

Communications in Mathematical Physics

Convergence rates for the Vlasov-Fokker-Planck equation and uniform in time propagation of chaos in non convex cases

--Manuscript Draft--

Manuscript Number:	CIMP-D-21-00590
Full Title:	Convergence rates for the Vlasov-Fokker-Planck equation and uniform in time propagation of chaos in non convex cases
Article Type:	Original research
Abstract:	We prove the existence of a contraction rate for Vlasov-Fokker-Planck equation in Wasserstein distance, provided the interaction potential is (locally) Lipschitz continuous and the confining potential is both Lipschitz continuous and greater than a quadratic function, thus requiring no convexity conditions. Our strategy relies on coupling methods suggested by A. Eberle [14] adapted to the kinetic setting enabling also to obtain uniform in time propagation of chaos in a non convex setting.

Convergence rates for the Vlasov-Fokker-Planck equation and uniform in time propagation of chaos in non convex cases

Arnaud Guillin · Pierre Le Bris · Pierre Monmarché

Received: date / Accepted: date

Abstract We prove the existence of a contraction rate for Vlasov-Fokker-Planck equation in Wasserstein distance, provided the interaction potential is (locally) Lipschitz continuous and the confining potential is both Lipschitz continuous and greater than a quadratic function, thus requiring no convexity conditions. Our strategy relies on coupling methods suggested by A. Eberle [14] adapted to the kinetic setting enabling also to obtain uniform in time propagation of chaos in a non convex setting.

Keywords Vlasov-Fokker-Planck equation · coupling · propagation of chaos

Mathematics Subject Classification (2020) 60J60 · 35K58 · 82B40

1 Introduction

1.1 Framework

Let U and W be two functions in $\mathcal{C}^1(\mathbb{R}^d)$. We consider the Vlasov-Fokker-Planck equation:

$$\partial_t \nu_t(x, v) = -\nabla_x \cdot (v \nu_t(x, v)) + \nabla_v \cdot ((v + \nabla U(x) + \nabla W * \mu_t(x)) \nu_t(x, v) + \nabla_v \nu_t(x, v)), \quad (1)$$

A. Guillin. Email
Laboratoire de Mathématiques Blaise Pascal - Université Clermont-Auvergne
E-mail: arnaud.guillin@uca.fr

P. Le Bris
Laboratoire Jacques-Louis Lions - Sorbonne Université
E-mail: pierre.lebris@sorbonne-universite.fr

P. Monmarché
Laboratoire Jacques-Louis Lions - Sorbonne Université
E-mail: pierre.monmarche@sorbonne-universite.fr

where $\nu_t(x, v)$ is a probability density in the space of positions $x \in \mathbb{R}^d$ and velocities $v \in \mathbb{R}^d$,

$$\mu_t(x) = \int_{\mathbb{R}^d} \nu_t(x, dv)$$

is the space marginal of ν_t and

$$\nabla W * \mu_t(x) = \int_{\mathbb{R}^d} \nabla W(x - y) \mu_t(dy).$$

It has the following probabilistic counterpart, the non linear stochastic differential equation of *McKean-Vlasov* type, i.e. ν_t is the density of the law at time t of the \mathbb{R}^{2d} -valued process $(X_t, V_t)_{t \geq 0}$ evolving as the mean field SDE (diffusive Newton's equations)

$$\begin{cases} dX_t = V_t dt \\ dV_t = \sqrt{2} dB_t - V_t dt - \nabla U(X_t) dt - \nabla W * \mu_t(X_t) dt \\ \mu_t = \text{Law}(X_t). \end{cases} \quad (2)$$

Here, $(X_t, V_t) \in \mathbb{R}^d \times \mathbb{R}^d$, $(B_t)_{t \geq 0}$ is a Brownian motion in dimension d on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, and μ_t is the law of the position X_t . The symbol ∇ refers to the gradient operator, and the symbol $*$ to the operation of convolution.

Both in the probability and in the partial differential equation community, existence and uniqueness of McKean-Vlasov processes have been well studied. See [27], [17], [33] for some historical milestones. In the specific case of (1) and (2), under the assumptions on U and W introduced in the next section, existence and uniqueness follow from [28] for square integrable initial data.

A related process is the N particles system in \mathbb{R}^d in mean field interaction

$$\forall i \in \llbracket 1, N \rrbracket, \quad \begin{cases} dX_t^i = V_t^i dt, \\ dV_t^i = \sqrt{2} dB_t^i - V_t^i dt - \nabla U(X_t^i) dt - \frac{1}{N} \sum_{j=1}^N \nabla W(X_t^i - X_t^j) dt, \end{cases} \quad (3)$$

where X_t^i and V_t^i are respectively the position and the velocity of the i -th particle, and $(B_t^i, 1 \leq i \leq N)$ are independent Brownian motions in dimension d . One can see equation (3) as an approximation of equation (2), where the law μ_t is replaced by the empirical measure $\mu_t^N = \frac{1}{N} \sum_{i=1}^N \delta_{X_t^i}$.

It is well known, at least in a non kinetic setting ([28], [33]), that, under some weak conditions on U and W , μ_t^N converges in some sense toward the law μ_t of X_t solution of (2). This phenomenon has been stated under the name *propagation of chaos*, an idea motivated by M. Kac [24]), and greatly developed by A.S. Sznitman [33].

In statistical physics, (3) is a Langevin equation that describes the motion of N particles subject to damping, random collisions and a *confining potential* U and interacting with one another through an *interaction potential* W , which can be polynomial (granular media), Newtonian (interacting stellar) or

Coulombian (charged matter). See for instance [26] for an english translation of P. Langevin's landmark paper on the physics behind the standard underdamped Langevin dynamics. Therefore, Equation (1) has the following natural interpretation: the solution ν_t is the density of the law at time t of the process $(X_t, V_t)_{t \geq 0}$ evolving according to (2), and thus describes the limit dynamic of a cloud of (charged) particles. In particular, it holds importance in plasma physics, see [36].

More recently, mean-field processes such as (3) have drawn much interest in the analysis of neuron networks in machine learning [12, 11]. In this context of stochastic algorithms, it is known that the underdamped Langevin dynamics (not necessarily with mean-field interactions) can converge faster than the overdamped (i.e non kinetic) Langevin dynamics [12, 20] toward its invariant measure. For example, the results on (2) could be applied to study the convergence of the Hamiltonian gradient descent algorithm for the overparametrized optimization as done in [25] for Generative Adversarial Network training.

The goal of the present work is twofold. We are interested, first, in the long-time convergence of the solution of (2) toward an equilibrium and, second, to a uniform in time convergence as $N \rightarrow +\infty$ of (3) toward (2). It is well known that such results cannot hold in full generality, as the non-linear equation (1) may have several equilibria. Here we will consider cases where the interaction is sufficiently small for the non-linear equilibrium to be unique and globally attractive, and for the propagation of chaos to be uniform in time.

There are various methods to study the long time behavior of kinetic type processes, such as Lyapunov conditions or hypocoercivity, and we will discuss these approaches and compare them with our results later on. We rely here on coupling methods following the guidelines of A. Eberle *et al.* in [15] where the convergence to equilibrium is established for (2) without interaction, and also extend the approach to handle only locally Lipschitz coefficient. In a second part, we also use reflection couplings (see [13]) for the propagation of chaos property.

Let us briefly describe the coupling method. The basic idea is that an upper bound on the Wasserstein distance between two probability distributions is given by the construction of any pair of random variables distributed respectively according to those. The goal is thus to construct simultaneously two solutions of (2) that have a trend to get closer with time. Have (X_t, V_t) be a solution of (2) driven by some Brownian motion $(B_t)_{t \geq 0}$ and let (X'_t, V'_t) solves

$$\begin{cases} dX'_t = V'_t dt \\ dV'_t = \sqrt{2} dB'_t - V'_t dt - \nabla U(X'_t) dt - \nabla W * \mu_t(X'_t) dt \\ \mu'_t = \text{Law}(X'_t). \end{cases}$$

with $(B'_t)_{t \geq 0}$ a d-dimensional Brownian motion. A coupling of (X, V) and (X', V') then follows from a coupling of the Brownian motions B and B' . Choosing $B = B'$ yields the so-called *synchronous* coupling, for which the Brownian noise cancels out in the infinitesimal evolution of the difference $(Z_t, W_t) = (X_t - X'_t, V_t - V'_t)$. In that case the contraction of a distance between the processes can only be induced by the deterministic drift, as in [6].

Such a deterministic contraction only holds under very restrictive conditions, in particular U should be strongly convex. Nevertheless, in more general cases, the calculation of the evolution of Z_t and W_t (see Section 3.1 below) shows that there is still some deterministic contraction when $Z_t + W_t = 0$. We can therefore use a synchronous coupling in the vicinity of this subspace.

Outside of $\{(z, w) \in \mathbb{R}^{2d}, z + w = 0\}$, it is necessary to make use of the noise to get the processes closer together, at least in the direction orthogonal to this space. In order to maximize the variance of this noise, we then use a so-called *reflection* coupling, which consists in B and B' being *symmetrical* (i.e. $B'_t = -B_t$) in the direction of space given by the difference of the processes, and synchronous in the orthogonal direction. In other words, writing

$$e_t = \begin{cases} \frac{Z_t + W_t}{|Z_t + W_t|} & \text{if } Z_t + W_t \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

we consider $dB'_t = (Id - 2e_t e_t^T) dB_t$. Levy's characterization then ensures that it is indeed a Brownian motion.

Finally we construct a Lyapunov function H to take into account the trend of each process to come back to some compact set of \mathbb{R}^{2d} . We are then led to the study of a suitable distance between the two processes, which will be of the form $\rho_t := f(r_t)(1 + \epsilon H(X_t, V_t) + \epsilon H(X'_t, V'_t))$, with $r_t = \alpha|Z_t| + |Z_t + W_t|$, where $\alpha, \epsilon > 0$ and the function f are some parameters to choose. More precisely, we have to choose these parameters carefully in order for $\mathbb{E}\rho_t$ to decay exponentially fast. This leads to several constraints on α, ϵ and on the parameters involved in the definition of f , and we have to prove that it is possible to meet all these conditions simultaneously. For the sake of clarity, in fact, we present the proof in a different order, namely we start by introducing very specific parameters and, throughout the proof, we check that our choice of parameters implies the needed constraints.

The study of the limit $N \rightarrow +\infty$ is based on a similar coupling, except that we couple a system of N interacting particles 3 with N independent non-linear processes 2.

The next subsections describe our main results and compare them to the few existing ones in the literature. Section 2 presents the precise construction of the aforementioned *ad hoc* Wasserstein distance. The proof of the long time behavior of the Vlasov-Fokker-Planck equation when confinement and interaction coefficient are Lipschitz continuous is done in Section 3, whereas the propagation of chaos property is proved in Section 4. An appendix gathers technical lemmas and the modifications of the main proofs when the confinement is only supposed locally Lipschitz continuous.

1.2 Main results

For μ and ν two probability measures on \mathbb{R}^{2d} , denote by $\Pi(\mu, \nu)$ the set of couplings of μ and ν , i.e. the set of probability measures Γ on $\mathbb{R}^{2d} \times \mathbb{R}^{2d}$ with

$\Gamma(A \times \mathbb{R}^{2d}) = \mu(A)$ and $\Gamma(\mathbb{R}^{2d} \times A) = \nu(A)$ for all Borel set A of \mathbb{R}^{2d} . We will define L^1 and L^2 Wasserstein distances as

$$\begin{aligned}\mathcal{W}_1(\mu, \nu) &= \inf_{\Gamma \in \Pi(\mu, \nu)} \int (|x - \tilde{x}| + |v - \tilde{v}|) \Gamma(dx dv d\tilde{x} d\tilde{v}), \\ \mathcal{W}_2(\mu, \nu) &= \left(\inf_{\Gamma \in \Pi(\mu, \nu)} \int (|x - \tilde{x}|^2 + |v - \tilde{v}|^2) \Gamma(dx dv d\tilde{x} d\tilde{v}) \right)^{1/2}.\end{aligned}$$

Our main results will be stated in terms of these distances, even if we work and get contraction in the Wasserstein distance defined with the aforementioned ρ . Let us detail the assumptions on the potentials U and W .

Proposition 1 *The potential U is non-negative and there exist $\lambda > 0$ and $A \geq 0$ such that*

$$\forall x \in \mathbb{R}^d, \quad \frac{1}{2} \nabla U(x) \cdot x \geq \lambda \left(U(x) + \frac{|x|^2}{4} \right) - A. \quad (4)$$

The condition (4) implies that the force $-\nabla U$ has a confining effect, bringing back particles toward some compact set. It implies the following:

Lemma 1 *Provided (4), there exists $\tilde{A} \geq 0$ such that for all $x \in \mathbb{R}^d$,*

$$U(x) \geq \frac{\lambda}{6} |x|^2 - \tilde{A}. \quad (5)$$

The proof is postponed to Appendix A.1. In particular, it implies that U goes to infinity at infinity and is bounded below. Since only the gradient of U is involved in the dynamics, the condition $U \geq 0$ is thus not restrictive as it can be enforced without loss of generality by adding a sufficient large constant to U . This condition is added in order to simplify some calculations.

Proposition 2 *There is a constant $L_U > 0$ such that*

$$\forall x, y \in \mathbb{R}^d \times \mathbb{R}^d, \quad |\nabla U(x) - \nabla U(y)| \leq L_U |x - y|.$$

Example 1 The double-well potential given by

$$U(x) = \begin{cases} (x^2 - 1)^2 & \text{if } |x| \leq 1, \\ (|x| - 1)^2 & \text{otherwise.} \end{cases}$$

satisfies Assumptions 1 and 2.

Proposition 3 *The potential W is even, i.e. $W(x) = W(-x)$ for all $x \in \mathbb{R}^d$, in particular $\nabla W(0) = 0$. Moreover, there exists $L_W < \lambda/8$ (where λ is given in Assumption 1) such that*

$$\forall x, y \in \mathbb{R}^d \times \mathbb{R}^d, \quad |\nabla W(x) - \nabla W(y)| \leq L_W |x - y|. \quad (6)$$

In particular $|\nabla W(x)| \leq L_W |x|$ for all $x \in \mathbb{R}^d$.

Here we consider an interaction force that is the gradient of a potential W , as we stick to the formalism of other related works (for instance [13]). Nevertheless, all the results and proofs still holds if ∇W is replaced by some $F : \mathbb{R}^d \mapsto \mathbb{R}^d$ satisfying the same conditions. The confinement potential may also be non gradient, however the fact that the confinement force ∇U is a gradient simplifies the construction of a Lyapunov function.

The condition $L_W \leq \lambda/8$ is related to the fact the interaction is considered as a perturbation of the non-interacting process studied in [15]. Therefore, ∇W has to be controlled by ∇U in some sense. Note that we immediately get the following bound on the non-linear drift:

Lemma 2 *Under Assumption 3, for all probability measures μ and ν on \mathbb{R}^d and $x, \tilde{x} \in \mathbb{R}^d$,*

$$|\nabla W * \mu(x) - \nabla W * \nu(\tilde{x})| \leq L_W |x - \tilde{x}| + L_W \mathcal{W}_1(\mu, \nu).$$

See Appendix A.2 for the proof.

Example 2 Assumption 3 is satisfied for an harmonic interaction $W(x) = \pm L_W |x|^2/2$, or a mollified Coulomb interaction for $a, b > 0$ and $k \in \mathbb{N}^*$

$$W(x) = \pm \frac{a}{(|x|^k + b^k)^{\frac{1}{k}}}, \quad \text{i.e.} \quad \nabla W(x) = \mp \frac{ax|x|^{k-2}}{(|x|^k + b^k)^{1+\frac{1}{k}}}.$$

We are also interested in cases where, instead of Assumption 2, ∇U is only assumed to be locally Lipschitz continuous.

Proposition 4 *There exist $L_U > 0$ and a function $\psi : \mathbb{R}^d \mapsto \mathbb{R}$ such that*

$$\forall x, y \in \mathbb{R}^d \times \mathbb{R}^d, \quad |\nabla U(x) - \nabla U(y)| \leq (L_U + \psi(x) + \psi(y)) |x - y|,$$

and

$$\forall x \in \mathbb{R}^d, \quad 0 \leq \psi(x) \leq L_\psi \sqrt{\lambda |x|^2 + 24U(x)},$$

where $L_\psi > 0$ is sufficiently small in the sense that

$$L_\psi \leq c_\psi(L_U, \lambda, \tilde{A}, d, a),$$

where c_ψ is an explicit function given below in (66), L_U is given in Assumption 2, λ by Assumption 1, \tilde{A} by Assumption 1 Lemma 1 and d is the dimension. Finally, a is used to bound an initial moment (see (65)).

The first of our main results concern the long-time convergence of the non-linear system (1).

Theorem 1 *Let $U \in \mathcal{C}^1(\mathbb{R}^d)$ satisfy Assumption 1 and either Assumption 2 or Assumption 4). There is an explicit $c^W > 0$ such that, for all $W \in \mathcal{C}^1(\mathbb{R}^d)$ satisfying Assumption 3 with $L_W < c^W$, there is an explicit $\tau > 0$ such that for all probability measures ν_0^1 and ν_0^2 on \mathbb{R}^{2d} with either a finite second moment (if Assumption 2 holds) or a finite Gaussian moment (if only Assumption 4 holds), there are explicit constants $C_1, C_2 > 0$ such that for all $t \geq 0$,*

$$\mathcal{W}_1(\nu_t^1, \nu_t^2) \leq e^{-\tau t} C_1, \quad \mathcal{W}_2(\nu_t^1, \nu_t^2) \leq e^{-\tau t} C_2$$

where ν_t^1 and ν_t^2 are solutions of (1) with respective initial distributions ν_0^1 and ν_0^2 .

In particular, we have existence and unicity of -as well as convergence towards - a stationary solution.

The second of our main results is a uniform in time convergence as $N \rightarrow +\infty$ of (3) toward (2).

Theorem 2 *Let $\tilde{\mathcal{C}}^0 > 0$ and $\tilde{a} > 0$. Let $U \in \mathcal{C}^1(\mathbb{R}^d)$ satisfy Assumptions 1 and 2. There is an explicit $c^W > 0$ such that, for all $W \in \mathcal{C}^1(\mathbb{R}^d)$ satisfying Assumption 3 with $L_W < c^W$, there exist explicit $B_1, B_2 > 0$, such that for all probability measures ν_0 on \mathbb{R}^{2d} satisfying $\mathbb{E}_{\nu_0}(e^{\tilde{a}(|X|+|V|)}) \leq \tilde{\mathcal{C}}^0$,*

$$\mathcal{W}_1(\nu_t^{k,N}, \bar{\nu}_t^{\otimes k}) \leq \frac{kB_1}{\sqrt{N}}, \quad \mathcal{W}_2(\nu_t^{k,N}, \bar{\nu}_t^{\otimes k}) \leq \frac{kB_2}{\sqrt{N}},$$

for all $k \in \mathbb{N}$, where $\nu_t^{k,N}$ is the marginal distribution at time t of the first k particles $((X_t^1, V_t^1), \dots, (X_t^k, V_t^k))$ of an N particle system (3) with initial distribution $(\nu_0)^{\otimes N}$, while $\bar{\nu}_t$ is a solution of (1) with initial distribution ν_0 .

1.3 Comparison to existing works

Space homogeneous models of diffusive and interacting granular media (see [3]) have attracted a lot of attention, usually named McKean-Vlasov diffusions, these past twenty years, by means of a stochastic interpretation and synchronous couplings as in [9] or the recent [13] by reflection couplings enabling to get rid of convexity conditions, but limited to small interactions. Remark however that small interactions are natural to get uniform in time propagation of chaos as for large interactions the non linear limit equation may have several stationary measures (see [23] for example). The granular media equations were interpreted as gradient flows in the space of probability measures in [8], leading to explicit exponential (or algebraic for non uniformly convex cases) rates of convergence to equilibrium of the non linear equation. Another approach relying on the dissipation of the Wasserstein distance and WJ inequalities was introduced in [5] handling small non convex cases. This approach was implemented in [32] to get propagation of chaos, under roughly the same type of assumptions.

Results on the long time behavior of the non-linear equation (2), i.e. space inhomogeneous, are few, as they combine the difficulty of getting explicit contraction rates for hypoelliptic diffusions as well as a non linear term. Concerning the uniform in time propagation of chaos, there are no results except in the strictly convex case (with very small perturbation). We however refer to [35] for a result on the torus with W bounded with continuous derivative of all orders, see also [7]. Based on functional inequalities (Poincaré or logarithmic Sobolev inequalities) for mean field models obtained in [19], other results were obtained provided the confining potential is a small perturbation of a quadratic function as in [29,18,21] which combines the hypocoercivity approach with independent of the number of particles constants appearing in the logarithmic Sobolev inequalities. Our results generalize [21]. Indeed, we may consider non gradient interactions whereas it is crucial in their approach to know explicitly the invariant measure of the particles system, and also we may handle only locally Lipschitz confinement potential, whereas they impose at most quadratic growth of the potentials, and non strictly convex at infinity potential. It is however difficult to compare the smallness of the interaction potentials needed in both approaches. Note however that they obtain convergence to equilibrium in entropy whereas we get it in Wasserstein distance (controlled by entropy through a Talagrand inequality). Using a coupling strategy, and more precisely synchronous couplings, results under strict convexity assumption were obtained in [6] for contraction rates in Wasserstein distance, see also [25] but only for the nonlinear system.

