

ON THE ACCEPT-REJECT MECHANISM FOR METROPOLIS-HASTINGS ALGORITHMS

BY NATHAN GLATT-HOLTZ^{1,a}, JUSTIN KROMETIS^{2,b} AND CECILIA MONDAINI^{3,c}

¹*Department of Mathematics, Tulane University, ^anegh@tulane.edu*

²*National Security Institute and Department of Mathematics, Virginia Tech, ^bjkrometi@vt.edu*

³*Department of Mathematics, Drexel University, ^ccf823@drexel.edu*

This work develops a powerful and versatile framework for determining acceptance ratios in Metropolis–Hastings-type Markov kernels widely used in statistical sampling problems. Our approach allows us to derive new classes of kernels which unify random walk or diffusion-type sampling methods with more complicated “extended phase space” algorithms based around ideas from Hamiltonian dynamics. Our starting point is an abstract result developed in the generality of measurable state spaces that addresses proposal kernels that possess a certain involution structure. Note that, while this underlying proposal structure suggests a scope which includes Hamiltonian-type kernels, we demonstrate that our abstract result is, in an appropriate sense, equivalent to an earlier general state space setting developed in (*Ann. Appl. Probab.* **8** (1998) 1–9) where the connection to Hamiltonian methods was more obscure.

On the basis of our abstract results we develop several new classes of extended phase space, HMC-like algorithms. First we tackle the classical finite-dimensional setting of a continuously distributed target measure. We then consider an infinite-dimensional framework for targets which are absolutely continuous with respect to a Gaussian measure with a trace-class covariance. Each of these algorithm classes can be viewed as “surrogate-trajectory” methods, providing a versatile methodology to bypass expensive gradient computations through skillful reduced order modeling and/or data driven approaches as we begin to explore in a forthcoming companion work (Glatt-Holtz et al. (2023)). On the other hand, along with the connection of our main abstract result to the framework in (*Ann. Appl. Probab.* **8** (1998) 1–9), these algorithm classes provide a unifying picture connecting together a number of popular existing algorithms which arise as special cases of our general frameworks under suitable parameter choices. In particular we show that, in the finite-dimensional setting, we can produce an algorithm class which includes the Metropolis adjusted Langevin algorithm (MALA) and random walk Metropolis method (RWM) alongside a number of variants of the HMC algorithm including the geometric approach introduced in (*J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** (2011) 123–214). In the infinite-dimensional situation, we show that the algorithm class we derive includes the preconditioned Crank–Nicolson (pCN), ∞ MALA and ∞ HMC methods considered in (*Stoch. Dyn.* **8** (2008) 319–350; *Stochastic Process. Appl.* **121** (2011) 2201–2230; *Statist. Sci.* **28** (2013) 424–446) as special cases.

CONTENTS

1. Introduction	5280
1.1. Background: Formulating the accept–reject mechanism	5281
1.2. Overview of our contribution	5284

Received October 2021; revised November 2022.

MSC2020 subject classifications. 65P10, 65C05.

Key words and phrases. Markov chain Monte Carlo (MCMC) algorithms, Metropolis–Hastings algorithms, sampling on abstract state spaces, Hamiltonian Monte Carlo, surrogate trajectory methods.

- 2. General formulation of the accept–reject mechanism 5286
 - 2.1. Preliminaries on measure theory 5286
 - 2.2. The main result 5288
- 3. Connection to the Tierney framework 5293
 - 3.1. Overview of Tierney’s formulation 5293
 - 3.2. Reduction to the Tierney formulation 5294
 - 3.3. An alternative proof of Theorem 2.1 5297
- 4. Approximate Hamiltonian Monte Carlo methods 5300
 - 4.1. The finite-dimensional case 5300
 - 4.1.1. Hamiltonian and surrogate extended phase space dynamics 5301
 - 4.1.2. Case 0: General extended phase space methods 5302
 - 4.1.3. Algebraic considerations for numerical splittings 5304
 - 4.1.4. Case 1: Separable surrogate dynamics 5305
 - 4.1.5. Case 2: Splitting into Hamiltonian sub-dynamics 5307
 - 4.1.6. Case 3: Nonseparable dynamics via implicit integrators 5309
 - 4.2. The infinite-dimensional case 5313
- 5. Connections with the classical algorithms 5318
 - 5.1. Finite-dimensional methods 5318
 - 5.1.1. The classical Metropolis–Hastings formulation 5319
 - 5.1.2. Random walk Monte Carlo 5319
 - 5.1.3. The MALA algorithm 5320
 - 5.1.4. Hamiltonian Monte Carlo 5321
 - 5.2. Hilbert space methods 5322
 - 5.2.1. The preconditioned Crank–Nicolson (pCN) algorithm 5322
 - 5.2.2. The infinite-dimensional Metropolis-adjusted Langevin algorithm (∞MALA) 5322
 - 5.2.3. The infinite-dimensional Hamiltonian Monte Carlo (∞HMC) algorithm 5323
- Appendix A: Proof of Theorem 4.12 5324
- Appendix B: Proof of Theorem 4.18 5325
- Appendix C: Proof of Corollary 4.19 5328
- Acknowledgments 5330
- Funding 5330
- References 5330

1. Introduction. A central concern in modern computational probability and statistics is the development of effective sampling methods. This is a nontrivial task particularly for high-dimensional, noncanonical probability distributions with elaborate correlation structures. Indeed, across a great diversity of situations in the pure and applied sciences and in engineering, it is crucial to be able to accurately resolve observable quantities from complex statistical models which naturally arise. As such, sampling is a subject of ubiquitous significance in a wide variety of application settings.

One of the most successful methodologies for sampling is the Markov chain Monte Carlo (MCMC) approach. Starting from a given “target” probability distribution μ sitting on a state space X one aims to find a Markov kernel $P(\mathbf{q}, d\tilde{\mathbf{q}})$ which holds μ as an invariant, that is,

(1.1)
$$\int P(\mathbf{q}, d\tilde{\mathbf{q}})\mu(d\mathbf{q}) = \mu(d\tilde{\mathbf{q}}).$$

By iteratively sampling from such a kernel P as $\mathbf{q}_n \sim P(\mathbf{q}_{n-1}, d\mathbf{r})$ one hopes that P maintains desirable mixing properties so that, for example,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \phi(\mathbf{q}_n) = \int \phi(\mathbf{r})\mu(d\mathbf{r})$$

across a variety of observables $\phi : X \rightarrow \mathbb{R}$.

Of course, deriving such kernels P and determining their effectiveness is an art, the subject of a wide and rapidly growing literature. Here the Metropolis–Hastings method, [43, 57], has

served as a core foundation for the subject. See also, for example, [12, 51, 67] for more general background. The basic idea of this method is to formulate a “proposal” Markov kernel $Q(\mathbf{q}, d\tilde{\mathbf{q}})$ which does not necessarily maintain the invariance (1.1) but which is computationally feasible to sample from. One then introduces an “accept–reject” mechanism which is used to correct for bias in Q with respect to the given target measure μ . In this fashion one builds a “Metropolis–Hastings” kernel P from Q which maintains (1.1) as follows: Starting from a given current state $\mathbf{q}_{n-1}$, one samples a proposal state $\tilde{\mathbf{q}}_n \sim Q(\mathbf{q}_{n-1}, d\mathbf{r})$. Next one determines an acceptance probability $\alpha \in [0, 1]$. The next step in the chain is then set to be $\mathbf{q}_n := \tilde{\mathbf{q}}_n$ with probability α and $\mathbf{q}_n := \mathbf{q}_{n-1}$ otherwise. Of course, the selection of Q and the derivation of the appropriate accept–reject mechanism α depends heavily on the structure of μ and underlying state space X which μ sits upon. Nevertheless, the Metropolis–Hastings methodology encompasses many popular and effective sampling methods, including random walk Monte Carlo (RWMC), the Metropolis adjusted Langevin (MALA) approach based on numerical discretizations of appropriate stochastic dynamics [7, 68], and the Hamiltonian (or Hybrid) Monte Carlo (HMC) algorithm [25, 59].

In this paper, we formulate a simple and quite flexible framework for determining the acceptance ratio α developed in an abstract setting applicable in the generality of measurable spaces. The main result of this work, Theorem 2.1, unifies random walk or diffusion-type approaches with more complicated “extended phase space” algorithms like HMC. In particular, while our framework appears “HMC-like” at first glance, we show that it in fact represents an alternative formulation of the measure-theoretic approach to Metropolis–Hastings kernels introduced in [72] where the connection to Hamiltonian algorithms is more obscure. On the other hand, Theorem 2.1 allows us to derive novel classes of HMC-like algorithms both in the finite and infinite-dimensional settings. These “extended phase space” algorithms provide a versatile methodology to bypass expensive gradient computations through skillful reduced order modeling and/or data driven approaches as we begin to explore in a forthcoming companion work, [34]. Moreover, our algorithms provide a unified picture connecting a number of popular existing algorithms which arise as special cases under suitable parameter choices. Altogether, the theoretical unity and reach of our main result provides a basis for deriving novel sampling algorithms while laying bare important relationships between existing methods.

1.1. Background: Formulating the accept–reject mechanism. Before outlining the main contributions of this work in more detail we first lay out some further background on the Metropolis–Hastings approach to give some context for our results here.

The original setting developed in [43, 57], addresses the case of a continuously distributed target measure $\mu(d\mathbf{r}) = p(\mathbf{r}) d\mathbf{r}$ on a state space $X = \mathbb{R}^N$. Here one considers proposal kernels which are also of the continuous form $Q(\mathbf{q}, d\mathbf{r}) = q(\mathbf{q}, \mathbf{r}) d\mathbf{r}$. A simple calculation aimed at establishing detailed balance for the resulting Markov transition kernel P as described above, namely that

$$(1.2) \quad P(\mathbf{q}, d\tilde{\mathbf{q}})\mu(d\mathbf{q}) = P(\tilde{\mathbf{q}}, d\mathbf{q})\mu(d\tilde{\mathbf{q}}),$$

a condition which is immediately seen to be sufficient for the invariance (1.1), yields the acceptance probability

$$(1.3) \quad \alpha(\mathbf{q}, \tilde{\mathbf{q}}) := 1 \wedge \frac{p(\tilde{\mathbf{q}})q(\tilde{\mathbf{q}}, \mathbf{q})}{p(\mathbf{q})q(\mathbf{q}, \tilde{\mathbf{q}})},$$

where $\mathbf{q}$ is the current state and $\tilde{\mathbf{q}}$ is the proposed next move.

While (1.3) encompasses a number of popular algorithms including the RWMC and MALA methods, the determination of the acceptance probability α can be a much more

complicated and delicate task in other cases of interest, particularly for popular Hamiltonian (HMC) algorithms. Moreover, one is often interested in settings where the target measure sits on a more general state space. Indeed, one class of Metropolis–Hastings algorithms motivating our work here addresses the situation where X is an infinite-dimensional Hilbert space and one considers target measures which are absolutely continuous with respect to a Gaussian probability measure μ_0 so that

$$(1.4) \qquad \mu(d\mathbf{q}) \propto e^{-\Phi(\mathbf{q})} \mu_0(d\mathbf{q}),$$

for an appropriate, μ_0 -integrable potential $\Phi : X \rightarrow \mathbb{R}$. This is an important and rich category of measures which arise naturally in the Bayesian approach to PDE inverse problems, [20, 21, 24, 55, 63, 70] and also in computational chemistry, [38, 39, 41, 42, 66].

One of our aims in this manuscript is to provide new insight into the derivation of a recently discovered collection of Metropolis–Hastings algorithms [8, 10, 11, 22] that provide a basis to effectively sample from such infinite-dimensional measures of the form (1.4). The idea in these works, [10, 11, 22], is to appropriately precondition stochastic or Hamiltonian dynamics related to the target measure μ and then to make a delicate choice for the numerical discretization of these equations so that one obtains an effective proposal kernel Q . As usual, an appropriate α is then introduced to correct for bias stemming from the original dynamics, the numerical discretization of these dynamics, or both. In particular, this infinite-dimensional approach results in an analogue of the random walk Monte Carlo algorithm with the so called preconditioned Crank–Nicolson (pCN) algorithm as well as infinite-dimensional formulations of the MALA [11, 22] and HMC [8, 10] algorithms. Each of these methods has shown great promise by partially beating the “curse of dimensionality” as borne out by recent theoretical developments [16, 26, 35, 40] and by effectively resolving certain challenging test problems [8, 15, 21, 34].

Regarding the derivation of the α for the pCN and for the infinite-dimensional MALA algorithms from [11, 22], one can make use of the abstract formulation due to Tierney [72] to determine α . Tierney’s approach may be seen as an extension of [43] to general state spaces. This is formulated as follows: Given a proposal kernel Q and a target measure μ , if the measures

$$(1.5) \qquad \eta(d\mathbf{q}, d\tilde{\mathbf{q}}) := Q(\mathbf{q}, d\tilde{\mathbf{q}})\mu(d\mathbf{q}), \qquad \eta^\perp(d\mathbf{q}, d\tilde{\mathbf{q}}) := \eta(d\tilde{\mathbf{q}}, d\mathbf{q})$$

are mutually absolutely continuous then one can define the acceptance probability α via the Radon–Nikodym derivative, namely

$$(1.6) \qquad \alpha(\mathbf{q}, \tilde{\mathbf{q}}) := 1 \wedge \frac{d\eta^\perp}{d\eta}(\mathbf{q}, \tilde{\mathbf{q}})$$

to achieve detailed balance à la (1.2). Here note that in the appropriate finite-dimensional setting (1.6) reduces to (1.3).

On the other hand, Tierney’s elegant formulation in [72] does not appear to cover HMC-type “extended phase space” algorithms in an obvious way, even in the original finite-dimensional formulation from [25]. In HMC sampling, the proposal is generated by interpreting the current state $\mathbf{q}$ as a position variable in a Hamiltonian system. One identifies a Hamiltonian $\mathcal{H}$ such that the marginal of the Gibbs measure

$$(1.7) \qquad \mathcal{M}(d\mathbf{q}, d\mathbf{v}) \propto e^{-\mathcal{H}(\mathbf{q}, \mathbf{v})} d\mathbf{q} d\mathbf{v}$$

onto position space corresponds to the desired target measure μ . Augmenting with a “momentum” (or sometimes “velocity”) variable $\mathbf{v}$ sampled from the $\mathbf{v}$ -marginal of $\mathcal{M}$ and then integrating the associate Hamiltonian dynamics in $(\mathbf{q}, \mathbf{v})$ -space via an appropriately chosen approximate integrator $\hat{S}$, one obtains

$$(1.8) \qquad (\tilde{\mathbf{q}}, \tilde{\mathbf{v}}) := \hat{S}(\mathbf{q}, \mathbf{v}).$$

The proposal $\tilde{\mathbf{q}}$ is then given by $\tilde{\mathbf{q}} = \Pi_1 \circ \hat{S}(\mathbf{q}, \mathbf{v})$ where Π_1 represents projection onto the position variable from the extended phase space. Then the acceptance ratio $\hat{\alpha}$ is specified as¹

$$(1.9) \quad \hat{\alpha}(\mathbf{q}, \mathbf{v}) := 1 \wedge \exp[\mathcal{H}(\mathbf{q}, \mathbf{v}) - \mathcal{H}(\hat{S}(\mathbf{q}, \mathbf{v}))],$$

where $\mathcal{H}$ is the Hamiltonian associated with the system.

Note carefully that $\hat{\alpha}$ in (1.9) depends not on the current and proposed states $\mathbf{q}, \tilde{\mathbf{q}}$ as in (1.6), but rather on the current state $\mathbf{q}$ and the initially sampled auxiliary variable $\mathbf{v}$. As such, for Hamiltonian Monte Carlo algorithms, the acceptance probability $\hat{\alpha}$ is not accommodated by the mechanism (1.6), at least not in an obvious fashion. It is also important to emphasize that the validity of this accept-reject mechanism for the classical HMC algorithm (1.9) requires one to formulate a numerical resolution $\hat{S}$ of the Hamiltonian dynamics which respects certain delicate structural properties of the original Hamiltonian system; namely one typically specifies $\hat{S}$ as a “geometric integration scheme,” one which preserves volumes on phase space and whose dynamics maintains certain “reversibility properties” (see (4.10) below). The latter reversibility condition reflects an indispensable underlying involutive structure that we exploit here.

Starting from [25, 59], the core ideas of this Hamiltonian approach have expanded into a profusion of “extended phase space” methods. This literature includes a number of variations on the Hamiltonian and symplectic structure leading to the formulation of (1.8) as well as use of “surrogate dynamics” methods to reduce the cost or complexity of numerically expensive gradient computations which arise. See, for example, [5, 8, 10, 33, 44, 47, 50, 51, 53, 56, 60, 61, 64, 65, 75, 76] and numerous other containing references. This large and rapidly expanding literature is reflective of the success and effectiveness of the Hamiltonian approach as, for example, can be seen in the wide adoption of the STAN software package, [30, 71], in recent years.

It is therefore of great interest to provide a unified theoretical foundation for these various works on extended phase space methods and, if possible, to place them in the context of the original RWMC methods dating all the way back to [43, 57]. A number of recent contributions preceding this current work are notable in this regard. For example, the monograph [17] building on [27] provides a lucid and partially self-contained survey explaining various underlying mechanisms involved in the derivation of (1.9) in the original finite-dimensional setting. On the other hand recent work [3, 49, 62, 64] has identified the connection between HMC methods and the involutive algorithms of Green [31, 32, 36]. Note that after publication of the initial draft of this manuscript, a related work [3] appeared, which contains an abstract result similar to results appearing herein. As such, [3] parallels our focus on providing a unified view of reversible extended phase space Metropolis–Hastings algorithms. Notwithstanding this overlapping scope, our two works diverge in addressing quite different classes of algorithms. [3] explores connections to well-known algorithms within the statistics literature, for example, NUTS [44], Bouncy particle sampler [18] and multiple-try Metropolis [52]. On the other hand, our work contains contemporary examples that are more relevant within applied mathematics by way of Bayesian inverse problems [24, 70]. In particular, the use of Bayesian methods to estimate functional parameters in ordinary and partial differential equations necessitates MCMC algorithms that generalize to (perhaps augmented) infinite-dimensional state spaces.

Regarding the infinite-dimensional version of the HMC algorithm developed in [10] (see also [8, 9]), the finite-dimensional setting is insufficient to derive an appropriate $\hat{\alpha}$ as in

¹For the infinite-dimensional formulation in [10], (1.7) is formal and made sense of with respect to a reference Gaussian measure while $\mathcal{H}$ is typically almost surely infinite. As such, the “change in energy” $\Delta\mathcal{H}(\mathbf{q}, \mathbf{v}) = \mathcal{H}(\mathbf{q}, \mathbf{v}) - \mathcal{H}(\hat{S}(\mathbf{q}, \mathbf{v}))$ in (1.9) is computed via a separate formula that accounts for a certain “cancellation of infinities” rather than by computing $\mathcal{H}$ directly. See Section 4.2 below and also [10] for further details.

(1.9). In the original framing from [10] this infinite-dimensional case was justified via a finite-dimensional approximation scheme reminiscent of an extended body of work on the invariance of the Gibbs measure for Hamiltonian PDEs, see [19] and more recently [6, 58] for a comprehensive survey of this research direction. On the other hand subsequent work [8, 9], provides notable insights into the underlying structural considerations at play in the formulation of the infinite-dimensional setting by making delicate use of the Cameron–Martin theorem.

1.2. *Overview of our contribution.* We turn finally to describe our contributions herein. The main result of this work is Theorem 2.1, which provides a single, simple formulation that subsumes all of the above algorithms and, we expect, many other methods of interest not directly addressed here. The result is developed in the generality of measurable state spaces (X, Σ_X) to reversibly sample from a given target probability measure μ defined on Σ_X .

To proceed we consider Markov kernels formulated as follows: select an additional measurable space (Y, Σ_Y) and form the extended phase space $X \times Y$. Fix any proposal map $S : X \times Y \rightarrow X \times Y$ which is an involution operation, namely such that $S^2 = I$, where we denote here and throughout $S^2 := S \circ S$, and any reference Markov kernel $\mathcal{V} : X \times \Sigma_Y \rightarrow [0, 1]$. We obtain a proposal kernel $\mathcal{Q}(\mathbf{q}, d\tilde{\mathbf{q}})$ from S and $\mathcal{V}$ by sampling $\mathbf{v} \sim \mathcal{V}(\mathbf{q}, d\mathbf{w})$ and then taking a proposed step as $\tilde{\mathbf{q}} = \Pi_1 \circ S(\mathbf{q}, \mathbf{v})$, where Π_1 is the projection onto the “position variable,” namely $\Pi_1(\tilde{\mathbf{q}}, \tilde{\mathbf{v}}) = \tilde{\mathbf{q}}$. Thus, in measure theoretic language, we have $\mathcal{Q}(\mathbf{q}, d\tilde{\mathbf{q}}) = (\Pi_1 \circ S(\mathbf{q}, \cdot))^* \mathcal{V}(\mathbf{q}, d\tilde{\mathbf{q}})$ where $f^* \nu$ denotes the *pushforward* of a measure ν by a function f ; see Section 2.1 below for details.

For such data μ , S and $\mathcal{V}$, our result then determines a suitable acceptance probability as a function of $\mathbf{q}$ and $\mathbf{v}$ given by

$$(1.10) \quad \hat{\alpha}(\mathbf{q}, \mathbf{v}) := 1 \wedge \left(\frac{dS^* \mathcal{M}}{d\mathcal{M}}(\mathbf{q}, \mathbf{v}) \right) \quad \text{where } \mathcal{M}(d\mathbf{q}, d\mathbf{v}) := \mathcal{V}(\mathbf{q}, d\mathbf{v}) \mu(d\mathbf{q})$$

and where $dS^* \mathcal{M}/d\mathcal{M}$ denotes the Radon–Nikodym derivative of the pushforward measure $S^* \mathcal{M}$ with respect to $\mathcal{M}$. Note that this acceptance ratio is well defined when $S^* \mathcal{M}$ is absolutely continuous with respect to $\mathcal{M}$ but this condition can be relaxed, see Remark 2.4 below. By adopting this measure theoretic language of pushforwards and Radon–Nikodym derivatives, as reviewed for our purposes in Section 2.1, Theorem 2.1 affords a simple proof based around the intuition of using $\hat{\alpha}$ to balance inflows and outflows between any two states $\mathbf{q}$ and $\tilde{\mathbf{q}}$ of the Markov chain.

We observe that Theorem 2.1 has a direct interpretation as an abstraction of the HMC method, but one which provides the crucial insight that the role of the “Gibbs measure” as in (1.7) and that of the “numerical integrator” as in (1.8) can be largely disconnected insofar as achieving the detailed balance condition, (1.2), is concerned. In this analogy $\mathcal{M}$, given as in (1.10), specifies the Gibbs measure. Regarding the involution S , we notice that the solution map $\hat{S}$ of a Hamiltonian system or any reasonable numerical resolution thereof is not typically an involution. However, by taking $S := R \circ \hat{S}$, where $R(\mathbf{q}, \mathbf{v}) = (\mathbf{q}, -\mathbf{v})$ is the momentum flip operation, one does indeed obtain an involution for a large class of “geometric” integration schemes $\hat{S}$. Thus, so long as $S^* \mathcal{M}$ is absolutely continuous with respect to $\mathcal{M}$, the algorithmic elements μ , $\mathcal{V}$ and $S = R \circ \hat{S}$ are otherwise unrelated insofar as the scope of Theorem 2.1 is concerned.

With Theorem 2.1 in hand, we proceed to demonstrate the reach and theoretical unity that this result provides by detailing how it can be used to derive and to analyze a variety of specific Metropolis–Hastings-type algorithms. We first observe that we can use Theorem 2.1 to recover the complete framework of the more traditional Metropolis–Hastings techniques up to and including the abstract generality of Tierney’s formulation in [72]. Indeed, in Section 3, we show that the formalisms in [72] and in Theorem 2.1 may be ultimately viewed

as having an equivalent scope. On the other hand the “HMC-like” character of Theorem 2.1 described in the previous paragraph suggests an immediate connection to Hamiltonian-type extended phase space methods. In Section 4, we use this observation as a crucial starting point for deriving a variety of HMC-like sampling methods culminating in the derivation of Algorithm 4.2, Algorithm 4.3, Algorithm 4.4 and Algorithm 4.5 provided below. The final Section 5 outlines multiple ways that a selection of important sampling methods widely considered in the literature may be subsumed under Theorem 2.1. Here to obtain our selection of $\mathcal{V}$ and S for a given target μ , we draw on various results in Section 4 and Section 3.

Regarding the equivalence of [72] and Theorem 2.1 we first consider in Section 3.2 proposal kernels of the form

$$(1.11) \quad Q(\mathbf{q}, d\tilde{\mathbf{q}}) = F(\mathbf{q}, \cdot)^* \mathcal{V}(\mathbf{q}, d\tilde{\mathbf{q}}),$$

for some $F : X \times Y \rightarrow X$ such that $F(\mathbf{q}, \cdot)$ is invertible for each fixed $\mathbf{q}$. We then demonstrate that there exists a unique involution S whose projection onto the position space is F . With this S in hand, we show that the main result in [72], namely the formulation (1.6), follows as a special case of Theorem 2.1. See Proposition 3.3, Theorem 3.4 below for precise details.

Note that the scope of (1.11) trivially encompasses the results in [72] as seen, by selecting, for any given proposal kernel Q , $F(\mathbf{q}, \mathbf{v}) := \mathbf{v}$ and $\mathcal{V} := Q$. On the other hand, numerous relevant examples including RWMC and MALA and even their Hilbert space counterparts, [11, 22], can be recovered from nontrivial formulations of F and $\mathcal{V}$. This leads to the interesting observation that algorithms of interest can be recovered through multiple, nonequivalent applications of Theorem 2.1.

Conversely, in the other direction, we observe in Section 3.3 that the results from [72] can be suitably employed to provide a second, independent, proof of Theorem 2.1. Here, from the given data of Theorem 2.1, namely the target μ , the involution S and kernel $\mathcal{V}$, we proceed by considering a suitable deterministic proposal kernel specified by S acting on the extended phase space. The results in [72] are then applied for $\mu = \mathcal{M}$ on this extended phase space. In this way we obtain a kernel which, when appropriately integrated against $\mathcal{V}$, yields the desired kernel specified by Theorem 2.1. These details are given in Section 3.3 below.

In Section 4 we use Theorem 2.1 and the intuition it provides to develop some classes of concrete, extended-phase space algorithms. We may view our methods as introducing additional “degrees of freedom” for the selection of parameters used for tuning HMC methods. Crucially the approach includes functional parameters which allow greater latitude in selecting the numerical integration procedure used for approximating Hamiltonian dynamics. In particular our formulation provides a flexible means of using reduced order modeling or data driven approaches to avoid expensive gradient computations which arise. As already mentioned above, we begin to explore such applications for these algorithms in a forthcoming companion work [34]. Note furthermore that our methods provide a unified view for an extensive existing literature around so called “surrogate trajectory methods” [8, 47, 50, 51, 56, 60, 61, 64, 65, 75, 76]. On the other hand, in Section 5, we show that a number of classical variants of HMC, MALA, RWMC as well as the Hilbert space methods ∞ HMC, pCN, ∞ MALA all fall as special cases of these algorithms introduced in Section 4, under an appropriate choice of algorithmic parameters.

In Section 4.1 we begin by developing the finite-dimensional setting of a continuously distributed target measure. To this end we provide a brief but self-contained tour of some geometric numerical methods for Hamiltonian systems highlighting how these methods connect back to the setting of Theorem 2.1. In particular this presentation emphasizes the fundamental role that volume preserving methods, palindromic splitting structures and reversibility play in determining the involutive mapping S at the center of Theorem 2.1, while suggesting further scope for deriving other Hamiltonian-type sampling methods in the finite-dimensional

setting. Notably we address situations where the integrator is not assumed to have a gradient structure in Algorithm 4.2 and generalize the use of some implicit integrators developed in [33] culminating in Algorithm 4.4. Furthermore Proposition 4.9 and Theorem 4.12 suggest possible variations on the usual setting of momentum flip reversibility which may be of some use in future algorithmic developments.

In Section 4.2 we turn to address the infinite-dimensional Gaussian setting where our target measure has the form (1.4). Here we derive a quite general surrogate trajectory method in Algorithm 4.5. Note that this algorithm includes ∞ HMC, pCN, ∞ MALA as well as geometric variations from [8] as important special cases. Our Hilbert space approach takes as its starting point the method introduced in [10]. In this Gaussian base measure setting, the implied Hamiltonian from [10] has the form $\mathcal{H} := \mathcal{H}_1 + \mathcal{H}_2$ with

$$(1.12) \qquad \mathcal{H}_1(\mathbf{q}, \mathbf{v}) := \frac{1}{2}|\mathcal{C}^{-1/2}\mathbf{q}|^2 + \frac{1}{2}|\mathcal{C}^{-1/2}\mathbf{v}|^2, \qquad \mathcal{H}_2(\mathbf{q}, \mathbf{v}) := \Phi(\mathbf{q}).$$

The crucial insight in [10] is to use an appropriate preconditioning operator J along with a particular Strang splitting defined around $\mathcal{H}_1$ and $\mathcal{H}_2$ to derive a numerical integrator of the form $\hat{S} = (\Xi_{\delta/2}^{(1)} \circ \Xi_{\delta}^{(2)} \circ \Xi_{\delta/2}^{(1)})^n$ for the algorithmic parameters $\delta > 0$, the size of the numerical time step, and $n = T/\delta$ where $T > 0$ is the total integration time. Here, for any $t > 0$,

$$(1.13) \qquad \begin{aligned} \Xi_t^{(1)}(\mathbf{q}, \mathbf{v}) &= (\mathbf{q}, \mathbf{v} - t\mathcal{C}D\Phi(\mathbf{q})), \\ \Xi_t^{(2)}(\mathbf{q}, \mathbf{v}) &= (\cos(t)\mathbf{q} + \sin(t)\mathbf{v}, -\sin(t)\mathbf{q} + \cos(t)\mathbf{v}). \end{aligned}$$

This delicate choice of splitting and preconditioning results, in the language of our framework in Theorem 2.1, in an involution $S = R \circ \hat{S}$, where again R is the momentum flip operation, such that $S^*(\mu \otimes \mu_0)$ is absolutely continuous with respect to $\mu \otimes \mu_0$ or, in more heuristic but concrete terms, in a “cancellation of infinities” as would appear in (1.9).

Our generalization of [10] to Algorithm 4.5 centers on the observation that we can replace the term $D\Phi$ in (1.13) with essentially any reasonable function $f : X \rightarrow X$ and replace the velocity proposal kernel $\mathcal{V} = \mu_0$ suggested by (1.12) with any $\mathcal{V}$ such that $\mathcal{V}(\mathbf{q}, d\mathbf{v})$ is absolutely continuous with respect to $\mu_0(d\mathbf{v})$ for any $\mathbf{q}$. Indeed, in this more general setting, we still obtain an involution S which maintains the absolute continuity of $S^*\mathcal{M}$ with respect to $\mathcal{M}(d\mathbf{q}, d\mathbf{v}) \propto e^{-\Phi(\mathbf{q})}\mu_0(d\mathbf{q})\mathcal{V}(\mathbf{q}, d\mathbf{v})$ required by Theorem 2.1. As already alluded to above, the proof of invariance of the target μ in [10] follows an involved spectral approximation approach analogous to [19]. Here in Theorem 4.18 we proceed similarly to [9] and provide a different and more direct proof based on the Cameron–Martin theorem which reduces the problem to computing $dS^*(\mu_0 \otimes \mu_0)/d(\mu_0 \otimes \mu_0)$ and making use of a simple identity for iterated pushforward maps, given as (2.8) below. Indeed, our involution S involves repeated applications of the maps $\Xi_t^{(1)}, \Xi_t^{(2)}$; the latter map $\Xi_t^{(2)}$ holds $\mu_0 \otimes \mu_0$ invariant, while the Radon–Nikodym derivative associated with the pushforward of $\mu_0 \otimes \mu_0$ by $\Xi_t^{(1)}$ can be computed via the Cameron–Martin theorem.

2. General formulation of the accept–reject mechanism. This section is devoted to the main result of the paper which we present as Theorem 2.1 in Section 2.2. Before stating and proving this main result we briefly recall some measure theoretic terminologies and facts in Section 2.1. For more details regarding this preliminary material, we refer the reader to, for example, [1, 14, 28].

2.1. Preliminaries on measure theory. Let $(\mathfrak{X}, \Sigma_{\mathfrak{X}})$ and $(\mathfrak{Y}, \Sigma_{\mathfrak{Y}})$ be measurable spaces. Given a measurable function $\phi : \mathfrak{X} \rightarrow \mathfrak{Y}$ and a measure ν on $(\mathfrak{X}, \Sigma_{\mathfrak{X}})$, the *pushforward* of ν by ϕ , denoted as $\phi^*\nu$, is defined as the measure on $\mathfrak{Y}$ given by

$$\phi^*\nu(A) := \nu(\phi^{-1}(A)) \quad \text{for any } A \in \Sigma_{\mathfrak{Y}}.$$

Clearly, for any measurable functions ϕ_1, ϕ_2 mapping between appropriate spaces, we have that

$$(2.1) \quad (\phi_1 \circ \phi_2)^* \nu = \phi_1^*(\phi_2^* \nu).$$

We recall moreover that, given a $(\phi^* \nu)$ -integrable function $\psi : \mathfrak{Y} \rightarrow \mathbb{R}$, that is, $\psi \in L^1(\phi^* \nu)$, it follows that the composition $\psi \circ \phi : \mathfrak{X} \rightarrow \mathbb{R}$ is in $L^1(\nu)$ and the following change of variables formula holds:

$$(2.2) \quad \int_{\mathfrak{Y}} \psi(\mathbf{w}) \phi^* \nu(d\mathbf{w}) = \int_{\mathfrak{X}} \psi(\phi(\mathbf{w})) \nu(d\mathbf{w}).$$

Let us observe that if $\bar{\mathbf{w}}$ is a random variable sampled from a probability measure ν then $\phi(\bar{\mathbf{w}})$ is distributed as $\phi^* \nu$. Furthermore it is worth noticing that, in the special case when $\mathfrak{X} = \mathbb{R}^N$ and ν is any Borel measure on $\mathbb{R}^N$ which is absolutely continuous with respect to Lebesgue measure, namely when

$$\nu(d\mathbf{w}) = p(\mathbf{w}) d\mathbf{w}$$

for some density function $p : \mathbb{R}^N \rightarrow \mathbb{R}$, then for any diffeomorphism $\phi : \mathbb{R}^N \rightarrow \mathbb{R}^N$, we have

$$(2.3) \quad \phi^* \nu(d\mathbf{w}) = p(\phi^{-1}(\mathbf{w})) |\det \nabla \phi^{-1}(\mathbf{w})| d\mathbf{w}.$$

Next recall that, given measures ν and ρ on a measurable space $(\mathfrak{X}, \Sigma_{\mathfrak{X}})$, we say that ν is *absolutely continuous with respect to* ρ , and write $\nu \ll \rho$, if $\nu(A) = 0$ whenever $\rho(A) = 0$, for $A \in \Sigma_{\mathfrak{X}}$. If ν and ρ are two sigma-finite measures on $(\mathfrak{X}, \Sigma_{\mathfrak{X}})$ such that $\nu \ll \rho$ then there exists a ρ -almost unique function $d\nu/d\rho \in L^1(\rho)$ such that

$$(2.4) \quad \nu(A) = \int_A \frac{d\nu}{d\rho}(\mathbf{w}) \rho(d\mathbf{w}), \quad A \in \Sigma_{\mathfrak{X}},$$

called the *Radon–Nikodym derivative* of ν with respect to ρ . In particular, if ν_1, ν_2 and ρ are sigma-finite measures on $(\mathfrak{X}, \Sigma_{\mathfrak{X}})$ with $\nu_1 \ll \rho$ and $\nu_2 \ll \rho$, namely

$$\nu_1(d\mathbf{w}) = \phi_1(\mathbf{w}) \rho(d\mathbf{w}), \quad \nu_2(d\mathbf{w}) = \phi_2(\mathbf{w}) \rho(d\mathbf{w}),$$

where $\phi_1 = d\nu_1/d\rho$ and $\phi_2 = d\nu_2/d\rho$, and if $\phi_2 > 0$ ρ -a.e. then $\nu_1 \ll \nu_2$ and

$$(2.5) \quad \frac{d\nu_1}{d\nu_2}(\mathbf{w}) = \frac{\phi_1(\mathbf{w})}{\phi_2(\mathbf{w})} \quad \text{for } \rho\text{-a.e. } \mathbf{w} \in \mathfrak{X}.$$

It also immediately follows from (2.4) that, given sigma-finite measures ν, ρ and γ on $(\mathfrak{X}, \Sigma_{\mathfrak{X}})$ such that $\nu \ll \rho$ and $\rho \ll \gamma$, then $\nu \ll \gamma$ and

$$(2.6) \quad \frac{d\nu}{d\gamma}(\mathbf{w}) = \frac{d\nu}{d\rho}(\mathbf{w}) \frac{d\rho}{d\gamma}(\mathbf{w}) \quad \text{for } \gamma\text{-a.e. } \mathbf{w} \in \mathfrak{X}.$$

Moreover, given a measurable and invertible mapping $\phi : \mathfrak{X} \rightarrow \mathfrak{X}$ with measurable inverse² $\phi^{-1} : \mathfrak{X} \rightarrow \mathfrak{X}$, and sigma-finite measures ν and ρ on $(\mathfrak{X}, \Sigma_{\mathfrak{X}})$ with $\nu \ll \rho$, it follows that $\phi^* \nu \ll \phi^* \rho$ and

$$(2.7) \quad \frac{d\phi^* \nu}{d\phi^* \rho}(\mathbf{w}) = \frac{d\nu}{d\rho}(\phi^{-1}(\mathbf{w})) \quad \text{for } (\phi^* \rho)\text{-a.e. } \mathbf{w} \in \mathfrak{X}.$$

²Here it is worth pointing out that when $\mathfrak{X}$ is a Polish space then, for any measurable and invertible mapping $\phi : \mathfrak{X} \rightarrow \mathfrak{X}$, its inverse ϕ^{-1} is always a measurable mapping, see, for example, [1], Theorem 12.29.

