Convergence of McKean-Vlasov processes and Markov Chain Monte Carlo methods for mean-field models

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von
Katharina Johanna Schuh
aus
Bad Kreuznach

Bonn, June 2022

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät
der Rheinischen Friedrich-Wilhelms-Universität Bonn

Gutachter: Prof. Dr. Andreas Eberle
Gutachterin: Prof. Dr. Franca Hoffmann

Tag der Promotion: 28.09.2022
Erscheinungsjahr: 2022

Abstract

In recent years, analysing the long-time behaviour of stochastic processes has received increasing interest. Firstly, efficient sampling of a given probability measure is an important task that arises in various fields such as Bayesian statistics or computational physics. Markov Chain Monte Carlo (MCMC) algorithms form a powerful class of sampling methods for which guarantees for fast mixing are of particular interest, especially for intractable target measures. Secondly, one would like to better understand the convergence behaviour of stochastic processes which have their origin in modelling phenomena in physics and are used in deep learning, among others.

In this thesis, we focus on specific high-dimensional problems. We are interested in sampling target measures of mean-field particle type consisting of a unary potential that is in general not strongly convex and of a pairwise interaction potential. Correspondingly, we consider a system of many particles moving according to an external confining force and a pairwise interaction force. Further, we address the connection between processes of mean-field particle type and their corresponding McKean-Vlasov process, where only one particle is considered and whose moves are determined by a nonlinear stochastic differential equation (SDE) with an external force and a distribution-dependent interaction force. We are interested in quantitative estimates between the laws of these two types of processes.

The thesis covers three projects. In the first part, we analyse the behaviour of the *unadjusted Hamiltonian Monte Carlo* (uHMC) algorithm which forms an MCMC method that samples approximately a given target measure. For a target measure of mean-field type, contraction in Wasserstein distance with dimension-free rates is established under certain conditions on the unary part and the interaction part of the mean-field potential. Furthermore, error estimates between the target measure and the measure sampled by uHMC are provided.

In the second part, we investigate nonlinear stochastic differential equations without confinement and their corresponding mean-field particle systems. To show contraction in Wasserstein distance, the so-called *sticky coupling* is established for nonlinear SDEs and a novel class of nonlinear one-dimensional SDEs with a sticky boundary behaviour at zero is introduced. For these equations, existence and uniqueness of a weak solution are proven and a phase transition from a unique to several invariant probability measures is analysed. Provided a unique invariant probability measure exists and contraction towards this measure holds, we deduce contraction in Wasserstein distance for the nonlinear SDE without confinement. Further, we establish uniform in time propagation of chaos estimates for the corresponding particle system.

In the final part, we study the long-time behaviour of diffusions given by the second-order Langevin dynamics with distribution-dependent forces. Global contraction in Wasserstein distance with dimension-free rates is shown via a coupling approach and a carefully constructed distance function. In addition, we analyse the optimal order of the contraction rates for the classical second-order Langevin dynamics with a strongly convex potential. Finally, we provide uniform in time propagation of chaos bounds for the corresponding mean-field particle system.

Acknowledgement

I would like to take the opportunity to thank several persons and institutions that supported me during my PhD and the development of the PhD thesis.

First of all, I want to thank my supervisor Andreas Eberle for his guidance and extensive support during the last years. He introduced me to probability theory and in particular to MCMC methods and drew my interest in this topic. I enjoyed the numerous instructive and helpful discussions from which I learned a lot. I am thankful for his advice, his patience and the effort he has put into me.

Many thanks go to my other co-authors: I want to thank Nawaf Bou-Rabee for his support during his long stay in Bonn. Moreover, I am grateful to Alain Durmus and Arnaud Guillin for their hospitality and support during my stays in Paris and Clermont-Ferrand.

Furthermore, I would like to thank Franca Hoffmann for willing to be the second referee of my thesis, as well as Carsten Burstedde and Carsten Urbach for being part of the committee.

Moreover, I want to thank all current and previous members of the probability theory group, including the secretary Mei-Ling Wang. I am very grateful to share an office with Saeda Marello and Francis Lörler. I enjoyed working in this pleasant environment. Special thanks go to Francis Lörler and Stefan Oberdörster for proofreading the thesis.

My PhD is financially supported by the DFG through the Bonner International Graduate School (BIGS) and by the Institute for Applied Mathematics. I am thankful for this support.

Finally, I thank my family and Konstantin for their extensive support.

Contents

List of Figures

1

Introduction

Obtaining quantitative estimates for the long-time behaviour of stochastic processes is a relevant issue occurring in many applications. Firstly, we are interested in getting guarantees for fast mixing of Markov Chain Monte Carlo (MCMC) methods which form a powerful class of sampling algorithms. In particular, efficient sampling of given intractable probability distribution is of great interest. Secondly, we want to better understand the convergence behaviour of stochastic processes that describe phenomena, for instance, in physics and are determined by stochastic differential equations.

In this PhD thesis, we study the Hamiltonian Monte Carlo (HMC) algorithm for mean-field models and two specific types of stochastic differential equations (SDEs) of McKean-Vlasov type and investigate their long-time behaviour using coupling methods as an analytic tool. The thesis covers three projects.

In the first project, we consider unadjusted HMC (uHMC) for mean-field models using the velocity Verlet discretisation of the Hamiltonian dynamics given by

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t} q_t^i = p^i_{\lfloor t \rfloor_h} - \frac{h}{2} \nabla_i U(q_{\lfloor t \rfloor_h}) \\ \frac{\mathrm{d}}{\mathrm{d}t} p_t^i = -\frac{1}{2}(\nabla_i U(q_{\lfloor t \rfloor_h}) + \nabla_i U(q_{\lceil t \rceil_h})), \end{cases} \qquad i = 1, \ldots, n, \tag{1.1}$$

where the mean-field potential $U : \mathbb{R}^{nd} \to \mathbb{R}$ is given by $U(x) = \sum_{i=1}^n (V(x^i) + \frac{\epsilon}{n} \sum_{j=1}^n W(x^i - x^j))$ and $\lfloor t \rfloor_h$ and $\lceil t \rceil_h$ denote the floor and ceiling function, respectively, with respect to the discretisation parameter $h > 0$. Using a particlewise coupling and a complementary particlewise metric, we establish dimension-free contraction bounds in L^1 Wasserstein distance. These bounds hold for unary potentials V including non-strongly convex functions provided the interaction parameter ϵ, the discretisation parameter h and the duration time of each uHMC step are sufficiently small. Moreover, we establish strong accuracy bounds for uHMC applied to mean-field models and derive quantitative error bounds between the target measure and the measure sampled using uHMC.

In the second project, we study nonlinear unconfined SDEs of McKean-Vlasov type on $\mathbb{R}^d$ given by

$$\mathrm{d}X_t = \Big(\int_{\mathbb{R}^d} b(X_t - x) \mu_t(\mathrm{d}x) \Big) \mathrm{d}t + \mathrm{d}B_t, \qquad \mu_t = \mathrm{Law}(X_t), \tag{1.2}$$

where $(B_t)_{t \geq 0}$ is a d-dimensional Brownian motion and the force $b : \mathbb{R}^d \to \mathbb{R}^d$ consists of a linear function and a bounded, Lipschitz continuous perturbation. We introduce a sticky coupling for nonlinear SDEs of McKean-Vlasov type and establish conditions under which contraction in Wasserstein distance holds. For this, we show that the distance process of the two copies of the coupling is controlled by the solution to a one-dimensional nonlinear equation with a sticky boundary at 0 given by

$$\mathrm{d}r_t = \Big(\tilde{b}(r_t) + \int_{\mathbb{R}^d} g(y)P_t(\mathrm{d}y)\Big)\mathrm{d}t + 2\mathbb{1}_{(0,\infty)}(r_t)\mathrm{d}W_t, \qquad P_t = \mathrm{Law}(r_t), \tag{1.3}$$

where $(W_t)_{t\geq 0}$ is a one-dimensional Brownian motion, $\tilde{b}$ is a Lipschitz continuous function and g is a bounded measurable function. For this novel class of SDEs, we prove existence and uniqueness in law of a weak solution. Further, we exhibit a phase transition for its invariant probability measures. In the case of a unique invariant probability measure, we establish conditions under which convergence to the invariant measure holds. Eventually, we show uniform in time propagation of chaos for the mean-field particle system corresponding to (1.2).

In the third project, we consider the Langevin dynamics with nonlinear interactions of McKean-Vlasov type given by

$$\begin{cases} \mathrm{d}X_t = Y_t\mathrm{d}t \\ \mathrm{d}Y_t = (-\gamma Y_t + ub(X_t) + u\int_{\mathbb{R}^d} \tilde{b}(X_t,x)\mu_t^x(\mathrm{d}x))\mathrm{d}t + \sqrt{2\gamma u}\mathrm{d}B_t, \qquad \mu_t^x = \mathrm{Law}(X_t), \end{cases} \tag{1.4}$$

where $(B_t)_{t\geq 0}$ is a d-dimensional Brownian motion, and where the confinement force $b:\mathbb{R}^d\to\mathbb{R}^d$ and the interaction force $\tilde{b}:\mathbb{R}^{2d}\to\mathbb{R}^d$ are Lipschitz continuous functions. We prove global contraction in Wasserstein distance under certain assumptions on b, $\tilde{b}$, the friction coefficient $\gamma > 0$ and the inverse mass $u > 0$ via a specifically designed distance function and a carefully aligned coupling approach. This distance is equivalent to the Euclidean distance and combines optimally two contraction results for large and small distances. In addition, we provide uniform in time propagation of chaos bounds for the corresponding particle systems.

Before we step into the details of the three projects, we introduce some notations and recall basic definitions and known facts. First, we define a distance between two probability measures and the notion of contraction in Wasserstein distance. In Section 1.2, we introduce the basic idea of Markov Chain Monte Carlo methods and define HMC. In Section 1.3, we present Langevin diffusions, before we recall some coupling techniques in Section 1.4 and present known contraction results for MCMC methods and SDEs in Section 1.5. Finally, we introduce nonlinear SDEs and establish the concept of propagation of chaos in Section 1.6. The references are directly provided in the respective sections.

1.1 Wasserstein distance and contraction

1.1.1 Wasserstein distance

Let $(\mathbb{X}, d)$ be some Polish space endowed with the Borel σ-algebra $\mathcal{B}(\mathbb{X})$. Often, we consider $\mathbb{X} = \mathbb{R}^d$ and the Euclidean distance $d(x,y) = |x-y|$ for all $x, y\in\mathbb{R}^d$. The set of all probability measures on $\mathbb{X}$ is denoted by $\mathcal{P}(\mathbb{X})$. To define the distance between two probability distributions $\mu,\nu\in\mathcal{P}(\mathbb{X})$, we first introduce the notion of a coupling between two probability measures. We call $\gamma\in\mathcal{P}(\mathbb{X}\times\mathbb{X})$ a *coupling* of the measures μ and ν if

$$\gamma(A\times\mathbb{X}) = \mu(A) \quad \text{and} \quad \gamma(\mathbb{X}\times B) = \nu(B) \qquad \text{for any } A, B\in\mathcal{B}(\mathbb{X}).$$

The set of all couplings of μ and ν is denoted by $\Pi(\mu,\nu)$. We say that the coupling is *realised* by random variables $X, Y : \Omega \to \mathbb{X}$ defined on a common probability space $(\Omega, \mathcal{A}, P)$ such that $(X,Y) \sim \gamma$.

Let $\rho : \mathbb{X} \times \mathbb{X} \to [0,\infty)$ be a metric on $\mathbb{X}$ that can differ from d. Fix $p \in [0,\infty)$. We define the L^p Wasserstein distance with respect to ρ on the set

$$\mathcal{P}^p_\rho(\mathbb{X}) = \Big\{\mu \in \mathcal{P}(\mathbb{X}) : \int_{\mathbb{R}^d} \rho(x,y)^p \mu(\mathrm{d}x) < \infty \text{ for some } y \in \mathbb{X}\Big\} \tag{1.5}$$

by

$$\mathcal{W}^p_\rho(\mu,\nu) = \inf_{\gamma\in\Pi(\mu,\nu)} \Big(\int_{\mathbb{X}\times\mathbb{X}} \rho(x,y)^p \gamma(\mathrm{d}x\mathrm{d}y)\Big)^{1/p} = \inf_{X\sim\mu, Y\sim\nu} E[\rho(X,Y)^p]^{1/p}. \tag{1.6}$$

In the case $\rho = d$, we write $\mathcal{W}^p$ and $\mathcal{P}^p$. It holds that $(\mathcal{P}^p(\mathbb{X}), \mathcal{W}^p)$ defines a Polish space, and if a sequence of measures $(\mu_n)_{n\in\mathbb{N}}$ in $\mathcal{P}^p(\mathbb{X})$ converges to a measure $\mu \in \mathcal{P}^p(\mathbb{X})$, then $\mu_n \to \mu$ weakly, see [192, Theorem 6.9]. Note that if ρ and d are equivalent, then $\mathcal{P}^p = \mathcal{P}^p_\rho$. For $p = 1$, $\mathcal{W}^1_\rho$ is called *Kantorovich distance*, and the exponent is often omitted for simplicity. We remark that the Kantorovich distance is often defined for a more general function ρ which only forms a semimetric, i.e. ρ satisfies $\rho(x,y) = \rho(y,x)$ for all $x, y \in \mathbb{X}$ and $\rho(x,y) = 0$ if and only if $x = y$. In that case, $\mathcal{W}_\rho$ defines a semimetric on the space $\mathcal{P}_\rho(\mathbb{X})$.

The Wasserstein distance with respect to d can easily be modified by considering a function $f : [0,\infty) \to [0,\infty)$ that is non-decreasing, concave and satisfies $f(0) = 0$ and $f'(0) > 0$. Then, $f \circ d$ defines again a metric and the corresponding Kantorovich distance is given by $\mathcal{W}_f(\mu,\nu) = \inf_{\gamma\in\Pi(\mu,\nu)} \int_{\mathbb{X}\times\mathbb{X}} (f\circ d)(x,y)\gamma(\mathrm{d}x\mathrm{d}y)$.

The *total variation distance* (TV distance) forms a prominent example for the Kantorovich distance, where the underlying distance is given by $\rho(x,y) = \mathbb{1}_{\{x\neq y\}}$. We denote the TV distance of two probability measure $\mu, \nu \in \mathcal{P}(\mathbb{X})$ by

$$\|\mu - \nu\|_{\mathrm{TV}} = \mathcal{W}_\rho(\mu,\nu) = \inf_{X\sim\mu, Y\sim\nu} P[X \neq Y].$$

We refer to [131, Chapter 4.2] for a detailed study on the TV distance.

We note that (1.6) represents a special case of the optimal transport problem where a general cost function $c : \mathbb{X}\times\mathbb{X} \to [0,\infty)$ is considered instead of the metric ρ. In the Monge formulation, a map $T : \mathbb{X} \to \mathbb{X}$ is searched that minimises $\int c(x, T(x))\mu(\mathrm{d}x)$ under the constraint $\nu = \mu \circ T^{-1}$, whereas in the less restrictive Kantorovich formulation, a coupling $\gamma \in \Pi(\mu,\nu)$ is searched that minimises $\int c(x,y)\gamma(\mathrm{d}x\mathrm{d}y)$.

The presented definitions and statements on the Wasserstein distance are taken from [192, Chapter 6].

1.1.2 Contraction in Wasserstein distance

Using the previously introduced distances for probability measures we are interested in the long-time behaviour of a given process. Next, we define the *Markov transition function* and introduce the concept of *contraction in Wasserstein distance* for a given transition function.

Let $I = \mathbb{N}$ or $I = \mathbb{R}_+$ be an index set. We denote by $(p_t)_{t\in I}$ a time-homogeneous transition function on $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$, which is a collection of probability kernels $p_t : \mathbb{X}\times\mathcal{B}(\mathbb{X}) \to [0,1]$ satisfying $p_0(x,\cdot) = \delta_x$ and $p_s p_t = p_{s+t}$ for all $s, t \in I$, where $(p_s p_t)(x, A) = \int_{\mathbb{X}} p_s(x, \mathrm{d}y) p_t(y, A)$ for all $x \in \mathbb{X}$ and $A \in \mathcal{B}(\mathbb{X})$. We write $\mu p_t(\mathrm{d}x) = \int_{\mathbb{X}} p_t(y, \mathrm{d}x)\mu(\mathrm{d}y)$ for all probability measures

μ and $(p_t f)(x) = \int_{\mathbb{X}} p_t(x, \mathrm{d}y) f(y)$ for all functions $f : \mathbb{X} \to \mathbb{R}$. Given a transition function $(p_t)_{t\in I}$ and a filtration $(\mathcal{F}_t)_{t\in I}$ on a probability space $(\Omega, \mathcal{A}, P)$, a stochastic process $(X_t)_{t\in I}$ on $(\Omega, \mathcal{A}, P)$ with values in $\mathbb{X}$ is called an $(\mathcal{F}_t)$-*Markov process with transition function* $(p_t)_{t\in I}$ if and only if $(X_t)_{t\in I}$ is $\mathcal{F}_t$-adapted and it holds that $P[X_t \in A|\mathcal{F}_s] = p_{t-s}(X_s, A)$ P-almost surely for all $s, t \in I$ with $s \leq t$ and for all $A \in \mathcal{B}(\mathbb{X})$. Given a transition function $(p_t)_{t\in I}$ and a probability measure $\nu \in \mathcal{P}(\mathbb{X})$, there exists a unique probability measure P_ν on the product space $(\Omega_{can}, \mathcal{A}_{can})$ such that $(X_t)_{t\in I}$, $X_t(\omega) = \omega(t)$ is a Markov process on $(\Omega_{can}, \mathcal{A}_{can}, P_\nu)$ with transition function $(p_t)_{t\in I}$ and $P_\nu \circ X_0^{-1} = \nu$, where $\Omega_{can} = \mathbb{X}^I = \{\omega : I \to \mathbb{X}\}$ is the product space and $\mathcal{A}_{can} = \sigma(X_t : t \in I)$ is the product σ-algebra on Ω_{can}. This result holds as a consequence of Kolmogorov's extension theorem [123]. Moreover, we assume that the transition function $(p_t)_{t\in I}$ is Feller, which means that for all functions $f \in \mathcal{C}_b(\mathbb{X})$ it holds that $p_t f \in \mathcal{C}_b(\mathbb{X})$, where $\mathcal{C}_b(\mathbb{X})$ denotes the set of continuous and bounded functions $f : \mathbb{X} \to \mathbb{R}$. The Feller property implies existence of a strong Markov process $(X_t)_{t\in I}$ with càdlàg paths (right continuous paths with left limits), cf. [123, Theorem 21.27].

For $1 \leq p < \infty$, we say *contraction in* L^p *Wasserstein distance with respect to the metric* $\rho : \mathbb{X} \times \mathbb{X} \to [0, \infty)$ holds if there exists a constant $c > 0$ such that

$$\mathcal{W}_\rho^p(\nu p_t, \eta p_t) \leq e^{-ct} \mathcal{W}_\rho^p(\nu, \eta) \qquad \text{for all probability measures } \nu, \eta \in \mathcal{P}_\rho^p(\mathbb{X}) \text{ and } t \in I. \tag{1.7}$$

The constant c is called *contraction rate*. Inequalities of this form, which were first studied by Dobrushin in [66] and which are also known as *Dobrushin uniqueness condition* in statistical mechanics, give results on the long-time behaviour of the Markov process corresponding to $(p_t)_{t\in I}$. Motivated by the concept of Ricci curvature bounds on Riemannian manifolds (cf. [9, 193]), the contraction rate is alternatively called *Ricci–Wasserstein curvature* or *Wasserstein curvature* with respect to ρ [117, 162, 180].

The issue of showing contraction in Wasserstein distance (1.7) is addressed in Section 1.4. Next, we present several consequences of (1.7). In the following, we assume that ρ is equivalent to d, i.e. there exist $C_1, C_2 > 0$ such that $C_1\rho(x, y) \leq d(x, y) \leq C_2\rho(x, y)$. As a direct consequence of contraction in L^p Wasserstein distance it holds:

Theorem 1.1 (Existence of a unique invariant measure and geometric ergodicity)**.** *There exists a unique invariant measure μ of $(p_t)_{t\in I}$ in $\mathcal{P}^p(\mathbb{X})$ and for every initial distribution $\nu \in \mathcal{P}^p(\mathbb{X})$, νp_t converges to μ, i.e.*

$$\mathcal{W}_\rho^p(\nu p_t, \mu) \leq e^{-ct} \mathcal{W}_\rho^p(\nu, \mu), \qquad \text{and} \qquad \mathcal{W}^p(\nu p_t, \mu) \leq M e^{-ct} \mathcal{W}^p(\nu, \mu),$$

where $M = C_2/C_1$.

Proof. Since $(\mathcal{P}^p(\mathbb{X}), \mathcal{W}^p)$ defines a Polish space the result holds by Banach fixed point theorem, [198, Chapter IV. 7]. □

To present another consequence, let us assume for a moment that $I = \mathbb{N}$ and let $(\mathbf{X}_n)_{n\in\mathbb{N}}$ be a Markov chain with transition function $(p_n)_{n\in\mathbb{N}}$, which admits a unique invariant probability measure μ. As we see in more detail in the next section on MCMC methods, one is often interested in approximating quantities of the form $\int f \mathrm{d}\mu$ for some observable $f : \mathbb{X} \to \mathbb{R}$ and some target measure μ on $\mathbb{X}$. Then, conversely, the Markov chain $(\mathbf{X}_n)_{n\in\mathbb{N}}$ is constructed such that μ is its unique invariant probability measure and the ergodic averages of the Markov chain converge to the desired quantity, i.e,

$$n^{-1} \sum_{i=1}^n f(\mathbf{X}_i) \to \int_{\mathbb{X}} f \mathrm{d}\mu \qquad \text{for } n \to \infty.$$

If (1.7) holds, we deduce quantitative bounds on the bias of the ergodic averages. This analysis is based on work by Joulin and Ollivier [118].

Corollary 1.2. *Let $g : \mathbb{X} \to \mathbb{R}$ be a Lipschitz continuous function with respect to ρ with Lipschitz constant $\|g\|_{\mathrm{Lip}(\rho)}$ given by*

$$\|g\|_{\mathrm{Lip}(\rho)} = \sup\{|g(x) - g(y)|/\rho(x, y) : \; x, y \in \mathbb{X}\}.$$

Then for any $n \in \mathbb{N}$ and $x \in \mathbb{X}$,

$$\Big|E_x\Big[n^{-1}\sum_{i=1}^{n} g(\mathbf{X}_i)\Big] - \int_{\mathbb{X}} g\mathrm{d}\mu\Big| \leq c^{-1}\|g\|_{\mathrm{Lip}(\rho)} \int_{\mathbb{X}} \rho(x, y)\mu(\mathrm{d}y),$$
$$\mathrm{Var}_x\Big[n^{-1}\sum_{i=1}^{n} g(\mathbf{X}_i)\Big] \leq \frac{1}{2(1-e^{-c})n}\|g\|^2_{\mathrm{Lip}(\rho)} \int_{\mathbb{X}}\int_{\mathbb{X}} \rho(y, z)^2 p_n(x, \mathrm{d}y)p_n(x, \mathrm{d}z),$$

where E_x and Var_x denote the expectation and the variance given the Markov chain $(\mathbf{X}_n)_{n\in\mathbb{N}}$ started in x.

A proof is given in [118].

We note that more consequences result from (1.7) such as bounds on the $L^2(\mu)$ spectral gap, see [48, 105], and concentration inequalities, see e.g. [118]. Furthermore, the results can be transferred to similar statements if ρ only constitutes a semimetric, i.e. the triangle inequality is not satisfied, see [201, Section 0.2].

1.1.3 Exponential decay in f-divergence and mixing time

Besides contraction in Wasserstein distance, there are further possibilities to control the long-time behaviour of stochastic processes. For this reason, we introduce another quantity to measure the difference between two probability measures $\mu, \nu \in \mathcal{P}(\mathbb{X})$, which was first studied by Rényi [171] and further developed by Csiszár [58] and Morimoto [154]. Assume that ν is absolutely continuous with respect to μ, $\nu \ll \mu$, and denote by $\rho = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}$ the Radon-Nikodym density. Let $f : (0, \infty) \to \mathbb{R}$ be a convex function with $f(1) = 0$ and we extend f to 0 by $f(0) = \lim_{t\downarrow 0} f(t)$, which is well-defined by convexity of f but can be infinite. The *f-divergence* of μ with respect to ν is given by

$$D_f(\nu|\mu) = \int_{\mathbb{X}} f(\rho(x))\mathrm{d}\mu(x). \tag{1.8}$$

For $f(t) = (t-1)^2$, $D_f(\nu|\mu) = \chi^2(\nu|\mu)$ denotes the χ^2-divergence, whereas for $f(t) = t\log(t)$, $D_f(\nu|\mu) = H(\nu|\mu)$ denotes the relative entropy or Kullback-Leibler divergence (KL divergence). For $f(t) = |t-1|/2$, we recover the total variation distance, $\|\mu - \nu\|_{\mathrm{TV}}$. The definition and the notation of the f-divergence are based on [84].

The f-divergence can be used to analyse the long-time behaviour of stochastic processes. As in the previous subsection, let $I = \mathbb{R}_+$ or $I = \mathbb{N}$ and let $(p_t)_{t\in I}$ denote a transition function of a time-homogeneous Markov process on $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ with invariant probability measure μ. For any probability measures $\nu \in \mathcal{P}(\mathbb{X})$, $t \to D_f(\nu p_t|\mu)$ is a non-increasing function [171]. For the relative entropy this result is known as H-theorem in statistical physics.

We remark that the χ^2-divergence and the relative entropy control the TV distance, i.e. $\|\mu - \nu\|_{\mathrm{TV}} \leq (1/2)\chi^2(\nu|\mu)^{1/2}$ by Jensen's inequality and $\|\mu - \nu\|_{\mathrm{TV}} \leq (H(\nu|\mu)/2)^{1/2}$ for any probability measure $\mu, \nu \in \mathcal{P}(\mathbb{X})$. The latter bound is known as Pinsker's inequality, see e.g. [191, Section 22]. Therefore if contraction holds for the transition semigroup $(p_t)_{t\in I}$ in χ^2-divergence or relative entropy with rate $c > 0$, exponential decay can be deduced for the TV distance with rate $c/2$, i.e

$$\|\nu p_t - \mu\|_{TV} \leq (1/2)\sqrt{\chi^2(\nu p_t|\mu)} \leq (1/2)e^{-ct/2}\sqrt{\chi^2(\nu|\mu)} \qquad \text{and}$$
$$\|\nu p_t - \mu\|_{TV} \leq \sqrt{H(\nu p_t|\mu)/2} \leq e^{-ct/2}\sqrt{H(\nu|\mu)/2}, \qquad \text{respectively.}$$

We remark that common analytic tools to obtain exponential decay in χ^2-divergence and relative entropy are given by functional inequalities such as the Logarithmic Sobolev inequality and the Poincaré inequality, see e.g. [179]. The estimates are obtained by differentiating the f-divergence in time and bound its derivative by the f-divergence itself by applying either the Poincaré inequality or the Logarithmic Sobolev inequality which provide a bound on the χ^2-divergence and the relative entropy, respectively. We note that in addition, if the Logarithmic Sobolev inequality holds, then Talagrand's inequality is satisfied, which bounds the L^2 Wasserstein distance by the relative entropy and exponential decay in Wasserstein distance holds [163], i.e

$$\mathcal{W}^2(\nu p_t, \mu) \leq \sqrt{2H(\nu p_t|\mu)} \leq e^{-ct/2}\sqrt{2H(\nu|\mu)}.$$

Here, we do not focus on the techniques relying on functional inequalities and further references can be found in Appendix B and Appendix C, where the results via analytic tools for the respective framework are discussed.

Finally, we briefly turn our attention to the TV distance and introduce the mixing time. To that end, we assume that the transition function $(p_t)_{t\in I}$ has a unique invariant probability distribution μ and we introduce the time it takes for the distance between the invariant probability distribution μ and the distribution $\delta_x p_t$ of the process started in $x \in \mathbb{X}$ to become smaller than a given value. For a set $K \in \mathcal{B}(\mathbb{X})$ and $t \in I$, we denote the maximal total variation distance to equilibrium at time t for the Markov process $(p_t)_{t\in I}$ started in K by

$$d(t, K) = \sup_{x\in K} \|p_t(x, \cdot) - \mu\|_{\mathrm{TV}}.$$

Fix $\epsilon > 0$. The ϵ-*mixing time* of the Markov process with starting point in K is defined by

$$t_{mix}(\epsilon, K) = \inf\{t \in I : d(t, K) \leq \epsilon\}.$$

We write $t_{mix}(\epsilon)$ for the global ϵ-mixing time $t_{mix}(\epsilon, \mathbb{X})$. A common choice is $\epsilon = 1/4$. Since for all $K \in \mathcal{B}(\mathbb{X})$, $d(t, K)$ is non-increasing in t, it holds that $d(t, K) \leq \epsilon$ for all $t \geq t_{mix}(\epsilon, K)$. The definition of the mixing time is stated, for instance, in [131, Chapter 4], where also a comprehensive study on mixing times to analyse the long-time behaviour of Markov chains is given.

1.2 Markov Chain Monte Carlo methods

In this section, we introduce the Markov Chain Monte Carlo (MCMC) methods which form a class of sampling algorithms going back to [147, 107].

In many applications, one aims to generate samples from a probability distribution μ on some space $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ and to approximately compute quantities of the form $\mu(f) = \int f \mathrm{d}\mu$. For a probability measure μ on $\mathbb{R}$, a direct sampling method is given by the generalised inverse of the cumulative distribution function F. This method is easy to implement if F^{-1} is accessible. Then for $U \sim \mathrm{Unif}(0,1)$, the random variable $X = F^{-1}(U)$ is distributed according to μ. However, exact sampling is often not possible due to the complexity of μ, the high-dimensionality of the state space or since μ is only known up to a multiplicative constant. This motivates us to consider other sampling methods.

The MCMC method simulates a time-homogeneous Markov chain $(\mathbf{X}_n, P)$ with a transition kernel π that leaves the target measure μ invariant, i.e. $\mu\pi = \mu$. More precisely, it holds that $\mu(B) = \int_{\mathbb{R}^d} \mu(\mathrm{d}x)\pi(x, B)$ for all $B \in \mathcal{B}(\mathbb{R}^d)$.

If the Markov chain is *reversible*, i.e. if its transition kernel π satisfies the *detailed balance condition*

$$\mu(\mathrm{d}x)\pi(x, \mathrm{d}y) = \mu(\mathrm{d}y)\pi(y, \mathrm{d}x), \tag{1.9}$$

it follows directly that π leaves the target measure μ invariant since by Fubini's theorem for all $B \in \mathcal{B}(\mathbb{R}^d)$,

$$\mu(B) = \int_B \mu(\mathrm{d}x) = \int_B \int_{\mathbb{R}^d} \mu(\mathrm{d}x)\pi(x, \mathrm{d}y) = \int_{\mathbb{R}^d} \int_B \mu(\mathrm{d}y)\pi(y, \mathrm{d}x) = \int_{\mathbb{R}^d} \mu(\mathrm{d}y)\pi(y, B).$$

Under appropriate ergodic properties for the transition kernel π, one can expect for large $n \in \mathbb{N}$, that the law of $\mathbf{X}_n$ gives a good approximation for the target measure μ and that the integral $\mu(f)$ can approximately be computed by ergodic averages, i.e.

$$\mu(f) \approx \frac{1}{m} \sum_{i=b}^{b+m-1} f(\mathbf{X}_i),$$

where b denotes the *burn-in time*. In simulations, one is often interested in choosing b and m sufficiently large so that the law of the Markov chain after b steps is sufficiently close to the invariant measure μ in an appropriate sense and that the ergodic average involving m steps of the Markov chain builds a good approximation of the quantity of interest.

Before we introduce the Metropolis-Hastings method, which is probably the most well-known MCMC method, let us note that in some MCMC methods the transition kernel leaves the target measure only approximately invariant, i.e. $\mu \approx \mu\pi$ in an appropriate sense. Hence, the law of $\mathbf{X}_n$ does not directly approximate μ and an additional error term occurs. In certain cases, this bias can be uniformly controlled for all steps $m \in \mathbb{N}$, as for uHMC in Appendix A.3 and for more general inexact MCMC methods in [73].

1.2.1 Metropolis-Hastings method

Next, we state the Metropolis-Hastings algorithm which forms the origin of the MCMC methods and which was introduced by Metropolis and his co-authors in [147] and further developed by Hastings in [107]. The basic idea of this method lies in adapting a given proposal transition kernel $p(x, \mathrm{d}y)$ such that the detailed balance condition is satisfied for the new transition kernel π. This modification is implemented by rejecting the proposal with an appropriate probability. We assume that the proposal kernel $p(x, \mathrm{d}y)$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$,

$$p(x, \mathrm{d}y) = p(x, y)\mathrm{d}y,$$

where $p(x,y)$ is a strictly positive density. A move from x to y that is proposed accordingly to $p(x,\mathrm{d}y)$ is accepted with probability

$$\alpha(x,y) = \min\Big(1, \frac{\mu(y)p(y,x)}{\mu(x)p(x,y)}\Big).$$

Otherwise the move is rejected and the Markov chain stays at x. This transition rule is encoded in the transition kernel π given by

$$\pi(x,\mathrm{d}y) = \alpha(x,y)p(x,\mathrm{d}y) + r(x)\delta_x(\mathrm{d}y),$$

where the rejection probability from x to a new state is given by

$$r(x) = \int_{\mathbb{R}^d} (1-\alpha(x,y))p(x,\mathrm{d}y).$$

For the transition kernel π the detailed balance condition holds, since

$$\begin{aligned}\mu(\mathrm{d}x)\pi(x,\mathrm{d}y) &= \mu(\mathrm{d}x)\alpha(x,y)p(x,\mathrm{d}y) + \mu(\mathrm{d}x)r(x)\delta_x(\mathrm{d}y)\\ &= \min(\mu(\mathrm{d}x)p(x,\mathrm{d}y), \mu(\mathrm{d}y)p(y,\mathrm{d}x)) + \mu(\mathrm{d}y)r(y)\delta_y(\mathrm{d}x) = \mu(\mathrm{d}y)\pi(y,\mathrm{d}x),\end{aligned}$$

and hence $\mu\pi = \mu$. The crucial advantage of this method is that it is sufficient to know the target measure only up to a multiplicative constant as this constant is cancelled out in the acceptance probability α.

Algorithm 1.2.1 Metropolis-Hastings algorithm

Require: proposal transition kernel $p(x,\mathrm{d}y)$, initial probability measure $\nu(\mathrm{d}x)$, acceptance probability $\alpha(x,y)$ corresponding to the desired target measure $\mu(\mathrm{d}x)$

$n \leftarrow 0$, sample $\mathbf{X}_0 \sim \nu$;
while Markov chain has not terminated **do**
  sample $\mathbf{Y}_{n+1} \sim p(\mathbf{X}_n, \cdot)$;
  sample $\mathrm{U}_{n+1} \sim \mathrm{Unif}[0,1]$;
  if $\mathrm{U}_{n+1} \leq \alpha(\mathbf{X}_n, \mathbf{Y}_{n+1})$ **then**
    $\mathbf{X}_{n+1} \leftarrow \mathbf{Y}_{n+1}$;
  else
    $\mathbf{X}_{n+1} \leftarrow \mathbf{Y}_n$;
  end if
  $n \leftarrow n+1$;
end while
return $\mathbf{X}_0, \mathbf{X}_1, ...$ Markov chain with initial law ν and invariant measure μ

If the proposal kernel $p(x,\mathrm{d}y)$ describes a random walk, the algorithm is called *Random Walk Metropolis* (RWM). A common choice for the proposal transition kernel $p(x,\mathrm{d}y)$ is given by the normal distribution centred at x with density $p(x,y) = 1/(2\pi)^{d/2}\exp(-|x-y|^2/2)$. In that case, states close to x are more likely to be chosen than states far away, and the acceptance probability simplifies to $\alpha(x,y) = \min(1, \mu(y)/\mu(x))$, since $p(x,y) = p(y,x)$. As stated later in Section 1.4, RWM exhibits a diffusive behaviour and displays slow convergence to the target measure, especially in a high dimensional setting, see [173].

Therefore, it is reasonable to look for a more sophisticated proposal transition kernel that exploits certain information of the target distribution if it is accessible. If, for instance, the gradient of the potential is known, the Hamiltonian Monte Carlo algorithm which is the object of the next section can be considered.

1.2.2 Hamiltonian Monte Carlo

Hamiltonian Monte Carlo, which was first established as *Hybrid Monte Carlo*, is a sampling method that relies on the *Hamiltonian dynamics*. It was originally developed to perform simulations for Lattice Quantum Chromodynamics [69]. In the '90s, Neal exploited the method for statistical computing [155].

Consider a twice differentiable function $U : \mathbb{R}^d \to \mathbb{R}$ satisfying $\int_{\mathbb{R}^d} \exp(-U(x))\mathrm{d}x < \infty$. Let μ be a probability measure on $\mathbb{R}^d$ of the form

$$\mu(\mathrm{d}x) = Z^{-1} \exp(-U(x))\mathrm{d}x, \tag{1.10}$$

where $Z = \int_{\mathbb{R}^d} \exp(-U(x))\mathrm{d}x$ is the normalising constant. The Hamiltonian is defined by $H(x, v) = U(x) + |v|^2/2$, where U corresponds to the potential energy and $|v|^2/2$ to the kinetic energy. Here, we omit an additional mass matrix M appearing often in the kinetic energy. To sample μ, we construct a Markov chain on $\mathbb{R}^d$ using the Hamiltonian dynamics given by

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t} q_t = \frac{\partial H(q_t,p_t)}{\partial p_t} = p_t \\ \frac{\mathrm{d}}{\mathrm{d}t} p_t = -\frac{\partial H(q_t,p_t)}{\partial q_t} = -\nabla U(q_t) \end{cases} \tag{1.11}$$

with initial condition $(q_0, p_0) = (x, v) \in \mathbb{R}^{2d}$. The transition step of *exact HMC* is given by $\mathbf{X}(x) = q_T(x, \xi)$, where $\xi \sim \mathcal{N}(0, \mathrm{I}_d)$ is a standard normally distributed random variable and $T > 0$ is the duration time. The transition kernel π for the time-homogeneous Markov chain corresponding to exact HMC is given by

$$\pi(x, A) = P[q_T(x, \xi) \in A] \qquad \text{for } x \in \mathbb{R}^d, A \in \mathcal{B}(\mathbb{R}^d). \tag{1.12}$$

We note that the Hamiltonian dynamics given by (1.11) preserves the Hamiltonian, i.e. $\mathrm{d}H/\mathrm{d}t = 0$ and it is symplectic and volume preserving [156]. Moreover, the Hamiltonian flow $\varphi_t = (q_t, p_t)$ forms a deterministic Markov process on $\mathbb{R}^{2d}$ with transition function $(\mathrm{p}_t)_{t\geq 0}$, that satisfies the *generalised detailed balance condition* (cf. [84, Section 9])

$$(\mu \otimes \mathcal{N}(0, \mathrm{I}_d))(\mathrm{d}x\mathrm{d}v)\mathrm{p}_t((x, v), \mathrm{d}y\mathrm{d}u) = (\mu \otimes \mathcal{N}(0, \mathrm{I}_d))(\mathrm{d}y\mathrm{d}u)\mathrm{p}_t(S^{-1}(y, u), S^{-1}(\mathrm{d}x\mathrm{d}v)), \tag{1.13}$$

where $\mathrm{p}_t((x, v), \cdot) = \delta_{\varphi_t(x,v)}$ and $\mathcal{S} : \mathbb{R}^{2d} \to \mathbb{R}^{2d}$ is a measurable transformation given by $\mathcal{S}(x, v) = (x, -v)$ for $(x, v) \in \mathbb{R}^{2d}$. More precisely, generalised reversibility

$$\varphi_{-t}(x, v) = \mathcal{S}(\varphi_t(\mathcal{S}(x, v))) \tag{1.14}$$

holds for the Hamiltonian flow and therefore (1.13) is satisfied. These properties of the Hamiltonian dynamics imply that the transition kernel of the Markov chain of exact HMC leaves the target measure μ invariant, i.e. $\mu = \mu\pi$, cf. [156].

Unfortunately, the Hamiltonian dynamics is numerically not exactly solvable. An implementable discretisation of the Hamiltonian dynamics is given by the velocity Verlet integrator,

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t} \tilde{q}_t = \tilde{p}_{\lfloor t \rfloor_h} - \frac{h}{2}\nabla U(\tilde{q}_{\lfloor t \rfloor_h}) \\ \frac{\mathrm{d}}{\mathrm{d}t} \tilde{p}_t = -\frac{1}{2}(\nabla U(\tilde{q}_{\lfloor t \rfloor_h} + \nabla U(\tilde{q}_{\lceil t \rceil_h})) \end{cases}$$

with $(\tilde{q}_0, \tilde{p}_0) = (x, v) \in \mathbb{R}^{2d}$, where $h > 0$ is the discretisation parameter and

$$\lfloor t \rfloor_h = \sup\{s \in h\mathbb{Z} : s \leq t\} \quad \text{and} \quad \lceil t \rceil_h = \inf\{s \in h\mathbb{Z} : s \geq t\}. \tag{1.15}$$

This discretisation scheme which is also called *Leapfrog method* is volume preserving and satisfies generalised reversibility (1.14), see [156, Section 5.2.3]. The transition step of *unadjusted HMC* (uHMC) is given by $x \to \mathbf{X}_h(x) = \tilde{q}_T(x,\xi)$ with $\xi \sim \mathcal{N}(0, \mathrm{I}_d)$. Analogously to exact HMC, the transition kernel for the time-homogeneous Markov chain induced by unadjusted HMC is given by $\pi_h(x, A) = P[\tilde{q}_T(x,\xi) \in A]$ for $x \in \mathbb{R}^d$, $A \in \mathcal{B}(\mathbb{R}^d)$. We note that in general $\mu\pi_h \neq \mu$, since the velocity Verlet integrator does not preserve the Hamiltonian operator. Under appropriate conditions on U, T and h, π_h leaves μ asymptotically invariant and the invariant measure corresponding to π_h is close to μ in an appropriate sense, see Appendix A.

A numerically implementable HMC method that leaves the target measure μ invariant is attained by taking π_h as a proposal kernel and adding a Metropolis adjustment step. The acceptance probability is given by

$$\alpha(x, \tilde{q}_T(x,\xi)) = \min\Big(1, \frac{\exp(-U(\tilde{q}_T(x,\xi)) - \frac{|\tilde{p}_T(x,\xi)|^2}{2})}{\exp(-U(x) - \frac{|\xi|^2}{2})}\Big) = \min\Big(1, \frac{\exp(-H(\tilde{q}_T(x,\xi), \tilde{p}_T(x,\xi)))}{\exp(-H(x,\xi))}\Big),$$

since $\xi \sim \mathcal{N}(0, \mathrm{I}_d)$ and since $(x, -\xi) = (\tilde{q}_T(y,-u), \tilde{p}_T(y,-u))$ for $(y,u) = (\tilde{q}_T(x,\xi), \tilde{p}_T(x,\xi))$ by symmetry of the Hamiltonian dynamics and of the discretisation scheme. The transition step of *Metropolis adjusted HMC* (MaHMC) is given by $x \to \tilde{\mathbf{X}}_h(x) = \tilde{q}_T(x,\xi)\mathbb{1}_{\mathcal{A}(x)} + x\mathbb{1}_{\mathcal{A}(x)^c}$ for $\xi \sim \mathcal{N}(0, \mathrm{I}_d)$, where the event $\mathcal{A}(x)$ is given by

$$\mathcal{A}(x) = \{\mathcal{U} \le \exp(H(x,\xi) - H(\tilde{q}_T(x,\xi), \tilde{p}_T(x,\xi))\} \quad \text{with } \mathcal{U} \sim \mathrm{Unif}[0,1].$$

The transition kernel for the time-homogeneous Markov chain induced by MaHMC is defined by $\tilde{\pi}_h(x, A) = P[\tilde{\mathbf{X}}_h(x) \in A] = P[\{\tilde{q}_T(x,\xi) \in A\} \cap \mathcal{A}(x)] + (1 - P[\mathcal{A}(x)])\delta_x(A)$.

The definitions and statements are based on the work [156] and on the lecture notes [27]. Results on the behaviour of uHMC and MaHMC are postponed to Section 1.5.

Algorithm 1.2.2 Unadjusted HMC/ Metropolis adjusted HMC

Require: initial probability measure ν, duration time $T > 0$, discretisation parameter $h > 0$ such that $T/h \in \mathbb{N}$, ∇U corresponding to the desired target measure μ

```
n ← 0, sample X_0 ∼ ν, K ← T/h;
while Markov chain has not terminated do
    sample ξ ∼ N(0, I_d), q_0 ← X_n, p_0 ← ξ;
    for k = 1, ..., K do
        q_k ← q_{k−1} + h p_{k−1} − (h^2/2) ∇U(q_{k−1});
        p_k ← p_{k−1} − (h/2)(∇U(q_{k−1}) + ∇U(q_k));
    end for
    X_{n+1} ← q_K;                                          {for uHMC}
    Y_{n+1} ← q_K, implement line 4-9 of Algorithm 1.2.1;   {for MaHMC}
    n ← n + 1;
end while
return X_0, X_1, ... Markov chain with initial law ν
```

To complete the description of HMC, let us mention further variants of HMC. Instead of updating the full velocity component in each step, a partial velocity randomisation can be considered [113]. In this case, the initial velocity for the Hamiltonian dynamics in each HMC step is given by $y = \delta y' + \sqrt{1-\delta^2}\xi$ for some $\delta \in (0,1)$, where $\xi \sim \mathcal{N}(0, \mathrm{I}_d)$ and y' is the velocity of the previous step at time T. If the duration time T is not a fixed constant but an exponentially distributed random variable, we refer to the method as *randomised HMC*. The corresponding Markov process leaves the measure $\mu \otimes \mathcal{N}(0, \mathrm{I}_d)$ invariant, see [34]. If additionally the velocity is only updated partially, i.e. only a few randomly selected components are updated, we obtain the so-called *Andersen dynamics*, see [8] and [30] for a recent analysis of this dynamics. We remark that both the randomised HMC and the Andersen dynamics are part of the broader class of Piecewise Deterministic Markov Processes (PDMPs). These processes are characterised by a deterministic flow, a jump or event rate and a transition kernel, determining the transition at the event, cf. [62]. They form a very interesting and promising class of processes and we refer, for instance, to [19, 35, 76] for more details on various PDMPs.

1.3 Stochastic Differential Equations

In this section, we consider diffusions on $\mathbb{R}^d$. We introduce the Langevin dynamics and its overdamped version and give the connection between them. We follow mainly the work of Pavliotis [165, Chapter 4 and 6]. We remark that we are particularly interested in the long-time behaviour of the Langevin diffusions, as the Langevin dynamics can be used to generate samples for given probability measures on $\mathbb{R}^d$ of the form (1.10).

Given $x \in \mathbb{R}^d$ and a d-dimensional standard Brownian motion $(B_t)_{t\geq 0}$, we consider the solution $(X_t)_{t\geq 0}$ of the first-order stochastic differential equation given by

$$\mathrm{d}X_t = -\nabla U(X_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t, \qquad X_0 = x. \tag{1.16}$$

A unique strong solution exists under mild conditions on the drift, e.g. if ∇U is Lipschitz continuous, see e.g. [172, Chapter 9] and [161, Chapter 5.2]. That means that there exists a stochastic process with continuous sample paths, $X_0 = x$, and

$$X_t = X_0 - \int_0^t \nabla U(X_s)\mathrm{d}s + \sqrt{2}B_t \qquad \text{for } t \geq 0.$$

The solution $(X_t)_{t\geq 0}$ is called *overdamped Langevin diffusion*.

Let ν_0 be some probability measure on $\mathbb{R}^d$ of the form $\nu_0(\mathrm{d}x) = \rho_0(x)\mathrm{d}x$ for some probability density function ρ_0. If X_0 is distributed according to ν_0, we remark that for each $t \geq 0$ the probability density function $\rho_t(x)$ of X_t solves the corresponding *Fokker-Planck equation*

$$\partial_t \rho_t(x) = \nabla \cdot (\nabla U(x)\rho_t(x)) + \Delta\rho_t(x), \tag{1.17}$$

see e.g. [165, Section 4.5]. The corresponding generator of the diffusion process $(X_t)_{t\geq 0}$ is given by

$$\mathcal{L} = -\nabla U(x) \cdot \nabla + \Delta. \tag{1.18}$$

Under appropriate conditions on the potential U, the Markov process with generator $\mathcal{L}$ is ergodic and the probability measure μ, given in (1.10), is the invariant distribution. In particular, μ is the unique invariant measure if U is a smooth potential satisfying $\lim_{|x|\to\infty} U(x) = +\infty$ and $e^{-U(x)} \in L^1(\mathbb{R}^d)$, see [165, Proposition 4.2].

A common numerically realisable approximation of (1.23) is given by the *unadjusted Langevin algorithm* (ULA) which uses the Euler discretisation with discretisation parameter $h > 0$ of (1.16) given by

$$\mathbf{X}_k = \mathbf{X}_{k-1} + h\nabla U(\mathbf{X}_{k-1}) + \sqrt{2h}\xi_k,$$

where $(\xi_k)_{k\in\mathbb{N}}$ is a sequence of independent normally distributed random variables, cf. [164, 175, 59, 77]. This method with fixed step size h generates a Markov chain $(\mathbf{X}_k)_{k\geq 0}$ on $\mathbb{R}^d$ that leaves the target measure μ only approximately invariant. Therefore, often an adaptive sequence for the step size $(h_k)_{k\in\mathbb{N}}$ with $h_k \to 0$ as $k \to \infty$ is considered [77, 78]. Alternatively, the implementation of an additional Metropolis-Hastings step produces the *Metropolis-adjusted Langevin algorithm* (MALA), whose corresponding transition kernel leaves the target measure invariant [175, 81, 79, 177].

Next, we consider the classical *Langevin dynamics*, whose origin goes back to modelling the evolution of a particle in statistical physics that is characterised by a position and a velocity component and undergoes damping and external forces [88, 126]. As for the Hamiltonian dynamics, an extended state space $\mathbb{R}^d \times \mathbb{R}^d$ is considered and the diffusion consists of a position (X_t) and a velocity (Y_t) which are driven by the stochastic differential equation

$$\begin{cases} \mathrm{d}X_t = Y_t \mathrm{d}t \\ \mathrm{d}Y_t = -\gamma Y_t \mathrm{d}t - u\nabla U(X_t)\mathrm{d}t + \sqrt{2\gamma u}\mathrm{d}B_t, \end{cases} \qquad (X_0, Y_0) = (x, y), \tag{1.19}$$

where the constant $\gamma \in (0,\infty)$ corresponds to friction, $u \in (0,\infty)$ to the inverse of the mass of the particle and $(x,y) \in \mathbb{R}^{2d}$ is the initial condition. As for the overdamped Langevin diffusion, a solution exists under appropriate mild assumptions on U, i.e. if ∇U is Lipschitz continuous, see e.g. [172, Chapter 9.2]. The generator of the Markov process $(X_t, Y_t)_{t\geq 0}$ is given by

$$\mathcal{L} = y \cdot \nabla_x - u\nabla_x U \cdot \nabla_y + \gamma(-y\nabla_y + u\Delta_y). \tag{1.20}$$

Let $\rho_0(x,y)$ be a probability density function on $\mathbb{R}^{2d}$ and let ν_0 be the probability measure on $\mathbb{R}^{2d}$ of the form $\nu_0(\mathrm{d}x\mathrm{d}y) = \rho_0(x,y)\mathrm{d}x\mathrm{d}y$. If (X_0, Y_0) is distributed according to ν_0, then the probability density function $\rho_t(x,y)$ corresponding to the diffusion process $(X_t, Y_t)_{t\geq 0}$ solves the kinetic Fokker-Planck equation given by

$$\partial_t \rho_t(x,y) = -y \cdot \nabla_x \rho_t(x,y) + u\nabla_x U(x) \cdot \nabla_y \rho_t(x,y) + \gamma(\nabla_y \cdot (y\rho_t(x,y)) + u\Delta_y \rho_t(x,y)).$$

We note that the overdamped Langevin dynamics (1.16) is obtained by taking the limit $\gamma \to \infty$ in (1.19). Under appropriate assumptions on U, e.g., if U is a smooth potential satisfying $\lim_{|x|\to\infty} U(x) = +\infty$ and $e^{-U(x)} \in L^1(\mathbb{R}^d)$, the diffusion $(X_t, Y_t)_{t\geq 0}$ with generator given in (1.20) is ergodic and the Boltzmann-Gibbs measure $\mu \otimes \mathcal{N}(0, u\mathrm{I}_d)$ with μ given in (1.10) is the unique invariant measure, see [165, Proposition 6.1].

As for the overdamped dynamics, this property is used to generate samples via discretised versions of the dynamics, see e.g. [37, 55, 60, 128, 153]. Often, the dynamics is split into the classical velocity Verlet integrator approximating the Hamiltonian dynamics and in the Ornstein-Uhlenbeck process and the steps are successively implemented. The generator $\mathcal{L}$ is decomposed in $\mathcal{L} = \mathcal{L}_A + \mathcal{L}_B + \mathcal{L}_O$ with

$$\mathcal{L}_A = y \cdot \nabla_x, \qquad \mathcal{L}_B = -\nabla_x U \cdot \nabla_y \qquad \text{and} \qquad \mathcal{L}_O = -\gamma y \cdot \nabla_y + \gamma\Delta_y,$$

where $u = 1$ is assumed for simplicity. Let $h > 0$ be the discretisation step and define $\delta = e^{-h\gamma/2}$. Making use of the decomposition, the sampling scheme referred as *OBABO algorithm* is given by

$$\begin{aligned}
\tilde{y}_0 &= \delta y_0 + \sqrt{1-\delta^2}\xi && (O)\\
y_{1/2} &= \tilde{y}_0 - (h/2)\nabla U(x_0) && (B)\\
x_1 &= x_0 + h y_{1/2} && (A)\\
\tilde{y}_1 &= y_{1/2} - (h/2)\nabla U(x_1) && (B)\\
y_1 &= \delta \tilde{y}_1 + \sqrt{1-\delta^2}\xi', && (O)
\end{aligned}$$

where $\xi, \xi' \sim \mathcal{N}(0, \mathrm{I}_d)$ are two independent random variables. This second order scheme satisfies generalised reversibility (1.14) and the generalised detailed balance condition (1.13) with transformation map $\mathcal{S}(x,v) = (x,-v)$ for $(x,v) \in \mathbb{R}^{2d}$. If we consider the scheme $O(BAB)^k O$ where the steps BAB are k-times repeated for some $k \in \mathbb{N}$, we directly recover uHMC for $\delta = 0$ with step size h and duration length $T = hk$. For $\delta > 0$, we obtain uHMC with partial velocity randomisation.

Finally, let us introduce a class of diffusions on $\mathbb{R}$ where the diffusion parameter of the corresponding SDE is not constant and which exhibits a sticky behaviour at 0. Consider the solution $(X_t)_{t\geq 0}$ of the stochastic differential equation with sticky boundary at 0 given by

$$\mathrm{d}X_t = b(X_t)\mathrm{d}t + \mathbb{1}_{\{X_t>0\}}\mathrm{d}W_t, \tag{1.21}$$

where $b : \mathbb{R} \to \mathbb{R}$ is a Lipschitz continuous function and $(W_t)_{t\geq 0}$ is a one-dimensional standard Brownian motion. Existence and uniqueness in law of weak solutions to (1.21) is investigated in e.g. [196, 197]. Sticky diffusions play an important role in analysing sticky couplings, see [87]. A variant of these diffusions is of particular interest in the second work given in Appendix B. An overview of the development of the sticky diffusions first studied by Feller [92] is given in [166] and we refer to the references given in Appendix B.

1.4 Couplings and contraction results

In this section, we address the question of how contraction in Wasserstein distance (1.7) can be established via couplings. Further, we describe direct coupling approaches for solutions to SDEs and exact HMC and show how these approaches are exploited to prove (1.7).

A *coupling of two stochastic processes* $((\mathbf{X}_t), P)$ and $((\mathbf{Y}_t), P')$ both with state space $\mathbb{X}$ is given by a process $((\bar{\mathbf{X}}_t, \bar{\mathbf{Y}}_t), \bar{P})$ with state space $\mathbb{X}\times\mathbb{X}$ such that the laws of $(\bar{\mathbf{X}}_t)_{t\in I}$ and $(\bar{\mathbf{Y}}_t)_{t\in I}$ under $\bar{P}$ coincide with laws of $(\mathbf{X}_t)_{t\in I}$ under P and $(\mathbf{Y}_t)_{t\in I}$ under P', respectively. We say that the coupling is *Markovian* iff the process $((\bar{\mathbf{X}}_t, \bar{\mathbf{Y}}_t), \bar{P})$ is a right-continuous strong Markov process, cf. [84, Definition 3.23]. We note that a coupling of two strong Markov processes is not Markovian in general. For time-discrete processes, the coupling is Markovian if the process satisfies the Markov property. In the time-continuous case, the coupling is Markovian if the transition semi-group is Feller and the process is right-continuous [176, Chapter III.2.8].

Given two Markov chains $(\mathbf{X}_n)_{n\in\mathbb{N}}$ and $(\mathbf{Y}_n)_{n\in\mathbb{N}}$ with transition kernels π and π' on $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$, the transition kernel $\bar{\pi}$ on $(\mathbb{X}\times\mathbb{X}, \mathcal{B}(\mathbb{X})\otimes\mathcal{B}(\mathbb{X}))$ is called a *coupling of* π *and* π' if the measure $\bar{\pi}((x,y), \mathrm{d}x'\mathrm{d}y')$ is a coupling of the measures $\pi(x, \mathrm{d}x')$ and $\pi(y, \mathrm{d}y')$ for all $x, y \in \mathbb{X}$. If γ is a coupling of two probability measures ν and η on $\mathbb{X}$, the canonical Markov chain $((\mathbf{X}_n, \mathbf{Y}_n), P_\gamma)$ with transition kernel $\bar{\pi}$ and initial distribution γ is a Markovian coupling of the Markov chains π and π' and initial distributions ν and η, respectively.

To give the strategy to prove (1.7) via couplings, we focus first on time-discrete Markov processes with one-step transition kernel π. The basic idea relies on finding a suitable coupling and a distance function $\rho : \mathbb{X} \times \mathbb{X} \to [0, \infty)$ such that the generator $\bar{\mathcal{L}} = \bar{\pi} - \mathrm{I}$ associated to the coupling transition kernel $\bar{\pi}$ of two copies of π satisfies

$$\bar{\mathcal{L}}\rho(x, y) \leq -c\rho(x, y) \qquad \text{for all } x, y \in \mathbb{X}. \tag{1.22}$$

Then,

$$\mathcal{W}_\rho(\nu\pi, \eta\pi) \leq E_{\bar{\mathbf{X}}_0 \sim \nu, \bar{\mathbf{Y}}_0 \sim \eta}[\rho(\bar{\mathbf{X}}_1, \bar{\mathbf{Y}}_1)] \leq (1 - c)E_{\bar{\mathbf{X}}_0 \sim \nu, \bar{\mathbf{Y}}_0 \sim \eta}[\rho(\bar{\mathbf{X}}_0, \bar{\mathbf{Y}}_0)] \quad \text{for all } \nu, \eta \in \mathcal{P}(\mathbb{X}).$$

Taking the infimum over all couplings γ of ν and η yields contraction in Wasserstein distance (1.7). To show contraction for a time-continuous Markov process, one is interested in aligning a Markovian coupling of two copies of the time-continuous Markov process with different initial conditions and a distance function ρ such that a similar equation as (1.22) can be proven, where $\bar{\mathcal{L}}$ is replaced by the generator of the time-continuous coupling process.

The idea of combining coupling and distance function is explored by Mu-Fa Chen and Feng-Yu Wang [49, 50] to obtain bounds for the spectral gap. Hairer, Mattingly and Scheutzow [104, 103, 102] and Eberle [83] used this ansatz to get contraction in Wasserstein distance. In particular, Harris type theorems are established in [104, 103, 102], which give contraction in Wasserstein distance by combining a local minorisation condition and a global Lyapunov condition. Eberle applied the interplay of coupling and distance to optimise the contraction rate for diffusions. The coupling and the distance construction of [83] are presented for specific SDEs in the next subsection.

1.4.1 Couplings and contraction results for first order SDEs

We consider the stochastic differential equation

$$\mathrm{d}X_t = b(X_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t, \qquad X_0 = x, \tag{1.23}$$

where $b : \mathbb{R}^d \to \mathbb{R}^d$ is a locally Lipschitz continuous function and $(B_t)_{t\geq 0}$ is a d-dimensional standard Brownian motion. We remark that there exists a unique strong solution $(X_t)_{t\geq 0}$ of (1.23) for a given $x \in \mathbb{R}^d$ and a given standard Brownian motion $(B_t)_{t\geq 0}$.

In the first instance, we impose for the function $b : \mathbb{R}^d \to \mathbb{R}^d$:

Assumption 1.1. *There exists $\kappa > 0$ such that*

$$\langle b(x) - b(y), x - y\rangle \leq -\kappa|x - y|^2 \qquad \textit{for all } x, y \in \mathbb{R}^d.$$

We note that Assumption 1.1 is satisfied for $b = -\nabla U$ where $U \in \mathcal{C}^2(\mathbb{R}^d)$ is some κ-strongly convex potential.

Let $(x, y) \in \mathbb{R}^{2d}$ and let $(B_t)_{t\geq 0}$ be a d-dimensional standard Brownian motion. We define the *synchronous coupling* of two solutions of (1.23) as a diffusion process $(X_t, Y_t)_{t\geq 0}$ on $\mathbb{R}^{2d}$ solving the the stochastic differential equation given by

$$\begin{cases} \mathrm{d}X_t = b(X_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t & X_0 = x, \\ \mathrm{d}Y_t = b(Y_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t, & Y_0 = y. \end{cases}$$

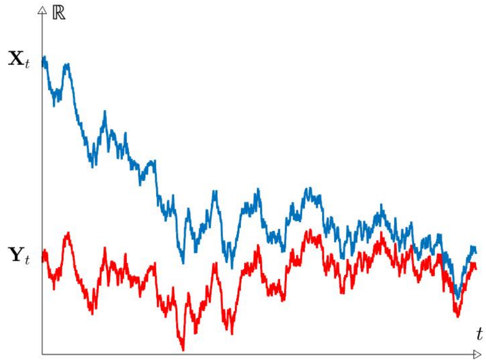

Figure 1.1: *Synchronous coupling of two diffusions on* $\mathbb{R}$ *given by the SDE* $\mathrm{d}X_t = -X_t\mathrm{d}t + \mathrm{d}B_t$ *with different initial values.*

Note that both copies $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ are solutions of (1.23) with different initial conditions and are driven by the same noise, see Figure 1.1.

We note that the process $(X_t - Y_t)_{t\geq 0}$ has a t-continuous sample path. By Assumption 1.1, it holds that

$$\mathrm{d}|X_t - Y_t|^2 = 2\langle X_t - Y_t, b(X_t) - b(Y_t)\rangle \mathrm{d}t \leq -2\kappa |X_t - Y_t|^2 \mathrm{d}t,$$

and hence, $|X_t - Y_t|^2 \leq e^{-2\kappa t}|x-y|^2$ for all $t \geq 0$. Therefore, for all $1 \leq p < \infty$ it holds that

$$\mathcal{W}^p(\delta_x p_t, \delta_y p_t) \leq e^{-\kappa t}\mathcal{W}^p(\delta_x, \delta_y) \qquad \text{for all } x, y \in \mathbb{R}^d \text{ and } t \geq 0.$$

We also obtain contraction for the purely deterministic dynamics when the Brownian motion is absent since Assumption 1.1 yields a global contractivity condition. Further, the contraction rate κ is independent of the dimension d and is often understood as a lower curvature bound.

If Assumption 1.1 is not globally satisfied, it is still possible to prove global contraction in Wasserstein distance by exploiting the noise and using a carefully constructed distance function. Here, we consider the reflection coupling introduced by Lindvall and Rogers in [133]. The *reflection coupling* or *mirror coupling* of two solutions of (1.23) is a diffusion process $(X_t, Y_t)_{t\geq 0}$ on $\mathbb{R}^{2d}$ solving the stochastic differential equation given by

$$\begin{aligned} \mathrm{d}X_t &= b(X_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t, && X_0 = x, \\ \mathrm{d}Y_t &= \begin{cases} b(Y_t)\mathrm{d}t + \sqrt{2}(\mathrm{I}_d - 2e_t e_t^T)\mathrm{d}B_t, & \text{if } t < \tau \\ b(Y_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t, & \text{if } t \geq \tau \end{cases} && Y_0 = y, \end{aligned} \tag{1.24}$$

where $x, y \in \mathbb{R}^d$, $(B_t)_{t\geq 0}$ is a d-dimensional standard Brownian motion $(B_t)_{t\geq 0}$ and $\tau = \inf\{t \geq 0 : X_t = Y_t\}$ denotes the coupling time. For $t < \tau$, the process e_t is given by $e_t = (X_t - Y_t)/|X_t - Y_t|$. In particular, for each time $t < \tau$ the Brownian motion is reflected at the hyperplane with normal vector e_t. This coupling given by (1.24) defines indeed a coupling for (1.23), since $\int_0^t (\mathrm{I}_d - 2e_s e_s^T)\mathrm{d}B_s$ is a standard Brownian motion by Levy's characterisation [172, Chapter 4, Theorem 3.6]. The difference process satisfies for $t < \tau$,

$$\mathrm{d}(X_t - Y_t) = (b(X_t) - b(Y_t))\mathrm{d}t + \sqrt{8}e_t e_t^T \mathrm{d}B_t. \tag{1.25}$$

Instead of Assumption 1.1, we impose the weaker condition:

Assumption 1.2. *There exist $\kappa > 0$, $L > 0$ and $R > 0$ such that*

$$\langle b(x) - b(y), x - y \rangle \leq (-\kappa \mathbb{1}_{\{|x-y| \geq R\}} + L \mathbb{1}_{\{|x-y|<R\}})|x-y|^2 \qquad \textit{for all } x, y \in \mathbb{R}^d.$$

If there exists $U \in \mathcal{C}^2(\mathbb{R}^d)$ such that $b = \nabla U$, this condition corresponds to U being strongly convex outside a Euclidean ball. The assumption is satisfied for instance for a double-well potential.

Then, under Assumption 1.2 the reflection coupling combined with a modified distance function gives contraction in average. Consider a non-decreasing, concave function $f : [0, \infty) \to [0, \infty)$ which is $\mathcal{C}^1$ and satisfies $f(0) = 0$. For $r_t = |X_t - Y_t|$, we obtain by (1.25), Assumption 1.2 and by Ito's formula,

$$\mathrm{d}f(r_t) \leq f'(r_t)(-\kappa \mathbb{1}_{\{r_t \geq R\}} + L \mathbb{1}_{\{r_t < R\}}) r_t \mathrm{d}t + \sqrt{2} f'(r_t) 2\mathrm{d}W_t + 4 f''(r_t)\mathrm{d}t \qquad \text{for } t < \tau,$$

where $W_t = \int_0^t {e_s}^T \mathrm{d}B_s$ is a one-dimensional Brownian motion by Levy's characterisation. If there exists a non-decreasing, concave f satisfying

$$f'(r)(-\kappa \mathbb{1}_{\{r \geq R\}} + L \mathbb{1}_{\{r<R\}}) r + 4 f''(r) \leq -c f(r) \tag{1.26}$$

for some constant $c > 0$, we can deduce contraction in L^1 Wasserstein distance with respect to the distance function $\rho : \mathbb{R}^d \times \mathbb{R}^d \to [0, \infty)$ given by $\rho(x, y) = f(|x - y|)$ for all $x, y \in \mathbb{R}^d$. Condition (1.26) is satisfied for

$$f(r) = \int_0^r g(s)\varphi(s)\mathrm{d}s, \qquad \text{where}$$
$$\varphi(r) = \exp(-L(r \wedge R)^2/8), \quad \Phi(r) = \int_0^r \varphi(s)\mathrm{d}s, \quad g(r) = 1 - \frac{1}{2}\int_0^{r \wedge R} \frac{\Phi(s)}{\varphi(s)}\mathrm{d}s \Big/ \int_0^R \frac{\Phi(s)}{\varphi(s)}\mathrm{d}s.$$

Then, by (1.26) for the above constructed function f,

$$\frac{\mathrm{d}}{\mathrm{d}t} E[f(r_t)] \leq -c E[f(r_t)]$$

with rate $c = \min(\kappa R \varphi(R)/2(\int_0^R g(s)\varphi(s)\mathrm{d}s)^{-1}, 2(\int_0^R \Phi(s)\varphi(s)^{-1}\mathrm{d}s)^{-1})$ and hence by Grönwall's inequality

$$\mathcal{W}_\rho(\delta_x p_t, \delta_y p_t) \leq E[f(|X_t - Y_t|)] \leq e^{-ct} E[f(|X_0 - Y_0|)].$$

Taking the infimum over all couplings yields contraction in L^1 Wasserstein distance with respect to ρ. Since f satisfies

$$r\varphi(R)/2 = r f'(R) \leq f(r) \leq \Phi(r) \leq r \qquad \text{for any } r \in \mathbb{R}_+,$$

we obtain contraction in L^1 Wasserstein distance with respect to the Euclidean distance,

$$\mathcal{W}(\delta_x p_t, \delta_y p_t) \leq M e^{-ct} \mathcal{W}(\delta_x, \delta_y) \qquad \text{for all } t \geq 0 \text{ and } x, y \in \mathbb{R}^d,$$

where $M = 2\varphi(R)^{-1} = 2\exp(LR^2/2)$.

Alternatively, there exist constants $\tilde{c}, a, M_1, R_1 \in (0, \infty)$ and a concave function $\tilde{f} : [0, \infty) \to [0, \infty)$ such that $\tilde{f}$ is $\mathcal{C}^1$, satisfies $\tilde{f}(0) = 0$, $\lim_{r \downarrow 0} \tilde{f}(r) = a$, $\lim_{r \downarrow 0} \tilde{f}'(r) = 1$, $f'(r) = 1/M_1$ for

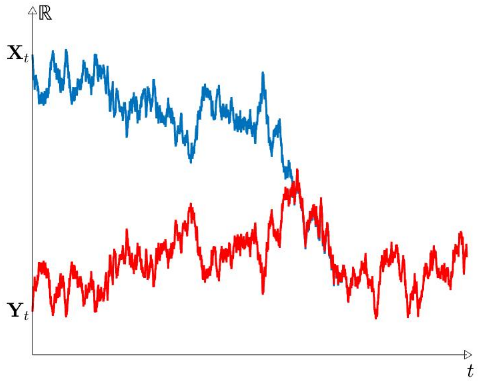

Figure 1.2: *Reflection coupling of two diffusions on* $\mathbb{R}$ *given by the SDE* $\mathrm{d}X_t = -4X_t(X_t-1)\mathrm{d}t + \mathrm{d}B_t$ *with different initial values.*

all $r \geq R_1$ and $\tilde{f}$ solves (1.26). This function has the advantage that we can deduce additionally bounds in TV distance of the form

$$\|\delta_x p_t, \delta_y p_t\|_{TV} \leq e^{-\tilde{c}t}(a + \mathcal{W}(\delta_x, \delta_y))/a \qquad \text{for all } t \geq 0 \text{ and } x, y \in \mathbb{R}^d.$$

The above calculations rely on [83, 84] where Eberle obtained contraction in Wasserstein distance and TV distance using a reflection coupling and a distance involving a carefully constructed concave function. The technique using couplings and concave functions is captured in different frameworks in subsequent works, see e.g. [31, 85].

1.4.2 Coupling and contraction for SDEs with degenerate noise

Next, we exhibit contraction for explicit SDEs with degenerate noise applying a synchronous coupling. We consider the Langevin dynamics (1.19), where the potential U is κ-strongly convex and has a Lipschitz continuous gradient, i.e. ∇U satisfies Assumption 1.1 and there exists $L \in (0, \infty)$ such that $|\nabla U(x) - \nabla U(y)| \leq L|x-y|$ for all $x, y \in \mathbb{R}^d$. This framework is also studied in detail in [181], see Appendix C. Since this case gives insight into the order of the contraction rate, the main calculations and the contraction results are provided here.

Given $x, y, x', y' \in \mathbb{R}^d$, a d-dimensional standard Brownian motion $(B_t)_{t\geq 0}$ and $\gamma, u > 0$, we consider the synchronous coupling $(X_t, Y_t, X'_t, Y'_t)_{t\geq 0}$ of (1.4) given by

$$\begin{cases} \mathrm{d}X_t = Y_t \mathrm{d}t \\ \mathrm{d}Y_t = (-\gamma Y_t - u\nabla U(X_t))\mathrm{d}t + \sqrt{2\gamma u}\mathrm{d}B_t, \end{cases} \qquad (X_0, Y_0) = (x, y),$$

$$\begin{cases} \mathrm{d}X'_t = Y'_t \mathrm{d}t \\ \mathrm{d}Y'_t = (-\gamma Y'_t - u\nabla U(X'_t))\mathrm{d}t + \sqrt{2\gamma u}\mathrm{d}B_t, \end{cases} \qquad (X'_0, Y'_0) = (x', y').$$

Under the above condition on U, there exists a unique strong solution $(X_t, Y_t)_{t\geq 0}$. We refer to Figure 1.3 for an illustration of the coupling.

Since U is κ-strongly convex with Lipschitz continuous gradients, there exists a positive definite matrix $K \in \mathbb{R}^{d\times d}$ with smallest eigenvalue $\kappa > 0$ and a convex function $G : \mathbb{R}^d \to \mathbb{R}^d$

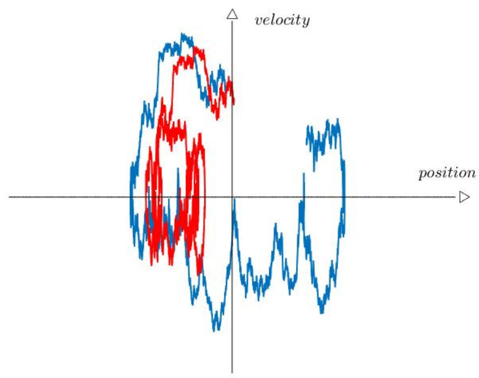

Figure 1.3: *Synchronous coupling of two Langevin diffusions on* $\mathbb{R}^2$.

with a Lipschitz continuous gradient such that

$$U(x) = (x \cdot (Kx))/2 + G(x). \tag{1.27}$$

Note that this splitting is not unique and a natural choice is given by $K = \kappa I_d$. As we see in later computations, we are particularly interested in a splitting, where the Lipschitz constant L_G of the gradient of G is minimised.

The convexity and the Lipschitz continuous gradient of G, i.e.

$$\langle \nabla G(x) - \nabla G(x'), x - x' \rangle \geq 0 \quad \text{and} \quad |\nabla G(x) - \nabla G(x')| \leq L_G |x - x'| \qquad \text{for all } x, x' \in \mathbb{R}^d,$$

imply co-coercivity of G (see [157, Theorem 2.1.5]),

$$|\nabla G(x) - \nabla G(x')|^2 \leq L_G \langle \nabla G(x) - \nabla G(x'), x - x' \rangle \qquad \text{for all } x, x' \in \mathbb{R}^d.$$

Let $(Z_t, W_t)_{t \geq 0} = (X_t - X_t', Y_t - Y_t')_{t \geq 0}$ denote the difference process and let $A, B, C \in \mathbb{R}^{d \times d}$ be positive definite matrices of the form

$$A = \gamma^{-2} u K + (1/2)(1 - 2\lambda)^2 \mathrm{I}_d, \qquad B = (1/2)(1 - 2\lambda)\gamma^{-1} \mathrm{I}_d \qquad \text{and} \qquad C = \gamma^{-2} \mathrm{I}_d$$

with $\lambda = \min(1/8, \kappa u \gamma^{-2}/2)$. We consider the function $\rho : \mathbb{R}^{2d} \times \mathbb{R}^{2d} \to [0, \infty)$ given by

$$\begin{aligned} \rho((x,y),(x',y')) &= ((x - x') \cdot (A(x - x')) + (x - x') \cdot (B(y - y')) + (y - y') \cdot (C(y - y'))^{1/2} \\ &= (\gamma^{-2} u (x - x') \cdot (K(x - x')) + (1/2)|(1 - \lambda)(x - x') + \gamma^{-1}(y - y')|^2 \\ &\quad + (1/2)\gamma^{-2}|y - y'|^2)^{1/2} \end{aligned}$$

for all $(x,y),(x',y') \in \mathbb{R}^{2d}$, which defines a metric that is equivalent to the Euclidean distance $d((x,y),(x',y')) = |(x,y) - (x',y')|$, i.e. there exists $C_1, C_2 \in (0, \infty)$ such that

$$C_1 d((x,y),(x',y')) \leq \rho((x,y),(x',y')) \leq C_2 d((x,y),(x',y')) \qquad \text{for all } (x,y),(x',y') \in \mathbb{R}^{2d}.$$

Then by Ito's formula, Young's inequality and co-coercivity of G, it holds that for $\rho_t^2 = \rho((X_t, Y_t), (X_t', Y_t'))^2$

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\rho_t^2 &\le 2W_t \cdot (AZ_t) + 2W_t \cdot (BW_t) - 2Z_t \cdot (B(uKZ_t - u(\nabla G(X_t) - \nabla G(X_t'))))\\ &- 2W_t \cdot (CW_t) - W_t \cdot (C(uKZ_t - u(\nabla G(X_t) - \nabla G(X_t'))))\\ &\le -(1-2\lambda)\gamma^{-1}uZ_t \cdot (KZ_t) - ((1-2\lambda)\gamma^{-1}u + L_G u^2\gamma^{-3})Z_t \cdot (\nabla G(X_t) - \nabla G(X_t'))\\ &- 2\lambda\gamma(2Z_t \cdot (BZ_t) + W_t \cdot (CW_t)).\end{aligned}$$

If $L_G u\gamma^{-2} \le 3/4$, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho_t^2 \le -2c\rho_t^2,$$

where $c = \lambda\gamma$, since $-(1-4\lambda)\gamma^{-1}uZ_t \cdot (KZ_t) \le -\gamma^{-1}u(\kappa/2)|Z_t|^2 \le -\lambda\gamma|Z_t|^2$ by the definition of λ. Then by Grönwall's inequality, $\rho_t \le e^{-ct}\rho_0$ for all $t \ge 0$. Therefore, if $L_G u\gamma^{-2} \le 3/4$, it holds that for all $1 \le p < \infty$

$$\mathcal{W}_\rho^p(\mu_t, \nu_t) \le e^{-ct}\mathcal{W}_\rho^p(\delta_{(x,y)}, \delta_{(x',y')}) \qquad \text{for all } (x,y),(x',y') \in \mathbb{R}^{2d} \text{ and } t \ge 0,$$

where $\mu_t = \mathrm{Law}(X_t, Y_t)$ and $\nu_t = \mathrm{Law}(X_t', Y_t')$. By the equivalence of ρ and d, we obtain

$$\mathcal{W}^p(\mu_t, \nu_t) \le Me^{-ct}\mathcal{W}^p(\delta_{(x,y)}, \delta_{(x',y')}) \qquad \text{for all } (x,y),(x',y') \in \mathbb{R}^{2d} \text{ and } t \ge 0,$$

where $M = C_2/C_1$.

A study of the contraction for Langevin dynamics in the above framework is also given in [60, 152] using a synchronous coupling approach and a different distance function. If U is not strongly convex, the analysis becomes more involved and we need to make use of the noise which is only present in the velocity component. In [85], a new coupling is established yielding local contraction for small distances in an appropriate distance. This approach combined with a semimetric involving a Lyapunov function yields a contraction result for strongly convex potentials outside a Euclidean ball. Another approach relying on a novel distance function is presented in Appendix C.

1.4.3 Coupling and contraction for exact HMC

Next, we present the synchronous coupling construction for exact HMC and its impact on obtaining contraction. The computations rely on [31, Lemma 3.4]. Given $(x,y) \in \mathbb{R}^{2d}$ and $\xi \sim \mathcal{N}(0, \mathrm{I}_d)$, the *synchronous coupling* of two transition kernels $\pi(x,\cdot)$ and $\pi(y,\cdot)$ is given by the transition step $(\mathbf{X}(x,y), \mathbf{Y}(x,y))$ satisfying

$$\mathbf{X}(x,y) = q_t(x,\xi) \qquad \text{and} \qquad \mathbf{Y}(x,y) = q_t(y,\xi).$$

The difference process $(z_t, w_t)_{t\ge0} = (q_t(x,\xi) - q_t(y,\xi), p_t(x,\xi) - p_t(y,\xi))_{t\ge0}$ is given by

$$\begin{cases}\frac{\mathrm{d}}{\mathrm{d}t}z_t = w_t\\ \frac{\mathrm{d}}{\mathrm{d}t}w_t = -(\nabla U(q_t(x,\xi)) - \nabla U(q_t(y,\xi)).\end{cases}$$

We suppose that U is twice differentiable and ∇U satisfies Assumption 1.1 and is Lipschitz continuous with Lipschitz constant L. Note that $\kappa \leq L$. Then for $(a(t), b(t)) = (|z_t|^2, 2z_t \cdot w_t)$ it holds that

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t} a(t) = b(t) & a(0) = |z_0|^2 = |x-y|^2, \\ \frac{\mathrm{d}}{\mathrm{d}t} b(t) = -2\kappa a(t) + \beta(t) & b(0) = 0, \end{cases}$$

where $\beta(t) = -(\nabla U(q_t(x,\xi)) - \nabla U(q_t(y,\xi)) \cdot z_t + 2\kappa a(t) + 2|w_t|^2 \leq 2|w_t|^2$. The initial value problem is solved by

$$a(t) \leq \cos(\sqrt{2\kappa} t) a(0) + \int_0^t (2\kappa)^{-1/2} \sin(\sqrt{2\kappa}(t-r)) \beta(r) \mathrm{d}r.$$

By Taylor expansion and if $Lt^2 \leq 1$, it holds that $\cos(\sqrt{2\kappa}t) \leq 1 - \kappa t^2 + (1/6)\kappa^2 t^4 \leq 1 - (5/6)\kappa t^2$, and $(2\kappa)^{-1/2} \sin(\sqrt{2\kappa}(t-r)) \leq (t-r)$. To bound $\beta(r) \leq |w_r|^2$, we note that

$$\begin{aligned} \max_{s \leq t} |z_s - z_0| &\leq \max_{s \leq t} \int_0^s \int_0^u |\nabla U(q_r(x,\xi)) - \nabla U(q_r(y,\xi)| \mathrm{d}r \mathrm{d}u \\ &\leq \frac{Lt^2}{2} \max_{s \leq t} |z_s| \leq \frac{Lt^2}{2} \max_{s \leq t} (|z_s - z_0| + |z_0|). \end{aligned}$$

For $Lt^2 \leq 1$, $\max_{s \leq t} |z_s - z_0| \leq Lt^2 |z_0|$, and hence, $\max_{s \leq t} |z_s| \leq (1 + Lt^2)|z_0|$. Then,

$$\max_{s \leq r} |w_s| \leq \max_{s \leq r} \int_0^s |\nabla U(q_u(x,\xi)) - \nabla U(q_u(y,\xi)| \mathrm{d}u \leq Lr \max_{s \leq r} |z_s| \leq Lr(1 + Lr^2)|z_0|.$$

Hence, if $Lt^2 \leq \min(1, \kappa/L) = \kappa/L$, then

$$\begin{aligned} a(t) &\leq (1 - (5/6)\kappa t^2)|z_0|^2 + \int_0^t \Big((t-r) 2 (Lr(1 + Lr^2)|z_0|)^2\Big) \mathrm{d}r \\ &\leq (1 - (5/6)\kappa t^2)|z_0|^2 + (1/6)t^4 (L(1 + Lt^2)^2 |z_0|^2 \leq (1 - (5/6)\kappa t^2)|z_0|^2. \end{aligned}$$

Therefore, we obtain for $LT^2 \leq \kappa/L$

$$|\mathbf{X}(x,y) - \mathbf{Y}(x,y)| \leq (1 - (1/12)\kappa T^2)|x - y|.$$

and for all $1 \leq p < \infty$ and $m \in \mathbb{N}$

$$\mathcal{W}^p(\delta_x \pi^m, \delta_y \pi^m) \leq e^{-cm} \mathcal{W}^p(\delta_x, \delta_y) \qquad \text{with } c = (1/12)\kappa T^2,$$

see Figure 1.4 illustrating the synchronous coupling and the contraction.

If Assumption 1.1 is relaxed to a more general assumption, the analysis of the convergence behaviour becomes more involved. If Assumption 1.2 is supposed, contraction in an L^1 Wasserstein distance is shown for exact HMC and local contraction is proven for MaHMC in [31] by applying a new coupling approach. Contraction in an L^1 Wasserstein distance for the uHMC is covered in Appendix A.

1.5 Summary of existing contraction results

In this section, we give an overview and comparison of existing contraction and complexity results. We begin with summarising and comparing the results for overdamped Langevin diffusions, classical Langevin diffusions and exact HMC before we state results for numerically implementable sampling algorithms.

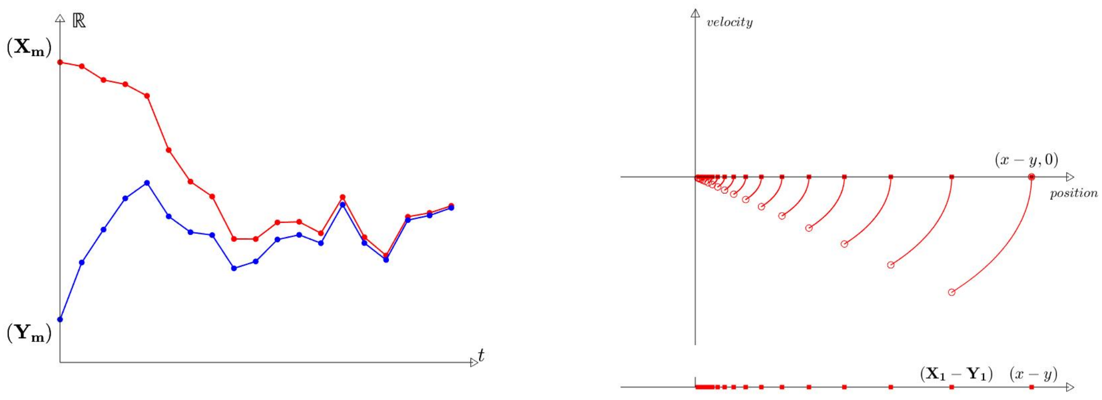

Figure 1.4: *Two synchronously coupled HMC chains* $(\mathbf{X}_m)_{m\geq 0}$ *and* $(\mathbf{Y}_m)_{m\geq 0}$ *on* $\mathbb{R}$ *for a quadratic potential. Left: Evolution in time of the two Markov chains. Right top: Difference of the two Hamiltonian flows* $(q_t(x,\xi)-q_t(y,\xi), p_t(x,\xi)-p_t(y,\xi))$ *in* $\mathbb{R}^2$. *Right bottom: Difference of the two corresponding Markov chains* $(\mathbf{X}_m-\mathbf{Y}_m)_{m\geq 0}$.

1.5.1 Analysis and comparison of results given in Section 1.4

By Section 1.4.1, contraction for the overdamped Langevin dynamics with $b = u\nabla U$ holds with contraction rate κ for κ-strongly convex potentials U (i.e. ∇U satisfies Assumption 1.1). An additional inverse mass u is considered in the SDE which is chosen to be of order $\mathcal{O}(L^{-1})$ to guarantee that the drift is of order $\mathcal{O}(1)$. Hence, the contraction rate is of order $\mathcal{O}(\mathcal{K}^{-1})$ where $\mathcal{K} = L/\kappa$ denotes the condition number. The contraction rate coincides with the contraction rate obtained via functional inequalities, cf. [11, Theorem 9.7.2]. If the potential U is only κ-strongly convex outside a Euclidean ball with radius R and has an L-Lipschitz lower bounded gradient (see Assumption 1.2), contraction holds with a rate that is upper bounded by κ and a quantity depending on L and R.

For the Langevin dynamics with quadratic potentials satisfying Assumption 1.1, we deduce from Section 1.4.2 that contraction holds for any friction coefficient $\gamma > 0$ with contraction rate $c = \gamma/8 \wedge \kappa u\gamma^{-1}/2$, which fits in with the spectral gap obtained in [165, Chapter 6.3] up to a constant. Then, for $\gamma = 2\sqrt{\kappa u}$ the contraction rate is optimised and given by $c = \sqrt{\kappa u}/4$. If an additional convex perturbation is considered as in Section 1.4.2, the additional constraint $L_G u\gamma^{-2} \leq 3/4$ appears where L_G denotes the Lipschitz constant of gradient of the convex perturbation. If $L_G > 3\kappa$, the rate is optimised for $\gamma = \sqrt{3L_G u/4}$ and is of order $\mathcal{O}(\kappa\sqrt{u}/\sqrt{L_G})$. In particular if L_G is of the same order as the Lipschitz constant L of ∇U, the rate is of the same order as in the work by Dalalyan and Riou-Durand [60]. If $L_G \leq 3\kappa$, we obtain a contraction rate of order $\mathcal{O}(\sqrt{\kappa u})$ for $\gamma = 2\sqrt{\kappa u}$ as for the quadratic potentials. This rate corresponds with the rate of the contraction result in L^2 norm obtained in [40]. To ensure that the drift is of order $\mathcal{O}(1)$ as in the overdamped case, we suppose $u \in \mathcal{O}(1/L)$. Then the rate is of order $\mathcal{O}(\mathcal{K}^{-1/2})$ for $L_G \leq 3\kappa$ and of order $\mathcal{O}(\mathcal{K}^{-1})$ for $L_G > 3\kappa$, respectively.

If Assumption 1.2 is supposed instead of Assumption 1.1, contraction is established in [85] by aligning a coupling approach and a semimetric based on a Lyapunov function. The approach is further developed in Appendix C observing a contraction rate that depends on κ, L_G, γ and R, but is independent of the dimension, see Section 2.3 for an outlook of the results.

By Section 1.4.3, we obtain a contraction rate of order $\mathcal{O}(\mathcal{K}^{-2})$ for exact HMC if T is of order $\mathcal{O}(\sqrt{\kappa}/L)$. In [27, Section 5], the constraint on T is improved to $LT^2 \leq 1/4$ leading to a

Dynamics	**rate**	**constraint**	**for**	**rate**
overdamped LD	$\mathcal{O}(\kappa u)$		$u \in \mathcal{O}(1/L)$	$\mathcal{O}(\mathcal{K}^{-1})$
LD (if $L_G \le 3\kappa$)	$\mathcal{O}((\gamma/4) \wedge (\kappa u \gamma^{-1}))$	$L_G u \gamma^{-2} \le 3/4$	$\gamma = 2\sqrt{\kappa u}$, $u \in \mathcal{O}(1/L)$	$\mathcal{O}(\mathcal{K}^{-1/2})$
LD (if $L_G > 3\kappa$)	$\mathcal{O}((\gamma/4) \wedge (\kappa u \gamma^{-1}))$	$L_G u \gamma^{-2} \le 3/4$	$\gamma = 2\sqrt{L_G u/3}$, $u \in \mathcal{O}(1/L)$	$\mathcal{O}(\mathcal{K}^{-1})$
HMC [Section 1.4.3]	$\mathcal{O}(\kappa T^2)$	$LT^2 \le (\kappa/L)$	$T \in \mathcal{O}(\sqrt{\kappa}/L)$	$\mathcal{O}(\mathcal{K}^{-2})$
HMC [27, Thm 5.4]	$\mathcal{O}(\kappa T^2)$	$LT^2 \le 1/4$	$T \in \mathcal{O}(1/\sqrt{L})$	$\mathcal{O}(\mathcal{K}^{-1})$

Table 1.1: *Contraction rates for contraction in Wasserstein distance of overdamped and classical Langevin diffusions (LD) and of HMC for κ-strongly convex potential with L-Lipschitz continuous gradients.*

contraction rate of order $\mathcal{O}(\mathcal{K}^{-1})$.

1.5.2 Results of numerically implementable sampling methods

Next, we focus on the contraction and complexity results for numerically implementable sampling algorithms.

First, let us mention the optimal scaling limits which are an often considered technique to accomplish a non-asymptotic analysis of MCMC methods in high dimensions. For this purpose, a target measure of product form $\mu(\mathrm{d}x) \propto \prod_{i=1}^n \exp(-U(x^i))\mathrm{d}x$ on $\mathbb{R}^{nd}$ is considered where $U : \mathbb{R}^d \to \mathbb{R}$ satisfies sufficient regularity conditions, and the optimal choice of the free parameters in the sampling algorithm is studied for $n \to \infty$ to balance the computational cost and to make acceptable moves. Although this approach is not further discussed in the thesis, we review some known results for completeness. First introduced in [173], Gelman, Gilks and Roberts showed that in the Random Walk Metropolis the step size should be chosen of order $\mathcal{O}(n^{-1})$ to obtain an average acceptance probability of $\mathcal{O}(1)$. For MALA, the optimal order for the step size improves to $\mathcal{O}(n^{-1/3})$, cf. [174, 168]. For Metropolis adjusted HMC, a discretisation step h of order $\mathcal{O}(n^{-1/4})$ leads to an average acceptance probability of $\mathcal{O}(1)$, cf. [17].

Furthermore, one is interested in non-asymptotic bounds of the error between the distribution obtained by running the sampling algorithm and the target measure μ. In the following, we assume that the potential $U : \mathbb{R}^d \to \mathbb{R}$ of the target measure $\mu(\mathrm{d}x) \propto \exp(-U(x))\mathrm{d}x$ is differentiable, κ-strongly convex and has an L-Lipschitz continuous gradient. As before, $\mathcal{K} = L/\kappa$ denotes the condition number. Given $M_0 \in \mathbb{R}_+$, we say that an initial measure μ_0 has a M_0-warm start if

$$\sup_{B \in \mathcal{B}(\mathbb{R}^d)} \frac{\mu_0(B)}{\mu(B)} = M_0$$

and the constant M_0 does not scale with the condition number $\mathcal{K}$ or the dimension d. Under a warm start, RWM has an ϵ-mixing in TV distance of order $\mathcal{O}(d\mathcal{K}^2)$ [81]. For MALA, [199] proved that given $\epsilon > 0$ an ϵ-mixing in TV distance occurs in $\mathcal{O}(\mathcal{K}d^{1/2})$ steps with step size scaling with $h \propto (L\sqrt{d})^{-1}$. These upper bounds improve previous results given in [56, 127] and are of the same order in the condition number $\mathcal{K}$ and the dimension d as the lower bounds. For Metropolis adjusted HMC, ϵ-mixing occurs in $\mathcal{O}(d^{2/3}\mathcal{K})$ steps, cf. [52]. In all three results, terms of logarithmic order in $\mathcal{K}$, d and ϵ are omitted and the proofs rely on conductance techniques

Method	ϵ-mixing	#gradient evaluations	step size
RWM [81]	$d\mathcal{K}^2$	—	$h \propto (dL\mathcal{K})^{-1}$
MALA [199]	$d^{1/2}\mathcal{K}$	$d^{1/2}\mathcal{K}$	$h \propto (L\sqrt{d})^{-1}$
MaHMC [52]	$d^{2/3}\mathcal{K}$	$d^{11/12}\mathcal{K}$	$h \propto (Ld^{7/6})^{-1/2}$
ULA [59]	$d\mathcal{K}^2/\epsilon^2$	$d\mathcal{K}^2/\epsilon^2$	$h \propto (L\mathcal{K}d^2\epsilon^{-2})^{-1}$

Table 1.2: *Upper bounds on ϵ-mixing in TV distance starting from a warm start. Here, logarithmic factors are omitted.*

which measure the bottleneck ratio of a Markov process in equilibrium and which give lower and upper bounds on the mixing time, cf. [131, 47].

