Reducing exit-times of diffusions with repulsive interactions

Paul-Eric Chaudru de Raynal, Manh Hong Duong, Pierre Monmarché, Milica Tomasevic, Julian Tugaut

ESAIM: PS 27 (2023) 723–748
https://doi.org/10.1051/ps/2023012

ESAIM: Probability and Statistics
www.esaim-ps.org

REDUCING EXIT-TIMES OF DIFFUSIONS WITH REPULSIVE INTERACTIONS

Paul-Eric Chaudru de Raynal[1], Manh Hong Duong[2,*], Pierre Monmarché[3], Milica Tomašević[4] and Julian Tugaut[5]

Abstract. In this work we prove a Kramers' type law for the low-temperature behavior of the exit-times from a metastable state for a class of self-interacting nonlinear diffusion processes. Contrary to previous works, the interaction is not assumed to be convex, which means that this result covers cases where the exit-time for the interacting process is smaller than the exit-time for the associated non-interacting process. The technique of the proof is based on the fact that, under an appropriate contraction condition, the interacting process is conveniently coupled with a non-interacting (linear) Markov process where the interacting law is replaced by a constant Dirac mass at the fixed point of the deterministic zero-temperature process.

Mathematics Subject Classification. 60F10, 60J60, 60H10.

Received September 2, 2022. Accepted June 14, 2023.

1. Introduction and main result

1.1. Overview

Let $M \in \mathbb{R}^{d\times d}$, $a : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma \geq 0$. Consider the exit-time $\mathcal{T}_\sigma(\mathcal{D}) := \inf\{t \geqslant 0 \ : \ Y_t^\sigma \notin \mathcal{D}\}$ from a domain $\mathcal{D}$ of the diffusion process that solves

$$\mathrm{d}Y_t^\sigma \ = \ a(Y_t^\sigma)\mathrm{d}t + \sigma M \mathrm{d}B_t, \quad Y_0^\sigma := y_0 \in \mathcal{D}\,. \tag{1.1-non-interacting}$$

A standard example is the overdamped Langevin process, which corresponds to $a = -\nabla U$, for $U : \mathbb{R}^d \to \mathbb{R}$ a given potential and $M = \sqrt{2}I$. Besides, the general form (1.1-non-interacting) covers various other cases of interest, like the kinetic Langevin process or coloured noise processes, as detailed in Section 3.1 below.

Assume for a while that a confines the dynamics in $\mathcal{D}$ and admits a unique stable equilibrium λ_0 therein. Under suitable additional conditions on a, y_0 and $\mathcal{D}$, it is known that the exit-time from $\mathcal{D}$ satisfies a so-called

Keywords and phrases: Exit-time problem, large deviations, self-interacting nonlinear diffusion, Kramer's law.

[1] Nantes Université, CNRS, Laboratoire de Mathématiques Jean Leray, LMJL, 44000 Nantes, France.
[2] School of Mathematics, University of Birmingham, B15 2TT Birmingham, UK.
[3] LJLL and LCT, Sorbonne Université, 4 place Jussieu, 75005 Paris, France.
[4] CMAP, CNRS, École polytechnique, Institut Polytechnique de Paris, 91120 Palaiseau, France.
[5] Université Jean Monnet, Institut Camille Jordan, 23, rue du docteur Paul Michelon, CS 82301, 42023 Saint-Étienne Cedex 2, France.

* Corresponding author: h.duong@bham.ac.uk

Kramers' type law: there exists $L > 0$ (given as the solution of a variational problem) such that, for all $\delta > 0$,

$$\lim_{\sigma\to 0} \mathbb{P}\left\{\exp\left(\frac{2}{\sigma^2}(L-\delta)\right) \leq \mathcal{T}_\sigma(\mathcal{D}) \leq \exp\left(\frac{2}{\sigma^2}(L+\delta)\right)\right\} = 1\,. \tag{1.2}$$

This estimate is strongly related to Large Deviations Principles (LDP). We briefly explain why. The underlying idea of the LDP is that the exit-time from $\mathcal{D}$ is a rare event. Indeed, as the drift tends to bring the process near λ_0, only the Brownian motion allows the process to leave the domain $\mathcal{D}$. As a consequence, when σ is small, the probability that the exit from $\mathcal{D}$ occurs before a given time $T > 0$ is small, namely, of the order e^{-2L/σ^2}, where L depends on a and $\mathcal{D}$. This somehow allows to obtain that the exit-time follows some exponential law (see [5]), so that (1.2) holds.

Let us just mention that under easy to check assumptions, if a is of gradient form $-\nabla U$ and if $M = I$, then the so-called exit-cost L is equal to $\inf_{\partial\mathcal{D}} U - \inf_{\mathcal{D}} U$ that corresponds to the difference between the potential at the boundary of $\mathcal{D}$ and its minimum in $\mathcal{D}$.

In the more general case where a is not of gradient form and where M is possibly degenerate, L is the infimum of the quasi-potential on the boundary of $\mathcal{D}$ namely

$$L = \inf_{y\in\partial\mathcal{D}} \inf_{t>0} \inf_{u} \frac{1}{4}\int_0^t |u(s)|^2 \mathrm{d}s,$$

where the last infimum runs over $u \in \mathrm{L}^2([0,t])$ such that $z_t = y$, the process z being the solution of $z_s = \lambda_0 + \int_0^s (a(z_w) + Mu_w)\mathrm{d}w$.

In this work we aim at establishing a similar result when (1.1-non-interacting) is replaced by some inhomogeneous processes, namely processes whose evolution at time t depends on a probability law μ_t^σ. Denoting by m_t^σ the law of the process at time t, the most classical case would be $\mu_t^\sigma = m_t^\sigma$, which gives a classical non-linear McKean-Vlasov diffusion. We have also in mind the case of so-called memorial McKean-Vlasov processes, where $\mu_t^\sigma = t^{-1}\int_0^t m_s^\sigma \mathrm{d}s$. A convenient way to gather these two examples, and more generally memorial processes with a non-uniform memory kernel which can be motivated by applications in stochastic algorithms, is to consider $\mu_t^\sigma = \int_0^t m_s^\sigma R(t,\mathrm{d}s)$ where, for all $t \geqslant 0$, $R(t,\cdot)$ is a probability measure on $[0,t]$. The classical non-linear case then corresponds to $R(t,\cdot) = \delta_t$ for all $t \geqslant 0$, and the memorial process to the case where $R(t,\cdot)$ is the uniform law on $[0,t]$ for all $t \geqslant 0$.

Let $\mathcal{P}_2(\mathbb{R}^d)$ be the set of probability measures on $\mathbb{R}^d$ with a finite second moment and $\mathcal{P}(\mathbb{R})$ the set of probability measures on $\mathbb{R}$. Given $b : \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^d$, an initial condition $m_0^\sigma \in \mathcal{P}_2(\mathbb{R}^d)$ and $R : \mathbb{R} \to \mathcal{P}(\mathbb{R})$, we are thus interested in a process $(X_t^\sigma)_{t\geqslant 0}$ with $X_0^\sigma \sim m_0^\sigma$ and such that, for all $t \geqslant 0$,

$$\left\{\begin{array}{rcll} \mathrm{d}X_t^\sigma &=& a(X_t^\sigma)\mathrm{d}t + b(X_t^\sigma,\mu_t^\sigma)\mathrm{d}t + \sigma M \mathrm{d}B_t, & \\ \mu_t^\sigma &:=& \int_0^t m_s^\sigma R(t,\mathrm{d}s) & m_t^\sigma := \mathcal{L}(X_t^\sigma). \end{array}\right. \tag{1.3-McKean-interaction}$$

Conditions on a, b and R that ensures the existence and uniqueness in distribution of such a process will be discussed below. The main subject of this work is the exit-time

$$\tau_\sigma(\mathcal{D}) := \inf\{t \geq 0 \ : \ X_t^\sigma \notin \mathcal{D}\} \tag{1.4}$$

of the above process from a given domain $\mathcal{D}$ on $\mathbb{R}^d$. We aim at proving a Kramers' law

$$\forall \delta > 0, \qquad \lim_{\sigma\to 0} \mathbb{P}\left\{\exp\left(\frac{2}{\sigma^2}(H-\delta)\right) \leq \tau_\sigma(\mathcal{D}) \leq \exp\left(\frac{2}{\sigma^2}(H+\delta)\right)\right\} = 1, \tag{1.5}$$

for some $H > 0$.

More precisely, our main contribution is Theorem 1.3 where we establish (1.5) under conditions that cover cases where $H < L$ (with L given in (1.2)), which means that the exit-time is shorter for the interacting process than for the non perturbed diffusion (1.1-non-interacting), contrary to the similar works [24, 25] which are restricted to convex interactions for which $H > L$. This is an important improvement since the interacting processes can then be viewed as a starting point for designing and analyzing efficient stochastic algorithms, see Section 1.4 for further details on the motivation.

Our general strategy to establish (1.5) is to prove, under some conditions (see next section), the following. First, at a fixed σ, μ_t^σ converges in large time to an equilibrium μ_∞^σ, at a speed *that is uniform in* σ (for the Wasserstein distance $\mathbb{W}_2$, see below). Second, as σ vanishes, μ_∞^σ converges to δ_λ for some $\lambda \in \mathbb{R}^d$. Hence, the interacting process (1.3-McKean-interaction) is expected to behave similarly to the *linear* (in the McKean–Vlasov sense) diffusion that solves

$$\mathrm{d}\tilde{X}_t^\sigma \;=\; a(\tilde{X}_t^\sigma)\mathrm{d}t + b(\tilde{X}_t^\sigma, \delta_\lambda)\mathrm{d}t + \sigma M \mathrm{d}B_t, \qquad \text{(1.6–equilibrium-interaction)}$$

for which the Kramers' law follows from classical results. In fact, more precisely, we can consider the two equations (1.3-McKean-interaction) and (1.6–equilibrium-interaction) simultaneously, driven by the same Brownian motion $(B_t)_{t\geqslant 0}$. A crucial point is then to prove that, at low temperature, the two processes will *deterministically* stay close one to the other, so that the exit of one of the process from an enlargement of $\mathcal{D}$ necessarily implies the exit of the other from $\mathcal{D}$.

1.2. Assumptions and main results

We divide the assumptions in three groups: the first group concerns the coefficients of the dynamics (1.3-McKean-interaction), ensuring in particular the well-posedness of the process. In the second group we gather basic conditions on the domain and the initial condition. The third one states a Kramers' law for a linear process of the form (1.6–equilibrium-interaction).

We start by introducing the conditions **(A)** that concern the drift functions a and b and the memory kernel R. Let us point out that we choose to split a and b despite the function a could be part of the function b. In fact, it is due to our initial aim to reduce the exit-cost (so of the exit-time) by adding a nonlinearity with the function b.

(A1) The function $a : \mathbb{R}^d \to \mathbb{R}^d$ (resp. $b : \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^d$) is locally (resp. globally) Lipschitz continuous.

(A2) There exist $\rho > \kappa \geqslant 0$ such that for all $z, y \in \mathbb{R}^d$ and all $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$,

$$(z-y)\cdot(a(z)+b(z,\mu)-a(y)-b(y,\nu)) \leqslant -\rho|z-y|^2 + \kappa \mathbb{W}_2^2(\nu,\mu)\,. \tag{1.7}$$

(A3) For all $s \geqslant 0$, $R(t,[0,s]) \to 0$ as $t \to +\infty$ and for all continuous $f : \mathbb{R} \to \mathbb{R}$, $t \mapsto \int_0^\infty f(s)R(t,\mathrm{d}s) = \int_0^t f(s)R(t,\mathrm{d}s)$ (providing that $R(t,\mathrm{d}s)$ is a measure on $[0,t]$) is measurable.

Let us point out that $R(t,\mathrm{d}s) = \delta_t(\mathrm{d}s)$ or $R(t,\mathrm{d}s) = \frac{1}{t}\mathbb{1}_{[0;t]}(s)\mathrm{d}s$ satisfy Assumption **(A3)**. Moreover, **(A2)** is true if a satisfies (1.7) (applied with $b = 0$, with $\kappa = 0$) and the Lipschitz constants of b, namely $\kappa_1, \kappa_2 \geqslant 0$ such that

$$|b(z,\mu) - b(y,\nu)| \leqslant \kappa_1|z-y| + \kappa_2 \mathbb{W}_2(\mu,\nu) \tag{1.8}$$

for all $z, y \in \mathbb{R}^d$ and all $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, satisfy $\kappa_1 + \kappa_2 < \rho$. However, **(A2)** is weaker when the interaction is attracting, for instance if $b(x,\nu) = \alpha(\int_{\mathbb{R}^d} y\nu(\mathrm{d}y) - x)$ with $\alpha > 0$.

Let us deduce a few preliminary results from these first assumptions. First, we check that the interacting process is indeed well defined under **(A)**.

Proposition 1.1. *Under Assumption* **(A)**, *the system* (1.3-McKean-interaction) *admits a unique weak solution.*

The proof is postponed to Section 2.2.

Next, in order to state the conditions on the domain $\mathcal{D}$, we need the following lemma.

Lemma 1.2. *Under* **(A)**, *there exists a unique* $\lambda \in \mathbb{R}^d$ *such that*

$$a(\lambda) + b(\lambda, \delta_\lambda) \;=\; 0\,. \tag{1.9}$$

The proof is postponed (see Lem. 2.2). Notice that (1.9) is equivalent to saying that the process given by $X_t = \lambda$ for all $t \geqslant 0$ is a constant solution of (1.3-McKean-interaction) in the zero noise case $\sigma = 0$. In all the rest of this section, we fix λ as given by Lemma 1.2.

The basic conditions **(D)** on the domain $\mathcal{D}$ are the following.

(D1) The domain $\mathcal{D} \subsetneq \mathbb{R}^d$ is open.
(D2) Moreover a is Lipschitz continuous on $\mathcal{D}_{+1} = \{z \in \mathbb{R}^d,\ d(z, \mathcal{D}) \leqslant 1\}$.
(D3) The initial distribution $m_0^\sigma = m_0$ is independent from σ and has a compact support included in a ball centered at λ (given in Lem. 1.2) included in $\mathcal{D}$.

