Large Deviation Principle for invariant distributions of Memory Gradient Diffusions

Sébastien Gadat, Fabien Panloup and Clément Pellegrini
Institut de Mathématiques
Université de Toulouse (UMR 5219)
31062 Toulouse, Cedex 9, France
{Sebastien.Gadat;fabien.panloup;clement.pellegrini}@math.univ-toulouse.fr

Abstract

In this paper, we consider a class of diffusion processes based on a memory gradient descent, *i.e.* whose drift term is built as the average all along the trajectory of the gradient of a coercive function U. Under some classical assumptions on U, this type of diffusion is ergodic and admits a unique invariant distribution. In view to optimization applications, we want to understand the behaviour of the invariant distribution when the diffusion coefficient goes to 0. In the non-memory case, the invariant distribution is explicit and the so-called Laplace method shows that a Large Deviation Principle (LDP) holds with an explicit rate function, that leads to a concentration of the invariant distribution around the global minimums of U. Here, excepted in the linear case, we have no closed formula for the invariant distribution but we show that a LDP can be obtained. Then, in the one-dimensional case, we get some bounds for the rate function that lead to the concentration around the global minimum under some assumptions on the second derivative of U.

1 Introduction

This work deals with the evolution of a dynamical system whose drift is an average over all past positions of the gradient of a potential. For a given choice of positive and increasing real maps h and k, we are interested by the randomisation of the deterministic dynamical system defined by

$$\frac{dx}{dt} = -\left(\frac{1}{k(t)}\int_0^t h(s)\nabla U(x(s))ds\right). \tag{1.1}$$

The maps h and k quantify the amount of memory $(x(t))_{t\geq 0}$ used to compute the drift at time t. In a recent work, Cabot, Engler, and Gadat (2009b) study the capacity of the deterministic regime (1.1) to minimize the potential U and the influence of memory functions h and k has been carefully studied. A great interest of such differential equation is its ability to avoid some local trap of U, even in the deterministic setting. In fact, such ability is usually obtained by addition of a small diffusive term. Here, without any diffusive effect, the deterministic process $(x(t))_{t\geq 0}$ may keep some inertia even when it reaches a local minima of U and this inertia leads to a larger exploration of the space than a classical gradient descent which cannot escape from local minima. This is why such special case of equation (1.1) has received a lot of interest in the optimisation community: see for instance the works of Alvarez (2000) on the Heavy Ball with Friction system which is concerned with the case $k(t) = h(t) = e^{\lambda t}$, or the more recent work of Cabot (2009) for general increasing non negative maps h and k. Here, we will assume from now on that $h = k'$, which means that the drift term is a weighted average of the gradient along the past of the trajectory.

In view to optimization procedures, it is quite tempting to include in the dynamical system a small diffusive effect and then to study the following stochastic differential equation :

$$dX_t^\varepsilon = \varepsilon dW_t - \left(\frac{1}{k(t)}\int_0^t k'(s)\nabla U(X_s^\varepsilon)ds\right)dt, \tag{1.2}$$

which is a noisy version of the deterministic differential equation. The study of such a small noise perturbation model is of great interest in order to obtain the convergence of a simulated annealing algorithm.

A major difference with the usual gradient diffusion is that the integration over the past of the trajectory makes the process $(X_t^\varepsilon)_{t\geq 0}$ non Markovian. This difficulty can be overcome by enlarging the state space and introducing an auxiliary process $(Y_t^\varepsilon)_{t\geq 0}$. Roughly speaking, the couple $(X_t^\varepsilon, Y_t^\varepsilon)_{t\geq 0}$ is a (generally non-homogeneous) Markov process which corresponds to the position and the speed of the heavy ball. The model described by (1.2) is a particular case of the so-called diffusion with memory gradient. In a more general setting, several results concerning the long time behaviour of such a diffusion are obtained by Gadat and Panloup (2011). Using a non-usual Lyapunov function, the authors show that, under some classical mean-reverting assumptions, the system is long-time stable and converges to a stationary regime if (and only if) $t \mapsto k(t)$ grows at least exponentially. In particular, when $k(t) = e^{\lambda t}$ (this corresponds to the case where $(X_t^\varepsilon, Y_t^\varepsilon)$ is an-homogeneous Markov process), the authors obtain that the system is hypo-elliptic and *approximatively controllable* (under some non-degeneracy assumptions on D^2U). These properties imply uniqueness for the invariant distribution ν_ε of the system and convergence results to ν_ε are obtained (including convergence rates).
In this work, we only focus on the (homogeneous) case $k(t) = e^{\lambda t}$ and our objective is to obtain some sharp estimations of the asymptotic behavior of (ν_ε) as $\varepsilon \to 0$. More precisely, we want in a first step to obtain some Large Deviation Principle (LDP) for $(\nu_\varepsilon)_{\varepsilon>0}$ and in a second step to get some sharp bounds for the associated rate function (also called quasi-potential). In fact, this second step will allow us to understand how the invariant probability is distributed as $\varepsilon \to 0$.

Our motivation is twofold. The first one is a forthcoming study of the simulated annealing problem. In this way, we want to show that the invariant distribution concentrates on the global minimum as $\varepsilon \to 0$. Such a result is one of the two main steps for proving the convergence in probability of the simulated annealing algorithm to the global minimum. The second one is a sharp study of the rate of convergence of the semi-group to the invariant distribution with a fixed ε. On this subject, we refer for instance to R. A. Holley, Kusuoka, and Stroock (1989) where the authors study the behaviour of the spectral gap of reversible operators (whose drift is a gradient of a potential) in the elliptic setting, and to Miclo (1992a) who extends the preceding results to the case of a general drift defined on a compact manifold. Note that this problem will not be tackled in this paper and that these works which are based on Sobolev inequalities seem to be difficult to adapt to our non compact and degenerated setting. Note also that the simulated annealing problem can also be directly investigated by studying the long time behaviour of the distribution of the diffusion with decreasing diffusing term $(\sigma(t))_{t\geq 0}$ (see e.g. Miclo (1992b)) but this approach seems also difficult to be adapted here.

The other motivation of this paper is to extend some results of Large Deviations for invariant distributions to a difficult context where the process is degenerated and the drift vector field is not the gradient of a potential. These two points and especially the second one strongly complicate the problem since explicit computations are generally impossible. This implies that the works on the elliptic Kolmogorov equation by Chiang, Hwang, and Sheu (1987), Miclo (1992b) or R. Holley and Stroock (1988) for instance, can not be extended to our context. In

the same spirit, one should also mention the more recent works on Mac-Kean Vlasov diffusions by Herrmann and Tugaut (2010) and on of self-interacting (with attractive potential) diffusions by Raimond (2009). These are two examples of similar studies in a non-Markovian setting.
Here, our roadmap to obtain a LDP for $(\nu_\varepsilon)_{\varepsilon\geq 0}$ (which is the first step of the work towards simulated annealing) is to adapt some results from Puhalskii (2003) and Freidlin and Wentzell (1979) to our degenerated context. Owing to a criterion of Puhalskii (2003) based on the finite-time LDP for the underlying stochastic process and to the control of some hitting times of compact sets, we deduce from Lyapunov-type arguments the exponential tightness of $(\nu_\varepsilon)_{\varepsilon\geq 0}$ and show that the associated rate function can be viewed as the solution of a control problem (or equivalently to an Hamilton-Jacobi equation). However, this approach has two drawbacks. The solution of the control problem is not unique and not very explicit. Then, adapting Freidlin and Wentzell (1979) to our hypoelliptic setting, we obtain a formulation of the rate function in terms of the costs to join stable critical points of our dynamical system (which implies in particular uniqueness of the limit). Then, the second step of the paper (sharp estimates of W) is investigated by the study of the cost to join stable critical points. More precisely, we obtain some upper and lower bounds for the cost which allow us to conclude (under some conditions on the second derivative of U) that the invariant distribution concentrates on the global minimum as $\varepsilon \to 0$.

The paper is then organized as follows. In Section 2, we recall some results about the long-time behaviour of the diffusion when ε is fixed. Then, we provide our main assumptions and summarize the results we obtained with the asymptotic behaviour as $\varepsilon \to 0$ of the dynamical system defined by (1.2). In Section 3, we prove the exponential tightness of (ν_ε) and show that any rate function W associated with a convergent subsequence is a solution of a finite or infinite time control problem. In Section 4, we prove the uniqueness of W by adapting the Freidlin and Wentzell approach to our context. As mentioned before, this approach gives us an explicit formulation of W in terms of the costs to join critical points. The study of the cost function being too difficult in a general setting, we focus in Section 5 on the case of a double-well potential U and obtain some upper and lower bounds for the associated cost function. Then, we provide some conditions on the second derivative of U and on the memory parameter λ which allow us to obtain the concentration of the invariant distribution around the global minimum. Note that, even if our assumptions in this part seem a little bit restrictive, the proofs of the bounds (especially the lower bound) appear to be obtained using Lyapunov functions in an original and almost optimal way.
Acknowledgments: The authors would like to thank Guy Barles for its hospitality, and are grateful to Guy Barles, Laurent Miclo and Christophe Prieur for helpful discussions and comments.

2 Setting and Main Results

2.1 Notations and background on Large Deviation theory

Before a precise definition of the dynamical system, let us list a short series of notations.
We consider a process living in $\mathbb{R}^d \times \mathbb{R}^d$, the scalar product and the Euclidean norm on $\mathbb{R}^d$ are respectively denoted by $\langle\, ,\, \rangle$ and $|.|$.
We denote by $\mathbb{H}(\mathbb{R}_+, \mathbb{R}^d)$ the Cameron-Martin space, *i.e.* the set of absolutely continuous functions $\varphi : \mathbb{R}_+ \to \mathbb{R}^d$ such that $\varphi(0) = 0$ and such that $\dot{\varphi} \in L^{2,loc}(\mathbb{R}_+, \mathbb{R}^d)$ (where $L^{2,loc}(\mathbb{R}_+, \mathbb{R}^d)$ denotes the set of locally square integrable functions from $\mathbb{R}^+$ to $\mathbb{R}^d$).

For a $\mathcal{C}^2$-function $f : \mathbb{R}^d \to \mathbb{R}$, ∇f and $D^2 f$ denote respectively the gradient of f and the

Hessian matrix of f. In the one-dimensional case, we will switch to the notation f' and f" in order to emphasize the difference with $d > 1$.

Given any $\mathcal{C}^2$-function $f : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, $\nabla_x f : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ and $D^2_x f : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{M}_d(\mathbb{R})$ denote the functions respectively defined by $(\nabla_x f(x,y))_i = \partial_{x_i} f(x,y)$ and $(D^2_x f(x))_{i,j} = \partial_{x_i}\partial_{x_j} f(x,y)$. Obviously these notations are naturally extended to $\nabla_y f$, $D^2_{x,y} f$ and $D^2_y f$. At last, for any vector $v \in \mathbb{R}^d$, $v^\star$ will refer to the transpose of v.
For a measure μ and a μ-measurable function f, we set $\mu(f) = \int f d\mu$.

Let us now recall some definitions relative to the Large Deviation theory (see Dembo and Zeitouni (2010) for further references on the subject). Let (E, d) denote a metric space. One says that a family of probability measures $(\nu_\varepsilon)_{\varepsilon>0}$ on E satisfies a Large Deviation Principle (denoted LDP in the sequel) with speed r_ε and rate function I if and only if for every open set O and closed set F,

$$\liminf_{\varepsilon\to 0} r_\varepsilon \log(\nu_\varepsilon(O)) \geq -\inf_{x\in O} I(x) \quad \text{and} \quad \limsup_{\varepsilon\to 0} r_\varepsilon \log(\nu_\varepsilon(F)) \leq -\inf_{x\in F} I(x).$$

The function I is refered to be *good* if for any $c \in \mathbb{R}$, $\{x \in E, I(x) \leq c\}$ is compact. In this paper, we will use some classical compactness results in Large Deviation theory. More precisely, a family of probability measures $(\nu_\varepsilon)_{\varepsilon>0}$ is said to be *exponentially tight of order* r_ε if

$$\forall a > 0,\ \exists K_a \text{ compact of } E \text{ such that } \limsup_{\varepsilon\to 0} r_\varepsilon \log(\nu_\varepsilon(K_a^c)) \leq -a.$$

Then, we recall the link between exponential tightness and the Large Deviation Principle (see Feng and Kurtz (2006), chapter 3 for instance).

Proposition 2.1 *Let* (S, d) *be a metric space and* $(\nu_\varepsilon)_{\varepsilon\geq 0}$ *a sequence of exponentially tight probability measures on the Borel* σ*-algebra of* S *with speed* r_ε*. Then there exists a subsequence* $(\varepsilon_k)_{k\geq 0}$ *such that* $\varepsilon_k \to 0$ *along which the Large Deviation Principle holds with good rate function* I *and speed* r_{ε_k}*.*

Definition 2.1 *A subsequence* $(\nu_{\varepsilon_k})_{k\geq 1}$ *satisfying Proposition 2.1 will be called a* (LD)*-convergent subsequence.*

2.2 Averaged gradient diffusions

Throughout this paper, we denote by $U : \mathbb{R}^d \mapsto \mathbb{R}$ a smooth (at least $\mathcal{C}^2$) function on $\mathbb{R}^d$ and coercive, *i.e.*

$$\inf_{x\in\mathbb{R}} U(x) > 0, \qquad \lim_{|x|\to+\infty} U(x) = +\infty, \quad \text{and} \quad \liminf_{|x|\to+\infty} \langle x, \nabla U(x)\rangle > 0. \tag{2.1}$$

We consider $\lambda > 0$ and we are interested in the behaviour of a process described through the following stochastic differential equation whose drift averages with memory a gradient of U over all the past of the trajectory. More precisely, let $(W_t)_{t\geq 0}$ be a standard d-dimensional brownian motion, we consider $(X^\varepsilon_t)_{t\geq 0}$ the process which lives in $\mathbb{R}^d$ described by

$$dX^\varepsilon_t = \varepsilon dW_t - \left(\lambda e^{-\lambda t}\int_0^t e^{\lambda s}\nabla U(X^\varepsilon_s)ds\right)dt.$$

If we consider a space enlargement through the definition of an auxiliary process $(Y^\varepsilon_t)_{t\geq 0}$ (which also lives in $\mathbb{R}^d$)

$$Y^\varepsilon_t = \lambda e^{-\lambda t}\int_0^t e^{\lambda s}\nabla U(X^\varepsilon_s)ds,$$

one can show (see Gadat and Panloup (2011) for instance) that $(Z_t^\varepsilon)_{t\geq 0} := ((X_t^\varepsilon, Y_t^\varepsilon))_{t\geq 0}$ is solution of:

$$\begin{cases} dX_t^\varepsilon = \varepsilon dW_t - Y_t^\varepsilon dt. \\ dY_t^\varepsilon = \lambda(\nabla U(X_t^\varepsilon) - Y_t^\varepsilon)dt. \end{cases} \tag{2.2}$$

In the sequel, we will also intensively use the deterministic system obtained setting $\varepsilon = 0$ in the above dynamical system (2.2), which is defined by

$$\begin{cases} \dot{x}(t) = -y(t). \\ \dot{y}(t) = \lambda(\nabla U(x(t)) - y(t)). \end{cases} \tag{2.3}$$

2.3 Assumptions

The function ∇U being not necessarily Lipschitz continuous, we assume in all the paper the following assumption:

$(\mathbf{H_0})$: There exists $C > 0$ such that for every $x \in \mathbb{R}^d$, $\|D^2U(x)\| \leq CU(x)$.

At this stage assumption $(\mathbf{H_0})$ ensures the non-explosion (in finite horizon) of $(Z_t^\varepsilon)_{t\geq 0}$ (see Proposition 2.1 of Gadat and Panloup (2011)). In particular, under $(\mathbf{H_0})$, existence and uniqueness hold for the solution of (2.2) and $(Z_t^\varepsilon)_{t\geq 0}$ is a Markov process with infinitesimal generator $\mathcal{A}^\varepsilon$ defined for every $f \in \mathcal{C}_c^2(\mathbb{R}^d \times \mathbb{R}^d)$, by:

$$\mathcal{A}^\varepsilon f(x,y) = -\langle y, \partial_x f\rangle + \lambda\langle \nabla U(x) - y, \partial_y f\rangle + \frac{\varepsilon^2}{2}\mathrm{Tr}\left(D_x^2 f\right). \tag{2.4}$$

We first recall some results obtained by Gadat and Panloup (2011) on existence and uniqueness for the invariant distribution of (2.2). To this end, we need to introduce a mean-reverting assumption denoted by $(\mathbf{H_{mr}})$ and some hypo-ellipticity assumption $(\mathbf{H_{Hypo}})$. The mean reverting assumption is expressed as follows:

$(\mathbf{H_{mr}})$: $\quad \lim_{|x|\to+\infty} -\langle x, \nabla U(x)\rangle = -\infty \quad$ and $\quad D^2U(x) = o(\langle x, \nabla U(x)\rangle) \quad$ as $|x| \to +\infty$.

Note that $(\mathbf{H_{mr}})$ implies Assumption $(\mathbf{H_1})$ of Gadat and Panloup (2011) in the particular case $\sigma = I_d$ and $r_\infty = \lambda$.
Concerning the hypo-ellipticity assumption, let us define $\mathcal{E}_U$ as

$$\mathcal{E}_U = \left\{x \in \mathbb{R}^d,\ \det\left(D^2U(x)\right) \neq 0\right\}, \tag{2.5}$$

and $\mathcal{M}_U$ the complementary manifold $\mathcal{M}_U = \mathbb{R}^d \setminus \mathcal{E}_U$. The hypothesis needed to obtain hypo-ellipticity of the process is given below.

$(\mathbf{H_{Hypo}})$: U is $\mathcal{C}^\infty(\mathbb{R}^d, \mathbb{R}), \lim_{|x|\to+\infty} \frac{U(x)}{|x|} = +\infty \quad$ and $\quad \dim(\mathcal{M}_U) \leq d-1$.

Note that the conditions concerning the limit $\lim_{|x|\to+\infty} U(x)/|x| = +\infty$ imposes that we consider potential which growth at least linearly at infinity. Under these assumptions, we deduce the following proposition from Theorems 2.3 and 3.2 of (Gadat & Panloup, 2011):

Proposition 2.2 *Assume* $(\mathbf{H_0})$ *and* $(\mathbf{H_{mr}})$. *Then, for every* $\varepsilon > 0$, *the solution of* (2.2) *admits an invariant distribution. Furthermore, if* $(\mathbf{H_{Hypo}})$ *holds, then the invariant distribution is unique and admits a* λ_{2d}*-a.s. positive density. We shall denote* ν_ε *this invariant distribution.*

Note that throughout the paper, we will adopt the same notation for the distribution ν_ε and for its density. Under the previous assumptions, we can now focus on the asymptotic behaviour of $(\nu_\varepsilon)_{\varepsilon>0}$ as ε goes to 0. In fact, as mentioned before, the aim is to obtain a Large Deviation Principle and in particular the exponential tightness of $(\nu_\varepsilon)_{\varepsilon>0}$. To this end, we need to introduce some more constraining mean-reverting assumptions than $(\mathbf{H_{mr}})$:

$(\mathbf{H_{Q+}})$: There exists $\rho \in (0,1)$, $\beta \in \mathbb{R}$ and $\alpha > 0$ such that

$$\begin{array}{ll} \text{(i)} & -\langle x, \nabla U(x)\rangle \leq \beta - \alpha U(x), \forall x \in \mathbb{R}^d \\ \text{(ii)} & |\nabla U| = O(U^{1-\rho}) \quad \text{and} \quad D^2 U = o(U) \quad \text{as } |x| \to +\infty. \end{array}$$

$(\mathbf{H_{Q-}})$: There exists $a \in (1/2, 1]$, $C > 0$, $\beta \in \mathbb{R}$ and $\alpha > 0$ such that

$$\begin{array}{ll} \text{(i)} & -\langle x, \nabla U(x)\rangle \leq \beta - \alpha |x|^{2a}, \forall x \in \mathbb{R}^d \\ \text{(ii)} & |\nabla U|^2 \leq C(1+U) \quad \text{and} \quad \sup_{x\in\mathbb{R}^d} \|D^2 U(x)\| < +\infty. \end{array}$$

Let us stress that the previous condition $(\mathbf{H_0})$ is covered by $(\mathbf{H_{Q+}})$ and $(\mathbf{H_{Q-}})$ and that we do not need it anymore for the sequel of the paper.

Remark 2.1 *Assumptions* $(\mathbf{H_{Q+}})$ *and* $(\mathbf{H_{Q-}})$ *correspond respectively to over-quadratic and subquadratic potentials. For instance, assume that* $U(x) = (1+|x|^2)^p$. *When* $p \geq 1$, $(\mathbf{H_{Q+}})$ *holds with* $\rho \in (1 - \frac{1}{2p}, 1)$ *and when* $p \in (1/2, 1]$, $(\mathbf{H_{Q-}})$ *holds with* $a = p$.
In fact, these assumptions are adapted to a large class of potentials U *with polynomial growth (more than linear). However, these assumptions do not cover the potentials with exponential growth (for which* $(\mathbf{H_{Q+}})$(ii) *is not fulfilled).*

2.4 Main results

2.4.1 Exponential tightness and Hamilton Jacobi equation

Let $\varphi \in \mathbb{H}$. When existence holds, we denote respectively by $\mathbf{z}_\varphi := (\mathbf{z}_\varphi(t))_{t\geq 0}$ and by $\tilde{\mathbf{z}}_\varphi := (\tilde{\mathbf{z}}_\varphi(t))_{t\geq 0}$, a solution of

$$\dot{\mathbf{z}}_\varphi = b(\mathbf{z}_\varphi) + \begin{pmatrix} \dot\varphi \\ 0 \end{pmatrix} \quad \text{and} \quad \dot{\tilde{\mathbf{z}}}_\varphi = -b(\tilde{\mathbf{z}}_\varphi) + \begin{pmatrix} \dot\varphi \\ 0 \end{pmatrix}. \tag{2.6}$$

Note that $(\mathbf{H_{Q+}})$ and $(\mathbf{H_{Q-}})$ ensure the finite-time non-explosion of $\mathbf{z}_\varphi$ and $\tilde{\mathbf{z}}_\varphi$ for every $\varphi \in \mathbb{H}$ (see e.g. Equation (3.4)). Thus, since ∇U is locally Lipschitz continuous, for every $z \in \mathbb{R}^{2d}$, the solutions starting from z respectively denoted by $\mathbf{z}_\varphi(z,.)$ and $\tilde{\mathbf{z}}_\varphi(z,.)$ exist and are unique.

