Several Remarks on the Numerical Integrator in Lagrangian Monte Carlo

James A. Brofos*
Department of Statistics and Data Science, Yale University

and

Roy R. Lederman
Department of Statistics and Data Science, Yale University

March 1, 2022

Abstract

Riemannian manifold Hamiltonian Monte Carlo (RMHMC) is a powerful method of Bayesian inference that exploits underlying geometric information of the posterior distribution in order to efficiently traverse the parameter space. However, the form of the Hamiltonian necessitates complicated numerical integrators, such as the generalized leapfrog method, that preserve the detailed balance condition. The distinguishing feature of these numerical integrators is that they involve solutions to implicitly defined equations. Lagrangian Monte Carlo (LMC) proposes to eliminate the fixed point iterations by transitioning from the Hamiltonian formalism to Lagrangian dynamics, wherein a fully explicit integrator is available. This work makes several contributions regarding the numerical integrator used in LMC. First, it has been claimed in the literature that the integrator is only first-order accurate for the Lagrangian equations of motion; to the contrary, we show that the LMC integrator enjoys second order accuracy. Second, the current conception of LMC requires four determinant computations in every step in order to maintain detailed balance; we propose a simple modification to the integration procedure in LMC in order to reduce the number of determinant computations from four to two while still retaining a fully explicit numerical integration scheme. Third, we demonstrate that the LMC integrator enjoys a certain robustness to human error that is not shared with the generalized leapfrog integrator, which can invalidate detailed balance in the latter case. We discuss these contributions within the context of several benchmark Bayesian inference tasks.

*The authors gratefully acknowledge the Yale Center for Research Computing for use of the research computing infrastructure. We thank Marcus A. Brubaker for helpful discussions. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. 1752134. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors(s) and do not necessarily reflect the views of the National Science Foundation. The work is also supported in part by NIH/NIGMS 1R01GM136780-01 and AFSOR FA9550-21-1-0317.
Keywords: Hamiltonian Markov Chain Monte Carlo Bayesian Posterior
1 Introduction

Let $\mathcal{L} : \mathbb{R}^m \to \mathbb{R}$ be the log-density of a smooth distribution, known up to an additive constant. A critical problem in Bayesian inference is the design of procedures that generate samples from a target density $\pi(q) \propto \exp(\mathcal{L}(q))$. Hamiltonian Monte Carlo (HMC) (Duane et al., 1987; Betancourt, 2017) is a Markov chain Monte Carlo (MCMC) method for sampling from arbitrary differentiable probability distributions based on numerical solutions to Hamilton’s equations of motion. Unlike random walk Metropolis or the Metropolis-adjusted Langevin algorithm, HMC can propose distant states, thereby dramatically decreasing the autocorrelation between states of the chain and increasing sampling efficiency.

The efficiency of HMC can be further improved by incorporating geometric concepts into the proposal mechanism. Information geometry (Amari, 2016) provides a framework for representing sets of probability densities as a Riemannian manifold whose metric can be chosen as the Fisher information matrix. When $\mathcal{L}$ decomposes into the sum of a log-likelihood and a log-prior terms, Girolami and Calderhead (2011) proposed Riemannian manifold Hamiltonian Monte Carlo (RMHMC), which sought to precondition Hamiltonian dynamics with the inverse of the sum of Fisher information of the log-likelihood and the negative Hessian of the log-prior. In general, both the Fisher information and the Hessian of the log-prior depend on the parameter, denoted here by $q$, to be sampled; this dependency produces a complicated Hamiltonian. In order to maintain the detailed balance condition of the RMHMC Markov chain, elaborate numerical integrators such as the generalized leapfrog or implicit midpoint methods must be used (Pourzanjani and Petzold, 2019; Girolami and Calderhead, 2011; Brofos and Lederman, 2021a). These numerical integrators involve implicitly defined updates, wherein updates are defined as the solution to a fixed point equation, which are typically resolved to a prescribed convergence tolerance by fixed point iteration.

Lan et al. (2015) introduced Lagrangian Monte Carlo (LMC) as an alternative to RMHMC. Unlike RMHMC, which is based on numerical solutions to Hamiltonian mechanics, LMC is instead inspired by the Lagrangian formalism of classical mechanics. Although Lagrangian and Hamiltonian mechanics are formally equivalent from a physical perspective,
methods of numerical integration assume a simplified form when expressed as Lagrangian mechanics. Indeed, the simplification is so significant that Lan et al. (2015) was able to devise a fully explicit numerical integrator for which detailed balance could be maintained when used as a proposal operator in MCMC. However, the elimination of fixed point iterations was replaced by the requirement that four Jacobian determinants be computed in a single step of the explicit integrator. Moreover, it was stated that the explicit method had only first-order accuracy as a numerical integrator, supposedly producing less accurate solutions than the generalized leapfrog method, whose accuracy is second-order.

The purpose of the present work is three-fold. First, we propose a simple mechanism by which to reduce the number of Jacobian determinant evaluations from four to two in a single step of the integrator. This is achieved by inverting the sequence in which position and velocity are integrated. Second, we will clarify that the order of the explicit integrator, with or without inversion, actually has second-order accuracy, the same as those integrators which are commonly used in RMHMC. Third, we discuss how LMC enjoys a greater robustness to human error than RMHMC. The outline of the remainder of this paper is as follows. In section 2 we discuss preliminary material on Hamiltonian and Lagrangian mechanics, numerical integrators, and MCMC methods based off of these physical models. In section 3 we review some related work in the literature on integration methods for RMHMC. In section 4 we proceed to our analysis of the LMC integrator, where we describe how one may reduce the number of Jacobian determinant computations while maintaining a fully explicit integration method and give a proof that the integrator has second-order accuracy. In section 5 we turn to the evaluation of the proposed modifications to LMC; we evaluate performance on several benchmark Bayesian inference tasks and give numerical evidence to support the claim of second-order accuracy.

2 Preliminaries

In this section we review the necessary background for Hamiltonian and Lagrangian Monte Carlo. Section 2.2 reviews the most important concepts from RMHMC and LMC, giving perspective on the varieties of integrators and how they are employed in a Markov chain
Monte Carlo procedure. We then proceed in appendix [B.1] to recall the Hamiltonian and Lagrangian formalisms from physics. Appendix [B.2] then treats the matter of numerical integration of the Hamiltonian and Lagrangian mechanics. In appendix [B.3] we then review Hamiltonian and Lagrangian Monte Carlo using the framework of involutive Monte Carlo, wherein we give special attention to the Jacobian determinant computations that are necessitated in the Lagrangian construction.

2.1 Notation

In the context of classical mechanics, we denote by $q$ the position variable, $v$ the velocity variable, $p$ the momentum variable, and $a$ the acceleration variable, all elements of $\mathbb{R}^m$. We adopt the notation $q^{(k)}$ to refer to the $k$-th element of $q$, with similar conventions being employed for $v$ and $p$. Denoting $z = (q, p)$; we call $z$ a point in phase space. We denote by $\text{PD}(m)$ the set of $m \times m$ positive definite matrices. Given a map $\Phi : \mathbb{R}^m \to \mathbb{R}^m$, we use the notation $\Phi^k$ to mean the composition $\Phi \circ \cdots \circ \Phi$ ($k$ times). We write $\text{Id}_m$ to denote the identity matrix of size $m \times m$. We denote the Borel $\sigma$-algebra on $\mathbb{R}^m$ by $\mathcal{B}(\mathbb{R}^m)$.

2.2 Lagrangian Monte Carlo: The Essentials

Lagrangian Monte Carlo (LMC) is a geometric method of Bayesian inference that seeks to incorporate second-order information about the posterior in order to produce effective proposals, similar to Riemannian manifold Hamiltonian Monte Carlo (RMHMC). In this section we review the fundamentals of these methods. For applications in HMC, an important class of Hamiltonians have the following form:

Definition 2.1. The Riemannian Hamiltonian

$$H(q, p) = U(q) + K(q, p) \tag{1}$$

where $U : \mathbb{R}^m \to \mathbb{R}$ is called the potential energy function and $K : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ is the kinetic energy function, having the form $K(q, p) = \frac{1}{2}p^\top G^{-1}(q)p$, where $G : \mathbb{R}^m \to \text{PD}(m)$ where $G$ is called the metric.
The Riemannian metric defines Christoffel symbols which convey information about the curvature and shape of $\mathbb{R}^m$ imbued with the metric $G$.

**Definition 2.2.** The Christoffel symbols are the $m^3$ functions defined by

$$
\Gamma^k_{ij}(q) = \frac{1}{2} \sum_{l=1}^{m} G^{-1}_{kl}(q) \left( \frac{\partial}{\partial q^{(i)}} G_{lj}(q) + \frac{\partial}{\partial q^{(j)}} G_{li}(q) - \frac{\partial}{\partial q^{(l)}} G_{ij}(q) \right).
$$

We note that the Christoffel symbols are symmetric in their lower indices (i.e. $\Gamma^k_{ij}(q) = \Gamma^k_{ji}(q)$). The Christoffel symbols play a prominent role in the development of Lagrangian mechanics. As a notational convenience, we will define the matrix-valued function $\Omega : \mathbb{R} \times \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}^{m \times m}$ whose $(i,j)$-th entry is

$$
\Omega_{ij}(\epsilon, q, v) = \frac{\epsilon}{2} \sum_{k=1}^{m} \Gamma^k_{ij}(q)v^{(k)}.
$$

The Riemannian Hamiltonian in definition 2.1 produces equations of motion (see appendix B.1 for details), which do not have closed-form solutions. This necessitates the use of numerical integrators. We now review two integrators that form the basis of our evaluations: the generalized leapfrog integrator, which is a reversible, volume-preserving, and second-order accurate, and the Lagrangian leapfrog method of [Lan et al. (2015)].

**Definition 2.3.** The **generalized leapfrog integrator** for the Hamiltonian equations of motion in eqs. (40) and (41) is a map $(q,p) \mapsto (\tilde{q}, \tilde{p})$ defined by,

$$
\tilde{p}^{(k)} = p^{(k)} - \frac{\epsilon}{2} \left( -\frac{1}{2} \hat{p}^{(k)} G^{-1}(q) \frac{\partial G}{\partial q^{(k)}}(q) \right) G^{-1}(q) \tilde{p} + \frac{\partial U}{\partial q^{(k)}}(q),
$$

$$
\tilde{q} = q + \frac{\epsilon}{2} \left( G^{-1}(q) \hat{p} + G^{-1}(\tilde{q}) \tilde{p} \right),
$$

$$
\tilde{p} = p - \frac{\epsilon}{2} \left( -\frac{1}{2} \hat{p} G^{-1}(q) \frac{\partial G}{\partial q^{(k)}}(q) \right) G^{-1}(q) \tilde{p} + \frac{\partial U}{\partial q^{(k)}}(q). \quad (4)
$$

Pseudo-code implementing the generalized leapfrog algorithm is given in algorithm in appendix A.

**Definition 2.4.** The **Lagrangian leapfrog integrator** for the Lagrangian equations of motion given in eq. (46) is a map $(q,v) \mapsto (\tilde{q}, \tilde{v})$ defined by,

$$
\tilde{v} = \left[ \text{Id}_m + \Omega(\epsilon, q, v) \right]^{-1} \left[ v - \frac{\epsilon}{2} G^{-1}(q) \nabla U(q) \right],
$$

$$
\tilde{q} = q + \epsilon \tilde{v},
$$

$$
\tilde{v} = \left[ \text{Id}_m + \Omega(\epsilon, \tilde{q}, \tilde{v}) \right]^{-1} \left[ \tilde{v} - \frac{\epsilon}{2} G^{-1}(\tilde{q}) \nabla U(\tilde{q}) \right].
$$

6
Pseudo-code implementing the Lagrangian leapfrog algorithm is given in algorithm [2] in appendix A. Unlike the generalized leapfrog integrator (definition [2.3]), which is a symplectic transformation and therefore necessarily volume-preserving, the Lagrangian leapfrog (definition [2.4]) is not volume-preserving. Its Jacobian determinant is computed in eq. (57).

A thorough treatment of numerical integrators is provided in appendix B. In the context of Monte Carlo, this means that Markov chains constructed from repeated applications of the Lagrangian leapfrog integrator will require a Jacobian determinant computation, whereas methods based on the generalized leapfrog will not (its Jacobian determinant is one). Such a Markov chain is the subject of the following example.

