v(t)) = \Xi^t(q(0), v(0)) \, .\]
The map $\Xi^t$ is globally Lipschitz with a Lipschitz constant of the form $\exp(K|t|)$, where $K$ depends only on $\pi_0$ and $\Phi$.
(iii) Accordingly, for each $T > 0$, there exists constant $C(T) > 0$ such that, for $0 \leq t \leq T$,
\[|q(t)|_\ell + |v(t)|_\ell \leq C(T)(1 + |q(0)|_\ell + |v(0)|_\ell) \, .\]

Our choices of measure (15) and dynamics (16) have been coordinated to ensure that $\Xi^t$ preserves $\Pi$.

**Theorem 3.1.** For any $t \in \mathbb{R}$, the flow $\Xi^t$ preserves the probability measure $\Pi$ given by (15).

The theorem implies that $\pi$ will be an invariant measure for the Markov chain for $q$ defined through the transitions $q^{(n)} \mapsto q^{(n+1)}$ determined by
\[(q^{(n+1)}, v^{(n+1)}) = \Xi^T(q^{(n)}, v^{(n)}) \, , \quad v^{(n)} \sim \pi_0 \, ,
\tag{17}\]
where the $v^{(n)}$ form an independent sequence. This chain is actually reversible:

**Theorem 3.2.** For any $t \in \mathbb{R}$, the Markov chain defined by (17) is reversible under the distribution $\pi(q)$ in (1).

We conclude this subsection by examining whether the dynamics of (16) preserve a suitable Hamiltonian function. The Hilbert-space counterpart of (9) is given by
\[H(q, v) = \frac{1}{2} \langle v, C^{-1}v \rangle + \frac{1}{2} \langle q, C^{-1}q \rangle + \Phi(q) \tag{18}\]
and it is in fact trivial to check that $H$ and therefore $\exp(-H)$ are formal invariants of (16). However the terms $\langle q, C^{-1}q \rangle$ and $\langle v, C^{-1}v \rangle$ are almost surely infinite in an infinite-dimensional context; this is because $|C^{-\frac{1}{2}}|\cdot$ is the Cameron-Martin norm for $\pi_0$, see e.g. [6, 7]. For further discussion on the Hamiltonian nature of (16) see the Appendix.

### 3.3. Numerical Integrator

Our next task is to study how to approximate numerically the flow $\Xi^T$. As in the derivation of the Verlet algorithm in Subsection 3.3 we resort to the idea of splitting; however the splitting that we choose is different, dictated by a desire to ensure that the resulting MCMC method is well-defined on Hilbert space. The system (16) is decomposed as (see the Appendix)
\[
\frac{dq}{dt} = 0 \, , \quad \frac{dv}{dt} = C f(q) \tag{19}
\]
and
\[
\frac{dq}{dt} = v \, , \quad \frac{dv}{dt} = -q \tag{20}
\]
with the explicitly computable flows
\begin{equation}
\Xi_t^1(q, v) = (q, v + tf(q)) ,
\end{equation}
and
\begin{equation}
\Xi_t^2(q, v) = \left( \cos(t)q + \sin(t)v, \ -\sin(t)q + \cos(t)v \right) .
\end{equation}
A time-step of length \( h > 0 \) of the integrator is carried out by the symmetric
composition (Strang’s splitting)
\begin{equation}
\Psi_h = \Xi_{h/2} \circ \Xi_h \circ \Xi_{h/2}
\end{equation}
and the exact flow \( \Xi_T, T > 0 \), of (16) is approximated by the map \( \Psi^{(T)}_h \) obtained
by concatenating \( \left\lfloor \frac{T}{h} \right\rfloor \) steps:
\begin{equation}
\Psi_h^{(T)} = \Psi_h^{\left\lfloor \frac{T}{h} \right\rfloor} .
\end{equation}
This integrator is reversible —due to the symmetric pattern in the Strang
splitting and the reversibility of \( \Xi_1^1 \) and \( \Xi_2^2 \)— and if applied in a finite-dimensional
setting would also preserve the volume element \( dq \, dv \). In the case where \( \Phi \equiv 0 \),
the integrator coincides with the rotation \( \Xi_2^2 \); it is therefore exact and preserves
exactly the measure \( \Pi_0 \). However, in general, \( \Psi^{(T)}_h \) does not preserve formally
the Hamiltonian (18), a fact that renders necessary the introduction of an ac-
cept/reject criterion, as we will describe in the following subsection.

The next result is analogous to Proposition 3.2:

**Proposition 3.3.**

(i) For any \( (q, v) \in H^t \times H^t \) we have \( \Psi_h(q, v) \in H^t \times H^t \) and therefore
\( \Psi^{(T)}_h(q, v) \in H^t \times H^t \).

(ii) \( \Psi_h \), and therefore \( \Psi^{(T)}_h \), preserves absolute continuity with respect to \( \Pi_0 \)
and \( \Pi \).

(iii) \( \Psi^{(T)}_h \) is globally Lipschitz as a map from \( H^t \times H^t \) onto itself with a
Lipschitz constant of the form \( \exp(KT) \) with \( K \) depending only on \( \pi_0 \) and \( \Phi \).

(iv) Accordingly, for each \( T > 0 \) there exists \( C(T) > 0 \) such that, for all
\( 0 \leq ih \leq T \),
\begin{equation}
|q_i|_\ell + |v_i|_\ell \leq C(T)(1 + |q_0|_\ell + |v_0|_\ell) ,
\end{equation}
where
\begin{equation}
(q_i, v_i) = \Psi_h^i(q_0, v_0) .
\end{equation}

3.4. Accept/Reject Rule

The analogy with the standard HMC would suggest the use of
\begin{equation}
1 \wedge \exp\left( H(q^{(n)}, v^{(n)}) - H(\Psi^{(T)}_h(q^{(n)}, v^{(n)})) \right)
\end{equation}
to define the acceptance probability. Unfortunately and as pointed out before, \( H \)
is almost surely infinite in our setting. We will bypass this difficulty by deriving
a well behaved expression for the energy difference
\begin{equation}
\Delta H(q, v) = H(\Psi^{(T)}_h(q, v)) - H(q, v) .
\end{equation}
A tedious but straightforward calculation using the definition of $\Psi_h(q, v)$ gives, for one time-step $(q', v') = \Psi_h(q, v)$:

$$
H(q', v') - \Phi(q') = H(q, v) - \Phi(q) + \frac{h^2}{8} \left( |C^{+}f(q)|^2 - |C^{+}f(q')|^2 \right) 
+ \frac{h}{2} \left( \langle f(q), v \rangle + \langle f(q'), v' \rangle \right).
$$