When $\mathcal{D}$ is bounded, of course **(D2)** is implied by **(A1)**, but there are cases of interest (as in the kinetic case, see Sect. 3.1.2) where $\mathcal{D}$ is not bounded.

As we will see (see Lem. 2.10), **(D3)** implies that the (deterministic) interacting process at zero temperature stays in $\mathcal{D}$, which is clearly a basic requirement to get a Kramers' law. Besides, this condition could be slightly weakened, see Remark 2.11.

The last assumption is that a Kramers' law holds true for the linear diffusion $\tilde{X}^\sigma$ solving equation (1.6–equilibrium-interaction) starting from λ, and the associated rate $H(\mathcal{D})$ is continuous with respect to $\mathcal{D}$ for monotonously and uniformly converging domains, in the following precise sense:

(K) Let $\tilde{X}^\sigma$ solve (1.6–equilibrium-interaction) with $\tilde{X}_0^\sigma \sim \delta_\lambda$. There exist two families of open domains $(\mathcal{D}_{i,\xi})_{\xi>0}$ and $(\mathcal{D}_{e,\xi})_{\xi>0}$ with $\mathcal{D}_{i,\xi} \subset \mathcal{D}_{i,0} = \mathcal{D} = \mathcal{D}_{e,0} \subset \mathcal{D}_{e,\xi}$ with the following properties. First,

$$\sup_{z \in \partial \mathcal{D}_{i,\xi}} \mathrm{d}\left(z\,;\, \mathcal{D}^c\right) + \sup_{z \in \partial \mathcal{D}_{e,\xi}} \mathrm{d}\left(z\,;\, \mathcal{D}\right) \underset{\xi \to 0}{\longrightarrow} 0.$$

Second, for all $\xi > 0$ small enough

$$\inf_{z \in \partial \mathcal{D}_{i,\xi}} \mathrm{d}\left(z\,;\, \mathcal{D}^c\right) > \xi \text{ and } \inf_{z \in \partial \mathcal{D}_{e,\xi}} \mathrm{d}\left(z\,;\, \mathcal{D}\right) > \xi.$$

Third, let $\tilde{\tau}_\sigma(\mathcal{D}_{u,\xi}) = \inf\{t \geqslant 0, \tilde{X}_t^\sigma \notin \mathcal{D}_{u,\xi}\}$ for $\xi \geq 0$ and $u \in \{e, i\}$. For all $\xi \geq 0$ and $u \in \{e, i\}$, there exists $H_{u,\xi} > 0$ such that it holds for all $\delta > 0$:

$$\mathbb{P}\left(\exp\left(\frac{2}{\sigma^2}\left(H_{u,\xi} - \delta\right)\right) \leqslant \tilde{\tau}_\sigma(\mathcal{D}_{u,\xi}) \leqslant \exp\left(\frac{2}{\sigma^2}\left(H_{u,\xi} + \delta\right)\right)\right) \underset{\sigma \to 0}{\longrightarrow} 1\,. \tag{1.10}$$

Moreover, for $u \in \{i, e\}$, $\lim_{\xi \to 0} H_{u,\xi} = H_{u,0} =: H$.

Since (1.6–equilibrium-interaction) is a standard (time-homogenous) diffusion process, Assumption **(K)** can be checked by applying standard Large Deviations results on processes (Schilder theorem, contraction principle and so Freidlin-Wentzell theory) in the small-noise limit, see [6]. We presented it as a black-box assumption for clarity (in particular to avoid a discussion on the characteristic boundary in non-elliptic cases) and refer to Section 3.3 for more details.

Let us just mention that if a is of gradient form $-\nabla U$, if $b(\cdot, \mu) = -\nabla_x F(\cdot, \mu)$ and if $M = I$ then the exit-cost H is equal to $\inf_{\partial \mathcal{D}} U(\cdot) + F(\cdot, \delta_\lambda) - U(\lambda) - F(\lambda, \delta_\lambda)$ where λ is given in Lemma 1.2.

In the more general case where the drifts are not of gradient form and where M is possibly degenerate, H is the infimum of the quasi-potential on the boundary of $\mathcal{D}$ that means

$$H = \inf_{y\in\mathcal{D}} \inf_{t>0} \inf_{u} \frac{1}{4}\int_0^t |u(s)|^2 \mathrm{d}s,$$

where the last infimum runs over $u \in \mathrm{L}^2([0,t])$ such that $z_t = y$, the process z being the solution of $z_s = \lambda + \int_0^s (a(z_w) + b(z_w, \delta_\lambda) + Mu_w)\mathrm{d}w$.

We can now state our main result.

Theorem 1.3. *Under Assumptions* **(A)**, **(D)** *and* **(K)**, *for all* $\delta > 0$,

$$\mathbb{P}\left(\exp\left(\frac{2}{\sigma^2}(H-\delta)\right) \leqslant \tau_\sigma(\mathcal{D}) \leqslant \exp\left(\frac{2}{\sigma^2}(H+\delta)\right)\right) \underset{\sigma\to 0}{\longrightarrow} 1\,.$$

In other words, the Kramers' law holds for the non-linear process (1.3-McKean-interaction), with the same rate H as the linear process (1.6–equilibrium-interaction).

Organization of the paper. The rest of the paper is organized as follows. We develop a first example in Section 1.3 and we give the algorithmic motivation in Section 1.4 then we conclude this introduction by a discussion of this result in Section 1.5. Section 2 is devoted to its proof. Examples of applications are provided in Section 3.

1.3. First example

To illustrate Theorem 1.3 and fix some ideas, consider the case of the overdamped Langevin process (see Sect. 3 for other applications). Let $U, W \in \mathcal{C}^2(\mathbb{R}^d)$ and, for $z \in \mathbb{R}^d$ and $\mu \in \mathcal{P}_2(\mathbb{R}^d)$,

$$a(z) = -\nabla U(z) \qquad b(z,\mu) = \int_{\mathbb{R}^d} \nabla W(z-y)\mu(\mathrm{d}y)\,.$$

Assume that U is ρ-strongly convex for some $\rho > 0$, that $W(y) = W(-y)$ for all $y \in \mathbb{R}^d$ and that $\nabla^2 W$ is bounded. Then a satisfies (1.7) (applied with $b = 0$) and, considering π a coupling of ν and μ,

$$\begin{aligned} |b(z,\nu) - b(y,\mu)| &= \left|\int_{\mathbb{R}^d\times\mathbb{R}^d} (\nabla W(z-v) - \nabla W(y-w))\,\pi(\mathrm{d}v,\mathrm{d}w)\right| \\ &\leqslant \|\nabla^2 W\|_\infty \left(|z-y| + \int_{\mathbb{R}^d} |v-w|\pi(\mathrm{d}v,\mathrm{d}w)\right), \end{aligned}$$

using the Jensen inequality and taking the infimum over all couplings yields (1.8) with $\kappa_1 = \kappa_2 = \|\nabla^2 W\|_\infty$. Hence, **(A2)** holds if $\|\nabla^2 W\|_\infty < \rho/2$.

Denote by λ the unique point where U attains its minimum. For $v \in \mathbb{R}^d$,

$$a(\cdot) + b(\cdot, \delta_v) = -\nabla U_v$$

with $U_v = U - W(\cdot - v)$. Remark that $\nabla^2 U_v \geqslant \rho/2 > 0$, so that U_v is convex, and moreover $\nabla U_\lambda(\lambda) = 0$. It means that, in Theorem 1.3, the linear process (1.6–equilibrium-interaction) reads

$$\mathrm{d}Z_t = -\nabla U_\lambda(Z_t) + \sigma \mathrm{d}B_t\,.$$

Take $\mathcal{D} = \mathbb{B}(\lambda, r)$ for some $r > 0$, so that **(D)** holds for all initial distribution with support included in $\mathcal{D}$. Setting $\mathcal{D}_{e,\xi} = \mathbb{B}(\lambda, r+\xi)$ and $\mathcal{D}_{i,\xi} = \mathbb{B}(\lambda, r-\xi)$ for $\xi < r$, **(K)** holds with

$$H_{u,\xi} = \inf_{x \in \partial \mathcal{D}_{u,\xi}} U_\lambda(x) - U(\lambda), \qquad H = \inf_{x \in \partial \mathcal{D}} U_\lambda(x) - U(\lambda),$$

see Section 3.3. Similarly, the rate for the initial linear process (1.1-non-interacting) is

$$L = \inf_{x \in \partial \mathcal{D}} U(x) - U(\lambda)\,.$$

As a consequence, as announced, our assumptions allow for situations where $H < L$, which is here the case as soon as $\inf_{x \in \partial \mathcal{D}} W(x-\lambda) > 0$.

For instance, if $U(z) = \rho/2|z-\lambda|^2$ and $W(z) = \alpha|z|^2$, then **(A2)** holds whenever $\alpha < \rho/4$, and $U_\lambda = (\rho/2 - \alpha)|z-\lambda|^2$. In that case, $L = \rho/2 r^2$ and $H = (\rho/2 - \alpha) r^2$, which can be made arbitrarily close to $L/2$ by taking α arbitrarily close to $\rho/4$.

Remark 1.4. More generally, we do not need to assume that $z \mapsto b(z, \delta_\lambda)$ derives from a potential. Considering the variational definition of H (see Sect. 3.3), we can see that the hypothesis $\langle b(z,\delta_\lambda), b(z,\delta_\lambda) - 4\nabla U(z)\rangle < 0$ for any $z \in \mathbb{R}^d$ is sufficient to ensure that the exit-cost is reduced. The idea is to consider the optimal trajectory for the linear process (1.1-non-interacting) and to show that, with the interaction, the cost of this trajectory gets strictly smaller than L.

1.4. An algorithmic motivation

In Monte Carlo Markov chain (MCMC) methods, high-dimensional expectations with respect to a given target probability distribution π, say on $\mathbb{R}^d$ with a Lebesgue density proportional to $\exp(-V)$ for some potential V, are estimated thanks to an ergodic law of large numbers:

$$\frac{1}{t}\int_0^t \varphi(X_s)\mathrm{d}s \underset{t\to+\infty}{\longrightarrow} \int_{\mathbb{R}^d} \varphi(x)\pi(\mathrm{d}x), \tag{1.11}$$

where $(X_t)_{t\geqslant 0}$ is a Markov process designed to be ergodic with respect to π, and φ is some observable of interest. A classical example is the overdamped Langevin diffusion

$$\mathrm{d}X_t \;=\; -\nabla V(X_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t,$$

where $(B_t)_{t\geqslant 0}$ is a standard Brownian motion, see *e.g.* [19]. The estimation given by the convergence (1.11) is reasonable if t is sufficiently large so that the process has visited during $[0,t]$ a representative sample (with respect to π) of the state space. In particular, if π is multimodal, *i.e.* if V has several local minima, then t should be at least large enough so that some transitions have occurred between the basins of attraction of the main minima (where *main* means: in term of contribution to the expectation). In the theoretical studies of this question, the difficulty of the problem is measured by the addition of a temperature parameter $\varepsilon > 0$, the target density being proportional to $\exp(-V/\varepsilon)$, so that the multimodality worsens as ε vanishes (*i.e.* at low temperature). The corresponding overdamped Langevin process is then

$$\mathrm{d}X_t \;=\; -\nabla V(X_t)\mathrm{d}t + \sqrt{2\varepsilon}\mathrm{d}B_t\,. \tag{1.12}$$

As ε vanishes, this is essentially a gradient descent, which means it converges quickly to a local minimum, and then has to wait a large deviation of the small Brownian noise to escape to another basin. As a consequence, transitions take a time that is exponentially large with ε^{-1}, which makes the convergence (1.11) very slow. This is a so-called metastable behaviour [16]. What is true for the overdamped Langevin diffusion is in fact generic

for all classical samplers: indeed, in practice, we only have a local knowledge of the energy landscape, which prevents the design of Markov processes that perform large jumps. Standard samplers have thus continuous trajectories, or perform small jumps (the exploration is local). This implies that, in practice, during a transition from one basin to another, the time spent in the low-probability area in between is lower bounded uniformly with ε. But then the ergodic property (1.11) implies that the ratio of the times spent in two areas tends to the ratio of the probabilities of those. As a consequence, we get that, roughly speaking, at low temperature, for a Markov process ergodic with respect to π and performing a local exploration, the time between transitions is in some sense necessarily of order at least $\exp(c/\varepsilon)$ where c is the difference of energy between the local minima and the lowest saddle point in between.

Besides, this issue of metastability in the exploration of a high-dimensional non-convex energy landscape also arises for optimization problems. Indeed, consider the process

$$\mathrm{d}X_t \;=\; -\nabla V(X_t)\mathrm{d}t + \sqrt{2\varepsilon_t}\mathrm{d}B_t,$$

where $t \mapsto \varepsilon_t$ is a decreasing vanishing map called the cooling schedule. This is a simple theoretic simulated annealing process. It is well-known that, provided ε_t vanishes sufficiently slowly with t, then the process converges in probability to the global minima of V [12]. The picture is very similar to the question of the convergence of an ergodic mean at constant (small) temperature: indeed, the process needs sufficient time to cross energy barriers and find global minima. In both cases (sampling and optimization), the real problem is exploration (discovering the main basins of attraction).

