Large deviations for singular and degenerate diffusion models in adaptive evolution

Nicolas Champagnat

Large deviations for singular and degenerate diffusion models in adaptive evolution

Nicolas Champagnat*

June 5, 2009

Abstract

In the course of Darwinian evolution of a population, punctualism is an important phenomenon whereby long periods of genetic stasis alternate with short periods of rapid evolutionary change. This paper provides a mathematical interpretation of punctualism as a sequence of change of basin of attraction for a diffusion model of the theory of adaptive dynamics. Such results rely on large deviation estimates for the diffusion process. The main difficulty lies in the fact that this diffusion process has degenerate and non-Lipschitz diffusion part at isolated points of the space and non-continuous drift part at the same points. Nevertheless, we are able to prove strong existence and the strong Markov property for these diffusions, and to give conditions under which pathwise uniqueness holds. Next, we prove a large deviation principle involving a rate function which has not the standard form of diffusions with small noise, due to the specific singularities of the model. Finally, this result is used to obtain asymptotic estimates for the time needed to exit an attracting domain, and to identify the points where this exit is more likely to occur.

AMS 2000 subject classification. Primary 60F10, 92D15; secondary 60J70, 60J60.

Key words and phrases: adaptive dynamics; punctualism; diffusion processes; degenerate diffusion; discontinuous drift; strong Markov property; probability to hit isolated points; large deviations; problem of exit from a domain.

1 Introduction

The Darwinian evolution of an asexual population is controlled by demographic (birth and death) rates, which are typically influenced by quantitative characters, called phenotypic traits: morphological traits like body size, physiological traits like the rate of food intake, life-history traits like the age at maturity. Such traits are heritable yet not perfectly transmitted from parents to offsprings, due to mutations of genes involved in their expression. The resulting variation of traits is then exposed to selection caused by ecological interactions between individuals competing for limited resources. Models of evolution of the dominant trait in the space of phenotypic traits are usually of two types: jump processes (often called "adaptive random walks" [36, 19]) or diffusion processes ([31, 26]). Diffusion

*EPI TOSCA, INRIA Sophia Antipolis – Méditerranée, 2004 route des Lucioles, BP. 93, 06902 Sophia Antipolis Cedex, France; e-mail: Nicolas.Champagnat@sophia.inria.fr.

models are usually more suited to finite populations, weak selection, or long time scales. These models usually involve a so-called "fitness function", which quantifies the selective ability of each possible phenotypic traits. Such models are also sometimes referred to as evolution models on a "fitness landscape" (an notion going back to Wright [42]).

In most cases, the parameters of these models (speed of evolution, fitness function,...) are based on heuristic considerations. However, since the early 1990's, adaptive dynamics theory [27, 34, 35] has been developed to give a firm basis to such models, starting from an individual-based description of the population with explicit ecological interactions. The combination of ecology and evolution allowed to obtain evolutionary models on a fitness landscape that depends on the current state of the population, and which is explicitly given in terms of individual parameters. The first model is an adaptive random walk, called the "trait substitution sequence" (TSS), first described in [36] (see also [15]). The mathematical derivation of this model from an individual-based model under specific asymptotics has been done in [6]. In the limit of small mutations, this stochastic jump process converges to a deterministic ordinary differential equation called "canonical equation of adaptive dynamics" [15, 7, 10]. Several diffusion models have also been obtained in this framework [8, 9], either as diffusion approximations of the TSS or in the case of weak selection in finite populations.

One evolutionary pattern that remains poorly understood among biologists is that of "punctualism", which goes back to Eldredge and Gould [22]. This is the phenomenon of Darwinian evolution whereby long periods of trait stasis alternate with periods of global, rapid changes in the trait values of the population, which can be due to a large mutation or to successive invasions of slightly disadvantaged mutants in the population [38]. In this paper, we interpret punctualism as phases of quick changes of basin of attraction for the canonical equation of adaptive dynamics, separated by long phases where the population state stays near the evolutionary equilibrium inside the current basin of attraction ("problem of exit from a domain" [24]). The TSS model is not well-suited to this study because it cannot jump in the direction of less fitted traits (i.e. traits having negative fitness). However, for punctualism to occur, a sequence of surviving unfitted mutations must occur. This is possible on long time scales because of the finiteness of the population. Therefore, we focus in this work on a diffusion model of adaptive dynamics that generalizes the one of [8] (see [5] for a general derivation of these models), where evolution can proceed in any direction of trait space. This model is obtained as a diffusion approximation (in the sense of [23, Ch. 11]) of the TSS.

This diffusion process on the trait space, assumed to be a subset of $\mathbb{R}^d$, is solution to the the following stochastic differential equation, with coefficients explicitly obtained in terms of biological parameters (see section 2):

$$dX_t^\varepsilon = (b(X_t^\varepsilon) + \varepsilon\tilde{b}(X_t^\varepsilon))dt + \sqrt{\varepsilon}\sigma(X_t^\varepsilon)dW_t, \tag{1.1}$$

where $b(x)$ and $\tilde{b}(x)$ are in $\mathbb{R}^d$, $\sigma(x)$ is a $d \times d$ symmetric positive real matrix, and $\varepsilon > 0$ is a small parameter scaling the size of mutation jumps.

The main difficulty of this model is that the standard regularity assumptions for stochastic differential equations (SDE) are not satisfied: the function b is (globally) Lipschitz, but $\tilde{b}$ is discontinuous at isolated points of the trait space, called *evolutionary singularities*, and σ is not globally Lipschitz, but is only 1/2-Hölder near the set Γ of evolutionary singularities. Moreover, $b(x) = \tilde{b}(x) = \sigma(x) = 0$ for $x \in \Gamma$.

Despite these difficulties, we are able to study the existence, strong Markov property and (partly) uniqueness for this SDE, to prove a large deviations principle (LDP) as $\varepsilon \to 0$, and to study the problem of diffusion exit from a domain of Freidlin and Wentzell [24], which is the key question for punctualism: what are the time and point of exit of X^ε from an attracting domain?

The original method for proving a LDP for the solution to a SDE with Lipschitz coefficients was based on discretization and continuous mapping techniques [24, 2] (transfer of the LDP for Brownian motion—Schilder's theorem—to the LDP for the diffusion). This technique has been extended to weaker assumptions on the coefficients (e.g. essentially locally-lipschitz in [3] or for a restricted class of two-dimensional diffusions in [30]) or to reflected diffusions [18]. Other techniques were more recently developed to study LDP for diffusions with irregular coefficients. The weak convergence approach of Dupuis and Ellis [20] is based on a combination of perturbation approach, discretization and representation formulas. They were in particular able to obtain upper bounds under very general assumptions [21] and to obtain the LDP for diffusions with discontinuous coefficients [4] (see also [11]). Another technique developed by de Acosta [1], is based on an abstract non-convex formulation of LDP, and allows one to deal with degenerate diffusion coefficients, but requires Lipschitz coefficients.

However, the existing results dealing with discontinuous coefficients are of a different nature as the singularity we consider (in [11, 4], the drift coefficient is discontinuous on a hyperplane), and these later methods require either the coefficients to be Lipschitz, or the diffusion parameter to be non-degenerate. Another reason why these methods seem not to apply easily to our situation is that the rate function arising naturally with these methods does not take into account the singularity of our model. Actually, the results of [21] can be used to obtain an large deviation upper bound, but, as appears in Section 4, with a non-optimal rate function. For these reasons, we adapt in this work the original methods based on discretization and path comparisons, allowing us to finely study the paths of the diffusion X^ε near Γ. Our proof follows the method of Azencott [2] (see also [18]). Interestingly, it also appears that, in contrast with what is usually observed in large deviations theory (see e.g. [21]), our upper bound is more difficult to obtain than the lower bound.

The paper is organized as follows. In Section 2, we describe precisely the model, study the regularity of the parameters $a = \sigma\sigma^*$, b and $\tilde{b}$ and give an example of biologically motivated parameters. In Sections 3.1 and 3.2, we establish strong existence and the strong Markov property for (1.1), by explicitly constructing a solution until the first time it hits Γ, and next setting X^ε constant after this time. Because of the bad regularity properties of $\tilde{b}$ and σ, uniqueness is a difficult problem. We are only able to prove pathwise uniqueness under the assumption that X^ε a.s. never hits Γ, and we give in Sections 3.3 and 3.4 explicit conditions ensuring this assumption and other conditions ensuring the converse. In Section 3.5, these conditions are applied to our example. In section 4, we prove the main result of this paper: a large deviation principle for X^ε as $\varepsilon \to 0$. In Section 5, we apply this result to the problem of diffusion exit from an attracting domain. We obtain in Section 5.1 a lower bound for the time of exit and we prove that the exit occurs with high probability near points of the boundary minimizing the quasi-potential. We finally show in Section 5.2 the implications of this result in terms of punctualism on our biological example.

2 Description of the model

We assume for simplicity that the space of phenotypic traits is $\mathbb{R}^d$ for some $d \geq 1$ (this may appear as a restrictive assumption, however see Remark 2.1 below). The coefficients b, $\tilde{b}$ and $\sigma\sigma^* = a$ of the SDE (1.1) are functions on $\mathbb{R}^d$, explicitly given in terms of two biological parameters: the fitness function, and the mutation law. In this section, we first define these parameters, and then study their regularity.

2.1 The fitness function

The function $g(y, x)$ from $\mathbb{R}^d \times \mathbb{R}^d$ to $\mathbb{R}$ is the *fitness function*, which measures the selective advantage (or disadvantage) of a single mutant individual with trait y in a population with dominant trait x (see [36, 6]). If $g(y, x) > 0$ (resp. $g(y, x) < 0$), then the mutant trait y is selectively advantaged (resp. disadvantaged) in a population of trait x. With this in mind, the fact that the fitness function satisfies

$$g(x, x) = 0, \quad \forall x \in \mathbb{R}^d \tag{2.1}$$

is natural (a mutant trait with trait x is neither advantaged nor disadvantaged in a population with the same trait).

When g is sufficiently regular, we will denote by $\nabla_1 g$ the gradient of $g(y, x)$ with respect to the first variable y, and by $H_{i,j}g$ the Hessian matrix of $g(y, x)$ with respect to the i-th and j-th variables $(1 \leq i, j \leq 2)$.

We introduce the sets

$$\Gamma = \{x \in \mathbb{R}^d : \nabla_1 g(x, x) = 0\}, \tag{2.2}$$

$$\text{and} \quad \forall \alpha > 0, \quad \Gamma_\alpha = \{x \in \mathbb{R}^d : d(x, \Gamma) \geq \alpha \text{ and } \|x\| \leq 1/\alpha\}, \tag{2.3}$$

where $\|\cdot\|$ denote the standard Euclidean norm in $\mathbb{R}^d$. The points of Γ are called *evolutionary singularities.*

We assume that

(H1) $g(y, x)$ is $\mathcal{C}^2$ on $\mathbb{R}^{2d}$ with respect to the first variable y, and $\nabla_1 g$ and $H_{1,1}g$ are bounded and Lipschitz on $\mathbb{R}^{2d}$.

Remark 2.1 *In most biological applications, the trait space is a compact subset $\mathcal{X}$ of $\mathbb{R}^d$. However, the boundary of the trait space usually corresponds to deleterious traits. In other words, $g(y, x) \leq 0$ for all y in the boundary of $\mathcal{X}$. Therefore, assuming that the trait space is unbounded is not restrictive, since one can extend the fitness function to $\mathbb{R}^d$ in such a way that $g(y, x) \leq 0$ for all $y \notin \mathcal{X}$ and $x \in \mathbb{R}^d$. This amounts to add fictive traits, such that individuals holding these traits cannot live.*

2.2 The mutation law

The second biological parameter, $p(x, h)dh$, is the law of $h = y - x$, where y is a mutant trait born from an individual with trait x. For all $x \in \mathbb{R}^d$, we assume that this law is absolutely continuous with respect to Lebesgue's measure and that it is symmetrical with respect to 0 for simplicity. This is a very frequent assumption in adaptive dynamics models (see e.g. [15, 16, 29]).

We also assume that

(H2) $p(x,h)dh$ has finite and bounded third-order moment, and there exists a measurable function $m : \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$\int (\|h\|^2 \vee \|h\|^3) m(\|h\|) dh < +\infty, \text{ or equivalently } \int_{\mathbb{R}_+} (r^{d+1} \vee r^{d+2}) m(r) dr < +\infty,$$

where $\|\cdot\|$ is the standard Euclidean norm in $\mathbb{R}^d$, and for any $x, y \in \mathbb{R}^d$ and $h \in \mathbb{R}^d$,

$$|p(x,h) - p(y,h)| \leq \|x - y\| m(\|h\|) \quad \text{and} \quad p(x,h) \leq m(\|h\|). \tag{2.4}$$

We will denote by (H) the two assumptions (H1) and (H2).

Assumption (2.4) is satisfied for classical jump measures taken in applications. For example, it holds when $p(x,h)dh$ is Gaussian for all $x \in \mathbb{R}^d$, with covariance matrix $K(x)$ uniformly non-degenerate, bounded and Lipschitz on $\mathbb{R}^d$.

Assumption (H2) trivially implies the following property.

Lemma 2.2 *Assume (H2). Let $S = \mathbb{R}^d$ or $S = \{h : h \cdot u > 0\}$ for some $u \in \mathbb{R}^d \setminus \{0\}$, and let f be a function from $\mathbb{R}^d$ to $\mathbb{R}$ such that $f(0) = 0$ and*

$$\forall x, y \in \mathbb{R}^d, \ |f(x) - f(y)| \leq K \|x - y\| \max\{\|x\|, \|y\|, \|x\|^2, \|y\|^2\} \tag{2.5}$$

for some constant K. Then, the function $\phi(x) = \int_S f(h) p(x, dh)$ is globally Lipschitz on $\mathbb{R}^d$.

Note that, in the previous statement, since $f(0) = 0$, $|f(h)| \leq K(\|h\|^2 \vee \|h\|^3)$. Thus, the function ϕ is well-defined.

As a consequence of this result, (H2) also implies the following property, needed in the sequel to control the non-degeneracy of the matrix $a(x)$:

$$\forall \alpha > 0, \inf_{\|x\| \leq 1/\alpha, \ u,v \in \mathbb{R}^d, \ \|u\| = \|v\| = 1} \int_{\mathbb{R}^d} |h \cdot u|^2 |h \cdot v| p(x,h) dh > 0, \tag{2.6}$$

where $u \cdot v$ denotes the standard Euclidean inner product between u and $v \in \mathbb{R}^d$. Indeed, $\int_{\mathbb{R}^d} |h \cdot u|^2 |h \cdot v| p(x,h) dh$ is a continuous and positive function of (x, u, v). Therefore, its minimum on a compact set is positive.

Remark 2.3 *Lemma 2.2 is the only consequence of (H2) that will be used below. Assumption (H2) could be replaced by any condition ensuring this result. In particular, it would be sufficient to assume regularity of the probability measure $p(x,h)dh$ with respect to appropriate Kantorovich metrics [37]. See [5] for such conditions.*

2.3 The diffusion model of adaptive dynamics

The diffusion model of [8] is given in dimension 1. However, the computation of its parameters can be easily generalized to a multidimensional setting (see [5] for details). The parameters $b = (b_1, \dots, b_d)$, $\tilde{b} = (\tilde{b}_1, \dots, \tilde{b}_d)$ and $a = \sigma\sigma^* = (a_{kl})_{1\leq k,l\leq d}$, where * denotes the matrix transpose operator, are given by the following expressions: for all $x \in \mathbb{R}^d$,

$$\begin{aligned} b_k(x) &= \int_{\mathbb{R}^d} h_k[\nabla_1 g(x,x)\cdot h]_+ p(x,h)dh, \\ \tilde{b}_k(x) &= \frac{1}{2}\int_{\{h\cdot\nabla_1 g(x,x)>0\}} h_k(h^* H_{1,1}g(x,x)h)p(x,h)dh \\ \text{and} \quad a_{kl}(x) &= \int_{\mathbb{R}^d} h_k h_l [h\cdot\nabla_1 g(x,x)]_+ p(x,h)dh. \end{aligned} \tag{2.7}$$

We also define

$$b^\varepsilon = b + \varepsilon\tilde{b}.$$

and the matrix σ appearing in (1.1) as the unique real symmetrical positive $d \times d$ square root of a.

Observe that, for all $x \in \Gamma$, $a(x) = b(x) = \tilde{b}(x) = 0$. Thus, points of Γ are possible rest points of solutions of (1.1).

The regularity of these parameters is given in the following result.

Proposition 2.4 *Assume (H).*

(i) *a and b are globally Lipschitz and bounded on $\mathbb{R}^d$, and $\tilde{b}$ is bounded on $\mathbb{R}^d$ and locally Lipschitz on $\mathbb{R}^d \setminus \Gamma$.*

(ii) *The matrix a is symmetrical and non-negative on $\mathcal{X}$, $a(x) = 0$ if $x \in \Gamma$, and $a(x)$ is positive definite if $x \in \mathbb{R}^d \setminus \Gamma$. For all $\alpha > 0$, there exists $c > 0$ such that $\Gamma_\alpha \subset \{x \in \mathbb{R}^d,\ \forall s \in \mathbb{R}^d,\ s^* a(x)s \geq c\|s\|^2\}$, where Γ_α is defined in (2.3).*

(iii) *The symmetrical square root σ of a is bounded, Hölder with exponent $1/2$ on $\mathbb{R}^d$ and locally Lipschitz on $\mathbb{R}^d \setminus \Gamma$.*

Proof In all this proof, the constant C may change from line to line.

Let us start with Point (i). The functions a, b and $\tilde{b}$ are trivially bounded. Fix x and y in $\mathbb{R}^d$. For $1 \leq k \leq d$,

$$\begin{aligned} |b_k(x) - b_k(y)| \leq & \left| \int_{\mathbb{R}^d} h_k([\nabla_1 g(x,x)\cdot h]_+ - [\nabla_1 g(y,y)\cdot h]_+)p(x,h)dh \right| \\ & + \left| \int_{\mathbb{R}^d} h_k[\nabla_1 g(x,x)\cdot h]_+ (p(x,h) - p(y,h))dh \right|. \end{aligned}$$

Since $|[a]_+ - [b]_+| \leq |a - b|$ and $\nabla_1 g$ is Lipschitz, the first term of the right-hand side is less than $C\|x - y\|M_2$, where M_2 is a bound for the second-order moments of $p(x,h)dh$. Since the second term is equal to

$$\left| \int_{\{h\cdot\nabla_1 g(x,x)>0\}} h_k \nabla_1 g(x,x)\cdot h(p(x,h) - p(y,h))dh \right|,$$

Lemma 2.2 can be applied to bound this term by $C\|\nabla_1 g(x,x)\|\|x-y\|$. Since $\nabla_1 g$ is bounded, it follows that b is Lipschitz on $\mathbb{R}^d$. Similarly, a is Lipschitz on $\mathbb{R}^d$.

