



**HAL**  
open science

## On the exit-problem for self-interacting diffusions

Ashot Aleksian, Pierre del Moral, Aline Kurtzmann, Julian Tugaut

► **To cite this version:**

Ashot Aleksian, Pierre del Moral, Aline Kurtzmann, Julian Tugaut. On the exit-problem for self-interacting diffusions. 2022. hal-01901145v2

**HAL Id: hal-01901145**

**<https://hal.science/hal-01901145v2>**

Preprint submitted on 25 Jan 2022 (v2), last revised 25 Mar 2023 (v3)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# On the exit-problem for self-interacting diffusions

Ashot Aleksian<sup>1</sup>, Pierre Del Moral<sup>2</sup>, Aline Kurtzmann<sup>3</sup>, and Julian Tugaut<sup>1</sup>

<sup>1</sup>Université Jean Monnet, Institut Camille Jordan, 23, rue du docteur Paul Michelon, CS 82301, 42023 Saint-Étienne Cedex 2, France.

<sup>2</sup>INRIA Bordeaux Research Center, France.

<sup>3</sup>Université de Lorraine, Institut Elie Cartan de Lorraine, CNRS, Institut Elie Cartan de Lorraine, UMR 7502, Vandoeuvre-lès-Nancy, F-54506, France.

January 25, 2022

## Abstract

We study the exit-time from a domain of a self-interacting diffusion, where the Brownian motion is replaced by  $\sigma B_t$  for a constant  $\sigma$ . The first part of this work consists in showing that the rate of convergence (of the occupation measure of the self-interacting process toward some explicit Gibbs measure) previously obtained in [KK12] for a convex confinement potential  $V$  and a convex interaction potential can be bounded uniformly with respect to  $\sigma$ . Then, we prove an Arrhenius-type law for the first exit-time from a domain (satisfying classical hypotheses of Freidlin-Wentzell theory).

*Keywords* : Self-interacting diffusion, exit-time, Kramers' law, deterministic flow.

*Mathematics Subject Classification* : 60K35, 60H10

## 1 Introduction

Path-interaction processes have been introduced by Norris, Rogers and Williams during the late 80s in [NRW87]. Since this period, they have been an intensive research area. Under the name of Brownian Polymers, Durrett and Rogers [DR92] studied a family of self-interacting diffusions, as a model for the shape of a growing polymer. Denoting by  $X_t$  the location of the end of the growing polymer at time  $t$ , the process  $X$  satisfies a stochastic differential equation

driven by a Brownian motion, with a drift term depending on its own occupation measure. One is then interested in finding the scale for which the process converges to a non trivial limit. Later, another model of growing polymer has been introduced by Benaïm, Ledoux and Raimond [BLR02], for which the drift term depends on its own empirical measure. Namely, they have studied the following process living in a compact smooth connected Riemannian manifold  $M$  without boundary:

$$dX_t = \sum_{i=1}^N F_i(X_t) \circ dB_t^i - \int_M \nabla_x W(X_t, y) \mu_t(dy) dt,$$

where  $W$  is a (smooth) interaction potential,  $(B^1, \dots, B^N)$  is a standard Brownian motion on  $\mathbb{R}^N$ ,  $\mu_t = \frac{1}{t} \int_0^t \delta_{X_s} ds$  and the symbol  $\circ$  stands for the Stratonovich stochastic integration. In the compact setting, they have shown that the asymptotic behaviour of the empirical measure of the process can be related to the analysis of some deterministic dynamical flow. Later, Benaïm and Raimond [BR05] gave sufficient conditions for the almost sure convergence of the empirical measure (again in the compact setting). More recently, Raimond [Rai09] has generalized the previous study and has proved that for the solution of the SDE living on a compact manifold

$$dX_t = dB_t - \frac{g(t)}{t} \int_0^t \nabla_x V(X_t, X_s) ds dt$$

unless  $g$  is constant, the approximation of the empirical measure by a deterministic flow is no longer valid.

Similar questions have also been answered in the non-compact setting, that is  $\mathbb{R}^d$ . Chambeu and Kurtzmann [CK11] have studied the ergodic behaviour of the self-interacting diffusion depending on the empirical mean of the process. They have proved, under some convexity assumptions (ensuring the non-explosion in finite time of the process), a convergence criterion for the diffusion solution to the SDE

$$dX_t = dB_t - g(t) \nabla V \left( X_t - \frac{1}{t} \int_0^t X_s ds \right) dt$$

where  $g$  is a positive function. This model could represent for instance the behaviour of some social insects, as ants who are marking their paths with the trails' pheromones. This paper shows in particular how difficult is the study of general self-interacting diffusions in non-compact spaces as in [Kur10], driven by the generic equation

$$dX_t = dB_t - \frac{1}{t} \int_0^t \nabla_x V(X_t, X_s) ds dt.$$

Nevertheless, if the interaction function  $V$  is symmetric and uniformly convex, then Kleptsyn and Kurtzmann [KK12] obtained the limit-quotient ergodic theorem for the self-attracting diffusion. Moreover, they managed to obtain a speed

of convergence. As the results of this former paper are essential for the present work, we will explain them more precisely in §2.1.

Another problem related to this paper is the diffusion corresponding to McKean-Vlasov's PDE. This corresponds to the Markov process governed by the SDE

$$dX_t = dB_t - \nabla W * \nu_t(X_t)dt \quad (1.1)$$

where  $\nu_t = \mathcal{L}(X_t)$ ,  $W$  is a smooth convex potential and  $*$  stands for the convolution. The asymptotic behaviour of  $X$  has been studied by various authors these last years, see for instance Cattiaux, Guillin and Malrieu [CGM08]. It turns out that under some assumptions, the law  $\nu_t$  converges to the (unique if  $W$  is strictly convex) probability measure solution to the equation  $\nu = \frac{1}{Z} e^{-2W * \nu}$  where  $Z = Z(\nu)$  is the normalisation constant. In the latter paper, the authors use a particle system to prove both a convergence result (with convergence rate) and a deviation inequality for solutions of granular media equation when the interaction potential is uniformly convex at infinity. To this end, they use a uniform propagation of chaos property and a control in Wasserstein distance of solutions starting from different initial conditions.

A related question to this problem concerns the exit-times from domains of attraction for the following motion

$$dX_t^\sigma = \sigma dB_t - \nabla V(X_t^\sigma)dt - \nabla W * \nu_t(X_t^\sigma)dt \quad (1.2)$$

where  $V$  is a potential,  $*$  stands for the convolution,  $\nu_t = \mathcal{L}(X_t^\sigma)$  and  $\sigma > 0$ . This was addressed by Herrmann, Imkeller and Peithmann [HIP08], who exhibited a Kramers' type law for the particle's exit from the potential's domains of attraction and a large deviations principle for the self-stabilizing (also named McKean-Vlasov) diffusion. To get this, they reconstructed the Freidlin-Wentzell theory for the self-stabilizing diffusion. More precisely, they established a large deviations principle with a good rate function. The exit-problem for the McKean-Vlasov diffusion has also been already studied recently, without using the Freidlin-Wentzell method. In [Tug12], Tugaut has analysed the exit-problem (time and location) in convex landscapes, showing the same result as Herrmann, Imkeller and Peithmann, but without reconstructing the proofs of Freidlin and Wentzell. He has then generalised very recently his results in the case of double-wells landscape in [Tug18]. In [Tug16, Tug19], Tugaut did not use large deviations principle but a coupling method between the time-homogeneous diffusion

$$dX_t = \sigma dB_t - \nabla V(X_t)dt - \nabla W(X_t - m)dt$$

(where  $m$  is the unique point at which the vector field  $\nabla V$  equals 0) and the McKean-Vlasov diffusion so that the results on the exit-time of  $X$  can be used for the exit-time of the self-stabilizing diffusion (1.2).

The present paper also deals with the exit-time problem of a specific diffusion, driven by the SDE

$$dX_t = \sigma dB_t - \left( \nabla V(X_t) + \frac{1}{t} \int_0^t \nabla W(X_t - X_s) ds \right) dt, \quad X_0 = x \in \mathbb{R}^d \quad (1.3)$$

where  $V, W$  are two potentials and  $\sigma > 0$ . We could adapt the techniques introduced by Herrmann, Imkeller and Peithmann but only in the case of a convex gradient  $V$ . Our aim is to generalize the study also to non-convex potentials. In the present work, we will solve the exit-problem (time and location) for the diffusion (1.3). Indeed, the exit-location will be easily obtained once we know the exit-time. The motivation for the study of such a diffusion is twofold. First, we wish to obtain the basin of attraction when the diffusion converges (if we know the speed of convergence, or at least a nice upper bound of it). And more challenging, our main aim consists in improving the simulated annealing method (even if we need a concave interaction in that case). This paper is a first step in this direction.

In the following, for readability issue, we will omit the *sigma*-exponent for the process  $X$  as well as the empirical measure  $\mu_t$ . Nevertheless, the reader has to keep in mind that the process and  $\mu_t$  depend on  $\sigma$ .