This follows by noticing that, for any $(\phi^* \nu)$ -integrable function $\psi : \mathfrak{X} \rightarrow \mathbb{R}$, we have from (2.2) and (2.4) that

$$\begin{aligned} \int_{\mathfrak{X}} \psi(\mathbf{w}) \phi^* \nu(d\mathbf{w}) &= \int_{\mathfrak{X}} \psi(\phi(\mathbf{w})) \nu(d\mathbf{w}) = \int_{\mathfrak{X}} \psi(\phi(\mathbf{w})) \frac{d\nu}{d\rho}(\mathbf{w}) \rho(d\mathbf{w}) \\ &= \int_{\mathfrak{X}} \psi(\mathbf{w}) \frac{d\nu}{d\rho}(\phi^{-1}(\mathbf{w})) \phi^* \rho(d\mathbf{w}). \end{aligned}$$

Finally we observe that, given a sigma-finite measure ν on $(\mathfrak{X}, \Sigma_{\mathfrak{X}})$, and a sequence $\phi_i : \mathfrak{X} \rightarrow \mathfrak{X}, i = 1, \dots, n$, of measurable and invertible functions with measurable inverses $\phi_i^{-1} : \mathfrak{X} \rightarrow \mathfrak{X}$ and such that $\phi_i^* \nu \ll \nu$ for each $i = 1, \dots, n$, then $(\phi_n \circ \dots \circ \phi_1)^* \nu \ll \nu$ and

$$(2.8) \quad \frac{d(\phi_n \circ \dots \circ \phi_1)^* \nu}{d\nu}(\mathbf{w}) = \frac{d\phi_n^* \nu}{d\nu}(\mathbf{w}) \prod_{i=1}^{n-1} \frac{d\phi_i^* \nu}{d\nu}((\phi_n \circ \dots \circ \phi_{i+1})^{-1}(\mathbf{w}))$$

for ν -a.e. $\mathbf{w} \in \mathfrak{X}$.

To derive this identity, notice that for every $((\phi_n \circ \dots \circ \phi_1)^* \nu)$ -integrable function $\psi : \mathfrak{X} \rightarrow \mathbb{R}$ we have

$$\begin{aligned} &\int \psi(\mathbf{w}) (\phi_n \circ \dots \circ \phi_2 \circ \phi_1)^* \nu(d\mathbf{w}) \\ &= \int \psi(\phi_n \circ \dots \circ \phi_2(\mathbf{w})) \phi_1^* \nu(d\mathbf{w}) = \int \psi(\phi_n \circ \dots \circ \phi_2(\mathbf{w})) \frac{d\phi_1^* \nu}{d\nu}(\mathbf{w}) \nu(d\mathbf{w}) \\ &= \int \psi(\phi_n \circ \dots \circ \phi_3(\mathbf{w})) \frac{d\phi_1^* \nu}{d\nu}(\phi_2^{-1}(\mathbf{w})) \phi_2^* \nu(d\mathbf{w}) \\ &= \int \psi(\phi_n \circ \dots \circ \phi_3(\mathbf{w})) \frac{d\phi_1^* \nu}{d\nu}(\phi_2^{-1}(\mathbf{w})) \frac{d\phi_2^* \nu}{d\nu}(\mathbf{w}) \nu(d\mathbf{w}) \\ &= \int \psi(\phi_n \circ \dots \circ \phi_4(\mathbf{w})) \frac{d\phi_1^* \nu}{d\nu}(\phi_2^{-1} \circ \phi_3^{-1}(\mathbf{w})) \frac{d\phi_2^* \nu}{d\nu}(\phi_3^{-1}(\mathbf{w})) \phi_3^* \nu(d\mathbf{w}) \\ &= \dots = \int \psi(\mathbf{w}) \left(\prod_{i=1}^{n-1} \frac{d\phi_i^* \nu}{d\nu}((\phi_n \circ \dots \circ \phi_{i+1})^{-1}(\mathbf{w})) \right) \frac{d\phi_n^* \nu}{d\nu}(\mathbf{w}) \nu(d\mathbf{w}). \end{aligned}$$

2.2. The main result. As alluded to above in the [Introduction](#), our main theorem shows how to define an acceptance probability that yields a reversible sampling algorithm when the proposal kernel is given in terms of an involution S defined on an extended parameter space, namely $X \times Y$ for measurable spaces (X, Σ_X) and (Y, Σ_Y) . More specifically, denoting by $\Pi_1 : X \times Y \rightarrow X$ the projection mapping onto the first component, that is,

$$(2.9) \quad \Pi_1(\mathbf{q}, \mathbf{v}) = \mathbf{q} \quad \text{for all } (\mathbf{q}, \mathbf{v}) \in X \times Y,$$

we consider the case of proposal kernels given as $Q(\mathbf{q}, d\tilde{\mathbf{q}}) = (\Pi_1 \circ S(\mathbf{q}, \cdot))^* \mathcal{V}(\mathbf{q}, \cdot)(d\tilde{\mathbf{q}})$, for some Markov kernel $\mathcal{V} : X \times \Sigma_Y \rightarrow [0, 1]$.

To avoid dealing with further technical measure theoretical details, in the statement below and throughout the manuscript we apply the general results from [Section 2.1](#) concerning sigma-finite measures in the particular context of probability measures.

THEOREM 2.1. *Let (X, Σ_X) and (Y, Σ_Y) be measurable spaces. Let μ be a probability measure on X , and let $\mathcal{V} : X \times \Sigma_Y \rightarrow [0, 1]$ be a Markov kernel. Let $\mathcal{M}$ be the probability measure on $X \times Y$ defined as*

$$(2.10) \quad \mathcal{M}(d\mathbf{q}, d\mathbf{v}) = \mathcal{V}(\mathbf{q}, d\mathbf{v}) \mu(d\mathbf{q}).$$

Suppose there exists a measurable mapping $S : X \times Y \rightarrow X \times Y$ satisfying the following properties:

- (P1) S is an involution, that is, $S^2 = I$;
- (P2) $S^*\mathcal{M}$ is absolutely continuous with respect to $\mathcal{M}$.

Let $\hat{\alpha} : X \times Y \rightarrow \mathbb{R}$ be the function defined by

$$(2.11) \quad \hat{\alpha}(\mathbf{q}, \mathbf{v}) := 1 \wedge \frac{dS^*\mathcal{M}}{d\mathcal{M}}(\mathbf{q}, \mathbf{v}), \quad (\mathbf{q}, \mathbf{v}) \in X \times Y,$$

and let $P : X \times \Sigma_X \rightarrow [0, 1]$ be the Markov transition kernel defined as

$$(2.12) \quad P(\mathbf{q}, d\tilde{\mathbf{q}}) = \int_Y \hat{\alpha}(\mathbf{q}, \mathbf{v}) \delta_{\Pi_1 \circ S(\mathbf{q}, \mathbf{v})}(d\tilde{\mathbf{q}}) \mathcal{V}(\mathbf{q}, d\mathbf{v}) + \delta_{\mathbf{q}}(d\tilde{\mathbf{q}}) \int_Y (1 - \hat{\alpha}(\mathbf{q}, \mathbf{v})) \mathcal{V}(\mathbf{q}, d\mathbf{v}),$$

for $\mathbf{q} \in X$. Then, P satisfies detailed balance with respect to μ , that is,

$$(2.13) \quad P(\mathbf{q}, d\tilde{\mathbf{q}}) \mu(d\mathbf{q}) = P(\tilde{\mathbf{q}}, d\mathbf{q}) \mu(d\tilde{\mathbf{q}}),$$

so that, in particular, μ is P invariant.

Before presenting the proof, we try to provide some intuition. Under the assumptions, if S maps $(\mathbf{q}, \mathbf{v})$ to $(\tilde{\mathbf{q}}, \tilde{\mathbf{v}})$, then S will map $(\tilde{\mathbf{q}}, \tilde{\mathbf{v}})$ back to $(\mathbf{q}, \mathbf{v})$. That is, if a proposal $\tilde{\mathbf{q}}$ can be generated from the Markov kernel starting from $\mathbf{q}$, then a proposal $\mathbf{q}$ can similarly be generated starting from $\tilde{\mathbf{q}}$. By appropriately selecting α to balance these two flows—from $(\mathbf{q}, \mathbf{v})$ to $S(\mathbf{q}, \mathbf{v})$ and vice versa—we can achieve detailed balance and therefore invariance. We now turn to the details.

PROOF. It suffices to show that for every bounded measurable function $\varphi : X \times X \rightarrow \mathbb{R}$ we have

$$(2.14) \quad \int_X \int_X \varphi(\mathbf{q}, \tilde{\mathbf{q}}) P(\mathbf{q}, d\tilde{\mathbf{q}}) \mu(d\mathbf{q}) = \int_X \int_X \varphi(\mathbf{q}, \tilde{\mathbf{q}}) P(\tilde{\mathbf{q}}, d\mathbf{q}) \mu(d\tilde{\mathbf{q}}).$$

From the definitions of P and $\mathcal{M}$ in (2.12), (2.10) and Fubini it follows that

$$\begin{aligned} & \int_X \int_X \varphi(\mathbf{q}, \tilde{\mathbf{q}}) P(\mathbf{q}, d\tilde{\mathbf{q}}) \mu(d\mathbf{q}) \\ &= \int_X \int_Y \varphi(\mathbf{q}, \Pi_1 \circ S(\mathbf{q}, \mathbf{v})) \hat{\alpha}(\mathbf{q}, \mathbf{v}) \mathcal{M}(d\mathbf{q}, d\mathbf{v}) + \int_X \int_Y \varphi(\mathbf{q}, \mathbf{q}) (1 - \hat{\alpha}(\mathbf{q}, \mathbf{v})) \mathcal{M}(d\mathbf{q}, d\mathbf{v}) \\ &=: (I) + (II). \end{aligned}$$

Analogously, for the right-hand side of (2.14) we have

$$\begin{aligned} & \int_X \int_X \varphi(\mathbf{q}, \tilde{\mathbf{q}}) P(\tilde{\mathbf{q}}, d\mathbf{q}) \mu(d\tilde{\mathbf{q}}) \\ &= \int_X \int_Y \varphi(\Pi_1 \circ S(\tilde{\mathbf{q}}, \mathbf{v}), \tilde{\mathbf{q}}) \hat{\alpha}(\tilde{\mathbf{q}}, \mathbf{v}) \mathcal{M}(d\tilde{\mathbf{q}}, d\mathbf{v}) + \int_X \int_Y \varphi(\tilde{\mathbf{q}}, \tilde{\mathbf{q}}) (1 - \hat{\alpha}(\tilde{\mathbf{q}}, \mathbf{v})) \mathcal{M}(d\tilde{\mathbf{q}}, d\mathbf{v}) \\ &=: (III) + (IV). \end{aligned}$$

Clearly, $(II) = (IV)$. We now show that $(I) = (III)$.

Invoking assumption (P1), we obtain

$$\begin{aligned} (I) &= \int_X \int_Y \varphi(\mathbf{q}, \Pi_1 \circ S(\mathbf{q}, \mathbf{v})) \hat{\alpha}(\mathbf{q}, \mathbf{v}) \mathcal{M}(d\mathbf{q}, d\mathbf{v}) \\ &= \int_X \int_Y \varphi(\Pi_1 \circ S^2(\mathbf{q}, \mathbf{v}), \Pi_1 \circ S(\mathbf{q}, \mathbf{v})) \hat{\alpha}(S^2(\mathbf{q}, \mathbf{v})) \mathcal{M}(d\mathbf{q}, d\mathbf{v}). \end{aligned}$$

Algorithm 2.1

-
- 1: Choose $\mathbf{q}_0 \in X$
 - 2: **for** $k \geq 0$ **do**
 - 3: Sample $\mathbf{v}_k \sim \mathcal{V}(\mathbf{q}_k, \cdot)$
 - 4: Propose $\bar{\mathbf{q}}_{k+1} = \Pi_1 \circ S(\mathbf{q}_k, \mathbf{v}_k)$
 - 5: Set $\mathbf{q}_{k+1} = \bar{\mathbf{q}}_{k+1}$ with probability $\hat{\alpha}(\mathbf{q}_k, \mathbf{v}_k)$ given by (2.11), otherwise $\mathbf{q}_{k+1} = \mathbf{q}_k$
-

Thus, by Fubini, the change of variables formula, (2.2), and then invoking assumption (P2), we write

$$\begin{aligned}
 (I) &= \int_{X \times Y} \varphi(\Pi_1 \circ S(\mathbf{q}, \mathbf{v}), \mathbf{q}) \hat{\alpha}(S(\mathbf{q}, \mathbf{v})) S^* \mathcal{M}(d\mathbf{q}, d\mathbf{v}) \\
 (2.15) \quad &= \int_{X \times Y} \varphi(\Pi_1 \circ S(\mathbf{q}, \mathbf{v}), \mathbf{q}) \hat{\alpha}(S(\mathbf{q}, \mathbf{v})) \frac{dS^* \mathcal{M}}{d\mathcal{M}}(\mathbf{q}, \mathbf{v}) \mathcal{M}(d\mathbf{q}, d\mathbf{v}).
 \end{aligned}$$

Hence, in order to conclude that $(I) = (III)$, it suffices to show that

$$(2.16) \quad \hat{\alpha}(S(\mathbf{q}, \mathbf{v})) \frac{dS^* \mathcal{M}}{d\mathcal{M}}(\mathbf{q}, \mathbf{v}) = \hat{\alpha}(\mathbf{q}, \mathbf{v}) \quad \text{for } \mathcal{M}\text{-a.e. } (\mathbf{q}, \mathbf{v}) \in X \times Y.$$

Since $dS^* \mathcal{M}/d\mathcal{M}(\mathbf{q}, \mathbf{v}) \geq 0$ $\mathcal{M}$ -a.e., from the definition of $\hat{\alpha}$ in (2.11) it follows that

$$(2.17) \quad \hat{\alpha}(S(\mathbf{q}, \mathbf{v})) \frac{dS^* \mathcal{M}}{d\mathcal{M}}(\mathbf{q}, \mathbf{v}) = \frac{dS^* \mathcal{M}}{d\mathcal{M}}(\mathbf{q}, \mathbf{v}) \wedge \left(\frac{dS^* \mathcal{M}}{d\mathcal{M}}(\mathbf{q}, \mathbf{v}) \frac{dS^* \mathcal{M}}{d\mathcal{M}}(S(\mathbf{q}, \mathbf{v})) \right)$$

for $\mathcal{M}$ -a.e. $(\mathbf{q}, \mathbf{v}) \in X \times Y$. Now from (2.8) and assumption (P1) we have that

$$(2.18) \quad \frac{dS^* \mathcal{M}}{d\mathcal{M}}(\mathbf{q}, \mathbf{v}) \frac{dS^* \mathcal{M}}{d\mathcal{M}}(S(\mathbf{q}, \mathbf{v})) = \frac{d(S^2)^* \mathcal{M}}{d\mathcal{M}}(\mathbf{q}, \mathbf{v}) = 1 \quad \text{for } \mathcal{M}\text{-a.e. } (\mathbf{q}, \mathbf{v}) \in X \times Y.$$

Plugging (2.18) into (2.17), we obtain (2.16). This concludes the proof. $\square$

The MCMC sampling scheme resulting from Theorem 2.1 is summarized in Algorithm 2.1.

We conclude this section with a number of remarks clarifying the scope of Theorem 2.1.

REMARK 2.2. If in Theorem 2.1 we assumed in addition that X and Y are Radon spaces (i.e., separable metric spaces on which every probability measure is tight), then we could take $\mathcal{M}$ to be any probability measure on $X \times Y$ with first marginal μ , that is, $\Pi_1^* \mathcal{M} = \mu$. Indeed, in this case it follows from the disintegration theorem that there exists a Markov kernel $\mathcal{V} : X \times \Sigma_Y \rightarrow [0, 1]$ such that $\mathcal{M}$ is written as in (2.10); see, for example, [2], Theorem 5.3.1.

REMARK 2.3. Assumptions (P1) and (P2) in Theorem 2.1 imply that the measures $S^* \mathcal{M}$ and $\mathcal{M}$ are in fact mutually absolutely continuous. Indeed, if $E \subset X \times Y$ is such that $S^* \mathcal{M}(E) = 0$, then by definition of pushforwards we have $\mathcal{M}(S^{-1}(E)) = 0$. Thus assumption (P2) implies that $S^* \mathcal{M}(S^{-1}(E)) = 0$, and with (P1) we obtain

$$0 = S^* \mathcal{M}(S^{-1}(E)) = \mathcal{M}(S^{-1}(S^{-1}(E))) = \mathcal{M}(S^2(E)) = \mathcal{M}(E).$$

Hence, $\mathcal{M}(E) = 0$, so that $\mathcal{M} \ll S^* \mathcal{M}$.

REMARK 2.4 (Generalizations of Theorem 2.1). We notice that the statement of Theorem 2.1 in fact holds under a more general form. Namely, similarly as in [72], we could disregard assumption (P2) and define the acceptance probability $\hat{\alpha}$ in (2.11) as $\hat{\alpha}(\mathbf{q}, \mathbf{v}) = 1 \wedge \frac{dS^* \mathcal{M}}{d\mathcal{M}}(\mathbf{q}, \mathbf{v})$ for all $(\mathbf{q}, \mathbf{v})$ in a measurable subset $\mathcal{O} \subset X \times Y$ where $S^* \mathcal{M}_{\mathcal{O}} \ll \mathcal{M}_{\mathcal{O}}$, and

0 otherwise. Here $\mathcal{M}_{\mathcal{O}}$ denotes the restriction of the measure $\mathcal{M}$ to the set $\mathcal{O}$. However, it is worth pointing out that in practice this set $\mathcal{O}$ could be empty, in which case $\hat{\alpha} \equiv 0$ and the corresponding MCMC algorithm would always reject the proposals, a clearly undesirable behavior.

Moreover, instead of defining $\hat{\alpha}$ explicitly as in (2.11), we could have taken $\hat{\alpha}$ to be any measurable function for which (2.16) holds (particularly on the set $\mathcal{O}$, under the setting of the previous paragraph). Indeed, (2.11) gives just one such example. It is worth pointing however that, as noticed in [72], Section 3, the standard choice given in (2.11) yields the maximum acceptance probability among all possible choices of $\hat{\alpha}$ satisfying (2.16), since

$$\hat{\alpha}(S(\mathbf{q}, \mathbf{v})) \frac{dS^* \mathcal{M}}{d\mathcal{M}}(\mathbf{q}, \mathbf{v}) = \hat{\alpha}(\mathbf{q}, \mathbf{v}) \leq 1 \wedge \frac{dS^* \mathcal{M}}{d\mathcal{M}}(\mathbf{q}, \mathbf{v}), \quad (\mathbf{q}, \mathbf{v}) \in X \times Y.$$

Finally, assumption (P1) can be replaced by the following more general condition:

(P1') $(S^2)^* \mathcal{M} = \mathcal{M}$, $\Pi_1 \circ S^2 = \Pi_1$, and S is invertible with measurable inverse $S^{-1} : X \times Y \rightarrow X \times Y$.

Indeed, we arrive at the same identity (2.15) by noticing that

$$\begin{aligned} (I) &= \int_X \int_Y \varphi(\mathbf{q}, \Pi_1 \circ S(\mathbf{q}, \mathbf{v})) \hat{\alpha}(\mathbf{q}, \mathbf{v}) \mathcal{M}(d\mathbf{q}, d\mathbf{v}) \\ &= \int_X \int_Y \varphi(\mathbf{q}, \Pi_1 \circ S(\mathbf{q}, \mathbf{v})) \hat{\alpha}(\mathbf{q}, \mathbf{v}) (S^2)^* \mathcal{M}(d\mathbf{q}, d\mathbf{v}) \\ &= \int_X \int_Y \varphi(\Pi_1 \circ S(\mathbf{q}, \mathbf{v}), \Pi_1 \circ S^2(\mathbf{q}, \mathbf{v})) \hat{\alpha}(S(\mathbf{q}, \mathbf{v})) S^* \mathcal{M}(d\mathbf{q}, d\mathbf{v}) \\ &= \int_X \int_Y \varphi(\Pi_1 \circ S(\mathbf{q}, \mathbf{v}), \mathbf{q}) \hat{\alpha}(S(\mathbf{q}, \mathbf{v})) S^* \mathcal{M}(d\mathbf{q}, d\mathbf{v}). \end{aligned}$$

Moreover the conclusion in (2.18) still holds since with the assumption $(S^2)^* \mathcal{M} = \mathcal{M}$ it follows that $S^* \mathcal{M} = (S^{-1})^* (S^2)^* \mathcal{M} = (S^{-1})^* \mathcal{M}$. Therefore, invoking (2.8), we obtain

$$\begin{aligned} \frac{dS^* \mathcal{M}}{d\mathcal{M}}(\mathbf{q}, \mathbf{v}) \frac{dS^* \mathcal{M}}{d\mathcal{M}}(S(\mathbf{q}, \mathbf{v})) &= \frac{d(S^{-1})^* \mathcal{M}}{d\mathcal{M}}(\mathbf{q}, \mathbf{v}) \frac{d(S^{-1})^* \mathcal{M}}{d\mathcal{M}}(S(\mathbf{q}, \mathbf{v})) \\ &= \frac{d((S^{-1})^2)^* \mathcal{M}}{d\mathcal{M}}(\mathbf{q}, \mathbf{v}) = 1 \end{aligned}$$

for $\mathcal{M}$ -a.e. $(\mathbf{q}, \mathbf{v}) \in X \times Y$, where the last identity holds since

$$((S^{-1})^2)^* \mathcal{M} = ((S^{-1})^2)^* (S^2)^* \mathcal{M} = \mathcal{M}.$$

Note that, all of the applications of Theorem 2.1 developed in Section 4, Section 5 below fall within the particular involutive structure of Theorem 2.1, (P1). It would thus be interesting to identify an example of MCMC algorithm that would require the more general framework from (P1').

REMARK 2.5. For many MCMC algorithms, the associated proposal kernel is defined from a mapping S as in Theorem 2.1 that is in turn given in terms of a numerical integrator of a suitably chosen dynamics. Denoting such an integrator by a measurable mapping $\hat{S} : X \times Y \rightarrow X \times Y$, one commonly seeks a corresponding measurable mapping $R : X \times Y \rightarrow X \times Y$ such that

$$(2.19) \quad R \circ \hat{S} = \hat{S}^{-1} \circ R.$$

In the context of vector spaces X, Y , the condition (2.19) is referred to in the theory of geometric integrators as the *reversibility of $\hat{S}$ with respect to R* , and R is often taken as a linear

invertible mapping (see, e.g., [37], Definition V.1.2). Frequently, one can find R commensurate with (2.19) which is also an involution, that is, $R^2 = I$. For example, in the case of the HMC algorithm, where $\hat{S}$ represents a numerical integrator of a suitable Hamiltonian dynamics, R is commonly taken as the “momentum-flip” involution $R(\mathbf{q}, \mathbf{v}) = (\mathbf{q}, -\mathbf{v})$, see Section 4 and Section 5.1.4, Section 5.2.3 below.

We thus define $S := R \circ \hat{S}$, and conclude from (2.19) and $R^2 = I$ that $S^2 = I$, that is, S is an involution, so that assumption (P1) of Theorem 2.1 is satisfied. In fact, if instead of $R^2 = I$ we assume more generally that $(R^2)^* \mathcal{M} = \mathcal{M}$, $\Pi_1 \circ R^2 = \Pi_1$, and R is invertible with measurable inverse, then clearly (2.19) implies that $S^2 = (R \circ \hat{S})^2 = R^2$, so that S satisfies the more general condition (P1') in Remark 2.4. Here again it would be interesting to identify an example of MCMC algorithm for which this more general setting for R is required.

REMARK 2.6 (Connection with the Metropolis–Hastings–Green algorithm). As already mentioned in the Introduction, Algorithm 2.1 may be seen as a generalization to abstract state spaces of the so called Metropolis–Hastings–Green algorithm [31, 32, 36]. We can make this connection explicit as follows. Let $X = \mathbb{R}^N$ and fix any continuously distributed target probability measure $\mu(d\mathbf{q}) = p(\mathbf{q}) d\mathbf{q}$. We consider any Markovian proposal kernel $\mathcal{V} : \mathbb{R}^N \times \mathcal{B}(\mathbb{R}^M) \rightarrow [0, 1]$ of the form $\mathcal{V}(\mathbf{q}, d\mathbf{v}) = q(\mathbf{q}, \mathbf{v}) d\mathbf{v}$ where $q : \mathbb{R}^{N \times M} \rightarrow [0, \infty)$ is such that $\int_{\mathbb{R}^M} q(\mathbf{q}, \mathbf{v}) d\mathbf{v} = 1$ for any $\mathbf{q} \in \mathbb{R}^N$. Finally select any $S : \mathbb{R}^{N+M} \rightarrow \mathbb{R}^{N+M}$ which is C^1 and is an involution, namely we assume that $S \circ S(\mathbf{z}) = \mathbf{z}$ for every $\mathbf{z} \in \mathbb{R}^{N+M}$. Then, referring back to (2.11) and recalling (2.3), (2.5), we obtain a Markov kernel P of the form (2.12) with the acceptance probability given by

$$(2.20) \quad \hat{\alpha}(\mathbf{q}, \mathbf{v}) = 1 \wedge \frac{\rho(S(\mathbf{q}, \mathbf{v})) |\det \nabla S(\mathbf{q}, \mathbf{v})|}{\rho(\mathbf{q}, \mathbf{v})} \quad \text{where } \rho(\mathbf{q}, \mathbf{v}) := p(\mathbf{q}) q(\mathbf{q}, \mathbf{v}).$$

Thus with these specifications for μ , $\mathcal{V}$ and S we see that the algorithm derived in [36] (see also [31], Section 1.2) falls out as a special case of Algorithm 2.1.

[36] also considered the possibility of combining proposals from a collection of different sampling mechanisms where the proposal mechanism employed is selected at random according to a state dependent probability. This “multi-kernel” approach developed in [36] can be generalized to an abstract context similar to Theorem 2.1 in a fashion which recovers the results in [36] as a special case. To see this we proceed as follows. Fix a measurable space (X, Σ_X) along with a collection of auxiliary measurable spaces (Y_j, Σ_{Y_j}) defined for $j = 1, \dots, L$. Our target measure μ is any probability measure on X . On each extended phase space $X \times Y_j$ for $j = 1, \dots, L$ we assume that we have defined a Markov proposal kernel $\mathcal{V}_j(\mathbf{q}, d\mathbf{v}_j)$ and a measurable involution $S_j : X \times Y_j \rightarrow X \times Y_j$ such that $S_j^* \mathcal{M}_j$ is absolutely continuous with respect to $\mathcal{M}_j$ where $\mathcal{M}_j(d\mathbf{q}, d\mathbf{v}) = \mathcal{V}_j(\mathbf{q}, d\mathbf{v}) \mu(d\mathbf{q})$. Finally we suppose we have defined, for each $j = 1, \dots, L$, a measurable function $\kappa_j : X \rightarrow [0, 1]$ in such a way that $\sum_{j=1}^L \kappa_j(\mathbf{q}) = 1$ for each $\mathbf{q} \in X$. With these ingredients in hand we define $P_j : X \times \Sigma_Y \rightarrow [0, 1]$, for $j = 1, \dots, L$ to be the Markov transition kernels given as

$$P_j(\mathbf{q}, d\tilde{\mathbf{q}}) = \int_Y \hat{\alpha}_j(\mathbf{q}, \mathbf{v}_j) \delta_{\Pi_1 \circ S_j(\mathbf{q}, \mathbf{v}_j)}(d\tilde{\mathbf{q}}) \mathcal{V}_j(\mathbf{q}, d\mathbf{v}_j) + \delta_{\mathbf{q}}(d\tilde{\mathbf{q}}) \int_Y (1 - \hat{\alpha}_j(\mathbf{q}, \mathbf{v}_j)) \mathcal{V}_j(\mathbf{q}, d\mathbf{v}_j),$$

with

$$\hat{\alpha}_j(\mathbf{q}, \mathbf{v}) := 1 \wedge \left(\frac{\kappa_j(\Pi_1 \circ S_j(\mathbf{q}, \mathbf{v}_j)) dS_j^* \mathcal{M}_j}{\kappa_j(\mathbf{q}) d\mathcal{M}_j}(\mathbf{q}, \mathbf{v}_j) \right), \quad (\mathbf{q}, \mathbf{v}_j) \in X \times Y_j.$$

We now define a compound kernel $P : X \times \Sigma_X \rightarrow [0, 1]$ as $P(\mathbf{q}, d\tilde{\mathbf{q}}) = \sum_{j=1}^L \kappa_j(\mathbf{q}) P_j(\mathbf{q}, \tilde{\mathbf{q}})$. Following the proof of Theorem 2.1 mutatis mutandis we obtain that P is reversible with

respect to μ . Moreover, in the special case when $\mu(d\mathbf{q}) = p(\mathbf{q}) d\mathbf{q}$, $Y_j = \mathbb{R}^{M_j}$, $\mathcal{V}_j(\mathbf{q}, d\mathbf{v}_j) = q_j(\mathbf{q}, \mathbf{v}_j) d\mathbf{v}_j$, we recover the formulation in [36] as described in [31], Section 1.3.

Note that the connection between the Metropolis–Hasting–Green algorithm and Hamiltonian Monte Carlo sampling, where the Hamiltonian dynamic or more precisely a well chosen numerical discretization of the dynamics provides an associated involution, seems to be more recent. See [49, 62]. Indeed, below in Section 4.1 we show how the usage of Hamiltonian dynamics can allow one to derive a broad class of involutions S and thus to recover numerous formulations of Hamiltonian Monte Carlo (essentially) as special cases of the general observation leading to (2.20).

We note that Theorem 2.1 has connections to the results presented in [3], Section 3.1, which we discuss further in Remark 3.8.

3. Connection to the Tierney framework. In this section, we connect our main result, Theorem 2.1, to the formulation of reversible Metropolis–Hastings algorithms on general state spaces $\mathcal{X}$ laid out previously in [72]. We show that the two frameworks connect or overlap in several different and interesting ways. On the one hand, in Section 3.2 with Theorem 3.4, we prove that Theorem 2.1 subsumes the main result in [72]. In fact we demonstrate that this connection can often be made in multiple, nonequivalent ways; namely, in cases of interest, a variety of different choices for S and $\mathcal{V}$ in (2.12) can ultimately yield the same desired Markov kernel $\bar{P}$ specified in Tierney’s formulation. In particular note that Theorem 3.4 is employed in Section 5 below to provide further insights for the derivation of various classical MCMC algorithms. On the other hand, in Section 3.3, we establish that the main results in [72] can be used to develop a proof of Theorem 2.1 distinct from the one presented above in Section 2.

3.1. Overview of Tierney’s formulation. Before turning to the main results in this section let us begin by recalling some of the framework and notation from [72]. As above in Section 2 we let $(\mathcal{X}, \Sigma_{\mathcal{X}})$ be any measurable space. Starting with a target probability measure μ on $\mathcal{X}$ and a proposal Markov kernel $Q : \mathcal{X} \times \Sigma_{\mathcal{X}} \rightarrow [0, 1]$, [72] considers the Metropolis–Hastings-type Markov kernel defined as

$$(3.1) \quad \bar{P}(\mathbf{q}, d\tilde{\mathbf{q}}) = \alpha(\mathbf{q}, \tilde{\mathbf{q}}) Q(\mathbf{q}, d\tilde{\mathbf{q}}) + \delta_{\mathbf{q}}(d\tilde{\mathbf{q}}) \int_{\mathcal{X}} (1 - \alpha(\mathbf{q}, \mathbf{r})) Q(\mathbf{q}, d\mathbf{r}), \quad \mathbf{q} \in \mathcal{X},$$

with the acceptance ratio α given by

$$(3.2) \quad \alpha(\mathbf{q}, \tilde{\mathbf{q}}) := 1 \wedge \frac{d\eta^{\perp}}{d\eta}(\mathbf{q}, \tilde{\mathbf{q}}), \quad \mathbf{q}, \tilde{\mathbf{q}} \in \mathcal{X}.$$

Here the measures $\eta, \eta^{\perp}$ are given by

$$(3.3) \quad \eta(d\mathbf{q}, d\tilde{\mathbf{q}}) = \mu(d\mathbf{q}) Q(\mathbf{q}, d\tilde{\mathbf{q}}), \quad \eta^{\perp}(d\mathbf{q}, d\tilde{\mathbf{q}}) = \eta(d\tilde{\mathbf{q}}, d\mathbf{q}) = \mu(d\tilde{\mathbf{q}}) Q(\tilde{\mathbf{q}}, d\mathbf{q})$$

and are assumed to be mutually absolutely continuous, so that in particular α in (3.2) is well defined.³ Tierney then shows, [72], Theorem 2, that (3.1), (3.2) yields a Markov kernel satisfying detailed balance with respect to μ , namely, cf. (2.13), $\bar{P}(\mathbf{q}, d\tilde{\mathbf{q}})\mu(d\mathbf{q}) = \bar{P}(\tilde{\mathbf{q}}, d\mathbf{q})\mu(d\tilde{\mathbf{q}})$.

As an important preliminary observation we notice that the formulation in [72], that is, (3.1), (3.2), can be recovered from Theorem 2.1 in a straightforward fashion as follows:

³In fact [72] allows for $\eta, \eta^{\perp}$ to be mutually absolutely continuous only on a subset $\mathcal{O} \subset \mathcal{X} \times \mathcal{X}$, in which case α is defined to be zero on $\mathcal{O}^c$. A similar generalization holds for Theorem 2.1; see Remark 2.4. But for simplicity of presentation we restrict ourselves to the case of mutual absolute continuity here.

REMARK 3.1. Given the inputs Q, μ we take $X := \mathcal{X}, Y := \mathcal{X}$ and set

$$(3.4) \quad \mathcal{V}(\mathbf{q}, d\mathbf{v}) := Q(\mathbf{q}, d\mathbf{v}), \quad S(\mathbf{q}, \mathbf{v}) := (\mathbf{v}, \mathbf{q}).$$

Clearly S is an involution and it is easy to see that, with these choices (cf. (2.10)), $\mathcal{M} = \eta$ and $S^*\mathcal{M} = \eta^\perp$ so that $dS^*\mathcal{M}/d\mathcal{M} = d\eta^\perp/d\eta$. Therefore (2.12) reduces to (3.1) under this $\mathcal{V}$ and S .

3.2. *Reduction to the Tierney formulation.* The observation in Remark 3.1, recovering (3.1) from (2.12), can be extended more broadly. Here a starting point is to notice that we are in fact writing the proposal kernel in (3.1) as $Q(\mathbf{q}, d\tilde{\mathbf{q}}) = F(\mathbf{q}, \cdot)^*\mathcal{V}(\mathbf{q}, \cdot)(d\tilde{\mathbf{q}})$ but in a trivial fashion where $Q = \mathcal{V}$ and $F(\mathbf{q}, \mathbf{v}) = \mathbf{v}$. The insight is that a desired Q , along with the involution S required for (2.12) which recovers the kernel (3.1), can often be identified from other, nontrivial, choices for F and $\mathcal{V}$.