As we mentioned, we adapt a proof from [15], which tackles (2) without interaction term. The article uses a Lyapunov condition that guarantees the recurrence of the process on a compact set. This idea is common when proving similar results through a probabilistic lens (see for instance [34] or [1]). Lyapunov conditions may also help to implement hypocoercivity techniques *à la Villani* to handle entropic convergence for non quadratic potentials, see [10]. Under the assumption U "greater than a quadratic function" at infinity and ∇W Lipschitz continuous, we too consider a Lyapunov function that allows us to construct a specific semimetric improving the convergence speed. But, and this is to our knowledge something new, when proving propagation of chaos we add a form of non linearity in the quantity we consider to tackle a part of the non linearity appearing in the dynamic (see Section 4 below).

2 Modified semimetrics

As mentioned in the introduction, the proofs rely on the construction of suitable semimetrics on \mathbb{R}^{2d} and \mathbb{R}^{2dN} . They are introduced in this section, together with some useful properties. In all this section, $\lambda, A, \tilde{A}, L_U$ and L_W are given by Assumptions 1, 2 and 3 and Lemma 1.

Before going into the details, let us highlight the main points of the construction of the semimetrics. It relies on the superposition of three ideas. The first idea is that, in order to deal with the kinetic process (2), the standard

Euclidean norm $|x|^2 + |v|^2$ is not suitable and one should consider a linear change of variables, like $(x, v) \mapsto (x, x + \beta v)$ for some $\beta \in \mathbb{R}$. This is the case when using coupling methods as in [15, 6] but also when using hypocoercive modified entropies involving mixed derivatives as in [35, 34, 2, 10], the link being made in [30]. This motivates the definition of r below. The second idea is a modification of this distance r by some concave function f , which is related to the fact we are using, at least in some parts of the space, a reflection coupling. The concavity is well adapted to Itô's formula enabling the diffusion to provide a contraction effect (in a compact). This method has been considered for elliptic diffusions in [14], see also [16]. Intuitively, the contraction is produced by the fact that a random decrease in r has more effect on $f(r)$ than a random increase of the same amount. Finally, the third idea is the multiplication of a distance by a Lyapunov function G , which has first been used for Wasserstein distances in [22]. That way, on average, $f(r)G$ tends to decay because, when r is small, $f(r)$ tends to decay and, when r is large, G tends to decay.

2.1 A Lyapunov function

Let

$$\gamma = \frac{\lambda}{2(\lambda + 1)}, \quad B = 24 \left(A + (\lambda - \gamma) \tilde{A} + d \right) \quad (7)$$

and, for $x, v \in \mathbb{R}^d$,

$$H(x, v) = 24U(x) + (6(1 - \gamma) + \lambda)|x|^2 + 12x \cdot v + 12|v|^2.$$

For μ a probability measure on \mathbb{R}^d with finite first moment, ∇W being assumed Lipschitz continuous, denote by \mathcal{L}_μ the generator given by

$$\mathcal{L}_\mu \phi(x, v) = v \cdot \nabla_x \phi(x, v) - (v + \nabla U(x) + \nabla W * \mu(x)) \cdot \nabla_v \phi(x, v) + \Delta_v \phi(x, v).$$

The main properties of H are the following.

Lemma 3 *Under Assumptions 1, 2 and 3, for all $x, v \in \mathbb{R}^d$ and μ ,*

$$H(x, v) \geq 24U(x) + \lambda|x|^2 + 12 \left| v + \frac{x}{2} \right|^2, \quad (8)$$

$$\mathcal{L}_\mu H(x, v) \leq B + L_W(6 + 8\lambda) \left(\int |y| d\mu(y) \right)^2 - \left(\frac{3}{4}\lambda + \lambda^2 \right) |x|^2 - \gamma H(x, v), \quad (9)$$

$$\mathcal{L}_\mu H(x, v) \leq B + \left(\left(\int |y| d\mu(y) \right)^2 - |x|^2 \right) \left(\frac{3}{4}\lambda + \lambda^2 \right) - \gamma H(x, v). \quad (10)$$

In particular H is non-negative and goes to $+\infty$ at infinity.

The proof follows from elementary computations and is detailed in Appendix A.3. Notice that the condition $L_W \leq \lambda/8$ is used here.

In the case of particular interest where $\mu = \mu_t$ is given by (2), taking the expectation in (9) and using Gronwall's lemma, we immediately get the following.

Lemma 4 *Under Assumptions 1, 2 and 3, let $(X_t, V_t)_{t \geq 0}$ be a solution of (2) with finite second moment at initial time. For all $t \geq 0$,*

$$\frac{d}{dt} \mathbb{E}H(X_t, V_t) \leq B - \gamma \mathbb{E}H(X_t, V_t), \quad (11)$$

$$\mathbb{E}H(X_t, V_t) - \frac{B}{\gamma} \leq \left(\mathbb{E}H(X_0, V_0) - \frac{B}{\gamma} \right) e^{-\gamma t}. \quad (12)$$

2.2 Change of variable and concave modification

We start by fixing the values of some parameters. The somewhat intricate expressions in this section are dictated by the computations arising in the proofs later on. Recall the definition of γ and B in (7). Set

$$\alpha = L_U + \frac{\lambda}{4}, \quad R_0 = \sqrt{\frac{24B}{5\gamma \min(3, \frac{\lambda}{3})}}, \quad R_1 = \sqrt{\frac{24((1+\alpha)^2 + \alpha^2)}{5\gamma \min(3, \frac{\lambda}{3})}} B.$$

For $x, \tilde{x}, v, \tilde{v} \in \mathbb{R}^d$, set

$$r(x, \tilde{x}, v, \tilde{v}) = \alpha|x - \tilde{x}| + |x - \tilde{x} + v - \tilde{v}|.$$

Lemma 5 *Under Assumptions 1, 2 and 3, for all $x, \tilde{x}, v, \tilde{v} \in \mathbb{R}^d$,*

$$r(x, \tilde{x}, v, \tilde{v})^2 \leq 2 \frac{(1+\alpha)^2 + \alpha^2}{\min(\frac{1}{3}\lambda, 3)} (H(x, v) + H(\tilde{x}, \tilde{v})), \quad (13)$$

so that, in particular,

$$r(x, \tilde{x}, v, \tilde{v}) \geq R_1 \quad \Rightarrow \quad \gamma H(x, v) + \gamma H(\tilde{x}, \tilde{v}) \geq \frac{12}{5} B.$$

We refer to Appendix A.4 for the proof. Let

$$c = \min \left\{ \frac{\gamma}{36}, \frac{B}{3}, \frac{1}{7} \min \left(\frac{1}{2} - \frac{L_U + L_W}{2\alpha}, 2\sqrt{\frac{L_U + L_W}{2\pi\alpha}} \right) \exp \left(-\frac{1}{8} \left(\frac{L_U + L_W}{\alpha} + \alpha + 96 \max \left(\frac{1}{2\alpha}, 1 \right) \right) R_1^2 \right) \right\}. \quad (14)$$

Set

$$\epsilon = \frac{3c}{B}, \quad \mathbf{C} = c + 2\epsilon B$$

and, for $s \geq 0$,

$$\phi(s) = \exp \left(-\frac{1}{8} \left(\frac{1}{\alpha} (L_U + L_W) + \alpha + 96\epsilon \max \left(\frac{1}{2\alpha}, 1 \right) \right) s^2 \right), \quad \Phi(s) = \int_0^s \phi(u) du$$

$$g(s) = 1 - \frac{\mathbf{C}}{4} \int_0^s \frac{\Phi(u)}{\phi(u)} du, \quad f(s) = \int_0^{\min(s, R_1)} \phi(u) g(u) du.$$

Remark 1 The parameters above are far from being optimal. They are somewhat roughly chosen as we only wish to convey the fact that every constant is explicit.

The next lemma, proved in Appendix B, gathers the intermediary bounds that will be useful in the proofs of the main results.

Lemma 6 *Under Assumptions 1, 2 and 3,*

$$c \leq \frac{\gamma}{6} \left(1 - \frac{\frac{5\gamma}{6}}{2\epsilon B + \frac{5\gamma}{6}} \right), \quad (15)$$

$$L_U + L_W < \alpha, \quad (16)$$

$$c + 2\epsilon B \leq \frac{1}{2} \left(1 - \frac{L_U + L_W}{\alpha} \right) \inf_{r \in]0, R_1]} \frac{r\phi(r)}{\Phi(r)}, \quad (17)$$

$$c + 2\epsilon B \leq 2 \left(\int_0^{R_1} \Phi(s) \phi(s)^{-1} ds \right)^{-1}, \quad (18)$$

$$\forall s \geq 0, \quad 0 = 4\phi'(s) + \left(\frac{1}{\alpha} (L_U + L_W) + \alpha + 96\epsilon \max\left(\frac{1}{2\alpha}, 1\right) \right) s\phi(s). \quad (19)$$

The main properties of f are the following.

Lemma 7 *The function f is \mathcal{C}^2 on $(0, R_1)$ with $f'_+(0) = 1$ and $f'_-(R_1) > 0$, and constant on $[R_1, \infty)$. Moreover, it is non-negative, non-decreasing and concave, and for all $s \geq 0$,*

$$\min(s, R_1) f'_-(R_1) \leq f(s) \leq \min(s, f(R_1)) \leq \min(s, R_1).$$

Proof First, notice that (18) ensures that $g(s) \geq \frac{1}{2}$ for all $s \geq 0$. Then, all the points immediately follow from the fact the functions ϕ and g are \mathcal{C}^2 , positive and decreasing, with $\phi(0) = g(0) = 1$.

2.3 The modified semimetrics

For $x, \tilde{x}, v, \tilde{v} \in \mathbb{R}^d$, set

$$G(x, v, \tilde{x}, \tilde{v}) = 1 + \epsilon H(x, v) + \epsilon H(\tilde{x}, \tilde{v}),$$

$$\rho(x, v, \tilde{x}, \tilde{v}) = f(r(x, v, \tilde{x}, \tilde{v})) G(x, v, \tilde{x}, \tilde{v}).$$

An immediate corollary of Lemmas 5 and 7 is that ρ is a semimetric on \mathbb{R}^{2d} which controls the usual L1 and L2 distances:

Lemma 8 *There are explicit constants $C_1, C_2, C_r, C_z > 0$ such that for all $x, x', v, v' \in \mathbb{R}^d$,*

$$\begin{aligned} |x - x'| + |v - v'| &\leq C_1 \rho((x, v), (x', v')) \\ |x - x'|^2 + |v - v'|^2 &\leq C_2 \rho((x, v), (x', v')) \\ r(x, v, x', v') &\leq C_r \rho((x, v), (x', v')) \\ |x - x'| &\leq C_z f(r(x, v, x', v')) \left(1 + \epsilon \sqrt{H(x, v)} + \epsilon \sqrt{H(x', v')}\right). \end{aligned}$$

We also mention a technical lemma, see Appendix A.6 for proof.

Lemma 9 *For all $x, v, \tilde{x}, \tilde{v} \in \mathbb{R}^d$*

$$|H(x, v) - H(\tilde{x}, \tilde{v})| \leq C_{dH,1} r(x, \tilde{x}, v, \tilde{v}) + C_{dH,2} r(x, \tilde{x}, v, \tilde{v}) \left(\sqrt{H(x, v)} + \sqrt{H(\tilde{x}, \tilde{v})} \right), \quad (20)$$

where

$$C_{dH,1} := \frac{24|\nabla U(0)|}{\alpha} \quad \text{and} \quad C_{dH,2} := \frac{24L_U}{\alpha\sqrt{\lambda}} + \frac{6(1-\gamma) + \lambda - 3}{\alpha\sqrt{\lambda}} + 2\sqrt{3} \max\left(1, \frac{1}{2\alpha}\right).$$

Finally, for μ and ν two probability measures on \mathbb{R}^{2d} and a measurable function $h : \mathbb{R}^{2d} \times \mathbb{R}^{2d} \rightarrow \mathbb{R}$, we define

$$\mathcal{W}_h(\mu, \nu) = \inf_{\Gamma \in \Pi(\mu, \nu)} \int h(x, v, \tilde{x}, \tilde{v}) \Gamma(dx, dv, d\tilde{x}, d\tilde{v}).$$

3 Proof of Theorem 1

In this section, for the sake of clarity, we only assume the potential U satisfies Assumption 1 and Assumption 2. We refer to Appendix C for the adjustment of the proof in the case ∇U locally Lipschitz continuous.

Our goal is to prove the following result

Theorem 3 *Let $C^0 > 0$. Let $U \in \mathcal{C}^2(\mathbb{R}^d)$ satisfy Assumption 1 and Assumption 2. Let*

$$\tilde{C}_K := C_1 \left(1 + \frac{2\epsilon B}{\gamma} + 2\epsilon C^0\right) + 2\epsilon \left(\frac{B}{\gamma} + C^0\right) \frac{6 + 8\lambda}{\lambda}.$$

For all $W \in \mathcal{C}^2(\mathbb{R}^d)$ satisfying Assumption 3 with $L_W < c/\tilde{C}_K$, for all probability measures ν_0^1 and ν_0^2 on \mathbb{R}^{2d} satisfying $\mathbb{E}_{\nu_0^1} H \leq C^0$ and $\mathbb{E}_{\nu_0^2} H \leq C^0$

$$\forall t \geq 0, \quad \mathcal{W}_\rho(\nu_t^1, \nu_t^2) \leq e^{-(c-L_W \tilde{C}_K)t} \mathcal{W}_\rho(\nu_0^1, \nu_0^2),$$

where ν_t^1 (resp. ν_t^2) is a solution of (1) with initial distribution ν_0^1 (resp. ν_0^2).

3.1 Step one: Coupling and evolution of the coupling semimetric

Let $\xi > 0$, and let $rc, sc : \mathbb{R}^{2d} \mapsto [0, 1]$ be two Lipschitz continuous functions such that :

$$rc^2 + sc^2 = 1,$$

$$rc(z, w) = 0 \text{ if } |z + w| \leq \frac{\xi}{2} \text{ or } \alpha|z| + |z + w| \geq R_1 + \xi,$$

$$rc(z, w) = 1 \text{ if } |z + w| \geq \xi \text{ and } \alpha|z| + |z + w| \leq R_1.$$

These two functions translate into mathematical terms the regions in which we use a reflection coupling (represented by $rc = 1$) and the ones where we use a synchronous coupling (represented by $sc = 1$). Finally, ξ is a parameter that will vanish to zero in the end. We therefore consider the following coupling :

$$\begin{cases} dX_t = V_t dt \\ dV_t = -V_t dt - \nabla U(X_t) dt - \nabla W * \mu_t(X_t) dt + \sqrt{2}rc(Z_t, W_t) dB_t^{rc} + \sqrt{2}sc(Z_t, W_t) dB_t^{sc} \\ \mu_t = \text{Law}(X_t) \\ d\tilde{X}_t = \tilde{V}_t dt \\ d\tilde{V}_t = -\tilde{V}_t dt - \nabla U(\tilde{X}_t) dt - \nabla W * \tilde{\mu}_t(\tilde{X}_t) dt + \sqrt{2}rc(Z_t, W_t) (Id - 2e_t e_t^T) dB_t^{rc} \\ \quad + \sqrt{2}sc(Z_t, W_t) dB_t^{sc} \\ \tilde{\mu}_t = \text{Law}(\tilde{X}_t), \end{cases}$$

where B^{rc} and B^{sc} are independent Brownian motions, and

$$Z_t = X_t - \tilde{X}_t, \quad W_t = V_t - \tilde{V}_t, \quad Q_t = Z_t + W_t, \quad e_t = \begin{cases} \frac{Q_t}{|Q_t|} & \text{if } Q_t \neq 0, \\ 0 & \text{otherwise,} \end{cases}$$

and e_t^T is the transpose of e_t . Then

$$\frac{dZ_t}{dt} = W_t = Q_t - Z_t. \quad (21)$$

So $\frac{d|Z_t|}{dt} = \frac{Z_t}{|Z_t|} (Q_t - Z_t)$ for every t such that $Z_t \neq 0$, and $\frac{d|Z_t|}{dt} \leq |Q_t|$ for every t such that $Z_t = 0$. In particular

$$\frac{d|Z_t|}{dt} \leq |Q_t| - |Z_t|.$$

We start by using Itô's formula to compute the evolution of $|Q_t|$. The following lemma is Lemma 7 of A. Durmus *et al.* [13] of which, for the sake of completeness, we give the proof.

Lemma 10 *Under Assumption 1, Assumption 2 and Assumption 3, we have almost surely for all $t \geq 0$.*

$$\begin{aligned} d|Q_t| = & -e_t \cdot \left(\nabla U(X_t) - \nabla U(\tilde{X}_t) \right) dt - e_t \cdot \left(\nabla W * \mu_t(X_t) - \nabla W * \tilde{\mu}_t(\tilde{X}_t) \right) dt \\ & + 2\sqrt{2}rc(Z_t, W_t) e_t \cdot dB_t^{rc} \end{aligned} \quad (22)$$

Proof Let $t \geq 0$. We begin by considering the dynamics of Z_t , W_t and Q_t . We have

$$\begin{aligned} dZ_t &= W_t dt \\ dW_t &= -W_t dt - \left(\nabla U(X_t) - \nabla U(\tilde{X}_t) \right) dt - \left(\nabla W * \mu_t(X_t) - \nabla W * \tilde{\mu}_t(\tilde{X}_t) \right) dt \\ &\quad + 2\sqrt{2}rc(Z_t, W_t) e_t e_t \cdot dB_t^{rc} \\ dQ_t &= - \left(\nabla U(X_t) - \nabla U(\tilde{X}_t) \right) dt - \left(\nabla W * \mu_t(X_t) - \nabla W * \tilde{\mu}_t(\tilde{X}_t) \right) dt + 2\sqrt{2}rc(Z_t, W_t) e_t e_t \cdot dB_t^{rc}. \end{aligned}$$

Therefore

$$\begin{aligned} d|Q_t|^2 &= -2Q_t \cdot \left(\nabla U(X_t) - \nabla U(\tilde{X}_t) \right) dt - 2Q_t \cdot \left(\nabla W * \mu_t(X_t) - \nabla W * \tilde{\mu}_t(\tilde{X}_t) \right) dt \\ &\quad + 4\sqrt{2}rc(Z_t, W_t) (Q_t \cdot e_t) e_t \cdot dB_t^{rc} + 8rc^2(Z_t, W_t) dt. \end{aligned}$$

We consider, for $\eta > 0$, the function $\psi_\eta(r) = (r + \eta)^{1/2}$ which is \mathcal{C}^∞ on $]0, \infty[$ and satisfies

$$\begin{aligned} \forall r \geq 0, \quad \lim_{\eta \rightarrow 0} \psi_\eta(r) &= r^{1/2}, \quad \lim_{\eta \rightarrow 0} 2\psi'_\eta(r) = r^{-1/2}, \quad \lim_{\eta \rightarrow 0} 4\psi''_\eta(r) = -r^{-3/2}, \\ \text{and thus } \lim_{\eta \rightarrow 0} 2r\psi''_\eta(r) + \psi'_\eta(r) &= 0. \end{aligned}$$

Then

$$\begin{aligned} d\psi_\eta(|Q_t|^2) &= -2\psi'_\eta(|Q_t|^2) Q_t \cdot \left(\nabla U(X_t) - \nabla U(\tilde{X}_t) \right) dt \\ &\quad - 2\psi'_\eta(|Q_t|^2) Q_t \cdot \left(\nabla W * \mu_t(X_t) - \nabla W * \tilde{\mu}_t(\tilde{X}_t) \right) dt \\ &\quad + 4\psi'_\eta(|Q_t|^2) \sqrt{2}rc(Z_t, W_t) (Q_t \cdot e_t) e_t \cdot dB_t^{rc} + 8\psi'_\eta(|Q_t|^2) rc^2(Z_t, W_t) dt \\ &\quad + 16\psi''_\eta(|Q_t|^2) rc^2(Z_t, W_t) |Q_t|^2 dt. \end{aligned}$$

We make sure each individual term converges almost surely as $\eta \rightarrow 0$. First, we notice that

$$2|Q_t|\psi'_\eta(|Q_t|^2) = \frac{|Q_t|}{(|Q_t|^2 + \eta)^{1/2}} \leq 1.$$

So

$$2\psi'_\eta(|Q_t|^2) Q_t \cdot \left(\nabla U(X_t) - \nabla U(\tilde{X}_t) \right) \leq |\nabla U(X_t) - \nabla U(\tilde{X}_t)| \leq L_U |Z_t|.$$

Then, by dominated convergence, for all $T \geq 0$ almost surely

$$\begin{aligned} \lim_{\eta \rightarrow 0} \int_0^T 2\psi'_\eta(|Q_t|^2) Q_t \cdot \left(\nabla U(X_t) - \nabla U(\tilde{X}_t) \right) dt &= \int_0^T \frac{Q_t}{|Q_t|} \cdot \left(\nabla U(X_t) - \nabla U(\tilde{X}_t) \right) dt \\ &= \int_0^T e_t \cdot \left(\nabla U(X_t) - \nabla U(\tilde{X}_t) \right) dt. \end{aligned}$$

Likewise for all $T \geq 0$

$$\begin{aligned} 2\psi'_\eta(|Q_t|^2) Q_t \cdot \left(\nabla W * \mu_t(X_t) - \nabla W * \tilde{\mu}_t(\tilde{X}_t) \right) &\leq \left| \nabla W * \mu_t(X_t) - \nabla W * \tilde{\mu}_t(\tilde{X}_t) \right| \\ &\leq L_W |Z_t| + L_W \mathbb{E} |Z_t|, \end{aligned}$$

hence

$$\begin{aligned} \lim_{\eta \rightarrow 0} \int_0^T 2\psi'_\eta(|Q_t|^2) Q_t \cdot \left(\nabla W * \mu_t(X_t) - \nabla W * \tilde{\mu}_t(\tilde{X}_t) \right) dt \\ = \int_0^T e_t \cdot \left(\nabla W * \mu_t(X_t) - \nabla W * \tilde{\mu}_t(\tilde{X}_t) \right) dt. \end{aligned}$$

Then, since $rc(Z_t, W_t) = 0$ for $|Q_t| \leq \frac{\xi}{2}$ and

$$\begin{aligned} 8\psi'_\eta(|Q_t|^2) + 16\psi''_\eta(|Q_t|^2) |Q_t|^2 &= 4 \left(\frac{1}{(|Q_t|^2 + \eta)^{1/2}} - \frac{|Q_t|^2}{(|Q_t|^2 + \eta)^{3/2}} \right) \\ &= 4 \frac{\eta}{(|Q_t|^2 + \eta)^{3/2}} \leq \frac{4\eta}{|Q_t|^3}, \end{aligned}$$

we have by dominated convergence

$$\lim_{\eta \rightarrow 0} \int_0^T (8d\psi'_\eta(|Q_t|^2) rc^2(Z_t, W_t) + 16\psi''_\eta(|Q_t|^2) rc^2(Z_t, W_t) |Q_t|^2) dt = 0.$$

Finally, by Theorem 2.12 chapter 4 of [31]

$$\lim_{\eta \rightarrow 0} \int_0^T 4\sqrt{2}\psi'_\eta(|Q_t|^2) rc(Z_t, W_t) (Q_t \cdot e_t) e_t \cdot dB_t^{rc} = \int_0^T 2\sqrt{2}rc(Z_t, W_t) e_t \cdot dB_t^{rc}.$$

For any t , we obtain the desired result almost surely. The continuity of $t \mapsto |Q_t|$ then allows us to conclude that (22) is almost surely true for all t .