## 1.1 Some useful notations

As usual, we denote by  $\mathcal{M}(\mathbb{R}^d)$  the space of signed (bounded) Borel measures on  $\mathbb{R}^d$  and by  $\mathcal{P}(\mathbb{R}^d)$  its subspace of probability measures. We will need the following measure space:

$$\mathcal{M}(\mathbb{R}^d; P) := \left\{ \mu \in \mathcal{M}(\mathbb{R}^d); \int_{\mathbb{R}^d} P(|y|) |\mu|(dy) < +\infty \right\}, \quad (1.4)$$

where  $|\mu|$  is the variation of  $\mu$  (that is  $|\mu| := \mu^+ + \mu^-$  with  $(\mu^+, \mu^-)$  the Hahn-Jordan decomposition of  $\mu$ :  $\mu = \mu^+ - \mu^-$ ) and  $P$  is some polynomial. Belonging to this space will enable us to always check the integrability of  $P$  (and therefore of  $V, W$  and their derivatives thanks to the domination condition (2.1)) with respect to the (random) measures to be considered. We endow this space with the dual weighted supremum norm (or dual  $P$ -norm) defined for  $\mu \in \mathcal{M}(\mathbb{R}^d; P)$  by

$$\|\mu\|_P := \sup_{\varphi \in \mathcal{C}(\mathbb{R}^d); |\varphi| \leq P} \left| \int_{\mathbb{R}^d} \varphi d\mu \right| = \int_{\mathbb{R}^d} P(|y|) |\mu|(dy), \quad (1.5)$$

where  $\mathcal{C}(\mathbb{R}^d)$  is the set of continuous functions  $\mathbb{R}^d \rightarrow \mathbb{R}$ . Without any loss of generality, we suppose that  $P(|x|) \geq 1$ , so that  $\|\mu\|_P \geq |\mu(\mathbb{R}^d)|$ . This norm makes  $\mathcal{M}(\mathbb{R}^d; P)$  a Banach space. Next, we consider  $\mathcal{P}(\mathbb{R}^d; P) = \mathcal{M}(\mathbb{R}^d; P) \cap \mathcal{P}(\mathbb{R}^d)$ . In the sequel,  $(\cdot, \cdot)$  stands for the Euclidean scalar product.

**Definition 1.1.** Let  $d$  be any positive integer. Let  $\mathcal{G}$  be a subset of  $\mathbb{R}^d$  and let  $U : \mathbb{R}^d \rightarrow \mathbb{R}^d$  be a vector field satisfying some “good assumptions”. For all  $x \in \mathbb{R}^d$ , we consider the dynamical system  $\rho_t(x) = x + \int_0^t U(\rho_s(x)) ds$ . We say that the domain  $\mathcal{G}$  is positively invariant for the flow generated by  $U$  if the orbit  $\{\rho_t(x); t \in \mathbb{R}_+\}$  is included in  $\mathcal{G}$  for all  $x \in \mathcal{G}$ .

## 1.2 Main results

The precise assumptions on the potentials will be given later in Section 2.

The goal of this paper consists in finding some precise upper and lower bounds for the exit-time of some positively invariant domain.

**Theorem 1.2.** *Let  $\mathcal{D}$  be a domain that is positively invariant for the flow  $x \mapsto -\nabla V(x) - \nabla W(x - m)$  and denote by  $\tau$  the first time the process  $X$  exits the domain  $\mathcal{D}$ . Let  $H := \inf_{x \in \partial \mathcal{D}} (V(x) + W(x - m) - V(m))$  be the exit cost from  $\mathcal{D}$ . Then  $\mathbb{P} - \lim_{\sigma \rightarrow 0} \frac{\sigma^2}{2} \log(\tau) = H$  that is for any  $\delta > 0$ , we have*

$$\lim_{\sigma \rightarrow 0} \mathbb{P} \left( \exp \left\{ \frac{2}{\sigma^2} (H - \delta) \right\} \leq \tau \leq \exp \left\{ \frac{2}{\sigma^2} (H + \delta) \right\} \right) = 1. \quad (1.6)$$

From Theorem 1.2, we immediately obtain the classical statement on the exit-location.

**Corollary 1.3.** *Under the same assumptions as the ones of Theorem 1.2, if  $\mathcal{N} \subset \partial \mathcal{D}$  is such that  $\inf_{z \in \mathcal{N}} (V(z) + W(z - m) - V(m)) > H$ , then*

$$\lim_{\sigma \rightarrow 0} \mathbb{P}(X_\tau \in \mathcal{N}) = 0. \quad (1.7)$$

### 1.3 Outline

Our paper is divided in two parts. First, Section 2 is devoted to the study of the self-interacting diffusion and more specifically, we will explain the former results of Kleptsyn and Kurtzmann [KK12] and, more precisely, how we adapt them in our context. In particular, we will show that the empirical measure of the studied process converges almost surely with a rate upper bounded independently of  $\sigma$ . After that, we will prove our main result, that is Theorem 1.2, in Section 3. To this aim, we will show that our process  $X$  is close to the solution of a given deterministic flow in §3.1. We then prove in §3.2 that the probability of leaving a positively invariant domain before the empirical mean remains stuck in the ball of center  $m$  and radius  $\kappa$  vanishes as  $\sigma$  goes to zero. Finally, a coupling permits to conclude the proof of the main theorem in §3.4 and we give the proof of the corollary in §3.5.

## 2 The self-interacting diffusion

We assume that  $V$  and  $W$  satisfy the following set of hypotheses denoted by **(H)**:

- i) (*regularity and positivity*)  $V \in \mathcal{C}^2(\mathbb{R}^d)$ ,  $W \in \mathcal{C}^2(\mathbb{R}^d)$  and  $V \geq 0$ ,  $W \geq 0$ ;
- ii) (*growth*)  $V$  and  $W$  have at most a polynomial growth: for some polynomial  $P$ , such that  $P(|x|) \geq 1$  for any  $x \in \mathbb{R}^d$ , we have  $\forall x \in \mathbb{R}^d$ :

$$|W(x)| + |\nabla W(x)| + \|\nabla^2 W(x)\| + |V(x)| + |\nabla V(x)| + \|\nabla^2 V(x)\| \leq P(|x|) \quad (2.1)$$

and  $\inf \nabla^2 V \geq \rho > 0$ ,  $\inf \nabla^2 W \geq \alpha > 0$ ,

$$\Delta V(x) \leq aV(x) \text{ and } \lim_{|x| \rightarrow \infty} \frac{|\nabla V(x)|^2}{V(x)} = \infty. \quad (2.2)$$

iv) (*curvature*)  $V$  and  $W$  are uniformly strictly convex functions.  $V$  has a unique minimum in  $m$ ;

v) (*spherical symmetry*)  $W(x) = G(|x|)$  for some function  $G$  from  $\mathbb{R}_+$  to  $\mathbb{R}$ .

**Remark 2.1.** By the growth condition (2.2),  $|\nabla V|^2 - \Delta V$  is bounded by below.

**Remark 2.2.** Note, that without loss of generality we can choose polynomials  $P$  to be of the form  $P(|x|) = C(1 + |x|^k)$ . Then the following property holds: there exists a constant  $\gamma > 0$  such that  $P(|x + y|) \leq \gamma(P(|x|) + P(|y|))$ .

Let us first show existence and uniqueness of the solution to the latter equation.

**Proposition 2.3.** *For any  $x \in \mathbb{R}^d$ , there exists a unique global strong solution  $(X_t, t \geq 0)$ .*

*Proof.* Local existence and uniqueness of the solution to (1.3) is standard (see for instance [DR92]). We only need to prove here that  $X$  does not explode in a finite time. Let us introduce the increasing sequence of stopping times  $\tau_0 = 0$  and

$$\tau_n := \inf \left\{ t \geq \tau_{n-1}; \mathcal{E}_t(X_t) + \int_0^t |\nabla \mathcal{E}_s(X_s)|^2 ds > n \right\}$$

where  $\mathcal{E}_t(X_t) := V(X_t) + \frac{1}{t} \int_0^t W(X_t - X_s) ds$ . In order to show that the solution never explodes, we use the Lyapunov functional  $(x, t) \mapsto \mathcal{E}_t(x)$ . As the process  $(t, x) \mapsto \mathcal{E}_{\mu_t}(x)$  is of class  $\mathcal{C}^2$  (in the space variable) and is a  $\mathcal{C}^1$ -semi-martingale (in the time variable), Itô-Ventzell formula applied to  $(x, t) \mapsto \mathcal{E}_{t \wedge \tau_n}(x)$  implies

$$\begin{aligned} \mathcal{E}_{t \wedge \tau_n}(X_{t \wedge \tau_n}) &= V(x) + \int_0^{t \wedge \tau_n} (\nabla \mathcal{E}_s(X_s), dB_s) - \int_0^{t \wedge \tau_n} |\nabla \mathcal{E}_s(X_s)|^2 ds \\ &\quad + \frac{\sigma^2}{2} \int_0^{t \wedge \tau_n} \Delta \mathcal{E}_s(X_s) ds - \int_0^{t \wedge \tau_n} \int_0^s W(X_s - X_u) du \frac{ds}{s^2}. \end{aligned} \quad (2.3)$$

