SUPPLEMENTARY MATERIALS

A Effective sample size

A common measure of the efficiency of a sampling algorithm is the *effective sample size* (ESS), representing the estimated equivalent number of iid samples drawn. For (unweighted) Markov chains, motivated by Markov chain central limit theory, it is typical to consider ESS in each dimension defined via

$$\mathrm{ESS}_{MC} := \frac{n}{1 + 2\sum_{i=1}^{T} \hat{\rho}_i}, \tag{17}$$

where $\hat{\rho}_i$ is an estimate of the lag-i autocorrelations of the Markov chain, and T is the stopping time given in Section 11.5 of Gelman et al. (2004). Following Girolami & Calderhead (2011), we account for the worst case performance in the chain by reporting the minimum of (17) taken over all dimensions in the chain.

There is some ambiguity when comparing notions of effective sample size for Markov chains that have been reweighted via importance sampling. For n samples reweighted by importance sampling, Kish's approximation to the ESS, given by

$$\mathrm{ESS}_K := \left(\sum_{j=1}^{n} \bar{w}_j^2\right)^{-1}, \tag{18}$$

where $\bar{w}_j = w_j / \sum_{k=1}^{n} w_k$, is most common. Roughly speaking, this accounts for the shift in sampling efficiency caused by imbalances in the importance sampling weights. One may verify for equally weighted samples ($\bar{w}_j = 1/n$ for all $j = 1, \dots, n$) that $\mathrm{ESS}_K = n$. Motivated by this approximation, in order to account for both the effects of both sample autocorrelation and reweighting via importance sampling, we approximated the effective sample size under importance sampling by

$$\mathrm{ESS} := \frac{\mathrm{ESS}_K}{n}\mathrm{ESS}_{MC} = \frac{\left(\sum_{j=1}^{n} \bar{w}_j^2\right)^{-1}}{1 + 2\sum_{i=1}^{T} \hat{\rho}_i}, \tag{19}$$

which also reduces to (17) when all the samples are weighted by unity. Once again, we report the minimum of (19) taken across all dimensions in the chain to account for worst case performance.

B Proofs

In this section, we provide proofs of Theorem 1 and Theorem 2.

Proof of Theorem 1. As outlined in the main text, the strategy is as follows. First, we consider a third–order Taylor expansion of the symplectic transformation implicitly defined by varying the step–size in the generalized leapfrog integrator. Since the integrator is reversible, Theorem 2.2 in Chapter IX of Hairer et al. (2006) guarantees that the fourth-order term vanishes. Theorem 2 then guarantees that this vector field corresponds to Hamilton's equations of a shadow Hamiltonian, which can be computed by integration. Throughout the proof, we will adopt the convention of summation over repeated indices.

Letting $(\theta(t), p(t))$ be the analytic solution to (3) with initial values (θ_0, p_0) and supposing that $(q(t), \hat{\theta}(t), \hat{p}(t))$ is a solution to the implicitly defined equations

$$\begin{cases} q(t) &= p_0 - \frac{t}{2}\nabla_\theta H\big(\theta_0, q(t)\big), \\ \hat{\theta}(t) &= \theta_0 + \frac{t}{2}\Big(\nabla_p H\big(\theta_0, q(t)\big) + \nabla_p H\big(\hat{\theta}(t), q(t)\big)\Big), \\ \hat{p}(t) &= q(t) - \frac{t}{2}\nabla_\theta H\big(\hat{\theta}(t), q(t)\big), \end{cases} \tag{20}$$

observe that (20) evaluated at $t = h$ retrieves (11). Differentiating each coordinate of (20) with respect to t,

$$
\begin{aligned}
q_i'(t) &= -\frac{1}{2}\frac{\partial}{\partial \theta_i} H\big(\theta_0, q(t)\big) + \mathcal{O}(t) \\
\hat{\theta}_i'(t) &= \frac{1}{2}\left(\frac{\partial}{\partial p_i} H\big(\theta_0, q(t)\big) + \frac{\partial}{\partial p_i} H\big(\hat{\theta}(t), q(t)\big)\right) + \mathcal{O}(t) \\
\hat{p}_i'(t) &= q_i'(t) - \frac{1}{2}\frac{\partial}{\partial \theta_i} H\big(\hat{\theta}(t), q(t)\big) + \mathcal{O}(t),
\end{aligned}
$$

which, evaluated at $t = 0$, becomes

$$
q_i'(0) = -\frac{1}{2}\frac{\partial}{\partial \theta_i} H\big(\theta_0, p_0\big), \qquad \hat{\theta}_i'(0) = \frac{\partial}{\partial p_i} H\big(\theta_0, p_0\big), \qquad \hat{p}_i'(0) = -\frac{\partial}{\partial \theta_i} H\big(\theta_0, p_0\big).
$$

