Submitted to Bernoulli

Mixing Time Guarantees for Unadjusted Hamiltonian Monte Carlo

NAWAF BOU-RABEE[1] and ANDREAS EBERLE[2]

[1]*Department of Mathematical Sciences*
Rutgers University Camden
311 N 5th Street
Camden, NJ 08102 USA
E-mail: nawaf.bourabee@rutgers.edu

[2]*Institut für Angewandte Mathematik*
Universität Bonn
Endenicher Allee 60
53115 Bonn, Germany
E-mail: eberle@uni-bonn.de

AbstractWe provide quantitative upper bounds on the total variation mixing time of the Markov chain corresponding to the unadjusted Hamiltonian Monte Carlo (uHMC) algorithm. For two general classes of models and fixed time discretization step size h, the mixing time is shown to depend only logarithmically on the dimension. Moreover, we provide quantitative upper bounds on the total variation distance between the invariant measure of the uHMC chain and the true target measure. As a consequence, we show that an ε-accurate approximation of the target distribution μ in total variation distance can be achieved by uHMC for a broad class of models with $O\left(d^{3/4}\varepsilon^{-1/2}\log(d/\varepsilon)\right)$ gradient evaluations, and for mean field models with weak interactions with $O\left(d^{1/2}\varepsilon^{-1/2}\log(d/\varepsilon)\right)$ gradient evaluations. The proofs are based on the construction of successful couplings for uHMC that realize the upper bounds.

MSC 2010 subject classifications: Primary 60J05; secondary 65C05,65P10.
Keywords: MCMC, Hamiltonian Monte Carlo, Mixing Time, Couplings, Variational Integrators.

1. Introduction

A central problem in Markov chain Monte Carlo (MCMC) is to determine the number of MCMC steps that guarantees that the distribution of the chain is a good approximation of its invariant probability measure. This "mixing time" is commonly measured in terms of the total variation distance. Many tools have been developed for bounding mixing times, see [26] for an overview in the case of discrete state spaces. For Markov processes on continuous state spaces, these tools include geometric and analytic approaches based on conductance and isoperimetric inequalities [37, 42], spectral gaps and functional inequalities [41, 2], and hypocoercivity [43, 10, 1, 29], as well as probabilistic approaches that are mainly based on couplings, possibly in combination with drift/minorization conditions, see for example [27, 28, 34, 19, 20, 14, 15, 16]. A focus of current research has become to understand the mixing properties of Hamiltonian Monte Carlo (HMC) in representative models that exhibit both high-dimensionality and non-convexity.

HMC is an MCMC method that is based on deterministic Hamiltonian dynamics combined with momentum randomizations [11, 39]. With few exceptions, the Hamiltonian dynamics is not analytically solvable, and therefore, it is normally discretized in time using an evenly spaced grid with time step size $h > 0$ and a numerical integrator; most often velocity Verlet [18, 25, 3, 32]. Time discretization leads to an error in the invariant measure of the HMC method based on Verlet, which can in principle be reduced by taking a smaller time step size or removed by adding a Metropolis accept-reject step per HMC step. The latter gives rise to *adjusted* HMC (or Metropolis-adjusted HMC, or Metropolized HMC), and the method without adjustment is called *unadjusted* HMC (uHMC). The method based on the exact Hamiltonian dynamics is called *exact* HMC and is mainly used as a theoretical tool.

In recent years, there has been considerable progress in developing quantitative Wasserstein convergence bounds for HMC. For strongly logconcave distributions, a synchronous coupling was used to derive L^1-Wasserstein convergence bounds for exact HMC in [31], and this result was subsequently sharpened in [8]. These papers use perturbative arguments to study the effect of time discretization error. For more general target distributions, a coupling-based argument for unadjusted and adjusted HMC was developed in [5] to obtain contractivity in a carefully designed Wasserstein distance equivalent to the standard L^1-Wasserstein distance $\boldsymbol{W}^1$. By applying this coupling componentwise, this result has been extended to mean-field models [6]. Moreover, by using a two-scale coupling, the approach has also been extended to perturbations of Gaussian measures in infinite dimension [4].

On the other hand, explicit total variation (TV) bounds for HMC are scarce. The few existing results [31, 7] hold only under restrictive conditions on the target and/or the initial distribution (strong logconcavity, warm start), and except for very special cases, the optimal order of upper bounds is unknown. This paper makes a contribution towards filling this gap in the literature. It provides explicit upper bounds on (i) a one step $\boldsymbol{W}^1$ to TV regularization of the uHMC transition kernel, (ii) the mixing time of the uHMC chain started at an arbitrary initial distribution with finite first moment, and (iii) the TV bias between the stationary distribution of uHMC and the true target distribution.

Recall that by the dual descriptions of the distances, TV bounds correspond to bounds for integrals of arbitrary bounded measurable functions ("observables"), while $\boldsymbol{W}^1$ bounds correspond to bounds for Lipschitz continuous functions. Thus, one advantage of TV bounds is that they require less regularity of the observables and in particular apply to indicator functions. Moreover, in contrast to $\boldsymbol{W}^1$ distances, the TV distance is scale invariant. If the transition kernels of a Markov chain have strong smoothing properties, then it is not difficult to deduce TV bounds from $\boldsymbol{W}^1$ bounds. For HMC, however, such regularizing properties and their precise dependence on the dimension are not obvious.

Although the results are stated below in a slightly different way, the main idea underlying the mixing time bounds for uHMC derived in this work is the construction of a successful coupling of two copies of uHMC starting from different initial conditions. This coupling builds on recently introduced couplings for uHMC which bring the two copies arbitrarily close in representative models having both non-convexity and high dimension [5, 6]. After the two copies are close, a "one-shot coupling" is used to get them to coalesce with high probability [30]. This strategy works provided for small distances between the two copies, there is a large overlap between their corresponding distributions in the next step. A key element of our proofs is correspondingly a precise upper bound on this overlap, which is equivalent to a $\boldsymbol{W}^1$ to TV regularization bound for the uHMC transition kernel, see Lemmas 16 and 17 below. The regularizing effect stems from the initial velocity randomization in each transition step of the uHMC chain. We stress that it is not trivial to quantify the overlap precisely because each HMC step involves many deterministic moves. A similar overlap bound can also be applied in combination with a triangle inequality trick and existing bounds for the Wasserstein bias in order to quantify the total variation bias of the invariant measure, see Lemmas 18 and 19.

Our work is closely related to the recent work by Chen, Dwivedi, Wainwright and Yu [7]. In the former paper, conductance methods are applied to prove TV convergence bounds for adjusted HMC assuming a warm start and that either an isoperimetric or a log-isoperimetric inequality holds. An important ingredient in their proofs is an overlap bound that is similar to the more refined bounds we develop for the one-shot coupling. For a warm start, it is shown in [7] that an ε-accurate approximation can be achieved by adjusted HMC with $O(d^{11/12}\log(\varepsilon^{-1}))$ gradient evaluations for strongly log-concave target distributions, and with $O(d^{4/3}\log(\varepsilon^{-1}))$ gradient evaluations for weakly non-log-concave target distributions. The convergence bounds for uHMC stated below have a substantially better dependence on the dimension d and on the initial law. In particular, we show that the mixing time of the uHMC Markov chain often depends only logarithmically on d. The price to pay is that both the required step size and, correspondingly, the required number of gradient evaluations per transition step of the Markov chain, depend substantially on the accuracy ε. Unfortunately, it is currently not clear how to develop a coupling approach that can provide sharp bounds for adjusted HMC. The problem is that straightforward couplings are not efficient in combination with accept-reject steps. See however the recent work [9] for promising first steps in this direction.

Overlap bounds for the one-step transition distributions of HMC on Riemannian manifolds were also developed by Lee and Vempala [24, 23]. This variant of HMC is stated in terms of the exact Hamiltonian flow, and their idea is to use an ODE solver to approximate this flow without Metropolis adjustment. The overlap bounds were used to estimate the conductance and in turn mixing time of Riemannian HMC with application to faster methods for polytope volume computation. For a flat space, their overlap bounds are stated in terms of the Frobenius norm of the Hessian of the potential energy, which can lead to an implicit dimension dependence in high dimensional models. In contrast, we assume only a bound on the operator norm of the Hessian. To see this difference, compare Assumption 2 below to Definition 17 of [23].

A related strategy as in our work has also recently been implemented for an "OBABO" discretization of second-order Langevin dynamics; see Proposition 3 and Proposition 22 of [36]. In this case, the proofs simplify substantially because they involve designing a one-shot coupling based on only one step of the OBABO scheme instead of many Verlet steps. The price to pay is that the resulting bounds in [36] are less sharp because in contrast to Lemma 16 below, the overlap bound for OBABO applies only if the two copies are very close to each other (within distance of order $O(h^{1/2})$, where h is the time step size).

We now outline the main results of this paper. Let $\pi(x,dy)$ denote the one-step transition kernel of exact HMC with fixed integration time $T > 0$ of the Hamiltonian flow during each transition step, and let $\tilde{\pi}(x,dy)$ denote the one-step transition kernel of uHMC operated with integration time T and discretization time step size $h > 0$ satisfying $T \in h\mathbb{Z}$. Denote by μ

and $\tilde{\mu}$ the corresponding invariant probability measures of exact HMC and unadjusted HMC, respectively, and let $\mathrm{TV}(\nu, \eta)$ denote the TV distance between probability measures ν, η on $\mathbb{R}^d$.

The main result of this paper is a quantitative bound on $\mathrm{TV}(\tilde{\mu}, \nu\tilde{\pi}^{m+1})$ that holds for any $m \geq 0$ and any probability measure ν on $\mathbb{R}^d$. More precisely, by using a one-shot coupling we first obtain

$$\mathrm{TV}(\tilde{\mu}, \nu\tilde{\pi}^{m+1}) \leq (3/4) \left(T^{-2} + 27\, d\, L_H^2\, T^4\right)^{1/2} \boldsymbol{W}^1(\tilde{\mu}, \nu\tilde{\pi}^m) \tag{1}$$

where $\boldsymbol{W}^1$ is the standard L^1-Wasserstein distance. Therefore, if the uHMC chain converges geometrically in $\boldsymbol{W}^1$, i.e., $\boldsymbol{W}^1(\tilde{\mu}, \nu\tilde{\pi}^m) \leq M_1 e^{-cm} \boldsymbol{W}^1(\tilde{\mu}, \nu)$, then

$$\mathrm{TV}(\tilde{\mu}, \nu\tilde{\pi}^{m+1}) \leq (3/4) \left(T^{-2} + 27\, d\, L_H^2\, T^4\right)^{1/2} M_1\, e^{-cm}\, \boldsymbol{W}^1(\tilde{\mu}, \nu)\,. \tag{2}$$

As long as the constant M_1 and the L^1-Wasserstein distance between the initial distribution of the chain and the invariant measure of uHMC $\boldsymbol{W}^1(\tilde{\mu}, \nu)$ depend polynomially on the dimension d, (2) implies that the mixing time of the uHMC chain depends at most logarithmically on d. Alternatively, it would also be possible to rephrase the bound in (2) in the spirit of [16] as a contraction bound in a Kantorovich distance that has both a TV and a Wasserstein part.

In a second step, we quantify the TV distance between $\tilde{\mu}$ and μ. By the triangle inequality

$$\begin{aligned} \mathrm{TV}(\mu, \tilde{\mu}) &\leq \mathrm{TV}(\mu\pi, \mu\tilde{\pi}) + \mathrm{TV}(\mu\tilde{\pi}, \tilde{\mu}\tilde{\pi}) \\ &\leq \mathrm{TV}(\mu\pi, \mu\tilde{\pi}) + (3/4) \left(T^{-2} + 27\, d\, L_H^2\, T^4\right)^{1/2} \boldsymbol{W}^1(\mu, \tilde{\mu})\,. \end{aligned} \tag{3}$$

By strong accuracy of the underlying integrator, i.e., $\boldsymbol{W}^1(\mu, \tilde{\mu}) \leq M_2\, h^2$, and using another one-shot coupling to estimate $\mathrm{TV}(\mu\pi, \mu\tilde{\pi})$, we obtain

$$\mathrm{TV}(\mu, \tilde{\mu}) \leq C\, h^2 + (3/4) \left(T^{-2} + 27\, d\, L_H^2\, T^4\right)^{1/2} M_2\, h^2\,, \tag{4}$$

where C is a constant given in (13) which grows like $d^{3/2}$ for general U and $d^{1/2}$ for quadratic U; and the constant M_2 grows like d for general U and $d^{1/2}$ for quadratic U. For mean-field U, the corresponding upper bound on $\mathrm{TV}(\mu, \tilde{\mu})$ grows like $O(dh^2)$. Applying the triangle inequality again, and inserting (2) and (4) gives an overall convergence bound on $\mathrm{TV}(\mu, \nu\tilde{\pi}^{m+1})$. These results are stated precisely in Theorems 5 and 7, and Corollary 8 in the next section for general U, and in Theorems 9 and 11, and Corollary 12 for mean-field U. The remaining sections contain detailed proofs. In the special case $T = h$, i.e., one integration step per HMC step, one recovers TV convergence bounds for the unadjusted Langevin algorithm (uLA) [13]. In this case, (2) shows that the mixing time again depends logarithmically on d, while (4) shows that the accuracy of the invariant measure is $O(dh)$ for general U.

We conclude this introduction by remarking that mixing time bounds based on coupling methods might be relevant to recently developed *unbiased* estimators based on couplings [21]. Both in theory and in practice, the usefulness of these unbiased estimators requires a successful coupling that realizes these bounds. Therefore, to understand the performance of these unbiased estimators, it is crucial to obtain quantitative TV convergence bounds and geometric tail bounds on the corresponding coupling times.

2. Main Results

2.1. Notation

Let $\mathcal{P}(\mathbb{R}^d)$ denote the set of probability measures on $\mathbb{R}^d$. Define the total variation (TV) distance between $\nu, \eta \in \mathcal{P}(\mathbb{R}^d)$ by

$$\mathrm{TV}(\nu,\eta) \ := \ \sup\{|\nu(A) - \eta(A)| : A \in \mathcal{B}(\mathbb{R}^d)\}$$

where $\mathcal{B}(\mathbb{R}^d)$ is the Borel σ-algebra on $\mathbb{R}^d$. Denote the set of all couplings of $\nu, \eta \in \mathcal{P}(\mathbb{R}^d)$ by $\mathrm{Couplings}(\nu,\eta)$. A useful property of the TV distance is the following coupling characterization

$$\mathrm{TV}(\nu,\eta) \ = \ \inf\left\{P[X \neq Y] \ : \ \mathrm{Law}(X,Y) \in \mathrm{Couplings}(\nu,\eta)\right\} . \tag{5}$$

Let d be a metric on $\mathbb{R}^d$. For $\nu, \eta \in \mathcal{P}(\mathbb{R}^d)$, define the L^1-Wasserstein distance with respect to d by

$$\boldsymbol{W}^1_{\mathsf{d}}(\nu,\eta) \ := \ \inf\left\{E\left[\mathsf{d}(X,Y)\right] \ : \ \mathrm{Law}(X,Y) \in \mathrm{Couplings}(\nu,\eta)\right\} .$$

In the special case where d is the standard Euclidean metric, we write the corresponding L^1-Wasserstein distance as $\boldsymbol{W}^1$.

2.2. Short Overview of Unadjusted Hamiltonian Monte Carlo

Unadjusted Hamiltonian Monte Carlo (uHMC) is an MCMC method for approximate sampling from a 'target' probability distribution on $\mathbb{R}^d$ of the form

$$\mu(dx) = \mathcal{Z}^{-1} \exp(-U(x))\, dx \ , \quad \mathcal{Z} = \int \exp(-U(x))\, dx \ , \tag{6}$$

where $U : \mathbb{R}^d \to \mathbb{R}$ is a twice continuously differentiable function such that $\mathcal{Z} < \infty$. The function U is interpreted as a potential energy. The uHMC algorithm generates a Markov chain on $\mathbb{R}^d$ with the help of: (i) a sequence $(\xi_k)_{k\in\mathbb{N}_0}$ of i.i.d. random variables $\xi_k \sim \mathcal{N}(0, I_d)$; and (ii) the velocity Verlet integrator with time step size $h > 0$ and initial condition $(x,v) \in \mathbb{R}^{2d}$ whose discrete solution takes values on an evenly spaced temporal grid $\{t_i := ih\}_{i\in\mathbb{N}_0}$ and is interpolated by $(\tilde{q}_t(x,v), \tilde{v}_t(x,v))$ which satisfies the following differential equations

$$\frac{d}{dt}\tilde{q}_t \ = \ \tilde{v}_{\lfloor t\rfloor_h} - (t - \lfloor t\rfloor_h)\nabla U(\tilde{q}_{\lfloor t\rfloor_h}), \qquad \frac{d}{dt}\tilde{v}_t \ = \ -\frac{1}{2}\left(\nabla U(\tilde{q}_{\lfloor t\rfloor_h}) + \nabla U(\tilde{q}_{\lceil t\rceil_h})\right) \tag{7}$$

with initial condition $(\tilde{q}_0(x,v), \tilde{v}_0(x,v)) = (x,v) \in \mathbb{R}^{2d}$. Here we have defined

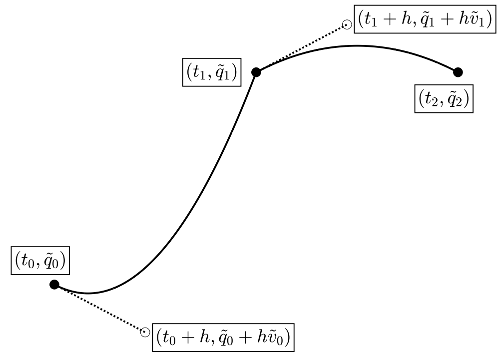
$$\lfloor t\rfloor_h = \max\{s \in h\mathbb{Z} \ : \ s \le t\} \quad \text{and} \quad \lceil t\rceil_h = \min\{s \in h\mathbb{Z} \ : \ s \ge t\} \quad \text{for } h > 0 \ . \tag{8}$$

For $h = 0$, we set $\lfloor t\rfloor_h = \lceil t\rceil_h = t$ and drop the tildes in the notation. Thus

$$\frac{d}{dt}q_t \ = \ v_t, \qquad \frac{d}{dt}v_t \ = \ -\nabla U(q_t), \tag{9}$$

and correspondingly, $(q_t(x,v), v_t(x,v))$ is the exact Hamiltonian flow with respect to the unit mass Hamiltonian $H(x,v) = (1/2)|v|^2 + U(x)$.

By integrating (7), note that $\tilde{q}_t$ is a piecewise quadratic function of time that interpolates between the points $\{(t_k, \tilde{q}_{t_k})\}$ and satisfies $\frac{d}{dt}\tilde{q}_t\big|_{t=t_k+} = \tilde{v}_{t_k}$, while $\tilde{v}_t$ is a piecewise linear function of time that interpolates between the points $\{(t_k, \tilde{v}_{t_k})\}$. In general, $\tilde{q}_t$ does not satisfy $\frac{d}{dt}\tilde{q}_t\big|_{t=t_k+} = \frac{d}{dt}\tilde{q}_t\big|_{t=t_k-}$, as illustrated in the figure. In the analysis of the convergence properties of uHMC, this continuous-time interpolation of the discrete solution produced by velocity Verlet is convenient to work with.



piecewise quadratic interpolation of positions

In the n-th step of the *unadjusted Hamiltonian Monte Carlo algorithm with complete velocity refreshment and duration parameter* $T \in (0,\infty)$, the initial velocity ξ_{n-1} is sampled independently of the previous development, and the current position and initial velocity are evolved by applying the Verlet approximation of the Hamiltonian flow over a time interval of length T.

Definition 1 (uHMC Markov chain). *Given an initial state $x \in \mathbb{R}^d$, duration $T > 0$, and time step size $h \geq 0$ with $T/h \in \mathbb{Z}$ for $h \neq 0$, define $\tilde{X}_0(x) := x$ and*

$$\tilde{X}_n(x) := \tilde{q}_T(\tilde{X}_{n-1}(x), \xi_{n-1}) \quad \text{for } n \in \mathbb{N}\,.$$

Let $\tilde{\pi}(x,A) = P[\tilde{X}_1(x) \in A]$ denote the corresponding one-step transition kernel.

For $h = 0$, we recover the *exact Hamiltonian Monte Carlo algorithm*. In this case, we drop all tildes in the notation, i.e., the n-th transition step is denoted by $X_n(x)$, and the corresponding transition kernel is denoted by π. The target measure μ is invariant under π, because the Hamiltonian flow (9) preserves the probability measure on $\mathbb{R}^{2d}$ with density proportional to $\exp(-H(x,v))$, and μ is the first marginal of this measure. When $h > 0$, the invariant probability measure for $\tilde{\pi}$ is denoted by $\tilde{\mu}$. In general, it does not agree with μ, but when it exists, it typically approaches μ as $h \downarrow 0$.

2.3. Assumptions

Let $\mathbf{H} \equiv D^2 U$ denote the Hessian of U. To prove our main results, we assume:

Assumption 2. *The function $U : \mathbb{R}^d \to \mathbb{R}$ satisfies the following conditions:*

(A1) *U has a global minimum at 0 and $U(0) = 0$.*
(A2) *$U \in \mathcal{C}^2(\mathbb{R}^d)$ with bounded second derivative, i.e., $L := \sup \|\mathbf{H}\| < \infty$.*

(A3) $U \in \mathcal{C}^3(\mathbb{R}^d)$ *with bounded third derivative, i.e.,* $L_H := \sup \|D\mathbf{H}\| < \infty$.
(A4) $U \in \mathcal{C}^4(\mathbb{R}^d)$ *with bounded fourth derivative, i.e.,* $L_I := \sup \|D^2\mathbf{H}\| < \infty$.

Here $\|\cdot\|$ *denotes the operator norm of a multilinear form.*

Remark 3 (Choice of norm and dimension dependence)**.** *The dimension dependence of the subsequent results depends on the choice of norm used in Assumptions (A2)-(A4). By the Cauchy-Schwarz inequality, note that the operator norm* $\|\cdot\|$ *of a* k*-multilinear form* $A = (a_{i_1\cdots i_k})$ *is always bounded by the Frobenius (or Hilbert-Schmidt) norm that is defined by* $\|A\|_F = (\sum_{i_1,\ldots,i_k=1}^{d,\ldots,d} a_{i_1\cdots i_k}^2)^{1/2}$, *i.e.,* $\|A\| \le \|A\|_F$, *see [17, Section 2]. Choosing the Frobenius norm in Assumptions (A2)-(A4) would give an improved dimension dependence in the results below.*

We additionally assume that the transition kernel $\tilde{\pi}$ of uHMC satisfies the following L^1-Wasserstein convergence bound and discretization error bound.

Assumption 4. *There exists an invariant probability measure* $\tilde{\mu}$ *of* $\tilde{\pi}$ *satisfying the following conditions.*

(A5) *There exist constants* $M_1, c \in (0,\infty)$ *such that for any* $m \in \mathbb{N}$ *and* $\nu \in \mathcal{P}(\mathbb{R}^d)$, $\boldsymbol{W}^1(\nu\tilde{\pi}^m, \tilde{\mu}) \le M_1\, e^{-cm}\, \boldsymbol{W}^1(\nu, \tilde{\mu})$.
(A6) *There exists a constant* $M_2 \in (0,\infty)$ *such that* $\boldsymbol{W}^1(\mu, \tilde{\mu}) \le M_2\, h^2$.

Note that by (A5), the invariant probability measure is unique. Assumption (A6) often holds as a consequence of (A5), a triangle-inequality trick [33, Remark 6.3] and strong accuracy of the Verlet integrator; see Section 3.2 of [6]. In general, the constants M_1, M_2 and c appearing in (A5) and (A6) will depend on the dimension d, but usually they can be chosen independently of the discretization step size h. In Section 2.5, we will see several classes of examples where c can be chosen independently of d, and the dimension dependence of M_1 and M_2 is explicit.