Example 1. Let $H : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ be as in definition [2.1] and define a probability density $\pi(q,p) \propto \exp(-H(q,p))$. Let $q \in \mathbb{R}^m$ be given; a single Markov chain step is constructed as follows. Sample $p \sim \text{Normal}(0, G(q))$. Fix $k \in \mathbb{N}$. We consider RMHMC and LMC separately:

**LMC** Let $\Phi_\epsilon$ denote the Lagrangian leapfrog (definition [2.4]). Compute the proposal $(\tilde{q}, \tilde{v}) = \Phi(q, G^{-1}(q)p)$ and set $\tilde{p} = G(\tilde{q})\tilde{v}$. Compute the Jacobian determinant $J$ of the map $(q,p) \mapsto (\tilde{q}, \tilde{p})$ using eq. (57).

**RMHMC** Let $\Phi_\epsilon$ denote the generalized leapfrog integrator (definition [2.3]). Compute the proposal $(\tilde{q}, \tilde{p}) = \Phi(q,p)$ and set $J = 1$.

Accept the proposal state $\tilde{q}$ with probability $\alpha((q,p), (\tilde{q}, \tilde{p}), J) = \min \left\{ 1, \frac{\pi(\tilde{q}, \tilde{p})}{\pi(q,p)} \cdot J \right\}$; otherwise remain at the current state $q$.

A more rigorous treatment of the LMC and RMHMC Markov chains is given in appendix B.3 using the framework of diffeomorphism Monte Carlo. Pseudo-code is provided in algorithm [3] in appendix A.

### 3 Related Work

The focus of the present work is to investigate the numerical methods of integration that were proposed in Lan et al. (2015). Brofos and Lederman (2021a) gave an evaluation of
the implicit midpoint integrator for RMHMC with special attention paid to the errors in reversibility and volume preservation that were produced by the implicit midpoint algorithm compared to the generalized leapfrog method, as well as the energy conservation properties enjoyed by the implicit midpoint integrator. Other mechanisms of explicit integration have been considered with applications to RMHMC foremost in mind, such as Cobb et al. (2019) which produced a reversible, volume-preserving numerical method in an expanded phase-space. Due to the expansion of phase-space, this integrator cannot be used to produce a Markov chain satisfying detailed balance. The work of Zhang and Sutton (2014) explored alternating blockwise Metropolis-within-Gibbs-like strategies with Riemannian metrics chosen to produce separable Hamiltonians within each block; each block can then be integrated using the standard leapfrog integrator. LMC has previously been criticized in the literature for having unfavorable performance in high dimensions; this failure of LMC in relation to RMHMC is discussed in Betancourt et al. (2014) and we will see evidence of this degradation in section 5.4.

4 Analytical Apparatus

In this section we describe an algorithmic recommendation for the numerical integrator used in Lagrangian Monte Carlo and we clarify certain statements around the order of this numerical method. Specifically, we show how to reduce the number of determinant computations from four to two, and that the integrator of Lagrangian dynamics has third-order local error, comparable to the error of the (generalized) leapfrog method used in HMC.

4.1 Inversion of the Integration Sequence

A disadvantage of the Lagrangian integrator is that it involves four Jacobian determinant computations at each step of the integrator. In general, computing the Jacobian determinant of an $m \times m$ matrix incurs a computational cost like $O(m^3)$. Therefore, it seems worthwhile to investigate mechanisms by which to reduce the number of these calculations.
that are required. Lan et al. (2015) proposed one method that has only two Jacobian
determinant computations, but necessitates the return to implicit methods of integration.
To retain the advantages of explicit integration, we propose a method that computes two
Jacobian determinants in every step and consists only of explicit integration steps. To
achieve this, we propose a conceptually simple procedure: invert the sequence in which
position and velocity are updated in the Lagrangian integrator so that position is updated
twice at the beginning and end of each step and velocity is updated once in between each
update to position. Formally:

Definition 4.1. The inverted Lagrangian leapfrog integrator for the Lagrangian equations of
motion given in eq. (46) is a map \((q, v) \mapsto (\tilde{q}, \tilde{v})\) defined by,

\[
\tilde{q} = q + \frac{\epsilon}{2} v \quad (10)
\]

\[
\tilde{v} = [\text{Id}_m + 2\Omega(\epsilon, \tilde{q}, v)]^{-1} \left[ v - \epsilon \mathbf{G}^{-1}(\tilde{q}) \nabla U(\tilde{q}) \right] \quad (11)
\]

\[
\tilde{q} = \tilde{q} + \frac{\epsilon}{2} \tilde{v} \quad (12)
\]

Pseudo-code for this procedure is provided in algorithm 4 in appendix A. The ba-
sic modification requires only two Jacobian determinant computations per step since the
two updates to position in eqs. (10) and (12), being shear transformations, are volume-


Approximations for this procedure is provided in algorithm 4 in appendix A. The ba-
sic modification requires only two Jacobian determinant computations per step since the
two updates to position in eqs. (10) and (12), being shear transformations, are volume-




Although this Jacobian determinant differs from that produced by algorithm 2 in ap-


Lemma 4.2. The inverted Lagrangian leapfrog integrator is self-adjoint and has at least
first-order local error.

A proof is given in appendix D.

Corollary 4.3. The inverted Lagrangian leapfrog integrator has at least second-order local
error.
Proof. This follows as an immediate corollary of lemma 4.2.

Definition 4.4. Let $\Phi_\epsilon$ be the inverted Lagrangian leapfrog integrator with step-size $\epsilon \in \mathbb{R}$ (definition 4.1). Let $k \in \mathbb{N}$ be the number of integration steps. The inverted Lagrangian Monte Carlo (ILMC) is an instance involutive Monte Carlo (definition B.16) with involution $F \circ \Phi_k^\epsilon$ where $F$ is the momentum flip operator (definition B.18).

Let us denote by $\hat{\Phi}_\epsilon$ the Lagrangian leapfrog integrator (definition 2.4) and $\tilde{\Phi}_\epsilon$ the inverted Lagrangian leapfrog (definition 4.1). Let $\Phi_\epsilon$ be the exact time $\epsilon$ solution of eq. (46).

Because both methods are second-order accurate, it follows that

$$\|\hat{\Phi}_\epsilon - \tilde{\Phi}_\epsilon\| \leq \|\hat{\Phi}_\epsilon - \Phi_\epsilon\| + \|\tilde{\Phi}_\epsilon - \Phi_\epsilon\|$$

$$= O(\epsilon^3)$$

Intuitively, in the limit of small step-sizes, the difference in proposals generated by the Lagrangian leapfrog and the inverted Lagrangian leapfrog will be minimal, but ILMC (definition 4.4) involves half the number of Jacobian determinant computations compared to LMC (definition B.20) and should therefore be preferred. On the other hand, for large step-sizes, the situation is less clear.

4.2 Aversions to Inverting the Integration Sequence

In the case when $G = \text{Id}_m$, the Lagrangian integrator devolve into the standard leapfrog integrator (definition E.1) that is ubiquitous in HMC. There are good reasons why implementations of HMC integrate in the order of an initial half-step in momentum, a full step in position, and then a second half-step of momentum. This is described visually by Bou-Rabee and Sanz-Serna (2018) for the Hamiltonian $H(q, p) = \frac{q^2}{2} + \frac{p^2}{2}$, who make the argument that the leapfrog integrator produces high acceptance probabilities along the $q$-axis, which are desirable. By contrast, inverted leapfrog (definition E.2) produces large acceptance probabilities along the $p$-axis, which are not valuable in HMC. Here we wish to expand on this example to consider what happens when HMC Markov chains based on the standard leapfrog and inverted leapfrog are initialized at stationarity. Based on the symmetric roles of $q$ and $p$ in the Hamiltonian, it is tempting to conclude that these
Markov chains would exhibit similar performance profiles; surprisingly, this is not the case as shown in the following example.

**Example 2.** Consider a Hamiltonian of the form $H(q, p) = \omega^2 q^2/2 + p^2/2$. This Hamiltonian corresponds to the distributions $q \sim \text{Normal}(0, 1/\omega^2)$ and $p \sim \text{Normal}(0, 1)$. Let $\Phi_\epsilon: \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m \times \mathbb{R}^m$ and $\tilde{\Phi}_\epsilon: \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m \times \mathbb{R}^m$ be the single step leapfrog (definition E.1) and inverted leapfrog (definition E.2) methods with step-size $\epsilon$, respectively. Let $\text{Proj}_q(q, p) = q$ be the projection onto the $q$-variables. Then,

$\mathbb{E} \left[ (\text{Proj}_q \circ \tilde{\Phi}_\epsilon(q, p) - q)^2 \right] = \frac{\epsilon^4 \omega^2}{4} + \epsilon^2$  \hspace{1cm} (16)

$\mathbb{E} \left[ (\text{Proj}_q \circ \Phi_\epsilon(q, p) - q)^2 \right] = \frac{\epsilon^4 \omega^2}{4} + \epsilon^2 \left( 1 - \frac{\epsilon^2 \omega^2}{4} \right)$.  \hspace{1cm} (17)

Both the leapfrog and inverted leapfrog methods are only numerically stable when $\epsilon^2 \omega^2 < 4$ (Leimkuhler and Reich, 2005). Hence, $\mathbb{E} \left[ (\text{Proj}_q \circ \tilde{\Phi}_\epsilon(q, p) - q)^2 \right] < \mathbb{E} \left[ (\text{Proj}_q \circ \Phi_\epsilon(q, p) - q)^2 \right]$, which we interpret to mean that, in a single step of the integrator, the expected squared distance between initial and terminal position is greater for the leapfrog integrator than for the inverted leapfrog. By deriving the propagator matrices (see appendix E) of the

Figure 1: Visualization of the difference in expected distance when integrating a trajectory of a separable, quadratic Hamiltonian for a varying number of steps and integration step-sizes. This difference is always non-negative, indicating that one expects a greater distance between position variables when integrating with the leapfrog, rather than inverted leapfrog, algorithm.
leapfrog and inverted leapfrog integrators, we may deduce exact formulas for the \( k \)-step quantities \( \mathbb{E} \left[ (\text{Proj}_q \circ \Phi^k(q,p) - q)^2 \right] \) and \( \mathbb{E} \left[ (\text{Proj}_q \circ \tilde{\Phi}^k(q,p) - q)^2 \right] \). In fig. 1 we show the difference of these expected squared distances as a function of the number of steps and for several step-sizes. Notably, this difference is always non-negative, indicating that one expects the leapfrog to produce more distant proposals than the inverted leapfrog in the Gaussian case.

Because autocorrelation is related to the distance to subsequent samples, one expects the inverted leapfrog method to exhibit fewer effective samples. This will be true even if LMC Markov chain is initialized in the stationary distribution. This illustrates an important short-coming of inverting the integration sequence, which must be balanced against computational savings obtained by the reduction in Jacobian determinant computations in the LMC setting.

4.3 The Order of the Explicit Integrators

In \cite{Lan2015}, the authors showed that the local error rate of the Lagrangian integrator is at least \( O(\epsilon^2) \). We now build on this foundation in order to deduce that the local error rate of the Lagrangian leapfrog is at least \( O(\epsilon^3) \). This means that the order of the LMC integrator matches the local and global error rates of the integrators used in HMC and RMHMC. Proofs of lemmas 4.5 and 4.6 may be found in appendix D.

**Lemma 4.5.** The explicit integrator of the Lagrangian dynamics has at least first-order local error.

**Lemma 4.6.** The explicit integrator of Lagrangian dynamics is self-adjoint.

**Proposition 4.7.** The explicit integrator of the Lagrangian dynamics has at least third-order local error.

**Proof.** From lemma 4.5 we know that \( r \geq 1 \). Suppose \( r = 1 \). From lemma 4.6 we know that the integrator is self-adjoint. From theorem B.12 we know that the order of a self-adjoint method must be even; hence \( r \) cannot be odd. But \( r = 1 \) by assumption, a contradiction.
Therefore, it must be that \( r \geq 2 \) so that the explicit integrator has, at least, third-order local error.

Denote by \( \hat{\Phi}_\epsilon \) the second-order integrator of Lagrangian dynamics. If \( \hat{\Phi}_\epsilon(q_0, v_0) = (\hat{q}_\epsilon, \hat{v}_\epsilon) \), one wonders if \((\hat{q}_\epsilon, \hat{p}_\epsilon)\) is a second-order approximation of the Hamiltonian dynamics, where \( \hat{p}_\epsilon = G(\hat{q}_\epsilon)\hat{v}_\epsilon \). Indeed, this is true and follows as an immediate consequence of proposition B.13 with diffeomorphism \((q, v) \mapsto (q, Gv)\).