Similarly, differentiating a second time,

$$
\begin{aligned}
q_i''(t) &= -\frac{\partial^2}{\partial p_j \partial \theta_i} H\big(\theta_0, q(t)\big) q_j'(t) + \mathcal{O}(t) \\
\hat{\theta}''(t) &= \frac{\partial^2}{\partial p_j \partial p_i}\Big(H\big(\theta_0, q(t)\big) + H\big(\hat{\theta}(t), q(t)\big)\Big) q_j'(t) + \frac{\partial^2}{\partial \theta_j \partial p_i} H\big(\hat{\theta}(t), q(t)\big)\hat{\theta}_j'(t) + \mathcal{O}(t) \\
\hat{p}_i''(t) &= q_i''(t) - \frac{\partial^2}{\partial p_j \partial \theta_i} H\big(\hat{\theta}(t), q(t)\big) q_j'(t) - \frac{\partial^2}{\partial \theta_j \partial \theta_i} H\big(\hat{\theta}(t), q(t)\big)\hat{\theta}_j'(t) + \mathcal{O}(t),
\end{aligned}
$$

which evaluated at $t = 0$ is just

$$
\begin{aligned}
q_i''(0) &= \frac{1}{2}\frac{\partial^2}{\partial p_j \partial \theta_i} H\big(\theta_0, p_0\big)\frac{\partial}{\partial \theta_j} H\big(\theta_0, p_0\big) \\
\hat{\theta}_i''(0) &= -\frac{\partial^2}{\partial p_j \partial p_i} H\big(\theta_0, p_0\big)\frac{\partial}{\partial \theta_j} H\big(\theta_0, p_0\big) + \frac{\partial^2}{\partial \theta_j \partial p_i} H\big(\theta_0, p_0\big)\frac{\partial}{\partial p_j} H\big(\theta_0, p_0\big) \\
\hat{p}_i''(0) &= \frac{\partial^2}{\partial p_j \partial \theta_i} H\big(\theta_0, p_0\big)\frac{\partial}{\partial \theta_j} H\big(\theta_0, p_0\big) - \frac{\partial^2}{\partial \theta_j \partial \theta_i} H\big(\theta_0, p_0\big)\frac{\partial}{\partial p_j} H\big(\theta_0, p_0\big).
\end{aligned}
$$

Moving on to the third order terms,

$$
\begin{aligned}
q_i'''(t) &= -\frac{3}{2}\frac{\partial^2}{\partial p_j \partial \theta_i} H\big(\theta_0, q(t)\big) q_j''(t) - \frac{3}{2}\frac{\partial^3}{\partial p_k \partial p_j \partial \theta_i} H\big(\theta_0, q(t)\big) q_j'(t) q_k'(t) + \mathcal{O}(t) \\
\hat{\theta}_i'''(t) &= \frac{3}{2}\frac{\partial^2}{\partial p_j \partial p_i}\Big(H\big(\theta_0, q(t)\big) + H\big(\hat{\theta}(t), q(t)\big)\Big) q_j''(t) + \frac{3}{2}\frac{\partial^2}{\partial \theta_j \partial p_i} H\big(\hat{\theta}(t), q(t)\big)\hat{\theta}_j''(t) \\
&\quad + \frac{3}{2}\frac{\partial^3}{\partial p_k \partial \theta_j \partial p_i} H\big(\hat{\theta}(t), q(t)\big)\hat{\theta}_j'(t) q_k'(t) + \frac{3}{2}\frac{\partial^3}{\partial \theta_k \partial \theta_j \partial p_i} H\big(\hat{\theta}(t), q(t)\big)\hat{\theta}_j'(t)\hat{\theta}_k'(t) \\
&\quad + \frac{3}{2}\frac{\partial^3}{\partial \theta_k \partial p_j \partial p_i} H\big(\hat{\theta}(t), q(t)\big) q_j'(t)\hat{\theta}_k'(t) + \frac{3}{2}\frac{\partial^3}{\partial p_k \partial p_j \partial p_i} H\big(\theta_0, q(t)\big) q_j'(t) q_k'(t) \\
&\quad + \frac{3}{2}\frac{\partial^3}{\partial p_k \partial p_j \partial p_i} H\big(\hat{\theta}(t), q(t)\big) q_j'(t) q_k'(t) + \mathcal{O}(t) \\
\hat{p}_i'''(t) &= q_i'''(t) - \frac{3}{2}\frac{\partial^2}{\partial \theta_j \partial \theta_i} H\big(\theta_0, q(t)\big)\hat{\theta}_j''(t) - \frac{3}{2}\frac{\partial^2}{\partial p_j \partial \theta_i} H\big(\theta_0, q(t)\big) q_j''(t) \\
&\quad - \frac{3}{2}\frac{\partial^3}{\partial \theta_k \partial \theta_j \partial \theta_i} H\big(\theta_0, q(t)\big)\hat{\theta}_j'(t)\hat{\theta}_k'(t) - \frac{3}{2}\frac{\partial^3}{\partial p_k \partial \theta_j \partial \theta_i} H\big(\theta_0, q(t)\big)\hat{\theta}_j'(t) q_k'(t) \\
&\quad - \frac{3}{2}\frac{\partial^3}{\partial \theta_k \partial p_j \partial \theta_i} H\big(\theta_0, q(t)\big) q_j'(t)\hat{\theta}_k'(t) - \frac{3}{2}\frac{\partial^3}{\partial p_k \partial p_j \partial \theta_i} H\big(\theta_0, q(t)\big) q_j'(t) q_k'(t) + \mathcal{O}(t).
\end{aligned}
$$