2.4. TV convergence bounds for uHMC

We are now in position to state our main results. For all of the following results, we assume that Assumptions 2 and 4 are satisfied. Moreover, we fix a duration parameter $T > 0$ and a time step size $h \ge 0$ such that $T/h \in \mathbb{Z}$ for $h \ne 0$ and

$$L(T^2 + Th) \le 1/6\,. \tag{10}$$

Theorem 5. *For any* $m \in \mathbb{N}_0$ *and for any initial law* $\nu \in \mathcal{P}(\mathbb{R}^d)$,

$$\mathrm{TV}(\tilde{\mu}, \nu\tilde{\pi}^{m+1}) \le (3/4)\left(T^{-2} + 27\, d\, L_H^2\, T^4\right)^{1/2} M_1\, \boldsymbol{W}^1(\tilde{\mu}, \nu)\, e^{-c\,m}\,. \tag{11}$$

Proof. For $m = 0$, the statement is a special case of a $\boldsymbol{W}^1$/TV regularization result that is proven in Lemma 17 below. The general case follows by combining Lemma 17 and Assumption 4 (A5). □

Theorem 5 immediately implies an upper bound on the ε-mixing time $t_{\mathrm{mix}}(\varepsilon,\nu) := \inf\{m \ge 0 : \mathrm{TV}(\tilde\mu, \nu\tilde\pi^m) \le \varepsilon\}$ of the uHMC Markov chain.

Corollary 6 (Upper bound for mixing time)**.** *For any $\varepsilon > 0$ and any $\nu \in \mathcal{P}(\mathbb{R}^d)$,*

$$t_{\mathrm{mix}}(\varepsilon,\nu) \;\le\; 2 + \frac{1}{c}\log\Big(\frac{3\left(T^{-2} + 27\, d\, L_H^2\, T^4\right)^{1/2}\; M_1\, \boldsymbol{W}^1(\tilde\mu,\nu)}{4\varepsilon}\Big)\,.$$

In particular, provided that M_1 and $\boldsymbol{W}^1(\tilde\mu,\nu)$ depend at most polynomially on d and c is independent of d, then the mixing time depends at most logarithmically on d and ε. Since simulating one step of the uHMC chain requires carrying out T/h gradient evaluations (for $h \neq 0$), an ε-approximation of $\tilde\mu$ in total variation distance can be achieved with $O\left(h^{-1}\log(d/\varepsilon)\right)$ gradient evaluations. However, in order to control the approximation error with respect to the target distribution μ, we also have to take into account the systematic error $\mathrm{TV}(\mu,\tilde\mu)$.

Theorem 7. *Under the assumptions made above,*

$$\mathrm{TV}(\mu,\tilde\mu) \;\le\; h^2\Big[(3/4)\,\left(T^{-2} + 27\, d\, L_H^2\, T^4\right)^{1/2}\, M_2 + C\Big], \quad \textit{where} \tag{12}$$

$$\begin{aligned} C \;:=\; (1/2)\,\Big[& d^3\left(4L_I^2T^4 + 14L_H^2L_IT^6 + 14L_H^4T^8\right) \\ & + d^2\left(35L_H^2T^2 + 8L_I^2T^4 + 28L_H^2L_IT^6 + 28L_H^4T^8\right) \\ & + d\left(16L^2 + 4L_H^2T^2\right) + \left(2dL_H^2 + L^2T^{-2}\right)\int |x|^2\,\mu(dx) \\ & + \left(dL_I^2 + dL_H^2L_IT^2 + dL_H^4T^4 + L_H^2T^{-2}\right)\int |x|^4\,\mu(dx)\Big]^{1/2}\,. \end{aligned} \tag{13}$$

Proof. By the triangle inequality,

$$\mathrm{TV}(\mu,\tilde\mu) \;\le\; \mathrm{TV}(\mu\pi,\mu\tilde\pi) + \mathrm{TV}(\mu\tilde\pi,\tilde\mu\tilde\pi)\,. \tag{14}$$

Moreover, by Lemma 19 below,

$$\mathrm{TV}(\mu\pi,\mu\tilde\pi) \;\le\; C\,h^2\,. \tag{15}$$

Moreover, by Lemma 17, and by Assumption 4 (A6),

$$\begin{aligned} \mathrm{TV}(\mu\tilde\pi,\tilde\mu\tilde\pi) \;&\le\; (3/4)\,\left(T^{-2} + 27\, d\, L_H^2\, T^4\right)^{1/2}\, \boldsymbol{W}^1(\mu,\tilde\mu) \\ &\le\; (3/4)\,\left(T^{-2} + 27\, d\, L_H^2\, T^4\right)^{1/2}\, M_2\, h^2\,. \end{aligned} \tag{16}$$

Inserting (16) and (15) into (14) gives (12). □

By the triangle inequality, note that

$$\mathrm{TV}(\nu\tilde\pi^{m+1},\mu) \;\le\; \mathrm{TV}(\nu\tilde\pi^{m+1},\tilde\mu) + \mathrm{TV}(\tilde\mu,\mu)\,. \tag{17}$$

Inserting (12) and (11) into (17) gives the following corollary.

Corollary 8. *For any $m \in \mathbb{N}_0$ and any initial law $\nu \in \mathcal{P}(\mathbb{R}^d)$,*

$$\begin{aligned} \mathrm{TV}(\mu, \nu\tilde{\pi}^{m+1}) \le\ & (3/4)\left(T^{-2} + 27\, d\, L_H^2\, T^4\right)^{1/2} M_1\, \boldsymbol{W}^1(\tilde{\mu}, \nu)\, e^{-c\, m} \\ & + h^2 \Big[(3/4)\left(T^{-2} + 27\, d\, L_H^2\, T^4\right)^{1/2} M_2 + C\Big]. \end{aligned} \tag{18}$$

The constant C appearing in (18) is typically of order $O(d^{3/2})$, although in the Gaussian case the dimension dependence improves further. Below, we will see examples where the constant M_2 appearing in (18) is of order $O(d)$. Then the TV accuracy $\mathrm{TV}(\mu, \tilde{\mu})$ is at most of order $O(h^2 d^{3/2})$, and correspondingly, for an ε-accurate approximation of μ by $\tilde{\mu}$, the step size h has to be chosen of order $O(d^{-3/4}\varepsilon^{1/2})$. Thus if additionally, c is independent of the dimension, and M_1 depends at most polynomially on d, then an ε-accurate approximation of the target distribution μ in TV distance can be achieved by uHMC with $O\left(d^{3/4}\varepsilon^{-1/2}\log(d/\varepsilon)\right)$ gradient evaluations.

2.5. Examples

By coupling methods, (A5) has been verified in the following models.

2.5.1. Asymptotically Strongly Logconcave Target

Suppose that $U : \mathbb{R}^d \to \mathbb{R}$ satisfies (A1) and (A2), and U is strongly convex outside a Euclidean ball, i.e., there exist constants $\mathcal{R} \in [0, \infty)$ and $K \in (0, \infty)$ such that

$$(x - y) \cdot (\nabla U(x) - \nabla U(y)) \ \ge\ K\,|x - y|^2 \quad \text{for all } x, y \in \mathbb{R}^d \text{ with } |x - y| \ge \mathcal{R}.$$

Using a synchronous coupling, and assuming that $T > 0$ satisfies $LT^2 \le 1/4$, Chen and Vempala (2019) [8] proved in the globally strongly logconcave case (i.e., for $\mathcal{R} = 0$) that for all initial distributions $\nu, \eta \in \mathcal{P}(\mathbb{R}^d)$, and for all $m \ge 0$, the transition kernel of exact HMC satisfies

$$\boldsymbol{W}^1(\nu\pi^m, \eta\pi^m) \ \le\ e^{-cm}\, \boldsymbol{W}^1(\nu, \eta) \quad \text{where} \quad c \ =\ KT^2/10.$$

Mangoubi and Smith (2017) [31] obtained a similar result under $LT^2 \le \min(1/4, K/L)$. Simple counterexamples demonstrate that a synchronous coupling is not contractive in the non-logconcave setting where $\mathcal{R} > 0$. Recently, a non-synchronous coupling tailored to HMC was introduced to obtain a corresponding result for non-convex potentials. Suppose that $T > 0$ and $h_1 \ge 0$ satisfy

$$L(T + h_1)^2 \le \min\left(\frac{3K}{10L}, \frac{1}{4}, \frac{3K}{256 \cdot 5 \cdot 2^6\, L\mathcal{R}^2(L + K)}\right) \quad \text{and} \quad h_1 \le \frac{KT}{525L + 235K}\,.$$

For any $h \in [0, h_1]$, Bou-Rabee, Eberle and Zimmer (2020) [5] essentially prove that for all $m \ge 0$ and all $\nu, \eta \in \mathcal{P}(\mathbb{R}^d)$,

$$\begin{aligned} &\boldsymbol{W}^1(\nu\tilde{\pi}^m, \eta\tilde{\pi}^m) \ \le\ M_1\, e^{-cm}\, \boldsymbol{W}^1(\nu, \eta), \qquad \text{where} \\ &M_1 \ =\ \exp\left(\frac{5}{2}\Big(1 + \frac{4\mathcal{R}}{T}\sqrt{\frac{L + K}{K}}\Big)\right) \quad \text{and} \quad c \ =\ \frac{KT^2}{156}\exp\left(-10\frac{\mathcal{R}}{T}\sqrt{\frac{L + K}{K}}\right). \end{aligned}$$

Intuitively speaking, the factor $L\mathcal{R}^2$ measures the degree of non-convexity of U. The bounds show that (A5) is satisfied with the explicit constants given above. Now suppose additionally that U is in $C^3(\mathbb{R}^d)$ and $L_H = \sup \|D^3 U\| < \infty$. Then by Corollary 7 of [6] with $n = 1$ and $\epsilon = 0$, we have

$$\boldsymbol{W}^1(\mu, \tilde{\mu}) \le \frac{1}{c}\tilde{C}_2 M_1\Big(d + \int |x|\,\mu(dx) + \int |x|^2\,\mu(dx)\Big)\,h^2$$

where $\tilde{C}_2$ depends only on K, L, L_H and T. Therefore, Assumption (A6) holds with

$$M_2 = \frac{1}{c}\tilde{C}_2 M_1\Big(d + \int |x|\,\mu(dx) + \int |x|^2\,\mu(dx)\Big)\,.$$

If the constants $\mathcal{R}$, K and L are fixed, then c and M_1 do not depend on the dimension. As a consequence, by Corollary 6, the mixing time is of order $O(\log d)$, and by Theorem 7, the TV accuracy $\mathrm{TV}(\mu, \tilde{\mu})$ is of order $O(d^{3/2}h^2)$. In particular, an ε-accurate approximation can be achieved with a step size of order $O(\varepsilon^{-1/2}h^{-3/4})$. The latter bound is not sharp for strongly logconcave product models with i.i.d. factors where one can prove by elementary methods that the correct order of $\mathrm{TV}(\mu, \tilde{\mu})$ is $\Theta(d^{1/2}h^2)$. In the general setup considered above, however, we do not expect a bound for TV accuracy of a similar order to hold. In a complimentary work, Durmus and Eberle [12] show that the L^1-Wasserstein accuracy $\boldsymbol{W}^1(\mu, \tilde{\mu})$ is of order $O(d^{1/2}h^2)$ for "nice" models and of order $O(dh^2)$ for general models satisfying the assumptions made above.

In general, we can not expect that constants $\mathcal{R}$, K and L as above can be chosen independently of the dimension d. Then also c and M_1 may depend implicitly on the dimension through these constants. For example, in mean-field models, which we consider next, $\mathcal{R}$ can increase with the number of particles; see Remark 1 of [6].

2.5.2. *Mean-Field Model*

Consider a mean-field model [22, 35, 6] consisting of n particles in dimension k with potential energy $U : \mathbb{R}^{nk} \to \mathbb{R}$ defined as

$$U(x) \;=\; \sum_{i=1}^{n}\Big(V(x^i) + \frac{\epsilon}{n}\sum_{\ell=1,\ell\neq i}^{n} W(x^i - x^\ell)\Big)\,, \quad x \;=\; (x^1, \ldots, x^n)\,, \quad x^i \in \mathbb{R}^k\,. \tag{19}$$

We assume that V, W are functions in $C^2(\mathbb{R}^k)$ satisfying:

- V has a local minimum at 0, and $V(0) = 0$;
- $L = \sup \|D^2 V\| < \infty$ and $\tilde{L} = \sup \|D^2 W\| < \infty$; and,
- there exist constants $\mathcal{R} \in [0, \infty)$ and $K \in (0, \infty)$ such that

$$(x^1 - y^1)\cdot(\nabla V(x^1) - \nabla V(y^1)) \;\ge\; K\left|x^1 - y^1\right|^2$$

  for all $x^1, y^1 \in \mathbb{R}^k$ with $\left|x^1 - y^1\right| \ge \mathcal{R}$.

Suppose that $T > 0$, $\epsilon \geq 0$ and $h_1 \geq 0$ satisfy

$$\begin{aligned} L(T+h_1)^2 &\leq \frac{3}{5}\min\left(\frac{3K}{10L},\frac{1}{4},\frac{3K}{256\cdot 5\cdot 2^6\,L\mathcal{R}^2(L+K)}\right), \\ |\epsilon|\tilde{L} &< \min\left(\frac{K}{6},\frac{1}{2}\left(\frac{K}{36\cdot 149}\right)^2\left(T+8\mathcal{R}\sqrt{\frac{L+K}{K}}\right)^2\exp\left(-40\frac{\mathcal{R}}{T}\sqrt{\frac{L+K}{K}}\right)\right), \\ h_1 &\leq \frac{KT}{525L+235K}\,. \end{aligned}$$

Then for all initial distributions $\nu,\eta \in \mathcal{P}(\mathbb{R}^{nk})$, for all $m \geq 0$ and for any $h \in [0,h_1]$, Bou-Rabee and Schuh (2020) [6] use a component-wise coupling method to prove that

$$\begin{aligned} &\boldsymbol{W}^1_{\ell_1}(\nu\tilde{\pi}^m,\eta\tilde{\pi}^m) \leq M\,e^{-cm}\,\boldsymbol{W}^1_{\ell_1}(\nu,\eta), \quad \text{where} \quad \ell_1(x,y) := \sum_{i=1}^n |x^i-y^i|, \\ &M = \exp\left(\frac{5}{2}(1+\frac{4\mathcal{R}}{T}\sqrt{\frac{L+K}{K}})\right) \quad \text{and} \quad c = \frac{KT^2}{156}\exp\left(-10\frac{\mathcal{R}}{T}\sqrt{\frac{L+K}{K}}\right). \end{aligned} \tag{20}$$

Since $|x-y| \leq \ell_1(x,y) \leq \sqrt{n}\,|x-y|$, we obtain

$$\boldsymbol{W}^1(\nu\tilde{\pi}^m,\eta\tilde{\pi}^m) \leq M\,e^{-cm}\,\boldsymbol{W}^1_{\ell_1}(\nu,\eta) \leq \sqrt{n}\,M\,e^{-cm}\,\boldsymbol{W}^1(\nu,\eta)\,.$$

Thus Assumption (A5) holds with $M_1 = \sqrt{n}M$ and c as above. In particular, c is independent of the number of particles n, and therefore, the mixing time depends only logarithmically on n. This result also implies that there exists a unique invariant probability measure $\tilde{\mu}$ of $\tilde{\pi}$; see Corollary 5 of [6]. In this mean-field context, we can also obtain slightly better dimension-dependence in the TV convergence bounds for uHMC. To state these bounds, we additionally assume that V, W are functions in $C^3(\mathbb{R}^k)$ such that $L_H = \sup\|D^3V\| < \infty$ and $\tilde{L}_H = \sup\|D^3W\| < \infty$. Moreover, analogous to (10), we assume that $T > 0$ and the time step size $h \geq 0$ are such that $T/h \in \mathbb{N}$ for $h \neq 0$ and

$$(L+4\epsilon\tilde{L})(T^2+Th) \leq 1/6\,. \tag{21}$$

Theorem 9. *For any $m \in \mathbb{N}_0$ and for any initial law $\nu \in \mathcal{P}(\mathbb{R}^{nk})$,*

$$\mathrm{TV}(\tilde{\mu},\nu\tilde{\pi}^{m+1}) \leq \frac{3}{4}\left(T^{-2}+34\,k\,(L_H+8\epsilon\tilde{L}_H)^2\,T^4\right)^{1/2}\,M\,\boldsymbol{W}^1_{\ell_1}(\tilde{\mu},\nu)\,e^{-cm}\,. \tag{22}$$

Proof. For $m = 0$, (22) is a special case of a $\boldsymbol{W}^1_{\ell_1}$/TV regularization result proven in Lemma 30 below. The general case follows by combining Lemma 30 and (20). □

As before, Theorem 9 immediately implies an upper bound on the ε-mixing time $t_{\mathrm{mix}}(\varepsilon,\nu)$ of the uHMC Markov chain applied to mean-field U. We stress that the following upper bound typically depends on the number of particles n through $\boldsymbol{W}^1_{\ell_1}(\tilde{\mu},\nu)$, but this dimension dependence is typically polynomial in n, and therefore, the upper bound typically depends logarithmically on n.

Corollary 10 (Upper bound for mixing time). *For any $\varepsilon > 0$ and any $\nu \in \mathcal{P}(\mathbb{R}^{nk})$,*

$$t_{\mathrm{mix}}(\varepsilon,\nu) \;\le\; 2 + \frac{1}{c}\log\left(\frac{3\left(T^{-2} + 34\,k\,(L_H + 8\epsilon\tilde{L}_H)^2\,T^4\right)^{1/2}\; M\,\boldsymbol{W}^1_{\ell_1}(\tilde{\mu},\nu)}{4\varepsilon}\right).$$

Moreover, for mean-field U, we can obtain a better dimension-dependence in the upper bounds for the TV accuracy $\mathrm{TV}(\mu,\tilde{\mu})$. To this end, we additionally assume that V, W are functions in $C^4(\mathbb{R}^k)$ such that $L_I = \sup\|D^4 V\| < \infty$ and $\tilde{L}_I = \sup\|D^4 W\| < \infty$. By Corollary 7 of [6],

$$\begin{aligned}\boldsymbol{W}^1(\mu,\tilde{\mu}) \;&\le\; \boldsymbol{W}^1_{\ell_1}(\mu,\tilde{\mu})\\ &\le\; \frac{1}{c}\tilde{C}_2 M\Big(nk + \sum_{\ell=1}^n \int |x^\ell|\,\mu(dx) + \sum_{\ell=1}^n \int |x^\ell|^2\,\mu(dx)\Big)\,h^2\end{aligned} \tag{23}$$

where $\tilde{C}_2$ depends only on K, L, $\mathcal{R}$, $\tilde{L}$, L_H, $\tilde{L}_H$, and T. Therefore, Assumption (A6) holds with

$$M_2 \;=\; \frac{1}{c}\tilde{C}_2 M\Big(nk + \sum_{\ell=1}^n \int |x^\ell|\,\mu(dx) + \sum_{\ell=1}^n \int |x^\ell|^2\,\mu(dx)\Big)\,.$$

Note that M_2 depends linearly on the number of particles and on the sum of the first and second moments of the n components of the target distribution.

Theorem 11. *Under the assumptions made above,*

$$\mathrm{TV}(\mu,\tilde{\mu}) \;\le\; h^2\Big[(3/4)\,\left(T^{-2} + 34\,k\,(L_H + 8\epsilon\tilde{L}_H)^2\,T^4\right)^{1/2}\,M_2 + C\Big], \quad \textit{where} \tag{24}$$

$$\begin{aligned} C \;:=\; \frac{1}{2}\Big[n^2 k\Big(&17(L + 4\epsilon\tilde{L})^2 + 28(L_H + 8\epsilon\tilde{L}_H)^2 T^2 + 104k(L_H + 8\epsilon\tilde{L}_H)^2 T^2\\ &+ 180(2k + k^2)T^4\big((L_I + 14\epsilon\tilde{L}_I) + (L_H + 8\epsilon\tilde{L}_H)^2 T^2\big)^2\Big)\\ &+ n\big(10k(L_H + 8\epsilon\tilde{L}_H)^2 + 7(L + 4\epsilon\tilde{L})^2 T^{-2}\big)\sum_{\ell=1}^n \int |x^\ell|^2\,\mu(dx)\\ &+ n\Big(40k\big((L_I + 14\epsilon\tilde{L}_I)^2 + (L_H + 8\epsilon\tilde{L}_H)^2 T^2\big)^2\\ &+ 7(L_H + 8\epsilon\tilde{L}_H)^2 T^{-2}\Big)\sum_{\ell=1}^n \int |x^\ell|^4\,\mu(dx)\Big]^{1/2}. \end{aligned} \tag{25}$$

For mean-field U, note that the constant C depends linearly on the number of particles n provided that $n^{-1}\sum_{\ell=1}^n \int |x^\ell|^4\,\mu(dx)$ does not depend on n.

Proof. By Lemma 32 below,

$$\mathrm{TV}(\mu\pi,\mu\tilde{\pi}) \;\le\; C\,h^2\,. \tag{26}$$

Moreover, by Lemma 30, and by (23),

$$\begin{aligned}\mathrm{TV}(\mu\tilde{\pi},\tilde{\mu}\tilde{\pi}) &\le (3/4)\left(T^{-2}+34\,k\,(L_H+8\epsilon\tilde{L}_H)^2\,T^4\right)^{1/2}\,\boldsymbol{W}^1_{\ell_1}(\mu,\tilde{\mu})\\ &\le (3/4)\left(T^{-2}+34\,k\,(L_H+8\epsilon\tilde{L}_H)^2\,T^4\right)^{1/2}\,M_2\,h^2\,. \qquad (27)\end{aligned}$$

Inserting (27) and (26) into (14) gives (24). □

Inserting (24) and (22) into (17) gives the following corollary.

Corollary 12. *For any $m\in\mathbb{N}_0$ and any initial law $\nu\in\mathcal{P}(\mathbb{R}^{nk})$,*

$$\begin{aligned}\mathrm{TV}(\mu,\nu\tilde{\pi}^{m+1}) \le\ & \frac{3}{4}\left(T^{-2}+34\,k\,(L_H+8\epsilon\tilde{L}_H)^2\,T^4\right)^{1/2}\,M\,\boldsymbol{W}^1_{\ell_1}(\tilde{\mu},\nu)\,e^{-cm}\\ &+h^2\left[\frac{3}{4}\left(T^{-2}+34\,k\,(L_H+8\epsilon\tilde{L}_H)^2\,T^4\right)^{1/2}\,M_2+C\right]. \end{aligned}\qquad (28)$$

3. Key ingredients of the proofs

3.1. Velocity Verlet as a Variational Integrator, Revisited

Here we recall a well-known variational characterization of the velocity Verlet integrator. In particular, velocity Verlet can be derived from a discrete-time variational principle, and hence, is a variational integrator [32]. A key tool in our analysis is a sufficient condition for convexity of the corresponding discrete action sum.

