Couplings for Andersen dynamics

Nawaf Bou-Rabee[1,a] and Andreas Eberle[2,b]

[1]*Department of Mathematical Sciences, Rutgers University Camden, 311 N 5th Street, Camden, NJ 08102, USA.* [a]*nawaf.bourabee@rutgers.edu*
[2]*Institut für Angewandte Mathematik, Universität Bonn, Endenicher Allee 60, 53115 Bonn, Germany.* [b]*eberle@uni-bonn.de*

Abstract. Andersen dynamics is a standard method for molecular simulations, and a precursor of the Hamiltonian Monte Carlo algorithm used in MCMC inference. The stochastic process corresponding to Andersen dynamics is a PDMP (piecewise deterministic Markov process) that iterates between Hamiltonian flows and velocity randomizations of randomly selected particles. Both from the viewpoint of molecular dynamics and MCMC inference, a basic question is to understand the convergence to equilibrium of this PDMP particularly in high dimension. Here we introduce a coupling approach to derive explicit convergence bounds in a Wasserstein sense. The bounds are dimension free for not necessarily convex potentials with weakly interacting components on a high dimensional torus, and for strongly convex and gradient Lipschitz potentials on a Euclidean product space.

Résumé. La dynamique d'Andersen est une méthode standard pour les simulations moléculaires et un précurseur de l'algorithme de Monte Carlo Hamiltonien utilisé dans l'inférence MCMC. Le processus stochastique correspondant à la dynamique d'Andersen est un PDMP (processus de Markov déterministe par morceaux) qui itère entre les écoulements hamiltoniens et les randomisations de vitesse de particules sélectionnées au hasard. Tant du point de vue de la dynamique moléculaire que de l'inférence MCMC, une question fondamentale est de comprendre la convergence vers l'équilibre de ce PDMP, surtout en dimension supérieure. Nous présentons ici des couplages pour obtenir des bornes de convergence au sens de Wasserstein qui ne nécessitent pas de convexité globale de l'énergie potentielle sous-jacente.

1. Introduction

A common task in molecular dynamics is to simulate a molecular system at a specified temperature [2,30]. The first method suggested for this purpose goes back to Andersen [3,26]. The stochastic process corresponding to Andersen dynamics is a piecewise deterministic Markov process (PDMP) [19,20] that combines Hamiltonian trajectories with velocity randomizations of randomly selected particles such that the resulting PDMP leaves the canonical or Boltzmann-Gibbs distribution invariant [26,38]. The durations between consecutive velocity randomizations are i.i.d. exponential random variables with constant mean determined by a collision frequency parameter, and in between these velocity randomizations, the PDMP follows pure Hamiltonian dynamics. Andersen dynamics is currently implemented in several molecular dynamics software packages including AMBER and GROMACS [1,31] and because of its simplicity and reliability continues to be employed in a wide variety of molecular dynamics simulations [5,10,46,55,56].

Besides molecular dynamics, Andersen dynamics plays an important conceptual role in Markov chain Monte Carlo (MCMC) inference. Indeed, Hamiltonian Monte Carlo (HMC) can be viewed as a refinement of Andersen dynamics to include a Metropolis accept/reject step [44]. Due to the ability of HMC to overcome the diffusive behavior that limits more conventional MCMC methods like Gibbs, random walk Metropolis and the Metropolis adjusted Langevin algorithm, HMC has garnered a great deal of attention in Bayesian statistics [12,25,32,35,39,42,45].

Both from the viewpoint of molecular dynamics and MCMC inference, a basic question with Andersen dynamics is to understand the convergence to equilibrium as a function of the collision frequency parameter particularly in high dimension. If the collision frequency is too small, then on average the integration times of the Hamiltonian trajectories are very long and the PDMP mainly follows Hamiltonian dynamics which by itself is not ergodic in general; whereas if

the collision frequency is too high, then the PDMP will exhibit diffusive behavior and it will again take a long time to sufficiently converge. Nevertheless, like other processes that involve Hamiltonian dynamics, one may hope that Andersen dynamics can achieve faster convergence than random walk based methods if the collision frequency is suitably chosen.

First steps to understand the convergence of Andersen dynamics in terms of the collision frequency have been taken. Mixing time bounds for Andersen dynamics on a torus were derived in [26] by showing that Doeblin's condition holds, and subsequently better bounds were obtained in Theorem 6.5 of [37] in the 'free-streaming' case where the potential is switched off.

On the other hand, the Fokker–Planck equation corresponding to Andersen dynamics is a linear Boltzmann equation with a specific collision kernel. The convergence to equilibrium of linear Boltzmann equations has been analyzed based on hypocoercivity techniques [23,33], see also [57]. The corresponding results rely on a standard Poincaré inequality, and as a consequence, the resulting convergence rates seem not to capture the improvements of kinetic versus diffusive behaviour in an adequate way. Recently, Armstrong and Mourrat [4] developed a new approach based on a Poincaré inequality in space-time that seems to overcome this problem. It has been shown by Cao, Lu and Wang [15,40] that for an appropriate class of potentials, this approach provides convergence rates of the correct order both for Langevin dynamics and for randomized HMC. An advantage of these analytic approaches is that they apply whenever certain functional inequalities are satisfied. A significant drawback is, however, that convergence to equilibrium can only be proven uniformly for initial laws with an L^2 bounded density, i.e., for a "warm start". This is relevant, because in the high dimensional case, the initial L^2 norm typically depends exponentially on the dimension. Whereas in the reversible case, the dependence on the initial law can at least be partially relaxed by applying logarithmic Sobolev inequalities, it is still open how to implement such an approach efficiently in the hypocoercive setup considered here.

Here we follow an alternative probabilistic approach based on coupling techniques. These techniques are based on the framework introduced in [27], and can be viewed as a continuous-time analogue on phase space of recently developed couplings for HMC applied to general non-convex models [12] and high-dimensional mean-field models [14]. The coupling used is itself a PDMP, and at least formally, the analysis is based on bounding the action of the generator of the coupling process on distances tailored to each system considered. An advantage of these coupling methods is that they provide explicit bounds on convergence to equilibrium in L^1 Wasserstein distances for arbitrary initial conditions and do not require a warm start. Moreover, couplings provide a different and more explicit understanding of the way in which convergence to equilibrium takes place. For example, it is possible to simulate couplings and to use them as an additional tool for convergence diagnostics, variance reduction, or unbiased estimation [32,50,55]. Finally, although the results below are stated in the case where the force is the gradient of a potential, they immediately carry over to the non-gradient case for which the current analytic approaches are not applicable.

We consider Andersen dynamics for systems with weakly anharmonic (i.e., strongly convex and gradient Lipschitz) potential energies in an unbounded space, and not necessarily convex, twice continuously differentiable potential energies on a high-dimensional torus with weak interactions between particles. In these settings, we obtain quantitative bounds for the convergence of Andersen dynamics in an L^1 Wasserstein sense with explicit rates that do not depend on the number of particles. The bounds reveal that if the collision frequency is suitably chosen, then Andersen dynamics can overcome diffusive convergence behavior.

We end this introduction by remarking that the tools developed in this paper might be relevant to quantify mixing times of related PDMPs proposed for molecular dynamics and MCMC inference algorithms including zig-zag and bouncy particle samplers [7,8,21,22,34].

2. Andersen dynamics and couplings

In this section, we briefly recall Andersen dynamics and its basic properties needed throughout the paper. Then we introduce a new class of couplings for two copies of the dynamics starting at different initial conditions.

2.1. *Andersen dynamics*

Andersen dynamics describes a molecular system at constant $\beta = (k_B T)^{-1}$ where T is the temperature and k_B is the Boltzmann constant. Here we consider a molecular system consisting of m particles each with n dimensions. A state of the molecular system is denoted by $(x, v) \in \mathbb{R}^{2\mathsf{mn}}$ where $x = (x_1, \ldots, x_\mathsf{m})$ represents the positions of the particles and $v = (v_1, \ldots, v_\mathsf{m})$ the corresponding velocities. Let $U \in C^2(\mathbb{R}^{\mathsf{mn}})$ denote the potential energy of the molecular system, and without loss of generality, suppose that $U(x) \geq 0$ for all $x \in \mathbb{R}^{\mathsf{mn}}$. For simplicity, suppose that all particles have unit

masses.[1] Hence, the Hamiltonian of the molecular system is

$$H(x, v) = (1/2)|v|^2 + U(x).$$

To precisely define Andersen dynamics, let

$$\phi_t(x, v) := \big(x_t(x, v), v_t(x, v)\big) \quad (t \in [0, \infty)) \tag{1}$$

denote the flow of the Hamiltonian dynamics

$$\frac{d}{dt}x_t = v_t, \qquad \frac{d}{dt}v_t = -\nabla U(x_t), \qquad \big(x_0(x, v), v_0(x, v)\big) = (x, v). \tag{2}$$

For $a \in \mathbb{R}^n$ and $i \in \{1, \ldots, m\}$, define the i-th particle velocity substitution

$$S(i, a)(x, v) := \big(x, (v_1, \ldots, v_{i-1}, a, v_{i+1}, \ldots, v_m)\big). \tag{3}$$

As seen below, this map is notationally convenient for describing the velocity randomization of a randomly selected particle in Andersen dynamics. On the same probability space, let $(N_t)_{t \geq 0}$ be a homogeneous Poisson process with intensity $\lambda > 0$ called the collision frequency in Andersen dynamics, and let $(T_k)_{k \in \mathbb{N}}$ be the corresponding strictly increasing sequence of jump times; let $(I_k)_{k \in \mathbb{N}}$ and $(\xi_k)_{k \in \mathbb{N}}$ be independent sequences of i.i.d. random variables $I_k \sim \mathrm{Unif}\{1, \ldots, m\}$ and $\xi_k \sim \mathcal{N}(0, \beta^{-1})^n$. The sequence of random variables $(I_k)_{k \in \mathbb{N}}$ represents the indices of the particles whose velocities get instantaneously randomized to $(\xi_k)_{k \in \mathbb{N}}$ at the jump times $(T_k)_{k \in \mathbb{N}}$ respectively.

With this notation, the stochastic process (X_t, V_t) corresponding to Andersen dynamics is defined as follows.

Definition 2.1 (Andersen Process). Given $t > 0$, $\lambda > 0$ and an initial condition $(x, v) \in \mathbb{R}^{2mn}$, define $T_0 = 0$, $\delta T_k = T_k - T_{k-1}$ for $k \geq 1$, $(X_0, V_0) = (x, v)$ and

$$(X_t, V_t) := \phi_{t-T_{N_t}} \circ S(I_{N_t}, \xi_{N_t}) \circ \phi_{\delta T_{N_t}} \circ \cdots \circ S(I_1, \xi_1) \circ \phi_{\delta T_1}(X_0, V_0).$$

The process (X_t, V_t) can also be defined piecewise. In particular, the process follows Hamiltonian dynamics in between two consecutive jump times, i.e.,

$$(X_s, V_s) = \phi_{s-T_{k-1}}(X_{T_{k-1}}, V_{T_{k-1}}) \quad \text{for } s \in [T_{k-1}, T_k) \text{ and } k \geq 1.$$

Moreover, at a jump time, $s = T_k$, the velocity of the I_k-th particle instantaneously changes to ξ_k, i.e.,

$$(X_{T_k}, V_{T_k}) = S(I_k, \xi_k)(X_{T_k-}, V_{T_k-})$$

where $(X_{T_k-}, V_{T_k-}) = \phi_{T_k-T_{k-1}}(X_{T_{k-1}}, V_{T_{k-1}})$.

The dynamics generates a Piecewise Deterministic Markov Process (PDMP) on the state space $\mathbb{R}^{2mn}$. The law of a PDMP is determined by one or several vector fields which govern its deterministic motion, a measurable function which gives the law of the random times between jumps, and a jump measure which gives the transition probability of its jumps [19,20]. In the case of Andersen dynamics, these are given by:

- the vector field

$$\mathfrak{X}(x, v) = \big(v, -\nabla U(x)\big), \quad (x, v) \in \mathbb{R}^{2mn},$$

  generating the deterministic Hamiltonian flow;
- the (constant) jump rate given by the collision frequency λ; and,
- the jump measure

$$Q\big((x, v), (dx'\, dv')\big) = \frac{1}{m} \sum_{i=1}^m \delta_x(dx') \varphi_\beta(v_i')\, dv_i' \prod_{j \neq i} \delta_{v_j}(dv_j'),$$

  where $\varphi_\beta(v_i') = (2\pi/\beta)^{-n/2} \exp(-(\beta/2)|v_i'|^2)$.

---

[1]The general case of a Hamiltonian system with a kinetic energy that is a positive-definite quadratic form can be treated by using mass-weighted coordinates; see Remark 3.6.

By [19, Theorem 5.5], the corresponding PDMP (X_t, V_t) with given initial condition (x, v) solves the local martingale problem for the extended generator $(\mathcal{G}, \mathcal{D}(\mathcal{G}))$ defined by

$$\mathcal{G}f = \mathcal{L}f + \mathcal{A}f, \quad f \in \mathcal{D}(\mathcal{G}). \tag{4}$$

Here $\mathcal{D}(\mathcal{G})$ is the set of all continuously differentiable functions $f : \mathbb{R}^{2mn} \to \mathbb{R}$,

$$\mathcal{L}f(x, v) = \mathfrak{X}(x, v) \cdot \nabla f(x, v) = v \cdot \nabla_x f(x, v) - \nabla U(x) \cdot \nabla_v f(x, v) \tag{5}$$

is the *Liouville operator* associated to the Hamiltonian dynamics, and

$$\mathcal{A}f(x, v) = \lambda \mathbb{E}\big\{f\big(\mathcal{S}(I, \xi)(x, v)\big) - f(x, v)\big\} \tag{6}$$

is the *Andersen collision operator* where the expectation in (6) is over the independent random variables $I \sim$ Unif$\{1, \ldots, m\}$ and $\xi \sim \mathcal{N}(0, \beta^{-1})^n$.

A key property of Andersen dynamics is that it leaves invariant the Boltzmann-Gibbs probability distribution

$$\Pi_{BG}(dx\, dv) \propto \exp\big(-\beta H(x, v)\big)\, dx\, dv. \tag{7}$$

Indeed, since the Hamiltonian flow preserves both the Hamiltonian function H and phase space volume (as a consequence of symplecticity), the Hamiltonian flow preserves Π_{BG}. Moreover, since the position component is held fixed and the ith velocity component is drawn from the v_i-marginal of Π_{BG}, the velocity randomizations also preserve Π_{BG}. This argument can be easily turned into a proof that Π_{BG} is infinitesimally invariant in the sense that

$$\int_{\mathbb{R}^{2mn}} \mathcal{G}f(z)\Pi_{BG}(dz) = 0$$

for any compactly supported C^1 function $f : \mathbb{R}^{2mn} \to \mathbb{R}$. To conclude that Π_{BG} is an invariant measure (not just infinitesimally invariant) requires additional assumptions on U, e.g., it is sufficient to show that an appropriate Foster–Lyapunov drift condition holds; see §3.5 of [13] for details.

Remark 2.2. In the case of one particle $m = 1$ with $\beta = 1$, Andersen dynamics becomes exact randomized Hamiltonian Monte Carlo which is geometrically ergodic under mild conditions on the potential energy U [13].

Remark 2.3. Andersen dynamics is related to second-order Langevin dynamics, but there are differences. First, note that Andersen dynamics does not incorporate explicit dissipation or diffusion. Second, although the velocity randomizations help ensure that Andersen dynamics is ergodic with respect to the Boltzmann-Gibbs probability distribution, they have the disadvantage of introducing jump discontinuities along the velocity of trajectories. In contrast, the velocity of trajectories for second-order Langevin dynamics is continuous.

2.2. *Couplings for Andersen dynamics*

A key tool in our analysis is a Markovian coupling $Y_t = ((X_t, V_t), (\tilde{X}_t, \tilde{V}_t))$ of two copies of the Andersen process starting from different initial conditions. This coupling is defined by suitably constructing the underlying random variables of the two copies on the same probability space. To precisely define this coupling, introduce the following Hamiltonian flow on $\mathbb{R}^{4mn}$

$$\phi_t^C\big((x, v), (\tilde{x}, \tilde{v})\big) := \big(\phi_t(x, v), \phi_t(\tilde{x}, \tilde{v}))\big) \quad \big(t \in [0, \infty)\big) \tag{8}$$

where ϕ_t is the Hamiltonian flow from (1). Let $\gamma \geq 0$ be a parameter of the coupling whose precise value will be specified in an appropriate way in subsequent sections. For $a \in \mathbb{R}^n$, $i \in \{1, \ldots, m\}$, and $u \in (0, 1)$, introduce

$$\mathcal{S}^C(i, a, u)\big((x, v), (\tilde{x}, \tilde{v})\big) := \big(\mathcal{S}(i, a)(x, v), \mathcal{S}(i, \tilde{a})(\tilde{x}, \tilde{v})\big) \tag{9}$$

where $\mathcal{S}(\cdot, \cdot)$ is the mapping in (3) and $\tilde{a} = \Phi(a, z_i, u)$ with $z_i = x_i - \tilde{x}_i$. Here we have introduced the function $\Phi : \mathbb{R}^n \times \mathbb{R}^n \times (0, 1) \to \mathbb{R}^n$ defined by

$$\Phi(a, b, u) := \begin{cases} a + \gamma b & \text{if } u < \frac{\varphi_\beta(a + \gamma b)}{\varphi_\beta(a)}, \\ a - 2(e_b \cdot a)e_b & \text{else,} \end{cases} \tag{10}$$

where $e_b = b/|b|$ for $b \neq 0$ and $e_0 = 0$. The function Φ is used to couple the randomized velocities of the two copies of the Andersen process. In particular, a simple calculation gives the following results that will be used below.

Lemma 2.4. *Let $b \in \mathbb{R}^n$, and let $\tilde{\xi} = \Phi(\xi, b, \mathcal{U})$ where $\xi \sim \mathcal{N}(0, \beta^{-1})^n$ and $\mathcal{U} \sim \mathrm{Unif}(0, 1)$. Then*

$$\mathrm{Law}(\tilde{\xi}) = \mathcal{N}(0, \beta^{-1})^n, \tag{11}$$

$$\mathbb{P}[\xi - \tilde{\xi} \neq -\gamma b] = d_{\mathrm{TV}}(\mathcal{N}(0, \beta^{-1})^n, \mathcal{N}(\gamma b, \beta^{-1})^n), \tag{12}$$

$$\mathbb{P}[\xi - \tilde{\xi} \neq -\gamma b] \leq \sqrt{\beta} \gamma |b| / \sqrt{2\pi}, \quad \text{and} \tag{13}$$

$$\mathbb{E}[|\xi|^2; \xi - \tilde{\xi} \neq -\gamma b] \leq (n+1)\gamma |b| / \sqrt{2\pi\beta}. \tag{14}$$

The estimates (13) and (14) are a refinement of Lemma 3.7 in [12]. A self-contained proof of these estimates is provided in Section 5.1. The results (11) and (12) indicate that the pair of random variables $(\xi, \tilde{\xi})$ is a realization of a coupling of two copies of $\mathcal{N}(0, \beta^{-1})^n$ such that $\tilde{\xi} = \xi + \gamma b$ with maximal possible probability. Proofs of (11) and (12) appear in [12, Section 2.3.2].