We denote

$$r_t := \alpha |X_t - \tilde{X}_t| + |X_t - \tilde{X}_t + V_t - \tilde{V}_t| = \alpha |Z_t| + |Q_t|, \quad (23)$$

$$\rho_t := f(r_t) G_t \text{ where } G_t = 1 + \epsilon H(X_t, V_t) + \epsilon H(\tilde{X}_t, \tilde{V}_t). \quad (24)$$

Since $H(x, v) \geq 0$ we have $G_t \geq 1$. We now state the main lemma of this section.

Lemma 11 *Under Assumption 1, Assumption 2 and Assumption 3, let $c \in]0, \infty[$. Then almost surely for all $t \geq 0$*

$$\forall t \geq 0, e^{ct} \rho_t \leq \rho_0 + \int_0^t e^{cs} K_s ds + M_t, \quad (25)$$

where $(M_t)_t$ is a continuous local martingale and

$$\begin{aligned} K_t = & 4f''(r_t)rc(Z_t, W_t)^2 G_t + cf(r_t)G_t + \left(\alpha \frac{d|Z_t|}{dt} + (L_U + L_W)|Z_t| + L_W \mathbb{E}|Z_t| \right) f'(r_t)G_t \\ & + \epsilon \left(2B - \gamma H(X_t, V_t) - \gamma H(\tilde{X}_t, \tilde{V}_t) \right) f(r_t) + \epsilon L_W (6 + 8\lambda) \left(\mathbb{E}(|X_t|)^2 + \mathbb{E}(|\tilde{X}_t|)^2 \right) f(r_t) \\ & + 96\epsilon \max \left(1, \frac{1}{2\alpha} \right) r_t f'(r_t) rc(Z_t, W_t)^2. \end{aligned}$$

Proof Using (22)

$$\begin{aligned} |Q_t| &= |Q_0| + A_t^Q + M_t^Q \text{ with} \\ dA_t^Q &= -e_t \cdot \left(\nabla U(X_t) - \nabla U(\tilde{X}_t) \right) dt - e_t \cdot \left(\nabla W * \mu_t(X_t) - \nabla W * \tilde{\mu}_t(\tilde{X}_t) \right) dt \\ dM_t^Q &= 2\sqrt{2}rc(Z_t, W_t) e_t \cdot dB_t^{rc}. \end{aligned}$$

Therefore $r_t = |Q_0| + \alpha|Z_t| + A_t^Q + M_t^Q$. Let $c > 0$. By Itô's formula

$$d(e^{ct}f(r_t)) = ce^{ct}f(r_t)dt + e^{ct}f'(r_t)dr_t + \frac{1}{2}e^{ct}f''(r_t)8rc^2(Z_t, W_t)dt.$$

Hence

$$\begin{aligned} e^{ct}f(r_t) &= f(r_0) + \hat{A}_t + \hat{M}_t \text{ with} \\ d\hat{A}_t &= \left(cf(r_t) + \alpha f'(r_t) \frac{d|Z_t|}{dt} - f'(r_t)e_t \cdot \left(\nabla U(X_t) - \nabla U(\tilde{X}_t) \right) \right. \\ &\quad \left. - f'(r_t)e_t \cdot \left(\nabla W * \mu_t(X_t) - \nabla W * \tilde{\mu}_t(\tilde{X}_t) \right) + 4f''(r_t)rc^2(Z_t, W_t) \right) e^{ct}dt \\ d\hat{M}_t &= e^{ct}2\sqrt{2}f'(r_t)rc(Z_t, W_t)e_t \cdot dB_t^{rc}. \end{aligned}$$

We now consider the evolution of $G_t = 1 + \epsilon H(X_t, V_t) + \epsilon H(\tilde{X}_t, \tilde{V}_t)$

$$\begin{aligned} dG_t &= \epsilon \left(\mathcal{L}_{\mu_t} H(X_t, V_t) + \mathcal{L}_{\tilde{\mu}_t} H(\tilde{X}_t, \tilde{V}_t) \right) dt \\ &\quad + \epsilon \sqrt{2}rc(Z_t, W_t) \left(\nabla_v H(X_t, V_t) - \nabla_v H(\tilde{X}_t, \tilde{V}_t) \right) \cdot e_t e_t^T dB_t^{rc} \\ &\quad + \epsilon \sqrt{2}rc(Z_t, W_t) \left(\nabla_v H(X_t, V_t) + \nabla_v H(\tilde{X}_t, \tilde{V}_t) \right) \cdot (Id - e_t e_t^T) dB_t^{rc} \\ &\quad + \epsilon \sqrt{2}sc(Z_t, W_t) \left(\nabla_v H(X_t, V_t) + \nabla_v H(\tilde{X}_t, \tilde{V}_t) \right) \cdot dB_t^{sc}. \end{aligned}$$

Therefore $e^{ct}\rho_t = e^{ct}f(r_t)G_t = \rho_0 + A_t + M_t$, where

$$\begin{aligned} dA_t &= G_t d\hat{A}_t + \epsilon e^{ct}f(r_t) \left(\mathcal{L}_{\mu_t} H(X_t, V_t) + \mathcal{L}_{\tilde{\mu}_t} H(\tilde{X}_t, \tilde{V}_t) \right) dt \\ &\quad + 4\epsilon e^{ct}f'(r_t)rc^2(Z_t, W_t) \left(\nabla_v H(X_t, V_t) - \nabla_v H(\tilde{X}_t, \tilde{V}_t) \right) \cdot e_t dt, \end{aligned}$$

and M_t is a continuous local martingale. This last equality uses the fact that B^{rc} and B^{sc} are independent Brownian motion and that $e_t \cdot (Id - e_t e_t^T) = 0$. Furthermore

$$\begin{aligned} |\nabla_v H(X_t, V_t) - \nabla_v H(\tilde{X}_t, \tilde{V}_t)| &= 12|X_t + 2V_t - \tilde{X}_t - 2\tilde{V}_t| = 12|2Q_t - Z_t| \\ &\leq 24 \left(\frac{1}{2}|Z_t| + |Q_t| \right) \\ &\leq 24 \max \left(1, \frac{1}{2\alpha} \right) r_t, \end{aligned}$$

so that $dA_t \leq e^{ct} \tilde{K}_t dt$, where

$$\begin{aligned} \tilde{K}_t &= \left(cf(r_t) + \alpha f'(r_t) \frac{d|Z_t|}{dt} - f'(r_t) e_t \cdot (\nabla U(X_t) - \nabla U(\tilde{X}_t)) \right. \\ &\quad \left. - f'(r_t) e_t \cdot (\nabla W * \mu_t(X_t) - \nabla W * \tilde{\mu}_t(\tilde{X}_t)) + 4f''(r_t) rc^2(Z_t, W_t) \right) G_t \\ &\quad + \epsilon \left(\mathcal{L}_t H(X_t, V_t) + \mathcal{L}_t H(\tilde{X}_t, \tilde{V}_t) \right) f(r_t) + 96\epsilon \max \left(1, \frac{1}{2\alpha} \right) r_t f'(r_t) rc^2(Z_t, W_t). \end{aligned}$$

And we conclude using Lemma 2, and Lemma 3.

3.2 Step two : Contractivity in various regions of space

At this point, we have

$$\forall t \geq 0, e^{ct} \rho_t \leq \rho_0 + \int_0^t e^{cs} K_s ds + M_t,$$

where M_t is a continuous local martingale and, by regrouping the terms according to how we will use them

$$K_t = \left(cf(r_t) + \left(\alpha \frac{d|Z_t|}{dt} + (L_U + L_W)|Z_t| \right) f'(r_t) \right) G_t \quad (26)$$

$$+ 4 \left(f''(r_t) G_t + 24\epsilon \max \left(1, \frac{1}{2\alpha} \right) r_t f'(r_t) \right) rc(Z_t, W_t)^2 \quad (27)$$

$$+ \epsilon \left(2B - \gamma H(X_t, V_t) - \gamma H(\tilde{X}_t, \tilde{V}_t) \right) f(r_t) \quad (28)$$

$$+ L_W f'(r_t) \mathbb{E}(|Z_t|) G_t + \epsilon L_W (6 + 8\lambda) \left(\mathbb{E}(|X_t|)^2 + \mathbb{E}(|\tilde{X}_t|)^2 \right) f(r_t). \quad (29)$$

Briefly,

- lines (26) and (27) will be non positive thanks to the construction of the function f when using the reflection coupling,
- when only using the synchronous coupling, i.e when the deterministic drift is contracting, line (26) alone will be sufficiently small,

- line (28) translates the effect the Lyapunov function has in bringing back processes that would have ventured at infinity,
- finally, line (29) contains the non linearity and will be tackled by taking L_W sufficiently small.

In this section, we thus prove the following lemma

Lemma 12 *Assume the confining potential U satisfies Assumption 1 and Assumption 2, there is a constant $c^W > 0$ such that for all interaction potential W satisfying Assumption 3 with $L_W < c^W$, the following holds for K_t defined in (26)-(29)*

$$\mathbb{E}K_t \leq (1 + \alpha) \xi \mathbb{E}G_t + L_W (\mathcal{C}_K + \mathcal{C}_K^0 e^{-\gamma t}) \mathbb{E}\rho_t,$$

with

$$\mathcal{C}_K = \mathcal{C}_1 \left(1 + \frac{2\epsilon B}{\gamma} \right) + \frac{2\epsilon B}{\gamma\lambda} (6 + 8\lambda), \quad (30)$$

$$\mathcal{C}_K^0 = \epsilon \left(\mathcal{C}_1 + \frac{6 + 8\lambda}{\lambda} \right) \left(\mathbb{E}H(X_0, V_0) + \mathbb{E}H(\tilde{X}_0, \tilde{V}_0) \right).$$

The constant c^W is explicit, as it will be shown in Appendix B.

To this end, we divide the space into three regions

$$\begin{aligned} \text{Reg}_1 &= \left\{ (X_t, V_t, \tilde{X}_t, \tilde{V}_t) \text{ s.t. } |Q_t| \geq \xi \text{ and } r_t \leq R_1 \right\}, \\ \text{Reg}_2 &= \left\{ (X_t, V_t, \tilde{X}_t, \tilde{V}_t) \text{ s.t. } |Q_t| < \xi \text{ and } r_t \leq R_1 \right\}, \\ \text{Reg}_3 &= \left\{ (X_t, V_t, \tilde{X}_t, \tilde{V}_t) \text{ s.t. } r_t > R_1 \right\}, \end{aligned}$$

and consider

$$\mathbb{E}K_t = \mathbb{E}(K_t \mathbf{1}_{\text{Reg}_1}) + \mathbb{E}(K_t \mathbf{1}_{\text{Reg}_2}) + \mathbb{E}(K_t \mathbf{1}_{\text{Reg}_3}).$$

3.2.1 First region : $|Q_t| \geq \xi$ and $r_t \leq R_1$

In this region of space, we use the Brownian motion through the reflection coupling and the construction of the function f to bring the processes closer together. Here we have $rc(Z_t, W_t) = 1$. Recall $\alpha|Z_t| + |Q_t| = r_t$ and $G_t \geq 1$.

- We have

$$\begin{aligned} \alpha \frac{d|Z_t|}{dt} + (L_U + L_W) |Z_t| &\leq \alpha|Q_t| - \alpha|Z_t| + (L_U + L_W) |Z_t| \\ &= \alpha r_t - \alpha^2 |Z_t| - \alpha|Z_t| + (L_U + L_W) |Z_t| \\ &\leq \left(\frac{1}{\alpha} (L_U + L_W) + \alpha \right) r_t. \end{aligned}$$

– Since $G_t = 1 + \epsilon H(X_t, V_t) + \epsilon H(\tilde{X}_t, \tilde{V}_t) \geq 1$,

$$cG_t + \epsilon \left(2B - \gamma H(X_t, V_t) - \gamma H(\tilde{X}_t, \tilde{V}_t) \right) \leq cG_t + 2\epsilon B G_t = \mathbf{C} G_t. \quad (31)$$

– We then have, by (19),

$$4\phi'(r_t) + \left(\frac{1}{\alpha} (L_U + L_W) + \alpha + 96\epsilon \max\left(\frac{1}{2\alpha}, 1\right) \right) r_t \phi(r_t) = 0.$$

Hence

$$\begin{aligned} 4f''(r_t) + \left(\frac{1}{\alpha} (L_U + L_W) + \alpha + 96\epsilon \max\left(\frac{1}{2\alpha}, 1\right) \right) r_t f'(r_t) \\ = 4\phi'(r_t) g(r_t) + 4\phi(r_t) g'(r_t) \\ + \left(\frac{1}{\alpha} (L_U + L_W) + \alpha + 96\epsilon \max\left(\frac{1}{2\alpha}, 1\right) \right) r_t \phi(r_t) g(r_t) \\ = 4\phi(r_t) g'(r_t), \end{aligned}$$

and

$$4\phi(r_t) g'(r_t) + \mathbf{C} f(r_t) \leq -4 \frac{\mathbf{C}}{4} \Phi(r_t) + \mathbf{C} \Phi(r_t) = 0.$$

– At this point, through this choice of function f , we are left with

$$K_t \mathbf{1}_{\text{Reg}_1} \leq L_W f'(r_t) \mathbb{E}(|Z_t|) G_t + \epsilon L_W (6 + 8\lambda) \left(\mathbb{E}(|X_t|)^2 + \mathbb{E}(|\tilde{X}_t|)^2 \right) f(r_t).$$

Using Lemma 8, $f'(r_t) \leq 1$ and $G_t \geq 1$,

$$\mathbb{E} \left(K_t \mathbf{1}_{\text{Reg}_1} \right) \leq L_W \mathcal{C}_1 \mathbb{E}(\rho_t) \mathbb{E}(G_t) + \epsilon L_W (6 + 8\lambda) \left(\mathbb{E}(|X_t|)^2 + \mathbb{E}(|\tilde{X}_t|)^2 \right) \mathbb{E}(\rho_t).$$

Recall (12)

$$\begin{aligned} \mathbb{E}(G_t) &= 1 + \epsilon \mathbb{E}H(X_t, V_t) + \epsilon \mathbb{E}H(\tilde{X}_t, \tilde{V}_t), \\ &\leq 1 + \frac{2\epsilon B}{\gamma} + \epsilon \left(\mathbb{E}H(X_0, V_0) + \mathbb{E}H(\tilde{X}_0, \tilde{V}_0) \right) e^{-\gamma t}, \end{aligned}$$

and, since $H(x, v) \geq \lambda|x|^2$,

$$\begin{aligned} \mathbb{E}(|X_t|)^2 + \mathbb{E}(|\tilde{X}_t|)^2 &\leq \frac{1}{\lambda} \mathbb{E}H(X_t, V_t) + \frac{1}{\lambda} \mathbb{E}H(\tilde{X}_t, \tilde{V}_t), \\ &\leq \frac{2B}{\gamma\lambda} + \frac{1}{\lambda} \left(\mathbb{E}H(X_0, V_0) + \mathbb{E}H(\tilde{X}_0, \tilde{V}_0) \right) e^{-\gamma t}. \end{aligned}$$

Hence

$$\begin{aligned} \mathbb{E} \left(K_t \mathbf{1}_{\text{Reg}_1} \right) &\leq L_W \left(\mathcal{C}_1 \left(1 + \frac{2\epsilon B}{\gamma} \right) + \frac{2\epsilon B}{\gamma\lambda} (6 + 8\lambda) \right) \mathbb{E}(\rho_t) \\ &\quad + L_W \epsilon \left(\mathcal{C}_1 + \frac{6 + 8\lambda}{\lambda} \right) \left(\mathbb{E}H(X_0, V_0) + \mathbb{E}H(\tilde{X}_0, \tilde{V}_0) \right) \mathbb{E}(\rho_t) e^{-\gamma t}. \end{aligned}$$

We thus obtain $\mathbb{E} \left(K_t \mathbf{1}_{\text{Reg}_1} \right) \leq L_W (\mathcal{C}_K + \mathcal{C}_K^0 e^{-\gamma t}) \mathbb{E}\rho_t$.

3.2.2 Second region : $|Q_t| < \xi$ and $r_t \leq R_1$

In this region of space, we use the naturally contracting deterministic drift thanks to a synchronous coupling. Here $R_1 \geq r_t \geq \alpha|Z_t| \geq r_t - \xi$ so that

$$\begin{aligned} K_t &\leq \mathbf{C}f(r_t)G_t + \left(\alpha\xi - r_t + \xi + \frac{1}{\alpha}(L_U + L_W)r_t\right)f'(r_t)G_t \\ &\quad + \left(4f''(r_t)G_t + 96\epsilon \max\left(\frac{1}{2\alpha}, 1\right)r_tf'(r_t)\right)rc(Z_t, W_t)^2 \\ &\quad + L_W f'(r_t)\mathbb{E}(|Z_t|)G_t + \epsilon L_W(6 + 8\lambda)\left(\mathbb{E}(|X_t|)^2 + \mathbb{E}(|\tilde{X}_t|)^2\right)f(r_t), \end{aligned}$$

where we use (31). First

$$4f''(r_t)G_t + 96\epsilon \max\left(\frac{1}{2\alpha}, 1\right)r_tf'(r_t) \leq 0.$$

We use (16) to obtain, since $f(r_t) \leq \Phi(r_t)$ and $\frac{1}{2}\phi(r_t) \leq f'(r_t) = \phi(r_t)g(r_t) \leq \phi(r_t)$ by (18),

$$\begin{aligned} K_t &\leq \xi(1 + \alpha)\phi(r_t)g(r_t)G_t + G_t\left(\mathbf{C}\Phi(r_t) + \frac{1}{2}\left(\frac{1}{\alpha}(L_U + L_W) - 1\right)r_t\phi(r_t)\right) \\ &\quad + L_W f'(r_t)\mathbb{E}(|Z_t|)G_t + \epsilon L_W(6 + 8\lambda)\left(\mathbb{E}(|X_t|)^2 + \mathbb{E}(|\tilde{X}_t|)^2\right)f(r_t). \end{aligned}$$

Then, thanks to (17), like in the first region of space

$$\begin{aligned} K_t \mathbf{1}_{\text{Reg}_2} &\leq \xi(1 + \alpha)\phi(r_t)g(r_t)G_t + L_W f'(r_t)\mathbb{E}(|Z_t|)G_t \\ &\quad + \epsilon L_W(6 + 8\lambda)\left(\mathbb{E}(|X_t|)^2 + \mathbb{E}(|\tilde{X}_t|)^2\right)f(r_t). \end{aligned}$$

Hence, since $\phi(r_t)g(r_t) \leq 1$

$$\mathbb{E}K_t \mathbf{1}_{\text{Reg}_2} \leq \xi(1 + \alpha)\mathbb{E}(G_t) + L_W(\mathcal{C}_K + \mathcal{C}_K^0 e^{-\gamma t})\mathbb{E}\rho_t.$$

3.2.3 Third region : $r_t > R_1$

In this region, we use the Lyapunov function. Here $f'(r_t) = f''(r_t) = 0$ so that

$$\begin{aligned} K_t \mathbf{1}_{\text{Reg}_3} &= \left(cG_t + \epsilon\left(2B - \gamma H(X_t, V_t) - \gamma H(\tilde{X}_t, \tilde{V}_t)\right)\right)f(r_t)\mathbf{1}_{\text{Reg}_3} \\ &\quad + \epsilon L_W(6 + 8\lambda)\left(\mathbb{E}(|X_t|)^2 + \mathbb{E}(|\tilde{X}_t|)^2\right)f(r_t)\mathbf{1}_{\text{Reg}_3} \\ &= \left[\epsilon(c - \gamma)\left(H(X_t, V_t) + H(\tilde{X}_t, \tilde{V}_t)\right) + 2\epsilon B + c\right. \\ &\quad \left.+ \epsilon L_W(6 + 8\lambda)\left(\mathbb{E}(|X_t|)^2 + \mathbb{E}(|\tilde{X}_t|)^2\right)\right]f(r_t)\mathbf{1}_{\text{Reg}_3} \end{aligned}$$

Since $c - \gamma < 0$ as a consequence of (15), and using Lemma 5

$$\begin{aligned} K_t &\leq \left((c - \gamma) \epsilon \frac{12B}{5\gamma} + 2\epsilon B + c \right) f(r_t) + \epsilon L_W (6 + 8\lambda) \left(\mathbb{E}(|X_t|)^2 + \mathbb{E}(|\tilde{X}_t|)^2 \right) f(r_t) \\ &\leq \left(c \left(\frac{12\epsilon B}{5\gamma} + 1 \right) - \frac{2}{5}\epsilon B \right) f(r_t) + \epsilon L_W (6 + 8\lambda) \left(\mathbb{E}(|X_t|)^2 + \mathbb{E}(|\tilde{X}_t|)^2 \right) f(r_t). \end{aligned}$$

Then, using (15), $\mathbb{E}K_t \mathbb{1}_{\text{Reg}_3} \leq L_W \mathcal{C}_K \mathbb{E}\rho_t + L_W \mathcal{C}_K^0 \mathbb{E}\rho_t e^{-\gamma t}$.

