COUPLING AND CONVERGENCE FOR HAMILTONIAN MONTE CARLO

By Nawaf Bou-Rabee[1], Andreas Eberle[2,*] and Raphael Zimmer[2,**]

[1]*Department of Mathematical Sciences, Rutgers University–Camden, nawaf.bourabee@rutgers.edu*

[2]*Institut für Angewandte Matehmatik, Universität Bonn, *eberle@uni-bonn.de; **raphael.zimmer@uni-bonn.de*

Based on a new coupling approach, we prove that the transition step of the Hamiltonian Monte Carlo algorithm is contractive w.r.t. a carefully designed Kantorovich (L^1 Wasserstein) distance. The lower bound for the contraction rate is explicit. Global convexity of the potential is not required, and thus multimodal target distributions are included. Explicit quantitative bounds for the number of steps required to approximate the stationary distribution up to a given error ϵ are a direct consequence of contractivity. These bounds show that HMC can overcome diffusive behavior if the duration of the Hamiltonian dynamics is adjusted appropriately.

1. Introduction. Markov chain Monte Carlo (MCMC) is a family of methods to approximately sample from an arbitrary probability distribution. In conjunction with Bayesian methods, MCMC has revolutionized statistics and enabled applications of statistical inference to machine learning, pattern recognition, and artificial intelligence [2, 4, 11, 21, 39]. Much of the classical research activity related to MCMC considered techniques based on random walks. Regrettably the meandering behavior of these random walks leads to sampling methods that are slow [12, 22]. Therefore, a more recent focus of MCMC research activity is to develop faster methods by overcoming this random walk or diffusive behavior. Hamiltonian Monte Carlo (HMC), in principle, provides one way to do this [6, 13, 32, 35, 36]. The basic idea is to "give the walker momentum." With this momentum, the walker intermixes periods of fast running with slower running to efficiently explore features of a probability distribution. However, beyond this intuition, the mathematical properties of HMC are not well understood.

In this paper, we use a coupling technique ([29], Chapter 14) to analyse the HMC algorithm. Let μ denote a probability measure on $\mathbb{R}^d$ with nonnormalized density $e^{-U(x)}$. In HMC, the function $U(x)$ is viewed as a potential energy. In its simplest form, the algorithm simulates a Markov chain with state space $\mathbb{R}^d \times \mathbb{R}^d$ and invariant probability measure $\hat{\mu}$ with nonnormalized density $e^{-H(x,v)}$ where $H(x, v) = \frac{1}{2}|v|^2 + U(x)$ is viewed as a Hamiltonian. Below we only consider the first component which is a Markov chain on $\mathbb{R}^d$ with invariant probability measure μ. A transition step of HMC inputs an initial position $x \in \mathbb{R}^d$ and a duration parameter $T > 0$, and outputs a final position by taking the following steps:

Step 1. Draw an initial velocity $\xi \sim \mathcal{N}(0, I_d)$.
Step 2. Run the Hamiltonian dynamics associated to the Hamiltonian function H for a duration T with initial position x and initial velocity ξ.
Step 3. Output the final position of this Hamiltonian dynamics.

We call the algorithm with this transition step *exact HMC*. In practice, the Hamiltonian dynamics in Step 2 is approximated by a numerical integrator. Furthermore, an accept/reject step can be added to remove the bias due to time discretization error. In order to distinguish

these from exact HMC, we call the resulting algorithms *unadjusted and adjusted* numerical HMC, respectively.

Below, we introduce a new coupling between the transition steps of two copies of exact HMC, or numerical HMC, respectively. The approach we use is based on the framework introduced in [18], and the specific coupling of the velocities is strongly inspired by a recently developed coupling for second-order Langevin dynamics [20]. The underlying idea is to couple two copies of HMC at different positions x and y by coupling their velocities ξ and η such that for a positive constant $\gamma \leq T^{-1}$, the event $\xi - \eta = -\gamma(x - y)$ happens with maximal probability, and otherwise, to apply a reflection coupling to the velocities. In particular, the coupling is designed such that in the free case when $U = 0$, the positions approach each other during the transition step with maximal probability. We use this contractive property of the coupling to obtain an explicit contraction rate for HMC in a specially designed Kantorovich (L^1 Wasserstein) metric.

To be more specific, we state a simpified version of one of our main results, which will later be reformulated rigorously as Corollary 2.8—a corollary of the fundamental contraction result in Theorem 2.3. Let $\pi(x, dy)$ denote the one-step transition kernel of exact HMC and let $\mathcal{W}^1$ denote the standard L^1-Wasserstein distance. Assuming sufficient regularity on the potential energy function U (see Assumption 2.1) including that ∇U is globally Lipschitz with Lipschitz constant L, and U is strongly convex outside a Euclidean ball of diameter $\mathcal{R}$ with strong convexity constant K, we prove that if

$$LT^2 \leq \min\left(\frac{K}{L}, \frac{1}{4}, \frac{1}{256L\mathcal{R}^2}\right)$$

then for all initial distributions ν and η, and for all $n \geq 0$,

$$\mathcal{W}^1(\nu\pi^n, \eta\pi^n) \leq Me^{-cn}\mathcal{W}^1(\nu, \eta)$$

$$\text{where } c = \frac{1}{10}\min\left(1, \frac{1}{2}KT^2\left(1 + \frac{\mathcal{R}}{T}\right)e^{-\mathcal{R}/(2T)}\right)e^{-2\mathcal{R}/T} \quad \text{and}$$

$$M = e^{\frac{5}{2}(1+\mathcal{R}/T)}.$$

More precisely, we prove in Theorem 2.3 that the transition kernel π is even contractive with contraction rate c w.r.t. an L^1 Wasserstein distance $\mathcal{W}_\rho$ based on an explicit metric ρ that is equivalent to the Euclidean distance. This statement can be used to quantify the speed of convergence of HMC to equilibrium, and it also directly implies completely explicit bias and variance bounds, as well as concentration inequalities for ergodic averages; see, for example, [27].

A remarkable feature of the contraction rate c is that under our hypothesis on LT^2, it only depends on K and $\mathcal{R}/T$. In particular, it does not depend explicitly on the dimension, although frequently there will be a hidden dimension dependence through the values of the constants K and $\mathcal{R}$. If we choose T proportional to $\mathcal{R}$, and assume that K and $L\mathcal{R}^2$ are fixed (which excludes the possibility of high energy barriers), then the rate does not deteriorate as $\mathcal{R}$ increases. Noting that the Hamiltonian dynamics is run for time T during each transition step, we can conclude that a given approximation accuracy can be obtained after running the dynamics for a total time of kinetic order $O(\mathcal{R})$, where $\mathcal{R}$ basically is the diameter of a ball where the distribution μ concentrates in. On the other hand, a Random Walk based method would require a time of diffusive order $O(\mathcal{R}^2)$. Hence if T is chosen adequately then HMC can indeed overcome diffusive behavior.

In Theorems 2.4 and 2.12, we extend our results to numerical HMC with a velocity Verlet integrator. The corresponding results are more involved than for exact HMC, but the bound c for the contraction rate is the same provided the time discretization step size h is chosen

sufficiently small depending on the other parameters, including the dimension. We consider both adjusted and unadjusted numerical HMC. In the unadjusted case, the contraction bounds are easier to derive and take a nicer form that is close to the corresponding results for exact HMC, but the price to pay is an additional bias due to the fact that the invariant measure for unadjusted HMC does not coincide with μ. The resulting error has to be controlled by other techniques that are out of the scope of this work; see [14, 34]. In Theorems 2.6 and 2.7, we also state versions of our main results where the asymptotic strong convexity assumption is replaced by a Foster–Lyapunov drift condition, which permits U that are only asymptotically convex.

Several recent works have studied ergodic properties of HMC methods. In [5], geometric ergodicity has been proven for a variant of exact HMC (called randomized HMC) where the lengths of the durations of the Hamiltonian dynamics at the different transitions of the Markov chain are i.i.d. exponential random variables with mean T. The proof relies on Harris' theorem, which requires a (local) version of Doeblin's condition: a minorization condition for the transition probabilities at a finite time and in a compact set. Unfortunately, given the complicated form of these transition probabilities, the minorization condition involves nonexplicit constants, and in particular, the dependence of the convergence rate on parameters in HMC is unclear. We remark that randomized HMC is related to Anderson's dynamics, which describes a molecular system interacting with a heat bath [1, 16, 30]. Convergence of Anderson's dynamics on an n-torus was proven in Ref. [16] by showing that Doeblin's condition holds. Recently, geometric ergodicity for HMC, but without explicit rates, has been shown by Durmus, Moulines and Saksman [15], cf. also [3] for a related work.

Closely related to our results is [34], which significantly extends ideas from [37]. In [34], Mangoubi and Smith apply coupling techniques in order to analyse the properties of HMC in high dimension under the assumption of strong convexity of U; see also [8] for related work on second-order Langevin dynamics. They obtain quite sharp results on the dimension dependence caused by different numerical integrators of the Hamiltonian dynamics, at the price of imposing very restrictive assumptions on U. Complementary to this, the approach presented here is more broadly applicable and offers significant flexibility for further extensions. It provides a suitable basic framework for studying dimension dependence caused by intrinsic properties of the underlying model, but in contrast to [34], we do not provide sharp bounds for the dimension dependence caused by time discretization. A major difference of our approach to [8, 34] is that these works rely on synchronous couplings of the initial velocities in HMC, that is, they set $\eta = \xi$. This simplifies the analysis considerably, but as a consequence, the couplings are contractive only if the stationary distribution is strongly log-concave. Another difference is that in [34] and [8], the coupling is applied to the exact dynamics, whereas the numerical discretization is controlled by a perturbative approach. In contrast, the coupling introduced below is contractive both for exact and numerical HMC. Its superiority to synchronous couplings is supported both by theoretical results and by numerical simulations. In connection with [25], the coupling may also be useful to parallelize HMC.

Let us finally remark that the Hamiltonian flow is what, in principle, enables HMC to make large moves in state space that reduce correlations in the resulting Markov chain. One might hope that, by increasing the duration T further, the final position moves even further away from the initial position, thus reducing correlation. However, simple examples show that this outcome is far from assured. For example, for a standard normal distribution, the corresponding Hamiltonian flow is a planar rotation with period 2π. It is easy to see that, if the initial position is taken from the target distribution, as T increases from 0 to $\pi/2$, the correlation between the initial and final positions decreases and for $T = \pi/2$, the initial and final positions are independent. However increasing T beyond $\pi/2$ will cause an increase in the correlation and for $T = \pi$, the chain is not even ergodic. For general distributions, it is likely that a small

T will lead to a highly correlated chain, while choosing T too large may cause the Hamiltonian trajectory to make a U-turn and fold back on itself, thus increasing correlation [26]. Generally speaking the performance of HMC may be very sensitive to changes in T as first noted by Mackenzie in [33]. This sensitivity is reflected in our conditions on the duration parameter T.

2. Main results.

2.1. *Hamiltonian Monte Carlo.* Fix a function $U \in C^4(\mathbb{R}^d)$ such that $\int \exp(-U(x)) \, dx < \infty$, and denote by

$$(1) \qquad H(x, v) = U(x) + \frac{1}{2}|v|^2, \quad x, v \in \mathbb{R}^d,$$

the corresponding Hamiltonian. Hamiltonian Monte Carlo (HMC) is an MCMC method for approximate sampling from probability measures of the form

$$(2) \qquad \mu(dx) = \mathcal{Z}^{-1} \exp(-U(x)) \, dx, \qquad \hat{\mu}(dx \, dv) = \hat{\mathcal{Z}}^{-1} \exp(-H(x, v)) \, dx \, dv,$$

on $\mathbb{R}^d$, $\mathbb{R}^d \times \mathbb{R}^d$, respectively, where $\mathcal{Z} = \int \exp(-U(x)) \, dx$ and $\hat{\mathcal{Z}} = (2\pi)^{d/2}\mathcal{Z}$.

We consider HMC as a Markov chain on $\mathbb{R}^d$ (not on the phase space $\mathbb{R}^d \times \mathbb{R}^d$). The transition step from x is given by $x \mapsto X'(x)$ with

$$(3) \qquad X'(x) = q_T(x, \xi)I_{A(x)} + xI_{A(x)^C}.$$

Here the duration $T : \Omega \to \mathbb{R}_+$ is in general a random variable on the underlying probability space $(\Omega, \mathcal{A}, P)$ with a given distribution ν (e.g., $\nu = \delta_s$ or $\nu = \text{Exp}(\lambda^{-1})$), $\xi \sim N(0, I_d)$ and $\mathcal{U} \sim \text{Unif}(0, 1)$ are independent random variables, and the acceptance event for a proposed transition is either

$$(4) \qquad A(x) = \{\mathcal{U} \le \exp(H(x, \xi) - H(q_T(x, \xi), p_T(x, \xi)))\} \quad \text{or} \quad A(x) = \Omega,$$

corresponding to adjusted and unadjusted HMC, respectively. Below we only consider the case where $T \in (0, \infty)$ is a given deterministic constant. Moreover,

$$\phi_t(x, v) = (q_t(x, v), p_t(x, v)) \quad (t \in [0, \infty), x, v \in \mathbb{R}^d)$$

is the exact Hamiltonian flow, or a numerical approximation of the Hamiltonian flow. The exact Hamiltonian flow is the solution of the ODE

$$(5) \qquad \frac{d}{dt}q_t = p_t, \qquad \frac{d}{dt}p_t = -\nabla U(q_t), \qquad (q_0(x, v), p_0(x, v)) = (x, v).$$

The corresponding Markov chain with transition step determined by (3), (4) and (5) is called *exact HMC*. Notice that for exact HMC, $H(q_T(x, \xi), p_T(x, \xi)) = H(x, \xi)$. Hence all proposed transitions are accepted without adjustment, and the transition step is simply given by

$$(6) \qquad X'(x) = q_T(x, \xi).$$

In practice, the Hamiltonian flow has to be approximated by a numerical integrator. Here, we focus on the *velocity Verlet* integrator with discretization step size $h > 0$. In this case, $\phi_t = (q_t, p_t)$ is the solution of the equation

$$(7) \qquad \frac{d}{dt}q_t = p_{\lfloor t \rfloor_h} - \frac{h}{2}\nabla U(q_{\lfloor t \rfloor_h}), \qquad \frac{d}{dt}p_t = -\frac{1}{2}(\nabla U(q_{\lfloor t \rfloor_h}) + \nabla U(q_{\lceil t \rceil_h}))$$

with initial condition $(q_0(x, v), p_0(x, v)) = (x, v)$, where

$$(8) \qquad \lfloor t \rfloor_h = \max\{s \in h\mathbb{Z} : s \le t\} \quad \text{and} \quad \lceil t \rceil_h = \min\{s \in h\mathbb{Z} : s \ge t\}.$$

The corresponding Markov chain with transition step determined by (3), (4) and (7) is called *adjusted resp. unadjusted* numerical HMC. For brevity, whenever $h > 0$ is fixed, we write $\lfloor t \rfloor$ and $\lceil t \rceil$ instead of $\lfloor t \rfloor_h$ and $\lceil t \rceil_h$, respectively. Since the velocity Verlet integrator does not preserve the Hamiltonian exactly, the rejection event $A(x)^C$ is not empty in general for adjusted numerical HMC. However, for fixed x, the rejection probability goes to 0 as $h \downarrow 0$.

The HMC algorithm induces a time-homogeneous Markov chain on $\mathbb{R}^d$ with transition kernel

$$
(9) \qquad
\begin{aligned}
\pi(x, B) &= P\big[X'(x) \in B\big] \\
&= P\big[\{q_T(x, \xi) \in B\} \cap A(x)\big] + (1 - P\big[A(x)\big])\delta_x(B).
\end{aligned}
$$

Here $1 - P[A(x)]$ is the rejection probability for a proposed transition from x. For exact and adjusted numerical HMC, the probability measure μ defined by (2) is invariant for π under the assumptions made below, cf. for example, [6, 35]. For unadjusted numerical HMC, the invariant probability measure for π in general does not agree with μ, but when it exists, typically approaches μ as $h \downarrow 0$.

2.2. *Assumptions.*

For our first main result we impose the following regularity condition on U:

ASSUMPTION 2.1. U is a function in $C^4(\mathbb{R}^d)$ satisfying the following conditions:

(A1) U has a local minimum at 0, and $U(0) = 0$.
(A2) U has bounded second, third and fourth derivatives. We set

$$
(10) \qquad L = \sup\|\nabla^2 U\|, \qquad M = \sup\|\nabla^3 U\|, \qquad N = \sup\|\nabla^4 U\|.
$$

(A3) U is strongly convex outside a Euclidean ball, that is, there exist constants $\mathcal{R} \in [0, \infty)$ and $K \in (0, \infty)$ s.t. for all $x, y \in \mathbb{R}^d$ with $|x - y| \geq \mathcal{R}$,

$$
(11) \qquad (x - y) \cdot \big(\nabla U(x) - \nabla U(y)\big) \geq K|x - y|^2.
$$

Notice that (A3) implies that U has a local minimum. Hence if (A3) holds then (A1) can always be satisfied by centering the coordinate system appropriately and subtracting a constant from U. Conditions (A1) and (A2) imply

$$
(12) \qquad \big|\nabla U(x)\big| = \big|\nabla U(x) - \nabla U(0)\big| \leq L|x| \quad \text{for any } x \in \mathbb{R}^d.
$$

Alternatively, it is possible to replace (A3) by a Lyapunov type drift condition.

ASSUMPTION 2.2. U is a function in $C^4(\mathbb{R}^d)$ satisfying (A1) and (A2). Moreover, there exists a measurable function $\Psi : \mathbb{R}^d \to [0, \infty)$ and constants $\lambda, \alpha \in (0, \infty)$ and $R_2 \in (0, \infty]$ such that:

(A4) The level set $\{x \in \mathbb{R}^d : \Psi(x) \leq 4\alpha/\lambda\}$ is compact.
(A5) For all $x \in \mathbb{R}^d$ with $|x| < R_2$,

$$
(\pi \Psi)(x) \leq (1 - \lambda)\Psi(x) + \alpha.
$$

If Assumption 2.2 is satisfied then we set

$$
(13) \qquad \mathcal{R} := \sup\{|x| : x \in \mathbb{R}^d \text{ such that } \Psi(x) \leq 4\alpha/\lambda\}.
$$

In the results below, this constant will play a similar rôle as the constant $\mathcal{R}$ in (A3). For exact and unadjusted HMC, we usually choose $R_2 = \infty$ in Assumption 2.2. For adjusted

numerical HMC, Lyapunov functions satisfying (A5) globally may fail to exist since on larger balls smaller step sizes may be required to ensure stability. In this case, we can still prove a contraction result on a large ball of radius $R_2 < \infty$ if a corresponding Lyapunov condition is satisfied.

Although the drift condition (A5) is not stated as explicitly as condition (A3), it can be verified in an explicit way for several important classes of examples.

EXAMPLE 2.1 (Quadratic Lyapunov functions). Suppose that Assumptions (A1) and (A2) are satisfied, and there exist constants $\kappa, C \in (0, \infty)$ such that

$$(14) \qquad x \cdot \nabla U(x) \geq \kappa |x|^2 - C \quad \text{for all } x \in \mathbb{R}^d.$$

Then there exists $n_0 \in \mathbb{N}$ such that for exact HMC and for unadjusted numerical HMC with step size $h = T/n$, the Lyapunov condition

$$(15) \qquad (\pi \Psi)(x) \leq (1 - \kappa T^2/8)\Psi(x) + (C + 2d)T^2 \quad \text{with } \Psi(x) := |x|^2$$

is satisfied for all $x \in \mathbb{R}^d$ provided $n \geq n_0$ and $L(T^2 + hT) \leq \kappa/(10L)$. For adjusted numerical HMC, the same assertion holds for $|x| \leq R_2$ where R_2 is an arbitrary finite constant, provided $h^2 R_2^3$ is sufficiently small.

EXAMPLE 2.2 (Exponential Lyapunov functions). Suppose that Assumptions (A1) and (A2) are satisfied and there exist constants $\kappa, C, Q \in (0, \infty)$ such that

$$(16) \qquad x \cdot \nabla U(x) \geq \kappa |x| - C \quad \text{and} \quad |\nabla U(x)| \leq Q \quad \text{for all } x \in \mathbb{R}^d.$$

Then there exists $\delta > 0$ and a smooth function $\Psi : \mathbb{R}^d \to (0, \infty)$ with $\Psi(x) = \exp(\delta |x|)$ for $|x| \geq 1/\delta$ such that for exact or unadjusted HMC

$$(17) \qquad (\pi \Psi)(x) \leq \delta \kappa T^2 (5 - \Psi(x)/7)$$

is satisfied for all $x \in \mathbb{R}^d$ provided $L(T^2 + hT) \leq 1$. Explicitly, one can choose $\delta = \kappa/(4C + 8d + Q^2 T^2)$. Again, a corresponding assertion holds for adjusted HMC provided $|x| \leq R_2$ and h is chosen sufficiently small depending on R_2.

A sketch of the proofs of (15) and (17) is included in the Appendix.

2.3. *Coupling*. We now introduce a coupling for the transition steps of two copies of the HMC chain starting at different initial conditions x and y. The coupling is defined in a different way depending on whether x and y are far apart or sufficiently close.