We note that  $\int_0^{t \wedge \tau_n} (\nabla \mathcal{E}_s(X_s), dB_s)$  is a true martingale. We then get

$$\mathbb{E} \mathcal{E}_{t \wedge \tau_n}(X_{t \wedge \tau_n}) \leq V(x) + a \int_0^t \mathbb{E} \mathcal{E}_{s \wedge \tau_n}(X_{s \wedge \tau_n}) ds.$$

So, Gronwall's lemma leads to:

$$\mathbb{E} V(X_{t \wedge \tau_n}) \leq \mathbb{E} \mathcal{E}_{t \wedge \tau_n}(X_{t \wedge \tau_n}) \leq V(x) e^{at}.$$

As  $\lim_{|x| \rightarrow \infty} V(x) = \infty$ , the process  $(X_t, t \geq 0)$  does not explode in a finite time and there exists a global strong solution.  $\square$

**Remark 2.4.** A large family of path-dependent process has been studied by Saporito, see for instance [JS17]. He proves, with his co-authors, existence and uniqueness of such processes. The difference with our process is that we normalize the occupation measure.

## 2.1 Speed of convergence

Our results are based on the paper of Kleptsyn and Kurtzmann [KK12]. More precisely, they have proved the following

**Theorem 2.5.** [KK12, Theorem 1.6] *Let  $X$  be the solution to the equation (1.3) with  $\sigma = \sqrt{2}$ .*

*Suppose, that  $V \in \mathcal{C}^2(\mathbb{R}^d)$  and  $W \in \mathcal{C}^2(\mathbb{R}^d)$ , satisfy **(H)** and either  $V$  or  $W$  is uniformly strictly convex that is there exists  $C > 0$  such that*

$$\forall x \in \mathbb{R}^d, \forall v \in \mathbb{S}^{d-1}, \left. \frac{\partial^2 V}{\partial v^2} \right|_x \geq C \quad \text{or} \quad \forall x \in \mathbb{R}^d, \forall v \in \mathbb{S}^{d-1}, \left. \frac{\partial^2 W}{\partial v^2} \right|_x \geq C.$$

*Then there exists a unique density  $\rho_\infty : \mathbb{R}^d \rightarrow \mathbb{R}_+$ , such that almost surely*

$$\mu_t = \frac{1}{t} \int_0^t \delta_{X_s} ds \xrightarrow[t \rightarrow +\infty]{*-weakly} \rho_\infty(x) dx.$$

Moreover, if  $V$  is symmetric with respect to some point  $q$ , then the corresponding density  $\rho_\infty$  is also symmetric with respect to the same point  $q$ . Remark that the density  $\rho_\infty$  is the same limit density as in the result of [CMV03], uniquely defined by the following property:  $\rho_\infty$  is a positive function, proportional to  $e^{-(V+W*\rho_\infty)}$ .

And what is more important, Kleptsyn and Kurtzmann obtained a speed of convergence in the following way. First, let us recall the definition of the Wasserstein distance.

**Definition 2.6.** For  $\mu_1, \mu_2 \in \mathcal{P}(\mathbb{R}^d; P)$ , the quadratic Wasserstein distance is defined as

$$\mathbb{W}_2(\mu_1, \mu_2) := (\inf\{\mathbb{E}(|\xi_1 - \xi_2|^2)\})^{1/2},$$

where the infimum is taken over all the random variables such that  $\{\text{law of } \xi_1\} = \mu_1$  and  $\{\text{law of } \xi_2\} = \mu_2$ . This corresponds to the minimal  $L^2$ -distance taken over all the couplings between  $\mu_1$  and  $\mu_2$ .

Similarly, the Wasserstein distance  $\mathbb{W}_{2k}$  is defined as

$$\mathbb{W}_{2k}(\mu_1, \mu_2) := (\inf\{\mathbb{E}(|\xi_1 - \xi_2|^{2k})\})^{1/(2k)}.$$

More precisely, in [KK12, Theorem 1.12], it is proved the existence of a constant  $a > 0$  such that almost surely, for  $t$  large enough one has

$$\mathbb{W}_2(\mu_t^c, \rho_\infty) = O(\exp\{-a^{2k+1}\sqrt{\log t}\}),$$

where  $2k$  is the degree of the polynomial  $P$ ,  $\mu_t^c$  is the translation of the empirical measure  $\mu_t$  such that  $\mathbb{E}(\mu_t^c) = 0$  and  $\mathbb{W}_2$  is the quadratic Wasserstein distance.

Of course, in the case  $W(x) = \alpha \frac{|x|^2}{2}$ , the polynomial  $P$  corresponds to the growth of  $V$ , and we have to replace the Brownian motion by the rescaled Brownian motion  $\sigma B_t$ . So that the density  $\rho_\infty$  is uniquely defined by the following



property:  $\rho_\infty$  is a positive function, proportional to  $e^{-2(V+W*\rho_\infty)/\sigma^2}$ . Let us sketch the proof of these results and explain how  $\sigma$  appears here.

First, note that the empirical measure  $\mu_t = \frac{1}{t} \int_0^t \delta_{X_s} ds$  evolves very slowly. Indeed, choose a deterministic sequence of times  $T_n \rightarrow +\infty$ , with  $T_n \gg T_{n+1} - T_n \gg 1$ , and consider the behaviour of the measures  $\mu_{T_n}$ . As  $T_n \gg T_{n+1} - T_n$ , it is natural to expect that the empirical measures  $\mu_t$  on the interval  $[T_n, T_{n+1}]$  almost do not change and thus stay close to  $\mu_{T_n}$ . So we can approximate, on this interval, the solution  $X_t$  of (1.3) with  $\sigma = \sqrt{2}$  by the solution of the same equation with  $\mu_t \equiv \mu_{T_n}$ :

$$dY_t = \sigma dB_t - (\nabla V + \nabla W * \mu_{T_n})(Y_t) dt, \quad t \in [T_n, T_{n+1}],$$

in other words, by a Brownian motion in a potential  $V + W * \mu_{T_n}$  that does not depend on time.

On the other hand, the series of general term  $T_{n+1} - T_n$  increases. So, using Birkhoff Ergodic Theorem, we see that the (normalized) distribution  $\mu_{[T_n, T_{n+1}]}$  of values of  $X_t$  on these intervals becomes (as  $n$  increases) close to the equilibrium measures  $\Pi(\mu_{T_n})$  for a Brownian motion in the potential  $V + W * \mu_{T_n}$ , where

$$\Pi(\mu)(dx) := \frac{1}{Z(\mu, \sigma)} e^{-2(V+W*\mu)(x)/\sigma^2} dx, \quad Z(\mu, \sigma) := \int_{\mathbb{R}^d} e^{-2(V+W*\mu)(x)/\sigma^2} dx.$$

As

$$\mu_{T_{n+1}} = \frac{T_n}{T_{n+1}} \mu_{T_n} + \frac{T_{n+1} - T_n}{T_{n+1}} \mu_{[T_n, T_{n+1}]},$$

we then have

$$\mu_{T_{n+1}} \approx \frac{T_n}{T_{n+1}} \mu_{T_n} + \frac{T_{n+1} - T_n}{T_{n+1}} \Pi(\mu_{T_n}) = \mu_{T_n} + \frac{T_{n+1} - T_n}{T_{n+1}} (\Pi(\mu_{T_n}) - \mu_{T_n}),$$

and

$$\frac{\mu_{T_{n+1}} - \mu_{T_n}}{T_{n+1} - T_n} \approx \frac{1}{T_{n+1}} (\Pi(\mu_{T_n}) - \mu_{T_n}).$$

This motivates Kleptsyn and Kurtzmann to approximate the behaviour of the measures  $\mu_t$  by trajectories of the flow (on the infinite-dimensional space of measures)

$$\dot{\mu} = \frac{1}{t} (\Pi(\mu) - \mu), \quad (2.4)$$

or after a logarithmic change of variable  $\theta = \log t$ ,

$$\mu' = \Pi(\mu) - \mu. \quad (2.5)$$

Indeed, choose an appropriate interval  $[T_n, T_{n+1}]$ . On this interval, fix the empirical measure  $\mu_t$  at  $\mu_{T_n}$ . Then construct a new process  $Y$ , coupled with  $X$  (the coupling is such that  $X$  and  $Y$  are driven by the same Brownian motion), such that for all  $t \in [T_n, T_{n+1}]$ , we have

$$dY_t = \sigma dB_t - (\nabla V(Y_t) + \nabla W * \mu_{T_n}(Y_t)) dt.$$

This new process has two advantages. First, it is Markovian (and its invariant probability measure is  $\Pi(\mu_{T_n})(dx) = \frac{1}{Z} e^{-2(V+W*\mu_{T_n})(x)/\sigma^2} dx$ ), and so is easier than  $X$  to study. Second, its evolution is very close to the evolution of the desired  $X$ . Indeed, we will use  $Y$  to prove that the transport distance between the empirical measure on  $[T_n, T_{n+1}]$ , denoted by  $\mu_{[T_n, T_{n+1}]}$  in [KK12, Proposition 3.2], and the probability measure  $\Pi(\mu_{T_n})$  (both measures being centered in  $c_{T_n}$ ) is controlled by  $T_n^{-\frac{1}{3} \min(8C_W, 1/5d)}$  and so, this distance vanishes as  $n \rightarrow +\infty$ . This has been done in [KK12, §3.1.1].