We formulate this generalization as follows. Let (X, Σ_X) and (Y, Σ_Y) be measurable spaces. We consider proposals $\tilde{\mathbf{q}} \sim Q(\mathbf{q}, d\tilde{\mathbf{q}})$ made from some starting point $\mathbf{q} \in X$ which are generated in the following fashion:

1. Draw $\mathbf{v} \sim \mathcal{V}(\mathbf{q}, \cdot)$, for some Markov kernel $\mathcal{V} : X \times \Sigma_Y \rightarrow [0, 1]$.
2. Compute $\tilde{\mathbf{q}} = F(\mathbf{q}, \mathbf{v})$ from a (measurable) deterministic map $F : X \times Y \rightarrow X$.

In other words we are considering the proposal kernel Q in the formulation

$$(3.5) \quad Q(\mathbf{q}, d\tilde{\mathbf{q}}) = \int_Y \delta_{F(\mathbf{q}, \mathbf{v})}(d\tilde{\mathbf{q}}) \mathcal{V}(\mathbf{q}, d\mathbf{v}) = F(\mathbf{q}, \cdot)^*\mathcal{V}(\mathbf{q}, \cdot)(d\tilde{\mathbf{q}}).$$

We assume throughout what follows that, for a given sample $\mathbf{q} \in X$ and proposal $\tilde{\mathbf{q}} \in X$, F can be inverted to obtain the $\mathbf{v} \in Y$ such that $F(\mathbf{q}, \mathbf{v}) = \tilde{\mathbf{q}}$, namely,

$$(3.6) \quad \text{for each } \mathbf{q} \in X, \text{ the map } F(\mathbf{q}, \cdot) : Y \rightarrow X \text{ is one-to-one.}$$

REMARK 3.2. One may formulate Q as in (3.5) maintaining (3.6) in a nontrivial fashion for MCMC methods with additive noise such as the RWMC, pCN and MALA schemes, all of which fall into the framework (3.1) outlined in [72]. In other words, for each of these examples, we may determine F and $\mathcal{V}$ in a form distinct from (3.4); see, for example, (5.1), (5.5), (5.10) and (5.12) in Section 5 below. By contrast, (3.6) does not encompass HMC, for which there may be more than one $\mathbf{v}$ leading to the same proposal.

We now show in Theorem 3.4 that in the formulation (3.5) under (3.6) we may obtain a suitable involution S which yields an equivalence between the Markov transition kernels in (2.12), Theorem 2.1, and (3.1), from [72]. As a preliminary step we show how to construct this involution S in $(\mathbf{q}, \mathbf{v})$ -space corresponding to any F satisfying (3.6).

PROPOSITION 3.3. *Let X, Y be any sets and let $F : X \times Y \rightarrow X$ be a mapping satisfying (3.6), that is, such that for each fixed $\mathbf{q} \in X$, $F(\mathbf{q}, \cdot) : Y \rightarrow X$ is one-to-one. Then, there exists a unique mapping $S : X \times Y \rightarrow X \times Y$ such that $\Pi_1 \circ S = F$ and $S^2 = I$, given by*

$$(3.7) \quad S(\mathbf{q}, \mathbf{v}) = (F(\mathbf{q}, \mathbf{v}), F(F(\mathbf{q}, \mathbf{v}), \cdot)^{-1}(\mathbf{q})) \quad \text{for all } (\mathbf{q}, \mathbf{v}) \in X \times Y.$$

PROOF. First, let us verify that S defined in (3.7) satisfies the required properties. Clearly, $\Pi_1 \circ S = F$, so it remains to show that $S^2 = I$. Define the maps $B_1 : X \times Y \rightarrow X \times X$ and $B_2 : X \times X \rightarrow X \times X$ by

$$(3.8) \quad B_1(\mathbf{q}, \mathbf{v}) = (\mathbf{q}, F(\mathbf{q}, \mathbf{v})), \quad B_2(\mathbf{q}, \tilde{\mathbf{q}}) = (\tilde{\mathbf{q}}, \mathbf{q})$$

for all $\mathbf{q} \in X$, $\tilde{\mathbf{q}} \in X$ and $\mathbf{v} \in Y$. Note that $B_2^2 = I$ trivially and by assumption (3.6) we have

$$B_1^{-1}(\mathbf{q}, \tilde{\mathbf{q}}) = (\mathbf{q}, F(\mathbf{q}, \cdot)^{-1}(\tilde{\mathbf{q}})) \quad \text{for all } (\mathbf{q}, \tilde{\mathbf{q}}) \in X \times X.$$

Then it is not difficult to check that S as in (3.7) can be written as

$$(3.9) \quad S(\mathbf{q}, \mathbf{v}) = B_1^{-1} \circ B_2 \circ B_1(\mathbf{q}, \mathbf{v}) \quad \text{for all } (\mathbf{q}, \mathbf{v}) \in X \times Y.$$

Then clearly $S^2 = I$ since $B_2^2 = I$.

Now suppose that $\bar{S}: X \times Y \rightarrow X \times Y$ is any mapping satisfying the required properties, that is, $\Pi_1 \circ \bar{S} = F$ and $\bar{S}^2 = I$. Let $G: X \times Y \rightarrow Y$ such that

$$\bar{S}(\mathbf{q}, \mathbf{v}) = (F(\mathbf{q}, \mathbf{v}), G(\mathbf{q}, \mathbf{v})) \quad \text{for all } (\mathbf{q}, \mathbf{v}) \in X \times Y.$$

Thus,

$$\bar{S}^2(\mathbf{q}, \mathbf{v}) = \bar{S}(\bar{S}(\mathbf{q}, \mathbf{v})) = (F(F(\mathbf{q}, \mathbf{v}), G(\mathbf{q}, \mathbf{v})), G(F(\mathbf{q}, \mathbf{v}), G(\mathbf{q}, \mathbf{v}))).$$

Since $\bar{S}^2 = I$, it follows in particular that

$$F(F(\mathbf{q}, \mathbf{v}), G(\mathbf{q}, \mathbf{v})) = \mathbf{q},$$

which implies

$$G(\mathbf{q}, \mathbf{v}) = F(F(\mathbf{q}, \mathbf{v}), \cdot)^{-1}(\mathbf{q}).$$

Therefore, $\bar{S} = S$. This concludes the proof. $\square$

With Proposition 3.3 in hand we now turn to our equivalence result:

THEOREM 3.4. *Let X and Y be measurable spaces. Take μ to be a probability measure on X and consider an associated proposal kernel Q satisfying (3.5), (3.6), that is, $Q(\mathbf{q}, d\tilde{\mathbf{q}}) = (F(\mathbf{q}, \cdot))^* \mathcal{V}(\mathbf{q}, \cdot)(d\tilde{\mathbf{q}})$, where $\mathcal{V}: X \times \Sigma_Y \rightarrow [0, 1]$ is a Markov kernel and $F: X \times Y \rightarrow X$ is a measurable mapping such that $F(\mathbf{q}, \cdot): Y \rightarrow X$ is one-to-one for each fixed $\mathbf{q} \in X$ and, additionally, its inverse $F(\mathbf{q}, \cdot)^{-1}$ is measurable.⁴ We define the probability measures η , $\eta^\perp$ on $X \times X$ as in (3.3) and let $S: X \times Y \rightarrow X \times Y$ be the unique mapping satisfying $\Pi_1 \circ S = F$ and $S^2 = I$, given by (3.7) in Proposition 3.3. Then, denoting by $\mathcal{M}$ the probability measure on $X \times Y$ defined as in (2.10), we have:*

(i) *the measure $\eta^\perp$ is absolutely continuous with respect to η if and only if $S^* \mathcal{M}$ is absolutely continuous with respect to $\mathcal{M}$.*

(ii) *Moreover, under either of the equivalent circumstances in (i), we have*

$$(3.10) \quad \frac{d\eta^\perp}{d\eta}(\mathbf{q}, F(\mathbf{q}, \mathbf{v})) = \frac{dS^* \mathcal{M}}{d\mathcal{M}}(\mathbf{q}, \mathbf{v})$$

for $\mathcal{M}$ -a.e. $(\mathbf{q}, \mathbf{v}) \in X \times Y$ or, equivalently,

$$(3.11) \quad \frac{d\eta^\perp}{d\eta}(\mathbf{q}, \tilde{\mathbf{q}}) = \frac{dS^* \mathcal{M}}{d\mathcal{M}}(\mathbf{q}, F(\mathbf{q}, \cdot)^{-1}(\tilde{\mathbf{q}}))$$

for η -a.e. $(\mathbf{q}, \tilde{\mathbf{q}}) \in X \times X$.

⁴As remarked in Section 2.1, here we notice again that if X and Y are Polish spaces, that is, separable and completely metrizable, then the fact that, for each fixed $\mathbf{q} \in X$, $F(\mathbf{q}, \cdot)$ is a measurable and one-to-one mapping between Polish spaces automatically implies that its inverse is measurable; see, for example, [1], Theorem 12.29.

(iii) Furthermore, under either of these equivalent absolute continuity conditions, the Markov kernels given by (2.12) and (3.1) coincide, where α and $\hat{\alpha}$ are as defined in (3.2) and (2.11), respectively. These acceptance functions α and $\hat{\alpha}$ maintain the relationships

$$(3.12) \quad \alpha(\mathbf{q}, \tilde{\mathbf{q}}) = \hat{\alpha}(\mathbf{q}, F(\mathbf{q}, \cdot)^{-1}(\tilde{\mathbf{q}})), \quad \hat{\alpha}(\mathbf{q}, \mathbf{v}) = \alpha(\mathbf{q}, F(\mathbf{q}, \mathbf{v}))$$

for $\mathbf{q} \in X$, $\tilde{\mathbf{q}} \in X$ and $\mathbf{v} \in Y$.

REMARK 3.5. Reiterating the discussion at the beginning of this subsection, notice that, modulo the requirement that $X = Y$, we see that Theorem 3.4 reduces to Remark 3.1 by taking $F(\mathbf{q}, \mathbf{v}) = \mathbf{v}$ and $\mathcal{V} = \mathcal{Q}$.

PROOF. Let $B_1 : X \times Y \rightarrow X \times X$ and $B_2 : X \times X \rightarrow X \times X$ be defined as in (3.8). Then it is not hard to see from (3.3) that

$$(3.13) \quad \eta^\perp(d\mathbf{q}, d\tilde{\mathbf{q}}) = B_2^* \eta(d\mathbf{q}, d\tilde{\mathbf{q}})$$

and similarly from (3.5) that

$$(3.14) \quad \eta(d\mathbf{q}, d\tilde{\mathbf{q}}) = B_1^* \mathcal{M}(d\mathbf{q}, d\tilde{\mathbf{q}}).$$

Then combining (3.9), (3.13) and (3.14) yields

$$(3.15) \quad \eta^\perp(d\mathbf{q}, d\tilde{\mathbf{q}}) = (B_2 \circ B_1)^* \mathcal{M}(d\mathbf{q}, d\tilde{\mathbf{q}}) = (B_1 \circ S)^* \mathcal{M}(d\mathbf{q}, d\tilde{\mathbf{q}}).$$

Let us first show that $\eta^\perp \ll \eta$ implies $S^* \mathcal{M} \ll \mathcal{M}$. Indeed, let A be a measurable subset of $X \times Y$ such that $\mathcal{M}(A) = 0$. By the assumptions on F it follows that B_1 is a one-to-one measurable mapping with measurable inverse. Then $B_1(A)$ is also a Borel set in $X \times X$. From (3.14), we thus obtain

$$0 = \mathcal{M}(A) = \mathcal{M}(B_1^{-1} \circ B_1(A)) = B_1^* \mathcal{M}(B_1(A)) = \eta(B_1(A)).$$

Since $\eta^\perp \ll \eta$, this implies that $\eta^\perp(B_1(A)) = 0$. With (3.15), we deduce

$$0 = \eta^\perp(B_1(A)) = (B_1 \circ S)^* \mathcal{M}(B_1(A)) = S^* \mathcal{M}(B_1^{-1} \circ B_1(A)) = S^* \mathcal{M}(A).$$

This shows that $S^* \mathcal{M} \ll \mathcal{M}$. The reciprocal claim, that $S^* \mathcal{M} \ll \mathcal{M}$ implies $\eta^\perp \ll \eta$, follows similarly by invoking (3.14) and (3.15). This concludes the proof of item (i).

Now assuming any of the equivalent circumstances in (i), notice that, for any bounded and measurable function $\varphi : X \times X \rightarrow \mathbb{R}$, application of (3.15) and (3.14) yields

$$\begin{aligned} \int_{X \times Y} \varphi(\mathbf{q}, \mathbf{v}) S^* \mathcal{M}(d\mathbf{q}, d\mathbf{v}) &= \int_{X \times X} \varphi(B_1^{-1}(\mathbf{q}, \tilde{\mathbf{q}})) (B_1 \circ S)^* \mathcal{M}(d\mathbf{q}, d\tilde{\mathbf{q}}) \\ &= \int_{X \times X} \varphi(B_1^{-1}(\mathbf{q}, \tilde{\mathbf{q}})) \eta^\perp(d\mathbf{q}, d\tilde{\mathbf{q}}) \\ &= \int_{X \times X} \varphi(B_1^{-1}(\mathbf{q}, \tilde{\mathbf{q}})) \frac{d\eta^\perp}{d\eta}(\mathbf{q}, \tilde{\mathbf{q}}) \eta(d\mathbf{q}, d\tilde{\mathbf{q}}) \\ &= \int_{X \times X} \varphi(B_1^{-1}(\mathbf{q}, \tilde{\mathbf{q}})) \frac{d\eta^\perp}{d\eta}(\mathbf{q}, \tilde{\mathbf{q}}) B_1^* \mathcal{M}(d\mathbf{q}, d\tilde{\mathbf{q}}) \\ &= \int_{X \times Y} \varphi(\mathbf{q}, \mathbf{v}) \frac{d\eta^\perp}{d\eta}(B_1(\mathbf{q}, \mathbf{v})) \mathcal{M}(d\mathbf{q}, d\mathbf{v}), \end{aligned}$$

so that

$$\frac{dS^* \mathcal{M}}{d\mathcal{M}}(\mathbf{q}, \mathbf{v}) = \frac{d\eta^\perp}{d\eta}(B_1(\mathbf{q}, \mathbf{v})) = \frac{d\eta^\perp}{d\eta}(\mathbf{q}, F(\mathbf{q}, \mathbf{v}))$$

for $\mathcal{M}$ -a.e. $(\mathbf{q}, \mathbf{v}) \in X \times Y$. Clearly, this implies (3.11). Indeed, if $E \subset X \times X$ is the set of points $(\mathbf{q}, \tilde{\mathbf{q}}) \in X \times X$ where (3.11) does not hold, then it is not difficult to see that (3.10) does not hold for every $(\mathbf{q}, \mathbf{v}) \in B_1^{-1}(E)$. But since (3.10) holds $\mathcal{M}$ -a.e., then $B_1^* \mathcal{M}(E) = \mathcal{M}(B_1^{-1}(E)) = 0$. Hence, from (3.14), $\eta(E) = 0$, so that (3.11) holds for η -a.e. $(\mathbf{q}, \tilde{\mathbf{q}}) \in X \times X$. Similarly, (3.11) implies (3.10), so that these are indeed equivalent.

Finally, concerning the coincidence of the kernels P and $\bar{P}$ in (iii), beginning from (3.1) and applying (3.5) followed by (3.10), we have

$$\begin{aligned}
 (3.16) \quad & \bar{P}(\mathbf{q}, d\tilde{\mathbf{q}}) \\
 &= \alpha(\mathbf{q}, \tilde{\mathbf{q}}) Q(\mathbf{q}, d\tilde{\mathbf{q}}) + \delta_{\mathbf{q}}(d\tilde{\mathbf{q}}) \int_X (1 - \alpha(\mathbf{q}, \mathbf{r})) Q(\mathbf{q}, d\mathbf{r}) \\
 &= \alpha(\mathbf{q}, \tilde{\mathbf{q}}) F(\mathbf{q}, \cdot)^* \mathcal{V}(\mathbf{q}, \cdot)(d\tilde{\mathbf{q}}) + \delta_{\mathbf{q}}(d\tilde{\mathbf{q}}) \int_X (1 - \alpha(\mathbf{q}, \mathbf{r})) F(\mathbf{q}, \cdot)^* \mathcal{V}(\mathbf{q}, \cdot)(d\mathbf{r}) \\
 &= \int_X \delta_{\mathbf{r}}(d\tilde{\mathbf{q}}) \alpha(\mathbf{q}, \mathbf{r}) F(\mathbf{q}, \cdot)^* \mathcal{V}(\mathbf{q}, \cdot)(d\mathbf{r}) + \delta_{\mathbf{q}}(d\tilde{\mathbf{q}}) \int_X (1 - \alpha(\mathbf{q}, \mathbf{r})) F(\mathbf{q}, \cdot)^* \mathcal{V}(\mathbf{q}, \cdot)(d\mathbf{r}) \\
 &= \int_X \delta_{F(\mathbf{q}, \mathbf{v})}(d\tilde{\mathbf{q}}) \alpha(\mathbf{q}, F(\mathbf{q}, \mathbf{v})) \mathcal{V}(\mathbf{q}, d\mathbf{v}) + \delta_{\mathbf{q}}(d\tilde{\mathbf{q}}) \int_X (1 - \alpha(\mathbf{q}, F(\mathbf{q}, \mathbf{v}))) \mathcal{V}(\mathbf{q}, d\mathbf{v}) \\
 &= \int_X \delta_{\Pi_1 \circ S(\mathbf{q}, \mathbf{v})}(d\tilde{\mathbf{q}}) \hat{\alpha}(\mathbf{q}, \mathbf{v}) \mathcal{V}(\mathbf{q}, d\mathbf{v}) + \delta_{\mathbf{q}}(d\tilde{\mathbf{q}}) \int_X (1 - \hat{\alpha}(\mathbf{q}, \mathbf{v})) \mathcal{V}(\mathbf{q}, d\mathbf{v}),
 \end{aligned}$$

which is (2.12). The proof is now complete. $\square$

3.3. An alternative proof of Theorem 2.1. Turning to our second task in this section we now show how the formulation in [72] can be employed to develop a second independent proof of Theorem 2.1. Here we proceed by defining an appropriate proposal kernel Q on a product space $X \times Y$, with (X, Σ_X) and (Y, Σ_Y) being any measurable spaces, and then taking an appropriate marginal of the corresponding transition kernel $\bar{P}$ as in (3.1).

We start with a more general result which is actually independent of this particular product structure.

THEOREM 3.6. *Let $(\mathcal{X}, \Sigma_{\mathcal{X}})$ be a measurable space and let $\mathcal{M}$ be any probability measure on $\mathcal{X}$. Suppose there exists a measurable mapping $S : \mathcal{X} \rightarrow \mathcal{X}$ satisfying the following properties:*

- (P1) S is an involution, that is, $S^2 = I$;
- (P2) $S^* \mathcal{M}$ is absolutely continuous with respect to $\mathcal{M}$.

Consider the Markov kernel Q on $\mathcal{X}$ defined as

$$(3.17) \quad Q(\mathbf{u}, d\tilde{\mathbf{u}}) = \delta_{S(\mathbf{u})}(d\tilde{\mathbf{u}}), \quad \mathbf{u} \in \mathcal{X},$$

and let $\eta, \eta^\perp$ be the measures on $\mathcal{X} \times \mathcal{X}$ given as

$$(3.18) \quad \eta(d\mathbf{u}, d\tilde{\mathbf{u}}) = Q(\mathbf{u}, d\tilde{\mathbf{u}}) \mathcal{M}(d\mathbf{u}), \quad \eta^\perp(d\mathbf{u}, d\tilde{\mathbf{u}}) = \eta(d\tilde{\mathbf{u}}, d\mathbf{u}).$$

Then η and $\eta^\perp$ are mutually absolutely continuous, with

$$(3.19) \quad \frac{d\eta^\perp}{d\eta}(\mathbf{u}, \tilde{\mathbf{u}}) = \frac{dS^* \mathcal{M}}{d\mathcal{M}}(\mathbf{u}) \quad \text{for } \eta\text{-a.e. } (\mathbf{u}, \tilde{\mathbf{u}}) \in \mathcal{X} \times \mathcal{X}.$$

Consequently, the Markov kernel $\bar{P}$ on $\mathcal{X}$ defined as in (3.1) for Q, α, η and $\eta^\perp$ as in (3.17), (3.2) and (3.18), respectively, written here as

$$(3.20) \quad \bar{P}(\mathbf{u}, d\tilde{\mathbf{u}}) = \alpha(\mathbf{u}, \tilde{\mathbf{u}}) \delta_{S(\mathbf{u})}(d\tilde{\mathbf{u}}) + \delta_{\mathbf{u}}(d\tilde{\mathbf{u}}) [1 - \alpha(\mathbf{u}, S(\mathbf{u}))],$$

satisfies detailed balance with respect to $\mathcal{M}$.

PROOF. Let $\varphi : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ be any bounded and measurable function. Recalling the definition of η and $\eta^\perp$ from (3.18), we have

$$\begin{aligned} \int_{\mathcal{X} \times \mathcal{X}} \varphi(\mathbf{u}, \tilde{\mathbf{u}}) \eta^\perp(d\mathbf{u}, d\tilde{\mathbf{u}}) &= \int_{\mathcal{X} \times \mathcal{X}} \varphi(\mathbf{u}, \tilde{\mathbf{u}}) \eta(d\tilde{\mathbf{u}}, d\mathbf{u}) = \int_{\mathcal{X} \times \mathcal{X}} \varphi(\mathbf{u}, \tilde{\mathbf{u}}) \delta_{S(\tilde{\mathbf{u}})}(d\mathbf{u}) \mathcal{M}(d\tilde{\mathbf{u}}) \\ &= \int_{\mathcal{X}} \varphi(S(\tilde{\mathbf{u}}), \tilde{\mathbf{u}}) \mathcal{M}(d\tilde{\mathbf{u}}). \end{aligned}$$

Now invoking properties (P1) and (P2) of S , we deduce that

$$\begin{aligned} \int_{\mathcal{X} \times \mathcal{X}} \varphi(\mathbf{u}, \tilde{\mathbf{u}}) \eta^\perp(d\mathbf{u}, d\tilde{\mathbf{u}}) &= \int_{\mathcal{X}} \varphi(S(\tilde{\mathbf{u}}), S^2(\tilde{\mathbf{u}})) \mathcal{M}(d\tilde{\mathbf{u}}) = \int_{\mathcal{X}} \varphi(\mathbf{u}, S(\mathbf{u})) S^* \mathcal{M}(d\mathbf{u}) \\ &= \int_{\mathcal{X}} \varphi(\mathbf{u}, S(\mathbf{u})) \frac{dS^* \mathcal{M}}{d\mathcal{M}}(\mathbf{u}) \mathcal{M}(d\mathbf{u}) \\ &= \int_{\mathcal{X} \times \mathcal{X}} \varphi(\mathbf{u}, \tilde{\mathbf{u}}) \frac{dS^* \mathcal{M}}{d\mathcal{M}}(\mathbf{u}) \delta_{S(\mathbf{u})}(d\tilde{\mathbf{u}}) \mathcal{M}(d\mathbf{u}) \\ &= \int_{\mathcal{X} \times \mathcal{X}} \varphi(\mathbf{u}, \tilde{\mathbf{u}}) \frac{dS^* \mathcal{M}}{d\mathcal{M}}(\mathbf{u}) \eta(d\mathbf{u}, d\tilde{\mathbf{u}}). \end{aligned}$$

This shows that $\eta^\perp \ll \eta$ and that (3.19) holds. Similarly, we can show that $\eta \ll \eta^\perp$, so that in fact η and $\eta^\perp$ are mutually absolutely continuous.

The observation that $\bar{P}$ defined in (3.1) can be written as (3.20) for Q as in (3.17) is clear. Finally, the fact that $\bar{P}$ is reversible with respect to $\mathcal{M}$ follows as a consequence of the general result given in [72], Theorem 2. The proof is complete. $\square$

Applying Theorem 3.6 combined with a suitable marginalization of the Markov kernel $\bar{P}$ in (3.20) now yields the result of Theorem 2.1.

COROLLARY 3.7. *Let (X, Σ_X) and (Y, Σ_Y) be measurable spaces. Let μ be a probability measure on X , and $\mathcal{V} : X \times \Sigma_Y \rightarrow [0, 1]$ be a Markov kernel. Consider $\mathcal{X} = X \times Y$ and suppose there exists a measurable mapping $S : \mathcal{X} \rightarrow \mathcal{X}$ satisfying properties (P1)–(P2) from Theorem 2.1 (with $\mathcal{M}$ specified as in (2.10) relative to inputs $\mathcal{V}$ and μ given here). Let $\bar{P}$ be the Markov kernel on $\mathcal{X}$ defined as in (3.20), and define P to be the Markov kernel on X given as*

$$(3.21) \quad P(\mathbf{q}, d\tilde{\mathbf{q}}) := \int_Y \int_Y \bar{P}((\mathbf{q}, \mathbf{v}), d\tilde{\mathbf{q}} d\tilde{\mathbf{v}}) \mathcal{V}(\mathbf{q}, d\mathbf{v}) \quad \text{for all } \mathbf{q} \in X \text{ and } A \in \Sigma_X,$$

or equivalently as $P(\mathbf{q}, A) := \int_Y \bar{P}((\mathbf{q}, \mathbf{v}), A \times Y) \mathcal{V}(\mathbf{q}, d\mathbf{v})$ for all $\mathbf{q} \in X$ and $A \in \Sigma_X$. Then P satisfies detailed balance with respect to μ . Moreover, $P(\mathbf{q}, \cdot)$ coincides with the definition given in (2.12) for μ -a.e. $\mathbf{q} \in X$.

PROOF. The statement that P defined in (3.21) satisfies detailed balance with respect to μ follows immediately from the fact that $\bar{P}$ satisfies detailed balance with respect to $\mathcal{M}$ defined in (2.10), as a consequence of Theorem 3.6. For the second claim, notice that for each fixed $\mathbf{q} \in X$ and $A \in \Sigma_X$ we have according to the definition of $\bar{P}$ in (3.20) that

$$\begin{aligned} P(\mathbf{q}, A) &= \int_Y \int_{A \times Y} \alpha((\mathbf{q}, \mathbf{v}), (\tilde{\mathbf{q}}, \tilde{\mathbf{v}})) \delta_{S(\mathbf{q}, \mathbf{v})}(d\tilde{\mathbf{q}}, d\tilde{\mathbf{v}}) \mathcal{V}(\mathbf{q}, d\mathbf{v}) \\ (3.22) \quad &+ \int_Y \int_{A \times Y} \delta_{(\mathbf{q}, \mathbf{v})}(d\tilde{\mathbf{q}}, d\tilde{\mathbf{v}}) [1 - \alpha((\mathbf{q}, \mathbf{v}), S(\mathbf{q}, \mathbf{v}))] \mathcal{V}(\mathbf{q}, d\mathbf{v}) \\ &=: (I_1) + (I_2). \end{aligned}$$

For the first term, we write

$$\begin{aligned}
 (I_1) &= \int_Y \int_{X \times Y} \mathbb{1}_{A \times Y}(\tilde{\mathbf{q}}, \tilde{\mathbf{v}}) \alpha((\mathbf{q}, \mathbf{v}), (\tilde{\mathbf{q}}, \tilde{\mathbf{v}})) \delta_{S(\mathbf{q}, \mathbf{v})}(d\tilde{\mathbf{q}}, d\tilde{\mathbf{v}}) \mathcal{V}(\mathbf{q}, d\mathbf{v}) \\
 (3.23) \quad &= \int_Y \mathbb{1}_{A \times Y}(S(\mathbf{q}, \mathbf{v})) \alpha((\mathbf{q}, \mathbf{v}), S(\mathbf{q}, \mathbf{v})) \mathcal{V}(\mathbf{q}, d\mathbf{v}) \\
 &= \int_Y \int_A \delta_{\Pi_1 \circ S(\mathbf{q}, \mathbf{v})}(d\tilde{\mathbf{q}}) \alpha((\mathbf{q}, \mathbf{v}), S(\mathbf{q}, \mathbf{v})) \mathcal{V}(\mathbf{q}, d\mathbf{v}),
 \end{aligned}$$

while for the second term

$$(3.24) \quad (I_2) = \int_A \delta_{\mathbf{q}}(d\tilde{\mathbf{q}}) \int_Y [1 - \alpha((\mathbf{q}, \mathbf{v}), S(\mathbf{q}, \mathbf{v}))] \mathcal{V}(\mathbf{q}, d\mathbf{v}).$$

Moreover, it is not difficult to show that (3.19) implies

$$\frac{d\eta^\perp}{d\eta}((\mathbf{q}, \mathbf{v}), S(\mathbf{q}, \mathbf{v})) = \frac{dS^* \mathcal{M}}{d\mathcal{M}}(\mathbf{q}, \mathbf{v}) \quad \text{for } \mathcal{M}\text{-a.e. } (\mathbf{q}, \mathbf{v}) \in X \times Y,$$

so that

$$(3.25) \quad \alpha((\mathbf{q}, \mathbf{v}), S(\mathbf{q}, \mathbf{v})) = \hat{\alpha}(\mathbf{q}, \mathbf{v}) \quad \text{for } \mathcal{M}\text{-a.e. } (\mathbf{q}, \mathbf{v}) \in X \times X,$$

for $\hat{\alpha}$ as defined in (2.11).

Consequently, from (3.22)–(3.24) and (3.25), we conclude that for μ -a.e. $\mathbf{q} \in X$ and every $A \in \Sigma_X$

$$P(\mathbf{q}, A) = \int_X \int_A \delta_{\Pi_1 \circ S(\mathbf{q}, \mathbf{v})}(d\tilde{\mathbf{q}}) \hat{\alpha}(\mathbf{q}, \mathbf{v}) \mathcal{V}(\mathbf{q}, d\mathbf{v}) + \int_A \delta_{\mathbf{q}}(d\tilde{\mathbf{q}}) \int_X [1 - \hat{\alpha}(\mathbf{q}, \mathbf{v})] \mathcal{V}(\mathbf{q}, d\mathbf{v}),$$

which coincides with the Markov kernel defined in (2.12). This finishes the proof. $\square$

REMARK 3.8 (Connection to [3]). We note that [3] uses a similar two-step extended phase space presentation to develop a general result for Metropolis–Hastings kernels. Specifically, Theorem 3 in [3] is analogous to Theorem 3.6, whereas Proposition 1 in [3] corresponds to Corollary 3.7.

REMARK 3.9. In regard to the results of Theorem 3.6 and Corollary 3.7, we notice that a slightly different approach to constructing a Markov transition kernel on a product space $X \times Y$ is provided in [27], Algorithm 1 (see also [17], Section 5.3), described as follows. Assume, similarly as in Remark 2.5, that the mapping S from Theorem 3.6 is given as $R \circ \hat{S}$, with $R, \hat{S} : X \times Y \rightarrow X \times Y$ being two measurable mappings such that $S = R \circ \hat{S}$ satisfies assumptions (P1)–(P2) from Theorem 2.1 and, additionally, that $R^* \mathcal{M} = \mathcal{M}$. Then, the Markov transition kernel resulting from [27], Algorithm 1, can be written as

$$(3.26) \quad \bar{P}'(\mathbf{u}, d\tilde{\mathbf{u}}) = \alpha'(\mathbf{u}) \delta_{\hat{S}(\mathbf{u})}(d\tilde{\mathbf{u}}) + \delta_{R(\mathbf{u})}(d\tilde{\mathbf{u}}) [1 - \alpha'(\mathbf{u})],$$

where

$$(3.27) \quad \alpha'(\mathbf{u}) = 1 \wedge \frac{d(\hat{S}^{-1})^* \mathcal{M}}{d\mathcal{M}}(\mathbf{u}) \quad \text{for } \mathbf{u} \in X \times Y,$$

with $\mathcal{M}$ as defined in (2.10). Here assumptions (P1)–(P2) together with $R^* \mathcal{M} = \mathcal{M}$ imply that the Radon–Nikodym derivative in (3.27) is well defined, since

$$(\hat{S}^{-1})^* \mathcal{M} = (S^{-1} \circ R)^* \mathcal{M} = (S^{-1})^* R^* \mathcal{M} = (S^{-1})^* \mathcal{M} = S^* \mathcal{M}.$$

In fact, in the setting from [27], X and Y are taken as finite-dimensional spaces, and α' is written in terms of the density of $\mathcal{M}$ with respect to the corresponding Lebesgue measure in $X \times Y$.

Most importantly, we notice that in contrast to $\bar{P}$ defined in (3.20), the transition kernel $\bar{P}'$ in (3.26) does *not* in general satisfy detailed balance with respect to the probability measure $\mathcal{M}$. Indeed, it is shown in [27], Section V, that $\bar{P}'$ satisfies a modified form of the standard detailed balance condition. Nevertheless, if we assume in addition that R is such that $\Pi_1 \circ R = \Pi_1$ then by taking the marginalization of $\bar{P}'$ as in (3.21) it is not difficult to show that we obtain a Markov kernel on X that also coincides a.e. with the one from (2.12), and thus satisfies detailed balance with respect to the first marginal of $\mathcal{M}$, that is, $\mu = \Pi_1^* \mathcal{M}$.

4. Approximate Hamiltonian Monte Carlo methods. This section derives and studies some “extended phase space” sampling methods. We first consider the classical finite-dimensional setting involving a continuously distributed target measure in Section 4.1 leading to Algorithm 4.1, Algorithm 4.2, Algorithm 4.3 and Algorithm 4.4. We then turn to the infinite-dimensional Gaussian–Hilbert space framework in Section 4.2 culminating in Algorithm 4.5.

The developments presented here are ultimately based on an application of Theorem 2.1 while drawing on an extended library of HMC samplers sitting on the foundation of a rich variety of numerical methods employed for the effective discretization of Hamiltonian systems. We provide an essentially self-contained presentation of some elements of this extensive and disparate literature laying out a toolbox which can be expanded upon to derive further algorithms in the future.

On the one hand the results below can be seen as representing a class of surrogate trajectory methods, reflective of a growing body of literature; cf. [5, 8, 47, 50, 51, 56, 60, 61, 64, 65, 75, 76]. These works seek to partially avoid expensive gradient computations dictated by classical HMC formulations. In this connection it is notable that our methods allow for a variety of gradient-free approximations where the dynamics may not be symplectic but are merely volume-preserving (see Definition 4.1 below). From a slightly different perspective we may see the results in this section as providing a large “parameter space” of possible samplers which include many popular and recently discovered methods as important special cases. We make explicit the parameter choices connecting back to existing methods below in Section 5, further illustrating and enriching the unifying outlook provided by Theorem 2.1.

4.1. The finite-dimensional case. We begin with the classical setting where the target measure μ sits on $\mathbb{R}^N$. Recall that, in this finite-dimensional context, our goal is to sample measures μ , which are presumed to be continuously distributed. For convenience, we also assume that μ is strictly positive, but see Remark 2.4. Thus, we may write μ in the potential form

$$(4.1) \quad \mu(d\mathbf{q}) = \frac{1}{Z_{\mathcal{U}}} e^{-\mathcal{U}(\mathbf{q})} d\mathbf{q}, \quad Z_{\mathcal{U}} = \int_{\mathbb{R}^N} e^{-\mathcal{U}(\mathbf{q})} d\mathbf{q},$$

where $\mathcal{U} : \mathbb{R}^N \rightarrow \mathbb{R}$ is any measurable function such that $e^{-\mathcal{U}(\mathbf{q})} \in L^1(\mathbb{R}^N)$. Typically we will additionally suppose that $\mathcal{U} \in C^1$.

In order to develop our results we need to introduce some elements from the theory of geometric integration and from Hamiltonian dynamical systems more broadly. We have tried to keep our discussion here as elementary and as self-contained as possible but it includes a number of results which are covered in much more detail and with a much wider scope elsewhere. See, for example, [17, 37, 48, 61] and we refer to [4, 45, 54] for the broader context of Hamiltonian systems.

