A. Momentum Negation Symmetry of Implicit Midpoint

Lemma 1. Given a Hamiltonian H in the form of eq. (1), step-size ϵ, and initial position (q, p), compute (q', p') according to algorithm 3 with $\delta = 0$. If one then computes (q'', p'') from initial position $(q', -p')$ using algorithm 3 a second time (with the same Hamiltonian, step-size, and $\delta = 0$), then $q'' = q$ and $-p'' = p$.

Lemma 1 establishes that the implicit midpoint integrator is suitable for HMC in that it satisfies properties (i) and (ii).

Proof. Consider the initial condition (q, p) and a fixed step-size of ϵ. For the Riemannian manifold Hamiltonian Monte Carlo, the implicit midpoint integrator computes the following updates:

$$q_i' = q_i + \epsilon\left(\sum_{j=1}^m \mathbb{G}_{ij}^{-1}\left(\frac{q'+q}{2}\right)\left(\frac{p_j'+p_j}{2}\right)\right) \tag{21}$$

$$\begin{aligned} p_i' = p_i + \epsilon\Bigg(&-\frac{\partial}{\partial q_i}\mathcal{L}\left(\frac{q'+q}{2}\right) - \frac{1}{2}\text{trace}\left(\mathbb{G}^{-1}\left(\frac{q'+q}{2}\right)\frac{\partial}{\partial q_i}\mathbb{G}\left(\frac{q'+q}{2}\right)\right) + \\ &\frac{1}{2}\left(\frac{p'+p}{2}\right)^\top \mathbb{G}^{-1}\left(\frac{q'+q}{2}\right)\frac{\partial}{\partial q_i}\mathbb{G}\left(\frac{q'+q}{2}\right)\mathbb{G}^{-1}\left(\frac{q'+q}{2}\right)\left(\frac{p'+p}{2}\right)\Bigg) \end{aligned} \tag{22}$$

What we want to show is that if we compute (q', p'), negate the momentum $(q', p') \mapsto (q', -p')$, and apply the implicit midpoint integrator a second time, then we arrive at $(q, -p)$. Thus, we need to establish that $(q, -p)$ is a fixed point of the relations,

$$q_i'' = q_i' + \epsilon\left(\sum_{j=1}^m \mathbb{G}_{ij}^{-1}\left(\frac{q''+q'}{2}\right)\left(\frac{p_j''+(-p_j')}{2}\right)\right) \tag{23}$$

$$\begin{aligned} p_i'' = -p_i' + \epsilon\Bigg(&-\frac{\partial}{\partial q_i}\mathcal{L}\left(\frac{q''+q'}{2}\right) - \frac{1}{2}\text{trace}\left(\mathbb{G}^{-1}\left(\frac{q''+q'}{2}\right)\frac{\partial}{\partial q_i}\mathbb{G}\left(\frac{q''+q'}{2}\right)\right) + \\ &\frac{1}{2}\left(\frac{p''+(-p')}{2}\right)^\top \mathbb{G}^{-1}\left(\frac{q''+q'}{2}\right)\frac{\partial}{\partial q_i}\mathbb{G}\left(\frac{q''+q'}{2}\right)\mathbb{G}^{-1}\left(\frac{q''+q'}{2}\right)\left(\frac{p''+(-p')}{2}\right)\Bigg). \end{aligned} \tag{24}$$

Plugging in we obtain,

$$q_i' + \epsilon\left(\sum_{j=1}^m \mathbb{G}_{ij}^{-1}\left(\frac{q+q'}{2}\right)\left(\frac{(-p_j)+(-p_j')}{2}\right)\right) = q_i' - \epsilon\left(\sum_{j=1}^m \mathbb{G}_{ij}^{-1}\left(\frac{q+q'}{2}\right)\left(\frac{p_j+p_j'}{2}\right)\right) \tag{25}$$

$$= q_i \tag{26}$$

by rearranging eq. (21). For notational simplicity let us define

$$U(q) \overset{\text{def.}}{=} -\frac{\partial}{\partial q_i}\mathcal{L}(q) - \frac{1}{2}\text{trace}\left(\mathbb{G}^{-1}(q)\frac{\partial}{\partial q_i}\mathbb{G}(q)\right) \tag{27}$$

$$R(q) \overset{\text{def.}}{=} \frac{1}{2}\mathbb{G}^{-1}(q)\frac{\partial}{\partial q_i}\mathbb{G}(q)\mathbb{G}^{-1}(q) \tag{28}$$

so that

$$p_i'' = -p_i' + \epsilon\left(U\left(\frac{q''+q}{2}\right) + \left(\frac{p''+(-p')}{2}\right)^\top R\left(\frac{q''+q}{2}\right)\left(\frac{p''+(-p')}{2}\right)\right). \tag{29}$$

Plugging in, we obtain,

$$
\begin{aligned}
& -p_i' + \epsilon\left(U\left(\frac{q+q}{2}\right) + \left(\frac{(-p)+(-p')}{2}\right)^\top R\left(\frac{q+q}{2}\right)\left(\frac{(-p)+(-p')}{2}\right)\right) && (30) \\
=& -p_i' + \epsilon\left(U\left(\frac{q+q}{2}\right) + \left(\frac{p+p'}{2}\right)^\top R\left(\frac{q+q}{2}\right)\left(\frac{p+p'}{2}\right)\right) && (31) \\
=& -p_i && (32)
\end{aligned}
$$

which follows from negating eq. (22) and rearranging. □

B. Implicit Midpoint Eigenvalues

Let $z = (q, p)$ and consider a quadratic Hamiltonian of the form,

$$H(z) = z^\top \mathbf{A} z \tag{33}$$

$$= \begin{pmatrix} q^\top & p^\top \end{pmatrix} \begin{pmatrix} \Sigma & 0 \\ 0 & \Sigma \end{pmatrix} \begin{pmatrix} q \\ p \end{pmatrix}. \tag{34}$$

The associated Hamiltonian vector field is,

$$\dot{z} = \begin{pmatrix} \dot{q} \\ \dot{p} \end{pmatrix} \tag{35}$$

$$= \begin{pmatrix} 0 & \mathrm{Id} \\ -\mathrm{Id} & 0 \end{pmatrix} \begin{pmatrix} \Sigma & 0 \\ 0 & \Sigma \end{pmatrix} \begin{pmatrix} q \\ p \end{pmatrix} \tag{36}$$

$$= \underbrace{\begin{pmatrix} 0 & \Sigma \\ -\Sigma & 0 \end{pmatrix}}_{\mathbf{J}} \begin{pmatrix} q \\ p \end{pmatrix} \tag{37}$$

$$= \mathbf{J} z. \tag{38}$$

From eq. (10), the implicit midpoint integrator computes the update,

$$z' = z + \epsilon \mathbf{J} \left(\frac{z' + z}{2} \right) \tag{39}$$

$$\implies z' = z + \frac{\epsilon}{2} \mathbf{J} z' + \frac{\epsilon}{2} \mathbf{J} z \tag{40}$$

$$\implies \left(\mathrm{Id} - \frac{\epsilon}{2} \mathbf{J} \right) z' = \left(\mathrm{Id} + \frac{\epsilon}{2} \mathbf{J} \right) z \tag{41}$$

$$\implies z' = \underbrace{\left(\mathrm{Id} - \frac{\epsilon}{2} \mathbf{J} \right)^{-1} \left(\mathrm{Id} + \frac{\epsilon}{2} \mathbf{J} \right)}_{\mathbf{Q}} z \tag{42}$$

The quantity $\mathbf{Q}$ is the Cayley transform of the linear transformation $\frac{\epsilon}{2}\mathbf{J}$. Noting that $\mathbf{J}$ is a skew-symmetric matrix, it is an established fact that the Cayley transform of a skew-symmetric matrix is an orthogonal matrix. This establishes that all of the eigenvalues of $\mathbf{Q}$, the linear transform representing the implicit midpoint integrator, have unit modulus.

C. Riemannian Metrics and a Silent Change-of-Variables?

The Riemannian volume measure is $\sqrt{\det(\mathbb{G}(q))}\,\mathrm{d}q$ on $\mathbb{R}^m$. When using HMC, does the fact that we have introduced a metric mean that we require a Jacobian correction to the posterior? Actually, the answer is no. The reason is that the Metropolis-Hastings accept-reject rule determines which density (specified with respect to the Lebesgue measure in the (q, p) phase-space) is sampled by the Markov chain. Just because the acceptance Hamiltonian (which is also the guidance Hamiltonian in RMHMC; see Duane et al. (1987)) involves computing the Riemannian metric does not mean that we have silently changed the underlying measure. Indeed, the log-determinant term appearing in eq. (1) is chosen so that the conditional distribution of p given q is multivariate normal *with respect to the Lebesgue measure*.

D. Numerical Stability

The stability of numerical integrators is defined by their long-term behavior on the harmonic oscillator, which is described by the following Hamiltonian system

$$\dot{q} = p \qquad \dot{p} = -\omega^2 q, \tag{43}$$

corresponding to the Hamiltonian $H(q, p) = \omega^2 \frac{q^2}{2} + \frac{p^2}{2}$ where ω^2 is a constant. Consider a single step of a numerical integrator for the harmonic oscillator with step-size ϵ that maps $(q, p) \mapsto (q', p')$. Because the harmonic oscillator is a *linear* differential equation, it is often possible find a matrix $\mathbf{R} \in \mathbb{R}^{2\times 2}$ such that $(q', p')^\top = \mathbf{R}(q, p)^\top$. A numerical method is called *stable* if the eigenvalues of $\mathbf{R}$ lie on the unit disk of the complex plane and are not repeated (Leimkuhler & Reich, 2005). We have the following result.

