



**HAL**  
open science

## Self-interacting diffusions: long-time behaviour and exit-problem in the convex case

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

► **To cite this version:**

Ashot Aleksian, Pierre del Moral, Aline Kurtzmann, Julian Tugaut. Self-interacting diffusions: long-time behaviour and exit-problem in the convex case. 2023. hal-01901145v3

**HAL Id: hal-01901145**

**<https://hal.science/hal-01901145v3>**

Preprint submitted on 25 Mar 2023

**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.

# Self-interacting diffusions: long-time behaviour and exit-problem in the convex case

A. Aleksian<sup>1,a</sup>, P. Del Moral<sup>2,b</sup>, A. Kurtzmann<sup>3,c</sup>  
and J. Tugaut<sup>1,d</sup>

<sup>1</sup>Université Jean Monnet, CNRS UMR 5208,  
Institut Camille Jordan, Maison de l'Université, 10 rue Tréfilerie,  
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 UMR 7502,  
Vandoeuvre-lès-Nancy, F-54506, France.

<sup>a</sup> ashot.aleksian@univ-st-etienne.fr, <sup>b</sup> pierre.del-moral@inria.fr  
<sup>c</sup> aline.kurtzmann@univ-lorraine.fr, <sup>d</sup> julian.tugaut@univ-st-etienne.fr

March 25, 2023

## Abstract

We study a class of time-inhomogeneous diffusion: the self-interacting one. We show a convergence result with a rate of convergence that does not depend on the diffusion coefficient. Finally, we establish a so-called Kramers' type law for the first exit-time of the process from domain of attractions when the landscapes are uniformly convex.

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

*Mathematics Subject Classification* : 60K35, 60H10, 60J60

## 1 Introduction

In this work, we are interested in a time-inhomogeneous diffusion. More precisely, we study the following specific diffusion, driven by the Stochastic Differential Equation (SDE):

$$dX_t = \sigma dB_t - \left( \nabla V(X_t) + \frac{1}{t} \int_0^t \nabla W(X_t - X_s) ds \right) dt, \quad X_0 = x \in \mathbb{R}^d \quad (1.1)$$

where  $V, W$  are two potentials on  $\mathbb{R}^d$  and  $\sigma > 0$ . The precise assumptions on the potentials will be given later in Subsection 2.1. We can already notice that the current position of the process  $X_t$  depends on the whole past trajectory of the

process  $(X_s)_{0 \leq s \leq t}$  through the interaction potential  $W$  appearing in the drift term. We call this kind of process a path-interaction process.

## 1.1 History

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 SDE 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 the rate of convergence.

## 1.2 Close processes

Another problem related to this paper is the diffusion corresponding to McKean-Vlasov partial differential equation. This corresponds to the measure-dependent drift diffusion governed by the SDE

$$dX_t = dB_t - \nabla W * \nu_t(X_t) dt \tag{1.2}$$

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, as  $t$  goes to infinity, to the (unique if  $W$  is strictly convex) probability measure solution to the equation

$$\nu := \frac{1}{Z} \exp(-2W * \nu),$$

where  $Z := Z(\nu)$  is the normalization constant. In the latter paper, the authors use a particles 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 strictly convex. 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 first exit-times from domains of attraction for the following motion

$$dX_t^\sigma = \sigma dB_t - \nabla V(X_t^\sigma) dt - \nabla W * \nu_t^\sigma(X_t^\sigma) dt, \tag{1.3}$$

where  $V$  is a potential,  $\nu_t^\sigma := \mathcal{L}(X_t^\sigma)$  and  $\sigma > 0$ . This was addressed by Herrmann, Imkeller and Peithmann [HIP08], who exhibited a Kramers' type law for the diffusion exit from the potential's domains of attraction by a large deviations principle for the self-stabilizing diffusion (which is a peculiar instance of 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, with a

different method. More precisely, in [Tug12], Tugaut analyses 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. Then, he generalized very recently his results in the case of double-wells landscape in [Tug18]. In [Tug16, Tug19], he does not use large deviations principle but a coupling method between the time-homogeneous diffusion

$$dY_t^\sigma = \sigma dB_t - \nabla V(Y_t^\sigma)dt - \nabla W(Y_t^\sigma - m)dt$$

(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  $Y^\sigma$  can be used for the exit-time of the self-stabilizing diffusion (1.3).

### 1.3 The questions that we study here

A large family of path-dependent processes has been studied by Saporito, see for instance [JS17]. He proves, with his co-authors, existence and uniqueness of such processes. The difference with the process studied here is that we normalize the occupation measure. In the current paper, we also prove the existence and uniqueness result for general potentials  $V$  and  $W$ , which are not necessarily convex.

A second result that we are obtaining is related to the asymptotic behaviour. Indeed, after proving the existence and the uniqueness of the solution to Diffusion (1.1), we are studying the convergence in long-time of the probability measure  $\mu_t$ . The idea is similar to the one in [KK12].

The present paper also deals with the exit-time problem of (1.1). We prove that the first exit-time  $\tau$  of the diffusion from some domain satisfies a Kramers' type law in the following sense:

$$\lim_{\sigma \rightarrow 0} \frac{\sigma^2}{2} \log(\tau) = H > 0,$$

where the convergence holds in probability.