Focusing on unadjusted sampling algorithms, Dalalyan [59] proved that for a warm start ULA achieves an ϵ-approximation of the target measure in total variation distance in $\mathcal{O}(d\mathcal{K}^2/\epsilon^2)$ steps where the logarithmic dependences are omitted. By [78], the same order of steps lead to an ϵ-approximation in L^2 Wasserstein distance. In [55], an ϵ-approximation in L^2 Wasserstein distance is shown in $\mathcal{O}(\sqrt{d}\mathcal{K}/\epsilon)$ steps for a discretisation scheme of Langevin diffusion. In Appendix A, contraction in Wasserstein distance and complexity bounds for uHMC are provided without assuming a warm start and further improved for the strongly convex case in [27, Section 5]. If the duration time satisfies $LT^2 \leq 1/4$, contraction is shown with rate $\kappa T^2/4$, which leads in combination with the restriction on the duration time to a rate of order $\mathcal{O}(\mathcal{K}^{-1})$ as in the exact case. Then complexity bounds for uHMC are obtained by applying the following general triangle trick, cf. [141]. If contraction in Wasserstein distance with respect to some distance ρ is satisfied for uHMC, it holds for the target measure μ and the invariant probability measures μ_h of uHMC that

$$\mathcal{W}_\rho(\mu,\mu_h) \leq \mathcal{W}_\rho(\mu\pi,\mu\pi_h) + \mathcal{W}_\rho(\mu\pi_h,\mu_h\pi_h) \leq \mathcal{W}_\rho(\mu\pi,\mu\pi_h) + (1-c)\mathcal{W}_\rho(\mu,\mu_h), \tag{1.28}$$

which implies

$$\mathcal{W}_\rho(\mu,\mu_h) \leq c^{-1}\mathcal{W}_\rho(\mu\pi,\mu\pi_h).$$

Note that the unique existence of μ_h holds by the contraction result. Therefore, the distance between μ and μ_h can be bounded from above by estimating the strong accuracy of the sampling algorithm. In the case of strongly convex potentials with bounded third derivatives, $\mathcal{W}^1(\mu\pi,\mu\pi_h)$ is of order $\mathcal{O}(h^2 d)$. Then given $\epsilon > 0$ and h of order $\mathcal{O}(d^{-1/2})$, an ϵ-approximation in Wasserstein distance of the target measure can be achieved by $\mathcal{O}(d^{1/2}\log(d/\epsilon))$ gradient evaluations if the Wasserstein distance between the initial measure and the target measure is of order $\mathcal{O}(d)$. In [28], for given $\epsilon > 0$, it is shown that an ϵ-mixing of uHMC to its invariant measure is of order $\mathcal{O}(\log(d/\epsilon))$ by applying the bounds in Wasserstein distance and a one-shot coupling. Hence, an ϵ-approximation of the invariant measure of uHMC is obtained in $\mathcal{O}(h^{-1}\log(d/\epsilon))$ gradient evaluations. An ϵ-approximation of the target measure is achieved in $\mathcal{O}(d^{3/4}\epsilon^{-1/2}\log(d/\epsilon))$ gradient evaluations provided the discretisation step h is chosen of order $\mathcal{O}(d^{-3/4}\epsilon^{1/2})$. This idea of estimating the complexity via the triangle trick is further developed for general inexact sampling algorithms in [73]. For ULA, the Wasserstein distance between the target measure and the invariant measure of ULA is of order $\mathcal{O}(hd)$. For h chosen of order $\mathcal{O}(d^{-1})$, an ϵ-approximation to the target measure can be achieved in $\mathcal{O}(d\log(d/\epsilon))$ gradient evaluations. For frameworks satisfying Assumption 1.2 instead of Assumption 1.1, we refer for uHMC to Section 2.1, where the results of Appendix A are summarised, and to [73, Example 15] for ULA.

1.6 Propagation of chaos

1.6.1 Introduction to propagation of chaos

In this section, we introduce the phenomenon *propagation of chaos* and describe the connection between solutions of nonlinear stochastic differential equations given by

$$\begin{cases} \mathrm{d}\bar{X}_t = \Big(b(\bar{X}_t) + \int_{\mathbb{R}^d} \tilde{b}(\bar{X}_t, x)\bar{\mu}_t(\mathrm{d}x)\Big)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t, \qquad \bar{\mu}_t = \mathrm{Law}(\bar{X}_t), \\ \bar{X}_0 = \bar{x} \end{cases} \tag{1.29}$$

with $\bar{x} \in \mathbb{R}^d$ and d-dimensional standard Brownian motion $(B_t)_{t\geq 0}$, and the mean-field particle system $(\{X_t^{i,N}\}_{i=1}^N)_{t\geq 0}$ with $N \in \mathbb{N}$ particles determined by

$$\begin{cases} \mathrm{d}X_t^{i,N} = \Big(b(X_t^{i,N}) + N^{-1}\sum_{j=1}^N \tilde{b}(X_t^{i,N}, X_t^{j,N})\Big)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t^i, \\ X_0^{i,N} = x^i, \end{cases} \qquad i = 1, \dots, N, \tag{1.30}$$

where $\{x^i\}_{i=1}^N \in \mathbb{R}^{Nd}$ and $\{(B_t)_{t\geq 0}\}_{i=1}^N$ are N independent d-dimensional standard Brownian motion. Here, $b : \mathbb{R}^d \to \mathbb{R}^d$ and $\tilde{b} : \mathbb{R}^{2d} \to \mathbb{R}^d$ are two locally Lipschitz continuous functions.

The notion *propagation of chaos* was formed by Kac, who was originally interested in describing mathematical rigorously the connection between the Boltzmann equation modelling a large system of interacting gas particles and the nonlinear Liouville equation [120]. Making use of the idea of linking particle systems with nonlinear equations, McKean studied a class of diffusions, which are unrelated to the Boltzmann theory, and observed that under specific assumptions the law of the particle system can be approximated by the law of the nonlinear equation in an appropriate sense [144].

Here, we concentrate on SDEs given by (1.29) and (1.30), where the drift consists of a confining force b and of a pairwise interaction force $\tilde{b}$. We remark that (1.29) forms the probabilistic description of the Fokker-Planck equation given by

$$\partial_t \rho_t(x) = -\nabla_x \cdot \Big(b(x)\rho_t(x) + \int_{\mathbb{R}^d} \tilde{b}(x, u)\bar{\mu}_t(\mathrm{d}u)\rho_t(x)\Big) + \Delta_x \rho_t(x),$$

where ρ_t denotes the time-dependent density function corresponding to the law $\bar{\mu}_t(\mathrm{d}x)$ of $\bar{X}_t$. We note that given $\bar{x}_0 \in \mathbb{R}^d$ and a d-dimensional standard Brownian motion $(B_t)_{t\geq 0}$ existence of a unique strong solution to (1.29) holds by [146, Theorem 2.2].

There are several ways to characterise propagation of chaos rigorously, see [45, Section 2.3] for an overview of different definitions. Here, we focus on describing the connection between (1.29) and (1.30) via probabilistic techniques and we follow the definition of chaos given in [45, Section 2.3.1], where a coupling between a solution to the mean-field model with N particles and N i.i.d. solutions to the nonlinear SDE (1.29) is considered and the trajectories of the coupling are compared.

Definition 1.3 (Propagation of chaos)**.** *Let ρ denote a distance on $\mathbb{R}^d$, let $p \in [1, \infty)$ and $T \in (0, \infty]$. We say that propagation of chaos holds if for all $N \in \mathbb{N}$, there exist a coupling $(\{X_t^{i,N}\}_{i=1}^N, \{\bar{X}_t\}_{i=1}^N)_{0\leq t\leq T}$ of a mean-field particle system $(\{X_t^{i,N}\}_{i=1}^N)_{0\leq t\leq T}$ driven by* (1.30) *with law $\mu_t^N = \mathrm{Law}(\{X_t\}_{i=1}^N)$ and of N independent processes $(\{\bar{X}_t\}_{i=1}^N)_{0\leq t\leq T}$ driven by* (1.29) *with law $\bar{\mu}_t = \mathrm{Law}(\bar{X}_t^i)$ for all $i = 1, ..., N$ and a constant $C_{(N,T)} > 0$ depending on N and T*

with $C_{(N,T)} \to 0$ *as* $N \to \infty$*, such that*

$$N^{-1}\sum_{i=1}^{N} E[\sup_{t\leq T}\rho(X_t^{i,N},\bar{X}_t^i)^p] \leq C_{(N,T)} \qquad \textit{(pathwise formulation)} \tag{1.31}$$

or

$$N^{-1}\sum_{i=1}^{N} \sup_{t\leq T} E[\rho(X_t^{i,N},\bar{X}_t^i)^p] \leq C_{(N,T)}. \qquad \textit{(pointwise formulation)} \tag{1.32}$$

We say uniform in time propagation of chaos holds if $C_{(N,T)}$ *is independent of* T *and* (1.31)*, respectively* (1.32)*, holds for all* $T \in (0,\infty]$.

We note that (1.31) implies (1.32), and (1.32) implies

$$\sup_{t\leq T}\mathcal{W}_\rho^p(\mu_t^N,\bar{\mu}_t^{\otimes N}) \leq C_{(N,T)} \to 0 \qquad \text{as } N\to\infty.$$

1.6.2 Propagation of chaos results

Next, we present some known propagation of chaos results. The difference of the trajectories of the particle system given by (1.30) with N copies of the process driven by the nonlinear SDE (1.29) was first studied by McKean [144] and Sznitman [187] by applying coupling methods. We adapt this propagation of chaos result with a fixed time-horizon presented in [45, Theorem 3.1] to the mean-field particle system given in (1.30).

Theorem 1.4 (Finite time propagation of chaos - McKean)**.** *Let* $b:\mathbb{R}^d\to\mathbb{R}^d$ *and* $\tilde{b}:\mathbb{R}^d\to\mathbb{R}^d$ *be Lipschitz continuous functions and let* $\tilde{b}$ *be bounded. Fix* $N\in\mathbb{N}$*. Let* $\{(B_t^i)_{t\geq 0}\}_{i=1}^N$ *be* N *independent* d*-dimensional standard Brownian motions. Let* $(\{\bar{X}_t^i\}_{i=1}^N,\{X_t^{i,N}\}_{i=1}^N)_{t\geq 0}$ *be the synchronous coupling of the mean-field particle system and* N *copies of the nonlinear stochastic differential equation, solving*

$$\begin{cases} \mathrm{d}\bar{X}_t^i = \Big(b(\bar{X}_t^i) + \int_{\mathbb{R}^d}\tilde{b}(\bar{X}_t^i - x)\bar{\mu}_t(\mathrm{d}x)\Big)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t^i, & \bar{\mu}_t = \mathrm{Law}(\bar{X}_t^i) \\ \mathrm{d}X_t^{i,N} = \Big(b(X_t^{i,N}) + N^{-1}\sum_{j=1}^N \tilde{b}(X_t^{i,N}-X_t^{j,N})\Big)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t^i \end{cases} \tag{1.33}$$

for $i=1,\dots N$*. Then for any* $T>0$*, propagation of chaos holds in the pathwise sense, i.e. there exists* $C_{(N,T)}>0$ *such that*

$$N^{-1}\sum_{i=1}^{N} E[\sup_{t\leq T}|X_t^{i,N}-\bar{X}_t^i|^2] \leq C_{(N,T)}.$$

The constant $C_{(N,T)}$ *is of order* $\mathcal{O}(N^{-1})$ *and grows exponentially fast in* T.

The original proof of this theorem by McKean is given in [144], see also [187] for an alternative proof. Extensions of McKean's finite propagation of chaos bounds are provided by Sznitman [187] and Méléard [146]. Relaxations of the Lipschitz condition of the drift including more singular drifts are addressed for instance in [23, 119, 159, 160]. In [108, 148, 149], uniform in time propagation of chaos estimates are established via analytic techniques.

Using probabilistic tools, uniform in time propagation of chaos estimates are established by Malrieu in [137] for the following framework:

Theorem 1.5 (Uniform in time propagation of chaos)**.** *Fix $N \in \mathbb{N}$. Let the confining force b be of the form $b = -\nabla V$ for some potential $V \in \mathcal{C}^2(\mathbb{R}^d)$, locally Lipschitz continuous and satisfy Assumption 1.1, i.e. V is a κ-strongly convex confinement potential. Let the interaction force $\tilde{b}$ be of the form $\tilde{b}(x,y) = -\nabla W(x-y)$ for some potential $W \in \mathcal{C}^2(\mathbb{R}^d)$ that is symmetric and convex. Moreover, ∇W is locally Lipschitz continuous and there exists $q \geq 1$ such that ∇W has polynomial growth of order q. Let $(\{\bar{X}_t^i\}_{i=1}^N, \{X_t^{i,N}\}_{i=1}^N)_{t\geq 0}$ be the solution to the synchronous coupling* (1.33)*, where the initial law $\bar{\mu}_0$ has bounded moments of order $2q$. Then, there exists a constant $C > 0$, depending on κ, such that*

$$\sup_{t\geq 0} N^{-1} \sum_{i=1}^N E[|X_t^{i,N} - \bar{X}_t^i|^2] \leq CN^{-1}.$$

A proof is given in [45, Section 3.1.3] and relies on a synchronous coupling, on applying the strong convexity of V and the convexity of W and on a uniform second moment bound for the law $\bar{\mu}_t$. Extensions to non-strongly convex confinement potentials are studied in [44]. In [75], a coupling approach involving a reflection coupling is applied to prove uniform in time propagation of chaos for non-strongly convex confinement potentials.

Let us generalise the idea of [75] how uniform in time propagation of chaos can be established if uniform in time second moment bounds hold for the solution to the nonlinear SDE (1.29) and if contraction in an L^1 Wasserstein distance holds for the mean-field particle system (1.30) with Lipschitz continuous interaction force $\tilde{b}$. We assume that there exists a coupling $(\{X_t^{i,N}\}_{i=1}^N, \{Y_t^{i,N}\}_{i=1}^N)_{t\geq 0}$ of two solutions to (1.30) and a distance function $\rho_N : \mathbb{R}^{Nd} \times \mathbb{R}^{Nd} \to [0,\infty)$ of the form $\rho_N(\{x^i\}_{i=1}^N, \{y^i\}_{i=1}^N) = N^{-1}\sum_{i=1}^N f(|x^i - y^i|)$ for some increasing, concave function f with $f(0) = 0$ and $f'(0) = 1$ such that

$$\mathrm{d}\rho_N(\{X_t^{i,N}\}_{i=1}^N, \{Y_t^{i,N}\}_{i=1}^N) \leq -c\rho_N(\{X_t^{i,N}\}_{i=1}^N, \{Y_t^{i,N}\}_{i=1}^N)\mathrm{d}t + \mathrm{d}M_t,$$

where $c > 0$ is a positive constant independent of N and $(M_t)_{t\geq 0}$ is some martingale. Following the proof approach of Section 1.4.1, it holds that

$$\mathcal{W}_{\rho_N}(\mu_t^N, \nu_t^N) \leq e^{-ct}\mathcal{W}_{\rho_N}(\mu_0^{\otimes N}, \nu_0^{\otimes N}),$$

where $X_0^i \sim \mu_0$ and $Y_0^i \sim \nu_0$ for all $i = 1, \ldots, N$ and $\mu_t^N = \mathrm{Law}(\{X_t^i\}_{i=1}^N)$ and $\nu_t^N = \mathrm{Law}(\{Y_t^i\}_{i=1}^N)$. In this case, we can establish uniform in time propagation of chaos by considering the same coupling approach for a mean-field particle system $(\{X_t^{i,N}\}_{i=1}^N)_{t\geq 0}$ and N i.i.d. copies of solutions $(\{\bar{X}_t^i\}_{i=1}^N)_{t\geq 0}$ to (1.29). Then,

$$\mathrm{d}\rho_N(\{X_t^{i,N}\}_{i=1}^N, \{\bar{X}_t^i\}_{i=1}^N) \leq -c\rho_N(\{X_t^{i,N}\}_{i=1}^N, \{\bar{X}_t^i\}_{i=1}^N)\mathrm{d}t + N^{-1}\sum_{i=1}^N A_t^i \mathrm{d}t + \mathrm{d}M_t, \tag{1.34}$$

where $(M_t)_{t\geq 0}$ is some martingale and $(A_t^i)_{t\geq 0}$ $(i = 1, \ldots, N)$ are adapted stochastic processes satisfying

$$A_t^i \leq \Big| \int_{\mathbb{R}^d} \tilde{b}(\bar{X}_t^i, x)\bar{\mu}_t(\mathrm{d}x) - N^{-1}\sum_{i=1}^N \tilde{b}(\bar{X}_t^i, \bar{X}_t^j)\Big|.$$

Given $\bar{X}_t^i$, we note that $\bar{X}_t^j$ $(j \neq i)$ are i.i.d. random variables with law $\bar{\mu}_t$. Then,

$$E\Big[\tilde{b}(\bar{X}_t^i, \bar{X}_t^j)|\bar{X}_t^i\Big] = \int_{\mathbb{R}^d} \tilde{b}(\bar{X}_t^i, x)\bar{\mu}_t(\mathrm{d}x)$$

and

$$E\Big[|\int_{\mathbb{R}^d}\tilde{b}(\bar{X}_t^i,x)\bar{\mu}_t(\mathrm{d}x)-N^{-1}\sum_{i=1}^N\tilde{b}(\bar{X}_t^i,\bar{X}_t^j)|^2\Big|\bar{X}_t^i\Big]$$
$$=\frac{N-1}{N^2}\mathrm{Var}_{\bar{\mu}_t}(\tilde{b}(\bar{X}_t^i,\cdot))+\frac{1}{N^2}E\Big[|\int_{\mathbb{R}^d}\tilde{b}(\bar{X}_t^i,x)\bar{\mu}_t(\mathrm{d}x)-\tilde{b}(\bar{X}_t^i,\bar{X}_t^i)|^2\Big|\bar{X}_t^i\Big]$$
$$+\frac{2}{N^2}\sum_{j=1,j\neq i}^N E\Big[|\int_{\mathbb{R}^d}\tilde{b}(\bar{X}_t^i,x)\bar{\mu}_t(\mathrm{d}x)-\tilde{b}(\bar{X}_t^i,\bar{X}_t^j)|\cdot|\int_{\mathbb{R}^d}\tilde{b}(\bar{X}_t^i,x)\bar{\mu}_t(\mathrm{d}x)-\tilde{b}(\bar{X}_t^i,\bar{X}_t^i)|\Big|\bar{X}_t^i\Big].$$

Hence, by Lipschitz continuity of $\tilde{b}$ and Young's inequality,

$$E\Big[|\int_{\mathbb{R}^d}\tilde{b}(\bar{X}_t^i,x)\bar{\mu}_t(\mathrm{d}x)-N^{-1}\sum_{i=1}^N\tilde{b}(\bar{X}_t^i,\bar{X}_t^j)|^2\Big]$$
$$\leq\frac{4\|\tilde{b}\|_{\mathrm{Lip}}^2}{N}\int_{\mathbb{R}^d}|x|^2\bar{\mu}_t(\mathrm{d}x)+\frac{4\|\tilde{b}\|_{\mathrm{Lip}}^2}{N^2}\int_{\mathbb{R}^d}|x|^2\bar{\mu}_t(\mathrm{d}x)+\frac{8\|\tilde{b}\|_{\mathrm{Lip}}^2}{N}\int_{\mathbb{R}^d}|x|^2\bar{\mu}_t(\mathrm{d}x).$$

Then by Jensen's inequality and the uniform in time second moment bound of the solution to (1.29), there exists $C>0$ such that

$$E[A_t^i]\leq\frac{4\|\tilde{b}\|_{\mathrm{Lip}}}{N^{1/2}}\Big(\int_{\mathbb{R}^d}|x|^2\bar{\mu}_t(\mathrm{d}x)\Big)^{1/2}\leq CN^{-1/2}.$$

Taking expectation in (1.34), inserting the upper bound and applying Grönwall's inequality yields

$$E[\rho_N(\{X_t^{i,N}\}_{i=1}^N,\{\bar{X}_t^i\}_{i=1}^N)]\leq e^{-ct}E[\rho(\{X_0^{i,N}\}_{i=1}^N,\{\bar{X}_0^i\}_{i=1}^N)]+CN^{-1/2}c^{-1}.$$

If for some probability measure μ_0 on $\mathbb{R}^d$ with finite second moment $X_0^{i,N},\bar{X}_0^i\sim\mu_0$ for all $i=1,\dots,N$,

$$N^{-1}\sum_{i=1}^N\sup_{t\geq0}E[f(|X_t^{i,N}-\bar{X}_t^i|)]\leq CN^{-1/2}c^{-1}$$

and hence uniform in time propagation of chaos holds in the pointwise sense.

This approach to show propagation of chaos is based on the results of [75]. Note that compared to the calculations in [75], the process A_t^i is controlled differently and the assumption $\tilde{b}(x,y)=\hat{b}(x-y)$ with $\hat{b}(0)=0$ for some Lipschitz continuous function $\hat{b}$ is not required here. A modification of the stated calculations is applied in [181], see Appendix C. The statements of this section up to and including the propagation of chaos result by Malrieu are based on the recent reviews [45, 46], where a detailed summary of the historical development and statements of propagation of chaos is given, see also the references therein.

1.6.3 Application to Deep Learning

The concept of propagation of chaos is applicable in the analysis of training a one-hidden layer neural network. We are interested in finding a function $f:\mathbb{R}^{d-1}\to\mathbb{R}$ such that for given input data $z=(z_1,\dots,z_{d-1})\in\mathbb{R}^{d-1}$ and output data $y\in\mathbb{R}$, $f(z)$ provides a good approximation of y. In the case of a one-hidden layer neural network, we consider f to be of the form $f(z)=$

$\frac{1}{N}\sum_{i=1}^N \beta_{i,N}\phi(\alpha_{N,i}\cdot z)$, where N represents the number of neurons in the hidden layer, $\phi:\mathbb{R}\to\mathbb{R}$ is a bounded, continuous, non-constant activation function. A typical example for ϕ is the sigmoid function $\phi(r)=1/(1+e^{-r})$ [178]. We are looking for optimal parameters $\alpha_{N,i}\in\mathbb{R}^{d-1}$ and $\beta_{i,N}\in\mathbb{R}$, $i=1,..,N$. More precisely, we are intersted in solving the non-convex optimisation problem

$$\min_{\alpha_{i,N},\beta_{i,N}}\Big\{\int_{\mathbb{R}\times\mathbb{R}^{d-1}}\Big|y-\frac{1}{N}\sum_{i=1}^N\beta_{i,N}\phi(\alpha_{i,N}\cdot z)\Big|^2\nu(\mathrm{d}y\mathrm{d}z)\Big\},$$

where ν is the measure with compact support of the data $(y,z)\in\mathbb{R}\times\mathbb{R}^{d-1}$. We denote by $\mu^N=\frac{1}{N}\sum_{i=1}^N\delta_{\beta_{i,N},\alpha_{i,N}}$ the empirical law of the parameters and we rewrite f to $f(z)=\int_{\mathbb{R}^d}\beta\phi(\alpha\cdot z)\mu^N(\mathrm{d}\beta\mathrm{d}\alpha)=\int_{\mathbb{R}^d}\varphi(x,z)\mu^N(\mathrm{d}x)$, where $x=(\alpha,\beta)$ and $\varphi(x,z)=\beta\phi(\alpha\cdot z)$. Then, instead of the non-convex minimisation problem on the finite dimensional parameter space, we consider the minimisation problem over probability distributions on $\mathbb{R}^d$

$$\bar{\mu}=\mathrm{argmin}_\mu\Big\{\int_{\mathbb{R}^d}|y-\int_{\mathbb{R}^d}\varphi(x,z)\mu(\mathrm{d}x)|^2\nu(\mathrm{d}y\mathrm{d}z)+H(\mu|\mathcal{N}(0,I_d))\Big\},$$

where an additional regulariser $H(\mu|\mathcal{N}(0,I_d))$ is considered, which is given by the relative entropy with respect the normal distribution $\mathcal{N}(0,I_d)$. By [114], this is a convex minimisation problem and the minimiser is given by

$$\bar{\mu}(\mathrm{d}x)\propto\exp\Big(-\frac{1}{2}|x|^2-\int_{\mathbb{R}^d}2\varphi(x,z)\Big(-y+\int_{\mathbb{R}^d}\varphi(\tilde{x},z)\bar{\mu}(\mathrm{d}\tilde{x})\Big)\nu(\mathrm{d}y\mathrm{d}z)\Big)\mathrm{d}x.$$

We note that $\bar{\mu}$ is an invariant probability measure of the nonlinear SDE (1.29) with

$$b(x)=-x-2\int_{\mathbb{R}^d}\nabla_x\varphi(x,z)y\nu(\mathrm{d}y\mathrm{d}z)\quad\text{and}\quad\tilde{b}(x,\bar{x})=2\int_{\mathbb{R}^d}\nabla_x\varphi(x,z)\nabla_{\bar{x}}\varphi(\bar{x},z)\nu(\mathrm{d}y\mathrm{d}z).$$

Further, we observe that given data points $(y,z)=(y,(z_1,\dots z_{d-1}))\in\mathbb{R}^d$ distributed according to ν, the Euler discretisation of the mean-field particle system (1.30) given by

$$\mathbf{X}^i_{k+1}=\mathbf{X}^i_k-h\Big(\mathbf{X}^i_k+2\nabla_x\varphi(\mathbf{X}^i_k,z)y-\frac{1}{N}\sum_{j=1}^N2\nabla_x\varphi(\mathbf{X}^i_k,z)\nabla_x\varphi(\mathbf{X}^j_k,z)\Big)+\sqrt{2h}\xi_{k+1}$$

with step size $h>0$ and $\xi_k\sim\mathcal{N}(0,I_d)$ provides a sampling algorithm of the distribution of the parameters $X_{i,N}=(\beta_{i,N},\alpha_{i,N})$. Therefore, if uniform in time propagation of chaos bounds, quantitative bounds on the contraction behaviour of the nonlinear SDE and error bounds on the disretisation scheme are given, we obtain quantitative estimates for the sampling behaviour.

The description of the application to deep learning relies on [114], see also [145, 178] for further information.

2

Outline of Projects

2.1 Outline of Project A

In the first project, we establish contraction with dimension-free rates for unadjusted HMC for mean-field models using a particlewise coupling approach. The results were distributed as a research article on the online-portal ArXiv:

N. Bou-Rabee and K. Schuh. Convergence of unadjusted Hamiltonian Monte Carlo for mean-field models, ArXiv preprint 2009.08735, September 2020.

The article is a joint work with Nawaf Bou-Rabee (Rutgers University Camden). Appendix A contains the article as stated on the online portal ArXiv. This subsection presents the main subject and results of the article. The precise statements, proofs and the context of the existing literature are given in Appendix A.

We consider the unadjusted Hamiltonian Monte Carlo method to sample the probability distribution $\mu(\mathrm{d}x) \propto \exp(-U(x))\mathrm{d}x$ on $\mathbb{R}^{nd}$ where the twice differentiable potential $U : \mathbb{R}^{nd} \to \mathbb{R}$ is of mean-field particle type given by

$$U(x) = \sum_{i=1}^{n} \Big(V(x^i) + \frac{\epsilon}{n} \sum_{j=1}^{n} W(x^i - x^j) \Big).$$

The function V denotes the unary potential, W the pairwise interaction potential and the positive constant $\epsilon > 0$ the interaction parameter. The parameter n corresponds to the number of particles and d to the dimension of one particle. Generating samples of the desired measure μ plays an essential role in understanding statistical properties of high-dimensional models which have applications in many areas such as chemical physics, material science and deep learning.

Unadjusted HMC generates a Markov chain on $\mathbb{R}^{nd}$ where the transition step of one uHMC step is given by $\mathbf{X}'(x) = q_T(x, \xi)$, where $q_T(x, \xi)$ denotes the position at duration time $T > 0$ of the velocity Verlet discretisation scheme of the Hamiltonian dynamics,

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t} q_t^i = p_{\lfloor t \rfloor_h}^i - \frac{h}{2} \nabla U_i(q_{\lfloor t \rfloor_h}) \\ \frac{\mathrm{d}}{\mathrm{d}t} p_t^i = -\frac{1}{2}(\nabla_i U(q_{\lfloor t \rfloor_h}) + \nabla_i U(q_{\lceil t \rceil_h})), \end{cases} \qquad i = 1, \dots, n$$

with initial value $(x, \xi) \in \mathbb{R}^{2nd}$, where $\xi \sim \mathcal{N}(0, \mathrm{I}_{nd})$ is a standard normally distributed random variable. The functions $\lfloor \cdot \rfloor_h$ and $\lceil \cdot \rceil_h$ are given in (1.15). Its transition kernel π_h is given by $\pi_h(x, A) = P[\mathbf{X}'(x) \in A]$ for any $x \in \mathbb{R}^{nd}$ and any $A \in \mathcal{B}(\mathbb{R}^{nd})$.

We recall that compared to the Metropolis adjusted HMC introduced in Section 1.2.2, uHMC omits the acceptance-rejection step leaving the target measure only approximately invariant. Therefore, besides studying the long-time behaviour of uHMC and verifying the existence of a unique invariant measure μ_h of uHMC, one is interested in controlling the difference between the target measure and μ_h in an appropriate sense.

We impose the following conditions for the unary potential $V : \mathbb{R}^d \to \mathbb{R}$ and the interaction potential $W : \mathbb{R}^d \to \mathbb{R}$:

Assumption 2.1. *It holds $V(0) = 0$, $V(x) \geq 0$ for all $x \in \mathbb{R}^d$, and V is strongly convex outside a Euclidean ball, i.e. there exist $K \in (0, \infty)$ and $R \in [0, \infty)$ such that*

$$(x - y) \cdot (\nabla V(x) - \nabla V(y)) \geq K|x - y|^2 \qquad \textit{for all } x, y \in \mathbb{R}^d \textit{ such that } |x - y| \geq R.$$

Further, V and W have bounded second and third derivatives, i.e. there exist $L, \tilde{L}, L_H, \tilde{L}_H \in [0, \infty)$ such that

$$\sup \|\nabla^2 V\| \leq L, \quad \sup \|\nabla^2 W\| \leq \tilde{L}, \quad \sup \|\nabla^3 V\| \leq L_H \quad \textit{and} \quad \sup \|\nabla^3 W\| \leq \tilde{L}_H.$$

In Appendix A, we consider a particlewise coupling approach that adapts ideas of the coupling approach of [31]. The coupling transition step $(\mathbf{X}(x, y), \mathbf{Y}(x, y))$ defined by $\mathbf{X}(x, y) = q_T(x, \xi)$ and $\mathbf{Y}(x, y) = q_T(y, \eta)$ is characterised by the pair of random variables (ξ, η) defined on a common probability space satisfying $\mathrm{Law}(\xi) = \mathrm{Law}(\eta) = \mathcal{N}(0, \mathrm{I}_{nd})$. For the particlewise construction we consider $\xi \sim \mathcal{N}(0, \mathrm{I}_{nd})$ and construct η^i for each $i = 1, \ldots, n$ separately. Namely, if the distance $|x^i - y^i|$ is large, we consider a synchronous coupling $\eta^i = \xi^i$. If the distance is small, we set $\eta^i = \xi^i + \gamma(x^i - y^i)$ for some $0 < \gamma < T^{-1}$ with maximal probability. Otherwise, we apply a reflection coupling, $\eta^i = \xi^i - 2(e^i \cdot \xi^i)e^i$ with $e^i = (x^i - y^i)/|x^i - y^i|$ if $|x^i - y^i| > 0$. Indeed, this construction satisfies $\mathrm{Law}(\xi) = \mathrm{Law}(\eta) = \mathcal{N}(0, \mathrm{I}_{nd})$ and defines a coupling of two uHMC transition steps.

We remark that for a large distance $|x^i - y^i|$ the strong convexity property of the unary potential leads to contraction for the i-th component of the transition step provided the interaction is sufficiently small. The definition of the coupling for small distances is motivated by the free dynamics when $U \equiv 0$. In this case, it holds that $|\mathrm{X}^i(x, y) - \mathrm{Y}^i(x, y)| = |x^i - y^i + T(\xi^i - \eta^i)| = |x^i - y^i||1 - T\gamma|$ with maximal probability. By the boundedness of the second derivatives of V and W, which corresponds to the Lipschitz continuity of the gradients of V and W, the deviation from the free dynamics can be bounded in terms of L and $\tilde{L}$ and contraction is obtained for this choice of ξ^i and η^i.

Corresponding to the coupling, a concave function f is aligned that puts more weight on a decrease in distance than an increase. This function in combination with the contraction for $\eta^i = \xi^i + \gamma(x^i - y^i)$ compensates for the missing contraction in the case when a reflection coupling occurs. Then, we establish contraction on average of one HMC transition step with respect to the l^1-distance $\rho(x, y) = \sum_{i=1}^n f(|x^i - y^i|)$, i.e.

$$E\Big[\sum_{i=1}^n f(|\mathrm{X}^i(x, y) - \mathrm{Y}^i(x, y)|)\Big] \leq (1 - c) \sum_{i=1}^n f(|x^i - y^i|) \qquad \text{for all } x, y \in \mathbb{R}^{nd},$$

provided the duration time T and the discretisation step h are sufficiently small and the interaction parameter ϵ is sufficiently small compared to the convexity parameter K of the unary potential. From that we can deduce existence of a unique invariant measure μ_h of uHMC and exponential contraction to the measure μ_h in L^1 Wasserstein distance with respect to ρ, i.e. for all initial distributions ν on $\mathbb{R}^d$,

$$\mathcal{W}_\rho(\nu {\pi_h}^m, \mu_h) \leq e^{-cm} \mathcal{W}_\rho(\nu, \mu_h) \qquad \text{for any } m \in \mathbb{N}.$$

We remark that the constraints on T, h and ϵ are independent of the number of particles and depend only on L, K and R.

Via the synchronous coupling, we verify that the L^1 Wasserstein distance with respect to ρ of the distribution after an exact and an unadjusted HMC step is of order $\mathcal{O}(h^2)$. Applying this strong accuracy bound of the velocity Verlet integrator, the contraction result and the triangle trick (1.28), we establish bounds in an L^1 Wasserstein distance between the target measure μ and the invariant measure μ_h of the uHMC. This bound provides quantitative estimates for the discretisation step and for the number of HMC steps needed to reach the target distribution in L^1 Wasserstein distance up to a given error ε.

Additionally, given an observable $g \in \mathcal{C}_b^1(\mathbb{R}^{nd})$, we are interested in quantitative bounds between $\mu(g) = \int_{\mathbb{R}^{nd}} g(x)\mu(\mathrm{d}x)$ and the estimator $A_{m,b}g = \frac{1}{m}\sum_{k=b}^{m+b-1} g(X_k)$, where b denotes the burn-in time and m denotes the number of steps of the Markov chain considered for the ergodic average. Using the contraction result and the strong accuracy bounds, we bound the bias of the estimator by

$$|E_\nu[A_{m,b}g] - \mu(g)| \leq \frac{1}{m} \max_i \|\nabla_i g\|_\infty M \frac{e^{-cb}}{1-e^{-c}} \mathcal{W}_{l^1}(\nu, \mu_h) + h^2 \max_i \|\nabla_i g\|_\infty \mathrm{C}, \tag{2.1}$$

where ν is the initial distribution of the Markov chain and M is a constant relating the distance ρ and the distance l^1 given by $l^1(x,y) = \sum_{i=1}^n |x^i - y^i|$. The distance $\mathcal{W}_{l^1}(\nu, \mu_h)$ between the initial distribution and the invariant measure of uHMC is often of order n. Similarly, the constant C is linear in n. Hence, if g is an intensive observable of the form $g(x) = \frac{1}{n}\sum_{i=1}^n \tilde{g}(x^i)$ for some $\tilde{g} \in \mathcal{C}_b^1(\mathbb{R}^d)$, the bias can be bounded from above in (2.1) by a given constant $\varepsilon > 0$ by choosing m, b sufficiently large and h sufficiently small. This choice is independent of the number of particles n.

Contribution by the author of the thesis: The idea of studying HMC applied to mean-field models was given to me by my advisor Andreas Eberle. The theoretical contraction result was established by me. Initially, I obtained error bounds for uHMC via an inductive argument leading to similar bounds with the same dependence on the discretisation parameter as presented in the paper. My co-author, Nawaf Bou-Rabee, brought up the idea of directly using the triangle inequality trick given in (1.28) which shortens the proof and is presented in the current version of the paper. The technical details, the numerical simulations and the writing were done by me getting advice and assistance from Nawaf Bou-Rabee.

2.2 Outline of Project B

In the second project, we establish conditions under which contraction for solutions of non-linear SDEs of McKean-Vlasov type using a sticky coupling holds, and we study nonlinear one-dimensional sticky SDEs. The results have been distributed through a research article on the online-portal ArXiv:

A. Durmus, A. Eberle, A. Guillin and K. Schuh. Sticky nonlinear SDEs and convergence of McKean-Vlasov equations without confinement, ArXiv preprint 2201.07652, January 2022.

The article is a joint work with Alain Durmus (Université Paris-Saclay), Andreas Eberle (University of Bonn) and Arnaud Guillin (Université Blaise Pascal). The work is presented in Appendix B as given on the online portal ArXiv. This subsection presents the main subject of the article. The precise statements, proofs and a survey of the existing literature are given in Appendix B.

We are interested in the long-time behaviour of the solution $(\bar{X}_t)_{t\geq 0}$ of the nonlinear stochastic differential equation of McKean-Vlasov type without confinement given by

$$\mathrm{d}\bar{X}_t = \Big(\int_{\mathbb{R}^d} b(\bar{X}_t - x)\mu_t(\mathrm{d}x)\Big)\mathrm{d}t + \mathrm{d}B_t, \qquad \mu_t = \mathrm{Law}(\bar{X}_t), \tag{2.2}$$

where $(B_t)_{t\geq 0}$ is a d-dimensional standard Brownian motion and $b:\mathbb{R}^d\to\mathbb{R}^d$ is a Lipschitz continuous drift function.

We establish a new coupling approach for nonlinear SDEs to prove contraction for the process $(\bar{X}_t)_{t\geq 0}$. Before we introduce the coupling and state the main results, we specify the conditions we impose on the nonlinear SDE:

Assumption 2.2. *The function $b:\mathbb{R}^d\to\mathbb{R}^d$ is Lipschitz continuous and anti-symmetric. Further, there exist a constant $L\in(0,\infty)$, a bounded function $\gamma:\mathbb{R}^d\to\mathbb{R}^d$ and a Lipschitz continuous function $\kappa:[0,\infty)\to\mathbb{R}$ such that*

$$b(x) = -Lx + \gamma(x)$$

and the following three conditions are satisfied for $x,\bar{x},y\in\mathbb{R}^d$:

$$\langle x-y, \gamma(x-\bar{x})-\gamma(y-\bar{x})\rangle \leq \kappa(|x-y|)|x-y|^2, \qquad \limsup_{r\to\infty}(\kappa(r)-L)<0, \qquad \text{and}$$
$$\|\gamma\|_\infty \leq \mathrm{C}_{(L,\kappa)},$$

where the constant $\mathrm{C}_{(L,\kappa)}\in(0,\infty)$ depends on L and κ. Moreover, the initial distribution μ_0 has bounded fourth moments and is centred, i.e. $\int_{\mathbb{R}^d} x\mu(\mathrm{d}x)=0$.

We note that the condition on the initial distribution combined with the anti-symmetric drift implies that the solution of (2.2) is centred for all $t\geq 0$. This is crucial to guarantee convergence to equilibrium.

To analyse the long-time behaviour of the process $(\bar{X}_t)_{t\geq 0}$, a sticky coupling of two solutions of (2.2) with different initial conditions is constructed in the following way: If the two solutions coincide, the noise is synchronised in the nonlinear SDE and otherwise, a reflection coupling is considered. We note that compared to the SDE analysed in Section 1.4.1, the drifts of the two copies of the SDEs do not coincide when the solutions coincide, since their laws differ in general. Therefore, the solutions are driven apart again after they are coupled and this coupling construction leads to a sticky behaviour of the two solutions. We observe that the distance process of the two solutions is controlled by a process $(r_t)_{t\geq 0}$ solving the following nonlinear SDE on $\mathbb{R}$ with a sticky boundary at zero

$$\mathrm{d}r_t = \bar{b}(r_t)\mathrm{d}t + aP[r_t>0]\mathrm{d}t + 2\mathbb{1}_{(0,\infty)}(r_t)\mathrm{d}W_t,$$

where $(W_t)_{t\geq 0}$ is a one-dimensional standard Brownian motion, $\bar{b}(r) = (\kappa(r)-L)r$ and $a = 2\|\gamma\|_\infty$. This SDE belongs to a new class of nonlinear SDEs with sticky boundary at zero, which is analysed carefully. In particular, existence of a solution and uniqueness in law are shown by considering a family of solutions $\{(r_t^{n,m})_{t\geq 0}\}_{n,m\in\mathbb{N}}$ of approximating SDEs and by taking the limit in two steps. More precisely, the nonlinear drift term $aP[r_t > 0]$ and the diffusion term $2\mathbb{1}_{(0,\infty)}(r_t)$ are approximated in two steps. Further, we detect that the SDE exhibits a phase transition. More precisely, if a is sufficiently small compared to $\tilde{b}$ and the nonlinear term contributes only a little to the drift, the Dirac measure at zero is the unique invariant measure. In this case, if we start outside equilibrium, we notice that if time evolves, more mass gets stuck at zero and we establish exponential convergence to equilibrium for the one-dimensional nonlinear sticky SDE. As in Section 1.4.1, the proof relies on a concave function f that is aligned to the drift of the nonlinear sticky SDE and causes a decrease of the process to have a larger impact than an increase. In particular, we prove

$$E[f(r_t)] \leq e^{-ct} f(r_0),$$

where the rate c depends on $\tilde{b}$ and a. Then, using the contraction result for the solution of the one-dimensional nonlinear sticky SDE which bounds the difference process of the sticky coupling from above and using that the distance function $\rho(x, y) = f(|x - y|)$ is equivalent to the Euclidean distance, we can deduce contraction in L^1 Wasserstein distance for the nonlinear unconfined SDE, i.e.

$$\mathcal{W}^1(\mu_t, \nu_t) \leq M e^{-ct} \mathcal{W}^1(\mu_0, \nu_0),$$

where μ_t and ν_t are the laws of the two copies of the coupling and the contraction rate c and the constant M depend on L, γ through a and $\tilde{b}$.

Additionally, we establish uniform in time propagation of chaos bounds for the corresponding mean-field particle system given by

$$\mathrm{d}X_t^i = \frac{1}{N} \sum_{j=1}^{N} b(X_t^i - X_t^j)\mathrm{d}t + \mathrm{d}B_t^i, \qquad i = 1, ..., N. \tag{2.3}$$

with i.i.d. initial values $X_0^1, \dots X_0^N$ and N independent d-dimensional standard Brownian motions $\{(B_t^i)_{t\geq 0}\}_{i=1}^N$. We consider a componentwise sticky coupling of the mean-field system (2.3) with N particles and of N copies of (2.2) and observe analogously that the componentwise difference process is controlled by the process $(\{r_t^i\}_{i=1}^N)_{t\geq 0}$ solving a system of N sticky one-dimensional SDEs. For this system of sticky SDEs, we prove existence and uniqueness analogously. Then in the same spirit as in the nonlinear case, the observation that the componentwise difference process is controlled by the process $(\{r_t^i\}_{i=1}^N)_{t\geq 0}$ is used to provide uniform in time propagation of chaos estimates for the mean-field system.

Finally, we transfer the sticky coupling approach for nonlinear confined SDEs on $\mathbb{R}^d$ to SDEs on the one-dimensional torus $\mathbb{T} = \mathbb{R}/(2\pi)$ and establish bounds on the contraction rate for the Kuramoto model, where the drift b is of the form $b(z) = -k\sin(z)$ for some $k \in \mathbb{R}$.

Contribution by the author of the thesis: My co-authors, Alain Durmus, Arnaud Guillin and Andreas Eberle, conceived of the idea to study the long-time behaviour of unconfined nonlinear SDEs of McKean-Vlasov type via a sticky coupling. We developed the main theory together during several research visits. Afterwards, I worked out the technical details getting advice and support from my co-authors. In particular, the details of the analysis of sticky nonlinear SDEs were elaborated by myself getting assistance from my co-authors. The main body of the paper and the proofs are written by me assisted by Alain Durmus and Andreas Eberle.

2.3 Outline of Project C

In the third project, we study the long-time behaviour of the classical Langevin dynamics and the Langevin dynamics with distribution-dependent forces. The results are summarised in a research article that is available on the online-portal ArXiv:

K. Schuh. Global contractivity for Langevin dynamics with distribution-dependent forces and uniform in time propagation of chaos, ArXiv preprint arXiv:2206.03082, June 2022.

The work is presented in Appendix C as it is stated on the online portal ArXiv. Here, the main subject and the results of the article are summarised. Precise statements, proofs and references to linked literature are given in Appendix C.

Given a probability measure $\bar{\mu}_0$ on $\mathbb{R}^{2d}$ and a d-dimensional Brownian motion $(B_t)_{t\geq 0}$, we consider the diffusion $(\bar{X}_t, \bar{Y}_t)_{t\geq 0}$, which is given as a solution to the Langevin dynamics with nonlinear McKean-Vlasov interactions of the form

$$\begin{cases} \mathrm{d}\bar{X}_t = \bar{Y}_t \mathrm{d}t \\ \mathrm{d}\bar{Y}_t = (-\gamma \bar{Y}_t + ub(\bar{X}_t) + u\int_{\mathbb{R}^d} \tilde{b}(\bar{X}_t, z)\bar{\mu}_t^X(\mathrm{d}z))\mathrm{d}t + \sqrt{2\gamma u}\mathrm{d}B_t, \end{cases} \qquad (\bar{X}_0, \bar{Y}_0) \sim \bar{\mu}_0, \tag{2.4}$$

where $\bar{\mu}_t^x$ is the marginal distribution in the first component of $\mu_t = \mathrm{Law}(\bar{X}_t, \bar{Y}_t)$, $\gamma, u > 0$ are positive constants and $b : \mathbb{R}^d \to \mathbb{R}^d$ and $\tilde{b} : \mathbb{R}^{2d} \to \mathbb{R}^d$ are two Lipschitz continuous functions. We are interested in the long-time behaviour of solutions to (2.4) and of the classical Langevin diffusion given by (2.4) with interaction force satisfying $\tilde{b} \equiv 0$. We impose the following assumption on the external force $b : \mathbb{R}^d \to \mathbb{R}^d$ and the interaction force $\tilde{b} : \mathbb{R}^{2d} \to \mathbb{R}^d$:

Assumption 2.3. *The function $b : \mathbb{R}^d \to \mathbb{R}^d$ is Lipschitz continuous and there exist a positive definite matrix $K \in \mathbb{R}^{d\times d}$ with smallest eigenvalue $\kappa > 0$ and largest eigenvalue L_K, a positive constant $R \geq 0$ and a Lipschitz continuous function $g : \mathbb{R}^d \to \mathbb{R}^d$ such that*

$$\begin{aligned} &b(x) = -Kx + g(x) && \textit{for all } x \in \mathbb{R}^d \textit{ and} \\ &(g(x) - g(y)) \cdot (x - y) \leq 0 && \textit{for all } x, y \in \mathbb{R}^d \textit{ such that } |x - y| \geq R \end{aligned} \tag{2.5}$$

and the Lipschitz constant L_g of g satisfies

$$L_g u \gamma^{-2} < \frac{\kappa}{2L_g}.$$

The function $\tilde{b} : \mathbb{R}^{2d} \to \mathbb{R}^d$ is Lipschitz continuous with Lipschitz constant $\tilde{L}$ satisfying

$$\tilde{L} \leq C_{(\kappa, L_K, R, L_g, u, \gamma)},$$

where $C_{(\kappa, L_K, R, L_g, u, \gamma)}$ is an explicit constant depending on κ, L_K, R, L_g, u and γ.

Note that (2.5) is equivalent to Assumption 1.2. Hence, b is not restricted to gradients of strongly convex potentials. But it includes gradients of potentials that are only strongly convex outside a Euclidean ball with a radius depending on R. Moreover, the constraints on L_g and $\tilde{L}$ in Assumption 2.3 are independent of the dimension d.

In this work, global contractivity is shown by combining two contraction results with respect to two different metrics for large and small distances and by exploiting a carefully aligned coupling approach. Let $(\bar{X}_t, \bar{Y}_t, \bar{X}_t', \bar{Y}_t')_{t\geq 0}$ be the coupling of $(\bar{X}_t, \bar{Y}_t)_{t\geq 0}$ and $(\bar{X}_t', \bar{Y}_t')_{t\geq 0}$ which are both driven by (2.4) with two Brownian motions such that the noise is either synchronised or reflected at a certain hyperplane. On the one hand, we consider the process $r_l(t) = ((\bar{X}_t - \bar{X}_t')\cdot(A(\bar{X}_t-\bar{X}_t'))+(\bar{X}_t-\bar{X}_t')\cdot(B(\bar{Y}_t-\bar{Y}_t'))+(\bar{Y}_t-\bar{Y}_t')\cdot(C(\bar{Y}_t-\bar{Y}_t')))^{1/2}$ which relies on a twisted 2-norm structure and where $A, B, C \in \mathbb{R}^{d\times d}$ are positive definite matrices depending on γ, u and K. If $r_l(t)$ is sufficiently large and hence either $|\bar{X}_t-\bar{X}_t'|$ or $|\bar{Y}_t-\bar{Y}_t'|$ is large, either the condition Assumption 2.3 or the damping term in (2.4) leads to contraction for the process $r_l(t)$. On the other hand, the distance process $r_s(t) = \alpha|\bar{X}_t-\bar{X}_t'|+|\bar{X}_t-\bar{X}_t'+\gamma^{-1}(\bar{Y}_t-\bar{Y}_t')|$ is considered with $\alpha > 0$ depending on γ, u, L_K and L_g. In that case, the coupling approach of [85] is used, where a synchronous coupling is considered for $\bar{X}_t-\bar{X}_t'+\gamma^{-1}(\bar{Y}_t-\bar{Y}_t') = 0$ since in that case contraction is observed for the first part of $r_s(t)$ and the second part vanishes. Otherwise a reflection coupling is considered, which returns the process to the hyperplane $\bar{X}_t-\bar{X}_t'+\gamma^{-1}(\bar{Y}_t-\bar{Y}_t') = 0$. Then, contraction is shown by exploiting this coupling and a concave function f that leads to contraction for $f(r_s(t))$ in a similar way as in Section 1.4.1 if the process $r_s(t)$ is sufficiently small.

The two processes $r_l(t)$ and $r_s(t)$ are glued together to $\rho_t = f(r_s(t)\wedge(D_K+\epsilon r_l(t)))$ where D_K, ϵ are positive constants such that one can make use of the contraction result for $f(r_s(t))$ for small distances and the contraction result for $r_l(t)$ for large distances. We refer to Figure C.2 in Appendix C sketching the construction of ρ. Since ρ finally defines a metric that is equivalent to the Euclidean distance, we deduce contraction in Wasserstein distance, i.e.

$$\mathcal{W}^1(\bar{\mu}_t,\bar{\nu}_t) \le Me^{-ct}\mathcal{W}^1(\bar{\mu}_0,\bar{\nu}_0),$$

where $\bar{\mu}_t$ and $\bar{\nu}_t$ are the laws of $(\bar{X}_t,\bar{Y}_t)_{t\ge0}$ and $(\bar{X}_t',\bar{Y}_t')_{t\ge0}$, respectively, and the constant c is a dimension-free contraction constant depending on κ, L_g, γ and R. The constant M is the quotient of the constants determining the equivalence between ρ and the Euclidean distance. This result holds provided $L_g u\gamma^{-2} < \kappa/(2L_g)$ is satisfied and $\tilde{L}$ is sufficiently small.

In particular for $R = 0$, the metric ρ reduces to r_l and global contraction with rate $c = \min(\gamma/16, \kappa u\gamma^{-1}/4 - L_g^3u^2\gamma^{-3}/2)$ is established. If the external force is additionally of gradient type and $\tilde{b} = 0$, we obtain the framework considered in Section 1.4.2. Then, the use of the co-coercivity property relaxes the restriction on γ to $L_g^2\gamma u \le 3/4$ and gives contraction in an L^p Wasserstein distance for $1 \le p < \infty$ with improved contraction rate $c = \min(\gamma/8, \kappa u\gamma^{-1}/2)$.

Moreover, for fixed $N \in \mathbb{N}$, we consider the corresponding particle system with N particles given by

$$\begin{cases} \mathrm{d}X_t^i = Y_t^i\mathrm{d}t \\ \mathrm{d}Y_t^i = \Big(-\gamma Y_t^i + ub(X_t^i) + uN^{-1}\sum_{j=1}^N \tilde{b}(X_t^i, X_t^j)\Big)\mathrm{d}t + \sqrt{2\gamma u}\mathrm{d}B_t^i, \end{cases} \qquad i = 1, ..., N, \tag{2.6}$$

with $(X_0^i, Y_0^i) \sim \mu_0$ for $i = 1, \ldots, N$ for some probability measure μ_0 on $\mathbb{R}^{2d}$. Applying a componentwise version of the coupling and an averaged l^1 distance of the form $\rho_N((x,y),(\bar{x},\bar{y})) = N^{-1}\sum_{i=1}^N \rho((x^i,y^i),(\bar{x}^i,\bar{y}^i))$ for $(x,y),(\bar{x},\bar{y}) \in \mathbb{R}^{2Nd}$, we prove propagation of chaos for the corresponding particle system, i.e. there exists a constant $C \in [0,\infty)$ such that for any $N \in \mathbb{N}$,

$$\sup_{t\ge0}\mathcal{W}_{l_N^1}(\bar{\mu}_t^{\otimes N}, \mu_t^N) \le CN^{-1/2}, \tag{2.7}$$

where $l_N^1((x,y),(x',y')) = N^{-1}\sum_{i=1}^N(|x-x'|+|y-y'|)$. Here, $\bar{\mu}_t^{\otimes N}$ denotes the product law of N copies $(\bar{X}_t,\bar{Y}_t)$ driven by (2.4) with initial distribution $\bar{\mu}_0 = \mu_0$, and μ_t^N denotes the law of $(X_t^i, Y_t^i)_{i=1}^N$ driven by (2.6) with initial distribution $\mu_0^N = \mu_0^{\otimes N}$. The constant C depends on the contraction rate c, on properties of the drift and on the second moment of the initial distribution μ_0.

Eventually, we adapt the construction of the distance function to study certain unconfined dynamics ($b \equiv 0$), where the interaction force is of the form

$$\tilde{b}(x,y) = \tilde{K}(x-y) + \tilde{g}(x,y) \qquad \text{for all } x,y \in \mathbb{R}^d,$$

where $\tilde{K} \in \mathbb{R}^{d\times d}$ is a positive definite matrix with smallest eigenvalue $\tilde{\kappa}$ and $\tilde{g} : \mathbb{R}^d \to \mathbb{R}^d$ is an anti-symmetric Lipschitz continuous function with Lipschitz constant $L_{\tilde{g}}$ satisfying $L_{\tilde{g}} \leq \tilde{\kappa}/8$. Then using a synchronous coupling, contraction in an L^1 and L^2 Wasserstein distance is shown and uniform in time propagation of chaos bounds are provided for the corresponding recentred mean-field particle system.

3

Conclusion

The thesis addressed the long-time behaviour of specific types of stochastic differential equations and Markov Chain Monte Carlo methods. Understanding especially the behaviour for high-dimensional frameworks and for invariant probability measures that are not log-concave is of wide interest since sampling of high-dimensional probability measures has many applications in various areas. Therefore, finding a sampling algorithm that produces good samples in a reasonable number of steps and that is additionally numerically implementable is an important task. Furthermore, stochastic processes are used to describe dynamics appearing, for instance, in physics and biology, and therefore, understanding their long-time behaviour is of relevance.

In this work, we concentrated on the mean-field particle model, where the potential of the target measure consists of a unary potential for each particle, that is not necessarily strongly convex, and of pairwise weak interaction potentials having Lipschitz continuous gradients. Correspondingly, we considered the dynamics of many particles moving according to an external force and a pairwise interaction force. We studied three stochastic processes in detail. First, we looked at the unadjusted Hamiltonian Monte Carlo method applied to mean-field models, which forms a numerically implementable sampling algorithm. Second, we studied first-order unconfined SDEs with McKean-Vlasov interaction forces. Finally, we investigated the Langevin dynamics forming a second-order SDE with McKean-Vlasov interactions. For the two latter processes, we were particularly interested in the corresponding nonlinear SDE and the connection between the nonlinear SDE and the mean-field particle system.

In all three projects, we established conditions under which we proved contraction in an L^1 Wasserstein distance with dimension-free rates via a coupling approach. More precisely, if $(p_t)_{t\in I}$ denotes the transition functions corresponding either to the time-discrete HMC method ($I=\mathbb{N}$) or to the time-continuous nonlinear diffusions ($I=\mathbb{R}_+$), then for any two initial distributions μ and ν satisfying the conditions imposed in the respective framework, we showed

$$\mathcal{W}^1(\mu p_t,\nu p_t)\leq Me^{-ct}\mathcal{W}^1(\mu,\nu)\qquad \text{for all } t\in I,$$

where the contraction rate c is dimension-free and M is a constant measuring the distance between the Euclidean distance and the distance function which is specifically aligned to the coupling approach. If we consider uHMC applied to the mean-field particle model with n particles, the l^1-distance is taken instead of the Euclidean distance.

In the following, let us highlight the precise contributions of the three individual projects:

Contribution of Project A: As mentioned above, our first contribution is the global contraction result in an L^1 Wasserstein distance for unadjusted HMC to its invariant measure with dimension-free rates. Since we considered a particlewise distance and a complementary particlewise coupling, the rate c and the constant M are independent of the number of particles. Furthermore, we established a bound on the distance between the invariant measure of uHMC and the desired target measure by establishing strong accuracy bounds for each uHMC step and using a triangle inequality trick. Combining the bound of the distance between the two probability measures and the contraction result for uHMC, we deduced a quantitative estimate of the number of uHMC steps needed to sample a probability measure whose difference in L^1 Wasserstein distance to the target measure is smaller than a given constant. We observed that the discretisation parameter of order $\mathcal{O}((dn)^{-1/2})$, where n denotes the number of particles and d is the dimension of each particle, guarantees that this bound does not degenerate. For this estimate, a warm start for the initial distribution is not required. Further, for the mean-field particle models, we showed that given an intensive observable such as the energy per particle, we can choose the discretisation parameter h and the number of steps independent of the number of particles to prove that the bias of the ergodic average is smaller than a given value. We supported the theoretical contraction result with numerical simulations.

Contribution of Project B: The contribution of the second work is threefold. Firstly, we established a contraction result in Wasserstein distance for nonlinear unconfined SDEs of McKean-Vlasov type, where the interaction force consists of a linear part and a bounded Lipschitz continuous function. The proof approach was based on introducing a sticky coupling for the nonlinear unconfined SDEs.

Secondly, in the analysis of the sticky coupling, we dealt with a class of one-dimensional nonlinear SDEs with a sticky boundary behaviour at 0. For this novel class of SDEs, we studied existence of a weak solution and uniqueness in law and established a result that provides a comparison between two solutions of one-dimensional nonlinear sticky SDEs with the same initial conditions and different drift functions. Additionally, we exhibited a phase transition for the appearance of multiple invariant probability measures and provided criteria for when a unique invariant measure exists and when the process converges to it in Wasserstein distance.

Thirdly, using a particlewise adaptation of the sticky coupling we gave uniform in time propagation of chaos bounds for the corresponding mean-field particle system.

Contribution of Project C: In the third project, we established a new approach to prove contraction in L^1 Wasserstein distance for the second-order Langevin dynamics. The proof relied on a novel construction of the underlying distance function that combines contraction results for different areas with respect to different distances and on aligning the coupling corresponding to the different areas. Via this construction, we improved existing contraction results by proving global contraction with dimension-free rates for non-strongly convex potentials. The results are further carried over to more general forces of non-gradient type. Additionally, this approach is applicable to provide contraction also for Langevin dynamics with nonlinear McKean-Vlasov interaction and to establish uniform in time propagation of chaos bounds for the corresponding particle system using a particlewise adaptation of the coupling and the distance function. Moreover, using the considered distance function for large distances, we proved global contraction in L^p Wasserstein distance ($1 \leq p < \infty$) with a contraction rate of order $\mathcal{O}(\sqrt{\kappa})$ for κ-strongly convex potentials, for which the deviation of a quadratic potential is of order $\mathcal{O}(\kappa)$. Finally, the approach is used to provide contraction in Wasserstein distance for certain unconfined Langevin dynamics with McKean-Vlasov interaction forces that form a small perturbation of a linear function.

Summarising the contributions of the individual project, on the one hand, this thesis gives a clear picture under which conditions contraction in L^1 Wasserstein distance holds for certain scenarios (i.e. for uHMC, the Langevin dynamics and nonlinear unconfined first-order SDEs). On the other hand, it also provides a couple of tools and techniques that may be relevant for the analysis of related problems.

First, the error bound analysis presented in the first project, where the difference between the invariant measure of the unadjusted measure and the true target measure is compared to the one-step error of the unadjusted method using a triangle trick and the contraction result, does not require a warm start for the initial distribution. Further, it provides a tool to analyse other unadjusted sampling methods (see also [73]).

The sticky coupling approach introduced in [87] and applied to nonlinear SDE in the second project provides an important tool to analyse and compare the long-time behaviour of SDEs with different drift and for which it is not guaranteed that the realisations stay together after they are coupled.

Eventually, the idea of combining two metrics, for which only partial contraction results are known, and constructing a new distance function as in the third project, can be transferred to other scenarios to obtain global contractivity there. We stress that this distance function benefits from avoiding a Lyapunov function in its construction and yields global contraction with dimension-free rates.

Outlook and open questions: Let us conclude by stating and discussing several questions and open problems that have arisen in the development of the thesis and which may be object of future research work.

In the first project, we focused mainly on the dimension-dependence in the choice of the discretisation parameter and the number of steps needed to obtain a good sample. Additionally, we are interested in working out the precise dependence on the condition number for strongly convex potentials. Particularly, we want to know how the optimal dependence on the condition number of uHMC compares with the dependence on the condition number of other sampling methods.

In addition, similarly to the nonlinear SDEs, one can ask whether we can make sense of a nonlinear HMC method and state analogously contraction bounds for the nonlinear HMC and uniform in time propagation of chaos results for HMC for mean-field models. In this case, target measures that are invariant to the transition kernel of nonlinear HMC can be approximated by uHMC applied to mean-field models. Indeed, we can prove propagation of chaos bounds which will be studied in detail in a future work.

It is essential in our analysis of nonlinear unconfined first order SDEs via sticky couplings that for the interaction forces the perturbation of the linear part is restricted to bounded Lipschitz continuous functions. Since the bound on the perturbation function is quite restrictive, it is of interest whether the contraction and propagation of chaos results can be extended to more general interaction forces.

For the one-dimensional nonlinear sticky SDE, we observed a phase transition from the existence of a unique invariant measure to the existence of multiple invariant measures. In the case of multiple invariant measures, we are interested in understanding better the behaviour of the nonlinear sticky SDE.

In the third project, for the classical Langevin dynamics, we established the contraction result in Wasserstein distance with a rate of order $\mathcal{O}(\sqrt{\kappa})$ for certain κ-strongly convex potentials that are not quadratic but form a small perturbation of a quadratic function. Since in [40] the order $\mathcal{O}(\sqrt{\kappa})$ for the optimal rate is proven for contraction in L^2 distance for all strongly convex potential via a Poincaré-type inequality, the natural question of whether using a coupling approach allows relaxing the assumption on the potential and obtaining a rate of the same order for general strongly convex potentials arises. This does not seem possible with the construction of the distance function considered here. Therefore, we wonder whether and how the distance function must be modified to obtain the desired order for the contraction rate.

Additionally, in our approach for both the strongly convex case and the more general case, there are restrictions on γ, which do not allow to take $\gamma \to 0$. Since for the quadratic potential the conditions disappear, we wonder whether it is possible to get rid of the constraint on γ and to obtain a contraction result in Wasserstein distance for the underdamped case via a coupling approach.

Last but not least, all results in our projects are restricted to Lipschitz continuous interaction forces. This restriction is an essential condition for the coupling approaches considered here. Since the interaction potentials in mean-field models are often singular in practice (as the Coulomb potential and the Newtonian potential), the question arises whether the condition on the Lipschitz condition of the interaction force can be removed, and whether and under which conditions contraction and uniform in time propagation of chaos bounds can still be shown via a coupling approach.

Bibliography

[1] Assyr Abdulle, Gilles Vilmart, and Konstantinos C. Zygalakis. High order numerical approximation of the invariant measure of ergodic SDEs. *SIAM J. Numer. Anal.*, 52(4):1600–1622, 2014.

[2] Assyr Abdulle, Gilles Vilmart, and Konstantinos C. Zygalakis. Long time accuracy of Lie-Trotter splitting methods for Langevin dynamics. *SIAM J. Numer. Anal.*, 53(1):1–16, 2015.

[3] Juan Acebron, Luis Bonilla, Conrado Pérez-Vicente, Fèlix Farran, and Renato Spigler. The Kuramoto model: A simple paradigm for synchronization phenomena. *Reviews of Modern Physics*, 77, 04 2005.

[4] Franz Achleitner, Anton Arnold, and Dominik Stürzer. Large-time behavior in non-symmetric Fokker-Planck equations. *Riv. Math. Univ. Parma (N.S.)*, 6(1):1–68, 2015.

[5] Elena Akhmatskaya and Sebastian Reich. GSHMC: an efficient method for molecular simulation. *J. Comput. Phys.*, 227(10):4934–4954, 2008.

[6] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids*. Clarendon Press, 1987.

[7] Adriano Amarante, Guedmiller Oliveira, Jéssica Ierich, Richard Cunha, Luiz Freitas, Eduardo Franca, and Fabio Leite. *Molecular Modeling Applied to Nanobiosystems*, pages 179–220. 12 2017.

[8] Hans Andersen. Molecular dynamics simulation at constant pressure and/or temperature. *J. Chem. Phys.*, 72:2384–2393, 02 1980.

[9] D. Bakry and Michel Émery. Diffusions hypercontractives. In *Séminaire de probabilités, XIX, 1983/84*, volume 1123 of *Lecture Notes in Math.*, pages 177–206. Springer, Berlin, 1985.

[10] Dominique Bakry, Patrick Cattiaux, and Arnaud Guillin. Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré. *J. Funct. Anal.*, 254(3):727–759, 2008.

[11] Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Cham, 2014.

[12] Richard F. Bass. A stochastic differential equation with a sticky point. *Electron. J. Probab.*, 19:no. 32, 22, 2014.

[13] D. Benedetto, E. Caglioti, J. A. Carrillo, and M. Pulvirenti. A non-Maxwellian steady distribution for one-dimensional granular media. *J. Statist. Phys.*, 91(5-6):979–990, 1998.

[14] Lorenzo Bertini, Giambattista Giacomin, and Khashayar Pakdaman. Dynamical aspects of mean field plane rotators and the Kuramoto model. *J. Stat. Phys.*, 138(1-3):270–290, 2010.

[15] Lorenzo Bertini, Giambattista Giacomin, and Christophe Poquet. Synchronization and random long time dynamics for mean-field plane rotators. *Probab. Theory Related Fields*, 160(3-4):593–653, 2014.

[16] A. Beskos, F. J. Pinski, J. M. Sanz-Serna, and A. M. Stuart. Hybrid Monte Carlo on Hilbert spaces. *Stochastic Process. Appl.*, 121(10):2201–2230, 2011.

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

[18] Alexandros Beskos, Gareth Roberts, and Andrew Stuart. Optimal scalings for local Metropolis-Hastings chains on nonproduct targets in high dimensions. *Ann. Appl. Probab.*, 19(3):863–898, 2009.

[19] Joris Bierkens, Paul Fearnhead, and Gareth Roberts. The zig-zag process and super-efficient sampling for Bayesian analysis of big data. *Ann. Statist.*, 47(3):1288–1320, 2019.

[20] Patrick Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

[21] Sergio Blanes, Fernando Casas, and J. M. Sanz-Serna. Numerical integrators for the hybrid Monte Carlo method. *SIAM J. Sci. Comput.*, 36(4):A1556–A1580, 2014.

[22] Peter G. Bolhuis. Transition path sampling on diffusive barriers. *Journal of Physics: Condensed Matter*, 15(1):S113, 2002.

[23] François Bolley, José A. Cañizo, and José A. Carrillo. Stochastic mean-field limit: non-Lipschitz forces and swarming. *Math. Models Methods Appl. Sci.*, 21(11):2179–2210, 2011.

[24] François Bolley, Ivan Gentil, and Arnaud Guillin. Convergence to equilibrium in Wasserstein distance for Fokker-Planck equations. *J. Funct. Anal.*, 263(8):2430–2457, 2012.

[25] François Bolley, Ivan Gentil, and Arnaud Guillin. Uniform convergence to equilibrium for granular media. *Arch. Ration. Mech. Anal.*, 208(2):429–445, 2013.

[26] François Bolley, Arnaud Guillin, and Florent Malrieu. Trend to equilibrium and particle approximation for a weakly selfconsistent Vlasov-Fokker-Planck equation. *M2AN Math. Model. Numer. Anal.*, 44(5):867–884, 2010.

[27] Nawaf Bou-Rabee and Andreas Eberle. Markov Chain Monte Carlo Methods. *Lecture Notes, University of Bonn, https://wt.iam.uni-bonn.de/eberle*, 2020.

[28] Nawaf Bou-Rabee and Andreas Eberle. Mixing Time Guarantees for Unadjusted Hamiltonian Monte Carlo. *arXiv preprint arXiv:2105.00887v1*, 2021.

[29] Nawaf Bou-Rabee and Andreas Eberle. Two-scale coupling for preconditioned Hamiltonian Monte Carlo in infinite dimensions. *Stoch. Partial Differ. Equ. Anal. Comput.*, 9(1):207–242, 2021.

[30] Nawaf Bou-Rabee and Andreas Eberle. Couplings for Andersen dynamics. *Ann. Inst. Henri Poincaré Probab. Stat.*, 58(2):916–944, 2022.

[31] Nawaf Bou-Rabee, Andreas Eberle, and Raphael Zimmer. Coupling and convergence for Hamiltonian Monte Carlo. *Ann. Appl. Probab.*, 30(3):1209–1250, 2020.

[32] Nawaf Bou-Rabee and Houman Owhadi. Long-run accuracy of variational integrators in the stochastic context. *SIAM J. Numer. Anal.*, 48(1):278–297, 2010.

[33] Nawaf Bou-Rabee and J. M. Sanz-Serna. Geometric integrators and the Hamiltonian Monte Carlo method. *Acta Numer.*, 27:113–206, 2018.

[34] Nawaf Bou-Rabee and Jesús María Sanz-Serna. Randomized Hamiltonian Monte Carlo. *Ann. Appl. Probab.*, 27(4):2159–2194, 2017.

[35] Alexandre Bouchard-Côté, Sebastian J. Vollmer, and Arnaud Doucet. The bouncy particle sampler: a nonreversible rejection-free Markov chain Monte Carlo method. *J. Amer. Statist. Assoc.*, 113(522):855–867, 2018.

[36] F. Bouchut and J. Dolbeault. On long time asymptotics of the Vlasov-Fokker-Planck equation and of the Vlasov-Poisson-Fokker-Planck system with Coulombic and Newtonian potentials. *Differential Integral Equations*, 8(3):487–514, 1995.

[37] Giovanni Bussi and Michele Parrinello. Accurate sampling using Langevin dynamics. *Phys. Rev. E*, 75:056707, May 2007.

[38] Cédric M. Campos and J. M. Sanz-Serna. Extra chance generalized hybrid Monte Carlo. *J. Comput. Phys.*, 281:365–374, 2015.

[39] Eric Cancès, Frédéric Legoll, and Gabriel Stoltz. Theoretical and numerical comparison of some sampling methods for molecular dynamics. *M2AN Math. Model. Numer. Anal.*, 41(2):351–389, 2007.

[40] Yu Cao, Jianfeng Lu, and Lihan Wang. On explicit L^2-convergence rate estimate for underdamped Langevin dynamics. *arXiv preprint arXiv:1908.04746v4*, 2019.

[41] J. A. Carrillo, R. S. Gvalani, G. A. Pavliotis, and A. Schlichting. Long-time behaviour and phase transitions for the McKean-Vlasov equation on the torus. *Arch. Ration. Mech. Anal.*, 235(1):635–690, 2020.

[42] José A. Carrillo, Robert J. McCann, and Cédric Villani. Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates. *Rev. Mat. Iberoamericana*, 19(3):971–1018, 2003.

[43] José A. Carrillo, Robert J. McCann, and Cédric Villani. Contractions in the 2-Wasserstein length space and thermalization of granular media. *Arch. Ration. Mech. Anal.*, 179(2):217–263, 2006.

[44] P. Cattiaux, A. Guillin, and F. Malrieu. Probabilistic approach for granular media equations in the non-uniformly convex case. *Probab. Theory Related Fields*, 140(1-2):19–40, 2008.

[45] Louis-Pierre Chaintron and Antoine Diez. Propagation of chaos: a review of models, methods and applications. II. Applications. *arXiv preprint arXiv:2106.14812v2*, 2021.

[46] Louis-Pierre Chaintron and Antoine Diez. Propagation of chaos: a review of models, methods and applications.I. Models and methods. *arXiv preprint arXiv:2203.00446*, 2022.

[47] Fang Chen, László Lovász, and Igor Pak. Lifting Markov chains to speed up mixing. In *Annual ACM Symposium on Theory of Computing (Atlanta, GA, 1999)*, pages 275–281. ACM, New York, 1999.

[48] Mu-Fa Chen. *From Markov chains to non-equilibrium particle systems*. World Scientific Publishing Co., Inc., River Edge, NJ, second edition, 2004.

[49] Mu Fa Chen and Feng Yu Wang. Estimation of the first eigenvalue of second order elliptic operators. *J. Funct. Anal.*, 131(2):345–363, 1995.

[50] Mu-Fa Chen and Feng-Yu Wang. Estimation of spectral gap for elliptic operators. *Trans. Amer. Math. Soc.*, 349(3):1239–1267, 1997.

[51] Tianqi Chen, Emily Fox, and Carlos Guestrin. Stochastic gradient Hamiltonian Monte Carlo. In *International conference on machine learning*, pages 1683–1691, 2014.

[52] Yuansi Chen, Raaz Dwivedi, Martin J. Wainwright, and Bin Yu. Fast mixing of metropolized Hamiltonian Monte Carlo: benefits of multi-step gradients. *J. Mach. Learn. Res.*, 21:Paper No. 92, 71, 2020.

[53] Zongchen Chen and Santosh S. Vempala. Optimal convergence rate of Hamiltonian Monte Carlo for strongly logconcave distributions. In *Approximation, randomization, and combinatorial optimization. Algorithms and techniques*, volume 145 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 64, 12. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2019.

[54] Xiang Cheng, Niladri S. Chatterji, Yasin Abbasi-Yadkori, Peter L. Bartlett, and Michael I. Jordan. Sharp convergence rates for Langevin dynamics in the nonconvex setting. *arXiv preprint arXiv:1805.01648v4*, 2018.

[55] Xiang Cheng, Niladri S. Chatterji, Peter L. Bartlett, and Michael I. Jordan. Underdamped Langevin MCMC: A non-asymptotic analysis. *arXiv preprint arXiv:1707.03663v7*, 2017.

[56] Sinho Chewi, Chen Lu, Kwangjun Ahn, Xiang Cheng, Thibaut Le Gouic, and Philippe Rigollet. Optimal dimension dependence of the Metropolis-Adjusted Langevin Algorithm. *arXiv preprint arXiv:2012.12810*, 2020.

[57] R. Chitashvili. On the nonexistence of a strong solution in the boundary problem for a sticky Brownian motion. *Proc. A. Razmadze Math. Inst.*, 115:17–31, 1997.

[58] Imre Csiszár. Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 8:85–108, 1963.

[59] Arnak S. Dalalyan. Theoretical guarantees for approximate sampling from smooth and log-concave densities. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 79(3):651–676, 2017.

[60] Arnak S. Dalalyan and Lionel Riou-Durand. On sampling from a log-concave density using kinetic Langevin diffusions. *Bernoulli*, 26(3):1956–1988, 2020.

[61] Masoumeh Dashti and Andrew M. Stuart. The Bayesian approach to inverse problems. In *Handbook of uncertainty quantification. Vol. 1, 2, 3*, pages 311–428. Springer, Cham, 2017.

[62] M. H. A. Davis. Piecewise-deterministic Markov processes: a general class of nondiffusion stochastic models. *J. Roy. Statist. Soc. Ser. B*, 46(3):353–388, 1984. With discussion.

[63] François Delarue and Alvin Tse. Uniform in time weak propagation of chaos on the torus. *arXiv preprint arXiv:2104.14973*, 2021.

[64] Matias G. Delgadino, Rishabh S. Gvalani, and Grigorios A. Pavliotis. On the diffusive-mean field limit for weakly interacting diffusions exhibiting phase transitions. *Arch. Ration. Mech. Anal.*, 241(1):91–148, 2021.

[65] George Deligiannidis, Alexandre Bouchard-Côté, and Arnaud Doucet. Exponential ergodicity of the bouncy particle sampler. *Ann. Statist.*, 47(3):1268–1287, 2019.

[66] R. L. Dobrushin. Prescribing a system of random variables by conditional distributions. *Theory of Probability & Its Applications*, 15(3):458–486, 1970.

[67] Jean Dolbeault, Clément Mouhot, and Christian Schmeiser. Hypocoercivity for kinetic equations with linear relaxation terms. *C. R. Math. Acad. Sci. Paris*, 347(9-10):511–516, 2009.

[68] Jean Dolbeault, Clément Mouhot, and Christian Schmeiser. Hypocoercivity for linear kinetic equations conserving mass. *Trans. Amer. Math. Soc.*, 367(6):3807–3828, 2015.

[69] Simon Duane, A. D. Kennedy, Brian J. Pendleton, and Duncan Roweth. Hybrid Monte Carlo. *Phys. Lett. B*, 195(2):216–222, 1987.

[70] D. B. Dunson and J. E. Johndrow. The Hastings algorithm at fifty. *Biometrika*, 107(1):1–23, 2020.

[71] M. H. Duong and J. Tugaut. Stationary solutions of the Vlasov-Fokker-Planck equation: existence, characterization and phase-transition. *Appl. Math. Lett.*, 52:38–45, 2016.

[72] Manh Hong Duong and Julian Tugaut. The Vlasov-Fokker-Planck equation in non-convex landscapes: convergence to equilibrium. *Electron. Commun. Probab.*, 23:Paper No. 19, 10, 2018.

[73] Alain Durmus and Andreas Eberle. Asymptotic bias of inexact Markov Chain Monte Carlo methods in high dimension. *arXiv preprint arXiv:2108.00682*, 2021.

[74] Alain Durmus, Andreas Eberle, Arnaud Guillin, and Katharina Schuh. Sticky nonlinear SDEs and convergence of McKean-Vlasov equations without confinement. *arXiv preprint arXiv:2201.07652*, 2022.

[75] Alain Durmus, Andreas Eberle, Arnaud Guillin, and Raphael Zimmer. An elementary approach to uniform in time propagation of chaos. *Proc. Amer. Math. Soc.*, 148(12):5387–5398, 2020.

[76] Alain Durmus, Arnaud Guillin, and Pierre Monmarché. Piecewise deterministic Markov processes and their invariant measures. *Ann. Inst. Henri Poincaré Probab. Stat.*, 57(3):1442–1475, 2021.

[77] Alain Durmus and Éric Moulines. Nonasymptotic convergence analysis for the unadjusted Langevin algorithm. *Ann. Appl. Probab.*, 27(3):1551–1587, 2017.

[78] Alain Durmus and Éric Moulines. High-dimensional Bayesian inference via the unadjusted Langevin algorithm. *Bernoulli*, 25(4A):2854–2882, 2019.

[79] Alain Durmus and Éric Moulines. On the geometric convergence for MALA under verifiable conditions. *arXiv preprint arXiv:2201.01951*, 2022.

[80] Alain Durmus, Eric Moulines, and Eero Saksman. On the convergence of Hamiltonian Monte Carlo. *arXiv preprint arXiv:1705.00166v2*, 2017.

[81] Raaz Dwivedi, Yuansi Chen, Martin J. Wainwright, and Bin Yu. Log-concave sampling: Metropolis-Hastings algorithms are fast. *J. Mach. Learn. Res.*, 20:Paper No. 183, 42, 2019.

[82] Andreas Eberle. Error bounds for Metropolis-Hastings algorithms applied to perturbations of Gaussian measures in high dimensions. *Ann. Appl. Probab.*, 24(1):337–377, 2014.

[83] Andreas Eberle. Reflection couplings and contraction rates for diffusions. *Probab. Theory Related Fields*, 166(3-4):851–886, 2016.

[84] Andreas Eberle. Markov processes. *Lecture Notes, University of Bonn, https://wt.iam.uni-bonn.de/eberle*, 2020.

[85] Andreas Eberle, Arnaud Guillin, and Raphael Zimmer. Couplings and quantitative contraction rates for Langevin dynamics. *Ann. Probab.*, 47(4):1982–2010, 2019.

[86] Andreas Eberle, Arnaud Guillin, and Raphael Zimmer. Quantitative Harris-type theorems for diffusions and McKean-Vlasov processes. *Trans. Amer. Math. Soc.*, 371(10):7135–7173, 2019.

[87] Andreas Eberle and Raphael Zimmer. Sticky couplings of multidimensional diffusions with different drifts. *Ann. Inst. Henri Poincaré Probab. Stat.*, 55(4):2370–2394, 2019.

[88] A. Einstein. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik*, 322(8):549–560, 1905.

[89] E. Emmrich. Discrete versions of Gronwall's lemma and their application to the numerical analysis of parabolic problems. *Preprint No. 637, Fachbereich Mathematik, TU Berlin*, 1999.

[90] Hans-Jürgen Engelbert and Goran Peskir. Stochastic differential equations for sticky Brownian motion. *Stochastics*, 86(6):993–1021, 2014.

[91] Youhan Fang, Jesus-Maria Sanz-Serna, and Robert D Skeel. Compressible generalized hybrid monte carlo. *The Journal of chemical physics*, 140(17):174108, 2014.

[92] William Feller. Diffusion processes in one dimension. *Trans. Amer. Math. Soc.*, 77:1–31, 1954.

[93] William Feller. The general diffusion operator and positivity preserving semi-groups in one dimension. *Ann. of Math. (2)*, 60:417–436, 1954.

[94] Daan Frenkel and Berend Smit. *Understanding Molecular Simulation: From Algorithms to Applications, Second Edition.* Academic Press, 2002.

[95] Tadahisa Funaki. A certain class of diffusion processes associated with nonlinear parabolic equations. *Z. Wahrsch. Verw. Gebiete*, 67(3):331–348, 1984.

[96] Mark Girolami and Ben Calderhead. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 73(2):123–214, 2011. With discussion and a reply by the authors.

[97] Arnaud Guillin, Pierre Le Bris, and Pierre Monmarché. Convergence rates for the Vlasov-Fokker-Planck equation and uniform in time propagation of chaos in non convex cases. *arXiv preprint arXiv:2105.09070v2*, 2021.

[98] Arnaud Guillin, Wei Liu, Liming Wu, and Chaoen Zhang. Uniform Poincaré and logarithmic Sobolev inequalities for mean field particles systems. *arXiv preprint arXiv:1909.07051v1*, 2019.

[99] Arnaud Guillin, Wei Liu, Liming Wu, and Chaoen Zhang. The kinetic Fokker-Planck equation with mean field interaction. *J. Math. Pures Appl. (9)*, 150:1–23, 2021.

[100] Arnaud Guillin and Pierre Monmarché. Uniform long-time and propagation of chaos estimates for mean field kinetic particles in non-convex landscapes. *J. Stat. Phys.*, 185(2):Paper No. 15, 20, 2021.

[101] Rajan Gupta, Gregory W. Kilcup, and Stephen R. Sharpe. Tuning the hybrid Monte Carlo algorithm. *Phys. Rev. D*, 38:1278–1287, Aug 1988.

[102] M. Hairer, J. C. Mattingly, and M. Scheutzow. Asymptotic coupling and a general form of Harris' theorem with applications to stochastic delay equations. *Probab. Theory Related Fields*, 149(1-2):223–259, 2011.

[103] Martin Hairer. Convergence of Markov processes. *Lecture notes, Mathematics Department, University of Warwick, http://www.hairer.org/notes/Convergence.pdf*, 2010.

[104] Martin Hairer and Jonathan C. Mattingly. Yet another look at Harris' ergodic theorem for Markov chains. In *Seminar on Stochastic Analysis, Random Fields and Applications VI*, volume 63 of *Progr. Probab.*, pages 109–117. Birkhäuser/Springer Basel AG, Basel, 2011.

[105] Martin Hairer, Andrew M. Stuart, and Sebastian J. Vollmer. Spectral gaps for a Metropolis-Hastings algorithm in infinite dimensions. *Ann. Appl. Probab.*, 24(6):2455–2490, 2014.

[106] William R. P. Hammersley, David Šiška, and Ł ukasz Szpruch. McKean-Vlasov SDEs under measure dependent Lyapunov conditions. *Ann. Inst. Henri Poincaré Probab. Stat.*, 57(2):1032–1057, 2021.

[107] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[108] Maxime Hauray and Stéphane Mischler. On Kac's chaos and related problems. *J. Funct. Anal.*, 266(10):6055–6157, 2014.

[109] Bernard Helffer and Francis Nier. *Hypoelliptic estimates and spectral theory for Fokker-Planck operators and Witten Laplacians*, volume 1862 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2005.

[110] Frédéric Hérau. Short and long time behavior of the Fokker-Planck equation in a confining potential and applications. *J. Funct. Anal.*, 244(1):95–118, 2007.

[111] Frédéric Hérau and Francis Nier. Isotropic hypoellipticity and trend to equilibrium for the Fokker-Planck equation with a high-degree potential. *Arch. Ration. Mech. Anal.*, 171(2):151–218, 2004.

[112] Matthew D. Hoffman and Andrew Gelman. The no-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.*, 15:1593–1623, 2014.

[113] Alan M. Horowitz. A generalized guided Monte Carlo algorithm. *Physics Letters B*, 268(2):247–252, 1991.

[114] Kaitong Hu, Zhenjie Ren, David Šiška, and Ł ukasz Szpruch. Mean-field Langevin dynamics and energy landscape of neural networks. *Ann. Inst. Henri Poincaré Probab. Stat.*, 57(4):2043–2065, 2021.

[115] Nobuyuki Ikeda. On the construction of two-dimensional diffusion processes satisfying Wentzell's boundary conditions and its application to boundary value problems. *Mem. Coll. Sci. Univ. Kyoto Ser. A. Math.*, 33:367–427, 1960/61.

[116] Nobuyuki Ikeda and Shinzo Watanabe. *Stochastic differential equations and diffusion processes*, volume 24 of *North-Holland Mathematical Library*. North-Holland Publishing Co., Amsterdam; Kodansha, Ltd., Tokyo, second edition, 1989.

[117] Aldéric Joulin. Poisson-type deviation inequalities for curved continuous-time Markov chains. *Bernoulli*, 13(3):782–798, 2007.

[118] Aldéric Joulin and Yann Ollivier. Curvature, concentration and error estimates for Markov chain Monte Carlo. *Ann. Probab.*, 38(6):2418–2442, 2010.

[119] B. Jourdain and S. Méléard. Propagation of chaos and fluctuations for a moderate model with smooth initial data. *Ann. Inst. H. Poincaré Probab. Statist.*, 34(6):727–766, 1998.

[120] M. Kac. Foundations of kinetic theory. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. III*, pages 171–197. University of California Press, Berkeley-Los Angeles, Calif., 1956.

[121] Olav Kallenberg. *Foundations of modern probability.* Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.

[122] Anna Kazeykina, Zhenjie Ren, Xiaolu Tan, and Junjian Yang. Ergodicity of the underdamped mean-field Langevin dynamics. *arXiv preprint arXiv:2007.14660v2*, 2020.

[123] Achim Klenke. *Probability theory.* Universitext. Springer, London, second edition, 2014. A comprehensive course.

[124] Roman Korol, Jorge L. Rosa-Raíces, Nawaf Bou-Rabee, and Thomas F. Miller. Dimension-free path-integral molecular dynamics without preconditioning. *The Journal of Chemical Physics*, 152(10):104102, 2020.

[125] S. C. Kou, Qing Zhou, and Wing Hung Wong. Equi-energy sampler with applications in statistical inference and statistical mechanics. *Ann. Statist.*, 34(4):1581–1652, 2006. With discussions and a rejoinder by the authors.

[126] Paul Langevin. Sur la theórie du mouvement brownien. *Comptes rendus de l'Académie des sciences (Paris)*, 146:530–533, 1908.

[127] Yin Tat Lee, Ruoqi Shen, and Kevin Tian. Logsmooth Gradient Concentration and Tighter Runtimes for Metropolized Hamiltonian Monte Carlo. In Jacob Abernethy and Shivani Agarwal, editors, *Proceedings of Thirty Third Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pages 2565–2597. PMLR, 09–12 Jul 2020.

[128] Benedict Leimkuhler and Charles Matthews. Rational construction of stochastic numerical methods for molecular sampling. *Appl. Math. Res. Express. AMRX*, (1):34–56, 2013.

[129] Benedict Leimkuhler, Charles Matthews, and Gabriel Stoltz. The computation of averages from equilibrium and nonequilibrium Langevin molecular dynamics. *IMA J. Numer. Anal.*, 36(1):13–79, 2016.

[130] Tony Lelièvre, Mathias Rousset, and Gabriel Stoltz. *Free energy computations.* Imperial College Press, London, 2010. A mathematical perspective.

[131] David A. Levin and Yuval Peres. *Markov chains and mixing times.* American Mathematical Society, Providence, RI, 2017. Second edition of [MR2466937], With contributions by Elizabeth L. Wilmer, With a chapter on "Coupling from the past" by James G. Propp and David B. Wilson.

[132] Faming Liang and Wing Hung Wong. Real-parameter evolutionary Monte Carlo with applications to Bayesian mixture models. *J. Amer. Statist. Assoc.*, 96(454):653–666, 2001.

[133] Torgny Lindvall and L. C. G. Rogers. Coupling of multidimensional diffusions by reflection. *Ann. Probab.*, 14(3):860–872, 1986.

[134] Jun S. Liu. *Monte Carlo strategies in scientific computing.* Springer Series in Statistics. Springer-Verlag, New York, 2001.

[135] Samuel Livingstone, Michael Betancourt, Simon Byrne, and Mark Girolami. On the geometric ergodicity of Hamiltonian Monte Carlo. *Bernoulli*, 25(4A):3109–3138, 2019.

[136] Paul B. Mackenzie. An Improved Hybrid Monte Carlo Method. *Phys. Lett.*, B226:369–371, 1989.

[137] F. Malrieu. Logarithmic Sobolev inequalities for some nonlinear PDE's. *Stochastic Process. Appl.*, 95(1):109–132, 2001.

[138] Florent Malrieu. Convergence to equilibrium for granular media equations and their Euler schemes. *Ann. Appl. Probab.*, 13(2):540–560, 2003.

[139] Oren Mangoubi and Aaron Smith. Rapid Mixing of Hamiltonian Monte Carlo on Strongly Log-Concave Distributions. *arXiv preprint arXiv:1708.07114v1*, 2017.

[140] 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):185–232, 2002.

[141] Jonathan C. Mattingly, Andrew M. Stuart, and M. V. Tretyakov. Convergence of numerical time-averaging and stationary measures via Poisson equations. *SIAM J. Numer. Anal.*, 48(2):552–577, 2010.

[142] H. P. McKean, Jr. A. Skorohod's stochastic integral equation for a reflecting barrier diffusion. *J. Math. Kyoto Univ.*, 3:85–88, 1963.

[143] H. P. McKean, Jr. A class of Markov processes associated with nonlinear parabolic equations. *Proc. Nat. Acad. Sci. U.S.A.*, 56:1907–1911, 1966.

[144] H. P. McKean, Jr. Propagation of chaos for a class of non-linear parabolic equations. In *Stochastic Differential Equations (Lecture Series in Differential Equations, Session 7, Catholic Univ., 1967)*, pages 41–57. Air Force Office Sci. Res., Arlington, Va., 1967.

[145] Song Mei, Andrea Montanari, and Phan-Minh Nguyen. A mean field view of the landscape of two-layer neural networks. *Proc. Natl. Acad. Sci. USA*, 115(33):E7665–E7671, 2018.

[146] Sylvie Méléard. Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models. In *Probabilistic models for nonlinear partial differential equations (Montecatini Terme, 1995)*, volume 1627 of *Lecture Notes in Math.*, pages 42–95. Springer, Berlin, 1996.

[147] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[148] Stéphane Mischler and Clément Mouhot. Kac's program in kinetic theory. *Invent. Math.*, 193(1):1–147, 2013.

[149] Stéphane Mischler, Clément Mouhot, and Bernt Wennberg. A new approach to quantitative propagation of chaos for drift, diffusion and jump processes. *Probab. Theory Related Fields*, 161(1-2):1–59, 2015.

[150] Yuliya Mishura and Alexander Veretennikov. Existence and uniqueness theorems for solutions of Mckean-Vlasov stochastic equations. *Theory Probab. Math. Statist.*, (103):59–101, 2020.

[151] Pierre Monmarché. Long-time behaviour and propagation of chaos for mean field kinetic particles. *Stochastic Process. Appl.*, 127(6):1721–1737, 2017.

[152] Pierre Monmarché. Almost sure contraction for diffusions on $\mathbb{R}^d$. Application to generalised Langevin diffusions. *arxiv preprint arXiv:2009.10828v5*, 2021.

[153] Pierre Monmarché. High-dimensional MCMC with a standard splitting scheme for the underdamped Langevin diffusion. *Electron. J. Stat.*, 15(2):4117–4166, 2021.

[154] Tetsuzo Morimoto. Markov processes and the H-theorem. *J. Phys. Soc. Japan*, 18:328–331, 1963.

[155] Radford M. Neal. *Bayesian Learning for Neural Networks.* Springer-Verlag, Berlin, Heidelberg, 1996.

[156] Radford M. Neal. MCMC using Hamiltonian dynamics. In *Handbook of Markov chain Monte Carlo*, Chapman & Hall/CRC Handb. Mod. Stat. Methods, pages 113–162. CRC Press, Boca Raton, FL, 2011.

[157] Yurii Nesterov. *Lectures on convex optimization*, volume 137 of *Springer Optimization and Its Applications.* Springer, Cham, 2018. Second edition of [MR2142598].

[158] Karl Oelschläger. A martingale approach to the law of large numbers for weakly interacting stochastic processes. *Ann. Probab.*, 12(2):458–479, 1984.

[159] Karl Oelschläger. A law of large numbers for moderately interacting diffusion processes. *Z. Wahrsch. Verw. Gebiete*, 69(2):279–322, 1985.

[160] Karl Oelschläger. A fluctuation theorem for moderately interacting diffusion processes. *Probab. Theory Related Fields*, 74(4):591–616, 1987.

[161] Bernt Øksendal. *Stochastic differential equations.* Universitext. Springer-Verlag, Berlin, sixth edition, 2003. An introduction with applications.

[162] Yann Ollivier. A survey of Ricci curvature for metric spaces and Markov chains. In *Probabilistic approach to geometry*, volume 57 of *Adv. Stud. Pure Math.*, pages 343–381. Math. Soc. Japan, Tokyo, 2010.

[163] F. Otto and C. Villani. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.*, 173(2):361–400, 2000.

[164] G. Parisi. Correlation functions and computer simulations. *Nuclear Phys. B*, 180(3, FS 2):378–384, 1981.

[165] Grigorios A. Pavliotis. *Stochastic processes and applications*, volume 60 of *Texts in Applied Mathematics.* Springer, New York, 2014. Diffusion processes, the Fokker-Planck and Langevin equations.

[166] Goran Peskir. *On boundary behaviour of one-dimensional diffusions: From Brown to Feller and beyond.* In William Feller – Selected Papers II. Springer, Berlin, 2015.

[167] Jakiw Pidstrigach. Convergence of Preconditioned Hamiltonian Monte Carlo on Hilbert Spaces. *arXiv preprint arXiv:2011.08578*, 2020.

[168] Natesh S. Pillai, Andrew M. Stuart, and Alexandre H. Thiéry. Optimal scaling and diffusion limits for the Langevin algorithm in high dimensions. *Ann. Appl. Probab.*, 22(6):2320–2356, 2012.

[169] F. J. Pinski and A. M. Stuart. Transition paths in molecules at finite temperature. *The Journal of Chemical Physics*, 132(18):184104, 2010.

[170] Philip E. Protter. *Stochastic integration and differential equations*, volume 21 of *Applications of Mathematics (New York).* Springer-Verlag, Berlin, second edition, 2004. Stochastic Modelling and Applied Probability.

[171] Alfréd Rényi. On measures of entropy and information. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I*, pages 547–561. Univ. California Press, Berkeley, Calif., 1961.

[172] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences].* Springer-Verlag, Berlin, third edition, 1999.

[173] G. O. Roberts, A. Gelman, and W. R. Gilks. Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.*, 7(1):110–120, 1997.

[174] Gareth O. Roberts and Jeffrey S. Rosenthal. Optimal scaling of discrete approximations to Langevin diffusions. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 60(1):255–268, 1998.

[175] Gareth O. Roberts and Richard L. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363, 1996.

[176] L. C. G. Rogers and David Williams. *Diffusions, Markov processes, and martingales. Vol. 1.* Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Ltd., Chichester, second edition, 1994. Foundations.

[177] Peter J. Rossky, Jimmie D. Doll, and Harold L. Friedman. Brownian dynamics as smart Monte Carlo simulation. *Journal of Chemical Physics*, 69:4628–4633, 1978.

[178] Grant M. Rotskoff and Eric Vanden-Eijnden. Trainability and Accuracy of Neural Networks: An Interacting Particle System Approach. *arXiv preprint arXiv:1805.00915v3*, 2018.

[179] Gilles Royer. *An initiation to logarithmic Sobolev inequalities*, volume 14 of *SMF/AMS Texts and Monographs.* American Mathematical Society, Providence, RI; Société Mathématique de France, Paris, 2007. Translated from the 1999 French original by Donald Babbitt.