Proposition 2. When $\epsilon < 2/\omega$, the (generalized) leapfrog integrator is stable. The implicit midpoint integrator is stable for any ϵ.

See Hairer et al. (2006); Leimkuhler & Reich (2005) for an introduction to stability analysis of numerical integrators.

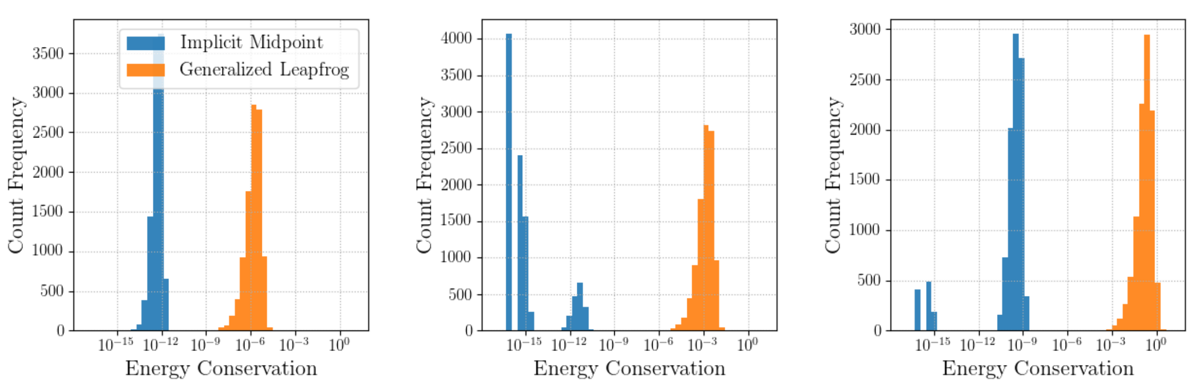
E. Quadratic Hamiltonian

We consider using HMC to draw samples from a Gaussian distribution in two dimensions. In particular, we aim to sample from the joint distribution of position and momentum defined by,

$$q \sim \mathrm{Normal}\left(\begin{pmatrix}1/2 \\ -1\end{pmatrix}, \begin{pmatrix}1 & 1/2 \\ 1/2 & 2\end{pmatrix}\right) \tag{44}$$

$$p \sim \mathrm{Normal}\left(\begin{pmatrix}0 \\ 0\end{pmatrix}, \begin{pmatrix}1 & 1/2 \\ 1/2 & 2\end{pmatrix}^{-1}\right) \tag{45}$$

These distributions correspond to the quadratic Hamiltonian $H(q, p) = \frac{1}{2}q^\top \Sigma^{-1} q + \frac{1}{2}p^\top \Sigma p$. This Hamiltonian can be interpreted in the Riemannian manifold setting as sampling from the posterior $q \sim \mathrm{Normal}(\mu, \Sigma)$ and the constant metric $\mathbb{G}(q) = \Sigma^{-1}$. The corresponding Hamiltonian is quadratic and therefore theorem 2 applies. We expect perfect conservation of the Hamiltonian regardless of step-size. To evaluate the conservation of the Hamiltonian energy, we consider drawing (q, p) from their joint distribution and integrating Hamilton's equations of motion for ten integration steps. We consider integration step-sizes in $\{0.01, 0.1, 1.0\}$. We then compare the initial Hamiltonian energy to the Hamiltonian energy at the terminal point of the integrator. We repeat this procedure 10,000 times and show the results in fig. 5, where the absolute difference in Hamiltonian energy is shown as a histogram. This experiment clearly shows that the implicit midpoint integrator has excellent conservation of the quadratic Hamiltonian energy and is orders of magnitude better than the generalized leapfrog integrator. Note that for a separable Hamiltonian, as is the case here, the steps of the generalized leapfrog integrator reduce to the standard leapfrog method.



(a) $\epsilon = 1/100$ (b) $\epsilon = 1/10$ (c) $\epsilon = 1$

Figure 5. Comparison between the energy conservation of the implicit midpoint integrator and the generalized leapfrog integrator on a quadratic Hamiltonian. We observe that for every step-size, the energy conservation of the implicit midpoint method is in the neighborhood of 1×10^{-10} whereas the generalized leapfrog has energy conservation that degrades with larger steps.

F. Bayesian Logistic Regression

Binary classification is a uniquitous task in the data sciences and logistic regression is the most popular algorithm for obtaining probabilistic estimates of class membership. Bayesian logistic regression simply equips each of the linear coefficients in the logistic regression model with a prior distribution. We consider Bayesian logistic regression as defined by the following generative model:

$$y_i|x_i, \beta \sim \text{Bernoulli}(\sigma(x_i^\top \beta)) \qquad \text{for } i = 1, \ldots, n \tag{46}$$

$$\beta_i \sim \text{Normal}(0, 1) \qquad \text{for } i = 1, \ldots, k, \tag{47}$$

where $x_i \in \mathbb{R}^k$ is vector of explanatory variables and $\sigma : \mathbb{R} \to (0, 1)$ is the sigmoid function. For the logistic regression model, let $\mathbf{x} \in \mathbb{R}^{n \times m}$ represent the matrix of features. The Riemannian metric formed by the sum of the Fisher information and the negative Hessian of the log-prior is $\mathbb{G}(\beta) = \mathbf{x}^\top \mathbf{\Lambda} \mathbf{x} + \text{Id}$ where $\mathbf{\Lambda}$ is a diagonal matrix whose i^{th} diagonal entry is $\sigma(x_i^\top \beta)(1 - \sigma(x_i^\top \beta))$.

We consider sampling from the posterior distribution of the linear coefficients for a breast cancer, heart disease, and diabetes dataset. We consider integration step-sizes in $\{1/10, 1\}$ and a number of integration steps in $\{5, 10, 50\}$; each configuration of step-size and number of steps is replicated ten times and we attempt to draw 10,000 samples from the posterior. Results are presented in tables 5 to 7. For the smaller step-size, both integrators enjoy very high acceptance rates and similar performance when *not* adjusted for timing; when adjusted for timing, the generalized leapfrog is often the better choice in the presence of a small step-size. For the larger step-size, only the implicit midpoint integrator is able to maintain a high acceptance rate; occasionally, the implicit midpoint is able to produce the optimal mean ESS and minimum ESS *per second*.

Step Size	Num. Steps	Method	Acc. Prob.	Mean ESS	Min. ESS	Mean ESS / Sec.	Min. ESS / Sec.
0.1	5	G.L.F.(a)	1.00 ± 0.00	657.70 ± 7.91	544.12 ± 14.20	5.10 ± 0.09	4.22 ± 0.13
		G.L.F.(b)	1.00 ± 0.00	656.40 ± 8.22	530.78 ± 14.66	11.64 ± 0.19	9.42 ± 0.31
		I.M.(a)	1.00 ± 0.00	644.86 ± 5.96	525.95 ± 12.06	5.18 ± 0.05	4.22 ± 0.10
		I.M.(b)	1.00 ± 0.00	653.46 ± 7.75	521.33 ± 15.77	5.98 ± 0.09	4.77 ± 0.15
	10	G.L.F.(a)	1.00 ± 0.00	3011.19 ± 22.56	2763.67 ± 35.50	12.33 ± 0.10	11.32 ± 0.17
		G.L.F.(b)	1.00 ± 0.00	3013.59 ± 13.75	2686.52 ± 48.17	29.85 ± 0.19	26.61 ± 0.44
		I.M.(a)	1.00 ± 0.00	3000.27 ± 17.29	2721.65 ± 36.14	12.60 ± 0.09	11.43 ± 0.17
		I.M.(b)	1.00 ± 0.00	2982.78 ± 15.14	2716.33 ± 43.65	14.42 ± 0.16	13.13 ± 0.27
	50	G.L.F.(a)	1.00 ± 0.00	5090.37 ± 28.81	4650.52 ± 57.69	4.29 ± 0.03	3.92 ± 0.06
		G.L.F.(b)	1.00 ± 0.00	5071.69 ± 26.96	4660.23 ± 57.15	11.08 ± 0.09	10.18 ± 0.15
		I.M.(a)	1.00 ± 0.00	5155.52 ± 38.72	4734.41 ± 52.70	4.48 ± 0.03	4.12 ± 0.04
		I.M.(b)	1.00 ± 0.00	5190.77 ± 22.02	4739.59 ± 64.52	5.23 ± 0.05	4.78 ± 0.09
1.0	5	G.L.F.(a)	0.19 ± 0.02	113.18 ± 10.39	61.40 ± 9.69	0.37 ± 0.03	0.20 ± 0.03
		G.L.F.(b)	0.21 ± 0.00	147.14 ± 5.94	79.51 ± 9.49	1.57 ± 0.06	0.85 ± 0.10
		I.M.(a)	0.88 ± 0.00	7338.66 ± 40.87	6284.89 ± 141.01	17.05 ± 0.15	14.60 ± 0.32
		I.M.(b)	0.87 ± 0.00	7355.17 ± 56.08	6418.95 ± 173.52	17.93 ± 0.16	15.65 ± 0.42
	10	G.L.F.(a)	0.14 ± 0.00	181.31 ± 22.14	87.52 ± 19.51	0.45 ± 0.05	0.21 ± 0.05
		G.L.F.(b)	0.14 ± 0.00	178.90 ± 19.79	90.50 ± 17.22	1.48 ± 0.16	0.75 ± 0.14
		I.M.(a)	0.86 ± 0.00	31253.10 ± 1766.12	17610.80 ± 2459.35	35.43 ± 2.34	20.10 ± 2.93
		I.M.(b)	0.86 ± 0.00	29505.12 ± 1650.54	17108.46 ± 2719.39	35.27 ± 2.06	20.49 ± 3.30
	50	G.L.F.(a)	0.06 ± 0.00	71.76 ± 9.80	20.62 ± 5.41	0.10 ± 0.01	0.03 ± 0.01
		G.L.F.(b)	0.06 ± 0.01	108.39 ± 25.71	44.92 ± 16.60	0.47 ± 0.10	0.19 ± 0.07
		I.M.(a)	0.83 ± 0.00	10127.12 ± 327.96	4430.28 ± 278.92	2.07 ± 0.07	0.91 ± 0.06
		I.M.(b)	0.83 ± 0.00	10162.12 ± 417.45	4173.36 ± 445.85	2.20 ± 0.10	0.91 ± 0.10