We could adapt the techniques introduced by Herrmann, Imkeller and Peithmann but only in the case of a convex potential  $V$ . 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 Diffusion (1.1). As will be shown later, the exit-location can be easily obtained once we know the asymptotics of the first exit-time.

### 1.4 Outline

Our paper is divided into three parts. First, Section 2 is devoted to the explanation of the precise assumptions and the statement of the main results. After that, in Section 3, we prove Theorem 2.12 and Corollary 2.13, that is we establish the Kramers' type law and the exit-location result. To do so, we first provide some intermediate results. Finally, some possible extensions are discussed in the

Appendix. Before finishing the introduction, we give the notations used in the paper.

## 1.5 Notations

The parameters  $\kappa$ ,  $\xi$ ,  $\epsilon$  and  $\delta$  are arbitrarily small. The constants are denoted by  $C$  as usual and generic.

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.

In the sequel,  $(\cdot, \cdot)$  stands for the Euclidean scalar product and  $|\cdot|$  is the associated norm.

We first introduce the notion of positively invariant domain.

**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  $x \in \mathcal{G}$ .

Let us recall the definition of the Wasserstein distance.

**Definition 1.2.** For  $\mu_1, \mu_2 \in \mathcal{P}(\mathbb{R}^d)$ , the quadratic Wasserstein distance is defined as

$$\mathbb{W}_2(\mu_1, \mu_2) := (\inf\{\mathbb{E}(|Z_1 - Z_2|^2)\})^{1/2},$$

where the infimum is taken over all the random variables such that  $\mathcal{L}(Z_1) = \mu_1$  and  $\mathcal{L}(Z_2) = \mu_2$ . 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

$$\mathbb{W}_{2k}(\mu_1, \mu_2) := (\inf\{\mathbb{E}(|Z_1 - Z_2|^{2k})\})^{1/(2k)}.$$

In the following, for readability issue, we will omit the  $\sigma$ -exponent for the process  $X$  as well as the occupation measure  $\mu_t$ . Nevertheless, the reader has to keep in mind that the process  $(X_t, t \geq 0)$  and  $\mu_t$  do depend on  $\sigma$ .

**Definition 1.3.** The minimizer of  $V$  is denoted as  $m$ .

We also introduce the following mapping on the probability measures:

$$\Pi_\sigma(\mu)(dx) := \frac{e^{-\frac{2}{\sigma^2}(V(x)+W*\mu(x))}}{\int_{\mathbb{R}^n} e^{-\frac{2}{\sigma^2}(V(y)+W*\mu(y))} dy} dx.$$

## 2 Assumptions and main results

### 2.1 Assumptions

In this section we introduce the assumptions considered in the paper.

**Assumption 2.1.** We assume some regularity for the potentials  $V$  and  $W$ :  $V \in \mathcal{C}^2(\mathbb{R}^d)$ ,  $W \in \mathcal{C}^2(\mathbb{R}^d)$ . Also, without loss of generality, we consider only potentials such that  $V \geq 0$ ,  $W \geq 0$ .

Assumption 2.1 is usual in SDE. Since we use Itô calculus techniques, the differentiability assumption above is necessary.

**Assumption 2.2.**  $V$  and  $W$  (and their first two derivatives) have at most a polynomial growth. In other words, there exists a polynomial function  $P$  of degree  $2k$  such that  $P(|x|) \geq 1$  for any  $x \in \mathbb{R}^n$ , and

$$|V(x)| + |W(x)| \leq P(|x|), \quad (2.1)$$

$$|\nabla V(x)| + |\nabla W(x)| \leq P(|x|), \quad (2.2)$$

$$\text{and } \|\nabla^2 V(x)\| + \|\nabla^2 W(x)\| \leq P(|x|). \quad (2.3)$$

This assumption is used in the paper [KK12] to establish the rate of convergence towards the invariant probability measure. We come back to this question in Section 2.3.

**Remark 2.3.** Note, that without any loss of generality, we can choose polynomial  $P$  to be such that  $P(|x|) = C(1 + |x|^k)$ . Then, the following property holds: there exists a constant  $\gamma > 0$  such that  $P(|x + y|) \leq \gamma(P(|x|) + P(|y|))$ .

We also need the following assumption to establish the exit-time result:

**Assumption 2.4.** There exist  $\rho, \alpha > 0$  such that for any  $x \in \mathbb{R}^n$ , we have  $\nabla^2 V(x) \geq \rho \text{Id}$  and  $\nabla^2 W(x) \geq \alpha \text{Id}$ . The unique minimizer of  $V$  is denoted as  $m$  and the unique minimizer of  $W$  is 0.

To ensure the existence of the process, we will also use the following assumption, that will help us to exhibit a Lyapunov function:

**Assumption 2.5.**  $\lim_{|x| \rightarrow \infty} V(x) = +\infty$ ,  $\lim_{|x| \rightarrow \infty} \frac{|\nabla V(x)|^2}{V(x)} = +\infty$  and there exists  $a > 0$  such that  $\Delta V(x) \leq aV(x)$ .