Using this result iteratively, we obtain for $I = \lfloor T/h \rfloor$ steps (subindices refer to time-levels along the numerical integration):

$$
\Delta H(q_0, v_0) = \Phi(q_I) - \Phi(q_0) + h^2 \left( |C^{+}f(q_0)|^2 - |C^{+}f(q_I)|^2 \right) 
+ h \sum_{i=1}^{I-1} \left( \langle f(q_i), v_i \rangle + \frac{h}{2} \left( \langle f(q_0), v_0 \rangle + \langle f(q_I), v_I \rangle \right) \right).
$$

(26)

(We note in passing that in the continuum $h \to 0$ limit, (26) gives formally:

$$
H(q(T), v(T)) - H(q(0), v(0)) = \Phi(q(T)) - \Phi(q(0)) + \int_0^T \langle f(q(t)), v(t) \rangle \, dt,
$$

with the right hand side here being identically 0: the gain in potential energy equals the power of the applied force. This is a reflection of the formal energy conservation by the flow (16) pointed out before.) Condition 3.2 and parts (ii) and (iv) of Lemma 4.1 in Section 4 now guarantee that $\Delta H(q, v)$, as defined in (26), is a $\Pi$-a.s. finite random variable; in fact $\Delta H: H^\ell \times H^\ell \to \mathbb{R}$ is continuous according to parts (iii) and (v) of that Lemma. We may therefore define the acceptance probability by

$$
a(q, v) = \min \left( 1, \exp \left( -\Delta H(q, v) \right) \right).
$$

(27)

We are finally ready to present an HMC algorithm on $H^\ell$ aiming at simulating from $\pi(q)$ in equilibrium. The pseudo-code is given in Table 2. Our main result asserts that the algorithm we have defined achieves its goal:

**Theorem 3.3.** For any choice of $T > 0$, the algorithm in Table 2 defines a Markov chain which is reversible under the distribution $\pi(q)$ in (1).

The practical application of the algorithm requires of course to replace $H$, $\pi_0$ and $\Phi$ by finite-dimensional approximations. Once these have been chosen, it is a trivial matter to write the corresponding versions of the differential system (16), of the integrator and of the accept/reject rule. The case where the discretization is performed by a spectral method is presented in Subsection 4.2, but of course many alternative possibilities exist. Indeed in Subsection 5.2 we present numerical results based on finite-dimensionalization using finite differences. For any finite-dimensional approximation of the state space, the fact that the algorithm is defined in the infinite-dimensional limit imparts robustness under refinement of finite-dimensional approximation.
HMC on $\mathcal{H}^\ell$:

(i) Pick $q^{(0)} \sim \Pi_0$ and set $n = 0$.
(ii) Given $q^{(n)}$, compute

$$ (q^*, v^*) = \Psi^T_h(q^{(n)}, v^{(n)}) $$

where $v^{(n)} \sim N(0, C)$ and propose $q^*$.
(iii) Using (27), define

$$ a = a(q^{(n)}, v^{(n)}) $$

(iv) Set $q^{(n+1)} = q^*$ with probability $a$; otherwise set $q^{(n+1)} = q^{(n)}$.
(v) Set $n \rightarrow n + 1$ and go to (ii).

Table 2: The HMC algorithm on a Hilbert space, for sampling from $\pi$ in (1).

4. Proofs

This section contains some auxiliary results and the proofs of the theorems and propositions presented in Section 3.

4.1. Preliminaries

The bounds for $f = -D\Phi$ provided in the following lemma will be used repeatedly. The proof relies on the important observation that Condition 3.1 implies that

$$ |C^{-s/2\kappa} \cdot | \simeq | \cdot |_s, \quad (28) $$

where we use the symbol $\simeq$ to denote that that two norms are equivalent.

**Lemma 4.1.** There exists a constant $K > 0$ such that

(i) for all $q, q' \in \mathcal{H}^\ell$,

$$ |Cf(q) - Cf(q')|_\ell \leq K|q - q'|_\ell ; $$

(ii) for all $q, v \in \mathcal{H}^\ell$,

$$ |\langle f(q), v \rangle| \leq K(1 + |q|_\ell)|v|_\ell ; $$

(iii) for all $q, q', v, v' \in \mathcal{H}^\ell$,

$$ |\langle f(q), v \rangle - \langle f(q'), v' \rangle| \leq K|v|_\ell q - q'|_\ell + K(1 + |q'|_\ell)|v - v'|_\ell ; $$

(iv) for all $q \in \mathcal{H}^\ell$,

$$ |C^2 f(q)| \leq K(1 + |q|_\ell) ; $$

$$ |C^3 f(q)| \leq K(1 + |q|_\ell) ; $$
(v) for all \( q, q' \in \mathcal{H}^\ell \),
\[
|\mathcal{C}^{1/2} f(q) - \mathcal{C}^{1/2} f(q')| \leq K |q - q'|^\ell .
\]

**Proof.** From (28)
\[
|\mathcal{C} \cdot |_\ell \simeq |\mathcal{C}^{1/2} \cdot | , \quad | \cdot |_\ell \simeq |\mathcal{C}^{1/2} \cdot | ,
\]
and, since \( \ell < \kappa - 1/2 \), we have that \( 1 - \ell/(2\kappa) > \ell/(2\kappa) \). Thus, there is a constant \( K \) such that
\[
|\mathcal{C} \cdot |_\ell \leq K | \cdot |_\ell \, .
\]
Item (i) now follows from Condition 3.2. For item (ii) note that, by (28) and Condition 3.2, we have
\[
|\langle f(q), v \rangle| = |\langle \mathcal{C}^{1/2} f(q), \mathcal{C}^{1/2} v \rangle| \leq |\mathcal{C}^{1/2} f(q)| |\mathcal{C}^{1/2} v| \leq K |f(q)|_{\ell} |v|_{\ell} \leq K (1 + |q|_{\ell}) |v|_{\ell} .
\]
The proof of item (iii) is similar. For (iv) we write, by (28) and since \( 0 < \ell < \kappa \),
\[
|\mathcal{C}^{1/2} f(q)| \leq |f(q)|_{\ell} \leq K |f(q)|_{\ell} .
\]
Item (v) is proved in an analogous way.

**Proof of Proposition 3.2.** Lemma 4.1 shows that \( \mathcal{C} f \) is a globally Lipschitz mapping from \( \mathcal{H}^\ell \) into itself. Therefore (16) is an ordinary differential equation in \( \mathcal{H}^\ell \times \mathcal{H}^\ell \) with globally Lipschitz right hand-side which proves directly the statement.

