OPTIMAL TUNING OF THE HYBRID MONTE-CARLO ALGORITHM

A. BESKOS, N.S. PILLAI, G.O.ROBERTS, J. M. SANZ-SERNA, AND A.M.STUART

Abstract. We investigate the properties of the Hybrid Monte-Carlo algorithm (HMC) in high dimensions. HMC develops a Markov chain reversible w.r.t. a given target distribution Π by using separable Hamiltonian dynamics with potential $-\log \Pi$. The additional momentum variables are chosen at random from the Boltzmann distribution and the continuous-time Hamiltonian dynamics are then discretised using the leapfrog scheme. The induced bias is removed via a Metropolis-Hastings accept/reject rule. In the simplified scenario of independent, identically distributed components, we prove that, to obtain an $\mathcal{O}(1)$ acceptance probability as the dimension d of the state space tends to ∞, the leapfrog step-size h should be scaled as $h = l \times d^{-1/4}$. Therefore, in high dimensions, HMC requires $\mathcal{O}(d^{1/4})$ steps to traverse the state space. We also identify analytically the asymptotically optimal acceptance probability, which turns out to be 0.651 (to three decimal places). This is the choice which optimally balances the cost of generating a proposal, which *decreases* as l increases, against the cost related to the average number of proposals required to obtain acceptance, which *increases* as l increases.

1. Introduction

The Hybrid Monte Carlo (HMC) algorithm originates from the physics literature [8] where it was introduced as a fast method for simulating molecular dynamics. It has since become popular in a number of application areas including statistical physics [10, 11, 28, 16, 1], computational chemistry [15, 19, 27, 30], data assimilation [2], geophysics [19] and neural networks [21, 31]. The algorithm has also been proposed as a generic tool for Bayesian statistical inference [20, 6, 9].

HMC has been proposed as a method to improve on traditional Markov Chain Monte Carlo (MCMC) algorithms. There are heuristic arguments to suggest why HMC might perform better, for example based on the idea that it breaks down *random walk-like* behaviour intrinsic to many MCMC algorithms such as Random-Walk Metropolis (RWM) algorithm. However there is very little theoretical understanding of this phenomenon (though see [7]). This lack of theoretical guidance of choosing the free parameters for the algorithm partly accounts for its relative obscurity in statistical applications. The aim of this paper is to provide insight into the behavior of HMC in high dimensions and develop theoretical tools for improving the efficiency of the algorithm.

HMC uses the derivative of the target probability log-density to guide the Monte-Carlo trajectory towards areas of high probability. The standard RWM algorithm [18] proposes *local*, symmetric moves around the current position. In many cases (especially in high dimensions) the variance of the proposal must be small for the corresponding acceptance probability to be satisfactory. However smaller proposal variance leads to higher autocorrelations, and large computing time to explore the state space. In contrast, and as discussed in the following sections, HMC exploits the information on the derivative of the log density to deliver guided, *global* moves, with higher acceptance probability.

HMC is closely related to the so-called Metropolis-adjusted Langevin algorithm (MALA) [25] which uses the derivative of the log-density to propose steepest-ascent moves in the state space. MALA employs *Langevin* dynamics; the proposal is derived from an Euler discretisation of a Langevin stochastic differential equation that leaves the target density invariant. On the other hand, HMC uses *Hamiltonian* dynamics. The original variable q is seen as a 'location' variable and an auxiliary 'momentum' variable p is introduced; Hamilton's ordinary differential equations are used to generate moves in the enlarged (q, p) phase space. These moves preserve the total energy, a fact that implies, in probability terms, that they preserve the target density Π of the original q variable, provided that the initial momentum is chosen randomly from an appropriate Gaussian distribution. Although seemingly of different origin, MALA can be thought of as a 'localised' version of HMC: we will return to this point in the main text.

In practice, continuous-time Hamiltonian dynamics are discretised by means of a numerical scheme; the popular *Störmer-Verlet* or *leapfrog* scheme [12, 17, 26, 29] is currently the scheme of choice. This integrator does not conserve energy exactly and the induced bias is corrected via a Metropolis-Hastings accept/reject rule. In this way, HMC develops a Markov chain reversible w.r.t. Π, whose transitions incorporate information on Π in a natural way.

In this paper we will investigate the properties of HMC in high dimensions and, in such a context, offer some guidance over the *optimal* specification of the free parameters of the algorithm. We assume that we wish to sample from a density Π on $\mathbb{R}^N$ with

$$\Pi(Q) = \exp\big(-\mathcal{V}(Q)\big) \ , \tag{1.1}$$

for $\mathcal{V} : \mathbb{R}^N \to \mathbb{R}$. We study the simplified scenario where $\Pi(Q)$ consists of $d \gg 1$ independent identically distributed (iid) vector components,

$$\Pi(Q) = \exp\big(-\sum_{i=1}^{d} V(q_i)\big) \ , \quad V : \mathbb{R}^m \to \mathbb{R} \ ; \quad N = m \times d \ . \tag{1.2}$$

For the leapfrog integrator, we show analytically that, under suitable hypotheses on V and as $d \to \infty$, HMC requires $\mathcal{O}(d^{1/4})$ steps to traverse the state space, and furthermore, identify the associated optimal acceptance probability.

To be more precise, if h is the step-size employed in the leapfrog integrator, then we show that the choice

$$\text{HMC}: \quad h = l \cdot d^{-1/4} \tag{1.3}$$

leads to an average acceptance probability which is of $\mathcal{O}(1)$ as $d \to \infty$: Theorem 3.6. This implies that $\mathcal{O}(d^{1/4})$ steps are required for HMC to make $\mathcal{O}(1)$ moves in state space. Furthermore we provide a result of perhaps greater practical relevance. We prove that, for the leapfrog integrator and as $d \to \infty$, the asymptotically *optimal* algorithm corresponds to a well-defined value of the acceptance probability, *independent of the particular target* Π in (1.2). This value is (to three decimal places) 0.651: Theorems 4.1 and 4.2. Thus, when applying HMC in high dimensions, one should try to tune the free algorithmic parameters to obtain an acceptance probability close to that value. We give the precise definition of optimality when stating the theorems but, roughly, it is determined by the choice of l which balances the cost of generating a proposal, which *decreases* as l increases, against the cost related to the average number of proposals required to obtain acceptance, which *increases* as l increases.

The scaling $\mathcal{O}(d^{1/4})$ to make $\mathcal{O}(1)$ moves in state space contrasts favorably with the corresponding scalings $\mathcal{O}(d)$ and $\mathcal{O}(d^{1/3})$ required in a similar context by RWM and MALA respectively (see the discussion below). Furthermore, the full analysis provided in this paper for the leapfrog scheme may be easily extended to high-order, volume-preserving, reversible integrators. For such an integrator the corresponding scaling would be $\mathcal{O}(d^{1/(2\nu)})$, where ν (an even integer) represents the order of the method. For the standard HMC algorithm, previous works have already established the relevance of the choice $h = \mathcal{O}(d^{-1/4})$ (by heuristic arguments, see [11]) and an optimal acceptance probability of around 0.7 (by numerical experiments, see [6]). Our analytic study of the scaling issues in HMC was prompted by these two papers.

The paper is organized as follows. Section 2 presents the HMC method and reviews the literature concerning scaling issues for the RWM and MALA algorithms. Section 3 studies the asymptotic behaviour of HMC as the dimensionality grows, $d \to \infty$, including the key Theorem 3.6. The optimal tuning of HMC is discussed in Section 4, including the key Theorems 4.1 and 4.2. Sections 5 and 6 are technical. The first of them contains the derivation of the required numerical analysis estimates on the leapfrog integrator, with careful attention paid to the dependence of constants in error estimates on the initial condition; estimates of this kind are not available in the literature and may be of independent interest. Section 6 gathers the probabilistic proofs. We finish with some conclusions and discussion in Section 7.

2. Hybrid Monte Carlo (HMC)

2.1. **Hamiltonian dynamics.**

Consider the Hamiltonian function:

$$\mathcal{H}(Q,P) = \frac{1}{2}\langle P, \mathcal{M}^{-1}P\rangle + \mathcal{V}(Q) \ ,$$

on $\mathbb{R}^{2N}$, where $\mathcal{M}$ is a symmetric positive definite matrix (the 'mass' matrix). One should think of Q as the *location* argument and $\mathcal{V}(Q)$ as the potential energy of the system; P as the *momenta*, and $(1/2)\langle P, \mathcal{M}^{-1}P\rangle$ as the kinetic energy. Thus $\mathcal{H}(Q,P)$ gives the total *energy*: the sum of the potential and the kinetic energy. The Hamiltonian dynamics associated with $\mathcal{H}$ are governed by

$$\frac{dQ}{dt} = \mathcal{M}^{-1}P, \quad \frac{dP}{dt} = -\nabla\mathcal{V}(Q) \ , \tag{2.1}$$

a system of ordinary differential equations whose solution flow Φ_t defined by

$$(Q(t), P(t)) = \Phi_t(Q(0), P(0))$$

possesses some key properties relevant to HMC:

- **1. Conservation of Energy:** The change in the potential becomes kinetic energy; *i.e.,* $\mathcal{H} \circ \Phi_t = \mathcal{H}$, for all $t > 0$, or $\mathcal{H}(\Phi_t(Q(0), P(0))) = \mathcal{H}(Q(0), P(0))$, for all $t > 0$ and all initial conditions $(Q(0), P(0))$.

- **2. Conservation of Volume:** The volume element $dP\,dQ$ of the phase space is conserved under the mapping Φ_t.

- **3. Time Reversibility:** If $\mathcal{S}$ denotes the symmetry operator:
$$\mathcal{S}(Q,P) = (Q,-P)$$
then $\mathcal{H}\circ\mathcal{S} = \mathcal{H}$ and
$$\mathcal{S}\circ(\Phi_t)^{-1}\circ\mathcal{S} = \Phi_t \ . \tag{2.2}$$
Thus, changing the sign of the initial velocity, evolving backwards in time, and changing the sign of the final velocity reproduces the forward evolution.

From the Liouville equation for equation (2.1) it follows that, if the initial conditions are distributed according a probability measure with Lebesgue density depending only on $\mathcal{H}(Q,P)$, then this probability measure is preserved by the Hamiltonian flow Φ_t. In particular, if the initial conditions $(Q(0),P(0))$ of (2.1) are distributed with a density (proportional to)
$$\exp(-\mathcal{H}(Q,P)) = \exp((1/2)\langle P,\mathcal{M}^{-1}P\rangle)\exp(-\mathcal{V}(Q)),$$
then, for all $t>0$, the marginal density of $Q(t)$ will also be (proportional to) $\exp(-\mathcal{V}(Q))$. This suggests that integration of equations (2.1) might form the basis for an exploration of the target density $\exp(-\mathcal{V}(Q))$.

2.2. **The HMC algorithm.** To formulate a practical algorithm, the continuous-time dynamics (2.1) must be discretised. The most popular *explicit* method is the Störmer-Verlet or leapfrog scheme (see [12, 17, 26] and the references therein) defined as follows. Assume a current state (Q_0,P_0); then, after one step of length $h>0$ the system (2.1) will be at a state (Q_h,P_h) defined by the three-stage procedure:

$$P_{h/2} = P_0 - \tfrac{h}{2}\,\nabla\mathcal{V}(Q_0)\ ; \tag{2.3a}$$

$$Q_h = Q_0 + h\,\mathcal{M}^{-1}P_{h/2}\ ; \tag{2.3b}$$

$$P_h = P_{h/2} - \tfrac{h}{2}\,\nabla\mathcal{V}(Q_h)\ . \tag{2.3c}$$

The scheme gives rise to a map:
$$\Psi_h\colon (Q_0,P_0)\mapsto(Q_h,P_h)$$
which approximates the flow Φ_h. The solution at time T is approximated by taking $\lfloor\frac{T}{h}\rfloor$ leapfrog steps:
$$(Q(T),P(T)) = \Phi_T((Q(0),P(0)) \approx \Psi_h^{\lfloor\frac{T}{h}\rfloor}((Q(0),P(0))\ .$$
Note that this is a *deterministic* computation. The map
$$\Psi_h^{(T)} := \Psi_h^{\lfloor\frac{T}{h}\rfloor}$$
may be shown to be volume preserving and time reversible (see [12, 17, 26]) but it does not exactly conserve energy. As a consequence the leapfrog algorithm does not share the property of equations (2.1) following from the Liouville equation, namely that any probability density function proportional to $\exp\big(-\mathcal{H}(Q,P)\big)$ is preserved. In order to restore this property an accept-reject step must be added. Paper [20] provides a clear derivation of the required acceptance ceriterion.

We can now describe the complete HMC algorithm. Let the current state be Q. The next state for the HMC Markov chain is determined by the dynamics described in Table 1.

HMC(Q):

(i) *Sample a momentum* $P \sim N(0, \mathcal{M})$.