**Remark 2.6.** By the latter growth condition,  $|\nabla V|^2 - \Delta V$  is bounded by below.

Finally, we will eventually assume the following, meaning that  $W$  is rotationally invariant:

**Assumption 2.7.** There exists a function  $G$  from  $\mathbb{R}_+$  to  $\mathbb{R}$  such that  $W(x) = G(|x|)$ .

The assumptions on the domain from which the diffusion  $X$  exits are the following.

**Assumption 2.8.** The domain  $\mathcal{D}$  is open and satisfies the following hypotheses:

i) Let  $\varphi$  be the solution to the following equation

$$\varphi_t = x_0 - \int_0^t \nabla V(\varphi_s) ds - \int_0^t \frac{1}{s} \int_0^s \nabla W(\varphi_s - \varphi_r) dr ds.$$

Then, for any  $t \in \mathbb{R}_+$ , we have  $\varphi_t \in \mathcal{D}$  and  $\lim_{t \rightarrow \infty} \varphi_t = m \in \mathcal{D}$ .

ii) The domain  $\mathcal{D}$  is positively invariant for the flow generated by the vector field  $x \mapsto -\nabla V(x) - \nabla W * \delta_m(x)$ .

iii) For any  $x \in \partial \mathcal{D}$ , define the flow  $\rho(x)$  as the solution to the equation

$$\rho_t(x) = x - \int_0^t \nabla V(\rho_s(x)) ds - \int_0^t \nabla W(\rho_s(x) - m) ds.$$

Moreover, assume that the following limit holds:

$$\lim_{t \rightarrow \infty} \rho_t(x) = m.$$

The assumption i) guarantees that, starting from fixed point  $x_0$ , the deterministic process defined by (1.1) with  $\sigma = 0$  converges towards the point of attraction  $m$ . Of course, we expect  $X_t$  to follow this path with high probability for small enough  $\sigma > 0$ . In fact, as will be shown later, its empirical measure  $\mu_t$  will also converge towards  $\delta_m$  and, after some deterministic time, with high probability, will stay inside a defined-in-advance neighbourhood of  $\delta_m$ . We call this effect: stabilisation of the empirical measure.

That leads to assumptions ii) and iii) above. After the “stabilisation time” we expect our drift term  $V + W * \mu_t$  to have a similar effect as  $V + W * \delta_m$ . Thus, the last two assumptions guarantee that the process  $X_t$  will forever tend to stay inside the domain  $\mathcal{D}$  and be attracted towards the point  $m$ . This is a necessary assumption when considering the exit from a stable domain of attraction solely under the influence of small noise.

## 2.2 Existence and uniqueness

The first results that we will provide are about the existence and the uniqueness of the solution to the SDE (1.1).

**Theorem 2.1.** *Under the Assumptions 2.1 and 2.5, for any  $x_0 \in \mathbb{R}^n$ , there exists a unique global strong solution  $(X_t, t \geq 0)$  to Equation (1.1).*

*Proof.* Local existence and uniqueness of the solution to (1.1) is standard under the locally Lipschitz assumptions on the vector fields (see for instance [RW00, Theorem 13.1]). 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  $(t, x) \mapsto \mathcal{E}_t(x)$  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

$$\begin{aligned} \mathcal{E}_{t \wedge \tau_n}(X_{t \wedge \tau_n}) &= V(x_0) + \int_0^{t \wedge \tau_n} (\nabla \mathcal{E}_s(X_s), dB_s) - \int_0^{t \wedge \tau_n} |\nabla \mathcal{E}_s(X_s)|^2 ds \\ &\quad + \frac{\sigma^2}{2} \int_0^{t \wedge \tau_n} \Delta \mathcal{E}_s(X_s) ds - \int_0^{t \wedge \tau_n} \int_0^s W(X_s - X_u) du \frac{ds}{s^2}. \end{aligned} \quad (2.4)$$

We note that  $\int_0^{t \wedge \tau_n} (\nabla \mathcal{E}_s(X_s), dB_s)$  is a true martingale. By removing the negative terms and by using Assumption 2.5, we get the following bound on the expectation of the Lyapunov functional:

$$\mathbb{E} \mathcal{E}_{t \wedge \tau_n}(X_{t \wedge \tau_n}) \leq V(x_0) + a \int_0^t \mathbb{E} \mathcal{E}_{s \wedge \tau_n}(X_{s \wedge \tau_n}) ds.$$

So, Gronwall's Lemma leads to:

$$\mathbb{E} V(X_{t \wedge \tau_n}) \leq \mathbb{E} \mathcal{E}_{t \wedge \tau_n}(X_{t \wedge \tau_n}) \leq V(x_0) e^{at}.$$

Since, by Assumption 2.5,  $\lim_{|x| \rightarrow \infty} V(x) = \infty$ , the process  $(X_t, t \geq 0)$  can not explode in a finite time, or else it leads to a contradiction with the inequality above. That proves existence of a unique global strong solution.  $\square$

### 2.3 Convergence rate

In this section we show the long time behaviour for the empirical measure of the self-interacting diffusion in the convex landscape. This framework was considered in the paper of Kleptsyn and Kurtzmann [KK12]. More precisely, they have proved the following

**Theorem 2.9.** [KK12, Theorem 1.6, Theorem 1.12, Proposition 2.5] *Let  $X$  be the solution to the equation (1.1) with  $\sigma = \sqrt{2}$ . Suppose that  $V$  and  $W$  satisfy Assumptions 2.1–2.5. Then:*

- i) *There exist  $\alpha, C > 0$  such that for any  $t \geq 0$  big enough:  $\mu_t \in K_{\alpha, C}$  a.s., where*

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

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

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

iii) There exists a constant  $a > 0$  such that almost surely, for  $t$  large enough one has

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

where  $2k$  is the degree of the polynomial  $P$  and  $\mathbb{W}_2$  is the quadratic Wasserstein distance.

