



# A Kramers' type law for self-interacting diffusions

Pierre del Moral, Aline Kurtzmann, Julian Tugaut

## ► To cite this version:

Pierre del Moral, Aline Kurtzmann, Julian Tugaut. A Kramers' type law for self-interacting diffusions. 2018. hal-01901145v1

**HAL Id: hal-01901145**

**<https://hal.science/hal-01901145v1>**

Preprint submitted on 22 Oct 2018 (v1), 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.

# A KRAMERS' TYPE LAW FOR SELF-INTERACTING DIFFUSIONS

P. DEL MORAL, A. KURTZMANN, J. TUGAUT

**ABSTRACT.** We study the exit time of a domain for a self-interacting diffusion, where the Brownian motion is replaced by  $\sigma B_t$  for a constant  $\sigma$ . We first show that the rate of convergence previously obtained for a convex confinement potential  $V$  and a convex interaction potential does not depend on  $\sigma$ . Then, we show a Kramers' type law for the first exit-time from a domain (satisfying classical hypotheses).

*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. The model is the following. 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. More recently, 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 [Ra09] 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

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

when  $g$  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 [AK10], 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 strictly 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 later in §2.2.

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

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

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 later 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

$$(1.2) \quad dX_t^\varepsilon = \sqrt{\varepsilon} dB_t - \nabla V(X_t^\varepsilon) dt - \nabla W * \nu_t(X_t^\varepsilon) dt$$

where  $V$  is a potential,  $*$  stands for the convolution,  $\nu_t = \mathcal{L}(X_t^\varepsilon)$  and  $\varepsilon > 0$ . This was addressed by Herrmann, Inkeller and Peithmann [HIP08], who exhibited a Kramer's type law for the particle's exit from the potential's domains of attraction and a large deviations principle for the self-stabilizing diffusion. To get this, they reconstructed the Freidlin-Wentzell theory for the self-stabilizing diffusion. More precisely, they established a large deviation 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 the papers [Tug16, Tug17], Tugaut did not use large deviation principle but a coupling method between the time-homogeneous diffusion

$$dX_t = \sqrt{\varepsilon} dB_t - \nabla V(X_t) dt - \nabla W(X_t - x_*) dt$$

(where  $x_*$  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 and exit location problem of a specific diffusion, driven by the SDE

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

where  $V$  is a potential and  $\sigma \in \mathbb{R}, \alpha > 0$ . 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 in the case when the diffusion converges (and if we know the speed of convergence). And more challenging, our main aim consists in improving the simulated annealing method (even if we need a negative interaction coefficient  $\alpha < 0$  for that). This paper is a first step in this direction.

**1.1. Some useful notation.** 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:

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

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 (2.3)) 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

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

where  $\mathcal{C}(\mathbb{R}^d)$  is the set of continuous functions  $\mathbb{R}^d \rightarrow \mathbb{R}$ . We recall 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 Euclidian scalar product.

**Definition 1.1.** Let  $d$  be any positive integer. Let  $\mathcal{G}$  be a subset of  $\mathbb{R}^d$  and let  $U$  be a vector field from  $\mathbb{R}^d$  to  $\mathbb{R}^d$  which satisfies 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 stable 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 result.** We study the following diffusion

$$(1.6) \quad dX_t = \sigma dB_t - \left( \nabla V(X_t) + \alpha \left( X_t - \frac{1}{t} \int_0^t X_s ds \right) \right) dt, \quad X_0 = x \in \mathbb{R}^d$$

where  $B$  is a Brownian motion and  $\alpha > 0, \sigma > 0$ . We also suppose that the potential  $V$  is regular and has a unique local minimum  $m$ . The precise assumptions on  $V$  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 stable domain.

**Theorem 1.2.** *Consider a domain  $\mathcal{D}$  that is stable by the flow  $x \mapsto -\nabla V(x) - \alpha(x - m)$  and denote by  $\tau$  the first time the process  $X$  exits the domain  $\mathcal{D}$ . Let  $H := \inf_{x \in \partial \mathcal{D}} V(x) + \frac{\alpha}{2} \|x - m\|^2 - V(m)$  be the exit cost from  $\mathcal{D}$ . For any  $\delta > 0$ , we have*

$$(1.7) \quad \lim_{\sigma \rightarrow 0} \mathbb{P} \left( \exp \left\{ \frac{2}{\sigma^2} (H - \delta) \right\} \leq \tau \leq \exp \left\{ \frac{2}{\sigma^2} (H + \delta) \right\} \right) = 1.$$

**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 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 stable 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 in §3.4.