4.4 Efficient Computation of the Jacobian Determinant and the Update to Velocity

The update to the velocity in eqs. (7), (9) and (11) and the associated change-in-volume in eq. (13) involve manipulations of a matrix of the form \( \text{Id} + \Omega(\epsilon, q, v) \): in the former case, we must solve a linear system involving this matrix while in the latter case we must compute the absolute value of its Jacobian determinant. An efficient procedure by which to achieve both of these objectives is to compute the PLU decomposition of \( \text{Id} + \Omega(\epsilon, q, v) = \text{PLU} \) where \( P \) is a permutation matrix, \( L \) is a lower-triangular matrix with unit diagonal, and \( U \) is an upper-triangular matrix. The computational cost of this decomposition is \( \frac{2}{3}m^3 + \mathcal{O}(m^2) \). Linear systems can be solved using the PLU decomposition by applying the permutation and solving the triangular systems via forward-backward substitution. Moreover, the required Jacobian determinant is simply \( \prod_{i=1}^{m} U_{ii} \), since the determinant of the permutation matrix has unit magnitude and the lower-triangular matrix \( L \) has unit Jacobian determinant since all of its diagonal elements are equal to one.

4.5 Built-In Robustness of the Lagrangian Integrator

As shown in Brofos and Lederman (2021b), the volume-preservation property (i.e. \( J = 1 \) in algorithm \( \text{A} \) in appendix \( \text{A} \)) of the generalized leapfrog integrator is predicated on the symmetry of partial derivatives \( \nabla_q^T \nabla_p H(q, p) = \nabla_p \nabla_q H(q, p) \). In an implementation of the generalized leapfrog integrator we may suppose that we have functions \( g_k(q) \) representing \( \frac{\partial G}{\partial q^{(k)}}(q) \). Substituting this function into the definition of the generalized leapfrog integrator
(definition 2.3) yields the following map \((q, p) \mapsto (\tilde{q}, \tilde{p})\)

\[
\begin{align*}
\tilde{p}^{(k)} &= p^{(k)} - \frac{\epsilon}{2} \delta_k(q, \tilde{p}) \\
\tilde{q} &= q + \frac{\epsilon}{2} (\Delta(q, \tilde{p}) + \Delta(\tilde{q}, \tilde{p})) \\
\tilde{p}^{(k)} &= \tilde{p}^{(k)} - \frac{\epsilon}{2} \delta_k(\tilde{q}, \tilde{p}),
\end{align*}
\]

where

\[
\begin{align*}
\delta_k(q, p) &= -\frac{1}{2} p^T G^{-1}(q) g_k(q) G^{-1}(q)p + \frac{\partial U}{\partial q^{(k)}}(q) \\
\Delta(q, p) &= G^{-1}(q)p
\end{align*}
\]

When \(g_k(q) = \frac{\partial G}{\partial q^{(k)}}(q)\), the resulting map is necessarily volume-preserving. However, we may then ask the question, “What happens when \(g_k(q)\) is *incorrectly implemented* so that, in fact, \(g_k(q) \neq \frac{\partial G}{\partial q^{(k)}}(q)\)?” The symmetry of partial derivatives has therefore been violated since

\[
\frac{\partial \Delta}{\partial q^{(k)}}(q, p) = \frac{\partial G^{-1}}{\partial q^{(k)}}(q)p \neq -G^{-1}(q)g_k(q)G^{-1}(q)p = \frac{\partial \delta_k}{\partial p}(q, p).
\]

In the case of the RMHMC Markov chain (definition B.19), detailed balance is no longer satisfied and there is no expectation that the RMHMC will converge to the target distribution.

The situation is different in the case of the LMC Markov chain (definition B.20). The fundamental difference is that LMC *expects* the transformation \((q, v) \mapsto (\tilde{q}, \tilde{v})\) to be non-volume-preserving, hence necessitating the Jacobian determinant correction in eq. (54). To see that the change-in-volume is still correctly computed even when \(g_k(q) \neq \frac{\partial G}{\partial q^{(k)}}(q)\), we observe that the Lagrangian leapfrog’s (definition 2.4) update to velocity in eq. (7) is a special case of the following map:

\[
\dot{v} = (\text{Id}_m + A(q, v))^{-1}(v - b(q))
\]

where \(A : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^{m \times m}\) and \(b : \mathbb{R}^m \to \mathbb{R}^m\). Under the assumption that \(A(q, v)\tilde{v} = A(q, v)v\) (which holds for the correctly implemented LMC with \(A = \Omega(\epsilon, q, v)\) by proposition B.8), the Jacobian determinant of the map \((q, v) \mapsto (q, \tilde{v})\) is

\[
\left| \det \left( \frac{\partial (q, \tilde{v})}{\partial (q, v)} \right) \right| = \frac{\det(\text{Id}_m + A(q, \tilde{v}))}{\det(\text{Id}_m + A(q, v))}.
\]
Hence, if we adopt the notation \( g_{k,ij}(q) \) as the \((i, j)\)-th element of \( g_k(q) \), and substitute

\[
\tilde{\Gamma}_{ij}^k(q) = \frac{1}{2} \sum_{l=1}^{m} G_{kl}^{-1}(q) \left( g_{i,lj}(q) + g_{j,li}(q) - g_{l,ij}(q) \right)
\]

for eq. (2) and define \( \tilde{\Omega}_{ij}(\epsilon, q, v) = \frac{\epsilon}{2} \sum_{k=1}^{m} \tilde{\Gamma}_{kl}^i(q) v^{(k)} \) then we still have \( \tilde{\Omega}(\epsilon, q, v)v = \tilde{\Omega}(\epsilon, q, \tilde{v})v\) using the fact that \( \tilde{\Gamma}_{ij}^k(q) = \tilde{\Gamma}_{ji}^k(q) \). Hence, with \( A(q, v) = \tilde{\Omega}(\epsilon, q, v) \), eq. (25) applies to compute the Jacobian determinant.

5 Experimentation

We now turn our attention to the empirical evaluation of the numerical integrator of Lagrangian dynamics in terms of its numerical order, its inverted variant, and its robustness to misspecification of the derivatives of the metric. We begin in section 5.1 by numerically validating the second-order behavior of the numerical integrator. In the subsequent material, we evaluate the integrator with and without inversion in a banana-shaped distribution, in Bayesian logistic regression, in a multiscale Student-\( t \) distribution, and in a stochastic volatility model. As baselines, we consider RMHMC and HMC. Code to reproduce these experiments may be found at [https://github.com/JamesBrofos/Rethinking-Lagrangian-Monte-Carlo](https://github.com/JamesBrofos/Rethinking-Lagrangian-Monte-Carlo).

We consider three metrics by which to assess the convergence of the Markov chain produced by ILMC and the baselines. First, we consider the expected squared jump distance (ESJD) as described in Gelman and Pasarica (2007); this measures the expected squared distance between the current state and the next state, where the expectation is computed over the acceptance probability. The larger the ESJD, the less serial autocorrelation in the Markov chain samples. We also consider the effective sample size (ESS) normalized by time elapsed, which gives an indication of the sampling efficiency of each method. We use the implementation of ESS as given in Kumar et al. (2019). We also consider the method of Brofos and Lederman (2021b) for measuring the ergodicity of the Markov chain given i.i.d. samples. Under this procedure, we project the i.i.d. samples and the Markov chain samples along one-hundred random directions and measure the average value of the Kolmogorov-Smirnov statistics of these one-dimensional distributions. By the Cramèr-Wold theorem,
Figure 2: Visualization of the second-order accuracy (i.e. third order local error) of the Lagrangian integrator on Lagrangian dynamics derived from a non-separable Hamiltonian. We see that the observed local error decreases on a log-log scale with a slope of three, corresponding to the claimed third-order local error. By contrast, we also show a line with a slope of two, indicating that the integrator exhibits accuracy better than first-order.

the closer these Kolmogorov-Smirnov statistics are concentrated toward zero, the higher the fidelity between the Markov chain samples and the i.i.d. samples. In implementing the generalized leapfrog integrator (definition 2.3), we resolve the fixed point equations eqs. (4) and (5) using fixed point iteration to a convergence tolerance of $1 \times 10^{-6}$, with convergence measured in $\| \cdot \|_{\infty}$.

5.1 Demonstration of Second-Order Error

We consider the following non-separable Hamiltonian $H(q, p) = \frac{q^2 p^2}{2}$, which describes geodesic motion in $\mathbb{R}$ when equipped with the metric $G(q) = 1/q^2$. After converting from momentum to velocity $v_t = \frac{p_t}{q_t^2}$, we obtain the second-order differential equation $a_t = \frac{v_t^2}{q_t}$. Given initial conditions $q_0$ and $p_0$ (in the Hamiltonian formalism), the exact solution to this differential equation is $q_t = q_0 \exp(q_0 p_0 t)$ and $v_t = q_0^2 p_0 \exp(q_0 p_0 t)$. If, as claimed, the Lagrangian integrator is indeed second-order, then it should exhibit third-order local error according to definition [B.10]. We evaluate this by examining the squared error between the output of the numerical integrator and the analytical solution for a decreasing sequence of
Figure 3: (Banana-Shaped Distribution) We show the ESJD, the minimum ESS per second, and the distribution of Kolmogorov-Smirnov statistics for sampling from the banana-shaped distribution using RMHMC, LMC, and ILMC. Surprisingly, we do not observe degradation of the ESJD when employing ILMC. Therefore, combined with its faster sampling iteration, ILMC enjoys the best ESS per second and a distribution of Kolmogorov-Smirnov statistics that is comparable to LMC.

step-sizes $\epsilon$; that is, denoting the output of a single step of the Lagrangian integrator by $(\hat{q}_\epsilon, \hat{v}_\epsilon)$, we measure $\|\hat{q}_\epsilon - q_\epsilon\|^2_2 + \|\hat{v}_\epsilon - v_\epsilon\|^2_2$. We observe in fig. 2 that this error decreases linearly on a log-log scale and, critically, the slope of this linear relation is three. This demonstrates numerically the third-order local error of the Lagrangian integrator and gives support to the claim that the method is of second-order accuracy.

5.2 Banana-Shaped Posterior Distribution

The banana-shaped distribution was proposed in Girolami and Calderhead (2011) by Cornebise and Julien as an example of Bayesian inference in non-identifiable models. In this example, a non-identifiable likelihood function in Bayesian linear regression is regularized according to a normal prior, the effect of which is to produce a density with symmetric, elongated tails. The generative model of this distribution is as follows:

$$ (\theta_1, \theta_2)^{\text{i.i.d.}} \sim \text{Normal}(0, \sigma^2_\theta) $$

$$ y_i|\theta_1, \theta_2^{\text{i.i.d.}} \sim \text{Normal}(\theta_1 + \theta_2^2, \sigma^2_y) \quad \text{for} \quad i = 1, \ldots, n. $$

$$(27) \quad (28)$$
This distribution also illustrates a short-coming of the generalized leapfrog method. For large step-sizes, the implicit update to the momentum variable will not have a solution; therefore, the generalized leapfrog integrator is compelled to adopt a significantly smaller step-size than can be used even by the standard leapfrog method. Indeed, an advantage of explicit numerical integrators is that one does not need to fret that constituent update steps in the integrator will not have solutions. We seek to draw 1,000,000 samples from this posterior.

In our experiments we set $\sigma_\theta^2 = \sigma_y^2 = 2$, $n = 100$, and set parameter values $\theta_1 = 1/2$ and $\theta_2 = \sqrt{1 - 1/2}$ for generating synthetic data. For HMC we use a step-size of 0.1 and ten integration steps, which was found to produce an acceptance probability between eighty and ninety percent. As for the Riemannian metric, we adopt the sum of the Fisher information of the log-likelihood and the negative Hessian of the log-prior. For RMHMC, we use a step-size of 0.04 and twenty integration steps, which produces an acceptance probability of around ninety percent. For LMC and ILMC, we use twenty integration steps with a step-size of 0.1, which yields an acceptance probability of around ninety percent. These parameter configurations were found to produce reasonable Markov chains based on hand-tuning. Results showing the effective sample size (ESS) per second are provided in fig. 3; we see that RMHMC struggles in this distribution, being even worse than ordinary HMC due to the requirement to use a small step-size. By contrast, LMC and ILMC do significantly better, with ILMC having the best ESS per second due to its elimination of two Jacobian determinant computations. We also show the distribution of this average value over ten trials for each sampling method. We find that the geometric methods based on the Lagrangian formalism perform similarly under this ergodicity measure and outperform competing methods.

