Biometrika (2016), **103**, 1, *pp*. 1–12
Printed in Great Britain Advance Access publication on 31 July 2016

Supplementary material for Kinetic energy choice in Hamiltonian/hybrid Monte Carlo

A. PROOFS OF PROPOSITIONS

Proof of proposition 1. (i) In Lemma B1 we show that $\int e^{\gamma\|x\|-U(x)}dx = \infty$ for any $\gamma > 0$, and in Lemma B2 that for any $\eta > 0$ there is an $r < \infty$ such that $P\{x, B_r(x)\} > 1 - \eta$. Theorem 2.2 of Jarner & Tweedie (2003) establishes that if these two conditions hold then the resulting Markov chain cannot be geometrically ergodic.

(ii) Lemmas B4, B5 and B6 show that when (7)-(12) hold, with probability one $\lim_{\|x_0\|\to\infty} \triangle(x_0, p_0) = \infty$, where $\triangle(x_0,p_0) = (\|x_{L\varepsilon}\| + \|p_{L\varepsilon}\|) - (\|x_0\| + \|p_0\|)$. Under (9) this implies that with probability one $\lim_{\|x_0\|\to\infty} \triangle H(x_0, p_0) = \infty$, where $\triangle H(x_0, p_0) = H(x_{L\varepsilon}, p_{L\varepsilon}) - H(x_0, p_0)$. This in turn implies that with probability one $\lim_{\|x_0\|\to\infty} \alpha(x_0, x_{L\varepsilon}) = 0$, which, using Proposition 5.1 of Roberts & Tweedie (1996), establishes the result. □

Proof of proposition 2. For the first part, note that the assumptions imply

$$\|\nabla K \circ \nabla U(x)\| \leq C(A\|x\|^q + B)^{1/q} + D,$$

which implies $\limsup_{\|x\|\to\infty} \|\nabla K \circ \nabla U(x)\|/\|x\| < \infty$ as required. We prove the second part by induction. Precisely, we show that assuming $\|p_{i\varepsilon}\| \leq E_i\|x_{i\varepsilon}\|^q + F_i$ for some $E_i, F_i < \infty$ implies $\|p_{(i+1)\varepsilon}\| \leq E_{i+1}\|x_{i\varepsilon}\|^q + F_{i+1}$ and $\|x_{(i+1)\varepsilon}\| \leq G_i\|x_{i\varepsilon}\| + H_i$ for $E_{i+1}, F_{i+1}, G_i, H_i < \infty$. These in turn imply the result. First note that

$$\begin{aligned}
\|x_{(i+1)\varepsilon} - x_{i\varepsilon}\| &= \varepsilon\|\nabla K\{p_{i\varepsilon} - \frac{\varepsilon}{2}\nabla U(x_{i\varepsilon})\}\| \\
&\leq \varepsilon C\|p_{i\varepsilon} - \frac{\varepsilon}{2}\nabla U(x_{i\varepsilon})\|^{1/q} + \varepsilon D \\
&\leq \varepsilon C\left\{\|p_{i\varepsilon}\| + \frac{\varepsilon}{2}\|\nabla U(x_{i\varepsilon})\|\right\}^{1/q} + \varepsilon D.
\end{aligned}$$

Using $\|\nabla U(x_{i\varepsilon})\| \leq A\|x_{i\varepsilon}\|^q + B$ gives

$$\|x_{(i+1)\varepsilon}\| \leq \|x_{i\varepsilon}\| + \varepsilon C\left\{(E_i + \varepsilon A/2)\|x_{i\varepsilon}\|^q + \varepsilon B/2\right\}^{1/q} + \varepsilon D.$$

Given this we can choose $G_i = \varepsilon C(E_i + \varepsilon A/2 + \varepsilon B/2)^{1/q} + 1$ and $H_i = \varepsilon C(E_i + \varepsilon A/2 + \varepsilon B/2)^{1/q} + \varepsilon D$ to see that

$$\|x_{(i+1)\varepsilon}\| \leq G_i\|x_{i\varepsilon}\| + H_i.$$

Iterating gives

$$\|x_{(i+1)\varepsilon}\| \leq \mathbf{G}_L\|x_0\| + \mathbf{H}_L,$$

where $\mathbf{G}_L = G_{L-1}G_{L-2}...G_0$ and $\mathbf{H}_L = H_{L-1} + G_{L-1}H_{L-2} + G_{L-1}G_{L-2}H_{L-3} + ... + G_{L-1}...G_1H_0$. Next recall that

