Evaluating the Implicit Midpoint Integrator for Riemannian Manifold Hamiltonian Monte Carlo

James A. Brofos [1] **Roy R. Lederman** [1]

Abstract

Riemannian manifold Hamiltonian Monte Carlo is traditionally carried out using the generalized leapfrog integrator. However, this integrator is not the only choice and other integrators yielding valid Markov chain transition operators may be considered. In this work, we examine the implicit midpoint integrator as an alternative to the generalized leapfrog integrator. We discuss advantages and disadvantages of the implicit midpoint integrator for Hamiltonian Monte Carlo, its theoretical properties, and an empirical assessment of the critical attributes of such an integrator for Hamiltonian Monte Carlo: energy conservation, volume preservation, and reversibility. Empirically, we find that while leapfrog iterations are faster, the implicit midpoint integrator has better energy conservation, leading to higher acceptance rates, as well as better conservation of volume and better reversibility, arguably yielding a more accurate sampling procedure.

1. Introduction

Riemannian manifold Hamiltonian Monte Carlo (RMHMC) is a powerful algorithm for sampling from Bayesian posterior distributions (Girolami & Calderhead, 2011). Given a log-posterior function $\mathcal{L} : \mathbb{R}^m \to \mathbb{R}$ and a Riemannian metric $\mathbb{G} : \mathbb{R}^m \to \mathbb{R}^{m \times m}$ (with the condition that $\mathbb{G}(q)$ is positive definite for each $q \in \mathbb{R}^m$), RMHMC considers the Hamiltonian dynamics corresponding to the Hamiltonian,

$$H(q,p) = -\mathcal{L}(q) + \frac{p^\top \mathbb{G}^{-1}(q) p}{2} + \frac{\log\det(\mathbb{G}(q))}{2}. \quad (1)$$

Riemannian metrics are incorporated into HMC in order to precondition dynamics and more efficiently explore the distribution. Irrespective of the choice of metric, the form of the Hamiltonian in eq. (1) corresponds to a Gibbs distribution proportional to $\exp(-H(q,p))$ and tractable conditional distribution $p|q \sim \text{Normal}(0, \mathbb{G}(q))$. However, the form of this Hamiltonian is such that it *cannot* be written as the sum of two functions, each a function of q or p alone; such a Hamiltonian is called "non-separable." The presence of a non-separable Hamiltonian presents unique challenges for numerical integration.

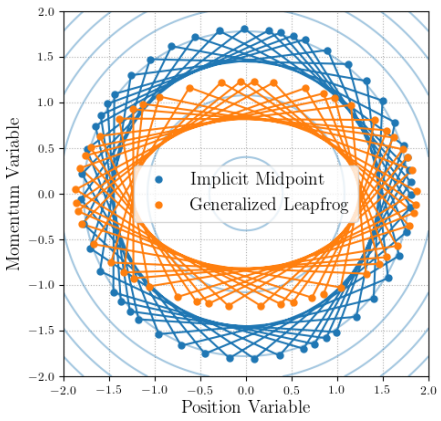


Figure 1. Trajectories computed by the leapfrog integrator and the implicit midpoint integrator for a quadratic, separable Hamiltonian $H(q,p) = q^\top q/2 + p^\top p/2$. Both integrators are stable but the leapfrog integrator deviates from the level sets of the Hamiltonian (faint blue circles) whereas every iterate of the implicit midpoint integrator lies on the same energy level set as the previous one.

The leapfrog integrator and its variants are a ubiquitous choice for the numerical integration of Hamiltonian mechanics for HMC; for instance see (Betancourt, 2012; Brubaker et al., 2012; Byrne & Girolami, 2013; Girolami & Calderhead, 2011; Neal, 2010; Tripuraneni et al., 2017), among many others. It may, therefore, not be apparent that numerical integrators other than the leapfrog method are applicable to HMC, provided that they exhibit two properties:

(i) The integrator has a unit Jacobian determinant so that it preserves volume in (q,p)-space.

(ii) The integrator is symmetric under negation of the momentum variable.

[1]Department of Statistics and Data Science, Yale University. Correspondence to: James A. Brofos <james.brofos@yale.edu>.

Proceedings of the 38^{th} International Conference on Machine Learning, PMLR 139, 2021.

These properties are sufficient to prove that HMC satisfies detailed balance, which in turn establishes that the stationary distribution has density proportional to $\exp(-H(q,p))$; see Bishop (2006); Neal (2010), or appendix H for a proof.

Why does the choice of numerical integrator matter? There are at least three reasons.

(a) Numerical integrators differ with respect to energy conservation and stability. The acceptance probability of HMC depends on the energy conservation and the ability of the HMC proposal to use large integration steps depends on stability.

(b) Numerical integrators may only satisfy properties (i) and (ii) above *approximately*, particularly if the integrators are defined as solutions to implicitly-defined equations. These approximate solvers of implicit equations will be discussed in more details below. Typically, the error of these methods will depend on a convergence tolerance δ used to find fixed-points of the integration step (algorithm 1). For a non-zero convergence tolerance, the degree to which properties (i) and (ii) are violated will depend on the integrator and the tolerance.

(c) Numerical integrators will differ in their efficiency in the sense that there may be structural properties of the Hamiltonian system that the integrator exploits. More efficient integrators will exhibit higher effective sample sizes *per second* when used in HMC.

The contribution of this work is to compare and contrast the generalized leapfrog integrator with the implicit midpoint method in application to RMHMC. We consider RMHMC because the non-separable Hamiltonian necessitates elaborate integration schemes which require solving implicitly-defined equations; this is in contrast to Euclidean HMC with constant $\mathbb{G}$ which produces a separable Hamiltonian that can be integrated explicitly. First, we compare the two integrators on the energy conservation, volume preservation, and reversibility as discussed in reason (a). Second, we study the breakdown of *exact* satisfaction of properties (i) and (ii) in implicitly-defined integrators as described in reason (b). Third, we consider multiple variants of the generalized leapfrog and implicit midpoint integrators that exhibit different efficiencies, relevant to reason (c). We conclude that the implicit midpoint integrator exhibits superior energy conservation, conservation of volume, and symmetry compared to the generalized leapfrog integrator. We explore inference in sophisticated Bayesian inference tasks wherein the implicit midpoint integrator is competitive with, or exceeds, the time-normalized performance of the generalized leapfrog method. We therefore argue that the implicit midpoint integrator is a procedure worth consideration in RMHMC.

