Supplementary Material for "Mixing of Hamiltonian Monte Carlo on strongly log-concave distributions 2: Numerical integrators"

Oren Mangoubi [2] and Aaron Smith [3]

7 Previous Results

In this section, we recall several results about the idealized HMC dynamics from the companion paper [33]. Consider two solutions $(q_t^{(1)}, p_t^{(1)})$ and $(q_t^{(2)}, p_t^{(2)})$ of Equation (2.2). Denote by $\tilde{q}_t := q_t^{(2)} - q_t^{(1)}$ and $\tilde{p}_t := p_t^{(2)} - p_t^{(1)}$ the differences between these solutions, and denote by $\hat{q}_t := \|\tilde{q}_t\|_2$ and $\hat{p}_t := \|\tilde{p}_t\|_2$ the magnitudes of these differences. The following is Theorem 3 of [33] (or, equivalently, Theorem 6 in the combined arXiv preprint [34]):

Theorem 3 (Contraction For Hamiltonian Mechanics with Convex Potentials)**.** *For* $0 \leq T \leq \frac{1}{2\sqrt{2}} \frac{\sqrt{m_2}}{M_2}$,

$$\hat{q}_T \leq \left[1 - \frac{1}{8}(\sqrt{m}_2 T)^2\right] \times \hat{q}_0. \tag{7.1}$$

We will also require the following intermediate bound on the stability of solutions to Hamilton's equations, which is the first part of Lemma 2.2 of [33, 34]:

Lemma 7.1. *For all* $t \geq 0$,

$$\begin{aligned} \hat{q}_t &\leq k_1 e^{t\sqrt{M_2}} + k_2 e^{-t\sqrt{M_2}} \\ \hat{p}_t &\leq k_1 \sqrt{M_2} e^{t\sqrt{M_2}} - k_2 \sqrt{M_2} e^{-t\sqrt{M_2}}, \end{aligned} \tag{7.2}$$

where $k_1 = \frac{1}{2}(\hat{q}_0 + \frac{\hat{p}_0}{\sqrt{M_2}})$, $k_2 = \frac{1}{2}(\hat{q}_0 - \frac{\hat{p}_0}{\sqrt{M_2}})$.

8 Leapfrog Integrator

In this section, we bound the error of the leapfrog integrator when it is used to approximate the continuous HMC dynamics. Recall the notation $q_\theta^\star$, $p_\theta^\star$ from Equation (2.4).

Lemma 8.1. *(Leapfrog method error)*

Fix $\theta > 0$. *We have*

$$\|q_\theta^\star(\mathbf{q}, \mathbf{p}) - q_\theta(\mathbf{q}, \mathbf{p})\| \leq \theta^2 \frac{M_2}{\sqrt{m_2}} \sqrt{U(\mathbf{q}) + \frac{1}{2}\|\mathbf{p}\|^2} \tag{8.1}$$

and

$$\|p_\theta^\star(\mathbf{q}, \mathbf{p}) - p_\theta(\mathbf{q}, \mathbf{p})\| \leq 3\theta^2 M_2 \sqrt{U(\mathbf{q}) + \frac{1}{2}\|\mathbf{p}\|^2}. \tag{8.2}$$

Proof. For all $t \geq 0$, we have by conservation of energy that

$$U(q_t(\mathbf{q}, \mathbf{p})) + \frac{1}{2}\|p_t(\mathbf{q}, \mathbf{p})\|^2 = U(\mathbf{q}) + \frac{1}{2}\|\mathbf{p}\|^2$$

and so we also have

$$\|q_t(\mathbf{q}, \mathbf{p})\| \leq \frac{1}{\sqrt{m_2}} \sqrt{U(\mathbf{q}) + \frac{1}{2}\|\mathbf{p}\|^2}. \tag{8.3}$$

[2] EPFL
[3] University of Ottawa

This implies

$$\|U'(q_t(\mathbf{q},\mathbf{p}))\| \le \left\| \int_0^{\|q_t(\mathbf{q},\mathbf{p})\|} D_{\frac{q_t(\mathbf{q},\mathbf{p})}{\|q_t(\mathbf{q},\mathbf{p})\|}} U'\big|_{\ell_s(0,q_t(\mathbf{q},\mathbf{p}))} \mathrm{d}s \right\| \tag{8.4}$$

$$\le \int_0^{\|q_t(\mathbf{q},\mathbf{p})\|} M_2 \mathrm{d}s = M_2 \|q_t(\mathbf{q},\mathbf{p})\|$$

$$\overset{\text{Eq. (8.3)}}{\le} \frac{M_2}{\sqrt{m_2}} \sqrt{U(\mathbf{q}) + \frac{1}{2}\|\mathbf{p}\|^2} \qquad \forall t \ge 0.$$

Therefore,

$$\|\mathbf{p} - p_t(\mathbf{q},\mathbf{p})\| = \left\| \int_0^t U'(q_s(\mathbf{q},\mathbf{p}))\mathrm{d}s \right\| \le \int_0^t \|U'(q_s(\mathbf{q},\mathbf{p}))\| \mathrm{d}s \le t \times \frac{M_2}{\sqrt{m_2}} \sqrt{U(\mathbf{q}) + \frac{1}{2}\|\mathbf{p}\|^2},$$

for all $t \ge 0$, and so

$$\|\mathbf{q} + \mathbf{p}\theta - q_\theta(\mathbf{q},\mathbf{p})\| = \left\| \int_0^\theta (\mathbf{p} - p_t(\mathbf{q},\mathbf{p}))\mathrm{d}t \right\| \le \int_0^\theta \|\mathbf{p} - p_t(\mathbf{q},\mathbf{p})\| \mathrm{d}t \tag{8.5}$$

$$\le \int_0^\theta t \times \frac{M_2}{\sqrt{m_2}} \sqrt{U(\mathbf{q}) + \frac{1}{2}\|\mathbf{p}\|^2} \mathrm{d}t = \frac{1}{2}\theta^2 \frac{M_2}{\sqrt{m_2}} \sqrt{U(\mathbf{q}) + \frac{1}{2}\|\mathbf{p}\|^2}.$$

Continuing,

$$\|\mathbf{q} + \mathbf{p}\theta - \frac{1}{2}\theta^2 U'(\mathbf{q}) - q_\theta(\mathbf{q},\mathbf{p})\| \le \|\mathbf{q} + \mathbf{p}\theta - q_\theta(\mathbf{q},\mathbf{p})\| + \frac{1}{2}\theta^2 \|U'(\mathbf{q})\| \tag{8.6}$$

$$\overset{\text{Eq. (8.4),(8.5)}}{\le} 2 \times \frac{1}{2}\theta^2 \frac{M_2}{\sqrt{m_2}} \sqrt{U(\mathbf{q}) + \frac{1}{2}\|\mathbf{p}\|^2}.$$

This completes the proof of Inequality (8.1).

We now prove Inequality (8.2). By the conservation of energy bound, $\|p_t(\mathbf{q},\mathbf{p})\| \le \sqrt{2}\sqrt{U(\mathbf{q}) + \frac{1}{2}\|\mathbf{p}\|^2}$ for all $t \ge 0$, so we have

$$\|q_t(\mathbf{q},\mathbf{p}) - \mathbf{q}\| \le t\,\sqrt{2}\sqrt{U(\mathbf{q}) + \frac{1}{2}\|\mathbf{p}\|^2}$$

for all $t \ge 0$. Applying this bound gives

$$\|U'(q_t(\mathbf{q},\mathbf{p})) - U'(\mathbf{q})\| = \left\| \int_0^{\|q_t(\mathbf{q},\mathbf{p}) - \mathbf{q}\|} D_{\frac{q_t(\mathbf{q},\mathbf{p}) - \mathbf{q}}{\|q_t \mathbf{q},\mathbf{p}) - \mathbf{q}\|}} U'\big|_{\ell_s(\mathbf{q},q_t(\mathbf{q},\mathbf{p}))} \mathrm{d}s \right\| \tag{8.7}$$

$$\le \int_0^{\|q_t(\mathbf{q},\mathbf{p}) - \mathbf{q}\|} \left\| D_{\frac{q_t(\mathbf{q},\mathbf{p}) - \mathbf{q}}{\|q_t \mathbf{q},\mathbf{p}) - \mathbf{q}\|}} U'\big|_{\ell_s(\mathbf{q},q_t(\mathbf{q},\mathbf{p}))} \right\| \mathrm{d}s$$

$$\le \int_0^{\|q_t(\mathbf{q},\mathbf{p}) - \mathbf{q}\|} M_2 \mathrm{d}s = M_2 \|q_t(\mathbf{q},\mathbf{p}) - \mathbf{q}\| \le M_2 t\,\sqrt{2}\sqrt{U(\mathbf{q}) + \frac{1}{2}\|\mathbf{p}\|^2}.$$

Applying this bound to the quantity of interest,

$$\|\mathbf{p} + \theta U'(\mathbf{q}) - p_\theta(\mathbf{q},\mathbf{p})\| = \left\| \int_0^\theta (U'(q_t(\mathbf{q},\mathbf{p})) - U'(\mathbf{q}))\mathrm{d}t \right\| \tag{8.8}$$

$$\le \int_0^\theta \|U'(q_t(\mathbf{q},\mathbf{p})) - U'(\mathbf{q})\| \mathrm{d}t$$

$$\le \int_0^\theta M_2 t\,\sqrt{2}\sqrt{U(\mathbf{q}) + \frac{1}{2}\|\mathbf{p}\|^2} \mathrm{d}t = \frac{1}{\sqrt{2}}\theta^2 M_2 \sqrt{U(\mathbf{q}) + \frac{1}{2}\|\mathbf{p}\|^2}.$$

We also have

$$\begin{aligned}\|U'(q_\theta^\star(\mathbf{q},\mathbf{p}))-U'(\mathbf{q})\| &\leq \|U'(q_\theta^\star(\mathbf{q},\mathbf{p}))-U'(q_\theta(\mathbf{q},\mathbf{p}))\| \\ &\qquad\qquad + \|U'(q_\theta(\mathbf{q},\mathbf{p}))-U'(\mathbf{q})\| \\ &\leq M_2\|q_\theta^\star(\mathbf{q},\mathbf{p}))-q_\theta(\mathbf{q},\mathbf{p})\| + \|U'(q_\theta(\mathbf{q},\mathbf{p}))-U'(\mathbf{q})\| \\ &\overset{\text{Eq. (8.6),(8.7)}}{\leq} M_2\left[\theta^2\frac{M_2}{\sqrt{m_2}}\sqrt{U(\mathbf{q})+\frac{1}{2}\|\mathbf{p}\|^2}\right] + M_2\theta\sqrt{2}\sqrt{U(\mathbf{q})+\frac{1}{2}\|\mathbf{p}\|^2} \\ &\leq 4M_2\theta\sqrt{U(\mathbf{q})+\frac{1}{2}\|\mathbf{p}\|^2},\end{aligned} \tag{8.9}$$

where the last inequality holds since $\theta \leq \frac{\sqrt{m_2}}{M_2}$.

Therefore,

$$\begin{aligned}&\left\|\mathbf{p}+\theta U'(\mathbf{q})-\frac{1}{2}\theta^2\frac{U'(q_\theta^\star(\mathbf{q},\mathbf{p}))-U'(\mathbf{q})}{\theta}-p_\theta(\mathbf{q},\mathbf{p})\right\| \\ &\qquad\leq \|\mathbf{p}+\theta U'(\mathbf{q})-p_\theta(\mathbf{q},\mathbf{p})\| + \left\|\frac{1}{2}\theta^2\frac{U'(q_\theta^\star(\mathbf{q},\mathbf{p}))-U'(\mathbf{q})}{\theta}\right\| \\ &\qquad\overset{\text{Eq. (8.8)}}{\leq} \frac{1}{\sqrt{2}}\theta^2 M_2\sqrt{U(\mathbf{q})+\frac{1}{2}\|\mathbf{p}\|^2} + \frac{1}{2}\theta\|U'(q_\theta^\star(\mathbf{q},\mathbf{p}))-U'(\mathrm{q}_j)\| \\ &\qquad\overset{\text{Eq. (8.9)}}{\leq} \frac{1}{\sqrt{2}}\theta^2 M_2\sqrt{U(\mathbf{q})+\frac{1}{2}\|\mathbf{p}\|^2} + 2\theta^2 M_2\sqrt{U(\mathbf{q})+\frac{1}{2}\|\mathbf{p}\|^2} \\ &\qquad\leq 3\theta^2 M_2\sqrt{U(\mathbf{q})+\frac{1}{2}\|\mathbf{p}\|^2}.\end{aligned}$$

This completes the proof of the lemma. □

We find it convenient to define dedicated notation $q_T^{\dagger\theta}: \mathbb{R}^d\times\mathbb{R}^d \to \mathbb{R}^d$ for the global leapfrog integrator as follows (the global leapfrog integrator is used in steps 4-7 in Algorithm 2, where $\mathrm{q}_{\lfloor\frac{T}{\theta}\rfloor} = q_T^{\dagger\theta}(X_i', \mathbf{p}_i)$). This will be used throughout the rest of the paper.

