Electronic Journal of Probability

Electron. J. Probab. **18** (2013), no. 81, 1–34.
ISSN: 1083-6489 DOI: 10.1214/EJP.v18-2031

Large deviation principle for invariant distributions of memory gradient diffusions

Sébastien Gadat* Fabien Panloup† Clément Pellegrini‡

Abstract

In this paper, we consider a class of diffusions based on a memory gradient descent, *i.e.* whose drift term is built as the average all along the past of 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. With the 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. In particular, such a result leads to a concentration of the invariant distribution around the global minima of U. Here, except in the linear case, we have no closed formula for the invariant distribution but we prove that a LDP can still be obtained. Then, in the one-dimensional case and under some assumptions on the second derivative of U, we get some bounds for the rate function that lead to the concentration around the global minima.

Keywords: Large Deviation Principle; Hamilton-Jacobi Equations; Freidlin and Wentzell Theory; small stochastic perturbations; hypoelliptic diffusions.
AMS MSC 2010: Primary 60F10 ; 60G10 ; 60J60 ; 60K35 , Secondary 35H10 ; 93D30.
Submitted to EJP on May 17, 2012, final version accepted on August 4, 2013.

1 Introduction

The aim of this paper is to study some small noise asymptotic properties of a diffusive stochastic model with memory gradient. The evolution is given by the following stochastic differential equation (SDE) on $\mathbb{R}^d$:

$$dX_t^\varepsilon = \varepsilon dB_t - \left(\frac{1}{k(t)}\int_0^t k'(s)\nabla U(X_s^\varepsilon)ds\right)dt, \tag{1.1}$$

where $\varepsilon > 0$ and (B_t) is a standard d-dimensional Brownian motion. A special feature of such an equation is the integration over the past of the trajectory depending on a

*Institut de Mathématiques, Université de Toulouse (UMR 5219), 31062 Toulouse, Cedex 9, France.
E-mail: sebastien.gadat@math.univ-toulouse.fr
†Institut de Mathématiques, Université de Toulouse, INSA Toulouse, F-31077 Toulouse Cedex 4, France.
E-mail: fabien.panloup@math.univ-toulouse.fr
‡Institut de Mathématiques, Université de Toulouse (UMR 5219), 31062 Toulouse, Cedex 9, France.
E-mail: clement.pellegrini@math.univ-toulouse.fr

function k which quantifies the amount of memory. Our work is mainly motivated by optimization applications. Indeed, in a recent work [8] have shown that the solution of the deterministic dynamical system ($\varepsilon = 0$) converges to the minima of the potential U. Without memory, that is without integration over the past of the trajectory, the model (1.1) reduces to the classical gradient descent model and such convergence results are well-known. Even in the deterministic framework, a potential interest of the gradient with memory is the capacity of the solution to avoid some local traps of U. Indeed, the solution of (1.1) (when $\varepsilon = 0$) may keep some inertia even when it reaches a local minimum of U. This implies a larger exploration of the space than a classical gradient descent which cannot escape from local minima (see [1] and [7]). Usually such a property is obtained by introducing a small noise term. In the classical case, this leads to the usual following SDE

$$dX_t^\varepsilon = \varepsilon dB_t - \nabla U(X_t^\varepsilon)dt. \tag{1.2}$$

As mentioned above, the behaviour of the invariant distribution of this model when ε goes to 0 is well-known. Using the so-called Laplace method, it can be proved that a Large Deviation Principle (LDP) holds and that the invariant distribution of (1.2) concentrates on the global minima of U when the parameter $\varepsilon \to 0$ (see *e.g.* [14]).

It is then natural to investigate the study of the stochastic memory gradient (1.1) in order to obtain similar results. 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 Markov. This can be overcome with the introduction of an auxiliary process (Y_t^ε) defined by

$$Y_t^\varepsilon = \left(\frac{1}{k(t)}\int_0^t k'(s)\nabla U(X_s^\varepsilon)ds\right). \tag{1.3}$$

In general, the couple $Z_t^\varepsilon = (X_t^\varepsilon, Y_t^\varepsilon)$ gives rise to a non-homogeneous Markov process (see [15]). In order to consider the notion of invariant measure, we concentrate on the case where $k(t) = e^{\lambda t}$ which turns (Z_t^ε) into a homogeneous Markov process. In this context, [15] have shown the existence and uniqueness of the invariant measure ν_ε for (Z_t^ε).

In the present work, our objective is to obtain some sharp estimations of the asymptotic behaviour of (ν_ε) as $\varepsilon \to 0$. More precisely, we shall first show that $(\nu_\varepsilon)_{\varepsilon>0}$ satisfies a Large Deviation Principle. Then, we shall try to obtain some sharp bounds for the associated rate function in order to understand how the invariant probability is distributed as $\varepsilon \to 0$. In particular, we shall establish the concentration around the global minima of U up to technical hypotheses. In the classical setting of (1.2), this is an essential step towards the implementation of the so-called *simulated annealing* strategy. The development of such an optimization procedure for the memory gradient diffusion is certainly a motivation of the study of (1.1). This will be addressed in a forthcoming work.

The paper is also motivated by extending some results of Large Deviations for invariant distributions to a difficult context where the process is not elliptic 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 of the invariant measure are generally impossible. This implies that the works on elliptic Kolmogorov equations by [10], [21] or [18] for instance, cannot be extended to our context. For similar considerations in other non-Markov models, one should also mention the recent works on Mac-Kean Vlasov diffusions by [17] and on self-interacting diffusions with attractive potential by [25].

Here, in order to obtain a LDP for $(\nu_\varepsilon)_{\varepsilon\geq 0}$ we adapt the strategy of [23] and [14] to our degenerated context. We shall first show a finite-time LDP for the underlying stochastic process. Second, we prove the exponential tightness of $(\nu_\varepsilon)_{\varepsilon\geq 0}$ by using Lyapunov type arguments. Finally, we show that the associated rate function, denoted as W in the paper, can be expressed as the solution of a control problem (in an equivalent way to the solution of a Hamilton-Jacobi equation). However, at first sight the solution of the control problem is not unique. This uniqueness property follows from an adaptation of the results of [14] to our framework. In particular, we obtain a formulation of the rate function in terms of the costs to join stable critical points of our dynamical system. Next, the second step of the paper (sharp estimates of W) is investigated by the study of the cost to join stable critical points.

The paper is organized as follows. In Section 2, we recall some results about the long-time behaviour of the diffusion when ε is fixed. Moreover, we provide the main assumptions needed for obtaining the LDP for (ν_ε). 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 (see also the works of [6] and [9] for other adaptations of this theory). Since the study of the cost function is quite hard in a general setting, we focus in Section 5 on the case of a double-well potential U. In this context, we obtain some upper and lower bounds for the associated quasi-potential function. Then, we provide some conditions on U and on the memory parameter λ which allow us to prove the concentration of the invariant distribution around the global minima. Note that, even if our assumptions in this part seem a little restrictive, the proofs of the bounds (especially the lower bound) are obtained by an original (and almost optimal) use of some Lyapunov functions associated with the dynamical system.

Acknowledgments: The authors would like to thank Guy Barles for his hospitality, and are grateful to Guy Barles, Laurent Miclo and Christophe Prieur for helpful discussions and comments. We also thank an anonymous referee for his careful reading and constructive suggestions. Authors wish to thank Marine Feron and Mathilde Rogez for helpful corrections of the text.

2 Setting and Main Results

2.1 Notations and background on Large Deviation theory

We respectively denote the scalar product and the Euclidean norm on $\mathbb{R}^d$ by $\langle\,,\,\rangle$ and $|\,.\,|$. The space of $d\times d$ real-valued matrices is referred as $\mathbb{M}_d(\mathbb{R})$ and we use the notation $\|\,.\,\|$ for the Frobenius norm on $\mathbb{M}_d(\mathbb{R})$.

We denote $\mathbb{H}(\mathbb{R}_+,\mathbb{R}^d)$ 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 $\dot\varphi\in L^{2,\mathrm{loc}}(\mathbb{R}_+,\mathbb{R}^d)$.

For a $\mathcal{C}^2$-function $f:\mathbb{R}^d\to\mathbb{R}$, ∇f and $D^2 f$ stand respectively for the gradient of f and the Hessian matrix of f. For a function $f\in\mathcal{C}^2(\mathbb{R}^d\times\mathbb{R}^d,\mathbb{R})$, we denote $\nabla_x f:\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{R}^d$ and $D^2_x f:\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{M}_d(\mathbb{R})$ the functions respectively defined by $(\nabla_x f(x,y))_i=\partial_{x_i}f(x,y)$ and $(D^2_x f(x,y))_{i,j}=\partial_{x_i}\partial_{x_j}f(x,y)$. These notations are naturally extended to $\nabla_y f$, $D^2_{x,y}f$ and $D^2_y f$. Finally, for any vector $v\in\mathbb{R}^d$, v^t 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 basic definitions Large Deviation Theory (see [12] for further references). Let (E,d) denote a metric space. A family of probability measures $(\nu_\varepsilon)_{\varepsilon>0}$ on E satisfies a Large Deviation Principle (shortened as LDP) with speed r_ε and rate function I if for any open set O and any 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 referred 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 classical compactness results of Large Deviation theory. A family of probability measures $(\nu_\varepsilon)_{\varepsilon>0}$ is said to be *exponentially tight of order* r_ε if for any $a > 0$, there exists a compact subset K_a of E such that

$$\limsup_{\varepsilon\to 0} r_\varepsilon \log(\nu_\varepsilon(K_a^c)) \leq -a.$$

We recall the following consequence of exponential tightness (see [13]).

Proposition 2.1. *Let (S, d) be a Polish space. Suppose that $(\nu_\varepsilon)_{\varepsilon\geq 0}$ is 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 $\varepsilon_k \to 0$ and $(\nu_{\epsilon_k})_{k\geq 0}$ satisfies a LDP with good rate function I and speed r_{ε_k}.*

Definition 2.2. *Such a subsequence $(\nu_{\varepsilon_k})_{k\geq 1}$ will be called a (LD)-convergent subsequence.*

2.2 Averaged gradient diffusions

As announced in Introduction with $k(t) = e^{\lambda t}$, we are interested in the stochastic evolution of

$$dX_t^\varepsilon = \varepsilon dB_t - \left(\lambda e^{-\lambda t}\int_0^t e^{\lambda s}\nabla U(X_s^\varepsilon)ds\right)dt,$$

where $\lambda > 0$, $(B_t)_{t\geq 0}$ is a standard d-dimensional Brownian motion and $U : \mathbb{R}^d \to \mathbb{R}$ is a smooth, positive and coercive function (see Subsection 2.3 for detailed assumptions).