$$\begin{aligned}
\|p_{(i+1)\varepsilon} - p_{i\varepsilon}\| &= \frac{\varepsilon}{2}\|\nabla U(x_{i\varepsilon}) + \nabla U(x_{(i+1)\varepsilon})\| \\
&\leq \frac{\varepsilon}{2}\left\{\|\nabla U(x_{i\varepsilon})\| + \|\nabla U(x_{(i+1)\varepsilon})\|\right\} \\
&\leq \frac{\varepsilon}{2}\left(A\|x_{i\varepsilon}\|^q + A\|x_{(i+1)\varepsilon}\|^q + 2B\right) \\
&\leq \frac{\varepsilon}{2}\left\{A\|x_{i\varepsilon}\|^q + A(G_i\|x_{i\varepsilon}\| + H_i)^q + 2B\right\} \\
&\leq \frac{\varepsilon}{2}\left[A\left\{1 + (G_i + H_i)^q\right\}\|x_{i\varepsilon}\|^q + A(G_i + H_i)^q + 2B\right].
\end{aligned}$$

Combining with the assumption that $\|p_{i\varepsilon}\| \leq E_i\|x_{i\varepsilon}\|^q + F_i$, gives

$$\|p_{(i+1)\varepsilon}\| \leq \left[E_i + \frac{\varepsilon}{2}A\left\{1 + (G_i + H_i)^q\right\}\right]\|x_{i\varepsilon}\|^q + \frac{\varepsilon}{2}\left\{A(G_i + H_i)^q + 2B\right\} + F_i.$$

Setting $E_{i+1} = [E_i + \varepsilon A\{1 + (G_i + H_i)^q/2\}]$ and $F_{i+1} = \varepsilon\{A(G_i + H_i)^q + 2B\}/2 + F_i$ then gives $\|p_{(i+1)\varepsilon}\| \leq E_{i+1}\|x_{i\varepsilon}\|^q + F_{i+1}$. Iterating then gives $\|p_{L\varepsilon}\| \leq E_L\|x_0\|^q + F_L$. Recalling that $\|p_0\| \leq E_0\|x_0\|^q + F_0$ by assumption completes the proof. $\square$

Proof of proposition 3. Consider the event $B = \{4\|p_0\| \leq \varepsilon\|\nabla U(x_0)\|\}$, and note that $\lim_{\|x\|\to\infty} \mathrm{pr}(B) = 1$. We use the facts that $\|x_{L\varepsilon} - x_0\| \leq \sum_{i=1}^{L-1}\|x_{(i+1)\varepsilon} - x_{i\varepsilon}\|$, and that for any $i \in \{0, ..., L-1\}$

$$\|x_{(i+1)\varepsilon} - x_{i\varepsilon}\| = \varepsilon \|\nabla K(p_{\frac{2i+1}{2}\varepsilon})\|. \tag{A1}$$

Taking $i = 0$ gives

$$\|x_\varepsilon - x_0\| = \varepsilon \|\nabla K \{p_0 - \varepsilon \nabla U(x_0)/2\} \|.$$

Since $4\|p_0 - \varepsilon\nabla U(x_0)/2\| \geq \varepsilon \|\nabla U(x_0)\|$ under B, it follows from the fact that $\pi(\cdot)$ is light-tailed and $\nu(\cdot)$ heavy-tailed that for every $\delta > 0$ there is an $M < \infty$ such that whenever $\|x_0\| > M$ then $\|\nabla K(p_{\varepsilon/2})\| < \delta/\varepsilon$. Thus $\|x_\varepsilon - x_0\|$ can be made arbitrarily small by choosing an x_0 with large enough norm.

Recall that $\nabla U(x)$ is continuous by assumption. It follows from the preceding argument that for any $\gamma_1 > 0$ we can choose an x_0 with large enough norm that $\|\nabla U(x_\varepsilon) - \nabla U(x_0)\| < \gamma_1$ under B.