[180] Marcus D. Sammer. Aspects of mass transportation in discrete concentration inequalities. *Ph.D. Thesis, Georgia: Georgia Institute of Technology. https://smartech.gatech.edu/handle/1853/7006.*, 2005.

[181] Katharina Schuh. Global contractivity for Langevin dynamics with distribution-dependent forces and uniform in time propagation of chaos. *ArXiv preprint arXiv:2206.03082*, 2022.

[182] Christof Schütte. *Conformational dynamics: Modeling, Theory, Algorithm, and Application to Biomolecules.* Habilitation, Free University Berlin, 1999.

[183] A. V. Skorokhod. Stochastic Equations for Diffusion Processes in a Bounded Region. *Theory of Probability & Its Applications*, 6(3):264–274, 1961.

[184] A. V. Skorokhod. Stochastic Equations for Diffusion Processes in a Bounded Region. II. *Theory of Probability & Its Applications*, 7(1):3–23, 1962.

[185] Gabriel Stoltz. *Some Mathematical Methods for Molecular and Multiscale Simulation*. PhD thesis, Ecole Nationale des Ponts et Chaussées, 2007.

[186] Daniel W. Stroock and S. R. S. Varadhan. Diffusion processes with boundary conditions. *Comm. Pure Appl. Math.*, 24:147–225, 1971.

[187] Alain-Sol Sznitman. Topics in propagation of chaos. In *École d'Été de Probabilités de Saint-Flour XIX—1989*, volume 1464 of *Lecture Notes in Math.*, pages 165–251. Springer, Berlin, 1991.

[188] D. Talay. Stochastic Hamiltonian systems: exponential convergence to the invariant measure, and discretization by the implicit Euler scheme. *Markov Process. Related Fields*, 8(2):163–198, 2002. Inhomogeneous random systems (Cergy-Pontoise, 2001).

[189] Julian Tugaut. Convergence to the equilibria for self-stabilizing processes in double-well landscape. *Ann. Probab.*, 41(3A):1427–1460, 2013.

[190] Cédric Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.

[191] Cédric Villani. Hypocoercivity. *Mem. Amer. Math. Soc.*, 202(950):iv+141, 2009.

[192] Cédric Villani. *Optimal transport*, volume 338 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.

[193] Max-K. von Renesse and Karl-Theodor Sturm. Transport inequalities, gradient estimates, entropy, and Ricci curvature. *Comm. Pure Appl. Math.*, 58(7):923–940, 2005.

[194] Maxime Vono, Daniel Paulin, and Arnaud Doucet. Efficient MCMC Sampling with Dimension-Free Convergence Rate using ADMM-type Splitting. *arXiv preprint arXiv:1905.11937v6*, 2019.

[195] David J Wales. Energy Landscapes of Clusters Bound by Short-Ranged Potentials. *ChemPhysChem*, 11(12):2491–2494, 2010.

[196] S. Watanabe. On stochastic differential equations for multi-dimensional diffusion processes with boundary conditions. *J. Math. Kyoto Univ.*, 11:169–180, 1971.

[197] S. Watanabe. On stochastic differential equations for multi-dimensional diffusion processes with boundary conditions. II. *J. Math. Kyoto Univ.*, 11:545–551, 1971.

[198] Dirk Werner. *Funktionalanalysis*. Springer-Verlag, Berlin, extended edition, 2000.

[199] Keru Wu, Scott Schmidler, and Yuansi Chen. Minimax Mixing Time of the Metropolis-Adjusted Langevin Algorithm for Log-Concave Sampling. *arXiv preprint arXiv:2109.13055*, 2021.

[200] Liming Wu. Large and moderate deviations and exponential convergence for stochastic damping Hamiltonian systems. *Stochastic Process. Appl.*, 91(2):205–238, 2001.

[201] Raphael Zimmer. *Couplings and Kantorovich contractions with explicit rates for diffusions.* PhD thesis, Rheinische Friedrich-Wilhelms-Universität Bonn, September 2017.

Appendix A

Convergence of unadjusted HMC for mean-field particle models

Nawaf Bou-Rabee and Katharina Schuh, Convergence of unadjusted Hamiltonian Monte Carlo for mean-field models. ArXiv e-print 2009.08735, September 2020.[1]

Abstract

We present dimension-free convergence and discretization error bounds for the unadjusted Hamiltonian Monte Carlo algorithm applied to high-dimensional probability distributions of mean-field type. These bounds require the discretization step to be sufficiently small, but do not require strong convexity of either the unary or pairwise potential terms present in the mean-field model. To handle high dimensionality, our proof uses a particlewise coupling that is contractive in a complementary particlewise metric.

Key words: Hamiltonian Monte Carlo, coupling, convergence to equilibrium, mean-field models.
Mathematics Subject Classification: Primary 60J05; secondary 65P10, 65C05.

A.1 Introduction

Markov Chain Monte Carlo (MCMC) methods are used to sample from a target probability distribution of the form $\mu(\mathrm{d}x) \propto \exp(-U(x))\mathrm{d}x$. The simplest methods (e.g., Gibbs and random walk Metropolis) display random walk behavior which slow their convergence to equilibrium. This slow convergence motivates the Hamiltonian Monte Carlo (HMC) method, first established in [69], which offers the potential to converge faster, particularly in high dimension [156, 96, 17, 54, 70].

The convergence properties of HMC have received increasing interest. Ergodicity was proven in [182, 39, 185]. By drift/minorization conditions, geometric ergodicity was demonstrated in [34, 135, 80]. In [31, 139, 53], the convergence behavior is analyzed for a strongly convex potential U and explicit bounds on convergence rates are obtained using a synchronous coupling

[1]The authors would like to thank Andreas Eberle for his insights and advice during the development of this work. N. B.-R. was supported by the National Science Foundation under Grant No. DMS-1816378 and the Alexander von Humboldt Foundation. K. S. was supported by *Bonn International Graduate School of Mathematics.* Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) im Rahmen der Exzellenzstrategie des Bundes und der Länder - GZ 2047/1, Projekt-ID 390685813.

approach. In [31], contraction bounds were obtained for more general potentials U by developing a coupling tailored to HMC. However, these convergence bounds deteriorate in high dimension for mean-field models (see, in particular, (A.11) for the precise form of these contraction bounds for high-dimensional mean-field models). Therefore, a new approach is needed to obtain convergence bounds for non-strongly convex potentials of mean-field type that are *dimension-free*, i.e., independent of the number of particles in the mean-field model.

Mean-field models play an important role in understanding statistical properties of high-dimensional systems. This connection was introduced by Kac in [120] as propagation of chaos and has been investigated amongst others in [143, 187, 146], for very recent related work on second-order mean-field Langevin dynamics see [98, 100]. A key component in Kac's program was to establish bounds on relaxation times of many-body dynamical systems that are dimension-free, see Section 1.4 of [148] for a fuller discussion.

The behavior of HMC in high-dimensional mean-field models is also relevant, at least conceptually, to molecular dynamics (MD), see [6] and [94], or [130] for a mathematical perspective. MD involves the time integration of high-dimensional Hamiltonian dynamics often coupled to a heat or pressure bath [6, 94]. The corresponding process typically admits a stationary distribution. Time discretization introduces an error in the numerically sampled stationary distribution. In general, one might hope that this discretization error is dimension-free for ergodic averages of measurable functions ("observables") that are intensive (e.g., energy per particle) as opposed to extensive (e.g., total energy). A key contribution of this paper is to demonstrate that this is indeed the case for particles with weak mean-field interactions (see Theorem A.13 and Remark A.14).

In this paper, we consider high-dimensional mean-field models, where the potential $U : \mathbb{R}^{dn} \to \mathbb{R}$ is a function of the form

$$U(x) = \sum_{i=1}^{n} \Big(V(x^i) + \frac{\epsilon}{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} W(x^i - x^j) \Big).$$

Here, $V : \mathbb{R}^d \to \mathbb{R}$ and $W : \mathbb{R}^d \to \mathbb{R}$ are twice differentiable functions, ϵ is a real constant and $x = (x^1, ..., x^n)$ where $x^i \in \mathbb{R}^d$ represents the position of the i-th particle. Usually, d is a small fixed number that represents the dimension per particle, whereas the number n of particles is large. We call the unary potential V the *confinement potential per particle* and the pairwise potential W the *interaction potential*. While we focus on mean-field U with pairwise interactions in this paper, our results can be readily extended to potentials U with more general mean-field interactions (see Remark A.1).

In its simplest form, every HMC step uses the Hamiltonian dynamics $(q_t(x, v), p_t(x, v))$ of the mean-field particle system with unit masses defined as the solution to the ordinary differential equations

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} q_t^i &= p_t^i \\ \frac{\mathrm{d}}{\mathrm{d}t} p_t^i &= -\nabla_i U(q_t) = -\nabla V(q_t^i) - \frac{\epsilon}{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \Big(\nabla W(q_t^i - q_t^j) - \nabla W(q_t^j - q_t^i) \Big), \end{aligned} \tag{A.1}$$

for $i = 1, ..., n$ with initial value $(q_0, p_0) = (x, v)$. The transition step of the Markov chain in $\mathbb{R}^{dn}$ corresponding to HMC is given by

$$\mathbf{X}(x) = q_T(x, \xi),$$

where the initial velocity $\xi \sim \mathcal{N}(0, I_{dn})$ is sampled independently per HMC step, and the integration time $T > 0$ is a fixed constant, determining the duration of the Hamiltonian dynamics per HMC step. The corresponding Markov chain is known as *exact HMC* because it uses the exact Hamiltonian dynamics and therefore, leaves invariant the target measure μ, cf. [33].

Generally, the choice of the duration T has a large impact on the performance per HMC step. If T is too small, we obtain a highly correlated chain indicative of random walk behavior. Whereas, if T is chosen too large, due to periodicities and near-periodicities, $q_T(x, v)$ can realize U-turns even as the computational cost of the algorithm increases. This issue was observed by Mackenzie in [136], and motivated duration randomization [156, 39, 34] and the No-U-Turn sampler [112]. In contraction bounds for HMC, this issue leads to conditions that limit the duration T of the Hamiltonian dynamics, e.g., for U stronlgy convex $LT^2 \leq$ constant where L is the Lipschitz constant of ∇U [53]. As we discuss more below, non-convexity of U leads to additional restrictions on the duration T.

Since the Hamiltonian dynamics cannot be simulated exactly in general, a numerical version of these dynamics comes into play to approximate the exact dynamics, and normally, the velocity Verlet algorithm is used, cf. [134, 33]. The numerical version contains an additional parameter, the discretization step $h > 0$ satisfying $T \in h\mathbb{Z}$. Note that in the numerical version of HMC without adjusting the algorithm by an additional acceptance-rejection step (see e.g. [156, 33]), the corresponding Markov chain does not exactly preserve the target measure. This chain is called *unadjusted HMC*. In this article we focus on unadjusted HMC because both from the viewpoint of theory and practice the acceptance-rejection step in adjusted HMC may lead to difficulties in high dimension. Indeed, in the product case (when $\epsilon = 0$), a dimension-dependent time step size ($h \propto n^{-1/4}$) is needed to ensure that the acceptance rate in adjusted HMC is bounded away from zero as $n \uparrow \infty$, cf. [17, 101]. Further, as far as we know only a local contraction result for adjusted HMC is known (see Remark A.5). We stress that both adjusted and unadjusted HMC are implementable on a computer, whereas exact HMC is not.

The main result of this paper gives dimension-free convergence bounds for unadjusted HMC applied to mean-field models, i.e., bounds that are independent of the number of particles in the mean-field model. Our proof is motivated by the coupling approach in [31], but with a new 'particlewise' coupling and a complementary particlewise metric. We now state a simplified version of our main result, which holds in the special case of exact HMC where $h = 0$.

We assume that ∇V and ∇W are Lipschitz continuous with Lipschitz constants L and $\tilde{L}$, respectively. Further, we assume that V is K-strongly convex outside a Euclidean ball of radius R, but possibly non-convex inside this ball. Let $\pi(x, dy)$ be the transition kernel of exact HMC, and let $\mathcal{W}_{\ell^1}$ denote the Kantorovich/L^1-Wasserstein distance on $\mathbb{R}^{dn}$ based on an ℓ^1-metric $\ell^1(x, y) = \sum_{i=1}^n |x^i - y^i|$. Then for any two probability measures η and ν on $\mathbb{R}^{dn}$, we show that

$$\mathcal{W}_{\ell^1}(\eta\pi^m, \nu\pi^m) \leq M e^{-cm} \mathcal{W}_{\ell^1}(\eta, \nu). \tag{A.2}$$

Here, $M = \exp\left(\frac{5}{2}\left(1 + \frac{4R}{T}\sqrt{\frac{L+K}{K}}\right)\right)$ and the contraction rate c is of the form

$$c = \frac{1}{156} K T^2 \exp\Big(- 10 \frac{R}{T} \sqrt{\frac{L+K}{K}} \Big).$$

This bound holds provided the duration T and the interaction parameter ϵ are sufficiently small, i.e.,

$$\frac{5}{3}LT^2 \leq \min\Big(\frac{1}{4}, \frac{3K}{10L}, \frac{3K}{256 \cdot 5 \cdot 2^6 LR^2(L+K)}\Big), \text{ and}$$
$$|\epsilon|\tilde{L} < \min\Big(\frac{K}{6}, \frac{1}{2}\Big(\frac{K}{36 \cdot 149}\Big)^2\Big(T + 8R\sqrt{\frac{L+K}{K}}\Big)^2 \exp\Big(-40\frac{R}{T}\sqrt{\frac{L+K}{K}}\Big)\Big).$$

Note that both the contraction rate c and the conditions above are dimension-free, i.e., independent of the number n of particles. A restriction on the strength of interactions ϵ cannot be avoided because for large values of ϵ multiple invariant measures and phase transition phenomena can occur, which typically leads to an exponential deterioration in the rate of convergence as the number of particles tends to infinity [158, 187, 189]. Roughly speaking, the factor LR^2 appearing in the condition on T measures the degree of non-convexity of U and excludes the possibility of high energy barriers. To obtain this result, we first show contraction for a modified Wasserstein distance that is based on a specially designed particlewise metric ρ on $\mathbb{R}^{dn}$, i.e. , $\mathcal{W}_\rho(\eta\pi^m, \nu\pi^m) \leq e^{-cm}\mathcal{W}_\rho(\eta, \nu)$, and by using that ρ is equivalent to ℓ^1, we obtain (A.2). From this result we deduce a quantitative bound for the number m of steps required to approximate the target measure μ up to a given error $\tilde{\epsilon}$, i.e., $\mathcal{W}_{\ell^1}(\eta\pi^m, \mu) \leq \tilde{\epsilon}$. This bound may depend logarithmically on the number n of particles through the distance between the initial distribution and the target measure. Finally, we show quantitative dimension-free bounds on the bias for ergodic averages of intensive observables of the form $f(x) = \frac{1}{n}\sum_i \hat{f}(x^i)$.

For unadjusted HMC, we show the same contraction result provided the discretization step h is chosen small enough and deduce that there exists a unique invariant measure μ_h of unadjusted HMC. Since unadjusted HMC does not exactly preserve the target measure μ, we prove that $\mathcal{W}_{\ell^1}(\mu, \mu_h) = \mathcal{O}(h^2 n)$ provided enough regularity for U is assumed, i.e., V and W are three times differentiable and have bounded third derivatives. If less regularity is assumed, i.e., V and W are only twice differentiable, an $\mathcal{O}(hn)$ bound is obtained. Invariant measure accuracy of numerical approximations for related second-order measure preserving dynamics has been extensively investigated in the literature [175, 140, 188, 141, 21, 32, 129, 1, 2], but according to our knowledge, it is new to obtain bounds on $\mathcal{W}_{l^1}$ with a precise dimension dependence (see Corollary A.9). Durmus and Eberle [73], using partially the same approach, generalize these results on invariant measure accuracy to a broader class of both models and inexact (or unadjusted) MCMC methods.

Other work on HMC in high dimension

The study of the behavior of HMC as dimensionality increases is carried out in other settings, too. For example, in Bayesian inference problems with a large number of observations where the posterior itself is not necessarily high-dimensional. In this setting, sampling the posterior directly using HMC is computationally intractable, which motivates stochastic gradient HMC [51], the zig-zag process [19] and the bouncy particle sampler [65]. In [194], an ADMM-type splitting of the posterior in conjunction with a split Gibbs sampler are proposed, and a dimension-free convergence rate for the split Gibbs sampler is obtained.

Considering the truncation of infinite dimensional probability distributions having a density with respect to a Gaussian reference measure leads to another class of high-dimensional target measures, which arises for instance in path integral MD, cf. [124, 22, 169], and statistical inverse problems, cf. [61]. Dimension-free convergence bounds are obtained for the Metropolis adjusted Langevin Algorithm [82] and for preconditioned Crank Nicholson (pCN) [105]. Moreover, preconditioned HMC was introduced in [16]. The convergence of pHMC was analyzed under strong convexity using a synchronous coupling [167], and by using a two-scale coupling, dimension-free convergence bounds are obtained for semi-discrete pHMC applied to potential energies that are not necessarily globally strongly convex [29].

Another standard approach to analyze convergence properties in high dimension is optimal scaling of MCMC, see [173, 174, 18, 70]. This theory of optimal scaling provides a general way to tune the time step size in HMC [101, 17].

While our object of study is the simplest version of HMC applied to mean-field models, there are other variants of HMC available including one that uses a general reversible approximation of the Hamiltonian dynamics [91], HMC with partial randomization of momentum [113, 5], preconditioned HMC using a position dependent mass matrix [96], and adjusted HMC with delayed rejection [38].

Outline

The rest of the paper is organized as follows. In Appendix A.2, we state the considered framework before presenting our main results in Appendix A.3. In Appendix A.4, estimates used to prove the main results are stated. Finally, Appendix A.5 and Appendix A.6 contain the proofs.

A.2 Preliminaries

We first give the definition of unadjusted HMC applied to mean-field models and state assumptions for the mean-field model before constructing the particlewise coupling used to obtain the contraction result in the next section.

A.2.1 Hamiltonian Monte Carlo Method

Consider a function $U \in \mathcal{C}^2(\mathbb{R}^{dn})$ of the form

$$U(x) = \sum_{i=1}^{n} \Big(V(x^i) + \frac{\epsilon}{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} W(x^i - x^j) \Big) \tag{A.3}$$

such that $\int \exp(-U(x))\mathrm{d}x < \infty$ holds. Assuming all particles have unit masses, the corresponding Hamiltonian is defined by $H(x,v) = U(x) + \frac{1}{2}|v|^2$ for $x, v \in \mathbb{R}^{dn}$. The HMC method is an MCMC method for sampling from a 'target' probability distribution

$$\mu(\mathrm{d}x) = Z^{-1} \exp(-U(x))\mathrm{d}x, \tag{A.4}$$

on $\mathbb{R}^{dn}$ with normalizing constant $Z = \int \exp(-U(x))\mathrm{d}x$. In particular, the HMC method generates a Markov chain on $\mathbb{R}^{dn}$.

Since (A.1) is not exactly solvable, a discretized version is considered. Here, we consider the velocity Verlet integrator with discretization step $h > 0$, cf. [33]. The numerical solution produced by the velocity Verlet integrator is interpolated by the flow $(q_t(x,v), p_t(x,v))$ of the ODE

$$\frac{\mathrm{d}}{\mathrm{d}t} q_t^i = p^i_{\lfloor t \rfloor_h} - \frac{h}{2} \nabla_i U(q_{\lfloor t \rfloor_h}), \qquad \frac{\mathrm{d}}{\mathrm{d}t} p_t^i = -\frac{1}{2}(\nabla_i U(q_{\lfloor t \rfloor_h}) + \nabla_i U(q_{\lceil t \rceil_h})) \tag{A.5}$$

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

$$\lfloor t \rfloor_h = \max\{s \in h\mathbb{Z} : s \leq t\}, \quad \lceil t \rceil_h = \min\{s \in h\mathbb{Z} : s \geq t\},$$

and where $\nabla_i U : \mathbb{R}^{dn} \to \mathbb{R}^d$ is the gradient in the x^i-th direction, i.e., $\frac{\partial U}{\partial x^i}$. The transition step of *unadjusted HMC* is given by $x \mapsto \mathbf{X}_h(x)$ where $\mathbf{X}_h(x) = q_T(x, \xi)$, $T/h \in \mathbb{Z}$ for $h > 0$ and $\xi \sim \mathcal{N}(0, I_{dn})$ is a random variable, where $\mathcal{N}(0, I_{dn})$ denotes the centered normal distribution on $\mathbb{R}^{dn}$ with covariance given by the $dn \times dn$ identity matrix. The transition kernel of the Markov chain on $\mathbb{R}^{dn}$ induced by the unadjusted HMC algorithm is denoted by $\pi_h(x, B) = P[\mathbf{X}_h(x) \in B]$.

If $h > 0$ is fixed, we write the abbreviation $\lfloor t \rfloor$ and $\lceil t \rceil$ instead of $\lfloor t \rfloor_h$ and $\lceil t \rceil_h$ and omit the h dependence in $\mathbf{X}_h(x)$. For $h = 0$ we consider the solution $(q_t(x, \xi), p_t(x, \xi))$ of (A.1) and obtain *exact HMC* with transition step $\mathbf{X}(x) := \mathbf{X}_0(x) = q_T(x, \xi)$ and transition kernel $\pi(x, B) := \pi_0(x, B)$. As the Hamiltonian is not preserved by the numerical flow with $h > 0$, unadjusted HMC does not preserve the target measure μ. Therefore, after we study convergence of unadjusted HMC, we then bound the error between exact and unadjusted HMC in Appendix A.3.

A.2.2 Mean-field particle model

Let $U : \mathbb{R}^{dn} \to \mathbb{R}$ be a potential function of the form (A.3) where $V : \mathbb{R}^d \to \mathbb{R}$ and $W : \mathbb{R}^d \to \mathbb{R}$ are twice continuously differentiable functions such that $\int \exp(-U(x))\mu(\mathrm{d}x) < \infty$. Without loss of generality we assume that ϵ is a non-negative constant. Otherwise we change the sign of the interaction potential W. The following conditions are imposed on the functions V and W for proving the contraction results for exact HMC.

Assumption A.1. *V has a global minimum at 0, $V(0) = 0$ and $V(\mathsf{x}) \geq 0$ for all $\mathsf{x} \in \mathbb{R}^d$.*

Assumption A.2. *V has bounded second derivatives, i.e., $L := \sup \|\nabla^2 V\| < \infty$.*

Assumption A.3. *V is strongly convex outside a Euclidean ball: There exists $K \in (0, \infty)$ and $R \in [0, \infty)$ such that for all $\mathsf{x}, \mathsf{y} \in \mathbb{R}^d$ with $|\mathsf{x} - \mathsf{y}| \geq R$,*

$$(\mathsf{x} - \mathsf{y}) \cdot (\nabla V(\mathsf{x}) - \nabla V(\mathsf{y})) \geq K|\mathsf{x} - \mathsf{y}|^2.$$

Assumption A.4. *W has bounded second derivatives, i.e., $\tilde{L} := \sup \|\nabla^2 W\| < \infty$.*

We note that Assumption A.1 is stated for simplicity, since Assumption A.3 implies that V has a local minimum and so Assumption A.1 can always be obtained by adjusting the coordinate system appropriately and adding a constant to V. Since V is a unary confinement potential per particle and W is a pairwise interaction potential, note that the strong convexity constant K, the Lipschitz constants L, $\tilde{L}$ and the radius R are dimension-free, i.e., independent of the number of particles. By Assumption A.1, Assumption A.2 and Assumption A.4,

$$|\nabla V(\mathsf{x})| = |\nabla V(\mathsf{x}) - \nabla V(0)| \leq L|\mathsf{x}|, \qquad \text{and} \tag{A.6}$$

$$|\nabla W(\mathsf{x} - \mathsf{y}) - \nabla W(\mathsf{y} - \mathsf{x})| \leq 2\tilde{L}|\mathsf{x} - \mathsf{y}| \leq 2\tilde{L}(|\mathsf{x}| + |\mathsf{y}|) \tag{A.7}$$

for all $\mathsf{x}, \mathsf{y} \in \mathbb{R}^d$. From (A.6) and Assumption A.3, it follows that K is smaller than L,

$$K/L \leq 1. \tag{A.8}$$

Further, we deduce from Assumption A.2 and Assumption A.3 that for all $\mathsf{x}, \mathsf{y} \in \mathbb{R}^d$,

$$(\mathsf{x} - \mathsf{y}) \cdot (\nabla V(\mathsf{x}) - \nabla V(\mathsf{y})) \geq K|\mathsf{x} - \mathsf{y}|^2 - \hat{C} \tag{A.9}$$

with $\hat{C} := R^2(L + K)$ and so V is asymptotically strongly convex.

Remark A.1. In this work, we focus on a pairwise mean-field interaction energy W. However, the results can be readily extended to the situation where the Hessian of the mean-field potential U satisfies:

$$L = \sup_{\substack{1\leq i\leq dn \\ x\in\mathbb{R}^{dn}}} \left|\frac{\partial^2 U}{\partial x_i^2}(x)\right| , \quad \tilde{L} = \sup_{\substack{1\leq i<j\leq dn \\ x\in\mathbb{R}^{dn}}} \left|\frac{\partial^2 U}{\partial x_i \partial x_j}(x)\right| \tag{A.10}$$

and the parameter $\tilde{L}$ scales like $1/n$ as $n \to \infty$ which corresponds to the standard mean-field limit [158, 187, 189, 75].

For proving discretization error bounds, we suppose additionally for the confinement potential V and for the interaction potential W:

Assumption A.5. *V is three times differentiable and has bounded third derivatives, i.e., $L_H := \sup \|\nabla^3 V\| < \infty$.*

Assumption A.6. *W is three times differentiable and has bounded third derivatives, i.e., $\tilde{L}_H = \sup \|\nabla^3 W\| < \infty$.*

This additional regularity gives a better order in the error bounds between exact HMC and unadjusted HMC, see Theorem A.8.

Possible interaction potentials meeting Assumption A.4 and Assumption A.6 are the Morse potential [195] and the harmonic (or linear) bonding potential [7, Section 7.4.1.1], which are both used to model interactions between particles in molecular dynamics.

Remark A.2. Let us note that by (A.7) and (A.9) it holds for the potential U that

$$\begin{aligned}
(x-y)\cdot(\nabla U(x)-\nabla U(y)) &= \sum_{i=1}^n \Big((x^i-y^i)\cdot(\nabla V(x^i)-\nabla V(y^i)) \\
&+ \frac{\epsilon}{n}\sum_{j\neq i}(x^i-y^i)\cdot(\nabla W(x^i-x^j)-\nabla W(y^i-y^j)-\nabla W(x^j-x^i)+\nabla W(y^j-y^i))\Big) \\
&\geq K|x-y|^2 - n(K+L)R^2 - \frac{2\epsilon\tilde{L}}{n}\sum_i\sum_{j\neq i}|x^i-y^i-(x^j-y^j)||x^i-y^i| \\
&\geq (K-4\epsilon\tilde{L})|x-y|^2 - n(K+L)R^2.
\end{aligned}$$

Hence, the potential U is strongly convex if $R = 0$ and $K - 4\epsilon\tilde{L} > 0$ holds. Moreover, a similar calculation shows that ∇U is globally Lipschitz continuous with an effective Lipschitz constant of $L+4\epsilon\tilde{L}$. In this case, [31, Theorem 2.1] and [139, Theorem 1] have already shown contraction for exact HMC with the dimension-free rate $c = (1/2)(K-4\epsilon\tilde{L})T^2$ if $(L+4\epsilon\tilde{L})T^2 \leq (K-4\epsilon\tilde{L})/(L+4\epsilon\tilde{L})$ holds. Recently, the latter condition on T has been improved to $(L+4\epsilon\tilde{L})T^2 \leq (1/4)$, cf. [53, Theorem 3]. Whereas, if $R > 0$, then the potential U is only asymptotically strongly convex provided $K - 4\epsilon\tilde{L} > 0$, and in this case,

$$(x-y)\cdot(\nabla U(x)-\nabla U(y)) \geq ((K-4\epsilon\tilde{L})/2)|x-y|^2$$

for all $|x-y| \geq R_n = R\sqrt{2n(L+K)/(K-4\epsilon\tilde{L})}$. Thus, by [31, Theorem 2.3] we obtain the following contraction rate for exact HMC

$$c_n = (1/10)\min(1,(1/4)(K-4\epsilon\tilde{L})T^2(1+(R_n/T))e^{-R_n/(2T)})e^{-2R_n/T} \tag{A.11}$$

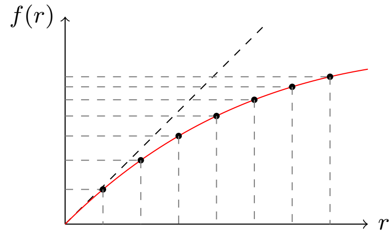

Figure A.1: Under an increasing concave distance function f, a decrease in r has a larger impact on $f(r)$ than an increase in r, i.e., $f(r) - f(r-\Delta) \geq f(r+\Delta) - f(r)$ for $r, \Delta > 0$.

provided $(L+4\epsilon\tilde{L})T^2 \leq \min(1/4, (K-4\epsilon\tilde{L})/(L+4\epsilon\tilde{L}), 1/(2^6(L+4\epsilon\tilde{L})R_n^2))$ holds. The condition on T is dependent on the number n of particles and the rate c_n decreases exponentially fast in the number of particles. This dimension dependence motivates the particlewise coupling stated next.

A.2.3 Construction of coupling

We establish a coupling between the transition probabilities $\pi_h(x, \cdot)$ and $\pi_h(y, \cdot)$ of unadjusted HMC with discretization step h for two states $x, y \in \mathbb{R}^{dn}$. The key idea for the coupling is to locally couple the velocity randomizations, i.e., for the i-th particles in each component of the coupling separately and independently of the other particles. A particlewise coupling approach was used before in [83, 75] and enables us here to show a dimension-free contraction rate, i.e. a rate that does not depend on the number n of particles. The idea for the construction for the i-th particles in each component of the coupling is adapted from [31], see also [85]. The coupling transition step for unadjusted HMC is given by

$$\mathbf{X}(x,y) = q_T(x,\xi) \quad \text{and} \quad \mathbf{Y}(x,y) = q_T(y,\eta) \tag{A.12}$$

with q_T defined in (A.5) and where ξ and η are the corresponding velocity refreshments for the position x and y given in the following way: Let $\xi \in \mathbb{R}^{dn}$ be a normally distributed random variable. Let $\mathcal{U}_i \sim \text{Unif}[0,1]$ be independent uniformly distributed random variables that are independent of ξ. Let γ be a constant that is specified later. If $|x^i - y^i| \geq \tilde{R}$, where $\tilde{R}$ is a positive constant specified later, we apply a synchronous coupling for the i-th particle by setting $\eta^i = \xi^i$. If $|x^i - y^i| < \tilde{R}$, the i-th velocity refreshment of y is given by

$$\eta^i := \begin{cases} \xi^i + \gamma z^i & \text{if } \mathcal{U}_i \leq \frac{\varphi_{0,1}(e^i \cdot \xi^i + \gamma |z^i|)}{\varphi_{0,1}(e^i \cdot \xi^i)}, \\ \xi^i - 2(e^i \cdot \xi^i)e^i & \text{otherwise,} \end{cases} \tag{A.13}$$

where $\varphi_{0,1}$ denotes the density of the standard normal distribution, $z^i = x^i - y^i$, and $e^i = z^i/|z^i|$ if $|z^i| \neq 0$. If $|z^i| = 0$, e^i is some arbitrary unit vector. If we consider the free dynamics, i.e., $U \equiv 0$, then the first case in (A.13) leads to a decrease in the difference of the positions in the i-th component provided the duration T is sufficiently small, i.e., $|\mathbf{X}^i(x,y) - \mathbf{Y}^i(x,y)| = |x^i - y^i||1 - T\gamma|$. When U does not vanish, we obtain contractivity of this coupling in a metric equivalent to the standard ℓ_1 metric that involves a concave distance function, see Figure A.1.

We note that each of the components η^i are normally distributed random variables by [31, Section 2.3] and that the components η^i are independent by the independent particlewise construction. This implies $\eta \sim \mathcal{N}(0, I_{dn})$, which is sufficient to verify that the constructed transition step given by (A.12) is a coupling of the transition probabilities $\pi_h(x, \cdot)$ and $\pi_h(y, \cdot)$.

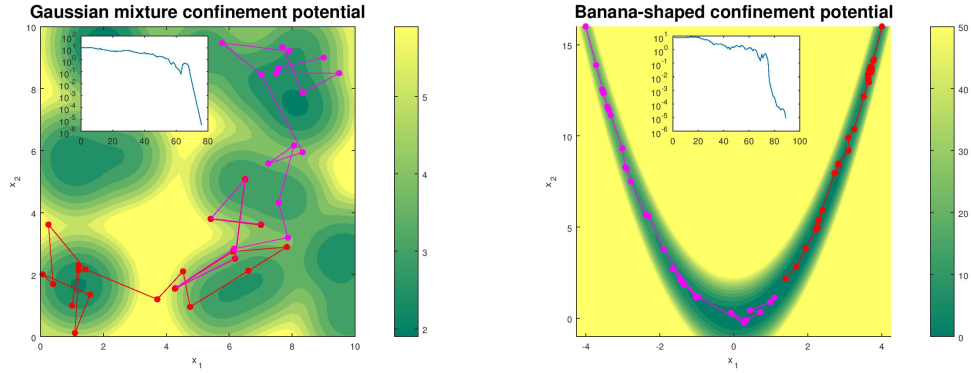

Figure A.2: *Coupling of HMC applied to mean-field models with $n = 10$ particles. The confinement potential is the potential of a Gaussian mixture distribution in the left plot and of a banana-shaped distribution in the right plot. The projection to one particle of the Markov chain is plotted on the contour graph of the potentials and connected by a linear interpolation; the inset shows the mean distance between the two components of the coupling on a log-scale.*

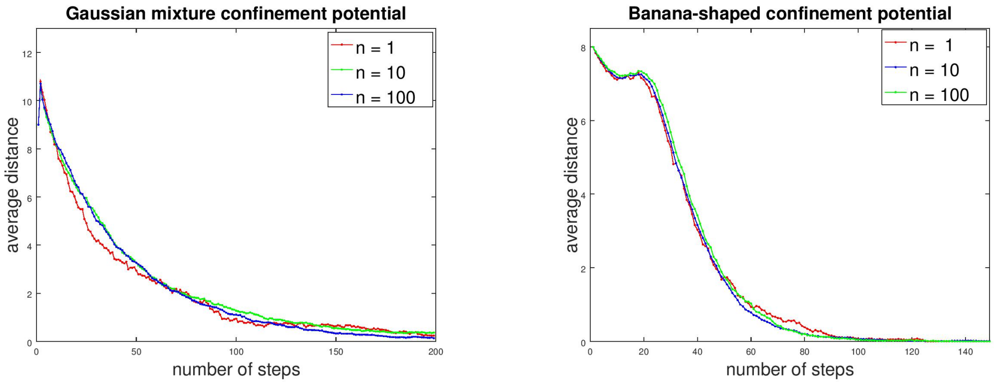

Figure A.3: *Evolution of the mean distance* $\frac{1}{n}\sum_{i=1}^{n}|\mathbf{X}_k^i - \mathbf{Y}_k^i|$ *between the two components of the coupling for HMC after k steps with $n \in \{1, 10, 100\}$ particles.*

A.2.4 Numerical simulations

We next present a numerical illustration of some properties of the particlewise coupling which supports the main results for unadjusted HMC stated in the next section.

We simulate the coupling for mean-field potentials with non-strongly convex confinement potential to illustrate the coupling and to support our theoretical results stated in the next subsection.

We consider two mean-field models with two different confinement potentials. The first potential is the negative logarithm of a Gaussian mixture distribution. Here, we take a mixture of 20 two-dimensional Gaussian distributions whose means are independent uniformly distributed random variables on the rectangle $[0, 10] \times [0, 10]$ and whose covariance matrices are the identity matrix, cf. [132, 125, 31]. The second confinement potential is the negative logarithm of a banana-shaped distribution. In particular, $V : \mathbb{R}^2 \to \mathbb{R}$ is given by the Rosenbrock function $V(\mathsf{x}) = (1 - \mathsf{x}_1)^2 + 10(\mathsf{x}_2 - (\mathsf{x}_1)^2)^2$, cf. [31].

For the interaction between particle i and j, we take the function $W(x^i - x^j) = (1/2)|x^i - x^j|^2$ and $\epsilon = 0.01$ in Figure A.2 and Figure A.3. In Figure A.4, we vary ϵ and W, as indicated in the

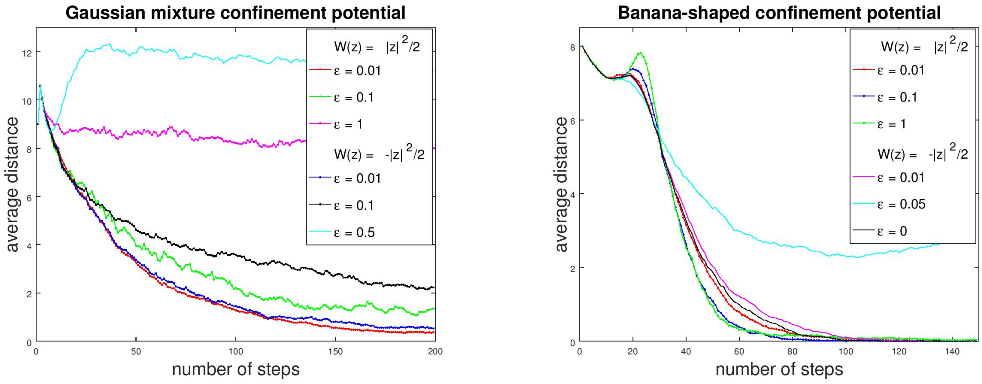

Figure A.4: *Evolution of the mean distance* $\frac{1}{n}\sum_{i=1}^{n}|\mathbf{X}_k^i - \mathbf{Y}_k^i|$ *between the two components of the coupling for HMC after* k *steps with* $n = 10$ *particles for various interaction parameters* ϵ*. This figure suggests that the particlewise coupling does not converge if the interaction is too large.*

legend.

The plots in Figure A.2 show realizations of the coupling with $T = 1$, $\gamma = 1$ and $n = 10$. The evolution of a selected particle of the coupling is drawn on a contour plot of the confinement potential. To visualize the order of the projected points they are connected by linear interpolation. The evolution of the distance function $\frac{1}{n}\sum_{i=1}^{n}|\mathbf{X}_k^i - \mathbf{Y}_k^i|$ is given in the inset. Here, $\mathbf{X}_k^i$ and $\mathbf{Y}_k^i$ are the positions of the i-th particles of the two realizations of the coupling after k HMC steps of duration $T = 1$. The simulation terminates when the distance is smaller than $\tilde{\epsilon} = 10^{-5}$. Figure A.3 shows the sample average of the mean distance $\frac{1}{n}\sum_{i=1}^{n}|\mathbf{X}_k^i - \mathbf{Y}_k^i|$ for different numbers $n \in \{1, 10, 100\}$ of particles. For $n \in \{1, 10\}$ we sampled the mean distance a hundred times and for $n = 100$ thirty times, since the statistical error is smaller for n large. We observe that the mean distance decreases exponentially fast after a short time, which reflects a factor M appearing in the bounds in Corollary A.7 given below, and that the rate is dimension-free, i.e., independent of the number of particles. In Figure A.4, the impact of the size of the interaction parameter ϵ is illustrated. We observe that for small attractive and repulsive interaction the mean coupling distance appears to converge to zero, whereas for larger interaction, particularly for large repulsive interaction (corresponding to $W(x^i - y^i) = -(1/2)|x^i - y^i|^2$) this convergence is not observed.

A.3 Main results

A.3.1 Dimension-free contraction rate for unadjusted HMC

To prove contraction for unadjusted HMC, we introduce a modified distance function. Define

$$\tilde{R} := 8R\sqrt{(L+K)/K}, \tag{A.14}$$

$$\gamma := \min(T^{-1}, \tilde{R}^{-1}/4), \tag{A.15}$$

$$R_1 := (5/4)(\tilde{R} + 2T). \tag{A.16}$$

Note that the constants are dimension-free, i.e. independent of the number of particles. Let $f : \mathbb{R}_+ \to \mathbb{R}_+$ be given by

$$f(r) := \int_0^r \exp(-\min(R_1, s)/T)\mathrm{d}s. \tag{A.17}$$

This function is concave and strictly increasing with $f(0) = 0$ and $f'(0) = 1$. We define a metric $\rho : \mathbb{R}^{dn} \times \mathbb{R}^{dn} \to [0, \infty)$ by

$$\rho(x, y) := \sum_{i=1}^{n} f(|x^i - y^i|). \tag{A.18}$$

This definition is motivated by [83] where it was introduced to obtain optimal contraction rates for weakly interacting diffusions. This metric is equivalent to the ℓ^1-metric,

$$\ell^1(x, y) := \sum_i |x^i - y^i|. \tag{A.19}$$

More precisely, since $rf'(r) \le f(r) \le r$,

$$\rho(x, y) \le \ell^1(x, y) \le M\rho(x, y), \qquad \text{with} \tag{A.20}$$

$$M = f'(R_1)^{-1} = \exp((5/4)(\tilde{R}/T + 2)). \tag{A.21}$$

The following theorem gives a contraction result for unadjusted HMC with respect to the metric ρ.

Theorem A.3 (Global contractivity for unadjusted HMC)**.** *Suppose that Assumption A.1, Assumption A.2, Assumption A.3 and Assumption A.4 hold. Let $\tilde{R}$, γ, R_1 and f be given as in* (A.14)*,* (A.15)*,* (A.16) *and* (A.17)*. Let $T \in (0, \infty)$ and $h_1 \in [0, \infty)$ satisfy*

$$L(T + h_1)^2 \le \frac{3}{5} \min\Big(\frac{1}{4}, \frac{3K}{10L}, \frac{3}{256 \cdot 5L\tilde{R}^2}\Big), \tag{A.22}$$

$$h_1 \le \frac{KT}{525L + 235K}. \tag{A.23}$$

Let $\epsilon \in [0, \infty)$ satisfy

$$\epsilon\tilde{L} < \min\Big(\frac{K}{6}, \frac{1}{2}\Big(\frac{K(\tilde{R} + T)}{36 \cdot 149}\Big)^2 \exp\Big(-5\frac{\tilde{R}}{T}\Big)\Big). \tag{A.24}$$

Then for all $x, y \in \mathbb{R}^{dn}$ and for any $h \in [0, h_1]$ such that $h = 0$ or $T/h \in \mathbb{N}$,

$$\mathbb{E}\Big[\rho(\mathbf{X}(x, y), \mathbf{Y}(x, y))\Big] \le (1 - c)\rho(x, y)$$

with contraction rate

$$c = \frac{1}{156} KT^2 \exp\Big(-\frac{5\tilde{R}}{4T}\Big). \tag{A.25}$$

A proof is given in Appendix A.6.1.

Remark A.4. The parameter c is dimension-free, i.e., independent of the number of particles, which is an improvement compared to the contraction rate given in (A.11) obtained by applying [31, Theorem 2.3]. However, it might depend implicitly on the number of degrees of freedom per particle d through the parameter $\tilde{R}$.

Further, note that the contraction result holds only if the interaction parameter ϵ is sufficiently small. For larger ϵ, contraction with a dimension-free contraction rate is not guaranteed, as illustrated in Figure A.4.

Remark A.5. For adjusted HMC one can show local contraction by precisely bounding the effect of the accept-reject step. The case is considered for a general potential in [31]. In the mean-field model for a large number n of particles, an analogous local contraction result for adjusted HMC is only obtained for a restrictive choice of h. In particular, using the estimate for the rejection probability of [31, Theorem 3.8] the discretization step h has to be chosen of order $\mathcal{O}(n^{-2})$.

Remark A.6. Theorem A.3 holds in particular for the product case with $\epsilon = 0$. As the interaction terms vanish and some calculations simplify in that case, the condition in T becomes $L(T+h_1)^2 \le \min(1/4, K/L, 1/(256L\tilde{R}^2))$ as in [31], the condition in h_1 relaxes to $h_1 \le 4KT/(165L)$ and the contraction rate improves to $c^{prod} = (1/39)KT^2 \exp(-5\tilde{R}/(4T))$. If V is a quadratic function, the mean-field model can be treated as a perturbation of the product model and the difference $|\mathbf{X}^{prod}(x,y) - \mathbf{Y}^{prod}(x,y) - (\mathbf{X}(x,y) - \mathbf{Y}(x,y))|$ of a coupling between to copies of the product model and two copies of the mean-field model can be bounded in terms of $\epsilon\tilde{L}\sum_{i=1}^n |x^i - y^i|$. This term can be controlled for sufficiently small ϵ by the obtained contraction for the product case. See Appendix A.7 for the complete argument.

A.3.2 Quantitative bounds for distance to the target measure

We deduce from Theorem A.3 global contractivity of the transition kernel $\pi_h(x, dy)$ with respect to the Kantorovich distance based on ρ

$$\mathcal{W}_\rho(\nu, \eta) = \inf_{\omega \in \Gamma(\nu,\eta)} \int \rho(x,y)\omega(\mathrm{d}x\mathrm{d}y)$$

on probability measures ν, η on $\mathbb{R}^{dn}$, where $\Gamma(\nu, \eta)$ denotes the set of all couplings of ν and η. Since the metric ρ is equivalent to the ℓ^1-distance ℓ^1 on $(\mathbb{R}^d)^n$ given in (A.19), contractivity with respect to $\mathcal{W}_\rho$ yields a quantitative bound on the Kantorovich distance based on ℓ^1 on $(\mathbb{R}^d)^n$,

$$\mathcal{W}_{\ell^1}(\nu{\pi_h}^m, \mu_h) := \inf_{\omega \in \Gamma(\nu{\pi_h}^m, \mu_h)} \int \sum_{i=1}^n |x^i - y^i| \omega(\mathrm{d}x\mathrm{d}y)$$

between the law after m HMC steps with initial distribution ν and invariant measure μ_h.

Corollary A.7. *Suppose that Assumption A.1, Assumption A.2, Assumption A.3 and Assumption A.4 hold. Let $T \in (0, \infty)$ and $h_1 \in [0, \infty)$ satisfy* (A.22) *and* (A.23). *Let $\epsilon \in [0, \infty)$ satisfy* (A.24). *Then, for any $m \in \mathbb{N}$, for any probability measures ν, η on $\mathbb{R}^{dn}$, and for any $h \in [0, h_1]$ such that $h = 0$ or $T/h \in \mathbb{N}$,*

$$\mathcal{W}_\rho(\nu{\pi_h}^m, \eta{\pi_h}^m) \le e^{-cm}\mathcal{W}_\rho(\nu, \eta), \tag{A.26}$$

$$\mathcal{W}_{\ell^1}(\nu{\pi_h}^m, \eta{\pi_h}^m) \le Me^{-cm}\mathcal{W}_{\ell^1}(\nu, \eta) \tag{A.27}$$

with c given by (A.25) *and M given by* (A.21). *Further, there exists a unique invariant probability measure μ_h on $\mathbb{R}^{dn}$ for the transition kernel π_h of unadjusted HMC and*

$$\mathcal{W}_{\ell^1}(\nu{\pi_h}^m, \mu_h) \le Me^{-cm}\mathcal{W}_{\ell^1}(\nu, \mu_h). \tag{A.28}$$

Thus, for any constant $\tilde{\epsilon} \in (0, \infty)$ and for any initial probability distribution ν the Kantorovich distance $\Delta(m) = \mathcal{W}_{\ell^1}(\nu{\pi_h}^m, \mu_h)$ satisfies $\Delta(m) \le \tilde{\epsilon}$ provided

$$m \ge \frac{1}{c}\Big(\frac{5}{2} + \frac{5\tilde{R}}{4T} + \log\Big(\frac{\Delta(0)}{\tilde{\epsilon}}\Big)\Big). \tag{A.29}$$

A proof is given in Appendix A.6.2. We note that we obtain the same bound as in (A.27) and (A.28) for the Kantorovich distance with respect to the ℓ^1-distance averaged over all particles, $\tilde{\ell}^1(x,y) = \frac{1}{n}\sum_i |x^i - y^i|$. Then, the term $\Delta(0)/\tilde{\epsilon}$ in (A.29) differs by a factor $1/n$. In this case, if we consider for example a product measure as initial distribution, the bound in terms of this metric does not depend logarthmically on the number of particles.

To give quantitative results of the accuracy of unadjusted HMC with respect to the target measure μ, we bound the strong accuracy of velocity Verlet. The exact dynamics started in (x,ξ) with $h=0$ is denoted by $(q_s(x,\xi), p_s(x,\xi))$ and the position of the dynamics started in (x,ξ) with $h>0$ is denoted by $(\tilde{q}_s(x,\xi), \tilde{p}_s(x,\xi))$.

Theorem A.8 (Strong accuracy of velocity Verlet)**.** *Suppose that Assumption A.1, Assumption A.2 and Assumption A.4 hold. Let $T \in (0,\infty)$ satisfy $(L + 4\epsilon\tilde{L})T^2 \le (1/4)$. For $x \in \mathbb{R}^{dn}$, for any $h \in (0,\infty)$ with $T/h \in \mathbb{N}$ and $k \in \mathbb{N}$ with $kh \le T$, it holds*

$$\mathbb{E}_{\xi\sim\mathcal{N}(0,I_{dn})}\Big[\sum_i |q^i_{kh}(x,\xi) - \tilde{q}^i_{kh}(x,\xi)|\Big] \le hC_2\Big(d^{1/2}n + \sum_i |x^i|\Big) \tag{A.30}$$

with C_2 depending on L, $\tilde{L}$, ϵ and T. If additionally Assumption A.5 and Assumption A.6 are supposed, then for $x \in \mathbb{R}^{dn}$, for any $h>0$ with $T/h \in \mathbb{N}$ and $k \in \mathbb{N}$ with $kh \le T$,

$$\mathbb{E}_{\xi\sim\mathcal{N}(0,I_{dn})}\Big[\sum_i |q^i_{kh}(x,\xi) - \tilde{q}^i_{kh}(x,\xi)|\Big] \le h^2\tilde{C}_2\Big(dn + \sum_i |x^i| + \sum_i |x^i|^2\Big) \tag{A.31}$$

with $\tilde{C}_2$ depending on L, $\tilde{L}$, ϵ, L_H, $\tilde{L}_H$ and T.

A proof is given in Appendix A.6.2.

We obtain a bound on the difference between the invariant measure μ_h and the target measure μ, by using the contraction result of Theorem A.3 and by applying a triangle inequality trick, which is mentioned in [141, Remark 6.3] and has been used in many other works. In particular, it holds

$$\begin{aligned}\mathcal{W}_\rho(\mu,\mu_h) = \mathcal{W}_\rho(\mu\pi,\mu_h\pi_h) &\le \mathcal{W}_\rho(\mu\pi,\mu\pi_h) + \mathcal{W}_\rho(\mu\pi_h,\mu_h\pi_h)\\ &\le \mathcal{W}_\rho(\mu\pi,\mu\pi_h) + (1-c)\mathcal{W}_\rho(\mu,\mu_h).\end{aligned}$$

Hence, by (A.20)

$$\mathcal{W}_{\ell^1}(\mu,\mu_h) \le Mc^{-1}\mathcal{W}_{\ell^1}(\mu\pi,\mu\pi_h) \le Mc^{-1}\mathbb{E}_{x\sim\mu,\ \xi\sim\mathcal{N}(0,I_{dn})}\Big[\sum_i |q^i_{kh}(x,\xi) - \tilde{q}^i_{kh}(x,\xi)|\Big]$$

with M given in (A.21). Inserting (A.30), respectively (A.31), yields the following result.

Corollary A.9 (Asymptotic Bias)**.** *Suppose that Assumption A.1, Assumption A.2, Assumption A.3 and Assumption A.4 hold. Let T and h_1 satisfy* (A.22)*. Let ϵ satisfy* (A.24)*. Let C_2 and $\tilde{C}_2$ be as in Theorem A.8. Then for $h \in (0,h_1]$ with $T/h \in \mathbb{N}$,*

$$\mathcal{W}_{\ell^1}(\mu,\mu_h) \le hc^{-1}MC_2\Big(d^{1/2}n + \int_{\mathbb{R}^{nd}}\sum_i |x^i|\mu(\mathrm{d}x)\Big)$$

with c given by (A.25) *and M given by* (A.21)*. If additionally Assumption A.5 and Assumption A.6 are assumed, then for $h \in (0,h_1]$ with $T/h \in \mathbb{N}$,*

$$\mathcal{W}_{\ell^1}(\mu,\mu_h) \le h^2c^{-1}M\tilde{C}_2\Big(dn + \int_{\mathbb{R}^{nd}}\sum_i |x^i|\mu(\mathrm{d}x) + \int_{\mathbb{R}^{nd}}\sum_i |x^i|^2\mu(\mathrm{d}x)\Big).$$

Note that the bound in Corollary A.9 is linear in the number n of particles.

For unadjusted HMC, Corollary A.7 gives exponential convergence to the invariant measure μ_h. In the next theorem, we give a bound on the number of steps to reach the target measure μ up to a given error.

Theorem A.10 (Complexity Guarantee)**.** *Suppose that Assumption A.1, Assumption A.2, Assumption A.3 and Assumption A.4 hold. Let $T \in (0,\infty)$ and $h_1 \in (0,\infty)$ satisfy* (A.22) *and* (A.23) *. Let $\epsilon \in [0,\infty)$ satisfy* (A.24)*. Let ν be a probability measure on $\mathbb{R}^{dn}$, and let $\Delta(m) = \mathcal{W}_{\ell^1}(\nu {\pi_h}^m, \mu)$ denote the Kantorovich distance with respect to ℓ^1 to the target probability measure μ after m steps with initial distribution ν. For some $\tilde{\epsilon} \in (0,\infty)$, let $m \in \mathbb{N}$ be such that*

$$m \geq \frac{1}{c}\Big(\frac{5}{2} + \frac{5R}{4T} + \log\Big(\frac{2\mathcal{W}_{\ell^1}(\mu_h,\nu)}{\tilde{\epsilon}}\Big)^+\Big) \tag{A.32}$$

with c given by (A.25)*. Then, there exists h_2 such that for $h \in (0, \min(h_1,h_2)]$ with $T/h \in \mathbb{Z}$,*

$$\Delta(m) \leq \tilde{\epsilon} \tag{A.33}$$

where for fixed K, L, $\tilde{L}$, ϵ, R and T, h_2^{-1} is of order $\mathcal{O}(\tilde{\epsilon}^{-1}(d^{1/2}n + \int \sum_i |x^i|\mu(\mathrm{d}x)))$. If additionally Assumption A.5 and Assumption A.6 are assumed, then there exists $\tilde{h}_2$ such that for $h \in (0, \min(h_1, \tilde{h}_2)]$ with $T/h \in \mathbb{Z}$, (A.33) *holds, where for fixed K, L, $\tilde{L}$, L_H, $\tilde{L}_H$, ϵ, R and T, $\tilde{h}_2^{-1}$ is of order $\mathcal{O}(\tilde{\epsilon}^{-1/2}((nd)^{1/2} + \sqrt{\int \sum_i |x^i|\mu(\mathrm{d}x)} + \sqrt{\int \sum_i |x^i|^2\mu(\mathrm{d}x)}))$.*

A proof is given in Appendix A.6.2. If we consider the averaged distance $\tilde{\ell}^1$ instead of ℓ^1, the argument in the logarithmic term in (A.32) changes by a factor $1/n$ and the logarithmic dependence on n in h_2 and $\tilde{h}_2$ vanishes.

Remark A.11. We note that h^{-1} is $\mathcal{O}(n^{1/2})$ in Theorem A.10 and hence it grows sublinear in n. Further, the constant $\tilde{C}_2$ obtained in the proof of Theorem A.8 is $\mathcal{O}(T^{-1})$. For the numerical method ULA, which forms a special case of unadjusted HMC with $h = T$ (see [156, Section 5.2]), we obtain that $h^{-1} = T^{-1}$ has to be chosen of order $\mathcal{O}(n)$, which corresponds to the results in [73, Example 18].

Remark A.12. From Theorem A.10, note that the number of evaluations of the gradient $\nabla U(x)$ in each step of duration T is $\mathcal{O}(n^{1/2})$ for fixed K, L, $\tilde{L}$, ϵ, T, R, d and h. If we assume that the computation of the gradient in one step is $\mathcal{O}(n)$, then the overall complexity of unadjusted HMC is $\mathcal{O}(n^{3/2})$.

A.3.3 Dimension-free bounds for ergodic averages of intensive observables

Next, we define the ergodic averages $A_{m,b}g$, which approximate $\mu(g) = \int g(x)\mu(\mathrm{d}x)$, by

$$A_{m,b}g := \frac{1}{m}\sum_{i=b}^{b+m-1} g(\mathbf{X}_i), \tag{A.34}$$

for some function $g : \mathbb{R}^{dn} \to \mathbb{R}$ and for $b, m \in \mathbb{N}$, where $(\mathbf{X}_n)$ is the Markov chain given by unadjusted HMC. The parameter b corresponds to the burn-in time. Here, we consider bounded and continuously differentiable observables, i.e., $g \in \mathcal{C}_b^1(\mathbb{R}^{dn})$. Quantitative bounds on the bias of the ergodic averages follow by the exponential convergence in the Kantorovich distance with respect to the ℓ^1 metric given in (A.19) and the bound on the accuracy of unadjusted HMC.

Theorem A.13 (Bias of Ergodic Averages)**.** *Let $g \in \mathcal{C}_b^1(\mathbb{R}^{dn})$ with $\max_i \|\nabla_i g\|_\infty < \infty$. Suppose that Assumption A.1, Assumption A.2, Assumption A.3 and Assumption A.4 hold. Let $T \in (0,\infty)$ and $h_1 \in [0,\infty)$ satisfy* (A.22) *and* (A.23)*. Let $\epsilon \in [0,\infty)$ satisfy* (A.24)*. Let ν be a probability measure on $\mathbb{R}^{dn}$. Let C_2 and $\tilde{C}_2$ be given as in Theorem A.8, and let c be given as in* (A.11)*. Then for $h \in [0,h_1]$ such that $h=0$ or $T/h \in \mathbb{N}$,*

$$|\mathbb{E}_\nu[A_{m,b}g] - \mu(g)| \le \frac{1}{m} \max_i \|\nabla_i g\|_\infty \frac{e^{-cb}}{1-e^{-c}} \mathcal{W}_{\ell^1}(\nu,\mu_h) + h \max_i \|\nabla_i g\|_\infty C_3,$$

where $C_3 = \exp(\frac{5}{4}(2+\tilde{R}/T))c^{-1}C_2\Big(d^{1/2}n + \int \sum_i |x^i|\mu(\mathrm{d}x)\Big)$. If additionally Assumption A.5 and Assumption A.6 are supposed, then

$$|\mathbb{E}_\nu[A_{m,b}g] - \mu(g)| \le \frac{1}{m} \max_i \|\nabla_i g\|_\infty \frac{e^{-cb}}{1-e^{-c}} \mathcal{W}_{\ell^1}(\nu,\mu_h) + h^2 \max_i \|\nabla_i g\|_\infty \tilde{C}_3,$$

where $\tilde{C}_3 = \exp(\frac{5}{4}(2+\tilde{R}/T))c^{-1}\tilde{C}_2\Big(dn + \int \sum_i |x^i|\mu(\mathrm{d}x) + \int \sum_i |x^i|^2\mu(\mathrm{d}x)\Big)$.

A proof is given in Appendix A.6.3.

Remark A.14. We note that provided $\max_i \|\nabla_i g\|_\infty$ is $\mathcal{O}(1/n)$ the bound of the bias of the ergodic averages is independent of the number n of particles. Hence for intensive observables of the form $g(x) = \frac{1}{n}\sum_i \hat{g}(x^i)$ where $\hat{g} \in \mathcal{C}_b^1(\mathbb{R}^d)$ with $\|\nabla \hat{g}\|_\infty < \infty$, Theorem A.13 gives quantitative bounds on the bias of their ergodic averages which are dimension-free, i.e., independent of the number n of particles. Whereas, for extensive observables, where $\max_i \|\nabla_i g\|_\infty$ is $\mathcal{O}(1)$, the bound depends on the number n of particles.

A.4 Estimates for the Hamiltonian dynamics

A.4.1 Deviation from free dynamics

Here we apply the Lipschitz conditions in Assumption A.2 and Assumption A.4 to obtain bounds on how far the dynamics in (A.5) deviates from the free dynamics, $U \equiv 0$. To obtain these bounds, we assume in the following that $t,h \in [0,\infty)$ are such that $t/h \in \mathbb{N}$ for $h>0$ and such that

$$(L + 4\epsilon\tilde{L})(t^2 + th) \le 1. \tag{A.35}$$

This condition essentially states that the duration of the Hamiltonian dynamics in (A.5) is small with respect to the fastest characteristic time-scale of the mean-field particle system represented by $\sqrt{\sup \|\mathrm{Hess}U\|} \le \sqrt{L + 4\epsilon\tilde{L}}$ (see Remark A.2). This bound follows from Assumption A.2 and Assumption A.4. The i-th component of the solution to (A.5) is denoted by (x_s^i, v_s^i).

Lemma A.15. *Let $x, v \in \mathbb{R}^{dn}$. Then for $i \in \{1,...,n\}$,*

$$\max_{s\le t} |x_s^i| \le (1 + (L+2\epsilon\tilde{L})(t^2+th)) \max(|x^i|, |x^i + tv^i|) + \frac{2\epsilon\tilde{L}(t^2+th)}{n} \max_{s\le t} \sum_{j\ne i} |x_s^j|, \tag{A.36}$$

$$\max_{s\le t} |v_s^i| \le |v^i| + (L+2\epsilon\tilde{L})t(1 + (L+2\epsilon\tilde{L})(t^2+th)) \max(|x^i|, |x^i+tv^i|) + \frac{2\epsilon\tilde{L}t}{n}(1 + (L+2\epsilon\tilde{L})(t^2+th)) \max_{s\le t} \sum_{j\ne i} |x_s^j|. \tag{A.37}$$

Moreover,

$$\max_{s\le t}\sum_i |x_s^i| \le (1+(L+4\epsilon\tilde{L})(t^2+th))\sum_i \max(|x^i|,|x^i+tv^i|), \tag{A.38}$$

$$\max_{s\le t}\sum_i |v_s^i| \le (L+4\epsilon\tilde{L})t(1+(L+4\epsilon\tilde{L})(t^2+th))\sum_i \max(|x^i|,|x^i+tv^i|)+\sum_i |v^i|. \tag{A.39}$$

A proof of Lemma A.15 is provided in Appendix A.5.

Let two processes (x_s,v_s), (y_s,u_s) with initial values (x,v) and (y,u) be driven by the Hamiltonian dynamics in (A.5). We set $(z_s,w_s):=(x_s-y_s,v_s-u_s)$. Since (x_s,v_s) and (y_s,u_s) depend on (x,v) and (y,u), respectively, (z_s,w_s) depends on (x,v,y,u). By (A.5), the dynamics of the i-th component of (z_s,w_s) is given by

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}z_t^i &= w_{\lfloor t\rfloor}^i-(h/2)(\nabla_i U(x_{\lfloor t\rfloor})-\nabla_i U(y_{\lfloor t\rfloor}))\\ \frac{\mathrm{d}}{\mathrm{d}t}w_t^i &= (1/2)(-\nabla_i U(x_{\lfloor t\rfloor})-\nabla_i U(x_{\lceil t\rceil})+\nabla_i U(y_{\lfloor t\rfloor})+\nabla_i U(y_{\lceil t\rceil})).\end{aligned} \tag{A.40}$$

Next, we bound the distance between the process (z_s^i,w_s^i) and the process given by the free dynamics, where $U\equiv 0$. As the particlewise coupling in Appendix A.2.3 is designed with respect to the free dynamics, this bound plays an important role in proving the contraction results of Appendix A.3. It explains why the particlewise coupling works when the distance between i-th particles is small, i.e., when $|x^i-y^i|<\tilde{R}$, and when the duration T and the time step h are small, i.e., when (A.35) is assumed.

Lemma A.16. *Let $x,y,v,u\in\mathbb{R}^{dn}$. Then for all $i\in\{1,...,n\}$,*

$$\begin{aligned}\max_{s\le t}|z_s^i-z^i-sw^i| &\le (L+2\epsilon\tilde{L})(t^2+th)\max(|z^i+tw^i|,|z^i|)\\ &\quad+\frac{2\epsilon\tilde{L}(t^2+th)}{n}\max_{s\le t}\sum_{j\neq i}|z_s^j|,\end{aligned} \tag{A.41}$$

$$\begin{aligned}\max_{s\le t}|z_s^i| &\le (1+(L+2\epsilon\tilde{L})(t^2+th))\max(|z^i+tw^i|,|z^i|)\\ &\quad+\frac{2\epsilon\tilde{L}(t^2+th)}{n}\max_{s\le t}\sum_{j\neq i}|z_s^j|,\end{aligned} \tag{A.42}$$

$$\begin{aligned}\max_{s\le t}|w_s^i-w^i| &\le (L+2\epsilon\tilde{L})t(1+(L+2\epsilon\tilde{L})(t^2+th))\max(|z^i+tw^i|,|z^i|))\\ &\quad+\frac{2\epsilon\tilde{L}t}{n}(1+(L+2\epsilon\tilde{L})(t^2+th))\max_{s\le t}\sum_{j\neq i}|z_s^j|,\end{aligned} \tag{A.43}$$

$$\begin{aligned}\max_{s\le t}|w_s^i| &\le |w^i|+(L+2\epsilon\tilde{L})t(1+(L+2\epsilon\tilde{L})(t^2+th))\max(|z^i+tw^i|,|z^i|))\\ &\quad+\frac{2\epsilon\tilde{L}t}{n}(1+(L+2\epsilon\tilde{L})(t^2+th))\max_{s\le t}\sum_{j\neq i}|z_s^j|.\end{aligned} \tag{A.44}$$

Moreover,

$$\max_{s\le t}\sum_i |z_s^i| \le (1+(L+4\epsilon\tilde{L})(t^2+th))\sum_i \max(|z^i+tw^i|,|z^i|), \tag{A.45}$$

$$\max_{s\le t}\sum_i |w_s^i| \le (L+4\epsilon\tilde{L})t(1+(L+4\epsilon\tilde{L})(t^2+th))\sum_i \max(|z^i+tw^i|,|z^i|)+\sum_i |w^i|. \tag{A.46}$$

A proof of Lemma A.16 is provided in Appendix A.5.

A.4.2 Bounds in region of strong convexity

Next, we obtain a bound for the difference between the positions of the i-th particles provided that $|x^i - y^i| > \tilde{R}$ and $v^i = u^i$. We assume that

$$(L + 4\epsilon\tilde{L})(t^2 + th) \leq \min\Big(\frac{\kappa}{L + 4\epsilon\tilde{L}}, \frac{1}{4}\Big), \tag{A.47}$$

where κ is given by

$$\kappa := K - 3\epsilon\tilde{L}. \tag{A.48}$$

Further, we assume that

$$h \leq \frac{Kt}{525L + 235K}. \tag{A.49}$$

Lemma A.17. *Suppose that Assumption A.1, Assumption A.2, Assumption A.3 and Assumption A.4 hold. Let $\epsilon \in [0, \infty)$ be such that $\epsilon\tilde{L} < K/6$ holds. Let $\tilde{R}$ be given in* (A.14)*. Let $t, h \in [0, \infty)$ be such that $h = 0$ or $t/h \in \mathbb{N}$, and such that* (A.47) *and* (A.49) *holds. Then, for all $x, y, v, u \in \mathbb{R}^{dn}$ and $i \in \{1, ..., n\}$ such that $|x^i - y^i| \geq \tilde{R}$ and $v^i = u^i$,*

$$|x^i_t - y^i_t|^2 \leq \Big(1 - \frac{1}{4}\kappa t^2\Big)|x^i - y^i|^2 + 2\frac{\epsilon\tilde{L}t^2}{n^2}\Big(\max_{s\leq t}\sum_{j\neq i}|x^j_s - y^j_s|\Big)^2. \tag{A.50}$$

A proof of Lemma A.17 is given in Appendix A.5.

A.5 Proofs of results from Section A.4

Before stating the proofs of Appendix A.4, note that by (A.6) and (A.7) for all $x, y \in \mathbb{R}^{dn}$,

$$|\nabla_i U(x)| \leq L|x^i| + \frac{2\epsilon\tilde{L}}{n}\sum_{j\neq i}|x^i - x^j| \leq (L + 2\epsilon\tilde{L})|x^i| + \frac{2\epsilon\tilde{L}}{n}\sum_{j\neq i}|x^j|, \tag{A.51}$$

and by Assumption A.2 and Assumption A.4

$$|\nabla_i U(x) - \nabla_i U(y)| \leq (L + 2\epsilon\tilde{L})|x^i - y^i| + \frac{2\epsilon\tilde{L}}{n}\sum_{j\neq i}|x^j - y^j|. \tag{A.52}$$

Further by (A.9) and (A.7), it holds for all $x, y \in \mathbb{R}^{dn}$,

$$\begin{aligned} -(x^i - y^i)\cdot(\nabla_i U(x) - \nabla_i U(y)) &\leq -(K - 2\epsilon\tilde{L})|x^i - y^i|^2 + 2\epsilon\tilde{L}|x^i - y^i|\frac{1}{n}\sum_{j\neq i}|x^j - y^j| + \hat{C} \\ &\leq -\kappa|x^i - y^i|^2 + \epsilon\tilde{L}\Big(\frac{1}{n}\sum_{j\neq i}|x^j - y^j|\Big)^2 + \hat{C}. \end{aligned} \tag{A.53}$$

It follows from the definition (A.14) of $\tilde{R}$ and the condition $\epsilon\tilde{L} < K/6$, which is assumed in Lemma A.17, that for all $\mathsf{x}, \mathsf{y} \in \mathbb{R}^d$ with $|\mathsf{x} - \mathsf{y}| \geq \tilde{R}$,

$$\hat{C} = R^2(L + K) < \frac{1}{64}K|\mathsf{x} - \mathsf{y}|^2 \leq \frac{1}{32}\kappa|\mathsf{x} - \mathsf{y}|^2. \tag{A.54}$$

Proof of Lemma A.15. Fix $x, v \in \mathbb{R}^{dn}$. Let $s \leq t$. We have from (A.5)

$$x_s^i - x^i - sv^i = \int_0^s \int_0^{\lfloor r \rfloor} \Big(-\frac{1}{2}\nabla_i U(x_{\lfloor u \rfloor}) - \frac{1}{2}\nabla_i U(x_{\lceil u \rceil})\Big) \mathrm{d}u \ \mathrm{d}r - \int_0^s \frac{h}{2}\nabla_i U(x_{\lfloor r \rfloor}) \mathrm{d}r.$$

We apply (A.51) to obtain

$$\begin{aligned}|x_s^i - x^i - sv^i| \leq \frac{(L+2\epsilon\tilde{L})(t^2+th)}{2} \max_{r\leq t}(|x_r^i - x^i - rv^i| + |x^i + rv^i|) \\ + \frac{2\epsilon\tilde{L}(t^2+th)}{2n} \max_{r\leq t} \sum_{j\neq i} |x_r^j|.\end{aligned}$$

Invoking condition (A.35), we get

$$\begin{aligned}\max_{s\leq t} |x_s^i - x^i - sv^i| &\leq (L+2\epsilon\tilde{L})(t^2+th) \max_{s\leq t} |x^i + sv^i| + \frac{2\epsilon\tilde{L}(t^2+th)}{n} \max_{s\leq t} \sum_{j\neq i} |x_s^j| \\ &= (L+2\epsilon\tilde{L})(t^2+th) \max(|x^i|, |x^i+tv^i|) + \frac{2\epsilon\tilde{L}(t^2+th)}{n} \max_{s\leq t} \sum_{j\neq i} |x_s^j|.\end{aligned}$$

By applying the triangle inequality, (A.36) is obtained. From (A.5) and (A.51), we have

$$|v_s^i - v^i| \leq \int_0^s \max_{u\leq t} |\nabla_i U(x_u)| \mathrm{d}r \leq (L+2\epsilon\tilde{L})t \max_{u\leq t} |x_u^i| + \frac{2\epsilon\tilde{L}t}{n} \max_{u\leq t} \sum_{j\neq i} |x_u^j|. \tag{A.55}$$

We insert (A.36) in (A.55) to obtain

$$\begin{aligned}|v_s^i - v^i| &\leq (L+2\epsilon\tilde{L})t(1+(L+2\epsilon\tilde{L})(t^2+th)) \max(|x^i|, |x^i+tv^i|) \\ &+ \frac{2\epsilon\tilde{L}t}{n}(1+(L+2\epsilon\tilde{L})(t^2+th)) \max_{u\leq t} \sum_{j\neq i} |x_u^j|.\end{aligned}$$

By applying the triangle inequality, (A.37) is obtained. Equation (A.38) and (A.39) follow by considering the sum over all particles, i.e., by (A.5) we have

$$\begin{aligned}\sum_i |x_s^i - x^i - sv^i| &\leq \int_0^s \int_0^r \frac{1}{2} \sum_i |\nabla_i U(x_{\lfloor u \rfloor}) + \nabla_i U(x_{\lceil u \rceil})| \mathrm{d}u\mathrm{d}r + \frac{h}{2}\int_0^s \sum_i |\nabla_i U(x_{\lfloor r \rfloor})| \mathrm{d}r \\ &\leq \frac{(L+4\epsilon\tilde{L})(t^2+th)}{2} \max_{r\leq t} \Big(\sum_i |x_r^i|\Big)\end{aligned}$$

and hence analogous to the estimate obtained for the i-th particle,

$$\begin{aligned}\max_{s\leq t} \sum_i |x_s^i - x^i - sv^i| &\leq (L+4\epsilon\tilde{L})(t^2+th) \max_{r\leq t} \sum_i |x^i + rv^i| \\ &\leq (L+4\epsilon\tilde{L})(t^2+th) \sum_i \max(|x^i|, |x^i+tv^i|).\end{aligned}$$

By applying the triangle inequality, (A.38) is obtained. By (A.5) and (A.38),

$$\begin{aligned}\sum_i |v_s^i - v^i| &\leq (L+4\epsilon\tilde{L})t \max_{r\leq t} \Big(\sum_i |x_r^i|\Big) \\ &\leq (L+4\epsilon\tilde{L})t(1+(L+4\epsilon\tilde{L})(t^2+th)) \sum_i \max(|x^i|, |x^i+tv^i|),\end{aligned}$$

and (A.39) is obtained by the triangle inequality. □

Proof of Lemma A.16. By (A.52) and (A.40),

$$
\begin{aligned}
&|z_s^i - z^i - sw^i| \\
&\le \int_0^s \int_0^r \max_{v \le t} |-\nabla_i U(x_v) + \nabla_i U(y_v)| \mathrm{d}u \ \mathrm{d}r + \frac{h}{2} \int_0^s \max_{v \le t} |-\nabla_i U(x_v) + \nabla_i U(y_v)| \mathrm{d}r \\
&\le \frac{(L + 2\epsilon \tilde{L})(t^2 + th)}{2} \max_{r \le t} |z_r^i| + \frac{2\epsilon \tilde{L}(t^2 + th)}{2n} \max_{r \le t} \sum_{j \ne i} |z_r^j|.
\end{aligned}
$$

Hence, we obtain similar to the previous proof

$$
\max_{s \le t} |z_s^i - z^i - sw^i| \le (L + 2\epsilon \tilde{L})(t^2 + th) \max(|z^i|, |z^i + tw^i|) + \frac{2\epsilon \tilde{L}(t^2 + th)}{n} \max_{s \le t} \sum_{j \ne i} |z_s^j|,
$$

which gives (A.41). Then (A.42) is obtained by applying triangle inequality. Next, we consider

$$
\begin{aligned}
|w_s^i - w^i| &\le \frac{1}{2} \int_0^s (|-\nabla_i U(x_{\lfloor r \rfloor}) + \nabla_i U(y_{\lfloor r \rfloor})| + |-\nabla_i U(x_{\lceil r \rceil}) + \nabla_i U(y_{\lceil r \rceil})|) \mathrm{d}r \\
&\le (L + 2\epsilon \tilde{L}) t \max_{r \le t} |z_r^i| + \frac{2\epsilon \tilde{L} t}{n} \max_{r \le t} \sum_{j \ne i} |z_r^j|,
\end{aligned}
$$

where we again used (A.52) and (A.40). Hence, we obtain by (A.42),

$$
\begin{aligned}
\max_{s \le t} |w_s^i - w^i| \le{}& (L + 2\epsilon \tilde{L}) t (1 + (L + 2\epsilon \tilde{L})(t^2 + th)) \max(|z^i|, |z^i + tw^i|) \\
&+ \frac{2\epsilon \tilde{L} t}{n} (1 + (L + 2\epsilon \tilde{L})(t^2 + th)) \max_{s \le t} \sum_{j \ne i} |z_s^j|,
\end{aligned}
$$

which gives (A.43) and (A.44). Estimates (A.45) and (A.46) hold similarly by considering the sum over all particles instead of considering only the i-th particle. □

Proof of Lemma A.17. As before, write $(z_s, w_s) = (x_s - y_s, v_s - u_s)$ whose dynamics is given by (A.40). Then, $z_0 = x - y$ and $w_0^i = 0$ since the velocities of the i-th component are synchronized.

Define $a^i(t) = |z_t^i|^2$ and $b^i(t) = 2z_t^i \cdot w_t^i$. We set up an initial value problem of the two deterministic processes $a^i(t)$ and $b^i(t)$ and solve it to obtain the required bound for $a^i(t)$. By (A.40), we have

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} a^i(t) &= b^i(t) + 2z_t^i \cdot (w_{\lfloor t \rfloor}^i - w_t^i) - hz_t^i \cdot (\nabla_i U(x_{\lfloor t \rfloor}) - \nabla_i U(y_{\lfloor t \rfloor})) = b^i(t) + \delta^i(t) \\
\frac{\mathrm{d}}{\mathrm{d}t} b^i(t) &= -z_t^i \cdot (\nabla_i U(x_{\lfloor t \rfloor}) - \nabla_i U(y_{\lfloor t \rfloor}) + \nabla_i U(x_{\lceil t \rceil}) - \nabla_i U(y_{\lceil t \rceil})) \\
&\quad + 2w_t^i \cdot w_{\lfloor t \rfloor}^i - hw_t^i \cdot (\nabla_i U(x_{\lfloor t \rfloor}) - \nabla_i U(y_{\lfloor t \rfloor})) \\
&= -z_{\lfloor t \rfloor}^i \cdot (\nabla_i U(x_{\lfloor t \rfloor}) - \nabla_i U(y_{\lfloor t \rfloor})) - z_{\lceil t \rceil}^i \cdot (\nabla_i U(x_{\lceil t \rceil}) - \nabla_i U(y_{\lceil t \rceil})) \\
&\quad + 2|w_t^i|^2 - 2\kappa |z_t^i|^2 + \kappa(|z_{\lfloor t \rfloor}^i|^2 + |z_{\lceil t \rceil}^i|^2) + \varepsilon^i(t)
\end{aligned}
$$

where $\varepsilon^i(t) = \varepsilon_1^i(t) + \varepsilon_2^i(t) + \varepsilon_3^i(t) + \varepsilon_4^i(t)$ and

$$\begin{aligned}
\delta^i(t) &= z_t^i \cdot (2(w_{\lfloor t\rfloor}^i - w_t^i) - h(\nabla_i U(x_{\lfloor t\rfloor}) - \nabla_i U(y_{\lfloor t\rfloor})))\\
\varepsilon_1^i(t) &= -(z_t^i - z_{\lfloor t\rfloor}^i)\cdot(\nabla_i U(x_{\lfloor t\rfloor}) - \nabla_i U(y_{\lfloor t\rfloor}))\\
\varepsilon_2^i(t) &= -(z_t^i - z_{\lceil t\rceil}^i)\cdot(\nabla_i U(x_{\lceil t\rceil}) - \nabla_i U(y_{\lceil t\rceil}))\\
\varepsilon_3^i(t) &= w_t^i \cdot (2(w_{\lfloor t\rfloor}^i - w_t^i) - h(\nabla_i U(x_{\lfloor t\rfloor}) - \nabla_i U(y_{\lfloor t\rfloor})))\\
\varepsilon_4^i(t) &= \kappa(2|z_t^i|^2 - |z_{\lfloor t\rfloor}^i|^2 - |z_{\lceil t\rceil}^i|^2).
\end{aligned}$$

By (A.53) the derivative of $b^i(t)$ is bounded by

$$\frac{\mathrm{d}}{\mathrm{d}t} b^i(t) \le -2\kappa|z_t^i|^2 + \frac{\epsilon\tilde{L}}{n^2}\Big(\sum_{j\neq i}|z_{\lfloor t\rfloor}^j|\Big)^2 + \frac{\epsilon\tilde{L}}{n^2}\Big(\sum_{j\neq i}|z_{\lceil t\rceil}^j|\Big)^2 + 2|w_t^i|^2 + \varepsilon^i(t) + 2\hat{C}.$$

The previous estimate leads to an initial value problem of the form

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} a^i(t) &= b^i(t) + \delta^i(t), && a^i(0) = |z_0^i|^2,\\
\frac{\mathrm{d}}{\mathrm{d}t} b^i(t) &= -2\kappa a^i(t) + \beta^i(t) + \varepsilon^i(t), && b^i(0) = 0,
\end{aligned}$$

where

$$\beta^i(t) \le \frac{\epsilon\tilde{L}}{n^2}\Big(\sum_{j\neq i}|z_{\lfloor t\rfloor}^j|\Big)^2 + \frac{\epsilon\tilde{L}}{n^2}\Big(\sum_{j\neq i}|z_{\lceil t\rceil}^j|\Big)^2 + 2|w_t^i|^2 + 2\hat{C}. \tag{A.56}$$

Note that when $h = 0$, $\varepsilon^i(t) = \delta^i(t) = 0$. By variation of parameters, $a^i(t)$ can be written as

$$\begin{aligned}
a^i(t) = \cos(\sqrt{2\kappa}\ t)|z_0^i|^2 &+ \int_0^t \cos(\sqrt{2\kappa}(t-r))\delta^i(r)\mathrm{d}r\\
&+ \int_0^t \frac{1}{\sqrt{2\kappa}}\sin(\sqrt{2\kappa}(t-r))(\beta^i(r) + \varepsilon^i(r))\mathrm{d}r.
\end{aligned} \tag{A.57}$$

Taylor's integral formula, i.e., $\cos(\sqrt{2\kappa}\ t) = 1 - \kappa t^2 + (1/6)\int_0^t (t-s)^3 \cos(\sqrt{2\kappa}\ s)(2\kappa)^2 ds \le 1 - \kappa t^2 + \kappa^2 t^4/6$, and the fact that by (A.47) and (A.8) $\kappa^2 t^4 \le (L + 2\epsilon\tilde{L})^2 t^4 \le \kappa t^2$ yield

$$\cos(\sqrt{2\kappa}t) \le 1 - (5/6)\kappa t^2. \tag{A.58}$$

Further, we get by (A.47) and (A.8)

$$\kappa t^2 \le (L + 2\epsilon\tilde{L})t^2 \le 1 \le \pi^2/2, \text{ and so } t \le (\pi/\sqrt{2\kappa}). \tag{A.59}$$

Therefore, $\sin(\sqrt{2\kappa}(t-r)) \ge 0$ for all $r \in [0,t]$. Further,

$$\frac{1}{\sqrt{2\kappa}}\sin(\sqrt{2\kappa}(t-r)) \le (t-r). \tag{A.60}$$

Inserting (A.58) and (A.60) in (A.57) yields

$$a^i(t) \le (1 - (5/6)\kappa t^2)|z_0^i|^2 + \int_0^t |\delta^i(r)|\mathrm{d}r + \int_0^t (t-r)(|\beta^i(r)| + |\varepsilon^i(r)|)\mathrm{d}r. \tag{A.61}$$

For $\beta^i(t)$, we note that by (A.56), (A.44) with $w^i = 0$ and (A.47),

$$\begin{aligned}|\beta^i(t)| &\leq 2\Big((L+2\epsilon\tilde{L})t\frac{5}{4}|z_0^i| + \frac{2\epsilon\tilde{L}t}{n}\frac{5}{4}\max_{s\leq t}\sum_{j\neq i}|z_s^j|\Big)^2 + \frac{2\epsilon\tilde{L}}{n^2}\Big(\max_{s\leq t}\sum_{j\neq i}|z_s^j|\Big)^2 + 2\hat{C} \\ &\leq \frac{25}{4}(L+2\epsilon\tilde{L})^2t^2|z_0^i|^2 + \Big(25\frac{\epsilon^2\tilde{L}^2t^2}{n^2} + \frac{2\epsilon\tilde{L}}{n^2}\Big)\Big(\max_{s\leq t}\sum_{j\neq i}|z_s^j|\Big)^2 + 2\hat{C}. \qquad \text{(A.62)}\end{aligned}$$

Note that by (A.8), (A.47) and since by assumption $\epsilon\tilde{L} < K/6$,

$$\epsilon\tilde{L}(t^2+th) \leq (1/10)(K+4\epsilon\tilde{L})(t^2+th) \leq (1/10)(L+4\epsilon\tilde{L})(t^2+th) \leq 40^{-1}. \qquad \text{(A.63)}$$

Hence, by (A.62) we obtain for the integral containing $\beta^i(t)$ in (A.61)

$$\begin{aligned}&\int_0^t (t-r)|\beta^i(r)|\mathrm{d}r \\ &\leq \int_0^t (t-r)\Big(\frac{25}{4}r^2(L+2\epsilon\tilde{L})^2|z_0^i|^2 + \Big(25\frac{r^2\epsilon^2\tilde{L}^2}{n^2} + \frac{2\epsilon\tilde{L}}{n^2}\Big)\Big(\max_{s\leq t}\sum_{j\neq i}|z_s^j|\Big)^2 + 2\hat{C}\Big)\mathrm{d}r \\ &\leq \frac{25}{48}t^4(L+2\epsilon\tilde{L})^2|z_0^i|^2 + \Big(\frac{25}{12\cdot 40}+1\Big)\frac{\epsilon\tilde{L}t^2}{n^2}\Big(\max_{s\leq t}\sum_{j\neq i}|z_s^j|\Big)^2 + \hat{C}t^2, \qquad \text{(A.64)}\end{aligned}$$

where the last step follows by (A.63).

Next, we bound $\delta^i(t)$ and $\varepsilon^i(t)$. To bound $\delta^i(t)$ and $\varepsilon_3^i(t)$, we note that by (A.40) and (A.52),

$$\begin{aligned}|w_{\lfloor t\rfloor}^i - w_t^i| &\leq \Big|\int_{\lfloor t\rfloor}^t \frac{\mathrm{d}}{\mathrm{d}s}w_s^i\mathrm{d}s\Big| \leq \frac{h}{2}|\nabla_iU(x_{\lfloor t\rfloor}) + \nabla_iU(x_{\lceil t\rceil}) - (\nabla_iU(y_{\lfloor t\rfloor}) + \nabla_iU(y_{\lceil t\rceil}))| \\ &\leq h\Big((L+2\epsilon\tilde{L})z_t^{i,*} + \frac{2\epsilon\tilde{L}}{n}\max_{s\leq t}\sum_{j\neq i}|z_s^j|\Big)\end{aligned}$$

where $z_t^{i,*} = \max_{s\leq t}|z_s^i|$. Hence, by (A.52), (A.42) with $w^i = 0$ and (A.47),

$$\begin{aligned}|2(w_{\lfloor t\rfloor}^i - w_t^i) - h(\nabla_iU(x_{\lfloor t\rfloor}) - \nabla_iU(y_{\lfloor t\rfloor}))| &\leq 3h\Big((L+2\epsilon\tilde{L})z_t^{i,*} + \frac{2\epsilon\tilde{L}}{n}\max_{s\leq t}\sum_{j\neq i}|z_s^j|\Big) \\ &\leq 3h\Big(\frac{5}{4}(L+2\epsilon\tilde{L})|z_0^i| + \frac{5}{2}\frac{\epsilon\tilde{L}}{n}\max_{s\leq t}\sum_{j\neq i}|z_s^j|\Big). \qquad \text{(A.65)}\end{aligned}$$

Hence by (A.65) and (A.42) with $w^i = 0$, and then by (A.63) and (A.47),

$$\begin{aligned}\max_{s\leq t}|\delta^i(s)| &\leq 3h\Big(\frac{5}{4}|z_0^i| + \frac{2\epsilon\tilde{L}(t^2+th)}{n}\max_{s\leq t}\sum_{j\neq i}|z_s^j|\Big)\Big(\frac{5}{4}(L+2\epsilon\tilde{L})|z_0^i| + \frac{5}{2}\frac{\epsilon\tilde{L}}{n}\max_{s\leq t}\sum_{j\neq i}|z_s^j|\Big) \\ &\leq 3h\Big(\frac{25}{16}(L+2\epsilon\tilde{L})|z_0^i|^2 + \frac{15}{4}\epsilon\tilde{L}|z_0^i|\frac{1}{n}\max_{s\leq t}\sum_{j\neq i}|z_s^j| + \frac{\epsilon\tilde{L}}{8n}\max_{s\leq t}\Big(\sum_{j\neq i}|z_s^j|\Big)^2\Big) \qquad \text{(A.66)} \\ &\leq h\Big(\frac{75L}{16} + 15\epsilon\tilde{L}\Big)|z_0^i|^2 + h\frac{6\epsilon\tilde{L}}{n^2}\max_{s\leq t}\Big(\sum_{j\neq i}|z_s^j|\Big)^2, \qquad \text{(A.67)}\end{aligned}$$

Note that Young's product inequality is used in (A.66) to bound the cross term. Similarly, by (A.65), (A.42) with $w^i = 0$, (A.44) with $w^i = 0$ and (A.63),

$$\begin{aligned}
\max_{s\le t} |\varepsilon_3^i(s)|\frac{t}{2} &\le 3h\frac{t}{2}\Big(\frac{5}{4}(L+2\epsilon\tilde{L})t|z_0^i| + \frac{5}{4}\frac{2\epsilon\tilde{L}t}{n}\max_{s\le t}\sum_{j\ne i}|z_s^j|\Big) \\
&\quad \cdot \Big(\frac{5}{4}(L+2\epsilon\tilde{L})|z_0^i| + \frac{5}{2}\frac{\epsilon\tilde{L}}{n}\max_{s\le t}\sum_{j\ne i}|z_s^j|\Big) \\
&\le 3h\frac{t^2}{2}\Big(\frac{25}{16}(L+2\epsilon\tilde{L})^2|z_0^i|^2 + \frac{25}{4}(L+2\epsilon\tilde{L})|z_0^i|\frac{\epsilon\tilde{L}}{n}\max_{s\le t}\sum_{j\ne i}|z_s^j| \qquad \text{(A.68)} \\
&\quad + \frac{25}{4}\Big(\frac{\epsilon\tilde{L}}{n}\max_{s\le t}\sum_{j\ne i}|z_s^j|\Big)^2\Big) \\
&\le h\frac{75(L+2\epsilon\tilde{L})}{64}|z_0^i|^2 + h\frac{15}{32}\frac{\epsilon\tilde{L}}{n^2}\max_{s\le t}\Big(\sum_{j\ne i}|z_s^j|\Big)^2. \qquad \text{(A.69)}
\end{aligned}$$

Note that Young's product inequality is used to bound the cross term in (A.68).

To bound $\varepsilon_1^i(t)$, $\varepsilon_2^i(t)$ and $\varepsilon_4^i(t)$, we note that by (A.40) and (A.52),

$$\begin{aligned}
|z_{\lfloor t\rfloor}^i - z_t^i| = \Big|\int_{\lfloor t\rfloor}^t \frac{\mathrm{d}}{\mathrm{d}s} z_s^i \mathrm{d}s\Big| &\le h|w_{\lfloor t\rfloor}^i - \frac{h}{2}(\nabla_i U(x_{\lfloor t\rfloor}) - \nabla_i U(y_{\lfloor t\rfloor}))| \\
&\le h w_t^{i,*} + \frac{h^2}{2}(L+2\epsilon\tilde{L})z_t^{i,*} + \frac{h^2\epsilon\tilde{L}}{n}\max_{s\le t}\sum_{j\ne i}|z_s^j| \qquad \text{(A.70)}
\end{aligned}$$

where $w_t^{i,*} = \max_{s\le t}|w_s^i|$. Similarly,

$$|z_{\lceil t\rceil}^i - z_t^i| \le h w_t^{i,*} + \frac{h^2}{2}(L+2\epsilon\tilde{L})z_t^{i,*} + \frac{h^2\epsilon\tilde{L}}{n}\max_{s\le t}\sum_{j\ne i}|z_s^j|. \qquad \text{(A.71)}$$

Hence, by applying (A.70), (A.71) and (A.52) in the first step, and (A.42) and (A.44) with

$w^i = 0$ in the second step,

$$\begin{aligned}
\max_{s\le t}(|\varepsilon_1^i(s)+\varepsilon_2^i(s)|)\frac{t}{2} &\le th\Big((L+2\epsilon\tilde{L})z_t^{i,*}+\frac{2\epsilon\tilde{L}}{n}\max_{s\le t}\sum_{j\neq i}|z_s^j|\Big)\\
&\quad\cdot\Big(w_t^{i,*}+\frac{h}{2}(L+2\epsilon\tilde{L})z_t^{i,*}+\frac{h\epsilon\tilde{L}}{n}\max_{s\le t}\sum_{j\neq i}|z_s^j|\Big)\\
&\le th\Big(\frac{5}{4}(L+2\epsilon\tilde{L})|z_0^i|+\frac{5\epsilon\tilde{L}}{2n}\max_{s\le t}\sum_{j\neq i}|z_s^j|\Big)\\
&\quad\cdot\Big((L+2\epsilon\tilde{L})\Big(t+\frac{h}{2}\Big)\frac{5}{4}|z_0|+\frac{2\epsilon\tilde{L}}{n}\frac{5}{4}\Big(t+\frac{h}{2}\Big)\max_{s\le t}\sum_{j\neq i}|z_s^j|\Big)\\
&\le ht\Big(t+\frac{h}{2}\Big)\Big((L+2\epsilon\tilde{L})^2\frac{25}{16}|z_0^i|^2+\frac{25}{4}(L+2\epsilon\tilde{L})\epsilon\tilde{L}|z_0^i|\max_{s\le t}\frac{1}{n}\sum_{j\neq i}|z_s^j|\\
&\quad+\frac{25}{4}\frac{(\epsilon\tilde{L})^2}{n^2}\max_{s\le t}\Big(\sum_{j\neq i}|z_s^j|\Big)^2\Big)\\
&\le ht\Big(t+\frac{h}{2}\Big)\Big(\frac{25}{8}(L+2\epsilon\tilde{L})^2|z_0^i|^2+\frac{25(\epsilon\tilde{L})^2}{2n^2}\max_{s\le t}\Big(\sum_{j\neq i}|z_s^j|\Big)^2\Big)\\
&\le h\frac{25}{32}(L+2\epsilon\tilde{L})|z_0^i|^2+h\frac{5\epsilon\tilde{L}}{16n^2}\max_{s\le t}\Big(\sum_{j\neq i}|z_s^j|\Big)^2. \qquad \text{(A.72)}
\end{aligned}$$

Note that Young's product inequality is used to bound the cross term in the third step and (A.47) and (A.63) are used in the last step. For $\varepsilon_4^i(t)$, we obtain by (A.70) and (A.71),

$$\begin{aligned}
\max_{s\le t}|\varepsilon_4^i(s)|\frac{t}{2} &\le \frac{t}{2}\kappa\max_{s\le t}|(z_s^i+z_{\lfloor s\rfloor}^i)\cdot(z_s^i-z_{\lfloor t\rfloor}^i)+(z_s^i+z_{\lceil s\rceil}^i)\cdot(z_s^i-z_{\lceil s\rceil}^i)|\\
&\le 2th\kappa z_t^{i,*}\Big(w_t^{i,*}+\frac{h}{2}(L+2\epsilon\tilde{L})z_t^{i,*}+\frac{h\epsilon\tilde{L}}{n}\max_{s\le t}\sum_{j\neq i}|z_s^j|\Big)\\
&\le 2th\Big(\frac{5}{4}(L+2\epsilon\tilde{L})|z_0^i|+\frac{5\epsilon\tilde{L}}{2n}\max_{s\le t}\sum_{j\neq i}|z_s^j|\Big) \qquad \text{(A.73)}\\
&\quad\cdot\Big((L+2\epsilon\tilde{L})\Big(t+\frac{h}{2}\Big)\frac{5}{4}|z_0^i|+\frac{2\epsilon\tilde{L}}{n}\frac{5}{4}\Big(t+\frac{h}{2}\Big)\max_{s\le t}\sum_{j\neq i}|z_s^j|\Big)\\
&\le 2ht\Big(t+\frac{h}{2}\Big)\Big((L+2\epsilon\tilde{L})^2\frac{25}{16}|z_0^i|^2+\frac{25}{4}(L+2\epsilon\tilde{L})\epsilon\tilde{L}|z_0^i|\max_{s\le t}\frac{1}{n}\sum_{j\neq i}|z_s^j|\\
&\quad+\frac{25}{4}\frac{(\epsilon\tilde{L})^2}{n^2}\max_{s\le t}\Big(\sum_{j\neq i}|z_s^j|\Big)^2\Big)\\
&\le 2ht\Big(t+\frac{h}{2}\Big)\Big(\frac{25}{8}(L+2\epsilon\tilde{L})^2|z_0^i|^2+\frac{25(\epsilon\tilde{L})^2}{2n^2}\max_{s\le t}\Big(\sum_{j\neq i}|z_s^j|\Big)^2\Big)\\
&\le h\frac{25}{16}(L+2\epsilon\tilde{L})|z_0^i|^2+h\frac{5\epsilon\tilde{L}}{8n^2}\max_{s\le t}\Big(\sum_{j\neq i}|z_s^j|\Big)^2 \qquad \text{(A.74)}
\end{aligned}$$

where (A.73) follows by (A.42) with $w^i = 0$ and (A.44) with $w^i = 0$ and since by (A.8) $\kappa \le$

$(L+2\epsilon\tilde{L})$. Note that Young's product inequality is used to bound the cross term in the third step.