3.3 Step three : Convergence

Let Γ be a coupling of ν_0^1 and ν_0^2 such that $\mathbb{E}_\Gamma \rho((x, v), (\tilde{x}, \tilde{v})) < \infty$. We consider the coupling of (X_t, V_t) and $(\tilde{X}_t, \tilde{V}_t)$, with initial distribution $((X_0, V_0), (\tilde{X}_0, \tilde{V}_0)) \sim \Gamma$, previously introduced. Using Lemma 11 and Lemma 12, by taking the expectation in (25) at stopping times τ_n increasingly converging to t , we have by Fatou's lemma for $n \rightarrow \infty$, $\forall \xi > 0, \forall t \geq 0$,

$$\begin{aligned} e^{ct} \mathbb{E}\rho_t &\leq \mathbb{E}\rho_0 + (1 + \alpha) \xi \int_0^t e^{cs} \mathbb{E}(G_s) ds \\ &\quad + L_W \mathcal{C}_K \int_0^t e^{cs} \mathbb{E}\rho_s ds + L_W \mathcal{C}_K^0 \int_0^t e^{(c-\gamma)s} \mathbb{E}\rho_s ds. \quad (32) \end{aligned}$$

Moreover, using Lemma 4 and the fact $\gamma > c$, for all $t \geq 0$,

$$\begin{aligned} \mathbb{E}(G_t) &\leq (1 + \epsilon \mathcal{C}_H^0 + \epsilon \mathcal{C}_{\tilde{H}}^0), \quad \int_0^t e^{(c-\gamma)s} \mathbb{E}\rho_s ds \leq \frac{f(R_1) (1 + \epsilon \mathcal{C}_H^0 + \epsilon \mathcal{C}_{\tilde{H}}^0)}{\gamma - c}, \\ \int_0^t e^{cs} ds &= \frac{c}{c - L_W \mathcal{C}_K} \frac{e^{ct} - 1}{c} - \frac{L_W \mathcal{C}_K}{c - L_W \mathcal{C}_K} \int_0^t e^{cs} ds, \end{aligned}$$

we get

$$\begin{aligned} &e^{ct} \left(\mathbb{E}\rho_t - \frac{(1 + \alpha) \xi}{c - L_W \mathcal{C}_K} (1 + \epsilon \mathcal{C}_H^0 + \epsilon \mathcal{C}_{\tilde{H}}^0) \right) \\ &\leq \mathbb{E}\rho_0 - \frac{(1 + \alpha) \xi}{c - L_W \mathcal{C}_K} (1 + \epsilon \mathcal{C}_H^0 + \epsilon \mathcal{C}_{\tilde{H}}^0) + L_W \mathcal{C}_K^0 \frac{f(R_1) (1 + \epsilon \mathcal{C}_H^0 + \epsilon \mathcal{C}_{\tilde{H}}^0)}{\gamma - c} \\ &\quad + L_W \mathcal{C}_K \int_0^t e^{cs} \left(\mathbb{E}\rho_s - \frac{(1 + \alpha) \xi}{c - L_W \mathcal{C}_K} (1 + \epsilon \mathcal{C}_H^0 + \epsilon \mathcal{C}_{\tilde{H}}^0) \right) ds. \end{aligned}$$

Gronwall's lemma yields, for all $t \geq 0$

$$e^{ct} \left(\mathbb{E}(\rho_t) - \frac{(1 + \alpha) \xi}{c - L_W \mathcal{C}_K} (1 + \epsilon \mathcal{C}_H^0 + \epsilon \mathcal{C}_{\tilde{H}}^0) \right) \leq \left(\mathbb{E}(\rho_0) + L_W \mathcal{C}_K^0 \frac{f(R_1) (1 + \epsilon \mathcal{C}_H^0 + \epsilon \mathcal{C}_{\tilde{H}}^0)}{\gamma - c} \right) e^{L_W \mathcal{C}_K t}.$$

Since $\mathcal{W}_\rho(\mu_t, \nu_t) \leq \mathbb{E}(\rho_t)$, we have thus obtained for all $t \geq 0$

$$\mathcal{W}_\rho(\nu_t^1, \nu_t^2) \leq \frac{(1+\alpha)\xi}{c - L_W \mathcal{C}_K} (1 + \epsilon \mathcal{C}_H^0 + \epsilon \mathcal{C}_{\tilde{H}}^0) + \left(\mathbb{E}(\rho_0) + L_W \mathcal{C}_K^0 \frac{f(R_1) (1 + \epsilon \mathcal{C}_H^0 + \epsilon \mathcal{C}_{\tilde{H}}^0)}{\gamma - c} \right) e^{(L_W \mathcal{C}_K - c)t}$$

Taking the infimum over all couplings Γ of the initial conditions and using the fact that the left hand side does not depend on ξ , so that we may take $\xi = 0$, we get finally that for all $t \geq 0$,

$$\mathcal{W}_\rho(\nu_t^1, \nu_t^2) \leq \left(\mathcal{W}_\rho(\nu_0^1, \nu_0^2) + L_W \mathcal{C}_K^0 \frac{f(R_1) (1 + \epsilon \mathcal{C}_H^0 + \epsilon \mathcal{C}_{\tilde{H}}^0)}{\gamma - c} \right) e^{(L_W \mathcal{C}_K - c)t}, \quad (33)$$

and since, by Lemma 8, $\mathcal{C}_1 \mathcal{W}_\rho(\nu_t^1, \nu_t^2) \geq \mathcal{W}_1(\nu_t^1, \nu_t^2)$ and $\mathcal{C}_2 \mathcal{W}_\rho(\nu_t^1, \nu_t^2) \geq \mathcal{W}_2^2(\nu_t^1, \nu_t^2)$,

$$\begin{aligned} \mathcal{W}_1(\nu_t^1, \nu_t^2) &\leq e^{-(c - L_W \mathcal{C}_K)t} \mathcal{C}_1^1 \nu_0^1, \nu_0^2, \\ \mathcal{W}_2^2(\nu_t^1, \nu_t^2) &\leq e^{-(c - L_W \mathcal{C}_K)t} \mathcal{C}_2^2 \nu_0^1, \nu_0^2. \end{aligned}$$

Then, for all W such that $L_W < c/\mathcal{C}_K$, there will be contraction at rate $\tau := c - L_W \mathcal{C}_K > 0$. So, it only remains for L_W to satisfy

$$L_W \leq \frac{c}{\mathcal{C}_1 \left(1 + \frac{2\epsilon B}{\gamma} \right) + \frac{2\epsilon B}{\gamma \lambda} (6 + 8\lambda)}, \quad (34)$$

with

$$\mathcal{C}_1 = \max \left(\frac{2}{\alpha}, 1 \right) \max \left(\frac{4 \left((1+\alpha)^2 + \alpha^2 \right)}{\epsilon \min \left(\frac{2}{3} \lambda, 6 \right) f(1)}, \frac{1}{\phi(R_1) g(R_1)} \right).$$

Remark 2 We draw the reader's attention to the fact that Theorem 3 is then a consequence of everything we have done so far : if we have an upper bound on $\mathbb{E}H(X_0, V_0) + \mathbb{E}H(\tilde{X}_0, \tilde{V}_0)$, the constant \mathcal{C}_K^0 in Lemma 12 can be chosen equal to 0 provided we modify \mathcal{C}_K .

Let us now show that there is existence and uniqueness of a stationary measure. Let $\mathcal{C}^0 > \frac{B}{\gamma}$ and μ_t a solution of (1) such that $\mathbb{E}_{\mu_0} H \leq \mathcal{C}^0$. Using (12), for all $t \geq 0$, $\mathbb{E}_{\mu_t} H \leq \mathcal{C}^0$. Thanks to Theorem 3, for L_W sufficiently small, there is $\tau > 0$ such that for all $t \geq s \geq 0$

$$\mathcal{W}_\rho(\mu_t, \mu_s) \leq e^{-\tau s} \mathcal{W}_\rho(\mu_{t-s}, \mu_0) \leq f(R_1) (1 + 2\epsilon \mathcal{C}^0) e^{-\tau s},$$

and thus

$$\mathcal{W}_1(\mu_t, \mu_s) \leq \mathcal{C}_1 f(R_1) (1 + 2\epsilon \mathcal{C}^0) e^{-\tau s}.$$

The space of probability measure with first moments, equipped with the \mathcal{W}_1 distance, being a complete metric space (see for instance [4]), and μ_t being a Cauchy sequence, there exists μ_∞ such that

$$\mathcal{W}_1(\mu_t, \mu_\infty) \rightarrow 0 \text{ as } t \rightarrow \infty,$$

and μ_∞ stationary. Theorem 1 then ensures uniqueness and convergence towards this stationary measure.

4 Proof of Theorem 2

In this section, we show how we obtain similar results for the convergence of the particle system to the non-linear kinetic Langevin diffusion using the same tools. We start by introducing the coupling, the new Lyapunov function, we give a new definition for the various quantities we consider, and then prove contraction of the coupling semimetric.

4.1 Coupling

We consider the following coupling

$$\begin{cases} d\bar{X}_t^i = \bar{V}_t^i dt \\ d\bar{V}_t^i = -\bar{V}_t^i dt - \nabla U(\bar{X}_t^i) dt - \nabla W * \bar{\mu}_t(\bar{X}_t^i) dt + \sqrt{2} \left(rc(Z_t^i, W_t^i) dB_t^{rc,i} + sc(Z_t^i, W_t^i) dB_t^{sc,i} \right) \\ \bar{\mu}_t = \mathcal{L}(\bar{X}_t^i) \\ dX_t^{i,N} = V_t^{i,N} dt \\ dV_t^{i,N} = -V_t^{i,N} dt - \nabla U(X_t^{i,N}) dt - \frac{1}{N} \sum_{j=1}^N \nabla W(X_t^{i,N} - X_t^{j,N}) dt \\ \quad + \sqrt{2} \left(rc(Z_t^i, W_t^i) (Id - 2e_t^i e_t^{i,T}) dB_t^{rc,i} + sc(Z_t^i, W_t^i) dB_t^{sc,i} \right), \end{cases}$$

with, similarly as before,

$$Z_t^i = \bar{X}_t^i - X_t^{i,N}, \quad W_t^i = \bar{V}_t^i - V_t^{i,N}, \quad Q_t^i = Z_t^i + W_t^i, \quad e_t^i = \begin{cases} \frac{Q_t^i}{|Q_t^i|} & \text{if } Q_t^i \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Let $\mu_t^N := \frac{1}{N} \sum_{i=1}^N \delta_{X_t^{i,N}}$ be the empirical distribution of the particle system, with i.i.d initial conditions $X_0^{i,N} \sim \nu_0$. We first notice that the particles are exchangeable. The generator of the process given by the particle system (3) is, for a function ϕ of $(x_1, \dots, x_N, v_1, \dots, v_N)$

$$\mathcal{L}^N \phi = \sum_{i=1}^N \mathcal{L}^{i,N} \phi,$$

with

$$\mathcal{L}^{i,N} \phi = v_i \cdot \nabla_{x_i} \phi - v_i \cdot \nabla_{v_i} \phi - \nabla U(x_i) \cdot \nabla_{v_i} \phi - \frac{1}{N} \sum_{j=1}^N \nabla W(x_i - x_j) \cdot \nabla_{v_i} \phi + \Delta_{v_i} \phi.$$

We define

$$r_t^i = \alpha |Z_t^i| + |Q_t^i|, \quad (35)$$

$$\tilde{H}(x, v) = \int_0^{H(x, v)} \exp(a\sqrt{u}) du \quad (36)$$

$$\begin{aligned} \rho_t = \frac{1}{N} \sum_{i=1}^N f(r_t^i) & \left(1 + \epsilon \tilde{H}(\bar{X}_t^i, \bar{V}_t^i) + \epsilon \tilde{H}(X_t^{i, N}, V_t^{i, N}) \right. \\ & \left. + \frac{\epsilon}{N} \sum_{j=1}^N \tilde{H}(\bar{X}_t^j, \bar{V}_t^j) + \frac{\epsilon}{N} \sum_{j=1}^N \tilde{H}(X_t^{j, N}, V_t^{j, N}) \right), \end{aligned} \quad (37)$$

$$:= \frac{1}{N} \sum_{i=1}^N f(r_t^i) G_t^i. \quad (38)$$

4.2 A modified Lyapunov function

Notice how in the expression of G^i above we did not consider the Lyapunov function H , but instead \tilde{H} . Let us assume there exist $\mathcal{C}_0, a > 0$ such that $\mathbb{E}_{\nu_0}(\tilde{H}(X, V)^2) \leq (\mathcal{C}_0)^2$ (which is equivalent to the existence of $\tilde{\mathcal{C}}^0, \tilde{a} > 0$ such that $\mathbb{E}_{\nu_0}(e^{\tilde{a}(|X|+|V|)}) \leq \tilde{\mathcal{C}}^0$, as it was stated in Theorem 2). First, notice

$$\tilde{H}(x, v) = \int_0^{H(x, v)} \exp(a\sqrt{u}) du = \frac{2}{a^2} \exp(a\sqrt{H(x, v)}) \left(a\sqrt{H(x, v)} - 1 \right) + \frac{2}{a^2}.$$

Direct calculations yield the following technical lemma.

Lemma 13 *We have, for all $x, v \in \mathbb{R}^d$*

$$H(x, v) \exp(a\sqrt{H(x, v)}) \geq \tilde{H}(x, v) \geq \exp(a\sqrt{H(x, v)}) - \frac{2}{a^2} \left(\exp\left(\frac{a^2}{2}\right) - 1 \right), \quad (39)$$

$$\frac{2}{a} \sqrt{H(x, v)} \exp(a\sqrt{H(x, v)}) \geq \tilde{H}(x, v) \geq \frac{1}{a} \sqrt{H(x, v)} \exp(a\sqrt{H(x, v)}) - \frac{1}{a^2} (e - 2), \quad (40)$$

$$\tilde{H}(x, v) \geq H(x, v) \quad (41)$$

We may calculate, using (8) and (9)

$$\begin{aligned}
\mathcal{L}_\mu(\tilde{H}) &= \exp(a\sqrt{H}) \mathcal{L}_\mu H + \frac{a}{2\sqrt{H}} \exp(a\sqrt{H}) |\nabla_v H|^2 \\
&= \exp(a\sqrt{H}) \mathcal{L}_\mu H + 24^2 \frac{a}{2\sqrt{H}} \exp(a\sqrt{H}) \left| \frac{x}{2} + v \right|^2 \\
&\leq \exp(a\sqrt{H}) \left(B + L_W (6 + 8\lambda) \mathbb{E}_\mu(|x|)^2 - \left(\frac{3}{4}\lambda + \lambda^2 \right) |x|^2 - \gamma H \right) + 24a\sqrt{H} \exp(a\sqrt{H}) \\
&\leq \exp(a\sqrt{H}) \left(B + \frac{288a^2}{\gamma} + L_W (6 + 8\lambda) \mathbb{E}_\mu(|x|)^2 - \frac{\gamma}{2} H \right), \quad (42)
\end{aligned}$$

where for this last inequality we used Young's inequality $24a\sqrt{H} \leq \frac{\gamma}{2}H + 288\frac{a^2}{\gamma}$.

Notice that (41) ensures that this new Lyapunov function also tends to bring back particle which ventured at infinity, and at an even greater rate. This new rate $H \exp(\sqrt{H})$ however comes at a cost : the initial condition must have a finite exponential moment, and not just a finite second moment as in Section 3.

First, by (12) and (40),

$$\mathbb{E}(|\bar{X}_t^i|)^2 \leq \frac{1}{\lambda} \mathbb{E}(H(\bar{X}_t^i, \bar{V}_t^i)) \leq \frac{1}{\lambda} \left(\frac{B}{\gamma} + \mathbb{E}H(\bar{X}_0^i, \bar{V}_0^i) \right) \leq \frac{1}{\lambda} \left(\frac{B}{\gamma} + \mathbb{E}\tilde{H}(\bar{X}_0^i, \bar{V}_0^i) \right) \leq \frac{1}{\lambda} \left(\frac{B}{\gamma} + C^0 \right).$$

Furthermore, the function $h \mapsto \exp(a\sqrt{h}) \left(\tilde{B} - \frac{\gamma}{4}h \right)$ is bounded from above for $h \geq 0$ and $\tilde{B} \in \mathbb{R}$. We therefore obtain from (41) the existence of \tilde{B} such that

$$\mathcal{L}_{\mu_t^{\otimes N}} \tilde{H}(x_i, v_i) \leq \tilde{B} - \frac{\gamma}{4} \left(H(x_i, v_i) \exp(a\sqrt{H(x_i, v_i)}) \right) \quad (43)$$

$$\frac{d}{dt} \mathbb{E} \tilde{H}(\bar{X}_t^i, \bar{V}_t^i) \leq \tilde{B} - \frac{\gamma}{4} \mathbb{E} \left(H(\bar{X}_t^i, \bar{V}_t^i) \exp(a\sqrt{H(\bar{X}_t^i, \bar{V}_t^i)}) \right) \quad (44)$$

$$\text{and } \frac{d}{dt} \mathbb{E} \tilde{H}(\bar{X}_t^i, \bar{V}_t^i) \leq \tilde{B} - \frac{\gamma}{4} \mathbb{E} \tilde{H}(\bar{X}_t^i, \bar{V}_t^i), \quad (45)$$

where for this last inequality, we used (38). While (42) and (43) will be useful in ensuring a sufficient restoring force, Equation (44) give us a uniform in time bound on $\mathbb{E} \tilde{H}(\bar{X}_t^i, \bar{V}_t^i)$, provided we have an initial bound.