After that, remark that if a.s. the empirical measure  $\mu_t$  converges weakly\* to  $\mu_\infty$ , then for  $t$  large enough, the process  $X$  shall be very close to  $Z$ , defined by

$$dZ_t = \sigma dB_t - (\nabla V + \nabla W * \mu_\infty)(Z_t) dt.$$

The process  $Z$  is obviously Markovian and the ergodic theorem can be applied:

$$\frac{1}{t} \int_0^t \delta_{Z_s} ds \xrightarrow[t \rightarrow +\infty]{} \Pi(\mu_\infty) \quad \text{a.s.}$$

for the weak\* convergence of measures. So when the limit  $\mu_\infty$  exists, it satisfies

$$\mu_\infty = \Pi(\mu_\infty). \quad (2.6)$$

This explains the idea of introducing the dynamical system  $\dot{\mu} = \Pi(\mu) - \mu$  (after the time-shift  $t \mapsto e^t$  in order to work with a time-homogeneous system) defined on the set of probability measures that are integrable for the polynomial  $P$ . Note that, instead of considering the latter dynamical system, Kleptsyn and Kurtzmann work with its discretized version, with the knots chosen at the moments  $T_n$ . They then prove, in [KK12, Proposition 3.5], that the transport distance between the deterministic trajectory induced by the smoothed (discrete) dynamical system and the (centered) random trajectory  $\mu_{T_n}$  is controlled and decreases to 0. This has been done in [KK12, §3.1.2].

Next, it remains to show that the free energy between this (centered) deterministic trajectory and the set of translates of  $\rho_\infty$  goes to 0. We recall that for the dynamics in presence of an exterior potential  $V$ , the free energy function is

$$\mathcal{F}_{V,W}(\mu) := -\frac{\sigma^2}{2} \mathcal{H}(\mu) + \int_{\mathbb{R}^d} V(x) \mu(x) dx + \frac{1}{2} \iint_{\mathbb{R}^d \times \mathbb{R}^d} \mu(x) W(x-y) \mu(y) dx dy.$$

and consider  $\mathcal{F}_{V+W*\mu} = -\frac{\sigma^2}{2} \mathcal{H}(\mu) + \int_{\mathbb{R}^d} (V(x) + W*\mu(x)) \mu(x) dx$  for the energy of “small parts”, where the entropy of the measure  $\mu$  is

$$\mathcal{H}(\mu) := - \int_{\mathbb{R}^d} \mu(x) \log \mu(x) dx. \quad (2.7)$$

As the free energy is controlled by the quadratic Wasserstein distance  $\mathbb{W}_2$ , this implies that the transport distance between the two previous quantities decreases, as asserted in [KK12, Proposition 3.6].

To conclude, it remains to put all the pieces together and use the triangle inequality:  $\mathbb{W}_2(\mu_t^c, \rho_\infty)$  is upper bounded by the sum of three distances, involving the flow  $\Phi_n$  induced by the discretization of the dynamical system  $\dot{\mu} = \Pi(\mu) - \mu$  on the interval  $[T_n, T_{n+1})$ , for  $n$  large enough. The first term of the summation bound will be  $\mathbb{W}_2(\mu_t^c, \Pi(\mu_{T_n}^c))$ , the second one  $\mathbb{W}_2(\Pi(\mu_{T_n}^c), \Phi_n^n(\mu_{T_n}^c))$  and the third one  $\mathbb{W}_2(\Phi_n^n(\mu_{T_n}^c), \rho_\infty)$ .

Finally, the previous decrease estimates will allow Kleptsyn and Kurtzmann to show the convergence of the center, after having made the appropriate choice  $T_n = n^{3/2}$ .

## 2.2 The speed of convergence for the solution of (1.3)

For this paper, the corresponding result of [KK12, Theorem 1.12] is the following

**Theorem 2.7.** *Let  $X$  be the solution to the equation (1.3). There exists a constant  $a > 0$  such that almost surely, we have for  $t$  large enough  $\mathbb{W}_{2k}(\mu_t, \rho_\infty) = O(\exp\{-a^{2k+1}\sqrt{\log t}\})$ , where  $2k$  is the degree of the polynomial  $P$ ,  $\mu_t$  is the empirical measure and  $\mathbb{W}_{2k}$  is the  $2k$ -Wasserstein distance, that is the minimal  $L^{2k}$ -distance taken over all the couplings between  $\mu_t$  and  $\rho_\infty$  that is the unique density proportional to  $e^{-\frac{2}{\sigma^2}(V+W*\rho_\infty)}$ .*

To prove this result, we mimic the proof of [KK12, Theorem 1.12], and show that the speed of convergence is less than the one of the case corresponding to  $\sigma = 1$ . We will not reproduce it here. We will only show how we handle this inequality for the following particular result, as it is representative of the difficulty and shows how  $\sigma$  appears in the calculation.

Let us prove for instance the exponential decrease for the centered measure  $\Pi(\mu)$  and show that, as  $\sigma^2 \ll 1$ , we can obtain a lower bound of the speed of convergence that does not depend on  $\sigma$ . We have seen previously that the centered  $\mu^c$  has the same asymptotic behaviour as  $\Pi(\mu)(\cdot + c_\mu)$ , up to a time-scale. An important step to prove the latter result consists in estimating the behaviour of the centered measures  $\mu_t^c$ . More precisely, one has to prove that the tail of these measures are exponentially decreasing. This is why we have to introduce the following sets:

**Definition 2.8.** Let  $\alpha, C > 0$  be given. Define

$$K_{\alpha, C}^0 := \{\mu \in \mathcal{P}(\mathbb{R}^d); \quad \forall R > 0, \mu(\{y; |y| > R\}) < Ce^{-\alpha R}\}, \quad (2.8)$$

$$K_{\alpha, C} := \{\mu \in \mathcal{P}(\mathbb{R}^d); \quad \mu^c \in K_{\alpha, C}^0\}. \quad (2.9)$$

We can now prove the exponential decrease of  $\Pi(\mu)(\cdot + c_\mu)$ .

**Proposition 2.9.** *There exist  $C_W, C_\Pi > 0$ , two constants independent of  $\sigma$ , such that for all  $\mu \in \mathcal{P}(\mathbb{R}; P)$ , we have  $\Pi(\mu)(\cdot + c_\mu) \in K_{C_W, C_\Pi}^0$  where  $c_\mu$  is defined by the equation  $(\nabla V + \nabla W * \mu)(c_\mu) = 0$ .*

*Proof.* Let us fix  $R > 0$ . Note first that, imposing a condition  $C_\Pi \geq e^{2C_W}$ , we can restrict ourselves only to  $R \geq 2$ : for  $R < 2$ , the estimate is obvious.

The measure  $\Pi(\mu)$  has the density  $\frac{1}{Z(\mu, \sigma)} e^{-2(V+W*\mu)(x)/\sigma^2}$ . To avoid working with the normalization constant  $Z(\mu, \sigma)$ , we will prove a stronger inequality, that is

$$\Pi(\mu)(|x - c_\mu| \geq R) \leq C_\Pi e^{-CR} \cdot \Pi(\mu)(|x - c_\mu| \leq 2), \quad (2.10)$$

which is equivalent to

$$\int_{|x-c_\mu| \geq R} e^{-2(V+W*\mu)(x)/\sigma^2} dx \leq C_\Pi e^{-CR} \int_{|x-c_\mu| \leq 2} e^{-2(V+W*\mu)(x)/\sigma^2} dx.$$

We use the polar coordinates, centered at the center  $c_\mu$ , and so we want to prove that

$$\begin{aligned} & \int_{\mathbb{S}^{d-1}} \int_R^{+\infty} e^{-2(V+W*\mu)(c_\mu+\lambda v)/\sigma^2} \lambda^{d-1} d\lambda dv \\ & \leq C_\Pi e^{-CR} \int_{\mathbb{S}^{d-1}} \int_0^2 e^{-2(V+W*\mu)(c_\mu+\lambda v)/\sigma^2} \lambda^{d-1} d\lambda dv. \end{aligned}$$