As we saw, at low temperature, exits from local minima are rare events. A general method to tackle rare events issues is importance sampling: the target π is replaced by a biased target $\tilde{\pi} \propto \exp(-(V-A)/\varepsilon)$ for some biasing potential A, in such a way the biased process

$$\mathrm{d}X_t \;=\; -\nabla(V-A)(X_t)\mathrm{d}t + \sqrt{2\varepsilon}\mathrm{d}B_t,$$

is less metastable than the initial process, so that the convergence

$$\frac{\int_0^t \varphi(X_s)e^{-A(X_s)/\varepsilon}\mathrm{d}s}{\int_0^t e^{-A(X_s)/\varepsilon}\mathrm{d}s} = \frac{t^{-1}\int_0^t \varphi(X_s)e^{-A(X_s)/\varepsilon}\mathrm{d}s}{t^{-1}\int_0^t e^{-A(X_s)/\varepsilon}\mathrm{d}s} \underset{t\to+\infty}{\longrightarrow} \frac{\int_{\mathbb{R}^d}\varphi(x)e^{-A(x)/\varepsilon}\tilde{\pi}(\mathrm{d}x)}{\int_{\mathbb{R}^d} e^{-A(x)/\varepsilon}\tilde{\pi}(\mathrm{d}x)} = \int_{\mathbb{R}^d}\varphi\pi$$

is faster than (1.11) ($\tilde{\pi}$ should still reasonably close to π, otherwise the weights e^{-A} in the estimator induce a large variance). Designing a good biasing potential A is a difficult question. Starting from a given local minimum a of V, a simple idea would be to take $A(x) = -\alpha|x-a|^2$ for some $\alpha > 0$, which would tend to repel the particle away from a. But this requires the knowledge of a; and then when the process has moved to another basin, another local minimum would have to be considered.

For this reason, in fact, as far as metastability is concerned, many strategies are based on *adaptive* biasing potentials, *i.e.* ∇A is not fixed a priori but evolves in time, depending of some current knowledge of the energy landscape. The basic idea is the following: if the process has already spent a long time in some area, then we should add a repulsion force from this area to accelerate the escape. In many cases (see *e.g.* [3, 9, 14, 17, 18] and references within) these adaptive biasing forces can be viewed as interactions with some probability law, *i.e.* the process is

$$\mathrm{d}X_t \;=\; -\nabla(V-A_{\mu_t})(X_t)\mathrm{d}t + \sqrt{2\varepsilon}\mathrm{d}B_t, \tag{1.13}$$

where μ_t is a probability measure: typically, the occupation measure of the past trajectory $(X_s)_{s\in[0,t]}$ as in [3], or the empirical measure of a system of N particles. If particles are repelled one from the other, the system will cover a larger area, enhancing the exploration. As N goes to infinity, according to the propagation of chaos

phenomenon, the particles become independent and the empirical measure of the system converges to the law of one of the particle, which leads to non-linear processes X that solves (1.13) with μ_t the law of X_t.

This leads to the question addressed in the present work, that is the question of reducing, through self-interaction, exit-times from the vicinity of local minima of V at low temperature.

1.5. Discussion on the result

In view of the motivations described in Section 1.4, the example discussed in Section 1.3 calls for a few remarks. In this example, we assumed that U is convex, which means that in fact the non-perturbed overdamped Langevin process is not metastable and the importance sampling scheme presented in Section 1.4 is not relevant. More generally, we see that the global contraction property **(A2)** is very strong and rules out any interesting practical case for the adaptive algorithms. In fact, Theorem 1.3 has to be understood as a first step toward proving a similar result where the convexity of U is only assumed locally on a neighborhood of λ. Indeed, if the initial distribution of the process is supported on a neighborhood of λ then the probability that the process exits in a time smaller than $e^{2(H-\delta)/\sigma^2}$ for some $\delta > 0$ will be small (and so will be the mass of m_t^σ far from λ) which means that only the local behavior of a is relevant up to these times.

In other words, with Theorem 1.3, we do not really prove that the metastability is reduced for a multi-modal probability target, which is the final goal. We only prove that, for a process attracted to a single point, the time to exit from a ball centered at the attractor can be reduced by adding a repelling interaction potential. This is still a new and far from trivial result and, as mentioned above, it should be possible to combine it with a localization procedure in order to treat a non-convex case. This will be the topic of a future work. More precisely, we expect the localization argument to enable the study of the case where U is not convex but $\mathcal{D}$ is still a domain on which U is convex. Going beyond this assumption will raise additional difficulties.

Let us make another remark on the local character of our result. In Theorem 1.3, the initial condition m_0 is concentrated on the domain $\mathcal{D}$, which in the example of the current section is a ball centered at the minimum of U. Now, consider the practical situation of a system of N interacting particles in a non-convex potential U. It is possible that, at some point, half the particles are in some well of U, and the other half is in another one. This is precisely why it is more natural to use an adaptive algorithm rather than a biasing potential $W(\cdot - \lambda)$ for a fixed λ. Using an interaction potential $W(y) = \alpha|y|^2$ would have the effect that the particles in the first well are repelled by the particles in the second one, even if the two wells are far away one from the other. Yet, there is no particular reason to believe that an efficient way to escape from the first well is to go in the direction opposite to the second one. In fact, in practice, localized repelling interactions are used, like $W(y) = -\alpha \exp(-\beta|y|^2)$ where both α and β are positive constants, as in the metadynamics algorithm [1, 13, 14]. It means that the effect of the particles in a given well on those in another well is negligible. In that case, Theorem 1.3 can be understood as a simplification of the problem, where each well is treated separately (which is reasonable and could possibly be made rigorous for domains $\mathcal{D}$ such that the exit-time is small with respect to the transition time from one well to another). A statement about a general situation with a non-localized interaction would be harder to interpret than the simple situation of a single well as in Theorem 1.3, which is why we focus on the latter in this work.

To be more precise, notice that, in a convex potential, even if a particle exits at some point from a ball centered at the stable point, it will fall back shortly after, since there is nowhere else to go. The mass around the stable point is thus constant. This is not at all the case if a particle which exits $\mathcal{D}$ falls in the basin of attraction of a different stable point: there is a mass leakage, which may diminish the strength of the repulsion, and thus slow down the exit of the remaining particles. There may be cases where the proof of Theorem 1.3 still works and yields a Kramers' law similar to the linear process with interaction $b(\cdot, p\delta_{\lambda_1} + (1-p)\delta_{\lambda_2})$ where λ_1 and λ_2 are the local minima in each well and p is the initial proportion of particles in the first well, but such a result can only hold if the transition from one well to another happens at a time much larger than the exit of $\mathcal{D}$ (which can be for instance the union of two balls centered at λ_1 and λ_2), in which case all particles will typically exit $\mathcal{D}$ before any transition from one well to the other is observed. From the point of view of numerical acceleration (where one indeed wants to accelerate the transitions), this is still not the interesting

framework. However, our strategy of proof could lead, in the case where exits of $\mathcal{D}$ correspond to transitions between different metastable states, to the following weaker result:

$$\forall \delta > 0, \qquad \liminf_{\sigma \to 0} \mathbb{P}\left(\tau_\sigma(\mathcal{D}) < e^{(H+\delta)/\sigma^2}\right) \; > \; 0,$$

where H is the rate of an explicit linear process (depending on the initial law).

For instance, let us consider a toy problem that gives a simplified picture of a system in $\mathbb{R}^d$ with two minima where the first particles to cross slow down the others, which gives a behaviour different than a simple Kramers' law with a modified height. Consider the Markov chain on $\{0,1\}$ with rates $\lambda(i \to j) = e^{-a_{ij}/\sigma^2}$ for some $a_{01}, a_{10} > 0$, with $\mathcal{D} = \{0\}$ and the initial distribution $m_0 = \delta_0$. Then $\tau_\sigma(\mathcal{D})$ is a geometric law with parameter e^{-a_{01}/σ^2}. Now, add some self-repulsion by considering the chain with rates $\lambda(i \to j) = e^{-b_{ij}(m_t^\sigma)/\sigma^2}$ where $b_{ij}(\nu) = a_{ij} + \alpha(\nu(j) - \nu(i))$ for some $\alpha > 0$. Then the law of $\tau_\sigma(\mathcal{D})$ is given by

$$\mathbb{P}\left(\tau_\sigma(\mathcal{D}) > t\right) \; = \; \exp\left(-\int_0^t e^{(\alpha(2x_s - 1) - a_{01})/\sigma^2} \mathrm{d}s\right)$$

where $x_t = m_t^\sigma(0)$ solves

$$\dot{x}_t \; = \; -e^{(\alpha(2x_t - 1) - a_{10})/\sigma^2} x_t + e^{(\alpha(1 - 2x_t) - a_{01})/\sigma^2}(1 - x_t)\,.$$

Suppose to fix ideas that $a_{01} = a_{10} = a$, so that $x_t \to 1/2$ as $t \to +\infty$ and $x_t \geqslant 1/2$ for all $t \geqslant 0$. Suppose also that $\alpha < a$. Then it is not difficult to see that for all $H \in [a - \alpha, a]$ and all $\delta > 0$,

$$\liminf_{\sigma \to 0} \mathbb{P}\left(e^{(H-\delta)/\sigma^2} < \tau_\sigma(\mathcal{D}) < e^{(H+\delta)/\sigma^2}\right) \; > \; 0\,.$$

Related works. The existing results devoted to a quantitative measure of the efficiency of adaptive algorithms are the following: in [17] the efficiency is expressed in term of a quantitative long-time convergence speed toward equilibrium (for a process interacting with its law). In [2, 8], for processes interacting with their occupation measures, it is written in term of the asymptotic variance in a CLT. To our knowledge, the only study concerning the exit times of such processes is conducted in [9], where the Wang-Landau algorithm is studied for the toy problem of a 3-states Markov chain (in which case the authors are able to establish that the exit times are sub-exponential in $1/\varepsilon$, which is not the case in our work).

Besides, concerning more generally the question of exit times for non-linear processes, we already mentioned [24, 25], where a result similar to Theorem 1.3 is established, but only for the usual elliptic McKean-Vlasov diffusion and in cases where the interaction is convex (and in particular the exit time is *larger* for the interacting process than for the initial dynamics). The general strategy of our proof is in the same spirit as the one of [25].

2. Proofs

In order to highlight the main steps of the proof of Theorem 1.3, this section is organised as follows. First, we present in Section 2.1 a general result based on the parallel coupling of two diffusion processes, which will be intensively used in all the remainder of the proof. The existence and uniqueness of the process (1.3-McKean-interaction) is addressed in Section 2.2. The long-time convergence toward equilibrium (at a speed that does not depend on the temperature) is established in Section 2.3, while the low noise asymptotics (both at equilibrium and in finite time intervals) is analysed in Section 2.4. Finally, building upon all these intermediary results, the proof of Theorem 1.3 is given in Section 2.5.

2.1. Parallel coupling

We consider in this section a general framework. First, let $a \in \mathcal{C}^1(\mathbb{R}^d, \mathbb{R}^d)$, $b \in \mathcal{C}^0(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d), \mathbb{R}^d)$, $M \in \mathbb{R}^{d\times d}$ and $(B_t)_{t\geqslant 0}$ be a standard Brownian motion on $\mathbb{R}^d$. In all Section 2.1, these parameters are fixed.

Second, let $\sigma \geqslant 0$, $m_0 \in \mathcal{P}_2(\mathbb{R}^d)$, $Y_0 \sim m_0$ and let $\nu = (\nu_t)_{t\geqslant 0} \in \mathcal{C}^0(\mathbb{R}_+, \mathcal{P}_2(\mathbb{R}^d))$. We say that $(Z_t)_{t\geqslant 0}$ is a process associated to Y_0, σ and ν if, almost surely, for all $t \geqslant 0$,

$$Z_t = Y_0 + \sigma M B_t + \int_0^t (a(Z_s) + b(Z_s, \nu_s))\,\mathrm{d}s. \tag{2.1}$$

Proposition 2.1. *Assume* **(A1)** *and* **(A2)**. *Let* $\sigma, \tilde\sigma \geqslant 0$, $m_0, \tilde m_0 \in \mathcal{P}_2(\mathbb{R}^d)$, $Y_0 \sim m_0$, $\tilde Y_0 \sim \tilde m_0$, $\nu, \tilde\nu \in \mathcal{C}^0(\mathbb{R}_+, \mathcal{P}_2(\mathbb{R}^d))$. *Let* $(Z_t)_{t\geqslant 0}$ *and* $(\tilde Z_t)_{t\geqslant 0}$ *be processes associated respectively to* σ, Y_0, ν *and* $\tilde\sigma, \tilde Y_0, \tilde\nu$, *in the sense of* (2.1). *Set* $\Delta Z_t = Z_t - \tilde Z_t$ *and* $f(t) = \mathbb{E}\left(|\Delta Z_t|^2\right)$.

1. *If* $\sigma = \tilde\sigma$, *then almost surely,* $t \mapsto |\Delta Z_t|^2$ *is* $\mathcal{C}^1$ *with for all* $t \geqslant 0$,

$$\partial_t |\Delta Z_t|^2 \;\leqslant\; -2\rho |\Delta Z_t|^2 + 2\kappa \mathbb{W}_2^2(\nu_t, \tilde\nu_t)\,.$$

2. *If* $\tilde\sigma = 0$ *then* f *is* $\mathcal{C}^1$ *with for all* $t \geqslant 0$,

$$f'(t) \leqslant d\sigma^2 \|M\|^2 - 2\rho f(t) + 2\kappa \mathbb{W}_2^2(\nu_t, \tilde\nu_t)\,.$$

Proof. If $\sigma = \tilde\sigma$,

$$\mathrm{d}\Delta Z_t \;=\; \Big(a(Z_t) + b(Z_t, \nu_t) - a(\tilde Z_t) - b(\tilde Z_t, \tilde\nu_t)\Big)\,\mathrm{d}t\,.$$

The right-hand side is continuous, and **(A2)** yields

$$\begin{aligned}\mathrm{d}|\Delta Z_t|^2 &= 2\Delta Z_t \cdot \Big(a(Z_t) + b(Z_t, \nu_t) - a(\tilde Z_t) - b(\tilde Z_t, \tilde\nu_t)\Big)\,\mathrm{d}t \\ &\leqslant -2\rho |\Delta Z_t|^2 \mathrm{d}t + 2\kappa \mathbb{W}_2^2(\nu_t, \tilde\nu_t)\mathrm{d}t.\end{aligned}$$

Using Ito's formula, the second point follows the same line. □

Let us now describe a first consequence of this general result in the deterministic case where $\sigma = \tilde\sigma = 0$ and ν and $\tilde\nu$ are constant Dirac masses.