Table 5. Comparison of the implicit midpoint and generalized leapfrog integrators on sampling from the Bayesian logistic regression posterior on the Breast Cancer dataset. For the larger step-size, the acceptance rate of the generalized leapfrog integrator completely deteriorates whereas the implicit midpoint integrator is more robust. The implicit midpoint integrator is able to achieve super-efficient sampling for a step-size of $\epsilon = 1$ and ten integration steps.

Step Size	Num. Steps	Method	Acc. Prob.	Mean ESS	Min. ESS	Mean ESS / Sec.	Min. ESS / Sec.
0.1	5	G.L.F.(a)	1.00 ± 0.00	651.60 ± 10.19	529.61 ± 24.09	4.21 ± 0.06	3.42 ± 0.16
		G.L.F.(b)	1.00 ± 0.00	647.30 ± 8.08	544.38 ± 14.30	10.13 ± 0.15	8.52 ± 0.23
		I.M.(a)	1.00 ± 0.00	649.67 ± 5.17	531.52 ± 18.59	3.84 ± 0.06	3.14 ± 0.11
		I.M.(b)	1.00 ± 0.00	647.56 ± 10.04	537.62 ± 16.70	4.41 ± 0.08	3.67 ± 0.13
	10	G.L.F.(a)	1.00 ± 0.00	2986.69 ± 23.91	2705.46 ± 58.58	10.16 ± 0.09	9.20 ± 0.19
		G.L.F.(b)	1.00 ± 0.00	2964.67 ± 17.29	2715.30 ± 34.13	26.06 ± 0.39	23.88 ± 0.51
		I.M.(a)	1.00 ± 0.00	2959.34 ± 21.18	2741.55 ± 35.33	9.18 ± 0.10	8.51 ± 0.12
		I.M.(b)	1.00 ± 0.00	2963.56 ± 17.48	2710.39 ± 58.89	10.78 ± 0.08	9.86 ± 0.21
	50	G.L.F.(a)	1.00 ± 0.00	5393.71 ± 31.87	5047.21 ± 52.74	3.80 ± 0.03	3.55 ± 0.04
		G.L.F.(b)	1.00 ± 0.00	5403.67 ± 40.17	4966.93 ± 71.68	10.53 ± 0.09	9.68 ± 0.14
		I.M.(a)	1.00 ± 0.00	5445.70 ± 35.37	5069.47 ± 87.22	3.51 ± 0.03	3.27 ± 0.06
		I.M.(b)	1.00 ± 0.00	5529.36 ± 24.08	5161.12 ± 39.27	4.14 ± 0.02	3.86 ± 0.03
1.0	5	G.L.F.(a)	0.71 ± 0.00	1507.07 ± 28.17	1051.41 ± 56.24	4.78 ± 0.10	3.34 ± 0.18
		G.L.F.(b)	0.71 ± 0.00	1533.13 ± 19.75	1136.48 ± 40.48	16.78 ± 0.17	12.45 ± 0.46
		I.M.(a)	0.97 ± 0.00	10484.51 ± 46.67	9934.74 ± 60.88	18.74 ± 0.08	17.76 ± 0.12
		I.M.(b)	0.97 ± 0.00	10446.28 ± 28.06	9870.19 ± 65.18	19.71 ± 0.13	18.62 ± 0.17
	10	G.L.F.(a)	0.60 ± 0.00	4878.12 ± 107.69	2912.57 ± 257.17	7.89 ± 0.18	4.71 ± 0.42
		G.L.F.(b)	0.60 ± 0.00	5088.12 ± 42.63	3495.29 ± 126.45	30.17 ± 0.28	20.71 ± 0.73
		I.M.(a)	0.97 ± 0.00	40000.00 ± 0.00	40000.00 ± 0.00	36.32 ± 0.13	36.32 ± 0.13
		I.M.(b)	0.96 ± 0.00	40000.00 ± 0.00	40000.00 ± 0.00	38.39 ± 0.08	38.39 ± 0.08
	50	G.L.F.(a)	0.64 ± 0.00	6454.94 ± 278.39	4318.50 ± 306.57	2.32 ± 0.10	1.55 ± 0.11
		G.L.F.(b)	0.64 ± 0.00	6357.53 ± 179.10	4245.83 ± 213.30	8.75 ± 0.24	5.84 ± 0.29
		I.M.(a)	0.97 ± 0.00	37135.14 ± 210.27	20941.08 ± 409.25	6.77 ± 0.04	3.82 ± 0.07
		I.M.(b)	0.97 ± 0.00	37049.46 ± 189.78	20932.03 ± 441.83	7.02 ± 0.09	3.97 ± 0.10

Table 6. Comparison of the implicit midpoint and generalized leapfrog integrators on sampling from the Bayesian logistic regression posterior on the Diabetes dataset.

Step Size	Num. Steps	Method	Acc. Prob.	Mean ESS	Min. ESS	Mean ESS / Sec.	Min. ESS / Sec.
0.1	5	G.L.F.(a)	1.00 ± 0.00	653.84 ± 8.93	541.18 ± 16.32	3.86 ± 0.05	3.19 ± 0.10
		G.L.F.(b)	1.00 ± 0.00	656.59 ± 6.98	533.13 ± 11.07	9.55 ± 0.15	7.75 ± 0.18
		I.M.(a)	1.00 ± 0.00	668.14 ± 4.07	564.02 ± 8.36	4.23 ± 0.04	3.57 ± 0.05
		I.M.(b)	1.00 ± 0.00	664.05 ± 6.83	517.75 ± 15.55	4.78 ± 0.08	3.73 ± 0.14
	10	G.L.F.(a)	0.99 ± 0.00	3019.18 ± 16.31	2659.45 ± 40.27	9.13 ± 0.17	8.05 ± 0.21
		G.L.F.(b)	0.99 ± 0.00	3021.32 ± 18.34	2686.55 ± 45.81	24.54 ± 0.28	21.83 ± 0.47
		I.M.(a)	1.00 ± 0.00	3019.84 ± 16.66	2728.01 ± 27.65	9.82 ± 0.07	8.86 ± 0.06
		I.M.(b)	1.00 ± 0.00	3029.96 ± 13.91	2683.39 ± 43.94	11.42 ± 0.10	10.12 ± 0.19
	50	G.L.F.(a)	0.99 ± 0.00	4934.18 ± 14.56	4500.43 ± 42.68	3.14 ± 0.02	2.86 ± 0.03
		G.L.F.(b)	0.99 ± 0.00	4952.56 ± 14.79	4431.26 ± 54.96	8.84 ± 0.07	7.91 ± 0.11
		I.M.(a)	1.00 ± 0.00	5107.34 ± 20.29	4626.57 ± 72.19	3.44 ± 0.01	3.12 ± 0.05
		I.M.(b)	1.00 ± 0.00	5089.77 ± 21.30	4642.54 ± 54.89	4.02 ± 0.02	3.67 ± 0.04
1.0	5	G.L.F.(a)	0.05 ± 0.00	25.97 ± 3.33	7.21 ± 1.36	0.08 ± 0.01	0.02 ± 0.00
		G.L.F.(b)	0.05 ± 0.00	29.48 ± 2.90	9.96 ± 2.03	0.37 ± 0.04	0.13 ± 0.03
		I.M.(a)	0.70 ± 0.00	4335.06 ± 63.32	3250.56 ± 138.84	6.53 ± 0.12	4.90 ± 0.21
		I.M.(b)	0.70 ± 0.00	4352.91 ± 83.82	3300.84 ± 186.04	6.51 ± 0.19	4.95 ± 0.31
	10	G.L.F.(a)	0.03 ± 0.00	44.35 ± 4.24	12.20 ± 1.90	0.12 ± 0.01	0.03 ± 0.00
		G.L.F.(b)	0.02 ± 0.00	44.07 ± 8.26	10.00 ± 1.91	0.51 ± 0.11	0.11 ± 0.02
		I.M.(a)	0.66 ± 0.00	9217.63 ± 305.79	5320.85 ± 661.55	5.45 ± 0.21	3.16 ± 0.41
		I.M.(b)	0.66 ± 0.00	9752.37 ± 315.58	5172.40 ± 382.12	5.72 ± 0.20	3.05 ± 0.24
	50	G.L.F.(a)	0.01 ± 0.00	19.19 ± 3.58	4.62 ± 0.55	0.05 ± 0.02	0.01 ± 0.00
		G.L.F.(b)	0.01 ± 0.00	24.34 ± 7.90	5.81 ± 1.41	0.21 ± 0.07	0.05 ± 0.01
		I.M.(a)	0.58 ± 0.00	2818.14 ± 244.09	1287.74 ± 212.68	0.22 ± 0.02	0.10 ± 0.02
		I.M.(b)	0.58 ± 0.01	3028.35 ± 370.09	1445.54 ± 329.95	0.25 ± 0.03	0.12 ± 0.03

Table 7. Comparison of the implicit midpoint and generalized leapfrog integrators on sampling from the Bayesian logistic regression posterior on the heart disease dataset.