As announced in the Introduction, since $(X_t^\varepsilon)_{t\geq 0}$ is not Markov, we introduce the auxiliary process $(Y_t^\varepsilon)_{t\geq 0}$

$$Y_t^\varepsilon = \lambda e^{-\lambda t}\int_0^t e^{\lambda s}\nabla U(X_s^\varepsilon)ds.$$

Then, the process $(Z_t^\varepsilon)_{t\geq 0} := ((X_t^\varepsilon, Y_t^\varepsilon))_{t\geq 0}$ is Markov and satisfies:

$$\begin{cases} dX_t^\varepsilon = \varepsilon dB_t - Y_t^\varepsilon dt, \\ dY_t^\varepsilon = \lambda(\nabla U(X_t^\varepsilon) - Y_t^\varepsilon)dt. \end{cases} \tag{2.1}$$

When necessary, we will denote by $(Z_t^{\varepsilon,z})_{t\geq 0}$ the solution starting from $z \in \mathbb{R}^d$ and by $\mathbb{P}_z^\varepsilon$ the distribution of this process on $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^d)$. In the sequel, we will also intensively use the deterministic system obtained when $\varepsilon = 0$ in (2.1). This deterministic system $(z(t))_{t\geq 0} := (x(t), y(t))_{t\geq 0}$ is described as follows

$$\dot{z}(t) = b(z(t)) \quad \text{with} \quad b(x, y) = \begin{pmatrix} 0 & -y \\ \lambda\nabla U(x) & -\lambda y \end{pmatrix}. \tag{2.2}$$

2.3 Assumptions

Throughout this paper, we assume that $U : \mathbb{R}^d \mapsto \mathbb{R}$ is a smooth (at least $\mathcal{C}^2$)-function on $\mathbb{R}^d$ such that

$$\inf_{x\in\mathbb{R}} U(x) > 0, \quad \lim_{|x|\to+\infty} U(x) = +\infty \quad \text{and} \quad \liminf_{|x|\to+\infty} \langle x, \nabla U(x)\rangle > 0. \tag{2.3}$$

Note that we do not suppose that ∇U is a sublinear function. In particular, D^2U is not necessarily bounded. However, in order to ensure the non-explosion (in finite horizon) of $(Z_t^\varepsilon)_{t\geq 0}$ (see Proposition 2.1 of [15]), we assume that there exists $C>0$ such that for all $x\in\mathbb{R}^d$, $\|D^2U(x)\|\leq CU(x)$. Then, since the function ∇U is locally Lipschitz continuous, strong existence and uniqueness hold for the solution of (2.1). In this case, $(Z_t^\varepsilon)_{t\geq 0}$ is a Markov process and we denote $(P_t^\varepsilon)_{t\geq 0}$ its semi-group. Its infinitesimal generator $\mathcal{A}^\varepsilon$ is defined by

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

We first recall some results obtained by [15] on the existence and uniqueness of the invariant distribution of (2.1). To this end, we need to introduce a mean-reverting assumption denoted by $(\mathbf{H_{mr}})$ and a hypoellipticity assumption $(\mathbf{H_{Hypo}})$. The mean reverting assumption is expressed as follows:

$(\mathbf{H_{mr}})$: $\quad \lim_{|x|\to+\infty}\langle x,\nabla U(x)\rangle = +\infty$ and $\lim_{|x|\to+\infty}\frac{\|D^2U(x)\|}{\langle x,\nabla U(x)\rangle}=0.$

Regarding the second assumption, let us define $\mathcal{E}_U$ by

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

and denote $\mathcal{M}_U := \mathbb{R}^d\setminus\mathcal{E}_U$. Assumption $(\mathbf{H_{Hypo}})$ is expressed as follows:

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

In particular, the smoothness of U and the fact that $\dim(\mathcal{M}_U)\leq d-1$ ensure that the Hörmander condition is satisfied on a sufficiently large subspace of $\mathbb{R}^{2d}$. The fact that $\lim|x|^{-1}U(x)=+\infty$ as $|x|\to+\infty$ is needed for the *topological irreducibility* of the semi-group (see [15] for details). These assumptions imply the uniqueness of the invariant distribution. We deduce the following proposition from Theorems 2.3 and 3.2 of [15]:

Proposition 2.3. *If U satisfies $(\mathbf{H_{mr}})$, then for any $\varepsilon>0$, the solution of (2.1) admits an invariant distribution. Furthermore, if $(\mathbf{H_{Hypo}})$ holds, the invariant distribution is unique and admits a λ_{2d}-a.s. positive density. We denote by ν_ε this invariant distribution.*

Note that $(\mathbf{H_{mr}})$ implies Assumption $(\mathbf{H_1})$ of [15] in the particular case $\sigma=I_d$ and $r_\infty=\lambda$.

Our goal is now to obtain a Large Deviation Principle for $(\nu_\varepsilon)_{\varepsilon>0}$ when $\varepsilon\to 0$. To this end, we need a more constraining mean-reverting assumption:

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

$$(i)\quad -\langle x,\nabla U(x)\rangle\leq\beta-\alpha U(x),\forall x\in\mathbb{R}^d$$

$$(ii)\qquad |\nabla U|^2\leq C(1+U^{2(1-\rho)})\quad\text{and}\quad \lim_{|x|\to+\infty}\frac{\|D^2U(x)\|}{U(x)}=0.$$

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

$$(i)\quad -\langle x,\nabla U(x)\rangle\leq\beta-\alpha|x|^{2a},\forall x\in\mathbb{R}^d$$

$$(ii)\quad |\nabla U|^2\leq C(1+U)\quad\text{and}\quad \sup_{x\in\mathbb{R}^d}\|D^2U(x)\|<+\infty.$$

Remark 1. *Assumptions $(\mathbf{H_{Q+}})$ and $(\mathbf{H_{Q-}})$ correspond respectively to super-quadratic and subquadratic potentials. For instance, suppose that $U(x)=(1+|x|^2)^p$. In the case $p\geq 1$, a moment's thought shows that $(\mathbf{H_{Q+}})$ holds with $\rho\in(0,\frac{1}{2p})$. If $p\in(1/2,1]$, $(\mathbf{H_{Q-}})$ holds with $a=p$. These assumptions are adapted to a large class of potentials U with polynomial growth (more than linear). However, they do not cover the case of potentials with exponential growth ($(\mathbf{H_{Q+}})(ii)$ is no longer fulfilled).*

2.4 Main results

2.4.1 Exponential tightness and Hamilton Jacobi equation

In this section, we provide our first Large Deviation results for $(\nu_\varepsilon)_{\varepsilon>0}$: the exponential tightness and a first characterization of the rate functions related to the LD-convergent subsequences (that we will sometimes call *LD-limits* in the sequel). To this end, we need to introduce some notations for the *controlled trajectories* related to the dynamical system (and to their time-reversed counterparts): for any function $\varphi \in \mathbb{H}$, we denote by $\mathbf{z}_\varphi := (\mathbf{z}_\varphi(t))_{t\geq 0}$ and by $\tilde{\mathbf{z}}_\varphi := (\tilde{\mathbf{z}}_\varphi(t))_{t\geq 0}$, the solutions of

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

For each $z \in \mathbb{R}^2$, we denote by $\mathbf{z}_\varphi(z,.)$ and $\tilde{\mathbf{z}}_\varphi(z,.)$ the solutions starting from z. Note that $(\mathbf{H}_{\mathbf{Q}+})$ and $(\mathbf{H}_{\mathbf{Q}-})$ ensure the finite-time non-explosion of $\mathbf{z}_\varphi$ and $\tilde{\mathbf{z}}_\varphi$ for all $\varphi \in \mathbb{H}$ (see $e.g.$ Equation (3.4)). Hence, since ∇U is locally Lipschitz continuous, these solutions exist and are uniquely determined.
Finally, with the view to the characterization of the LD-limits, we introduce the following assumption:

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

We can now state our first main result.

Theorem 1. *Suppose that $(\mathbf{H}_{\mathbf{Hypo}})$ holds and that either $(\mathbf{H}_{\mathbf{Q}+})$ or $(\mathbf{H}_{\mathbf{Q}-})$ is satisfied. Then,*

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

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

$$\forall t \geq 0, \quad W(z) = \inf_{\varphi\in\mathbb{H}} \left[\frac{1}{2}\int_0^t |\dot{\varphi}(s)|^2 ds + W(\tilde{\mathbf{z}}_\varphi(z,t)) \right]. \tag{2.7}$$

(iii) Furthermore, assume that $(\mathbf{H}_\mathbf{D})$ is fulfilled. Then,

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

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 $i = 1,\ldots,l$.

Equation (2.7) satisfied by W may be seen as an Hamilton-Jacobi equation (see $e.g.$ [3] 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 of W is only available conditionally to the values of $W(z_i^\star)$, $i = 1,\ldots\ell$. Thus, in order to obtain a LDP, we must establish that the values of $W(z_i^\star)$ are in fact uniquely determined. Such a result is obtained following the [14] approach. We first give some useful elements of Freidlin and Wentzell theory.

$\{i\}$-Graphs Following the notations of Theorem 1, we denote the finite set of equilibria by $\{z_1^\star,\dots,z_\ell^\star\}$. We recall that for any $i\in\{1,\dots,\ell\}$, an oriented graph on $\{z_1^\star,\dots,z_\ell^\star\}$ is called an $\{i\}$-Graph if it 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 edges starting at state $z_j^\star$ and leading to the state $z_i^\star$.

L^2 control cost between equilibria For any $(\xi_1,\xi_2)\in(\mathbb{R}^d\times\mathbb{R}^d)^2$, the minimal L^2 cost to join ξ_2 from ξ_1 in a *finite time t* is

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

Then, we denote by I the function given by:

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

The function I is usually called the *quasipotential* and yields the following representation of $W(z_i^\star)$, $i=1,\dots,\ell$:

Theorem 2. *Assume that $(\mathbf{H_{Hypo}})$ and $(\mathbf{H_D})$ hold, and that either $(\mathbf{H_{Q+}})$ or $(\mathbf{H_{Q-}})$ is satisfied. For any (LD)-convergent subsequence $(\nu_{\varepsilon_n})_{n\geq1}$, the associated rate function W 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

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

The next corollary follows immediately from Theorem 1 and Theorem 2.

Corollary 1. *Assume that $(\mathbf{H_{Hypo}})$ and $(\mathbf{H_D})$ hold and that either $(\mathbf{H_{Q+}})$ or $(\mathbf{H_{Q-}})$ is satisfied. Then, (ν_ε) satisfies a large deviation principle with speed ε^{-2} and good rate function W such that*

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

where $\mathcal{W}(z_i^\star)$ is given by (2.9).

Case of a double-well potential In the sequel, we are interested by the location of the global minima of W. More precisely, we expect the first coordinate of this minimum to be located on the set of global minima of U. Using Equation (2.8), this point is clear when U is a strictly convex potential. Regarding now the non-convex case, the situation is more involved. 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 $x_1^\star$ and $x_2^\star$. We obtain the following result:

Theorem 3. *Suppose the hypothesis of Corollary 1 hold and that U satisfies* (2.10)*, then,*
(i) $\mathcal{W}$ *satisfies*

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

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

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

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

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

In particular, $(\nu_\varepsilon)_{\varepsilon\ge 0}$ weakly converges towards $\delta_{z_1^\star}$ as $\varepsilon \to 0$.

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 except in Section 5, we will prove these 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 is devoted to the proof of Theorem 1. In Subsection (3.2), we focus on the exponential tightness of the invariant measures $(\nu_\varepsilon)_{\varepsilon\in(0,1]}$. The proof of this property is based on two main ingredients: the *finite-time* LDP for $(Z^\epsilon)_{\epsilon>0}$ stated below and a uniform control of the return times to compact sets obtained with some Lyapunov-type arguments. Then, in Subsections (3.3) and (3.4), we obtain successively the representations of the LD-limits of $(\nu_\varepsilon)_{\varepsilon\in(0,1]}$ given by (2.7) and (2.8).

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

In the next lemma, we provide a LDP for $((Z_t^\varepsilon)_{t\ge 0})_{\varepsilon>0}$ on $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^{2d})$ (space of continuous functions from $\mathbb{R}_+$ to $\mathbb{R}^{2d}$). In the sequel, when we refer to this property, we will call it the "finite-time interval LDP for $(Z^\varepsilon)_{\epsilon>0}$".