**Proof of Proposition 3.3.** Part (i) is a consequence of (i) in Lemma 4.1. For part (ii) it is clearly sufficient to address the case of the Gaussian law \( \Pi_0 \). From the definition of \( \Psi_h \) as a composition, it is enough to show that \( \Xi_1^1 \) and \( \Xi_2^1 \) defined in (21) and (22) preserve absolute continuity with respect to \( \Pi_0 \). The rotation \( \Xi_2^1 \) leaves \( q \) invariant and thus it suffices to establish that for, any fixed \( q \in \mathcal{H}^\ell \), the mapping \( v \mapsto v + t \mathcal{C} f(q) \) preserves absolute continuity with respect to \( N(0, \mathcal{C}) \). Writing \( \mathcal{C} f(q) = \mathcal{C}^{1/2} \{ \mathcal{C}^{1/2} f(q) \} \), we see from Lemma 4.1(iv) that \( \mathcal{C}^{1/2} f(q) \) is an element of \( \mathcal{H} \); then, the second application of \( \mathcal{C}^{1/2} \) projects \( \mathcal{H} \) onto the Cameron-Martin space of the Gaussian measure \( N(0, \mathcal{C}) \). It is well-known (see e.g. [6, 7]) that translations by elements of the Cameron-Martin space preserve absolute continuity of the Gaussian measure.

Parts (iii) and (iv) are simple consequences of the fact that both \( \Xi_1^1 \) and \( \Xi_2^1 \) are globally Lipschitz continuous with constants of the form \( 1 + \mathcal{O}(|t|) \) as \( t \to 0 \).

14
4.2 Finite-Dimensional Approximations

The proofs of the main Theorems 3.1, 3.2 and 3.3, to be presented in the next subsection, and involve demonstration of invariance or reversibility properties of the algorithmic dynamics, rely on the use of finite-dimensional approximations.

Taking into account the spectral decomposition (12), we introduce the subspaces \((N \in \mathbb{N})\):

\[
\mathcal{H}_N = \{ q \in \mathcal{H} : q = \sum_{j=1}^{N} q_j \phi_j, q_j \in \mathbb{R} \},
\]

and denote by \(\text{proj}_{\mathcal{H}_N}\) the projection of \(\mathcal{H}\) onto \(\mathcal{H}_N\). For \(q, v \in \mathcal{H}\), we also employ the notations:

\[
q^N = \text{proj}_{\mathcal{H}_N}(q), \quad v^N = \text{proj}_{\mathcal{H}_N}(v).
\]

We will make use of the standard isomorphism \(\mathcal{H}_N \leftrightarrow \mathbb{R}^N\) and will sometimes treat a map \(\mathcal{H}_N \to \mathcal{H}_N\) as one \(\mathbb{R}^N \to \mathbb{R}^N\); this should not create any confusion. If we think of elements of \(\mathcal{H}\) as functions of ‘spatial’ variables, then the process of replacing \(\mathcal{H}\) by \(\mathcal{H}_N\) corresponds to space discretization by means of a spectral method.

We introduce the distributions in \(\mathcal{H}_N\) (equivalently, \(\mathbb{R}^N\)) given by

\[
\pi_{0,N}(q) = N(0, C_N), \quad \pi_N(q) \propto \exp\{-\frac{1}{2}(q, C_N^{-1}q) - \Phi_N(q)\}, \quad (29)
\]

where \(C_N\) is the \(N \times N\) diagonal matrix

\[
C_N = \text{diag}\{\lambda_1^2, \lambda_2^2, \ldots, \lambda_N^2\}
\]

and \(\Phi_N\) is the restriction of \(\Phi\) to \(\mathcal{H}_N\), i.e.

\[
\Phi_N(q) = \Phi(q), \quad \text{for } q \in \mathcal{H}_N.
\]

To sample from \(\pi_N\) we reformulate the algorithm in Table 2 in the present finite-dimensional setting. Once more we discuss the flow, the integrator and the accept/reject rule, now for the finite dimensional approximation.

Since, for \(q \in \mathcal{H}_N\), \(D\Phi_N(q) \equiv \text{proj}_{\mathcal{H}_N} D\Phi(q)\), instead of the system (16) we now consider:

\[
\frac{dq}{dt} = v, \quad \frac{dv}{dt} = -q + C \text{proj}_{\mathcal{H}_N} f(q) \quad (30)
\]

(for convenience we have written \(C\) here instead of \(C_N\); both coincide in \(\mathcal{H}_N\)).

The following result, similar to Proposition 3.2 holds:

**Proposition 4.1.**

(i) For any initial condition \((q(0), v(0)) \in \mathcal{H}_N \times \mathcal{H}_N\) and any \(T > 0\) there exists a unique solution of (30) in the space \(C^1([-T, T], \mathcal{H}_N \times \mathcal{H}_N)\).
(ii) Let \( \Xi_N : \mathcal{H}_N \times \mathcal{H}_N \rightarrow \mathcal{H}_N \times \mathcal{H}_N, \ t \in \mathbb{R} \) denote the group flow of (30). The map \( \Xi_N \) is globally Lipschitz with respect to the norm induced by \( \mathcal{H} \times \mathcal{H}^\ell \), with Lipschitz constant of the form \( \exp(K|t|) \) where \( K \) is independent of \( N \) and depends only on \( \pi_0 \) and \( \Phi \).

(iii) For each \( T > 0 \), there exists \( C(T) > 0 \) independent of \( N \) such that for \( 0 \leq t \leq T \) and \( q(0), v(0) \in \mathcal{H}^\ell \), if we set
\[
(q^N(t), v^N(t)) = \Xi_N(q(0), v(0)),
\]
then
\[
|q^N(t)|_\ell + |v^N(t)|_\ell \leq C(T) \left( 1 + |q^N(0)|_\ell + |v^N(0)|_\ell \right)
\leq C(T) \left( 1 + |q(0)|_\ell + |v(0)|_\ell \right),
\]
and, for any \( s \in (\ell, \kappa - 1/2) \),
\[
|q^N(t) - q(t)|_\ell + |v^N(t) - v(t)|_\ell \leq C(T) \left( \frac{1}{N^{\kappa - s - 1}} |q(0)|_s + |v(0)|_s + \frac{1}{N} \left( 1 + |q(0)|_\ell + |v(0)|_\ell \right) \right),
\]
where \( (q(t), v(t)) = \Xi^s(q(0), v(0)) \) is as specified in Proposition 3.2.