We are now able to state our first main result:

Theorem 2.1 *Assume* $(\mathbf{H_{Hypo}})$ *and* $(\mathbf{H_{Q+}})$ *or* $(\mathbf{H_{Q-}})$. *Then,*

(i) $(\nu_\varepsilon)_{\varepsilon\in(0,1]}$ *is exponentially tight on* $\mathbb{R}^{2d}$ *with speed* ε^{-2}.

(ii) Let $(\varepsilon_n)_{n\geq 1}$ *be a* (LD)*-convergent subsequence and denote by W the associated (good) rate function. Then, W satisfies for every* $t \geq 0$ *and any* $z \in \mathbb{R}^d \times \mathbb{R}^d$*:*

$$W(z) = \inf_{\varphi\in\mathbb{H}} \left[\frac{1}{2}\int_0^t |\dot\varphi|^2 + W(\tilde{\mathbf{z}}_\varphi(z,t))\right]. \tag{2.7}$$

(iii) Furthermore, assume that $\{x \in \mathbb{R}^d, \nabla U(x) = 0\} = \{x_1^\star, \dots, x_\ell^\star\}$ *(*$\ell \in \mathbb{N}$*) and that for every* $i \in \{1, \dots, \ell\}$*,* $D^2U(x_i^\star)$ *is invertible. Then,*

$$W(z) = \min_{1 \le i \le \ell} \inf_{\begin{cases} \varphi \in \mathbb{H} \\ \tilde{\mathbf{z}}_\varphi(z, +\infty) = z_i^\star \end{cases}} \left[\frac{1}{2} \int_0^\infty |\dot{\varphi}|^2 + W(z_i^\star) \right]. \tag{2.8}$$

where $\tilde{\mathbf{z}}_\varphi(z, +\infty) := \lim_{t \to +\infty} \tilde{\mathbf{z}}_\varphi(z, t)$ *(when exists) and* $z_i^\star = (x_i^\star, 0)$ *for all* $i = 1, \dots, l$.

Equation (2.7) satisfied by W may be seen as an Hamilton-Jacobi equation (see *e.g.* Barles (1994) for further details on such equations).

2.4.2 Freidlin and Wentzell estimates

Let us stress that the main problem in the expression (2.8) is that the uniqueness is only available conditionally to the values of $W(z_i^\star)$, $i = 1, \dots \ell$. However, using that W is also defined as a rate function related to a (LD)-convergent subsequence of (ν_ε), we are going to show that the values of $W(z_i^\star)$ are uniquely determined. This is provided by the Freidlin and Wentzell (1979) approach in the case of the finite number of equilibriums of the ordinary differential equation (2.3). We thus make the next assumption.

(**D**) : The set of critical points $(x_i^\star)_{i=1\dots\ell}$ of U is discrete and each $D^2U(x_i^\star)$ is invertible.

Under the assumption $\lim_{|x| \to +\infty} \langle x, \nabla U(x) \rangle > 0$, it follows that the set of critical points is finite. In what follows we recall some useful elements of Freidlin and Wentzell theory needed to ensure the uniqueness of W in Theorem 2.1.

{i}-Graphs Following the notations of Theorem 2.1, we denote $\{z_1^\star, \dots, z_\ell^\star\}$ this finite set of equilibriums. For sake of completeness, we recall here the definition of {i}-Graphs defined on the finite set $\{z_1^\star, \dots, z_\ell^\star\}$. For any $i \in \{1, \dots, \ell\}$, we denote $\mathcal{G}(i)$ the set of oriented graphs with vertices $\{z_1^\star, \dots, z_\ell^\star\}$ that satisfies the three following properties.

(i) Each state $z_j^\star \neq z_i^\star$ is the initial point of exactly one oriented edge in the graph.

(ii) The graph does not have any cycle.

(iii) For any $z_j^\star \neq z_i^\star$, there exists a (unique) path composed of oriented edge starting at state $z_j^\star$ and leading to the state $z_i^\star$.

L^2 control cost between between equilibriums We now define for any couple of points $(\xi_1, \xi_2) \in (\mathbb{R}^d \times \mathbb{R}^d)^2$ the minimal L^2 cost to go from ξ_1 to ξ_2 within a finite time t as

$$I_t(\xi_1, \xi_2) = \inf_{\begin{cases} \varphi \in \mathbb{H} \\ \mathbf{z}_\varphi(\xi_1, t) = \xi_2 \end{cases}} \frac{1}{2} \int_0^t |\dot{\varphi}(s)|^2 ds,$$

and also the minimal L^2 cost to go from ξ_1 to ξ_2 within any time:

$$I(\xi_1, \xi_2) = \inf_{t \ge 0} I_t(\xi_1, \xi_2).$$

The function I is the so-called usual *quasipotential.* With these definitions, one can obtain the Freidlin and Wentzell estimate which gives another representation of $W(z_i^\star)$, $i = 1, \dots, \ell$.

Theorem 2.2 *Assume* $(\mathbf{H_{Hypo}})$ *and* $(\mathbf{H_{Q+}})$ *or* $(\mathbf{H_{Q-}})$*. If* $(\boldsymbol{D})$ *holds, then any adherence point* W *obtained with a (LD)-convergent subsequence satisfies*

$$\forall i \in \{1 \dots \ell\} \qquad W(z_i^\star) = \mathcal{W}(z_i^\star) - \min_{j \in \{1,\dots,\ell\}} \mathcal{W}(z_j^\star)$$

where

$$\forall i \in \{1 \dots \ell\} \qquad \mathcal{W}(z_i^\star) = \min_{g \in \mathcal{G}(i)} \sum_{(z_m^\star \to z_n^\star) \in g} I(z_m^\star, z_n^\star). \tag{2.9}$$

The next corollary follows immediately from Theorem 2.1 and Theorem 2.2.

Corollary 1 *Assume* $(\mathbf{H_{Hypo}})$ *and* $(\mathbf{H_{Q+}})$ *or* $(\mathbf{H_{Q-}})$*. If* $(\boldsymbol{D})$ *holds,* (ν_ε) *satisfies a large deviation principle with speed* ε^{-2} *and good rate function* W *such that*

$$W(z) = \min_{1 \le i \le \ell} \inf_{\begin{cases} \varphi \in \mathbb{H} \\ \tilde{\mathbf{z}}_\varphi(z, +\infty) = z_i^\star \end{cases}} \left[\frac{1}{2} \int_0^\infty |\dot{\varphi}|^2 + W(z_i^\star) \right],$$

where $W(z_i^\star)$ *is given by (2.9).*

Case of a double-well non-convex potential In the sequel, we are interested by the location of the minimum of W. More precisely, we expect that this minimum is located on the set of global minimums of U. This point is clear in the case of strictly convex potential U using Equation (2.8). Regarding now the non-convex case, the situation is more complicated. Thus, we only focus on the double-well one-dimensional case. Without loss of generality, we assume that U has two local minima denoted by $x_1^\star$ and $x_2^\star$ with

$$x_1^\star < x^\star < x_2^\star \quad \text{and} \quad U(x_1^\star) < U(x_2^\star), \tag{2.10}$$

where $x^\star$ is the unique local maximum between x_1 and x_2. We obtain the following result:

Theorem 2.3 *Assume the hypothesis of Corollary 1 and that* U *satisfies* (2.10)*. Then,*
(i) $\mathcal{W}$ *is bounded and we have*

$$\mathcal{W}(z_1^\star) = I(z_2^\star, z_1^\star) \le 2[U(x^\star) - U(x_2)].$$

(ii) *For every* $\alpha \in [0, 2]$*, there exists an explicit constant* $m_\lambda(\alpha)$ *such that*

$$\|U''\|_\infty \le m_\lambda(\alpha) \Longrightarrow \mathcal{W}(z_2^\star) = I(z_1^\star, z_2^\star) \ge \alpha[U(x^\star) - U(x_1)].$$

(iii) *As a consequence, if* U *satisfies* $\|U''\|_\infty \le m_\lambda\left(2\frac{U(x^\star)-U(x_2)}{U(x^\star)-U(x_1)}\right)$*, then*

$$\mathcal{W}(z_1^\star) < \mathcal{W}(z_2^\star),$$

and finally,

$$\nu_\varepsilon \xrightarrow{\varepsilon \to 0} \delta_{z_1^\star}.$$

In the next sections, we prove the above statements. Note that throughout the rest of the paper, C will stand for any non-explicit constant. Note also that excepted in Section 5, we will prove all the results with $\lambda = 1$ for sake of convenience (one can deduce similar convergences with small modifications for any $\lambda > 0$).

3 Large Deviation Principle for invariant measures $(\nu_\varepsilon)_{\varepsilon\in(0,1]}$

This section describes the proof of Theorem 2.1 which contains two important parts. The first one concerns the exponential tightness of the invariant measures $(\nu_\varepsilon)_{\varepsilon\in(0,1]}$ although the second result is a functional equality for *any* good rate function associated to *any* (LD)-convergent subsequence $(\nu_{\varepsilon_k})_{k\geq 0}$.

We first establish a trajectorial finite time LDP for the stochastic process $(Z^\varepsilon_t)_{\varepsilon\geq 0}$ and then we detail how one can derive the exponential tightness property of $(\nu_\varepsilon)_{\varepsilon\in(0,1]}$ using the existence of Lyapunov functions for our dynamical system. At last, we show that a functional equality such as (2.7) holds.

3.1 Large Deviation Principle for $(Z^\varepsilon)_{\varepsilon>0}$

The next lemma establishes a Large Deviation Principle for trajectories of the process $((Z^\varepsilon_t)_{t\geq 0})_{\varepsilon>0}$ within a finite time.

Lemma 3.1 *Assume $(\mathbf{H_{Q+}})$ or $(\mathbf{H_{Q-}})$, for every $z\in\mathbb{R}^d$ and sequence $(z_\varepsilon)_{\varepsilon>0}$ such that $z_\varepsilon\to z$ as $\varepsilon\to 0$, the coupled process $Z^\varepsilon=(X^\varepsilon,Y^\varepsilon)$ satisfies a large deviation principle on $\mathcal{C}(\mathbb{R}_+,\mathbb{R}^{2d})$ (endowed with the topology of uniform convergence on compact sets) with speed ε^{-2} and (good) rate function $\mathcal{I}_z$ defined for every absolutely continuous $(\mathbf{z}(t))_{t\geq 0}$ by*

$$\mathcal{I}_z((\mathbf{z}(t))_{t\geq 0})=\frac{1}{2}\inf_{\begin{cases}\varphi\in\mathbb{H}\\ \forall t\geq 0,\mathbf{z}_\varphi(z,t)=\mathbf{z}(t)\end{cases}}\int_0^\infty|\dot\varphi(s)|^2ds.$$

In particular, for every $t\geq 0$, for every $z\in\mathbb{R}^{2d}$, $(P^\varepsilon_t(z_\varepsilon,.))_{\varepsilon>0}$ satisfies a LDP with speed ε^{-2} and rate function $I_t(z,.)$ defined for every $z,z'\in\mathbb{R}^{2d}$ by:

$$I_t(z,z')=\inf_{\mathbf{z}(.)\in\mathcal{Z}_t(z,z')}\mathcal{I}_z(\mathbf{z}(.)),\tag{3.1}$$

where $\mathcal{Z}_t(z,z')$ denotes the set of absolutely continuous functions $\mathbf{z}(.)$ such that $\mathbf{z}(0)=z$, $\mathbf{z}(t)=z'$. Furthermore, the function I_t can be written as

$$I_t(z,z')=\frac{1}{2}\inf_{\begin{cases}\varphi\in\mathbb{H}\\ \mathbf{z}_\varphi(z,t)=z'\end{cases}}\int_0^t|\dot\varphi(s)|^2ds.$$

Remark 3.1 *Note that such result is quite classical when $z_\varepsilon=z$ and when the coefficients are Lipschitz continuous functions (see e.g. Azencott (1980) for instance), but in our setting, we may be interested by vector fields that may possess some over linear growth. Note also that our process is degenerate in the second coordinate y and then the classical approach of Freidlin and Wentzell (1979) does not apply in our setting. At last, remark that one step of this lemma is similar to a contraction principle and is proved using a suitable Lyapunov function.*

<u>Proof :</u> We wish to apply Theorem 5.2.12 of Puhalskii (2001) and we first stand for hypothesis needed to apply this theorem. We recall that the assumptions are as follows.

- Let $\mathbb{D}$ the Skorohod space of processes $\mathbb{R}_+\mapsto\mathbb{R}^d$, $F=\{F_t(z),t\geq 0,z\in\mathbb{D}\}$ is said to satisfy the local majoration condition if forall $b>0$, there exists an increasing continuous map $\bar F$ such that

$$\forall 0\leq s\leq t\qquad \sup_{\mathbf{z}\in\mathbb{D},\mathbf{z}_\infty<+\infty}(F_t(\mathbf{z})-F_s(\mathbf{z}))\leq\bar F_t-\bar F_s.$$

- The Non-Explosion condition, referred as (NE) in the sequel, holds if
  - the function π_z defined by $\pi_z := \exp(-\mathcal{I}_z(\mathbf{z}))$ is upper-compact,
  - for all $t \geq 0$ and for all $a \in (0,1]$, the set $\bigcup_{s\leq t}\left\{\sup_{u\leq s}|\mathbf{z}(u)| \mid \pi_{z,s}(\mathbf{z}) \geq a\right\}$ is bounded where

$$\pi_{z,s}(\mathbf{z}) = \exp\left(-\inf_{\varphi,\forall s\in[0,t],\mathbf{z}_\varphi(z,s)=\mathbf{z}(s)} \int_0^t |\dot\varphi(s)|^2 ds\right).$$

Local majoration condition: To this end, we introduce the predictable characteristics of Z^ε:

$$B_\varepsilon(t) := \int_0^t b(X_s^\varepsilon, Y_s^\varepsilon)ds \qquad \text{and} \qquad C_\varepsilon(t) := C(t) = t.$$

First, one has to prove that for every $t \geq 0$ the function ϕ_t from $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^{2d})$ to $\mathbb{R}^{2d}$ defined by $\phi_t(\mathbf{z}) = \int_0^t b(\mathbf{z}(s))ds$ is a continuous function of $\mathbf{z}$. Let $\mathbf{z}^{(n)}$ be a sequence of $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^{2d})$ such that $\mathbf{z}^{(n)} \to \mathbf{z}$ as $n \to +\infty$ where $\mathbf{z} \in \mathcal{C}(\mathbb{R}_+, \mathbb{R}^{2d})$. Then, $\|\mathbf{z}^{(n)} - \mathbf{z}\|_{[0,t]} \to 0$ and in particular $M := \|\mathbf{z}(s)\|_{[0,t]} \vee \sup_{n\in\mathbb{N}} \|\mathbf{z}^{(n)}\|_{[0,t]}$ is finite. Since b is a local Lipschitz function, we know that b is Lipschitz continuous on $B(0, M)$ and it follows that for every $t \geq 0$,

$$|\int_0^t b(\mathbf{z}^{(n)}(s))ds - \int_0^t b(\mathbf{z}(s))ds| \to 0 \text{ as } n \to +\infty.$$

As well, the so-called local majoration condition figured by the function $\bar F_t^\alpha$ in Theorem 5.2.12 of Puhalskii (2001) is satisfied with $\bar F_t^\alpha = \sup_{|z|\leq\alpha} |b(z)|t$ in our context.

Non-Explosion condition:

The property that π_z is upper-compact means that for every $a \in (0,1]$, the set $K_a := \{\mathbf{z}, \pi_z(\mathbf{z}) \geq a\}$ is a compact set (for the topology of uniform convergence on compact sets). For this, we wish to apply the Ascoli Theorem. We first show the boundedness property for the paths of K_a. From the definition of π_z, we observe that for any $\mathbf{z}$ of K_a, there exists a control $\varphi \in \mathbb{H}$ with a uniform bound on $\|\dot\varphi\|_2$:

$$\int_0^\infty |\dot\varphi(s)|^2 ds \leq -2\log a + 1 \tag{3.2}$$

and such that each $\mathbf{z}$ is a φ controlled trajectory: $\mathbf{z} = \mathbf{z}_\varphi$. If we denote $\mathcal{E}(x,y) = U(x) + \frac{|y|^2}{2}$, one checks that for every $p > 0$,

$$\begin{aligned}\frac{d}{dt}(\mathcal{E}^p(\mathbf{z}(t)) &= p\mathcal{E}(\mathbf{z}(t))^{p-1}\left(|\mathbf{y}(t)|^2 + \langle\nabla U(\mathbf{x}(t)), \dot\varphi(t)\rangle\right)\\ &\leq C\left(\mathcal{E}(\mathbf{z}(t))^p + \mathcal{E}(\mathbf{z}(t))^{2p-2}|\nabla U(x)|^2 + |\dot\varphi(t)|^2\right)\end{aligned}$$

Under $(\mathbf{H_{Q+}})$ and $(\mathbf{H_{Q-}})$, we have respectively $|\nabla U(x)|^2 = O(U^{2-2\rho})$ and $|\nabla U(x)|^2 = O(U)$. Thus, applying the inequalities with $\bar p = \rho$ (resp. $\bar p = 1$) under $(\mathbf{H_{Q+}})$ (resp. $(\mathbf{H_{Q-}})$) yields:

$$\frac{d}{dt}\left\{\mathcal{E}^{\bar p}(\mathbf{z}(t)\right\} \leq C\left(\mathcal{E}^{\bar p}(\mathbf{z}(t)) + |\dot\varphi(t)|^2\right), \tag{3.3}$$

and the Gronwall lemma implies that

$$\forall t > 0 \quad \exists C_t > 0 \qquad \forall s \in [0,t] \qquad \mathcal{E}^{\bar p}(\mathbf{z}(s)) \leq C_t\left(\mathcal{E}^{\bar p}(z) + C\int_0^s |\dot\varphi(u)|^2 du\right) \leq C_{t,a,z}. \tag{3.4}$$

Finally, (3.2) combined with (3.4) and the fact that $\lim_{|z|\to+\infty} \mathcal{E}(z) = +\infty$ yields the boundedness result:

$$\sup_{\mathbf{z}\in K_a} \sup_{s\in[0,t]} |\mathbf{z}(s)| < +\infty. \tag{3.5}$$

The equicontinuity property can be derived from the continuity of b and the Cauchy-Schwarz inequality as follows: for every $t > 0$, and for every $u, v \in [0, t]$ with $u \le v$, for every $\mathbf{z} \in K_a$, we know that for a suitable constant $\tilde{C}_{t,a,z}$, the controlled trajectories of K_a are a priori bounded: $\|\mathbf{z}\|_{[0,t],\infty} \le \tilde{C}_{t,a,z}$. Thus, we can write

$$|\mathbf{z}(v) - \mathbf{z}(u)| \le \int_u^v |b(\mathbf{z}(s))|ds + \int_u^v |\dot{\varphi}(s)|ds \le \sup_{|z|\le \tilde{C}_{t,a,z}} |b(z)|(v-u) + \sqrt{1-2\log a}\sqrt{v-u}$$

and the compactness of K_a follows from Ascoli's Theorem.

Now we shall show the second part of Condition (NE), that is, for all $t \ge 0$, for all $a \in (0, 1]$, $\cup_{s\le t}\{\sup_{u\le s} |\mathbf{z}(u)| \mid \pi_{z,s}(\mathbf{z}) \ge a\}$ is a bounded subset of $\mathbb{R}$. This point follows easily from the controls established previously (see (3.4)).

Finally, other conditions needed in Theorem 5.2.12 of Puhalskii (2001) are trivially satisfied. This concludes the proof of the lemma. □

3.2 Exponential tightness (Proof ot i) of Theorem 2.1)

The exponential tightness of $(\nu_\varepsilon)_{\varepsilon\in(0,1]}$ is the purpose of the next proposition. Our approach consists in showing sufficiently sharper estimates for hitting time of the process $(Z_t^\varepsilon)_{t\ge0}$.

Proposition 3.1 *Assume* $(\mathbf{H_{Q+}})$ *or* $(\mathbf{H_{Q-}})$, *then there exists a compact set* B *of* $\mathbb{R}^{2d}$, *such that the first hitting time* τ_ε *of* B *defined as* $\tau_\varepsilon = \inf\{t > 0, Z_t^\varepsilon \in B\}$ *satisfies the following three properties:*

i) *For every compact set* K *of* $\mathbb{R}^{2d}$,

$$\limsup_{\varepsilon\to0} \sup_{z\in K} \mathbb{E}_z[(\tau_\varepsilon)^2] < \infty. \tag{3.6}$$

ii) *There exists* $\delta > 0$ *such that for every compact set* K *of* $\mathbb{R}^{2d}$,

$$\limsup_{\varepsilon\to0} \sup_{z\in K} \sup_{t\ge0} \mathbb{E}_z\left[|Z_{t\wedge\tau_\varepsilon}^\varepsilon|^{\frac{\delta}{\varepsilon^2}}\right]^{\varepsilon^2} < +\infty. \tag{3.7}$$

iii) *For evey compact set* K *of* $\mathbb{R}^{2d}$ *such that* $K \cap B = \emptyset$,

$$\liminf_{\varepsilon\to0} \inf_{z\in K} \mathbb{E}_z[\tau_\varepsilon] > 0. \tag{3.8}$$

Therefore, in view of Lemma 7 of Puhalskii (2003), the family of invariant distributions $(\nu^\varepsilon)_{\varepsilon\in(0,1]}$ *is exponentially tight.*

A fundamental step of the proof of Proposition 3.1 is the next lemma which shows some mean-reverting properties for the process (with some constants that do not depend on ε). Its technical proof is postponed in the appendix. Note that such lemma uses a Lyapunov function V which is rather not standard due to the kinetic form of the coupled process. Such Lyapunov estimates are key ingredient for the rest of the paper.