Therefore, by (A.67), (A.69), (A.72) and (A.74),

$$\begin{aligned}
&\int_0^t \Big((t-r)|\varepsilon^i(r)|+|\delta^i(r)|\Big)\mathrm{d}r\\
&\leq \frac{t^2}{2}(\max_{s\leq t}(|\varepsilon_1^i(s)+\varepsilon_2^i(s)|)+\max_{s\leq t}|\varepsilon_3^i(s)|+\max_{s\leq t}|\varepsilon_4^i(s)|)+t\max_{s\leq t}|\delta^i(s)|\\
&\leq ht\Big(\frac{25}{32}(L+2\epsilon\tilde{L})+\frac{75}{64}(L+2\epsilon\tilde{L})+\frac{25}{16}(L+2\epsilon\tilde{L})+\frac{75}{16}L+15\epsilon\tilde{L}\Big)|z_0^i|^2\\
&+ht\Big(\frac{5}{16}+\frac{15}{32}+\frac{5}{8}+6\Big)\frac{\epsilon\tilde{L}}{n^2}\max_{s\leq t}\Big(\sum_{j\neq i}|z_s^j|\Big)^2\\
&= ht\Big(\frac{525}{64}L+\frac{235}{64}K\Big)|z_0^i|^2+ht\frac{237}{32}\frac{\epsilon\tilde{L}}{n^2}\max_{s\leq t}\Big(\sum_{j\neq i}|z_s^j|\Big)^2
\end{aligned}\tag{A.75}$$

where we used $\epsilon\tilde{L}<K/6$ in (A.75). We note that by (A.49), (A.8) and since by assumption $\epsilon\tilde{L}<K/6$,

$$h\Big(\frac{525}{64}L+\frac{235}{64}K\Big)\leq\frac{Kt}{64}\leq\frac{\kappa t}{32}\tag{A.76}$$

and

$$h\frac{237}{32}\leq\frac{237}{32}\frac{Kt}{525L+235K}\leq\frac{1}{2}t.\tag{A.77}$$

Therefore, by (A.75), (A.76) and (A.77)

$$\int_0^t\Big((t-r)|\varepsilon^i(r)|+|\delta^i(r)|\Big)\mathrm{d}r\leq t^2\Big(\frac{\kappa}{32}|z_0|^2+\frac{\epsilon\tilde{L}}{2n^2}\max_{s\leq t}\Big(\sum_{j\neq i}|z_s^j|\Big)^2\Big).\tag{A.78}$$

Inserting (A.64) and (A.78) in (A.61) and applying (A.47) yields,

$$\begin{aligned}
a^i(t)\leq&\Big(1-\frac{5}{6}\kappa t^2\Big)|z_0^i|^2+\frac{25}{48}\kappa t^2|z_0^i|^2+\Big(\frac{5}{96}+1\Big)\frac{\epsilon\tilde{L}t^2}{n^2}\Big(\max_{s\leq t}\sum_{j\neq i}|z_s^j|\Big)^2+\hat{C}t^2\\
&+t^2\Big(\frac{\kappa}{32}|z_0^i|^2+\frac{\epsilon\tilde{L}}{2n^2}\Big(\max_{s\leq t}\sum_{j\neq i}|z_s^j|\Big)^2\Big).
\end{aligned}$$

By (A.54), we obtain for $x,y\in\mathbb{R}^{dn}$ with $|x^i-y^i|>\tilde{R}$,

$$\begin{aligned}
|z_t^i|^2&\leq\Big(1-\Big(\frac{5}{6}-\frac{25}{48}-\frac{2}{32}\Big)\kappa t^2\Big)|z_0^i|^2+\Big(1+\frac{5}{96}+\frac{1}{2}\Big)\frac{\epsilon\tilde{L}t^2}{n^2}\Big(\max_{s\leq t}\sum_{j\neq i}|z_s^j|\Big)^2\\
&\leq\Big(1-\frac{1}{4}\kappa t^2\Big)|z_0^i|^2+2\frac{\epsilon\tilde{L}t^2}{n^2}\Big(\max_{s\leq t}\sum_{j\neq i}|z_s^j|\Big)^2,
\end{aligned}$$

as required.

□

A.6 Proofs of main results

A.6.1 Proof of main contraction result

For the proof of Theorem A.3, we write R^i and r^i for $r^i(x,y) = |x^i - y^i|$ and $R^i(x,y) = |\mathbf{X}^i(x,y) - \mathbf{Y}^i(x,y)|$ for fixed $x, y \in \mathbb{R}^{dn}$. Further, we write $r_s^i = |q_s^i(x,\xi) - q_s^i(y,\eta)|$ for the distance between the two positions at time s satisfying (A.5) where ξ, η are the velocities coupled using the construction given in Appendix A.2.3. Further, we denote $z = x - y$ and $w = \xi - \eta$.

Proof of Theorem A.3. Note that (A.22), (A.24) and (A.48) imply

$$\kappa \geq (1/2)K \quad \text{and} \quad L + 4\epsilon\tilde{L} \leq L + (2K/3) \leq (5/3)L. \tag{A.79}$$

Hence, we obtain by (A.22)

$$(L + 4\epsilon\tilde{L})(T + h_1)^2 \leq \min\Big(\frac{1}{4}, \frac{\kappa}{L + 4\epsilon\tilde{L}}, \frac{1}{256(L + 4\epsilon\tilde{L})\tilde{R}^2}\Big). \tag{A.80}$$

Moreover, the following inequalities are satisfied,

$$\gamma T \leq 1, \tag{A.81}$$
$$(L + 4\epsilon\tilde{L})(T + h) \leq \gamma/4, \tag{A.82}$$
$$\gamma\tilde{R} \leq 1/4, \tag{A.83}$$
$$\exp(T^{-1}(R_1 - \tilde{R})) \geq 12. \tag{A.84}$$

Inequalities (A.81) and (A.83) follow by (A.15), (A.82) follows by (A.15) and (A.80), and the inequality (A.84) follows by (A.16).

We first prove a bound on $\mathbb{E}[f(R^i) - f(r^i)]$ for each particle i similarly to the strategy to bound $\mathbb{E}[f(R) - f(r)]$ in [31, Proof of Theorem 2.4]. We split the calculation of this expectation in two cases depending on the applied coupling.

Case 1: $r^i = |x^i - y^i| \geq \tilde{R}$. In this case, the initial velocities of the i-th particles are synchronized, i.e., $w^i = 0$. By concavity of the function f, by Lemma A.17 and since

$$\sqrt{1 - \mathsf{a}} \leq 1 - \mathsf{a}/2 \qquad \text{for } \mathsf{a} \in [0, 1), \tag{A.85}$$

we obtain

$$\begin{aligned} \mathbb{E}[f(R^i) - f(r^i)] &\leq f'(r^i)\mathbb{E}[R^i - r^i] \\ &\leq f'(r^i)\Big(-\frac{1}{8}\kappa T^2\Big)r^i + f'(r^i)\sqrt{2\epsilon\tilde{L}}\frac{T}{n}\mathbb{E}\Big[\max_{s\leq T}\sum_{j\neq i} r_s^j\Big]. \end{aligned} \tag{A.86}$$

Case 2: $r^i = |x^i - y^i| < \tilde{R}$. In this case, since the distance between the i-th particles is smaller than $\tilde{R}$, the initial velocities of the i-th particles satisfy $w^i = -\gamma z^i$ with maximal possible probability and otherwise a reflection is applied. These disjoint possibilities motivate splitting the expectation $E[f(R^i) - f(r^i)]$ as follows

$$\begin{aligned} \mathbb{E}[f(R^i) - f(r^i)] &= \mathbb{E}[f(R^i) - f(r^i), \{w^i = -\gamma z^i\}] \\ &+ \mathbb{E}[f(R_1 \wedge R^i) - f(r^i), \{w^i \neq -\gamma z^i\}] \\ &+ \mathbb{E}[f(R^i) - f(R_1 \wedge R^i), \{w^i \neq -\gamma z^i\}] = \mathrm{I} + \mathrm{II} + \mathrm{III}. \end{aligned}$$

First, we bound the probability $\mathbb{P}[w^i \neq -\gamma z^i]$, which equals the total variation distance between a standard normal distribution with zero mean and a normal distribution with mean γz^i and unit variance, cf. Lemma 4.4 of [29]. Note using the coupling characterization of the TV distance, this representation shows that the coupling $\xi^i - \eta^i = -\gamma z^i$ holds with maximal probability. By (A.83),

$$\begin{aligned}\mathbb{P}[w^i \neq -\gamma z^i] &= \int_{-\infty}^{\gamma|z^i|/2} \frac{1}{\sqrt{2\pi}}\Big(\exp\Big(-\frac{1}{2}\mathsf{x}^2\Big) - \exp\Big(-\frac{1}{2}\big(\mathsf{x}-\gamma|z^i|\big)^2\Big)\Big)^+ \mathrm{d}\mathsf{x} \\ &= \int_{-\infty}^{\gamma|z^i|/2} \frac{1}{\sqrt{2\pi}}\exp\Big(-\frac{1}{2}\mathsf{x}^2\Big)\mathrm{d}\mathsf{x} - \int_{-\infty}^{-\gamma|z^i|/2} \frac{1}{\sqrt{2\pi}}\exp\Big(-\frac{1}{2}\mathsf{x}^2\Big)\mathrm{d}\mathsf{x} \\ &\leq \frac{2}{\sqrt{2\pi}}\int_0^{\gamma|z^i|/2}\exp\Big(-\frac{1}{2}\mathsf{x}^2\Big)\mathrm{d}\mathsf{x} \leq \frac{2}{\sqrt{2\pi}}\frac{\gamma|z^i|}{2} < \frac{1}{10}. \end{aligned} \tag{A.87}$$

Next, we bound $\mathbf{I}$, $\mathbf{II}$ and $\mathbf{III}$. For $\mathbf{I}$, we note that on the set $\{w^i = -\gamma z^i\}$, by (A.41) and (A.82)

$$\begin{aligned} R^i &\leq (1-\gamma T)|z^i| + (L+2\epsilon\tilde{L})(T^2+Th)|z^i| + \frac{2\epsilon\tilde{L}}{n}(T^2+Th)\max_{s\leq T}\sum_{j\neq i}|z_s^j| \\ &\leq |z^i|\Big(1-\gamma T+\frac{\gamma T}{4}\Big) + \frac{2\epsilon\tilde{L}}{n}(T^2+Th)\max_{s\leq T}\sum_{j\neq i} r_s^j \\ &= \Big(1-\frac{3\gamma T}{4}\Big)r^i + \frac{2\epsilon\tilde{L}}{n}(T^2+Th)\max_{s\leq T}\sum_{j\neq i} r_s^j. \end{aligned}$$

Hence by concavity of f and by (A.87),

$$\begin{aligned} \mathbf{I} &\leq -f'(r^i)\frac{3}{4}\gamma T r^i \mathbb{P}[w^i = -\gamma z^i] + f'(r^i)\mathbb{E}\Big[\frac{2\epsilon\tilde{L}}{n}(T^2+Th)\max_{s\leq T}\sum_{j\neq i} r_s^j\Big] \\ &\leq -f'(r^i)\frac{27}{40}\gamma T r^i + f'(r^i)\frac{2\epsilon\tilde{L}}{n}(T^2+Th)\mathbb{E}\Big[\max_{s\leq T}\sum_{j\neq i} r_s^j\Big]. \end{aligned} \tag{A.88}$$

To bound $\mathbf{II}$, note that by (A.17) for $r, s \leq R_1$,

$$f(s) - f(r) = \int_r^s e^{-t/T}\mathrm{d}t = T(e^{-r/T} - e^{-s/T}) \leq Te^{-r/T} = Tf'(r).$$

Therefore, by (A.87)

$$\mathbf{II} \leq Tf'(r^i)\mathbb{P}[w^i \neq -\gamma z^i] \leq Tf'(r^i)\frac{\gamma r^i}{\sqrt{2\pi}} < \frac{2}{5}\gamma T r^i f'(r^i). \tag{A.89}$$

where we used the bound $1/\sqrt{2\pi} < 2/5$. For $\mathbf{III}$, we get by concavity of f

$$\mathbf{III} \leq f'(R_1)\mathbb{E}[(R^i - R_1)^+, \{w^i \neq -\gamma z^i\}]. \tag{A.90}$$

If $w^i \neq -\gamma z^i$, then $w^i = 2(e^i\cdot\xi^i)e^i$ with $e^i = z^i/|z^i|$ and hence $|z^i + Tw^i| = |r^i + 2Te^i\cdot\xi^i|$. This computation and (A.42) yield

$$R^i \leq (1+(L+2\epsilon\tilde{L})(T^2+Th))\max(|r^i+2Te^i\cdot\xi^i|, r^i) + \frac{2\epsilon\tilde{L}(T^2+Th)}{n}\max_{s\leq T}\sum_{j\neq i} r_s^j.$$

Hence by (A.80) and since $(5/4)r^i - R_1 \le (5/4)\tilde{R} - R_1 \le 0$,

$$\begin{aligned}
&\mathbb{E}[(R^i - R_1)^+, \{w^i \neq -z^i\gamma\}] \\
&\quad \le \mathbb{E}\Big[\Big(\frac{5}{4}\max(|r^i + 2Te^i\cdot\xi^i|, r^i) + \frac{2\epsilon\tilde{L}(T^2+Th)}{n}\max_{s\le T}\sum_{j\neq i} r^j_s - R_1\Big)^+, \{w^i \neq -\gamma z^i\}\Big] \\
&\quad \le \mathbb{E}\Big[\Big(\frac{5}{4}|r^i + 2Te^i\cdot\xi^i| - R_1\Big)^+, \{w^i \neq -\gamma z^i\}\Big] + \mathbb{E}\Big[\frac{2\epsilon\tilde{L}(T^2+Th)}{n}\max_{s\le T}\sum_{j\neq i} r^j_s\Big]. \qquad \text{(A.91)}
\end{aligned}$$

For the first term, where only the i-th particle is involved, we follow the calculations in the proof of [31, Theorem 2.4],

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\frac{5}{4}|r^i + 2Te^i\cdot\xi^i| - R_1\Big)^+, \{w^i \neq -\gamma z^i\}\Big] \\
&\quad = \int_{-\infty}^{\infty}\Big(\frac{5}{4}|r^i + 2T\mathsf{u}| - R_1\Big)^+\frac{1}{\sqrt{2\pi}}\Big(\exp\Big(-\frac{\mathsf{u}^2}{2}\Big) - \exp\Big(-\frac{(\mathsf{u}+\gamma r^i)^2}{2}\Big)\Big)^+ d\mathsf{u} \\
&\quad = \int_{-\frac{\gamma r^i}{2}}^{\infty}\Big(\frac{5}{4}|r^i + 2T\mathsf{u}| - R_1\Big)^+\frac{1}{\sqrt{2\pi}}\Big(\exp\Big(-\frac{\mathsf{u}^2}{2}\Big) - \exp\Big(-\frac{(\mathsf{u}+\gamma r^i)^2}{2}\Big)\Big) d\mathsf{u} \\
&\quad = \int_{-\frac{\gamma r^i}{2}}^{\frac{\gamma z^i}{2}}\Big(\frac{5}{4}|r^i + 2T\mathsf{u}| - R_1\Big)^+\frac{1}{\sqrt{2\pi}}\exp\Big(-\frac{\mathsf{u}^2}{2}\Big) d\mathsf{u} \\
&\quad + \int_{\frac{\gamma r^i}{2}}^{\infty}\Big(\Big(\frac{5}{4}|r^i + 2T\mathsf{u}| - R_1\Big)^+ - \Big(\frac{5}{4}|r^i + 2(\mathsf{u}-\gamma r^i)T| - R_1\Big)^+\Big)\frac{1}{\sqrt{2\pi}}\exp\Big(-\frac{\mathsf{u}^2}{2}\Big) d\mathsf{u} \\
&\quad \le \int_{\frac{\gamma r^i}{2}}^{\infty}\Big(\frac{5}{4}2\gamma r^i T\Big)\frac{1}{\sqrt{2\pi}}\exp\Big(-\frac{\mathsf{u}^2}{2}\Big) d\mathsf{u} \le \frac{5}{4}\gamma T r^i. \qquad \text{(A.92)}
\end{aligned}$$

Hence by (A.84), (A.90), (A.91) and (A.92),

$$\begin{aligned}
\mathrm{III} &\le f'(R_1)\Big(\frac{5}{4}\gamma T r^i + \mathbb{E}\Big[\frac{2\epsilon\tilde{L}(T^2+Th)}{n}\max_{s\le T}\sum_{j\neq i} r^j_s\Big]\Big) \\
&\le f'(r^i)\Big(\frac{5}{48}\gamma T r^i + \frac{\epsilon\tilde{L}(T^2+Th)}{6n}\mathbb{E}\Big[\max_{s\le T}\sum_{j\neq i} r^j_s\Big]\Big). \qquad \text{(A.93)}
\end{aligned}$$

We combine the bounds on I, II and III in (A.88), (A.89) and (A.93) respectively, to obtain for $r^i \le \tilde{R}$,

$$\begin{aligned}
\mathbb{E}[f(R^i) - f(r^i)] &\le -f'(r^i)\frac{27}{40}\gamma T r^i + f'(r^i)\frac{2}{5}\gamma T r^i + f'(r^i)\frac{5}{48}\gamma T r^i \\
&\quad + f'(r^i)\Big(\frac{2\epsilon\tilde{L}(T^2+Th)}{n} + \frac{\epsilon\tilde{L}(T^2+Th)}{6n}\Big)\mathbb{E}\Big[\max_{s\le T}\sum_{j\neq i} r^j_s\Big] \\
&\le -f'(r^i)\frac{41}{240}\gamma T r^i + f'(r^i)\frac{13\epsilon\tilde{L}(T^2+Th)}{6n}\mathbb{E}\Big[\max_{s\le T}\sum_{j\neq i} r^j_s\Big]. \qquad \text{(A.94)}
\end{aligned}$$

Next, we combine (A.86) and (A.94) and sum over i to obtain

$$\begin{aligned}
\mathbb{E}\Big[\sum_i (f(R^i) - f(r^i))\Big] &\le -\min\Big(\frac{41}{240}\gamma T, \frac{1}{8}\kappa T^2\Big)\sum_i r^i f'(r^i) \\
&\quad + \max\Big(\frac{13\epsilon\tilde{L}(T^2+Th)}{6}, \sqrt{2\epsilon\tilde{L}}T\Big)\frac{1}{n}\sum_i f'(r^i)\mathbb{E}\Big[\max_{s\le T}\sum_{j\neq i} r^j_s\Big]. \qquad \text{(A.95)}
\end{aligned}$$

To bound the expectation in the last term of (A.95) we note that when $w^j \neq -\gamma z^j$, then $w^j = 2(e^j \cdot \xi^j)e^j$ with $e^j = z^j/|z^j|$, and hence by (A.83),

$$\begin{aligned}
&\mathbb{E}\Big[|w^j| 1_{\{r^j<\tilde{R}\}\cap\{w^j\neq-\gamma z^j\}}\Big] \\
&\quad = 1_{\{r^j\leq\tilde{R}\}} \int_{-\frac{\gamma r^j}{2}}^{\infty} \frac{2|\mathsf{x}|}{\sqrt{2\pi}}\Big(\exp\Big(-\frac{\mathsf{x}^2}{2}\Big) - \exp\Big(-\frac{(\mathsf{x}+\gamma r^j)^2}{2}\Big)\Big)d\mathsf{x} \\
&\quad = 1_{\{r^j\leq\tilde{R}\}}\Big(\int_{-\frac{\gamma r^j}{2}}^{\infty} \frac{1}{\sqrt{2\pi}} 2|\mathsf{x}| \exp\Big(-\frac{\mathsf{x}^2}{2}\Big)d\mathsf{x} - \int_{\frac{\gamma r^j}{2}}^{\infty} \frac{1}{\sqrt{2\pi}} 2|\mathsf{x}-\gamma r^j| \exp\Big(-\frac{\mathsf{x}^2}{2}\Big)d\mathsf{x}\Big) \\
&\quad \leq 1_{\{r^j\leq\tilde{R}\}}\Big(\int_{-\frac{\gamma r^j}{2}}^{\frac{\gamma r^j}{2}} \frac{1}{\sqrt{2\pi}} 2|\mathsf{x}| \exp\Big(-\frac{\mathsf{x}^2}{2}\Big)d\mathsf{x} + \int_{\frac{\gamma r^j}{2}}^{\infty} \frac{1}{\sqrt{2\pi}} 2\gamma r^j \exp\Big(-\frac{\mathsf{x}^2}{2}\Big)d\mathsf{x}\Big) \\
&\quad \leq 1_{\{r^j\leq\tilde{R}\}}\Big((\gamma r^j)^2 + \gamma r^j\Big) \leq 1_{\{r^j\leq\tilde{R}\}}\Big(\frac{1}{4}\gamma r^j + \gamma r^j\Big) \leq \frac{5}{4}\gamma r^j. \qquad \text{(A.96)}
\end{aligned}$$

Then we obtain by (A.45), by (A.80), and since by (A.81) for $w^j = -\gamma z^j$, $|z^j + Tw^j| \leq |z^j|$,

$$\begin{aligned}
\mathbb{E}\Big[\max_{s\leq T} \sum_{j\neq i} r_s^j\Big] &\leq \frac{5}{4}\mathbb{E}\Big[\sum_j \max(|z^j + Tw^j|, |z^j|)\Big] \\
&\leq \frac{5}{4}\mathbb{E}\Big[\sum_j |z^j| + \sum_j T|w^j| 1_{\{r^j<\tilde{R}\}\cap\{w^j\neq-\gamma z^j\}}\Big] \\
&= \frac{5}{4}\sum_j |z^j| + \frac{5}{4}T\sum_j \mathbb{E}\Big[|w^j| 1_{\{r^j<\tilde{R}\}\cap\{w^j\neq-\gamma z^j\}}\Big] \leq \frac{45}{16}\sum_j r^j, \qquad \text{(A.97)}
\end{aligned}$$

where last step holds by (A.96) and (A.81). Hence inserting (A.97) in (A.95),

$$\begin{aligned}
\mathbb{E}\Big[\sum_i (f(R^i) - f(r^i))\Big] &\leq -\min\Big(\frac{41}{240}\gamma T, \frac{1}{8}\kappa T^2\Big)\sum_i f'(r^i) r^i \\
&\quad + \sum_i f'(r^i) \max\Big(\frac{13\epsilon\tilde{L}(T^2+Th)}{6}, \sqrt{2\epsilon\tilde{L}}T\Big)\frac{1}{n}\frac{45}{16}\sum_j r^j. \qquad \text{(A.98)}
\end{aligned}$$

Since by (A.82) $\kappa T^2 \leq T\gamma/4$, the minimum in (A.98) is attained at $\frac{1}{8}\kappa T^2$. Since (A.8), (A.80) and (A.24) imply (A.63) with $t = T$, it holds $(13/6)\epsilon\tilde{L}(T^2 + Th) \leq 13/6\sqrt{\epsilon\tilde{L}(T^2 + Th)/40}$. Hence, the maximum in (A.98) is attained at $\sqrt{2\epsilon\tilde{L}}T$. The minimum of $\frac{r^i f'(r^i)}{f(r^i)}$ is attained at R_1 defined in (A.16),

$$\inf_{r^i} \frac{r^i f'(r^i)}{f(r^i)} = \frac{R_1 \exp(-R_1/T)}{T(1-\exp(-R_1/T))} \geq \frac{5}{4}\Big(\frac{\tilde{R}}{T} + 2\Big)\exp\Big(-\frac{5\tilde{R}}{4T}\Big)\exp\Big(-\frac{5}{2}\Big), \qquad \text{(A.99)}$$

and it holds by (A.16) that

$$\sum_i f'(r^i)\frac{1}{n} r^j \leq \frac{f(r^j)}{f'(r^j)} \leq \exp\Big(\frac{R_1}{T}\Big) f(r^j) = \exp\Big(\frac{5\tilde{R}}{4T}\Big)\exp\Big(\frac{5}{2}\Big) f(r^j) \qquad \text{(A.100)}$$

where we used that $f(r^j) \geq r^j f'(r^j)$ and $\exp(-R_1/T) \leq f'(r^i) \leq 1$. Hence,

$$\begin{aligned}\mathbb{E}\Big[\sum_i (f(R^i) - f(r^i))\Big] &\leq -\frac{1}{8}\kappa T^2 \frac{5}{4}\Big(\frac{\tilde{R}}{T} + 2\Big)\exp\Big(-\frac{5\tilde{R}}{4T}\Big)\exp\Big(-\frac{5}{2}\Big)\sum_i f(r^i) \\ &\quad + \sqrt{2\epsilon}\tilde{L}T\frac{45}{16}\exp\Big(\frac{5}{2}\Big)\exp\Big(\frac{5\tilde{R}}{4T}\Big)\sum_i f(r^i) \\ &\leq -\frac{1}{78}\kappa T^2 \exp\Big(-\frac{5\tilde{R}}{4T}\Big)\sum_i f(r^i),\end{aligned}$$

where the last step holds by (A.24). □

A.6.2 Proofs of results from Section A.3.2

Proof of Corollary A.7. This proof works analogously to the proof of [31, Corollary 2.6] and uses essentially [31, Lemma 6.1]. By Theorem A.3, the contractivity condition

$$\mathbb{E}[\rho(\mathbf{X}(x,y), \mathbf{Y}(x,y))] \leq e^{-c}\rho(x,y) \tag{A.101}$$

is satisfied for the coupling $(\mathbf{X}(x,y), \mathbf{Y}(x,y))$. Let ν, η be probability measures on $\mathbb{R}^{dn}$ and let ω be an arbitrary coupling of ν and η. By [31, Lemma 6.1], there exists a Markov chain $(\mathbf{X}_m, \mathbf{Y}_m)_{m\geq 0}$ on a probability space $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P})$ such that $(\mathbf{X}_0, \mathbf{Y}_0) \sim \omega$, $(\mathbf{X}_m)$, $(\mathbf{Y}_m)$ are Markov chains each having transition kernel π_h and initial distributions ν and η, respectively, and $M_m = e^{cm}\rho(\mathbf{X}_m, \mathbf{Y}_m)$ is a non-negative supermartingale. Then, for all $m \in \mathbb{N}$,

$$\mathcal{W}_\rho(\nu{\pi_h}^m, \eta{\pi_h}^m) \leq \mathbb{E}[\rho(\mathbf{X}_m, \mathbf{Y}_m)] \leq e^{-cm}\mathbb{E}[\rho(\mathbf{X}_0, \mathbf{Y}_0)] = e^{-cm}\int \rho \mathrm{d}\omega.$$

Since ω is chosen arbitrary, we take the infimum over all couplings $\omega \in \Gamma(\nu, \eta)$ and obtain (A.26). The bound (A.27) follows by (A.20). The existence of a unique probability measure μ_h on $\mathbb{R}^{dn}$ holds by (A.27) and by Banach fixed-point theorem, cf. [84, Theorem 3.9]. Since $\mu_h{\pi_h}^m = \mu_h$ for all m, $\Delta(m) \leq e^{R_1/T}e^{-cm}\Delta(0)$. Hence, for a given $\tilde{\epsilon} > 0$, $\Delta(m) \leq \tilde{\epsilon}$ holds for (A.29) by (A.16). □

Proof of Theorem A.8. This proof uses essentially standard numerical analysis techniques and a priori estimates given in Lemma A.15. Fix $x, \xi \in \mathbb{R}^{dn}$. Denote by $(x_s, v_s) = (q_s(x,\xi), p_s(x,\xi))$ the Hamiltonian dynamics driven by (A.1). Set $\mathbf{x}_k^i := q_{kh}^i(x,\xi)$, $\tilde{\mathbf{x}}_k^i := \tilde{q}_{kh}^i(x,\xi)$, $\mathbf{v}_k^i := p_{kh}^i(x,\xi)$ and $\tilde{\mathbf{v}}_k^i := \tilde{p}_{kh}^i(x,\xi)$. By (A.1) and (A.5), it holds

$$\begin{aligned}|\mathbf{x}_{k+1}^i - \tilde{\mathbf{x}}_{k+1}^i| &\leq |\mathbf{x}_k^i - \tilde{\mathbf{x}}_k^i| + h|\mathbf{v}_k^i - \tilde{\mathbf{v}}_k^i| + \Big|\int_{kh}^{(k+1)h}\int_{kh}^{u}\Big(\nabla_i U(x_r) - \nabla_i U(\tilde{\mathbf{x}}_k)\Big)\mathrm{d}r\mathrm{d}u\Big|, \\ |\mathbf{v}_{k+1}^i - \tilde{\mathbf{v}}_{k+1}^i| &\leq |\mathbf{v}_k^i - \tilde{\mathbf{v}}_k^i| + \Big|\int_{kh}^{(k+1)h}\Big(\frac{1}{2}\nabla_i U(\tilde{\mathbf{x}}_k) - \nabla_i U(x_u) + \frac{1}{2}\nabla_i U(\tilde{\mathbf{x}}_{k+1})\Big)\mathrm{d}u\Big|.\end{aligned}$$

By (A.52) and (A.5),

$$\begin{aligned}\sum_i \Big|\nabla_i U(x_r) - \nabla_i U(\tilde{\mathbf{x}}_k)\Big| &\leq \sum_i \Big|\nabla_i U(x_r) - \nabla_i U(\mathbf{x}_k)\Big| + \sum_i \Big|\nabla_i U(\mathbf{x}_k) - \nabla_i U(\tilde{\mathbf{x}}_k)\Big| \\ &\leq \sum_i \Big| \int_{kh}^r v_s \cdot \nabla\nabla_i U(x_s) \mathrm{d}s\Big| + (L + 4\epsilon\tilde{L}) \sum_i |\mathbf{x}_k^i - \tilde{\mathbf{x}}_k^i| \\ &\leq \sum_i \Big| \int_{kh}^r (L + 4\epsilon\tilde{L}) v_s^i \mathrm{d}s\Big| + (L + 4\epsilon\tilde{L}) \sum_i |\mathbf{x}_k^i - \tilde{\mathbf{x}}_k^i| \\ &\leq \sum_i (L + 4\epsilon\tilde{L})\Big(h\Big(\frac{21}{16}|v_0^i| + \frac{5}{4}(L + 4\epsilon\tilde{L})T|x_0^i|\Big) + |\mathbf{x}_k^i - \tilde{\mathbf{x}}_k^i|\Big),\end{aligned}$$

where (A.39) and $(L + 4\epsilon\tilde{L})T^2 \leq (1/4)$ is used in the last step. Analogously,

$$\begin{aligned}&\sum_i \Big(- \nabla_i U(x_u) + \frac{1}{2}\nabla_i U(\tilde{\mathbf{x}}_k) + \frac{1}{2}\nabla_i U(\tilde{\mathbf{x}}_{k+1})\Big) \\ &\quad \leq \sum_i (L + 4\epsilon\tilde{L})\Big(h\Big(\frac{21}{16}|v_0^i| + \frac{5}{4}(L + 4\epsilon\tilde{L})T|x_0^i|\Big) + \frac{1}{2}|\mathbf{x}_k^i - \tilde{\mathbf{x}}_k^i| + \frac{1}{2}|\mathbf{x}_{k+1}^i - \tilde{\mathbf{x}}_{k+1}^i|\Big).\end{aligned} \tag{A.102}$$

Then for any initial position $x \in \mathbb{R}^{dn}$,

$$\begin{aligned}\mathbb{E}\Big[\sum_i |\mathbf{x}_{k+1}^i - \tilde{\mathbf{x}}_{k+1}^i|\Big] &\leq \Big(1 + \frac{h^2(L + 4\epsilon\tilde{L})}{2}\Big)\mathbb{E}\Big[\sum_i |\mathbf{x}_k^i - \tilde{\mathbf{x}}_k^i|\Big] \\ &\quad + h\mathbb{E}\Big[\sum_i |\mathbf{v}_k^i - \tilde{\mathbf{v}}_k^i|\Big] + \frac{h^3}{2}M_1,\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}\Big[\sum_i |\mathbf{v}_{k+1}^i - \tilde{\mathbf{v}}_{k+1}^i|\Big] &\leq \mathbb{E}\Big[\sum_i |\mathbf{v}_k^i - \tilde{\mathbf{v}}_k^i|\Big] + h^2 M_1 \\ &\quad + \frac{(L + 4\epsilon\tilde{L})h}{2}\Big(\mathbb{E}\Big[\sum_i |\mathbf{x}_{k+1}^i - \tilde{\mathbf{x}}_{k+1}^i|\Big] + \mathbb{E}\Big[\sum_i |\mathbf{x}_k^i - \tilde{\mathbf{x}}_k^i|\Big]\Big)\end{aligned} \tag{A.103}$$

with $M_1 := \mathbb{E}_{\xi\sim\mathcal{N}(0,I_{dn})}[\sum_i (L + 4\epsilon\tilde{L})(\frac{21}{16}|\xi^i| + \frac{5}{4}(L + 4\epsilon\tilde{L})T|x^i|)]$. Set $a_k := \mathbb{E}[\sum_i |\mathbf{x}_k^i - \tilde{\mathbf{x}}_k^i|]$ and $b_k := \mathbb{E}[\sum_i |\mathbf{v}_k^i - \tilde{\mathbf{v}}_k^i|]$. The goal is to bound a_k from above using the discrete Gronwall lemma [89, Proposition 3.2]. Note that this sequence (a_k, b_k) with $a_0 = b_0 = 0$ satisfies

$$\begin{aligned}a_{k+1} &\leq (1 + (L + 4\epsilon\tilde{L})h^2/2)a_k + hb_k + (h^3 M_1/2) \\ b_{k+1} &\leq b_k + h^2 M_1 + ((L + 4\epsilon\tilde{L})h/2)(a_{k+1} + a_k).\end{aligned}$$

We deduce for b_{k+1}

$$b_{k+1} \leq (L + 4\epsilon\tilde{L})h\sum_{l=1}^k a_l + \frac{(L + 4\epsilon\tilde{L})h}{2}a_{k+1} + (k+1)h^2 M_1.$$

Inserting this estimate in a_{k+1} yields

$$a_{k+1} \leq (1 + (L + 4\epsilon\tilde{L})h^2)a_k + (kh^3 M_1 + h^3 M_1/2) + (L + 4\epsilon\tilde{L})h^2 \sum_{l=1}^{k-1} a_l. \tag{A.104}$$

Note that the sequence $(\tilde{a}_k)$ satisfying

$$\tilde{a}_{k+1} = (1 + (L + 4\epsilon\tilde{L})h^2)\tilde{a}_k + (k + (1/2))h^3 M_1 + (L + 4\epsilon\tilde{L})h^2 \sum_{l=1}^{k-1} \tilde{a}_l \tag{A.105}$$

is an upper bound of the sequence (a_k), i.e. $a_k \leq \tilde{a}_k$. Moreover, it holds $\tilde{a}_k \leq \tilde{a}_{k+1}$. Hence,

$$\tilde{a}_{k+1} \leq (1 + (L + 4\epsilon\tilde{L})kh^2)\tilde{a}_k + (k + 1/2)h^3 M_1 \leq (1 + (L + 4\epsilon\tilde{L})Th)\tilde{a}_k + Th^2 M_1.$$

Applying the discrete Grönwall lemma to $\tilde{a}_k$ yields for all $k \leq (T/h)$,

$$\begin{aligned} a_k \leq \tilde{a}_k &\leq \frac{1}{(L + 4\epsilon\tilde{L})T}\Big((1 + (L + 4\epsilon\tilde{L})hT)^k - 1\Big)ThM_1 \\ &\leq h\frac{\exp((L + 4\epsilon\tilde{L})T^2) - 1}{(L + 4\epsilon\tilde{L})}M_1 \leq h\frac{\exp(1/4) - 1}{(L + 4\epsilon\tilde{L})}M_1, \end{aligned} \tag{A.106}$$

where we applied $(L + 4\epsilon\tilde{L})T^2 \geq 1/4$ in the last step.

Hence, there exists a constant C_2 depending on L, $\tilde{L}$, ϵ and T such that for all $k \in \mathbb{N}$ with $kh \leq T$ and for any initial value $x \in \mathbb{R}^{dn}$,

$$\mathbb{E}\Big[\sum_i |\mathbf{x}_k^i - \tilde{\mathbf{x}}_k^i|\Big] \leq h \cdot C_2\Big(d^{1/2}n + \sum_i |x^i|\Big)$$

and so (A.30) holds. Note that the term $d^{1/2}n$ comes from $\mathbb{E}[\sum |\xi^i|]$ since $\xi^i \sim \mathcal{N}(0, I_d)$.

If we assume additionally Assumption A.5 and Assumption A.6, then we can instead of (A.102) bound using (A.51) and the trapezoidal rule,

$$\begin{aligned} &\Big|\int_{kh}^{(k+1)h} \sum_i \Big(-\nabla_i U(x_u) + \frac{1}{2}\nabla_i U(\tilde{\mathbf{x}}_k) + \frac{1}{2}\nabla_i U(\tilde{\mathbf{x}}_{k+1})\Big)\mathrm{d}u\Big| \\ &\leq \Big|\int_{kh}^{(k+1)h} \sum_i \Big(-\nabla_i U(x_u) + \frac{1}{2}\nabla_i U(\mathbf{x}_k) + \frac{1}{2}\nabla_i U(\mathbf{x}_{k+1})\Big)\mathrm{d}u\Big| \\ &+ \frac{h}{2}\sum_i (L + 4\epsilon\tilde{L})(|\mathbf{x}_k - \tilde{\mathbf{x}}_k| + |\mathbf{x}_{k+1} - \tilde{\mathbf{x}}_{k+1}|) \\ &\leq \frac{h}{2}\sum_i (L + 4\epsilon\tilde{L})(|\mathbf{x}_k^i - \tilde{\mathbf{x}}_k^i| + |\mathbf{x}_{k+1}^i - \tilde{\mathbf{x}}_{k+1}^i|) + \frac{h^3}{12}\sum_i \sup_{u \in [kh,(k+1)h]}\Big|\frac{\mathrm{d}^2}{\mathrm{d}u^2}\nabla_i U(x_u)\Big|. \end{aligned} \tag{A.107}$$

The last term is bounded using (A.5), (A.51), Assumption A.5 and Assumption A.6 by

$$\sum_i \sup_{u \in [kh,(k+1)h]}\Big|\frac{\mathrm{d}^2}{\mathrm{d}u^2}\nabla_i U(x_u)\Big| \leq \sum_i (L_H + 8\epsilon\tilde{L}_H)\max_{s \leq T}|v_s^i|^2 + \sum_i (L + 4\epsilon\tilde{L})^2 \max_{s \leq T}|x_s^i|.$$

Since we can bound $\sum_i \max_{s \leq T}|v_s^i|^2$ and $\sum_i \max_{s \leq T}|x_s^i|$ by Lemma A.15 and Young's product inequality in terms of $\sum_i |\xi^i|$, $\sum_i |\xi^i|^2$, $\sum_i |x^i|$ and $\sum_i |x^i|^2$, we can bound the last term in (A.107) after taking expectation over $\xi \sim \mathcal{N}(0, I_{dn})$ by a constant $h^3 M_2$ where M_2 is a constant depending on L, $\tilde{L}$, L_H, $\tilde{L}_H$, ϵ, d, n, $\sum_i |x^i|$ and $\sum_i |x^i|^2$. More precisely, the dependence of M_2 is linear in nd, $\sum_i |x^i|$ and $\sum_i |x^i|^2$. Replacing $h^2 M_1$ in (A.103) by $h^3 M_2$ leads to the fact that a_k in (A.106) is bounded from above by $a_{k+1} \leq h^2(\exp(1/(4k)) - 1)/(L + 4\epsilon\tilde{L})(M_2 + M_1/(2T))$. Hence, there exists a constant $\tilde{C}_2$ of order $\mathcal{O}(T^{-1})$ depending on L, $\tilde{L}$, ϵ, L_H and $\tilde{L}_H$ such that for all $k \in \mathbb{N}$ with $kh \leq T$ and for any initial value $x \in \mathbb{R}^{dn}$ (A.31) holds, which concludes the proof. □

Proof of Theorem A.10. Let ν be an arbitrary probability measure on $\mathbb{R}^{dn}$. Recall that by Corollary A.7, it holds $\mathcal{W}_{\ell^1}(\mu_h {\pi_h}^m, \nu {\pi_h}^m) \leq \exp((5/4)(2 + (\tilde{R}/T))) \exp(-cm) \mathcal{W}_{\ell^1}(\mu_h, \nu)$. By (A.20) and Corollary A.7,

$$\begin{aligned}\Delta(m) &:= \mathcal{W}_{\ell^1}(\mu, \nu {\pi_h}^m) \leq \mathcal{W}_{\ell^1}(\mu, \mu_h) + \mathcal{W}_{\ell^1}(\mu_h, \nu {\pi_h}^m) \leq \mathrm{I} + \mathrm{II}, \qquad \text{where} \\ \mathrm{I} &= \exp\Big(\frac{5}{4}\Big(2 + \frac{\tilde{R}}{T}\Big)\Big) \mathcal{W}_\rho(\mu, \mu_h) \\ \mathrm{II} &= \exp\Big(\frac{5}{4}\Big(2 + \frac{\tilde{R}}{T}\Big) - cm\Big) \mathcal{W}_{\ell^1}(\mu_h, \nu).\end{aligned}$$

For m chosen as in (A.32), $\mathrm{II} \leq \tilde{\epsilon}/2$. To obtain $\mathrm{I} \leq \tilde{\epsilon}/2$, we use the results of Corollary A.9. Then there exists h_2 such that for $h \leq \min(h_1, h_2)$, $\mathrm{I} \leq \tilde{\epsilon}/2$ holds. In particular, we choose $h_2^{-1} = 2C_2(d^{1/2}n + \int \sum_i |x^i| \mu(\mathrm{d}x))/(c\tilde{\epsilon})$. Hence, for fixed L, $\tilde{L}$, ϵ, K, R, T, h_2^{-1} is of order $\mathcal{O}(\tilde{\epsilon}^{-1}(d^{1/2}n + \int \sum_i |x^i| \mu(\mathrm{d}x)))$. If additionally Assumption A.5 and Assumption A.6 are assumed, then for $h \leq \min(h_1, \tilde{h}_2)$ where $\tilde{h}_2^{-1} = (2\tilde{C}_2(dn + \int \sum_i |x^i| \mu(\mathrm{d}x) + \int \sum_i |x^i|^2 \mu(\mathrm{d}x))/(c\tilde{\epsilon}))^{1/2}$, $\mathrm{I} \leq \tilde{\epsilon}/2$ holds. Note that $\tilde{h}_2^{-1}$ is for fixed L, $\tilde{L}$, L_H, $\tilde{L}_H$, ϵ, K, R, T of order $\mathcal{O}(\tilde{\epsilon}^{-1/2}((nd)^{1/2} + (\int \sum_i |x^i| \mu(\mathrm{d}x))^{1/2} + (\int \sum_i |x^i|^2 \mu(\mathrm{d}x))^{1/2}))$.

Let us finally remark that $\sum |x^i| \mu(\mathrm{d}x) = \int |x^1| \mu(\mathrm{d}x)$ and $\int \sum_i |x^i|^2 \mu(\mathrm{d}x)$ are finite. This holds, since by Assumption A.3 and Assumption A.4 $\exp(-U(x))$ can be bounded from above by a density function of a Gaussian product measure which has finite first and second moments. □

A.6.3 Proofs of results from Section A.3.3

Proof of Theorem A.13. The proof follows [84, Proof of Theorem 3.17]. It holds for $m, b \in \mathbb{N}$ by (A.34),

$$\mathbb{E}_\nu[A_{m,b} g] = \frac{1}{m} \sum_{k=b}^{b+m-1} (\nu {\pi_h}^k)(g).$$

For all $g \in \mathcal{C}_b^1(\mathbb{R}^{nd})$ with $\max_{l \in \{1,\ldots,n\}} \|\nabla_l g\| \leq \infty$,

$$\begin{aligned}|g(x) - g(y)| &= \sum_{i=1}^n |g(x^1, \ldots, x^i, y^{i+1}, \ldots, y^n) - g(x^1, \ldots, x^{i-1}, y^i, \ldots, y^n)| \\ &\leq \max_l \|\nabla_l g\| \sum_{i=1}^n |(x^1, \ldots, x^i, y^{i+1}, \ldots, y^n) - (x^1, \ldots, x^{i-1}, y^i, \ldots, y^n)| \\ &= \max_l \|\nabla_l g\| \sum_{i=1}^n |x^i - y^i|.\end{aligned}$$

Then for all $k \in \mathbb{N}$ and for all couplings $\omega \in \Gamma(\nu \pi_h^k, \mu)$,

$$|(\nu \pi_h^k)(g) - \mu(g)| \leq \max_l \|\nabla_l g\| \int \sum_{i=1}^n |x^i - y^i| \omega(\mathrm{d}x \mathrm{d}y).$$

Hence by the triangle inequality, by (A.28) and by (A.20),

$$
\begin{aligned}
&|\mathbb{E}_\nu[A_{m,b}g] - \mu(g)| \\
&\quad \le \frac{1}{m}\sum_{k=b}^{b+m-1} |(\nu{\pi_h}^k)(g) - \mu(g)| \le \frac{1}{m}\sum_{k=b}^{b+m-1} \max_i \|\nabla_i g\|_\infty \mathcal{W}_{\ell^1}(\nu{\pi_h}^k, \mu) \\
&\quad \le \frac{1}{m}\sum_{k=b}^{b+m-1} \max_i \|\nabla_i g\|_\infty \mathcal{W}_{\ell^1}(\nu{\pi_h}^k, \mu_h) + \max_i \|\nabla_i g\|_\infty \mathcal{W}_{\ell^1}(\mu_h, \mu) \\
&\quad \le \frac{1}{m}\sum_{k=b}^{b+m-1} \max_i \|\nabla_i g\|_\infty M e^{-ck} \mathcal{W}_{\ell^1}(\nu, \mu_h) + \max_i \|\nabla_i g\|_\infty \mathcal{W}_{\ell^1}(\mu_h, \mu) \\
&\quad \le \frac{1}{m}\max_i \|\nabla_i g\|_\infty M \frac{e^{-cb}}{1-e^{-c}} \mathcal{W}_{\ell^1}(\nu, \mu_h) + \max_i \|\nabla_i g\|_\infty \mathcal{W}_{\ell^1}(\mu_h, \mu)
\end{aligned}
$$

with $M = \exp(\frac{5}{4}(2 + \frac{\tilde{R}}{T}))$. Applying Corollary A.9 yields the result. □

A.7 Appendix: Perturbation of the product model

If the confinement potential is a quadratic potential, i.e., $V(\mathsf{x}) = K/2|\mathsf{x}|^2$ for all $\mathsf{x} \in \mathbb{R}^d$, the mean-field model can be treated as a perturbation of the product model. Given $x, y \in \mathbb{R}^{dn}$ we consider the synchronous coupling of four transition kernels $\pi_h(x,\cdot)$, $\pi_h(y,\cdot)$, $\pi_h^{prod}(x,\cdot)$ and $\pi_h^{prod}(y,\cdot)$, where $\pi_h(x,\cdot)$ and $\pi_h(y,\cdot)$ denote the two transition kernels with a mean-field interaction, i.e., $\epsilon > 0$, and $\pi_h^{prod}(x,\cdot)$ and $\pi_h^{prod}(y,\cdot)$ are transition kernels of the product model, i.e., $\epsilon = 0$. Then the coupling HMC step is given by

$$
\begin{aligned}
\mathbf{X}(x,y) &= q_T(x,\xi), & \mathbf{Y}(x,y) &= q_T(y,\xi), \\
\mathbf{X}^{prod}(x,y) &= \hat{q}_T(x,\xi), & \mathbf{Y}^{prod}(x,y) &= \hat{q}_T(y,\xi),
\end{aligned}
$$

where $\xi \sim \mathcal{N}(0, I_{dn})$ and $\hat{q}_T$ denotes the position component of the Hamiltonian dynamics given by (A.5) for the product model.

Theorem A.18. *Suppose that $V(\mathsf{x}) = (K/2)|\mathsf{x}|^2$ for all $\mathsf{x} \in \mathbb{R}^d$ and Assumption A.4 hold. Let $T \in (0,\infty)$, $h_1 \in [0,\infty)$ and $\epsilon \in (0,\infty)$ satisfy*

$$
(K + 4\epsilon)(T^2 + Th_1) \le 1. \tag{A.108}
$$

Then for any $h \in [0, h_1]$ such that $h = 0$ or $T/h \in \mathbb{N}$ and any $x, y \in \mathbb{R}^{dn}$,

$$
\sum_{i=1}^n |\mathbf{X}^i(x,y) - \mathbf{Y}^i(x,y) - (\mathbf{X}^{i,prod}(x,y) - \mathbf{Y}^{i,prod}(x,y))| \le 8\epsilon\tilde{L}(T^2 + Th)\sum_{i=1}^n |x^i - y^i|.
$$

Proof. Fix $x, y, v \in \mathbb{R}^d$. For $t \in [0,T]$, we write $x_t^i = q_t^i(x,v)$ and $y_t^i = q_t^i(y,v)$ for the i-th position component of the solution to (A.5) with initial values (x,v) and (y,v), respectively, and with potential $U(x) = \sum_{i=1}^n((K/2)|x^i| + \epsilon n^{-1}\sum_{j=1,j\neq i}^n W(x^i - x^j))$. Analogously, we write $\hat{x}_t^i = \hat{q}_t^i(x,v)$ and $\hat{y}_t^i = \hat{q}_t^i(y,v)$ for the i-th position component of the solution to (A.5) with initial values (x,v) and (y,v), respectively, and with potential $\hat{U}(x) = \sum_{i=1}^n (K/2)|x^i|$. We set

$z_t^i = x_t^i - y_t^i$ and $\hat{z}_t^i = \hat{x}_t^i - \hat{y}_t^i$ for all $i = 1, \dots, n$ and $t \in [0, T]$. By (A.5) and Assumption A.4 it holds for $t \in [0, T]$,

$$\begin{aligned}
&\max_{s \le t} \sum_{i=1}^n |z_s^i - \hat{z}_s^i| \\
&\quad = \max_{s\le t} \sum_{i=1}^n \Big| \int_0^s \int_0^r \Big(-\frac{1}{2}(\nabla_i U(x^i_{\lfloor u \rfloor}) - \nabla_i U(y^i_{\lfloor u \rfloor}) + \nabla_i U(x^i_{\lceil u \rceil}) - \nabla_i U(y^i_{\lceil u \rceil})) \\
&\quad + \frac{1}{2}(\nabla_i \hat{U}(\hat{x}^i_{\lfloor u \rfloor}) - \nabla_i \hat{U}(\hat{y}^i_{\lfloor u \rfloor}) + \nabla_i \hat{U}(\hat{x}^i_{\lceil u \rceil}) - \nabla_i \hat{U}(\hat{y}^i_{\lceil u \rceil}))\Big) \mathrm{d}u \mathrm{d}r \\
&\quad - \frac{h}{2} \int_0^s \Big(\nabla_i U(x^i_{\lfloor u \rfloor}) - \nabla_i U(y^i_{\lfloor u \rfloor}) - (\nabla_i \hat{U}(\hat{x}^i_{\lfloor u \rfloor}) - \nabla_i \hat{U}(\hat{y}^i_{\lfloor u \rfloor}))\Big) \mathrm{d}u \Big| \\
&\quad \le \frac{K}{2}(t^2 + th) \max_{s\le t} \sum_{i=1}^n |z_s^i - \hat{z}_s^i| + 2\epsilon \tilde{L}(t^2 + th) \max_{s \le t} \sum_{i=1}^n |z_s^i|. \qquad \text{(A.109)}
\end{aligned}$$

By (A.108) and (A.45),

$$\max_{s\le t} \sum_{i=1}^n |z_s^i - \hat{z}_s^i| \le 4\epsilon \tilde{L}(t^2 + th) \max_{s\le t} \sum_{i=1}^n |z_s^i| \le 8 \epsilon \tilde{L}(t^2 + th) \sum_{i=1}^n |x^i - y^i|.$$

Thus, the result holds for $t = T$. □

We note that the step (A.109) uses crucially that the third derivative of V vanishes.

As some calculations simplify in the product case with quadratic confinement potential, (A.50) in Lemma A.17 holds for all $i = 1, \dots, n$ provided $K(t^2 + th) \le 1/4$ and $h \le (4/165)t$ is satisfied. Hence by (A.85),

$$\sum_{i=1}^n |\mathbf{X}^{i,prod}(x, y) - \mathbf{Y}^{i,prod}(x, y)| \le (1 - (1/8)KT^2) \sum_{i=1}^n |x^i - y^i|$$

for $K(T^2 + Th) \le 1/4$ and $h \le (4/165)T$. Combining the contraction result for the product model with the perturbation result yields the following consequence.

Corollary A.19. *Suppose that $V(\mathsf{x}) = (K/2)|\mathsf{x}|^2$ for all $\mathsf{x} \in \mathbb{R}^d$ and Assumption A.4 hold. Let $T \in (0, \infty)$, $h_1 \in (0, \infty)$ and $\epsilon \in (0, \infty)$ satisfy*

$$\begin{aligned}
&K(T^2 + Th_1) \le 1/4, \qquad h \le (4/165)T, \qquad \textit{and} \\
&\epsilon \tilde{L} \le K/256. \qquad \text{(A.110)}
\end{aligned}$$

Then, for any $h \in [0, h_1]$ such that $h = 0$ or $T/h \in \mathbb{N}$ and for any $x, y \in \mathbb{R}^{dn}$

$$\sum_{i=1}^n |\mathbf{X}^i(x, y) - \mathbf{Y}^i(x, y)| \le (1 - KT^2/16) \sum_{i=1}^n |x^i - y^i|,$$

and for any two probability measures ν and η on $\mathbb{R}^{dn}$ and any $m \in \mathbb{N}$,

$$\mathcal{W}_{l^1}(\nu \pi_h^m, \eta \pi_h^m) \le e^{-KT^2 m/16} \mathcal{W}_{l^1}(\nu, \eta).$$

Proof. The result is a direct consequence of the contraction result and Theorem A.18, i.e.,

$$
\begin{aligned}
\sum_{i=1}^{n} |\mathbf{X}^i(x,y) - \mathbf{Y}^i(x,y)| &\leq \sum_{i=1}^{n} |\mathbf{X}^{i,prod}(x,y) - \mathbf{Y}^{i,prod}(x,y)| \\
&\quad + \sum_{i=1}^{n} |\mathbf{X}^i(x,y) - \mathbf{Y}^i(x,y) - (\mathbf{X}^{i,prod}(x,y) - \mathbf{Y}^{i,prod}(x,y))| \\
&\leq (1 - KT^2/8) \sum_{i=1}^{n} |x^i - y^i| + (8\epsilon \tilde{L}(T^2 + Th) \sum_{i=1}^{n} |x^i - y^i| \\
&\leq (1 - KT^2/16) \sum_{i=1}^{n} |x^i - y^i|,
\end{aligned}
$$

where the last step follows by (A.110). The second bound in Corollary A.19 holds in the same line as the proof of Corollary A.7. ☐

Appendix B

Sticky nonlinear SDEs and convergence of McKean-Vlasov equations without confinement

Alain Durmus, Andreas Eberle, Arnaud Guillin and Katharina Schuh, Sticky nonlinear SDEs and convergence of McKean-Vlasov equations without confinement. ArXiv e-print 2201.07652, January 2022.[1]

Abstract

We develop a new approach to study the long time behaviour of solutions to nonlinear stochastic differential equations in the sense of McKean, as well as propagation of chaos for the corresponding mean-field particle system approximations. Our approach is based on a sticky coupling between two solutions to the equation. We show that the distance process between the two copies is dominated by a solution to a one-dimensional nonlinear stochastic differential equation with a sticky boundary at zero. This new class of equations is then analyzed carefully. In particular, we show that the dominating equation has a phase transition. In the regime where the Dirac measure at zero is the only invariant probability measure, we prove exponential convergence to equilibrium both for the one-dimensional equation, and for the original nonlinear SDE. Similarly, propagation of chaos is shown by a componentwise sticky coupling and comparison with a system of one dimensional nonlinear SDEs with sticky boundaries at zero. The approach applies to equations without confinement potential and to interaction terms that are not of gradient type.

Key words: sticky coupling, McKean-Vlasov equation, unconfined dynamics, convergence to equilibrium, sticky nonlinear SDE.
Mathematics Subject Classification: 60H10, 60J60, 82C31.

[1] A. E. and K. S. have been supported by the *Hausdorff Center for Mathematics.* Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) im Rahmen der Exzellenzstrategie des Bundes und der Länder - GZ 2047/1, Projekt-ID 390685813. The work of A.G. has been (partially) supported by the Project EFI ANR-17-CE40-0030 of the French National Research Agency A.D. acknowledges support of the Lagrange Mathematical and Computing Research Center.

B.1 Introduction

The main objective of this paper is to study and quantify convergence to equilibrium for McKean-Vlasov type nonlinear stochastic differential equations of the form

$$\mathrm{d}\bar{X}_t = \left[\int_{\mathbb{R}^d} b(\bar{X}_t - x)\mathrm{d}\bar{\mu}_t(x)\right]\mathrm{d}t + \mathrm{d}B_t\ , \qquad \bar{\mu}_t = \mathrm{Law}(\bar{X}_t)\ , \tag{B.1}$$

where $(B_t)_{t\geq 0}$ is a d-dimensional standard Brownian motion and $b : \mathbb{R}^d \to \mathbb{R}^d$ is a Lipschitz continuous function. This nonlinear SDE is the probabilistic counterpart of the *Fokker-Planck equation*

$$\frac{\partial}{\partial t}u_t = \nabla \cdot \left[(1/2)\nabla u_t - (b * u_t)u_t\right]\ , \tag{B.2}$$

which describes the time evolution of the density u_t of $\bar{\mu}_t$ with respect to the Lebesgue measure on $\mathbb{R}^d$. Moreover, we also study uniform in time propagation of chaos for the approximating mean-field interacting particle systems

$$\mathrm{d}X_t^{i,N} = \frac{1}{N}\sum_{j=1}^{N} b(X_t^{i,N} - X_t^{j,N})\mathrm{d}t + \mathrm{d}B_t^i\ , \qquad\qquad i \in \{1, \ldots, N\}\ , \tag{B.3}$$

with i.i.d. initial values $X_0^{1,N}, \ldots, X_0^{N,N}$, and driven by independent d-dimensional Brownian motions $\{(B_t^i)_{t\geq 0}\}_{i=1}^N$. Our results are based on a new probabilistic approach relying on sticky couplings and comparison with solutions to a class of nonlinear stochastic differential equations on the real interval $[0, \infty)$ with a sticky boundary at 0. The study of this type of equations carried out below might also be of independent interest.

The equations (B.1) and (B.2) have been studied in many works. Often a slightly different setup is considered, where the interaction b is assumed to be of gradient type, i.e., $b = -\nabla W$ for an *interaction potential* function $W : \mathbb{R}^d \to \mathbb{R}$, and an additional *confinement potential* function $V : \mathbb{R}^d \to \mathbb{R}$ satisfying $\lim_{|x|\to\infty} V(x) = \infty$ is included in the equations. The corresponding Fokker-Planck equation

$$\frac{\partial}{\partial t}u_t \quad = \quad \nabla \cdot \left[(1/2)\nabla u_t + (\nabla V + \nabla W * u_t)u_t\right]\ , \tag{B.4}$$

occurs for example in the modelling of granular media, see [190, 13] and the references therein. Existence and uniqueness of solutions to (B.1), (B.2) and (B.4) have been studied intensively. Introductions to this topic can be found for example in [95, 143, 146, 187], while recent results have been established in [150, 106]. Under appropriate conditions, it can be shown that the solutions converge to a unique stationary distribution at some given rate, see e.g. [42, 43, 25, 86, 75, 98]. In the case without confinement considered here, convergence to equilibrium of $(\bar{\mu}_t)_{t\geq 0}$ defined by (B.1) can only be expected for centered solutions, or after recentering around the center of mass of $\bar{\mu}_t$. It has first been analyzed in [42, 43] by an analytic approach and under the assumption that $b = -\nabla W$ for a convex function W. In particular, exponential convergence to equilibrium has been established under the strong convexity assumption $\mathrm{Hess}(W) \geq \rho\,\mathrm{Id}$ for some $\rho > 0$, and polynomial convergence in the case where W is only degenerately strictly convex. Similar results and some extensions have been derived in [138, 44] using a probabilistic approach.

Our first contribution aims at complementing these results, and extending them to non-convex interaction potentials and interaction functions that are not of gradient type. More precisely, suppose that

$$b(x) = -Lx + \gamma(x) \ , \qquad x \in \mathbb{R}^d \ , \tag{B.5}$$

where $L \in (0,\infty)$ is a positive real constant, and $\gamma : \mathbb{R}^d \to \mathbb{R}^d$ is a bounded function. Then we give conditions on γ ensuring exponential convergence of centered solutions to (B.1) to a unique stationary distribution in the standard L^1 Wasserstein metric. More generally, we show in Theorem B.1 that under these conditions there exist constants $M, c \in (0,\infty)$ that depend only on L and γ such that if $(\bar{\mu}_t)_{t\geq 0}$ and $(\bar{\nu}_t)_{t\geq 0}$ are the marginal distributions of two solutions of (B.1), then for all $t \geq 0$,

$$\mathcal{W}_1(\bar{\mu}_t, \bar{\nu}_t) \leq M\mathrm{e}^{-ct}\mathcal{W}_1(\bar{\mu}_0, \bar{\nu}_0) \ .$$

Using a coupling approach, related results have been derived in the previous works [86, 75] for the case where an additional confinement term is included in the equations. However, the arguments in these works rely on treating the equation with confinement and interaction term as a perturbation of the corresponding equation without interaction term, which has good ergodic properties. In the unconfined case this approach does not work, since the equation without interaction is transient and hence does not admit an invariant probability measure. Therefore, we have to develop a new approach for analyzing the equation without confinement.

Our approach is based on sticky couplings, an idea first developed in [87] to control the total variation distance between the marginal distributions of two non degenerate diffusion processes with identical noise but different drift coefficients. Since two solutions of (B.1) differ only in their drifts, we can indeed couple them using a sticky coupling in the sense of [87]. It can then be shown that the coupling distance process can be controlled by the solution $(r_t)_{t\geq 0}$ of a nonlinear SDE on $[0,\infty)$ with a sticky boundary at 0 of the form

$$\mathrm{d}r_t = [\tilde{b}(r_t) + a\mathbb{P}(r_t > 0)]\mathrm{d}t + 2\mathbb{1}_{(0,\infty)}(r_t)\mathrm{d}W_t \ , \tag{B.6}$$

Here $\tilde{b}$ is a real-valued function on $[0,\infty)$ satisfying $\tilde{b}(0) = 0$, a is a positive constant, and $(W_t)_{t\geq 0}$ is a one-dimensional standard Brownian motion. Solutions to SDEs with diffusion coefficient $r \mapsto \mathbb{1}_{(0,\infty)}(r)$, as in (B.6), have a sticky boundary at 0, i.e., if the drift at 0 is strictly positive, then the set of all time points $t \in [0,\infty)$ such that $r_t = 0$ is a fractal set with strictly positive Lebesgue measure that does not contain any open interval. Sticky SDEs have attracted wide interest, starting from [92, 93] in the one-dimensional case. Multivariate extensions have been considered in [115, 196, 197] building upon results obtained in [142, 183, 184], while corresponding martingale problems have been investigated in [186]. Note that in general no strong solution for this class of SDEs exists as illustrated in [57]. We refer to [90, 12] and the references therein for recent contributions on this topic. Note, however, that in contrast to standard sticky SDEs, the equation (B.6) is nonlinear in the sense of McKean. We are not aware of previous studies of such nonlinear sticky equations, which seems to be a very interesting topic on its own.

Intuitively, one would expect that as time evolves, more mass gets stuck at 0, i.e., $\mathbb{P}(r_t > 0)$ decreases. As a consequence, the drift at 0 in Equation (B.6) decreases, which again forces even more mass to get stuck at 0. Therefore, one could hope that $\mathbb{P}(r_t = 0)$ converges to 1 as $t \to \infty$. On the other hand, if a is too large then the drift at 0 might be too strong so that not all of the mass gets stuck at 0 eventually. This indicates that there might be a *phase transition* for the nonlinear sticky SDE depending on the size of the constant a compared to $\tilde{b}$. In Section

B.3, we prove rigorously that this intuition is correct. Under appropriate conditions on $\tilde{b}$, we show at first that existence and uniqueness in law holds for solutions of (B.6). Then we prove that for a sufficiently small, the Dirac measure at 0 is the unique invariant probability measure, and geometric ergodicity holds. As a consequence, under corresponding assumptions, the sticky coupling approach yields exponential convergence to equilibrium for the original nonlinear SDE (B.1). On the other hand, we prove the existence of multiple invariant probability measures for (B.6) if the smallness condition on a is not satisfied. Our results for (B.1) can also be adapted to deal with nonlinear SDEs over the torus $\mathbb{T} = \mathbb{R}/(2\pi\mathbb{Z})$, as considered in [64]. As an example, we discuss the application to the Kuramoto model for which a more explicit analysis is available [3, 14, 15, 41].

Finally, in addition to studying the long-time behaviour of the nonlinear SDE (B.1), we are also interested in establishing *propagation of chaos* for the mean-field particle system approximation (B.3). The propagation of chaos phenomenon first introduced by Kac [120] describes the convergence of the empirical measure of the mean-field particle system (B.3) to the solution (B.1). More precisely, in [187, 146] it has been shown under weak assumptions on W that for i.i.d. initial laws, the random variables $X_t^{i,N}$, $i \in \{1, \dots, N\}$, become asymptotically independent as $N \to \infty$, and the common law μ_t^N of each of these random variables converges to $\bar{\mu}_t$. However, the original results are only valid uniformly over a finite time horizon. Quantifying the convergence uniformly for all times $t \in \mathbb{R}_+$ is an important issue. The case with a confinement potential has been studied for example in [75], see also the references therein. Again, the case when there is only interaction is more difficult. Malrieu [138] seems the first to consider the case without confinement. By applying a synchronous coupling, he proved uniform in time propagation of chaos for strongly convex interaction potentials. Later on, assuming that the interaction potential is loosing strict convexity only in a finite number of points (e.g., $W(x) = |x|^3$), Cattiaux, Guillin and Malrieu [44] have shown uniform in time propagation of chaos with a rate getting worse with the degeneracy in convexity. In a very recent work, Delarue and Tse [63] prove uniform in time weak propagation of chaos (i.e., observable by observable) on the torus via Lions derivative methods. Remarkably, their results are not limited to the unique invariant measure case.

Our contribution is in the same vein using probabilistic tools in place of analytic ones. We endow the space $\mathbb{R}^{Nd}$ consisting of N particle configurations $x = (x^i)_{i=1}^N$ with the semi-metric $l^1 \circ \pi$, where

$$l^1(x, y) \;=\; \frac{1}{N} \sum\nolimits_{i=1}^N \left| x^i - y^i \right| \tag{B.7}$$

is a normalized l^1-distance between configurations $x, y \in \mathbb{R}^{Nd}$, and

$$\pi(x, y) \;=\; \left(\left(x^i - \frac{1}{N} \sum\nolimits_{j=1}^N x^j \right)_{i=1}^N, \left(y^i - \frac{1}{N} \sum\nolimits_{j=1}^N y^j \right)_{i=1}^N \right) \;, \tag{B.8}$$

is a projection from $\mathbb{R}^{Nd} \times \mathbb{R}^{Nd}$ to the subspace $\mathsf{H}_N \times \mathsf{H}_N$, where

$$\mathsf{H}_N \;=\; \{ x \in \mathbb{R}^{Nd} : \sum\nolimits_{i=1}^N x^i = 0 \} \;. \tag{B.9}$$

Let $\mathcal{W}_{l^1 \circ \pi}$ denote the L^1 Wasserstein semimetric on probability measures on $\mathbb{R}^{Nd}$ corresponding to the cost function $l^1 \circ \pi$. Then under assumptions stated below, we prove uniform in time propagation of chaos for the mean-field particle system in the following sense: Suppose that $(X_t^{1,N}, \dots, X_t^{N,N})_{t \ge 0}$ is a solution of (B.3) such that $X_0^{1,N}, \dots, X_0^{N,N}$ are i.i.d. with distribution

$\bar{\mu}_0$ having finite second moment. Let ν_t^N denote the joint law of the random variables $X_t^{i,N}$, $i \in \{1, \dots N\}$, and let $\bar{\mu}_t$ denote the law of the solution of (B.1) with initial law $\bar{\mu}_0$. Then there exists a constant $C \in [0, \infty)$ such that for any $N \in \mathbb{N}$,

$$\sup_{t \geq 0} \mathcal{W}_{l^1 \circ \pi}(\bar{\mu}_t^{\otimes N}, \nu_t^N) \leq C N^{-1/2} \ . \tag{B.10}$$

The proof is based on a componentwise sticky coupling, and a comparison of the coupling difference process with a system of one-dimensional sticky nonlinear SDEs.

The paper is organised as follows. In Appendix B.2, we state our main results regarding the long-time behaviour of (B.1). The main results on one-dimensional nonlinear SDEs with a sticky boundary at zero are stated in Section B.3. Sections B.4 and B.5 contain the corresponding results on uniform (in time) propagation of chaos and mean-field systems of sticky SDEs. All the proofs are given in Appendix B.6. In Appendix B.7, we carry the results over to nonlinear sticky SDEs over $\mathbb{T}$ and consider the application to the Kuramoto model.

Notation The Euclidean norm on $\mathbb{R}^d$ is denoted by $|\cdot|$. For $x \in \mathbb{R}$, we write $x_+ = \max(0, x)$. For some space $\mathbb{X}$, which here is either $\mathbb{R}^d$, $\mathbb{R}^{Nd}$ or $\mathbb{R}_+$, we denote its Borel σ-algebra by $\mathcal{B}(\mathbb{X})$. The space of all probability measures on $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ is denoted by $\mathcal{P}(\mathbb{X})$. Let $\mu, \nu \in \mathcal{P}(\mathbb{X})$. A coupling ξ of μ and ν is a probability measure on $(\mathbb{X} \times \mathbb{X}, \mathcal{B}(\mathbb{X}) \otimes \mathcal{B}(\mathbb{X}))$ with marginals μ and ν. $\Gamma(\mu, \nu)$ denotes the set of all couplings of μ and ν. The L^1 Wasserstein distance with respect to a distance function $d : \mathbb{X} \times \mathbb{X} \to \mathbb{R}_+$ is defined by

$$\mathcal{W}_d(\mu, \nu) = \inf_{\xi \in \Gamma(\mu, \nu)} \int_{\mathbb{X} \times \mathbb{X}} d(x, y) \xi(\mathrm{d}x \mathrm{d}y) \ .$$

We write $\mathcal{W}_1$ if the underlying distance function is the Euclidean distance.

We denote by $\mathcal{C}(\mathbb{R}_+, \mathbb{X})$ the set of continuous functions from $\mathbb{R}_+$ to $\mathbb{X}$, and by $\mathcal{C}^2(\mathbb{R}_+, \mathbb{X})$ the set of twice continuously differentiable functions.

Consider a probability space $(\Omega, \mathcal{A}, P)$ and a measurable function $r : \Omega \to \mathcal{C}(\mathbb{R}_+, \mathbb{X})$. Then $\mathbb{P} = P \circ r^{-1}$ denotes the law on $\mathcal{C}(\mathbb{R}_+, \mathbb{X})$, and $P_t = P \circ {r_t}^{-1}$ the marginal law on $\mathbb{X}$ at time t.

B.2 Long-time behaviour of McKean-Vlasov diffusions

We establish our results regarding (B.1) and (B.3) under the following assumption on b.

Assumption B.1. *The function $b : \mathbb{R}^d \to \mathbb{R}^d$ is Lipschitz continuous and anti-symmetric, i.e., $b(z) = -b(-z)$, and there exist $L \in (0, \infty)$, a function $\gamma : \mathbb{R}^d \to \mathbb{R}^d$ and a Lipschitz continuous function $\kappa : [0, \infty) \to \mathbb{R}$ such that*

$$b(z) = -Lz + \gamma(z) \qquad \textit{for all } z \in \mathbb{R}^d \ , \tag{B.11}$$

and the following conditions are satisfied for all $x, \tilde{x}, y \in \mathbb{R}^d$:

$$\langle x - y, \gamma(x - \tilde{x}) - \gamma(y - \tilde{x}) \rangle \leq \kappa(|x - y|)|x - y|^2 \ , \tag{B.12}$$

and

$$\limsup_{r \to \infty} (\kappa(r) - L) < 0 \ . \tag{B.13}$$

Let $\bar{b}(r) = (\kappa(r) - L)r$. If (B.13) holds, then there exist $R_0, R_1 \geq 0$ such that for

$$\bar{b}(r) < 0 \,, \qquad \text{for any } r \geq R_0 \,, \tag{B.14}$$

$$\bar{b}(r)/r \leq -4/[R_1(R_1 - R_0)] \,, \qquad \text{for any } r \geq R_1 \;. \tag{B.15}$$

In addition, we assume

Assumption B.2.

$$\|\gamma\|_\infty \leq \Big(4\int_0^{R_1} \exp\Big(\frac{1}{2}\int_0^s \bar{b}(r)_+ \mathrm{d}r\Big)\mathrm{d}s\Big)^{-1} \,.$$

We consider the following condition on the initial distribution.

Assumption B.3. *The initial distribution μ_0 satisfies $\int_{\mathbb{R}^d} \|x\|^4 \mu_0(\mathrm{d}x) < +\infty$ and $\int_{\mathbb{R}^d} x\, \mu_0(\mathrm{d}x) = 0$.*

Note that under conditions Assumption B.1 and Assumption B.3, unique strong solutions $(\bar{X}_t)_{t\geq 0}$ and $(\{X_t^{i,N}\}_{i=1}^N)_{t\geq 0}$ exist for (B.1) and (B.3), see e.g. [44, Theorem 2.6]. In addition, note that since b is assumed to be anti-symmetric, by an easy localisation argument, we get that $\mathrm{d}\mathbb{E}[\bar{X}_t]/\mathrm{d}t = \mathbb{E}[b * \mu_t(\bar{X}_t)] = 0$ and $\mathrm{d}\mathbb{E}[N^{-1}\sum_{i=1}^N X_t^{i,N}]/\mathrm{d}t = 0$. Thus, if $\bar{X}_0$ and $\{X_0^{i,N}\}_{i=1}^N$ have distribution μ_0 and $\mu_0^{\otimes N}$, respectively, with μ_0 satisfying Assumption B.3, then it holds $\mathbb{E}[\bar{X}_t] = 0$ and $\mathbb{E}[N^{-1}\sum_{i=1}^N X_t^{i,N}] = 0$ for all $t \geq 0$.

Suppose $f : \mathbb{R}_+ \to \mathbb{R}_+$ is an increasing, concave function vanishing at zero. Then $d(x, y) = f(|x - y|)$ defines a distance. The corresponding L^1 Wasserstein distance is denoted by $\mathcal{W}_f$. Note that in the case $f(t) = t$ for any $t \geq 0$, $\mathcal{W}_f$ is simply $\mathcal{W}_1$.

Theorem B.1 (Contraction for nonlinear SDE)**.** *Assume Assumption B.1. Let $\bar{\mu}_0, \bar{\nu}_0$ be probability measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ satisfying Assumption B.3. For any $t \geq 0$, let $\bar{\mu}_t$ and $\bar{\nu}_t$ denote the laws of $\bar{X}_t$ and $\bar{Y}_t$ where $(\bar{X}_s)_{s\geq 0}$ and $(\bar{Y}_s)_{s\geq 0}$ are solutions of* (B.1) *with initial distribution $\bar{\mu}_0$ and $\bar{\nu}_0$, respectively. Then, for all $t \geq 0$,*

$$\mathcal{W}_f(\bar{\mu}_t, \bar{\nu}_t) \leq \mathrm{e}^{-\tilde{c}t}\mathcal{W}_f(\bar{\mu}_0, \bar{\nu}_0) \qquad \textit{and} \qquad \mathcal{W}_1(\bar{\mu}_t, \bar{\nu}_t) \leq M_1 \mathrm{e}^{-\tilde{c}t}\mathcal{W}_1(\bar{\mu}_0, \bar{\nu}_0) \,, \tag{B.16}$$

where the function f is defined by (B.37) *and the constants $\tilde{c}$ and M_1 are given by*

$$\tilde{c}^{-1} = 2\int_0^{R_1}\int_0^s \exp\Big(\frac{1}{2}\int_r^s \bar{b}(u)_+ \;\mathrm{d}u\Big)\mathrm{d}r\mathrm{d}s \,, \tag{B.17}$$

$$M_1 = 2\exp\Big(\frac{1}{2}\int_0^{R_0} \bar{b}(s)_+ \mathrm{d}s\Big) \,. \tag{B.18}$$

Proof. The proof is postponed to Appendix B.6.2. □

The construction and definition of the underlying distance function $f(|x - y|)$ mentioned in Theorem B.1 is based on the one introduced by [83].

To prove Theorem B.1 we use a coupling $(\bar{X}_t, \bar{Y}_t)_{t\geq 0}$ of two copies of solutions to the nonlinear stochastic differential equation (B.1) with different initial conditions. The coupling $(\bar{X}_t, \bar{Y}_t)_{t\geq 0}$ will be defined as the weak limit of a family of couplings $(\bar{X}_t^\delta, \bar{Y}_t^\delta)_{t\geq 0}$, parametrized by $\delta > 0$. Roughly, this family is mixture of synchronous and reflection couplings and can be described as follows. For $\delta > 0$, $(\bar{X}_t^\delta, \bar{Y}_t^\delta)_{t\geq 0}$ behaves like a reflection coupling if $|\bar{X}_t^\delta - \bar{Y}_t^\delta| \geq \delta$, and like a synchronous coupling if $|\bar{X}_t^\delta - \bar{Y}_t^\delta| = 0$. For $|\bar{X}_t^\delta - \bar{Y}_t^\delta| \in (0, \delta)$ we take an interpolation of synchronous and reflection coupling. We argue that the family of couplings $\{(\bar{X}_t^\delta, \bar{Y}_t^\delta)_{t\geq 0} : \delta > 0\}$ is tight and that a subsequence $\{(\bar{X}_t^{\delta_n}, \bar{Y}_t^{\delta_n})_{t\geq 0} : n \in \mathbb{N}\}$ converges to a limit $(\bar{X}_t, \bar{Y}_t)_{t\geq 0}$. This limit is a coupling which we call the sticky coupling associated to (B.1).

To carry out the construction rigorously, we take two Lipschitz continuous functions $\mathrm{rc}^\delta, \mathrm{sc}^\delta : \mathbb{R}_+ \to [0,1]$ for $\delta > 0$ such that

$$\mathrm{rc}^\delta(0) = 0\ ,\ \mathrm{rc}^\delta(r) = 1 \text{ for } r \geq \delta\ , \mathrm{rc}^\delta(r) > 0 \text{ for } r > 0 \text{ and } \mathrm{rc}^\delta(r)^2 + \mathrm{sc}^\delta(r)^2 = 1 \text{ for } r \geq 0\ . \tag{B.19}$$

Further, we assume that there exists $\epsilon_0 > 0$ such that for any $\delta \leq \epsilon_0$, rc^δ satisfies

$$\mathrm{rc}^\delta(r) \geq \frac{\|\gamma\|_{\mathrm{Lip}}}{2\|\gamma\|_\infty} r \quad \text{for any } r \in (0,\delta)\ , \tag{B.20}$$

where $\|\gamma\|_{\mathrm{Lip}} < \infty$ denotes the Lipschitz norm of γ. This assumption is satisfied for example if $\mathrm{rc}^\delta(r) = \sin((\pi/2\delta)r)\mathbb{1}_{r<\delta} + \mathbb{1}_{r\geq\delta}$ and $\mathrm{sc}^\delta(r) = \cos((\pi/2\delta)r)\mathbb{1}_{r<\delta}$ with $\delta \leq \epsilon_0 = 2\|\gamma\|_\infty/\|\gamma\|_{\mathrm{Lip}}$.

Let $(B_t^1)_{t\geq 0}$ and $(B_t^2)_{t\geq 0}$ be two d-dimensional Brownian motions. We define the coupling $(\bar{X}_t^\delta, \bar{Y}_t^\delta)_{t\geq 0}$ as a process in $\mathbb{R}^{2d}$ satisfying the following nonlinear stochastic differential equation

$$\begin{aligned} \mathrm{d}\bar{X}_t^\delta &= b * \bar{\mu}_t^\delta(\bar{X}_t^\delta)\mathrm{d}t + \mathrm{rc}^\delta(\bar{r}_t^\delta)\mathrm{d}B_t^1 + \mathrm{sc}^\delta(\bar{r}_t^\delta)\mathrm{d}B_t^2\ , & \bar{\mu}_t^\delta &= \mathrm{Law}(\bar{X}_t^\delta)\ , \\ \mathrm{d}\bar{Y}_t^\delta &= b * \bar{\nu}_t^\delta(\bar{Y}_t^\delta)\mathrm{d}t + \mathrm{rc}^\delta(\bar{r}_t^\delta)(\mathrm{Id} - 2\bar{e}_t^\delta(\bar{e}_t^\delta)^T)\mathrm{d}B_t^1 + \mathrm{sc}^\delta(\bar{r}_t^\delta)\mathrm{d}B_t^2\ , & \bar{\nu}_t^\delta &= \mathrm{Law}(\bar{Y}_t^\delta) \end{aligned} \tag{B.21}$$

with initial condition $(\bar{X}_0^\delta, \bar{Y}_0^\delta) = (x_0, y_0)$. Here we set $\bar{Z}_t^\delta = \bar{X}_t^\delta - \bar{Y}_t^\delta$, $\bar{r}_t^\delta = |\bar{Z}_t^\delta|$ and $\bar{e}_t^\delta = \bar{Z}_t^\delta/\bar{r}_t^\delta$ if $\bar{r}_t^\delta \neq 0$. For $\bar{r}_t^\delta = 0$, $\bar{e}_t^\delta$ is some arbitrary unit vector, whose exact choice is irrelevant since $\mathrm{rc}^\delta(0) = 0$.

Theorem B.2. *Assume Assumption B.1. Let $\bar{\mu}_0$ and $\bar{\nu}_0$ be probability measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ satisfying Assumption B.3. Then, $(\bar{X}_t, \bar{Y}_t)_{t\geq 0}$ is a subsequential limit in distribution as $\delta \to 0$ of $\{(\bar{X}_t^\delta, \bar{Y}_t^\delta)_{t\geq 0} : \delta > 0\}$ where $(\bar{X}_t)_{t\geq 0}$ and $(\bar{Y}_t)_{t\geq 0}$ are solutions of* (B.1) *with initial distribution $\bar{\mu}_0$ and $\bar{\nu}_0$. Further, there exists a process $(r_t)_{t\geq 0}$ defined on the same probability space as $(\bar{X}_t, \bar{Y}_t)_{t\geq 0}$ satisfying for any $t \geq 0$, $|\bar{X}_t - \bar{Y}_t| \leq r_t$ almost surely and which is a weak solution of*

$$\mathrm{d}r_t = (\bar{b}(r_t) + 2\|\gamma\|_\infty \mathbb{P}(r_t > 0))\mathrm{d}t + 2\mathbb{1}_{(0,\infty)}(r_t)\mathrm{d}\tilde{W}_t\ , \tag{B.22}$$

where $(\tilde{W}_t)_{t\geq 0}$ is a one-dimensional Brownian motion.

Proof. The proof is postponed to Appendix B.6.2. □

Therefore, next we study sticky nonlinear SDEs given by (B.6).

B.3 Nonlinear SDEs with sticky boundaries

Consider nonlinear SDEs with a sticky boundary at 0 of the form

$$\mathrm{d}r_t = (\tilde{b}(r_t) + P_t(g))\mathrm{d}t + 2\mathbb{1}_{(0,\infty)}(r_t)\mathrm{d}W_t\ , \qquad P_t = \mathrm{Law}(r_t)\ , \tag{B.23}$$

where $\tilde{b} : [0,\infty) \to \mathbb{R}$ is some continuous function and $P_t(g) = \int_{\mathbb{R}_+} g(r)P_t(\mathrm{d}r)$ for some measurable function $g : [0,\infty) \to \mathbb{R}$.

In this section we establish existence, uniqueness in law and comparison results for solutions of (B.6). Consider a filtered probability space $(\Omega, \mathcal{A}, (\mathcal{F}_t)_{t\geq 0}, P)$ and a probability measure μ on $\mathbb{R}_+$. We call an $(\mathcal{F}_t)_{t\geq 0}$ adapted process $(r_t, W_t)_{t\geq 0}$ a *weak solution* of (B.23) with initial distribution μ if the following holds: $\mu = P \circ r_0^{-1}$, the process $(W_t)_{t\geq 0}$ is a one-dimensional $(\mathcal{F}_t)_{t\geq 0}$ Brownian motion w.r.t. P, the process $(r_t)_{t\geq 0}$ is non-negative and continuous, and satisfies almost-surely

$$r_t - r_0 = \int_0^t \Big(\tilde{b}(r_s) + P_s(g)\Big)\mathrm{d}s + \int_0^t 2 \cdot \mathbb{1}_{(0,\infty)}(r_s)\mathrm{d}W_s \ , \qquad \text{for } t \in \mathbb{R}_+ \ .$$

Note that the sticky nonlinear SDE given in (B.6) is a special case of (B.23) with $g(r) = a\mathbb{1}_{(0,\infty)}(r)$ since $\mathbb{P}(r_t > 0) = \int_{\mathbb{R}_+} \mathbb{1}_{(0,\infty)}(y)P_t(\mathrm{d}y)$ with $P_t = P \circ r_t^{-1}$.

B.3.1 Existence, uniqueness in law, and a comparison result

Let $\mathbb{W} = \mathcal{C}(\mathbb{R}_+, \mathbb{R})$ be the space of continuous functions endowed with the topology of uniform convergence on compact sets, and let $\mathcal{B}(\mathbb{W})$ be the corresponding Borel σ-algebra. Suppose $(r_t, W_t)_{t\geq 0}$ is a solution of (B.23) on $(\Omega, \mathcal{A}, P)$, then we denote by $\mathbb{P} = P \circ r^{-1}$ its law on $(\mathbb{W}, \mathcal{B}(\mathbb{W}))$. We say that *uniqueness in law* holds for (B.23) if for any two solutions $(r_t^1)_{t\geq 0}$ and $(r_t^2)_{t\geq 0}$ of (B.23) with the same initial law, the distributions of $(r_t^1)_{t\geq 0}$ and $(r_t^2)_{t\geq 0}$ on $(\mathbb{W}, \mathcal{B}(\mathbb{W}))$ are equal.

We impose the following assumptions on $\tilde{b}$, g and the initial condition μ:

Assumption B.4. *$\tilde{b}$ is a Lipschitz continuous function with Lipschitz constant $\tilde{L}$ and $\tilde{b}(0) = 0$.*

Assumption B.5. *g is a left-continuous, non-negative, non-decreasing and bounded function.*

Assumption B.6. *There exists $p > 2$ such that the p-th order moment of the law μ is finite.*

Note that for (B.6), the condition Assumption B.5 is satisfied if a is a positive constant. It follows from Assumption B.4 and Assumption B.5 that there is a constant $C < \infty$ such that for all $r \in \mathbb{R}_+$, the following linear growth condition holds,

$$\tilde{b}(r) + \sup_{p\in\mathcal{P}(\mathbb{R}_+)} p(g) \leq C(1 + |r|) \ . \tag{B.24}$$

In order to get a solution to (B.23) on $\mathbb{R}_+$ we extend the function $\tilde{b}$ to $\mathbb{R}$ by setting $\tilde{b}(r) = 0$ for $r < 0$. Note that any solution $(r_t)_{t\geq 0}$ with initial distribution supported on $\mathbb{R}_+$ satisfies almost surely $r_t \geq 0$ for all $t \geq 0$. This follows from Ito-Tanaka formula applied to $F(r) = \mathbb{1}_{(-\infty,0)}(r)r$, cf. [172, Chapter 6, Theorem 1.1]. Indeed

$$\begin{aligned}
\mathbb{1}_{(-\infty,0)}(r_t)r_t &= \int_0^t \mathbb{1}_{(-\infty,0)}(r_s)\mathrm{d}r_s - \frac{1}{2}\ell_t^0(r) \\
&= \int_0^t \mathbb{1}_{(-\infty,0)}(r_s)(\tilde{b}(r_s) + P_s(g))\mathrm{d}s + \int_0^t \mathbb{1}_{(-\infty,0)}2\mathbb{1}_{(0,\infty)}(r_s)\mathrm{d}W_s - \frac{1}{2}\ell_t^0(r) \\
&= \int_0^t \mathbb{1}_{(-\infty,0)}P_s(g)\mathrm{d}s > 0 \ ,
\end{aligned}$$

where $\ell_t^0(r)$ is the local time at 0, which vanishes, since $\mathrm{d}[r]_s = \mathbb{1}_{(0,\infty)}(r_s)\mathrm{d}s$.

Existence and uniqueness in law of (B.23) is a direct consequence of a stronger result that we now introduce. To study existence and uniqueness and to compare two solutions of (B.23) with different drifts, we establish existence of a synchronous coupling of two copies of (B.23),

$$\begin{aligned}
\mathrm{d}r_t &= (\tilde{b}(r_t) + P_t(g))\mathrm{d}t + 2\mathbb{1}_{(0,\infty)}(r_t)\mathrm{d}W_t \ , \\
\mathrm{d}s_t &= (\hat{b}(s_t) + \hat{P}_t(h))\mathrm{d}t + 2\mathbb{1}_{(0,\infty)}(s_t)\mathrm{d}W_t \ , \qquad \mathrm{Law}(r_0, s_0) = \eta \ ,
\end{aligned} \tag{B.25}$$

where $P_t = P \circ r_t^{-1}$, $\hat{P}_t = P \circ s_t^{-1}$, $(W_t)_{t\geq 0}$ is a Brownian motion and where $\eta \in \Gamma(\mu, \nu)$ for $\mu, \nu \in \mathcal{P}(\mathbb{R}_+)$.

Theorem B.3. *Suppose that* $(\tilde{b}, g)$ *and* $(\hat{b}, h)$ *satisfy Assumption B.4 and Assumption B.5. Let* $\eta \in \Gamma(\mu, \nu)$ *where the probability measures* μ *and* ν *on* $\mathbb{R}_+$ *satisfy Assumption B.6. Then there exists a weak solution* $(r_t, s_t)_{t \geq 0}$ *of the sticky stochastic differential equation* (B.25) *with initial distribution* η *defined on a probability space* $(\Omega, \mathcal{A}, P)$ *with values in* $(\mathbb{W} \times \mathbb{W}, \mathcal{B}(\mathbb{W}) \otimes \mathcal{B}(\mathbb{W}))$. *If additionally,*

$$\begin{aligned} &\tilde{b}(r) \leq \hat{b}(r) \quad \text{and} \quad g(r) \leq h(r) && \text{for any } r \in \mathbb{R}_+, \text{ and} \\ &P[r_0 \leq s_0] = 1, \end{aligned}$$

then $P[r_t \leq s_t \text{ for all } t \geq 0] = 1$.

Proof. The proof is postponed to Appendix B.6.3. □

Remark B.4. We note that by the comparison result we can deduce uniqueness in law for the solution of (B.23).

B.3.2 Invariant measures and phase transition

Under the following conditions on the drift function $\tilde{b}$ we exhibit a phase transition phenomenon for the model (B.6).

Theorem B.5. *Suppose Assumption B.4 holds and* $\limsup_{r \to \infty}(r^{-1}\tilde{b}(r)) < 0$. *Then, the Dirac measure at* 0, δ_0, *is an invariant probability measure for* (B.6). *If there exists* $p \in (0, 1)$ *solving*

$$(2/a) = (1-p) I(a, p) \tag{B.26}$$

with

$$I(a, p) = \int_0^\infty \exp\Big(\frac{1}{2} apx + \frac{1}{2} \int_0^x \tilde{b}(r) \mathrm{d}r\Big) \mathrm{d}x \ , \tag{B.27}$$

then the probability measure π *on* $[0, \infty)$ *given by*

$$\pi(\mathrm{d}x) \propto \frac{1}{Z}\Big(\frac{2}{ap}\delta_0(\mathrm{d}x) + \exp\Big(\frac{1}{2} apx + \frac{1}{2}\int_0^x \tilde{b}(r)\mathrm{d}r\Big) \lambda_{(0,\infty)}(\mathrm{d}x)\Big) \tag{B.28}$$

is another invariant probability measure for (B.6).

Proof. The proof is postponed to Appendix B.6.3. □

In our next result we specify a necessary and sufficient condition for the existence of a solution of (B.26).

Proposition B.6. *Suppose that* $\tilde{b}(r)$ *in* (B.6) *is of the form* $\tilde{b}(r) = -\tilde{L}r$ *with constant a* $\tilde{L} > 0$. *If* $a/\sqrt{\tilde{L}} > 2/\sqrt{\pi}$, *then there exists a unique* $\hat{p}$ *solving* (B.27). *In particular, the Dirac measure* δ_0 *and the measure* π *given in* (B.28) *with* $\hat{p}$ *are invariant measures for* (B.6). *On the other hand, if* $a/\sqrt{\tilde{L}} \leq 2/\sqrt{\pi}$, *then there exists no* $\hat{p}$ *solving* (B.27).

Proof. The proof is postponed to Appendix B.6.3. □

B.3.3 Convergence for sticky nonlinear SDEs

Under Assumption B.4 and the following additional assumption we establish geometric convergence in Wasserstein distance for the marginal law of the solution r_t of (B.6) to the Dirac measure at 0:

Assumption B.7. *It holds* $\limsup_{r\to\infty}(r^{-1}\tilde{b}(r)) < 0$ *and* $a \le (2\int_0^{\tilde{R}_1}\exp\left(\frac{1}{2}\int_0^s \tilde{b}(u)_+ \mathrm{d}u\right)ds)^{-1}$ *with* $\tilde{R}_0, \tilde{R}_1$ *defined by*

$$\tilde{R}_0 = \inf\{s \in \mathbb{R}_+ : \tilde{b}(r) \le 0 \ \forall r \ge s\} \qquad \textit{and} \tag{B.29}$$

$$\tilde{R}_1 = \inf\{s \ge \tilde{R}_0 : -\frac{s}{r}(s-\tilde{R}_0)\tilde{b}(r) \ge 4 \ \forall r \ge s\}\ . \tag{B.30}$$

Theorem B.7. *Suppose Assumption B.4 and Assumption B.7 holds. Then, the Dirac measure at* 0*,* δ_0*, is the unique invariant probability measure of* (B.6)*. Moreover if* $(r_s)_{s\ge 0}$ *is a solution of* (B.6) *with* r_0 *distributed with respect to an arbitrary probability measure* μ *on* $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$*, it holds for all* $t \ge 0$*,*

$$\mathbb{E}[f(r_t)] \le \mathrm{e}^{-ct}\mathbb{E}[f(r_0)]\ , \tag{B.31}$$

where f *and* c *are given by* (B.37) *and* (B.36) *with* a *and* $\tilde{b}$ *given in* (B.6) *and* $\tilde{R}_0$ *and* $\tilde{R}_1$ *given in* (B.29) *and* (B.30)*.*

Proof. The proof is postponed to Appendix B.6.3. □

B.4 Uniform in time propagation of chaos

To prove uniform in time propagation of chaos, we consider the L^1 Wasserstein distance with respect to the cost function $\bar{f}_N \circ \pi : \mathbb{R}^{Nd} \times \mathbb{R}^{Nd} \to \mathbb{R}_+$ with π given in (B.8), and $\bar{f}_N$ given by

$$\bar{f}_N((x^{i,N})_{i=1}^N, (y^{i,N})_{i=1}^N) = \frac{1}{N}\sum_{i=1}^N f\left(\left|x^i - y^i\right|\right)\ , \tag{B.32}$$

with $f : \mathbb{R}_+ \to \mathbb{R}_+$ defined in (B.37). This distance is denoted by $\mathcal{W}_{f,N}$. Note that $\bar{f}_N$ is equivalent to l^1 defined in (B.7).