4.1.1. *Hamiltonian and surrogate extended phase space dynamics.* The starting point of the HMC approach involves selecting a Hamiltonian function $\mathcal{H} : \mathbb{R}^{2N} \rightarrow \mathbb{R}$ such that the marginal of the associated Gibbs measure

$$(4.2) \quad \mathcal{M}(d\mathbf{q}, d\mathbf{v}) := \frac{1}{Z_{\mathcal{H}}} e^{-\mathcal{H}(\mathbf{q}, \mathbf{v})} d\mathbf{q} d\mathbf{v}, \quad Z_{\mathcal{H}} = \int_{\mathbb{R}^{2N}} e^{-\mathcal{H}(\mathbf{q}, \mathbf{v})} d\mathbf{q} d\mathbf{v},$$

with respect to the “position” variable $\mathbf{q}$ coincides with the target measure μ defined in (4.1). As such, it is natural to consider a Hamiltonian given by

$$(4.3) \quad \mathcal{H}(\mathbf{q}, \mathbf{v}) = \mathcal{U}(\mathbf{q}) + \mathcal{K}(\mathbf{q}, \mathbf{v}) + \ln Z_{\mathcal{K}}(\mathbf{q}), \quad Z_{\mathcal{K}}(\mathbf{q}) := \int_{\mathbb{R}^N} e^{-\mathcal{K}(\mathbf{q}, \mathbf{v})} d\mathbf{v}$$

for some C^1 function $\mathcal{K} : \mathbb{R}^{2N} \rightarrow \mathbb{R}$ such that $\mathbf{v} \mapsto e^{-\mathcal{K}(\mathbf{q}, \mathbf{v})} \in L^1(\mathbb{R}^N)$, for each fixed $\mathbf{q} \in \mathbb{R}^N$. Here the term $\ln Z_{\mathcal{K}}(\mathbf{q})$ is included precisely to ensure that the marginal of $\mathcal{M}$ with respect to $\mathbf{q}$ coincides with μ . Therefore, $\mathcal{M}$ from (4.2) can be written as

$$(4.4) \quad \mathcal{M}(d\mathbf{q}, d\mathbf{v}) = \mathcal{V}(\mathbf{q}, d\mathbf{v}) \mu(d\mathbf{q}), \quad \mathcal{V}(\mathbf{q}, d\mathbf{v}) = \frac{1}{Z_{\mathcal{K}}(\mathbf{q})} e^{-\mathcal{K}(\mathbf{q}, \mathbf{v})} d\mathbf{v},$$

with $Z_{\mathcal{H}} = Z_{\mathcal{U}}$ from (4.1). Denoting by $\mathcal{B}(\mathbb{R}^N)$ the σ -algebra of Borel sets in $\mathbb{R}^N$, it follows by construction that $\mathcal{V} : \mathbb{R}^N \times \mathcal{B}(\mathbb{R}^N) \rightarrow [0, 1]$ is a Markov kernel, and $\mathcal{M}$ defines a probability measure on $\mathbb{R}^{2N}$.

Note that typically one considers

$$(4.5) \quad \mathcal{K}(\mathbf{q}, \mathbf{v}) = \frac{1}{2} \langle M(\mathbf{q})^{-1} \mathbf{v}, \mathbf{v} \rangle$$

for an appropriately chosen symmetric positive definite “mass matrix” M , so that $\mathcal{K}(\mathbf{q}, \mathbf{v})$ corresponds to the negative log-density of the $\mathbb{R}^N$ -valued gaussian distribution $N(0, M(\mathbf{q}))$. Classically, M is $\mathbf{q}$ independent and often simply taken to be the identity (but see the infinite-dimensional formulation in Section 4.2 and Section 5.2 below). On the other hand, the Riemannian manifold HMC method introduced in [33] considers cases where we introduce a dependence on $\mathbf{q}$ in M in (4.5), thus providing an important motivating example for allowing “position-dependence” in the formulation of the kinetic portion of the Hamiltonian in (4.3). Note also that “non-Gaussian” choices for $\mathcal{K}$ are also relevant. See, for example, the relativistic HMC algorithm developed in [53]. Both of these HMC variants are briefly described in Section 5.1.4 below where they are connected back to the generalized frameworks we consider here.

Having determined $\mathcal{H}$ as in (4.3) one now considers the associated Hamiltonian dynamics for the pair $\mathbf{z} = (\mathbf{q}, \mathbf{v}) \in \mathbb{R}^{2N}$ as

$$(4.6) \quad \frac{d\mathbf{z}}{dt} = J^{-1} \nabla \mathcal{H}(\mathbf{z}), \quad \mathbf{z}(0) = (\mathbf{q}_0, \mathbf{v}_0),$$

where J is any $2N \times 2N$ real matrix which is antisymmetric and invertible. Here the typical form for J is

$$(4.7) \quad J := \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix}$$

but other “noncanonical” choices for J are relevant for sampling. For example, in the infinite-dimensional version of HMC derived in [10], which we consider in Section 4.2, Section 5.2.3

below, J is used to “precondition” the dynamics, see (4.52). Other possibilities for J are studied, for example, with the so called magnetic HMC methods introduced in [73].

Since $\mathcal{H}$ is invariant under the flow associated to (4.6), the Gibbs measure $\mathcal{M}$ given as (4.2) is invariant with respect to this flow. This implies that the Markov transition kernel associated to the dynamics of the $\mathbf{q}$ variable in (4.6), given by

$$(4.8) \qquad P^t(\mathbf{q}_0, A) := \mathbb{P}(\mathbf{q}(t; \mathbf{q}_0, \mathbf{v}_0) \in A), \quad \mathbf{v}_0 \sim \mathcal{V}(\mathbf{q}_0, \cdot),$$

defined for some $t \geq 0$, for all $\mathbf{q}_0 \in \mathbb{R}^N$ and any Borel set $A \subset \mathbb{R}^N$, holds the $\mathbf{q}$ -marginal of $\mathcal{M}$, that is, the target μ , as an invariant measure. For a fixed integration time $T > 0$, P^T as given in (4.8) defines the Markov kernel for what is known as the exact HMC algorithm, which is of theoretical interest as an idealization of HMC, [16, 35]. However this P^T is of much less practical significance because it is typically impossible to exactly resolve the solution operator for (4.6). Instead, one resorts to a skillfully chosen numerical approximation $\hat{S}(\mathbf{q}_0, \mathbf{v}_0)$ for the solution of (4.6) at the time T that is commensurate with the setting of Theorem 2.1.

In view of obtaining a wider class of HMC-like algorithms to sample from μ in (4.1), we replace (4.6) with the following general dynamics:

$$(4.9) \qquad \frac{d\mathbf{q}}{dt} = f_1(\mathbf{q}, \mathbf{v}), \qquad \frac{d\mathbf{v}}{dt} = f_2(\mathbf{q}, \mathbf{v}), \qquad (\mathbf{q}(0), \mathbf{v}(0)) = (\mathbf{q}_0, \mathbf{v}_0),$$

for suitably regular functions $f_1 : \mathbb{R}^{2N} \rightarrow \mathbb{R}^N$ and $f_2 : \mathbb{R}^{2N} \rightarrow \mathbb{R}^N$. Here an underlying idea for considering the more general dynamic (4.9) is that we may aim to replace the right-hand side of (4.6), $J^{-1} \nabla \mathcal{H}(\mathbf{z})$, with an artfully chosen approximation $(f_1(\mathbf{q}, \mathbf{v}), f_2(\mathbf{q}, \mathbf{v}))$, one that is computationally cheaper to evaluate while maintaining essential features of $J^{-1} \nabla \mathcal{H}(\mathbf{z})$. Therefore (4.9) is the starting point for a methodology to resolve the target μ with a lower overall computational cost. In this connection notice that, in contrast to (4.6), the system (4.9) for general functions f_1 and f_2 may not be a Hamiltonian system, and also may not be expected to hold $\mathcal{M}$ as an invariant measure.

Below we illustrate some classes of MCMC algorithms resulting from such f_1, f_2 which still preserve the target measure μ as invariant, as long as the integrator $\hat{S}$ and accept–reject function $\hat{\alpha}$ are chosen appropriately. In Proposition 4.3 below, we show how Theorem 2.1 can be used to derive an appropriate accept–reject mechanism assuming certain natural structural properties of the map $\hat{S}$. Then, in a series of subsections, we introduce three classes of algorithms developed around different considerations for approximating (4.6) with (4.9); see (4.23), (4.30) and (4.36)–(4.37) below.

4.1.2. *Case 0: General extended phase space methods.* Before introducing our first and most general algorithm class, let us first recall some basic definitions from the theory of Hamiltonian dynamical systems which we need here and below.

DEFINITION 4.1. Let $\hat{S} : \mathbb{R}^{2N} \rightarrow \mathbb{R}^{2N}$ be a C^1 map.

(i) Fix any linear invertible map $R : \mathbb{R}^{2N} \rightarrow \mathbb{R}^{2N}$. We say that $\hat{S}$ is *reversible with respect to R* (or simply *R -reversible*) if $\hat{S}$ is itself invertible and

$$(4.10) \qquad R \circ \hat{S}(\mathbf{z}) = \hat{S}^{-1} \circ R(\mathbf{z})$$

for every $\mathbf{z} \in \mathbb{R}^{2N}$.

(ii) We say that $\hat{S}$ is *symplectic*, with respect to an invertible (typically antisymmetric) matrix J , if

$$(4.11) \quad (\nabla \hat{S}(\mathbf{z}))^* J \nabla \hat{S}(\mathbf{z}) = J$$

for every $\mathbf{z} \in \mathbb{R}^{2N}$, where A^* denotes the conjugate transpose of a matrix A .⁵

Let us collect a few elementary properties of symplectic and R -reversible mappings under Definition 4.1 whose proofs are immediate.

LEMMA 4.2. (i) Under (4.10) it follows that if additionally R is an involution, that is, $R \circ R = I$, then $R \circ \hat{S}$ is an involution, that is,

$$(4.12) \quad R \circ \hat{S} \circ R \circ \hat{S}(\mathbf{z}) = \mathbf{z}$$

for every $\mathbf{z} \in \mathbb{R}^{2N}$.

(ii) If $\hat{S}$ is symplectic, with respect to any invertible matrix J , then $\hat{S}$ is volume preserving in $\mathbb{R}^{2N}$, namely

$$(4.13) \quad |\det \nabla \hat{S}(\mathbf{z})| = 1$$

for every $\mathbf{z} \in \mathbb{R}^{2N}$.

(iii) If $\hat{S}_1, \hat{S}_2$ are two symplectic mappings, then their composition $\hat{S}_1 \circ \hat{S}_2$ is also symplectic. Similarly, under the weaker condition that $\hat{S}_1, \hat{S}_2$ are both volume preserving, à la (4.13), then so too $\hat{S}_1 \circ \hat{S}_2$ must be volume preserving.

It is immediately clear that these two properties introduced in Definition 4.1 together with Lemma 4.2 are tailor-made for Theorem 2.1. See also Remark 2.6 and the identity (2.20). We formalize this as follows.

PROPOSITION 4.3. Fix any C^1 potential functions $\mathcal{U} : \mathbb{R}^N \rightarrow \mathbb{R}$ and $\mathcal{K} : \mathbb{R}^{2N} \rightarrow \mathbb{R}$ so that we can define a probability measure $\mu(d\mathbf{q}) = Z_{\mathcal{U}}^{-1} e^{-\mathcal{U}(\mathbf{q})} d\mathbf{q}$ and a Markov kernel $\mathcal{V}(\mathbf{q}, d\mathbf{v}) = Z_{\mathcal{K}}(\mathbf{q})^{-1} e^{-\mathcal{K}(\mathbf{q}, \mathbf{v})} d\mathbf{v}$ as in (4.1) and (4.4), respectively. Consider the associated Hamiltonian $\mathcal{H} = \mathcal{U} + \mathcal{K} + \ln Z_{\mathcal{U}}$ as defined in (4.3). Let $\hat{S} : \mathbb{R}^{2N} \rightarrow \mathbb{R}^{2N}$ be a C^1 mapping which is reversible with respect to a linear involution R as in Definition 4.1(i). Then:

(i) the kernel P defined as in (2.12) with $S = R \circ \hat{S}$ and with $\hat{\alpha}$ defined as

$$(4.14) \quad \hat{\alpha}(\mathbf{q}, \mathbf{v}) = 1 \wedge [\exp(-\mathcal{H}(R \circ \hat{S}(\mathbf{q}, \mathbf{v})) + \mathcal{H}(\mathbf{q}, \mathbf{v})) |\det \nabla \hat{S}(\mathbf{q}, \mathbf{v})|]$$

satisfies detailed balance with respect to μ as in (2.13).

(ii) If we furthermore assume that $\hat{S}$ is symplectic, à la (4.11), or merely volume-preserving as in (4.13) then $\hat{\alpha}$ reduces to

$$(4.15) \quad \hat{\alpha}(\mathbf{q}, \mathbf{v}) = 1 \wedge [\exp(-\mathcal{H}(R \circ \hat{S}(\mathbf{q}, \mathbf{v})) + \mathcal{H}(\mathbf{q}, \mathbf{v}))].$$

⁵Equivalently, one may consider the symplectic form $\Omega(\tilde{\mathbf{z}}, \bar{\mathbf{z}}) := \langle \tilde{\mathbf{z}}, J\bar{\mathbf{z}} \rangle$, for $\tilde{\mathbf{z}}, \bar{\mathbf{z}} \in \mathbb{R}^{2N}$ and assert that $\hat{S}$ is symplectic if Ω is invariant under the pullback by $\hat{S}$, that is, we have that $\hat{S}^* \Omega = \Omega$. Here recall that, in this context, $(\hat{S}^* \Omega)_{\mathbf{w}}(\tilde{\mathbf{z}}, \bar{\mathbf{z}}) := \Omega(\nabla \hat{S}(\mathbf{w})\tilde{\mathbf{z}}, \nabla \hat{S}(\mathbf{w})\bar{\mathbf{z}})$, for $\tilde{\mathbf{z}}, \bar{\mathbf{z}}, \mathbf{w} \in \mathbb{R}^{2N}$. We observe that, in the case where J is canonical, namely when J is of the form (4.7), then we may write $\Omega = d\mathbf{q} \wedge d\mathbf{v} := \sum_{j=1}^N dq_j \wedge dv_j$ with d the exterior derivative and $\wedge$ the wedge product so that $dq_j \wedge dv_j(\tilde{\mathbf{z}}, \bar{\mathbf{z}}) = \tilde{q}_j \bar{v}_j - \bar{q}_j \tilde{v}_j$ for $\tilde{\mathbf{z}} = (\tilde{\mathbf{q}}, \tilde{\mathbf{v}})$, $\bar{\mathbf{z}} = (\bar{\mathbf{q}}, \bar{\mathbf{v}})$. See, for example, [74] for basic definitions. As identified in, for example, [48], Chapter 4.1, we therefore have that $\hat{S}$ is symplectic with respect to the canonical form Ω when $d\mathbf{q} \wedge d\mathbf{v} = d\bar{\mathbf{q}} \wedge d\bar{\mathbf{v}}$ where $(\mathbf{q}, \mathbf{v}) = \hat{S}(\bar{\mathbf{q}}, \bar{\mathbf{v}})$.

Algorithm 4.1 Generic extended phase space algorithm to sample from $\mu(d\mathbf{q}) = Z_{\mathcal{U}}^{-1} e^{-\mathcal{U}(\mathbf{q})} d\mathbf{q}$

1: Select the algorithm parameters:

- (i) The momentum kernel $\mathcal{V}(\mathbf{q}, d\mathbf{v}) = Z_{\mathcal{K}}(\mathbf{q})^{-1} e^{-\mathcal{K}(\mathbf{q}, \mathbf{v})} d\mathbf{v} \in \text{Pr}(\mathbb{R}^N)$, for each fixed $\mathbf{q} \in \mathbb{R}^N$.
- (ii) The integrator $\hat{S}$, and the linear involution R such that $\hat{S}$ is R -reversible (see (4.10), (4.12)).

2: Choose $\mathbf{q}_0 \in \mathbb{R}^N$.

3: **for** $k \geq 0$ **do**

4: Sample $\mathbf{v}_k \sim \mathcal{V}(\mathbf{q}_k, \cdot)$.

5: Propose $\bar{\mathbf{q}}_{k+1} := \Pi_1 \circ R \circ \hat{S}(\mathbf{q}_k, \mathbf{v}_k)$, where $\Pi_1(\mathbf{q}, \mathbf{v}) = \mathbf{q}$.

6: Set $\mathbf{q}_{k+1} := \bar{\mathbf{q}}_{k+1}$ with probability $\hat{\alpha}(\mathbf{q}_k, \mathbf{v}_k)$ for $\hat{\alpha}$ given by (4.14), and otherwise take $\mathbf{q}_{k+1} := \mathbf{q}_k$.

(iii) On the other hand, if we assume that $\mathcal{H}$ is invariant under R , namely

$$(4.16) \quad \mathcal{H}(\mathbf{q}, \mathbf{v}) = \mathcal{H}(R(\mathbf{q}, \mathbf{v})),$$

then $\hat{\alpha}$ becomes

$$(4.17) \quad \hat{\alpha}(\mathbf{q}, \mathbf{v}) = 1 \wedge [\exp(-\mathcal{H}(\hat{S}(\mathbf{q}, \mathbf{v})) + \mathcal{H}(\mathbf{q}, \mathbf{v})) |\det \nabla \hat{S}(\mathbf{q}, \mathbf{v})|].$$

(iv) Finally if both $\hat{S}$ is volume-preserving and (4.16) holds we can take $\hat{\alpha}$ as

$$(4.18) \quad \hat{\alpha}(\mathbf{q}, \mathbf{v}) = 1 \wedge \exp(-\mathcal{H}(\hat{S}(\mathbf{q}, \mathbf{v})) + \mathcal{H}(\mathbf{q}, \mathbf{v})).$$

PROOF. Let $\mathcal{M}$ be the probability measure on $\mathbb{R}^{2N}$ defined as in (4.2), (4.4), namely $\mathcal{M}(d\mathbf{q}, d\mathbf{v}) = \mathcal{V}(\mathbf{q}, d\mathbf{v}) \mu(d\mathbf{q}) = Z_{\mathcal{H}}^{-1} e^{-\mathcal{H}(\mathbf{q}, \mathbf{v})} d\mathbf{q} d\mathbf{v}$. Since $\mathcal{M}$ has a strictly positive density with respect to the Lebesgue measure and, by Lemma 4.2(i), $S = R \circ \hat{S}$ is an involution, then the result follows directly from Theorem 2.1 once we compute $dS^* \mathcal{M} / d\mathcal{M}$. Here we use (2.3) and (2.5) to obtain that

$$(4.19) \quad \begin{aligned} \frac{dS^* \mathcal{M}}{d\mathcal{M}}(\mathbf{q}, \mathbf{v}) &= \exp(-\mathcal{H}(S^{-1}(\mathbf{q}, \mathbf{v})) + \mathcal{H}(\mathbf{q}, \mathbf{v})) |\det \nabla S^{-1}(\mathbf{q}, \mathbf{v})| \\ &= \exp(-\mathcal{H}(S(\mathbf{q}, \mathbf{v})) + \mathcal{H}(\mathbf{q}, \mathbf{v})) |\det \nabla R(\hat{S}(\mathbf{q}, \mathbf{v}))| |\det \nabla \hat{S}(\mathbf{q}, \mathbf{v})|. \end{aligned}$$

Since R is a linear involution then clearly $|\det \nabla R(\mathbf{z})| = 1$ for any $\mathbf{z} \in \mathbb{R}^{2N}$, so that (4.14) follows from (2.11) and (4.19). Now (4.15) follows immediately from Lemma 4.2(ii), and the remaining identities (4.17) and (4.18) are clear, completing the proof. $\square$

We summarize the algorithm resulting from Proposition 4.3 in Algorithm 4.1.

4.1.3. *Algebraic considerations for numerical splittings.* With the above formulations in place we next present some approaches to constructing suitable numerical integrators $\hat{S}$ from the general system (4.9) that are commensurate with the setting of Proposition 4.3. Here the reversibility condition (4.10) leading to (4.12) is indispensable. Our starting point for constructing such reversible integrators out of a well chosen discretization of (4.9) are dictated by the following basic algebraic observations which will be useful here and again further on in Section 4.2.

LEMMA 4.4. *Let $\mathcal{R} : \mathcal{X} \rightarrow \mathcal{X}$ be a mapping on a set $\mathcal{X}$.*

(i) Suppose that $\mathcal{T}_j, \bar{\mathcal{T}}_j : \mathcal{X} \rightarrow \mathcal{X}$ are such that $\mathcal{R} \circ \mathcal{T}_j = \bar{\mathcal{T}}_j \circ \mathcal{R}$ for $j = 1, \dots, n$. Then, taking $\mathcal{T} := \mathcal{T}_1 \circ \dots \circ \mathcal{T}_n$ and $\bar{\mathcal{T}} := \bar{\mathcal{T}}_1 \circ \dots \circ \bar{\mathcal{T}}_n$, we have that $\mathcal{R} \circ \mathcal{T} = \bar{\mathcal{T}} \circ \mathcal{R}$.

(ii) If $\mathcal{S}_j : \mathcal{X} \rightarrow \mathcal{X}$, $j = 1, \dots, n$, are invertible mappings satisfying $\mathcal{R} \circ \mathcal{S}_j = \mathcal{S}_j^{-1} \circ \mathcal{R}$, then the “palindromic” composition $\mathcal{S}$ defined as

$$(4.20) \quad \mathcal{S} := \mathcal{S}_1 \circ \mathcal{S}_2 \circ \dots \circ \mathcal{S}_{n-1} \circ \mathcal{S}_n \circ \mathcal{S}_{n-1} \circ \dots \circ \mathcal{S}_2 \circ \mathcal{S}_1$$

satisfies $\mathcal{R} \circ \mathcal{S} = \mathcal{S}^{-1} \circ \mathcal{R}$.

PROOF. The first item, (i), is obvious. For (ii) notice that, due to the palindromic structure of $\mathcal{S}$, its inverse $\mathcal{S}^{-1}$ maintains the same palindromic structure, with

$$\mathcal{S}_1^{-1} \circ \mathcal{S}_2^{-1} \circ \dots \circ \mathcal{S}_{n-1}^{-1} \circ \mathcal{S}_n^{-1} \circ \mathcal{S}_{n-1}^{-1} \circ \dots \circ \mathcal{S}_2^{-1} \circ \mathcal{S}_1^{-1} = \mathcal{S}^{-1},$$

so that $\mathcal{R} \circ \mathcal{S} = \mathcal{S}^{-1} \circ \mathcal{R}$ follows directly from item (i). $\square$

Next let us introduce some definitions.

DEFINITION 4.5. Fix $\delta_0 > 0$ and suppose that for each $\delta \in (-\delta_0, \delta_0)$ we have an invertible map $\hat{\mathcal{S}}_\delta : \mathbb{R}^{2N} \rightarrow \mathbb{R}^{2N}$.

- (i) We define the *adjoint* of $\hat{\mathcal{S}}_\delta$, denoted as $\hat{\mathcal{S}}_\delta^*$, according to $\hat{\mathcal{S}}_\delta^* := (\hat{\mathcal{S}}_{-\delta})^{-1}$.
- (ii) We say $\hat{\mathcal{S}}_\delta$ is *symmetric* (or self-adjoint) if $\hat{\mathcal{S}}_\delta^* = \hat{\mathcal{S}}_\delta$.

The following desirable properties around this adjoint operation are immediate.

LEMMA 4.6. (i) Given maps collections of invertible maps $\hat{\mathcal{S}}, \hat{\mathcal{T}}$ as in Definition 4.5 we have $\hat{\mathcal{S}}^{**} = \hat{\mathcal{S}}$ and also $(\hat{\mathcal{S}}\hat{\mathcal{T}})^* = \hat{\mathcal{T}}^*\hat{\mathcal{S}}^*$. In particular $\hat{\mathcal{S}}^*\hat{\mathcal{S}}$ is symmetric.

(ii) If $\hat{\mathcal{S}}_\delta$ is symmetric for some $\delta \in \mathbb{R}$ then $\hat{\mathcal{S}}_\delta^{-1} = \hat{\mathcal{S}}_{-\delta}$.

We now have everything in hand to implement algorithms around the dynamics (4.9) discretized in a suitable form to apply Proposition 4.3 in three specific cases.

4.1.4. *Case 1: Separable surrogate dynamics.* One particular case of (4.9) we consider is when f_1 depends only on $\mathbf{v}$ and f_2 depends only on $\mathbf{q}$. Namely, in this case, (4.9) reduces to

$$(4.21) \quad \frac{d\mathbf{q}}{dt} = f_1(\mathbf{v}), \quad \frac{d\mathbf{v}}{dt} = f_2(\mathbf{q}), \quad (\mathbf{q}(0), \mathbf{v}(0)) = (\mathbf{q}_0, \mathbf{v}_0),$$

for some C^1 functions $f_1, f_2 : \mathbb{R}^N \rightarrow \mathbb{R}^N$. Here we have in mind the situation where our Hamiltonian $\mathcal{H}$ and matrix J in (4.6) have the separable form

$$(4.22) \quad \mathcal{H}(\mathbf{q}, \mathbf{v}) := \mathcal{U}(\mathbf{q}) + \mathcal{K}(\mathbf{v}) \quad \text{and} \quad J := \begin{pmatrix} 0 & -A \\ A^* & 0 \end{pmatrix},$$

for some invertible matrix $A \in \mathbb{R}^{N \times N}$, so that f_1 and f_2 would serve as suitable approximations of the form

$$(4.23) \quad f_1(\mathbf{v}) \approx (A^*)^{-1} \nabla \mathcal{K}(\mathbf{v}), \quad f_2(\mathbf{q}) \approx -A^{-1} \nabla \mathcal{U}(\mathbf{q}).$$

In this case, (4.21) we may consider the classical *leapfrog* integrator. This scheme is defined by splitting the dynamics (4.21) into

$$(4.24) \quad \frac{d\mathbf{q}}{dt} = 0, \quad \frac{d\mathbf{v}}{dt} = f_2(\mathbf{q}) \quad \text{and} \quad \frac{d\mathbf{q}}{dt} = f_1(\mathbf{v}), \quad \frac{d\mathbf{v}}{dt} = 0,$$

with analytical solutions given explicitly for any time t and initial datum $(\mathbf{q}_0, \mathbf{v}_0) \in \mathbb{R}^{2N}$ as

$$(4.25) \quad \Xi_t^{(1)}(\mathbf{q}_0, \mathbf{v}_0) = (\mathbf{q}_0, \mathbf{v}_0 + t f_2(\mathbf{q}_0)) \quad \text{and} \quad \Xi_t^{(2)}(\mathbf{q}_0, \mathbf{v}_0) = (\mathbf{q}_0 + t f_1(\mathbf{v}_0), \mathbf{v}_0),$$

respectively. Given time steps $\delta_1, \delta_2 > 0$ and a number of iterations $n \in \mathbb{N}$, a leapfrog-type integrator for (4.21) is then defined according to the following Strang splitting:

$$(4.26) \quad \hat{\mathcal{S}}(\mathbf{q}_0, \mathbf{v}_0) = \hat{\mathcal{S}}_{n, \delta_1, \delta_2}(\mathbf{q}_0, \mathbf{v}_0) := (\Xi_{\delta_1}^{(1)} \circ \Xi_{\delta_2}^{(2)} \circ \Xi_{\delta_1}^{(1)})^n(\mathbf{q}_0, \mathbf{v}_0) \\ \text{for all } (\mathbf{q}_0, \mathbf{v}_0) \in \mathbb{R}^{2N}.$$

Under minimal condition on f_1, f_2 we have the following result placing the map (4.26) defined from (4.21) in the setting of Proposition 4.3.

THEOREM 4.7. *Consider the dynamics in (4.24), resolved as (4.25), for any given C^1 functions $f_1, f_2 : \mathbb{R}^N \rightarrow \mathbb{R}^N$. For any time steps $\delta_1, \delta_2 > 0$ and any number of iterations $n \in \mathbb{N}$, we take $\hat{\mathcal{S}}_{n, \delta_1, \delta_2}$ to be the integrator defined in (4.26). Then:*

- (i) $\hat{\mathcal{S}}_{n, \delta_1, \delta_2}$ is volume-preserving, that is, $|\det \nabla \hat{\mathcal{S}}_{n, \delta_1, \delta_2}(\mathbf{z})| = 1$ for all $\mathbf{z} \in \mathbb{R}^{2N}$.
- (ii) If we furthermore assume that

$$(4.27) \quad f_1(-\mathbf{v}) = -f_1(\mathbf{v}) \quad \text{for every } \mathbf{v} \in \mathbb{R}^N$$

then $\hat{\mathcal{S}}_{n, \delta_1, \delta_2}$ is reversible with respect to the momentum flip involution R in the sense of Definition 4.1, (4.10). Here R is given by

$$(4.28) \quad R(\mathbf{q}, \mathbf{v}) = (\mathbf{q}, -\mathbf{v}), \quad (\mathbf{q}, \mathbf{v}) \in \mathbb{R}^{2N}.$$

PROOF. Start with the first item, (i). From (4.25) it follows that

$$\nabla \Xi_t^{(1)}(\mathbf{q}, \mathbf{v}) = \begin{pmatrix} I & 0 \\ t \nabla f_2(\mathbf{q}) & I \end{pmatrix} \quad \text{and} \quad \nabla \Xi_t^{(2)}(\mathbf{q}, \mathbf{v}) = \begin{pmatrix} I & t \nabla f_1(\mathbf{v}) \\ 0 & I \end{pmatrix},$$

so that, clearly, $|\det \nabla \Xi_t^{(1)}(\mathbf{z})| = |\det \nabla \Xi_t^{(2)}(\mathbf{z})| = 1$ for every $\mathbf{z} \in \mathbb{R}^{2N}$ and each $t \geq 0$. From (4.26) it thus follows that $|\det \nabla \hat{\mathcal{S}}_{n, \delta_1, \delta_2}(\mathbf{z})| = 1$ for all $\mathbf{z} \in \mathbb{R}^{2N}$. Regarding the reversibility claim in item (ii) we observe that $\Xi^{(j)}$ is symmetric in the sense of Definition 4.5(ii) for $j = 1, 2$. On the other hand it is also direct to check that $R \Xi_t^{(j)} = \Xi_{-t}^{(j)} R$ for any $t \geq 0$ and $j = 1, 2$. Here note the use of the condition (4.27) for $\Xi^{(2)}$. Thus with Lemma 4.6(ii) and Lemma 4.4(ii), noting the palindromic structure in (4.26), we obtain the desired reversibility claim for $\hat{\mathcal{S}}_{n, \delta_1, \delta_2}$, completing the proof. $\square$

We summarize and present the algorithm directly resulting from Theorem 4.7 and Proposition 4.3(ii). As we make explicit in Section 5.1 below this algorithm includes the classical random walk Monte Carlo (RWMC), Metropolis adjusted Langevin (MALA), and Hamiltonian Monte Carlo (HMC) algorithms as special cases under suitable parameter choices.

REMARK 4.8. Algorithm 4.2 derived from (4.21) can be further generalized in several directions. For example, (4.26) can be replaced with any palindromic splitting involving the maps $\Xi^{(1)}, \Xi^{(2)}$ defined in (4.25). The reversibility condition (4.27) as suitable for the momentum flip operation R in (4.28) can be replaced with a more general condition (4.41) as we explore in Theorem 4.12 below.

Algorithm 4.2 Generalized leapfrog for surrogate dynamics to sample $\mu(d\mathbf{q}) = Z_{\mathcal{U}}^{-1} e^{-\mathcal{U}(\mathbf{q})} d\mathbf{q}$

1: Select the algorithm parameters:

- (i) The proposal distribution $\mathcal{V}(\mathbf{q}, d\mathbf{v}) = Z_{\mathcal{K}}(\mathbf{q})^{-1} e^{-\mathcal{K}(\mathbf{q}, \mathbf{v})} d\mathbf{v} \in \text{Pr}(\mathbb{R}^N)$ for each $\mathbf{q} \in \mathbb{R}^N$.
- (ii) The surrogate functions $f_1, f_2 : \mathbb{R}^N \rightarrow \mathbb{R}^N$ defining $\Xi_t^{(1)}(\mathbf{q}, \mathbf{v}) := (\mathbf{q}, \mathbf{v} + t f_2(\mathbf{q}))$, and $\Xi_t^{(2)}(\mathbf{q}, \mathbf{v}) := (\mathbf{q} + t f_1(\mathbf{v}), \mathbf{v})$, where f_1 satisfies $f_1(-\mathbf{v}) = -f_1(\mathbf{v})$ for all $\mathbf{v} \in \mathbb{R}^N$.
- (iii) The time step sizes $\delta_1, \delta_2 > 0$.
- (iv) The number of iterations n .

2: Choose $\mathbf{q}_0 \in \mathbb{R}^N$.

3: **for** $k \geq 0$ **do**

4: Sample $\mathbf{v}_k \sim \mathcal{V}(\mathbf{q}_k, d\mathbf{v})$.

5: Propose $(\bar{\mathbf{q}}_{k+1}, \bar{\mathbf{v}}_{k+1}) := (\Xi_{\delta_1}^{(1)} \circ \Xi_{\delta_2}^{(2)} \circ \Xi_{\delta_1}^{(1)})^n(\mathbf{q}_k, \mathbf{v}_k)$.

6: Set $\mathbf{q}_{k+1} := \bar{\mathbf{q}}_{k+1}$ with probability $1 \wedge [\exp(-\mathcal{H}(\bar{\mathbf{q}}_{k+1}, -\bar{\mathbf{v}}_{k+1}) + \mathcal{H}(\mathbf{q}_k, \mathbf{v}_k))]$, otherwise take $\mathbf{q}_{k+1} := \mathbf{q}_k$.

4.1.5. *Case 2: Splitting into Hamiltonian sub-dynamics.* A second case of interest arises when we consider a surrogate of the form

$$(4.29) \quad J^{-1} \nabla \mathcal{H}(\mathbf{z}) \approx \sum_{j=1}^m J_j^{-1} \nabla \mathcal{H}_j(\mathbf{z})$$

for (4.6) where each J_j is antisymmetric and invertible and each $\mathcal{H}_j : \mathbb{R}^N \rightarrow \mathbb{R}$ is suitably smooth. Corresponding to each element in this sum, (4.29), we consider mappings $\Xi^{(j)}$ where

$$(4.30) \quad \{\Xi_t^{(j)}\}_{t \geq 0} \text{ is the solution map for } \frac{d\mathbf{z}}{dt} = J_j^{-1} \nabla \mathcal{H}_j(\mathbf{z}),$$

defined so long as each of the associated Hamiltonian systems admits a globally defined dynamic. The idea here is that we might formulate the approximation in (4.29) so that each $\{\Xi_t^{(j)}\}_{t \geq 0}$ has an explicitly solvable form while in any case preserving crucial structural properties of Hamiltonian systems commensurate with the setting of Proposition 4.3. As such, (4.29) would suggest that a suitable composition of the maps $\Xi^{(j)}$ would yield a reasonable approximation for (4.6). Keeping in mind Lemma 4.4, we select for some $l \in \mathbb{N}$, $n \in \mathbb{N}$, any $j_1, \dots, j_l \in \{1, \dots, m\}$, any $\delta_1, \dots, \delta_l > 0$, and define

$$(4.31) \quad \hat{\mathcal{S}}(\mathbf{q}_0, \mathbf{v}_0) := (\Xi_{\delta_1}^{(j_1)} \circ \dots \circ \Xi_{\delta_l}^{(j_l)} \circ \Xi_{\delta_l}^{(j_l)} \circ \dots \circ \Xi_{\delta_1}^{(j_1)})^n(\mathbf{q}_0, \mathbf{v}_0)$$

for all $(\mathbf{q}_0, \mathbf{v}_0) \in \mathbb{R}^{2N}$.

Before providing conditions under which (4.31) yields a suitable class of sampling algorithms we first recall some basic properties of Hamiltonian dynamical systems that will be needed.

PROPOSITION 4.9. *Suppose that $\{\Xi_t\}_{t \geq 0}$ is the solution operator of*

$$(4.32) \quad \frac{d\mathbf{z}}{dt} = J^{-1} \nabla \tilde{\mathcal{H}}(\mathbf{z})$$

for a matrix J which is antisymmetric and invertible and a C^2 Hamiltonian function $\tilde{\mathcal{H}} : \mathbb{R}^{2N} \rightarrow \mathbb{R}$ such that the dynamics (4.32) are uniquely and globally defined. Then,

- (i) for every $t \geq 0$, Ξ_t is symplectic (with respect to J) as in Definition 4.1(ii).
- (ii) Suppose we have a linear involution $R : \mathbb{R}^{2N} \rightarrow \mathbb{R}^{2N}$ such that

(4.33)
$$RJ^{-1}R^* = -J^{-1}$$

and

(4.34)
$$\tilde{\mathcal{H}}(R\mathbf{z}) = \tilde{\mathcal{H}}(\mathbf{z}) \quad \text{for every } \mathbf{z} \in \mathbb{R}^{2N}.$$

Then, for every $t \geq 0$, Ξ_t is reversible with respect to R in the sense of Definition 4.1(i).