2.3.1. *Synchronous coupling for $|x - y| \geq 2\mathcal{R}$.* The easiest way to couple the transition probabilities $\pi(x, \cdot)$ and $\pi(y, \cdot)$ for two states $x, y \in \mathbb{R}^d$ is to use the same random variables ξ and $\mathcal{U}$ in both cases for the momentum refreshment and to decide whether a proposed move is accepted. The corresponding coupling transition is given by $(x, y) \mapsto (X'(x, y), Y'(x, y))$ where

$$(18) \qquad \begin{aligned} X'(x, y) &= q_T(x, \xi) I_{A(x)} + x I_{A(x)^c}, \\ Y'(x, y) &= q_T(y, \xi) I_{A(y)} + y I_{A(y)^c}, \end{aligned}$$

with $A(x), A(y)$ defined as in (4) above with the same ξ and $\mathcal{U}$ in both cases. We will apply synchronous coupling for $|x - y| \geq 2\mathcal{R}$. Here we can exploit the strong convexity condition (A3) to ensure contractivity for the coupling transition.
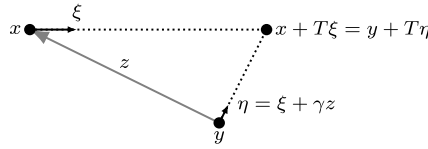
FIG. 1.   *A diagram showing the basic idea behind the coupling in the case $\gamma = T^{-1}$. The dotted lines connect the initial positions x and y with the final position $q_T(x, \xi) = q_T(y, \eta)$ for $U = 0$. When $U \neq 0$, $q_t(x, \xi) - q_t(y, \eta)$ is still contracting for small t.*

2.3.2. *A contractive coupling for $|x - y| < 2\mathcal{R}$.*   For $|x - y| < 2\mathcal{R}$ we use a different coupling that enables us to derive a weak form of contractivity even in the absence of convexity. Let $\gamma > 0$ be a positive constant. The precise value of the parameter γ will be chosen in an appropriate way below. The coupling transition step is now given by

$$
(19) \qquad
\begin{aligned}
X'(x, y) &= q_T(x, \xi) I_{A(x)} + x I_{A(x)^C}, \\
Y'(x, y) &= q_T(y, \eta) I_{\hat{A}(y)} + y I_{\hat{A}(y)^C},
\end{aligned}
$$

with the event $A(x)$ defined as in (4) above, and

$$
(20) \qquad \hat{A}(y) = \{\mathcal{U} \leq \exp(H(y, \eta) - H(q_T(y, \eta), p_T(y, \eta)))\} \quad \text{or} \quad \hat{A}(y) = \Omega
$$

in the adjusted and unadjusted case, respectively. Here the *same* random variable $\mathcal{U}$ as in (4) is used to decide whether the proposed move to $q_T(y, \eta)$ is accepted. Moreover, we set

$$
(21) \qquad \eta :=
\begin{cases}
\xi + \gamma z & \text{if } \widetilde{\mathcal{U}} \leq \dfrac{\varphi_{0,1}(e \cdot \xi + \gamma |z|)}{\varphi_{0,1}(e \cdot \xi)}, \\
\xi - 2(e \cdot \xi)e & \text{otherwise,}
\end{cases}
$$

where $z = x - y$, $e = z/|z|$, $\varphi_{0,1}$ denotes the density of the standard normal distribution, and $\widetilde{\mathcal{U}} \sim \text{Unif}(0, 1)$ is independent of T, ξ and $\mathcal{U}$.

This coupling is partially motivated by a coupling for second order Langevin diffusions introduced in [20]. It is defined in such a way that $\xi - \eta = -\gamma z$ holds with the maximal possible probability, and a reflection coupling is applied otherwise. As illustrated in Figure 1, the reason for this choice is that the difference process $q_t(x, \xi) - q_t(y, \eta)$ is contracting in a time interval $[0, t_0]$ if the difference $\xi - \eta$ of the initial velocities is negatively proportional to the difference of the initial positions.

In order to verify that $(X'(x, y), Y'(x, y))$ is indeed a coupling of the transition probabilities $\pi(x, \cdot)$ and $\pi(y, \cdot)$, we remark that the distribution of η is $N(0, I_d)$ since, by definition of η in (21) and a change of variables,

$$
\begin{aligned}
P[\eta \in B] &= E\left[I_B(\xi + \gamma z)\frac{\varphi_{0,1}(e \cdot \xi + \gamma |z|)}{\varphi_{0,1}(e \cdot \xi)} \wedge 1\right] \\
&\quad + E\left[I_B(\xi - 2(e \cdot \xi)e)\left(1 - \frac{\varphi_{0,1}(e \cdot \xi + \gamma |z|)}{\varphi_{0,1}(e \cdot \xi)}\right)^+\right] \\
&= \int I_B(x + \gamma z)\varphi_{0,I_d}(x + \gamma z) \wedge \varphi_{0,I_d}(x)\, dx \\
&\quad + \int I_B(x - 2(e \cdot x)e)(\varphi_{0,I_d}(x) - \varphi_{0,I_d}(x + \gamma z))^+\, dx \\
&= \int I_B(x)\varphi_{0,I_d}(x)\, dx = P[\xi \in B]
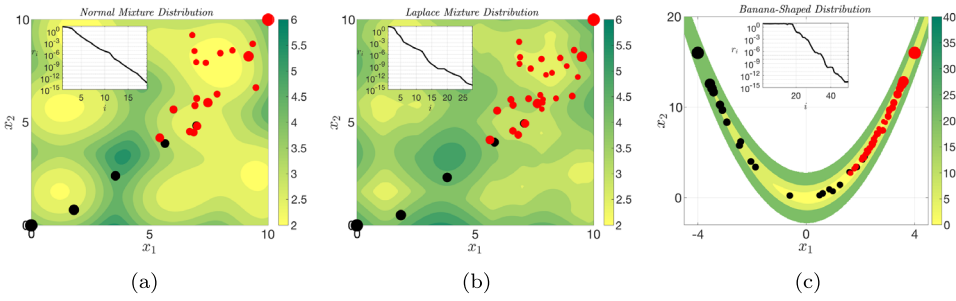\end{aligned}
$$

FIG. 2. *This figure illustrates realizations of the coupling with $T = 1$ and $\gamma = 1$. The different components of the coupling are shown as different color dots. The size of the dots is related to the number of steps: points along the trajectory corresponding to a larger number of steps have smaller markers. A contour plot of the underlying potential energy function is shown in the background. The inset plots the distance r_i between the components of the coupling as a function of the step index i. The simulation is terminated when this distance first reaches 10^{-14}. In (a), (b) and (c), this occurs in 20, 29, and 51 steps, respectively.*

for any measurable set B. Here $a \wedge b$ denotes the minimum of real numbers a and b, and we have used that $\varphi_{0, I_d}(y - 2(e \cdot y)e) = \varphi_{0, I_d}(y) = \varphi_{0, I_d}(-y)$. As a byproduct of this calculation, note also that

$$P[\eta \neq \xi + \gamma z] = \int (\varphi_{0, I_d}(x) - \varphi_{0, I_d}(x + \gamma z))^+ \, dx = d_{\mathrm{TV}}(\mathcal{N}(0, I_d), \mathcal{N}(\gamma z, I_d)),$$

where d_{TV} is the total variation distance. Hence, by the coupling characterization of the total variation distance, $\xi - \eta = -\gamma z$ does indeed hold with maximal possible probability.

2.4. *Numerical illustration of couplings.* Before stating our theoretical results, we test the coupling defined by (19) numerically on the following distributions:

- A normal mixture distribution where the mixture components are twenty two-dimensional Gaussian distributions with covariance matrix given by the 2×2 identity matrix and with mean vectors given by 20 independent samples from the uniform distribution over the rectangle $[0, 10] \times [0, 10]$. The energy barriers are not large. This example is adapted from [28, 31].
- A Laplace mixture distribution where the mixture components are twenty two-dimensional (regularized) Laplace distributions using the same covariance matrix and mean vectors as in the preceding example. However, unlike the preceding example, in this example the underlying potential is not strongly convex outside a ball, but instead satisfies Assumption 2.2 with respect to an exponential Foster–Lyapunov function as in Example 2.2.
- A banana-shaped distribution whose associated potential energy $U : \mathbb{R}^2 \to \mathbb{R}$ is given by the Rosenbrock function $U(x, y) = (1 - x)^2 + 10(y - x^2)^2$. This function is highly nonconvex and unimodal with a global minimum at the point $(1, 1)$ where $U(1, 1) = 0$. This minimum lies in a long, narrow, banana shaped valley.

For simplicity, we apply the coupling globally and choose the step size h to integrate the Hamiltonian dynamics small enough to ensure that essentially all proposed moves are accepted. Realizations of the coupling process with $T = 1$ and $\gamma = 1$ are shown in Figure 2. We chose these parameters only for visualization purposes. The different components of the coupling are shown as different color dots. The insets of the figures show the distance between the components of the coupling as a function of the number of steps.

Figure 3 shows the average time after which the distance between the components of the coupling is for the first time within 10^{-9}. To produce this figure, we generated 10^5 samples of the coupled process for one hundred different values of the duration parameter T. We
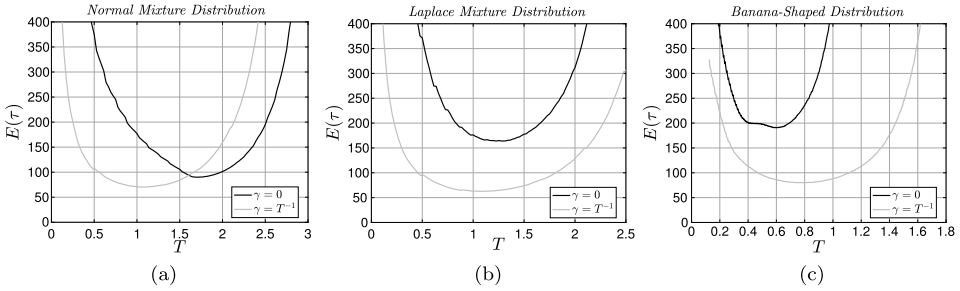
FIG. 3. *This figure illustrates the average of the random time τ after which the distance between the components of the coupling is for the first time within 10^{-9}. The estimated average is plotted as a function of the duration T of the Hamiltonian dynamics for $\gamma = 0$ (black) and $\gamma = T^{-1}$ (gray). The latter choice is motivated by Figure 1. From (a), note that the minimum of the function is smaller and occurs at a smaller value of T when $\gamma = T^{-1}$. This difference is more pronounced in (b) and (c), because the underlying potential is not strongly convex outside a ball in (b) and is highly nonconvex in (c). The kinks in the graphs are due to an artificial periodicity phenomena with respect to the time step size, and can be reduced by reducing the time step size.*

chose the coupling parameter γ equal to either T^{-1}, or equal to zero which corresponds to a synchronous coupling. The former choice is motivated by Figure 1.

2.5. *Contractivity.* We now state our main contraction bounds for the coupling introduced above. For given $x, y \in \mathbb{R}^d$ let

$$r(x, y) = |x - y|, \qquad R'(x, y) = |X'(x, y) - Y'(x, y)|,$$

denote the coupling distance before and after the transition step. For exact HMC we set $h := 0$, whereas for numerical HMC, $h > 0$ is the discretization step size.

2.5.1. *Contractivitiy by strong convexity.* The assumed strong convexity of U outside of a euclidean ball directly implies contractivity of a transition step for exact HMC for initial values x and y that are sufficiently far apart.

THEOREM 2.1 (Contractivity for exact HMC, strongly convex case). *Suppose that Assumption 2.1 is satisfied, and let $h = 0$. Then for any $x, y \in \mathbb{R}^d$ and $T \in \mathbb{R}_+$ such that $|x - y| \geq 2\mathcal{R}$ and $LT^2 \leq K/L$,*

$$(22) \qquad R'(x, y) \leq \left(1 - \frac{1}{2}KT^2\right)r(x, y).$$

The proof is a direct consequence of Lemma 3.4 below; see Section 5. A similar result is proven in [34].

Notice that contractivity is only guaranteed for LT^2 smaller than the conditioning number K/L. Sometimes, contraction bounds for longer durations can be obtained. However, as discussed in the Introduction, due to possible periodicity of the Hamiltonian flow, in general these do hot hold for arbitrary T.

EXAMPLE 2.3 (Bivariate normal target). Consider $U(x) = \frac{1}{2}x^T \Sigma^{-1} x$, where $\Sigma = \begin{bmatrix} \sigma_{\max}^2 & 0 \\ 0 & \sigma_{\min}^2 \end{bmatrix}$ with $\sigma_{\max} \geq \sigma_{\min} > 0$. In this case, Assumption 2.1 is satisfied with $\mathcal{R} = 0$, $L = \sigma_{\min}^{-2}$ and $K = \sigma_{\max}^{-2}$. Theorem 2.1 gives a global contraction for synchronous coupling with rate $KT^2/2$ provided that $T^2 \leq \sigma_{\min}^4/\sigma_{\max}^2$. In particular, a necessary condition is that T is no greater than $\sigma_{\min}$, which avoids periodicities in the Hamiltonian dynamics [33].

Next we consider both adjusted and unadjusted numerical HMC. We fix an upper bound $h_1 > 0$ for the discretization step size h. We assume that

$$(23) \qquad\qquad LT(T + h_1) \le K/L.$$

Under similar conditions as in Theorem 2.1 we obtain contractivity on average for coupled HMC transition steps:

THEOREM 2.2 (Contractivity for numerical HMC, strongly convex case). *Suppose that Assumption* 2.1 *is satisfied, and fix* $T, R_2, h_1 \in (0, \infty)$ *such that* (23) *holds. Then there exists* $h_0 > 0$ *depending only on* K, L, M, N, T, R_2 *and* d *such that for any* $h \in (0, \min(h_0, h_1)]$ *with* $T/h \in \mathbb{Z}$ *and for any* $x, y \in \mathbb{R}^d$ *with* $|x - y| \ge 2\mathcal{R}$ *and* $\max(|x|, |y|) \le R_2$,

$$(24) \qquad\qquad E[R'(x, y)] \le \left(1 - \frac{1}{4} KT^2\right) r(x, y).$$

Moreover, for fixed K, L, M *and* N, h_0 *can be chosen such that* h_0^{-1} *is of order* $O(R_2)$ *for unadjusted numerical HMC and of order* $O((1 + T^{-1/2})(R_2^2 + d))$ *for adjusted numerical HMC.*

The proof is given in Section 5. Key ingredients are the bound for contractivity of the proposal in Lemma 3.4 and a bound for the probability that the proposal move gets accepted for one of the components of the coupling and rejected for the other component, cf. Theorem 3.8. For unadjusted HMC, the proof simplifies considerably, and h_0 can be chosen independently of the dimension.

2.5.2. *Contractivity without global convexity.* Even if we do not assume convexity, we can still obtain contractivity for x, y at a bounded distance if we replace $r(x, y) = |x - y|$ by a modified metric. To this end we consider a distance function of the form

$$(25) \qquad\qquad \rho(x, y) = f(r(x, y)), \quad x, y \in \mathbb{R}^d,$$

where $f : [0, \infty) \to [0, \infty)$ is a concave function given by

$$(26) \qquad\qquad f(r) = \int_0^r \exp(-a \min(s, R_1)) \, ds$$

with parameters $a, R_1 \in (0, \infty)$ to be specified below.

We again fix an upper bound $h_1 > 0$ for the discretization step size h. We now replace (23) by the more stringent assumption

$$(27) \qquad\qquad L(T + h_1)^2 \le \min\left(\frac{K}{L}, \frac{1}{4}, \frac{1}{16\Lambda}\right),$$

where $\Lambda := 16L\mathcal{R}^2$. Under this condition we obtain contractivity on average w.r.t. the metric ρ if the parameters γ, a and R_1 defining the coupling and the metric are adjusted appropriately. Explicitly, we set

$$(28) \qquad\qquad \gamma := \min(T^{-1}, \mathcal{R}^{-1}/4),$$

$$(29) \qquad\qquad a := T^{-1},$$

$$(30) \qquad\qquad R_1 := \frac{5}{2}(\mathcal{R} + T).$$

For exact HMC, we can prove a global contraction under Assumption 2.1.

THEOREM 2.3 (Contractivity for exact HMC, general case). *Suppose that Assumption* 2.1 *is satisfied, and fix* $T \in (0, \infty)$ *such that* $LT^2 \leq \min(\frac{K}{L}, \frac{1}{4}, \frac{1}{16\Lambda})$. *Let* γ, a *and* R_1 *be given by* (28), (29) *and* (30), *respectively. Then for exact HMC, for all* $x, y \in \mathbb{R}^d$,

$$E[f(R'(x, y))] \leq (1 - c)f(r(x, y)), \tag{31}$$

$$\text{where } c = \frac{1}{10}\min\left(1, \frac{1}{2}KT^2\left(1 + \frac{\mathcal{R}}{T}\right)e^{-\mathcal{R}/(2T)}\right)e^{-2\mathcal{R}/T}. \tag{32}$$

For numerical HMC, a similar contraction property holds on every finite ball in $\mathbb{R}^d$. The contraction rate does not depend on the radius R_2 of the ball, but on a larger ball, a smaller step size is required to ensure contractivity.

THEOREM 2.4 (Contractivity for numerical HMC, general case). *Suppose that Assumption* 2.1 *is satisfied, and fix* $T, R_2, h_1 \in (0, \infty)$ *such that* (27) *holds. Let* γ, a *and* R_1 *be given by* (28), (29) *and* (30), *respectively. Then there exists* $h_\star > 0$ *depending only on* K, L, M, N, $\mathcal{R}, T, R_2$ *and* d *s.t. for any* $h \in (0, \min(h_\star, h_1)]$ *with* $T/h \in \mathbb{Z}$ *and for any* $x, y \in \mathbb{R}^d$ *with* $\max(|x|, |y|) \leq R_2$, (31) *holds with the contraction rate* c *given by* (32). *Moreover, for fixed* K, L, M *and* N, $h_\star$ *can be chosen such that* $h_\star^{-1}$ *is of order* $O(R_2)$ *for unadjusted, and of order* $O((1 + T^{-1/2} + \mathcal{R}^{1/2})(R_2^2 + d))$ *for adjusted numerical HMC.*

The proofs of Theorems 2.3 and 2.4 are given in Section 5.

The following simple example demonstrates that our results provide applicable bounds in multimodal situations if T is adjusted appropriately:

EXAMPLE 2.4 (Gaussian mixture). Consider a mixture of two Gaussians with means $\pm 2\sigma$ and variances σ^2 where $\sigma > 0$. The corresponding potential is

$$U(x) = -\log\left(\exp\left(-\frac{(x - 2\sigma)^2}{2\sigma^2}\right) + \exp\left(-\frac{(x + 2\sigma)^2}{2\sigma^2}\right)\right).$$

In this case, $\sigma^2 U''(x) = 1 - 4\operatorname{sech}(2x/\sigma)^2$, $L = \sup|U''| = 3/\sigma^2$, and $U''(x) \geq 2/(3\sigma^2)$ for all $|x| > \sigma$, which allows us to choose $K = 2/(3\sigma^2)$ and $\mathcal{R} = 2\sigma$. Hence, the condition on LT^2 in Theorem 2.3 reduces to $LT^2 \leq 1/3072$, and the rate in (32) reduces to

$$c = \frac{1}{10}\min\left(1, \frac{1}{3}\frac{T^2}{\sigma^2}\left(1 + 4\frac{\sigma}{T}\right)e^{-2\sigma/T}\right)e^{-8\sigma/T}.$$

If we choose T proportional to σ, then this rate is constant.

REMARK 2.5 (Scope of coupling approach). There are various possible extensions of the basic coupling approach presented here. These include componentwise coupling on product spaces, two-scale couplings on high dimensional spaces, and geometry-adapted couplings. See [18, 38, 40] for corresponding approaches for reflection coupling of diffusion processes—we expect that similar extensions exist for our couplings of HMC. For example, it might be sufficient to apply asynchronous coupling to a few components and couple the other components synchronously [40], and for models with a nontrivial geometry such as the banana shaped distribution in Figure 2, the coupling should be designed with straight lines replaced by geodesics in an appropriate geometry [9].

In Assumption 2.1, at least the global Lipschitz condition on $\nabla U(x)$ is crucial for our proofs. Nevertheless, convergence may still hold under a relaxed condition if the step size is adjusted appropriately, possibly depending on x. The asymptotic strong convexity assumption can be replaced by a Lyapunov condition; see below. Finally, Example 2.3 shows that

condition (27) on T cannot be avoided in general. For appropriate classes of models, it could possibly be improved by considering an adequate non-Euclidean geometry. More generally, we point out that even if some of the assumptions are not satisfied, the coupling is still well defined, and one can check empirically if two copies with distant starting points approach each other. Of course, this does not guarantee convergence but it may serve as a diagnostic tool for HMC.

2.5.3. *Contractivity under Foster–Lyapunov condition.* It is possible to replace the asymptotic strong convexity condition in Assumption 2.1 by the Lyapunov drift condition in Assumption 2.2. In this case we consider a semimetric of the form

$$(33) \qquad \rho_\epsilon(x, y) = \sqrt{f\left(\min(|x - y|, 2\mathcal{R})\right) \cdot \left(1 + \epsilon\Psi(x) + \epsilon\Psi(y)\right)},$$

where $\epsilon > 0$ is a positive constant, $\mathcal{R}$ is defined by (13), and f is again given by (26). The idea of using corresponding semimetrics for deriving contraction properties under Lyapunov conditions goes back to [23] and [7]; see also [19].