(ii) *Accept the proposed update* Q' *defined via* $(Q', P') = \Psi_h^{(T)}(Q, P)$ *w.p.:*

$$a((Q,P),(Q',P')) := 1 \wedge \exp\{\mathcal{H}(Q,P) - \mathcal{H}(Q',P')\} \ .$$

TABLE 1. The Markov transition for the Hybrid Monte-Carlo algorithm. Iterative application for a given starting location Q^0, will yield a Markov chain $Q^0, Q^1, \ldots$

Due to the time reversibility and volume conservation properties of the integrator map $\Psi_h^{(T)}$, the recipe in Table 1 defines (see [8, 20]) a Markov chain reversible w.r.t $\Pi(Q)$; sampling this chain up to equilibrium will provide correlated samples Q^n from $\Pi(Q)$. We note that the momentum P is merely an auxiliary variable and that the user of the algorithm is free to choose h, T and the mass matrix $\mathcal{M}$. In this paper we concetrate on the optimal choice of h, for high dimensional targets.

2.3. **Connection with other Metropolis-Hastings algorithms.** Earlier research has studied the optimal tuning of other Metropolis-Hastings algorithms, namely the Random-Walk Metropolis (RWM) and the Metropolis-adjusted Langevin algorithm (MALA). In contrast with HMC, whose proposals involve a deterministic element, those algorithms use updates that are purely stochastic. For the target density $\Pi(Q)$ in (1.1), RWM is specified through the proposed update

$$Q' = Q + \sqrt{h}\, Z \ ,$$

with $Z \sim N(0, I)$ (this sample case suffices for our exposition, but note that Z may be allowed to have an arbitrary mean zero distribution), while MALA is determined through the proposal

$$Q' = Q + \frac{h}{2}\,\nabla \log \Pi(Q) + \sqrt{h}\, Z \ .$$

The density Π is invariant for both algorithms when the proposals are accepted with probability

$$a(Q, Q') = 1 \wedge \frac{\Pi(Q')T(Q', Q)}{\Pi(Q)T(Q, Q')} \ ,$$

where

$$T(x, y) = \mathrm{P}\left[\, Q' \in dy \mid Q = x \,\right] / \, dy$$

is the transition density of the proposed update (note that for RWM the symmetry of the proposal implies $T(Q, Q') = T(Q', Q)$).

The proposal distribution for MALA corresponds to the Euler discretization of the stochastic differential equation (SDE)

$$dQ = \frac{1}{2}\nabla \log \Pi(Q)\, dt + dW,$$

for which Π is an invariant density (here W denotes a standard Brownian motion). One can easily check that HMC and MALA are connected because HMC reduces to MALA when $T \equiv h$, *i.e.,* when the algorithm makes only a single leapfrog step at each transition of the chain.

Assume now that RWL and MALA are applied with the scalings

$$\text{RWM}: \quad h = l \cdot d^{-1}, \qquad \text{MALA}: \quad h = l \cdot d^{-1/3}, \tag{2.4}$$

for some constant $l > 0$, in the simplified scenario where the target Π has the iid structure (1.2) with $m = 1$. The papers [22], [23] prove that, as $d \to \infty$ and under regularity conditions on V (the function V must be seven times differentiable[1], with all derivatives having polynomial growth bounds, and all moments of $\exp(-V)$ must be finite), the acceptance probability approaches a nontrivial value:

$$\mathbb{E}\left[\, a(Q, Q')\,\right] \to a(l) \in (0, 1)$$

(the limit $a(l)$ is different for each of the two algorithms). Furthermore, if $q_1^0, q_1^1, \ldots$ denotes the projection of the trajectory $Q^0, Q^1, \ldots$ onto its first coordinate, in the above scenario it is possible to show ([22], [23]) the convergence of the continuous-time interpolation

$$\text{RWM}: \; t \mapsto q_1^{[\, t \cdot d\,]}, \qquad \text{MALA}: \; t \mapsto q_1^{[\, t \cdot d^{1/3}\,]} \tag{2.5}$$

to the diffusion process governed by the SDE

$$dq = -\frac{1}{2}\, l\, a(l)\, V'(q)\, dt + \sqrt{l\, a(l)}\, dw, \tag{2.6}$$

(w represents a standard Brownian motion). In view of (2.4), (2.5) and (2.6) we deduce that the RWM and MALA algorithms cost $\mathcal{O}(d)$ and $\mathcal{O}(d^{1/3})$ respectively to explore the invariant measure in stationarity. Furthermore, as the product $l\, a(l)$ determines the *speed* of the limiting diffusion the state space will be explored faster for the choice l_{opt} of l that maximises $l\, a(l)$. While l_{opt} depends on the target distribution, it turns out that the optimal acceptance probability $a(l_{opt})$ is independent of V. In fact, with three decimal places, one finds:

$$\text{RWM}: \; a(l_{opt}) = 0.234, \qquad \text{MALA}: \; a(l_{opt}) = 0.574\;.$$

Asymptotically as $d \to \infty$, this analysis identifies algorithms that may be regarded as *uniformly* optimal, because, as discussed in [24], ergodic averages of trajectories corresponding to $l = l_{opt}$ provide optimal estimation of expectations $\mathbb{E}\left[\, f(q)\,\right]$, $q \sim \exp(-V)$, irrespectively of the choice of the (regular) function f. These investigations of the optimal tuning of RWL and MALA have been subsequently extended in [3] and [4] to non-product target distributions.

For HMC we show that the scaling (1.3) leads to an average acceptance probability of $\mathcal{O}(1)$ and hence to a cost of $\mathcal{O}(d^{1/4})$ to make the $\mathcal{O}(1)$ moves necessary to explore the invariant measure. However, in constrast to RWM and MALA, we are not able to provide a simple description of the limiting dynamics of a single coordinate of the Markov chain. Consequently optimality is harder to define.

---

[1] although this is a technical requirement which may be relaxed

3. Hybrid Monte Carlo in the limit $d \to \infty$.

The primary aim of this section is to prove Theorem 3.6 concerning the scaling of the step-size h in HMC. We also provide some insight into the limiting behaviour of the resulting Markov chain, under this scaling, in Propositions 3.8 and 3.9.

3.1. **HMC in the iid scenario.** We now study the asymptotic behaviour of the HMC algorithm in the iid scenario (1.2), when the number d of 'particles' goes to infinity. We write $Q = (q_i)_{i=1}^d$ and $P = (p_i)_{i=1}^d$ to distinguish the individual components, and use the following notation for the combination location/momentum:

$$X = (x_i)_{i=1}^d; \quad x_i := (q_i, p_i) \in \mathbb{R}^{2m}.$$

We denote by $\mathcal{P}_q$ and $\mathcal{P}_p$ the projections onto the position and momentum components of x, *i.e.* $\mathcal{P}_q(q,p) = q$, $\mathcal{P}_p(q,p) = p$.

We have:

$$\mathcal{H}(Q,P) = \sum_{i=1}^d H(q_i,p_i); \quad H(q,p) := \frac{1}{2}\langle p, M^{-1}p\rangle + V(q) \ ,$$

where M is a $m \times m$ symmetric, positive definite matrix. The Hamiltonian differential equations for a single (m-dimensional) particle are then

$$\frac{dq}{dt} = M^{-1}p, \quad \frac{dp}{dt} = -\nabla V(q) \ , \tag{3.1}$$

where $V : \mathbb{R}^m \to \mathbb{R}$. We denote the corresponding flow by φ_t and the leapfrog solution operator over one h-step by ψ_h.

Thus the acceptance probability for the evolution of the d particles is given by (see Table 1):

$$a(X,Y) = 1 \wedge \exp\Big(\sum_{i=1}^d \big[H(x_i) - H(\psi_h^{(T)}(x_i))\big]\Big) \tag{3.2}$$

with $Y = (y_i)_{i=1}^d = \Psi_h^{(T)}(X)$ denoting the HMC proposal. Note that the leapfrog scheme (2.3) is applied independently for each of the d particles (q_i, p_i); the different co-ordinates are only connected through the accept/reject decision based on (3.2).

3.2. **Energy increments.** Our first aim is to estimate (in an analytical sense) the exponent in the right-hand side of (3.2). Since the d particles play the same role, it is sufficient to study a single term $H(x_i) - H(\psi_h^{(T)}(x_i))$. We set

$$\Delta(x,h) := H(\psi_h^{(T)}(x)) - H(\varphi_T(x)) = H(\psi_h^{(T)}(x)) - H(x) \ . \tag{3.3}$$

This is the energy change, due to the leapfrog scheme, over $0 \le t \le T$, with step-size h and initial condition x, which by conservation of energy under the true dynamics, is simply the energy error at time T. We will study the first and second moments:

$$\mu(h) := \mathbb{E}\,[\,\Delta(x,h)\,] = \int_{\mathbb{R}^{2m}} \Delta(x,h)\, e^{-H(x)} dx \ ,$$
$$s^2(h) := \mathbb{E}\,[|\Delta(x,h)|^2] \ ,$$

and the corresponding variance

$$\sigma^2(h) = s^2(h) - \mu^2(h) \ .$$

If the integrator were exactly energy-preserving, one would have $\Delta \equiv 0$ and all proposals would be accepted. However it is well known that the size of $\Delta(x,h)$ is in general no better than the size of the integration error $\psi_h^{(T)}(x) - \varphi_T(x)$, *i.e.* $\mathcal{O}(h^2)$. In fact, under natural smoothness assumptions on V the following condition holds (see Section 5 for a proof):

Condition 3.1. *There exist functions $\alpha(x)$, $\rho(x,h)$ such that*

$$\Delta(x,h) = h^2\alpha(x) + h^2\rho(x,h) \tag{3.4}$$

with $\lim_{h\to 0}\rho(x,h) = 0$.

Furthermore in the proofs of the theorems below we shall use an additional condition to control the variation of Δ as a function of x. This condition will be shown in Section 5 to hold under suitable assumptions on the growth of V and its derivatives.

Condition 3.2. *There exists a function $D : \mathbb{R}^{2m} \to \mathbb{R}$ such that*

$$\sup_{0\le h\le 1} \frac{|\Delta(x,h)|^2}{h^4} \le D(x)\ ,$$

with

$$\int_{\mathbb{R}^{2m}} D(x)\, e^{-H(x)} dx < \infty\ .$$

Key to the proof of Theorem 3.6 is the fact that the average energy increment scales as $\mathcal{O}(h^4)$. We show this in Proposition 3.4 using the following simple lemma that holds for general volume preserving, time reversible integrators:

Lemma 3.3. *Let $\psi_h^{(T)}$ be any volume preserving, time reversible numerical integrator of the Hamiltonian equations (3.1) and $\Delta(x,h) : \mathbb{R}^{2m}\times\mathbb{R}_+ \to \mathbb{R}$ be as in (3.3). If $\varphi : \mathbb{R}\to\mathbb{R}$ is an odd function then:*

$$\int_{\mathbb{R}^{2m}} \varphi(\Delta(x,h))\, e^{-H(x)}\, dx = -\int_{\mathbb{R}^{2m}} \varphi(\Delta(x,h))\, e^{-H(\psi_h^{(T)}(x))}\, dx$$

provided at least one of the integrals above exist. If φ is an even function, then:

$$\int_{\mathbb{R}^{2m}} \varphi(\Delta(x,h))\, e^{-H(x)}\, dx = \int_{\mathbb{R}^{2m}} \varphi(\Delta(x,h))\, e^{-H(\psi_h^{(T)}(x))}\, dx\ ,$$

provided at least one of the integrals above exist.

Proof. See Section 6. □

Applying this lemma with $\varphi(u) = u$, we obtain

$$\mu(h) = -\int_{\mathbb{R}^{2m}} \Delta(x,h)\, e^{-H(\psi_h^{(T)}(x))}\, dx\ ,$$

which implies that

$$2\mu(h) = \int_{\mathbb{R}^{2m}} \Delta(x,h)\left[1 - \exp(-\Delta(x,h))\right] e^{-H(x)}\, dx\ . \tag{3.5}$$

We now use first the inequality $|e^u - 1| \le |u|(e^u + 1)$ and then Lemma 3.3 with $\varphi(u) = u^2$ to conclude that

$$|2\,\mu(h)| \le \int_{\mathbb{R}^{2m}} |\Delta(x,h)|^2\, e^{-H(\psi_h^{(T)}(x))} dx + \int_{\mathbb{R}^{2m}} |\Delta(x,h)|^2\, e^{-H(x)} dx$$

$$\le 2\int_{\mathbb{R}^{2m}} |\Delta(x,h)|^2 e^{-H(x)} dx = 2\, s^2(h)\ . \tag{3.6}$$

The bound in (3.6) is important: it shows that the average of $\Delta(x,h)$ is actually of the order of (the average of) $\Delta(x,h)^2$. Since for the second-order leapfrog scheme $\Delta(x,h) = \mathcal{O}(h^2)$, we see from (3.6) that we may expect the average $\mu(h)$ to actually behave as $\mathcal{O}(h^4)$. This is made precise in the following theorem.