We note that since π defines a projection from $\mathbb{R}^{Nd}$ to the hyperplane $\mathsf{H}_N \subset \mathbb{R}^{Nd}$ given in (B.9), for $\hat{\mu}$ and $\hat{\nu}$ on H_N, $\mathcal{W}_{f,N}(\hat{\mu},\hat{\nu})$ coincides with the Wasserstein distance given by

$$\hat{\mathcal{W}}_{f,N}(\hat{\mu},\hat{\nu}) = \inf_{\xi\in\Gamma(\hat{\mu},\hat{\nu})}\int_{\mathsf{H}_N\times\mathsf{H}_N}\bar{f}_N(x,y)\xi(\mathrm{d}x\mathrm{d}y) \tag{B.33}$$

and $\mathcal{W}_{l^1\circ\pi}(\hat{\mu},\hat{\nu}) = \hat{\mathcal{W}}_{l^1}(\hat{\mu},\hat{\nu})$, where $\bar{f}_N$ and l^1 are given in (B.32) and (B.7), respectively, and where $\hat{\mathcal{W}}_{l^1}(\hat{\mu},\hat{\nu})$ is defined as in (B.33) with respect to the distance l^1.

Theorem B.8 (Uniform in time propagation of chaos)**.** *Let* $N \in \mathbb{N}$ *and assume Assumption B.1. Let* $\bar{\mu}_0$ *and* ν_0 *be probability measures on* $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ *satisfying Assumption B.3. For* $t \ge 0$*, denote by* $\bar{\mu}_t$ *and* ν_t^N *the law of* $\bar{X}_t$ *and* $\{X_t^{i,N}\}_{i=1}^N$ *where* $(\bar{X}_s)_{s\ge0}$ *and* $(\{X_s^{i,N}\}_{i=1}^N)_{s\ge0}$ *are solutions of* (B.1) *and* (B.3)*, respectively, with initial distributions* $\bar{\mu}_0$ *and* $\nu_0^{\otimes N}$*. Then for all* $t \ge 0$*,*

$$\mathcal{W}_{f,N}(\bar{\mu}_t^{\otimes N}, \nu_t^N) \le \mathrm{e}^{-\tilde{c}t}\mathcal{W}_{f,N}(\bar{\mu}_0^{\otimes N}, \nu_0^{\otimes N}) + \tilde{C}\tilde{c}^{-1}N^{-1/2}\ ,$$
$$\mathcal{W}_{l^1\circ\pi}(\bar{\mu}_t^{\otimes N}, \nu_t^N) \le M_1\mathrm{e}^{-\tilde{c}t}\mathcal{W}_{l^1\circ\pi}(\bar{\mu}_0^{\otimes N}, \nu_0^{\otimes N}) + M_1\tilde{C}\tilde{c}^{-1}N^{-1/2}\ ,$$

where f is defined by (B.37)*,* M_1 *by* (B.18)*,* $\tilde{c}$ *by* (B.17) *and* $\tilde{C}$ *is a finite constant depending on* $\|\gamma\|_\infty$*,* L *and the second moment of* $\bar{\mu}_0$ *and given in* (B.77)*.*

Proof. The proof is postponed to Appendix B.6.4. □

Remark B.9. Denote by μ_t^N and ν_t^N the distribution of $\{X_t^{i,N}\}_{i=1}^N$ and $\{Y_t^{i,N}\}_{i=1}^N$ where the two processes $(\{X_s^{i,N}\}_{i=1}^N)_{s\geq 0}$ and $(\{Y_s^{i,N}\}_{i=1}^N)_{s\geq 0}$ are solutions of (B.3) with initial probability distributions $\mu_0^N, \nu_0^N \in \mathcal{P}(\mathbb{R}^{Nd})$, respectively, with finite forth moment. An easy inspection and adaptation of the proof of Theorem B.8 show that if Assumption B.1 holds, then

$$\mathcal{W}_{f,N}(\mu_t^N, \nu_t^N) \leq \mathrm{e}^{-\tilde{c}t}\mathcal{W}_{f,N}(\mu_0^{\otimes N}, \nu_0^{\otimes N}) , \qquad \mathcal{W}_{l^1\circ\pi}(\mu_t^N, \nu_t^N) \leq 2M_1\mathrm{e}^{-\tilde{c}t}\mathcal{W}_{l^1\circ\pi}(\mu_0^{\otimes N}, \nu_0^{\otimes N}) ,$$

where f, $\tilde{c}$ and M_1 are defined as in Theorem B.8.

B.5 System of N sticky SDEs

Consider a systerm of N one-dimensional SDEs with sticky boundaries at 0 given by

$$\mathrm{d}r_t^i = \Big(\tilde{b}(r_t^i) + \frac{1}{N}\sum_{j=1}^N g(r_t^j)\Big)\mathrm{d}t + 2\mathbb{1}_{(0,\infty)}(r_t^i)\mathrm{d}W_t^i , \qquad i = 1, \dots, N. \tag{B.34}$$

The results on existence, uniqueness and the comparison theorem for solutions of sticky nonlinear SDEs mostly carry directly over to a solution of (B.34) and are applied to prove propagation of chaos in Theorem B.8.

Let μ be a probability distribution on $\mathbb{R}_+$. For $N \in \mathbb{N}$, $(\{r_t^i, W_t^i\}_{i=1}^N)_{t\geq 0}$ is a weak solution on the filtered probability space $(\Omega, \mathcal{A}, (\mathcal{F}_t)_{t\geq 0}, P)$ of (B.34) with initial distribution $\mu^{\otimes N}$ if the following hold: $\mu^{\otimes N} = P \circ (\{r_0\}_{i=1}^N)^{-1}$, $(\{W_t\}_{i=1}^N)_{t\geq 0}$ is a N-dimensional $(\mathcal{F}_t)_{t\geq 0}$ Brownian motion w.r.t. P, the process $(r_t^i)_{t\geq 0}$ is non-negative, continuous and satisfies almost surely for any $i \in \{1, \dots, N\}$ and $t \in \mathbb{R}_+$,

$$r_t^i - r_0^i = \int_0^t \Big(\tilde{b}(r_s^i) + \frac{1}{N}\sum_{j=1}^N g(r_s^j)\Big)\mathrm{d}s + \int_0^t 2\mathbb{1}_{(0,\infty)}(r_s^i)\mathrm{d}W_s^i .$$

To show existence and uniqueness in law of a weak solution $(\{r_t^i, W_t^i\}_{i=1}^N)_{t\geq 0}$, we suppose Assumption B.4 and Assumption B.5 for $\tilde{b}$ and g.

It follows that there exists a constant $C < \infty$ such that for all $\{r^i\}_{i=1}^N \in \mathbb{R}_+^N$, it holds $\sum_{i=1}^N |\tilde{b}(r^i)| + |g(r^i)| \leq C(1 + \sum_{i=1}^N |r^i|)$, and a possible solution $(\{r_t^i\}_{i=1}^N)_{t\geq 0}$ is non-explosive. If the initial distribution is supported on $\mathbb{R}_+^N$, then in the same line as for the nonlinear SDE in Appendix B.3.1, the solution $(\{r_t^i\}_{i=1}^N)_{t\geq 0}$ satisfies $r_t^i > 0$ almost surely for any $i = 1, \dots, N$ and $t \geq 0$ by Assumption B.4 and Assumption B.5.

Existence and uniqueness in law of (B.34) is a direct consequence of a stronger result that we now introduce. To study existence and uniqueness and to compare two solutions of (B.34) with different drifts, we establish existence of a synchronous coupling of two copies of (B.34),

$$\begin{aligned} \mathrm{d}r_t^i &= \Big(\tilde{b}(r_t^i) + \frac{1}{N}\sum_{j=1}^N g(r_t^j)\Big)\mathrm{d}t + 2\mathbb{1}_{(0,\infty)}(r_t^i)\mathrm{d}W_t^i , \\ \mathrm{d}s_t^i &= \Big(\hat{b}(s_t^i) + \frac{1}{N}\sum_{j=1}^N h(s_t^j)\Big)\mathrm{d}t + 2\mathbb{1}_{(0,\infty)}(s_t^i)\mathrm{d}W_t^i , \qquad \text{for } i \in \{1, \dots, N\} \\ \mathrm{Law}(r_0^i, s_0^i) &= \eta , \end{aligned} \tag{B.35}$$

where $(\{W_t^i\}_{i=1}^N)_{t\geq 0}$ are N i.i.d.1-dimensional Brownian motions and where $\eta \in \Gamma(\mu,\nu)$ for $\mu,\nu \in \mathcal{P}(\mathbb{R}_+)$.

Let $\mathbb{W}^N = \mathcal{C}(\mathbb{R}_+,\mathbb{R}^N)$ be the space of continuous functions from $\mathbb{R}_+$ to $\mathbb{R}^N$ endowed with the topology of uniform convergence on compact sets, and let $\mathcal{B}(\mathbb{W}^N)$ denote its Borel σ-Algebra.

Theorem B.10. *Assume that $(\tilde{b},g)$ and $(\hat{b},h)$ satisfy Assumption B.4 and Assumption B.5. Let $\eta \in \Gamma(\mu,\nu)$ where μ and ν are the probability measure on $\mathbb{R}_+$ satisfying Assumption B.6. Then there exists a weak solution $(\{r_t^i,s_t^i\}_{i=1}^N)_{t\geq 0}$ of the sticky stochastic differential equation* (B.35) *with initial distribution $\eta^{\otimes N}$ defined on a probability space $(\Omega,\mathcal{A},P)$ with values in $\mathbb{W}^N\times\mathbb{W}^N$. If additionally,*

$$\begin{aligned}&\tilde{b}(r)\leq\hat{b}(r) \quad \text{and} \quad g(r)\leq h(r)\;, && \text{for any } r\in\mathbb{R}_+\;,\\ &P[r_0^i\leq s_0^i \text{ for all } i=1,\dots,N]=1\;,\end{aligned}$$

then $P[r_t^i\leq s_t^i$ for all $t\geq 0$ and $i=1,\dots,N]=1$.

Proof. The proof is postponed to Appendix B.6.5. □

Remark B.11. We note that by the comparison result we can deduce uniqueness in law for the solution of (B.34).

B.6 Proofs

B.6.1 Definition of the metrics

In Theorem B.1, Theorem B.7 and Theorem B.8 we consider Wasserstein distances based on a carefully designed concave function $f:\mathbb{R}_+\to\mathbb{R}_+$ that we now define. In addition we derive useful properties of this function that will be used in our proofs of Theorem B.1, Theorem B.8 and Theorem B.7. Let $a\in\mathbb{R}_+$ and $\tilde{b}:\mathbb{R}_+\to\mathbb{R}$ be such that Assumption B.7 is satisfied with $\tilde{R}_0$ and $\tilde{R}_1$ defined in (B.29). We define

$$\begin{aligned}&\varphi(r)=\exp\left(-\int_0^r\{\tilde{b}(s)_+/2\}\mathrm{d}s\right)\;, \qquad \Phi(r)=\int_0^r\varphi(s)\mathrm{d}s\;, \qquad \text{and}\\ &g(r)=1-\frac{c}{2}\int_0^{r\wedge\tilde{R}_1}\{\Phi(s)/\varphi(s)\}\mathrm{d}s-\frac{a}{2}\int_0^{r\wedge\tilde{R}_1}\{1/\varphi(s)\}\mathrm{d}s\;,\end{aligned}$$

where

$$c=\left(2\int_0^{\tilde{R}_1}\{\Phi(s)/\varphi(s)\}\mathrm{d}s\right)^{-1}\;, \tag{B.36}$$

and $\tilde{R}_1$ is given in (B.30). It holds $\varphi(r)=\varphi(\tilde{R}_0)$ for $r\geq\tilde{R}_0$ with $\tilde{R}_0$ given in (B.29), $g(r)=g(\tilde{R}_1)\in[1/2,3/4]$ for $r\geq\tilde{R}_1$ and $g(r)\in[1/2,1]$ for all $r\in\mathbb{R}_+$ by (B.36) and Assumption B.7. We define the increasing function $f:[0,\infty)\to[0,\infty)$ by

$$f(t)=\int_0^t\varphi(r)g(r)\mathrm{d}r\;. \tag{B.37}$$

Note that f is concave, since φ and g are decreasing. Since for all $r\in\mathbb{R}_+$

$$\varphi(\tilde{R}_0)r/2\leq\Phi(r)/2\leq f(r)\leq\Phi(r)\leq r\;, \tag{B.38}$$

$(x, y) \mapsto f(|x-y|)$ defines a distance on $\mathbb{R}^d$ equivalent to the Euclidean distance on $\mathbb{R}^d$.

Moreover, f satisfies

$$2f''(0) = -\tilde{b}(0)_+ - a = -a \ , \tag{B.39}$$

and

$$2f''(r) \leq 2f''(0) - f'(r)\tilde{b}(r) - cf(r) \ , \qquad \text{for all } r \in \mathbb{R}_+ \backslash \{\tilde{R}_1\} \ . \tag{B.40}$$

Indeed by construction of f, $f''(r) = -\tilde{b}(r)_+ f'(r)/2 - c\Phi(r)/2 - a/2$ for $0 \leq r < \tilde{R}_1$ and so (B.40) holds for $0 \leq r < \tilde{R}_1$ by (B.38). To show (B.40) for $r > \tilde{R}_1$ note that $f''(r) = 0$ and $f'(r) \geq \varphi(\tilde{R}_0)/2$ hold for $r > \tilde{R}_1$. Hence, by the definition (B.30) of $\tilde{R}_1$, for $r > \tilde{R}_1$,

$$f''(r) + f'(r)\tilde{b}(r)/2 \leq \varphi(\tilde{R}_0)\tilde{b}(r)/4 \leq -(\tilde{R}_1(\tilde{R}_1 - \tilde{R}_0))^{-1}\varphi(\tilde{R}_0) r \ . \tag{B.41}$$

Since $\varphi(r) = \varphi(\tilde{R}_0)$ for $r \geq \tilde{R}_0$, it holds $\Phi(r) = \Phi(\tilde{R}_0) + (r - \tilde{R}_0)\varphi(\tilde{R}_0)$ for $r \geq \tilde{R}_0$. Further, it holds $\Phi(R_0) \geq \tilde{R}_0\varphi(\tilde{R}_0)$ since φ is decreasing for $r \leq \tilde{R}_0$. Hence,

$$\frac{r}{\tilde{R}_1} = \frac{(r - \tilde{R}_1)(\Phi(\tilde{R}_0) + (\tilde{R}_1 - \tilde{R}_0)\varphi(\tilde{R}_0))}{\tilde{R}_1\Phi(\tilde{R}_1)} + 1 \geq \frac{(r - \tilde{R}_1)\tilde{R}_1\varphi(\tilde{R}_0)}{\tilde{R}_1\Phi(\tilde{R}_1)} + 1 = \frac{\Phi(r)}{\Phi(\tilde{R}_1)} \ . \tag{B.42}$$

Furthermore, we have

$$\begin{aligned} \int_{\tilde{R}_0}^{\tilde{R}_1} \{\Phi(s)/\varphi(s)\} \mathrm{d}s &= \int_{\tilde{R}_0}^{\tilde{R}_1} \frac{\Phi(\tilde{R}_0) + (s - \tilde{R}_0)\varphi(\tilde{R}_0)}{\varphi(\tilde{R}_0)} \mathrm{d}s \\ &= (\tilde{R}_1 - \tilde{R}_0)\frac{\Phi(\tilde{R}_0)}{\varphi(\tilde{R}_0)} + \frac{1}{2}(\tilde{R}_1 - \tilde{R}_0)^2 \geq \frac{1}{2}(\tilde{R}_1 - \tilde{R}_0)\frac{\Phi(\tilde{R}_1)}{\varphi(\tilde{R}_0)} \ . \end{aligned} \tag{B.43}$$

We insert (B.42) and (B.43) in (B.41) and use (B.36) to obtain

$$f''(r) + f'(r)\tilde{b}(r)/2 \leq -\Phi(r)\Phi(\tilde{R}_1)^{-1}(\tilde{R}_1 - \tilde{R}_0)^{-1}\varphi(\tilde{R}_0) \tag{B.44}$$

$$\leq -\frac{\Phi(r)}{2\int_{\tilde{R}_0}^{\tilde{R}_1}\{\Phi(s)/\varphi(s)\}\mathrm{d}s} \leq -\frac{cf(r)}{2} - \frac{c\Phi(r)}{2} \ . \tag{B.45}$$

By Assumption B.7 and (B.36), we get

$$-\frac{c\Phi(r)}{2} \leq -\frac{\Phi(\tilde{R}_1)}{4\int_0^{\tilde{R}_1}\{\Phi(s)/\varphi(s)\}\mathrm{d}s} \leq -\frac{1}{4\int_0^{\tilde{R}_1}\{1/\varphi(s)\}\mathrm{d}s} \leq -\frac{a}{2} = f''(0) \ .$$

Combining this estimate with (B.44) gives (B.40) for $r > \tilde{R}_1$.

B.6.2 Proof of Appendix B.2

Proof of Theorem B.1

Proof of Theorem B.1. We consider the process $(\bar{X}_t, \bar{Y}_t, r_t)_{t \geq 0}$ defined in Theorem B.2 and satisfying $|\bar{X}_t - \bar{Y}_t| \leq r_t$ for any $t \geq 0$, and $(r_t)_{t \geq 0}$ is a weak solution of (B.22). Set $a = 2\|\gamma\|_\infty$ and $\tilde{b}(r) = \bar{b}(r)$. With this notation, Assumption B.1 and Assumption B.2 imply Assumption B.7

and $\tilde{R}_0 = R_0$ and $\tilde{R}_1 = R_1$ by (B.14), (B.15), (B.29) and (B.30). By Ito-Tanaka formula, cf. [172, Chapter 6, Theorem 1.1], using that f' is absolutely continuous, we have,

$$\begin{aligned}\mathrm{d}f(r_t) &\leq f'(r_t)(\bar{b}(r_t) + 2\|\gamma\|_\infty \mathbb{P}(r_t > 0))\mathrm{d}t + 2f''(r_t)\mathbb{1}_{(0,\infty)}(r_t)\mathrm{d}t \\ &\quad + f'(r_t)2\mathbb{1}_{(0,\infty)}(r_t)\mathrm{d}W_t \ .\end{aligned}$$

Taking expectation we obtain by (B.39) and (B.40)

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[f(r_t)] \leq \mathbb{E}[f'(r_t)\tilde{b}(r_t)_+ + 2(f''(r_t) - f''(0))] + \mathbb{E}[(a + 2f''(0))\mathbb{1}_{r_t>0}] \leq -\tilde{c}\mathbb{E}[f(r_t)] \ ,$$

where $\tilde{c}$ is given by (B.17). Therefore by Grönwall's lemma,

$$\mathbb{E}[f(|\bar{X}_t - \bar{Y}_t|)] \leq \mathbb{E}[f(r_t)] \leq \mathrm{e}^{-\tilde{c}t}\mathbb{E}[f(r_0)] = \mathrm{e}^{-\tilde{c}t}\mathbb{E}[f(|\bar{X}_0 - \bar{Y}_0|)] \ .$$

Hence, it holds

$$\mathcal{W}_f(\bar{\mu}_t, \bar{\nu}_t) \leq \mathbb{E}[f(|\bar{X}_t - \bar{Y}_t|)] \leq \mathrm{e}^{-\tilde{c}t}\int_{\mathbb{R}^d\times\mathbb{R}^d} f(|x-y|)\xi(\mathrm{d}x\mathrm{d}y)$$

for an arbitrary coupling $\xi \in \Gamma(\mu_0, \nu_0)$. Taking the infimum over all couplings $\xi \in \Gamma(\mu_0, \nu_0)$, we obtain the first inequality of (B.16). By (B.38), we get the second inequality of (B.16). □

Proof of Theorem B.2

Note that the nonlinear SDE (B.21) has Lipschitz continuous coefficients. The existence and the uniqueness of the coupling $(\bar{X}_t^\delta, \bar{Y}_t^\delta)_{t\geq 0}$ follows from [146, Theorem 2.2]. By Levy's characterization, $(\bar{X}_t^\delta, \bar{Y}_t^\delta)_{t\geq 0}$ is indeed a coupling of two copies of solutions of (B.1). Further, we remark that $W_t^\delta = \int_0^t (\bar{e}_s^\delta)^T \mathrm{d}B_s^1$ is a one-dimensional Brownian motion. In the next step, we analyse $|\bar{X}_t^\delta - \bar{Y}_t^\delta|$.

Lemma B.12. *Suppose that the conditions Assumption B.1 and Assumption B.3 are satisfied. Then, it holds for any $\epsilon < \epsilon_0$, where ϵ_0 is given by* (B.20)*, setting $\bar{r}_t^\delta = |\bar{X}_t^\delta - \bar{Y}_t^\delta|$*

$$\mathrm{d}\bar{r}_t^\delta = \Big(-L\bar{r}_t^\delta + \Big\langle \bar{e}_t^\delta, \int_{\mathbb{R}}\int_{\mathbb{R}^d} \gamma(\bar{X}_t^\delta - x) - \gamma(\bar{Y}_t^\delta - y)\mu_t^\delta(\mathrm{d}x)\nu_t^\delta(\mathrm{d}y)\Big\rangle\Big)\mathrm{d}t + 2\mathrm{rc}^\delta(\bar{r}_t^\delta)\mathrm{d}W_t^\delta \quad \text{(B.46)}$$

$$\leq \Big(\bar{b}(\bar{r}_t^\delta) + 2\|\gamma\|_\infty \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mathrm{rc}^\epsilon(|x-y|)\bar{\mu}_t^\delta(\mathrm{d}x)\bar{\nu}_t^\delta(\mathrm{d}y)\Big)\mathrm{d}t + 2\mathrm{rc}^\delta(\bar{r}_t^\delta)\mathrm{d}W_t^\delta \ , \quad \text{(B.47)}$$

almost surely for all $t \geq 0$, where $\bar{\mu}_t^\delta$ and $\bar{\nu}_t^\delta$ are the laws of $\bar{X}_t^\delta$ and $\bar{Y}_t^\delta$, respectively.

Proof. Using (B.21), Assumption B.1 and Assumption B.3, the stochastic differential equation of the process $((\bar{r}_t^\delta)^2)_{t\geq 0}$ is given by

$$\begin{aligned}\mathrm{d}((\bar{r}_t^\delta)^2) &= 2\Big\langle Z_t^\delta, -LZ_t^\delta + \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \gamma(\bar{X}_t^\delta - x) - \gamma(\bar{Y}_t^\delta - y)\bar{\mu}_t^\delta(\mathrm{d}x)\bar{\nu}_t^\delta(\mathrm{d}y)\Big\rangle\mathrm{d}t \\ &\quad + 4\mathrm{rc}^\delta(\bar{r}_t^\delta)^2\mathrm{d}t + 4\mathrm{rc}^\delta(\bar{r}_t^\delta)\langle Z_t^\delta, e_t^\delta\rangle\mathrm{d}W_t^\delta \ .\end{aligned}$$

For $\varepsilon > 0$ we define as in [87, Lemma 8] a $\mathcal{C}^2$ approximation of the square root by

$$S_\varepsilon(r) = \begin{cases}(-1/8)\varepsilon^{-3/2}r^2 + (3/4)\varepsilon^{-1/2}r + (3/8)\varepsilon^{1/2} & \text{for } r < \varepsilon \\ \sqrt{r} & \text{otherwise} \ .\end{cases}$$

Then, by Ito's formula,

$$\begin{aligned}\mathrm{d}S_\varepsilon((\bar r_t^\delta)^2) &= S'_\varepsilon((\bar r_t^\delta)^2)\mathrm{d}(\bar r_t^\delta)^2 + \frac{1}{2}S''_\varepsilon((\bar r_t^\delta)^2)\mathrm{d}[(\bar r^\delta)^2]_t \\ &= 2S'_\varepsilon((\bar r_t^\delta)^2)\Big\langle Z_t^\delta, -LZ_t^\delta + \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\gamma(\bar X_t^\delta - x) - \gamma(\bar Y_t^\delta - y)\bar\mu_t^\delta(\mathrm{d}x)\bar\nu_t^\delta(\mathrm{d}y)\Big\rangle \mathrm{d}t \\ &+ S'_\varepsilon((\bar r_t^\delta)^2)4\mathrm{rc}^\delta(\bar r_t^\delta)^2\mathrm{d}t + S'_\varepsilon((\bar r_t^\delta)^2)4\mathrm{rc}^\delta(\bar r_t^\delta)\langle Z_t^\delta, e_t^\delta\rangle \mathrm{d}W_t^\delta + 8S''_\varepsilon((\bar r_t^\delta)^2)(\mathrm{rc}^\delta(\bar r_t^\delta))^2(\bar r_t^\delta)^2\mathrm{d}t\ .\end{aligned}$$

We take the limit $\varepsilon \to 0$. Then $\lim_{\varepsilon\to 0} S'_\varepsilon(r) = (1/2)r^{-1/2}$ and $\lim_{\varepsilon\to 0} S''_\varepsilon(r) = -(1/4)r^{-3/2}$ for $r > 0$. Since $\sup_{0\le r\le\varepsilon}|S'_\varepsilon(r)| \lesssim \varepsilon^{-1/2}$, $\sup_{0\le r\le\epsilon}|S''_{\bar\varepsilon}(r)| \lesssim \bar\varepsilon^{-3/2}$ and rc^δ is Lipschitz continuous with $\mathrm{rc}^\delta(0) = 0$, we apply Lebesgue's dominated convergence theorem to show convergence for the integrals with respect to time t. More precisely, we note that the integrand $(4S'_\varepsilon((\bar r_t^\delta)^2) + 8S''_\varepsilon((\bar r_t^\delta)^2))\mathrm{rc}^\delta(\bar r_t^\delta))^2(\bar r_t^\delta)^2$ is dominated by $3\varepsilon^{1/2}\|\mathrm{rc}^\delta\|_{\mathrm{Lip}}$. For any $\varepsilon < \varepsilon_0$ for fixed $\varepsilon_0 > 0$, the integrand $2S'_\varepsilon((\bar r_t^\delta)^2)\langle Z_t^\delta, -LZ_t^\delta + \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}(\gamma(\bar X_t^\delta - x) - \gamma(\bar Y_t^\delta - y))\bar\mu_t^\delta(\mathrm{d}x)\bar\nu_t^\delta(\mathrm{d}y)\rangle$ is dominated by $(3/2)(L\max(\varepsilon_0^{(1/2)}, \bar r_t^\delta) + 2\|\gamma\|_\infty)$.

For the stochastic integral it holds $|S'_\varepsilon((\bar r_t^\delta)^2)4\mathrm{rc}^\delta(\bar r_t^\delta)\bar r_t^\delta| \le 3$. Hence, the stochastic integral converges along a subsequence almost surely, to $\int_0^t 2\mathrm{rc}^\delta(\bar r_s^\delta)\mathrm{d}W_s^\delta$, see [172, Chapter 4, Theorem 2.12]. Hence, we obtain (B.46). By Assumption B.1 and (B.20), we obtain for $\epsilon < \epsilon_0$

$$\begin{aligned}&\Big\langle \bar e_t^\delta, \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}(\gamma(\bar X_t^\delta - x) - \gamma(\bar Y_t^\delta - y))\mu_t^\delta(\mathrm{d}x)\nu_t^\delta(\mathrm{d}y)\Big\rangle \\ &\quad\le \Big\langle \bar e_t^\delta, \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}(\gamma(\bar X_t^\delta - x) - \gamma(\bar Y_t^\delta - x) + \gamma(\bar Y_t^\delta - x) - \gamma(\bar Y_t^\delta - y))\mu_t^\delta(\mathrm{d}x)\nu_t^\delta(\mathrm{d}y)\Big\rangle \\ &\quad\le \kappa(\bar r_t^\delta)\bar r_t^\delta + \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} 2\|\gamma\|_\infty \mathrm{rc}^\epsilon(|x-y|)\mu_t^\delta(\mathrm{d}x)\nu_t^\delta(\mathrm{d}y)\ ,\end{aligned}$$

and hence (B.47) holds. □

We define a one-dimensional process $(r_t^{\delta,\epsilon})_{t\ge 0}$ by

$$\mathrm{d}r_t^{\delta,\epsilon} = \Big(\bar b(r_t^{\delta,\epsilon}) + 2\|\gamma\|_\infty\int_{\mathbb{R}_+}\mathrm{rc}^\epsilon(u)P_t^{\delta,\epsilon}(\mathrm{d}u)\Big)\mathrm{d}t + 2\mathrm{rc}^\delta(r_t^{\delta,\epsilon})\mathrm{d}W_t^\delta \tag{B.48}$$

with initial condition $r_0^{\delta,\epsilon} = \bar r_0^\delta$, $P_t^{\delta,\epsilon} = \mathrm{Law}(r_t^{\delta,\epsilon})$ and $W_t^\delta = \int_0^t(\bar e_s^\delta)^T\mathrm{d}B_s^1$. This process will allow us to control the distance of $\bar X_t^\delta$ and $\bar Y_t^\delta$.

By [146, Theorem 2.2], under Assumption B.1 and Assumption B.3, $(U_t^{\delta,\epsilon})_{t\ge 0} = (\bar X_t^\delta, \bar Y_t^\delta, r_t^{\delta,\epsilon})_{t\ge 0}$ exists and is unique, where $(\bar X_t^\delta, \bar Y_t^\delta)_{t\ge 0}$ solves uniquely (B.21), $(\bar r_t^\delta)_{t\ge 0}$ and $(r_t^{\delta,\epsilon})_{t\ge 0}$ solve uniquely (B.46) and (B.48), respectively, with $W_t^\delta = \int_0^t(\bar e_s^\delta)^T\mathrm{d}B_s^1$.

Lemma B.13. *Assume Assumption B.1 and Assumption B.3. Then,* $|\bar X_t^\delta - \bar Y_t^\delta| = \bar r_t^\delta \le r_t^{\delta,\epsilon}$*, almost surely for all* t *and* $\epsilon < \epsilon_0$*.*

Proof. Note that $(\bar r_t^\delta)_{t\ge 0}$ and $(r_t^{\delta,\epsilon})_{t\ge 0}$ have the same initial distribution and are driven by the same noise. Since the drift of $(\bar r_t^\delta)_{t\ge 0}$ is smaller than the drift of $(r_t^{\delta,\epsilon})_{t\ge 0}$ for $\epsilon < \epsilon_0$, the result follows by Lemma B.14.

□

Proof of Theorem B.2. We consider the nonlinear process $(U_t^{\delta,\epsilon})_{t\ge 0} = (\bar X_t^\delta, \bar Y_t^\delta, r_t^{\delta,\epsilon})_{t\ge 0}$ on $\mathbb{R}^{2d+1}$ for each $\epsilon, \delta > 0$. We denote by $\mathbb{P}^{\delta,\epsilon}$ the law of $U^{\delta,\epsilon}$ on the space $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^{2d+1})$. We define by $\mathbf{X}, \mathbf{Y} : \mathcal{C}(\mathbb{R}_+, \mathbb{R}^{2d+1}) \to \mathcal{C}(\mathbb{R}_+, \mathbb{R}^d)$ and $\mathbf{r} : \mathcal{C}(\mathbb{R}_+, \mathbb{R}^{2d+1}) \to \mathcal{C}(\mathbb{R}_+, \mathbb{R})$ the canonical projections onto the first d components, onto the second d components and onto the last component, respectively. By Assumption B.1 and Assumption B.3 following the same line as the proof of Lemma B.15, see (B.56), it holds for each $T > 0$

$$E[|U_{t_2}^{\delta,\epsilon} - U_{t_1}^{\delta,\epsilon}|^4] \leq C|t_2 - t_1|^2 \qquad \text{for } t_1, t_2 \in [0,T] \;, \tag{B.49}$$

for some constant C depending on T, L, $\|\gamma\|_{\mathrm{Lip}}$, $\|\gamma\|_\infty$ and on the fourth moment of μ_0 and ν_0. As in Lemma B.15 the law $\mathbb{P}_T^{\delta,\epsilon}$ of $(U_t^{\delta,\epsilon})_{0\leq t\leq T}$ on $\mathcal{C}([0,T],\mathbb{R}^{2d+1})$ is tight for each $T > 0$ by [121, Corollary 14.9] and for each $\epsilon > 0$ there exists a subsequence $\delta_n \to 0$ such that $(\mathbb{P}_T^{\delta_n,\epsilon})_{n\in\mathbb{N}}$ on $\mathcal{C}([0,T],\mathbb{R}^{2d+1})$ converge to a measure $\mathbb{P}_T^\epsilon$ on $\mathcal{C}([0,T],\mathbb{R}^{2d+1})$. By a diagonalization argument and since $\{\mathbb{P}_T^\epsilon : T \geq 0\}$ is a consistent family, cf. [121, Theorem 5.16], there exists a probability measure $\mathbb{P}^\epsilon$ on $\mathcal{C}(\mathbb{R}_+,\mathbb{R}^{2d+1})$ such that for all ϵ there exists a subsequence δ_n such that $(\mathbb{P}^{\delta_n,\epsilon})_{n\in\mathbb{N}}$ converges along this subsequence to $\mathbb{P}^\epsilon$. As in the proof of Lemma B.16 we repeat this argument for the family of measures $(\mathbb{P}^\epsilon)_{\epsilon>0}$. Hence, there exists a subsequence $\epsilon_m \to 0$ such that $(\mathbb{P}^{\epsilon_m})_{m\in\mathbb{N}}$ converges to a measure $\mathbb{P}$. Let $(\bar{X}_t, \bar{Y}_t, r_t)_{t\geq 0}$ be some process on $\mathbb{R}^{2d+1}$ with distribution $\mathbb{P}$ on $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P})$.

Since $(\bar{X}_t^\delta)_{t\geq 0}$ and $(\bar{Y}_t^\delta)_{t\geq 0}$ are solutions of (B.1) which are unique in law, we have that for any $\epsilon, \delta > 0$, $\mathbb{P}^{\delta,\epsilon} \circ \mathbf{X}^{-1} = \mathbb{P} \circ \mathbf{X}^{-1}$ and $\mathbb{P}^{\delta,\epsilon} \circ \mathbf{Y}^{-1} = \mathbb{P} \circ \mathbf{Y}^{-1}$. And therefore $(\bar{X}_t)_{t\geq 0}$ and $(\bar{Y}_t)_{t\geq 0}$ are solutions of (B.1) as well with the same initial condition. Hence $\mathbb{P} \circ (\mathbf{X}, \mathbf{Y})^{-1}$ is a coupling of two copies of (B.1).

Similarly to the proof of Lemma B.15 and Lemma B.16 there exist an extended probability space and a one-dimensional Brownian motion $(W_t)_{t\geq 0}$ such that $(r_t, W_t)_{t\geq 0}$ is a solution to

$$\mathrm{d}r_t = (\bar{b}(r_t) + 2\|\gamma\|_\infty \mathbb{P}(r_t > 0))\mathrm{d}t + 2\mathbb{1}_{(0,\infty)}(r_t)\mathrm{d}W_t \;.$$

In addition, the statement of Lemma B.13 carries over to the limiting process $(r_t)_{t\geq 0}$, i.e., $|\bar{X}_t - \bar{Y}_t| \leq r_t$ for all $t \geq 0$, since by the weak convergence along the subsequences $(\delta_n)_{n\in\mathbb{N}}$ and $(\epsilon_m)_{m\in\mathbb{N}}$ and the Portmanteau theorem, $P(|\bar{X}_t - \bar{Y}_t| \leq r_t) \geq \limsup_{m\to\infty} \limsup_{n\to\infty} P(|\bar{X}_t^{\delta_n} - \bar{Y}_t^{\delta_n}| \leq r_t^{\delta_n,\epsilon_m}) = 1$.

□

B.6.3 Proof of Appendix B.3

Proof of Theorem B.3

We show Theorem B.3 via a family of stochastic differential equations, indexed by $n, m \in \mathbb{N}$, with Lipschitz continuous coefficients,

$$\begin{aligned} \mathrm{d}r_t^{n,m} &= (\tilde{b}(r_t^{n,m}) + P_t^{n,m}(g^m))\mathrm{d}t + 2\theta^n(r_t^{n,m})\mathrm{d}W_t \\ \mathrm{d}s_t^{n,m} &= (\hat{b}(s_t^{n,m}) + \hat{P}_t^{n,m}(h^m))\mathrm{d}t + 2\theta^n(s_t^{n,m})\mathrm{d}W_t \;, \quad \mathrm{Law}(r_0^{n,m}, s_0^{n,m}) = \eta_{n,m} \;, \end{aligned} \tag{B.50}$$

where $P_t^{n,m} = \mathrm{Law}(r_t^{n,m})$, $\hat{P}_t^{n,m} = \mathrm{Law}(s_t^{n,m})$, $P_t^{n,m}(g^m) = \int_{\mathbb{R}_+} g^m(r) P_t^{n,m}(\mathrm{d}x)$ and $\hat{P}_t^{n,m}(h^m) = \int_{\mathbb{R}_+} h^m(r)\hat{P}_t^{n,m}(\mathrm{d}x)$ for some measurable functions $(g^m)_{m\in\mathbb{N}}$ and $(h^m)_{m\in\mathbb{N}}$, and where $\eta_{n,m} \in \Gamma(\mu_{n,m}, \nu_{n,m})$ for $\mu_{n,m}, \nu_{n,m} \in \mathcal{P}(\mathbb{R}_+)$. We identify the weak limit for $n \to \infty$ as solution of a family of stochastic differential equations, indexed by $m \in \mathbb{N}$, given by

$$\begin{aligned} \mathrm{d}r_t^m &= (\tilde{b}(r_t^m) + P_t^m(g^m))\mathrm{d}t + 2\mathbb{1}_{(0,\infty)}(r_t^m)\mathrm{d}W_t \\ \mathrm{d}s_t^m &= (\hat{b}(s_t^m) + \hat{P}_t^m(h^m))\mathrm{d}t + 2\mathbb{1}_{(0,\infty)}(s_t^m)\mathrm{d}W_t \;, \quad \mathrm{Law}(r_0^m, s_0^m) = \eta_m \;. \end{aligned} \tag{B.51}$$

with $P_t^m = \text{Law}(r_t^m)$ and $\hat{P}_t^m = \text{Law}(s_t^m)$, and where $\eta_m \in \Gamma(\mu_m, \nu_m)$ for $\mu_m, \nu_m \in \mathcal{P}(\mathbb{R}_+)$. Taking the limit $m \to \infty$, we show in the next step that the solution of (B.51) converges to a solution of (B.25).

We assume for $(g^m)_{m\in\mathbb{N}}$, $(h^m)_{m\in\mathbb{N}}$, $(\theta^n)_{n\in\mathbb{N}}$ and the initial distributions:

Assumption B.8. *$(g^m)_{m\in\mathbb{N}}$ and $(h^m)_{m\in\mathbb{N}}$ are sequences of non-decreasing non-negative uniformly bounded Lipschitz continuous functions such that for all $r \geq 0$, $g^m(r) \leq g^{m+1}(r)$ and $h^m(r) \leq h^{m+1}(r)$ and $\lim_{m\to+\infty} g^m(r) = g(r)$ and $\lim_{m\to+\infty} h^m(r) = h(r)$ where g, h are left-continuous non-negative non-decreasing bounded functions. In addition, there exists $K_m < \infty$ for any m such that for all $r, s \in \mathbb{R}$*

$$|g^m(r) - g^m(s)| \leq K_m|r - s| \qquad \text{and} \qquad |h^m(r) - h^m(s)| \leq K_m|r - s| \ .$$

Assumption B.9. *$(\theta^n)_{n\in\mathbb{N}}$ is a sequence of Lipschitz continuous functions from $\mathbb{R}_+$ to $[0, 1]$ with $\theta^n(0) = 0$, $\theta^n(r) = 1$ for all $r \geq 1/n$ and $\theta^n(r) > 0$ for all $r > 0$.*

Assumption B.10. *$(\mu_{n,m})_{m,n\in\mathbb{N}}$, $(\nu_{n,m})_{m,n\in\mathbb{N}}$, $(\mu_m)_{m\in\mathbb{N}}$, $(\nu_m)_{m\in\mathbb{N}}$ are families of probability distributions on $\mathbb{R}_+$ and $(\eta_{n,m})_{n,m\in\mathbb{N}}$, $(\eta_m)_{m\in\mathbb{N}}$ families of probability distributions on $\mathbb{R}_+^2$ such that for any $n, m \in \mathbb{N}$ $\eta_{n,m} \in \Gamma(\mu_{n,m}, \nu_{n,m})$ and $\eta_m \in \Gamma(\mu_m, \nu_m)$ and for any $m \in \mathbb{N}$, $(\eta_{n,m})_{n\in\mathbb{N}}$ converges weakly to η_m and $(\eta_m)_{m\in\mathbb{N}}$ converges weakly to η. Further, the p-th order moments of $(\mu_{n,m})_{n,m\in\mathbb{N}}$, $(\nu_{n,m})_{n,m\in\mathbb{N}}$, $(\mu_m)_{m\in\mathbb{N}}$ and $(\nu_m)_{m\in\mathbb{N}}$ are uniformly bounded for $p > 2$ given in Assumption B.6.*

Note that by Assumption B.8 for any non-decreasing sequence $(u_m)_{m\in\mathbb{N}}$, which converges to $u \in \mathbb{R}_+$, $g^m(u_m)$ and $h^m(u_m)$ converge to $g(u)$ and $h(u)$, respectively. More precisely, it holds for for all $m \in \mathbb{N}$, $g^m(u_m) - g(u) \leq 0$ and for $m \geq n$, $g^m(u_m) \geq g^m(u_n)$ and therefore, $\lim_{m\to\infty} g^m(u_n) - g(u) \geq \lim_{n\to\infty} \lim_{m\to\infty} = \lim_{n\to\infty} g(u_n) - g(u) = 0$ by left-continuity of g. Hence, $\lim_{m\to\infty} g^m(u_m) - g(u) = 0$ and analogously $\lim_{m\to\infty} h^m(u_m) - h(u) = 0$. By Assumption B.8, $\Gamma = \max(\|h\|_\infty, \|g\|_\infty)$ is a uniform upper bound of $(g^m)_{m\in\mathbb{N}}$ and $(h^m)_{m\in\mathbb{N}}$.

Consider a probability space $(\Omega_0, \mathcal{A}_0, Q)$ and a one-dimensional Brownian motion $(W_t)_{t\geq 0}$. Under Assumption B.8, Assumption B.9 and Assumption B.10, for all $m, n \in \mathbb{N}$, there exists random variables $r^{n,m}, s^{n,m} : \Omega_0 \to \mathbb{W}$ for each n, m such that $(r_t^{n,m}, s_t^{n,m})_{t\geq 0}$ is a unique strong solution to (B.50) associated to $(W_t)_{t\geq 0}$ by [146, Theorem 2.2]. We denote by $\mathbb{P}^{n,m} = Q \circ (r^{n,m}, s^{n,m})^{-1}$ the corresponding distribution on $\mathbb{W} \times \mathbb{W}$.

Before studying the two limits $n, m \to \infty$ and proving Theorem B.3, we state a modification of the comparison theorem by Ikeda and Watanabe to compare two solutions of (B.50), cf. [116, Section VI, Theorem 1.1].

Lemma B.14. *Let $(r_t^{n,m}, s_t^{n,m})_{t\geq 0}$ be a solution of (B.50) for fixed $n, m \in \mathbb{N}$. Assume Assumption B.4, Assumption B.8 and Assumption B.9. If $Q[r_0^{n,m} \leq s_0^{n,m}] = 1$, $\tilde{b}(r) \leq \hat{b}(r)$ and $g^m(r) \leq h^m(r)$ for any $r \in \mathbb{R}_+$, then*

$$Q[r_t^{n,m} \leq s_t^{n,m} \textit{ for all } t \geq 0] = 1 \ . \tag{B.52}$$

Proof. For simplicity, we drop the dependence on n, m in $(r_t^{n,m})$ and $(s_t^{n,m})$. Denote by ρ the Lipschitz constant of θ^n. Let $(a_k)_{k\in\mathbb{N}}$ be a decreasing sequence, $1 > a_1 > a_2 > \ldots > a_k > \ldots > 0$, such that $\int_{a_1}^1 \rho^{-2}x^{-1}\mathrm{d}x = 1$, $\int_{a_2}^{a_1} \rho^{-2}x^{-1}\mathrm{d}x = 2, \ldots, \int_{a_k}^{a_{k-1}} \rho^{-2}x^{-1}\mathrm{d}x = k$. We choose a sequence $\Psi_k(u)$, $k = 1, 2, \ldots$, of continuous functions such that its support is contained in (a_k, a_{k-1}), $\int_{a_k}^{a_{k-1}} \Psi_k(u)\mathrm{d}u = 1$ and $0 \leq \Psi_k(u) \leq 2/k \cdot \rho^{-2}u^{-2}$. Such a function exists. We set

$$\varphi_k(x) = \begin{cases} \int_0^x \mathrm{d}y \int_0^y \Psi_k(u)\mathrm{d}u & \text{if } x \geq 0, \\ 0 & \text{if } x < 0 \ . \end{cases}$$

Note that for any $k \in \mathbb{N}$, $\varphi_k \in \mathcal{C}^2(\mathbb{R}_+)$, $|\varphi_k'(x)| \leq 1$, $\varphi_k(x) \to x_+$ as $k \uparrow \infty$ and $\varphi_k'(x) \uparrow \mathbb{1}_{(0,\infty)}(x)$. Applying Ito's formula to $\varphi_k(r_t - s_t)$, we obtain

$$\varphi_k(r_t - s_t) = \varphi_k(r_0 - s_0) + I_1(k) + I_2(k) + I_3(k) \;,$$

where

$$\begin{aligned} I_1(k) &= \int_0^t \varphi_k'(r_u - s_u)[\theta^n(r_u) - \theta^n(s_u)]\mathrm{d}B_u \;, \\ I_2(k) &= \int_0^t \varphi_k'(r_u - s_u)[b(r_u) - \hat{b}(s_u) + P_u(g^m) - \hat{P}_u(h^m)]\mathrm{d}u \;, \\ I_3(k) &= \frac{1}{2}\int_0^t \varphi_k''(r_u - s_u)[\theta^n(r_u) - \theta^n(s_u)]^2\mathrm{d}u \;, \end{aligned}$$

with $P_u = Q \circ r_u^{-1}$ and $\hat{P}_u = Q \circ s_u^{-1}$. It holds by boundedness and Lipschitz continuity of θ^n

$$\mathbb{E}[I_1(k)] = 0 \;, \quad \text{ and } \quad \mathbb{E}[I_3(k)] \leq \frac{1}{2}\mathbb{E}\Big[\int_0^t \varphi_k''(r_u - s_u)\rho^2|r_u - s_u|^2\mathrm{d}u\Big] \leq \frac{t}{k} \;.$$

We note that by Assumption B.8 $\mathbb{E}[(g^m(r_u) - h^m(s_u))\mathbb{1}_{r_u - s_u < 0}] \leq 0$ and

$$\begin{aligned} \mathbb{E}[(g^m(r_u) - h^m(s_u))\mathbb{1}_{r_u - s_u \geq 0}] &\leq \mathbb{E}[(g^m(r_u) - g^m(s_u) + g^m(s_u) - h^m(s_u))\mathbb{1}_{r_u - s_u \geq 0}] \\ &\leq \mathbb{E}[(g^m(r_u) - g^m(s_u))\mathbb{1}_{r_u - s_u \geq 0}] \\ &\leq K_m\mathbb{E}[|r_u - s_u|\mathbb{1}_{r_u - s_u \geq 0}] \end{aligned} \tag{B.53}$$

by Lipschitz continuity of g^m, by $g^m(r) \leq h^m(r)$ and since g^m and h^m are non-decreasing. Hence for I_2, we obtain

$$\begin{aligned} I_2(k) &= \int_0^t \varphi_k'(r_u - s_u)[\tilde{b}(r_u) - \hat{b}(r_u) + \hat{b}(r_u) - \hat{b}(s_u)]\mathrm{d}u \\ &+ \int_0^t \varphi_k'(r_u - s_u)\Big(\mathbb{E}[(g^m(r_u) - h^m(s_u))\mathbb{1}_{r_u - s_u \geq 0}] + \mathbb{E}[(g^m(r_u) - h^m(s_u))\mathbb{1}_{r_u - s_u < 0}]\Big)\mathrm{d}u \\ &\leq \int_0^t \varphi_k'(r_u - s_u)\tilde{L}|r_u - s_u|\mathrm{d}u + \int_0^t \varphi_k'(r_u - s_u)K_m\mathbb{E}[|r_u - s_u|\mathbb{1}_{r_u - s_u \geq 0}]\mathrm{d}u \;. \end{aligned}$$

Taking the limit $k \to \infty$ and using that $\mathbb{E}[r_0 - s_0] = 0$, we obtain

$$\mathbb{E}[(r_t - s_t)_+] \leq \tilde{L}\mathbb{E}\Big[\int_0^t (r_u - s_u)_+\mathrm{d}u\Big] + K_m\mathbb{E}\Big[\int_0^t \mathbb{1}_{(0,\infty)}(r_u - s_u)\mathbb{E}[(r_u - s_u)_+]\mathrm{d}u\Big] \;, \tag{B.54}$$

by the monotone convergence theorem and since $(\varphi_k')_{k\in\mathbb{N}}$ is a monotone increasing sequence which converges pointwise to $\mathbb{1}_{(0,\infty)}(x)$. Assume there exists $t^* = \inf\{t \geq 0 : \mathbb{E}[(r_t - s_t)_+] > 0\} < \infty$. Then, $\int_0^{t^*} \mathbb{E}[(r_u - s_u)_+]\mathrm{d}u > 0$ or $\int_0^{t^*} \mathbb{E}[\mathbb{1}_{(0,\infty)}(r_u - s_u)]\mathbb{E}[(r_u - s_u)_+]\mathrm{d}u > 0$. By definition of t^*, $\mathbb{E}[(r_u - s_u)_+] = 0$ for all $u < t^*$ and hence both terms are zero. This contradicts the definition of t^*. Hence, (B.52) holds. □

Next, we show that the distribution of the solution of (B.50) converges as $n \to \infty$.

Lemma B.15. *Assume that $\tilde{b}$, $\hat{b}$, g and h satisfy Assumption B.4 and Assumption B.5. Let $\eta \in \Gamma(\mu, \nu)$ where the probability measures μ and ν on $\mathbb{R}_+$ satisfy Assumption B.6. Assume that $(g^m)_{m\in\mathbb{N}}$, $(h^m)_{m\in\mathbb{N}}$, $(\theta^n)_{n\in\mathbb{N}}$, $(\mu_{n,m})_{m,n\in\mathbb{N}}$, $(\nu_{n,m})_{m,n\in\mathbb{N}}$ and $(\eta_{n,m})_{m,n\in\mathbb{N}}$ satisfy condition*

Assumption B.8, Assumption B.10 and Assumption B.9. Then for any $m \in \mathbb{N}$, there exists a random variable (r^m, s^m) defined on some probability space $(\Omega^m, \mathcal{A}^m, P^m)$ with values in $\mathbb{W}\times\mathbb{W}$, such that $(r_t^m, s_t^m)_{t\geq 0}$ is a weak solution of the stochastic differential equation (B.51). *More precisely, for all $m \in \mathbb{N}$ the sequence of laws $Q\circ(r^{n,m}, s^{n,m})^{-1}$ converges weakly to the distribution $P^m \circ (r^m, s^m)^{-1}$. If additionally,*

$$\begin{aligned} &\tilde{b}(r) \leq \hat{b}(r) \quad \text{and} \quad g^m(r) \leq h^m(r)\ , && \text{for any } r \in \mathbb{R}_+ \text{ and} \\ &Q[r_0^{n,m} \leq s_0^{n,m}] = 1 && \text{for any } n, m \in \mathbb{N}, \end{aligned}$$

then $P^m[r_t^m \leq s_t^m$ for all $t \geq 0] = 1$.

Proof. Fix $m \in \mathbb{N}$. The proof is divided in three parts. First we show tightness of the sequences of probability measures. Then we identify the limit of the sequence of stochastic processes. Finally, we compare the two limiting processes.

Tightness: We show that the sequence of probability measures $(\mathbb{P}^{n,m})_{n\in\mathbb{N}}$ on $(\mathbb{W}\times\mathbb{W}, \mathcal{B}(\mathbb{W})\otimes\mathcal{B}(\mathbb{W}))$ is tight by applying Kolmogorov's continuity theorem. Consider $p > 2$ such that the p-th moment in Assumption B.6 and Assumption B.10 are uniformly bounded. Fix $T > 0$. Then the p-th moment of $r_t^{n,m}$ for $t < T$ can be bounded using Ito's formula,

$$\begin{aligned} \mathrm{d}|r_t^{n,m}|^p &\leq p|r_t^{n,m}|^{p-2}\langle r_t^{n,m}, (\tilde{b}(r_t^{n,m}) + P_t^{n,m}(g^m))\rangle \mathrm{d}t + 2\theta^n(r_t^{n,m})p|r_t^{n,m}|^{p-2}r_t^{n,m}\mathrm{d}W_t \\ &\quad + p(p-1)|r_t^{n,m}|^{p-2}2\theta^n(r_t^n)^2\mathrm{d}t \\ &\leq p\Big(|r_t^{n,m}|^p\tilde{L} + \Gamma|r_t^{n,m}|^{p-1} + 2(p-1)|r_t^{n,m}|^{p-2}\Big)\mathrm{d}t + 2\theta^n(r_t^{n,m})p(r_t^{n,m})^{p-1}\mathrm{d}W_t \\ &\leq p\Big(\tilde{L} + \Gamma + 2(p-1)\Big)|r_t^{n,m}|^p\mathrm{d}t + p(\Gamma + 2(p-1))\mathrm{d}t + 2\theta^n(r_t^{n,m})p(r_t^{n,m})^{p-1}\mathrm{d}W_t\ , \end{aligned}$$

where $\Gamma = \max(\|g\|_\infty, \|h\|_\infty)$. Taking expectation yields

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[|r_t^{n,m}|^p] \leq p\Big(\tilde{L} + \Gamma + 2(p-1)\Big)\mathbb{E}|r_t^{n,m}|^p + p(\Gamma + 2(p-1))\ .$$

Then by Gronwall's lemma

$$\sup_{t\in[0,T]} \mathbb{E}[|r_t^{n,m}|^p] \leq \mathrm{e}^{p(\tilde{L}+\Gamma+2(p-1))T}(\mathbb{E}[|r_0^{n,m}|^p] + Tp(\Gamma + 2(p-1))) < C_p < \infty\ , \tag{B.55}$$

where C_p depends on T and the p-th moment of the initial distribution, which is finite by Assumption B.9. Similarly, it holds $\sup_{t\in[0,T]}\mathbb{E}[|s_t^{n,m}|^p] < C_p$ for $t \leq T$. Using this moment bound, it holds for all $t_1, t_2 \in [0, T]$ by Assumption B.4, Assumption B.8 and Assumption B.9,

$$\begin{aligned} \mathbb{E}[|r_{t_2}^{n,m} - r_{t_1}^{n,m}|^p] &\leq C_1(p)\Big(\mathbb{E}[|\int_{t_1}^{t_2}\tilde{b}(r_u^{n,m}) + P_u^{n,m}(g^m)\mathrm{d}u|^p] + \mathbb{E}[|\int_{t_1}^{t_2}2\theta^n(r_u^{n,m})\mathrm{d}W_u|^p]\Big) \\ &\leq C_2(p)\Big(\Big(\mathbb{E}\Big[\frac{\tilde{L}^p}{|t_2-t_1|}\int_{t_1}^{t_2}|r_u^{n,m}|^p\mathrm{d}u\Big] + \Gamma^p\Big)|t_2-t_1|^p + \mathbb{E}[|\int_{t_1}^{t_2}2\theta^n(r_u^{n,m})\mathrm{d}u|^{p/2}]\Big) \\ &\leq C_2(p)\Big(\Big(\frac{\tilde{L}^p}{|t_2-t_1|}\int_{t_1}^{t_2}\mathbb{E}[|r_u^{n,m}|^p]\mathrm{d}u + \Gamma^p\Big)|t_2-t_1|^p + 2^{p/2}|t_2-t_1|^{p/2}\Big) \\ &\leq C_3(p, T, \tilde{L}, \Gamma, C_p)|t_2-t_1|^{p/2}\ , \end{aligned}$$

where $C_i(\cdot)$ are constants depending on the stated argument and which are independent of n, m. Note that in the second step, we used Burkholder-Davis-Gundy inequality, see [170, Chapter IV, Theorem 48]. It holds similarly, $\mathbb{E}[|s_{t_2}^{n,m} - s_{t_1}^{n,m}|^p] \leq C_3(p, T, \tilde{L}, \Gamma, C_p)|t_2-t_1|^{p/2}$. Hence,

$$\mathbb{E}[|(r_{t_2}^{n,m}, s_{t_2}^{n,m}) - (r_{t_1}^{n,m}, s_{t_1}^{n,m})|^p] \leq C_4(p, T, \tilde{L}, \Gamma, C_p)|t_2-t_1|^{p/2} \tag{B.56}$$

for all $t_1, t_2 \in [0,T]$. Hence, by Kolmogorov's continuity criterion, cf. [121, Corollary 14.9], there exists a constant $\tilde{C}$ depending on p and γ such that

$$\mathbb{E}\Big[[(r^{n,m}, s^{n,m})]_\gamma^p\Big] \leq \tilde{C} \cdot C_4(p, T, \tilde{L}, \Gamma, C_p) \ , \tag{B.57}$$

where $[\cdot]_\gamma^p$ is given by $[x]_\gamma = \sup_{t_1,t_2\in[0,T]} \frac{|x(t_1)-x(t_2)|}{|t_1-t_2|^\gamma}$ and $(r_t^{n,m}, s_t^{n,m})_{n\in\mathbb{N},t\geq 0}$ is tight in $\mathcal{C}([0,T],\mathbb{R}^2)$. Hence, for each $T > 0$ there exists a subsequence $n_k \to \infty$ and a probability measure $\mathbb{P}_T^m$ on $\mathcal{C}([0,T],\mathbb{R}^2)$. Since $\{\mathbb{P}_T^m\}_T$ is a consistent family, there exists by [121, Theorem 5.16] a probability measure $\mathbb{P}^m$ on $(\mathbb{W}\times\mathbb{W}, \mathcal{B}(\mathbb{W})\otimes\mathcal{B}(\mathbb{W}))$ such that there is a subsequence $(n_k)_{k\in\mathbb{N}}$ such that $\mathbb{P}^{n_k,m}$ converges along this subsequence to $\mathbb{P}^m$. Note that here we can take by a diagonalization argument the same subsequence $(n_k)_{k\in\mathbb{N}}$ for all m.

Characterization of the limit measure: In the following we drop for simplicity the index k in the subsequence. Denote by $(\mathbf{r}_t, \mathbf{s}_t)(\omega) = \omega(t)$ the canonical process on $\mathbb{W}\times\mathbb{W}$. Since $\mathbb{P}^{n,m}\circ(\mathbf{r}_0,\mathbf{s}_0)^{-1} = \eta_{n,m}$ converges weakly to η_m by Assumption B.10, it holds $\mathbb{P}^m\circ(\mathbf{r}_0,\mathbf{s}_0)^{-1} = \eta_m$. We define the maps $M^{n,m}, N^{n,m} : \mathbb{W}\times\mathbb{W}\to\mathbb{W}$ by

$$M_t^{n,m} = \mathbf{r}_t - \mathbf{r}_0 - \int_0^t (\tilde{b}(\mathbf{r}_u) + P_u^n(g^m))\mathrm{d}u \text{ and } N_t^{n,m} = \mathbf{s}_t - \mathbf{s}_0 - \int_0^t (\hat{b}(\mathbf{s}_u) + \hat{P}_u^n(h^m))\mathrm{d}u \ ,$$

where $P_u^n = \mathbb{P}^{n,m}\circ(\mathbf{r}_u)^{-1}$ and $\hat{P}_u^n = \mathbb{P}^{n,m}\circ(\mathbf{s}_u)^{-1}$. For each $m,n\in\mathbb{N}$, $(M_t^{n,m}, \mathcal{F}_t, \mathbb{P}^{n,m})$ and $(N_t^{n,m}, \mathcal{F}_t, \mathbb{P}^{n,m})$ are martingales with respect to the canonical filtration $\mathcal{F}_t = \sigma((\mathbf{r}_u,\mathbf{s}_u)_{0\leq u\leq t})$ by Ito's formula and the moment estimate (B.55). Further the family $(M_t^{n,m}, \mathbb{P}^{n,m})_{n\in\mathbb{N},t\geq0}$ and $(N_t^{n,m}, \mathbb{P}^{n,m})_{n\in\mathbb{N},t\geq0}$ are uniformly integrable by Lipschitz continuity of $\tilde{b}$ and $\hat{b}$ and by boundedness of g^m and h^m. Further, the mappings $M^{n,m}$ and $N^{n,m}$ are continuous in $\mathbb{W}$. We show that $\mathbb{P}^{n,m}\circ(\mathbf{r},\mathbf{s},M^{n,m},N^{n,m})^{-1}$ converges weakly to $\mathbb{P}^m\circ(\mathbf{r},\mathbf{s},M^m,N^m)^{-1}$ as $n\to\infty$, where

$$M_t^m = \mathbf{r}_t - \mathbf{r}_0 - \int_0^t (\tilde{b}(\mathbf{r}_u) + P_u(g^m))\mathrm{d}u \quad \text{and} \quad N_t^m = \mathbf{s}_t - \mathbf{s}_0 - \int_0^t (\hat{b}(\mathbf{s}_u) + \hat{P}_u(h^m))\mathrm{d}u \ , \tag{B.58}$$

with $P_u = \mathbb{P}^m\circ\mathbf{r}_u^{-1}$ and $\hat{P}_u = \mathbb{P}^m\circ\mathbf{s}_u^{-1}$. To show weak convergence to $\mathbb{P}^m\circ(\mathbf{r},\mathbf{s},M^m,N^m)^{-1}$, we note that (M^m,N^m) is continuous in $\mathbb{W}$ and we consider for a Lipschitz continuous and bounded function $G:\mathbb{W}\to\mathbb{R}$,

$$\begin{aligned}
&\left|\int_{\mathbb{W}} G(\omega)\mathrm{d}\mathbb{P}^{n,m}\circ(M^{n,m})^{-1}(\omega) - \int_{\mathbb{W}} G(\omega)\mathrm{d}\mathbb{P}^{m}\circ(M^{m})^{-1}(\omega)\right| \\
&\leq \left|\int_{\mathbb{W}} G(\omega)\mathrm{d}\mathbb{P}^{n,m}\circ(M^{n,m})^{-1}(\omega) - \int_{\mathbb{W}} G(\omega)\mathrm{d}\mathbb{P}^{n,m}\circ(M^{m})^{-1}(\omega)\right| \\
&+ \left|\int_{\mathbb{W}} G(\omega)\mathrm{d}\mathbb{P}^{n,m}\circ(M^{m})^{-1}(\omega) - \int_{\mathbb{W}} G(\omega)\mathrm{d}\mathbb{P}^{m}\circ(M^{m})^{-1}(\omega)\right| \ .
\end{aligned}$$

The second term converges to 0 as $n\to\infty$, since (M^m) is continuous. For the first term it holds

$$\begin{aligned}
&\left|\int_{\mathbb{W}} G(\omega)\mathrm{d}\mathbb{P}^{n,m}\circ(M^{n,m})^{-1}(\omega) - \int_{\mathbb{W}} G(\omega)\mathrm{d}\mathbb{P}^{n,m}\circ(M^{m})^{-1}(\omega)\right| \\
&= \left|\int_{\mathbb{W}} (G\circ M^{n,m})(\omega)\mathrm{d}\mathbb{P}^{n,m}(\omega) - \int_{\mathbb{W}} (G\circ M^{m})(\omega)\mathrm{d}\mathbb{P}^{n,m}(\omega)\right| \\
&\leq \|G\|_{\mathrm{Lip}} \sup_{\omega\in\mathbb{W}} d_{\mathbb{W}}(M^{n,m}(\omega), M^m(\omega)) \ ,
\end{aligned}$$

where $d_{\mathbb{W}}(f,g) = \sum_{k=1}^{\infty} \sup_{t\in[0,k]} 2^{-k}|f(t)-g(t)|$. This term converges to 0 for $n \to \infty$, since for all $T > 0$ and $\omega \in \mathbb{W}$, for $n \to \infty$

$$\sup_{t\in[0,T]} |M_t^{n,m}(\omega) - M_t^m(\omega)| \leq \int_0^T \left|(\mathbb{P}^{n,m} \circ \mathbf{r}_s^{-1})(g^m) - (\mathbb{P}^m \circ \mathbf{r}_s^{-1})(g^m)\right| \mathrm{d}s \to 0 \ ,$$

by Lebesgue dominated convergence theorem, since g is bounded. Hence,

$$\left|\int_{\mathbb{W}} G(\omega)\mathrm{d}\mathbb{P}^{n,m} \circ (M^{n,m})^{-1}(\omega) - \int_{\mathbb{W}} G(\omega)\mathrm{d}\mathbb{P}^m \circ (M^m)^{-1}(\omega)\right| \to 0 \quad \text{for } n \to \infty,$$

and similarly for $(N^{n,m})$, and therefore by the Portmanteau theorem [123, Theorem 13.16], weak convergence of $\mathbb{P}^{n,m} \circ (\mathbf{r}, \mathbf{s}, M^{n,m}, N^{n,m})^{-1}$ to $\mathbb{P}^m \circ (\mathbf{r}, \mathbf{s}, M^m, N^m)^{-1}$ holds.

Let $G : \mathbb{W} \to \mathbb{R}_+$ be a $\mathcal{F}_s$-measurable, bounded, non-negative function. By uniformly integrability of $(M_t^{n,m}, \mathbb{P}^{n,m})_{n\in\mathbb{N},t\geq 0}$, for any $s \leq t$,

$$\begin{aligned}
\mathbb{E}^m[G(M_t^m - M_s^m)] &= \mathbb{E}^m[G(\int_s^t (\tilde{b}(\mathbf{r}_u) + P_u(g^m))\mathrm{d}u)] \\
&= \lim_{n\to\infty} \mathbb{E}^{n,m}[G(\int_s^t (\tilde{b}(\mathbf{r}_u) + P_u^n(g^m))\mathrm{d}u)] \\
&= \lim_{n\to\infty} \mathbb{E}^{n,m}[G(M_t^{n,m} - M_s^{n,m})] = 0 \ ,
\end{aligned} \tag{B.59}$$

and analogously for $(N_t^{n,m})_{t\geq 0}$ and hence, $(M_t^m, \mathcal{F}_t, \mathbb{P}^m)$ and $(N_t^m, \mathcal{F}_t, \mathbb{P}^m)$ are continuous martingales. The quadratic variation $([(M^m, N^m)]_t)$ exists $\mathbb{P}^m$-almost surely. To complete the identification of the limit, it suffices to note that the quadratic variation is given by

$$\begin{aligned}
[M^m] &= 4\int_0^{\cdot} \mathbb{1}_{(0,\infty)}(\mathbf{r}_u)\mathrm{d}u && \mathbb{P}^m\text{-almost surely,} \\
[N^m] &= 4\int_0^{\cdot} \mathbb{1}_{(0,\infty)}(\mathbf{s}_u)\mathrm{d}u && \mathbb{P}^m\text{-almost surely, and} \\
[M^m, N^m] &= 4\int_0^{\cdot} \mathbb{1}_{(0,\infty)}(\mathbf{r}_u)\mathbb{1}_{(0,\infty)}(\mathbf{s}_u)\mathrm{d}u && \mathbb{P}^m\text{-almost surely,}
\end{aligned} \tag{B.60}$$

which holds following the computations in the proof of [87, Theorem 22]. We show that $((M_t^m)^2 - 4\int_0^t \mathbb{1}_{(0,\infty)}\mathbf{r}_u\mathrm{d}u)$ is a sub- and a supermartingale and hence a martingale using a monotone class argument by noting first that for any bounded continuous and non-negative function $G : \mathbb{W} \to \mathbb{R}_+$,

$$\mathbb{E}^m[G(M_t^m)^2] = \lim_{n\to\infty} \mathbb{E}^{n,m}[G(M_t^{n,m})^2] \tag{B.61}$$

holds using uniform integrability of $((M_t^{n,m})^2, \mathbb{P}^{n,m})_{n\in\mathbb{N},t\geq 0}$ which holds similarly as above. Note that

$$\mathbb{E}^m\left[G\int_s^t \mathbb{1}_{(0,\infty)}(\mathbf{r}_u)\mathrm{d}u\right] \leq \lim_{\epsilon\downarrow 0} \liminf_{n\to\infty} \mathbb{E}^{n,m}\left[G\int_s^t \mathbb{1}_{(\epsilon,\infty)}(\mathbf{r}_u)\mathrm{d}u\right] \tag{B.62}$$

holds by lower semicontinuity of $\omega \to \int_0^{\cdot} \mathbb{1}_{(\epsilon,\infty)}(\omega_s)\mathrm{d}s$ for each $\epsilon > 0$, Fatou's lemma and the Portmanteau theorem. For any fixed $\epsilon > 0$,

$$\liminf_{n\to\infty} \mathbb{E}^{n,m}\left[G\left(\int_s^t \theta^n(\mathbf{r}_u)^2\mathrm{d}u - \int_s^t \mathbb{1}_{(\epsilon,\infty)}(\mathbf{r}_u)\mathrm{d}u\right)\right] . \tag{B.63}$$

Then by (B.61), (B.62) and (B.63)

$$\begin{aligned}&\mathbb{E}^m\left[G\left((M_t^m)^2-(M_s^m)^2-4\int_s^t \mathbb{1}_{(0,\infty)}(\mathbf{r}_u)\mathrm{d}u\right)\right]\\&\quad\geq \lim_{\epsilon\downarrow 0}\liminf_{n\to\infty}\mathbb{E}^{n,m}\left[G\left((M_t^{n,m})^2-(M_s^{n,m})^2-4\int_s^t \theta^n(\mathbf{r}_u)^2\mathrm{d}u\right)\right]=0\end{aligned}$$

and by a monotone class argument, cf. [170, Chapter 1, Theorem 8], $((M_t^m)^2-4\int_0^t \mathbb{1}_{(0,\infty)}(\mathbf{r}_u)\mathrm{d}u,\mathbb{P}^m)$ is a submartingale. To show that it is also a supermartingale we note that $((M_t^m)^2-4t,\mathbb{P}^m)$ is a supermartingale by (B.61). By the uniqueness of the Doob-Meyer decomposition, cf. [170, Chapter 3, Theorem 8], $t\to[M^m]_t-4t$ is $\mathbb{P}^m$-almost surely decreasing. Note further, that $(\mathbf{r}_t,\mathcal{F}_t,\mathbb{P}^m)$ is a continuous semimartingale with $[\mathbf{r}]=[M^m]$. Then by Ito-Tanaka formula, cf. [172, Chapter 6, Theorem 1.1],

$$\int_0^t \mathbb{1}_{\{0\}}(\mathbf{r}_u)\mathrm{d}[M^m]_u=\int_0^t \mathbb{1}_{\{0\}}(\mathbf{r}_u)\mathrm{d}[\mathbf{r}]_u=\int_0^t \mathbb{1}_{\{0\}}(y)\ell_t^y(\mathbf{r})\mathrm{d}y=0\ ,$$

where $\ell_t^y(\mathbf{r})$ is the local time of $\mathbf{r}$ in y. Therefore, for any $0\leq s<t$,

$$[M^m]_t-[M^m]_s=\int_0^t \mathbb{1}_{(0,\infty)}(\mathbf{r}_u)\mathrm{d}[M^m]_u\leq 4\int_0^t \mathbb{1}_{(0,\infty)}(\mathbf{r}_u)\mathrm{d}u$$

and hence, for any $\mathcal{F}_s$-measurable, bounded, non-negative function $G:\mathbb{W}\to\mathbb{R}_+$,

$$\mathbb{E}^m\left[G((M_t^m)^2-(M_s^m)^2-4\int_s^t \mathbb{1}_{(0,\infty)}(\mathbf{r}_u)\mathrm{d}u)\right]\leq 0\ .$$

As before, by a monotone class argument, $((M_t^m)^2-4\int_0^t \mathbb{1}_{(0,\infty)}(\mathbf{r}_u)\mathrm{d}u,\mathbb{P}^m)$ is a supermartingale, and hence a martingale.

Hence, we obtain the quadratic variation $[M^m]_t$ given in (B.60). The other characterizations in (B.60) follow by analogous arguments. Then by a martingale representation theorem, see [116, Chapter II, Theorem 7.1], we conclude, that there are a probability space $(\Omega^m,\mathcal{A}^m,P^m)$ and a Brownian motion motion W and random variables (r^m,s^m) on this space such that $P^m\circ(r^m,s^m)^{-1}=\mathbb{P}^m\circ(\mathbf{r}^m,\mathbf{s}^m)^{-1}$ and such that (r^m,s^m,W) is a weak solution of (B.51). Finally, note that we have weak convergence of $Q\circ(r^{n,m},s^{n,m})^{-1}$ to $P^m\circ(r^m,s^m)^{-1}$ not only along a subsequence since the characterization of the limit holds for any subsequence $(n_k)_{k\in\mathbb{N}}$.

Comparison of two solutions: To show $P^m[r_t^m\leq s_t^m$ for all $t\geq 0]=1$ we note that by Lemma B.14, $Q[r_t^n\leq s_t^n$ for all $t\geq 0]=1$. The monotonicity carries over to the limit by the Portmanteau theorem for closed sets, since we have weak convergence of $\mathbb{P}^{n,m}\circ(\mathbf{r},\mathbf{s})^{-1}$ to $\mathbb{P}^m\circ(\mathbf{r},\mathbf{s})^{-1}$. □

We show in the next step that the distribution of the solution of (B.51) converges as $m\to\infty$. For each $m\in\mathbb{N}$ let $(\Omega^m,\mathcal{A}^m,P^m)$ be a probability space and random variables $r^m,s^m:\Omega^m\to\mathbb{W}$ such that $(r_t^m,s_t^m)_{t\geq 0}$ is a solution of (B.51). Let $\mathbb{P}^m=P^m\circ(r^m,s^m)^{-1}$ denote the law on $\mathbb{W}\times\mathbb{W}$.

Lemma B.16. *Assume that $(\tilde{b},g)$ and $(\hat{b},h)$ satisfy Assumption B.4 and Assumption B.5. Let $\eta\in\Gamma(\mu,\nu)$ where the probability measures μ and ν on $\mathbb{R}_+$ satisfy Assumption B.6. Assume that $(g^m)_{m\in\mathbb{N}}$, $(h^m)_{m\in\mathbb{N}}$, $(\mu_m)_{m\in\mathbb{N}}$, $(\nu_m)_{m\in\mathbb{N}}$ and $(\eta_m)_{m\in\mathbb{N}}$ satisfy conditions Assumption B.8 and Assumption B.10. Then there exists a random variable (r,s) defined on some probability space $(\Omega,\mathcal{A},P)$ with values in $\mathbb{W}\times\mathbb{W}$, such that $(r_t,s_t)_{t\geq 0}$ is a weak solution of the sticky stochastic* *differential equation* (B.25)*. Furthermore, the sequence of laws* $P^m \circ (r^m, s^m)^{-1}$ *converges weakly to the law* $P \circ (r, s)^{-1}$*. If additionally,*

$$\begin{aligned} &\tilde{b}(r) \le \hat{b}(r)\ , \qquad g(r) \le h(r) \quad \text{and} \qquad g^m(r) \le h^m(r) && \text{for any } r \in \mathbb{R}_+, \text{ and} \\ &P^m[r_0^m \le s_0^m] = 1 && \text{for any } m \in \mathbb{N} \end{aligned}$$

then $P[r_t \le s_t \text{ for all } t \ge 0] = 1$*.*

Proof. The proof is structured as the proof of Lemma B.15. First analogously to the proof of (B.55) we show under Assumption B.4, Assumption B.8 and Assumption B.10,

$$\sup_{t\in[0,T]} \mathbb{E}[|r_t^m|^p] < \infty\ . \tag{B.64}$$

Tightness of the sequence of probability measures $(\mathbb{P}^m)_{m\in\mathbb{N}}$ on $(\mathbb{W}\times\mathbb{W}, \mathcal{B}(\mathbb{W})\otimes\mathcal{B}(\mathbb{W}))$ holds adapting the steps of the proof of Lemma B.15 to (B.51). Note that (B.55) and (B.56) hold analogously for $(r_t^m, s_t^m)_{m\in\mathbb{N}}$ by Assumption B.4, Assumption B.8 and Assumption B.10. Hence by Kolmogorov's continuity criterion, cf. [121, Corollary 14.9], we can deduce that there exists a probability measure $\mathbb{P}$ on $(\mathbb{W}\times\mathbb{W}, \mathcal{B}(\mathbb{W})\otimes\mathcal{B}(\mathbb{W}))$ such that there is a subsequence $(m_k)_{k\in\mathbb{N}}$ along which $\mathbb{P}^{m_k}$ converge towards $\mathbb{P}$. To characterize the limit, we first note that by Skorokhod representation theorem, cf. [20, Chapter 1, Theorem 6.7], without loss of generality we can assume that (r^m, s^m) are defined on a common probability space $(\Omega, \mathcal{A}, P)$ with expectation E and converge almost surely to (r, s) with distribution $\mathbb{P}$. By Assumption B.8, $P_t^m(g^m) = E[g^m(r_t^m)]$ and the monotone convergence theorem, $P_t^m(g^m)$ converges to $P_t(g)$ for any $t \ge 0$. Then, by Lebesgue convergence theorem it holds almost surely for all $t \ge 0$

$$\lim_{m\to\infty} \int_0^t \Big(\tilde{b}(r_t^m) + P_u^m(g^m)\Big)\mathrm{d}u = \int_0^t \Big(\tilde{b}(r_t) + P_u(g)\Big)\mathrm{d}u\ , \tag{B.65}$$

where $P_u^m = P\circ(r_u^m)^{-1}$ and $P_u = P\circ(r_u)^{-1}$. A similar statement holds for $(s_t)_{t\ge0}$.

Consider the mappings $M^m, N^m : \mathbb{W}\times\mathbb{W} \to \mathbb{W}$ given by (B.58). Then for all $m \in \mathbb{N}$, $(M_t^m, \mathcal{F}_t, \mathbb{P}^m)$ and $(N_t^m, \mathcal{F}_t, \mathbb{P}^m)$ are martingales with respect to the canonical filtration $\mathcal{F}_t = \sigma((\mathbf{r}_u, \mathbf{s}_u)_{0\le u\le t})$. Further the family $(M_t^m, \mathbb{P}^m)_{m\in\mathbb{N},t\ge0}$ and $(N_t^m, \mathbb{P}^m)_{m\in\mathbb{N},t\ge0}$ are uniformly integrable by (B.64). In the same line as in the proof of Lemma B.15 and by (B.65), $\mathbb{P}^m\circ(\mathbf{r}, \mathbf{s}, M^m, N^m)$ converges weakly to $\mathbb{P}\circ(\mathbf{r}, \mathbf{s}, M, N)$ where

$$M_t = \mathbf{r}_t - \mathbf{r}_0 - \int_0^t (\tilde{b}(\mathbf{r}_u) + P_u(g))\mathrm{d}u \qquad \text{and} \qquad N_t = \mathbf{s}_t - \mathbf{s}_0 - \int_0^t (\hat{b}(\mathbf{s}_u) + \hat{P}_u(h))\mathrm{d}u\ .$$

Let $G : \mathbb{W} \to \mathbb{R}_+$ be a $\mathcal{F}_s$-measurable bounded, non-negative function. By uniform integrability, for any $s \le t$,

$$\begin{aligned} \mathbb{E}[G(M_t - M_s)] &= \mathbb{E}[G(\int_s^t (\tilde{b}(\mathbf{r}_u) + P_u(g))\mathrm{d}u)] = \lim_{m\to\infty} \mathbb{E}^m[G(\int_s^t (\tilde{b}(\mathbf{r}_u) + P_u(g^m))\mathrm{d}u)] \\ &= \lim_{m\to\infty} \mathbb{E}^m[G(M_t^m - M_s^m)] = 0\ , \end{aligned}$$

and analogously for $(N_t)_{t\ge0}$. Hence, $(M_t, \mathcal{F}_t, \mathbb{P})$ and $(N_t, \mathcal{F}_t, \mathbb{P})$ are martingales. Further, the quadratic variation $([(M, N)]_t)$ exists $\mathbb{P}$-almost surely and is given by (B.60) $\mathbb{P}$-almost surely, which holds following the computations in the proof of Lemma B.15. As in Lemma B.15, we conclude by a martingale representation theorem that there are a probability space $(\Omega, \mathcal{A}, P)$ and a Brownian motion W and random variables (r, s) on this space such that $P\circ(r, s)^{-1} = \mathbb{P}\circ(\mathbf{r}, \mathbf{s})^{-1}$

and such that (r, s, W) is a weak solution of (B.25). Note that the limit identification holds for all subsequences $(m_k)_{k\in\mathbb{N}}$ and hence $P^m \circ (r^m, s^m)^{-1}$ converges weakly to $P \circ (r, s)^{-1}$ for $m \to \infty$. The monotonicity $P^m[r_t^m \le s_t^m$ for all $t \ge 0] = 1$ carries over to the limit by Portmanteau theorem, since $\mathbb{P}^m \circ (\mathbf{r}, \mathbf{s})^{-1}$ converges weakly to $\mathbb{P} \circ (\mathbf{r}, \mathbf{s})^{-1}$. □

Proof of Theorem B.3. The proof is a direct consequence of Lemma B.15 and Lemma B.16. □

Proof of Theorem B.5

Proof of Theorem B.5. Note that the Dirac at 0, δ_0, is by definition an invariant measure of $(r_t)_{t\ge 0}$ solving (B.6). Assume that the process starts from an invariant probability measure π, hence $\mathbb{P}(r_t > 0) = p = \pi((0,\infty))$ for any $t \ge 0$. Note that for $p = 0$ the drift vanishes. If the initial measure is the Dirac measure in 0, δ_0, then the diffusion coefficient disappears. Hence, $\mathrm{Law}(r_t) = \delta_0$ for any $t \ge 0$. It remains to investigate the case $p \ne 0$. Here, we are in the regime of [87, Lemma 24] where an invariant measure is of the form (B.28). Since $p = \mathbb{P}(r_t > 0)$, the invariant measure π satisfies additionally the necessary condition

$$p = \pi((0,\infty)) = \frac{I(a,p)}{2/(ap) + I(a,p)} \tag{B.66}$$

with $I(a,p)$ given in (B.27). For $p \ne 0$, this expression is equivalent to (B.26). □

Proof of Proposition B.6. By Theorem B.5, it suffices to study the solutions of (B.26). By (B.27) and since $\tilde{b}(r) = -\tilde{L}r$, it holds for $\hat{I}(a,p) = (1-p)I(a,p)$,

$$\hat{I}(a,p) = \Big(\sqrt{\frac{\pi}{2}} + \int_0^{\frac{ap}{\sqrt{2\tilde{L}}}} \exp(-x^2/2)\mathrm{d}x\Big)\sqrt{\frac{2}{\tilde{L}}}\exp\Big(\frac{a^2p^2}{4\tilde{L}}\Big)(1-p)\ . \tag{B.67}$$

In the case $a/\sqrt{\tilde{L}} \le 2/\sqrt{\pi}$, $\hat{I}(a,0) = \sqrt{\pi/\tilde{L}}$ by (B.67). Further, by $1+x \le \mathrm{e}^x$ and $a/\sqrt{\tilde{L}} \le 2/\sqrt{\pi}$,

$$\begin{aligned}\Big(\sqrt{\frac{\pi}{2}} + \int_0^{\frac{ap}{\sqrt{2\tilde{L}}}} \mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x\Big)(1-p)\mathrm{e}^{\frac{a^2p^2}{4\tilde{L}}} &\le \sqrt{\frac{\pi}{2}}\Big(1 + \sqrt{\frac{2}{\pi}}\int_0^{\frac{ap}{\sqrt{2L}}} \mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x\Big)\mathrm{e}^{-p}\mathrm{e}^{\frac{p^2}{\pi}} \\ &\le \sqrt{\frac{\pi}{2}}\Big(1 + \frac{2p}{\pi}\Big)\mathrm{e}^{-p}\mathrm{e}^{\frac{p^2}{\pi}} \le \sqrt{\frac{\pi}{2}}\mathrm{e}^{p(\frac{3}{\pi}-1)} < \sqrt{\frac{\pi}{2}}\end{aligned}$$

for $p \in (0,1]$. Hence, $\hat{I}(a,p) < \hat{I}(a,0)$ by (B.67). Therefore, $\hat{I}(a,p) < \hat{I}(a,0) \le \frac{2}{a}$ for all $p \in (0,1]$ and so δ_0 is the unique invariant probability measure for $a/\sqrt{\tilde{L}} \le 2/\sqrt{\pi}$.

To show that for $a/\sqrt{\tilde{L}} > 2/\sqrt{\pi}$, there exists a unique p solving (B.26), we note that $\hat{I}(a,p)$ is continuous with $\hat{I}(a,0) > 2/a$ and $\hat{I}(a,1) = 0$. By the mean value theorem, there exists at least one $p \in (0,1)$ satisfying (B.26). In the following we drop the dependence on a in $I(a,p)$ and $\hat{I}(a,p)$. We show uniqueness of the solution p by contradiction. Assume that $p_1 < p_2$ are the two smallest solutions of (B.26). Hence, it holds either $\hat{I}'(p_1) < 0$ or $\hat{I}'(p) = 0$ for p_1. Note that the derivative is given by

$$\begin{aligned}\hat{I}'(p_i) &= -I(p_i) + (1-p_i)I'(p_i) = -I(p_i) + (1-p_i)\Big(p_i\frac{a^2}{2\tilde{L}}I(p_i) + \frac{a}{\tilde{L}}\Big) \\ &= -\frac{2}{a(1-p_i)} + (1-p_i)\frac{a}{\tilde{L}}\Big(\frac{p_i}{1-p_i} + 1\Big) = -\frac{2}{a(1-p_i)} + \frac{a}{\tilde{L}}\ .\end{aligned} \tag{B.68}$$

Then, for $p_2 > p_1$, it holds

$$\hat{I}'(p_2) = -\frac{2}{a(1-p_2)} + \frac{a}{\tilde{L}} < -\frac{2}{a(1-p_1)} + \frac{a}{\tilde{L}} = \hat{I}'(p_1) \le 0 \,.$$

If $\hat{I}'(p_1) < 0$, it holds $\hat{I}'(p_2) < 0$ which contradicts that p_1 and p_2 are the two smallest solutions. In the second case, when $\hat{I}'(p_1) = 0$, we note that the second derivative of $\hat{I}(p)$ at p_1 is given by

$$\begin{aligned}\hat{I}''(p_1) &= -2I'(p_1) + (1-p_1)I''(p_1)\\ &= \Big(-2 + (1-p_1)\frac{a^2 p_1}{2\tilde{L}}\Big)\Big(I(p_1)\frac{a^2 p_1}{2\tilde{L}} + \frac{a}{\tilde{L}}\Big) + (1-p_1)I(p_1)\frac{a^2}{2\tilde{L}}\\ &= \Big(-2 + (1-p_1)\frac{a^2 p_1}{2\tilde{L}}\Big)\frac{a}{\tilde{L}(1-p_1)} + \frac{a}{\tilde{L}} = -\frac{a}{\tilde{L}(1-p_1)} < 0 \,.\end{aligned}$$

Hence, in this case there is a maximum at p_1, which contradicts that p_1 is the smallest solution. Thus, there exists a unique solution p_1 of (B.26) for $a/\sqrt{\tilde{L}} > 2/\sqrt{\pi}$.

□

Proof of Theorem B.7

Proof of Theorem B.7. To show (B.31) we extend the function f to a concave function on $\mathbb{R}$ by setting $f(x) = x$ for $x < 0$. Note that f is continuously differentiable and f' is absolutely continuous and bounded. Using Ito-Tanaka formula, c.f. [172, Chapter 6, Theorem 1.1] we obtain

$$\mathrm{d}f(r_t) = f'(r_t)(\tilde{b}(r_t) + a\mathbb{P}(r_t > 0))\mathrm{d}t + 2f''(r_t)\mathbb{1}_{(0,\infty)}(r_t)\mathrm{d}t + \mathrm{d}M_t \,,$$

where $M_t = 2\int_0^t f'(r_s)1_{(0,\infty)}(r_s)\mathrm{d}B_s$ is a martingale. Taking expectation, we get

$$\begin{aligned}\frac{d}{dt}\mathbb{E}[f(r_t)] &= \mathbb{E}[f'(r_t)(\tilde{b}(r_t) + a\mathbb{P}(r_t > 0))] + 2\mathbb{E}[f''(r_t)\mathbb{1}_{(0,\infty)}(r_t)]\\ &= \mathbb{E}[f'(r_t)\tilde{b}(r_t) + 2(f''(r_t) - f''(0))] + \mathbb{E}[af'(r_t) + 2f''(0)]\mathbb{P}(r_t > 0)\\ &\le -c\mathbb{E}[f(r_t)] \,,\end{aligned}$$

where the last step holds by (B.39) and (B.40). By applying Gronwall's lemma, we obtain (B.31).

□

B.6.4 Proof of Appendix B.4

To show Theorem B.8, we first give a uniform in time bound for the second moment of the process $(\bar{X}_t)_{t\ge0}$ solving (B.1).

Lemma B.17. *Let $(\bar{X}_t)_{t\ge0}$ be a solution of* (B.1) *with $\mathbb{E}[|\bar{X}_0|^2] < \infty$. Assume Assumption B.1. Then there exists $C \in (0,\infty)$ depending on d, W and the second moment of $\bar{X}_0$ such that*

$$C = \sup_{t\ge0}\mathbb{E}[|\bar{X}_t|^2] < \infty \,. \tag{B.69}$$

Proof of Lemma B.17. By Ito's formula, it holds

$$\frac{1}{2}\mathrm{d}|\bar{X}_t|^2 = \langle \bar{X}_t, b * \bar{\mu}_t(\bar{X}_t)\rangle \mathrm{d}t + \bar{X}_t^T \mathrm{d}B_t + \frac{1}{2}d \ \mathrm{d}t \ .$$

Taking expectation and using symmetry, we get

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[|\bar{X}_t|^2] &= \mathbb{E}[\langle \bar{X}_t - \tilde{X}_t, b(\bar{X}_t - \tilde{X}_t)\rangle] + d \\ &= -\mathbb{E}[\langle \bar{X}_t - \tilde{X}_t, L(\bar{X}_t - \tilde{X}_t) - \gamma(\bar{X}_t - \tilde{X}_t)\rangle \mathbb{1}_{|\bar{X}_t - \tilde{X}_t| > R_0}] \\ &\quad - \mathbb{E}[\langle \bar{X}_t - \tilde{X}_t, L(\bar{X}_t - \tilde{X}_t) - \gamma(\bar{X}_t - \tilde{X}_t)\rangle \mathbb{1}_{|\bar{X}_t - \tilde{X}_t| \le R_0}] + d \\ &\le \mathbb{E}[|\bar{X}_t|^2(-2L + \kappa(|\bar{X}_t - \tilde{X}_t|)\mathbb{1}_{|\bar{X}_t - \tilde{X}_t| > R_0})] + \|\gamma\|_\infty R_0 + d \ .\end{aligned}$$

Hence by definition (B.14) of R_0 and by Gronwall's lemma we obtain the result (B.69). □

Let $N \in \mathbb{N}$. We construct a sticky coupling of N i.i.d. realizations of solutions $(\{\bar{X}_t^i\}_{i=1}^N)_{t\ge 0}$ to (B.1) and of the solution $(\{Y_t^i\}_{i=1}^N)_{t\ge 0}$ to the mean field particle system (B.3). Then, we consider a weak limit for $\delta \to 0$ of Markovian couplings which are constructed similar as in Appendix B.2. Let rc^δ, sc^δ satisfy (B.19) and (B.20). The coupling $(\{\bar{X}_t^{i,\delta}, Y^{i,\delta}\}_{i=1}^N)_{t\ge 0}$ is defined as process in $\mathbb{R}^{2Nd}$ satisfying a system of SDEs given by

$$\begin{aligned}\mathrm{d}\bar{X}_t^{i,\delta} &= b * \bar{\mu}_t^\delta(\bar{X}_t^{i,\delta})\mathrm{d}t + \mathrm{rc}^\delta(\tilde{r}_t^{i,\delta})\mathrm{d}B_t^{i,1} + \mathrm{sc}^\delta(\tilde{r}_t^{i,\delta})\mathrm{d}B_t^{i,2} \\ \mathrm{d}Y_t^{i,\delta} &= \frac{1}{N}\sum_{j=1}^N b(Y_t^{i,\delta} - Y_t^{j,\delta})\mathrm{d}t + \mathrm{rc}^\delta(\tilde{r}_t^{i,\delta})(\mathrm{Id} - 2\tilde{e}_t^{i,\delta}(\tilde{e}_t^{i,\delta})^T)\mathrm{d}B_t^{i,1} + \mathrm{sc}^\delta(\tilde{r}_t^{i,\delta})\mathrm{d}B_t^{i,2} \ ,\end{aligned} \tag{B.70}$$

where $\mathrm{Law}(\{\bar{X}_0^{i,\delta}, Y_0^{i,0}\}_{i=1}^N) = \bar{\mu}_0^{\otimes N} \otimes \nu_0^{\otimes N}$, and where $(\{B_t^{i,1}\}_{i=1}^N)_{t\ge 0}, (\{B_t^{i,2}\}_{i=1}^N)_{t\ge 0}$ are i.i.d. d-dimensional standard Brownian motions. We set $\tilde{X}_t^{i,\delta} = \bar{X}_t^{i,\delta} - \frac{1}{N}\sum_{j=1}^N \bar{X}_t^{j,\delta}$, $\tilde{Y}_t^{i,\delta} = Y_t^{i,\delta} - \frac{1}{N}\sum_{j=1}^N Y_t^{j,\delta}$, $\tilde{Z}_t^{i,\delta} = \tilde{X}_t^{i,\delta} - \tilde{Y}_t^{i,\delta}$, $\tilde{r}_t^{i,\delta} = |\tilde{Z}_t^{i,\delta}|$ and $\tilde{e}_t^{i,\delta} = \tilde{Z}_t^{i,\delta}/\tilde{r}_t^{i,\delta}$ for $\tilde{r}_t^{i,\delta} \neq 0$. The value $\tilde{e}_t^{i,\delta}$ for $\tilde{r}_t^{i,\delta} = 0$ is irrelevant as $\mathrm{rc}^{i,\delta}(0) = 0$. By Levy's characterization $(\{\bar{X}_t^{i,\delta}, Y_t^{i,\delta}\}_{i=1}^N)_{t\ge 0}$ is indeed a coupling of (B.1) and (B.3). Existence and uniqueness of the coupling given in (B.70) hold by [146, Theorem 2.2]. In the next step we analyse $\tilde{r}_t^{i,\delta}$.