Similarly as above, we fix an upper bound $h_1 > 0$ for the discretization step size h such that

$$(34) \qquad L(T + h_1)^2 \leq \min\left(\frac{1}{4}, \frac{1}{16\Lambda}\right),$$

where $\Lambda := 16L\mathcal{R}^2$. We choose the parameters γ, a and R_1 as in (28), (29) and (30) respectively, and we set

$$(35) \qquad \epsilon := \frac{1}{40A}e^{-2\mathcal{R}/T}.$$

Then under a global Lyapunov condition, we can prove global contractivity on average w.r.t. the semimetric ρ_ϵ for exact and for unadjusted numerical HMC.

THEOREM 2.6 (Contractivity for exact and unadjusted HMC under Lyapunov condition). *Consider exact HMC or unadjusted numerical HMC with step size $h \in (0, h_1]$ such that $T/h \in \mathbb{Z}$. Suppose that Assumption 2.2 is satisfied with $R_2 = \infty$, and (34) holds. Let γ, a, R_1 and ϵ be given by (28), (29), (30) and (35), respectively. Then for all $x, y \in \mathbb{R}^d$,*

$$(36) \qquad E\left[\rho_\epsilon\left(X'(x, y), Y'(x, y)\right)\right] \leq (1 - c)\rho_\epsilon(x, y),$$

$$(37) \qquad \text{where } c = \frac{1}{80}\min(10\lambda, e^{-2\mathcal{R}/T}).$$

For adjusted numerical HMC, we again obtain contractivity on a given ball for sufficiently small step sizes.

THEOREM 2.7 (Contractivity for numerical HMC under Lyapunov condition). *Consider adjusted numerical HMC with step size $h \in (0, h_1]$ such that $T/h \in \mathbb{Z}$. Suppose that Assumption 2.2 is satisfied for some $R_2 \in (0, \infty)$, and (34) holds. Let γ, a, R_1 and ϵ be given by (28), (29), (30) and (35), respectively. Then there exists $h_\star > 0$ depending only on L, M, N, $\mathcal{R}$, R_2 and d s.t. for any $h \in (0, \min(h_\star, h_1)]$ with $T/h \in \mathbb{Z}$ and for any $x, y \in \mathbb{R}^d$ with $\max(|x|, |y|) \leq R_2$, (31) holds with the contraction rate c given by (32). Moreover, for fixed L, M and N, $h_\star$ can be chosen such that $h_\star^{-1}$ is of order $O((1 + \mathcal{R}^{1/2})(R_2^2 + d))$ for adjusted numerical HMC.*

The proofs of Theorems 2.6 and 2.7 are given in Section 5.

2.5.4. *Dimension dependence of contraction rates.*   Dependence on the dimension of the bounds for the contraction rates derived above arises by two different effects:

(a) *Dimension dependence due to model properties.* The rate c in Theorems 2.3 and 2.4 does not depend explicitly on the dimension d. However, a dimension dependence arises if the underlying parameters in Assumptions 2.1 and 2.2 are dimension dependent. For example, the parameter $\mathcal{R}$ in (A3) may depend on d which would then cause an implicit (and possibly exponential) dimension dependence of the bound c for the contraction rate in Theorems 2.3 and 2.4. Similarly, the constant α in (A5) will usually depend on d and cause an exponential degeneration of the rates in Theorems 2.6 and 2.7. This possible degeneration in high dimensions cannot be avoided in the general setup considered here, and it occurs in a very similar form for other Markov processes with invariant measure μ; see, for example, [18–20]. Indeed, note that our assumptions do not exclude models with a phase transition in the limit $d \to \infty$.

For more specific model classes, modifications of our coupling approach can avoid a possible dimension dependence. We mention two such situations that will be analysed in detail in forthcoming work; see also the related results for other classes of Markov processes in [17, 18, 40].

- For models with weak interactions (e.g., perturbations of product measures), a componentwise coupling similar to the one suggested for Langevin diffusions in [18] can be applied to obtain dimension free contraction rates for exact or unadjusted numerical Hamiltonian Monte Carlo.
- For probability measures μ that have a sufficiently nice density w.r.t. a Gaussian measure on $\mathbb{R}^d$ or on an infinite dimensional Hilbert space, a two scale coupling approach has been applied in [40] to derive dimension free contraction rates in the case of overdamped Langevin diffusions with invariant measure μ. Such measures arise for example in transition path sampling and Bayesian inverse problems [10, 24]. In forthcoming work, we show that in combination with the results in this paper, the approach in [40] can be carried over to HMC.

(b) *Dimension dependence due to numerical discretization.* For numerical HMC, the dimension also affects the upper bound $h_\star$ for the discretization step size. This is relevant for resulting upper bounds of the computational complexity which are typically of order $1/(h_\star c)$.

In the results above, this dimension dependence of the maximal step size does not occur for unadjusted HMC. Although in this case, the contraction properties of the Markov chain are completely dimension free, a dimension dependence arises because the step size has to be chosen sufficiently small to ensure that the invariant measure of the unadjusted Markov chain is close to μ. The analysis of this dimension dependence requires different techniques that are not the content of this work. Under rather strong assumptions (including strong convexity), quite sharp bounds for the resulting total dimension dependence for several integrators have been derived by Mangoubi and Smith [34]. In a more general context, sharp bounds for the distance between the invariant measures are derived in the forthcoming work [14].

On the other hand, for adjusted HMC, our bounds require h^{-1} of order d, thus causing a dimension dependence of at least the same order for the computational complexity. It is an open question if these bounds can be improved. Scaling limits and the results in [34] suggest that at least under more restrictive assumptions, a better dimension dependence (possibly up to $d^{1/4}$) might hold for the computational complexity. On the other hand, we are not aware of a result proving a contraction under a bound on h^{-1} of order d^α with $\alpha < 1$.

2.6. *Quantitative bounds for distance to the invariant measure.* For exact HMC, Theorem 2.3 establishes global contractivity of the transition kernel $\pi(x, dy)$ w.r.t. the Kantorovich (L^1 Wasserstein) distance

$$\mathcal{W}_\rho(\nu, \eta) = \inf_{\gamma \in C(\nu, \eta)} \int \rho(x, y)\gamma(dx\, dy)$$

on probability measures ν, η on $\mathbb{R}^d$. Here the infimum is over all couplings γ of ν and η. Since the metric ρ is comparable to the Euclidean Distance on $\mathbb{R}^d$, contractivity w.r.t. $\mathcal{W}_\rho$ immediately implies a quantitative bound on the standard L^1-Wasserstein distance

$$\mathcal{W}^1(\nu\pi^n, \mu) = \inf_{\gamma \in C(\nu\pi^n, \mu)} \int |x - y|\gamma(dx\, dy)$$

between the law of the HMC chain after n steps and the invariant probability measure μ.

COROLLARY 2.8. *Suppose that Assumption* 2.1 *is satisfied, and let* $T \in (0, \infty)$ *such that*

$$(38) \qquad LT^2 \leq \min\left(\frac{K}{L}, \frac{1}{4}, \frac{1}{256LR^2}\right).$$

Then for any $n \in \mathbb{N}$ *and for any probability measures* ν, η *on* $\mathbb{R}^d$,

$$(39) \qquad \mathcal{W}_\rho(\nu\pi^n, \eta\pi^n) \leq e^{-cn}\mathcal{W}_\rho(\nu, \eta) \quad and$$

$$(40) \qquad \mathcal{W}^1(\nu\pi^n, \eta\pi^n) \leq Me^{-cn}\mathcal{W}^1(\nu, \eta),$$

where c *is given by* (32), *and* $M = \exp(\frac{5}{2}(1 + \mathcal{R}/T))$. *In particular, for a given constant* $\epsilon \in (0, \infty)$, *the standard* L^1 *Wasserstein distance* $\Delta(n) = \mathcal{W}^1(\nu\pi^n, \mu)$ *w.r.t.* μ *after* n *steps of the chain with initial distribution* ν *satisfies* $\Delta(n) \leq \epsilon$ *provided*

$$(41) \qquad n \geq \frac{1}{c}\left(\frac{5}{2} + \frac{5\mathcal{R}}{2T} + \log\frac{\Delta(0)}{\epsilon}\right).$$

The corollary is a rather direct consequence of Theorem 2.3. The proof is given in Section 6.

REMARK 2.9 (Kinetic bounds). One remarkable feature of the result in Corollary 2.8 is that for a given initial error $\Delta(0)$, the number of steps required to stay below a certain error bound ϵ can be chosen universally provided T is chosen proportional to $\mathcal{R}$. Notice, however, that by Condition (38), it is only possible to choose T proportional to $\mathcal{R}$ with a fixed proportionality constant if $L\mathcal{R}^2$ is bounded by a fixed constant!

REMARK 2.10 (Quantitative bounds for ergodic averages). MCMC methods are often applied to approximate expectation values w.r.t. the target distribution by ergodic averages of the Markov chain. Our results (e.g., (39)) directly imply completely explicit bounds for biasses and variances, as well as explicit concentration inequalities for these ergodic averages in the case of HMC. Indeed, the general results by Joulin and Ollivier [27] show that such bounds follow directly from an L^1 Wasserstein contraction w.r.t. an arbitrary metric ρ, which is precisely the statement shown above.

We now return to numerical HMC. Here, our main results in Theorems 2.4 and 2.7 only establish contractivity w.r.t. $\mathcal{W}_\rho$ on a ball of given radius R_2. In order to derive bounds for the distance to the invariant measure of the law after n steps, we additionally have to control exit probabilities from the ball. This is achieved by another Lyapunov bound that we first state in a general form. Suppose that $\pi(x, dy)$ is the transition kernel of a Markov chain on a complete separable metric space (S, ρ), and let $\mathcal{W}_\rho$ denote the corresponding Kantorovich distance on probability measures on S.

ASSUMPTION 2.3. The following conditions are satisfied for a constant $C \in (0, \infty)$ and measurable functions $\psi, \varphi : S \to (0, \infty)$:

(C1) *Main Lyapunov condition*: There is a constant $\lambda \in [1, \infty)$ such that

$$(\pi \psi)(x) \leq \lambda \psi(x) \quad \text{for any } x \in S \text{ s.t. } \psi(x) \leq C.$$

(C2) *Additional global Lyapunov condition*: There is a constant $\beta \in [1, \infty)$ s.t.

$$(\pi \varphi)(x) \leq \beta \varphi(x) \quad \text{and} \quad \rho(x, y) \leq \varphi(x) + \varphi(y) \quad \text{for any } x, y \in S.$$

(C3) *Local contractivity*: There are a measurable map $(X', Y') : S \times S \times \Omega \to S \times S$ defined on a probability space $(\Omega, \mathcal{A}, P)$ and a constant $c \in (0, \infty)$ such that for any $x, y \in S$, $(X'(x, y, \cdot), Y'(x, y, \cdot))$ is a realization of a coupling of $\pi(x, \cdot)$ and $\pi(y, \cdot)$ satisfying

$$E\big[\rho\big(X'(x, y, \cdot), Y'(x, y, \cdot)\big)\big] \leq e^{-c} \rho(x, y) \quad \text{if } \psi(x) \leq C \text{ and } \psi(y) \leq C.$$

The proof of the following theorem is given in Section 6.

THEOREM 2.11. *Suppose that Assumption 2.3 is satisfied. Then for any $n \in \mathbb{N}$ and for any probability measures v, η on $(S, \mathcal{B}(S))$,*

$$(42) \qquad \mathcal{W}_\rho(v\pi^n, \eta\pi^n) \leq e^{-cn} \mathcal{W}_\rho(v, \eta) + \beta^n \lambda^{n-1} \left(\int \psi \, dv + \int \psi \, d\eta \right) \delta(C),$$

where

$$(43) \qquad \delta(C) := \sup\left\{ \frac{\varphi(x) + \varphi(y)}{\psi(x) + \psi(y)} : x, y \in S \text{ s.t. } \psi(x) > C \text{ or } \psi(y) > C \right\}.$$

Theorem 2.11 can be applied to bound the distance to the invariant measure μ after n steps of adjusted numerical HMC. Suppose that π is the corresponding transition kernel for a given discretization step size $h > 0$, and let ρ denote the metric on $S = \mathbb{R}^d$ defined by (25), (26), (29) and (30). We will then apply Theorem 2.11 with Lyapunov functions of the form $\varphi(x) = 2Td^{1/2} + |x|$ and $\psi(x) = \exp(U(x)^{2/3})$. The right side of (42) can then be bounded by a given constant $\epsilon > 0$ by first choosing n such that the first term is bounded by $\epsilon/2$, and then choosing C such that the second term is bounded by $\epsilon/2$ as well. As a consequence of Theorem 2.4 and Theorem 2.11, we can prove that a similar number of steps as for exact HMC is also sufficient for an approximation of the invariant measure by numerical HMC, provided h is chosen sufficiently small.

THEOREM 2.12. *Suppose that Assumption 2.1 is satisfied. Let $T, h_1 \in (0, \infty)$ such that (27) holds, let v be a probability measure on $\mathbb{R}^d$, and let $\Delta(n) = \mathcal{W}^1(v\pi^n, \mu)$ denote the standard L^1 Wasserstein distance to the invariant probability measure after n steps of adjusted numerical HMC with initial distribution v. Let $\epsilon \in (0, \infty)$ and $n \in \mathbb{N}$ such that*

$$(44) \qquad n \geq \frac{1}{c}\left(\frac{5}{2} + \frac{5\mathcal{R}}{2T} + \log^+\left(\frac{2\Delta(0)}{\epsilon} \right) \right),$$

where c is given by (32). Then there exists $h_{\star\star} > 0$ depending only on $K, L, M, N, \mathcal{R}, T, d, v$ and n, such that for any $h \in (0, \min(h_{\star\star}, h_1))$ with $T/h \in \mathbb{Z}$,

$$(45) \qquad \Delta(n) \leq \epsilon.$$

Furthermore, $h_{\star\star}$ can be chosen such that for fixed values of K, L, M, N, $h_{\star\star}^{-1}$ is of order

$$O\big((1 + T^{-\frac{1}{2}} + \mathcal{R}^{\frac{1}{2}})(d^{\frac{3}{2}} n^{\frac{3}{2}} + (1 + \mathcal{R}/T)^{\frac{3}{2}}(d + A(v) + \log \epsilon^{-1})^{\frac{3}{2}} + \mathcal{R}^2)\big),$$

where $A(v) = \log \int \exp(U^{2/3}) \, dv$.

We finally remark that for unadjusted HMC, the above result does not apply in the same form. In this case, there is an additional error due to the fact that the invariant measure does not coincide with μ. The careful control of this error requires different techniques that are out of scope of this work; see [14, 34].

3. A priori estimates. In this section we state several bounds for the Hamiltonian flow, for the coupling, and for acceptance-rejection probabilities that will be crucial in the proof of our main result. The proofs of all the results stated in this section are included in Section 4.

3.1. *Bounds for the Hamiltonian flow and for velocity Verlet.* In the following, we consider $t \in [0, \infty)$ and $h \in [0, 1]$ such that $t/h \in \mathbb{Z}$ if $h > 0$. We assume throughout that Assumption 2.1 is satisfied, and

$$(46) \qquad\qquad L(t^2 + ht) \leq 1.$$

Recall that $\phi_t = (q_t, p_t)$ denotes the Hamiltonian flow for $h = 0$, and the flow of the velocity Verlet integrator for $h > 0$. The proofs of the following statements are provided in Section 4.

LEMMA 3.1. *For any $x, v \in \mathbb{R}^d$,*

$$(47) \qquad \max_{s \leq t} |q_s(x, v) - (x + sv)| \leq L(t^2 + th) \max(|x|, |x + tv|) \quad and$$

$$(48) \qquad \begin{aligned} \max_{s \leq t} |p_s(x, v) - v| &\leq Lt \max_{s \leq t} |q_s(x, v)| \\ &\leq Lt(1 + L(t^2 + th)) \max(|x|, |x + tv|). \end{aligned}$$

In particular,

$$(49) \qquad \max_{s \leq t} |q_s(x, v)| \leq 2 \max(|x|, |x + tv|) \quad and$$

$$(50) \qquad \max_{s \leq t} |p_s(x, v)| \leq |v| + 2Lt \max(|x|, |x + tv|).$$

LEMMA 3.2. *For any $x, y, u, v \in \mathbb{R}^d$,*

$$(51) \qquad \begin{aligned} &\max_{s \leq t} |q_s(x, u) - q_s(y, v) - (x - y) - s(u - v)| \\ &\leq L(t^2 + th) \max(|x - y|, |(x - y) + t(u - v)|) \quad and \end{aligned}$$

$$(52) \qquad \begin{aligned} &\max_{s \leq t} |q_s(x, u) - q_s(y, v) - (x - y) - s(u - v)| \\ &\leq Lt(1 + L(t^2 + th)) \max(|x - y|, |(x - y) + t(u - v)|). \end{aligned}$$

REMARK 3.3. The lemma shows that on sufficiently short time intervals, the first variation of velocity Verlet can be controlled by that of the corresponding motion with constant velocity. In particular, contractivity for small times holds if $u - v = -\gamma(x - y)$ for some $\gamma > 0$.

We will show next that in the region of strong convexity, the bounds in Lemma 3.2 can be improved if the initial velocities coincide. For such initial conditions, (51) and (53) imply

$$(53) \qquad |q_t(x, v) - q_t(y, v) - (x - y)| \leq L(t^2 + ht)|x - y|,$$

$$(54) \qquad |p_t(x, v) - p_t(y, v)| \leq Lt(1 + L(t^2 + ht))|x - y|.$$

For $|x - y| \geq 2\mathcal{R}$, the bound in (53) can be improved considerably:

LEMMA 3.4. *There exists a finite constant $C \in (0, \infty)$, depending only on L and M, such that the bound*

$$(55) \qquad \left| q_t(x, v) - q_t(y, v) \right|^2 \leq (1 - Kt^2/2)|x - y|^2$$

holds for any t, h as above such that

$$(56) \qquad L(t^2 + th) < K/L,$$

and for any $x, y, v \in \mathbb{R}^d$ such that

$$(57) \qquad |x - y| \geq 2\mathcal{R} \quad and \quad (1 + |x| + |v|)h \leq K/C.$$

REMARK 3.5. The lemma does not provide a bound if $|v|$ is very large. However, in this case we still have the upper bound

$$(58) \qquad \left| q_t(x, v) - q_t(y, v) \right| \leq (1 + L(t^2 + th))|x - y|$$

that follows from (53). Hence if $|v|$ is large with small probability, then we still get a contraction on average. For the exact Hamiltonian dynamics, there is no corresponding restriction on $|x|$ and $|v|$. Here, the lemma immediately yields a contraction result for synchronous coupling.

In the case of the exact Hamiltonian flow, that is, for $h = 0$, we have

$$(59) \qquad H\big(\phi_t(x, v)\big) = H(x, v) \quad \text{for any } t \in \mathbb{R} \text{ and } x, v \in \mathbb{R}^d.$$

We are now going to quantify the error in (59) in the case where the exact flow is replaced by the flow of the velocity Verlet integrator. This is crucial to quantify the acceptance-rejection probabilities.

LEMMA 3.6. *There exist finite constants $C_1, C_2 \in (0, \infty)$ that depend only on L, M and N such that the bounds*

$$(60) \qquad \left| (H \circ \phi_t - H)(x, v) \right| \leq C_1 th^2 \max(|x|, |v|)^3,$$

$$(61) \qquad \left| \partial_{(z,w)}(H \circ \phi_t - H)(x, v) \right| \leq C_2 th^2 \max(|x|, |v|)^3 \max(|z|, |w|),$$

hold for any $x, v, z, w \in \mathbb{R}^d$ and t, h as above satisfying (46).

3.2. *Bounds for acceptance-rejection probabilities.* We now provide some crucial bounds for probabilities and expectations that involve acceptance-rejection events and the coupling. Recall that the coupling that we consider for $|x - y| < 2\mathcal{R}$ ensures that $\xi - \eta = -\gamma z$ with the maximal possible probability, where $z = x - y$. The following lemma enables us to control probabilities and expectations when $\xi - \eta \neq -\gamma z$.

LEMMA 3.7. *For any $p \geq 1$ there exist finite constants C_p and $\widetilde{C}_p$ such that for any choice of γ,*

$$(62) \qquad P[\xi - \eta \neq -\gamma z] \leq |\gamma z|/\sqrt{2\pi},$$

$$(63) \qquad E\big[|e \cdot \xi|^p; \xi - \eta \neq -\gamma z\big] \leq C_p|\gamma z| \max(|\gamma z|, 1)^p,$$

$$(64) \qquad E\big[|\xi|^{2p}; \xi - \eta \neq -\gamma z\big] \leq \widetilde{C}_p|\gamma z|((d - 1)^p + \max(|\gamma z|, 1)^{2p}).$$

The a priori bounds (60) and (61) for velocity Verlet can be used to obtain a rather precise control for the rejection probabilities in HMC, and for the probability that in a coupling for HMC, the proposal is accepted for one component and rejected for the other. The resulting bounds are crucial to prove contractivity on average for the coupling.