Proposition 3.4. *If the potential V is such that Conditions 3.1 and 3.2 hold for the leapfrog integrator $\psi_h^{(T)}$, then*

$$\lim_{h\to 0} \frac{\mu(h)}{h^4} = \mu\ , \quad \lim_{h\to 0} \frac{\sigma^2(h)}{h^4} = \Sigma\ ,$$

for the constants:

$$\Sigma = \int_{R^{2m}} \alpha^2(x)\, e^{-H(x)}\, dx\ ; \quad \mu = \Sigma/2\ .$$

Proof. See Section 6. □

Next, we perform explicit calculations for the harmonic oscillator and verify the conclusions of Proposition 3.4.

Example 3.5 (Harmonic Oscillator)**.** *Consider the Hamiltonian*

$$H(q,p) = \frac{1}{2}p^2 + \frac{1}{2}q^2$$

that gives rise to the system

$$\begin{pmatrix} dq/dt \\ dp/dt \end{pmatrix} = \begin{pmatrix} p \\ -q \end{pmatrix}\ ,$$

with solutions

$$\begin{pmatrix} q(t) \\ p(t) \end{pmatrix} = \begin{pmatrix} \cos(t) & \sin(t) \\ -\sin(t) & \cos(t) \end{pmatrix} \begin{pmatrix} q(0) \\ p(0) \end{pmatrix}\ .$$

In this case, the leapfrog integration can be written as:

$$\psi_h = \psi_h(q,p) = \begin{pmatrix} 1 - h^2/2 & h \\ -h + h^3/4 & 1 - h^2/2 \end{pmatrix} \begin{pmatrix} q \\ p \end{pmatrix} = \Xi \begin{pmatrix} q \\ p \end{pmatrix}\ ,$$

and, accordingly, the numerical solution after $\lfloor \frac{1}{h} \rfloor$ steps is given by:

$$\psi_h^{(1)}(q,p) = \Xi^{\lfloor \frac{1}{h} \rfloor} \begin{pmatrix} q \\ p \end{pmatrix}\ .$$

Diagonalizing Ξ and exponentiating yields:

$$\Xi^n = \begin{pmatrix} \cos(\theta n) & \frac{1}{\sqrt{1-h^2/4}} \sin(\theta n) \\ -\sqrt{1 - h^2/4}\, \sin(\theta n) & \cos(\theta n) \end{pmatrix}$$

where $\theta = \cos^{-1}(1 - h^2/2)$. Using, for instance, MATHEMATICA, one can now obtain the Taylor expansion:

$$\Delta(x,h) = H(\psi_h^{(1)}(x)) - H(x) = h^2\alpha(x) + h^4\beta(x) + \mathcal{O}(h^6)$$

where:

$$\alpha(q,p) = \big((p^2-q^2)\sin^2(1) + pq\sin(2)\big)/8 \ ;$$
$$\beta(q,p) = \Big(-q^2\sin(2) + pq\big(2\cos(2)+3\sin(2)\big) + p^2\big(3-3\cos(2)+\sin(2)\big)\Big)/192 \ .$$

Notice that, in the stationary regime, q, p are standard normal variables. Therefore, the expectation of $\alpha(x)$ is 0. Tedious calculations give:

$$\mathrm{Var}\,[\,\alpha(x)\,] = \frac{1}{16}\sin^2(1)\ , \quad \mathbb{E}\,[\,\beta(x)\,] = \frac{1}{32}\sin^2(1)\ ,$$

in agreement with Proposition 3.4.

3.3. **Expected acceptance probability.** We are now in a position to identify the scaling for h that gives non-trivial acceptance probability as $d \to \infty$.

Theorem 3.6. *Assume that the potential V is such that the leapfrog integrator $\psi_h^{(T)}$ satisfies Conditions 3.1 and 3.2 and that*

$$h = l \cdot d^{-1/4}\ , \tag{3.7}$$

for a constant $l > 0$. Then in stationarity, i.e., for $X \sim \exp(-\mathcal{H})$,

$$\lim_{d\to\infty} \mathbb{E}\,[\,a(X,Y)\,] = 2\,\Phi(-l^2\sqrt{\Sigma}/2) =: a(l)$$

where the constant Σ is as defined in Proposition 3.4.

Proof. To grasp the main idea, note that the acceptance probability (3.2) is given by

$$a(X,Y) = 1 \wedge e^{R_d}\ ; \qquad R_d = -\sum_{i=1}^{d}\Delta(x_i,h)\ . \tag{3.8}$$

Due to the simple structure of the target density and stationarity, the terms $\Delta(x_i,h)$ being added in (3.8) are iid random variables. Since the expectation and standard deviation of $\Delta(x,h)$ are both $\mathcal{O}(h^4)$ and we have d terms, the natural scaling to obtain a distributional limit is given by (3.7). Then $R_d \approx N(-\frac{1}{2}l^4\Sigma, l^4\Sigma)$ and the desired result follows. See Section 6 for a detailed proof. □

In Theorem 3.6 the limit acceptance probability arises from the use of the Central Limit Theorem. If Condition 3.2 is not satisfied and $\sigma^2(h) = \infty$, then a Gaussian limit is not guaranteed and it may be necessary to consider a different scaling to obtain a heavy tailed limiting distribution such as a stable law.

The scaling (3.7) is a direct consequence of the fact that the leapfrog integrator possesses second order accuracy. Arguments similar to those used above prove that the use of a volume-preserving, symmetric ν-th order integrator would result in a scaling $h = \mathcal{O}(d^{-1/(2\nu)})$ (ν is an even integer) to obtain an acceptance probability of $\mathcal{O}(1)$.

3.4. **The displacement of one particle in a transition.** We now turn our attention to the displacement $q_1^{n+1} - q_1^n$ of a single particle in a transition $n \to n+1$ of the chain. Note that clearly

$$q_1^{n+1} = I^n \cdot \mathcal{P}_q\,\psi_h^{(T)}(q_1^n,p_1^n) + \big(1-I^n\big)q_1^n; \quad I^n = \mathbb{I}_{U^n \le a(X^n,Y^n)}\ . \tag{3.9}$$

While Conditions 3.1 and 3.2 above refer to the error in energy, the proof of the next results requires a condition on the leapfrog integration error in the dynamic variables q and p. In Section 5 we describe conditions on V that guarantee the fulfillment of this condition.

Condition 3.7. *There exists a function* $E : \mathbb{R}^{2m} \to \mathbb{R}$ *such that*

$$\sup_{0\le h\le 1} \frac{|\psi_h^{(T)}(x) - \varphi_T(x)|}{h^2} \le E(x) \ ,$$

with

$$\int_{\mathbb{R}^{2m}} E(x)^4 \, e^{-H(x)} dx < \infty \ .$$

Under the scaling (3.7) and at stationarity, the second moment $\mathbb{E}\,[\,(q_1^{n+1} - q_1^n)^2\,]$ will also approach a nontrivial limit:

Proposition 3.8. *Assume that the hypotheses of Theorem 3.6 and Condition 3.7 hold and, furthermore, that the density* $\exp(-V(q))$ *possesses finite fourth moments. Then, in stationarity,*

$$\lim_{d\to\infty} \mathbb{E}\,[\,(q_1^{n+1} - q_1^n)^2\,] = C_J \cdot a(l)$$

where the value of the constant C_J *is given by*

$$C_J = \mathbb{E}\,[\,(\mathcal{P}_q\varphi_T(q,p) - q)^2\,]\ ; \quad (q,p) \sim \exp\bigl(-H(q,p)\bigr) \ .$$

Proof. See Section 6. □

We will use this proposition in Section 4.

3.5. **The limit dynamics.** We now discuss the limiting dynamics of the Markov chain, under the same assumptions made in Proposition 3.8. For HCM (as for RWM or MALA) the marginal process $\{q_1^n\}_{n\ge 0}$ is not Markovian w.r.t. its own filtration since its dynamics depend on the current position of all d particles via the acceptance probability $a(X^n, Y^n)$ (see (3.9)). In the case of MALA and RWM, $\{q_1^n\}_{n\ge 0}$ is *asymptotically* Markovian: as $d \to \infty$ the effect of the rest of the particles gets averaged to a constant via the Strong Law of Large Numbers. This allows for the interpolants of (2.5) to converge to solutions of the SDE (2.6), which defines a Markov process. We will now argue that for HCM $\{q_1^n\}_{n\ge 0}$ cannot be expected to be *asymptotically* Markovian. In order to simplify the exposition we will not present all the technicalities of the argument that follows.

It is well known (see for instance [29]) that, due to time reversibility and under suitable smoothness assumptions on V, the energy increments of the leapfrog integrator may be expanded in even powers of h as follows (cf. (3.4)):

$$\Delta(x,h) = h^2\alpha(x) + h^4\beta(x) + \mathcal{O}(h^6) \ .$$

Here $\mathbb{E}\,[\,\alpha(x)\,] = 0$ because from Proposition 3.4 we know that $\mathbb{E}\,[\,\Delta(x,h)\,] = \mathcal{O}(h^4)$. Ignoring $\mathcal{O}(h^6)$-terms, we can write:

$$a(X^n, Y^n) = 1 \wedge e^{R_{1,d}^n + R_{2,d}^n}$$

with

$$R_{1,d}^n = -h^2 \sum_{i=1}^d \bigl\{\alpha(x_i^n) - \mathbb{E}\,[\,\alpha(x_i^n)\,|\,q_i^n\,]\bigr\} - h^4 \sum_{i=1}^d \beta(x_i^n) \ ,$$

$$R_{2,d}^n = -h^2 \sum_{i=1}^d \mathbb{E}\,[\,\alpha(x_i^n)\,|\,q_i^n\,] \ .$$

Under appropriate conditions, $R^n_{1,d}$ converges, as $d \to \infty$, to a Gaussian limit independent of the σ-algebra $\sigma(q^n_1, q^n_2, \ldots)$. To see that, note that, due to the Strong Law of Large Numbers and since $h^4 = l^4/d$, the second sum in $R^n_{1,d}$ converges a.s. to a constant. Conditionally on $\sigma(q^n_1, q^n_2, \ldots)$, the distributional limit of the first term in $R^n_{1,d}$ is Gaussian with zero mean and a variance determined by the the a.s. limit of $h^4 \sum_{i=1}^d \big\{\alpha(x^n_i) - \mathbb{E}\,[\,\alpha(x^n_i)\,|\,q^n_i\,]\big\}^2$; this follows from the Martingale Central Limit Theorem (see e.g. Theorem 3.2 of [14]). On the other hand, the limit distribution of $R^n_{2,d}$ is Gaussian with zero mean but, in general, cannot be asymptotically independent of $\sigma(q^0_1, q^0_2, \ldots)$. In the case of RWM or MALA, the conditional expectations that play the role played here by $\mathbb{E}\,[\,\alpha(x^n_i)\,|\,q^n_i\,]$ are identically zero (see the expansions for the acceptance probability in [22] and [23]) and this implies that the corresponding acceptance probabilities are asymptotically independent from $\sigma(q^n_1, q^n_2, \ldots)$ and that the marginal processes $\{q^n_1\}_{n\geq 0}$ are asymptotically Markovian.

The last result in this section provides insight into the limit dynamics of $\{q^n_1\}_{n\geq 0}$:

Proposition 3.9. *Let $Q^n \sim \Pi(Q)$, define*

$$\mathsf{q}^{n+1}_1 = \mathsf{I}^n \cdot \mathcal{P}_q\varphi_T(q^n_1, p^n_1) + \big(1 - \mathsf{I}^n\big)q^n_1; \quad \mathsf{I}^n = \mathbb{I}_{\,U^n \leq a(l)} \ ,$$

and consider q^{n+1}_1 in (3.9). Then, under the hypotheses of Proposition 3.8, as $d \to \infty$:

$$(q^n_1, q^{n+1}_1) \xrightarrow{\ \mathcal{L}\ } (q^n_1, \mathsf{q}^{n+1}_1) \ .$$

Proof. See Section 6. □

This proposition provides a simple description of the asymptotic behaviour of the one-transition dynamics of the marginal trajectories of HMC. As $d \to \infty$, with probability $a(l)$, the HMC particle moves under the *correct* Hamiltonian dynamics. However, the deviation from the true Hamiltonian dynamics, due to the energy errors accumulated from leapfrog integration of all d particles, gives rise to the alternative event of staying at the current position q^n, with probability $1 - a(l)$.

4. Optimal tuning of HMC

In the previous section we addressed the question of how to scale the step-size in the leapfrog integration in terms of the dimension d, leading to Theorem 3.6. In this section we refine this analysis and study the choice of constant l in (3.7). Regardless of the metrics used to measure the efficiency of the algorithm, a good choice of l in (3.7) has to balance the amount of work needed to simulate a full T-leg (interval of length T) of the Hamiltonian dynamics and the probability of accepting the resulting proposal. Increasing l decreases the acceptance probability but also decreases the computational cost of each T-leg integration; decreasing l will yield the opposite effects, suggesting an optimal value of l. In this section we present an analysis that avoids the complex calculations typically associated with the estimation of mixing times of Markov chains, but still provides useful guidance regarding the choice of l. We provide two alternative ways of doing this, summarized in Theorems 4.1 and Theorem 4.2.