G. Volume Preservation and Symmetry Metrics

Here we describe how we compute metrics related to volume preservation and symmetry. Let $(q_1, \ldots, q_n)$ be samples generated by Hamiltonian Monte Carlo with numerical integrator $\Phi : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m \times \mathbb{R}^m$; each q_i is an element of $\mathbb{R}^m$.

G.1. Reversibility

For each sample q_i, generate $p_i|q_i \sim \mathrm{Normal}(0, \mathbb{G}(q_i))$ and compute $(q_i', p_i') = \Phi(q_i, p_i)$. Now compute $(q_i'', -p_i'') = \Phi(q_i', -p_i')$. The *violation of reversibility* is defined by

$$\sqrt{\|(q_i, p_i') - (q_i'', p_i'')\|_2^2}. \tag{48}$$

If the numerical integrator is reversible, this norm will be zero. In our metrics, we report the median violation of reversibility.

G.2. Volume Preservation

Let $z = (q, p)$ and identify $\Phi(z) \equiv \Phi(q, p)$. Define $f_j(z) = \frac{\Phi(z_1,\ldots,z_j+\eta/2,\ldots,z_{2m}) - \Phi(z_1,\ldots,z_j-\eta/2,\ldots,z_{2m})}{\eta}$ for $\eta = 1 \times 10^{-5}$ (except for the Fitzhugh-Nagumo model where we set $\eta = 1 \times 10^{-3}$ for numerical reasons), which is the central difference formula that approximates $\frac{\partial}{\partial z_j}\Phi(z)$. We compute the approximation to the Jacobian of Φ by constructing,

$$\nabla\Phi(z) \approx F(z) \stackrel{\text{def.}}{=} \begin{pmatrix} f_1(z) & f_2(z) & \cdots & f_{2m}(z) \end{pmatrix} \in \mathbb{R}^{2m \times 2m}. \tag{49}$$

For each sample q_i, generate $p_i|q_i \sim \mathrm{Normal}(0, \mathbb{G}(q_i))$ and set $z_i = (q_i, p_i)$. The *violation of volume preservation* is defined by

$$\left|\det(F(z_i)) - 1\right|. \tag{50}$$

If the numerical integrator is volume preserving, this difference will be zero. In our metrics, we report the median violation of volume preservation.

H. Detailed Balance from Reversibility and Volume Preservation

Let $\Phi : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m \times \mathbb{R}^m$ be a numerical integrator satisfying the following two properties as described in the main text:

(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.

Given $(q, p) \in \mathbb{R}^m \times \mathbb{R}^m$, define the *momentum flip operator* by $\Psi(q, p) = (q, -p)$. For Markov chain Monte Carlo, we then define the Markov chain *proposal* operator by $\mathfrak{P} \stackrel{\text{def.}}{=} \Psi \circ \Phi$; because Φ satisfies property (ii) we have that $\mathfrak{P} \circ \mathfrak{P} = \Psi \circ \Phi \circ \Psi \circ \Phi = \mathrm{Id}$ so that the proposal operator is self-inverse $\mathfrak{P}^{-1} = \mathfrak{P}$. Moreover, Ψ has unit Jacobian determinant: $|\det(\nabla\Psi)| = 1$; therefore, since Φ has unit Jacobian determinant by property (i), $\mathfrak{P}$ also has unit Jacobian determinant:

$$
\begin{aligned}
|\det(\nabla\mathfrak{P}(q, p))| &= |\det(\nabla\Psi(\Phi(q, p))\nabla\Phi(q, p))| && (51)\\
&= |\det(\nabla\Psi(\Phi(q, p)) \cdot \det(\nabla\Phi(q, p))| && (52)\\
&= 1 && (53)
\end{aligned}
$$

Given a current position (q, p) of the Markov chain, the proposal for the next state of the Markov chain $(q', p') \stackrel{\text{def.}}{=} \mathfrak{P}(q, p)$ is accepted with probability

$$
\min\left\{1, \frac{\pi(q', p')}{\pi(q, p)}\right\}, \tag{54}
$$

where $\pi : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ is the target distribution. (Notice that π must only be specified up to a constant.) Thus, the Markov chain *transition* operator is defined by,

$$
\mathfrak{T}(q, p) = \begin{cases} (q', p') & \text{w.p. } \min\left\{1, \frac{\pi(q', p')}{\pi(q, p)}\right\} \\ (q, p) & \text{else.} \end{cases} \tag{55}
$$

We say that detailed balance holds if for all sets $A, B \subset \mathbb{R}^m \times \mathbb{R}^m$ we have,

$$
\Pr_{(q,p)\sim\pi}\left[(q, p) \in A \text{ and } \mathfrak{T}(q, p) \in B\right] = \Pr_{(q,p)\sim\pi}\left[(q, p) \in B \text{ and } \mathfrak{T}(q, p) \in A\right] \tag{56}
$$

Let $z = (q, p)$. Expanding we compute,

$$
\begin{aligned}
\Pr_{z\sim\pi}\left[z \in A \text{ and } \mathfrak{T}(z) \in B\right] &= \int_A \pi(z) \cdot \mathbf{1}\left\{\mathfrak{P}(z) \in B\right\} \cdot \min\left\{1, \frac{\pi(\mathfrak{P}(z))}{\pi(z)}\right\} \mathrm{d}z && (57)\\
&= \int_{A\cap\mathfrak{P}(B)} \pi(z) \cdot \min\left\{1, \frac{\pi(\mathfrak{P}(z))}{\pi(z)}\right\} \mathrm{d}z && (58)\\
&\stackrel{z'=\mathfrak{P}(z)}{=} \int_{\mathfrak{P}(A)\cap B} \pi(\mathfrak{P}(z')) \cdot \min\left\{1, \frac{\pi(z')}{\pi(\mathfrak{P}(z'))}\right\} \mathrm{d}z' && (59)\\
&= \int_{\mathfrak{P}(A)\cap B} \pi(z') \cdot \min\left\{1, \frac{\pi(\mathfrak{P}(z'))}{\pi(z')}\right\} \mathrm{d}z' && (60)\\
&= \int_B \pi(z') \cdot \mathbf{1}\left\{\mathfrak{P}(z') \in A\right\} \cdot \min\left\{1, \frac{\pi(\mathfrak{P}(z'))}{\pi(z')}\right\} \mathrm{d}z' && (61)\\
&= \Pr_{z\sim\pi}\left[z \in B \text{ and } \mathfrak{T}(z) \in A\right] && (62)
\end{aligned}
$$

showing that detailed balance holds. The change-of-variables in eq. (59) does not incur a Jacobian determinant correction since $\mathfrak{P}$ has unit Jacobian determinant.

I. Implementation of Integrators

I.1. Implementations of the Generalized Leapfrog Integrator

As the purpose of this research is to compare two integrators for Hamiltonian Monte Carlo, we wish to be precise about how these numerical methods have been implemented.

The first implementation of the generalized leapfrog integrator is presented in algorithm 2. The system eqs. (7) to (9) was described as "naive" because it appears to ignore important structural properties of the equations of motion in eqs. (2) and (3) that would accelerate a step of the generalized leapfrog integrator. For instance, in eq. (7), the metric $\mathbb{G}(q)$ may be precomputed because it is an invariant of the fixed-point relation. Another example is that $\mathbb{G}^{-1}(q)\bar{p}$ is an invariant quantity of eq. (8) and need not be recomputed in each fixed-point iteration. Thus, we see that the generalized leapfrog integrator, when efficiently implemented, has an important computational advantage in that certain invariant quantities can be "cached" when finding fixed points. A more complicated implementation of the generalized leapfrog integrator with caching is presented in given in algorithm 4. We stress that algorithms 2 and 4 perform the same calculation.

I.2. Implementations of the Implicit Midpoint Integrator

In addition to the implementation of the implicit midpoint method described in algorithm 3, we also consider a variant advocated by Leimkuhler & Reich (2005) and present the implementation in algorithm 5. The essential difference between algorithms 3 and 5 is whether or not the implicitly-defined update computes the terminal point of the step (algorithm 3) or the midpoint of the step (algorithm 5). We stress that algorithms 3 and 5 perform the same calculation when $\delta = 0$, which is easily verified by plugging eq. (66) into eq. (67) and comparing to eq. (10). Unlike the generalized leapfrog integrator, the implicit midpoint integrator does not enjoy the ability to cache intermediate computations.

I.3. Remarks on Computational Complexity

In discussing the computational complexity of the generalized leapfrog and implicit midpoint methods, we will assume that the functions defining the fixed-point relations are *contraction maps*. Contraction maps in a complete metric space

Algorithm 4 (G.L.F.(b)) The procedure for a single step of integrating Hamiltonian dynamics using the efficient implementation of the generalized leapfrog integrator.