To this end, we introduce the *discrete Lagrangian* $L_h:\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{R}$ corresponding to velocity Verlet

$$L_h(x,y) = \frac{h}{2}\left(\frac{|y-x|^2}{h^2}-U(y)-U(x)\right)\,.$$

Given $\mathsf{a},\mathsf{b}\in\mathbb{R}^d$, define the *discrete path space*

$$\mathcal{C}_h=\left\{\tilde{q}:\{t_k\}_{k=0}^N\to\mathbb{R}^d\ \text{ with endpoint conditions } \tilde{q}_0=\mathsf{a}\text{ and }\tilde{q}_T=\mathsf{b}\right\}\,,$$

and the *discrete action sum* $S_h:\mathcal{C}_h\to\mathbb{R}$ by

$$S_h(\tilde{q})=\sum_{k=0}^{N-1}L_h(\tilde{q}_{t_k},\tilde{q}_{t_{k+1}})\,.$$

Note that $t_0=0$ and $t_N=T$. Computing the directional derivative of S_h in the direction of

$u : \{t_k\}_{k=0}^N \to \mathbb{R}^d$ with $u_0 = u_T = 0$ gives the first variation of S_h

$$\begin{aligned}\partial_u S_h(\tilde{q}) &= \frac{\partial}{\partial \epsilon} S_h(\tilde{q} + \epsilon u)\Big|_{\epsilon=0} \\ &= \sum_{k=0}^{N-1} [D_1 L_h(\tilde{q}_{t_k}, \tilde{q}_{t_{k+1}}) \cdot u_{t_k} + D_2 L_h(\tilde{q}_{t_k}, \tilde{q}_{t_{k+1}}) \cdot u_{t_{k+1}}] \\ &= \sum_{k=1}^{N-1} (D_1 L_h(\tilde{q}_{t_k}, \tilde{q}_{t_{k+1}}) + D_2 L_h(\tilde{q}_{t_{k-1}}, \tilde{q}_{t_k})) \cdot u_{t_k}\end{aligned}$$

where in the last step we used summation by parts and $u_0 = u_T = 0$. Stationarity of this action sum gives the discrete Euler-Lagrange equations

$$D_1 L_h(\tilde{q}_{t_k}, \tilde{q}_{t_{k+1}}) + D_2 L_h(\tilde{q}_{t_{k-1}}, \tilde{q}_{t_k}) = 0 , \quad k \in \{1, \dots, N-1\} . \tag{29}$$

Introducing discrete velocities and simplifying yields the velocity Verlet integrator

$$\left.\begin{aligned}\tilde{v}_{t_k} &= -D_1 L_h(\tilde{q}_{t_k}, \tilde{q}_{t_{k+1}}) \\ \tilde{v}_{t_{k+1}} &= D_2 L_h(\tilde{q}_{t_k}, \tilde{q}_{t_{k+1}})\end{aligned}\right\} \implies \begin{cases}\tilde{q}_{t_{k+1}} = \tilde{q}_{t_k} + h\tilde{v}_{t_k} - (h^2/2)\nabla U(\tilde{q}_{t_k}) , \\ \tilde{v}_{t_{k+1}} = \tilde{v}_{t_k} - (h/2)[\nabla U(\tilde{q}_{t_k}) + \nabla U(\tilde{q}_{t_{k+1}})] .\end{cases}$$

The following lemma indicates that for sufficiently short time intervals the discrete action sum S_h corresponding to velocity Verlet is strongly convex.

Lemma 13 (Müller & Ortiz 2004 [38]). *Suppose that (A2) holds and $T > 0$ satisfies $LT^2 \le (2/5)\pi^2$. For any $\mathsf{a}, \mathsf{b} \in \mathbb{R}^d$ and $h > 0$ satisfying $T/h \in \mathbb{N}$, the discrete action sum S_h is strongly convex. Moreover, for $h = 0$, the corresponding action integral is strongly convex.*

Proof. For $h = 0$, the proof of strong convexity of the action integral is given in Lemma 2.1 of [38]. For $h > 0$, the proof that follows shows that the second derivative of the action sum is positive definite. The second derivative of the action sum is given by

$$D^2 S_h(\tilde{q})(u,u) = \frac{\partial^2}{\partial \epsilon^2} S_h(\tilde{q} + \epsilon u)\Big|_{\epsilon=0} = I_1 + I_2 ,$$

where I_1 and I_2 are defined and bounded as follows. Using summation by parts, $u_0 = u_T = 0$, and $\|D^2 U\| < L$ (by assumption (A2)) yields

$$\begin{aligned}I_1 &:= \frac{-h}{2} \sum_{k=0}^{N-1} \big(D^2 U(\tilde{q}_{t_k})(u_{t_k}, u_{t_k}) + D^2 U(\tilde{q}_{t_{k+1}})(u_{t_{k+1}}, u_{t_{k+1}})\big) \\ &= -h \sum_{k=1}^{N-1} D^2 U(\tilde{q}_{t_k})(u_{t_k}, u_{t_k}) \ \ge\ -L \sum_{k=1}^{N-1} h|u_{t_k}|^2 .\end{aligned}$$

Applying a Poincaré inequality yields

$$I_2 := h \sum_{k=0}^{N-1} \frac{|u_{t_{k+1}} - u_{t_k}|^2}{h^2} \ \ge\ \frac{\pi^2}{T^2} \left(\frac{\sin(\pi/(2N))}{\pi/(2N)}\right)^2 \sum_{k=1}^{N-1} h|u_{t_k}|^2 \ \ge\ \frac{2}{5}\frac{\pi^2}{T^2} \sum_{k=1}^{N-1} h|u_{t_k}|^2 .$$

Thus, $D^2 S_h(\tilde{q})(u,u) \ge ((2/5)(\pi^2/T^2) - L) \sum_{k=1}^{N-1} h|u_{t_k}|^2$. □

As an immediate corollary to Lemma 13 we obtain.

Corollary 14. *Suppose that (A2) holds and $T > 0$ satisfies $LT^2 \le (2/5)\pi^2$. For any $\mathsf{a}, \mathsf{b} \in \mathbb{R}^d$, and for any $h > 0$ satisfying $T/h \in \mathbb{N}$, there exists a unique solution $\tilde{q}^\star : \{t_k\}_{k=0}^N \to \mathbb{R}^d$ of the discrete Euler-Lagrange equations (29) with endpoint conditions $\tilde{q}_0^\star = \mathsf{a}$ and $\tilde{q}_T^\star = \mathsf{b}$. Moreover, for $h = 0$, there exists a unique solution to the corresponding Euler-Lagrange equations.*

Note that $q^\star$ in Cor. 14 is the minimum of the corresponding action sum S_h.

3.2. Overlap between Reference and Perturbed Gaussian Measure

Let $\xi \sim \mathcal{N}(0,1)^d$ and $\Phi : \mathbb{R}^d \to \mathbb{R}^d$ be a differentiable, near identity map. Here we provide a general upper bound for the TV distance between the *reference* Gaussian measure $\mathrm{Law}(\xi)$ and the *perturbed* Gaussian measure $\mathrm{Law}(\Phi(\xi))$. The couple $(\xi, \Phi(\xi))$ is an example of a deterministic coupling [44, Definition 1.2]. In conjunction with Corollary 14, the general bound given below is crucial to proving that two copies of uHMC can meet in one step.

Lemma 15. *Let $\xi \sim \mathcal{N}(0,1)^d$ and suppose that $\Phi : \mathbb{R}^d \to \mathbb{R}^d$ is an invertible and differentiable map satisfying $\|D\Phi(v) - I_d\| \le 1/2$ for all $v \in \mathbb{R}^d$. Then*

$$\mathrm{TV}(\mathrm{Law}(\xi), \mathrm{Law}(\Phi(\xi))) \;\le\; \frac{1}{2}\sqrt{E[|\Phi(\xi) - \xi|^2 + 2\,\|D\Phi(\xi) - I_d\|_F^2]}\,. \tag{30}$$

Let φ denote the probability density function of the standard d-dimensional normal distribution satisfying

$$\mathrm{Law}(\xi)(dv) \;=\; \mathcal{N}(0, I_d)(dv) \;=\; \varphi(v)dv \;=\; (2\pi)^{-d/2}\exp(-|x|^2/2)dx\,.$$

By change of variables, note that

$$\mathrm{Law}(\Phi(\xi))(dv) \;=\; |\det(D\Phi^{-1}(v))|\varphi(\Phi^{-1}(v))dv\,. \tag{31}$$

To prove Lemma 15, we recall Pinsker's inequality. For probability measures $\nu_1, \nu_2 \in \mathcal{P}(\mathbb{R}^d)$ with probability densities p_1 and p_2 with respect to a common reference measure λ, define the Kullback-Leibler divergence by

$$\mathrm{KL}(\nu_1 \parallel \nu_2) \;:=\; \int_{\mathbb{R}^d} p_2(v) G\Big(\frac{p_1(v)}{p_2(v)}\Big)\lambda(dv) \;\text{ where }\; G(x) = \begin{cases} x\log(x) & \text{if } x \neq 0, \\ 0 & \text{else.} \end{cases}$$

Although non-negative, the Kullback-Leibler divergence is not a metric on $\mathcal{P}(\mathbb{R}^d)$ because it is not symmetric and does not satisfy the triangle inequality. In this situation to bound the TV distance between ν_1 and ν_2, it is convenient to use Pinsker's inequality

$$\mathrm{TV}(\nu_1, \nu_2) \;\le\; \sqrt{(1/2)\,\mathrm{KL}(\nu_1 \parallel \nu_2)}\,. \tag{32}$$

Proof. Since $\mathrm{TV}(\mathrm{Law}(\xi),\mathrm{Law}(\Phi(\xi)) = \mathrm{TV}(\mathrm{Law}(\xi),\mathrm{Law}(\Phi^{-1}(\xi))$, by Pinsker's inequality (32), it suffices to bound

$$\mathrm{KL}(\mathrm{Law}(\xi) \parallel \mathrm{Law}(\Phi^{-1}(\xi)) = \int_{\mathbb{R}^d} \frac{e^{-\frac{|v|^2}{2}}}{(2\pi)^{d/2}} \left[\frac{|\Phi(v)|^2}{2} - \frac{|v|^2}{2} - \log|\det D\Phi(v)| \right] dv$$
$$= \int_{\mathbb{R}^d} \frac{e^{-\frac{|v|^2}{2}}}{(2\pi)^{d/2}} \left[\frac{1}{2} |\Phi(v) - v|^2 + (\Phi(v) - v) \cdot v - \log|\det D\Phi(v)| \right] dv$$
$$= \int_{\mathbb{R}^d} \frac{e^{-\frac{|v|^2}{2}}}{(2\pi)^{d/2}} \left[\frac{1}{2} |\Phi(v) - v|^2 + \mathrm{Trace}(D\Phi(v) - I_d) - \log|\det D\Phi(v)| \right] dv \tag{33}$$

where in the last step we used the following integration by parts identity

$$\int_{\mathbb{R}^d} e^{-\frac{|v|^2}{2}} (\Phi(v) - v) \cdot v dv = \int_{\mathbb{R}^d} e^{-\frac{|v|^2}{2}} \mathrm{Trace}(D\Phi(v) - I_d) dv \,.$$

Since $\|D\Phi(v) - I_d\| \le 1/2$, the spectral radius of $D\Phi(v) - I_d$ does not exceed $1/2$. Therefore, we can invoke Theorem 1.1 of [40], to obtain

$$\mathrm{Trace}(D\Phi(v) - I_d) - \log|\det D\Phi(v)| \le \frac{\|D\Phi(v) - I_d\|_F^2 /2}{1 - \|D\Phi(v) - I_d\|} \le \|D\Phi(v) - I_d\|_F^2 \,. \tag{34}$$

Inserting (34) into (33) yields

$$\mathrm{KL}(\mathrm{Law}(\xi) \parallel \mathrm{Law}(\Phi^{-1}(\xi)) \le E\left[(1/2)\,|\Phi(\xi) - \xi|^2 + \|D\Phi(\xi) - I_d\|_F^2 \right] .$$

Applying Pinsker's inequality (32) gives (30). □

3.3. TV Bounds and Regularization by One-Shot Couplings

By using the one-shot coupling illustrated in Figure 1 (a), here we prove that the transition kernel of uHMC has a regularizing effect. To this end, we fix a duration parameter $T > 0$ and a time step size $h \ge 0$ such that $T/h \in \mathbb{Z}$ for $h \ne 0$ and

$$L(T^2 + Th) \le 1/6 \,. \tag{35}$$

Lemma 16. *Suppose Assumption 2 (A1)-(A3) hold. For any $x, y, v \in \mathbb{R}^d$, let $\Phi : \mathbb{R}^d \to \mathbb{R}^d$ be the map defined by $\tilde{q}_T(x, v) = \tilde{q}_T(y, \Phi(v))$. Then*

$$\mathrm{TV}(\delta_x \tilde{\pi}, \delta_y \tilde{\pi}) \le \mathrm{TV}(\mathrm{Law}(\xi), \mathrm{Law}(\Phi(\xi))) \le \frac{3}{4} \left(T^{-2} + 27 d L_H^2 T^4 \right)^{1/2} |x - y|. \tag{36}$$

This lemma can be viewed as a nontrivial refinement of (35) in the proof of Lemma 6 of [7], which presents overlap bounds for the transition kernel of uHMC.

Proof. Since $\tilde{q}_T : \mathbb{R}^{2d} \to \mathbb{R}^d$ is deterministic and measurable [30, Lemma 3],

$$\begin{aligned} \mathrm{TV}(\delta_x \tilde{\pi}, \delta_y \tilde{\pi}) &= \mathrm{TV}(\mathrm{Law}(\tilde{q}_T(y,\xi)), \mathrm{Law}(\tilde{q}_T(y,\Phi(\xi)))) \\ &\le \mathrm{TV}(\mathrm{Law}(\xi), \mathrm{Law}(\Phi(\xi)))\ , \end{aligned} \tag{37}$$

which gives the first inequality in (36). Moreover, by Lemmas 15, 25 and 26, we have

$$\mathrm{TV}(\mathrm{Law}(\xi), \mathrm{Law}(\Phi(\xi))^2 \ \le\ (1/4)\left((9/4)\, T^{-2} + (121/2)\, d\, L_H^2\, T^4\right) |x-y|^2$$

where we used $\|D\Phi(v) - I_d\|_F^2 \ \le\ d\, \|D\Phi(v) - I_d\|^2$. Taking square roots and inserting $27 > (4/9)(121/2)$ gives the second inequality in (36). □

Lemma 16 implies that the transition kernel of uHMC has a regularizing effect in the following sense (cf. Theorem 12 (a) of [30]).

Lemma 17. *Suppose Assumption 2 (A1)-(A3) hold. For any $\nu, \eta \in \mathcal{P}(\mathbb{R}^d)$,*

$$\mathrm{TV}(\eta\tilde{\pi}, \nu\tilde{\pi}) \ \le\ \frac{3}{4}\left(T^{-2} + 27 d L_H^2 T^4\right)^{1/2} \boldsymbol{W}^1(\eta,\nu)\ . \tag{38}$$

Proof. Let ω be an arbitrary coupling of ν, η. By the coupling characterization of the TV distance in (5),

$$\mathrm{TV}(\eta\tilde{\pi}, \nu\tilde{\pi}) \le E[\mathrm{TV}(\delta_X\tilde{\pi}, \delta_Y\tilde{\pi})] \le \frac{3}{4}\left(T^{-2} + 27 d L_H^2 T^4\right)^{1/2} E[|X-Y|]$$

where $\mathrm{Law}(X,Y) = \omega$ and in the last step we inserted (36) in Lemma 16. Since ω is arbitrary, we can take the infimum over all $\omega \in \mathrm{Couplings}(\nu,\eta)$ to obtain (38). □

To bound the bias in the invariant measure of uHMC, we bound the TV distance between $\mu\tilde{\pi}$ and $\mu\pi$ by using the one-shot coupling illustrated in Figure 1(b).

Lemma 18. *Suppose Assumption 2 holds. For any $x, v \in \mathbb{R}^d$, let $\Phi : \mathbb{R}^d \to \mathbb{R}^d$ be the map defined by $\tilde{q}_T(x,v) = q_T(x,\Phi(v))$. Then*

$$\begin{aligned} &\mathrm{TV}(\delta_x \pi, \delta_x \tilde{\pi}) \ \le\ \mathrm{TV}(\mathrm{Law}(\xi), \mathrm{Law}(\Phi(\xi))) \\ &\le\ \frac{h^2}{2}\Big[d^3\big(4L_I^2 T^4 + 14 L_H^2 L_I T^6 + 14 L_H^4 T^8\big) \\ &\qquad + d^2\big(35 L_H^2 T^2 + 8 L_I^2 T^4 + 28 L_H^2 L_I T^6 + 28 L_H^4 T^8\big) \\ &\qquad + d\big(16 L^2 + 4 L_H^2 T^2\big) + \big(2 d L_H^2 + L^2 T^{-2}\big)|x|^2 \\ &\qquad + \big(d L_I^2 + d L_H^2 L_I T^2 + d L_H^4 T^4 + L_H^2 T^{-2}\big)|x|^4 \Big]^{1/2} . \end{aligned} \tag{39}$$

Proof. Since $q_T : \mathbb{R}^{2d} \to \mathbb{R}^d$ is deterministic and measurable [30, Lemma 3],

$$\begin{aligned} \mathrm{TV}(\delta_x\pi, \delta_x\tilde{\pi}) &= \mathrm{TV}(\mathrm{Law}(q_T(x,\xi)), \mathrm{Law}(q_T(x,\Phi(\xi)))) \\ &\le \mathrm{TV}(\mathrm{Law}(\xi), \mathrm{Law}(\Phi(\xi))) \,, \end{aligned} \tag{40}$$

which gives the first inequality in (39). By Lemmas 15, 27 and 28, $\|D\Phi(v) - I_d\|_F^2 \le d\,\|D\Phi(v) - I_d\|^2$, and the Cauchy-Schwarz inequality we have

$$\begin{aligned} \mathrm{TV}(\mathrm{Law}(\xi), \mathrm{Law}(\Phi(\xi))^2 \;\le\; & \frac{1}{4} E\Big[|\Phi(\xi) - \xi|^2 + 2d\,\|D\Phi(\xi) - I_d\|^2\Big] \\ \le\; \frac{h^4}{4}\, E\Big[& 10dL^2 + \big(2dL_H^2 + L^2T^{-2}\big)|x|^2 + \big(6L^2 + 33dL_H^2T^2\big)|\xi|^2 \\ & + \big(dL_I^2 + L_H^2T^{-2} + dL_H^2L_IT^2 + dL_H^4T^4\big)|x|^4 \\ & + \big(L_H^2T^2 + 2dL_I^2T^4 + 7dL_H^2L_IT^6 + 7dL_H^4T^8\big)|\xi|^4 \Big] \\ \le\; \frac{h^4}{4}\Big[& d^3\big(4L_I^2T^4 + 14L_H^2L_IT^6 + 14L_H^4T^8\big) \\ & + d^2\big(35L_H^2T^2 + 8L_I^2T^4 + 28L_H^2L_IT^6 + 28L_H^4T^8\big) \\ & + d\big(16L^2 + 4L_H^2T^2\big) + \big(2dL_H^2 + L^2T^{-2}\big)|x|^2 \\ & + \big(dL_I^2 + dL_H^2L_IT^2 + dL_H^4T^4 + L_H^2T^{-2}\big)|x|^4 \Big] \end{aligned}$$

where in the last step we used $E[|\xi|^2] = d$ and $E[|\xi|^4] = 2d(d+2)$. Taking square roots gives the second inequality in (39). □

Lemma 18 implies the following bound on the TV distance between $\mu\pi$ and $\mu\tilde{\pi}$.

Lemma 19. *Suppose Assumption 2 holds. Then* $\mathrm{TV}(\mu\pi, \mu\tilde{\pi}) \le Ch^2$*, where* C *is defined in* (13).

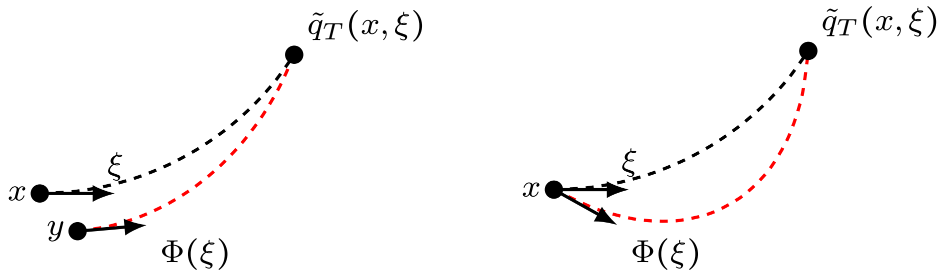
The proof of this result is similar to the proof of Lemma 17 and therefore omitted.

3.4. Successful Coupling for uHMC

The TV bound in Theorem 5 can be realized by a successful coupling for uHMC which combines a coupling that brings the two copies arbitrarily close together and the following one-shot coupling construction.

Let $\mathcal{U} \sim \mathrm{Unif}(0,1)$ and $\xi \sim \mathcal{N}(0, I_d)$ be independent random variables. For notational convenience, set

$$\alpha(v) = \min\Big(\frac{\varphi(\Phi(v))|\det(D\Phi(v))|}{\varphi(v)}, 1\Big)\,, \tag{41}$$



(a) $\tilde{q}_T(x,\xi) = \tilde{q}_T(y,\Phi(\xi))$ (b) $\tilde{q}_T(x,\xi) = q_T(x,\Phi(\xi))$

Figure 1. **One-shot couplings.** (a) To obtain TV convergence bounds for uHMC, the initial velocities are coupled such that $\tilde{q}_T(x,\xi) = \tilde{q}_T(y,\Phi(\xi))$ with maximal possible probability. (b) For TV accuracy of the invariant measure of uHMC, the initial velocities are coupled such that $\tilde{q}_T(x,\xi) = q_T(x,\Phi(\xi))$ with maximal possible probability.

where for any $x, y, v \in \mathbb{R}^d$ we define Φ via

$$\tilde{q}_T(x,v) = \tilde{q}_T(y,\Phi(v)) \ . \tag{42}$$

The transition step of the one-shot coupling is given by

$$\begin{cases} \tilde{X}(x,y) &= \tilde{q}_T(x,\xi), \\ \tilde{Y}(x,y) &= \tilde{q}_T(y,\eta), \end{cases} \qquad \text{where} \quad \eta = \begin{cases} \Phi(\xi) & \text{if } \mathcal{U} \le \alpha(\xi), \\ \tilde{\eta} & \text{otherwise,} \end{cases} \tag{43}$$

where $\tilde{\eta}$ is independent of ξ and $\mathcal{U}$ and satisfies

$$\mathrm{Law}(\tilde{\eta})(dv) = \frac{(\varphi(v) - \varphi(\Phi^{-1}(v))|\det(D\Phi^{-1}(v))|)^+}{1-\gamma} dv \tag{44}$$

where $\gamma = E[\alpha(\xi)]$. By definition of Φ, $\tilde{q}_T(x,\xi) = \tilde{q}_T(x,\Phi(\xi))$.

In order to verify that $(\tilde{X}(x,y), \tilde{Y}(x,y))$ is indeed a coupling of the transition probabilities $\tilde{\pi}(x,\cdot)$ and $\tilde{\pi}(y,\cdot)$, we remark that the distribution of η is $\mathcal{N}(0, I_d)$ since, by definition of η in (43) and a change of variables,

$$\begin{aligned} P[\eta \in B] &= E\left[I_B(\Phi(\xi))\,\alpha(\xi)\right] + E\left[I_B(\tilde{\eta})\right] E\left[1-\alpha(\xi)\right] \\ &= \int I_B(v)\, \min(\varphi(\Phi^{-1}(v))|\det(D\Phi^{-1}(v))|, \varphi(v))\, dv \\ &\quad + \int I_B(v)(\varphi(v) - \varphi(\Phi^{-1}(v))|\det(D\Phi^{-1}(v))|)^+ \, dv \\ &= \int I_B(v)\, \varphi(v)\, dv \ = \ P[\xi \in B] \end{aligned}$$

for any measurable set B. As a byproduct of this calculation, note also that

$$\begin{aligned} P[\eta \ne \Phi(\xi)] &= \int \left(\varphi(v) - \varphi(\Phi^{-1}(v))|\det(D\Phi^{-1}(v))|\right)^+ dv \\ &= \mathrm{TV}(\mathrm{Law}(\xi), \mathrm{Law}(\Phi(\xi))) \ . \end{aligned}$$

This calculation shows that the one-shot coupling in (43) ensures that $\eta = \Phi(\xi)$ with maximal possible probability, and with remaining probability, $\eta = \tilde{\eta}$ with $\tilde{\eta} \perp \xi$. Thus, the coupling coalesces in one step whenever $\eta = \Phi(\xi)$, as illustrated in Figure 1(a).