## 2. THE LINEAR SELF-INTERACTING DIFFUSION

We remind the reader that  $X$  is solution to

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

Consider the potential  $V : \mathbb{R}^d \rightarrow \mathbb{R}_+$ . Let  $Max = \{M_1, \dots, M_p\}$  be the (finite) set of saddle points and local maxima of  $V$  and denote by  $m$  the unique local minimum of  $V$ . So  $\mathcal{CP} := m \cup Max$  is the set of critical points of  $V$ . We assume that  $V$  satisfies:

- 1) (*regularity and positivity*)  $V \in C^2(\mathbb{R}^d)$  and  $V > 0$ ;
- 2) (*convexity*)  $V = \chi + W$  where  $\chi$  is a compactly supported function and there exists  $c > \alpha$  such that  $\nabla^2 W \geq cId$ ;
- 3) (*growth*) there exists  $a > 0$  such that for all  $x \in \mathbb{R}^d$ , we have

$$(2.1) \quad \Delta V(x) \leq aV(x) \text{ and } \lim_{|x| \rightarrow \infty} \frac{|\nabla V(x)|^2}{V(x)} = \infty;$$

- 4) (*critical points*)  $\forall \xi \in \mathbb{R}^d$ ,  $(\nabla^2 V(m)\xi, \xi) > \rho|\xi|^2$  with  $\rho > \alpha$  and for all  $M_i$ ,  $\nabla^2 V(M_i)$  admits a negative eigenvalue.

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

**Proposition 2.1.** *For any  $x \in \mathbb{R}^d$ , there exists a unique global strong solution  $(X_t, t \geq 0)$ .*

*Proof.* The local existence and uniqueness of the solution to (1.6) is standard. We only need to prove here that  $X$  does not explode in a finite time. To this aim, apply Itô's formula to the function  $x \mapsto V(x)$ :

$$\begin{aligned} dV(X_t) &= \sigma(\nabla V(X_t), dB_t) \\ &\quad - \left( |\nabla V(X_t)|^2 - \alpha(X_t, \nabla V(X_t)) + \frac{\alpha}{t} \int_0^t (\nabla V(X_t), X_s) ds - \frac{\sigma^2}{2} \Delta V(X_t) \right) dt, \end{aligned}$$

and introduce the sequence of stopping times  $\tau_0 = 0$  and

$$\tau_n = \inf\{t \geq 0; V(X_t) + \frac{\alpha}{t} \int_0^t (X_s, \nabla V(X_t)) ds > n\}.$$

The convexity condition and the growth condition (2.1) imply that

$$\begin{aligned} \mathbb{E}V(X_{t \wedge \tau_n}) &\leq V(x) + a \frac{\sigma^2}{2} \mathbb{E} \int_0^{t \wedge \tau_n} V(X_s) ds - \alpha \mathbb{E} \int_0^{t \wedge \tau_n} \frac{1}{s} \int_0^s (X_u, \nabla V(X_s)) du ds \\ &\leq V(x) + a \frac{\sigma^2}{2} \mathbb{E} \int_0^{t \wedge \tau_n} V(X_s) ds + d\alpha \mathbb{E} \int_0^{t \wedge \tau_n} \left( \frac{1}{s} \int_0^s |X_u|^2 du + |\nabla V(X_s)|^2 \right) ds \\ &\leq V(x) + a \frac{\sigma^2}{2} \mathbb{E} \int_0^{t \wedge \tau_n} V(X_s) ds + d\alpha \mathbb{E} \int_0^{t \wedge \tau_n} \left( \frac{1}{s} \int_0^s V(X_u) du + aV(X_s) \right) ds \\ &\leq V(x) + a \left( \frac{\sigma^2}{2} + d\alpha \right) \mathbb{E} \int_0^{t \wedge \tau_n} V(X_s) ds + d\alpha \mathbb{E} \int_0^{t \wedge \tau_n} V(X_s) \log(t/s) ds. \end{aligned}$$

We finally conclude that

$$\mathbb{E}V(X_{t \wedge \tau_n}) \leq V(x) + \beta t \log t e^{\beta t \log t}$$

with  $\beta = a \frac{\sigma^2}{2} + d\alpha(a+1)$ . □

**2.1. Behaviour near the critical points of  $V$ .** We first prove that  $X$  gets close to the critical points of  $V$ .