5.3 Bayesian Logistic Regression

We consider Bayesian logistic regression on a breast cancer and a thyroid cancer dataset. The breast cancer dataset has 277 observations and ten covariates while the thyroid cancer dataset has 215 observations and six covariates. The Bayesian generative model is assumed
Figure 4: (Logistic Regression - Breast Cancer) We show the ESJD and the minimum and mean ESS per second for the breast cancer dataset using RMHMC, LMC, and ILMC. We observe that LMC and ILMC have degraded movement through the sample space as computed by the ESJD; however, this is compensated for by their superior computational efficiency, ultimately yielding more effectively independent samples per second and RMHMC.

$$\beta \sim \text{Normal}(0, \alpha^{-1}I_d)$$ (29)

$$y_i | x_i, \beta \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(\sigma(x_i^\top \beta)) \quad \text{for} \ i = 1, \ldots, n, \quad (30)$$

Figure 5: (Logistic Regression - Thyroid Cancer) We show the ESJD and the minimum and mean ESS per second for the thyroid cancer dataset using RMHMC, LMC, and ILMC. We observe that LMC and ILMC have significantly degraded movement through the sample space as computed by the ESJD; however, this is compensated for by their superior computational efficiency, ultimately yielding more effectively independent samples per second and RMHMC.
where $\sigma : \mathbb{R} \to (0,1)$ denotes the sigmoid function. As the Riemannian metric, we adopt the sum of the Fisher information of the log-likelihood and the negative Hessian of the log-prior. We show in figs. 4 and 5 the minimum ESS per second, where we have also considered varying the convergence threshold used to solve fixed point iterations in RMHMC; as discussed in detail by Brofos and Lederman (2021b), implementations of RMHMC require that the implicit updates to momentum and position be resolved using an iterative procedure such as fixed point iteration or Newton’s method. The tolerance in these numerical methods directly controls the degree to which reversibility and volume preservation are violated by the implementation of the generalized leapfrog method: For small values of the threshold, these theoretical properties are closely respected by the numerical method, while for large thresholds, violations occur. Reversibility and volume preservation imply detailed balance in HMC, and violations imply that the detailed balance may not hold in an implementation of RMHMC with large thresholds. By contrast, LMC and ILMC are fully explicit and detailed balance is respected to machine precision for either method. We observe that both sampling methods based on Lagrangian mechanics exhibit higher ESS per second than their counterpart based on the Hamiltonian formalism except for the largest thresholds (with the greatest bias); we refer the interested reader to Brofos and Lederman (2021b) for a detailed discussion on the effects of the convergence threshold on the bias of the RMHMC Markov chain. Moreover, ILMC outperforms LMC on this metric. When comparing the ESJD, we observe that, consistent with our understanding in the Gaussian case, the ILMC method has the smallest expected distance traveled. However, this is offset by a faster sampling iteration due to only requiring two Jacobian determinant calculations instead of four.

5.4 Multiscale Student’s $t$-Distribution

Multiscale distributions can be challenging for HMC since trajectories integrated by the leapfrog method will exhibit significant oscillations along directions of the distributions possessing the smallest spatial scale. To investigate this phenomenon, we consider sampling from a multivariate Student-$t$ distribution with a multiscale covariance; in particular we
We show the ESJD, the minimum ESS per second, and the distribution of Kolmogorov-Smirnov statistics for sampling from the multiscale Student-\(t\) distribution with \(\nu = 5 \times 10^9\) using RMHMC, LMC, and ILMC. We observe that RMHMC enjoys the best ESJD, time-normalized ESS, and Kolmogorov-Smirnov statistics. This illustrates an important limitation of the Lagrangian procedure in that its performance may actually be unable to match RMHMC in certain circumstances.

We show the ESJD, the minimum ESS per second, and the distribution of Kolmogorov-Smirnov statistics for sampling from the multiscale Student-\(t\) distribution with \(\nu = 5 \times 10^3\) using RMHMC, LMC, and ILMC. In this example, despite having a smaller ESJD, ILMC ultimately produces a slight benefit in terms of ESS per second. Among the geometric methods, the distribution of the Kolmogorov-Smirnov statistics are nearly identical.

consider a distribution with density function,

\[
\pi(x) \propto \left[1 + \frac{1}{\eta} x^\top \Sigma^{-1} x \right]^{-(\eta+m)/2},
\]  

(31)
where $x \in \mathbb{R}^m$, $\eta > 2$ is the degrees-of-freedom, and $\Sigma$ is the scale matrix. We consider scale matrices of the form $\Sigma = \text{diag}(1, \ldots, 1, \sigma^2) \in \mathbb{R}^{m \times m}$. In our experiments we set $m = 20$ and $\eta \in \{5 \times 10^0, 5 \times 10^3\}$ and consider multiscale distributions for $\sigma^2 \in \{1 \times 10^1, 1 \times 10^2, 1 \times 10^3, 1 \times 10^4\}$. We choose these two values of the degrees-of-freedom to demonstrate two distinct sampling behaviors. For the Riemannian methods, we consider a step-size of 0.7 and twenty integration steps. As the Riemannian metric, we use the positive definite term in the negative Hessian of the log-density of the distribution.

For $\nu = 5 \times 10^0$, we observe that the ESJD is largest for RMHMC; this occurs because RMHMC enjoys a far superior acceptance probability in this scenario at 95% whereas LMC and ILMC have acceptance probabilities of 40% and 62%, respectively. This degraded performance of the Lagrangian methods is then reflected in the time-normalized ESS, which shows RMHMC dominating LMC, ILMC, and HMC. For the case of $\eta = 5 \times 10^3$, circumstances are more favorable to the Lagrangian methods, with ILMC exhibiting the best performance in terms of time-normalized ESS, with the LMC method giving the second best results. In terms of ESJD, we observe that ILMC moves less far in sample space than LMC, but that this is offset by the faster sampling. As in the case of the banana-shaped distribution, we may sample from this target density analytically in order to assess the ergodicity properties of the samplers. We find that the ergodicity of the geometric methods is essentially constant with respect to the multiscale parameter, whereas the performance of Euclidean HMC noticeably degenerates. For both $\nu = 5 \times 10^3$ and $\nu = 5 \times 10^0$, one observes that the performance of the geometric methods is essentially constant over the multiple scales of the target distribution, demonstrating the beneficial effect of capturing the geometry of the target.
5.5 Stochastic Volatility Model

We consider Bayesian inference in a stochastic volatility model. We consider the following generative model:

\[ x_t | x_{t-1}, \phi, \sigma^2 \sim \text{Normal}(\phi x_{t-1}, \sigma^2) \quad (32) \]
\[ y_t | \beta, x_t \sim \text{Normal}(0, \beta^2 \exp(x_t)) \quad (33) \]

for \( t = 2, \ldots, T \) in eq. (32) and \( t = 1, \ldots, T \) in eq. (33) and with priors \( x_1 \sim \text{Normal}(0, \sigma^2/(1-\phi^2)), (\phi+1)/2 \sim \text{Beta}(20, 3/2), 1/\sigma^2 \sim \xi^2(10, 1/20) \), and the prior over \( \beta \) being proportional to \( 1/\beta^2 \). Given \( (y_1, \ldots, y_T) \), we seek to sample the posterior of \( (x_1, \ldots, x_T, \phi, \beta, \sigma^2) \). We follow [Girolami and Calderhead (2011)] and employ a Metropolis-within-Gibbs-like alternating procedure for sampling the posteriors of \( (x_1, \ldots, x_T) \) and \( (\phi, \beta, \sigma^2) \). In our experiments we set \( T = 1,000 \) and use values of \( \phi = 0.98, \beta = 0.65, \) and \( \sigma^2 = 0.15^2 \). For HMC, we use a step-size of 0.01 and fifty integration steps when sampling \( (\phi, \beta, \sigma^2) \); for the geometric methods, we use a step-size of 0.5 and six integration steps. As for the Riemannian metric, we adopt the sum of the Fisher information of the log-likelihood and the negative Hessian of the log-prior. We compare the average ESS per second among the three latent variables.

Figure 8: (Stochastic Volatility Model) We show the ESJD and the minimum and mean ESS per second for the stochastic volatility model using HMC, RMHMC, LMC, and ILMC. In this example, ILMC moves less efficiently through the same space as measured by ESJD, and ultimately produces fewer effective samples per second relative to LMC. However, it is intriguing to note that ILMC nonetheless produces more effective transitions than RMHMC and HMC.
Figure 9: (Fitzhugh-Nagumo Model) We show the minimum ESS per second and the Kolmogorov-Smirnov ergodicity metric for inferring the posterior distribution of \((a, b, c)\) in the Fitzhugh-Nagumo posterior distribution. We observe that LMC has the strongest performance in terms of time-normalized ESS and that all methods exhibit comparable performance with respect to ergodicity. When we break the symmetry of partial derivatives, RMHMC no longer has the target distribution as its invariant distribution; this causes a degradation of the ergodicity metric. However, the geometric methods LMC and ILMC are robust to this misspecification because they are non-volume-preserving by design.

\((\phi, \beta, \sigma^2)\) with results reported in fig. 8. We find that LMC and ILMC are the strongest performing methods, with LMC having better ESS per second due to its more efficient traversal of the sample space. Indeed, ILMC has degraded performance in this example, owing to its greater autocorrelation, but nevertheless outperforms HMC and RMHMC.

### 5.6 Fitzhugh-Nagumo Model

We now investigate the LMC and RMHMC in the Fitzhugh-Nagumo posterior distribution. The Fitzhugh-Nagumo posterior distribution has the following generative model,

\[
(a, b, c) \overset{i.i.d.}{\sim} \text{Normal}(0, 1) \quad \text{(34)}
\]

\[
\hat{r}_{tk} \overset{i.i.d.}{\sim} \text{Normal}(r_{tk}, \sigma^2) \quad \text{(35)}
\]

\[
\hat{v}_{tk} \overset{i.i.d.}{\sim} \text{Normal}(v_{tk}, \sigma^2) \quad \text{for } k = 1, \ldots, n, \quad \text{(36)}
\]
where $t_1, \ldots, t_n$ are evenly spaced time points in $[0, 10)$ and $r_t$ and $v_t$ obey the differential equations,

\begin{align}
\dot{r}_t &= -\frac{v_t - a + br_t}{c}, \\
\dot{v}_t &= cv_t - \frac{cv_t^3}{3} + cr_t.
\end{align}

(37)  

(38)

In our experiments we give initial conditions $v_0 = 1$, $r_0 = -1$, $\sigma^2 = 1/4$, and $n = 200$. The objective is to infer the posterior of $q = (a, b, c)$ given observations $\{(\hat{v}_{t_k}, \hat{r}_{t_k})\}_{k=1}^{200}$.

Here we emphasize a different aspect of the computation: its robustness to misspecification. As described in section 4.5, the proof that the generalized leapfrog integrator is a volume-preserving transformation when used to integrate Hamiltonian vector fields is the symmetry of partial derivatives. We therefore consider the repercussions of an error in the implementation of partial derivatives that invalidates this requirement. To give further context to this experiment, we quote from Dahlquist and Björck (2003): “In all numerical work, one must expect that clerical errors, errors in hand calculation, and misunderstandings will occur. [...] Most of the errors depend on the so-called human factor. [...] We take up these sources of error in order to emphasize that both the person who carries out a calculation and the person who guides the work of others can plan so that such sources of error are not damaging.” It seems to us, therefore, that the robustness of a Markov chain procedure to human misspecification is a most desirable circumstance.

As described in Girolami and Calderhead (2011), computing the gradient of the log-posterior, the Riemannian metric, and the Jacobian of the Riemannian metric requires us to compute sensitivity equations of the form $\frac{d}{dt} \frac{\partial r_t}{\partial q_i}$, $\frac{d}{dt} \frac{\partial v_t}{\partial q_i}$, $\frac{d}{dt} \frac{\partial^2 r_t}{\partial q_i \partial q_j}$, and $\frac{d}{dt} \frac{\partial^2 v_t}{\partial q_i \partial q_j}$ for $i = 1, 2, 3$ and $j = 1, 2, 3$. We employ a Riemannian metric of the form,

\begin{equation}
G_{ij}(q) = \frac{1}{\sigma^2} \sum_{k=1}^{n} \left( \frac{\partial v_{t_k}}{\partial q_i} \frac{\partial v_{t_k}}{\partial q_j} + \frac{\partial r_{t_k}}{\partial q_i} \frac{\partial r_{t_k}}{\partial q_j} \right) + 1 \{i = j\}.
\end{equation}

(39)

If any of these sensitivity equations are misspecified, then we will destroy the symmetry of partial derivatives required by the leapfrog integrator to preserve volume. However, the integrator used in LMC and ILMC is non-volume-preserving by design, and the appropriate volume correction is readily computable during sampling. We expect, therefore, that
the volume correction employed in LMC and ILMC will provide robustness against these incorrectly computed quantities. We note that, in this case, there is no obvious mechanism by which to compute the required Jacobian determinant when the generalized leapfrog integrator is no longer a symplectic transformation.