It suffices to prove such an inequality “directionwise”: for all  $v \in \mathbb{S}^{d-1}$ , for all  $R \geq 2$

$$\int_R^{+\infty} e^{-2(V+W*\mu)(c_\mu+\lambda v)/\sigma^2} \lambda^{d-1} d\lambda \leq C_\Pi e^{-CR} \int_0^2 e^{-2(V+W*\mu)(c_\mu+\lambda v)/\sigma^2} \lambda^{d-1} d\lambda.$$

But from the uniform convexity of  $V$  and  $W$  and the definition of the center, the function  $f(\lambda) = 2(V + W * \mu)(c_\mu + \lambda v)$  satisfies  $f'(0) = 0$  and  $\forall r > 0$ ,  $f''(r) \geq C$ . Hence,  $f$  is monotone increasing on  $[0, +\infty)$ , and in particular,

$$\int_0^2 e^{-f(\lambda)/\sigma^2} \lambda^{d-1} d\lambda \geq e^{-f(2)/\sigma^2} \int_0^2 \lambda^{d-1} d\lambda =: C_1 e^{-f(2)/\sigma^2}. \quad (2.11)$$

On the other hand, for all  $\lambda \geq 2$ ,  $f'(\lambda) \geq f'(2) \geq 2C$ , and thus  $f(\lambda) \geq 2C(\lambda - 2) + f(2)$ . Hence, as  $\sigma^2 \ll 1$ , we have

$$\begin{aligned} \int_R^{+\infty} e^{-f(\lambda)/\sigma^2} \lambda^{d-1} d\lambda & \leq e^{-f(2)/\sigma^2} \int_R^{+\infty} \lambda^{d-1} e^{-2C(\lambda-2)/\sigma^2} d\lambda \\ & \leq e^{-\frac{f(2)}{\sigma^2}} \int_R^{+\infty} \lambda^{d-1} e^{-2C(\lambda-2)} d\lambda. \end{aligned} \quad (2.12)$$

So finally, using that  $\sigma \ll 1$

$$\int_R^{+\infty} e^{-f(\lambda)/\sigma^2} \lambda^{d-1} d\lambda \leq C_2 R^{d-1} e^{-2CR} \cdot e^{-\frac{f(2)}{\sigma^2}} \leq C_3 e^{-CR} \cdot e^{-\frac{f(2)}{\sigma^2}}. \quad (2.13)$$

Comparing (2.11) and (2.13), we obtain the desired exponential decrease. Indeed, as  $\sigma$  only appears in the exponential under the form  $\exp(-const/\sigma^2)$ , we can upper bound the previous quantity by  $e^{-const}$ , so that the constant  $C_W$  is upper bounded by a constant that is independent of  $\sigma$ . By an abuse of notation, we call again this constant  $C_W$ .  $\square$

**Remark 2.10.** This result corresponds to [KK12, Proposition 2.9] for the case  $\sigma = 1$ .

### 2.3 Small-noise limit

In the following, we remind the reader that we don't emphasize the dependence on  $\sigma$ , but it will appear everywhere in the computations.

As was mentioned above, the invariant probability measure of self-interacting diffusion is the unique solution to the equation

$$\rho_\infty = \Pi(\rho_\infty),$$

where  $\Pi(\mu)(x) = e^{-2(V+W*\mu)(x)/\sigma^2} / \int e^{-2(V+W*\mu)(z)/\sigma^2} dz$ . The same invariant probability measure appears in the self-stabilizing diffusion, small-noise limit of which was studied in [HT10]. There, authors studied the case of double-wells potentials which is more general than our diffusion. In this paper the result, that can be transformed in our context as following, was proved. If the moments of invariant probability measures  $\rho_\infty$  are uniformly bounded with respect to  $\sigma$ , then  $\delta_m$  is the weak-\* limit of  $\rho_\infty$  with  $\sigma \rightarrow 0$  a.s. Note, that indeed, moments of  $\rho_t$  are uniformly bounded, since  $\mu_t \in K_{\alpha,C}$  for any  $t > 0$  and some  $\alpha, C$  that do not depend on  $\sigma$ .

Thus, consider the following deterministic time, representing the time of stabilization of the occupation measure, if it occurs, around its supposed limit  $\delta_m$ :

$$T_\kappa(\sigma) := \inf \{t_0 \geq 0 : \forall t \geq t_0, \mathbb{E}(\mathbb{W}_{2k}(\mu_t; \delta_m)) \leq \kappa\}. \quad (2.14)$$

First, let us discuss why expectation  $\mathbb{E}(\mathbb{W}_{2k}(\mu_t; \delta_m))$  exists in the first place. To show that, we use the fact that  $\mu_t \in K_{\alpha,C}$  almost surely and get

$$\mathbb{W}_{2k}(\mu_t; \delta_m) \leq \left(2^{2k-1} \int |x|^{2k} \mu_t(dx) + 2^{2k-1} |m|^{2k}\right)^{1/(2k)} \leq \text{Const},$$

where the last constant depends only on  $\alpha, C, m$  and  $k$ . Therefore, since the random variable is bounded by a constant almost surely, expectation exists.

Second, let us show that the definition of the time  $T_\kappa(\sigma)$  makes sense. Indeed,

$$\mathbb{E}\mathbb{W}_{2k}(\mu_t; \delta_m) \leq \mathbb{E}\mathbb{W}_{2k}(\mu_t; \rho_\infty) + \mathbb{E}\mathbb{W}_{2k}(\rho_\infty; \delta_m) \xrightarrow[\sigma \rightarrow 0]{t \rightarrow \infty} 0,$$

where the limit is not just iterated, but holds for the pair  $(t, \sigma)$ , since the speed of convergence of  $\mu_t$  towards  $\rho_\infty$  in time does not depend on  $\sigma$ , which was shown in Theorem 2.7. Therefore, for any  $\kappa > 0$  we can find  $\sigma_0$  small enough and  $t_0$  big enough such that  $T_\kappa(\sigma) < T_\kappa < \infty$  for any  $\sigma < \sigma_0$ , which does not only prove existence and finiteness of  $T_\kappa(\sigma)$ , but also its uniformness with respect to  $\sigma$ .

## 3 Proof of Theorem 1.2

In this section, we prove our main result. First, we show that the process  $X$  solution to (1.3) is close to the solution of a deterministic flow  $(\psi_t)_{t \geq 0}$  in §3.1.

Using that, we prove in §3.2 that the probability of leaving a positively invariant domain before the empirical mean remains stuck in the ball of center  $m$  and radius  $\kappa$  vanishes as  $\sigma$  goes to zero. Then, we consider the coupling between the studied diffusion and the one where the empirical measure is frozen to  $\delta_m$  and we show that these diffusions are close in §3.3. We conclude the proof in §3.4.

### 3.1 Upper bound

We remind the reader that in this work, the noise vanishes. Consequently, it is natural to introduce the deterministic flow  $(\psi_t)_t$  defined by the following

$$\dot{\psi}_t = -\nabla V(\psi_t) - \frac{1}{t} \int_0^t \nabla W(\psi_t - \psi_s) ds, \quad \psi_0 = x_0. \quad (3.1)$$

We will show that  $X_t$  and  $\psi_t$  are uniformly close while the noise goes to zero. Namely

**Proposition 3.1.** *For any  $\xi > 0$  and for any  $T > 0$ , we have:*

$$\lim_{\sigma \rightarrow 0} \mathbb{P} \left( \sup_{t \in [0; T]} \|X_t - \psi_t(x_0)\|^2 > \xi \right) = 0. \quad (3.2)$$

*Proof.* First of all, we fix some  $\xi$  and introduce the following stopping time  $\mathcal{T} := \inf\{t : |X_t^\sigma - \psi_t|^2 \geq \xi\}$ . We apply Itô formula and get the following result, for  $\omega \in \{\mathcal{T} > t\}$  (the choice of this event will be clear further) :

$$\begin{aligned} |X_t - \psi_t|^2 &= 2 \int_0^t (X_s - \psi_s, dX_s - d\psi_s) + d\sigma^2 t \\ &\leq d\sigma^2 t - 2C_V \int_0^t |X_s - \psi_s|^2 ds + \sigma \int_0^t (X_s - \psi_s, dB_s) \\ &\quad - \int_0^t \frac{1}{s} \int_0^s (X_s - \psi_s, \nabla W(X_s - X_z) - \nabla W(\psi_s - \psi_z)) dz ds. \end{aligned}$$

Let  $\text{Lip}_{\nabla W}^{K'}$  be a Lipschitz constant of  $\nabla W$  inside the following compact  $K' := \{x : |x - \psi_t|^2 \leq \xi, \text{ for some } t > 0\}$  (due to our assumptions, this set is indeed a compact at least for small  $\xi$ , which we can decrease without loss of generality), and  $C_V$  is the convexity constant of  $V$ . We thus have

$$\begin{aligned} |X_t - \psi_t|^2 &\leq d\sigma^2 t - 2C_V \int_0^t |X_s - \psi_s|^2 ds + \sigma \int_0^t (X_s - \psi_s, dB_s) \\ &\quad + \text{Lip}_{\nabla W}^{K'} \int_0^t \frac{1}{s} \int_0^s (|X_s - \psi_s|^2 + |X_s - \psi_s| \cdot |X_z - \psi_z|) dz \\ &\leq d\sigma^2 t - 2C_V \int_0^t |X_s - \psi_s|^2 ds + \sigma \int_0^t (X_s - \psi_s, dB_s) \quad (3.3) \\ &\quad + \frac{\text{Lip}_{\nabla W}^{K'}}{2} \int_0^t \frac{1}{s} \int_0^s (3|X_s - \psi_s|^2 + |X_z - \psi_z|^2) dz. \end{aligned}$$