Lemma 3.1. *Assume that $(\mathbf{H_{Q+}})$ or $(\mathbf{H_{Q-}})$ is satisfied. Let $z \in \mathbb{R}^{2d}$ and $(z_\varepsilon)_{\varepsilon>0}$ be a net of $\mathbb{R}^{2d}$ such that $z_\varepsilon \xrightarrow{\varepsilon\to 0} z$. Then, $(Z^{\varepsilon,z_\varepsilon})_{\varepsilon>0}$ satisfies a LDP on $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^{2d})$ (endowed with the topology of uniform convergence on compact sets) with speed ε^{-2}. The corresponding (good) rate function $\mathcal{I}_z$ is defined for all absolutely continuous $(\mathbf{z}(t))_{t\ge 0} = (\mathbf{x}(t), \mathbf{y}(t))_{t\ge 0}$ by*

$$\mathcal{I}_z((\mathbf{z}(t))_{t\ge 0}) = \inf_{\varphi\in\mathbb{H}, \mathbf{z}_\varphi(z,.)=\mathbf{z}(.)} \frac{1}{2}\int_0^\infty |\dot\varphi(s)|^2 ds = \frac{1}{2}\int_0^\infty |\dot{\mathbf{x}}(s) + \mathbf{y}(s)|^2\, ds,$$

where $\mathbf{z}_\varphi(z,.) = \mathbf{z}(.)$ means that $\mathbf{z}_\varphi(z,t) = \mathbf{z}(t)$, for all $t \ge 0$,. In particular, for all $t \ge 0$, $(P_t^\varepsilon(z_\varepsilon,.))_{\varepsilon>0}$ satisfies a LDP with speed ε^{-2}. The corresponding rate function $I_t(z,.)$ is defined for all $z, z' \in \mathbb{R}^{2d}$ 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 $\mathbf{z}(0) = z$, $\mathbf{z}(t) = z'$. Consequently, the function I_t can be written as

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

Remark 2. *Note that such a result is quite classical when $z_\varepsilon = z$ and when the coefficients are Lipschitz continuous functions (see e.g. [2] for instance). Here, we have to handle the possibly super-linear growth of the drift vector field b (and also the degeneracy of the diffusion).*

Proof. We wish to apply Theorem 5.2.12 of [22]. For this purpose, we need to prove the following four points:

• Uniqueness for the *maxingale problem*: This step is an identification of the (potential) LD-limits of $(Z^\varepsilon)_{\varepsilon>0}$. More precisely, we need to prove that the *idempotent probability* $\pi_z(.) := \exp(-\mathcal{I}_z(.))$ is the unique solution to the *maxingale problem* (z, G) where G : $\mathbb{R}^{2d} \times \mathcal{C}(\mathbb{R}_+, \mathbb{R}^{2d}) \to \mathcal{C}(\mathbb{R}_+, \mathbb{R}^{2d})$ is given by:

$$\forall\, \lambda = (\lambda_1, \lambda_2) \in \mathbb{R}^d \times \mathbb{R}^d,\ \forall \mathbf{z} \in \mathcal{C}(\mathbb{R}_+, \mathbb{R}^{2d}),\ \forall t \geq 0, \quad G_t(\lambda, \mathbf{z}) = \int_0^t b(\mathbf{z}(s))ds + \frac{1}{2}\lambda_1^2.$$

The fact that π_z solves the maxingale problem follows from Theorem 3.1 and Lemma 3.2 of [24]. Setting $\mathcal{E}(x,y) = U(x) + \dfrac{|y|^2}{2}$, note that Lemma 3.2 can be applied since $\langle \nabla \mathcal{E}(x,y), b(x,y) \rangle \leq 0$ (see condition (3.6a) of [24]). Furthermore, since b is locally Lipschitz continuous, for any $\varphi \in \mathbb{H}$, the ordinary differential equation

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

has a unique solution. Thus, uniqueness for the maxingale problem is a consequence of the second point of Lemma 2.6.17 of [22] and of Theorem 3.1 of [24].

• Continuity condition for the characteristics of the diffusion: Since the diffusive component is constant, we only have to focus on the drift component. We need to show that for all $t \geq 0$ 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 $\mathbf{z}$. Since b is Lipschitz continuous on every compact set of $\mathbb{R}^{2d}$, this point is obvious.

• Local majoration condition: In this step, we have to check that for all $M > 0$, there exists an increasing continuous map $\bar{F}^M : \mathbb{R}_+ \to \mathbb{R}$ such that

$$\forall s \in [0,t] \qquad \sup_{\mathbf{z}\in\mathcal{C}(\mathbb{R}_+,\mathbb{R}^{2d}),\|\mathbf{z}\|_\infty \leq M} (\phi_t(\mathbf{z}) - \phi_s(\mathbf{z})) \leq \bar{F}^M(t) - \bar{F}^M(s),$$

with $\|\mathbf{z}\|_\infty = \sup_{t\geq 0} |\mathbf{z}(t)|$. Since b is locally bounded, the previous inequality holds with

$$\bar{F}^M(t) := \sup_{z\in\mathbb{R}^{2d}, |z|\leq M} |b(z)|t.$$

• Non-Explosion condition (NE): The Non-Explosion condition holds if the following two points are satisfied

(i) The function π_z defined by $\pi_z : \mathbf{z} \longmapsto \exp(-\mathcal{I}_z(\mathbf{z}))$ is upper-compact,

(ii) For all $t \geq 0$ and for all $a \in (0,1]$, the set $\displaystyle\bigcup_{s\leq t} \left\{ \sup_{u\leq s} |\mathbf{z}(u)|,\ \pi_{z,s}(\mathbf{z}) \geq a \right\}$ is bounded where

$$\forall\, z \in \mathbb{R}^{2d},\ \forall\, t \geq 0, \quad \pi_{z,t}(\mathbf{z}) = \exp\left(- \inf_{\varphi\in\mathbb{H}, \mathbf{z}_\varphi(z,.)=\mathbf{z}(.)} \frac{1}{2}\int_0^t |\dot\varphi(s)|^2 ds \right).$$

Point (i): The property that π_z is upper-compact means that for all $a \in (0,1]$, 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). In order to show that, we use 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 $\varphi \in \mathbb{H}$ such that $\mathbf{z} = \mathbf{z}_\varphi$ and

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

Using the previously defined function $\mathcal{E}$, one checks that for every $p > 0$,

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

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

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

By the Gronwall Lemma, it follows that

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

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

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

Now, let us prove that K_a is equicontinuous: for all $t > 0$, $u, v \in [0,t]$ with $u \leq v$ and $\mathbf{z} \in K_a$, we know that for a suitable constant $\tilde C_{t,a,z}$, the controlled trajectories of K_a are *a priori* bounded: $\sup_{s\in[0,t]}|\mathbf{z}(s)| \leq \tilde C_{t,a,z}$. Since b is continuous, the Cauchy-Schwarz Inequality yields:

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

The two conditions of the Ascoli Theorem being satisfied, the compactness of K_a follows.

Point (ii): We do not detail this item which easily follows from the controls established in the proof of (i) (see (3.4)). Finally, the other conditions of Theorem 5.2.12 of [22] being trivially satisfied, the lemma follows.

□

3.2 Exponential tightness (Proof of i) of Theorem 1)

In the next proposition, we investigate the exponential tightness of $(\nu_\varepsilon)_{\varepsilon\in(0,1]}$. Our approach consists in providing sufficiently sharp estimates for hitting times of the process $(Z_t^\varepsilon)_{t\geq 0}$.

Proposition 3.2. *Suppose that* $(\mathbf{H_{Q+}})$ *or* $(\mathbf{H_{Q-}})$ *holds. Then, there exists a compact set* B *of* $\mathbb{R}^{2d}$*, such that the first hitting time* τ_ε *of* B $\tau_\varepsilon = \inf\{t > 0, Z_t^\varepsilon \in B\}$ *satisfies the three properties:*

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

$$\limsup_{\varepsilon\to 0}\sup_{z\in K}\mathbb{E}_z[(\tau_\varepsilon)^2]<\infty. \tag{3.6}$$

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

$$\limsup_{\varepsilon\to 0}\sup_{z\in K}\sup_{t\geq 0}\mathbb{E}_z\left[|Z^\varepsilon_{t\wedge\tau_\varepsilon}|^{\frac{\delta}{\varepsilon^2}}\right]^{\varepsilon^2}<+\infty. \tag{3.7}$$

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

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

Proposition 3.2 and Lemma 7 of [23] then imply the following corollary.

Corollary 2. *The family of invariant distributions $(\nu^\varepsilon)_{\varepsilon\in(0,1]}$ is exponentially tight.*

A fundamental step of the proof of Proposition 3.2 is the next lemma in which we establish 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 a lemma uses a key Lyapunov function V which is rather not standard due to the kinetic form of the coupled process.

Lemma 3.3. *Assume that $(\mathbf{H_{Q+}})$ or $(\mathbf{H_{Q-}})$ is satisfied 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 any $p>0$, $\delta>0$ and $\varepsilon>0$, we set

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

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

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

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

where $\mathcal{A}^\varepsilon$ is the infinitesimal generator of $(X^\varepsilon_t,Y^\varepsilon_t)$ defined in (2.4) and where

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

Proof of Proposition 3.2. For sake of simplicity, we omit the ε-dependence and write (X_t,Y_t) instead of $(X^\varepsilon_t,Y^\varepsilon_t)$.

• Proof of (i): We use a Lyapunov method to bound the second moment of the hitting time τ_ε. 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

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

Since V is a positive function, we have

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

Note that in the previous 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 l.h.s. of (3.13) satisfies

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

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

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

Let us prove the above inequality under condition $(\mathbf{H_{Q+}})$ or $(\mathbf{H_{Q-}})$. First, since $m\in(0,1)$, one can check that there exists $C>0$ such that

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

As a consequence, we have

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

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

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

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

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

where $\bar{p}$ is defined in Lemma 3.3. Under (3.17), one checks that $2p-1<\bar{p}$. Thus, uniformly in ε,

$$\lim_{|(x,y)|\to+\infty} H_{p,\varepsilon}(x,y) = +\infty,$$

and (3.14) follows.

Now, let M_1 be such that (3.14) holds and define τ_ε by $\tau_\varepsilon = \inf\{t \geq 0, Z_t^\varepsilon \in \bar{B}(0, M_1)\}$. We then have:

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

Computing the integral and using the Fatou Lemma, we get

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

The Jensen Inequality applied to $x \to x^{\frac{1}{\varepsilon^2}}$ yields that

$$\forall (x,y) \in \mathbb{R}^{2d}, \forall \varepsilon \in (0,1], \quad \mathbb{E}_{(x,y)}[(1+\tau_\varepsilon)^2] \leq \exp\left(V^p(x,y)\right).$$

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

• Proof of (ii): Thanks to (3.15), we have for all $p > 0$ and for large enough $|(x,y)|$,

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

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 the Itô formula to $\psi_\varepsilon(x,y) := \exp(\delta V^p(x,y)/\varepsilon^2)$, we get:

$$\forall t \geq 0, \quad \psi_\varepsilon(X_t, Y_t) = \psi_\varepsilon(x,y) + \int_0^t \mathcal{A}\psi_\varepsilon(X_s, Y_s)ds + M_t, \tag{3.19}$$

where $(M_t)_{t\geq 0}$ is a local martingale that we do not need to make explicit. Let us choose $p \in (0,1)$ such that inequality (3.10) of Lemma 3.3 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 any real positive number δ, there exists $M_2 > 0$ such that

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

Let $\tau_\varepsilon = \inf\{t \geq 0, (X_t, Y_t) \in \bar{B}(0, M_2)\}$. A standard localization argument in (3.19) yields

$$\forall\, (x,y) \in \mathbb{R}^{2d}, \quad \mathbb{E}_{(x,y)}[\psi_\varepsilon(X_{t\wedge\tau_\varepsilon}, Y_{t\wedge\tau_\varepsilon})] \leq \psi_\varepsilon(x,y).$$

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

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

From the above inequality, we finally deduce (3.7).