- (iii) In particular, consider

$$J = \begin{pmatrix} 0 & -A \\ A^* & 0 \end{pmatrix},$$

for an invertible A and we assume that $\tilde{\mathcal{H}}$ is symmetric in its second variable, namely $\tilde{\mathcal{H}}(\mathbf{q}, \mathbf{v}) = \tilde{\mathcal{H}}(\mathbf{q}, -\mathbf{v})$, for any $(\mathbf{q}, \mathbf{v}) \in \mathbb{R}^{2N}$. Then, for every $t \geq 0$, Ξ_t is reversible with respect to the momentum flip involution R as in (4.28).

PROOF. For the first item let $B(t)(\mathbf{z}) := (\nabla \Xi_t(\mathbf{z}))^* J \nabla \Xi_t(\mathbf{z})$ and notice that $B(0)(\mathbf{z}) = J\mathbf{z}$ while $dB(\mathbf{z})/dt = 0$.

Regarding the second item, given any solution $\mathbf{z}(t)$ of (4.32) we consider $\tilde{\mathbf{z}}(t) := R\mathbf{z}(-t)$. Observe that, with the assumption that R is an involution,

$$\frac{d\tilde{\mathbf{z}}(t)}{dt} = -RJ^{-1}\nabla\tilde{\mathcal{H}}(\mathbf{z}(-t)) = -RJ^{-1}\nabla\tilde{\mathcal{H}}(R\tilde{\mathbf{z}}(t)).$$

Now, from (4.34), we have

$$R^*\nabla\tilde{\mathcal{H}}(R\bar{\mathbf{z}}) = \nabla\tilde{\mathcal{H}}(\bar{\mathbf{z}}) \quad \text{for every } \bar{\mathbf{z}} \in \mathbb{R}^{2N}$$

and so with the fact that R^* is an involution and our assumption (4.33) we conclude that $\tilde{\mathbf{z}}(t)$ must also obey (4.32). From this symmetry observation and the uniqueness of solutions of (4.32), we infer

$$\Xi_t(R\mathbf{z}_0) = R\Xi_{-t}(\mathbf{z}_0) \quad \text{for any } t \in \mathbb{R}, \text{ and any } \mathbf{z}_0 \in \mathbb{R}^{2N}.$$

Thus

$$(R \circ \Xi_t \circ R \circ \Xi_t)(\mathbf{z}_0) = R\Xi_t(R\Xi_t(\mathbf{z}_0)) = \Xi_{-t}(\Xi_t(\mathbf{z}_0)) = \mathbf{z}_0$$

as desired for the second item. The third item follows from the second with a direct computation showing that (4.33) holds for these specific choices of J and R . The proof is complete. $\square$

REMARK 4.10. See [73], Lemma 2, for variations on the theme of Proposition 4.9(ii) with some interesting applications in deriving further HMC-type sampling algorithms.

THEOREM 4.11. Fix a collection of C^2 functions $\mathcal{H}_j : \mathbb{R}^{2N} \rightarrow \mathbb{R}$ along with corresponding antisymmetric and invertible $2N \times 2N$ real matrices J_j for $j = 1, \dots, m$. Assume the Hamiltonian dynamics associated to each of these pairs, à la (4.32), are uniquely and globally defined. Let $\{\Xi_t^{(j)}\}_{t \in \mathbb{R}}$ be the associated solution maps as in (4.30). Furthermore suppose that there exists a linear involution R such that

(4.35)
$$RJ_j^{-1}R^* = -J_j^{-1} \quad \text{and} \quad \mathcal{H}_j \circ R = \mathcal{H}_j \quad \text{for each } j = 1, \dots, m.$$

Select any $l \in \mathbb{N}$, $n \in \mathbb{N}$, along with orderings $j_1, \dots, j_l \in \{1, \dots, m\}$, and time step sizes $\delta_1, \dots, \delta_l > 0$, and define $\hat{S}$ as in (4.31). Then, according to this definition (4.31), $\hat{S}$ is both symplectic and reversible with respect to R in the sense given in Definition 4.1.

Algorithm 4.3 Palindromic iterations of Hamiltonian surrogates to sample $\mu(d\mathbf{q}) = Z_{\mathcal{U}}^{-1} e^{-\mathcal{U}(\mathbf{q})} d\mathbf{q}$

1: Select the algorithm parameters:

- (i) The proposal distribution $\mathcal{V}(\mathbf{q}, d\mathbf{v}) = Z_{\mathcal{K}}(\mathbf{q})^{-1} e^{-\mathcal{K}(\mathbf{q}, \mathbf{v})} d\mathbf{v} \in \text{Pr}(\mathbb{R}^N)$ for each $\mathbf{q} \in \mathbb{R}^N$.
- (ii) For some $j = 1, \dots, m$ determine a set of surrogate $\mathcal{H}_j : \mathbb{R}^{2N} \rightarrow \mathbb{R}$ and associated symplectic matrices J_j yielding solution maps $\{\Xi^{(j)}\}_{t \in \mathbb{R}}$ defined according to $d\mathbf{z}/dt = J_j^{-1} \nabla \mathcal{H}_j(\mathbf{z})$.
- (iii) Identify a linear involution R such that $RJ_j^{-1}R^* = -J_j^{-1}$ and $\mathcal{H}_j \circ R = \mathcal{H}_j$ for $j = 1, \dots, m$.
- (iv) An operation ordering $j_1, \dots, j_l \in \{1, \dots, m\}$ along with associated time step sizes $\delta_1, \delta_2, \dots, \delta_l > 0$.
- (v) The number of iterations n .

2: Choose $\mathbf{q}_0 \in \mathbb{R}^N$.

3: **for** $k \geq 0$ **do**

4: Sample $\mathbf{v}_k \sim \mathcal{V}(\mathbf{q}_k, d\mathbf{v})$.

5: Propose $(\bar{\mathbf{q}}_{k+1}, \bar{\mathbf{v}}_{k+1}) := (\Xi_{\delta_1}^{(j_1)} \circ \dots \circ \Xi_{\delta_l}^{(j_l)} \circ \Xi_{\delta_l}^{(j_l)} \circ \dots \circ \Xi_{\delta_1}^{(j_1)})^n(\mathbf{q}_k, \mathbf{v}_k)$.

6: Set $\mathbf{q}_{k+1} := \bar{\mathbf{q}}_{k+1}$ with probability $1 \wedge [\exp(-\mathcal{H}(R(\bar{\mathbf{q}}_{k+1}), \bar{\mathbf{v}}_{k+1})) + \mathcal{H}(\mathbf{q}_k, \mathbf{v}_k)]$ and otherwise take $\mathbf{q}_{k+1} := \mathbf{q}_k$.

PROOF. According to Proposition 4.9 each $\Xi^{(j_i)}$ is symplectic and is reversible with respect to R for $i = 1, \dots, m$. Keeping in mind the palindromic structure of $\hat{\mathcal{S}}$ in regards to reversibility, it is therefore clear that these two properties extend to $\hat{\mathcal{S}}$. The proof is complete. $\square$

We now summarize our second class of sampling methods in Algorithm 4.3. This algorithm is derived from Theorem 4.11 with Proposition 4.3. As previously with Algorithm 4.2, this class includes classical formulations of RWMC, MALA and HMC as notable special cases in a fashion which we make precise in Section 5.1 below. Note also that a special case of (4.31) yields the type of “nonstandard” splittings which proves to be desirable for the setting of (4.51) in Section 4.2, Section 5.2.3 below.

4.1.6. *Case 3: Nonseparable dynamics via implicit integrators.* We turn to our final case where (4.9) is assumed to have a “nonseparable” form unsuitable for either of the formulations previously considered in Section 4.1.4 or Section 4.1.5. This situation arises, for example, from the consideration of position dependent kinetic energy terms in (4.3) as developed previously in [33]. Here, in this position dependent case, we have

$$(4.36) \quad f_1(\mathbf{q}, \mathbf{v}) \approx (A^*)^{-1} \nabla_{\mathbf{v}} \mathcal{K}(\mathbf{q}, \mathbf{v}),$$

$$(4.37) \quad f_2(\mathbf{q}, \mathbf{v}) \approx -A^{-1} (\nabla_{\mathbf{q}} (\mathcal{U}(\mathbf{q}) + \mathcal{K}(\mathbf{q}, \mathbf{v})) + Z_{\mathcal{K}}(\mathbf{q})^{-1} \nabla_{\mathbf{q}} Z_{\mathcal{K}}(\mathbf{q})),$$

when, as above, (4.6) is defined with a matrix J of the separated form given in (4.22). In this nonseparable case, implicit numerical discretizations for (4.9) provides a means of maintaining indispensable structural properties, namely the reversibility and volume preservation conditions, required by Proposition 4.3. Following ideas from [33, 37, 48], we consider schemes starting from the so-called *Euler-B* and *Euler-A* methods applied to (4.9).

Start with the *Euler-B* scheme which is defined, for a fixed time step $\delta > 0$, number of iterations $n \in \mathbb{N}$ and any initial point $(\mathbf{q}_0, \mathbf{v}_0) \in \mathbb{R}^{2N}$, by $(\mathbf{q}_n, \mathbf{v}_n) := (\hat{S}_{\delta}^{(B)})^n(\mathbf{q}_0, \mathbf{v}_0)$. Here each

iteration step $(\mathbf{q}, \mathbf{v}) := \hat{S}_\delta^{(B)}(\bar{\mathbf{q}}, \bar{\mathbf{v}})$ is specified, for a given $\bar{\mathbf{q}}, \bar{\mathbf{v}} \in \mathbb{R}^N$ through the following implicit system of equations:

$$(4.38) \quad \mathbf{q} = \bar{\mathbf{q}} + \delta f_1(\bar{\mathbf{q}}, \mathbf{v}), \quad \mathbf{v} = \bar{\mathbf{v}} + \delta f_2(\bar{\mathbf{q}}, \mathbf{v}).$$

It is not difficult to check that the adjoint of $\hat{S}_\delta^{(B)}$, cf. Definition 4.5, is the *Euler-A method* which is defined implicitly for a single step as $(\mathbf{q}, \mathbf{v}) := \hat{S}_\delta^{(A)}(\bar{\mathbf{q}}, \bar{\mathbf{v}})$ where

$$(4.39) \quad \mathbf{q} = \bar{\mathbf{q}} + \delta f_1(\mathbf{q}, \bar{\mathbf{v}}), \quad \mathbf{v} = \bar{\mathbf{v}} + \delta f_2(\mathbf{q}, \bar{\mathbf{v}}),$$

for given $\bar{\mathbf{q}}, \bar{\mathbf{v}} \in \mathbb{R}^N$. Balancing these two schemes, (4.38) and (4.39), leads to the consideration of a symmetric integrator called the *generalized Störmer–Verlet method*. Relative to the parameters $\delta > 0$ and $n \geq 1$ we define $\hat{S} = \hat{S}_{n,\delta} := (\hat{S}_{\delta/2}^{(A)} \circ \hat{S}_{\delta/2}^{(B)})^n$ so that, for any given $(\mathbf{q}_0, \mathbf{v}_0) \in \mathbb{R}^{2N}$, $(\mathbf{q}_n, \mathbf{v}_n) := \hat{S}_{n,\delta}(\mathbf{q}_0, \mathbf{v}_0)$ is computed inductively according to

$$(4.40) \quad \begin{aligned} \mathbf{v}_{m+1/2} &= \mathbf{v}_m + \frac{\delta}{2} f_2(\mathbf{q}_m, \mathbf{v}_{m+1/2}), \\ \mathbf{q}_{m+1} &= \mathbf{q}_m + \frac{\delta}{2} [f_1(\mathbf{q}_m, \mathbf{v}_{m+1/2}) + f_1(\mathbf{q}_{m+1}, \mathbf{v}_{m+1/2})], \\ \mathbf{v}_{m+1} &= \mathbf{v}_{m+1/2} + \frac{\delta}{2} f_2(\mathbf{q}_{m+1}, \mathbf{v}_{m+1/2}) \end{aligned}$$

for $m = 0, \dots, n-1$. Note carefully that (4.40) reduces to (4.26) in the special case when f_2 depends only on $\mathbf{q}$, f_1 only on $\mathbf{v}$, and we take $\delta_1 = \delta/2$, $\delta_2 = \delta$. See also Remark 4.16 below for further commentary around this point.

We next establish conditions on f_1 and f_2 which yield desirable reversibility and volume-preservation properties for the scheme (4.40). In the statement below, $\Pi_1, \Pi_2 : \mathbb{R}^{2N} \rightarrow \mathbb{R}^{2N}$ denote the projections onto the first and second components, respectively, that is,

$$\Pi_1(\mathbf{q}, \mathbf{v}) = \mathbf{q}, \quad \Pi_2(\mathbf{q}, \mathbf{v}) = \mathbf{v} \quad \text{for all } (\mathbf{q}, \mathbf{v}) \in \mathbb{R}^{2N}.$$

THEOREM 4.12. *Suppose that $f_1, f_2 : \mathbb{R}^{2N} \rightarrow \mathbb{R}^N$ are C^1 functions and assume that, for some $\delta > 0$, the maps $\hat{S}_{\delta/2}^{(B)}, \hat{S}_{\delta/2}^{(A)}$ specified implicitly from (4.38), (4.39) are uniquely and globally defined and are C^1 ; cf. Remark 4.13 below. For any $n \geq 1$ we consider the generalized Störmer–Verlet implicit integration scheme $\hat{S} = \hat{S}_{n,\delta} := (\hat{S}_{\delta/2}^{(A)} \circ \hat{S}_{\delta/2}^{(B)})^n$ as given by (4.40).*

(i) *Suppose that, for some linear invertible matrix R , the mapping $\mathbf{f} = (f_1, f_2)$ satisfies*

$$(4.41) \quad R\mathbf{f}(\Pi_1 \mathbf{z}, \Pi_2 \tilde{\mathbf{z}}) = -\mathbf{f}(\Pi_1 \circ R\mathbf{z}, \Pi_2 \circ R\tilde{\mathbf{z}}) \quad \text{for all } \mathbf{z}, \tilde{\mathbf{z}} \in \mathbb{R}^{2N}.$$

Then $\hat{S}$ is reversible with respect to R , in the sense of Definition 4.1(i). In particular, if f_1 and f_2 maintain

$$(4.42) \quad f_1(\mathbf{q}, \mathbf{v}) = -f_1(\mathbf{q}, -\mathbf{v}), \quad f_2(\mathbf{q}, \mathbf{v}) = f_2(\mathbf{q}, -\mathbf{v}) \quad \text{for all } (\mathbf{q}, \mathbf{v}) \in \mathbb{R}^{2N},$$

then $\hat{S}$ is reversible with respect to the “momentum-flip involution operation” (4.28).

(ii) *Now suppose that $f_1, f_2 : \mathbb{R}^{2N} \rightarrow \mathbb{R}^N$ are such that*

$$(4.43) \quad \begin{aligned} \nabla_{\mathbf{q}} f_1(\mathbf{q}, \mathbf{v}) + (\nabla_{\mathbf{v}} f_2)^*(\mathbf{q}, \mathbf{v}) &= 0 \quad \text{and} \\ \nabla_{\mathbf{v}} f_1(\mathbf{q}, \mathbf{v}), \nabla_{\mathbf{q}} f_2(\mathbf{q}, \mathbf{v}) &\text{ are symmetric matrices,} \end{aligned}$$

for all $(\mathbf{q}, \mathbf{v}) \in \mathbb{R}^{2N}$. Then $\hat{S}$ is symplectic relative to the canonical form J as in (4.7) in the sense of Definition 4.1(ii).

(iii) Finally consider the case where $\mathbf{f}(\mathbf{z}) = \tilde{J}^{-1} \nabla \tilde{\mathcal{H}}(\mathbf{z})$, for each $\mathbf{z} \in \mathbb{R}^{2N}$ where $\tilde{\mathcal{H}} \in C^2$ and where $\tilde{J}$ is of the form

$$(4.44) \quad \tilde{J} = \begin{pmatrix} 0 & -E^{-1} \\ E^{-1} & 0 \end{pmatrix}$$

for some invertible matrix $E \in \mathbb{R}^{N \times N}$. Then, in this circumstance, $\hat{S}$ is volume-preserving, that is, $|\det \nabla \hat{S}(\mathbf{z})| = 1$ for all $\mathbf{z} \in \mathbb{R}^{2N}$.

The proof of Theorem 4.12 is presented in Appendix A.

Let us make several further technical remarks concerning Theorem 4.12.

REMARK 4.13. It is natural to ask if there are reasonable conditions which may be placed on $\mathbf{f} := (f_1, f_2)$ and on $\delta > 0$ in (4.40) guaranteeing a unique solution to this implicitly defined scheme. Here, suppose, for example, that $\mathbf{f}$ satisfies a global Lipschitz condition, namely

$$(4.45) \quad |\mathbf{f}(\tilde{\mathbf{z}}) - \mathbf{f}(\mathbf{z})| \leq K |\tilde{\mathbf{z}} - \mathbf{z}| \quad \text{for any } \tilde{\mathbf{z}}, \mathbf{z} \in \mathbb{R}^N,$$

where $K > 0$ is an absolute constant independent of $\tilde{\mathbf{z}}, \mathbf{z}$. Then, under (4.45), for any $\delta < K^{-1}$ the scheme (4.40) is uniquely defined for any $n \geq 1$, as may be readily demonstrated using the Banach fixed point theorem. Indeed, noting that (4.40) is a composition of the Euler A and B subschemes, (4.38) and (4.39), we may apply this fixed point argument to each of these substeps in turn. Regarding the Euler B step, (4.38), consider, for any $\mathbf{q}, \mathbf{v}, \tilde{\mathbf{q}}, \tilde{\mathbf{v}} \in \mathbb{R}^N$,

$$F_\delta(\mathbf{q}, \mathbf{v}, \tilde{\mathbf{q}}, \tilde{\mathbf{v}}) := (\tilde{\mathbf{q}} + \delta f_1(\mathbf{q}, \tilde{\mathbf{v}}), \tilde{\mathbf{v}} + \delta f_2(\mathbf{q}, \tilde{\mathbf{v}})).$$

Under (4.45) it is clear that, for any fixed $\tilde{\mathbf{q}}, \tilde{\mathbf{v}} \in \mathbb{R}^N$ the map $(\mathbf{q}, \mathbf{v}) \mapsto F_\delta(\mathbf{q}, \mathbf{v}, \tilde{\mathbf{q}}, \tilde{\mathbf{v}})$ is contractive on $\mathbb{R}^{2N}$ so long as $\delta < K^{-1}$ and hence, for any given $\tilde{\mathbf{q}}, \tilde{\mathbf{v}}$ and δ we obtain a unique $(\mathbf{q}^*, \mathbf{v}^*)$ such that $(\mathbf{q}^*, \mathbf{v}^*) = F_\delta(\mathbf{q}^*, \mathbf{v}^*, \tilde{\mathbf{q}}, \tilde{\mathbf{v}})$. In other words we have shown that (4.38) must have a unique solution. Of course precisely the same argument may be applied also to (4.39). Finally we note that, under (4.45) it is direct to show that $\hat{S}_\delta^{(A)}$ and $\hat{S}_\delta^{(B)}$ are C^1 functions.

REMARK 4.14. In the case where $\mathbf{f}(\mathbf{z}) = \tilde{J}^{-1} \nabla \tilde{\mathcal{H}}(\mathbf{z})$ the reversibility condition (4.41) in Theorem 4.12(i) can be characterized as follows. Given any linear invertible mapping $R : \mathbb{R}^{2N} \rightarrow \mathbb{R}^{2N}$, which we write in the block form

$$R = \begin{pmatrix} A & B \\ C & D \end{pmatrix}, \quad A, B, C, D \in \mathbb{R}^{N \times N},$$

we suppose that

$$(4.46) \quad \tilde{\mathcal{H}}(\Pi_1 \circ R\mathbf{z}, \Pi_2 \circ R\tilde{\mathbf{z}}) = \tilde{\mathcal{H}}(\Pi_1\mathbf{z}, \Pi_2\tilde{\mathbf{z}}) \quad \text{for all } \mathbf{z}, \tilde{\mathbf{z}} \in \mathbb{R}^{2N},$$

and that

$$(4.47) \quad R\tilde{J}^{-1}R_0^* = -\tilde{J}^{-1} \quad \text{where } R_0 := \begin{pmatrix} A & 0 \\ 0 & D \end{pmatrix}.$$

To see that these conditions yield (4.41) observe that using (4.46) we obtain after a direct calculation that $\nabla \tilde{\mathcal{H}}(\Pi_1\mathbf{z}, \Pi_2\tilde{\mathbf{z}}) = R_0^* \nabla \tilde{\mathcal{H}}(\Pi_1 \circ R\mathbf{z}, \Pi_2 \circ R\tilde{\mathbf{z}})$, for any $\mathbf{z}, \tilde{\mathbf{z}} \in \mathbb{R}^{2N}$. Thus

$$\begin{aligned} R\mathbf{f}(\Pi_1\mathbf{z}, \Pi_2\tilde{\mathbf{z}}) &= R\tilde{J}^{-1} \nabla \tilde{\mathcal{H}}(\Pi_1\mathbf{z}, \Pi_2\tilde{\mathbf{z}}) = R\tilde{J}^{-1} R_0^* \nabla \tilde{\mathcal{H}}(\Pi_1 \circ R\mathbf{z}, \Pi_2 \circ R\tilde{\mathbf{z}}) \\ &= -\tilde{J}^{-1} \nabla \tilde{\mathcal{H}}(\Pi_1 \circ R\mathbf{z}, \Pi_2 \circ R\tilde{\mathbf{z}}) = -\mathbf{f}(\Pi_1 \circ R\mathbf{z}, \Pi_2 \circ R\tilde{\mathbf{z}}), \end{aligned}$$

where the third equality follows from (4.47). Note that this condition (4.46), (4.47) is comparable to (4.33), (4.34) in Proposition 4.9.

Algorithm 4.4 Surrogates for nonseparable Hamiltonians to sample $\mu(d\mathbf{q}) = Z_{\mathcal{U}}^{-1} e^{-\mathcal{U}(\mathbf{q})} d\mathbf{q}$

1: Select the algorithm parameters:

- (i) The proposal distribution $\mathcal{V}(\mathbf{q}, d\mathbf{v}) = Z_{\mathcal{K}}(\mathbf{q})^{-1} e^{-\mathcal{K}(\mathbf{q}, \mathbf{v})} d\mathbf{v} \in \text{Pr}(\mathbb{R}^N)$ for each $\mathbf{q} \in \mathbb{R}^N$.
- (ii) Surrogate functions $f_1, f_2 : \mathbb{R}^{2N} \rightarrow \mathbb{R}^N$ specified so that either (4.43) holds or $(f_1(\mathbf{z}), f_2(\mathbf{z})) = J^{-1} \nabla \mathcal{H}(\mathbf{z})$ where J is as in (4.44).
- (iii) Relative to f_1, f_2 identify a linear involution R maintaining (4.41) (see in particular (4.42)).
- (iv) The time step $\delta > 0$ and number of iterations $n \geq 1$.

2: Choose $\mathbf{q}_0 \in \mathbb{R}^N$.

3: **for** $k \geq 0$ **do**

4: Sample $\mathbf{v}_k \sim \mathcal{V}(\mathbf{q}_k, d\mathbf{v})$.

5: Propose $(\bar{\mathbf{q}}_{k+1}, \bar{\mathbf{v}}_{k+1}) := \hat{S}_{n,\delta}(\mathbf{q}_k, \mathbf{v}_k)$ where $\hat{S}_{n,\delta} = (\hat{S}_{\delta/2}^{(A)} \circ \hat{S}_{\delta/2}^{(B)})^n$ is defined implicitly via (4.40).

6: Set $\mathbf{q}_{k+1} := \bar{\mathbf{q}}_{k+1}$ with probability $1 \wedge [\exp(-\mathcal{H}(R(\bar{\mathbf{q}}_{k+1}, \bar{\mathbf{v}}_{k+1})) + \mathcal{H}(\mathbf{q}_k, \mathbf{v}_k))]$ and otherwise take $\mathbf{q}_{k+1} := \mathbf{q}_k$.

REMARK 4.15. In [37], Theorem 3.3, it is shown that if $\tilde{J}$ is of the form in (4.44) with $E = I$, then $\hat{S}^{(A)}$ and $\hat{S}^{(B)}$ are in fact symplectic integrators with respect to $\tilde{J}$. Namely, $(\nabla \hat{S}_{\delta}^{(A)}(\mathbf{z}))^* \tilde{J} \nabla \hat{S}_{\delta}^{(A)}(\mathbf{z}) = \tilde{J}$, for all $\mathbf{z} \in \mathbb{R}^{2N}$, and analogously for $\hat{S}^{(B)}$. This can be verified directly from (A.4) by taking $E = I$ or by repeating the approach given in Theorem 4.12(ii). However, the same does not seem to hold for a general invertible $E \in \mathbb{R}^{N \times N}$.

Combining Theorem 4.12 with Proposition 4.3 we now formulate the final algorithm of this section as Algorithm 4.4.

REMARK 4.16. The methods described in Algorithm 4.2, Algorithm 4.3 and Algorithm 4.4 have quite a bit of overlapping scope and can be generalized in a number of immediate ways working from the frameworks developed here. Regarding the scope of the methods notice, for example, that Algorithm 4.2 is actually a special case of Algorithm 4.3 when $f_1(\mathbf{v}) = (A^*)^{-1} \nabla \tilde{\mathcal{K}}(\mathbf{v})$, $f_2(\mathbf{q}) = -A^{-1} \nabla \tilde{\mathcal{U}}(\mathbf{q})$, for any $\tilde{\mathcal{K}}, \tilde{\mathcal{U}} \in C^1$ and any invertible A , or a special case of Algorithm 4.4 when, more generally, $f_1 = f_1(\mathbf{v})$, $f_2 = f_2(\mathbf{q})$, with appropriate choices for the time steps. In fact, as already observed above, (4.40) reduces to (4.26) with $\delta_1 = \delta/2$ and $\delta_2 = \delta$ whenever f_1 only depends on $\mathbf{v}$ and f_2 only on $\mathbf{q}$.

Regarding further generalizations we observe that Algorithm 4.2 and Algorithm 4.4 are easily extended to any palindromic splitting of $\Xi^{(1)}, \Xi^{(2)}$ or of $\hat{S}^{(A)}, \hat{S}^{(B)}$ in the fashion of (4.31) in Algorithm 4.3. Note that the efficacy of such extended palindromic settings have been explored in more depth recently in [17]. In a different direction we note that in all of the above algorithms we have identified conditions guaranteeing that the integrator is symplectic or at least volume preserving. In principle, Algorithm 4.1 allows for schemes $\hat{S}$ where $|\det(\nabla \hat{S}(\mathbf{z}))| \neq 1$. However, at least within a naive set-up, the computational requirements around such gradient terms would appear to be lethal. Nevertheless, see the recent work [49].

REMARK 4.17. People who have applied HMC to complicated problems may have encountered scenarios where the implementation of the numerical integrator was incorrect, but the sampler still produced correct results. The structure of Algorithm 4.1, Algorithm 4.2, Algorithm 4.3 and Algorithm 4.4 (see also [61]) provides an explanation for this phenomenon— as long as the integrator and acceptance ratio are consistent with each other, the accept/reject

step will correct for any errors in the propagation in the dynamical system, whether they be due to discretization errors or bugs in the implementation.

4.2. The infinite-dimensional case. We turn now to introduce our second class of generalized trajectory methods. These methods, summarized in Algorithm 4.5 below, are adapted to sampling from measures which are defined on a separable Hilbert space and which are absolutely continuous with respect to a certain class of Gaussian base measures. As mentioned above, our approach here takes its inspiration from a preconditioned HMC method introduced in [10] and later revisited in [8, 9]. Actually, as we show below in Section 5.2, the algorithms introduced in [8, 10] along with the pCN and ∞ -MALA samplers from, for example, [11, 22] all fall out as special parameter choices in our class of samplers summarized in Algorithm 4.5.

Let us begin by making our infinite-dimensional setting precise. Throughout this section we assume that the parameter space X is a separable Hilbert space, with associated norm and inner product denoted by $|\cdot|$ and $\langle \cdot, \cdot \rangle$, respectively. We take $\mathcal{B}_X$ to signify the associated σ -algebra of Borel subsets of X . On X we consider target probability measures of the form

$$(4.48) \quad \mu(d\mathbf{q}) = \frac{1}{Z} e^{-\Phi(\mathbf{q})} \mu_0(d\mathbf{q}), \quad Z = \int_X e^{-\Phi(\mathbf{q})} \mu_0(d\mathbf{q}),$$

for some potential function $\Phi : X \rightarrow \mathbb{R}$ such that $e^{-\Phi(\mathbf{q})}$ is μ_0 -integrable, and with μ_0 given as a centered Gaussian measure on X with covariance operator $\mathcal{C} : X \rightarrow X$, that is, $\mu_0 = \mathcal{N}(0, \mathcal{C})$.

Regarding the covariance structure for the Gaussian measure μ_0 , we assume throughout what follows that $\mathcal{C}$ is a trace-class, symmetric and strictly positive definite operator. By the spectral theorem, it follows that X admits a complete orthonormal basis $\{\mathbf{e}_i\}_{i \in \mathbb{N}}$ consisting of eigenfunctions of $\mathcal{C}$, corresponding to a nonincreasing sequence of positive eigenvalues $\{\lambda_i\}_{i \in \mathbb{N}}$.⁶ We thus define the fractional powers $\mathcal{C}^\gamma$ of $\mathcal{C}$ for any $\gamma \in \mathbb{R}$ as

$$(4.49) \quad \mathcal{C}^\gamma \mathbf{q} = \sum_{i=1}^{\infty} \lambda_i^\gamma \langle \mathbf{q}, \mathbf{e}_i \rangle \mathbf{e}_i \quad \text{for all } \mathbf{q} \in \text{Dom}(\mathcal{C}^\gamma),$$

where $\text{Dom}(\mathcal{C}^\gamma)$ can be characterized as the set of all $\mathbf{q} \in X$ such that $\sum_{i=1}^{\infty} \lambda_i^{2\gamma} \langle \mathbf{q}, \mathbf{e}_i \rangle^2 < \infty$ when $\gamma \leq 0$ while taking $\text{Dom}(\mathcal{C}^\gamma)$ as the dual relative to X of $\text{Dom}(\mathcal{C}^{-\gamma})$ when $\gamma > 0$. We refer the reader to [13, 23] for further general background on the wider theory of Gaussian measures on function spaces.

The preconditioned Hamiltonian Monte Carlo algorithm introduced in [10] to sample from $\mu(d\mathbf{q}) = Z^{-1} e^{-\Phi(\mathbf{q})} \mu_0(d\mathbf{q})$ is based on the following Hamiltonian function:

$$(4.50) \quad \mathcal{H}(\mathbf{q}, \mathbf{v}) := \frac{1}{2} |\mathcal{C}^{-1/2} \mathbf{v}|^2 + \frac{1}{2} |\mathcal{C}^{-1/2} \mathbf{q}|^2 + \Phi(\mathbf{q}).$$

The potential function Φ is assumed to be Fréchet differentiable, with its Fréchet derivative denoted by $D\Phi : X \rightarrow X$, where the Hilbert space X is identified with its dual. It thus follows that (4.50) has an associated (noncanonical) Hamiltonian dynamic given by

$$(4.51) \quad \frac{d\mathbf{q}}{dt} = \mathbf{v}, \quad \frac{d\mathbf{v}}{dt} = -\mathbf{q} - \mathcal{C} D\Phi(\mathbf{q}),$$

which formally corresponds to the more compact formulation (4.6) with the following choice of “preconditioning” operator:

$$(4.52) \quad J := \begin{pmatrix} 0 & -\mathcal{C}^{-1} \\ \mathcal{C}^{-1} & 0 \end{pmatrix}.$$

⁶Here we note that the trace-class condition amounts to requiring that $\sum_{j=1}^{\infty} \lambda_j < \infty$.

Under suitable assumptions on the potential function Φ , one can show that the Hamiltonian flow associated to (4.51) holds $\mu \otimes \mu_0$ as an invariant measure, see [10], Theorem 3.1. As a consequence, the Markov transition kernel associated to the evolution of the $\mathbf{q}$ variable in (4.51), defined as

$$P^t(\mathbf{q}_0, A) := \mathbb{P}(\mathbf{q}(t; \mathbf{q}_0, \mathbf{v}_0) \in A), \quad \mathbf{v}_0 \sim \mu_0,$$

for all $\mathbf{q}_0 \in X$ and $A \in \mathcal{B}_X$, holds the marginal of $\mu \otimes \mu_0$ onto the $\mathbf{q}$ variable, that is, μ , as an invariant measure.

A crucial insight in [10] later revisited in [8, 9], is that each of the components of a certain Strang splitting of (4.51) given by

$$(4.53) \quad \frac{d\mathbf{q}}{dt} = 0, \quad \frac{d\mathbf{v}}{dt} = -CD\Phi(\mathbf{q}) \quad \text{and} \quad \frac{d\mathbf{q}}{dt} = \mathbf{v}, \quad \frac{d\mathbf{v}}{dt} = -\mathbf{q},$$

yield maps $\Xi_t^{(1)}, \Xi_t^{(2)}$ such that $\Xi_t^{(i)}(\mu_0 \otimes \mu_0)$ are absolutely continuous with respect to $\mu_0 \otimes \mu_0$ for $i = 1, 2$ and any $t \geq 0$.⁷ Taking (4.50) and the splitting (4.53) of (4.51) as a starting point therefore suggests the following more general class of trajectory methods.

First we may replace $\mathcal{H}(\mathbf{q}, \mathbf{v})$ in (4.50) with a more general class of nonseparable Hamiltonians which allows for a “position”-dependent kinetic energy component, namely

$$(4.54) \quad \mathcal{H}(\mathbf{q}, \mathbf{v}) := \hat{\Psi}(\mathbf{q}, \mathbf{v}) + \frac{1}{2}|C^{-1/2}\mathbf{v}|^2 + \frac{1}{2}|C^{-1/2}\mathbf{q}|^2 + \Phi(\mathbf{q})$$

for a measurable function $\hat{\Psi} : X \times X \rightarrow \mathbb{R}$. In view of obtaining a well-defined probability measure on the extended state space $X \times X$, we require the following integrability condition:

$$(4.55) \quad \int_{X \times X} \exp(-\hat{\Psi}(\mathbf{q}, \mathbf{v}) - \Phi(\mathbf{q})) \mu_0 \otimes \mu_0(d\mathbf{q}, d\mathbf{v}) < \infty.$$

We further assume that

$$(4.56) \quad \int_X \exp(-\hat{\Psi}(\mathbf{q}, \mathbf{v})) \mu_0(d\mathbf{v}) = 1, \quad \mu_0 \text{ almost surely in } \mathbf{q}.$$

This later assumption ensures that the position marginal of the Gibbs measure $\mathcal{M}$ associated to $\mathcal{H}$ given by (4.54) is the desired target measure μ as in (4.48). We can therefore write, in the notational formulation of Theorem 2.1,

$$(4.57) \quad \mathcal{M}(d\mathbf{q}, d\mathbf{v}) = \mathcal{V}(\mathbf{q}, d\mathbf{v})\mu(d\mathbf{q}) \quad \text{where } \mathcal{V}(\mathbf{q}, d\mathbf{v}) = \exp(-\hat{\Psi}(\mathbf{q}, \mathbf{v}))\mu_0(d\mathbf{v}),$$

so that (4.55) implies mutual absolute continuity between $\mathcal{M}$ and $\mathcal{M}_0 := \mu_0 \otimes \mu_0$.⁸

A Hamiltonian dynamic associated to the nonseparable Hamiltonian in (4.54) can be formally written as

$$(4.58) \quad \frac{d\mathbf{z}}{dt} = J^{-1}D\mathcal{H}(\mathbf{z}),$$

for some general “preconditioning” operator J , and where $D\mathcal{H}$ denotes the Fréchet derivative of $\mathcal{H}$. Under appropriate assumptions on J and $\mathcal{H}$, and following a similar finite-dimensionalization argument as in [10], one can show that such dynamic holds the probability measure $\mathcal{M}$ from (4.57) as invariant. However, notwithstanding Section 4.1.6 above,

⁷In fact it is not hard to see that $\Xi_t^{(2)}$ leaves $\mu_0 \otimes \mu_0$ fixed as is shown in the proof of Theorem 4.18 below.
⁸One can formally relate (4.3), (4.2) and (4.54), (4.57) as follows: We assume that $Z_{\mathcal{K}}(\mathbf{q}) := \int \exp(-\hat{\Psi}(\mathbf{q}, \mathbf{v}) - \frac{1}{2}|C^{-1/2}\mathbf{v}|^2) d\mathbf{v} \equiv 1$ so that (4.3) reduces to (4.54) by setting $\mathcal{K}(\mathbf{q}, \mathbf{v}) = \hat{\Psi}(\mathbf{q}, \mathbf{v}) + \frac{1}{2}|C^{-1/2}\mathbf{v}|^2$ and $\mathcal{U}(\mathbf{q}) = \frac{1}{2}|C^{-1/2}\mathbf{q}|^2 + \Phi(\mathbf{q})$. Of course these relationships are purely formal when $\dim(X) = \infty$ since the Lebesgue measure is ill-defined in infinite dimensions.

deriving a suitable numerical integrator for (4.58) that is well-defined particularly in this infinite-dimensional context is a nontrivial task.