Algorithm 3 Global leapfrog integrator

parameters: Potential U, trajectory time $T>0$, and accuracy $\theta>0$.
input: Initial point $\mathbf{q}\in\mathbb{R}^d$, initial momentum $\mathbf{p}\in\mathbb{R}^d$.
output: $q_T^{\dagger\theta}(\mathbf{q},\mathbf{p})$, $p_T^{\dagger\theta}(\mathbf{q},\mathbf{p})$.

1: Set $q_\dagger^0=\mathbf{q}$ and $p_\dagger^0=\mathbf{p}$.
2: **for** $i=0$ to $\lceil\frac{T}{\theta}-1\rceil$ **do**
3: Compute $q_\dagger^{i+1}=q_\theta^\star(q_\dagger^i,p_\dagger^i)$ and $p_\dagger^{i+1}=p_\theta^\star(q_\dagger^i,p_\dagger^i)$.
4: **end for**
5: Set $q_T^{\dagger\theta}(\mathbf{q},\mathbf{p})=q_\dagger^{i+1}$, $p_T^{\dagger\theta}(\mathbf{q},\mathbf{p})=p_\dagger^{i+1}$.

Lemma 8.2. *Fix parameters $T,\theta>0$ satisfying $7\theta\leq T\leq\frac{1}{2\sqrt{2}}\frac{\sqrt{m_2}}{M_2}$ and $\frac{T}{\theta}\in\mathbb{N}$, and let $q_T^{\dagger\theta}$ be the function defined Algorithm 3 with these parameters. Then we have the bound on position error*

$$\|q_T^{\dagger\theta}(\mathbf{q},\mathbf{p})-q_T(\mathbf{q},\mathbf{p})\| \leq 6\theta\times T\times\frac{M_2}{\sqrt{m_2}}\sqrt{H(\mathbf{q},\mathbf{p})} \tag{8.10}$$

and a bound on energy error

$$|H(q_T^{\dagger\theta}(\mathbf{q},\mathbf{p}),p_T^{\dagger\theta}(\mathbf{q},\mathbf{p}))-H(\mathbf{q},\mathbf{p})| \leq 7\frac{\theta}{T}H(\mathbf{q},\mathbf{p}). \tag{8.11}$$

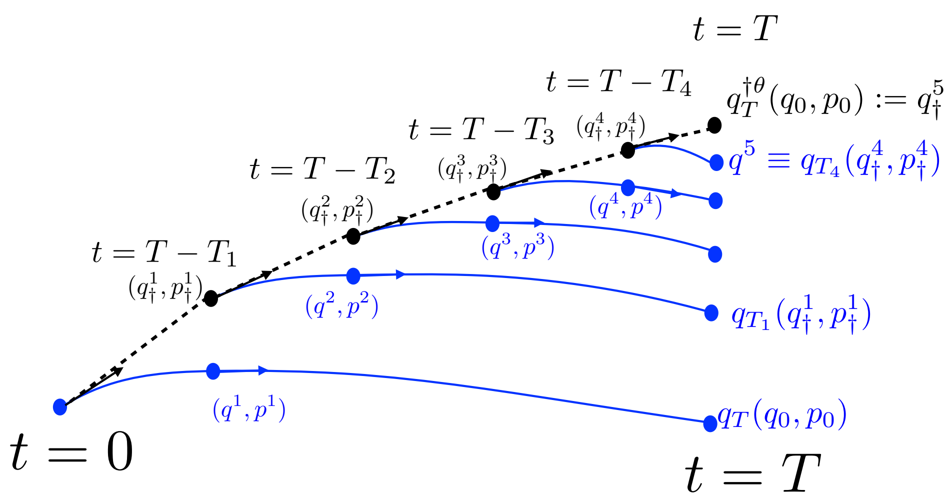

Figure 2: This is an illustration of the proof of Lemma 8.2. Steps taken by the leapfrog integrator are represented by black dashed lines. The true Hamiltonian trajectories are blue curves. Only the Hamiltonian trajectory $q_t(\mathbf{q}, \mathbf{p})$ on the bottom belongs to the idealized HMC Markov chain. We imagine the other Hamiltonian trajectories to help us bound the error. The distance between (q^i, p^i) (blue dot, blue arrow) and $(q_\dagger^i, p_\dagger^i)$ (black dot, black arrow) at each time $t = T - T_i$ are bounded using Lemma 8.1. The distance between any blue dot at $t = T$ to the blue dot directly below it is bounded using Lemma 7.1.

Proof. In this proof we will use the notation of Algorithm 3. Define $q^{i+1} := q_\theta(q_\dagger^i, p_\dagger^i)$ and $p^{i+1} = p_\theta(q_\dagger^i, p_\dagger^i)$ for every $i \in 0, 1, \ldots, \frac{T}{\theta}$. Also set $q^0 = \mathbf{q}$ and $p^0 = \mathbf{p}$. Set $E := H(\mathbf{q}, \mathbf{p}) = U(\mathbf{q}) + \frac{1}{2}\|\mathbf{p}\|^2$.

We will now prove the following claim by induction: For all $0 \le j \le \frac{T}{\theta}$, the inequality

$$|H(q_\dagger^j, p_\dagger^j) - E| \le j \times 7\left(\frac{\theta}{T}\right)^2 E \tag{8.12}$$

is satisfied. The case $i = 0$ is obvious, since $H(q_\dagger^0, p_\dagger^0) = H(\mathbf{q}, \mathbf{p}) = E$. We now fix i and assume that Inequality (8.12) is satisfied for all $0 \le j \le i$; we then show that it is satisfied for $j = i + 1$.

Inductive assumption: Suppose that $|H(q_\dagger^i, p_\dagger^i) - E| \le i \times 7\left(\frac{\theta}{T}\right)^2 E$, and that $i \le \frac{T}{\theta} - 1$.

Since $i \le \frac{T}{\theta} - 1 \le \frac{T}{\theta}$ and $\frac{\theta}{T} \le \frac{1}{7}$, Inequality (8.12) implies that

$$H(q^{i+1}, p^{i+1}) \overset{\text{Conservation of Energy}}{=} H(q_\dagger^i, p_\dagger^i) \le E + 7\frac{\theta}{T}E \le 2E. \tag{8.13}$$

Then Lemma 8.1 and Inequality (8.13) together imply that

$$\|q_\dagger^{i+1} - q^{i+1}\| \le \theta^2 \frac{M_2}{\sqrt{m_2}} \sqrt{E} \tag{8.14}$$

and

$$\|p_\dagger^{i+1} - p^{i+1}\| \le \theta^2 M_2 \sqrt{E}. \tag{8.15}$$

But by Equation (8.13),

$$\|p^{i+1}\| \le 2\sqrt{E}. \tag{8.16}$$

Therefore, by the triangle inequality,

$$\begin{aligned}
\|p_\dagger^{i+1}\|^2 - \|p^{i+1}\|^2 &\leq (\|p_\dagger^{i+1} - p^{i+1}\| + \|p^{i+1}\|)^2 - \|p^{i+1}\|^2 \\
&= \|p_\dagger^{i+1} - p^{i+1}\|^2 + 2\|p_\dagger^{i+1} - p^{i+1}\| \times \|p^{i+1}\| \\
&\overset{\text{Eq. 8.16}}{\leq} \|p_\dagger^{i+1} - p^{i+1}\|^2 + 2\|p_\dagger^{i+1} - p^{i+1}\| \times 2\sqrt{E} \\
&= \|p_\dagger^{i+1} - p^{i+1}\| \times (\|p_\dagger^{i+1} - p^{i+1}\| + 4\sqrt{E}) \\
&\overset{\text{Eq. 8.15}}{\leq} \theta^2 M_2 \sqrt{E} \times (\theta^2 M_2 \sqrt{E} + 4\sqrt{E}) \\
&= E\theta^2 M_2 \times (\theta^2 M_2 + 4).
\end{aligned} \tag{8.17}$$

Also, Equation (8.13), together with Assumption 2.1, implies that

$$\|q^{i+1}\| \leq \frac{1}{\sqrt{m_2}}\sqrt{2E}, \tag{8.18}$$

and thus (again by Assumption 2.2) that

$$\|U'(q^{i+1})\| \overset{\text{Assumption 2.2}}{\leq} M_2 \|q^{i+1}\| \overset{\text{Eq. 8.18}}{\leq} \frac{M_2}{\sqrt{m_2}}\sqrt{2E}. \tag{8.19}$$

Therefore,

$$\begin{aligned}
U(q_\dagger^{i+1}) - U(q^{i+1}) &\leq \max_{x \in \text{Convex Hull}(\{q^{i+1}, q_\dagger^{i+1}\})} \|U'(x)\| \times \|q_\dagger^{i+1} - q^{i+1}\| \\
&\overset{\text{Assumption 2.2}}{\leq} \left(\|U'(q^{i+1})\| + M_2\|q_\dagger^{i+1} - q^{i+1}\|\right) \times \|q_\dagger^{i+1} - q^{i+1}\| \\
&\overset{\text{Eq. 8.19}}{\leq} \left(\frac{M_2}{\sqrt{m_2}}\sqrt{2E} + M_2\|q_\dagger^{i+1} - q^{i+1}\|\right) \times \|q_\dagger^{i+1} - q^{i+1}\| \\
&\overset{\text{Eq. 8.14}}{\leq} \left(\frac{M_2}{\sqrt{m_2}}\sqrt{2E} + M_2\theta^2\frac{M_2}{\sqrt{m_2}}\sqrt{E}\right) \times \theta^2\frac{M_2}{\sqrt{m_2}}\sqrt{E} \\
&\leq \left(\sqrt{2} + M_2\theta^2\right) \times \theta^2 \frac{M_2^2}{m_2} E.
\end{aligned} \tag{8.20}$$

Hence the total change in energy is bounded by

$$\begin{aligned}
&|H(q_\dagger^{i+1}, p_\dagger^{i+1}) - H(q_\dagger^{i}, p_\dagger^{i})| \\
&\overset{\text{Conservation of Energy}}{=} |H(q_\dagger^{i+1}, p_\dagger^{i+1}) - H(q^{i+1}, p^{i+1})| \\
&\overset{\text{Eq. 8.20, 8.17}}{\leq} \left(\sqrt{2} + M_2\theta^2\right) \times \theta^2\frac{M_2^2}{m_2}E + \frac{1}{2}E\theta^2 M_2 \times (\theta^2 M_2 + 4) \\
&\leq \left(4 + 1.5 M_2\theta^2\right) \times \theta^2 \frac{M_2^2}{m_2}E \leq (4 + 1.5) \times \theta^2\frac{M_2^2}{m_2}E = 5.5[\frac{\theta}{T} \times T]^2 \frac{M_2^2}{m_2}E \\
&\leq 5.5\left(\frac{\theta}{T} \times \frac{1}{2\sqrt{2}}\frac{\sqrt{m_2}}{M_2}\right)^2 \frac{M_2^2}{m_2}E \leq 7\left(\frac{\theta}{T}\right)^2 E,
\end{aligned} \tag{8.21}$$

where the second inequality is true since $0 < \frac{M_2}{m_2} \leq 1$, and the third and fourth inequalities are true since $\theta \leq T \leq \frac{1}{2\sqrt{2}}\frac{\sqrt{m_2}}{M_2} \leq \frac{1}{\sqrt{M_2}}$.