• Proof of (iii): With the notations of the two previous parts of the proof, the properties (3.6) and (3.7) hold with $\tau_\varepsilon := \inf\{t \geq 0, (X_t, Y_t) \in B\}$ for each 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 $\varepsilon_n \to 0$, $z_n \in K$ for all $n \geq 1$ and

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

Up to an extraction, we can assume that $(z_n)_{n\geq 1}$ is a convergent sequence. Let $\tilde{z}$ denote its limit. Lemma 3.1 implies that $(\mathcal{L}((Z^{\varepsilon_n,z_n})_{t\in[0,1]})_{n\geq 1}$ is exponentially tight, and then tight on $\mathcal{C}([0,1],\mathbb{R}^d)$. Using a second extraction, we can assume that $(Z^{\varepsilon_n,z_n})_{n\geq 1}$ converges in distribution to $Z^{(\infty)}$. Furthermore, since $\varepsilon_n \to 0$, the limit process $Z^{(\infty)}$ is $a.s.$ a solution of the o.d.e. $\dot{\mathbf{z}} = b(\mathbf{z})$ starting at $\tilde{z}$. Since the function b is locally Lipschitz continuous, the 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 $\mathbf{z}(\tilde{z},.)$ (here $\mathbf{z}(\tilde{z},.)$ denotes the unique solution of $\dot{\mathbf{z}} = b(\mathbf{z})$ starting from $\tilde{z}$). Since the function $\mathbf{z}(\tilde{z},.)$ is 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 $n \geq n_0$,

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

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 $n \geq n_0$,

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

Using the Fatou 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 $d(K,\bar{B}(0,M+\frac{\delta}{2})) > 0$, the stopping time $\tau_{\tilde{z},\frac{\delta}{2}}$ is clearly positive. The result follows.

□

3.3 Hamilton-Jacobi equation (Proof of $ii)$ of Theorem 1)

This point is a consequence of the finite time large deviation principle for $(Z^\varepsilon)_{\varepsilon\geq 0}$ (Lemma 3.1) and of the exponential tightness of $(\nu_\varepsilon)_{\varepsilon\geq 0}$ (Proposition 3.2). This is the purpose of the next proposition which is an adaptation of Corollary 1 of [23].

Proposition 3.4. *For all $\varepsilon > 0$, let $(P_t^\varepsilon(z,.))_{t\geq 0, z\in\mathbb{R}^{2d}}$ be the semi-group associated to* (2.1) *whose unique invariant distribution is denoted by ν_ε. Suppose that the following assumptions hold:*

(i) $(\nu_\varepsilon)_{\varepsilon>0}$ is exponentially tight of order ε^{-2} on $\mathbb{R}^{2d}$.

(ii) For all $t \geq 0$ and $z \in \mathbb{R}^{2d}$, there exists a function $I_t(z,.) : \mathbb{R}^{2d} \to \mathbb{R}$, such that for all $(z_\varepsilon)_{\varepsilon>0}$ $z_\varepsilon \to z$ as $\varepsilon \to 0$ and $P_t^\varepsilon(z_\varepsilon,.)$ satisfy a LDP with speed ε^{-2} and rate function $I_t(z,.)$.

Then, $(\nu_\varepsilon)_{\varepsilon>0}$ admits a (LD)-convergent subsequence. For such a subsequence $(\nu_{\varepsilon_k})_{k\geq 0}$ (with $\varepsilon_k \to 0$ as $k \to +\infty$), the associated rate function W satisfies for all $z_0 \in \mathbb{R}^{2d}$,

$$\forall t \geq 0, \qquad W(z_0) = \inf_{z\in\mathbb{R}^{2d}} \left(I_t(z,z_0) + W(z)\right). \tag{3.20}$$

With the terminology of [23], Equation (3.20) means that $\tilde{W}$ defined for all $\Gamma \in \mathcal{B}(\mathbb{R}^{2d})$ by $\tilde{W}(\Gamma) = \sup_{y\in\Gamma} \exp(-W(y))$ is an invariant deviability for $(P_t^\varepsilon(z,.))_{t\geq 0, z\in\mathbb{R}^{2d}}$. In Corollary 1 of [23], this result is stated with a uniqueness assumption on the invariant deviabilities. The above proposition is in fact an extension of this corollary to the case where uniqueness is not fulfilled. We refer to Appendix A for details.

Owing to Proposition 3.2 and to Lemma 3.1, Proposition 3.4 can be applied with $I_t(z,.)$ defined by (3.1). Thus, in order to prove *ii)* of Theorem 1, it remains to check that a solution W of (3.20) (with $I_t(z,.)$ defined by (3.1)) also satisfies Equation (2.7). This is the purpose of the next proposition.

Proposition 3.5. *Assume that either* $(\mathbf{H_{Q+}})$ *or* $(\mathbf{H_{Q-}})$ *is fulfilled. Then, for any (LD)-convergent subsequence* $(\varepsilon_n)_{n\geq 1}$ *and (good) rate function* W*:*

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

Proof. We know that W satisfies (3.20). Thus, for any $z_0 \in \mathbb{R}^{2d}$

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

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

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

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

3.4 Infinite horizon Hamilton-Jacobi equation

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 Theorem 4 of [5]. The main novelty of our proof is the second step. Indeed, using arguments based on asymptotic pseudo-trajectories and Lyapunov functions, we prove that the optimal controlled trajectory is attracted by a critical point of the drift vector field.

Proof of (iii) *of Theorem 1.* The proof is divided in three parts. We first build an optimal path $t \mapsto \tilde{z}_\psi(z,t)$ for the Hamilton Jacobi equation of interest. Then, we focus in the second step on its long time behaviour and obtain that $\tilde{z}_\psi(z,t)$ converges to $z^\star$ which belongs to $\{z, b(z) = 0\}$. In order to conclude, we need to prove the continuity of W at each point of $\{z, b(z) = 0\}$. This is the purpose of the third step.
• Step 1: We show that we can build a function $\psi \in \mathbb{H}$ such that for all $z \in \mathbb{R}^{2d}$ the couple $(\tilde{\mathbf{z}}_\psi(z,t), \dot{\psi}(t))_{t\geq 0}$ on $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^{2d}) \times L^{2,loc}(\mathbb{R}_+, \mathbb{R}^d)$ satisfies for all $t > 0$,

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

First, let $T>0$ 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

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

Since W is non negative, it is clear that $(\int_0^T|\dot{\varphi}^{(n)}(s)|^2ds)_{n\geq 0}$ is bounded. It follows 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. This also implies that

$$M:=\sup_{n\geq 1}\sup_{t\in[0,T]}|\tilde{\mathbf{z}}^{(n)}_{\varphi^{(n)}}(t)|<+\infty. \tag{3.22}$$

Actually, under $(\mathbf{H_{Q-}})$, b is Lipschitz continuous and this point is classical. Now, under $(\mathbf{H_{Q+}})$, we have $|\nabla U|=O(U^{1-\rho})$ with $\rho\in(0,1)$. Since $\sup_{n\geq 1}\int_0^T|\dot{\varphi}^{(n)}(s)|^2ds<+\infty$, Inequality (3.4) implies that

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

Since $\lim_{|z|\to+\infty}\mathcal{E}(z)=+\infty$, Equation (3.22) also follows in this case.

Now, since b is locally Lipschitz, b is then Lipschitz continuous on $B(0,M)$ and 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 then there exists a convergent subsequence to $(\hat{\mathbf{z}}^T,\dot{\psi}_{_T})$ 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 $\hat{\mathbf{z}}^T(t)=\tilde{\mathbf{z}}_{\psi_{_T}}(z,t)$, for all $t\in[0,T]$ and $(\hat{z}^T,\dot{\psi}_{_T})$ satisfies (3.21) (for a fixed T). Furthermore, for all $t\in[0,T]$, we have

$$W(\tilde{\mathbf{z}}_{\psi_{_T}}(z,t))=\frac{1}{2}\int_t^T|\dot{\psi}_{_T}(s)|^2ds+W(\tilde{\mathbf{z}}_{\psi_{_T}}(z,T)), \tag{3.23}$$

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

• Step 2: Dropping the initial condition z, we show that $(\tilde{\mathbf{z}}_{\psi}(t+.))_{t\geq 0}$ converges as $t\to+\infty$ to a stationary solution of $\dot{\mathbf{z}}=-b(\mathbf{z})$. First, as in (3.23),

$$W(\tilde{\mathbf{z}}_{\psi}(t+s))-W(\tilde{\mathbf{z}}_{\psi}(t))=-\frac{1}{2}\int_t^{t+s}|\dot{\psi}(u)|^2du. \tag{3.24}$$

It follows that $(W(\tilde{\mathbf{z}}_{\psi}(t)))_{t\geq 0}$ is a non-increasing and thus bounded function. Since W is a good rate function, the quantity $W^{-1}([0,M])$ is a compact subset of $\mathbb{R}^{2d}$, for all $M>0$. This means that $(\tilde{\mathbf{z}}_{\psi}(t))_{t\geq 0}$ is bounded. From (3.24), we deduce that $(\int_t^{t+T}|\dot{\psi}(s)|^2ds)_{t\geq 0}$ is also bounded. The argument described in Step 1 can be used again since b is locally Lipschitz continuous. We deduce from the Ascoli Theorem that $(\tilde{\mathbf{z}}_{\psi}(t+.))$ is relatively compact (for the topology of uniform convergence on compact sets).

We denote now by $\tilde{\mathbf{z}}^{\infty}_{\psi}(.)$ the limit of a convergent subsequence. Let us show that $(\tilde{\mathbf{z}}^{\infty}_{\psi}(t))_{t\geq 0}$ is a solution of $\dot{\mathbf{z}}=-b(\mathbf{z})$. First, since $(W(\tilde{\mathbf{z}}_{\psi}(t)))_{t\geq 0}$ is non-increasing (and non negative as a rate function), we deduce from (3.24) that $\int_t^{t+T}|\dot{\psi}(u)|^2du\xrightarrow{t\to+\infty}0$. As a consequence, using that for any $s\geq 0$, 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 for all $t\geq 0$, $T\geq 0$ and $s\in[0,T]$,

$$\left|\tilde{\mathbf{z}}_{\psi}(t+s)-\tilde{\mathbf{z}}_{\psi}(t)+\int_0^s b(\tilde{\mathbf{z}}_{\psi}(t+u))du\right|\leq C_T\left(\int_t^{t+T}|\dot{\psi}(u)|^2du\right)^{\frac{1}{2}},$$

we obtain that $(\tilde{\mathbf{z}}^{\infty}_{\psi}(t))_{t\geq 0}$ is a solution of $\dot{\mathbf{z}} = -b(\mathbf{z})$. It remains to show that $(\tilde{\mathbf{z}}^{\infty}_{\psi}(t))_{t\geq 0}$ is stationary, *i.e.* that every limit point of $(\tilde{\mathbf{z}}_{\psi}(t))_{t\geq 0}$ belongs to $\{z \in \mathbb{R}^{2d}, b(z) = 0\}$. Denote by $(\Phi_t(z))_{t,z}$ the flow associated to the o.d.e. $\dot{\mathbf{z}} = -b(\mathbf{z})$. Again, owing to the fact that for all $T > 0$, $\int_t^{t+T} |\dot{\psi}(u)|^2 du \xrightarrow{t\to+\infty} 0$, we can deduce that for all $T > 0$,

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

This means that $(\tilde{\mathbf{z}}_{\psi}(t))_{t\geq 0}$ is an *asymptotic pseudo-trajectory* for Φ (see [4]). As a consequence, by Proposition 5.3 and Theorem 5.7 of [4], the set K of limit points of $(\tilde{\mathbf{z}}_{\psi}(t))_{t\geq 0}$ is a (compact) invariant set for Φ such that $\Phi_{|K}$ has no proper attractor. This means that there is no strict invariant subset A of K such that for all $z \in K$, $d(\Phi_t(z), A) \xrightarrow{t\to+\infty} 0$.
Thus, 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 $\Phi_{|\bar{K}}$. For that, let ρ be a positive real number and 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 D^2U is locally bounded, we can find a small enough ρ and a positive α_ρ such that for all $(\mathbf{z}(t))$ solution of $\dot{\mathbf{z}} = -b(\mathbf{z})$ with $\mathbf{z}(0) \in K$,

$$\frac{d}{dt} L(\mathbf{z}(t)) \geq \alpha_\rho |\mathbf{y}(t)|^2 + \rho|\nabla U(\mathbf{x}(t))|^2. \tag{3.25}$$

For all 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 the Ascoli Theorem yields that $(\mathbf{z}(t+.))$ is relatively compact. 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 all $t \geq 0$ and thus

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

By (3.25), 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 $(\tilde{\mathbf{z}}_{\psi}(t))_{t\geq 0}$ belongs to $\{z, b(z) = 0\}$. Finally, since the set of limit points of $(\tilde{\mathbf{z}}_{\psi}(t))_{t\geq 0}$ is a compact connected set and since the set $\{x, \nabla U(x) = 0\}$ is finite, it follows that $\tilde{\mathbf{z}}_{\psi}(t) \to z^\star = (x^\star, 0)$ when $t \to +\infty$. Note that here $x^\star \in \{x, \nabla U(x) = 0\}$. Then, by (3.21) if we prove that W is continuous at $z^\star$, we shall deduce the announced result. This is the purpose of the next step.