An alternative consists in defining a suitable “surrogate dynamics” for (4.58) that can be more easily integrated. Here we consider the following general and possibly gradient-free “surrogate dynamics,” namely

$$(4.59) \quad \frac{d\mathbf{q}}{dt} = \mathbf{v}, \quad \frac{d\mathbf{v}}{dt} = -\mathbf{q} - f(\mathbf{q}),$$

where $f : X \rightarrow X$ is a measurable mapping. In order to maintain absolute continuity with respect to μ_0 , we assume that $f(\mathbf{q})$ is in the Cameron–Martin space of μ_0 for every $\mathbf{q} \in X$, or in other words

$$(4.60) \quad f(\mathbf{q}) \in \text{Dom}(\mathcal{C}^{-1/2}) \quad \text{for every } \mathbf{q} \in X.$$

Notably, in the particular case that the kernel $\mathcal{V}$ from (4.57) is given by $\mathcal{V}(\mathbf{q}, \cdot) = \mu_0 = \mathcal{N}(0, \mathcal{C})$ for every $\mathbf{q} \in X$, as in (4.50), we may view f as a surrogate for $\mathcal{C}D\Phi$. Then, following (4.53), we consider a numerical discretization of (4.59) by first splitting the dynamics into

$$(4.61) \quad \frac{d\mathbf{q}}{dt} = 0, \quad \frac{d\mathbf{v}}{dt} = -f(\mathbf{q})$$

and

$$(4.62) \quad \frac{d\mathbf{q}}{dt} = \mathbf{v}, \quad \frac{d\mathbf{v}}{dt} = -\mathbf{q}.$$

Clearly, the solution of (4.61) at any time t starting from $(\mathbf{q}_0, \mathbf{v}_0) \in X \times X$ is given by

$$(4.63) \quad \Xi_{f,t}^{(1)}(\mathbf{q}_0, \mathbf{v}_0) = \Xi_t^{(1)}(\mathbf{q}_0, \mathbf{v}_0) = (\mathbf{q}_0, \mathbf{v}_0 - tf(\mathbf{q}_0)),$$

while the corresponding solution of (4.62) is given by

$$(4.64) \quad \Xi_t^{(2)}(\mathbf{q}_0, \mathbf{v}_0) = (\cos(t)\mathbf{q}_0 + \sin(t)\mathbf{v}_0, -\sin(t)\mathbf{q}_0 + \cos(t)\mathbf{v}_0).$$

For any fixed time steps $\delta_1 > 0$ and $\delta_2 > 0$ for (4.63) and (4.64), respectively, and a fixed number $n \in \mathbb{N}$ of iterative steps, we consider a numerical integrator for (4.59) starting from $(\mathbf{q}_0, \mathbf{v}_0) \in X \times X$ given by the following Strang-type splitting:

$$(4.65) \quad \hat{S}(\mathbf{q}_0, \mathbf{v}_0) = (\Xi_{\delta_1}^{(1)} \circ \Xi_{\delta_2}^{(2)} \circ \Xi_{\delta_1}^{(1)})^n(\mathbf{q}_0, \mathbf{v}_0).$$

Here note that, with R taken as the usual “momentum-flip” operator defined in (4.28), and invoking Lemma 4.4.(i) and Lemma 4.2.(i), it is direct to verify that $S = R \circ \hat{S}$ is an involution. For convenience, we also denote by $\Gamma^i(\mathbf{q}_0, \mathbf{v}_0) = (\mathbf{q}_i, \mathbf{v}_i)$ the solution arising from i steps taken in the process of (4.65), that is,

$$(4.66) \quad \Gamma^i(\mathbf{q}_0, \mathbf{v}_0) = (\mathbf{q}_i, \mathbf{v}_i) = (\Xi_{\delta_1}^{(1)} \circ \Xi_{\delta_2}^{(2)} \circ \Xi_{\delta_1}^{(1)})^i(\mathbf{q}_0, \mathbf{v}_0),$$

for $i = 1, \dots, n$.

We notice that, in contrast to (4.51), the system (4.59) for general f may not generate a Hamiltonian dynamics, and also may not hold $\mu \otimes \mu_0$ as an invariant measure. Similarly with a more general class of Gibbsian measures $\mathcal{M}$ as in (4.54), (4.57) it is not clear how to select an f so that the resulting flow (4.59) holds $\mathcal{M}$ as an invariant.⁹ Nevertheless, as we now demonstrate, the class of numerical integrators $\hat{S}$ defined by (4.65) and by “momentum selection” mechanisms $\mathcal{V}$ given in (4.57) allow for the identification of an accept–reject function $\hat{\alpha}$ as in (2.11) which generates a sampling algorithm of the form (2.12) that is reversible with respect to the desired target measure μ .

⁹In this sense we may understand the algorithms introduced in [8] as themselves being surrogate methods.

THEOREM 4.18. *Let X be a separable Hilbert space, and take $\mathcal{C} : X \rightarrow X$ to be a trace-class, symmetric and strictly positive definite operator. Let $\mu_0 = \mathcal{N}(0, \mathcal{C})$ and consider a measurable mapping $\Phi : X \rightarrow \mathbb{R}$ with $e^{-\Phi(\cdot)} \in L^1(\mu_0)$. Under these assumptions, we consider a target probability measure μ as in (4.48). Fix also any measurable $f : X \rightarrow X$ such that $f(\mathbf{q})$ is in the Cameron–Martin space of μ_0 for every $\mathbf{q} \in X$, namely (4.60). Let $\hat{S}$ be the numerical integrator for the dynamics (4.59) defined in (4.65) for any fixed discretization parameters $\delta_1 > 0$, $\delta_2 > 0$, and number of iteration steps $n \geq 1$. Finally, consider the Markov kernel $\mathcal{V}$ and probability measure $\mathcal{M}$ defined in (4.57), with $\hat{\Psi}$ satisfying conditions (4.55) and (4.56).*

Then, taking R to be the “momentum-flip” operator as defined in (4.28), $S := R \circ \hat{S}$ is an involution and $(R \circ \hat{S})^ \mathcal{M}$ is absolutely continuous with respect to $\mathcal{M}$. Furthermore, defining the function $\hat{\alpha} : X \times X \rightarrow [0, 1]$ as*

$$(4.67) \quad \hat{\alpha}(\mathbf{q}_0, \mathbf{v}_0) := 1 \wedge \frac{d(R \circ \hat{S})^* \mathcal{M}}{d\mathcal{M}}(\mathbf{q}_0, \mathbf{v}_0), \quad (\mathbf{q}_0, \mathbf{v}_0) \in X \times X,$$

it follows that the associated Markov kernel $P : X \times \mathcal{B}_X \rightarrow [0, 1]$ given as in (2.12), with these choices of S and $\mathcal{V}$, is reversible with respect to μ . Moreover, invoking the notation from (4.66), we have

$$(4.68) \quad \begin{aligned} & \frac{d(R \circ \hat{S})^* \mathcal{M}}{d\mathcal{M}}(\mathbf{q}_0, \mathbf{v}_0) \\ &= \exp \left(\Phi(\mathbf{q}_0) + \hat{\Psi}(\mathbf{q}_0, \mathbf{v}_0) - \Phi(\mathbf{q}_n) - \hat{\Psi}(\mathbf{q}_n, -\mathbf{v}_n) \right. \\ & \quad - \frac{\delta_1^2}{2} [|\mathcal{C}^{-1/2} f(\mathbf{q}_0)|^2 - |\mathcal{C}^{-1/2} f(\mathbf{q}_n)|^2] + 2\delta_1 \sum_{i=1}^{n-1} \langle \mathcal{C}^{-1/2} \mathbf{v}_i, \mathcal{C}^{-1/2} f(\mathbf{q}_i) \rangle \\ & \quad \left. + \delta_1 [\langle \mathcal{C}^{-1/2} \mathbf{v}_0, \mathcal{C}^{-1/2} f(\mathbf{q}_0) \rangle + \langle \mathcal{C}^{-1/2} \mathbf{v}_n, \mathcal{C}^{-1/2} f(\mathbf{q}_n) \rangle] \right), \end{aligned}$$

for $\mathcal{M}$ -a.e. $(\mathbf{q}_0, \mathbf{v}_0) \in X \times X$.

We summarize the class of algorithms to sample from measures of the form $\mu(d\mathbf{q}) \propto e^{-\Phi(\mathbf{q})} \mu_0(d\mathbf{q})$ for $\mu_0 = \mathcal{N}(0, \mathcal{C})$ resulting from Theorem 4.18 as Algorithm 4.5.

Our proof of Theorem 4.18 is based around the basic properties of pushforward measures (2.6), (2.7) and (2.8), but shares similar ideas to the argument given in [9], Proposition 3.1, so we save our presentation for Appendix B. To provide some general intuition beyond the technical details of the proof, we notice that the integrator $S = R \circ \hat{S}$ in this case involves repeated applications of the maps $\Xi_{\delta_1}^{(1)}$, $\Xi_{\delta_2}^{(2)}$ and R . The latter two maps hold $\mu_0 \otimes \mu_0$ invariant, while the Radon–Nikodym derivative associated with the pushforward of $\mu_0 \otimes \mu_0$ by $\Xi_{\delta_1}^{(1)}$ can be computed via the Cameron–Martin theorem. We then repeatedly apply property (2.8) to compute the Radon–Nikodym derivative associated with the pushforward of $\mu_0 \otimes \mu_0$ by S , which, combined with a simple decomposition using (2.6) and (2.7), yields the desired acceptance probability given by (4.67), (4.68).

To finish this section, we notice that by using a similar idea as in Theorem 4.18, we can also obtain a rigorous reversibility result for a generalized/surrogate version of the infinite-dimensional Metropolis-adjusted Langevin algorithm (∞ MALA) for sampling from measures μ of the form (4.48), as introduced in [11, 22]. In explaining this connection, let us consider, for simplicity, only the case when $\mathcal{V}(\mathbf{q}, \cdot) = \mathcal{N}(0, \mathcal{C})$, for every $\mathbf{q} \in X$, with a suitable covariance operator $\mathcal{C}$.

Algorithm 4.5 Splittings of preconditioned dynamics for Gaussian–Hilbert space targets $\frac{e^{-\Phi(\mathbf{q})}}{Z} \mu_0(d\mathbf{q})$

1: Select the algorithm parameters:

- (i) The form of $\hat{\Psi}(\mathbf{q}, \mathbf{v})$ for the proposal distribution $\mathcal{V}(\mathbf{q}, d\mathbf{v}) := \exp(-\hat{\Psi}(\mathbf{q}, \mathbf{v})) \mu_0(d\mathbf{v})$.
- (ii) The approximation f of $D\Phi$.
- (iii) The discretization time step parameters $\delta_1 > 0$, $\delta_2 > 0$ in the splittings (4.63) and (4.64) respectively.
- (iv) The number of iterative steps $n \in \mathbb{N}$ now defining $\hat{S}$ in (4.65).

2: Choose $\mathbf{q}_0 \in X$

3: **for** $k \geq 0$ **do**

4: Sample $\mathbf{v}_k \sim \mathcal{V}(\mathbf{q}_k, d\mathbf{v})$

5: Propose $\bar{\mathbf{q}}_{k+1} := \Pi_1 \circ \hat{S}(\mathbf{q}_k, \mathbf{v}_k)$, with $\hat{S}$ as defined in (4.65)

6: Set $\mathbf{q}_{k+1} := \bar{\mathbf{q}}_{k+1}$ with probability $\hat{a}(\mathbf{q}_k, \mathbf{v}_k)$ given by (4.67), computed via (4.68), otherwise take $\mathbf{q}_{k+1} := \mathbf{q}_k$.

We recall that such Langevin-type algorithms are defined for a given Fréchet differentiable potential function $\Phi : X \rightarrow \mathbb{R}$ and covariance operator $\mathcal{C} : X \rightarrow X$ as in (4.48), through the following Langevin dynamic:

$$(4.69) \quad d\mathbf{q} + \frac{1}{2} \mathcal{K}(\mathcal{C}^{-1}\mathbf{q} + D\Phi(\mathbf{q})) dt = \sqrt{\mathcal{K}} dW$$

for some “preconditioning” operator $\mathcal{K} : X \rightarrow X$ and a cylindrical Brownian motion W evolving on X (see, e.g., [23] for the general setting of such stochastic evolution equations). Under appropriate assumptions on the potential Φ , it follows that such dynamics holds $\mu(d\mathbf{q}) \propto e^{-\Phi(\mathbf{q})} \mu_0(d\mathbf{q})$ as an invariant measure, and thus effective MCMC proposals can be generated by taking suitable numerical discretizations of (4.69) [11, 22].

Similarly as in (4.59), here we consider the following more general dynamics:

$$(4.70) \quad d\mathbf{q} + \frac{1}{2} \mathcal{K} \mathcal{C}^{-1} \mathbf{q} + f(\mathbf{q}) dt = \sqrt{\mathcal{K}} dW,$$

for some measurable mapping $f : X \rightarrow X$. In practice we would expect f to be taken as a suitable approximation of $\mathcal{K} D\Phi$. Choosing $\mathcal{K} = \mathcal{C}$ and taking a semi-implicit Euler time discretization of (4.70) where the linear part of the drift is approximated in a Crank–Nicolson fashion, yields, for a fixed time step $\delta > 0$,

$$(4.71) \quad \frac{\tilde{\mathbf{q}} - \mathbf{q}}{\delta} = -\frac{1}{2} \left(\frac{\mathbf{q} + \tilde{\mathbf{q}}}{2} + f(\mathbf{q}) \right) + \frac{1}{\sqrt{\delta}} \mathbf{v}, \quad \mathbf{v} \sim \mu_0 = \mathcal{N}(0, \mathcal{C}).$$

Solving (4.71) for $\tilde{\mathbf{q}}$ and denoting $\rho = (4 - \delta)/(4 + \delta)$ thus yields the following proposal map:

$$(4.72) \quad F(\mathbf{q}, \mathbf{v}) := \rho \mathbf{q} + \sqrt{1 - \rho^2} \left(\mathbf{v} - \frac{\sqrt{\delta}}{2} f(\mathbf{q}) \right), \quad \mathbf{q} \in X, \mathbf{v} \sim \mu_0,$$

which then defines the proposal kernel $Q(\mathbf{q}, d\tilde{\mathbf{q}}) = F(\mathbf{q}, \cdot)^* \mu_0(d\tilde{\mathbf{q}})$.

To account for the bias introduced both by replacing $\mathcal{K} D\Phi$ with the surrogate f ¹⁰ and from the numerical discretization, the resulting proposal kernel Q may be complemented with an

¹⁰Here, in contrast to (4.69), the dynamics (4.70) are not guaranteed to hold μ as an invariant measure in general.

appropriate accept–reject function α so as to yield a Markov transition kernel $\bar{P}$ as in (3.1) for which μ is invariant. Following the approach by [72] recalled in Section 3, such an α can be obtained as in (3.2), by computing the Radon–Nikodym derivative $d\eta^\perp/d\eta$ directly, for $\eta(d\mathbf{q}, d\tilde{\mathbf{q}}) = F(\mathbf{q}, \cdot)^* \mu_0(d\tilde{\mathbf{q}}) \mu(d\mathbf{q})$ and $\eta^\perp(d\mathbf{q}, d\tilde{\mathbf{q}}) = \eta(d\tilde{\mathbf{q}}, d\mathbf{q})$, similarly as done in [11].

Alternatively, we may obtain $d\eta^\perp/d\eta$ explicitly as a special case of (4.68), by invoking (3.11) and defining an associated involution mapping S on $X \times X$ according to (3.7). Indeed, denoting $\hat{S} = R \circ S$, where R is the “momentum-flip” map (4.28), so that $S = R \circ \hat{S}$, it follows from a similar observation in [9] that $\hat{S}$ is the particular case of (4.65) for $n = 1$ and specific choices of δ_1, δ_2 . This yields the following corollary, whose complete proof is presented in Appendix C.

COROLLARY 4.19. *Let X be a separable Hilbert space, and take $\mathcal{C} : X \rightarrow X$ to be a trace-class, symmetric and strictly positive definite operator. Let $\mu_0 = \mathcal{N}(0, \mathcal{C})$ and consider a measurable mapping $\Phi : X \rightarrow \mathbb{R}$ with $e^{-\Phi(\cdot)} \in L^1(\mu_0)$. Under these assumptions, we consider a target probability measure μ as in (4.48). Fix also any measurable $f : X \rightarrow X$ such that $f(\mathbf{q})$ is in the Cameron–Martin space of μ_0 for every $\mathbf{q} \in X$, namely (4.60).*

Let $Q : X \times \mathcal{B}_X \rightarrow [0, 1]$ be the Markov kernel $Q(\mathbf{q}, d\tilde{\mathbf{q}}) = F(\mathbf{q}, \cdot)^ \mu_0(d\tilde{\mathbf{q}})$, with F as defined in (4.72). Let $\eta, \eta^\perp \in \text{Pr}(X \times X)$ as in (3.3), namely $\eta(d\mathbf{q}, d\tilde{\mathbf{q}}) = Q(\mathbf{q}, d\tilde{\mathbf{q}}) \mu(d\mathbf{q})$, and $\eta^\perp(d\mathbf{q}, d\tilde{\mathbf{q}}) := \eta(d\tilde{\mathbf{q}}, d\mathbf{q})$. Then, it follows that*

$$\frac{d\eta^\perp}{d\eta}(\mathbf{q}, \tilde{\mathbf{q}}) = \frac{\beta(\tilde{\mathbf{q}}, \mathbf{q})}{\beta(\mathbf{q}, \tilde{\mathbf{q}})},$$

where

$$\beta(\mathbf{q}, \tilde{\mathbf{q}}) := \exp\left(-\Phi(\mathbf{q}) - \frac{\delta}{8}|C^{-1/2}f(\mathbf{q})|^2 - \frac{\sqrt{\delta}}{2}\left\langle \frac{C^{-1/2}(\tilde{\mathbf{q}} - \rho\mathbf{q})}{\sqrt{1-\rho^2}}, C^{-1/2}f(\mathbf{q}) \right\rangle\right).$$

Consequently, defining $\alpha : X \times X \rightarrow [0, 1]$ as in (3.2), that is,

$$(4.73) \quad \alpha(\mathbf{q}, \tilde{\mathbf{q}}) := 1 \wedge \frac{d\eta^\perp}{d\eta}(\mathbf{q}, \tilde{\mathbf{q}}) = 1 \wedge \frac{\beta(\tilde{\mathbf{q}}, \mathbf{q})}{\beta(\mathbf{q}, \tilde{\mathbf{q}})},$$

then it follows that, with these choices of α and Q , the corresponding Markov kernel $\bar{P} : X \times \mathcal{B}_X \rightarrow [0, 1]$ given as in (3.1) satisfies detailed balance with respect to μ .

5. Connections with the classical algorithms. In this final section we make explicit how a variety of established MCMC method can be derived from Theorem 2.1. Here it is notable and interesting that many of the algorithms discussed here can be derived by multiple, nonequivalent, applications of Theorem 2.1; namely different choices of $S, \mathcal{V}$ in (2.12) may ultimately lead to the same sampling kernel. In fact the machinery introduced in Section 3, Section 4 plays an important role here, identifying and/or clarifying various connections between a number of popular sampling methods.

First in Section 5.1 we consider methods defined for sampling continuous distributions on $\mathbb{R}^N$. Further on in Section 5.2 we consider the Hilbert space methods introduced more recently in, for example, [10, 11, 22].

5.1. Finite-dimensional methods. For all examples in this subsection, we assume a finite-dimensional state space $X = \mathbb{R}^N$, and consider a continuously distributed target measure $\mu(d\mathbf{q}) = p(\mathbf{q}) d\mathbf{q}$. We often write μ in the potential form (4.1) convenient for Hamiltonian methods, that is, $\mu(d\mathbf{q}) \propto e^{-\mathcal{U}(\mathbf{q})} d\mathbf{q}$ for some suitably regular potential function $\mathcal{U} : \mathbb{R}^N \rightarrow \mathbb{R}$.

5.1.1. *The classical Metropolis–Hastings formulation.* Start first with the situation when we consider a general continuously distributed Markov kernel $Q(\mathbf{q}, d\tilde{\mathbf{q}}) := q(\mathbf{q}, \tilde{\mathbf{q}}) d\tilde{\mathbf{q}}$, where $q : \mathbb{R}^N \times \mathbb{R}^N \rightarrow \mathbb{R}^+$ is a measurable mapping such that $\int q(\mathbf{q}, \tilde{\mathbf{q}}) d\tilde{\mathbf{q}} = 1$ for any $\mathbf{q} \in \mathbb{R}^N$. Recalling the identity (2.5), this is a special case of Tierney’s formulation [72]. Therefore, precisely as a special case of Remark 3.1, we derive the classical Hastings ratio [43], that is, (1.3), from Theorem 2.1 as follows: take $S(\mathbf{q}, \mathbf{v}) = (\mathbf{v}, \mathbf{q})$ which is clearly a volume preserving involution. Set $\mathcal{V} = Q$ so that $\mathcal{M}(d\mathbf{q}, d\mathbf{v}) = p(\mathbf{q})q(\mathbf{q}, \mathbf{v}) d\mathbf{q} d\mathbf{v}$. Thus, comparing (2.11) with (2.3) we obtain (1.3). Here, strictly speaking, we would need that both $p > 0$ and $q > 0$ almost everywhere to obtain the condition (P2) in Theorem 2.1, but see Remark 2.4.

Note that, with this choice for $S, \mathcal{V}$, we may view [43] as a special case of the Metropolis–Hastings–Green algorithm, [36], discussed above in Remark 2.6. Also, it is amusing to observe that the classical Hastings ratio, (1.3), falls out as special case of Algorithm 4.2. To see this write the proposal density q in the potential form $q(\mathbf{q}, \mathbf{v}) = Z_{\mathcal{K}}(\mathbf{q})^{-1} e^{-\mathcal{K}(\mathbf{q}, \mathbf{v})}$ and set $\mathcal{V} = Q$. In (4.26), we select $f_1(\mathbf{v}) = \mathbf{v}$, $f_2(\mathbf{q}) = -\mathbf{q}$ and set $\delta_1 = \delta_2 = 1$. Similar connections between each of the other algorithms in Section 4 and the acceptance probability (1.3) can be drawn as well.

5.1.2. *Random walk Monte Carlo.* Perhaps the simplest Metropolis–Hastings method is the random walk Monte Carlo (RWMC) algorithm: Take a proposal map $F : \mathbb{R}^N \times \mathbb{R}^N \rightarrow \mathbb{R}^N$ given by

$$(5.1) \quad F(\mathbf{q}, \mathbf{v}) := \mathbf{q} + \mathbf{v}, \quad \mathbf{v} \sim \nu,$$

where $\nu(d\tilde{\mathbf{q}}) = q(\tilde{\mathbf{q}}) d\tilde{\mathbf{q}} = Z_{\mathcal{K}}^{-1} e^{-\mathcal{K}(\tilde{\mathbf{q}})} d\tilde{\mathbf{q}}$ is some jumping probability distribution on $\mathbb{R}^N$. With this definition, (2.3) and noting that $F(\mathbf{q}, \cdot)^{-1}(\tilde{\mathbf{q}}) = \tilde{\mathbf{q}} - \mathbf{q}$ for any fixed $\mathbf{q} \in \mathbb{R}^N$, the associated proposal kernel takes the form

$$(5.2) \quad Q(\mathbf{q}, d\tilde{\mathbf{q}}) := F(\mathbf{q}, \cdot)^* \nu(d\tilde{\mathbf{q}}) = q(\tilde{\mathbf{q}} - \mathbf{q}) d\tilde{\mathbf{q}} = Z_{\mathcal{K}}^{-1} e^{-\mathcal{K}(\tilde{\mathbf{q}} - \mathbf{q})} d\tilde{\mathbf{q}}.$$

Clearly this case falls under the rubric of the usual Hastings ratio, (1.3), which translates to

$$(5.3) \quad \alpha(\mathbf{q}, \tilde{\mathbf{q}}) := 1 \wedge \frac{p(\tilde{\mathbf{q}})q(\mathbf{q} - \tilde{\mathbf{q}})}{p(\mathbf{q})q(\tilde{\mathbf{q}} - \mathbf{q})} = 1 \wedge \exp(\mathcal{U}(\mathbf{q}) - \mathcal{U}(\tilde{\mathbf{q}}) + \mathcal{K}(\tilde{\mathbf{q}} - \mathbf{q}) - \mathcal{K}(\mathbf{q} - \tilde{\mathbf{q}}));$$

cf., for example, [46], equation 3.38 or [29], equation 11.2. Typically, ν is taken to be a zero-mean Gaussian distribution $\mathcal{N}(0, C)$, for some covariance matrix C on $\mathbb{R}^{N \times N}$. Note that, in this particular symmetric case and in general, the terms involving q cancel in (1.3) which is the classical Metropolis case.

Let us observe that the machinery in Section 3.2 reveals a slightly different S and $\mathcal{V}$ which yield the same RWMC algorithm from Theorem 2.1. Comparing (5.2) with (3.7) and using that $F(\mathbf{q}, \cdot)^{-1}(\tilde{\mathbf{q}}) = \tilde{\mathbf{q}} - \mathbf{q}$ we obtain the involution $S(\mathbf{q}, \mathbf{v}) := (\mathbf{q} + \mathbf{v}, -\mathbf{v})$ and then set $\mathcal{V}(\mathbf{q}, d\mathbf{v}) := \nu(d\mathbf{v})$. Thus, according to (2.20), it follows that

$$\begin{aligned} \hat{\alpha}(\mathbf{q}, \mathbf{v}) &= 1 \wedge \frac{dS^*(\mu \otimes \nu)}{d(\mu \otimes \nu)}(\mathbf{q}, \mathbf{v}) = 1 \wedge \frac{p(\mathbf{q} + \mathbf{v})q(-\mathbf{v})}{p(\mathbf{q})q(\mathbf{v})} \\ &= 1 \wedge \exp(-\mathcal{U}(\mathbf{q} + \mathbf{v}) - \mathcal{K}(-\mathbf{v}) + \mathcal{U}(\mathbf{q}) + \mathcal{K}(\mathbf{v})), \end{aligned}$$

which thus recovers α by invoking (3.12). Finally note that this choice of S and $\mathcal{V}$ turns out to be a special case of Algorithm 4.2. To see this take $f_1 := 0$ (and/or $\delta_1 = 0$), set $f_2(\mathbf{v}) := \mathbf{v}$, $\delta_2 = 1$ and choose $n = 1$. According to (4.26), we have $S = R \circ \hat{S}$ with R the usual momentum-flip involution as in (4.28).

5.1.3. *The MALA algorithm.* As introduced in [7, 68], the MALA algorithm uses as its starting point a numerical resolution of the Langevin dynamic

$$(5.4) \quad d\mathbf{q} + \frac{1}{2} \nabla \mathcal{U}(\mathbf{q}) dt = dW,$$

so chosen in order that the target measure $\mu(d\mathbf{q}) \propto e^{-\mathcal{U}(\mathbf{q})} d\mathbf{q}$ is an invariant. Here W is an N -dimensional Brownian motion so that (5.4) is an Itô stochastic differential equation evolving on $\mathbb{R}^N$. By taking an explicit Euler numerical discretization of (5.4) with time step δ^2 , we obtain the proposal kernel $Q(\mathbf{q}, d\tilde{\mathbf{q}}) := F(\mathbf{q}, \cdot)^* \nu(d\tilde{\mathbf{q}})$ where $\nu(d\mathbf{v}) = (2\pi)^{-N/2} e^{-\frac{1}{2}|\mathbf{v}|^2} d\mathbf{v}$, that is, normally distributed with unit covariance, and

$$(5.5) \quad F(\mathbf{q}, \mathbf{v}) := \mathbf{q} - \frac{\delta^2}{2} \nabla \mathcal{U}(\mathbf{q}) + \delta \mathbf{v}.$$

A corresponding acceptance probability α can be directly obtained as in (3.2), that is, $\alpha(\mathbf{q}, \tilde{\mathbf{q}}) = 1 \wedge d\eta^\perp/d\eta(\mathbf{q}, \tilde{\mathbf{q}})$, with

$$\eta(d\mathbf{q}, d\tilde{\mathbf{q}}) = F(\mathbf{q}, \cdot)^* \nu(d\tilde{\mathbf{q}}) \mu(d\mathbf{q}) \propto \exp\left(-\mathcal{U}(\mathbf{q}) - \frac{1}{2\delta^2} \left| \tilde{\mathbf{q}} - \mathbf{q} + \frac{\delta^2}{2} \nabla \mathcal{U}(\mathbf{q}) \right|^2\right) d\mathbf{q} d\tilde{\mathbf{q}},$$

so that

$$(5.6) \quad \alpha(\mathbf{q}, \tilde{\mathbf{q}}) = 1 \wedge \exp\left(-\mathcal{U}(\tilde{\mathbf{q}}) - \frac{1}{2\delta^2} \left| \mathbf{q} - \tilde{\mathbf{q}} + \frac{\delta^2}{2} \nabla \mathcal{U}(\tilde{\mathbf{q}}) \right|^2 + \mathcal{U}(\mathbf{q}) + \frac{1}{2\delta^2} \left| \tilde{\mathbf{q}} - \mathbf{q} + \frac{\delta^2}{2} \nabla \mathcal{U}(\mathbf{q}) \right|^2\right).$$

Alternatively, we notice that under this setting the MALA algorithm can also be seen as a particular case of Algorithm 2.1 associated to Theorem 2.1, by defining an involution map S as in Proposition 3.3 and invoking Theorem 3.4(ii). Indeed, observe that, for each fixed $\mathbf{q}$, $F(\mathbf{q}, \cdot)$ is invertible in $\mathbf{v}$ with

$$F(\mathbf{q}, \cdot)^{-1}(\tilde{\mathbf{q}}) = \frac{1}{\delta} \left(\tilde{\mathbf{q}} - \mathbf{q} + \frac{\delta^2}{2} \nabla \mathcal{U}(\mathbf{q}) \right).$$

Invoking Proposition 3.3, one obtains the desired involution S from (3.7) as

$$(5.7) \quad S(\mathbf{q}, \mathbf{v}) = \left(\mathbf{q} - \frac{\delta^2}{2} \nabla \mathcal{U}(\mathbf{q}) + \delta \mathbf{v}, \frac{\delta}{2} \nabla \mathcal{U}(\mathbf{q}) - \mathbf{v} + \frac{\delta}{2} \nabla \mathcal{U} \left(\mathbf{q} - \frac{\delta^2}{2} \nabla \mathcal{U}(\mathbf{q}) + \delta \mathbf{v} \right) \right).$$

The acceptance probability $\hat{\alpha}$ from Theorem 2.1, that is, $\hat{\alpha}(\mathbf{q}, \mathbf{v}) = 1 \wedge dS^*(\mu \otimes \nu)/d(\mu \otimes \nu)(\mathbf{q}, \mathbf{v})$, is given particularly in this finite-dimensional context by formula (2.20), with $p(\mathbf{q}) \propto e^{-\mathcal{U}(\mathbf{q})}$ and $q(\mathbf{q}, \mathbf{v}) \propto e^{-\frac{1}{2}|\mathbf{v}|^2}$. However, computing ∇S directly from the definition in (5.7) might turn out to be an impractical task.

Instead, we notice that MALA is a special case of Algorithm 4.2 and that S in (5.7) can be written as $R \circ \hat{S}_{1,\delta}$, where R is the velocity-flip involution $R(\mathbf{q}, \mathbf{v}) = (\mathbf{q}, -\mathbf{v})$ and $\hat{S}_{1,\delta}$ is the numerical integrator defined in (4.26) with $n = 1$, $2\delta_1 = \delta_2 = \delta$ and applied to the Hamiltonian dynamic (4.6) with canonical J as in (4.7) in $\mathbb{R}^{2N}$ corresponding to the separated Hamiltonian $\mathcal{H}(\mathbf{q}, \mathbf{v}) = \mathcal{U}(\mathbf{q}) + \frac{1}{2}|\mathbf{v}|^2$, $(\mathbf{q}, \mathbf{v}) \in \mathbb{R}^{2N}$. This leads us to choose $f_1(\mathbf{v}) = \mathbf{v}$ and $f_2(\mathbf{q}) = -\nabla \mathcal{U}(\mathbf{q})$. It thus follows from Theorem 4.7 (or alternatively Proposition 4.9(i)) that S is volume-preserving, so that $|\det \nabla S(\mathbf{q}, \mathbf{v})| = 1$, for all $(\mathbf{q}, \mathbf{v}) \in \mathbb{R}^{2N}$.

Now from (2.20), or equivalently (4.15), we have

$$\begin{aligned} \hat{\alpha}(\mathbf{q}, \mathbf{v}) &= 1 \wedge e^{-\mathcal{H}(S(\mathbf{q}, \mathbf{v})) + \mathcal{H}(\mathbf{q}, \mathbf{v})} \\ &= 1 \wedge \exp\left(-\mathcal{U}(F(\mathbf{q}, \mathbf{v})) - \frac{1}{2} |F(F(\mathbf{q}, \mathbf{v}), \cdot)^{-1}(\mathbf{q})|^2 + \mathcal{U}(\mathbf{q}) + \frac{1}{2} |\mathbf{v}|^2\right). \end{aligned}$$

Invoking (3.12), we thus recover α as in (5.6). This shows that in this setting MALA is a particular case of Algorithm 2.1, or equivalently Algorithm 4.2, corresponding to the involution S as given in (5.7) and the kernel $\mathcal{V}(\mathbf{q}, d\mathbf{v}) = \nu(d\mathbf{v}) \propto e^{-\frac{1}{2}|\mathbf{v}|^2} d\mathbf{v}$.

5.1.4. Hamiltonian Monte Carlo. In the classical HMC algorithm from [25, 61], one takes $\mathcal{K}$ in the Hamiltonian (4.3) as in (4.5), that is, $\mathcal{K}(\mathbf{v}) = \frac{1}{2}\langle M^{-1}\mathbf{v}, \mathbf{v} \rangle$, for a so that the Markov kernel $\mathcal{V}(\mathbf{q}, \cdot)$ is the ($\mathbf{q}$ -independent) $\mathbb{R}^N$ -valued gaussian $\nu = \mathcal{N}(0, M)$. Here, M is a positive-definite mass matrix in $\mathbb{R}^{N \times N}$ which is an algorithmic parameter to be chosen. The corresponding Hamiltonian dynamics is given as in (4.6) with the standard choice of matrix J from (4.7), thus written as

$$(5.8) \quad \frac{d\mathbf{q}}{dt} = M^{-1}\mathbf{v}, \quad \frac{d\mathbf{v}}{dt} = -\nabla\mathcal{U}(\mathbf{q}).$$

A typical choice of numerical integrator $\hat{S}$ for such dynamics is given by the leapfrog integrator $\hat{S}_{n,\delta} = (\Xi_{\delta/2}^{(1)} \circ \Xi_{\delta}^{(2)} \circ \Xi_{\delta/2}^{(1)})^n$ defined as in (4.26) for a given time step $2\delta_1 = \delta_2 = \delta > 0$, number of iterations $n \in \mathbb{N}$ and taking $f_1(\mathbf{v}) = M^{-1}\mathbf{v}$, $f_2(\mathbf{q}) = -\nabla\mathcal{U}(\mathbf{q})$. Here note that $T := \delta \cdot n$ is understood as the total integration time of our numerical approximation of (5.8). A direct calculation (or see Proposition 4.9, Theorem 4.7) yields that each of the solution maps $\Xi_{\delta/2}^{(1)}$ and $\Xi_{\delta}^{(2)}$ are reversible with respect to the “momentum”-flip involution $R(\mathbf{q}, \mathbf{v}) = (\mathbf{q}, -\mathbf{v})$. Hence, by Lemma 4.4(i), $\hat{S}_{n,\delta}$ is reversible with respect to R , so that $S := R \circ \hat{S}_{n,\delta}$ is an involution. The classical HMC algorithm thus follows as a special case of Algorithm 4.2 under these choices.