Lemma 2.2. *Under* **(A)**,

1. *for all* $v \in \mathbb{R}^d$ *and* $z \in \mathbb{R}^d$, *the solution of the Cauchy problem* $\dot z_t = a(z_t) + b(z_t, \delta_v)$ *with* $z_0 = z$ *is defined for all positive times. We denote by* $(\psi_t^v)_{t\geqslant 0}$ *the associated flow (so that* $z_t = \psi_t^v(z)$*).*
2. *for all* $v \in \mathbb{R}^d$, $z, y \in \mathbb{R}^d$ *and* $t \geqslant 0$,

$$|\psi_t^v(z) - \psi_t^v(y)| \;\leqslant\; e^{-\rho t}|z - y|\,.$$

3. *there exists a unique* $\lambda \in \mathbb{R}^d$ *such that* $\psi_t^\lambda(\lambda) = \lambda$ *for all* $t \geqslant 0$.

Proof. Using **(A2)**, for $v \in \mathbb{R}^d$, a solution of $\dot z_t = a(z_t) + b(z_t, \delta_v)$ is such that

$$\partial\left(|z_t|^2\right) \leqslant -2\rho |z_t|^2 + 2|z_t|\left(|a(0)| + |b(0, \delta_v)|\right),$$

which by Grönwall's Lemma implies non-explosion, hence the flow is defined for all positive times.

Let $z, y, v \in \mathbb{R}^d$. Consider the settings of Proposition 2.1 with $\sigma = \tilde{\sigma} = 0$, $\nu_t = \tilde{\nu}_t = \delta_v$ for all $t \geqslant 0$, $Y_0 = z$ and $\tilde{Y}_0 = y$. Write $g(t) = |\psi_t^v(z) - \psi_t^v(y)|^2$ for $t \geqslant 0$. Using that $\mathbb{W}_2(\nu_s, \tilde{\nu}_s) = 0$ for all $s \geqslant 0$ in that case, we immediately get from the first point of Proposition 2.1 that $g'(t) \leqslant -2\rho g(t)$ for all $t \geqslant 0$, which proves the second point. Moreover, this implies that, for all $v \in \mathbb{R}^d$, the flow ψ^v admits a unique equilibrium, that we denote $\Pi(v)$.

Fix $\lambda_1, \lambda_2 \in \mathbb{R}^d$ and set $Z_t = \Pi(\lambda_1)$ and $\tilde{Z}_t = \Pi(\lambda_2)$ for all $t \geqslant 0$. Then Z (resp. $\tilde{Z}$) solves (2.1) with $\sigma = 0$, $Y_0 = \Pi(\lambda_1)$ (resp. $\Pi(\lambda_2)$) and $\nu_t = \delta_{\lambda_1}$ (resp. δ_{λ_2}) for all $t \geqslant 0$. The first point of Proposition 2.1 then reads

$$|\Pi(\lambda_1) - \Pi(\lambda_2)|^2 \leqslant \frac{\kappa}{\rho} |\lambda_1 - \lambda_2|^2 .$$

The condition $\rho > \kappa$ implies that Π is a contraction. As a consequence, it admits a unique fixed point λ. □

2.2. Existence of the process

This section is devoted to the proof of Proposition 1.1. Before entering the proof, let us state the following crucial Lemma.

Lemma 2.3. *For $t \geqslant 0$, let $R(t, \cdot)$ be a probability measure on $[0, t]$ and, for $i = 1, 2$, let $(m_{s,i})_{s \in [0,t]}$ be a family of probability measures in $\mathcal{P}_2(\mathbb{R}^d)$ and $\mu_{t,i} = \int_0^t m_{s,i} R(t, \mathrm{d}s)$. Then*

$$\mathbb{W}_2^2(\mu_{t,1}, \mu_{t,2}) \leqslant \int_0^t \mathbb{W}_2^2(m_{s,1}, m_{s,2}) \, R(t, \mathrm{d}s)$$

Proof. Let S be a random variable distributed according to $R(t, \cdot)$. Conditionally to S, let (Y_1, Y_2) be an optimal $\mathbb{W}_2$-coupling of $m_{S,1}$ and $m_{S,2}$. Then (Y_1, Y_2) is a coupling of $\mu_{t,1}$ and $\mu_{t,2}$, so that

$$\mathbb{W}_2^2(\mu_{t,1}, \mu_{t,2}) \leqslant \mathbb{E}\left(|Y_1 - Y_2|^2\right) = \int_0^t \mathbb{W}_2^2(m_{s,1}, m_{s,2}) \, R(t, \mathrm{d}s) .$$

□

Proof of Proposition 1.1. Let $T < \infty$. Denote first a linearized version of the initial process (1.3-McKean-interaction), where the dependence of the law in the drift is replaced by the family $\bar{m} \in \mathcal{C}([0, T], \mathcal{P}_2(\mathbb{R}^d))$, by $\bar{X}$:

$$\mathrm{d}\bar{X}_t = a(\bar{X}_t)\mathrm{d}t + \bar{b}(t, \bar{X}_t)\mathrm{d}t + \sigma M \mathrm{d}B_t \tag{2.2}$$

where for $(t, x) \in \mathbb{R}_+ \times \mathbb{R}^d$ we set

$$\bar{b}(t, x) = b\left(x, \int_0^t \bar{m}_s R(t, \mathrm{d}s)\right).$$

From **(A1)** and **(A2)** one easily has for any $t \leq T$ and $x \in \mathbb{R}^d$

$$x \cdot (a(x) + \bar{b}(t, x)) \leq -\rho|x|^2 + x \cdot (a(0) + \bar{b}(t, 0)).$$

Hence, on the one hand, if $|x|$ is large enough then

$$x \cdot (a(x) + \bar{b}(t, x)) \leq 0.$$

On the other hand, since, by **(A)**,

$$|\bar{b}(t,0)| \le |b(0,\delta_0)| + C\mathbb{W}_2\left(\int_0^t \bar{m}_s R(t,\mathrm{d}s)),\delta_0\right) \le |b(0,\delta_0)| + C\sup_{s\le T}\left(\int_{\mathbb{R}^d}|y|^2\bar{m}_s(\mathrm{d}y)\right)^{1/2}$$

for some $C>0$, we get that for $|x|$ small,

$$x\cdot(a(x)+\bar{b}(t,x)) \le |x|\left(|a(0)|+|b(0,\delta_0)| + C\sup_{s\le T}\left(\int_{\mathbb{R}^d}|y|^2\bar{m}_s(\mathrm{d}y)\right)^{1/2}\right).$$

Thus, we can apply Theorem 10.2.2, p. 255 of [23] and conclude the existence of a weak solution to the SDE in (2.2). As the coefficients in (2.2) are locally Lipschitz continuous by **(A)**, this SDE admits strong uniqueness. Hence, by Yamada and Watanabe principle, (2.2) admits a unique strong solution up to any time horizon $T>0$.

Now, fix an initial condition $m_0\in\mathcal{P}_2(\mathbb{R}^d)$ and, for $\bar{m}\in\mathcal{C}([0,T],\mathcal{P}_2(\mathbb{R}^d))$ define $\Phi(\bar{m}) = \left(\mathcal{L}(X_t^{\sigma,\bar{m}})\right)_{t\in[0,T]}$, where $X^{\sigma,\bar{m}}$ denotes the solution of (2.2) on $[0,T]$ with initial condition $X_0^{\sigma,\bar{m}}\sim m_0$. Note that, from the computations done in the proof of Proposition 2.1, Φ maps $\mathcal{C}([0,T],\mathcal{P}_2(\mathbb{R}^d))$ onto itself. Let us now see that Φ is a contraction of this space for T small enough, which will conclude the proof as the solutions of (1.3-McKean-interaction) are exactly the fixed points of Φ. For $m,\bar{m}\in\mathcal{C}([0,T],\mathcal{P}_2(\mathbb{R}^d))$, consider $X_t^{\sigma,m}, X_t^{\sigma,\bar{m}}$ two associated solutions of (2.2), with the same initial condition and driven by the same Brownian motion, and write $\Delta Z_t := X_t^{\sigma,m} - X_t^{\sigma,\bar{m}}$. The first point of Proposition 2.1 implies that

$$\partial_t|\Delta Z_t|^2 \;\leqslant\; -2\rho|\Delta Z_t|^2 + 2\kappa\mathbb{W}_2^2(m_t,\bar{m}_t) \;\leqslant\; 2\kappa\mathbb{W}_2^2(m_t,\bar{m}_t)\,.$$

Since $(X_t^{\sigma,m}, X_t^{\sigma,\bar{m}})$ is a coupling of $\Phi(m)_t$ and $\Phi(\bar{m})_t$ for all $t\geqslant 0$, using moreover that $\Delta Z_0 = 0$, we get

$$\sup_{0\le s\le T}\mathbb{W}_2^2(\Phi(m)_s,\Phi(\bar{m})_s) \le \sup_{0\le s\le T}\mathbb{E}\left(|\Delta Z_s|^2\right) \le 2\kappa T\sup_{0\le s\le T}\mathbb{W}_2^2(m_s,\bar{m}_s),$$

which concludes the proof. □

2.3. Long-time behavior

We start by proving a Grönwall-type lemma in the presence of a memory kernel R.

Lemma 2.4. *Let R satisfy* **(A3)** *and $\alpha,\beta\geqslant 0$ with $\alpha>\beta$. There exists a decreasing $x:\mathbb{R}_+\to\mathbb{R}_+$ with $x(t)\to 0$ as $t\to+\infty$ such that, for all $f\in\mathcal{C}^1(\mathbb{R}_+,\mathbb{R}_+)$ and $\gamma\geqslant 0$ such that for all $t\geqslant 0$,*

$$f'(t) \;\leqslant\; -\alpha f(t) + \beta\int_0^t f(s)R(t,\mathrm{d}s) + \gamma,$$

then for all $t\geqslant 0$

$$f(t) \;\leqslant\; \frac{\gamma}{\alpha-\beta} + x(t)\left(f(0)-\frac{\gamma}{\alpha-\beta}\right)_+.$$

Proof. Set $a=\gamma/(\alpha-\beta)$, and let $b>a$. As a first step, let us prove that, if $f(0)\leqslant b$, then $f(t)\leqslant b$ for all $t\geqslant 0$. Suppose that $s=\inf\{t\geqslant 0, f(t)>b\}$ is finite. Then $f(s)=b$ and $\int_0^s f(u)R(s,\mathrm{d}u)\leqslant b$, and thus

$$f'(s) \;\leqslant\; (-\alpha+\beta)(b-a) \;<\; 0,$$

which yields a contradiction, and prove the claim. Since we can take any arbitrary $b > a$, we obtain that, if $f(0) \leqslant a$, then $f(t) \leqslant a$ for all $t \geqslant 0$, which concludes the proof of the lemma in this case.

For the rest of the proof, we assume that $f(0) > a$. The previous result applied with $b = f(0)$ shows that $f(t) \leqslant f(0)$. Let $g(t) = (f(t) - a)/(f(0) - a)$, which is such that, for all $t \geqslant 0$, $g(t) \leqslant g(0) = 1$ and

$$g'(t) \;\leqslant\; -\alpha g(t) + \beta \int_0^t g(s) R(t, \mathrm{d}s)\,.$$

In particular, $g' \leqslant -\alpha g + \beta$, and thus $g \leqslant x_0$ where, for all $t \geqslant 0$,

$$x_0(t) \;=\; e^{-\alpha t} + \frac{\beta}{\alpha}(1 - e^{-\alpha t})\,.$$

Let $c = \sqrt{\beta/\alpha}$, so that $\beta/\alpha < c < 1$, and $t_0 = 0$. Suppose by induction that a function $x_n : \mathbb{R}_+ \to \mathbb{R}_+$ and a time $t_n \geqslant n$ have been defined for some $n \in \mathbb{N}$, with $x_n(t) \to c^{n+2}$ as $t \to +\infty$. Let $t_{n+1} \geqslant 1 + t_n$ be large enough so that

$$x_n(t_{n+1}) \leqslant c^{n+1} \qquad \text{and} \qquad \int_0^t x_n(s) R(t, \mathrm{d}s) \leqslant c^{n+1} \;\forall t \geqslant t_{n+1},$$

which is possible thanks to **(A3)** and the fact $c^{n+1} > c^{n+2}$. Define $x_{n+1}(t) = x_n(t)$ for $t < t_{n+1}$ and

$$x_{n+1}(t) = e^{-\alpha(t - t_{n+1})} c^{n+1} + \left(1 - e^{-\alpha(t - t_{n+1})}\right) c^{n+3}$$

for $t \geqslant t_{n+1}$. Then, $x_{n+1}(t) \to c^{n+3}$ as $t \to +\infty$, which concludes the definition by induction of x_n and t_n for all $n \in \mathbb{N}$. Remark that this construction only involves α, β, R. Let us prove that $g \leqslant x_n$ for all $n \in \mathbb{N}$. We have already treated the case $n = 0$, suppose that this is true for some $n \in \mathbb{N}$. For $t < t_{n+1}$, $g(t) \leqslant x_n(t) = x_{n+1}(t)$. For $t \geqslant t_{n+1}$,

$$g'(t) \;\leqslant\; -\alpha g(t) + \beta \int_0^t g(s) R(t, \mathrm{d}s) \;\leqslant\; -\alpha g(t) + \beta \int_0^t x_n(s) R(t, \mathrm{d}s) \;\leqslant\; -\alpha g(t) + \beta c^{n+1},$$

and thus, using that $g(t_{n+1}) \leqslant x_n(t_{n+1}) \leqslant c^{n+1}$,

$$g(t) \;\leqslant e^{-\alpha(t - t_n)} g(t_n) + \left(1 - e^{-\alpha(t - t_n)}\right) c^{n+3} \;\leqslant\; x_{n+1}(t),$$

which concludes the proof that $g \leqslant x_n$ for all $n \in \mathbb{N}$. Define $x(t) = c^n$ for $t \in [t_n, t_{n+1})$ (remark that $t_n \geqslant n \to +\infty$, so that $x(t)$ is indeed defined for all $t \geqslant 0$). Then $g \leqslant x$, which concludes the proof of the lemma. □

Remark 2.5. In the case where $R_t = \delta_t$ for all $t \geqslant 0$, of course the result holds with $x(t) = e^{-(\alpha - \beta)t}$. In the uniform case, assuming that $f' = -\alpha f + \beta F$ with $F(t) = t^{-1} \int_0^t f(s)\mathrm{d}s$, we see that, for large t, $F(t)$ evolves slowly, so the evolution of f is approximately $f'(t) \simeq -\alpha f(t) + \beta F(t_0)$ for with $0 \leqslant t - t_0 \ll t_0$, so that $f(t) \simeq \beta/\alpha F(t)$ and $F'(t) \simeq -(1 - \beta/\alpha)/tF(t)$. As a consequence, F (hence f) goes to 0 as $t^{\beta/\alpha - 1}$.