• Step 3: We prove that for each $z^\star \in \{z, b(z) = 0\}$, *i.e.* for each $z^\star = (x^\star, 0)$ with $x^\star \in \{x, \nabla U(x) = 0\}$, W is continuous at $z^\star$. Since $D^2U(x^\star)$ is invertible, we deduce from Lemma 4.3 that the dynamical system is *locally controllable* around $z^\star$, *i.e.* that for all $T > 0$ and $\varepsilon > 0$, 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, the definition of W implies that $W(z^\star) \leq W(z) + \varepsilon$ and $W(z) \leq W(z^\star) + \varepsilon$. The continuity of W follows. Letting t go to $+\infty$ in (3.21) ends the proof of (iii). □

4 Freidlin and Wentzell theory

In this section, we derive some sharp estimations of the behaviour of (ν_ε) when $\varepsilon \to 0$. To this end, we adapt the strategy of [14] to our setting. Our goal is twofold: first, we aim at establishing some uniqueness property for the rate function W defined in Theorem 1. This will then lead to a large deviation principle for (ν_ε). Second, we want to obtain a more explicit formulation of W in order to characterize, at least in some particular cases, the limit behaviour of (ν_ε) for some non-convex potential U. In the paper, we assume that the potential U satisfies Assumption $(\mathbf{H_D})$ defined in Section 2.4.1, that is, the set of critical points of U is finite. We can then set $\{x \in \mathbb{R}^d, \nabla U(x) = 0\} = \{x_1^\star, \dots, x_\ell^\star\}$.

First, we classify the critical points, *i.e.*, we connect the critical points of the vector field b to those of U and we determine their stability. Then, with respect to these critical points, we construct the so-called skeleton Markov chain associated to the process $(X_t^\varepsilon, Y_t^\varepsilon)$. Using this, we finally derive the LDP for (ν_ε).

4.1 Classification of critical points.

We first need to classify the equilibria of the dynamical system $\dot{z} = b(z)$. We recall that $\{z \in \mathbb{R}^{2d}, b(z) = 0\} = \{z_1^\star, \dots, z_\ell^\star\}$ where for every $i \in \{1, \dots, \ell\}$, $z_i^\star = (x_i^\star, 0)$. The following proposition characterizes the nature of $z_i^\star$ with respect to that of $x_i^\star$.

Proposition 4.1. *Assume that $D^2U(x_i^\star)$ is invertible for all $i \in \{1, \dots, \ell\}$. 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 first define $I = \{i \in \{1 \dots \ell\} | x_i^\star \text{ is a local minimum}\}$ and $J = \{1 \dots \ell\} \setminus I$. Let us compute the Jacobian matrix of the vector field b: and for each $i \in \{1, \dots \ell\}$

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

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

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

where $\sqrt{1/4-\mu}$ denotes $i\sqrt{|1/4-\mu|}$ if $1/4 - \mu \leq 0$. When $i \in I$, one can note that $D^2U(x_i^\star)$ is a positive definite matrix. It follows that $\mu \in Sp(D^2U(x_i^\star))$ is positive and

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

so that $z_i^\star$ is a stable equilibrium in this case.
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). Thus, $z_i^\star$ is an unstable equilibrium of the deterministic dynamical system. $\square$

4.2 Skeleton representation

The [14] description of the invariant measure ν_ε of the continuous time Markov process is based on its representation through the invariant measure of a specific skeleton Markov chain. This formula, due to Khas'minskiĭ (see [20], chapter 4) in the uniformly elliptic case, will remain true in our framework even if the original process is hypoelliptic and defined on a non compact manifold. This is the purpose of Proposition 4.2 below but before a precise statement, 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 by

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

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 $(g_i)_i, g, \Gamma$ as well as the positions of the critical points $z_i^\star$. An example of a path $(Z_t^\epsilon)_{t\geq 0}$ is also depicted (K will be defined in the sequel).

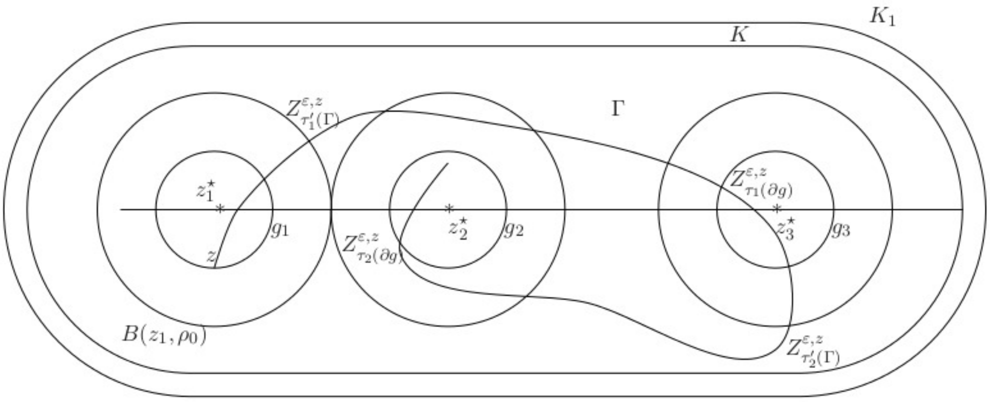

Figure 1: Graphical representation of the neighbourhood g_i, the process $(Z_t^{\varepsilon,z})_{t\geq 0}$, the skeleton chain and the compact sets K and K_1.

Now, let us define the skeleton Markov chain $(\tilde{Z}_n)_{n\in\mathbb{N}}$. We consider a path of Z^ε starting from $z \in \partial g$ and we set $\tilde{Z}_0 = z$. Then, the sequel of the skeleton chain is defined through the hitting and exit times of the neighbourhoods defined above: we set $\tau_0(\partial g) = 0$,

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

Then, for every n, τ_n' and τ_n are defined inductively by:

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

We will show in Proposition 4.2 that for all $n \geq 0$, $\tau_n(\partial g) < +\infty\ a.s.$ The skeleton is then defined for all $n \in \mathbb{N}$ by, $\tilde{Z}_n = Z^{\varepsilon,z}_{\tau_n(\partial g)}$. 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 a distribution by $\tilde{\mu}_\varepsilon^{\partial g}$. The proposition states that ν_ε may be related to $\tilde{\mu}_\varepsilon^{\partial g}$.

Proposition 4.2. *(i) Following the notations introduced before, we have*

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

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

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

is invariant for the process $(Z_t^\varepsilon)_{t\geq 0}$. Hence, $\mu_\varepsilon^{\partial g}$ is a finite measure proportional to ν_ε.

Proof. Proof of (i). Using Proposition 3.2, one first check that one can find two compact sets K, K_1 such that $g \subseteq K$ and $g \subsetneq K_1$ and such that 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 hypoelliptic setting the proofs of Lemma 4.1 and 4.3 of [20] given under some elliptic assumptions. Let $z \in \partial g$ and set $\tilde{\tau}_0 = \inf\{t \geq 0,\ Z_t^{\varepsilon,z} \in \partial K\}$,

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

and recursively for all $n \geq 2$,

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

By construction, we have $a.s.$:

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

Then, by the strong Markov property and (4.6), a careful adaptation of the proofs of Lemma 4.1 and 4.3 of [20] shows that $\sup_{z\in\partial g} \mathbb{E}_z^{\varepsilon}[\tau_1(\partial g)] < \infty$ as soon as the two following points hold for all $\varepsilon > 0$:

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

Let us focus on the first point. By Remark 5.2 of [27], it is enough to check that there exists $T > 0$ and 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 [26], we obtain that $\sup_{z\in K} \mathbb{P}(\tau(\partial K_1) \leq T) < 1$. The first point follows from the strong Markov property (see Remark 5.2 of [27] 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 definition that for all $M > 0$, there exists $T > 0$, such that for all $z \in K$, $|\mathbf{x}_\varphi(T)| > M$. The first point easily follows.

It is well-known (see for instance [27]) that for all $\varepsilon > 0$, p_ε is a solution of

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

Thus, since $\sup_{z\in K_1\backslash g}\mathbb{E}[\tau(\delta g\cup\delta K_1)]<+\infty$ and since h is 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 [27] with $k=f=0$ to obtain that $z\mapsto p_\varepsilon(z)$ is a continuous map. Furthermore, for any starting point $z\in K\backslash g$, we can build a controlled trajectory starting at any $z\in\partial K$ which hits ∂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 $\sup_{(x,y)\in K}\mathcal{E}(x,y)<\inf_{(x,y)\in K_1^c}\mathcal{E}(x,y)$, leads to an available control for all $z\in K$. Finally, using again the support theorem of [26], w deduce that for each $z\in\partial K$, $p_\varepsilon(z)<1$. The second point then follows from the continuity of $z\mapsto p_\varepsilon(z)$. This ends the proof of $i)$.

Proof of (ii). As argued in the paragraph before the statement of this proposition, $(\tilde Z_n)_{n\in\mathbb{N}}$ admits a unique invariant measure $\tilde\mu_\varepsilon^{\partial g}$. The fact that $\mu_\varepsilon^{\partial g}$ is invariant for $(Z_t^\varepsilon)_{t\geq 0}$ is classical and relies on the strong Markov property of the process (see $e.g.$ in [19]).

□

Remark 3. *One could also have used a uniqueness argument for viscosity solutions to obtain the continuity of $z\mapsto p_\varepsilon(z)$ using the maximum principle on $\mathcal{A}^\varepsilon$ (as it is already used by [27]). One may refer to [3] for further details.*

4.3 Transitions of the skeleton Markov chain

This paragraph is devoted to the description of estimations obtained through the Freidlin and Wentzell theory for the Markov skeleton chain defined above. These estimations as well as Proposition 4.2 are then used to obtain the asymptotic behaviour of (ν_ε). By Theorem 1, we know that there exists a subsequence $(\varepsilon_n)_{n\in\mathbb{N}}$ such that (ν_{ε_n}) satisfies a large deviation principle with speed ε_n^2 and good rate function W. For the sake of simplicity, we keep the notation ε. Hence, $\varepsilon\to 0$ means $\varepsilon_n\to 0$ as $n\to+\infty$ (where $(\varepsilon_n)_n$ is a suitable (LD) convergent subsequence).