which implies

$$
q_i'''(0) = -\frac{3}{4}\frac{\partial^2 H}{\partial p_j \partial \theta_i}\frac{\partial^2 H}{\partial p_k \partial \theta_j}\frac{\partial H}{\partial \theta_k}(\theta_0, p_0) - \frac{3}{8}\frac{\partial^3 H}{\partial p_k \partial p_j \partial \theta_i}\frac{\partial H}{\partial \theta_j}\frac{\partial H}{\partial \theta_k}(\theta_0, p_0)
$$

$$\begin{aligned}\hat{\theta}_i'''(0) &= -\frac{3}{2}\frac{\partial^2 H}{\partial\theta_j\partial p_i}\frac{\partial^2 H}{\partial p_k\partial p_j}\frac{\partial H}{\partial\theta_k}(\theta_0,p_0) + \frac{3}{2}\frac{\partial^2 H}{\partial\theta_j\partial p_i}\frac{\partial^2 H}{\partial\theta_k\partial p_j}\frac{\partial H}{\partial p_k}(\theta_0,p_0) + \frac{3}{2}\frac{\partial^2 H}{\partial p_j\partial p_i}\frac{\partial^2 H}{\partial p_k\partial\theta_j}\frac{\partial H}{\partial\theta_k}(\theta_0,p_0)\\ &\quad - \frac{3}{4}\frac{\partial^3 H}{\partial p_k\partial\theta_j\partial p_i}\frac{\partial H}{\partial p_j}\frac{\partial H}{\partial\theta_k}(\theta_0,p_0) + \frac{3}{2}\frac{\partial^3 H}{\partial\theta_k\partial\theta_j\partial p_i}\frac{\partial H}{\partial p_j}\frac{\partial H}{\partial p_k}(\theta_0,p_0) - \frac{3}{4}\frac{\partial^3 H}{\partial\theta_k\partial p_j\partial p_i}\frac{\partial H}{\partial\theta_j}\frac{\partial H}{\partial p_k}(\theta_0,p_0)\\ &\quad + \frac{3}{4}\frac{\partial^3 H}{\partial p_k\partial p_j\partial p_i}\frac{\partial H}{\partial\theta_j}\frac{\partial H}{\partial\theta_k}(\theta_0,p_0)\\ \hat{p}_i'''(0) &= -\frac{3}{2}\frac{\partial^2 H}{\partial p_j\partial\theta_i}\frac{\partial^2 H}{\partial p_k\partial\theta_j}\frac{\partial H}{\partial\theta_k}(\theta_0,p_0) + \frac{3}{2}\frac{\partial^2 H}{\partial\theta_j\partial\theta_i}\frac{\partial^2 H}{\partial p_k\partial p_j}\frac{\partial H}{\partial\theta_k}(\theta_0,p_0) - \frac{3}{2}\frac{\partial^2 H}{\partial\theta_j\partial\theta_i}\frac{\partial^2 H}{\partial\theta_k\partial p_j}\frac{\partial H}{\partial p_k}(\theta_0,p_0)\\ &\quad - \frac{3}{2}\frac{\partial^3 H}{\partial\theta_k\partial\theta_j\partial\theta_i}\frac{\partial H}{\partial p_j}\frac{\partial H}{\partial p_k}(\theta_0,p_0) + \frac{3}{2}\frac{\partial^3 H}{\partial p_k\partial\theta_j\partial\theta_i}\frac{\partial H}{\partial p_j}\frac{\partial H}{\partial\theta_k}(\theta_0,p_0) - \frac{3}{4}\frac{\partial^3 H}{\partial p_k\partial p_j\partial\theta_i}\frac{\partial H}{\partial\theta_j}\frac{\partial H}{\partial\theta_k}(\theta_0,p_0)\end{aligned}$$