Take x and y in $\mathbb{R}^d \setminus \Gamma$ and let $S = \{h \in \mathbb{R}^d : h \cdot \nabla_1 g(x,x) > 0\}$ and $S' = \{h : h \cdot \nabla_1 g(y,y) > 0\}$. We also denote by S^c (resp. S'^c) the complement of S (resp. S') in $\mathbb{R}^d$. Then,

$$\begin{aligned} 2|\tilde{b}_k(x) - \tilde{b}_k(y)| \leq & \left| \int_{S\cap S'} h_k[h^*(H_{1,1}g(x,x) - H_{1,1}g(y,y))h]p(y,h)dh \right| \\ & + \left| \int_S h_k(h^* H_{1,1}g(x,x)h)(p(x,h) - p(y,h))dh \right| \\ & + \left| \int_{S\cap S'^c} h_k(h^* H_{1,1}g(x,x)h)p(y,h)dh \right| \\ & + \left| \int_{S^c\cap S'} h_k(h^* H_{1,1}g(y,y)h)p(y,h)dh \right|. \end{aligned} \tag{2.8}$$

By Lemma 2.2, the first two terms of the right-hand side are both bounded by $C\|x-y\|$ for some constant C. The third term can be bounded by

$$C \int_{S\cap S'^c} \|h\|^3 m(\|h\|)dh.$$

Making an appropriate spherical coordinates change of variables, this quantity can be bounded by

$$C\theta \int_{\mathbb{R}_+} r^{d+2} m(r)dr \leq C'\theta,$$

where θ is the angle between the vectors $\nabla_1 g(x,x)$ and $\nabla_1 g(y,y)$.

Now, fix $\alpha > 0$. For all $z \in \Gamma_\alpha$, $\nabla_1 g(z,z) \neq 0$. Therefore, $\beta := \inf_{z\in\Gamma_\alpha} \|\nabla_1 g(z,z)\| > 0$. Let K be such that $\nabla_1 g(x,x)$ is K-Lipschitz and let $u = \nabla_1 g(x,x)/\|\nabla_1 g(x,x)\|$ and $v = \nabla_1 g(y,y)/\|\nabla_1 g(y,y)\|$. Then

$$\|u-v\| \leq \frac{\|\nabla_1 g(x,x) - \nabla_1 g(y,y)\|}{\|\nabla_1 g(x,x)\|} + \|\nabla_1 g(y,y)\| \left| \frac{1}{\|\nabla_1 g(x,x)\|} - \frac{1}{\|\nabla_1 g(y,y)\|} \right| \leq \frac{K\|x-y\|}{\beta}.$$

Now, on the one hand $\sin(\theta/2) = \|u-v\|/2$ and on the other hand, $\sin x \geq (2\sqrt{2}/\pi)x$ for all $0 \leq x \leq \pi/4$. Therefore, for any $x, y \in \Gamma_\alpha$ such that $\|x-y\| \leq \sqrt{2}\beta/K$, we have

$$\theta \leq \frac{K\pi}{2\beta}\|x-y\|.$$

Therefore, for any $x, y \in \Gamma_\alpha$ such that $\|x-y\| \leq \sqrt{2}\beta/K$,

$$\left| \int_{S\cap S'^c} h_k(h^* H_{1,1}g(x,x)h)p(y,h)dh \right| \leq C_\alpha \|x-y\|,$$

where the constant C_α depends only on α. Proceeding as before for the last term of (2.8), we obtain that $\tilde{b}$ is uniformly Lipschitz on any convex compact subset of $\mathbb{R}^d \setminus \Gamma$, ending the proof of Point (i).

Concerning Point (ii), a is obviously symmetrical, and for any $s = (s_1, \dots, s_d) \in \mathbb{R}^d$, using the symmetry of $p(x, h)dh$,

$$\begin{aligned} s^* a(x) s &= \int_{\mathbb{R}^d} (h \cdot s)^2 [h \cdot \nabla_1 g(x, x)]_+ p(x, h) dh \\ &= \frac{1}{2} \int_{\mathbb{R}^d} (h \cdot s)^2 |h \cdot \nabla_1 g(x, x)| p(x, h) dh. \end{aligned}$$

This is non-negative for all $s \in \mathbb{R}^d$, and is non-zero if $s \neq 0$ and $x \notin \Gamma$.

Fix $\alpha > 0$, $x \in \Gamma_\alpha$, and $s = (s_1, \dots, s_d) \in \mathbb{R}^d$. We denote by u and v the unit vectors of $\mathbb{R}^d$ such that $s = \|s\| u$ and $\nabla_1 g(x, x) = \|\nabla_1 g(x, x)\| v$. Then

$$\begin{aligned} s^* a(x) s &= \frac{1}{2} \|s\|^2 \|\nabla_1 g(x, x)\| \int_{\mathbb{R}^d} |h \cdot u|^2 |h \cdot v| p(x, dh) \\ &\geq C_\alpha \|s\|^2 \|\nabla_1 g(x, x)\| \end{aligned} \tag{2.9}$$

where $C_\alpha > 0$ by (2.6). Since Γ_α is a compact subset of $\mathbb{R}^d$, we also have $\inf_{x \in \Gamma_\alpha} \|\nabla_1 g(x, x)\| > 0$, completing the proof of Point (ii).

Finally, Point (iii) follows from the facts that a is globally Lipschitz on $\mathbb{R}^d$ and that the symmetric square root function on the set of symmetric positive $d \times d$ matrices is globally 1/2-Hölder, and Lipschitz in $\{a \in \mathcal{S}_+ : \forall s \in \mathbb{R}^d,\ s^* a s \geq c\|s\|^2\}$ for any $c > 0$. A proof of these facts can be found for example in [41]. $\square$

2.4 A biological example

As explained in the introduction, the SDE (1.1) is a generalization of a classical evolutionary model, called the "canonical equation of adaptive dynamics" [15] $\dot{x}(t) = b(x(t))$ which describes the dynamics of the dominant trait of a population.

The situation we want to address in this article is the case where this ODE has several stable steady states (each of them having disjoint basins of attractions) and where the small diffusion part in (1.1) implies a change of basin of attraction for the population dominant trait (see Section 5). From a macroscopic viewpoint, this corresponds to a quick "jump" of the population state from (a neighborhood of) a stable steady state to (a neighborhood of) another one. This biological phenomenon is called *punctualism* [38].

Many biological works have described models where the canonical equation of adaptive dynamics has several stable steady states (see e.g. [25] in the context of seed size evolution, [33, 14] in the context of predator-prey interactions or [32] in the context of multispecific asymmetric competition). Since they include ecological specificities of the populations under consideration, these models are often quite complicated. To keep the presentation clear, we are going to illustrate our results with an bistable and multidimensional extension of a very classical and simple ecological model of competition for resources of Roughgarden [40], which was adapted to the framework of adaptive dynamics in [16].

We assume that the population is characterized by a d-dimensional trait $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ and is subject to competition for resources which are not exploited uniformly by different traits (for example, when $d = 2$, one could think of a population of birds that eat seeds, and that x_1 represents the beak size of a bird and x_2 the beak strength: a bird with bigger and stronger beak can eat bigger and harder seeds). We assume that the reproductive efficiency of an individual, which is a consequence of its resource consumption, depends on its trait value as

$$\lambda(x) = \exp\left(-\frac{\|x-e_1\|^2\|x+e_1\|^2}{2\sigma_b^2}\right),$$

where $e_1 = (1, 0, \dots, 0) \in \mathbb{R}^d$. Thus, there are two optimal trait values at e_1 and $-e_1$. The parameter $\sigma_b > 0$ controls the distance around these points where $\lambda(x)$ is of order 1. We also assume that the competition exerted by individuals of trait y on individuals of trait x is of logistic type and is given by the competition kernel

$$\alpha(x, y) = \exp\left(-\frac{\|x-y\|^2}{2\sigma_\alpha^2}\right)$$

(for precise definitions and details, see e.g. [36, 6]). Thus, competition for resources is symmetric, and the competition range is given by the parameter $\sigma_\alpha > 0$. Finally, as often done in biological models (see e.g. [16]), we assume that the mutation law $p(x, h)dh$ is Gaussian with mean 0 and with covariance matrix $\rho^2\mathrm{Id}$ for all $x \in \mathbb{R}^2$, where Id is the d-dimensional identity matrix. Without loss of generality, we can (and will) assume that $\rho = 1$. These parameters clearly satisfy Assumptions (H1) and (H2).

The fitness function is obtained from these parameters as (see e.g. [36, 6])

$$g(y, x) = \lambda(y) - \frac{\alpha(y, x)}{\alpha(x, x)}\lambda(x).$$

Thus,

$$\nabla_1 g(x, x) = \nabla\lambda(x) = -\frac{2\lambda(x)}{\sigma_b^2}\Big(x_1(\|x\|^2 - 1),\ x_2(\|x\|^2 + 1), \dots, x_d(\|x\|^2 + 1)\Big)^* \tag{2.10}$$

and it follows from elementary computations using the invariance of the Gaussian law with respect to rotations that $b(x) = \nabla_1 g(x, x)/2 = \nabla\lambda(x)/2$.

In particular $\Gamma = \{-e_1, 0, e_1\}$. Moreover, $-\lambda(x)$ is a strict Lyapunov function (see e.g. [12]) for the canonical equation of adaptive dynamics. It is then elementary to prove that the canonical equation admits no limit cycle and that every solution converges when time goes to inifinity to a point of Γ. More precisely, for symmetry reasons, any solution to the canonical equation starting in $(-\infty, 0) \times \mathbb{R}^{d-1}$ converges to $-e_1$, any solution starting from $(0, +\infty) \times \mathbb{R}^{d-1}$ converges to e_1, and any solution starting from $\{0\} \times \mathbb{R}^{d-1}$ converges to 0.

Elementary computations give

$$a(x) = \frac{\|\nabla_1 g(x, x)\|}{\sqrt{2\pi}}\mathrm{Id} + \frac{1}{\sqrt{2\pi}\|\nabla_1 g(x, x)\|}\nabla_1 g(x, x)\nabla_1 g(x, x)^*, \tag{2.11}$$

where $\|\nabla_1 g(x, x)\| = 2\lambda(x)\|x\|\,\|x - e_1\|\,\|x + e_1\|/\sigma_b^2$. The matrix $a(x)$ is clearly uniformly elliptic away from evolutionary singularities.

The computation of $\tilde{b}(x)$ is more tricky. First,

$$H_{11}g(x, x) = \left(\frac{1}{\sigma_\alpha^2} - \frac{2(\|x\|^2 + 1)}{\sigma_b^2}\right)\lambda(x)\mathrm{Id} + \frac{1}{\lambda(x)}\nabla_1 g(x, x)\nabla_1 g(x, x)^* - \frac{4\lambda(x)}{\sigma_b^2}(xx^* - e_1 e_1^*). \tag{2.12}$$

The three terms of the right-hand side give three parts for $\tilde{b}(x)$, denoted $\tilde{b}^{(1)}(x)$, $\tilde{b}^{(2)}(x)$ and $\tilde{b}^{(3)}(x)$. It can be easily deduced from the invariance of the Gaussian law with respect to rotation that

$$\tilde{b}^{(1)}(x) = \frac{(d+1)\lambda(x)}{2\sqrt{2\pi}} \left(\frac{1}{\sigma_\alpha^2} - \frac{2(\|x\|^2+1)}{\sigma_b^2} \right) \frac{\nabla_1 g(x,x)}{\|\nabla_1 g(x,x)\|} \tag{2.13}$$

and

$$\tilde{b}^{(2)}(x) = \frac{\|\nabla_1 g(x,x)\|}{\sqrt{2\pi}\lambda(x)} \nabla_1 g(x,x). \tag{2.14}$$

Now, assume that $\nabla_1 g(x,x) \neq 0$ and let R be an orthogonal matrix such that $R\nabla_1 g(x,x) = \|\nabla_1 g(x,x)\| e_1$. Then

$$\begin{aligned} R\tilde{b}^{(3)}(x) &= -\frac{2\lambda(x)}{\sigma_b^2} \int_{\{h_1>0\}} h(h^* M(x) h) p(x,h) dh \\ &= -\frac{2\lambda(x)}{\sigma_b^2} \Big(\|x\|^2 - 1 + M_{11}(x),\ 2M_{12}(x), \dots,\ 2M_{1d}(x) \Big)^*, \end{aligned}$$

where $M(x) = (M_{ij}(x))_{1\leq i,j\leq d} = R(xx^* - e_1 e_1^*)R^*$ and where we used that $\mathrm{Tr}(M(x)) = \mathrm{Tr}(xx^* - e_1e_1^*) = \|x\|^2 - 1$, where Tr denotes the trace operator on square matrices.

In particular, $\tilde{b} = \tilde{b}^{(1)} + \tilde{b}^{(2)} + \tilde{b}^{(3)}$ is clearly bounded in $\mathbb{R}^d$ and continuous in $\mathbb{R}^2 \setminus \Gamma$. We will actually prove in Section 3.5 that $\tilde{b}$ is always discontinuous at the points of Γ.

3 Strong existence, pathwise uniqueness and strong Markov property

Our goal in this section is to construct a particular, strong Markov solution of the SDE (1.1), identify the difficulty for pathwise uniqueness and give some conditions solving this difficulty, both in the one-dimensional case and the general case.

We fix $\varepsilon > 0$ until the end of this section.

3.1 Strong existence and pathwise uniqueness: construction of a particular solution of (1.1)

Proposition 3.1 *Assume (H). For any filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}, W)$ equipped with a d-dimensional standard Brownian motion W, and for any $x \in \mathbb{R}^d$, there exists a $\mathcal{F}_t$-adapted process $X^{\varepsilon,x}$ on Ω a.s. solution of (1.1) with initial state x, such that $X_t^{\varepsilon,x}$ is constant after τ, where*

$$\tau = \inf\{t \geq 0 : X_t^{\varepsilon,x} \in \Gamma\}. \tag{3.1}$$

Moreover, this process is the unique solution of (1.1) up to indistinguishability satisfying $X_t^{\varepsilon,x} = X_\tau^{\varepsilon,x}$ for all $t \geq \tau$ a.s.

Proof By Proposition 2.4, the functions $\tilde{b}$ and σ are bounded and locally Lipschitz on $\mathbb{R}^d \setminus \Gamma$. Moreover, b is bounded and globally Lipschitz on $\mathbb{R}^d$.

Assume that $x \notin \Gamma$ and fix $\alpha > 0$ such that $x \in \Gamma_\alpha$. Since Γ_α is a compact subset of $\mathbb{R}^d \setminus \Gamma$. one can construct $\tilde{b}^\alpha$ (resp. σ^α) an extension to $\mathbb{R}^d$ of $\tilde{b}$ (resp. σ) restricted to Γ_α

such that $\tilde{b}^\alpha$ (resp. σ^α) is bounded and globally Lipschitz on $\mathbb{R}^d$ (resp. bounded, globally Lipschitz and uniformly non-degenerate on $\mathbb{R}^d$). Then, strong existence and pathwise uniqueness for the SDE

$$d\tilde{X}_t^{\varepsilon,\alpha} = (b(\tilde{X}_t^{\varepsilon,\alpha}) + \varepsilon\tilde{b}^\alpha(\tilde{X}_t^{\varepsilon,\alpha}))dt + \sqrt{\varepsilon}\sigma^\alpha(\tilde{X}_t^{\varepsilon,\alpha})dW_t \tag{3.2}$$

with initial condition $\tilde{X}_0^{\varepsilon,\alpha} = x$ are well-known results. Let

$$\tau_\alpha = \inf\{t \geq 0 : \tilde{X}_t^{\varepsilon,\alpha} \notin \Gamma_\alpha\}.$$

By pathwise uniqueness, for any $\alpha, \alpha' > 0$, $\tilde{X}_t^{\varepsilon,\alpha} = \tilde{X}_t^{\varepsilon,\alpha'}$ for all $t \leq \tau_\alpha \wedge \tau_{\alpha'}$ a.s. Therefore, the process $X^{\varepsilon,x}$ defined by $X_t^{\varepsilon,x} = \tilde{X}_t^{\varepsilon,\alpha}$ for $t \leq \tau_\alpha$ is a solution of (1.1) for $t < \sup_{\alpha>0}\tau_\alpha = \tau$.

On the event $\{\tau = +\infty\}$, this gives a strong solution of (1.1). On the event $\{\tau < \infty\}$, as a solution to (1.1), the semimartingale $(X_t^{\varepsilon,x}, t < \tau)$ has a uniformly Lipschitz finite variation part (since b^ε is bounded), and a local martingale part which is uniformly in L^2, and thus uniformly integrable, on finite time intervals (since σ is bounded). Therefore, on the event $\{\tau < \infty\}$, the random variable

$$X_\tau^{\varepsilon,x} := \lim_{t\uparrow\tau} X_t^{\varepsilon,x}$$

is a.s. well-defined and finite. Since $b(x) = \tilde{b}(x) = \sigma(x) = 0$ for all $x \in \Gamma$, defining $X_t^{\varepsilon,x} = X_\tau^{\varepsilon,x}$ for $t \geq \tau$. provides a strong solution of (1.1).

In the case where $x \in \Gamma$, setting $X_t^{\varepsilon,x} = x$ for all $t \geq 0$ trivially provides a strong solution of (1.1).

Now, by pathwise uniqueness for (3.2), there is pathwise uniqueness for (1.1) until time τ. Therefore, the process $X^{\varepsilon,x}$ we constructed above is the unique solution of (1.1) constant after time τ. □

The following result is a trivial consequence of the previous one.

Proposition 3.2 *With the same assumption and notation as in Proposition 3.1, assume that, for some $x \in \mathbb{R}^d \setminus \Gamma$,*

$$\mathbb{P}(X_t^{\varepsilon,x} \notin \Gamma,\ \forall t \geq 0) = \mathbb{P}_x(\tau = \infty) = 1, \tag{3.3}$$

where $\mathbb{P}_x$ is the law of $X^{\varepsilon,x}$. Then, pathwise uniqueness holds for (1.1) with initial state x.

The question whether pathwise uniqueness also holds for the whole trajectory when it can hit Γ in finite time is difficult. Because of the singularities of our diffusion ($\tilde{b}$ discontinuous and σ degenerate and non-Lipschitz), no standard technique apply in dimension two or more. In the one-dimensional case, general criterions of Engelbert and Schmidt exist on pathwise uniqueness (see [28]). However, the nature of our singularity corresponds precisely to a situation where the criterion does not allow to conclude. The combination of our singularities is also incompatible with classical results about uniqueness in law.

Therefore, it is desirable to have conditions ensuring (3.3) or its converse. This is done is Sections 3.3 and 3.4. These results will also be useful in Section 5.

3.2 Strong Markov property

The strong Markov property for solutions of SDEs is known to be linked to the uniqueness of solutions to the corresponding martingale problem. Here, we cannot prove uniqueness in general, but the strong Markov property can be easily proved.