2. Background

For Bayesian inference tasks, $\mathcal{L}$ is the sum of the log-likelihood and the log-prior; in this circumstance, the typical form of the Riemannian metric $\mathbb{G}$ is the sum of the Fisher information of the log-likelihood and negative Hessian of the log-prior; this choice of Riemannian metric is motivated by information geometry (Amari, 2016). The Hamiltonian in eq. (1) leads to the equations of motion,

$$\dot{q}_i = \sum_{j=1}^{m} \mathbb{G}^{-1}_{ij}(q) p_j \tag{2}$$

$$\begin{aligned} \dot{p}_i = &-\frac{\partial}{\partial q_i}\mathcal{L}(q) - \frac{1}{2}\text{trace}\left(\mathbb{G}^{-1}(q)\frac{\partial}{\partial q_i}\mathbb{G}(q)\right) \\ &+ \frac{1}{2}p^\top \mathbb{G}^{-1}(q)\frac{\partial}{\partial q_i}\mathbb{G}(q)\mathbb{G}^{-1}(q)p \end{aligned} \tag{3}$$

As stated in section 1, the standard integrator for RMHMC is the (generalized) leapfrog integrator. A naive implementation of a single step of the generalized leapfrog integrator with step-size ϵ and initial position (q,p) is presented in algorithm 2. Notice that eqs. (7) and (8) are *implicitly defined* in the sense that the quantities appearing on the left-hand side also appear on the right-hand side; these equations are typically solved to a given tolerance $\delta \geq 0$ (in the sense defined in the fixed point iteration algorithm algorithm 1). When $\delta = 0$, the generalized leapfrog integrator satisfies properties (i) and (ii), however, in practice, the tolerance is often chosen to be larger than machine precision in order to reduce the number of fixed point iterations; therefore properties (i) and (ii) are no longer satisfied accurately.

The implicit midpoint method, an alternative to the generalize leapfrog integrator, is presented in algorithm 3; the implicit midpoint integrator also involves the solution to an implicitly-defined eq. (10). When $\delta = 0$, it is well-known that the implicit midpoint integrator satisfies property (i); see Leimkuhler & Reich (2005). It also satisfies property (ii) for Hamiltonians of the form eq. (1); see appendix A.

We turn now to discussing a theoretical property of numerical integrators related to conserved quantities.

Definition 1. Let $z = (q,p)$ where the time evolution of q_i is given by eq. (2) and of p_i by eq. (3) for $i = 1, \ldots, m$. A conserved quantity of z is a real-valued function $z \mapsto \mathfrak{G}(z)$ for which $\frac{\mathrm{d}}{\mathrm{d}t}\mathfrak{G}(z) = 0$.

It is important to notice that definition 1 is a statement about the underlying dynamics and *has nothing to do with the integrator used to approximate these dynamics*. The properties of the integrators will be discussed in the next paragraph. For Hamiltonian systems, the canonical example of a conserved quantity is the Hamiltonian energy itself; see Marsden & Ratiu (2010). Hamiltonian flows are also

Algorithm 1 (Fixed Point Iteration) Procedure for solving the equation $z = f(z)$ via fixed point interation to a given tolerance.

1: **Input**: Function $f : \mathbb{R}^m \to \mathbb{R}^m$, initial guess $z \in \mathbb{R}^m$, fixed point convergence tolerance $\delta \geq 0$.
2: Set $\Delta z = \infty$ and $z' = z$.
3: **While**: $\Delta z > \delta$ compute

$$z'' = f(z') \tag{4}$$
$$\Delta z = \max_{i \in \{1,\ldots,m\}} |z''_i - z'_i| \tag{5}$$
$$z' = z'' \tag{6}$$

4: **Return**: $z' \in \mathbb{R}^m$.

Algorithm 2 (G.L.F.(a)) The procedure for a single step of integrating Hamiltonian dynamics using the generalized leapfrog integrator.

1: **Input**: Hamiltonian $H : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$, initial position and momentum variables $(q, p) \in \mathbb{R}^m \times \mathbb{R}^m$, integration step-size size $\epsilon \in \mathbb{R}$, fixed-point convergence tolerance $\delta \geq 0$.
2: Use algorithm 1 with tolerance δ and initial guess p to solve for $\bar{p}$,

$$\bar{p} \overset{\text{def.}}{=} \underbrace{p - \frac{\epsilon}{2}\nabla_q H(q, \bar{p})}_{f(\bar{p})} \tag{7}$$

3: Use algorithm 1 with tolerance δ and initial guess q to solve for q',

$$q' \overset{\text{def.}}{=} \underbrace{q + \frac{\epsilon}{2}\left(\nabla_p H(q, \bar{p}) + \nabla_p H(q', \bar{p})\right)}_{f(q')} \tag{8}$$

4: Compute the explicit update

$$p' \overset{\text{def.}}{=} \bar{p} - \frac{\epsilon}{2}\nabla_q H(q', \bar{p}) \tag{9}$$

5: **Return**: $(q', p') \in \mathbb{R}^m \times \mathbb{R}^m$.

symplectic (Hairer et al., 2006) which implies conservation of volume (in the same sense as that of property (i)).

A numerical integrator cannot preserve all of the conserved quantities as the underlying ODE, but it may be able to conserve some simple ones. The following two results may be found in Leimkuhler & Reich (2005).

Theorem 1. Let $z = (q, p)$. The generalized leapfrog integrator (algorithm 2) with $\delta = 0$ preserves any conserved quantity of the form $\mathfrak{G}(z) = q^\top \mathbf{A} p + \mathbf{b}^\top z$ where $\mathbf{A} \in \mathbb{R}^{m \times m}$ is a symmetric matrix and $\mathbf{b} \in \mathbb{R}^{2m}$.

Theorem 2. Let $z = (q, p)$. The implicit midpoint integrator (algorithm 3) with $\delta = 0$ preserves any conserved quantity of the form $\mathfrak{G}(z) = z^\top \mathbf{A} z + \mathbf{b}^\top z$ where $\mathbf{A} \in \mathbb{R}^{2m \times 2m}$ is a symmetric matrix and $\mathbf{b} \in \mathbb{R}^{2m}$.