Now, for the system of particle, we have, using (41), $\forall i \in \{1, \dots, N\}$, $\forall x_i, v_i \in \mathbb{R}^d$,

$$\mathcal{L}^{i,N} \tilde{H}(x_i, v_i) \leq \exp(a\sqrt{H(x_i, v_i)}) \left(B + \frac{288a^2}{\gamma} + L_W (6 + 8\lambda) \left(\frac{\sum_{j=1}^N |x_j|}{N} \right)^2 - \frac{\gamma}{2} H(x_i, v_i) \right).$$

Summing over $i \in \{1, \dots, N\}$, we may calculate

$$\begin{aligned} & L_W (6 + 8\lambda) \sum_{j=1}^N \left(\frac{\sum_{j=1}^N |x_j|}{N} \right)^2 \sum_{i=1}^N \frac{\exp \left(a\sqrt{H(x_i, v_i)} \right)}{N} - \frac{\gamma}{8} \sum_{i=1}^N \frac{H(x_i, v_i) \exp \left(a\sqrt{H(x_i, v_i)} \right)}{N} \\ & \leq \frac{\gamma}{8} \left(\sum_{i,j=1}^N \frac{H(x_i, v_i) \exp \left(a\sqrt{H(x_j, v_j)} \right)}{N} - \sum_{i=1}^N \frac{H(x_i, v_i) \exp \left(a\sqrt{H(x_i, v_i)} \right)}{N} \right) \leq 0. \end{aligned} \quad (46)$$

Here, we used (8), the fact that $\forall x, y \geq 0 \quad xe^{a\sqrt{y}} + ye^{a\sqrt{x}} - xe^{a\sqrt{x}} - ye^{a\sqrt{y}} = (e^{a\sqrt{x}} - e^{a\sqrt{y}})(y - x) \leq 0$ and assumed

$$6 \frac{L_W}{\lambda} \left(1 + \frac{4}{3} \lambda \right) \leq \frac{\gamma}{8} \quad \text{i.e.} \quad L_W \leq \frac{\gamma \lambda}{16(3 + 4\lambda)}.$$

Likewise, there is a constant, which for the sake of clarity we will also denote \tilde{B} (as we may take the maximum of the previous constants), such that we get

$$\begin{aligned} \mathcal{L}^{i,N} \tilde{H}(x_i, v_i) & \leq \tilde{B} + L_W (6 + 8\lambda) \left(\frac{\sum_{j=1}^N |x_j|}{N} \right)^2 \exp \left(a\sqrt{H(x_i, v_i)} \right) \\ & \quad - \frac{\gamma}{4} H(x_i, v_i) \exp \left(a\sqrt{H(x_i, v_i)} \right) \end{aligned} \quad (47)$$

$$\mathcal{L}^N \left(\frac{1}{N} \sum_{i=1}^N \tilde{H}(x_i, v_i) \right) \leq \tilde{B} - \frac{\gamma}{4} \left(\frac{1}{N} \sum_{i=1}^N H(x_i, v_i) \exp \left(a\sqrt{H(x_i, v_i)} \right) \right) \quad (48)$$

and

$$\mathcal{L}^N \left(\frac{1}{N} \sum_{i=1}^N \tilde{H}(x_i, v_i) \right) \leq \tilde{B} - \frac{\gamma}{4} \left(\frac{1}{N} \sum_{i=1}^N \tilde{H}(x_i, v_i) \right) \quad (49)$$

Once again, (46) and (47) will ensure a sufficient restoring force, and (48) ensures a uniform in time bound on the expectation of $\tilde{H}(X_t^{i,N}, V_t^{i,N})$, since $\mathbb{E} \left(\frac{1}{N} \sum_{j=1}^N \tilde{H}(X_t^{j,N}, V_t^{j,N}) \right) = \mathbb{E} \left(\tilde{H}(X_t^{i,N}, V_t^{i,N}) \right)$ by exchangeability of the particles.

More precisely, we obtain from (44) and (48) the direct corollary

Lemma 14 *Provided the initial expectations $\mathbb{E}(G_0^1)$ and $\mathbb{E}((G_0^1)^2)$ are finite, there are constants $\mathcal{C}_{G,1}$ and $\mathcal{C}_{G,2}$, depending on initial conditions, such that for all $t \geq 0$, for all $N \geq 0$, and all i*

$$\mathbb{E}(G_t^i) \leq \mathcal{C}_{G,1} \quad \text{and} \quad \mathbb{E}((G_t^i)^2) \leq \mathcal{C}_{G,2}.$$

Finally, since $\tilde{H}(x, v) \geq H(x, v)$, Lemma 8 still holds for our new semimetric.

4.3 New parameters

For the sake of completeness, and since this is similar to Section 2.2, we quickly give some explicit parameters that satisfy the various conditions arising from calculation. These parameters are far from optimal, and are just given to show that every constant is explicit. Let \tilde{B} be given by (42)-(44), and (46)-(48). Define

$$\alpha = L_U + \frac{\lambda}{4}, \quad R_0 = \sqrt{\frac{160\tilde{B}}{\gamma \min(\frac{\lambda}{3}, 3)}} \quad \text{and} \quad R_1 = \sqrt{(1 + \alpha)^2 + \alpha^2} R_0.$$

Recall the definition of $C_{dH,1}$ and $C_{dH,2}$ in (9). Denoting

$$C_{f,1} = 8 \left(\left(\frac{96}{a^2} \max \left(1, \frac{1}{2\alpha} \right) + \frac{16\sqrt{3}}{a} C_{dH,1} \right) \left(\exp \left(\frac{a^2}{2} \right) - 1 \right) + 16\sqrt{3}(e-2)C_{dH,2} \right)$$

$$C_{f,2} = 8 \left(24 \max \left(1, \frac{1}{2\alpha} \right) + 4\sqrt{3}C_{dH,1}a + 8\sqrt{3}C_{dH,2}a^2 \right)$$

we set

$$c = \left\{ \frac{1}{12} \min \left(2\sqrt{\frac{L_U + L_W}{2\pi\alpha R_1^2}}, \frac{1}{2} \left(1 - \frac{L_U + L_W}{\alpha} \right) \right) \exp \left(-\frac{1}{8} \left(\frac{L_U + L_W}{\alpha} + \alpha + C_{f,1} + C_{f,2} \right) R_1^2 \right), \right. \\ \left. \frac{2\tilde{B}}{5}, \frac{\gamma}{800} \right\},$$

and $\epsilon = \frac{5c}{2\tilde{B}}$. For $s \geq 0$,

$$\phi(s) = \exp \left(-\frac{1}{8} \left(\frac{1}{\alpha} (L_U + L_W) + \alpha + \epsilon C_{f,1} + C_{f,2} \right) s^2 \right), \quad \Phi(s) = \int_0^s \phi(u) du$$

$$g(s) = 1 - \frac{c + 2\epsilon\tilde{B}}{2} \int_0^s \frac{\Phi(u)}{\phi(u)} du, \quad f(s) = \int_0^{\min(s, R_1)} \phi(u) g(u) du.$$

This way we satisfy the following conditions

$$c \leq \frac{\gamma}{160} \left(1 - \frac{\gamma}{80\epsilon\tilde{B} + \gamma} \right)$$

$$\alpha > L_U + L_W$$

$$\epsilon \leq 1$$

$$2c + 4\epsilon\tilde{B} \leq 2 \left(\int_0^{R_1} \frac{\Phi(u)}{\phi(u)} du \right)^{-1}$$

$$2c + 4\epsilon\tilde{B} \leq \frac{1}{2} \left(1 - \frac{L_U + L_W}{\alpha} \right) \inf_{r \in [0, R_1]} \frac{r\phi(r)}{\Phi(r)}$$

$$\forall s \geq 0, 0 = 4\phi'(s) + \left(\frac{1}{\alpha} (L_U + L_W) + \alpha + \epsilon C_{f,1} + C_{f,2} \right) s\phi(s)$$

4.4 Convergence

The goal of the section is to prove the following result

Theorem 4 *Let $U \in \mathcal{C}^1(\mathbb{R}^d)$ satisfy Assumption 1 and Assumption 2. For all $W \in \mathcal{C}^1(\mathbb{R}^d)$ satisfying Assumption 3 with*

$$L_W \leq \min \left(\frac{\gamma\lambda}{16(3+4\lambda)}, \frac{c}{\mathcal{C}_1}, \frac{\gamma}{64\mathcal{C}_z}, \frac{\gamma a}{256\mathcal{C}_z\epsilon} \right), \quad (50)$$

and for all probability measures $\bar{\nu}_0$ on \mathbb{R}^{2d} such that $\mathbb{E}_{\bar{\nu}_0} \tilde{H}^2(X, V) \leq (\mathcal{C}^0)^2$, for all N , $\xi > 0$, and $t \geq 0$,

$$e^{ct} \mathbb{E}(\rho_t) \leq \mathbb{E}(\rho_0) + \xi(1+\alpha)\mathcal{C}_{G,1} \int_0^t e^{cs} ds + L_W \frac{\mathcal{C}^0 \mathcal{C}_{G,2}^{1/2}}{\lambda} \sqrt{\frac{8}{N}} \int_0^t e^{cs} ds.$$

4.4.1 Proof of Theorem 2 using Theorem 4

We first show how Theorem 2 is a consequence of Theorem 4. Let Γ be a coupling of $\nu_0^{\otimes N}$ and $\bar{\nu}_0^{\otimes N}$, such that $\mathbb{E}\rho_0 < \infty$. We consider the coupling previously introduced. For clarity, let us denote

$$\frac{A}{\sqrt{N}} = L_W \frac{\mathcal{C}^0 \mathcal{C}_{G,2}^{1/2}}{\lambda} \sqrt{\frac{8}{N}}, \quad B = (1+\alpha)\mathcal{C}_{G,1},$$

i.e

$$e^{ct} \mathbb{E}(\rho_t) \leq \mathbb{E}(\rho_0) + \xi B \int_0^t e^{cs} ds + \frac{A}{\sqrt{N}} \int_0^t e^{cs} ds.$$

Let us consider

$$u(t) = e^{ct} \left(\mathbb{E}(\rho_t) - \frac{A}{c} \frac{1}{\sqrt{N}} - \xi \frac{B}{c} \right)$$

Then $u(t) \leq u(0)$ i.e

$$\mathbb{E}(\rho_t) \leq \mathbb{E}(\rho_0) e^{-ct} + \frac{A}{c} \frac{1}{\sqrt{N}} (1 - e^{-ct}) + \xi \frac{B}{c} (1 - e^{-ct}).$$

We thus obtain the desired result, by taking the limit as $\xi \rightarrow 0$ uniformly in time, and by using the exchangeability of the particles to have $\mathbb{E}(\rho_t) = \mathbb{E}\left(\frac{1}{N} \sum_{i=1}^N \rho_t^i\right) = \mathbb{E}\left(\frac{1}{k} \sum_{i=1}^k \rho_t^i\right)$ for all $k \in \mathbb{N}$.

4.4.2 Evolution of the coupling semimetric for the particle system

We thus need to start by considering the dynamic of ρ_t . Like in Lemma 10, we have almost surely for all $t \geq 0$

$$d|Q_t^i| = -e_t^{i,T} \left(\nabla U(\bar{X}_t^i) - \nabla U(X_t^{i,N}) \right) dt - e_t^{i,T} \left(\nabla W * \bar{\mu}_t(\bar{X}_t^i) - \nabla W * \bar{\mu}_t^N(X_t^{i,N}) \right) dt \\ + 2\sqrt{2}rc(Z_t^i, W_t^i) e_t^{i,T} dB_t^{rc,i}.$$

Hence $e^{ct}f(r_t^i) = f(r_0) + \hat{A}_t^i + \hat{M}_t^i$ with

$$d\hat{A}_t^i = \left(cf(r_t^i) + \alpha f'(r_t^i) \frac{d|Z_t^i|}{dt} - f'(r_t^i) e_t^{i,T} \left(\nabla U(X_t^i) - \nabla U(X_t^{i,N}) \right) \right. \\ \left. - f'(r_t^i) e_t^{i,T} \left(\nabla W * \mu_t(X_t^i) - \frac{1}{N} \sum_{j=1}^N \nabla W(X_t^{i,N} - X_t^{j,N}) \right) + 4f''(r_t^i) rc^2(Z_t^i, W_t^i) \right) e^{ct} dt, \\ d\hat{M}_t^i = e^{ct} 2\sqrt{2}f'(r_t^i) rc(Z_t^i, W_t^i) e_t^{i,T} dB_t^{rc,i}.$$

We now consider the evolution of

$$G_t^i = 1 + \epsilon \tilde{H}(\bar{X}_t^i, \bar{V}_t^i) + \epsilon \tilde{H}(X_t^{i,N}, V_t^{i,N}) + \frac{\epsilon}{N} \sum_{j=1}^N \tilde{H}(\bar{X}_t^j, \bar{V}_t^j) + \frac{\epsilon}{N} \sum_{j=1}^N \tilde{H}(X_t^{j,N}, V_t^{j,N}).$$

Notice how we have added new terms in G_t^i . Those additional quantities will help us in dealing with the non linearity, as will be shown later.

$$dG_t^i = \epsilon \left(\mathcal{L}_{\bar{\mu}_t^{\otimes N}} \tilde{H}(\bar{X}_t^i, \bar{V}_t^i) + \mathcal{L}^N \tilde{H}(X_t^{i,N}, V_t^{i,N}) \right) dt \\ + \epsilon \sqrt{2}rc(Z_t^i, W_t^i) \left(\nabla_v \tilde{H}(\bar{X}_t^i, \bar{V}_t^i) - \nabla_v \tilde{H}(X_t^{i,N}, V_t^{i,N}) \right) \cdot e_t^i e_t^{i,T} dB_t^{rc,i} \\ + \epsilon \sqrt{2}rc(Z_t^i, W_t^i) \left(\nabla_v \tilde{H}(\bar{X}_t^i, \bar{V}_t^i) + \nabla_v \tilde{H}(X_t^{i,N}, V_t^{i,N}) \right) \cdot (Id - e_t^i e_t^{i,T}) dB_t^{rc,i} \\ + \epsilon \sqrt{2}sc(Z_t^i, W_t^i) \left(\nabla_v \tilde{H}(\bar{X}_t^i, \bar{V}_t^i) + \nabla_v \tilde{H}(X_t^{i,N}, V_t^{i,N}) \right) \cdot dB_t^{sc,i} \\ + \frac{\epsilon}{N} \sum_{j=1}^N \left(\mathcal{L}_{\bar{\mu}_t^{\otimes N}} \tilde{H}(\bar{X}_t^j, \bar{V}_t^j) + \mathcal{L}^N \tilde{H}(X_t^{j,N}, V_t^{j,N}) \right) dt \\ + \frac{\epsilon \sqrt{2}}{N} \sum_{j=1}^N rc(Z_t^j, W_t^j) \left(\nabla_v \tilde{H}(\bar{X}_t^j, \bar{V}_t^j) - \nabla_v \tilde{H}(X_t^{j,N}, V_t^{j,N}) \right) \cdot e_t^j e_t^{j,T} dB_t^{rc,j} \\ + \frac{\epsilon \sqrt{2}}{N} \sum_{j=1}^N rc(Z_t^j, W_t^j) \left(\nabla_v \tilde{H}(\bar{X}_t^j, \bar{V}_t^j) + \nabla_v \tilde{H}(X_t^{j,N}, V_t^{j,N}) \right) \cdot (Id - e_t^j e_t^{j,T}) dB_t^{rc,j} \\ + \frac{\epsilon \sqrt{2}}{N} \sum_{j=1}^N sc(Z_t^j, W_t^j) \left(\nabla_v \tilde{H}(\bar{X}_t^j, \bar{V}_t^j) + \nabla_v \tilde{H}(X_t^{j,N}, V_t^{j,N}) \right) \cdot dB_t^{sc,j}.$$

Therefore

$$e^{ct}\rho_t^i = e^{ct}f(r_t^i) G_t^i = \rho_0 + A_t^i + M_t^i, \quad (51)$$

with

$$\begin{aligned} dA_t^i = & G_t^i d\hat{A}_t^i + \epsilon e^{ct} f(r_t^i) \left(\mathcal{L}_{\bar{\mu}_t^{\otimes N}} \tilde{H}(\bar{X}_t^i, \bar{V}_t^i) + \mathcal{L}^N \tilde{H}(X_t^{i,N}, V_t^{i,N}) \right. \\ & + \frac{1}{N} \sum_{j=1}^N \mathcal{L}_{\bar{\mu}_t^{\otimes N}} \tilde{H}(\bar{X}_t^j, \bar{V}_t^j) + \frac{1}{N} \mathcal{L}^N \sum_{j=1}^N \tilde{H}(X_t^{j,N}, V_t^{j,N}) \Big) dt \\ & + 4\epsilon \left(1 + \frac{1}{N} \right) e^{ct} f'(r_t^i) r_t^i c^2 (Z_t^i, W_t^i) \left(\nabla_v \tilde{H}(\bar{X}_t^i, \bar{V}_t^i) - \nabla_v \tilde{H}(X_t^{i,N}, V_t^{i,N}) \right) \cdot e_t^i dt \end{aligned}$$

and M_t^i is a continuous local martingale. Let us deal with this last line. For the sake of conciseness, from now on we denote for all i

$$\bar{H}_i := H(\bar{X}_t^i, \bar{V}_t^i), \quad \text{and} \quad H_i^N := H(X_t^{i,N}, V_t^{i,N})$$

We have

$$\begin{aligned} & |\nabla_v \tilde{H}(\bar{X}_t^i, \bar{V}_t^i) - \nabla_v \tilde{H}(X_t^{i,N}, V_t^{i,N})| \\ &= \left| \nabla_v \bar{H}_i \exp(a\sqrt{\bar{H}_i}) - \nabla_v H_i^N \exp(a\sqrt{H_i^N}) \right| \\ &\leq \left| 12\bar{X}_t^i + 24\bar{V}_t^i - 12X_t^{i,N} - 24V_t^{i,N} \right| \left(\exp(a\sqrt{\bar{H}_i}) + \exp(a\sqrt{H_i^N}) \right) \\ &\quad + a \left| 12\bar{X}_t^i + 24\bar{V}_t^i \right| \left| \sqrt{\bar{H}_i} - \sqrt{H_i^N} \right| \left(\exp(a\sqrt{\bar{H}_i}) + \exp(a\sqrt{H_i^N}) \right) \\ &\leq 24 \max \left(1, \frac{1}{2\alpha} \right) r_t^i \left(\exp(a\sqrt{\bar{H}_i}) + \exp(a\sqrt{H_i^N}) \right) \\ &\quad + 4a\sqrt{3} \left| \bar{H}_i - H_i^N \right| \left(\exp(a\sqrt{\bar{H}_i}) + \exp(a\sqrt{H_i^N}) \right). \end{aligned}$$

Now, using Lemma 9, we get

$$\begin{aligned} & |\nabla_v \tilde{H}(\bar{X}_t^i, \bar{V}_t^i) - \nabla_v \tilde{H}(X_t^{i,N}, V_t^{i,N})| \\ &\leq \left(24 \max \left(1, \frac{1}{2\alpha} \right) + 4\sqrt{3}C_{dH,1}a \right) r_t^i \left(\exp(a\sqrt{\bar{H}_i}) + \exp(a\sqrt{H_i^N}) \right) \\ &\quad + 4\sqrt{3}C_{dH,2}ar_t^i \left(\sqrt{\bar{H}_i} + \sqrt{H_i^N} \right) \left(\exp(a\sqrt{\bar{H}_i}) + \exp(a\sqrt{H_i^N}) \right) \\ &\leq \left(24 \max \left(1, \frac{1}{2\alpha} \right) + 4\sqrt{3}C_{dH,1}a \right) r_t^i \left(\exp(a\sqrt{\bar{H}_i}) + \exp(a\sqrt{H_i^N}) \right) \\ &\quad + 8\sqrt{3}C_{dH,2}ar_t^i \left(\sqrt{\bar{H}_i} \exp(a\sqrt{\bar{H}_i}) + \sqrt{H_i^N} \exp(a\sqrt{H_i^N}) \right). \end{aligned}$$

Hence why, using (38) and (39), we get

$$\begin{aligned} & |\nabla_v \tilde{H}(\bar{X}_t^i, \bar{V}_t^i) - \nabla_v \tilde{H}(X_t^{i,N}, V_t^{i,N})| \\ &\leq \left(24 \max \left(1, \frac{1}{2\alpha} \right) + 4\sqrt{3}C_{dH,1}a \right) r_t^i \left(\frac{4}{a^2} \left(\exp\left(\frac{a^2}{2}\right) - 1 \right) + \tilde{H}(\bar{X}_t^i, \bar{V}_t^i) + \tilde{H}(X_t^{i,N}, V_t^{i,N}) \right) \\ &\quad + 8\sqrt{3}C_{dH,2}a^2r_t^i \left(\frac{2}{a^2}(e-2) + \tilde{H}(\bar{X}_t^i, \bar{V}_t^i) + \tilde{H}(X_t^{i,N}, V_t^{i,N}) \right) \end{aligned}$$

and thus

$$\begin{aligned}
& 4\epsilon \left(1 + \frac{1}{N}\right) e^{ct} f'(r_t^i) r c^2 (Z_t^i, W_t^i) \left(\nabla_v \tilde{H}(\bar{X}_t^i, \bar{V}_t^i) - \nabla_v \tilde{H}(X_t^{i,N}, V_t^{i,N}) \right) \cdot e_t^i dt \\
& \leq 8\epsilon r_t^i f'(r_t^i) e^{ct} r c^2 (Z_t^i, W_t^i) \left(\left(\frac{96}{a^2} \max\left(1, \frac{1}{2\alpha}\right) + \frac{16\sqrt{3}}{a} C_{dH,1} \right) \left(\exp\left(\frac{a^2}{2}\right) - 1 \right) + 16\sqrt{3}(e-2)C_{dH,2} \right) \\
& \quad + 8r_t^i f'(r_t^i) e^{ct} r c^2 (Z_t^i, W_t^i) \left(24 \max\left(1, \frac{1}{2\alpha}\right) + 4\sqrt{3}C_{dH,1}a + 8\sqrt{3}C_{dH,2}a^2 \right) \left(\epsilon \tilde{H}(\bar{X}_t^i, \bar{V}_t^i) + \epsilon \tilde{H}(X_t^{i,N}, V_t^{i,N}) \right) \\
& \leq (\epsilon C_{f,1} + C_{f,2}) r_t^i f'(r_t^i) r c^2 (Z_t^i, W_t^i) G_t^i.
\end{aligned}$$