Proposition 3.3 *Assume (H). Then the family $(X^{\varepsilon,x})_{x\in\mathbb{R}^d}$ of solutions of (1.1) constructed in Proposition 3.1 satisfy the strong Markov property.*

Proof Let x be a fixed point of $\mathbb{R}^d$, S be a $\mathcal{F}_t$-stopping time and φ be a bounded and continuous function from $\mathbb{R}^d$ to $\mathbb{R}$. We want to prove that

$$\mathbb{E}(\varphi(X^{\varepsilon,x}_{S+t}) \mid \mathcal{F}_S) = \mathbb{E}(\varphi(X^{\varepsilon,x}_{S+t}) \mid X^{\varepsilon,x}_S).$$

Since $X^{\varepsilon,x}_t$ is constant after time τ, this is equivalent to the existence of a Lebesgue-measurable function $f:\mathbb{R}^d\to\mathbb{R}$ such that

$$\mathbb{E}(\varphi(X^{\varepsilon,x}_{(S+t)\wedge\tau}) \mid \mathcal{F}_S) = f(X^{\varepsilon,x}_S).$$

Recall the definition of τ_α and $\tilde{X}^{\varepsilon,\alpha}$ with initial condition x in the proof of Proposition 3.1. Since there is strong existence and pathwise uniqueness for (3.2), the strong Markov property holds for $\tilde{X}^{\varepsilon,\alpha}$ [28, Thm. 5.4.20]. Therefore, for any $\alpha>0$, there is a bounded Lebesgue-measurable function f_α such that

$$\mathbb{E}(\mathbf{1}_{\tau_\alpha>S}\varphi(\tilde{X}^{\varepsilon,\alpha}_{(S+t)\wedge\tau_\alpha}) \mid \mathcal{F}_S) = \mathbf{1}_{\tau_\alpha>S} f_\alpha(\tilde{X}^{\varepsilon,\alpha}_S).$$

Since $\tilde{X}^{\alpha,\varepsilon}_t = X^{\varepsilon,x}_t$ for all $t\leq\tau_\alpha$, this yields

$$\mathbb{E}(\mathbf{1}_{\tau_\alpha>S}\varphi(X^{\varepsilon,x}_{(S+t)\wedge\tau_\alpha}) \mid \mathcal{F}_S) = \mathbf{1}_{\tau_\alpha>S} f_\alpha(X^{\varepsilon,x}_S).$$

Observing that $\mathbf{1}_{\tau>S} = \mathbf{1}_{X^{\varepsilon,x}_S\not\in\Gamma}$ is $\sigma(X^{\varepsilon,x}_S)$-measurable, we deduce that

$$\mathbb{E}(\mathbf{1}_{\tau_\alpha>S}\varphi(X^{\varepsilon,x}_{(S+t)\wedge\tau_\alpha}) \mid \mathcal{F}_S) + \mathbf{1}_{\tau>S\geq\tau_\alpha} f_\alpha(X^{\varepsilon,x}_S)$$

is $\sigma(X^{\varepsilon,x}_S)$-measurable for all $\alpha>0$. Letting α go to 0, it follows from Lebesgue's theorem for conditional expectations that this random variable (in short, r.v.) a.s. converges to $\mathbb{E}(\mathbf{1}_{\tau>S}\varphi(X^{\varepsilon,x}_{(S+t)\wedge\tau})|\mathcal{F}_S)$. As an a.s. limit of $\sigma(X^{\varepsilon,x}_S)$-measurable r.v., this r.v. is also $\sigma(X^{\varepsilon,x}_S)$-measurable.

Now,

$$\mathbb{E}(\mathbf{1}_{\tau\leq S}\varphi(X^{\varepsilon,x}_{(S+t)\wedge\tau}) \mid \mathcal{F}_S) = \mathbb{E}(\mathbf{1}_{\tau\leq S}\varphi(X^{\varepsilon,x}_S) \mid \mathcal{F}_S) = \mathbf{1}_{X^{\varepsilon,x}_S\in\Gamma}\varphi(X^{\varepsilon,x}_S),$$

which is also $\sigma(X^{\varepsilon,x}_S)$-measurable. This ends the proof of Proposition 3.3. □

3.3 Study of $\mathbb{P}(\tau = \infty)$: the case of dimension 1

As we saw above, the uniqueness of $X^{\varepsilon,x}$ relies on the fact that $\mathbb{P}_x(\tau = \infty) = 1$, where τ has been defined in (3.1) and where $\mathbb{P}_x$ is the law of $X^{\varepsilon,x}$. Our goal in this section and the following one is to give conditions under which this is true (or false).

In this section, we assume that $d = 1$. In this case, an elementary calculation gives the following formulas for a, b and $\tilde{b}$:

$$b(x) = \frac{M_2(x)}{2}\partial_1 g(x,x), \quad \tilde{b}(x) = \frac{M_3(x)}{4}\text{sign}[\partial_1 g(x,x)]\partial^2_{1,1}g(x,x),$$
$$a(x) = \frac{M_3(x)}{2}|\partial_1 g(x,x)|, \quad \text{where} \quad M_k(x) = \int_{\mathbb{R}} |h|^k p(x,h)dh$$
$$\text{and} \quad \text{sign}(x) = -1 \text{ if } x < 0;\ 0 \text{ if } x = 0;\ 1 \text{ if } x > 0.$$

In the following result, we use the fact that $\partial^2_{1,1}g(x,x) + 2\partial^2_{1,2}g(x,x) + \partial^2_{2,2}g(x,x) = 0$ for all $x \in \mathbb{R}$, which follows from differentiation of (2.1).

Theorem 3.4 *Assume (H). Assume also that $d = 1$ and g is $\mathcal{C}^3$ with bounded third-order derivatives. Let $x \notin \Gamma$ and define $c = \sup\{y \in \Gamma, y < x\}$, $c' = \inf\{y \in \Gamma, y > x\}$, and assume that $-\infty < c < c' < \infty$, $\partial^2_{1,1}g(c,c) + \partial^2_{1,2}g(c,c) \neq 0$ and $\partial^2_{1,1}g(c',c') + \partial^2_{1,2}g(c',c') \neq 0$. We can define*

$$\begin{aligned}
\alpha &:= \frac{\partial^2_{1,1}g(c,c)}{\partial^2_{1,1}g(c,c) + \partial^2_{1,2}g(c,c)} = \frac{2\partial^2_{1,1}g(c,c)}{\partial^2_{1,1}g(c,c) - \partial^2_{2,2}g(c,c)} \\
\beta &:= \frac{\partial^2_{1,1}g(c',c')}{\partial^2_{1,1}g(c',c') + \partial^2_{1,2}g(c',c')} = \frac{2\partial^2_{1,1}g(c',c')}{\partial^2_{1,1}g(c',c') - \partial^2_{2,2}g(c',c')}.
\end{aligned} \tag{3.4}$$

(a) *If $\alpha \geq 1$ and $\beta \leq -1$, then $\mathbb{P}_x(\tau = \infty) = 1$ and the process $X^{\varepsilon,x}$ is recurrent in (c,c').*

(b) *If $\alpha \geq 1$ and $\beta > -1$, then $\mathbb{P}_x(\tau < \infty) = 1$ and $\mathbb{P}(\lim_{t\to\tau} X^{\varepsilon,x}_t = c') = 1$.*

(c) *If $\alpha < 1$ and $\beta \leq -1$, then $\mathbb{P}_x(\tau < \infty) = 1$ and $\mathbb{P}(\lim_{t\to\tau} X^{\varepsilon,x}_t = c) = 1$.*

(d) *If $\alpha < 1$ and $\beta > -1$, then $\mathbb{P}_x(\tau < \infty) = 1$ and*
$\mathbb{P}(\lim_{t\to\tau} X^{\varepsilon,x}_t = c) = 1 - \mathbb{P}(\lim_{t\to\tau} X^{\varepsilon}_t = c') \in (0,1)$.

Remarks 3.5

- *When $c = -\infty$ or $c' = \infty$, the calculation below depends on the precise behaviour of g and M_k near infinity, and no simple general result can be stated.*
- *The biological theory of adaptive dynamics gives a classification of evolutionary singularities in dimension $d = 1$, depending on the values of $\partial^2_{1,1}g$ and $\partial^2_{2,2}g$ at these points. Here, the condition $\alpha \geq 1$ corresponds, when $\partial^2_{1,1}g(c,c) - \partial^2_{2,2}g(c,c) > 0$, to the case $\partial^2_{1,1}g(c,c) + \partial^2_{2,2}g(c,c) \geq 0$, which corresponds in the biological terminology (see* e.g. *Diekmann [17]) to a converging stable strategy with mutual invasibility, which include the evolutionary branching condition; and when $\partial^2_{1,1}g(c,c) - \partial^2_{2,2}g(c,c) < 0$, to the case $\partial^2_{1,1}g(c,c) + \partial^2_{2,2}g(c,c) \leq 0$, which corresponds biologically to a repelling strategy without mutual invasibility.*

Proof We will use the classical method of removal of drift of Engelbert and Schmidt and the explosion criterion of Feller (see e.g. [28]), which can be applied to $X^{\varepsilon,x}$, considered as a process with value in (c,c') killed when it hits c or c'. These methods involve the two following functions, defined for a fixed $\gamma \in (c,c')$:

$$\begin{aligned} p(x) &= \int_\gamma^x \exp\left[-2\int_\gamma^y \frac{b^\varepsilon(z)dz}{\varepsilon\sigma^2(z)}\right]dy, \ \forall x \in (c,c'), \\ \text{and} \quad v(x) &= \int_\gamma^x p'(y)\int_\gamma^y \frac{2dz}{\varepsilon p'(z)\sigma^2(z)}dy, \ \forall x \in (c,c'). \end{aligned} \tag{3.5}$$

Then [28, pp. 345–351], the statements about the limit of the process X_t^ε when $t \to \tau$ and about the recurrence of X^ε depend on whether $p(x)$ is finite or not when $x \to c$ and c', and the statements about τ depends on whether $v(x)$ is finite or not when $x \to c$ and c'.

Let us compute these limits. We will use the standard notation $f(x) = o(g(x))$ (resp. $f(x) = O(g(x))$, resp. $f(x) \sim g(x)$) when $x \to a$, if $f(x)/g(x) \to 0$ when $x \to a$ (resp. $|f(x)| \leq Cg(x)$ for some constant C in a neighborhood of a, resp. $f(x)/g(x) \to 1$ when $x \to a$).

$$\frac{b^\varepsilon(x)}{\varepsilon\sigma^2(x)} = \frac{b^\varepsilon(x)}{\varepsilon a(x)} = \frac{M_2(x)}{\varepsilon M_3(x)}\mathrm{sign}[\partial_1 g(x,x)] + \frac{1}{2}\frac{\partial^2_{1,1}g(x,x)}{\partial_1 g(x,x)}, \tag{3.6}$$

so, for $x < y < \gamma$, the quantity inside the exponential appearing in the definition of p is

$$\int_y^\gamma \frac{2M_2(z)}{\varepsilon M_3(z)}\mathrm{sign}[\partial_1 g(z,z)]dz + \int_y^\gamma \frac{\partial^2_{1,1}g(z,z)}{\partial_1 g(z,z)}dz.$$

Since $c \neq -\infty$, the first term is bounded for $c < y < \gamma$ (by Assumption (H), M_3 is positive and continuous on $[c,c']$), so we only have to study the second term.

When $y \to c$, an easy calculation gives

$$\frac{\partial^2_{1,1}g(z,z)}{\partial_1 g(z,z)} = \frac{\alpha}{z-c} + C + o(1),$$

where α is defined in (3.4), and where C is a constant depending on the derivatives of g at (c,c) up to order 3. Consequently, when $y \to c$,

$$\begin{aligned} \exp\left[-2\int_\gamma^y \frac{b^\varepsilon(z)dz}{\varepsilon\sigma^2(z)}\right] &= \exp\left[C' + o(1) + \int_y^\gamma \left(\frac{\alpha}{y-c} + C + o(1)\right)dz\right] \\ & e^{C''}(y-c)^{-\alpha}, \end{aligned} \tag{3.7}$$

as $y \to c$.

Therefore, if $\alpha < 1$, $p(c+) > -\infty$, and if $\alpha \geq 1$, $p(c+) = -\infty$. The same computation gives the same result when $x \to c'$, replacing α by β.

Now let us compute the limit of v at c and c'. Since $p(c'-) = \infty \Rightarrow v(c'-) = \infty$ and $p(c+) = -\infty \Rightarrow v(c+) = \infty$ [28, p. 348], we only have to deal with the cases $\alpha < 1$ and $\beta > -1$.

Equation (3.7) yields $p'(y) \sim e^C(y-c)^{-\alpha}$, so, for some constant C,

$$\frac{2}{\varepsilon p'(z)a(z)} \sim C(z-c)^{\alpha-1},$$

since

$$a(z) = \frac{M_3(z)}{2}|\partial_1 g(z,z)| \sim \frac{M_3(c)}{2}|\partial^2_{1,1}g(c,c) + \partial^2_{1,2}g(c,c)|(z-c).$$

If $\alpha < 0$, when $y \to c$, $p'(y)\int_y^\gamma \frac{2dz}{\varepsilon p'(z)a(z)} \sim -Cp'(y)(y-c)^\alpha$ is bounded on (c,γ), and so $v(c+) < \infty$. If $\alpha = 0$, $p'(y)\int_y^\gamma \frac{2dz}{\varepsilon p'(z)a(z)} \sim C\log(y-c)$, which has a finite integral on (c,γ), so $v(c+) < \infty$. Finally, if $0 < \alpha < 1$, $\int_y^\gamma \frac{2dz}{\varepsilon p'(z)a(z)}$ is bounded, so $v(c+) < \infty$ is equivalent to the convergence of the integral $\int_c^\gamma p'(y)dy$, which holds since $p'(y) \sim \frac{C}{(y-a)^\alpha}$ and $\alpha < 1$. $\square$

3.4 Study of $\mathbb{P}(\tau = \infty)$: the general case

Let us turn now to the case $d \geq 2$. The following result gives conditions under which $\mathbb{P}_x(\tau = \infty) = 1$, based on a comparison of $d(X^{\varepsilon,x}, \Gamma)$ with Bessel processes.

Theorem 3.6 *Assume (H). Assume also that g is $\mathcal{C}^2$ on $\mathbb{R}^{2d}$ and that the points of Γ are isolated. For any $y \in \Gamma$, let $\mathcal{U}_y$ be a neighborhood of y and $a^y > 0$ and $a_y > 0$ two constants such that a is a^y-Lipschitz on $\mathcal{U}_y$ and $s^*a(x)s \geq a_y\|s\|^2\|x-y\|$ for all $x \in \mathcal{U}_y$ and $s \in \mathbb{R}^d$. Define also*

$$\tilde{b}_y = \inf_{x \in \mathcal{U}_y \setminus \{y\}} \frac{x-y}{\|x-y\|} \cdot \tilde{b}(x)$$
$$\textit{and} \quad \tilde{b}^y = \sup_{x \in \mathcal{U}_y \setminus \{y\}} \frac{x-y}{\|x-y\|} \cdot \tilde{b}(x).$$

(a) *If for any $y \in \Gamma$, $\frac{\tilde{b}_y + da_y/2}{a^y} \geq 1$, then, for all $x \notin \Gamma$, $\mathbb{P}_x(\tau = \infty) = 1$ and $\mathbb{P}(\lim_{t\to+\infty} X_t^{\varepsilon,x} \in \Gamma) = 0$.*

(b) *If there exists $y \in \Gamma$ such that $\frac{\tilde{b}^y + da^y/2}{a_y} < 1$, then, for all $x \notin \Gamma$, $\mathbb{P}(\lim_{t\to\tau} X_t^{\varepsilon,x} = y) > 0$.*

Before proving Theorem 3.6, let us give some bounds for the constants involved in this Theorem. This result makes use of the notation $B(x,r)$ for the open Euclidean ball of $\mathbb{R}^d$ centered at x with radius r.

Proposition 3.7 *Assume (H). Assume also that g is $\mathcal{C}^2$ on $\mathbb{R}^{2d}$ and that the points of Γ are isolated. Fix $y \in \Gamma$ and $\alpha > 0$ such that $B(y,\alpha) \cap \Gamma = \{y\}$. Define*

$$C = \inf_{u,v \in \mathbb{R}^d : \|u\|=\|v\|=1} \int |h \cdot u|^2 |h \cdot v| p(x,h)dh.$$

$C > 0$ by (2.6). Let M_3 be a bound for the third-order moment of $p(x,h)dh$ on $B(y,\alpha)$. Let $D = H_{1,1}g(y,y) + H_{1,2}g(y,y)$, and denote by λ^y (resp. λ_y) the greatest (resp. the smallest)

eigenvalue of D^*D. Suppose also that D is invertible ($\lambda_y > 0$). Then, for any $\delta > 0$ there exists a neighborhood $\mathcal{U}_y$ of y such that, in the statement of Theorem 3.6, we can take

$$a^y = M_3\sqrt{\lambda^y} + \delta, \quad a_y = C\sqrt{\lambda_y} - \delta,$$
$$\tilde{b}^y < \frac{M_3}{2}\|H_{1,1}g(y,y)\| + \delta \quad \text{and} \quad \tilde{b}_y > -\frac{M_3}{2}\|H_{1,1}g(y,y)\| - \delta.$$

Proof It follows from the definition (2.7) of $\tilde{b}$ that for $x \neq y$,

$$\frac{x-y}{\|x-y\|}\cdot\tilde{b}(x) = \int_{\{\nabla_1 g(x,x)\cdot h>0\}}\left(\frac{x-y}{\|x-y\|}\cdot h\right)(h^* H_{1,1}g(x,x)h)p(x,h)dh. \tag{3.8}$$

By assumption, the quantity inside the integral can be bounded by $\|h\|^3[\|H_{1,1}g(y,y)\| + O(\|x-y\|)]p(x,h)$. Therefore,

$$\begin{aligned}\frac{x-y}{\|x-y\|}\cdot\tilde{b}(x) &\leq [\|H_{1,1}g(y,y)\| + O(\|x-y\|)]\int_{\{\nabla_1 g(x,x)\cdot h>0\}}\|h\|^3 p(x,h)dh \\ &= \frac{M_3}{2}[\|H_{1,1}g(y,y)\| + O(\|x-y\|)].\end{aligned}$$

This gives the required bounds for $\tilde{b}^y$ and $\tilde{b}_y$.

It follows from equation (2.9) in the proof of Proposition 2.4, that, for all $s \in \mathbb{R}^d$ and $x \in \mathbb{R}^d$

$$C\|s\|^2\|\nabla_1 g(x,x)\| \leq s^* a(x) s \leq M_3\|s\|^2\|\nabla_1 g(x,x)\|.$$

Considering an orthonormal basis of $\mathbb{R}^d$ in which D^*D is diagonal, one has $\lambda_y\|v\|^2 \leq \|Dv\|^2 = v\cdot D^*Dv \leq \lambda^y\|v\|^2$ for any $v \in \mathbb{R}^d$. It remains to observe that $\nabla_1 g(x,x) \sim D(x-y)$ when $x \to y$ to obtain the required bounds for a^y and a_y. $\square$

Proof of Theorem 3.6 Fix $y \in \Gamma$. Let us assume for convenience that $y = 0$. By assumption, to this point of Γ is associated a neighborhood $\mathcal{U}_0$ of 0 and four constants $a_0 > 0$, $a^0 > 0$, $\tilde{b}_0$ and $\tilde{b}^0$. Let ρ be small enough for $B(\rho) := \{x \in \mathbb{R}^d : \|x\| \leq \rho\} \subset \mathcal{U}_0$ and $\Gamma \cap B(\rho) = \{0\}$, and define $\tau_\rho := \inf\{t \geq 0 : \|X^\varepsilon_t\| = \rho\}$ and $\tau_0 = \inf\{t \geq 0 : X^\varepsilon_t = 0\}$, where we omitted the dependence of $X^{\varepsilon,x}$ with respect to the initial condition. Recall also the notation $\mathbb{P}_x$ for the law of X^ε when $X^\varepsilon_0 = x$.

Theorem 3.6 can be deduced from the next lemma.

Lemma 3.8
(a) *If $\frac{\tilde{b}_0 + da_0/2}{a^0} \geq 1$, then, for all $x \in B(\rho)\setminus\{0\}$, $\mathbb{P}_x(\tau_\rho \leq \tau_0) = 1$.*

(b) *If $\frac{\tilde{b}^0 + da^0/2}{a_0} < 1$, then, there exists a constant $c > 0$ such that, for all $x \in B(\rho/2)\setminus\{0\}$, $\mathbb{P}_x(\{\tau_0 < \tau_\rho\} \cup \{\tau_0 = \tau_\rho = \infty$ and $\lim_{t\to+\infty} X^\varepsilon_t = 0\}) \geq c$.*

Together with the strong Markov property of Proposition 3.3, Point (a) of this lemma easily implies Theorem 3.6 (a), and part (b) implies Theorem 3.6 (b) if we can prove that for any $x \in \mathbb{R}^d\setminus\Gamma$, $\mathbb{P}_x(\tau_{\rho/2} < \infty) > 0$. This can be done as follows.