4.1. **Asymptotically optimal acceptance probability.** The number of leapfrog steps of length h needed to compute a proposal is obviously given by $\lceil T/h \rceil$. Furthermore, at each step of the chain, it is necessary to evaluate $a(X,Y)$ and sample P. Thus the computing time for a single proposal will be

$$\mathsf{C}_{l,d} := \lceil \frac{T\, d^{1/4}}{l} \rceil \cdot d \cdot C_{LF} + d \cdot C_O \ , \tag{4.1}$$

for some constants C_{LF}, C_O that measure, for one particle, the leapfrog costs and the overheads. Let $\mathsf{E}_{l,d}$ denote the expected computing time until the first accepted T-leg, in stationarity. If N denotes the number of proposals until (and including) the first to be accepted, then

$$\mathsf{E}_{l,d} = \mathsf{C}_{l,d}\ \mathbb{E}\,[\,\mathsf{N}\,] = \mathsf{C}_{l,d}\ \mathbb{E}\,[\,\mathbb{E}\,[\,\mathsf{N}\,|\,Q\,]\,] = \mathsf{C}_{l,d}\ \mathbb{E}\,\Big[\ \frac{1}{\mathbb{E}\,[\,a(X,Y)\,|\,Q\,]}\ \Big]\ .$$

Here we have used the fact that, given the locations Q, the number of proposed T-legs follows a geometric distribution with probability of success $\mathbb{E}\,[\,a(X,Y)\,|\,Q\,]$. Jensen's inequality yields

$$\mathsf{E}_{l,d} \geq \frac{\mathsf{C}_{l,d}}{\mathbb{E}\,[\,a(X,Y)\,]} =: \mathsf{E}^*_{l,d}\ , \tag{4.2}$$

and, from (4.1) and Theorem 3.6, we conclude that:

$$\lim_{d\to\infty} d^{-5/4} \times \mathsf{E}^*_{l,d} = \frac{T\, C_{LF}}{a(l)\, l}\ .$$

A sensible choice for l is that which minimizes the asymptotic cost $\mathsf{E}^*_{l,d}$, that is:

$$l_{opt} = \arg\max_{l>0}\ \mathrm{eff}(l)\,; \quad \mathrm{eff}(l) := a(l)\, l\ .$$

The value of l_{opt} will in general depend on the specific target distribution under consideration. However, by expressing eff as a function of $a = a(l)$, we may write

$$\mathrm{eff} = \Big(\frac{\sqrt{2}}{\Sigma^{\frac{1}{4}}}\Big) \cdot a \cdot \Big(\Phi^{-1}\big(1 - \frac{a}{2}\big)\Big)^{\frac{1}{2}} \tag{4.3}$$

and this equality makes it apparent that $a(l_{opt})$ *does not vary with the selected target.* Fig.1 illustrates the mapping $a \mapsto \mathrm{eff}(a)$; different choices of target distribution only change the vertical scale. In summary, we have:

Theorem 4.1. *Under the hypotheses of Theorem 3.6 and as $d \to \infty$, the measure of cost $\mathsf{E}^*_{l,d}$ defined in (4.2) is minimised for the choice l_{opt} of l that leads to the value of $a = a(l)$ that maximises (4.3). Rounded to 3 decimal places the, target independent, optimal value of the limit probability a is*

$$a(l_{opt}) = 0.651\ .$$

The optimal value identified in the preceding theorem is based on the quantity $\mathsf{E}^*_{l,d}$ that underestimates the expected number of proposals. It may be assumed that the practical optimal average acceptance probability is in fact *greater than* or equal to 0.651. In the next subsection we use an alternative measure of efficiency: the expected squared jumping distance. Consideration of this alternative metric will also lead to the same asymptotically optimal acceptance probability of precisely 0.651 as did the minimisation of $\mathsf{E}^*_{l,d}$. This suggests that, as $d \to \infty$, the consequences of the fact that $\mathsf{E}^*_{l,d}$ underestimates $\mathsf{E}_{l,d}$ become negligible; proving

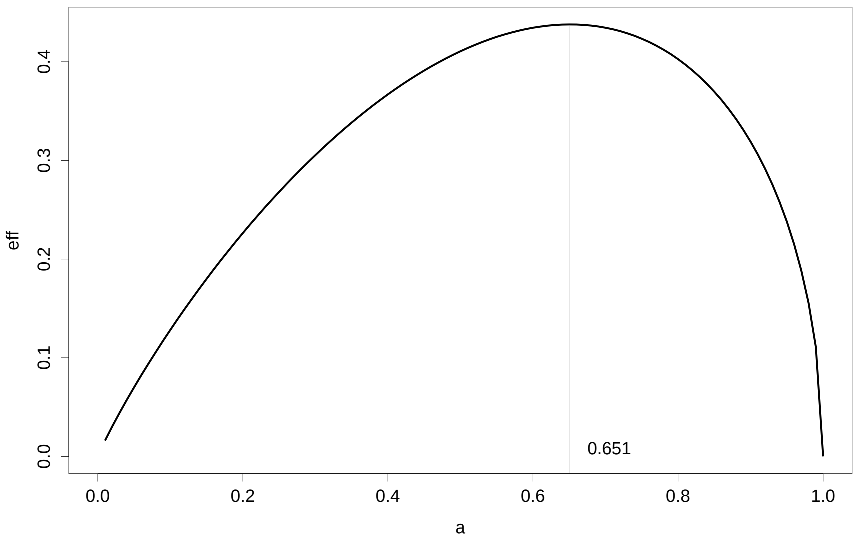


FIGURE 1. The efficiency function $\text{eff} = \text{eff}(a)$.

analytically such a conjecture seems hard given our current understanding of the limiting HMC dynamics.

4.2. **Squared jumping distance.** We now consider the chain $Q^0, Q^1, \ldots$ in stationarity (*i.e.* $Q^0 \sim \Pi(Q)$) and account for the computing cost $\mathsf{C}_{l,d}$ in (4.1) by introducing the continuous-time process $Q^{N(t)}$, where $\{N(t); t \geq 0\}$ denotes a Poisson process of intensity $\lambda_d = 1/\mathsf{C}_{l,d}$. If $q_d(t) := q_1^{N(t)}$ denotes the projection of $Q^{N(t)}$ onto the first particle and $\delta > 0$ is a parameter (the jumping time), we measure the efficiency of HMC algorithms by using the expected squared jump distance:

$$\mathcal{SJD}_d(\delta) = \mathbb{E}\,[\,(q_d(t+\delta) - q_d(t))^2\,]\ .$$

The following result shows that $\mathcal{SJD}_d(\delta)$ is indeed asymptotically maximized by maximizing $a(l)\,l$:

Theorem 4.2. *Under the hypotheses of Proposition 3.8:*

$$\lim_{d\to\infty} d^{5/4} \times \mathcal{SJD}_d = \frac{C_J\,\delta}{T\,C_{LF}} \times a(l)\,l\ .$$

Proof. See Section 6. □

4.3. **Optimal acceptance probability in practice.** As $d \to \infty$, the computing time required for a proposal scales as $1/l$ (see (4.1)) and the number of proposals that may be performed in a given amount of time scales as l. Inspection of (4.1) reveals however that selecting a big value of l gives the full benefit of a proportional increase of the number of proposals only asymptotically, and at the slow rate of $\mathcal{O}(d^{-1/4})$. On the other hand, the average acceptance probability converges at the faster rate $\mathcal{O}(d^{-1/2})$ (this is an application of Stein's method). These considerations suggest that unless $d^{-1/4}$ is very small the algorithm will tend to benefit from average acceptance probabilities higher than 0.651.

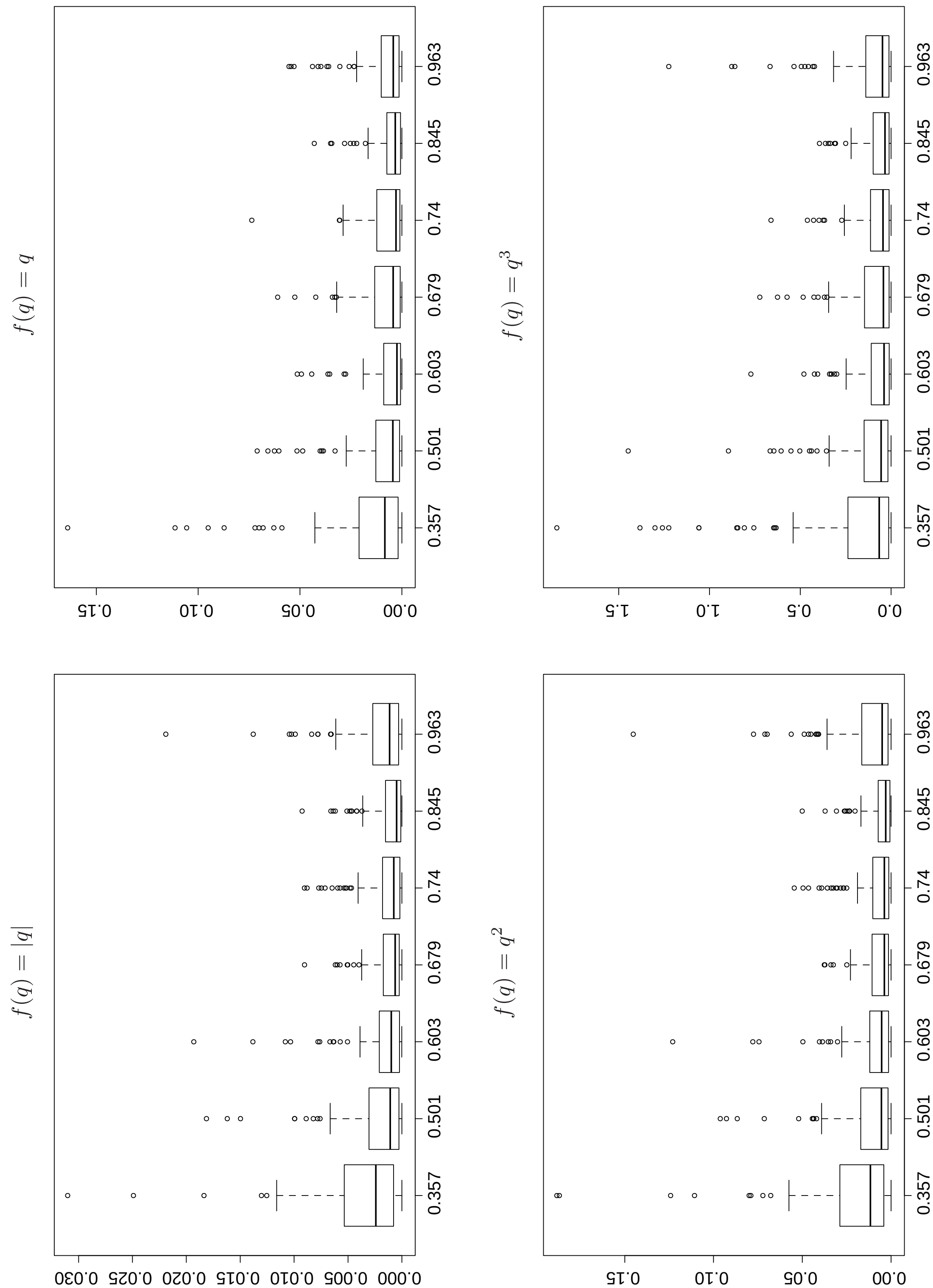


FIGURE 2. Boxplots of Squared Errors (SEs) from Monte-Carlo averages of HMC. For 7 different selections of the leapfrog step-size h (corresponding to the different boxplots in each panel); the values of h are not shown. We ran HMC 120 times; every run was allowed a computing time of 30s. Each boxplot corresponds to the 120 SEs in estimating $\mathbb{E}[\,f(q)\,]$, for a particular h and $f(\cdot)$. Written at the bottom of each boxplots is the median of the 120 empirical average acceptance probabilities for the corresponding h.

Fig.2 shows the results of a numerical study on HMC. The target distribution is a product of $d = 10^5$ standard Gaussian densities $N(0,1)$. We have applied HMC with different choices of the step-size h and, in all cases, allowed the algorithm to run during a computational time t_{comp} of 30 seconds. We used Monte-Carlo averages of the output

$$\hat{f} = \frac{1}{N_{t_{comp}}} \sum_{n=1}^{N_{t_{comp}}} f(q_1^n)$$

to estimate, for different choices of f, the expectation $\mathbb{E}[\, f\,] = \mathbb{E}[\, f(q)\,]$, $q \sim N(0,1)$; here $N_{t_{comp}}$ denotes the number of T-legs carried out within the allowed time t_{comp}. For each choice of h we ran the HMC algorithm 120 times.