Note then, that by the BDG inequality we get for some constant  $C > 0$ :

$$\begin{aligned} \mathbb{E} \left( \sup_{[0, t \wedge \mathcal{T}]} \left| \sigma \int_0^s (X_z - \psi_z, dB_z) \right| \right) &\leq C\sigma^2 \mathbb{E} \sqrt{\int_0^{t \wedge \mathcal{T}} |X_s - \psi_s|^2 ds} \\ &\leq C\sigma^2 \sqrt{\int_0^t \mathbb{E} \left( \sup_{z \in [0, s \wedge \mathcal{T}]} (|X_z - \psi_z|^2) \right) ds}. \end{aligned}$$

Let us consider the following random variable  $\sup_{s \in [0; t \wedge \mathcal{T}]} |X_s - \psi_s|^2$ . The fact that we consider the supremum before time  $t \wedge \mathcal{T}$  gives us that for any  $\omega$  we consider only such  $s$ , that  $s \leq \mathcal{T}(\omega)$ , which in turn means that we can apply estimation (3.3) for any  $s \in [0, t \wedge \mathcal{T}]$ . We also remind that  $t \leq T$  and derive:

$$\begin{aligned} \mathbb{E} \left( \sup_{s \in [0; t \wedge \mathcal{T}]} |X_s - \psi_s|^2 \right) &\leq d\sigma^2 T + C\sigma^2 \sqrt{\int_0^t \mathbb{E} \left( \sup_{z \in [0, s \wedge \mathcal{T}]} (|X_z - \psi_z|^2) \right) dz} \\ &\quad + 2\text{Lip}_{\nabla W}^{K'} \int_0^t \mathbb{E} \left( \sup_{z \in [0, s \wedge \mathcal{T}]} |X_z - \psi_z|^2 \right) ds \\ &\leq d\sigma^2 T + \frac{C\sigma^2}{2} \left[ 1 + T \mathbb{E} \left( \sup_{s \in [0, t \wedge \mathcal{T}]} (|X_s - \psi_s|^2) \right) \right] \\ &\quad + 2\text{Lip}_{\nabla W}^{K'} \int_0^t \mathbb{E} \left( \sup_{z \in [0, s \wedge \mathcal{T}]} |X_z - \psi_z|^2 \right) ds, \end{aligned}$$

where in the last inequality we used  $\sqrt{x} \leq (1+x)/2$ . Now, if we denote  $u_t := \mathbb{E} \left( \sup_{s \in [0; t \wedge \mathcal{T}]} |X_s - \psi_s|^2 \right)$ , we have

$$u_t \leq \frac{1}{1 - CT\sigma^2/2} \left( \frac{2Td + C}{2} \sigma^2 + 2\text{Lip}_{\nabla W}^{K'} \int_0^t u_s ds \right),$$

for small enough  $\sigma$  (such that  $1 - CT\sigma^2/2 > 0$ ). Thus, using Grönwall lemma, we get

$$u_t \leq \frac{(2Td + C)\sigma^2}{2(1 - CT\sigma^2/2)} \exp \left\{ \frac{\text{Lip}_{\nabla W}^{K'}}{1 - CT\sigma^2/2} T \right\} = O(\sigma^2). \quad (3.4)$$

This in particular means, that  $\mathbb{E} \left( \sup_{s \in [0; T \wedge \mathcal{T}]} |X_s - \psi_s|^2 \right) \leq O(\sigma^2)$ . Nevertheless, to show the necessary result, we have to get rid of the stopping time  $\mathcal{T}$  in the previous equation. It is sufficient to show, that  $\mathbb{P}(\mathcal{T} \leq T) \xrightarrow{\sigma \rightarrow 0} 0$ .

Indeed, by its definition,  $\mathcal{T}$  is the first time when the difference  $|X_t - \psi_t|^2$  reaches  $\xi$ . But under the assumption  $\mathcal{T} \leq T$  and due to (3.4), by decreasing  $\sigma$  we can control  $|X_t - \psi_t|^2$  and make it small enough, such that  $|X_{\mathcal{T}} - \psi_{\mathcal{T}}|^2 < \xi$  (in some sense), which contradicts the definition of  $\mathcal{T}$ . Rigorously,

$$\mathcal{T} < T \Rightarrow \sup_{[0, \mathcal{T} \wedge T]} |X_s - \psi_s|^2 = \sup_{[0, \mathcal{T}]} |X_s - \psi_s|^2 \geq \xi.$$

Thereby,

$$\mathbb{P}(\mathcal{T} < T) \leq \mathbb{P}\left(\sup_{[0, \mathcal{T} \wedge T]} |X_s - \psi_s|^2 \geq \xi\right) \leq O(\sigma^2),$$

by Markov inequality.

To conclude the proof of the Proposition, we consider

$$\begin{aligned} \mathbb{P}\left(\sup_{t \in [0; T]} |X_t - \psi_t(x_0)|^2 > \xi\right) &\leq \mathbb{P}\left(\sup_{t \in [0; T]} |X_t - \psi_t(x_0)|^2 > \xi, \mathcal{T} > T\right) \\ &\quad + \mathbb{P}(\mathcal{T} \leq T) \\ &\leq \mathbb{P}\left(\sup_{t \in [0; T \wedge \mathcal{T}]} |X_t - \psi_t(x_0)|^2 > \xi\right) + O(\sigma^2) \\ &\leq O(\sigma^2), \end{aligned}$$

by Markov inequality and (3.4), which completes the proof.  $\square$

### 3.2 Probability of leaving before $T_\kappa(\sigma)$

Remind that we denoted in (2.14) by  $T_\kappa(\sigma)$  the first time at which the expectation of the  $2k$ -Wasserstein distance between the occupation measure of the process and  $\delta_m$  is smaller than  $\kappa$ . By  $\mathbb{B}(m; \kappa)$ , we denote the ball of center  $\delta_m$  and radius  $\kappa$  for  $\mathbb{W}_{2k}$ .

**Proposition 3.2.** *We put  $\tau := \inf\{t \geq 0 : X_t \notin \mathcal{D}\}$  where  $\mathcal{D}$  is a domain that is positively invariant for the flow  $x \mapsto -\nabla V(x) - \nabla W(x - m)$ . For any  $\kappa > 0$ ,*

$$\lim_{\sigma \rightarrow 0} \mathbb{P}(\tau \leq T_\kappa(\sigma)) = 0. \quad (3.5)$$

*Proof.* First, we remind that if  $\sigma$  is small enough,  $T_\kappa(\sigma) \leq T_\kappa$  where  $T_\kappa$  does not depend on  $\sigma$ . Moreover,  $T_\kappa$  and  $T_\kappa(\sigma)$  are deterministic. As a consequence, we have

$$\mathbb{P}(\tau \leq T_\kappa(\sigma)) \leq \mathbb{P}(\tau \leq T_\kappa).$$

We now prove that for any  $T > 0$ ,  $\mathbb{P}(\tau \leq T) \rightarrow 0$  as  $\sigma$  goes to 0. According to the assumption  $\{\psi_t(x_0); 0 \leq t \leq T\} \subset \mathcal{D}$ , we know (since  $\mathcal{D}$  is an open set) that there exists  $\epsilon > 0$  such that  $\mathbb{B}(0; \epsilon) + \{\psi_t(x_0); 0 \leq t \leq T\} \subset \mathcal{D}$ .

Now, on the event  $\{\tau \leq T\}$ , we deduce that  $\sup_{t \in [0; T]} \|X_t - \psi_t(x_0)\|^2 > \epsilon^2$ . As a consequence:

$$\mathbb{P}(\tau \leq T) \leq \mathbb{P}\left(\sup_{t \in [0; T]} \|X_t - \psi_t(x_0)\|^2 > \epsilon^2\right),$$

which goes to 0 as  $\sigma$  goes to 0, thanks to (3.2). This concludes the proof.  $\square$



### 3.3 Coupling for $t \geq T_\kappa(\sigma)$

In [Tug16, Tug18], Tugaut has proved the Kramers' type law for the exit-time. He has used a coupling between the diffusion of interest ( $X$  here) and another diffusion that is expected to be close from  $X$  if the time is sufficiently large. The main difficulty with the considered self-stabilizing diffusion is in fact that we do not have a uniform (with respect to the time) control of the law.