Let D be any connected bounded open domain D with smooth boundary containing $B(\rho/2)$. The process $\tilde{X}^{\varepsilon,\alpha}$ of the proof of Proposition 3.1 has smooth and uniformly non-degenerate coefficients. Therefore, it is standard to prove that such a process hits $B(\rho/2)$

before hitting ∂D with positive probability, starting from any $y \in D$. (This may be proved for example by applying Feynman-Kac's formula to obtain the elliptic PDE solved by this probability in $D \setminus B(\rho/2)$, and next applying the strong maximum principle to this PDE.) Choosing α and D such that $x \in D$ and $D \setminus B(\rho/2) \subset \Gamma_\alpha$, we easily obtain the required estimate. $\square$

Before coming to the proof of Lemma 3.8, we need to introduce a few notation: it follows from Itô's formula that, for all $t < \tau$,

$$\|X_t^\varepsilon\| = \|x\| + \int_0^t \frac{1}{\|X_s^\varepsilon\|}\Bigg[X_s^\varepsilon \cdot (b(X_s^\varepsilon) + \varepsilon\tilde{b}(X_s^\varepsilon)) \\ + \frac{\varepsilon}{2}\mathrm{Tr}(a(X_s^\varepsilon)) - \frac{\varepsilon}{2}\frac{(X_s^\varepsilon)^*}{\|X_s^\varepsilon\|}a(X_s^\varepsilon)\frac{X_s^\varepsilon}{\|X_s^\varepsilon\|}\Bigg]ds + M_t,$$

where Tr is the trace operator on $d \times d$ matrices, and where, for $t < \tau$,

$$M_t := \sqrt{\varepsilon}\int_0^t \frac{(X_s^\varepsilon)^*}{\|X_s^\varepsilon\|}\sigma(X_s^\varepsilon)dW_s.$$

Let us extend M_t to $t \geq \tau$ by setting $M_t = M_{t\wedge\tau}$ for all $t \geq 0$. Since σ is bounded, M_t is a $\mathbb{L}^2$-martingale in $\mathbb{R}$ with quadratic variation

$$\langle M\rangle_t = \varepsilon\int_0^{t\wedge\tau} \frac{(X_s^\varepsilon)^*}{\|X_s^\varepsilon\|}a(X_s^\varepsilon)\frac{X_s^\varepsilon}{\|X_s^\varepsilon\|}ds. \tag{3.9}$$

It follows from Dubins-Schwartz's Theorem (see e.g. [28]) that for any $t \geq 0$, $M_t = B_{\langle M\rangle_t}$, where B is a one-dimensional Brownian motion.

Define the time change $T_t = \inf\{s \geq 0 : \langle M\rangle_s > t\}$ for all $t \geq 0$. If $t < \langle M\rangle_\infty := \lim_{t\to+\infty}\langle M\rangle_t = \langle M\rangle_\tau$, then $T_t < \infty$ and $\langle M\rangle_{T_t} = t$. For $t < \langle M\rangle_\infty$, define $Y_t = X_{T_t}^\varepsilon$. An easy change of variable shows that for $t < \langle M\rangle_\infty$, $Y_t \notin \Gamma$, and

$$\|Y_t\| = \|x\| + \int_0^t c(Y_s)ds + B_t,$$

where

$$c(z) = \|z\|\frac{z \cdot (b(z) + \varepsilon\tilde{b}(z)) + \varepsilon\mathrm{Tr}(a(z))/2}{\varepsilon z^* a(z) z} - \frac{1}{2\|z\|}.$$

Using the constants defined in the statement of Theorem 3.6, the fact that b is K-Lipschitz on $\mathbb{R}^d$, and the fact that $\mathrm{Tr}(a) = \sum_{i=1}^d e_i^* a e_i$, where e_i is the i^{th} vector of the canonical basis of $\mathbb{R}^d$, one easily obtains that, for all $z \in \mathcal{U}_0$,

$$c_1(\|z\|) < c(z) < c_2(\|z\|),$$

where, for $u > 0$,

$$c_1(u) = \left(\frac{da_0/2 + \tilde{b}_0}{a^0} - \frac{1}{2}\right)\frac{1}{u} - \frac{2K}{\varepsilon a_0}$$

$$\text{and} \quad c_2(u) = \left(\frac{da^0/2 + \tilde{b}^0}{a_0} - \frac{1}{2}\right)\frac{1}{u} + \frac{2K}{\varepsilon a_0}.$$

Define also the processes Z^1 and Z^2 strong solutions in $(0, \infty)$ to the SDEs

$$Z_t^i = \|x\| + \int_0^t c_i(Z_s^i) ds + B_t$$

for $i = 1, 2$, and stopped when they reach 0. As strong solutions, these processes can be constructed on the same probability space than X^ε (and Y). Finally, define for $1 \leq i \leq 2$ the stopping times

$$\theta_0^i = \inf\{t \geq 0 : Z^i = 0\}$$
$$\text{and} \quad \theta_\rho^i = \inf\{t \geq 0 : Z^i = \rho\}.$$

The proof of Lemma 3.8 relies on the following three lemmas. The first one is a comparison result between Z^1, Z^2 and Y.

Lemma 3.9 *Almost surely, $Z_t^1 \leq \|Y_t\|$ for all $t < \theta_\rho^1 \wedge \langle M \rangle_\infty$, and $\|Y_t\| \leq Z_t^2$ for all $t < \theta_\rho^2 \wedge \langle M \rangle_\infty$.*

The processes Z^1 and Z^2 are Bessel processes with additional drifts. The second lemma examines whether these processes hit 0 in finite time or not.

Lemma 3.10
(a) *Z^1 is recurrent in $(0, +\infty)$ if and only if $\frac{\tilde{b}_0 + da_0/2}{a^0} \geq 1$.*

(b) *Let $\mathbf{P}_u$ be the law of Z^2 with initial state $u > 0$. If $\frac{\tilde{b}^0 + da^0/2}{a_0} < 1$, then, for any $u < \rho$, $\mathbf{P}_u(\theta_0^2 < \theta_\rho^2) > 0$.*

The last lemma states that, when $\langle M \rangle_\infty < \infty$, X^ε reaches Γ in finite or infinite time.

Lemma 3.11 *$\{\langle M \rangle_\infty < \infty\} \subset \{\tau < \infty\} \cup \{\tau = \infty$ and $\lim_{t \to +\infty} X_t^\varepsilon \in \Gamma\}$ a.s.*

Proof of Lemma 3.8 Assume first that $\frac{\tilde{b}_0 + da_0/2}{a^0} \geq 1$, and fix $x \in B(\rho) \setminus \{0\}$. Then, by Lemma 3.10 (a), $\theta_0^1 = \infty$ and $\theta_\rho^1 < \infty$ a.s. Moreover, by Lemma 3.9, for all $t < T_{\theta_\rho^1}$, $\|X_t^\varepsilon\| = \|Y_{\langle M \rangle_t}\| \geq Z^1_{\langle M \rangle_t}$.

Then, $\langle M \rangle_\infty = \infty$ implies a.s. that there exists $t \geq 0$ such that $\langle M \rangle_t = \theta_\rho^1$ and thus $\tau_\rho < \tau_0$. Conversely, by Lemma 3.11, $\langle M \rangle_\infty < \infty$ implies a.s. that $\lim_{t \to \tau} X_t^\varepsilon \in \Gamma \setminus \{0\}$, and thus that $\tau_\rho < \tau_0$. This completes the proof of Lemma 3.8 (a).

Now, assume that $\frac{\tilde{b}^0 + da^0/2}{a_0} < 1$ and fix $x \in B(\rho/2)$. By Lemma 3.9, for all $t < T_{\theta_\rho^2}$, $\|X_t^\varepsilon\| = \|Y_{\langle M \rangle_t}\| \leq Z^2_{\langle M \rangle_t}$.

Then, on the event $\{\theta_0^2 < \theta_\rho^2\}$, $\langle M \rangle_\infty = \infty$ implies a.s. that $\tau_0 < \tau_\rho$. Conversely, on the event $\{\theta_0^2 < \theta_\rho^2\}$, by Lemma 3.11, $\langle M \rangle_\infty < \infty$ implies a.s. that $\lim_{t \to \tau} X_t^\varepsilon = 0$ (where τ may be finite or infinite), and thus that $\tau_0 < \tau_\rho$ or that $\tau_0 = \tau_\rho = \infty$ and $\lim_{t \to +\infty} X_t^\varepsilon = 0$. Hence,

$$\mathbb{P}_x(\{\tau_0 < \tau_\rho\} \cup \{\tau_0 = \tau_\rho = \infty \text{ and } \lim_{t \to +\infty} X_t^\varepsilon = 0\}) \geq \mathbf{P}_{\|x\|}(\theta_0^2 < \theta_\rho^2).$$

But, applying the Markov property to Z^2, $\mathbf{P}_{\|x\|}(\theta_0^2 < \theta_\rho^2) \geq \mathbf{P}_{\rho/2}(\theta_0^2 < \theta_\rho^2)$ for any $x \in B(\rho/2)$. Since this is positive by Lemma 3.10 (b), the proof of Lemma 3.8 (b) is completed. $\square$

Proof of Lemma 3.9 First, remind that Y_t is defined only for $t < \langle M\rangle_\infty$. Observe that for $t < \theta_0^1 \wedge \langle M\rangle_\infty$,

$$\|Y_t\| - Z_t^1 = \int_0^t (c(Y_s) - c_1(Z_s^1))ds.$$

If there exists $t_0 < \theta_\rho^1 \wedge \theta_0^1 \wedge \langle M\rangle_\infty$ such that $\|Y_{t_0}\| = Z_{t_0}^1$, then $(\|Y\| - Z^1)'(t_0) = c(Y_{t_0}) - c_1(Z_{t_0}^1) = c(Y_{t_0}) - c_1(\|Y_{t_0}\|) > 0$, and therefore, $\|Y_t\| > Z_t^1$ for $t > t_0$ in a neighborhood of t_0. Consequently, $Z_t^1 \leq \|Y_t\|$ for any $t < \theta_\rho^1 \wedge \theta_0^1 \wedge \langle M\rangle_\infty$. Since $Z_t^1 = 0$ for $t \geq \theta_0^1$, this inequality actually holds for $t < \theta_\rho^1 \wedge \langle M\rangle_\infty$. The proof of the other inequality is similar. □

Proof of Lemma 3.10 The proof relies on the same functions p and v as in the proof of Theorem 3.4. They are given by (3.5), where b^ε has to be replaced by c_i, and εa by 1. For the process Z^1, if we fix $\gamma > 0$, then, for any $x > 0$,

$$\begin{aligned} p(y) &= \int_\gamma^y \exp\left[-2\int_\gamma^u c_1(z)dz\right] du = -\int_y^\gamma \exp\left[2k\int_u^\gamma \frac{dz}{z} - k'(\gamma - u)\right] du \\ &= -C\int_y^\gamma u^{-2k} e^{k'u} du, \end{aligned}$$

where we have used the constants $k = \frac{\tilde{b}_0 + da_0/2}{a^0} - \frac{1}{2}$ and $k' = \frac{4K}{\varepsilon a_0}$. Consequently, $p(0+) = -\infty$ if and only if $2k \geq 1$, and $p(+\infty) = +\infty$, which yields (a). A similar computation for Z^2 gives that $p(0+) > -\infty$ if and only if $\frac{\tilde{b}^0 + da^0/2}{a_0} < 1$, which yields Lemma 3.10 (b). □

Proof of Lemma 3.11 Assume that $\mathbb{P}(\{\langle M\rangle_\infty < \infty\} \cap \{\lim_{t\to+\infty} X_t^\varepsilon \in \Gamma\}^c) > 0$. Then, there exists $\alpha > 0$ such that

$$\delta := \mathbb{P}(\langle M\rangle_\infty < \infty,\ \limsup_{t\to+\infty} d(X_t^\varepsilon, \Gamma) \geq \alpha) > 0.$$

Define for any $t > 0$ the stopping time $\tau_{\alpha,t} = \inf\{s \geq t : d(X_s^\varepsilon, \Gamma) \geq \alpha\}$. Then , for any $t > 0$,

$$\mathbb{P}(\langle M\rangle_\infty < \infty,\ \tau_{\alpha,t} < \infty) \geq \delta. \tag{3.10}$$

We will obtain a contradiction from this statement thanks to the following inequality: for any $\varepsilon < 1$, $h \in (0,1)$ and any stopping time S a.s. finite,

$$\mathbb{E}\left[\sup_{0<u<h} \|X_{S+u}^\varepsilon - X_S^\varepsilon\|^2\right] \leq 10C^2 h,$$

where C is a bound for b, $\tilde{b}$ and σ on $\mathbb{R}^d$. This is a straightforward consequence of the inequality

$$\|X_{S+u}^\varepsilon - X_S^\varepsilon\|^2 \leq 2\left(\int_S^{S+u} \|b(X_s^\varepsilon) + \varepsilon\tilde{b}(X_s^\varepsilon)\|ds\right)^2 + 2\sqrt{\varepsilon}\left\|\int_S^{S+u} \sigma(X_s^\varepsilon)dW_s\right\|^2$$

and of Doob's inequality.

Taking $h = \delta\alpha^2/80C^2$ and $S = \tau_{\alpha,t} \wedge T$, we get

$$\mathbb{P}\left(\sup_{0<u<h} \|X^\varepsilon_{\tau_{\alpha,t}\wedge T+u} - X^\varepsilon_{\tau_{\alpha,t}\wedge T}\| > \frac{\alpha}{2}\right) \le \frac{\delta}{2}.$$

Letting $T \to +\infty$,

$$\mathbb{P}\left(\tau_{\alpha,t} < \infty,\ \sup_{0<u<h} \|X^\varepsilon_{\tau_{\alpha,t}+u} - X^\varepsilon_{\tau_{\alpha,t}}\| > \frac{\alpha}{2}\right) \le \frac{\delta}{2}.$$

Together with inequality (3.10), this yields the first line of the following inequality, and the last line makes use of (3.9) and a constant $C > 0$ such that $s^*a(x)s \ge C\|s\|^2$ for any $s \in \mathbb{R}^d$ and $x \in \Gamma_{\alpha/2}$.

$$\begin{aligned}\frac{\delta}{2} &\le \mathbb{P}\left(\langle M\rangle_\infty < \infty,\ \sup_{0<u<h} \|X^\varepsilon_{\tau_{\alpha,t}+u} - X^\varepsilon_{\tau_{\alpha,t}}\| \le \frac{\alpha}{2}\right)\\ &\le \mathbb{P}\left(\langle M\rangle_\infty < \infty,\ \inf_{0<u<h} \|X^\varepsilon_{\tau_{\alpha,t}+u}\| \ge \frac{\alpha}{2}\right)\\ &\le \mathbb{P}\left(\langle M\rangle_\infty < \infty,\ \langle M\rangle_{\tau_{\alpha,t}+h} - \langle M\rangle_{\tau_{\alpha,t}} \ge \varepsilon Ch\right).\end{aligned}$$

Therefore,

$$\mathbb{P}\left(\langle M\rangle_\infty < \infty,\ \langle M\rangle_\infty - \langle M\rangle_t \ge \varepsilon Ch\right) \ge \frac{\delta}{2}$$

holds for any $t > 0$, which is impossible. $\square$

3.5 Example

Let us apply Theorem 3.6 to the model described in Section 2.4. First, it follows from (2.11) that

$$s^*a(x)s = \frac{\|\nabla_1 g(x,x)\|\ \|s\|^2}{\sqrt{2\pi}} + \frac{(s\cdot\nabla_1 g(x,x))^2}{\sqrt{2\pi}\|\nabla_1 g(x,x)\|}$$

for all $s, x \in \mathbb{R}^d$. Since $\|\nabla_1 g(x,x)\| = 2\lambda(x)\|x\|\ \|x-e_1\|\ \|x+e_1\|/\sigma_b^2$, for any $\delta > 0$, there exists $\eta > 0$ such that

$$\begin{cases}\mathcal{U}_y = B(y,\eta), \quad a_y = \frac{4}{\sqrt{2\pi}\sigma_b^2} - \delta, \quad a^y = \frac{8}{\sqrt{2\pi}\sigma_b^2} + \delta & \text{if } y \in \{-e_1, e_1\}\\ \mathcal{U}_y = B(y,\eta), \quad a_y = \frac{2\beta}{\sqrt{2\pi}} - \delta, \quad a^y = \frac{4\beta}{\sqrt{2\pi}} + \delta & \text{if } y = 0\end{cases}$$

are neighborhoods and constants satisfying the assumptions of Theorem 3.6, where $\beta := \exp(-1/2\sigma_b^2)/\sigma_b^2$.

Second, observe that

$$H_{12}g(x,x) = -\frac{\lambda(x)}{\sigma_\alpha^2}\mathrm{Id}.$$

Thus, it can be checked from (2.12) that, for all $y \in \Gamma$,

$$\begin{aligned}\nabla_1 g(x,x) &= (H_{11}g(y,y) + H_{12}g(y,y))(x-y) + O(\|x-y\|^2)\\ &= \begin{cases}-\frac{4}{\sigma_b^2}(x-y) + O(\|x-y\|^2) & \text{if } y \in \{-e_1, e_1\}\\ -2\beta(x - 2x_1e_1) + O(\|x\|^2) & \text{if } y = 0.\end{cases}\end{aligned} \tag{3.11}$$

In the case where $y \in \{-e_1, e_1\}$, since

$$\tilde{b}^{(3)}(x) = -\frac{2\lambda(x)}{\sigma_b^2} \int_{\{h \cdot \nabla_1 g(x,x) > 0\}} h(h^*(xx^* - e_1 e_1^*)h) p(x,h) dh \tag{3.12}$$

and $2(xx^* - e_1 e_1^*) = (x - e_1)(x + e_1)^* + (x + e_1)(x - e_1)^*$, we have $\|\tilde{b}^{(3)}(x)\| = O(\|x - y\|)$. Moreover, by (2.14), we also have $\|\tilde{b}^{(2)}(x)\| = O(\|x - y\|^2)$. Therefore, combining (2.13) and (3.11) and reducing η if necessary, one has in Theorem 3.6

$$\tilde{b}_y \geq -\frac{d+1}{2\sqrt{2\pi}} \left(\frac{1}{\sigma_\alpha^2} - \frac{4}{\sigma_b^2} \right) - \delta \quad \text{and} \quad \tilde{b}^y \leq -\frac{d+1}{2\sqrt{2\pi}} \left(\frac{1}{\sigma_\alpha^2} - \frac{4}{\sigma_b^2} \right) + \delta.$$

Therefore, combining all the previous inequalities, the process X^ε a.s. never hits or converges to $\{-e_1, e_1\}$ as soon as

$$\frac{2d+1}{4} - \frac{(d+1)\sigma_b^2}{16\sigma_\alpha^2} > 1,$$

i.e. if d is large enough and σ_b/σ_α small enough. Similarly, the process X^ε has a positive probability to hit $\{-e_1, e_1\}$ as soon as

$$\frac{3d+1}{2} - \frac{(d+1)\sigma_b^2}{8\sigma_\alpha^2} < 1,$$

i.e. if d is small enough and σ_b/σ_α large enough. Note that this is possible for any dimension d.