**Proposition 2.2.** *For  $\varepsilon > 0$ , let  $T_t^\varepsilon := \inf\{s \geq t : d(X_s, \mathcal{CP}) < \varepsilon\}$ . Then, for all  $\varepsilon > 0$  and all  $t \geq 0$ , we have  $\mathbb{P}(T_t^\varepsilon < +\infty) = 1$ .*

*Proof.* Let  $\varepsilon > 0$ . We have

$$\begin{aligned} d(V(X_t) + \frac{\alpha}{2} |X_t - \frac{1}{t} \int_0^t X_s ds|^2) &= \sigma \left( \nabla V(X_t) + \alpha(X_t - \frac{1}{t} \int_0^t X_s ds), \sigma dB_t \right) \\ &\quad - \left( |\nabla V(X_t)|^2 - (\alpha^2 - \frac{\alpha}{t}) |x_t - \frac{1}{t} \int_0^t X_s ds|^2 - \frac{\sigma^2}{2} \Delta V(X_t) - \alpha d \right) dt. \end{aligned}$$

There exists  $t_0$  such that for any  $t \geq t_0$ , we have that

$$\begin{aligned} &\left( V(X_{s \wedge T_t^\varepsilon}) + \frac{\alpha}{2} |X_{s \wedge T_t^\varepsilon} - \frac{1}{s \wedge T_t^\varepsilon} \int_0^{s \wedge T_t^\varepsilon} X_u du|^2 + \frac{\sigma^2}{2} \int_0^{s \wedge T_t^\varepsilon} \Delta V(X_u) du + \right. \\ &\quad \left. + \alpha^2 \int_0^{s \wedge T_t^\varepsilon} |X_u - \frac{1}{u} \int_0^u X_v dv|^2 du + \alpha ds \wedge T_t^\varepsilon \right)_{s \geq t} \end{aligned}$$

is a positive supermartingale. It thus converges a.s. as  $s \rightarrow +\infty$ . The same calculus also applies to the positive super-martingale

$$\left( V(X_{s \wedge T_t^\varepsilon}) + \frac{\alpha}{2} |X_{s \wedge T_t^\varepsilon} - \frac{1}{s \wedge T_t^\varepsilon} \int_0^{s \wedge T_t^\varepsilon} X_u du|^2 + \frac{\sigma^2}{2} \int_0^{s \wedge T_t^\varepsilon} \Delta V(X_u) du + \alpha^2 \int_0^{s \wedge T_t^\varepsilon} |X_u - \frac{1}{u} \int_0^u X_v dv|^2 du + \alpha ds \wedge T_t^\varepsilon + \frac{1}{2} \int_0^{s \wedge T_t^\varepsilon} |\nabla V(X_u)|^2 du \right)_{s \geq t}.$$

We thus conclude that  $\left( \int_0^{s \wedge T_t^\varepsilon} |\nabla V(X_u)|^2 du \right)_{s \geq t}$  converges a.s.

Suppose now that we are on the set  $\{T_t^\varepsilon = +\infty\}$ . It follows that  $|\nabla V(X_{s \wedge T_t^\varepsilon})|$  converges a.s. to 0. This implies that  $X_{s \wedge T_t^\varepsilon}$  gets close to  $\mathcal{CP}$ , leading to a contradiction. We conclude that  $\mathbb{P}(T_t^\varepsilon < +\infty) = 1$ .  $\square$

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

**Theorem 2.3.** [KK12, Theorem 1.6] *Let  $X$  be the solution to the equation*

$$(2.2) \quad dX_t = \sqrt{2} dB_t - \left( \nabla V(X_t) + \frac{1}{t} \int_0^t \nabla W(X_t - X_s) ds \right) dt.$$

*Suppose, that  $V \in C^2(\mathbb{R}^d)$  and  $W \in C^2(\mathbb{R}^d)$ , and:*

- i) *spherical symmetry:  $W(x) = W(|x|)$ ;*
- ii)  *$V$  and  $W$  are convex,  $\lim_{|x| \rightarrow +\infty} V(x) = +\infty$ , and either  $V$  or  $W$  is uniformly convex:*

$$\exists C > 0 : \quad \forall x \in \mathbb{R}^d, \forall v \in \mathbb{S}^{d-1}, \quad \left| \frac{\partial^2 V}{\partial v^2} \right|_x \geq C \quad \text{or} \quad \forall x \in \mathbb{R}^d, \forall v \in \mathbb{S}^{d-1}, \quad \left| \frac{\partial^2 W}{\partial v^2} \right|_x \geq C;$$

- iii)  *$V$  and  $W$  have at most a polynomial growth: for some polynomial  $P$ , we have  $\forall x \in \mathbb{R}^d$*

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

*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 remind the definition of the Wasserstein distance: for  $\mu_1, \mu_2 \in \mathcal{P}(\mathbb{R}^d; P)$ , the quadratic Wasserstein distance is defined as

$$W_2(\mu_1, \mu_2) := (\inf \{ \mathbb{E}(|\xi_1 - \xi_2|^2) \})^{1/2},$$

where the infimum is taken over all the random variables such that  $\{\text{law of } \xi_1\} = \mu_1$  and  $\{\text{law of } \xi_2\} = \mu_2$ .