4.3.1 Controllability and exit times estimates

In order to obtain some estimates related to the transition of the 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 [14]). In our hypoelliptic framework, such a bound of the exit times is strongly based on the controllability around the equilibria. We have the following property:

Lemma 4.3. *Let $i\in\{1,\ldots,\ell\}$ such that $D^2U(x_i^\star)$ is invertible. Let $T>0$. Then, for all $\delta>0$, there exists $\rho(\delta)>0$ such that*

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

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 $u=(\dot\varphi,\psi)^t$ with $\psi\in\mathbb{H}(\mathbb{R}^d)$. Since $D^2U(x_i^\star)$ 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. [11]) is satisfied and it follows from Theorems 1.16 and 3.8 of [11] that (4.9) is *locally exactly controllable* at $z_i^\star$. The lemma is then proved. □

We are now able to obtain the following estimation.

Lemma 4.4. *Suppose that* $(\mathbf{H_D})$ *holds and that either* $(\mathbf{H_{Q+}})$ *or* $(\mathbf{H_{Q-}})$ *is satisfied. Then, for all* $\gamma > 0$*, 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*

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

Proof. Let $i \in \{1 \dots \ell\}$ and fix $\gamma > 0$. By Lemma 4.3 applied with $T = 1$, one can find $\rho > 0$ such that

$$\forall (a, b) \in B(z_i^\star, 2\rho),\ \exists \varphi \in \mathbb{H} \text{ such that } \mathbf{z}_\varphi(a, 1) = b \text{ and } \frac{1}{2} \int_0^1 |\dot{\varphi}(s)|^2 ds \leq \frac{\gamma}{2}.$$

Now, we set $\delta = \rho/2$, $G = B(z_i^\star, \delta)$ and we fix $a = z$ and take b such that $|z_i^\star - b| = \rho$. This implies that for every $z \in B(z_i^\star, \delta)$, $|z - b| \leq 3\delta < 3\rho/2$. Thus, we can find $\varphi_z \in \mathbb{H}$ such that $\mathbf{z}_{\varphi_z}(z, 1) = b$ and $\frac{1}{2} \int_0^1 |\dot{\varphi}_z(s)|^2 ds \leq \gamma/2$.

It is now possible to follow the proof of Lemma 1.7, chapter 6 of [14]: first, remark that

$$\mathbb{P}_z^\varepsilon[\tau_{G^c} \leq 1] \geq \mathbb{P}\left[\sup_{t \in [0,1]} |Z_t^{\varepsilon, z} - \mathbf{z}_{\varphi_z}(z, t)| \leq \delta\right]. \tag{4.10}$$

Second, since 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

$$\liminf_{\varepsilon \to 0} \inf_{z \in G} \varepsilon^2 \ln \mathbb{P}_z^\varepsilon[\tau_{G^c} \leq 1] = \lim_{k \to +\infty} \ln(\mathbb{P}_{z_k}^{\varepsilon_k}[\tau_{G^c} \leq 1]).$$

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

$$\liminf_{\varepsilon \to 0} \inf_{z \in G} \varepsilon^2 \ln \mathbb{P}_z^\varepsilon[\tau_{G^c} \leq 1] \geq -\frac{1}{2} \int_0^1 |\dot{\varphi}_{z_\infty}(s)|^2 ds \geq -\frac{\gamma}{2},$$

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

$$\mathbb{P}_z^\varepsilon[\tau_{G^c} \leq 1] \geq e^{-\gamma \varepsilon^{-2}}.$$

Then, the strong Markov property implies that

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

so that for every $z \in G$ and $\varepsilon \in (0, \varepsilon_0]$

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

□

Following the same kind of argument based on Lemma 4.3 and on the finite time large deviation principle, we also obtain that Lemma 1.8, chapter 6 of [14] still holds. In our context, this leads to the following lemma.

Lemma 4.5. *Suppose that* $(\mathbf{H_D})$ *holds and that either* $(\mathbf{H_{Q+}})$ *or* $(\mathbf{H_{Q-}})$ *is satisfied. For any* $\rho > 0$ *and any equilibrium* $z_i^\star$ *of (2.2), we define* $G := B(z_i^\star, \rho)$*. Then, for any small enough* ε *and any* $\gamma > 0$*, there exists* $\delta \in (0, \rho]$ *such that the exit time of* G *satisfies:*

$$\inf_{z \in B(z_i^\star, \delta)} \mathbb{E}\left[\int_0^{\tau_{G^c}} \chi_{B(z_i^\star, \delta)}(Z_t^{\varepsilon, z}) dt\right] > e^{-\gamma \varepsilon^{-2}}.$$

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 $(\xi_1,\xi_2)\in(\mathbb{R}^d\times\mathbb{R}^d)^2$, $I_t(\xi_1,\xi_2)$ denotes the L^2-minimal cost to join ξ_2 from ξ_1 in a finite time t:

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

and $I(\xi_1,\xi_2)=\inf_{t\geq 0}I_t(\xi_1,\xi_2)$. We also introduce $\tilde{I}(z_i^\star,z_j^\star)$ defined for all $(i,j)\in\{1\dots\ell\}^2$ by:

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

The quantity $\tilde{I}(z_i^\star,z_j^\star)$ is the minimal cost to join $z_j^\star$ from $z_i^\star$ avoiding other equilibria of (2.2). In the following proposition, we prove that $\tilde{I}(z_i^\star,z_j^\star)$ is always finite.

Proposition 4.6. *For all $(i,j)\in\{1\dots\ell\}^2$, $\tilde{I}(z_i^\star,z_j^\star)<+\infty$.*

Proof. In the proof, we assume that $i\neq j$ and we build a controlled trajectory starting at $z_i^\star$, ending at $z_j^\star$ (in a finite time) and which avoids the other equilibria neighbourhoods $\cup_{k\neq(i,j)}g_k$.

We first assume that $d>1$. In this case, for any 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 the initial condition $\mathbf{y}_0(0)=0$ and let $\varphi_0\in\mathbb{H}$ satisfying $\dot{\varphi}_0(t)=\dot{\mathbf{x}}_0(t)+\mathbf{y}_0(t)$. We 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 all $t\in[0,t_0]$. This way, we have

$$\mathbf{x}_{\varphi_0}(z_i^\star,t_0)=x_j^\star\text{ and }\forall s\in[0;t_0]\quad\mathbf{z}_{\varphi_0}(z_i^\star,s)\notin\cup_{k\neq i,j}g_k.$$

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 all $t\geq t_0$ 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, $\mathbf{z}_{\varphi_1}((x_j^\star,\mathbf{y}_0(t_0)),t)\xrightarrow{t\to+\infty}(x_j^\star,0)$. As a consequence, 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. Moreover, this trajectory avoids the other ρ_1-neighbourhood of $\cup_{k\neq(i,j)}g_k$. It remains to use Lemma 4.3 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 is 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 consider $x_i^\star,x_j^\star$ any two critical points of U. Without loss of generality, one may suppose that $x_i^\star<x_j^\star$. From $(\mathbf{H_D})$, the number of critical points which belong to $[x_i^\star,x_j^\star]$ is finite (denoted by p):

$$x_i^\star<x_{i_1}^\star<\cdots<x_{i_p}^\star<x_{i_{p+1}}^\star:=x_j^\star.$$

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, $\mathbf{y}_\alpha(t)$ 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 increasing 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 $t_1, \dots t_p$ 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 α is increasing, we first show that one can find a monotone α satisfying $\mathbf{y}_\alpha(t_1) \neq 0$. Let α be 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), $\mathbf{y}_\alpha(t_1) \neq 0$. 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 $\mathbf{x}(\tilde{t}_1) = \xi_1$ and α remains constant on $[\tilde{t}_1; \tilde{t}_1 + \delta t_1]$. Expanding the integral given in Equation (4.11) between $[0, t_1], [t_1, \tilde{t}_1]$ and $[\tilde{t}_1, \tilde{t}_1 + \delta t_1]$, a simple computation yields

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

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 t_2, time at which $x_{i_2}^\star$ is reached. By construction, $\mathbf{y}_\alpha(t_2) > 0$. Now, one can repeat the same argument by induction to find α such that $\mathbf{y}_\alpha(t_k) \neq 0$ for all $k \leq p+1$ such that $\mathbf{x}_\alpha(t_k) = x_{i_k}^\star$.

Thus, at time t_{p+1}, $\mathbf{x}_\alpha(t_{p+1}) = x_j^\star$ and $\mathbf{y}_\alpha(t_{p+1}) \neq 0$. It remains now to join $z_j^\star = (x_j^\star, 0)$ without hitting $\cup_{k\neq i,j} g_k$. This concluding step can be achieved exactly as in dimension $d > 1$. □

It is now possible to obtain the estimation of the invariant measure $\tilde{\mu}_\varepsilon^{\partial g}$ of the skeleton chain. This point follows from the estimation of the transition probability of $(\tilde{Z}_n)_{n\in\mathbb{N}}$ (denoted $\tilde{P}^\varepsilon(z, .)$). From Lemmas 4.3, 4.4, 4.5 and Proposition 4.6, we deduce the following result from a simple adaptation of the proof of Lemmas 2.1 and 2.2, chapter 6 of [14].

Proposition 4.7. *For any $\gamma > 0$, there exist some sufficiently small ρ_0 and ρ_1 satisfying $0 < \rho_1 < \rho_0$ such that we have for any small enough ε*

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

4.4 $\{i\}$-Graphs and invariant measure estimation

We recall that $\{i\}$-Graphs for Markov chains are defined in paragraph 2.4.2, and that the set of all possible $\{i\}$-Graphs is referred as $\mathcal{G}(i)$. Recall that

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

As pointed in Lemma 4.1 of [14], one can check that

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

We are now able to obtain the main result of this paragraph. From the skeleton representation (Proposition 4.2) and the estimations given by Lemma 4.5 and Proposition 4.7, we obtain the asymptotic behaviour of (ν_ε) as $\varepsilon \longrightarrow 0$. The result is as follows.

Theorem 4. *For any $\gamma > 0$, there exist some real numbers ρ_0 and ρ_1 satisfying $0 < \rho_1 < \rho_0$ such that if $g_j = B(z_j^\star, \rho_1)$:*

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

for all $i \in \{1 \dots \ell\}$. As well, in terms of W, we get that

$$e^{-\varepsilon^{-2}\left[W(z_i^\star) + \gamma\right]} \leq \nu_\varepsilon(g_j) \leq e^{-\varepsilon^{-2}\left[W(z_i^\star) - \gamma\right]}, \quad \forall i \in \{1 \dots \ell\}.$$

The proof of this theorem is straightforward according to the previous results: the invariant measure ν_ε only weights small neighbourhoods of global minima of W, when $\varepsilon \to 0$. Such global minima are appropriately described using the quasipotential I and the function $\mathcal{W}$ obtained through the $\{i\}$-graph structures.

5 Lower and upper bound of the rate function with a double-well landscape in $\mathbb{R}$

This last part is devoted to the proof of Theorem 3. Here, we focus on a one dimensional potential U with a double-well profile and on the memory gradient system with fixed memory parameter λ. In this case we push further our study of (ν_ε) when $\varepsilon \to 0$. From Freidlin and Wentzell estimates, ν_ε concentrates on the minima of W (this set is also the minima of $\mathcal{W}$). Here we derive that (ν_ε) concentrates on the global minimum of U. For this purpose, we consider of a *double-well potential* U whose the two minima are denoted by $x_1^\star$ and $x_2^\star$. One needs to compare the costs $I(z_1^\star, z_2^\star)$ and $I(z_2^\star, z_1^\star)$ for the two stable equilibria $z_1^\star = (x_1^\star, 0)$ or $z_2^\star = (x_2^\star, 0)$. Without loss of generality, we fix $x_1^\star < x_2^\star$ and we assume that there exists a unique local 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 without loss of generality that $U(x_1^\star) < U(x_2^\star)$.