In fig. 9 we show the minimum ESS per second and Kolmogorov-Smirnov statistics for sampling from the Fitzhugh-Nagumo posterior. We observe that LMC outperforms ILMC in this example, but that both Lagrangian methods outperform HMC and RMHMC. In terms of ergodicity, all methods perform similarly. When we introduce changes to the sensitivity equations, we observe that RMHMC’s ergodicity severely degrades. For HMC, for which higher-order sensitivities are not required (because of the Euclidean metric), ergodicity is identical to the previous case. However, for LMC and ILMC, in which the higher-order sensitivities are required to compute the Christoffel symbols, the fact that the sensitivities have been incorrectly specified has not noticeably degraded ergodicity. This is a virtue of the LMC and ILMC Markov chains that they are more robust to human errors of this variety.

6 Conclusion

This work has examined the numerical integrator for Lagrangian Monte Carlo (LMC). Motivated by the observation that LMC requires four Jacobian determinant evaluations, mechanisms by which this number may be reduced were examined. By inverting the sequence of integration so that position, rather than velocity, is updated twice, the number of Jacobian determinant evaluations in each step was reduced from four to two while still maintaining a fully explicit method. Empirical evaluations of this method were provided to show several situations in which the proposed integration strategy enjoys the best time-normalized performance among several alternatives. Moreover, it has been demonstrated in this work that the local error of the Lagrangian leapfrog (and its inverted counterpart) is third order, which improves the previously known order. Additionally, an important robustness property was characterized that LMC possesses and an illustration wherein human error will invalidate stationarity in RMHMC but not in LMC was given.
Algorithm 1 Pseudo-code implementing the generalized leapfrog integrator for computing approximate solutions to Hamilton’s equations of motion. We assume a Hamiltonian in the form of eq. (1).

1: **Input**: An integration step-size $\epsilon \in \mathbb{R}$, an initial position in phase space $(q, p) \in \mathbb{R}^m \times \mathbb{R}^m$.
2: Compute an implicit half-step update to the momentum variable to obtain $\tilde{p}$ using eq. (4).
3: Compute an implicit full-step to the position variable to obtain $\tilde{q}$ using eq. (5).
4: Compute an explicit half-step update to the momentum variable to obtain $\tilde{p}$ using eq. (6).
5: Compute the Jacobian determinant $J = 1$.
6: **Output**: The updated position and momentum $(\tilde{q}, \tilde{p})$ and the Jacobian determinant $J$.

Methods of Bayesian inference that incorporate geometric understanding exhibit a pleasing aesthetic, yet they are burdened by numerical considerations – among which are fixed points and cubic complexity – that have limited their adoption. We hope that this research sparks renewed interest in mechanisms of improving these geometric methods.

SUPPLEMENTARY MATERIAL

A Algorithms

In this appendix we give pseudo-code implementations of algorithms featured in this work. In algorithm 1 we show the generalized leapfrog, which contrasts with the Lagrangian leapfrog in algorithm 2 in the presence of implicitly-defined integration steps. Algorithm 4 shows the Lagrangian integrator but with an inverted sequence of integration (updating position before velocity). In algorithm 3 we show an algorithmic implementation of a single-step of the Hamiltonian Monte Carlo Markov chain, which can be carried out using any of the integrators considered in this work.
Algorithm 2 Pseudo-code implementing the Lagrangian leapfrog integrator for computing approximate solutions to the Lagrangian equations of motion. We assume a Lagrangian in the form of eq. (42). The conversion from momentum to velocity and back again is included for notational simplicity; in practice, when applying multiple steps of this integrator, these conversions need only be computed for the first conversion from momentum to velocity and the last conversion from velocity to momentum.

1: **Input**: An integration step-size \( \epsilon \in \mathbb{R} \), an initial position in phase space \((q,p) \in \mathbb{R}^m \times \mathbb{R}^m\).

2: Use the inverse Legendre transform (definition B.5) to convert from momentum to velocity \( v = G^{-1}(q)p \).

3: Compute an explicit half-step update to the velocity variable to obtain \( \ddot{v} \) using eq. (7).

4: Compute an explicit full-step to the position variable to obtain \( \ddot{q} \) using eq. (8).

5: Compute an explicit half-step update to the velocity variable to obtain \( \dot{v} \) using eq. (9).

6: Use the Legendre transform (definition B.4) to convert from velocity to momentum \( \ddot{p} = G(\ddot{q})\dot{v} \).

7: Compute the Jacobian determinant \( J = \frac{\partial(\ddot{q},\ddot{p})}{\partial(q,p)} \) using eq. (57).

8: **Output**: The updated position and momentum \((\ddot{q}, \ddot{p})\) and the Jacobian determinant \( J \).

B Expanded Preliminaries

B.1 Hamiltonian and Lagrangian Mechanics

*Definition* B.1. Let \( H : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \) be a smooth function, which we call the Hamiltonian. Hamilton’s equations of motion are defined as the solutions to the initial value problem

\[
v_t = \nabla_p H(q_t, p_t) \quad (40)
\]

\[
\dot{p}_t = -\nabla_q H(q_t, p_t), \quad (41)
\]

with \((q_0, p_0)\) a given initial position in phase space.
Algorithm 3 Single step of the Hamiltonian / Lagrangian Monte Carlo transition kernel.

1: **Input:** The current state of the Markov chain \((q_n, p_n)\), an integration step-size \(\epsilon \in \mathbb{R}\), a number of integration steps \(k \in \mathbb{N}\), a numerical integrator \(\Phi\) (as described in any of algorithms 1, 2 and 4).

2: Resample the momentum \(p_n|q_n \sim \text{Normal}(0, G(q_n))\).

3: Initialize the Jacobian determinant \(J = 1\) and initial integration state \((\tilde{q}, \tilde{p}) = (q_n, p_n)\).

4: for \(i = 1, \ldots, k\) do

5: Compute the proposal by applying the numerical integrator: \(((\tilde{q}, \tilde{p}), \tilde{J}) = \Phi_\epsilon(\tilde{q}, \tilde{p})\).

6: Update the Jacobian determinant \(J = J \times \tilde{J}\)

7: end for

8: Apply the momentum flip operator (definition B.18) to obtain \((\tilde{q}, \tilde{p}) = F(\tilde{q}, \tilde{p})\).

9: Sample \(u \sim \text{Uniform}(0, 1)\) and compute the acceptance probability \(a = \alpha((q_n, p_n), (\tilde{q}, \tilde{p}), J)\) using eq. (54).

10: if \(u < a\) then

11: Accept the proposal \((q_{n+1}, p_{n+1}) = (\tilde{q}, \tilde{p})\).

12: else

13: Reject the proposal \((q_{n+1}, p_{n+1}) = (q_n, p_n)\).

14: end if

15: **Output:** The next state of the Markov chain \((q_{n+1}, p_{n+1})\).

Definition B.1 gives us a system of coupled first-order differential equations. Hamilton’s equations of motion exhibit several key properties (Marsden and Ratiu, 2010), which we summarize.

**Theorem B.2.** The Hamiltonian mechanics given in definition B.1 possess the following three properties:

1. They preserve the Hamiltonian \(\frac{d}{dt}H(q_t, p_t) = 0\).

2. Denoting \(\dot{z}_t = (v_t, \dot{p}_t)\) the equations of motion in phase space, Hamiltonian mechanics preserve volume in phase space: \(\text{div}(\dot{z}_t) = 0\).
**Algorithm 4** Pseudo-code implementing the inverted Lagrangian leapfrog integrator for computing approximate solutions to the Lagrangian equations of motion. We assume a Lagrangian in the form of eq. (42). This version of leapfrog integration for Lagrangian dynamics requires only two Jacobian determinant computations in each step, as opposed to the four required by algorithm 2.

1: **Input**: An integration step-size $\epsilon \in \mathbb{R}$, an initial position in phase space $(q, p) \in \mathbb{R}^m \times \mathbb{R}^m$.

2: Use the inverse Legendre transform (definition B.5) to convert from momentum to velocity $v = G^{-1}(q)p$.

3: Compute an explicit half-step to the position variable to obtain $\dot{v}$ using eq. (10).

4: Compute an explicit full-step update to the velocity variable to obtain $\dot{v}$ using eq. (11).

5: Compute an explicit half-step to the position variable to obtain $\ddot{q}$ using eq. (12).

6: Use the Legendre transform (definition B.4) to convert from velocity to momentum $\dot{p} = G(\dot{q})\dot{v}$.

7: Compute the Jacobian determinant $J = \frac{\partial (\dot{q}, \dot{p})}{\partial (q, p)}$ using eq. (57).

8: **Output**: The updated position and momentum $(\ddot{q}, \ddot{p})$ and the Jacobian determinant $J$.

3. Under the conditions that $\nabla_q H(q, -p) = \nabla_q H(q, p)$ and $-\nabla_p H(q, p) = \nabla_p H(q, -p)$, the equations of motion are reversible via negation of the momentum variable.

A proof is given in appendix C. In the special case when $U(q) = 0$ in eq. (11) for all $q \in \mathbb{R}^m$, Hamilton’s equations of motion produce geodesic motion on the Riemannian manifold $(\mathbb{R}^m, G)$, where $G$ assumes the role of the Riemannian metric (Calin and Chang, 2004).

**Definition B.3**. Given a Riemannian Hamiltonian (definition 2.1), define the associated Lagrangian function by,

$$L(q, v) = \tilde{K}(q, v) - U(q),$$

(42)

where $\tilde{K}(q, v) = \frac{1}{2}v^\top G(q)v$. 

30
Hamiltonian mechanics are related to Lagrangian dynamics by the Legendre transform which converts between momentum and velocity (Marsden and West, 2001).

**Definition B.4.** Let $L : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ be a Lagrangian of the form in eq. (42). The **Legendre transform** relates the momentum and velocity according to $p = \frac{\partial L}{\partial \dot{v}}(q,v) = G(q)v$.

**Definition B.5.** Let $H : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ be a Hamiltonian of the form in eq. (1). The **inverse Legendre transform** relates velocity and momentum according to $v = \frac{\partial H}{\partial p}(q,p) = G^{-1}(q)p$.

The Lagrangian then determines equations of motion in accordance with the following physical principle.

**Definition B.6.** *Hamilton’s principle* states that the equations of motion $(q_t, v_t)$ over an interval of time $[a,b]$, with known boundary conditions $q_a$ and $q_b$, are solutions of the variational equation

$$\frac{\delta S[q_t]}{\delta q_t} = 0$$

where

$$S[q_t] = \int_a^b L(q_t, v_t) \, dt$$

where $q_t$ and $v_t$ are related by $v_t = \frac{d}{dt}q_t$.

Hamilton’s principle states that equations of motion (as specified by $q$ and $v$) should *extremize* the Lagrangian subject to the boundary conditions on $q_a$ and $q_b$.

**Theorem B.7.** Under Hamilton’s principle (definition B.6), the equations of motion $(q_t, v_t)$ must be solutions of the Euler-Lagrange equation

$$\nabla_q L(q_t, v_t) = \frac{d}{dt} \nabla_v L(q_t, v_t).$$

The equations of motion from theorem B.7 (or definition B.6) are, in fact, equivalent to the motion produced under Hamiltonian mechanics in definition B.1 (Marsden and Ratiu, 2010). For Lagrangians of the form in eq. (42), the $k$-th element of the acceleration is,

$$a^{(k)}_t = - \sum_{i=1}^{m} \sum_{j=1}^{m} \Gamma^k_{ij}(q_t) v^{(i)}_t v^{(j)}_t - \sum_{i=1}^{m} G^{-1}_{kt}(q_t) \frac{\partial U}{\partial q^{(i)}}(q_t),$$

where $\Gamma^k_{ij}$ are the Christoffel symbols (definition 2.2). We conclude this section by noting that $\Omega$ in eq. (3) enjoys the following properties:
Proposition B.8. The function $\Omega : \mathbb{R} \times \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^{m \times m}$ in eq. (3) satisfies
\[
\Omega(\epsilon, q, v)\ddot{v} = \Omega(\epsilon, q, \dot{v})v
\]  
(47)
\[
\frac{\partial}{\partial v} (\Omega(\epsilon, q, v)\ddot{v}) = \Omega(\epsilon, q, \dot{v}).
\]  
(48)

Proof. The result in eq. (47) can be seen from
\[
(\Omega(\epsilon, q, v)\ddot{v})_i = \frac{\epsilon}{2} \sum_{j=1}^{m} \sum_{k=1}^{m} \Gamma_{kj}^{i} (q) v^{(k)} \ddot{v}^{(j)}
\]  
(49)
\[
= \frac{\epsilon}{2} \sum_{j=1}^{m} \sum_{k=1}^{m} \Gamma_{jk}^{i} (q) v^{(k)} \dot{v}^{(j)}
\]  
(50)
\[
= \frac{\epsilon}{2} \sum_{k=1}^{m} \sum_{j=1}^{m} \Gamma_{jk}^{i} (q) v^{(k)} \ddot{v}^{(j)}
\]  
(51)
\[
= (\Omega(\epsilon, q, \dot{v})v)_i.
\]  
(52)
Equation (48) then follows as an immediate corollary.