Then we use

$$\begin{aligned}
& \left| \nabla W * \bar{\mu}_t(\bar{X}_t^i) - \frac{1}{N} \sum_{j=1}^N \nabla W(X_t^{i,N} - X_t^{j,N}) \right| \\
& \leq \left| \nabla W * \bar{\mu}_t(\bar{X}_t^i) - \frac{1}{N} \sum_{j=1}^N \nabla W(\bar{X}_t^i - \bar{X}_t^j) \right| + \left| \frac{1}{N} \sum_{j=1}^N \nabla W(\bar{X}_t^i - \bar{X}_t^j) - \frac{1}{N} \sum_{j=1}^N \nabla W(X_t^{i,N} - X_t^{j,N}) \right|, \\
& \leq \left| \nabla W * \bar{\mu}_t(\bar{X}_t^i) - \frac{1}{N} \sum_{j=1}^N \nabla W(\bar{X}_t^i - \bar{X}_t^j) \right| + \frac{1}{N} \sum_{j=1}^N \left| \nabla W(\bar{X}_t^i - \bar{X}_t^j) - \nabla W(X_t^{i,N} - X_t^{j,N}) \right|, \\
& \leq \left| \nabla W * \bar{\mu}_t(\bar{X}_t^i) - \frac{1}{N} \sum_{j=1}^N \nabla W(\bar{X}_t^i - \bar{X}_t^j) \right| + \frac{L_W}{N} \sum_{j=1}^N \left(|\bar{X}_t^i - X_t^{i,N}| + |\bar{X}_t^j - X_t^{j,N}| \right).
\end{aligned}$$

Thus

$$\begin{aligned}
& \left| \nabla W * \bar{\mu}_t(\bar{X}_t^i) - \frac{1}{N} \sum_{j=1}^N \nabla W(X_t^{i,N} - X_t^{j,N}) \right| \\
& \leq \left| \nabla W * \bar{\mu}_t(\bar{X}_t^i) - \frac{1}{N} \sum_{j=1}^N \nabla W(\bar{X}_t^i - \bar{X}_t^j) \right| + L_W |Z_t^i| + L_W \frac{\sum_{j=1}^N |Z_t^j|}{N}.
\end{aligned}$$

And finally we use (42), (46) and (47) to have

$$\begin{aligned}
& \mathcal{L}_{\bar{\mu}_t^{\otimes N}} \tilde{H}(\bar{X}_t^i, \bar{V}_t^i) + \mathcal{L}^N \tilde{H}(X_t^{i,N}, V_t^{i,N}) + \frac{1}{N} \sum_{j=1}^N \mathcal{L}_{\bar{\mu}_t^{\otimes N}} \tilde{H}(\bar{X}_t^j, \bar{V}_t^j) + \frac{1}{N} \mathcal{L}^N \sum_{j=1}^N \tilde{H}(X_t^{j,N}, V_t^{j,N}) \\
& \leq 4\tilde{B} + L_W (6 + 8\lambda) \left(\frac{\sum_{j=1}^N |X_t^{j,N}|}{N} \right)^2 \exp\left(a\sqrt{H_i^N}\right) - \frac{\gamma}{4} \bar{H}_i \exp\left(a\sqrt{\bar{H}_i}\right) - \frac{\gamma}{4} H_i^N \exp\left(a\sqrt{H_i^N}\right) \\
& \quad - \frac{\gamma}{4N} \sum_{j=1}^N \left(\bar{H}_j \exp\left(a\sqrt{\bar{H}_j}\right) + H_j^N \exp\left(a\sqrt{H_j^N}\right) \right).
\end{aligned}$$

We thus obtain

$$dA_t^i \leq e^{ct} K_t^i dt \quad (52)$$

with

$$K_t^i = f'(r_t^i) G_t^i \left(\alpha \frac{d|Z_t^i|}{dt} + (L_U + L_W) |Z_t^i| + (\epsilon C_{f,1} + C_{f,2}) r_t^i r c^2 (Z_t^i, W_t^i) \right) + 2cf(r_t^i) G_t^i \quad (53)$$

$$+ 4f''(r_t^i) G_t^i r c^2 (Z_t^i, W_t^i) + \left| \nabla W * \bar{\mu}_t(\bar{X}_t^i) - \frac{1}{N} \sum_{j=1}^N \nabla W(\bar{X}_t^i - \bar{X}_t^j) \right| f'(r_t^i) G_t^i \quad (54)$$

$$+ \epsilon f(r_t^i) \left(4\tilde{B} - \frac{\gamma}{16} \tilde{H}(\bar{X}_t^i, \bar{V}_t^i) - \frac{\gamma}{16} \tilde{H}(X_t^{i,N}, V_t^{i,N}) - \frac{\gamma}{16N} \sum_{j=1}^N \tilde{H}(\bar{X}_t^j, \bar{V}_t^j) \right. \\ \left. - \frac{\gamma}{16N} \sum_{j=1}^N \tilde{H}(X_t^{j,N}, V_t^{j,N}) \right) \quad (55)$$

$$+ L_W \frac{\sum_{j=1}^N |Z_t^j|}{N} f'(r_t^i) G_t^i - cf(r_t^i) G_t^i - \epsilon f(r_t^i) \left(\frac{\gamma}{16} \bar{H}_i \exp(a\sqrt{\bar{H}_i}) + \frac{\gamma}{16} H_i^N \exp(a\sqrt{H_i^N}) \right) \\ + \frac{\gamma}{16N} \sum_{j=1}^N \bar{H}_j \exp(a\sqrt{\bar{H}_j}) + \frac{\gamma}{16N} \sum_{j=1}^N H_j^N \exp(a\sqrt{H_j^N}) \quad (56)$$

$$+ \epsilon L_W (6 + 8\lambda) f(r_t^i) \left(\frac{\sum_{j=1}^N |X_t^{j,N}|}{N} \right)^2 \exp(a\sqrt{H_i^N}) \\ - \frac{\gamma\epsilon}{8} f(r_t^i) \left(H_i^N \exp(a\sqrt{H_i^N}) + \frac{1}{N} \sum_{j=1}^N H_j^N \exp(a\sqrt{H_j^N}) \right). \quad (57)$$

This formulation of K_t^i might seem cumbersome (and to some degree it is...) but we have actually grouped the various terms based on how we will have them compensate one another. Thus,

- lines (52) and (53) will be managed thanks to the construction of the function f like before, with a special care given to the last term of line (53), on which we will use a law of large number,
- line (54) will come into play when considering the "last region of space" introduced previously,
- line (55) will, under some conditions on L_W , be nonpositive when summing up all $(K_t^j)_j$,
- and finally, line (56) will be nonpositive thanks to Lemma 3, provided L_W is sufficiently small.

This shows an important idea in the construction of the function ρ : we added in G_t^i the empirical mean of $H(X_t^{i,N}, V_t^{i,N}) + H(\bar{X}_t^i, \bar{V}_t^i)$. This will allow us to tackle the non linearity appearing through $\frac{\sum_{j=1}^N |Z_t^j|}{N}$.

4.4.3 Some calculations

Like previously, we now have to show contraction in all three regions of space. Recall $f'(r_t^i) \leq 1$. The same calculations as before will be used, we only detail here the differences.

- First, since $\frac{L_W}{\lambda} (6 + 8\lambda) \leq \frac{\gamma}{8}$, by using Lemma 3 and since

$$H_j^N \exp\left(a\sqrt{H_i^N}\right) \leq H_i^N \exp\left(a\sqrt{H_i^N}\right) + H_j^N \exp\left(a\sqrt{H_j^N}\right)$$

we obtain

$$\begin{aligned} \epsilon L_W (6 + 8\lambda) f(r_t^i) \left(\frac{\sum_{j=1}^N |X_t^{j,N}|}{N} \right)^2 \exp\left(a\sqrt{H_i^N}\right) \\ - \frac{\gamma\epsilon}{8N} f(r_t^i) \left(N H_i^N \exp\left(a\sqrt{H_i^N}\right) + \sum_{j=1}^N H_j^N \exp\left(a\sqrt{H_j^N}\right) \right) \leq 0. \end{aligned}$$

This takes care of (56).

- We have, since $f'(r_t^i) \leq 1$: $\frac{1}{N} \sum_{i=1}^N \frac{\sum_{j=1}^N |Z_t^j|}{N} f'(r_t^i) G_t^i \leq \frac{\sum_{i,j=1}^N |Z_t^j| G_t^i}{N^2}$.
Then, using Lemma 8

$$\begin{aligned} \frac{1}{N^2} \sum_{i,j=1}^N |Z_t^i| G_t^j &= \frac{1}{N} \sum_{i=1}^N |Z_t^i| + \frac{2\epsilon}{N^2} \sum_{i,j=1}^N |Z_t^i| \tilde{H}(\bar{X}_t^j, \bar{V}_t^j) + \frac{2\epsilon}{N^2} \sum_{i,j=1}^N |Z_t^i| \tilde{H}(X_t^{j,N}, V_t^{j,N}) \\ &\leq \frac{C_1}{N} \sum_{i=1}^N \rho_t^i + \frac{2C_z\epsilon}{N^2} \sum_{i,j=1}^N f(r_t^i) \left(\tilde{H}(\bar{X}_t^j, \bar{V}_t^j) + \tilde{H}(X_t^{j,N}, V_t^{j,N}) \right) \\ &\quad + \frac{2C_z\epsilon^2}{N^2} \frac{2}{a} \sum_{i,j=1}^N f(r_t^i) \left(\sqrt{\bar{H}_i} + \sqrt{H_i^N} \right) \left(\sqrt{\bar{H}_j} \exp\left(a\sqrt{\bar{H}_j}\right) + \sqrt{H_j^N} \exp\left(a\sqrt{H_j^N}\right) \right). \end{aligned}$$

First, using (38)

$$\frac{2C_z\epsilon}{N^2} \sum_{i,j=1}^N f(r_t^i) \left(\tilde{H}(\bar{X}_t^j, \bar{V}_t^j) + \tilde{H}(X_t^{j,N}, V_t^{j,N}) \right) \leq \frac{2C_z\epsilon}{N^2} \sum_{i,j=1}^N f(r_t^i) \left(\bar{H}_j \exp\left(a\sqrt{\bar{H}_j}\right) + H_j^N \exp\left(a\sqrt{H_j^N}\right) \right).$$

Since

$$\begin{aligned} \left(\sqrt{\bar{H}_i} + \sqrt{H_i^N} \right) \left(\sqrt{\bar{H}_j} \exp\left(a\sqrt{\bar{H}_j}\right) + \sqrt{H_j^N} \exp\left(a\sqrt{H_j^N}\right) \right) \\ \leq 2 \left(\bar{H}_i \exp\left(a\sqrt{\bar{H}_i}\right) + \bar{H}_j \exp\left(a\sqrt{\bar{H}_j}\right) + H_i^N \exp\left(a\sqrt{H_i^N}\right) + H_j^N \exp\left(a\sqrt{H_j^N}\right) \right), \end{aligned}$$

we have

$$\begin{aligned} & \frac{4\mathcal{C}_z\epsilon^2}{aN^2} \sum_{i,j=1}^N f(r_t^i) \left(\sqrt{\bar{H}_i} + \sqrt{H_i^N} \right) \left(\sqrt{\bar{H}_j} \exp \left(a\sqrt{\bar{H}_j} \right) + \sqrt{H_j^N} \exp \left(a\sqrt{H_j^N} \right) \right) \\ & \leq \frac{8\mathcal{C}_z\epsilon^2}{aN^2} \sum_{i,j=1}^N f(r_t^i) \left(\bar{H}_i \exp \left(a\sqrt{\bar{H}_i} \right) + \bar{H}_j \exp \left(a\sqrt{\bar{H}_j} \right) + H_i^N \exp \left(a\sqrt{H_i^N} \right) + H_j^N \exp \left(a\sqrt{H_j^N} \right) \right). \end{aligned}$$

This way, since $2\mathcal{C}_z L_W \leq \frac{\gamma}{32}$, $L_W \epsilon \frac{8\mathcal{C}_z}{a} \leq \frac{\gamma}{32}$, and $L_W \mathcal{C}_1 \leq c$, we get

$$\begin{aligned} & \frac{1}{N} \sum_{i=1}^N \left(L_W \frac{\sum_{j=1}^N |Z_t^j|}{N} f'(r_t^i) G_t^i - c f(r_t^i) G_t^i - \epsilon f(r_t^i) \left(\frac{\gamma}{16} \bar{H}_i \exp \left(a\sqrt{\bar{H}_i} \right) + \frac{\gamma}{16} H_i^N \exp \left(a\sqrt{H_i^N} \right) \right. \right. \\ & \quad \left. \left. + \frac{\gamma}{16N} \sum_{j=1}^N \bar{H}_j \exp \left(a\sqrt{\bar{H}_j} \right) + \frac{\gamma}{16N} \sum_{j=1}^N H_j^N \exp \left(a\sqrt{H_j^N} \right) \right) \right) \leq 0 \end{aligned}$$

– Using Cauchy-Schwarz inequality

$$\begin{aligned} & \mathbb{E} \left(\left| G_t^i \nabla W * \bar{\mu}_t(\bar{X}_t^i) - \frac{1}{N} \sum_{j=1}^N \nabla W(\bar{X}_t^i - \bar{X}_t^j) \right| \right) \\ & \leq \mathbb{E} \left(G_t^{i2} \right)^{1/2} \mathbb{E} \left(\left| \nabla W * \bar{\mu}_t(\bar{X}_t^i) - \frac{1}{N} \sum_{j=1}^N \nabla W(\bar{X}_t^i - \bar{X}_t^j) \right|^2 \right)^{1/2}, \\ & \leq \mathbb{E} \left(G_t^{i2} \right)^{1/2} \mathbb{E} \left(\mathbb{E} \left(\left| \nabla W * \bar{\mu}_t(\bar{X}_t^i) - \frac{1}{N} \sum_{j=1}^N \nabla W(\bar{X}_t^i - \bar{X}_t^j) \right|^2 \middle| \bar{X}_t^i \right) \right)^{1/2}. \end{aligned}$$

Moreover, we notice that given \bar{X}_t^i , the random variables \bar{X}_t^j for $j \neq i$ are i.i.d with law $\bar{\mu}_t$. Hence

$$\begin{aligned} & \mathbb{E} \left(\left| \nabla W * \bar{\mu}_t(\bar{X}_t^i) - \frac{1}{N-1} \sum_{j=1}^N \nabla W(\bar{X}_t^i - \bar{X}_t^j) \right|^2 \middle| \bar{X}_t^i \right) = \frac{1}{N-1} \text{Var}_{\bar{\mu}_t}(\nabla W(\bar{X}_t^i - \cdot)), \\ & \leq \frac{2L_W^2}{N-1} \mathbb{E}_{\bar{\mu}_t}(|\cdot|^2), \end{aligned}$$

so

$$\begin{aligned}
& \mathbb{E} \left(\left| \nabla W * \bar{\mu}_t (\bar{X}_t^i) - \frac{1}{N} \sum_{j=1}^N \nabla W (\bar{X}_t^i - \bar{X}_t^j) \right|^2 \right) \\
& \leq \mathbb{E} \left(\left| \nabla W * \bar{\mu}_t (\bar{X}_t^i) - \frac{1}{N-1} \sum_{j=1}^N \nabla W (\bar{X}_t^i - \bar{X}_t^j) \right|^2 \right) \\
& \quad + \mathbb{E} \left(\left| \frac{1}{N} \sum_{j=1}^N \nabla W (\bar{X}_t^i - \bar{X}_t^j) - \frac{1}{N-1} \sum_{j=1}^N \nabla W (\bar{X}_t^i - \bar{X}_t^j) \right|^2 \right), \\
& \leq \frac{2L_W^2}{N-1} \mathbb{E}_{\bar{\mu}_t} (|\cdot|^2) + \left(\frac{1}{N-1} - \frac{1}{N} \right)^2 N \sum_{j=1}^N L_W^2 \mathbb{E} (|\bar{X}_t^i - \bar{X}_t^j|^2), \\
& \leq 2L_W^2 \left(\frac{1}{N-1} + \frac{1}{(N-1)^2} \right) \mathbb{E}_{\bar{\mu}_t} (|\cdot|^2).
\end{aligned}$$

We may then use $\mathbb{E}_{\bar{\mu}_t} (|\cdot|^2) \leq \frac{c^0}{\lambda}$.

Thus, by the same exact construction as before, we can obtain the existence of a function f and a constant $c > 0$ such that in all regions of space, for L_W sufficiently small,

$$\mathbb{E} \left(\frac{1}{N} \sum_i K_t^i \right) \leq \xi (1 + \alpha) \mathcal{C}_{G,1} + L_W \frac{\mathcal{C}^0 \mathcal{C}_{G,2}^{1/2}}{\lambda} \left(\frac{2}{N-1} + \frac{2}{(N-1)^2} \right)^{1/2}.$$

By taking the expectation in the dynamic of ρ_t given by (50) and (51) at stopping times τ_n increasingly converging to t , we prove Theorem 4 by using Fatou's lemma for $n \rightarrow \infty$.

A Various results

A.1 Proof of lemma 1

The property only depends on the distance to the the origin, not the direction. We therefore only need to prove it in dimension 1, making sure the constant A is independent of the direction. There is $x_0 > 0$ such that $\frac{\lambda}{2} x_0^2 = 2A$. Therefore, for $x \geq 0$, using (4):

$$U' (x_0 + x) (x_0 + x) \geq 2\lambda U (x_0 + x) + \frac{\lambda}{2} (x_0 + x)^2 - 2A = 2\lambda U (x_0 + x) + \frac{\lambda}{2} x^2 + \lambda x x_0.$$

Then, for $x \geq 0$:

$$\begin{aligned} U(x_0 + x) - U(x_0) &= \int_0^1 U'(x_0 + tx) x dt = \int_0^1 U'(x_0 + tx) (x_0 + tx) \frac{x}{x_0 + tx} dt \\ &\geq \frac{x}{x_0 + x} \int_0^1 2\lambda U(x_0 + tx) + \frac{\lambda}{2} t^2 x^2 + \lambda t x x_0 dt \\ &\geq \frac{x}{x_0 + x} \left(\frac{\lambda}{6} x^2 + \frac{\lambda}{2} x x_0 \right) \quad \text{since } U \geq 0 \\ &= \frac{\lambda}{6} \frac{x^3}{x_0 + x} + \frac{\lambda}{2} \frac{x^2 x_0}{x_0 + x}. \end{aligned}$$

We thus have for all $x \geq x_0$:

$$\begin{aligned} U(x) - U(x_0) &\geq \frac{\lambda}{6} \frac{(x - x_0)^3}{x} + \frac{\lambda}{2} \frac{(x - x_0)^2 x_0}{x} \\ &= \frac{\lambda}{6} x^2 - \frac{\lambda}{2} x x_0 + \frac{\lambda}{2} x_0^2 - \frac{\lambda}{6} \frac{x_0^3}{x} + \frac{\lambda}{2} x x_0 - \lambda x_0^2 + \frac{\lambda}{2} \frac{x_0^3}{x} \\ &= \frac{\lambda}{6} x^2 - \frac{\lambda}{2} x_0^2 + \frac{\lambda}{3} \frac{x_0^3}{x}. \end{aligned}$$

However, $-\frac{\lambda}{2} x_0^2 + \frac{\lambda}{3} \frac{x_0^3}{x} \geq -\frac{\lambda}{2} x_0^2 = -2A$ for $x \geq x_0$. We therefore have the desired result for $x \geq x_0$. The same reasoning gives us the result for $x \leq -x_0$.

Hence, if $|x| \geq |x_0| = \sqrt{\frac{4A}{\lambda}}$, $U(x) - U(x_0) \geq \frac{\lambda}{6} x^2 - 2A$. We then use the fact that $U(x)$ is continuous on the sphere of center 0 and radius $\sqrt{\frac{4A}{\lambda}}$, hence bounded on this set, to give a lower bound on $U(x_0)$ independent of the direction. Finally, for $|x| \in [-x_0, x_0]$, the function $x \mapsto U(x) - \frac{\lambda}{6} x^2$ is continuous, therefore bounded.

A.2 Proof of lemma 2

We have

$$\nabla W * \mu(x) - \nabla W * \nu(\tilde{x}) = \nabla W * \mu(x) - \nabla W * \mu(\tilde{x}) + \nabla W * \mu(\tilde{x}) - \nabla W * \nu(\tilde{x})$$

Let (X, \tilde{X}) be a coupling of μ and ν . Then

$$\begin{aligned} |\nabla W * \mu_t(x) - \nabla W * \tilde{\mu}_t(\tilde{x})| &= \left| \mathbb{E} \left(\nabla W(x - X) - \nabla W(\tilde{x} - \tilde{X}) \right) \right| \\ &\leq L_W \mathbb{E}(|x - X - \tilde{x} + \tilde{X}|) \\ &\leq L_W \mathbb{E}(|x - \tilde{x}| + |X - \tilde{X}|) \end{aligned}$$

This being true for all coupling, we obtain the desired result.

A.3 Proof of Lemma 3

Remark 3 With γ given by (7), we have $\gamma \leq \frac{1}{2}$.