Lemma B.18. *Assume Assumption B.1 holds. Then, for $\epsilon < \epsilon_0$, where ϵ_0 is given in* (B.20), *and for any $i \in \{1, \dots, N\}$, it holds almost surely,*

$$\begin{aligned}\mathrm{d}\tilde{r}_t^{i,\delta} &= -L\tilde{r}_t^{i,\delta}\mathrm{d}t + \langle \tilde{e}_t^{i,\delta}, \frac{1}{N}\sum_{j=1}^N \gamma(\tilde{X}_t^{i,\delta} - \tilde{X}_t^{j,\delta}) - \gamma(\tilde{Y}_t^{i,\delta} - \tilde{Y}_t^{j,\delta})\rangle \mathrm{d}t \\ &\quad + 2\sqrt{1 + \frac{1}{N}}\mathrm{rc}^\delta(\tilde{r}_t^{i,\delta})\mathrm{d}W_t^{i,\delta} + \Big\langle \tilde{e}_t^{i,\delta}, \Theta_t^{i,\delta} + \frac{1}{N}\sum_{k=1}^N \Theta_t^{k,\delta}\Big\rangle \mathrm{d}t \\ &\le \Big(\bar{b}(\tilde{r}_t^{i,\delta}) + 2\|\gamma\|_\infty \frac{1}{N}\sum_{j=1}^N \mathrm{rc}^\epsilon(\tilde{r}_t^{j,\delta})\Big)\mathrm{d}t + 2\sqrt{1 + \frac{1}{N}}\mathrm{rc}^\delta(\tilde{r}_t^{i,\delta})\mathrm{d}W_t^{i,\delta} \\ &\quad + \Big(A_t^{i,\delta} + \frac{1}{N}\sum_{k=1}^N A_t^{k,\delta}\Big)\mathrm{d}t \ .\end{aligned} \tag{B.71}$$

with $\Theta_t^{i,\delta} = b * \bar{\mu}_t^\delta(\bar{X}_t^{i,\delta}) - \frac{1}{N}\sum_{j=1}^N b(\bar{X}_t^{i,\delta} - \bar{X}_t^{j,\delta})$ *and*

$$A_t^{i,\delta} = \Big|\Theta_t^{i,\delta}\Big| = \Big|b * \bar{\mu}_t^\delta(\bar{X}_t^{i,\delta}) - \frac{1}{N}\sum_{j=1}^N b(\bar{X}_t^{i,\delta} - \bar{X}_t^{j,\delta})\Big| \tag{B.72}$$

and where $(\{W_t^{i,\delta}\}_{i=1}^N)_{t\geq 0}$ *are* N *one-dimensional Brownian motions given by*

$$W_t^{i,\delta} = \sqrt{\frac{N}{N+1}}\left(\int_0^t (\tilde{e}_s^{i,\delta})^T \mathrm{d}B_s^{i,1} + \frac{1}{N}\sum_{j=1}^N \int_0^t (\tilde{e}_s^{j,\delta})^T \mathrm{d}B_s^{j,1}\right), \quad i = 1,\dots,N. \tag{B.73}$$

Proof. By (B.70) and since γ is anti-symmetric, it holds by Ito's formula for any $i \in \{1,\dots,N\}$,

$$\begin{aligned}
\mathrm{d}(\tilde{r}_t^{i,\delta})^2 = &-2L(\tilde{r}_t^{i,\delta})^2\mathrm{d}t + 2\langle \tilde{Z}_t^{i,\delta}, \frac{1}{N}\sum_{j=1}^N \gamma(\tilde{X}_t^{i,\delta} - \tilde{X}_t^{j,\delta}) - \gamma(\tilde{Y}_t^{i,\delta} - \tilde{Y}_t^{j,\delta})\rangle \mathrm{d}t \\
&+ 4\Big(1+\frac{1}{N}\Big)\mathrm{rc}^\delta(\tilde{r}_t^{i,\delta})^2\mathrm{d}t + 4\sqrt{1+\frac{1}{N}}\mathrm{rc}^\delta(\tilde{r}_t^{i,\delta})\langle \tilde{Z}_t^{i,\delta}, \tilde{e}_t^{i,\delta}\rangle \mathrm{d}W_t^{i,\delta} \\
&+ 2\langle \tilde{Z}_t^{i,\delta}, b * \bar{\mu}_t^\delta(\bar{X}_t^{i,\delta}) - \frac{1}{N}\sum_{j=1}^N b(\bar{X}_t^{i,\delta} - \bar{X}_t^{j,\delta})\rangle \mathrm{d}t \\
&+ 2\langle \tilde{Z}_t^{i,\delta}, -\frac{1}{N}\sum_{k=1}^N \Big(b * \bar{\mu}_t^\delta(\bar{X}_t^{k,\delta}) - \frac{1}{N}\sum_{j=1}^N b(\bar{X}_t^{k,\delta} - \bar{X}_t^{j,\delta})\Big)\rangle \mathrm{d}t \ .
\end{aligned}$$

where $(\{W_t^i\}_{i=1}^N)_{t\geq 0}$ are N i.i.d.one-dimensional Brownian motions given by (B.73). Note that the prefactor $(N/(N+1))^{1/2}$ ensures that the quadratic variation satisfies $[W^i]_t = t$ for $t \geq 0$, and hence $(\{W_t^i\}_{i=1}^N)_{t\geq 0}$ are Brownian motions. This definition of $(\{W_t^i\}_{i=1}^N)_{t\geq 0}$ leads to $(1+1/N)^{1/2}$ in the diffusion term of the SDE. Applying the $\mathcal{C}^2$ approximation of the square root used in the proof of Lemma B.12 and taking $\varepsilon \to 0$ in the approximation yields the stochastic differential equations of $(\{\tilde{r}_t^{i,\delta}\}_{i=1}^N)_{t\geq 0}$. We obtain its upper bound for $\epsilon < \epsilon_0$ by Assumption B.1 and (B.20) similarly to the proof of Lemma B.12. □

Next, we state a bound for (B.72).

Lemma B.19. *Under the same assumption as in Lemma B.20, it holds for any* $i = 1,\dots,N$

$$E\Big[|A_t^{i,\delta}|^2\Big] \leq C_1 N^{-1} \text{ and } E\Big[A_t^{i,\delta}\Big] \leq C_2 N^{-1/2} \ ,$$

where $A_t^{i,\delta}$ *is given in* (B.72) *and* C_1 *and* C_2 *are constants depending on* $\|\gamma\|_\infty$*,* L *and* C *given in Lemma B.17.*

Proof. By Assumption B.3, it holds $\mathbb{E}[|\bar{X}_0^{i,\delta}|^2] < \infty$ for $i = 1,\dots,N$. Note that given $\bar{X}_t^{i,\delta}$, $\bar{X}_t^{j,\delta}$ are i.i.d.with law $\bar{\mu}_t^\delta$ for all $j \neq i$. Hence,

$$\mathbb{E}[b(\bar{X}_t^{i,\delta} - \bar{X}_t^{j,\delta})|\bar{X}_t^{i,\delta}] = b * \bar{\mu}_t^\delta(\bar{X}_t^{i,\delta}) \ .$$

Since γ is anti-symmetric, $b(0) = 0$, and we have

$$\begin{aligned}
&\mathbb{E}\Big[|b * \bar{\mu}_t^\delta(\bar{X}_t^{i,\delta}) - \frac{1}{N-1}\sum_{j=1}^N b(\bar{X}_t^{i,\delta} - \bar{X}_t^{j,\delta})|^2\Big|\bar{X}_t^{i,\delta}\Big] \\
&= \mathbb{E}\Big[|\frac{1}{N-1}\sum_{j=1}^N \mathbb{E}[b(\bar{X}_t^{i,\delta} - \bar{X}_t^{j,\delta})|\bar{X}_t^{i,\delta}] - \frac{1}{N-1}\sum_{j=1}^N b(\bar{X}_t^{i,\delta} - \bar{X}_t^{j,\delta})|^2\Big|\bar{X}_t^{i,\delta}\Big] \\
&= \frac{1}{N-1}\mathrm{Var}_{\bar{\mu}_t^\delta}(b(\bar{X}_t^{i,\delta} - \cdot)) \ .
\end{aligned}$$

By (B.11), Assumption B.1, Assumption B.3 and Lemma B.17, we obtain

$$\begin{aligned}\mathrm{Var}_{\bar{\mu}_t^\delta}(b(\bar{X}_t^{i,\delta}-\cdot)) &= \int_{\mathbb{R}^d}\Big|\Big(-L(\bar{X}_t^{i,\delta}-x)+\int_{\mathbb{R}^d}L(\bar{X}_t^{i,\delta}-\tilde{x})\bar{\mu}_t^\delta(\mathrm{d}\tilde{x})\Big) \\ &\quad+\Big(\gamma(\bar{X}_t^{i,\delta}-x)-\int_{\mathbb{R}^d}\gamma(\tilde{X}_t^{i,\delta}-\tilde{x})\bar{\mu}_t^\delta(\mathrm{d}\tilde{x})\Big)\Big|^2\bar{\mu}_t^\delta(\mathrm{d}x) \\ &= \int_{\mathbb{R}^d}\Big|Lx+\Big(\gamma(\bar{X}_t^{i,\delta}-x)-\int_{\mathbb{R}^d}\gamma(\tilde{X}_t^{i,\delta}-\tilde{x})\bar{\mu}_t^\delta(\mathrm{d}\tilde{x})\Big)\Big|^2\bar{\mu}_t^\delta(\mathrm{d}x) \\ &\le 2L^2\int_{\mathbb{R}^d}|x|^2\bar{\mu}_t^\delta(\mathrm{d}x)+8\|\gamma\|_\infty^2 \le 2L^2C^2+8\|\gamma\|_\infty^2\ .\end{aligned}$$

By the Cauchy-Schwarz inequality, we have

$$\begin{aligned}\mathbb{E}[(A_t^{i,\delta})^2] &\le 2\mathbb{E}\Big[|b*\bar{\mu}_t(\bar{X}_t^{i,\delta})-\frac{1}{N-1}\sum_{j=1}^N b(\bar{X}_t^{i,\delta}-\bar{X}_t^{j,\delta})|^2\Big] \\ &\quad+2\Big(\frac{1}{N-1}-\frac{1}{N}\Big)^2\mathbb{E}\Big[|\sum_{j=1}^N b(\bar{X}_t^{i,\delta}-\bar{X}_t^{j,\delta})|^2\Big] \\ &\le 2\frac{1}{N-1}\mathbb{E}[\mathrm{Var}_{\bar{\mu}_t^\delta}(b(\bar{X}_t^{i,\delta}-\cdot))]+\frac{1}{N^2(N-1)}\mathbb{E}\Big[\sum_{j=1}^N|b(X_t^{i,\delta}-X_t^{j,\delta})|^2\Big] \\ &\le \frac{4L^2}{N-1}C+\frac{16\|\gamma\|_\infty^2}{N-1}+\frac{1}{N^2}\Big(8CL^2+4\|\gamma\|_\infty^2\Big) \\ &\le N^{-1}C_1<\infty\ ,\end{aligned}$$

where C_1 depends on $\|\gamma\|_\infty$, L and the second moment bound C. Similarly, it holds

$$\begin{aligned}\mathbb{E}[A_t^{i,\delta}] &\le \mathbb{E}\Big[|b*\bar{\mu}_t^\delta(\bar{X}_t^{i,\delta})-\frac{1}{N-1}\sum_{j=1}^N b(\bar{X}_t^{i,\delta}-\bar{X}_t^{j,\delta})|\Big] \\ &\quad+\Big(\frac{1}{N-1}-\frac{1}{N}\Big)\sum_{j=1}^N\mathbb{E}\Big[|b(\bar{X}_t^{i,\delta}-\bar{X}_t^{j,\delta})|\Big] \\ &\le \frac{\sqrt{2}L}{\sqrt{N-1}}C^{1/2}+\frac{\sqrt{8}\|\gamma\|_\infty}{\sqrt{N-1}}+\frac{1}{N}\Big(\sqrt{2}C^{1/2}L+\|\gamma\|_\infty\Big) \\ &\le N^{-1/2}C_2<\infty\ ,\end{aligned}$$

where $C_2=2LC^{1/2}+4\|\gamma\|_\infty+(\sqrt{2}C^{1/2}+\|\gamma\|_\infty)$. □

To control $(\{\tilde{r}_t^{i,\delta}\}_{i=1}^N)_{t\ge0}$, we consider $(\{r_t^{i,\delta,\epsilon}\}_{i=1}^N)_{t\ge0}$ given as solution of

$$\begin{aligned}\mathrm{d}r_t^{i,\delta,\epsilon} &= \bar{b}(r_t^{i,\delta,\epsilon})\mathrm{d}t+\frac{1}{N}\sum_{j=1}^N 2\|\gamma\|_\infty\mathrm{rc}^\epsilon(r_t^{j,\delta,\epsilon})\mathrm{d}t+\Big(A_t^{i,\delta}+\frac{1}{N}\sum_{k=1}^N A_t^{k,\delta}\Big)\mathrm{d}t \\ &\quad+2\sqrt{1+\frac{1}{N}}\mathrm{rc}^\delta(r_t^{i,\delta,\epsilon})\mathrm{d}W_t^{i,\delta}\end{aligned} \tag{B.74}$$

with initial condition $r_0^{i,\delta,\epsilon}=\tilde{r}_0^{i,\delta}$ for all $i=1,\dots,N$, $A_t^{i,\delta}$ given in (B.72) and $W_t^{i,\delta}$ given in (B.73).

By [146, Theorem 2.2], under Assumption B.1 and Assumption B.3, $(\{U_t^{i,\delta,\epsilon}\}_{i=1}^N)_{t\geq 0} = (\{\bar{X}_t^{i,\delta}, Y_t^{i,\delta}, r_t^{i,\delta,\epsilon}\}_{i=1}^N)_{t\geq 0}$ exists and is unique, where $(\{\bar{X}_t^{i,\delta}, \bar{Y}_t^{i,\delta}\}_{i=1}^N)_{t\geq 0}$ solves uniquely (B.70), $(\{\bar{r}_t^{i\delta}\}_{i=1}^N)_{t\geq 0}$ and $(\{r_t^{i,\bar{\delta},\epsilon}\}_{i=1}N)_{t\geq 0}$ solve uniquely (B.71) and (B.74), respectively, with $(\{W_t^{i,\delta}\}_{i=1}^N)_{t\geq 0}$ given by (B.73).

Lemma B.20. *Assume Assumption B.1 and Assumption B.3. Then for any $i = 1, \dots, N$, $|\bar{X}_t^{i,\delta} - Y_t^{i,\delta} - \frac{1}{N}\sum_j(\bar{X}_t^{j,\delta} - Y_t^{j,\delta})| = \tilde{r}_t^{i,\delta} \leq r_t^{i,\delta,\epsilon}$, almost surely for all $t \geq 0$ and $\epsilon < \epsilon_0$.*

Proof. Note, that both processes $(\{\tilde{r}_t^{i,\delta}\}_{i=1}^N)_{t\geq 0}$ and $(\{r_t^{i,\delta,\epsilon}\}_{i=1}^N)_{t\geq 0}$ have the same initial condition and are driven by the same noise. Since the drift for $(\{r_t^{i,\delta,\epsilon}\}_{i=1}^N)_{t\geq 0}$ is larger than the drift for $(\{\tilde{r}_t^{i,\delta}\}_{i=1}^N)_{t\geq 0}$ for $\epsilon < \epsilon_0$ by (B.20), we can conclude $\tilde{r}_t^{i,\delta} \leq r_t^{i,\delta,\epsilon}$ almost surely for all $t \geq 0$, $\epsilon < \epsilon_0$ and $i = 1, \dots N$ by Lemma B.21. □

Proof of Theorem B.8. Consider the process $(\{U_t^{i,\delta,\epsilon}\}_{i=1}^N)_{t\geq 0} = (\{\bar{X}_t^{i,\delta}, Y_t^{i,\delta}, r_t^{i,\delta,\epsilon}\}_{i=1}^N)_{t\geq 0}$ on $\mathbb{R}^{N(2d+1)}$ for each $\epsilon, \delta > 0$. We denote by $\mathbb{P}^{\delta,\epsilon}$ the law of $\{U^{\bar{\delta},\epsilon}\}_{i=1}^N$ on $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^{N(2d+1)})$. We define the canonical projections $\mathbf{X}, \mathbf{Y}, \mathbf{r}$ onto the first Nd, second Nd and last N components.

By Assumption B.1 and Assumption B.3 it holds in the same line as in the proof of Lemma B.22 for each $T > 0$

$$E[|\{U_{t_2}^{i,\delta,\epsilon} - U_{t_1}^{i,\delta,\epsilon}\}_{i=1}^N|^4] \leq C|t_2 - t_1|^2 \qquad \text{for } t_1, t_2 \in [0, T], \tag{B.75}$$

for some constant C depending on T, L, $\|\gamma\|_{\mathrm{Lip}}$, $\|\gamma\|_\infty$, N and on the fourth moment of μ_0 and ν_0. Note that we used here that the additional drift terms $(A_t^{i,\delta})_{t\geq 0}$ occurring in the SDE of $(\{r_t^{i,\delta,\epsilon}\}_{i=1}^N)_{t\geq 0}$ are Lipschitz continuous in $(\{\bar{X}_t^{i,\delta}\}_{i=1}^N)_{t\geq 0}$. Then as in the proofs of Lemma B.22 and Lemma B.23, $\mathbb{P}^{\delta,\epsilon}$ is tight and converges weakly along a subsequence to a measure $\mathbb{P}$ by Kolmogorov's continuity criterion, cf. [121, Corollary 14.9].

As in Lemma B.22 the law $\mathbb{P}_T^{\delta,\epsilon}$ of $(\{U_t^{i,\delta,\epsilon}\}_{i=1}^N)_{0\leq t\leq T}$ on $\mathcal{C}([0,T], \mathbb{R}^{N(2d+1)})$ is tight for each $T > 0$ by [121, Corollary 14.9] and for each $\epsilon > 0$ there exists a subsequence $\delta_n \to 0$ such that $(\mathbb{P}_T^{\delta_n,\epsilon})_{n\in\mathbb{N}}$ on $\mathcal{C}([0,T], \mathbb{R}^{N(2d+1)})$ converge to a measure $\mathbb{P}_T^\epsilon$ on $\mathcal{C}([0,T], \mathbb{R}^{N(2d+1)})$. By a diagonalization argument and since $\{\mathbb{P}_T^\epsilon : T \geq 0\}$ is a consistent family, cf. [121, Theorem 5.16], there exists a probability measure $\mathbb{P}^\epsilon$ on $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^{N(2d+1)})$ such that for all ϵ there exists a subsequence δ_n such that $(\mathbb{P}^{\delta_n,\epsilon})_{n\in\mathbb{N}}$ converges along this subsequence to $\mathbb{P}^\epsilon$. As in the proof of Lemma B.23 we repeat this argument for the family of measures $(\mathbb{P}^\epsilon)_{\epsilon>0}$. Hence, there exists a subsequence $\epsilon_m \to 0$ such that $(\mathbb{P}^{\epsilon_m})_{m\in\mathbb{N}}$ converges to a measure $\mathbb{P}$. Let $(\{\bar{X}_t^i, Y_t^i, r_t^i\}_{i=1}^N)_{t\geq 0}$ be some process on $\mathbb{R}^{N(2d+1)}$ with distribution $\mathbb{P}$ on $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P})$.

Since $(\{\bar{X}_t^{i,\delta}\}_{i=1}^N)_{t\geq 0}$ and $(\{Y_t^{i,\delta}\}_{i=1}^N)_{t\geq 0}$ are solutions that are unique in law, we have that for any $\delta, \epsilon > 0$, $\mathbb{P}^{\delta,\epsilon} \circ \mathbf{X}^{-1} = \mathbb{P} \circ \mathbf{X}^{-1}$ and $\mathbb{P}^{\delta,\epsilon} \circ \mathbf{Y}^{-1} = \mathbb{P} \circ \mathbf{Y}^{-1}$. Hence, $\mathbb{P} \circ (\mathbf{X}, \mathbf{Y})^{-1}$ is a coupling of (B.1) and (B.3).

Similarly to the proof of Lemma B.22 and Lemma B.23 there exist an extended underlying probability space and N i.i.d.one-dimensional Brownian motion $(\{W_t^i\}_{i=1}^N)_{t\geq 0}$ such that $(\{r_t^i, W_t^i\}_{i=1}^N)_{t\geq 0}$ is a solution of

$$\begin{aligned} \mathrm{d}r_t^i = \bar{b}(r_t^i)\mathrm{d}t + \frac{1}{N}\sum_{j=1}^N 2\|\gamma\|_\infty \mathbb{1}_{(0,\infty)}(r_t^j)\mathrm{d}t + \Big(A_t^i + \frac{1}{N}\sum_{k=1}^N A_t^k\Big)\mathrm{d}t \\ + 2\sqrt{1 + \frac{1}{N}}\mathbb{1}_{(0,\infty)}(r_t^i)\mathrm{d}W_t^i , \end{aligned}$$

where $A_t^i = |b * \bar{\mu}_t(\bar{X}_t^i) - \frac{1}{N}\sum_{j=1}^N b(\bar{X}_t^i - \bar{X}_t^j)|$.

In addition, the statement of Lemma B.20 carries over to the limiting process $(\{r_t^i\}_{i=1}^N)_{t\geq 0}$, since by the weak convergence along the subsequences $(\delta_n)_{n\in\mathbb{N}}$ and $(\epsilon_m)_{m\in\mathbb{N}}$ and the Portmanteau theorem, $P(|\tilde{X}_t^i - \tilde{Y}_t^i| \leq r_t^i \text{ for } i=1,\dots,N) \geq \limsup_{m\to\infty}\limsup_{n\to\infty} P(|\tilde{X}_t^{i,\delta_n} - \tilde{Y}_t^{i,\delta_n}| \leq r_t^{i,\delta_n,\epsilon_m} \text{ for } i=1,\dots,N) = 1$, where $\tilde{X}_t^i = \bar{X}_t^i - (1/N)\sum_{j=1}^N \bar{X}_t^j$ and $\tilde{Y}_t^i = \bar{X}_t^i - (1/N)\sum_{j=1}^N \bar{Y}_t^j$ for all $t\geq 0$ and $i=1,\dots,N$.

Using Ito-Tanaka formula, c.f. [172, Chapter 6, Theorem 1.1], and f' is absolutely continuous, we obtain for f defined in (B.37) with $\tilde{b}(r) = (\kappa(r)-L)r$ and $a = 2\|\gamma\|_\infty$,

$$\begin{aligned} \mathrm{d}\Big(\frac{1}{N}\sum_{i=1}^N f(r_t^i)\Big) &= \frac{1}{N}\sum_{i=1}^N \Big(\tilde{b}(r_t^i)f'(r_t^i) + f''(r_t^i)2\frac{N+1}{N}\mathbb{1}_{(0,\infty)}(r_t^i)\Big)\mathrm{d}t \\ &+ \frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N 2f'(r_t^i)\|\gamma\|_\infty \mathbb{1}_{(0,\infty)}(r_t^j)\mathrm{d}t \\ &+ \frac{1}{N}\sum_{i=1}^N f'(r_t^i)2\sqrt{1+\frac{1}{N}}\mathbb{1}_{(0,\infty)}(r_t^i)\mathrm{d}W_t^i + \frac{1}{N}\sum_{i=1}^N f'(r_t^i)\Big(A_t^i + \frac{1}{N}\sum_{k=1}^N A_t^k\Big)\mathrm{d}t\,. \end{aligned}$$

Taking expectation, we get using $f'(r)\leq 1$ for all $r\geq 0$,

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\Big[\frac{1}{N}\sum_{i=1}^N f(r_t^i)\Big] &\leq \frac{1}{N}\sum_{i=1}^N \Big\{\mathbb{E}\Big[\tilde{b}(r_t^i)f'(r_t^i) + 2\frac{N+1}{N}(f''(r_t^i) - f''(0))\Big] \\ &+ \mathbb{E}\Big[2\Big(\|\gamma\|_\infty + \frac{N+1}{N}f''(0)\Big)\mathbb{1}_{(0,\infty)}(r_t^i)\Big] + \mathbb{E}\Big[2A_t^i\Big]\Big\}\,. \end{aligned} \tag{B.76}$$

By (B.39) and (B.40), the first two terms are bounded by $-\tilde{c}\frac{1}{N}\sum_i f(r_t^i)$ with $\tilde{c}$ given in (B.17).

By Lemma B.19 the last term in (B.76) is bounded by

$$2E[A_t^i] \leq \tilde{C}N^{-1/2}\,,$$

where

$$\tilde{C} = 2C_2 = 4LC^{1/2} + 8\|\gamma\|_\infty + 2(\sqrt{2}C^{1/2}L + \|\gamma\|_\infty)\,. \tag{B.77}$$

Hence, we obtain

$$\frac{d}{dt}\mathbb{E}\Big[\frac{1}{N}\sum_i f(r_t^i)\Big] \leq -\tilde{c}\frac{1}{N}\sum_i \mathbb{E}[f(r_t^i)] + \tilde{C}N^{-1/2}$$

for $t\geq 0$ which leads by Grönwall's lemma to

$$\mathbb{E}\Big[\frac{1}{N}\sum_i f(r_t^i)\Big] \leq \mathrm{e}^{-\tilde{c}t}\mathbb{E}\Big[\frac{1}{N}\sum_i f(r_0^i)\Big] + \frac{1}{\tilde{c}}\tilde{C}N^{-1/2}\,.$$

For an arbitrary coupling $\xi\in\Gamma(\bar{\mu}_0^{\otimes N}, \nu_0^{\otimes N})$, we have

$$\mathcal{W}_{f,N}((\bar{\mu}_t)^{\otimes N}, \nu_t^N) \leq \mathrm{e}^{-\tilde{c}t}\int_{\mathbb{R}^{2Nd}} \frac{1}{N}\sum_{i=1}^N f\left(\left|x^i - y^i - \frac{1}{N}\sum_{j=1}^N (x^j - y^j)\right|\right)\xi(\mathrm{d}x\mathrm{d}y) + \frac{\tilde{C}}{\tilde{c}N^{1/2}}\,,$$

as $\mathbb{E}[f(r_0^i)] \leq \int_{\mathbb{R}^{2Nd}} \frac{1}{N}\sum_{i=1}^N f(|x^i - y^i - \frac{1}{N}\sum_{j=1}^N (x^j - y^j)|)\xi(\mathrm{d}x\mathrm{d}y)$. Taking the infimum over all couplings $\xi\in\Gamma(\bar{\mu}_0^{\otimes N}, \nu_0^{\otimes})$ gives the first bound. By (B.38), the second bound follows. □

B.6.5 Proof of Appendix B.5

As for the nonlinear case we show Theorem B.10 via a family of stochastic differential equations, with Lipschitz continuous coefficients,

$$\begin{aligned}
\mathrm{d}r_t^{i,n,m} &= \Big(\tilde{b}(r_t^{i,n,m}) + \frac{1}{N}\sum_{j=1}^N g^m(r_t^{j,n,m})\Big)\mathrm{d}t + 2\theta^n(r_t^{i,n,m})\mathrm{d}W_t^i \\
\mathrm{d}s_t^{i,n,m} &= \Big(\hat{b}(s_t^{i,n,m}) + \frac{1}{N}\sum_{j=1}^N h^m(s_t^{j,n,m})\Big)\mathrm{d}t + 2\theta^n(s_t^{i,n,m})\mathrm{d}W_t^i \\
\mathrm{Law}(r_0^{i,n,m}, s_0^{i,n,m}) &= \eta_{n,m}\ , \qquad i \in \{1,\dots,N\}\ ,
\end{aligned} \tag{B.78}$$

where $\eta_{n,m} \in \Gamma(\mu_{n,m}, \nu_{n,m})$. Under Assumption B.4, Assumption B.5, Assumption B.8, Assumption B.9 and Assumption B.10 we identify the weak limit of $(\{r_t^{i,n,m}, s_t^{i,n,m}\}_{i=1,n,m\in\mathbb{N}}^N)_{t\geq 0}$ solving (B.78) for $n\to\infty$ by $(\{r_t^{i,m}, s_t^{i,m}\}_{i=1,m\in\mathbb{N}}^N)_{t\geq 0}$ solving the family of SDEs given by

$$\begin{aligned}
\mathrm{d}r_t^{i,m} &= \Big(\tilde{b}(r_t^{i,m}) + \frac{1}{N}\sum_{j=1}^N g^m(r_t^{j,m})\Big)\mathrm{d}t + 2\mathbb{1}_{(0,\infty)}(r_t^{i,m})\mathrm{d}W_t^i\ , \\
\mathrm{d}s_t^{i,m} &= \Big(\hat{b}(s_t^{i,m}) + \frac{1}{N}\sum_{j=1}^N h^m(s_t^{j,m})\Big)\mathrm{d}t + 2\mathbb{1}_{(0,\infty)}(s_t^{i,m})\mathrm{d}W_t^i\ , \\
\mathrm{Law}(r_0^{i,m}, s_0^{i,m}) &= \eta_m\ , \qquad i \in \{1,\dots,N\}\ ,
\end{aligned} \tag{B.79}$$

where $\eta_m \in \Gamma(\mu_m, \nu_m)$.

Taking the limit $m\to\infty$, we obtain (B.35) as the weak limit of (B.79). In the case $g(r) = \mathbb{1}_{(0,\infty)}(r)$, we can choose $g^m = \theta^m$.

Consider a probability space $(\Omega_0, \mathcal{A}_0, Q)$ and N i.i.d.1-dimensional Brownian motions $(\{W_t^i\}_{i=1}^N)_{t\geq 0}$. Note that under Assumption B.4–Assumption B.10, there are random variables $\{r^{i,n,m}\}_{i=1}^N, \{s^{i,n,m}\}_{i=1}^N : \Omega_0 \to \mathbb{W}^N$ for each n, m such that $(\{r^{i,n,m}, s^{i,n,m}\}_{i=1}^N)$ is a unique solution to (B.78) by [146, Theorem 2.2]. We denote by $\mathbb{P}^{n,m} = Q \circ (\{r^{i,n,m}, s^{i,n,m}\}_{i=1}^N)^{-1}$ the law on $\mathbb{W}^N \times \mathbb{W}^N$.

Before taking the two limits and proving Theorem B.10, we introduce a modification of Ikeda and Watanabe's comparison theorem, to compare two solutions of (B.78), cf. [116, Section VI, Theorem 1.1].

Lemma B.21. *Suppose a solution $(\{r_t^{i,n,m}, s_t^{i,n,m}\}_{i=1}^N)_{t\geq 0}$ of (B.78) is given for fixed $n, m \in \mathbb{N}$. Assume Assumption B.8 for g^m and h^m, Assumption B.4 for $\tilde{b}$ and $\hat{b}$, Assumption B.9 for θ^n. If $Q[r_0^{i,n,m} \leq s_0^{i,n,m}$ for all $i = 1,\dots,N] = 1$, $\tilde{b}(r) \leq \hat{b}(r)$ and $g^m(r) \leq h^m(r)$ for any $r \in \mathbb{R}_+$, then*

$$Q[r_t^{i,n,m} \leq s_t^{i,n,m} \text{ for all } t \geq 0 \text{ and } i = 1,\dots,N] = 1$$

Proof. The proof is similar for each component $i = 1,\dots,N$ to the proof of Lemma B.14. It holds for the interaction part similarly to (B.53) using the properties of g^m and h^m,

$$\frac{1}{N}\sum_{j=1}^N (g^m(r_t^{j,n,m}) - h^m(s_t^{j,n,m})) \leq K_m \frac{1}{N}\sum_{j=1}^N |r_t^{j,n,m} - s_t^{j,n,m}| \mathbb{1}_{(0,\infty)}(r_t^{j,n,m} - s_t^{j,n,m})\ .$$

Hence, we obtain analogously to (B.54),

$$\mathbb{E}[(r_t^{i,n,m}-s_t^{i,n,m})_+]\leq \tilde{L}\mathbb{E}\Big[\int_0^t (r_u^{i,n,m}-s_u^{i,n,m})_+\mathrm{d}u\Big]+K_m\mathbb{E}\Big[\int_0^t \frac{1}{N}\sum_{j=1}^N (r_u^{j,n,m}-s_u^{j,n,m})_+\mathrm{d}u\Big]$$

for all $i=1,\dots,N$. Assume $t^*=\inf\{t\geq 0:\mathbb{E}[(r_t^{i,n,m}-s_t^{i,n,m})_+]>0 \text{ for some } i\}<\infty$. Then, there exists $i\in\{1,\dots,N\}$ such that $\int_0^{t^*}\mathbb{E}[(r_u^{i,n,m}-s_u^{i,n,m})_+]\mathrm{d}u>0$. But, by definition of t^*, for all i, $u<t^*$, $\mathbb{E}[(r_u^{i,n,m}-s_u^{i,n,m})_+]=0$. This contradicts the definition of t^*. Hence, $Q[r_t^{i,n,m}\leq s_t^{i,n,m}$ for all $i,\ t\geq 0]=1$. □

In the next step, we prove that the distribution of the solution of (B.78) converges as $n\to\infty$.

Lemma B.22. *Assume that Assumption B.4 and Assumption B.5 is satisfied for* $(\tilde{b},g)$ *and* $(\hat{b},h)$. *Further, let* $(\theta^n)_{n\in\mathbb{N}}$, $(g^m)_{m\in\mathbb{N}}$, $(h^m)_{m\in\mathbb{N}}$, $(\mu_{n,m})_{n,m\in\mathbb{N}}$, $(\nu_{n,m})_{n,m\in\mathbb{N}}$ *and* $(\eta_{n,m})_{n,m\in\mathbb{N}}$ *be such that Assumption B.8, Assumption B.9 and Assumption B.10 hold. Let* $m\in\mathbb{N}$. *Then there exists a random variable* $(\{r^{i,m},s^{i,m}\}_{i=1}^N)$ *defined on some probability space* $(\Omega^m,\mathcal{A}^m,P^m)$ *with values in* $\mathbb{W}^N\times\mathbb{W}^N$ *such that* $(\{r_t^{i,m},s_t^{i,m}\}_{i=1}^N)_{t\geq 0}$ *is a weak solution of* (B.79). *Moreover, the laws* $Q\circ(\{r^{i,n,m},s^{i,n,m}\}_{i=1}^N)^{-1}$ *converge weakly to* $P^m\circ(\{r^{i,m},s^{i,m}\}_{i=1}^N)^{-1}$. *If in addition,*

$$\begin{aligned}&\tilde{b}(r)\leq\hat{b}(r)\quad\textit{and}\quad g^m(r)\leq h^m(r) &&\textit{for any } r\in\mathbb{R}_+,\\ &Q[r_0^{i,n,m}\leq s_0^{i,n,m}]=1 &&\textit{for any } n\in\mathbb{N}, i=1,\dots,N,\end{aligned}$$

then $P^m[r_t^{i,m}\leq s_t^{i,m}$ *for all* $t\geq 0$ *and* $i\in\{1,\dots,N\}]=1$.

Proof. Fix $m\in\mathbb{N}$. The proof is divided in three parts and is similar to the proof of Lemma B.15. First we show tightness of the sequences of probability measures. Then we identify the limit of the sequence of stochastic processes. Finally, we compare the two limiting processes.

Tightness: We show analogously as in the proof of Lemma B.15 that the sequence of probability measures $(\mathbb{P}^{n,m})_{n\in\mathbb{N}}$ on $(\mathbb{W}^N\times\mathbb{W}^N,\mathcal{B}(\mathbb{W}^N)\otimes\mathcal{B}(\mathbb{W}^N))$ is tight by applying Kolmogorov's continuity theorem. We consider $p>2$ such that the p-th moment in Assumption B.10 are uniformly bounded. Fix $T>0$. Then the p-th moment of $r_t^{i,n,m}$ and $s_t^{i,n,m}$ for $t<T$ is bounded using Ito's formula,

$$\begin{aligned}\mathrm{d}|r_t^{i,n,m}|^p &\leq p|r_t^{i,n,m}|^{p-2}\langle r_t^{i,n,m},(\tilde{b}(r_t^{i,n,m})+\frac{1}{N}\sum_{j=1}^N g^m(r_t^{j,n,m}))\rangle\mathrm{d}t\\ &\quad+2\theta^n(r_t^{i,n,m})p|r_t^{n,m}|^{p-2}r_t^{i,n,m}\mathrm{d}W_t^i+p(p-1)|r_t^{i,n,m}|^{p-2}2\theta^n(r_t^{i,n,m})^2\mathrm{d}t\\ &\leq p\Big(|r_t^{i,n,m}|^p\tilde{L}+\Gamma|r_t^{i,n,m}|^{p-1}+2(p-1)|r_t^{i,n,m}|^{p-2}\Big)\mathrm{d}t+2\theta^n(r_t^{i,n,m})p(r_t^{i,n,m})^{p-1}\mathrm{d}W_t^i\\ &\leq p\Big(\tilde{L}+\Gamma+2(p-1)\Big)|r_t^{i,n,m}|^p\mathrm{d}t+p(\Gamma+2(p-1))\mathrm{d}t+2\theta^n(r_t^{i,n,m})p(r_t^{n,m})^{p-1}\mathrm{d}W_t^i\,.\end{aligned}$$

Taking expectation yields

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[|r_t^{i,n,m}|^p]\leq p\Big(L+\Gamma+2(p-1)\Big)\mathbb{E}|r_t^{i,n,m}|^p+p(\Gamma+2(p-1))\,.$$

Then by Gronwall's lemma

$$\sup_{t\in[0,T]}\mathbb{E}[|r_t^{i,n,m}|^p]\leq \mathrm{e}^{p(L+\Gamma+2(p-1))T}(\mathbb{E}[|r_0^{i,n,m}|^p]+Tp(\Gamma+2(p-1)))<C_p<\infty\,, \tag{B.80}$$

where C_p depends on T and the p-th moment of the initial distribution, which is by assumption finite. Similarly, it holds $\sup_{t\in[0,T]} \mathbb{E}[|s_t^{i,n,m}|^p] < C_p$ for $t \leq T$. Using these moment bounds, it holds for all $t_1, t_2 \in [0,T]$ by Assumption B.4, Assumption B.8 and Assumption B.9,

$$\begin{aligned}
&\mathbb{E}[|r_{t_2}^{i,n,m} - r_{t_1}^{i,n,m}|^p] \\
&\leq C_1(p)\Big(\mathbb{E}[|\int_{t_1}^{t_2} \tilde{b}(r_u^{i,n,m}) + \frac{1}{N}\sum_{j=1}^N g^m(r_t^{j,n,m})\mathrm{d}u|^p] + \mathbb{E}[|\int_{t_1}^{t_2} 2\theta^n(r_u^{i,n,m})\mathrm{d}W_u^i|^p]\Big) \\
&\leq C_2(p)\Big(\Big(\mathbb{E}\Big[\frac{\tilde{L}^p}{|t_2-t_1|}\int_{t_1}^{t_2} |r_u^{i,n,m}|^p \mathrm{d}u\Big] + \Gamma^p\Big)|t_2-t_1|^p + \mathbb{E}[|\int_{t_1}^{t_2} 2\theta^n(r_u^{i,n,m})\mathrm{d}u|^{p/2}]\Big) \\
&\leq C_2(p)\Big(\Big(\frac{\tilde{L}^p}{|t_2-t_1|}\int_{t_1}^{t_2} \mathbb{E}[|r_u^{i,n,m}|^p]\mathrm{d}u + \Gamma^p\Big)|t_2-t_1|^p + 2^{p/2}|t_2-t_1|^{p/2}\Big) \\
&\leq C_3(p,T,\tilde{L},\Gamma,C_p)|t_2-t_1|^{p/2} \ ,
\end{aligned}$$

where $C_k(\cdot)$ are constants depending on the stated arguments, but independent of n, m. Note that in the second step, we use Burkholder-Davis-Gundy inequality, see [170, Chapter IV, Theorem 48]. It holds similarly, $\mathbb{E}[|s_{t_2}^{i,n,m} - s_{t_1}^{i,n,m}|^p] \leq C_3(p,T,\tilde{L},\Gamma,C_p)|t_2-t_1|^{p/2}$. Hence,

$$\begin{aligned}
\mathbb{E}[|(\{r_{t_2}^{i,n,m}, s_{t_2}^{i,n,m}\}_{i=1}^N) - (\{r_{t_1}^{i,n,m}, s_{t_1}^{i,n,m}\}_{i=1}^N)|^p] & \\
&\leq C_4(p,N)(\sum_{i=1}^N(\mathbb{E}[|r_{t_2}^{i,n,m} - r_{t_1}^{i,n,m}|^p] + \mathbb{E}[|s_{t_2}^{i,n,m} - s_{t_1}^{i,n,m}|^p])) \\
&\leq C_5(p,N,T,\tilde{L},\Gamma,C_p)|t_2-t_1|^{p/2}
\end{aligned}$$

for all $t_1, t_2 \in [0,T]$. Hence, by Kolmogorov's continuity criterion, cf. [121, Corollary 14.9], there exists a constant $\tilde{C}$ depending on p and γ such that

$$\mathbb{E}\Big[[(\{r^{i,n,m}, s^{i,n,m}\}_{i=1}^N)]_\gamma^p\Big] \leq \tilde{C} \cdot C_5(p,N,T,\tilde{L},\Gamma,C_p) \ . \tag{B.81}$$

where $[\cdot]_\gamma^p$ is defined by $[x]_\gamma = \sup_{t_1,t_2\in[0,T]} \frac{|x(t_1)-x(t_2)|}{|t_1-t_2|^\gamma}$ and $(\{r_t^{i,n,m}, s_t^{i,n,m}\}_{i=1}^N)_{n\in\mathbb{N},t\geq 0}$ is tight in $\mathcal{C}([0,T],\mathbb{R}^{2N})$. Hence, for each $T > 0$ there exists a subsequence $n_k \to \infty$ and a probability measure $\mathbb{P}_T$ on $\mathcal{C}([0,T],\mathbb{R}^{2N})$. Since $\{\mathbb{P}_T^m\}_T$ is a consistent family, there exists by [121, Theorem 5.16] a probability measure $\mathbb{P}^m$ on $(\mathbb{W}^N \times \mathbb{W}^N, \mathcal{B}(\mathbb{W}^N) \otimes \mathcal{B}(\mathbb{W}^N))$ such that $\mathbb{P}^{n_k,m}$ converges weakly to $\mathbb{P}^m$. Note that we can take here the same subsequence (n_k) for all m using a diagonalization argument.

Characterization of the limit measure: Denote by $(\{\mathbf{r}_t^i, \mathbf{s}_t^i\}_{i=1}^N) = \omega(t)$ the canonical process on $\mathbb{W}^N \times \mathbb{W}^N$. To characterize the measure $\mathbb{P}^m$ we first note that $\mathbb{P}^m \circ (\mathbf{r}_0^i, \mathbf{s}_0^i)^{-1} = \eta_m$ for all $i \in \{1,\dots,N\}$, since $\mathbb{P}^{n,m}(\mathbf{r}_0^i, \mathbf{s}_0^i)^{-1} = \eta_{n,m}$ converges weakly to η_m by assumption. We define maps $M^{i,m}, N^{i,m} : \mathbb{W}^N \times \mathbb{W}^N \to \mathbb{W}$ by

$$\begin{aligned}
M_t^{i,m} &= \mathbf{r}_t^i - \mathbf{r}_0^i - \int_0^t \Big(\tilde{b}(\mathbf{r}_u^i) + \frac{1}{N}\sum_{j=1}^N g^m(\mathbf{r}_u^j)\Big)\mathrm{d}u \ , \quad \text{and} \\
N_t^{i,m} &= \mathbf{s}_t^i - \mathbf{s}_0^i - \int_0^t \Big(\hat{b}(\mathbf{s}_u^i) + \frac{1}{N}\sum_{j=1}^N h^m(\mathbf{s}_u^j)\Big)\mathrm{d}u \ .
\end{aligned} \tag{B.82}$$

For each $n, m \in \mathbb{N}$ and $i = 1,\dots,N$, $(M_t^{i,m}, \mathcal{F}_t, \mathbb{P}^{n,m})$ is a martingale with respect to the filtration $\mathcal{F}_t = \sigma((\mathbf{r}_u^j, \mathbf{s}_u^j) : j = 1,\dots,N, 0 \leq u \leq t)$. Note that the families $(\{M_t^{i,m}\}_{i=1}^N, \mathbb{P}^{n,m})_{n\in\mathbb{N},t\geq 0}$ and $(\{N_t^{i,m}\}_{i=1}^N, \mathbb{P}^{n,m})_{n\in\mathbb{N},t\geq 0}$ are uniformly integrable.

Since the mappings $M^{i,m}$ and $N^{i,m}$ are continuous in $\mathbb{W}$, $\mathbb{P}^{n,m}\circ(\{\mathbf{r}^i,\mathbf{s}^i,M^{i,m},N^{i,m}\}_{i=1}^N)^{-1}$ converges weakly to $\mathbb{P}^m\circ(\{\mathbf{r}^i,\mathbf{s}^i,M^{i,m},N^{i,m}\}_{i=1}^N)^{-1}$ by the continuous mapping theorem. Then applying the same argument as in (B.59), $(M_t^{m,i},\mathcal{F}_t,\mathbb{P}^m)$ and $(N_t^{m,i},\mathcal{F}_t,\mathbb{P}^m)$ are continuous martingales for all $i=1,\dots,N$ and the quadratic variation $([\{M^{i,m},N^{i,m}\}_{i=1}^N]_t)_{t\geq 0}$ exists $\mathbb{P}^m$-almost surely. To complete the identification of the limit, it suffices to identify the quadratic variation. Similar to the computations in the proof of Lemma B.15, it holds

$$\begin{aligned}
&[M^{i,m}]=4\int_0^\cdot \mathbb{1}_{(0,\infty)}(\mathbf{r}_u^i)\mathrm{d}u && \mathbb{P}^m\text{-almost surely,}\\
&[N^{i,m}]=4\int_0^\cdot \mathbb{1}_{(0,\infty)}(\mathbf{s}_u^i)\mathrm{d}u && \mathbb{P}^m\text{-almost surely, and}\\
&[M^{i,m},N^{i,m}]=4\int_0^\cdot \mathbb{1}_{(0,\infty)}(\mathbf{r}_u^i)\mathbb{1}_{(0,\infty)}(\mathbf{s}_u^i)\mathrm{d}u && \mathbb{P}^m\text{-almost surely,}
\end{aligned}\tag{B.83}$$

Further, $[M^{i,m},M^{j,m}]_t=[N^{i,m},N^{j,m}]_t=[M^{i,m},N^{j,m}]_t=0$ $\mathbb{P}^{n,m}$-almost surely for $i\neq j$ and $(M_t^{i,m}M_t^{j,m},\mathbb{P}^{n,m})$, $(N_t^{i,m}N_t^{j,m},\mathbb{P}^{n,m})$ and $(M_t^{i,m}N_t^{j,m},\mathbb{P}^{n,m})$ are martingales. For any bounded, continuous non-negative function $G:\mathbb{W}\to\mathbb{R}$, it holds

$$E^m[G(M_t^{i,m}M_t^{j,m}-M_s^{i,m}M_s^{j,m})]=\lim_{n\to\infty}E^{n,m}[G(M_t^{i,m}M_t^{j,m}-M_s^{i,m}M_s^{j,m})]=0\ ,$$

respectively, $E^m[G(N_t^{i,m}N_t^{j,m}-N_s^{i,m}N_s^{j,m})]=0$ and $E^m[G(M_t^{i,m}N_t^{j,m}-M_s^{i,m}N_s^{j,m})]=0$. Then

$$[M^{i,m},M^{j,m}]=[N^{i,m},N^{j,m}]=[M^{i,m},N^{j,m}]=0\qquad \mathbb{P}^m\text{-almost surely, for all } i\neq j\ . \tag{B.84}$$

Then by a martingale representation theorem, cf. [116, Chapter II, Theorem 7.1], there is a probability space $(\Omega^m,\mathcal{A}^m,P^m)$ and a Brownian motion $\{W^i\}_{i=1}^N$ and random variables $(\{r^{i,m},s^{i,m}\}_{i=1}^N)$ on this space, such that it holds $P^m\circ(\{r^{i,m},s^{i,m}\}_{i=1}^N)^{-1}=\mathbb{P}^m\circ(\{\mathbf{r}^i,\mathbf{s}^i\}_{i=1}^N)^{-1}$ and such that $(\{r^{i,m},s^{i,m},W^i\}_{i=1}^N)$ is a weak solution of (B.79).
Comparison of two solutions: To show $P^m[r_t^{i,m}\leq s_t^{i,m}$ for all $t\geq 0$ and $i=1,\dots,N]=1$ it suffices to note that $P^{n,m}[r_t^{i,n,m}\leq s_t^{i,n,m}$ for all $t\geq 0$ and $i=1,\dots,N]=1$, which holds by Lemma B.21, carries over to the limit by the Portmanteau theorem, since we have weak convergence of $\mathbb{P}^{n,m}\circ(\{\mathbf{r}^i,\mathbf{s}^i\}_{i=1}^N)^{-1}$ to $\mathbb{P}^m\circ(\{\mathbf{r}^i,\mathbf{s}^i\}_{i=1}^N)^{-1}$. ☐

In the next step we show that the distribution of the solution of (B.79) converges as $m\to\infty$. Consider a probability space $(\Omega^m,\mathcal{A}^m,P^m)$ for each $m\in\mathbb{N}$ and random variables $\{r^{i,m}\}_{i=1}^N,\{s^{i,m}\}_{i=1}^N:\Omega^m\to\mathbb{W}^N$ such that $(\{r_t^{i,m},s_t^{i,m}\}_{i=1}^N)_{t\geq 0}$ is a solution to (B.79). Denote by $\mathbb{P}^m=P^m\circ(\{r^{i,m},s^{i,m}\}_{i=1}^N)^{-1}$ the law on $\mathbb{W}^N\times\mathbb{W}^N$.

Lemma B.23. *Assume that Assumption B.4 and Assumption B.5 is satisfied for $(\tilde{b},g)$ and $(\hat{b},h)$. Let $\eta\in\Gamma(\mu,\nu)$ where the probability measures μ and ν on $\mathbb{R}_+$ satisfy Assumption B.6. Further, let $(g^m)_{m\in\mathbb{N}}$, $(h^m)_{m\in\mathbb{N}}$, $(\mu_m)_{m\in\mathbb{N}}$, $(\nu_m)_{m\in\mathbb{N}}$ and $(\eta_m)_{m\in\mathbb{N}}$ be such that Assumption B.8 and Assumption B.10 hold. Then there exists a random variable $(\{r^i,s^i\}_{i=1}^N)$ defined on some probability space $(\Omega,\mathcal{A},P)$ with values in $\mathbb{W}^N\times\mathbb{W}^N$ such that $(\{r_t^i,s_t^i\}_{i=1}^N)$ is a weak solution of (B.35). Moreover, the laws $P^m\circ(\{r^{i,m},s^{i,m}\}_{i=1}^N)^{-1}$ converge weakly to $P\circ(\{r^i,s^i\}_{i=1}^N)^{-1}$. If in addition,*

$$\begin{aligned}
&\tilde{b}(r)\leq\hat{b}(r),\quad g(r)\leq h(r),\quad \text{and}\quad g^m(r)\leq h^m(r) && \text{for any } r\in\mathbb{R}_+,\ \text{and}\\
&P^m[r_0^{i,m}\leq s_0^{i,m}\ \text{for all } t\geq 0 \text{ and } i\in\{1,\dots,N\}]=1 && \text{for any } m\in\mathbb{N},
\end{aligned}$$

then $P[r_t^i\leq s_t^i$ for all $t\geq 0$ and $i\in\{1,\dots,N\}]=1$.

Proof. The proof is structured as the proof of Lemma B.22. Tightness of the sequence of probability measures $(\mathbb{P}^m)_{m\in\mathbb{N}}$ on $(\mathbb{W}^N \times \mathbb{W}^N, \mathcal{B}(\mathbb{W}^N)\otimes\mathcal{B}(\mathbb{W}^N))$ holds adapting the steps of the proof of Lemma B.22 to (B.79). Note that (B.80) and (B.81) hold analogously for $(\{r_t^{i,m}, s_t^{i,m}\}_{i=1}^N)$ by Assumption B.4, Assumption B.8 and Assumption B.10. Hence by Kolmogorov's continuity criterion, cf. [121, Corollary 14.9], we can deduce that there exists a probability measure $\mathbb{P}$ on $(\mathbb{W}^N \times \mathbb{W}^N, \mathcal{B}(\mathbb{W}^N)\otimes\mathcal{B}(\mathbb{W}^N))$ such that there is a subsequence $(m_k)_{k\in\mathbb{N}}$ along which $\mathbb{P}^{m_k}$ converge towards $\mathbb{P}$.

To characterize the limit, we first note that by Skorokhod representation theorem, cf. [20, Chapter 1, Theorem 6.7], without loss of generality we can assume that $(\{r^{i,m}, s^{i,m}\}_{i=1}^N)$ are defined on a common probability space $(\Omega, \mathcal{A}, P)$ with expectation E and converge almost surely to $(\{r^i, s^i\}_{i=1}^N)$ with distribution $\mathbb{P}$. Then, by Assumption B.8 and Lebesgue convergence theorem it holds almost surely for all $t \geq 0$,

$$\lim_{m\to\infty}\int_0^t \tilde{b}(r_t^{i,m}) + \frac{1}{N}\sum_{j=1}^N g^m(r_u^{j,m})\mathrm{d}u = \int_0^t \tilde{b}(r_t^i) + \frac{1}{N}\sum_{j=1}^N g^m(r_u^j)\mathrm{d}u \ . \tag{B.85}$$

Consider the mappings $M^{i,m}, N^{i,m} : \mathbb{W}^N\times\mathbb{W}^N\times\mathcal{P}(\mathbb{W}^N\times\mathbb{W}^N)\to\mathbb{W}$ defined by (B.82) Then for all $m\in\mathbb{N}$ and $i = 1,\dots,N$, $(M_t^{i,m}, \mathcal{F}_t, \mathbb{P}^m)$ and $(N_t^{i,m}, \mathcal{F}_t, \mathbb{P}^m)$ are martingales with respect to the canonical filtration $\mathcal{F}_t = \sigma((\{\mathbf{r}_u^i, \mathbf{s}_u^i\}_{i=1}^N)_{0\leq u\leq t})$. Further the family $(\{M_t^{i,m}\}_{i=1}^N, \mathbb{P}^m)_{m\in\mathbb{N},t\geq0}$ and $(\{N_t^{i,m}\}_{i=1}^N, \mathbb{P}^m)_{m\in\mathbb{N},t\geq0}$ are uniformly integrable. In the same line as weak convergence is shown in the proof of Lemma B.15 and by (B.85), $\mathbb{P}^m\circ(\{\mathbf{r}^i, \mathbf{s}^i, M^{i,m}, N^{i,m}\}_{i=1}^N)^{-1}$ converges weakly to $\mathbb{P}\circ(\{\mathbf{r}^i, \mathbf{s}^i, M^i, N^i\}_{i=1}^N)^{-1}$ where

$$M_t^i = \mathbf{r}_t^i - \mathbf{r}_0^i - \int_0^t \Big(\tilde{b}(\mathbf{r}_u^i) + \frac{1}{N}\sum_{j=1}^N g(\mathbf{r}_u^j)\Big)\mathrm{d}u \ , \qquad \text{and}$$

$$N_t^i = \mathbf{s}_t^i - \mathbf{s}_0^i - \int_0^t \Big(\hat{b}(\mathbf{s}_u^i) + \frac{1}{N}\sum_{j=1}^N h(\mathbf{s}_u^j)\Big)\mathrm{d}u \ .$$

Then $(\{M_t^i\}_{i=1}^N, \mathcal{F}_t, \mathbb{P})$ and $(\{N_t^i\}_{i=1}^N, \mathcal{F}_t, \mathbb{P})$ are continuous martingales using the same argument as in (B.59). Further, the quadratic variation $([\{M_t^i, N_t^i\}_{i=1}^N]_t)_{t\geq0}$ exists $\mathbb{P}$-almost surely and is given by (B.83) and (B.84) $\mathbb{P}$-almost surely, which holds following the computations in the proof of Lemma B.15 and Lemma B.22. As in Lemma B.22, we conclude by a martingale representation theorem that there are a probability space $(\Omega, \mathcal{A}, P)$ and a Brownian motion $\{W^i\}_{i=1}^N$ and random variables $(\{r^i\}_{i=1}^N, \{s^i\}_{i=1}^N)$ on this space such that $P\circ(\{r^i, s^i\}_{i=1}^N)^{-1} = \mathbb{P}\circ(\{\mathbf{r}^i, \mathbf{s}^i\}_{i=1}^N)^{-1}$ and such that $(\{r^i, s^i, W^i\}_{i=1}^N)$ is a weak solution of (B.25). By the Portmanteau theorem the monotonicity carries over to the limit, since $\mathbb{P}^m\circ(\{\mathbf{r}^i, \mathbf{s}^i\}_{i=1}^N)^{-1}$ converges weakly to $\mathbb{P}\circ(\{\mathbf{r}^i, \mathbf{s}^i\}_{i=1}^N)^{-1}$. □

Proof of Theorem B.10. The proof is a direct consequence of Lemma B.22 and Lemma B.23. □

B.7 Appendix

B.7.1 Kuramoto model

Lower bounds on the contraction rate can also be shown for nonlinear SDEs on the one-dimensional torus using the same approach. Here, we consider the Kuramoto model given by

$$\mathrm{d}X_t = -k\left[\int_{\mathbb{T}} \sin(X_t - x)\mathrm{d}\mu_t(x)\right]\mathrm{d}t + \mathrm{d}B_t \tag{B.86}$$

on the torus $\mathbb{T} = \mathbb{R}/(2\pi\mathbb{Z})$.

Theorem B.24. *Let μ_t and ν_t be laws of X_t and Y_t where $(X_s)_{s\geq 0}$ and $(Y_s)_{s\geq 0}$ are two solutions of* (B.86) *with initial distributions μ_0 and ν_0 on $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$, respectively. If*

$$4k\int_0^\pi \exp(2k - 2k\cos(r/2))\mathrm{d}r \leq 1 \tag{B.87}$$

holds, then for all $t \geq 0$,

$$\mathcal{W}_{\tilde{f}}(\mu_t, \nu_t) \leq \mathrm{e}^{-c_{\mathbb{T}}t}\mathcal{W}_{\tilde{f}}(\mu_0, \nu_0) \qquad \text{and} \qquad \mathcal{W}_1(\mu_t, \nu_t) \leq 2\exp(2k)\mathrm{e}^{-c_{\mathbb{T}}t}\mathcal{W}_1(\mu_0, \nu_0)\ ,$$

where

$$c_{\mathbb{T}} = 1/(2\int_0^\pi \int_0^r \exp[2k(\cos(r/2) - \cos(s/2))]\mathrm{d}s\mathrm{d}r) \tag{B.88}$$

and $\tilde{f}$ is a concave, increasing function given in (B.92).

In [64, Appendix A], a contraction result is stated for a general drift using a similar approach.

We prove Theorem B.24 via a sticky coupling approach. In the same line as in Appendix B.2 the coupling $(X_t, Y_t)_{t\geq 0}$ is defined as the weak limit of Markovian couplings $\{(X_t^\delta, Y_t^\delta)_{t\geq 0} : \delta > 0\}$ on $\mathbb{T}\times\mathbb{T} = \mathbb{R}/(2\pi\mathbb{Z}) \times \mathbb{R}/(2\pi\mathbb{Z})$ given by

$$\begin{aligned}\mathrm{d}X_t^\delta &= -k\left[\int_{\mathbb{T}} \sin(X_t^\delta - x)\mathrm{d}\mu_t^\delta(x)\right]\mathrm{d}t + \mathrm{rc}^\delta(\bar{r}_t^\delta)\mathrm{d}B_t^1 + \mathrm{sc}^\delta(\bar{r}_t^\delta)\mathrm{d}B_t^2 \\ \mathrm{d}Y_t^\delta &= -k\left[\int_{\mathbb{T}} \sin(Y_t^\delta - x)\mathrm{d}\nu_t^\delta(x)\right]\mathrm{d}t - \mathrm{rc}^\delta(\bar{r}_t^\delta)\mathrm{d}B_t^1 + \mathrm{sc}^\delta(\bar{r}_t^\delta)\mathrm{d}B_t^2\ ,\end{aligned} \tag{B.89}$$

where $\bar{r}_t^\delta = d_{\mathbb{T}}(X_t^\delta, Y_t^\delta)$ with $d_{\mathbb{T}}(\cdot,\cdot)$ defined by

$$d_{\mathbb{T}}(x,y) = \begin{cases}(|x-y| \bmod 2\pi) & \text{if } (|x-y| \bmod 2\pi) \leq \pi\ , \\ (2\pi - |x-y| \bmod 2\pi) & \text{otherwise}\ .\end{cases} \tag{B.90}$$

The functions $\mathrm{rc}^\delta, \mathrm{sc}^\delta$ are given by (B.19) and satisfy that there exists $\epsilon_0 > 0$ such that $\mathrm{rc}^\delta(r) \geq r/2$ for any $0 \leq r \leq \delta \leq \epsilon_0$.

Theorem B.25. *Assume* (B.87)*. Let μ_0 and ν_0 be probability measures on $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$ having finite forth moment. Then, $(X_t, Y_t)_{t\geq 0}$ is a subsequential limit in distribution as $\delta \to 0$ of $\{(X_t^\delta, Y_t^\delta)_{t\geq 0} : \delta > 0\}$, where $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ are solutions of* (B.86) *with initial distributions μ_0 and ν_0, respectively. Further, there exists a process $(r_t)_{t\geq 0}$ satisfying for any $t \geq 0$, $d_{\mathbb{T}}(X_t, Y_t) \leq r_t$ almost surely, and which is a weak solution of*

$$\mathrm{d}r_t = (2k\sin(r_t/2) + 2k\mathbb{P}(r_t))\mathrm{d}t + 2\mathbb{1}_{(0,\pi]}(r_t)\mathrm{d}W_t - 2\mathrm{d}\ell_t^\pi\ , \tag{B.91}$$

where $(W_t)_{t\geq 0}$ is a one-dimensional Brownian motion on $\mathbb{T}$ and ℓ^π is the local time at π.

Proof. The proof works analogously to the proof of Theorem B.2 stated in Appendix B.6.2. It holds similarly to Lemma B.12 by Meyer-Tanaka's formula, cf. [172, Chapter 6, Theorem 1.1], and using (B.90),

$$\bar{r}_t^\delta - \bar{r}_0^\delta = \int_0^t \mathrm{sgn}(X_t^\delta - Y_t^\delta)(-k)e_t \left[\int_{\mathbb{T}} \sin(X_t^\delta - x)\mathrm{d}\mu_t(x) - \int_{\mathbb{T}} \sin(Y_t^\delta - x)\mathrm{d}\nu_t(x)\right] \mathrm{d}t$$
$$+ \int_0^t \mathrm{sgn}(X_t^\delta - Y_t^\delta) 2\mathrm{rc}^\delta(\bar{r}_t^\delta)e_t \mathrm{d}B_t^1 + \int_{\mathbb{R}} 2\mathrm{rc}^\delta(\bar{r}_t^\delta)^2 \ell_t^a(\delta_0 - \delta_\pi)(\mathrm{d}a) ,$$

where $\mathrm{sgn}(x) = \mathbb{1}_{(0,\pi]}(x) - \mathbb{1}_{(\pi,2\pi]}(x)$, $(\ell_t^a)_{t\geq 0}$ is the local time at a associated with $(X_t^\delta - Y_t^\delta)_{t\geq 0}$ and $e_t = (X_t^\delta - Y_t^\delta)/d_{\mathbb{T}}(X_t^\delta, Y_t^\delta)$ for $\bar{r}_t^\delta \neq 0$. For $\bar{r}_t^\delta = 0$, e_t is some arbitrary unit vector. For any a the support of ℓ_t^a as a function of t is a subset of the set of t such that $r_t = a$ [121, Theorem 19.1], hence $\mathbb{1}_{(0,\pi]}(r_t)\ell_t^0 = 0$ almost surely and so the term involving the local time reduces to $-2\ell_t^\pi$. Further, we note that $W_t = \int_0^t \mathrm{sgn}(X_t^\delta - Y_t^\delta)e_t \mathrm{d}B_t^1$ is a Brownian motion. As in Lemma B.12, it holds for the process $(\bar{r}_t^\delta)_{t\geq 0}$ for $\epsilon < \epsilon_0$ with ϵ_0 given by (B.20),

$$\mathrm{d}\bar{r}_t^\delta \leq (2k\sin(\bar{r}_t^\delta/2) + 2k\mathbb{E}_{x\sim\mu_t^\delta, y\sim\nu_t^\delta}(\mathrm{rc}^\epsilon(d_{\mathbb{T}}(x,y))))\mathrm{d}t + 2\mathrm{rc}^\delta(\bar{r}_t^\delta)\mathrm{d}W_t - 2\mathrm{d}\ell_t^\pi ,$$

where we used the properties of rc^δ and

$$(x-y)\cdot(\sin(x-\tilde{x}) - \sin(y-\tilde{x})) \leq 2\sin(|x-y|/2)|x-y|$$

for any $x, y, \tilde{x} \in \mathbb{T}$. Consider $(r_t^{\delta,\epsilon})_{t\geq 0}$ given by

$$\mathrm{d}r_t^{\delta,\epsilon} = (2k\sin(r_t^{\delta,\epsilon}/2) + 2k\int_0^\pi \mathrm{rc}^\epsilon(u)\mathrm{d}P_t^{\delta,\epsilon}(u))\mathrm{d}t + 2\mathrm{rc}^\delta(r_t^{\delta,\epsilon})\mathrm{d}W_t - 2\mathrm{d}\ell_t^\pi ,$$

where $P_t^{\delta,\epsilon}$ is the law of $r_t^{\delta,\epsilon}$. Then as in Lemma B.13, for the processes $(\bar{r}_t^\delta)_{t\geq 0}$ and $(r_t^{\delta,\epsilon})_{t\geq 0}$ with the same initial condition and driven by the same noise it holds $\bar{r}_t^\delta \leq r_t^{\delta,\epsilon}$ almost surely for every t and $\epsilon < \epsilon_0$.

Consider the process $(U_t^{\delta,\epsilon})_{t\geq 0} = (X_t^\delta, Y_t^\delta, r_t^{\delta,\epsilon})_{t\geq 0}$ on $\mathbb{T}^2 \times [0,\pi]$ for each $\epsilon, \delta > 0$. We define by $\mathbf{X}, \mathbf{Y} : \mathcal{C}(\mathbb{R}_+, \mathbb{T}^2 \times [0,\pi]) \to \mathcal{C}(\mathbb{R}_+, \mathbb{T})$ and $\mathbf{r} : \mathcal{C}(\mathbb{R}_+, \mathbb{T}^2 \times [0,\pi]) \to \mathcal{C}(\mathbb{R}_+, [0,\pi])$ the canonical projections onto the first component, onto the second component and onto the last component, respectively. Analogously to the proof of Theorem B.2, the law $\mathbb{P}^{\delta,\epsilon}$ of the process $(U_t^{\delta,\epsilon})_{t\geq 0}$ converges along a subsequence $(\delta_k, \epsilon_k)_{k\in\mathbb{N}}$ to a probability measure $\mathbb{P}$. Let $(X_t, Y_t, r_t)_{t\geq 0}$ be some process on $\mathbb{T}^2 \times [0,\pi]$ with distribution $\mathbb{P}$ on $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P})$. Since $(X_t^\delta)_{t\geq 0}$ and $(Y_t^\delta)_{t\geq 0}$ are solutions of (B.86) which are unique in law, we have that for any $\epsilon, \delta > 0$, $\mathbb{P}^{\delta,\epsilon} \circ \mathbf{X}^{-1} = \mathbb{P} \circ \mathbf{X}^{-1}$ and $\mathbb{P}^{\delta,\epsilon} \circ \mathbf{Y}^{-1} = \mathbb{P} \circ \mathbf{Y}^{-1}$. And therefore $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ are solutions of (B.86) as well with the same initial condition. Hence $\mathbb{P} \circ (\mathbf{X}, \mathbf{Y})^{-1}$ is a coupling of two copies of (B.86).

Further, the monotonicity $\bar{r}_t^\delta \leq r_t^{\delta,\epsilon}$ carries over to the limit by the Portmanteau theorem. Finally, similarly to the proof of Lemma B.15 and Lemma B.16 there exist an extended probability space and a one-dimensional Brownian motion $(W_t)_{t\geq 0}$ such that $(r_t, W_t)_{t\geq 0}$ is a solution to (B.97).

□

Proof of Theorem B.24. Similarly to (B.37) we consider a function $\tilde{f}$ on $[0,\pi]$ defined by

$$\tilde{f}(t) = \int_0^t \tilde{\varphi}(r)\tilde{g}(r)\mathrm{d}r , \tag{B.92}$$

where

$$\begin{aligned}
&\tilde{\varphi}(r) = \exp\{2k(\cos(r/2)-1)\}\ , \qquad \tilde{\Phi}(r) = \int_0^r \tilde{\varphi}(s)\mathrm{d}s\ ,\\
&\tilde{g}(r) = 1 - \frac{c_{\mathbb{T}}}{2}\int_0^r \{\tilde{\Phi}(s)/\tilde{\varphi}(s)\}\mathrm{d}s - k\int_0^r \{1/\tilde{\varphi}(s)\}\mathrm{d}s\ ,\\
&c_{\mathbb{T}} = \left(2\int_0^\pi \{\tilde{\Phi}(s)/\tilde{\varphi}(s)\}\mathrm{d}s\right)^{-1}\ .
\end{aligned}$$

Then for k satisfying (B.87), $\tilde{g}(r)\in[1/2,1]$ and $\tilde{f}$ is a concave function satisfying similarly to (B.38)

$$\exp(-2k)/2r \le \tilde{f} \le \tilde{\Phi}(r) \le r \tag{B.93}$$

and

$$\begin{aligned}
&\tilde{f}''(0) = -k\\
&2(\tilde{f}''(r)-\tilde{f}''(0)) \le -2k\sin(r/2)\tilde{f}'(r) - c_{\mathbb{T}}\tilde{f}(r) \qquad \text{for all } r\in[0,\pi]\ .
\end{aligned} \tag{B.94}$$

By Ito's formula it holds

$$\begin{aligned}
\mathrm{d}\tilde{f}(r_t) = {}& \tilde{f}'(r_t)(2k\sin(r/2)+2k\mathbb{P}(r_t>0))\mathrm{d}t + 2\tilde{f}'(r_t)\mathbb{1}_{(0,\pi]}(r_t)\mathrm{d}W_t - 2\tilde{f}'(r_t)\mathrm{d}\ell_t^\pi\\
&+ 2\tilde{f}''(r_t)\mathbb{1}_{(0,\pi]}(r_t)\mathrm{d}t\ .
\end{aligned}$$

Taking expectation and using that the term involving the local time is negative, we obtain

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\tilde{f}(r_t)] &\le \mathbb{E}[2(\tilde{f}''(r_t)-f''(0)) + \tilde{f}'(r_t)2k\sin(r_t/2)] + (2\tilde{f}''(0)+2k)\mathbb{P}(r_t>0)\\
&\le -c_{\mathbb{T}}\mathbb{E}[\tilde{f}(r_t)]\ ,
\end{aligned}$$

where the last step holds by (B.94). Then

$$\mathbb{E}[\tilde{f}(d_{\mathbb{T}}(\bar{X}_t,\bar{Y}_t))] \le \mathbb{E}[\tilde{f}(r_t)] \le \mathrm{e}^{-c_{\mathbb{T}}t}\mathbb{E}[\tilde{f}(r_0)] = \mathrm{e}^{-c_{\mathbb{T}}t}\mathbb{E}[\tilde{f}(d_{\mathbb{T}}(\bar{X}_0,\bar{Y}_0))]\ , \tag{B.95}$$

provided (B.87) holds. Thus

$$\mathcal{W}_{\tilde{f}}(\mu_t,\nu_t) \le \mathrm{e}^{-c_{\mathbb{T}}t}\mathcal{W}_{\tilde{f}}(\mu_0,\nu_0)\ ,$$

and by (B.93)

$$\mathcal{W}_1(\mu_t,\nu_t) \le 2\exp(2k)\mathrm{e}^{-c_{\mathbb{T}}t}\mathcal{W}_1(\mu_0,\nu_0)\ .$$

□

Remark B.26. Let us finally remark that we can relax the condition (B.87) and we can obtain contraction with a modified contraction rate $c_{\mathbb{T}}$ for all $k<k_0$, where k_0 is given by

$$k_0\int_0^\pi \exp(2k_0 - 2k_0\cos(r/2))\mathrm{d}r = 1\ . \tag{B.96}$$

More precisely, set $\zeta = 1-k\int_0^\pi \exp(2k-2k\cos(r/2))\mathrm{d}r$ and $c_{\mathbb{T}} = \zeta\left(\int_0^\pi\{\tilde{\Phi}(s)/\tilde{\varphi}(s)\}\mathrm{d}s\right)^{-1}$. Then, $\tilde{g}(r)\in[\zeta/2,1]$ and $\zeta\exp(-2k)/2r \le \tilde{f}(r) \le r$. Following the previous computations, we obtain

$$\mathcal{W}_1(\mu_t,\nu_t) \le 2\exp(2k)/\zeta\mathrm{e}^{-c_{\mathbb{T}}t}\mathcal{W}_1(\mu_0,\nu_0)\ ,$$

where for k close to k_0, the contraction rate becomes small and the prefactor $2\exp(2k)/\zeta$ explodes.

B.7.2 Sticky nonlinear SDEs on bounded state space

In the same line as in Theorem B.3, existence, uniqueness in law and comparison results hold for solutions to the sticky SDE on $[0,\pi]$ given by

$$\mathrm{d}r_t = (\tilde{b}(r_t) + 2k\mathbb{P}(r_t > 0))\mathrm{d}t + 2\mathbb{1}_{(0,\pi)}(r_t)\mathrm{d}W_t - 2\mathrm{d}\ell_t^{\pi}, \tag{B.97}$$

where $k \in \mathbb{R}_+$ and ℓ^{π} is the local time at π.

The analysis of invariant measures and phase transitions can be easily adapted to the case of the sticky SDE on $[0,\pi]$ given by (B.97).

Theorem B.27. *Let* $(r_t)_{t\geq 0}$ *be a solution of* (B.97) *with drift* $\tilde{b}$ *satisfying Assumption B.4. Then, the Dirac measure at zero,* δ_0*, is an invariant probability measure on* $[0,\pi]$ *for* (B.97)*. If there exists* $p \in (0,1)$ *solving* $(1/k) = (1-p)I(k,p)$ *where*

$$I(k,p) = \int_0^{\pi} \exp\Big(kpx + \frac{1}{2}\int_0^x \tilde{b}(r)\mathrm{d}r\Big)\mathrm{d}x \ ,$$

then the probability measure π *on* $[0,\pi]$ *given by*

$$\pi(\mathrm{d}x) \propto \frac{1}{kp}\delta_0(\mathrm{d}x) + \exp\Big(kpx + \frac{1}{2}\int_0^x \tilde{b}(r)\mathrm{d}r\Big)\lambda_{(0,\pi)}(\mathrm{d}x) \tag{B.98}$$

is another invariant probability measure for (B.97)*.*

Proof of Theorem B.27. The proof works analogously to the proof of Theorem B.5 for sticky SDEs on $\mathbb{R}_+$. Note that here the condition (B.66) transforms for $p \in (0,1]$ to

$$p = \pi((0,\pi)) = \frac{I(k,p)}{1/(kp) + I(k,p)} \Leftrightarrow (1-p)I(k,p) = 1/k \ .$$

□

Example B.28. Consider a solution $(r_t)_{t\geq 0}$ of (B.97) with drift $\tilde{b}(r) = 2k\sin(r/2)$. Consider a solution $p \in (0,1]$ solving $1/k = (1-p)I(k,p)$ with

$$I(k,p) = \int_0^{\pi} \exp\Big(kpx + \int_0^x k\sin(r/2)\mathrm{d}r\Big)\mathrm{d}x = \int_0^{\pi} \exp\Big(kpx + 2k - 2k\cos(x/2)\Big)\mathrm{d}x \ .$$

Then by Theorem B.27, the Dirac measure at zero, δ_0 and the probability measure

$$\pi(\mathrm{d}x) \propto \frac{1}{kp}\delta_0(\mathrm{d}x) + \exp(kpx + 2k - 2k\cos(x/2))\lambda_{(0,\pi)}(\mathrm{d}x) \tag{B.99}$$

are invariant probability measures for (B.97). We specify a necessary and sufficient condition for the existence of a solution p satisfying $1/k = (1-p)I(k,p)$. We define $\hat{I}(k,p) = (1-p)I(k,p)$. We first consider the case $1/k < \hat{I}(k,0) = \int_0^{\pi} \exp(2k - 2k\cos(x/2))\mathrm{d}x$. Then since $1/k > \hat{I}(k,1) = 0$ and by the mean value theorem there exists a p solving $1/k = \hat{I}(k,p)$ and therefore there exist multiple invariant distributions for (B.99). On the other hand, if $1/k > \hat{I}(k,0) = \int_0^{\pi} \exp(2k - 2k\cos(x/2))\mathrm{d}x$, since $\pi \leq \int_0^{\pi} \exp(2k - 2k\cos(x/2))\mathrm{d}x$ and for $k < 1/\pi$, it holds

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}p}\hat{I}(k,p) &= -I(k,p) + (1-p)\int_0^{\pi} kx\exp(kpx + 2k - 2k\cos(x/2))\mathrm{d}x \\ &= \int_0^{\pi} ((1-p)kx - 1)\exp(kpx + 2k - 2k\cos(x/2))\mathrm{d}x \leq 0 \ ,\end{aligned}$$

there is no p satisfying (B.99).

Remark B.29. The contraction result given in Theorem B.7 carries over to the sticky diffusion (r_t) given by (B.97) on $[0,\pi]$ with $\tilde{b}(r) = 2k\sin(r/2)$. If (B.87) holds, then for $t \geq 0$, (B.31) holds with $\tilde{f}$ defined in (B.92) and $c_{\mathbb{T}}$ defined in (B.88) using (B.95). Moreover by Remark B.26, we can deduce that if (B.96) holds, the Dirac measure at zero, δ_0, is the unique invariant measure and contraction towards δ_0 holds.

Appendix C

Global contractivity for Langevin dynamics with distribution-dependent forces and uniform in time propagation of chaos

Katharina Schuh, Global contractivity for Langevin dynamics with distribution-dependent forces and uniform in time propagation of chaos. ArXiv e-print arXiv:2206.03082, June 2022.[1]

Abstract

We study the long-time behaviour of both the classical second-order Langevin dynamics and the nonlinear second-order Langevin dynamics of McKean-Vlasov type. By a coupling approach, we establish global contraction in an L^1 Wasserstein distance with an explicit dimension-free rate for pairwise weak interactions. For external forces corresponding to a κ-strongly convex potential, a contraction rate of order $\mathcal{O}(\sqrt{\kappa})$ is obtained in certain cases. But the contraction result is not restricted to these forces. It rather includes multi-well potentials and non-gradient-type external forces as well as non-gradient-type repulsive and attractive interaction forces. The proof is based on a novel distance function which combines two contraction results for large and small distances and uses a coupling approach adjusted to the distance. By applying a componentwise adaptation of the coupling we provide uniform in time propagation of chaos bounds for the corresponding mean-field particle system.

Key words: Langevin dynamics, coupling, convergence to equilibrium, Wasserstein distance, Vlasov-Fokker-Planck equation, propagation of chaos
Mathematics Subject Classification: 60H10, 60J60, 82C31

[1]The author would like to thank her supervisor Andreas Eberle for bringing up the idea of glueing two metrics to combine two local contraction results and for his support and advice during the development of this work. Support by the *Hausdorff Center for Mathematics* has been gratefully acknowledged. Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) im Rahmen der Exzellenzstrategie des Bundes und der Länder - GZ 2047/1, Projekt-ID 390685813.

C.1 Introduction

In this paper, we are interested in the long-time behaviour of the Langevin diffusion $(X_t, Y_t)_{t\geq 0}$ of McKean-Vlasov type on $\mathbb{R}^{2d}$ given by the stochastic differential equation

$$\begin{cases} \mathrm{d}\bar{X}_t = \bar{Y}_t \mathrm{d}t \\ \mathrm{d}\bar{Y}_t = (ub(\bar{X}_t) + u\int_{\mathbb{R}^d} \tilde{b}(\bar{X}_t, z)\bar{\mu}_t^x(\mathrm{d}z) - \gamma \bar{Y}_t)\mathrm{d}t + \sqrt{2\gamma u}\mathrm{d}B_t, \qquad \bar{\mu}_t^x = \mathrm{Law}(\bar{X}_t), \end{cases} \tag{C.1}$$

where $b : \mathbb{R}^d \to \mathbb{R}^d$ and $\tilde{b} : \mathbb{R}^{2d} \to \mathbb{R}^d$ are two Lipschitz continuous functions, $u, \gamma > 0$ are two positive constants and $(B_t)_{t\geq 0}$ is a d-dimensional standard Brownian motion. The functions b and $\tilde{b}$ denote the *external force* and the *interaction force*, respectively. If $\tilde{b} \equiv 0$, (C.1) corresponds to the classical Langevin dynamics, which is also of particular interest and whose long-time behaviour will separately be studied in detail. Existence of a solution and uniqueness in law hold provided the initial conditions have bounded second moments and b and $\tilde{b}$ are Lipschitz continuous [146, Theorem 2.2].

Equation (C.1) is the probabilistic description of the *Vlasov-Fokker-Planck equation* given by

$$\partial_t f_t(x,y) = \nabla_y \cdot \Big[\gamma \nabla_y f_t(x,y) + \gamma y f_t(x,y) + u\Big(b(x) + \int_{\mathbb{R}^d} \tilde{b}(x,z)\bar{\mu}_t^x(\mathrm{d}z)\Big)\bar{\mu}_t(x,y)\Big] - u\nabla_x \cdot [y f_t(x,y)], \tag{C.2}$$

where f_t is the time dependent density function on $\mathbb{R}^{2d}$ and $\bar{\mu}_t^x$ is the marginal distribution in the first component of $\bar{\mu}_t(\mathrm{d}x\mathrm{d}y) = f_t(x,y)\mathrm{d}x\mathrm{d}y$. The solution $(f_t)_{t\geq 0}$ of (C.2) describes the density function of the process $(\bar{X}_t, \bar{Y}_t)_{t\geq 0}$ which moves according to (C.1). Often, b and $\tilde{b}$ are of the form $b(x) = -\nabla \mathrm{V}(x)$ and $\tilde{b}(x, x') = -\nabla_x \mathrm{W}(x, x')$ for all $x, x' \in \mathbb{R}^d$ and for some functions $\mathrm{V} \in \mathcal{C}^1(\mathbb{R}^d)$ and $\mathrm{W} \in \mathcal{C}^1(\mathbb{R}^{2d})$, which are called *confinement potential* and *interaction potential*, respectively.

Besides the long-time behaviour of (C.1), we study the mean-field particle system corresponding to (C.1) with $N \in \mathbb{N}$ particles which is given by

$$\begin{cases} \mathrm{d}X_t^{i,N} = Y_t^{i,N}\mathrm{d}t \\ \mathrm{d}Y_t^{i,N} = (ub(X_t^{i,N}) + N^{-1}\sum_{j=1}^N u\tilde{b}(X_t^{i,N}, X_t^{j,N}) - \gamma Y_t^{i,N})\mathrm{d}t + \sqrt{2\gamma u}\mathrm{d}B_t^i, \qquad i = 1, ..., N. \end{cases} \tag{C.3}$$

We are interested in establish conditions on b and $\tilde{b}$ such that for all $t \geq 0$ for $N \to \infty$ the law of the particles converges to the law of $(\bar{X}_t, \bar{Y}_t)$. This phenomenon was stated under the name *propagation of chaos* and was first introduced by Kac for the Boltzmann equation in [120]. For finite time horizon, bounds on the difference between the law of the particle system and the law of N independent solutions to (C.1) are established by McKean [143] provided b and $\tilde{b}$ are Lipschitz continuous and bounded. This result is further developed in e.g. [187, 146], see [45, 46] for a overview and the references therein.

The equations (C.1), (C.2), (C.3) and its variants have various applications in physics. If $\tilde{b} \equiv 0$, the solution of (C.1) can be interpreted as a particle having a position $\bar{X}_t$ and a velocity $\bar{Y}_t$ and which moves according to the external force. The constant $\gamma > 0$ corresponds to the friction parameter and $u > 0$ denotes the inverse of the mass per particle. Equation (C.3) describes many particles whose moves are additionally determined by pairwise interactions given by the interaction force. Equation (C.2) describes the limit distribution as the number of particles tends to infinity.

In the deep learning community, Langevin dynamics with a mean-field interaction provide a tool to prove trainability of neural networks [145, 178]. Algorithms using Langevin dynamics have a better long-time behaviour compared to the overdamped Langevin dynamics [55, 60], which forms a degenerated special case of the Langevin dynamics, where the limit for γ to infinity is taken [165, Section 6.5.1]. Therefore, nonlinear Langevin dynamics became recently popular for training networks as the Generative Adversarial Network (GAN) [122].

If $\tilde{b} \equiv 0$ and $b = \nabla \mathrm{V}$, then under some mild conditions on V the unique invariant measure is given by the Boltzmann-Gibbs distribution

$$\mu_\infty(\mathrm{d}x\ \mathrm{d}y) \propto \exp(-\mathrm{V}(x) - |y|^2/(2u)),$$

see e.g. [165, Proposition 6.1]. Otherwise, i.e., if b is not of gradient-type or $\tilde{b} \neq 0$, it is often not clear if uniqueness of an invariant probability measure holds (see [71]) and how fast the marginal law of a solution of (C.1) converges towards it.

Getting a clear picture of the long-time behaviour of processes given by stochastic differential equations with and without nonlinear forces of McKean-Vlasov type is of wide interest and the objective of many works. For the overdamped Langevin dynamics forming a first-order equation, the long-time behaviour is studied using both analytic approaches as functional inequalities (e.g. [11, 24]) and probabilistic approaches as coupling techniques. Via a reflection coupling, Eberle [83] established contraction in L^1 Wasserstein distance with respect to a carefully aligned distance function with explicit rates for locally non-convex potentials. For the dynamics with an additional nonlinear drift term, which appears to model for example granular media (see [13]), exponential convergence rates have been investigated for uniformly convex potentials in [42] using gradient flow structure, Logarithmic Sobolev inequalities and transportation cost inequalities (see [43, 137, 44] for relaxations to certain non-uniformly convex potentials). Further, [137, 44] provide uniform in time propagation of chaos estimates for the corresponding particle system. Based on a coupling approach consisting of a mixture of a synchronous and a reflection coupling, uniform in time propagation of chaos is shown in [75] for possibly non strongly convex confinement potentials and possibly non-convex interaction potentials. For the unconfined dynamics (i.e., $b = 0$) exponential convergence is studied in [44, 25] for convex interaction potentials applying analytic tools. If the convexity assumption on the interaction potential is removed, exponential convergence and propagation of chaos can still be established for unconfined overdamped Langevin dynamics via a sticky coupling approach (see [74]) for a class of interaction forces that split in a linear term and a perturbation part.

Proving contraction rates for second-order SDEs given by (C.1) is more delicate as additionally one has to deal with the hypoellipticity of the diffusion. In the case of the classical Langevin dynamics with a gradient-type force, i.e., when $b = \nabla \mathrm{V}$ and $\tilde{b} \equiv 0$ hold, exponential convergence is studied in e.g. [4, 67, 68, 109, 111, 110, 191] using analytic methods including the Witten Laplacian, semigroups, functional inequalities and hypocoercivity. To our knowledge, the best-known contraction rate is obtained for κ-strongly convex potentials V in [40], where contraction in L^2 distance is shown with a rate of order $\mathcal{O}(\sqrt{\kappa})$ via a Poincaré type inequality. Harris type theorems, involving a Lyapunov drift condition, provide a probabilistic technique to analyse the long-time behaviour of Langevin dynamics, see [10, 200, 140, 188]. An alternative powerful probabilistic approach, which provides quantitative rates, is based on couplings. Via a synchronous coupling approach, Dalalyan and Riou-Durand [60] showed contraction in Wasserstein distance with rate of order $\mathcal{O}(\kappa/\sqrt{L})$ for κ-strongly convex potentials with L-Lipschitz continuous gradients if $L\gamma^{-2}u \leq 1$ holds. In [85], Eberle, Guillin and Zimmer introduced a coupling for the Langevin dynamics including non-convex confinement potentials and showed exponential convergence with explicit rates. There, contraction is shown in a specific L^1 Wasserstein distance with respect to a semimetric involving a Lyapunov function. More precisely, for large distances, a synchronous coupling is considered and the Lyapunov function in the semimetric yields contraction. For small distances, the noise is synchronized on a line, where contraction for the position is observed, and reflected otherwise to force the dynamics to return to that line. Combining the results of the different areas, contraction in average is obtained for a carefully aligned semimetric. Due to the Lyapunov function, the contraction rate depends on the dimension and the semimetric is not applicable for nonlinear Langevin dynamics, which suggests getting rid of the Lyapunov function and treating the area of large distances differently.

To get results on the long-time behaviour for nonlinear Langevin diffusions given by (C.1), we have to handle both the difficulties coming from the nonlinearity and the hypoellipticity of the equation. Beginning with the analytic approaches, let us mention the work by Villani [191], where the hypocoercivity is extended to the framework on the torus with small interactions, see also the work by Bouchut and Dolbeault [36]. Using a free energy approach, convergence to equilibrium is studied in [72] for specific non-convex confining potentials and convex polynomial interaction potentials. Applying functional inequalities for mean-field models, established in [99] to prove convergence to equilibrium in weighted Sobolev norm, Monmarché and Guillin proved propagation of chaos for (C.3) in [151, 100]. There, they considered both strongly convex confinement potentials and more general confinement potentials and attractive interaction potentials with at most quadratic growth.

Coupling techniques are also employed in the study of the nonlinear dynamics (C.1). In [26], convergence to equilibrium is shown via a synchronous coupling for small Lipschitz interactions and a quadratic-like friction term. The combination of the coupling approach of [85] and a Lyapunov function is used in [122] to prove exponential contraction in the case of certain small mean-field potentials of non-convolution-type. There, the results are applied to the numerical discretized version of the dynamics corresponding to the Hamiltonian Stochastic Gradient Descent, and the connection to the analysis of deep neural networks is drawn, see [114] for further references on the connection to deep learning. Very closely related to this work is the recent preprint [97] by Guillin, Le Bris and Monmarché, which has been prepared independently in parallel. They considered non-globally convex confinement potentials and Lipschitz continuous even interaction potentials and extended the approach by [85]. More precisely, they modified the semimetric by a sophisticated Lyapunov function to treat the nonlinear Langevin dynamics and to obtain propagation of chaos bounds. The main differences between this work and [97] are that here we include forces that are not necessarily of gradient type and that we establish global contractivity with dimension-free rates by constructing a novel distance function and modifying the coupling approach of [85] appropriately. In particular, we consider two separate metrics r_l and r_s for large and small distances instead of a semimetric involving a Lyapunov function and establish contraction for both metrics separately. For small distances we make use of the results by [85], whereas for large distances we consider a twisted 2-norm structure for the metric r_l of the form $(x \cdot (Ax) + x \cdot (By) + y \cdot (Cy)$ with positive definite matrices $A, B, C \in \mathbb{R}^{d\times d}$. This structure is similar to the structure appearing in the Lyapunov function in [140, 188] and to the norm used in e.g. [4] to prove contraction for certain strongly convex potentials.

Then, our first main contribution is a global contraction result in Wasserstein distance with respect to a distance ρ that is carefully glued of r_s and r_l and that is equivalent to the Euclidean distance. More precisely, we impose b to be a sum of a linear function $-Kx$, where $K \in \mathbb{R}^{d\times d}$ is a positive definite matrix with smallest eigenvalue κ, and a certain Lipschitz continuous function $g(x)$ with Lipschitz constant L_g which is such that b includes gradients of asymptotically strongly convex potentials. If the friction parameter γ is sufficiently large, i.e., $\gamma^2 > 2L_g^2 u/\kappa$, and if the Lipschitz constant $\tilde{L}$ of the interaction force $\tilde{b}$ is sufficiently small, we prove for two probability measures $\bar{\mu}_0$ and $\bar{\nu}_0$ on $\mathbb{R}^{2d}$ with finite second moment,

$$\mathcal{W}_\rho(\bar{\mu}_t, \bar{\nu}_t) \leq e^{-ct}\mathcal{W}_\rho(\bar{\mu}_0, \bar{\nu}_0), \qquad \text{and} \qquad \mathcal{W}_1(\bar{\mu}_t, \bar{\nu}_t) \leq M_1 e^{-ct}\mathcal{W}_1(\bar{\mu}_0, \bar{\nu}_0), \tag{C.4}$$

where $\bar{\mu}_t$ and $\bar{\nu}_t$ are the laws of the solutions $(\bar{X}_t, \bar{Y}_t)$ and $(\bar{X}_t', \bar{Y}_t')$ to (C.1) with initial distribution $\bar{\mu}_0$ and $\bar{\nu}_0$, respectively. The dimension-free constants c and M_1 depend on κ, γ, u, on the largest eigenvalue of K and on properties of g. Note that the additional constant M_1 in the second bound measures the difference between the distance ρ and the Euclidean distance.

These bounds are established using a modification of the coupling introduced in [85], which is a synchronous coupling for large distances and mainly a reflection coupling for small distances except on one line the noise is synchronized. In this work, we adjust the transition from synchronous coupling for large distances to reflection coupling for small distances to suit the underlying distance function. Namely, the synchronous coupling is applied when r_l is considered and the coupling approach of [85] when r_s is considered.

This approach which does not rely on a Lyapunov function has the advantage that the upper bound in (C.4) depends only on the Wasserstein distance between the two initial distributions and is independent of the two distributions themselves (cf. [85, 122, 97]). Further, the metric r_l is chosen such that the rate of the contraction result for large distances is optimized up to a constant. We emphasize that these bounds give also global contractivity for the classical Langevin dynamics and improve the result obtained in [85].

Moreover, using the ansatz for large distances, we contribute to the analysis of the optimal contraction rate for strongly convex potentials and improve the results of [60]. If the drift corresponds to a κ-strongly convex potential, we can split V in a linear part $x \cdot (Kx)$, where K is a positive definite matrix with smallest eigenvalue κ, and a convex function G with L_G Lipschitz continuous gradients. We prove contraction in Wasserstein distance with respect to a distance function of the same form as r_l with rate $c = \gamma/2 \min(1/4, \kappa u \gamma^{-2})$ provided $L_G u \gamma^{-2} \leq 3/4$ holds. If the perturbation G is sufficiently small, i.e., $L_G \leq 3\kappa$, we obtain for optimized γ a rate of order $\mathcal{O}(\sqrt{\kappa})$, that coincides with the order given in the L^2 contraction result in [40], and otherwise we obtain a rate of the same order as in [60].

Finally, applying a componentwise version of the preceding coupling we establish a uniform in time propagation of chaos bound for the corresponding particle system (C.3), i.e., we show for a probability measure μ_0 on $\mathbb{R}^{2d}$ with finite second moment,

$$\mathcal{W}_{1,\ell^1_N}(\bar{\mu}_t^{\otimes N}, \mu_t^N) \leq C_1 c^{-1} N^{-1/2},$$

where μ_t^N is the law of the particles driven by (C.3) with initial distribution $\mu_0^N = \mu_0^{\otimes N}$ and $\bar{\mu}_t^{\otimes N}$ is the product law of N independent solutions to (C.1) with initial distribution μ_0. Here, C_1 is a constant depending on κ, γ, u, d, on properties of g, and on the second moment of μ_0. The normalized ℓ^1-distance ℓ^1_N is given by

$$\ell^1_N((x,y),(\bar{x},\bar{y})) = N^{-1}\sum_{i=1}^N (|x^i - \bar{x}^i| + |y^i - \bar{y}^i|), \qquad \text{for all } x, y, \bar{x}, \bar{y} \in \mathbb{R}^{Nd}, \tag{C.5}$$

where $|\cdot|$ denotes the Euclidean metric.

Eventually, we note that the construction of the metric for large distance can be applied to prove contraction to specific unconfined cases, where $b \equiv 0$ and $\tilde{b}$ is a small perturbation of a linear force.

Notation: For some space $\mathbb{X}$, which is here either $\mathbb{R}^{2d}$ or $\mathbb{R}^{2Nd}$, we denote its Borel σ-algebra by $\mathcal{B}(\mathbb{X})$. The space of all probability measures on $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ is denoted by $\mathcal{P}(\mathbb{X})$. Let $\mu, \nu \in \mathcal{P}(\mathbb{X})$. A coupling ω of μ and ν is a probability measure on $(\mathbb{X} \times \mathbb{X}, \mathcal{B}(\mathbb{X}) \otimes \mathcal{B}(\mathbb{X}))$ with marginals μ and ν. The L^p Wasserstein distance with respect to a distance function $d : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ is defined by

$$\mathcal{W}_{p,d}(\mu, \nu) = \inf_{\omega \in \Pi(\mu,\nu)} \Big(\int_{\mathbb{X}\times\mathbb{X}} d(x,y)^p \omega(\mathrm{d}x\mathrm{d}y) \Big)^{1/p},$$

where $\Pi(\mu, \nu)$ denotes the set of all couplings of μ and ν. We write $\mathcal{W}_p$ if the underlying distance function is the Euclidean distance.

Outline of the paper: In Appendix C.2, we state the contraction results for the classical Langevin dynamics and give an informal construction of the coupling and the metric. In Appendix C.3, we state the framework and the contraction results for Langevin dynamics of McKean-Vlasov type before defining rigorously the metric and the coupling approach in Appendix C.4. Uniform in time propagation of chaos is established in Appendix C.5. The proofs are postponed to Appendix C.6.

C.2 Contraction for classical Langevin dynamics

C.2.1 Contraction for Langevin dynamics with strongly convex confinement potential

First, we consider the Langevin dynamics without a non-linear drift and with confinement potential V given by the stochastic differential equation

$$\begin{cases} \mathrm{d}X_t = Y_t \mathrm{d}t, \\ \mathrm{d}Y_t = (-\gamma Y_t - u\nabla \mathrm{V}(X_t))\mathrm{d}t + \sqrt{2\gamma u}\mathrm{d}B_t, \end{cases} \tag{C.6}$$

with initial condition $(X_0, Y_0) = (x, y) \in \mathbb{R}^{2d}$ and with d-dimensional standard Brownian motion $(B_t)_{t\geq 0}$. We impose for $\mathrm{V} \in \mathcal{C}^2(\mathbb{R}^d)$:

Assumption C.1. *There exist a positive definite matrix $K \in \mathbb{R}^{d\times d}$ with smallest eigenvalue $\kappa > 0$ and a convex function $G : \mathbb{R}^d \to \mathbb{R}$ with L_G-Lipschitz continuous gradients, i.e.,*

$$\begin{aligned} &\langle \nabla G(x) - \nabla G(\bar{x}), x - \bar{x}\rangle \geq 0 && \text{and} \\ &|\nabla G(x) - \nabla G(\bar{x})| \leq L_G |x - \bar{x}| && \text{for all } x, \bar{x} \in \mathbb{R}^d, \end{aligned} \tag{C.7}$$

such that

$$\mathrm{V}(x) = x \cdot (Kx)/2 + G(x) \qquad \text{for any } x \in \mathbb{R}^d.$$

We note that Assumption C.1 is satisfied for all κ-strongly convex functions V with L_V-Lipschitz continuous gradients, i.e.,

$$\begin{aligned} &\langle \nabla \mathrm{V}(x) - \nabla \mathrm{V}(y), x - y\rangle \geq \kappa |x - y|^2 && \text{and} \\ &|\nabla \mathrm{V}(x) - \nabla \mathrm{V}(y)| \leq L_V |x - y| && \text{for all } x, y \in \mathbb{R}^d. \end{aligned}$$

Note that the splitting of V in K and G is in general not unique. A natural choice is given by $K = \kappa\mathrm{Id}$ and $G(x) = \mathrm{V}(x) - (\kappa/2)|x|^2$, where Id is the $d \times d$ identity matrix. As we see later, we often want a splitting of V such that the Lipschitz constant L_G is minimized.