Algorithm 3 (I.M.(a)) The procedure for a single step of integrating Hamiltonian dynamics using the implicit midpoint integrator.

1: **Input**: Hamiltonian $H : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$, initial position and momentum variables $(q, p) \in \mathbb{R}^m \times \mathbb{R}^m$, integration step-size size $\epsilon \in \mathbb{R}$, fixed-point convergence tolerance $\delta \geq 0$.
2: Use algorithm 1 with tolerance δ and initial guess (q, p) to solve for (q', p')

$$\begin{pmatrix} q' \\ p' \end{pmatrix} \overset{\text{def.}}{=} \underbrace{\begin{pmatrix} q \\ p \end{pmatrix} + \epsilon \begin{pmatrix} \nabla_p H(\bar{q}, \bar{p}) \\ -\nabla_q H(\bar{q}, \bar{p}) \end{pmatrix}}_{f(q', p')} \tag{10}$$

where $\bar{q} \overset{\text{def.}}{=} (q' + q)/2$ and $\bar{p} \overset{\text{def.}}{=} (p' + p)/2$.
3: **Return**: $(q', p') \in \mathbb{R}^m \times \mathbb{R}^m$.

Notice that theorem 2 contains a strictly more general class of conserved quantity than theorem 1. We come now to a hypothesis that would justify the consideration of the implicit midpoint integrator within the context of HMC. Before stating the hypothesis, we provide some initial motivation for how the implicit midpoint integrator performs in the presence of a quadratic Hamiltonian.

Proposition 1. Let $H(q, p) \equiv H(z) = z^\top \mathbf{A} z$ be a quadratic Hamiltonian. Then, *for any step-size*, the proposals generated by Hamiltonian Monte Carlo using the implicit midpoint integrator with $\delta = 0$ will be accepted.

A proof is given in appendix J. Note, however, that perfect conservation of the Hamiltonian energy does *not* imply that the implicit midpoint integrator is the exact solution of the Hamilton's equations of motion. Nevertheless, Proposition 1 suggests an important difference between the generalized leapfrog integrator and the implicit midpoint method in terms of their conservation properties. Although Bayesian posterior distributions are unlikely to be Gaussian, it is widely accepted that Gaussian approximations are useful. Such notions materialize, for example, in the central limit theorem and the Laplace approximation. We therefore speculate that the Gaussian case may be useful for providing intuition for the more general case we examine in our experimental results. Provided the posterior is approximately Gaussian, therefore, this leads us to the following hypothesis.

Hypothesis. The implicit midpoint algorithm will exhibit higher acceptance probabilities than the generalized leapfrog integrator for the same step-size.

If true, and if the fixed point iterations required by the implicit midpoint procedure are not too burdensome relative to the generalized leapfrog integrator, then the higher acceptance rate may produce more favorable effective sample sizes for the Markov chain whose transitions are computed using the implicit midpoint algorithm. In this scenario, the implicit midpoint integrator may be worth consideration as an alternative to the generalized leapfrog integrator.

We wish to emphasize that the behavior of the implicit midpoint integrator in the presence of a quadratic Hamiltonian is not a definitive explanation of all differences in sampling behaviors that may arise when using it as a transition operator in RMHMC. However, we believe that the setting of quadratic Hamiltonians, corresponding to Gaussian densities, can provide helpful intuition. In section 4 we will turn to the empirical evaluation of the implicit midpoint method to examine the extent to which this alternative integrator offers an advantage over the generalized leapfrog method in the non-Gaussian regime.

For a brief introduction to the stability of numerical integrators, see appendix D.

3. Related Work

Most relevant to our discussion is Pourzanjani & Petzold (2019). In this work, the authors examine the relationship between the (non-generalized) leapfrog integrator and the implicit midpoint integrator; the authors make the argument that the implicit midpoint integrator is more stable in the presence of posteriors whose dimensions exhibit large differences in their variability ("multi-scale"). The presence of multi-scale posterior dimensions necessitates a small step-size for the leapfrog integrator, which is found to be unnecessary for the implicit midpoint algorithm. As the authors note, however, "RMHMC uses local Hessian evaluations of the potential energy surface to adaptively change this step-size based on the local curvature;" therefore, their experiments instead focus on the circumstance where a constant mass matrix is utilized, corresponding to Euclidean HMC with no local adaptation of the step-size. Indeed, as observed in Martens (2020), the Fisher information captures the second-order geometry of the posterior and actually exhibits properties that make it preferable to the Hessian of the posterior in optimization. Therefore, the present work differs from Pourzanjani & Petzold (2019) in its focus on Riemannian geometry wherein the metric compensates (at least locally) for multi-scale dimensions; moreover, our *empirical* analysis of reversibility and volume preservation is, to the best of our knowledge, novel. Before proceeding to the experimental results, we note that stability alone *cannot* account for the high acceptance rate enjoyed by the implicit midpoint integrator: even in the regime wherein the generalized leapfrog integrator is stable, it is not able to perfectly conserve the Hamiltonian energy as the implicit midpoint integrator does. This phenomenon is visualized in fig. 1.

4. Experimental Results

We turn now to evaluating the implicit midpoint integrator in several Bayesian inference tasks. We consider inference in a banana-shaped posterior, sampling from Neal's funnel distribution, a stochastic volatility model, and Bayesian inference in the Fitzhugh-Nagumo differential equation model. We have additional experimental results in our appendices. In appendix E, we seek to verify theorem 2 in the presence of a truly quadratic Hamiltonian. In appendix F, we examine Bayesian inference in a logistic regression posterior. To define a stopping condition for the fixed point iterations used by the implicit midpoint and generalized leapfrog methods, we demand that the change in each coordinate be less than a threshold; we let $\delta \in \left\{1 \times 10^{-9}, 1 \times 10^{-6}, 1 \times 10^{-3}\right\}$ when considering reversibility and volume preservation. When reporting performance metrics such as effective sample size, we report results corresponding to a threshold of $\delta = 1 \times 10^{-6}$. We implemented all methods in 64-bit precision using NumPy and SciPy (Harris et al., 2020; Virtanen et al., 2020). We compute effective sample sizes (ESS) using Kumar et al. (2019). Additional experiments with a randomized number of integration steps are included in appendix K. Code for our experiments can be found at `https://github.com/JamesBrofos/Evaluating-the-Implicit-Midpoint-Integrator`.