Let $(N_t)_{t \geq 0}$ be a Poisson counting process with intensity λ and let $(T_k)_{k \in \mathbb{N}}$ be the corresponding strictly increasing sequence of jump times; let $(I_k)_{k \in \mathbb{N}}$, $(\xi_k)_{k \in \mathbb{N}}$, and $(\mathcal{U}_k)_{k \in \mathbb{N}}$ be independent sequences of i.i.d. random variables $I_k \sim \mathrm{Unif}\{1, \ldots, m\}$, $\xi_k \sim \mathcal{N}(0, \beta^{-1})^n$, and $\mathcal{U}_k \sim \mathrm{Unif}(0, 1)$, all defined on a joint probability space.

With this notation, we define the following coupling for Andersen dynamics.

Definition 2.5 (Coupling for Andersen Dynamics). Given $t > 0$, $\lambda > 0$, $\gamma \geq 0$, and an initial condition $y \in \mathbb{R}^{4mn}$, define $T_0 = 0$, $\delta T_k = T_k - T_{k-1}$ for $k \geq 1$, $Y_0 = y$, and

$$Y_t := \phi^C_{t-T_{N_t}} \circ \mathcal{S}^C(I_{N_t}, \xi_{N_t}, \mathcal{U}_{N_t}) \circ \phi^C_{\delta T_{N_t}} \circ \cdots \circ \mathcal{S}^C(I_1, \xi_1, \mathcal{U}_1) \circ \phi^C_{\delta T_1}(Y_0).$$

The process Y_t can also be defined piecewise. In particular, the components of the coupling follow Hamiltonian dynamics in between two consecutive jump times,

$$Y_s = \phi^C_{s-T_{k-1}}(Y_{T_{k-1}}), \quad \text{for } s \in [T_{k-1}, T_k).$$

Moreover, at a jump time, $s = T_k$, the velocities of the I_k-th particles in the first and second components of the coupling process instantaneously change to $V^{I_k}_{T_k} = \xi_k$ and $\tilde{V}^{I_k}_{T_k} = \Phi(\xi_k, X^{I_k}_{T_k} - \tilde{X}^{I_k}_{T_k}, \mathcal{U}_k)$ respectively, i.e.,

$$Y_{T_k} = \mathcal{S}^C(I_k, \xi_k, \mathcal{U}_k)(Y_{T_k-})$$

where $Y_{T_k-} = \phi^C_{T_k - T_{k-1}}(Y_{T_{k-1}})$. We stress that $X_{T_k} = X_{T_k-}$, $\tilde{X}_{T_k} = \tilde{X}_{T_k-}$, and $V^j_{T_k} = V^j_{T_k-}$, $\tilde{V}^j_{T_k} = \tilde{V}^j_{T_k-}$ for $j \in \{1, \ldots, m\} \setminus \{I_k\}$. Moreover, all of the underlying random variables of the two copies of the Andersen process are synchronously coupled except possibly the randomized velocities.

The coupling of the randomized velocities is inspired by recently introduced couplings for Hamiltonian Monte Carlo [12] and second-order Langevin dynamics [29]. It is motivated by the observation that the free-streaming Hamiltonian dynamics is contractive for small time durations if the difference in the initial velocities is chosen negatively proportional to the difference in the initial positions [12, Figure 1]. Therefore, the randomized velocities of the I_k-th particles at a jump time T_k, i.e., $V^{I_k}_{T_k} = \xi_k$ and $\tilde{V}^{I_k}_{T_k} = \Phi(\xi_k, X^{I_k}_{T_k} - \tilde{X}^{I_k}_{T_k}, \mathcal{U}_k)$, are defined such that the difference in the velocities satisfies $V^{I_k}_{T_k} - \tilde{V}^{I_k}_{T_k} = \xi_{I_k} - \tilde{\xi}_{I_k} = -\gamma(X^{I_k}_{T_k} - \tilde{X}^{I_k}_{T_k})$ with maximal possible probability, as illustrated in Figure 1(a). Otherwise, a reflection coupling is applied, as illustrated in Figure 1(b).

The coupling process Y_t is itself a PDMP on the state space $\mathbb{R}^{4mn}$ with the following characteristics:

- the vector field

$$\mathfrak{X}^C((x, v), (\tilde{x}, \tilde{v})) = (v, -\nabla U(x), \tilde{v}, -\nabla U(\tilde{x}));$$

- the (constant) jump rate given by the collision frequency λ; and,

(a) $\tilde{\xi} = \xi + \gamma(x_I - \tilde{x}_I)$
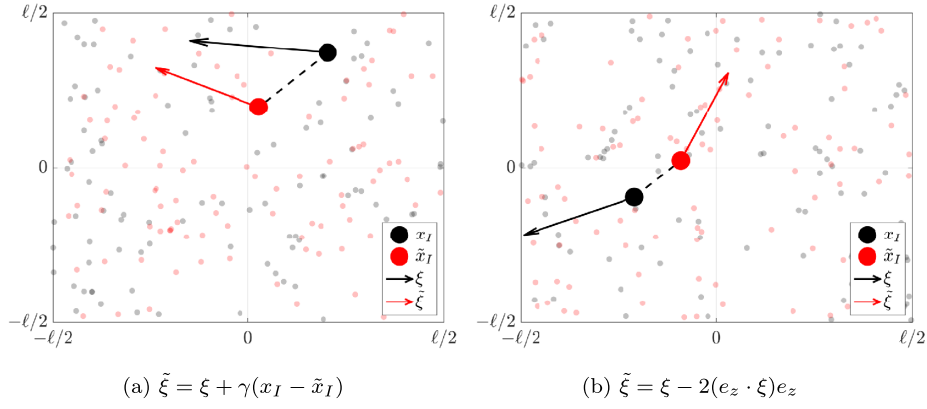
(b) $\tilde{\xi} = \xi - 2(e_z \cdot \xi)e_z$

Fig. 1. Illustrations of a velocity randomization step where the difference in velocities of the I-th particles (enlarged dots), $\xi - \tilde{\xi}$, is: (a) negatively proportional to the difference in their positions $z = x_I - \tilde{x}_I$; and (b) reflected about the hyperplane passing through the origin, orthogonal to $e_z = z/|z|$.

- the jump measure

$$Q^C(y, dy')$$

$$= \frac{1}{\mathsf{m}} \sum_{i=1}^{\mathsf{m}} \delta_x(dx')\delta_{\tilde{x}}(d\tilde{x}') Q_i^C\big((v_i, \tilde{v}_i), (dv_i'\, d\tilde{v}_i')\big) \prod_{j \neq i} \delta_{v_j}(dv_j')\delta_{\tilde{v}_j}(d\tilde{v}_j') \quad \text{where,}$$

$$Q_i^C\big((v_i, \tilde{v}_i), (dv_i'\, d\tilde{v}_i')\big) = \big(\varphi_\beta(v_i') \wedge \varphi_\beta(v_i' + \gamma z_i)\big)\delta_{v_i' + \gamma z_i}(d\tilde{v}_i')\, dv_i'$$

$$+ \big(\varphi_\beta(v_i') - \varphi_\beta(v_i' + \gamma z_i)\big)^+ \delta_{v_i' - 2(e_{z_i} \cdot v_i')e_{z_i}}(d\tilde{v}_i')\, dv_i'$$

and where $e_{z_i} = z_i/|z_i|$ for $z_i \neq 0$ and $e_0 = 0$.

Since the coupling process is again a PDMP, the results in [19] show that it solves a local martingale problem for an extended generator

$$\mathcal{G}_\gamma^C = \mathcal{L}^C + \mathcal{A}_\gamma^C \tag{15}$$

which is the sum of the Liouville operator $\mathcal{L}^C$ for the Hamiltonian vector field $\mathfrak{X}^C$ and a velocity randomization operator $\mathcal{A}_\gamma^C$, and whose domain $\mathcal{D}(\mathcal{G}_\gamma^C)$ consists of continuously differentiable functions on $\mathbb{R}^{4\mathsf{mn}}$. For a function $F : \mathbb{R}^{4\mathsf{mn}} \to \mathbb{R}$ that is differentiable at y, the Liouville operator $\mathcal{L}^C$ is given by

$$\mathcal{L}^C F(y) = \mathfrak{X}^C(y) \cdot \nabla F(y). \tag{16}$$

The action of the coupled velocity randomization operator $\mathcal{A}_\gamma^C$ on a function $F : \mathbb{R}^{4\mathsf{mn}} \to \mathbb{R}$ is defined as

$$\mathcal{A}_\gamma^C F(y) = \lambda \mathbb{E}\big\{F\big(\mathcal{S}^C(I, \xi, \mathcal{U})y\big) - F(y)\big\} \tag{17}$$

$$= \lambda \int_{\mathbb{R}^{4\mathsf{mn}}} \big(F(y') - F(y)\big) Q^C(y, dy').$$

where the expectation is taken over the independent random variables $I \sim \text{Unif}\{1, \ldots, \mathsf{m}\}$, $\xi \sim \mathcal{N}(0, \beta^{-1})^{\mathsf{n}}$ and $\mathcal{U} \sim \text{Unif}(0, 1)$.

It can be easily verified that the process $(Y_t)_{t \geq 0}$ is indeed a coupling of two copies of Andersen dynamics. Indeed, by uniqueness of the local martingale problem for the Andersen process, it is sufficient to check that $\mathcal{G}_\gamma^C F(y)$ reduces to $\mathcal{G}F(x, v)$ or $\mathcal{G}F(\tilde{x}, \tilde{v})$ for functions independent of the first or second component of $y = ((x, v), (\tilde{x}, \tilde{v}))$, respectively. For such functions, it immediately follows that $\mathcal{L}^C F = \mathcal{L}F$, and using a similar calculation to the one performed in Section 2.3.2 of [12], $\mathcal{A}_\gamma^C F = \mathcal{A}F$; hence, $\mathcal{G}_\gamma^C F = \mathcal{G}F$.

Remark 2.6 (Synchronous coupling). When $\gamma = 0$ in Definition 2.5 the velocities of the I_k-th particles are synchronously randomized, i.e., $V_{T_k}^{I_k} = \tilde{V}_{T_k}^{I_k} = \xi_k$. For the corresponding generators, we write $\mathcal{A}_{\text{sync}}^C = \mathcal{A}_0^C$ and $\mathcal{G}_{\text{sync}}^C = \mathcal{G}_0^C$. By itself, a synchronous coupling is insufficient to obtain contractivity for non-strongly-convex potentials.

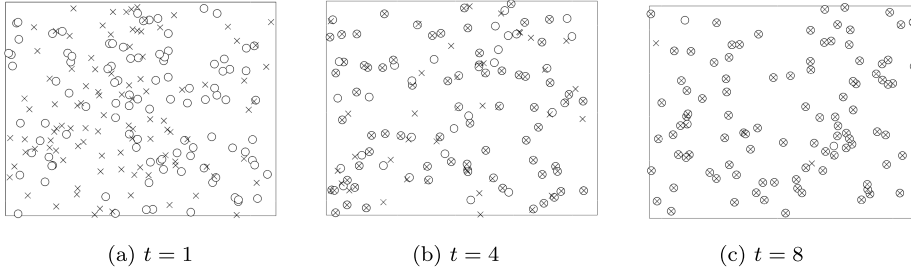(a) $t = 1$ (b) $t = 4$ (c) $t = 8$

Fig. 2. Snapshots at the indicated times along a realization of the coupling for Andersen dynamics on $(\mathbb{T}_\ell^{2m} \times \mathbb{R}^{2m})^2$ in the free-streaming case with $\lambda = 600$, m $= 100$, $\ell = 1$, $U \equiv 0$, $\beta = 1$, and $\gamma = 1/(\ell/2 + \mathsf{m}/\lambda)$.

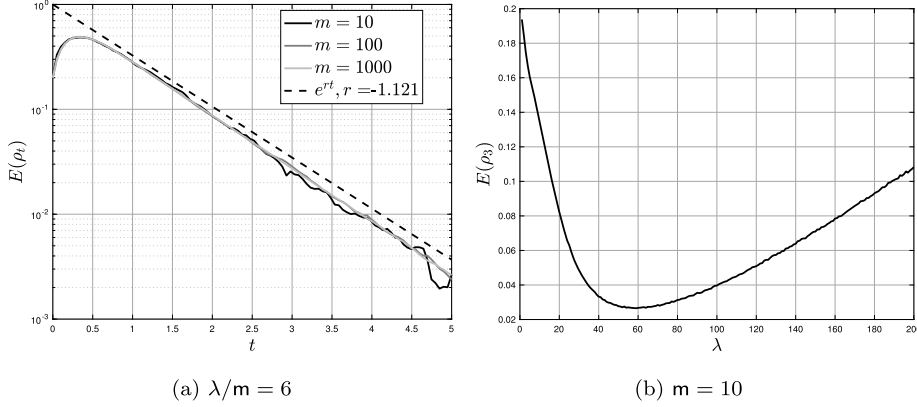(a) $\lambda/\mathsf{m} = 6$ (b) $\mathsf{m} = 10$

Fig. 3. Evolution of a Monte Carlo estimate of $\mathbb{E}(\rho_t)$ using 10^5 realizations of the coupling for Andersen dynamics on $(\mathbb{T}_\ell^m \times \mathbb{R}^m)^2$ where $\ell = 1$, $U \equiv 0$, $\beta = 1$, $\rho_t = (1/\mathsf{m}) \sum_{i=1}^m \sqrt{\zeta_i(Z_t, W_t)^2 + (W_t^i)^2}$, and $\gamma = 1/(\ell/2 + \mathsf{m}/\lambda)$. The choice of γ is motivated by Thm. 3.9. In (a), λ/m is fixed while m is increased from 10 to 1000; note that the observed convergence rate is dimension-free and consistent with Thm. 3.9 which implies contractivity with respect to an equivalent metric. In (b), $\mathbb{E}(\rho_3)$ is plotted as a function of λ; note that λ in (a) is approximately the minimizer of $\mathbb{E}(\rho_3)$.

2.3. *Andersen dynamics on a torus*

Molecular dynamics simulations routinely employ periodic boundary conditions [1,2,30,31,36]. In particular, the configuration space of the molecular system is typically a flat torus $\mathbb{T}_\ell^{mn}$. Here $\mathbb{T}_\ell = \mathbb{R}/(\ell\mathbb{Z})$ denotes the circle with circumference $\ell > 0$. The canonical projection from the covering space $\mathbb{R}^{mn}$ to the torus $\mathbb{T}_\ell^{mn}$ is denoted by π, and $\tau_z(x) \in \mathbb{T}_\ell^{mn}$ denotes the translation of a point $x \in \mathbb{T}_\ell^{mn}$ by a tangent vector $z \in \mathbb{R}^{mn}$.

Let $U \in C^2(\mathbb{T}_\ell^{mn})$, and without loss of generality, suppose $U(x) \geq 0$ for all $x \in \mathbb{T}_\ell^{mn}$. Andersen dynamics on the torus T_ℓ^{mn} with potential U is the PDMP with state space $T_\ell^{mn} \times \mathbb{R}^{mn}$ defined by Definition 2.1, where ϕ_t is now the flow of Hamiltonian dynamics (2) on the torus, and $\mathcal{S}$ is again defined by (3) as above. The process can also be obtained by projection from Andersen dynamics on Euclidean space. Indeed, let $\hat{U}$ denote the periodic function in $C^2(\mathbb{R}^{mn})$ defined by $\hat{U}(x) = U(\pi(x))$ for all x. Then the Andersen process (X_t, V_t) on the torus with inital condition $(x_0, v_0) \in T_\ell^{mn} \times \mathbb{R}^{mn}$ is given by $X_t = \pi(\hat{X}_t)$ and $\hat{V}_t = V_t$, where $(\hat{X}_t, \hat{V}_t)$ is the Andersen process on $\mathbb{R}^{mn} \times \mathbb{R}^{mn}$ with initial condition $(\hat{x}_0, v_0)$ for an arbitrary $\hat{x}_0 \in \pi^{-1}(x_0)$.

2.4. *Coupling for Andersen dynamics on a torus*

We now introduce a coupling for two copies of the Andersen process on the torus. The coupling is a piecewise deterministic Markov process $((X_t, V_t), (\tilde{X}_t, \tilde{V}_t))$ with state space $(\mathbb{T}_\ell^{mn} \times \mathbb{R}^{mn})^2$. Although in spirit, the construction is similar to the construction of a coupling for Andersen dynamics on euclidean space in Section 2.2, some technical difficulties arise in the torus case. Therefore, we require a slightly different setup.

In order to construct the coupling process, we consider another PDMP $Y_t = (X_t, V_t, Z_t, W_t)$ with state space $\mathbb{T}_\ell^{mn} \times \mathbb{R}^{mn} \times \mathbb{R}^{mn} \times \mathbb{R}^{mn}$. The coupling is obtained from this process by setting

$$\big((X_t, V_t), (\tilde{X}_t, \tilde{V}_t)\big) := \pi^C(Y_t), \tag{18}$$
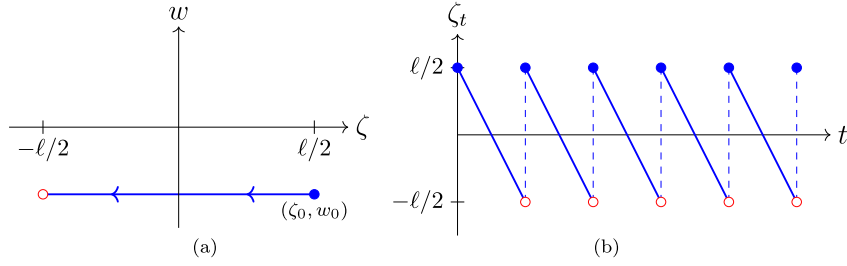
Fig. 4. Plots of $\zeta_t = \zeta(z_t, w_t)$ initially at $(z_0, w_0) = (\ell/2, -1)$ with constant w_t in (a) phase space and (b) as a function of time. Note that $t \mapsto \zeta_t$ is a càdlàg trajectory.

where $\pi^C : \mathbb{T}_\ell^{mn} \times \mathbb{R}^{3mn} \to (\mathbb{T}_\ell^{mn} \times \mathbb{R}^{mn})^2$ is the projection map defined by

$$\pi^C(x, v, z, w) = \big((x, v), (\tau_{-z}(x), v - w)\big). \tag{19}$$

Thus $W_t = V_t - \tilde{V}_t$ and $\tilde{X}_t = \tau_{-Z_t}(X_t)$, i.e., W_t and Z_t correspond to the differences between the coupling components. Let $\phi_t^C = (x_t, v_t, z_t, w_t)$ denote the flow on $\mathbb{T}_\ell^{mn} \times \mathbb{R}^{3mn}$ of the ODE

$$\frac{d}{dt}x_t = v_t, \qquad \frac{d}{dt}z_t = w_t,$$
$$\frac{d}{dt}v_t = -\nabla U(x_t), \qquad \frac{d}{dt}w_t = \nabla U\big(\tau_{-z_t}(x_t)\big) - \nabla U(x_t). \tag{20}$$

For $(z, w) \in \mathbb{R}^{2mn}$, we also define $\zeta(z, w) \in [-\ell/2, \ell/2]^{mn}$ by

$$\zeta_{i,j}(z, w) = \begin{cases} z_{i,j} - \lfloor (z_{i,j} + \ell/2)/\ell \rfloor \ell & \text{if } z_{i,j} \notin \ell/2 + \ell\mathbb{Z}, \\ \ell/2 & \text{if } w_{i,j} < 0 \text{ and } z_{i,j} \in \ell/2 + \ell\mathbb{Z}, \\ -\ell/2 & \text{if } w_{i,j} \geq 0 \text{ and } z_{i,j} \in \ell/2 + \ell\mathbb{Z}. \end{cases} \tag{21}$$

One should think of $\zeta = \zeta(z, w)$ as a minimal difference vector between the corresponding components x and $\tilde{x}$ on the torus. In particular, $\zeta_{i,j} \equiv z_{i,j} \bmod \ell$ for all $1 \leq i \leq m$ and $1 \leq j \leq n$, and thus

$$x = \tau_z(\tilde{x}) = \tau_\zeta(\tilde{x}). \tag{22}$$

The motivation for the special definition of $\zeta_{i,j}$ for $z_{i,j} \in \ell/2 + \ell\mathbb{Z}$ is that it ensures that if (x_t, v_t, z_t, w_t) is a solution of (20) then $t \mapsto \zeta_{i,j}(z_t, w_t)$ is càdlàg (right continuous with left limits) for all t such that $w_t^{i,j} \neq 0$, see Figure 4. This will imply that the coupling distance introduced further below is also a càdlàg function of t, see Lemma 3.8.