Therefore,

$$|H(q_\dagger^{i+1}, p_\dagger^{i+1}) - E| \overset{\text{Eq. 8.21}}{\leq} |H(q_\dagger^{i}, p_\dagger^{i}) - E| + 7\left(\frac{\theta}{T}\right)^2 E \tag{8.22}$$

$$\overset{\text{by inductive assumption}}{\leq} i \times 7\left(\frac{\theta}{T}\right)^2 E + 7\left(\frac{\theta}{T}\right)^2 E = (i+1)\times 7\left(\frac{\theta}{T}\right)^2 E.$$

This completes the proof by induction of Inequality (8.12) (and in particular this implies Inequality (8.11)). Therefore, since $\frac{\theta}{T} \leq \frac{1}{7}$,

$$|H(q_\dagger^i, p_\dagger^i) - E| \leq i \times 7\left(\frac{\theta}{T}\right)^2 E \leq E, \qquad \forall\, 0 \leq i \leq \frac{T}{\theta}. \tag{8.23}$$

Therefore, by Lemma 8.1

$$\|q_\dagger^i - q^i\| \leq \theta^2 \frac{M_2}{\sqrt{m_2}}\sqrt{E} \quad \text{and} \quad \|p_\dagger^i - p^i\| \leq \theta^2 M_2 \sqrt{E} \tag{8.24}$$

for all $0 \leq i \leq \frac{T}{\theta}$.

Define $T_i := T - \theta \times i$ for all i. Therefore, since $T_{i+1} \leq T \leq \frac{1}{2\sqrt{2}}\frac{\sqrt{m_2}}{M_2} \leq \frac{1}{\sqrt{M_2}}$, for all $i \leq \frac{T}{\theta} - 1$, we have by Lemma 7.1

$$\begin{aligned}
&\|q_{T_{i+1}}(q_\dagger^{i+1}, p_\dagger^{i+1}) - q_{T_i}(q_\dagger^i, p_\dagger^i)\| \\
&\quad = \|q_{T_{i+1}}(q_\dagger^{i+1}, p_\dagger^{i+1}) - q_{T_{i+1}}(q^{i+1}, p^{i+1})\| \\
&\quad \overset{\text{Lemma 7.1}}{\leq} \frac{1}{2}\left(\|q_\dagger^{i+1} - q^{i+1}\| + \frac{\|p_\dagger^{i+1} - p^{i+1}\|}{\sqrt{M_2}}\right) e^{T_{i+1}\sqrt{M_2}} \\
&\qquad\qquad + \frac{1}{2}\left(\|q_\dagger^{i+1} - q^{i+1}\| - \frac{\|p_\dagger^{i+1} - p^{i+1}\|}{\sqrt{M_2}}\right) e^{-T_{i+1}\sqrt{M_2}} \\
&\quad \leq \frac{1}{2}\left(\|q_\dagger^{i+1} - q^{i+1}\| + \frac{\|p_\dagger^{i+1} - p^{i+1}\|}{\sqrt{M_2}}\right) e \\
&\qquad\qquad + \frac{1}{2}\left(\|q_\dagger^{i+1} - q^{i+1}\| - \frac{\|p_\dagger^{i+1} - p^{i+1}\|}{\sqrt{M_2}}\right) e^0 \\
&\quad \leq \left(\|q_\dagger^{i+1} - q^{i+1}\| + \frac{\|p_\dagger^{i+1} - p^{i+1}\|}{\sqrt{M_2}}\right) e \\
&\quad \overset{\text{Eq. 8.24}}{\leq} \left(\theta^2 \frac{M_2}{\sqrt{m_2}}\sqrt{E} + \frac{\theta^2 M_2 \sqrt{E}}{\sqrt{M_2}}\right) e \leq 6\theta^2 \frac{M_2}{\sqrt{m_2}}\sqrt{E},
\end{aligned} \tag{8.25}$$

where the second inequality is true since $0 \leq T_{i+1} \leq \frac{1}{\sqrt{M_2}}$ and since the functions $e^t + e^{-t}$ and $e^t - e^{-t}$ are both nondecreasing in t for $t \geq 0$; the fifth inequality is true since $\sqrt{m_2} \leq \sqrt{M_2}$.

Therefore, since $q_\dagger^{\frac{T}{\theta}} = q_T^{\dagger\theta}(\mathbf{q}, \mathbf{p})$, $T_0 = T$, and $(q_\dagger^0, p_\dagger^0) = (\mathbf{q}, \mathbf{p})$, by the triangle inequality (see Figure 2) we have

$$\begin{aligned}
\|q_T^{\dagger\theta}(\mathbf{q}, \mathbf{p}) - q_T(\mathbf{q}, \mathbf{p})\| &= \|q_\dagger^{\frac{T}{\theta}} - q_{T_0}(q_\dagger^0, p_\dagger^0)\| \\
&\leq \|q_\dagger^{\frac{T}{\theta}} - q_{T_{(\frac{T}{\theta}-1)}}(q_\dagger^{\frac{T}{\theta}-1}, p_\dagger^{\frac{T}{\theta}-1})\| + \sum_{i=0}^{\frac{T}{\theta}-2} \|q_{T_{i+1}}(q_\dagger^{i+1}, p_\dagger^{i+1}) - q_{T_i}(q_\dagger^i, p_\dagger^i)\| \\
&= \|q_\dagger^{\frac{T}{\theta}} - q^{\frac{T}{\theta}}\| + \sum_{i=0}^{\frac{T}{\theta}-2} \|q_{T_{i+1}}(q_\dagger^{i+1}, p_\dagger^{i+1}) - q_{T_i}(q_\dagger^i, p_\dagger^i)\| \\
&\overset{\text{Eq. 8.24}}{\leq} \theta^2 \frac{M_2}{\sqrt{m_2}}\sqrt{E} + \sum_{i=0}^{\frac{T}{\theta}-2} \|q_{T_{i+1}}(q_\dagger^{i+1}, p_\dagger^{i+1}) - q_{T_i}(q_\dagger^i, p_\dagger^i)\|
\end{aligned}$$

$$\overset{\text{Eq. 8.25}}{\leq} \theta^2 \frac{M_2}{\sqrt{m_2}}\sqrt{E} + \sum_{i=0}^{\frac{T}{\theta}-2} 6\theta^2 \frac{M_2}{\sqrt{m_2}}\sqrt{E} \leq \frac{T}{\theta} \times 6\theta^2 \frac{M_2}{\sqrt{m_2}}\sqrt{E}.$$

This completes the proof of Inequality (8.10). □

9 Mixing of Approximate HMC

In this section, we prove the main bounds used in the proof of Theorem 1. We first need the generic bounds in the following subsection:

9.1 Perturbation Bounds for Markov Chains

We give some simple general bounds on small perturbations of Markov chains. As this section is not specifically about HMC, the notation used here is essentially independent of the notation used in the remainder of the paper. We begin by recalling the definition of the *trace* of a Markov chain on a set:

Definition 9.1 (Trace Chain)**.** Let K be the transition kernel of an ergodic Markov chain on state space Ω with stationary measure μ, and let $S \subset \Omega$ be a subset with $\mu(S) > 0$. Let $\{X_t\}_{t \geq 0}$ be a Markov chain evolving according to K, and iteratively define

$$c_0 = \inf\{t \geq 0 \, : \, X_t \in S\}, \qquad c_{i+1} = \inf\{t > c_i \, : \, X_t \in S\}.$$

Then

$$\hat{X}_t = X_{c_t}, \quad t \geq 0 \tag{9.1}$$

is the *trace of* $\{X_t\}_{t \geq 0}$ *on* S. Note that $\{\hat{X}_t\}_{t \geq 0}$ is a Markov chain with state space S, and so this procedure also defines a transition kernel with state space S. We call this kernel the *trace of the kernel* K *on* S.

Note that Equation (9.1) defines a coupling between $\{X_t\}_{t \geq 0}$ and its trace chain $\{\hat{X}_t\}_{t \geq 0}$; we call this the "natural" coupling between a chain and its trace.

Recall the following bound from [35]:

Lemma 9.2. *Let K be a transition kernel on $\mathbb{R}^d$ with unique stationary measure μ and contraction coefficient $\kappa > 0$ satisfying*

$$W_1(K(x, \cdot), K(y, \cdot)) \leq (1 - \kappa) \|x - y\| \tag{9.2}$$

for all $x, y \in \mathbb{R}^d$. Let Q be a transition kernel on $\mathbb{R}^d$. Assume that

$$\sup_{x \in \mathbb{R}^d} W_1(K(x, \cdot), Q(x, \cdot)) < \delta \tag{9.3}$$

for some fixed $\delta \geq 0$. Then Q satisfies

$$W_1(\pi_1 \, Q^{\mathcal{I}}, \pi_2 \, K^{\mathcal{I}}) \leq (1 - \kappa)^{\mathcal{I}} \, W_1(\pi_1, \pi_2) + \frac{\delta}{\kappa} \tag{9.4}$$

for all measures π_1, π_2 and $\mathcal{I} \in \mathbb{N}$. Furthermore, if Q is ergodic with stationary measure ν , then

$$W_1(\mu, \nu) \leq \frac{\delta}{\kappa}. \tag{9.5}$$

We add the following short lemma:

Lemma 9.3. *Let K, Q be two ergodic transition kernels on $\mathbb{R}^d$ with stationary measures μ and ν, let $d \, : \, (\mathbb{R}^d)^2 \mapsto [0, \infty)$ be a nonnegative function, and let $S \subset \mathbb{R}^d$ be a measurable set with $\mu(S), \nu(S) > 0$. Let $x, y \in S$, let $s \in \mathbb{N}$ and let $\hat{K}, \hat{Q}$ be the traces of K, Q on S. Next, fix any coupling of $\{X_t\}_{t \geq 0} \sim K$, $\{Y_t\}_{t \geq 0} \sim Q$ with $X_0 = x$,*

$Y_0 = y$, and let the pairs $\{\hat{X}_t\}_{t\geq 0} \sim \hat{K}$, $\{X_t\}_{t\geq 0}$ and $\{\hat{Y}_t\}_{t\geq 0} \sim \hat{Q}$, $\{Y_t\}_{t\geq 0}$ be coupled according to the natural coupling of a chain to its trace (as in Definition 9.1).

Then

$$\mathbb{E}[d(\hat{X}_s, \hat{Y}_s)\mathbf{1}_{\phi > s}] \leq \mathbb{E}[d(X_s, Y_s)],$$

where

$$\phi = \min\{t \,:\, X_t \notin S \text{ or } Y_t \notin S\}.$$

Proof. We have

$$\begin{aligned}\mathbb{E}[d(\hat{X}_s, \hat{Y}_s)\mathbf{1}_{\tau > s}] &= \mathbb{E}[d(X_s, Y_s)\mathbf{1}_{\phi > s}] \\ &\leq \mathbb{E}[d(X_s, Y_s)],\end{aligned}$$

where the equality follows from the fact that $\{X_t\}_{t=0}^{\tau} = \{\hat{X}_t\}_{t=0}^{\tau}$ (and similarly for $\{Y_t\}_{t=0}^{\tau}$), and the inequality follows from the fact that $d \geq 0$ and $\mathbf{1}_{\phi > s} \in [0, 1]$. □

We use these to prove the following bound, which we can use if the approximation of K by Q is not uniformly good:

Lemma 9.4. *Let K be a transition kernel on metric space $\mathbb{R}^d$ with unique stationary measure μ and contraction coefficient $\kappa > 0$ satisfying*

$$W_1(K(x, \cdot), K(y, \cdot)) \leq (1 - \kappa)\|x - y\| \tag{9.6}$$

for all $x, y \in \mathbb{R}^d$. Let Q be a transition kernel on $\mathbb{R}^d$. Fix $\lambda > 0$ and define $V(x) \equiv e^{\lambda\|x\|}$. Assume that there exists $0 < \alpha < 1$, $0 \leq \beta < \infty$ so that

$$\begin{aligned}Q(x, \cdot)[V] &\leq (1 - \alpha)V(x) + \beta, \\ K(x, \cdot)[V] &\leq (1 - \alpha)V(x) + \beta \qquad \forall x \in \mathbb{R}^d.\end{aligned} \tag{9.7}$$