**Theorem 2.4.** [KK12, Theorem 1.12] *There exists a constant  $a > 0$  such that almost surely, we have for  $t$  large enough  $W_2(\mu_t^c, \rho_\infty) = O(\exp\{-a^{k+1}\sqrt{\log t}\})$ , where  $k$  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  $W_2$  is the quadratic Wasserstein distance, that is the minimal  $L^2$ -distance taken over all the couplings between  $\mu_t^c$  and  $\rho_\infty$ .*

Of course, we study here the case  $W(x) = \alpha \frac{|x|^2}{2}$  and thus  $P$  corresponds to the polynomial 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  $\mu_t = \frac{1}{t} \int_0^t X_s ds$  of  $X_t$  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 (2.2) 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<sup>1</sup>, 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

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

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)

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

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

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

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

<sup>1</sup>see for instance [RY98], chap. X

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  $\Pi(\mu_{T_n})(dx) = \frac{1}{Z} e^{-(V+W*\mu_{T_n})(x)} dx$ ), 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

$$dZ_t = \sqrt{2}dB_t - (\nabla V + \nabla W * \mu_\infty)(Z_t)dt.$$

The process  $Z$  is obviously Markovian and the limit-quotient theorem applies (see [RY98]):

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

for the weak\* convergence of measures. So when the limit  $\mu_\infty$  exists, it satisfies  $\mu_\infty = \Pi(\mu_\infty)$ . 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 smoothened (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 remind that for the dynamics in presence of an exterior potential  $V$ , the free energy function is

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

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

$$(2.6) \quad \mathcal{H}(\mu) := \int_{\mathbb{R}^d} \mu(x) \log \mu(x) dx.$$

As the free energy is controlled by the quadratic Wasserstein distance  $W_2$ , 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:  $W_2(\mu_t^c, \rho_\infty)$  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  $W_2(\mu_t^c, \Pi(\mu_{T_n}^c))$ , the second one  $W_2(\Pi(\mu_{T_n}^c), \Phi_n^n(\mu_{T_n}^c))$  and the third one  $W_2(\Phi_n^n(\mu_{T_n}^c), \rho_\infty)$ .

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

Let us now prove for instance the exponential decrease for the centered measure  $\Pi(\mu)$  and show that, as  $\sigma^2 \ll 1$ , we can obtain a minoration of the speed of convergence that does not depend on  $\sigma$ . First, define the following spaces:

**Definition 2.5.** Let  $\alpha, C > 0$  be given. Define

$$(2.7) \quad K_{\alpha, C}^0 := \{\mu \in \mathcal{P}(\mathbb{R}^d); \quad \forall r > 0, \mu(\{y; |y| > r\}) < C e^{-\alpha r}\},$$

$$(2.8) \quad K_{\alpha, C} := \{\mu \in \mathcal{P}(\mathbb{R}^d); \quad \mu^c \in K_{\alpha, C}^0\}.$$

**Proposition 2.6.** [KK12, Proposition 2.9] *There exist  $C_W, C_\Pi > 0$  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  $c_\mu := \mathbb{E}(\mu)$ .*

*Proof.* 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

$$(2.9) \quad \Pi(\mu)(|x - c_\mu| \geq R) \leq C_\Pi e^{-CR} \cdot \Pi(\mu)(|x - c_\mu| \leq 2),$$

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

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

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  $f(\lambda) = (V + W * \mu)(c_\mu + \lambda v)$  satisfies  $f'(0) = 0$  and  $\forall r > 0, f''(r) \geq C$ . Hence,  $f$  is monotone increasing on  $[0, +\infty)$ , and in particular,

$$(2.10) \quad \int_0^2 e^{-f(\lambda)} \lambda^{d-1} d\lambda \geq e^{-f(2)} \int_0^2 \lambda^{d-1} d\lambda =: C_1 e^{-f(2)}.$$

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

$$\begin{aligned} \int_R^{+\infty} e^{-f(\lambda)} \lambda^{d-1} d\lambda &\leq e^{-f(2)} \int_R^{+\infty} \lambda^{d-1} e^{-2C(\lambda-2)/\sigma^2} d\lambda \\ &\leq e^{-f(2)} \int_R^{+\infty} \lambda^{d-1} e^{-2C(\lambda-2)} d\lambda \\ (2.11) \quad &\leq C_2 R^{d-1} e^{-2CR} \cdot e^{-f(2)} \leq C_3 e^{-CR} \cdot e^{-f(2)}. \end{aligned}$$

Comparing (2.10) and (2.11), we obtain the desired exponential decrease.  $\square$

**2.3. Conclusion.** Let us introduce the stopping time

$$(2.12) \quad T_\kappa(\sigma) := \inf \left\{ t_0 \geq 0 : \forall t \geq t_0, \left\| \frac{1}{t} \int_0^t X_s ds - m \right\|^2 \leq \kappa^2 \right\}.$$

By the convergence result, we know that for any  $\sigma > 0$ ,  $T_\kappa < \infty$ . Moreover, by the rate of convergence result, we even know that for any  $\kappa > 0$ , there exists  $T_\kappa < \infty$  such that if  $\sigma < 1$ , then  $T_\kappa(\sigma) < T_\kappa$ .