4.1. Summary of Integrators

We consider two variants of the generalized leapfrog method and two variants of the implicit midpoint integrator, which we summarily describe as follows.

G.L.F.(a) An implementation of the generalized leapfrog integrator as presented in algorithm 2.

G.L.F.(b) An implementation of the generalized leapfrog integrator that caches repeated calculations and which is specific to Hamiltonians in the form of eq. (1). See algorithm 4 in appendix I. G.L.F.(b) is mathematically identical to G.L.F.(a), but this implementation avoids some redundant computation. Differences between the outputs of G.L.F.(a) and G.L.F.(b) are due to random seeds and machine error in computation.

I.M.(a) An implementation of the implicit midpoint integrator as presented in algorithm 3.

I.M.(b) An implementation of the implicit midpoint integrator that implicitly computes the midpoint followed by an explicit Euler step, as advocated by (Leimkuhler & Reich, 2005). See algorithm 5 in appendix I.

In all of our implementations, we use fixed point iterations in order to find solutions to implicitly-defined relations. This is the approach advocated by Hairer et al. (2006). Additional details are presented in appendix I.

4.2. Banana-Shaped Distribution

The banana-shaped distribution was proposed in a discussion to (Girolami & Calderhead, 2011) as a representative example of the ridge-like posterior structure that can manifest in non-identifiable models. The banana-shaped distribution is defined by the following generative model.

$$y_i|\theta_1,\theta_2 \sim \text{Normal}(\theta_1+\theta_2^2,\sigma_y^2) \quad \text{for } i=1,\ldots,n \tag{11}$$
$$\theta_i \sim \text{Normal}(0,\sigma_\theta^2) \quad \text{for } i\in\{1,2\}\,. \tag{12}$$

For the banana-shaped distribution, the Riemannian metric is

$$\mathbb{G}(\theta_1,\theta_2)=\begin{pmatrix}\frac{n}{\sigma_y^2}+\frac{1}{\sigma_\theta^2} & \frac{2n\theta_2}{\sigma_y^2}\\ \frac{2n\theta_2}{\sigma_y^2} & \frac{4n\theta_2^2}{\sigma_y^2}+\frac{1}{\sigma_\theta^2}\end{pmatrix}. \tag{13}$$

In our experiments, we take $n=100$. We generate observations $\{y_1,\ldots,y_{100}\}$ from the banana-shaped distribution by setting $\theta_1=1/2$, $\theta_2=1/\sqrt{2}$, and $\sigma_y=\sigma_\theta=2$. We then attempt to sample the posterior distribution of (θ_1,θ_2) using RMHMC when integration is performed using the implicit midpoint algorithm or the generalized leapfrog method. We consider two step-sizes $\{0.01,0.1\}$ and a number of integration steps in $\{5,10,50\}$. We attempt to draw 10,000 samples from the posterior. Each of these configurations is replicated ten times.

Results are shown in table 1, demonstrating that the I.M.(a) and (b) integrators are able to maintain high energy conservation at step-sizes for which the G.L.F.(a) and (b) variants cannot. As a consequence, Markov chains using I.M.(a) or I.M.(b) are able to achieve very high effective sample sizes (ESS); moreover, because the cost of evaluating the gradients of the banana-shaped posterior is not too large, these Marko chains also exhibits superior performance *on the timing comparisons*. We find that I.M.(a) and I.M.(b) perform similarly. In addition to energy conservation, an essential component of HMC are volume preservation and reversibility (recall properties (i) and (ii) from section 1). Using the samples drawn by the Markov chains with either integrator, we may compute numerical estimates of the degree to which these properties are satisfied. We give a detailed description of the volume preservation and reversibility metrics in appendix G. We use one-hundred randomly selected samples generated from the Markov chains in order to compute these statistics. Results showing the *median* reversibility versus the *median* difference from unit Jacobian are shown in fig. 2. These results show that the median symmetry and volume preservation of the implicit midpoint integrator is approximately an order of magnitude more faithfully preserved than is the case for the generalized leapfrog method.

4.3. Hierarchical Neal's Funnel Distribution

As an example of a hierarchical Bayesian posterior, we consider Neal's funnel distribution defined by,

$$x_i \sim \text{Normal}(0,\exp(-v)) \;\text{ for } i=1,\ldots,10 \tag{14}$$
$$v \sim \text{Normal}(0,9). \tag{15}$$

Due to the hierarchical structure of the distribution, the Hessian of the distribution is not convex and therefore cannot be used to construct a Riemannian metric on its own. Instead, we follow the approach proposed in Betancourt (2012) and adopt the SoftAbs transformation of the Hessian in order to construct a positive definite Riemannian metric. This allows us to sample all variables of the hierarchical distribution jointly. For RMHMC, we consider an integration step-size in $\{0.1,0.2,0.5\}$ and we attempt to draw 10,000 samples of $(x_1,\ldots,x_{10},v)$ from Neal's funnel distribution.

Results are presented in table 2. For the largest step-size, the I.M.(a) and (b) integrators are able to maintain high acceptance probabilities. Markov chains using the I.M.(a) or (b) method achieve the best *minimum* ESS per-second. This example presents a circumstance wherein the time-normalized best-case performance of mean and minimum effective sample sizes did not co-occur in the same parameter configuration. Nevertheless, when optimizing for the highest *mean* ESS per-second, the I.M.(a) and (b) methods also outperform the G.L.F.(a) integrator. We visualize the symmetry and volume preservation in fig. 3; the implicit midpoint integrator exhibits better symmetry and volume preservation for the same convergence criterion.