4. A priori estimates

Here we state several bounds for the dynamics (7) that are crucial in the proof of our main results. Throughout this section, we fix a duration parameter $T > 0$ and a time step size $h \geq 0$ such that $T/h \in \mathbb{N}$ for $h \neq 0$ and

$$L(T^2 + Th) \leq 1/6\,. \tag{45}$$

4.1. A priori estimates for the dynamics

Lemma 20. *Suppose that Assumption 2 (A1)-(A2) hold. For any $x, y, u, v \in \mathbb{R}^d$,*

$$\max_{s\leq T} |\tilde{q}_s(x,v)| \leq (1 + L(T^2 + Th)) \max(|x|, |x + Tv|), \tag{46}$$

$$\max_{s\leq T} |\tilde{v}_s(x,v)| \leq |v| + LT(1 + L(T^2 + Th)) \max(|x|, |x + Tv|), \tag{47}$$

$$\begin{aligned} &\max_{s\leq T} |\tilde{q}_s(x,u) - \tilde{q}_s(y,v)| \\ &\quad \leq (1 + L(T^2 + Th)) \max\left(|x - y|, |x - y + T(u - v)|\right). \end{aligned} \tag{48}$$

Lemma 20 is contained in Lemmas 3.1 and 3.2 of [5] and hence a proof is omitted.

4.2. A priori estimates for velocity derivative flow

The next lemma provides similar bounds for $D_2\tilde{q}_T(x,u) := \partial\tilde{q}_T(x,u)/\partial u$.

Lemma 21. *Suppose that Assumption 2 (A1)-(A3) hold. For any $x, y, u, v \in \mathbb{R}^d$,*

$$\max_{s\leq T} \|D_2\tilde{q}_s(x,v)\| \leq (6/5)\, T\,, \tag{49}$$

$$\max_{s\leq T} \|D_2\tilde{v}_s(x,v)\| \leq (6/5)\,, \quad \textit{and} \tag{50}$$

$$\begin{aligned} &\max_{s\leq T} \|D_2\tilde{q}_s(x,u) - D_2\tilde{q}_s(y,v)\| \\ &\quad \leq L_H (6/5)^2 T^3 (1 + L(T^2 + Th)) \max\left(|x - y|, |(x - y) + T(u - v)|\right). \end{aligned} \tag{51}$$

Proof. From (7), note that $D_2\tilde{q}_t(x,u)$ and $D_2\tilde{v}_t(x,u)$ satisfy:

$$\begin{aligned} D_2\tilde{q}_T(x,u) &= T I_d - \frac{1}{2}\int_0^T (T - s)\left[\mathbf{H}(\tilde{q}_{\lfloor s\rfloor_h}) D_2\tilde{q}_{\lfloor s\rfloor_h} + \mathbf{H}(\tilde{q}_{\lceil s\rceil_h}) D_2\tilde{q}_{\lceil s\rceil_h}\right] ds \\ &\quad + \frac{1}{2}\int_0^T (s - \lfloor s\rfloor_h)\left[\mathbf{H}(\tilde{q}_{\lceil s\rceil_h}) D_2\tilde{q}_{\lceil s\rceil_h} - \mathbf{H}(\tilde{q}_{\lfloor s\rfloor_h}) D_2\tilde{q}_{\lfloor s\rfloor_h}\right] ds\,, \end{aligned} \tag{52}$$

$$D_2\tilde{v}_T(x,u) = I_d - \frac{1}{2}\int_0^T \left[\mathbf{H}(\tilde{q}_{\lfloor s\rfloor_h}) D_2\tilde{q}_{\lfloor s\rfloor_h} + \mathbf{H}(\tilde{q}_{\lceil s\rceil_h}) D_2\tilde{q}_{\lceil s\rceil_h}\right] ds\,.$$

Applying Assumption 2 (A2) and inserting $LT^2 \le 1/6$, we obtain

$$\begin{aligned}\max_{s\le T}\|D_2\tilde{q}_s(x,u)\| &\le T\,\|I_d\| + T^2\max_{s\le T}\|\mathbf{H}(\tilde{q}_s)D_2\tilde{q}_s(x,u)\|\\ &\le T + LT^2\max_{s\le T}\|D_2\tilde{q}_s(x,u)\| \le T + (1/6)\max_{s\le T}\|D_2\tilde{q}_s(x,u)\| \le (6/5)T\,,\\ \max_{s\le T}\|D_2\tilde{v}_s(x,u)\| &\le 1 + LT\max_{s\le T}\|D_2\tilde{q}_s(x,u)\| \le 1 + LT^2(6/5) \le 6/5\,,\end{aligned}$$

which gives (49) and (50). For (51), it is notationally convenient to introduce the shorthand $\tilde{q}_s^{(1)} := \tilde{q}_s(x,u)$ and $\tilde{q}_s^{(2)} := \tilde{q}_s(y,v)$ for any $s \in [0,T]$. Then from (52)

$$\begin{aligned}D_2\tilde{q}_T^{(1)} - D_2\tilde{q}_T^{(2)} &= -\frac{1}{2}\int_0^T (T-s)\left[\mathbf{H}(\tilde{q}^{(1)}_{\lfloor s\rfloor_h})D_2\tilde{q}^{(1)}_{\lfloor s\rfloor_h} - \mathbf{H}(\tilde{q}^{(2)}_{\lfloor s\rfloor_h})D_2\tilde{q}^{(2)}_{\lfloor s\rfloor_h}\right]ds\\ &\quad-\frac{1}{2}\int_0^T (T-s)\left[\mathbf{H}(\tilde{q}^{(1)}_{\lceil s\rceil_h})D_2\tilde{q}^{(1)}_{\lceil s\rceil_h} - \mathbf{H}(\tilde{q}^{(2)}_{\lceil s\rceil_h})D_2\tilde{q}^{(2)}_{\lceil s\rceil_h}\right]ds\\ &\quad+\frac{1}{2}\int_0^T (s-\lfloor s\rfloor_h)\left[\mathbf{H}(\tilde{q}^{(1)}_{\lceil s\rceil_h})D_2\tilde{q}^{(1)}_{\lceil s\rceil_h} - \mathbf{H}(\tilde{q}^{(2)}_{\lceil s\rceil_h})D_2\tilde{q}^{(2)}_{\lceil s\rceil_h}\right]ds\\ &\quad+\frac{1}{2}\int_0^T (s-\lfloor s\rfloor_h)\left[\mathbf{H}(\tilde{q}^{(2)}_{\lfloor s\rfloor_h})D_2\tilde{q}^{(2)}_{\lfloor s\rfloor_h} - \mathbf{H}(\tilde{q}^{(1)}_{\lfloor s\rfloor_h})D_2\tilde{q}^{(1)}_{\lfloor s\rfloor_h}\right]ds\\ =& \int_0^T \frac{T-s}{2}\left[\mathbf{H}(\tilde{q}^{(2)}_{\lfloor s\rfloor_h})(D_2\tilde{q}^{(2)}_{\lfloor s\rfloor_h} - D_2\tilde{q}^{(1)}_{\lfloor s\rfloor_h}) - (\mathbf{H}(\tilde{q}^{(1)}_{\lfloor s\rfloor_h}) - \mathbf{H}(\tilde{q}^{(2)}_{\lfloor s\rfloor_h}))D_2\tilde{q}^{(1)}_{\lfloor s\rfloor_h}\right]ds\\ +& \int_0^T \frac{T-s}{2}\left[\mathbf{H}(\tilde{q}^{(2)}_{\lceil s\rceil_h})(D_2\tilde{q}^{(2)}_{\lceil s\rceil_h} - D_2\tilde{q}^{(1)}_{\lceil s\rceil_h}) - (\mathbf{H}(\tilde{q}^{(1)}_{\lceil s\rceil_h}) - \mathbf{H}(\tilde{q}^{(2)}_{\lceil s\rceil_h}))D_2\tilde{q}^{(1)}_{\lceil s\rceil_h}\right]ds\\ -& \int_0^T \frac{s-\lfloor s\rfloor_h}{2}\left[\mathbf{H}(\tilde{q}^{(2)}_{\lceil s\rceil_h})(D_2\tilde{q}^{(2)}_{\lceil s\rceil_h} - D_2\tilde{q}^{(1)}_{\lceil s\rceil_h}) - (\mathbf{H}(\tilde{q}^{(1)}_{\lceil s\rceil_h}) - \mathbf{H}(\tilde{q}^{(2)}_{\lceil s\rceil_h}))D_2\tilde{q}^{(1)}_{\lceil s\rceil_h}\right]ds\\ -& \int_0^T \frac{s-\lfloor s\rfloor_h}{2}\left[\mathbf{H}(\tilde{q}^{(1)}_{\lfloor s\rfloor_h})(D_2\tilde{q}^{(1)}_{\lfloor s\rfloor_h} - D_2\tilde{q}^{(2)}_{\lfloor s\rfloor_h}) - (\mathbf{H}(\tilde{q}^{(2)}_{\lfloor s\rfloor_h}) - \mathbf{H}(\tilde{q}^{(1)}_{\lfloor s\rfloor_h}))D_2\tilde{q}^{(2)}_{\lfloor s\rfloor_h}\right]ds\,.\end{aligned}$$

By using Assumption 2 (A2), Assumption 2 (A3), and (49), we then get

$$\begin{aligned}&\max_{s\le T}\left\|D_2\tilde{q}_s^{(1)} - D_2\tilde{q}_s^{(2)}\right\|\\ &\qquad\le LT^2\max_{s\le T}\left\|D^2\tilde{q}_s^{(1)} - D^2\tilde{q}_s^{(2)}\right\| + L_H(6/5)T^3\max_{s\le T}\left|\tilde{q}_s^{(1)} - \tilde{q}_s^{(2)}\right|\,.\end{aligned}$$

Inserting $LT^2 \le 1/6$ and (48), and simplifying yields (51). □

4.3. Discretization error bounds for Verlet

In this part we gather standard estimates of the error between the exact solution $q_T(x,v)$ and Verlet $\tilde{q}_T(x,v)$. These estimates are a key ingredient in quantifying the invariant measure accuracy in TV distance of uHMC.

To this end, the following error estimate for a variant of the trapezoidal rule is useful because (7) involves this particular trapezoidal rule approximation.

Lemma 22. *Fix $T > 0$. Let $(\mathbf{V}, \|\cdot\|)$ be a normed space. Let $f : [0,T] \to \mathbf{V}$ be a twice differentiable function such that $\max_{s\in[0,T]} \|f'(s)\| \le B_1$ and $\max_{s\in[0,T]} \|f''(s)\| \le B_2$. Then for any $h \ge 0$ such that $T/h \in \mathbb{N}$ for $h \neq 0$*

$$\begin{aligned}&\left\| \int_0^T (T-s) f(s) ds - \frac{1}{2}\int_0^T (T-s)\left[f(\lfloor s \rfloor_h) + f(\lceil s \rceil_h)\right] ds \right\| \\ &\qquad \le \frac{h^2}{12}(B_2 T^2 + B_1 T)\,.\end{aligned} \tag{53}$$

Proof. The error over $[t_k, t_{k+1}]$ is given by

$$\epsilon_k \ := \ \left\| \int_{t_k}^{t_{k+1}} (T-s) f(s) ds - \frac{1}{2}\int_{t_k}^{t_{k+1}} (T-s)\left[f(\lfloor s \rfloor_h) + f(\lceil s \rceil_h)\right] ds \right\| \,.$$

By integrating by parts,

$$\epsilon_k \ = \ \left\| \int_{t_k}^{t_{k+1}} \left[\frac{s^2}{2} - Ts - \alpha_k \right] f'(s) ds \right\|$$

where $\alpha_k = (1/4)(t_{k+1}^2 - 2Tt_{k+1} - 2Tt_k + t_k^2)$. A second integration by parts gives

$$\epsilon_k = \left\| \int_{t_k}^{t_{k+1}} \left[\frac{s^3}{6} - \frac{Ts^2}{2} - \alpha_k s + \beta_k \right] f''(s) ds - \frac{1}{12} f'(t_k) h^3 \right\| \le \frac{1}{12}(B_2 T + B_1) h^3$$

where $\beta_k = (1/12)(-6Tt_k + 3t_k^2 + t_{k+1}^2) t_{k+1}$ and in the last step we used

$$\int_{t_k}^{t_{k+1}} \left| \frac{s^3}{6} - \frac{Ts^2}{2} - \alpha_k s + \beta_k \right| ds = \frac{1}{12}(T - t_k) h^3 \le \frac{1}{12} T h^3 \,.$$

Summing these errors over the T/h subintervals gives the upper bound in (53). □

Lemma 23. *Suppose Assumption 2 (A1)-(A3) hold. For any $x, v \in \mathbb{R}^d$,*

$$\begin{aligned}&\max_{s\le T} |q_s(x,v) - \tilde{q}_s(x,v)| \\ &\qquad \le h^2 \Big(\frac{7}{45} L\,|x| + \frac{1547}{1800} LT\,|v| + \frac{1}{120} L_H\,|x|^2 + \frac{3}{10} L_H T^2\,|v|^2 \Big)\,.\end{aligned} \tag{54}$$

Proof. By (46) and (47) with $h = 0$, and since $LT^2 \le 1/6$, note that

$$\max_{s\le T} |q_s| \ \le \ (7/6)\,(|x| + T\,|v|)\,, \tag{55}$$

$$\max_{s\le T} |v_s| \ \le \ |v| + (7/6)\,LT\,(|x| + T\,|v|) \ \le \ (7/6)\,LT\,|x| + (6/5)\,|v|\,, \tag{56}$$

$$\max_{s\le T} |v_s|^2 \ \le \ 2(7/6)^2 L^2 T^2 \,|x|^2 + 2(6/5)^2\,|v|^2 \ \le \ (1/2) L\,|x|^2 + 3\,|v|^2\,. \tag{57}$$

Introduce the shorthand $q_T := q_T(x,v)$ and $\tilde{q}_T := \tilde{q}_T(x,v)$. Using (7), note that

$$
\begin{aligned}
q_T - \tilde{q}_T &= \mathrm{I} + \mathrm{II} + \mathrm{III} \quad \text{where} && (58)\\
\mathrm{I} &:= \frac{1}{2}\int_0^T (T-s)[\nabla U(\tilde{q}_{\lfloor s\rfloor_h}) - \nabla U(q_{\lfloor s\rfloor_h}) + \nabla U(\tilde{q}_{\lceil s\rceil_h}) - \nabla U(q_{\lceil s\rceil_h})]ds \\
&\quad - \frac{1}{2}\int_0^T (s-\lfloor s\rfloor_h)[\nabla U(\tilde{q}_{\lceil s\rceil_h}) - \nabla U(q_{\lceil s\rceil_h}) - (\nabla U(\tilde{q}_{\lfloor s\rfloor_h}) - \nabla U(q_{\lfloor s\rfloor_h}))]ds\,, \\
\mathrm{II} &:= -\frac{1}{2}\int_0^T (s-\lfloor s\rfloor_h)[\nabla U(q_{\lceil s\rceil_h}) - \nabla U(q_{\lfloor s\rfloor_h})]ds\,, \quad \text{and} \\
\mathrm{III} &:= -\int_0^T (T-s)\nabla U(q_s)ds + \frac{1}{2}\int_0^T (T-s)[\nabla U(q_{\lfloor s\rfloor_h}) + \nabla U(q_{\lceil s\rceil_h})]ds\,.
\end{aligned}
$$

By Assumption 2 (A2) and the condition $LT^2 \le 1/6$,

$$
\begin{aligned}
|\mathrm{I}| &\le LT^2 \max_{s\le T}|q_s - \tilde{q}_s| \le (1/6)\max_{s\le T}|q_s - \tilde{q}_s|\,, && (59)\\
|\mathrm{II}| &\le (1/2)hL\int_0^T \left|q_{\lceil s\rceil_h} - q_{\lfloor s\rfloor_h}\right|ds = (1/2)hL\int_0^T \left|\int_{\lfloor s\rfloor_h}^{\lceil s\rceil_h} v_r dr\right| ds\,, \\
&\le (1/2)Lh^2 T\max_{s\le T}|v_s| \le Lh^2\left((7/72)\,|x| + (3/5)\,T\,|v|\right), && (60)
\end{aligned}
$$

where for (60) we used (56). Next, we apply Lemma 22 with $f(s) = -\nabla U(q_s)$ and

$$
\begin{aligned}
&|f'(s)| = |\mathbf{H}(q_s)v_s| \le L\,|v_s| \le L(LT(7/6)|x| + (6/5)|v|) =: B_1\,, \\
&|f''(s)| = \left|-D^3U(q_s)(v_s,v_s) + \mathbf{H}(q_s)\nabla U(q_s)\right|\,, \\
&\quad\le L_H\,|v_s|^2 + L^2\,|q_s| \le L_H(L(1/2)\,|x|^2 + 3\,|v|^2) + L^2(7/6)(|x| + T\,|v|) =: B_2\,,
\end{aligned}
$$

where we used Assumption 2 (A2), (A3) and $LT^2 \le 1/6$. Thus,

$$
|\mathrm{III}| \le \frac{h^2}{12}\left(\frac{7}{18}L\,|x| + \frac{251}{180}LT\,|v| + \frac{1}{12}L_H\,|x|^2 + 3L_H T^2\,|v|^2)\right). \qquad (61)
$$

Inserting (59), (60) and (61) into (58) and simplifying gives (54). □

Lemma 24. *Suppose Assumption 2 (A1)-(A4) hold. For any $x, v \in \mathbb{R}^d$,*

$$
\begin{aligned}
\max_{s\le T}\|D_2 q_s(x,v) - D_2\tilde{q}_s(x,v)\| &\le h^2\Big(\frac{43}{50}LT + \frac{1183}{4500}L_H T\,|x| + \frac{1847}{1250}L_H T^2\,|v| \\
&\quad + (\frac{3}{250}L_H^2 T^3 + \frac{1}{100}L_I T)\,|x|^2 + (\frac{54}{125}L_H^2 T^5 + \frac{9}{25}L_I T^3)\,|v|^2\Big)\,. \qquad (62)
\end{aligned}
$$

Proof. Using (7), write the difference as

$$
\begin{aligned}
D_2 q_T(x,v) - D_2\tilde{q}_T(x,v) &= \mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV} \quad \text{where} \qquad (63)\\
\mathrm{I} &:= \frac{1}{2}\int_0^T (T-s)\mathbf{H}(\tilde{q}_{\lfloor s\rfloor_h})(D_2\tilde{q}_{\lfloor s\rfloor_h} - D_2 q_{\lfloor s\rfloor_h})ds \\
&\quad + \frac{1}{2}\int_0^T (T-s)\mathbf{H}(\tilde{q}_{\lceil s\rceil_h})(D_2\tilde{q}_{\lceil s\rceil_h} - D_2 q_{\lceil s\rceil_h})ds \\
&\quad - \frac{1}{2}\int_0^T (s-\lfloor s\rfloor_h)\mathbf{H}(\tilde{q}_{\lceil s\rceil_h})(D_2\tilde{q}_{\lceil s\rceil_h} - D_2 q_{\lceil s\rceil_h})ds \\
&\quad + \frac{1}{2}\int_0^T (s-\lfloor s\rfloor_h)\mathbf{H}(\tilde{q}_{\lfloor s\rfloor_h})(D_2\tilde{q}_{\lfloor s\rfloor_h} - D_2 q_{\lfloor s\rfloor_h})ds \\
\mathrm{II} &:= \frac{1}{2}\int_0^T (T-s)(\mathbf{H}(\tilde{q}_{\lfloor s\rfloor_h}) - \mathbf{H}(q_{\lfloor s\rfloor_h}))D_2 q_{\lfloor s\rfloor_h} ds \\
&\quad + \frac{1}{2}\int_0^T (T-s)(\mathbf{H}(\tilde{q}_{\lceil s\rceil_h}) - \mathbf{H}(q_{\lceil s\rceil_h}))D_2 q_{\lceil s\rceil_h} ds \\
&\quad - \frac{1}{2}\int_0^T (s-\lfloor s\rfloor_h)(\mathbf{H}(\tilde{q}_{\lceil s\rceil_h}) - \mathbf{H}(q_{\lceil s\rceil_h}))D_2 q_{\lceil s\rceil_h} ds \\
&\quad + \frac{1}{2}\int_0^T (s-\lfloor s\rfloor_h)(\mathbf{H}(\tilde{q}_{\lfloor s\rfloor_h}) - \mathbf{H}(q_{\lfloor s\rfloor_h}))D_2 q_{\lfloor s\rfloor_h} ds \\
\mathrm{III} &:= -\frac{1}{2}\int_0^T (s-\lfloor s\rfloor_h)[(\mathbf{H}(q_{\lceil s\rceil_h}) - \mathbf{H}(q_{\lfloor s\rfloor_h}))D_2 q_{\lceil s\rceil_h}] \\
&\quad - \frac{1}{2}\int_0^T (s-\lfloor s\rfloor_h)[\mathbf{H}(q_{\lfloor s\rfloor_h})(D_2 q_{\lceil s\rceil_h} - D_2 q_{\lfloor s\rfloor_h})]ds\,, \quad \text{and} \\
\mathrm{IV} &:= -\int_0^T (T-s)\mathbf{H}(q_s)D_2 q_s ds \\
&\quad + \frac{1}{2}\int_0^T (T-s)[\mathbf{H}(q_{\lfloor s\rfloor_h})D_2 q_{\lfloor s\rfloor_h} + \mathbf{H}(q_{\lceil s\rceil_h})D_2 q_{\lceil s\rceil_h}]ds\,.
\end{aligned}
$$

By Assumptions 2 (A2) and (A3), and the condition $LT^2 \le 1/6$,

$$
\begin{aligned}
\|\mathrm{I}\| &\le LT^2 \max_{s\le T}\|D_2 q_s - D_2\tilde{q}_s\| \le (1/6)\max_{s\le T}\|D_2 q_s - D_2\tilde{q}_s\|\,, \qquad (64)\\
\|\mathrm{II}\| &\le L_H T^2 \max_{s\le T}|q_s - \tilde{q}_s| \max_{s\le T}\|D_2 q_s\| \le (6/5)L_H T^3 \max_{s\le T}|q_s - \tilde{q}_s|\,, \qquad (65)\\
\|\mathrm{III}\| &\le \frac{hL_H}{2}\max_{s\le T}\|D_2 q_s\|\int_0^T\left|\int_{\lfloor s\rfloor_h}^{\lceil s\rceil_h} v_r dr\right| ds + \frac{hL}{2}\int_0^T\left\|\int_{\lfloor s\rfloor_h}^{\lceil s\rceil_h} D_2 v_r dr\right\| ds\,, \\
&\le \frac{h^2 L_H T^2}{2}\frac{6}{5}\max_{s\le T}|v_s| + \frac{h^2 LT}{2}\max_{s\le T}\|D_2 v_s\|\,, \\
&\le \frac{3}{5}h^2 LT + h^2 L_H\left(\frac{7}{60}T|x| + \frac{18}{25}T^2|v|\right)\,, \qquad (66)
\end{aligned}
$$

where for (65) we used (49) and for (66) we used (56) and (50). Next, we apply Lemma 22 with

$f(s) = -\mathbf{H}(q_s)D_2 q_s$ and

$$\begin{aligned}
\|f'(s)\| &= \left\|(D^3U(q_s)v_s)D_2q_s + \mathbf{H}(q_s)D_2v_s\right\| \le L_H|v_s|\,\|D_2q_s\| + L\,\|D_2v_s\| \;, \\
&\le \frac{6}{5}(L_H(\frac{7}{36}|x| + \frac{6}{5}T|v|) + L) \;=:\; B_1 \;, \\
\|f''(s)\| &= \left\|(D^4U(v_s,v_s) - D^3U\nabla U - \mathbf{H}^2)D_2q_s + 2(D^3Uv_s)D_2v_s\right\| \;, \\
&\le (L_I|v_s|^2 + L_HL|q_s| + L^2)\,\|D_2q_s\| + 2L_H|v_s|\,\|D_2v_s\| \;, \\
&\le [L_I(\frac{L}{2}|x|^2 + 3|v|^2) + L_HL\frac{7}{6}(|x| + T|v|) + L^2]\frac{6}{5}T + L_H[\frac{7}{3}LT|x| + \frac{12}{5}|v|] =: B_2 \;,
\end{aligned}$$

where we used $LT^2 \le 1/6$, Assumption 2, (56), (57), (49) and (50). Thus,

$$\|\mathbb{IV}\| \le \frac{h^2}{12}\left(\frac{7}{5}LT + \frac{77}{90}L_HT|x| + \frac{611}{150}L_HT^2|v| + \frac{1}{10}L_IT|x|^2 + \frac{18}{5}L_IT^3|v|^2\right). \tag{67}$$

Inserting (64), (65), (66), (67) and Lemma 23 into (63) and simplifying gives (62). □

5. One-shot Coupling Bounds

Here we prove bounds related to the one-shot coupling. Throughout this section, we fix a duration parameter $T > 0$ and a time step size $h \ge 0$ such that $T/h \in \mathbb{N}$ for $h \ne 0$ and

$$L(T^2 + Th) \;\le\; 1/6\,. \tag{68}$$

5.1. One-shot for $\tilde{q}_T(x,\xi) = \tilde{q}_T(y,\Phi(\xi))$

By Corollary 14, for any $x, y, v \in \mathbb{R}^d$, there exists a unique minimum $(\tilde{q}_t^\star)_{0\le t\le T}$ of the action sum satisfying the endpoint conditions $\tilde{q}_0^\star = y$ and $\tilde{q}_T^\star = \tilde{q}_T(x,v)$. In terms of this minimum, we introduce a function $\Phi : \mathbb{R}^d \to \mathbb{R}^d$ defined as $\Phi(v) := -D_1L_h(\tilde{q}_0^\star, \tilde{q}_h^\star)$, which by definition satisfies $\tilde{q}_T(x,v) = \tilde{q}_T^\star(y,\Phi(v))$. Here we develop some bounds for $\Phi(v)$ and $D\Phi(v)$.