We first describe how one can provide a lower bound of the cost $I(z_1^\star, z_2^\star)$. We propose two approaches using sharp estimates of some particular Lyapunov functions. In the next subsection, we adopt a non-degenerate approach where the main idea is to project the drift vector field onto the gradient of a Lyapunov function. However, even if the idea seems to be original, the bounds are not very satisfactory (see Proposition 5.1). In Subsection 5.2 we propose a second approach which provides better bounds (see Proposition (5.2)).

5.1 Lower-Bound using a non-degenerate approach

In this section, we consider the following Lyapunov function defined by

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

when $(\beta, \gamma) \in \mathbb{R}^2$. For the sake of simplicity, we omit the dependence on β and γ and denote by $\mathcal{L}$ this function. Here, the main idea relies on the fact that $\nabla\mathcal{L}$ corresponds to a favoured direction of the drift b. This will allow us to control the L^2 cost to move from $z_1^\star$ to $z_2^\star$. First let us remark that the cost I is necessarily bounded from below by the L^2 cost for an elliptic system. In the elliptic context, the L^2 cost is defined by

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

which can also been written as

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

As a consequence, since the set of admissible control for the degenerate cost I_T is included in the set of admissible controls for $I_{\mathcal{E},T}$ (v is forced to be 0 in Equation (5.1)), we deduce that I_T is greater than $I_{\mathcal{E},T}$. This way, a lower bound for $I_{\mathcal{E},T}$ 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 of [10], 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 line generated by $\nabla\mathcal{L}$. 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 $\mathbf{z}$. Let us now remark that

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

If one can find β and γ such that there exists $\alpha > 0$ satisfying

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

then, it would be possible to conclude that for all $T > 0$

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

and so

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

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

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

In the case $\beta \leq 0$, 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 (that is under H_{Q-}).

Proposition 5.1. *Let $U : \mathbb{R}^d \to \mathbb{R}$ be a $\mathcal{C}^2$ one-dimensional double-well potential such that U'' is bounded and set $M = \|U"\|_\infty$. 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

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

At last, we have

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

Proof. The idea is to optimize the ratio given in Equation (5.4) for the largest possible α. Such a ratio can be written as a quadratic form on y and $U'(x)$. This way, an algebraic argument based on a simultaneous reduction of these quadratic forms yields a suitable calibration for M and α. We refer to [16] for the missing technical details. □

Let us remark that this bound strongly depends on the second derivative of U. In particular, when $M = \|U"\|_\infty$ is large, the lower bound vanishes as $M \to +\infty$ and becomes useless.

5.2 Lower bound using a degenerate approach (Proof of (ii) of Theorem 3)

Our original dynamical system is indeed degenerated on the first coordinate. In order to take into account the degeneracy, we derive a lower bound of $I_T(z_1, z_2)$ by a gradient of a suitable Lyapunov function. This approach will lead to some better estimates than those of the previous paragraph obtained with an elliptic argument.

For this purpose, we consider the Lyapunov function defined by

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

where α, β and γ are real numbers such that $\alpha > 0$. We are looking for a suitable choice of (α, β, γ). Consider $\varphi \in \mathbb{H}(\mathbb{R}_+, \mathbb{R}^d)$ and denote by $(\mathbf{z}_\varphi(t))_{t\geq 0}$ the associated controlled trajectory. Our objective is to obtain the following kind of lower bound:

$$\forall \varphi \in \mathbb{H}(\mathbb{R}_+, \mathbb{R}^d), \quad \forall t \geq 0 \qquad \dot{\varphi}^2(t) \geq 2\frac{\mathrm{d}\mathcal{L}_{\alpha,\beta,\gamma}(\mathbf{z}_\varphi(t))}{\mathrm{d}t}. \tag{5.5}$$

Such a lower bound is useful especially if α is positive, largest as possible and β is non-negative. Indeed, denote by $t^\star$ the first time at which a controlled trajectory $(\mathbf{z}(t))_{t\geq 0}$ reaches $z^\star = (x^\star, 0)$ where $x^\star$ denotes the (unique) local maximum of U. Then, if Inequality (5.5) holds, we have for all $T > 0$

$$\begin{aligned}
I_T(z_1^\star, z_2^\star) &= \inf_{u\in\mathbb{L}^2([0;T])} \left\{ \frac{1}{2}\int_0^T u^2(s)ds,\ \dot{\mathbf{z}} = b(\mathbf{z}) + \begin{pmatrix} u \\ 0 \end{pmatrix}, \mathbf{z}(0) = z_1^\star, \mathbf{z}(T) = z_2^\star \right\} \\
&\geq \inf_{u\in\mathbb{L}^2([0;T])} \left\{ \frac{1}{2}\int_0^{t^\star} u^2(s)ds,\ \dot{\mathbf{z}} = b(\mathbf{z}) + \begin{pmatrix} u \\ 0 \end{pmatrix}, \mathbf{z}(0) = z_1^\star, \mathbf{z}(T) = z_2^\star \right\} \\
&\geq \alpha[U(x^\star) - U(x_1^\star)] + \frac{\beta}{2}\mathbf{y}(t^\star)^2 \\
&\geq \alpha[U(x^\star) - U(x_1^\star)]. \qquad (5.6)
\end{aligned}$$

The next proposition shows that indeed, Inequality (5.5) holds for a suitable choice of β and γ. In some case, the lower bound is almost optimal.

Proposition 5.2. *For any $\alpha \in [0;2[$, there exist some explicit constants $m_\lambda(\alpha), \beta^\star(\alpha), \gamma^\star(\alpha)$ such that (5.5) is true for $\beta = \beta^\star(\alpha), \gamma = \gamma^\star(\alpha)$ and for all one-dimensional double well potential U satisfying $\|U"\|_\infty = M < m_\lambda(\alpha)$. In this case, we get*

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

Proof. Again, we only give the sketch of the proof. Dropping the time parameter, we write $u^2 - \mathrm{d}\mathcal{L}_{\alpha,\beta,\gamma}(\mathbf{z})/\mathrm{d}t$ as a quadratic form that depends on $\dot{\mathbf{x}}, \mathbf{y}$ and $U'(\mathbf{x})$. We know that this quadratic form is positive if and only if all its principal minors are positive. Then, a careful calibration of the several parameters yields the desired results. We refer to [16] for a detailed proof of this result.

□

When M is large, the admissible values of α vanish and our lower bound becomes useless. When $M \to 0$, 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). A comparison of the two methods is depicted in Figure 2 through the size of admissible α for several values of λ.

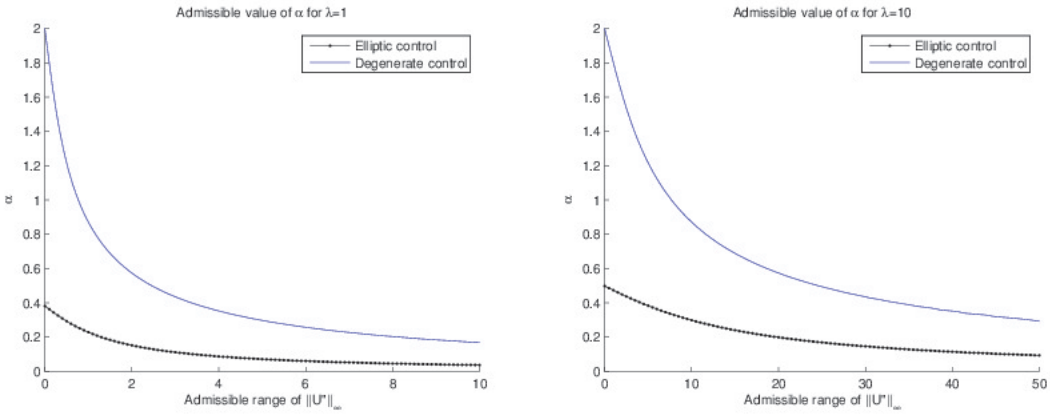

Figure 2: Maximum size of α with $\|U"\|_\infty$ when $\lambda = 1$ (left) and $\lambda = 10$ (right) for both approaches (Proposition 5.1 and 5.2).

5.3 Upper-Bound for the cost function (Proof of i) of Theorem 3)

Remind that we assume that there are two local minima 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 $x^\star$. Again, we set $z_1^\star = (x_1^\star, 0)$, $z_2^\star = (x_2^\star, 0)$ and $z^\star = (x^\star, 0)$. We are looking for a upper-bound of $I(z_2^\star, z_1^\star)$ 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 $U(x_1^\star) < U(x_2^\star)$. Then, for all $\lambda > 0$,*

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

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.

This proposition is a consequence of the following Lemma 5.4 and Lemma 5.5 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 5.4 and Lemma 5.5 rely both on Lemma 5.6.

Lemma 5.4. *Under the assumptions of Proposition 5.3,* $I(z^\star, z_i^\star) = 0 \quad \text{for } i = 1, 2.$

Proof. We establish the result for $i = 1$. We first show that for all starting points $z_\varepsilon = (x_\varepsilon, y_\varepsilon)$ such that $U(x_\varepsilon) + |y_\varepsilon|^2/(2\lambda) < U(x^\star)$ and $x_\varepsilon \in [x_1^\star, x^\star]$, we have $I(z_\varepsilon, z_1^\star) = 0$. Second, applying Lemma 5.6, we will find $(z_\varepsilon)_{\varepsilon>0}$ such that for all $\varepsilon > 0$, $I(z^\star, z_\varepsilon) \leq \varepsilon$ and the result will follow when $\varepsilon \to 0$.

Let $(\mathbf{z}(t))_{t\geq 0} = (\mathbf{x}(t), \mathbf{y}(t))_{t\geq 0}$ be a solution of $\dot{\mathbf{z}} = b(\mathbf{z})$ starting from z_ε and let F be the function 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 and thus convergent,when t goes to infinity. As a consequence, F is bounded on $\mathbb{R}_+$. Then, using that $\mathcal{E}$ is coercive, it follows that $(\mathbf{z}(t))_{t\geq 0}$ is bounded. Since U" is continuous, we deduce that U' is Lipschitz continuous on the set where $(\mathbf{x}, \mathbf{y})$ is living. Using standard arguments, this implies that the family of shifted trajectories $(\mathbf{z}(t + .))_{t\geq 0}$ is relatively compact for the topology of uniform convergence on compact sets. If $\mathbf{z}^\infty$ stands for the limit of a convergent subsequence $(\mathbf{z}(t_n + .))_{n\geq 0}$, then $\mathbf{z}^\infty$ is a solution of $\dot{\mathbf{z}} = b(\mathbf{z})$. Since F is continuous and converges as $t \to +\infty$ to some limit l, we deduce that $\mathcal{E}(\mathbf{x}^\infty(s), \mathbf{y}^\infty(s)) = l$ for all $s \geq 0$. Thus we get

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

Using that $F'(t) = -\mathbf{y}(t)^2$, we then obtain $\mathbf{y}^\infty(t) = 0$ for all $t \geq 0$. Thus, $\mathbf{x}^\infty$ is constant and $\mathbf{z}^\infty$ is a *stationary* solution of $\dot{\mathbf{z}} = b(\mathbf{z})$. We can deduce that every accumulation point of $(\mathbf{x}(t), \mathbf{y}(t))_{t\geq 0}$ belongs to $\{z \in \mathbb{R}^{2d}, b(z) = 0\}$. Under the assumption $U(x_\varepsilon) + |y_\varepsilon|^2/(2\lambda) < U(x^\star)$, and since F is non increasing, the only possible limit 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