Proposition 2.6. *Under* **(A)**, *there exists a positive function Q on $\mathbb{R}_+$ that depends only on ρ and κ and vanishes at infinity such that the following holds. Let $(X_t^\sigma, \mu_t^\sigma, m_t^\sigma)_{t \geqslant 0}$ and $(\tilde{X}_t^\sigma, \tilde{\mu}_t^\sigma, \tilde{m}_t^\sigma)_{t \geqslant 0}$ solve* (1.3-McKean-interaction) *with respective initial distributions $m_0^\sigma, \tilde{m}_0^\sigma \in \mathcal{P}_2(\mathbb{R}^d)$. Then for all $t \geqslant 0$,*

$$\mathbb{W}_2(m_t^\sigma, \tilde{m}_t^\sigma) + \mathbb{W}_2(\mu_t^\sigma, \tilde{\mu}_t^\sigma) \;\leqslant\; Q(t) \mathbb{W}_2(m_0^\sigma, \tilde{m}_0^\sigma)\,.$$

Proof. We can consider a copy of the processes since the statement only concerns the distributions. We consider $(Y_0, \tilde{Y}_0)$ an optimal $\mathbb{W}_2$ coupling of m_0^σ and $\tilde{m}_0^\sigma$. Then $(X_t^\sigma)_{t\geqslant 0}$ (resp. $(\tilde{X}_t^\sigma)_{t\geqslant 0}$) has the same law as the solution $(Z_t)_{t\geqslant 0}$ (resp. $(\tilde{Z}_t)_{t\geqslant 0}$) of (2.1) associated to Y_0, σ and $\nu = (\mu_t^\sigma)_{t\geqslant 0}$ (resp. $\tilde{Y}_0$, σ and $(\tilde{\mu}_t^\sigma)_{t\geqslant 0}$). In particular, $(Z_t, \tilde{Z}_t)$ is a coupling of m_t^σ and $\tilde{m}_t^\sigma$ for all $t \geqslant 0$ which, together with Lemma 2.3, implies that

$$\mathbb{W}_2^2(\mu_t^\sigma, \tilde{\mu}_t^\sigma) \leqslant \int_0^t f(s) R(t, \mathrm{d}s)$$

with $f(s) = \mathbb{E}(|Z_s - \tilde{Z}_s|^2)$ for all $s \geqslant 0$. Taking the expectation in the first point of Proposition 2.1, we get that, for all $t \geqslant 0$,

$$f'(t) \leqslant -2\rho f(t) + 2\kappa \int_0^t f(s) R(t, \mathrm{d}s). \tag{2.3}$$

The conclusion follows, thanks to **(A3)**, by applying Lemmas 2.4 and 2.3 to get

$$\mathbb{W}_2(m_t^\sigma, \tilde{m}_t^\sigma) + \mathbb{W}_2(\mu_t^\sigma, \tilde{\mu}_t^\sigma) \leqslant f(t) + \int_0^t f(s) R(t, \mathrm{d}s) \leqslant \left[x(t) + \int_0^t x(s) R(t, \mathrm{d}s) f(0) \right].$$

□

Corollary 2.7. *Under* **(A)**, *there exists a unique* $m_\infty^\sigma \in \mathcal{P}_2(\mathbb{R}^d)$ *which is stationary for the process* (1.3-McKean-interaction), *in the sense that the process* $(Z_t)_{t\geqslant 0}$ *solving* (2.1) *with* $Y_0 \sim m_\infty^\sigma$ *and* $\nu_t = m_\infty^\sigma$ *for all* $t \geqslant 0$ *is such that* $Z_t \sim m_\infty^\sigma$ *for all* $t \geqslant 0$.

Proof. Remark that the fact that $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ is a fixed point for (1.3-McKean-interaction) does not depend on R. As a consequence, we only consider the classical non-linear McKean-Vlasov case, *i.e.* $R(t, \cdot) = \delta_t$ for all $t \geqslant 0$. For $s \geqslant 0$, let $\Phi_s : \mathcal{P}_2(\mathbb{R}^d) \to \mathcal{P}_2(\mathbb{R}^d)$ be defined by $\Phi_s(m_0^\sigma) = m_s^\sigma$. In the classical non-linear case, we have $\Phi_s(m_t^\sigma) = m_{s+t}^\sigma$ for all $t, s \geqslant 0$. Let $s \geqslant 0$ be large enough so that, by Proposition 2.6, Φ_s is a contraction of $(\mathcal{P}_2(\mathbb{R}^d), \mathbb{W}_2)$, which is a complete metric space. Denote m_∞^σ the unique fixed point of Φ_s. Using that $\Phi_{t+s} = \Phi_t \circ \Phi_s = \Phi_s \circ \Phi_t$ for all $t \geqslant 0$, we get that $\Phi_t(m_\infty^\sigma)$ is a fixed point of Φ_s for all $t \geqslant 0$ which, by uniqueness, implies that $\Phi_t(m_\infty^\sigma) = m_\infty^\sigma$ for all $t \geqslant 0$. □

2.4. Low noise asymptotics

First, we state the low noise convergence of the stationary distribution.

Proposition 2.8. *Under* **(A)**, *let* $\lambda \in \mathbb{R}^d$ *be given by Lemma 1.2 and, for* $\sigma \geqslant 0$, *let* m_∞^σ *be given by Corollary 2.7. For all* $\sigma \geqslant 0$,

$$\mathbb{W}_2^2(m_\infty^\sigma, \delta_\lambda) \leqslant \frac{d\|M\|^2}{2(\rho - \kappa)} \sigma^2 .$$

Proof. Let Z (resp. $\tilde{Z}$) solve (2.1) with $Y_0 \sim m_\infty^\sigma$, σ and $\nu_t = m_\infty^\sigma$ for all $t \geqslant 0$ (resp. $\tilde{Y}_0 = \lambda$, $\tilde{\sigma} = 0$ and $\nu_t = \delta_\lambda$ for all $t \geqslant 0$). In particular, for all for all $t \geqslant 0$, $Z_t \sim m_\infty^\sigma$, $\tilde{Z}_t = \lambda$ and $\mathbb{W}_2^2(m_\infty^\sigma, \delta_\lambda) = \mathbb{E}(|Z_t - \lambda|^2)$. The result then is a straightforward consequence of the second point of Proposition 2.1. □

Second, we consider the low noise convergence of the process on finite time intervals.

Proposition 2.9. *Assume* **(A)**, *let* $m_0 \in \mathcal{P}_2(\mathbb{R}^d)$ *and* $Y_0 \sim m_0$. *There exists* $\mathcal{K} \in \mathcal{C}^0(\mathbb{R}_+, \mathbb{R}_+)$ *such that the following holds. For a fixed Brownian motion* $(B_t)_{t\geqslant 0}$, *for all* $\sigma \geqslant 0$, *let* $(X_t^\sigma)_{t\geqslant 0}$ *be the solution of*

(1.3-McKean-interaction) *with* $X_0^\sigma = Y_0$. *For all* $t \geqslant 0$, $\sigma \geqslant 0$ *and* $\varepsilon \in (0,1]$,

$$\mathbb{P}\left(\sup_{s\in[0,t]} |X_s^\sigma - X_s^0| > \varepsilon\right) \leqslant \frac{\sigma \mathcal{K}(t)}{\varepsilon}. \tag{2.4}$$

Proof. Let $f(t) = \mathbb{E}(|X_s^\sigma - X_s^0|^2)$ for $t \geqslant 0$. In particular, $\mathbb{W}_2^2(m_t^\sigma, m_t^0) \leqslant f(t)$ for all $t \geqslant 0$ and, applying Lemma 2.3 and the second point of Proposition 2.1,

$$f'(t) \leqslant \gamma - 2\rho f(t) + 2\kappa \int_0^t f(s) R(t, \mathrm{d}s)$$

with $\gamma = \mathrm{d}\sigma^2 \|M\|^2$. Since $f(0) = 0$, Lemma 2.4 implies that $f(t) \leqslant \gamma/(2(\rho - \kappa))$ for all $t \geqslant 0$, and thus

$$\mathbb{W}_2^2(m_t^\sigma, m_t^0) \leqslant \frac{d\|M\|^2\sigma^2}{2(\rho-\kappa)}$$

for all $t \geqslant 0$, which also implies

$$\mathbb{W}_2^2(\mu_t^\sigma, \mu_t^0) \leqslant \frac{d\|M\|^2\sigma^2}{2(\rho-\kappa)}$$

for all $t \geqslant 0$.

Recall that, from **(A1)** and **(D2)**, a and b are Lipschitz continuous on $\mathcal{D}_{+1}$. Moreover, **(D3)** implies that $X_t^0 \in \mathcal{D}$ for all $t \geqslant 0$, see Lemma 2.10. Let $\tau_{+1} = \inf\{t \geqslant 0, X_t^\sigma \notin \mathcal{D}_{+1}\}$. Then, for some $L > 0$, for all $t \leqslant \tau_{+1}$,

$$|X_t^\sigma - X_t^0| \leqslant \sigma\|M\||B_t| + L\int_0^t \left(|X_s^\sigma - X_s^0| + \mathbb{W}_2\left(\mu_s^\sigma, \mu_s^0\right)\right) \mathrm{d}s\,.$$

Grönwall's Lemma yields

$$\text{almost surely, } \forall t \in [0, \tau_{+1}], \qquad \sup_{s\in[0,t]} |X_s^\sigma - X_s^0| \leqslant \sigma e^{Lt}\left(\|M\| \sup_{s\in[0,t]} |B_s| + Ct\right),$$

with $C = L\|M\|\sqrt{d/(2(\rho-\kappa))}$.

Now, fix $t \geqslant 0$ and $\varepsilon \in (0,1]$. Remark that $|X_s^\sigma - X_s^0| \leqslant \varepsilon$ for some $s \geqslant 0$ implies that $d(X_s^\sigma, \mathcal{D}) \leqslant 1$. Hence,

$$\left\{\sigma e^{Lt}\left(\|M\| \sup_{s\in[0,t]} |B_s| + Ct\right) \leqslant \varepsilon\right\} \subset \left(\{\tau_{+1} \geqslant t\} \cap \left\{\sup_{s\in[0,t]} |X_s^\sigma - X_s^0| \leqslant \varepsilon\right\}\right).$$

As a conclusion, if $\sigma e^{Lt} Ct > \varepsilon/2$ we simply bound

$$\mathbb{P}\left(\sup_{s\in[0,t]} |X_s^\sigma - X_s^0| > \varepsilon\right) \leqslant 1 \leqslant \frac{2\sigma e^{Lt} Ct}{\varepsilon},$$

while, if $\sigma e^{Lt} Ct \leqslant \varepsilon/2$, we can bound

$$\mathbb{P}\left(\sup_{s\in[0,t]} |X_s^\sigma - X_s^0| > \varepsilon\right) \leqslant \mathbb{P}\left(\sup_{s\in[0,t]} |B_s| > \frac{\varepsilon e^{-Lt}}{2\sigma\|M\|}\right) \leqslant \frac{2\sigma\|M\|\sqrt{t}e^{Lt}}{\varepsilon},$$

thanks to Doob's inequality. □

From the previous result, the question of the exit time of X^σ before a given fixed time T (independent from σ) in the low noise regime boils down to the question of the exit time for the deterministic process X^0, which is addressed in the next lemma.

Lemma 2.10. *Under Assumptions* **(A)** *and* **(D)**, *the process* X^0 *with initial law* m_0 *and solving the deterministic equation* (1.3-McKean-interaction) *with* $\sigma = 0$ *satisfies:*

$$\mathbb{P}\left(X_t^0 \in \mathcal{D}\, \forall t \geqslant 0\right) = 1\,.$$

Proof. Let $r > 0$ be such that the support of m_0 is in the ball $\mathbb{B}(\lambda, r) \subset \mathcal{D}$. Applying the second point of Proposition 2.1 with $\sigma = 0$ and using that

$$f(t) := \mathbb{E}|X_t^0 - \lambda|^2 = \mathbb{W}_2^2(m_t^0, \delta_\lambda)$$

for all $t \geqslant 0$, we get that $f'(t) \leqslant 0$, and thus $f(t) \leqslant f(0) \leqslant r^2$ for all $t \geqslant 0$. Using now the first point of Proposition 2.1, we get that, almost surely, for all $t \geqslant 0$,

$$\partial_t |X_t^0 - \lambda|^2 \leqslant -2\rho |X_t^0 - \lambda|^2 + 2\kappa r^2$$

with $|X_0^0 - \lambda|^2 \leqslant r^2$, which implies that $X_t^0 \in \mathbb{B}(\lambda, r) \subset \mathcal{D}$ for all $t \geqslant 0$.