The process (Y_t) is now defined by Definition 2.5 above where ϕ_t^C is the flow of (20), and

$$\mathcal{S}^C(i, a, u)(x, v, z, w) = \big(\mathcal{S}(i, a)(x, v), \mathcal{S}(i, a - \tilde{a})(z, w)\big)$$

$$\text{with } \tilde{a} = \Phi\big(a, \zeta_i(z, w), u\big). \tag{23}$$

Again, (Y_t) is a piecewise deterministic Markov process with generator given by (15), (16) and (17), where now the vector field generating the deterministic flow is $\mathfrak{X}^C(x, v, z, w) = (v, -\nabla U(x), w, \nabla U(\tau_{-z}(x)) - \nabla U(x))$, $\mathcal{S}^C$ is defined by (23), and the jump measure Q^C is adapted correspondingly.

3. Main results

We now apply the couplings introduced above to derive contraction results and bounds on Wasserstein distances to the invariant measure for Andersen dynamics. We first consider a strongly convex potential energy function on $\mathbb{R}^{mn}$. In this case, relatively precise bounds can be derived by synchronous coupling. Then we consider Andersen dynamics on a high dimensional torus, which is a common setup in molecular dynamics. In that case, synchronous coupling can not be applied since the potential energy function is not convex. In general, phase transitions can cause slow mixing as the dimension goes to infinity. Using the couplings introduced above, we are able to show that rapid mixing still holds for weak interactions between the particles.

3.1. *Andersen dynamics for weakly anharmonic molecular systems*

Here we consider potentials $U(x)$ that satisfy the following assumption.

Assumption 3.1. The potential energy is weakly anharmonic, i.e.,

$$U(x) = \frac{1}{2}x^T \mathcal{C}^{-1} x + G(x), \quad \text{for all } x \in \mathbb{R}^{\mathsf{mn}}, \tag{24}$$

where $\mathcal{C}$ is an $\mathsf{mn} \times \mathsf{mn}$ symmetric positive definite matrix; and the perturbation $G(x)$ is a convex, continuously differentiable and L_G-gradient Lipschitz function, i.e., there exists $L_G \geq 0$ such that

$$\left|\nabla G(x) - \nabla G(\tilde{x})\right| \leq L_G |x - \tilde{x}|, \quad x, \tilde{x} \in \mathbb{R}^{\mathsf{mn}}. \tag{25}$$

Let $\sigma_{\max}^2$ and $\sigma_{\min}^2$ denote the largest and smallest eigenvalues of $\mathcal{C}$, respectively. Any K-strongly convex, continuously differentiable and gradient Lipschitz function $U(x)$ can be put in the form of (24) with $\mathcal{C} = K^{-1}\mathbf{1}_{\mathsf{mn}}$ where $\mathbf{1}_{\mathsf{mn}}$ is the $\mathsf{mn} \times \mathsf{mn}$ identity matrix and $G(x) = U(x) - K|x|^2/2$. Moreover, it follows from this assumption that $U(x)$ is itself strongly convex

$$\left(\nabla U(x) - \nabla U(\tilde{x})\right) \cdot (x - \tilde{x}) \geq \sigma_{\max}^{-2} |x - \tilde{x}|^2 \quad \text{for all } x, \tilde{x} \in \mathbb{R}^{\mathsf{mn}}. \tag{26}$$

Here we used the convexity of $G(x)$ which implies that $(\nabla G(x) - \nabla G(\tilde{x})) \cdot (x - \tilde{x}) \geq 0$. The contraction result given below uses a synchronous coupling of velocities to exploit the convexity of the perturbation $G(x)$; see Remark 2.6. Let

$$H_0(x, v) := (1/2)\left(|v|^2 + x^T \mathcal{C}^{-1} x\right)$$

be the unperturbed Hamiltonian. The Hamiltonian of the weakly anharmonic system is $H(x, v) = H_0(x, v) + G(x)$. In terms of H_0, define the metric $\rho : \mathbb{R}^{4\mathsf{mn}} \to \mathbb{R}^+$ by

$$\rho(y)^2 := H_0\big(z(y), w(y)\big) + \frac{\lambda}{4\mathsf{m}} z(y) \cdot w(y) + \frac{\lambda^2}{8\mathsf{m}^2}\big|z(y)\big|^2 \tag{27}$$

$$= \begin{pmatrix} z(y) & w(y) \end{pmatrix} \mathsf{G} \begin{pmatrix} z(y) \\ w(y) \end{pmatrix}, \quad \mathsf{G} := \begin{bmatrix} \frac{\lambda^2}{8\mathsf{m}^2}\mathbf{1}_{\mathsf{mn}} + \frac{1}{2}\mathcal{C}^{-1} & \frac{\lambda}{8\mathsf{m}}\mathbf{1}_{\mathsf{mn}} \\ \frac{\lambda}{8\mathsf{m}}\mathbf{1}_{\mathsf{mn}} & \frac{1}{2}\mathbf{1}_{\mathsf{mn}} \end{bmatrix}, \tag{28}$$

where for $y = (x, v, \tilde{x}, \tilde{v})$, we set $z(y) = x - \tilde{x}$ and $w(y) = v - \tilde{v}$. Note that $\rho(y)$ only depends on $\mathcal{C}$ and the intensity of the velocity randomizations per particle λ/m. Moreover, by completing the square in (27), it is easy to show that $\rho(y)^2$ is positive definite. This twisted metric involves the "qp trick" behind Foster–Lyapunov functions for (i) dissipative Hamiltonian systems with random impulses [51]; (ii) second-order Langevin processes [43,53]; and (iii) exact randomized HMC [13].

In the sequel, we will sometimes write the y dependence in z, w, ρ, etc. and sometimes suppress it in the notation, depending on what is more convenient. Let $(p_t)_{t\geq 0}$ denote the transition semigroup of Andersen dynamics, and for all probability measures μ, ν on $\mathbb{R}^{2\mathsf{mn}}$ let $\mathcal{W}_2(\mu, \nu)$ denote the standard 2-Wasserstein distance.

Theorem 3.2. *Suppose that Assumption 3.1 holds and $\lambda > 0$ satisfies*

$$\lambda/\mathsf{m} \geq 4L_G \sigma_{\max}. \tag{29}$$

Then

$$\mathcal{G}_{\mathrm{sync}}^C \rho^2 \leq -c\rho^2, \quad \text{where } c := \min\left(\frac{1}{8}\frac{\lambda}{\mathsf{m}}, \frac{8}{5}\frac{1}{\sigma_{\max}^2}\frac{\mathsf{m}}{\lambda}\right). \tag{30}$$

Thus, the process $t \mapsto e^{ct}\rho(Y_t)^2$ is a nonnegative supermartingale, and

$$\mathcal{W}_2(\mu p_t, \nu p_t) \leq \varkappa(\mathsf{G})^{1/2} e^{-ct/2}\mathcal{W}_2(\mu, \nu) \tag{31}$$

where $\varkappa(\mathsf{G})$ is the condition number of the matrix G which satisfies

$$\varkappa(\mathsf{G}) \leq \frac{\max(\lambda^2/\mathsf{m}^2 + 4\sigma_{\max}^{-2}, 4)\max(\lambda^2/\mathsf{m}^2 + 4\sigma_{\min}^{-2}, 4)}{3\lambda^2/\mathsf{m}^2 + 16\sigma_{\max}^{-2}}. \tag{32}$$
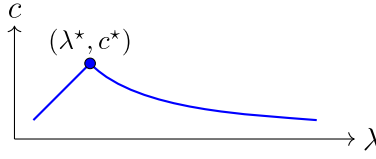
Fig. 5. The rate appearing in Example 3.4 is maximized at $(\lambda^\star, c^\star)$.

A proof of this theorem is provided in Section 4. In the unperturbed case $G \equiv 0$, a similar result is proven for exact randomized HMC in Proposition 4 of [21]. Related results have been proven for HMC in [12,16,42] and second order Langevin dynamics in [17,18], though an important difference in Theorem 3.2 is that condition (29) and the rate in (30) do not deteriorate in the limit that the condition number of $\mathcal{C}$ becomes large for fixed $\sigma_{\max}$. For fixed $\sigma_{\max}$, note also that the rate in (30) is of the form $\min(\lambda/\mathsf{m}, \mathsf{m}/\lambda)$, which is the typical rate one encounters with second-order Langevin dynamics and the kinetic Fokker–Planck equation, though possibly using different distances to quantify convergence to equilibrium e.g. weighted L^2 or Sobolev norms or relative entropy [29,48].

Example 3.3 (Strongly Convex Potential). For a K-strongly convex, continuously differentiable and gradient Lipschitz function $U(x)$, Theorem 3.2 gives a rate of $c = \min((1/8)(\lambda/\mathsf{m}), (8/5)K(\mathsf{m}/\lambda))$ provided that λ satisfies $\lambda/\mathsf{m} \geq 4L_G/\sqrt{K}$ where L_G is a Lipschitz constant for the gradient of $G(x) = U(x) - K|x|^2/2$.

The next example can be viewed as the potential energy corresponding to a truncation of an infinite-dimensional Gaussian measure [6,9,11]. This model problem illustrates the importance of duration randomization when the underlying Hamiltonian dynamics is highly oscillatory.

Example 3.4 (Neal's Example). Let $\mathsf{m} = 1$ and $U(x) = 2^{-1}\sum_{i=1}^{\mathsf{n}} i^2 x_i^2$; hence, $\sigma_{\max} = 1$. The corresponding Hamiltonian dynamics is highly oscillatory when the dimension n is large [49]. Noting that condition (29) always holds when $L_G = 0$, Theorem 3.2 gives a dimension-free rate of $c = \min(\lambda/8, (2/5)\lambda^{-1})$ which is maximized at $\lambda^\star = 4\sqrt{5}/5$ where $c^\star = \sqrt{5}/10$.

More generally, when $L_G = 0$, the rate from Theorem 3.2 is maximized at $\lambda^\star/\mathsf{m} = 4\sqrt{5}/(5\sigma_{\max})$ where $c^\star = \sqrt{5}/(10\sigma_{\max})$. This conclusion remains true when L_G is small; specifically, when $L_G \leq \sqrt{5}/(5\sigma_{\max}^2)$. However, when L_G is larger than that, i.e., $L_G > \sqrt{5}/(5\sigma_{\max}^2)$, the rate is maximized at $\lambda^\star/\mathsf{m} = 4L_G\sigma_{\max}$ where $c^\star = 1/(10L_G\sigma_{\max}^3)$.

Remark 3.5 (Duration Randomization). Due to possible periodicity of the Hamiltonian flow, contraction bounds for HMC in the strongly convex case typically require that the duration parameter is short enough [12,16,42]. On the other hand, since duration randomized Hamiltonian flows avoid periodicities almost surely, contraction bounds for exact randomized HMC allow longer mean durations as illustrated in Example 3.4 [13,21,41,45].

Remark 3.6 (Preconditioned Andersen Dynamics). Theorem 3.2 applies to Andersen dynamics with respect to kinetic energies that are positive definite quadratic forms by using mass-weighted coordinates. In particular, if the Hamiltonian is of the form $\check{H}(\check{x}, \check{v}) = (1/2)\check{v}^T M\check{v} + \check{U}(\check{x})$, then the change of variables $(\check{x}, \check{v}) \mapsto (M^{1/2}\check{x}, M^{1/2}\check{v})$ results in a Hamiltonian system with $H(x, v) = (1/2)|v|^2 + U(x)$ where $U(x) = \check{U}(M^{-1/2}x)$ and Theorem 3.2 would hold if $D^2 U(x) \geq \mathcal{C}_M^{-1}$ where $\mathcal{C}_M = M^{1/2}\mathcal{C}M^{1/2}$. The choice $M = \mathcal{C}^{-1/2}$ would then correspond to "preconditioned" Andersen dynamics because it normalizes to one all of the eigenvalues of the matrix $\mathcal{C}$ [6,11].

3.2. *Contractivity of Andersen dynamics with weak interactions on a high-dimensional torus*

In this part, in order to avoid overloading the notation, we assume $n = 1$. However, we stress that the results below can be extended without essential changes to the case $n \neq 1$. In the following, we assume that the potential energy of the molecular system $U : \mathbb{T}^{\mathsf{m}} \to \mathbb{R}$ is twice continuously differentiable, and without loss of generality, nonnegative.

Assumption 3.7. The potential energy $U \in C^2(\mathbb{T}_\ell^{\mathsf{m}})$ satisfies $U(x) \geq 0$ for all $x \in \mathbb{T}_\ell^{\mathsf{m}}$.

Assumption 3.7 implies that the following constants are finite:

$$L = \sup_{\substack{1 \le i \le m \\ x \in \mathbb{T}_\ell^m}} \left| \frac{\partial^2 U}{\partial x_i^2}(x) \right|, \qquad J = \sup_{\substack{1 \le i < j \le m \\ x \in \mathbb{T}_\ell^m}} \left| \frac{\partial^2 U}{\partial x_i \partial x_j}(x) \right|. \tag{33}$$

Fix $i \in \{1, \dots, m\}$ and define $\nabla_i U(x) := \frac{\partial U}{\partial x_i}(x)$. Then for all $x, \tilde{x} \in \mathbb{T}_\ell^m$,

$$\begin{aligned}
\left| \nabla_i U(x) - \nabla_i U(\tilde{x}) \right| &= \left| \int_0^1 \frac{d}{ds} \nabla_i U \left(\tau_{s\zeta}(\tilde{x}) \right) ds \right| \\
&\le \left| \int_0^1 \frac{\partial^2 U}{\partial x_i^2} \left(\tau_{s\zeta}(\tilde{x}) \right) \zeta_i \, ds \right| + \sum_{k \ne i} \left| \int_0^1 \frac{\partial^2 U}{\partial x_i \partial x_k} \left(\tau_{s\zeta}(\tilde{x}) \right) \zeta_k \, ds \right| \\
&\le L |\zeta_i| + J \sum_{k \ne i} |\zeta_k| \tag{34}
\end{aligned}$$

where $\zeta \in \mathbb{R}^m$ is an arbitrary tangent vector such that $x = \tau_\zeta(\tilde{x})$. The parameter J quantifies the strength of interaction between particles, and "weak interactions" specifically means that J scales like $1/m$ as $m \to \infty$, which corresponds to the standard mean-field limit [24,47,52,54].

The choice of an adequate metric in order to prove contraction properties on the torus is quite tricky. It combines ideas from several previous works including in particular the results on contractive couplings for Langevin dynamics and HMC in [27,29] and [12], as well as the derivation of dimension-free contraction rates for mean-field models with weak interactions in [27] and [14]. Besides combining these approaches, they have to be adapted to the special setup on the torus.

To each pair $((x, v), (\tilde{x}, \tilde{v})) \in (\mathbb{T}_\ell^m \times \mathbb{R}^m)^2$, we assign $y = (x, v, z, w) \in \mathbb{T}_\ell^m \times \mathbb{R}^{3m}$ such that $w = v - \tilde{v}$ and $x = \tau_z(\tilde{x})$, i.e., $((x, v), (\tilde{x}, \tilde{v})) = \pi^C(y)$. We define $\zeta = \zeta(z, w)$ by (21), and we set

$$q(z, w) = \zeta(z, w) + \gamma^{-1} w.$$

Although the choice of z and hence y is not unique, the definitions of ζ and q do not depend on this choice. Since ζ is in $[-\ell/2, \ell/2]^m$, it is a tangent vector of a minimal geodesic from $\tilde{x}$ to x. Let $\alpha > 0$ and let $i \in \{1, \dots, m\}$. With a slight abuse of notation, we now define a weighted ℓ_2-distance between the i-th components of the coupling by

$$r_i\big((x, v), (\tilde{x}, \tilde{v})\big) = r_i(y) = \sqrt{\big|\zeta_i(z, w)\big|^2 + \alpha^{-2} \big|q_i(z, w)\big|^2}. \tag{35}$$

Given an initial condition $y \in \mathbb{T}_\ell^m \times \mathbb{R}^{3m}$, and for any $t \ge 0$, let $y_t = (x_t, v_t, w_t, z_t)$ be the solution to (20) with $y_0 = y$, and for any $i \in \{1, \dots, m\}$, let $r_t^i := r_i(y_t)$ and $\zeta_t^i := \zeta_i(z_t, w_t)$. As illustrated in Figure 6, and as presented in the lemma below, the definition in (35) is motivated by the property that $t \mapsto r_t^i := r_i(y_t)$ is a càdlàg trajectory.

Lemma 3.8. *The function $t \mapsto r_t^i$ is càdlàg and lower semi-continuous, i.e., $r_t^i = \lim_{s \downarrow t} r_s^i \le \lim_{s \uparrow t} r_s^i$ for any $t \ge 0$. Moreover, it is continuous at points t such that $|\zeta_t^i| < \ell/2$ or $w_t^i = 0$.*
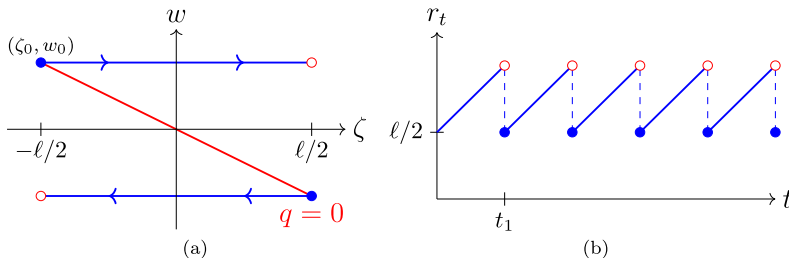
A proof of Lemma 3.8 is provided in Section 5.3.