Other more recent versions of HMC can also be seen to be special cases of Algorithm 4.2. For example, in the relativistic HMC method from [53] the kinetic portion $\mathcal{K}$ of the Hamiltonian in (4.3) is given as

$$(5.9) \quad \mathcal{K}(\mathbf{v}) = mc^2 \left(\frac{|\mathbf{v}|^2}{m^2 c^2} + 1 \right)^{1/2}$$

up to the user-determined algorithmic parameters $m, c > 0$. In the physical analogy drawn here these parameters correspond to “mass” and the “speed of light,” respectively. To connect back to Algorithm 4.2 we keep the same choices for $\mathcal{V}$, R as the classical case and we now take $f_1(\mathbf{v}) = \mathcal{K}(\mathbf{v})$ as in (5.9) leaving $f_2(\mathbf{q}) = -\nabla\mathcal{U}(\mathbf{q})$ to once again define $\hat{S} = (\Xi_{\delta/2}^{(1)} \circ \Xi_{\delta}^{(2)} \circ \Xi_{\delta/2}^{(1)})^n$ according to the associated leap-frog steps $\Xi^{(1)}$, $\Xi^{(2)}$.

Another version is the Riemannian manifold Hamiltonian Monte Carlo (RMHMC) introduced in [33]. Here one considers $\mathcal{K}$ in (4.3) as the negative log-density of a normal distribution with “position”-dependent covariance matrix. More specifically,

$$\mathcal{K}(\mathbf{q}, \mathbf{v}) = \frac{1}{2}\langle M(\mathbf{q})^{-1}\mathbf{v}, \mathbf{v} \rangle,$$

so that $\mathcal{V}(\mathbf{q}, \cdot) = \mathcal{N}(0, M(\mathbf{q}))$. The corresponding Hamiltonian dynamics is written as in (4.6) with J taken in the canonical form as in (4.7). The standard numerical integrator in this case is given by the implicit mapping $\hat{S}_{n,\delta}$ defined in (4.40), for given choices of time step $\delta > 0$ and number of steps $n \in \mathbb{N}$ and f_1, f_2 taken as equalities in (4.36)–(4.37) and note that here $Z_{\mathcal{K}}(\mathbf{q}) = \frac{1}{2} \ln(2\pi) + \frac{N}{2} \ln(\det M(\mathbf{q}))$. From Theorem 4.12, it follows that $\hat{S}_{n,\delta}$ is symplectic (and hence volume preserving) as well as reversible with respect to the momentum-flip involution R in (4.28). Therefore, the RMHMC method follows as a special case of Algorithm 4.4 under this setting.

5.2. *Hilbert space methods.* As in the setting from Section 4.2, here we consider examples where X is a separable Hilbert space, and the target measure μ is of the form (4.48). Namely, $\mu(d\mathbf{q}) \propto e^{-\Phi(\mathbf{q})} \mu_0(d\mathbf{q})$, with $\mu_0 = \mathcal{N}(0, \mathcal{C})$, for a suitable potential function $\Phi : X \rightarrow \mathbb{R}$ and covariance operator $\mathcal{C} : X \rightarrow X$ which is symmetric, positive and trace-class. In particular we will see that all of the algorithms presented in this subsection fall out as special case of Algorithm 4.5 under suitable parameter choices.

5.2.1. *The preconditioned Crank–Nicolson (pCN) algorithm.* The preconditioned Crank–Nicolson MCMC algorithm [11, 22] is derived from the dynamics (4.70) in the particular case when $f = 0$ and $\mathcal{K} = \mathcal{C}$, so that

$$d\mathbf{q} = -\frac{1}{2}\mathbf{q} \, dt + \sqrt{\mathcal{C}} \, dW,$$

which defines an Ornstein–Uhlenbeck process, and holds $\mu_0 = \mathcal{N}(0, \mathcal{C})$ as an invariant measure.

Following a similar derivation as in (4.71)–(4.72), the pCN proposal map is written as

$$(5.10) \qquad F(\mathbf{q}, \mathbf{v}) = \rho \mathbf{q} + \sqrt{1 - \rho^2} \mathbf{v}, \quad \mathbf{q} \in X, \, \mathbf{v} \sim \mu_0 = \mathcal{N}(0, \mathcal{C}),$$

namely we can write the proposal kernel in the form (3.5) with this F and $\mathcal{V}(\mathbf{q}, d\mathbf{v}) = \mu_0(d\mathbf{v})$. Here we recall that $\rho := (4 - \delta)/(4 + \delta)$ for a fixed time step $\delta > 0$.

As in [72], a suitable accept–reject function is given as $\alpha(\mathbf{q}, \tilde{\mathbf{q}}) = 1 \wedge d\eta^\perp/d\eta(\mathbf{q}, \tilde{\mathbf{q}})$, for $\mathbf{q}, \tilde{\mathbf{q}} \in X$, where $\eta(d\mathbf{q}, d\tilde{\mathbf{q}}) = F(\mathbf{q}, \cdot)^* \mu_0(d\tilde{\mathbf{q}}) \mu(d\mathbf{q})$ and $\eta^\perp(d\mathbf{q}, d\tilde{\mathbf{q}}) = \eta(d\tilde{\mathbf{q}}, d\mathbf{q})$, with F as in (5.10), $\mu_0 = \mathcal{N}(0, \mathcal{C})$ and μ as in (4.48). Such α can be computed by noticing that

$$\frac{d\eta^\perp}{d\eta}(\mathbf{q}, \tilde{\mathbf{q}}) = \exp[\Phi(\mathbf{q}) - \Phi(\tilde{\mathbf{q}})] \frac{d\eta_0^\perp}{d\eta_0}(\mathbf{q}, \tilde{\mathbf{q}}),$$

where $\eta_0(d\mathbf{q}, d\tilde{\mathbf{q}}) := F(\mathbf{q}, \cdot)^* \mu_0(d\tilde{\mathbf{q}}) \mu_0(d\mathbf{q})$. Further, it is not difficult to check, for example, via equivalence of characteristic functionals, that $\eta_0^\perp = \eta_0$, so that $d\eta_0^\perp/d\eta_0(\mathbf{q}, \tilde{\mathbf{q}}) = 1$ for η_0 -a.e. $(\mathbf{q}, \tilde{\mathbf{q}}) \in X \times X$ and thus

$$(5.11) \qquad \alpha(\mathbf{q}, \tilde{\mathbf{q}}) = 1 \wedge \exp[\Phi(\mathbf{q}) - \Phi(\tilde{\mathbf{q}})]$$

(cf., e.g., [8, 11, 22]).

As an alternate route, α can also be computed by invoking formula (3.12) with S given by Proposition 3.3, as done in (C.7) below. Indeed, taking $f = 0$ in (C.7) yields immediately (5.11). Regarding the connection to Algorithm 4.5, from (C.3) below, here we notice that the associated involution S derived from (3.7) can be written as $S = R \circ \Xi_{\delta_2}^{(2)}$, where R is the momentum-flip operator (4.28), $\delta_2 > 0$ is such that $\rho = \cos \delta_2$, and $\Xi_t^{(2)}$ is as defined in (4.64).

5.2.2. *The infinite-dimensional Metropolis-adjusted Langevin algorithm (∞ MALA).* The infinite-dimensional Metropolis-adjusted Langevin algorithm (∞ MALA) [11, 22] is derived from the Langevin dynamic (4.69) via the numerical discretization (4.71). In this case, for a fixed time step $\delta > 0$, it follows from (4.72) with $f = \mathcal{C} D\Phi$ that the proposal map is given by

$$(5.12) \qquad F(\mathbf{q}, \mathbf{v}) := \rho \mathbf{q} + \sqrt{1 - \rho^2} \left(\mathbf{v} - \frac{\sqrt{\delta}}{2} \mathcal{C} D\Phi(\mathbf{q}) \right), \quad \mathbf{q} \in X, \, \mathbf{v} \sim \mu_0,$$

where we recall that $\rho = (4 - \delta)/(4 + \delta)$. Together with the decomposition in (C.3) below, it thus follows that ∞ MALA is the particular case of Algorithm 4.5 with the specific choices $\tilde{\Psi} = 0$ (i.e., $\mathcal{V}(\mathbf{q}, \cdot) = \mu_0$ for all $\mathbf{q} \in X$), $f = \mathcal{C} D\Phi$, $\delta_1 := \sqrt{\delta}/2$, $\delta_2 := \cos^{-1} \rho$ and $n = 1$.

Moreover, from (4.73) it follows that an accept-reject function for ∞ MALA is given by

$$(5.13) \quad \alpha(\mathbf{q}, \tilde{\mathbf{q}}) = 1 \wedge \frac{\beta(\tilde{\mathbf{q}}, \mathbf{q})}{\beta(\mathbf{q}, \tilde{\mathbf{q}})},$$

where

$$\beta(\mathbf{q}, \tilde{\mathbf{q}}) := \exp\left(-\Phi(\mathbf{q}) - \frac{\delta}{8} |\mathcal{C}^{1/2} D\Phi(\mathbf{q})|^2 - \frac{\sqrt{\delta}}{2} \left\langle \frac{\tilde{\mathbf{q}} - \rho \mathbf{q}}{\sqrt{1 - \rho^2}}, D\Phi(\mathbf{q}) \right\rangle\right),$$

for η -a.e. $(\mathbf{q}, \tilde{\mathbf{q}}) \in X \times X$. Here, $\eta(d\mathbf{q}, d\tilde{\mathbf{q}}) = F(\mathbf{q}, \cdot)^* \mu_0(d\tilde{\mathbf{q}}) \mu(d\mathbf{q})$ with F as given in (5.12).

5.2.3. The infinite-dimensional Hamiltonian Monte Carlo (∞ HMC) algorithm. The infinite-dimensional Hamiltonian Monte Carlo algorithm introduced in [10] is the particular case of Algorithm 4.5 when $\hat{\Psi} = 0$ (i.e., $\mathcal{V}(\mathbf{q}, \cdot) = \mu_0$ for all $\mathbf{q} \in X$), $f = \mathcal{C} D\Phi$, and when the discretization parameters δ_1, δ_2 from (4.65) are defined as $\delta_1 := \delta/2$ and $\delta_2 := \delta$ for some fixed time step $\delta > 0$. From (4.68), it thus follows that an acceptance probability in this case is given by

$$\begin{aligned} \hat{\alpha}(\mathbf{q}_0, \mathbf{v}_0) &= 1 \wedge \frac{d(R \circ \hat{S})^*(\mu \otimes \mu_0)}{d(\mu \otimes \mu_0)}(\mathbf{q}_0, \mathbf{v}_0) \\ &= 1 \wedge \exp\left(\Phi(\mathbf{q}_0) - \Phi(\mathbf{q}_n) - \frac{\delta^2}{8} [|\mathcal{C}^{1/2} D\Phi(\mathbf{q}_0)|^2 - |\mathcal{C}^{1/2} D\Phi(\mathbf{q}_n)|^2] \right. \\ &\quad \left. + \delta \sum_{i=1}^{n-1} \langle \mathbf{v}_i, D\Phi(\mathbf{q}_i) \rangle + \frac{\delta}{2} [\langle \mathbf{v}_0, D\Phi(\mathbf{q}_0) \rangle + \langle \mathbf{v}_n, D\Phi(\mathbf{q}_n) \rangle] \right), \end{aligned}$$

where R is the “momentum-flip” map (4.28), and we invoked similar notation as in (4.66).

A geometric version of this ∞ HMC algorithm that considers a certain position-dependent kinetic energy as in (4.54) was developed in [8]. Specifically, the kernel $\mathcal{V}(\mathbf{q}, \cdot)$ at each $\mathbf{q} \in X$ is taken to be a Gaussian $\mathcal{N}(0, \mathcal{K}(\mathbf{q}))$, for a position-dependent covariance operator $\mathcal{K}(\mathbf{q}) : X \rightarrow X$ that is assumed to be trace-class, symmetric and strictly positive definite. Here one assumes additionally that $\text{Im}(\mathcal{K}(\mathbf{q})^{1/2}) = \text{Im}(\mathcal{C}^{1/2})$, and that $(\mathcal{C}^{-1/2} \mathcal{K}(\mathbf{q})^{1/2})(\mathcal{C}^{-1/2} \mathcal{K}(\mathbf{q})^{1/2})^* - I$ is a Hilbert–Schmidt operator on X , for μ_0 -a.e. $\mathbf{q} \in X$. Under these assumptions, it follows from the Feldman–Hajek theorem that $\mathcal{V}(\mathbf{q}, \cdot) = \mathcal{N}(0, \mathcal{K}(\mathbf{q}))$ and $\mu_0 = \mathcal{N}(0, \mathcal{C})$ are mutually absolutely continuous for μ_0 -a.e. $\mathbf{q} \in X$ (see, e.g., [23]), so that $\hat{\Psi}$ in (4.57) is well defined. The authors then consider the following surrogate dynamic from the full corresponding Hamiltonian system:

$$\frac{d\mathbf{q}}{dt} = \mathbf{v}, \quad \frac{d\mathbf{v}}{dt} = -\mathcal{K}(\mathbf{q})[\mathcal{C}^{-1} \mathbf{q} + D\Phi(\mathbf{q})],$$

which is thus a particular case of the general system (4.59) with

$$(5.14) \quad f(\mathbf{q}) = \mathcal{K}(\mathbf{q})[(\mathcal{C}^{-1} - \mathcal{K}^{-1}(\mathbf{q}))\mathbf{q} + D\Phi(\mathbf{q})].$$

Due to the fixed assumptions on $\mathcal{K}(\mathbf{q})$ and $\mathcal{C}$, it follows that such f satisfies (4.60). Proceeding with a numerical splitting as in (4.61)–(4.62), (4.65), we obtain that this geometric version of the ∞ HMC algorithm is a particular case of Algorithm 4.5 with the choices $\mathcal{V}(\mathbf{q}, \cdot) = \mathcal{N}(0, \mathcal{K}(\mathbf{q}))$, f as in (5.14), $\delta_1 = \delta$ and $\delta_2 = \delta/2$ for a given time step $\delta > 0$, and any number of iterative steps $n \in \mathbb{N}$.

APPENDIX A: PROOF OF THEOREM 4.12

We begin with the first item and prove that $\hat{S}$ is R -reversible under assumption (4.41). Since $\hat{S}$ is the n -fold composition of $\hat{S}_{\delta/2}^{(A)} \circ \hat{S}_{\delta/2}^{(B)}$, due to Lemma 4.4(ii), it suffices to show that $\hat{S}_{\delta/2}^{(A)} \circ \hat{S}_{\delta/2}^{(B)}$ is R -reversible. Moreover, as we already observed above, it is direct to verify that the adjoint of the Euler-B scheme is the Euler-A scheme in sense given in Definition 4.5(i). In other words we have that $(\hat{S}_{\delta/2}^{(B)})^* = \hat{S}_{\delta/2}^{(A)}$ and so, as observed in Lemma 4.6(i), we infer that $\hat{S}_{\delta/2}^{(A)} \circ \hat{S}_{\delta/2}^{(B)}$ is symmetric. Invoking Lemma 4.6(ii) and Lemma 4.4(i), it therefore suffices to show that

$$(A.1) \quad R \circ \hat{S}_{\delta/2}^{(A)} = \hat{S}_{-\delta/2}^{(A)} \circ R \quad \text{and} \quad R \circ \hat{S}_{\delta/2}^{(B)} = \hat{S}_{-\delta/2}^{(B)} \circ R.$$

Let us verify the statement in (A.1) for $\hat{S}_{\delta/2}^{(A)}$. Let $\bar{\mathbf{z}} = (\bar{\mathbf{q}}, \bar{\mathbf{v}}) \in \mathbb{R}^{2N}$ and denote $\mathbf{z} = (\mathbf{q}, \mathbf{v}) := \hat{S}_{\delta/2}^{(A)}(\bar{\mathbf{z}})$. From (4.39), we have

$$\mathbf{z} = \bar{\mathbf{z}} + \frac{\delta}{2} \mathbf{f}(\mathbf{q}, \bar{\mathbf{v}}) = \bar{\mathbf{z}} + \frac{\delta}{2} \mathbf{f}(\Pi_1 \mathbf{z}, \Pi_2 \bar{\mathbf{z}}).$$

Applying R and invoking (4.41), it follows that

$$(A.2) \quad R\mathbf{z} = R\bar{\mathbf{z}} + \frac{\delta}{2} R\mathbf{f}(\Pi_1 \mathbf{z}, \Pi_2 \bar{\mathbf{z}}) = R\bar{\mathbf{z}} - \frac{\delta}{2} \mathbf{f}(\Pi_1 \circ R\mathbf{z}, \Pi_2 \circ R\bar{\mathbf{z}}).$$

On the other hand, denoting $\tilde{\mathbf{z}} := \hat{S}_{-\delta/2}^{(A)}(R\bar{\mathbf{z}})$, we have again from (4.39) that

$$(A.3) \quad \tilde{\mathbf{z}} = R\bar{\mathbf{z}} - \frac{\delta}{2} \mathbf{f}(\Pi_1 \tilde{\mathbf{z}}, \Pi_2 \circ R\bar{\mathbf{z}}).$$

From (A.2), (A.3) we conclude by uniqueness of solutions for (4.38) that $R\mathbf{z} = \tilde{\mathbf{z}}$ or in other words $R \circ \hat{S}_{\delta/2}^{(A)}(\bar{\mathbf{z}}) = \hat{S}_{-\delta/2}^{(A)} \circ R(\bar{\mathbf{z}})$, for every $\bar{\mathbf{z}} \in \mathbb{R}^{2N}$ as desired. The proof for $\hat{S}^{(B)}$ is analogous. Finally, notice that if R is the momentum-flip involution (4.28), then (4.41) reduces to the requirement (4.42). This finishes the proof of the first item.

Turning to the second item we follow the approach in [48], Chapter 4.1, and show that $d\mathbf{q} \wedge d\mathbf{v} = d\bar{\mathbf{q}} \wedge d\bar{\mathbf{v}}$ for $(\mathbf{q}, \mathbf{v}) = \hat{S}_{\delta/2}^{(A)}(\bar{\mathbf{q}}, \bar{\mathbf{v}})$ and for $(\mathbf{q}, \mathbf{v}) = \hat{S}_{\delta/2}^{(B)}(\bar{\mathbf{q}}, \bar{\mathbf{v}})$. Here d is the exterior derivative and $\wedge$ is the wedge product. See the footnote in Definition 4.1 above and reference, for example, [48, 74] for further details and proper definitions. Starting with Euler-A we have, referring back to (4.39), that

$$\begin{aligned} d\mathbf{q} &= d\bar{\mathbf{q}} + \frac{\delta}{2} \nabla_{\mathbf{q}} f_1(\mathbf{q}, \bar{\mathbf{v}}) d\mathbf{q} + \frac{\delta}{2} \nabla_{\mathbf{v}} f_1(\mathbf{q}, \bar{\mathbf{v}}) d\bar{\mathbf{v}}, \\ d\mathbf{v} &= d\bar{\mathbf{v}} + \frac{\delta}{2} \nabla_{\mathbf{q}} f_2(\mathbf{q}, \bar{\mathbf{v}}) d\mathbf{q} + \frac{\delta}{2} \nabla_{\mathbf{v}} f_2(\mathbf{q}, \bar{\mathbf{v}}) d\bar{\mathbf{v}}. \end{aligned}$$

We now compute $d\mathbf{q} \wedge d\mathbf{v}$ by expanding first in $d\mathbf{v}$ then in $d\mathbf{q}$ appropriately. As illustrated in [48], Chapter 3.6, we make use of the identity that $d\mathbf{w} \wedge (A d\tilde{\mathbf{w}}) = (A^* d\mathbf{w}) \wedge d\tilde{\mathbf{w}}$ for any $\mathbf{w}, \tilde{\mathbf{w}} \in \mathbb{R}^N$ and any $N \times N$ matrix A so that in particular $d\mathbf{w} \wedge (A d\mathbf{w}) = 0$ when A is symmetric. With our standing assumption (4.43) we find

$$\begin{aligned} d\mathbf{q} \wedge d\mathbf{v} &= d\mathbf{q} \wedge \left(d\bar{\mathbf{v}} + \frac{\delta}{2} \nabla_{\mathbf{q}} f_2(\mathbf{q}, \bar{\mathbf{v}}) d\mathbf{q} + \frac{\delta}{2} \nabla_{\mathbf{v}} f_2(\mathbf{q}, \bar{\mathbf{v}}) d\bar{\mathbf{v}} \right) \\ &= d\mathbf{q} \wedge d\bar{\mathbf{v}} + d\mathbf{q} \wedge \left(\frac{\delta}{2} \nabla_{\mathbf{v}} f_2(\mathbf{q}, \bar{\mathbf{v}}) d\bar{\mathbf{v}} \right) \end{aligned}$$

$$\begin{aligned}
&= d\bar{\mathbf{q}} \wedge d\bar{\mathbf{v}} + \left(\frac{\delta}{2} \nabla_{\mathbf{q}} f_1(\mathbf{q}, \bar{\mathbf{v}}) d\mathbf{q} + \frac{\delta}{2} \nabla_{\mathbf{v}} f_1(\mathbf{q}, \bar{\mathbf{v}}) d\bar{\mathbf{v}} \right) \wedge d\bar{\mathbf{v}} + d\mathbf{q} \wedge \left(\frac{\delta}{2} \nabla_{\mathbf{v}} f_2(\mathbf{q}, \bar{\mathbf{v}}) d\bar{\mathbf{v}} \right) \\
&= d\bar{\mathbf{q}} \wedge d\bar{\mathbf{v}},
\end{aligned}$$

as desired in the first case. Regarding Euler-B, (4.38), we have

$$\begin{aligned}
d\mathbf{q} &= d\bar{\mathbf{q}} + \frac{\delta}{2} \nabla_{\mathbf{q}} f_1(\bar{\mathbf{q}}, \mathbf{v}) d\bar{\mathbf{q}} + \frac{\delta}{2} \nabla_{\mathbf{v}} f_1(\bar{\mathbf{q}}, \mathbf{v}) d\mathbf{v}, \\
d\mathbf{v} &= d\bar{\mathbf{v}} + \frac{\delta}{2} \nabla_{\mathbf{q}} f_2(\bar{\mathbf{q}}, \mathbf{v}) d\bar{\mathbf{q}} + \frac{\delta}{2} \nabla_{\mathbf{v}} f_2(\bar{\mathbf{q}}, \mathbf{v}) d\mathbf{v}.
\end{aligned}$$

Similar to the previous case but now expanding first in $d\mathbf{q}$ and then in $d\mathbf{v}$ we have:

$$\begin{aligned}
d\mathbf{q} \wedge d\mathbf{v} &= \left(d\bar{\mathbf{q}} + \frac{\delta}{2} \nabla_{\mathbf{q}} f_1(\bar{\mathbf{q}}, \mathbf{v}) d\bar{\mathbf{q}} + \frac{\delta}{2} \nabla_{\mathbf{v}} f_1(\bar{\mathbf{q}}, \mathbf{v}) d\mathbf{v} \right) \wedge d\mathbf{v} \\
&= d\bar{\mathbf{q}} \wedge d\mathbf{v} + \left(\frac{\delta}{2} \nabla_{\mathbf{q}} f_1(\bar{\mathbf{q}}, \mathbf{v}) d\bar{\mathbf{q}} \right) \wedge d\mathbf{v} \\
&= d\bar{\mathbf{q}} \wedge \left(d\bar{\mathbf{v}} + \frac{\delta}{2} \nabla_{\mathbf{q}} f_2(\bar{\mathbf{q}}, \mathbf{v}) d\bar{\mathbf{q}} + \frac{\delta}{2} \nabla_{\mathbf{v}} f_2(\bar{\mathbf{q}}, \mathbf{v}) d\mathbf{v} \right) + \left(\frac{\delta}{2} \nabla_{\mathbf{q}} f_1(\bar{\mathbf{q}}, \mathbf{v}) d\bar{\mathbf{q}} \right) \wedge d\mathbf{v} \\
&= d\bar{\mathbf{q}} \wedge d\bar{\mathbf{v}},
\end{aligned}$$

completing the proof of the second item.

We address the final item (iii) showing in this case that $\hat{S}$ is volume-preserving by explicitly computing $\det \nabla \hat{S}$. Since $\hat{S}$ is the composition of half-steps of $\hat{S}^{(B)}$ and $\hat{S}^{(A)}$, it suffices to verify that each of these integrators is volume-preserving. We show this only for $\hat{S}^{(A)}$, since the proof for $\hat{S}^{(B)}$ follows analogously. Taking $(f_1(\mathbf{z}), f_2(\mathbf{z})) = \tilde{J}^{-1} \nabla \tilde{\mathcal{H}}(\mathbf{z})$ in (4.39) and differentiating $(\mathbf{q}^*, \mathbf{v}^*) = \hat{S}_\delta^{(A)}(\bar{\mathbf{q}}, \bar{\mathbf{v}})$ with respect to $\bar{\mathbf{q}}$ and $\bar{\mathbf{v}}$, it follows that

$$\begin{pmatrix} I & -\delta E \tilde{\mathcal{H}}_{\mathbf{v}\mathbf{v}} \\ 0 & I + \delta E \tilde{\mathcal{H}}_{\mathbf{q}\mathbf{v}} \end{pmatrix} \nabla \hat{S}_{1,\delta}^{(A)}(\bar{\mathbf{q}}, \bar{\mathbf{v}}) = \begin{pmatrix} I + \delta E \tilde{\mathcal{H}}_{\mathbf{v}\mathbf{q}} & 0 \\ -\delta E \tilde{\mathcal{H}}_{\mathbf{q}\mathbf{q}} & I \end{pmatrix},$$

where $\tilde{\mathcal{H}}_{\mathbf{q}\mathbf{q}}, \tilde{\mathcal{H}}_{\mathbf{q}\mathbf{v}}, \tilde{\mathcal{H}}_{\mathbf{v}\mathbf{q}}, \tilde{\mathcal{H}}_{\mathbf{v}\mathbf{v}} \in \mathbb{R}^{N \times N}$ are evaluated at $(\mathbf{q}^*, \mathbf{v}^*)$ and denote the matrices of second-order partial derivatives of $\tilde{\mathcal{H}}$ with respect to the variables $\mathbf{q}$ and/or $\mathbf{v}$. Therefore,

$$\begin{aligned}
&\nabla \hat{S}_\delta^{(A)}(\bar{\mathbf{q}}, \bar{\mathbf{v}}) \\
&= \begin{pmatrix} I + \delta E \tilde{\mathcal{H}}_{\mathbf{v}\mathbf{q}} - \delta^2 E \tilde{\mathcal{H}}_{\mathbf{v}\mathbf{v}} (I + \delta E \tilde{\mathcal{H}}_{\mathbf{q}\mathbf{v}})^{-1} E \tilde{\mathcal{H}}_{\mathbf{q}\mathbf{q}} & \delta E \tilde{\mathcal{H}}_{\mathbf{v}\mathbf{v}} (I + \delta E \tilde{\mathcal{H}}_{\mathbf{q}\mathbf{v}})^{-1} \\ -\delta (I + \delta E \tilde{\mathcal{H}}_{\mathbf{q}\mathbf{v}})^{-1} E \tilde{\mathcal{H}}_{\mathbf{q}\mathbf{q}} & (I + \delta E \tilde{\mathcal{H}}_{\mathbf{q}\mathbf{v}})^{-1} \end{pmatrix}.
\end{aligned}
\tag{A.4}$$

Invoking the formula for the determinant of block matrices, namely

$$\det \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det(AD - BD^{-1}CD) \quad \text{whenever } A, B, C, D \in \mathbb{R}^{N \times N} \text{ and } D \text{ is invertible}
\tag{A.5}$$

(see, e.g., [69]), a direct calculation yields that $|\det \nabla \hat{S}_\delta^{(A)}(\bar{\mathbf{q}}, \bar{\mathbf{v}})| = 1$ for any $(\bar{\mathbf{q}}, \bar{\mathbf{v}}) \in \mathbb{R}^{2N}$. The proof is now complete.

APPENDIX B: PROOF OF THEOREM 4.18

As usual we proceed by establishing conditions (P1), (P2) in Theorem 2.1 for the elements $S = R \circ \hat{S}$ and $\mathcal{M}$ specified according to (4.65) and (4.57), respectively. In the process of establishing (P2) we demonstrate (4.68) by making suitable use of various identities from

Section 2.1. Regarding (P1), from the definitions of $\Xi_t^{(1)}$ and $\Xi_t^{(2)}$ in (4.63) and (4.64), it is not difficult to check that

$$(B.1) \quad (R \circ \Xi_t^{(1)})^2 = I \quad \text{and} \quad (R \circ \Xi_t^{(2)})^2 = I \quad \text{for any } t \geq 0.$$

Hence, from Lemma 4.4(i) and Lemma 4.2(i) it follows that $(R \circ \hat{S})^2 = I$, so that S is indeed an involution and thus condition (P1) is verified.

Turning to condition (P2) of Theorem 2.1 we introduce an intermediate measure $\mathcal{M}_0 = \mu_0 \otimes \mu_0$. Notice that, embedded in the assumption (4.55), we have that $\mathcal{M}$ and $\mathcal{M}_0$ must be mutually absolutely continuous and hence it follows that $S^*\mathcal{M}$ and $S^*\mathcal{M}_0$ must also be mutually absolutely continuous. Thus, in order to establish (P2), it suffices to prove that $S^*\mathcal{M}_0$ is absolutely continuous with respect to $\mathcal{M}_0$. This being true, notice moreover that, with the fact that S is an involution and the aide of the identities (2.6), (2.7), we have

$$(B.2) \quad \frac{dS^*\mathcal{M}}{d\mathcal{M}}(\mathbf{q}, \mathbf{v}) = \frac{d\mathcal{M}_0}{d\mathcal{M}}(\mathbf{q}, \mathbf{v}) \frac{d\mathcal{M}}{d\mathcal{M}_0}(S(\mathbf{q}, \mathbf{v})) \frac{dS^*\mathcal{M}_0}{d\mathcal{M}_0}(\mathbf{q}, \mathbf{v})$$

for $\mathcal{M}$ -a.e. $(\mathbf{q}, \mathbf{v}) \in X \times X$. Here, using (2.5), we have

$$(B.3) \quad \frac{d\mathcal{M}_0}{d\mathcal{M}}(\mathbf{q}_0, \mathbf{v}_0) = Z \exp(\Phi(\mathbf{q}_0) + \hat{\Psi}(\mathbf{q}_0, \mathbf{v}_0))$$

and similarly, invoking the notation introduced in (4.66),

$$(B.4) \quad \frac{d\mathcal{M}}{d\mathcal{M}_0}(S(\mathbf{q}_0, \mathbf{v}_0)) = \frac{d\mathcal{M}}{d\mathcal{M}_0}(\mathbf{q}_n, -\mathbf{v}_n) = \frac{1}{Z} \exp(-\Phi(\mathbf{q}_n) - \hat{\Psi}(\mathbf{q}_n, -\mathbf{v}_n)).$$

We now show that indeed $S^*\mathcal{M}_0 \ll \mathcal{M}_0$ and compute $dS^*\mathcal{M}_0/d\mathcal{M}_0$ by breaking S up into its constituent maps and making repeated usage of (2.8). Start by noticing that the mappings $\Xi_{\delta_1}^{(1)}$ and $\Xi_{\delta_2}^{(2)}$ defined in (4.63)–(4.64) satisfy $(\Xi_{\delta_i}^{(i)})^*\mathcal{M}_0 \ll \mathcal{M}_0$, $i = 1, 2$. Indeed, it is not difficult to check via, for example, the equivalence of characteristic functionals that the rotation mapping $\Xi_{\delta_2}^{(2)}$ preserves the measure $\mathcal{M}_0$, that is, $(\Xi_{\delta_2}^{(2)})^*\mathcal{M}_0 = \mathcal{M}_0$, so that in particular

$$(B.5) \quad \frac{d(\Xi_{\delta_2}^{(2)})^*\mathcal{M}_0}{d\mathcal{M}_0}(\mathbf{q}, \mathbf{v}) = 1 \quad \text{for } \mathcal{M}_0\text{-a.e. } (\mathbf{q}, \mathbf{v}) \in X \times X.$$

Regarding $\Xi_{\delta_1}^{(1)}$ we denote by $G : X \times X \rightarrow X$ the mapping corresponding to the second component of $\Xi_{\delta_1}^{(1)}$, that is,

$$G(\mathbf{q}, \mathbf{v}) := \mathbf{v} - \delta_1 f(\mathbf{q}), \quad (\mathbf{q}, \mathbf{v}) \in X \times X,$$

and thus obtain

$$(\Xi_{\delta_1}^{(1)})^*\mathcal{M}_0(d\mathbf{q}, d\mathbf{v}) = G(\mathbf{q}, \cdot)^*\mu_0(d\mathbf{v})\mu_0(d\mathbf{q}).$$

Now, for $\mathbf{v} \sim \mu_0 = \mathcal{N}(0, C)$ it follows that for each fixed $\mathbf{q} \in X$, $G(\mathbf{q}, \mathbf{v}) \sim \mathcal{N}(-\delta_1 f(\mathbf{q}), C)$. Hence, for each $\mathbf{q} \in X$, $G(\mathbf{q}, \cdot)^*\mu_0 = \text{Law}(G(\mathbf{q}, \mathbf{v})) = \mathcal{N}(-\delta_1 f(\mathbf{q}), C)$. Under our assumption (4.60), namely that $f(\mathbf{q}) \in \text{Dom}(C^{-1/2})$, we have that μ_0 and $G(\mathbf{q}, \cdot)^*\mu_0$ are mutually absolutely continuous and, consequently, $(\Xi_{\delta_1}^{(1)})^*\mathcal{M}_0 \ll \mathcal{M}_0$, with

$$(B.6) \quad \begin{aligned} \frac{d(\Xi_{\delta_1}^{(1)})^*\mathcal{M}_0}{d\mathcal{M}_0}(\mathbf{q}, \mathbf{v}) &= \frac{dG(\mathbf{q}, \cdot)^*\mu_0}{d\mu_0}(\mathbf{v}) \\ &= \exp\left(-\delta_1 \langle C^{-1/2} f(\mathbf{q}), C^{-1/2} \mathbf{v} \rangle - \frac{\delta_1^2}{2} |C^{-1/2} f(\mathbf{q})|^2\right) \end{aligned}$$

for $\mathcal{M}_0$ -a.e. $(\mathbf{q}, \mathbf{v}) \in X \times X$ (see, e.g., [23], Theorem 2.23).

From (2.8) and (B.5) and recalling the notation (4.66), we deduce that the mapping $\Gamma := \Xi_{\delta_1}^{(1)} \circ \Xi_{\delta_2}^{(2)} \circ \Xi_{\delta_1}^{(1)}$ satisfies $\Gamma^* \mathcal{M}_0 \ll \mathcal{M}_0$ and

$$(B.7) \quad \frac{d\Gamma^* \mathcal{M}_0}{d\mathcal{M}_0}(\mathbf{q}, \mathbf{v}) = \frac{d(\Xi_{\delta_1}^{(1)})^* \mathcal{M}_0}{d\mathcal{M}_0}((\Xi_{\delta_1}^{(1)} \circ \Xi_{\delta_2}^{(2)})^{-1}(\mathbf{q}, \mathbf{v})) \frac{d(\Xi_{\delta_1}^{(1)})^* \mathcal{M}_0}{d\mathcal{M}_0}(\mathbf{q}, \mathbf{v}),$$

for $\mathcal{M}_0$ -a.e. $(\mathbf{q}, \mathbf{v}) \in X \times X$. Consequently, invoking (2.8) once again, we obtain that $\hat{S} = \Gamma^n$ satisfies $\hat{S}^* \mathcal{M}_0 \ll \mathcal{M}_0$, with

$$(B.8) \quad \frac{d\hat{S}^* \mathcal{M}_0}{d\mathcal{M}_0}(\mathbf{q}, \mathbf{v}) = \prod_{i=1}^n \frac{d\Gamma^* \mathcal{M}_0}{d\mathcal{M}_0}((\Gamma^{i-1})^{-1}(\mathbf{q}, \mathbf{v})),$$

for $\mathcal{M}_0$ -a.e. $(\mathbf{q}, \mathbf{v}) \in X \times X$, where we set $\Gamma^0 := I$.