Each of the four panels in Fig.2 corresponds to a different choice of $f(\cdot)$. In each of the panels, the various boxplots correspond to choices of h; at the bottom of each boxplot we have written the median of the 120 empirical average acceptance probabilities. The boxplots themselves use the 120 realizations of the squared distances: $(\hat{f} - \mathbb{E}[\, f\,])^2$. The shape of the boxplots endorses the point made above, that the optimal acceptance probability for large (but finite) d is larger than the asymptotically optimal value of 0.651.

5. Estimates for the leapfrog algorithm

In this section we identify hypotheses on V under which Conditions 3.1, 3.2 and 3.7 in Section 3 hold.

We set $f := -\nabla V$ (the 'force') and denote by $f'(q) := f^{(1)}(q), f^{(2)}(q), \ldots$ the successive Fréchet derivatives of f at q. Thus, at a fixed q, $f^{(k)}(q)$ is a multilinear operator from $(\mathbb{R}^m)^{k+1}$ to $\mathbb{R}$. For the rest of this section we will use the following assumptions on V:

Assumptions 5.1. *The function $V : \mathbb{R}^m \to \mathbb{R}$ satisfies:*

- *(i) $V \in C^4(\mathbb{R}^m \to \mathbb{R}_+)$.*
- *(ii) $f', f^{(2)}, f^{(3)}$ are uniformly bounded by a constant B.*

These assumptions imply that the potential $V(q)$ can grow at most quadratically at infinity as $|q| \to \infty$. (If the growth of V is more than quadratic, then the leapfrog algorithm as applied with a constant value of h throughout the phase space is in fact unstable whenever the initial condition is large.) The case where V takes negative values but is bounded from below can be reduced to the case $V \geq 0$ by adding a suitable constant to V. In terms of the target measure this just involves changing the normalization constant and hence is irrelevant in the HMC algorithm.

5.1. Preliminaries.

Differentiating (3.1) with respect to t, we find successively:

$$\begin{aligned}
\ddot{p}(t) &= f'(q(t))M^{-1}p(t) \; , \\
\ddot{q}(t) &= M^{-1}f(q(t)) \; , \\
\dddot{p}(t) &= f^{(2)}(q(t))(M^{-1}p(t), M^{-1}p(t)) + f'(q(t))M^{-1}f(q(t)) \; , \\
\dddot{q}(t) &= M^{-1}f'(q(t))M^{-1}p(t) \; , \\
\ddddot{p}(t) &= f^{(3)}(q(t))(M^{-1}p(t), M^{-1}p(t), M^{-1}p(t)) + \\
&\qquad 3f^{(2)}(q(t))(M^{-1}f(q(t)), M^{-1}p(t)) + f'(q(t))M^{-1}f'(q(t))M^{-1}f(q(t)) \; , \\
\ddddot{q}(t) &= M^{-1}f^{(2)}(q(t))(M^{-1}p(t), M^{-1}p(t)) + M^{-1}f'(q(t))M^{-1}f(q(t)) \; .
\end{aligned}$$

In this section the letter K will denote a generic constant which may vary from one appearance to the next, but will depend only on B, T, $\|M\|$, $\|M^{-1}\|$. From the above equations for the derivatives and using the assumptions on V, we obtain the following bounds:

(5.1)
$$
\begin{aligned}
&|\dot p(t)| \le |f(q(t))| \ , && |\dot q(t)| \le K|p(t)| \ ,\\
&|\ddot p(t)| \le K|p(t)| \ , && |\ddot q(t)| \le K|f(q(t))| \ ,\\
&|\dddot p(t)| \le K(|p(t)|^2 + |f(q(t))|) \ , && |\dddot q(t)| \le K|p(t)| \ ,\\
&|\ddddot p(t)| \le K(|p(t)|^3 + |p(t)||f(q(t))| + |f(q(t))|) \ , && |\ddddot q(t)| \le K(|p(t)|^2 + |f(q(t))|) \ .
\end{aligned}
$$

5.2. **Asymptotic expansion for the leapfrog solution.** In previous sections we have used a subscript to denote the different particles comprising our state space. Here we consider leapfrog integration of a single particle and use the subscript to denote the time-level in this integration. The leapfrog scheme can then be compactly written as

$$q_{n+1} = q_n + hM^{-1}p_n + \frac{h^2}{2}M^{-1}f(q_n) \ , \tag{5.2}$$

$$p_{n+1} = p_n + \frac{h}{2}f(q_n) + \frac{h}{2}f\Big(q_n + hM^{-1}p_n + \frac{h^2}{2}M^{-1}f(q_n)\Big) \ . \tag{5.3}$$

We define the truncation error in the usual way:

$$
\begin{aligned}
-\tau_n^{(q)} &:= q(t_{n+1}) - \Big(q(t_n) + hM^{-1}p(t_n) + \frac{h^2}{2}M^{-1}f(q(t_n))\Big) \ ,\\
-\tau_n^{(p)} &:= p(t_{n+1}) - \Big(p(t_n) + \frac{h}{2}f(q_n) + \frac{h}{2}f\big(q(t_n) + hM^{-1}p(t_n) + \frac{h^2}{2}M^{-1}f(q(t_n))\big)\Big) \ ,
\end{aligned}
$$

where we have set $t_n = nh \in [0,T]$. Expanding (see [12]) we obtain:

$$
\begin{aligned}
\tau_n^{(q)} &= \frac{1}{6}\,h^3\,\dddot q(t_n) + h^4\,\mathcal{O}(\|\ddddot q(\cdot)\|_\infty) \ ,\\
\tau_n^{(p)} &= -\frac{1}{12}\,h^3\,\dddot p(t_n) + h^4\,\mathcal{O}(\|\ddddot p(\cdot)\|_\infty) + h\,\mathcal{O}(\tau_n^{(q)}) \ ,
\end{aligned}
$$

where, for arbitrary function g:

$$\|g(\cdot)\|_\infty := \sup_{0\le t\le T}|g(t)| \ .$$

In view of these estimates, $(1/6)\,h^3\dddot q(t_n)$ and $-(1/12)\,h^3\,\dddot p(t_n)$ are the leading terms in the asymptotic expansion of the truncation error. Standard results (see, for instance, [13], Section II.8) show that the numerical solution possesses an asymptotic expansion:

$$
\begin{aligned}
q_n &= q(t_n) + h^2 v(t_n) + \mathcal{O}(h^3) \ ,\\
p_n &= p(t_n) + h^2 u(t_n) + \mathcal{O}(h^3) \ ,
\end{aligned}
\tag{5.4}
$$

where functions $u(\cdot)$ and $v(\cdot)$ are the solutions, with initial condition $u(0) = v(0) = 0$, of the *variational* system

$$\begin{pmatrix}\dot u(t)\\ \dot v(t)\end{pmatrix} = \begin{pmatrix}0 & M^{-1}f'(q(t))\\ I & 0\end{pmatrix}\begin{pmatrix}u(t)\\ v(t)\end{pmatrix} + \begin{pmatrix}\frac{1}{12}\dddot p(t)\\ -\frac{1}{6}\dddot q(t)\end{pmatrix}. \tag{5.5}$$

Remark 5.2. *Notice here that $u(\cdot), v(\cdot)$ depend on the initial conditions $(q(0), p(0))$ via $(q(\cdot), p(\cdot))$ but this dependence is not reflected in the notation. One should keep in mind that most of the norms appearing in the sequel are functions of $(q(0), p(0))$.*

Applying Gronwall's lemma and using the estimates (5.1), we obtain the bound:

$$\|u(\cdot)\|_\infty + \|v(\cdot)\|_\infty \leq K(\|p(\cdot)\|_\infty^2 + \|f(q(\cdot))\|_\infty) \tag{5.6}$$

and, by differentiating (5.5) with respect to t, expressing $\dot{u}, \dot{v}$ in terms of u, v, and using (5.1) again, we obtain in turn:

$$\|\ddot{u}(\cdot)\|_\infty \leq K(\|p(\cdot)\|_\infty^3 + \|p(\cdot)\|_\infty \|f(q(\cdot))\|_\infty + \|f(q(\cdot))\|_\infty) \ , \tag{5.7}$$

$$\|\ddot{v}(\cdot)\|_\infty \leq K(\|p(\cdot)\|_\infty^2 + \|f(q(\cdot))\|_\infty) \ . \tag{5.8}$$

5.3. **Estimates for the global error.** With the leading coefficients u, v of the global errors $q_n - q(t_n)$, $p_n - p(t_n)$ estimated in (5.6), our task now is to obtain an explicit bound for the constants implied in the $\mathcal{O}(h^3)$ remainder in (5.4). To this end, we define the quantities

$$z_n := q(t_n) + h^2 v(t_n) \ ,$$
$$w_n := p(t_n) + h^2 u(t_n) \ ,$$

and denote by $\tau_n^{(q)*}$, $\tau_n^{(p)*}$ the residuals they generate when substituted in (5.2), (5.3) respectively, *i.e.*,

$$-\tau_n^{(q)*} = z_{n+1} - z_n - hM^{-1}w_n - \frac{h^2}{2}M^{-1}f(z_n) \ ,$$
$$-\tau_n^{(p)*} = w_{n+1} - w_n - \frac{h}{2}f(z_n) - \frac{h}{2}f\Big(z_n + hM^{-1}w_n + \frac{h^2}{2}M^{-1}f(z_n)\Big) \ .$$

Since the leapfrog scheme is stable, we have

$$\max_{0\leq t_n \leq T}(|q_n - z_n| + |p_n - w_n|) \leq \frac{C}{h} \max_{0 \leq t_n \leq T}(|\tau_n^{(q)*}| + |\tau_n^{(p)*}|) \tag{5.9}$$

with the constant C depending only on T and Lipschitz constant of the map $(q_n, p_n) \mapsto (q_{n+1}, p_{n+1})$, which in turn depends on $\|M^{-1}\|$ and the bound for f'. The stability bound (5.9) is the basis of the proof of the following estimation of the global error:

Proposition 5.3. *If the potential V satisfies Assumptions 5.1, then for $0 \leq t_n \leq T$,*

$$|p_n - \big(p(t_n) + h^2 u(t_n)\big)| \leq Kh^3(\|p(\cdot)\|_\infty^4 + \|f(q(\cdot))\|_\infty^2 + 1) \ ,$$
$$|q_n - \big(q(t_n) + h^2 v(t_n)\big)| \leq Kh^3(\|p(\cdot)\|_\infty^4 + \|f(q(\cdot))\|_\infty^2 + 1) \ .$$

Proof. Our task is reduced to estimating $\tau_n^{(q)*}, \tau_n^{(p)*}$. We only present the estimation for $\tau_n^{(p)*}$, since the computations for $\tau_n^{(q)*}$ are similar but simpler.

Indeed, after regrouping the terms,

$$
\begin{aligned}
-\tau_n^{(p)*} &= \underbrace{p(t_{n+1}) - p(t_n) - \frac{h}{2} f(q(t_n)) - \frac{h}{2} f(q(t_{n+1})) + \frac{h^3}{12} \ddddot{p}(t)}_{I_1} \\
&\quad + \underbrace{h^2 \Big(u(t_{n+1}) - u(t_n) - h f'(q(t_n)) v(t_n) - \frac{h}{12} \dddot{p}(t) \Big)}_{I_2} \\
&\quad + \underbrace{\frac{h}{2} \Big(f(q(t_n)) - f(z_n) + h^2 f'(q(t_n)) v(t_n) \Big)}_{I_3} \\
&\quad - \underbrace{\frac{h}{2} \Big(f\big(z_n + h M^{-1} w_n + \frac{h^2}{2} M^{-1} f(q(t_n)) \big) - f(q(t_{n+1}) - h^2 f'(q(t_n)) v(t_n) \Big)}_{I_4} \\
&\quad + \underbrace{\frac{h}{2} \Big(f\big(z_n + h M^{-1} w_n + \frac{h^2}{2} M^{-1} f(q(t_n)) \big) - f\big(z_n + h M^{-1} w_n + \frac{h^2}{2} M^{-1} f(z_n) \big) \Big)}_{I_5}
\end{aligned}
$$

Now we estimate the above five terms separately.
I_1: We note that

$$
p(t_{n+1}) - p(t_n) - \frac{h}{2} f(q(t_n)) - \frac{h}{2} f(q(t_{n+1})) = p(t_{n+1}) - p(t_n) - \frac{h}{2} \dot{p}(t_{n+1}) - \frac{h}{2} \dot{p}(t_n) \ .
$$

and by using the estimates in (5.1) it follows that

$$
|I_1| \leq K h^4 (\|p(\cdot)\|_\infty + \|p(\cdot)\|_\infty \|f(q(\cdot))\|_\infty + \|f(q(\cdot))\|_\infty) \ .
$$

I_2: Here we write $I_2 = h^2(u(t_{n+1}) - u(t_n) - h\,\dot{u}(t_n))$ so that by (5.7)

$$
|I_2| \leq K h^4 (\|p(\cdot)\|_\infty^3 + \|p(\cdot)\|_\infty \|f(q(\cdot))\|_\infty + \|f(q(\cdot))\|_\infty).
$$

I_3 : This term is estimated, after Taylor expanding $f(z_n)$ near $f(q(t_n))$, by

$$
|I_3| \leq K h^5 (\|p(\cdot)\|_\infty + \|f(q(\cdot))\|_\infty)^2.
$$

I_4 : We rewrite this as

$$
\frac{h}{2} \Big(f\big(q(t_{n+1}) + \tau_n^{(q)} + h^2 v(t_n) + h^3 M^{-1} v(t_n) \big) - f(q(t_{n+1})) - h^2 f'(q(t_n)) v(t_n) \Big)
$$

and Taylor expand around $f(q(t_n))$ to derive the bound:

$$
|I_4| \leq K h^4 (\|p(\cdot)\|_\infty^4 + \|f(q(\cdot))\|_\infty^2).
$$

I_5 : This term is easily estimated as:

$$
|I_5| \leq K h^5 \|v(\cdot)\|_\infty \leq K h^5 (\|p(\cdot)\|_\infty^2 + \|f(q(\cdot))\|_\infty) \ .
$$

Combining all the above estimates, we have the bound

$$
|\tau_n^{(p)*}| \leq K h^4 (\|p(\cdot)\|_\infty^4 + \|f(q(\cdot))\|_\infty^2) \ .
$$

A similar analysis for $\tau_n^{(q)*}$ yields the bound

$$
|\tau_n^{(q)*}| \leq K h^4 (\|p(\cdot)\|_\infty^4 + \|f(q(\cdot))\|_\infty^2) \ .
$$

The proof is completed by substituting the above estimates in (5.9). □

We now use the estimates in Proposition 5.3 to derive the asymptotic expansion for the energy increment for the leapfrog scheme (cf. Condition 1).