On the other hand, differentiating (3),

$$\begin{aligned}\theta''(t) &= \frac{\partial^2}{\partial p_j\partial p_i}H\big(\theta(t),p(t)\big)p_j'(t) + \frac{\partial^2}{\partial\theta_j\partial p_i}H\big(\theta(t),p(t)\big)\theta_j'(t)\\ &= -\frac{\partial^2}{\partial p_j\partial p_i}H\big(\theta(t),p(t)\big)\frac{\partial}{\partial\theta_j}H\big(\theta(t),p(t)\big) + \frac{\partial^2}{\partial\theta_j\partial p_i}H\big(\theta(t),p(t)\big)\frac{\partial}{\partial p_j}H\big(\theta(t),p(t)\big)\end{aligned}$$

$$\begin{aligned}p_i''(t) &= -\frac{\partial^2}{\partial p_j\partial\theta_i}H\big(\theta(t),p(t)\big)p_j'(t) - \frac{\partial^2}{\partial\theta_j\partial\theta_i}H\big(\theta(t),p(t)\big)\theta_j'(t)\\ &= \frac{\partial^2}{\partial p_j\partial\theta_i}H\big(\theta(t),p(t)\big)\frac{\partial}{\partial\theta_j}H\big(\theta(t),p(t)\big) - \frac{\partial^2}{\partial\theta_j\partial\theta_i}H\big(\theta(t),p(t)\big)\frac{\partial}{\partial p_j}H\big(\theta(t),p(t)\big)\end{aligned}$$

and

$$\begin{aligned}\theta_i'''(t) &= -\frac{\partial^2 H}{\partial\theta_j\partial p_i}\frac{\partial^2 H}{\partial p_k\partial p_j}\frac{\partial H}{\partial\theta_k}\big(\theta(t),p(t)\big) + \frac{\partial^2 H}{\partial\theta_j\partial p_i}\frac{\partial^2 H}{\partial\theta_k\partial p_j}\frac{\partial H}{\partial p_k}\big(\theta(t),p(t)\big) + \frac{\partial^2 H}{\partial p_j\partial p_i}\frac{\partial^2 H}{\partial p_k\partial\theta_j}\frac{\partial H}{\partial\theta_k}\big(\theta(t),p(t)\big)\\ &\quad - \frac{\partial^2 H}{\partial p_j\partial p_i}\frac{\partial^2 H}{\partial\theta_k\partial\theta_j}\frac{\partial H}{\partial p_k}\big(\theta(t),p(t)\big) - 2\frac{\partial^3 H}{\partial p_k\partial\theta_j\partial p_i}\frac{\partial H}{\partial p_j}\frac{\partial H}{\partial\theta_k}\big(\theta(t),p(t)\big) + \frac{\partial^3 H}{\partial\theta_k\partial\theta_j\partial p_i}\frac{\partial H}{\partial p_j}\frac{\partial H}{\partial p_k}\big(\theta(t),p(t)\big)\\ &\quad + \frac{\partial^3 H}{\partial p_k\partial p_j\partial p_i}\frac{\partial H}{\partial\theta_j}\frac{\partial H}{\partial\theta_k}\big(\theta(t),p(t)\big)\\ p_i'''(t) &= -\frac{\partial^2 H}{\partial p_j\partial\theta_i}\frac{\partial^2 H}{\partial p_k\partial\theta_j}\frac{\partial H}{\partial\theta_k}\big(\theta(t),p(t)\big) + \frac{\partial^2 H}{\partial\theta_j\partial\theta_i}\frac{\partial^2 H}{\partial p_k\partial p_j}\frac{\partial H}{\partial\theta_k}\big(\theta(t),p(t)\big) - \frac{\partial^2 H}{\partial\theta_j\partial\theta_i}\frac{\partial^2 H}{\partial\theta_k\partial p_j}\frac{\partial H}{\partial p_k}\big(\theta(t),p(t)\big)\\ &\quad + \frac{\partial^2 H}{\partial p_j\partial\theta_i}\frac{\partial^2 H}{\partial\theta_k\partial\theta_j}\frac{\partial H}{\partial p_k}\big(\theta(t),p(t)\big) - \frac{\partial^3 H}{\partial\theta_k\partial\theta_j\partial\theta_i}\frac{\partial H}{\partial p_j}\frac{\partial H}{\partial p_k}\big(\theta(t),p(t)\big) + 2\frac{\partial^3 H}{\partial p_k\partial\theta_j\partial\theta_i}\frac{\partial H}{\partial p_j}\frac{\partial H}{\partial\theta_k}\big(\theta(t),p(t)\big)\\ &\quad - \frac{\partial^3 H}{\partial p_k\partial p_j\partial\theta_i}\frac{\partial H}{\partial\theta_j}\frac{\partial H}{\partial\theta_k}\big(\theta(t),p(t)\big).\end{aligned}$$