Lemma 3.2 *Assume* $(\mathbf{H_{Q+}})$ *or* $(\mathbf{H_{Q-}})$ *with* $a \in (1/2, 1]$ *and let* $V : \mathbb{R}^{2d} \to \mathbb{R}$ *be defined by*

$$V(x,y) = U(x) + \frac{|y|^2}{2} + m\left(\frac{|x|^2}{2} - \langle x, y\rangle\right),$$

with $m \in (0,1)$*. For* $p > 0$*,* $\delta > 0$ *and* $\varepsilon > 0$*, set*

$$\psi_\varepsilon(x,y) = \exp\left(\frac{\delta V^p(x,y)}{\varepsilon^2}\right).$$

Then, if $p \in (0,1)$ *under* $(\mathbf{H_{Q+}})$ *and* $p \in (1-a, a)$ *under* $(\mathbf{H_{Q-}})$ *and* δ *is a positive number, there exist* $\alpha, \beta, \alpha', \beta'$ *positive such that for all* $(x,y) \in \mathbb{R}^{2d}$ *and* $\varepsilon \in (0,1]$

$$\mathcal{A}^\varepsilon V^p(x,y) \le \beta - \alpha V^{\bar{p}}(x,y) \quad \text{and,} \tag{3.9}$$

$$\mathcal{A}^\varepsilon \psi_\varepsilon(x,y) \le \frac{\delta}{\varepsilon^2}\psi_\varepsilon(x,y)(\beta' - \alpha' V^{\bar{p}}(x,y)), \tag{3.10}$$

where

$$\bar{p} = \begin{cases} p & \text{under } (\mathbf{H_{Q+}}) \\ p + a - 1 & \text{under } (\mathbf{H_{Q-}}). \end{cases}$$

Proof of Proposition 3.1: For sake of simplicity, we omit the ε dependence and write (X_t, Y_t) instead of $(X_t^\varepsilon, Y_t^\varepsilon)$. We first show the first point i) and study the bound of Equation (3.6) . Let $p \in (0,1)$. By the Itô formula, we have

$$\frac{V^p(X_t,Y_t)}{1+t} = V^p(x,y) + \int_0^t -\frac{V^p(x,y)}{(1+s)^2} + \frac{\mathcal{A}^\varepsilon V^p(x,y)}{1+s} ds + \varepsilon M_t, \tag{3.11}$$

where (M_t) is the local martingale defined by

$$M_t \;=\; \int_0^t p\frac{V^{p-1}(X_s,Y_s)}{1+s}\langle \nabla U(X_s) + m(X_s - Y_s), dW_s\rangle \tag{3.12}$$

Since V is a positive function, we can deduce that

$$\frac{1}{\varepsilon^2}\int_0^t -\frac{\mathcal{A}^\varepsilon V^p(X_s,Y_s)}{1+s} ds - \frac{1}{2}\left\langle \frac{M_t}{\varepsilon}, \frac{M_t}{\varepsilon}\right\rangle \;\le\; \frac{1}{\varepsilon^2}V^p(x,y) + \frac{M_t}{\varepsilon} - \frac{1}{2}\left\langle \frac{M_t}{\varepsilon}, \frac{M_t}{\varepsilon}\right\rangle \tag{3.13}$$

Note that in the above expression, the martingale $(\frac{M_t}{\varepsilon})_{t\ge 0}$ has been compensated by its stochastic bracket in order to use further exponential martingale properties. The left hand side of (3.13) satisfies

$$\begin{aligned}
&\frac{1}{\varepsilon^2}\int_0^t -\frac{\mathcal{A}^\varepsilon V^p(X_s,Y_s)}{1+s} ds - \frac{1}{2}\langle \frac{M_t}{\varepsilon}, \frac{M_t}{\varepsilon}\rangle \\
&= \frac{1}{\varepsilon^2}\int_0^t \frac{1}{1+s}\left(-\mathcal{A}^\varepsilon V^p(X_s,Y_s) - \frac{p^2 V^{2p-2}(X_s,Y_s)}{1+s}|\nabla U(X_s) + m(X_s - Y_s)|^2\right) ds \\
&\ge \frac{1}{\varepsilon^2}\int_0^t \frac{H_{p,\varepsilon}(X_s,Y_s)}{1+s} ds.
\end{aligned}$$

with $H_{p,\varepsilon}(x,y) = -\mathcal{A}^\varepsilon V^p(x,y) - p^2 V^{2p-2}(x,y)|\nabla U(x) + m(x-y)|^2$. Then, a localization of (M_t) combined with Fatou's Lemma yields for every stopping time τ

$$\mathbb{E}\left[\exp\left(\frac{1}{\varepsilon^2}\int_0^{t\wedge\tau} \frac{H_{p,\varepsilon}(X_s,Y_s)}{1+s} ds\right)\right] \le \exp\left(\frac{1}{\varepsilon^2}V^p(x,y)\right)$$

The final step relies on the fact that there exists $p \in (0,1)$ and $M_1 > 0$ such that:

$$\forall (x,y) \in \bar{B}(0,M_1)^c \text{ and } \forall \varepsilon \in (0,1], \quad H_{p,\varepsilon}(x,y) \geq 2. \tag{3.14}$$

Let us prove the above inequality under condition $((\mathbf{H_{Q+}}))$ or $(\mathbf{H_{Q-}})$. First, using that V can be written

$$V(x,y) = U(x) + \frac{|y|^2}{2}(1-m^2) + \frac{m}{2}|x-y|^2 = U(x) + \frac{1}{2}|y-mx|^2 + m\frac{|x|^2}{2}(1-m)$$

and that $1-m^2$ and $1-m$ belong to $(0,1)$ (since $m \in (0,1)$), we deduce that there exists $C > 0$ such that

$$\forall (x,y) \in \mathbb{R}^{2d}, \quad |x|^2 + |y|^2 \leq CV(x,y). \tag{3.15}$$

Note that we also deduce that

$$\lim_{|(x,y)| \to +\infty} V(x,y) = +\infty. \tag{3.16}$$

Now, owing to the assumptions on ∇U, it follows that,

$$V^{2p-2}|\nabla U(x) + m(x-y)|^2 = \begin{cases} O(V^{2(p-\rho)}(x,y)) + O(V^{2p-1}(x,y)) & \text{under } ((\mathbf{H_{Q+}})) \\ O(V^{2p-1}(x,y)) & \text{under } ((\mathbf{H_{Q-}})). \end{cases}$$

From now on, assume that

$$\begin{cases} 0 < p < 2\rho \wedge 1 & \text{under } ((\mathbf{H_{Q+}})) \\ 1-a < p < a & \text{under } ((\mathbf{H_{Q-}})). \end{cases} \tag{3.17}$$

By Lemma 3.2, we then obtain that for every $(x,y) \in \mathbb{R}^{2d}$ and $\varepsilon \in (0,1]$

$$H_{p,\varepsilon}(x,y) \geq -\beta + \alpha V^{\bar{p}}(x,y) - O(V^{2p-1}).$$

where $\bar{p}$ is defined in Lemma 3.2. Under (3.17), one checks that $2p-1 < \bar{p}$ and then

$$\lim_{|(x,y)| \to +\infty} H_p(x,y) = +\infty$$

and (3.14) follows. Next, we consider (3.2) with τ being $\tau_\varepsilon = \inf\{t \geq 0, Z_t^\varepsilon \in \bar{B}(0,M_1)\}$ where M_1 is such that (3.14) holds, we have

$$\mathbb{E}\left[\exp\left(\frac{1}{\varepsilon^2}\int_0^{t\wedge\tau_\varepsilon} \frac{2}{1+s}ds\right)\right] \leq \mathbb{E}\left[\exp\left(\frac{1}{\varepsilon^2}\int_0^{t\wedge\tau_\varepsilon} \frac{H_{p,\varepsilon}(X_s,Y_s)}{1+s}\right)ds\right] \leq \exp\left(\frac{V^p(x,y)}{\varepsilon^2}\right)$$

and then computing the integral and using Fatous's Lemma, we get

$$\mathbb{E}_{(x,y)}\left[(1+\tau_\varepsilon)^{\frac{2}{\varepsilon^2}}\right] \leq \exp\left(\frac{1}{\varepsilon^2}V^p(x,y)\right).$$

Applying now Jensen's inequality to $x \to x^{\frac{1}{\varepsilon^2}}$, we obtain that and every $(x,y) \in \mathbb{R}^{2d}$, for every $\varepsilon \in (0,1]$

$$\mathbb{E}_{(x,y)}[(1+\tau_\varepsilon)^2] \leq \exp\left(V^p(x,y)\right).$$

The first statement follows using that V^p is locally bounded.

We study the second point ii) of Proposition 3.1. Let us first notice that thanks to (3.15), we have for every $p > 0$, for $|(x,y)|$ large enough,

$$\ln(|(x,y)|) \leq \frac{1}{2}\ln(CV(x,y)) \leq V^p(x,y). \tag{3.18}$$

Multiplying by δ/ε^2, this inequality suggests the computation of

$$\mathbb{E}\left[\exp\left(\frac{\delta}{\varepsilon^2}V^p(X_{t\wedge\tau}, Y_{t\wedge\tau})\right)\right],$$

for appropriate p and τ. Applying Itô formula to the function $\psi_\varepsilon(x,y) = \exp(\delta V^p(x,y)/\varepsilon^2)$, we get for all t

$$\psi_\varepsilon(X_t, Y_t) = \psi_\varepsilon(x,y) + \int_0^t \mathcal{A}\psi_\varepsilon(X_s, Y_s)ds + M_t,$$

where $(M_t)_{t\geq 0}$ is a local martingale that we do not need to make explicit. Denoting by $(T_n)_{n\geq 1}$ a sequence of stopping times such that $T_n \xrightarrow{n\to+\infty} +\infty$ a.s. and such that for every $n \geq 1$, $(M_{t\wedge T_n})$ is a martingale. We deduce from Fatou's Lemma that for every stopping time τ,

$$\mathbb{E}_{(x,y)}[\psi_\varepsilon(X_{t\wedge\tau}, Y_{t\wedge\tau})] \leq \psi_\varepsilon(x,y) + \liminf_{n\to+\infty} \mathbb{E}_{(x,y)}\left[\int_0^{t\wedge\tau\wedge T_n} \mathcal{A}\psi_\varepsilon(X_s, Y_s)ds\right] \tag{3.19}$$

Let us choose $p \in (0,1)$ such that inequality (3.10) of Lemma 3.2 holds. Since $V(x,y) \to +\infty$ as $|(x,y)| \to +\infty$ and since $\bar{p} > 0$, we deduce that

$$\beta' - \alpha' V^{\bar{p}}(x,y) \xrightarrow{|(x,y)|\to+\infty} -\infty.$$

As a consequence, for every positive δ, there exists $M_2 > 0$ such that

$$\forall \varepsilon \in (0,1], \quad \forall (x,y) \in \bar{B}(0, M_2)^c, \quad \mathcal{A}\psi_\varepsilon \leq 0.$$

By (3.19) applied with $\tau_\varepsilon = \inf\{t \geq 0, (X_t, Y_t) \in \bar{B}(0, M_2)\}$, we deduce that

$$\mathbb{E}_{(x,y)}[\psi_\varepsilon(X_{t\wedge\tau_\varepsilon}, Y_{t\wedge\tau_\varepsilon})] \leq \psi_\varepsilon(x,y).$$

Without loss of generality, we can also assume that M_2 is such that (3.18) is valid for every $(x,y) \in \bar{B}(0,M_2)^c$. It follows that for every $\varepsilon \in (0,1]$, for every $t \geq 0$, for every $(x,y) \in \bar{B}(0,M_2)^c$,

$$\left(\mathbb{E}_{(x,y)}\left[|X_{t\wedge\tau_\varepsilon}, Y_{t\wedge\tau_\varepsilon}|^{\frac{\delta}{\varepsilon^2}}\right]\right)^{\varepsilon^2} \leq e^{\delta V^p(x,y)}.$$

As a consequence, for every compact subset K of $\mathbb{R}^d$,

$$\limsup_{\varepsilon\to+\infty} \sup_{(x,y)\in K} \sup_{t\geq 0} \left(\mathbb{E}_{(x,y)}\left[|X_{t\wedge\tau_\varepsilon}, Y_{t\wedge\tau_\varepsilon}|^{\frac{\delta}{\varepsilon^2}}\right]\right)^{\varepsilon^2} < +\infty.$$

At last, we consider the third point iii) and note that we do not follow the roadmap of Puhalskii (2003) for this last bound. With the notations of the two previous parts of the proof, (3.6) and (3.7) hold with $\tau_\varepsilon := \inf\{t \geq 0, (X_t, Y_t) \in B\}$ for every compact set B such that $\bar{B}(0, M_1 \vee M_2) \subset B$. In this last part of the proof, we then set $B = \bar{B}(0, M)$ where $M \geq M_1 \vee M_2$. Second, remark that it is enough to show that the result holds with $\tau_\varepsilon \wedge 1$ instead of τ_ε. Now, let K be a compact set of $\mathbb{R}^{2d}$ such that $B \cap K = \emptyset$ and let $(\varepsilon_n, z_n)_{n\geq 1}$ be a sequence such that $\varepsilon_n \to 0$, such that $z_n \in K$ for every $n \geq 1$ and such that

$$\mathbb{E}_{z_n}[\tau_{\varepsilon_n} \wedge 1] \xrightarrow{n\to+\infty} \liminf_{\varepsilon\to 0} \inf_{z\in K} \mathbb{E}_z[\tau_\varepsilon \wedge 1].$$

Up to an extraction, we can assume that $(z_n)_{n\geq 1}$ is a convergent sequence. Let $z^\star$ denote its limit. Set $P^{\varepsilon,z} = \mathcal{L}(Z_t^{(\varepsilon),z})_{t\geq 0}$. Thanks to Lemma 3.1, $(P^{\varepsilon_n,z_n})_{n\geq 1}$ is exponentially tight and thus tight on $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^d)$. At the price of a second extraction, we can thus assume that $(Z^{(\varepsilon_n),z_n})_{n\geq 1}$

converges in distribution to $Z^{(\infty)}$. Furthermore, since $\varepsilon_n \to 0$, $Z^{(\infty)}$ is a.s. a solution of the o.d.e. $\dot{\mathbf{z}} = b(\mathbf{z})$ starting at $\tilde{z}$. The function b being locally Lipschitz continuous, uniqueness holds for the solutions of this o.d.e. and we can conclude that $(Z^{(\varepsilon_n),z_n})_{n\geq 1}$ converges in distribution to $\mathbf{z}(\tilde{z},.)$ (where $\mathbf{z}(\tilde{z},.)$ denotes the unique solution of $\dot{\mathbf{z}} = b(\mathbf{z})$ starting from $\tilde{z}$. The function $\mathbf{z}(\tilde{z},.)$ being deterministic, the convergence holds in fact in probability and at the price of a last extraction, we can assume without loss of generality that $(Z^{(\varepsilon_n),z_n})_{n\geq 1}$ converges a.s. to $\mathbf{z}(\tilde{z},.)$. In particular, setting $\delta := d(K,B)$ ($\delta > 0$), there exists $n_0 \in \mathbb{N}$ such that for every $n \geq n_0$,

$$\sup_{t\in[0,1]} |Z_t^{(\varepsilon_n),z_n} - \mathbf{z}(\tilde{z},t)| \leq \frac{\delta}{4} \quad \text{a.s.}$$

Setting now,

$$\tau_{\tilde{z},\frac{\delta}{2}} := \inf\{t \geq 0, d(\mathbf{z}(\tilde{z},t),B) \leq \frac{\delta}{2}\} \wedge 1,$$

we deduce that for every $n \geq n_0$, a.s.

$$\inf_{t\in[0,\tau_{\tilde{z},\frac{\delta}{2}}]} d(Z_t^{(\varepsilon_n),z_n}, B) \geq \frac{\delta}{4} \implies \tau_{\varepsilon_n}^{z_n} \geq \tau_{\tilde{z},\frac{\delta}{2}} \quad \text{a.s.}$$

Using Fatou's lemma, we can conclude that

$$\lim_{n\to+\infty} \mathbb{E}_{z_n}[\tau_{\varepsilon_n} \wedge 1] \geq \mathbb{E}_{z_n}[\liminf_{n\to+\infty} \tau_{\varepsilon_n} \wedge 1] \geq \tau_{\tilde{z},\frac{\delta}{2}}.$$

Finally, since $t \mapsto \mathbf{z}(\tilde{z},t)$ is a continuous function and since $d(K, \bar{B}(0, M + \frac{\delta}{2})) > 0$, $\tau_{\tilde{z},\frac{\delta}{2}}$ is clearly positive. The result follows and this last point ends the proof of Proposition 3.1. $\square$

Using some lower bound obtained on the Lyapunov function $\mathcal{E}(x,y) = U(x) + y^2/2$ in Cabot, Engler, and Gadat (2009a), one could improve the minoration of the mean hitting time (point iii) of the last Proposition) but such refinement will not be necessary in the sequel.

3.3 Hamilton-Jacobi equation (Proof of ii) of Theorem 2.1)

This point is a consequence of the finite time large deviation principle which holds for $(Z^\varepsilon)_{\varepsilon\geq 0}$ (Lemma 3.1) and of the exponential tightness of $(\nu_\varepsilon)_{\varepsilon\geq 0}$ (Proposition 3.1). This is the purpose of the next proposition which is an adaptation of Corollary 1 of Puhalskii (2003).

Proposition 3.2 *For every $\varepsilon > 0$, let $(P_t^\varepsilon(z,.))_{t\geq 0, z\in\mathbb{R}^{2d}}$ denote the semi-group associated to* (2.2) *and assume that* (2.2) *admits a unique invariant distribution denoted by ν_ε. Then, if the following assumptions hold:*
(i) $(\nu_\varepsilon)_{\varepsilon>0}$ is exponentially tight of order ε^2 on $\mathbb{R}^{2d}$.

(ii) For every $t \geq 0$, for every $z \in \mathbb{R}^{2d}$, there exists a function $I_t(z,.) : \mathbb{R}^{2d} \to \mathbb{R}$, such that for every $(z_\varepsilon)_{\varepsilon>0}$ such that $z_\varepsilon \to z$ as $\varepsilon \to 0$, $P_t^\varepsilon(z_\varepsilon,.)$ satisfy a LDP with speed ε^{-2} and rate function $I_t(z,.)$.
Then, $(\nu_\varepsilon)_{\varepsilon>0}$ admits a (LD)-convergent subsequence and for such subsequence $(\nu_{\varepsilon_k})_{k\geq 0}$ (with $\varepsilon_k \to 0$ as $k \to +\infty$), the associated rate function W satisfies for every $z_0 \in \mathbb{R}^{2d}$,

$$W(z_0) = \inf_{z\in\mathbb{R}^{2d}} (I_t(z,z_0) + W(z)). \tag{3.20}$$

With the terminology of (Puhalskii, 2003), (3.20) says that $\tilde{W}$ defined for every $\Gamma \in \mathcal{B}(\mathbb{R}^{2d})$ by $\tilde{W}(\Gamma) = \sup_{y\in\Gamma} \exp(-W(y))$ is an invariant deviability for $(\mathbb{P}_t^\varepsilon(z,.))_{t\geq 0, z\in\mathbb{R}^{2d}}$. In Corollary 1 of (Puhalskii, 2003), this result is stated with a uniqueness assumption on the invariant deviabilities. Following carefully the proof of this corollary yields the previous proposition when the uniqueness assumption is not fulfilled. We refer to Appendix A for details.

Owing to Proposition 3.1 and Lemma 3.1, Proposition 3.2 can be applied with $I_t(z,.)$ defined in (3.1). The rate function W is solution of (3.20) and equation (2.7) is satisfied. Thus, next result proves the assertion ii) of Theorem 2.1.

Proposition 3.3 *Assume* $(\mathbf{H_{Q+}})$ *or* $(\mathbf{H_{Q-}})$ *with* $a(1/2,1]$*, then any (good) rate function* W *associated to any (LD)-convergent subsequence* $(\varepsilon_n)_{n\geq 1}$ *satisfies for every* $t \geq 0$ *and for every* $z \in \mathbb{R}^{2d}$

$$W(z_0) = \inf_{\begin{cases} \varphi \in \mathbb{H} \\ \tilde{\mathbf{z}}_\varphi(0) = z_0 \end{cases}} \left[\frac{1}{2} \int_0^t |\dot{\varphi}|^2 + W(\tilde{\mathbf{z}}_\varphi(t)) \right].$$

Proof : We know that W satisfies (3.20) and thus for any $z_0 \in \mathbb{R}^{2d}$

$$W(z_0) = \inf_{v \in \mathbb{R}^{2d}} (I_t(v,z) + W(v)) = \inf_{v \in \mathbb{R}^{2d}} \left(\inf_{\begin{cases} \varphi \in \mathbb{H} \\ \mathbf{z}_\varphi(0) = v \text{ and } \mathbf{z}_\varphi(t) = z_0 \end{cases}} \int_0^t |\dot{\varphi}(s)|^2 ds + W(\mathbf{z}_\varphi(0)) \right)$$

Noting $g : [0,t] \to \mathbb{R}^{2d}$ defined by $g(s) = \mathbf{z}_\varphi(t-s)$ is a controlled trajectory associated to $-b$ and $-\varphi$, we deduce that for every $t \geq 0$

$$W(z_0) = \inf_{v \in \mathbb{R}^{2d}} \left(\inf_{\begin{cases} \varphi \in \mathbb{H} \\ \tilde{\mathbf{z}}_{-\varphi}(0) = z_0 \text{ and } \tilde{\mathbf{z}}_{-\varphi}(t) = v \end{cases}} \int_0^t |\dot{\varphi}(s)|^2 ds + W(\tilde{\mathbf{z}}_{-\varphi}(t)) \right).$$

The result follows from the change of variable $\tilde{\varphi} = -\varphi$. □

3.4 Hamilton Jacobi equation with infinite time

The aim of this part is to show that when there is a finite number of critical points, we can "replace t by $+\infty$" in (2.7). This proof is an adaptation of that of Theorem 4 of Biswas and Borkar (2009). The main novelty of our proof is the second step where we prove with arguments based on asymptotic pseudo-trajectories and Lyapunov functions that the optimal controlled trajectory is attracted by a critical point of the drift vector field.