Fig. 6. Plots of a piecewise-constant-velocity trajectory initially at $\zeta_0 = -\ell/2$, $w_0 > 0$ and $r_0 = |\zeta_0|$ that jumps at $t = t_1$ to $w_{t_1} = -w_0$ where $\zeta_{t_1} = \ell/2$ and $r_{t_1} = |\zeta_{t_1}|$ in (a) phase space and (b) r_t as a function of time. Note that $t \mapsto r_t$ is a càdlàg trajectory.

Let $a, \mathcal{R} > 0$, and define a function $f : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ as

$$f(r) := \int_0^r e^{-at} \mathbb{1}_{\{t \leq \mathcal{R}\}} \, dt = \frac{1}{a}\left(1 - e^{-ar \wedge \mathcal{R}}\right). \tag{36}$$

Note that f is nondecreasing, concave, bounded, and both constant and maximal when $r \geq \mathcal{R}$. Moreover, for all $s, r \geq 0$,

$$f(s) - f(r) \leq f'_-(r) \min(s - r, a^{-1}), \tag{37}$$

where $f'_-(r)$ is the left derivative of $f(r)$; see Lemma 5.2 for a proof of (37). To measure the distance between the components of the coupling process, we use the following distance function

$$\rho\big((x, v), (\tilde{x}, \tilde{v})\big) = \rho(y) = \sum_{i=1}^m f\big(r_i(y)\big). \tag{38}$$

This definition is motivated by [27] and [14] where similar distance functions have been introduced to obtain dimension-free contraction rates for (resp.) Langevin dynamics and HMC applied to models with weak interactions. For probability measures μ, ν on $\mathbb{T}_\ell^m \times \mathbb{R}^m$, we define

$$\mathcal{W}_\rho(\mu, \nu) := \inf_{\substack{(X, V) \sim \mu \\ (\tilde{X}, \tilde{V}) \sim \nu}} \mathbb{E}\big[\rho\big((X, V), (\tilde{X}, \tilde{V})\big)\big] \tag{39}$$

where the infimum is over all couplings of μ and ν.

We can now state our main contraction result for Andersen dynamics on $\mathbb{T}_\ell^m \times \mathbb{R}^m$. Let $(p_t)_{t \geq 0}$ denote the transition semigroup. The parameters defining the coupling and the metric are defined in the following way:

$$\mathcal{R} = \ell/2 + m/\big(\beta^{1/2}\lambda\big), \tag{40}$$

$$\gamma = 1/\big(\beta^{1/2}\mathcal{R}\big), \tag{41}$$

$$a = \beta^{1/2}\lambda/m, \text{ and} \tag{42}$$

$$\alpha = \sqrt{1 + \beta L \mathcal{R}^2}. \tag{43}$$

The choice of γ is motivated by the bound in (13) (see (62)), and the choice of the other parameters is motivated by the proof of the following theorem.

Theorem 3.9. *Suppose that Assumption* 3.7 *holds and let* $\lambda > 0$ *satisfy*

$$\beta^{1/2} \frac{\lambda}{m} \frac{\ell}{2} \geq \frac{25}{6} + 11\beta L \left(\frac{\ell}{2}\right)^2. \tag{44}$$

Suppose moreover that

$$J \leq \frac{e^{-1} - e^{-5}}{20(m - 1)\beta\ell^2} \max\big(\sqrt{\beta L \ell^2}, 1\big) \exp\left(-\beta^{1/2} \frac{\lambda}{m} \frac{\ell}{2}\right). \tag{45}$$

Then for all $y \in \mathbb{T}_\ell^m \times \mathbb{R}^{3m}$, *the process* $e^{c_A t} \rho(Y_t)$ *is a nonnegative supermartingale where*

$$c_A := \frac{1}{40e} \frac{\lambda}{m} \exp\left(-\beta^{1/2} \frac{\lambda}{m} \frac{\ell}{2}\right). \tag{46}$$

Moreover, for all probability measures ν *and* μ *on* $\mathbb{T}_\ell^m \times \mathbb{R}^m$ *we have*

$$\mathcal{W}_\rho(\mu p_t, \nu p_t) \leq e^{-c_A t} \mathcal{W}_\rho(\mu, \nu). \tag{47}$$

A proof of this theorem is provided in Section 5. Remarkably, the result captures the correct order of the dimension dependence for Andersen dynamics in the free-streaming case where $L = J = 0$ and condition (44) reduces to $\beta^{1/2}(\lambda/m)(\ell/2) \geq 25/6$. A corresponding bound holds for weak interactions, i.e., when J satisfies Condition (45). On

the other hand, a restriction on J can not be avoided. Indeed, for large values of the interaction parameter J, multiple invariant measures and phase transition phenomena in the mean field limit can cause an exponential degeneration of the rate of convergence to equilibrium as the number of particles m goes to infinity, even if λ is increased linearly with m [47,52,54].

4. Proofs in the weakly anharmonic case

Proof of Theorem 3.2. Here we apply the synchronous coupling described in Remark 2.6. Let $(Z_t, W_t) := (X_t - \tilde{X}_t, V_t - \tilde{V}_t)$. In between two consecutive jump times, $t \in [T_k, T_{k+1})$, note that the time derivative of (Z_t, W_t) satisfies

$$\frac{d}{dt}Z_t = W_t, \qquad \frac{d}{dt}W_t = -C^{-1}Z_t - \left(\nabla G(X_t) - \nabla G(\tilde{X}_t)\right),$$

with $Z_{T_k} = Z_{T_k-}$, $W_{T_k}^{I_k} = 0$, and $W_{T_k}^j = W_{T_k-}^j$ for $j \neq I_k$. In particular, when $G \equiv 0$ these differential equations become Hamiltonian with respect to the unperturbed Hamiltonian function $H_0(z, w)$, and hence for $y = (x, v, \tilde{x}, \tilde{v})$, $Z = x - \tilde{x}$ and $w = v - \tilde{v}$,

$$\left(\mathcal{L}^C H_0\right)(y) = -\left(\nabla G(x) - \nabla G(\tilde{x})\right) \cdot w \leq L_G |w||z| \tag{48}$$

where we applied in turn the definition of $\mathcal{L}^C$ in (16) and (25) in Assumption 3.1. Similarly, applying $\mathcal{L}^C$ to $\Psi(y) := z \cdot w$ gives

$$\left(\mathcal{L}^C \Psi\right)(y) = |w|^2 - z \cdot \left(\nabla U(x) - \nabla U(\tilde{x})\right) \leq |w|^2 - z^T C^{-1} z \tag{49}$$

where in the last step we used convexity of the perturbation $G(x)$. Applying $\mathcal{L}^C$ to ρ^2 in (27), and then inserting (48) and (49) yields

$$\left(\mathcal{L}^C \rho^2\right)(y) \leq \frac{\lambda}{4m}|w|^2 + \frac{\lambda^2}{4m^2}z \cdot w - \frac{\lambda}{4m}z^T C^{-1} z + L_G |w||z|. \tag{50}$$

By definition of $\mathcal{A}_{\text{sync}}^C$ in (17) and Remark 2.6,

$$\left(\mathcal{A}_{\text{sync}}^C \rho^2\right)(y) = -\frac{\lambda^2}{4m^2}z \cdot w - \frac{\lambda}{2m}|w|^2. \tag{51}$$

Combining (51) and (50) yields

$$\begin{aligned}
\left(\mathcal{G}_{\text{sync}}^C \rho^2\right)(y) &\leq -\frac{\lambda}{4m}\left(|w|^2 + z^T C^{-1} z - \frac{4L_G m}{\lambda}|w||z|\right) \\
&\leq -\frac{\lambda}{4m}\left(\frac{1}{2}|w|^2 + \left(z^T C^{-1} z - \frac{8L_G^2 m^2}{\lambda^2}|z|^2\right)\right) \\
&\leq -\frac{\lambda}{4m}H_0(z, w)
\end{aligned} \tag{52}$$

where in the last step we applied condition (29) and $\sigma_{\max}^{-2}|z|^2 \leq z^T C^{-1} z$ which together imply that $(8L_G^2 m^2/\lambda^2)|z|^2 \leq (1/2)\sigma_{\max}^{-2}|z|^2 \leq (1/2)z^T C^{-1} z$. Note that

$$\begin{aligned}
\rho(y)^2 &= H_0(z, w) + \frac{\lambda}{4m}z \cdot w + \frac{\lambda^2}{8m^2}|z|^2 \leq 2H_0(z, w) + \frac{5\lambda^2}{32m^2}|z|^2 \\
&\leq \max\left(2, \frac{5\sigma_{\max}^2 \lambda^2}{32m^2}\right)H_0(z, w)
\end{aligned} \tag{53}$$

where in the last step we again used $\sigma_{\max}^{-2}|z|^2 \leq z^T C^{-1} z$. Inserting (53) into (52) gives the required infinitesimal contraction result in (30).

For the corresponding Wasserstein bound, first, note from (28)

$$\lambda_{\min}(\mathsf{G})\big(|z|^2 + |w|^2\big) \leq \rho(y)^2 \leq \lambda_{\max}(\mathsf{G})\big(|z|^2 + |w|^2\big), \tag{54}$$

where $\lambda_{\min}(\mathsf{G})$ and $\lambda_{\max}(\mathsf{G})$ are the smallest and largest eigenvalues of the matrix G, respectively. Let $g(t, y) := e^{ct}\rho(y)^2$. Then by (30), $\frac{\partial g}{\partial t} + \mathcal{G}^C_{\text{sync}}g \leq 0$. Hence by [19, Theorem 5.5], the process $g(t, Y_t)$ is a non-negative supermartingale, and thus, $\mathbb{E}[\rho(Y_t)^2] \leq e^{-ct}\rho(y)^2$. Therefore, by the coupling characterization of the 2-Wasserstein metric and (54),

$$\mathcal{W}_2(\mu p_t, \nu p_t)^2 \leq \lambda_{\min}(\mathsf{G})^{-1}\mathbb{E}\big[\rho(Y_t)^2\big] \leq \varkappa(\mathsf{G})e^{-ct}\mathcal{W}_2(\mu, \nu)^2,$$

where $\varkappa(\mathsf{G}) = \lambda_{\max}(\mathsf{G})\lambda_{\min}(\mathsf{G})^{-1}$ is the condition number of G. By taking square roots, we obtain the bound in (31).

To bound the condition number of G, note that the eigenvalues of a 2×2 symmetric positive definite matrix $\mathsf{S} = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ satisfy:

$$\frac{a+c}{2} \pm \sqrt{\left(\frac{a-c}{2}\right)^2 + b^2} \leq \frac{a+c}{2} + \frac{|a+c|}{2} + b \leq 2\max(a, c) \tag{55}$$

where in the last step we used $b \leq \sqrt{ac} \leq (a+c)/2$. Moreover, from (28) note that $\mathsf{G}_{lo} \leq \mathsf{G} \leq \mathsf{G}_{up}$ where

$$\mathsf{G}_{lo} := \begin{bmatrix} (\frac{\lambda^2}{8m^2} + \frac{\sigma_{\max}^{-2}}{2})\mathbf{1}_{mn} & \frac{\lambda}{8m}\mathbf{1}_{mn} \\ \frac{\lambda}{8m}\mathbf{1}_{mn} & \frac{1}{2}\mathbf{1}_{mn} \end{bmatrix}, \qquad \mathsf{G}_{up} := \begin{bmatrix} (\frac{\lambda^2}{8m^2} + \frac{\sigma_{\min}^{-2}}{2})\mathbf{1}_{mn} & \frac{\lambda}{8m}\mathbf{1}_{mn} \\ \frac{\lambda}{8m}\mathbf{1}_{mn} & \frac{1}{2}\mathbf{1}_{mn} \end{bmatrix}.$$

Therefore, we obtain

$$\varkappa(\mathsf{G}) = \lambda_{\max}(\mathsf{G})\lambda_{\max}\big(\mathsf{G}^{-1}\big) \leq \lambda_{\max}(\mathsf{G}_{up})\lambda_{\max}\big(\mathsf{G}_{lo}^{-1}\big)$$

$$\leq \frac{\max(\lambda^2/m^2 + 4\sigma_{\max}^{-2}, 4)\max(\lambda^2/m^2 + 4\sigma_{\min}^{-2}, 4)}{3\lambda^2/m^2 + 16\sigma_{\max}^{-2}}$$

where in the last step we used (55). This gives the bound in (32). □

5. Proofs for Andersen dynamics on a high-dimensional torus

To prove contractivity of Andersen dynamics on $\mathbb{T}_\ell^m$, and as illustrated in Figure 7, we use the distance function $r_i(y)$ in (35) to decompose $\mathbb{T}_\ell^m \times \mathbb{R}^{3m}$ into the following sets: $\{r_i > \mathcal{R}\}$, $\{0 < r_i \leq \mathcal{R}\}$, and $Z_i := \{r_i = 0\}$. In addition, we introduce the following subset

$$B_i := \big\{y = (x, v, z, w) \in \mathbb{T}_\ell^m \times \mathbb{R}^{3m} : \zeta_i(z, w) = \ell/2 \text{ and } w_i = 0\big\}. \tag{56}$$
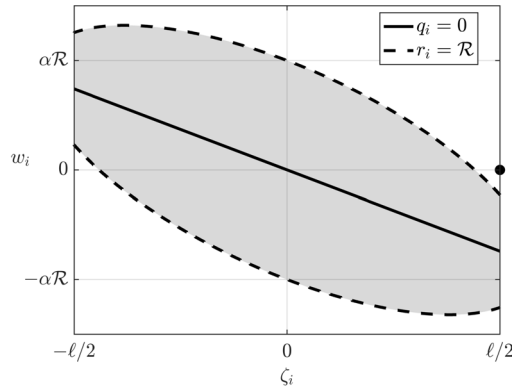

Fig. 7. The grey-shaded region corresponds to the sublevel set $\{r_i \leq \mathcal{R}\}$ whose boundary is the truncated ellipse indicated by the dashed black line. The solid black line corresponds to the line segment $q_i = \zeta_i + \gamma^{-1}w_i = 0$, which is on the long axis of the truncated ellipse. As noted in Remark 5.1, the point $(\zeta_i, w_i) = (\ell/2, 0)$ is not in the grey-shaded region.

The following remark shows that with the definition of $\mathcal{R}$ in (40), and under condition (44), $B_i \subset \{r_i > \mathcal{R}\}$, see also Figure 7.

Remark 5.1. In Theorem 3.9, the condition on λ in (44) implies

$$\beta^{1/2}\lambda\mathcal{R}/\mathsf{m} \geq 4 + 6\beta L\mathcal{R}^2.$$

Under this condition, by definition of $\mathcal{R}$ in (40),

$$\frac{\mathcal{R} - \ell/2}{\mathcal{R}} = \frac{\mathsf{m}}{\beta^{1/2}\lambda\mathcal{R}} \leq \frac{1}{4 + 6\beta L\mathcal{R}^2}$$

which implies that $\ell/(2\mathcal{R}) \geq (3 + 6\beta L\mathcal{R}^2)/(4 + 6\beta L\mathcal{R}^2)$ and hence

$$\frac{\ell}{2} < \mathcal{R} \leq \frac{4 + 6\beta L\mathcal{R}^2}{3 + 6\beta L\mathcal{R}^2}\frac{\ell}{2} = \frac{\ell}{2} + \frac{1}{3 + 6\beta L\mathcal{R}^2}\frac{\ell}{2}$$

In particular, $\mathcal{R} \leq (4/3)\ell/2$, and consequently, for all $y \in B_i$,

$$r_i(y) = \sqrt{1 + \alpha^{-2}}|\zeta_i| = \sqrt{1 + \alpha^{-2}}(\ell/2) \geq \left(1 + (\sqrt{2} - 1)\alpha^{-2}\right)(\ell/2)$$

$$\geq \left(1 + \frac{\sqrt{2} - 1}{1 + \beta L\mathcal{R}^2}\right)\frac{\ell}{2} \geq \left(1 + \frac{1/3}{1 + \beta L\mathcal{R}^2}\right)\frac{\ell}{2} \geq \mathcal{R},$$

where we used the inequality $\sqrt{1 + \mathsf{x}} \geq 1 + (\sqrt{2} - 1)\mathsf{x}$ valid for all $\mathsf{x} \in [0, 1]$, and (43) to eliminate α.

By Remark 5.1, for all $y \in B_i$, $f(r_i(y)) = f(\mathcal{R})$ is constant and maximal. As we will see below in Lemma 5.7, this observation simplifies the bounds on the metric along the deterministic flow of (20) starting at $y \in B_i$.

Lemma 5.2. *For all $a > 0$, $\mathcal{R} > 0$ and $s, r \geq 0$, the function f in (36) satisfies*

$$f(s) - f(r) \leq f'_-(r)\min(a^{-1}, s - r).$$

Proof. The function f is concave, and hence, for all $s, r \geq 0$,

$$f(s) - f(r) \leq f'_-(r)(s - r). \tag{57}$$

In the case $s \leq r$, the inequality $f(s) - f(r) \leq f'_-(r)a^{-1}$ is trivial since f is nondecreasing. In the case $s > r > \mathcal{R}$, $f(s) = f(r) = f(\mathcal{R})$ and $f'(r) = 0$, and the inequality holds with equality. Otherwise, if $s > r$ and $r \leq \mathcal{R}$, from (36)

$$f(s) - f(r) = \int_r^s e^{-at}\mathbb{1}_{\{t \leq \mathcal{R}\}}\,dt = e^{-ar}\int_r^s e^{-a(t-r)}\mathbb{1}_{\{t \leq \mathcal{R}\}}\,dt$$

$$\leq f'_-(r)\frac{1 - e^{-a(s-r)}}{a} \leq f'_-(r)a^{-1}.$$

Thus, $f(s) - f(r) \leq f'(r)a^{-1}$ holds for all $s, r \geq 0$. Combining this with (57) gives the required inequality. □

5.1. *Bounds for coupling of velocities*

Proof of Lemma 2.4. Let $\mathbf{1}_\mathsf{n}$ be the $\mathsf{n} \times \mathsf{n}$ identity matrix and introduce $w' = \xi - \tilde{\xi}$. Noting that $\mathbb{P}(w' \neq -\gamma\mathsf{b}) = d_{\mathrm{TV}}(\mathcal{N}(0, \beta^{-1}\mathbf{1}_\mathsf{n}), \mathcal{N}(\gamma\mathsf{b}, \beta^{-1}\mathbf{1}_\mathsf{n}))$ [12, Section 2.3.2], scale invariance of the total variation distance implies

$$\mathbb{P}(w' \neq -\gamma\mathsf{b}) = d_{\mathrm{TV}}(\mathcal{N}(0, \mathbf{1}_\mathsf{n}), \mathcal{N}(\sqrt{\beta}\gamma\mathsf{b}, \mathbf{1}_\mathsf{n})) = d_{\mathrm{TV}}(\mathcal{N}(0, 1), \mathcal{N}(\sqrt{\beta}\gamma|\mathsf{b}|, 1))$$

$$= 2\mathcal{N}(0, 1)[(0, \sqrt{\beta}\gamma|\mathsf{b}|/2)] \leq \sqrt{\beta}\gamma|\mathsf{b}|/\sqrt{2\pi}.$$

Hence (13) holds. Figure 4 of [11] illustrates the second to last step. To be sure, we note that $\mathcal{N}(0, 1)[(a, b)] = (2\pi)^{-1/2}\int_a^b e^{-t^2/2}\,dt$ for any $a, b \in \mathbb{R}$ with $a \leq b$.