In the case where $y = 0$, by (3.12),

$$\|\tilde{b}^{(3)}(x)\| \leq O(\|x\|^2) + \frac{\lambda(x)}{\sigma_b^2} \left(\sum_{k=1}^d \left(\int_{\mathbb{R}^d} |h_k| h_1^2 p(x,h) dh \right)^2 \right)^{1/2} = \beta \sqrt{\frac{d+3}{2\pi}} + O(\|x\|),$$

and by (2.14), $\|\tilde{b}^{(2)}(x)\| = O(\|x|^2)$. Then, we get as above that, in Theorem 3.6,

$$\tilde{b}_0 \geq -\frac{(d+1)\beta}{2\sqrt{2\pi}} \left| \frac{\sigma_b^2}{\sigma_\alpha^2} - 2 \right| - \beta\sqrt{\frac{d+3}{2\pi}} - \delta \quad \text{and} \quad \tilde{b}^0 \leq \frac{(d+1)\beta}{2\sqrt{2\pi}} \left| \frac{\sigma_b^2}{\sigma_\alpha^2} - 2 \right| + \beta\sqrt{\frac{d+3}{2\pi}} + \delta.$$

Hence, X^ε a.s. never hits or converges to 0 as soon as

$$\frac{d}{4} - \frac{d+1}{8} \left| \frac{\sigma_b^2}{\sigma_\alpha^2} - 2 \right| - \frac{\sqrt{d+3}}{4} > 1,$$

i.e. if d is large enough and σ_b/σ_α not too far from $\sqrt{2}$. However, Theorem 3.6 and the previous estimates do not allow us to assess whether there exist parameters d, σ_b and σ_α such that the process X^ε has a positive probability to hit 0.

4 Large deviations for X^ε as $\varepsilon \to 0$

Our large deviation result will be obtained by a transfer technique to carry the LDP from the family $\{\sqrt{\varepsilon}W\}_{\varepsilon>0}$, where W is a standard d-dimensional Brownian motion (Schilder's Theorem, e.g. [13, p. 185]) to the family $\{X^\varepsilon\}_{\varepsilon>0}$, where X^ε is the solution to the SDE (1.1) defined in section 3.1. The method of the proof, adapted from Azencott [2], consists in constructing a function S mapping (in some sense) the paths of $\sqrt{\varepsilon}W$ to the paths of X^ε.

4.1 Statement of the result

We denote by $\mathcal{C}([0,T],\mathbb{R}^d)$ (resp. $\mathcal{C}^{ac}([0.T],\mathbb{R}^d)$) the set of continuous (resp. absolutely continuous) functions from $[0,T]$ to $\mathbb{R}^d$. Fix $T>0$ and $x\in\mathbb{R}^d$, and define

$$\begin{aligned} &\forall\psi\in\mathcal{C}\left([0,T],\mathbb{R}^d\right),\quad t_\psi=\inf\{t\in[0,T]:\psi(t)\in\Gamma\}\wedge T \\ \text{and}\quad &\tilde{\mathcal{C}}^{ac}_x([0,T],\mathbb{R}^d)=\{\psi\in\mathcal{C}^{ac}([0,T],\mathbb{R}^d)\text{ constant on }[t_\psi,T]\text{ such that }\psi(0)=x\}.\end{aligned}$$

Then, we define for $\psi\in\mathcal{C}([0,T],\mathbb{R}^d)$

$$I_{T,x}(\psi)=\begin{cases}\dfrac{1}{2}\displaystyle\int_0^{t_\psi}[\dot\psi(t)-b(\psi(t))]^*a^{-1}(\psi(t))[\dot\psi(t)-b(\psi(t))]dt \\ \qquad\qquad\qquad\qquad\qquad\qquad\text{if}\quad\psi\in\tilde{\mathcal{C}}^{ac}_x([0,T],\mathbb{R}^d) \\ +\infty\qquad\qquad\qquad\qquad\qquad\text{otherwise.}\end{cases}\tag{4.1}$$

By Proposition 2.4 (ii), the inverse matrix $a^{-1}(x)$ of $a(x)$ is well-defined, symmetric and non-negative for all $x\notin\Gamma$, so $I_{T,x}(\psi)$ is well-defined and belongs to $\mathbb{R}_+\cup\{+\infty\}$. When $t_\psi=T$, $I_{T,x}(\psi)$ takes the classical form of rate functions for diffusion processes.

This original form of rate function will appear naturally in the proof. However, as shown in Proposition 4.5 below, this function is *not* lower semicontinuous. Therefore, it is natural to introduce for all $\psi\in\mathcal{C}([0,T],\mathbb{R}^d)$

$$\tilde I_{T,x}(\psi)=\liminf_{\tilde\psi\to\psi}I_{T,x}(\tilde\psi),\tag{4.2}$$

which is the biggest lower semicontinuous function on $\mathcal{C}([0,T],\mathbb{R}^d)$ smaller than $I_{T,x}$.

Theorem 4.1 *Assume (H). Assume also that the points of Γ are isolated in $\mathbb{R}^d$. Fix $T>0$. Then, for any $x\in\mathbb{R}^d$ and any open subset O of $\mathcal{C}([0,T],\mathbb{R}^d)$,*

$$\liminf_{\varepsilon\to0,y\to x}\varepsilon\ln\mathbb{P}(X^{\varepsilon,y}\in O)\geq-\inf_{\psi\in O}\tilde I_{T,x}(\psi),\tag{4.3}$$

and for any $x\notin\Gamma$ and any closed subset C of $\mathcal{C}([0,T],\mathbb{R}^d)$,

$$\limsup_{\varepsilon\to0,y\to x}\varepsilon\ln\mathbb{P}(X^{\varepsilon,y}\in C)\leq-\inf_{\psi\in C}\tilde I_{T,x}(\psi).\tag{4.4}$$

The general form of the lower and upper bounds (4.3) and (4.4) (where the limit is taken over $y\to x$) will be useful in Section 5. This general form requires the restriction that $x\notin\Gamma$ for the upper bound for technical reasons. However, this result implies that the following standard form of LDP holds without any restriction.

Corollary 4.2 *Assume the conditions of Theorem 4.1. Then, for any $x\in\mathbb{R}^d$, for any open $O\subset\mathcal{C}([0,T],\mathbb{R}^d)$, and for any closed $C\subset\mathcal{C}([0,T],\mathbb{R}^d)$,*

$$\liminf_{\varepsilon\to0}\varepsilon\ln\mathbb{P}(X^{\varepsilon,x}\in O)\geq-\inf_{\psi\in O}\tilde I_{T,x}(\psi),\tag{4.5}$$

$$\limsup_{\varepsilon\to0}\varepsilon\ln\mathbb{P}(X^{\varepsilon,x}\in C)\leq-\inf_{\psi\in C}\tilde I_{T,x}(\psi).\tag{4.6}$$

Proof The lower bound (4.5) is a trivial consequence of (4.3) and the upper bound (4.6) for $x \notin \Gamma$ also trivially follows from (4.4). If $x \in \Gamma$, let us denote by x the constant function of $\mathcal{C}([0,T],\mathbb{R}^d)$ equal to x. In this case, $X^{\varepsilon,x}_t = x$ for all $t \geq 0$. Therefore, $\mathbb{P}(X^{\varepsilon,x} \in C)$ equals 1 if the function x belongs to C, and equals 0 otherwise. Since $\tilde{I}_{T,x}(x) \leq I_{T,x}(x) = 0$, the upper bound (4.6) is clear when $x \in \Gamma$. □

Remark 4.3 *As usual for large deviation principles, Corollary 4.2 implies the convergence in probability of $X^{\varepsilon,x}$ to the solution with initial state x of the deterministic ODE*

$$\dot{\phi} = b(\phi)$$

as $\varepsilon \to 0$. This ODE is known as the canonical equation of adaptive dynamics *[15, 7, 10].*

In Section 5, we will use the following classical consequence of Theorem 4.6, which can be proved exactly as Corollary 5.6.15 of [13]:

Corollary 4.4 *Assume the conditions of Theorem 4.1. Then, for any compact set $K \subset \mathbb{R}^d$ and for any open $O \subset \mathcal{C}([0,T],\mathbb{R}^d)$,*

$$\liminf_{\varepsilon\to 0} \varepsilon \ln \inf_{y\in K} \mathbb{P}(X^{\varepsilon,y} \in O) \geq -\sup_{y\in K}\inf_{\psi\in O} \tilde{I}_{T,y}(\psi),$$

and if $K \cap \Gamma = \emptyset$, for any closed $C \subset \mathcal{C}([0,T],\mathbb{R}^d)$,

$$\limsup_{\varepsilon\to 0} \varepsilon \ln \sup_{y\in K} \mathbb{P}(X^{\varepsilon,y} \in C) \leq -\inf_{y\in K,\ \psi\in C} \tilde{I}_{T,y}(\psi).$$

We end this subsection with some remarks on the rate functions we obtain and their links with the classical form of rate functions for diffusion processes with small noise.

Proposition 4.5 *Assume the conditions of Theorem 4.1. Assume also that there exists an isolated point y of Γ such that g is $\mathcal{C}^2$ at (y,y), and that $H_{1,1}g(y,y) + H_{1,2}g(y,y)$ is invertible. Then, for any $x \notin \Gamma$ and $T > 0$, $I_{T,x}$ is* not *lower semicontinuous.*

We postpone the proof of this result at the end of this subsection.

General large deviation estimates are known for diffusions in $\mathbb{R}^d$ with small noise using different techniques. For example, Dupuis, Ellis and Weiss [21] have obtained upper bounds under very general assumptions. We could have applied their result in our case (with some modifications since they consider a drift that does not depend on ε, see Remark 1.2 in [21]) to obtain a large deviations upper bound with lower semicontinuous rate function

$$\hat{I}_{T,x}(\psi) = \begin{cases} \dfrac{1}{2}\displaystyle\int_0^T \mathbf{1}_{\psi(t)\notin\Gamma}[\dot{\psi}(t) - b(\psi(t))]^* a^{-1}(\psi(t))[\dot{\psi}(t) - b(\psi(t))]dt \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{if } \psi \in \mathcal{C}^{ac}([0,T],\mathbb{R}^d) \text{ and } \psi(0) = x \\ +\infty \qquad\qquad\qquad\qquad\qquad \text{otherwise,} \end{cases}$$

for all $\psi \in \mathcal{C}([0,T],\mathbb{R}^d)$.

Since obviously $\hat{I}_{T,x} \leq I_{T,x}$ and $\hat{I}_{T,x}$ is lower semicontinuous, we have $\hat{I}_{T,x} \leq \tilde{I}_{T,x}$. Since $\tilde{I}_{T,x} \leq I_{T,x}$, this immediately implies that $\hat{I}_{T,x}$ and $\tilde{I}_{T,x}$ coincide on $\mathcal{C}([0,T],\mathbb{R}^d \setminus \Gamma)$. Unfortunately, because of the degeneracy of a on Γ, we are not able to obtain in general an explicit expression for $\tilde{I}_{T,x}(\psi)$ when $\psi_t \in \Gamma$ for some $t \in [0,T]$. However, it is possible to find simple examples where these two rate function are not equal: Assume that $d = 1$ and 0 is an isolated point of Γ, and consider a function ψ such that $\psi(0) < 0$, $\psi(T) > 0$ and $\hat{I}_{T,x}(\psi) < +\infty$ (such a function can be easily obtained by adapting the construction of the function ψ in the proof of Proposition 4.5 below). Obviously, $\tilde{I}_{T,x}(\psi) = +\infty$, giving the required counter-example.

Therefore, our upper bound is more precise than the one obtained by classical general methods. This also explains why we have to use a method based on a precise study of the paths of $X^{\varepsilon,x}$ to obtain our result.

Proof of Proposition 4.5 Take y as in Proposition 4.5. By translation, we can suppose that $y = 0$. Then, Proposition 3.7 implies that there exists a neighborhood $\mathcal{N}_0$ of 0 and a constant $a_0 > 0$ such that for all $s \in \mathbb{R}^d$ and $x \in \mathcal{N}_0$, $s^* a(x) s \geq a_0 \|x\| \|s\|^2$, *i.e.* each eigenvalue of $a(x)$ is greater than $a_0\|x\|$. Therefore, for all $s \in \mathbb{R}^d$ and $x \in \mathcal{N}_0$,

$$s^* a^{-1}(x) s \leq \frac{\|s\|^2}{a_0\|x\|}. \tag{4.7}$$

Take $x_0 \in \mathbb{R}^d \setminus \Gamma$ such that the segment $(0, x_0]$ is included in $(\mathbb{R}^d \setminus \Gamma) \cup \mathcal{N}_0$, and define for $0 \leq t \leq T$

$$\psi(t) = \left(1 - \frac{2t}{T}\right)^2 x_0, \tag{4.8}$$

and for all $n \geq 1$

$$\psi_n(t) = \begin{cases} \psi(t) & \text{if} \quad t \in \left[0, \dfrac{T}{2} - \dfrac{1}{n}\right] \cup \left[\dfrac{T}{2} + \dfrac{1}{n}, T\right] \\ \psi\left(\dfrac{T}{2} - \dfrac{1}{n}\right) & \text{otherwise.} \end{cases}$$

Since $\psi(T/2 - 1/n) = \psi(T/2 + 1/n)$, ψ_n is continuous and piecewise differentiable. Note that $\psi(t)$ and $\psi_n(t)$ belong to $[0, x_0]$ for all $t \in [0,T]$, that $\psi(t) \notin \Gamma$ except if $t = T/2$, and that $\psi_n(t) \notin \Gamma$ for any $t \in [0,T]$. Therefore, $I_{T,x_0}(\psi) = \infty$, and $I_{T,x_0}(\psi_n) < \infty$.

It follows from (4.7) and from the fact that b is K-Lipschitz that

$$\begin{aligned} \hat{I}_{T,x_0}(\psi) &\leq \frac{1}{2a_0} \int_0^T \frac{\|(1-2t/T)2x_0/T + b(\psi(t))\|^2}{\|\psi(t)\|} dt \\ &\leq \frac{1}{2a_0} \int_0^T \frac{2(1-2t/T)^2 4\|x_0\|^2/T^2 + 2K^2\|\psi(t)\|^2}{\|\psi(t)\|} dt \\ &\leq \frac{1}{2a_0} \int_0^T \left(\frac{8}{T^2}\|x_0\| + 2K^2\|\psi(t)\|\right) dt < \infty. \end{aligned} \tag{4.9}$$

Now, for all $n \geq 1$,

$$\begin{aligned} I_{T,x_0}(\psi_n) &\leq \hat{I}_{T,x_0}(\psi) + \frac{1}{2a_0}\int_{T/2-1/n}^{T/2+1/n} \frac{\|b(\psi_n(t))\|^2}{\|\psi_n(t)\|} dt \\ &\leq \hat{I}_{T,x_0}(\psi) + \frac{1}{2a_0}\int_{T/2-1/n}^{T/2+1/n} K^2\|\psi_n(t)\| dt, \end{aligned}$$

which is uniformly bounded in n. Hence $\limsup I_{T,x_0}(\psi_n) < +\infty = I_{T,x_0}(\psi)$.

Let us extend this result to an arbitrary $x \notin \Gamma$. Since the points of Γ are isolated in $\mathbb{R}^d$, there exists $\alpha > 0$ and $\phi \in \mathcal{C}^1([0,T], \Gamma_\alpha)$ such that $\phi(0) = x$ and $\phi(T) = x_0$. Since a is uniformly non-degenerate on Γ_α, $I_{T,x}(\phi) < \infty$. Therefore, it suffices to concatenate ϕ and ψ to obtain a function $\tilde{\psi}$ defined on $[0, 2T]$ such that $\limsup \tilde{I}_{2T,x}(\tilde{\psi}) < I_{2T,x}(\tilde{\psi})$. Since this can be done for all $T > 0$, this ends the proof of Proposition 4.5. $\square$

4.2 Proof of Theorem 4.1

We first give some notation used throughout the proof.

- $\mathcal{C}_x(I,E)$(*resp.* $\mathcal{C}_x^{ac}(I,E)$, $\mathcal{C}_x^1(I,E)$) is the set of continuous functions from $I \subset \mathbb{R}_+$ to $E \subset \mathbb{R}^d$ (resp. absolutely continuous, resp. $\mathcal{C}^1$) with value x at 0, endowed with the L^∞ norm.
- For $\varphi \in \mathcal{C}([0,T], \mathbb{R}^d)$ and $0 \leq a < b \leq T$, define

$$\|\varphi\|_{a,b} = \sup_{a \leq t \leq b} \|\varphi(t)\|, \tag{4.10}$$

  and

$$B_b(\varphi, \delta) = \{\tilde{\varphi} \in \mathcal{C}([0,T], \mathbb{R}^d) : \|\tilde{\varphi} - \varphi\|_{0,b} \leq \delta\}. \tag{4.11}$$

  When $a = 0$ and $b = T$, $\|\cdot\|_{0,T}$ is the usual L^∞ norm in $\mathcal{C}([0,T], \mathbb{R}^d)$, and $B_T(\varphi, \delta)$ is the usual closed ball centered at φ with radius δ for this norm, also simply denoted $B(\varphi, \delta)$.

We are actually going to prove the following result.

Theorem 4.6 *Assume the conditions of Theorem 4.1. Then, for any $x \in \mathbb{R}^d$ and any open subset O of $\mathcal{C}([0,T], \mathbb{R}^d)$,*

$$\liminf_{\varepsilon \to 0, y \to x} \varepsilon \ln \mathbb{P}(X^{\varepsilon,y} \in O) \geq -\inf_{\psi \in O} I_{T,x}(\psi), \tag{4.12}$$

and for any $x \notin \Gamma$ and any closed subset C of $\mathcal{C}([0,T], \mathbb{R}^d)$ such that $\mathcal{C}_x^1([0,T], \mathbb{R}^d \setminus \Gamma)$ is dense in $C \cap \mathcal{C}_x([0,T], \mathbb{R}^d)$,

$$\limsup_{\varepsilon \to 0, y \to x} \varepsilon \ln \mathbb{P}(X^{\varepsilon,y} \in C) \leq -\inf_{\psi \in C} I_{T,x}(\psi). \tag{4.13}$$

This is an incomplete LDP involving the non-lower semicontinuous rate function $I_{T,x}$. From this can be deduced the LDP involving the rate function $\tilde{I}_{T,x}$ (Theorem 4.1) as follows.

First, by definition of $\tilde{I}_{T,x}$, for any open $O \subset \mathcal{C}([0,T],\mathbb{R}^d)$,

$$\inf_{\psi\in O} I_{T,x}(\psi) = \inf_{\psi\in O} \tilde{I}_{T,x}(\psi).$$

Therefore, (4.3) is immediate.

Moreover, $\tilde{I}_{T,x} \leq I_{T,x}$, so (4.4) obviously holds for the same closed sets as in Theorem 4.6. Now, let K be any compact subset of $\mathcal{C}([0,T],\mathbb{R}^d)$. Since $\tilde{I}_{T,x}$ is lower semicontinuous, for any $\eta > 0$, there exists $\alpha > 0$ such that

$$\inf_{\psi\in K} \tilde{I}_{T,x}(\psi) \leq \inf_{\psi\in K_\alpha} \tilde{I}_{T,x}(\psi) + \eta,$$

where

$$K_\alpha = \bigcup_{\psi\in K} B(\psi,\alpha).$$

Indeed, if this would fail, there would exist $\eta > 0$ and two sequences $(\psi_n)_{n\geq 1}$ and $(\tilde{\psi}_n)_{n\geq 1}$ such that $\tilde{\psi}_n \in K$, $\|\psi_n - \tilde{\psi}_n\|_{0,T} \leq 1/n$ and $\tilde{I}_{T,x}(\psi_n) \leq \tilde{I}_{T,x}(\tilde{\psi}_n) - \eta$ for all $n \geq 1$. Since K is compact, we could then extract a subsequence $(\tilde{\psi}_{i_n})$ of $(\tilde{\psi}_n)$ converging to some $\tilde{\psi} \in K$. Since $\tilde{I}_{T,x}$ is lower semicontinuous, this would imply that

$$\tilde{I}_{T,x}(\tilde{\psi}) \leq \liminf_{n\to+\infty} \tilde{I}_{T,x}(\psi_{i_n}) \leq \inf_{\psi\in K} \tilde{I}_{T,x}(\psi) - \eta,$$

which is a contradiction.