4.4. Stochastic Volatility Model

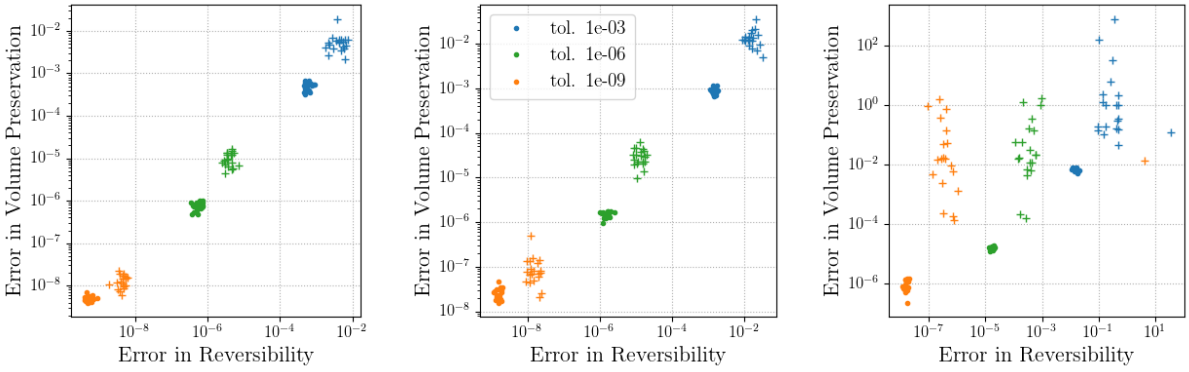
While Neal's funnel is a hierarchical distribution, it is not sampled in a hierarchical manner, instead sampling all variables jointly using the SoftAbs Riemannian metric (Betancourt, 2012). Here, we consider a stochastic volatility model whose posterior includes the stochastic volatilities as well as latent hyperparameters of the model; we will sample these variables using an alternating Gibbs procedure. Following Girolami & Calderhead (2011), the stochastic volatility model is defined, for $t=1,\ldots,T$, by,

$$y_t|\beta,x_t \sim \text{Normal}(0,\beta^2 e^{x_t}) \tag{16}$$
$$x_t|x_{t-1},\phi,\sigma^2 \sim \text{Normal}(\phi x_{t-1},\sigma^2) \tag{17}$$

where $x_1 \sim \text{Normal}(0,\sigma^2/(1-\phi^2))$, $\pi(\beta)\propto 1/\beta$, $\sigma^2 \sim \text{Inv}{-}\chi^2(10,0.05)$, and $(\phi+1)/2 \sim \text{Beta}(20,1.5)$. The sampler proceeds by alternating between sampling the conditional posteriors of $(x_1,\ldots,x_T)|(y_1,\ldots,y_T),\phi,\beta,\sigma^2$ and $\phi,\beta,\sigma^2|(x_1,\ldots,x_T),(y_1,\ldots,y_T)$. The Riemannian

Step Size	Num. Steps	Method	Acc. Prob.	Time (Sec.)	Mean ESS	Min. ESS	Mean ESS / Sec.	Min. ESS / Sec.
0.1	5	G.L.F.(a)	0.62 ± 0.01	400.33 ± 6.83	486.32 ± 17.89	286.28 ± 16.09	1.21 ± 0.04	0.71 ± 0.04
		G.L.F.(b)	0.61 ± 0.01	145.99 ± 2.20	491.07 ± 22.03	301.90 ± 16.21	3.36 ± 0.13	2.07 ± 0.10
		I.M.(a)	0.98 ± 0.00	102.37 ± 1.13	884.09 ± 27.39	620.26 ± 30.08	8.65 ± 0.31	6.07 ± 0.32
		I.M.(b)	0.98 ± 0.00	95.28 ± 2.28	857.61 ± 24.99	619.91 ± 27.09	9.03 ± 0.28	6.53 ± 0.31
	10	G.L.F.(a)	0.50 ± 0.01	615.14 ± 6.51	1038.47 ± 25.68	778.80 ± 30.92	1.69 ± 0.05	1.27 ± 0.05
		G.L.F.(b)	0.49 ± 0.01	231.55 ± 2.48	1027.78 ± 25.19	782.85 ± 14.87	4.44 ± 0.12	3.38 ± 0.07
		I.M.(a)	0.98 ± 0.00	194.82 ± 2.39	3018.50 ± 70.23	2518.65 ± 83.19	15.51 ± 0.39	12.94 ± 0.43
		I.M.(b)	0.98 ± 0.00	172.12 ± 1.45	3025.14 ± 55.19	2540.88 ± 99.95	17.58 ± 0.34	14.76 ± 0.56
	50	G.L.F.(a)	0.13 ± 0.00	2133.63 ± 64.84	192.82 ± 24.27	79.12 ± 14.55	0.09 ± 0.01	0.04 ± 0.01
		G.L.F.(b)	0.14 ± 0.00	786.40 ± 18.50	247.58 ± 28.90	119.23 ± 18.68	0.31 ± 0.03	0.15 ± 0.02
		I.M.(a)	0.95 ± 0.00	938.79 ± 15.17	4173.70 ± 199.89	3207.59 ± 113.94	4.47 ± 0.25	3.43 ± 0.14
		I.M.(b)	0.95 ± 0.00	834.63 ± 13.42	3928.27 ± 159.13	3158.40 ± 93.76	4.73 ± 0.24	3.80 ± 0.15

Table 1. Comparison of the implicit midpoint and generalized leapfrog integrators on sampling from the banana-shaped distribution. To assess performance of the sampler, we measure the effective sample size (ESS) and present per-second timing comparisons for the mean and minimum ESS. The hypothesis that the implicit midpoint integrator should exhibit better energy conservation is captured in the acceptance probability of the Markov chain. Results are averaged over ten trials.



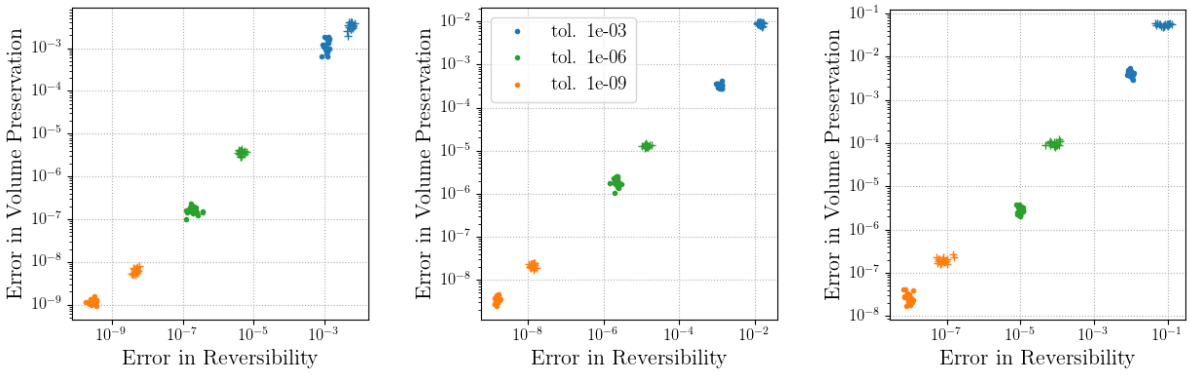
(a) $\epsilon = 1/10$ and 5 steps (b) $\epsilon = 1/10$ and 10 steps (c) $\epsilon = 1/10$ and 50 steps