When $b = 0$, the set $\{w' \neq -\gamma b\}$ is empty and (14) holds. Thus, suppose that $b \neq 0$. Then the set $\{x \in \mathbb{R}^n : x \cdot b = 0\}$ defines an $n - 1$ dimensional hyperplane. By (10),

$$
\begin{aligned}
\mathbb{E}\big(|\xi|^2; w' \neq -\gamma b\big) &= \int_{\mathbb{R}^n} |x|^2 \big(\varphi_\beta(x) - \varphi_\beta(x) \wedge \varphi_\beta(x + \gamma b)\big)\, dx \\
&= \int_{\mathbb{R}^n} |x|^2 \big(\varphi_\beta(x) - \varphi_\beta(x + \gamma b)\big)^+ dx \\
&= \int_{\mathbb{R}^n} \big|x - (1/2)\gamma b\big|^2 \big(\varphi_\beta\big(x - (1/2)\gamma b\big) - \varphi_\beta\big(x + (1/2)\gamma b\big)\big)^+ dx \\
&= \int_{\{b \cdot x \geq 0\}} \big|x - (1/2)\gamma b\big|^2 \big(1 - e^{-\beta \gamma b \cdot x}\big) \varphi_\beta\big(x - (1/2)\gamma b\big)\, dx = \mathrm{I} + \mathrm{II}
\end{aligned}
$$

where we introduced I and II

$$
\mathrm{I} = \int_{\{b \cdot x_\| \geq 0\}} |x_\perp|^2 \big(1 - e^{-\beta \gamma b \cdot x_\|}\big) \varphi_\beta\Big(x_\| + x_\perp - \frac{1}{2}\gamma b\Big)\, dx_\| \, dx_\perp, \tag{58}
$$

$$
\mathrm{II} = \int_{\{b \cdot x_\| \geq 0\}} \Big|x_\| - \frac{1}{2}\gamma b\Big|^2 \big(1 - e^{-\beta \gamma b \cdot x_\|}\big) \varphi_\beta\Big(x_\| + x_\perp - \frac{1}{2}\gamma b\Big)\, dx_\| \, dx_\perp, \tag{59}
$$

that involve a change of variables given by $x = x_\| + x_\perp$ with $x_\perp \cdot b = 0$.

Now let $\phi(s) = \exp(-(1/2)\beta|s|^2)/\sqrt{2\pi\beta^{-1}}$. Integration over the $x_\perp$ variable yields $\mathrm{I} = \beta^{-1}(n-1)\mathbb{P}(w' \neq -\gamma b)$ and

$$
\begin{aligned}
\mathrm{II} &= \int_0^\infty \big(1 - e^{-\beta\gamma|b|s}\big)|s - \gamma|b|/2|^2 \phi\big(s - \gamma|b|/2\big)\, ds \\
&= \int_0^\infty \big(1 - e^{-\beta\gamma|b|s}\big)\Big(\beta^{-2}\frac{d^2}{ds^2}\phi\big(s - \gamma|b|/2\big) + \beta^{-1}\phi\big(s - \gamma|b|/2\big)\Big)\, ds \\
&= \int_0^\infty \big(1 - e^{-\beta\gamma|b|s}\big)\beta^{-2}\frac{d^2}{ds^2}\phi\big(s - \gamma|b|/2\big)\, ds + \beta^{-1}\mathbb{P}(w' \neq -\gamma b) \\
&= -\gamma^2|b|^2 \int_0^\infty \phi\big(s + \gamma|b|/2\big)\, ds + \frac{\gamma|b|\phi(\gamma|b|/2) + \mathbb{P}(w' \neq -\gamma b)}{\beta} \\
&= -\frac{\gamma^2|b|^2}{2}\mathbb{P}\big(w' = -\gamma b\big) + \frac{\gamma|b|\phi(\gamma|b|/2) + \mathbb{P}(w' \neq -\gamma b)}{\beta} \tag{60}
\end{aligned}
$$

where in the last step integration by parts was used twice. Combining I and II with (13) gives $\mathbb{E}(|\xi|^2; w' \neq -\gamma b) \leq (n+1)\gamma|b|/\sqrt{2\pi\beta}$; hence, (14) holds. $\qquad\square$

5.2. *Bounds for Andersen collision operator acting on metric*

Lemma 5.3. *Suppose that $\lambda > 0$ satisfies condition* (44). *For any $i \in \{1, \ldots, m\}$,*

$$
\mathcal{A}_\gamma^C(f \circ r_i) \leq \begin{cases} -\frac{9}{25}\frac{\lambda}{m}\exp(-\beta^{1/2}\frac{\lambda}{m}\frac{\ell}{2})f \circ r_i & \text{if } r_i > \mathcal{R}, \\ -\gamma\big(-\frac{2}{5}|\zeta_i|^2 + (\frac{3}{10}\frac{\lambda}{\gamma m} - \frac{1}{5})\big)\frac{f' \circ r_i}{r_i} & \text{if } 0 < r_i \leq \mathcal{R}, \\ 0 & \text{if } r_i = 0. \end{cases} \tag{61}
$$

This lemma is a key ingredient to the (weak) infinitesimal contraction result developed in Theorem 5.8.

Proof. Fix $y \in \mathbb{T}_\ell^m \times \mathbb{R}^{3m}$. Let $I \sim \mathrm{Unif}\{1, \ldots, m\}$, $\xi \sim \mathcal{N}(0, \beta^{-1})^n$ and $\mathcal{U} \sim \mathrm{Unif}(0, 1)$ be independent random variables. We set $\tilde{\xi} = \Phi(\xi, \zeta_i, \mathcal{U})$ and introduce the shorthand $w_i' = \xi - \tilde{\xi}$. Inserting $\gamma = \beta^{-1/2}\mathcal{R}^{-1}$ from (41) into (13),

$$
\mathbb{P}\big(w_i' \neq -\gamma\zeta_i\big) \leq \sqrt{\beta}\gamma|z_i|/\sqrt{2\pi} \leq \sqrt{\beta}\gamma\mathcal{R}/\sqrt{2\pi} \leq 1/\sqrt{2\pi} < 2/5. \tag{62}
$$

Bound for $r_i(y) > \mathcal{R}$. On $[\mathcal{R}, \infty)$, f is constant and takes its maximum value. Therefore, $f(r_i(y)) = f(\mathcal{R})$ and

$$
\begin{aligned}
\mathcal{A}_\gamma^C(f \circ r_i)(y) &= \lambda \mathbb{E}\big(f\big(r_i\big(\mathcal{S}(I, \xi, \mathcal{U})y\big)\big) - f\big(r_i(y)\big)\big) \\
&\le \frac{\lambda}{\mathsf{m}} \mathbb{E}\big(f\big(r_i\big(\mathcal{S}(i, \xi, \mathcal{U})y\big)\big) - f\big(r_i(y)\big); w_i' = -\gamma \zeta_i\big) \\
&\le \frac{\lambda}{\mathsf{m}} \big(f\big(|\zeta_i|\big) - f(r_i)\big) \mathbb{P}\big(w_i' = -\gamma \zeta_i\big) \le -\frac{3}{5} \frac{\lambda}{\mathsf{m}} \left(1 - \frac{f(\ell/2)}{f(\mathcal{R})}\right) f\big(r_i(y)\big)
\end{aligned}
\tag{63}
$$

where in the last step we used $\mathcal{R} \ge \ell/2$ and (62) which implies that $\mathbb{P}(w_i' = -\gamma \zeta_i) \ge 3/5$. Since, by (40), $\mathcal{R} = 1/a + \ell/2$, and using $1 - e^{-1} \ge 3/5$, we have

$$
1 - f(\ell/2)/f(\mathcal{R}) = \big(1 - e^{-1}\big)/\big(e^{a\ell/2} - e^{-1}\big) \ge (3/5)e^{-a\ell/2}.
$$

Inserting this in (63) and eliminating a using (42) gives the required bound.

Bound for $0 < r_i(y) \le \mathcal{R}$. Let $r_i' = r_i(\mathcal{S}(i, \xi, \mathcal{U})y)$ and write

$$
\mathcal{A}_\gamma^C(f \circ r_i) = \mathrm{I} + \mathrm{II} \quad \text{where} \quad \begin{cases} \mathrm{I} := \frac{\lambda}{\mathsf{m}} \mathbb{E}(f(r_i') - f(r_i); w_i' = -\gamma \zeta_i) \\ \mathrm{II} := \frac{\lambda}{\mathsf{m}} \mathbb{E}(f(r_i') - f(r_i); w_i' \ne -\gamma \zeta_i) \end{cases}
\tag{64}
$$

For I, note that on $\{w_i' = -\gamma \zeta_i\}$,

$$
r_i' - r_i = |\zeta_i| - r_i = \big(|\zeta_i|^2 - r_i^2\big)/\big(|\zeta_i| + r_i\big) \le -\alpha^{-2}|q_i|^2/(2r_i).
$$

Combining this bound with (37) and (62), we obtain

$$
\mathrm{I} \le -\frac{\lambda}{\mathsf{m}} \alpha^{-2}|q_i|^2 \frac{f_-'(r_i)}{2r_i} \mathbb{P}\big(w_i' = -\gamma \zeta_i\big) \le -\frac{3}{10} \frac{\lambda}{\mathsf{m}} \alpha^{-2}|q_i|^2 \frac{f_-'(r_i)}{r_i}
\tag{65}
$$

For II, use (37), (62) and (42) to obtain

$$
\begin{aligned}
\mathrm{II} &\le \frac{\lambda}{\mathsf{m}} a^{-1} f_-'(r_i) \mathbb{P}\big(w_i' \ne -\gamma \zeta_i\big) \le \frac{\lambda}{\mathsf{m}} a^{-1} f_-'(r_i) \frac{\sqrt{\beta}\gamma |\zeta_i|}{\sqrt{2\pi}} \\
&= \frac{\gamma f_-'(r_i)|\zeta_i| r_i}{r_i \sqrt{2\pi}} \le \frac{2}{5} \gamma \left(|\zeta_i|^2 + \frac{\alpha^{-2}}{2}|q_i|^2\right) \frac{f_-'(r_i)}{r_i}.
\end{aligned}
\tag{66}
$$

Inserting (65) and (66) into (64) gives the required bound.

Bound for $r_i(y) = 0$. In this case, $\zeta_i = w_i = 0$, and thus, $\mathcal{S}(i, \xi, \mathcal{U})y = y$, i.e., $\mathcal{A}_\gamma^C(f \circ r_i)(y) = 0$, as required. $\qquad \square$

5.3. *Regularity of distance function under flow of* (20)

Here we prove Lemma 3.8 – a key ingredient to controlling boundary effects for $|\zeta^i| = \ell/2$ and $w^i \ne 0$. The following remark is useful in the proof.

As illustrated in Fig. 8, $(z_i, w_i) \mapsto |\zeta_i|$ is a contraction in the sense that

$$
\big||\zeta_i| - |\tilde{\zeta}_i|\big| \le |z_i - \tilde{z}_i|.
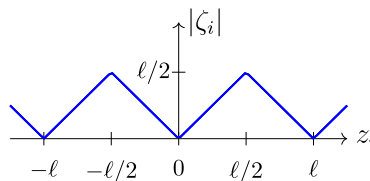\tag{67}
$$

Fig. 8. This figure illustrates that $(z_i, w_i) \mapsto |\zeta_i|$ is contractive.

Proof of Lemma 3.8. Let $r_t^i = r_i(y_t)$ and $\zeta_t^i = \zeta_i(z_t, w_t)$ where $i \in \{1, \ldots, \mathsf{m}\}$ is fixed and y_t is the deterministic solution of (20) starting at $y_0 = y$. Recall from (35) that $r_t^i = \sqrt{|\zeta_t^i|^2 + \alpha^{-2}|\zeta_t^i + \gamma^{-1}w_t^i|^2}$. The function $t \mapsto (z_t^i, w_t^i)$ is continuous. Moreover, ζ_i in (21) is continuous at points where $|\zeta_i| < \ell/2$. Therefore, if $|\zeta_t^i| < \ell/2$, then ζ_t^i and r_t^i are continuous at t.

Suppose, next, that at time t, $\zeta_t^i = -\ell/2$ and $w_t^i > 0$. Since $w_t^i > 0$, z_t^i is strictly increasing in an open interval containing t. Therefore, for sufficiently small $h > 0$,

$$\zeta_{t+h}^i - \zeta_t^i = z_{t+h}^i - z_t^i, \quad \text{and} \quad \zeta_t^i - \zeta_{t-h}^i = (z_t^i - \ell) - z_{t-h}^i.$$

Hence, $\lim_{h\downarrow 0}(\zeta_{t+h}^i - \zeta_t^i) = 0$ while $\lim_{h\downarrow 0}(\zeta_t^i - \zeta_{t-h}^i) = -\ell$, and in particular, (ζ_t^i, w_t^i) is càdlàg, and hence, r_t^i is càdlàg as well. Moreover, $(z_i, w_i) \mapsto |\zeta_i|$ is continuous because $(z_i, w_i) \mapsto |\zeta_i|$ is a contraction by (67), and thus,

$$|q_t^i|^2 = |\zeta_t^i + \gamma^{-1}w_t^i|^2 = |\zeta_t^i|^2 + \gamma^{-2}|w_t^i|^2 + 2\gamma^{-1}\zeta_t^i w_t^i \leq \lim_{s\uparrow t}|q_s^i|^2$$

because $w_t^i > 0$ and $\zeta_t^i \leq \zeta_{t-}^i$. Therefore, $r_t^i \leq \lim_{s\uparrow t} r_s^i$. The case $\zeta_t^i = \ell/2$ and $w_t^i < 0$ can be treated similarly; in this case $\lim_{h\downarrow 0}(\zeta_t^i - \zeta_{t-h}^i) = \ell$.

Finally, suppose that at time t, $\zeta_t^i = \ell/2$ and $w_t^i = 0$. In this case, r_i is itself continuous at y_t, and therefore, $s \mapsto r_s^i = r_i(y_s)$ is continuous at t. Continuity of r_i at y_t follows from $|q_i(z, w)| \to |\zeta_i(z, w)|$ as $w_i \to 0$. □

Remark 5.4. By Lemma 3.8, $t \mapsto r_t^i$ is a càdlàg trajectory. Therefore, for any $\epsilon > 0$ and for any $T > 0$, the number of jumps of size greater than ϵ, i.e., $\#\{t \in [0, T] : |r_t^i - r_{t-}^i| > \epsilon\}$, is finite [28]. However, for a trajectory starting in B_i where $(\zeta_0^i, w_0^i) = (\ell/2, 0)$, it is still possible that there are infinitely many jumps in every interval $(0, h)$ with $h > 0$, i.e., the underlying trajectory $t \mapsto (\zeta_t^i, w_t^i)$ may wind around the point $(\ell/2, 0)$ infinitely often. For the bounds on the deterministic part of the dynamics, we avoid this potential complication by selecting $\mathcal{R}$ and λ such that $B_i \subset \{r_i(y) > \mathcal{R}\}$ where $f(r_i(y)) = f(\mathcal{R})$ is constant and maximal; see Remark 5.1.

5.4. *Bounds for Liouville operator acting on metric*

Since r_i in (35) lacks continuity at boundary points where $|\zeta_i| = \ell/2$, the domain of $\mathcal{L}^C$ excludes ρ. Nonetheless, by Lemma 3.8, $t \mapsto r_t^i$ is a càdlàg trajectory. This càdlàg time regularity motivates defining the following right-sided directional derivative of a function along the deterministic flow of (20).

Definition 5.5. For a function $g : \mathbb{T}_\mathsf{m}^\ell \to \mathbb{R}$, define

$$\mathcal{L}^C g(y) := \lim_{h\downarrow 0} \frac{g(\phi_h^C(y)) - g(y)}{h} \quad \text{whenever the limit exists.}$$

According to this definition, $\mathcal{L}^C(f \circ r_i)$ is well-defined at most boundary points, and in particular,

$$\mathcal{L}^C \zeta_i(y) = w_i(y) \quad \text{for all } y \in (\mathbb{T}_\ell^\mathsf{m} \times \mathbb{R}^{3\mathsf{m}}) \setminus B_i. \tag{68}$$

This is because when the deterministic flow is at a boundary point at time t with either $\zeta_i = -\ell/2$ and $w_i > 0$, or $\zeta_i = \ell/2$ and $w_i < 0$, there exists a time interval $[t, t+h)$ such that the trajectory $s \mapsto \zeta_s^i$ is streatly increasing, (respectively, strictly decreasing) on $[t, t+h)$, and hence, there exists an integer k such that $\zeta_s^i = z_s^i + k\ell$ for all $s \in [t, t+h)$. Moreover,

$$\mathcal{L}^C w_i(y) = \nabla_i U(\tilde{x}) - \nabla_i U(x) \quad \text{for all } y \in \mathbb{T}_\ell^\mathsf{m} \times \mathbb{R}^{3\mathsf{m}}. \tag{69}$$

Since $r_i = \sqrt{|\zeta_i|^2 + \alpha^{-2}|\zeta_i + \gamma^{-1}w_i|^2}$, r_i^2 is a smooth function of (ζ_i, w_i), and r_i is a smooth function of (ζ_i, w_i) except at $(\zeta_i, w_i) = (0, 0)$. Thus, $\mathcal{L}^C(r_i^2)$ exists for all $y \in (\mathbb{T}_\ell^\mathsf{m} \times \mathbb{R}^{3\mathsf{m}}) \setminus B_i$ and $\mathcal{L}^C r_i$ exists for all $y \in (\mathbb{T}_\ell^\mathsf{m} \times \mathbb{R}^{3\mathsf{m}}) \setminus (Z_i \cup B_i)$.