To complete the proof we show that if $\sum_{j=1}^i \|x_{j\varepsilon} - x_{(j-1)\varepsilon}\| < \delta/2$ then $\|x_{(i+1)\varepsilon} - x_{i\varepsilon}\| \leq \delta/2$ under B. Combining this with the previous paragraphs establishes that for any $\delta > 0$ then there is an x_0 with large enough norm that $\|x_{L\varepsilon} - x_0\| < \delta$ if event B holds, establishing the result.

From equation (A1) the key factor in controlling $\|x_{(i+1)\varepsilon} - x_{i\varepsilon}\|$ is $\|p_{(i+1/2)\varepsilon}\|$, which can be lower bounded using

$$\|p_{(i+1/2)\varepsilon}\| \geq \frac{2i+1}{2}\varepsilon\|\nabla U(x_0)\| + \varepsilon\sum_{j=1}^{i} \|\nabla U(x_{j\varepsilon}) - \nabla U(x_0)\| - \|p_0\| \tag{A2}$$

If for any $\delta > 0$ we can choose an x_0 with large enough norm that $\sum_{j=1}^i \|x_{j\varepsilon} - x_{(j-1)\varepsilon}\| < \delta/2$ then $\sum_{j=1}^i \|\nabla U(x_{j\varepsilon}) - \nabla U(x_0)\|$ can be made arbitrarily small through the same continuity argument made above. Thus, under B it holds that $\|p_{(i+1/2)\varepsilon}\| \geq i\varepsilon\|\nabla U(x_0)\|$, from which it follows that $\|x_{(i+1)\varepsilon} - x_{i\varepsilon}\|$ can be made arbitrarily small by choosing $\|x_0\|$ large enough. $\square$

Proof of proposition 4. It is shown in chapter 16 of Meyn & Tweedie (1993) that a geometric convergence bound is equivalent to the drift condition $\int V(y)P(x, dy) \leq \lambda V(x)$ whenever x is outside some small set C, where $\lambda < 1$. Lemma B7 establishes that if (14) holds then any small set must be bounded.

Hence if a geometric bound holds here then

$$\limsup_{\|x\|\to\infty} \frac{\int V(y)P(x,dy)}{V(x)} < 1. \tag{A3}$$

For any $\delta > 0$ we can write

$$\begin{aligned}\int V(y)P(x,dy) &= \int_{B_\delta(x)} V(y)P(x,dy) + \int_{B_\delta^c(x)} V(y)P(x,dy),\\ &\geq \int_{B_\delta(x)} V(y)P(x,dy) + \epsilon,\end{aligned}$$

where $\epsilon = P\{x, B_\delta^c(x)\}$. If (i) holds then we can choose a $\delta < \delta'$, so that

$$\int_{\mathcal{B}_\delta(x)} e^{\log V(y) - \log V(x)} P(x,dy) + \epsilon \geq \int_{\mathcal{B}_\delta(x)} e^{-\epsilon'} P(x,dy) + \epsilon = e^{-\epsilon'}(1-\epsilon) + \epsilon.$$

Noting that both ϵ and ϵ' can be made arbitrarily small as $\|x\| \to \infty$, this expression tends to 1 in the same limit, proving the result. If (ii) holds, note that $\liminf_{\|x\|\to\infty} V(x)e^{-s\|x\|} = c$ implies that $\forall \epsilon' > 0$ there is an $M < \infty$ such that $V(x)e^{-s\|x\|} \geq c - \epsilon'$ whenever $\|x\| \geq M$. This means that when $\|x\| > M$

$$\int_{\mathcal{B}_\delta(x)} V(y)P(x,dy) + \epsilon \geq (c - \epsilon') \int_{\mathcal{B}_\delta(x)} e^{-s\|y\|} P(x,dy) + \epsilon.$$

Condition (ii) also implies that for all $\epsilon' > 0$, there is a sequence $\{x_i\}_{i\geq 1}$ for which $\|x_i\| \to \infty$ as $i \to \infty$ such that whenever $i \geq N$ for some $N < \infty$ then $\|x_i\| > M$ and the condition $V(x_i)e^{-s\|x_i\|} \leq c + \epsilon'$ holds. Combining gives that for all $i \geq N$

$$\int \frac{V(y)}{V(x_i)} P(x_i, dy) \geq \frac{(c-\epsilon')}{(c+\epsilon')} e^{-s\delta}(1-\epsilon) + \frac{\epsilon}{V(x_i)}.$$