We have

$$\begin{aligned}\mathcal{L}_\mu H(x, v) &= v \cdot \nabla_x H(x, v) - v \cdot \nabla_v H(x, v) - \nabla U(x) \cdot \nabla_v H(x, v) - \nabla W * \mu(x) \cdot \nabla_v H(x, v) \\ &\quad + \Delta_v H(x, v) \\ &= v \cdot (24 \nabla U(x) + 12(1 - \gamma)x + 2\lambda x + 12v) - v \cdot (12x + 24v) \\ &\quad - \nabla U(x) \cdot (12x + 24v) - \nabla W * \mu(x) \cdot (12x + 24v) + 24d \\ &= 24d - 12 \nabla U(x) \cdot x + x \cdot v (12(1 - \gamma) + 2\lambda - 12) - \nabla W * \mu(x) \cdot (12x + 24v) - 12|v|^2,\end{aligned}$$

with

$$\begin{aligned}-\gamma H(x, v) &= -24\gamma U(x) - 6\gamma(1 - \gamma)|x|^2 - \gamma\lambda|x|^2 - 12\gamma x \cdot v - 12\gamma|v|^2 \\ -12 \nabla U(x) \cdot x &\leq -24\lambda U(x) - 6\lambda|x|^2 + 24A \\ -\nabla W * \mu(x) \cdot (12x + 24v) &\leq (L_W|x| + L_W \mathbb{E}_\mu(|\cdot|)) (12|x| + 24|v|) \\ &\leq 12L_W|x|^2 + 24L_W|x||v| + L_W \mathbb{E}_\mu(|\cdot|) \left(6 \frac{|x|^2}{a_x \mathbb{E}_\mu(|\cdot|)} + 6a_x \mathbb{E}_\mu(|\cdot|) \right) \\ &\quad + 12 \frac{|v|^2}{a_v \mathbb{E}_\mu(|\cdot|)} + 12a_v \mathbb{E}_\mu(|\cdot|),\end{aligned}$$

where this last inequality holds for any $a_x, a_v > 0$. Therefore

$$\begin{aligned}\mathcal{L}_\mu H(x, v) &\leq 24A + 24d + 6L_W \mathbb{E}_\mu(|\cdot|)^2 (a_x + 2a_v) - \gamma H(x, v) + 24\gamma U(x) + 6\gamma(1 - \gamma)|x|^2 \\ &\quad + \gamma\lambda|x|^2 + 12\gamma x \cdot v + 12\gamma|v|^2 - 24\lambda U(x) - 6\lambda|x|^2 + 12L_W|x|^2 + 24L_W|x||v| \\ &\quad + \frac{6L_W}{a_x}|x|^2 + \frac{12L_W}{a_v}|v|^2 + x \cdot v (12(1 - \gamma) + 2\lambda - 12) - 12|v|^2,\end{aligned}$$

and then

$$\begin{aligned}\mathcal{L}_\mu H(x, v) &\leq 24A + 24d + 6L_W \mathbb{E}_\mu(|\cdot|)^2 (a_x + 2a_v) - \gamma H(x, v) + 24U(x)(\gamma - \lambda) \\ &\quad + |x||v| (12\gamma + 12(1 - \gamma) + 2\lambda - 12) + 24L_W \\ &\quad + |x|^2 \left(6\gamma(1 - \gamma) + \gamma\lambda - 6\lambda + 12L_W + \frac{6L_W}{a_x} \right) \\ &\quad + |v|^2 \left(12\gamma - 12 + \frac{12L_W}{a_v} \right).\end{aligned}$$

We now use $|x||v| \leq \frac{\lambda}{3}|x|^2 + \frac{3}{4\lambda}|v|^2$, and $|12\gamma\lambda + 12(1 - \gamma)\lambda + 2\lambda - 12| = 2\lambda$.

We have $(\gamma - \lambda) < 0$. Hence $24U(x)(\gamma - \lambda) \leq 4\lambda(\gamma - \lambda)|x|^2 - 24(\gamma - \lambda)\tilde{A}$ using Lemma 1. Then

$$\begin{aligned}\mathcal{L}_\mu H(x, v) &\leq 24A - 24(\gamma - \lambda)\tilde{A} + 24d + 6L_W \mathbb{E}_\mu(|\cdot|)^2 (a_x + 2a_v) - \gamma H(x, v) \\ &\quad + |x|^2 \left(4\lambda(\gamma - \lambda) + 6\gamma(1 - \gamma) + \gamma\lambda - 6\lambda + \frac{6L_W}{a_x} + 12L_W + \frac{2\lambda^2}{3} + \frac{24L_W\lambda}{3} \right) \\ &\quad + |v|^2 \left(12\gamma - 12 + \frac{12L_W}{a_v} + \frac{3}{2} + \frac{3}{4\lambda} 24L_W \right).\end{aligned}$$

We now consider each term individually.

Coefficient of $|x|^2$. We have, using $0 < \gamma < 1$ and $L_W \leq \frac{\lambda}{8}$

$$\begin{aligned}4\lambda(\gamma - \lambda) + 6\gamma(1 - \gamma) + \gamma\lambda - 6\lambda + \frac{2\lambda^2}{3} + \frac{24L_W\lambda}{3} + 12L_W \\ \leq \gamma(5\lambda + 6) - \left(4\lambda^2 + 6\lambda - \frac{2\lambda^2}{3} - \lambda^2 - \frac{3\lambda}{2} \right).\end{aligned}$$

Therefore, it is sufficient that

$$\gamma \leq \lambda \frac{\frac{7}{3}\lambda + \frac{9}{2}}{5\lambda + 6}.$$

We check this holds for $\gamma = \frac{\lambda}{2\lambda+2}$. Then

$$\begin{aligned} & 4\lambda(\gamma - \lambda) + 6\gamma(1 - \gamma) + \gamma\lambda - 6\lambda + \frac{2\lambda^2}{3} + \frac{24L_W\lambda}{3} + 12L_W + \frac{6L_W}{a_x} \\ & \leq (5\lambda + 6) \left(\gamma - \frac{\frac{7}{3}\lambda + \frac{9}{2}\lambda}{5\lambda + 6} \right) + \frac{3\lambda}{4a_x}. \end{aligned}$$

We therefore choose

$$\frac{3\lambda}{4a_x} \leq -(5\lambda + 6) \left(\gamma - \frac{\frac{7}{3}\lambda + \frac{9}{2}\lambda}{5\lambda + 6} \right) = \frac{7}{3}\lambda^2 + \frac{9}{2}\lambda - \frac{5\lambda^2 + 6\lambda}{2\lambda + 2}.$$

It is, for that, sufficient to take

$$\frac{3\lambda}{4a_x} = \frac{3}{4}\lambda, \quad \text{i.e.} \quad a_x = 1.$$

Furthermore

$$\begin{aligned} & 6\lambda(\gamma - \lambda) + 6\gamma(1 - \gamma) + \gamma\lambda - 6\lambda + \frac{2\lambda^2}{3} + \frac{24L_W\lambda}{3} + 12L_W + \frac{6L_W}{a_x} \\ & \leq \frac{5\lambda^2 + 6\lambda}{2\lambda + 2} - \frac{7}{3}\lambda^2 - \frac{9}{2}\lambda + \frac{3}{4}\lambda = \frac{5\lambda^2 + 6\lambda}{2\lambda + 2} - \frac{7}{3}\lambda^2 - \frac{15}{4}\lambda. \end{aligned}$$

We then observe

$$6\lambda(\gamma - \lambda) + 6\gamma(1 - \gamma) + \gamma\lambda - 6\lambda + \frac{6L_W}{a_x} + \frac{2\lambda^2}{3} + \frac{24L_W\lambda}{3} + 12L_W \leq -\lambda^2 - \frac{3}{4}\lambda.$$

And finally

$$\begin{aligned} \forall \lambda > 0, \forall x, \quad |x|^2 \left(6\lambda(\gamma - \lambda) + 6\gamma(1 - \gamma) + \gamma\lambda - 6\lambda + \frac{6L_W}{a_x} + \frac{2\lambda^2}{3} + \frac{48L_W\lambda}{3} + 12L_W \right) \\ \leq -\lambda^2|x|^2 - \frac{3}{4}\lambda|x|^2 \end{aligned}$$

Coefficient of $|v|^2$. We have, using $0 < \gamma \leq \frac{1}{2}$ and $L_W \leq \lambda/8$

$$12\gamma - 12 + \frac{3}{2} + \frac{3}{4\lambda}24L_W \leq -6 + \frac{3}{2} + \frac{18}{\lambda} \cdot \frac{\lambda}{8} = -6 + \frac{3}{2} + \frac{9}{4} = -\frac{9}{4}.$$

We then choose

$$\frac{12\lambda}{8a_v} = \frac{9}{4}, \quad \text{i.e.} \quad a_v = \frac{2}{3}\lambda.$$

Therefore

$$\forall \lambda > 0, \forall v, \quad |v|^2 \left(12\gamma - 12 + \frac{3}{2} + \frac{3}{4\lambda}24L_W + \frac{12L_W}{a_v} \right) \leq 0.$$

We thus obtain

$$\mathcal{L}_\mu H(x, v) \leq 24 \left(A - (\gamma - \lambda) \tilde{A} + d \right) + 6L_W \mathbb{E}_\mu(|\cdot|)^2 \left(1 + \frac{4}{3}\lambda \right) - \lambda^2|x|^2 - \frac{3}{4}\lambda|x|^2 - \gamma H(x, v),$$

i.e

$$\mathcal{L}_\mu H(x, v) \leq 24 \left(A - (\gamma - \lambda) \tilde{A} + d \right) + \mathbb{E}_\mu(|\cdot|)^2 \left(\frac{3}{4}\lambda + \lambda^2 \right) - \lambda^2|x|^2 - \frac{3}{4}\lambda|x|^2 - \gamma H(x, v). \quad (58)$$

A.4 Proof of Lemma 5

Using $1 - \gamma \geq \frac{1}{2}$, we get

$$H(x, v) \geq 24U(x) + (3 + \lambda)|x|^2 + 12\left|v + \frac{x}{2}\right|^2 - 3|x|^2,$$

which is (8). We then have

$$H(x, v) \geq \min\left(\frac{2}{3}\lambda, 6\right)(|v|^2 + |x + v|^2).$$

Thus

$$\begin{aligned} r(x, \tilde{x}, v, \tilde{v})^2 &\leq ((1 + \alpha)|x - \tilde{x} + v - \tilde{v}| + \alpha|v - \tilde{v}|)^2 \\ &\leq 2(1 + \alpha)^2|x - \tilde{x} + v - \tilde{v}|^2 + 2\alpha^2|v - \tilde{v}|^2 \\ &\leq 4\left((1 + \alpha)^2 + \alpha^2\right)(|x + v|^2 + |v|^2 + |\tilde{x} + \tilde{v}|^2 + |\tilde{v}|^2). \end{aligned}$$

Therefore we obtain the final point.

A.5 Proof of control of L1 and L2 Wasserstein distances

We prove Lemma 8. Using the definition of R_1 and (7), and since $B \geq d \geq 1$ and $\gamma \leq \frac{1}{2}$, we have $R_1 \geq 1$.

– First for the L1-Wasserstein distance

$$|x - x'| + |v - v'| \leq |v - v' + x - x'| + 2|x - x'| \leq \max\left(\frac{2}{\alpha}, 1\right)r((x, v), (x', v')).$$

If $r((x, v), (x', v')) \leq 1 \leq R_1$

$$r((x, v), (x', v')) \leq \frac{f(r)}{f'_-(R_1)} \leq \frac{\rho((x, v), (x', v'))}{\phi(R_1)g(R_1)}.$$

If $r((x, v), (x', v')) \geq 1$, we have shown (13)

$$r((x, v), (x', v')) \leq r^2((x, v), (x', v')) \leq 4\frac{(1 + \alpha)^2 + \alpha^2}{\min(\frac{2}{3}\lambda, 6)}(H(x, v) + H(x', v')).$$

Thus

$$\begin{aligned} r((x, v), (x', v')) &\leq \frac{4(1 + \alpha)^2 + \alpha^2}{\epsilon \min(\frac{2}{3}\lambda, 6)}(\epsilon H(x, v) + \epsilon H(x', v')) \\ &\leq \frac{4(1 + \alpha)^2 + \alpha^2}{\epsilon \min(\frac{2}{3}\lambda, 6)}\frac{\rho((x, v), (x', v'))}{f(r)} \\ &\leq \frac{4(1 + \alpha)^2 + \alpha^2}{\epsilon \min(\frac{2}{3}\lambda, 6)}\frac{\rho((x, v), (x', v'))}{f(1)}. \end{aligned}$$

Therefore

$$|x - x'| + |v - v'| \leq \max\left(\frac{2}{\alpha}, 1\right)\max\left(\frac{4((1 + \alpha)^2 + \alpha^2)}{\epsilon \min(\frac{2}{3}\lambda, 6)f(1)}, \frac{1}{\phi(R_1)g(R_1)}\right)\rho((x, v), (x', v')).$$

– Then for the L2-Wasserstein distance

$$|v - v'|^2 = |v - v' + x - x' - (x - x')|^2 \leq 2|v - v' + x - x'|^2 + 2|x - x'|^2.$$

Hence

$$|x - x'|^2 + |v - v'|^2 \leq 3(|v - v' + x - x'|^2 + |x - x'|^2).$$

But

$$\begin{aligned} r^2((x, v), (x', v')) &= (\alpha|x - x'| + |x - x' + v - v'|)^2 \\ &\geq \alpha^2|x - x'|^2 + |x - x' + v - v'|^2 \\ &\geq (1 + \alpha^2)(|x - x'|^2 + |x - x' + v - v'|^2) \\ &\geq \frac{1 + \alpha^2}{3}(|x - x'|^2 + |v - v'|^2). \end{aligned}$$

If $r((x, v), (x', v')) \leq 1 \leq R_1$

$$r^2((x, v), (x', v')) \leq r((x, v), (x', v')) \leq \frac{f(r)}{f'_-(R_1)} \leq \frac{\rho((x, v), (x', v'))}{\phi(R_1)g(R_1)}.$$

If $r((x, v), (x', v')) \geq 1$, we have shown (13)

$$r^2((x, v), (x', v')) \leq 4 \frac{(1 + \alpha)^2 + \alpha^2}{\min(\frac{2}{3}\lambda, 6)} (H(x, v) + H(x', v')).$$

Thus

$$\begin{aligned} r((x, v), (x', v')) &\leq \frac{4}{\epsilon} \frac{(1 + \alpha)^2 + \alpha^2}{\min(\frac{2}{3}\lambda, 6)} (\epsilon H(x, v) + \epsilon H(x', v')) \\ &\leq \frac{4}{\epsilon} \frac{(1 + \alpha)^2 + \alpha^2}{\min(\frac{2}{3}\lambda, 6)} \frac{\rho((x, v), (x', v'))}{f(r)} \\ &\leq \frac{4}{\epsilon} \frac{(1 + \alpha)^2 + \alpha^2}{\min(\frac{2}{3}\lambda, 6)} \frac{\rho((x, v), (x', v'))}{f(1)}. \end{aligned}$$

Therefore

$$|x - x'|^2 + |v - v'|^2 \leq \frac{3}{1 + \alpha^2} \max \left(\frac{4((1 + \alpha)^2 + \alpha^2)}{\epsilon \min(\frac{2}{3}\lambda, 6) f(1)}, \frac{1}{\phi(R_1)g(R_1)} \right) \rho((x, v), (x', v')).$$

A.6 Proof of Lemma 9

We have

$$\begin{aligned} H(x, v) - H(\tilde{x}, \tilde{v}) &= 24(U(x) - U(\tilde{x})) + (6(1 - \gamma) + \lambda)(|x|^2 - |\tilde{x}|^2) + 12(x \cdot v - \tilde{x} \cdot \tilde{v}) + 12(|v|^2 - |\tilde{v}|^2) \\ &= 24(U(x) - U(\tilde{x})) + (6(1 - \gamma) + \lambda - 3)(|x|^2 - |\tilde{x}|^2) + 12 \left(\left| v + \frac{x}{2} \right|^2 - \left| \tilde{v} + \frac{\tilde{x}}{2} \right|^2 \right). \end{aligned}$$

We first have

$$||x|^2 - |\tilde{x}|^2| \leq |x - \tilde{x}|(|x| + |\tilde{x}|) \leq \frac{r(x, v, \tilde{x}, \tilde{v})}{\alpha\sqrt{\lambda}} \left(\sqrt{H(x, v)} + \sqrt{H(\tilde{x}, \tilde{v})} \right).$$

Then

$$\begin{aligned} \left| \left| v + \frac{x}{2} \right|^2 - \left| \tilde{v} + \frac{\tilde{x}}{2} \right|^2 \right| &\leq \left| v + \frac{x}{2} - \tilde{v} - \frac{\tilde{x}}{2} \right| \left(\left| v + \frac{x}{2} \right| + \left| \tilde{v} + \frac{\tilde{x}}{2} \right| \right) \\ &\leq \frac{1}{\sqrt{12}} |v - \tilde{v} + \frac{1}{2}(x - \tilde{x})| \left(\sqrt{H(x, v)} + \sqrt{H(\tilde{x}, \tilde{v})} \right) \\ &\leq \frac{1}{2\sqrt{3}} \max \left(1, \frac{1}{2\alpha} \right) r(x, v, \tilde{x}, \tilde{v}) \left(\sqrt{H(x, v)} + \sqrt{H(\tilde{x}, \tilde{v})} \right). \end{aligned}$$

And finally

$$\begin{aligned} |U(x) - U(\tilde{x})| &= \left| \int_0^1 \nabla U(\tilde{x} + t(x - \tilde{x})) \cdot (x - \tilde{x}) dt \right| \\ &\leq \sup_{t \in [0, 1]} |\nabla U(\tilde{x} + t(x - \tilde{x}))| |x - \tilde{x}| \\ &\leq (\nabla U(0) + L_U(|x| + |\tilde{x}|)) |x - \tilde{x}| \\ &\leq \left(\nabla U(0) + \frac{L_U}{\sqrt{\lambda}} \left(\sqrt{H(x, v)} + \sqrt{H(\tilde{x}, \tilde{v})} \right) \right) \frac{r(x, v, \tilde{x}, \tilde{v})}{\alpha}. \end{aligned}$$

These three inequalities yield the desired result.

B Proof of Lemma 6

We first rewrite the various conditions on the parameters.

- Since for all $u \geq 0$, $0 < \phi(u) \leq 1$, we have $0 < \Phi(s) = \int_0^s \phi(u) du \leq s$, i.e $s/\Phi(s) \geq 1$.
Therefore

$$\inf_{r \in [0, R_1]} \frac{r\phi(r)}{\Phi(r)} \geq \inf_{r \in [0, R_1]} \phi(r) = \phi(R_1).$$

It is thus sufficient for (17) that

$$c + 2\epsilon B \leq \frac{1}{2} \left(1 - \frac{1}{\alpha} (L_U + L_W) \right) \phi(R_1). \quad (59)$$

- We have

$$\phi(r) \leq \exp \left(-\frac{L_U + L_W}{8\alpha} r^2 \right).$$

So

$$\Phi(r) \leq \int_0^\infty \exp \left(-\frac{L_U + L_W}{8\alpha} s^2 \right) ds = \sqrt{\frac{2\pi\alpha}{L_U + L_W}}.$$

Then

$$\int_0^{R_1} \frac{\Phi(r)}{\phi(r)} dr \leq \sqrt{\frac{2\pi\alpha}{L_U + L_W}} R_1 \frac{1}{\phi(R_1)}.$$

It is thus sufficient for (18) that

$$c + 2\epsilon B \leq 2\sqrt{\frac{L_U + L_W}{2\pi\alpha}} \frac{\phi(R_1)}{R_1}. \quad (60)$$

At this point, we have now proven that under Assumption 1, Assumption 2 and Assumption 3, for the parameters to satisfy Lemma 6 it is sufficient for them to satisfy

$$\alpha > L_U + L_W, \quad (61)$$

$$c \leq \frac{\gamma}{6} \left(1 - \frac{\frac{5}{6}\gamma}{2\epsilon B + \frac{5}{6}\gamma} \right), \quad (62)$$

$$c + 2\epsilon B \leq \frac{1}{2} \left(1 - \frac{1}{\alpha} (L_U + L_W) \right) \phi(R_1), \quad (63)$$

$$c + 2\epsilon B \leq 2\sqrt{\frac{L_U + L_W}{2\pi\alpha}} \frac{\phi(R_1)}{R_1}, \quad (64)$$

with, again

$$B = 24 \left(A + (\lambda - \gamma) \tilde{A} + d \right), \quad R_1 = \sqrt{(1 + \alpha)^2 + \alpha^2} \sqrt{\frac{24}{5\gamma \min(3, \frac{1}{3}\lambda)}} B.$$

Let us show that there are positive parameters ϵ , α , L_W and c satisfying those conditions.