THEOREM 3.8. *There exist finite constants* $C_1, C_2, C_3 \in (0, \infty)$ *that depend only on* L, M *and* N *such that the following bounds hold for any* $x, y \in \mathbb{R}^d$ *and* $T \in [0, \infty)$, $h \in [0, 1]$ *such that* $L(T^2 + hT) \leq 1$:

$$P[A(x)^C | \xi] \leq C_1 T(1 + T) \max(|x|, |\xi|)^3 h^2, \tag{65}$$

$$P[A(x)^C] \leq C_1 T(1 + T)(|x|^3 + 2d^{3/2})h^2, \tag{66}$$

$$P[\hat{A}(y)^C | \eta] \leq C_1 T(1 + T) \max(|x|, |\eta|)^3 h^2, \tag{67}$$

$$P[\hat{A}(y)^C] \leq C_1 T(1 + T)(|x|^3 + 2d^{3/2})h^2, \tag{68}$$

$$P[A(x)\Delta A(y) | \xi] \leq C_2 T(1 + T) \max(|x|, |y|, |\xi|)^3 |x - y| h^2, \tag{69}$$

$$P[A(x)\Delta A(y)] \leq C_2 T(1 + T)(\max(|x|, |y|)^3 + 2d^{3/2})|x - y| h^2, \tag{70}$$

$$P[A(x)\Delta \hat{A}(y) | \xi, \eta] \tag{71}$$
$$\leq C_2 T(1 + T) \max(|x - y|, |\xi - \eta|)h^2 \cdot \max(|x|, |y|, |\xi|, |\eta|)^3.$$

Furthermore, if $\gamma|x - y| \leq 1$, *then*

$$E[\max(|x|, |y|, |\xi|, |\eta|); A(x)\Delta \hat{A}(y)] \tag{72}$$
$$\leq C_3 \max(1, \gamma)T(1 + T)(\max(|x|, |y|)^4 + d^2)|x - y| h^2.$$

4. Proofs of a priori bounds. If $h > 0$ then we define $\lfloor t \rfloor = \lfloor t \rfloor_h$ and $\lceil t \rceil = \lceil t \rceil_h$ by (8). For $h = 0$ we set $\lfloor t \rfloor = \lceil t \rceil = t$. In both cases, (q_t, p_t) solves (7).

PROOF OF LEMMA 3.1. We fix $x, v \in \mathbb{R}^d$. Let $x_s = q_s(x, v)$ and $v_s = p_s(x, v)$. By (7), we have for any $s \in [0, t]$ that

$$x_s = x + \int_0^s v_{\lfloor r \rfloor} \, dr - \frac{h}{2} \int_0^s \nabla U(x_{\lfloor r \rfloor}) \, dr$$

$$= x + sv - \frac{1}{2} \int_0^s \int_0^{\lfloor r \rfloor} (\nabla U(x_{\lfloor u \rfloor}) + \nabla U(x_{\lceil u \rceil})) \, du \, dr - \frac{h}{2} \int_0^s \nabla U(x_{\lfloor r \rfloor}) \, dr.$$

By (12) and since $t \in h\mathbb{Z}$,

$$|x_s - x - sv| \leq \frac{L}{2} \int_0^s \int_0^r (|x_{\lfloor u \rfloor}| + |x_{\lceil u \rceil}|) \, du \, dr + \frac{hL}{2} \int_0^s |x_{\lfloor r \rfloor}| \, dr$$

$$\leq \frac{L}{2}(t^2 + th) \max_{s \leq t} |x_s| \quad \text{and thus}$$

$$\max_{s \leq t} |x_s - x - sv| \leq \frac{L}{2}(t^2 + th)\left(\max_{s \leq t} |x_s - x - sv| + \max(|x|, |x + vt|)\right).$$

By (46), we obtain:

$$\max_{s \leq t} |x_s - x - sv| \leq L(t^2 + th) \max(|x|, |x + vt|),$$

$$\max_{s \leq t} |x_s| \leq (1 + L(t^2 + th)) \max(|x|, |x + vt|) \tag{73}$$

$$\leq 2 \max(|x|, |x + vt|).$$

We now derive bounds for v_s. By (7) and (12),

$$v_s = v - \frac{1}{2} \int_0^s ((\nabla U)(x_{\lfloor r \rfloor}) + (\nabla U)(x_{\lceil r \rceil})) \, dr,$$

$$|v_s - v| \le \frac{L}{2} \int_0^s \left(|x_{\lfloor r \rfloor}| + |x_{\lceil r \rceil}|\right) dr.$$

Since $t \in h\mathbb{Z}$, we obtain by (73) and (46),

$$\max_{s \le t} |v_s - v| \le Lt \max_{s \le t} |x_s| \le Lt\left(1 + L(t^2 + th)\right) \max(|x|, |x + vt|),$$

$$\max_{s \le t} |v_s| \le |v| + 2Lt \max(|x|, |x + vt|). \qquad \square$$

PROOF OF LEMMA 3.2. The proof can be carried out in a similar way to the proof of Lemma 3.1, where instead of (12), we directly apply the Lipschitz bound $|\nabla U(x) - \nabla U(y)| \le L|x - y|$ for $x, y \in \mathbb{R}^d$. $\square$

PROOF OF LEMMA 3.4. Notice that we are in the case where the initial velocities coincide. We fix $x, y, v \in \mathbb{R}^d$ such that (57) holds true and set $x_s = q_s(x, v)$, $v_s = p_s(x, v)$, $y_s = q_s(y, v), z_s = q_s(x, v) - q_s(y, v)$ and $w_s = p_s(x, v) - p_s(y, v)$. In particular, $z_0 = x - y$ and $w_0 = 0$. Let

$$z_t^\star = \max_{s \le t} |z_s| \quad \text{and} \quad w_t^\star = \max_{s \le t} |w_s|.$$

By Lemma 3.2, we have

(74) $$|z_t - z_0| \le L(t^2 + ht)|z_0| \quad \text{and}$$

(75) $$w_t^\star \le Lt z_t^\star \le 2Lt \quad \text{for any } t \in h\mathbb{Z}_+ \text{ s.t. } L(t^2 + ht) \le 1.$$

The following computations are valid for $t \in \mathbb{R}_+$ such that $|z_s| \ge \mathcal{R}$ for $s \in [0, t]$. Recall that by (7),

(76) $$\dot{z}_t = w_{\lfloor t \rfloor} - \frac{h}{2}\left(\nabla U(x_{\lfloor t \rfloor}) - \nabla U(y_{\lfloor t \rfloor})\right),$$

(77) $$\dot{w}_t = -\frac{1}{2}\left(\nabla U(x_{\lfloor t \rfloor}) - \nabla U(y_{\lfloor t \rfloor}) + \nabla U(x_{\lceil t \rceil}) - \nabla U(y_{\lceil t \rceil})\right).$$

Let $a(t) := |z_t|^2$ and $b(t) := 2z_t \cdot w_t$. Our goal is to derive an upper bound for $a(t)$. To this end we note that a and b satisfy the following differential equations:

$$\dot{a}(t) = b(t) + \delta(t),$$

$$\dot{b}(t) = -z_t \cdot \left(\nabla U(x_{\lfloor t \rfloor}) - \nabla U(y_{\lfloor t \rfloor}) + \nabla U(x_{\lceil t \rceil}) - \nabla U(y_{\lceil t \rceil})\right)$$
$$\qquad + 2w_{\lfloor t \rfloor} \cdot w_t - hw_t \cdot \left(\nabla U(x_{\lfloor t \rfloor}) - \nabla U(y_{\lfloor t \rfloor})\right)$$
$$\qquad = -2z_t \cdot \left(\nabla U(x_t) - \nabla U(y_t)\right) + 2|w_t|^2 + \epsilon(t),$$

where

$$\delta(t) = 2z_t \cdot (w_{\lfloor t \rfloor} - w_t) - hz_t \cdot \left(\nabla U(x_{\lfloor t \rfloor}) - \nabla U(y_{\lfloor t \rfloor})\right)$$

$$\qquad = \delta_1(t) + \delta_2(t) + \delta_3(t) \quad \text{with}$$

$$\delta_1(t) = 2\left(t - \lfloor t \rfloor - h/2\right)z_{\lfloor t \rfloor} \cdot \left(\nabla U(x_{\lfloor t \rfloor}) - \nabla U(y_{\lfloor t \rfloor})\right),$$

$$\delta_2(t) = 2\left(t - \lfloor t \rfloor - h/2\right)(z_t - z_{\lfloor t \rfloor}) \cdot \left(\nabla U(x_{\lfloor t \rfloor}) - \nabla U(y_{\lfloor t \rfloor})\right),$$

$$\delta_3(t) = \left(t - \lfloor t \rfloor\right)z_t \cdot \left(\nabla U(x_{\lceil t \rceil}) - \nabla U(y_{\lceil t \rceil}) - \nabla U(x_{\lfloor t \rfloor}) + \nabla U(y_{\lfloor t \rfloor})\right),$$

$$\epsilon(t) = \epsilon_1(t) + \epsilon_2(t) + \epsilon_3(t) \quad \text{with}$$

$$\epsilon_1(t) = 2z_t \cdot \left(\nabla U(x_t) - \nabla U(y_t)\right)$$

$$+ z_t \cdot (-\nabla U(x_{\lfloor t \rfloor}) + \nabla U(y_{\lfloor t \rfloor}) - \nabla U(x_{\lceil t \rceil}) + \nabla U(y_{\lceil t \rceil})),$$

$$\epsilon_2(t) = 2w_t \cdot (w_{\lfloor t \rfloor} - w_t),$$

$$\epsilon_3(t) = -hw_t \cdot (\nabla U(x_{\lfloor t \rfloor}) - \nabla U(y_{\lfloor t \rfloor})).$$

We see that

$$\dot{b}(t) = -2Ka(t) + \beta(t)$$

with a function β satisfying

(78) $$\beta(t) \le 2|w_t|^2 + \epsilon(t).$$

The initial value problem

$$\dot{a} = b + \delta, \qquad a(0) = |z_0|^2,$$

$$\dot{b} = -2Ka + \beta, \qquad b(0) = 0,$$

has a unique solution that is given by

(79)
$$a(t) = \cos(\sqrt{2K}t)|z_0|^2 + \int_0^t \cos(\sqrt{2K}(t-r))\delta(r)\,dr$$

$$+ \int_0^t \frac{1}{\sqrt{2K}} \sin(\sqrt{2K}(t-r))\beta(r)\,dr.$$

We now bound the terms δ, ϵ and β. Note first that the assumptions imply that $Kt^2 \le Lt^2 \le 1 \le \pi^2/2$. Hence $t \le \pi/\sqrt{2K}$, and thus $\sin(\sqrt{2K}(t-r)) \ge 0$ for any $r \in [0, t]$. Let $\bar{t} := (\lfloor t \rfloor + \lceil t \rceil)/2 = \lfloor t \rfloor + h/2$. Then for $f \in C^1$,

$$\left| \int_{\lfloor t \rfloor}^{\lceil t \rceil} (r - \bar{t}) f(r)\,dr \right| = \left| \int_{\lfloor t \rfloor}^{\lceil t \rceil} (r - \bar{t})(f(r) - f(\bar{t}))\,dr \right|$$

$$\le \int_{\lfloor t \rfloor}^{\lceil t \rceil} (r - \bar{t})^2\,dr \sup|f'| = \frac{h^3}{12} \sup|f'|.$$

Therefore, we obtain

$$\int_{\lfloor t \rfloor}^{\lceil t \rceil} \cos(\sqrt{2K}(t-r))\delta_1(r)\,dr \le \frac{h^3}{6}\sqrt{2K}L|z_{\lfloor t \rfloor}|^2.$$

In particular, for $t \in h\mathbb{Z}_+$,

(80) $$\int_0^t \cos(\sqrt{2K}(t-r))\delta_1(r)\,dr \le th^2 \frac{\sqrt{2K}L}{6} z_t^{\star,2},$$

where $z_t^{\star,2} := (z_t^\star)^2$. Moreover, $\delta_2(t)$ is given by

$$2(t - \bar{t})(t - \lfloor t \rfloor)\left(w_{\lfloor t \rfloor} - \frac{h}{2}(\nabla U(x_{\lfloor t \rfloor}) - \nabla U(y_{\lfloor t \rfloor}))\right) \cdot (\nabla U(x_{\lfloor t \rfloor}) - \nabla U(y_{\lfloor t \rfloor})),$$

and hence by (54), for $t \in h\mathbb{Z}_+$,

(81) $$\int_0^t \cos(\sqrt{2K}(t-r))\delta_2(r)\,dr \le \frac{1}{2}h^2 t\left(Lw_t^\star z_t^\star + \frac{h}{2}L^2 z_t^{\star,2}\right) \le \frac{5}{4}t^2 h^2 L^2 z_t^{\star,2}.$$

In order to control δ_3 in an efficient way note that

$$
\begin{aligned}
&\bigl|\nabla U(x_{\lceil t \rceil}) - \nabla U(x_{\lfloor t \rfloor}) - \nabla U(y_{\lceil t \rceil}) + \nabla U(y_{\lfloor t \rfloor})\bigr| \\
&= \left| \int_{\lfloor t \rfloor}^{\lceil t \rceil} (\nabla^2 U(x_r)\dot{x}_r - \nabla^2 U(y_r)\dot{y}_r)\, dr \right| \\
&\le M \int_{\lfloor t \rfloor}^{\lceil t \rceil} |z_r||\dot{x}_r|\, dr + L \int_{\lfloor t \rfloor}^{\lceil t \rceil} |\dot{z}_r|\, dr \\
&\le M h z_{\lceil t \rceil}^{\star}\Bigl(|v_{\lfloor t \rfloor}| + \frac{hL}{2}|x_{\lfloor t \rfloor}|\Bigr) + Lh\Bigl(|w_{\lfloor t \rfloor}| + \frac{hL}{2}|z_{\lfloor t \rfloor}|\Bigr).
\end{aligned}
$$

(82)

Therefore, we obtain for $t \in h\mathbb{Z}_+$, by (54),

(83)
$$
\begin{aligned}
&\int_0^t \cos(\sqrt{2K}(t-r))\delta_3(r)\, dr \\
&\le t\delta_3^{\star}(t) \le th^2\Bigl(Mz_t^{\star,2}\Bigl(v_t^{\star} + \frac{hL}{2}x_t^{\star}\Bigr) + 2L^2 t z_t^{\star,2} + \frac{hL^2}{2}z_t^{\star,2}\Bigr).
\end{aligned}
$$

Next, we derive bounds for $\beta(t)$. We first observe that by (78) and (54),

$$
\beta(t) \le 2L^2 t^2 z_t^{\star,2} + \epsilon^{\star}(t),
$$

and hence

(84)
$$
\begin{aligned}
&\int_0^t \frac{1}{\sqrt{2K}} \sin(\sqrt{2K}(t-r))\beta(r)\, dr \\
&\le 2L^2 \int_0^t (t-r)r^2\, dr\, z_t^{\star,2} + \int_0^t (t-r)\, dr\, \epsilon^{\star}(t) \\
&= \frac{1}{6}L^2 t^4 z_t^{\star,2} + \frac{1}{2}t^2 \epsilon^{\star}(t).
\end{aligned}
$$

The terms ϵ_1, ϵ_2 and ϵ_3 can be controlled similarly to δ_1, δ_2 and δ_3. Analogously to (82), we obtain

$$
\begin{aligned}
\bigl|\epsilon_1(t)\bigr| &\le z_t^{\star}\bigl|\nabla U(x_t) - \nabla U(y_t) - \nabla U(x_{\lfloor t \rfloor}) + \nabla U(y_{\lfloor t \rfloor})\bigr| \\
&\quad + z_t^{\star}\bigl|\nabla U(x_{\lceil t \rceil}) - \nabla U(y_{\lceil t \rceil}) - \nabla U(x_t) + \nabla U(y_t)\bigr| \\
&\le 2Mh z_{\lceil t \rceil}^{\star,2}\Bigl(v_{\lceil t \rceil}^{\star} + \frac{hL}{2}x_{\lceil t \rceil}^{\star}\Bigr) + 2Lh z_{\lceil t \rceil}^{\star}\Bigl(w_{\lceil t \rceil}^{\star} + \frac{hL}{2}z_{\lceil t \rceil}^{\star}\Bigr),
\end{aligned}
$$

and thus by (54), for $t \in h\mathbb{Z}_+$,

$$
\begin{aligned}
\epsilon_1^{\star}(t) &\le h z_t^{\star,2}(2Mv_t^{\star} + 4L^2 t + hLMx_t^{\star} + hL^2), \\
\epsilon_2^{\star}(t) &\le 2Lh w_t^{\star} z_t^{\star} \le 4L^2 ht z_t^{\star,2}, \\
\epsilon_3^{\star}(t) &\le Lh w_t^{\star} z_t^{\star} \le 2L^2 ht z_t^{\star,2}.
\end{aligned}
$$

Thus in total, we obtain by (84),

$$
\begin{aligned}
&\int_0^t \frac{1}{\sqrt{2K}} \sin(\sqrt{2K}(t-r))\beta(r)\, dr \\
&\le \Bigl(\frac{1}{6}L^2 t^4 + \frac{1}{2}t^2 h(2Mv_t^{\star} + 10L^2 t + hLMx_t^{\star} + hL^2)\Bigr)z_t^{\star,2}.
\end{aligned}
$$

By (79), (80), (81) and (84), we obtain

$$(85) \quad |z_t|^2 = a(t) \le \cos(\sqrt{2K}t)|z_0|^2 + \left(\frac{1}{6}L^2t^4 + A_t\right)z_t^{\star,2} \quad \text{where}$$

$$A_t = (t^2h + th^2)(Mv_t^\star + 5L^2t + hLMx_t^\star + hL^2) + L^{3/2}th^2 + 2L^2h^2t^2.$$

By Lemma 3.1, there is a finite constant C that depends only on K, L and M such that for any $t \in h\mathbb{Z}_+$ satisfying (56), we have

$$(86) \quad A_t \le Ct^2h(1 + |x| + |v|).$$

Noting that $\cos(\sqrt{2K}t) \le 1 - Kt^2 + K^2t^4/6$, and $K^2t^4 \le L^2t^4 \le Kt^2$ by (56), we obtain by (85) and (57),

$$(87) \quad |z_t|^2 \le (1 - Kt^2)|z_0|^2 + \left(\frac{1}{3}Kt^2 + Cht^2(1 + |x| + |v|)\right)z_t^{\star,2}$$

$$(88) \quad \le (1 - Kt^2)|z_0|^2 + \frac{1}{2}Kt^2z_t^{\star,2}$$

for any $t \in h \cdot \mathbb{Z}_+$ s.t. (56) holds. This inequality then implies

$$(89) \quad |z_t|^2 \le \left(1 - \frac{1}{2}Kt^2\right)|z_0|^2$$

for t as before. Indeed, suppose first that $h > 0$. Then (89) follows directly from (88) if $|z_t| \le |z_0|$ holds for any $t > 0$ satisfying (56). Now suppose for a contradiction that there exists $t > 0$ s.t. (56) holds and $|z_t| > |z_0|$. Since z_t is linear on each partition interval, we may assume that $t \in h\mathbb{Z}_+$. Let t_0 denote the smallest $s \in h\mathbb{Z}_+$ for which $|z_s| > |z_0|$. Then $z_{t_0}^\star = |z_{t_0}|$, and hence by (88), $|z_{t_0}| \le |z_0|$ in contradiction to the definition of t_0. Thus (89) holds for all t as above. For $h = 0$, we can argue similarly by the intermediate value theorem.