Expanding on this point, by (68) and (69), for all $y \in (\mathbb{T}_\ell^\mathsf{m} \times \mathbb{R}^{3\mathsf{m}}) \setminus B_i$,

$$\mathcal{L}^C(r_i^2) = 2\left(\zeta_i w_i + \alpha^{-2} q_i w_i + \frac{1}{\gamma\alpha^2} q_i (\nabla_i U(\tilde{x}) - \nabla_i U(x))\right)$$

$$= 2\gamma\left(-|\zeta_i|^2 + \frac{|q_i|^2}{\alpha^2} + (1 - \alpha^{-2})\zeta_i q_i + \frac{1}{\gamma^2\alpha^2} q_i (\nabla_i U(\tilde{x}) - \nabla_i U(x))\right)$$

$$\leq 2\gamma\left(-|\zeta_i|^2 + \frac{|q_i|^2}{\alpha^2} + \left(1 - \alpha^{-2} + \frac{L}{\gamma^2\alpha^2}\right)|\zeta_i||q_i| + \frac{J|q_i|}{\gamma^2\alpha^2}\sum_{k\neq i}|\zeta_k|\right)$$

$$\leq 2\gamma\left(-|\zeta_i|^2 + \frac{|q_i|^2}{\alpha^2} + 2\frac{\alpha^2-1}{\alpha^2}|\zeta_i||q_i| + \frac{J|q_i|}{\gamma^2\alpha^2}\sum_{k\neq i}|\zeta_k|\right) \tag{70}$$

where, in turn, we eliminated w_i using $w_i = \gamma(q_i - \zeta_i)$, used (34) to bound $|\nabla_i U(\tilde{x}) - \nabla_i U(x)|$, and used $L\gamma^{-2} = \alpha^2 - 1$ which follows from (41) and (43). For all $y \in (\mathbb{T}_\ell^m \times \mathbb{R}^{3m}) \setminus (B_i \cup Z_i)$, the chain rule and (70) imply

$$\mathscr{L}^C r_i \leq \frac{\gamma}{r_i}\left(-|\zeta_i|^2 + \frac{|q_i|^2}{\alpha^2} + 2\frac{\alpha^2-1}{\alpha^2}|\zeta_i||q_i| + \frac{J|q_i|}{\gamma^2\alpha^2}\sum_{k\neq i}|\zeta_k|\right) \tag{71}$$

$$\leq \gamma\alpha r_i + \frac{J}{\gamma\alpha}\sum_{k\neq i}\min(r_k, \ell/2) \tag{72}$$

where we used $|q_i|/\alpha \leq r_i$, $|\zeta_k| \leq \min(r_k, \ell/2)$, and the inequality

$$-|\zeta_i|^2 + \frac{|q_i|^2}{\alpha^2} + 2\frac{\alpha^2-1}{\alpha^2}|\zeta_i||q_i| \leq \left(\frac{1}{\delta}\frac{\alpha^2-1}{\alpha^2} - 1\right)|\zeta_i|^2 + \left(1 + \delta(\alpha^2-1)\right)\alpha^{-2}|q_i|^2$$

$$\leq \max\left(\frac{\alpha^2-1}{\delta\alpha^2} - 1, 1 + \delta(\alpha^2-1)\right)r_i^2 \leq \alpha r_i^2$$

where in the first step we used Young's inequality with $\delta > 0$ and in the second step we optimize the bound by choosing $\delta > 0$ such that $(\alpha^2 - 1)/(\delta\alpha^2) - 1 = 1 + \delta(\alpha^2 - 1) = \sqrt{\alpha^{-2} - 1 + \alpha^2} \leq \alpha$, since $\alpha \geq 1$.

Remark 5.6. For $y \in (\mathbb{T}_\ell^m \times \mathbb{R}^{3m}) \setminus B_i$, (72) holds in a weak sense. To see this, approximate $r_i(y)$ by $r_{i,\epsilon}(y) := \varphi_\epsilon((r_i(y))^2)$ where $\epsilon > 0$ is a small parameter and φ_ϵ is a C^1 function defined by $\varphi_\epsilon(x) = \sqrt{x}$ for $x \geq \epsilon^2$ and $\varphi_\epsilon(x) = \epsilon/2 + x/(2\epsilon)$ for $x \leq \epsilon^2$. By the standard chain rule,

$$\mathscr{L}^C r_{i,\epsilon} = \varphi_\epsilon'\left((r_i)^2\right)\mathscr{L}^C (r_i)^2 = \frac{1}{2\max(r_i, \epsilon)}\mathscr{L}^C (r_i)^2,$$

and thus for any $y_0 \in (\mathbb{T}_\ell^m \times \mathbb{R}^{3m}) \setminus B_i$ and $t \geq 0$ sufficiently small,

$$r_{i,\epsilon}(y_t) - r_{i,\epsilon}(y_0) = \int_0^t \frac{\mathscr{L}^C (r_i)^2(y_s)}{2\max(r_s^i, \epsilon)}\,ds \leq \int_0^t \left(\gamma\alpha r_s^i + \frac{J}{\gamma\alpha}\sum_{k\neq i}\min(r_s^k, \ell/2)\right)ds.$$

As $\epsilon \downarrow 0$, $r_{i,\epsilon}(y_t) \downarrow r_t^i$ and thus we obtain the same bound for r_t^i, i.e., (72) holds in a weak sense.

Lemma 5.7. *Suppose that $\lambda > 0$ satisfies condition* (44). *For any $i \in \{1, \ldots, m\}$, for any initial condition $y \in \mathbb{T}_\ell^m \times \mathbb{R}^{3m}$, for any $t > 0$, and for any $s \in [0, t]$,*

$$f(r_t^i) - f(r_s^i) \leq \int_s^t g_i(y_u)\,du, \quad where\ g_i : \mathbb{T}_\ell^m \times \mathbb{R}^{3m} \to \mathbb{R}\ is\ given\ by \tag{73}$$

$$g_i(y) := \begin{cases} 0 & for\ r_i > \mathcal{R}, \\ f'(r_i(y))\mathscr{L}^C r_i(y) & for\ 0 < r_i \leq \mathcal{R}, \\ J/(\gamma\alpha)\sum_{k\neq i}\min(r_k(y), \ell/2) & for\ r_i = 0. \end{cases} \tag{74}$$

Lemma 5.7 states that in the weak sense $\mathscr{L}^C(f \circ r_i) \leq g_i$. We know this holds with equality for $y \in \mathbb{T}_\ell^m \times \mathbb{R}^{3m}$ such that $|\zeta_i| < \ell/2$ and $y \notin Z_i$. This lemma extends this equality to an inequality that is valid globally.

Note that condition (44) is used below in order to ensure that the point $(\zeta_i, w_i) = (\ell/2, 0)$ lies in the set where $r_i > \mathcal{R}$ and therefore $f(r_i)$ is constant and maximal; see Remark 5.1 and Figure 7. This substantially simplifies the proof, since we can only ensure that $t \mapsto \zeta_i(z_t, w_t)$ is a càdlàg function if $w_t^i \neq 0$.

Proof. It suffices to prove (73) for $s = 0$, and use the flow property to write $f(r_t^i) - f(r_s^i) = f \circ r_i(\phi_t^C(y)) - f \circ r_i(\phi_s^C(y)) = f \circ r_i(\phi_{t-s}^C(y_s)) - f \circ r_i(y_s)$, i.e., start the underlying flow with initial condition y_s instead of y. From now on, we assume w.l.o.g. that $s = 0$. Fix an $\epsilon > 0$ and introduce the function $g_{i,\epsilon} : \mathbb{T}_\ell^m \times \mathbb{R}^{3m} \to \mathbb{R}$ defined by

$$g_{i,\epsilon}(y) := \begin{cases} g_i(y) & r_i > \epsilon, \\ \gamma \alpha r_i(y) + J/(\gamma \alpha) \sum_{k \neq i} \min(r_k(y), \ell/2) & r_i \leq \epsilon. \end{cases}$$

Below we prove (73) holds with g_i replaced with $g_{i,\epsilon}$, i.e.,

$$f(r_t^i) - f(r_0^i) \leq \int_0^t g_{i,\epsilon}(y_u) \, du. \tag{75}$$

Then (73) follows since as $\epsilon \downarrow 0$ we have $g_{i,\epsilon} \downarrow g_i$. Define

$$\tau := \sup\{u \geq 0 : (75) \text{ holds for all } t \in [0, u]\}.$$

We will prove $\tau = \infty$ by contradiction. Hence suppose $\tau < \infty$. By Lemma 3.8 and monotonicity of f, (75) holds for all $t \in [0, \tau]$ with τ included. Indeed, by definition of τ, (75) holds for $t < \tau$. Moreover, by Lemma 3.8,

$$f(r_\tau^i) - f(r_0^i) \leq \lim_{t \uparrow \tau} f(r_t^i) - f(r_0^i).$$

Thus, since the r.h.s. of (75) is continuous in t, this bound extends from $t < \tau$ to $t = \tau$.

Now we distinguish several cases depending on the size of r_τ^i.

Case (i): $r_\tau^i > \mathcal{R}$. Note, first, that this case includes $r_\tau^i \in B_i$ by Remark 5.1. In this case, by right continuity of $t \mapsto r_t^i$, there exists $h > 0$ such that for all $t \in [\tau, \tau + h]$ we have $r_t^i > \mathcal{R}$, and hence, $f'(r_t^i) = 0$ and $f(r_t^i) = f(\mathcal{R}) = f(r_\tau^i)$. Inserting these results into (75) gives $f(r_t^i) - f(r_0^i) \leq \int_0^\tau g_{i,\epsilon}(y_u) \, du = \int_0^t g_{i,\epsilon}(y_u) \, du$. Thus, (75) holds for all $t \in [\tau, \tau + h]$, which contradicts the definition of τ.

Case (ii): $\epsilon < r_\tau^i \leq \mathcal{R}$ *and* $|\zeta_\tau^i| < \ell/2$. In this case, there exists $h > 0$ such that $|\zeta_t^i| < \ell/2$ and $r_t^i > \epsilon$ for all $t \in [\tau - h, \tau + h]$. Therefore, $t \mapsto r_t^i$ is smooth on this interval, and thus for $t \in [\tau - h, \tau + h]$, since f is Lipschitz continuous,[2] it is also absolutely continuous, and therefore, $f(r_t^i) - f(r_\tau^i) = \int_\tau^t f'(r_u^i) \mathscr{L}^C r_i(y_u) \, du$. Thus, (75) holds for all $t \in [\tau, \tau + h]$, which contradicts the definition of τ. Here we used that every Lipschitz continuous function is absolutely continuous.

Case (iii): $\epsilon < r_\tau^i \leq \mathcal{R}$ *and* $|\zeta_\tau^i| = \ell/2$. This case can be treated similarly to case (ii). Note, first, that $\ell/2 \leq r_\tau^i \leq \mathcal{R}$, and hence, $r_\tau^i \notin B_i$ by Remark 5.1. Suppose, for example, that $\zeta_\tau^i = -\ell/2$ and $w_\tau^i > 0$. Then z_t^i is strictly increasing for t near τ. Therefore, for $t \in [\tau, \tau + h]$ with h sufficiently small, ζ_t^i is strictly increasing, $\zeta_t^i \in (-\ell/2, 0)$ and $r_t^i > \epsilon$. In particular, for $t \in [\tau, \tau + h]$, $\zeta_t^i = z_t^i + k\ell$ for a fixed integer k, and thus, r_t^i is a smooth function for all $t \in [\tau, \tau + h]$. (The only difference to case (ii) is that now, smoothness of r_t^i holds only for $t \in [\tau, \tau + h]$, and not for $t \in [\tau - h, \tau + h]$.) Now we can argue completely analogously to case (ii) to conclude that (75) holds for all $t \in [\tau, \tau + h]$, which contradicts the definition of τ.

Case (iv): $0 < r_\tau^i \leq \epsilon$. In this case, there exists $h > 0$, such that r_t^i is smooth for $t \in [\tau - h, \tau + h]$, and similar to case (ii), $f(r_t^i) - f(r_\tau^i) = \int_\tau^t f'(r_u^i) \mathscr{L}^C r_i(y_u) \, du$. However, we only have $r_t^i > 0$. This motivates inserting $0 \leq f' \leq 1$ into the bound in (72) to obtain that for $t \in [\tau - h, \tau + h]$

$$f'(r_t^i) \mathscr{L}^C r_i(y_t) \leq \gamma \alpha r_t^i + J/(\gamma \alpha) \sum_{k \neq i} \min(r_t^k, \ell/2).$$

Thus, (75) holds for all $t \in [\tau, \tau + h]$, which contradicts the definition of τ.

---

[2]From (36), note that f is a composition of two Lipschitz functions, and hence,

$$\left| f(r) - f(s) \right| = a^{-1} \left| e^{-a(r \wedge \mathcal{R})} - e^{-a(s \wedge \mathcal{R})} \right| \leq a |r \wedge \mathcal{R} - s \wedge \mathcal{R}| \leq a |r - s| \quad \text{for all } r, s \in [0, \infty).$$

Case (v): $r_\tau^i = 0$.   In this case r_t^i is not smooth at τ, but as noted in Remark 5.6 the bound in (72) can still be applied in a weak sense. In particular, for $t \in [\tau, \tau + h]$ with h sufficiently small we have $r_t^i < \epsilon$ and

$$f(r_t^i) - f(r_\tau^i) \leq \int_\tau^t \left(\gamma \alpha r_u^i + J/(\gamma \alpha) \sum_{k \neq i} \min(r_u^k, \ell/2) \right) du.$$

Hence, we see again that for h sufficiently small, (75) extends to $t \in [\tau, \tau + h]$, but this contradicts the definition of τ. $\square$

5.5. *Combined bounds for generator of Andersen dynamics on $\mathbb{T}_\ell^m$ acting on metric*

The following theorem uses Lemmas 5.3 and 5.7 to bound, in the weak sense, the generator of Andersen dynamics acting on ρ.

Theorem 5.8. *Suppose that $\lambda > 0$ satisfies condition (44), and (45) holds. Then for every $y \in \mathbb{T}_\ell^m \times \mathbb{R}^{3m}$*

$$\sum_i \left(g_i + \mathcal{A}_\gamma^C (f \circ r_i) \right) \leq -c_A \sum_i f(r_i), \quad \text{where } c_A \text{ is defined in (46)}.$$

Theorem 5.8 states that in the weak sense

$$\frac{d}{dt} \rho(Y_t) = \mathcal{G}_\gamma^C \rho(Y_t) \leq -c_A \rho(y),$$

for every initial condition $y \in \mathbb{T}_\ell^m \times \mathbb{R}^{3m}$.

Proof. First we combine the bounds from Lemmas 5.3 and 5.7 to obtain a global component-wise bound, and then sum over these bounds to obtain an overall global bound on the generator of Andersen dynamics acting on ρ.

Bound for $r_i > \mathcal{R}$.   Applying Lemmas 5.3 and 5.7 in this case gives

$$g_i + \mathcal{A}_\gamma^C (f \circ r_i) \leq -\tilde{c}_0 f(r_i), \text{ where } \tilde{c}_0 := \frac{9}{25} \frac{\lambda}{m} \exp\left(-\beta^{1/2} \frac{\lambda}{m} \frac{\ell}{2} \right). \tag{76}$$

Bound for $0 < r_i \leq \mathcal{R}$.   By Lemma 5.7 and (71),

$$g_i = f'(r_i) \mathcal{L}^C r_i \leq -\gamma \frac{f'(r_i)}{r_i} \left(|\zeta_i|^2 - \frac{|q_i|^2}{\alpha^2} + 2\frac{1 - \alpha^2}{\alpha^2} |\zeta_i||q_i| - \frac{J|q_i|}{\gamma^2 \alpha^2} \sum_{k \neq i} |\zeta_k| \right)$$

Combining this bound with Lemma 5.3, we obtain

$$g_i + \mathcal{A}_\gamma^C (f \circ r_i) \leq -\gamma \frac{f'(r_i)}{r_i} \left(\frac{1}{20} r_i^2 + Q(|\zeta_i|, |q_i|) - \frac{J}{\gamma^2 \alpha^2} \sum_{k \neq i} |\zeta_k||q_i| \right).$$

Here we have introduced the quadratic form

$$Q(|\zeta_i|, |q_i|) := \frac{11}{20} |\zeta_i|^2 + \left(\frac{3}{10} \frac{\lambda}{\gamma m} - \frac{5}{4} \right) \frac{|q_i|^2}{\alpha^2} - 2\frac{\alpha^2 - 1}{\alpha^2} |\zeta_i||q_i|$$

$$= \frac{11}{20} |\zeta_i|^2 + \left(\frac{3}{10} \frac{\beta^{1/2} \lambda \mathcal{R}}{m} - \frac{5}{4} \right) \frac{|q_i|^2}{1 + \beta L \mathcal{R}^2} - 2\frac{\beta L \mathcal{R}^2}{1 + \beta L \mathcal{R}^2} |\zeta_i||q_i|$$

where in the last expression we eliminated γ and α using (41) and (43). This quadratic form is nonnegative provided that

$$\left(\frac{\beta L \mathcal{R}^2}{1 + \beta L \mathcal{R}^2} \right)^2 \leq \frac{33}{200} \left(\frac{\beta^{1/2} \lambda \mathcal{R}}{m} - \frac{25}{6} \right) \frac{1}{1 + \beta L \mathcal{R}^2}.$$

A sufficient condition for this condition to hold is

$$\frac{\beta^{1/2} \lambda \mathcal{R}}{m} \geq \frac{25}{6} + \frac{200}{33} \beta L \mathcal{R}^2. \tag{77}$$

Moreover, since (as noted in Remark 5.1) $\mathcal{R} \leq (4/3)(\ell/2)$, condition (44) implies condition (77) because

$$\frac{\beta^{1/2}\lambda\mathcal{R}}{\mathsf{m}} \geq \frac{\beta^{1/2}\lambda}{\mathsf{m}}\frac{\ell}{2} \geq \frac{25}{6} + 11\frac{9}{16}\beta L\mathcal{R}^2 \geq \frac{25}{6} + \frac{200}{33}\beta L\mathcal{R}^2.$$

Thus, under condition (44), we obtain

$$\mathsf{g}_i + \mathcal{A}_\gamma^C(f \circ r_i) \leq -\gamma\frac{f'(r_i)}{r_i}\left(\frac{1}{20}r_i^2 - \frac{J}{\gamma^2\alpha^2}\sum_{k \neq i}|\zeta_k||q_i|\right)$$

$$\leq -\frac{\gamma}{20}r_i f'(r_i) + \frac{J}{\gamma\alpha}\sum_{k \neq i}|\zeta_k| = -\frac{\gamma}{20}\frac{ar_i}{e^{ar_i}-1}f(r_i) + \frac{J}{\gamma\alpha}\sum_{k \neq i}|\zeta_k|$$

$$\leq -\tilde{\tilde{c}}_0 f(r_i) + \frac{J}{\gamma\alpha}\sum_{k \neq i}\min(r_k, \ell/2), \quad \text{where } \tilde{\tilde{c}}_0 := \frac{\gamma}{20}\frac{a\mathcal{R}}{e^{a\mathcal{R}}-1},$$

where we used monotonicity of $\mathsf{x}/(e^\mathsf{x}-1)$ for $\mathsf{x} > 0$. By (40) and (42),

$$c_0 := \frac{\lambda}{20e\mathsf{m}}e^{-\beta^{1/2}\frac{\lambda}{\mathsf{m}}\frac{\ell}{2}} \leq \tilde{\tilde{c}}_0 = \frac{\lambda}{20\mathsf{m}}\frac{e^{-\beta^{1/2}\frac{\lambda}{\mathsf{m}}\frac{\ell}{2}}}{e - e^{-\beta^{1/2}\frac{\lambda}{\mathsf{m}}\frac{\ell}{2}}} \leq \frac{\lambda}{20\mathsf{m}}e^{-\beta^{1/2}\frac{\lambda}{\mathsf{m}}\frac{\ell}{2}} < \tilde{c}_0$$

where we used $1 \leq e - e^{-\beta^{1/2}\frac{\lambda}{\mathsf{m}}\frac{\ell}{2}} \leq e$. Hence, the following holds for $r_i > 0$,

$$\mathsf{g}_i + \mathcal{A}_\gamma^C(f \circ r_i) \leq -c_0 f(r_i) + J/(\gamma\alpha)\sum_{k \neq i}\min(r_k, \ell/2). \tag{78}$$

Bound for $r_i = 0$. Applying Lemmas 5.3 and 5.7 in this case gives

$$\mathsf{g}_i + \mathcal{A}_\gamma^C(f \circ r_i) \leq J/(\gamma\alpha)\sum_{k \neq i}\min(r_k, \ell/2). \tag{79}$$

Thus, the component-wise bound in (78) holds globally under condition (44).