Since ϵ, ϵ' and δ can all be made arbitrarily small and $V(x_i) \to \infty$ as $\|x_i\| \to \infty$, then this proves the result. $\square$

Proof of proposition 5. Assume $H(x_0, p_0) = E$ and $x_0 = 0$, $p_0 = (\beta E)^{\frac{1}{\beta}}$. Take $4T$ to be the period length, and note that by the symmetry of the Hamiltonian in question this implies that $p_T = 0$ and $x_T = (\alpha E)^{\frac{1}{\alpha}}$. Then

$$\mathcal{P}(E) = 4\int_0^T dt = 4\int_0^{x_T} \frac{dt}{dx_t} dx_t = 4\int_0^{x_T} p_t^{1-\beta} dx_t.$$

Setting $b = (1-\beta)/\beta$, $c_\beta = \beta^b$ and noting that $p_t^{1-\beta} = c_\beta(E - \alpha^{-1}x_t^\alpha)^b$ for $t \in [0, T]$, then the expression can be written

$$\mathcal{P}(E) = 4c_\beta \int_0^{x_T} \left(E - \alpha^{-1}x_t^\alpha\right)^b dx_t.$$

Applying the change of variables $y_t = (\alpha E)^{-1/\alpha}x_t$ and setting $c_\alpha = \alpha^{1/\alpha}$ gives

$$P(E) = 4c_\beta c_\alpha E^{b+1/\alpha} \int_0^1 (1 - y_t^\alpha)^b dy_t,$$

where . Now, we have that $\mathcal{P}(E) = f(E^\eta)$, for some function f, where

$$\eta = \frac{1-\beta}{\beta} + \frac{1}{\alpha} = \frac{1-(\beta-1)(\alpha-1)}{\alpha\beta}.$$

Setting $\alpha = 1+\gamma$ and $\beta = 1+\gamma^{-1}$ for some $\gamma > 0$ gives

$$\eta = \frac{1-\gamma\gamma^{-1}}{(1+\gamma)(1+\gamma^{-1})} = 0,$$

as required. $\square$

Proof of proposition 6. Set $\gamma(x) = \min\left[\frac{\varepsilon}{4}\|\nabla U(x)\|, \left\|\nabla^2 K\left\{\frac{\varepsilon}{4}\nabla U(x)\right\}\right\|^{-1/2}\right]$, and note that $\lim_{\|x\|\to\infty} \mathrm{pr}_\nu\left\{\|p\| \leq \gamma(x)\right\} = 1$. For $\|p\| \leq \gamma(x)$, as a direct consequence of the mean value inequality (Dieudonné, 1961)

$$\left\|\nabla K\left\{\frac{\varepsilon}{2}\nabla U(x) - p\right\} - \nabla K\left\{\frac{\varepsilon}{2}\nabla U(x)\right\}\right\| \leq M(x)\|p\|,$$

where $M(x) = \sup_{\{4\|p\|\geq\varepsilon\|\nabla U(x)\|\}} \|\nabla^2 K(p)\|$. As the right-hand side tends to 0 as $\|x\| \to \infty$, then the result follows. $\square$

B. Technical results

Lemma B1. *If $\pi(\cdot)$ is heavy-tailed then for every $\gamma > 0$*

$$\int e^{\gamma\|x\| - U(x)} dx = \infty.$$

Proof. Choose $\delta < \gamma$. Let B be a Euclidean ball centred at the origin such that $\|\nabla U(x)\| \leq \delta$ whenever $x \notin B$. By continuity of $U(x)$, there is an $M < \infty$ such that $U(x) \leq M$ for all $x \in \partial B$. Then for all $x \notin B$ the integrand is bounded below by $e^{(\gamma-\delta)\|x\|-M}$, which diverges uniformly and hence is not integrable. □

LEMMA B2. *If $\pi(x)$ is heavy-tailed then for any $\eta > 0$ there is an $r < \infty$ such that*

$$P\{x, B_r(x)\} > 1 - \eta.$$

Proof. We need to show that $Q\{x, B_r(x)\} > 1 - \eta$, for any x. After one leapfrog step we have

$$x_\varepsilon = x_0 + \varepsilon \nabla K \left\{ p_0 - \frac{\varepsilon}{2} \nabla U(x_0) \right\},$$
$$p_\varepsilon = p_0 - \frac{\varepsilon}{2} \nabla U(x_0) - \frac{\varepsilon}{2} \nabla U(x_\varepsilon).$$