Proof. We only derive the approximation result (32): the other statements are standard. We begin with the chain of inequalities (\( K \) will denote a constant independent of \( N \) whose value may vary from one occurrence to the next):
\[
|\mathcal{C}f(q^N(t)) - \mathcal{C}\text{proj}_{\mathcal{H}_N}f(q^N(t))|_\ell^2
\leq K|I - \text{proj}_{\mathcal{H}_N}f(q^N(t))|_\ell^2
\leq \frac{K}{N^{4(\kappa - \ell)}} |f(q^N(t))|_\ell^2
\leq \frac{K}{N^{4(\kappa - \ell)}} \left( 1 + |q^N(t)|_\ell \right)^2
\leq \frac{K}{N^2} \left( 1 + |q^N(t)|_\ell \right)^2
\leq \frac{K}{N^2} \left( 1 + |q(0)|_\ell + |v(0)|_\ell \right)^2,
\]
where we have used successively (28) with \( s = -2\kappa \), basic approximation theory, the facts that (recalling Condition 3.2) \( |D\Phi(q^N(t))|_{-\ell} \leq K \left( 1 + |q^N(t)|_\ell \right) \) and \( 2(\kappa - \ell) > 1 \), and finally (31). Using triangle inequality and Lemma 4.1(i) we may now write
\[
|\mathcal{C}f(q(t)) - \mathcal{C}\text{proj}_{\mathcal{H}_N}f(q^N(t))|_\ell \leq K \left( |q(t) - q^N(t)|_\ell + \frac{1}{N} \left( 1 + |q(0)|_\ell + |v(0)|_\ell \right) \right).
\]
Subtracting the differential equations satisfied by \((q(t), v(t))\) and \((q^N(t), v^N(t))\), a standard Gronwall argument leads to
\[
|q(t) - q^N(t)|_\ell + |v(t) - v^N(t)|_\ell \leq C(T) \left( |q(0) - q^N(0)|_\ell + |v(0) - v^N(0)|_\ell + \frac{1}{N} \left( 1 + |q(0)|_\ell + |v(0)|_\ell \right) \right).
\]
and (32) follows from the approximation theory estimate
\[ |(I - \text{proj}_{H})q(0)|^2 \leq \frac{K}{N^{2(s-\ell)}}|q(0)|^2 \]
in tandem with the corresponding estimate for \( v(0) \).

Clearly (30) is the Hamiltonian system associated with the following Hamiltonian function in \( \mathbb{R}^{2N} \)
\[ H_N(q,v) = \Phi_N(q) + \frac{1}{2} \langle q, C^{-1} q \rangle + \frac{1}{2} \langle v, C^{-1} v \rangle \]
thus, we immediately have the following result:

**Proposition 4.2.** For any \( t \in \mathbb{R} \), the flow \( \Xi_{t/\hbar} \) preserves the probability measure \( \pi_N(dq_0, dv_0) = \exp(-H_N(q,v)) dq dv \).

Note also that \( H_N \) is the restriction to \( H_N \times H_N \) of the Hamiltonian \( H \) in (18).

We will also need to deal with the integrator of (30) and the relevant acceptance probability. By splitting (30) as we did in the case of (16), we construct mappings similar to \( \Psi_h \) and \( \Psi_{\hbar}^{T_h} \) in (23) and (24) respectively. The following definitions will be useful in this context.

**Definition 4.1.**

i) Let \( \Psi_{h,N} : H_N \times H_N \to H_N \times H_N \) be as \( \Psi_h \) in (23) with the only difference that the former has \( C \text{proj}_{H_N} f(q) \) wherever the latter has \( C f(q) \) in (21)). Also, let
\[ \Psi_{h,N}^{(T_h)} = \Psi_{h,N}^{\lfloor T_h \rfloor} . \]

ii) Let \( a_N : H_N \times H_N \to [0,1] \) be defined as \( a \) in (27), (26) but with \( f(\cdot) \) replaced by \( \text{proj}_{H_N} f(\cdot) \) in the latter formula (and with the \( q_i,s, v_i,s \) appearing in (26) now derived by applying iteratively the integrator \( \Psi_{h,N} \)).

The bounds in the following proposition are the discrete-time counterparts of those in Proposition 4.1:

**Proposition 4.3.**

(i) \( \Psi_{h,N}^{(T_h)} \) is a globally Lipschitz map in \( H_N \times H_N \) with respect to the norm induced by \( H^s \times H^s \) with Lipschitz constant of the form \( \exp(KT) \), where \( K \) is independent of \( N \) and depends only on \( \pi_0 \) and \( \Phi \).

(ii) For each \( T > 0 \), there exists \( C(T) > 0 \) independent of \( N \) such that for \( 0 \leq i \leq \lfloor T/\hbar \rfloor \), and \( q_0, v_0 \in H^s \), if we set
\[ (q_i^N, v_i^N) = \Psi_{h,N}^{i}(\text{proj}_{H_N} q_0, \text{proj}_{H_N} v_0) \]
then
\[ |q_i^N| + |v_i^N| \leq C(T)(1 + |q_0^N| + |v_0^N|) \leq C(T)(1 + |q_0| + |v_0|) , \]
and, for \( s \in (\ell, \kappa - \frac{1}{2}) \),
\[ |q_i^N - q_i| + |v_i^N - v_i| \leq C(T)
\left( \frac{1}{N^{2s-\ell}}(|q_0| + |v_0|) + \frac{1}{N}(1 + |q_0| + |v_0|) \right) . \]
Proof. The convergence bound (35) is established by an argument similar to that used for (32). The role played there by Gronwall’s lemma is now played by the stability of the numerical scheme, i.e. by the property in item (i).

The integrator $\Psi_{h,N}^{(T)}$ is time-reversible and, as a composition of Hamiltonian flows, symplectic. As a consequence, it also preserves volume in $\mathbb{R}^{2N}$. In total, $\Psi_{h,N}^{(T)}$ and the acceptance probability $a_N$ can be brought together to formulate an HMC sampling algorithm in $\mathbb{R}^N$ similar to that in Table 2.

**Proposition 4.4.** The algorithm in $\mathbb{R}^N$ with proposal $(q^*, p^*) = \Psi_{h,N}^{(T)}(q^{(n)}, v^{(n)})$, where $v^{(n)} \sim N(0, C_N)$, and acceptance probability $a_N(q^{(n)}, v^{(n)})$ gives rise to a Markov chain reversible under the distribution $\pi_N$ in (29).

Proof. In view of the reversibility and volume preservation of the integrator, this algorithm corresponds to a standard HMC algorithm on Euclidean space, so the required result follows directly from known general properties of the HMC algorithm, see e.g. [8].

4.3. Invariance of Measures

This subsection contains the proofs of Theorems 3.1 and 3.3. The proof of Theorem 3.2 is similar and will be omitted.