Moreover, it was proved that, 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  $q$ . The authors showed 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  $e^{-(V+W*\rho_\infty)}$ .

We note that the results of [KK12] were established for the case  $\sigma = \sqrt{2}$ . Nevertheless, one can check that each step of the proof can be reformulated with  $\sigma$ . Moreover, all the asymptotic (with respect to time  $t$ ) results for small  $\sigma$  can be upper bounded without loss of generality by the case of a constant  $\sigma$ , that one can take to be equal to  $\sqrt{2}$ . That means that for smaller  $\sigma > 0$  we have faster convergence towards the invariant probability measure  $\rho_\infty$ .

We put together the observations above in the form of the following two results.

**Proposition 2.10.** *Under Assumptions 2.1–2.5, there exist  $\alpha, C > 0$ , such that for any  $t \geq 0$  and for any  $\sigma > 0$  small enough  $\mu_t \in K_{\alpha, C}$  a.s.*

**Proposition 2.11.** *Under Assumptions 2.1–2.5, there exists a constant  $a > 0$  such that for any  $x \in \mathbb{R}^n$ , almost surely, the following asymptotics holds:*

$$\mathbb{W}_{2k}(\mu_t, \rho_\infty) = O\left(\exp\{-a(\log t)^{\frac{1}{2k+1}}\}\right),$$

where  $\rho_\infty$  is the unique probability measure such that  $\rho_\infty = \Pi_\sigma(\rho_\infty)$ .

We stress that the convergence rate that we establish in Proposition 2.11 does not depend on  $\sigma$ .

## 2.4 Main result on exit-problem

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

**Theorem 2.12.** *We assume that the potentials  $V, W$  and an open domain  $\mathcal{D}$  satisfy the Assumptions 2.1–2.8. By  $\tau := \inf\{t \geq 0 : X_t \notin \mathcal{D}\}$ , we denote the first time the process  $X$  exits the domain  $\mathcal{D}$ . We introduce the so-called exit-cost:*

$$H := \inf_{x \in \partial\mathcal{D}} (V(x) + W(x - m) - V(m)). \quad (2.5)$$

Then  $\mathbb{P} - \lim_{\sigma \rightarrow 0} \frac{\sigma^2}{2} \log(\tau) = H$  that is for any  $\delta > 0$ , we have

$$\lim_{\sigma \rightarrow 0} \mathbb{P}\left(\exp\left\{\frac{2}{\sigma^2}(H - \delta)\right\} \leq \tau \leq \exp\left\{\frac{2}{\sigma^2}(H + \delta)\right\}\right) = 1. \quad (2.6)$$

This statement about the exit-time corresponds to what we denote as the Kramers' type law.

From Theorem 2.12, we immediately obtain the classical statement on the exit-location.

**Corollary 2.13.** Under the same assumptions as the ones of Theorem 2.12, if  $\mathcal{N}$  is a subset of  $\partial\mathcal{D}$  such that  $\inf_{z \in \mathcal{N}} (V(z) + W(z - m) - V(m)) > H$ , then

$$\lim_{\sigma \rightarrow 0} \mathbb{P}(X_\tau \in \mathcal{N}) = 0. \quad (2.7)$$

This means that the diffusion avoids to exit from a part of the boundary where the cost of exiting exceeds the exit-cost of  $\mathcal{D}$ .

### 3 Exit-problem

In this section, we prove our main result. First, we give the necessary intermediate results in Section 3.1. More precisely, we show that there exists a time of stabilisation around  $\delta_m$  for the occupation measure, in terms of Wasserstein distance. Then, we show that the process  $X$  solution to (1.1) is close to the solution of the deterministic flow  $(\varphi_t)_{t \geq 0}$ . Using that, we prove in Corollary 3.4 that the probability of leaving a positively invariant domain before the occupation measure remains stuck in the ball of center  $\delta_m$  and radius  $\kappa$  for  $\mathbb{W}_{2k}$  vanishes as  $\sigma$  goes to zero. Then, we consider the coupling between the studied diffusion and the one where the occupation measure is frozen to  $\delta_m$  and we show that these diffusions are close.

After, we provide the proof of Theorem 2.12 then we give the proofs of the intermediate results. Finally, we apply Theorem 2.12 to level sets so that we are in position to prove the exit-location result in Section 3.5.

#### 3.1 Intermediate results

We first introduce a deterministic time, representing the time of stabilisation of the occupation measure, if it occurs, around its supposed limit  $\delta_m$ :

**Definition 3.1.** For any  $\sigma > 0$  and for any  $\kappa > 0$ , we introduce:

$$T_\kappa(\sigma) := \inf \left\{ t_0 \geq 0 : \forall t \geq t_0, \mathbb{E}(\mathbb{W}_{2k}(\mu_t; \delta_m)) \leq \kappa \right\}, \quad (3.1)$$

where we remind that  $2k$  is the degree of the polynomial function  $P$ , introduced in Assumption 2.2.