Write $\|x\|_\infty$ for the supremum norm, and note that by equivalence of norms in finite dimensions we can write $\|x\|_\infty \le C\|x\|$ for all x, for some $C < \infty$. We have that $\nabla U(x) \in C_0(\mathbb{R}^d)$, which implies $\|\nabla U(x)\| < M/C$ for some $M < \infty$ which does not depend on x, so that $\|\nabla U(x)\|_\infty < M$. The class of distributions for $\{p_0 - \varepsilon \nabla U(x_0)/2\}$ is therefore tight. Now recall that if f is a locally bounded function, and $\mathcal{F}$ a tight family of probability measures, then the resulting family of probability measures induced by pushing forward each element of $\mathcal{F}$ through f is also tight. So since ∇K is continuous and hence locally bounded, the result follows. □

LEMMA B3. *If* (7) *and* (10) *hold then*

$$\lim_{\|x\| \to \infty} \frac{\|\nabla K\{\frac{\varepsilon}{4} \nabla U(x)\}\|}{\|x\|} = \infty. \tag{B1}$$

Proof. First we re-write the expression

$$\lim_{\|x\| \to \infty} \frac{\|\nabla K\{\frac{\varepsilon}{4} \nabla U(x)\}\|}{\|x\|} = \lim_{\|x\| \to \infty} \frac{\|\nabla K\{\frac{\varepsilon}{4} \nabla U(x)\}\|}{\|\nabla K \circ \nabla U(x)\|} \frac{\|\nabla K \circ \nabla U(x)\|}{\|x\|},$$

Now, (10) implies that the first term will be bounded below by a finite positive constant, while (7) ensures that the second will have an infinite limit, proving the result. □

LEMMA B4. *If $\pi(\cdot)$ is light-tailed,* (7) *and either of* (11) *or* (12) *hold and $\|p_0\| \le \frac{\varepsilon}{4} \min\{\|\nabla U(x_0)\|, \|\nabla U(x_0)\|_\infty\}$ then there is a $\gamma_M < \infty$ such that, provided $\|x_0\| \ge \gamma_M$, it holds that $\|x_\varepsilon\| \ge M\|x_0\|$, for any $M < \infty$.*

Proof. Note

$$\|x_\varepsilon\| = \|x_0 + \varepsilon \nabla K \left\{p_0 - \frac{\varepsilon}{2}\nabla U(x_0)\right\}\| \geq \varepsilon \|\nabla K \left\{p_0 - \frac{\varepsilon}{2}\nabla U(x_0)\right\}\| - \|x_0\|.$$

It is therefore sufficient to show that for any $M < \infty$ we can choose an $\|x_0\|$ large enough that

$$\|\nabla K \left\{p_0 - \frac{\varepsilon}{2}\nabla U(x_0)\right\}\| \geq \frac{(M+1)}{\varepsilon}\|x_0\|.$$

Under (11), note that

$$\|p_0 - \frac{\varepsilon}{2}\nabla U(x_0)\| \geq \frac{\varepsilon}{2}\|\nabla U(x_0)\| - \|p_0\| \geq \frac{\varepsilon}{4}\|\nabla U(x_0)\|,$$

which implies

$$\|\nabla K \left\{p_0 - \frac{\varepsilon}{2}\nabla U(x_0)\right\}\| \geq \|\nabla K \left\{\frac{\varepsilon}{4}\nabla U(x_0)\right\}\|.$$

By (B1), therefore, if $\|x_0\|$ is chosen to be large enough then this can be made $\geq (M+1)\|x_0\|/\varepsilon$, for any finite M, proving the result.