□

Remark 2.11. In fact, the same proof works if we only assume that m_0 has a compact support in a ball $\mathbb{B}\,(x_0, r') \subset \mathcal{D}$ where x_0 is such that the solution of

$$\partial_t z_t = a(z_t) + b(z_t, \delta_{z_t}), \qquad z_0 = x_0$$

stays in $\mathcal{D}$ for all $t \geqslant 0$.

2.5. Proof of Theorem 1.3

Fix a and b that satisfy Assumption **(A)**, and an initial condition m_0. For all $\sigma \geqslant 0$ we consider $(X_t^\sigma)_{t \geqslant 0}$ that solves (1.3-McKean-interaction) where $m_0^\sigma = \mathcal{L}(X_0^\sigma) = m_0$. Consider $\lambda \in \mathbb{R}^d$ as given by Lemma 1.2 and, for all $\sigma \geqslant 0$, $(\tilde{X}_t^\sigma)_{t \geqslant 0}$ that solves (1.6–equilibrium-interaction) with $\tilde{X}_0 = \lambda$.

Now, let $\delta > 0$. Let $\xi > 0$ be small enough so that $|H_{u,\xi} - H| \leqslant \delta/2$ for $u \in \{i, e\}$. We also take take ξ small enough such that $\tau_0(\mathcal{D}_{i,\xi})$ is equal to infinity, which is ensured by Lemma 2.10. Then, for all $\sigma > 0$, according to **(K)** one has

$$\begin{aligned}\mathbb{P}\left(\tau_\sigma(\mathcal{D}) > \exp\left(\frac{2}{\sigma^2}\,(H + \delta)\right)\right) \;\leqslant\; & \mathbb{P}\left(\tilde{\tau}_\sigma(\mathcal{D}_{e,\xi}) > \exp\left(\frac{2}{\sigma^2}\,(H + \delta)\right)\right) \\ & + \mathbb{P}\left(\tau_\sigma(\mathcal{D}) > \tilde{\tau}_\sigma(\mathcal{D}_{e,\xi})\right)\,.\end{aligned}$$

The choice of ξ and (1.10) imply that the first term of the right-hand side vanishes with σ, since $H + \delta \geqslant H_{e,\xi} + \delta/2$. Similarly,

$$\mathbb{P}\left(\tau_\sigma(\mathcal{D}) < \exp\left(\frac{2}{\sigma^2}\,(H - \delta)\right)\right) \;\leqslant\; \mathbb{P}\left(\tau_\sigma(\mathcal{D}) < \tilde{\tau}_\sigma(\mathcal{D}_{i,\xi})\right) + \underset{\sigma \to 0}{o}(1)\,.$$

As a consequence, the proof will be concluded if we show that

$$\mathbb{P}\left(\tau_\sigma(\mathcal{D}) > \tilde{\tau}_\sigma(\mathcal{D}_{e,\xi})\right) + \mathbb{P}\left(\tau_\sigma(\mathcal{D}) < \tilde{\tau}_\sigma(\mathcal{D}_{i,\xi})\right) \underset{\sigma\to 0}{\longrightarrow} 0\,.$$

We will consider T large enough such that the initial condition will be forgotten (since there will be stabilization around λ) and the coupling will hold true for any larger time.

Let us prove this in two steps: considering for $T, \varepsilon > 0$ the event

$$\mathcal{A}_{T,\varepsilon} \;=\; \Big\{\big[\tau_\sigma(\mathcal{D}) \wedge \tilde{\tau}_\sigma(\mathcal{D}_{i,\xi})\big] > T \text{ and } |X_T^\sigma - \tilde{X}_T^\sigma| \leqslant \varepsilon\Big\},$$

we will prove in Step 1 that there exist $T_0, \varepsilon, \sigma_0 > 0$ such that for all $\sigma \in (0, \sigma_0]$ and $T \geqslant T_0$,

$$\mathbb{P}\left(\{\tau_\sigma(\mathcal{D}) > \tilde{\tau}_\sigma(\mathcal{D}_{e,\xi})\} \cap \mathcal{A}_{T,\varepsilon}\right) + \mathbb{P}\left(\{\tau_\sigma(\mathcal{D}) < \tilde{\tau}_\sigma(\mathcal{D}_{i,\xi})\} \cap \mathcal{A}_{T,\varepsilon}\right) \;=\; 0\,. \tag{2.5}$$

Then in Step 2 we will show that for all $\varepsilon > 0$,

$$\lim_{T\to+\infty} \liminf_{\sigma\to 0} \mathbb{P}\left(\mathcal{A}_{T,\varepsilon}\right) \;=\; 1\,.$$

The combination of Step 1 and Step 2 concludes the whole proof.

Step 1. Remark that, on the event $\mathcal{A}_{T,\varepsilon}$, $\tilde{\tau}_\sigma(\mathcal{D}_{e,\xi}) > T$ (since $\tilde{\tau}_\sigma(\mathcal{D}_{e,\xi}) \geqslant \tilde{\tau}_\sigma(\mathcal{D}_{i,\xi})$). As a consequence,

$$\Big(\{\tau_\sigma(\mathcal{D}) > \tilde{\tau}_\sigma(\mathcal{D}_{e,\xi})\} \cap \mathcal{A}_{T,\varepsilon}\Big) \;\subset\; \Big\{\sup_{t\geqslant T} |X_t^\sigma - \tilde{X}_t^\sigma| \geqslant \xi\Big\} \;:=\; \mathcal{B}_{T,\xi}\,.$$

Indeed, if $T < s := \tilde{\tau}_\sigma(\mathcal{D}_{e,\xi}) < \tau_\sigma(\mathcal{D})$, then $X_s^\sigma \in \mathcal{D}$ while $\tilde{Z}_s^\sigma \in \partial\mathcal{D}_{e,\xi}$, so that $|X_s^\sigma - \tilde{X}_s^\sigma| \geqslant \mathrm{d}(\mathcal{D}, \mathcal{D}_{e,\xi}^c) \geqslant \xi$. With a similar argument we see that

$$\Big(\{\tau_\sigma(\mathcal{D}) < \tilde{\tau}_\sigma(\mathcal{D}_{i,\xi})\} \cap \mathcal{A}_{T,\varepsilon}\Big) \;\subset\; \mathcal{B}_{T,\xi}\,.$$

It only remains to prove that $\mathbb{P}(\mathcal{A}_{T,\varepsilon} \cap \mathcal{B}_{T,\xi}) = 0$ for T large enough and ε, σ small enough.

From the first part of Proposition 2.1 and Gronwall's inequality, for all $t \geqslant T \geqslant 0$, $\sigma > 0$, almost surely,

$$|X_t^\sigma - \tilde{X}_t^\sigma|^2 \leqslant e^{-2\rho(t-T)}|X_T^\sigma - \tilde{X}_T^\sigma|^2 + \left(1 - e^{-2\rho(t-T)}\right)\frac{\kappa}{\rho}\sup_{t\geqslant T}\mathbb{W}_2^2(\mu_t^\sigma, \delta_\lambda),$$

and thus, using that $\kappa < \rho$, almost surely,

$$\sup_{t\geqslant T}|X_t^\sigma - \tilde{X}_t^\sigma| \;\leqslant\; \max\left(|X_T^\sigma - \tilde{X}_T^\sigma|, \sup_{t\geqslant T}\mathbb{W}_2(\mu_t^\sigma, \delta_\lambda)\right),$$

A keypoint of the proof is that the last term will be small when T is large and σ is small, since μ_t^σ converges (uniformly in σ) to an equilibrium m_∞^σ which is itself close to δ_λ at low temperature. More precisely, in view of Propositions 2.6 and 2.8, choose $T_0, \sigma_0 > 0$ such that

$$\frac{\xi}{2} \;\geqslant\; \sup_{t\geqslant T_0}\sup_{\sigma\in(0,\sigma_0]}\left[\mathbb{W}_2\left(\mu_t^\sigma, m_\infty^\sigma\right) + \mathbb{W}_2\left(m_\infty^\sigma, \delta_\lambda\right)\right] \;\geqslant\; \sup_{t\geqslant T_0}\sup_{\sigma\in(0,\sigma_0]}\mathbb{W}_2\left(\mu_t^\sigma, \delta_\lambda\right).$$

Set $\varepsilon = \xi/2$. Then, for all $T \geqslant T_0$ and all $\sigma \in (0, \sigma_0]$,

$$\mathcal{A}_{T,\varepsilon} \subset \left\{ \sup_{t \geqslant T} |X_t^\sigma - \tilde{X}_t^\sigma| \leqslant \xi/2 \right\} \subset \mathcal{B}_{T,\xi}^c,$$

which concludes the proof of (2.5).

Step 2. Fix $T, \varepsilon > 0$. We bound

$$\mathbb{P}\left(\mathcal{A}_{T,\varepsilon}^c\right) \leqslant \mathbb{P}\left(\tau_\sigma(\mathcal{D}) \leqslant T\right) + \mathbb{P}\left(\tilde{\tau}_\sigma(\mathcal{D}_{i,\xi}) \leqslant T\right) + \mathbb{P}\left(|X_T^\sigma - \lambda| > \varepsilon/2\right) + \mathbb{P}\left(|\tilde{X}_T^\sigma - \lambda| > \varepsilon/2\right)$$

and treat each term separately. The second term vanishes with σ according to (1.10) (see **(K)**). The last two terms are similar: from the Markov inequality we get

$$\mathbb{P}\left(|\tilde{X}_T^\sigma - \lambda| > \varepsilon/2\right) \leqslant \frac{4}{\varepsilon^2} \mathbb{W}_2^2\left(\mathcal{L}(\tilde{X}_T^\sigma), \delta_\lambda\right).$$

Notice that, as $\tilde{X}_0^\sigma = \lambda$, we have that for all $t \geq 0$, $\mathcal{L}(\tilde{X}_t^0) = \delta_\lambda$. Thus, from the second part of Proposition 2.1 we easily get that the right-hand side vanishes with σ. Similarly,

$$\mathbb{P}\left(|X_T^\sigma - \lambda| > \varepsilon/2\right) \leqslant \frac{4}{\varepsilon^2} \mathbb{W}_2^2\left(m_T^\sigma, \delta_\lambda\right) \leqslant \frac{4}{\varepsilon^2} \left(\mathbb{W}_2\left(m_T^\sigma, m_\infty^\sigma\right) + \mathbb{W}_2\left(m_\infty^\sigma, \delta_\lambda\right)\right)^2$$

where m_∞^σ is the equilibrium given by Corollary 2.7. Now we apply Proposition 2.6 for the following two processes: the process X^σ and the process solving (1.3-McKean-interaction) with initial condition m_∞^σ, so that $m_t^\sigma = m_\infty^\sigma$ for all $t \geqslant 0$. It comes

$$\mathbb{W}_2\left(m_T^\sigma, m_\infty^\sigma\right) \leqslant Q(T) \mathbb{W}_2\left(m_0^\sigma, m_\infty^\sigma\right) \leqslant Q(T)\left(\mathbb{W}_2\left(m_0^\sigma, \delta_\lambda\right) + \mathbb{W}_2\left(\delta_\lambda, m_\infty^\sigma\right)\right).$$

From Proposition 2.8, we see that $\mathbb{W}_2\left(\delta_\lambda, \mu_\infty^\sigma\right)$ vanishes with σ, so that

$$\limsup_{\sigma \to 0} \mathbb{P}\left(|X_T^\sigma - \lambda| > \varepsilon/2\right) \leqslant \frac{4}{\varepsilon^2} Q^2(T) \mathbb{W}_2^2\left(m_0, \delta_\lambda\right).$$

Finally, since $\tau_0(\mathcal{D}_{i,\xi})$ is equal to infinity, we can then bound the first term of the right hand as in Step 1 of the proof by

$$\mathbb{P}\left(\tau_0(\mathcal{D}_{i,\xi}) > T \geqslant \tau_\sigma(\mathcal{D})\right) \leqslant \mathbb{P}\left(\sup_{s \in [0,T]} |X_s^\sigma - X_s^0| \geqslant \xi\right),$$

which, from Proposition 2.9, vanishes with σ.

Gathering all these bounds we have obtained that

$$\limsup_{\sigma \to 0} \mathbb{P}\left(\mathcal{A}_{T,\varepsilon}^c\right) \leqslant \frac{4}{\varepsilon^2} Q^2(T) \mathbb{W}_2^2\left(m_0, \delta_\lambda\right) \underset{T \to \infty}{\longrightarrow} 0,$$

which concludes the proof since $Q(t)$ vanishes at infinity, see Proposition 2.6.

3. Application to particular models

As discussed in Section 1.2, **(A2)** is implied by the condition on a

$$\exists \rho > 0,\ \forall z, y \in \mathbb{R}^d, \qquad (z-y)\cdot(a(z)-a(y)) \leqslant -\rho|z-y|^2 \tag{3.1}$$

and by taking an interaction $b = \varepsilon\tilde{b}$ where $\tilde{b}$ is Lipschitz continuous and ε is sufficiently small depending on ρ and the Lipschitz constants of $\tilde{b}$. Moreover, this condition does not involve the memory kernel R. For this reason, we now present, separately, some non-interacting diffusions (solving (1.1-non-interacting)), then some interacting forces b, and check that the conditions hold.

3.1. Some underlying (linear) diffusions

3.1.1. The overdamped Langevin diffusion

The overdamped Langevin diffusion is the Markov diffusion on $\mathbb{R}^d$ which solves

$$\mathrm{d}X_t = -\nabla V(X_t)\mathrm{d}t + \sqrt{2\sigma}\mathrm{d}W_t$$

where $V \in \mathcal{C}^1(\mathbb{R}^d)$, which corresponds to (1.3-McKean-interaction) with $b = 0$, $a = -\nabla V$ and $M = I_d$. The following is clear:

Proposition 3.1. *Suppose that V is strongly convex. Then $a = -\nabla V$ satisfies (3.1).*

Remark 3.2. For V uniformly convex and $\mathcal{C}^2$, the unique fixed point of $\dot{z} = -\nabla V(z)$ is the unique minimizer of V.