Overall global bound. Summing over the component-wise bounds in (78),

$$\sum_i\left(\mathsf{g}_i + \mathcal{A}_\gamma^C(f \circ r_i)\right) \leq -c_0\sum_i f(r_i) + J/(\gamma\alpha)(\mathsf{m}-1)\sum_i\min(r_i, \ell/2),$$

$$\leq \left(-c_0 + \frac{J}{\gamma\alpha}(\mathsf{m}-1)\frac{\ell/2}{f(\ell/2)}\right)\sum_i f(r_i),$$

$$\leq \left(-c_0 + \frac{J}{\gamma\alpha}(\mathsf{m}-1)\frac{\beta^{1/2}\frac{\lambda}{\mathsf{m}}\frac{\ell}{2}}{1 - e^{-\beta^{1/2}\frac{\lambda}{\mathsf{m}}\frac{\ell}{2}}}\right)\sum_i f(r_i),$$

$$\leq -\left(c_0 - \frac{\mathsf{m}-1}{1 - e^{-4}}\frac{J}{\gamma\alpha}\beta^{1/2}\frac{\lambda}{\mathsf{m}}\frac{\ell}{2}\right) \leq -c_A\sum_i f(r_i), \quad \text{as required.} \tag{80}$$

Here, in turn, we used that $\min(r_i, \ell/2) \leq (\ell/2)(f(r_i)/f(\ell/2))$; condition (77), which implies $\beta^{1/2}(\lambda/\mathsf{m})(\ell/2) \geq 25/6 > 4$ and hence $e^{-\beta^{1/2}(\lambda/\mathsf{m})(\ell/2)} < e^{-4}$ and by (40), $\mathcal{R} \leq \ell/2 + \ell/8 < \ell$. In the last step, we used $c_A = c_0/2$ by (46), and inserted (81) from the following upper bounds on J

$$J \leq \frac{e^{-1} - e^{-5}}{40(\mathsf{m}-1)}\frac{2}{\ell}\max\left(L^{1/2}, \beta^{-1/2}\mathcal{R}^{-1}\right)\beta^{-1/2}\exp\left(-\beta^{1/2}\frac{\lambda}{\mathsf{m}}\frac{\ell}{2}\right)$$

$$\leq \frac{1 - e^{-4}}{40e(\mathsf{m}-1)}\frac{2}{\ell}\beta^{-1/2}\gamma\alpha\exp\left(-\beta^{1/2}\frac{\lambda}{\mathsf{m}}\frac{\ell}{2}\right) = \frac{1 - e^{-4}}{\mathsf{m}-1}\frac{\gamma\alpha}{\beta^{1/2}\frac{\lambda}{\mathsf{m}}\frac{\ell}{2}}\frac{c_0}{2} \tag{81}$$

where the first step follows from the condition (45) and the second step uses $\gamma\alpha = \beta^{-1/2}\mathcal{R}^{-1}\sqrt{1 + \beta L\mathcal{R}^2} \geq \max(L^{1/2}, \beta^{-1/2}\mathcal{R}^{-1})$ and the definition of c_0. □

5.6. *Proof of main contraction result for Andersen dynamics on* $\mathbb{T}_\ell^m$

In this part, we show that $M_t = e^{c_A t} \sum_i f(r_i(Y_t))$ is a nonnegative supermartingale. To this end, we develop a Dynkin-like inequality for $e^{c_A t} f(r_i(Y_t))$, and as an intermediate step, we prove an analogous result for the deterministic solution y_t of (20) starting at $y_0 = y$.

Lemma 5.9. *Let* $r_t^i := r_i(y_t)$, *fix* $c \in (0, \infty)$ *and suppose that* $g_i : \mathbb{T}_\ell^m \times \mathbb{R}^{3m} \to \mathbb{R}$ *satisfies* $f(r_t^i) - f(r_s^i) \le \int_s^t g_i(y_u)\, du$. *For all* $t > 0, 0 \le s \le t, y \in \mathbb{T}_\ell^m \times \mathbb{R}^{3m}$ *with* $y_0 = y$, *and* $i \in \{1, \ldots, m\}$,

$$e^{ct} f(r_t^i) - e^{cs} f(r_s^i) \le \int_s^t e^{cu}(g_i + cf \circ r_i)(y_u)\, du.$$

Formally, Lemma 5.9 follows by the chain rule, but since $f(r_t^i)$ is not differentiable, we give a direct proof.

Proof. Fix a sequence $(\Pi_n)_{n \in \mathbb{N}}$ of partitions of $[s, t]$ such that $\Pi_n \subseteq \Pi_{n+1}$ such that the mesh size $\|\Pi_n\| \to 0$ as $n \to \infty$. Let $v := \min\{r \in \Pi_n : r > u\}$ denote the next partition point after u. Let $A_t := e^{ct}$, $F_t := f(r_t^i)$, and $\sum_u := \sum_{\substack{u \in \Pi_n \\ u < t}}$. Then $A_t F_t - A_s F_s = \sum_u (A_v F_v - A_u F_u)$, and hence,

$$A_t F_t - A_s F_s = \mathrm{I} + \mathrm{II} + \mathrm{III} \quad \text{where} \quad \begin{cases} \mathrm{I} := \sum_u F_u(A_v - A_u), \\ \mathrm{II} := \sum_u A_u(F_v - F_u), \\ \mathrm{III} := \sum_u (A_v - A_u)(F_v - F_u). \end{cases} \tag{82}$$

As $n \to \infty$, we have:

$$\mathrm{I} = \int_{[s,t]} \sum_u F_u \mathbb{1}_{[u,v]}(r)\, dA_r \to \int_{[s,t]} F_{r-}\, dA_r = \int_{[s,t]} F_r \dot{A}_r\, dr,$$

by dominated convergence and continuity of $\dot{A}_r$;

$$\mathrm{II} \le \sum_u A_u \int_u^v g_i(y_r)\, dr \to \int_{[s,t]} A_r g_i(y_r)\, dr,$$

by continuity of A_r and dominated convergence; and,

$$\mathrm{III} \le ce^{ct} \sum_u (v - u) \int_u^v g_i(y_r)\, dr \to 0.$$

Hence, $A_t F_t - A_s F_s \le \liminf_{n \to \infty}(\mathrm{I} + \mathrm{II} + \mathrm{III}) \le \int_{[s,t]}(A_r g_i(y_r) + \dot{A}_r F_r)\, dr$, as required. $\qquad \square$

The next lemma applies Lemma 5.9 to obtain a Dynkin-like inequality for $e^{c_A t} f \circ r_i(Y_t)$ where Y_t is the coupling process.

Lemma 5.10. *Suppose that* $\lambda > 0$ *satisfies condition* (44). *Let* $R_t^i := r_i(Y_t)$. *There exists* $C_d > 0$ *such that for all* $t \in [0, 1]$, $y \in \mathbb{T}_\ell^m \times \mathbb{R}^{3m}$ *with* $Y_0 = y$, *and* $i \in \{1, \ldots, m\}$,

$$\mathbb{E}\left(e^{c_A t} f(R_t^i) - f(R_0^i)\right) \le e^{-\lambda t} \int_0^t e^{c_A s}\left(g_i + \mathcal{A}_\gamma^C (f \circ r_i) + c_A f \circ r_i\right)(y_s)\, ds + C_d t^2$$

where y_s *denotes the deterministic solution to* (20) *with the same initial condition* $y_0 = y$.

Proof. Recall from Definition 2.5, N_t represents the number of velocity randomizations that have occurred over $[0, t]$, and T_1 is the first jump time. Introduce the decomposition $e^{ct} f(R_t^i) - f(R_0^i) = \mathrm{I} + \mathrm{II} + \mathrm{III}$ where

$$\begin{cases} \mathrm{I} := (e^{ct} f(R_t^i) - f(R_t^0))\mathbb{1}_{\{N_t = 0\}}, \\ \mathrm{II} := e^{cT_1}(f(R_{T_1}^i) - f(R_{T_1-}^i))\mathbb{1}_{\{N_t \ge 1\}}, \quad \text{and} \\ \mathrm{III} := (e^{ct} f(R_t^i) - e^{cT_1} f(R_{T_1}^i) + e^{cT_1} f(R_{T_1-}^i) - f(R_t^0))\mathbb{1}_{\{N_t \ge 1\}}. \end{cases}$$

We now bound the expectations of I, II, and III.

On $N_t = 0$, we have $R_s^i = r_s^i$ for all $s \leq t$, where recall $r_s^i = r_i(y_s)$ denotes the corresponding distance function for the deterministic solution. Hence, for all $y \in \mathbb{T}_\ell^m \times \mathbb{R}^{3m}$, by Lemma 5.9,

$$\mathbb{E}(\mathrm{I}) = \mathbb{P}(N_t = 0)\left(e^{c_A t} f\left(r_t^i\right) - f\left(r_0^i\right)\right)$$

$$\leq e^{-\lambda t} \int_0^t \left(e^{c_A s} \mathrm{g}_i(y_s) + c_A e^{c_A s} f\left(r_s^i\right)\right) ds. \tag{83}$$

To bound II, note that the event $\{N_t \geq 1\}$ is equivalent to the event $\{T_1 \leq t\}$, and that $Y_s = y_s$ for $s < T_1$, and hence, $Y_{T_1-} = y_{T_1-} = y_{T_1}$. Thus, we can write

$$\mathrm{II} = e^{c_A T_1}\left(f \circ r_i\left(\mathcal{S}^C(I_1, \xi_1, \mathcal{U}_1)y_{T_1}\right) - f \circ r_i(y_{T_1})\right)\mathbb{1}_{\{T_1 \leq t\}}$$

$$= e^{c_A T_1}\left(f \circ r_i\left(\mathcal{S}^C(i, \xi_1, \mathcal{U}_1)y_{T_1}\right) - f \circ r_i(y_{T_1})\right)\mathbb{1}_{\{T_1 \leq t, I_1 = i\}}$$

Since $T_1, \xi_1, \mathcal{U}_1$ and I_1 are independent, the conditional expectation of II given T_1 is given by

$$\mathbb{E}(\mathrm{II} \mid T_1 = s) = \frac{1}{\mathsf{m}}\mathbb{E}\left(f \circ r_i\left(\mathcal{S}^C(i, \xi_1, \mathcal{U}_1)y_s\right) - f \circ r_i(y_s)\right)e^{c_A s}\mathbb{1}_{\{s \leq t\}}$$

$$= \frac{1}{\lambda}\mathcal{A}_\gamma^C(f \circ r_i)(y_s)e^{c_A s}\mathbb{1}_{\{s \leq t\}}, \quad \text{and thus,}$$

$$\mathbb{E}(\mathrm{II}) = \int_0^\infty \mathbb{E}(\mathrm{II} \mid T_1 = s)\lambda e^{-\lambda s}\, ds = \int_0^t e^{(c_A - \lambda)s}\mathcal{A}_\gamma^C(f \circ r_i)(y_s)\, ds. \tag{84}$$

Now we show that $\mathbb{E}(\mathrm{III})$ is of order $O(t^2)$ for small t. For this purpose, we introduce the decomposition $\mathrm{III} = \mathrm{III}_a + \mathrm{III}_b + \mathrm{III}_c$ where

$$\begin{cases} \mathrm{III}_a := (e^{c T_1} f(R_{T_1-}^i) - f(R_t^0))\mathbb{1}_{\{N_t \geq 1\}}, \\ \mathrm{III}_b := (e^{ct} f(R_t^i) - e^{c T_1} f(R_{T_1}^i))\mathbb{1}_{\{N_t = 1\}}, \quad \text{and} \\ \mathrm{III}_c := (e^{ct} f(R_t^i) - e^{c T_1} f(R_{T_1}^i))\mathbb{1}_{\{N_t \geq 2\}}. \end{cases}$$

To bound $\mathbb{E}(\mathrm{III}_a)$, note from (72) that g_i in (74) is globally bounded by a constant C_{g} and that $R_s^i = r_s^i$ for $s < T_1$. Thus, by Lemma 5.9, there exists a constant $C_a > 0$ such that

$$\mathbb{E}(\mathrm{III}_a) = \mathbb{E}\left(e^{c_A T_1} f\left(r_{T_1}^i\right) - f\left(r_0^i\right); T_1 \leq t\right)$$

$$\leq t\left(e^{c_A t} C_{\mathrm{g}} + c_A e^{c_A t} f(\mathcal{R})\right)\left(1 - e^{-\lambda t}\right) \leq C_a t^2 \quad \text{for all } t \leq 1.$$

Here we have used that by Lemma 5.9 and since $\mathrm{g}_i \leq C_{\mathrm{g}}$,

$$e^{c_A s} f\left(r_s^i\right) - f\left(r_0^i\right) \leq \int_0^s \left(e^{c_A u} C_{\mathrm{g}} + c_A e^{c_A u} f\left(r_u^i\right)\right) du,$$

$$\leq s\left(e^{c_A s} C_{\mathrm{g}} + c_A e^{c_A s} f(\mathcal{R})\right) \quad \text{for all } s \geq 0.$$

A similar bound holds for $\mathbb{E}(\mathrm{III}_b)$, since on $\{N_t = 1\}$, $R_s^i = r^i(\phi_{s-T_1}^C(Y_{T_1})) =: \tilde{r}_{s-T_1}^i$ for all $s \in [T_1, t]$ where $\tilde{r}_u^i$ is the distance for the deterministic solution $\tilde{y}_u$ with initial condition $\tilde{y}_0 = Y_{T_1}$. Hence, by Lemma 5.9, on $\{N_t = 1\}$,

$$e^{c_A t} r\left(R_t^i\right) - e^{c_A T_1} f(R_{T_1}^i = e^{c_A T_1}\left(e^{c_A(t-T_1)} f\left(\tilde{r}_{t-T_1}^i\right) - f(\tilde{r}_0)\right)$$

$$\leq t\left(e^{c_A, t} C_{\mathrm{g}} + c_A e^{c_A t} f(\mathcal{R})\right).$$

Thus, we obtain similarly as in $\mathbb{E}(\mathrm{III}_a)$,

$$\mathbb{E}(\mathrm{III}_b) = \mathbb{E}\left(e^{c_A t} f\left(r_{T_1}^i\right) - e^{c_A T_1} f\left(r_0^i\right); T_1 \leq t\right)$$

$$\leq t\left(e^{c_A t} C_{\mathrm{g}} + c_A e^{c_A t} f(\mathcal{R})\right)\left(1 - e^{-\lambda t}\right) \leq C_a t^2 \quad \text{for all } t \leq 1.$$

To bound $\mathbb{E}(\mathrm{III}_c)$, a rough bound suffices,

$$\mathbb{E}(\mathrm{III}_c) \leq \mathbb{P}(N_t \geq 2)e^{c_A t} f(R)$$

$$\leq \left(1 - e^{-\lambda t} - \lambda e^{-\lambda t}\right)e^{c_A t} f(R) \leq C_c t^2 \quad \text{for all } t \leq 1,$$

with a finite constant $C_c > 0$. In sum, we obtain

$$\mathbb{E}(\mathrm{III}) \leq (2C_a + C_c)t^2 \quad \text{for all } t \leq 1. \tag{85}$$

Combining (83), (84), and (85) we obtain for $t \leq 1$:

$$\mathbb{E}\left(e^{c_A t} f\left(R_t^i\right) - f\left(R_0^i\right)\right) \leq e^{-\lambda t} \int_0^t e^{c_A s}\left(g_i + \mathcal{A}_\gamma^C(f \circ r_i) + c_A f \circ r_i\right)(y_s)\, ds$$

$$+ \int_0^t e^{c_A s}\left(e^{-\lambda s} - e^{-\lambda t}\right)\mathcal{A}_\gamma^C(f \circ r_i)(y_s)\, ds + (2C_a + C_c)t^2$$

$$\leq e^{-\lambda t} \int_0^t e^{c_A s}\left(g_i + \mathcal{A}_\gamma^C(f \circ r_i) + c_A f \circ r_i\right)(y_s)\, ds + C_d t^2,$$

with a finite constant C_d depending only on the parameters λ, m, etc. In the last step, we used $|e^{-\lambda s} - e^{-\lambda t}| \leq \lambda(t - s)$ and the crude bound $\mathcal{A}_\gamma^C(f \circ r_i)(y_s) \leq \lambda f(\mathcal{R})$ to obtain $\int_0^t e^{c_A s}(e^{-\lambda s} - e^{-\lambda t})\mathcal{A}_\gamma^C(f \circ r_i)(y_s)\, ds \lesssim t^2$. $\qquad\square$

Now we combine Lemma 5.10 and Theorem 5.8 to prove that M_t is a nonnegative supermartingale for every initial condition $Y_0 = y \in \mathbb{T}_\ell^{\mathsf{m}} \times \mathbb{R}^{3\mathsf{m}}$.