Under (12), recall that there exists global constants $C, c > 0$ such that $C\|\nabla U(x)\| \geq \|\nabla U(x)\|_\infty \geq c\|\nabla U(x)\|$ for all $x \in \mathbb{R}^d$. It suffices in this setting therefore to show that we can choose an $\|x_0\|$ large enough that

$$\|\nabla K \left\{p_0 - \frac{\varepsilon}{2}\nabla U(x_0)\right\}\|_\infty \geq \frac{C(M+1)}{\varepsilon}\|x_0\|.$$

We have

$$\|\nabla K \left\{p_0 - \frac{\varepsilon}{2}\nabla U(x_0)\right\}\|_\infty = \max_j |k'\{p_0(j) - \partial_j U(x_0)\}|.$$

Write i^* and j^* to denote the indices for which $\|p_0 - \nabla U(x_0)\|_\infty = |p_0(i^*) - \partial_{i^*}U(x_0)|$ and $\|\nabla U(x_0)\|_\infty = |\partial_{j^*}U(x_0)|$. We have:

$$\begin{aligned}
\|p_0 - \frac{\varepsilon}{2}\nabla U(x_0)\|_\infty &= |p_0(i^*) - \partial_{i^*}U(x_0)| \\
&\geq |p_0(j^*) - \partial_{j^*}U(x_0)| \\
&\geq \frac{\varepsilon}{2}|\partial_{j^*}U(x_0)| - |p_0(j^*)| \\
&\geq \frac{\varepsilon}{4}|\partial_{j^*}U(x_0)|.
\end{aligned}$$

Now, $|p_0(i^*) - (\varepsilon/2)\partial_{i^*}U(x_0)| \geq (\varepsilon/4)|\partial_{j^*}U(x_0)|$ implies that $|k'\{p_0(i^*) - (\varepsilon/2)\partial_{i^*}U(x_0)\}| \geq |k'\{(\varepsilon/4)\partial_{j^*}U(x_0)\}| = \|\nabla K\{(\varepsilon/4)\nabla U(x_0)\}\|_\infty$. Using the global bounds then we see that for any $M < \infty$ we can choose an $\|x_0\|$ large enough that

$$\frac{\|\nabla K\{(\varepsilon/4)\nabla U(x_0)\}\|_\infty}{\|x_0\|} \geq \frac{C(M+1)}{\varepsilon},$$

establishing the result. □

LEMMA B5. *If $\pi(\cdot)$ is light-tailed and* (7)-(10) *and one of* (11) *and* (12) *hold, and provided that for any fixed $i \geq 0$*

(i) $\|x_0\| \geq \gamma_M$ for some $\gamma_M < \infty$,

(ii) $\|p_0\| \leq (\varepsilon/4)\min\{\|\nabla U(x_0)\|, \|\nabla U(x_0)\|_\infty\}$,

(iii) M is large enough that $\phi(M) \geq 7/3$ with ϕ as in (8)*,*

(iv) $\|x_{j\varepsilon}\| \geq M\|x_{(j-1)\varepsilon}\|$ for all $j \leq i$,

it holds for any finite $M < \infty$ that $\|x_{(i+1)\varepsilon}\| \geq M\|x_{i\varepsilon}\|$.

Proof. We first show

$$\|p_{(i+\frac{1}{2})\varepsilon}\| \geq \frac{\varepsilon}{4}\|\nabla U(x_{i\varepsilon})\|. \tag{B2}$$

To show (B2), first note by iterating (4) and noting $p_{(i+1/2)\varepsilon} = p_{i\varepsilon} - \varepsilon\nabla U(x_{i\varepsilon})/2$ that

$$\begin{aligned}\|p_{(i+\frac{1}{2})\varepsilon}\| &= \|p_0 - \frac{\varepsilon}{2}\nabla U(x_0) - \varepsilon\sum_{j=1}^{i}\nabla U(x_{j\varepsilon})\| \\ &\geq \varepsilon\|\nabla U(x_{i\varepsilon})\| - \frac{\varepsilon}{2}\|\nabla U(x_0)\| - \varepsilon\sum_{j=1}^{i-1}\|\nabla U(x_{j\varepsilon})\| - \|p_0\|.\end{aligned}$$

Using the stated assumption that $\|p_0\| \leq (\varepsilon/4)\|\nabla U(x_0)\|$ then gives

$$\|p_{(i+\frac{1}{2})\varepsilon}\| \geq \varepsilon\left\{\|\nabla U(x_{i\varepsilon})\| - \sum_{j=0}^{i-1}\|\nabla U(x_{j\varepsilon})\|\right\}.$$

Now, (8) implies that if $\|x_{(j-1)\varepsilon}\| \leq M^{-1}\|x_{j\varepsilon}\|$, $\|\nabla U(x_{(j-1)\varepsilon})\| \leq \phi(M)^{-1}\|\nabla U(x_{j\varepsilon})\|$. Substituting into the above expression gives

$$\|p_{(i+\frac{1}{2})\varepsilon}\| \geq \varepsilon\left\{1 - \sum_{j=0}^{i-1}\phi(M)^{j-i}\right\}\|\nabla U(x_{i\varepsilon})\|.$$