Invoking Taylor's theorem, as $t \downarrow 0$,

$$\begin{aligned}\theta(t) &= \theta_0 + t\theta'(0) + \frac{t^2}{2}\theta''(0) + \frac{t^3}{6}\theta'''(0) + \mathcal{O}(t^4), & p(t) &= p_0 + tp'(0) + \frac{t^2}{2}p''(0) + \frac{t^3}{6}p'''(0) + \mathcal{O}(t^4),\\ \hat{\theta}(t) &= \theta_0 + t\hat{\theta}'(0) + \frac{t^2}{2}\hat{\theta}''(0) + \frac{t^3}{6}\hat{\theta}'''(0) + \mathcal{O}(t^4), & \hat{p}(t) &= p_0 + t\hat{p}'(t) + \frac{t^2}{2}\hat{p}''(0) + \frac{t^3}{6}\hat{p}'''(0) + \mathcal{O}(t^4).\end{aligned}$$

Collecting terms, we see that

$$\begin{aligned}\big(\hat{\theta}_i(t) - \theta_i(t)\big) = \frac{t^3}{6}\Big(&-\frac{1}{2}\frac{\partial^2 H}{\partial\theta_j\partial p_i}\frac{\partial^2 H}{\partial p_k\partial p_j}\frac{\partial H}{\partial\theta_k} + \frac{1}{2}\frac{\partial^2 H}{\partial\theta_j\partial p_i}\frac{\partial^2 H}{\partial\theta_k\partial p_j}\frac{\partial H}{\partial p_k} + \frac{1}{2}\frac{\partial^2 H}{\partial p_j\partial p_i}\frac{\partial^2 H}{\partial p_k\partial\theta_j}\frac{\partial H}{\partial\theta_k}\\ &+ \frac{1}{4}\frac{\partial^3 H}{\partial p_k\partial\theta_j\partial p_i}\frac{\partial H}{\partial p_j}\frac{\partial H}{\partial\theta_k} + \frac{1}{2}\frac{\partial^3 H}{\partial\theta_k\partial\theta_j\partial p_i}\frac{\partial H}{\partial p_j}\frac{\partial H}{\partial p_k} + \frac{1}{4}\frac{\partial^3 H}{\partial\theta_k\partial p_j\partial p_i}\frac{\partial H}{\partial\theta_j}\frac{\partial H}{\partial p_k}\\ &- \frac{1}{4}\frac{\partial^3 H}{\partial p_k\partial p_j\partial p_i}\frac{\partial H}{\partial\theta_j}\frac{\partial H}{\partial\theta_k} + \frac{\partial^2 H}{\partial p_j\partial p_i}\frac{\partial^2 H}{\partial\theta_k\partial\theta_j}\frac{\partial H}{\partial p_k}\Big) + \mathcal{O}(t^4),\end{aligned}$$

which can be written

$$\begin{aligned}\big(\hat{\theta}_i(t) - \theta_i(t)\big) &= \frac{\partial}{\partial p_i}\frac{t^3}{6}\Big(\frac{1}{2}\frac{\partial^2 H}{\partial\theta_k\partial\theta_j}\frac{\partial H}{\partial p_j}\frac{\partial H}{\partial p_k} - \frac{1}{4}\frac{\partial^2 H}{\partial p_k\partial p_j}\frac{\partial H}{\partial\theta_j}\frac{\partial H}{\partial\theta_k} + \frac{1}{2}\frac{\partial^2 H}{\partial\theta_k\partial p_j}\frac{\partial H}{\partial\theta_j}\frac{\partial H}{\partial p_k}\Big) + \mathcal{O}(t^4)\\ &= \frac{\partial}{\partial p_i}\frac{t^3}{6}\Big(\frac{1}{2}\nabla_p H\nabla_{\theta\theta}H\nabla_p H - \frac{1}{4}\nabla_\theta H\nabla_{pp}H\nabla_\theta H + \frac{1}{2}\nabla_\theta H\nabla_{\theta p}H\nabla_p H\Big) + \mathcal{O}(t^4).\end{aligned}$$