Now, let $\psi_1, \dots, \psi_n$ be such that

$$\tilde{K}_\alpha = \bigcup_{i=1}^{n} B(\psi_i, \alpha) \supset K.$$

Since $\tilde{K}_\alpha \subset K_\alpha$,

$$\inf_{\psi\in K} \tilde{I}_{T,x}(\psi) \leq \inf_{\psi\in \tilde{K}_\alpha} \tilde{I}_{T,x}(\psi) + \eta.$$

Moreover, the points of Γ are isolated, and thus any point of the interior of $\tilde{K}_\alpha$ is obviously limit of elements of $\tilde{K}_\alpha \cap \mathcal{C}^1([0,T],\mathbb{R}^d \setminus \Gamma)$. Since $\tilde{K}_\alpha$ is the closure of its interior, any point of $\partial\tilde{K}_\alpha$ is also limit of elements of $\tilde{K}_\alpha \cap \mathcal{C}^1([0,T],\mathbb{R}^d \setminus \Gamma)$ by a diagonal procedure. Moreover, $\tilde{K}_\alpha$ is closed. Therefore, one can apply (4.13) to this set:

$$\begin{aligned}\limsup_{\varepsilon\to 0,\, y\to x} \varepsilon \ln \mathbb{P}(X^{\varepsilon,y} \in K) &\leq \limsup_{\varepsilon\to 0,\, y\to x} \varepsilon \ln \mathbb{P}(X^{\varepsilon,y} \in \tilde{K}_\alpha) \leq -\inf_{\psi\in\tilde{K}_\alpha} I_{T,x}(\psi) \\ &\leq -\inf_{\psi\in\tilde{K}_\alpha} \tilde{I}_{T,x}(\psi) \leq -\inf_{\psi\in K} \tilde{I}_{T,x}(\psi) + \eta.\end{aligned}$$

Since this holds for all $\eta > 0$, (4.4) is proved for compact sets.

The extension to any closed sets is classically deduced from the following uniform exponential tightness estimate.

Lemma 4.7 *For any $k > 0$ and $y \in \mathbb{R}^d$, define the compact set*

$$K_k^y = \left\{ \psi \in \mathcal{C}_y([0,T], \mathbb{R}^d) : \forall l \geq k,\ \omega\left(\psi, \frac{1}{l^3}\right) \leq \frac{1}{l} \right\}, \tag{4.14}$$

where $\omega(\psi, \delta) = \sup_{|t-s|\leq\delta} \|\psi(t) - \psi(s)\|$. Then, there exists k_0 and ε_0, such that for all $y \in \mathbb{R}^d$, $k \geq k_0$ and $\varepsilon \leq \varepsilon_0$,

$$\varepsilon \ln \mathbb{P}(X^{\varepsilon,y} \notin K_k^y) \leq -\frac{k}{64 d \Sigma^2}, \tag{4.15}$$

where $\Sigma := \sup_{x\in\mathbb{R}^d} \|\sigma(x)\|$.

Then, taking any closed $C \in \mathcal{C}([0,T], \mathbb{R}^d)$ and choosing k large enough,

$$\begin{aligned} \limsup_{\varepsilon\to 0, y\to x} \varepsilon \ln \mathbb{P}(X^{\varepsilon,y} \in C) &\leq \sup \Big\{ \limsup_{\varepsilon\to 0, y\to x} \varepsilon \ln \mathbb{P}(X^{\varepsilon,y} \in C \cap K_k^y), \\ &\qquad \limsup_{\varepsilon\to 0, y\to x} \varepsilon \ln \mathbb{P}(X^{\varepsilon,y} \notin K_k^y) \Big\} \\ &\leq -\inf_{\psi\in C} I_{T,x}(\psi), \end{aligned} \tag{4.16}$$

ending the proof of Theorem 4.1. □

The proof of Lemma 4.7 makes use of the following classical exponential inequality for stochastic integrals, of which the proof is omitted. This result will be also used in the proof of Theorem 4.6 below. Let $\mathcal{M}_{d,d}$ denote the set of real $d \times d$ matrices.

Lemma 4.8 *Let Y_t be a $\mathcal{F}_t$-martingale with values in $\mathbb{R}^d$ on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$, and suppose that its quadratic covariation process $\langle Y \rangle_t$ satisfies $\sup_{t\leq T} \|\langle Y \rangle_t\| \leq A$. Let τ be a $\mathcal{F}_t$ stopping time, and let $Z : \mathbb{R}_+ \times \Omega \to \mathcal{M}_{d,d}$ be a progressively measurable process such that $\sup_{t\leq\tau} \|Z_t\| \leq B$. Then for any $R > 0$,*

$$\mathbb{P}\left(\sup_{t\leq T} \left\| \int_0^{t\wedge\tau} Z_s dY_s \right\| \geq R \right) \leq 2d \exp\left(-\frac{R^2}{2dTAB^2} \right).$$

Proof of Lemma 4.7 It follows from (1.1) that, for any $y \in \mathbb{R}^d$, $s > 0$ and $t \in [0,T]$,

$$\|X_{t+s}^{\varepsilon,y} - X_t^{\varepsilon,y}\| \leq Cs + \sqrt{\varepsilon} \left\| \int_t^{t+s} \sigma(X_u^{\varepsilon,y}) dW_u \right\|.$$

Fix $h > 0$ and $R \geq Ch$. Applying Lemma 4.8, we have

$$\mathbb{P}\left(\sup_{0\leq s\leq h} \|X_{t+s}^{\varepsilon,y} - X_t^{\varepsilon,y}\| \geq R \right) \leq 2d \exp\left(-\frac{(R - Ch)^2}{2dh\varepsilon\Sigma^2} \right).$$

Writing this for $t = ih$ for $0 \leq i < T/h$, we deduce that

$$\mathbb{P}\left(\omega(X^\varepsilon, h) \geq 2R \right) \leq 2d \left(\frac{T}{h} + 1 \right) \exp\left(-\frac{(R - Ch)^2}{2d\varepsilon\Sigma^2 h} \right). \tag{4.17}$$

For any $l \geq 1$, set $R_l = 1/2l$ and $h_l = 1/l^3$. Then, for sufficiently large l, $R_l \geq Ch_l$ and

$$\frac{(R_l - Ch_l)^2}{2d\varepsilon\Sigma^2 h_l} = \frac{(\sqrt{l} - 2C/l^{3/2})^2}{8d\varepsilon\Sigma^2} \geq \frac{(\sqrt{l}/2)^2}{8d\varepsilon\Sigma^2} = \frac{l}{32d\varepsilon\Sigma^2}. \tag{4.18}$$

Observing that

$$K_k^y = \{\psi \in \mathcal{C}_y([0,T],\mathcal{X}) : \forall l \geq k, \omega(\psi, h_l) \leq 2R_l\},$$

inequality (4.15) easily follows from (4.17) and (4.18). $\square$

4.3 Proof of Theorem 4.6

The proof of Theorem 4.6 makes use of the function $I_{T,x}$ and of the (good) rate function of Schilder's theorem (LDP for Brownian motion)

$$J_T(\varphi) = \begin{cases} \dfrac{1}{2}\displaystyle\int_0^T \|\dot{\varphi}(t)\|^2 dt & \text{if } \quad \varphi \in \mathcal{C}_0^{ac}([0,T],\mathbb{R}^d) \\ +\infty & \text{otherwise.} \end{cases}$$

First, we need to construct the function S "mapping" Brownian paths to the paths of X^ε. For any $\varphi \in \mathcal{C}_0^{ac}([0,T],\mathbb{R}^d)$, let $S(\varphi)$ be the solution on $[0,T]$ to

$$S(\varphi)(t) = x + \int_0^t b(S(\varphi)(s))ds + \int_0^t \sigma(S(\varphi)(s))\dot{\varphi}(s)ds, \tag{4.19}$$

obtained as follows: by Proposition 2.4 (i) and (iii), b and σ are bounded and locally Lipschitz on $\mathbb{R}^d \setminus \Gamma$. Therefore, Cauchy-Lipschitz's theorem implies local existence and uniqueness in $\mathbb{R}^d \setminus \Gamma$ of a solution to $\dot{y} = b(y) + \sigma(y)\dot{\varphi}$. This defines properly $S(\varphi)$ until the time $t_{S(\varphi)}$ where it reaches Γ. In the case where $t_{S(\varphi)} < T$, set $S(\varphi)(t) = S(\varphi)(t_{S(\varphi)})$ for $t_{S(\varphi)} \leq t \leq T$. The function $S(\varphi)$ obtained this way is a solution to (4.19) on $[0,T]$ and belongs to $\tilde{\mathcal{C}}_x^{ac}([0,T],\mathbb{R}^d)$.

The proof of Theorem 4.6 is based on the following two lemmas. The first one gives a precise sense to the fact that the function S maps the paths of $\sqrt{\varepsilon}W$ to the paths of $X^{\varepsilon,x}$. The second one gives the relation between S, $I_{T,x}$ and J_T. Their proof is postponed after the proof of the theorem.

Lemma 4.9

(i) *Fix $\varphi \in \mathcal{C}_0^{ac}([0,T],\mathbb{R}^d)$ such that $\psi := S(\varphi)$ takes no value in Γ and such that $J_T(\varphi) < +\infty$. Then, for all $\eta > 0$ and $R > 0$, there exists $\delta > 0$ such that*

$$\limsup_{\varepsilon\to 0, y\to x} \varepsilon \ln \mathbb{P}\left(\|X^{\varepsilon,y} - S(\varphi)\|_{0,T} \geq \eta, \|\sqrt{\varepsilon}W - \varphi\|_{0,T} \leq \delta\right) \leq -R. \tag{4.20}$$

(ii) *Fix $\tilde{\varphi} \in \mathcal{C}_0^{ac}([0,T],\mathbb{R}^d)$ such that $\psi(t) := S(\tilde{\varphi})(t) \in \Gamma$ for some $t \in [0,T]$. Define $\varphi(t) = \tilde{\varphi}(t)$ for $t < t_\psi$ and $\varphi(t) = \tilde{\varphi}(t_\psi)$ for $t_\psi \leq t \leq T$. Then $S(\varphi) = S(\tilde{\varphi}) = \psi$. Suppose that $J_T(\varphi) < +\infty$. Then, for all $\eta > 0$ and $R > 0$, there exists $\delta > 0$ such that (4.20) holds.*

(iii) *With the same φ as in (i), for all $\delta > 0$ and $R > 0$, there exists $\eta > 0$ such that*

$$\limsup_{\varepsilon\to 0, y\to x} \varepsilon \ln \mathbb{P}\left(\|X^{\varepsilon,y} - S(\varphi)\|_{0,T} \leq \eta, \|\sqrt{\varepsilon}W - \varphi\|_{0,T} \geq \delta\right) \leq -R. \tag{4.21}$$

Lemma 4.10

(i) *For all $\psi \in \mathcal{C}_x([0,T],\mathbb{R}^d)$,*

$$I_{T,x}(\psi) = \inf\{J_T(\varphi), S(\varphi) = \psi\}$$

and when $I_{T,x}(\psi) < +\infty$, there is a unique $\varphi \in \mathcal{C}_0^{ac}([0,T],\mathbb{R}^d)$ that realizes this infimum, and this function is constant after t_ψ.

(ii) *$\mathcal{C}^1([0,T],\mathbb{R}^d \setminus \Gamma)$ is dense in $S(\{J_T < \infty\})$.*

In [2, 18], b^ε and σ are assumed Lipschitz, and thus Point (i) of Lemma 4.9 can be proved for all $\varphi \in \mathcal{C}_0^{ac}([0,T],\mathbb{R}^d)$, which is enough to conclude. In our case, because of the bad regularity of the coefficients of the SDE, we cannot prove (i) for all $\varphi \in \mathcal{C}_0^{ac}([0,T],\mathbb{R}^d)$. As a consequence, we are only able to obtain the large deviations lower bound from Lemma 4.9 (i) and (ii). In order to prove the large deviations upper bound, we use an original method based on Lemma 4.9 (iii).

Lemma 4.10 is an extension to our case of very similar lemmas in [2, 18].

Proof of Theorem 4.6: lower bound It is well-known that the lower bound (4.12) for any open set O is equivalent to the fact that, for all $\psi \in \mathcal{C}_x([0,T],\mathbb{R}^d)$ and $\eta > 0$,

$$\liminf_{\varepsilon\to 0, y\to x} \varepsilon \ln \mathbb{P}(\|X^{\varepsilon,y} - \psi\|_{0,T} \leq \eta) \geq -I_{T,x}(\psi). \tag{4.22}$$

Fix ψ and η as above, and assume that $I_{T,x}(\psi) < +\infty$ (otherwise, there is nothing to prove). By Lemma 4.10 (i), there is a unique $\varphi \in \mathcal{C}_0^{ac}([0,T],\mathbb{R}^d)$ such that $S(\varphi) = \psi$ and $u := J_T(\varphi) = I_{T,x}(\psi)$. Choose $R > u$. If the image of ψ has empty intersection with Γ, apply Lemma 4.9 (i). Otherwise, apply Lemma 4.9 (ii). In both cases, there exists $\delta > 0$ such that

$$\limsup_{\varepsilon\to 0, y\to x} \varepsilon \ln \mathbb{P}\left(\|X^{\varepsilon,y} - \psi\|_{0,T} \geq \eta, \|\sqrt{\varepsilon}W - \varphi\|_{0,T} \leq \delta\right) \leq -R.$$

Since

$$\begin{aligned}\mathbb{P}(\|\sqrt{\varepsilon}W - \varphi\|_{0,T} \leq \delta) \leq \mathbb{P}(&\|X^{\varepsilon,y} - \psi\|_{0,T} < \eta) \\ &+ \mathbb{P}(\|X^{\varepsilon,y} - \psi\|_{0,T} \geq \eta, \|\sqrt{\varepsilon}W - \varphi\|_{0,T} \leq \delta),\end{aligned}$$

we deduce from Schilder's theorem that

$$\begin{aligned}-u = -J_T(\varphi) &\leq -\inf\{J_T(\tilde\varphi), \tilde\varphi \in B_T(\varphi,\delta)\} \\ &\leq \liminf_{\varepsilon\to 0, y\to x} \varepsilon \ln \mathbb{P}(\|\sqrt{\varepsilon}W - \varphi\|_{0,T} < \delta) \\ &\leq \sup\Bigg\{\liminf_{\varepsilon\to 0, y\to x} \varepsilon \ln \mathbb{P}(\|X^{\varepsilon,y} - \psi\|_{0,T} < \eta), \\ &\qquad \liminf_{\varepsilon\to 0, y\to x} \varepsilon \ln \mathbb{P}(\|X^{\varepsilon,y} - \psi\|_{0,T} \geq \eta, \|\sqrt{\varepsilon}W - \varphi\|_{0,T} \leq \delta)\Bigg\} \\ &\leq \sup\left\{\liminf_{\varepsilon\to 0, y\to x} \varepsilon \ln \mathbb{P}(\|X^{\varepsilon,y} - \psi\|_{0,T} < \eta), -R\right\},\end{aligned}$$

and since $R > u$, (4.22) is established. □

Proof of Theorem 4.6: upper bound We first prove (4.13) for particular compact sets: let K be a non-empty compact set of $\mathcal{C}([0,T],\mathbb{R}^d)$ such that $S(\{J_T<+\infty\})$ is dense in K_x, where $K_x := K\cap\mathcal{C}_x([0,T],\mathbb{R}^d)$. By Lemma 4.10 (i), $S(\{J_T<+\infty\}) = \{I_{T,x}<+\infty\}$, and so $u:=\inf\{I_{T,x}(\psi),\psi\in K\}<+\infty$.

Fix $\rho>0$. For any $\psi\in K\cap S(\{J_T<+\infty\})$, by Lemma 4.10 (i), there exists a unique $\varphi\in\mathcal{C}_0^{ac}([0,T],\mathbb{R}^d)$ constant after t_ψ such that $S(\varphi)=\psi$ and $I_{T,x}(\psi)=J_T(\varphi)<\infty$. We intend to use Lemma 4.9 (iii), which holds only if ψ takes no value in Γ. So we have to introduce $\alpha_\psi>0$ such that

$$\frac{1}{2}\int_{t_\psi-\alpha_\psi}^{t_\psi}\|\dot\varphi_s\|^2ds<\frac{\rho}{2},$$

so that $J_T(\varphi)\le J_{t_\psi-\alpha_\psi}(\varphi)+\rho/2$. Since $J_{t_\psi-\alpha_\psi}$ is lower semicontinuous, there exists $\delta_\psi>0$ such that

$$\forall\tilde\varphi\in B_{t_\psi-\alpha_\psi}(\varphi,\delta_\psi),\quad J_{t_\psi-\alpha_\psi}(\tilde\varphi)\ge J_{t_\psi-\alpha_\psi}(\varphi)-\frac{\rho}{2}\ge J_T(\varphi)-\rho, \tag{4.23}$$

where $B_t(\varphi,\delta)$ has been defined in (4.11).

Applying Lemma 4.9 (ii) to ψ with $T=t_\psi-\alpha_\psi$, $\delta=\delta_\psi$ and $R>u$, there exists $\eta_\psi>0$ such that

$$\limsup_{\varepsilon\to0,y\to x}\varepsilon\ln\mathbb{P}\left(\|X^{\varepsilon,y}-\psi\|_{0,t_\psi-\alpha_\psi}\le\eta_\psi,\|\sqrt{\varepsilon}W-\varphi\|_{0,t_\psi-\alpha_\psi}\ge\delta_\psi\right)\le-R. \tag{4.24}$$

Since we have assumed that $K_x\cap S(\{J_T<+\infty\})$ is dense in K_x,

$$K_x\subset\bigcup_{\psi\in K_x\cap S(\{J_T<+\infty\})}B_T(\psi,\eta_\psi).$$

Since K_x is compact, there exists a finite number of functions $\psi_1,\dots,\psi_n$ in $K_x\cap S(\{J_T<+\infty\})$ such that

$$K_x\subset\bigcup_{i=1}^n B_T(\psi_i,\eta_i),$$

where we wrote η_i instead of η_{ψ_i}. Since K is compact, there exists a neighborhood $\mathcal{N}_x$ of x such that

$$K_{\mathcal{N}_x}\subset\bigcup_{i=1}^n B_T(\psi_i,\eta_i),$$

where $K_{\mathcal{N}_x}=\{\psi\in K:\psi(0)\in\mathcal{N}_x\}$.