**Proof of Theorem 3.1.** We wish to show that, for any bounded continuous function $g : H^\ell \times H^\ell \to \mathbb{R}$ and for any $t \in \mathbb{R}$,

$$\int_{H^\ell \times H^\ell} g(\Xi_t(q,v)) \Pi(dq, dv) = \int_{H^\ell \times H^\ell} g(q,v) \Pi(dq, dv) .$$

or, equivalently, that:

$$\int_{H^\ell \times H^\ell} g(\Xi_t(q,v)) e^{-\Phi(q)} \Pi_0(dq, dv) = \int_{H^\ell \times H^\ell} g(q,v) e^{-\Phi(q)} \Pi_0(dq, dv) .$$  \hspace{1cm} (36)

First observe that it suffices to prove that

$$\int_{H^\ell \times H^\ell} g(\Xi_N(q^N,v^N)) e^{-\Phi(q^N)} \Pi_0(dq, dv) = \int_{H^\ell \times H^\ell} g(q^N,v^N) e^{-\Phi(q^N)} \Pi_0(dq, dv) .$$  \hspace{1cm} (37)

This follows by dominated convergence: analytically, from Condition 3.3 the integrands $g(\Xi_N(q^N,v^N)) e^{-\Phi(q^N)}$ and $g(q^N,v^N) e^{-\Phi(q^N)}$ are both dominated by $K \exp(\epsilon |q_1^2|)$ for some $K = K(\epsilon)$, which is integrable with respect to $\Pi_0$ by Fernique’s theorem; they also converge pointwise $\Pi_0$-a.s. to their counterparts in (36) by virtue of (32), continuity of $g$ and continuity of $\Phi$ from Condition 3.2 (to see this, recall that $\Pi_0$ is supported in $H^s \times H^s$ for any $s \in (\ell, \kappa - \frac{1}{2})$, cf. Proposition 3.1).
Thus it remains to establish (37). This identity may be rewritten as
\[
\int_{\mathbb{R}^N \times \mathbb{R}^N} g(\Xi) (d\mu_N(dq)v_N)\pi_N(dq)\pi_0(dv) = \int_{\mathbb{R}^N \times \mathbb{R}^N} g(q,v)\pi_N(dq)\pi_0(dv).
\]
But \(\pi_N(dq)\pi_0(dv) = \exp(-H_N(q,v))dq\ dv\) and, from Proposition 4.2, this measure is preserved by \(\Xi\).

We now turn to Theorem 3.3. The dynamics of the HMC Markov chain on \(H_e\) described in Table 2 correspond to the following one-step transitions:

\[
q_1 = \begin{cases} 1 \text{ if } U \leq a \end{cases} P(q,\Psi_T(q, v)) + \begin{cases} 1 \text{ if } U > a \end{cases} q_0, \tag{38}
\]

where \(U \sim U[0,1]\), \(v_0 \sim N(0,\mathcal{C})\) and \(a = 1 \wedge \exp(-\Delta H(q_0, v_0))\); also \(P(q)\) denotes the projection \(P(q, v) = q\). Here \(\Delta H\) is given by (26) and \(U\) and \(v_0\) are independent. Our last goal is to prove that this Markov chain is reversible under \(\pi(q)\), that is:

\[
\pi(dq) P(q, dq') = \pi(dq') P(dq', q) \tag{39}
\]

with \(P(\cdot, \cdot)\) being the transition kernel corresponding to the Markov dynamics (38). We begin with the following standard result:

**Lemma 4.2.** The detailed balance equation (39) is satisfied if and only:

\[
I(g) = I(g^\top)
\]

for any continuous bounded function \(g: H_e \times H_e \to \mathbb{R}\), where:

\[
I(g) = \int_{H_e \times H_e} g(q, P(q, \Psi_T(q, v))) a(q, v) e^{-\Phi(q)} \pi_0(dq)\pi_0(dv) \tag{40}
\]

and \(g^\top(q, q') := g(q', q)\).

**Proof.** The two probability measures in (39) are equal if and only if (see e.g. [15]):

\[
\int_{H_e \times H_e} g(q, q') \pi(dq) P(q, dq') = \int_{H_e \times H_e} g(q, q') \pi(dq') P(dq', q)
\]

for any continuous bounded \(g\). In terms of expectations, this can be equivalently re-written as:

\[
E[g(q_0, q_1)] = E[g^\top(q_0, q_1)],
\]

with \(q_0 \sim \pi\) and \(q_1 | q_0\) determined via (38). Integrating out \(U\), we get that:

\[
E[g(q_0, q_1)] = E[g(q_0, P(q, \Psi_T(q, v_0))) a(q_0, v_0)] + E[g(q_0, (1 - a(q_0, v_0))] .
\]

The desired result follows from the fact that the second expectation on the right hand side will not change if we replace \(g \leftrightarrow g^\top\). \(\square\)
We first apply this lemma in the discretized finite-dimensional setting. Recall the definition of $a_N$ in Definition 4.1. Taking into account the reversibility of the discretized chain with respect to $\pi_N$ (Proposition 4.4) we have that, for any continuous and bounded function $\hat{g} : \mathbb{R}^N \times \mathbb{R}^N \rightarrow \mathbb{R}$:

$$\int_{\mathbb{R}^{2N}} \hat{g}(q, P_q(\Psi_{h,N}^{(q,v)}(q,v)))a_N(q,v)\pi_N(dq)\pi_N,0(dv) =$$

$$\int_{\mathbb{R}^{2N}} \hat{g}^\top(q, P_q(\Psi_{h,N}^{(q,v)}(q,v)))a_N(q,v)\pi_N(dq)\pi_N,0(dv)$$

and after selecting

$$\hat{g}(q,v) = g\left(\sum_{j=1}^{N} q_j \phi_j, \sum_{i=1}^{N} v_j \phi_j\right)$$

we reach the conclusion

$$I_N(g) = I_N(g^\top) \quad (41)$$

where

$$I_N(g) = \int_{\mathcal{H}^s \times \mathcal{H}^t} g(q^N, P_q \Psi_{h,N}^{(q,v)}(q^N,v^N))a_N(q^N,v^N)e^{-\Phi(q^N)} \pi_0(dq)\pi_0(dv). \quad (42)$$

The idea now is to conclude the proof by taking the limit $N \rightarrow \infty$ in (41) to show that $I(g) = I(g^\top)$.

**Lemma 4.3.** As $N \rightarrow \infty$ then

$$P_q \Psi_{h,N}^{(q,v)}(q^N,v^N) \rightarrow P_q \Psi_{h}^{(q,v)}, \quad a_N(q^N,v^N) \rightarrow a(q,v), \quad \Pi_0\text{-almost surely.}$$

**Proof.** The first result follows directly from the bound (35), since $\Pi_0$ is concentrated in $\mathcal{H}^s$ for any $s < \kappa - 1/2$. We proceed to the second result. We define $q_i^N, v_i^N$ as in (33) and $q_i, v_i$ as in (25), where (for both cases) now the starting positions are $q$ and $v$ (instead of $q_0, v_0$ appearing in the definitions). As a direct consequence of the definition of $a(q,v)$ and $a_N(q^N,v^N)$, to prove the required result it suffices to show the following statements are true $\Pi_0$-a.s.:

$$\Phi(q_i^N) - \Phi(q_i) \rightarrow 0;$$

$$|C^2_{\text{proj}}_{\mathcal{H}_N} f(q_i^N)|^2 \rightarrow |C^2 f(q_i)|^2;$$

$$\langle \text{proj}_{\mathcal{H}_N} f(q_i^N), v_i^N \rangle \rightarrow \langle f(q_i), v_i \rangle.$$