**Proposition 3.2.** For any  $\sigma, \kappa > 0$ , the time  $T_\kappa(\sigma)$  is finite. Moreover, for any  $\kappa > 0$ , there exists  $T_\kappa > 0$  such that

$$\sup_{0 < \sigma < 1} T_\kappa(\sigma) \leq T_\kappa.$$

The proof of Proposition 3.2 is postponed to Section 3.4.1.

Next, we show that the probability for the process  $X$  to exit from  $\mathcal{D}$  before the time  $T_\kappa(\sigma)$  tends to 0 as  $\sigma$  goes to 0.

We remind the reader that in this work, the noise vanishes. Consequently, it is natural to introduce the deterministic flow  $(\varphi_t)_{t \geq 0}$  defined by the following zero-noise process

$$\dot{\varphi}_t = -\nabla V(\varphi_t) - \frac{1}{t} \int_0^t \nabla W(\varphi_t - \varphi_s) ds, \quad \varphi_0 = x_0. \quad (3.2)$$

We will state that for any  $T > 0$ ,  $(X_t, 0 \leq t \leq T)$  and  $(\varphi_t, 0 \leq t \leq T)$  are uniformly close while the noise goes to zero. Namely,

**Proposition 3.3.** *We assume that the potentials  $V$ ,  $W$  and an open domain  $\mathcal{D}$  satisfy the Assumptions 2.1–2.8. Then, for any  $\xi > 0$  and for any  $T > 0$ , we have:*

$$\lim_{\sigma \rightarrow 0} \mathbb{P} \left( \sup_{t \in [0; T]} |X_t - \varphi_t(x_0)|^2 > \xi \right) = 0. \quad (3.3)$$

The proof of Proposition 3.3 is postponed to Section 3.4.2. We deduce immediately the following.

**Corollary 3.4.** *We assume that the potentials  $V$ ,  $W$  and an open domain  $\mathcal{D}$  satisfy the Assumptions 2.1–2.8. Then:*

$$\lim_{\sigma \rightarrow 0} \mathbb{P}(\tau \leq T_\kappa(\sigma)) = 0. \quad (3.4)$$

*Proof.* It is sufficient to consider  $T := T_\kappa$  and  $\xi := \inf_{t \in [0; T_\kappa]} d(\mathcal{D}^c; \varphi_t) > 0$  in Proposition 3.3. We thus have:

$$\begin{aligned} \mathbb{P}(\tau \leq T_\kappa(\sigma)) &= \mathbb{P} \left( \inf \left\{ d(\mathcal{D}^c; X_t) : t \in [0; T_\kappa(\sigma)] \right\} = 0 \right) \\ &\leq \mathbb{P} \left( \sup_{t \in [0; T_\kappa(\sigma)]} |X_t - \varphi_t| > \xi \right), \end{aligned}$$

which converges towards 0 as noise vanishes.  $\square$

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 to  $X$  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 time) control of the law.

Here, we have proved that the nonlinear quantity appearing in (1.1) (that is  $\frac{1}{t} \int_0^t \delta_{X_s} ds$ ) remains stuck - with high probability - in a small ball (for the  $\mathbb{W}_{2k}$ -distance) of center  $\delta_m$  and radius  $\kappa$  for any  $t \geq T_\kappa(\sigma)$ . The idea is thus to

substitute  $\frac{1}{t} \int_0^t \delta_{X_s} ds$  by  $\delta_m$  and to compare the new diffusion with the initial one.

We introduce the diffusion  $(Y_t)_{t \geq 0}$  such that  $Y_t = X_t$  if  $t \leq T_\kappa(\sigma)$  and for any  $t \geq T_\kappa(\sigma)$

$$dY_t = \sigma dB_t - \nabla V(Y_t) dt - \nabla W(Y_t - m) dt. \quad (3.5)$$

**Proposition 3.5.** *We assume that the potentials  $V$ ,  $W$  and an open domain  $\mathcal{D}$  satisfy the Assumptions 2.1–2.8. Then, if  $\kappa$  is small enough, we have for any  $\xi > 0$*

$$\limsup_{\sigma \rightarrow 0} \mathbb{P} \left( \sup_{T_\kappa(\sigma) \leq t \leq \exp\left[\frac{2H+10}{\sigma^2}\right]} |X_t - Y_t| \geq \xi \right) \leq \kappa^k, \quad (3.6)$$

where we remind that  $2k$  is the degree of the polynomial function  $P$  introduced in Assumption 2.2.

The proof of Proposition 3.5 is postponed to Section 3.4.3.