Like Hamiltonian mechanics, Lagrangian dynamics conserve the Hamiltonian (when $p_t = G(q_t) v_t$) and are reversible. However, Lagrangian dynamics do not conserve volume in $(q, v)$-space.

B.2 Background on Numerical Integrators

The Hamiltonian equations of motion in eqs. (40) and (41) and the Lagrangian motion described in eq. (46) rarely have closed-form solutions. Therefore, it is necessary to investigate methods of numerical integration to produce approximate solutions to these initial value problems. In the following section, we review key ideas from Hairer et al. (2006).

Definition B.9. Let $g : \mathbb{R}^m \to \mathbb{R}^m$. The solution to an initial value problem is a function $z(\cdot) : \mathbb{R} \to \mathbb{R}^m$ for which $\frac{d}{dt} z_t = g(z_t)$ and for which the initial value $z_0 \in \mathbb{R}^m$ is known. In this case $g$ is called a (time-homogenous) vector field.

Definition B.10. Let $z_t$ be the solution to an initial value problem (definition B.9) with initial value $z_0$. A numerical integrator with step-size $\epsilon$ is a mapping $\Phi_{\epsilon} : \mathbb{R}^m \to \mathbb{R}^m$ and is said to have order $p$ if $\| \Phi_{\epsilon}(z_0) - z_\epsilon \| = \mathcal{O}(\epsilon^{p+1})$. 
Definition B.11. The adjoint of a numerical method $\Phi_\epsilon : \mathbb{R}^m \to \mathbb{R}^m$ is defined by the relation $\Phi_\epsilon^*(z) = \Phi^{-1}_\epsilon(z)$. The numerical method $\Phi_\epsilon$ is said to be self-adjoint if $\Phi_\epsilon^* = \Phi_\epsilon$.

Theorem B.12 (Hairer et al. (2006)). Consider an initial value problem $\dot{z}_t = g(z_t)$ with initial condition $z_0 \in \mathbb{R}^m$. Let $\Phi_\epsilon$ be a one-step numerical integrator of (maximal) order $r \in \mathbb{N}$. If $\Phi_\epsilon$ is self-adjoint, then $r$ is even.

Proposition B.13. Let $g : \mathbb{R}^m \to \mathbb{R}^m$ be a time-homogenous vector field. Let $\Xi : \mathbb{R}^m \to \mathbb{R}^m$ be a diffeomorphism. Let $z_t$ be the solution to the initial value problem $\dot{z}_t = g(z_t)$ given $z_0$. Let $\Phi_\epsilon : \mathbb{R}^m \to \mathbb{R}^m$ be the flow map of $z_\epsilon$ and suppose that $\hat{\Phi}_\epsilon$ is a $p$-th order approximation of $\Phi_\epsilon$. Then $\Xi \circ \hat{\Phi}_\epsilon$ is a $p$-th order approximation of $\Xi \circ \Phi_\epsilon$.

A proof is given in appendix D.

Proposition B.14. The generalized leapfrog integrator is a second-order (definition B.10), self-adjoint (definition B.11) numerical method.

B.3 Hamiltonian and Lagrangian Monte Carlo

Our objective in Bayesian inference is to draw samples from the distribution whose density is $\pi(q) \propto \exp(\mathcal{L}(q))$. We now review basic concepts from Markov chain Monte Carlo.

Definition B.15. A Markov chain transition kernel is a map $K : \mathbb{R}^m \times \mathfrak{B}(\mathbb{R}^m) \to [0,1]$ satisfying (i) for every $x \in \mathbb{R}^m$ the map $A \mapsto K(x,A)$ is a probability measure and (ii) for every $A \in \mathfrak{B}(\mathbb{R}^m)$, the map $x \mapsto K(x,A)$ is measurable. Given a Markov chain transition kernel $K$, a Markov chain is defined inductively by $x_{n+1}|x_n \sim K(x_n, \cdot)$.

HMC accomplishes this by artificially expanding the distribution to incorporate a momentum variable. Defining $U(q) = -\mathcal{L}(q) + \frac{1}{2} \log \det(G(q))$, let $H(q,p)$ be the Hamiltonian given in eq. (1) and observe that the density $\pi(q,p) \propto \exp(-H(q,p))$ has $\pi(q) = \int_{\mathbb{R}^m} \pi(q,p) \, dp$ and $p|q \sim \text{Normal}(0, G(q))$. In order to unify methods of Bayesian inference based on either the generalized leapfrog integrator (definition 2.3) or the Lagrangian leapfrog (definition 2.4) under one framework, we now introduce the Markov chain transition kernel based on smooth involutions.
Definition B.16 (Neklyudov et al. (2020)). Let $\Phi : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m \times \mathbb{R}^m$ be a smooth involution (i.e. $\Phi = \Phi^{-1}$). Let $\pi : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}_+$ be a probability density with respect to Lebesgue measure. Then we define the Markov chain transition kernel of involutive Monte Carlo by

\[ K((q, p), (A, B)) = \alpha((q, p), (\tilde{q}, \tilde{p}), J) \mathbb{1} \{ (\tilde{q}, \tilde{p}) \in (A, B) \} \]
\[ + (1 - \alpha((q, p), (\tilde{q}, \tilde{p}), J)) \mathbb{1} \{ (q, p) \in (A, B) \} \]
\[ \alpha((q, p), (\tilde{q}, \tilde{p}), J) = \min \left\{ 1, \frac{\exp(-H(\tilde{q}, \tilde{p}))}{\exp(-H(q, p))} \cdot J \right\} \]
\[ J = \left| \det \left( \frac{\partial (\tilde{q}, \tilde{p})}{\partial (q, p)} \right) \right| , \]

where $(A, B) \in \mathcal{B}(\mathbb{R}^m \times \mathbb{R}^m)$ and $(\tilde{q}, \tilde{p}) = \Phi(q, p)$.

Proposition B.17. The Markov chain transition kernel of involutive Monte Carlo satisfies detailed balance with respect to the distribution whose density is $\pi(q, p) \propto \exp(-H(q, p))$.

A proof is provided in appendix F. Central to the construction of involutions of interest to us is the momentum flip operator, defined as follows.

Definition B.18. The momentum flip operator is the map $F(q, p) = (q, -p)$.

The fact that numerical integrators can be combined with the momentum flip operator in order to produce involutions is covered in appendix F in the case of the Lagrangian leapfrog; other integrators are handled similarly. We provide pseudo-code implementing a single step of involutive Monte Carlo in algorithm [3] with involutions provided by the generalized leapfrog or Lagrangian leapfrog integrator. Given an initial point $(q_0, p_0)$ in phase space drawn from the distribution $\pi(q, p)$, the sequence of states $((q_1, p_1), (q_2, p_2), \ldots)$ computed by algorithm [3] are guaranteed to have $\pi(q, p)$ as their marginal distributions.

Under the additional conditions that the HMC Markov chain is irreducible and aperiodic, HMC also produces an ergodic chain.

The generalized leapfrog integrator is volume preserving. However, the integrator of Lagrangian dynamics is not. The required change-of-volume can be deduced as follows. First, observe that the update eq. (8) is immediately volume-preserving since it is merely
the translation of the position variable by a quantity. The change of volume incurred in
eq (7) has a Jacobian determinant given by,
\begin{align}
\left| \det \left( \frac{\partial (q, \dot{v})}{\partial (q, v)} \right) \right| &= \left| \frac{\det (\text{Id}_m - \Omega(\epsilon, q, \dot{v}))}{\det (\text{Id}_m + \Omega(\epsilon, q, v))} \right|.
\end{align}
(56)
The update in eq. (9) incurs a similar change of volume. Thus, when employing algorithm 2, the Jacobian determinant of the transformation \((q, p) \mapsto (\tilde{q}, \tilde{p})\) is,
\begin{align}
\left| \det \left( \frac{\partial (\tilde{q}, \tilde{p})}{\partial (q, p)} \right) \right| &= \det (\mathbf{G}^{-1}(q)) \det (\mathbf{G}(\tilde{q})) \times \\
&\left| \frac{\det (\text{Id}_m - \Omega(\epsilon, \tilde{q}, \dot{\tilde{v}}))\det (\text{Id}_m - \Omega(\epsilon, q, \dot{v}))}{\det (\text{Id}_m + \Omega(\epsilon, \tilde{q}, \dot{v}))\det (\text{Id}_m + \Omega(\epsilon, q, v))} \right|.
\end{align}
(57)

\textbf{Definition B.19.} Let \(\Phi_{\epsilon}\) be the generalized leapfrog integrator with step-size \(\epsilon \in \mathbb{R}\) (definition 2.3). Let \(k \in \mathbb{N}\) be a number of integration steps. The \textit{Riemannian manifold Hamiltonian Monte Carlo (RMHMC) Markov chain} is an instance involutive Monte Carlo (definition B.16) with involution \(\mathbf{F} \circ \Phi^k_{\epsilon}\).

\textbf{Definition B.20.} Let \(\Phi_{\epsilon}\) be the Lagrangian leapfrog integrator with step-size \(\epsilon \in \mathbb{R}\) (definition 2.4). Let \(k \in \mathbb{N}\) be a number of integration steps. The \textit{Lagrangian Monte Carlo (LMC) Markov chain} is an instance involutive Monte Carlo (definition B.16) with involution \(\mathbf{F} \circ \Phi^k_{\epsilon}\).

\section{C Proofs Concerning Mechanics Systems}

We give a proof of theorem B.2.

\textbf{Proof.} We first show that the Hamiltonian energy is conserved.
\begin{align}
\frac{d}{dt} H(q_t, p_t) &= \nabla_q H(q_t, p_t) \cdot v_t + \nabla_p H(q_t, p_t) \cdot \dot{p}_t \\
&= \nabla_q H(q_t, p_t) \cdot \nabla_p H(q_t, p_t) - \nabla_p H(q_t, p_t) \cdot \nabla_q H(q_t, p_t) \\
&= 0.
\end{align}
(58) (59) (60)

Next we show that Hamiltonian mechanics conserve volume in \((q, p)\)-space. This is equiv-
alent to the vector field \((v_t, \dot{p}_t)\) having zero divergence, which we now verify.

\[
\text{div}(v_t, \dot{p}_t) = \nabla_q \cdot v_t + \nabla_p \cdot \dot{p}_t \tag{61}
\]
\[
= \nabla_q \cdot \nabla_p H(q_t, p_t) - \nabla_p \cdot \nabla_q H(q_t, p_t) \tag{62}
\]
\[
= \sum_{i=1}^{m} \frac{\partial^2 H}{\partial q_i \partial p_i} H(q_t, p_t) - \sum_{i=1}^{m} \frac{\partial^2 H}{\partial p_i \partial q_i} H(q_t, p_t) \tag{63}
\]
\[
= 0, \tag{64}
\]

by symmetry of partial derivatives. Finally we show that under the conditions that
\(\nabla_q H(q, -p) = \nabla_q H(q, p)\) and \(\nabla_p H(q, -p) = -\nabla_p H(q, p)\) that the equations of motion are reversible under negation of the momentum variable. To see this, fix \(\tau \in \mathbb{R}_+\) and consider \((q_t, p_t)\) satisfying eqs. (40) and (41) for \(t \in [0, \tau]\). Let \(\tilde{p}_t = -p_{\tau-t}\) and \(\tilde{q}_t = q_{\tau-t}\).

We find that \(\tilde{q}_t\) and \(\tilde{p}_t\) obey the following equations of motion:

\[
\frac{d}{dt} \tilde{q}_t = -v_{\tau-t} \tag{65}
\]
\[
= -\nabla_p H(q_{\tau-t}, p_{\tau-t}) \tag{66}
\]
\[
= \nabla_p H(\tilde{q}_t, -p_{\tau-t}) \tag{67}
\]
\[
= \nabla_p H(\tilde{q}_t, \tilde{p}_t) \tag{68}
\]
\[
\frac{d}{dt} \tilde{p}_t = \dot{p}_{\tau-t} \tag{69}
\]
\[
= -\nabla_q H(q_{\tau-t}, p_{\tau-t}) \tag{70}
\]
\[
= -\nabla_q H(q_{\tau-t}, -p_{\tau-t}) \tag{71}
\]
\[
= -\nabla_q H(\tilde{q}_t, \tilde{p}_t). \tag{72}
\]

Thus we see that \((\tilde{q}_t, \tilde{p}_t)\) are also solutions to Hamilton’s equations of motion and satisfy \(\tilde{q}_\tau = q_0\) and \(\tilde{p}_\tau = -p_0\), demonstrating reversibility. \(\square\)
D Proofs Concerning the Numerical Order of the Lagrangian Integrators

The claim of first order accuracy was derived in [Lan et al. 2015]; we have included a proof for completeness.