### 3. PROOF OF THEOREM 1.2

In this section, we will prove our main result. First, we will show that our process is close to the solution of a deterministic flow  $(\psi_t)_{t \geq 0}$  in §3.1. Using that, we will prove in §3.2 that the probability of leaving a stable 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 mean is frozen to  $m$  and we show that these diffusions are close in §3.3. We conclude the proof in §3.4.

**3.1. Majoration.** We remind the reader that in our work, the noise elapses. Consequently, it is natural to introduce the deterministic flow  $(\psi_t)_t$  defined by the following

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

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:*

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

*Proof.* We apply the Itô formula to obtain

$$\begin{aligned} \frac{1}{2} d|X_t - \psi_t|^2 &= (X_t - \psi_t, d(X_t - \psi_t)) + \frac{1}{2} d \langle X_t - \psi_t, X_t - \psi_t \rangle_t \\ &= \sigma(X_t - \psi_t, dB_t) + (X_t - \psi_t, -\nabla V(X_t) + \nabla V(\psi_t) - \\ &\quad - \alpha \left( X_t - \frac{1}{t} \int_0^t X_s ds - \psi_t + \frac{1}{t} \int_0^t \psi_s ds \right)) dt + \frac{\sigma^2}{2} dt \\ &\leq \sigma(X_t - \psi_t, dB_t) - (\rho + \alpha) |X_t - \psi_t|^2 dt + \frac{\alpha}{t} \int_0^t (X_t - \psi_t, X_s - \psi_s) ds dt + \frac{\sigma^2}{2} dt. \end{aligned}$$

The triangular inequality applied to the scalar product  $(X_t - \psi_t, X_s - \psi_s)$  now leads to

$$\begin{aligned} \frac{1}{2} d|X_t - \psi_t|^2 &\leq \sigma(X_t - \psi_t, dB_t) - (\rho + \frac{\alpha}{2}) |X_t - \psi_t|^2 dt + \frac{\alpha}{2t} \int_0^t |X_s - \psi_s|^2 ds dt + \frac{\sigma^2}{2} dt \\ &\leq \sigma(X_t - \psi_t, dB_t) + \frac{\alpha}{2t} \int_0^t |X_s - \psi_s|^2 ds dt + \frac{\sigma^2}{2} dt. \end{aligned}$$

Thus, we will find an upper bound for each term of the preceding inequality. First, we apply the BDG inequality for the local martingale term to get that there exists a positive constant  $C$  such that

$$\mathbb{E} \left( \sup_{[0,T]} \left| \sigma \int_0^t (X_s - \psi_s, dB_s) \right|^2 \right) \leq C\sigma^2 \mathbb{E} \sqrt{\int_0^T |X_s - \psi_s|^2 ds} \leq C\sigma^2 \sqrt{\int_0^T \mathbb{E}(|X_s - \psi_s|^2) ds}.$$

Then, we remark that

$$\int_0^T \frac{1}{t} \int_0^t \mathbb{E}(|X_s - \psi_s|^2) ds dt \leq \int_0^T \mathbb{E}(|X_t - \psi_t|^2) \log t dt.$$

Putting all the pieces together, we have

$$\xi^2 \mathbb{P} \left( \sup_{[0,T]} \|X_t - \psi_t(x_0)\|^2 > \xi \right) \leq C\sigma^2 \sqrt{\int_0^T \mathbb{E}(|X_s - \psi_s|^2) ds} + \frac{\alpha}{2} \log T \int_0^T \mathbb{E}(|X_u - \psi_u|^2) du + \frac{\sigma^2}{2} T.$$

As  $T$  is given and both  $X$  and  $\psi$  do not explode in a finite time, this leads to the result.  $\square$

**3.2. Probability of leaving before  $T_\kappa(\sigma)$ .** Remind that we denoted in (2.12) by  $T_\kappa(\sigma)$  the first time at which the empirical mean of the process remains stuck in the ball centered in  $m$  with radius  $\kappa$ .