3.1.2. The kinetic Langevin diffusion

The kinetic Langevin diffusion corresponds to Markov diffusion on $\mathbb{R}^d$ (with $d = 2n$ for some $n \geqslant 1$) that solves (1.3-McKean-interaction) with $b = 0$ and (decomposing $z = (x, y) \in \mathbb{R}^n \times \mathbb{R}^n$)

$$a(x,y) \;=\; \begin{pmatrix} y \\ -\nabla V(x) - \gamma y \end{pmatrix}, \qquad M \;=\; \begin{pmatrix} 0 & 0 \\ 0 & \sqrt{2\gamma} I_n \end{pmatrix} \tag{3.2}$$

where $V \in \mathcal{C}^1(\mathbb{R}^n)$ and $\gamma > 0$. The condition (3.1) is not satisfied directly but, as proven in [20], provided V is convex, its gradient is Lipschitz continuous and the friction γ is high enough, then it is satisfied up to a linear change of variable. More precisely, Proposition 4 of [20] reads:

Proposition 3.3. *Suppose that $V \in \mathcal{C}^2(\mathbb{R}^n)$ and that there exist $\lambda, \Lambda > 0$ with*

$$\lambda I_n \;\leqslant\; \nabla^2 V(x) \;\leqslant\; \Lambda I_n$$

for all $x \in \mathbb{R}^n$. Assume furthermore that $\Lambda - \lambda < \gamma(\sqrt{\lambda} + \sqrt{\Lambda})$. Then there exists an invertible $d \times d$ matrix D such that $\tilde{a}$ given by $\tilde{a}(z) = D^{-1}a(Dz)$ for $z \in \mathbb{R}^d$ satisfies (3.1).

Remark 3.4. In the setting of Proposition 3.3, the unique fixed point of $\dot{z} = a(z)$ is $z_* = (x_*, 0)$ where x_* is the unique minimum of V (in other words, z_* is the unique minimum of the Hamiltonian $V(x) + \gamma|y|^2/2$).

3.1.3. Overdamped Langevin diffusion with coloured noise

Consider a variation of the overdamped Langevin process where the white noise is replaced by a diffusion process, *i.e.*

$$\mathrm{d}X_t = -\nabla V(X_t)\mathrm{d}t + B\eta_t\,\mathrm{d}t \tag{3.3a}$$

$$\mathrm{d}\eta_t = F(\eta_t)\,\mathrm{d}t + \sigma D \mathrm{d}W_t, \tag{3.3b}$$

where $X_t \in \mathbb{R}^d, \eta_t \in \mathbb{R}^n$, $B \in \mathbb{R}^{d\times n}$, $F \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R}^n)$, $D \in \mathbb{R}^{n\times n}$. It corresponds to coefficients

$$a(x,\eta) \;=\; \begin{pmatrix} -\nabla V(x) + B\eta \\ F(\eta) \end{pmatrix}, \qquad M \;=\; \begin{pmatrix} 0 & 0 \\ 0 & D \end{pmatrix} \tag{3.4}$$

Important instances of this model include

1. Scalar OU process:

$$\mathrm{d}X_t = -\nabla V(X_t)\mathrm{d}t + \sqrt{2}\eta_t\,\mathrm{d}t \tag{3.5a}$$
$$\mathrm{d}\eta_t = -\eta_t\,\mathrm{d}t + \sqrt{2}\sigma\,\mathrm{d}W_t, \tag{3.5b}$$

2. Langevin noise:

$$\mathrm{d}X_t = -\nabla V(X_t)\mathrm{d}t + \sqrt{2}\,y_t\,\mathrm{d}t \tag{3.6a}$$
$$\mathrm{d}y_t = v_t \tag{3.6b}$$
$$\mathrm{d}v_t = -y_t\,\mathrm{d}t - \gamma v_t\,\mathrm{d}t + \sqrt{2}\sigma\,\mathrm{d}W_t. \tag{3.6c}$$

Notice that, in these two examples, F is linear, so that by considering a rescaled auxiliary variable $\tilde{\eta}_t = \eta_t/\sigma$ we end up with

$$\mathrm{d}X_t = -\nabla V(X_t)\mathrm{d}t + \sqrt{2}\sigma B\tilde{\eta}_t,$$

where the dynamics of $\tilde{\eta}$ is independent from σ. This is more similar to the standard overdamped Langevin with small white noise, however it doesn't fit in this form in the framework of our results where σ is the intensity of the white noise, which is why we wrote it with η rather than $\tilde{\eta}$.

Proposition 3.5. *Assume that V is strongly convex, that F satisfies (3.1) and let a be given by (3.4). Then there exists a diagonal $d \times d$ matrix D such that $\tilde{a}$ given by $\tilde{a}(z) = D^{-1}a(Dz)$ for $z \in \mathbb{R}^d$ satisfies (3.1).*

Remark 3.6. In the case of a Langevin noise, F does not satisfies (3.1) directly but, as discussed in the previous section, we can enforce (3.1) by a linear change of variable. More generally, when F is linear, given by a matrix with spectrum in $\{\lambda \in \mathbb{C}, \Re e(\lambda) < 0\}$, it is known that (3.1) holds up to a linear change of variable.

Proof. Assume that (3.1) holds for F for some ρ and that V is ρ'-strongly convex. Let $\tilde{a}(x,\eta) = a(cx,\eta)$ with $c = |B|^2/(\rho\rho')$. Then, for all $z = (x,\eta), z' = (x',\eta') \in \mathbb{R}^d \times \mathbb{R}^n$,

$$\begin{aligned}(z - z') \cdot (\tilde{a}(z) - \tilde{a}(z')) &= (x - x') \cdot (\nabla V(cx') - \nabla V(cx) + B(\eta - \eta')) + (\eta - \eta')\,(F(\eta) - F(\eta')) \\ &\leqslant -c\rho'|x - x'|^2 + |B||x - x'||\eta - \eta'| - \rho|\eta - \eta'|^2 \\ &\leqslant -\frac{c\rho'}{2}|x - x'|^2 - \frac{\rho}{2}|\eta - \eta'|^2\,.\end{aligned}$$

□

3.1.4. The generalized Langevin diffusion

The generalized Langevin diffusion corresponds to Markov diffusion on $\mathbb{R}^d$ (with $d = 2n+p$ for some $n, p \geqslant 1$) that solves (1.3-McKean-interaction) with $b = 0$ and (decomposing $z = (x, y, w) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^p$)

$$a(x,y,w) \;=\; \begin{pmatrix} y \\ -\nabla V(x) \\ 0 \end{pmatrix} - \gamma \begin{pmatrix} 0 & 0 & 0 \\ 0 & B_{11} & B_{12} \\ 0 & B_{21} & B_{22} \end{pmatrix} \begin{pmatrix} x \\ y \\ w \end{pmatrix}, \quad M \;=\; \sqrt{\gamma} \begin{pmatrix} 0 & 0 & 0 \\ 0 & \Sigma_{11} & \Sigma_{12} \\ 0 & \Sigma_{21} & \Sigma_{22} \end{pmatrix} \tag{3.7}$$

where $V \in \mathcal{C}^1(\mathbb{R}^n)$, $\gamma > 0$ and the $\Sigma_{i,j}$'s and $B_{i,j}$'s are constant matrices. These processes arise in Monte Carlo method [15, 20, 21] and in effective dynamics problem in molecular dynamics [4, 22]. The fluctuation-dissipation relation is said to be satisfied if

$$\Sigma^T \Sigma \;=\; B^T + B \qquad \text{where} \qquad B \;=\; \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}, \qquad \Sigma \;=\; \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

In that case, the invariant measure of the process is the probability density proportional to $\exp(-[V(x)+|y|^2/2+|w|^2/2]/\sigma^2)$. Classical examples are K^{th} order Generalized Langevin processes for $K \geqslant 3$, which corresponds to $p = (K-2)n$, *i.e.* the total dimension is $d = Kn$, and, decomposing B in $n \times n$ blocks,

$$B = \gamma \begin{pmatrix} 0 & -I_n & 0 & \dots & 0 \\ I_n & 0 & -I_n & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & I_n & 0 & -I_n \\ 0 & \dots & 0 & I_n & I_n \end{pmatrix}.$$

In other words, the process solves

$$\begin{aligned} \mathrm{d}X_t &= Y_t \mathrm{d}t \\ \mathrm{d}Y_t &= -\nabla V(X_t)\mathrm{d}t + \gamma Z_t^{(1)} \mathrm{d}t \\ \mathrm{d}Z_t^{(1)} &= \gamma \left(Z_t^{(2)} - Y_t \right) \mathrm{d}t \\ \mathrm{d}Z_t^{(2)} &= \gamma \left(Z_t^{(3)} - Z_t^{(1)} \right) \mathrm{d}t \\ &\vdots \\ \mathrm{d}Z_t^{(K-3)} &= \gamma \left(Z_t^{(K-2)} - Z_t^{(K-3)} \right) \mathrm{d}t \\ \mathrm{d}Z_t^{(K-2)} &= -\gamma \left(Z_t^{(K-2)} + Z_t^{(K-3)} \right) \mathrm{d}t + \sqrt{2\gamma} \mathrm{d}W_t \,. \end{aligned}$$

Here, the SDE is very degenerated, since a d-dimensional noise drives a Kd-dimensional process. See [15, 21] and references for more detailed discussions and more examples.

Similarly to the kinetic Langevin case, assuming that V is strongly convex with bounded Hessian and that the friction is large enough, it can be proven that, up to a linear change of variables, a given by (3.7) satisfies (3.1), see Proposition 5 of [20] for details.

Remark 3.7. The unique fixed point of $\dot{z} = a(z)$ where a is given by (3.7) is $z_* = (x_*, 0, 0)$ where x_* is the unique minimum of V (in other words, z_* is the unique minimum of the Hamiltonian $V(x) + |y|^2/2 + |w|^2/2$).

3.2. Examples of interactions

3.2.1. Interaction potential

One of the most classical interaction mechanism is given by

$$b(z,\mu) \;=\; A\int_{\mathbb{R}^d}\nabla_z W(z,y)\mu(\mathrm{d}y) \tag{3.8}$$

where $W\in\mathcal{C}^1(\mathbb{R}^d\times\mathbb{R}^d)$ and A is a $d\times d$ constant matrix. For instance, in the case of the overdamped Langevin diffusion (see Sect. 3.1.1), typically $A=-I_d$, and in the case of the kinetic Langevin diffusion (with $d=2n$, see Sect. 3.1.2),

$$A \;=\; \begin{pmatrix} 0 & 0\\ -I_n & 0\end{pmatrix}$$

and W is given by $W((x,y),(x',y'))=\tilde{W}(x,x')$ for some $\tilde{W}\in\mathcal{C}^1(\mathbb{R}^n\times\mathbb{R}^n)$.

Proposition 3.8. *Suppose that $\nabla_z W$ is a Lipschitz function. Then b given by (3.8) satisfies Assumption* **(A)**.

Proof. Let $L>0$ be such that

$$|\nabla_z W(z,y)-\nabla_z W(z',y')| \;\leqslant\; L\left(|z-z'|+|y-y'|\right)$$

for all $z,z',y,y'\in\mathbb{R}^d$.

Let $z,y\in\mathbb{R}^d$ and $\mu,\nu\in\mathcal{P}_2(\mathbb{R}^d)$. Let (Z,Y) be a $\mathbb{W}_2$-optimal coupling of μ and ν. Then

$$\begin{aligned}|b(z,\mu)-b(y,\nu)| \;&=\; |\mathbb{E}\left(A\nabla_z W(z,Z)-A\nabla_z W(y,Y)\right)|\\ &\leqslant\; \|A\|L\left(|y-z|+\mathbb{E}\left(|Z-Y|\right)\right) \;\leqslant\; \|A\|L\left(|y-z|+\mathbb{W}_2(\nu,\mu)\right)\,.\end{aligned}$$

□

Remark 3.9. In most classical cases, $W(z,y)=\tilde{W}(z-y)$ with an even $\tilde{W}\in\mathcal{C}^1(\mathbb{R}^d)$. Then, $b(z,\delta_z)=0$ for all $z\in\mathbb{R}^d$, and thus $a(\lambda)+b(\lambda,\delta_\lambda)=0$ if and only if λ is a fixed point of $\dot{z}=a(z)$.

3.2.2. The Adaptive Biasing Potential algorithm

The standard adaptive biasing algorithms used in molecular dynamics do not use a repulsive bias which is isotropic in the whole space. Rather, they tend to bias only a small dimensional part of the space, described by so-called reaction coordinates or collective variables. Reaction coordinates are given by $\xi:\mathbb{R}^d\to\mathbb{R}^p$ with $p\ll d$ (typically, in a full-atom simulation, $d=3N$ with N the number of atoms which may be of order 10^6, while $p=1$ or 2). For instance, if z is the positions of N atoms, $\xi(z)$ may be the distance between two particular atoms. We refer to [17] for more details and motivations of the use and design of reaction coordinates, in particular for enhanced sampling with self-biasing processes. A key point is that good reaction coordinates are meant to encode most of the metastability of the system, namely, if all the statistically representative values of $\xi(z)$ along a trajectory have been visited then this should also be the case for z. For this reason, adaptive algorithms based on reaction coordinates are designed so that, at stationarity, the law of $\xi(z)$ is unimodal. This is the case for algorithms such as the metadynamics [1, 13, 14], Adaptive Biasing Force (ABF) [17] or Adaptive Biasing Potential (ABP) [3] methods.