The first of these results follows directly from the continuity of $\Phi$ in Condition 3.2 and (35). For the other two limits, we observe that from the continuity properties of the involved functions in Lemma 4.1(iii) and (iv), it suffices to prove the following:

$$|C^2_{\text{proj}}_{\mathcal{H}_N} f(q_i^N)|^2 - |C^2 f(q_i^N)|^2 \rightarrow 0;$$
\[ \langle \text{proj}_{\mathcal{H}_N} f(q^N_i), v^N_i \rangle - \langle f(q^N_i), v^N_i \rangle \to 0. \]

For the first of these, note that:

\[ |C^\sharp \text{proj}_{\mathcal{H}_N} f(q^N_i)| + |C^\sharp f(q^N_i)| \leq 2|C^\sharp f(q^N_i)| \leq K \left( 1 + |q_0|\ell + |v_0|\ell \right) \]

by Lemma 4.1(iv) and (34). Now, using in succession Condition 3.1, standard approximation theory, and Condition 3.2 with (34), we obtain:

\[ |C^\sharp (I - \text{proj}_{\mathcal{H}_N}) f(q^N_i)| \leq K |(I - \text{proj}_{\mathcal{H}_N}) f(q^N_i)|_{-\kappa} \leq \frac{K}{N^{\kappa-\ell}} |f(q^N_i)|_{-\ell} \leq \frac{K}{N^{\kappa-\ell}} (1 + |q_0|\ell + |v_0|\ell). \]

Since \( \kappa - \ell > \frac{1}{2} \), the desired convergence follows. For the remaining limit, note that the difference can be bounded by

\[ |(I - \text{proj}_{\mathcal{H}_N}) f(q^N_i)|_{-\ell} \leq |v^N_i|_\ell. \]

Both \( |v^N_i|_\ell \) and \( |q^N_i|_\ell \) are bounded independently of \( N \) and hence Condition 3.2 shows that \( |f(q^N_i)|_{-\ell} \) is bounded independently of \( N \). It follows that

\[ |(I - \text{proj}_{\mathcal{H}_N}) f(q^N_i)|_{-\ell} \to 0 \]

as \( N \to \infty \) and hence the desired result holds true.

\[ \square \]

**Proof of Theorem 3.3.** Using Lemma 4.3 and the continuity of \( \Phi \) from Condition 3.2, the integrand in \( I_N(g) \) (see (42)) converges \( \Pi_0 \)-a.s. to the integrand of \( I(g) \) (see (40)). Also, for every \( \epsilon > 0 \), there is \( K = K(\epsilon) \) such that the former integrand is dominated by \( K \exp(\epsilon |q_0|^2) \), by Condition 3.3. Since \( \pi_0(\mathcal{H}_\ell) = 1 \), Fernique’s theorem enables us to employ dominated convergence to deduce that

\[ I_N(g) \to I(g). \]

Equation (41) and Lemma 4.2 now prove the theorem.

\[ \square \]

5. Numerical Illustrations

We present two sets of numerical experiments which illustrate the performance of the function space HMC algorithm suggested in this paper. In Subsection 5.1 we compare the new algorithm with the standard HMC method; this experiment illustrates that use of the new algorithm removes the \( N^2 \) dependence on the number steps that is required for application of the standard HMC method to measures arising in high dimension \( N \) [1]. The experiment in Subsection 5.2 compares the new HMC method on Hilbert space with a Hilbert space Langevin MCMC algorithm introduced in [3]. Neither of these Hilbert space algorithms exhibit \( N \) dependence in the required number of steps, precisely because they are defined in the limit \( N \to \infty \); however the experiments show the clear advantage of the HMC method in alleviating the random walk behaviour of algorithms, such as those using Langevin proposals, which are based on local moves.
5.1. Comparison with Standard HMC

Consider the target distribution \( \pi \) in the space \( \ell^2 \) of squared integrable sequences (see (13)):

\[
\frac{d\pi}{d\pi_0}(q) \propto \exp\left\{ -\frac{1}{2} \langle q, C^{-\alpha/2} q \rangle \right\}
\]  

(43)

with the reference Gaussian measure given by:

\[
\pi_0 = N(0, \mathcal{C}); \quad \mathcal{C} = \text{diag}\{j^{-2\kappa}; j \geq 1\}.
\]

Since \( \pi \) is itself Gaussian, with independent co-ordinates, it may be easily sampled using standard approaches. However, it provides a useful test case on which to illustrate the differences between standard HMC and our new HMC method. We start by discussing Conditions 3.1, 3.2 and 3.3 for this problem. Clearly, \( \{j^{-2\kappa}\} \) are precisely the eigenvalues of \( \mathcal{C} \) and Condition 3.1, which is necessary and sufficient for \( \mathcal{C} \) to be a trace-class operator on \( \ell^2 \), becomes

\[
\kappa > 1/2;
\]

(44)

recall also that \( \pi_0 \) will be concentrated in \( \mathcal{H}^s \) for any \( s < \kappa - 1/2 \). Notice now that \( \Phi(q) < \infty \) if and only if \( q \in \mathcal{H}^{\kappa \alpha/2} \), thus we must restrict \( \alpha \) to satisfy \( \kappa \alpha/2 < \kappa - 1/2 \), ie.

\[
\alpha < 2 - \frac{1}{\kappa},
\]

(45)

to ensure that \( \Phi(q) < \infty \), \( \pi_0 \)-a.s.. So, in this set-up we can chose any

\[
\ell \in [\kappa \alpha/2, \kappa - 1/2) \cap \mathbb{R}_+
\]

and work on \( \mathcal{H}^\ell \). With regard to Condition 3.2, clearly \( \Phi \) is continuous on \( \mathcal{H}^\ell \) as an inner product; also, it may be checked that the fact that \( \ell > \kappa \alpha/4 \) implies the global Lipschitz continuity property stated for

\[
f(q) = -D\Phi(q) = \mathcal{C}^{-\alpha/2} q
\]

in Condition 3.2 with the Lipschitz constant \( K = 1 \). Condition 3.3 is trivially satisfied since \( \Phi \) here is lower bounded. In total, specification of \( \kappa \) and \( \alpha \) under the restrictions (44) and (45) sets the target (43) into the context of the general algorithm presented in the previous sections, with all relevant conditions satisfied.

The problem is discretized by the spectral technique which we introduced in Subsection 4.2 for theoretical purposes. Because of the product structure of the target, the resulting sampling methods then correspond to applying either the standard or the new HMC as described in Tables 1 or 2 to sample from the marginal distribution of the first \( N \) co-ordinates of \( \pi \):

\[
\pi_N(q) \propto \exp\left\{ -\frac{1}{2} \langle q, C_N^{-1} q \rangle - \frac{1}{2} \langle q, C_N^{-\alpha/2} q \rangle \right\}
\]

(46)

where \( C_N = \text{diag}\{j^{-2\kappa}; j = 1, 2, \ldots, N\} \).
We applied the algorithms in Tables 1 (with mass matrix $M = C_N^{-1}$) and 2 to sample, for various choices of $N$, from the target distribution $\pi_N$ in (46) with $\kappa = 1$, $\alpha = 1/2$. We have chosen the following algorithmic parameters: length of integration of Hamiltonian dynamics $T = 1$; discretisation increment $h = 0.2$; number of MCMC iterations $n = 5,000$. Fig. 5.1 shows empirical average acceptance probabilities from applications of the MCMC algorithms for increasing $N = 2^{10}, 2^{11}, \ldots, 2^{20}$. Execution times for the Hilbert-space algorithm were about $2.5 - 3$ times greater than for the standard HMC.