To finish the argument we need to therefore show that $\{1 - \sum_{j=0}^{i-1} \phi(M)^{j-i}\} \geq 1/4$. First note that

$$\sum_{j=0}^{i-1} \phi(M)^{j-i} = \phi(M)^{-i} + \phi(M)^{-(i-1)} + ... + \phi(M)^{-1} = \sum_{j=1}^{i} \phi(M)^{-j}.$$

Using simple geometric series identities gives

$$\begin{aligned}
\sum_{j=1}^{i} \phi(M)^{-j} &= \frac{1 - \phi(M)^{-(i+1)}}{1 - \phi(M)^{-1}} - 1 \\
&= \frac{\phi(M) - \phi(M)^{-i}}{\phi(M) - 1} - \frac{\phi(M) - 1}{\phi(M) - 1} \\
&= \frac{1 - \phi(M)^{-i}}{\phi(M) - 1}.
\end{aligned}$$

Hence $\{1 - \sum_{j=0}^{i-1} \phi(M)^{j-i}\} \geq 1/4$ if

$$1 - \frac{1 - \phi(M)^{-i}}{\phi(M) - 1} \geq 1/4,$$

which will hold if $\phi(M) \geq 7/3$ for any $i \geq 1$. The stated assumption that $\phi(M) \geq 7/3$ therefore establishes (B2).

If (11) holds, then (B2) can be combined with (B1) directly to show that for any $M < \infty$ and $\varepsilon \in (0, \infty)$, if $\|x_{i\varepsilon}\| \geq \gamma_M$ for some suitably chosen $\gamma_M < \infty$ it will hold that

$$\|\nabla K\{p_{(i+\frac{1}{2})\varepsilon}\}\| \geq \|\nabla K\left\{\frac{\varepsilon}{4}\nabla U(x_{i\varepsilon})\right\}\| \geq \frac{(M+1)}{\varepsilon}\|x_{i\varepsilon}\|.$$

If (12) holds instead of (11), then the same result can be established using a similar argument to that given in the last paragraph of the proof of Lemma B4. Hence, provided that $\|x_0\| \geq \gamma_M$, for all $i \geq 0$ it holds that

$$\begin{aligned}
\|x_{(i+1)\varepsilon}\| &= \|x_{i\varepsilon} + \varepsilon \nabla K\{p_{(i+\frac{1}{2})\varepsilon}\}\| \\
&\geq \varepsilon\|\nabla K\{p_{(i+\frac{1}{2})\varepsilon}\}\| - \|x_{i\varepsilon}\| \\
&\geq M\|x_{i\varepsilon}\|,
\end{aligned}$$

which proves the result.

LEMMA B6. *Under the same conditions as Lemma B5 and provided M is such that $\phi(M) \geq 5$, then $\|p_{L\varepsilon}\| \geq \|p_0\|$ for any $L \geq 1$.*

Proof. We have

$$\begin{aligned}\|p_{L\varepsilon}\| &= \|p_0 - \frac{\varepsilon}{2}\{\nabla U(x_0) + \nabla U(x_{L\varepsilon})\} - \varepsilon \sum_{i=1}^{L-1} \nabla U(x_{i\varepsilon})\| \\ &\geq \frac{\varepsilon}{2}\left\{\|\nabla U(x_{L\varepsilon})\| - 2\sum_{i=0}^{L-1}\|\nabla U(x_{i\varepsilon})\|\right\},\end{aligned}$$

by recalling that as in Lemma B5 $\|p_0\| \leq (\varepsilon/4)\|\nabla U(x_0)\|$. Using (8) and the stated assumptions we have for any $i \leq L$ that

$$\|\nabla U(x_{i\varepsilon})\| \leq \phi(M)^{i-L}\|\nabla U(x_{L\varepsilon})\|,$$

which implies

$$\|p_{L\varepsilon}\| \geq \frac{\varepsilon}{2}\left\{1 - 2\sum_{i=0}^{L-1}\phi(M)^{i-L}\right\}\|\nabla U(x_{L\varepsilon})\|$$