We establish a global contraction result for (C.6) in L^p Wasserstein distance with respect to the distance function $r : \mathbb{R}^{2d} \times \mathbb{R}^{2d} \to [0, \infty)$ given by

$$r((x,y),(\bar{x},\bar{y})) = \gamma^{-2}u(x-\bar{x}) \cdot (K(x-\bar{x})) + \frac{1}{2}|(1-2\lambda)(x-\bar{x}) + \gamma^{-1}(y-\bar{y})|^2 + \frac{1}{2}\gamma^{-2}|y-\bar{y}|^2 \tag{C.8}$$

for $(x,y),(\bar{x},\bar{y}) \in \mathbb{R}^{2d}$ with

$$\lambda = \min(1/8, \kappa u \gamma^{-2}). \tag{C.9}$$

Theorem C.1 (Contractivity for strongly convex potentials)**.** *For $t \geq 0$, let μ_t and ν_t be the law at time t of the processes (X_t, Y_t) and (X'_t, Y'_t), respectively, where $(X_s, Y_s)_{s\geq 0}$ and $(X'_s, Y'_s)_{s\geq 0}$ are solutions to* (C.6) *with initial distributions μ_0 and ν_0 on $\mathbb{R}^{2d}$, respectively. Suppose Assumption C.1 holds and*

$$L_G u \gamma^{-2} \leq 3/4. \tag{C.10}$$

Then, for any $1 \leq p < \infty$

$$\mathcal{W}_{p,r}(\mu_t, \nu_t) \leq e^{-ct}\mathcal{W}_{p,r}(\mu_0, \nu_0) \qquad \text{and} \qquad \mathcal{W}_p(\mu_t, \nu_t) \leq M e^{-ct}\mathcal{W}_p(\mu_0, \nu_0),$$

where the contraction rate c is given by

$$c = \gamma\lambda = \min(\gamma/8, \kappa u \gamma^{-1}/2). \tag{C.11}$$

The constant M is given by

$$M = \sqrt{\max(uL_K + \gamma^2, 3/2)\max(1/(u\kappa), 2)}, \tag{C.12}$$

where L_K denotes the largest eigenvalue of K.

Proof. The proof is based on a synchronous coupling and is postponed to Appendix C.6.1. □

Remark C.2. If V is a quadratic function, then $L_G = 0$ and the restriction on γ vanishes. In this case, the L^2 spectral gap of the corresponding generator is given by

$$c_{\text{gap}} = (1 - \sqrt{(1-4\kappa u\gamma^{-2})^+})(\gamma/2),$$

cf., [165, Section 6.3]. More precisely, $c_{\text{gap}} = \gamma/2$ if $4\kappa u \gamma^{-2} \geq 1$, and $\kappa u \gamma^{-1} \leq c_{\text{gap}} \leq 2\kappa u \gamma^{-1}$ if $4\kappa u \gamma^{-2} < 1$. Hence, the contraction rate is of the same order as the spectral gap. In particular for $\gamma = 2\sqrt{\kappa u}$ the optimal contraction rate $c = \sqrt{\kappa u}/8$ is obtained. If $L_G \leq 3\kappa$, $\gamma = 2\sqrt{\kappa u}$ satisfies condition (C.10) and yields the optimal contraction rate of order $\mathcal{O}(\sqrt{\kappa})$. Otherwise, for $\gamma = \sqrt{(4/3)L_G u}$ the contraction rate is optimized and of order $\mathcal{O}(\kappa/\sqrt{L_G})$.

C.2.2 Framework for classical Langevin dynamics with general external forces

Next, we consider the classical Langevin dynamics $(X_t, Y_t)_{t\geq 0}$ with a general external drift given by the stochastic differential equation

$$\begin{cases} \mathrm{d}X_t = Y_t \mathrm{d}t, \\ \mathrm{d}Y_t = (-\gamma Y_t + ub(X_t))\mathrm{d}t + \sqrt{2\gamma u}\mathrm{d}B_t, \end{cases} \tag{C.13}$$

with initial condition $(X_0, Y_0) = (x, y) \in \mathbb{R}^{2d}$.

We impose the following assumption on the force b:

Assumption C.2. *The function $b : \mathbb{R}^d \to \mathbb{R}^d$ is Lipschitz continuous and there exist a positive definite matrix $K \in \mathbb{R}^{d\times d}$ with smallest eigenvalue $\kappa \in (0,\infty)$ and largest eigenvalue $L_K \in (0,\infty)$, a constant $R \in [0,\infty)$ and a function $g : \mathbb{R}^d \to \mathbb{R}^d$ with Lipschitz constant $L_g \in (0,\infty)$ such that*

$$b(x) = -Kx + g(x) \qquad \text{for all } x \in \mathbb{R}^d, \tag{C.14}$$

and

$$\langle g(x) - g(\bar{x}), x - \bar{x}\rangle \le 0 \qquad \text{for all } x, \bar{x} \in \mathbb{R}^d \text{ such that } |x-\bar{x}| \ge R. \tag{C.15}$$

Remark C.3. Suppose that $b = -\nabla V$ where V is a potential function with a L_V-Lipschitz continuous gradient and that is k-strongly convex outside a Euclidean ball of radius $\tilde{R}$, i.e.,

$$\langle \nabla \mathrm{V}(x) - \nabla \mathrm{V}(\bar{x}), x-\bar{x}\rangle \ge k|x-\bar{x}|^2 \qquad \text{for all } x,\bar{x} \in \mathbb{R}^d \text{ such that } |x|, |\bar{x}| \ge \tilde{R}.$$

Note that ∇V can be split in $\nabla \mathrm{V}(x) = kx + h(x)$ where $h : \mathbb{R}^d \to \mathbb{R}^d$ is an L_h-Lipschitz continuous function with $L_h \le L_V + k$ and $\langle h(x) - h(\bar{x}), x - \bar{x}\rangle \ge 0$ for all $x, \bar{x} \in \mathbb{R}^d$ such that $|x|, |\bar{x}| \ge \tilde{R}$. Then for $l \le \frac{1}{2}\min(1, \frac{L_h}{k})$, $b = -\nabla$V satisfies Assumption C.2 with $L_g \le L_V + (1-l)k$, $\kappa = (1-l)k \ge \max(\frac{1}{2}k, k - \frac{L_h}{2})$ and $R = 2\tilde{R}\frac{L_h}{lk}$.

Example C.4 (Double-well potential). For $\beta > 0$, we consider the double-well potential V $\in \mathcal{C}^1(\mathbb{R})$ defined by

$$\mathrm{V}(x) = \begin{cases} \beta\Big(\frac{|x|^4}{4} - \frac{|x|^2}{2}\Big) & \text{for } |x| \le 2, \\ \beta\Big(\frac{3|x|^2}{2} - 4\Big) & \text{for } |x| > 2. \end{cases} \tag{C.16}$$

This potential has a Lipschitz continuous gradient and is strongly convex with convexity constant $k = 3\beta$ outside a Euclidean ball with radius $\tilde{R} = 2$. We consider the splitting $-\nabla \mathrm{V}(x) = -\kappa x + g(x)$ with $\kappa = (2/3)k = 2\beta$ and

$$g(x) = \begin{cases} -\beta(x^3 - 3x) & \text{for } |x| \le 2, \\ -\beta x & \text{for } |x| > 2. \end{cases}$$

Then, the function g is Lipschitz continuous with Lipschitz constant $L_g = 9\beta$ and (C.15) is satisfied for sufficiently large R.

C.2.3 Construction of the metric and the coupling

We provide an informal construction of the coupling and the complementary metric. Given two Brownian motions $(B_t)_{t\ge0}$, $(B'_t)_{t\ge0}$ and $(x,y),(x',y') \in \mathbb{R}^{2d}$, let $((X_t,Y_t),(X'_t,Y'_t))_{t\ge0}$ be an arbitrary coupling of two solutions to (C.13). It holds for the difference process $(Z_t,W_t)_{t\ge0} = (X_t - X'_t, Y_t - Y'_t)_{t\ge0}$,

$$\begin{cases} \mathrm{d}Z_t = W_t \\ \mathrm{d}W_t = (-\gamma W_t + ub(X_t) - ub(X'_t))\mathrm{d}t + \sqrt{2\gamma u}\,\mathrm{d}(B_t - B'_t). \end{cases}$$

Adapting the idea of the coupling construction from [85], the process $(Z_t,Q_t)_{t\ge0} = (Z_t, Z_t + \gamma^{-1}W_t)_{t\ge0}$ satisfies the stochastic differential equation

$$\begin{cases} \mathrm{d}Z_t = -\gamma Z_t\mathrm{d}t + \gamma Q_t\mathrm{d}t \\ \mathrm{d}Q_t = \gamma^{-1}u(b(X_t) - b(X'_t))\mathrm{d}t + \sqrt{2\gamma^{-1}u}\,\mathrm{d}(B_t - B'_t). \end{cases} \tag{C.17}$$

As in [85], we apply a synchronous coupling for $Q_t = 0$, since in this case the first equation of (C.17) is contractive and the absence of the noise ensures that the dynamics is not driven away from this area by random fluctuations. Apart from $Q_t = 0$, we want to apply a reflection coupling, which guarantees that the dynamics returns to the line $Q_t = 0$. Note that this construction leads to a coupling that is sticky on the hyperplane $\{((x,y),(x',y')) \in \mathbb{R}^{4d} : x - x' + \gamma^{-1}(y-y') = 0\}$. However, since it is technically hard to construct this sticky coupling, we consider approximations of the coupling, which are rigorously stated in Appendix C.4.2 and which suffice for our purpose. Similarly as in [85], we show for $r_s(t) = \alpha|Z_t| + |Q_t| < R_1$ with appropriately chosen constants α, R_1 that there exists a concave increasing function f depending on α and R_1 such that $f(r_s(t))$ is contractive on average. Note that the application of a concave function has the effect that a decrease in r_s has a larger impact than an increase in r_s.

On the other hand, if the difference process $(Z_t, W_t)_{t\geq 0}$ is sufficiently far away from the origin, we obtain under Assumption C.2 for the force b contractivity for the process $r_l(t) = (\gamma^{-2}uZ_t \cdot (KZ_t) + (1/2)|(1-2\tau)Z_t + \gamma^{-1}W_t|^2 + (1/2)|\gamma^{-1}W_t|^2)^{1/2}$, where $\tau > 0$ is a constant depending on κ, γ, u and L_g. More precisely, we obtain local contractivity with contraction rate $\gamma\tau$ for $r_l(t)^2 > \mathcal{R}$ for some $\mathcal{R} > 0$ depending on R, κ, γ, u and L_g. The process $r_l(t)$ is designed such that the local contraction rate is optimized up to some constant, see Lemma C.19.

We construct a metric which is globally contractive on average using the previously established coupling. The key idea lies in combining r_s and r_l in such a way, that the two local contraction results imply global contractivity in the new metric. Note that for simplicity, we write r_l and r_s both for the norm $r_l(z,w)$ (respectively $r_s(z,w)$) of $(z,w) \in \mathbb{R}^{2d}$ and for the distance $r_l((x,y),(x',y'))$ (respectively $r_s((x,y),(x',y'))$) of $(x,y),(x',y') \in \mathbb{R}^{2d}$.

As we see in Appendix C.6.2, the lower bound $\mathcal{R}$ in the contraction result for large distances is fixed due to the dependence on the drift assumptions, whereas the upper bound R_1 in the result for small distances is flexible with the drawback that the contraction rate gets smaller for larger R_1. To benefit from the local contraction results, we want for all $(z,w) \in \mathbb{R}^{2d}$ that $r_s(z,w) \leq R_1$ or $r_l(z,w)^2 \geq \mathcal{R}$ holds, which we achieve by choosing R_1 sufficiently large. We construct a continuous transition between r_s and r_l by considering $r_s \wedge (D_{\mathcal{K}} + \epsilon r_l)$, where the constant ϵ satisfies $2\epsilon r_l \leq r_s$ and the constant $D_{\mathcal{K}}$ is given such that $r_s(z,w) \wedge (D_{\mathcal{K}} + \epsilon r_l(z,w)) = r_s(z,w)$ for (z,w) with $r_l(z,w)^2 \leq \mathcal{R}$. Then, we set R_1 such that $r_s(z,w) \wedge (D_{\mathcal{K}} + \epsilon r_l(z,w)) = D_{\mathcal{K}} + \epsilon r_l(z,w)$ for (z,w) with $r_s(z,w) \leq R_1$ is guaranteed.

In particular, in this construction the level set $r_s(z,w) - \epsilon r_l(z,w) = D_{\mathcal{K}}$ is optimally encompassed by the level set $r_s(z,w) = R_1$ and $r_l(z,w)^2 = \mathcal{R}$, as illustrated in Figure C.1, and $r_s(z,w) \leq R_1$ or $r_l(z,w)^2 \geq \mathcal{R}$ is ensured. We define the metric $\rho((x,y),(x',y')) = f(r_s((x,y),(x',y')) \wedge \{D_{\mathcal{K}} + \epsilon r_l((x,y),(x',y'))\})$. As illustrated in Figure C.2, we obtain $f(r_s)$ for small distances and $f(D_{\mathcal{K}} + \epsilon r_l((x,y),(x',y')))$ for large distances. A detailed rigorous construction and a proof showing that ρ defines a metric are given in Appendix C.4.

C.2.4 A global contraction result for the classical Langevin dynamics with general external force

We establish the main contraction result for the classical Langevin dynamics given by (C.13).

Theorem C.5. *For $t \geq 0$, let μ_t and ν_t be the law at time t of the processes (X_t, Y_t) and (X'_t, Y'_t), respectively, where $(X_s, Y_s)_{s\geq 0}$ and $(X'_s, Y'_s)_{s\geq 0}$ are solutions to* (C.13) *with initial distributions μ_0 and ν_0 on $\mathbb{R}^{2d}$, respectively. Suppose Assumption C.2 holds and*

$$L_g u \gamma^{-2} < \frac{\kappa}{2L_g}. \tag{C.18}$$

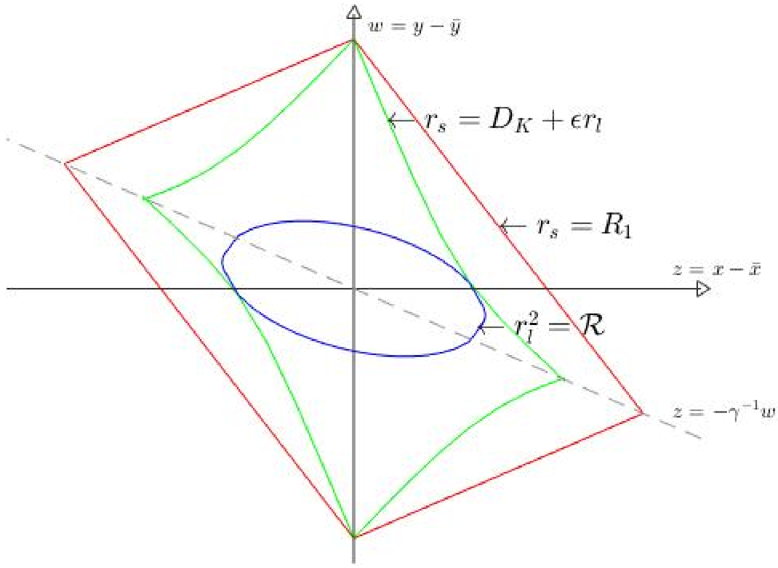

Figure C.1: Level sets of the metrics r_l and r_s.

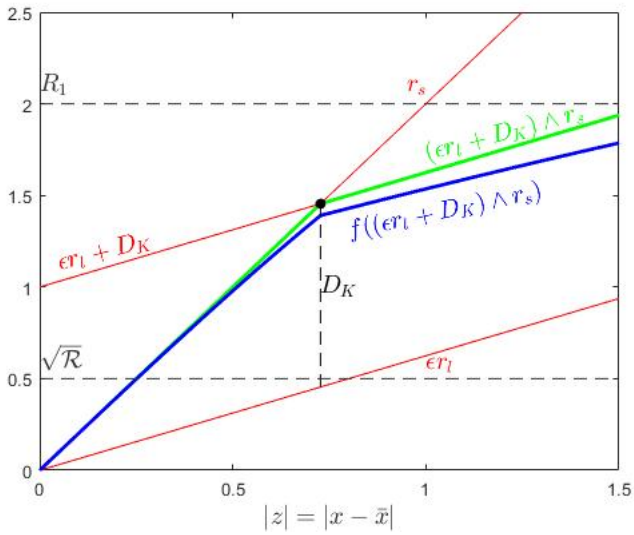

Figure C.2: Sketch of the metric construction $f((\epsilon r_l + D_\mathcal{K}) \wedge r_s)$. Here the metric is evaluated for $z = -\gamma^{-1}w$ (i.e., along the dashed line in Figure C.1).

Then,

$$\mathcal{W}_{1,\rho}(\mu_t,\nu_t) \le e^{-ct}\mathcal{W}_{1,\rho}(\mu_0,\nu_0) \quad \textit{and} \quad \mathcal{W}_1(\mu_t,\nu_t) \le M_1 e^{-ct}\mathcal{W}_1(\mu_0,\nu_0),$$

where the distance ρ *is defined precisely in* (C.35) *below and the contraction rate* c *is given by*

$$c = \gamma \exp(-\Lambda) \min\Big(\frac{(L_K+L_g)u\gamma^{-2}}{4}\Lambda^{1/2}, \frac{1}{8}\Lambda^{1/2}, \frac{\tau E}{2}\Big) \quad \textit{with} \tag{C.19}$$

$$\Lambda = \frac{L_K+L_g}{4}R_1^2, \qquad \textit{and}, \tag{C.20}$$

$$\tau := \min(1/8, \gamma^{-2}u\kappa/2 - \gamma^{-4}L_g^2u^2), \qquad \textit{and}, \tag{C.21}$$

$$E := \frac{1}{6}\min\Big(1, \frac{\sqrt{\kappa}\gamma}{\sqrt{8u}(L_K+L_g)}, \sqrt{\frac{\kappa u}{2}}\gamma^{-1}, 2(L_K+L_g)u\gamma^{-2}\Big). \tag{C.22}$$

The constants R_1 *satisfies*

$$\begin{aligned}\frac{2}{3}\min(1, 2(L_K+L_g)u\gamma^{-2})\sqrt{\frac{8u\mathbb{1}_{\{R>0\}}+L_guR^2}{\tau\gamma^2}} &\le R_1 \\ &\le 4\max\Big(\frac{\sqrt{8}(L_K+L_g)u}{\gamma\sqrt{\kappa}}, 1\Big)\sqrt{\frac{8u\mathbb{1}_{\{R>0\}}+L_guR^2}{\tau\gamma^2}},\end{aligned} \tag{C.23}$$

and is explicitly stated in (C.38)*. The constant* M_1 *is given by*

$$M_1 = \max(2(L_K+L_g)u\gamma^{-1}+\gamma, 1)\frac{1}{2}\exp(\Lambda)\max\Big(3, \frac{3\gamma^2}{2(L_K+L_g)u}\Big)\max(\sqrt{2/(\kappa u)}, 2). \tag{C.24}$$

Proof. The proof is postponed to Appendix C.6.2. □

Remark C.6. Compared to the contraction result obtained in [85, Theorem 2.3], global contractivity in Wasserstein distance is obtained with rate c given in (C.19) which is independent of the dimension d.

Remark C.7 (Kinetic behaviour). If γ is chosen such that $\kappa u\gamma^{-2}$, $L_gu\gamma^{-2}$ and $L_Ku\gamma^{-2}$ are fixed and further L_KR^2 and L_gR^2 are fixed, we obtain similarly to [85, Corollary 2.9] that the contraction rate is of order $\Omega(R^{-1})$.

Remark C.8. If $R=0$, the metric ρ defined in (C.35) reduces to $\rho((x,y),(\bar{x},\bar{y})) = (\gamma^{-2}(x-\bar{x})\cdot(K(x-\bar{x})) + (1/2)|(1-2\tau)(x-\bar{x})+\gamma^{-1}(y-\bar{y})|^2 + (1/2)|\gamma^{-1}(y-\bar{y})|^2)^{1/2}$ and the coupling given in Appendix C.4.2 becomes the synchronous coupling. This metric differs from r defined in (C.8) by the constant τ, since here the drift b is not necessarily of gradient-type and we can not make use of the co-coercivity property as in the proof of Theorem C.1. Following the proof given in Appendix C.6.2, we obtain contraction in L^1 Wasserstein distance, with contraction rate $c = \min(\gamma/16, \kappa\gamma^{-1}/4 - 8\gamma^{-3}L_g^2u^2)$. We remark that the constant E vanishes in the contraction rate, which measures the difference between the two metrics that are considered in general for ρ. If $L_g \le \sqrt{2}\kappa$, the contraction rate is maximized for $\gamma = u^{1/2}(2\kappa + (4\kappa^2 - 8L_g^2)^{1/2})^{1/2}$ and satisfies $c = u^{1/2}(2\kappa + (4\kappa^2 - 8L_g^2)^{1/2})^{1/2}/16$, i.e., in this case the rate is of order $\mathcal{O}(\sqrt{\kappa})$.

Example C.9 (Double-well potential)*.* For the model given in Example C.4, we obtain contraction with respect to the designed Wasserstein distance if $\gamma > 9\sqrt{\beta}$ is satisfied.

C.3 Contraction for nonlinear Langevin dynamics of McKean-Vlasov type

Consider the Langevin dynamics of McKean-Vlasov type given in (C.1). We require Assumption C.2 for the function $b : \mathbb{R}^d \to \mathbb{R}^d$. For the function $\tilde{b} : \mathbb{R}^{2d} \to \mathbb{R}^d$ we impose:

Assumption C.3. *The function $\tilde{b} : \mathbb{R}^{2d} \to \mathbb{R}^d$ is $\tilde{L}$-Lipschitz continuous.*

Example C.10 (Quadratic interaction potential). Consider $\tilde{b}(x, y) = ky$ with $k \in \mathbb{R}$. Then $\tilde{L} = |k|$ and $\tilde{b}$ corresponds to the interaction potential $\mathrm{W}(x, y) = -kx \cdot y$. This potential is attractive for $k > 0$ and repulsive for $k < 0$.

Example C.11 (Mollified Coulomb, Newtonian and logarithmic potentials). The gradients of the Coulomb potential and of the Newtonian potential, which describe charged and self-gravitating particles [36], are not Lipschitz continuous. However, the gradient of a mollified version (see [97]) given by

$$\mathrm{W}(x, y) = \frac{\tilde{k}}{(|x-y|^p + q^p)^{1/p}} \qquad \text{for } p \geq 2, q \in \mathbb{R}_+ \text{ and } \tilde{k} \in \mathbb{R}$$

satisfies Assumption C.3, since $\|\text{Hess } W\| < \infty$, and therefore $\nabla_x \mathrm{W}$ is Lipschitz continuous. In the same line, the gradient of the mollified version of the logarithmic potential given by

$$\mathrm{W}(x, y) = -2\log((|x-y|^p + q^p)^{1/p}) \qquad \text{for } p \geq 2, q \in \mathbb{R}_+$$

satisfies Assumption C.3.

Under the above conditions, we establish contraction in an L^1 Wasserstein distance.

Theorem C.12 (Contraction for nonlinear Langevin dynamics)**.** *Let $\bar{\mu}_0$ and $\bar{\nu}_0$ be two probability distributions on $\mathbb{R}^{2d}$ with finite second moment. For $t \geq 0$, let $\bar{\mu}_t$ and $\bar{\nu}_t$ be the law at time t of the processes $(\bar{X}_t, \bar{Y}_t)$ and $(\bar{X}'_t, \bar{Y}'_t)$, respectively, where $(\bar{X}_s, \bar{Y}_s)_{s\geq 0}$ and $(\bar{X}'_s, \bar{Y}'_s)_{s\geq 0}$ are solutions to* (C.1) *with initial distribution $\bar{\mu}_0$ and $\bar{\nu}_0$, respectively. Suppose Assumption C.2, Assumption C.3 and* (C.18) *hold. Let $\tilde{L}$ satisfy*

$$\tilde{L} \leq \exp(-\Lambda) \min\Big\{\frac{\gamma\tau}{12}\sqrt{\frac{\kappa}{u}} \min(1, 2(L_K + L_g)u\gamma^{-2}), \frac{L_K + L_g}{4}\Big\}, \tag{C.25}$$

where Λ and τ are given in (C.20) *and* (C.21)*, respectively. Then*

$$\mathcal{W}_{1,\rho}(\bar{\mu}_t, \bar{\nu}_t) \leq e^{-\bar{c}t}\mathcal{W}_{1,\rho}(\bar{\mu}_0, \bar{\nu}_0) \quad \text{and} \quad \mathcal{W}_1(\bar{\mu}_t, \bar{\nu}_t) \leq M_1 e^{-\bar{c}t}\mathcal{W}_1(\bar{\mu}_0, \bar{\nu}_0),$$

where the distance ρ is given in (C.35) *and $\bar{c} = c/2$ with c given in* (C.19)*. The constant M_1 is given in* (C.24)*. Moreover, there exists a unique invariant probability measure $\bar{\mu}_\infty$ for* (C.1) *and convergence in L^1 Wasserstein distance to $\bar{\mu}_\infty$ holds.*

Proof. The proof is based on the coupling approach and the metric construction given in Appendix C.4.1 and Appendix C.4.2, respectively, and is postponed to Appendix C.6.2. □

Remark C.13. In comparison to [97, Theorem 3.1], global contractivity is established with a contraction rate and a restriction on the Lipschitz constant $\tilde{L}$ that are independent of the dimension d.

Remark C.14. Compared to the contraction result in Theorem C.5 for classical Langevin dynamics, the contraction rate deteriorates by a factor of 2 to compensate for the nonlinear interaction terms.

If $R = 0$, (C.25) reduces to $\tilde{L} \leq \tau\gamma\sqrt{\kappa/u}/8$ and contraction holds with rate $\bar{c} = \min(\gamma/32, \kappa u\gamma^{-1}/8 - L_g^2 u^2\gamma^{-3}/2)$ by Lemma C.19 and (C.67). If $L_g \leq \sqrt{2}\kappa$, the contraction rate is maximized for $\gamma = \sqrt{u}(2\kappa + (4\kappa^2 - 8L_g^2)^{1/2})^{1/2}$ yielding $\bar{c} = \sqrt{u}(2\kappa + (4\kappa^2 - 8L_g^2)^{1/2})^{1/2}/32$. If the drift is additionally of gradient-type, we can adapt the proof of Theorem C.1 and use the co-coercivity property to obtain a contraction rate of order $\mathcal{O}(\sqrt{\kappa})$ for $L_g \leq 3\kappa$ and a rate of order $\mathcal{O}(\kappa/\sqrt{L_g})$ for $L_g > 3\kappa$.

Remark C.15. The contraction results can be extended to unconfined Langevin dynamics. Consider $b \equiv 0$ and $\tilde{b} : \mathbb{R}^{2d} \to \mathbb{R}^d$ given by $\tilde{b}(x, y) = -\tilde{K}(x-y)+\tilde{g}(x-y)$ where $\tilde{K} \in \mathbb{R}^{d\times d}$ is a positive definite matrix with smallest eigenvalue $\tilde{\kappa}$ and where $\tilde{g} : \mathbb{R}^d \to \mathbb{R}^d$ is an anti-symmetric, $L_{\tilde{g}}$-Lipschitz continuous function $\tilde{g} : \mathbb{R}^d \to \mathbb{R}^d$. If $L_{\tilde{g}} \leq (\gamma/2)\sqrt{\tilde{\kappa}/u}\min(1/8, \tilde{\kappa}u\gamma^{-2}/2)$, contraction in an L^1 Wasserstein distance can be shown via a synchronous coupling approach. The underlying distance function in the Wasserstein distance is based on a similar twisted 2-norm structure as the distance r_l given in (C.26). We note that the conditions on L_g and $\tilde{L}$ are combined in the restrictive condition on $L_{\tilde{g}}$, which implies $L_{\tilde{g}} \leq \tilde{\kappa}/8$ and which gives only contraction for small perturbations of linear interaction forces. A detailed analysis of the unconfined dynamics is given in Appendix C.7.

C.4 Metric and coupling

C.4.1 Metric construction

For both the classical Langevin dynamics and the nonlinear Langevin dynamics, i.e., when Assumption C.2 holds, we consider the metrics $r_l, r_s : \mathbb{R}^{2d} \times \mathbb{R}^{2d} \to [0, \infty)$ given by

$$\begin{aligned} r_l((x, y), (\bar{x}, \bar{y}))^2 &:= \frac{u}{\gamma^2}(x - \bar{x}) \cdot (K(x - \bar{x})) + \frac{(1 - 2\tau)^2}{2}|x - \bar{x}|^2 + \gamma^{-1}(1 - 2\tau)(x - \bar{x})(y - \bar{y}) \\ &\quad + \gamma^{-2}|y - \bar{y}|^2 \\ &= \gamma^{-2}u(x - \bar{x}) \cdot (K(x - \bar{x})) + \frac{1}{2}|(1 - 2\tau)(x - \bar{x}) + \gamma^{-1}(y - \bar{y})|^2 + \frac{1}{2}\gamma^{-2}|y - \bar{y}|^2, \end{aligned} \tag{C.26}$$

and

$$r_s((x, y), (\bar{x}, \bar{y})) := \alpha|x - \bar{x}| + |x - \bar{x} + \gamma^{-1}(y - \bar{y})|, \tag{C.27}$$

for $(x, y), (\bar{x}, \bar{y}) \in \mathbb{R}^{2d}$, where the constants τ and α are given by (C.21) and

$$\alpha := 2(L_K + L_g)u\gamma^{-2}, \tag{C.28}$$

respectively. Next, we state the rigorous construction of the metric $\rho : \mathbb{R}^{2d} \times \mathbb{R}^{2d} \to [0, \infty)$, that is applied in Theorem C.5 and Theorem C.12, and that is glued together of r_l and r_s in an appropriate way. Note that r_l and r_s are equivalent metrics. More precisely, for all $(x, y), (\bar{x}, \bar{y}) \in \mathbb{R}^{2d}$ it holds $2\epsilon r_l((x, y), (\bar{x}, \bar{y})) \leq r_s((x, y), (\bar{x}, \bar{y}))$ with

$$\epsilon = (1/2)\min(1, (2/3)\alpha/(\sqrt{L_K u}\gamma^{-1}), \alpha). \tag{C.29}$$

Indeed, for $(z,w)=(x-\bar{x},y-\bar{y})$

$$r_l^2((x,y),(\bar{x},\bar{y}))\le L_K\gamma^{-2}u|z|^2+\frac{1}{2}|z+\gamma^{-1}w|^2+2\tau|z||z+\gamma w|+2\tau^2|z|^2+\frac{1}{2}|\gamma^{-1}w|^2 \qquad \text{and}$$
$$\begin{aligned}r_s^2((x,y),(\bar{x},\bar{y}))&\ge\frac{1}{2}(\alpha|z|+|z+\gamma^{-1}w|)^2+\frac{1}{2}\min(\alpha^2,1)\gamma^{-2}|w|^2\\&\ge\frac{\alpha^2}{2}|z|^2+\alpha|z||z+\gamma^{-1}w|+\frac{1}{2}|z+\gamma^{-1}w|^2+\frac{1}{2}\min(1,\alpha^2)\gamma^{-2}|w|^2,\end{aligned}$$

and

$$4\epsilon^2\le\min\Big(\frac{\alpha^2}{2(L_Ku\gamma^{-2}+2\tau L_Ku\gamma^{-2}/2)},1,\alpha^2\Big)\le\min\Big(\frac{\alpha^2}{2(L_Ku\gamma^{-2}+2\tau^2)},1,\frac{\alpha}{2\tau},\alpha^2\Big),$$

since $\alpha>\kappa\gamma^{-2}$ and $\tau\le\min(1/8,L_K\gamma^{-2}u/2)$ by (C.28) and (C.21). Further, for all $(x,y),(\bar{x},\bar{y})\in\mathbb{R}^{2d}$ it holds $\mathcal{E}r_s((x,y),(\bar{x},\bar{y}))\le r_l((x,y),(\bar{x},\bar{y}))$ with

$$\mathcal{E}=\min(\sqrt{\kappa u}\gamma^{-1}/(\sqrt{8}\alpha),1/2), \tag{C.30}$$

since

$$\frac{r_l(t)}{r_s(t)}\ge\Big(\frac{\kappa u\gamma^{-2}|\bar{Z}_t|^2+(1/2)|(1-2\tau)\bar{Z}_t+\gamma^{-1}\bar{W}_t|^2}{2(a+2\tau)^2|\bar{Z}_t|^2+2|(1-2\tau)\bar{Z}_t+\gamma^{-1}\bar{W}_t|^2}\Big)^{1/2}\ge\min\Big(\frac{\sqrt{\kappa u}\gamma^{-1}}{\sqrt{8}\alpha},\frac{1}{2}\Big).$$

Define

$$\Delta((x,y),(\bar{x},\bar{y})):=r_s((x,y),(\bar{x},\bar{y}))-\epsilon r_l((x,y),(\bar{x},\bar{y})) \tag{C.31}$$

for $(x,y),(\bar{x},\bar{y})\in\mathbb{R}^{2d}$ and

$$D_{\mathcal{K}}:=\sup_{((x,y),(\bar{x},\bar{y}))\in\mathbb{R}^{4d}:(x-\bar{x},y-\bar{y})\in\mathcal{K}}\Delta((x,y),(\bar{x},\bar{y})), \tag{C.32}$$

where the compact set $\mathcal{K}\subset\mathbb{R}^{2d}$ is given by

$$\mathcal{K}:=\{(z,w)\in\mathbb{R}^{2d}:\gamma^{-2}uz\cdot(Kz)+(1/2)|(1-2\tau)z+\gamma^{-1}w|^2+(1/2)|\gamma^{-1}w|^2\le\mathcal{R}\}. \tag{C.33}$$

with

$$\mathcal{R}=(1/\tau)(8u\mathbb{1}_{\{R>0\}}+L_guR^2)\gamma^{-2}. \tag{C.34}$$

We define the metric $\rho:\mathbb{R}^{2d}\times\mathbb{R}^{2d}\to[0,\infty)$ by

$$\rho((x,y),(\bar{x},\bar{y})):=f((\Delta((x,y),(\bar{x},\bar{y}))\wedge D_{\mathcal{K}})+\epsilon r_l((x,y),(\bar{x},\bar{y}))) \tag{C.35}$$

for $(x,y),(\bar{x},\bar{y})\in\mathbb{R}^{2d}$, where Δ and $D_{\mathcal{K}}$ are given in (C.31) and (C.32). The function f is an increasing concave function defined by

$$f(r):=\int_0^r\phi(s)\psi(s)\mathrm{d}s, \tag{C.36}$$

where

$$\begin{aligned}\phi(s)&:=\exp\Big(-\frac{\alpha\gamma^2}{4u}\frac{(s\wedge R_1)^2}{2}\Big), & \Phi(s)&=\int_0^s\phi(x)\mathrm{d}x,\\ \psi(s)&:=1-\frac{\hat{c}}{2}\gamma u^{-1}\int_0^{s\wedge R_1}\Phi(x)\phi(x)^{-1}\mathrm{d}x, & \hat{c}&=\frac{1}{u^{-1}\gamma\int_0^{R_1}\Phi(s)\phi(s)^{-1}\mathrm{d}s},\end{aligned} \tag{C.37}$$

and where R_1 is given by

$$R_1 := \sup_{((x,y),(\bar{x},\bar{y})):\Delta((x,y),(\bar{x},\bar{y}))\le D_{\mathcal{K}}} r_s(((x,y),(\bar{x},\bar{y}))). \tag{C.38}$$

The construction of the function f is adapted from [83]. Since $\psi(s) \in [1/2, 1]$, it holds for $r \geq 0$

$$f'(R_1)r = (\phi(R_1)/2)r \le \Phi(r)/2 \le f(r) \le \Phi(r) \le r. \tag{C.39}$$

Note that the constant R_1 is finite and $R_1 \le \sup_{\Delta((x,y),(\bar{x},\bar{y}))\le D_{\mathcal{K}}} 2\Delta((x,y),(\bar{x},\bar{y})) \le 2D_{\mathcal{K}}$ holds, since $\Delta((x,y),(\bar{x},\bar{y})) = r_s((x,y),(\bar{x},\bar{y})) - \epsilon r_l((x,y),(\bar{x},\bar{y})) \ge (1/2)r_s((x,y),(\bar{x},\bar{y}))$ for any $(x,y),(\bar{x},\bar{y}) \in \mathbb{R}^{2d}$ by (C.29). Hence, $\hat{c}$ given in (C.37) and f are well-defined. Further,

$$R_1 \le 2D_{\mathcal{K}} \le 2 \sup_{((x,y),(\bar{x},\bar{y}))\in\mathbb{R}^{4d}:(x-\bar{x},y-\bar{y})\in\mathcal{K}} (\mathcal{E}^{-1} - 2\epsilon)r_l((x,y),(\bar{x},\bar{y})) \le 2(\mathcal{E}^{-1} - 2\epsilon)\sqrt{\mathcal{R}}.$$

The constant R_1 is also bounded from below by

$$R_1 \ge \sup_{((x,y),(\bar{x},\bar{y})):\Delta((x,y),(\bar{x},\bar{y}))\le D_{\mathcal{K}}} 2\epsilon r_l(((x,y),(\bar{x},\bar{y}))) \ge 2\epsilon\sqrt{\mathcal{R}},$$

since $\Delta((x,y),(\bar{x},\bar{y})) \le D_{\mathcal{K}}$ for all $(x,y),(\bar{x},\bar{y}) \in \mathbb{R}^{2d}$ such that $r_l((x,y),(\bar{x},\bar{y}))^2 = \mathcal{R}$. By (C.34), (C.29), (C.30), the two bounds on R_1 imply the relation (C.23) of R and R_1 given in Theorem C.5.

By this construction for the metric ρ, it holds $(\Delta((x,y),(\bar{x},\bar{y})) \wedge D_{\mathcal{K}}) + \epsilon r_l((x,y),(\bar{x},\bar{y})) = r_s((x,y),(\bar{x},\bar{y}))$ for $\Delta((x,y),(\bar{x},\bar{y})) \le D_{\mathcal{K}}$, and in particular for $r_l((x,y),(\bar{x},\bar{y}))^2 \le \mathcal{R}$. Further, $(\Delta((x,y),(\bar{x},\bar{y})) \wedge D_{\mathcal{K}}) + \epsilon r_l((x,y),(\bar{x},\bar{y})) = D_{\mathcal{K}} + \epsilon r_l((x,y),(\bar{x},\bar{y}))$ for $\Delta((x,y),(\bar{x},\bar{y})) > D_{\mathcal{K}}$ and in particular for $r_s((x,y),(\bar{x},\bar{y})) > R_1$.

If $R = 0$, then $\mathcal{K} = \{(0,0)\}$ and hence $D_{\mathcal{K}} = R_1 = 0$ and $f(r) = r$. In this case, we can omit the factor ϵ in (C.35) and (C.46) and set $\rho((x,y),(\bar{x},\bar{y})) = r_l((x,y),(\bar{x},\bar{y}))$ for simplicity.

Lemma C.16. *The function ρ given in* (C.35) *defines a metric on $\mathbb{R}^{2d}$ and is equivalent to the Euclidean distance on $\mathbb{R}^{2d}$.*

Proof. Symmetry and positive definiteness holds directly. Hence, ρ is a semimetric. To prove the triangle inequality, we note that for $(x,y),(\bar{x},\bar{y}),(\hat{x},\hat{y}) \in \mathbb{R}^{2d}$,

$$\begin{aligned}
&(\Delta((x,y),(\bar{x},\bar{y})) \wedge D_{\mathcal{K}}) + \epsilon r_l((x,y),(\bar{x},\bar{y})) \\
&\quad = r_s((x,y),(\bar{x},\bar{y})) \wedge (D_{\mathcal{K}} + \epsilon r_l((x,y),(\bar{x},\bar{y}))) \\
&\quad \le (r_s((x,y),(\hat{x},\hat{y})) + r_s((\hat{x},\hat{y}),(\bar{x},\bar{y}))) \wedge (D_{\mathcal{K}} + \epsilon r_l((x,y),(\hat{x},\hat{y})) + \epsilon r_l((\hat{x},\hat{y}),(\bar{x},\bar{y}))) \\
&\quad \le (r_s((x,y),(\hat{x},\hat{y})) + r_s((\hat{x},\hat{y}),(\bar{x},\bar{y}))) \wedge (D_{\mathcal{K}} + \epsilon r_l((x,y),(\hat{x},\hat{y})) + D_{\mathcal{K}} + \epsilon r_l((\hat{x},\hat{y}),(\bar{x},\bar{y}))) \\
&\qquad \wedge (D_{\mathcal{K}} + \epsilon r_l((x,y),(\hat{x},\hat{y})) + (1/2)r_s((\hat{x},\hat{y}),(\bar{x},\bar{y}))) \\
&\qquad \wedge (D_{\mathcal{K}} + (1/2)r_s((x,y),(\hat{x},\hat{y})) + \epsilon r_l((\hat{x},\hat{y}),(\bar{x},\bar{y}))) \\
&\quad \le (\Delta((x,y),(\bar{x},\bar{y})) \wedge D_{\mathcal{K}}) + \epsilon r_l((x,y),(\bar{x},\bar{y})) + (\Delta((x,y),(\hat{x},\hat{y})) \wedge D_{\mathcal{K}}) + \epsilon r_l((\hat{x},\hat{y}),(\bar{x},\bar{y})),
\end{aligned}$$

since r_l and r_s are metrics on $\mathbb{R}^{2d}$ and $\epsilon r_l((x,y),(\bar{x},\bar{y})) \le (1/2)r_s((x,y),(\bar{x},\bar{y}))$. Since f given in (C.36) is a concave function, $\rho((x,y),(\bar{x},\bar{y})) \le \rho((x,y),(\hat{x},\hat{y})) + \rho((\hat{x},\hat{y}),(\bar{x},\bar{y}))$ for $(x,y),(\bar{x},\bar{y}),(\hat{x},\hat{y}) \in \mathbb{R}^{2d}$. Hence, ρ defines a metric.

Further, it holds for all $(x,y),(\bar{x},\bar{y}) \in \mathbb{R}^{2d}$,

$$\begin{aligned}
\Delta((x,y),(\bar{x},\bar{y})) \wedge D_{\mathcal{K}} + \epsilon r_l((x,y),(\bar{x},\bar{y})) &\le r_s((x,y),(\bar{x},\bar{y})) \\
&\le \max(\alpha+1,\gamma^{-1})(|x-\bar{x}| + |y-\bar{y}|) \qquad \text{(C.40)} \\
&\le \max(\alpha+1,\gamma^{-1})\sqrt{2}|(x,y)-(\bar{x},\bar{y})|.
\end{aligned}$$

and

$$\begin{aligned}\Delta((x,y),(\bar{x},\bar{y}))\wedge D_{\mathcal{K}}+\epsilon r_l((x,y),(\bar{x},\bar{y}))&\geq \epsilon r_l((x,y),(\bar{x},\bar{y}))\\ &\geq \epsilon(\kappa u\gamma^{-2}|x-\bar{x}|^2+\frac{1}{2}\gamma^{-2}|y-\bar{y}|^2)^{1/2}\\ &\geq \epsilon\gamma^{-1}\min(\sqrt{\kappa u},1/\sqrt{2})|(x,y)-(\bar{x},\bar{y})|\\ &\geq \epsilon\gamma^{-1}\min(\sqrt{\kappa u/2},1/2)(|x-\bar{x}|+|y-\bar{y}|).\end{aligned}\tag{C.41}$$

Then, by (C.39),

$$\mathbf{C}_1|(x,y)-(\bar{x},\bar{y})|\leq \rho((x,y),(\bar{x},\bar{y}))\leq \mathbf{C}_2|(x,y)-(\bar{x},\bar{y})|\tag{C.42}$$

with $\mathbf{C}_1=f'(R_1)\epsilon\gamma^{-1}\min(\sqrt{\kappa u},1/\sqrt{2})$ and $\mathbf{C}_2=\sqrt{2}\max(\alpha+1,\gamma^{-1})$.

□

C.4.2 Coupling for Langevin dynamics

To prove Theorem C.5 and Theorem C.12 we construct a coupling of two solutions to (C.1). The construction is partially adapted from the coupling approach introduced in [85]. Recall that $\tilde{b}\equiv 0$ in Theorem C.5.

Let ξ be a positive constant, which we take finally to the limit $\xi\to 0$. Let $(B_t^{\mathrm{rc}})_{t\geq 0}$ and $(B_t^{\mathrm{sc}})_{t\geq 0}$ be two independent d-dimensional Brownian motions and let $\bar{\mu}_0,\bar{\nu}_0$ be two probability measures on $\mathbb{R}^{2d}$. The coupling $((\bar{X}_t,\bar{Y}_t),(\bar{X}_t',\bar{Y}_t'))_{t\geq 0}$ of two copies of solutions to (C.1) is a solution to the SDE on $\mathbb{R}^{2d}\times\mathbb{R}^{2d}$ given by

$$\begin{aligned}&\begin{cases}\mathrm{d}\bar{X}_t &=\bar{Y}_t\mathrm{d}t\\ \mathrm{d}\bar{Y}_t &=(-\gamma\bar{Y}_t+ub(\bar{X}_t)+u\int_{\mathbb{R}^d}\tilde{b}(\bar{X}_t,z)\bar{\mu}_t^x(\mathrm{d}z))\mathrm{d}t+\sqrt{2\gamma u}\mathrm{sc}(Z_t,W_t)\mathrm{d}B_t^{\mathrm{sc}}\\ &\quad+\sqrt{2\gamma u}\mathrm{rc}(Z_t,W_t)\mathrm{d}B_t^{\mathrm{rc}}\end{cases}\\ &\begin{cases}\mathrm{d}\bar{X}_t' &=\bar{Y}_t'\mathrm{d}t\\ \mathrm{d}\bar{Y}_t' &=(-\gamma\bar{Y}_t'+ub(\bar{X}_t')+u\int_{\mathbb{R}^d}\tilde{b}(\bar{X}_t',z)\bar{\nu}_t^x(\mathrm{d}z))\mathrm{d}t+\sqrt{2\gamma u}\mathrm{sc}(Z_t,W_t)\mathrm{d}B_t^{\mathrm{sc}}\\ &\quad+\sqrt{2\gamma u}\mathrm{rc}(Z_t,W_t)(\mathrm{Id}-2e_te_t^T)\mathrm{d}B_t^{\mathrm{rc}},\end{cases}\\ &(\bar{X}_0,\bar{Y}_0)\sim\bar{\mu}_0,\quad (\bar{X}_0',\bar{Y}_0')\sim\bar{\nu}_0,\end{aligned}\tag{C.43}$$

where $\bar{\mu}_t^x=\mathrm{Law}(\bar{X}_t)$ and $\bar{\nu}_t^x=\mathrm{Law}(\bar{X}_t')$. Further, $Z_t=\bar{X}_t-\bar{X}_t'$, $W_t=\bar{Y}_t-\bar{Y}_t'$, $Q_t=Z_t+\gamma^{-1}W_t$ and $e_t=Q_t/|Q_t|$ if $Q_t\neq 0$ and $e_t=0$ otherwise. The functions $\mathrm{rc},\mathrm{sc}:\mathbb{R}^{2d}\to[0,1)$ are Lipschitz continuous and satisfy $\mathrm{rc}^2+\mathrm{sc}^2\equiv 1$ and

$$\begin{aligned}&\mathrm{rc}(z,w)=0 &&\text{if } |z+\gamma^{-1}w|=0 \text{ or } (r_s(z,w))-\epsilon(r_l(z,w))\geq D_{\mathcal{K}}+\xi\cdot\mathbb{1}_{\{D_{\mathcal{K}}>0\}},\\ &\mathrm{rc}(z,w)=1 &&\text{if } |z+\gamma^{-1}w|\geq\xi \text{ and } (r_s(z,w))-\epsilon(r_l(z,w))\leq D_{\mathcal{K}} \text{ and } D_{\mathcal{K}}>0\end{aligned}\tag{C.44}$$

for $(z,w)\in\mathbb{R}^{2d}$, where ϵ is given in (C.29). Analogously to (C.26) and (C.27), $r_l(z,w)^2=\gamma^{-2}uz\cdot(Kz)+(1/2)|(1-2\tau)z+\gamma^{-1}w|^2+(1/2)\gamma^{-2}|w|^2$ and $r_s(z,w)=\alpha|z|+|z+\gamma^{-1}w|$.

We note that by Levy's characterization, for any solution to (C.74) the processes

$$\begin{aligned}B_t&:=\int_0^t\mathrm{sc}(Z_s,W_s)\mathrm{d}B_s^{\mathrm{sc}}+\int_0^t\mathrm{rc}(Z_s,W_s)\mathrm{d}B_s^{\mathrm{rc}}\qquad\text{and}\\ \tilde{B}_t&:=\int_0^t\mathrm{sc}(Z_s,W_s)\mathrm{d}B_s^{\mathrm{sc}}+\int_0^t\mathrm{rc}(Z_s,W_s)(\mathrm{Id}-e_s{e_s}^T)\mathrm{d}B_s^{\mathrm{rc}}\end{aligned}$$

are d-dimensional Brownian motions. Therefore, (C.74) defines a coupling between two solutions to (C.1). The constructed coupling denotes a *reflection coupling* for $\mathrm{rc} \equiv 1$ and $\mathrm{sc} \equiv 0$ and a *synchronous coupling* for $\mathrm{sc} \equiv 1$ and $\mathrm{rc} \equiv 0$. Note that we obtain a synchronous coupling if $D_{\mathcal{K}} = 0$.

The processes $(Z_t)_{t\geq 0}$, $(W_t)_{t\geq 0}$ and $(Q_t)_{t\geq 0}$ satisfy the following SDEs:

$$\begin{aligned} \mathrm{d}Z_t &= W_t \mathrm{d}t = (Q_t - \gamma Z_t)\mathrm{d}t, \\ \mathrm{d}W_t &= -\gamma W_t \mathrm{d}t + u\Big(b(\bar{X}_t) - b(\bar{X}'_t) + \int_{\mathbb{R}^d} \tilde{b}(\bar{X}_t, z)\bar{\mu}^x_t(\mathrm{d}z) - \int_{\mathbb{R}^d} \tilde{b}(\bar{X}'_t, \tilde{z})\bar{\nu}^x_t(\mathrm{d}\tilde{z})\Big)\mathrm{d}t \\ &\quad + \sqrt{8\gamma u}\mathrm{rc}(Z_t, W_t) e_t {e_t}^T \mathrm{d}B^{\mathrm{rc}}_t, \\ \mathrm{d}Q_t &= \gamma^{-1} u\Big(b(\bar{X}_t) - b(\bar{X}'_t) + \int_{\mathbb{R}^d} \tilde{b}(\bar{X}_t, z)\bar{\mu}^x_t(\mathrm{d}z) - \int_{\mathbb{R}^d} \tilde{b}(\bar{X}'_t, \tilde{z})\bar{\nu}^x_t(\mathrm{d}\tilde{z})\Big)\mathrm{d}t \\ &\quad + \sqrt{8\gamma^{-1} u}\mathrm{rc}(Z_t, W_t) e_t {e_t}^T \mathrm{d}B^{\mathrm{rc}}_t. \end{aligned} \tag{C.45}$$

If $Q_t = 0$, we note that Z_t is contractive, which we exploit in the proof of Lemma C.20.

C.5 Uniform in time propagation of chaos

We provide uniform in time propagation of chaos bounds for the mean-field particle system corresponding to the nonlinear Langevin dynamics of McKean-Vlasov type.

Fix $N \in \mathbb{N}$. We consider the metric $\rho_N : \mathbb{R}^{2Nd} \times \mathbb{R}^{2Nd} \to [0, \infty)$ given by

$$\rho_N((x,y),(\bar{x},\bar{y})) := N^{-1} \sum_{i=1}^N \rho((x^i, y^i), (\bar{x}^i, \bar{y}^i)) \qquad \text{for } ((x,y),(\bar{x},\bar{y})) \in \mathbb{R}^{2Nd} \times \mathbb{R}^{2Nd}, \tag{C.46}$$

where ρ is given in (C.35). Since ρ is a metric on $\mathbb{R}^{2d} \times \mathbb{R}^{2d}$ by Lemma C.16, ρ_N defines a metric on $\mathbb{R}^{2Nd} \times \mathbb{R}^{2Nd}$. By (C.40) and (C.41), ρ_N is equivalent to l^1_N given in (C.5), i.e.,

$$\mathbf{C}_1/\sqrt{2}\ell^1_N((x,y),(\bar{x},\bar{y})) \leq \rho_N((x,y),(\bar{x},\bar{y})) \leq \mathbf{C}_2/\sqrt{2}\ell^1_N((x,y),(\bar{x},\bar{y})) \tag{C.47}$$

with $\mathbf{C}_1 = \exp(-\Lambda)\min(1, 2(L_K + L_g)u\gamma^{-2})/3\gamma^{-1}\min(\sqrt{\kappa u}, 1/\sqrt{2})$ and $\mathbf{C}_2 = \sqrt{2}\max(2(L_K + L_g)u\gamma^{-2} + 1, \gamma^{-1})$.

For $t \geq 0$, we denote by $\bar{\mu}_t$ the law of the process $(\bar{X}_t, \bar{Y}_t)$, where $(\bar{X}_s, \bar{Y}_s)_{s\geq 0}$ is a solution to (C.1) with initial distribution $\bar{\mu}_0$. We denote by μ^N_t the law of $\{X^{i,N}_t, Y^{i,N}_t\}_{i=1}^N$, where $(\{X^{i,N}_s, Y^{i,N}_s\}_{i=1}^N)_{s\geq 0}$ is a solution to (C.3) with initial distribution $\mu^N_0 = \mu_0^{\otimes N}$.

Theorem C.17 (Propagation of chaos for Langevin dynamics)**.** *Suppose Assumption C.2 and Assumption C.3 hold. Let $\bar{\mu}_0$ and μ_0 be two probability distributions on $\mathbb{R}^{2d}$ with finite second moment. Suppose that* (C.18) *holds. If $\tilde{L}$ satisfies* (C.25)*, then*

$$\begin{aligned} &\mathcal{W}_{1,\rho_N}(\bar{\mu}_t^{\otimes N}, \mu^N_t) \leq e^{-\tilde{c}t}\mathcal{W}_{1,\rho_N}(\bar{\mu}_0^{\otimes N}, \mu^N_0) + \mathcal{C}_1\tilde{c}^{-1}N^{-1/2} \qquad \text{and} \\ &\mathcal{W}_{1,\ell^1_N}(\bar{\mu}_t^{\otimes N}, \mu^N_t) \leq M_1 e^{-\tilde{c}t}\mathcal{W}_{1,\ell^1_N}(\bar{\mu}_0^{\otimes N}, \mu^N_0) + M_2\mathcal{C}_1\tilde{c}^{-1}N^{-1/2}, \end{aligned}$$

where the distance ρ_N is defined in (C.46) *and $\tilde{c} = c/2$ with c given in* (C.19)*. The constant $\mathcal{C}_1$ depends on γ, d, u, R, κ, L_g, $\tilde{L}$ and on the second moment of $\bar{\mu}_0$. The constants M_1 and is given in* (C.24) *and* (C.48) *and M_2 is given by*

$$M_2 = 3\exp(\Lambda)\max\Big(1, \frac{\gamma^2}{2(L_K + L_g)u}\Big)\gamma\max(\sqrt{2/(\kappa u)}, 2). \tag{C.48}$$

Proof. The proof is postponed to Appendix C.6.3. □

Remark C.18. For $t \geq 0$, let μ_t^N and ν_t^N be the law of $\{X_t^{i,N}, Y_t^{i,N}\}_{i=1}^N$ and $\{X_t'^{i,N}, Y_t'^{i,N}\}_{i=1}^N$ where the processes $(\{X_s^{i,N}, Y_s^{i,N}\}_{i=1}^N)_{s\geq 0}$ and $(\{X_s'^{i,N}, Y_s'^{i,N}\}_{i=1}^N)_{s\geq 0}$ are solutions to (C.3) with initial distributions μ_0^N and ν_0^N, respectively. An easy adaptation of the proof of Theorem C.17 shows that if Assumption C.2, Assumption C.3, (C.18) and (C.25) hold, then

$$\mathcal{W}_{1,\rho_N}(\mu_t^N, \nu_t^N) \leq e^{-\tilde{c}t}\mathcal{W}_{1,\rho_N}(\mu_0^N, \nu_0^N) \qquad \text{and} \qquad \mathcal{W}_{1,\ell_N^1}(\mu_t^N, \nu_t^N) \leq M_1 e^{-\tilde{c}t}\mathcal{W}_{1,\ell_N^1}(\mu_0^N, \nu_0^N),$$

where ρ_N and M_1 are given in (C.46), and (C.24), respectively, and $\tilde{c} = c/2$ with c given in (C.19). To adapt the proof, a coupling between two copies of N particle systems is applied which is constructed in the same line as (C.74).

C.6 Proofs

C.6.1 Proof of Section C.2.1

Proof of Theorem C.1. Given a d-dimensional standard Brownian motion on $(B_t)_{t\geq 0}$ and $(x,y),(x',y') \in \mathbb{R}^{2d}$, we consider the synchronous coupling $((X_t,Y_t),(X_t',Y_t'))_{t\geq 0}$ of two copies of solutions to (C.6) on $\mathbb{R}^{2d}\times\mathbb{R}^{2d}$ given by

$$\begin{aligned}
&\begin{cases} \mathrm{d}X_t &= Y_t \mathrm{d}t \\ \mathrm{d}Y_t &= (-\gamma Y_t - u\nabla \mathrm{V}(X_t))\mathrm{d}t + \sqrt{2\gamma u}\,\mathrm{d}B_t, \qquad (X_0,Y_0) = (x,y) \end{cases}\\
&\begin{cases} \mathrm{d}X_t' &= Y_t' \mathrm{d}t \\ \mathrm{d}Y_t' &= (-\gamma Y_t' - u\nabla \mathrm{V}(X_t'))\mathrm{d}t + \sqrt{2\gamma u}\,\mathrm{d}B_t, \qquad (X_0',Y_0') = (x',y'). \end{cases}
\end{aligned} \tag{C.49}$$

Then, the difference process $(Z_t,W_t)_{t\geq 0} = (X_t - X_t', Y_t - Y_t')_{t\geq 0}$ satisfies

$$\begin{cases} \mathrm{d}Z_t = W_t \mathrm{d}t \\ \mathrm{d}W_t = (-\gamma W_t - uKZ_t - u(\nabla G(X_t) - \nabla G(X_t')))\mathrm{d}t. \end{cases}$$

We note that since by Assumption C.1, G is continuously differentiable, convex and has L_G-Lipschitz continuous gradients, G is co-coercive (see e.g. [157, Theorem 2.1.5]), i.e., it holds

$$|\nabla G(x) - \nabla G(x')|^2 \leq L_G(\nabla G(x) - \nabla G(x'))\cdot(x - x') \qquad \text{for all } x,x' \in \mathbb{R}^d. \tag{C.50}$$

Let $A,B,C \in \mathbb{R}^{d\times d}$ be positive definite matrices given by

$$A = \gamma^{-2}uK + (1/2)(1-2\lambda)^2\mathrm{Id}, \qquad B = (1-2\lambda)\gamma^{-1}\mathrm{Id}, \qquad C = \gamma^{-2}\mathrm{Id},$$

where λ is given in (C.9) and Id is the $d\times d$ identity matrix. Then by Ito's formula and Young's inequality, we obtain

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}(Z_t\cdot(AZ_t) + Z_t\cdot(BW_t) + W_t\cdot(CW_t))\\
&\quad = 2W_t\cdot(AZ_t) + W_t\cdot(BW_t) + Z_t\cdot(B(-\gamma W_t - uKZ_t - u(\nabla G(X_t) - \nabla G(X_t'))))\\
&\quad + 2W_t\cdot(C(-\gamma W_t - uKZ_t - u(\nabla G(X_t) - \nabla G(X_t'))))\\
&\quad \leq -u\gamma^{-1}(1-2\lambda)Z_t\cdot(KZ_t) - (1-2\lambda)\gamma^{-1}uZ_t(\nabla G(X_t) - \nabla G(X_t'))\\
&\quad + \gamma^{-3}u^2|\nabla G(X_t) - \nabla G(X_t')|^2 + Z_t\cdot((2A - \gamma B - 2uKC)W_t) + ((1-2\lambda)\gamma^{-1} - \gamma^{-1})|W_t|^2.
\end{aligned} \tag{C.51}$$

By (C.50), (C.9) and (C.10), it holds

$$\begin{aligned}-(1-2\lambda)\gamma^{-1}uZ_t\cdot(\nabla G(X_t)-\nabla G(X_t'))+\gamma^{-3}u^2|\nabla G(X_t)-\nabla G(X_t')|^2 \\ \leq -((1-2\lambda)\gamma^{-1}u-\gamma^{-3}L_G u^2)Z_t(\nabla G(X_t)-\nabla G(X_t'))\leq 0.\end{aligned} \tag{C.52}$$

Further by (C.9), it holds

$$\begin{aligned}-u\gamma^{-1}(1-4\lambda)Z_t\cdot(KZ_t)\leq -u(\gamma^{-1}/2)Z_t\cdot(KZ_t)\leq -u(\gamma^{-1}/2)\kappa|Z_t|^2 \\ \leq -\lambda\gamma|Z_t|^2\leq -\lambda\gamma(1-2\lambda)^2|Z_t|^2\end{aligned}$$

and hence, $-u\gamma^{-1}(1-2\lambda)Z_t\cdot(KZ_t)\leq -2\gamma\lambda Z_t\cdot(AZ_t)$. Set $r(t)=r((X_t,Y_t),(X_t',Y_t'))$ with r defined in (C.8). Then by (C.51) and (C.52), we obtain

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}r(t)^2&=\frac{\mathrm{d}}{\mathrm{d}t}(Z_t\cdot(AZ_t)+Z_t\cdot(BW_t)+W_t\cdot(CW_t)) \\ &\leq -2\lambda\gamma(Z_t\cdot(AZ_t)+Z_t\cdot(BW_t)+W_t\cdot(CW_t))=-2\lambda\gamma r(t)^2.\end{aligned}$$

Taking the square root and applying Grönwall's inequality yields

$$r(t)\leq e^{-ct}r(0)$$

with c given in (C.11). Then for all $p\geq 1$ it holds

$$\mathcal{W}_{p,r}(\mu_t,\nu_t)\leq \mathbb{E}[r(t)^p]^{1/p}\leq e^{-ct}\mathbb{E}[r(0)^p]^{1/p}.$$

We take the infimum over all couplings $\gamma\in\Pi(\mu_0,\nu_0)$ and obtain the first bound. For the second bound we note that for any $(x,y),(x',y')\in\mathbb{R}^{2d}$

$$\begin{aligned}\sqrt{\min(u\gamma^{-2}\kappa,\gamma^{-2}/2)}(|x-x'|^2+|y-y'|^2)^{1/2}\leq r((x,y),(x',y')) \\ \leq\sqrt{\max(u\gamma^{-2}L_K+1,3/2\gamma^{-2})}(|x-x'|^2+|y-y'|^2)^{1/2}.\end{aligned}$$

Hence, the second bound in Theorem C.1 holds with M given in (C.12). □

C.6.2 Proofs of Section C.2.4 and Section C.3

To show Theorem C.12, we prove two local contraction results using the coupling defined in (C.43). We write $r_l(t)=r_l((\bar{X}_t,\bar{Y}_t),(\bar{X}_t',\bar{Y}_t'))$, $r_s(t)=r_s((\bar{X}_t,\bar{Y}_t),(\bar{X}_t',\bar{Y}_t'))$ and $\Delta(t)=\Delta((\bar{X}_t,\bar{Y}_t),(\bar{X}_t',\bar{Y}_t'))$.

Lemma C.19. *Suppose Assumption C.2, Assumption C.3 and* (C.18) *hold.*
Let $((\bar{X}_s,\bar{Y}_s),(\bar{X}_s',\bar{Y}_s'))_{s\geq 0}$ *be a solution to* (C.43). *Then for* $t\geq 0$ *with* $\Delta(t)\geq D_{\mathcal{K}}$, *it holds*

$$\begin{aligned}\mathrm{d}r_l(t)\leq -c_1r_l(t)\mathrm{d}t+\frac{|(1-2\tau)Z_t+2\gamma^{-1}W_t|}{2\gamma r_l(t)}\tilde{L}u(\mathbb{E}[|Z_t|]+|Z_t|)\mathrm{d}t \\ +\sqrt{8\gamma^{-1}u}\mathrm{rc}(Z_t,W_t)\frac{(1-2\tau)Z_t+2\gamma^{-1}W_t}{2r_l(t)}\cdot e_te_t{}^T\mathrm{d}B_t,\end{aligned} \tag{C.53}$$

where $c_1=\tau\gamma/2$ *with* τ *given in* (C.21).

Proof. Let $A, B, C \in \mathbb{R}^{d\times d}$ be positive definite matrices given by

$$A = \gamma^{-2}uK + (1/2)(1-2\tau)^2\mathrm{Id}, \qquad B = (1-2\tau)\gamma^{-1}\mathrm{Id}, \qquad \text{and} \qquad C = \gamma^{-2}\mathrm{Id}, \tag{C.54}$$

where τ is given by (C.21) and Id is the $d \times d$ identity matrix. By (C.45) and Ito's formula, it holds

$$\begin{aligned}
&\mathrm{d}(Z_t \cdot (AZ_t) + Z_t \cdot (BW_t) + W_t \cdot (CW_t)) \\
&\quad \le 2(AZ_t)\cdot W_t \mathrm{d}t + \Big(W_t \cdot (BW_t) - \gamma(BZ_t)\cdot W_t - u(BZ_t)\cdot(KZ_t)\Big)\mathrm{d}t \\
&\quad + \Big(-2\gamma W_t \cdot (CW_t) - 2u(CW_t)\cdot(KZ_t) + 2\gamma^{-2}L_g u|W_t||Z_t|\Big)\mathrm{d}t \\
&\quad + L_g u(1-2\tau)\gamma^{-1}|Z_t|^2 \cdot \mathbb{1}_{\{|Z_t|<R\}}\mathrm{d}t + |BZ_t + 2CW_t|\tilde{L}u(\mathbb{E}[|Z_t|] + |Z_t|)\mathrm{d}t \\
&\quad + \gamma^{-2}8\gamma u \mathrm{rc}(Z_t, W_t)^2\mathrm{d}t + \sqrt{8\gamma u}\,\mathrm{rc}(Z_t, W_t)(BZ_t + 2CW_t)\cdot e_t {e_t}^T \mathrm{d}B_t \\
&\quad \le Z_t \cdot ((-uBK + \gamma^{-1}uL_g^2 C)Z_t)\mathrm{d}t + Z_t \cdot (2A - \gamma B - 2uKC)W_t \mathrm{d}t \\
&\quad + ((1-2\tau)\gamma^{-1} - \gamma^{-1})|W_t|^2 \mathrm{d}t \\
&\quad + (1-2\tau)\gamma^{-1}uL_g|Z_t|^2\mathbb{1}_{\{|Z_t|<R\}}\mathrm{d}t + |(1-2\tau)\gamma^{-1}Z_t + 2\gamma^{-2}W_t|\tilde{L}u(\mathbb{E}[|Z_t|] + |Z_t|)\mathrm{d}t \\
&\quad + 8\gamma^{-1}u(\mathrm{rc}(Z_t, W_t))^2\mathrm{d}t + \sqrt{8\gamma u}\,\mathrm{rc}(Z_t, W_t)((1-2\tau)\gamma^{-1}Z_t + 2\gamma^{-1}W_t)\cdot e_t {e_t}^T \mathrm{d}B_t \\
&\quad \le -2\tau\gamma(Z_t \cdot (AZ_t) + Z_t \cdot (BW_t) + W_t \cdot (CW_t))\mathrm{d}t \\
&\quad + (1-2\tau)\gamma^{-1}uL_g|Z_t|^2\mathbb{1}_{\{|Z_t|<R\}}\mathrm{d}t + |(1-2\tau)\gamma^{-1}Z_t + 2\gamma^{-2}W_t|\tilde{L}u(\mathbb{E}[|Z_t|] + |Z_t|)\mathrm{d}t \\
&\quad + 8\gamma^{-1}u(\mathrm{rc}(Z_t, W_t))^2\mathrm{d}t + \sqrt{8\gamma u}\,\mathrm{rc}(Z_t, W_t)((1-2\tau)\gamma^{-1}Z_t + 2\gamma^{-1}W_t)\cdot e_t {e_t}^T \mathrm{d}B_t,
\end{aligned}$$

where we used (C.21) in the last step. More precisely, the definition of τ implies for all $z \in \mathbb{R}^d$,

$$\begin{aligned}
z \cdot ((-(1-4\tau)\gamma^{-1}uK + \gamma^{-3}L_g^2u^2\mathrm{Id})z) &\le (-(1/2)\kappa u\gamma^{-1} + \gamma^{-3}L_g^2u^2)|z|^2 \\
&\le (-\tau\gamma)|z|^2 \le (-\tau\gamma(1-2\tau)^2)|z|^2.
\end{aligned} \tag{C.55}$$

Note that $r_l(t)^2 = Z_t \cdot (AZ_t) + Z_t \cdot (BW_t) + W_t \cdot (CW_t)$. Then,

$$\begin{aligned}
\mathrm{d}r_l(t)^2 &\le -2\tau\gamma r_l(t)^2\mathrm{d}t + \gamma^{-1}(1-2\tau)L_g u|Z_t|^2\mathbb{1}_{\{|Z_t|<R\}}\mathrm{d}t \\
&\quad + \gamma^{-1}|(1-2\tau)Z_t + 2\gamma^{-1}W_t|\tilde{L}u(\mathbb{E}[|Z_t|] + |Z_t|)\mathrm{d}t \\
&\quad + 8\gamma^{-1}u\mathrm{rc}(Z_t, W_t)^2\mathrm{d}t + \sqrt{8\gamma^{-1}u}\,\mathrm{rc}(Z_t, W_t)((1-2\tau)Z_t + 2\gamma^{-1}W_t)\cdot e_t {e_t}^T \mathrm{d}B_t.
\end{aligned}$$

Since $\Delta(t) \ge D_{\mathcal{K}}$, it holds $r_l(t)^2 \ge \mathcal{R}$ by (C.32) and (C.33). By (C.44), $\mathrm{rc}(Z_t, W_t)^2 \le \mathbb{1}_{\{R>0\}}$, and hence, by (C.34)

$$\begin{aligned}
-\tau\gamma r_l(t)^2 + \gamma^{-1}(1-2\tau)L_g u|Z_t|^2\mathbb{1}_{\{|Z_t|<R\}} &+ 8\gamma^{-1}u\mathrm{rc}(Z_t, W_t)^2 \\
&\le -\tau\gamma\mathcal{R} + L_g uR^2\gamma^{-1} + 8\gamma^{-1}u\mathbb{1}_{\{R>0\}} \le 0.
\end{aligned}$$

We obtain by Ito's formula and since the second derivative of the square root is negative,

$$\begin{aligned}
\mathrm{d}r_l(t) &\le (2r_l(t))^{-1}\mathrm{d}r_l(t)^2 \\
&\le -c_1 r_l(t)\mathrm{d}t + \gamma^{-1}|(1-2\tau)Z_t + 2\gamma^{-1}W_t|(2r_l(t))^{-1}\tilde{L}u(\mathbb{E}[|Z_t|] + |Z_t|)\mathrm{d}t \\
&\quad + \sqrt{8\gamma^{-1}}\mathrm{rc}(Z_t, W_t)(2r_l(t))^{-1}((1-2\tau)Z_t + 2\gamma^{-1}W_t)\cdot e_t {e_t}^T \mathrm{d}B_t,
\end{aligned}$$

which concludes the proof. ☐

Lemma C.20. *Suppose Assumption C.2 and Assumption C.3 hold. Fix* $\xi > 0$. *Let* $((\bar{X}_s, \bar{Y}_s), (\bar{X}'_s, \bar{Y}'_s))_{s\geq 0}$ *be a solution to* (C.43). *Let* r_s *be given by* (C.27) *with* α *given in* (C.28). *Then for* $t \geq 0$ *with* $\Delta(t) < D_{\mathcal{K}}$, *it holds*

$$\mathrm{d}f(r_s(t)) \leq -c_2 f(r_s(t))\mathrm{d}t + \gamma^{-1}\tilde{L}u(\mathbb{E}[|Z_t|] + |Z_t|)\mathrm{d}t - \frac{\gamma\alpha}{4}f'(R_1)|Z_t|\mathrm{d}t + (1+\alpha)\xi\gamma\mathrm{d}t + \mathrm{d}M_t,$$

where f *is given in* (C.36), $(M_t)_{t\geq 0}$ *is a martingale and* c_2 *is given by*

$$c_2 := \min\Big(\frac{2}{\gamma\int_0^{R_1}\Phi(s)\phi(s)^{-1}ds}, \frac{\gamma}{8}\frac{R_1\phi(R_1)}{\Phi(R_1)}\Big). \tag{C.56}$$

Proof. The proof is an adaptation of the proof of [85, Lemma 3.1]. First, we note that, $(Z_t)_{t\geq 0}$ given in (C.45) is almost surely continuously differentiable with derivative $\mathrm{d}Z_t/\mathrm{d}t = -\gamma Z_t + \gamma Q_t$ and hence $t \to |Z_t|$ is almost surely absolutely continuous with

$$\frac{\mathrm{d}}{\mathrm{d}t}|Z_t| = \frac{Z_t}{|Z_t|}\cdot(-\gamma Z_t + \gamma Q_t) \qquad \text{for a.e. } t \text{ such that } Z_t \neq 0 \text{ and}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}|Z_t| \leq \gamma|Q_t| \qquad \text{for a.e. } t \text{ such that } Z_t = 0.$$

and therefore

$$\frac{\mathrm{d}}{\mathrm{d}t}|Z_t| \leq -\gamma|Z_t| + \gamma|Q_t| \text{ for a.e. } t \geq 0. \tag{C.57}$$

By Ito's formula and by Assumption C.2 and Assumption C.3, we obtain for $|Q_t|$,

$$\begin{aligned}\mathrm{d}|Q_t| &= \gamma^{-1}ue_t\cdot\Big(b(\bar{X}_t) - b(\bar{X}'_t) + \int_{\mathbb{R}^d}\tilde{b}(\bar{X}_t, z)\bar{\mu}^x_t(\mathrm{d}z) - \int_{\mathbb{R}^d}\tilde{b}(\bar{X}'_t, \tilde{z})\bar{\nu}^x_t(\mathrm{d}\tilde{z})\Big)\mathrm{d}t \\ &\quad + \sqrt{8\gamma^{-1}u}\mathrm{rc}(Z_t, W_t){e_t}^T\mathrm{d}B_t \\ &\leq \gamma^{-1}u(L_K + L_g + \tilde{L})|Z_t|\mathrm{d}t + \gamma^{-1}\tilde{L}u\mathbb{E}[|Z_t|]\mathrm{d}t + \sqrt{8\gamma^{-1}u}\mathrm{rc}(Z_t, W_t){e_t}^T\mathrm{d}B^{\mathrm{rc}}_t.\end{aligned}$$

Note that there is no Ito correction term, since $\partial^2_{q/|q|}|q| = 0$ for $q \neq 0$ and $\mathrm{rc} = 0$ for $Q_t = 0$. Combining this bound with (C.57) yields for $r_s(t)$,

$$\begin{aligned}\mathrm{d}r_s(t) &\leq \Big(((L_K + L_g)u\gamma^{-2} - \alpha)\gamma|Z_t| + \alpha\gamma|Q_t| + \gamma^{-1}\tilde{L}u(\mathbb{E}[|Z_t|] + |Z_t|)\Big)\mathrm{d}t \\ &\quad + \sqrt{8\gamma^{-1}u}\mathrm{rc}(Z_t, W_t){e_t}^T\mathrm{d}B^{\mathrm{rc}}_t.\end{aligned}$$

By Ito's formula,

$$\begin{aligned}\mathrm{d}f(r_s(t)) &\leq f'(r_s(t))\Big(((L_K + L_g)u\gamma^{-2} - \alpha)\gamma|Z_t| + \alpha\gamma|Q_t| + \gamma^{-1}\tilde{L}u(\mathbb{E}[|Z_t|] + |Z_t|)\Big)\mathrm{d}t \\ &\quad + f'(r_s(t))\sqrt{8\gamma^{-1}u}\mathrm{rc}(Z_t, W_t){e_t}^T\mathrm{d}B^{\mathrm{rc}}_t + f''(r_s(t))4\gamma^{-1}u\mathrm{rc}(Z_t, W_t)^2\mathrm{d}t.\end{aligned}$$

Case 1: Consider $\Delta(t) < D_{\mathcal{K}}$ and $|Q_t| > \xi$, then $\mathrm{rc}(Z_t, W_t) = 1$ and $r_s(t) < R_1$. Hence, we obtain

$$\begin{aligned}\mathrm{d}f(r_s(t)) &\leq f'(r_s(t))\alpha\gamma r_s(t)\mathrm{d}t + f''(r_s(t))4\gamma^{-1}u\mathrm{d}t + \gamma^{-1}\tilde{L}u(\mathbb{E}[|Z_t|] + |Z_t|)\mathrm{d}t \\ &\quad - \frac{\alpha\gamma}{2}|Z_t|f'(r_s(t))\mathrm{d}t + \mathrm{d}M_t \\ &\leq -2\hat{c}f(r_s(t))\mathrm{d}t + \gamma^{-1}\tilde{L}u(\mathbb{E}[|Z_t|] + |Z_t|)\mathrm{d}t - \frac{\alpha\gamma}{2}|Z_t|f'(R_1)\mathrm{d}t + \mathrm{d}M_t \\ &\leq -c_2 f(r_s(t))\mathrm{d}t + \gamma^{-1}\tilde{L}u(\mathbb{E}[|Z_t|] + |Z_t|)\mathrm{d}t - \frac{\alpha\gamma}{2}|Z_t|f'(R_1)\mathrm{d}t + \mathrm{d}M_t,\end{aligned}$$

where $(M_t)_{t\geq 0}$ is a martingale and $\hat{c}$ is given in (C.37). Note that the second step holds since by (C.36) and (C.39),

$$f'(r)\alpha\gamma r + f''(r)4\gamma^{-1}u \leq -2\hat{c}f(r) \qquad \text{for all } r \in [0, R_1). \tag{C.58}$$

Case 2: Consider $\Delta(t) < D_{\mathcal{K}}$ and $|Q_t| \leq \xi$, then $\alpha|Z_t| = r_s(t) - |Q_t| \geq r_s(t) - \xi$. We note that

$$((L_K + L_g)u\gamma^{-2} - \alpha)|Z_t| + \alpha|Q_t| \leq -\frac{1}{2}r_s(t) + (1+\alpha)\xi.$$

Since the second derivative of f is negative and $\psi(s) \in [1/2, 1]$, it holds

$$\begin{aligned}
\mathrm{d}f(r_s(t)) &\leq -\frac{\gamma}{2}r_s(t)f'(r_s(t))\mathrm{d}t + (1+\alpha)\gamma\xi\mathrm{d}t + \gamma^{-1}u\tilde{L}(\mathbb{E}[|Z_t|] + |Z_t|)\mathrm{d}t + \mathrm{d}M_t \\
&\leq -\frac{\gamma}{8}\inf_{r\leq R_1}\frac{r\phi(r)}{\Phi(r)}f(r_s(t))\mathrm{d}t - \frac{\gamma}{4}f'(R_1)\alpha|Z_t|\mathrm{d}t + (1+\alpha)\gamma\xi\mathrm{d}t \\
&\quad + \gamma^{-1}\tilde{L}u(\mathbb{E}[|Z_t|] + |Z_t|)\mathrm{d}t + \mathrm{d}M_t \\
&\leq -\frac{\gamma}{8}\frac{R_1\phi(R_1)}{\Phi(R_1)}f(r_s(t))\mathrm{d}t - \frac{\gamma\alpha}{4}f'(R_1)|Z_t|\mathrm{d}t + (1+\alpha)\gamma\xi\mathrm{d}t \\
&\quad + \gamma^{-1}\tilde{L}u(\mathbb{E}[|Z_t|] + |Z_t|)\mathrm{d}t + \mathrm{d}M_t.
\end{aligned} \tag{C.59}$$

Combining the two cases, we obtain the result with c_2 given in (C.56). □

Proof of Theorem C.17. To prove contraction, we consider the coupling $((\bar{X}_t, \bar{Y}_t), (\bar{X}'_t, \bar{Y}'_t))_{t\geq 0}$ given in (C.43) and combine the results of Lemma C.19 and Lemma C.20. We abbreviate $\rho(t) = f((\Delta(t) \wedge D_{\mathcal{K}}) + \epsilon r_l(t))$. We distinguish two cases:
Case 1: Consider $\Delta(t) < D_{\mathcal{K}}$. Then $r_s(t) \leq R_1$ and $\rho(t) = f(r_s(t))$. By Lemma C.20, it holds for $\xi > 0$

$$\begin{aligned}
\mathrm{d}\rho(t) &= \mathrm{d}f(r_s(t)) \\
&\leq -c_2 f(r_s(t))\mathrm{d}t + \gamma^{-1}\tilde{L}u(\mathbb{E}[|Z_t|] + |Z_t|)\mathrm{d}t - \frac{\alpha\gamma}{4}f'(R_1)|Z_t|\mathrm{d}t + (1+\alpha)\gamma\xi\mathrm{d}t + \mathrm{d}M_t \\
&\leq -c_2 f(r_s(t))\mathrm{d}t + \gamma^{-1}\tilde{L}u\mathbb{E}[|Z_t|]\mathrm{d}t - \frac{\alpha\gamma}{8}f'(R_1)|Z_t|\mathrm{d}t + (1+\alpha)\gamma\xi\mathrm{d}t + \mathrm{d}M_t,
\end{aligned} \tag{C.60}$$

where c_2 is given by (C.56) and $(M_t)_{t\geq 0}$ is a martingale. The second step holds by (C.25).
Case 2: Consider $\Delta(t) \geq D_{\mathcal{K}}$. We obtain by Lemma C.19,

$$\begin{aligned}
\mathrm{d}r_l(t) \leq -c_1 r_l(t)\mathrm{d}t &+ \frac{|(1-2\tau)Z_t + 2\gamma^{-1}W_t|}{2\gamma r_l(t)}\tilde{L}u(\mathbb{E}[|Z_t|] + |Z_t|)\mathrm{d}t \\
&+ \sqrt{8\gamma^{-1}u}\mathrm{rc}(Z_t, W_t)\frac{(1-2\tau)Z_t + 2\gamma^{-1}W_t}{2r_l(t)} \cdot e_t {e_t}^T \mathrm{d}B_t,
\end{aligned}$$

where c_1 is given in Lemma C.19. Note that $\frac{\mathrm{d}}{\mathrm{d}x}f(D_{\mathcal{K}} + \epsilon x) = \epsilon f'(D_{\mathcal{K}} + \epsilon x)$. Further, since $f(D_{\mathcal{K}} + \epsilon x)$ is a concave function, $\frac{\mathrm{d}^2}{\mathrm{d}x^2}f(D_{\mathcal{K}} + \epsilon x)$ is negative. By Ito's formula, we obtain

$$\begin{aligned}
\mathrm{d}\rho(t) &= \mathrm{d}f(D_{\mathcal{K}} + \epsilon r_l(t)) \\
&\leq \epsilon f'(D_{\mathcal{K}} + \epsilon r_l(t))\Big(-c_1 r_l(t) + \frac{|(1-2\tau)Z_t + 2\gamma^{-1}W_t|}{2\gamma r_l(t)}\tilde{L}u(\mathbb{E}[|Z_t|] + |Z_t|)\Big)\mathrm{d}t + \mathrm{d}\tilde{M}_t,
\end{aligned} \tag{C.61}$$

where $\tilde{M}_t$ is a martingale given by

$$\tilde{M}_t = \int_0^t \frac{\epsilon f'(D_{\mathcal{K}} + \epsilon r_l(s))}{2r_l(s)} \sqrt{8\gamma^{-1}u}\mathrm{rc}(Z_s, W_s)((1-2\tau)Z_s + 2\gamma^{-1}W_s) \cdot e_s {e_s}^T \mathrm{d}B_s. \tag{C.62}$$

We split the first term of (C.61) and bound each part applying (C.39),

$$-\frac{\epsilon f'(D_{\mathcal{K}} + \epsilon r_l(t))}{2} c_1 r_l(t) \leq -\Big\{\inf_{q\geq 0} \frac{f'(q)q}{f(q)}\Big\} \frac{\epsilon c_1 r_l(t)}{2(D_{\mathcal{K}} + \epsilon r_l(t))}\rho(t) \leq -f'(R_1)\frac{\epsilon c_1 r_l(t)}{2(D_{\mathcal{K}} + \epsilon r_l(t))}\rho(t) \tag{C.63}$$

and

$$-\frac{\epsilon f'(D_{\mathcal{K}} + \epsilon r_l(t))}{2} c_1 r_l(t) \leq -f'(R_1)\frac{\epsilon c_1}{2} r_l(t). \tag{C.64}$$

We note that since $\Delta(t) > D_{\mathcal{K}}$ it holds,

$$\frac{r_l(t)}{D_{\mathcal{K}} + \epsilon r_l(t)} \geq \frac{r_l(t)}{r_s(t)} \geq \mathcal{E}, \tag{C.65}$$

where $\mathcal{E}$ is given in (C.30). Hence, we obtain for the first term of (C.61), by (C.63), (C.64) and (C.65)

$$-\epsilon f'(D_{\mathcal{K}} + \epsilon r_l(t))c_1 r_l(t)^2 \leq -f'(R_1)\frac{c_1\epsilon\mathcal{E}}{2}\rho(t) - f'(R_1)\frac{c_1\epsilon}{2} r_l(t). \tag{C.66}$$

For the second term of (C.61), we note

$$\begin{aligned} &\epsilon f'(D_{\mathcal{K}} + \epsilon r_l(t)) \frac{|(1-2\tau)Z_t + 2\gamma^{-1}W_t|}{2\gamma r_l(t)} \\ &\qquad \leq \frac{\epsilon}{2\gamma}\sqrt{\frac{(1-2\tau)^2|Z_t|^2 + 4(1-2\tau)\gamma^{-1}Z_t\cdot W_t + 4\gamma^{-2}|W_t|^2}{(1/2)(1-2\tau)^2|Z_t|^2 + (1-2\tau)\gamma^{-1}Z_t\cdot W_t + \gamma^{-2}|W_t|^2}} \leq \frac{\epsilon}{\gamma}. \end{aligned} \tag{C.67}$$

Combining (C.66) and (C.67) yields,

$$\begin{aligned} \mathrm{d}\rho(t) &\leq -f'(R_1)\frac{c_1\epsilon\mathcal{E}}{2}\rho(t)\mathrm{d}t - f'(R_1)\frac{c_1\epsilon}{2} r_l(t)\mathrm{d}t + \epsilon\gamma^{-1}\tilde{L}u(\mathbb{E}[|Z_t|] + |Z_t|)\mathrm{d}t + \mathrm{d}\tilde{M}_t. \\ &\leq -f'(R_1)\frac{c_1\epsilon\mathcal{E}}{2}\rho(t)\mathrm{d}t - f'(R_1)\frac{c_1\epsilon}{2}\sqrt{\kappa u\gamma^{-2}}|Z_t|\mathrm{d}t + \frac{1}{2}\gamma^{-1}\tilde{L}u(\mathbb{E}[|Z_t|] + |Z_t|)\mathrm{d}t + \mathrm{d}\tilde{M}_t, \end{aligned} \tag{C.68}$$

where $r_l(t) \geq \sqrt{\kappa u \gamma^{-2}}|Z_t|$ and $2\epsilon \leq 1$ are applied and where $(\tilde{M}_t)_{t\geq 0}$ is given in (C.62).

Combining (C.60) and (C.68), taking expectation and $\xi \to 0$, yields

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\rho(t)] \leq &- \min\Big(c_2, f'(R_1)\frac{c_1\epsilon\mathcal{E}}{2}\Big)\mathbb{E}[\rho(t)] - \min\Big(f'(R_1)\frac{\alpha\gamma}{8}, f'(R_1)\frac{c_1\epsilon}{2}\sqrt{\kappa u\gamma^{-2}}\Big)\mathbb{E}[|Z_t|] \\ &+ \gamma^{-1}\tilde{L}u\mathbb{E}[|Z_t|] \\ \leq &- \min\Big(c_2, f'(R_1)\frac{c_1\epsilon\mathcal{E}}{2}\Big)\mathbb{E}[\rho(t)], \end{aligned}$$

where we used (C.25) and (C.28) in the second step. By applying Grönwall's inequality, we obtain

$$\mathcal{W}_{1,\rho}(\bar{\mu}_t, \bar{\nu}_t) \leq \mathbb{E}[\rho(t)] \leq e^{-c_3 t}\mathbb{E}[\rho(0)]$$

with

$$c_3 = \min\Big(\frac{2}{\gamma\int_0^{R_1}\Phi(s)\phi(s)^{-1}ds}, \frac{\gamma}{8}\frac{R_1\phi(R_1)}{\Phi(R_1)}, f'(R_1)\gamma\tau\frac{\epsilon\mathcal{E}}{4}\Big). \tag{C.69}$$

The term $\epsilon\mathcal{E}$ is bounded from below by E given in (C.22). For the first two arguments in the minimum we note that

$$\begin{aligned}\int_0^{R_1}\int_0^s \exp\Big(-\frac{\alpha\gamma^2}{4u}\frac{r^2}{2}\Big)\mathrm{d}r\exp\Big(\frac{\alpha\gamma^2}{4u}\frac{s^2}{2}\Big)\mathrm{d}s &\le \sqrt{\frac{\pi}{2}}\Big(\frac{\alpha\gamma^2}{4u}\Big)^{-1/2}\int_0^{R_1}\exp\Big(\frac{\alpha\gamma^2}{4u}\frac{s^2}{2}\Big)\mathrm{d}s\\ &\le \sqrt{\frac{\pi}{2}}\Big(\frac{\alpha\gamma^2}{4u}\Big)^{-1/2}2\Big(\frac{\alpha\gamma^2}{4u}R_1\Big)^{-1}\exp\Big(\frac{\alpha\gamma^2}{4u}\frac{R_1^2}{2}\Big) \le 4\Big(\frac{\alpha\gamma^2}{4u}\Big)^{-1}\Big(\frac{\alpha\gamma^2}{4u}\frac{R_1^2}{2}\Big)^{-1/2}\exp\Big(\frac{\alpha\gamma^2}{4u}\frac{R_1^2}{2}\Big)\end{aligned} \tag{C.70}$$

since $\int_0^x \exp(r^2/2)\mathrm{d}r \le 2x^{-1}\exp(x^2/2)$, and

$$\begin{aligned}\frac{R_1\phi(R_1)}{\Phi(R_1)} &\ge \frac{R_1\exp(-\frac{\alpha\gamma^2}{4u}\frac{R_1^2}{2})}{\sqrt{\frac{\pi}{2}}(\frac{\alpha\gamma^2}{4u})^{-1/2}} = \frac{2}{\sqrt{\pi}}\Big(\frac{\alpha\gamma^2}{4u}\frac{R_1^2}{2}\Big)^{1/2}\exp\Big(-\frac{\alpha\gamma^2}{4u}\frac{R_1^2}{2}\Big)\\ &\ge \Big(\frac{\alpha\gamma^2}{4u}\frac{R_1^2}{2}\Big)^{1/2}\exp\Big(-\frac{\alpha\gamma^2}{4u}\frac{R_1^2}{2}\Big).\end{aligned} \tag{C.71}$$

Hence, $\mathcal{W}_{1,\rho}(\bar{\mu}_t,\bar{\nu}_t) \le \mathbb{E}[\rho(t)] \le e^{-\bar{c}t}\mathbb{E}[\rho(0)]$ with c given by

$$\bar{c} = \gamma\exp(-\Lambda)\min\Big(\frac{(L_K+L_g)u\gamma^{-2}}{4}\Lambda^{1/2}, \frac{1}{8}\Lambda^{1/2}, \frac{\tau E}{4}\Big) \tag{C.72}$$

with Λ, τ and E given in (C.20), (C.21) and (C.22). Taking the infimum over all couplings $\omega \in \Pi(\bar{\mu}_0,\bar{\nu}_0)$ concludes the proof of the first result.