Lemma 25. *Suppose that Assumption 2 (A1)-(A2) hold. Then for any $x, y, v \in \mathbb{R}^d$ such that $\tilde{q}_T(x,v) = \tilde{q}_T(y,\Phi(v))$, we have*

$$T\,|\Phi(v) - v| \;\le\; (3/2)\;|x - y| \;. \tag{69}$$

Proof. Let $u = \Phi(v)$. Integrating (7) yields

$$\begin{aligned}
\tilde{q}_T(x,u) = x + Tu - \frac{1}{2}\int_0^T (T-s)\left[\nabla U(\tilde{q}_{\lfloor s\rfloor_h}) + \nabla U(\tilde{q}_{\lceil s\rceil_h})\right]ds \\
+ \frac{1}{2}\int_0^T (s - \lfloor s\rfloor_h)\left[\nabla U(\tilde{q}_{\lceil s\rceil_h}) - \nabla U(\tilde{q}_{\lfloor s\rfloor_h})\right]ds
\end{aligned}$$

and since $\tilde{q}_T(x,u) = \tilde{q}_T(y,v)$, we obtain

$$\begin{aligned} T\,|u-v| &\le |x-y| + LT^2 \max_{s\le T} |\tilde{q}_s(x,u) - \tilde{q}_s(y,v)| \\ &\le |x-y| + LT^2(1+L(T^2+Th))\max\left(|x-y|\,,|(x-y)+T(u-v)|\right) \\ &\le |x-y| + (7/36)(|x-y| + T\,|u-v|)\;, \end{aligned}$$

where in the second to last step we used (48) from Lemma 20 and in the last step we used $L(T^2+Th) \le 1/6$. Simplifying and using $(1+7/36)/(1-7/36) < 3/2$ gives (69). □

Lemma 26. *Suppose that Assumption 2 (A1)-(A3) hold. Then for any $x,y,v \in \mathbb{R}^d$ such that $\tilde{q}_T(x,v) = \tilde{q}_T(y,\Phi(v))$, we have*

$$\|D\Phi(v) - I_d\| \;\le\; (1/2)\min(1, 11\, L_H T^2\,|x-y|)\;. \tag{70}$$

Proof. Introduce the shorthand $\tilde{q}_s^{(1)} := \tilde{q}_s(x,v)$ and $\tilde{q}_s^{(2)} := \tilde{q}_s(y,\Phi(v))$ for any $s \in [0,T]$. Differentiating both sides of $\tilde{q}_T(x,v) = \tilde{q}_T(y,\Phi(v))$ with respect to v yields

$$\begin{aligned} &T(D\Phi(v) - I_d) = \\ &+ \frac{1}{2}\int_0^T (T-s)\left[\mathbf{H}(\tilde{q}^{(2)}_{\lfloor s\rfloor_h})D_2\tilde{q}^{(2)}_{\lfloor s\rfloor_h} - \mathbf{H}(\tilde{q}^{(1)}_{\lfloor s\rfloor_h})D_2\tilde{q}^{(1)}_{\lfloor s\rfloor_h}\right]ds \\ &+ \frac{1}{2}\int_0^T (T-s)\left[\mathbf{H}(\tilde{q}^{(2)}_{\lceil s\rceil_h})D_2\tilde{q}^{(2)}_{\lceil s\rceil_h} - \mathbf{H}(\tilde{q}^{(1)}_{\lceil s\rceil_h})D_2\tilde{q}^{(1)}_{\lceil s\rceil_h}\right]ds \\ &- \frac{1}{2}\int_0^T (s-\lfloor s\rfloor_h)\left[\mathbf{H}(\tilde{q}^{(2)}_{\lceil s\rceil_h})D_2\tilde{q}^{(2)}_{\lceil s\rceil_h} - \mathbf{H}(\tilde{q}^{(1)}_{\lceil s\rceil_h})D_2\tilde{q}^{(1)}_{\lceil s\rceil_h}\right]ds \\ &- \frac{1}{2}\int_0^T (s-\lfloor s\rfloor_h)\left[\mathbf{H}(\tilde{q}^{(1)}_{\lfloor s\rfloor_h})D_2\tilde{q}^{(1)}_{\lfloor s\rfloor_h} - \mathbf{H}(\tilde{q}^{(2)}_{\lfloor s\rfloor_h})D_2\tilde{q}^{(2)}_{\lfloor s\rfloor_h}\right]ds \\ &+ \frac{1}{2}\int_0^T (T-s)\left[\mathbf{H}(\tilde{q}^{(2)}_{\lfloor s\rfloor_h})D_2\tilde{q}^{(2)}_{\lfloor s\rfloor_h} + \mathbf{H}(\tilde{q}^{(2)}_{\lceil s\rceil_h})D_2\tilde{q}^{(2)}_{\lceil s\rceil_h}\right](D\Phi(v) - I_d)ds \\ &- \frac{1}{2}\int_0^T (s-\lfloor s\rfloor_h)\left[\mathbf{H}(\tilde{q}^{(2)}_{\lceil s\rceil_h})D_2\tilde{q}^{(2)}_{\lceil s\rceil_h} - \mathbf{H}(\tilde{q}^{(2)}_{\lfloor s\rfloor_h})D_2\tilde{q}^{(2)}_{\lfloor s\rfloor_h}\right](D\Phi(v) - I_d)ds\;. \end{aligned} \tag{71}$$

By using Assumption 2 (A2), (49), and $LT^2 \le 1/6$, note that

$$\|D\Phi(v) - I_d\| \le (6/5)2LT^2/(1-(6/5)LT^2) \le 1/2\;. \tag{72}$$

We can also rewrite (71) as

$$
\begin{aligned}
&T(D\Phi(v) - I_d) = \\
&+\frac{1}{2}\int_0^T (T-s)\Big[\mathbf{H}(\tilde{q}^{(2)}_{\lfloor s\rfloor_h})(D_2\tilde{q}^{(2)}_{\lfloor s\rfloor_h} - D_2\tilde{q}^{(1)}_{\lfloor s\rfloor_h}) - (\mathbf{H}(\tilde{q}^{(1)}_{\lfloor s\rfloor_h}) - \mathbf{H}(\tilde{q}^{(2)}_{\lfloor s\rfloor_h}))D_2\tilde{q}^{(1)}_{\lfloor s\rfloor_h}\Big]ds \\
&+\frac{1}{2}\int_0^T (T-s)\Big[\mathbf{H}(\tilde{q}^{(2)}_{\lceil s\rceil_h})(D_2\tilde{q}^{(2)}_{\lceil s\rceil_h} - D_2\tilde{q}^{(1)}_{\lceil s\rceil_h}) - (\mathbf{H}(\tilde{q}^{(1)}_{\lceil s\rceil_h}) - \mathbf{H}(\tilde{q}^{(2)}_{\lceil s\rceil_h}))D_2\tilde{q}^{(1)}_{\lceil s\rceil_h}\Big]ds \\
&-\frac{1}{2}\int_0^T (s-\lfloor s\rfloor_h)\Big[\mathbf{H}(\tilde{q}^{(2)}_{\lceil s\rceil_h})(D_2\tilde{q}^{(2)}_{\lceil s\rceil_h} - D_2\tilde{q}^{(1)}_{\lceil s\rceil_h}) - (\mathbf{H}(\tilde{q}^{(1)}_{\lceil s\rceil_h}) - \mathbf{H}(\tilde{q}^{(2)}_{\lceil s\rceil_h}))D_2\tilde{q}^{(1)}_{\lceil s\rceil_h}\Big]ds \\
&+\frac{1}{2}\int_0^T (s-\lfloor s\rfloor_h)\Big[\mathbf{H}(\tilde{q}^{(2)}_{\lfloor s\rfloor_h})(D_2\tilde{q}^{(2)}_{\lfloor s\rfloor_h} - D_2\tilde{q}^{(1)}_{\lfloor s\rfloor_h}) - (\mathbf{H}(\tilde{q}^{(1)}_{\lfloor s\rfloor_h}) - \mathbf{H}(\tilde{q}^{(2)}_{\lfloor s\rfloor_h}))D_2\tilde{q}^{(1)}_{\lfloor s\rfloor_h}\Big]ds \\
&+\frac{1}{2}\int_0^T (T-s)\Big[\mathbf{H}(\tilde{q}^{(2)}_{\lfloor s\rfloor_h})D_2\tilde{q}^{(2)}_{\lfloor s\rfloor_h} + \mathbf{H}(\tilde{q}^{(2)}_{\lceil s\rceil_h})D_2\tilde{q}^{(2)}_{\lceil s\rceil_h}\Big](D\Phi(v) - I_d)ds \\
&-\frac{1}{2}\int_0^T (s-\lfloor s\rfloor_h)\Big[\mathbf{H}(\tilde{q}^{(2)}_{\lceil s\rceil_h})D_2\tilde{q}^{(2)}_{\lceil s\rceil_h} - \mathbf{H}(\tilde{q}^{(2)}_{\lfloor s\rfloor_h})D_2\tilde{q}^{(2)}_{\lfloor s\rfloor_h}\Big](D\Phi(v) - I_d)ds\,.
\end{aligned}
$$

By using Assumption 2 (A2), (49), Assumption 2 (A3), (51), and $LT^2 \le 1/6$, we get

$$
\begin{aligned}
(4/5)\,T\,\|D\Phi(v) - I_d\| &\le (1 - LT^2(6/5))\,T\,\|D\Phi(v) - I_d\| \\
&\le LT^2 \max_{s\le T}\left\|D^2\tilde{q}^{(1)}_s - D^2\tilde{q}^{(2)}_s\right\| + L_H(6/5)T^3\max_{s\le T}\left|\tilde{q}^{(1)}_s - \tilde{q}^{(2)}_s\right| \\
&\le (42/25)\,L_H T^3\,(|x-y| + T\,|\Phi(v) - v|)\,.
\end{aligned}
$$

Inserting (69), using $21/4 < 11/2$, simplifying and inserting (72) gives (70). □

5.2. One-shot for $\tilde{q}_T(x,\xi) = q_T(x,\Phi(\xi))$

By Corollary 14, for any $x, v \in \mathbb{R}^d$, there exists a unique minimum $(q^\star_t)_{0\le t\le T}$ of the action integral satisfying the endpoint conditions $q^\star_0 = x$ and $q^\star_T = \tilde{q}_T(x,v)$. In terms of this minimum, we introduce a function $\Phi : \mathbb{R}^d \to \mathbb{R}^d$ defined as $\Phi(v) := v^\star_0$, which by definition satisfies $\tilde{q}_T(x,v) = q^\star_T(x,\Phi(v))$. Here we develop some bounds for $\Phi(v)$ and $D\Phi(v)$.

Lemma 27. *Suppose that Assumption 2 (A1)-(A3) hold. Then for any $x, v \in \mathbb{R}^d$ such that $\tilde{q}_T(x,v) = q_T(x,\Phi(v))$, we have*

$$
T\,|\Phi(v) - v| \;\le\; \frac{7}{6}h^2\Big(\frac{7}{45}L\,|x| + \frac{1547}{1800}LT\,|v| + \frac{1}{120}L_H\,|x|^2 + \frac{3}{10}L_H T^2\,|v|^2\Big)\,. \tag{73}
$$

Proof. Introduce the shorthand $\tilde{q}_T = \tilde{q}_T(x,v)$, $q^{(1)}_T = q_T(x,v)$ and $q^{(2)}_T = q_T(x,\Phi(v))$. Since $\tilde{q}_T = q^{(2)}_T$ we have $q^{(2)}_T - q^{(1)}_T = \tilde{q}_T - q^{(1)}_T$ which implies that

$$
\begin{aligned}
T\,|\Phi(v) - v| \;&=\; \left|\int_0^T (T-s)[\nabla U(q^{(2)}_s) - \nabla U(q^{(1)}_s)]ds + \tilde{q}_T - q^{(1)}_T\right| \\
&\le\; \frac{LT^2}{2}\max_{s\le T}\left|q^{(2)}_s - q^{(1)}_s\right| + \left|\tilde{q}_T - q^{(1)}_T\right|\,.
\end{aligned} \tag{74}
$$

Moreover, by (48) with $h = 0$, we obtain $\max_{s\le T}\left|q_s^{(2)} - q_s^{(1)}\right| \le (7/6)T\,|\Phi(v) - v|$. Inserting this latter inequality into (74) and simplifying gives

$$T\,|\Phi(v) - v| \;\le\; (7/6)\max_{s\le T}\left|\tilde q_s - q_s^{(1)}\right| \;=\; (7/6)\max_{s\le T}|\tilde q_s(x,v) - q_s(x,v)|\;.$$

Inserting (54) gives the required result. □

Lemma 28. *Suppose that Assumption 2 (A1)-(A4) hold. Then for any $x, v \in \mathbb{R}^d$ such that $\tilde q_T(x,v) = q_T(x,\Phi(v))$, we have*

$$\begin{aligned}\|D\Phi(v) - I_d\| \;\le\; \min\Bigg(&\frac{1}{2}, h^2\Big(\frac{43}{45}L + \frac{1946}{6075}L_H\,|x| + \frac{873707}{486000}L_H T\,|v| \\ &+ \big(\frac{121}{5400}L_H^2T^2 + \frac{1}{90}L_I\big)\,|x|^2 + \big(\frac{121}{150}L_H^2T^4 + \frac{2}{5}L_IT^2\big)\,|v|^2\Big)\Bigg)\,.\end{aligned} \tag{75}$$

Proof. Introduce the shorthand $\tilde q_T = \tilde q_T(x,v)$, $q_T^{(1)} = q_T(x,v)$ and $q_T^{(2)} = q_T(x,\Phi(v))$. The derivative of $q_T^{(2)} - q_T^{(1)} = \tilde q_T - q_T^{(1)}$ with respect to v yields

$$\begin{aligned}T\,(D\Phi(v) - I_d) \;=\; & D_2\tilde q_T - D_2 q_T^{(1)} + \int_0^T (T-s)[\mathbf{H}(q_s^{(2)}) - \mathbf{H}(q_s^{(1)})]D_2 q_s^{(1)}ds \\ &+ \int_0^T (T-s)\mathbf{H}(q_s^{(2)})[D_2q_s^{(2)} - D_2q_s^{(1)} + D_2q_s^{(2)}(D\Phi(v) - I_d)]ds\,.\end{aligned}$$

By Assumption 2 (A2)-(A3) and (49),

$$\begin{aligned}T\,\|D\Phi(v) - I_d\| \;\le\; & \left\|D_2\tilde q_T - D_2q_T^{(1)}\right\| + \frac{L_HT^2}{2}\frac{6T}{5}\max_{s\le T}\left|q_s^{(2)} - q_s^{(1)}\right| \\ &+ \frac{LT^2}{2}\max_{s\le T}\left\|D_2q_s^{(2)} - D_2q_s^{(1)}\right\| + \frac{LT^2}{2}\frac{6T}{5}\,\|D\Phi(v) - I_d\|\;.\end{aligned}$$

Since $LT^2 \le 1/6$, and inserting (48) and (51) with $h = 0$, we obtain

$$T\,\|D\Phi(v) - I_d\| \le \frac{10}{9}\left\|D_2\tilde q_T - D_2q_T^{(1)}\right\| + L_HT^3\frac{14}{15}\,(T\,|\Phi(v) - v|)\;.$$

Inserting (73) and (62), simplifying, and combining with a bound similar to (72) (and therefore omitted), gives (75). □

Appendix A: Proofs of Results for Mean-Field U

In this section, we provide the remaining ingredients needed to prove Theorem 9 and Theorem 11 from Section 2.5.2.

A.1. Preliminaries

Consider the mean-field model with potential energy function given in (19). Throughout this section, we assume that V, W are functions in $C^4(\mathbb{R}^k)$ with

$$L = \sup \left\|D^2 V\right\|, \quad L_H = \sup \left\|D^3 V\right\|, \quad L_I = \sup \left\|D^4 V\right\|,$$
$$\tilde{L} = \sup \left\|D^2 W\right\|, \quad \tilde{L}_H = \sup \left\|D^3 W\right\|, \quad \tilde{L}_I = \sup \left\|D^4 W\right\|,$$

which we assume are all bounded. This assumption is not needed in every statement given in this section, but for simplicity, we assume it throughout. Moreover, we assume that $T > 0$ and the time step size $h \geq 0$ are such that $T/h \in \mathbb{N}$ for $h \neq 0$ and

$$(L + 4\epsilon\tilde{L})(T^2 + Th) \leq 1/6 . \tag{76}$$

As we discuss next, the constant $(L + 4\epsilon\tilde{L})$ represents an effective Lipschitz constant for the gradient of U.

Define $\mathbf{H}_{ij}(x) = \nabla_{ij} U(x) = \partial^2 U(x)/\partial x^i \partial x^j$ for $i, j \in \{1, \ldots, n\}$. From (19), note that for all $x, y \in \mathbb{R}^{nk}$ and $i, j \in \{1, \ldots, n\}$,

$$|\nabla_i U(x)| \leq (L + 2\epsilon\tilde{L}) \left|x^i\right| + \frac{2\epsilon\tilde{L}}{n} \sum_{\ell \neq i} \left|x^\ell\right|, \tag{77}$$

$$|\nabla_i U(x) - \nabla_i U(y)| \leq (L + 2\epsilon\tilde{L}) \left|x^i - y^i\right| + \frac{2\epsilon\tilde{L}}{n} \sum_{\ell \neq i} \left|x^\ell - y^\ell\right|, \tag{78}$$

$$\|\mathbf{H}_{ij}(x)\| \leq \begin{cases} L + 2\epsilon\tilde{L} & \text{if } i = j, \\ 2\epsilon\tilde{L}/n & \text{else,} \end{cases} \tag{79}$$

$$\|\mathbf{H}_{ij}(x) - \mathbf{H}_{ij}(y)\| \leq \begin{cases} (L_H + 2\epsilon\tilde{L}_H) \left|x^i - y^i\right| + \frac{2\epsilon\tilde{L}_H}{n} \sum_{\ell \neq i} \left|x^\ell - y^\ell\right| & \text{if } i = j, \\ \frac{2\epsilon\tilde{L}_H}{n} (\left|x^i - y^i\right| + \left|x^j - y^j\right|) & \text{else,} \end{cases} \tag{80}$$

where we used $\sum_{\ell \neq i} = \sum_{\ell=1, \ell \neq i}^n$. Additionally, for any $i_1, i_2, i_3, i_4 \in \{1, \ldots, n\}$,

$$\left\| \frac{\partial^3 U(x)}{\partial x^{i_1} \partial x^{i_2} \partial x^{i_3}} \right\| \leq \begin{cases} L_H + 2\epsilon\tilde{L}_H & \text{if all indices are equal,} \\ 2\epsilon\tilde{L}_H/n & \text{if exactly two indices are equal,} \\ 0 & \text{else,} \end{cases} \tag{81}$$

$$\left\| \frac{\partial^4 U(x)}{\partial x^{i_1} \partial x^{i_2} \partial x^{i_3} \partial x^{i_4}} \right\| \leq \begin{cases} L_I + 2\epsilon\tilde{L}_I & \text{if all indices are equal,} \\ 2\epsilon\tilde{L}_I/n & \text{if exactly three indices are equal,} \\ 0 & \text{else.} \end{cases} \tag{82}$$

As in Section 5 for general U, since we assume (76), we can invoke Corollary 14, to obtain the existence/uniqueness of a function $\Phi : \mathbb{R}^{nk} \to \mathbb{R}^{nk}$ that satisfies either $\tilde{q}_T(x, v) = \tilde{q}_T(y, \Phi(v))$ for any $x, y, v \in \mathbb{R}^{nk}$ or $\tilde{q}_T(x, v) = q_T(x, \Phi(v))$ for any $x, v \in \mathbb{R}^{nk}$.

A.2. TV bounds and regularization by one-shot couplings

Here we prove that the transition kernel of uHMC for mean-field U has a regularizing effect with a better dimension dependence than for general U.

Lemma 29. *For any $x, y, v \in \mathbb{R}^{nk}$, let $\Phi : \mathbb{R}^{nk} \to \mathbb{R}^{nk}$ be the map defined by $\tilde{q}_T(x, v) = \tilde{q}_T(y, \Phi(v))$. Then*

$$\begin{aligned} \mathrm{TV}(\delta_x\tilde{\pi}, \delta_y\tilde{\pi}) \;&\le\; \mathrm{TV}(\mathrm{Law}(\xi), \mathrm{Law}(\Phi(\xi))) \\ &\le\; \frac{3}{2}\left(T^{-2} + 34k(L_H + 8\epsilon\tilde{L}_H)^2T^4\right)^{1/2} \sum_{\ell=1}^{n} \left|x^\ell - y^\ell\right|. \end{aligned} \tag{83}$$

Note that the prefactor in the upper bound of (83) does not depend on the number n of particles. In this sense (83) is an improvement over (36).