Analogous calculations give us

$$\begin{aligned}\big(\hat{p}_i(t) - p_i(t)\big) &= -\frac{\partial}{\partial\theta_i}\frac{t^3}{6}\Big(\frac{1}{2}\frac{\partial^2 H}{\partial\theta_k\partial\theta_j}\frac{\partial H}{\partial p_j}\frac{\partial H}{\partial p_k} - \frac{1}{4}\frac{\partial^2 H}{\partial p_k\partial p_j}\frac{\partial H}{\partial\theta_j}\frac{\partial H}{\partial\theta_k} + \frac{1}{2}\frac{\partial^2 H}{\partial\theta_k\partial p_j}\frac{\partial H}{\partial\theta_j}\frac{\partial H}{\partial p_k}\Big) + \mathcal{O}(t^4)\\ &= -\frac{\partial}{\partial\theta_i}\frac{t^3}{6}\Big(\frac{1}{2}\nabla_p H\nabla_{\theta\theta}H\nabla_p H - \frac{1}{4}\nabla_\theta H\nabla_{pp}H\nabla_\theta H + \frac{1}{2}\nabla_\theta H\nabla_{\theta p}H\nabla_p H\Big) + \mathcal{O}(t^4).\end{aligned}$$

Taken together,

$$\big(\hat{z}(t) - z(t)\big) = J\nabla\frac{t^3}{6}\Big(\frac{1}{2}\nabla_p H\nabla_{\theta\theta}H\nabla_p H - \frac{1}{4}\nabla_\theta H\nabla_{pp}H\nabla_\theta H + \frac{1}{2}\nabla_\theta H\nabla_{\theta p}H\nabla_p H\Big) + \mathcal{O}(t^4).$$

Fixing $t = h$, and keeping in mind that the generalized leapfrog integrator is reversible, Theorems 1.2 and 2.2 from Chapter IX in Hairer et al. (2006) imply that

$$\begin{aligned}\hat{z}'(t) &= z'(t) + \big(\hat{z}'(t) - z'(t)\big)\\ &= J\nabla H + J\nabla\frac{h^2}{6}\Big(\frac{1}{2}\nabla_p H\nabla_{\theta\theta}H\nabla_p H - \frac{1}{4}\nabla_\theta H\nabla_{pp}H\nabla_\theta H + \frac{1}{2}\nabla_\theta H\nabla_{\theta p}H\nabla_p H\Big) + \mathcal{O}(h^4)\\ &= J\nabla\mathcal{H}^{[4]} + \mathcal{O}(h^4),\end{aligned}$$

as required. □

Proof of Theorem 2. Assume M is a smooth, simply connected Riemannian manifold with metric g, and that ϕ_H^t is a symplectic integrator on T^*M, smooth in its step size t. Denote by X the corresponding (and typically time-inhomogeneous) symplectic vector field on T^*M. Note that this vector field can be locally computed by asymptotic expansion after differentiating ϕ_H^t in t. Since X is symplectic, it corresponds to a closed differential 1-form via the canonical isomorphism induced by the symplectic form.

In order to show existence of a shadow Hamiltonian, we will now demonstrate that this closed form must be exact. To do so, we first argue that the first de Rham cohomology group on the vector bundle T^*M must be trivial (see Barp et al. (2018) for a statistical introduction). For $(\theta, p) \in T^*M$ define a deformation retraction onto the zero section of T^*M via $(\theta, p) \mapsto (\theta, (1-t)p)$ for $t \in [0, 1]$. The homotopy invariance of the de Rham cohomology now guarantees that $H^1_{dR}(T^*M) \simeq H^1_{dR}(M)$.