Proof of Theorem 3.9. Let $\rho_t := \sum_{i=1}^{\mathsf{m}} f \circ r_i(Y_t)$ with $Y_0 = y$ and let $\mathbb{E}_y$ denote expectation conditional on $Y_0 = y$; as the notation indicates, the underlying measure space depends on the initial point. By Lemma 5.10 and Theorem 5.8, we obtain

$$\mathbb{E}_y(M_t - M_0) \leq \mathsf{m} C_d t^2 \quad \text{for all } t \in [0, 1] \text{ and } y \in \mathbb{T}_\ell^{\mathsf{m}} \times \mathbb{R}^{3\mathsf{m}}.$$

Fix $h \in [0, 1]$. Then

$$\mathbb{E}_y(M_t - M_0) \leq \mathsf{m} C_d h t \quad \text{for all } t \in [0, h] \text{ and } y \in \mathbb{T}_\ell^{\mathsf{m}} \times \mathbb{R}^{3\mathsf{m}}.$$

For $s \geq 0$, the Markov property implies

$$\mathbb{E}_y(M_t - M_s) = e^{c_A s}\mathbb{E}_y\left(e^{c_A(t-s)}\rho_t - \rho_s\right)$$

$$= e^{c_A s}\mathbb{E}_y\left(\mathbb{E}_{Y_s}\left(e^{c_A(t-s)}\rho_{t-s} - \rho_0\right)\right) \leq e^{c_A s}\mathsf{m} C_d h(t - s)$$

for all $t \in [s, s + h]$ and $y \in \mathbb{T}_\ell^{\mathsf{m}} \times \mathbb{R}^{3\mathsf{m}}$. Hence, with $t_k := kh$ for all $k \in \mathbb{N}_0$,

$$\mathbb{E}_y(M_t - M_0) = \sum_{k=1}^{\infty} \mathbb{E}_y(M_{t_k \wedge t} - M_{t_{k-1} \wedge t}) \leq e^{c_A t}\mathsf{m} C_d h t$$

for all $t \geq 0$ and $y \in \mathbb{T}_\ell^{\mathsf{m}} \times \mathbb{R}^{3\mathsf{m}}$. Letting $h \downarrow 0$ we obtain $\mathbb{E}_y(M_t - M_0) \leq 0$, and thus,

$$\mathbb{E}_y(M_t - M_s \mid \mathcal{F}_s) = e^{c_A s}\mathbb{E}_y\left(e^{c_A(t-s)}\rho_t - \rho_s \mid \mathcal{F}_s\right)$$

$$= e^{c_A s}\mathbb{E}_{Y_s}\left(e^{c_A(t-s)}\rho_{t-s} - \rho_0\right) = e^{c_A s}\mathbb{E}_{Y_s}(M_{t-s} - M_0) \leq 0.$$

Hence, $M_t = e^{c_A t}\rho_t$ is a nonnegative supermartingale.

Finally, we consider the process Y_t with initial distribution given by an optimal coupling of the initial distributions ν and η w.r.t. the distance $\mathcal{W}_\rho$, i.e., the law of $(x, v, \tilde{x}, \tilde{v})$ has marginals μ and ν and $Y_0 = (x, v, x - \tilde{x}, v - \tilde{v})$, and $\mathcal{W}_\rho(\mu, \nu) = \mathbb{E}[\rho(Y_0)]$. Then, for all $t \geq 0$, the law of Y_t represents a coupling of μp_t and νp_t, and hence by (39),

$$\mathcal{W}_\rho(\mu p_t, \nu p_t) \leq \mathbb{E}\left[\rho(Y_t)\right] \leq e^{-c_A t}\mathbb{E}\left[\rho(Y_0)\right] \leq e^{-c_A t}\mathcal{W}_\rho(\mu, \nu),$$

which proves (47), as required. $\qquad\square$

Appendix

This Appendix briefly reviews and slightly adapts [19, Theorem 5.5] to prove a supermartingale theorem for PDMPs needed in the proof of our main results. To state this result, let $(X_t)_{t \geq 0}$ be a PDMP with the following characteristics

  (i) boundaryless state space S;
 (ii) deterministic flow $\zeta_t : S \to S$ generated by a vector field $\mathfrak{X} : S \to \mathbb{R}^n$;
(iii) jump rates $J(x)$ where $J : S \to \mathbb{R}_{>0}$; and,
 (iv) jump measure $Q(x, dy)$.

On continuously differentiable functions f, define the generator of X_t as the operator $\mathcal{G}$ that outputs the function $\mathcal{G} f : S \to \mathbb{R}$ defined as

$$\mathcal{G} f(x) = (\mathfrak{X} \cdot \nabla f)(x) + J(x) \int_S \big(f(y) - f(x) \big) Q(x, dy).$$

Let $g : [0, \infty) \times S \to \mathbb{R}$ be a space-time-dependent function. For any $x \in S$, suppose that the function $t \mapsto g(t, \zeta_t(x))$ is absolutely continuous in time except at jump discontinuities where it is cádlág and nonincreasing. In this context, we prove that the process

$$g(t, X_t) - \int_0^t \left(\frac{\partial g}{\partial t} + \mathcal{G} g \right)(s, X_s) \, ds$$

is a local supermartingale.

First, we recall that X_t solves the following time-dependent martingale problem [19,20].

Lemma A.1. *For any $g : [0, \infty) \times S \to \mathbb{R}$ such that g is differentiable in its first variable, $\mathbb{E} \sum_{s \leq t} |g(s, X_s) - g(s-, X_{s-})| < \infty$ for each $t \geq 0$, and the function $t \mapsto g(t, \zeta_t(x))$ is absolutely continuous for all $x \in S$, the process*

$$g(t, X_t) - \int_0^t \left(\frac{\partial g}{\partial t} + \mathcal{G} g \right)(s, X_s) \, ds$$

is a local martingale.

Proof. This is a special case of Theorem 5.5 of Ref. [19] when the state space S is boundaryless. □

Next we apply this result to functions of the process X_t that are piecewise absolutely continuous functions (in time) and that have nonincreasing jumps along the deterministic part of X_t.

Lemma A.2. *For any $g : [0, \infty) \times S \to \mathbb{R}$ such that g is differentiable in its first variable; $\mathbb{E} \sum_{s \leq t} |g(s, X_s) - g(s-, X_{s-})| < \infty$ for each $t \geq 0$ and for all $X_0 \in S$; and the function $G(t) : t \mapsto g(t, \zeta_t(x))$ is piecewise absolutely continuous, cádlág, and $\Delta G(t) \leq 0$ for all $t > 0$ and for all $x \in S$; then the process*

$$g(t, X_t) - \int_0^t \left(\frac{\partial g}{\partial t} + \mathcal{G} g \right)(s, X_s) \, ds$$

is a local supermartingale.

Proof. The proof of this result is almost identical to the proof of Theorem 5.5 of [19] except that we must include the jumps in $g(t, X_t)$ along the deterministic parts of X_t. Let $\{t_i\}$ denote the jump times of the process X_t. Then we have the following representation

$$g(t, X_t) - g(0, X_0) = \sum_{s \leq t} \sum_{t_i \leq t} \Delta g \big(s - t_i, \zeta_{s - t_i}(X_{t_i}) \big) 1_{s \in [t_i, t_{i+1})}$$

$$+ \int_0^t \left(\frac{\partial g}{\partial t} + \mathcal{G} g \right)(s, X_s) \, ds + M_t^g$$

where M_t^g is a local martingale. Since the jumps in g along the deterministic parts of X_t are nonincreasing everywhere,

$$g(t, \mathsf{X}_t) - \int_0^t \left(\frac{\partial g}{\partial t} + \mathcal{G}g \right)(s, \mathsf{X}_s)\, ds \le g(0, \mathsf{X}_0) + M_t^g.$$

It follows that $g(t, \mathsf{X}_t) - \int_0^t (\frac{\partial g}{\partial t} + \mathcal{G}g)(s, \mathsf{X}_s)\, ds$ is a local supermartingale. $\square$

Theorem A.3. *Suppose that $g : [0, \infty) \times S \to \mathbb{R}$ is nonnegative, satisfies the conditions of Lemma A.2, and satisfies $\int_0^t (\frac{\partial g}{\partial t} + \mathcal{G}g)(s, \mathsf{X}_s)\, ds \le 0$ for all $t \ge 0$ and for all $\mathsf{X}_0 \in S$. Then $g(t, \mathsf{X}_t)$ is a supermartingale.*

Proof. Since the conditions for Lemma A.2 hold, the process

$$g(t, \mathsf{X}_t) - \int_0^t \left(\frac{\partial g}{\partial t} + \mathcal{G}g \right)(s, \mathsf{X}_s)\, ds$$

is a local supermartingale. Moreover, since $\int_0^t (\frac{\partial g}{\partial t} + \mathcal{G}g)(s, \mathsf{X}_s)\, ds \le 0$ by assumption, $g(t, \mathsf{X}_t)$ is also a local supermartingale. Since the function g is also nonnegative by assumption, Fatou's lemma implies that $g(t, \mathsf{X}_t)$ is a supermartingale, as required. $\square$

Acknowledgements

References

[1]   2019 Amber Reference Manual. https://ambermd.org/doc12/Amber19.pdf. Accessed: 2020-02-27.

[2]   M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids* 9. Clarendon, Oxford, 1987.

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

[4]   S. Armstrong and J.-C. Mourrat. *Variational Methods for the Kinetic Fokker–Planck Equation*, 2019. Available at arXiv:1902.04037 [math.AP]. MR4278431 https://doi.org/10.1007/s00028-021-00684-4

[5]   K. A. Ball et al. Conformational dynamics of the HIV-vif protein complex. *Biophys. J.* **116** (8) (2019) 1432–1445.

[6]   A. Beskos et al. Hybrid Monte-Carlo on Hilbert spaces. *Stochastic Process. Appl.* **121** (10) (2011) 2201–2230. MR2822774 https://doi.org/10.1016/j.spa.2011.06.003

[7]   J. Bierkens, P. Fearnhead and G. Roberts. The zig-zag process and super-efficient sampling for Bayesian analysis of big data. *Ann. Statist.* **47** (3) (2019) 1288–1320. MR3911113 https://doi.org/10.1214/18-AOS1715

[8]   J. Bierkens, G. O. Roberts and P.-A. Zitt. Ergodicity of the zigzag process. *Ann. Appl. Probab.* **29** (4) (2019) 2266–2301. MR3983339 https://doi.org/10.1214/18-AAP1453

[9]   S. Blanes, F. Casas and J. M. Sanz-Serna. Numerical integrators for the Hybrid Monte Carlo method. *SIAM J. Sci. Comput.* **36** (4) (2014) A1556–A1580. MR3233942 https://doi.org/10.1137/130932740

[10] P. G. Bolhuis. Transition path sampling on diffusive barriers. *J. Phys., Condens. Matter* **15** (1) (2002), S113.

[11] N. Bou-Rabee and A. Eberle. Two-scale coupling for preconditioned Hamiltonian Monte Carlo in infinite dimensions. *Stoch. Partial Differ. Equ. Anal. Comput.* **9** (1) (2021), 207242. MR4218791 https://doi.org/10.1007/s40072-020-00175-6

[12] N. Bou-Rabee, A. Eberle and R. Zimmer. Coupling and convergence for Hamiltonian Monte Carlo. *Ann. Appl. Probab.* **30** (3) (2020) 1209–1250. MR4133372 https://doi.org/10.1214/19-AAP1528

[13] N. Bou-Rabee and J. M. Sanz-Serna. Randomized Hamiltonian Monte Carlo. *Ann. Appl. Probab.* **27** (4) (2017) 2159–2194. MR3693523 https://doi.org/10.1214/16-AAP1255

[14] N. Bou-Rabee and K. Schuh Convergence of Unadjusted Hamiltonian Monte Carlo for Mean-Field Models. arXiv preprint, 2020. Available at arXiv:2009.08735.

[15] Y. Cao, J. Lu and L. Wang. On explicit L^2-convergence rate estimate for underdamped Langevin dynamics, 2020. Available at arXiv:1908.04746 [math.AP].

[16] Z. Chen and S. S. Vempala. Optimal convergence rate of Hamiltonian Monte Carlo for strongly logconcave distributions, 2019. Available at arXiv:1905.02313.

[17] X. Cheng et al. Underdamped Langevin MCMC: A non-asymptotic analysis. In *Conference on Learning Theory* 300–323, 2018.

[18] A. S. Dalalyan and L. Riou-Durand. On sampling from a log-concave density using kinetic Langevin diffusions. *Bernoulli* **26** (3) (2020), 19561988. MR4091098 https://doi.org/10.3150/19-BEJ1178

[19] M. H. A. Davis. Piecewise-deterministic Markov processes: A general class of non-diffusion stochastic models. *J. Roy. Statist. Soc. Ser. B* **46** (3) (1984) 353–388. MR0790622

[20] M. H. A. Davis. *Markov Models and Optimization*, **49**. CRC Press, Boca Raton, 1993. MR1283589 https://doi.org/10.1007/978-1-4899-4483-2

[21] G. Deligiannidis et al. Randomized Hamiltonian Monte Carlo as scaling limit of the bouncy particle sampler and dimension-free convergence rates. arXiv preprint, 2018. Available at arXiv:1808.04299.

[22] G. Deligiannidis, A. Bouchard-Cote and A. Doucet. Exponential ergodicity of the bouncy particle sampler. *Ann. Statist.* **47** (3) (2019) 1268–1287. MR3911112 https://doi.org/10.1214/18-AOS1714

[23] J. Dolbeault, C. Mouhot and C. Schmeiser. Hypocoercivity for linear kinetic equations conserving mass. *Trans. Amer. Math. Soc.* **367** (6) (2015) 3807–3828. MR3324910 https://doi.org/10.1090/S0002-9947-2015-06012-7

[24] A. Durmus et al. An elementary approach to uniform in time propagation of chaos. *Proc. Amer. Math. Soc.* **148** (12) (2020) 5387–5398. MR4163850 https://doi.org/10.1090/proc/14612

[25] A. Durmus, E. Moulines and E. Saksman. On the convergence of Hamiltonian Monte Carlo, 2017. Available at arXiv:1705.00166 [stat.CO].

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

[27] A. Eberle. Reflection couplings and contraction rates for diffusions. *Probab. Theory Related Fields* **166** (3–4) (2016) 851–886. MR3568041 https://doi.org/10.1007/s00440-015-0673-1

[28] A. Eberle Bonn University Lecture Notes: Stochastic Analysis, 2019. Available at https://wt.iam.uni-bonn.de/eberle/skripten/. Last. visited on 2020/05/12,.

[29] A. Eberle, A. Guillin and R. Zimmer. Couplings and quantitative contraction rates for Langevin dynamics. *Ann. Probab.* **47** (4) (2019) 1982–2010. MR3980913 https://doi.org/10.1214/18-AOP1299

[30] D. Frenkel and B. Smit. *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd edition. 9. Academic Press, San Diego, 2002.

[31] GROMACS Documentation Release 2019. http://manual.gromacs.org/documentation/2019/manual-2019.pdf. Accessed: 2020-02-27 (cit. on pp. 2, 9).

[32] J. Heng and P. E. Jacob. Unbiased Hamiltonian Monte Carlo with couplings. *Biometrika* **106** (2) (2019) 287–302. MR3949304 https://doi.org/10.1093/biomet/asy074

[33] F. Herau. Hypocoercivity and exponential time decay for the linear in- homogeneous relaxation Boltzmann equation. *Asymptot. Anal.* **46** (3) (2006) 349–359. MR2215889

[34] P. Holderrieth Cores for Piecewise-Deterministic Markov Processes used in Markov Chain Monte Carlo. Preprint, 2019. Available at arXiv:1910.11429. MR1220886 https://doi.org/10.1080/07362999308809317

[35] M. D. Homan and A. Gelman. The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* **15** (1) (2014) 1593–1623. MR3214779

[36] T. Lelievre, M. Rousset and G. Stoltz. *Free Energy Computations: A Mathematical Perspective*, 1st edition. Imperial College Press, London, 2010. MR2681239 https://doi.org/10.1142/9781848162488

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

[38] J. S. Liu. *Monte Carlo Strategies in Scientific Computing*, 2nd. edition. Springer, Berlin, 2008. MR2401592

[39] S. Livingstone et al. On the geometric ergodicity of Hamiltonian Monte Carlo. *Bernoulli* **25** (4A) (2019) 3109–3138. MR4003576 https://doi.org/10.3150/18-BEJ1083

[40] J. Lu and L. Wang. On explicit L^2-convergence rate estimate for piecewise deterministic Markov processes in MCMC algorithms, 2021. Available at arXiv:2007.14927 [math.PR]. MR1220886 https://doi.org/10.1080/07362999308809317

[41] P. B. Mackenzie. An improved hybrid Monte Carlo method. *Phys. Lett. B* **226** (3) (1989) 369–371.

[42] O. Mangoubi and A. Smith Rapid mixing of Hamiltonian Monte Carlo on strongly log-concave distributions. Preprint, 2017. Available at arXiv:1708.07114.

[43] J. C. Mattingly, A. M. Stuart and D. J. Higham. Ergodicity for SDEs and approximations: Locally Lipschitz vector fields and degenerate noise. *Stochastic Process. Appl.* **101** (2) (2002) 185–232. MR1931266 https://doi.org/10.1016/S0304-4149(02)00150-3

[44] R. M. Neal. Bayesian Learning for Neural Networks. PhD thesis, University of Toronto, 1995.

[45] R. M. Neal. MCMC using Hamiltonian dynamics. In *Handbook of Markov Chain Monte Carlo* 113–162, 2011. MR2858447

[46] I. S. Novikov, A. V. Shapeev and Y. V. Suleimanov. Ring polymer molecular dynamics and active learning of moment tensor potential for gas-phase barrierless reactions: Application to S + H2. *J. Chem. Phys.* **151** (22) (2019), 224105.

[47] K. Oelschlager. A martingale approach to the law of large numbers for weakly interacting stochastic processes. *Ann. Probab.* **12** (2) (1984) 458–479. MR0735849

[48] G. A. Pavliotis. *Stochastic Processes and Applications*. *Texts in Applied Mathematics* **60**. Springer, Berlin, 2014. MR3288096 https://doi.org/10.1007/978-1-4939-1323-7

[49] L. R. Petzold, L. O. Jay and J. Yen. Numerical solution of highly oscillatory ordinary differential equations. *Acta Numer.* **6** (1997) 437–483. ISSN: 0962-4929.

[50] D. Piponi, M. Hoffman and P. Sountsov. Hamiltonian Monte Carlo Swindles. In *International Conference on Artificial Intelligence and Statistics* 3774–3783, PMLR, 2020.

[51] J. M. Sanz-Serna and A. M. Stuart. Ergodicity of dissipative differential equations subject to random impulses. *J. Differential Equations* **155** (2) (1999) 262–284. MR1698555 https://doi.org/10.1006/jdeq.1998.3594

[52] A.-S. Sznitman. Topics in propagation of chaos. In *Ecole d'ete de probabilites de Saint-Flour XIX – 1989* 165–251. Springer, Berlin, 1991. MR1108185 https://doi.org/10.1007/BFb0085169

[53] D. Talay. Stochastic Hamiltonian systems: Exponential convergence to the invariant measure, and discretization by the implicit Euler scheme. *Markov Process. Related Fields* **8** (2002) 1–36. MR1924934

[54] J. Tugaut et al. Convergence to the equilibria for self-stabilizing processes in double-well landscape. *Ann. Probab.* **41** (3A) (2013) 1427–1460. MR3098681 https://doi.org/10.1214/12-AOP749

[55] B. P. Uberuaga, M. Anghel and A. F. Voter. Synchronization of trajectories in canonical molecular-dynamics simulations: Observation, explanation, and exploitation. *J. Chem. Phys.* **120** (14) (2004) 6363–6374.

[56] C. L. Vaillant, D. J. Wales and S. C. Althorpe. Tunneling splittings from path-integral molecular dynamics using a Langevin thermostat. *J. Chem. Phys.* **148** (23) (2018), 234102.

[57] C. Villani. Hypocoercivity. *Mem. Amer. Math. Soc.* **202** (950) (2009), 0065. MR2562709 https://doi.org/10.1090/S0065-9266-09-00567-5