Assume that there exists some $\frac{4\beta}{\alpha} < C < \infty$ and $\delta > 0$ so that

$$\sup_{x \,:\, V(x) \leq C} W_1(K(x, \cdot), Q(x, \cdot)) < \delta. \tag{9.8}$$

Then Q satisfies

$$W_1(Q^{\mathcal{I}}(x, \cdot), \mu) \leq \frac{2\log(C)}{\lambda}(1 - \kappa)^s + \frac{\delta}{\kappa} + \frac{\beta(s+1)}{\lambda\alpha C}(8 + 2\log(C)) \qquad , \tag{9.9}$$

for all $0 \leq s \leq \mathcal{I} \in \mathbb{N}$ and all x satisfying $V(x) \leq \frac{\beta}{\alpha}$. Furthermore, if Q is ergodic with stationary measure ν,

$$W_1(\mu, \nu) \leq \inf_{s \in \{0, 1, \ldots\}} \frac{2\log(C)}{\lambda}(1 - \kappa)^s + \frac{\delta}{\kappa} + \frac{\beta(s+1)}{\lambda\alpha C}(8 + 2\log(C)) \qquad . \tag{9.10}$$

Proof. We make some initial estimates. Let $\{X_{\mathcal{I}}\}_{\mathcal{I}\geq 0}$ be a Markov chain evolving according to Q and started at a (possibly random) point X_0 satisfying $\mathbb{E}[V(X_0)] \leq \frac{\beta}{\alpha}$. By Inequality (9.7),

$$\begin{aligned}\mathbb{E}[V(X_{\mathcal{I}})] &\leq (1 - \alpha)\mathbb{E}[V(X_{\mathcal{I}-1})] + \beta \leq \ldots \\ &\leq (1 - \alpha)^{\mathcal{I}}\mathbb{E}[V(X_0)] + \frac{\beta}{\alpha} \leq \frac{2\beta}{\alpha} \qquad \forall \mathcal{I} \in \mathbb{N}.\end{aligned} \tag{9.11}$$

By Markov's inequality, then,

$$\mathbb{P}[\sup_{\mathcal{I}-s\leq h\leq \mathcal{I}} V(X_h) > r] \leq \frac{2\beta(s+1)}{\alpha r} \tag{9.12}$$

for any fixed integers $0 \leq s \leq \mathcal{I}$ and any $r > 0$. Rewriting this,

$$\mathbb{P}[\sup_{\mathcal{I}-s\leq h\leq \mathcal{I}} e^{\lambda\|X_h\|} > r] \leq \frac{2\beta(s+1)}{\alpha r} \qquad \forall\, r > 0, \tag{9.13}$$

and so

$$\mathbb{P}[\sup_{\mathcal{I}-s\leq h\leq \mathcal{I}} \|X_h\| > r] \leq \frac{2\beta(s+1)}{\alpha} e^{-\lambda r} \qquad \forall r > 0, \tag{9.14}$$

which gives

$$\begin{aligned} &\mathbb{E}[(\sup_{\mathcal{I}-s\leq h\leq \mathcal{I}} \|X_h\|) \times \mathbb{1}\{\sup_{\mathcal{I}-s\leq h\leq \mathcal{I}} V(X_h) \geq C\}] \\ &\quad = \mathbb{E}[(\sup_{\mathcal{I}-s\leq h\leq \mathcal{I}} \|X_h\|) \times \mathbb{1}\{\sup_{\mathcal{I}-s\leq h\leq \mathcal{I}} \|X_h\| \geq \frac{1}{\lambda}\log(C)\}] \\ &\quad \overset{\text{Eq. 9.14}}{\leq} \int_{\frac{1}{\lambda}\log(C)}^{\infty} \frac{2\beta(s+1)}{\alpha} e^{-\lambda r} \mathrm{d}r = \frac{2\beta(s+1)}{\lambda\alpha C}. \end{aligned} \tag{9.15}$$

Also, by Jensen's inequality, Equation (9.11) implies that,

$$e^{\lambda\mathbb{E}[\|X_{\mathcal{I}}\|]} \leq \mathbb{E}[e^{\lambda\|X_{\mathcal{I}}\|}] = \mathbb{E}[V(X_{\mathcal{I}})] \overset{\text{Eq. 9.11}}{\leq} \frac{2\beta}{\alpha},$$

and so

$$\mathbb{E}[\|X_{\mathcal{I}}\|] \leq \lambda^{-1}\log(\frac{2\beta}{\alpha}). \tag{9.16}$$

Now let $\{Y_{\mathcal{I}}\}_{\mathcal{I}\geq 0}$ be a Markov chain evolving according to K and started at $Y_0 \sim \mu$. Then by Inequality (9.7),

$$\mathbb{E}[V(Y_0)] = \mathbb{E}[V(Y_1)] \leq (1-\alpha)\mathbb{E}[V(Y_0)] + \beta,$$

which implies that $\mathbb{E}[V(Y_0)] \leq \frac{\beta}{\alpha}$. Therefore, by the same argument we used to show Inequalities (9.13), (9.15), and (9.16) (but replacing K with Q and X_h with Y_h), we must have, respectively,

$$\mathbb{E}[\sup_{\mathcal{I}-s\leq h\leq \mathcal{I}} \|Y_h\| \times \mathbb{1}\{\sup_{\mathcal{I}-s\leq h\leq \mathcal{I}} V(Y_h) \geq C\}] \leq \frac{2\beta(s+1)}{\lambda\alpha C} \tag{9.17}$$

and

$$\mathbb{P}[\sup_{\mathcal{I}-s\leq h\leq \mathcal{I}} e^{\lambda\|Y_h\|} > r] \leq \frac{2\beta(s+1)}{\alpha r} \qquad \forall r > 0 \tag{9.18}$$

and

$$\mathbb{E}[\|Y_{\mathcal{I}}\|] \leq \lambda^{-1}\log(\frac{2\beta}{\alpha}). \tag{9.19}$$

We now prove Inequality (9.9) using an explicit coupling. Fix integers $0 \leq s \leq \mathcal{I} < \infty$. Let $\{X_h\}_{h=0}^{\mathcal{I}-s}$, $\{Y_h\}_{h=0}^{\mathcal{I}-s}$ be Markov chains with transition kernels Q and K respectively, and starting points $X_0 = x$ and $Y \sim \mu$. We begin to construct our coupling of these two chains by allowing them to evolve independently over the time interval $\{0, 1, \ldots, \mathcal{I} - s\}$.

Next, let $\tilde{Q}$, $\tilde{K}$ be the traces of Q, K on the set $\{z \in \mathbb{R}^d : V(z) \leq C\}$. Fix $\gamma > 0$. By Lemmas 9.2 and 9.3, it is possible to couple two Markov chains $\{\tilde{X}_i\}_{i\geq 0}$, $\{\tilde{Y}_i\}_{i\geq 0}$ with transition kernels $\tilde{Q}$, $\tilde{K}$ and starting points

$$\tilde{X}_0 = \begin{cases} X_{\mathcal{I}-s}, & V(X_{\mathcal{I}-s}) \leq C \\ 0, & V(X_{\mathcal{I}-s}) > C \end{cases} \tag{9.20}$$

$$\tilde{Y}_0 = \begin{cases} Y_{\mathcal{I}-s}, & V(Y_{\mathcal{I}-s}) \leq C \\ 0, & V(Y_{\mathcal{I}-s}) > C \end{cases}$$

so that

$$\mathbb{E}[\|\tilde{X}_s - \tilde{Y}_s\| \mathbf{1}_{\phi > s}] \leq (1-\kappa)^s \sup_{x,y : V(x), V(y) \leq C} \|x - y\| + \frac{\delta}{\kappa} + \gamma, \tag{9.21}$$

where

$$\phi = \min\{t \geq 0 : X_s \notin S \text{ or } Y_s \notin S\}.$$

Next, couple $\{X_h\}_{h=\mathcal{I}-s}^{\mathcal{I}}$ to $\{\tilde{X}_h\}_{h=0}^{s}$ (respectively, couple $\{Y_h\}_{h=\mathcal{I}-s}^{\mathcal{I}}$ to $\{\tilde{Y}_h\}_{h=0}^{s}$) according to the natural coupling of a Markov chain to its trace on a set (that is, the coupling in Definition 9.1). Defining $\tau_X = \min\{h \geq \mathcal{I} - s : V(X_h) > C\}$, $\tau_Y = \min\{h \geq \mathcal{I} - s : V(Y_h) > C\}$, and $\tau = \min\{\tau_X, \tau_Y\}$, we have:

$$\begin{aligned}
\mathbb{E}[\|X_{\mathcal{I}} - Y_{\mathcal{I}}\|] &= \mathbb{E}[\|X_{\mathcal{I}} - Y_{\mathcal{I}}\| \, \mathbb{1}_{\tau > \mathcal{I}}] + \mathbb{E}[\|X_{\mathcal{I}} - Y_{\mathcal{I}}\| \, \mathbb{1}_{\tau \leq \mathcal{I}}] \\
&\leq \mathbb{E}[\|X_{\mathcal{I}} - Y_{\mathcal{I}}\| \, \mathbb{1}_{\tau > \mathcal{I}}] + \mathbb{E}[(\|X_{\mathcal{I}}\| + \|Y_{\mathcal{I}}\|) \, \mathbb{1}_{\tau \leq \mathcal{I}}] \\
&\leq \mathbb{E}[\|X_{\mathcal{I}} - Y_{\mathcal{I}}\| \, \mathbb{1}_{\tau > \mathcal{I}}] + \mathbb{E}[(\|X_{\mathcal{I}}\| + \|Y_{\mathcal{I}}\|) \, \mathbb{1}_{\tau_X \leq \mathcal{I}}] + \mathbb{E}[(\|X_{\mathcal{I}}\| + \|Y_{\mathcal{I}}\|) \, \mathbb{1}_{\tau_Y \leq \mathcal{I}}] \\
&\leq \mathbb{E}[\|X_{\mathcal{I}} - Y_{\mathcal{I}}\| \, \mathbb{1}_{\tau > \mathcal{I}}] + \mathbb{E}[(\|X_{\mathcal{I}}\| + \|Y_{\mathcal{I}}\|) \, \mathbb{1}_{\tau_X \leq \mathcal{I}} \times (\mathbb{1}_{\tau_Y \leq \mathcal{I}} + \mathbb{1}_{\tau_Y > \mathcal{I}})] \\
&\quad + \mathbb{E}[(\|X_{\mathcal{I}}\| + \|Y_{\mathcal{I}}\|) \, \mathbb{1}_{\tau_Y \leq \mathcal{I}} \times (\mathbb{1}_{\tau_X \leq \mathcal{I}} + \mathbb{1}_{\tau_X > \mathcal{I}})] \\
&\leq \mathbb{E}[\|X_{\mathcal{I}} - Y_{\mathcal{I}}\| \, \mathbb{1}_{\tau > \mathcal{I}}] + \mathbb{E}[\|X_{\mathcal{I}}\| \, \mathbb{1}_{\tau_X \leq \mathcal{I}}] \\
&\quad + \mathbb{E}[\|Y_{\mathcal{I}}\| \, \mathbb{1}_{\tau_Y \leq \mathcal{I}}] + \mathbb{E}[\|Y_{\mathcal{I}}\| \, \mathbb{1}_{\tau_X \leq \mathcal{I}} \times \mathbb{1}_{\tau_Y > \mathcal{I}}] \\
&\qquad + \mathbb{E}[\|Y_{\mathcal{I}}\| \, \mathbb{1}_{\tau_Y \leq \mathcal{I}}] + \mathbb{E}[\|X_{\mathcal{I}}\| \, \mathbb{1}_{\tau_X \leq \mathcal{I}}] + \mathbb{E}[\|X_{\mathcal{I}}\| \, \mathbb{1}_{\tau_Y \leq \mathcal{I}} \times \mathbb{1}_{\tau_X > \mathcal{I}}] \\
&\leq \mathbb{E}[\|X_{\mathcal{I}} - Y_{\mathcal{I}}\| \, \mathbb{1}_{\tau > \mathcal{I}}] + \mathbb{E}[\|X_{\mathcal{I}}\| \, \mathbb{1}_{\tau_X \leq \mathcal{I}}] \\
&\qquad + \mathbb{E}[\|Y_{\mathcal{I}}\| \, \mathbb{1}_{\tau_Y \leq \mathcal{I}}] + \mathbb{E}[\frac{1}{\lambda} \log(C) \, \mathbb{1}_{\tau_X \leq \mathcal{I}} \times \mathbb{1}_{\tau_Y > \mathcal{I}}] \\
&\qquad + \mathbb{E}[\|Y_{\mathcal{I}}\| \, \mathbb{1}_{\tau_Y \leq \mathcal{I}}] + \mathbb{E}[\|X_{\mathcal{I}}\| \, \mathbb{1}_{\tau_X \leq \mathcal{I}}] + \mathbb{E}[\frac{1}{\lambda} \log(C) \, \mathbb{1}_{\tau_Y \leq \mathcal{I}} \times \mathbb{1}_{\tau_X > \mathcal{I}}] \\
&\leq \mathbb{E}[\|\tilde{X}_s - \tilde{Y}_s\| \mathbb{1}_{\tau > \mathcal{I}}] + 2\mathbb{E}[\sup_{\mathcal{I}-s \leq h \leq \mathcal{I}} \|X_h\| \times \mathbb{1}\{\sup_{\mathcal{I}-s \leq h \leq \mathcal{I}} V(X_h) \geq C\}] \\
&\quad + \frac{1}{\lambda} \log(C) \, \mathbb{P}[\sup_{\mathcal{I}-s \leq h \leq \mathcal{I}} V(X_h) \geq C\}] \\
&\quad + 2\mathbb{E}[\sup_{\mathcal{I}-s \leq h \leq \mathcal{I}} \|Y_h\| \times \mathbb{1}\{\sup_{\mathcal{I}-s \leq h \leq \mathcal{I}} V(Y_h) \geq C\}] \\
&\qquad + \frac{1}{\lambda} \log(C) \times \mathbb{P}[\sup_{\mathcal{I}-s \leq h \leq \mathcal{I}} V(Y_h) \geq C\}] \\
&\overset{\text{Eq. 9.13,9.18}}{\leq} \mathbb{E}[\|\tilde{X}_s - \tilde{Y}_s\| \mathbb{1}_{\tau > \mathcal{I}}] + 2\mathbb{E}[\sup_{\mathcal{I}-s \leq h \leq \mathcal{I}} \|X_h\| \times \mathbb{1}\{\sup_{\mathcal{I}-s \leq h \leq \mathcal{I}} V(X_h) \geq C\}] \\
&\quad + 2\mathbb{E}[\sup_{\mathcal{I}-s \leq h \leq \mathcal{I}} \|Y_h\| \times \mathbb{1}\{\sup_{\mathcal{I}-s \leq h \leq \mathcal{I}} V(Y_h) \geq C\}] + 2\lambda^{-1} \log(C) \times \frac{2\beta(s+1)}{\alpha C}.
\end{aligned}$$