Here, we have proved that the nonlinear quantity appearing in the equation (that is  $\frac{1}{t} \int_0^t \delta_{X_s} ds$ ) remains stuck - with high probability - in a small ball (for  $\mathbb{W}_{2k}$ ) of center  $\delta_m$  and radius  $\kappa$  for any  $t \geq T_\kappa(\sigma)$ . The idea is thus to replace  $\frac{1}{t} \int_0^t \delta_{X_s} ds$  by  $\delta_m$  and to compare the new diffusion with the self-interacting one.

In other words, we consider the diffusion

$$Y_t = X_{T_\kappa(\sigma)} + \sigma (B_t - B_{T_\kappa(\sigma)}) - \int_{T_\kappa(\sigma)}^t \nabla V(Y_s) ds - \int_{T_\kappa(\sigma)}^t \nabla W(Y_s - m) ds, \quad (3.6)$$

for any  $t \geq T_\kappa(\sigma)$  and  $Y_t = X_t$  if  $t \leq T_\kappa(\sigma)$ .

**Proposition 3.3.** *For any  $\xi > 0$ , if  $\kappa$  is small enough, we have*

$$\limsup_{\sigma \rightarrow 0} \mathbb{P} \left( \sup_{T_\kappa(\sigma) \leq t \leq \exp[\frac{2H+10}{\sigma^2}]} |X_t - Y_t| \geq \xi \right) \leq \sqrt{\kappa}. \quad (3.7)$$

*Proof.* For any  $t \geq T_\kappa(\sigma)$ , we have

$$\begin{aligned} & d |X_t - Y_t|^2 \\ &= -2 (X_t - Y_t; (\nabla V(X_t) + \nabla W * \mu_t(X_t)) - (\nabla V(Y_t) + \nabla W(Y_t - m))) dt, \end{aligned}$$

with the empirical measure  $\mu_t := \frac{1}{t} \int_0^t \delta_{X_s} ds$ . Let us define  $W_m(x) := V(x) + W(x - m)$  and  $W_{\mu_t}(x) := V(x) + W * \mu_t(x)$ . We thus have

$$\begin{aligned} \frac{d}{dt} |X_t - Y_t|^2 &= -2 (X_t - Y_t; \nabla W_{\mu_t}(X_t) - \nabla W_{\mu_t}(Y_t)) \\ &\quad + 2 (X_t - Y_t; \nabla W(Y_t - m) - \nabla W * \mu_t(Y_t)). \end{aligned}$$

However,  $\nabla^2 W_{\mu_t} = \nabla^2 V + \nabla^2 W * \mu_t \geq \rho + \alpha > 0$ . So, putting  $\gamma(t) := |X_t - Y_t|^2$ , Cauchy-Schwarz inequality yields to

$$\gamma'(t) \leq -2(\alpha + \rho) \gamma(t) + 2\sqrt{\gamma(t)} |\nabla W(Y_t - m) - \nabla W * \mu_t(Y_t)|.$$

But by the growth assumption (2.1) on  $W$ , we have for any probability measures  $\mu, \nu$  the following control

$$|\nabla W * \mu(x) - \nabla W * \nu(x)| \leq C (1 + |x|^{2k}) \mathbb{W}_{2k}^{2k}(\mu; \nu)$$

where  $2k$  is the degree of the polynomial  $P$ . We introduce the set

$$\mathcal{A}_\kappa := \left\{ \omega \in \Omega : \mathbb{W}_{2k}^{2k}(\mu_t; \delta_m) \leq \kappa^k \right\}.$$

By Markov inequality, we have  $\mathbb{P}(\mathcal{A}_\kappa) \geq 1 - \kappa^k$ . This implies for any  $t \geq T_\kappa(\sigma)$  and for any  $\omega \in \mathcal{A}_\kappa$ :

$$\gamma'(t) \leq -2(\alpha + \rho)\gamma(t) + 2C\kappa^k \sqrt{\gamma(t)} \left(1 + |Y_t|^{2k}\right).$$

However,  $\gamma(t) = 0$  for any  $t \leq T_\kappa(\sigma)$ . This means that

$$\left\{ t \geq 0 : \gamma(t) > \frac{C^2 \kappa^{2k} \left(1 + |Y_t|^{2k}\right)^2}{(\alpha + \rho)^2} \right\} \subset \{t \geq 0 : \gamma'(t) < 0\}.$$

We deduce that

$$\sup_{T_\kappa(\sigma) \leq t \leq \exp\left[\frac{2H+10}{\sigma^2}\right]} \gamma(t) \leq \frac{C^2 \kappa^{2k} \left(1 + \sup_{T_\kappa(\sigma) \leq t \leq \exp\left[\frac{2H+10}{\sigma^2}\right]} |Y_t|^{2k}\right)^2}{(\alpha + \rho)^2},$$

if  $\omega \in \mathcal{A}_\kappa$ . We now consider  $R > 0$  such that the exit cost of the diffusion  $Y$  from the ball of center  $m$  and radius  $R$  is at least  $H + 6$ , meaning that  $\inf\{V(x) + W(x - m) - V(m) : x \in B(m, R)\} \geq H + 6$ .

Then, by Freidlin-Wentzell theory, we deduce that

$\lim_{\sigma \rightarrow 0} \mathbb{P}\left(\sup_{T_\kappa(\sigma) \leq t \leq \exp\left[\frac{2H+10}{\sigma^2}\right]} |Y_t - m| \geq R\right) = 0$ . However, we have

$$\begin{aligned} & \mathbb{P}\left(\sup_{T_\kappa(\sigma) \leq t \leq \exp\left[\frac{2H+10}{\sigma^2}\right]} |X_t - Y_t| \geq \xi\right) \\ & \leq \mathbb{P}\left(\sup_{T_\kappa(\sigma) \leq t \leq \exp\left[\frac{2H+10}{\sigma^2}\right]} |Y_t - m| \geq R\right) \\ & + \mathbb{P}\left(\sup_{T_\kappa(\sigma) \leq t \leq \exp\left[\frac{2H+10}{\sigma^2}\right]} \gamma(t) \geq \xi^2, \sup_{T_\kappa(\sigma) \leq t \leq \exp\left[\frac{2H+10}{\sigma^2}\right]} |Y_t - m| < R\right) \\ & + \mathbb{P}(\mathcal{A}_\kappa^c). \end{aligned}$$

The first term tends to 0 as  $\sigma$  goes to 0. The second term is equal to 0 provided that  $\xi > \frac{C\kappa^k(1+2^{2k-1}(R+|m|^{2k}))}{\alpha+\rho}$ . In other words, if  $\kappa$  is small enough, the second term is equal to 0 uniformly with respect to  $\sigma$ . The third term is less than  $\sqrt{\kappa}$ . This concludes the proof.  $\square$

### 3.4 Proof of Theorem 1.2

Now we will prove the main Theorem 1.2. The idea of the proof is to use the fact that diffusions  $Y$  and  $X$  are close to each other after the deterministic stabilization time  $T_\kappa(\sigma)$  and until some fixed deterministic time  $\exp\left[\frac{2(H+5)}{\sigma^2}\right]$ , which we chose to be sufficiently big for our line of reasoning. We can control

the proximity of these two diffusions by parameter  $\kappa$ , which represents how close time-limit of our occupation measure and  $\delta_m$  are. It was already shown in Proposition 3.2, that with  $\sigma \rightarrow 0$  probability of exiting before time  $T_\kappa(\sigma)$  tends to zero, which means that we can only focus on our dynamics after the stabilization of occupation measure happens. After that, for the upper bound we show, that the event  $\tau > \exp\{\frac{2(H+\delta)}{\sigma^2}\}$  is not very likely due to the fact that, for small  $\sigma$ , diffusion  $Y$  can even leave some bigger domain in a smaller time, which contradicts the closeness of  $X$  and  $Y$ . Same type of reasoning takes place for the lower bound. Let us now provide the rigorous proof.