### 3.2 Idea of the proof concerning the exit-time

The idea of the proof is to use the fact that diffusions  $Y$  and  $X$  are close to each other at least after the deterministic stabilisation time  $T_\kappa(\sigma)$  and until some fixed deterministic time  $\exp\left\{\frac{2(H+5)}{\sigma^2}\right\}$ . We choose this time 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 the empirical measure  $\mu_t$  and  $\delta_m$  are. It was already shown in Corollary 3.4, that with  $\sigma \rightarrow 0$  probability of exiting before time  $T_\kappa(\sigma)$  tends to zero. That means that we can focus on the dynamics after the stabilisation of the occupation measure happens. For the upper bound, we show that the event  $\left\{\tau > \exp\left[\frac{2(H+\delta)}{\sigma^2}\right]\right\}$  is unlikely due to the fact that, for small  $\sigma > 0$ , the diffusion  $Y$  leaves a bigger (than  $\mathcal{D}$ ) domain before the time  $\exp\left[\frac{2(H+\delta)}{\sigma^2}\right]$ , which, given the closeness of  $X$  and  $Y$ , gives  $\left\{\tau \leq \exp\left[\frac{2(H+\delta)}{\sigma^2}\right]\right\}$ . Same type of reasoning is used to prove the lower bound  $\exp\left\{\frac{2(H-\delta)}{\sigma^2}\right\}$ .

Let us now provide the rigorous proof in the next section. The proof of the intermediate lemmas are given in Subsection 3.4.

### 3.3 Proof for the exit-time result

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

$$\mathbb{P}(\tau > e^{\frac{2(H+\delta)}{\sigma^2}}) \leq \mathbb{P}(\tau > e^{\frac{2(H+\delta)}{\sigma^2}}, \tau_{\mathcal{D}^c}^Y \leq e^{\frac{2(H+\delta)}{\sigma^2}}) + \mathbb{P}(\tau_{\mathcal{D}^c}^Y > e^{\frac{2(H+\delta)}{\sigma^2}}), \quad (3.7)$$

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 the first 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 [DZ10]) and  $d_e := d(\mathcal{D}, \partial\mathcal{D}^e) > 0$ . By classical result of Freidlin-Wentzell theory,

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

Let us decrease  $\sigma_\kappa$  if necessary, such that the quantity above will be less than  $\sqrt{\kappa}$  for any  $\sigma < \sigma_\kappa$ . Moreover, the first probability in (3.7) 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 3.5 and decrease  $\kappa$  and  $\sigma_\kappa$  if necessary.

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

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

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 boundary of the initial domain, that is  $d_c := d(\mathcal{D}^c, \partial\mathcal{D}) > 0$ . We introduce the 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 Corollary 3.4, with  $\kappa$  and  $\sigma_\kappa$  small enough. This 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.

### 3.4 Proof of the intermediate results

Several propositions are proven here.

#### 3.4.1 Proof of Proposition 3.2

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

1. As was mentioned above, the invariant probability measure of self-interacting diffusion (and, at the same time, the weak-\* limit of its empirical measure a.s.) is the unique solution to the equation

$$\mu_\infty = \Pi_\sigma(\mu_\infty),$$

where  $\Pi_\sigma$  is defined as:

$$\Pi_\sigma(\mu)(x) = e^{-2(V+W*\mu)(x)/\sigma^2} / \int e^{-2(V+W*\mu)(z)/\sigma^2} dz.$$

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  $\mu_\infty$  are uniformly bounded with respect to  $\sigma$ , then  $\delta_m$  is the weak-\* limit of  $\mu_\infty$  with  $\sigma \rightarrow 0$  a.s. Note, that indeed, moments of  $\mu_t$  are uniformly bounded for any  $t > 0$ . Indeed, this is due to the fact that  $\mu_t \in K_{\alpha,C}$  for any  $t > 0$  and for some  $\alpha, C$  that do not depend on  $\sigma$  (Proposition 2.10). It proves that

$$\mu_\infty \xrightarrow[\sigma \rightarrow 0]{\text{weak-*}} \delta_m \text{ a.s.}$$

2. Let us consider the expectation  $\mathbb{E}(\mathbb{W}_{2k}(\mu_t; \delta_m))$ . First, let us show its existence. To do that, we use the fact that for any  $t > 0$ ,  $\mu_t \in K_{\alpha,C}$  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.

3. Now, we can finish the proof by separating the expectation of the distance between  $\mu_t$  and  $\delta_m$  into two parts and find the limit:

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

where the limit is not just iterated, but holds for the pair  $(t, \sigma)$ , since the rate of convergence of  $\mu_t$  towards  $\mu_\infty$  in time does not depend on  $\sigma$ , which was shown in Proposition 2.11. 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  $\sigma$ .

### 3.4.2 Proof of Proposition 3.3

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) :

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

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 compact at least for small  $\xi$ , which we can decrease without any loss of generality. We remind that  $\rho$  is the convexity constant of  $V$ . We thus have

$$\begin{aligned} |X_t - \psi_t|^2 &\leq d\sigma^2 t - 2\rho \int_0^t |X_s - \psi_s|^2 ds + 2\sigma \int_0^t (X_s - \psi_s, dB_s) \\ &\quad + \text{Lip}_{\nabla W}^{K'} \int_0^t \frac{2}{s} \int_0^s (|X_s - \psi_s|^2 + |X_s - \psi_s| \cdot |X_z - \psi_z|) dz \\ &\leq d\sigma^2 t - 2\rho \int_0^t |X_s - \psi_s|^2 ds + 2\sigma \int_0^t (X_s - \psi_s, dB_s) \quad (3.8) \\ &\quad + \text{Lip}_{\nabla W}^{K'} \int_0^t \frac{1}{s} \int_0^s (3|X_s - \psi_s|^2 + |X_z - \psi_z|^2) dz. \end{aligned}$$