Proof of (iii) of Theorem 2.1): *Step 1:* Here we show that we can build a function $\hat{\varphi} \in \mathbb{H}$ such that for all $z \in \mathbb{R}^{2d}$ the couple $(\tilde{\mathbf{z}}_{\hat{\varphi}}(z,t), \dot{\hat{\varphi}}(t))_{t\geq 0}$ on $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^{2d}) \times L^{2,loc}(\mathbb{R}_+, \mathbb{R}^d)$ satisfies $\forall t > 0$,

$$W(z) = \frac{1}{2} \int_0^t |\dot{\hat{\varphi}}(t)|^2(s) ds + W(\tilde{\mathbf{z}}_{\hat{\varphi}}(z,t)). \tag{3.21}$$

First, consider a fixed positive T and let $(\tilde{\mathbf{z}}^{(n)}_{\varphi^{(n)}}, \varphi^{(n)})_{n\geq 1}$ be a minimizing sequence of $\mathcal{C}([0,T], \mathbb{R}^{2d}) \times \mathbb{H}$ such that

$$\frac{1}{2} \int_0^T |\dot{\varphi}^{(n)}(s)|^2 ds + W(\tilde{\mathbf{z}}^{(n)}_{\varphi^{(n)}}(z,t)) \xrightarrow{n \to +\infty} \inf_{\begin{cases} \varphi \in \mathbb{H}, \\ \mathbf{z}_\varphi(0) = z \end{cases}} \frac{1}{2} \int_0^T |\dot{\varphi}(s)|^2 ds + W(\tilde{\mathbf{z}}_\varphi(z,T)).$$

Since W is non negative, it is clear that $(\int_0^T |\dot\varphi^{(n)}(s)|^2 ds)_{n\geq 0}$ is bounded. It follows on the one hand that $(\dot\varphi^{(n)})_{n\geq 1}$ is relatively compact on $L^2_w([0,T],\mathbb{R}^{2d})$ which denotes the set of square-integrable functions on $[0,T]$ endowed with the weak topology. Moreover, we will show that

$$\sup_{n\geq 1}\sup_{t\in[0,T]} |\mathbf{z}^{(n)}(t)| < +\infty.$$

Actually, in the subquadratic case, *i.e.* if D^2U is bounded, b is Lipschitz continuous and this point is classical. Now, consider the case where ∇U satisfies $|\nabla U| = O(U^{1-\rho})$ with $\rho \in (0,1)$. Since $\sup_{n\geq 1}\int_0^T |\dot\varphi^{(n)}(s)|^2 ds < +\infty$, inequality (3.4) implies that

$$\sup_{n\geq 1}\sup_{t\in[0,T]} \mathcal{E}^\rho\left(\tilde{\mathbf{z}}^{(n)}_{\varphi^{(n)}}(t)\right) < +\infty.$$

As a consequence, since $\mathcal{E}^\rho(z) \to +\infty$ as $|z| \to +\infty$, we conclude that

$$M := \sup_{n\geq 1}\sup_{t\in[0,T]} |\tilde{\mathbf{z}}^{(n)}_{\varphi^{(n)}}(t)| < +\infty.$$

Since b is a locally Lipschitz, b is then Lipschitz continuous on $B(0,M)$ and again a classical argument based on the Ascoli Theorem shows that $(\tilde{\mathbf{z}}^{(n)}_{\varphi^{(n)}})_{n\geq 1}$ is relatively compact on $\mathcal{C}([0,T],\mathbb{R}^{2d})$. It follows that $(\tilde{\mathbf{z}}^{(n)}_{\varphi^{(n)}}, \dot\varphi^{(n)})_{n\geq 1}$ is relatively compact on $\mathcal{C}([0,T],\mathbb{R}^{2d}) \times L^2_w([0,T],\mathbb{R}^{2d})$ and thus there exists a convergent subsequence to $(\hat{\mathbf{z}}^T, \dot{\hat\varphi}^T)$ which belongs to $\mathcal{C}([0,T],\mathbb{R}^{2d})\times L^2_w([0,T],\mathbb{R}^{2d})$. Using that b is a continuous function, one checks that $\hat{\mathbf{z}}^T(t) = \tilde{\mathbf{z}}_{\hat\varphi^T}(z,t)$, for every $t\in[0,T]$ and the couple $(\hat z^T, \dot{\hat\varphi}^T)$ satisfies (3.21) (for a fixed T). Furthermore, using the dynamic programming principle, for every $t\in[0,T]$, we have

$$W(\tilde{\mathbf{z}}_{\hat\varphi^T}(z,t)) = \int_t^T |\dot{\hat\varphi}^T(s)|^2 ds + W(\tilde{\mathbf{z}}_{\hat\varphi^T}(z,T)), \tag{3.22}$$

and it follows that (3.21) holds for every $t\in[0,T]$. As a consequence, we can build $(\tilde{\mathbf{z}}_{\hat\varphi}(z,.), \dot{\hat\varphi}) \in \mathcal{C}(\mathbb{R}_+,\mathbb{R}^{2d}) \times L^{2,loc}(\mathbb{R}_+,\mathbb{R}^{2d})$ (by concatenation) that satisfies (3.21) (for every $t\geq 0$).

Step 2: Dropping the initial condition z, we show that $(\tilde{\mathbf{z}}_{\hat\varphi}(t+.))_{t\geq 0}$ converges as $t\to+\infty$ to a stationary solution of $\dot{\mathbf{z}} = -b(\mathbf{z})$.
First, as in (3.22),

$$W(\tilde{\mathbf{z}}_{\hat\varphi}(t+s)) - W(\tilde{\mathbf{z}}_{\hat\varphi}(t)) = -\int_t^{t+s} |\dot{\hat\varphi}(u)|^2 du \tag{3.23}$$

and it follows that $(W(\tilde{\mathbf{z}}_{\hat\varphi}(t))_{t\geq 0}$ is a non-increasing and thus bounded function. Since W is a good rate function, for every $M>0$, $W^{-1}([0,M])$ is a compact subset of $\mathbb{R}^{2d}$. This means that $(\tilde{\mathbf{z}}_{\hat\varphi}(t))_{t\geq 0}$ is bounded. From (3.23), we deduce that $(\int_t^{t+T}|\dot{\hat\varphi}(s)|^2ds)_{t\geq 0}$ is also bounded. Thus, as in Step 1, owing to the previous statements and to the fact that b is locally Lipschitz continuous, we deduce from the Ascoli Theorem that $(\tilde{\mathbf{z}}_{\hat\varphi}(t+.))$ is relatively compact (for the topology of uniform convergence on compact sets).

We denote now by $\tilde{\mathbf{z}}^\infty_{\hat\varphi}(.)$ the limit of a convergent subsequence. Let us show that $(\tilde{\mathbf{z}}^\infty_{\hat\varphi}(t))_{t\geq 0}$ is a solution of $\dot{\mathbf{z}} = -b(\mathbf{z})$. First, since $(W(\tilde{\mathbf{z}}_{\hat\varphi}(t)))_{t\geq 0}$ is non-increasing (and non negative as a rate function), we again deduce from (3.23) that $\int_t^{t+T}|\dot{\hat\varphi}(u)|^2du \xrightarrow{t\to+\infty} 0$. As a consequence, using that for every $s\geq 0$, the map $\mathbf{z}\mapsto \mathbf{z}(s)-\mathbf{z}(0)+\int_0^s b(\mathbf{z}(u))du$ (from $\mathcal{C}(\mathbb{R}_+,\mathbb{R}^{2d})$ to $\mathbb{R}^{2d}$) is continuous and that

$$\forall T\geq 0, t\geq 0, s\in[0,T], \left|\tilde{\mathbf{z}}_{\hat\varphi}(t+s) - \tilde{\mathbf{z}}_{\hat\varphi}(t) + \int_0^s b(\tilde{\mathbf{z}}_{\hat\varphi}(t+u))du\right| \leq C_T\int_t^{t+T}|\dot{\hat\varphi}(u)|^2du,$$

we obtain that $(\tilde{\mathbf{z}}^\infty_{\hat\varphi}(t))_{t\geq 0}$ is a solution of $\dot{\mathbf{z}} = -b(\mathbf{z})$. It remains to show that $(\tilde{\mathbf{z}}^\infty_{\hat\varphi}(t))_{t\geq 0}$ is stationary, i.e. that every limit point of $(\tilde{\mathbf{z}}_{\hat\varphi}(t))_{t\geq 0}$ belongs to $\{z\in\mathbb{R}^{2d}, b(z)=0\}$. Denote by

$(\varphi_t(z))_{t,z}$ the flow associated with the o.d.e. $\dot{\mathbf{z}} = -b(\mathbf{z})$. Owing again to the fact that for every $T > 0$, $\int_t^{t+T} |\dot{\varphi}(u)|^2 du \xrightarrow{t\to+\infty} 0$, we can deduce that for every $T > 0$,

$$\sup_{s\in[0,T]} |\tilde{\mathbf{z}}_{\hat{\varphi}}(t+s) - \varphi_s(\tilde{\mathbf{z}}_{\hat{\varphi}}(t))| \xrightarrow{t\to+\infty} 0.$$

This means that $(\tilde{\mathbf{z}}_{\hat{\varphi}}(t))_{t\geq 0}$ is an *asymptotic pseudo-trajectory* for φ (see (Benaim, 1996)). As a consequence, by Proposition 5.3 and Theorem 5.7 of Benaim (1996), the set K of limit points of $(\tilde{\mathbf{z}}_{\hat{\varphi}}(t))_{t\geq 0}$ is a (compact) invariant set for φ such that $\varphi_{|K}$ has no proper attractor. This means that there is no strict invariant subset A of K such that for every $z \in K$, $d(\varphi_t(z), A) \xrightarrow{t\to+\infty} 0$. It follows that in order to conclude that K is included in $\{z, b(z) = 0\}$, it is now enough to show that $A = \{z, b(z) = 0\} \cap K$ is an attractor for $\varphi_{|\bar{K}}$. To this end for a positive ρ, we consider $L : \mathbb{R}^{2d} \mapsto \mathbb{R}$ defined by

$$L(z) = U(x) + (1-\rho)\frac{|y|^2}{2} - \rho\langle \nabla U(x), y\rangle \quad \text{with } z = (x, y).$$

If $\mathbf{z}$ is solution of $\dot{\mathbf{z}} = -b(\mathbf{z})$, we have :

$$\frac{d}{dt} L(\mathbf{z}(t)) = \mathbf{y}(t)^t \Big((1-\rho)I_d - \rho D^2 U(\mathbf{x}(t))\Big)\mathbf{y}(t) + \rho|\nabla U(\mathbf{x}(t))|^2.$$

Since K is a bounded invariant set and that D^2U is locally bounded, we can choose ρ small enough and $\alpha_\rho > 0$ such that for every $(\mathbf{z}(t))$ solution of $\dot{\mathbf{z}} = -b(\mathbf{z})$ with $\mathbf{z}(0) \in K$,

$$\frac{d}{dt} L(\mathbf{z}(t)) \geq \alpha_\rho |\mathbf{y}(t)|^2 + \rho|\nabla U(\mathbf{x}(t))|^2. \tag{3.24}$$

For every starting point $z \in K$, the function $t \mapsto L(\mathbf{z}(t))$ is then non-decreasing and thus convergent to $\ell_\infty \in \mathbb{R}$. Since $(\mathbf{z}(t))_{t\geq 0}$ is bounded, an argument similar to the one developed in Step 1 combined with Ascoli's theorem yields again that $(\mathbf{z}(t + .))$ is relatively compact. Then, if $(\mathbf{z}(t_n + .))_{n\geq 0}$ denotes a subsequence of $(\mathbf{z}(t + .))$, we can assume (at the price of a potential extraction) that $(\mathbf{z}(t_n + .))_{n\geq 0}$ converges to $\mathbf{z}^\infty(.)$. We have necessarily $L(\mathbf{z}^\infty(t)) = \ell_\infty$ for every $t \geq 0$ and it follows that

$$\frac{d}{dt} L(\mathbf{z}^\infty(t)) = 0.$$

By (3.24), we deduce that $\mathbf{y}^\infty(t) = \nabla U(\mathbf{x}^\infty(t)) = 0$. This means that $\mathbf{z}^\infty(.)$ is a stationary solution and that every limit point of $(\mathbf{z}(t))_{t\geq 0}$ is an equilibrium point of the o.d.e.
Thus, we can conclude that every limit point of $(\tilde{\mathbf{z}}_{\hat{\varphi}}(t))_{t\geq 0}$ belongs to $\{z, b(z) = 0\}$. Finally, since the set of limit points of $(\tilde{\mathbf{z}}_{\hat{\varphi}}(t))_{t\geq 0}$ is compact connected and $\{z, b(z) = 0\}$ is finite, it follows that $\tilde{\mathbf{z}}_{\hat{\varphi}}(t) \to z_i$ as $t \to +\infty$ where $z_i = (x_i, 0)$ with $x_i \in \{x, \nabla U(x) = 0\}$. Then, by (3.22), we will deduce the announced result if we prove that W is continuous at z_i. This is the purpose of the next step.

Step 3: We now use a continuity argument and take the limit when $t \to +\infty$ in (3.21). For every $x^\star \in \{x, \nabla U(x) = 0\}$, W is continuous at $z^\star = (x^\star, 0)$. Indeed, since $D^2U(x^\star)$ is invertible, we deduce from Lemma 4.1 below that the dynamical system is *locally controllable* around $z^\star$, i.e. that for every $T > 0$, for every $\varepsilon > 0$, there exists $\eta > 0$ such that for every $z \in B(z^\star, \eta)$, $I_T(z, z^\star) \leq \varepsilon$ and $I_T(z^\star, z) \leq \varepsilon$. Now, owing to the definition of W, it follows that $W(z^\star) \leq W(z) + \varepsilon$ and that $W(z) \leq W(z^\star) + \varepsilon$. The continuity of W follows and it ends the proof of (iii) in Theorem 2.1 by taking the limit $t \to +\infty$ in (3.21). □

4 Freidlin and Wentzell theory

In this section, we are interested by some sharp estimations of the behaviour of the invariant distribution of the averaged diffusion using the roadmap of Freidlin and Wentzell (1979). Our goal is twofold: first, we aim to obtain some uniqueness property for the rate function W defined in Theorem 2.1 and thus to derive a large deviation principle for $(\nu_\varepsilon)_\varepsilon$. Second, we want to obtain a more explicit formumlation of W in order to characterize at least in some particular cases, the limit behaviour of $(\nu_\varepsilon)_\varepsilon$ for some non-convex potential U . In the rest of the paper, we will assume that the potential U satisfies Assumption $(\mathbf{D})$ defined in Section 2.4.2. We recall that under the assumptions on U, the set of critical points of U is in fact finite. Thus, we set:

$$\{x \in \mathbb{R}^d, \nabla U(x) = 0\} = \{x_1^\star, \ldots, x_\ell^\star\}.$$

More precisely, in this section, we start by classifying the critical points that is we link the critical points of the vector field b with the one of U and we determine their nature (stable or unstable). Next, with respect to this critical points, we construct the so called skeleton Markov chain associated to the process $(X_t^\varepsilon, Y_t^\varepsilon)$. With all these ingredients, we finally derive the large deviation principle for $(\nu_\varepsilon)_\varepsilon$.

4.1 Classification of critical points.

We first need to classify the equilibrium points of the dynamical system $\dot{\mathbf{z}} = b(z)$. We recall that

$$\{z \in \mathbb{R}^{2d}, b(z) = 0\} = \{z_1^\star, \ldots, z_\ell^\star\}$$

where for every $i \in \{1, \ldots, \ell\}$, $z_i^\star = (x_i^\star, 0)$. The following proposition characterizes the nature of $z_i^\star$ with respect to the one of $x_i^\star$.

Proposition 4.1 *Assume that $D^2U(x_i^\star)$ is invertible for every $i \in \{1, \ldots, \ell\}$. Then, if $x_i^\star$ is a minimum of U, then $z_i^\star$ is a stable equilibrium of the deterministic dynamical system. Otherwise, $z_i^\star$ is an unstable equilibrium.*

Proof : We denote by I the minima of U and by J the other critical points. Let us compute the differential of the vector field b: for each $i \in \{1, \ldots \ell\}$

$$Db(z_i^\star) = \begin{pmatrix} 0 & -I_d \\ D^2U(x_i^\star) & -I_d \end{pmatrix}.$$

Now, simple linear algebra yields the characterization of the spectrum of the linearized vector field near each equilibrium $z_i^\star$:

$$\begin{aligned} Sp(Db(z_i^\star)) &= \left\{\lambda | -\lambda(\lambda+1) \in Sp(D^2U(x_i^\star))\right\} \\ &= \left\{-1/2 \pm \sqrt{1/4-\mu} \qquad \mu \in Sp(D^2U(x_i^\star))\right\}, \end{aligned}$$

where $\sqrt{1/4-\mu}$ denotes $i\sqrt{|1/4-\mu|}$ if $1/4-\mu \leq 0$. Since $D^2U(x_i^\star)$ is a positive definite matrix, it follows that when $x_i^\star$ is a local minima of U, $\mu \in Sp(D^2U(x_i^\star))$ is then positive and

$$\forall i \in I \qquad \Re\left(Sp(Db(z_i^\star))\right) \subset (-1, 0).$$

Hence, $z_i^\star$ is a stable equilibrium when $x_i^\star$ is a local minimum of U.
When $x_i^\star$ is another equilibrium point, $D^2U(x_i^\star)$ has some negative eigenvalues μ. Then, $Db(z_i^\star)$ has some positive eigenvalues (since $\sqrt{1/4-\mu} < 1/2$ in this case) and $z_i^\star$ is thus an unstable equilibrium of the deterministic dynamical system. This ends the proof of the proposition. □

4.2 Skeleton representation

The Freidlin and Wentzell (1979) description of the invariant measure ν_ε of the continuous time Markov strongly depends on its representation using the invariant measure of a specific skeleton Markov chain. This formula, due to Khas'minskii (see Has'minskii (1980), chapter 4) in the uniform elliptic case, will remain true in our framework even if the original process is hypo-elliptic and defined on a non compact manifold. This is the purpose of Proposition 4.2 below but before stating it, we first need to define the skeleton Markov chain associated to our process.

Let ρ_0 be the half of the minimum distance between two critical points:

$$\rho_0 = \frac{1}{2} \min_{i \neq j} d(z_i^\star, z_j^\star). \tag{4.1}$$

Now, let $0 < \rho_1 < \rho_0$ and set $g_i = B(z_i^\star, \rho_1)$. Each boundary ∂g_i is smooth as well as the one of the set g defined as

$$g = \cup_i g_i. \tag{4.2}$$

Note that by construction, $g_i \cap g_j = \emptyset$ if $i \neq j$. Finally, we denote by Γ the complementary set of the ρ_0-neighbourhood of the set of the critical points $z_i^\star$:

$$\Gamma = \left(\mathbb{R}^d \times \mathbb{R}^d\right) \setminus \cup_i B(z_i^\star, \rho_0). \tag{4.3}$$

We provide in Figure 1 a short summary of the construction of the sets $(g_i)_{i\in}, g, \Gamma$ as well as the positions of the critical points $z_i^\star$. Moreover, Figure 1 provides an example of the trajectory $(Z_t^{z,\varepsilon})_{t\geq 0}$ (K will be defined in the sequel).

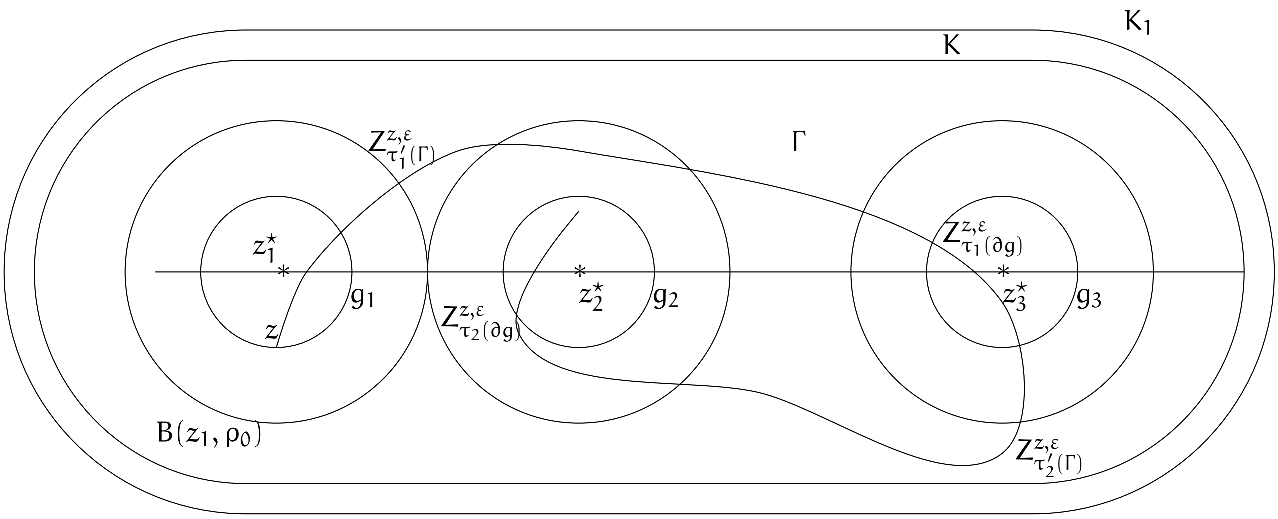

Figure 1: Graphical representation of the neighbourhood g_i, the process $(Z_t^{z,\varepsilon})_{t\geq 0}$, the skeleton chain and the compact sets K and K_1.

Now, we consider any initialisation on the boundary of the neighbourhoods of critical points $z \in \partial g$ (in our figure, $Z_0^{z,\varepsilon} = z \in \partial g_1$), and we define $(\tilde{Z}_n)_{n\in\mathbb{N}}$ the skeleton Markov chain which lives in ∂g through the classical construction of hitting and exit times of the neighbourhoods defined above. First, we set $\tau_0(\partial g) = 0$ and we also define

$$\tau_1'(\Gamma) = \inf\{t \geq 0,\ Z_t^{z,\varepsilon} \in \Gamma\}, \quad \tau_1(\partial g) = \inf\{t > \tau_1',\ Z_t^{z,\varepsilon} \in \partial g\}. \tag{4.4}$$

We then follow the natural recursion

$$\tau'_n(\Gamma) = \inf\{t > \tau_{n-1},\ Z_t^{z,\varepsilon} \in \Gamma\}, \quad \tau_n(\partial g) = \inf\{t > \tau'_n,\ Z_t^{z,\varepsilon} \in \partial g\}.$$

Owing for instance to Proposition 4.2 below, one checks that for every $n \geq 0$, $\tau_n(\partial g) < +\infty$ a.s. The skeleton is then defined for every $n \in \mathbb{N}$ by, $\tilde{Z}_n = Z^{z,\varepsilon}_{\tau_n(\partial g)}$. Note that $(\tilde{Z}_n)_{n\geq 0}$ belongs to ∂g and that $(\tilde{Z}_n)_{n\geq 0}$ is a Markov chain (this is actually a consequence of the strong Markov property). The set ∂g being compact, existence holds for the invariant distribution $(\tilde{Z}_n)_{n\in\mathbb{N}}$. We denote such distribution by $\tilde{\mu}_\varepsilon^{\partial g}$. The next proposition states that ν_ε may be related to $\tilde{\mu}_\varepsilon^{\partial g}$.