Summarizing, we have shown that (89) holds for $t \in h\mathbb{Z}_+$ satisfying (56) provided $|z_s| \ge \mathcal{R}$ for all $s \in [0, t]$ and h satisfies (57). To conclude the proof suppose that $|z_0| \ge 2\mathcal{R}$. We claim that then $|z_t| \ge \mathcal{R}$ holds for all t satisfying (56). Indeed let $t_1 := \inf\{t : |z_t| < \mathcal{R}\}$. Then $|z_s| \ge \mathcal{R}$ on $[0, t_1]$. Suppose for a contradiction that $L(t_1^2 + t_1h) < K/L \le 1$. Then by (89), $|z_s| \le |z_0|$ for $s \in [0, t_1]$. Hence by (76) and (75),

$$|z_{t_1} - z_0| = \left|\int_0^{t_1}\left(w_{\lfloor s\rfloor} - \frac{h}{2}(\nabla U(x_{\lfloor s\rfloor}) - \nabla U(y_{\lfloor s\rfloor}))\right)ds\right|$$

$$\le \frac{Lt_1^2}{2}|z_0| + \frac{Lht_1}{2}|z_0| = \frac{1}{2}L(t_1^2 + ht_1)|z_0| < \frac{1}{2}|z_0|,$$

and thus $|z_{t_1}| > \frac{1}{2}|z_0| \ge \mathcal{R}$ in contradiction to the definition of t_1. $\square$

PROOF OF LEMMA 3.6.   Fix $x, v, z, w \in \mathbb{R}^d$. We set $x_t = q_t(x, v)$, $v_t = p_t(x, v)$ and $H_t = H(x_t, v_t) = \frac{1}{2}|v_t|^2 + U(x_t)$. Then

$$\frac{d}{dt}H_t = v_t \cdot \frac{d}{dt}v_t + \nabla U(x_t) \cdot \frac{d}{dt}x_t$$

$$= -\frac{1}{2}v_t \cdot (\nabla U(x_{\lfloor t\rfloor}) + \nabla U(x_{\lceil t\rceil})) + v_{\lfloor t\rfloor} \cdot \nabla U(x_t)$$

$$- \frac{h}{2}\nabla U(x_{\lfloor t\rfloor}) \cdot \nabla U(x_t),$$

$$= \mathrm{I}_t + \mathrm{II}_t + \mathrm{III}_t + \mathrm{IV}_t, \quad \text{where}$$

$$(90) \qquad \mathrm{I}_t = -\frac{1}{2} v_{\lfloor t \rfloor} \cdot (\nabla U(x_{\lfloor t \rfloor}) + \nabla U(x_{\lceil t \rceil}) - 2\nabla U(x_t)),$$

$$(91) \qquad \mathrm{II}_t = -\frac{1}{2}(v_t - v_{\lfloor t \rfloor}) \cdot (\nabla U(x_{\lfloor t \rfloor}) + \nabla U(x_{\lceil t \rceil}) - 2\nabla U(x_t)),$$

$$(92) \qquad \mathrm{III}_t = (v_{\lfloor t \rfloor} - v_t) \cdot \nabla U(x_t) - \frac{h}{2}|\nabla U(x_t)|^2,$$

$$(93) \qquad \mathrm{IV}_t = \frac{h}{2}(\nabla U(x_t) - \nabla U(x_{\lfloor t \rfloor})) \cdot \nabla U(x_t).$$

Furthermore,

$$v_{\lfloor t \rfloor} - v_t = \frac{t - \lfloor t \rfloor}{2}(\nabla U(x_{\lfloor t \rfloor}) + \nabla U(x_{\lceil t \rceil}))$$

$$= (t - \lfloor t \rfloor)\nabla U(x_t) + \frac{t - \lfloor t \rfloor}{2}(\nabla U(x_{\lfloor t \rfloor}) + \nabla U(x_{\lceil t \rceil}) - 2\nabla U(x_t)),$$

$$\nabla U(x_t) - \nabla U(x_{\lfloor t \rfloor})$$

$$= \int_{\lfloor t \rfloor}^t \nabla^2 U(x_s) \cdot \left(v_{\lfloor t \rfloor} - \frac{h}{2}\nabla U(x_{\lfloor t \rfloor}) \right) ds$$

$$= (t - \lfloor t \rfloor)\nabla^2 U(x_{\lfloor t \rfloor}) \cdot v_{\lfloor t \rfloor} - \int_{\lfloor t \rfloor}^t (\nabla^2 U(x_s) - \nabla^2 U(x_{\lfloor t \rfloor})) \cdot v_{\lfloor t \rfloor}\, ds$$

$$- \frac{h}{2}\int_{\lfloor t \rfloor}^t \nabla^2 U(x_s) \cdot \nabla U(x_{\lfloor t \rfloor})\, ds \quad \text{and hence,}$$

$$2\nabla U(x_t) - \nabla U(x_{\lfloor t \rfloor}) - \nabla U(x_{\lceil t \rceil})$$

$$= (t - \lfloor t \rfloor + t - \lceil t \rceil)\nabla^2 U(x_{\lfloor t \rfloor}) \cdot v_{\lfloor t \rfloor} + \int_{\lfloor t \rfloor}^t (\nabla^2 U(x_s) - \nabla^2 U(x_{\lfloor t \rfloor})) \cdot v_{\lfloor t \rfloor}\, ds$$

$$(94) \qquad - \int_t^{\lceil t \rceil} (\nabla^2 U(x_s) - \nabla^2 U(x_{\lfloor t \rfloor})) \cdot v_{\lfloor t \rfloor}\, ds - \frac{h}{2}\int_{\lfloor t \rfloor}^t \nabla^2 U(x_s) \cdot \nabla U(x_{\lfloor t \rfloor})\, ds$$

$$+ \frac{h}{2}\int_t^{\lceil t \rceil} \nabla^2 U(x_s) \cdot \nabla U(x_{\lfloor t \rfloor})\, ds$$

$$= 2(t - \bar{t})\nabla^2 U(x_{\lfloor t \rfloor}) \cdot v_{\lfloor t \rfloor} + \mathrm{V}_t,$$

where $\bar{t} = (\lfloor t \rfloor + \lceil t \rceil)/2$ and

$$(95) \qquad |\mathrm{V}_t| \le M|v_{\lfloor t \rfloor}| \int_{\lfloor t \rfloor}^{\lceil t \rceil} |x_s - x_{\lfloor t \rfloor}|\, ds + \frac{1}{2}L^2 h^2 |x_{\lfloor t \rfloor}|$$

$$= \frac{1}{2}Mh^2|v_{\lfloor t \rfloor}|\left| v_{\lfloor t \rfloor} - \frac{h}{2}\nabla U(x_{\lfloor t \rfloor}) \right| + \frac{1}{2}L^2 h^2 |x_{\lfloor t \rfloor}|.$$

Consequently, $\mathrm{I}_t = \mathrm{I}_t^a + \mathrm{I}_t^b$, where

$$\mathrm{I}_t^a = (t - \bar{t})v_{\lfloor t \rfloor} \cdot \nabla^2 U(x_{\lfloor t \rfloor})v_{\lfloor t \rfloor}, \qquad \mathrm{I}_t^b = \frac{1}{2}v_{\lfloor t \rfloor} \cdot \mathrm{V}_t.$$

In particular, for any $t \in h\mathbb{Z}_+$, $\int_0^t \mathrm{I}_s^a\, ds = 0$, and

$$\mathrm{I}_t^{b,\star} = \sup_{s \le t}|\mathrm{I}_s^b| \le \frac{h^2}{4}\left(M(v_t^\star)^3 + \frac{hLM}{2}(v_t^\star)^2 x_t^\star + L^2 x_t^\star v_t^\star \right).$$

Similarly, we obtain

$$\begin{aligned}
|\text{II}_t| &\le \frac{L}{2}|v_t - v_{\lfloor t \rfloor}|\big(|x_{\lfloor t \rfloor} - x_t| + |x_{\lceil t \rceil} - x_t|\big) \\
&= \frac{L}{4}h^2\big|\big(\nabla U(x_{\lfloor t \rfloor}) + \nabla U(x_{\lceil t \rceil})\big)\big| \cdot \Big|v_{\lfloor t \rfloor} - \frac{h}{2}\nabla U(x_{\lfloor t \rfloor})\Big| \quad \text{for } t \ge 0,
\end{aligned}$$

$$\text{II}_t^\star \le \frac{L^2}{2}h^2\Big(x_t^\star v_t^\star + \frac{Lh}{2}(x_t^\star)^2\Big) \quad \text{for } t \in h\mathbb{Z}_+,$$

$$\text{III}_t = \text{III}_t^a + \text{III}_t^b + \text{III}_t^c \quad \text{where}$$

$$(96) \qquad \text{III}_t^a = -\Big(t - \lfloor t \rfloor - \frac{h}{2}\Big)|\nabla U(x_{\lfloor t \rfloor})|^2,$$

$$(97) \qquad \text{III}_t^b = \Big(v_t - v_{\lfloor t \rfloor} - \frac{h}{2}\nabla U(x_t)\Big) \cdot \big(\nabla U(x_t) - \nabla U(x_{\lfloor t \rfloor})\big),$$

$$(98) \qquad \text{III}_t^c = \frac{h}{2}\big(\nabla U(x_t) - \nabla U(x_{\lfloor t \rfloor})\big) \cdot \nabla U(x_t) = \text{IV}_t.$$

In particular, for any $t \in h\mathbb{Z}_+$, $\int_0^t \text{III}_s^a \, ds = 0$, and

$$\text{III}_t^{b,\star} \le \frac{3}{2}h^2 L^2\Big(v_t^\star x_t^\star + \frac{hL}{2}(x_t^\star)^2\Big),$$

$$\text{III}_t^{c,\star} = \text{IV}_t^\star \le \frac{1}{2}h^2 L^2\Big(v_t^\star x_t^\star + \frac{hL}{2}(x_t^\star)^2\Big).$$

By combining the bounds, we obtain for $t \in h\mathbb{Z}_+$:

$$\begin{aligned}
|H_t - H_0| &= \Big|\int_0^t \big(\text{I}_s^a + \text{I}_s^b + \text{II}_s + \text{III}_s^a + \text{III}_s^b + \text{III}_s^c + \text{IV}_s\big)\,ds\Big| \\
&\le th^2\Big(\frac{M}{4}(v_t^\star)^3 + \frac{hL}{8}(v_t^\star)^2 x_t^\star + 3L^2 v_t^\star x_t^\star + \frac{3}{2}hL^3(x_t^\star)^2\Big).
\end{aligned}$$

This implies the first claim (60), since for $t \in h\mathbb{Z}_+$ satisfying (46), both $x_t^\star$ and $v_t^\star$ are bounded by a constant multiple of $\max(|x|, |v|)$.

Next, we consider the derivative flow

$$x_t' = (\partial_{(z,w)} q_t)(x, v), \qquad v_t' = (\partial_{(z,w)} p_t)(x, v),$$

where the derivatives are taken w.r.t. the initial condition. We have

$$(99) \qquad \frac{d}{dt}x_t' = \partial_{(z,w)}\dot{x}_t = v_{\lfloor t \rfloor}' - \frac{h}{2}\nabla^2 U(x_{\lfloor t \rfloor})x_{\lfloor t \rfloor}',$$

$$(100) \qquad \frac{d}{dt}v_t' = \partial_{(z,w)}\dot{v}_t = -\frac{1}{2}\big(\nabla^2 U(x_{\lfloor t \rfloor})x_{\lfloor t \rfloor}' + \nabla^2 U(x_{\lceil t \rceil})x_{\lceil t \rceil}'\big)$$

with initial condition $(x_0', v_0') = (z, w)$. In particular, for $s, t \in \mathbb{R}_+$ s.t. $s \in [\lfloor t \rfloor, \lceil t \rceil]$,

$$|x_t' - x_s'| = |t - s| \cdot \Big|v_{\lfloor t \rfloor}' - \frac{h}{2}\nabla^2 U(x_{\lfloor t \rfloor})x_{\lfloor t \rfloor}'\Big| \le h\Big(v_t'^\star + \frac{hL}{2}x_t'^\star\Big),$$

$$|v_t' - v_s'| = \frac{1}{2}|t - s||\nabla^2 U(x_{\lfloor t \rfloor})x_{\lfloor t \rfloor}' + \nabla^2 U(x_{\lceil t \rceil})x_{\lceil t \rceil}'| \le hLx_{\lceil t \rceil}'^\star.$$

We now first derive a priori bounds for $x_t'^\star$ and $v_t'^\star$. By (99) and (100),

$$|x_t' - z - wt|$$

$$= \left| \frac{1}{2} \int_0^t \int_0^{\lfloor s \rfloor} \left(\nabla^2 U(x_{\lfloor r \rfloor}) x'_{\lfloor r \rfloor} + \nabla^2 U(x_{\lceil r \rceil}) x'_{\lceil r \rceil} \right) dr\, ds \right.$$

$$\left. - \frac{h}{2} \int_0^t \nabla^2 U(x_{\lfloor s \rfloor}) x'_{\lfloor s \rfloor}\, ds \right|$$

$$\leq \frac{L}{2} \int_0^t \int_0^{\lfloor s \rfloor} \left(|x'_{\lfloor r \rfloor}| + |x'_{\lceil r \rceil}| \right) dr\, ds + \frac{hL}{2} \int_0^t |x'_{\lfloor s \rfloor}|\, ds.$$

For $t \in h\mathbb{Z}_+$, we obtain

$$\max_{s \leq t} |x'_s - z - ws| \leq \frac{L}{2}(t^2 + ht)\left(\max_{s \leq t} |z + ws| + \max_{s \leq t} |x'_s - z - ws| \right).$$

Hence if $L(t^2 + ht) \leq 1$ then

(101) $$\max_{s \leq t} |x'_s - z - ws| \leq L(t^2 + ht) \max(|z|, |z + wt|).$$

Similarly, by (100) and (101),

(102) $$\max_{s \leq t} |v'_s - w| \leq Lt \max_{s \leq t} |x'_s| \leq 2Lt \max(|z|, |z + wt|).$$

Now we can derive bounds for H'_t. We have

(103) $$\frac{d}{dt} H'_t = \left(\frac{d}{dt} H_t \right)' = \mathrm{I}'_t + \mathrm{II}'_t + \mathrm{III}'_t + \mathrm{IV}'_t.$$

Similarly as above, we bound the terms I'_t, II'_t, III'_t and IV'_t individually. By (90), $\mathrm{I}'_t = \mathrm{VI}_t - \frac{1}{2} v_{\lfloor t \rfloor} \mathrm{VII}_t$, where

$$\mathrm{VI}_t = -\frac{1}{2} v'_{\lfloor t \rfloor} \left(\nabla U(x_{\lfloor t \rfloor}) + \nabla U(x_{\lceil t \rceil}) - 2\nabla U(x_t) \right),$$

$$\mathrm{VII}_t = \nabla^2 U(x_{\lfloor t \rfloor}) x'_{\lfloor t \rfloor} + \nabla^2 U(x_{\lceil t \rceil}) x'_{\lceil t \rceil} - 2\nabla^2 U(x_t) x'_t.$$

Similarly to the decomposition of I_t above, we have $\mathrm{VI}_t = \mathrm{VI}^a_t + \mathrm{VI}^b_t$ where $\mathrm{VI}^a_t = (t - \bar{t}) v'_{\lfloor t \rfloor} \cdot \nabla^2 U(x_{\lfloor t \rfloor}) v_{\lfloor t \rfloor}$ and $\mathrm{VI}^b_t = \frac{1}{2} v'_{\lfloor t \rfloor} \cdot \mathrm{V}_t$ In particular, for $t \in h\mathbb{Z}_+$,

$$\int_{\lfloor t \rfloor}^{\lceil t \rceil} \mathrm{VI}^a_s\, ds = 0 \quad \text{and} \quad \mathrm{VI}^{b,\star}_t \leq \frac{Mh^2}{4} \left(v'^\star_t v^\star_t + \frac{hL}{2} v'^\star_t x^\star_t \right).$$

Furthermore, $\mathrm{VII}_t = \mathrm{VII}^a_t + \mathrm{VII}^b_t + \mathrm{VII}^c_t$ with

$$\mathrm{VII}^a_t = \nabla^2 U(x_t)\left(x'_{\lfloor t \rfloor} + x'_{\lceil t \rceil} - 2x'_t \right)$$

$$= 2(\bar{t} - t) \nabla^2 U(x_{\lfloor t \rfloor}) \left(v'_{\lfloor t \rfloor} - \frac{h}{2} \nabla^2 U(x_{\lfloor t \rfloor}) x'_{\lfloor t \rfloor} \right),$$

$$\mathrm{VII}^b_t = \left(\nabla^2 U(x_{\lfloor t \rfloor}) + \nabla^2 U(x_{\lceil t \rceil}) - 2\nabla^2 U(x_t) \right) x'_{\lfloor t \rfloor},$$

$$\mathrm{VII}^c_t = \left(\nabla^2 U(x_{\lceil t \rceil}) - \nabla^2 U(x_t) \right) \left(x'_{\lceil t \rceil} - x'_{\lfloor t \rfloor} \right).$$

For $t \in h\mathbb{Z}_+$, $\int_0^t \mathrm{VII}^a_s\, ds = 0$. Moreover, similarly to (94) and (95), we have

$$2\nabla^2 U(x_t) - \nabla^2 U(x_{\lfloor t \rfloor}) - \nabla^2 U(x_{\lceil t \rceil}) = 2(t - \bar{t}) \nabla^3 U(x_{\lfloor t \rfloor}) \cdot v_{\lfloor t \rfloor} + \mathrm{VIII}_t,$$

where $|\mathrm{VII}_t| \leq \frac{1}{2} Nh^2 |v_{\lfloor t \rfloor}| |v_{\lfloor t \rfloor} - \frac{h}{2} \nabla U(x_{\lfloor t \rfloor})| + \frac{1}{2} LMh^2 |x_{\lfloor t \rfloor}|$. Therefore, we can decompose $\mathrm{VII}^b_t = \mathrm{VII}^d_t + \mathrm{VII}^e_t$ where $\int_0^t \mathrm{VII}^d_s\, ds = 0$ for $t \in h\mathbb{Z}_+$, and

$$\mathrm{VII}^{e,\star}_t \leq \frac{h^2}{4} \left(N v^{\star,3}_t + \frac{hLN}{2} v^{\star,2}_t x^\star_t + LM v^\star_t x^\star_t \right) x'^\star_t.$$

Furthermore, by (99), we have

$$\mathrm{VII}_t^{c,\star} \le Mh^2\Big(v_t^\star + \frac{hL}{2}x_t^\star\Big) \cdot \Big(v_t'^\star + \frac{hM}{2}x_t'^\star\Big).$$

For the second term we have $\mathrm{II}_t' = \mathrm{IX}_t + \mathrm{X}_t + \mathrm{XI}_t$ where

$$\mathrm{IX}_t = -\frac{1}{2}(v_t' - v_{\lfloor t \rfloor}')(\nabla U(x_{\lfloor t \rfloor}) + \nabla U(x_{\lceil t \rceil}) - 2\nabla U(x_t)),$$

$$\mathrm{X}_t = -\frac{1}{2}(v_t - v_{\lfloor t \rfloor})(\nabla^2 U(x_{\lfloor t \rfloor}) + \nabla^2 U(x_{\lfloor t \rfloor}) - 2\nabla^2 U(x_t))x_t',$$

$$\mathrm{XI}_t = -\frac{1}{2}(v_t - v_{\lfloor t \rfloor})(\nabla^2 U(x_{\lfloor t \rfloor})(x_{\lfloor t \rfloor}' - x_t') + \nabla^2 U(x_{\lceil t \rceil})(x_{\lceil t \rceil}' - x_t')).$$

For $t \in h\mathbb{Z}_+$, we obtain by (99) and (100),

$$\mathrm{IX}_t^\star \le LMh^2 x_t'^\star(v_t^\star + hLx_t^\star/2)/2,$$

$$\mathrm{X}_t^\star \le LMh^2 x_t^\star(v_t^\star + hLx_t^\star/2)x_t'^\star/2,$$

$$\mathrm{XI}_t^\star \le L^2h^2 x_t^\star(v_t'^\star + hLx_t'^\star/2)/2.$$

Furthermore, $\mathrm{III}_t' = (\mathrm{III}_t^a)' + (\mathrm{III}_t^b)' + (\mathrm{III}_t^c)'$. By (96) and the chain rule, $\int_0^t (\mathrm{III}_t^a)'\, ds = 0$ for $t \in h\mathbb{Z}_+$. Moreover, by (97) and the chain rule,

$$\begin{aligned}
|(\mathrm{III}_t^b)'| &\le \Big(|v_t' - v_{\lfloor t \rfloor}'| + \frac{hL}{2}|x_t'|\Big)L|x_t - x_{\lfloor t \rfloor}| \\
&\quad + \Big(|v_t - v_{\lfloor t \rfloor}| + \frac{hL}{2}x_{\lfloor t \rfloor}\Big)(L|x_t' - x_{\lfloor t \rfloor}'| + M|x_t - x_{\lfloor t \rfloor}||x_t'|) \\
&\le \frac{3}{2}L^2h^2|x_{\lfloor t \rfloor}'|\Big|v_{\lfloor t \rfloor} + \frac{hL}{2}x_{\lfloor t \rfloor}\Big| \\
&\quad + \frac{3}{2}L^2h^2|x_t|\Big|v_{\lfloor t \rfloor}' + \frac{hL}{2}x_{\lfloor t \rfloor}'\Big| \\
&\quad + \frac{3}{2}LMh^2|x_t|\Big|v_{\lfloor t \rfloor} + \frac{hL}{2}x_{\lfloor t \rfloor}\Big||x_t'|, \\
(\mathrm{III}_t^b)'^\star &\le \frac{3}{2}L^2h^2\Big(x_t'^\star\Big(v_t^\star + \frac{hL}{2}x_t^\star\Big) + x_t^\star\Big(v_t'^\star + \frac{hL}{2}x_t'^\star\Big)\Big) \\
&\quad + \frac{3}{2}LMh^2 x_t^\star\Big(v_t^\star + \frac{hL}{2}x_t^\star\Big)x_t'^\star.
\end{aligned}$$

Finally, a similar computation as for $(\mathrm{III}_t^b)'$ shows that

$$\begin{aligned}
(\mathrm{III}_t^c)'^\star = \mathrm{IV}_t'^\star &\le \frac{1}{2}L^2h^2\Big(x_t'^\star\Big(v_t^\star + \frac{hL}{2}x_t^\star\Big) + x_t^\star\Big(v_t'^\star + \frac{hL}{2}x_t'^\star\Big)\Big) \\
&\quad + \frac{1}{2}LMh^2 x_t^\star\Big(v_t^\star + \frac{hL}{2}x_t^\star\Big)x_t'^\star.
\end{aligned}$$

Collecting all the bounds derived above, we eventually obtain

$$\begin{aligned}
|H_t' - H_0'| &\le th^2\Big(\frac{L_3}{4}(v_t^\star)^3 x_t'^\star + \frac{L_3 L}{8}h(v_t^\star)^2 x_t^\star x_t'^\star \\
&\quad + Q_1(v_t^\star, x_t^\star)v_t'^\star + x_t Q_2(v_t^\star, x_t^\star)x_t'^\star\Big)
\end{aligned}$$

for $t \in h\mathbb{Z}$, where Q_1 and Q_2 are explicit quadratic forms. We can conclude that

$$|H_t' - H_0'| \le C_2 t (1+t) h^2 \max(|x_0|, |v_0|)^3 \max(|x_0'|, |v_0'|). \qquad \square$$