Let us focus on the ABP method here. For a small $\varepsilon>0$, consider a Gaussian kernel $K_\varepsilon(x)=e^{-|x|^2/(2\varepsilon)}/(2\pi\varepsilon)^{p/2}$ on $\mathbb{R}^p$. If Z is a random variable on $\mathbb{R}^d$ with law μ, then a smooth approximation for

the law of $\xi(Z)$ is given by the density

$$\rho_\mu(x) = \int_{\mathbb{R}^d} K_\varepsilon\left(x - \xi(z)\right) \mu(\mathrm{d}z)\,.$$

The ABP algorithm (or more precisely its mean field limit) can then be written as the self-interacting process (1.3-McKean-interaction) with interaction given by

$$b(z,\mu) = \omega A \nabla \ln(\varepsilon' + \rho_\mu \circ \xi)(z),$$

where the matrix A is as in the previous section (depending whether we consider an overdamped or a kinetic Langevin dynamics), $\varepsilon' > 0$ is another small regularisation parameter and $\omega \in [0,1]$ parametrizes the strength of the bias. It is straightforward to check that b is Lipschitz continuous. However, it should be noticed that the Lipschitz constants depends on $\varepsilon, \varepsilon'$ and blow up as these regularization parameters vanish. Due to the condition $\rho > \kappa$ in **(A2)**, for a fixed value of ω, applying our results to the ABP algorithm prevents ε and ε' to be too small, which is a limitation in term of practical application of the algorithm.

3.2.3. Interaction between two different species

In [7], a two-species model in the form of a coupled system of nonlinear stochastic differential equations has been studied. It corresponds to the function b on $\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$ with $\mathbb{R}^d = \mathbb{R}^n \times \mathbb{R}^n$ of the form

$$b((x,y),\nu) = \begin{pmatrix} \int_{\mathbb{R}^d} \left(b_{11}(x-x') + b_{12}(x-y')\right) \nu(\mathrm{d}x', \mathrm{d}y') \\ \int_{\mathbb{R}^d} \left(b_{21}(y-x') + b_{12}(y-y')\right) \nu(\mathrm{d}x', \mathrm{d}y') \end{pmatrix}$$

where the functions $b_{ij} : \mathbb{R}^n \to \mathbb{R}^n$ are Lipschitz continuous. It follows that b is Lipschitz continuous, as in Section 3.2.1.

3.3. Kramers' law for linear processes

In this section, we gather some classical results which can be used to check Assumption **(K)**. Standard references for large deviation principles and Kramers' law are the books of Freidlin and Wentzell [11] and of Dembo and Zeitouni [6].

Following [6], we now define the quasi-potential associated to the process solution to Equation (1.6–equilibrium-interaction). First, for $x, y \in \mathbb{R}^d$ and $t > 0$, consider the cost function

$$V(x,y,t) = \inf \frac{1}{4} \int_0^t |u(s)|^2 \mathrm{d}s,$$

where the infimum runs over all $u \in L^2([0,t])$ such that $z_t = y$ where $(z_s)_{s\in[0,t]}$ is the solution of $z_s = x + \int_0^s (a(z_w) + b(z_w, \delta_\lambda) + Mu_w)dw$, $s \in [0,t]$. The quasi-potential for the stochastic process (1.6–equilibrium-interaction) is then defined as

$$\overline{V}(x) = \inf_{t>0} V(\lambda, x, t)\,.$$

In our framework, Theorem 5.7.11 of [6] reads as follows[1].

Theorem 3.10. *Under* **(A)**, *assume furthermore the following:*

[1]Pay attention that there is a difference between our setting and the one of [6]: they consider σ^2 to be the parameter that controls the rate of convergence of the LDP and we consider $\sigma^2/2$.

1. *The domain $\mathcal{D}$ is open, bounded and positively invariant for the vector field $a + b(\cdot, \delta_\lambda)$.*
2. *$H := \inf_{z\in\partial\mathcal{D}} \overline{V}(z) < \infty$.*
3. *There exists $M > 0$ and a function $T : \mathbb{R}_+ \to \mathbb{R}_+$ with $T(s) \to 0$ as $s \to 0$ such that, for all $\varepsilon > 0$ small enough and all $x, y \in \mathbb{R}^d$ with $|x - z| + |y - z| \leqslant \varepsilon$ for some $z \in \partial\mathcal{D} \cup \{\lambda\}$, there exists a function $u : [0, T(\varepsilon)] \to \mathbb{R}^d$ with $\|u\|_{L^2} \leqslant M$ such that $z_{T(\varepsilon)} = y$ where z solves $z_s = x + \int_0^s (a(z_w) + b(z_w, \delta_\lambda) + \Sigma u_w)dw$, $s \in [0, T(\varepsilon)]$.*

Then, for all $x \in \mathcal{D}$ and all $\delta > 0$:

$$\mathbb{P}_x\left(\exp\left(\frac{2}{\sigma^2}(H-\delta)\right) \leqslant \tilde{\tau}_\sigma(\mathcal{D}) \leqslant \exp\left(\frac{2}{\sigma^2}(H+\delta)\right)\right) \underset{\sigma\to 0}{\longrightarrow} 1\,. \tag{3.9}$$

Indeed, $\mathcal{D}$ being positively invariant for $a + b(\cdot, \delta_\lambda)$, from Lemma 2.2, we get that necessarily $\lambda \in \mathcal{D}$ and all solutions of $\dot{z} = a(z) + b(z, \delta_\lambda)$ converges to λ, which is required to apply Theorem 5.7.11 of [6]. The third condition of the theorem is a controllability assumption, which is in particular always satisfied if Σ is non-singular, see Exercise 5.7.29 of [6].

In order to get **(K)**, we need to apply Theorem 3.10 to the domains $\mathcal{D}_{u,\xi}$. In particular these approximations of $\mathcal{D}$ should be positively invariant, which can be done following the construction of Definition 2.1, Proposition 2.2 in [24]. The last thing to check is that $H_{u,\xi} = \inf_{z\in\partial\mathcal{D}_{u,\xi}} \overline{V}(z)$ converge to H as ξ vanishes, for $u \in \{e, i\}$. This follows from the continuity of $\bar{V}$ in the vicinity of ∂D, which is a consequence of the third assumption of Theorem 3.10, as proven in Lemma 5.7.8 of [6].

In the kinetic Langevin case (as in Sect. 3.1.2) the third assumption of Theorem 3.10 does not hold, and one would like to consider unbounded domains $\mathcal{D} = \mathcal{D}' \times \mathbb{R}^n$ where $\mathcal{D}'$ is bounded (*i.e.* we are interested in the exit time of the position of the kinetic process from a given domain). In the case where

$$a(x, y) + b((x, y), \delta_\lambda) = \begin{pmatrix} y \\ -c(x) - \gamma y \end{pmatrix}, \qquad M = \sqrt{2\gamma}\begin{pmatrix} 0 & 0 \\ 0 & I_n \end{pmatrix}, \tag{3.10}$$

and $\mathcal{D}'$ is a smooth bounded domain with $c(x) \cdot \mathrm{n}(x) < 0$ for all $x \in \partial\mathcal{D}'$ where n is the exterior normal to $\partial\mathcal{D}'$ (which, under **(A)**, necessarily implies that $\lambda \in \mathcal{D}' \times \mathbb{R}^n$; besides, necessarily $\lambda = (\lambda', 0)$ for some $\lambda' \in \mathbb{R}^n$), it is proven in [10] that, in particular if the initial condition is λ, then (3.9) holds with

$$H = \inf_{x\in\partial\mathcal{D}'} S(x)$$

where, for $x \in \mathbb{R}^n$,

$$S(x) = \inf\{I_{[0T]}(\varphi) :\ \varphi_0 = x_*,\ \varphi_T = x,\ \dot{\varphi}_0 = 0,\ T \geq 0,\ \varphi \in C_{0T}\}$$

$$I_{[0,T]}(\varphi) = \begin{cases} \frac{1}{4}\int_0^T |\ddot{\varphi}_t + \gamma\dot{\varphi}_t + c(\varphi_t)|^2 \,\mathrm{d}t, & \text{if } \dot{\varphi} \text{ is abs. cont.} \\ +\infty \quad \text{otherwise.} \end{cases}$$

Finally, notice that if the condition $c \cdot \mathrm{n} < 0$ is satisfied at the boundary of $\mathcal{D}'$, then families of domains $\mathcal{D}'_{u,\xi}$ for $u \in \{e, i\}$ and $\xi \geqslant 0$ are easily constructed to satisfy the same condition and to meet the requirements of **(K)**. Moreover, it is not difficult to see that S is continuous, from which $H_{i,\xi} := \inf_{x\in\mathcal{D}_{u,\xi}} S(x) \to H$ as ξ vanishes. As a consequence, in the framework presented here, **(K)** holds.

To conclude, let us recall two classical cases where the quasi-potential is explicit.

1. **The equilibrium elliptic case**. If $a + b(\cdot, \delta_\lambda) = -\nabla U$ for some $U \in \mathcal{C}^2(\mathbb{R}^d)$ and $\Sigma = I_d$, then it is well-known (see *e.g.* [11]) that $\overline{V} = U - U(\lambda)$ (notice that λ is necessarily the unique global minimum of U since all solutions to $\dot{z} = -\nabla U(z)$ converges to λ). In this case, in Theorem 1.3, $H = \inf_{z\in\partial\mathcal{D}} U - U(\lambda)$.

2. **The equilibrium kinetic case**. If $d = 2n$ and a and Σ are given by (3.10) with $c = \nabla U$ for some $U \in \mathcal{C}^2$, then $S = U - U(\lambda')$, as proven in [10]. Again, in this case, if $\mathcal{D} = \mathcal{D}' \times \mathbb{R}^n$ with $-\nabla U \cdot \mathrm{n} < 0$ at the boundary of $\mathcal{D}'$ then, in Theorem 1.3, $H = \inf_{z \in \partial \mathcal{D}'} U - U(\lambda')$.

Acknowledgements. This work is supported by the French ANR grant METANOLIN (ANR-19-CE40-0009). P. Monmarché acknowledges financial support by the French ANR grant SWIDIMS (ANR-20-CE40-0022). The research of M.H. Duong was supported by EPSRC Grants EP/V038516/1 and EP/W008041/1. M. Tomašević was supported by *Fondation Mathématique Jacques Hadamard*.

References

[1] A. Barducci, G. Bussi and M. Parrinello, Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **100** (2008) 020603.
[2] M. Benaïm and C.-E. Bréhier, Convergence analysis of adaptive biasing potential methods for diffusion processes. *Commun. Math. Sci.* **17** (2019) 81–130.
[3] M. Benaïm, C.-E. Bréhier and P. Monmarché, Analysis of an Adaptive Biasing Force method based on self-interacting dynamics. *Electron. J. Probab.* (2020), in press.
[4] M. Ceriotti, G. Bussi and M. Parrinello, Colored-noise thermostats á la carte. *J. Chem. Theory Comput.* **6** (2010) 1170–1180.
[5] M.V. Day, On the exponential exit law in the small parameter exit problem. *Stochastics* **8** (1983) 297–323.
[6] A. Dembo and O. Zeitouni, Large Deviations Techniques and Applications, Vol. 38 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin (2010). Corrected reprint of the second (1998) edition.
[7] M.H. Duong and J. Tugaut, Coupled McKean–Vlasov diffusions: wellposedness, propagation of chaos and invariant measures. *Stochastics* **92** (2020) 900–943.
[8] V. Ehrlacher, T. Lelièvre and P. Monmarché. Adaptive force biasing algorithms: new convergence results and tensor approximations of the bias. Working paper or preprint, July 2020.
[9] G. Fort, B. Jourdain, T. Lelièvre and G. Stoltz, Convergence and efficiency of adaptive importance sampling techniques with partial biasing. *J. Stat. Phys.* **171** (2018) 220–268.
[10] M.I. Freidlin, Some remarks on the Smoluchowski-Kramers approximation. *J. Stat. Phys.* **117** (2004) 617–634.
[11] M.I. Freidlin and A.D. Wentzell, Random Perturbations of Dynamical Systems, Vol. 260 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*, 2nd edn. Springer-Verlag, New York (1998). Translated from the 1979 Russian original by Joseph Szücs.
[12] R.A. Holley, S. Kusuoka and D.W. Stroock, Asymptotics of the spectral gap with applications to the theory of simulated annealing. *J. Funct. Anal.* **83** (1989) 333–347.
[13] B. Jourdain, T. Lelièvre and P.-A. Zitt, Convergence of metadynamics: discussion of the adiabatic hypothesis. arXiv preprint arXiv:1904.08667, 2019.
[14] A. Laio and M. Parrinello, Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.* **99** (2002) 12562–12566.
[15] B. Leimkuhler and M. Sachs, Efficient Numerical Algorithms for the Generalized Langevin Equation. arXiv e-prints, page arXiv:2012.04245, December 2020.
[16] T. Lelièvre, Two mathematical tools to analyze metastable stochastic processes, in *Numerical Mathematics and Advanced Applications 2011*. Springer, Heidelberg (2013) 791–810.
[17] T. Lelièvre, M. Rousset and G. Stoltz, Long-time convergence of an adaptive biasing force method. *Nonlinearity* **21** (2008) 1155–1181.
[18] T. Lelièvre, M. Rousset and G. Stoltz, Free Energy Computations: A Mathematical Perspective. Imperial College Press (2010).
[19] T. Lelièvre and G. Stoltz, Partial differential equations and stochastic methods in molecular dynamics. *Acta Numerica* **25** (2016) 681–880.
[20] P. Monmarché, Almost sure contraction for diffusions on $\mathbb{R}^d$. Application to generalised Langevin diffusions. arXiv e-prints, page arXiv:2009.10828, September 2020.
[21] M. Ottobre and G.A. Pavliotis, Asymptotic analysis for the generalized Langevin equation. *Nonlinearity* **24** (2011) 1629–1653.
[22] L. Stella, C.D. Lorenz and L. Kantorovich, Generalized langevin equation: An efficient approach to nonequilibrium molecular dynamics of open systems. *Phys. Rev. B* **89** (2014) 134303.
[23] D.W. Stroock and S.R. Srinivasa Varadhan, Multidimensional Diffusion Processes, Vol. 233 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin-New York (1979).
[24] J. Tugaut, Exit problem of McKean-Vlasov diffusions in convex landscapes. *Electron. J. Probab.* **17** (2012) 26.
[25] J. Tugaut, A simple proof of a Kramers' type law for self-stabilizing diffusions. *Electron. Commun. Probab.* **21** (2016) 7.