1: **Input**: Log-posterior $\mathcal{L} : \mathbb{R}^m \to \mathbb{R}$, Riemannian metric $\mathbb{G} : \mathbb{R}^m \to \mathbb{R}^{m\times m}$, initial position and momentum variables $(q, p) \in \mathbb{R}^m \times \mathbb{R}^m$, integration step-size size $\epsilon \in \mathbb{R}$.
2: Precompute $\mathbb{G}^{-1}(q)$, $\frac{\partial}{\partial q_i}\mathbb{G}(q)$, and define

$$A_i \stackrel{\text{def.}}{=} \frac{\partial}{\partial q_i}\mathcal{L}(q) + \frac{1}{2}\text{trace}\left(\mathbb{G}^{-1}(q)\frac{\partial}{\partial q_i}\mathbb{G}(q)\right) \tag{63}$$

for $i = 1, \ldots, m$.
3: Use algorithm 1 with tolerance δ and initial guess p to solve for $\bar{p}$,

$$\bar{p} = \underbrace{p - \frac{\epsilon}{2}\left(A_1 - \frac{1}{2}(\mathbb{G}^{-1}(q)\bar{p})^\top \frac{\partial}{\partial q_1}\mathbb{G}(q)(\mathbb{G}^{-1}(q)\bar{p}), \ldots, A_m - \frac{1}{2}(\mathbb{G}^{-1}(q)\bar{p})^\top \frac{\partial}{\partial q_m}\mathbb{G}(q)(\mathbb{G}^{-1}(q)\bar{p})\right)^\top}_{f(\bar{p})}. \tag{64}$$

4: Precompute $\mathbb{G}(q)^{-1}\bar{p}$.
5: Use algorithm 1 with tolerance δ and initial guess q to solve for q',

$$q' = \underbrace{q + \frac{\epsilon}{2}\left(\mathbb{G}^{-1}(q)\bar{p} + \mathbb{G}^{-1}(q')\bar{p}\right)}_{f(q')}. \tag{65}$$

6: Compute p' using eq. (9), which is an explicit update.
7: **Return**: $(q', p') \in \mathbb{R}^m \times \mathbb{R}^m$.

Algorithm 5 (I.M.(b)) The procedure for a single step of integrating Hamiltonian dynamics using the implicit midpoint integrator as advocated by Leimkuhler & Reich (2005).

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 $(\bar{q}, \bar{p})$,

$$\begin{pmatrix} \bar{q} \\ \bar{p} \end{pmatrix} \overset{\text{def.}}{=} \underbrace{\begin{pmatrix} q \\ p \end{pmatrix} + \frac{\epsilon}{2} \begin{pmatrix} \nabla_p H(\bar{q}, \bar{p}) \\ -\nabla_q H(\bar{q}, \bar{p}) \end{pmatrix}}_{f(\bar{q}, \bar{p})}. \tag{66}$$

3: Compute the explicit update

$$\begin{pmatrix} q' \\ p' \end{pmatrix} = \begin{pmatrix} \bar{q} \\ \bar{p} \end{pmatrix} + \frac{\epsilon}{2} \begin{pmatrix} \nabla_p H(\bar{q}, \bar{p}) \\ -\nabla_q H(\bar{q}, \bar{p}) \end{pmatrix}. \tag{67}$$

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

guarantee that fixed point equations have unique solutions and that these solutions are reached via fixed point iteration for any initial condition. Moreover, contraction maps have known convergence rates, which are convenient for analysis.

Definition 2. The L_∞ distance on $\mathbb{R}^m$ is defined by

$$d_\infty(z, z') \overset{\text{def.}}{=} \max_{i=1,\ldots,m} |z_i - z_i'|. \tag{68}$$

We will restrict our attention to the L_∞ distance as it is used as our convergence criterion in algorithm 1.

Definition 3. A contraction map on $\mathbb{R}^m$ is a Lipschitz function $f : \mathbb{R}^m \to \mathbb{R}^m$ whose Lipschitz constant is less than one; that is,

$$d_\infty(f(z), f(z')) \leq L \cdot d_\infty(z, z') \tag{69}$$

for all $z, z' \in \mathbb{R}^m$ and where $L \in [0, 1)$.

Theorem 3 (Banach Fixed Point Theorem)**.** Let $f : \mathbb{R}^m \to \mathbb{R}^m$ be a contraction map with Lipschitz constant $L \in [0, 1)$. Then (i) the equation $z = f(z)$ has a unique solution; (ii) the fixed point iterations $z_{n+1} = f(z_n)$ converge to z from any $z_0 \in \mathbb{R}^m$; (iii) the distance between iterates satisfies $d_\infty(z_{n+1}, z_n) \leq L^n d_\infty(z_1, z_0)$.

In assessing convergence to the unique solution using fixed point iteration, algorithm 1 demands that $d_\infty(z_{n+1}, z_n) \leq \delta$ for some convergence tolerance $\delta > 0$. Assuming that $d_\infty(z_1, z_0) > 0$, we can rearrange $L^n d_\infty(z_1, z_0) \leq \delta$ to give a sufficient condition on the number of fixed point iterates n to guarantee that $d_\infty(z_{n+1}, z_n) \leq \delta$. Namely:

Lemma 2. Assume $d_\infty(z_1, z_0) > 0$. Then the choice

$$n = \left\lceil \frac{\log \delta - \log d_\infty(z_1, z_0)}{\log L} \right\rceil \tag{70}$$

is sufficient to ensure that $d_\infty(z_{n+1}, z_n) \leq \delta$.

Proof. From the Banach fixed point theorem we know that $d_\infty(z_{n+1}, z_n) \leq L^n d_\infty(z_1, z_0)$. Therefore, we seek to establish when the right-hand side of the inequality is less than δ. Rearranging and noting that $\log L$ is negative (since f is a contraction map) yields

$$n \geq \frac{\log \delta - \log d_\infty(z_1, z_0)}{\log L}. \tag{71}$$

It makes sense to use the smallest integer n such that the $d_\infty(z_{n+1}, z_n) \leq \delta$ so computing the ceiling of the right-hand side yields the result. □

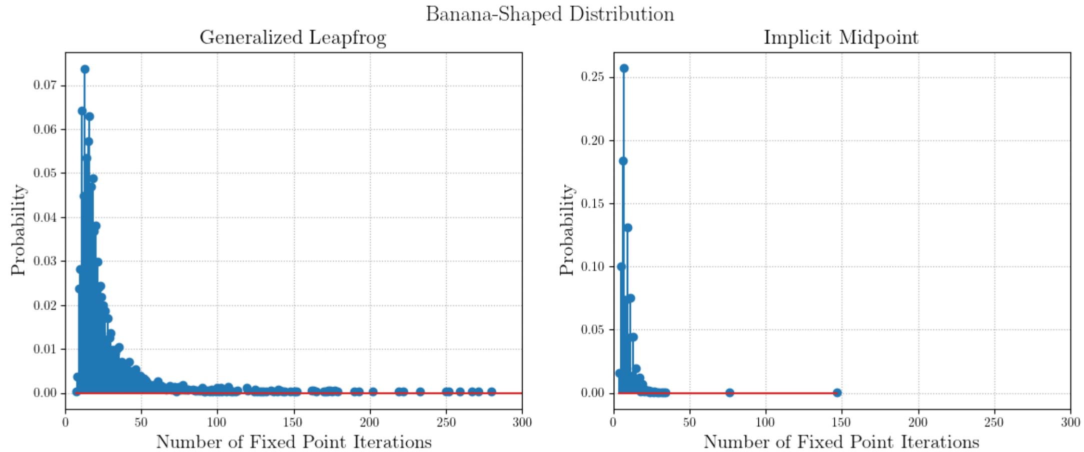


Figure 6. Visualization of the relative frequency of the number of fixed point iterations in the implicit midpoint and generalized leapfrog integration procedure for the banana-shaped distribution.

Under the assumption that functions defining fixed point equations are contractions, we can give an approximate comparison of the computational complexity of the implicit midpoint and generalized leapfrog algorithms. In integrating the dynamics corresponding to a Hamiltonian as in eq. (1) (refer to eqs. (2) and (3)), there are several operations of notable computational complexity; these are (i) computing the gradient of the log-likelihood, (ii) computing the inverse of the Riemannian metric, and (iii) computing the derivatives of the Riemannian metric. To make this slightly more formal, let $\mathrm{Cost}(\nabla\mathcal{L})$, $\mathrm{Cost}(\mathbb{G}^{-1})$, and $\mathrm{Cost}(\nabla\mathbb{G})$ denote some notion of computational complexity associated to these three quantities; we assume that any remaining arithmetic operations used in the integration of Hamilton's mechanics have negligible computational cost.