For inequality (60) it is sufficient, as $L_W < \frac{\lambda}{8}$, to consider

$$\alpha = L_U + \frac{\lambda}{4},$$

while inequality (61) first invites us to consider $2\epsilon B$ of a comparable order to c

$$2\epsilon B = \delta c.$$

We have thus switched parameter ϵ for δ . First we translate (61) into our new parameter:

$$\begin{aligned} c \leq \frac{\gamma}{6} \left(1 - \frac{\frac{5}{6}\gamma}{2\epsilon B + \frac{5}{6}\gamma} \right) &\iff c \leq \frac{\gamma}{6} \frac{\delta c}{\delta c + \frac{5}{6}\gamma} \\ &\iff 1 \leq \frac{\gamma}{6} \frac{\delta}{\delta c + \frac{5}{6}\gamma} \quad (\text{since } c \geq 0) \\ &\iff c \leq \frac{\gamma}{6} \frac{\delta - 5}{\delta}. \end{aligned}$$

The appearance of $\phi(R_1)$ in (62) and (63) suggests we should try to minimize it. Let us assume, for simplicity, that $\epsilon \leq 1$, which is equivalent to having $c \leq \frac{2B}{\delta}$. We then have

$$\begin{aligned} \phi(r) &= \exp \left(-\frac{1}{8} \left(\frac{1}{\alpha} (L_U + L_W) + \alpha + 96\epsilon \max \left(\frac{1}{2\alpha}, 1 \right) \right) r^2 \right) \\ &\geq \exp \left(-\frac{1}{8} \left(\frac{L_U + L_W}{\alpha} + \alpha + 96 \max \left(\frac{1}{2\alpha}, 1 \right) \right) r^2 \right) \quad \text{on } [0, R_1]. \end{aligned} \quad (65)$$

Now, using (64), we have for (62) and (63) that it is sufficient that

$$c \leq \frac{1}{2(\delta + 1)} \left(1 - \frac{1}{\alpha} (L_U + L_W) \right) \exp \left(-\frac{1}{8} \left(\frac{L_U + L_W}{\alpha} + \alpha + 96 \max \left(\frac{1}{2\alpha}, 1 \right) \right) R_1^2 \right),$$

and

$$c \leq \frac{2}{\delta + 1} \sqrt{\frac{L_U + L_W}{2\pi\alpha}} \exp \left(-\frac{1}{8} \left(\frac{L_U + L_W}{\alpha} + \alpha + 96 \max \left(\frac{1}{2\alpha}, 1 \right) \right) R_1^2 \right).$$

We could now optimize parameter δ , but for the sake of conciseness, we choose $\delta = 6$.

Recall $0 \leq L_W < \frac{\lambda}{8}$. This way, both in (14) and (33), c and C_1 can be bounded independently of L_W . Hence why L_W , in (33) and (49), can be chosen last, and every other quantities can be chosen independently of L_W .

C ∇U locally Lipschitz-continuous

In this section we replace Assumption 2 with Assumption 4. We assume, for ν_0^1 and ν_0^2 the initial conditions,

$$\forall i \in \{1, 2\}, \mathbb{E}_{\nu_0^i} \left(\left(\int_0^{H(X,V)} e^{a\sqrt{u}} du \right)^2 \right) \leq (C^0)^2 \quad (66)$$

We show how the proof can be modified to still obtain contraction. We make the following assumption

$$L_\psi \leq c_\psi(L_U, \lambda, \tilde{A}, d, a) := L_U \frac{\gamma}{8} \min \left(\frac{a}{8}, \epsilon_{\psi,*} \right) \quad (67)$$

with \tilde{B} given by (42)-(44) and $\gamma = \frac{\lambda}{2(\lambda+1)}$. Then, choose $\epsilon_{\psi,*}$ to be smaller than ϵ , independently of L_W , using the assumption that $0 \leq L_W \leq \frac{\lambda}{8}$. Like previously, we consider

$$G_t = 1 + \epsilon \tilde{H}(X_t, V_t) + \epsilon \tilde{H}(\tilde{X}_t, \tilde{V}_t).$$

Hence following the same method as previously we obtain

$$K_t \leq G_t \left(c f(r_t) + \alpha f'(r_t) \frac{d|Z_t|}{dt} + (L_U + L_W) f'(r_t) |Z_t| + L_W f'(r_t) \mathbb{E}(|Z_t|) \right) \quad (68)$$

$$+ 4 f''(r_t) r c^2(Z_t, W_t) + \frac{1}{2} (\epsilon C_{f,1} + C_{f,2}) r_t f'(r_t) r c(Z_t, W_t)^2 \quad (69)$$

$$+ \epsilon \left(2\tilde{B} - \frac{\gamma}{8} (\tilde{H}(X_t, V_t) + \tilde{H}(\tilde{X}_t, \tilde{V}_t)) \right) f(r_t) \quad (70)$$

$$+ \left(\psi(X_t) + \psi(\tilde{X}_t) \right) |Z_t| f'(r_t) G_t - \frac{\gamma \epsilon}{8} \left(H(X_t, V_t) \exp(a\sqrt{H(X_t, V_t)}) + H(\tilde{X}_t, \tilde{V}_t) \exp(a\sqrt{H(\tilde{X}_t, \tilde{V}_t)}) \right) \quad (71)$$

$$+ \epsilon \frac{L_W}{\lambda} (6 + 8\lambda) \left(\exp(a\sqrt{H(X_t, V_t)}) \mathbb{E}H(X_t, V_t) + \exp(a\sqrt{H(\tilde{X}_t, \tilde{V}_t)}) \mathbb{E}H(\tilde{X}_t, \tilde{V}_t) \right) f(r_t). \quad (72)$$

We describe briefly how the terms will compensate each other before writing the calculations that are different.

- Like previously, lines (67) and (68) will be dealt with through the choice of function f , with the non linearity appearing at the end of (67) giving us a remaining expectation (cf bullet **1** below),
- line (69) will intervene like before in the last region of space (where we use that for all x in \mathbb{R} , $\tilde{H} \geq H$ to come back to calculations we've made in Section 3.2.3, cf bullet **2** below) and in the first two region of space to compensate line (70) (cf bullet **3** below),
- and line (71) will give us a remaining expectation (cf bullet **4** below).

Notice how we use the Lyapunov function to compensate ψ appearing when considering ∇U only locally Lipschitz continuous.

- **1.** We can find a constant $C_{1,e}$ such that for all $x, v, \tilde{x}, \tilde{v} \in \mathbb{R}^d$,

$$|x - \tilde{x}| + |v - \tilde{v}| \leq C_{1,e} \rho(x, v, \tilde{x}, \tilde{v}),$$

and thus

$$\mathbb{E}(\mathbb{E}(|Z_t|) G_t f'(r_t)) \leq C_{1,e} \mathbb{E}(\rho_t) \mathbb{E}(G_t).$$

- **2.** In the last region of space, we use the fact that

$$K_t \mathbb{1}_{R_3} \leq \left(\left(c - \frac{\gamma}{8} \right) G_t + 2\epsilon \tilde{B} + \frac{\gamma}{8} \right) f(r_t) \quad (73)$$

We deal with (72) exactly like in Section 3.2.3.

- 3. To deal with the only locally Lipschitz continuous aspect, we use the upper bound on ψ given in (66). In the first two regions of space we use $f'(r_t)|Z_t| \leq f'(r_t)r_t/\alpha \leq f(r_t)/\alpha$

$$\begin{aligned}
& G_t \left(\psi(X_t) + \psi(\tilde{X}_t) \right) f'(r_t) |Z_t| \\
& - \epsilon \frac{\gamma}{8} \left(H(X_t, V_t) \exp \left(a \sqrt{H(X_t, V_t)} \right) + H(\tilde{X}_t, \tilde{V}_t) \exp \left(a \sqrt{H(\tilde{X}_t, \tilde{V}_t)} \right) \right) f(r_t) \\
& \leq \left(\psi(X_t) + \psi(\tilde{X}_t) \right) f(r_t) f'(r_t) |Z_t| \\
& + \frac{2\epsilon}{a\alpha} \left(\psi(X_t) + \psi(\tilde{X}_t) \right) \left(\sqrt{H(X_t, V_t)} \exp \left(a \sqrt{H(X_t, V_t)} \right) + \sqrt{H(\tilde{X}_t, \tilde{V}_t)} \exp \left(a \sqrt{H(\tilde{X}_t, \tilde{V}_t)} \right) \right) f(r_t) \\
& - \epsilon \frac{\gamma}{8} \left(H(X_t, V_t) \exp \left(a \sqrt{H(X_t, V_t)} \right) + H(\tilde{X}_t, \tilde{V}_t) \exp \left(a \sqrt{H(\tilde{X}_t, \tilde{V}_t)} \right) \right) f(r_t),
\end{aligned}$$

where we used (39). On one hand, since $\psi(x) \leq L_\psi \sqrt{H(x, v)}$, we have

$$\psi(x) + \psi(\tilde{x}) \leq L_\psi \sqrt{H(x, v)} + L_\psi \sqrt{H(\tilde{x}, \tilde{v})} \leq L_\psi \left(\frac{H(x, v) + H(\tilde{x}, \tilde{v})}{2} + 1 \right),$$

and thus

$$\begin{aligned}
& (\psi(x) + \psi(\tilde{x})) f'(r_t) |Z_t| \\
& \leq \frac{L_\psi}{2} \left(H(x, v) \exp \left(a \sqrt{H(x, v)} \right) + H(\tilde{x}, \tilde{v}) \exp \left(a \sqrt{H(\tilde{x}, \tilde{v})} \right) \right) \frac{f(r_t)}{\alpha} + L_\psi f'(r_t) |Z_t|
\end{aligned}$$

On the other hand

$$\begin{aligned}
& (\psi(x) + \psi(\tilde{x})) \left(\sqrt{H(x, v)} \exp \left(a \sqrt{H(x, v)} \right) + \sqrt{H(\tilde{x}, \tilde{v})} \exp \left(a \sqrt{H(\tilde{x}, \tilde{v})} \right) \right) \\
& \leq L_\psi \left(\sqrt{H(x, v)} + \sqrt{H(\tilde{x}, \tilde{v})} \right) \left(\sqrt{H(x, v)} \exp \left(a \sqrt{H(x, v)} \right) + \sqrt{H(\tilde{x}, \tilde{v})} \exp \left(a \sqrt{H(\tilde{x}, \tilde{v})} \right) \right) \\
& \leq 2L_\psi \left(H(x, v) \exp \left(a \sqrt{H(x, v)} \right) + H(\tilde{x}, \tilde{v}) \exp \left(a \sqrt{H(\tilde{x}, \tilde{v})} \right) \right).
\end{aligned}$$

This way, since $\frac{L_\psi}{2\alpha} \leq \epsilon \frac{\gamma}{16}$ and $\frac{4L_\psi \epsilon}{a\alpha} \leq \epsilon \frac{\gamma}{16}$ (recall $\alpha > L_U$), we get, since in the third region of space $f''(r_t) = 0$,

$$\begin{aligned}
& G_t \left(\psi(X_t) + \psi(\tilde{X}_t) \right) f'(r_t) |Z_t| \\
& - \epsilon \frac{\gamma}{8} \left(H(X_t, V_t) \exp \left(a \sqrt{H(X_t, V_t)} \right) + H(\tilde{X}_t, \tilde{V}_t) \exp \left(a \sqrt{H(\tilde{X}_t, \tilde{V}_t)} \right) \right) f(r_t) \leq L_\psi |Z_t| f'(r_t) G_t.
\end{aligned}$$

The righthand side is dealt with through the choice of the concave function f (we consider $L_U + L_\psi$ instead of L_U).

- 4. Likewise, we can bound

$$\begin{aligned}
& \mathbb{E} \left(\epsilon \left(\exp \left(a \sqrt{H(X_t, V_t)} \right) \mathbb{E} H(X_t, V_t) + \exp \left(a \sqrt{H(\tilde{X}_t, \tilde{V}_t)} \right) \mathbb{E} H(\tilde{X}_t, \tilde{V}_t) \right) f(r_t) \right) \\
& \leq C_{H, \tilde{H}} \mathbb{E}(\rho_t) + C_{H, \tilde{H}}^0 \mathbb{E}(\rho_t) e^{-\gamma t},
\end{aligned}$$

with $C_{H, \tilde{H}}$ a constant independent of initial conditions and $C_{H, \tilde{H}}^0$ another constant, possibly depending on initial conditions. Here, we used (12) and (38).

We can thus construct a function f and a constant c , through the same calculations as before, such that there are C and C^0 constants (resp. independent and dependent on initial conditions) such that

$$\forall t, \quad e^{ct} \mathbb{E}(\rho_t) \leq \mathbb{E}(\rho_0) + \xi(1 + \alpha) \mathbb{E}(G_t) e^{ct} + L_W C \int_0^t e^{cs} \mathbb{E}(\rho_s) ds + L_W C^0 \int_0^t e^{(c-\gamma\lambda)s} \mathbb{E}(\rho_s) ds.$$

Since $\mathbb{E}G_t$ is bounded uniformly in time, we may now conclude using Gronwall's lemma.

Acknowledgements This work has been funded by the project EFI ANR-17-CE40-0030.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Bakry, D., Cattiaux, P., Guillin, A.: Rate of convergence for ergodic continuous markov processes: Lyapunov versus poincaré. *Journal of Functional Analysis* **254**(3), 727 – 759 (2008). DOI <https://doi.org/10.1016/j.jfa.2007.11.002>. URL <http://www.sciencedirect.com/science/article/pii/S0022123607004259>
2. Baudoin, F.: Bakry-Émery meet Villani. *J. Funct. Anal.* **273**(7), 2275–2291 (2017). DOI [10.1016/j.jfa.2017.06.021](https://doi.org/10.1016/j.jfa.2017.06.021). URL <https://doi.org/10.1016/j.jfa.2017.06.021>
3. Benedetto, D., Caglioti, E., Carrillo, J.A., Pulvirenti, M.: A non Maxwellian steady distribution for one-dimensional granular media. *J. Stat. Physics* **91**(5/6), 979–990 (1998)
4. Bolley, F.: Separability and completeness for the Wasserstein distance. *Séminaire de Probabilités XLI, Lecture notes in mathematics* **1934**, 371–377 (2008). URL <https://hal.archives-ouvertes.fr/hal-00453887>
5. Bolley, F., Gentil, I., Guillin, A.: Uniform convergence to equilibrium for granular media. *Arch. Ration. Mech. Anal.* **208**(2), 429–445 (2013)
6. Bolley, F., Guillin, A., Malrieu, F.: Trend to equilibrium and particle approximation for a weakly selfconsistent Vlasov-Fokker-Planck equation. *ESAIM, Math. Model. Numer. Anal.* **44**(5), 867–884 (2010)
7. Bouchut, F., Dolbeault, J.: 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)
8. Carrillo, J.A., McCann, R.J., Cédric, V.: Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates. *Rev. Mat. Iberoam.* **19**(3), 971–1018 (2003)
9. Cattiaux, P., Guillin, A., Malrieu, F.: Probabilistic approach for granular media equations in the non-uniformly convex case. *Probab. Theory Related Fields* **140**(1-2), 19–40 (2008). DOI [10.1007/s00440-007-0056-3](https://doi.org/10.1007/s00440-007-0056-3). URL <https://doi.org/10.1007/s00440-007-0056-3>
10. Cattiaux, P., Guillin, A., Monmarché, P., Zhang, C.: Entropic multipliers method for Langevin diffusion and weighted log Sobolev inequalities. *J. Funct. Anal.* **277**(11), 108288, 24 (2019). DOI [10.1016/j.jfa.2019.108288](https://doi.org/10.1016/j.jfa.2019.108288). URL <https://doi.org/10.1016/j.jfa.2019.108288>
11. Cheng, X., Chatterji, N.S., Abbasi-Yadkori, Y., Bartlett, P.L., Jordan, M.I.: Sharp convergence rates for Langevin dynamics in the nonconvex setting. *arXiv e-prints* [arXiv:1805.01648](https://arxiv.org/abs/1805.01648) (2018)
12. Cheng, X., Chatterji, N.S., Bartlett, P.L., Jordan, M.I.: Underdamped langevin mcmc: A non-asymptotic analysis. In: S. Bubeck, V. Perchet, P. Rigollet (eds.) *Proceedings of the 31st Conference On Learning Theory, Proceedings of Machine Learning Research*, vol. 75, pp. 300–323. PMLR (2018). URL <http://proceedings.mlr.press/v75/cheng18a.html>
13. Durmus, A., Eberle, A., Guillin, A., Zimmer, R.: An elementary approach to uniform in time propagation of chaos. *Proc. Amer. Math. Soc.* **148**(12), 5387–5398 (2020). DOI [10.1090/proc/14612](https://doi.org/10.1090/proc/14612). URL <https://doi.org/10.1090/proc/14612>
14. Eberle, A.: Reflection couplings and contraction rates for diffusions. *Probab. Theory Relat. Fields* **166**(3-4), 851–886 (2016)
15. Eberle, A., Guillin, A., Zimmer, R.: Couplings and quantitative contraction rates for Langevin dynamics. *Ann. Probab.* **47**(4), 1982–2010 (2019)
16. Eberle, A., Guillin, A., Zimmer, R.: Quantitative Harris-type theorems for diffusions and McKean-Vlasov processes. *Trans. Am. Math. Soc.* **371**(10), 7135–7173 (2019)
17. Funaki, T.: A certain class of diffusion processes associated with nonlinear parabolic equations. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **67**(3), 331–348 (1984). DOI [10.1007/BF00535008](https://doi.org/10.1007/BF00535008). URL <https://doi.org/10.1007/BF00535008>

18. Guillin, A., Liu, W., Wu, L., Zhang, C.: The kinetic fokker-planck equation with mean field interaction. To appear in *Journal de Mathématiques Pures et Appliquées*, arXiv preprint arXiv:1912.02594 (2019)
19. Guillin, A., Liu, W., Wu, L., Zhang, C.: Uniform poincaré, and logarithmic sobolev inequalities for mean field models. *Thal-02287711* (2019)
20. Guillin, A., Monmarché, P.: Optimal linear drift for the speed of convergence of an hypoelliptic diffusion. *Electron. Commun. Probab.* **21**, Paper No. 74, 14 (2016). DOI 10.1214/16-ECP25. URL <https://doi.org/10.1214/16-ECP25>
21. Guillin, A., Monmarché, P.: Uniform long-time and propagation of chaos estimates for mean field kinetic particles in non-convex landscapes. *Hal-02504451* (2020)
22. Hairer, M., Mattingly, J.C.: Spectral gaps in Wasserstein distances and the 2D stochastic Navier-Stokes equations. *Ann. Probab.* **36**(6), 2050–2091 (2008). DOI 10.1214/08-AOP392. URL <https://doi.org/10.1214/08-AOP392>
23. Herrmann, S., Tugaut, J.: Non-uniqueness of stationary measures for self-stabilizing processes. *Stochastic Process. Appl.* **120**(7), 1215–1246 (2010). DOI 10.1016/j.spa.2010.03.009. URL <https://doi.org/10.1016/j.spa.2010.03.009>
24. Kac, M.: Foundations of kinetic theory. In: *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 3: Contributions to Astronomy and Physics*, pp. 171–197. University of California Press, Berkeley, Calif. (1956). URL <https://projecteuclid.org/euclid.bsmmsp/1200502194>
25. Kazeykina, A., Ren, Z., Tan, X., Yang, J.: Ergodicity of the underdamped mean-field Langevin dynamics. arXiv e-prints arXiv:2007.14660 (2020)
26. Lemons, D.S., Gythiel, A.: Paul Langevin’s 1908 paper “On the Theory of Brownian Motion” [“Sur la théorie du mouvement brownien,” *C. R. Acad. Sci. (Paris)* 146, 530–533 (1908)]. *American Journal of Physics* **65**(11), 1079–1081 (1997). DOI 10.1119/1.18725
27. McKean, H.P.: A class of markov processes associated with nonlinear parabolic equations. *Proceedings of the National Academy of Sciences of the United States of America* **56**(6), 1907–1911 (1966). URL <http://www.jstor.org/stable/57643>
28. Méléard, S.: Asymptotic behaviour of some interacting particle systems; mckean-vlasov and boltzmann models. In: *Probabilistic models for nonlinear partial differential equations*, pp. 42–95. Springer (1996)
29. Monmarché, P.: Long-time behaviour and propagation of chaos for mean field kinetic particles. *Stochastic Process. Appl.* **127**(6), 1721–1737 (2017). DOI 10.1016/j.spa.2016.10.003. URL <https://doi.org/10.1016/j.spa.2016.10.003>
30. Monmarché, P.: Generalized Γ calculus and application to interacting particles on a graph. *Potential Anal.* **50**(3), 439–466 (2019). DOI 10.1007/s11118-018-9689-3. URL <https://doi.org/10.1007/s11118-018-9689-3>
31. Revuz, D., Yor, M.: *Continuous Martingales and Brownian Motion*, vol. 293, 3rd edn. Springer-Verlag Berlin Heidelberg (1999)
32. Salem, S.: A gradient flow approach of propagation of chaos. *Discrete Contin. Dyn. Syst.* **40**(10), 5729–5754 (2020). DOI 10.3934/dcds.2020243. URL <https://doi.org/10.3934/dcds.2020243>
33. Sznitman, A.S.: Topics in propagation of chaos. In: P.L. Hennequin (ed.) *Ecole d’Eté de Probabilités de Saint-Flour XIX — 1989*, pp. 165–251. Springer Berlin Heidelberg, Berlin, Heidelberg (1991)
34. Talay, D.: 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)
35. Villani, C.: Hypocoercivity, vol. 202. *Mem. Amer. Math. Soc* (2009)
36. Vlasov, A.A.: The vibrational properties of an electron gas. *Soviet Physics Uspekhi* **10**(6), 721–733 (1968). DOI 10.1070/pu1968v010n06abeh003709. URL <https://doi.org/10.1070/pu1968v010n06abeh003709>