Proof. This proof is similar to the proof of Lemma 16, except here we directly insert the results in Lemma 37 and 38 into Lemma 15 to obtain

$$\mathrm{TV}(\mathrm{Law}(\xi), \mathrm{Law}(\Phi(\xi))^2 \;\le\; \left(\frac{9}{4}T^{-2} + \frac{2401}{32}\,k\,(L_H + 8\epsilon\tilde{L}_H)^2\,T^4\right)\left(\sum_{\ell=1}^{n}\left|x^\ell - y^\ell\right|\right)^2 .$$

Taking square roots and inserting $34 > (4/9)(2401/32)$ gives (36). □

Analogous to Lemmas 16 and 17, Lemma 29 similarly implies that the transition kernel of uHMC has a regularizing effect in the following sense.

Lemma 30. *For any $\nu, \eta \in \mathcal{P}(\mathbb{R}^{nk})$,*

$$\mathrm{TV}(\eta\tilde{\pi}, \nu\tilde{\pi}) \;\le\; \frac{3}{2}\left(T^{-2} + 34k(L_H + 8\epsilon\tilde{L}_H)^2T^4\right)^{1/2} \boldsymbol{W}^1_{\ell_1}(\eta, \nu)\,. \tag{84}$$

The proof of Lemma 30 is very similar to the proof of Lemma 17 except that it involves the $\boldsymbol{W}^1_{\ell_1}$ distance rather than the $\boldsymbol{W}^1$ distance, and therefore, omitted.

Lemma 31. *For any $x, v \in \mathbb{R}^{nk}$, let $\Phi : \mathbb{R}^{nk} \to \mathbb{R}^{nk}$ be the map defined by $\tilde{q}_T(x, v) = q_T(x, \Phi(v))$. Then*

$$\begin{aligned} &\mathrm{TV}(\delta_x\pi, \delta_x\tilde{\pi}) \;\le\; \mathrm{TV}(\mathrm{Law}(\xi), \mathrm{Law}(\Phi(\xi))) \\ &\le\; h^2\Big[n^2k\Big(17(L + 4\epsilon\tilde{L})^2 + 28(L_H + 8\epsilon\tilde{L}_H)^2T^2 + 104k(L_H + 8\epsilon\tilde{L}_H)^2T^2 \\ &+ 180(2k + k^2)\big((L_I + 14\epsilon\tilde{L}_I)^2T^2 + (L_H + 8\epsilon\tilde{L}_H)^2T^4\big)^2\Big) \\ &+ n\Big(10k(L_H + 8\epsilon\tilde{L}_H)^2 + 7(L + 4\epsilon\tilde{L})^2T^{-2}\Big)\sum_{\ell=1}^{n}\left|x^\ell\right|^2 \\ &+ n\Big(40k\big((L_I + 14\epsilon\tilde{L}_I) + (L_H + 8\epsilon\tilde{L}_H)^2T^2\big)^2 \\ &+ 7(L_H + 8\epsilon\tilde{L}_H)^2T^{-2}\Big)\sum_{\ell=1}^{n}\left|x^\ell\right|^4\Big]^{1/2}. \end{aligned} \tag{85}$$

Proof. The proof is similar to the proof of Lemma 18 except here we directly insert the results of Lemmas 39 and 40 into Lemma 15, and then use the Cauchy-Schwarz inequality, to obtain

$$\begin{aligned}
&\mathrm{TV}(\mathrm{Law}(\xi),\mathrm{Law}(\Phi(\xi))^2 \;\le\; E\Big[|\Phi(\xi)-\xi|^2 + 2\,\|D\Phi(\xi)-I_d\|_F^2\Big] \\
&\le\; h^4\, E\Big[10n^2k(L+4\epsilon\tilde{L})^2 + n\big(10k(L_H+8\epsilon\tilde{L}_H)^2 + 7(L+4\epsilon\tilde{L})^2T^{-2}\big)\sum_{\ell=1}^n |x^\ell|^2 \\
&\quad + n\big(90k(L_H+8\epsilon\tilde{L}_H)^2T^2 + 7(L+4\epsilon\tilde{L})^2\big)\sum_{\ell=1}^n |v^\ell|^2 \\
&\quad + n\Big(10k(L_I+14\epsilon\tilde{L}_I)^2 + 7(L_H+8\epsilon\tilde{L}_H)^2T^{-2} \\
&\qquad + 40k(L_H+8\epsilon\tilde{L}_H)^2(L_I+14\epsilon\tilde{L}_I)T^2 + 40k(L_H+8\epsilon\tilde{L}_H)^4T^4\Big)\sum_{\ell=1}^n |x^\ell|^4 \\
&\quad + n\Big(10k(L_I+14\epsilon\tilde{L}_I)^2T^4 + 7(L_H+8\epsilon\tilde{L}_H)^2T^2 \\
&\qquad + 60k(L_H+8\epsilon\tilde{L}_H)^2(L_I+14\epsilon\tilde{L}_I)T^6 + 90k(L_H+8\epsilon\tilde{L}_H)^4T^8\Big)\sum_{\ell=1}^n |v^\ell|^4\Big].
\end{aligned}$$

We then use $E[|\xi^\ell|^2] = k$ and $E[|\xi^\ell|^4] = 2k(k+2)$ to obtain

$$\begin{aligned}
&\mathrm{TV}(\mathrm{Law}(\xi),\mathrm{Law}(\Phi(\xi))^2 \\
&\le\; h^4\Big[n^2k\Big(17(L+4\epsilon\tilde{L})^2 + 28(L_H+8\epsilon\tilde{L}_H)^2T^2 + 104k(L_H+8\epsilon\tilde{L}_H)^2T^2 \\
&\qquad + (40k+20k^2)(L_I+14\epsilon\tilde{L}_I)^2T^4 + (240k+120k^2)(L_H+8\epsilon\tilde{L}_H)^2(L_I+14\epsilon\tilde{L}_I)T^6 \\
&\qquad + (360k+180k^2)(L_H+8\epsilon\tilde{L}_H)^4T^8\Big) \\
&\quad + n\big(10k(L_H+8\epsilon\tilde{L}_H)^2 + 7(L+4\epsilon\tilde{L})^2T^{-2}\big)\sum_{\ell=1}^n |x^\ell|^2 \\
&\quad + n\Big(10k(L_I+14\epsilon\tilde{L}_I)^2 + 7(L_H+8\epsilon\tilde{L}_H)^2T^{-2} \\
&\qquad + 40k(L_H+8\epsilon\tilde{L}_H)^2(L_I+14\epsilon\tilde{L}_I)T^2 + 40k(L_H+8\epsilon\tilde{L}_H)^4T^4\Big)\sum_{\ell=1}^n |x^\ell|^4\Big].
\end{aligned}$$

Taking square roots and simplifying gives (85). □

Lemma 31 implies the following bound on the TV distance between $\mu\pi$ and $\mu\tilde{\pi}$.

Lemma 32. *We have* $\mathrm{TV}(\mu\pi,\mu\tilde{\pi}) \le Ch^2$*, where* C *is defined in* (25).

The proof of this result is similar to the proof of Lemma 17 and therefore omitted.

A.3. A priori estimates for the dynamics

Here we develop estimates on the dynamics (7) for the special case of the mean-field model. Let $\tilde{q}_t(x,v) = (\tilde{q}_t^1(x,v), \dots, \tilde{q}_t^n(x,v))$ and similarly for $\tilde{v}_t(x,v)$.

Lemma 33. *For any $x, y, u, v \in \mathbb{R}^{nk}$ and $i \in \{1, \dots, n\}$,*

$$\max_{s \le T} \left|\tilde{q}_s^i(x,v)\right| \le (1 + (L + 2\epsilon\tilde{L})(T^2 + Th)) \max(\left|x^i\right|, \left|x^i + Tv^i\right|) + \frac{2\epsilon\tilde{L}(T^2 + Th)}{n} \sum_{\ell \ne i} \max_{s \le T} \left|\tilde{q}_s^i(x,v)\right|, \tag{86}$$

$$\max_{s \le T} \left|\tilde{v}_s^i(x,v)\right| \le \left|v^i\right| + \frac{2\epsilon\tilde{L}T}{n}(1 + (L + 2\epsilon\tilde{L})(T^2 + Th)) \sum_{\ell \ne i} \max_{s \le T} \left|\tilde{q}_s^i(x,v)\right| + (L + 2\epsilon\tilde{L})T(1 + (L + 2\epsilon\tilde{L})(T^2 + Th)) \max(\left|x^i\right|, \left|x^i + Tv^i\right|), \tag{87}$$

$$\sum_{\ell=1}^n \max_{s \le T} \left|\tilde{q}_s^\ell(x,v)\right| \le (1 + (L + 4\epsilon\tilde{L})(T^2 + Th)) \sum_{\ell=1}^n \max(\left|x^\ell\right|, \left|x^\ell + Tv^\ell\right|), \tag{88}$$

$$\sum_{\ell=1}^n \max_{s \le T} \left|\tilde{v}_s^\ell(x,v)\right| \le \sum_{\ell=1}^n \left|v^\ell\right| + (L + 4\epsilon\tilde{L})T(1 + (L + 4\epsilon\tilde{L})(T^2 + Th)) \sum_{\ell=1}^n \max(\left|x^\ell\right|, \left|x^\ell + Tv^\ell\right|), \tag{89}$$

$$\sum_{\ell=1}^n \max_{s \le T} \left|\tilde{q}_s^\ell(x,u) - \tilde{q}_s^\ell(y,v)\right| \le (1 + (L + 4\epsilon\tilde{L})(T^2 + Th)) \sum_{\ell=1}^n \max\left(\left|x^\ell - y^\ell\right|, \left|x^\ell - y^\ell + T(u^\ell - v^\ell)\right|\right). \tag{90}$$

Lemma 33 is the mean-field analog of Lemma 20 and is contained in Lemmas 12 and 13 of [6]. Hence, a proof is omitted.

A.4. A priori estimates for the velocity derivative flow

Here we develop estimates on the velocity derivative flows $\partial_{v^j} \tilde{q}_s^i = \partial \tilde{q}_s^i / \partial v^j$.

Lemma 34. *For any $x, y, u, v \in \mathbb{R}^{nk}$ and $j \in \{1, \dots, n\}$,*

$$\sum_{\ell=1}^n \max_{s \le T} \left\| \partial_{v^j} \tilde{q}_s^\ell(x, v) \right\| \le (6/5)\, T\,, \tag{91}$$

$$\sum_{\ell=1}^n \max_{s \le T} \left\| \partial_{v^\ell} \tilde{q}_s^j(x, v) \right\| \le (7/5)\, T\,, \tag{92}$$

$$\sum_{\ell=1}^n \max_{s \le T} \left\| \partial_{v^j} \tilde{v}_s^\ell(x, v) \right\| \le (6/5)\,, \quad \textit{and} \tag{93}$$

$$\begin{aligned}\sum_{i=1}^n \sum_{\ell=1}^n \max_{s \le T} \left\| \partial_{v^i} \tilde{q}_s^\ell(x, u) - \partial_{v^i} \tilde{q}_s^\ell(y, v) \right\| &\le (42/25)(L_H + 8\epsilon \tilde{L}_H) T^3 \\ \times (1 + (L + 4\epsilon\tilde{L})(T^2 + Th)) &\sum_{\ell=1}^n \max\left(\left|x^\ell - y^\ell\right|, \left|x^\ell - y^\ell + T(u^\ell - v^\ell)\right| \right).\end{aligned} \tag{94}$$

Lemma 34 is the mean-field analog of Lemma 21.

Proof. From (7), note that for any $\ell, j \in \{1, \dots, n\}$,

$$\begin{aligned}\partial_{v^j} \tilde{q}_T^\ell(x, u) &= T I_k \delta_{\ell j} \\ &\quad - \frac{1}{2} \int_0^T (T - s) \sum_{i=1}^n \left[\mathbf{H}_{\ell i}(\tilde{q}_{\lfloor s \rfloor_h}) \partial_{v^j} \tilde{q}^i_{\lfloor s \rfloor_h} + \mathbf{H}_{\ell i}(\tilde{q}_{\lceil s \rceil_h}) \partial_{v^j} \tilde{q}^i_{\lceil s \rceil_h} \right] ds \\ &\quad + \frac{1}{2} \int_0^T (s - \lfloor s \rfloor_h) \sum_{i=1}^n \left[\mathbf{H}_{\ell i}(\tilde{q}_{\lceil s \rceil_h}) \partial_{v^j} \tilde{q}^i_{\lceil s \rceil_h} - \mathbf{H}_{\ell i}(\tilde{q}_{\lfloor s \rfloor_h}) \partial_{v^j} \tilde{q}^i_{\lfloor s \rfloor_h} \right] ds\,, \\ \partial_{v^j} \tilde{v}_T^\ell(x, u) &= I_k \delta_{\ell j} - \frac{1}{2} \int_0^T \sum_{i=1}^n \left[\mathbf{H}_{\ell i}(\tilde{q}_{\lfloor s \rfloor_h}) \partial_{v^j} \tilde{q}^i_{\lfloor s \rfloor_h} + \mathbf{H}_{\ell i}(\tilde{q}_{\lceil s \rceil_h}) \partial_{v^j} \tilde{q}^i_{\lceil s \rceil_h} \right] ds.\end{aligned} \tag{95}$$

Using (79) and inserting $(L + 4\epsilon\tilde{L})T^2 \le 1/6$, we obtain

$$\begin{aligned}\sum_{\ell=1}^n \max_{s \le T} \left\| \partial_{v^j} \tilde{q}_T^\ell(x, u) \right\| &\le T \left\| I_k \right\| + T^2 \sum_{\ell=1}^n \sum_{i=1}^n \max_{s \le T} \left\| \mathbf{H}_{\ell i}(\tilde{q}_s) \partial_{v^j} \tilde{q}_s^i(x, u) \right\| \\ &\le T + (L + 4\epsilon\tilde{L})T^2 \sum_{i=1}^n \max_{s \le T} \left\| \partial_{v^j} \tilde{q}_s^i(x, u) \right\| \le (6/5)T\,, \\ \sum_{\ell=1}^n \max_{s \le T} \left\| \partial_{v^j} \tilde{v}_s^\ell(x, u) \right\| &\le 1 + (L + 4\epsilon\tilde{L})T \sum_{\ell=1}^n \max_{s \le T} \left\| \partial_{v^j} \tilde{q}_T^\ell(x, u) \right\| \le 6/5\,,\end{aligned}$$

which gives (91) and (93). For (92), by (79) and $(L + 4\epsilon\tilde{L})T^2 \le 1/6$ imply that,

$$\begin{aligned}\sum_{\ell=1}^n \max_{s \le T} \left\| \partial_{v^\ell} \tilde{q}_T^j(x, u) \right\| &\le T + T^2 \sum_{\ell=1}^n \sum_{i=1}^n \max_{s \le T} \left\| \mathbf{H}_{ji}(\tilde{q}_s) \partial_{v^\ell} \tilde{q}_s^i(x, u) \right\| \\ &\le T + (L + 2\epsilon\tilde{L})T^2 \sum_{\ell=1}^n \max_{s \le T} \left\| \partial_{v^\ell} \tilde{q}_s^j(x, u) \right\| + T^2 \frac{2\epsilon\tilde{L}}{n} \sum_{\ell=1}^n \sum_{i=1}^n \max_{s \le T} \left\| \partial_{v^\ell} \tilde{q}_s^i(x, u) \right\| \\ &\le (6/5)T + 2(6/5)^2 T^3 \epsilon\tilde{L} \le (7/5)T\,,\end{aligned}$$

where in the second to last step we used (91), and in the last step, we used $\epsilon\tilde{L}T^2 \leq 1/24$ and $33/25 \leq 7/5$. For (94), it is notationally convenient to introduce $\tilde{q}_s^{(1),\ell} := \tilde{q}_s^\ell(x,u)$ and $\tilde{q}_s^{(2),\ell} := \tilde{q}_s^\ell(y,v)$ for any $s \in [0,T]$ and $\ell \in \{1,\dots,n\}$. Then from (95)

$$\begin{aligned}
\partial_{v^j}\tilde{q}_T^{(1),\ell} - \partial_{v^j}\tilde{q}_T^{(2),\ell} = &-\int_0^T \frac{T-s}{2}\sum_{i=1}^n \Big[\mathbf{H}_{\ell i}(\tilde{q}^{(1)}_{\lfloor s\rfloor_h})\partial_{v^j}\tilde{q}^{(1),i}_{\lfloor s\rfloor_h} - \mathbf{H}_{\ell i}(\tilde{q}^{(2)}_{\lfloor s\rfloor_h})\partial_{v^j}\tilde{q}^{(2),i}_{\lfloor s\rfloor_h}\Big]ds \\
&-\int_0^T \frac{T-s}{2}\sum_{i=1}^n \Big[\mathbf{H}_{\ell i}(\tilde{q}^{(1)}_{\lceil s\rceil_h})\partial_{v^j}\tilde{q}^{(1),i}_{\lceil s\rceil_h} - \mathbf{H}_{\ell i}(\tilde{q}^{(2)}_{\lceil s\rceil_h})\partial_{v^j}\tilde{q}^{(2),i}_{\lceil s\rceil_h}\Big]ds \\
&+\int_0^T \frac{s-\lfloor s\rfloor_h}{2}\sum_{i=1}^n \Big[\mathbf{H}_{\ell i}(\tilde{q}^{(1)}_{\lceil s\rceil_h})\partial_{v^j}\tilde{q}^{(1),i}_{\lceil s\rceil_h} - \mathbf{H}_{\ell i}(\tilde{q}^{(2)}_{\lceil s\rceil_h})\partial_{v^j}\tilde{q}^{(2),i}_{\lceil s\rceil_h}\Big]ds \\
&+\int_0^T \frac{s-\lfloor s\rfloor_h}{2}\sum_{i=1}^n \Big[\mathbf{H}_{\ell i}(\tilde{q}^{(2)}_{\lfloor s\rfloor_h})\partial_{v^j}\tilde{q}^{(2),i}_{\lfloor s\rfloor_h} - \mathbf{H}_{\ell i}(\tilde{q}^{(1)}_{\lfloor s\rfloor_h})\partial_{v^j}\tilde{q}^{(1),i}_{\lfloor s\rfloor_h}\Big]ds
\end{aligned}$$

$$\begin{aligned}
&= \int_0^T \frac{T-s}{2}\sum_{i=1}^n \Big[\mathbf{H}_{\ell i}(\tilde{q}^{(2)}_{\lfloor s\rfloor_h})(\partial_{v^j}\tilde{q}^{(2),i}_{\lfloor s\rfloor_h} - \partial_{v^j}\tilde{q}^{(1),i}_{\lfloor s\rfloor_h}) - (\mathbf{H}_{\ell i}(\tilde{q}^{(1)}_{\lfloor s\rfloor_h}) - \mathbf{H}_{\ell i}(\tilde{q}^{(2)}_{\lfloor s\rfloor_h}))\partial_{v^j}\tilde{q}^{(1)}_{\lfloor s\rfloor_h}\Big]ds \\
&+ \int_0^T \frac{T-s}{2}\sum_{i=1}^n \Big[\mathbf{H}_{\ell i}(\tilde{q}^{(2)}_{\lceil s\rceil_h})(\partial_{v^j}\tilde{q}^{(2),i}_{\lceil s\rceil_h} - \partial_{v^j}\tilde{q}^{(1),i}_{\lceil s\rceil_h}) - (\mathbf{H}_{\ell i}(\tilde{q}^{(1)}_{\lceil s\rceil_h}) - \mathbf{H}_{\ell i}(\tilde{q}^{(2)}_{\lceil s\rceil_h}))\partial_{v^j}\tilde{q}^{(1),i}_{\lceil s\rceil_h}\Big]ds \\
&- \int_0^T \frac{s-\lfloor s\rfloor_h}{2}\sum_{i=1}^n \Big[\mathbf{H}_{\ell i}(\tilde{q}^{(2)}_{\lceil s\rceil_h})(\partial_{v^j}\tilde{q}^{(2),i}_{\lceil s\rceil_h} - \partial_{v^j}\tilde{q}^{(1),i}_{\lceil s\rceil_h}) - (\mathbf{H}_{\ell i}(\tilde{q}^{(1)}_{\lceil s\rceil_h}) - \mathbf{H}_{\ell i}(\tilde{q}^{(2)}_{\lceil s\rceil_h}))\partial_{v^j}\tilde{q}^{(1),i}_{\lceil s\rceil_h}\Big]ds \\
&- \int_0^T \frac{s-\lfloor s\rfloor_h}{2}\sum_{i=1}^n \Big[\mathbf{H}_{\ell i}(\tilde{q}^{(1)}_{\lfloor s\rfloor_h})(\partial_{v^j}\tilde{q}^{(1),i}_{\lfloor s\rfloor_h} - \partial_{v^j}\tilde{q}^{(2),i}_{\lfloor s\rfloor_h}) - (\mathbf{H}_{\ell i}(\tilde{q}^{(2)}_{\lfloor s\rfloor_h}) - \mathbf{H}_{\ell i}(\tilde{q}^{(1)}_{\lfloor s\rfloor_h}))\partial_{v^j}\tilde{q}^{(2),i}_{\lfloor s\rfloor_h}\Big]ds.
\end{aligned}$$

By using (79), (80), and (92), we then get

$$\begin{aligned}
\sum_{j=1}^n\sum_{\ell=1}^n \max_{s\leq T}\left\|\partial_{v^j}\tilde{q}_T^{(1),\ell} - \partial_{v^j}\tilde{q}_T^{(2),\ell}\right\| &\leq (L+4\epsilon\tilde{L})T^2\sum_{j=1}^n\sum_{\ell=1}^n \max_{s\leq T}\left\|\partial_{v^j}\tilde{q}_T^{(1),\ell} - \partial_{v^j}\tilde{q}_T^{(2),\ell}\right\| \\
&+ (L_H + 8\epsilon\tilde{L}_H)(7/5)T^3\sum_{\ell=1}^n \max_{s\leq T}\left|\tilde{q}_s^{(1),\ell} - \tilde{q}_s^{(2),\ell}\right| .
\end{aligned}$$

Inserting $(L+4\epsilon\tilde{L})T^2 \leq 1/6$ and (90), and simplifying yields (94). □

A.5. Discretization error bounds

The following lemma is the mean-field analog of Lemma 23.