PROOF OF LEMMA 3.7.   Let $p = 0$ or $p \ge 1$. Then by definition of η,

$$E\big[|e \cdot \xi|^p; \xi - \eta \ne -\gamma z\big]$$
$$\le E\big[|e \cdot \xi|^p; \widetilde{U} > \varphi_{0,1}(e \cdot \xi + \gamma|z|)/\varphi_{0,1}(e \cdot \xi)\big]$$
$$= \int_{-\infty}^{\infty} |t|^p \big(\varphi_{0,1}(t) - \varphi_{0,1}(t + |\gamma z|)\big)^+ dt$$
$$= \int_{-|\gamma z|/2}^{\infty} |t|^p \big(\varphi_{0,1}(t) - \varphi_{0,1}(t + |\gamma z|)\big)^+ dt$$
$$= \int_{-|\gamma z|/2}^{\infty} |t|^p \phi_{0,1}(t) \, dt - \int_{\gamma|z|/2}^{\infty} |t - |\gamma z||^p \varphi_{0,1}(t) \, dt$$
$$= \int_{-|\gamma z|/2}^{|\gamma z|/2} |t|^p \varphi_{0,1}(t) \, dt + \int_{|\gamma z|/2}^{\infty} \big(|t|^p - |t - |\gamma z||^p\big)\varphi_{0,1}(t) \, dt.$$

For $p = 0$, we directly obtain (62), and for $p \ge 1$,

$$E\big[|e \cdot \xi|^p; \xi - \eta \ne -\gamma z\big]$$
$$\le \frac{1}{\sqrt{2\pi}} 2^{-p} |\gamma z|^{p+1} + \int_{|\gamma z|/2}^{\infty} p|t|^{p-1} |\gamma z| \varphi_{0,1}(t) \, dt$$
$$\le C_p |\gamma z| \max(|\gamma z|, 1)^p,$$

where $C_p = \max(2^{-p}/\sqrt{2\pi}, 2pm_{p-1})$ with m_p denoting the pth moment of the standard normal distribution. Finally, since $\xi \sim \mathcal{N}(0, I_d)$ and the event $\{\xi - \eta \ne -\gamma z\}$ is measurable w.r.t. $\sigma(e \cdot \xi)$, we obtain

$$E\big[|\xi|^{2p}; \xi - \eta \ne -\gamma z\big]$$
$$\le 2^{p-1} E\big[|e \cdot \xi|^{2p}; \xi - \eta \ne -\gamma z\big] + 2^{p-1} E\big[|\xi - (e \cdot \xi)e|^{2p}\big] P[\xi - \eta \ne -\gamma z]$$
$$\le 2^{p-1} C_{2p} |\gamma z| \max(1, |\gamma z|)^{2p} + 2^{p-1}(d-1)^p m_{2p} |\gamma z|/\sqrt{2\pi}. \qquad \square$$

PROOF OF THEOREM 3.8.   Recall that

(104) $$A(x) = \big\{U \le \exp(-H(\phi_T(x, \xi)) + H(x, \xi))\big\} \quad \text{and}$$

(105) $$\hat{A}(x) = \big\{U \le \exp(-H(\phi_T(y, \eta)) + H(y, \eta))\big\}.$$

Therefore, and since U is independent of ξ, we obtain by Lemma 3.6,

$$P\big[A(x)^C|\xi\big] = |1 - \exp(-(H(\phi_T(x, \xi)) - H(x, \xi))^+)|$$
$$\le (H(\phi_T(x, \xi)) - H(x, \xi))^+$$
$$\le C_1 T (1+T) h^2 \max(|x|, |\xi|)^3, \quad \text{and hence}$$
$$P\big[A(x)^C\big] \le C_1 T (1+T) h^2 E\big[\max(|x|^3, |\xi|^3)\big]$$
$$\le C_1 T (1+T) h^2 \big(|x|^3 + d^{3/2}\sqrt{8/\pi}\big).$$

Since $\hat{A}(y)$ is defined similarly to $A(x)$ with x, ξ replaced by y, η and $\eta \sim \xi$, we obtain corresponding bounds for $P[\hat{A}(y)^C|\eta]$ and $P[\hat{A}(y)^C]$.

Next, we derive the corresponding bounds for the probabilities that the proposed move is accepted for one of the two components and rejected for the other. By independence of U from ξ and η and by Lemma 3.6, we have

$$P[A(x)\Delta\hat{A}(y)|\xi, \eta]$$

$$= |\exp(-(H(\phi_T(x, \xi)) - H(x, \xi))^+) - \exp(-(H(\phi_T(y, \eta)) - H(y, \eta))^+)|$$

$$\leq |[H(\phi_T(x, \xi)) - H(x, \xi)] - [H(\phi_T(y, \eta)) - H(y, \eta)]|$$

$$\leq \int_0^1 |\partial_{(x-y,\xi-\eta)}(H \circ \phi_T)(x_u, \xi_u) - \partial_{(x-y,\xi-\eta)}H(x_u, \xi_u)| \, du$$

$$\leq C_2 T(1 + T)h^2 \int_0^1 \max(|x_u|, |\xi_u|)^3 \, du \max(|x - y|, |\xi - \eta|)$$

$$\leq C_2 T(1 + T)h^2 \max(|x - y|, |\xi - \eta|) \max(|x|, |y|, |\xi|, |\eta|)^3,$$

where $x_u = ux + (1 - u)y$, $\xi_u = u\xi + (1 - u)\eta$, $z = x - y$ and $W = \xi - \eta$. This proves (71), and (69) can be shown similarly with η replaced by ξ.

Next, we bound the unconditioned probabilities of acceptance rejection events. At first we observe that by (69) and since $\xi \sim \mathcal{N}(0, I_d)$,

$$P[A(x)\Delta A(y)] \leq C_2 T(1 + T)h^2|x - y|E[\max(|x|, |y|, |\xi|)^3]$$

$$\leq C_2 T(1 + T)h^2|x - y|(\max(|x|, |y|)^3 + d^{3/2}\sqrt{8/\pi}),$$

which implies (70). The proof of a corresponding bound for the expectation in (72) is slightly more complicated. We first note that by (71),

$$E[\max(|x|, |y|, |\xi|, |\eta|); (A(x)\Delta\hat{A}(y)) \cap \{W = -\gamma z\}]$$

(106)
$$\leq C_2 T(1 + T)h^2 \max(1, \gamma)|z|E[\max(|x|, |y|, |\xi|, |\xi + \gamma z|)^4]$$

$$\leq C_2 T(1 + T)h^2 \max(1, \gamma)|z|(\max(|x|, |y|)^4 + E[(|\xi| + |\gamma z|)^4]).$$

Secondly, on $\{W \neq -\gamma z\}$, we have $\eta = \xi - 2(e \cdot \xi)e$ where $e = z/|z|$. In particular, $\eta = \xi$. Therefore, by (71),

$$E[\max(|x|, |y|, |\xi|, |\eta|); (A(x)\Delta\hat{A}(y)) \cap \{W \neq -\gamma z\}]$$

(107)
$$\leq C_2 T(1 + T)h^2 E[\max(|z|, 2|e \cdot \xi|) \max(|x|, |y|, |\xi|)^4; W \neq -\gamma z]$$

$$\leq C_2 T(1 + T)h^2 E[(|z| + 2|e \cdot \xi|)(\max(|x|, |y|)^4 + |\xi|^4); W \neq -\gamma z].$$

By (106), (107), and by the bounds in Lemma 3.7, we can conclude that there is a finite constant C_3 depending only on L, M and N such that for $|\gamma z| \leq 1$,

$$E[\max(|x|, |y|, |\xi|, |\eta|); A(x)\Delta\hat{A}(y)]$$

$$\leq C_3 T(1 + T)h^2 \max(1, \gamma)|z|(\max(|x|, |y|)^4 + d^2).$$

This proves the last assertion of the theorem. $\quad\square$

5. Proofs of main results.

PROOF OF THEOREM 2.1.   We fix $x, y \in \mathbb{R}^d$ such that $|x - y| \geq 2\mathcal{R}$ and $\max(|x|, |y|) \leq R_2$. Since synchronous coupling is applied for $|x - y| \geq 2\mathcal{R}$, we have $\eta = \xi$. Hence by (51) and Lemma 3.4 with $h = 0$, we obtain

$$R'(x, y) = |q_T(x, \xi) - q_T(y, \xi)| \leq (1 - KT^2/2)r(x, y)$$

provided $LT^2 \leq K/L$. $\quad\square$

PROOF OF THEOREM 2.2.   We directly give the more involved proof for the case of adjusted numerical HMC. For unadjusted numerical HMC, the proof simplifies considerably, see the comments at the end of the proof.

Without loss of generality, we may assume that R_2 is chosen sufficiently large such that

$$(108) \qquad P\big[|\xi| > R_2\big] \le \frac{K}{70L}.$$

We fix $x, y \in \mathbb{R}^d$ such that $|x - y| \ge 2\mathcal{R}$ and $\max(|x|, |y|) \le R_2$. Since synchronous coupling is applied for $|x - y| \ge 2\mathcal{R}$, we have $\eta = \xi$, and hence

$$R'(x, y) = \big|q_T(x, \xi) - q_T(y, \xi)\big| \quad \text{on } A(x) \cap A(y),$$

$$R'(x, y) = r(x, y) \quad \text{on } A(x)^C \cap A(y)^C.$$

Moreover, on $A(x) \cap A(y)^C$ we have $Y' = y$, and thus

$$R'(x, y) - r(x, y) = \big|q_T(x, \xi) - y\big| - |x - y| \le \big|q_t(x, \xi) - x\big|.$$

Similarly, on $A(x)^C \cap A(y)$,

$$R' - r \le \big|q_T(y, \xi) - y\big|.$$

Therefore, we obtain

$$
\begin{aligned}
(109) \qquad E\big[R'(x, y) - r(x, y)\big] \le{}& E\big[\big|q_T(x, \xi) - q_T(y, \xi)\big| - |x - y|; A(x) \cap A(y)\big] \\
& + E\big[\big|q_T(x, \xi) - x\big|; A(x) \cap A(y)^C\big] \\
& + E\big[\big|q_T(y, \xi) - y\big|; A(x)^C \cap A(y)\big] \\
={}& : \mathrm{I} + \mathrm{II} + \mathrm{III}.
\end{aligned}
$$

In order to control the first term, we choose a constant $C \in (0, \infty)$ as in Lemma 3.4, and we assume $h \le \min(h_1, h_2)$ where $h_2 := \frac{K}{C(1+2R_2)}$. Then by Lemma 3.4,

$$\big|q_T(x, \xi) - q_T(y, \xi)\big| \le \Big(1 - \frac{1}{4} K T^2\Big)|x - y| \quad \text{if } |\xi| \le R_2.$$

Therefore, by (58), and since $K \le L$,

$$
\begin{aligned}
\mathrm{I} \le{}& -\frac{1}{4} K T^2 r(x, y) P\big[A(x) \cap A(y) \cap \{|\xi| \le R_2\}\big] \\
& + (LT^2 + LTh) r(x, y) P\big[|\xi| > R_2\big] \\
\le{}& -\frac{1}{4} K T^2 r(x, y) P\big[A(x) \cap A(y)\big] + \Big(\frac{5}{4} LT^2 + LTh\Big) r(x, y) P\big[|\xi| > R_2\big] \\
\le{}& -\frac{1}{4} K T^2 r(x, y) P\big[A(x) \cap A(y)\big] + \frac{9}{4} LT^2 r(x, y) P\big[|\xi| > R_2\big]
\end{aligned}
$$

for $T \in h \cdot \mathbb{N}$. For $h \le h_1$ we have

$$(110) \qquad L(T^2 + Th) \le K/L \le 1.$$

Therefore, by Theorem 3.8, for $|x|, |y| \le R_2$,

$$(111) \qquad P\big[A(x)^C\big] + P\big[A(y)^C\big] \le 2C_1 T(1 + T)(R_2^3 + 2d^{3/2})h^2.$$

We choose $h_3 > 0$ such that for $h \le h_3$, the expression on the r.h.s. is smaller than $1/5$. Because of (23), this can be achieved with h_3^{-2} of order $O(R_2^3 + d^{3/2})$. For $h \le h_3$, we obtain

$$P\big[A(x) \cap A(y)\big] \ge 1 - \frac{1}{5} = \frac{4}{5}.$$

Therefore, and by (108),

$$I \le -\frac{1}{5}KT^2r + \frac{9}{4}LT^2r\,P[|\xi| > R_2] \le -\frac{1}{6}KT^2r.$$

In order to control II, we note that by Lemma 3.1,

$$|q_T(x,\xi) - x| \le T|\xi| + \max(|x|, |x + T\xi|) \le |x| + 2T|\xi|$$

provided $L(T^2 + Th) \le 1$. Hence in this case we obtain

$$\mathrm{II} \le E[|x| + 2T|\xi|; A(x) \cap A(y)^C].$$

A corresponding bound with x and y interchanged holds for III. Hence by the bound (69) for the conditional AR probability given ξ,

$$\mathrm{II} + \mathrm{III} \le E[\max(|x|, |y|) + 2T|\xi|; A(x)\Delta A(y)]$$
$$\le C_2 T(1 + T)h^2|x - y|(1 + 2T)E[\max(|x|, |y|, |\xi|)^4]$$
$$\le 2C_2 T(1 + T)^2h^2(R_2^4 + 3d^2)r.$$

We choose a strictly positive constant h_4 such that for $h \le h_4$, the right hand side is smaller than $\frac{1}{24}KT^2r$. By (23), this can be achieved with h_4^{-1} of order $O((R_2^2 + d)K^{-1/2}T^{-1/2})$. Let $h_0 = \min(h_2, h_3, h_4)$. Then for $h \le \min(h_0, h_1)$, we obtain

$$\mathrm{I} + \mathrm{II} + \mathrm{III} \le -\frac{1}{5}KT^2r + \frac{1}{24}KT^2r \le -\frac{1}{8}KT^2r.$$

This completes the proof for adjusted numerical HMC.

For unadjusted numerical HMC, the argument simplifies since the rejection events $A(x)^C$ and $A(y)^C$ are empty. Thus it suffices to bound I, which can be done similarly as above for $h \le \min(h_0, h_1)$ where $h_0 := h_2$. $\square$

Next, we directly prove our main result for numerical HMC. Afterwards, we will give the corresponding proof of Theorem 2.3 for exact HMC, which essentially is an easy special case of the more difficult proof for numerical HMC.

PROOF OF THEOREM 2.4.   As above, we directly prove the result for adjusted numerical HMC, and we comment on the simplifications in the unadjusted case in the end. The parameters γ, a and R_1 have been chosen in (28), (29) and (30) such that the following conditions are satisfied:

$$(112) \qquad\qquad\qquad\qquad \gamma T \le 1,$$

$$(113) \qquad\qquad\qquad\qquad L(T + h) \le \gamma/4,$$

$$(114) \qquad\qquad\qquad\qquad \gamma \mathcal{R} \le 1/4,$$

$$(115) \qquad\qquad\qquad\qquad aT \ge 1,$$

$$(116) \qquad\qquad\qquad\qquad R_1 \ge \frac{5}{2}(1 + \gamma T)\mathcal{R},$$

$$(117) \qquad\qquad\qquad \exp(a(R_1 - 2\mathcal{R})) \ge 20.$$

Indeed, (112) and (114) hold by (28), (113) holds by (28) and (27), (115) holds by (29), (116) holds by (30) and (28), and (117) holds by (30) and (29). The bounds (112)–(117) will be essential in the following arguments. We have chosen γ and a as large resp. small as possible such that (112), (114) and (115) hold. Then (113) implies the additional constraints on T in (27), and R_1 is chosen such that (116) and (117) are satisfied.

To prove contractivity, we fix $x, y \in \mathbb{R}^d$ such that $\max(|x|, |y|) \leq R_2$. Since x and y are fixed, we briefly write r and R' instead of $r(x, y)$ and $R'(x, y)$. We consider separately the cases where $|x - y| \geq 2\mathcal{R}$ and $|x - y| < 2\mathcal{R}$.

(i) *Contractivity for $|x - y| \geq 2\mathcal{R}$.* For $|x - y| \geq 2\mathcal{R}$, we can apply the result of Theorem 2.2. Indeed, choose h_0 as in Theorem 2.2. Then for $h \leq h_0$, by concavity of f and by (24),

$$(118) \qquad E[f(R') - f(r)] \leq f'(r) E[R' - r] \leq -\frac{1}{4} KT^2 r f'(r) \leq -c_1 f(r),$$

where the lower bound c_1 for the contraction rate is given by

$$(119) \qquad c_1 = \frac{1}{4} KT^2 \inf_{r > 0} \frac{r f'(r)}{f(r)}.$$

Recall that f is concave with $f(0) = 0$, and f is linear for $r \geq R_1$. Hence the function $r \mapsto r f'(r)/f(r)$ attains its minimum at R_1, where $f'(R_1) = e^{-aR_1}$ and $f(R_1) = \int_0^{R_1} e^{-as} \, ds \leq \min(R_1, a^{-1})$. Therefore, by (29) and (30),

$$
\begin{aligned}
(120) \qquad c_1 &\geq \frac{1}{4} KT^2 \max(1, aR_1) e^{-aR_1} \geq \frac{1}{4} KT^2 \frac{5}{2}\left(1 + \frac{\mathcal{R}}{T}\right) e^{-5/2} e^{-\frac{5\mathcal{R}}{2T}} \\
&> \frac{1}{20} KT^2 \left(1 + \frac{\mathcal{R}}{T}\right) e^{-\frac{5\mathcal{R}}{2T}}.
\end{aligned}
$$

(ii) *Contractivity for $|x - y| < 2\mathcal{R}$.* For $|x - y| < 2\mathcal{R}$, we apply the coupling defined by (19) and (20). Let $z = x - y$ and $W = \xi - \eta$. Since $R' = r$ on $A(x)^C \cap \hat{A}(y)^C$, we have

$$E[f(R') - f(r)] = \text{I} + \text{II} + \text{III} + \text{IV} \quad \text{where}$$

$$\text{I} = E[f(R') - f(r); A(x) \cap \hat{A}(y) \cap \{W = -\gamma z\}],$$

$$\text{II} = E[f(R' \wedge R_1) - f(r); A(x) \cap \hat{A}(y) \cap \{W \neq -\gamma z\}],$$

$$\text{III} = E[f(R') - f(R' \wedge R_1); A(x) \cap \hat{A}(y) \cap \{W \neq -\gamma z\}],$$

$$\text{IV} = E[f(R') - f(r); A(x) \triangle \hat{A}(y)].$$

Only the first term is responsible for contractivity. The other terms are perturbations that have to be controlled. We will now derive upper bounds for each of the four terms. We remark at first that on $A(x) \cap \hat{A}(y)$,

$$(121) \qquad R' = |q_T(x, \xi) - q_T(y, \eta)| \leq |z + WT| + \max(|z|, |z + WT|) LT(T + h)$$

by Lemma 3.2.

I. On $A(x) \cap \hat{A}(y) \cap \{W = -\gamma z\}$, we obtain by (121), (112) and (113),

$$
\begin{aligned}
R' &\leq |(1 - \gamma T)z| + \max(|z|, |(1 - \gamma T)z|) LT(T + h) \\
&\leq \left(1 - \gamma T + \frac{1}{4}\gamma T\right) |z| = \left(1 - \frac{3}{4}\gamma T\right) r.
\end{aligned}
$$

Therefore, by concavity of f,

$$
\begin{aligned}
\text{I} &\leq f'(r) E[R' - r; A(x) \cap \hat{A}(y) \cap \{W = -\gamma z\}] \\
&\leq -\frac{3}{4}\gamma T r f'(r)(1 - P[W \neq -\gamma z] - P[A(x)^C] - P[\hat{A}(y)^C]).
\end{aligned}
$$

By Lemma 3.7 and by (114),

$$P[W \neq -\gamma z] \leq \frac{\gamma r}{\sqrt{2\pi}} \leq \frac{1}{4\sqrt{2\pi}} < \frac{1}{10}.$$

Furthermore, by Theorem 3.8, there is a finite constant C_1 depending only on L, M and N such that

$$P[A(x)^C] + P[\hat{A}(y)^C] \le C_1 T(1+T)(|x|^3 + |y|^3 + 4d^{3/2})h^2.$$

Since $\max(|x|, |y|) \le R_2$ and (27) holds, we can conclude that there is a constant $h_5 > 0$ depending only on L, M, N, d and R_2 such that for $h \le h_5$,

$$(122) \qquad\qquad I \le -\frac{27}{40}\gamma T r f'(r).$$

Furthermore, for fixed L, M and N, the constant h_5 can be chosen by (27) such that h_5^{-2} is of order $O(R_2^3 + d^{3/2})$.