Example 1. For the implicit midpoint integrator, each of the above quantities (i)-(iii) must be computed within each fixed point iteration. If L_{imp} denotes the Lipschitz constant associated to the map f in eq. (10), then the total computational cost of the implicit midpoint integrator is

$$\left\lceil \frac{\log\delta - \log d_\infty(f(z_0), z_0)}{\log L_{\mathrm{imp}}} \right\rceil \cdot \left(\mathrm{Cost}(\nabla\mathcal{L}) + \mathrm{Cost}(\mathbb{G}^{-1}) + \mathrm{Cost}(\nabla\mathbb{G})\right). \tag{72}$$

Example 2. The situation is rather different for the generalized leapfrog integrator as expressed in algorithm 4. The fixed point equation in eq. (64) requires that we compute quantities (i)-(iii); however, these are invariant of the fixed point equation in eq. (64), which therefore does not incur additional cost beyond the sum of $\mathrm{Cost}(\nabla\mathcal{L})$, $\mathrm{Cost}(\mathbb{G}^{-1})$, and $\mathrm{Cost}(\nabla\mathbb{G})$. For the second fixed point equation, let L_{glf} denote the Lipschitz constant of the contraction map f in eq. (65); each iteration requires computing the inverse of the Riemannian metric. Thus, the computational cost incurred by the second fixed point equation is

$$\left\lceil \frac{\log\delta - \log d_\infty(f(q_0), q_0)}{\log L_{\mathrm{glf}}} \right\rceil \cdot \mathrm{Cost}(\mathbb{G}^{-1}). \tag{73}$$

The final step of the generalized leapfrog integrator requires an explicit update wherein we compute quantities (i)-(iii) using the updated position variable. In total, the cost of a step of the generalized leapfrog integrator is therefore,

$$\left\lceil \frac{\log\delta - \log d_\infty(f(q_0), q_0)}{\log L_{\mathrm{glf}}} \right\rceil \cdot \mathrm{Cost}(\mathbb{G}^{-1}) + 2\cdot\left(\mathrm{Cost}(\nabla\mathcal{L}) + \mathrm{Cost}(\mathbb{G}^{-1}) + \mathrm{Cost}(\nabla\mathbb{G})\right). \tag{74}$$

I.4. Counts of Fixed Point Iterations

In this section we compute the number of fixed point iterations performed by both the implicit midpoint and generalized leapfrog integrators on the banana, stochastic volatility, and Fitzhugh-Nagumo experiments. We report the relative frequency of each number of the fixed point iterations over the course of sampling. We note that this is for context only, and that the

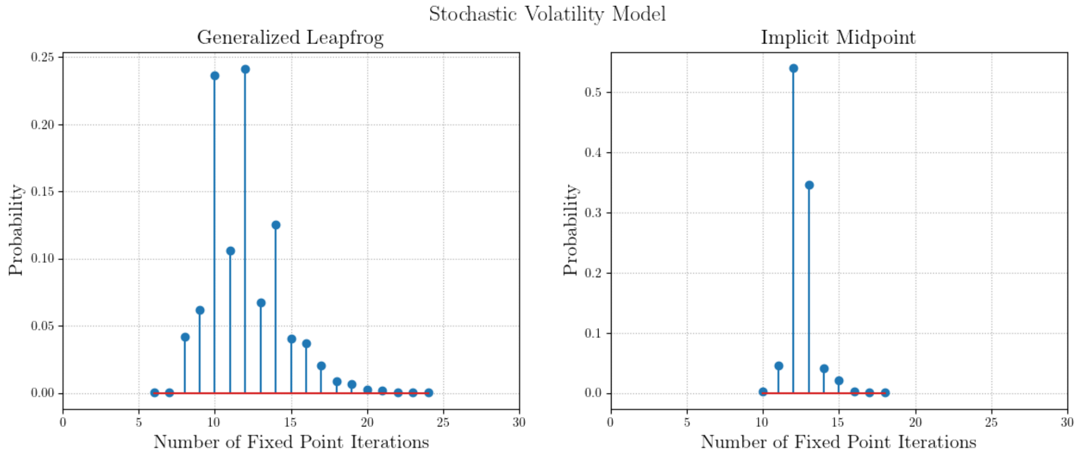


Figure 7. Visualization of the relative frequency of the number of fixed point iterations in the implicit midpoint and generalized leapfrog integration procedure for the stochastic volatility model.

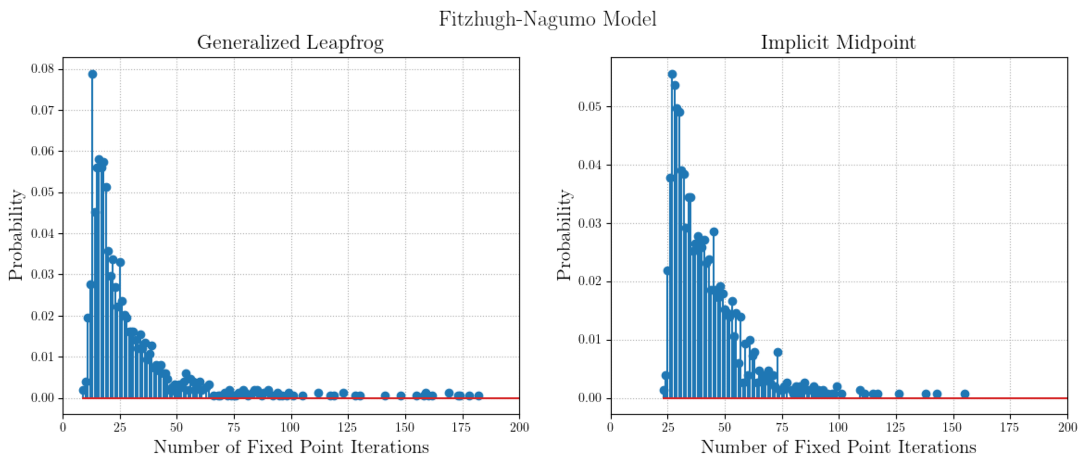


Figure 8. Visualization of the relative frequency of the number of fixed point iterations in the implicit midpoint and generalized leapfrog integration procedure for the Fitzhugh-Nagumo model.

precise number of fixed point iterations consumed by either integrator are not comparable to one another; the reason for this is that the amount of computation differs for the implicit midpoint and generalized leapfrog methods as described in appendix I.3. For the generalized leapfrog method, we report the total number of iterations to solve both fixed point relations to the prescribed tolerance. Results are shown in figs. 6 to 8 for a convergence tolerance of $\delta = 1 \times 10^{-6}$.

J. Proof of Proposition 1

Proof. The implicit midpoint integrator conserves quadratic first integrals (see theorem 2) and, in this case, the Hamiltonian energy is itself a quadratic function. □

K. Randomized Step Experimental Design

In the main text we have considered a grid search over the number of integration steps and use this number of steps in generating every proposal using either the implicit midpoint or generalized leapfrog integrators. An alternative experimental design is to randomize the number of integration steps for each proposal by sampling the number of steps uniformly between one and some upper bound. When the Hamiltonian is separable, the leapfrog integrator with a single step recovers the Metropolis-adjusted Langevin algorithm, which has known ergodicity properties. Therefore, it may be anticipated that randomizing the number of integration steps may lead to improved ergodicity when the proposal is computed by the generalized leapfrog method. In this appendix, we consider randomizing the number of integration steps for both the implicit midpoint and generalized leapfrog integrators.

K.1. Banana-Shaped Distribution

Results for the banana-shaped distribution when using the randomized number of steps experimental design are presented in table 8.

K.2. Neal Funnel

Results for Neal's funnel distribution when using the randomized number of steps experimental design are presented in table 9.

K.3. Stochastic Volatility Model

Results for the stochastic volatility model when using the randomized number of steps experimental design are presented in table 10.

Step Size	Max. Steps	Method	Acc. Prob.	Time (Sec.)	Mean ESS	Min. ESS	Mean ESS / Sec.	Min. ESS / Sec.
0.1	5	G.L.F.(a)	0.66 ± 0.01	298.11 ± 4.99	240.50 ± 13.01	140.54 ± 10.54	0.81 ± 0.04	0.47 ± 0.03
		G.L.F.(b)	0.66 ± 0.01	103.90 ± 1.64	246.99 ± 10.57	149.70 ± 8.26	2.37 ± 0.08	1.44 ± 0.07
		I.M.(a)	0.99 ± 0.00	65.48 ± 0.97	372.91 ± 20.52	276.00 ± 18.41	5.69 ± 0.28	4.20 ± 0.24
		I.M.(b)	0.99 ± 0.00	59.90 ± 0.98	375.99 ± 23.34	268.26 ± 15.80	6.32 ± 0.45	4.51 ± 0.30
	10	G.L.F.(a)	0.60 ± 0.01	420.38 ± 8.68	541.71 ± 13.63	373.56 ± 16.10	1.29 ± 0.05	0.89 ± 0.04
		G.L.F.(b)	0.59 ± 0.00	153.94 ± 3.34	560.73 ± 12.99	370.12 ± 13.86	3.66 ± 0.12	2.41 ± 0.11
		I.M.(a)	0.98 ± 0.00	111.84 ± 1.64	1205.85 ± 38.74	945.21 ± 35.13	10.80 ± 0.37	8.47 ± 0.34
		I.M.(b)	0.98 ± 0.00	100.24 ± 1.79	1212.95 ± 41.28	948.95 ± 47.69	12.12 ± 0.42	9.47 ± 0.46
	50	G.L.F.(a)	0.31 ± 0.00	1158.96 ± 13.35	744.51 ± 34.97	603.15 ± 17.87	0.64 ± 0.03	0.52 ± 0.02
		G.L.F.(b)	0.31 ± 0.01	451.75 ± 4.30	754.07 ± 37.32	602.83 ± 42.83	1.67 ± 0.09	1.34 ± 0.10
		I.M.(a)	0.97 ± 0.00	473.24 ± 4.62	6306.89 ± 136.15	2568.88 ± 34.70	13.33 ± 0.28	5.43 ± 0.07
		I.M.(b)	0.97 ± 0.00	419.67 ± 3.65	6345.83 ± 165.83	2597.81 ± 50.49	15.14 ± 0.46	6.20 ± 0.14

Table 8. Comparison of the implicit midpoint and generalized leapfrog integrators on sampling from the banana-shaped distribution when a randomized number of steps is used.