**Proof of lemma 4.5.** Let \( q_t \) and \( v_t \) be solutions to the initial value problem,

\[
\dot{q}_t = v_t \tag{73}
\]

\[
\dot{v}_t = -\Omega(2, q_t, v_t)v_t - G^{-1}(q_t)\nabla U(q_t). \tag{74}
\]

Given the initial value \( q_0 = q \) and \( v_0 = v \), we can expand the solution in a Taylor series about \( t = 0 \).

\[
q_t = q_0 + t\dot{q}_0 + \mathcal{O}(t^2) \tag{75}
\]

\[
= q + tv + \mathcal{O}(t^2) \tag{76}
\]

\[
v_t = v_0 - t(\Omega(2, q_0, v_0)v_0 - G^{-1}(q_0)\nabla U(q_0)) + \mathcal{O}(t^2) \tag{77}
\]

\[
= v - \Omega(2t, q, v)v - tG^{-1}(q)\nabla U(q) + \mathcal{O}(t^2). \tag{78}
\]

Letting \( t = \epsilon \) gives the following approximations,

\[
q_{\epsilon} = q + \epsilon v + \mathcal{O}(\epsilon^2) \tag{79}
\]

\[
v_{\epsilon} = v - \Omega(2\epsilon, q, v)v - \epsilon G^{-1}(q)\nabla U(q) + \mathcal{O}(\epsilon^2) \tag{80}
\]

Now we expand the steps of the explicit numerical integrator. We begin with the first update to velocity.

\[
\bar{v} = (I_d + \Omega(\epsilon, q, v))^{-1}\left(v - \frac{\epsilon}{2}G^{-1}(q)\nabla U(q)\right) \tag{81}
\]

\[
= (I_d - \Omega(\epsilon, q, v) + \mathcal{O}(\epsilon^2))\left(v - \frac{\epsilon}{2}G^{-1}(q)\nabla U(q)\right) \tag{82}
\]

\[
= v - \Omega(\epsilon, q, v)v - \frac{\epsilon}{2}G^{-1}(q)\nabla U(q) + \mathcal{O}(\epsilon^2) \tag{83}
\]

Now we expand the update to position.

\[
\tilde{q} = q + \epsilon \bar{v} \tag{84}
\]

\[
= q + \epsilon v + \mathcal{O}(\epsilon^2) \tag{85}
\]
We conclude by expanding the second update to velocity in terms of the initial conditions.

\[
\tilde{v} = (\text{Id}_m + \Omega(\epsilon, \tilde{q}, \tilde{v}))^{-1} \left( \tilde{v} - \frac{\epsilon}{2} G^{-1}(\tilde{q}) \nabla U(\tilde{q}) \right) \tag{86}
\]

\[
\tilde{v} = \tilde{v} - \Omega(\epsilon, \tilde{q}, \tilde{v}) - \frac{\epsilon}{2} G^{-1}(\tilde{q}) \nabla U(\tilde{q}) + \mathcal{O}(\epsilon^2) \tag{87}
\]

\[
\tilde{v} = \tilde{v} - \Omega(\epsilon, q, v) - \frac{\epsilon}{2} G^{-1}(q) \nabla U(q) + \mathcal{O}(\epsilon^2) \tag{88}
\]

\[
\tilde{v} = v - \Omega(2\epsilon, q, v) v - \epsilon G^{-1}(q) \nabla U(q) + \mathcal{O}(\epsilon^2). \tag{89}
\]

Therefore, a single step of the numerical integrator with step-size \(\epsilon\) agrees with the analytical solution to the initial value problem (from the same initial condition) to at least first order in \(\epsilon\). \(\square\)

**Proof of Lemma 4.2.** To demonstrate that the inverted integrator has at least first-order error, we expand the steps of the integrator as follows. First,

\[
\tilde{q} = q + \frac{\epsilon}{2} v \tag{90}
\]

\[
\tilde{v} = (\text{Id}_m + \Omega(2\epsilon, \tilde{q}, v))^{-1} \left( v - \epsilon G^{-1}(\tilde{q}) \nabla U(\tilde{q}) \right) \tag{91}
\]

\[
\tilde{v} = (\text{Id}_m + \Omega(2\epsilon, \tilde{q}, v)) \left( v - \epsilon G^{-1}(\tilde{q}) \nabla U(\tilde{q}) \right) + \mathcal{O}(\epsilon^2) \tag{92}
\]

\[
\tilde{v} = v - \Omega(2\epsilon, q, v) v - \epsilon G^{-1}(q) \nabla U(q) + \mathcal{O}(\epsilon^2) \tag{93}
\]

\[
\tilde{q} = \tilde{q} + \frac{\epsilon}{2} \tilde{v} \tag{94}
\]

\[
\tilde{q} = q + \frac{\epsilon}{2} v + \frac{\epsilon}{2} v + \mathcal{O}(\epsilon^2) \tag{95}
\]

\[
\tilde{q} = q + \epsilon v. \tag{96}
\]

This verifies that the inverted Lagrangian leapfrog has at least first order accuracy.

In order to show that the inverted leapfrog integrator is symmetric, we proceed as follows. Recall that the three steps of the inverted Lagrangian leapfrog are

\[
\tilde{q} = q + \frac{\epsilon}{2} v \tag{97}
\]

\[
\tilde{v} = (\text{Id}_m + \Omega(2\epsilon, \tilde{q}, v))^{-1} \left( v - \epsilon G^{-1}(\tilde{q}) \nabla U(\tilde{q}) \right) \tag{98}
\]

\[
\tilde{q} = \tilde{q} + \frac{\epsilon}{2} \tilde{v}. \tag{99}
\]
Therefore, we consider integrating from initial position \((\tilde{q}, \tilde{v})\) with a negated step-size \(-\epsilon\) as follows:

\[
\tilde{q}' = \tilde{q} - \frac{\epsilon}{2} \tilde{v} \tag{100}
\]

\[
= \tilde{q} \tag{101}
\]

For the velocity we have,

\[
\tilde{v}' = (\text{Id}_m - \Omega(2\epsilon, \tilde{q}, \tilde{v}))^{-1} \left( \tilde{v} + \epsilon G^{-1}(\tilde{q})\nabla U(\tilde{q}) \right) \tag{102}
\]

\[
\implies (\text{Id}_m - \Omega(2\epsilon, \tilde{q}, \tilde{v})) \tilde{v}' = \tilde{v} + \epsilon G^{-1}(\tilde{q})\nabla U(\tilde{q}) \tag{103}
\]

\[
\implies \tilde{v} + \Omega(2\epsilon, \tilde{q}, \tilde{v})\tilde{v} = \tilde{v}' - \epsilon G^{-1}(\tilde{q})\nabla U(\tilde{q}) \tag{104}
\]

\[
\implies \tilde{v} = (\text{Id}_m + \Omega(2\epsilon, \tilde{q}, \tilde{v}))^{-1} \left( \tilde{v}' - \epsilon G^{-1}(\tilde{q})\nabla U(\tilde{q}) \right) \tag{105}
\]

\[
\implies \tilde{v}' = \tilde{v} \tag{106}
\]

Finally, the last update to the position is,

\[
\tilde{q}' = \tilde{q} - \frac{\epsilon}{2} v \tag{107}
\]

\[
= q. \tag{108}
\]

Hence we see that the inverted Lagrangian leapfrog is also self-adjoint. As noted in the main text, the combination of self-adjointness and at least first-order accuracy immediately imply second-order accuracy.

**Proof of lemma 4.6** Self-adjointness of a numerical integrator follows immediately from the condition \(\Phi_{-\epsilon} \circ \Phi_{\epsilon} = \text{Id}\). Therefore, to demonstrate that a numerical method is self-adjoint it suffices to establish this condition. Consider the first update to the velocity:

\[
\tilde{v} = (\text{Id}_m + \Omega(\epsilon, q, v))^{-1} \left( v - \frac{\epsilon}{2} G(q)^{-1}\nabla U(q) \right) \tag{109}
\]

\[
\iff \tilde{v} + \Omega(\epsilon, q, v)\tilde{v} = v - \frac{\epsilon}{2} G(q)^{-1}\nabla U(q) \tag{110}
\]

\[
\iff \tilde{v} + \frac{\epsilon}{2} G(q)^{-1}\nabla U(q) = v - \Omega(\epsilon, q, \tilde{v})v \tag{111}
\]

\[
\iff \tilde{v} + \frac{\epsilon}{2} G(q)^{-1}\nabla U(q) = (\text{Id}_m - \Omega(\epsilon, q, \tilde{v}))v \tag{112}
\]

\[
\iff v = (\text{Id}_m - \Omega(\epsilon, q, \tilde{v}))^{-1} \left( \tilde{v} + \frac{\epsilon}{2} G^{-1}(q)\nabla U(q) \right). \tag{113}
\]
An identical series of computations reveals,
\[
\dot{v} = (\text{Id}_m - \Omega(\epsilon, \tilde{q}, \tilde{v}))^{-1} \left( \tilde{v} + \frac{\epsilon}{2} \mathbf{G}^{-1}(\tilde{q}) \nabla U(\tilde{q}) \right).
\] (114)

Hence, applying $\Phi_{-\epsilon}$ to $(\tilde{q}, \tilde{v})$ yields the following series of updates,
\[
\dot{v}' = (\text{Id}_m - \Omega(\epsilon, \tilde{q}, \tilde{v}))^{-1} \left( \tilde{v} + \frac{\epsilon}{2} \mathbf{G}^{-1}(\tilde{q}) \nabla U(\tilde{q}) \right)
\] (115)
\[
= \dot{v}
\] (116)
\[
q' = \tilde{q} - \epsilon \tilde{v}'
\] (117)
\[
= \tilde{q} - \epsilon \dot{v}
\] (118)
\[
= q
\] (119)
\[
v' = (\text{Id}_m - \Omega(\epsilon, q, \tilde{v}))^{-1} \left( \tilde{v} + \frac{\epsilon}{2} \mathbf{G}^{-1}(q) \nabla U(q) \right)
\] (120)
\[
= v.
\] (121)

Hence we return to the initial condition $(q, v)$. This verifies that the explicit integrator employed in LMC is self-adjoint.

$\square$

**Proof of Proposition B.13** Since $\hat{\Phi}_\epsilon$ is $p$-th order accurate for $\Phi_\epsilon$ we have, by Taylor series expansion, that,

\[
\hat{\Phi}_0 = \Phi_0
\] (122)
\[
\frac{d^k}{d\epsilon^k} \hat{\Phi}_\epsilon(z_0) \bigg|_{\epsilon=0} = \frac{d^k}{d\epsilon^k} \Phi_\epsilon(z_0) \bigg|_{\epsilon=0},
\] (123)

for $k = 1, \ldots, p$. Let $f : \mathbb{R} \rightarrow \mathbb{R}^m$, then the Taylor series of expansion of $\Xi \circ f(\epsilon)$ is,

\[
\Xi \circ f(\epsilon) = \Xi \circ f(0) + \sum_{k=1}^{p} \epsilon^k g^{[k]} \left( f(0), \frac{d}{d\epsilon} f(0), \ldots, \frac{d^k}{d\epsilon^k} f(0) \right) + \mathcal{O}(\epsilon^{p+1}),
\] (124)

where $g^{[k]}$ are functions determining the Taylor series coefficients which depend on the derivatives of $f$. For instance,

\[
g^{[1]} \left( f(0), \frac{d}{d\epsilon} f(0) \right) = \nabla \Xi(f(0)) \cdot \frac{d}{d\epsilon} f(0)
\] (125)
\[
g^{[2]} \left( f(0), \frac{d}{d\epsilon} f(0), \frac{d^2}{d\epsilon^2} f(0) \right) = \nabla^2 \Xi(f(0)) \left( \frac{d}{d\epsilon} f(0), \frac{d}{d\epsilon} f(0) \right) + \nabla \Xi(f(0)) \cdot \frac{d^2}{d\epsilon^2} f(0).
\] (126)
By the equality of the Taylor series expansion coefficients in eqs. (122) and (123) it follows from eq. (124) that,

\[ \Xi \circ \Phi_\epsilon(z_0) - \Xi \circ \hat{\Phi}_\epsilon(z_0) = \mathcal{O}(\epsilon^{p+1}). \]  

(127)

This proves that \( \Xi \circ \hat{\Phi}_\epsilon \) is also \( p \)-th order accurate for \( \Xi \circ \Phi_\epsilon \). \qed
E Propagator Matrices for the Leapfrog and Inverted Leapfrog