Proposition 4.2 *For any borelian set $A \in \mathcal{B}(\mathbb{R}^d \times \mathbb{R}^d)$ and for any $\rho_1 \in (0, \rho_0)$ the measure*

$$\mu_\varepsilon^{\partial g}(A) = \int_{\partial g} \tilde{\mu}_\varepsilon^{\partial g}(dz)\mathbb{E}_z \int_0^{\tau_1(\partial g)} \mathbf{1}_{Z_s^{z,\varepsilon} \in A} ds \tag{4.5}$$

is invariant for the process $(Z_t^\varepsilon)_{t\geq 0}$. Furthermore,

$$\forall \varepsilon > 0 \qquad \sup_{z\in\partial g} \mathbb{E}_z^\varepsilon[\tau_1(\partial g)] < \infty.$$

As a consequence, $\mu_\varepsilon^{\partial g}$ is a finite measure proportional to ν_ε.

Proof : The fact that $\mu_\varepsilon^{\partial g}$ is invariant for $(Z_t^\varepsilon)_{t\geq 0}$ is standard and relies on the strong Markov property of the process (see e.g. in Has'minskii (1960)).
Thus, we only show that $\sup_{z\in\partial g} \mathbb{E}_z^\varepsilon[\tau_1(\partial g)] < \infty$. Owing to Proposition 3.1, we first check that one can find a compact set K such that $g \subseteq K$ an such that for every compact set K_1 such that $K \subseteq K_1$, the first hitting time $\tau(K)$ of K satisfies

$$\sup_{z\in\partial K_1} \mathbb{E}_z^\varepsilon[\tau(K)] < +\infty. \tag{4.6}$$

Then, the idea of the proof is to extend to our hypo-elliptic context the proofs of Lemma 4.1 and 4.3 of Has'minskii (1980) given under some elliptic assumptions. Let $z \in \partial g$ and set $\tilde{\tau}_0 = \inf\{t \geq 0, \quad Z_t^{z,\varepsilon} \in g^c\}$,

$$\tilde{\tau}'_1 = \inf\{t > \tilde{\tau}_0,\ Z_t^{z,\varepsilon} \in \partial g \cup \partial K_1\}, \quad \tilde{\tau}_1 = \inf\{t > \tilde{\tau}'_1,\ Z_t^{z,\varepsilon} \in \partial K\},$$

and recursively for every $n \geq 2$,

$$\tilde{\tau}'_n = \inf\{t > \tilde{\tau}_{n-1},\ Z_t^{z,\varepsilon} \in \partial g \cup \partial K_1\} \quad \tilde{\tau}_n(\partial g) = \inf\{t > \tilde{\tau}'_n,\ Z_t^{z,\varepsilon} \in K\}.$$

By construction, we have a.s.:

$$\tau_1(\partial g) \leq \inf\{\tilde{\tau}'_k,\ Z^{z,\varepsilon}_{\tilde{\tau}'_k} \in \partial g\}.$$

Then, by the strong Markov property and (4.6), it follows from a careful adaptation of the proofs of Lemma 4.1 and 4.3 of Has'minskii (1980) that $\sup_{z\in\partial g} \mathbb{E}_z^\varepsilon[\tau_1(\partial g)] < \infty$ if the two following points hold for every $\varepsilon > 0$:

- $\sup_{z\in K} \mathbb{E}_z^\varepsilon[\tau(\partial K_1)] < +\infty$.
- $\sup_{z\in K\setminus g} p_\varepsilon(z) < 1$ where $p_\varepsilon(z) := \mathbb{P}(Z^{z,\varepsilon}_{\tau(\partial g\cup\partial K_1)} \in \partial K_1)$.

As concerns the first point, it follows from Remark 5.2 of Stroock and Varadhan (1972) that it is enough to check that there exists $T > 0$, a control $(\varphi(t))_{t\in[0,T]}$ such that

$$\forall z \in K, \quad \inf\{t \geq 0, \mathbf{z}_\varphi(t,z) \in K_1^c\} \leq T. \tag{4.7}$$

Indeed, in this case, using the support theorem of Stroock and Varadhan (1970), we obtain that $\sup_{z\in K} \mathbb{P}(\tau(\partial K_1) \leq T) < 1$ and the first point follows from the strong Markov property (see Remark 5.2 of Stroock and Varadhan (1972) for details). Now, we build $(\varphi(t))_{t\geq 0}$ as follows. Let us consider the system:

$$\begin{cases} \dot{\mathbf{x}} = I_d \\ \dot{\mathbf{y}} = \nabla U(\mathbf{x}) - \mathbf{y} \end{cases}.$$

Setting $\dot{\varphi} = \mathbf{y} + I_d$, we obtain a controlled trajectory $\mathbf{z}_\varphi(z,.)$ and it is clear from its design that for every $M > 0$, there exists $T > 0$, such that for every $z \in K$, $|\mathbf{x}_\varphi(T)| > M$. The first point easily follows.

It is well-known (see for instance Stroock and Varadhan (1972)) that for every $\varepsilon > 0$, p_ε is a solution of

$$\mathcal{A}^\varepsilon p_\varepsilon = 0 \qquad \text{with} \qquad p_\varepsilon|_{\partial g} = 0 \qquad \text{and} \qquad p_\varepsilon|_{\partial K_1} = 1. \tag{4.8}$$

Thus, since $\sup_{z\in K_1\backslash g} \mathbb{E}[\tau(\delta g \cup \delta K_1)] < +\infty$, since h defined by $h(x) = 1$ on ∂K_1 and $h(x) = 0$ on ∂g is obviously continuous on $\partial g \cup \partial K_1$, we can apply Theorem 9.1 of Stroock and Varadhan (1972) with $k = f = 0$ to obtain that $z \mapsto p_\varepsilon(z)$ is a continuous map. Furthermore, for every $z \in K\backslash g$, we can build a controlled trajectory starting at any $z \in \partial K$ which hit ∂g before ∂K_1. Taking for instance $\dot{\varphi} = 0$, we check that $(\mathcal{E}(\mathbf{x}_0(t), \mathbf{y}_0(t)))_{t\geq 0}$ is non-increasing (with $\mathcal{E}(x,y) = U(x) + |y|^2/2$) and that the accumulation points of $(\mathbf{x}_0(t), \mathbf{y}_0(t))$ lie in $\{z, b(z) = 0\}$. Thus, taking K_1 large enough in order that $\sup_{(x,y)\in K} \mathcal{E}(x,y) < \inf_{(x,y)\in K_1} \mathcal{E}(x,y)$, leads to an available control for every $z \in K$. Finally, using again the support theorem of Stroock and Varadhan (1970) implies that for each $z \in \partial K$, $p_\varepsilon(z) < 1$. The second point then follows from the continuity of $z \mapsto p_\varepsilon(z)$.

□

Remark 4.1 *We could also have used some uniqueness argument of viscosity solutions to obtain the continuity of $z \mapsto p_\varepsilon(z)$ with the maximum principle on $\mathcal{A}^\varepsilon$ (as it is already used by Stroock and Varadhan (1972)). One may refer to Barles (1994) for further details.*

4.3 Transitions of the skeleton Markov chain

This paragraph is dedicated to the description of estimations obtained through the Freidlin and Wentzell theory for the Markov skeleton chain defined above. These estimations and Proposition 4.2 are then used to obtain the asymptotic behaviour of ν_ε. In view of Theorem 2.1, we know that there exists a subsequence $(\varepsilon_n)_{n\in\mathbb{N}}$ such that ν_{ε_n} satisfies a large deviation principle of rate ε_n^2 with good rate function W. We then consider in the sequel this extracted subsequence but keep the notation ε. Hence, when we write $\varepsilon \to 0$, the reader may consider that we refer to $\varepsilon_n \to 0$ as $n \to +\infty$ with our appropriate subsequence along with the large deviation principle holds. In the same asymptotic setting, ε small enough will correspond to n large enough.

4.3.1 Controllability and exit times estimates

In order to obtain some estimates related to the transition of skeleton Markov chain, the first step is to control the exit times of some balls $B(z_i^\star, \delta)$ where $z_i^\star$ denotes a critical point of $\dot{\mathbf{z}} = b(\mathbf{z})$ (similarly to Section 1, Chapter 6 of Freidlin and Wentzell (1979)). In our hypo-elliptic framework, such controls of the exit times are strongly based on the controllability around the equilibrium points. We have the following property:

Lemma 4.1 *Let* $T > 0$. *Assume that for every* $i \in \{1, \dots, \ell\}$, $D^2U(x_i^\star)$ *is invertible. Then, for every* $\delta > 0$, *there exists* $\rho(\delta) > 0$ *small enough such that:*

$$\forall i \in \{1 \dots \ell\}, \quad \forall (a, b) \in B(z_i^\star, \rho(\delta)) \quad \exists \varphi \in \mathbb{H}, \text{ such that } z_\varphi(a, T) = b$$

and such that,

$$\int_0^T |\dot\varphi(s)|^2 ds \le \delta.$$

Proof : Setting

$$A = \begin{pmatrix} 0 & -I_d \\ D^2U(x_i^\star) & -I_d \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} I_d & 0 \\ 0 & 0 \end{pmatrix},$$

the linearized system (at $z_i^\star$) associated with the controlled system

$$\dot{\mathbf{z}} = b(\mathbf{z}) + \begin{pmatrix} \dot\varphi \\ 0 \end{pmatrix} \tag{4.9}$$

can be written $\dot{\mathbf{z}} = A\mathbf{z} + Bu$ where $u = (\dot\varphi, \psi)^\star$ with $\psi \in \mathbb{H}(\mathbb{R}^d)$. Using that $D^2U(x_i^\star)$ is invertible, one easily checks that $\mathrm{Span}(Bu, ABu, u \in \mathbb{R}^{2d}) = \mathbb{R}^{2d}$. As a consequence, the Kalman condition (see e.g. (Coron, 2007)) is satisfied and it follows from Theorems 1.16 and 3.8 of (Coron, 2007) that the system (4.9) is *locally exactly controllable* at $z_i^\star$, i.e. that the statement of the lemma is true. □

We are now able to obtain the following estimation:

Lemma 4.2 *Assume the hypothesis of Lemma 3.1 and that for every* $i \in \{1, \dots, \ell\}$, $D^2U(x_i^\star)$ *is invertible. Then, for every* $\gamma > 0$, *there exists* $\delta > 0$ *and* ε_0 *small enough such that if we define* $G = B(z_i^\star, \delta)$, *the first exit time of* G *denoted* τ_{G^c} *satisfies*

$$\forall \varepsilon \in (0, \varepsilon_0], \ \sup_{z \in G} \mathbb{E}_z^\varepsilon \tau_{G^c} < e^{\gamma \varepsilon^{-2}}.$$

Proof : Let $i \in \{1 \dots \ell\}$ and fix any $\gamma > 0$. By Lemma 4.1, one can find $\rho > 0$ such that

$$\forall (a, b) \in B(z_i^\star, 2\rho) \qquad \exists \varphi \in \mathbb{H}, \quad \text{such that} \qquad \mathbf{z}_\varphi(a, T) = b, \quad \text{and} \quad \int_0^T |\dot\varphi(s)|^2 ds \le \frac{\gamma}{2}.$$

Now, we set $\delta = \rho/2$, $G = B(z_i^\star, \delta)$ and we fix $a = z$ and take b such that $|z_i^\star - b| = \rho$ so that for every $z \in B(z_i^\star, \delta)$, $|z - b| \le 3\delta < \rho$. Thus, for every $z \in B(z_i^\star, \delta)$, we can find $\varphi_z \in \mathbb{H}$ such that $\mathbf{z}_{\varphi_z}(z, T) = b$ and $\int_0^T |\dot\varphi_z(s)|^2 ds \le \gamma/2$.

Now, it is possible to follow the proof of Lemma 1.7, chapter 6 of Freidlin and Wentzell (1979): remark that

$$\mathbb{P}_z^\varepsilon[\tau_{G^c} \le T] \ge \mathbb{P}\left[\|Z_.^{z,\varepsilon} - \mathbf{z}_{\varphi_z}(z, .)\|_{\infty,[0;T]} \le \delta\right]. \tag{4.10}$$

Second, using that G is a compact set,there exists a convergent sequence (z_k) of G and a sequence (ε_k) such that $\varepsilon_k \to 0$ and such that,

$$\liminf_{\varepsilon \to 0} \inf_{z \in G} \varepsilon^2 \ln \mathbb{P}_z^\varepsilon[\tau_{G^c} \le T] = \lim_{k \to +\infty} \ln(\mathbb{P}_{z_k}^{\varepsilon_k}[\tau_{G^c} \le T]).$$

Now, owing to Lemma 3.1 and to (4.10), we deduce that

$$\liminf_{\varepsilon \to 0} \inf_{z \in G} \varepsilon^2 \ln \mathbb{P}_z^\varepsilon[\tau_{G^c} \le T] \ge -\int_0^T |\dot\varphi_{z_\infty}(s)|^2 ds \ge -\frac{\gamma}{2}$$

where $z_\infty := \lim_{k \to +\infty} z_k$. As a consequence, there exists $\varepsilon_0 > 0$ such that for every $\varepsilon \in (0, \varepsilon_0]$, for every $z \in G$,

$$\mathbb{P}_z^\varepsilon[\tau_{G^c} \le T] \ge e^{-\gamma \varepsilon^{-2}}$$

and the strong Markov property implies that

$$\forall n \in \mathbb{N} \qquad \mathbb{P}[\tau_{G^c} > nT] \leq [1 - e^{-\gamma\varepsilon^{-2}}]^n.$$

Thus, applying the inequality with $T = 1$, we obtain for every $z \in G$ and $\varepsilon \in (0, \varepsilon_0]$

$$\mathbb{E}^{\varepsilon}\tau_{G^c} \leq \sum_{n=0}^{\infty}[1 - e^{-\gamma\varepsilon^{-2}}]^n \leq e^{-\gamma\varepsilon^{-2}}.$$

□

Following the same kind of argument using again the key Lemma 4.1 and the finite time large deviation principle, we also obtain that Lemma 1.8, chapter 6 of Freidlin and Wentzell (1979) still holds:

Lemma 4.3 *For any $i \in \{1 \dots \ell\}$ and $z_i^\star$ an equilibrium of (2.3), for any $\gamma > 0$, there exists $\delta > 0$ and ε small enough such that if we define $G = B(z_i^\star, \delta)$, the first exit time of G denoted τ_{G^c} satisfies*

$$\inf_{z \in G} \mathbb{E}_z \int_0^{\tau_{G^c}} \chi_G(Z_t^\varepsilon)dt > e^{-\gamma\varepsilon^{-2}}.$$

4.3.2 Transitions of the Markov chain skeleton

By Proposition 4.2, the idea is now to deduce the behaviour of ν_ε from the control of the transitions of the skeleton chain $(\tilde{Z}_n)_{n\in\mathbb{N}}$. We recall that for any $(\xi_1, \xi_2) \in (\mathbb{R}^d \times \mathbb{R}^d)^2$, $I_t(\xi_1, \xi_2)$ is the L^2 minimal cost to go from ξ_1 to ξ_2 in a finite time T:

$$I_T(\xi_1, \xi_2) = \inf_{\begin{cases} \varphi \in \mathbb{H} \\ \mathbf{z}_\varphi(\xi_1, T) = \xi_2 \end{cases}} \frac{1}{2}\int_0^T |\dot{\varphi}(s)|^2 ds,$$

and $I(\xi_1, \xi_2)$ is defined by:

$$I(\xi_1, \xi_2) = \inf_{T \geq 0} I_T(\xi_1, \xi_2).$$

In the sequel, we will also need to introduce $\tilde{I}(z_i^\star, z_j^\star)$ defined for every $\forall (i, j) \in \{1 \dots \ell\}^2$ by:

$$\tilde{I}(z_i^\star, z_j^\star) = \inf_{T>0} \inf \left\{ \frac{1}{2}\int_0^T |\dot{\varphi}(s)|^2 ds, \varphi \in \mathbb{H}, \mathbf{z}_\varphi(z_i^\star, T) = z_j^\star, \forall s \in [0, T],\ \mathbf{z}_\varphi(z_i^\star, s) \notin \cup_{k\neq i,j} g_k \right\}.$$

This represents the minimum cost to join $z_j^\star$ from $z_i^\star$ avoiding other equilibriums of (2.3). With our particular dynamical system, one can show that it is always possible to find a controlled trajectory starting at $z_i^\star$ and ending at $z_j^\star$ that avoids the other equilibria neigbourhoods $\cup_{k\neq(i,j)} g_k$, and then deduce the following proposition.

Proposition 4.3 *For every $(i, j) \in \{1 \dots \ell\}^2$, $\tilde{I}(z_i^\star, z_j^\star) < +\infty$.*

<u>*Proof :*</u> We first assume that $d > 1$. In this case, for every fixed $t_0 > 0$, for any ρ_1-neighbourhood g_k of $z_k^\star$, one can find a smooth trajectory $(\mathbf{x}_0(t))_{t\geq 0}$ satisfying $\mathbf{x}_0(0) = x_i^\star$, $\mathbf{x}_0(t_0) = x_j^\star$ and

$$\forall s \in [0; t_0] \qquad \inf_{k\neq i,j} |\mathbf{x}_0(s) - x_k^\star| > \rho_1.$$

Then, denote by $(\mathbf{y}_0(t))_{t\geq 0}$ a solution of $\dot{\mathbf{y}}_0(t) = \nabla U(\mathbf{x}_0(t)) - \mathbf{y}_0(t)$ with initial condition $\mathbf{y}_0(0) = 0$ and let $\varphi_0 \in \mathbb{H}$ satisfies $\dot{\varphi}_0(t) = \dot{\mathbf{x}}_0(t) + \mathbf{y}_0(t)$. We thus obtain a controlled trajectory $\mathbf{z}_{\varphi_0}(z_i^\star, .)$ which satisfies $\mathbf{z}_{\varphi_0}(z_i^\star, t) = (\mathbf{x}_0(t), \mathbf{y}_0(t))$ for every $t \in [0, t_0]$ and thus,

$$\mathbf{x}_\varphi(z_i^\star, t_0) = x_j^\star \quad \text{and} \quad \forall s \in [0; t_0] \qquad \mathbf{z}_\varphi(z_i^\star, s) \notin \cup_{k\neq i,j} g_k.$$

It remains now to join $(x_j^\star, 0)$ from $(x_j^\star, y_0(t_0))$ without hitting $\cup_{k\neq i,j} g_k$. Let $(\mathbf{x}_1(t), \mathbf{y}_1(t))_{t\geq t_0}$ be defined for every $t \geq t_0$ by $\mathbf{x}_1(t) = x_j^\star$ and $\mathbf{y}_1(t) = \mathbf{y}_0(t_0)e^{t_0-t}$ (so that $\mathbf{y}_1$ is a solution of $\dot{\mathbf{y}_1} = -\mathbf{y}_1$ with $\mathbf{y}_1(t_0) = \mathbf{y}_0(t_0)$). Once again, $(\mathbf{x}_1(t), \mathbf{y}_1(t))_{t\geq t_0}$ can be viewed as a controlled trajectory $\mathbf{z}_{\varphi_1}((x_j^\star, \mathbf{y}_0(t_0), .)$ by setting $\dot{\varphi}_1(t) = \mathbf{y}_1(t)$.

Furthermore, $\mathbf{z}_{\varphi_1}((x_j^\star, \mathbf{y}_0(t_0)), t) \xrightarrow{t\to+\infty} 0$. Thus, there exists T such that $\mathbf{z}_{\varphi_1}((x_j^\star, \mathbf{y}_0(t_0), T) \in g_j$. Hence, one can find a controlled trajectory starting from z_i and ending into any sufficiently small neighbourhood of z_j in a finite time and avoids the other ρ_1 neighbourhood of $\cup_{k\neq(i,j)} g_k$. It remains to use Lemma 4.1 to obtain a controlled trajectory starting at $\mathbf{z}_{\varphi_1}((x_j^\star, \mathbf{y}_0(t_0)), T)$ and ending at point $z_j^\star$ within a finite time. The global controlled trajectory initialized at $z_i^\star$ ends at $z_j^\star$ with a finite L^2 control cost. The result then follows when $d > 1$.