Figure 2. Comparison between the error in symmetry and error in volume preservation properties of the implicit midpoint integrator and the generalized leapfrog integrator on the banana-shaped distribution. We observe that the implicit midpoint integrator tends to produce transitions whose median reversibility and volume preservation can be an order of magnitude, or more, better than the generalized leapfrog integrator. Each point is the median of one-hundred measurements of symmetry and reversibility shown for each of the ten trials. For $\epsilon = 1/10$ and fifty integration steps, the generalized leapfrog integrator exhibits severely divergent behavior. **The implicit midpoint is represented by the symbol $(\cdot)$ and the generalized leapfrog by the symbol $(+)$.**

Num. Steps	Step Size	Method	Acc. Prob.	Time (Sec.)	Mean ESS	Min. ESS	Mean ESS / Sec.	Min. ESS / Sec.
20	0.1	G.L.F.(a)	0.99 ± 0.00	1317.57 ± 16.89	15614.52 ± 177.89	340.56 ± 17.41	11.86 ± 0.16	0.26 ± 0.01
		I.M.(a)	1.00 ± 0.00	1315.24 ± 57.79	16199.00 ± 157.60	383.48 ± 22.16	12.47 ± 0.44	0.30 ± 0.02
		I.M.(b)	1.00 ± 0.00	1147.45 ± 64.92	15697.78 ± 148.81	392.22 ± 23.11	14.02 ± 0.74	0.35 ± 0.03
	0.2	G.L.F.(a)	0.96 ± 0.00	1933.64 ± 97.74	29457.72 ± 353.68	1511.75 ± 28.61	15.55 ± 0.79	0.79 ± 0.03
		I.M.(a)	0.99 ± 0.00	1669.69 ± 78.93	32055.29 ± 257.52	1588.73 ± 54.01	19.52 ± 0.86	0.96 ± 0.05
		I.M.(b)	0.99 ± 0.00	1441.81 ± 61.06	31585.28 ± 197.07	1558.54 ± 36.83	22.17 ± 0.79	1.10 ± 0.05
	0.5	G.L.F.(a)	0.36 ± 0.00	2389.27 ± 49.42	2144.93 ± 103.34	1771.92 ± 140.71	0.90 ± 0.05	0.74 ± 0.06
		I.M.(a)	0.85 ± 0.00	2981.80 ± 71.09	10984.29 ± 123.63	10147.33 ± 121.32	3.70 ± 0.09	3.42 ± 0.07
		I.M.(b)	0.85 ± 0.00	2728.61 ± 82.50	10563.94 ± 210.16	9711.29 ± 264.30	3.90 ± 0.13	3.58 ± 0.13

Table 2. Comparison of the implicit midpoint and the naive generalized leapfrog integrators on sampling from Neal's funnel distribution. We see that the implicit midpoint integrator is able to take large steps and produce an effective sample size that outperforms the generalized leapfrog integrator even in the time-normalized performance.

metric of this first posterior is constant with respect to $(x_1, \ldots, x_T)$; therefore, sampling is carried out using the standard leapfrog integrator. The second distribution has a position-dependent Riemannian metric, necessitating the use of implicitly-defined integrators; here, we compare the implicit midpoint and generalized leapfrog integrators. For details of the Riemannian structures of the conditional posteriors, see Girolami & Calderhead (2011).



(a) $\epsilon = 1/10$ (b) $\epsilon = 1/5$ (c) $\epsilon = 1/2$

Figure 3. Comparison of the degree to which the implicit midpoint and generalized leapfrog integrators violate reversibility and volume preservation on Neal's funnel distribution. We observe that the implicit midpoint integrator exhibits better symmetry and volume preservation. **The implicit midpoint is represented by the symbol $(\cdot)$ and the generalized leapfrog by the symbol $(+)$.**

Method	Acc. Prob.	Time (Sec.)	Mean ESS	Min. ESS	Mean ESS (Sec.)	Min. ESS (Sec.)	Volume Preservation		Symmetry	
							Median	90^{th}-Per.	Median	90^{th}-Per.
G.L.F.(a)	0.78 ± 0.0	2616.70 ± 16.5	294.51 ± 10.7	121.79 ± 7.7	0.11 ± 0.0	0.05 ± 0.0	7.2e−07	3.1e−06	4.9e−06	2.5e−05
G.L.F.(b)	0.78 ± 0.0	2339.78 ± 9.5	297.82 ± 9.2	132.79 ± 6.6	0.13 ± 0.0	0.06 ± 0.0	6.7e−07	2.8e−06	5.0e−06	2.4e−05
I.M.(a)	0.80 ± 0.0	2761.61 ± 8.3	309.10 ± 10.5	133.32 ± 7.8	0.11 ± 0.0	0.05 ± 0.0	8.7e−08	2.2e−07	1.3e−06	2.4e−06
I.M.(b)	0.80 ± 0.0	2710.47 ± 8.4	296.72 ± 9.2	127.35 ± 7.3	0.11 ± 0.0	0.05 ± 0.0	1.7e−07	4.5e−07	2.8e−06	4.7e−06

Table 3. Comparison of the implicit midpoint generalized leapfrog integrators on the stochastic volatility model. We see that the implicit midpoint integrator is competitive on the effective sample size and timing comparisons. We also evaluate the median reversibility and volume preservation of the implicit midpoint and the generalized leapfrog integrators on the stochastic volatility model with $\delta = 1 \times 10^{-6}$. Here we see that the implicit midpoint integrator enjoys better conservation of volume and reversibility.