**Definition E.1.** The generalized leapfrog integrator for the Hamiltonian equations of motion in eqs. (40) and (41) with Hamiltonian $H(q, p) = U(q) + p^\top p/2$ is a map $(q, p) \mapsto (\tilde{q}, \tilde{p})$ defined by,

\[
\begin{align*}
\tilde{p} &= p - \frac{\epsilon}{2} \nabla U(q) \\
\tilde{q} &= q + \epsilon \tilde{p} \\
\tilde{p} &= \tilde{p} - \frac{\epsilon}{2} \nabla U(\tilde{q}).
\end{align*}
\]

**Definition E.2.** The generalized leapfrog integrator for the Hamiltonian equations of motion in eqs. (40) and (41) with Hamiltonian $H(q, p) = U(q) + p^\top p/2$ is a map $(q, p) \mapsto (\tilde{q}, \tilde{p})$ defined by,

\[
\begin{align*}
\tilde{q} &= q + \frac{\epsilon}{2} p \\
\tilde{p} &= p - \epsilon \nabla U(q) \\
\tilde{q} &= \tilde{q} + \frac{\epsilon}{2} \tilde{p}.
\end{align*}
\]

For Hamiltonians of the form $H(q, p) = \omega^2 q^2/2 + p^2/2$, the action of the leapfrog and inverted leapfrog integrators are linear. This means that there are matrices, called “propagator matrices,” which, when acting on the vector $(q, p) \in \mathbb{R}^2$, produce the same position in phase space as the integrators themselves. Computing integer matrix powers of these matrices can then produce the multi-step output of the integrators. In the case of the leapfrog integrator, the propagator matrix is (Leimkuhler and Reich, 2005),

\[
\begin{pmatrix}
\tilde{q} \\
\tilde{p}
\end{pmatrix} = \begin{pmatrix}
1 - \frac{\epsilon^2 \omega^2}{2} & \epsilon \\
\epsilon \omega^2 \left(1 - \frac{\epsilon^2 \omega^2}{4}\right) & 1 - \frac{\epsilon^2 \omega^2}{2}
\end{pmatrix} \begin{pmatrix}
q \\
p
\end{pmatrix}.
\]

The propagator matrix for the inverted leapfrog is,

\[
\begin{pmatrix}
\tilde{q} \\
\tilde{p}
\end{pmatrix} = \begin{pmatrix}
1 - \frac{\epsilon^2 \omega^2}{2} & \epsilon \left(1 - \frac{\epsilon^2 \omega^2}{4}\right) \\
-\epsilon \omega^2 & 1 - \frac{\epsilon^2 \omega^2}{2}
\end{pmatrix} \begin{pmatrix}
q \\
p
\end{pmatrix}.
\]
At stationarity,

\[
\begin{pmatrix}
q \\
p
\end{pmatrix}
\sim \text{Normal}
\begin{pmatrix}
\begin{pmatrix}
0 \\
0
\end{pmatrix},
\begin{pmatrix}
1/\omega^2 & 0 \\
0 & 1
\end{pmatrix}
\end{pmatrix}.
\]

(136)

Therefore, for an integrator with propagator matrix \( R \), the distribution of the \( k \)-step transition is,

\[
\text{Normal}
\begin{pmatrix}
\begin{pmatrix}
0 \\
0
\end{pmatrix},
R^k
\begin{pmatrix}
1/\omega^2 & 0 \\
0 & 1
\end{pmatrix}(R^k)^\top
\end{pmatrix},
\]

(137)
F Involution Monte Carlo

Proof of Proposition B.17. The detailed balance condition states that for any Borel sets 
\((A, B), (C, D) \in \mathcal{B}(\mathbb{R}^m \times \mathbb{R}^m)\) we have,

\[
\int_{(A, B)} K((q, p), (C, D)) \pi(q, p) \, dq \, dp = \int_{(C, D)} K((q, p), (A, B)) \pi(q, p) \, dq \, dp. \tag{138}
\]

Let \(\pi(q, p)\) be the probability density on \(\mathbb{R}^m \times \mathbb{R}^m\) that is proportional to \(\exp(-H(q, p))\).

Using the fact that \(\Phi\) is an involution, it suffices to verify

\[
\int_{(A, B)} \alpha((q, p), \Phi(q, p), |\det (\nabla \Phi(x))|) 1 \{((\tilde{q}, \tilde{p}) \in (C, D)\} \pi(q, p) \, dq \, dp \tag{139}
\]

\[
= \int_{(A, B)} \min \left\{ 1, \frac{\pi(\Phi(q, p))}{\pi(q, p)} |\det (\nabla \Phi(x))| \right\} 1 \{\Phi(q, p) \in (C, D)\} \pi(q, p) \, dq \, dp \tag{140}
\]

\[
= \int_{(A, B)} \min \{\pi(q, p), \pi(\Phi(q, p)) |\det (\nabla \Phi(x))| \} 1 \{\Phi(q, p) \in (C, D)\} \pi(q, p) \, dq \, dp \tag{141}
\]

\[
= \int_{(A, B)} \min \left\{ \frac{\pi(q, p)}{\pi(\Phi(q, p))} |\det (\nabla \Phi(x))|, 1 \right\} 1 \{\Phi(q, p) \in (C, D)\} \pi(q, p) |\det (\nabla \Phi(x))| \, dq \, dp \tag{142}
\]

\[
= \int_{\Phi(A, B)} \min \left\{ 1, \frac{\pi(\Phi(\tilde{q}, \tilde{p}))}{\pi(\tilde{q}, \tilde{p})} |\det (\nabla \Phi(\tilde{q}, \tilde{p}))| \right\} 1 \{((\tilde{q}, \tilde{p}) \in (C, D)\} \pi(\tilde{q}, \tilde{p}) \, d\tilde{q} \, d\tilde{p} \tag{143}
\]

\[
= \int_{(C, D)} \alpha((q, p), \Phi(q, p), |\det (\nabla \Phi(q, p))|) 1 \{(q, p) \in \Phi(A, B)\} \pi(q, p) \, dq \, dp \tag{145}
\]

\[
= \int_{(C, D)} \alpha((q, p), \Phi(q, p), |\det (\nabla \Phi(q, p))|) 1 \{(q, p) \in (A, B)\} \pi(q, p) \, dq \, dp \tag{146}
\]

Moreover, integrating over the rejection components of the transition kernel already have symmetry in \((A, B)\) and \((C, D)\):

\[
\int_{(A, B)} (1 - \alpha((q, p), (\tilde{q}, \tilde{p}), |\det (\nabla \Phi(x))|)) 1 \{(q, p) \in (C, D)\} \pi(q, p) \, dq \, dp \tag{147}
\]

\[
= \int_{(C, D)} (1 - \alpha((q, p), (\tilde{q}, \tilde{p}), |\det (\nabla \Phi(x))|)) 1 \{(q, p) \in (A, B)\} \pi(q, p) \, dq \, dp. \tag{148}
\]

Detailed balance follows as a consequence. \qed
Lemma F.1. Let $\Phi_\epsilon : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m \times \mathbb{R}^m$ be the Lagrangian leapfrog integrator with step-size $\epsilon \in \mathbb{R}$. Then $F \circ \Phi_\epsilon$ is an involution.

Proof. Let $(\tilde{q}, \tilde{v}) = \Phi_\epsilon(q, v)$. Now we apply the Lagrangian leapfrog to $(\tilde{q}, -\tilde{v})$.

\[
\tilde{v}'' = (\text{Id}_m + \Omega(\epsilon, \tilde{q}, -\tilde{v}))^{-1} \left(-\tilde{v} - \frac{\epsilon}{2} G^{-1}(\tilde{q}) \nabla U(\tilde{q})\right) 
\]

from eq. (115). Moreover,

\[
\tilde{q}'' = \tilde{q} + \epsilon \tilde{v}'' 
\]

\[
= \tilde{q} - \epsilon \tilde{v}
\]

\[
= q. 
\]

Finally, \[\tilde{v}'' = (\text{Id}_m + \Omega(\epsilon, q, -v))^{-1} \left(-v - \frac{\epsilon}{2} G^{-1}(q) \nabla U(q)\right) \]

from eq. (120). \hfill \square

References

S.-i. Amari. *Information Geometry and Its Applications*. Springer Publishing Company, Incorporated, 1st edition, 2016. ISBN 4431559779.

M. Betancourt. A conceptual introduction to hamiltonian monte carlo, 2017. URL http://arxiv.org/abs/1701.02434 cite arxiv:1701.02434Comment: 60 pages, 42 figures.

M. Betancourt, S. Byrne, S. Livingstone, and M. Girolami. The geometric foundations of hamiltonian monte carlo. *Stat. Sci.*, 23, 10 2014. doi: 10.3150/16-BEJ810.
N. Bou-Rabee and J. M. Sanz-Serna. Geometric integrators and the hamiltonian monte carlo method. *Acta Numerica*, 27:113–206, 2018.

J. Brofos and R. R. Lederman. Evaluating the implicit midpoint integrator for riemannian hamiltonian monte carlo. In M. Meila and T. Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 1072–1081. PMLR, 18–24 Jul 2021a. URL [https://proceedings.mlr.press/v139/brofos21a.html](https://proceedings.mlr.press/v139/brofos21a.html).

J. A. Brofos and R. R. Lederman. On numerical considerations for riemannian manifold hamiltonian monte carlo, 2021b.

O. Calin and D.-C. Chang. *Geometric Mechanics on Riemannian Manifolds: Applications to Partial Differential Equations*. Birkhäuser, 2004.

A. D. Cobb, A. G. Baydin, A. Markham, and S. J. Roberts. Introducing an explicit symplectic integration scheme for riemannian manifold hamiltonian monte carlo, 2019.

G. Dahlquist and Å. Björck. *Numerical Methods*. Dover Books on Mathematics. Dover Publications, 2003. ISBN 9780486428079. URL [https://books.google.com/books?id=armfeHkJlwAC](https://books.google.com/books?id=armfeHkJlwAC).

S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth. Hybrid monte carlo. *Physics Letters B*, 195(2):216 – 222, 1987. ISSN 0370-2693. URL [http://www.sciencedirect.com/science/article/B6TVN-46YSWPH-2XF/2/0f89cdc6cf214a2169b03df7414f3df4](http://www.sciencedirect.com/science/article/B6TVN-46YSWPH-2XF/2/0f89cdc6cf214a2169b03df7414f3df4).

A. Gelman and C. Pasarica. Adaptively scaling the metropolis algorithm using expected squared jumped distance. *Statistica Sinica*, 2007. URL [http://dx.doi.org/10.2139/ssrn.1010403](http://dx.doi.org/10.2139/ssrn.1010403).

M. Girolami and B. Calderhead. Riemann manifold langevin and hamiltonian monte carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(2):123–214, 2011. URL [https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2010.00765.x](https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2010.00765.x).
E. Hairer, C. Lubich, and G. Wanner. *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations; 2nd ed.* Springer, Dordrecht, 2006. URL https://cds.cern.ch/record/1250576.

R. Kumar, C. Carroll, A. Hartikainen, and O. A. Martin. ArviZ a unified library for exploratory analysis of Bayesian models in Python. *The Journal of Open Source Software*, 2019. URL http://joss.theoj.org/papers/10.21105/joss.01143.

S. Lan, V. Stathopoulos, B. Shahbaba, and M. Girolami. Markov chain monte carlo from lagrangian dynamics. *Journal of Computational and Graphical Statistics*, 24(2): 357–378, Apr 2015. ISSN 1537-2715. URL http://dx.doi.org/10.1080/10618600.2014.902764.

B. Leimkuhler and S. Reich. *Simulating Hamiltonian Dynamics*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, 2005.

J. E. Marsden and T. S. Ratiu. *Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems*. Springer Publishing Company, Incorporated, 2010. ISBN 1441931430.

J. E. Marsden and M. West. Discrete mechanics and variational integrators. *Acta Numerica*, 10:357–514, 2001.

C. Modi, A. Barnett, and B. Carpenter. Delayed rejection hamiltonian monte carlo for sampling multiscale distributions, 2021.

K. Neklyudov, M. Welling, E. Egorov, and D. P. Vetrov. Involution MCMC: a unifying framework. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, volume 119 of *Proceedings of Machine Learning Research*, pages 7273–7282. PMLR, 2020. URL http://proceedings.mlr.press/v119/neklyudov20a.html.

A. A. Pourzanjani and L. R. Petzold. Implicit hamiltonian monte carlo for sampling multiscale distributions, 2019.
Y. Zhang and C. Sutton. Semi-separable hamiltonian monte carlo for inference in bayesian hierarchical models. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014. URL https://proceedings.neurips.cc/paper/2014/file/a87ff679a2f3e71d9181a67b7542122c-Paper.pdf.