II. By definition of f, we have for $s \le R_1$,

$$f(s) - f(r) = \int_r^s e^{-at}\, dt \le \frac{1}{a} e^{-ar} = \frac{1}{a} f'(r).$$

Therefore, by (115) and by Lemma 3.7, the second term can be bounded by

$$(123) \qquad\qquad II \le \frac{1}{a} f'(r) P[W \ne -\gamma z] \le \frac{\gamma T}{\sqrt{2\pi}} r f'(r) < \frac{2}{5}\gamma T r f'(r).$$

III. If $W \ne -\gamma z$ then by definition of the coupling,

$$W = \xi - \eta = 2(e \cdot \xi)e \quad \text{where } e = z/|z|,$$

and hence $|z + WT| = |r + 2e \cdot \xi T|$. Therefore on $A(x) \cap \hat{A}(y) \cap \{W \ne -\gamma z\}$,

$$R' \le (1 + LT(T+h))|r + 2e \cdot \xi T| \le \frac{5}{4}|r + 2e \cdot \xi T|$$

by (121), (113) and (112). Thus

$$E[(R' - R_1)^+; A(x) \cap \hat{A}(y) \cap \{W \ne -\gamma z\}]$$

$$\le E\left[\left(\frac{5}{4}|r + 2e \cdot \xi T| - R_1\right)^+; W \ne -\gamma z\right]$$

$$= \int_{-\infty}^{\infty}\left(\frac{5}{4}|r + 2uT| - R_1\right)^+ (\varphi_{0,1}(u) - \varphi_{0,1}(u + \gamma r))^+\, du$$

$$= \int_{-\gamma r/2}^{\infty}\left(\frac{5}{4}|r + 2uT| - R_1\right)^+ (\varphi_{0,1}(u) - \varphi_{0,1}(u + \gamma r))\, du$$

$$= \int_{\gamma r/2}^{\infty}\left\{\left(\frac{5}{4}|r + 2uT| - R_1\right)^+ - \left(\frac{5}{4}|r + 2(u - \gamma r)T| - R_1\right)^+\right\}\varphi_{0,1}(u)\, du$$

$$\le \frac{5}{2}\gamma r T \int_{\gamma r/2}^{\infty} \varphi_{0,1}(u)\, du \le \frac{5}{4}\gamma r T.$$

Here we have used that by (21),

$$P[W \ne -\gamma z | \xi] = (\varphi_{0,1}(e \cdot \xi) - \varphi_{0,1}(e \cdot \xi + \gamma r))^+ / \varphi_{0,1}(e \cdot \xi).$$

Moreover, we have used that by (116), $R_1 \ge \frac{5}{4}(1 + \gamma T)r$. By concavity of f and by (117), we obtain

$$(124) \qquad\qquad \begin{aligned} III &\le f'(R_1) E[(R' - R_1)^+; A(x) \cap \hat{A}(y) \cap \{W \ne -\gamma z\}] \\ &\le \frac{5}{4}\gamma T r f'(R_1) \le \frac{5}{4}e^{-a(R_1 - 2\mathcal{R})}\gamma T r f'(r) \le \frac{1}{16}\gamma T r f'(r). \end{aligned}$$

IV. By a similar argument as in the proof of Theorem 2.2, we obtain

$$E[R' - r; A(x) \triangle \hat{A}(y)]$$

$$\leq E[|q_T(x, \xi) - x|; A(x) \cap \hat{A}(y)^C] + E[|q_T(y, \eta) - y|; A(x)^C \cap \hat{A}(y)]$$

$$\leq E[|x| + 2T|\xi|; A(x) \cap \hat{A}(y)^C] + E[|y| + 2T|\eta|; A(x)^C \cap \hat{A}(y)]$$

$$\leq (1 + 2T)E[\max(|x|, |y|, |\xi|, |\eta|); A(x) \triangle \hat{A}(y)]$$

$$\leq 2C_3 \max(1, \gamma)T(1 + T)^2(R_2^4 + d^2)h^2 r.$$

Here we have used (72) in the last step. By concavity of f we obtain

(125)
$$\begin{aligned} \text{IV} &\leq f'(r)E[R' - r; A(x) \triangle \hat{A}(y)] \\ &\leq 2C_3 \max(1, \gamma)T(1 + T)^2(R_2^4 + d^2)h^2 r f'(r) \\ &\leq \frac{1}{80}\gamma T r f'(r) \end{aligned}$$

for $h \leq h_6$ where h_6 is a positive constant depending only on L, M, N, $\mathcal{R}$, R_2 and d that by (27) can be chosen such that h_6^{-2} is of order $O((1 + \mathcal{R})(R_2^4 + d^2))$.

Combining the bounds for the terms I, II, III and IV in (122), (123), (124) and (125), we obtain for $h \leq \min(h_5, h_6)$,

(126)
$$\begin{aligned} E[f(R') - f(r)] &\leq \left(-\frac{27}{40} + \frac{2}{5} + \frac{1}{16} + \frac{1}{80}\right)\gamma T r f'(r) \\ &= -\frac{1}{5}\gamma T r f'(r) \leq -c_2 f(r), \end{aligned}$$

where the contraction rate c_2 satisfies

(127)
$$\begin{aligned} c_2 &= \frac{1}{5}\gamma T \inf_{r \leq 2\mathcal{R}} \frac{r f'(r)}{f(r)} \geq \frac{1}{5}\gamma T \max(1, 2a\mathcal{R})e^{-2a\mathcal{R}} \\ &= \frac{1}{5}\min\left(1, \frac{T}{4\mathcal{R}}\right)\max\left(1, \frac{2\mathcal{R}}{T}\right)e^{-2\mathcal{R}/T} \geq \frac{1}{10}e^{-2\mathcal{R}/T}. \end{aligned}$$

(iii) *Global contraction.* Let $h_\star := \min(h_0, h_5, h_6)$. Then by combining the bounds in (119) and (127), we see that for $h \leq \min(h_1, h_\star)$ and for any $x, y \in \mathbb{R}^d$ with $\max(|x|, |y|) \leq R_2$,

$$E[f(R')] \leq (1 - c)f(r),$$

where $c := \min(c_1, c_2)$. Moreover, by (120) and (128),

$$c \geq \frac{1}{10}e^{-2\mathcal{R}/T}\min\left(1, \frac{1}{2}KT^2(1 + \mathcal{R}/T)e^{-\mathcal{R}/(2T)}\right).$$

This completes the proof for adjusted numerical HMC.

For unadjusted numerical HMC, the proof simplifies because the rejection events $A(x)^C$ and $\hat{A}(y)^C$ are empty. Thus it suffices to control I, II and III, and this can be done similarly as above whenever $h \leq \min(h_1, h_\star)$, where now we can choose $h_\star := h_0$. $\square$

PROOF OF THEOREM 2.3. The contraction bound for exact HMC can be derived similarly to the proof of Theorem 2.4. In this case, instead of Theorem 2.2, we apply Theorem 2.1 in Step (i). Furthermore, the rejection events $A(x)^C$ and $\hat{A}(y)^C$, $x, y \in \mathbb{R}^d$, are empty for exact HMC. Therefore, the corresponding terms do not have to be taken into account in Step (ii). Consequently, the resulting bound (31) is valid for all $x, y \in \mathbb{R}^d$ with the same rate c as above. $\square$

Next, we prove the contraction result under Assumption 2.2. Again, we directly give the proof for adjusted numerical HMC, and we comment on the simplifications in the other cases afterwards.

PROOF OF THEOREM 2.7.   We fix $x, y \in \mathbb{R}^d$ such that $\max(|x|, |y|) \leq R_2$ and we set

$$F(x, y) = f\big(\min(|x - y|, 2\mathcal{R})\big), \qquad F'(x, y) = F\big(X'(x, y), Y'(x, y)\big),$$

$$G(x, y) = 1 + \epsilon \Psi(x) + \epsilon \Psi(y), \qquad G'(x, y) = G\big(X'(x, y), Y'(x, y)\big).$$

Since x and y are fixed, we omit the dependences on x and y in the notation. Thus $\rho_\epsilon = \sqrt{FG}$ and $\rho_\epsilon(X', Y') = \sqrt{F'G'}$. We consider separately the cases where $|x - y| < 2\mathcal{R}$ and $|x - y| \geq 2\mathcal{R}$.

(i) *Contractivity for* $|x - y| \leq 2\mathcal{R}$. In this case, the same arguments as in the proof of Theorem 2.4 show that for $h \leq \min(h_5, h_6)$,

$$E[F'] \leq E[f(R')] \leq (1 - c_2) f(r) = (1 - c_2) F$$

for a constant $c_2 \geq \frac{1}{10} e^{-2\mathcal{R}/T}$. Furthermore, (A4) implies

$$E[G'] = 1 + \epsilon E[\Psi(X') + \Psi(Y')] = 1 + \epsilon (\pi \Psi)(x) + \epsilon (\pi \Psi)(y)$$

$$\leq 1 + \epsilon \Psi(x) + \epsilon \Psi(y) + 2\epsilon A \leq (1 + 2\epsilon A) G.$$

Thus by the choice of ϵ in (35),

$$(128) \qquad \begin{aligned} E[\rho_\epsilon(X', Y')] = E[\sqrt{F'G'}] &\leq E[F']^{1/2} E[G']^{1/2} \\ &\leq (1 - c_2)^{1/2} (1 + 2\epsilon A)^{1/2} \sqrt{FG} \leq (1 - c_2/2 + \epsilon A) \rho_\epsilon \\ &\leq (1 - c_2/4) \rho_\epsilon. \end{aligned}$$

(ii) *Contractivity for* $|x - y| > 2\mathcal{R}$. If $|x - y| > 2\mathcal{R}$ then $|x| > \mathcal{R}$ or $|y| > \mathcal{R}$. Hence by (13), $\Psi(x) + \Psi(y) > 4A/\lambda$, and thus by (A4),

$$\begin{aligned} E[G'] &= 1 + \epsilon (\pi \Psi)(x) + \epsilon (\pi \Psi)(y) \\ &\leq 1 + \epsilon \big((1 - \lambda)(\Psi(x) + \Psi(y)) + 2A\big) \\ &\leq 1 + \epsilon \big((1 - \lambda/4)(\Psi(x) + \Psi(y)) - A\big) \\ &\leq \max(1 - \lambda/4, 1 - \epsilon A) G = (1 - c_3) G, \end{aligned}$$

where $c_3 := \min(\lambda/4, \epsilon A) = \min(c_2, \lambda)/4$. Since $F' \leq f(2\mathcal{R}) \leq f(|x - y|) = F$, we thus obtain

$$(129) \qquad \begin{aligned} E[\rho_\epsilon(X', Y')] = E[\sqrt{F'G'}] &\leq \sqrt{F} E[G']^{1/2} \leq (1 - c_3)^{1/2} \sqrt{FG} \\ &\leq (1 - c_3/2) \rho_\epsilon. \end{aligned}$$

(iii) *Global contraction.* Let $h_\star := \min(h_5, h_6)$. Then by combining the bounds in (128) and (129), we see that for $h \leq \min(h_1, h_\star)$ and for any $x, y \in \mathbb{R}^d$ with $\max(|x|, |y|) \leq R_2$,

$$E[(\rho_\epsilon(X', Y')] \leq (1 - c) \rho_\epsilon(x, y),$$

where $c := \min(c_2/4, c_3/2) = \min(c_2/8, \lambda/4) \geq \min(e^{-2\mathcal{R}/T}, 20\lambda)/80$. $\square$

PROOF OF THEOREM 2.6.   The proof is similar to Theorem 2.7. Since the rejection events are empty, the bound holds for all $h \in [0, h_1]$. $\square$

6. Proofs of results in Section 2.6. All bounds in Section 2.6 are based on the following observation:

LEMMA 6.1 (Locally contractive couplings and supermartingales). *Let $\pi(x, dy)$ be a Markov transition kernel on a complete separable metric space (S, ρ). Suppose that there exist a constant $c \in (0, \infty)$, a measurable subset $A \subseteq S$, a probability space $(\Omega, \mathcal{A}, P)$, and a measurable map*

$$(x, y, \omega) \mapsto \big(X'(x, y)(\omega), Y'(x, y)(\omega)\big)$$

from $S \times S \times \Omega$ to $S \times S$ such that for any $x, y \in S$, $(X'(x, y), Y'(x, y))$ is a realization of a coupling of $\pi(x, \cdot)$ and $\pi(y, \cdot)$, and

$$(130) \qquad E\big[\rho\big(X'(x, y), Y'(x, y)\big)\big] \leq e^{-c}\rho(x, y) \quad \text{for } x, y \in A.$$

Then, for any probability measure γ on $S \times S$, there is a Markov chain $(X_n, Y_n)_{n \geq 0}$ defined on a probability space $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{P})$ such that $(X_0, Y_0) \sim \gamma$, both marginal processes $(X_n)_{n \geq 0}$ and $(Y_n)_{n \geq 0}$ are Markov chains on S with transition kernel π, and such that the process

$$(131) \qquad M_n = e^{c(n \wedge T)}\rho(X_{n \wedge T}, Y_{n \wedge T}), \qquad T = \min\{n \geq 0 : (X_n, Y_n) \notin A \times A\},$$

is a nonnegative supermartingale w.r.t. the filtration generated by $(X_n, Y_n)_{n \geq 0}$.

PROOF. For $x, y \in S \times S$ let

$$k\big((x, y), \cdot\big) := P \circ \big(X'(x, y), Y'(x, y)\big)^{-1}$$

denote the joint law of $X'(x, y)$ and $Y'(x, y)$. Then k is a transition kernel on $S \times S$ with marginals $\pi(x, \cdot)$ and $\pi(y, \cdot)$, and by (130),

$$(132) \qquad (k\rho)(x, y) \leq e^{-c}\rho(x, y) \quad \text{for any } x, y \in A.$$

Now let $(X_n, Y_n)_{n \geq 0}$ be a time-homogeneous Markov chain on a probability space $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{P})$ with initial distribution $(X_0, Y_0) \sim \gamma$ and transition kernel k, and let $\mathcal{F}_n = \sigma((X_i, Y_i) : 0 \leq i \leq n)$. Then for any $n \geq 0$,

$$\widetilde{E}\big[\rho(X_{n+1}, Y_{n+1}) | \mathcal{F}_n\big] = (k\rho)(X_n, Y_n) \leq e^{-c}\rho(X_n, Y_n)$$

holds $\widetilde{P}$-almost surely on $\{(X_n, Y_n) \in A \times A\}$. Therefore, the process (M_n) defined by (131) is a nonnegative $(\mathcal{F}_n)$-supermartingale. $\square$

The error bound for exact HMC in Corollary 2.8 is a direct consequence of Theorem 2.3 and Lemma 6.1:

PROOF OF COROLLARY 2.8. For exact HMC, by Theorem 2.3, the local contractivity condition (130) in Lemma 6.1 is satisfied for $S = A = \mathbb{R}^d$, ρ and c given by (25) and (32), and the coupling $(X'(x, y), Y'(x, y))$ introduced above. Now let ν and η be probability measures on $\mathbb{R}^d$, and let γ be an arbitrary coupling of ν and η. Then by Lemma 6.1, there is a Markov chain $(X_n, Y_n)_{n \geq 0}$ on a probability space $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{P})$ such that $(X_0, Y_0) \sim \gamma$, both (X_n) and (Y_n) are Markov chains with transition kernel π and initial laws ν and η, respectively, and $M_n = e^{cn}\rho(X_n, Y_n)$ is a nonnegative supermartingale. Hence for any $n \in \mathbb{N}$,

$$\mathcal{W}_\rho\big(\nu\pi^n, \mu\pi^n\big) \leq \widetilde{E}\big[\rho(X_n, Y_n)\big] \leq e^{-cn}E\big[\rho(X_0, Y_0)\big] = e^{-cn}\int \rho \, d\gamma.$$

Taking the infimum over all couplings $\gamma \in \Pi(\nu, \eta)$, we see that (39) holds. Furthermore, by (25) and (26),

$$(133) \qquad e^{-aR_1}|x - y| \leq \rho(x, y) \leq |x - y| \quad \text{for any } x, y \in \mathbb{R}^d.$$

Therefore, (39) implies

$$\mathcal{W}^1(\nu\pi^n, \eta\pi^n) \le e^{aR_1}e^{-cn}\mathcal{W}^1(\nu, \eta).$$

Choosing $\eta = \mu$, we have $\eta\pi^n = \mu$ for all n. Hence

$$\Delta(n) \le \exp(aR_1 - cn)\Delta(0).$$

The second part of the assertion now follows, because by (29) and (30), $aR_1 = \frac{5}{2}(1 + \mathcal{R}/T)$.
$\square$

Now suppose again that $\pi(x, dy)$ is an arbitrary Markov transition kernel on a complete separable metric space (S, ρ). For proving Theorem 2.11, we combine Lemma 6.1 with a Lyapunov bound for exit probabilities:

PROOF OF THEOREM 2.11. Let ν and η be probability measures on S, and let γ be a coupling of ν and η. By (C3) in Assumption 2.3, the conditions in Lemma 6.1 are satisfied with $A = \{\psi > C\}$. Hence on some probability space $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{P})$, there is a coupling $(X_n, Y_n)_{n \ge 0}$ of Markov chains with transition kernel π and joint initial law $(X_0, Y_0) \sim \gamma$ such that $M_n = e^{c(n \wedge T)}\rho(X_{n \wedge T}, Y_{n \wedge T})$ is a nonnegative supermartingale stopped at

$$T = \min\{n \ge 0 : \psi(X_n) > C \text{ or } \psi(Y_n) > C\}.$$

In particular, we obtain

$$(134) \quad e^{cn}E[\rho(X_n, Y_n); n \le T] \le E[e^{c(n \wedge T)}\rho(X_{n \wedge T}, Y_{n \wedge T})$$
$$\le E[\rho(X_0, Y_0)] = \int \rho \, d\gamma$$

for any $n \in \mathbb{N}$. In order to bound the corresponding expectation on the complement $\{n < T\}$, we observe that by Condition (C2) in Assumption 2.3 and by the definition of $\delta(C)$ in (43),

$$(135) \quad E[\rho(X_n, Y_n); n > T] \le E[\varphi(X_n) + \varphi(Y_n); n > T]$$
$$\le \beta^n E[\varphi(X_T) + \varphi(Y_T); n > T]$$
$$\le \beta^n E[\psi(X_T) + \psi(Y_T); n > T]\delta(C).$$

Here we have used that by (C2), for any $k \le n$;

$$E[\varphi(X_n); T = k] = E[(\pi^{n-k}\varphi)(X_k); T = k] \le \beta^{n-k}E[\varphi(X_k); T = k]$$
$$\le \beta^n E[\varphi(X_T); T = k],$$

and a corresponding inequality holds for $E[\varphi(Y_n); T = k]$.

Furthermore, the Lyapunov condition (C1) in Assumption 2.3 implies that the stopped process $N_n = \psi(X_{n \wedge T})/\lambda^{n \wedge T}$ is a nonnegative supermartingale. Therefore,

$$E[\psi(X_T); n > T] \le \lambda^{n-1}E[\psi(X_T)/\lambda^T] \le \lambda^{n-1}E[\psi(X_0)] = \lambda^{n-1}\int \psi \, d\nu.$$

A corresponding bound holds for $E[\psi(Y_T); n > T]$, and thus

$$(136) \quad E[\psi(X_T) + \psi(Y_T); n > T] \le \lambda^{n-1}\left(\int \psi \, d\nu + \int \psi \, d\eta\right).$$

By combining the bounds in (134), (135) and (136), we obtain

$$E[\rho(X_n, Y_n)] \le e^{-cn}\int \rho \, d\gamma + \beta^n\lambda^{n-1}\left(\int \psi \, d\nu + \int \psi \, d\eta\right)\delta(C)$$

for any $n \in \mathbb{N}$. The bound for the Kantorovich distance in (42) now follows by taking the infimum over all couplings $\gamma \in \Pi(\nu, \eta)$. $\square$

From now on, we consider numerical HMC. Let π denote the transition kernel for a given step size $h > 0$, and let ρ be the metric defined by (25), (26), (29) and (30). To be able to apply Theorem 2.11, we first identify appropriate Lyapunov functions.