By (C.42), the second result holds with $M_1 = \mathbf{C}_2/\mathbf{C}_1$ given by (C.24). □

Proof of Theorem C.5. Theorem C.5 forms a special case of Theorem C.12. We obtain analogously to Lemma C.19 for $\Delta(t) \ge D_{\mathcal{K}}$,

$$\mathrm{d}r_l(t) \le -c_1 r_l(t)\mathrm{d}t + \sqrt{8\gamma^{-1}u}\mathrm{rc}(Z_t,W_t)(r_l(t)^{-1}/2)((1-2\tau)Z_t + 2\gamma^{-1}W_t)\cdot e_t {e_t}^T\mathrm{d}B_t,$$

where $c_1 = \tau\gamma/2$ with τ given in (C.21). Similarly as in Lemma C.20, we get for $\Delta(t) < D_{\mathcal{K}}$ using $\tilde{L} = 0$

$$\mathrm{d}f(r_s(t)) \le -c_2 f(r_s(t))\mathrm{d}t + (1+\alpha)\xi\gamma\mathrm{d}t + \mathrm{d}M_t,$$

where M_t is a martingale, α is defined in (C.28), f is defined in (C.36) and c_2 is given in (C.56). Combining the two local contraction results as in the proof of Theorem C.12 gives the desired result with contraction rate

$$c = \min\Big(\frac{2}{u^{-1}\gamma\int_0^{R_1}\Phi(s)\phi(s)^{-1}ds}, \frac{\gamma}{8}\frac{R_1\phi(R_1)}{\Phi(R_1)}, f'(R_1)\gamma\tau\frac{\epsilon\mathcal{E}}{2}\Big). \tag{C.73}$$

Note that the last two terms in the minimum differ by a factor of 2 from the last two terms in (C.69), as the first terms in (C.61) and (C.59) are not split up to compensate for the interaction term as in the nonlinear term. □

C.6.3 Proof of Section C.5

Fix $N \in \mathbb{N}$. To show propagation in chaos in Theorem C.17 we construct in the same line as in Appendix C.4.2 a coupling between a solution to (C.3) and N copies of solutions to (C.1). We fix a positive constant ξ, which we take in the end to the limit $\xi \to 0$. Let $\{(B^{i,\mathrm{rc}})_{t\geq 0} : i = 1, \ldots, N\}$ and $\{(B^{i,\mathrm{sc}})_{t\geq 0} : i = 1, \ldots, N\}$ be $2N$ independent d-dimensional Brownian motions and let μ_0 and $\bar{\mu}_0$ be two probability measures on $\mathbb{R}^{2d}$. The coupling $(\{(\bar{X}^i_t, \bar{Y}^i_t), (X^i_t, Y^i_t)\}_{i=1}^N)_{t\geq 0}$ is a solution to the SDE on $\mathbb{R}^{2Nd} \times \mathbb{R}^{2Nd}$ given by

$$
\begin{aligned}
&\begin{cases}
\mathrm{d}\bar{X}^i_t &= \bar{Y}^i_t \mathrm{d}t \\
\mathrm{d}\bar{Y}^i_t &= (-\gamma \bar{Y}^i_t + ub(\bar{X}^i_t) + u\int_{\mathbb{R}^d} \tilde{b}(\bar{X}^i_t, z)\bar{\mu}^x_t(\mathrm{d}z))\mathrm{d}t + \sqrt{2\gamma u}\,\mathrm{sc}(Z^i_t, W^i_t)\mathrm{d}B^{i,\mathrm{sc}}_t \\
&\quad + \sqrt{2\gamma u}\,\mathrm{rc}(Z^i_t, W^i_t)\mathrm{d}B^{i,\mathrm{rc}}_t
\end{cases} \\
&\begin{cases}
\mathrm{d}X^i_t &= Y^i_t \mathrm{d}t \\
\mathrm{d}Y^i_t &= (-\gamma Y^i_t + ub(X^i_t) + uN^{-1}\sum_{j=1}^N \tilde{b}(X^i_t, X^j_t))\mathrm{d}t + \sqrt{2\gamma u}\,\mathrm{sc}(Z^i_t, W^i_t)\mathrm{d}B^{i,\mathrm{sc}}_t \\
&\quad + \sqrt{2\gamma u}\,\mathrm{rc}(Z^i_t, W^i_t)(\mathrm{Id} - 2e^i_t {e^i_t}^T)\mathrm{d}B^{i,\mathrm{rc}}_t
\end{cases} \\
&(\bar{X}^i_0, \bar{Y}^i_0) \sim \bar{\mu}_0, \quad (X^i_0, Y^i_0) \sim \mu_0
\end{aligned}
\tag{C.74}
$$

for $i = 1, ..., N$, where $\bar{\mu}^x_t = \mathrm{Law}(\bar{X}^i_t)$ for all i. Further, $Z^i_t = \bar{X}^i_t - X^i_t$, $W^i_t = \bar{Y}^i_t - Y^i_t$, $Q^i_t = Z^i_t + \gamma^{-1} W^i_t$, and $e^i_t = Q^i_t/|Q^i_t|$ if $Q^i_t \neq 0$ and $e^i_t = 0$ if $Q^i_t = 0$. As in Appendix C.4.2, the functions $\mathrm{rc}, \mathrm{sc} : \mathbb{R}^{2d} \to [0, 1)$ are Lipschitz continuous and satisfy $\mathrm{rc}^2 + \mathrm{sc}^2 \equiv 1$ and (C.44). We note that by Levy's characterization, for any solution of (C.74) the processes

$$
\begin{aligned}
B^i_t &:= \int_0^t \mathrm{sc}(Z^i_s, W^i_s)\mathrm{d}B^{i,\mathrm{sc}}_s + \int_0^t \mathrm{rc}(Z^i_s, W^i_s)\mathrm{d}B^{i,\mathrm{rc}}_s \\
\tilde{B}^i_t &:= \int_0^t \mathrm{sc}(Z^i_s, W^i_s)\mathrm{d}B^{i,\mathrm{sc}}_s + \int_0^t \mathrm{rc}(Z^i_s, W^i_s)(\mathrm{Id} - e^i_s {e^i_s}^T)\mathrm{d}B^{i,\mathrm{rc}}_s
\end{aligned}
$$

are d-dimensional Brownian motions. Therefore, (C.74) defines a coupling between N copies of solutions to (C.1) and a solution to (C.3). The processes $(\{Z^i_t\}_{i=1}^N)_{t\geq 0}$, $(\{W^i_t\}_{i=1}^N)_{t\geq 0}$ and $(\{Q^i_t\}_{i=1}^N)_{t\geq 0}$ satisfy the stochastic differential equations given by

$$
\begin{aligned}
\mathrm{d}Z^i_t &= W^i_t \mathrm{d}t = (Q^i_t - \gamma Z^i_t)\mathrm{d}t \\
\mathrm{d}W^i_t &= \Big(-\gamma W^i_t + u\Big(b(\bar{X}^i_t) - b(X^i_t) + \int_{\mathbb{R}^d} \tilde{b}(\bar{X}^i_t, z)\bar{\mu}^x_t(\mathrm{d}z) - N^{-1}\sum_{j=1}^N \tilde{b}(X^i_t, X^j_t)\Big)\Big)\mathrm{d}t \\
&\quad + \sqrt{8\gamma u}\,\mathrm{rc}(Z^i_t, W^i_t) e^i_t {e^i_t}^T \mathrm{d}B^{i,\mathrm{rc}}_t \\
\mathrm{d}Q^i_t &= \gamma^{-1} u\Big(b(\bar{X}^i_t) - b(X^i_t) + \int_{\mathbb{R}^d} \tilde{b}(\bar{X}^i_t, z)\bar{\mu}^x_t(\mathrm{d}z) - N^{-1}\sum_{j=1}^N \tilde{b}(X^i_t, X^j_t)\Big)\mathrm{d}t \\
&\quad + \sqrt{8\gamma^{-1} u}\,\mathrm{rc}(Z^i_t, W^i_t) e^i_t {e^i_t}^T \mathrm{d}B^{i,\mathrm{rc}}_t,
\end{aligned}
\tag{C.75}
$$

for all $i = 1, ..., N$.

The proof of Theorem C.17 relies on three auxiliary lemmata. We abbreviate $r^i_l(t) = r_l((\bar{X}^i_t, \bar{Y}^i_t), (X^i_t, Y^i_t))$, $r^i_s(t) = r_s((\bar{X}^i_t, \bar{Y}^i_t), (X^i_t, Y^i_t))$ and $\Delta^i(t) = \Delta((\bar{X}^i_t, \bar{Y}^i_t), (X^i_t, Y^i_t))$.

Lemma C.21. *Suppose Assumption C.2 and Assumption C.3 hold. Suppose that* (C.18) *holds. Let* $\tau > 0$ *be given by* (C.21). *Let* $(\{(\bar{X}^i_t, \bar{Y}^i_t), (X^i_t, Y^i_t)\}_{i=1}^N)_{t\geq 0}$ *be a solution to* (C.74). *Then for*

$i \in \{1, \dots, N\}$ with $\Delta^i(t) \ge D_{\mathcal{K}}$, it holds

$$\begin{aligned} \mathrm{d}r_l^i(t) \le -c_1 r_l^i(t)\mathrm{d}t + \frac{|(1-2\tau)Z_t^i + 2\gamma^{-1}W_t^i|}{2\gamma r_l^i(t)} u\Big(\tilde{L}N^{-1}\sum_{j=1}^N (|Z_t^j| + |Z_t^i|) + A_t^i\Big)\mathrm{d}t \\ + \sqrt{2\gamma^{-1}}\mathrm{rc}(Z_t^i, W_t^i)\frac{(1-2\tau)Z_t^i + 2\gamma^{-1}W_t^i}{r_l^i(t)} \cdot e_t^i {e_t^i}^T \mathrm{d}B_t^i, \end{aligned} \tag{C.76}$$

where $c_1 = \tau\gamma/2$ and $\{A_t^i\}_{i=1}^N$ is given by

$$A_t^i := \Big| \int_{\mathbb{R}^d} \tilde{b}(\bar{X}_t^i, z)\bar{\mu}_t^x(\mathrm{d}z) - N^{-1}\sum_{j=1}^N \tilde{b}(\bar{X}_t^i, \bar{X}_t^j)\Big| \qquad \textit{with } \bar{\mu}_t^x = \mathrm{Law}(\bar{X}_t^i). \tag{C.77}$$

Proof. By Ito's formula, it holds for $(\{Z_t^i, W_t^i\}_{i=1}^N)_{t\ge 0} = (\{\bar{X}_t^i - X_t^i, \bar{Y}_t^i - Y_t^i\}_{i=1}^N)_{t\ge 0}$,

$$\begin{cases} \mathrm{d}Z_t^i &= W_t^i \mathrm{d}t \\ \mathrm{d}W_t^i &= (-\gamma W_t^i + u(b(\bar{X}_t^i) - b(X_t^i) + N^{-1}\sum_{j=1}^N(\tilde{b}(\bar{X}_t^i, \bar{X}_t^j) - \tilde{b}(X_t^i, X_t^j)) + \tilde{A}_t^i))\mathrm{d}t \\ & \quad + \sqrt{8\gamma u}\mathrm{rc}(Z_t^i, W_t^i) e_t^i {e_t^i}^T \mathrm{d}B_t^i, \end{cases}$$

where

$$\tilde{A}_t^i := \Big(\int_{\mathbb{R}^d} \tilde{b}(\bar{X}_t^i, z)\bar{\mu}_t^x(\mathrm{d}z) - N^{-1}\sum_{j=1}^N \tilde{b}(\bar{X}_t^i, \bar{X}_t^j)\Big) \qquad \text{with } \bar{\mu}_t^x = \mathrm{Law}(\bar{X}_t^i)$$

for all $i = 1, ..., N$. Hence, by Ito's formula it holds for the positive matrices A, B, C given in (C.54),

$$\begin{aligned} &\mathrm{d}(Z_t^i \cdot (AZ_t^i) + Z_t^i \cdot (BW_t^i) + W_t^i \cdot (CW_t^i)) \\ &\le 2(AZ_t^i)\cdot W_t^i \mathrm{d}t + \Big(W_t^i \cdot (BW_t^i) - (BZ_t^i)\cdot(\gamma W_t^i + uKZ_t^i)\Big)\mathrm{d}t \\ &+ \Big(-2\gamma W_t^i \cdot (CW_t^i) - 2u(CW_t^i)\cdot(KZ_t^i) + 2L_g u|C^{1/2}W_t^i||C^{1/2}Z_t^i|\Big)\mathrm{d}t \\ &+ |BZ_t^i + 2CW_t^i| u\Big(\tilde{L}N^{-1}\sum_{j=1}^N(|Z_t^j| + |Z_t^i|) + A_t^i\Big)\mathrm{d}t + L_g u(1-2\tau)\gamma^{-1}|Z_t^i|^2 \cdot \mathbb{1}_{\{|Z_t^i|<R\}}\mathrm{d}t \\ &+ \gamma^{-2} 8\gamma u \mathrm{rc}(Z_t^i, W_t^i)^2 \mathrm{d}t + \sqrt{8\gamma u}\mathrm{rc}(Z_t^i, W_t^i)(BZ_t^i + CW_t^i)\cdot e_t^i {e_t^i}^T \mathrm{d}B_t^i \\ &\le Z_t^i \cdot ((-uKB + \gamma^{-1}L_g^2 u^2 C)Z_t^i)\mathrm{d}t + Z_t^i \cdot ((2A - \gamma B - 2uKC)W_t^i)\mathrm{d}t \\ &+ W_t^i \cdot ((B - \gamma C)W_t^i)\mathrm{d}t \\ &+ |BZ_t^i + 2CW_t^i| u\Big(\tilde{L}N^{-1}\sum_{j=1}^N(|Z_t^j| + |Z_t^i|) + A_t^i\Big)\mathrm{d}t + (1-2\tau)\gamma^{-1}L_g|Z_t^i|^2 \cdot \mathbb{1}_{\{|Z_t^i|<R\}}\mathrm{d}t \\ &+ 8\gamma^{-1}u(\mathrm{rc}(Z_t^i, W_t^i))^2\mathrm{d}t + \sqrt{8\gamma u}\mathrm{rc}(Z_t^i, W_t^i)(BZ_t^i + 2CW_t^i)\cdot e_t^i {e_t^i}^T \mathrm{d}B_t^i \end{aligned}$$

with $\{A_t^i\}_{i=1}^N$ given by (C.77). By (C.21) and (C.55),

$$\begin{aligned} \mathrm{d}r_l^i(t)^2 &= \mathrm{d}(Z_t^i \cdot (AZ_t^i) + Z_t^i \cdot (BW_t^i) + W_t^i \cdot (CW_t^i)) \\ &\le -2\tau\gamma r_l^i(t)^2 \mathrm{d}t + |(1-2\tau)Z_t^i + 2\gamma^{-1}W_t^i| \frac{u}{\gamma}\Big(\tilde{L}N^{-1}\sum_{j=1}^N(|Z_t^j| + |Z_t^i|) + A_t^i\Big)\mathrm{d}t \\ &+ 8\gamma^{-1}u\mathrm{rc}(Z_t^i, W_t^i)^2\mathrm{d}t + \gamma^{-1}(1-2\tau)L_g u|Z_t^i|^2 \cdot \mathbb{1}_{\{|Z_t^i|<R\}}\mathrm{d}t \\ &+ \sqrt{8\gamma^{-1}u}\mathrm{rc}(Z_t^i, W_t^i)((1-2\tau)Z_t^i + 2\gamma^{-1}W_t^i)\cdot e_t^i {e_t^i}^T \mathrm{d}B_t^i. \end{aligned}$$

Since $\Delta^i(t) \geq D_{\mathcal{K}}$, it holds $r_l^i(t)^2 > \mathcal{R}$ by (C.32) and (C.33). By (C.34) and (C.44),

$$-\tau\gamma r_l^i(t)^2 + \gamma^{-1}(1-2\tau)L_g u|Z_t^i|^2 \mathbb{1}_{\{|Z_t^i|<R\}} + 8\gamma^{-1}u\mathrm{rc}(Z_t^i, W_t^i)^2 \\ \leq -\tau\gamma\mathcal{R} + L_g u R^2\gamma^{-1} + 8\gamma^{-1}u\mathbb{1}_{\{R>0\}} \leq 0.$$

By Ito's formula and since the second derivative of the square root is negative,

$$\begin{aligned} \mathrm{d}r_l^i(t) &\leq (2r_l^i(t))^{-1}\mathrm{d}r_l^i(t)^2 \\ &\leq -c_1 r_l^i(t)\mathrm{d}t + \frac{|(1-2\tau)Z_t^i + 2\gamma^{-1}W_t^i|}{2\gamma r_l^i(t)} u\Big(\tilde{L}N^{-1}\sum_{j=1}^N(|Z_t^j| + |Z_t^i|) + A_t^i\Big)\mathrm{d}t \\ &+ \sqrt{2\gamma^{-1}}\mathrm{rc}(Z_t^i, W_t^i) r_l^i(t)^{-1}((1-2\tau)Z_t^i + 2\gamma^{-1}W_t^i)\cdot e_t^i {e_t^i}^T \mathrm{d}B_t^i, \end{aligned}$$

which concludes the proof. □

Lemma C.22. *Suppose Assumption C.2 and Assumption C.3 hold. Let* $(\{\bar{X}_t^i, \bar{Y}_t^i, X_t^i, Y_t^i\}_{i=1}^N)_{t\geq 0}$ *be a solution to* (C.74). *Let* r_s *be given in* (C.27) *with* α *defined in* (C.28). *If* $\Delta^i(t) < D_{\mathcal{K}}$ *with* $D_{\mathcal{K}}$ *given in* (C.32), *it holds*

$$\begin{aligned} \mathrm{d}f(r_s^i(t)) &\leq -c_2 f(r_s^i(t))\mathrm{d}t + \gamma^{-1}\tilde{L}uN^{-1}\sum_{j=1}^N(|Z_t^j| + |Z_t^i|)\mathrm{d}t - \frac{\alpha\gamma}{4}f'(R_1)|Z_t^i|\mathrm{d}t \\ &+ \gamma^{-1}u\Big|\int_{\mathbb{R}^d}\tilde{b}(\bar{X}_t^i, z)\bar{\mu}_t(\mathrm{d}z) - N^{-1}\sum_{j=1}^N \tilde{b}(\bar{X}_t^i, \bar{X}_t^j)\Big|\mathrm{d}t + (1+\alpha)\gamma\xi\mathrm{d}t + \mathrm{d}M_t^i, \end{aligned}$$

where f *is given in* (C.36), $(M_t^i)_{t\geq 0}$ *is a martingale and* c_2 *is given in* (C.56).

Proof. The proof works similarly as the proof of Lemma C.20. First, note that for all i, $(Z_t^i)_{t\geq 0}$ is almost surely continuously differentiable with derivative $\mathrm{d}Z^i/\mathrm{d}t = -\gamma Z^i + \gamma Q^i$ and hence $t \to |Z_t^i|$ is almost surely absolutely continuous with

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t}|Z_t^i| = \frac{Z_t^i}{|Z_t^i|}\cdot(-\gamma Z_t^i + \gamma Q_t^i) && \text{for a.e. } t \text{ such that } Z_t^i \neq 0 \text{ and} \\ &\frac{\mathrm{d}}{\mathrm{d}t}|Z_t^i| \leq \gamma|Q_t^i| && \text{for a.e. } t \text{ such that } Z_t^i = 0. \end{aligned}$$

and therefore

$$\frac{\mathrm{d}}{\mathrm{d}t}|Z_t^i| \leq -\gamma|Z_t^i| + \gamma|Q_t^i| \qquad \text{for a.e. } t \geq 0. \tag{C.78}$$

By Ito's formula and by Assumption C.2 and Assumption C.3, we obtain for $|Q_t^i|$,

$$\begin{aligned} \mathrm{d}|Q_t^i| &= \gamma^{-1}ue_t^i\cdot\Big(b(\bar{X}_t^i) - b(X_t^i) + \int_{\mathbb{R}^d}\tilde{b}(\bar{X}_t^i, z)\bar{\mu}_t^x(\mathrm{d}z) - N^{-1}\sum_{j=1}^N\tilde{b}(X_t^i, X_t^j)\Big)\mathrm{d}t \\ &+ \sqrt{8\gamma^{-1}u}\mathrm{rc}(Z_t^i, W_t^i){e_t^i}^T\mathrm{d}B_t^i \\ &\leq \gamma^{-1}u(L_K + L_g)|Z_t^i|\mathrm{d}t + \gamma^{-1}u(A_t^i + N^{-1}\sum_{j=1}^N\tilde{L}(|Z_t^j| + |Z_t^i|))\mathrm{d}t \\ &+ \sqrt{8\gamma^{-1}u}\mathrm{rc}(Z_t^i, W_t^i){e_t^i}^T\mathrm{d}B_t^{i,\mathrm{rc}}, \end{aligned}$$

where A_t^i is given by (C.77). Note that there is no Ito correction term, since $\partial^2_{q/|q|}|q| = 0$ for $q \neq 0$ and $\mathrm{rc} = 0$ for $Q_t = 0$. Combining this bound and (C.78) yields for $f(r_s^i(t))$ by Ito's formula,

$$\begin{aligned}\mathrm{d}f(r_s^i(t)) = f'(r_s^i(t))\Big(&((L_K + L_g)u\gamma^{-2} - \alpha)\gamma|Z_t^i| + \alpha\gamma|Q_t^i| \\ &+ \gamma^{-1}u\Big(A_t^i + N^{-1}\sum_{j=1}^N \tilde{L}(|Z_t^j| + |Z_t^i|)\Big)\Big)\mathrm{d}t \\ &+ f'(r_s^i(t))\sqrt{8\gamma^{-1}u}\mathrm{rc}(Z_t^i, W_t^i)(e_t^i)^T\mathrm{d}B_t^{i,\mathrm{rc}} + f''(r_s^i(t))4\gamma^{-1}u\mathrm{rc}(Z_t^i, W_t^i)^2\mathrm{d}t.\end{aligned}$$

Case 1: Consider $\Delta^i(t) < D_{\mathcal{K}}$ and $|Q_t^i| > \xi$, then $\mathrm{rc}(Z_t^i, W_t^i) = 1$ and $r_s^i(t) < R_1$. Hence, by (C.58) we obtain

$$\begin{aligned}\mathrm{d}f(r_s^i(t)) &\leq f'(r_s^i(t))\alpha\gamma r_s^i(t)\mathrm{d}t + f''(r_s^i(t))4\gamma^{-1}u\mathrm{d}t + \gamma^{-1}u\Big(A_t^i + N^{-1}\sum_{j=1}^N \tilde{L}(|Z_t^j| + |Z_t^i|)\Big)\mathrm{d}t \\ &\quad - f'(R_1)\frac{1}{2}\gamma\alpha|Z_t^i|\mathrm{d}t + \mathrm{d}M_t^i \\ &\leq -2\hat{c}f(r_s^i(t))\mathrm{d}t + \gamma^{-1}u\Big(A_t^i + N^{-1}\sum_{j=1}^N \tilde{L}(|Z_t^j| + |Z_t^i|)\Big)\mathrm{d}t - f'(R_1)\frac{\gamma\alpha}{2}|Z_t^i|\mathrm{d}t + \mathrm{d}M_t^i \\ &\leq -c_2 f(r_s^i(t))\mathrm{d}t + \gamma^{-1}u\Big(A_t^i + N^{-1}\sum_{j=1}^N \tilde{L}(|Z_t^j| + |Z_t^i|)\Big)\mathrm{d}t - f'(R_1)\frac{\gamma\alpha}{2}|Z_t^i|\mathrm{d}t + \mathrm{d}M_t^i.\end{aligned}$$

Case 2: Consider $\Delta^i(t) < D_{\mathcal{K}}$ and $|Q_t^i| \leq \xi$, then $\alpha|Z_t^i| = r_s^i(t) - |Q_t^i| \geq r_s^i(t) - \xi$. We note that

$$((L_K + L_g)u\gamma^{-2} - \alpha)|Z_t^i| + \alpha|Q_t^i| \leq -\frac{1}{2}r_s^i(t) + (1+\alpha)\xi.$$

Since the second derivative of f is negative and $\psi(s) \in [1/2, 1]$, it holds

$$\begin{aligned}\mathrm{d}f(r_s^i(t)) &\leq -\frac{\gamma}{2}r_s^i(t)f'(r_s(t))\mathrm{d}t + (1+\alpha)\gamma\xi\mathrm{d}t + \gamma^{-1}u\Big(A_t^i + N^{-1}\sum_{j=1}^N \tilde{L}(|Z_t^j| + |Z_t^i|)\Big)\mathrm{d}t + \mathrm{d}M_t^i \\ &\leq -\frac{\gamma}{8}\inf_{r \leq R_1}\frac{r\phi(r)}{\Phi(r)}f(r_s^i(t))\mathrm{d}t - \frac{\gamma\alpha}{4}|Z_t^i|f'(R_1)\mathrm{d}t + (1+\alpha)\gamma\xi\mathrm{d}t \\ &\quad + \gamma^{-1}u\Big(A_t^i + N^{-1}\sum_{j=1}^N \tilde{L}(|Z_t^j| + |Z_t^i|)\Big)\mathrm{d}t + \mathrm{d}M_t^i \\ &\leq -\frac{\gamma}{8}\frac{R_1\phi(R_1)}{\Phi(R_1)}f(r_s^i(t))\mathrm{d}t - \frac{\gamma\alpha}{4}|Z_t^i|f'(R_1)\mathrm{d}t + (1+\alpha)\gamma\xi\mathrm{d}t \\ &\quad + \gamma^{-1}u\Big(A_t^i + N^{-1}\sum_{j=1}^N \tilde{L}(|Z_t^j| + |Z_t^i|)\Big)\mathrm{d}t + \mathrm{d}M_t^i.\end{aligned}$$

Combining the two cases, we obtain the result by using the definition of c_2 given in (C.56). □

Lemma C.23. *(Moment control for Langevin dynamics) Suppose that Assumption C.2 and Assumption C.3 hold. Suppose that* (C.18) *and* (C.25) *hold. Let* $(\bar{X}_t, \bar{Y}_t)_{t\geq 0}$ *be a solution to* (C.1) *with* $\mathbb{E}[|\bar{X}_0|^2 + |\bar{Y}_0|^2] \leq \infty$. *Then there exists a finite constant* $\mathcal{C}_2 > 0$ *such that*

$$\sup_{t\geq 0}\mathbb{E}[|\bar{X}_t|^2] \leq \mathcal{C}_2.$$

The constant $\mathcal{C}_2$ depends on γ, $\mathbb{E}[|\bar{X}_0|^2 + |\bar{Y}_0|^2]$, d, R, κ, L_g, u and $\tilde{L}$.

Proof. We adapt the proof idea from [75, Lemma 8]. By Ito's formula, by Assumption C.2 and by Assumption C.3, it holds

$$\begin{aligned}
&\mathrm{d}(\gamma^{-2}u\bar{X}_t\cdot(K\bar{X}_t)+\frac{1}{2}|(1-2\tau)\bar{X}_t+\gamma^{-1}\bar{Y}_t|^2+\frac{1}{2}\gamma^{-2}|\bar{Y}_t|^2)\\
&\leq\Big(2\gamma^{-2}u\bar{X}_t\cdot(K\bar{Y}_t)+(1-2\tau)^2\bar{X}_t\cdot\bar{Y}_t+\gamma^{-1}(1-2\tau)|\bar{Y}_t|^2\Big)\mathrm{d}t\\
&+\gamma^{-1}(1-2\tau)\Big(-u\bar{X}_t\cdot(K\bar{X}_t)-\gamma\bar{X}_t\cdot\bar{Y}_t\Big)\mathrm{d}t+2\gamma^{-2}\Big(-u(K\bar{Y}_t)\cdot\bar{X}_t+L_g|\bar{Y}_t||\bar{X}_t|-\gamma|\bar{Y}_t|^2\Big)\mathrm{d}t\\
&+\frac{u}{\gamma}|(1-2\tau)\bar{X}_t+2\gamma^{-1}\bar{Y}_t|\Big(\tilde{L}(\mathbb{E}[|\bar{X}_t|]+|X_t|)+|\tilde{b}(0,0)|\Big)\mathrm{d}t\\
&+(1-2\tau)\gamma^{-1}u(L_g|\bar{X}_t|^2+|g(0)||\bar{X}_t|)\mathbb{1}_{\{|\bar{X}_t|<R\}}\mathrm{d}t+2\gamma^{-2}u|\bar{Y}_t||g(0)|\mathrm{d}t+2\gamma^{-1}ud\mathrm{d}t\\
&+\sqrt{2\gamma^{-1}u}((1-2\tau)\bar{X}_t+2\gamma^{-1}\bar{Y}_t)\mathrm{d}B_t\\
&\leq-\gamma^{-1}u(1-2\tau)\bar{X}_t\cdot(K\bar{X}_t)-2\tau\gamma(\gamma^{-2}|\bar{Y}_t|^2+(1-2\tau)\gamma^{-1}\bar{X}_t\cdot\bar{Y}_t)+\gamma^{-3}u^2L_g^2|\bar{X}_t|^2\\
&+\frac{u}{\gamma}|(1-2\tau)\bar{X}_t+2\gamma^{-1}\bar{Y}_t|(\tilde{L}(\mathbb{E}[|\bar{X}_t|]+|X_t|)+|\tilde{b}(0,0)|)\mathrm{d}t\\
&+(1-2\tau)\gamma^{-1}u(L_g|\bar{X}_t|^2+|g(0)||\bar{X}_t|)\mathbb{1}_{\{|\bar{X}_t|<R\}}\mathrm{d}t\\
&+2\gamma^{-2}u|\bar{Y}_t||g(0)|\mathrm{d}t+2\gamma^{-1}ud\mathrm{d}t+\sqrt{2\gamma^{-1}u}((1-2\tau)\bar{X}_t+2\gamma^{-1}\bar{Y}_t)\mathrm{d}B_t
\end{aligned}$$

Taking expectation, we obtain

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\gamma^{-2}u\bar{X}_t\cdot(K\bar{X}_t)+\frac{1}{2}|(1-2\tau)\bar{X}_t+\gamma^{-1}\bar{Y}_t|^2+\frac{1}{2}\gamma^{-2}|\bar{Y}_t|^2]\\
&\quad\leq-\gamma^{-1}u(1-2\tau)\mathbb{E}[\bar{X}_t\cdot(K\bar{X}_t)]+\gamma^{-3}u^2L_g^2\mathbb{E}[|\bar{X}_t|^2]\\
&\quad-2\tau\gamma\Big(\gamma^{-2}\mathbb{E}[|\bar{Y}_t|^2]+(1-2\tau)\gamma^{-1}\mathbb{E}[\bar{X}_t\cdot\bar{Y}_t]\Big)+(1-2\tau)\gamma^{-1}u(L_gR^2+R|g(0)|)+2\gamma^{-1}ud\\
&\quad+u\gamma^{-1}\mathbb{E}\Big[|(1-2\tau)\bar{X}_t+2\gamma^{-1}\bar{Y}_t|\Big(\tilde{L}(\mathbb{E}[|X_t|]+|X_t|)+|\tilde{b}(0,0)|\Big)\Big]+2\gamma^{-2}u\mathbb{E}[|\bar{Y}_t|]|g(0)|.
\end{aligned}$$

We note that by (C.25) and by Young's inequality,

$$\begin{aligned}
&\gamma^{-1}\mathbb{E}[|(1-2\tau)\bar{X}_t+2\gamma^{-1}\bar{Y}_t|u(\tilde{L}(\mathbb{E}[|\bar{X}_t|]+|\bar{X}_t|)+|\tilde{b}(0,0)|)]\\
&\quad\leq\frac{\tau\sqrt{\kappa u}}{8}\mathbb{E}[|(1-2\tau)\bar{X}_t+2\gamma^{-1}\bar{Y}_t|(\mathbb{E}[|\bar{X}_t|]+|\bar{X}_t|)]+\gamma^{-1}u\mathbb{E}[|(1-2\tau)\bar{X}_t+2\gamma^{-1}\bar{Y}_t|]|\tilde{b}(0,0)|\\
&\quad\leq\frac{\tau\gamma}{4}\Big(\kappa u\gamma^{-2}\mathbb{E}[|\bar{X}_t|^2]+\frac{1}{4}\mathbb{E}[|(1-2\tau)\bar{X}_t+2\gamma^{-1}\bar{Y}_t|^2]\Big)+\frac{\tau\gamma}{4}\frac{1}{4}\mathbb{E}[|(1-2\tau)\bar{X}_t+2\gamma^{-1}\bar{Y}_t|^2]\\
&\quad+\frac{4u^2}{\tau\gamma^3}|\tilde{b}(0,0)|^2\\
&\quad\leq\frac{\tau\gamma}{2}\Big(\kappa u\gamma^{-2}\mathbb{E}[|\bar{X}_t|^2]+\frac{1}{2}\mathbb{E}[|(1-2\tau)\bar{X}_t+\gamma^{-1}\bar{Y}_t|^2]+\frac{1}{2}\mathbb{E}[|\gamma^{-1}\bar{Y}_t|^2]\Big)+\frac{4u^2}{\tau\gamma^3}|\tilde{b}(0,0)|^2
\end{aligned}$$

and

$$2\gamma^{-2}u\mathbb{E}[|\bar{Y}_t|]|g(0)|\leq\frac{\tau\gamma}{2}\frac{1}{2}\mathbb{E}[|\gamma^{-1}\bar{Y}_t|^2]+\frac{4u^2}{\tau\gamma^3}|g(0)|^2.$$

Then by (C.55),

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\Big[\gamma^{-2}u\bar{X}_t\cdot(K\bar{X}_t)+\frac{1}{2}|(1-2\tau)\bar{X}_t+\gamma^{-1}\bar{Y}_t|^2+\frac{1}{2}\gamma^{-2}|\bar{Y}_t|^2\Big]\\
&\quad\le -2\tau\gamma\mathbb{E}\Big[\gamma^{-2}u\bar{X}_t\cdot(K\bar{X}_t)+\frac{1}{2}|(1-2\tau)\bar{X}_t+\gamma^{-1}\bar{Y}_t|^2+\frac{1}{2}\gamma^{-2}|\bar{Y}_t|^2\Big]+(1-2\tau)\gamma^{-1}uL_gR^2\\
&\qquad+2\gamma^{-1}ud+\tau\gamma\Big(\kappa\gamma^{-2}\mathbb{E}[|\bar{X}_t|^2]+\frac{1}{2}\mathbb{E}[|(1-2\tau)\bar{X}_t+\gamma^{-1}\bar{Y}_t|^2+|\gamma^{-1}\bar{Y}_t|^2]\Big)\\
&\qquad+4\tau^{-1}\gamma^{-3}u^2(|\tilde{b}(0,0)|^2+|g(0)|^2)\\
&\quad\le -\tau\gamma\mathbb{E}\Big[\gamma^{-2}u\bar{X}_t\cdot(K\bar{X}_t)+\frac{1}{2}|(1-2\tau)\bar{X}_t+\gamma^{-1}\bar{Y}_t|^2+\frac{1}{2}\gamma^{-2}|\bar{Y}_t|^2\Big]+(1-2\tau)\gamma^{-1}L_guR^2\\
&\qquad+2\gamma^{-1}ud+4\tau^{-1}\gamma^{-3}u^2(|\tilde{b}(0,0)|^2+|g(0)|^2).
\end{aligned}$$

By Grönwall's inequality, there exists a constant $\mathbf{C}$ such that

$$\sup_{t\ge0}\mathbb{E}\Big[\gamma^{-2}u\bar{X}_t\cdot(K\bar{X}_t)+\frac{1}{2}|(1-2\tau)\bar{X}_t+\gamma^{-1}\bar{Y}_t|^2+\frac{1}{2}\gamma^{-2}|\bar{Y}_t|^2\Big]\le\mathbf{C}<\infty.$$

Thus, we obtain the result for $\mathcal{C}_2=\mathbf{C}/(\kappa u\gamma^{-2})$. □

Proof of Theorem C.17. To prove uniform in time propagation of chaos, we consider the coupling $(\{(\bar{X}_t^i,\bar{Y}_t^i),(X_t^i,Y_t^i)\}_{i=1}^N)_{t\ge0}$ given in (C.74) and combine the results of Lemma C.21 and Lemma C.22. The second moment control given in Lemma C.23 will be essential to bound the terms involving the non-linearity. We write here $r_s^i(t)=r_s^i((\bar{X}_t^i,\bar{Y}_t^i),(X_t^i,Y_t^i))$, $r_l^i(t)=r_l^i((\bar{X}_t^i,\bar{Y}_t^i),(X_t^i,Y_t^i))$, $\Delta^i(t)=r_s^i(t)-\epsilon r_l^i(t)$ and $\rho^i(t)=f((\Delta^i(t)\wedge D_{\mathcal{K}})+\epsilon r_l^i(t))$. We distinguish two cases for all particles $i=1,...,N$:
Case 1: Consider $\Delta^i(t)<D_{\mathcal{K}}$. Then $\rho^i(t)=f(r_s^i(t))$, and by Lemma C.22 it holds for $\xi>0$

$$\begin{aligned}
\mathrm{d}\rho^i(t)=\mathrm{d}f(r_s^i(t))&\le -c_2f(r_s^i(t))\mathrm{d}t+\gamma^{-1}u\Big(A_t^i+N^{-1}\sum_{j=1}^N\tilde{L}(|Z_t^j|+|Z_t^i|)\Big)\mathrm{d}t-\frac{\alpha\gamma}{4}f'(R_1)|Z_t^i|\mathrm{d}t\\
&\quad+(1+\alpha)\gamma\xi\mathrm{d}t+\mathrm{d}M_t^i\\
&\le -c_2f(r_s^i(t))\mathrm{d}t+\gamma^{-1}u\Big(A_t^i+N^{-1}\sum_{j=1}^N\tilde{L}|Z_t^j|\Big)\mathrm{d}t-\frac{\alpha\gamma}{8}f'(R_1)|Z_t^i|\mathrm{d}t+(1+\alpha)\gamma\xi\mathrm{d}t+\mathrm{d}M_t^i,
\end{aligned}\tag{C.79}$$

where A_t^i is given in (C.77) and c_2 is given by (C.56). Note the last step holds by (C.25).
Case 2: Consider $\Delta^i(t)\ge D_{\mathcal{K}}$. We obtain by Lemma C.21,

$$\begin{aligned}
\mathrm{d}r_l^i(t)&\le -c_1r_l^i(t)\mathrm{d}t+\frac{|(1-2\tau)Z_t^i+2\gamma^{-2}W_t^i|}{2\gamma r_l^i(t)}u\Big(A_t^i+N^{-1}\sum_{j=1}^N\tilde{L}(|Z_t^j|+|Z_t^i|)\Big)\mathrm{d}t\\
&\quad+\sqrt{2\gamma^{-1}u}\mathrm{rc}(Z_t^i,W_t^i)r_l^i(t)^{-1}((1-2\tau)Z_t^i+2\gamma^{-1}W_t^i)\cdot e_t^ie_t^{iT}\mathrm{d}B_t^i
\end{aligned}$$

with c_1 given in Lemma C.21. Note that $\frac{\mathrm{d}}{\mathrm{d}x}f(D_{\mathcal{K}}+\epsilon x)=\epsilon f'(D_{\mathcal{K}}+\epsilon x)$. Further, since $f(D_{\mathcal{K}}+\epsilon x)$

is a concave function, $\frac{\mathrm{d}^2}{\mathrm{d}x^2} f(D_{\mathcal{K}} + \epsilon x)$ is negative. By Ito's formula, we obtain

$$\begin{aligned}
\mathrm{d}\rho^i(t) &= \mathrm{d}f(D_{\mathcal{K}} + \epsilon r_l^i(t)) \\
&\leq \epsilon f'(D_{\mathcal{K}} + \epsilon r_l^i(t))\Big(- c_1 r_l^i(t)^2 + \frac{|(1-2\tau)Z_t^i + 2\gamma^{-1}W_t^i|}{2\gamma r_l^i(t)} u\Big(A_t^i + N^{-1}\sum_{j=1}^N \tilde{L}(|Z_t^j| + |Z_t^i|)\Big)\Big)\mathrm{d}t \\
&\quad + \frac{\epsilon f'(D_{\mathcal{K}} + \epsilon r_l^i(t))}{r_l^i(t)} \sqrt{2\gamma^{-1} u}\mathrm{rc}(Z_t^i, W_t^i)((1-2\tau)Z_t^i + 2\gamma^{-1}W_t^i)\cdot e_t^i {e_t^i}^T \mathrm{d}B_t^i.
\end{aligned}$$

By (C.66) and (C.67), which holds in the same line as in the proof of Theorem C.12, it holds

$$\begin{aligned}
\mathrm{d}\rho^i(t) &\leq -f'(R_1)\frac{c_1\epsilon}{2}\min\Big(\frac{\sqrt{\kappa u}\gamma^{-1}}{\sqrt{8}\alpha}, \frac{1}{2}\Big)\rho^i(t)\mathrm{d}t - f'(R_1)\frac{c_1\epsilon}{2}\sqrt{\kappa u \gamma^{-2}}|Z_t^i|\mathrm{d}t \\
&\quad + 2\epsilon\gamma^{-1}u\Big(A_t^i + N^{-1}\sum_{j=1}^N \tilde{L}(|Z_t^j| + |Z_t^i|)\Big)\mathrm{d}t + \mathrm{d}M_t^i,
\end{aligned} \tag{C.80}$$

where $(\{M_t^i\}_{i=1}^N)_{t\geq 0}$ is some martingale.

Combining (C.79) and (C.80), taking expectations and summing over $i = 1, \ldots, N$ yields

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\Big[N^{-1}\sum_{i=1}^N \rho^i(t)\Big] \\
&\leq -\min\Big(c_2, f'(R_1)\frac{c_1\epsilon}{2}\min\Big(\frac{\sqrt{\kappa u}\gamma^{-1}}{\sqrt{8}\alpha}, \frac{1}{2}\Big)\Big)\mathbb{E}\Big[N^{-1}\sum_{i=1}^N \rho^i(t)\Big] + \gamma^{-1}u\mathbb{E}\Big[N^{-1}\sum_{i=1}^N A_t^i\Big] \\
&\quad - \min\Big(f'(R_1)\frac{\gamma\alpha}{8}, f'(R_1)\frac{c_1\epsilon}{2}\sqrt{\kappa u\gamma^{-2}}\Big)\mathbb{E}\Big[N^{-1}\sum_{i=1}^N |Z_t^i|\Big] + \tilde{L}u\gamma^{-1}\mathbb{E}\Big[N^{-1}\sum_{i=1}^N |Z_t^i|\Big] \\
&\leq -\min\Big(c_2, f'(R_1)\frac{c_1\epsilon}{4}\min\Big(\frac{\sqrt{\kappa u}\gamma^{-1}}{\sqrt{8}\alpha}, \frac{1}{2}\Big)\Big)\mathbb{E}\Big[N^{-1}\sum_{i=1}^N \rho^i(t)\Big] + \gamma^{-1}u\mathbb{E}\Big[N^{-1}\sum_{i=1}^N A_t^i\Big],
\end{aligned} \tag{C.81}$$

where we used $2\epsilon \leq 1$ for the last term and (C.25).

To bound $\mathbb{E}[A_t^i]$, we note that given $\bar{X}_t^i$, $\bar{X}_t^j$, $j \neq i$ are identically and independent distributed with law $\bar{\mu}_t^x$ and

$$\mathbb{E}[\tilde{b}(\bar{X}_t^i, \bar{X}_t^j)|\bar{X}_t^i] = \int_{\mathbb{R}^d} \tilde{b}(\bar{X}_t^i, z)\bar{\mu}_t^x(\mathrm{d}z). \tag{C.82}$$

Hence,

$$\begin{aligned}
&\mathbb{E}\Big[|\int_{\mathbb{R}^d}\tilde{b}(\bar{X}_t^i, z)\bar{\mu}_t^x(\mathrm{d}z) - \frac{1}{N}\sum_{j=1}^N \tilde{b}(\bar{X}_t^i, \bar{X}_t^j)|^2\Big|\bar{X}_t^i\Big] \\
&= \frac{N-1}{N^2}\mathrm{Var}_{\bar{\mu}_t^x}(\tilde{b}(\bar{X}_t^i, \cdot)) + \frac{1}{N^2}\mathbb{E}\Big[|\int_{\mathbb{R}^d}\tilde{b}(\bar{X}_t^i, z)\bar{\mu}_t^x(\mathrm{d}z) - \tilde{b}(\bar{X}_t^i, \bar{X}_t^i)|^2\Big|\bar{X}_t^i\Big] \\
&\quad + \frac{2}{N^2}\sum_{j=1, j\neq i}^N \mathbb{E}\Big[|\int_{\mathbb{R}^d}\tilde{b}(\bar{X}_t^i, z)\bar{\mu}_t^x(\mathrm{d}z) - \tilde{b}(\bar{X}_t^i, \bar{X}_t^j)|\cdot|\int_{\mathbb{R}^d}\tilde{b}(\bar{X}_t^i, z)\bar{\mu}_t^x(\mathrm{d}z) - \tilde{b}(\bar{X}_t^i, \bar{X}_t^i)|\Big|\bar{X}_t^i\Big]
\end{aligned}$$

By Assumption C.3, Cauchy inequality and Young's inequality

$$\mathbb{E}\Big[\Big|\int_{\mathbb{R}^d}\tilde{b}(\bar{X}_t^i,z)\bar{\mu}_t^x(\mathrm{d}z)-\frac{1}{N}\sum_{j=1}^N\tilde{b}(\bar{X}_t^i,\bar{X}_t^j)\Big|^2\Big]\leq\frac{4\tilde{L}^2}{N}\int_{\mathbb{R}^d}|x|^2\bar{\mu}_t^x(\mathrm{d}x)+\frac{4\tilde{L}^2}{N^2}\int_{\mathbb{R}^d}|x|^2\bar{\mu}_t^x(\mathrm{d}x)+\frac{8\tilde{L}^2}{N}\int_{\mathbb{R}^d}|x|^2\bar{\mu}_t^x(\mathrm{d}x). \tag{C.83}$$

Then, by Jensen's inequality

$$\mathbb{E}[A_t^i]\leq\frac{4\tilde{L}}{N^{1/2}}\Big(\int_{\mathbb{R}^d}|x|^2\bar{\mu}_t^x(\mathrm{d}x)\Big)^{1/2}.$$

By Lemma C.23, there exists a finite constant $\mathcal{C}_1$ such that for $N\geq 2$ and all $i=1,...,N$,

$$\sup_{t\geq 0}\mathbb{E}[A_t^i]\leq\gamma u^{-1}\mathcal{C}_1N^{-1/2}. \tag{C.84}$$

Note that $\mathcal{C}_1$ depends on γ, $\mathbb{E}[|\bar{X}_0|^2+|\bar{Y}_0|^2]$, d, u, R, κ, L_g and $\tilde{L}$. Inserting the bound for $\mathbb{E}[A_t^i]$ in (C.81) yields

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\Big[N^{-1}\sum_{i=1}^N\rho^i(t)\Big]\leq-\min\Big(c_2,f'(R_1)\frac{c_1\epsilon}{2}\min\Big(\frac{\sqrt{\kappa u}\gamma^{-1}}{\sqrt{8}\alpha},\frac{1}{2}\Big)\Big)\mathbb{E}\Big[N^{-1}\sum_{i=1}^N\rho^i(t)\Big]+\frac{\mathcal{C}_1}{N^{1/2}}.$$

Applying Grönwall's inequality and (C.70) and (C.71) yields

$$\mathcal{W}_{1,\rho_N}(\bar{\mu}_t^{\otimes N},\mu_t^N)\leq\mathbb{E}\Big[N^{-1}\sum_{i=1}^N\rho^i(t)\Big]\leq e^{-\tilde{c}t}\mathbb{E}\Big[N^{-1}\sum_{i=1}^N\rho^i(0)\Big]+\mathcal{C}_1N^{-1/2}\tilde{c}^{-1}.$$

with $\tilde{c}$ given in (C.72). Taking the infimum over all couplings $\omega\in\Pi(\bar{\mu}_0^{\otimes},\mu_0^N)$ concludes the proof of the first result.

The second bound holds by (C.47) with M_1 given in (C.24) and $M_2=\sqrt{2}/\mathbf{C}_1$ given in (C.48).

□

C.7 Appendix: Unconfined nonlinear Langevin dynamics

C.7.1 Contraction for unconfined nonlinear Langevin dynamics

Consider the unconfined nonlinear Langevin dynamics given by

$$\begin{cases}\mathrm{d}\bar{X}_t=\bar{Y}_t\mathrm{d}t\\ \mathrm{d}\bar{Y}_t=(-\gamma\bar{Y}_t+u\int_{\mathbb{R}^d}\tilde{b}(\bar{X}_t,z)\bar{\mu}_t^x(\mathrm{d}z))\mathrm{d}t+\sqrt{2\gamma u}\mathrm{d}B_t,\qquad(\bar{X}_0,\bar{Y}_0)\sim\bar{\mu}_0,\end{cases} \tag{C.85}$$

where $\gamma,u>0$, $\bar{\mu}_0$ is a probability measure on $\mathbb{R}^{2d}$, $\bar{\mu}_t^x=\mathrm{Law}(\bar{X}_t)$ and $(B_t)_{t\geq 0}$ is a d-dimensional standard Brownian motion. We impose for the function $\tilde{b}$ and for the initial distribution:

Assumption C.4. *The function $\tilde{b}:\mathbb{R}^{2d}\to\mathbb{R}^d$ is Lipschitz continuous, and there exist a function $\tilde{g}:\mathbb{R}^d\to\mathbb{R}^d$ and a positive definite matrix $\tilde{K}\in\mathbb{R}^{d\times d}$ with smallest eigenvalue $\tilde{\kappa}\in(0,\infty)$ and largest eigenvalue $L_{\tilde{K}}\in(0,\infty)$ such that*

$$\tilde{b}(x,y)=-\tilde{K}(x-y)+\tilde{g}(x-y)\qquad\text{for all } x,y\in\mathbb{R}^d,$$

and $\tilde{g}$ is Lipschitz continuous with Lipschitz constant $L_{\tilde{g}}\in(0,\infty)$ and anti-symmetric, i.e., $\tilde{g}(-z)=-\tilde{g}(z)$ for all $z\in\mathbb{R}^d$.

Assumption C.5. *Let* $\bar{\mu}_0 \in \mathcal{P}(\mathbb{R}^{2d})$ *satisfy* $\int_{\mathbb{R}^{2d}} |(x,y)|^2 \bar{\mu}_0(\mathrm{d}x\mathrm{d}y) < \infty$ *and* $\int_{\mathbb{R}^{2d}} (x,y)\bar{\mu}_0(\mathrm{d}x\mathrm{d}y) = 0$.

By Assumption C.4, it holds $\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[(X_t,Y_t)] = \mathbb{E}[(Y_t,-\gamma Y_t)]$ and hence by Assumption C.5 $\mathbb{E}[(X_t,Y_t)] = 0$ for all $t \geq 0$. Note that this observation is crucial in our analysis, since in general convergence to equilibrium can not be guaranteed for the unconfined dynamics unless the solution is centered or a recentering of the center of mass is considered.

We establish contraction in Wasserstein distance with respect to the distance function $\tilde{r} : \mathbb{R}^{2d} \times \mathbb{R}^{2d} \to [0,\infty)$ given by

$$\begin{aligned}\tilde{r}((x,y),(\bar{x},\bar{y}))^2 = \gamma^{-2}u(x-\bar{x})\cdot(\tilde{K}(x-\bar{x})) + \frac{1}{2}|(1-2\sigma)(x-\bar{x}) + \gamma^{-1}(y-\bar{y})|^2 \\ + \frac{1}{2}\gamma^{-2}|y-\bar{y}|^2,\end{aligned} \tag{C.86}$$

for $(x,y),(\bar{x},\bar{y}) \in \mathbb{R}^{2d}$ where σ is given by

$$\sigma = \min(1/8, \tilde{\kappa}u\gamma^{-2}/2). \tag{C.87}$$

Theorem C.24 (Contraction for nonlinear unconfined Langevin dynamics in L^2 and L^1 Wasserstein distance)**.** *Suppose Assumption C.4 holds. Let* $\bar{\mu}_0$ *and* $\bar{\nu}_0$ *be two probability distributions on* $\mathbb{R}^{2d}$ *satisfying Assumption C.5. For* $t \geq 0$, *let* $\bar{\mu}_t$ *and* $\bar{\nu}_t$ *be the law of the processes* $(\bar{X}_t, \bar{Y}_t)$ *and* $(\bar{X}'_t, \bar{Y}'_t)$, *respectively, where* $(\bar{X}_s,\bar{Y}_s)_{s\geq 0}$ *and* $(\bar{X}'_s,\bar{Y}'_s)_{s\geq 0}$ *are solutions to* (C.85) *with initial distribution* $\bar{\mu}_0$ *and* $\bar{\nu}_0$, *respectively. If*

$$L_{\tilde{g}} \leq \sqrt{\tilde{\kappa}/u}(\gamma/2)\min(1/8, \tilde{\kappa}u\gamma^{-2}/2), \tag{C.88}$$

then

$$\mathcal{W}_{2,\tilde{r}}(\bar{\mu}_t,\bar{\nu}_t) \leq e^{-\hat{c}t}\mathcal{W}_{2,\tilde{r}}(\bar{\mu}_0,\bar{\nu}_0) \qquad \text{and} \qquad \mathcal{W}_2(\bar{\mu}_t,\bar{\nu}_t) \leq M_3 e^{-\hat{c}t}\mathcal{W}_2(\bar{\mu}_0,\bar{\nu}_0), \tag{C.89}$$

where $\tilde{r}$ *is defined in* (C.86) *and where the contraction rate* $\hat{c}$ *is given by*

$$\hat{c} = \min(\gamma/16, \tilde{\kappa}\gamma^{-1}/4). \tag{C.90}$$

The constant M_3 *is given by*

$$M_3 = \max(\sqrt{L_{\tilde{K}}u + \gamma^2}, \sqrt{3/2})\max(\sqrt{(\tilde{\kappa}u)^{-1}}, \sqrt{2}). \tag{C.91}$$

Moreover, there exists a unique invariant probability measure $\bar{\mu}_\infty$ *for* (C.85) *and convergence in* L^2 *Wasserstein distance to* $\bar{\mu}_\infty$ *holds.*

If

$$L_{\tilde{g}} \leq \sqrt{\tilde{\kappa}/u}(\gamma/4)\min(1/8, \tilde{\kappa}\gamma^{-2}/2), \tag{C.92}$$

then

$$\mathcal{W}_{1,\tilde{r}}(\bar{\mu}_t,\bar{\nu}_t) \leq e^{-\hat{c}t}\mathcal{W}_{1,\tilde{r}}(\bar{\mu}_0,\bar{\nu}_0) \qquad \text{and} \qquad \mathcal{W}_1(\bar{\mu}_t,\bar{\nu}_t) \leq M_3 e^{-\hat{c}t}\mathcal{W}_1(\bar{\mu}_0,\bar{\nu}_0) \tag{C.93}$$

and convergence in L^1 *Wasserstein distance to* $\bar{\mu}_\infty$ *holds.*

Proof. The proof uses a synchronous coupling and is postponed to Appendix C.7.3. □

Remark C.25. Note that (C.89) implies directly a bound in L^p Wasserstein distance for $1 \leq p < 2$, i.e., by Jensen's inequality it holds $\mathcal{W}_p(\bar{\mu}_t, \bar{\nu}_t) \leq \mathcal{W}_2(\bar{\mu}_t, \bar{\nu}_t) \leq M_0 M_3 e^{-\hat{c}t} \mathcal{W}_p(\bar{\mu}_0, \bar{\nu}_0)$, where $M_0 = \mathcal{W}_2(\bar{\mu}_0, \bar{\nu}_0)/\mathcal{W}_p(\bar{\mu}_0, \bar{\nu}_0)$. The additional constant M_0 is finite by Assumption C.5, but it might be very large. Here, contraction in L^1 Wasserstein distance is stated separately and (C.93) is proven directly.

Remark C.26. By (C.88) and (C.92), it holds $L_{\tilde{g}} \leq \tilde{\kappa}/8$ and $L_{\tilde{g}} \leq \tilde{\kappa}/16$, respectively. Hence, contraction is proven for $\tilde{b}$ being a small perturbation of a linear function. Further, the contraction rate is maximized for $\gamma = 2\sqrt{\tilde{\kappa} u}$.

Remark C.27. Note that the underlying distance $\tilde{r}$ is defined similarly as r_l in (C.26) and coincides with ρ defined in (C.35) if $\tilde{K} = K$, $\sigma = \tau$ and $\mathcal{K} = \{(0,0)\}$. Moreover, $\tilde{r}$ is equivalent to the Euclidean distance on $\mathbb{R}^{2d}$, i.e.,

$$\begin{aligned}\min(\tilde{\kappa}u/2, 1/4)\gamma^{-2}(|x-\bar{x}|+|y-\bar{y}|)^2 &\leq \min(\tilde{\kappa}u, 1/2)\gamma^{-2}|(x,y)-(\bar{x},\bar{y})|^2 \leq \tilde{r}((x,y),(\bar{x},\bar{y}))^2 \\ &\leq \max(L_{\tilde{K}}u\gamma^{-2}+1, (3/2)\gamma^{-2})|(x,y)-(\bar{x},\bar{y})|^2 \\ &\leq \max(L_{\tilde{K}}u\gamma^{-2}+1, (3/2)\gamma^{-2})(|x-\bar{x}|+|y-\bar{y}|)^2. \end{aligned} \tag{C.94}$$

C.7.2 Uniform in time propagation of chaos in the unconfined case

Next, we establish uniform in time propagation of chaos bounds for the unconfined Langevin dynamics. Fix $N \in \mathbb{N}$. We consider the functions $\hat{\rho}_N, \tilde{\rho}_N : \mathbb{R}^{2Nd} \times \mathbb{R}^{2Nd} \to [0, \infty)$ given by

$$\hat{\rho}_N((x,y),(\bar{x},\bar{y}))^2 := N^{-1}\sum_{i=1}^N \tilde{r}(\pi(x,y),\pi(\bar{x},\bar{y}))^2, \qquad \text{and} \tag{C.95}$$

$$\tilde{\rho}_N((x,y),(\bar{x},\bar{y})) := N^{-1}\sum_{i=1}^N \tilde{r}(\pi(x,y),\pi(\bar{x},\bar{y})) \qquad \text{for all } x,y,\bar{x},\bar{y} \in \mathbb{R}^{Nd}, \tag{C.96}$$

where $\tilde{r}$ is given in (C.86) and $\pi : \mathbb{R}^{2Nd} \to \mathbb{R}^{2Nd}$ is given by

$$\pi(x,y) = \Big(x^i - N^{-1}\sum_{j=1}^N x^j, y^i - N^{-1}\sum_{j=1}^N y^j\Big)_{i=1}^N \qquad \text{for } (x,y) \in \mathbb{R}^{2Nd}. \tag{C.97}$$

The function π defines a projection from $\mathbb{R}^{2Nd}$ to the hyperplane $\mathsf{H}^N = \{(x,y) \in \mathbb{R}^{2Nd} : (\sum_i x^i, \sum_i y^i) = 0\}$. We note that distances $\hat{\rho}_N$ and $\tilde{\rho}_N$ are equivalent to $\tilde{\ell}^p_N$ given by

$$\tilde{\ell}^p_N((x,y),(\bar{x},\bar{y})) = \ell^p_N(\pi(x,y),\pi(\bar{x},\bar{y})), \qquad \text{for all } x,y,\bar{x},\bar{y} \in \mathbb{R}^{Nd}, \tag{C.98}$$

with $p = 1$ and $p = 2$, respectively.

Theorem C.28 (Propagation of chaos for unconfined Langevin dynamics in L^2 and L^1 Wasserstein distance)**.** *Suppose Assumption C.4 holds. Let $\bar{\mu}_0$ and μ_0 be two probability distributions on $\mathbb{R}^{2d}$ satisfying Assumption C.5. For $t \geq 0$, let $\bar{\mu}_t$ be the law of the process $(\bar{X}_t, \bar{Y}_t)$, where $(\bar{X}_s, \bar{Y}_s)_{s\geq 0}$ is a solution to* (C.85) *with initial distribution $\bar{\mu}_0$. Let μ_t^N be the law of $\{X_t^{i,N}, Y_t^{i,N}\}_{i=1}^N$, where $(\{X_s^{i,N}, Y_s^{i,N}\}_{i=1}^N)_{s\geq 0}$ is a solution to* (C.3) *with $b = 0$ and with initial distribution $\mu_0^N = \mu_0^{\otimes N}$. If $L_{\tilde{g}}$ satisfies* (C.88)*, then*

$$\mathcal{W}_{2,\hat{\rho}_N}(\bar{\mu}_t^{\otimes N}, \mu_t^N) \leq e^{-\hat{c}/2t}\mathcal{W}_{2,\hat{\rho}_N}(\bar{\mu}_0^{\otimes N}, \mu_0^N) + \hat{c}^{-1/2}\mathcal{C}_3 N^{-1/2} \qquad \textit{and}$$
$$\mathcal{W}_{2,\tilde{\ell}^2_N}(\bar{\mu}_t^N, \mu_t^N) \leq \sqrt{2}M_3 e^{-\hat{c}/2t}\mathcal{W}_{2,\tilde{\ell}^2_N}(\bar{\mu}_0^N, \mu_0^N) + M_4\hat{c}^{-1/2}\mathcal{C}_3 N^{-1/2},$$

where $\hat{c}$, $\tilde{l}^2_N$ and M_3 are given in (C.90), (C.98) *and* (C.91), *respectively. The constant M_4 is given by*

$$M_4 = \gamma \max(\sqrt{2/\tilde{\kappa}}, 2). \tag{C.99}$$

and $\mathcal{C}_3$ is a positive constant depending on γ, d, $\tilde{\kappa}$, $L_{\tilde{K}}$, $L_{\tilde{g}}$, u and on the second moment of $\bar{\mu}_0$. If $L_{\tilde{g}}$ satisfies (C.92), *then*

$$\mathcal{W}_{1,\tilde{\rho}_N}(\bar{\mu}_t^{\otimes N}, \mu_t^N) \le e^{-\hat{c}t}\mathcal{W}_{1,\tilde{\rho}_N}(\bar{\mu}_0^{\otimes N}, \mu_0^N) + \hat{c}^{-1}\mathcal{C}_4 N^{-1/2} \qquad \text{and}$$
$$\mathcal{W}_{1,\tilde{\ell}^1_N}(\bar{\mu}_t^{\otimes N}, \mu_t^N) \le \sqrt{2}M_3 e^{-\hat{c}t}\mathcal{W}_{1,\tilde{\ell}^1_N}(\bar{\mu}_0^{\otimes N}, \mu_0^N) + M_4\hat{c}^{-1}\mathcal{C}_4 N^{-1/2},$$

where $\mathcal{C}_4$ is a positive constant depending on γ, d, $\tilde{\kappa}$, $L_{\tilde{K}}$, $L_{\tilde{g}}$, u and on the second moment of $\bar{\mu}_0$.

Proof. The proof is postponed to Appendix C.7.3. □

Remark C.29. For $t \ge 0$, let μ_t^N and ν_t^N denote the law of $\{X_t^{i,N}, Y_t^{i,N}\}_{i=1}^N$ and $\{X_t'^{i,N}, Y_t'^{i,N}\}_{i=1}^N$, where the processes $(\{X_s^{i,N}, Y_s^{i,N}\}_{i=1}^N)_{s\ge 0}$ and $(\{X_s'^{i,N}, Y_s'^{i,N}\}_{i=1}^N)_{s\ge 0}$ are solutions to (C.3) with initial distributions μ_0^N and ν_0^N, respectively, and for which Assumption C.4 is supposed. An easy adaptation of the proof of Theorem C.17 shows that if (C.88) holds, then

$$\mathcal{W}_{2,\hat{\rho}_N}(\mu_t^N, \nu_t^N) \le e^{-\hat{c}t}\mathcal{W}_{2,\hat{\rho}_N}(\mu_0^N, \nu_0^N) \qquad \text{and} \qquad \mathcal{W}_{2,\tilde{\ell}^2_N}(\mu_t^N, \nu_t^N) \le \sqrt{2}M_3 e^{-\hat{c}t}\mathcal{W}_{2,\tilde{\ell}^2_N}(\mu_0^N, \nu_0^N),$$

and if (C.92) holds, then

$$\mathcal{W}_{1,\tilde{\rho}_N}(\mu_t^N, \nu_t^N) \le e^{-\hat{c}t}\mathcal{W}_{1,\tilde{\rho}_N}(\mu_0^N, \nu_0^N) \qquad \text{and} \qquad \mathcal{W}_{1,\tilde{\ell}^1_N}(\mu_t^N, \nu_t^N) \le \sqrt{2}M_3 e^{-\hat{c}t}\mathcal{W}_{1,\tilde{\ell}^1_N}(\mu_0^N, \nu_0^N),$$

where $\hat{c}$ and M_3 are given in (C.90) and (C.91), respectively. For the proof, a coupling of two copies of N particle systems is constructed in the same line as (C.108). As it will clarify by an inspection of the proof of Theorem C.17, we can obtain a slightly better contraction rate in L^2 Wasserstein distance for the particle system compared to the rate in the propagation of chaos result.

C.7.3 Proof of Section C.7.1 and Section C.7.2

Proof of Theorem C.24. Given two probability measures $\bar{\mu}_0, \bar{\nu}_0$ on $\mathbb{R}^{2d}$ and a d-dimensional Brownian motion $(B_t)_{t\ge 0}$, we consider the synchronous coupling $((\bar{X}_t, \bar{Y}_t), (\bar{X}_t', \bar{Y}_t'))_{t\ge 0}$ of two copies of solutions to (C.85) on $\mathbb{R}^{2d} \times \mathbb{R}^{2d}$ given by

$$\begin{cases} \mathrm{d}\bar{X}_t &= \bar{Y}_t \mathrm{d}t \\ \mathrm{d}\bar{Y}_t &= (-\gamma\bar{Y}_t + u\int_{\mathbb{R}^d}\tilde{b}(\bar{X}_t, z)\bar{\mu}_t^x(\mathrm{d}z))\mathrm{d}t + \sqrt{2\gamma u}\mathrm{d}B_t, \qquad (\bar{X}_0, \bar{Y}_0) \sim \bar{\mu}_0, \end{cases}$$
$$\begin{cases} \mathrm{d}\bar{X}_t' &= \bar{Y}_t' \mathrm{d}t \\ \mathrm{d}\bar{Y}_t' &= (-\gamma\bar{Y}_t' + u\int_{\mathbb{R}^d}\tilde{b}(\bar{X}_t', \tilde{z})\bar{\nu}_t^x(\mathrm{d}\tilde{z}))\mathrm{d}t + \sqrt{2\gamma u}\mathrm{d}B_t, \qquad (\bar{X}_0', \bar{Y}_0') \sim \bar{\nu}_0, \end{cases} \tag{C.100}$$

where $\bar{\mu}_t^x = \mathrm{Law}(\bar{X}_t)$, $\bar{\nu}_t^x = \mathrm{Law}(\bar{X}_t')$. We set $\tilde{Z}_t = \bar{X}_t - \bar{X}_t'$ and $\tilde{W}_t = \bar{Y}_t - \bar{Y}_t'$. By Assumption C.4 the process $(\tilde{Z}_t, \tilde{W}_t)_{t\ge 0}$ satisfies

$$\begin{cases} \mathrm{d}\tilde{Z}_t &= \tilde{W}_t \mathrm{d}t \\ \mathrm{d}\tilde{W}_t &= (-\gamma\tilde{W}_t + u\int_{\mathbb{R}^d}\tilde{b}(\bar{X}_t, z)\bar{\mu}_t^x(\mathrm{d}z) - u\int_{\mathbb{R}^d}\tilde{b}(\bar{X}_t', \tilde{z})\bar{\nu}_t^x(\mathrm{d}\tilde{z}))\mathrm{d}t \\ &= (-\gamma\tilde{W}_t - u\tilde{K}\tilde{Z}_t + u\int_{\mathbb{R}^d}\tilde{g}(\bar{X}_t - z)\bar{\mu}_t(\mathrm{d}z) - u\int_{\mathbb{R}^d}\tilde{g}(\bar{X}_t' - \tilde{z})\bar{\nu}_t(\mathrm{d}\tilde{z}))\mathrm{d}t, \end{cases} \tag{C.101}$$

where we used that $\mathbb{E}[\tilde{Z}_t] = 0$, which holds by Assumption C.4 and Assumption C.5. Let $\tilde{A}, \tilde{B}, \tilde{C} \in \mathbb{R}^{d\times d}$ be positive definite matrices given by

$$\tilde{A} = \gamma^{-2} u\tilde{K} + (1/2)(1-2\sigma)^2 \mathrm{Id}, \qquad \tilde{B} = (1-2\sigma)\gamma^{-1}\mathrm{Id}, \qquad \text{and} \qquad \tilde{C} = \gamma^{-2}\mathrm{Id}, \tag{C.102}$$

where σ is given by (C.87). Then, by Ito's formula,

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}(\tilde{Z}_t \cdot (\tilde{A}\tilde{Z}_t) + \tilde{Z}_t \cdot (\tilde{B}\tilde{W}_t) + \tilde{W}_t \cdot (\tilde{C}\tilde{W}_t)) \\
&\quad \le 2(\tilde{A}\tilde{Z}_t)\cdot \tilde{W}_t \mathrm{d}t + (\tilde{W}_t \cdot (\tilde{B}\tilde{W}_t) - (B\tilde{Z}_t)\cdot(\gamma\tilde{W}_t + u\tilde{K}\tilde{Z}_t))\mathrm{d}t - 2(\tilde{C}\tilde{W}_t)\cdot(\gamma\tilde{W}_t + u\tilde{K}\tilde{Z}_t)\mathrm{d}t \\
&\qquad + L_{\tilde{g}} u|\tilde{B}\tilde{Z}_t + 2\tilde{C}\tilde{W}_t|(|\tilde{Z}_t| + \mathbb{E}[|\tilde{Z}_t|])\mathrm{d}t \\
&\quad \le \tilde{Z}_t \cdot ((-u\tilde{K}\tilde{B})\tilde{Z}_t) + \tilde{Z}_t \cdot (2\tilde{A} - \gamma\tilde{B} - 2u\tilde{K}\tilde{C})\tilde{W}_t + \tilde{W}_t \cdot ((\tilde{B} - \gamma\tilde{C})\tilde{W}_t) \\
&\qquad + L_{\tilde{g}} u|\tilde{B}\tilde{Z}_t + 2\tilde{C}\tilde{W}_t|(|\tilde{Z}_t| + \mathbb{E}[|\tilde{Z}_t|]) \\
&\quad \le -2\sigma\gamma(\tilde{Z}_t \cdot (\tilde{A}\tilde{Z}_t) + \tilde{Z}_t \cdot (\tilde{B}\tilde{W}_t) + \tilde{W}_t \cdot (\tilde{C}\tilde{W}_t)) + L_{\tilde{g}} u|\tilde{B}\tilde{Z}_t + 2\tilde{C}\tilde{W}_t|(|\tilde{Z}_t| + \mathbb{E}[|\tilde{Z}_t|]),
\end{aligned}$$

where we applied (C.87) in the last step More precisely, it holds for all $z \in \mathbb{R}^d$

$$z \cdot ((-u\tilde{K}(1-4\sigma)\gamma^{-1})z) \le -(\tilde{\kappa}u/2)\gamma^{-1}|z|^2 \le -\gamma\sigma|z|^2 \le -\gamma\sigma(1-2\sigma)^2|z|^2 \tag{C.103}$$

and therefore $z \cdot ((-u\tilde{K}(1-2\sigma)\gamma^{-1})z) \le -2\gamma\sigma(\tilde{\kappa}u\gamma^{-2} + (1/2)(1-2\sigma)^2)|z|^2$.

Then for $\tilde{r}(t) = \tilde{r}((\bar{X}_t, \bar{Y}_t), (\bar{X}'_t, \bar{Y}'_t)) = (\tilde{Z}_t \cdot (\tilde{A}\tilde{Z}_t) + \tilde{Z}_t \cdot (\tilde{B}\tilde{W}_t) + \tilde{W}_t \cdot (\tilde{C}\tilde{W}_t))^{1/2}$ given in (C.86),

$$\mathrm{d}\tilde{r}(t)^2 \le -2\sigma\gamma\tilde{r}(t)^2 \mathrm{d}t + L_{\tilde{g}} u\gamma^{-1}|(1-2\sigma)\tilde{Z}_t + 2\gamma^{-1}\tilde{W}_t|(|\tilde{Z}_t| + \mathbb{E}[|\tilde{Z}_t|])\mathrm{d}t. \tag{C.104}$$

By taking expectation, it holds

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\tilde{r}(t)^2] \le -2\sigma\gamma\mathbb{E}[\tilde{r}(t)^2] + L_{\tilde{g}} u\gamma^{-1}\mathbb{E}[|(1-2\sigma)\tilde{Z}_t + 2\gamma^{-1}\tilde{W}_t|(|\tilde{Z}_t| + \mathbb{E}[|\tilde{Z}_t|])]. \tag{C.105}$$

By (C.88), (C.87) and Young's inequality, we obtain for the last term

$$\begin{aligned}
&L_{\tilde{g}} u\gamma^{-1}\mathbb{E}[|(1-2\sigma)\tilde{Z}_t + 2\gamma^{-1}\tilde{W}_t|(|\tilde{Z}_t| + \mathbb{E}[|\tilde{Z}_t|])] \\
&\qquad \le \frac{\sigma\sqrt{\tilde{\kappa}u}}{2}\mathbb{E}[|(1-2\sigma)\tilde{Z}_t + 2\gamma^{-1}\tilde{W}_t|(|\tilde{Z}_t| + \mathbb{E}[|\tilde{Z}_t|]] \\
&\qquad \le \sigma\gamma\Big(\tilde{\kappa}u\gamma^{-2}\mathbb{E}[|\tilde{Z}_t|^2] + \frac{1}{4}\mathbb{E}[|(1-2\sigma)\tilde{Z}_t + 2\gamma^{-1}\tilde{W}_t|^2]\Big) \\
&\qquad \le \sigma\gamma\Big(\tilde{\kappa}u\gamma^{-2}\mathbb{E}[|\tilde{Z}_t|^2] + \frac{1}{2}\mathbb{E}[|(1-2\sigma)\tilde{Z}_t + \gamma^{-1}\tilde{W}_t|^2] + \frac{1}{2}\mathbb{E}[|\tilde{W}_t|^2]\Big) \le \sigma\gamma\mathbb{E}[\tilde{r}(t)^2].
\end{aligned} \tag{C.106}$$

By inserting this bound in (C.105), we obtain by Grönwall's inequality,

$$\mathcal{W}_{2,\tilde{r}}(\bar{\mu}_t, \bar{\nu}_t)^2 \le \mathbb{E}[\tilde{r}(t)^2] \le e^{-2\hat{c}t}\mathbb{E}[\tilde{r}(0)^2]$$

with $\hat{c}$ given in (C.90). By taking the square root and the infimum over all couplings $\omega \in \Pi(\bar{\mu}_0, \bar{\nu}_0)$, we obtain the first result in L^2 Wasserstein distance. The second bound holds by (C.94) with M_3 given by (C.91). To obtain contraction in L^1 Wasserstein distance, we take the square root in (C.104),

$$\begin{aligned}
\mathrm{d}\tilde{r}(t) &\le -\sigma\gamma\tilde{r}(t)\mathrm{d}t + L_{\tilde{g}} u\gamma^{-1}\frac{|(1-2\sigma)\tilde{Z}_t + 2\gamma^{-1}\tilde{W}_t|}{2\tilde{r}(t)}(|\tilde{Z}_t| + \mathbb{E}[|\tilde{Z}_t|])\mathrm{d}t \\
&\le -\sigma\gamma\tilde{r}(t)\mathrm{d}t + L_{\tilde{g}} u\gamma^{-1}(|\tilde{Z}_t| + \mathbb{E}[|\tilde{Z}_t|])\mathrm{d}t,
\end{aligned}$$

where the last step holds by

$$\begin{aligned}&\frac{|(1-2\sigma)\tilde{Z}_t+2\gamma^{-1}\tilde{W}_t|}{2\tilde{r}(t)}\\&\leq\frac{1}{2}\Big(\frac{(1-2\sigma)^2|\tilde{Z}_t|^2+4(1-2\sigma)\gamma^{-1}\tilde{Z}_t\cdot\tilde{W}_t+4\gamma^{-2}|\tilde{W}_t|^2}{(\tilde{\kappa}u\gamma^{-2}+(1/2)(1-2\sigma)^2)|\tilde{Z}_t|^2+(1-2\sigma)\gamma^{-1}\tilde{Z}_t\cdot\tilde{W}_t+\gamma^{-2}|\tilde{W}_t|^2}\Big)^{1/2}\leq 1.\end{aligned}\tag{C.107}$$

Taking expectation and applying (C.92) we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\tilde{r}(t)]\leq-\sigma\gamma\mathbb{E}[\tilde{r}(t)]+2L_{\tilde{g}}u\gamma^{-1}\mathbb{E}[|\tilde{Z}_t|]\leq-\sigma\gamma\mathbb{E}[\tilde{r}(t)]+\frac{\sigma\gamma}{2}\mathbb{E}[\sqrt{\tilde{\kappa}u\gamma^{-2}}|\tilde{Z}_t|]\leq-\frac{\sigma\gamma}{2}\mathbb{E}[\tilde{r}(t)].$$

Hence by Grönwall's inequality,

$$\mathcal{W}_{1,\tilde{r}}(\bar{\mu}_t,\bar{\nu}_t)\leq e^{-\hat{c}t}\mathbb{E}[\tilde{r}(0)],$$

where $\hat{c}$ is given in (C.90). Taking the infimum over all couplings $\omega\in\Pi(\bar{\mu}_0,\bar{\nu}_0)$, we obtain the first bound in L^1 Wasserstein distance. The second bound follows by (C.94) with M_3 given in (C.91). □

To prove Theorem C.28, we establish a second moment bound of the solution to the nonlinear unconfined Langevin equation.

Lemma C.30 (Moment control for unconfined Langevin dynamics)**.** *Suppose that Assumption C.4 and* (C.88) *hold. Let* $(\bar{X}_t,\bar{Y}_t)_{t\geq 0}$ *be a solution to* (C.85) *with initial distribution satisfying Assumption C.5. Then there exists a finite constant* $\mathcal{C}_5>0$ *such that*

$$\sup_{t\geq 0}\mathbb{E}[|\bar{X}_t|^2]\leq\mathcal{C}_5.$$

The constant $\mathcal{C}_5$ *depends on* γ*,* d*,* $\tilde{\kappa}$*,* $L_{\tilde{g}}$*,* u *and on the second moment of the initial distribution.*

Proof. As in the proof of Lemma C.23, we adapt the proof idea from [75, Lemma 8]. First, we note that by Assumption C.4 and Assumption C.5, $\mathbb{E}[\bar{X}_t]=\mathbb{E}[\bar{Y}_t]=0$ for all $t\geq 0$, since by anti-symmetry of $\tilde{g}$

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\bar{X}_t]=\mathbb{E}[\bar{Y}_t],\quad\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\bar{Y}_t]=-\gamma\mathbb{E}[\bar{Y}_t],$$

and $\mathbb{E}[\bar{X}_0]=\mathbb{E}[\bar{Y}_0]=0$. Hence, $\bar{X}_t\cdot\mathbb{E}[\bar{X}_t]=\bar{Y}_t\cdot\mathbb{E}[\bar{X}_t]=0$. Further, we bound $|\mathbb{E}_{x\sim\bar{\mu}_t}[\tilde{g}(\bar{X}_t,x)]|\leq L_{\tilde{g}}(|\bar{X}_t|+\mathbb{E}[|\bar{X}_t|])$. By Ito's formula and Assumption C.4, it holds for $\sigma\in(0,1/2)$,

$$\begin{aligned}&\mathrm{d}(\gamma^{-2}u\bar{X}_t\cdot(\tilde{K}\bar{X}_t)+(1/2)|(1-2\sigma)\bar{X}_t+\gamma^{-1}\bar{Y}_t|^2+(1/2)\gamma^{-2}|\bar{Y}_t|^2)\\&\leq(2\gamma^{-2}u\bar{X}_t\cdot(\tilde{K}\bar{Y}_t)+(1-2\sigma)^2\bar{X}_t\cdot\bar{Y}_t)\mathrm{d}t+(1-2\sigma)\gamma^{-1}(|\bar{Y}_t|^2-\bar{X}_t\cdot(u\tilde{K}\bar{X}_t)-\gamma\bar{X}_t\cdot\bar{Y}_t)\mathrm{d}t\\&+\gamma^{-2}(-2\gamma|\bar{Y}_t|^2-2(u\tilde{K}\bar{Y})_t\cdot\bar{X}_t)\mathrm{d}t+L_{\tilde{g}}u|(1-2\sigma)\gamma^{-1}\bar{X}_t+2\gamma^{-2}\bar{Y}_t|(|\bar{X}_t|+\mathbb{E}[|\bar{X}_t|])\mathrm{d}t\\&+2\gamma^{-1}ud\mathrm{d}t+\sqrt{2\gamma^{-1}u}((1-2\sigma)\bar{X}_t+2\gamma^{-1}\bar{Y}_t)\mathrm{d}B_t\\&\leq-(1-2\sigma)\gamma^{-1}u\bar{X}_t\cdot(\tilde{K}\bar{X}_t)\mathrm{d}t-2\sigma\gamma((1-2\sigma)\gamma^{-1}\bar{X}_t\cdot\bar{Y}_t+\gamma^{-2}|\bar{Y}_t|^2)\mathrm{d}t+2\gamma^{-1}ud\mathrm{d}t\\&+L_{\tilde{g}}u|(1-2\sigma)\gamma^{-1}\bar{X}_t+2\gamma^{-2}\bar{Y}_t|(|\bar{X}_t|+\mathbb{E}[|\bar{X}_t|])\mathrm{d}t+\sqrt{2\gamma^{-1}u}((1-2\sigma)\bar{X}_t+2\gamma^{-1}\bar{Y}_t)\mathrm{d}B_t.\end{aligned}$$

Then by (C.103) we obtain after taking expectation

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\gamma^{-2}u\bar{X}_t\cdot(\tilde{K}\bar{X}_t)+\frac{1}{2}|(1-2\sigma)\bar{X}_t+\gamma^{-1}\bar{Y}_t|^2+\frac{1}{2}\gamma^{-2}|\bar{Y}_t|^2]\\&\quad\leq-2\sigma\gamma\mathbb{E}[\gamma^{-2}u\bar{X}_t\cdot(\tilde{K}\bar{X}_t)+\frac{1}{2}|(1-2\sigma)\bar{X}_t+\gamma^{-1}\bar{Y}_t|^2+\frac{1}{2}\gamma^{-2}|\bar{Y}_t|^2]+2\gamma^{-1}ud\\&\qquad+L_{\tilde{g}}u\gamma^{-1}\mathbb{E}[|(1-2\sigma)\bar{X}_t+2\gamma^{-1}\bar{Y}_t|(|\bar{X}_t|+\mathbb{E}[|\bar{X}_t|])].\end{aligned}$$

By (C.88) and Young's inequality, we bound the last term similarly as (C.106) by

$$\begin{aligned}&L_{\tilde{g}}u\gamma^{-1}\mathbb{E}[|(1-2\sigma)\bar{X}_t+2\gamma^{-1}\bar{Y}_t|(|\bar{X}_t|+\mathbb{E}[|\bar{X}_t|])]\\&\qquad\leq\sigma\gamma\Big(\tilde{\kappa}u\gamma^{-2}\mathbb{E}[|\bar{X}_t|^2]+\frac{1}{2}\mathbb{E}[|(1-2\sigma)\bar{X}_t+\gamma^{-1}\bar{Y}_t|^2]+\frac{1}{2}\mathbb{E}[\bar{Y}_t|^2]\Big).\end{aligned}$$

Hence,

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\Big[\gamma^{-2}u\bar{X}_t\cdot(\tilde{K}\bar{X}_t)+\frac{1}{2}|(1-2\sigma)\bar{X}_t+\gamma^{-1}\bar{Y}_t|^2+\frac{1}{2}\gamma^{-2}|\bar{Y}_t|^2\Big]\\&\quad\leq-\sigma\gamma\mathbb{E}\Big[\gamma^{-2}u\bar{X}_t\cdot(\tilde{K}\bar{X}_t)+\frac{1}{2}|(1-2\sigma)\bar{X}_t+\gamma^{-1}\bar{Y}_t|^2+\frac{1}{2}\gamma^{-2}|\bar{Y}_t|^2\Big]+2\gamma^{-1}ud.\end{aligned}$$

Then by Grönwall's inequality, there exists a constant $\mathbf{C}$ such that

$$\sup_{t\geq0}\mathbb{E}[\gamma^{-2}u\bar{X}_t\cdot(\tilde{K}\bar{X}_t)\frac{1}{2}|(1-2\sigma)\bar{X}_t+\gamma^{-1}\bar{Y}_t|^2+\frac{1}{2}\gamma^{-2}|\bar{Y}_t|^2]\leq\mathbf{C}<\infty$$

and we obtain the result for $\mathcal{C}_5=\mathbf{C}/(\tilde{\kappa}\gamma^{-2}u)$. □

Proof of Theorem C.28. We consider a synchronous coupling approach of solutions to (C.85) and (C.3) with $b\equiv0$. Fix $N\in\mathbb{N}$. Let $\{(B_t^i)_{t\geq0}\}_{i=1}^N$ be N independent d-dimensional Brownian motions and let μ_0 and $\bar{\mu}_0$ be two probability measrues on $\mathbb{R}^{2d}$. The coupling $(\{(\bar{X}_t^i,\bar{Y}_t^i),(X_t^i,Y_t^i)\}_{i=1}^N)_{t\geq0}$ of N copies of a solution to (C.85) and a solution to (C.3) with $b\equiv0$ is given on $\mathbb{R}^{2Nd}\times\mathbb{R}^{2Nd}$ by

$$\begin{aligned}&\begin{cases}\mathrm{d}\bar{X}_t^i&=\bar{Y}_t^i\mathrm{d}t\\\mathrm{d}\bar{Y}_t^i&=(-\gamma\bar{Y}_t^i+u\int_{\mathbb{R}^d}\tilde{b}(\bar{X}_t^i,z)\bar{\mu}_t^x(\mathrm{d}z))\mathrm{d}t+\sqrt{2\gamma u}\mathrm{d}B_t^i,\qquad(\bar{X}_0^i,\bar{Y}_0^i)\sim\bar{\mu}_0,\end{cases}\\&\begin{cases}\mathrm{d}X_t^i&=Y_t^i\mathrm{d}t\\\mathrm{d}Y_t^i&=(-\gamma Y_t^i+uN^{-1}\sum_{j=1}^N\tilde{b}(X_t^i,X_t^j))\mathrm{d}t+\sqrt{2\gamma u}\mathrm{d}B_t^i,\qquad(X_0^i,Y_0^i)\sim\mu_0\end{cases}\end{aligned}\tag{C.108}$$

for $i=1,...,N$, where $\bar{\mu}_t^x=\mathrm{Law}(\bar{X}_t^i)$ for all i. For simplicity, we omitted the parameter N in the index of (X_t^i,Y_t^i) in the particle model. We set $\tilde{Z}_t^i=\bar{X}_t^i-X_t^i-N^{-1}\sum_{j=1}^N(\bar{X}_t^j-X_t^j)$ and $\tilde{W}_t^i=\bar{Y}_t^i-Y_t^i-N^{-1}\sum_{j=1}^N(\bar{Y}_t^j-Y_t^j)$. By Assumption C.4, the process $(\{\tilde{Z}_t^i,\tilde{W}_t^i\}_{i=1}^N)_{t\geq0}$ satisfies

$$\begin{cases}\mathrm{d}\tilde{Z}_t^i&=\tilde{W}_t^i\mathrm{d}t\\\mathrm{d}\tilde{W}_t^i&=-\gamma\tilde{W}_t^i\mathrm{d}t+u\Big(\int_{\mathbb{R}^d}\tilde{b}(\bar{X}_t^i,z)\bar{\mu}_t^x(\mathrm{d}z)-N^{-1}\sum_{j=1}^N\int_{\mathbb{R}^d}\tilde{b}(\bar{X}_t^j,\tilde{z})\bar{\mu}_t^x(\mathrm{d}\tilde{z})\\&\qquad-N^{-1}\sum_{j=1}^N\tilde{b}(X_t^i,X_t^j)+N^{-2}\sum_{j,k=1}^N\tilde{b}(X_t^j,X_t^k)\Big)\mathrm{d}t\\&=-\gamma\tilde{W}_t^i\mathrm{d}t+u\Big(-\tilde{K}\tilde{Z}_t^i+N^{-1}\sum_{j=1}^N(\tilde{g}(\bar{X}_t^i-\bar{X}_t^j)-\tilde{g}(X_t^i-X_t^j))\\&\qquad+\tilde{A}_t^i+N^{-1}\sum_{j=1}^N\tilde{A}_t^j\Big)\mathrm{d}t,\end{cases}\tag{C.109}$$

where $\tilde{A}_t^k = \int_{\mathbb{R}^d} \tilde{b}(\bar{X}_t^k, z)\bar{\mu}_t^x(\mathrm{d}z) - N^{-1}\sum_{j=1}^N \tilde{b}(\bar{X}_t^k, \bar{X}_t^j)$ for all $k = 1, ..., N$. Hence, for the positive definite matrices $\tilde{A}, \tilde{B}, \tilde{C}$ given in (C.102), we obtain for $i = 1, ..., N$,

$$\begin{aligned}
&\mathrm{d}(\tilde{Z}_t^i \cdot (\tilde{A}\tilde{Z}_t^i) + \tilde{Z}_t^i \cdot (\tilde{B}\tilde{W}_t^i) + \tilde{W}_t^i \cdot (\tilde{C}\tilde{W}_t^i) \\
&\le 2\tilde{Z}_t^i \cdot (\tilde{A}\tilde{W}_t^i)\mathrm{d}t + (\tilde{W}_t^i \cdot (\tilde{B}\tilde{W}_t^i) - \gamma\tilde{Z}_t^i \cdot (\tilde{B}\tilde{W}_t^i) - (\tilde{B}\tilde{Z}_t^i) \cdot (u\tilde{K}\tilde{W}_t^i))\mathrm{d}t \\
&+ (\tilde{C}W_t^i) \cdot (-2\gamma W_t^i - 2u\tilde{K}\tilde{Z}_t^i)\mathrm{d}t \\
&+ |\tilde{B}\tilde{Z}_t^i + 2\tilde{C}\tilde{W}_t^i| u\Big(L_{\tilde{g}} N^{-1}\sum_{j=1}^N (|\tilde{Z}_t^j| + |\tilde{Z}_t^i|) + A_t^i + N^{-1}\sum_{i=j}^N A_t^j\Big)\mathrm{d}t \\
&\le \Big(-(u\tilde{K}\tilde{Z}_t^i) \cdot (\tilde{B}\tilde{Z}_t^i) + \tilde{Z}_t^i \cdot ((2\tilde{A} - \gamma\tilde{B} - 2u\tilde{K}\tilde{C})\tilde{W}_t^i) + \tilde{W}_t^i \cdot ((\tilde{B} - 2\gamma\tilde{C})\tilde{W}_t^i)\Big)\mathrm{d}t \\
&+ |\tilde{B}\tilde{Z}_t^i + 2\tilde{C}\tilde{W}_t^i| u\Big(L_{\tilde{g}} N^{-1}\sum_{j=1}^N (|\tilde{Z}_t^j| + |\tilde{Z}_t^i|) + A_t^i + N^{-1}\sum_{i=j}^N A_t^j\Big)\mathrm{d}t,
\end{aligned}$$

where $A_t^k = |\int_{\mathbb{R}^d} \tilde{b}(\bar{X}_t^k, z)\bar{\mu}_t^x(\mathrm{d}z) - N^{-1}\sum_{j=1}^N \tilde{b}(\bar{X}_t^k, \bar{X}_t^j)|$ for all $k = 1, ..., N$. Then by (C.104) for $\tilde{r}^i(t) = \tilde{r}((\bar{X}_t^i, \bar{Y}_t^i), (X_t^i, Y_t^i))$

$$\begin{aligned}
\mathrm{d}\tilde{r}^i(t)^2 &= \mathrm{d}(\tilde{Z}_t^i \cdot (\tilde{A}\tilde{Z}_t^i) + \tilde{Z}_t^i \cdot (\tilde{B}\tilde{W}_t^i) + \tilde{W}_t^i \cdot (\tilde{C}\tilde{W}_t^i)) \\
&\le -2\sigma\gamma\tilde{r}^i(t)^2\mathrm{d}t + \gamma^{-1}|(1-2\sigma)\tilde{Z}_t^i + 2\gamma^{-1}\tilde{W}_t^i|\Big(L_{\tilde{g}} N^{-1}\sum_{j=1}^N (|\tilde{Z}_t^j| + |\tilde{Z}_t^i|) + A_t^i + N^{-1}\sum_{i=j}^N A_t^j\Big)\mathrm{d}t
\end{aligned} \tag{C.110}$$

and hence, for $\hat{\rho}_t := \hat{\rho}_N((X_t, Y_t), (\bar{X}_t, \bar{Y}_t))$ given in (C.95),

$$\begin{aligned}
\mathrm{d}\hat{\rho}_t \le -2\sigma\gamma\hat{\rho}_t\mathrm{d}t + \frac{u}{\gamma}N^{-1}\sum_{i=1}^N \Big(|(1-2\sigma)\tilde{Z}_t^i + 2\gamma^{-1}\tilde{W}_t^i| \\
\Big(L_{\tilde{g}} N^{-1}\sum_{j=1}^N (|\tilde{Z}_t^j| + |\tilde{Z}_t^i|) + A_t^i + N^{-1}\sum_{j=1}^N A_t^j\Big)\Big)\mathrm{d}t.
\end{aligned} \tag{C.111}$$

For the last term, we obtain by (C.88) and Young's inequality

$$L_{\tilde{g}} u\gamma^{-1}\frac{1}{N^2}\sum_{i,j=1}^N |(1-2\sigma)\tilde{Z}_t^i + 2\gamma^{-1}\tilde{W}_t^i|(|\tilde{Z}_t^j| + |\tilde{Z}_t^i|) \le \sigma\gamma\hat{\rho}_t$$

similarly as in (C.106) and

$$\begin{aligned}
&\frac{u}{\gamma}\frac{1}{N^2}\sum_{i,j=1}^N |(1-2\sigma)\tilde{Z}_t^i + 2\gamma^{-1}\tilde{W}_t^i|(A_t^i + A_t^j) \\
&\qquad \le \frac{\sigma\gamma}{2}\frac{1}{N}\sum_{i=1}^N \Big(\frac{1}{4}|(1-2\sigma)\tilde{Z}_t^i + 2\gamma^{-1}\tilde{W}_t^i|^2\Big) + \frac{8u^2}{\gamma^3\sigma}\frac{1}{N}\sum_{i=1}^N (A_t^i)^2 \\
&\qquad \le \frac{\sigma\gamma}{2}\hat{\rho}_t + \frac{8}{\gamma^3\sigma}\frac{1}{N}\sum_{i=1}^N (A_t^i)^2.
\end{aligned}$$

Inserting these estimates in (C.111) and taking expectation yields

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\hat{\rho}_t] \le -\frac{\sigma\gamma}{2}\mathbb{E}[\hat{\rho}_t] + \frac{8}{\gamma^3\sigma}\frac{1}{N}\sum_{i=1}^N \mathbb{E}[(A_t^i)^2].$$

We bound $\mathbb{E}[{A_t^i}^2]$ similar as in the proof of Theorem C.17. Note that by Assumption C.4, $\tilde{b}$ is Lipschitz continuous with a Lipschitz constant which is bounded from above by $L_{\tilde{K}}+L_{\tilde{g}}$. Hence, (C.82) and (C.83) hold here with $L_{\tilde{K}}+L_{\tilde{g}}$ instead of $\tilde{L}$. Then,

$$\mathbb{E}[{A_t^i}^2] \leq \mathbb{E}\Big[\Big|\int_{\mathbb{R}^d}\tilde{b}(\bar{X}_t^i,z)\bar{\mu}_t^x(\mathrm{d}z)-\frac{1}{N}\sum_{j=1}\tilde{b}(\bar{X}_t^i,\bar{X}_t^j)\Big|^2\Big] \leq \frac{16(L_{\tilde{K}}+L_{\tilde{g}})^2}{N}\int_{\mathbb{R}^d}|x|^2\bar{\mu}_t(\mathrm{d}x).$$

By Lemma C.30, there exists a constant $\mathcal{C}_6$ depending on γ, $\mathbb{E}[|\bar{X}_0|^2+|\bar{Y}_0|^2]$, d, $\tilde{\kappa}$, $L_{\tilde{K}}$, $L_{\tilde{g}}$, u such that for $N\geq 2$ and $i=1,...,N$,

$$\sup_{t\geq 0}\mathbb{E}[{A_t^i}^2] \leq \mathcal{C}_6 N^{-1}.$$

Hence,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\hat{\rho}_t^2] \leq -2\sigma\gamma\mathbb{E}[\hat{\rho}_0^2]+\mathcal{C}_3^2N^{-1}/2,$$

where $\mathcal{C}_3^2=\frac{8u^2}{\sigma\gamma^3}\mathcal{C}_6$. By Grönwall's inequality,

$$\mathcal{W}_{2,\hat{\rho}_N}(\mathrm{Law}(X_t^1,...,X_t^N),(\bar{\mu}_t)^{\otimes N})^2 \leq \mathbb{E}[\hat{\rho}_t^2] \leq e^{-\hat{c}t}\mathbb{E}[\hat{\rho}_0^2]+\hat{c}^{-1}\mathcal{C}_3^2N^{-1}$$

with $\hat{c}$ given in (C.90). By taking the infimum over all couplings $\omega\in\Pi(\mu_0^N,\bar{\mu}_0^{\otimes N})$, we obtain the first result in L^2 Wasserstein distance. The second bound holds by (C.94) with M_3 and M_4 given by (C.91) and (C.99), respectively. To obtain the bound in L^1 Wasserstein distance, we note that by (C.110)

$$\begin{aligned}
\mathrm{d}\tilde{r}^i(t) &= \frac{1}{2r^i(t)}\mathrm{d}\tilde{r}^i(t)^2 \\
&\leq -\sigma\gamma\tilde{r}^i(t)\mathrm{d}t+\frac{|(1-2\sigma)\tilde{Z}_t^i+2\gamma^{-1}\tilde{W}_t^i|}{2\gamma\tilde{r}^i(t)}u\Big(\frac{L_{\tilde{g}}}{N}\sum_j(|\tilde{Z}_t^j|+|\tilde{Z}_t^i|)+A_t^i+\frac{1}{N}\sum_{j=1}^N A_t^j\Big)\mathrm{d}t \\
&\leq -\sigma\gamma\tilde{r}^i(t)\mathrm{d}t+\gamma^{-1}u\Big(\frac{L_{\tilde{g}}}{N}\sum_j(|\tilde{Z}_t^j|+|\tilde{Z}_t^i|)+A_t^i+\frac{1}{N}\sum_{j=1}^N A_t^j\Big)\mathrm{d}t,
\end{aligned}$$

where the last step holds by (C.107). By summing over i and taking expectation, we obtain by (C.92) for $\tilde{\rho}_t := \tilde{\rho}_N((X_t,Y_t),(\bar{X}_t,\bar{Y}_t))$ given in (C.96),

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\tilde{\rho}_t] \leq -\sigma\gamma/2\mathbb{E}[\tilde{\rho}_t]+\gamma^{-1}uN^{-1}\sum_{i=1}^N\mathbb{E}[A_t^i].$$

By Assumption C.4 and Lemma C.30, there exists a constant $\mathcal{C}_4$ depending on γ, $\mathbb{E}[|\bar{X}_0|^2+|\bar{Y}_0|^2]$, d, $\tilde{\kappa}$, $L_{\tilde{K}}$, u and $L_{\tilde{g}}$ such that

$$\sup_{t\geq 0}\mathbb{E}[A_t^i] \leq \mathcal{C}_4\gamma N^{-1/2}$$

similarly as in (C.84). Hence,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[\tilde{\rho}_t] \leq -\frac{\sigma\gamma}{2}\mathbb{E}[\tilde{\rho}_t]+\mathcal{C}_4N^{-1/2}.$$

By Grönwall's inequality,

$$\mathcal{W}_{1,\tilde{\rho}_N}(\bar{\mu}_t^{\otimes N}, \mu_t^N) \leq \mathbb{E}[\tilde{\rho}_t] \leq e^{-\hat{c}t}\mathbb{E}[\tilde{\rho}_0] + \hat{c}^{-1}\mathcal{C}_4 N^{-1/2}$$

for $\hat{c}$ given in (C.90). Taking the infimum over all couplings $\omega \in \Pi(\bar{\mu}_0^{\otimes N}, \mu_0^N)$, we obtain the first result in L^1 Wasserstein distance. The second bound holds by (C.94) with M_3 and M_4 given in (C.91) and (C.99).

□