Next, invoke de Rham's Theorem (5.36 in Warner (1983)), which tells us that the first de Rham cohomology group is isomorphic to the first *singular cohomology* group with real coefficients (see Warner (1983) or Hatcher (2000)). That is, $H^1_{dR}(M) \simeq H^1(M, \mathbb{R})$. From here, the *universal coefficient theorem for cohomology* (see Theorem 3.2 in Hatcher (2000)) suggests that $H^1(M, \mathbb{R}) \simeq \mathrm{Hom}\big(H_1(M)\big)$, where $H_1(M)$ is the first homology group of M. The Hurewicz theorem (see Theorem 4.1 in Whitehead (1978)) now implies that the Hurewicz map $\rho : \pi_1(M) \to H_1(M)$ is a homomorphism, where $\pi_1(M)$ is the first fundamental group of M. Moreover this homomorphism is an isomorphism whenever $\pi_1(M)$ is abelian.

Keeping in mind that M is simply connected and so has trivial first fundamental group, we can conclude that the first de Rham cohomology group of T^*M is also trivial. This can be seen in the chain of isomorphisms

$$H^1_{dR}(T^*M) \simeq H^1_{dR}(M) \simeq H^1(M, \mathbb{R}) \simeq \mathrm{Hom}\big(H_1(M)\big) \simeq \mathrm{Hom}\big(\pi_1(M)\big) \simeq 0.$$

This shows that the symplectic vector field X is not only closed, but exact. That is, it is a Hamiltonian vector field corresponding to some Hamiltonian $\mathcal{H}^\phi$.

Now since the symplectic integrator is reversible, its maximal error must be of even order, say n, for some $n \in 2\mathbb{N}$ (see Chapter V of Hairer et al. (2006)). Symplecticity guarantees the Hamiltonian H is preserved up to this order.

By taking the $n+1$-th order Taylor approximation of X, in the same vein as Theorem 1, one can integrate this vector field approximation to compute the $(n+2)$-th order shadow Hamiltonian, $\mathcal{H}^{n+2}$.

Assume now that we have such a shadow Hamiltonian of order k, for some even integer $k > n+2$. That is, the symplectic integrator φ_H^t's approximation of the Hamiltonian trajectories corresponding to H, conserves the shadow Hamiltonian $\mathcal{H}^{2k}$ and tracks its Hamiltonian trajectories up to order $2k$. We know that the integrator ϕ_H^t is reversible, so its error must be even. By computing the $2k+1$-th order Taylor approximation of X, we can integrate to find $\mathcal{H}^{2k+1}$. The inductive hypothesis implies the theorem. □

C Additional Numerical Experiments

C.1 Momentum Refreshment

To investigate the effects of varying the momentum refreshment parameter ρ, we ran ten independent chains of RMHMC and SMHMC on the 30-dimensional funnel for each $\rho \in \{0.0, 0.1, \ldots, 0.8\}$. Each chain is comprised of 1000 samples with 100 burn-in steps, but otherwise used the same parameters as those in Section 4.3. The results are displayed in Figure 4. RMHMC shows a general trend towards improved acceptance and performance as ρ increases. The same trend is shared by SMHMC, but to a lesser extent, possibly due to its universally high acceptance probabilities.