LEMMA 6.2. *Let $T, h_1 \in (0, \infty)$ such that* (27) *holds. Then there exists $C_1 \in (0, \infty)$ depending only on L, M and N such that for $C \in (1, \infty)$ and $h \in (0, h_1)$ with*

$$(137) \qquad C_1 T h^2 \leq \min\left((\mathcal{R} + \sqrt{2/K} (\log C)^{3/4})^{-3}, (1/3)^{3/2}\right).$$

Conditions (C1) *and* (C2) *in Assumption* 2.3 *are satisfied with*

$$(138) \qquad \varphi(x) = |x| + 2T d^{1/2}, \qquad \psi(x) = \exp(U(x)^{2/3}),$$

$$(139) \qquad \beta = 2, \qquad \lambda = E\left[\exp\left(|\xi|^{4/3} + \frac{1}{3}|\xi|^2 + 1\right)\right] \quad \text{with } \xi \sim N(0, I_d).$$

PROOF OF THEOREM 2.12. Let $C \in [e, \infty)$, that is, $\log C \geq 1$. Then by Lemma 6.2, Conditions (C1) and (C2) in Assumption 2.3 are satisfied for φ, ψ, β and λ given by (138) and (139), provided (137) holds. This is the case for $h \leq h_0$ where $h_0 > 0$ can be chosen such that h_0^{-1} is of order $O(\mathcal{R}^{3/2} + (\log C)^{9/8})$ for fixed values of K and L. Furthermore, by (152), $\psi(x) \leq C$ implies $|x| \leq R_2$, where we set

$$(140) \qquad R_2 = \mathcal{R} + \sqrt{2/K} (\log C)^{3/4}.$$

Therefore, by Theorem 2.4, the local contractivity condition (C3) is satisfied with c given by (32) provided $h \leq \min(h_\star, h_1)$ where $h_\star$ can be chosen such that $h_\star^{-1}$ is of order $O((1 + T^{-1/2} + \mathcal{R}^{1/2})(d + \mathcal{R}^2 + (\log C)^{3/2}))$. Hence for $h \leq h_{\star\star} = \min(h_\star, h_0, h_1)$, all parts of Assumption 2.3 are satisfied, and thus we can apply Theorem 2.11. By (42), and since $\mu \pi^n = \mu$, we obtain

$$\mathcal{W}_\rho(\nu \pi^n, \mu) \leq e^{-cn} \mathcal{W}_\rho(\nu, \mu) + \beta^n \lambda^{n-1}\left(\int \psi \, d\nu + \int \psi \, d\mu\right) \delta(C),$$

where $\delta(C)$ is given by (43). By (133), (29) and (30), this implies

$$(141) \qquad \Delta(n) = \mathcal{W}^1(\nu \pi^n, \mu) \leq \mathrm{I} + \mathrm{II}, \quad \text{where}$$

$$\mathrm{I} = \exp(a R_1 - cn) \Delta(0) = \exp\left(\frac{5}{2}(1 + \mathcal{R}/T) - cn\right) \Delta(0) \quad \text{and}$$

$$\mathrm{II} = \exp\left(\frac{5}{2}(1 + \mathcal{R}/T)\right) \beta^n \lambda^{n-1}\left(\int \psi \, d\nu + \int \psi \, d\mu\right) \delta(C).$$

Choosing n as in (44), we obtain $\mathrm{I} \leq \epsilon/2$. Furthermore, we can ensure $\mathrm{II} \leq \epsilon/2$ and thus $\Delta(n) \leq \epsilon$ by choosing C sufficiently large. Indeed, by (43),

$$(142) \qquad \delta(C) \leq 2 \sup\left\{\frac{\max(\varphi(x), \varphi(y))}{\max(\psi(x), \psi(y))} : x, y \in S \text{ s.t. } \max(\psi(x), \psi(y)) > C\right\}.$$

Moreover, by (152) and (138), for any $x \in S$,

$$\varphi(x) = |x| + 2T d^{1/2} \leq \mathcal{R} + 2T d^{1/2} + \sqrt{2/K} (\log \psi(x))^{3/4}.$$

Let $x, y \in S$ such that $\max(\psi(x), \psi(y)) > C$. Without loss of generality, we assume $\max(\psi(x), \psi(y)) = \psi(x)$. Then $\log \psi(x) > \log C \geq 1$, and hence

$$\frac{\max(\varphi(x), \varphi(y))}{\max(\psi(x), \psi(y))} \leq \frac{\mathcal{R} + 2T\sqrt{d} + \sqrt{2/K}\log\psi(x)}{\psi(x)}$$

$$\leq \frac{\mathcal{R} + 2T\sqrt{d} + \sqrt{2/K}\log C}{C}.$$

Here we have used that $t \mapsto t^{-1}\log t$ is decreasing for $\log t \geq 1$. By (142), we see that

(143) $$\delta(C) \leq 2(\mathcal{R} + 2T\sqrt{d} + \sqrt{2/K}\log C)/C.$$

Consequently, we have $\mathrm{II} \leq \epsilon/2$ if

(144) $$C/(u + v\log C) \geq w(\beta\lambda)^n,$$

where $u := \mathcal{R} + 2T\sqrt{d}$, $v := \sqrt{2/K}$, and

$$w := 4\epsilon^{-1}\exp\left(\frac{5}{2}(1 + \mathcal{R}/T)\right) \cdot \left(\int \psi \, dv + \int \psi \, d\mu\right).$$

Condition (144) holds if and only if

(145) $$\log C \geq \log(u + v\log C) + \log w + n\log(\beta\lambda).$$

In particular, since

$$\log(u + v\log C) \leq \log^+(2u) + \log^+(2v\log C) \leq \log^+ u + \log^+ v + 2 + \log\log C,$$

there is a universal finite constant C_0 such that (144) is satisfied if $C \geq C_0$ and

(146) $$\log C \geq \log^+ u + \log^+ v + \log w + n\log(\beta\lambda).$$

We have $\log u = \log(\mathcal{R} + 2T\sqrt{d})$, $\log v = \frac{1}{2}\log(2/K)$, and

$$\log w = \frac{5}{2}(1 + \mathcal{R}/T)\log\left(4\frac{\int \psi \, d\mu + \int \psi \, dv}{\epsilon}\right).$$

Furthermore, by (139), $\log(\beta\lambda)$ is of order $O(d)$, and n satisfies (44). Combining these observations, we see that we can ensure $\mathrm{II} \leq \epsilon/2$ and thus $\Delta(n) \leq \epsilon$ by choosing $\log C$ proportional to $dn + (1 + \mathcal{R}/T)\log^+(\frac{\int \psi \, d\mu + \int \psi \, dv}{\epsilon})$. The assertion follows since $\log^+ \int \psi \, d\mu = O(d)$ and

$$h_{\star\star}^{-1} = O((1 + T^{-1/2} + \mathcal{R}^{1/2})(d + \mathcal{R}^2 + (\log C)^{3/2})). \qquad \square$$

APPENDIX: EXPLICIT LYAPUNOV FUNCTIONS

In this appendix, we prove the results on Lyapunov functions for HMC stated in Examples 2.1 and 2.2 and Lemma 6.2.

PROOF FOR EXAMPLE 2.1. Let $x_t = q_t(x, \xi)$ and $v_t = p_t(x, \xi)$ where $x \in \mathbb{R}^d$ and $\xi \sim N(0, I_d)$. A simple computation shows that if (14) holds then for $t \leq T$,

$$\frac{d}{dt}|x_t|^2 \leq 2x_t \cdot v_t + 2hLx_T^{\star,2}, \quad \text{and}$$

$$\frac{d}{dt}(x_t \cdot v_t) \leq |v_t|^2 - \kappa|x_t|^2 + C + \frac{3}{2}hLx_T^{\star}v_T^{\star} + \frac{1}{4}h^2L^2x_T^{\star,2}.$$

Therefore,

$$(147) \quad \begin{aligned} |x_T|^2 &\leq |x|^2 + 2Tx \cdot \xi + 2\int_0^T \int_0^t |v_s|^2 \, ds \, dt - 2\kappa \int_0^T \int_0^t |x_s|^2 \, ds \, dt \\ &\quad + CT^2 + 2hLTx_T^{\star,2} + \frac{3}{2}hLT^2 x_T^\star v_T^\star + \frac{1}{4}h^2L^2T^2 x_T^{\star,2}. \end{aligned}$$

Now for $s \leq T$, we can apply the a priori bounds

$$(148) \quad |v_s| \leq |\xi| + 2Ls \max(|x|, |x + s\xi|),$$

$$(149) \quad |x_s - (x + s\xi)| \leq \max(|x|, |x + s\xi|)/10,$$

which follow from Lemma 3.1. With a short computation these imply

$$2\int_0^T \int_0^t |v_s|^2 \, ds \, dt \leq \left(2T^2 + \frac{16}{15}L^2T^6\right)|\xi|^2 + \frac{8}{3}L^2T^4|x|^2,$$

$$2\int_0^T \int_0^t |x_s|^2 \, ds \, dt \geq \frac{79}{200}T^2|x|^2 + \frac{27}{100}T^3 x \cdot \xi + \frac{27}{400}T^4|\xi|^2.$$

By bounding the corresponding terms in (147) and noting that $x_T^\star \leq \frac{11}{10}(|x| + T|\xi|)$, $v_T^\star \leq \frac{6}{5}|\xi| + 2LT|x|$ and $L(T^2 + hT) \leq \kappa/(10L)$, we conclude that

$$(150) \quad |x_T|^2 \leq \left(1 - \frac{76}{600}\kappa T^2\right)|x|^2 + \left(2T - \frac{27}{100}\kappa T^3\right)x \cdot \xi + (C + 2|\xi|^2)T^2$$

provided that $T/h \geq n_0$ for an appropriate constant $n_0 \in \mathbb{N}$ depending on κ and L. Taking expectations in (150), we conclude that (15) holds for exact and unadjusted HMC with $\Psi(x) = |x|^2$. For adjusted numerical HMC, we obtain

$$\begin{aligned} (\pi\Psi)(x) &= E[|x_T|^2; A(x)] + |x|^2 P[A(x)^C] \\ &\leq \left(1 - \frac{76}{600}\kappa T^2 + P[A(x)^C]\right)\Psi(x) + (C + 2d)T^2. \end{aligned}$$

Hence by the a priori bound on the rejection probability in Theorem 3.8, we see that (15) holds for $|x| \leq R_2$ if $R_2^3 h^2$ is sufficiently small. $\square$

PROOF FOR EXAMPLE 2.2.   We only sketch the proof for exact HMC. Let $\delta > 0$ and choose $\Psi(x) = g(|x|^2)$ where $g : \mathbb{R}_+ \to \mathbb{R}_+$ is a smooth increasing convex function such that $g(0) = 0$, $g(s) = \exp(\delta\sqrt{s})$ for $s \geq \delta^{-2}$, and $g''(s) \leq 2\delta^4$ for $s \leq \delta^{-2}$, and let x_t and v_t be defined as in the proof above. Assumption (16) implies that for all t,

$$(151) \quad |v_t - \xi| \leq Qt \quad \text{and} \quad |x_t - (x + \xi t)| \leq Qt^2/2.$$

Noting that $\Psi(x) = \exp(\delta|x|)$ if $|x| \geq \delta^{-1}$, an explicit computation shows that if $|x_t| > \delta^{-1}$ then by (16) and (151),

$$\begin{aligned} \frac{d^2}{dt^2}\Psi(x_t) &= \left[\delta|x_t|^{-1}\left(\left|v_t - \frac{x_t \cdot v_t}{|x_t|^2}x_t\right|^2 - x_t \cdot \nabla U(x_t)\right) + \delta^2\frac{(x_t \cdot v_t)^2}{|x_t|^2}\right]\Psi(x_t) \\ &\leq \left[-\kappa + \delta C + 2\delta(|\xi|^2 + Q^2t^2)\right]\delta\Psi(x_t) \\ &\leq \left[-\kappa e^{-\delta t|\xi| - \delta Qt^2/2} + (C + 2|\xi|^2 + Q^2t^2)\delta e^{\delta t|\xi| + \delta Qt^2/2}\right]\delta\Psi(x). \end{aligned}$$

Similarly, one verifies that if $|x_t| \leq \delta^{-1}$ then

$$\frac{d^2}{dt^2}\Psi(x_t) = \frac{d^2}{dt^2}g(|x_t|^2) \leq (16|\xi|^2 + 16Q^2t^2 + eC)\delta^2.$$

We now choose $\delta := \kappa/(4A + 8d + Q^2 T^2)$. Since by (16), $\kappa \leq Q$ and $\xi \sim N(0, I_d)$, this choice implies in particular that $E[\exp(\delta T|\xi| + \delta Q T^2/2)] \leq \sqrt{2}$. Noting that

$$\Psi(x_T) = \Psi(x) + T g'(|x|^2)x \cdot \xi + \int_0^T \int_0^t \frac{d^2}{ds^2}\Psi(x_s)\,ds\,dt,$$

and combining the bounds above, we then obtain

$$(\pi\Psi)(x) = \Psi(x) + \int_0^T \int_0^t E\left[\frac{d^2}{ds^2}\Psi(x_s)\right]ds\,dt \leq \delta\kappa T^2\left(5 - \frac{1}{4\sqrt{2}}\Psi(x)\right). \qquad \square$$

PROOF OF LEMMA 6.2.   We first remark that by Assumption 2.1, $x \cdot \nabla U(x) \geq K|x|^2$ for $|x| \geq \mathcal{R}$. Therefore, for any $x \in \mathbb{R}^d$,

(152) $$U(x) \geq \frac{K}{2}\min(|x| - \mathcal{R}, 0)^2 \quad \text{and} \quad |x| \leq \mathcal{R} + \sqrt{2U(x)/K}.$$

Furthermore, by (25) and (26),

$$\rho(x, y) \leq |x - y| \leq |x| + |y| \leq \varphi(x) + \varphi(y) \quad \text{for any } x, y \in \mathbb{R}^d.$$

To verify the Lyapunov conditions recall that

$$(\pi\varphi)(x) = E[\varphi(q_T(x, \xi)); A(x)] + \varphi(x)P[A(x)^C] \quad \text{with } \xi \sim N(0, I_d).$$

By Lemma 3.1, $|q_T(x, \xi)| \leq 2(|x| + T|\xi|)$, and thus for any $x \in \mathbb{R}^d$,

$$(\pi\varphi)(x) \leq 2E[|x| + T|\xi| + 2Td^{1/2}] \leq 2|x| + 4Td^{1/2} = 2\varphi(x).$$

Hence (C2) is satisfied.

Furthermore, by Lemma 3.6, there is a finite constant C_1 such that

(153) $$U(q_T(x, \xi)) \leq H(\phi_T(x, \xi)) \leq H(x, \xi) + C_1 T h^2 \max(|x|, |\xi|)^3$$

$$\leq U(x) + \frac{1}{2}|\xi|^2 + C_1 T h^2 |\xi|^3 + C_1 T h^2 |x|^3.$$

Suppose that $\psi(x) \leq C$. Then $U(x) \leq (\log C)^{3/2}$, and hence by (152), $|x| \leq \mathcal{R} + \sqrt{2/K} \times (\log C)^{3/4}$. Therefore, if (137) holds then by (153), we obtain

$$(\pi\psi)(x) \leq E[\exp(U(x)^{2/3} + |\xi|^{4/3} + |\xi|^2/3 + 1)] = \lambda\psi(x)$$

for any $x \in S$ such that $\psi(x) \leq C$. Hence (C1) is satisfied as well. $\square$

REFERENCES

[1] ANDERSEN, H. C. (1980). Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.* **72** 2384.

[2] ANDRIEU, C., DE FREITAS, N., DOUCET, A. and JORDAN, M. I. (2003). An introduction to MCMC for machine learning. *Mach. Learn.* **50** 5–43.

[3] BETANCOURT, M., BYRNE, S., LIVINGSTONE, S. and GIROLAMI, M. (2017). The geometric foundations of Hamiltonian Monte Carlo. *Bernoulli* **23** 2257–2298. MR3648031 https://doi.org/10.3150/16-BEJ810

[4] BISHOP, C. M. (2006). *Pattern Recognition and Machine Learning. Information Science and Statistics.* Springer, New York. MR2247587 https://doi.org/10.1007/978-0-387-45528-0

[5] BOU-RABEE, N. and SANZ-SERNA, J. M. (2017). Randomized Hamiltonian Monte Carlo. *Ann. Appl. Probab.* **27** 2159–2194. MR3693523 https://doi.org/10.1214/16-AAP1255

[6] BOU-RABEE, N. and SANZ-SERNA, J. M. (2018). Geometric integrators and the Hamiltonian Monte Carlo method. *Acta Numer.* **27** 113–206. MR3826507 https://doi.org/10.1017/s0962492917000101

[7] BUTKOVSKY, O. (2014). Subgeometric rates of convergence of Markov processes in the Wasserstein metric. *Ann. Appl. Probab.* **24** 526–552. MR3178490 https://doi.org/10.1214/13-AAP922

[8] CHENG, X., CHATTERJI, N. S., BARTLETT, P. L. and JORDAN, M. I. (2018). Underdamped Langevin MCMC: A non-asymptotic analysis. In *Conference on Learning Theory.*

[9] CRANSTON, M. (1991). Gradient estimates on manifolds using coupling. *J. Funct. Anal.* **99** 110–124. MR1120916 https://doi.org/10.1016/0022-1236(91)90054-9

[10] DASHTI, M. and STUART, A. M. (2017). The Bayesian approach to inverse problems. In *Handbook of Uncertainty Quantification. Vol.* 1, 2, 3 311–428. Springer, Cham. MR3839555

[11] DIACONIS, P. (2009). The Markov chain Monte Carlo revolution. *Bull. Amer. Math. Soc.* (*N.S.*) **46** 179–205. MR2476411 https://doi.org/10.1090/S0273-0979-08-01238-X

[12] DIACONIS, P., HOLMES, S. and NEAL, R. M. (2000). Analysis of a nonreversible Markov chain sampler. *Ann. Appl. Probab.* **10** 726–752. MR1789978 https://doi.org/10.1214/aoap/1019487508

[13] DUANE, S., KENNEDY, A. D., PENDLETON, B. J. and ROWETH, D. (1987). Hybrid Monte Carlo. *Phys. Lett. B* **195** 216–222. MR3960671

[14] DURMUS, A. and EBERLE, A. Asymptotic bias for inexact Markov chain Monte Carlo methods in high dimensions. In preparation.

[15] DURMUS, A., MOULINES, E. and SAKSMAN, E. (2017). On the convergence of Hamiltonian Monte Carlo. Preprint. Available at arXiv:1705.00166.

[16] E, W. and LI, D. (2008). The Andersen thermostat in molecular dynamics. *Comm. Pure Appl. Math.* **61** 96–136. MR2361305 https://doi.org/10.1002/cpa.20198

[17] EBERLE, A. (2014). Error bounds for Metropolis–Hastings algorithms applied to perturbations of Gaussian measures in high dimensions. *Ann. Appl. Probab.* **24** 337–377. MR3161650 https://doi.org/10.1214/13-AAP926

[18] EBERLE, A. (2016). Reflection couplings and contraction rates for diffusions. *Probab. Theory Related Fields* **166** 851–886. MR3568041 https://doi.org/10.1007/s00440-015-0673-1

[19] EBERLE, A., GUILLIN, A. and ZIMMER, R. (2019). Quantitative Harris-type theorems for diffusions and McKean–Vlasov processes. *Trans. Amer. Math. Soc.* **371** 7135–7173. MR3939573 https://doi.org/10.1090/tran/7576

[20] EBERLE, A., GUILLIN, A. and ZIMMER, R. (2019). Couplings and quantitative contraction rates for Langevin dynamics. *Ann. Probab.* **47** 1982–2010. MR3980913 https://doi.org/10.1214/18-AOP1299

[21] GHAHRAMANI, Z. (2015). Probabilistic machine learning and artificial intelligence. *Nature* **521** 452.

[22] GUSTAFSON, P. (1998). A guided walk Metropolis algorithm. *Stat. Comput.* **8** 357–364.

[23] HAIRER, M., MATTINGLY, J. C. and SCHEUTZOW, M. (2011). Asymptotic coupling and a general form of Harris' theorem with applications to stochastic delay equations. *Probab. Theory Related Fields* **149** 223–259. MR2773030 https://doi.org/10.1007/s00440-009-0250-6

[24] HAIRER, M., STUART, A. M. and VOSS, J. (2007). Analysis of SPDEs arising in path sampling. II. The nonlinear case. *Ann. Appl. Probab.* **17** 1657–1706. MR2358638 https://doi.org/10.1214/07-AAP441

[25] HENG, J. and JACOB, P. E. (2019). Unbiased Hamiltonian Monte Carlo with couplings. *Biometrika* **106** 287–302. MR3949304 https://doi.org/10.1093/biomet/asy074

[26] HOFFMAN, M. D. and GELMAN, A. (2014). The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* **15** 1593–1623. MR3214779

[27] JOULIN, A. and OLLIVIER, Y. (2010). Curvature, concentration and error estimates for Markov chain Monte Carlo. *Ann. Probab.* **38** 2418–2442. MR2683634 https://doi.org/10.1214/10-AOP541

[28] KOU, S. C., ZHOU, Q. and WONG, W. H. (2006). Equi-energy sampler with applications in statistical inference and statistical mechanics. *Ann. Statist.* **34** 1581–1652. MR2283711 https://doi.org/10.1214/009053606000000515

[29] LEVIN, D. A. and PERES, Y. (2017). *Markov Chains and Mixing Times*. Amer. Math. Soc., Providence, RI. MR3726904

[30] LI, D. (2007). On the rate of convergence to equilibrium of the Andersen thermostat in molecular dynamics. *J. Stat. Phys.* **129** 265–287. MR2358805 https://doi.org/10.1007/s10955-007-9391-0

[31] LIANG, F. and WONG, W. H. (2001). Real-parameter evolutionary Monte Carlo with applications to Bayesian mixture models. *J. Amer. Statist. Assoc.* **96** 653–666. MR1946432 https://doi.org/10.1198/016214501753168325

[32] LIU, J. S. (2008). *Monte Carlo Strategies in Scientific Computing*. *Springer Series in Statistics*. Springer, New York. MR2401592

[33] MACKENZIE, P. B. (1989). An improved hybrid Monte Carlo method. *Phys. Lett. B* **226** 369–371.

[34] MANGOUBI, O. and SMITH, A. (2017). Rapid mixing of Hamiltonian Monte Carlo on strongly log-concave distributions. Preprint. Available at arXiv:1708.07114.

[35] NEAL, R. M. (2011). MCMC using Hamiltonian dynamics. In *Handbook of Markov Chain Monte Carlo* (S. Brooks, A. Gelman, G. Jones and X.-L. Meng, eds.) *Chapman & Hall/CRC Handb. Mod. Stat. Methods* 113–162. CRC Press, Boca Raton, FL. MR2858447

[36] SANZ-SERNA, J. M. (2014). Markov chain Monte Carlo and numerical differential equations. In *Current Challenges in Stability Issues for Numerical Differential Equations* (L. Dieci and N. Guglielmi, eds.) *Lecture Notes in Math.* **2082** 39–88. Springer, Cham. MR3204991 https://doi.org/10.1007/978-3-319-01300-8_2

[37] SEILER, C., RUBINSTEIN-SALZEDO, S. and HOLMES, S. (2014). Positive curvature and Hamiltonian Monte Carlo. In *Advances in Neural Information Processing Systems* 27 (Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence and K. Q. Weinberger, eds.) 586–594. Curran Associates, Red Hook.

[38] WANG, F.-Y. (2005). *Functional Inequalities, Markov Semigoups and Spectral Theory*. Science Press.

[39] WEBB, A. R. (2002). *Statistical Pattern Recognition*, 2nd ed. Wiley, Chichester. MR2191640 https://doi.org/10.1002/0470854774

[40] ZIMMER, R. (2017). Explicit contraction rates for a class of degenerate and infinite-dimensional diffusions. *Stoch. Partial Differ. Equ. Anal. Comput.* **5** 368–399. MR3689216 https://doi.org/10.1007/s40072-017-0091-8