Applying Inequalities (9.15), (9.17), and (9.21),

$$\mathbb{E}[\|X_{\mathcal{I}} - Y_{\mathcal{I}}\|]$$

$$
\begin{aligned}
&\leq (1-\kappa)^s \sup_{x,y\,:\,V(x),\,V(y)\leq C} \|x-y\| + \frac{\delta}{\kappa} + \frac{8\beta(s+1)}{\lambda\alpha C} \\
&+ 2\lambda^{-1}\log(C) \times \frac{2\beta(s+1)}{\alpha C} + \gamma \\
&\leq (1-\kappa)^s \frac{2\log(C)}{\lambda} + \frac{\delta}{\kappa} + \frac{\beta(s+1)}{\lambda\alpha C}\,(8+2\log(C)) + \gamma,
\end{aligned}
$$

where the second line uses the fact $\sup_{x,y\,:\,V(x),\,V(y)\leq C}\|x-y\| \leq \frac{2\log(C)}{\lambda}$. Since $\gamma > 0$ was arbitrary, this completes the proof of Inequality (9.9).

To prove Inequality (9.10), we note by the ergodicity of Q and Inequality (9.9) that, for fixed x satisfying $V(x) \leq \frac{\beta}{\alpha}$,

$$
\begin{aligned}
W_1(\nu,\mu) &= \lim_{\mathcal{I}\to\infty} W_1(Q^{\mathcal{I}}(x,\cdot),\mu) \\
&\leq \lim_{\mathcal{I}\to\infty} \inf_{0\leq s\leq \mathcal{I}} \left[\frac{2\log(C)}{\lambda}(1-\kappa)^s + \frac{\delta}{\kappa} + \frac{\beta(s+1)}{\lambda\alpha C}\,(8+2\log(C))\right] \\
&= \inf_{s\in\{0,1,\ldots\}} \left[\frac{2\log(C)}{\lambda}(1-\kappa)^s + \frac{\delta}{\kappa} + \frac{\beta(s+1)}{\lambda\alpha C}\,(8+2\log(C))\right],
\end{aligned}
$$

which completes the proof. □

9.2 Bounds on unadjusted HMC Algorithm

We now prove a mixing bound for the unadjusted HMC algorithm described in Algorithm 2. The proof strategy is:

- We use the error bounds on the leapfrog integrator proved in Section 8 to show that the unadjusted HMC Markov chain described in Algorithm 2 is a "small perturbation" of the ideal HMC Markov chain described in Algorithm 1.
- We use the "small perturbation" results of Section 9.1 to show that the unadjusted HMC Markov chain inherits the good mixing properties of the ideal HMC chain (the latter are summarized in Section 7).

We set some notation to be used throughout this section. We always fix a target potential U that satisfies Assumptions 2.1 and 2.2. We also fix integration time $0 \leq T \leq \frac{1}{2\sqrt{2}}\frac{\sqrt{m_2}}{M_2}$ and let K be the transition kernel defined in Algorithm 1 with these parameters.

For any $0 \leq \theta \leq \frac{T}{5}$, we denote by Q the transition kernel defined by Algorithm 2 with these parameters. We use the notation from Algorithms 1 and 2 in our analyses of these algorithms. We begin with coarse bounds on the behavior of the HMC chain:

Lemma 9.5. *Set notation as above. Fix $C' > 0$ and $0 < \lambda \leq \frac{1}{32}\sqrt{2m_2}$ and define $V(x) \equiv e^{\lambda\|x\|}$ for all $x \in \mathbb{R}^d$. Let $\kappa := \frac{1}{8}(\sqrt{m}_2 T)^2$ and fix $\theta > 0$ such that $6\theta\, T\, \frac{M_2}{\sqrt{m_2}}\sqrt{M_2} < \kappa$. Then for all $x \in \mathbb{R}^d$ we have the drift condition*

$$
\begin{aligned}
K(x,\cdot)[V] &\leq (1-\alpha)V(x) + \beta, \\
Q(x,\cdot)[V] &\leq (1-\alpha)V(x) + \beta,
\end{aligned}
\tag{9.22}
$$

where

$$
\begin{aligned}
\alpha &= 1 - (1+e^{-\frac{d}{8}})e^{\lambda\left(-\kappa+6\theta\, T\, \frac{M_2}{\sqrt{m_2}}\sqrt{M_2}\right)C' + \lambda(1+6\theta\, T\, M_2)\sqrt{\frac{d}{m_2}}} \\
\beta &= e^{\lambda C'}\,(1-\alpha).
\end{aligned}
$$

Moreover,

$$
\|q_T^{\dagger\theta}(x,\mathbf{p})\| \leq \left(1-\kappa+6\theta\, T\, \frac{M_2^{1.5}}{\sqrt{m_2}}\right)\|x\| + (1+6\theta\, T\, M_2)\,\frac{\|\mathbf{p}\|}{\sqrt{2m_2}} \qquad , \tag{9.23}
$$

for any $x, \mathbf{p} \in \mathbb{R}^d$.

Proof. We begin by coupling two Markov chains evolving according to Algorithm 1 to a Markov chain evolving according to Algorithm 2.

Let $\mathbf{p}_0, \mathbf{p}_1, \ldots$ be independent $\mathcal{N}(0, I_d)$ Gaussians. Recalling the forward mapping representation (2.3) of Algorithm 1, we set initial conditions $X_0 = x$, $Y_0 = 0$, and inductively define

$$X_{i+1} = \mathcal{Q}_T^{X_i}(\mathbf{p}_i), \qquad Y_{i+1} = \mathcal{Q}_T^{Y_i}(\mathbf{p}_i)$$

for $i \in \mathbb{N}$. This gives a coupling of two copies $\{X_i\}_{i\geq 0}$, $\{Y_i\}_{i\geq 0}$ of the idealized HMC chain defined in Algorithm 1. We couple these two chains to a Markov chain $(X_0', X_1', \ldots)$ generated by Algorithm 2 with starting point $X_0' = X_0 = x$ by inductively setting

$$X_{i+1}' = q_T^{\dagger\theta}(X_i', \mathbf{p}_i)$$

for all $i \geq 0$. Note that the resulting Markov chain evolves according to Algorithm 2.

Define $\mathfrak{p}_0 = \mathbf{p}_0$, so $\mathfrak{p}_0 \sim \mathcal{N}(0, I_d)$. By Theorem 3 we have

$$\|X_1 - Y_1\| \leq (1-\kappa)\,\|X_0 - Y_0\| = (1-\kappa)\,\|X_0\|. \tag{9.24}$$

Also, by conservation of energy for Hamilton's equations,

$$U(Y_1) \leq H(Y_0, \mathbf{p}_0) = 0 + \frac{1}{2}\|\mathfrak{p}_0\|^2.$$

By strong convexity of U (Assumption 2.1), we have $m_2\|Y_1\|^2 \leq U(Y_1)$, so

$$\|Y_1\| \leq \frac{1}{\sqrt{m_2}}\sqrt{U(Y_1)} \leq \frac{\|\mathfrak{p}_0\|}{\sqrt{2m_2}}. \tag{9.25}$$

Therefore, combining Inequalities (9.24) and (9.25),

$$\|X_1\| \leq \|X_1 - Y_1\| + \|Y_1\| \leq (1-\kappa)\,\|X_0\| + \|Y_1\| \leq (1-\kappa)\,\|X_0\| + \frac{\|\mathfrak{p}_0\|}{\sqrt{2m_2}}$$

and so,

$$e^{\lambda\|X_1\|} \leq e^{\lambda(1-\kappa)\times\|X_0\|} e^{\lambda\frac{\|\mathfrak{p}_0\|}{\sqrt{2m_2}}}. \tag{9.26}$$

Inequality (9.26) implies that

$$e^{\lambda\|X_1\|} \leq \begin{cases} (1-\alpha')e^{\lambda\|X_0\|}, & \|X_0\| \geq C' \\ \beta', & \|X_0\| \leq C', \end{cases}$$

where $\alpha' = 1 - e^{-\lambda\kappa C' + \lambda\frac{\|\mathfrak{p}_0\|}{\sqrt{2m_2}}}$ and $\beta' = e^{\lambda(1-\kappa)C' + \lambda\frac{\|\mathfrak{p}_0\|}{\sqrt{2m_2}}}$ are random variables that depend on $\mathfrak{p}_0$. Thus, for any X_0,

$$e^{\lambda\|X_1\|} \leq (1-\alpha')e^{\lambda\|X_0\|} + \beta'. \tag{9.27}$$

Now, $\|\mathfrak{p}_0\|^2 \sim \chi_d^2$, so by the Hanson-Wright concentration inequality (see, for instance, [23, 45]), we have

$$\mathbb{P}(\|\mathfrak{p}_0\|^2 > s + d) \leq e^{-\frac{s}{8}} \qquad \text{for } s > d.$$

Therefore, for any γ such that $0 < \gamma < \frac{1}{16}\sqrt{2m_2}$,

$$\mathbb{P}\left[e^{\gamma\frac{\|\mathfrak{p}_0\|}{\sqrt{2m_2}}} > e^{\gamma\frac{\sqrt{s+d}}{\sqrt{2m_2}}}\right] \leq e^{-\frac{s}{8}} \qquad \text{for } s > d$$

and so

$$\mathbb{P}\left[e^{\gamma\frac{\|\mathfrak{p}_0\|}{\sqrt{2m_2}}} > r\right] \le e^{-\frac{1}{8}((\frac{\sqrt{2m_2}}{\gamma}\log(r))^2-d)} \qquad \text{for } r > e^{\gamma\frac{\sqrt{2d}}{\sqrt{2m_2}}}. \tag{9.28}$$

However, for $r > e^{\gamma\frac{\sqrt{2d}}{\sqrt{2m_2}}}$, we have

$$\begin{aligned} e^{-\frac{1}{8}((\frac{\sqrt{2m_2}}{\gamma}\log(r))^2-d)} &= e^{\frac{1}{8}d}[e^{\log(r)}]^{-\frac{1}{8}(\frac{\sqrt{2m_2}}{\gamma})^2\log(r)} = e^{\frac{1}{8}d}r^{-\frac{1}{8}(\frac{\sqrt{2m_2}}{\gamma})^2\log(r)} \\ &\le e^{\frac{1}{8}d}r^{-\frac{1}{8}\frac{\sqrt{2m_2}}{\gamma}\sqrt{2d}}. \end{aligned} \tag{9.29}$$