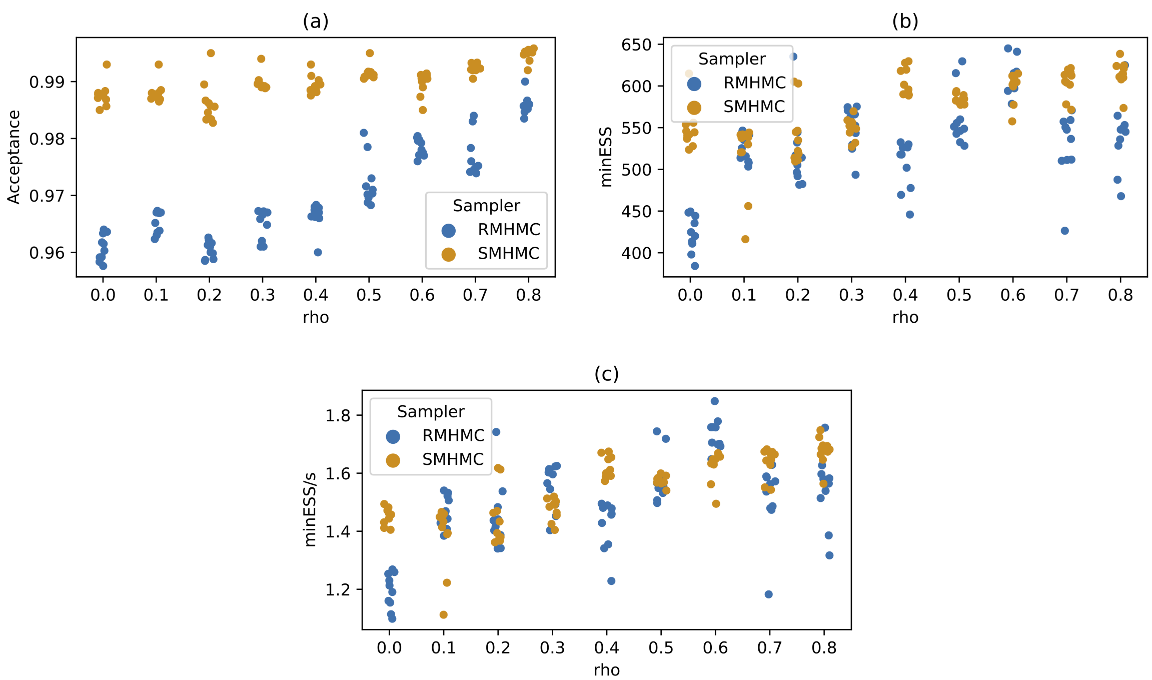

Figure 4: (a) Acceptance probabilities; (b) minESS; and (c) minESS/s for 10 chains of RMHMC and SMHMC on the 30-dimensional funnel with varying choices of ρ.

C.2 Varying Tail Behaviour

In a similar vein, we also investigate the effects of the tail parameter c in (15). We ran ten independent chains of SMHMC on the Bayesian logistic regression problem with the German dataset. Each chain is comprised of 1000 samples with 100 burn-in steps — otherwise, we use the same parameters seen in Section 4.2. The results are displayed in Figure 5. We recall that increasing c interpolates between RMHMC and SMHMC, and this can be readily seen in the rising acceptance probabilities. Such monotonicity is not shared by the minESS. Indeed, worst-case performance appears when $c = -0.5$, when roughly half of the modified Hamiltonians $\mathcal{H}(\theta, p)$ coincide with the classical Hamiltonians $H(\theta, p)$.

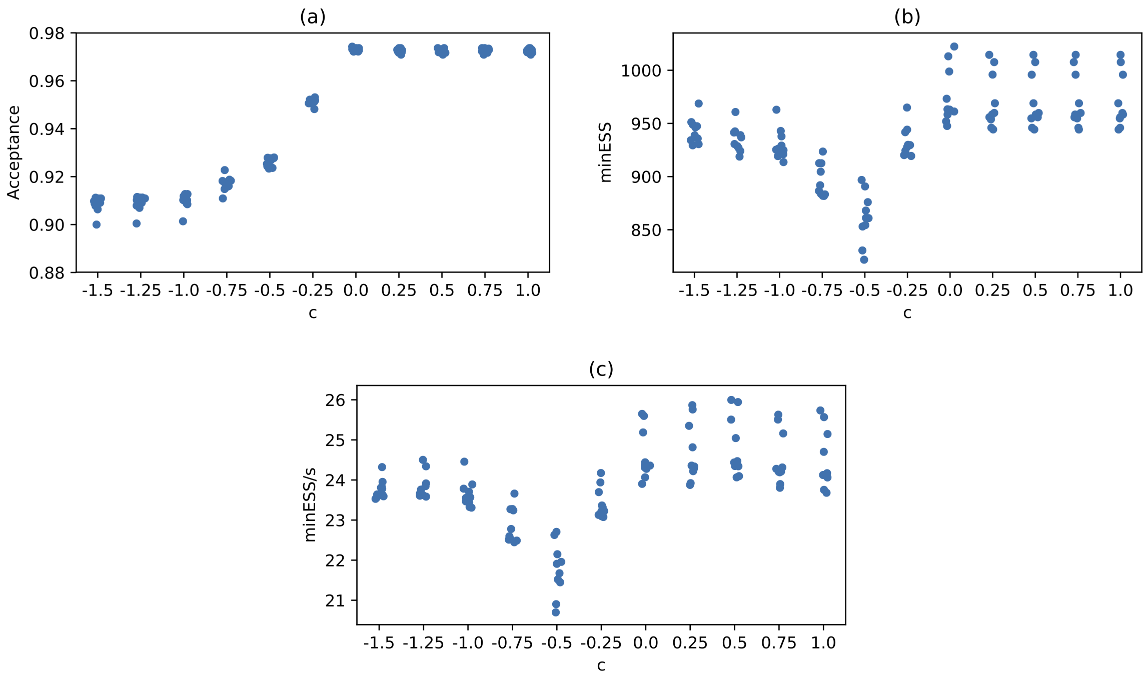

Figure 5: (a) Acceptance probabilities; (b) minESS; and (c) minESS/s for 10 chains of SMHMC on the German dataset with varying choices of c.