Max. 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	786.17 ± 62.99	3535.77 ± 45.31	120.34 ± 15.07	4.69 ± 0.31	0.16 ± 0.02
		I.M.(a)	1.00 ± 0.00	655.70 ± 33.04	3538.75 ± 28.40	132.25 ± 12.73	5.48 ± 0.21	0.21 ± 0.02
		I.M.(b)	1.00 ± 0.00	579.39 ± 40.26	3545.94 ± 32.44	133.26 ± 10.57	6.29 ± 0.31	0.23 ± 0.02
	0.2	G.L.F.(a)	0.96 ± 0.00	968.70 ± 58.97	11305.65 ± 109.81	525.33 ± 15.22	11.96 ± 0.59	0.56 ± 0.04
		I.M.(a)	0.99 ± 0.00	818.02 ± 32.74	11809.38 ± 128.38	502.90 ± 33.26	14.59 ± 0.51	0.63 ± 0.05
		I.M.(b)	0.99 ± 0.00	728.94 ± 21.80	11959.25 ± 115.00	522.38 ± 22.78	16.52 ± 0.51	0.72 ± 0.04
	0.5	G.L.F.(a)	0.43 ± 0.01	1420.67 ± 49.91	1857.57 ± 172.54	754.50 ± 85.59	1.30 ± 0.11	0.53 ± 0.06
		I.M.(a)	0.89 ± 0.00	1596.95 ± 48.55	8468.24 ± 54.32	3018.45 ± 93.75	5.34 ± 0.15	1.91 ± 0.11
		I.M.(b)	0.89 ± 0.00	1458.31 ± 53.48	8422.27 ± 57.72	3034.97 ± 81.87	5.83 ± 0.19	2.10 ± 0.09

Table 9. Comparison of the implicit midpoint and the naive generalized leapfrog integrators on sampling from Neal's funnel distribution when using a randomized number of integration steps.

Method	Acc. Prob.	Time (Sec.)	Mean ESS	Min. ESS	Mean ESS (Sec.)	Min. ESS (Sec.)	Volume Preservation Median	Volume Preservation 90^{th}-Per.	Symmetry Median	Symmetry 90^{th}-Per.
G.L.F.(a)	0.85 ± 0.0	2367.93 ± 16.1	335.80 ± 15.3	132.96 ± 3.9	0.14 ± 0.0	0.06 ± 0.0	6.8e−07	2.9e−06	4.9e−06	2.4e−05
G.L.F.(b)	0.85 ± 0.0	2196.52 ± 8.5	331.84 ± 12.1	139.30 ± 3.7	0.15 ± 0.0	0.06 ± 0.0	7.1e−07	2.9e−06	4.8e−06	2.7e−05
I.M.(a)	0.86 ± 0.0	2434.74 ± 7.9	316.98 ± 15.4	131.79 ± 3.4	0.13 ± 0.0	0.05 ± 0.0	8.7e−08	2.3e−07	1.3e−06	2.4e−06
I.M.(b)	0.87 ± 0.0	2399.22 ± 6.4	342.11 ± 14.0	133.76 ± 3.7	0.14 ± 0.0	0.06 ± 0.0	1.7e−07	4.4e−07	2.8e−06	4.7e−06

Table 10. Comparison of the implicit midpoint generalized leapfrog integrators on the stochastic volatility model when using a randomized number of integration steps.

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)	0.75 ± 0.01	4084.16 ± 138.81	289.87 ± 12.89	251.36 ± 12.33	0.07 ± 0.01	0.06 ± 0.00
		G.L.F.(b)	0.74 ± 0.01	1193.08 ± 61.74	301.44 ± 16.67	262.36 ± 15.38	0.26 ± 0.02	0.23 ± 0.02
		I.M.(a)	0.95 ± 0.00	4353.40 ± 104.50	237.73 ± 8.67	219.04 ± 11.18	0.05 ± 0.00	0.05 ± 0.00
		I.M.(b)	0.95 ± 0.00	4539.73 ± 132.69	228.33 ± 8.69	191.36 ± 9.91	0.05 ± 0.00	0.04 ± 0.00
	2	G.L.F.(a)	0.75 ± 0.01	6265.35 ± 187.70	620.36 ± 17.86	553.05 ± 20.54	0.10 ± 0.00	0.09 ± 0.00
		G.L.F.(b)	0.74 ± 0.01	1743.26 ± 61.19	592.04 ± 12.90	518.75 ± 16.25	0.35 ± 0.02	0.30 ± 0.02
		I.M.(a)	0.94 ± 0.00	7051.31 ± 190.41	612.50 ± 23.74	543.62 ± 27.06	0.09 ± 0.00	0.08 ± 0.00
		I.M.(b)	0.94 ± 0.00	6549.13 ± 146.01	626.26 ± 16.04	566.83 ± 19.62	0.10 ± 0.00	0.09 ± 0.00
	5	G.L.F.(a)	0.74 ± 0.01	12447.27 ± 560.25	726.90 ± 53.55	629.11 ± 64.83	0.06 ± 0.01	0.05 ± 0.01
		G.L.F.(b)	0.74 ± 0.01	2974.54 ± 90.05	756.32 ± 49.99	637.53 ± 61.54	0.26 ± 0.02	0.22 ± 0.02
		I.M.(a)	0.94 ± 0.00	13331.24 ± 263.47	1320.95 ± 77.86	1106.66 ± 81.97	0.10 ± 0.01	0.08 ± 0.01
		I.M.(b)	0.93 ± 0.00	12501.60 ± 457.80	1441.09 ± 54.15	1229.29 ± 63.45	0.12 ± 0.01	0.10 ± 0.01

Table 11. Comparison of the implicit midpoint and generalized leapfrog integrators on sampling from the posterior of the Fitzhugh-Nagumo model when a randomized number of integration steps are used.

Step Size	Max. Steps	Method	Acc. Prob.	Mean ESS	Min. ESS	Mean ESS / Sec.	Min. ESS / Sec.
0.1	5	G.L.F.(a)	1.00 ± 0.00	286.35 ± 4.21	208.68 ± 9.49	3.50 ± 0.06	2.55 ± 0.12
		G.L.F.(b)	1.00 ± 0.00	283.91 ± 5.17	212.44 ± 8.27	7.18 ± 0.16	5.39 ± 0.25
		I.M.(a)	1.00 ± 0.00	283.77 ± 4.26	222.26 ± 9.88	3.58 ± 0.06	2.80 ± 0.13
		I.M.(b)	1.00 ± 0.00	286.73 ± 4.68	221.40 ± 12.74	4.15 ± 0.07	3.20 ± 0.19
	10	G.L.F.(a)	1.00 ± 0.00	1010.60 ± 11.69	893.84 ± 20.10	7.22 ± 0.10	6.39 ± 0.16
		G.L.F.(b)	1.00 ± 0.00	1002.32 ± 15.93	873.10 ± 21.47	16.45 ± 0.27	14.33 ± 0.35
		I.M.(a)	1.00 ± 0.00	1005.16 ± 12.23	871.72 ± 15.41	7.50 ± 0.07	6.50 ± 0.10
		I.M.(b)	1.00 ± 0.00	1021.59 ± 8.10	887.65 ± 27.54	8.48 ± 0.19	7.37 ± 0.28
	50	G.L.F.(a)	1.00 ± 0.00	14410.27 ± 147.75	13303.56 ± 189.57	23.89 ± 0.19	22.05 ± 0.26
		G.L.F.(b)	1.00 ± 0.00	14411.21 ± 90.58	13450.61 ± 91.65	60.17 ± 0.33	56.16 ± 0.29
		I.M.(a)	1.00 ± 0.00	14548.15 ± 121.78	13483.46 ± 126.40	24.79 ± 0.19	22.98 ± 0.22
		I.M.(b)	1.00 ± 0.00	14519.88 ± 118.88	13317.24 ± 127.50	28.52 ± 0.25	26.16 ± 0.29
1.0	5	G.L.F.(a)	0.23 ± 0.00	904.47 ± 54.85	542.07 ± 69.66	4.10 ± 0.23	2.45 ± 0.31
		G.L.F.(b)	0.23 ± 0.00	900.64 ± 23.10	596.22 ± 40.82	12.66 ± 0.37	8.37 ± 0.56
		I.M.(a)	0.89 ± 0.00	12294.41 ± 127.45	10991.38 ± 217.01	47.59 ± 0.55	42.55 ± 0.87
		I.M.(b)	0.89 ± 0.00	12506.83 ± 182.42	11280.48 ± 316.77	50.82 ± 0.68	45.82 ± 1.21
	10	G.L.F.(a)	0.19 ± 0.00	663.43 ± 35.45	417.15 ± 38.40	2.25 ± 0.11	1.41 ± 0.12
		G.L.F.(b)	0.19 ± 0.00	568.08 ± 31.76	333.88 ± 39.54	6.32 ± 0.31	3.69 ± 0.40
		I.M.(a)	0.87 ± 0.00	8312.69 ± 53.82	7449.40 ± 86.13	17.09 ± 0.23	15.31 ± 0.23
		I.M.(b)	0.88 ± 0.00	8306.90 ± 46.45	7407.88 ± 136.86	18.24 ± 0.19	16.27 ± 0.36
	50	G.L.F.(a)	0.10 ± 0.00	218.25 ± 33.24	119.54 ± 22.69	0.41 ± 0.06	0.22 ± 0.04
		G.L.F.(b)	0.10 ± 0.00	199.52 ± 25.14	110.61 ± 21.56	1.26 ± 0.15	0.70 ± 0.13
		I.M.(a)	0.85 ± 0.00	6422.28 ± 65.69	5107.53 ± 212.97	2.63 ± 0.04	2.09 ± 0.08
		I.M.(b)	0.85 ± 0.00	6693.87 ± 72.47	5565.30 ± 272.50	2.90 ± 0.05	2.41 ± 0.13

Table 12. Comparison of the implicit midpoint and generalized leapfrog integrators on sampling from the Bayesian logistic regression posterior on the Breast Cancer dataset when using a randomized number of integration steps.