For $N = 2^{10}$ the standard HMC algorithm gives an average acceptance probability of 0.89, whereas the Hilbert-space HMC gives 0.965. Thus the new integrator, with $h = 0.2$, appears to be marginally more accurate than the standard Verlet integrator. Critically, as the dimension increases, the acceptance probability deteriorates for the standard HMC method until it eventually becomes 0. In contrast, the Hilbert-space algorithm is well-defined even in the limit $N = \infty$ and the algorithm does not deteriorate as $N$ grows. Indeed we have proved in Lemma 4.3 that the limit of the acceptance probability as $N \to \infty$ exists and this limiting behaviour is apparent in Figure 5.1. In practice, when applying the standard HMC method a user would have to use smaller $h$ for increasing $N$ (with $h \to 0$ as $N \to \infty$) to attain similar decorrelation to that given by the new HMC method with a fixed step-size $h$. The result is that the new method has smaller asymptotic variance, for given computational work; and this disparity increases with dimension.

We conclude this subsection with some remarks concerning choice of the mass matrix for the standard HMC method. We have given the algorithm the benefit of the choice $M = C_N^{-1}$. This equalizes the frequencies of the Hamiltonian oscillator underlying the HMC method to one, when applied to sample the Gaussian reference measure $\pi_0$. Had we made the choice $M = I$ then the frequencies of this oscillator would have been $\{1, 2, \cdots, N\}$ (for $\kappa = 1$) resulting in the need to scale $h \propto N^{-5/4} = N^{-1} \times N^{-1/4}$ to obtain an order one acceptance probability. Intuitively the factor $N^{-1}$ comes from a stability requirement\(^3\) required to control integration of fast frequencies, whilst the factor of $N^{-1/4}$ comes (as for the choice $M = C_N^{-1}$) from an accuracy requirement related to controlling deviations in the energy in the large $N$ limit. The choice $M = C_N^{-1}$ may be viewed as a preconditioner and the choice $M = I$ the unpreconditioned case. This viewpoint is discussed for MALA algorithms in [3] where, for the same $\pi_0$ used here, the preconditioned method requires the relevant proposal timestep $\Delta t$ to be chosen as $\Delta t \propto N^{-1/3}$ whilst the unpreconditioned method requires $\Delta t \propto N^{-7/3}$.

5.2. Comparison with a MALA Algorithm

This subsection is devoted to a comparison of the new method with a MALA (Langevin based) MCMC method, also defined on the infinite dimensional space,

\[^3\]Analogous to a Courant-Friedrichs-Lewy condition in the numerical approximation of PDEs
Figure 1: Empirical average acceptance probabilities corresponding to implementations of the standard and Hilbert-space HMC (with $h = 0.2$ and $T = 1$, and $n = 5,000$ iterations) with target distribution $\pi_N$ in (46) (with $\kappa = 1, \alpha = 1/2$), for $N = 2^{10}, 2^{11}, \ldots, 2^{20}$.

as introduced in [3]. As the methods are both defined on Hilbert space no $N-$dependence is expected for either of them. We illustrate the fact that the HMC method breaks the random walk type behaviour resulting from the local moves used in the Langevin algorithm and can consequently be far more efficient, in terms of asymptotic variance per unit of computational cost. It is pertinent in this regard to note that the cost of implementing one step of the new HMC Markov chain from Table 2 is roughly equivalent (in fact slightly less than) the cost of $T/h$ Langevin steps with time-step $\Delta t \propto h^2$. This is because the HMC algorithm can be implemented by a straightforward adaptation of the Langevin code, as noted for the standard HMC and Langevin methods in [13]. This follows from the fact that using the HMC method (standard or Hilbert space versions) with one step of integration ($T = h$) is equivalent to use of the MALA method with a time-step $\Delta t \propto h^2$.

The target measure that we study is defined via a bridge diffusion. Consider the stochastic differential equation

$$dq(\tau) = -V'(q(\tau))d\tau + \sqrt{20}dW(\tau)$$

subject to the end-point conditions $q(0) = q(20) = 0$ and with $V(u) = (u^2 - 1)^2$. Use of the Girsanov formula, together with an integration by parts using the Itô formula, shows [3] that the resulting probability measure for $u \in L^2((0, 20); \mathbb{R})$ may be written in the form (1) with $\pi_0$ Brownian bridge measure on $(0, 20)$ and

$$\Phi(q) = \int_0^{20} \frac{1}{2} \left( |V'(q(\tau))|^2 - 20 V''(q(\tau)) \right) d\tau.$$
The precision operator for Brownian bridge is the second order differential operator \( L = -d^2/d\tau^2 \) with domain \( H^2_0(I) \cap H^2(I) \) and \( I = (0, 20) \). The reference measure \( \pi_0 = N(0, L^{-1}) \) hence satisfies Condition 3.1 with \( \kappa = 1 \). Using the polynomial properties of \( \Phi \) and Sobolev embedding it is then possible to show that Conditions 3.2 and 3.3 are satisfied for suitably chosen \( \ell \), and with the proviso that the Lipschitz property of the force \( f \) is only local. Because of the symmetry of \( V \) and \( \pi_0 \) about the origin it is clear that \( \pi \) is also symmetric about the origin. Thus we may use the HMC and Langevin Markov chains to compute, via the ergodic theorem, an approximation to the mean function under \( \pi \), knowing that the true mean is zero. In this manner we obtain a measure of the asymptotic variance of the two methods.

To implement this comparison the target measure is approximated by a finite difference method employing \( 10^5 \) points in \([0, 20]\) and a value of \( \Delta \tau \) given by \( (\Delta \tau)^2 = 2.0 \times 10^{-4} \). The HMC algorithm from Table 2 is run with a value of \( h = 8.944272 \times 10^{-3} \) (leading to an acceptance rate of 94%), and the Langevin algorithm from [3] with a value of \( \Delta t = 8 \times 10^{-5} \) (leading to an acceptance rate of 78%). In practice we find that a good rule of thumb is the choice \( T \approx 3.14159 \ldots \) since for \( \Phi = 0 \) this gives anti-correlated samples from the Gaussian reference measure. Hence we use a value of \( T \) close to this. We run both algorithms for a number of steps determined so that the computational work for each is almost identical. Comparison of Figures 2 and 3 clearly shows the advantage of the HMC method over the Langevin method, in terms of asymptotic variance for fixed computational cost. This is primarily manifest in the empirical mean which is much closer to the true mean function 0 for HMC than for Langevin. Furthermore, the empirical standard deviations are much closer to their true values, which are shown in Figure 4 (in fact these are computed by running the HMC algorithm until it the ergodic averages have converged, around 10 times as many steps as for Figure 2).