Inequalities (9.28) and (9.29) together give

$$\mathbb{P}\left[e^{\gamma\frac{\|\mathfrak{p}_0\|}{\sqrt{2m_2}}} > r\right] \le e^{\frac{1}{8}d}r^{-\frac{1}{8}\frac{\sqrt{2m_2}}{\gamma}\sqrt{2d}} \qquad \text{for } r > e^{\gamma\frac{\sqrt{2d}}{\sqrt{2m_2}}}, \tag{9.30}$$

and hence

$$\begin{aligned} \mathbb{E}\left[e^{\gamma\frac{\|\mathfrak{p}_0\|}{\sqrt{2m_2}}}\right] &\overset{\text{Eq. 9.30}}{\le} e^{\gamma\frac{\sqrt{2d}}{\sqrt{2m_2}}} + \int_{e^{\gamma\frac{\sqrt{2d}}{\sqrt{2m_2}}}}^{\infty} e^{\frac{1}{8}d}r^{-\frac{1}{8}\frac{\sqrt{2m_2}}{\gamma}\sqrt{2d}}\mathrm{d}r \\ &= e^{\gamma\frac{\sqrt{2d}}{\sqrt{2m_2}}} + \frac{1}{\frac{1}{8}\frac{\sqrt{2m_2}}{\gamma}\sqrt{2d}-1}e^{-\frac{1}{8}d+\gamma\frac{\sqrt{2d}}{\sqrt{2m_2}}} \le (1+e^{-\frac{d}{8}})\,e^{\gamma\frac{\sqrt{2d}}{\sqrt{2m_2}}} \end{aligned}$$

where the inequality uses the fact that $\gamma \le \frac{1}{16}\sqrt{2m_2}$. So we have

$$\mathbb{E}\left[e^{\gamma\frac{\|\mathfrak{p}_0\|}{\sqrt{2m_2}}}\right] \le (1+e^{-\frac{d}{8}})e^{\gamma\frac{\sqrt{2d}}{\sqrt{2m_2}}} \qquad \text{for } 0<\gamma\le\frac{1}{16}\sqrt{2m_2}. \tag{9.31}$$

Therefore,

$$\begin{aligned} \mathbb{E}[e^{\lambda\|X_1\|}] &\overset{\text{Eq. 9.27}}{\le} \mathbb{E}[(1-\alpha')e^{\lambda\|X_0\|}] + \mathbb{E}[\beta'] \\ &= \mathbb{E}[e^{-\lambda\kappa C'+\lambda\frac{\|\mathfrak{p}_0\|}{\sqrt{2m_2}}}e^{\lambda\|X_0\|}] + \mathbb{E}[e^{\lambda(1-\kappa)C'+\lambda\frac{\|\mathfrak{p}_0\|}{\sqrt{2m_2}}}] \\ &\overset{\text{Eq. 9.31}}{\le} (1+e^{-\frac{d}{8}})e^{-\lambda\kappa C'}e^{\lambda\frac{\sqrt{2d}}{\sqrt{2m_2}}}e^{\lambda\|X_0\|} + (1+e^{-\frac{d}{8}})e^{\lambda(1-\kappa)C'}e^{\lambda\frac{\sqrt{2d}}{\sqrt{2m_2}}}, \end{aligned} \tag{9.32}$$

where the assumption of Inequality (9.31) is satisfied because $0<\lambda\le\frac{1}{32}\sqrt{2m_2}<\frac{1}{16}\sqrt{2m_2}$.

Next, since $6\theta\,T\,\frac{M_2}{\sqrt{m_2}}>0$, we have

$$\begin{aligned} 1-\alpha &= (1+e^{-\frac{d}{8}})e^{\lambda\left(-\kappa+6\theta\,T\,\frac{M_2}{\sqrt{m_2}}\sqrt{M_2}\right)C'+\lambda(1+6\theta\,T\,M_2)\sqrt{\frac{d}{m_2}}} \\ &\ge (1+e^{-\frac{d}{8}})e^{-\lambda\kappa C'+\lambda\frac{\sqrt{2d}}{\sqrt{2m_2}}} \end{aligned} \tag{9.33}$$

and

$$\beta = e^{\lambda C'}(1-\alpha) \ge (1+e^{-\frac{d}{8}})e^{\lambda(1-\kappa)C'+\lambda\frac{\sqrt{2d}}{\sqrt{2m_2}}}. \tag{9.34}$$

Therefore, substituting Inequalities (9.33) and (9.34) into Inequality (9.32), we get

$$\mathbb{E}[V(X_1)] \le (1-\alpha)V(X_0)+\beta,$$

and hence

$$K(x,\cdot)[V] \le (1-\alpha)\times V(x)+\beta. \tag{9.35}$$

This proves the first line of Ineqality (9.22). We now prove Inequality (9.23), before finally proving the second line of Inequality (9.22). By Lemma 8.2, for every $\mathbf{p} \in \mathbb{R}^d$ we have

$$\begin{aligned}\|q_T^{\dagger\theta}(x,\mathbf{p}) - q_T(x,\mathbf{p})\| &\leq 6\theta\, T\, \frac{M_2}{\sqrt{m_2}} \sqrt{H(X_0,\mathbf{p})} \\ &\overset{\text{Assumption 2.2}}{\leq} 6\theta\, T\, \frac{M_2}{\sqrt{m_2}} \sqrt{M_2\|X_0\|^2 + \frac{1}{2}\|\mathbf{p}\|^2} \leq 6\theta\, T\, \frac{M_2}{\sqrt{m_2}} \left[\sqrt{M_2\|X_0\|^2} + \sqrt{\frac{1}{2}\|\mathbf{p}\|^2}\right].\end{aligned} \tag{9.36}$$

Also, by the same calculation that was used to prove Inequality (9.26), we have

$$e^{\lambda\|q_T(x,\mathbf{p})\|} \leq e^{\lambda(1-\kappa)\times\|x\|} e^{\lambda\frac{\|\mathbf{p}\|}{\sqrt{2m_2}}} \qquad \forall \mathbf{p} \in \mathbb{R}^d. \tag{9.37}$$

And so for all $\mathbf{p} \in \mathbb{R}^d$ we have,

$$\begin{aligned}\|q_T^{\dagger\theta}(x,\mathbf{p})\| &\leq \|q_T(x,\mathbf{p})\| + \|q_T^{\dagger\theta}(x,\mathbf{p}) - q_T(x,\mathbf{p})\| \\ &\overset{\text{Eq. 9.37, 9.36}}{\leq} (1-\kappa)\,\|x\| + \frac{\|\mathbf{p}\|}{\sqrt{2m_2}} + 6\theta\, T\, \frac{M_2}{\sqrt{m_2}}\left[\sqrt{M_2}\|x\| + \frac{1}{\sqrt{2}}\|\mathbf{p}\|\right] \\ &\leq \left(1-\kappa + 6\theta\, T\, \frac{M_2}{\sqrt{m_2}}\sqrt{M_2}\right)\|x\| + \left(\frac{1}{\sqrt{2m_2}} + \frac{1}{\sqrt{2}} 6\theta\, T\, \frac{M_2}{\sqrt{m_2}}\right)\|\mathbf{p}\|.\end{aligned}$$

This proves Inequality (9.23). Inequality (9.23) in turn implies that

$$\|X_1'\| \leq \left(1-\kappa+6\theta\, T\, \frac{M_2^{1.5}}{\sqrt{m_2}}\right)\|X_0\| + (1+6\theta\, T\, M_2)\,\frac{\|\mathfrak{p}_0\|}{\sqrt{2m_2}}$$

and so

$$e^{\lambda\|X_1'\|} \leq e^{\lambda\|X_0\|} e^{\lambda\left(-\kappa+6\theta\, T\, \frac{M_2^{1.5}}{\sqrt{m_2}}\right)\|X_0\| + \lambda(1+6\theta\, T\, M_2)\,\frac{\|\mathfrak{p}_0\|}{\sqrt{2m_2}}}.$$

Since $6\theta\, T\, \frac{M_2}{\sqrt{m_2}}\sqrt{M_2} < \kappa$, this implies

$$e^{\lambda\|X_1'\|} \leq \begin{cases}(1-\alpha'')e^{\lambda\|X_0\|}, & \|X_0\| \geq C' \\ e^{\lambda\|X_1'\|} \leq \beta'' & \|X_0\| \leq C',\end{cases}$$

where

$$\begin{aligned}\alpha'' &:= 1 - e^{\lambda\left(-\kappa+6\theta\, T\, \frac{M_2^{1.5}}{\sqrt{m_2}}\right)C' + \lambda(1+6\theta\, T\, M_2)\,\frac{\|\mathfrak{p}_0\|}{\sqrt{2m_2}}} \\ \beta'' &:= e^{\lambda\left(1-\kappa+6\theta\, T\, \frac{M_2^{1.5}}{\sqrt{m_2}}\right)C' + \lambda(1+6\theta\, T\, M_2)\,\frac{\|\mathfrak{p}_0\|}{\sqrt{2m_2}}}\end{aligned}$$

are random variables that depend on $\mathfrak{p}_0$. Thus,

$$e^{\lambda\|X_1'\|} \leq (1-\alpha'')e^{\lambda\|X_0\|} + \beta''$$

for any X_0, and so

$$\begin{aligned}\mathbb{E}[e^{\lambda\|X_1'\|}] &\leq \mathbb{E}[(1-\alpha'')e^{\lambda\|X_0\|} + \beta''] \\ &= \mathbb{E}[1-\alpha'']\, e^{\lambda\|X_0\|} + \mathbb{E}[\beta''] \\ &= \mathbb{E}\left[e^{\lambda\left(-\kappa+6\theta\, T\, \frac{M_2^{1.5}}{\sqrt{m_2}}\right)C' + \lambda(1+6\theta\, T\, M_2)\,\frac{\|\mathfrak{p}_0\|}{\sqrt{2m_2}}}\right] e^{\lambda\|X_0\|} \\ &\qquad + \mathbb{E}\left[e^{\lambda\left(1-\kappa+6\theta\, T\, \frac{M_2^{1.5}}{\sqrt{m_2}}\right)C' + \lambda(1+6\theta\, T\, M_2)\,\frac{\|\mathfrak{p}_0\|}{\sqrt{2m_2}}}\right]\end{aligned} \tag{9.38}$$

$$\overset{\text{Eq. 9.31}}{\leq} (1+e^{-\frac{d}{8}})e^{\lambda\left(-\kappa+6\theta\, T\, \frac{M_2^{1.5}}{\sqrt{m_2}}\right) C'+\lambda\left(\frac{1}{\sqrt{2m_2}}+\frac{1}{\sqrt{2}}6\theta\, T\, \frac{M_2}{\sqrt{m_2}}\right)\sqrt{2d}} e^{\lambda\|X_0\|}$$
$$+ (1+e^{-\frac{d}{8}})e^{\lambda\left(1-\kappa+6\theta\, T\, \frac{M_2^{1.5}}{\sqrt{m_2}}\right) C'+\lambda\left(\frac{1}{\sqrt{2m_2}}+\frac{1}{\sqrt{2}}6\theta\, T\, \frac{M_2}{\sqrt{m_2}}\right)\sqrt{2d}},$$

where Inequality (9.31) is applied in the fourth line with $\gamma = \lambda(1 + 6\theta T\, M_2)$ (note that the condition on γ is satisfied, since $6\theta T\, M_2 \leq 6\theta T\, \frac{M_2^{1.5}}{\sqrt{m_2}} < \kappa < 1$ implies $\gamma < 2\lambda \leq 2 \times \frac{1}{32}\sqrt{2m_2} = \frac{1}{16}\sqrt{2m_2}$). Therefore, since $X_0 = X_0' = x$,

$$\mathbb{E}[V(X_1')] \leq (1-\alpha)V(X_0') + \beta,$$

and hence

$$Q(x,\cdot)[V] \leq (1-\alpha) \times V(x) + \beta. \tag{9.39}$$

This proves the second line of Inequality (9.22), completing the proof of the lemma. □

We show that we can use the approximation bound in Lemma 9.4:

Lemma 9.6. *Set notation and parameters as in the statement of Lemma 9.5. Then, using the same notation, the assumptions* (9.6), (9.7) *and* (9.8) *of Lemma 9.4 hold for any choice of* $C > 0$ *and the choice*

$$\delta = 6\theta\, T\, \frac{M_2}{\sqrt{m_2}} \left(\sqrt{M_2}\lambda^{-1}\log(C) + \frac{\sqrt{d}}{\sqrt{2}}\right).$$

Proof. Inequalities (9.6) and (9.7) follow immediately from Theorem 3 and Lemma 9.5, respectively. So we need only prove Inequality (9.8). Fix $x \in \mathbb{R}^d$ and let $\mathfrak{p}_0 \sim \mathcal{N}(0, I_d)$ be a standard spherical Gaussian. Set $X_1 := q_T(x, \mathfrak{p}_0)$ and $X_1' := q_T^{\dagger\theta}(x, \mathfrak{p}_0)$, so that $X_1 \sim K(x,\cdot)$ and $X_1' \sim Q(x,\cdot)$. By Lemma 8.2,

$$\|X_1' - X_1\| = \|q_T^{\dagger\theta}(x, \mathfrak{p}_0) - q_T(x, \mathfrak{p}_0)\| \leq 6\theta\, T\, \frac{M_2}{\sqrt{m_2}}\sqrt{H(x, \mathfrak{p}_0)}. \tag{9.40}$$

Therefore,

$$\begin{aligned}
\sup_{x\,:\,V(x)\leq C} W_1(K(x,\cdot), Q(x,\cdot)) &\leq \sup_{x\,:\,V(x)\leq C} W_2(K(x,\cdot), Q(x,\cdot)) \\
&\overset{\text{Eq. 9.40}}{\leq} \sup_{x\,:\,e^{\lambda\|x\|}\leq C} \mathbb{E}\left[\left(6\theta \times T \times \frac{M_2}{\sqrt{m_2}}\sqrt{H(x,\mathfrak{p}_0)}\right)^2\right]^{\frac{1}{2}} \\
&\overset{\text{Assumption 2.2}}{\leq} \sup_{x\,:\,\|x\|\leq\lambda^{-1}\log(C)} \mathbb{E}\left[\left(6\theta \times T \times \frac{M_2}{\sqrt{m_2}}\sqrt{M_2\|x\|^2 + \frac{1}{2}\|\mathfrak{p}_0\|^2}\right)^2\right]^{\frac{1}{2}} \\
&= \mathbb{E}\left[\left(6\theta \times T \times \frac{M_2}{\sqrt{m_2}}\sqrt{M_2(\lambda^{-1}\log(C))^2 + \frac{1}{2}\|\mathfrak{p}_0\|^2}\right)^2\right]^{\frac{1}{2}} \\
&= 6\theta \times T \times \frac{M_2}{\sqrt{m_2}} \times \left(M_2(\lambda^{-1}\log(C))^2 + \frac{1}{2}d\right)^{\frac{1}{2}} \\
&\leq 6\theta \times T \times \frac{M_2}{\sqrt{m_2}} \times \left(\sqrt{M_2}\lambda^{-1}\log(C) + \frac{\sqrt{d}}{\sqrt{2}}\right).
\end{aligned}$$

This completes the proof of the lemma. □

We conclude with a bound on the approximation error of Q after many steps:

Lemma 9.7. *Set notation and parameters as in Lemma 9.5, fix $C > \frac{4\beta}{\alpha}$, and let*

$$\delta = 6\theta\, T\, \frac{M_2}{\sqrt{m_2}} \left(\sqrt{M_2}\lambda^{-1}\log(C) + \frac{1}{\sqrt{2}}\sqrt{d} \right).$$

Then Q satisfies

$$W_1(Q^{\mathcal{I}}(x,\cdot),\pi) \leq (1-\kappa)^s\, 2\lambda^{-1}\log(C) + \frac{\delta}{\kappa} + \frac{\beta(s+1)}{\lambda\alpha C}\,(8+2\log(C))$$

for all $0 \leq s \leq \mathcal{I} \in \mathbb{N}$ and all x satisfying $V(x) \leq \frac{\beta}{\alpha}$.

Proof. Set $V(x) \equiv e^{\lambda\|x\|}$ for all $x \in \mathbb{R}^d$. The proof now follows by applying Lemmas 9.4 and 9.6, with constants given in the statement of Lemma 9.6. □

Define the function

$$\Gamma(a,b) := \left\lceil 2e^{\frac{1}{32}(4-2a)\,\frac{16}{a}\log(1+e^{-\frac{b}{8}})-\frac{7}{8}} \right\rceil \tag{9.41}$$

for $a, b > 0$. The following is essentially a restatement of Lemma 9.7:

Lemma 9.8. *Choose $\epsilon > 0$. Fix notation as in Lemma 9.5, with the additional constraints $0 \leq T \leq \frac{1}{2\sqrt{2}}\frac{\sqrt{m_2}}{M_2}$ and $\lambda = \frac{\kappa}{64}\sqrt{\frac{m_2}{d}}$. Set*

$$\begin{aligned} s &= \frac{1}{\kappa}\log\left(\frac{24\log(15000\,\Gamma(\kappa,d)\,\lambda^{-1}\epsilon^{-1}\kappa^{-2}) + 24\log(\kappa^{-1})}{\lambda\epsilon}\right) \\ C &= (15000\,\Gamma(\kappa,d)\,\lambda^{-1}\epsilon^{-1}(s+1)\kappa^{-2}+4)^2 \\ \theta &\leq \frac{\kappa\epsilon\sqrt{m_2}}{18TM_2(\sqrt{M_2}\lambda^{-1}\log(C)+\frac{1}{\sqrt{2}}\sqrt{d})}. \end{aligned}$$

Then

$$W_1(Q^{\mathcal{I}}(x,\cdot),\pi) \leq \epsilon \tag{9.42}$$

for all $\mathcal{I} \geq s$ and all x satisfying $\|x\| \leq \frac{\sqrt{d}}{\sqrt{m_2}}$. Furthermore, if Q is ergodic with stationary measure ν,

$$W_1(\pi,\nu) \leq \epsilon. \tag{9.43}$$

Proof. By our assumption about the value of θ, and recalling that $0 < \kappa < 1$ and $0 < m_2 \leq M_2$, we have

$$6\theta\, T\, \frac{{M_2}^{1.5}}{\sqrt{m_2}} \leq \frac{1}{2}\kappa \quad \text{and} \quad 6\theta\, T\, \frac{M_2}{\sqrt{m_2}} \leq \frac{1}{2\sqrt{M_2}}\kappa \leq \frac{1}{\sqrt{m_2}}. \tag{9.44}$$

We now choose $C' = (1+\frac{16}{\kappa}\log(1+e^{-\frac{d}{8}}))\,\frac{8}{\kappa}\sqrt{\frac{d}{m_2}}$. By Inequalities (9.44), we have after some algebraic manipulation

$$\begin{aligned} 1-\alpha &= (1+e^{-\frac{d}{8}})e^{\lambda\left(-\kappa+6\theta\,T\,\frac{{M_2}^{1.5}}{\sqrt{m_2}}\right)C'+\lambda(1+6\theta\,T\,M_2)\sqrt{\frac{d}{m_2}}} \\ &\leq (1+e^{-\frac{d}{8}})e^{-\frac{1}{2}\lambda\kappa C'+\lambda\left(\frac{1}{\sqrt{2m_2}}+\frac{1}{\sqrt{2m_2}}\right)\times\sqrt{2d}} \\ &= (1+e^{-\frac{d}{8}})e^{-\frac{1}{2}\lambda\kappa\times(1+\frac{16}{\kappa}\log(1+e^{-\frac{d}{8}}))\times\frac{4}{\kappa}\times\left(\frac{1}{\sqrt{2m_2}}+\frac{1}{\sqrt{2m_2}}\right)\times\sqrt{2d}+\lambda\left(\frac{1}{\sqrt{2m_2}}+\frac{1}{\sqrt{2m_2}}\right)\times\sqrt{2d}} \\ &\leq e^{-\frac{1}{32}\kappa}. \end{aligned}$$

Also by Inequalities (9.44),

$$\begin{aligned}
\beta &= e^{\lambda C'}(1-\alpha) \leq (1+e^{-\frac{d}{8}})e^{\lambda\left(1-\frac{1}{2}\kappa\right)\times C' + 2\lambda\sqrt{\frac{d}{m_2}}} \\
&= (1+e^{-\frac{d}{8}})e^{\lambda\left(1-\frac{1}{2}\kappa\right)\times(1+\frac{16}{\kappa}\log(1+e^{-\frac{d}{8}}))\times\frac{4}{\kappa}\times\left(\frac{1}{\sqrt{2m_2}}+\frac{1}{\sqrt{2m_2}}\right)\times\sqrt{2d}+\lambda\left(\frac{1}{\sqrt{2m_2}}+\frac{1}{\sqrt{2m_2}}\right)\times\sqrt{2d}} \\
&= (1+e^{-\frac{d}{8}})e^{\frac{1}{32}(4-2\kappa)\times(1+\frac{16}{\kappa}\log(1+e^{-\frac{d}{8}}))+\frac{1}{32}\kappa} \\
&= (1+e^{-\frac{d}{8}})e^{\frac{1}{32}(4-2\kappa)\times\frac{16}{\kappa}\log(1+e^{-\frac{d}{8}})} \times e^{\frac{1}{32}(4-\kappa)} \\
&\leq 2e^{\frac{1}{32}(4-2\kappa)\times\frac{16}{\kappa}\log(1+e^{-\frac{d}{8}})} \times e^{\frac{1}{32}(4-\kappa)}.
\end{aligned}$$

Combining these two calculations,

$$\begin{aligned}
\frac{\beta}{\alpha} &\leq \frac{2e^{\frac{1}{32}(4-2\kappa)\times\frac{16}{\kappa}\log(1+e^{-\frac{d}{8}})} \times e^{\frac{1}{32}(4-\kappa)}}{1-e^{-\frac{1}{32}\kappa}} \\
&= \left[2e^{\frac{1}{32}(4-2\kappa)\times\frac{16}{\kappa}\log(1+e^{-\frac{d}{8}})-\frac{7}{8}}\right] \times \frac{e^{1-\frac{1}{32}\kappa}}{1-e^{-\frac{1}{32}\kappa}} = \Gamma(\kappa,d)\times\frac{e^{1-\frac{1}{32}\kappa}}{1-e^{-\frac{1}{32}\kappa}}.
\end{aligned}$$

Since $T \leq \frac{1}{2\sqrt{2}}\frac{\sqrt{m_2}}{M_2}$, we have $0 < \frac{1}{32}\kappa \leq \frac{1}{32} \times \frac{1}{16} < \frac{1}{3}$ and so this implies

$$\frac{\beta}{\alpha} \leq \Gamma(\kappa,d)\times\frac{e^{1-\frac{1}{32}\kappa}}{1-e^{-\frac{1}{32}\kappa}} \leq \Gamma(\kappa,d)\times\frac{1}{(\frac{1}{32}\kappa)^2}. \tag{9.45}$$

Note that $C > 16$, so we have the easy inequality

$$\begin{aligned}
C &= (15000\,\Gamma(\kappa,d)\,\lambda^{-1}\epsilon^{-1}(s+1)\kappa^{-2}+4)^2 \\
&\geq 15000\,\Gamma(\kappa,d)\,\lambda^{-1}\epsilon^{-1}(s+1)\kappa^{-2}\,\left(8+\log(C)\right).
\end{aligned}$$

Applying Inequality (9.45), this implies

$$\frac{\beta(s+1)}{\lambda\alpha C}\left(8+2\log(C)\right) \leq \frac{1}{3}\epsilon. \tag{9.46}$$

Our assumption on θ also implies that

$$\delta \leq \frac{1}{3}\kappa\epsilon. \tag{9.47}$$

Finally, our assumption about s implies that $s \geq \frac{1}{\kappa}\log\left(\frac{6\log(C)}{\lambda\epsilon}\right)$ and hence that

$$(1-\kappa)^s \times 2\lambda^{-1}\log(C) \leq \frac{1}{3}\epsilon. \tag{9.48}$$

Therefore, Inequalities (9.46), (9.47), and (9.48) together imply that

$$(1-\kappa)^s \times 2\lambda^{-1}\log(C) + \frac{\delta}{\kappa} + \frac{\beta(s+1)}{\lambda\alpha C}\left(8+2\log(C)\right) \leq \frac{1}{3}\epsilon+\frac{1}{3}\epsilon+\frac{1}{3}\epsilon = \epsilon \qquad . \tag{9.49}$$

Inequality (9.45) also implies that $C > 4\frac{\beta}{\alpha}$. Moreover, since $0 < \kappa \leq \frac{1}{16}$, our choice of λ satisfies $0 < \lambda \leq \frac{1}{32}\sqrt{2m_2}$. Therefore, by Lemma 9.7 and Inequality (9.49), for all x satisfying $V(x) \leq \frac{\beta}{\alpha}$, we have

$$W_1(Q^{\mathcal{I}}(x,\cdot),\pi) \leq \epsilon \quad \text{and} \quad W_1(\pi,\nu) \leq \epsilon.$$

This completes the proof of the desired mixing bounds for starting point x satisfying $V(x) \leq \frac{\beta}{\alpha}$. To complete the proof of the theorem, then, it is enough to show that $V(x) \leq \frac{\beta}{\alpha}$ for all x satisfying $\|x\| \leq \sqrt{\frac{d}{m_2}}$. To do so, we must find a lower bound for $\frac{\beta}{\alpha}$.