**Proposition 3.2.** *We put  $\tau := \inf\{t \geq 0 : X_t \notin \mathcal{D}\}$  where  $\mathcal{D}$  is a domain stable by the drift  $x \mapsto -\nabla V(x) - \alpha(x - m)$ . For any  $\kappa > 0$ ,*

$$(3.3) \quad \lim_{\sigma \rightarrow 0} \mathbb{P}(\tau \leq T_\kappa(\sigma)) = 0.$$

*Proof.* First, we remark that  $T_\kappa(\sigma) \leq T_\kappa$  where  $T_\kappa$  does not depend on  $\sigma$ . Indeed, it has been proved that the exponential decrease is uniform with respect to the noise. 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  $\mathbb{B}(0; \epsilon) + \{\psi_t(x_0); 0 \leq t \leq T\} \subset \mathcal{D}$ , where  $\mathbb{B}(0; \epsilon)$  is the ball centered in 0 with radius  $\epsilon$ .

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 achieves 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 by making a coupling between the diffusion of interest ( $X$  here) and a diffusion that we expect to be close from  $X$  if the time is sufficiently large. The main difficulty with the self-stabilizing diffusion that was under consideration is in fact that we did 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  $\frac{1}{t} \int_0^t X_s ds$ ) remains stuck in a small ball of center  $m$  and radius  $\kappa$  for any  $t \geq T_\kappa(\sigma)$ . The idea thus is to replace  $\frac{1}{t} \int_0^t X_s ds$  by  $m$  and to compare the new diffusion with the self-interacting one.

In other words, we consider the diffusion

$$(3.4) \quad Y_t = X_{T_\kappa(\sigma)} + \sigma (B_t - B_{T_\kappa(\sigma)}) - \int_{T_\kappa(\sigma)}^t \nabla V(Y_s) ds - \alpha \int_{T_\kappa(\sigma)}^t (Y_s - m) ds,$$

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*

$$(3.5) \quad \mathbb{P} \left\{ \sup_{t \geq T_\kappa(\sigma)} \|X_t - Y_t\| \geq \xi \right\} = 0$$

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

$$d\|X_t - Y_t\|^2 = -2(X_t - Y_t; (\nabla V(X_t) + \alpha(X_t - \bar{\mu}_t)) - (\nabla V(Y_t) + \alpha(Y_t - m))) dt,$$

with  $\bar{\mu}_t := \frac{1}{t} \int_0^t X_s ds$ . We put  $W_m(x) := V(x) + \frac{\alpha}{2} \|x - m\|^2$ . We thus have

$$\begin{aligned} \frac{d}{dt} \|X_t - Y_t\|^2 &= -2(X_t - Y_t; \nabla W_m(X_t) - \nabla W_m(Y_t)) \\ &\quad + 2\alpha(X_t - Y_t; \bar{\mu}_t - m). \end{aligned}$$

However,  $\nabla^2 W_m = \nabla^2 V + \alpha \geq \rho + \alpha > 0$ . So, by putting  $\gamma(t) := \|X_t - Y_t\|^2$ , Cauchy-Schwarz inequality yields

$$\gamma'(t) \leq -2(\alpha + \rho)\gamma(t) + 2\sqrt{\gamma(t)}\kappa,$$

for any  $t \geq T_\kappa(\sigma)$ . However,  $\gamma(t) = 0$  for any  $t \leq T_\kappa(\sigma)$ . This means that  $\left\{ t \geq 0 : \gamma(t) > \frac{\kappa^2}{(\alpha + \rho)^2} \right\} \subset \{ t \geq 0 : \gamma'(t) < 0 \}$ . We deduce that

$$\sup_{t \geq T_\kappa(\sigma)} \gamma(t) \leq \frac{\kappa^2}{(\alpha + \rho)^2}.$$

The last term is smaller than  $\xi^2$  provided that  $\kappa < (\alpha + \rho)\xi$ . This achieves the proof.  $\square$

**3.4. Conclusion: proof of Theorem 1.2.** We now are in position to obtain the Kramers' type law for the exit-time and prove Theorem 1.2.

We decompose the proof in several parts. Let  $\xi > 0$ . It has been proved in [Tug12] that there exist two families of domains  $(\mathcal{D}_{i,\xi})_{\xi > 0}$  and  $(\mathcal{D}_{e,\xi})_{\xi > 0}$  such that

- $\mathcal{D}_{i,\xi} \subset \mathcal{D} \subset \mathcal{D}_{e,\xi}$ .
- $\mathcal{D}_{i,\xi}$  and  $\mathcal{D}_{e,\xi}$  are stable by  $W_m$ . The terminology “stable by” has been introduced in Definition 1.1.

- $\sup_{z \in \partial \mathcal{D}_{i,\xi}} d(z; \mathcal{D}^c) + \sup_{z \in \partial \mathcal{D}_{e,\xi}} d(z; \mathcal{D})$  tends to 0 when  $\xi$  goes to 0.
- $\inf_{z \in \partial \mathcal{D}_{i,\xi}} d(z; \mathcal{D}^c) = \inf_{z \in \partial \mathcal{D}_{e,\xi}} d(z; \mathcal{D}) = \xi$ .