Note then that by the Burkholder-Davis-Gundy inequality, we get for some constant  $C > 0$ :

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

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 (3.8) for any  $s \in [0, t \wedge \mathcal{T}]$ . We also remind that  $t \leq T$  and derive:

$$\begin{aligned} \mathbb{E} \left( \sup_{s \in [0; t \wedge \mathcal{T}]} |X_s - \psi_s|^2 \right) &\leq d\sigma^2 T + C\sigma^2 \sqrt{\int_0^t \mathbb{E} \left( \sup_{z \in [0, s \wedge \mathcal{T}]} (|X_z - \psi_z|^2) \right) dz} \\ &\quad + 4\text{Lip}_{\nabla W}^{K'} \int_0^t \mathbb{E} \left( \sup_{z \in [0, s \wedge \mathcal{T}]} |X_z - \psi_z|^2 \right) ds \\ &\leq d\sigma^2 T + \frac{C\sigma^2}{2} \left[ 1 + T \mathbb{E} \left( \sup_{s \in [0, t \wedge \mathcal{T}]} (|X_s - \psi_s|^2) \right) \right] \\ &\quad + 4\text{Lip}_{\nabla W}^{K'} \int_0^t \mathbb{E} \left( \sup_{z \in [0, s \wedge \mathcal{T}]} |X_z - \psi_z|^2 \right) ds, \end{aligned}$$

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

$$u_t \leq \frac{1}{1 - CT\sigma^2/2} \left( \frac{2Td + C}{2} \sigma^2 + 4\text{Lip}_{\nabla W}^{K'} \int_0^t u_s ds \right),$$

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

$$u_t \leq \frac{(2Td + C)\sigma^2}{2(1 - CT\sigma^2/2)} \exp \left\{ \frac{4\text{Lip}_{\nabla W}^{K'}}{1 - CT\sigma^2/2} T \right\} = O(\sigma^2). \quad (3.9)$$

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 (3.9), 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 Proposition 3.3, 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 (3.9), which completes the proof.

### 3.4.3 Proof of Proposition 3.5

Let us define

$$W_m(x) := V(x) + W(x - m) \quad \text{and} \quad W_{\mu_t}(x) := V(x) + W * \mu_t(x),$$

with the occupation measure  $\mu_t := \frac{1}{t} \int_0^t \delta_{X_s} ds$ . For any  $t \geq T_\kappa(\sigma)$ , we have

$$d|X_t - Y_t|^2 = -2(X_t - Y_t; \nabla W_{\mu_t}(X_t) - \nabla W_m(Y_t)) dt.$$

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

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

However, by the growth condition (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  $P$  introduced in Assumption 2.2. We introduce the set

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

By Markov inequality, we have  $\mathbb{P}(\mathcal{A}_\kappa^c) \leq 7\kappa^k$  then  $\mathbb{P}(\mathcal{A}_\kappa) \geq 1 - \kappa^k$ . This implies for any  $t \geq T_\kappa(\sigma)$  and for any  $\omega \in \mathcal{A}_\kappa$ :

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

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

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

By [BRTV98, Lemma 3.7], we deduce that

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

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

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

However, we have

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

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.

### 3.5 Proof for the exit-location result

We can apply Theorem 2.12 to the level sets of the potential  $W_m := V + W * \delta_m$ .

By definition of  $\mathcal{N}$  in Corollary 2.13, 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

$$\lim_{\sigma \rightarrow 0} \mathbb{P} \left\{ \exp \left[ \frac{2}{\sigma^2} (H + 2\xi - \eta) \right] < \tau_\xi < \exp \left[ \frac{2}{\sigma^2} (H + 2\xi + \eta) \right] \right\} = 1, \quad (3.10)$$

for any  $\eta > 0$ . 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 (3.10) with  $\eta := \xi$  to the first term and Theorem 2.12 to the second one, we obtain the result.

## A Discussions on extension

In this Section, we provide some ideas on how the results of the paper can be extended.

First, we can modify our equation by adding a reflection at the boundary as it was done for McKean-Vlasov case, for example in [AdRR+22]. Note, that if the boundary of reflection contains the closure of the domain from which we want to exit, the result for exit-time does not change and is immediate, since, unlike in McKean-Vlasov case, there is no interaction with the law of the process for self-interacting diffusion.

Second, in this paper we take a diffusion coefficient which is proportional to the identity matrix. However, it could be relevant for some problems related to optimization to consider a more general diffusion coefficient.

Third, in the current work we do not derive an Arrhenius law, that is to say the convergence of  $\frac{\sigma^2}{2} \log(\mathbb{E}[\tau])$  towards  $H$ . To obtain such a result, it requires to use the large deviations techniques instead of the coupling method that is used here.

Fourth, we point out that the potentials  $V$  and  $W$  are both assumed to be uniformly convex. The techniques used in this paper are not adapted for a more general case. One shall use the techniques close to [DZ10, FW98] to relax the convexity assumptions.