Finally, it is not difficult to check (e.g., via equivalence of characteristic functionals) that the momentum-flip operator R , defined as in (4.28), preserves the measure $\mathcal{M}_0$, so that

$$\frac{dR^* \mathcal{M}_0}{d\mathcal{M}_0}(\mathbf{q}, \mathbf{v}) = 1 \quad \text{for } \mathcal{M}_0\text{-a.e. } (\mathbf{q}, \mathbf{v}) \in X \times X.$$

Therefore, with a third application of (2.8), we deduce that $(R \circ \hat{S})^* \mathcal{M}_0 \ll \mathcal{M}_0$ and

$$(B.9) \quad \begin{aligned} \frac{d(R \circ \hat{S})^* \mathcal{M}_0}{d\mathcal{M}_0}(\mathbf{q}, \mathbf{v}) &= \frac{d\hat{S}^* \mathcal{M}_0}{d\mathcal{M}_0}(R^{-1}(\mathbf{q}, \mathbf{v})) \frac{dR^* \mathcal{M}_0}{d\mathcal{M}_0}(\mathbf{q}, \mathbf{v}) \\ &= \frac{d\hat{S}^* \mathcal{M}_0}{d\mathcal{M}_0}(R(\mathbf{q}, \mathbf{v})), \end{aligned}$$

for $\mathcal{M}_0$ -a.e. $(\mathbf{q}, \mathbf{v}) \in X \times X$, where in the last equality we used that $R^2 = I$.

Moreover, from (B.8) and (B.9) it follows that for $\mathcal{M}_0$ -a.e. $(\mathbf{q}, \mathbf{v}) \in X \times X$

$$(B.10) \quad \begin{aligned} \frac{d(R \circ \hat{S})^* \mathcal{M}_0}{d\mathcal{M}_0}(\mathbf{q}, \mathbf{v}) &= \prod_{i=1}^n \frac{d\Gamma^* \mathcal{M}_0}{d\mathcal{M}_0}((\Gamma^{i-1})^{-1} \circ R(\mathbf{q}, \mathbf{v})) \\ &= \prod_{i=1}^n \frac{d\Gamma^* \mathcal{M}_0}{d\mathcal{M}_0}(R \circ \Gamma^{i-1}(\mathbf{q}, \mathbf{v})). \end{aligned}$$

To justify the second equality, observe that from (B.1), (4.66) together with Lemma 4.4 and the fact that $R^2 = I$ we deduce that $(\Gamma^{i-1})^{-1} \circ R = R \circ \Gamma^{i-1}$, for all $i = 1, \dots, n$. With similar logic we observe that $(\Xi_{\delta_1}^{(1)} \circ \Xi_{\delta_2}^{(2)})^{-1} \circ R = \Xi_{\delta_1}^{(1)} \circ R \circ \Gamma$ and obtain, invoking the notation in (4.66), that for $i = 1, \dots, n$ and $(\mathbf{q}_0, \mathbf{v}_0) \in X \times X$

$$\begin{aligned} (\Xi_{\delta_1}^{(1)} \circ \Xi_{\delta_2}^{(2)})^{-1} \circ R \circ \Gamma^{i-1}(\mathbf{q}_0, \mathbf{v}_0) &= \Xi_{\delta_1}^{(1)} \circ R \circ \Gamma^i(\mathbf{q}_0, \mathbf{v}_0) = \Xi_{\delta_1}^{(1)} \circ R(\mathbf{q}_i, \mathbf{v}_i) \\ &= (\mathbf{q}_i, -\mathbf{v}_i - \delta_1 f(\mathbf{q}_i)). \end{aligned}$$

Therefore, with these observations (B.10), (B.7) and then (B.6), we obtain for $\mathcal{M}_0$ -a.e. $(\mathbf{q}_0, \mathbf{v}_0) \in X \times X$

$$\begin{aligned} &\frac{d(R \circ \hat{S})^* \mathcal{M}_0}{d\mathcal{M}_0}(\mathbf{q}_0, \mathbf{v}_0) \\ &= \prod_{i=1}^n \frac{d(\Xi_{\delta_1}^{(1)})^* \mathcal{M}_0}{d\mathcal{M}_0}(\Xi_{\delta_1}^{(1)} \circ R \circ \Gamma^i(\mathbf{q}_0, \mathbf{v}_0)) \frac{d(\Xi_{\delta_1}^{(1)})^* \mathcal{M}_0}{d\mathcal{M}_0}(R \circ \Gamma^{i-1}(\mathbf{q}_0, \mathbf{v}_0)) \\ &= \prod_{i=1}^n \exp\left(-\delta_1 \langle C^{-1/2} f(\mathbf{q}_i), -C^{-1/2}(\mathbf{v}_i + \delta_1 f(\mathbf{q}_i)) \rangle - \frac{\delta_1^2}{2} |C^{-1/2} f(\mathbf{q}_i)|^2\right) \end{aligned}$$

$$\begin{aligned}
& \times \exp\left(-\delta_1 \langle \mathcal{C}^{-1/2} f(\mathbf{q}_{i-1}), -\mathcal{C}^{-1/2} \mathbf{v}_{i-1} \rangle - \frac{\delta_1^2}{2} |\mathcal{C}^{-1/2} f(\mathbf{q}_{i-1})|^2\right) \\
& = \exp\left(\frac{\delta_1^2}{2} |\mathcal{C}^{-1/2} f(\mathbf{q}_n)|^2 - \frac{\delta_1^2}{2} |\mathcal{C}^{-1/2} f(\mathbf{q}_0)|^2\right. \\
& \quad \left. + \delta_1 \sum_{i=1}^n [\langle \mathcal{C}^{-1/2} f(\mathbf{q}_{i-1}), \mathcal{C}^{-1/2} \mathbf{v}_{i-1} \rangle + \langle \mathcal{C}^{-1/2} f(\mathbf{q}_i), \mathcal{C}^{-1/2} \mathbf{v}_i \rangle]\right).
\end{aligned}$$

Thus, together with (B.2), this identity combines with (B.3), (B.4) to conclude (4.68). The proof is now complete.

REMARK B.1. As similarly pointed out in [9], we emphasize that the proof of Theorem 4.18 provides some additional clarification regarding the particular choice of numerical integrator for (4.51) that allowed the derivation of ∞ HMC in [10] as a well-defined algorithm in infinite dimensions. Indeed, if we choose instead the classical “leapfrog” integrator described in (4.26) within the finite-dimensional setting, and apply it to the infinite-dimensional dynamics (4.59), we obtain the following splitting:

$$\frac{d\mathbf{q}}{dt} = 0, \quad \frac{d\mathbf{v}}{dt} = -\mathbf{q} - f(\mathbf{q}) \quad \text{and} \quad \frac{d\mathbf{q}}{dt} = \mathbf{v}, \quad \frac{d\mathbf{v}}{dt} = 0.$$

The corresponding solution mappings for a given time $t \geq 0$ are given respectively by

$$\Xi_t^{(1)}(\mathbf{q}_0, \mathbf{v}_0) = (\mathbf{q}_0, \mathbf{v}_0 - t\mathbf{q}_0 - tf(\mathbf{q}_0)) \quad \text{and} \quad \Xi_t^{(2)}(\mathbf{q}_0, \mathbf{v}_0) = (\mathbf{q}_0 + t\mathbf{v}_0, \mathbf{v}_0).$$

Following the proof of Theorem 4.18, we notice that here the measures $(\Xi_{\delta_1}^{(1)})^*(\mu_0 \otimes \mu_0)$ and $(\Xi_{\delta_2}^{(2)})^*(\mu_0 \otimes \mu_0)$ can be written as

$$(\Xi_{\delta_1}^{(1)})^*(\mu_0 \otimes \mu_0)(d\mathbf{q}, d\mathbf{v}) = G(\mathbf{q}, \cdot)^* \mu_0(d\mathbf{v}) \mu_0(d\mathbf{q}) \quad \text{for } G(\mathbf{q}, \mathbf{v}) := \mathbf{v} - t\mathbf{q} - tf(\mathbf{q})$$

and

$$(\Xi_{\delta_2}^{(2)})^*(\mu_0 \otimes \mu_0)(d\mathbf{q}, d\mathbf{v}) = F(\cdot, \mathbf{v})^* \mu_0(d\mathbf{q}) \mu_0(d\mathbf{v}) \quad \text{for } F(\mathbf{q}, \mathbf{v}) := \mathbf{q} + t\mathbf{v}.$$

Now, since $\mu_0 = \mathcal{N}(0, \mathcal{C})$, it follows that $G(\mathbf{q}, \cdot)^* \mu_0 = \mathcal{N}(-t\mathbf{q} - tf(\mathbf{q}), \mathcal{C})$ and $F(\cdot, \mathbf{v})^* \mu_0 = \mathcal{N}(t\mathbf{v}, \mathcal{C})$. However, $|\mathcal{C}^{-1/2}(\mathbf{q} + f(\mathbf{q}))|$ and $|\mathcal{C}^{-1/2}\mathbf{v}|$ are infinite for μ_0 -a.e. $\mathbf{q} \in X$ and μ_0 -a.e. $\mathbf{v} \in X$, respectively, and thus $-t\mathbf{q} - tf(\mathbf{q}) \notin \text{Dom}(\mathcal{C}^{-1/2})$ and $t\mathbf{v} \notin \text{Dom}(\mathcal{C}^{-1/2})$ almost surely. This implies that the Cameron–Martin formula (see [23], Theorem 2.23) cannot be applied as done in (B.6), a crucial step for the remainder of the proof.

In summary, this shows that the validity of Theorem 4.18 relies on utilizing maps $\Xi_t^{(1)}$ and $\Xi_t^{(2)}$ for which their corresponding pushforwards on the product of Gaussians, namely $(\Xi_{\delta_1}^{(1)})^*(\mu_0 \otimes \mu_0)$ and $(\Xi_t^{(2)})^*(\mu_0 \otimes \mu_0)$, with $\mu_0 = \mathcal{N}(0, \mathcal{C})$, are absolutely continuous with respect to the product $\mu_0 \otimes \mu_0$. In Theorem 4.18, this is achieved with a “rotation” mapping, $\Xi_t^{(2)}$ as in (4.64), and a mapping given by a “shift” within $\text{Dom}(\mathcal{C}^{-1/2})$, $\Xi_t^{(1)}$ as in (4.63).

APPENDIX C: PROOF OF COROLLARY 4.19

Notice that for each fixed $\mathbf{q} \in X$, the mapping $F(\mathbf{q}, \cdot) : X \rightarrow X$ defined in (4.72) is invertible, with

$$(C.1) \quad F(\mathbf{q}, \cdot)^{-1}(\tilde{\mathbf{q}}) = \frac{1}{\sqrt{1 - \rho^2}}(\tilde{\mathbf{q}} - \rho\mathbf{q}) + \frac{\sqrt{\delta}}{2}f(\mathbf{q}).$$

Now, in view of (3.7), we compute

$$\begin{aligned}
 & F(F(\mathbf{q}, \mathbf{v}), \cdot)^{-1}(\mathbf{q}) \\
 &= \frac{1}{\sqrt{1-\rho^2}}(\mathbf{q} - \rho F(\mathbf{q}, \mathbf{v})) + \frac{\sqrt{\delta}}{2} f(F(\mathbf{q}, \mathbf{v})) \\
 (C.2) \quad &= \frac{1}{\sqrt{1-\rho^2}} \left[\mathbf{q} - \rho \left(\rho \mathbf{q} + \sqrt{1-\rho^2} \left(\mathbf{v} - \frac{\sqrt{\delta}}{2} f(\mathbf{q}) \right) \right) \right] + \frac{\sqrt{\delta}}{2} f(F(\mathbf{q}, \mathbf{v})) \\
 &= \sqrt{1-\rho^2} \mathbf{q} - \rho \mathbf{v} + \rho \frac{\sqrt{\delta}}{2} f(\mathbf{q}) + \frac{\sqrt{\delta}}{2} f(F(\mathbf{q}, \mathbf{v})).
 \end{aligned}$$

Thus, it follows by Proposition 3.3 or direct calculation that

$$S(\mathbf{q}, \mathbf{v}) = (F(\mathbf{q}, \mathbf{v}), F(F(\mathbf{q}, \mathbf{v}), \cdot)^{-1}(\mathbf{q}))$$

is an involution in $X \times X$.

Moreover, denote $\hat{S} = R \circ S$, where R is the “momentum-flip” map (4.28), so that $S = R \circ \hat{S}$. Following an observation from [9], we notice that such $\hat{S}$ is the particular one-step case of (4.65) regarding Algorithm 4.5 for specific choices of δ_1, δ_2 . Namely, taking $\delta_1 := \sqrt{\delta}/2$ and $\delta_2 := \cos^{-1} \rho$, we have

$$(C.3) \quad S = R \circ \hat{S}, \quad \hat{S} = \Xi_{\delta_1}^{(1)} \circ \Xi_{\delta_2}^{(2)} \circ \Xi_{\delta_1}^{(1)},$$

where $\Xi_t^{(1)}$ and $\Xi_t^{(2)}$ are as defined in (4.63) and (4.64), respectively. It thus follows from (4.68) with $\mathcal{M} = \mu \otimes \mu_0$ that for $(\mu \otimes \mu_0)$ -a.e. $(\mathbf{q}, \mathbf{v}) \in X \times X$

$$\begin{aligned}
 (C.4) \quad \frac{dS^*(\mu \otimes \mu_0)}{d(\mu \otimes \mu_0)}(\mathbf{q}, \mathbf{v}) &= \exp \left(\Phi(\mathbf{q}) - \Phi(\tilde{\mathbf{q}}) - \frac{\delta}{8} [|\mathcal{C}^{-1/2} f(\mathbf{q})|^2 - |\mathcal{C}^{-1/2} f(\tilde{\mathbf{q}})|^2] \right. \\
 &\quad \left. + \frac{\sqrt{\delta}}{2} [|\mathcal{C}^{-1/2} \mathbf{v}, \mathcal{C}^{-1/2} f(\mathbf{q})| + |\mathcal{C}^{-1/2} \tilde{\mathbf{v}}, \mathcal{C}^{-1/2} f(\tilde{\mathbf{q}})|] \right),
 \end{aligned}$$

where $(\tilde{\mathbf{q}}, \tilde{\mathbf{v}}) = \Xi_{\delta_1}^{(1)} \circ \Xi_{\delta_2}^{(2)} \circ \Xi_{\delta_1}^{(1)}(\mathbf{q}, \mathbf{v}) = R \circ S(\mathbf{q}, \mathbf{v})$, so that

$$(C.5) \quad \tilde{\mathbf{q}} = F(\mathbf{q}, \mathbf{v}) \quad \text{and} \quad \tilde{\mathbf{v}} = -F(F(\mathbf{q}, \mathbf{v}), \cdot)^{-1}(\mathbf{q}),$$

with F as defined in (4.72). Plugging $\tilde{\mathbf{v}}$ as given in (C.5) and (C.2) into (C.4), we obtain after rearranging terms that for $(\mu \otimes \mu_0)$ -a.e. $(\mathbf{q}, \mathbf{v}) \in X \times X$

$$\begin{aligned}
 (C.6) \quad & \frac{dS^*(\mu \otimes \mu_0)}{d(\mu \otimes \mu_0)}(\mathbf{q}, \mathbf{v}) \\
 &= \exp \left(\Phi(\mathbf{q}) - \Phi(\tilde{\mathbf{q}}) - \frac{\delta}{8} [|\mathcal{C}^{-1/2} f(\mathbf{q})|^2 + |\mathcal{C}^{-1/2} f(\tilde{\mathbf{q}})|^2] + \frac{\sqrt{\delta}}{2} \langle \mathcal{C}^{-1/2} \mathbf{v}, \mathcal{C}^{-1/2} f(\mathbf{q}) \rangle \right. \\
 &\quad \left. - \frac{\sqrt{\delta}}{2} \left\langle \mathcal{C}^{-1/2} (\sqrt{1-\rho^2} \mathbf{q} - \rho \mathbf{v}) + \rho \frac{\sqrt{\delta}}{2} \mathcal{C}^{-1/2} f(\mathbf{q}), \mathcal{C}^{-1/2} f(\tilde{\mathbf{q}}) \right\rangle \right).
 \end{aligned}$$

In view of (3.12), we plug $\mathbf{v} = F(\mathbf{q}, \cdot)^{-1}(\tilde{\mathbf{q}})$ as given in (C.1) into (C.6) to obtain that

$$(C.7) \quad \alpha(\mathbf{q}, \tilde{\mathbf{q}}) := 1 \wedge \frac{dS^*(\mu \otimes \mu_0)}{d(\mu \otimes \mu_0)}(\mathbf{q}, F(\mathbf{q}, \cdot)^{-1}(\tilde{\mathbf{q}})) = 1 \wedge \frac{\beta(\tilde{\mathbf{q}}, \mathbf{q})}{\beta(\mathbf{q}, \tilde{\mathbf{q}})},$$

where

$$\beta(\mathbf{q}, \tilde{\mathbf{q}}) := \exp\left(-\Phi(\mathbf{q}) - \frac{\delta}{8}|C^{-1/2}f(\mathbf{q})|^2 - \frac{\sqrt{\delta}}{2}\left\langle \frac{C^{-1/2}(\tilde{\mathbf{q}} - \rho\mathbf{q})}{\sqrt{1-\rho^2}}, C^{-1/2}f(\mathbf{q}) \right\rangle\right),$$

for η -a.e. $(\mathbf{q}, \tilde{\mathbf{q}}) \in X \times X$, where we recall that $\eta(d\mathbf{q}, d\tilde{\mathbf{q}}) = F(\mathbf{q}, \cdot)^* \mu_0(d\tilde{\mathbf{q}}) \mu(d\mathbf{q})$. This completes the proof. We notice that formula (C.7) for the accept–reject function concurs with [11, 22] in the particular case $f = CD\Phi$.

Acknowledgments. This work was partially conceived over the course of two research visits: by CFM to New Orleans, LA in January 2020 and by NEGH and CFM to Blacksburg, VA in February 2020. We are grateful to the Tulane and Virginia Tech math departments for hosting these productive visits.

We would like to thank a number of our colleagues for extensive and invaluable feedback surrounding this work: Geordie Richards for additional references concerning recent developments in the study of Gibbs measures for dispersive PDEs, Andrew Holbrook for extensive references from the machine learning and statistics literature and specifically for pointing us to the Metropolis–Hastings–Green algorithm as well as a body of recent work on surrogate trajectory methods, and Gideon Simpson for suggesting the extension of Theorem 2.1 to include a more general proposal space Y in the formulation found in the second draft of this manuscript. Finally we would like to thank Radford Neal, Michael Betancourt, Shiwei Lan and Christophe Andrieu for further inspiring conversations that helped us contextualize this work.

Funding. Our efforts are supported under the Grants DMS-1816551 and NSF-DMS-2108790 (NEGH, DMS-2009859 (CFM), and NSF-DMS-2108791 (JAK).

REFERENCES

- [1] ALIPRANTIS, C. D. and BORDER, K. C. (2013). *Infinite-Dimensional Analysis: A Hitchhiker’s Guide*, 2nd ed. Springer, Berlin. [MR1717083](#) <https://doi.org/10.1007/978-3-662-03961-8>
- [2] AMBROSIO, L., GIGLI, N. and SAVARÉ, G. (2008). *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, 2nd ed. *Lectures in Mathematics ETH Zürich*. Birkhäuser, Basel. [MR2401600](#)
- [3] ANDRIEU, C., LEE, A. and LIVINGSTONE, S. (2020). A general perspective on the Metropolis–Hastings kernel. Preprint. Available at [arXiv:2012.14881](#).
- [4] ARNOLD, V. I. (2013). *Mathematical Methods of Classical Mechanics. Graduate Texts in Mathematics* **60**. Springer, New York. [MR0690288](#)
- [5] ATCHADÉ, Y. F., FORT, G. and MOULINES, E. (2017). On perturbed proximal gradient algorithms. *J. Mach. Learn. Res.* **18** Paper No. 10, 33 pp. [MR3634877](#)
- [6] BÉNYI, Á., OH, T. and POCOVNICU, O. (2019). On the probabilistic Cauchy theory for nonlinear dispersive PDEs. In *Landscapes of Time-Frequency Analysis. Appl. Numer. Harmon. Anal.* 1–32. Birkhäuser/Springer, Cham. [MR3889875](#)
- [7] BESAG, J. E. (1994). Comments on “Representations of knowledge in complex systems” by U. Grenander and M. I. Miller. *J. Roy. Statist. Soc. Ser. B* **56** 591–592.
- [8] BESKOS, A., GIROLAMI, M., LAN, S., FARRELL, P. E. and STUART, A. M. (2017). Geometric MCMC for infinite-dimensional inverse problems. *J. Comput. Phys.* **335** 327–351. [MR3612501](#) <https://doi.org/10.1016/j.jcp.2016.12.041>
- [9] BESKOS, A., KALOGEROPOULOS, K. and PAZOS, E. (2013). Advanced MCMC methods for sampling on diffusion pathspace. *Stochastic Process. Appl.* **123** 1415–1453. [MR3016228](#) <https://doi.org/10.1016/j.spa.2012.12.001>
- [10] BESKOS, A., PINSKI, F. J., SANZ-SERNA, J. M. and STUART, A. M. (2011). Hybrid Monte Carlo on Hilbert spaces. *Stochastic Process. Appl.* **121** 2201–2230. [MR2822774](#) <https://doi.org/10.1016/j.spa.2011.06.003>
- [11] BESKOS, A., ROBERTS, G., STUART, A. and VOSS, J. (2008). MCMC methods for diffusion bridges. *Stoch. Dyn.* **8** 319–350. [MR2444507](#) <https://doi.org/10.1142/S0219493708002378>

- [12] BETANCOURT, M. (2019). The convergence of Markov chain Monte Carlo methods: From the Metropolis method to Hamiltonian Monte Carlo. *Ann. Phys.* **531** 1700214, 6 pp. MR3925439 <https://doi.org/10.1002/andp.201700214>
- [13] BOGACHEV, V. I. (1998). *Gaussian Measures. Mathematical Surveys and Monographs* **62**. Amer. Math. Soc., Providence, RI. MR1642391 <https://doi.org/10.1090/surv/062>
- [14] BOGACHEV, V. I. (2007). *Measure Theory. Vol. I, II*. Springer, Berlin. MR2267655 <https://doi.org/10.1007/978-3-540-34514-5>
- [15] BORGGAARD, J., GLATT-HOLTZ, N. and KROMETIS, J. (2020). A Bayesian approach to estimating background flows from a passive scalar. *SIAM/ASA J. Uncertain. Quantificat.* **8** 1036–1060. MR4133486 <https://doi.org/10.1137/19M1267544>
- [16] BOU-RABEE, N. and EBERLE, A. (2021). Two-scale coupling for preconditioned Hamiltonian Monte Carlo in infinite dimensions. *Stoch. Partial Differ. Equ. Anal. Comput.* **9** 207–242. MR4218791 <https://doi.org/10.1007/s40072-020-00175-6>
- [17] BOU-RABEE, N. and SANZ-SERNA, J. M. (2018). Geometric integrators and the Hamiltonian Monte Carlo method. *Acta Numer.* **27** 113–206. MR3826507 <https://doi.org/10.1017/s0962492917000101>
- [18] BOUCHARD-CÔTÉ, A., VOLLMER, S. J. and DOUCET, A. (2018). The bouncy particle sampler: A non-reversible rejection-free Markov chain Monte Carlo method. *J. Amer. Statist. Assoc.* **113** 855–867. MR3832232 <https://doi.org/10.1080/01621459.2017.1294075>
- [19] BOURGAIN, J. (1994). Periodic nonlinear Schrödinger equation and invariant measures. *Comm. Math. Phys.* **166** 1–26. MR1309539
- [20] BUI-THANH, T. and GHATTAS, O. (2014). An analysis of infinite dimensional Bayesian inverse shape acoustic scattering and its numerical approximation. *SIAM/ASA J. Uncertain. Quantificat.* **2** 203–222. MR3283906 <https://doi.org/10.1137/120894877>
- [21] BUI-THANH, T. and NGUYEN, Q. P. (2016). FEM-based discretization-invariant MCMC methods for PDE-constrained Bayesian inverse problems. *Inverse Probl. Imaging* **10** 943–975. MR3610747 <https://doi.org/10.3934/ipi.2016028>
- [22] COTTER, S. L., ROBERTS, G. O., STUART, A. M. and WHITE, D. (2013). MCMC methods for functions: Modifying old algorithms to make them faster. *Statist. Sci.* **28** 424–446. MR3135540 <https://doi.org/10.1214/13-STS421>
- [23] DA PRATO, G. and ZABCZYK, J. (2014). *Stochastic Equations in Infinite Dimensions*, 2nd ed. *Encyclopedia of Mathematics and Its Applications* **152**. Cambridge Univ. Press, Cambridge. MR3236753 <https://doi.org/10.1017/CBO9781107295513>
- [24] DASHTI, M. and STUART, A. M. (2017). The Bayesian approach to inverse problems. In *Handbook of Uncertainty Quantification. Vol. 1, 2*, 3 311–428. Springer, Cham. MR3839555
- [25] DUANE, S., KENNEDY, A. D., PENDLETON, B. J. and ROWETH, D. (1987). Hybrid Monte Carlo. *Phys. Lett. B* **195** 216–222. MR3960671 [https://doi.org/10.1016/0370-2693\(87\)91197-x](https://doi.org/10.1016/0370-2693(87)91197-x)
- [26] EBERLE, A. (2014). Error bounds for Metropolis–Hastings algorithms applied to perturbations of Gaussian measures in high dimensions. *Ann. Appl. Probab.* **24** 337–377. MR3161650 <https://doi.org/10.1214/13-AAP926>
- [27] FANG, Y., SANZ-SERNA, J. M. and SKEEL, R. D. (2014). Compressible generalized hybrid Monte Carlo. *J. Chem. Phys.* **140** 174108.
- [28] FOLLAND, G. B. (1999). *Real Analysis: Modern Techniques and Their Applications*, 2nd ed. *Pure and Applied Mathematics (New York)*. Wiley, New York. MR1681462
- [29] GELMAN, A., CARLIN, J. B., STERN, H. S., DUNSON, D. B., VEHTARI, A. and RUBIN, D. B. (2014). *Bayesian Data Analysis*, 3rd ed. *Texts in Statistical Science Series*. CRC Press, Boca Raton, FL. MR3235677
- [30] GELMAN, A., LEE, D. and GUO, J. (2015). Stan: A probabilistic programming language for Bayesian inference and optimization. *J. Educ. Behav. Stat.* **40** 530–543.
- [31] GEYER, C. J. (2003). The Metropolis–Hastings–Green algorithm.
- [32] GEYER, C. J. (2011). Introduction to Markov chain Monte Carlo. In *Handbook of Markov Chain Monte Carlo. Chapman & Hall/CRC Handb. Mod. Stat. Methods* 3–48. CRC Press, Boca Raton, FL. MR2858443
- [33] GIROLAMI, M. and CALDERHEAD, B. (2011). Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 123–214. MR2814492 <https://doi.org/10.1111/j.1467-9868.2010.00765.x>
- [34] GLATT-HOLTZ, N., KROMETIS, J. and MONDAINI, C. (2023). A reduced order modeling approach to Hamiltonian Monte Carlo sampling for infinite-dimensional problems. To appear.
- [35] GLATT-HOLTZ, N. E. and MONDAINI, C. F. (2022). Mixing rates for Hamiltonian Monte Carlo algorithms in finite and infinite dimensions. *Stoch. Partial Differ. Equ. Anal. Comput.* **10** 1318–1391. MR4503169 <https://doi.org/10.1007/s40072-021-00211-z>

- [36] GREEN, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82** 711–732. [MR1380810](#) <https://doi.org/10.1093/biomet/82.4.711>
- [37] HAIRER, E., LUBICH, C. and WANNER, G. (2006). *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*, 2nd ed. *Springer Series in Computational Mathematics* **31**. Springer, Berlin. [MR2221614](#)
- [38] HAIRER, M., STUART, A. and VOSS, J. (2009). Sampling conditioned diffusions. In *Trends in Stochastic Analysis. London Mathematical Society Lecture Note Series* **353** 159–185. Cambridge Univ. Press, Cambridge. [MR2562154](#)
- [39] HAIRER, M., STUART, A. and VOSS, J. (2011). Signal processing problems on function space: Bayesian formulation, stochastic PDEs and effective MCMC methods. In *The Oxford Handbook of Nonlinear Filtering* 833–873. Oxford Univ. Press, Oxford. [MR2884617](#)
- [40] HAIRER, M., STUART, A. M. and VOLLMER, S. J. (2014). Spectral gaps for a Metropolis–Hastings algorithm in infinite dimensions. *Ann. Appl. Probab.* **24** 2455–2490. [MR3262508](#) <https://doi.org/10.1214/13-AAP982>
- [41] HAIRER, M., STUART, A. M. and VOSS, J. (2007). Analysis of SPDEs arising in path sampling. II. The nonlinear case. *Ann. Appl. Probab.* **17** 1657–1706. [MR2358638](#) <https://doi.org/10.1214/07-AAP441>
- [42] HAIRER, M., STUART, A. M., VOSS, J. and WIBERG, P. (2005). Analysis of SPDEs arising in path sampling. I. The Gaussian case. *Commun. Math. Sci.* **3** 587–603. [MR2188686](#)
- [43] HASTINGS, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57** 97–109. [MR3363437](#) <https://doi.org/10.1093/biomet/57.1.97>
- [44] HOFFMAN, M. D. and GELMAN, A. (2014). The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* **15** 1593–1623. [MR3214779](#)
- [45] JOSÉ, J. V. and SALETAN, E. J. (1998). *Classical Dynamics: A Contemporary Approach*. Cambridge Univ. Press, Cambridge. [MR1640663](#) <https://doi.org/10.1017/CBO9780511803772>
- [46] KAIPIO, J. and SOMERSALO, E. (2005). *Statistical and Computational Inverse Problems. Applied Mathematical Sciences* **160**. Springer, New York. [MR2102218](#)
- [47] LAN, S., BUI-THANH, T., CHRISTIE, M. and GIROLAMI, M. (2016). Emulation of higher-order tensors in manifold Monte Carlo methods for Bayesian inverse problems. *J. Comput. Phys.* **308** 81–101. [MR3448239](#) <https://doi.org/10.1016/j.jcp.2015.12.032>
- [48] LEIMKUHLE, B. and REICH, S. (2004). *Simulating Hamiltonian Dynamics. Cambridge Monographs on Applied and Computational Mathematics* **14**. Cambridge Univ. Press, Cambridge. [MR2132573](#)
- [49] LEVY, D., HOFFMAN, M. D. and SOHL-DICKSTEIN, J. (2017). Generalizing Hamiltonian Monte Carlo with neural networks. Preprint. Available at [arXiv:1711.09268](#).
- [50] LI, L., HOLBROOK, A., SHAHBABA, B. and BALDI, P. (2019). Neural network gradient Hamiltonian Monte Carlo. *Comput. Statist.* **34** 281–299. [MR3920582](#) <https://doi.org/10.1007/s00180-018-00861-z>
- [51] LIU, J. S. (2008). *Monte Carlo Strategies in Scientific Computing. Springer Series in Statistics*. Springer, New York. [MR2401592](#)
- [52] LIU, J. S., LIANG, F. and WONG, W. H. (2000). The multiple-try method and local optimization in Metropolis sampling. *J. Amer. Statist. Assoc.* **95** 121–134. [MR1803145](#) <https://doi.org/10.2307/2669532>
- [53] LU, X., PERRONE, V., HASENCLEVER, L., TEH, Y. W. and VOLLMER, S. (2017). Relativistic Monte Carlo. In *Artificial Intelligence and Statistics* 1236–1245. PMLR.
- [54] MARSDEN, J. E. and RATIU, T. S. (1995). Introduction to mechanics and symmetry. *Phys. Today* **48** 65.
- [55] MARTIN, J., WILCOX, L. C., BURSTEDDE, C. and GHATTAS, O. (2012). A stochastic Newton MCMC method for large-scale statistical inverse problems with application to seismic inversion. *SIAM J. Sci. Comput.* **34** A1460–A1487. [MR2970260](#) <https://doi.org/10.1137/110845598>
- [56] MEEDS, E. and WELLING, M. (2014). GPS-ABC: Gaussian process surrogate approximate Bayesian computation. Preprint. Available at [arXiv:1401.2838](#).
- [57] METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. and TELLER, E. (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21** 1087–1092.
- [58] NAHMOD, A. R. and STAFFILANI, G. (2019). Randomness and nonlinear evolution equations. *Acta Math. Sin. (Engl. Ser.)* **35** 903–932. [MR3952697](#) <https://doi.org/10.1007/s10114-019-8297-5>
- [59] NEAL, R. M. (1993). *Probabilistic Inference Using Markov Chain Monte Carlo Methods*. Department of Computer Science, Univ. Toronto, ON, Canada.
- [60] NEAL, R. M. (1999). Regression and classification using Gaussian process priors. In *Bayesian Statistics, 6 (Alcoceber)* 475–501. Oxford Univ. Press, New York. [MR1723510](#)
- [61] NEAL, R. M. (2011). MCMC using Hamiltonian dynamics. In *Handbook of Markov Chain Monte Carlo. Chapman & Hall/CRC Handb. Mod. Stat. Methods* 113–162. CRC Press, Boca Raton, FL. [MR2858447](#)
- [62] NEKLYUDOV, K., WELLING, M., EGOROV, E. and VETROV, D. (2020). Involutive MCMC: A unifying framework. In *International Conference on Machine Learning* 7273–7282. PMLR.

- [63] PETRA, N., MARTIN, J., STADLER, G. and GHATTAS, O. (2014). A computational framework for infinite-dimensional Bayesian inverse problems, Part II: Stochastic Newton MCMC with application to ice sheet flow inverse problems. *SIAM J. Sci. Comput.* **36** A1525–A1555. [MR3233941](#) <https://doi.org/10.1137/130934805>
- [64] RADIVOJEVIĆ, T. and AKHMATSKAYA, E. (2020). Modified Hamiltonian Monte Carlo for Bayesian inference. *Stat. Comput.* **30** 377–404. [MR4064627](#) <https://doi.org/10.1007/s11222-019-09885-x>
- [65] RASMUSSEN, C. E. (2003). Gaussian processes to speed up hybrid Monte Carlo for expensive Bayesian integrals. In *Bayesian Statistics, 7 (Tenerife, 2002)* 651–660. Oxford Univ. Press, New York. [MR2003529](#)
- [66] REZNIKOFF, M. G. and VANDEN-EIJNDEN, E. (2005). Invariant measures of stochastic partial differential equations and conditioned diffusions. *C. R. Math. Acad. Sci. Paris* **340** 305–308. [MR2121896](#) <https://doi.org/10.1016/j.crma.2004.12.025>
- [67] ROBERT, C. P. and CASELLA, G. (2013). *Monte Carlo Statistical Methods. Springer Texts in Statistics.* Springer, New York. [MR1707311](#) <https://doi.org/10.1007/978-1-4757-3071-5>
- [68] ROBERTS, G. O. and TWEEDIE, R. L. (1996). Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli* **2** 341–363. [MR1440273](#) <https://doi.org/10.2307/3318418>
- [69] SILVESTER, J. R. (2000). Determinants of block matrices. *Math. Gaz.* **84** 460–467.
- [70] STUART, A. M. (2010). Inverse problems: A Bayesian perspective. *Acta Numer.* **19** 451–559. [MR2652785](#) <https://doi.org/10.1017/S0962492910000061>
- [71] TEAM, S. D. (2016). Stan modeling language users guide and reference manual. Technical report.
- [72] TIERNEY, L. (1998). A note on Metropolis–Hastings kernels for general state spaces. *Ann. Appl. Probab.* **8** 1–9. [MR1620401](#) <https://doi.org/10.1214/aoap/1027961031>
- [73] TRIPURANENI, N., ROWLAND, M., GHAHRAMANI, Z. and TURNER, R. (2017). Magnetic Hamiltonian Monte Carlo. In *International Conference on Machine Learning* 3453–3461.
- [74] TU, L. W. (2011). *An Introduction to Manifolds*, 2nd ed. *Universitext.* Springer, New York. [MR2723362](#) <https://doi.org/10.1007/978-1-4419-7400-6>
- [75] ZHANG, C., SHAHBABA, B. and ZHAO, H. (2017). Precomputing strategy for Hamiltonian Monte Carlo method based on regularity in parameter space. *Comput. Statist.* **32** 253–279. [MR3604217](#) <https://doi.org/10.1007/s00180-016-0683-1>
- [76] ZHANG, C., SHAHBABA, B. and ZHAO, H. (2017). Hamiltonian Monte Carlo acceleration using surrogate functions with random bases. *Stat. Comput.* **27** 1473–1490. [MR3687321](#) <https://doi.org/10.1007/s11222-016-9699-1>