We do not provide the proof since it is technical and we refer to [Tug12, Definition 2.1, Proposition 2.2] for the details.

**Step 2.** By  $\tau'_{i,\xi}(\sigma)$  (resp.  $\tau'_{e,\xi}(\sigma)$ ), we denote the first exit time of the diffusion  $Y$  from the domain  $\mathcal{D}_{i,\xi}$  (resp.  $\mathcal{D}_{e,\xi}$ ).

**Step 3.** We prove here the upper bound:

$$\begin{aligned}
\mathbb{P}\left(\tau \geq e^{\frac{2}{\sigma^2}(H+\delta)}\right) &= \mathbb{P}\left(\tau \geq e^{\frac{2}{\sigma^2}(H+\delta)}; \tau'_{e,\xi}(\sigma) \geq e^{\frac{2}{\sigma^2}(H+\delta)}\right) \\
&\quad + \mathbb{P}\left(\tau \geq e^{\frac{2}{\sigma^2}(H+\delta)}; \tau'_{e,\xi}(\sigma) < e^{\frac{2}{\sigma^2}(H+\delta)}\right) \\
&\leq \mathbb{P}\left(\tau'_{e,\xi}(\sigma) \geq e^{\frac{2}{\sigma^2}(H+\delta)}\right) \\
&\quad + \mathbb{P}\left(\tau \geq e^{\frac{2}{\sigma^2}(H+\delta)}; \tau'_{e,\xi}(\sigma) < e^{\frac{2}{\sigma^2}(H+\delta)}\right) \\
&=: a_\xi(\sigma) + b_\xi(\sigma).
\end{aligned}$$

This will be decomposed in three parts.

**Step 3.1.** First, by taking  $\xi$  sufficiently small, the exit-cost of the diffusion  $Y$  from the domain  $\mathcal{D}_{e,\xi}$  will be less than  $H + \frac{\delta}{3}$ . Since the diffusion  $Y$  is time homogeneous, we can apply the classical results of Freidlin-Wentzell theory and we thus obtain:

$$a_\xi(\sigma) = \mathbb{P}\left(\tau'_{e,\xi}(\sigma) \geq e^{\frac{2}{\sigma^2}(H+\frac{\delta}{3}+2\frac{\delta}{3})}\right) \longrightarrow 0$$

as  $\sigma > 0$  since  $2\frac{\delta}{3} > 0$ . Indeed, we remind the reader that the probability for a classical diffusion (with exit-cost  $H_0$ ) to exit at a time more than  $\exp\left\{\frac{2}{\sigma^2}(H_0 + \theta)\right\}$  tends to 0 as  $\sigma$  goes to 0, for any  $\theta > 0$ . We deduce that there exists  $\xi_1 > 0$  such that for all  $0 < \xi < \xi_1$ , we have:  $\lim_{\sigma \rightarrow 0} \mathbb{P}\left(\tau'_{e,\xi}(\sigma) \geq \exp\left[\frac{2}{\sigma^2}(H + \delta)\right]\right) = 0$ . Therefore, if  $\xi$  is small enough, then the first term  $a_\xi(\sigma)$  tends to 0 as  $\sigma$  goes to 0.

**Step 3.2.** Let us now control the second term  $b_\xi(\sigma)$ . On the event

$$\left\{\tau \geq e^{\frac{2}{\sigma^2}(H+\delta)}; \tau'_{e,\xi}(\sigma) < e^{\frac{2}{\sigma^2}(H+\delta)}\right\},$$

we know that at time  $\tau'_{e,\xi}(\sigma)$ , the distance between  $X$  and  $Y$  is at least  $\xi$  (since  $X$  is in  $\mathcal{D}$  whilst  $Y$  is outside  $\mathcal{D}_{e,\xi}$ ). Now, according to classical results on Freidlin-Wentzell theory, we know that the time  $\tau'_{e,\xi}(\sigma)$  is larger than  $T_\kappa(\sigma)$  with high probability (indeed,  $\tau'_{e,\xi}(\sigma)$  is at least  $\exp\left\{\frac{2}{\sigma^2}(H - \delta)\right\}$  for any  $\delta > 0$  and we have proved that  $T_\kappa(\sigma)$  is uniformly bounded with respect to  $\sigma$ ). Consequently, we have

$$\begin{aligned}
&\lim_{\sigma \rightarrow 0} \mathbb{P}\left(\tau \geq e^{\frac{2}{\sigma^2}(H+\delta)}; \tau'_{e,\xi}(\sigma) \leq e^{\frac{2}{\sigma^2}(H+\delta)}\right) \\
&= \lim_{\sigma \rightarrow 0} \mathbb{P}\left(\sup_{T_\kappa(\epsilon) \leq t} \|X_t - Y_t\| \geq \xi\right).
\end{aligned}$$

According to Theorem 3.3, there exists  $\kappa > 0$  such that the previous term is 0.