By Inequality (9.34) and the trivial bounds $\alpha \leq 1$ and $\lambda, (1-\kappa), C' \geq 0$, we have

$$\frac{\beta}{\alpha} \geq \beta \geq (1 + e^{-\frac{d}{8}}) e^{\lambda \frac{\sqrt{2d}}{\sqrt{2m_2}}} e^{\lambda(1-\kappa)C'} \geq e^{\lambda \frac{\sqrt{d}}{\sqrt{m_2}}}.$$

Hence, the inequality $V(x) \leq \frac{\beta}{\alpha}$ holds if $e^{\lambda \|x\|} \leq e^{\lambda \frac{\sqrt{d}}{\sqrt{m_2}}}$. That is,

$$\{x : \|x\| \leq \sqrt{\frac{d}{m_2}}\} \subset \{x : V(x) \leq \frac{\beta}{\alpha}\}. \tag{9.50}$$

This completes the proof. ☐

We can now prove Theorem 1:

Proof of Theorem 1. Set notation as in Lemma 9.8 and let

$$\theta_0 = \frac{\kappa \epsilon \sqrt{m_2}}{18 T M_2 (\sqrt{M_2} \lambda^{-1} \log(C) + \frac{1}{\sqrt{2}} \sqrt{d})}.$$

This theorem is then an immediate consequence of Lemma 9.8. ☐

References

[1] Pierre Alquier, Nial Friel, Richard Everitt, and Aidan Boland. Noisy Monte Carlo: Convergence of Markov chains with approximate transition kernels. *Statistics and Computing*, 26(1-2):29–47, 2016.

[2] Alexandros Beskos, Natesh Pillai, Gareth Roberts, Jesus-Maria Sanz-Serna, and Andrew Stuart. Optimal tuning of the hybrid Monte Carlo algorithm. *Bernoulli*, 19(5A):1501–1534, 2013.

[3] François Bolley, Arnaud Guillin, and Florent Malrieu. Trend to equilibrium and particle approximation for a weakly selfconsistent Vlasov-Fokker-Planck equation. *ESAIM: Mathematical Modelling and Numerical Analysis*, 44(5):867–884, 2010.

[4] Christian Borgs, Jennifer T Chayes, Alan Frieze, Jeong Han Kim, Prasad Tetali, Eric Vigoda, and Van Ha Vu. Torpid mixing of some Monte Carlo Markov chain algorithms in statistical physics. In *40th Annual Symposium on Foundations of Computer Science*, pages 218–229. IEEE, 1999.

[5] Nawaf Bou-Rabee, Andreas Eberle, and Raphael Zimmer. Coupling and convergence for Hamiltonian Monte Carlo. *arXiv preprint arXiv:1805.00452*, 2018.

[6] Roger Brockett. Oscillatory descent for function minimization. In *Current and future directions in applied mathematics*, pages 65–82. Springer, 1997.

[7] Russ Bubley and Martin Dyer. Path coupling: A technique for proving rapid mixing in markov chains. In *Foundations of Computer Science, 1997. Proceedings., 38th Annual Symposium on*, pages 223–231. IEEE, 1997.

[8] Simone Calogero. Exponential convergence to equilibrium for kinetic fokker-planck equations. *Communications in Partial Differential Equations*, 37(8):1357–1390, 2012.

[9] Xiang Cheng, Niladri S. Chatterji, Peter L. Bartlett, and Michael I. Jordan. Underdamped Langevin MCMC: A non-asymptotic analysis. In Sébastien Bubeck, Vianney Perchet, and Philippe Rigollet, editors, *Proceedings of the 31st Conference On Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pages 300–323. PMLR, 06–09 Jul 2018.

[10] Sai Hung Cheung and James L Beck. Bayesian model updating using hybrid Monte Carlo simulation with application to structural dynamic models with many uncertain parameters. *Journal of engineering mechanics*, 135(4):243–255, 2009.

[11] Arnak S Dalalyan. Theoretical guarantees for approximate sampling from smooth and log-concave densities. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(3):651–676, 2017.

[12] Persi Diaconis. The Markov chain Monte Carlo revolution. *Bulletin of the American Mathematical Society*, 46(2):179–205, 2009.

[13] Persi Diaconis, Kshitij Khare, and Laurent Saloff-Coste. Gibbs sampling, exponential families and orthogonal polynomials. *Statistical Science*, 23(2):151–178, 2008.

[14] Jean Dolbeault, Clément Mouhot, and Christian Schmeiser. Hypocoercivity for linear kinetic equations conserving mass. *Transactions of the American Mathematical Society*, 367(6):3807–3828, 2015.

[15] C dric Villani. *Hypocoercivity*. Number 949-951. American Mathematical Soc., 2009.

[16] Alain Durmus and Eric Moulines. High-dimensional Bayesian inference via the unadjusted Langevin algorithm. *arXiv preprint arXiv:1605.01559*, 2016.

[17] Alain Durmus and Eric Moulines. Non-asymptotic convergence analysis for the Unadjusted Langevin Algorithm. *The Annals of Applied Probability*, 27(3):1551–1587, 2017.

[18] Raaz Dwivedi, Yuansi Chen, Martin J Wainwright, and Bin Yu. Log-concave sampling: Metropolis-Hastings algorithms are fast! *To appear in Proceedings of COLT 2018, arXiv preprint arXiv:1801.02309*, 2018.

[19] Andreas Eberle, Arnaud Guillin, and Raphael Zimmer. Couplings and quantitative contraction rates for Langevin dynamics. *arXiv preprint arXiv:1703.01617*, 2017.

[20] D. Ferré, L. Hervé, and J. Ledoux. Regular perturbation of V-geometrically ergodic Markov chains. *Journal of Applied Probability*, 50(1):184–194, 2013.

[21] Mark Girolami and Ben Calderhead. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(2):123–214, 2011.

[22] Martin Hairer, Andrew M Stuart, and Sebastian J Vollmer. Spectral gaps for a Metropolis–Hastings algorithm in infinite dimensions. *The Annals of Applied Probability*, 24(6):2455–2490, 2014.

[23] David Lee Hanson and Farroll Tim Wright. A bound on tail probabilities for quadratic forms in independent random variables. *The Annals of Mathematical Statistics*, 42(3):1079–1083, 1971.

[24] Jeremy Heng and Pierre E Jacob. Unbiased Hamiltonian Monte Carlo with couplings. *arXiv preprint arXiv:1709.00404*, 2017.

[25] Frédéric Hérau and Francis Nier. Isotropic hypoellipticity and trend to equilibrium for the Fokker-Planck equation with a high-degree potential. *Archive for Rational Mechanics and Analysis*, 171(2):151–218, 2004.

[26] James E Johndrow, Jonathan C Mattingly, Sayan Mukherjee, and David Dunson. Approximations of Markov chains and high-dimensional Bayesian inference. *arXiv preprint arXiv:1508.03387*, 2015.

[27] Galin L Jones and James P Hobert. Honest exploration of intractable probability distributions via Markov chain Monte Carlo. *Statistical Science*, 16(4):312–334, 2001.

[28] Yin Tat Lee and Santosh S Vempala. Convergence rate of Riemannian Hamiltonian Monte Carlo and faster polytope volume computation. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 1115–1121. ACM, 2018.

[29] David Asher Levin, Yuval Peres, and Elizabeth Lee Wilmer. *Markov chains and mixing times*. American Mathematical Soc., 2009.

[30] Samuel Livingstone, Michael Betancourt, Simon Byrne, and Mark Girolami. On the geometric ergodicity of Hamiltonian Monte Carlo. *arXiv preprint arXiv:1601.08057*, 2016.

[31] Oren Mangoubi, Natesh Pillai, and Aaron Smith. Does Hamiltonian Monte Carlo mix faster than a random walk on multimodal densities? *arXiv preprint arXiv:1808.03230*, 2018.

[32] Oren Mangoubi, Natesh Pillai, and Aaron Smith. Simple conditions for metastability of continuous Markov chains. *arXiv preprint arXiv:1808.03239*, 2018.

[33] Oren Mangoubi and Aaron Smith. Mixing of Hamiltonian Monte Carlo on strongly log-concave distributions 1: Continuous dynamics. *preprint*, 2017.

[34] Oren Mangoubi and Aaron Smith. Rapid mixing of Hamiltonian Monte Carlo on strongly log-concave distributions. *arXiv preprint*, 2017.

[35] Oren Mangoubi and Aaron Smith. Rapid mixing of geodesic walks on manifolds with positive curvature. *The Annals of Applied Probability*, 28(4):2501–2543, 2018.

[36] Oren Mangoubi and Nisheeth K Vishnoi. Dimensionally tight running time bounds for second-order Hamiltonian Monte Carlo. In *Advances in Neural Information Processing Systems*, 2018.

[37] B Mehlig, DW Heermann, and BM Forrest. Hybrid Monte Carlo method for condensed-matter systems. *Physical Review B*, 45(2):679, 1992.

[38] Stéphane Mischler and Clément Mouhot. Exponential stability of slowly decaying solutions to the kinetic Fokker-Planck equation. *Archive for rational mechanics and analysis*, 221(2):677–723, 2016.

[39] Radford M Neal. MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, 2:113–162, 2011.

[40] Jeffrey Negrea and Jeffrey S Rosenthal. Error bounds for approximations of geometrically ergodic Markov chains. *arXiv preprint arXiv:1702.07441*, 2017.

[41] Yann Ollivier. Ricci curvature of Markov chains on metric spaces. *Journal of Functional Analysis*, 256(3):810–864, 2009.

[42] Michela Ottobre, Natesh S Pillai, Frank J Pinski, and Andrew M Stuart. A function space HMC algorithm with second order Langevin diffusion limit. *Bernoulli*, 22(1):60–106, 2016.

[43] Natesh S Pillai and Aaron Smith. Ergodicity of approximate MCMC chains with applications to large data sets. *arXiv preprint arXiv:1405.0182*, 2014.

[44] Gareth O Roberts and Jeffrey S Rosenthal. Complexity bounds for Markov chain Monte Carlo algorithms via diffusion limits. *J. Appl. Prob*, 53(2):1–11, 2016.

[45] Mark Rudelson and Roman Vershynin. Hanson-wright inequality and sub-Gaussian concentration. *Electron. Commun. Probab*, 18(82):1–9, 2013.

[46] Daniel Rudolf and Nikolaus Schweizer. Perturbation theory for Markov chains via Wasserstein distance. *Bernoulli*, 24(4A):2610–2639, 2018.

[47] Christof Seiler, Simon Rubinstein-Salzedo, and Susan Holmes. Positive curvature and Hamiltonian Monte Carlo. In *Advances in Neural Information Processing Systems*, pages 586–594, 2014.

[48] Santosh Vempala. Geometric random walks: a survey. *Combinatorial and computational geometry*, 52(573-612):2, 2005.

[49] Cédric Villani. Hypocoercive diffusion operators. In *International Congress of Mathematicians*, volume 3, pages 473–498, 2006.

[50] Raphael Zimmer. Explicit contraction rates for a class of degenerate and infinite-dimensional diffusions. *Stochastics and Partial Differential Equations: Analysis and Computations*, 5(3):368–399, 2017.