Fix some  $\delta, \kappa > 0$ , decrease it if necessary to be  $\delta < 5$ . For the upper bound, consider the following inequality:

$$\mathbb{P}(\tau > e^{\frac{2(H+\delta)}{\sigma^2}}) \leq \mathbb{P}(\tau > e^{\frac{2(H+\delta)}{\sigma^2}}, \tau_{\mathcal{D}^e}^Y \leq e^{\frac{2(H+\delta)}{\sigma^2}}) + \mathbb{P}(\tau_{\mathcal{D}^e}^Y > e^{\frac{2(H+\delta)}{\sigma^2}}), \quad (3.8)$$

where  $\mathcal{D}^e$  is some enlargement of domain  $\mathcal{D}$  such that its exit cost is equal to  $H + \frac{\delta}{2}$ , i.e.:

$$\mathcal{D}^e := \{x \in \mathbb{R}^d : V(x) + W(x - m) - V(m) < H + \frac{\delta}{2}\};$$

and  $\tau_{\mathcal{D}^e}^Y$  is exit time of diffusion  $Y$  from this domain, i.e.:

$$\tau_{\mathcal{D}^e}^Y := \inf\{t : Y_t \notin \mathcal{D}^e\}.$$

Note, that domain  $\mathcal{D}^e$  (since both  $V$  and  $W$  are continuous and convex) satisfies the usual assumptions (see [DZ98]) and  $d_e := d(\mathcal{D}, \mathcal{D}^e) > 0$ . By classical result of Freidlin-Wentzell theory,

$$\mathbb{P}(\tau_{\mathcal{D}^e}^Y > e^{\frac{2((H+\delta/2)+\delta/2)}{\sigma^2}}) \xrightarrow{\sigma \rightarrow 0} 0.$$

Let us decrease  $\sigma_\kappa$  if necessary, such that the quantity above will be less then  $\sqrt{\kappa}$  for any  $\sigma < \sigma_\kappa$ . Moreover, the first probability in (3.8) can be bounded by:

$$\mathbb{P}(\tau_{\mathcal{D}^e}^Y \leq e^{\frac{2(H+\delta)}{\sigma^2}} < \tau) \leq \mathbb{P}(|X_{\tau_{\mathcal{D}^e}^Y} - Y_{\tau_{\mathcal{D}^e}^Y}| \geq d_e) \leq 2\kappa^k,$$

where we use Proposition 3.3 and decrease  $\kappa$  and  $\sigma_\kappa$  if necessary.

We approach the lower bound similarly and introduce the contraction of the domain  $\mathcal{D}$ :

$$\mathcal{D}^c := \inf\{x \in \mathbb{R}^d : V(x) + W(x - m) - V(m) < H - \frac{\delta}{2}\}.$$

If  $\mathcal{D}^c$  turns out to be empty, decrease  $\delta$ . As previously, the domain  $\mathcal{D}^c$  satisfies usual properties and has positive distance with the initial domain, that is  $d_c := d(\mathcal{D}^c, \mathcal{D}) > 0$ . We introduce exit-time from the contracted domain for diffusion  $Y$ :

$$\tau_{\mathcal{D}^c}^Y := \inf\{t : Y_t \notin \mathcal{D}^c\},$$

and have the following estimate:

$$\begin{aligned}
\mathbb{P}(\tau < e^{\frac{2(H-\delta)}{\sigma^2}}) &\leq \mathbb{P}(T_\kappa(\sigma) < \tau < e^{\frac{2(H-\delta)}{\sigma^2}} \leq \tau_{\mathcal{D}^c}^Y) \\
&\quad + \mathbb{P}(\tau \leq T_\kappa(\sigma)) + \mathbb{P}(\tau_{\mathcal{D}^c}^Y \geq e^{\frac{2((H-\delta/2)-\delta/2)}{\sigma^2}}) \\
&\leq \mathbb{P}(|X_\tau - Y_\tau| \geq d_c) + \mathbb{P}(\tau \leq T_\kappa(\sigma)) + 2\kappa^k \\
&\leq 3\kappa^k + \mathbb{P}(\tau \leq T_\kappa(\sigma)) \leq 4\kappa^k,
\end{aligned}$$

by Proposition 3.2 (3.5), with  $\kappa$  and  $\sigma_\kappa$  small enough. That leads to:

$$\mathbb{P}(e^{\frac{2(H-\delta)}{\sigma^2}} \leq \tau \leq e^{\frac{2(H+\delta)}{\sigma^2}}) \geq 1 - 7\kappa^k,$$

which proves the theorem if we consider  $\kappa \rightarrow 0$ , parameter that uniformly controls the convergence of  $\sigma$  towards 0.

### 3.5 Proof of Corollary 1.3

We can apply Theorem 1.2 to the level sets of the potential  $V + W * \delta_m$ .

By definition of  $\mathcal{N}$  in Corollary 1.3, there exists a constant  $\xi > 0$  such that  $\inf_{z \in \mathcal{N}} (V(z) + W(z - m) - V(m)) = H + 3\xi$ . We introduce the set

$$\mathcal{K}_{H+2\xi} := \{x \in \mathbb{R}^d : V(x) + W(x - m) - V(m) < H + 2\xi\}.$$

If we denote by  $\tau_\xi$  the first exit time of  $X$  from  $\mathcal{K}_{H+2\xi}$ , then we obtain

$$\lim_{\sigma \rightarrow 0} \mathbb{P} \left\{ \exp \left[ \frac{2}{\sigma^2} (H + 2\xi - \rho) \right] < \tau_\xi < \exp \left[ \frac{2}{\sigma^2} (H + 2\xi + \rho) \right] \right\} = 1 \quad (3.9)$$

for any  $\rho > 0$ . By construction of  $\mathcal{K}_{H+2\xi}$ ,  $\mathcal{N} \subset \mathcal{K}_{H+2\xi}^c$ , which implies

$$\begin{aligned}
\mathbb{P} \{X_\tau \in \mathcal{N}\} &\leq \mathbb{P} \{X_\tau \notin \mathcal{K}_{H+2\xi}\} \\
&\leq \mathbb{P} \{\tau_\xi \leq \tau\} \\
&\leq \mathbb{P} \left\{ \tau_\xi \leq \exp \left[ \frac{2(H + 3\xi)}{\sigma^2} \right] \right\} + \mathbb{P} \left\{ \exp \left[ \frac{2H + \xi}{\sigma^2} \right] \leq \tau \right\}.
\end{aligned}$$

Applying (3.9) with  $\rho := \xi$  to the first term and Theorem 1.2 to the second one, we obtain the result.

## References

- [BLR02] Benaïm M., Ledoux M. and Raimond O. (2002), Self-interacting diffusions, *Prob. Th. Rel. Fields* **122**, 1-41.
- [BR05] Benaïm M. and Raimond O. (2005), Self-interacting diffusions III: symmetric interactions, *Ann. Prob.* **33**(5), 1716-1759.
- [CMV03] Carrillo J.A., McCann R.J. and Villani C. (2003), Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates, *Rev. math. Iberoam.* **19**(3), 971-1018.

- [CGM08] Cattiaux P., Guillin A., and Malrieu F. (2008), *Probab. Theory Relat. Fields* **140**: 19.
- [CK11] Chambeu S. and Kurtzmann A. (2011), Some particular self-interacting diffusions: ergodic behavior and almost sure convergence, *Bernoulli* **17**(4), 1248–1267.
- [DZ98] Dembo A. and Zeitouni O., *Large deviations techniques and applications*, volume 38 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2010. Corrected reprint of the second (1998) edition.
- [DR92] Durrett R.T. and Rogers L.C.G. (1992), Asymptotic behaviour of Brownian polymers, *Prob. Th. Rel. Fields* **92**(3), 337–349.
- [FW98] Freidlin M. I. and Wentzell A. D. (1998), *Random perturbations of dynamical systems*, volume 260 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York, second edition. Translated from the 1979 Russian original by Joseph Szücs.
- [HIP08] Herrmann S., Imkeller P., and Peithmann D. (2008), Large deviations and a Kramers’ type law for self-stabilizing diffusions, *Ann. Appl. Probab.*, **18**(4):1379–1423, 2008.
- [HT10] Herrmann S. and Tugaut J. (2010), Stationary measures for self-stabilizing processes: asymptotic analysis in the small noise limit, *Electron. J. Probab.* **15**, Paper No. 69, 2087–2116.
- [JS17] Jazaerli S., Saporito Y.F. (2017), Functional Itô calculus, path-dependence and the computation of Greeks. *Stochastic Processes and their Applications*, vol. 127, no.12, 3997–4028.
- [KK12] Kleptsyn V. and Kurtzmann A. (2012), Ergodicity of self-attracting motion, *Elect. J. Probab.*, Vol.17, paper no. 50.
- [Kur10] Kurtzmann A. (2010), The ODE method for some self-interacting diffusions on  $\mathbb{R}^d$ , *Ann. Inst. H. Poincaré, Probab. Statist.* **46**(3), 618–643.
- [NRW87] Norris J.R., Rogers L.C.G. and Williams D. (1987), Self-avoiding random walk: a Brownian motion model with local time drift, *Prob. Th. Rel. Fields* **74**(2), 271–287.
- [Rai09] Raimond O. (2009), Self-interacting diffusions: a simulated annealing version, *Prob. Th. Rel. Fields* **144**, 247–279.
- [Tug12] Tugaut J. (2012), Exit problem of McKean-Vlasov diffusions in convex landscapes. *Electron. J. Probab.*, Vol. 17, no. 76, 1–26.
- [Tug16] Tugaut J. (2016), A simple proof of a Kramers’ type law for self-stabilizing diffusions. *Electron. Commun. Probab.*, Vol. 21, no. 11, 1–7.
- [Tug18] Tugaut J. (2018), Exit problem of McKean-Vlasov diffusion in double-wells landscape. *J. Theoret. Probab.* **31**, no. 2, 1013–1023.
- [Tug19] Tugaut J. (2019), A simple proof of a Kramers’ type law for self-stabilizing diffusions in double-wells landscape. *ALEA, Lat. Am. J. Probab. Math. Stat.* **16**, 389–398.