K.4. Fitzhugh-Nagumo Differential Equation Model

Results for the Fitzhugh-Nagumo posterior when using the randomized number of steps experimental design are presented in table 11.

K.5. Bayesian Logistic Regression

Results for the Bayesian logistic regression models with a randomized number of steps are presented in tables 12 to 14.

Step Size	Max. Steps	Method	Acc. Prob.	Mean ESS	Min. ESS	Mean ESS / Sec.	Min. ESS / Sec.
0.1	5	G.L.F.(a)	1.00 ± 0.00	286.98 ± 4.54	234.68 ± 7.46	2.97 ± 0.05	2.43 ± 0.07
		G.L.F.(b)	1.00 ± 0.00	280.75 ± 3.93	214.84 ± 8.31	6.39 ± 0.09	4.89 ± 0.19
		I.M.(a)	1.00 ± 0.00	285.13 ± 6.19	227.96 ± 13.22	2.73 ± 0.07	2.18 ± 0.13
		I.M.(b)	1.00 ± 0.00	273.54 ± 6.51	217.14 ± 11.43	2.99 ± 0.07	2.37 ± 0.12
	10	G.L.F.(a)	1.00 ± 0.00	1015.63 ± 14.83	877.03 ± 31.16	6.05 ± 0.09	5.23 ± 0.19
		G.L.F.(b)	1.00 ± 0.00	1009.53 ± 9.46	862.28 ± 17.84	14.88 ± 0.13	12.72 ± 0.31
		I.M.(a)	1.00 ± 0.00	987.06 ± 10.31	832.63 ± 24.35	5.44 ± 0.07	4.59 ± 0.13
		I.M.(b)	1.00 ± 0.00	1010.27 ± 9.23	867.30 ± 19.17	6.38 ± 0.11	5.47 ± 0.10
	50	G.L.F.(a)	1.00 ± 0.00	14735.31 ± 75.07	13759.10 ± 141.77	20.13 ± 0.17	18.80 ± 0.26
		G.L.F.(b)	1.00 ± 0.00	14774.00 ± 99.12	13680.31 ± 124.54	54.81 ± 0.38	50.75 ± 0.49
		I.M.(a)	1.00 ± 0.00	14748.16 ± 100.96	13445.33 ± 111.76	18.49 ± 0.13	16.85 ± 0.14
		I.M.(b)	1.00 ± 0.00	14721.71 ± 120.84	13834.19 ± 162.68	21.64 ± 0.16	20.34 ± 0.26
1.0	5	G.L.F.(a)	0.68 ± 0.00	7161.52 ± 72.09	6678.78 ± 98.85	37.13 ± 0.33	34.63 ± 0.48
		G.L.F.(b)	0.68 ± 0.00	7132.61 ± 58.58	6617.77 ± 98.49	118.67 ± 1.61	110.15 ± 2.30
		I.M.(a)	0.97 ± 0.00	16750.80 ± 134.04	15466.43 ± 76.75	49.44 ± 0.42	45.65 ± 0.26
		I.M.(b)	0.97 ± 0.00	16635.79 ± 131.54	15402.22 ± 176.98	51.11 ± 0.39	47.31 ± 0.44
	10	G.L.F.(a)	0.68 ± 0.00	6498.21 ± 53.62	5885.25 ± 55.32	18.86 ± 0.19	17.08 ± 0.20
		G.L.F.(b)	0.68 ± 0.00	6655.24 ± 61.17	6222.74 ± 75.56	67.37 ± 0.79	62.99 ± 0.91
		I.M.(a)	0.97 ± 0.00	10858.03 ± 109.54	10245.67 ± 118.98	17.69 ± 0.18	16.70 ± 0.20
		I.M.(b)	0.97 ± 0.00	10672.12 ± 108.02	10006.57 ± 128.94	18.54 ± 0.20	17.38 ± 0.24
	50	G.L.F.(a)	0.66 ± 0.00	4999.26 ± 50.94	4518.13 ± 54.31	3.41 ± 0.04	3.08 ± 0.04
		G.L.F.(b)	0.66 ± 0.00	5098.68 ± 72.88	4601.08 ± 121.45	13.05 ± 0.27	11.78 ± 0.36
		I.M.(a)	0.97 ± 0.00	9176.25 ± 69.76	8681.47 ± 129.00	3.27 ± 0.03	3.10 ± 0.05
		I.M.(b)	0.97 ± 0.00	9176.41 ± 58.95	8657.74 ± 87.61	3.44 ± 0.03	3.24 ± 0.03

Table 13. Comparison of the implicit midpoint and generalized leapfrog integrators on sampling from the Bayesian logistic regression posterior on the Diabetes dataset when using a randomized number of steps.

Step Size	Max. Steps	Method	Acc. Prob.	Mean ESS	Min. ESS	Mean ESS / Sec.	Min. ESS / Sec.
0.1	5	G.L.F.(a)	1.00 ± 0.00	288.33 ± 4.45	213.23 ± 12.18	2.74 ± 0.05	2.03 ± 0.12
		G.L.F.(b)	1.00 ± 0.00	285.05 ± 2.89	208.93 ± 12.85	6.13 ± 0.05	4.49 ± 0.28
		I.M.(a)	1.00 ± 0.00	284.53 ± 2.43	200.77 ± 9.67	2.86 ± 0.02	2.02 ± 0.10
		I.M.(b)	1.00 ± 0.00	282.89 ± 4.35	209.47 ± 7.53	3.26 ± 0.05	2.41 ± 0.08
	10	G.L.F.(a)	1.00 ± 0.00	1026.88 ± 6.25	877.98 ± 13.29	5.50 ± 0.08	4.70 ± 0.08
		G.L.F.(b)	1.00 ± 0.00	1010.32 ± 7.44	840.99 ± 18.85	13.45 ± 0.24	11.20 ± 0.33
		I.M.(a)	1.00 ± 0.00	1023.10 ± 9.64	857.23 ± 17.40	5.89 ± 0.05	4.94 ± 0.09
		I.M.(b)	1.00 ± 0.00	1026.88 ± 11.64	865.57 ± 27.24	6.75 ± 0.10	5.70 ± 0.21
	50	G.L.F.(a)	0.99 ± 0.00	14285.39 ± 94.94	13034.74 ± 212.10	17.93 ± 0.16	16.36 ± 0.28
		G.L.F.(b)	1.00 ± 0.00	14327.81 ± 86.32	13322.33 ± 153.96	49.54 ± 0.37	46.07 ± 0.61
		I.M.(a)	1.00 ± 0.00	14465.89 ± 123.72	13374.24 ± 126.08	19.04 ± 0.22	17.60 ± 0.21
		I.M.(b)	1.00 ± 0.00	14527.88 ± 110.84	13225.19 ± 143.07	22.29 ± 0.16	20.29 ± 0.18
1.0	5	G.L.F.(a)	0.09 ± 0.00	348.13 ± 17.73	221.83 ± 20.56	1.35 ± 0.08	0.86 ± 0.08
		G.L.F.(b)	0.09 ± 0.00	302.71 ± 17.13	194.10 ± 18.07	4.44 ± 0.25	2.84 ± 0.25
		I.M.(a)	0.76 ± 0.00	8260.70 ± 56.57	6971.22 ± 181.18	21.65 ± 0.37	18.28 ± 0.62
		I.M.(b)	0.76 ± 0.00	8376.43 ± 83.15	6986.00 ± 196.62	22.27 ± 0.33	18.59 ± 0.61
	10	G.L.F.(a)	0.06 ± 0.00	201.29 ± 13.39	121.49 ± 12.69	0.69 ± 0.04	0.42 ± 0.04
		G.L.F.(b)	0.06 ± 0.00	212.65 ± 19.55	107.86 ± 17.94	2.68 ± 0.24	1.37 ± 0.23
		I.M.(a)	0.72 ± 0.00	5371.69 ± 83.35	4242.38 ± 174.21	6.51 ± 0.12	5.15 ± 0.23
		I.M.(b)	0.71 ± 0.00	5464.59 ± 73.10	4472.08 ± 119.62	6.82 ± 0.10	5.58 ± 0.16
	50	G.L.F.(a)	0.02 ± 0.00	54.25 ± 4.75	16.96 ± 3.01	0.15 ± 0.01	0.05 ± 0.01
		G.L.F.(b)	0.02 ± 0.00	63.55 ± 6.00	21.88 ± 5.09	0.65 ± 0.06	0.22 ± 0.05
		I.M.(a)	0.64 ± 0.00	3384.35 ± 162.98	2354.23 ± 228.25	0.57 ± 0.03	0.40 ± 0.04
		I.M.(b)	0.63 ± 0.00	3381.84 ± 58.00	2217.38 ± 190.97	0.59 ± 0.01	0.39 ± 0.03

Table 14. Comparison of the implicit midpoint and generalized leapfrog integrators on sampling from the Bayesian logistic regression posterior on the heart disease dataset when using a randomized number of steps.