6. Conclusions

We have suggested (see Table 2) and analyzed a generalized HMC algorithm that may be applied to sample from Hilbert-space probability distributions \( \pi \) defined by a density with respect to a Gaussian measure \( \pi_0 \) as in (1). In practice the algorithm has to be applied to a discretized \( N \)-dimensional version \( \pi_N \) of \( \pi \), but the fact that the algorithm is well defined in the limit case \( N = \infty \) ensures that its performance when applied to sample from \( \pi_N \) does not deteriorate as \( N \) increases. In this way, and as shown experimentally in Section 5, the new algorithm eliminates a shortcoming of the standard HMC when used for large values of \( N \). On the other hand, we have also illustrated numerically how the algorithm suggested here benefits from the rationale behind all HMC methods: the capability of taking nonlocal steps when generating proposals alleviates the random-walk behaviour of other MCMC algorithms; more precisely we have shown that the Hilbert-space HMC method clearly improves on a Hilbert-space MALA counterpart in an example involving conditioned diffusions.

In order to define the algorithm we have successively addressed three issues:
Figure 2: Empirical mean function, and empirical standard deviation functions, for Hilbert space-valued HMC algorithm.

Figure 3: Empirical mean function, and empirical standard deviation functions, for Hilbert space-valued Langevin algorithm.
• The definition of a suitable enlarged phase space for the variables $q$ and $v = dq/dt$ and a corresponding measures $\Pi_0$ and $\Pi$ having $\pi_0$ and $\pi$, respectively, as marginals on $q$. The probability measure $\Pi$ is left invariant by the flow of an appropriate Hamiltonian system. Since the Hamiltonian itself is almost surely infinite under $\Pi$ this result is proved by finite dimensional approximation and passage to the limit.

• A geometric numerical integrator to simulate the flow. This integrator is reversible and symplectic and, when applied in finite-dimensions, also volume-preserving. It preserves the measure $\Pi$ exactly in the particular case $\Pi = \Pi_0$ and approximately in general. The integrator is built on the idea of Strang’s splitting, see e.g. [18], [10]; more sophisticated splittings are now commonplace and it would be interesting to consider them as possible alternatives to the method used here.

• We have provided an accept/reject strategy that results in an algorithm that generates a chain reversible with respect to $\pi$. Here we note that straightforward generalizations of the formulae employed to accept/reject in finite dimensions are not appropriate in the Hilbert space context, as the Hamiltonian (energy) is almost surely infinite. However for our particular splitting method the energy difference along the trajectory is finite almost surely, enabling the algorithm to be defined in the Hilbert space setting.

There many interesting avenues for future research opened up by the research.

Figure 4: Mean function and standard deviation functions under the target measure.
contained herein. In particular we intend to use the new HMC method to study a variety of applications with the structure given in (1), such as molecular dynamics and inverse problems in partial differential equations, with applications to fluid mechanics and subsurface geophysics. We also intend to study further enhancements of the new HMC method, for example by means of more sophisticated time-integration methods.

Appendix: Hamiltonian formalism

In this appendix we have collected a number of well-known facts from the Hamiltonian formalism (see e.g. [18]) that, while being relevant to the paper, are not essential for the definition of the Hilbert-space algorithm.

To each real-valued function $H(z)$ on the Euclidean space $\mathbb{R}^{2N}$ there correspond a canonical Hamiltonian system of differential equations:
\[
\frac{dz}{dt} = J^{-1} \nabla_z H(z),
\]
where $J$ is the skew-symmetric matrix
\[
J = \begin{pmatrix} 0_N & -I_N \\ I_N & 0_N \end{pmatrix}.
\]
This system conserves the value of $H$, i.e. $H(z(t))$ remains constant along solutions of the system. Of more importance is the fact that the flow of the canonical equations preserves the standard or canonical symplectic structure in $\mathbb{R}^{2N}$, defined by the matrix $J$, or, in the language of differential forms, preserves the associated canonical differential form $\Omega$. As a consequence the exterior powers $\Omega^n$, $n = 2, \ldots, N$ are also preserved (Poincaré integral invariants). The conservation of the $N$-th power corresponds to conservation of the volume element $dz$. For the Hamiltonian function (3), with $z = (q, p)$, the canonical system is given by (4).

There are many non-canonical symplectic structures in $\mathbb{R}^{2N}$. For instance, the matrix $J$ may be replaced by
\[
\tilde{J} = \begin{pmatrix} 0_N & -L \\ L & 0_N \end{pmatrix},
\]
with $L$ an invertible symmetric $N \times N$ real matrix. Then the Hamiltonian system corresponding to the Hamiltonian function $H$ is given by
\[
\frac{dz}{dt} = \tilde{J}^{-1} \nabla_z H(z).
\]
Again $H$ is an invariant of the system and there is a differential form $\tilde{\Omega}$ that is preserved along with its exterior powers $\tilde{\Omega}^n$, $n = 2, \ldots, N$. The $N$-th power is a constant multiple of the volume element $dz$ and therefore the standard

---

\footnote{In fact, any two non-zero $2N$-forms in $\mathbb{R}^{2N}$ differ only by a constant factor.}
volume is also preserved. With this terminology, the system (8) for the unknown $z = (q, v)$, that was introduced through a change of variables in the canonical Hamiltonian system (7), is a (non-canonical) Hamiltonian system on its own right for the Hamiltonian function (9).

These considerations may be extended to a Hilbert space setting in an obvious way. Thus (16) is the Hamiltonian system in $\mathcal{H} \times \mathcal{H}$ arising from the Hamiltonian function $H$ in (18) and the structure operator matrix

$$\hat{J} = \begin{pmatrix} 0 & -C^{-1} \\ C^{-1} & 0 \end{pmatrix}. $$

However both $H$ and the bilinear symplectic form defined by $\hat{J}$, though densely defined in $\mathcal{H} \times \mathcal{H}$, are almost surely infinite in our context, as they only make sense in the Cameron-Martin space.

The splitting of (16) into (19) and (20) used to construct the Hilbert space integrator corresponds to the splitting

$$H = H_1 + H_2, \quad H_1(q, v) = \Phi(q), \quad H_2(q, v) = \frac{1}{2} \langle v, C^{-1} v \rangle + \frac{1}{2} \langle q, C^{-1} q \rangle$$

of the Hamiltonian function and therefore the flows $\Xi_1$ and $\Xi_2$ in (21) and (22) are symplectic. The integrator $\Psi_h$ is then symplectic as composition of symplectic mappings.

References