Proposition 5.4. *Let potential V satisfy Assumptions 5.1. Then, for the leapfrog scheme, we get*

$$\Delta(x,h) = h^2\alpha(x) + h^2\rho(x,h) \ ,$$

with

$$\begin{aligned} \alpha(x) &= \langle M^{-1}p(T), u(T)\rangle - \langle f(q(T)), v(T)\rangle \ , \\ |\alpha(x)| &\leq K(\|p(\cdot)\|_\infty^3 + \|f(q(\cdot))\|_\infty^2 + 1) \ , \\ |\rho(x,h)| &\leq Kh(\|p(\cdot)\|_\infty^8 + \|f(q(\cdot))\|_\infty^2 + 1), \quad 0 < h \leq 1 \ , \end{aligned}$$

where $(q(\cdot), p(\cdot))$ denotes the solution of (3.1) with initial data $x \equiv (q(0), p(0))$ and $u(\cdot), v(\cdot)$ are the solutions of the corresponding variational system given in (5.5) with $u(0) = v(0) = 0$.

Proof. We only consider the case when T/h is an integer. The general case follows with minor adjustments. By Proposition 5.3,

$$\begin{aligned} \Delta(x,h) &= H(\psi_h^{(T)}(x)) - H(x) = H(\psi_h^{(T)}(x)) - H(\varphi_T(x)) = \\ &= \langle M^{-1}p(T), h^2u(T) + h^3R_1\rangle + \frac{1}{2}\left\langle M^{-1}(h^2u(T) + h^3R_1, (h^2u(T) + h^3R_1)\right\rangle \\ &\qquad + V\Big(q(T) + h^2v(T) + h^3R_2\Big) - V(q(T)) \ , \end{aligned}$$

where R_1, R_2 are remainders with

$$|R_1| + |R_2| \leq K(\|p(\cdot)\|_\infty^4 + \|f(q(\cdot))\|_\infty^2 + 1) \ .$$

By Taylor expanding $V(\cdot)$ around $q(T)$ we obtain,

$$\Delta(x,h) = h^2\big(\langle M^{-1}p(T), u(T)\rangle - \langle f(q(T)), v(T)\rangle\big) + \rho(x,h) \ ,$$

with

$$|\rho(x,h)| \leq Kh^3(\|p(\cdot)\|_\infty^8 + \|f(q(\cdot))\|_\infty^2 + 1)$$

for $0 \leq h \leq 1$. From the bound (5.6) it follows that

$$\begin{aligned} |\alpha(x)| &\leq K(\|p(\cdot)\|_\infty \|u(\cdot)\|_\infty + \|f(q(\cdot))\|_\infty \|v(\cdot)\|_\infty) \\ &\leq K(\|p(\cdot)\|_\infty^3 + \|f(\cdot)\|_\infty^2 + 1) \end{aligned}$$

and the theorem is proved. □

Our analysis is completed by estimating the quantities $\|p(\cdot)\|_\infty$ and $\|q(\cdot)\|_\infty$, that feature in the preceding theorems, in terms of the initial data $(q(0), p(0))$. We obtain these estimates for two families of potentials which include most of the interesting/useful target distributions. The corresponding estimates for other potentials may be obtained using similar methods.

Proposition 5.5. *Let potential V satisfy Assumptions 5.1. If V satisfies, in addition, either of the following conditions:*

(i) *f is bounded and*

$$\int_{\mathbb{R}^m} |V(q)|^8 e^{-V(q)} dq < \infty \ ; \tag{5.10}$$

(ii) *there exist constants* $C_1, C_2 > 0$ *and* $0 < \gamma \leq 1$ *such that for all* $|q| \geq C_2$, *we have* $V(q) \geq C_1 |q|^\gamma$ *;*

then Conditions 3.1, 3.2 and 3.7 all hold.

Proof. We only present the treatment of Conditions 3.1 and 3.2. The derivation of Condition 3.7 is similar and simpler.

From Proposition 5.4 we observe that function $D(x)$ in Condition 3.2 may be taken to be

$$D(x) = K(\|p(\cdot)\|_\infty^{16} + \|f(q(\cdot))\|_\infty^4 + 1) \ .$$

Thus, to prove integrability of $D(\cdot)$ we need to estimate $\|p(\cdot)\|_\infty$ and $\|f(q(\cdot))\|_\infty$. Estimating $\|p(\cdot)\|_\infty$ is easier. Indeed, by conservation of energy,

$$\frac{1}{2}\langle p(t), M^{-1}p(t)\rangle \leq \frac{1}{2}\langle p(0), M^{-1}p(0)\rangle + V(q(0)) \ ,$$

which implies

$$|p(t)|^{16} \leq K(|p(0)|^{16} + |V(q(0))|^8) \ . \tag{5.11}$$

Now, we prove integrability of $D(\cdot)$ under each of the two stated hypothesis.

Under hypothesis (i): Suppose f is bounded. In this case we obtain that $|D(x)| \leq K(\|p(\cdot)\|_\infty^{16} + 1)$, therefore it is enough to estimate $\|p(\cdot)\|_\infty$. Since the Gaussian distribution has all moments, integrability of D follows from (5.10) and (5.11).

Under hypothesis (ii): Using the stated hypothesis on $V(q)$ we obtain

$$C_1 |q(t)|^\gamma \leq V(q(t)) \leq \frac{1}{2}\langle p(0), M^{-1}p(0)\rangle + V(q(0)) \ ,$$

which implies that:

$$|q(t)| \leq K\Big(|p(0)|^{\frac{2}{\gamma}} + |V(q(0))|^{\frac{1}{\gamma}}\Big) \ .$$

By Assumptions 5.1(ii), $|f(q(t))| \leq K(1 + |q(t)|)$ and arguing as above and using the bound (5.11), integrability of D follows if we show that

$$\int_{\mathbb{R}^m} |V(q)|^\delta \, e^{-V(q)} dq < \infty \ , \quad \delta = \max(8, \frac{4}{\gamma}) \ .$$

Since $|V(q)| \leq K(1 + |q|^2)$,

$$\int_{\mathbb{R}^m} |V(q)|^\delta \, e^{-V(q)} dq \leq K \int_{\mathbb{R}^m} (1 + |q|^{2\delta}) \, e^{-B|q|^\gamma} dq < \infty$$

and we are done. □

6. Proofs of Probabilistic Results

Proof of Lemma 3.3. The volume preservation property of $\psi_h^{(T)}(\cdot)$ implies that the associated Jacobian is unit. Thus, setting $x = (\psi_h^{(T)})^{-1}(y)$ we get:

$$\begin{aligned}\int_{\mathbb{R}^{2m}} \varphi(\Delta(x,h)) \, e^{-H(x)} \, dx &= \int_{\mathbb{R}^{2m}} \varphi\big(H(\psi_h^{(T)}(x)) - H(x)\big) \, e^{-H(x)} dx \\ &= \int_{\mathbb{R}^{2m}} \varphi\big[H(y) - H((\psi_h^{(T)})^{-1}(y))\big] \, e^{-H((\psi_h^{(T)})^{-1}(y))} dy \ .\end{aligned}$$

Following the definition of time reversibility in (2.2), we have:

$$S \circ \psi_h^{(T)} = (\psi_h^{(T)})^{-1} \circ S$$

for the symmetry operator S such that $S(q,p)=(q,-p)$. Using now the volume presenving transformation $y=Sz$ and continuing from above, we get:

$$\begin{aligned}&\int_{\mathbb{R}^{2m}}\varphi(\Delta(x,h))\,e^{-H(x)}\,dx\\&\qquad=\int_{\mathbb{R}^{2m}}\varphi\big(H(Sz)-H((\psi_h^{(T)})^{-1}(Sz))\big)\,e^{-H((\psi_h^{(T)})^{-1}(Sz))}dz\\&\qquad=\int_{\mathbb{R}^{2m}}\varphi\big(H(Sz)-H(S\psi_h^{(T)}(z))\big)\,e^{-H(S(\psi_h^{(T)}(z)))}dz\\&\qquad=\int_{\mathbb{R}^{2m}}\varphi\big(H(z)-H(\psi_h^{(T)}(z))\big)\,e^{-H(\psi_h^{(T)}(z))}dz,\end{aligned}$$

where in the last equation we have used the identity $H(Sz)=H(z)$. □

Proof of Proposition 3.4. We will first find the limit of $\sigma^2(h)/h^4$. Conditions 3.1 and 3.2 imply that:

$$\frac{\Delta^2(x,h)}{h^4}=\alpha^2(x)+\rho^2(x,h)+2\rho(x,h)\alpha(x)\leq D(x)$$

and since, for fixed x, $\Delta^2(x,h)/h^4\to\alpha^2(x)$, the dominated convergence theorem shows:

$$\lim_{h\to0}\frac{s^2(h)}{h^4}=\int_{\mathbb{R}^{2m}}\alpha^2(x)\,e^{-H(x)}dx=\Sigma\ .$$

Now, (3.6) implies that:

$$\lim_{h\to0}\frac{\mu^2(h)}{h^4}=0\ , \tag{6.1}$$

and the required limit for $\sigma^2(h)/h^4$ follows directly. Then, from (3.5) we obtain

$$\begin{aligned}&\frac{2\mu(h)-\sigma^2(h)}{h^4}=\\&\qquad-\int_{\mathbb{R}^{2m}}\frac{\Delta(x,h)}{h^2}\,\frac{\big[\exp(-\Delta(x,h))-1+\Delta(x,h)\big]}{h^2}\,e^{-H(x)}\,dx+\frac{\mu^2(h)}{h^4}\ .\end{aligned}$$

Since for any fixed x, Conditions 3.1 and 3.2 imply that $\Delta(x,h)\to0$ as $h\to0$ and $\Delta^2(x,h)=\mathcal{O}(h^4)$, we have the pointwise limit

$$\lim_{h\to0}\frac{\exp(-\Delta(x,h))-1+\Delta(x,h)}{h^2}=0\ .$$

Using the inequality $|u||e^u-1-u|\leq|u|^2(e^u+2)$, we deduce that for all sufficiently small h,

$$\begin{aligned}&\int_{\mathbb{R}^{2m}}\frac{|\Delta(x,h)|}{h^2}\,\frac{\big|\exp(-\Delta(x,h))\big|}{h^2}\,e^{-H(x)}\,dx\\&\leq\int_{\mathbb{R}^{2m}}\frac{|\Delta^2(x,h)|}{h^4}\exp(-\Delta(x,h))\,e^{-H(x)}\,dx+2\int_{\mathbb{R}^{2m}}\frac{|\Delta^2(x,h)|}{h^4}\,e^{-H(x)}\,dx\\&\leq3\int_{\mathbb{R}^{2m}}D(x)\,e^{-H(x)}dx<\infty,\end{aligned}$$

where the last line follows from applying Lemma 3.3 with $\varphi(x)=x^2$ and Condition 3.2. So, the dominated convergence theorem yields

$$\lim_{h\to0}\frac{2\mu(h)-\sigma^2(h)}{h^4}=0\ .$$

This completes the proof of the proposition. □

Proof of Theorem 3.6. We continue from (3.8). In view of the scaling $h = l \cdot d^{-1/4}$ we obtain, after using Proposition 3.4:

$$\mathbb{E}\,[\,R_d\,] = -d \cdot \mu(h) \to -\frac{l^4\,\sigma}{2}$$

and

$$\mathrm{Var}\,[\,R_d\,] = d \cdot \sigma^2(h) \to l^4\,\Sigma\ .$$

The Lindeberg condition is easily seen to hold and therefore:

$$R_d \xrightarrow{\mathcal{L}} R_\infty := N(-\tfrac{l^4\,\Sigma}{2}, l^4\,\Sigma)\ .$$