Finally, we could also study SDEs where the nonlinear part of the drift is more general. that is to say:

$$X_t = x_0 + \sigma B_t - \int_0^t \nabla V(X_s) ds - \int_0^t b(X_s, \mu_s) ds, \quad (\text{A.1})$$

where  $\mu_s := \frac{1}{t} \int_0^t \delta_{X_s} ds$  and  $x \mapsto b(x, \mu)$  is differentiable whereas  $\mu \mapsto b(x, \mu)$  is  $L$ -differentiable, see [RW19] and references therein. However, this will require some adaptations of our methods.

**Acknowledgments:** *This work is supported by the French ANR grant METANOLIN (ANR-19-CE40-0009).*

## References

- [AdRR<sup>+</sup>22] Daniel Adams, Gonalo dos Reis, Romain Ravaille, William Salkeld, and Julian Tugaut. Large deviations and exit-times for reflected McKean-Vlasov equations with self-stabilising terms and superlinear drifts. *Stochastic Process. Appl.*, 146:264–310, 2022.
- [BLR02] Michel Benaïm, Michel Ledoux, and Olivier Raimond. Self-interacting diffusions. *Probab. Theory Related Fields*, 122(1):1–41, 2002.
- [BR05] Michel Benaïm and Olivier Raimond. Self-interacting diffusions. III. Symmetric interactions. *Ann. Probab.*, 33(5):1717–1759, 2005.
- [BRTV98] S. Benachour, B. Roynette, D. Talay, and P. Vallois. Nonlinear self-stabilizing processes. I. Existence, invariant probability, propagation of chaos. *Stochastic Process. Appl.*, 75(2):173–201, 1998.
- [CGM08] P. Cattiaux, A. Guillin, and F. Malrieu. Probabilistic approach for granular media equations in the non-uniformly convex case. *Probab. Theory Related Fields*, 140(1-2):19–40, 2008.
- [CK11] Sebastien Chambeu and Aline Kurtzmann. Some particular self-interacting diffusions: ergodic behaviour and almost sure convergence. *Bernoulli*, 17(4):1248–1267, 2011.
- [CMV03] Jose A. Carrillo, Robert J. McCann, and Cedric Villani. Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates. *Rev. Mat. Iberoamericana*, 19(3):971–1018, 2003.
- [DR92] R. T. Durrett and L. C. G. Rogers. Asymptotic behavior of Brownian polymers. *Probab. Theory Related Fields*, 92(3):337–349, 1992.

- [DZ10] Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications*, volume 38 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2010. Corrected reprint of the second (1998) edition.
- [FW98] M. I. Freidlin and A. D. Wentzell. *Random perturbations of dynamical systems*, volume 260 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York, second edition, 1998. Translated from the 1979 Russian original by Joseph Szücs.
- [HIP08] Samuel Herrmann, Peter Imkeller, and Dierk Peithmann. Large deviations and a Kramers' type law for self-stabilizing diffusions. *Ann. Appl. Probab.*, 18(4):1379–1423, 2008.
- [HT10] S. Herrmann and J. Tugaut. Stationary measures for self-stabilizing processes: asymptotic analysis in the small noise limit. *Electron. J. Probab.*, 15:no. 69, 2087–2116, 2010.
- [JS17] Samy Jazaerli and Yuri F. Saporito. Functional Itô calculus, path-dependence and the computation of Greeks. *Stochastic Process. Appl.*, 127(12):3997–4028, 2017.
- [KK12] Victor Kleptsyn and Aline Kurtzmann. Ergodicity of self-attracting motion. *Electron. J. Probab.*, 17:no. 50, 37, 2012.
- [Kur10] Aline Kurtzmann. The ODE method for some self-interacting diffusions on  $\mathbb{R}^d$ . *Ann. Inst. Henri Poincaré Probab. Stat.*, 46(3):618–643, 2010.
- [NRW87] J. R. Norris, L. C. G. Rogers, and David Williams. Self-avoiding random walk: a Brownian motion model with local time drift. *Probab. Theory Related Fields*, 74(2):271–287, 1987.
- [Rai09] Olivier Raimond. Self-interacting diffusions: a simulated annealing version. *Probab. Theory Related Fields*, 144(1-2):247–279, 2009.
- [RW00] L. C. G. Rogers and David Williams. *Diffusions, Markov processes and martingales: Itô calculus*, volume 2. Cambridge university press, 2000.
- [RW19] Panpan Ren and Feng-Yu Wang. Bismut formula for Lions derivative of distribution dependent SDEs and applications. *J. Differential Equations*, 267(8):4745–4777, 2019.
- [Tug12] Julian Tugaut. Exit problem of McKean-Vlasov diffusions in convex landscapes. *Electron. J. Probab.*, 17:no. 76, 26, 2012.
- [Tug16] Julian Tugaut. A simple proof of a Kramers' type law for self-stabilizing diffusions. *Electron. Commun. Probab.*, 21:Paper No. 11, 7, 2016.

- [Tug18] Julian Tugaut. Exit-problem of McKean-Vlasov diffusions in double-well landscape. *J. Theoret. Probab.*, 31(2):1013–1023, 2018.
- [Tug19] Julian Tugaut. A simple proof of a Kramers’ type law for self-stabilizing diffusions in double-wells landscape. *ALEA Lat. Am. J. Probab. Math. Stat.*, 16(1):389–398, 2019.