Now, define

$$U=\bigcup_{i=1}^n B_{t_i-\alpha_i}(\varphi_i,\delta_i),$$

where $t_i=t_{\psi_i}$, $\alpha_i=\alpha_{\psi_i}$ and $\delta_i=\delta_{\psi_i}$, and where φ_i is the function satisfying $S(\varphi_i)=\psi_i$

and $I_{T,x}(\psi_i) = J_T(\varphi_i)$. Then, for any $y \in \mathcal{N}_x$,

$$\begin{aligned}
\mathbb{P}(X^{\varepsilon,y} \in K) &\leq \mathbb{P}(\sqrt{\varepsilon}W \in U) + \mathbb{P}(\sqrt{\varepsilon}W \notin U,\ X^{\varepsilon,y} \in K_{\mathcal{N}_x}) \\
&\leq \sum_{i=1}^{n} \mathbb{P}(\sqrt{\varepsilon}W \in B_{t_i-\alpha_i}(\varphi_i, \delta_i)) \\
&\quad + \sum_{i=1}^{n} \mathbb{P}(\|X^{\varepsilon,y} - \psi_i\|_{0,T} \leq \eta_i, \sqrt{\varepsilon}W \notin U) \\
&\leq \sum_{i=1}^{n} \mathbb{P}(\|\sqrt{\varepsilon}W - \varphi_i\|_{0,t_i-\alpha_i} < \delta_i) \\
&\quad + \sum_{i=1}^{n} \mathbb{P}(\|X^{\varepsilon,y} - \psi_i\|_{0,t_i-\alpha_i} \leq \eta_i, \|\sqrt{\varepsilon}W - \varphi_i\|_{0,t_i-\alpha_i} \geq \delta_i).
\end{aligned}$$

Since by Schilder's Theorem and (4.23)

$$\limsup_{\varepsilon\to 0, y\to x} \varepsilon \ln \mathbb{P}(\|\sqrt{\varepsilon}W - \varphi_i\|_{0,t_i-\alpha_i} \leq \delta_i) \leq -\inf_{\varphi \in B_{t_i-\alpha_i}(\varphi_i,\delta_i)} J_{t_i-\alpha_i}(\varphi) \leq -J_T(\varphi_i) + \rho,$$

we finally deduce from (4.24) that

$$\begin{aligned}
\limsup_{\varepsilon\to 0, y\to x} \varepsilon \ln \mathbb{P}(X^{\varepsilon,y} \in K) &\leq \sup\left\{ \sup_{1\leq i\leq n} (-J_T(\varphi_i) + \rho), -R \right\} \\
&\leq \sup\left\{ -\inf\{I_{T,x}(\psi), \psi \in K\} + \rho, -R \right\} \leq -u + \rho.
\end{aligned}$$

Since this holds for any $\rho > 0$, the proof of (4.13) for the set K is completed.

Now, let C be a closed subset of $\mathcal{C}([0,T],\mathbb{R}^d)$ such that $\mathcal{C}^1_x([0,T],\mathbb{R}^d \setminus \Gamma)$ is dense in $C \cap \mathcal{C}_x([0,T],\mathbb{R}^d)$. Define the compact set

$$\begin{aligned}
K_k &= \{\psi \in \mathcal{C}([0,T],\mathbb{R}^d) : \|\psi(0) - x\| \leq 1, \forall l \geq k, \omega(\psi, 1/k^3) \leq 1/k\} \\
&= \bigcup_{\|y-x\|\leq 1} K_k^y,
\end{aligned}$$

where K_k^y is defined in (4.14). In order to apply the previous upper bound for compact sets, we are going to construct a compact set $\tilde{K}_k \supset K_k$ such that $S\{(J_T < \infty\})$ is dense in $C \cap \tilde{K}_k \cap \mathcal{C}_x([0,T],\mathbb{R}^d)$. This will be enough to conclude since, by Lemma 4.7,

$$\limsup_{\varepsilon\to 0, y\to x} \varepsilon \ln \mathbb{P}(X^{\varepsilon,y} \notin \tilde{K}_k) \leq \limsup_{\varepsilon\to 0, y\to x} \varepsilon \ln \mathbb{P}(X^{\varepsilon,y} \notin K_k) \leq -k/64d\Sigma^2, \tag{4.25}$$

so that the upper bound (4.13) will be proved as in (4.16).

The set $\tilde{K}_k$ can be constructed as follows. The set $C \cap K_k \cap \mathcal{C}_x([0,T],\mathbb{R}^d)$ is compact, so it is separable. Let $(\psi_n)_{n\geq o}$ be a sequence of functions dense in this set. For all $n \geq 0$, $\psi_n \in C$, so, by assumption, there exists a sequence $(\psi_{n,p})_{p\geq 0}$ in $C \cap \mathcal{C}^1_x([0,T],\mathbb{R}^d \setminus \Gamma)$ converging to ψ_n, such that $\|\psi_{n,p} - \psi_n\|_{0,T} \leq 2^{-p}$ for all $p \geq 0$. Let us define

$$\tilde{K}_k = K_k \cup \left(\bigcup_{n\geq 0} \{\psi_{n,p} : p \geq n\} \right),$$

and let us prove that $\tilde{K}_k$ is compact. Let (ϕ_m) be a sequence of $\tilde{K}_k$. Extracting a converging subsequence is trivial, except in the case where $\{m : \phi_m \in K_k\}$ is finite, and when for all $n \geq 0$, $\{m : \phi_m \in \{\psi_{n,p} : p \geq n\}\}$ is finite. In this case, there exists two increasing sequences of integers (α_m) and (β_m) such that for all $m \geq 0$, $\phi_{\alpha_m} \in \{\psi_{\beta_m,p} : p \geq \beta_m\}$. For all $m \geq 0$, ψ_{β_m} belongs to the compact set $C \cap K_k \cap \mathcal{C}_x([0,T],\mathbb{R}^d)$, so, extracting a subsequence from (β_m), we can assume that $\psi_{\beta_m} \to \psi \in C \cap K_k \cap \mathcal{C}_x([0,T],\mathbb{R}^d)$. Then

$$\|\phi_{\alpha_m} - \psi\|_{0,T} \leq 2^{-\beta_m} + \|\psi_{\beta_m} - \psi\| \to 0$$

when $m \to \infty$. Hence $\tilde{K}_k$ is compact. Moreover, $\tilde{K}_k$ has been constructed in such a way that $\mathcal{C}^1_x([0,T],\mathbb{R}^d \setminus \Gamma)$ is dense in $C \cap \tilde{K}_k \cap \mathcal{C}_x([0,T],\mathbb{R}^d)$, as required. This ends the proof of Theorem 4.6. □

4.4 Proof of Lemmas 4.9 and 4.10

Proof of Lemma 4.9 Let φ be as in any point of Lemma 4.9. We will first restrict ourselves to the case $\varphi = 0$ by means of Girsanov's Theorem. Define on $(\Omega, \mathcal{F}_T)$ the probability measure $\mathbb{P}^{\varepsilon,y}$ by

$$\frac{d\mathbb{P}^{\varepsilon,y}}{d\mathbb{P}} = \exp\left(\frac{1}{\sqrt{\varepsilon}}\int_0^T \dot{\varphi}_s dW_s - \frac{1}{2\varepsilon}\int_0^T \|\dot{\varphi}_s\|^2 ds\right).$$

Since in all cases $J_T(\varphi) = 1/2\int_0^T \|\dot{\varphi}_t\|^2 dt < +\infty$, by Novikov's criterion, Girsanov's Theorem is applicable and implies that

$$\tilde{W}^\varepsilon_t := W_t - \frac{\varphi_t}{\sqrt{\varepsilon}}$$

is a $\mathbb{P}^{\varepsilon,y}$-Brownian motion for $t \leq T$ and that, $\mathbb{P}^{\varepsilon,y}$-a.s., for any $t \leq T$,

$$X^{\varepsilon,y}_t = y + \int_0^t (b^\varepsilon(X^{\varepsilon,y}_s) + \sigma(X^{\varepsilon,y}_s)\dot{\varphi}_s)ds + \sqrt{\varepsilon}\int_0^t \sigma(X^{\varepsilon,y}_s)d\tilde{W}^\varepsilon_s. \tag{4.26}$$

Let

$$\begin{aligned} F^{\varepsilon,y} &= \{\|X^{\varepsilon,y} - S(\varphi)\|_{0,T} \geq \eta,\ \|\sqrt{\varepsilon}W - \varphi\|_{0,T} \leq \delta\} \\ &= \{\|X^{\varepsilon,y} - S(\varphi)\|_{0,T} \geq \eta,\ \|\sqrt{\varepsilon}\tilde{W}^\varepsilon\|_{0,T} \leq \delta\}. \end{aligned}$$

It follows from Cauchy-Schwartz's inequality that

$$\mathbb{P}(F^{\varepsilon,y}) = \int \mathbf{1}_{F^{\varepsilon,y}} \frac{d\mathbb{P}}{d\mathbb{P}^{\varepsilon,y}} d\mathbb{P}^{\varepsilon,y} \leq (\mathbb{P}^{\varepsilon,y}(F^{\varepsilon,y}))^{\frac{1}{2}} \left(\int \left(\frac{d\mathbb{P}}{d\mathbb{P}^{\varepsilon,y}}\right)^2 d\mathbb{P}^{\varepsilon,y}\right)^{\frac{1}{2}}. \tag{4.27}$$

Now,

$$\begin{aligned} \left(\frac{d\mathbb{P}}{d\mathbb{P}^{\varepsilon,y}}\right)^2 &= \exp\left(-\frac{2}{\sqrt{\varepsilon}}\int_0^T \dot{\varphi}_s d\tilde{W}^\varepsilon_s - \frac{1}{\varepsilon}\int_0^T \|\dot{\varphi}_s\|^2 ds\right) \\ &= \exp\left(\int_0^T \left(-\frac{2\dot{\varphi}_s}{\sqrt{\varepsilon}}\right) d\tilde{W}^\varepsilon_s - \frac{1}{2}\int_0^T \left\|\frac{2\dot{\varphi}_s}{\sqrt{\varepsilon}}\right\|^2 ds\right) \\ &\quad \times \exp\left(\frac{1}{\varepsilon}\int_0^T \|\dot{\varphi}_s\|^2 ds\right). \end{aligned}$$

The first term in the product of the right-hand side is a $\mathbb{P}^{\varepsilon,y}$-martingale (by Novikov's criterion), and the second term is equal to $\exp(2J_T(\varphi)/\varepsilon)$. Therefore, (4.27) implies

$$\varepsilon \ln \mathbb{P}(F^{\varepsilon,y}) \leq \frac{\varepsilon}{2} \ln \mathbb{P}^{\varepsilon,y}(F^{\varepsilon,y}) + J_T(\varphi).$$

Therefore, Lemma 4.9 follows from the next result. $\square$

Lemma 4.11 *The three points of Lemma 4.9 hold when (4.20) and (4.21) are replaced respectively by*

$$\limsup_{\varepsilon\to 0, y\to x} \varepsilon \ln \mathbb{P}^{\varepsilon,y}\left(\|X^{\varepsilon,y} - S(\varphi)\|_{0,T} \geq \eta, \|\sqrt{\varepsilon}\tilde{W}^{\varepsilon}\|_{0,T} \leq \delta\right) \leq -R \tag{4.28}$$

and

$$\limsup_{\varepsilon\to 0, y\to x} \varepsilon \ln \mathbb{P}^{\varepsilon,y}\left(\|X^{\varepsilon,y} - S(\varphi)\|_{0,T} \leq \eta, \|\sqrt{\varepsilon}\tilde{W}^{\varepsilon}\|_{0,T} \geq \delta\right) \leq -R. \tag{4.29}$$

Lemma 4.11 relies on the following lemma, of which the proof is postponed after the proof of Lemma 4.11.

Lemma 4.12 *With the previous notation, let Y_t be a $\mathbb{P}^{\varepsilon,y}$-martingale in L^2 such that $\sup_{t\leq T}\|\langle Y\rangle_t\| \leq A$, let τ be a stopping time, and let ξ be a uniformly continuous bounded function on $\mathbb{R}^d$. Then, for any $\eta > 0$ and $R > 0$, there exists $\delta > 0$ and $\varepsilon_0 > 0$ both depending on Y only through A and both independent of τ, such that for any $y \in \mathbb{R}^d$ and $\varepsilon < \varepsilon_0$,*

$$\varepsilon \ln \mathbb{P}^{\varepsilon,y}\left(\left\|\sqrt{\varepsilon}\int_0^{\cdot\wedge\tau} \xi(X_s^{\varepsilon,y})dY_s\right\|_{0,T} \geq \eta,\ \|\sqrt{\varepsilon}Y\|_{0,T} \leq \delta\right) \leq -R. \tag{4.30}$$

Proof of Lemma 4.11 (i) The function $\psi = S(\varphi)$ does not take any value in Γ on $[0,T]$, so there exists $\alpha > 0$ such that $\forall t \in [0,T]$, $\psi_t \in \Gamma_\alpha$. Suppose without loss of generality that $\eta < \alpha/2$, and define for $y \in \mathbb{R}^d$

$$\tau^{\varepsilon,y} = \inf\{t : d(X_t^{\varepsilon,y}, \Gamma) \leq \alpha/2\} \wedge T.$$

When $\tau^{\varepsilon,y} < T$, $\|X^{\varepsilon,y}_{\tau^{\varepsilon,y}} - S(\varphi)_{\tau^{\varepsilon,y}}\| \geq d(S(\varphi)_{\tau^{\varepsilon,y}}, \Gamma) - d(X^{\varepsilon,y}_{\tau^{\varepsilon,y}}, \Gamma) \geq \alpha/2 > \eta$, so

$$\|X^{\varepsilon,y} - S(\varphi)\|_{0,T} \geq \eta \Rightarrow \|X^{\varepsilon,y} - S(\varphi)\|_{0,\tau^{\varepsilon,y}} \geq \eta.$$

Consequently, (4.28) will be proved if we find $\delta > 0$ such that

$$\limsup_{\varepsilon\to 0, y\to x} \varepsilon \ln \mathbb{P}^{\varepsilon,y}(\|X^{\varepsilon,y} - S(\varphi)\|_{0,\tau^{\varepsilon,y}} \geq \eta,\ \|\sqrt{\varepsilon}W^{\varepsilon}\|_{0,T} \leq \delta) \leq -R.$$

Take C such that σ and b are C-Lipschitz and $\tilde{b}$ is bounded by C on $\Gamma_{\alpha/2}$. It follows from (4.26) that, for $t \leq \tau^{\varepsilon,y}$,

$$\begin{aligned}
\|X_t^{\varepsilon,y} - S(\varphi)_t\| &\leq \sqrt{\varepsilon}\left\|\int_0^t \sigma(X_s^{\varepsilon,y})dW_s^{\varepsilon}\right\| + \varepsilon\int_0^t \|\tilde{b}(X_s^{\varepsilon,y})\|ds + \|x-y\| \\
&\quad + \int_0^t \|b(X_s^{\varepsilon,y}) - b(S(\varphi)_s)\|ds + \int_0^t \|\sigma(X_s^{\varepsilon,y}) - \sigma(S(\varphi)_s)\|\,\|\dot{\varphi}_s\|ds \\
&\leq \sqrt{\varepsilon}\left\|\int_0^t \sigma(X_s^{\varepsilon,y})dW_s^{\varepsilon}\right\| + \varepsilon CT + \|x-y\| + C\int_0^t (1 + \|\dot{\varphi}_s\|)\|X_s^{\varepsilon,y} - S(\varphi)_s\|ds.
\end{aligned}$$

Since $u := \int_0^T \|\dot\varphi_s\|^2 ds < +\infty$, by Gronwall's lemma and the Cauchy-Schwartz's inequality, for $t \leq \tau^{\varepsilon,y}$

$$\|X^{\varepsilon,y}_t - S(\varphi)_t\| \leq \left(\sqrt{\varepsilon}\left\|\int_0^t \sigma(X^{\varepsilon,y}_s)dW^\varepsilon_s\right\| + \varepsilon CT + \|x-y\|\right)\exp\left(C\left(T+\sqrt{uT}\right)\right).$$

Therefore, it suffices to find $\delta > 0$ such that

$$\limsup_{\varepsilon\to 0, y\to x} \varepsilon \ln \mathbb{P}^{\varepsilon,y}\left(\sqrt{\varepsilon}\left\|\int_0^t \sigma(X^{\varepsilon,y}_s)dW^\varepsilon_s\right\|_{0,\tau^{\varepsilon,y}} \geq \eta\beta,\ \sqrt{\varepsilon}\|W\|_{0,T} \leq \delta\right) \leq -R,$$

where $\beta = \exp[-C(T+\sqrt{uT})]/2$. This is an direct consequence of Lemma 4.12 with $Y = W^\varepsilon$, $A = 1$, $\xi = \sigma$ and $\tau = \tau^{\varepsilon,y}$. □

Proof of Lemma 4.11 (ii) In Lemma 4.11 (ii), φ is defined from $\tilde\varphi$ by $\varphi_t = \tilde\varphi_t$ for $t \leq t_\psi$, and $\varphi_t = \tilde\varphi_{t_\psi}$ otherwise, where $\psi = S(\tilde\varphi) = S(\varphi)$. By Cauchy-Schwartz's inequality, $\int_0^{t_\psi} \|\dot\varphi_s\|ds \leq (2TJ_T(\varphi))^{1/2} < +\infty$, so there exists $\rho > 0$ small enough such that

$$\int_{t_\psi-\rho}^{t_\psi} \|\dot\varphi_s\|ds \leq \frac{\eta e^{-CT}}{8C}, \tag{4.31}$$

where C is a constant bounding b, $\tilde b$ and σ, and such that b is C-Lipschitz.

Now, we have

$$\{\|X^{\varepsilon,y}-\psi\|_{0,T} \geq \eta,\ \|\sqrt{\varepsilon}W^\varepsilon\|_{0,T} \leq \delta\} \subset D^{\varepsilon,y} \cup E^{\varepsilon,y},$$

where

$$D^{\varepsilon,y} = \left\{\|X^{\varepsilon,y}-\psi\|_{0,t_\psi-\rho} \leq \frac{\eta e^{-CT}}{4},\ \|X^{\varepsilon,y}-\psi\|_{t_\psi-\rho,T} \geq \eta, \|\sqrt{\varepsilon}W^\varepsilon\|_{0,T} \leq \delta\right\}$$
$$\text{and}\quad E^{\varepsilon,y} = \left\{\|X^{\varepsilon,y}-\psi\|_{0,t_\psi-\rho} \geq \frac{\eta e^{-CT}}{4},\ \|\sqrt{\varepsilon}W^\varepsilon\|_{0,t_\psi-\rho} \leq \delta\right\}.$$

Part (i) of Lemma 4.11 shows that $\mathbb{P}^{\varepsilon,y}(E^{\varepsilon,y})$ has the required exponential decay if δ is small enough. Let us estimate $\mathbb{P}^{\varepsilon,y}(D^{\varepsilon,y})$.