In our experiments, we set $T = 1,000$ and use fifty integration steps with a step-size of 0.1 to sample $(x_1, \ldots, x_T)|(y_1, \ldots, y_T), \phi, \beta, \sigma^2$ and six integration steps with a step-size of 0.5 to sample $\phi, \beta, \sigma^2|(x_1, \ldots, x_T), (y_1, \ldots, y_T)$. We seek to sample $20,000$ times from the posterior and use a burn-in period of $10,000$ iterations. We repeat this experiment one-hundred times for each integrator. Effective sample size metrics and measures of the volume preservation and symmetry are are presented in table 3 *for the parameters ϕ, β, and σ^2*. We find that the I.M.(a) and (b) integrators are comparable to the G.L.F.(a) and (b) methods in terms of their time-normalized performance. However, volume preservation and symmetry are better for the I.M.(a) and (b) integrators.

4.5. Fitzhugh-Nagumo ODE Model

The Fitzhugh-Nagumo ordinary differential equation is a model of neural spiking activity. It is described by two time-varying measurements whose dynamics obey,

$$\dot{v} = \theta_3 \left(v - \frac{v^3}{3} + r \right) \tag{18}$$

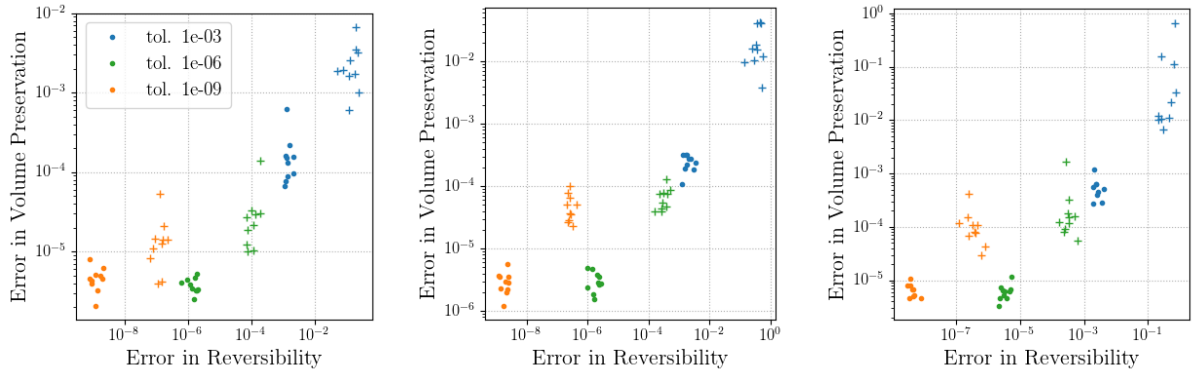
$$\dot{r} = -\left(\frac{v - \theta_1 + \theta_2 r}{\theta_3} \right). \tag{19}$$

Consider the setting wherein one has 200 observations of the Fitzhugh-Nagumo dynamics at equally-spaced times between zero and ten. Assume moreover that these observations have been corrupted by i.i.d. Gaussian noise with a known standard deviation of $\sigma = 1/2$. If we equip the parameters θ_1, θ_2, and θ_3 with standard normal priors, we may use the dynamics of eqs. (18) and (19) and the assumed noise distribution in order to sample the posterior of $(\theta_1, \theta_2, \theta_3)$. Let (v_n, r_n) be the solution of the Fitzhugh-Nagumo ODE at the n^{th} time period. For the Fitzhugh-Nagumo differential equation model, the (i, j)-entry of the metric is,

$$\mathbb{G}_{ij}(\theta_1, \theta_2, \theta_3) = \frac{1}{\sigma^2} \left(\sum_{n=1}^{200} \frac{\partial v_n}{\partial \theta_i} \frac{\partial v_n}{\partial \theta_j} + \frac{\partial r_n}{\partial \theta_i} \frac{\partial r_n}{\partial \theta_j} \right) + \delta_{ij}. \tag{20}$$

In generating data from the Fitzhugh-Nagumo model, we set $\theta_1 = \theta_2 = 0.2$ and $\theta_3 = 3$. The dynamics are integrated using SciPy's `odeint` function and gradients are *approximated* by forward sensitivity analysis as in Girolami & Calderhead (2011). We consider an integration step-size of $\epsilon = 1$ and a number of integration steps in $\{1, 2, 5\}$; each configuration is replicated ten times. We sample 1,000 times from the posterior.

We expect the Fitzhugh-Nagumo ODE model to favor the



(a) One integration step (b) Two integration steps (c) Five integration steps

Figure 4. Comparison of the degree to which the implicit midpoint and generalized leapfrog integrators violate reversibility and volume preservation on the Fitzhugh-Nagumo posterior. We observe that for the chosen convergence tolerances, the implicit midpoint integrator tends to exhibit better reversibility and volume preservation except for the smallest tolerance where the generalized leapfrog has slightly better volume preservation but worse reversibility. **The implicit midpoint is represented by the symbol $(\cdot)$ and the generalized leapfrog by the symbol $(+)$.**

Step Size	Num. Steps	Method	Acc. Prob.	Time (Sec.)	Mean ESS	Min. ESS	Mean ESS / Sec.	Min. ESS / Sec.
1.0	1	G.L.F.(a)	0.74 ± 0.01	4441.16 ± 117.73	276.39 ± 13.48	235.39 ± 10.55	0.06 ± 0.00	0.05 ± 0.00
		G.L.F.(b)	0.73 ± 0.01	1343.63 ± 76.12	272.72 ± 7.96	235.52 ± 8.99	0.21 ± 0.01	0.18 ± 0.01
		I.M.(a)	0.95 ± 0.00	4523.49 ± 160.00	227.01 ± 13.47	199.10 ± 13.70	0.05 ± 0.00	0.04 ± 0.00
		I.M.(b)	0.95 ± 0.00	4486.29 ± 147.33	248.11 ± 8.74	225.67 ± 9.11	0.06 ± 0.00	0.05 ± 0.00
	2	G.L.F.(a)	0.74 ± 0.01	8100.90 ± 423.17	1089.36 ± 55.17	907.30 ± 75.02	0.14 ± 0.01	0.11 ± 0.01
		G.L.F.(b)	0.75 ± 0.01	1912.73 ± 48.99	1109.78 ± 71.50	917.52 ± 98.66	0.58 ± 0.04	0.48 ± 0.05
		I.M.(a)	0.94 ± 0.00	8704.37 ± 250.22	1387.96 ± 40.53	1212.91 ± 42.68	0.16 ± 0.01	0.14 ± 0.01
		I.M.(b)	0.94 ± 0.00	8423.40 ± 139.60	1366.07 ± 56.59	1213.01 ± 55.18	0.16 ± 0.00	0.14 ± 0.01
	5	G.L.F.(a)	0.70 ± 0.01	20096.64 ± 384.16	188.59 ± 27.78	90.00 ± 19.15	0.01 ± 0.00	0.00 ± 0.00
		G.L.F.(b)	0.71 ± 0.01	5138.31 ± 300.83	150.03 ± 23.06	102.32 ± 22.10	0.03 ± 0.00	0.02 ± 0.00
		I.M.(a)	0.94 ± 0.00	22069.48 ± 650.03	954.94 ± 93.17	818.16 ± 91.14	0.04 ± 0.01	0.04 ± 0.01
		I.M.(b)	0.94 ± 0.00	20575.78 ± 506.21	1095.88 ± 130.14	821.77 ± 80.15	0.05 ± 0.01	0.04 ± 0.00