Lemma 35. *For any $x, v \in \mathbb{R}^{nk}$,*

$$\sum_{\ell=1}^n \max_{s\le T} \left| q_s^\ell(x,v) - \tilde{q}_s^\ell(x,v) \right| \le h^2 \Big((L+4\epsilon\tilde{L}) \sum_{\ell=1}^n \left|x^\ell\right| + (L+4\epsilon\tilde{L})T \sum_{\ell=1}^n \left|v^\ell\right| + (L_H + 8\epsilon\tilde{L}_H) \sum_{\ell=1}^n \left|x^\ell\right|^2 + (L_H+8\epsilon\tilde{L}_H)T \sum_{\ell=1}^n \left|v^\ell\right|^2 \Big) . \tag{96}$$

Proof. By (88) and (89) with $h = 0$, and since $(L+4\epsilon\tilde{L})T^2 \le 1/6$, note that

$$\sum_{\ell=1}^n \max_{s\le T} \left|q_s^\ell\right| \le (7/6) \sum_{\ell=1}^n \left|x^\ell\right| + (7/6)T \sum_{\ell=1}^n \left|v^\ell\right| , \tag{97}$$

$$\sum_{\ell=1}^n \max_{s\le T} \left|v_s^\ell\right| \le (7/6)\,(L+4\epsilon\tilde{L})T \sum_{\ell=1}^n \left|x^\ell\right| + (6/5) \sum_{\ell=1}^n \left|v^\ell\right| , \tag{98}$$

$$\sum_{\ell=1}^n \max_{s\le T} \left|q_s^\ell\right|^2 \le (9/2) \sum_{\ell=1}^n \left|x^\ell\right|^2 + (9/2)T^2 \sum_{\ell=1}^n \left|v^\ell\right|^2 , \tag{99}$$

$$\sum_{\ell=1}^n \max_{s\le T} \left|v_s^\ell\right|^2 \le (9/2)(L+4\epsilon\tilde{L}) \sum_{\ell=1}^n \left|x^\ell\right|^2 + (9/2) \sum_{\ell=1}^n \left|v^\ell\right|^2 . \tag{100}$$

Introduce the shorthand $q_T := q_T(x,v)$ and $\tilde{q}_T := \tilde{q}_T(x,v)$. Using (7), note that

$$q_T^\ell - \tilde{q}_T^\ell = \mathrm{I}^\ell + \mathrm{II}^\ell + \mathrm{III}^\ell \quad \text{where} \tag{101}$$

$$\mathrm{I}^\ell := \frac{1}{2}\int_0^T (T-s)[\nabla_\ell U(\tilde{q}_{\lfloor s\rfloor_h}) - \nabla_\ell U(q_{\lfloor s\rfloor_h}) + \nabla_\ell U(\tilde{q}_{\lceil s\rceil_h}) - \nabla_\ell U(q_{\lceil s\rceil_h})]ds - \frac{1}{2}\int_0^T (s-\lfloor s\rfloor_h)[\nabla_\ell U(\tilde{q}_{\lceil s\rceil_h}) - \nabla_\ell U(q_{\lceil s\rceil_h}) - (\nabla_\ell U(\tilde{q}_{\lfloor s\rfloor_h}) - \nabla_\ell U(q_{\lfloor s\rfloor_h}))]ds ,$$

$$\mathrm{II}^\ell := -\frac{1}{2}\int_0^T (s-\lfloor s\rfloor_h)[\nabla_\ell U(q_{\lceil s\rceil_h}) - \nabla_\ell U(q_{\lfloor s\rfloor_h})]ds , \quad \text{and}$$

$$\mathrm{III}^\ell := -\int_0^T (T-s)\nabla_\ell U(q_s)ds + \frac{1}{2}\int_0^T (T-s)[\nabla_\ell U(q_{\lfloor s\rfloor_h}) + \nabla_\ell U(q_{\lceil s\rceil_h})]ds .$$

By (78) and since $(L+4\epsilon\tilde{L})T^2 \le 1/6$,

$$\sum_{\ell=1}^n \left|\mathrm{I}^\ell\right| \le (L+4\epsilon\tilde{L})T^2 \sum_{\ell=1}^n \max_{s\le T} \left|q_s^\ell - \tilde{q}_s^\ell\right| \le (1/6) \sum_{\ell=1}^n \max_{s\le T}\left|q_s^\ell - \tilde{q}_s^\ell\right| , \tag{102}$$

$$\sum_{\ell=1}^n \left|\mathrm{II}^\ell\right| \le \frac{h(L+4\epsilon\tilde{L})}{2} \sum_{\ell=1}^n \int_0^T \left|q_{\lceil s\rceil_h}^\ell - q_{\lfloor s\rfloor_h}^\ell\right| ds = \frac{h(L+4\epsilon\tilde{L})}{2} \sum_{\ell=1}^n \int_0^T \left| \int_{\lfloor s\rfloor_h}^{\lceil s\rceil_h} v_r^\ell dr \right| ds, \le \frac{h^2(L+4\epsilon\tilde{L})T}{2} \sum_{\ell=1}^n \max_{s\le T}\left|v_s^\ell\right| \le (L+4\epsilon\tilde{L})h^2 \left(\frac{7}{72} \sum_{\ell=1}^n \left|x^\ell\right| + \frac{3}{5} T \sum_{\ell=1}^n \left|v^\ell\right| \right) , \tag{103}$$

where for (103) we used (89). Applying Lemma 22 with $f(s) = -\nabla_\ell U(q_s)$ we obtain

$$\begin{aligned}\sum_{\ell=1}^n \left|\mathrm{III}^\ell\right| \le \frac{h^2}{12}\Big((L+4\epsilon\tilde{L})T\sum_{\ell=1}^n \max_{s\le T}\left|v_s^\ell\right| + (L+4\epsilon\tilde{L})^2T^2\sum_{\ell=1}^n \max_{s\le T}\left|q_s^\ell\right| \\ + (L_H+8\epsilon\tilde{L}_H)T^2\sum_{\ell=1}^n \max_{s\le T}\left|v_s^\ell\right|^2\Big)\,. \end{aligned} \tag{104}$$

Inserting (102), (103) and (104) into the norm of (101) summed over ℓ; then inserting (97), (98), and (100); and then simplifying gives (96). □

Lemma 36. *For any $x, v \in \mathbb{R}^{nk}$,*

$$\begin{aligned}&\sum_{i=1}^n\sum_{\ell=1}^n \max_{s\le T}\left\|\partial_{v^i} q_s^\ell(x,v) - \partial_{v^i}\tilde{q}_s^\ell(x,v)\right\| \\ &\le\ h^2\Big((L+4\epsilon\tilde{L})Tn + (L_H+8\epsilon\tilde{L}_H)T\sum_{\ell=1}^n\left|x^\ell\right| + 2(L_H+8\epsilon\tilde{L}_H)T^2\sum_{\ell=1}^n\left|v^\ell\right| \\ &\qquad + \big(2(L_H+8\epsilon\tilde{L}_H)^2T^3 + (L_I+14\epsilon\tilde{L}_I)T\big)\sum_{\ell=1}^n\left|x^\ell\right|^2 \\ &\qquad + \big(2(L_H+8\epsilon\tilde{L}_H)^2T^5 + (L_I+14\epsilon\tilde{L}_I)T^3\big)\sum_{\ell=1}^n\left|v^\ell\right|^2\Big)\,. \end{aligned} \tag{105}$$

Proof. Using (7), write the difference as

$$\partial_{v^i} q_T^\ell(x,v) - \partial_{v^i} \tilde{q}_T^\ell(x,v) \ = \ \mathrm{I}_i^\ell + \mathrm{II}_i^\ell + \mathrm{III}_i^\ell + \mathrm{IV}_i^\ell \quad \text{where} \tag{106}$$

$$\begin{aligned}
\mathrm{I}_i^\ell \ &:= \ \frac{1}{2}\int_0^T (T-s) \sum_{m=1}^n \mathbf{H}_{\ell m}(\tilde{q}_{\lfloor s \rfloor_h})(\partial_{v^i}\tilde{q}^m_{\lfloor s \rfloor_h} - \partial_{v^i} q^m_{\lfloor s \rfloor_h}) ds \\
&+ \frac{1}{2}\int_0^T (T-s) \sum_{m=1}^n \mathbf{H}_{\ell m}(\tilde{q}_{\lceil s \rceil_h})(\partial_{v^i}\tilde{q}^m_{\lceil s \rceil_h} - \partial_{v^i} q^m_{\lceil s \rceil_h}) ds \\
&- \frac{1}{2}\int_0^T (s-\lfloor s \rfloor_h) \sum_{m=1}^n \mathbf{H}_{\ell m}(\tilde{q}_{\lceil s \rceil_h})(\partial_{v^i}\tilde{q}^m_{\lceil s \rceil_h} - \partial_{v^i} q^m_{\lceil s \rceil_h}) ds \\
&+ \frac{1}{2}\int_0^T (s-\lfloor s \rfloor_h) \sum_{m=1}^n \mathbf{H}_{\ell m}(\tilde{q}_{\lfloor s \rfloor_h})(\partial_{v^i}\tilde{q}^m_{\lfloor s \rfloor_h} - \partial_{v^i} q^m_{\lfloor s \rfloor_h}) ds \\
\mathrm{II}_i^\ell \ &:= \ \frac{1}{2}\int_0^T (T-s) \sum_{m=1}^n (\mathbf{H}_{\ell m}(\tilde{q}_{\lfloor s \rfloor_h}) - \mathbf{H}_{\ell m}(q_{\lfloor s \rfloor_h}))\partial_{v^i} q^m_{\lfloor s \rfloor_h} ds \\
&+ \frac{1}{2}\int_0^T (T-s) \sum_{m=1}^n (\mathbf{H}_{\ell m}(\tilde{q}_{\lceil s \rceil_h}) - \mathbf{H}_{\ell m}(q_{\lceil s \rceil_h}))\partial_{v^i} q^m_{\lceil s \rceil_h} ds \\
&- \frac{1}{2}\int_0^T (s-\lfloor s \rfloor_h) \sum_{m=1}^n (\mathbf{H}_{\ell m}(\tilde{q}_{\lceil s \rceil_h}) - \mathbf{H}_{\ell m}(q_{\lceil s \rceil_h}))\partial_{v^i} q^m_{\lceil s \rceil_h} ds \\
&+ \frac{1}{2}\int_0^T (s-\lfloor s \rfloor_h) \sum_{m=1}^n (\mathbf{H}_{\ell m}(\tilde{q}_{\lfloor s \rfloor_h}) - \mathbf{H}_{\ell m}(q_{\lfloor s \rfloor_h}))\partial_{v^i} q^m_{\lfloor s \rfloor_h} ds \\
\mathrm{III}_i^\ell \ &:= \ -\frac{1}{2}\int_0^T (s-\lfloor s \rfloor_h) \sum_{m=1}^n [(\mathbf{H}_{\ell m}(q_{\lceil s \rceil_h}) - \mathbf{H}_{\ell m}(q_{\lfloor s \rfloor_h}))\partial_{v^i} q^m_{\lceil s \rceil_h}] \\
&- \frac{1}{2}\int_0^T (s-\lfloor s \rfloor_h) \sum_{m=1}^n [\mathbf{H}_{\ell m}(q_{\lfloor s \rfloor_h})(\partial_{v^i} q^m_{\lceil s \rceil_h} - \partial_{v^i} q^m_{\lfloor s \rfloor_h})] ds \ , \quad \text{and} \\
\mathrm{IV}_i^\ell \ &:= \ -\int_0^T (T-s) \sum_{m=1}^n \mathbf{H}_{\ell m}(q_s)\partial_{v^i} q^m_s ds \\
&+ \frac{1}{2}\int_0^T (T-s) \sum_{m=1}^n [\mathbf{H}_{\ell m}(q_{\lfloor s \rfloor_h})\partial_{v^i} q^m_{\lfloor s \rfloor_h} + \mathbf{H}_{\ell m}(q_{\lceil s \rceil_h})\partial_{v^i} q^m_{\lceil s \rceil_h}] ds \ .
\end{aligned}$$

To obtain the following bounds, we use (79) and $(L + 4\epsilon\tilde{L})T^2 \le 1/6$ for (107); (80) and (92) for

(108); and (79), (80), (92), and for (109).

$$\sum_{i=1}^{n}\sum_{\ell=1}^{n}\left\|\mathrm{I}_i^{\ell}\right\| \le (1/6)\max_{s\le T}\sum_{i=1}^{n}\sum_{\ell=1}^{n}\left\|\partial_{v^i}q_s^{\ell}-\partial_{v^i}\tilde{q}_s^{\ell}\right\| , \tag{107}$$

$$\sum_{i=1}^{n}\sum_{\ell=1}^{n}\left\|\mathrm{II}_i^{\ell}\right\| \le (7/5)(L_H+8\epsilon\tilde{L}_H)T^3\sum_{\ell=1}^{m}\max_{s\le T}\left|q_s^{\ell}-\tilde{q}_s^{\ell}\right| , \tag{108}$$

$$\begin{aligned}\sum_{i=1}^{n}\sum_{\ell=1}^{n}\left\|\mathrm{III}_i^{\ell}\right\| &\le (7/5)\frac{h(L_H+8\epsilon\tilde{L}_H)T}{2}\sum_{\ell=1}^{n}\int_0^T\left|\int_{\lfloor s\rfloor_h}^{\lceil s\rceil_h}v_r^{\ell}dr\right|ds \\ &\quad+\frac{h(L+4\epsilon\tilde{L})}{2}\sum_{i=1}^{n}\sum_{\ell=1}^{n}\int_0^T\left\|\int_{\lfloor s\rfloor_h}^{\lceil s\rceil_h}\partial_{v^i}v_r^{\ell}dr\right\|ds \\ &\le \frac{7h^2(L_H+8\epsilon\tilde{L}_H)T^2}{10}\sum_{\ell=1}^{n}\max_{s\le T}\left|v_s^{\ell}\right|+\frac{h^2(L+4\epsilon\tilde{L})T}{2}\sum_{i=1}^{n}\sum_{\ell=1}^{n}\max_{s\le T}\left\|\partial_{v^i}v_s^{\ell}\right\| .\end{aligned} \tag{109}$$

Applying Lemma 22 with $f(s)=\mathbf{H}_{\ell m}(q_s)\partial_{v^i}q_s^m$ and using $(L+4\epsilon\tilde{L})T^2\le 1/6$, (79), (81), (82), (92) and (93), we obtain

$$\begin{aligned}&\sum_{i=1}^{n}\sum_{\ell=1}^{n}\left\|\mathrm{IV}_i^{\ell}\right\| \\ &\le \frac{h^2}{12}\Big(\frac{43}{30}(L+4\epsilon\tilde{L})Tn+4(L_H+8\epsilon\tilde{L}_H)T^2\sum_{\ell=1}^{n}\max_{s\le T}\left|v_s^{\ell}\right| \\ &\quad+\frac{7}{30}(L_H+8\epsilon\tilde{L}_H)T\sum_{\ell=1}^{n}\max_{s\le T}\left|q_s^{\ell}\right|+\frac{7}{5}(L_I+14\epsilon\tilde{L}_I)T^3\sum_{\ell=1}^{n}\max_{s\le T}\left|v_s^{\ell}\right|^2\Big) .\end{aligned} \tag{110}$$

Insert (107), (108), (109), (110) and Lemma 35 into norm of the double sum of (106) over i and ℓ; use (97), (98), (100), and (93); and simplify to obtain (105). □

A.6. One-shot coupling bounds for $\tilde{q}_T(x,\xi)=\tilde{q}_T(y,\Phi(\xi))$

The following lemmas are the mean-field analogs of Lemmas 25 and 26.

Lemma 37. *For any $x,y,v\in\mathbb{R}^{nk}$ such that $\tilde{q}_T(x,v)=\tilde{q}_T(y,\Phi(v))$, we have*

$$T\left|\Phi(v)-v\right| \le T\sum_{\ell=1}^{n}\left|\Phi^{\ell}(v)-v^{\ell}\right| \le (3/2)\sum_{\ell=1}^{n}\left|x^{\ell}-y^{\ell}\right| . \tag{111}$$

Proof. Let $u=\Phi(v)$. From $\tilde{q}_T(x,v)=\tilde{q}_T(y,u)$,

$$\begin{aligned}T\left|u^{\ell}-v^{\ell}\right| &\le \left|x^{\ell}-y^{\ell}\right|+(L+2\epsilon\tilde{L})T^2\max_{s\le T}\left|\tilde{q}_s^i(x,v)-\tilde{q}_s^i(y,u)\right| \\ &\quad+\frac{2\epsilon\tilde{L}T^2}{n}\sum_{i=1,i\ne\ell}^{n}\left|\tilde{q}_s^i(x,v)-\tilde{q}_s^i(y,u)\right| .\end{aligned}$$

Summing over ℓ and using (90) and $(L+4\epsilon\tilde{L})(T^2+Th)\le 1/6$ gives

$$\begin{aligned} T\sum_{\ell=1}^n \left|u^\ell - v^\ell\right| &\le \sum_{\ell=1}^n \left|x^\ell - y^\ell\right| + (L+4\epsilon\tilde{L})T^2 \sum_{\ell=1}^n \max_{s\le T}\left|\tilde{q}_s^\ell(x,v)-\tilde{q}_s^\ell(y,u)\right| \\ &\le \sum_{\ell=1}^n \left|x^\ell - y^\ell\right| + (7/36)\sum_{\ell=1}^n \left(\left|x^\ell - y^\ell\right| + T\left|u^\ell - v^\ell\right|\right) \le (3/2)\sum_{\ell=1}^n \left|x^\ell - y^\ell\right| \end{aligned}$$

where in the last step we used $(1+7/36)/(1-7/36)<3/2$.

□

Lemma 38. *For any $x,y,v\in\mathbb{R}^{nk}$ such that $\tilde{q}_T(x,v)=\tilde{q}_T(y,\Phi(v))$, we have that $\|D\Phi(v)-I_d\|\le 1/2$ and*

$$\|D\Phi(v)-I_d\|_F \le \frac{49}{8}\sqrt{k}\,(L_H+8\epsilon\tilde{L}_H)\,T^2\sum_{\ell=1}^n \left|x^\ell - y^\ell\right| . \tag{112}$$

Proof. First, note that

$$\|D\Phi(v)-I_d\|_F = \left(\sum_{\ell=1}^n\sum_{i=1}^n \left\|\partial_{v^\ell}\Phi^i(v)-\delta_{i\ell}I_k\right\|_F^2\right)^{1/2} \le \sqrt{k}\sum_{\ell=1}^n\sum_{i=1}^n \left\|\partial_{v^\ell}\Phi^i(v)-\delta_{i\ell}I_k\right\| .$$

Next, and in turn, we upper bound $\|D\Phi(v)-I_d\|$ and $\sum_{\ell=1}^n\sum_{i=1}^n \left\|\partial_{v^\ell}\Phi^i(v)-\delta_{i\ell}I_k\right\|$. Introduce the shorthand $\tilde{q}_s^{(1)} := \tilde{q}_s(x,v)$ and $\tilde{q}_s^{(2)} := \tilde{q}_s(y,\Phi(v))$ for any $s\in[0,T]$. Differentiating both sides of $\tilde{q}_T^\ell(x,v)=\tilde{q}_T^\ell(y,\Phi(v))$ with respect to v^j yields

$$\begin{aligned} &T(\partial_{v^\ell}\Phi^i(v)-\delta_{i\ell}I_k) = \\ &+\int_0^T \frac{T-s}{2}\sum_{j=1}^n \Big[\mathbf{H}_{ij}(\tilde{q}^{(2)}_{\lfloor s\rfloor_h})\partial_{v^\ell}\tilde{q}^{(2),j}_{\lfloor s\rfloor_h} - \mathbf{H}_{ij}(\tilde{q}^{(1)}_{\lfloor s\rfloor_h})\partial_{v^\ell}\tilde{q}^{(1),j}_{\lfloor s\rfloor_h}\Big]ds \\ &+\int_0^T \frac{T-s}{2}\sum_{j=1}^n \Big[\mathbf{H}_{ij}(\tilde{q}^{(2)}_{\lceil s\rceil_h})\partial_{v^\ell}\tilde{q}^{(2),j}_{\lceil s\rceil_h} - \mathbf{H}_{ij}(\tilde{q}^{(1)}_{\lceil s\rceil_h})\partial_{v^\ell}\tilde{q}^{(1),j}_{\lceil s\rceil_h}\Big]ds \\ &-\int_0^T \frac{s-\lfloor s\rfloor_h}{2}\sum_{j=1}^n \Big[\mathbf{H}_{ij}(\tilde{q}^{(2)}_{\lceil s\rceil_h})\partial_{v^\ell}\tilde{q}^{(2),j}_{\lceil s\rceil_h} - \mathbf{H}_{ij}(\tilde{q}^{(1)}_{\lceil s\rceil_h})\partial_{v^\ell}\tilde{q}^{(1),j}_{\lceil s\rceil_h}\Big]ds \\ &-\int_0^T \frac{s-\lfloor s\rfloor_h}{2}\sum_{j=1}^n \Big[\mathbf{H}_{ij}(\tilde{q}^{(1)}_{\lfloor s\rfloor_h})\partial_{v^\ell}\tilde{q}^{(1),j}_{\lfloor s\rfloor_h} - \mathbf{H}_{ij}(\tilde{q}^{(2)}_{\lfloor s\rfloor_h})\partial_{v^\ell}\tilde{q}^{(2),j}_{\lfloor s\rfloor_h}\Big]ds \\ &+\int_0^T \frac{T-s}{2}\sum_{j,m} \Big[\mathbf{H}_{ij}(\tilde{q}^{(2)}_{\lfloor s\rfloor_h})\partial_{v^m}\tilde{q}^{(2),j}_{\lfloor s\rfloor_h} + \mathbf{H}_{ij}(\tilde{q}^{(2)}_{\lceil s\rceil_h})\partial_{v^m}\tilde{q}^{(2),j}_{\lceil s\rceil_h}\Big](\partial_{v^\ell}\Phi^m(v)-\delta_{m\ell}I_k)ds \\ &-\int_0^T \frac{s-\lfloor s\rfloor_h}{2}\sum_{j,m} \Big[\mathbf{H}_{ij}(\tilde{q}^{(2)}_{\lceil s\rceil_h})\partial_{v^m}\tilde{q}^{(2),j}_{\lceil s\rceil_h} - \mathbf{H}_{ij}(\tilde{q}^{(2)}_{\lfloor s\rfloor_h})\partial_{v^m}\tilde{q}^{(2),j}_{\lfloor s\rfloor_h}\Big](\partial_{v^\ell}\Phi^m(v)-\delta_{m\ell}I_k)ds. \end{aligned} \tag{113}$$

By using (79), (91), and $(L+4\epsilon\tilde{L})T^2 \le 1/6$, for fixed $\ell \in \{1,\dots,n\}$ note that

$$\sum_{i=1}^{n} \left\| \partial_{v^\ell} \Phi^i(v) - \delta_{i\ell} I_k \right\| \le \frac{(6/5)2(L+4\epsilon\tilde{L})T^2}{(1-(6/5)(L+4\epsilon\tilde{L})T^2)} = 1/2 \, ,$$

and therefore, for any $z = (z^1,\dots,z^n) \in \mathbb{R}^{nk}$, we have

$$\left|(D\Phi(v) - I_d)z\right|^2 \le \sum_{\ell=1}^{n}\sum_{i=1}^{n} \left|z^\ell\right|^2 \left\|\partial_{v^\ell}\Phi^i(v) - \delta_{i\ell}^2 I_k\right\|^2 \le \frac{1}{4}\sum_{\ell=1}^{n}\left|z^\ell\right|^2 .$$