Consider now the case $d = 1$ and let $x_i^\star, x_j^\star$ be two critical points of U. Without loss of generality, one may suppose that $x_i^\star < x_j^\star$. From Assumption **(D)**, the number of critical points which belong to $[x_i^\star, x_j^\star]$ is finite (denoted by p):

$$x_i^\star < x_{i_1}^\star < \cdots < x_{i_p}^\star < x_j^\star.$$

Now, we consider a path which joins $x_i^\star$ to $x_j^\star$ parametrised as

$$\mathbf{x}_\alpha(t) = x_i^\star + \alpha(t)[x_j^\star - x_i^\star],$$

with $\alpha(0) = 0$ and $\alpha(T) = 1$ for T large enough which will be given later. Of course, $\mathbf{y}(t)$ is then defined as

$$\forall t \in [0; T] \qquad \mathbf{y}_\alpha(t) = \int_0^t e^{s-t} U'\left(x_i^\star + \alpha(s)[x_j^\star - x_i^\star]\right) ds. \tag{4.11}$$

For the sake of simplicity, we consider only non decreasing maps α. If $p = 0$, we know that $(\mathbf{x}_\alpha(t), \mathbf{y}_\alpha(t))_{t\in[0;1]}$ avoids $\cup_{k\neq(i,j)}(x_k^\star, 0)$ and then $\tilde{I}(z_i^\star, z_j^\star) < +\infty$ which proves the proposition. If $p > 0$, there exists $t_1, \ldots t_p$ such that $\mathbf{x}_\alpha(t_k) = x_{i_k}^\star$ and we shall prove that one can find α such that $\mathbf{y}_\alpha(t_k) \neq 0$. Since we consider only increasing paths, we first show that one can find a monotone α such that $\mathbf{y}_\alpha(t_1) \neq 0$. Let α any $\mathcal{C}^1$ increasing parametrisation defined on $[0; t_1]$. We know that U' does not vanish on $]x_i^\star, x_{i_1}^\star[$ and from equation (4.11), it is immediate to see that $\mathbf{y}_\alpha(t_1) \neq 0$ since x is monotone on $[0; t_1]$ and defines an homeomorphism $\psi : [0; t_1] \mapsto [x_i^\star, x_{i_1}^\star]$ and

$$\mathbf{y}_\alpha(t_1) = e^{-t_1} \int_{x_i}^{x_{t_1}} \frac{e^{\psi^{-1}(x)}}{\psi'(\psi^{-1}(x))} U'(x) dx.$$

Suppose without loss of generality that $\mathbf{y}_\alpha(t_1) < 0$, which means that $U' < 0$ on $]x_i^\star, x_{i_1}^\star[$. Since we know that $U"(x_{i_1}^\star) \neq 0$, one can find $\delta > 0$ small enough such that $U' > 0$ on $]x_{i_1}^\star; x_{i_1}^\star + \delta[$. Let $\xi_1 \in]x_{i_1}^\star; x_{i_1}^\star + \delta[$, we continue the parametrisation α from t_1 to $\tilde{t}_1$ such that $x(\tilde{t}_1) = \xi_1$ and α remains constant on $[\tilde{t}_1; \tilde{t}_1 + \delta t_1]$. Expanding the integral that defines $\mathbf{y}$ (see equation (4.11)) between $[0, t_1], [t_1, \tilde{t}_1]$ and $[\tilde{t}_1, \tilde{t}_1 + \delta t_1]$, simple computation yields

$$\begin{aligned} \mathbf{y}_\alpha(\tilde{t}_1 + \delta t_1) \quad &= \quad \mathbf{y}_\alpha(t_1)e^{t_1 - \tilde{t}_1 - \delta t_1} \\ &\quad + \int_{t_1}^{\tilde{t}_1} e^{s - \tilde{t}_1 + \delta t_1} U'(x_i^\star + \alpha(s)[x_j^\star - x_i^\star]) ds \\ &\quad + U'(\xi_1)[1 - e^{-\delta t_1}]. \end{aligned}$$

Hence, it is obvious to see that we can find a sufficiently large δt_1 such that $y_\alpha(\tilde{t}_1 + \delta t_1) > 0$ since $U'(\xi_1) > 0$. We continue the parametrisation α until $x_{i_2}^\star$ is reached at time t_2 and by construction, $\mathbf{y}_\alpha(t_2) > 0$. Now, one can repeat the same argument by induction to find α such that $\mathbf{y}_\alpha(t_j) \neq 0$ forall j such that $\mathbf{x}_\alpha(t_j) = x_{i_j}^\star$.

We now must end the trajectory so that $\mathbf{y}_\alpha(T) = 0$ and $\mathbf{x}_\alpha(T) = x_j^\star$. Without loss of generality, we can assume that $\mathbf{y}_\alpha(t_p+2) > 0$ and $U''(x_j^\star) \neq 0$: the sign of U' is changing in the neighbourhood of $x_j^\star$ although $\mathbf{x}_\alpha(t_p+2) = x_j^\star$. There exists ζ close enough of $x_j^\star$ such that $U'(\zeta) < 0$ and $(x_j^\star, \zeta)$ does not contain another critical point of U'. For any time $\Delta > 0$, we consider the parametrisation such that x is monotone on $[t_{p+2}, t_{p+2}+\nu]$ and reach ζ, then remains on ζ a time Δ and come back to $x_j^\star$ monotonically on $[t_{p+2}+\nu+\Delta, t_{p+2}+\Delta+2\nu]$. From (4.11), $(\Delta, \nu) \longmapsto \mathbf{y}_\alpha(t_{p+2}+\Delta+2\nu)$ is continuous and for ν small enough and $\Delta = 0$, we have $\mathbf{y}_\alpha(t_{p+2}+\Delta+2\nu)$ is closed to $\mathbf{y}_\alpha(t_{p+2}) < 0$. We then choose $\nu^\star$ sufficiently small so that

$$\mathbf{y}_\alpha(t_{p+2}+\Delta+2\nu^\star) < 0 \quad \text{when} \quad \Delta = 0.$$

Now for large Δ, $\mathbf{y}_\alpha(t_{p+2}+\Delta+2\nu^\star)$ is positive. Thus there exists $\Delta^\star$ so that

$$\mathbf{y}_\alpha(t_{p+2}+\Delta^\star+2\nu^\star) = 0.$$

The global trajectory is obtained by concatenation of each parts of the trajectories between $[t_j, \tilde{t}_j+\delta t_j] \cup [\tilde{t}_j+\delta t_j, t_{j+1}]$, and $[t_{p+2}, t_{p+2}+\Delta^\star+2\nu^\star]$, this ends the proof. □

It is now possible to compute the estimations of the invariant measure $\tilde{\mu}_\varepsilon^{\partial g}$ of the skeleton chain. The key estimation of the transition probability of $(\tilde{Z}_n)_{n\in\mathbb{N}}$ denoted $\tilde{\mathbb{P}}^\varepsilon$ is as follows.

Proposition 4.4 *For any $\gamma > 0$, there exists a sufficiently small ρ_0 and ρ_1 satisfying $0 < \rho_1 < \rho_0$ such that with the definition (4.2) and (4.4), we have for ε small enough*

$$\forall (i,j) \in \{1\ldots\ell\}^2 \quad \forall x \in \partial g_i \qquad 0 < e^{-\varepsilon^{-2}[\tilde{I}(z_i^\star, z_j^\star)+\gamma]} \leq \tilde{\mathbb{P}}^\varepsilon(x, \partial g_j) \leq e^{-\varepsilon^{-2}[\tilde{I}(z_i^\star, z_j^\star)-\gamma]}.$$

The proof is a simple adaptation of the proof of Lemma 2.1 and 2.2, chapter 6 of (Freidlin & Wentzell, 1979) in view of our three Lemmas 4.1,4.2,4.3 and our Proposition 4.3.

4.4 {i}-Graphs and invariant measure estimation

For sake of completeness, we recall here the {i}-Graphs definition for Markov chains living in a space which is divided into a finite union of states. For our skeleton Markov chain $(\tilde{Z}_n)_{n\in\mathbb{N}}$, these partition is $\cup g_i$. For any $i \in \{1\ldots\ell\}$, we define $\mathcal{G}(i)$ the set of oriented graphs with vertices $\{1,\ldots,\ell\}$ that satisfies the three following properties.

(i) Each state $j \neq i$ is the initial point of exactly one oriented edge in the graph.

(ii) The graph does not have any cycle.

(iii) For any $j \neq i$, there exists a (unique) path composed of oriented edge starting at state j and leading to the state i.

Given this definition, we can define

$$\mathcal{W}(z_i^\star) = \min_{g\in\mathcal{G}(i)} \sum_{(m\to n)\in g} \tilde{I}(z_m^\star, z_n^\star).$$

and as pointed in Lemma 4.1 of Freidlin and Wentzell (1979), one can check that

$$\mathcal{W}(z_i^\star) = \min_{g\in\mathcal{G}(i)} \sum_{(m\to n)\in g} I(z_m^\star, z_n^\star).$$

One may now deduce from the skeleton representation (Proposition 4.2) and from the estimations given by Lemma 4.3 and Proposition 4.4 the following theorem:

Theorem 4.1 *For any* $\gamma > 0$*, there exists* ρ_1 *satisfying* $0 < \rho_1 < \rho_0$ *such that*

$$e^{-\varepsilon^{-2}\left[\mathcal{W}(z_i^\star) - \min_{j\in\{1,\dots,\ell\}} \mathcal{W}(z_j^\star) + \gamma\right]} \leq \nu_\varepsilon(g_j) \leq e^{-\varepsilon^{-2}\left[\mathcal{W}(z_i^\star) - \min_{j\in\{1,\dots,\ell\}} \mathcal{W}(z_j^\star) - \gamma\right]},$$

for all $i \in \{1 \dots \ell\}$.

As well, in terms of W*, we get that*

$$e^{-\varepsilon^{-2}\left[W(z_i^\star) + \gamma\right]} \leq \nu_\varepsilon(g_j) \leq e^{-\varepsilon^{-2}\left[W(z_i^\star) - \gamma\right]},$$

for all $i \in \{1 \dots \ell\}$.

5 Minoration and Majoration of the rate function with a double-well landscape in $\mathbb{R}$

This last part is devoted to the proof of Theorem 2.3.

We describe in this part the behaviour of the memory gradient system with fixed memory parameter λ when U is a one-dimensional potential with a double-well profile. More precisely, we describe the asymptotic behaviour of our measure ν_ε when $\varepsilon \to 0$. We know from the last paragraph that ν_ε concentrates on the minimums of $\mathcal{W}$. If we want now to prove that ν_ε concentrates on the global minimum in the simplest case of a double-well potential with two minima $x_1^\star$ and $x_2^\star$, one needs to compare for the two stable equilibrias $z_1^\star = (x_1^\star, 0)$ or $z_2^\star = (x_2^\star, 0)$ the costs $I(z_1^\star, z_2^\star)$ and $I(z_2^\star, z_1^\star)$. Without loss of generality, we fix $x_1^\star < x_2^\star$ and there exists a unique maximum $x^\star$ of U such that $x_1^\star < x^\star < x_2^\star$ and $U'(x^\star) = 0, U''(x^\star) < 0$. We assume moreover that $U(x_1^\star) < U(x_2^\star)$ and that the memory effect is described through $k(t) = e^{\lambda t}$. Such potential U is represented in Figure 2.

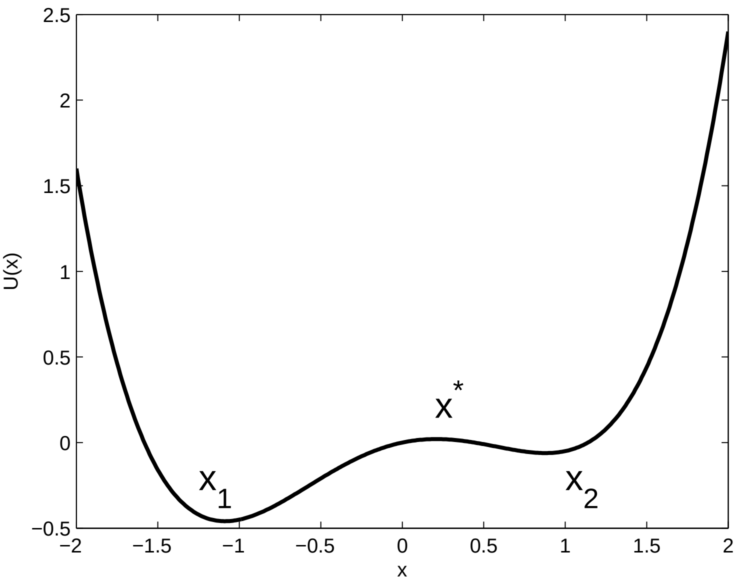

Figure 2: Example of double-well potential (here $U(x) = x^4/4 - x^2/2 + x/5$).

We first describe how one can provide a lower bound of the cost $I(z_1^\star, z_2^\star)$. To obtain this minoration, we follow the idea that in this context where the drift vector field is not a gradient, a Lyapunov function of the dynamical system may be useful to control the L^2 cost to move the sytem from $z_1^\star$ to $z_2^\star$ since $\nabla\mathcal{L}$ corresponds to a favored direction for the drift b. In this view, we consider the Lyapunov function $\mathcal{L}$ defined as

$$\mathcal{L}_{\beta,\gamma}(z) := \mathcal{L}_{\beta,\gamma}(x,y) = U(x) + \beta y^2/2 - \gamma U'(x)y,$$

that depends on parameters (β,γ) which will be properly chosen in the sequel. For the sake of simplicity, we will omit the dependence on β and γ and denote by $\mathcal{L}$ this function.
We propose two approachs: in the next subsection, we choose to use a non-degenerate approach where the main idea is to project the drift vector field on the gradient of the Lyapunov function. However, even if the idea seems to be original, the bounds are not very satisfactory (see Proposition 5.1). That is why, in Subsection 5.2 we propose a second approach which provides better bounds, where we optimize "directly" the choice of the parameters in the Lyapunov function (see Proposition (5.2)).

5.1 Minoration using a non-degenerate approach

The first idea is to use the simple remark that the cost I is necessarily bounded from below by the L^2 cost for an elliptic system. In particular, in the elliptic context the L^2 cost is defined as

$$I_{\mathcal{E},T}(z_1^\star,z_2^\star) = \inf_{\varphi\in\mathbb{H}}\left\{\frac{1}{2}\int_0^T |\dot\varphi_1 - b(\varphi)_1|^2 + |\dot\varphi_2 - b(\varphi)_2|^2 \quad | \quad \varphi(0)=z_1^\star, \varphi(T)=z_2^\star\right\},$$

which can also been written as

$$I_{\mathcal{E},T}(z_1^\star,z_2^\star) = \inf_{(u,v)\in\mathbb{L}^2([0;T])}\left\{\frac{1}{2}\int_0^T u^2(s)+v^2(s)ds \quad | \quad \dot{\mathbf{z}} = b(\mathbf{z}) + \begin{pmatrix}u\\ v\end{pmatrix}, \quad \mathbf{z}(0)=z_1^\star, \mathbf{z}(T)=z_2^\star\right\}. \tag{5.1}$$

As a consequence, since the set of admissible control for the degenerate cost I_T is contained in the set of admissible controls for $I_{\mathcal{E},T}$ (v is forced to be 0 in Equation (5.1)), we easily deduce that I_T is greater than $I_{\mathcal{E},T}$ and obtaining a lower bound of $I_{\mathcal{E},T}$ will yield a lower bound for I_T.

Now, let u and v be admissible controls for $I_{\mathcal{E},T}$, we have

$$u^2+v^2 = |\dot{\mathbf{z}} - b(\mathbf{z})|^2. \tag{5.2}$$

Adapting the approach developed in Chiang et al. (1987), we shall use the Lyapunov function $\mathcal{L}$ to bound from below the term above (somehow the Lyapunov function $\mathcal{L}$ will play the role of U). Indeed, if $\nabla\mathcal{L}\neq 0$, one can decompose b as follows

$$b(\mathbf{z}) = b_{\nabla\mathcal{L}(\mathbf{z})} + b_{\nabla\mathcal{L}(\mathbf{z})^\perp}, \tag{5.3}$$

where $b_{\nabla\mathcal{L}(\mathbf{z})}$ is the orthogonal projection of b on the direction $\nabla\mathcal{L}$. In the special case $\nabla\mathcal{L}=0$, we fix $b_{\nabla\mathcal{L}(\mathbf{z})}=0$ so that Equation (5.3) makes sense for any $\mathbf{z}$. One can now expand the L^2 cost using the simple remark that

$$\begin{aligned}
|\dot{\mathbf{z}} - b(\mathbf{z})|^2 &= |\dot{\mathbf{z}} - b_{\nabla\mathcal{L}(\mathbf{z})} - b_{\nabla\mathcal{L}(\mathbf{z})^\perp}|^2\\
&= |\dot{\mathbf{z}} - b_{\nabla\mathcal{L}(\mathbf{z})^\perp}|^2 + |b_{\nabla\mathcal{L}(\mathbf{z})}|^2 - 2\langle\dot{\mathbf{z}}; b_{\nabla\mathcal{L}(\mathbf{z})}\rangle\\
&\geq -2\frac{\langle b(\mathbf{z});\nabla\mathcal{L}(\mathbf{z})\rangle}{|\nabla\mathcal{L}(\mathbf{z})|^2}\langle\dot{\mathbf{z}};\nabla\mathcal{L}(\mathbf{z})\rangle.
\end{aligned}$$

Hence, if one can find β and γ such that there exists $\alpha>0$ such that

$$\forall z\in\mathbb{R}^2 \qquad -\frac{\langle b(z);\nabla\mathcal{L}(z)\rangle}{|\nabla\mathcal{L}(z)|^2}\geq\alpha, \tag{5.4}$$

then it is possible to conclude that

$$\begin{aligned}
\forall T>0 \qquad I_{\mathcal{E},T}(z_1,z_2) &= \inf_{(u,v)\in\mathbb{L}^2([0;T])}\left\{\frac{1}{2}\int_0^T u^2(s)+v^2(s)ds \;\mid\; \dot{\mathbf{z}} = b(\mathbf{z})+\begin{pmatrix}u\\ v\end{pmatrix}, z(0)=z_1, z(T)=z_2\right\}\\
&\geq \inf_{\varphi}\left\{\int_0^T -\frac{\langle b(\mathbf{z}(s));\nabla\mathcal{L}(\mathbf{z}(s))\rangle}{|\nabla\mathcal{L}(\mathbf{z}(s))|^2}\langle\dot{\mathbf{z}}(s);\nabla\mathcal{L}(\mathbf{z}(s))\rangle ds \;\mid\; \varphi(0)=z_1,\varphi(T)=z_2\right\},\\
&\geq \alpha[\mathcal{L}(\mathbf{z}(t))-\mathcal{L}(z_1^\star)], \qquad \forall t\in[0;T].
\end{aligned}$$

Now, remark that for admissible controls, $(\mathbf{z}(t))_{t\geq 0}$ moves continuously from z_1 to z_2 and there exists t^* such that $\mathbf{x}(t)=x^\star$. We then obtain

$$I_{\mathcal{E},T}(z_1^\star,z_2^\star)\geq \alpha[\mathcal{L}(\mathbf{z}(t^\star))-\mathcal{L}(z_1^\star)].$$

In the definition of $\mathcal{L}$ if $\beta\geq 0$, one obtains a lower bound of the cost of the form

$$I_{\mathcal{E},T}(z_1^\star,z_2^\star)\geq \alpha[U(x^\star)-U(x_1^\star)].$$

If $\beta\leq 0$, the only available minoration is obtained taking $t=T$ and we then get the weaker bound

$$I_{\mathcal{E},T}(z_1,z_2)\geq \alpha[U(x_2^\star)-U(x_1^\star)].$$

The next proposition provides a lower bound of the cost in the (restrictive) case of subquadratic potential U following the above idea.

Proposition 5.1 *Suppose* $U\in\mathcal{C}^2(\mathbb{R},\mathbb{R})$ *and define* $M=\|U"\|_\infty$*, then*

$$I_{\mathcal{E}}(z_1^\star,z_2^\star)\geq \alpha_\lambda(M)[U(x^\star)-U(x_1^\star)],$$

where $\alpha_\lambda(M)$ *satisfies the asymptotic properties*

$$\lim_{M\to 0}\alpha(M)=\frac{1}{1+\frac{1}{2\lambda^2}+\sqrt{1+\frac{1}{4\lambda^4}}}, \qquad \alpha(M)\sim_{M\to\infty}\frac{\lambda}{M\left[1+\frac{1}{2\lambda^2}+\sqrt{1+\frac{1}{4\lambda^4}}\right]}.$$

At last, we have

$$\forall M>0 \qquad \lim_{\lambda\mapsto+\infty}\alpha_\lambda(M)=2.$$

We remark that this bound strongly depends on the second derivative of U. In particular, when $M=\|U"\|_\infty$ is large, the lower bounds become very bad since it vanishes as $M\to+\infty$.
Proof : Let us first compute the projection of $b(z)$ on $\nabla\mathcal{L}(z)$ when it does not vanish. We can expand $\langle b(z);\nabla\mathcal{L}(z)\rangle$ as a quadratic form on variables $(U'(x),y)$.

$$\begin{aligned}
\langle b(z);\nabla\mathcal{L}(z)\rangle &= \langle\begin{pmatrix}-y\\ \lambda[U'(x)-y]\end{pmatrix};\begin{pmatrix}U'(x)-\gamma U"(x)y\\ \beta y-\gamma U'(x)\end{pmatrix}\rangle\\
&= -[\beta\lambda-\gamma U"(x)]y^2-\gamma\lambda U'(x)^2+yU'(x)[\gamma\lambda+\beta\lambda-1]
\end{aligned}$$

If we set

$$M_1=\begin{pmatrix}\gamma\lambda & -\frac{\gamma\lambda+\beta\lambda-1}{2}\\ -\frac{\gamma\lambda+\beta\lambda-1}{2} & [\beta\lambda-\gamma U"(x)]\end{pmatrix},$$

we then obtain

$$\langle b(z);\nabla\mathcal{L}(z)\rangle = -\begin{pmatrix}U'(x)\\ y\end{pmatrix}^t M_1\begin{pmatrix}U'(x)\\ y\end{pmatrix}. \tag{5.5}$$

In the same way, one can compute that

$$|\nabla\mathcal{L}(z)|^2 = U'(x)^2(1+\gamma^2) + y^2(\gamma^2 U"(x)^2 + \beta^2) - 2U'(x)y(\gamma U"(x) + \beta\gamma),$$

so that

$$|\nabla\mathcal{L}(z)|^2 = \begin{pmatrix} U'(x) \\ y \end{pmatrix}^t M_2 \begin{pmatrix} U'(x) \\ y \end{pmatrix}, \tag{5.6}$$

where

$$M_2 = \begin{pmatrix} 1+\gamma^2 & -\gamma U"(x) - \beta\gamma \\ -\gamma U"(x) - \beta\gamma & \beta^2 + \gamma^2 U"(x)^2 \end{pmatrix}.$$

In order to have an equal equilibrium between the repelling effect on $U'(x)^2$ and y^2 on the quadratic defined by M_1, a natural choice for β and γ would be

$$\begin{cases} \min_{x\in\mathbb{R}}[\beta\lambda - \gamma U"(x)] = \gamma\lambda \\ 1 - \gamma\lambda - \beta\lambda \quad = 0 \end{cases}$$

Hence, we set

$$\beta = \frac{\lambda + M}{\lambda(2\lambda+M)} \qquad \text{and} \qquad \gamma = \frac{1}{2\lambda+M}.$$

The end of the proof then falls into an algebraic argument : denote $(a,b) = (U'(x), y)$, we are looking for a bound similar to (5.4) with the larger possible α . The projection of b on $\nabla\mathcal{L}$ can be expressed as

$$-\frac{\langle b(z); \nabla\mathcal{L}(z)\rangle}{|\nabla\mathcal{L}(z)|^2} = \frac{q_{M_1}(a,b)}{q_{M_2}(a,b)},$$

where q_{M_1} and q_{M_2} are the two quadratic forms defined from expressions (5.5) and (5.6). To bound the ratios of these two quadratic forms, remark that M_1 is invertible except if $M = 0$ which is a rather trivial case. Then, M_1 is symetric and positive definite as well as M_2 is non-negative and symetric. It is possible to use a simultaneous reduction of q_{M_1} and q_{M_2}. We denote ρ_1 and ρ_2 the eigenvalues of $M_1^{-1}M_2$ associated to eigenvectors e_1, e_2 which are an orthonormal basis for q_{M_1}, if $(\tilde{a}, \tilde{b})$ are the coordinates in this basis we have

$$\frac{q_{M_1}(a,b)}{q_{M_2}(a,b)} = \frac{\tilde{a}^2 + \tilde{b}^2}{\rho_1\tilde{a}^2 + \rho_2\tilde{b}^2},$$

and the minimum of q_{M_1}/q_{M_2} is then

$$\min_{(a,b)} \frac{q_{M_1}(a,b)}{q_{M_2}(a,b)} = \frac{1}{\rho_1} \wedge \frac{1}{\rho_2} = \frac{1}{\rho_1 \vee \rho_2}.$$

With our choice of β and γ and setting $\xi = U"(x)/M \in [-1;1]$, simple algebra yields

$$M_1^{-1}M_2(\xi) = \begin{pmatrix} 2 + \frac{M}{\lambda} + \frac{1}{\lambda(2\lambda+M)} & -\left(\frac{M}{\lambda}\xi + \frac{1+\frac{M}{\lambda}}{\lambda(2\lambda+M)}\right) \\ -\left(\frac{M\xi + \frac{\lambda+M}{\lambda(2\lambda+M)}}{\lambda+M(1-\xi)}\right) & \frac{\xi^2M^2+(1+M/\lambda)^2}{(\lambda+M(1-\xi))(2\lambda+M)} \end{pmatrix}.$$

For small M, it is immediate to show that $M_1^{-1}M_2$ becomes independent of ξ and that

$$\frac{1}{\rho_1 \vee \rho_2} = \frac{1}{1 + \frac{1}{2\lambda^2} + \sqrt{1 + \frac{1}{4\lambda^4}}}.$$

For large M, the maximum eigenvalue of $M_1^{-1}M_2$ is reached for $\xi = 1$ and one obtains again after tedious computations that

$$\frac{1}{\rho_1 \vee \rho_2} = \frac{\lambda}{M\left[1 + \frac{1}{2\lambda^2} + \sqrt{1 + \frac{1}{4\lambda^4}}\right]}.$$

For any $M > 0$, the coefficient $\alpha(M)$ is then obtained by

$$\alpha(M) = \min_{\xi \in [-1;1], \rho \in \mathrm{Sp}(M_1^{-1}M_2(\xi))} \frac{1}{\rho}$$

and the result holds.