Table 4. Comparison of the implicit midpoint and generalized leapfrog integrators on sampling from the posterior of the Fitzhugh-Nagumo model. In this example, the higher acceptance probability of the implicit midpoint integrator did not produce a performance increase for a single integration step. The implicit midpoint integrator becomes super-efficient in the two-step regime, but cannot compete with the generalized leapfrog integrator's computational advantages.

generalized leapfrog integrator because of the complexity of evaluating the log-posterior, the gradient of the log-posterior, the Riemannian metric, and the gradient of the Riemannian metric, each of which involves solving a system of differential equations. Therefore, the caching behavior associated to the G.L.F.(b) integrator gives it an advantage here. Table 4 shows the results of inferences in the Fitzhugh-Nagumo posterior. We observe that for a single-step, the I.M.(a) and (b) integrators appears to perform somewhat worse than the G.L.F.(a) and (b) variants, even on the measures of ESS that ignore timing; this occurs despite the larger acceptance rate enjoyed by the implicit midpoint integrator. For two integration steps, the inferences produced by I.M.(a) and (b) become super-efficient; however, G.L.F.(a) and (b) are also efficient and the computational advantage of the (b) variant cause it to have superior performance in the timing metrics. For the largest number of steps, the performance of G.L.F.(a) and (b) deteriorates so that the I.M.(a) and (b) integrators outperform them even on the timing comparison.

We also evaluate the degree to which the numerical integrator possesses the properties of symmetry and volume preservation. The results are shown in fig. 4. We see that the implicit midpoint integrator offers a clear advantage in numerical symmetry, and performs better on volume preservation as well.

5. Conclusion

This work has considered the implicit midpoint integrator as a substitute for the generalized leapfrog integrator for use in RMHMC. Inspired by the theory of the conserved quantities of numerical integrators, we hypothesized that the implicit midpoint integrator would have better energy conservation in posterior distributions that are approximately Gaussian.

Hamiltonian Monte Carlo requires that its integrators are volume preserving and reversible; we give numerical assessments of the extent to which these properties are present in implementations of these integrators, which differ from their theoretical representation when a convergence tolerance is used to halt a fixed point iteration. We find that the implicit midpoint integrator has superior energy conservation, conservation of volume, and reversibility across several Bayesian inference tasks. In three of the four example applications, the implicit midpoint integrator met or exceeded the *time-normalized* performance of the generalized leapfrog integrator. This, combined with its better volume preservation and reversibility, leads us to conclude that it is a method worth consideration when implementing RMHMC.

Acknowledgments

The authors would like to 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. RRL was supported in part by NIH/NIGMS 1R01GM136780-01.

References

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

Betancourt, M. A general metric for riemannian manifold hamiltonian monte carlo. 8085, 12 2012. doi: 10.1007/978-3-642-40020-9_35.

Bishop, C. M. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2006. ISBN 0387310738.

Brubaker, M., Salzmann, M., and Urtasun, R. A family of mcmc methods on implicitly defined manifolds. In Lawrence, N. D. and Girolami, M. (eds.), *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics*, volume 22 of *Proceedings of Machine Learning Research*, pp. 161–172, La Palma, Canary Islands, April 2012. PMLR. URL http://proceedings.mlr.press/v22/brubaker12.html.

Byrne, S. and Girolami, M. Geodesic monte carlo on embedded manifolds. *Scandinavian Journal of Statistics*, 40(4):825–845, Sep 2013. ISSN 0303-6898. doi: 10.1111/sjos.12036. URL http://dx.doi.org/10.1111/sjos.12036.

Duane, S., Kennedy, A. D., Pendleton, B. J., and Roweth, D. Hybrid monte carlo. *Physics Letters B*, 195(2):216 – 222, 1987. ISSN 0370-2693. doi: DOI:10.1016/0370-2693(87)91197-X. URL http://www.sciencedirect.com/science/article/B6TVN-46YSWPH-2XF/2/0f89cdc6cf214a2169b03df7414f3df4.

Girolami, M. and Calderhead, B. Riemann manifold langevin and hamiltonian monte carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(2):123–214, 2011. doi: 10.1111/j.1467-9868.2010.00765.x. URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2010.00765.x.

Hairer, E., Lubich, C., and Wanner, G. *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations; 2nd ed.* Springer, Dordrecht, 2006. doi: 10.1007/3-540-30666-8. URL https://cds.cern.ch/record/1250576.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del R'ıo, J. F., Wiebe, M., Peterson, P., G'erard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020. doi: 10.1038/s41586-020-2649-2. URL https://doi.org/10.1038/s41586-020-2649-2.

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

Leimkuhler, B. and Reich, S. *Simulating Hamiltonian Dynamics*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, 2005. doi: 10.1017/CBO9780511614118.

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

Martens, J. New insights and perspectives on the natural gradient method. *Journal of Machine Learning Research*, 21(146):1–76, 2020. URL http://jmlr.org/papers/v21/17-678.html.

Neal, R. M. MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, 54:113–162, 2010.

Pourzanjani, A. A. and Petzold, L. R. Implicit hamiltonian monte carlo for sampling multiscale distributions, 2019.

Tripuraneni, N., Rowland, M., Ghahramani, Z., and Turner, R. Magnetic Hamiltonian Monte Carlo. In Precup, D. and Teh, Y. W. (eds.), *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 3453–3461, International Convention Centre, Sydney, Australia, August 2017. PMLR. URL http://proceedings.mlr.press/v70/tripuraneni17a.html.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2.