**Step 3.3.** Finally, choosing  $\xi$  and  $\kappa$  arbitrarily small, we obtain the upper bound

$$\lim_{\sigma \rightarrow 0} \mathbb{P} \left( \tau \geq \exp \left[ \frac{2}{\sigma^2} (H + \delta) \right] \right) = 0.$$

**Step 4.** Analogous arguments show that  $\lim_{\sigma \rightarrow 0} \mathbb{P} \left( T_\kappa(\sigma) \leq \tau \leq e^{\frac{2}{\sigma^2} (H - \delta)} \right) = 0$ . However, we have  $\lim_{\sigma \rightarrow 0} \mathbb{P}(\tau \leq T_\kappa(\sigma)) = 0$ . This ends the proof.

**Remark 3.4.** If we consider a non-convex potential  $V$  (but still convex at infinity), we have (1.7) if the domain  $\mathcal{D}$  is included into a domain in which  $V$  is convex (and contains a wells  $m$ ). Indeed, let us consider two potentials  $V_1$  and  $V_2$  such that  $V_1 = V_2$  on a compact  $\mathcal{K}$ . Then, for the self-interacting diffusion,  $X_t \mathbb{1}_{t \leq \tau_{\mathcal{K}}} = Y_t \mathbb{1}_{t \leq \tau_{\mathcal{K}}}$  where  $X$  and  $Y$  are two self-interacting diffusions defined with confining potentials  $V_1$  and  $V_2$  respectively. We also remark that this does not hold with the self-stabilizing diffusion since all the trajectories are taken into account.

#### REFERENCES

- [BLR02] Benaïm M., Ledoux M. & Raimond O. (2002), Self-interacting diffusions, *Prob. Th. Rel. Fields* **122**, 1-41.
- [BR05] Benaïm M. & 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. MR 2053570
- [CGM08] Cattiaux P., Guillin A. & Malrieu F. (2008), *Probab. Theory Relat. Fields* 140: 19.
- [CK11] Chambeu S. & Kurtzmann A. (2011), Some particular self-interacting diffusions: ergodic behavior and almost sure convergence, *Bernoulli* **17**(4), 1248-1267.
- [DZ98] A. Dembo and O. 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.
- [DR92] Durrett R.T. & Rogers L.C.G. (1992), Asymptotic behaviour of Brownian polymers, *Prob. Th. Rel. Fields* **92**(3), 337-349.
- [FW98] M. I. Freidlin and A. D. Wentzell (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] Samuel Herrmann, Peter Imkeller, and Dierk Peithmann (2008), Large deviations and a Kramers' type law for self-stabilizing diffusions, *Ann. Appl. Probab.*, 18(4):1379-1423, 2008.
- [HT18] Samuel Herrmann, Julian Tugaut, A Kramers' type law for self-stabilizing diffusions without convexity assumption. Preprint.
- [KK12] Victor Kleptsyn and Aline Kurtzmann (2012), Ergodicity of self-attracting motion, *Elect. J. Probab.*, Vol.17, paper no. 50.
- [AK10] 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. & Williams D. (1987), Self-avoiding random walk: a Brownian motion model with local time drift, *Prob. Th. Rel. Fields* **74**(2), 271-287.
- [Pe07] Pemantle R. (2007), A survey of random processes with reinforcement, *Probab. Surveys*, 1-79.
- [Ra09] Raimond O. (2009), Self-interacting diffusions: a simulated annealing version, *Prob. Th. Rel. Fields* **144**, 247-279.
- [RY98] Revuz D. & Yor M. (1998), *Continuous Martingales and Brownian Motion*, 3rd edition, Springer.
- [Tug12] J. Tugaut (2012), Exit problem of McKean-Vlasov diffusions in convex landscapes. *Electron. J. Probab.*, Vol. 17, no. 76, 1-26.
- [Tug16] J. Tugaut (2016), A simple proof of a Kramers' type law for self-stabilizing diffusions. *Electron. Commun. Probab.*, Vol. 21, no. 11, 1-7.

- [Tug17] J. Tugaut. A simple proof of a Kramers' type law for self-stabilizing diffusions in double-wells landscape. Preprint <https://hal.archives-ouvertes.fr/hal-01598978>
- [Tug18] J. Tugaut (2018), Exit problem of McKean-Vlasov diffusion in double-wells landscape. *J. Theoret. Probab.* 31, no. 2, 1013–1023.