At last, remark that for any large λ, the matrix $M_1^{-1}M_2$ becomes diagonal with the two eigenvalues 2 and 0. This proves that $\lim_{\lambda \to +\infty} \alpha_\lambda(M) = 2$. □

5.2 Minoration using a degenerate approach (Proof of ii) of Theorem 2.3)

To take into account the degeneracy of the dynamical system for the control cost, we directly bound the terms in the integral of $I_T(z_1, z_2)$ by a gradient of a suitable Lyapunov function. This may lead to better estimates since obviously in the previous paragraph we use a minoration technique based on elliptic argument.

Given any $\alpha > 0$ and $(\beta, \gamma) \in \mathbb{R}^2$, we define

$$\mathcal{L}_{\alpha,\beta,\gamma}(z) = \mathcal{L}_{\alpha,\beta,\gamma}(x, y) := \alpha U(x) + \beta y^2/2 - \gamma y U'(x).$$

We are looking for an ideal choice of (α, β, γ). For every $\varphi \in \mathbb{H}(\mathbb{R}_+, \mathbb{R}^d)$, we set $u = \dot{\varphi}$. If u denotes any admissible control and $(\mathbf{z}(t))_{t \geq 0}$ is the controlled trajectory, we aim to obtain a bound such as for every $\varphi \in \mathbb{H}(\mathbb{R}_+, \mathbb{R}^d)$, we have

$$\forall t \geq 0 \qquad u^2(t) \geq 2\frac{d\mathcal{L}_{\alpha,\beta,\gamma}(\mathbf{z}(t))}{dt}. \tag{5.7}$$

Recall that $t^\star$ is the first time such that z reaches the local maximum of U (*i.e.* $\mathbf{x}(t^\star) = x^\star$). Such lower bound is usefull especially if α is positive and large and β non-negative. Indeed assume that we can obtain lower bound of the form (5.7), we then have for all T:

$$\begin{aligned}
I_T(z_1^\star, z_2^\star) &= \inf_{u \in \mathbb{L}^2([0;T])} \left\{ \frac{1}{2}\int_0^T u^2(s)ds \quad | \quad \dot{\mathbf{z}} = b(\mathbf{z}) + \begin{pmatrix} u \\ 0 \end{pmatrix}, \quad \mathbf{z}(0) = z_1^\star, \mathbf{z}(T) = z_2^\star \right\} \\
&\geq \inf_{u \in \mathbb{L}^2([0;T])} \left\{ \frac{1}{2}\int_0^{t^\star} u^2(s)ds \quad | \quad \dot{\mathbf{z}} = b(\mathbf{z}) + \begin{pmatrix} u \\ 0 \end{pmatrix}, \quad \mathbf{z}(0) = z_1^\star, \mathbf{z}(T) = z_2^\star \right\} \\
&\geq \alpha[U(x^\star) - U(x_1^\star)] + \beta \mathbf{y}(t^\star)^2 \\
&\geq \alpha[U(x^\star) - U(x_1^\star)].
\end{aligned} \tag{5.8}$$

The next proposition shows that indeed such minoration (5.7) holds for some suitable choice of β, γ and in some case, this minoration is almost optimal.

Proposition 5.2 *For every $\alpha \in [0; 2[$, there exist explicit constants $m_\lambda(\alpha), \beta^\star(\alpha), \gamma^\star(\alpha)$ such that (5.7) is true for $\beta = \beta^\star(\alpha), \gamma = \gamma^\star(\alpha)$ and for every one-dimensional double well potential U satisfying $\|U''\|_\infty = M < m_\lambda(\alpha)$. In this case, let U such that $\|U''\|_\infty = M < m_\lambda(\alpha)$, we get*

$$I(z_1^\star, z_2^\star) \geq \alpha\Big[U(x^\star) - U(x_1^\star)\Big].$$

Proof : In order to obtain a minoration such as (5.7), we fix any $t > 0$ and any admissible control u. Dropping the time parameter, note that, we have

$$u^2 = |\dot{\mathbf{x}} + \mathbf{y}|^2 = \dot{\mathbf{x}}^2 + \mathbf{y}^2 + 2\dot{\mathbf{x}}\mathbf{y},$$

and since $\dot{\mathbf{y}} = \lambda[U'(\mathbf{x}) - \mathbf{y}]$, one can also compute

$$\begin{aligned}
\frac{d\mathcal{L}_{\alpha,\beta,\gamma}(\mathbf{z})}{dt} &= \alpha\dot{\mathbf{x}}U'(\mathbf{x}) + \lambda\beta\mathbf{y}(U'(\mathbf{x}) - \mathbf{y}) - \gamma\mathbf{y}\dot{\mathbf{x}}U''(\mathbf{x}) - \lambda\gamma(U'(\mathbf{x}) - \mathbf{y})U'(\mathbf{x}) \\
&= -\lambda\beta\mathbf{y}^2 - \lambda\gamma U'(\mathbf{x})^2 + \alpha\dot{\mathbf{x}}U'(\mathbf{x}) + \lambda\mathbf{y}U'(\mathbf{x})(\beta+\gamma) + \dot{\mathbf{x}}\mathbf{y}(-\gamma U''(\mathbf{x})).
\end{aligned}$$

Let us now define M_1 and M_2 as

$$M_1 = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \qquad M_2 = \begin{pmatrix} 0 & \frac{\alpha}{2} & -\frac{\gamma}{2}U''(x) \\ \frac{\alpha}{2} & -\lambda\gamma & \frac{\beta+\gamma}{2}\lambda \\ -\frac{\gamma}{2}U''(x) & \frac{\beta+\gamma}{2}\lambda & -\beta\lambda \end{pmatrix},$$

for all $x \in \mathbb{R}$. This way, we can write

$$u^2 = (\dot{\mathbf{x}}, U'(\mathbf{x}), \mathbf{y})M_1(\dot{\mathbf{x}}, U'(\mathbf{x}), \mathbf{y})^t, \qquad \text{and} \qquad \frac{d\mathcal{L}_{\alpha,\beta,\gamma}(\mathbf{z})}{dt} = (\dot{\mathbf{x}}, U'(\mathbf{x}), \mathbf{y})M_2(\dot{\mathbf{x}}, U'(\mathbf{x}), \mathbf{y})^t.$$

Remark that again, the product $yU'(x)$ is essential in the structure of the Lyapunov function since it creates some repelling effect in M_2 in variables y^2 and $U'(x)^2$. Without this term, there is no chance to obtain positiveness of $M_1 - 2M_2$. Moreover, we immediately get that β and γ should be positives.

It is then sufficient to obtain that the symmetric matrix $S := M_1 - 2M_2$ is positive. We again introduce a parameter ξ but this time it is easier to manipulate $\xi = U''(x) \in [-M; M]$. The matrix S becomes:

$$S = \begin{pmatrix} 1 & -\alpha & 1+\gamma\xi \\ -\alpha & 2\gamma\lambda & -(\beta+\gamma)\lambda \\ 1+\gamma\xi & -(\beta+\gamma)\lambda & 1+2\beta\lambda \end{pmatrix}.$$

S is positive if and only if principal minors are positives. We trivially have $\Delta_1 = 1 > 0$. Hence, we compute

$$\Delta_2 := \det\begin{pmatrix} 1 & -\alpha \\ -\alpha & 2\gamma \end{pmatrix} = 2\gamma - \alpha^2.$$

Hence, imposing $2\gamma\lambda > \alpha^2$ implies the positiveness of Δ_2. Regarding now $\Delta_3 := \det S$, after several computations, we obtain

$$\Delta_3(\xi) = C + B\xi - A\xi^2$$

where

$$A = 2\lambda\gamma^3, B = 2\alpha\gamma\lambda(\beta+\gamma) - 4\lambda\gamma^2, \quad \text{and} \quad C = (1+2\beta\lambda)\left(2\lambda\gamma - \alpha^2\right) - \lambda^2(\beta+\gamma)^2 + 2\alpha\lambda(\beta+\gamma) - 2\gamma\lambda.$$

Hence, Δ_3 is a quadratic polynomial of ξ, it is impossible to obtain the positiveness of $\Delta_3(\xi)$ for all $\xi \in \mathbb{R}$ (this justifies we suppose that $\|U''\|_\infty < +\infty$). For any $\alpha > 0$, we aim to maximise the absolute values of the roots of Δ_3 among the convenient choices of β and γ. By a symmetry argument, it is easy to check that one should have necessarily $B = 0$ since in this case the roots of Δ_3 are opposite.

Thus the parameter β can be expressed in terms of α and γ:

$$\beta = \beta(\alpha,\gamma) = \gamma\,(2/\alpha - 1),$$

and for this choice, the roots of Δ_3 are

$$x_{\pm}(\alpha,\gamma) = \pm\sqrt{\frac{-2(1/\alpha-1)^2\lambda}{\gamma} + \frac{1+(\alpha-1)^2}{\gamma^2} - \frac{\alpha^2}{2\lambda\gamma^3}}.$$

Note also that for this choice, we obtain

$$\Delta_3(0) = -\alpha^2 + 2\lambda\gamma(1+(\alpha-1)^2) - \gamma^2(4(1/\alpha-1)^2)\lambda^2.$$

Since $\Delta_3(0)$ must be positive, it is easy to show that $\alpha < 2$ (when $\alpha = 2$, $\Delta_3(0) = -(\gamma\lambda - 2)^2$ which is the limiting case). It is possible to maximise $|x_{\pm}(\alpha,\gamma)|$ with respect to γ for $\alpha \in [0;2[$. Differentiating with respect to γ, the optimal $\gamma(\alpha)$ is solution of

$$\lambda(1/\alpha-1)^2\gamma^2 - [1+(\alpha-1)^2]\gamma + 3\alpha^2/(4\lambda) = 0.$$

Solving this equation, we then successively obtain

$$\gamma^\star_\lambda(\alpha) = \frac{[1+(\alpha-1)^2] - \sqrt{[1+(\alpha-1)^2]^2 - 3\alpha^2(1/\alpha-1)^2}}{2(1/\alpha-1)^2\lambda},$$

and

$$\beta^\star(\alpha) = \beta(\alpha,\gamma^\star(\alpha)).$$

The maximum admissible value for $\|U''\|_\infty$ defined as $m_\lambda(\alpha)$ in this proposition is then obtained using $\gamma^\star(\alpha)$ in $x_{\pm}(\alpha,\gamma)$, that is,

$$m_\lambda(\alpha) = x_+(\alpha,\gamma^\star_\lambda(\alpha)).$$

Now since (5.8) is true for all T the result holds for the cost I. □

Note that when M is large, the admissible values for α vanish and our lower bound becomes useless. When $M \mapsto 0$, we obtain $I(z_1,z_2) \geq 2[U(x^\star) - U(x_1)]$ which is optimal in view of the upper bound constructed in the next paragraph (it is obviously better than the bound obtained in Proposition 5.1). The evolution of admissible α is shown in Figures 3 for several values of λ.

As announced in the beginning of the section, one may remark that the second approach is clearly more efficient than the first one. However, we chose to keep the first approach since the idea can be of interest in a more general context, especially in an elliptic case with a drift vector field which is not a gradient.

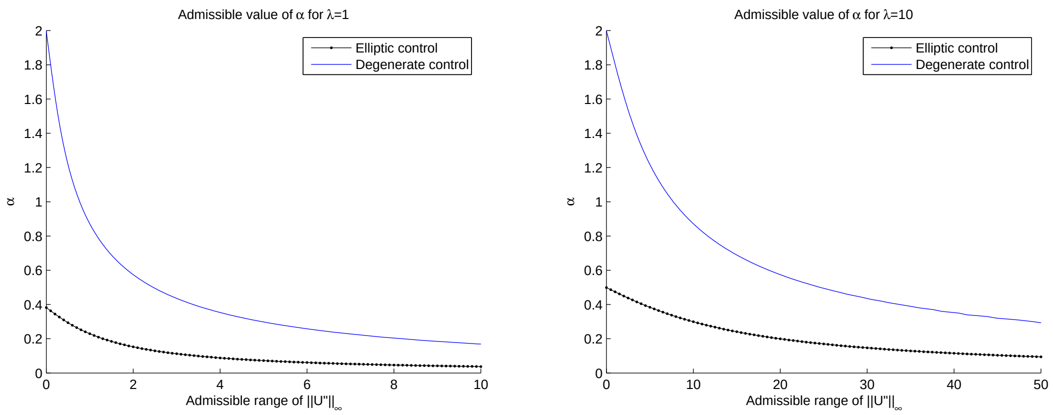

Figure 3: Evolution of the maximum size of α with $\|U''\|_\infty$ when $\lambda = 1$ (left) and $\lambda = 10$ (right) for both approaches (Proposition 5.1 and 5.2.

5.3 Upper-Bound for the cost function (Proof of i) of Theorem 2.3)

Remind that we assume that there are two local minimums for U denoted by $x_1^\star$ and $x_2^\star$ with $U(x_1^\star) < U(x_2^\star)$ and a local maximum denoted by $x^\star$. We set $z_1^\star = (x_1^\star, 0)$, $z_2^\star = (x_2^\star, 0)$ and $z^\star = (x^\star, 0)$. In this particular setting, we want to obtain an upper-bound for $I(z_2^\star, z_1^\star)$ and then for W. This is the purpose of the next proposition.

Proposition 5.3 *Assume that U is a one-dimensional double well potential defined as above such that $U(x_1^\star) < U(x_2^\star)$ and such that U'' is not equal to 0 at equilibrium points. Then, for every $\lambda > 0$,*

$$W(z_1^\star) = I(z_2^\star, z_1^\star) \leq 2(U(x^\star) - U(x_2^\star)).$$

where $z_1^\star = (x_1^\star, 0)$, $z_2^\star = (x_2^\star, 0)$ and $x^\star$ denotes the unique local maximum of U.

Essentially, the previous proposition is a consequence of Lemma 5.1 and Lemma 5.2 combined with the fact that $I(z_2^\star, z_1^\star) \leq I(z_2^\star, z^\star) + I(z^\star, z_1^\star)$. Let us stress that the proofs of Lemma 5.1 and Lemma 5.2 rely both on Lemma 5.3.

Lemma 5.1 *Under the assumptions of Proposition 5.3, we have*

$$I(z^\star, z_i^\star) = 0 \quad \text{for } i = 1, 2.$$

Proof : We show the result for $i = 1$. The last part of the proof of this Lemma relies on Lemma 5.3. Indeed, first, we prove that for every initial point $z_\varepsilon = (x_\varepsilon, y_\varepsilon)$ with $U(x_\varepsilon) + |y_\varepsilon|^2/2 < U(x^\star)$ and $x_\varepsilon \in [x_1^\star, x^\star]$, $I(z_\varepsilon, z_1^\star) = 0$. Second, applying Lemma 5.3 we will have $I(z^\star, z_\varepsilon) \leq \varepsilon$, for all $\varepsilon > 0$ and taking ε goes to zero the result follows.

In order to prove the first part, we adopt a similar strategy as the proof used in Section 3.4. Indeed, assume that $(\mathbf{x}(t), \mathbf{y}(t))$ is solution of

$$\begin{cases} \dot{\mathbf{x}}(t) = -\mathbf{y}(t) \\ \dot{\mathbf{y}}(t) = \lambda(U'(\mathbf{x}(t)) - \mathbf{y}(t)) \end{cases} \tag{5.9}$$

starting from z_ε. Considering the function F defined by $F(t) = \mathcal{E}(\mathbf{x}(t), \mathbf{y}(t)) = U(\mathbf{x}(t)) + |\mathbf{y}(t)|^2/(2\lambda)$, one can check that $F'(t) = -\mathbf{y}(t)^2$. In particular, F is a positive non-increasing function thus convergent. Then, $(\mathbf{x}(t), \mathbf{y}(t))_{t\geq 0}$ is bounded and the fact that $\mathcal{E}(z_\varepsilon) \leq U(x^\star)$ shows that z_ε belongs to a compact set, as well as all the trajectories initialized in $\mathcal{E}^{-1}(]-\infty; U(x^\star)])$. Since U" is continuous, we then deduce that U' is Lipschitz continuous on the set where $(\mathbf{x}, \mathbf{y})$ is living. It implies classically that the family of shifted trajectories $(\mathbf{z}(t + .))_{t\geq 0}$ is relatively compact for the topology of uniform convergence on compact sets. Now, let $\mathbf{z}^\infty$ denote the limit of a convergent subsequence $(\mathbf{z}(t_n + .))_{n\geq 0}$, $\mathbf{z}^\infty$ is a solution of (5.9) and since F is continuous and converges as $t \to +\infty$ to some limit l, we have necessarily $\mathcal{E}(\mathbf{x}^\infty(s), \mathbf{y}^\infty(s)) = l$ for all $s \geq 0$. Thus we get

$$\frac{d}{dt}\mathcal{E}(\mathbf{x}^\infty(t), \mathbf{y}^\infty(t)) = 0,$$

for all $t \geq 0$. Using that $F'(t) = -\mathbf{y}(t)^2$ for a solution of (5.9), we then obtain that $\mathbf{y}^\infty(t) = 0$ for every $t \geq 0$.

Thus, $\mathbf{x}^\infty$ is constant and $\mathbf{z}^\infty$ is a *stationary* solution of the ordinary differential equation (5.9). We can deduce that every accumulation point of $(\mathbf{x}(t), \mathbf{y}(t))$ belongs to $\{(x, y) \in \mathbb{R}, b(x, y) = 0\}$. Under the assumption $U(x_\varepsilon) + |y_\varepsilon|^2/2 < U(x^\star)$, and since F is non increasing, the only possible accumulation point is $z_1^\star$. Then, $(\mathbf{x}(t), \mathbf{y}(t)) \xrightarrow{t\to+\infty} z_1^\star$ and $I(z_\varepsilon, z_1^\star) = 0$. As a consequence, we have

$$I(z^\star, z_1^\star) \leq I(z^\star, z_\varepsilon).$$

We shall now prove the second part. As announced at the beginning of the proof, applying Lemma 5.3, we get that for every $\varepsilon > 0$, one may find $z_\varepsilon = (x_\varepsilon, y_\varepsilon)$ such that we both have $U(x_\varepsilon) + |y_\varepsilon|^2/2 < U(x^\star)$, $x_\varepsilon \in [x_1, x^\star]$ and $I(z^\star, z_\varepsilon) \leq \varepsilon$. Taking finally $\varepsilon \to 0$, the result holds $\square$

Lemma 5.2 *Assume the assumptions of Proposition 5.3. Then,*

$$I(z_2^\star, z^\star) \leq 2(U(x^\star) - U(x_2^\star)).$$

Proof : The idea of the proof is to use the " reverse" differential flow which transports the trajectory on $z^\star$ with a computable cost. Of course, since $z_2^\star$ is a stationary point of (5.9), we shall first move the trajectory out of this stationary point and we use an auxiliary point $\tilde{z}_\varepsilon$ sufficiently close to $z_2^\star$.