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

Now, by Lemma 5.6, we get that for any $\varepsilon > 0$, one can find $z_\varepsilon = (x_\varepsilon, y_\varepsilon)$ such that we both have $U(x_\varepsilon) + |y_\varepsilon|^2/(2\lambda) < U(x^\star)$, $x_\varepsilon \in [x_1^\star, x^\star]$ and $I(z^\star, z_\varepsilon) \leq \varepsilon$ (taking for instance z_ε on the segment $[z_1^\star, z^\star]$ sufficiently close to $z^\star$). The result follows. □

Lemma 5.5. *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 (see (5.7) below). Since $z_2^\star$ is an equilibrium point of $\dot{\mathbf{z}} = b(\mathbf{z})$, it follows from Lemma 5.6 that for all $\varepsilon > 0$, there exists $\tilde{z}_\varepsilon$ such that $\tilde{z}_\varepsilon = z_2^\star + \delta_\varepsilon(z^\star - z_2^\star) = (x_2^\star + \delta_\varepsilon(x^\star - x_2^\star), 0)$ with $\delta_\varepsilon \in (0, 1)$, $U(\tilde{x}_\varepsilon) > U(x_2^\star)$ and $I(z_2^\star, \tilde{z}_\varepsilon) \leq \varepsilon$. It is now enough to prove that

$$\forall\, \varepsilon > 0, \quad I(\tilde{z}_\varepsilon, z^\star) \leq 2(U(x^\star) - U(\tilde{x}_\varepsilon)).$$

Let us consider $\mathbf{z}_\varepsilon = (\mathbf{x}_\varepsilon, \mathbf{y}_\varepsilon)$ defined as the solution of

$$\begin{cases} \dot{\mathbf{x}}_\varepsilon(t) = \mathbf{y}_\varepsilon(t) = -\mathbf{y}_\varepsilon(t) + 2\mathbf{y}_\varepsilon(t) \\ \dot{\mathbf{y}}_\varepsilon(t) = \lambda(U'(\mathbf{x}_\varepsilon(t)) - \mathbf{y}_\varepsilon(t)) \end{cases} \tag{5.7}$$

starting from $\tilde{z}_\varepsilon$. Note that setting $\varphi(t) = 2\int_0^t \mathbf{y}_\varepsilon(s)ds$, $\mathbf{z}_\varepsilon = \mathbf{z}_\varphi$. Let us study its asymptotic behavior. To this end, we now introduce the function $\tilde{F}$ defined by $\tilde{F}(t) = \frac{\mathbf{y}_\varepsilon(t)^2}{2\lambda} - U(\mathbf{x}_\varepsilon(t))$. We observe that $\tilde{F}'(t) = -\mathbf{y}_\varepsilon(t)^2$ and thus that $\tilde{F}$ is non-increasing.

We first show that $(\mathbf{x}_\varepsilon(t), \mathbf{y}_\varepsilon(t))$ starting from $\tilde{z}_\varepsilon$ satisfies $x_1^\star < \mathbf{x}_\varepsilon(t) < x_2^\star$ for all $t \geq 0$. 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

$$\frac{\mathbf{y}_\varepsilon(\tau_1)^2}{2\lambda} - U(x_1^\star) = -U(\tilde{x}_\varepsilon) - \int_0^{\tau_1} \mathbf{y}_\varepsilon(s)^2 ds. \tag{5.8}$$

It would imply 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 similarly 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 decreasing property of $\tilde{F}$ implies that

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

As a consequence, $(\mathbf{x}_\varepsilon(t), \mathbf{y}_\varepsilon(t))_{t\geq 0}$ is bounded and an argument closed to the one used in the proof of Lemma 5.4 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 set of stationary solutions of (5.7). Thus, we deduce that $\mathbf{x}_\varepsilon^\infty \in \{x_1^\star, x^\star, x_2^\star\}$.

Now, Equation (5.8) and the fact that $U(\tilde{x}_\varepsilon) > \max(U(x_1^\star), U(x_2^\star))$ 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, since $\mathbf{z}_\varepsilon = \mathbf{z}_\varphi$ with $\dot{\varphi} = 2\mathbf{y}_\varepsilon$, we deduce that

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

The announced result follows. □

Lemma 5.6. *Let x_0 be 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 $|v| = 1$ and set $z_0 = (x_0, 0)$. Then, for any positive ε and ρ, there exist $\tilde{\rho} > 0$ and $\tau \geq 0$ such that $z_\varepsilon := z_0 + \tilde{\rho}v$ satisfies $I_\tau(z_0, z_\varepsilon) \leq \varepsilon$.*

This result is a particular case of Lemma 4.3 if $U''(x_0) \neq 0$. In the degenerate case, $i.e.$ when $U''(x_0) = 0$, this result can be proved using the fact that there exists an infinite number of oscillations of the dynamical system around the stable points of $\dot{\mathbf{z}} = b(\mathbf{z})$ (see [16] for a detailed proof).

Appendix A: Proof of Proposition 3.4

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

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

Since h is bounded continuous, it follows from Assumption (ii) and Lemma 3.1.12 of [22] that for all $z \in \mathbb{R}^{2d}$, for all $(z_\varepsilon)_{\varepsilon>0}$ such that $z_\varepsilon \to z$,

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

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 [22], 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 [22], we obtain that

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

It follows that for all bounded continuous function h,

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

By Theorem 1.7.27 of [22], the above equality holds in fact for all bounded measurable function h. Applying this equality with $h = \mathbf{1}_{\{z_0\}}$

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

and the result follows. □

Appendix B: Proof of Lemma 3.3

Proof. 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 $D_x^2V(x,y) = D^2U(x)+mI_d$, $\rho\in(0,1)$ and $\lim_{|(x,y)|\to+\infty} V(x,y)=+\infty$ (see (3.16)), we have

$$\frac{Tr\Big[p(p-1)V^{p-2}\nabla_x V\otimes\nabla_x V + pV^{p-1}D_x^2V\Big]}{V^p(x,y)} \longrightarrow 0 \quad \text{as } |(x,y)|\longrightarrow+\infty.$$

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

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

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

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

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 $|(x,y)| \to +\infty$,

$$\text{(B.3)} = \begin{cases} F_1 + F_2 & \text{under } (\mathbf{H_{Q+}}) \\ F_1 & \text{under } (\mathbf{H_{Q-}}), \end{cases}$$

where $F_1 \le C(1 + V^{2p-1})$ for a suitable constant C and $F_2/V^p \to 0$ as $|(x,y)| \to +\infty$. 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

[1] Alvarez F., On the minimizing property of a second order dissipative system in hilbert spaces, SIAM J. Control Optim. **38** (2000), 1102–1119. MR-1760062

[2] Azencott R., Grandes déviations et applications, Eighth Saint Flour Probability Summer School. (1978), 1–176. MR-0590626

[3] Barles G., Solutions de viscosité des équations de Hamilton-Jacobi, Mathématiques & applications, vol. 17, Springer-Verlag, 1994. MR-1613876

[4] Benaim M., A dynamical system approach to stochastic approximations, SIAM J. Control Optim. **34** (1996), no. 2, 437–472. MR-1377706

[5] Biswas A. and Borkar V.S., Small noise asymptotics for invariant densities for a class of diffusions: a control theoretic view, J. Math. Anal. Appl. **360** (2009), no. 2, 476–484. MR-2561245

[6] Biswas A. and Budhiraja A., Exit time and invariant measure asymptotics for small noise constrained diffusions, Stochastic Process. Appl. **121** (2011), no. 5, 899–924. MR-2775101

[7] Cabot A., Asymptotics for a gradient system with memory term., Proc. of the American Mathematical Society **9** (2009), 3013–3024. MR-2506460

[8] Cabot A., Engler H., and Gadat S., Second-order differential equations with asymptotically small dissipation and piecewise flat potentials., Electronic Journal of Differential Equation **17** (2009), 33–38. MR-2605582

[9] Cerrai S. and Röckner M., Large deviations for invariant measures of stochastic reaction-diffusion systems with multiplicative noise and non-Lipschitz reaction term, Ann. Inst. H. Poincaré Probab. Statist. **41** (2005), no. 1, 69–105. MR-2110447

[10] Chiang T.S., Hwang C.R., and Sheu S.J., Diffusion for global optimization in R^n, SIAM J. Control Optim. **25** (1987), no. 3, 737–753. MR-885196

[11] Coron J.M., Control and nonlinearity, Mathematical Surveys and Monographs, vol. 136, American Mathematical Society, Providence, RI, 2007. MR-2302744

[12] Dembo A. and Zeitouni O., Large deviations techniques and applications, Stochastic Modelling and Applied Probability, vol. 38, Springer-Verlag, Berlin, 2010, Corrected reprint of the second (1998) edition. MR-2571413

[13] Feng J. and Kurtz T.G., Large deviations for stochastic processes, Mathematical Surveys and Monographs, vol. 131, American Mathematical Society, Providence, RI, 2006. MR-2260560

[14] Freidlin M. I. and Wentzell A. D., Random perturbations of dynamical systems, second ed., Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 260, Springer-Verlag, New York, 1979, Translated from the 1979 Russian original by Joseph Szücs. MR-1652127

[15] Gadat S. and Panloup F., Long time behaviour and stationary regime of memory gradient diffusions., Ann. Inst. H. Poincaré Probab. Statist., to appear (2013), 1–36.

[16] Gadat S., Panloup F., and Pellegrini C., Large deviation principle for invariant distributions of memory gradient diffusions (long version)., Preprint hal-00759188 (2012), 1–40.

[17] Herrmann S. and Tugaut J., *Stationary measures for self-stabilizing processes: asymptotic analysis in the small noise limit*, Electron. J. Probab. **15** (2010), no. 69, 2087–2116. MR-2745727

[18] Holley R. and Stroock D., Simulated annealing via Sobolev inequalities, Comm. Math. Phys. **115** (1988), no. 4, 553–569. MR-933455

[19] Khas'minskiĭ R.Z., 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** (1960), 179 –196. MR-0133871

[20] Khas'minskiĭ R.Z., Stochastic stability of differential equations, Sijthoff & Noordhoff, Alphen aan den Rijn (The Nederlands), 1980. MR-0600653

[21] Miclo L., Recuit simulé sur $\mathbf{R}^n$. Étude de l'évolution de l'énergie libre, Ann. Inst. H. Poincaré Probab. Statist. **28** (1992), no. 2, 235–266. MR-1162574

[22] Puhalskii A., Large deviations and idempotent probability, Monographs and Surveys in Pure and Applied Mathematics, vol. 119, Chapman & Hall/CRC, Boca Raton, FL, 2001. MR-1851048

[23] Puhalskii A., On large deviation convergence of invariant measures, J. Theoret. Probab. **16** (2003), no. 3, 689–724. MR-2009199

[24] Puhalskii A., On some degenerate large deviation problems, Electron. J. Probab. **9** (2004), no. 28, 862–886. MR-2110021

[25] Raimond 0., Self-interacting diffusions: a simulated annealing version, Probab. Theory Related Fields **144** (2009), no. 1-2, 247–279. MR-2480791

[26] Stroock D. and Varadhan S. R. S., On the support of diffusion processes with applications to the strong maximum principle, Symposium of Math. Statist. Prob. III (1970), 333–359. MR-0400425

[27] Stroock D. and Varadhan S. R. S., On degenerate elliptic-parabolic operators of second order and their associated diffusions, Comm. Pure Appl. Math. **25** (1972), 651–713. MR-0387812