From the boundedness of $u \mapsto 1 \wedge e^u$ we may write:

$$\mathbb{E}\,[\,a(X,Y)\,] \to \mathbb{E}\,[\,1 \wedge e^{R_\infty}\,]\ ,$$

where the last expectation can be found analytically (see e.g. [22]) to be:

$$\mathbb{E}\,[\,1 \wedge e^{R_\infty}\,] = 2\Phi(-l^2\sqrt{\Sigma}/2)\ .$$

This completes the proof. □

Proof of Proposition 3.8. For simplicity, we will write just q^n, q^{n+1} and p^n instead of q_1^n, q_1^{n+1}, p_1^n respectively. Using (3.9), we get:

$$(q^{n+1} - q^n)^2 = I^n\,(\mathcal{P}_q\psi_h^{(T)}(q^n,p^n) - q^n)^2\ .$$

We define:

$$a^-(X^n,Y^n) := 1 \wedge \exp\big\{-\sum_{i=2}^{d}\Delta(x_i^n,h)\big\}\,; \quad I^{n-} := \mathbb{I}_{\,U^n < a^-(X^n,Y^n)}\ , \tag{6.2}$$

and set

$$\xi^n = I^{n-}(\,\mathcal{P}_q\,\psi_h^{(T)}(q^n,p^n) - q^n\,)^2\ .$$

Using the Lipschitz continuity of $u \mapsto \mathbb{I}_{\,U\,\leq\,1\wedge\,e^u}$ and the Cauchy-Schwartz inequality we get:

$$\mathbb{E}\,|(q^{n+1}-q^n)^2 - \xi^n| \leq |\Delta(x_1,h)|_{L_2}\,|(\mathcal{P}_q\psi_h^{(T)}(q^n,p^n) - q^n)^2|_{L_2}$$

Now, Conditions 3.1 and 3.2 imply that

$$|\Delta(x_1,h)|_{L_2} = \mathcal{O}(h^2)\ .$$

Also, from Condition 3.7 and the stated hypothesis on the density $\exp(-V)$, q^n and $\mathcal{P}_q\psi_h^{(T)}(q^n,p^n)$ have bounded fourth moments uniformly in h, so:

$$|(\mathcal{P}_q\psi_h^{(T)}(q^n,p^n) - q^n)^2|_{L_2} \leq C\ ,$$

for some constant $C > 0$. The last two statements imply that:

$$\mathbb{E}\,|(q^{n+1}-q^n)^2 - \xi^n| = \mathcal{O}(h^2)\ . \tag{6.3}$$

Exploiting the independence between I^{n-} and the first particle:

$$\mathbb{E}\,[\,\xi_n\,] = \mathbb{E}\,[\,a^-(X,Y)\,] \times \mathbb{E}\,[\,(\mathcal{P}_q\psi_h^{(T)}(q^n,p^n) - q^n)^2\,] \longrightarrow a(l)\cdot\mathbb{E}\,[\,(\mathcal{P}_q\varphi_T(q^n,p^n) - q^n)^2\,]\ ,$$

where, for the first factor we used its limit from Theorem 3.6; for the second factor the limit is a consequence by Condition 3 and the dominated convergence theorem. Equation (6.3) completes the proof. □

Proof of Proposition 3.9. Fix some $q_1^n \in \mathbb{R}^m$. We define $a^-(X^n, Y^n)$ and I^{n-} as in (6.2). For simplicity, we will write just q^n, q^{n+1}, q^{n+1} and p^n instead of q_1^n, q_1^{n+1}, q_1^{n+1} and p_1^n respectively.

We set

$$g^{n+1} = I^{n-} \cdot \mathcal{P}_q \varphi_T(q^n, p^n) + \big(1 - I^{n-}\big)\, q^n \ .$$

Adding and subtracting $I^n \cdot \mathcal{P}_q(\varphi_T(q^n, p^n))$ yields:

$$\begin{aligned} |q^{n+1} - g^{n+1}| \le\ & |\mathcal{P}_q(\psi_h^{(T)}(q^n, p^n)) - \mathcal{P}_q(\varphi_T(q^n, p^n))| \\ & + |I^{n-} - I^n| \,\big(|\mathcal{P}_q(\varphi_T(q^n, p^n))| + |q^n|\big) \ . \end{aligned} \tag{6.4}$$

Using the Lipschitz continuity (with constant 1) of $u \mapsto \mathbb{I}_{U \le 1 \wedge \exp(u)}$:

$$|I^{n-} - I^n| \le |\Delta(x_1, h)| \ . \tag{6.5}$$

Now, Condition 3.7 implies that the first term on the right-hand side of (6.4) vanishes w.p.1 and Condition 3.1 implies (via (6.5)) that also the second term vanishes w.p.1. Therefore, as $d \to \infty$:

$$q^{n+1} - g^{n+1} \to 0, \ \text{ a.s. .}$$

Theorem 3.6 immediately implies that $I^{n-} \xrightarrow{\mathcal{L}} \mathsf{I}^n$, thus:

$$g^{n+1} \xrightarrow{\mathcal{L}} \mathsf{q}^{n+1} \ .$$

From these two limits, we have $q^{n+1} \xrightarrow{\mathcal{L}} \mathsf{q}^{n+1}$, and this completes the proof. □

Proof of Theorem 4.2. To simplify the notation we again drop the subscript 1. Conditionally on the trajectory $q^0, q^1, \ldots$ we get:

$$(q(t+\delta) - q(t))^2 = \begin{cases} 0\,, & \text{w.p. } 1 - \lambda_d\delta + \mathcal{O}((\lambda_d\delta)^2)\ , \\ (q^{N(t)+1} - q^{N(t)})^2\,, & \text{w.p. } \lambda_d\delta + \mathcal{O}((\lambda_d\delta)^2)\ , \\ (q^{N(t)+1+j} - q^{N(t)})^2\,,\ j \ge 1\,, & \text{w.p. } \mathcal{O}((\lambda_d\delta)^{j+1})\ . \end{cases}$$

Therefore,

$$\begin{aligned} \mathcal{SJD}_d = {} & \mathbb{E}\,[\,(q^{N(t)+1} - q^{N(t)})^2\,]\,(\lambda_d\delta + \mathcal{O}((\lambda_d\delta)^2)) \\ & + \sum_{j\ge 1} \mathbb{E}\,[\,(q^{N(t)+1+j} - q^{N(t)})^2\,]\,\mathcal{O}((\lambda_d\delta)^{j+1})\ . \end{aligned} \tag{6.6}$$

Note now that:

$$\begin{aligned} \mathbb{E}\,[\,(q^{N(t)+1+j} - q^{N(t)})^2\,] &\le \Big(\sum_{k=1}^{j+1} |q^{N(t)+k} - q^{N(t)+k-1}|_{L_2}\Big)^2 \\ &= (j+1)^2\,\mathbb{E}\,[\,(q^{n+1} - q^n)^2\,]\ , \end{aligned}$$

since we have assumed stationarity. From (4.1):

$$\lambda_d = d^{-5/4}\,\frac{l}{T\,C_{LF}} + \mathcal{O}(d^{-6/4})\ .$$

and, from Proposition 3.8, $\mathbb{E}\,[\,(q^{n+1} - q^n)^2\,] = \mathcal{O}(1)$. Therefore,

$$d^{5/4} \times \sum_{j\ge 1} \mathbb{E}\,[\,(q^{N(t)+1+j} - q^{N(t)})^2\,]\,\mathcal{O}((\lambda_d\delta)^{j+1})$$

is of the same order in d as

$$\lambda_d^2 \cdot d^{5/4} \times \sum_{j\geq 1} (j+1)^2 \, \mathcal{O}(\lambda_d^{j-1}) \ ,$$

thus:

$$d^{5/4} \times \sum_{j\geq 1} \mathbb{E}\left[\, (q^{N(t)+1+j} - q^{N(t)})^2 \,\right] \mathcal{O}((\lambda_d \delta)^{j+1}) = \mathcal{O}(\lambda_d) \ .$$

Using this result, and continuing from (6.6), Proposition 3.8 provides the required statement. □

7. Conclusions

The HMC methodology provides a promising framework for the study of a number of sampling problems, especially in high dimensions. There are a number of directions in which the research direction taken in this paper could be developed further. We list some of them.

- The overall optimization involves tuning *three* free parameters (l, T, M), and since M is a matrix, the number of parameters to be optimized over, is even more in general. In this paper, we have fixed M and T, and focussed on optimizing the HMC algorithm over choice of step-size h. The natural next step would be to study the algorithm for various choices of the mass matrix M and the integration time T.
- We have concentrated on explicit integration by the leapfrog method. For measures which have density with respect to a Gaussian measure (in the limit $d \to \infty$) it may be of interest to use semi-implicit integrators. This idea has been developed for the MALA algorithm (see [4] and the references therein) and could also be developed for HMC methods. It has the potential of leading to methods which explore state space in $\mathcal{O}(1)$ steps.
- The issue of irreducibility for the transition kernel of HMC is subtle, and requires further investigation, as certain exceptional cases can lead to non-ergodic behaviour (see [5, 27] and the references therein).
- There is evidence that the limiting properties of MALA for high-dimensional target densities do not appear to depend critically on the tail behaviour of the target (see [23]). However in the present paper for HMC, we have considered densities that are no lighter than Gaussian at innity. It would thus be interesting to extend the work to light-tailed densities. This links naturally to the question of using variable step size integration for HMC since light tailed densities will lead to superlinear vector fields at infinity in (2.1).
- There is interesting recent computational work [9] concerning exploration of state space by means of nonseparable Hamiltonian dynamics; this work opens up several theoretical research directions.
- We have shown how to scale the HMC method to obtain $\mathcal{O}(1)$ acceptance probabilities as the dimension of the target product measure grows. We have also shown how to minimize a reasonable measure of computational cost, defined as the work needed to make an $\mathcal{O}(1)$ move in state space. However, in contrast to similar work for RWM and MALA ([22, 23]) we have not completely identified the limiting Markov process which arises in the infinite dimensional limit. This remains an interesting and technically demanding challenge.

Acknowledgements

We thank Sebastian Reich for drawing our attention to the paper [11] which sparked our initial interest in the scaling issue for HMC. Further thanks also to Gabriel Stoltz for stimulating discussions. Part of this paper was written when NSP was visiting JMS at the University of Valladolid and we thank this institution for its warm hospitality.

References

1. Elena Akhmatskaya, Nawaf Bou-Rabee, and Sebastian Reich, *A comparison of generalized hybrid monte carlo methods with and without momentum flip*, J. Comput. Phys. **228** (2009), no. 6, 2256–2265.
2. F.J. Alexander, G.L. Eyink, and J.M. Restrepo, *Accelerated monte carlo for optimal estimation of time series*, J. Stat. Phys. **119** (2005), no. 5-6, 1331–1345.
3. A. Beskos, G.O. Roberts, and A.M. Stuart, *Optimal scalings for local Metropolis-Hastings chains on non-product targets in high dimensions*, Ann. Appl. Probab. **19** (2009), no. 3, 863–898.
4. A. Beskos and A.M. Stuart, *MCMC methods for sampling function space*, Proceedings of the International Congress of Industrial and Applied Mathematicians (2009).
5. E. Cancès, F. Legoll, and G. Stoltz, *Theoretical and numerical comparison of some sampling methods for molecular dynamics*, M2AN Math. Model. Numer. Anal. **41** (2007), no. 2, 351–389.
6. L. Chen, Z. Qin, and J. Liu, *Exploring Hybrid Monte Carlo in Bayesian computation*, Proceedings, ISBA, 2000.
7. P. Diaconis, S. Holmes, and R.M. Neal, *Analysis of a nonreversible markov chain sampler*, Annals of Applied Probability **10** (2000), no. 3, 726–752.
8. S. Duane, A.D. Kennedy, B. Pendleton, and D. Roweth, *Hybrid Monte Carlo*, Phys. Lett. B **195** (1987), no. 2, 216–222.
9. M. Girolami, B. Calderhead, and S. Chin, *Riemannian manifold hamiltonian monte carlo*, Tech. report, University of Glasgow, Department of Computing Science, 2009.
10. R. Gupta, G.W. Kilcup, and S.R. Sharpe, *Tuning the hybrid monte carlo algorithm*, Phys. Rev. D **38** (1988), no. 4, 1278–1287.
11. S. Gupta, A. Irbäck, F. Karsch, and B. Petersson, *The acceptance probability in the hybrid Monte Carlo method*, Phys. Lett. B **242** (1990), no. 3-4, 437–443.
12. Ernst Hairer, Christian Lubich, and Gerhard Wanner, *Geometric numerical integration*, second ed., Springer Series in Computational Mathematics, vol. 31, Springer-Verlag, Berlin, 2006, Structure-preserving algorithms for ordinary differential equations.
13. Ernst. Hairer, Syvert. P. Nørsett, and Gerhard. Wanner, *Solving ordinary differential equations. I*, Springer Series in Computational Mathematics, vol. 8, Springer-Verlag, Berlin, 1987, Nonstiff problems.
14. P. Hall and C. C. Heyde, *Martingale limit theory and its application*, Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1980, Probability and Mathematical Statistics.
15. U. H. E. Hansmann, Y. Okamoto, and F. Eisenmenger, *Molecular dynamics, Langevin and hydrid Monte Carlo simulations in a multicanonical ensemble*, Chemical Physics Letters **259** (1996), no. 3-4, 321 – 330.
16. Martin Hasenbusch, *Speeding up the hybrid Monte Carlo algorithm for dynamical fermions*, Phys. Lett. B **519** (2001), no. 1-2, 177–182.
17. Benedict Leimkuhler and Sebastian Reich, *Simulating Hamiltonian dynamics*, Cambridge Monographs on Applied and Computational Mathematics, vol. 14, Cambridge University Press, Cambridge, 2004. MR MR2132573 (2006a:37078)
18. N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller, *Equation of state calculations by fast computing machines*, The Journal of Chemical Physics **21** (1953), no. 6, 1087–1092.
19. L. Mohamed, M. Christie, and V. Demyanov, *Comparison of stochastic sampling algorithms for uncertainty quantification*, Tech. report, Institute of Petroleum Engineering, Heriot-Watt University, Edinburgh, 2009, SPE Reservoir Simulation Symposium.