In this view, let us define $\tilde{z}_\varepsilon = z_2^\star + \varepsilon(z^\star - z_2^\star)$ with $\varepsilon \in (0,1)$ being such that $U(\tilde{x}_\varepsilon) > U(x_2^\star)$. We then have $\tilde{x}_\varepsilon = x_2^\star + \varepsilon(x^\star - x_2^\star)$ and $\tilde{y}_\varepsilon = 0$. We want to show that

$$I(\tilde{z}_\varepsilon, z^\star) \leq 2(U(x^\star) - U(\tilde{x}_\varepsilon)).$$

Let us consider the controlled trajectory denoted $\mathbf{z}_\varepsilon = (\mathbf{x}_\varepsilon, \mathbf{y}_\varepsilon)$ obtained with $\varphi = 2\mathbf{y}_\varepsilon$, this trajectory is then solution of

$$\begin{cases} \dot{\mathbf{x}}(t) = \mathbf{y}(t) = -\mathbf{y}(t) + 2\mathbf{y}(t) \\ \dot{\mathbf{y}}(t) = \lambda(U'(\mathbf{x}(t)) - \mathbf{y}(t)) \end{cases} \tag{5.10}$$

starting from $\tilde{z}_\varepsilon$. We shall now study its asymptotic behaviour. To this end, we introduce now the function $\tilde{F}$ defined by $\tilde{F}(t) = \frac{\mathbf{y}_\varepsilon(t)^2}{2\lambda} - U(\mathbf{x}_\varepsilon(t))$ ($\tilde{F}$ is the equivalent of function F defined above for the uncontrolled trajectory). We observe that $\tilde{F}'(t) = -\mathbf{y}_\varepsilon(t)^2$ and thus $\tilde{F}$ is non-increasing. We first show that the solution $(\mathbf{x}_\varepsilon(t), \mathbf{y}_\varepsilon(t))$ to (5.10) starting from $\tilde{z}_\varepsilon$ satisfies necessarily $x_1^\star < \mathbf{x}_\varepsilon(t) < x_2^\star$ for every $t \geq 0$. Actually, on the one hand, let $\tau_1 := \inf\{t \geq 0, \mathbf{x}_\varepsilon(t) = x_1^\star\}$. Then, if $\tau_1 < +\infty$, remark that $\tilde{y}_\varepsilon = 0$ and one should have

$$\frac{\mathbf{y}_\varepsilon(\tau_1)^2}{2\lambda} - U(x_1^\star) = -U(\tilde{x}_\varepsilon) - \int_0^{\tau_1} \mathbf{y}_\varepsilon(s)^2 ds \tag{5.11}$$

and it would follow that $U(\tilde{x}_\varepsilon) - U(x_1^\star) < 0$, which is impossible under our assumptions. Thus, $\tau_1 = +\infty$. On the other hand, the fact that $U(\tilde{x}_\varepsilon) > U(x_2^\star)$ implies in a same way that $\tau_2 := \inf\{t \geq 0, \mathbf{x}_\varepsilon(t) = x_2^\star\}$ satisfies $\tau_2 = +\infty$. Thus, we obtain that $(\mathbf{x}_\varepsilon(t))_{t\geq 0}$ belongs to the interval $(x_1^\star, x_2^\star)$.
This point combined with the decrease of $\tilde{F}$ implies that

$$\sup_{t\geq 0} |\mathbf{y}_\varepsilon(t)| \leq \sqrt{2\lambda \left[\tilde{F}(0) + \sup_{x\in[x_1,x_2]} |U|(x)\right]} < +\infty.$$

As a consequence, $(\mathbf{x}_\varepsilon(t), \mathbf{y}_\varepsilon(t))_{t\geq 0}$ is bounded and a similar argument as in the proof of Lemma 5.1 yields that the limit $(\mathbf{x}_\varepsilon^\infty, \mathbf{y}_\varepsilon^\infty)$ of any convergent subsequence $(\mathbf{x}_\varepsilon(t_n + .), \mathbf{y}_\varepsilon(t_n + .))_{n\geq 1}$ lies in the stationary solutions of (5.10). Thus, we deduce that $\mathbf{x}_\varepsilon^\infty \in \{x_1^\star, x^\star, x_2^\star\}$.
Now, equation (5.11) and the fact that $U(\tilde{x}_\varepsilon) > U(x_1^\star)$ and $U(\tilde{x}_\varepsilon) > U(x_2^\star)$ imply that $\mathbf{x}_\varepsilon^\infty$ can not be $x_1^\star$ or $x_2^\star$. Thus,

$$(\mathbf{x}_\varepsilon(t), \mathbf{y}_\varepsilon(t)) \xrightarrow{t\to+\infty} z^\star.$$

Finally, recall that writing $\varphi(t) = 2\mathbf{y}(t)$, the differential system (5.10) can be written

$$\begin{cases} \dot{\mathbf{x}}(t) = -\mathbf{y}(t) + \varphi(t) \\ \dot{\mathbf{y}}(t) = \lambda(U'(\mathbf{x}(t)) - \mathbf{y}(t)). \end{cases} \tag{5.12}$$

We then deduce that

$$I(\tilde{z}_\varepsilon, z^\star) \le \frac{1}{2}\int_0^{+\infty} |\varphi(s)|^2 ds = 2\int_0^{+\infty} |\mathbf{y}_\varepsilon(s)|^2 ds = 2(U(x^\star) - U(\tilde{x}_\varepsilon)).$$

Since $I(z_2^\star, \tilde{z}_\varepsilon) \to 0$ as $\varepsilon \to 0$ by Lemma 5.3, the announced result follows . □

We finish this section by showing the result used in the above proofs. This key lemma exploits the infinite number of oscillations of the dynamical system close to its stable points.

Lemma 5.3 *Let x_0 denote an equilibrium point for U such that, in the neighbourhood of x_0, U is strictly convex (resp. strictly concave) if x_0 is a local maximum (resp. a local minimum). Let $v \in \mathbb{R}^2$ with $|v| = 1$. Then, for every $\varepsilon > 0$ for every $\rho > 0$, there exists z_ε on the segment $[x_0, x_0 + \rho v]$ and $\tau \ge 0$ such that $I_\tau(z^\star, z_\varepsilon) \le \varepsilon$.*

Note that this point is a direct consequence of Lemma 4.1 if $U''(x_0) \neq 0$. However, we choose to give below an alternative proof under a less constraining assumption, based on the turning dynamics of the system.

Proof : Set $z_0 = (x_0, 0)$. Since $b(x_0, 0) = 0$, we first remark that for every $\varepsilon > 0$, there exists $\delta_\varepsilon > 0$, such that the solution $(\mathbf{z}_1(t))_{t\ge 0}$ of the controlled trajectory (with control $\varphi = 1$)

$$\begin{cases} \dot{\mathbf{x}}(t) = -\mathbf{y}(t) + 1 \\ \dot{\mathbf{y}}(t) = \lambda(U'(\mathbf{x}(t)) - \mathbf{y}(t)) \end{cases}$$

satisfies $\mathbf{x}(\delta_\varepsilon) \neq x_0$, $|\mathbf{z}_1(\delta_\varepsilon) - z_0| \le \varepsilon$ and $I(z_0, \mathbf{z}_1(\delta_\varepsilon)) \le \varepsilon$. Let us now consider separately the cases where x_0 is either a local minimum or a local maximum. In the first case, we consider the solution $(\mathbf{z}_2(t))_{t\ge\delta_\varepsilon}$ to (5.9) satisfying $\mathbf{z}_2(\delta_\varepsilon) = \mathbf{z}_1(\delta_\varepsilon)$. We already know from the proof of Lemma 5.1 that $(\mathbf{z}_2(t))_{t\ge\delta_\varepsilon}$ converges to z_0 for ε small enough (this part of the proof of Lemma 5.1 did not rely on the result of Lemma 5.3). Using that U is strictly convex in the neighbourhood of x_0, if we let δ_ε small enough, we also deduce from Theorem 6.1 of Cabot et al. (2009a) that the number of sign changes (oscillations of the dynamical system) of $(\dot{\mathbf{x}}(t))$ and thus of $(\mathbf{y}(t))$ is infinite. These two points imply that $(\mathbf{z}_2(t))$ converges to z_0 by turning around z_0 and for every $v \in \mathbb{R}^2$ with $|v| = 1$, for every $\rho > 0$, there exist $\tau \ge \delta_\varepsilon$ and $\tilde{\rho} \in (0, \rho)$ such that $\mathbf{z}_2(\tau) = x_0 + \tilde{\rho} v := z_\varepsilon$. The result follows in this case using that $I_\tau(z_0, \mathbf{z}_2(\tau)) \le I_{\delta_\varepsilon}(z_0, \mathbf{z}_1(\delta_\varepsilon)) \le \varepsilon$.

Consider now the case where x_0 is a local maximum, the proof is similar but $(\mathbf{z}_2(t))_{t\ge\delta_\varepsilon}$ is replaced by the solution of (5.10) starting from $\mathbf{z}_1(\delta_\varepsilon)$ and U by $-U$. Following the proof of Lemma 5.2, we obtain in this case that,

$$\begin{aligned} I(z_0, \mathbf{z}_2(\tau)) &\le I_{\delta_\varepsilon}(z_0, \mathbf{z}_1(\delta_\varepsilon)) + 2\int_{\delta_\varepsilon}^{\tau} |\mathbf{y}_2(s)|^2 ds \\ &\le \varepsilon + 2\int_{\delta_\varepsilon}^{+\infty} |\mathbf{y}(s)|^2 ds \le \varepsilon + 2\left(U(x_0) - U(\mathbf{x}_1(\delta_\varepsilon)) - \frac{|\mathbf{y}_1(\delta_\varepsilon)|^2}{2}\right) \end{aligned}$$

and the result follows taking δ_ε small enough. □

Appendix A: Proof of Proposition 3.2

Let $\varepsilon > 0$ and h be a bounded continuous function. Since ν_ε is an invariant distribution, we have for every $t > 0$,

$$\int h^{\frac{1}{\varepsilon^2}} d\nu_\varepsilon = \int h_{\varepsilon,t} d\nu_\varepsilon \quad \text{where} \quad h_{\varepsilon,t}(z) = \mathbb{E}[h^{\frac{1}{\varepsilon^2}}(Z_t^{(\varepsilon),z})].$$

Since h is bounded continuous, it follows from Assumption (ii) and Lemma 3.1.12 of (Puhalskii, 2001) that for every $z \in \mathbb{R}^{2d}$, for every $(z_\varepsilon)_{\varepsilon>0}$ such that $z_\varepsilon \to z$,

$$\lim_{\varepsilon \to 0} (h_{\varepsilon,t})^{\varepsilon^2}(z_\varepsilon) = \sup_{\nu \in \mathbb{R}^{2d}} h(\nu) \exp(-I_t(z,\nu)) \tag{A.1}$$

Now, since $(\nu_\varepsilon)_{\varepsilon>0}$ is exponentially tight, $(\nu_\varepsilon)_{\varepsilon>0}$ admits some (LD)-convergent subsequence. Let $(\nu_{\varepsilon_n})_{n\geq 1}$ denote such a subsequence. Then, $(\nu_{\varepsilon_n})_{n\geq 1}$ satisfies a large deviation principle with speed ε^{-2} and rate function denoted by W. Then, by Lemma 3.1.12 of (Puhalskii, 2001), we have

$$\left(\int h^{\frac{1}{\varepsilon_n^2}} d\nu_{\varepsilon_n}\right)^{\varepsilon_n^2} \xrightarrow{n\to+\infty} \sup_{z\in\mathbb{R}^{2d}} h(z)\exp(-W(z)).$$

and by (A.1) and Lemma 3.1.13 of (Puhalskii, 2001), we obtain that

$$\left(\int h_{\varepsilon_n,t} d\nu_{\varepsilon_n}\right)^{\varepsilon_n^2} \xrightarrow{n\to+\infty} \sup_{z\in\mathbb{R}^{2d}} \left(\sup_{\nu\in\mathbb{R}^{2d}} h(\nu)\exp(-I_t(z,\nu))\exp(-W(z))\right).$$

It follows that for every bounded continuous function h,

$$\sup_{z\in\mathbb{R}^{2d}} h(z)\exp(-W(z)) = \sup_{z\in\mathbb{R}^{2d}} h(z) \left(\sup_{\nu\in\mathbb{R}^{2d}} \exp(-I_t(\nu,z))\exp(-W(\nu))\right).$$

By Theorem 1.7.27 of (Puhalskii, 2001), the above equality holds in fact for every bounded measurable function h. Applying this equality with $h = \mathbf{1}_{\{z_0\}}$

$$\exp(-W(z_0)) = \sup_{\nu\in\mathbb{R}^{2d}} \exp(-I_t(\nu,z_0))\exp(-W(\nu)),$$

and the result follows. □

Appendix B: Proof of Lemma 3.2

The explicit computation of $\mathcal{A}^\varepsilon V^p$ gives for all (x,y),

$$\begin{aligned}\mathcal{A}^\varepsilon V^p(x,y) &= pV^{p-1}(x,y)\Big(-m\langle x, \nabla U(x)\rangle - (1-m)|y|^2\Big) \\ &\quad + \frac{\varepsilon^2}{2}\mathrm{Tr}\Big[p(p-1)V^{p-2}\nabla_x V \otimes \nabla_x V + pV^{p-1}D_x^2 V\Big],\end{aligned} \tag{B.1}$$

where for $u, v \in \mathbb{R}^d$, $u \otimes v$ is the $d \times d$ matrix defined by $(u\otimes v)_{i,j} = u_i v_j$.

Then, let us prove (3.9) under Assumption $(\mathbf{H_{Q+}})$. Since $m \in (0,1)$, we have

$$-m\langle x, \nabla U(x)\rangle - |y|^2(1-m) \leq m\beta - m\alpha U(x) - (1-m)|y|^2 \leq \beta_1 - \alpha_1 V(x,y),$$

for some constants $\beta_1 \in \mathbb{R}$ and $\alpha_1 > 0$. Moreover, since $D_x^2 V(x,y) = D^2U(x) + mI_d$, $\rho \in (0,1)$ and $\lim_{|(x,y)|\to+\infty} V(x,y) = +\infty$ (see (3.16)), we have

$$\mathrm{Tr}\Big[p(p-1)V^{p-2}\nabla_x V \otimes \nabla_x V + pV^{p-1}D_x^2 V\Big] = o(V^p(x,y)) \quad \text{as } |(x,y)| \to +\infty.$$

It follows that there exists $\beta_2 > 0$ such that for every $\varepsilon \in (0,1]$,

$$\frac{\varepsilon^2}{2}\mathrm{Tr}\Big[p(p-1)V^{p-2}\nabla_x V \otimes \nabla_x V + pV^{p-1}D_x^2 V\Big] \leq \beta_2 + \frac{p\alpha_1}{2}V^p(x,y).$$

Therefore, we get for every $\varepsilon \in (0,1]$

$$\mathcal{A}V^p(x,y) \le p\beta_1 V^{p-1}(x,y) + \beta_2 - \frac{p\alpha_1}{2}V^p(x,y).$$

Using again that $\lim_{|(x,y)|\to+\infty} V(x,y) = +\infty$, we deduce that $p\beta_1 V^{p-1} \le \beta_3 + \frac{p\alpha_1}{4}V^p$ and we deduce there exist some positive $\tilde{\beta}$ and $\tilde{\alpha}$ such that

$$\forall \varepsilon \in (0,1],\ \forall (x,y) \in \mathbb{R}^{2d}, \quad \mathcal{A}V^p(x,y) \le \tilde{\beta} - \tilde{\alpha}V^p(x,y).$$

Let us now consider (3.9) under assumption $(\mathbf{H_{Q-}})$. Here, we fix $p \in (a-1,a)$. Since $m \in (0,1)$, we have that

$$-m\langle x, \nabla U(x)\rangle - |y|^2(1-m) \le m\beta - m\alpha(|x|^2)^a - (1-m)|y|^2 \le \beta_1 - \alpha_1 V^a(x,y) \quad (\beta' \in \mathbb{R}, \alpha' > 0)$$

where in the second inequality, we used the elementary inequalities $u^{2a} \le 1 + u^2$ for $u \ge 0$ and $(u+v)^a \le u^a + v^a$ for $u, v \ge 0$, and the fact that $V(x,y) \le C(1+|x|^2) + |y|^2)$ ($|\nabla U|^2 \le C(1+U)$ implies that $\sqrt{U}$ is sublinear).

Under $(\mathbf{H_{Q-}})$ we also have

$$\sup_{x\in\mathbb{R}^d} \|D_x^2 V(x,y)\| < +\infty,$$

and since $p \in 0,1$ we have

$$\mathrm{Tr}\Big[p(p-1)V^{p-2}\nabla_x V \otimes \nabla_x V + pV^{p-1}D_x^2 V\Big] \quad \le \quad CV^{p-1}(x,y)$$

This way, there exist $\tilde{\alpha} > 0, \tilde{\beta}$ such that for all $\varepsilon \in [0,1]$ and all (x,y), we have

$$\mathcal{A}V^p(x,y) \le \beta - \alpha V^{p+a-1}(x,y)$$

Now we prove inequality (3.10). One can check that

$$\mathcal{A}^\varepsilon \psi_\varepsilon = \frac{\delta}{\varepsilon^2}\psi_\varepsilon \Bigg(- pV^{p-1}\Big(m\langle x, \nabla U(x)\rangle + (1-m)|y|^2\Big) \tag{B.2}$$

$$+ \frac{1}{2}\mathrm{Tr}\left[\varepsilon^2\left(p(p-1)V^{p-2} + \delta p^2 V^{2p-2}\right)\nabla_x V \otimes \nabla_x V + \varepsilon^2 pV^{p-1}D_x^2 V\right]\Bigg). \tag{B.3}$$

We recall that $\nabla_x V = \nabla U + m(x-y)$ and that $D_x^2 V = D^2 U + mI_d$. Thus, using $(\mathbf{H_{Q+}})(ii)$ and $(\mathbf{H_{Q-}})(ii)$, we obtain that when $|(x,y)| \to +\infty$,

$$\text{(B.3)} = \begin{cases} O(1+V^{2p-1}) + o(V^p) & \text{under } (\mathbf{H_{Q+}}) \\ O(1+V^{2p-1}) & \text{under } (\mathbf{H_{Q-}}). \end{cases}$$

Then, since $2p-1 < p$ if $p \in (0,1)$ and that $2p-1 < p+a-1$ if $p < a$, we obtain easily (3.10) by following the lines of the part of the proof. □

References

Alvarez, F. (2000). On the minimizing property of a second order dissipative system in hilbert spaces. *SIAM J. Control Optim.*, *38*, 110–1119.

Azencott, R. (1980). *Large deviations theory and applications* (Vol. 774). Springer-Verlag.

Barles, G. (1994). *Solutions de viscosité des équations de hamilton-jacobi* (Vol. 17). Springer-Verlag.

Benaim, M. (1996). A dynamical system approach to stochastic approximations. *SIAM J. Control Optim.*, *34*(2), 437–472.

Biswas, A., & Borkar, V. S. (2009). Small noise asymptotics for invariant densities for a class of diffusions: a control theoretic view. *J. Math. Anal. Appl.*, *360*(2), 476–484.

Cabot, A. (2009). Asymptotics for a gradient system with memory term. *Proc. of the American Mathematical Society*, *9*, 3013–3024.

Cabot, A., Engler, H., & Gadat, S. (2009a). On the long time behavior of second order differential equations with asymptotically small dissipation. *Trans. of the American Mathematical Society*, *361*, 5983–6017.

Cabot, A., Engler, H., & Gadat, S. (2009b). Second-order differential equations with asymptotically small dissipation and piecewise flat potentials. *Electronic Journal of Differential Equation*, *17*, 33–38.

Chiang, T.-S., Hwang, C.-R., & Sheu, S. J. (1987). Diffusion for global optimization in $\mathbb{R}^n$. *SIAM J. Control Optim.*, *25*(3), 737–753.

Coron, J.-M. (2007). *Control and nonlinearity* (Vol. 136). Providence, RI: American Mathematical Society.

Dembo, A., & Zeitouni, O. (2010). *Large deviations techniques and applications* (Vol. 38). Berlin: Springer-Verlag. (Corrected reprint of the second (1998) edition)

Feng, J., & Kurtz, T. G. (2006). *Large deviations for stochastic processes* (Vol. 131). Providence, RI: American Mathematical Society.

Freidlin, M. I., & Wentzell, A. D. (1979). *Random perturbations of dynamical systems* (Second ed., Vol. 260). New York: Springer-Verlag. (Translated from the 1979 Russian original by Joseph Szücs)

Gadat, S., & Panloup, F. (2011). Long time behaviour and stationary regime of memory gradient diffusions. *submitted*, 1–36.

Has'minskii, R. (1960). Ergodic properties of recurrent diffusion processes and stabilization of the solution to the cauchy problem for parabolic equations. *Theory of probability and its applications*, *5*, 179 –196.

Has'minskii, R. (1980). *Stochastic stability of differential equations.* Sijthoff & Noordhoff, Alphen aan den Rijn (The Nederlands).

Herrmann, S., & Tugaut, J. (2010). Stationary measures for self-stabilizing processes: asymptotic analysis in the small noise limit. *Electron. J. Probab.*, *15*, no. 69, 2087–2116.

Holley, R., & Stroock, D. (1988). Simulated annealing via Sobolev inequalities. *Comm. Math. Phys.*, *115*(4), 553–569.

Holley, R. A., Kusuoka, S., & Stroock, D. W. (1989). Asymptotics of the spectral gap with applications to the theory of simulated annealing. *J. Funct. Anal.*, *83*(2), 333–347.

Miclo, L. (1992a). Recuit simulé sans potentiel sur une variété riemannienne compacte. *Stochastics Stochastics Rep.*, *41*(1-2), 23–56.

Miclo, L. (1992b). Recuit simulé sur $\mathbf{R}^n$. Étude de l'évolution de l'énergie libre. *Annales de l'Institut Henri Poincaré. Probabilités et Statistiques*, *28*(2), 235–266.

Puhalskii, A. (2001). *Large deviations and idempotent probability* (Vol. 119). Chapman & Hall/CRC, Boca Raton, FL.

Puhalskii, A. (2003). On large deviation convergence of invariant measures. *J. Theoret. Probab.*, *16*(3), 689–724.

Raimond, O. (2009). Self-interacting diffusions: a simulated annealing version. *Probab. Theory Related Fields*, *144*(1-2), 247–279.

Stroock, D., & Varadhan, S. R. S. (1970). On the support of diffusion processes with applications to the strong maximum principle. *Symposium of Math. Statist. Prob. III*, 333–359.

Stroock, D., & Varadhan, S. R. S. (1972). On degenerate elliptic-parabolic operators of second order and their associated diffusions. *Comm. Pure Appl. Math.*, *25*, 651–713.