Thus, $\|D\Phi(v) - I_d\| \le 1/2$. We can also rewrite (113) as

$$\begin{aligned}
&T(\partial_{v^\ell}\Phi^i(v) - \delta_{i\ell} I_k) = \\
&+ \int_0^T \frac{T-s}{2} \sum_{j=1}^{n} \Big[\mathbf{H}_{ij}(\tilde{q}^{(2)}_{\lfloor s\rfloor_h})(\partial_{v^\ell}\tilde{q}^{(2),j}_{\lfloor s\rfloor_h} - \partial_{v^\ell}\tilde{q}^{(1),j}_{\lfloor s\rfloor_h}) - (\mathbf{H}_{ij}(\tilde{q}^{(1)}_{\lfloor s\rfloor_h}) - \mathbf{H}_{ij}(\tilde{q}^{(2)}_{\lfloor s\rfloor_h}))\partial_{v^\ell}\tilde{q}^{(1),j}_{\lfloor s\rfloor_h} \Big] ds \\
&+ \int_0^T \frac{T-s}{2} \sum_{j=1}^{n} \Big[\mathbf{H}_{ij}(\tilde{q}^{(2)}_{\lceil s\rceil_h})(\partial_{v^\ell}\tilde{q}^{(2),j}_{\lceil s\rceil_h} - \partial_{v^\ell}\tilde{q}^{(1),j}_{\lceil s\rceil_h}) - (\mathbf{H}_{ij}(\tilde{q}^{(1)}_{\lceil s\rceil_h}) - \mathbf{H}_{ij}(\tilde{q}^{(2)}_{\lceil s\rceil_h}))\partial_{v^\ell}\tilde{q}^{(1),j}_{\lceil s\rceil_h} \Big] ds \\
&- \int_0^T \frac{s-\lfloor s\rfloor_h}{2} \sum_{j=1}^{n} \Big[\mathbf{H}_{ij}(\tilde{q}^{(2)}_{\lceil s\rceil_h})(\partial_{v^\ell}\tilde{q}^{(2),j}_{\lceil s\rceil_h} - \partial_{v^\ell}\tilde{q}^{(1),j}_{\lceil s\rceil_h}) - (\mathbf{H}_{ij}(\tilde{q}^{(1)}_{\lceil s\rceil_h}) - \mathbf{H}_{ij}(\tilde{q}^{(2)}_{\lceil s\rceil_h}))\partial_{v^\ell}\tilde{q}^{(1),j}_{\lceil s\rceil_h} \Big] ds \\
&+ \int_0^T \frac{s-\lfloor s\rfloor_h}{2} \sum_{j=1}^{n} \Big[\mathbf{H}_{ij}(\tilde{q}^{(2)}_{\lfloor s\rfloor_h})(\partial_{v^\ell}\tilde{q}^{(2),j}_{\lfloor s\rfloor_h} - \partial_{v^\ell}\tilde{q}^{(1),j}_{\lfloor s\rfloor_h}) - (\mathbf{H}_{ij}(\tilde{q}^{(1)}_{\lfloor s\rfloor_h}) - \mathbf{H}_{ij}(\tilde{q}^{(2)}_{\lfloor s\rfloor_h}))\partial_{v^\ell}\tilde{q}^{(1),j}_{\lfloor s\rfloor_h} \Big] ds \\
&+ \int_0^T \frac{T-s}{2} \sum_{j,m} \Big[\mathbf{H}_{ij}(\tilde{q}^{(2)}_{\lfloor s\rfloor_h})\partial_{v^m}\tilde{q}^{(2),j}_{\lfloor s\rfloor_h} + \mathbf{H}_{ij}(\tilde{q}^{(2)}_{\lceil s\rceil_h})\partial_{v^m}\tilde{q}^{(2),j}_{\lceil s\rceil_h} \Big] (\partial_{v^\ell}\Phi^m(v) - \delta_{m\ell} I_k) ds \\
&- \int_0^T \frac{s-\lfloor s\rfloor_h}{2} \sum_{j,m} \Big[\mathbf{H}_{ij}(\tilde{q}^{(2)}_{\lceil s\rceil_h})\partial_{v^m}\tilde{q}^{(2),j}_{\lceil s\rceil_h} - \mathbf{H}_{ij}(\tilde{q}^{(2)}_{\lfloor s\rfloor_h})\partial_{v^m}\tilde{q}^{(2),j}_{\lfloor s\rfloor_h} \Big] (\partial_{v^\ell}\Phi^m(v) - \delta_{m\ell} I_k) ds.
\end{aligned}$$

By using (79), (80), (91), (92), (94), and $(L+4\epsilon\tilde{L})T^2 \le 1/6$, we get

$$\begin{aligned}
(4/5)T\sum_{i,\ell}\left\|\partial_{v^\ell}\Phi^i(v) - \delta_{i\ell} I_k\right\| &\le (L+4\epsilon\tilde{L})T^2 \sum_{j,\ell} \max_{s\le T} \left\|\partial_{v^\ell}\tilde{q}^{(2),j}_s - \partial_{v^\ell}\tilde{q}^{(1),j}_s\right\| \\
&\quad + (7/5)T(L_H + 8\epsilon\tilde{L}_H)T^2 \sum_{\ell=1}^{n} \max_{s\le T}\left|\tilde{q}^{(1),\ell}_s - \tilde{q}^{(2),\ell}_s\right| \\
&\le (49/25)(L_H + 8\epsilon\tilde{L}_H)T^3 \sum_{\ell=1}^{n} (\left|x^\ell - y^\ell\right| + T\left|\Phi^\ell(v) - v^\ell\right|) \\
&\le (49/10)(L_H + 8\epsilon\tilde{L}_H)T^3 \sum_{\ell=1}^{n} \left|x^\ell - y^\ell\right|
\end{aligned}$$

where in the last step we inserted (111). Simplifying gives (112). □

A.7. One-shot coupling bounds for $\tilde{q}_T(x,\xi) = q_T(x,\Phi(\xi))$

The following lemmas are the mean-field analogs of Lemmas 27 and 28.

Lemma 39. *For any $x, v \in \mathbb{R}^{nk}$ such that $\tilde{q}_T(x,v) = q_T(x,\Phi(v))$, we have*

$$\begin{aligned} T\left|\Phi(v)-v\right| &\le T\sum_{\ell=1}^n \left|\Phi^\ell(v)-v^\ell\right| \le \frac{72}{65}h^2\Big((L+4\epsilon\tilde{L})\sum_{\ell=1}^n\left|x^\ell\right| \\ &+ (L+4\epsilon\tilde{L})T\sum_{\ell=1}^n\left|v^\ell\right| + (L_H+8\epsilon\tilde{L}_H)\sum_{\ell=1}^n\left|x^\ell\right|^2 + (L_H+8\epsilon\tilde{L}_H)T\sum_{\ell=1}^n\left|v^\ell\right|^2\Big). \end{aligned} \tag{114}$$

Proof. Introduce the shorthand $\tilde{q}_T = \tilde{q}_T(x,v)$, $q_T^{(1)} = q_T(x,v)$ and $q_T^{(2)} = q_T(x,\Phi(v))$. Noting that $|\Phi(v)-v| \le \sum_{\ell=1}^n \left|\Phi^\ell(v)-v^\ell\right|$, and using $\tilde{q}_T = q_T^{(2)}$ or $q_T^{(2)} - q_T^{(1)} = \tilde{q}_T - q_T^{(1)}$ and then applying (78), we obtain

$$\begin{aligned} T\sum_{\ell=1}^n\left|\Phi^\ell(v)-v^\ell\right| &\le \sum_{\ell=1}^n\left|\int_0^T (T-s)[\nabla_\ell U(q_s^{(2)}) - \nabla_\ell U(q_s^{(1)})]ds + \tilde{q}_T^\ell - q_T^{(1),\ell}\right| \\ &\le (L+4\epsilon\tilde{L})\frac{T^2}{2}\sum_{\ell=1}^n\max_{s\le T}\left|q_s^{(2),\ell}-q_s^{(1),\ell}\right| + \sum_{\ell=1}^n\left|\tilde{q}_T^\ell - q_T^{(1),\ell}\right| \\ &\le \frac{7}{72}T\sum_{\ell=1}^n\left|\Phi^\ell(v)-v^\ell\right| + \sum_{\ell=1}^n\left|\tilde{q}_T^\ell - q_T^{(1),\ell}\right| \le \frac{72}{65}\sum_{\ell=1}^n\left|q_T^{(1),\ell}-\tilde{q}_T^\ell\right| \end{aligned}$$

where in the next to last step we used (90) and $(L+4\epsilon\tilde{L})T^2 \le 1/6$. Inserting (96) into this last inequality gives (114). □

Lemma 40. *For any $x, v \in \mathbb{R}^{nk}$ such that $\tilde{q}_T(x,v) = q_T(x,\Phi(v))$, we have that $\|D\Phi(v) - I_d\| \le 1/2$ and*

$$\begin{aligned} T\left\|D\Phi(v)-I_d\right\|_F &\le \sqrt{k}\sum_{\ell=1}^n\sum_{i=1}^n\left\|\partial_{v^\ell}\Phi^i(v)-\delta_{i\ell}I_k\right\| \\ &\le \sqrt{k}h^2\Bigg((L+4\epsilon\tilde{L})Tn + (L_H+8\epsilon\tilde{L}_H)T\sum_{\ell=1}^n\left|x^\ell\right| + 3(L_H+8\epsilon\tilde{L}_H)T^2\sum_{\ell=1}^n\left|v^\ell\right| \\ &\quad + \Big((L_I+14\epsilon\tilde{L}_I)T + 2(L_H+8\epsilon\tilde{L}_H)^2T^3\Big)\sum_{\ell=1}^n\left|x^\ell\right|^2 \\ &\quad + \Big((L_I+14\epsilon\tilde{L}_I)T^3 + 3(L_H+8\epsilon\tilde{L}_H)^2T^5\Big)\sum_{\ell=1}^n\left|v^\ell\right|^2\Bigg). \end{aligned} \tag{115}$$

Proof. The proof that $\|D\Phi(v) - I_d\| \le 1/2$ is similar to the proof in Lemma 38 and therefore omitted. Note that

$$\left\|D\Phi(v)-I_d\right\|_F = \left(\sum_{\ell=1}^n\sum_{i=1}^n\left\|\partial_{v^\ell}\Phi^i(v)-\delta_{i\ell}I_k\right\|_F^2\right)^{1/2} \le \sqrt{k}\sum_{\ell=1}^n\sum_{i=1}^n\left\|\partial_{v^\ell}\Phi^i(v)-\delta_{i\ell}I_k\right\|.$$

Next, we upper bound $\sum_{\ell=1}^n \sum_{i=1}^n \left\| \partial_{v^\ell} \Phi^i(v) - \delta_{i\ell} I_k \right\|$. Introduce the shorthand $\tilde{q}_T = \tilde{q}_T(x,v)$, $q_T^{(1)} = q_T(x,v)$ and $q_T^{(2)} = q_T(x,\Phi(v))$. The derivative of $q_T^{(2),i} - q_T^{(1),i} = \tilde{q}_T^i - q_T^{(1),i}$ with respect to v^ℓ yields

$$
\begin{aligned}
T\left(\partial_{v^\ell}\Phi^i(v) - \delta_{i\ell} I_k\right) \;&=\; \partial_{v^\ell}\tilde{q}_T^i - \partial_{v^\ell} q_T^{(1),i} \\
&+ \int_0^T (T-s) \sum_{j=1}^n [\mathbf{H}_{ij}(q_s^{(2)}) - \mathbf{H}_{ij}(q_s^{(1)})] \partial_{v^\ell} q_s^{(1),j} ds \\
&+ \int_0^T (T-s) \sum_{j=1}^n \mathbf{H}_{ij}(q_s^{(2)}) [\partial_{v^\ell} q_s^{(2),j} - \partial_{v^\ell} q_s^{(1),j}] ds \\
&+ \int_0^T (T-s) \sum_{j=1}^n \sum_{m=1}^n \mathbf{H}_{ij}(q_s^{(2)}) \partial_{v^m} q_s^{(2),j} (\partial_{v^\ell}\Phi^m(v) - \delta_{\ell m} I_k)] ds.
\end{aligned}
$$

By (79), (80), (91), (92) and $(L + 4\epsilon\tilde{L})T^2 \le 1/6$,

$$
\begin{aligned}
& T \sum_{i,\ell} \left\| \partial_{v^\ell}\Phi^i(v) - \delta_{i\ell} I_k \right\| \\
&\le \sum_{i,\ell} \left\| \partial_{v^\ell}\tilde{q}_T^i - \partial_{v^\ell} q_T^{(1),i} \right\| + \frac{7}{5} T (L_H + 8\epsilon \tilde{L}_H) \frac{T^2}{2} \sum_\ell \max_{s \le T} \left| q_s^{(2),\ell} - q_s^{(1),\ell} \right| \\
&\quad + (L + 4\epsilon\tilde{L}) \frac{T^2}{2} \sum_{i,\ell} \left| \partial_{v^\ell} q_s^{(2),i} - \partial_{v^\ell} q_s^{(1),i} \right| + \frac{6}{5} T (L + 4\epsilon\tilde{L}) \frac{T^2}{2} \sum_{i,\ell} \left\| \partial_{v^\ell}\Phi^i(v) - \delta_{i\ell} I_k \right\| \\
&\le \frac{10}{9} \sum_{i,\ell} \left\| \partial_{v^\ell}\tilde{q}_T^i - \partial_{v^\ell} q_T^{(1),i} \right\| + \frac{10}{9}\frac{7}{5} T (L_H + 8\epsilon\tilde{L}_H) \frac{T^2}{2} \sum_\ell \max_{s\le T} \left| q_s^{(2),\ell} - q_s^{(1),\ell} \right| \\
&\quad + \frac{10}{9}\frac{1}{12} \sum_{i,\ell} \left\| \partial_{v^\ell} q_s^{(2),i} - \partial_{v^\ell} q_s^{(1),i} \right\| \\
&\le \frac{10}{9} \sum_{i,\ell} \left\| \partial_{v^\ell}\tilde{q}_T^i - \partial_{v^\ell} q_T^{(1),i} \right\| + \frac{49}{45} (L_H + 8\epsilon\tilde{L}_H) T^3 \left(T \sum_{\ell=1}^n \left| \Phi^\ell(v) - v^\ell \right| \right)
\end{aligned}
$$

where in the last step we used (90) and (94). Inserting (105) and (114) and simplifying gives (115). □

Acknowledgements

We wish to acknowledge Katharina Schuh for useful discussions.

N. Bou-Rabee has been supported by the Alexander von Humboldt foundation and the National Science Foundation under Grant No. DMS-1816378.

A. Eberle has been supported by the Hausdorff Center for Mathematics. Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) im Rahmen der Exzellenzstrategie des Bundes und der Länder - GZ 2047/1, Projekt-ID 390685813.

References

[1] S. Armstrong and J.-C. Mourrat. *Variational methods for the kinetic Fokker-Planck equation*. 2019. arXiv: 1902.04037 [math.AP] (cit. on p. 1).

[2] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and Geometry of Markov Diffusion Operators*. Grundlehren der mathematischen Wissenschaften. Springer, 2013. ISBN: 9783319002279. URL: https://books.google.de/books?id=gU3ABAAAQBAJ (cit. on p. 1).

[3] N. Bou-Rabee and J. M. Sanz-Serna. "Geometric Integrators and the Hamiltonian Monte Carlo Method". In: *Acta Numerica* 27 (2018), pp. 113–206 (cit. on p. 2).

[4] N. Bou-Rabee and A. Eberle. "Two-scale coupling for preconditioned Hamiltonian Monte Carlo in infinite dimensions". In: *Stochastics and Partial Differential Equations: Analysis and Computations* 9.1 (2021), pp. 207–242. DOI: 10.1007/s40072-020-00175-6 (cit. on p. 2).

[5] N. Bou-Rabee, A. Eberle, and R. Zimmer. "Coupling and convergence for Hamiltonian Monte Carlo". In: *Ann. Appl. Probab.* 30.3 (2020), pp. 1209–1250 (cit. on pp. 2, 9, 20).

[6] N. Bou-Rabee and K. Schuh. "Convergence of Unadjusted Hamiltonian Monte Carlo for Mean-Field Models". arXiv preprint arXiv:2009.08735. 2020 (cit. on pp. 2, 7, 10–12, 32).

[7] Y. Chen, R. Dwivedi, M. J. Wainwright, and B. Yu. "Fast mixing of Metropolized Hamiltonian Monte Carlo: Benefits of multi-step gradients". In: *Journal of Machine Learning Research* 21.92 (2020), pp. 1–72 (cit. on pp. 2, 3, 16).

[8] Z. Chen and S. S. Vempala. "Optimal convergence rate of Hamiltonian Monte Carlo for strongly logconcave distributions". In: *arXiv preprint arXiv:1905.02313* (2019) (cit. on pp. 2, 9).

[9] S. Chewi, C. Lu, K. Ahn, X. Cheng, T. L. Gouic, and P. Rigollet. "Optimal dimension dependence of the Metropolis-Adjusted Langevin Algorithm". In: *arXiv preprint arXiv:2012.12810* (2020) (cit. on p. 3).

[10] J. Dolbeault, C. Mouhot, and C. Schmeiser. "Hypocoercivity for linear kinetic equations conserving mass". In: *Trans. Amer. Math. Soc.* 367.6 (2015), pp. 3807–3828. ISSN: 0002-9947. DOI: 10.1090/S0002-9947-2015-06012-7. URL: https://doi.org/10.1090/S0002-9947-2015-06012-7 (cit. on p. 1).

[11] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth. "Hybrid Monte-Carlo". In: *Phys Lett B* 195 (1987), pp. 216–222 (cit. on p. 2).

[12] A. Durmus and A. Eberle. "Asymptotic bias for Markov Chain Monte Carlo Methods in High Dimension". arXiv: 2021 (cit. on p. 10).

[13] A. Durmus and É. Moulines. "High-dimensional Bayesian inference via the unadjusted Langevin algorithm". In: *Bernoulli* 25.4A (Nov. 2019), pp. 2854–2882 (cit. on p. 4).

[14] A. Eberle. "Reflection couplings and contraction rates for diffusions". In: *Probability theory and related fields* 166.3-4 (2016), pp. 851–886 (cit. on p. 1).

[15] A. Eberle, A. Guillin, and R. Zimmer. "Couplings and quantitative contraction rates for Langevin dynamics". In: *Ann. Probab.* 47.4 (July 2019), pp. 1982–2010. DOI: 10.1214/18-AOP1299. URL: https://doi.org/10.1214/18-AOP1299 (cit. on p. 1).

[16] A. Eberle and M. B. Majka. "Quantitative contraction rates for Markov chains on general state spaces". In: *Electronic Journal of Probability* 24 (2019) (cit. on pp. 1, 4).

[17] S. Friedland and L.-H. Lim. "Nuclear norm of higher-order tensors". In: *Mathematics of Computation* 87.311 (2018), pp. 1255–1281 (cit. on p. 7).

[18] E. Hairer, C. Lubich, and G. Wanner. *Geometric Numerical Integration*. Springer, 2010 (cit. on p. 2).

[19] M. Hairer and J. C. Mattingly. "Ergodicity of the 2D Navier-Stokes equations with degenerate stochastic forcing". In: *Annals of Mathematics* (2006), pp. 993–1032 (cit. on p. 1).

[20] M. Hairer, J. C. Mattingly, and M. Scheutzow. "Asymptotic coupling and a general form of Harris' theorem with applications to stochastic delay equations". In: *Probab. Theory Related Fields* 149.1-2 (2011), pp. 223–259 (cit. on p. 1).

[21] J. Heng and P. E. Jacob. "Unbiased Hamiltonian Monte Carlo with couplings". In: *Biometrika* 106.2 (2019), pp. 287–302 (cit. on p. 4).

[22] M. Kac. "Foundations of Kinetic Theory". In: *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 3: Contributions to Astronomy and Physics*. Berkeley, Calif.: University of California Press, 1956, pp. 171–197. URL: https://projecteuclid.org/euclid.bsmsp/1200502194 (cit. on p. 10).

[23] Y. T. Lee and S. S. Vempala. "Convergence rate of Riemannian Hamiltonian Monte Carlo and faster polytope volume computation". In: *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*. 2018, pp. 1115–1121 (cit. on p. 3).

[24] Y. T. Lee and S. S. Vempala. "Geodesic walks in polytopes". In: *Proceedings of the 49th Annual ACM SIGACT Symposium on theory of Computing*. 2017, pp. 927–940 (cit. on p. 3).

[25] B. Leimkuhler and S. Reich. *Simulating Hamiltonian Dynamics*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, 2004 (cit. on p. 2).

[26] D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov chains and mixing times*. American Mathematical Soc., 2009 (cit. on p. 1).

[27] T. Lindvall. *Lectures on the coupling method*. Corrected reprint of the 1992 original. Dover Publications, Inc., Mineola, NY, 2002, pp. xiv+257. ISBN: 0-486-42145-7 (cit. on p. 1).

[28] T. Lindvall and L. C. G. Rogers. "Coupling of multidimensional diffusions by reflection". In: *Ann. Probab.* 14.3 (1986), pp. 860–872. ISSN: 0091-1798. URL: http://links.jstor.org/sici?sici=0091-1798(198607)14:3%3C860:COMDBR%3E2.0.CO;2-V&origin=MSN (cit. on p. 1).

[29] J. Lu and L. Wang. *On explicit L^2-convergence rate estimate for piecewise deterministic Markov processes in MCMC algorithms*. 2021. arXiv: 2007.14927 [math.PR] (cit. on p. 1).

[30] N. Madras and D. Sezer. "Quantitative bounds for Markov chain convergence: Wasserstein and total variation distances". In: *Bernoulli* 16.3 (2010), pp. 882–908 (cit. on pp. 2, 17, 18).

[31] O. Mangoubi and A. Smith. "Rapid mixing of Hamiltonian Monte Carlo on strongly log-concave distributions". In: *arXiv preprint arXiv:1708.07114* (2017) (cit. on pp. 2, 9).

[32] J. E. Marsden and M. West. "Discrete Mechanics and Variational Integrators". In: *Acta Numerica* 10 (2001), pp. 357–514 (cit. on pp. 2, 13).

[33] J. C. Mattingly, A. M. Stuart, and M. V. Tretyakov. "Convergence of numerical time-averaging and stationary measures via Poisson equations". In: *SIAM J Num Anal* 48.2 (2010), pp. 552–577 (cit. on p. 7).

[34] S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability*. 2nd. New York, NY: Cambridge University Press, 2009 (cit. on p. 1).

[35] S. Mischler and C. Mouhot. "Kac's program in kinetic theory". In: *Inventiones mathematicae* 193.1 (2013), pp. 1–147. DOI: 10.1007/s00222-012-0422-3. URL: https://doi.org/10.1007/s00222-012-0422-3 (cit. on p. 10).

[36] P. Monmarché. "High-dimensional MCMC with a standard splitting scheme for the underdamped Langevin". In: *arXiv preprint arXiv:2007.05455* (2020) (cit. on p. 3).

[37] R. Montenegro and P. Tetali. "Mathematical aspects of mixing times in Markov chains". In: *Found. Trends Theor. Comput. Sci.* 1.3 (2006), pp. x+121. ISSN: 1551-305X. DOI: 10.1561/0400000003. URL: https://doi.org/10.1561/0400000003 (cit. on p. 1).

[38] S. Müller and M. Ortiz. "On the Γ-convergence of discrete dynamics and variational integrators". In: *Journal of Nonlinear Science* 14.3 (2004), pp. 279–296 (cit. on p. 14).

[39] R. M. Neal. "MCMC using Hamiltonian dynamics". In: *Handbook of Markov Chain Monte Carlo* 2 (2011), pp. 113–162 (cit. on p. 2).

[40] S. M. Rump. "Estimates of the determinant of a perturbed identity matrix". In: *Linear algebra and its applications* 558 (2018), pp. 101–107 (cit. on p. 16).

[41] L. Saloff-Coste. "Lectures on finite Markov chains". In: *Lectures on probability theory and statistics (Saint-Flour, 1996)*. Vol. 1665. Lecture Notes in Math. Springer, Berlin, 1997, pp. 301–413. DOI: 10.1007/BFb0092621. URL: https://doi.org/10.1007/BFb0092621 (cit. on p. 1).

[42] S. Vempala. "Geometric random walks: a survey". In: *Combinatorial and computational geometry*. Vol. 52. Math. Sci. Res. Inst. Publ. Cambridge Univ. Press, Cambridge, 2005, pp. 577–616 (cit. on p. 1).

[43] C. Villani. "Hypocoercivity". In: *Mem. Amer. Math. Soc.* 202.950 (2009), pp. iv+141. ISSN: 0065-9266. DOI: 10.1090/S0065-9266-09-00567-5. URL: https://doi.org/10.1090/S0065-9266-09-00567-5 (cit. on p. 1).

[44] C. Villani. *Optimal transport: old and new*. Vol. 338. Springer Science & Business Media, 2008 (cit. on p. 15).