20. R.M. Neal, *Probabilistic inference using Markov chain Monte Carlo methods*, Tech. report, Department of Computer Science, University of Toronto, 1993.
21. ______, *Bayesian learning for neural networks*, Springer-Verlag, 1996.
22. G. O. Roberts, A. Gelman, and W. R. Gilks, *Weak convergence and optimal scaling of random walk Metropolis algorithms*, Ann. Appl. Probab. **7** (1997), no. 1, 110–120.
23. G.O. Roberts and J.S. Rosenthal, *Optimal scaling of discrete approximations to Langevin diffusions*, J. R. Stat. Soc. Ser. B Stat. Methodol. **60** (1998), no. 1, 255–268.
24. ______, *Optimal scaling for various Metropolis-Hastings algorithms*, Statist. Sci. **16** (2001), no. 4, 351–367.
25. G.O. Roberts and R.L. Tweedie, *Exponential convergence of Langevin distributions and their discrete approximations*, Bernoulli **2** (1996), no. 4, 341–363.
26. Jesus-Maria Sanz-Serna and Mari-Paz Calvo, *Numerical Hamiltonian problems*, Applied Mathematics and Mathematical Computation, vol. 7, Chapman & Hall, London, 1994.
27. Christof Schütte, *Conformational dynamics: Modelling, theory, algorithm, and application of biomolecules*, (1998), Habilitation Thesis, Dept. of Mathematics and Computer Science, Free University Berlin, Available at http://proteomics-berlin.de/89/.
28. J. C. Sexton and D. H. Weingarten, *Hamiltonian evolution for the hybrid Monte Carlo algorithm*, Nuclear Physics B **380** (1992), no. 3, 665–677.
29. R.D. Skeel, *Integration schemes for molecular dynamics and related applications*, The graduate student's guide to numerical analysis '98 (Leicester), Springer Ser. Comput. Math., vol. 26, Springer, Berlin, 1999, pp. 119–176.
30. M.E. Tuckerman, B.J. Berne, G.J. Martyna, and M.L. Klein, *Efficient molecular dynamics and hybrid monte carlo algorithms for path integrals*, The Journal of Chemical Physics **99** (1993), no. 4, 2796–2808.
31. M. Zlochin and Y. Baram, *Manifold stochastic dynamics for bayesian learning*, Neural Computation **13** (2001), no. 11, 2549–2572.

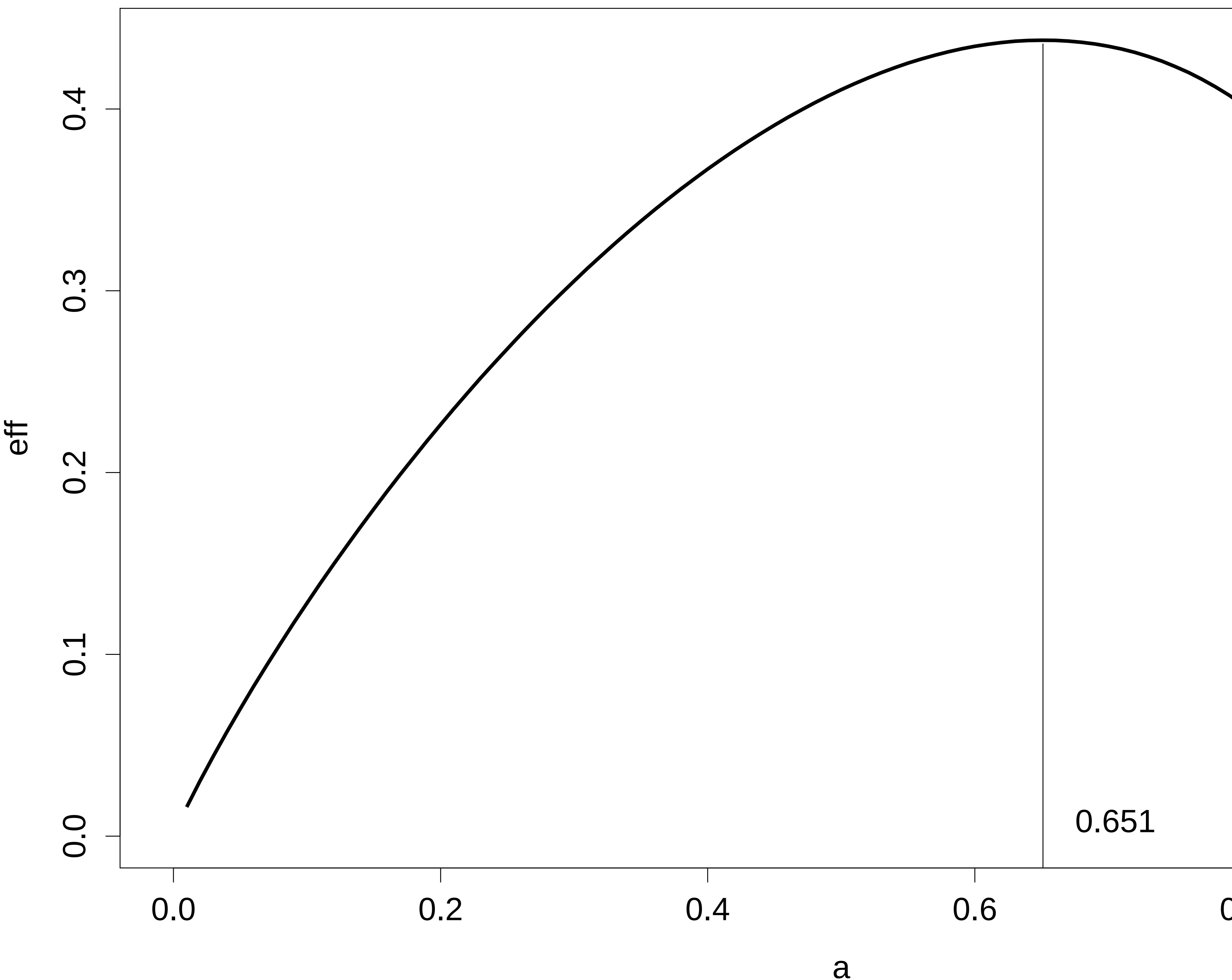
Department of Statistical Science, UCL, Gower Street, London, WC1E 6BT, UK
E-mail address: `alex@stats.ucl.ac.uk`

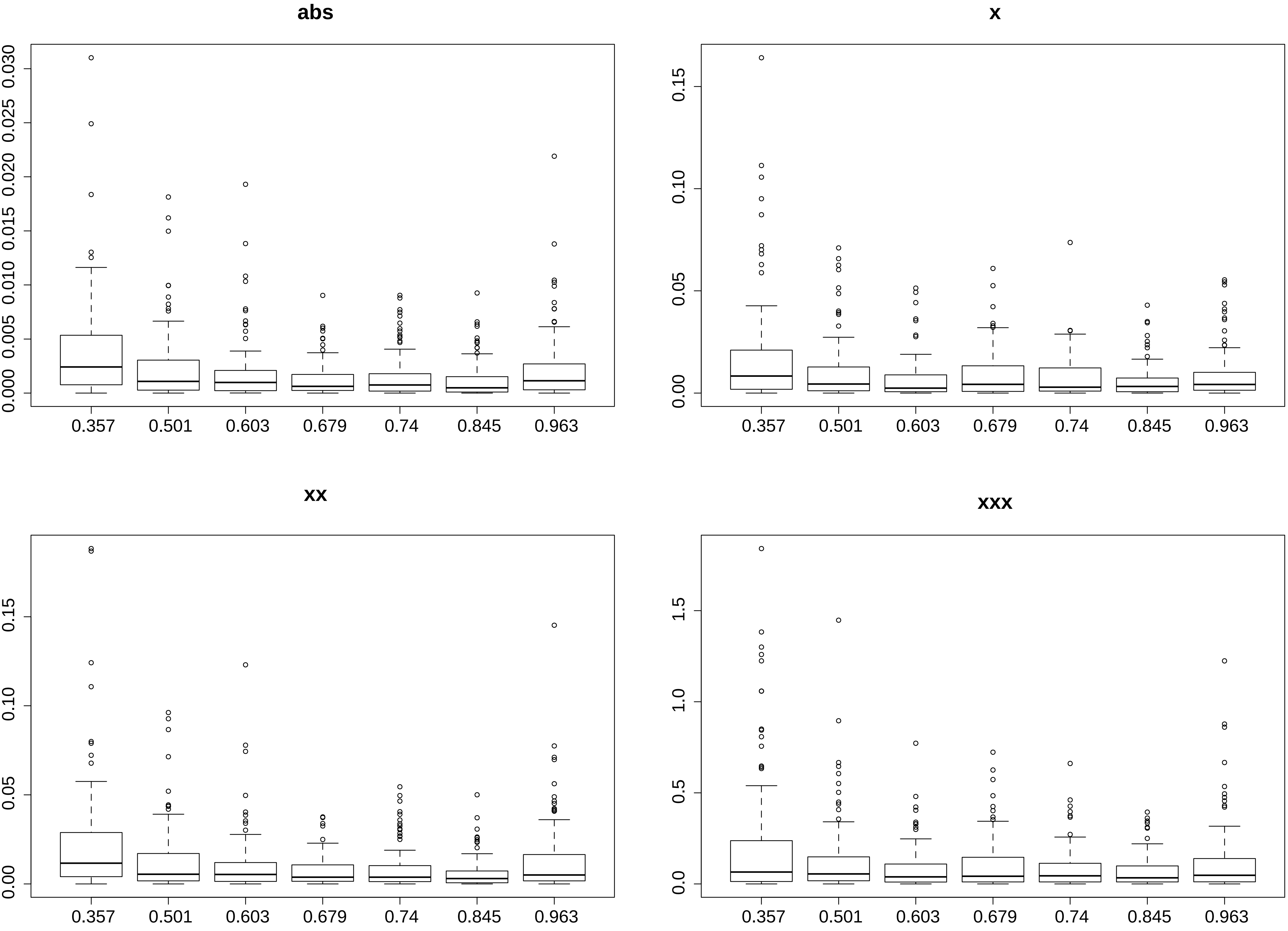
CRiSM, Department of Statistics, University of Warwick, Coventry, CV4 7AL, UK
E-mail address: `n.pillai@warwick.ac.uk`

Department of Statistics, University of Warwick, Coventry, CV4 7AL, UK
E-mail address: `gareth.o.roberts@warwick.ac.uk`

Departamento de Matematica Aplicada, Facultad de Ciencias, Universidad de Valladolid, Spain
E-mail address: `sanzsern@mac.uva.es`

Mathematics Institute, University of Warwick, Coventry, CV4 7AL, UK
E-mail address: `a.m.stuart@warwick.ac.uk`


eff
0.0
0.1
0.2
0.3
0.4
0.651
0.0
0.2
0.4
0.6
a


abs
0.000
0.005
0.010
0.015
0.020
0.025
0.030
0.357
0.501
0.603
0.679
0.74
0.845
0.963
x
0.00
0.05
0.10
0.15
0.357
0.501
0.603
0.679
0.74
0.845
0.963
xx
0.00
0.05
0.10
0.15
0.357
0.501
0.603
0.679
0.74
0.845
0.963
xxx
0.0
0.5
1.0
1.5
0.357
0.501
0.603
0.679
0.74
0.845
0.963