The stated assumptions $\|p_0\| \leq (\varepsilon/4)\|\nabla U(x_0)\|$ and $\phi(M)^L\|\nabla U(x_0)\| \leq \|\nabla U(x_{L\varepsilon})\|$ lead to the bound

$$\|\nabla U(x_{L\varepsilon})\| \geq \phi(M)^L\frac{4}{\varepsilon}\|p_0\|,$$

which, when combined with the above inequality, give

$$\|p_{L\varepsilon}\| \geq 2\left\{1 - 2\sum_{i=0}^{L-1}\phi(M)^{i-L}\right\}\phi(M)^L\|p_0\|.$$

As shown in the proof of Lemma B5 $\sum_{i=0}^{L-1}\phi(M)^{i-L} = \{1 - \phi(M)^{-L}\}/\{\phi(M) - 1\}$. Hence the result is proven if

$$2\left\{1 - 2\frac{1 - \phi(M)^{-L}}{\phi(M) - 1}\right\}\phi(M)^L \geq 1,$$

which will indeed be true under the stated assumption that $\phi(M) > 5$. □

LEMMA B7. *If* (14) *holds for a Hamiltonian Monte Carlo method then any small set must be bounded.*

Proof. Since $\nabla K(p)$ and $\nabla U(x)$ are continuous and $\nabla K \circ \nabla U(x)$ is vanishing at infinity, then $\nabla K \circ \nabla U(x)$ is also bounded, implying that the collection of Hamiltonian Monte Carlo increments $\{P(x, \cdot) - x\}$ is uniformly tight, using a similar argument to that employed in the proof of Lemma B2. Hence, any small set must be bounded, following Lemma 2.2 of Jarner & Hansen (2000). □

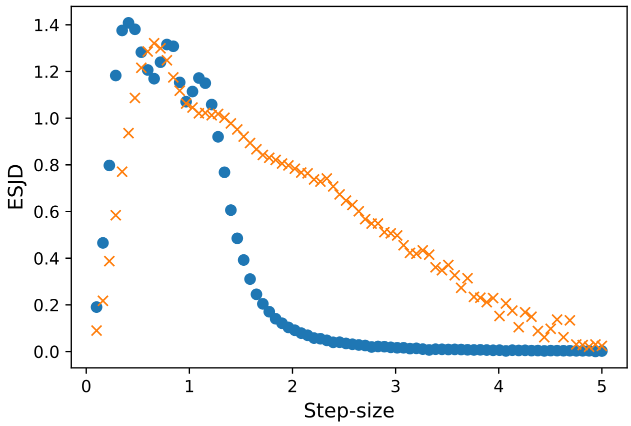

Fig. 1. Empirical expected squared jump distance versus step-size for the relativistic kinetic energy, denoted by orange crosses, and quadratic kinetic energy, denoted by blue dots.

C. Numerical example of the efficiency-robustness trade-off

We consider the distribution $\pi(x) \propto \exp\left(-\beta^{-1}|x|^{\beta}\right)$ with $\beta = 1.5$, and compare the quadratic and relativistic kinetic energy choices. We test 80 evenly spaced choices of step-size ε spanning from 0.1 to 5, and in each case begin the sampler at equilibrium and compute the empirical expected squared jump distance from a chain of length 200,000, with the number of leapfrog steps randomly selected uniformly between 1 and 5 at each iteration. The results are shown in Figure 1. As can be seen, the quadratic choice leads to a higher optimal value, but when step-sizes are chosen to be too large the jump distance drops quickly. The relativistic choice, by comparison, exhibits a larger degree of robustness to bigger step-sizes.

References

Dieudonné, J. (1961). *Foundations of modern analysis*. Academic Press New York.

Jarner, S. F. & Hansen, E. (2000). Geometric ergodicity of Metropolis algorithms. *Stochastic processes and their applications* **85**, 341–361.

Jarner, S. F. & Tweedie, R. L. (2003). Necessary conditions for geometric and polynomial ergodicity of random-walk-type Markov chains. *Bernoulli* **9**, 559–578.

Meyn, S. P. & Tweedie, R. L. (1993). *Markov chains and stochastic stability*. London: Springer–Verlag.

ROBERTS, G. O. & TWEEDIE, R. L. (1996). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika* , 95–110.

[*Received April* 2012. *Revised October* 2015]