It follows from (4.26) and from the fact that $\dot\varphi_t = 0$ for $t > t_\psi$ that, for any $t \geq t_\psi - \rho$

$$\begin{aligned}\|X^{\varepsilon,y}_t - \psi_t\| \leq \|X^{\varepsilon,y}_{t_\psi-\rho} - \psi_{t_\psi-\rho}\| + \sqrt{\varepsilon}\left\|\int_{t_\psi-\rho}^t \sigma(X^{\varepsilon,y}_s)dW^\varepsilon_s\right\| \\ + C\int_{t_\psi-\rho}^t \|X^{\varepsilon,y}_s - \psi_s\|ds + \varepsilon CT + \int_{t_\psi-\rho}^{t_\psi\wedge t} \|\sigma(X^{\varepsilon,y}_s) - \sigma(\psi_s)\|\,]\|\dot\varphi_s\|ds.\end{aligned}$$

On the event $D^{\varepsilon,y}$, the first term of the right-hand side is smaller than $\eta e^{-CT}/4$, and, since σ is bounded by C, the last term is smaller than $2C\int_{t_\psi-\rho}^{t_\psi}\|\dot\varphi\|ds \leq \eta e^{-CT}/4$ by (4.31). Moreover, we can assume ε small enough to have $\varepsilon CT \leq \eta e^{-CT}/4$. So, on the event $D^{\varepsilon,y}$, by Gronwall's Lemma, for $t \geq t_\psi - \rho$,

$$\|X^{\varepsilon,y}_t - \psi_t\| \leq \left(\frac{3}{4}\eta e^{-CT} + \sqrt{\varepsilon}\left\|\int_{t_\psi-\rho}^t \sigma(X^{\varepsilon,y}_s)dW^\varepsilon_s\right\|\right)e^{CT}.$$

Since $\|X^{\varepsilon,y} - \psi\|_{t_\psi-\rho,T} \geq \eta$ on $D^{\varepsilon,y}$, we finally obtain

$$D^{\varepsilon,y} \subset \left\{ \left\| \sqrt{\varepsilon} \int_{t_\psi-\rho}^{\cdot} \sigma(X_s^{\varepsilon,y}) dW_s^\varepsilon \right\|_{t_\psi-\rho,T} \geq \frac{\eta e^{-CT}}{4},\ \|\sqrt{\varepsilon}(W_\cdot^\varepsilon - W_{t_\psi-\rho}^\varepsilon)\|_{t_\psi-\rho,T} \leq 2\delta \right\}$$

Equation (4.28) is now a consequence of Lemma 4.12. □

Proof of Lemma 4.11 (iii) As for Point (i), take $\alpha > 0$ such that $S(\varphi)_t \in \Gamma_\alpha$ for all $t \in [0,T]$. Fix $\eta \leq \alpha/2$. Then, on the event $\{\|X^{\varepsilon,y} - S(\varphi)\|_{0,T} \leq \eta\}$, for any $t \in [0,T]$, $X_t^{\varepsilon,y} \in \Gamma_{\alpha/2}$. Take C such that b and σ are C-Lipschitz and $\tilde{b}$ is bounded by C on $\Gamma_{\alpha/2}$. It follows from (4.26) that, on the event $\{\|X^{\varepsilon,y} - S(\varphi)\|_{0,T} \leq \eta\}$, for any $t \in [0,T]$,

$$\begin{aligned}
\sqrt{\varepsilon}\left\| \int_0^t \sigma(X_s^{\varepsilon,y}) dW_s^\varepsilon \right\| &\leq \|X_t^{\varepsilon,y} - S(\varphi)_t\| + \|y - x\| + \left\| \int_0^t [\sigma(X_s^{\varepsilon,y}) - \sigma(S(\varphi)_s)]\dot{\varphi}_s ds \right\| \\
&\quad + \left\| \int_0^t [b(X_s^{\varepsilon,y}) - b(S(\varphi)_s)] ds \right\| - \left\| \varepsilon \int_0^t \tilde{b}(X_s^{\varepsilon,y}) ds \right\| \\
&\leq 2\eta + C \int_0^T (1 + \|\dot{\varphi}_s\|)\|X_s^\varepsilon - S(\varphi)_s\| ds + \varepsilon CT \\
&\leq \eta(2 + 2CT + C\sqrt{uT})
\end{aligned}$$

if $\varepsilon < \eta$. Therefore,

$$\begin{aligned}
&\{\|X^{\varepsilon,y} - S(\varphi)\|_{0,T} \leq \eta,\ \|\sqrt{\varepsilon}W^\varepsilon\|_{0,T} \geq \delta\} \\
&\subset \left\{ \forall t \in [0,T],\ X_t^{\varepsilon,y} \in \Gamma_{\frac{\alpha}{2}},\ \sqrt{\varepsilon}\left\| \int_0^t \sigma(X_s^{\varepsilon,y}) dW_s^\varepsilon \right\|_{0,T} \leq \eta\beta,\ \sqrt{\varepsilon}\|W^\varepsilon\|_{0,T} \geq \delta \right\},
\end{aligned} \tag{4.32}$$

where $\beta = 2 + 2CT + C\sqrt{uT}$.

Let

$$\begin{aligned}
\tau^{\varepsilon,y} &= \inf\{t : d(X_t^{\varepsilon,y}, \Gamma) \leq \alpha/2\} \wedge T, \\
Y_t^{\varepsilon,y} &= \int_0^t \sigma(X_s^{\varepsilon,y}) dW_s^\varepsilon, \\
\xi &= \chi\sigma^{-1},
\end{aligned}$$

where χ is a Lipschitz function from $\mathbb{R}^d$ to $[0,1]$ such that $\chi(x) = 0$ if $d(x,\Gamma) \leq \alpha/4$ and $\chi(x) = 1$ if $d(x,\Gamma) \geq \alpha/2$. With these notations, (4.32) implies

$$\begin{aligned}
&\{\|X^{\varepsilon,y} - S(\varphi)\|_{0,T} \leq \eta,\ \|\sqrt{\varepsilon}W^\varepsilon\|_{0,T} \geq \delta\} \\
&\qquad \subset \left\{ \sqrt{\varepsilon}\|Y^{\varepsilon,y}\|_{0,T} \leq \eta\beta,\ \sqrt{\varepsilon}\left\| \int_0^{t\wedge\tau^{\varepsilon,y}} \xi(X_s^{\varepsilon,y}) dY_s^{\varepsilon,y} \right\|_{0,T} \geq \delta \right\}.
\end{aligned}$$

Equation (4.29) is now a direct consequence of Lemma 4.12: ξ is Lipschitz and bounded on $\mathbb{R}^d$ by Proposition 2.4 (iii), and for any $t \leq \tau^{\varepsilon,y}$, $\langle Y^{\varepsilon,y}\rangle_t = \int_0^t a(X_s^{\varepsilon,y}) ds$ which is bounded by a constant A independent of y and ε, by Proposition 2.4 (i). □

Let us come to the proof of Lemmas 4.12. It is adapted from the proof of Lemma 1.3 of [18], and makes use of Lemma 4.8.

Proof of Lemma 4.12 We use a discretization technique: for any $p \in \mathbb{N}$, we define $X_t^{\varepsilon,y,p} = X_{k2^{-p}}^{\varepsilon,y}$, where $k \in \mathbb{N}$ is such that $k \leq t2^p < k+1$. Given $\gamma > 0$, $p \geq 1$ and $\delta > 0$, we can write

$$\left\{ \left\| \sqrt{\varepsilon} \int_0^{\cdot\wedge\tau} \xi(X_s^{\varepsilon,y}) dY_s \right\|_{0,T} \geq \eta, \ \|\sqrt{\varepsilon} Y\|_{0,T} \leq \delta \right\} \subset A^\varepsilon \cup B^\varepsilon \cup C^\varepsilon,$$

where

$$A^\varepsilon = \{\|X^{\varepsilon,y} - X^{\varepsilon,y,p}\|_{0,\tau} \geq \gamma\},$$
$$B^\varepsilon = \left\{ \|X^{\varepsilon,y} - X^{\varepsilon,y,p}\|_{0,\tau} \leq \gamma, \ \left\| \sqrt{\varepsilon} \int_0^{\cdot\wedge\tau} [\xi(X_s^{\varepsilon,y}) - \xi(X_s^{\varepsilon,y,p})] dY_s \right\|_{0,T} \geq \frac{\eta}{2} \right\}$$
$$\text{and} \quad C^\varepsilon = \left\{ \left\| \sqrt{\varepsilon} \int_0^{\cdot\wedge\tau} \xi(X_s^{\varepsilon,y,p}) dY_s \right\|_{0,T} \geq \frac{\eta}{2}, \ \|\sqrt{\varepsilon} Y\|_{0,T} \leq \delta \right\}.$$

We will choose γ such that $\mathbb{P}^{\varepsilon,y}(B^\varepsilon)$ is sufficiently small, next $p \geq 1$ to control $\mathbb{P}^{\varepsilon,y}(A^\varepsilon)$, and finally $\delta > 0$ such that $C^\varepsilon = \emptyset$.

First, we apply Lemma 4.8 with $Z_t = \sqrt{\varepsilon}[\xi(X_t^{\varepsilon,y}) - \xi(X_t^{\varepsilon,y,p})]$. Let $M_\gamma := \sup_{\|x-y\| \leq \gamma} \|\xi(x) - \xi(y)\|$, which is finite since ξ is uniformly continuous. Then, on B^ε, $\|Z_t\| \leq \sqrt{\varepsilon} M_\gamma$ for all $t \leq \tau$. Therefore,

$$\mathbb{P}^{\varepsilon,y}(B^\varepsilon) \leq 2d \exp\left(-\frac{\eta^2/4}{2dTA\varepsilon M_\gamma^2} \right).$$

Now, $M_\gamma \to 0$ when $\gamma \to 0$ since ξ is absolutely continuous. Therefore, choosing γ small enough, $\varepsilon \ln \mathbb{P}^{\varepsilon,y}(B^\varepsilon) \leq -2R$ for all $\varepsilon \leq 1$.

Second, $\gamma > 0$ being fixed as above, (4.26) yields

$$\begin{aligned}
&\mathbb{P}^{\varepsilon,y}(\|X^{\varepsilon,y} - X^{\varepsilon,y,p}\|_{0,\tau} \geq \gamma) \\
&\quad \leq \sum_{k=0}^{T2^p-1} \mathbb{P}^{\varepsilon,y} \left(\sup_{k2^{-p} \leq t \leq (k+1)2^{-p}} \left\| \int_{k2^{-p}\wedge\tau}^{t\wedge\tau} \sqrt{\varepsilon} \sigma(X_s^{\varepsilon,y}) dW_s^\varepsilon \right\| \geq \frac{\gamma}{2} \right) \\
&\qquad + \sum_{k=0}^{T2^p-1} \mathbb{P}^{\varepsilon,y} \left(\sup_{k2^{-p} \leq t \leq (k+1)2^{-p}} \left\| \int_{k2^{-p}\wedge\tau}^{t\wedge\tau} [b^\varepsilon(X_s^{\varepsilon,y}) + \sigma(X_s^{\varepsilon,y}) \dot{\varphi}_s] \, ds \right\| \geq \frac{\gamma}{2} \right) \\
&\quad \leq \sum_{k=0}^{T2^p-1} \mathbb{P}^{\varepsilon,y} \left(\sup_{k2^{-p} \leq t \leq (k+1)2^{-p}} \left\| \int_{k2^{-p}\wedge\tau}^{t\wedge\tau} \sqrt{\varepsilon} \sigma(X_s^{\varepsilon,y}) dW_s^\varepsilon \right\| \geq \frac{\gamma}{2} \right) \\
&\qquad + \sum_{k=0}^{T2^p-1} \mathbb{P}^{\varepsilon,y} \left(C2^{-p} + C2^{-p/2}\sqrt{u} \geq \frac{\gamma}{2} \right),
\end{aligned}$$

where C is a bound for b^ε and σ and $u = \int_0^T \|\dot{\varphi}_s\|^2 ds < +\infty$. For p big enough, the second sum of the right-hand side equals 0. For the first sum, Lemma 4.8 with $\tau = T = 2^{-p}$, $Y = W^\varepsilon$, $A = 1$, $R = \gamma/2$ and $B = \sqrt{\varepsilon} C$ gives that

$$\mathbb{P}^{\varepsilon,y} \left(\sup_{k2^{-p} \leq t \leq (k+1)2^{-p}} \left\| \int_{k2^{-p}\wedge\tau}^{t\wedge\tau} \sqrt{\varepsilon} \sigma(X_s^{\varepsilon,y}) dW_s^\varepsilon \right\| \geq \frac{\gamma}{2} \right) \leq 2d \exp\left(-\frac{\gamma^2/4}{2d2^{-p}C^2\varepsilon} \right)$$

for all $0 \leq k < T2^p$. Therefore, taking p large enough, $\varepsilon \ln \mathbb{P}^{\varepsilon,y}(A^\varepsilon) \leq -2R$ for all $\varepsilon \leq 1$.

Third, with $p \geq 1$ and $\gamma > 0$ as above, for $t \leq T$,

$$\sqrt{\varepsilon} \int_0^{t\wedge\tau} \xi(X_s^{\varepsilon,y,p}) dY_s = \sum_{i=0}^{T2^p-1} \sqrt{\varepsilon}\xi(X^{\varepsilon,y}_{i2^{-p}\wedge\tau})[Y_{(i+1)2^{-p}\wedge t\wedge\tau} - Y_{i2^{-p}\wedge t\wedge\tau}].$$

Therefore, since $\|\sqrt{\varepsilon}Y\|_{[0,T]} \leq \delta$ on the event C^ε, we have for all $t \leq T$

$$\left\|\sqrt{\varepsilon} \int_0^{t\wedge\tau} \xi(X_s^{\varepsilon,y,p}) dY_s\right\| \leq \sum_{i=0}^{T2^p-1} 2\delta C,$$

where C is a bound for ξ. Hence $C^\varepsilon = \emptyset$ as soon as $\delta < \eta 2^{-(p+2)}/CT$.

We finally obtain that $\varepsilon \ln \mathbb{P}^{\varepsilon,y}(A^\varepsilon \cup B^\varepsilon \cup C^\varepsilon) \leq \varepsilon \ln 2 - 2R$, which yields (4.30) for ε small enough.

This argument is true for any $y \in \mathbb{R}^d$ and for any stopping time τ. It remains to observe that A is the only information about Y that we used to estimate $\mathbb{P}^{\varepsilon,y}(B^\varepsilon)$, that Y does not appear in A^ε, and that no assumption about Y is necessary to obtain $C^\varepsilon = \emptyset$. Hence, the constant A is the only information about Y required to obtain δ and ε_0. $\square$

The proof of Lemma 4.9 is now completed. It only remains to prove Lemmas 4.10.

Proof of Lemma 4.10 Let us first prove Point (i). Take $\psi \in \tilde{\mathcal{C}}^{ac}_x([0,T],\mathbb{R}^d)$. Any $\varphi \in \mathcal{C}^{ac}_0([0,T],\mathbb{R}^d)$ such that $S(\varphi) = \psi$ must satisfy for any $t \in [0, t_\psi)$

$$\dot{\psi}_t = b(\psi_t) + \sigma(\psi_t)\dot{\varphi}_t.$$

Therefore, such a φ is uniquely defined for $t < t_\psi$ by

$$\dot{\varphi}_t = \sigma^{-1}(\psi_t)[\dot{\psi}_t - b(\psi_t)]. \tag{4.33}$$

Thus

$$I_{T,x}(\psi) = \frac{1}{2}\int_0^{t_\psi} \|\sigma^{-1}(\psi_t)[\dot{\psi}_t - b(\psi_t)]\|^2 dt = \frac{1}{2}\int_0^{t_\psi} \|\dot{\varphi}_t\|^2 dt \leq J_T(\varphi)$$

for any φ such that $S(\varphi) = \psi$, and $I_{T,x}(\psi) = J_T(\varphi)$ if and only if $\dot{\varphi}_t = 0$ for all $t > t_\psi$.

This trivially implies that $I_{T,x}(\psi) = \inf\{J_T(\varphi),\ S(\varphi) = \psi\}$ when $I_{T,x}(\psi) = +\infty$. In the case where $I_{T,x}(\psi) < +\infty$, we clearly have $I_{T,x}(\psi) \leq \inf\{J_T(\varphi),\ S(\varphi) = \psi\}$. To prove the converse inequality, it suffices to check that there exists an absolutely continuous function φ satisfying (4.33) for $t < t_\psi$ and $\dot{\varphi}_t = 0$ for $t \geq t_\psi$. This is equivalent to the fact that $\sigma^{-1}(\psi_t)[\dot{\psi}_t - b(\psi_t)]$ is L^1 on $[0, t_\psi]$. Since $I_{T,x}(\psi) < +\infty$, this function is actually L^2, which ends the proof of Point (i).

For Point (ii), remind that σ is uniformly non-degenerate on Γ_α for any $\alpha > 0$. Therefore, the fact that $\mathcal{C}^1([0,T],\mathbb{R}^d \setminus \Gamma) \subset S(\{J_T < \infty\})$ follows from (4.33). Since $S(\{J_T < \infty\}) \subset \tilde{\mathcal{C}}^{ac}_x([0,T],\mathbb{R}^d)$ and any function of $\tilde{\mathcal{C}}^{ac}_x([0,T],\mathbb{R}^d)$ is the limit of elements of $\mathcal{C}^1([0,T],\mathbb{R}^d \setminus \Gamma)$, Point (ii) is clear. $\square$

5 Application to the problem of exit from a domain

We study in this section the biological phenomenon of punctualism. We first state our result on the problem of exit from a domain in Section 5.1. Next (Section 5.2) we explain how it applies to the example of Sections 2.4 and 3.5. Finally, we prove this result in Section 5.3.

5.1 The result

We consider a bounded open subset G of $\mathbb{R}^d$ containing a unique, stable equilibrium of the canonical equation of adaptive dynamics $\dot{\phi} = b(\phi)$. We will assume for convenience that this equilibrium is 0. Note that the equilibria of the canonical equation are exactly the points of Γ. As observed in Remark 4.3, when ε is small, $X^{\varepsilon,x}$ is close to the solution of this ODE with initial state x on bounded time inervals with high probability. Yet, the diffusion phenomenon may almost surely drive $X^{\varepsilon,x}$ out of G. Our next result gives estimates of the time and position of exit of X^ε from G ("problem of exit from a domain" [24]).

We will follow closely Section 5.7 of Dembo and Zeitouni [13], where a similar result is proved for non-degenerate diffusions.

When the initial condition of the solution of the SDE (1.1) constructed in Proposition 3.1 is not precised, it will by denoted by X^ε. The value of X^ε at time 0 will then be specified by considering the probability of events involving X^ε under $\mathbb{P}_x$, which is the law of the process $X^{\varepsilon,x}$. Expectations with respect to $\mathbb{P}_x$ will be denoted $\mathbb{E}_x$. We will also use throughout this section the notation $B(\rho) := \{y \in \mathbb{R}^d : \|y\| \leq \rho\}$ and $S(\rho) = \{y \in \mathbb{R}^d : \|y\| = \rho\}$ for $\rho > 0$. It will always be implicitly assumed that $\rho > 0$ is small enough to have $B(\rho) \subset G$ and $S(\rho) \subset G$.

We will assume $d \geq 2$. Otherwise, the problem has few interest: if $G = (c, c') \subset \mathbb{R}$ contains a unique point x of Γ, and if $y > x$ (say), the process $X^{\varepsilon,y}$ can exit G, only at c', and the probability of reaching x before c' can be computed explicitly using classical results on one-dimensional diffusion processes [28, Prop. 5.5.22].

Let

$$V(y, z, t) = \inf_{\{\psi \in \mathcal{C}([0,t],\mathbb{R}^d):\psi(0)=y,\psi(t)=z\}} \tilde{I}_{t,y}(\psi),$$

which is, heuristically, the cost of forcing $X^{\varepsilon,y}$ to be at z at time t. Define also

$$V(y, z) = \inf_{t>0} V(y, z, t).$$

The function $V(0, z)$ is called the *quasi-potential* [24].

Six assumptions are required for our result:

(Ha) G is a bounded open subset of $\mathbb{R}^d$ such that $G \cap \Gamma = \{0\}$ and with sufficiently smooth boundary ∂G for

$$\tau^\varepsilon = \inf\{t > 0 : X_t^\varepsilon \in \partial G\}$$

to be a well-defined stopping time. Moreover, for any solution of

$$\dot{\phi} = b(\phi) \tag{5.1}$$

such that $\phi(0) \in G$, we have $\phi(t) \in G$ for all $t > 0$ and $\lim_{t\to\infty} \phi(t) = 0$.

(Hb) $\bar{V} := \inf_{z\in\partial G} V(0,z) < \infty$.

(Hc) For any $\varepsilon > 0$ and $y \in G \setminus \{0\}$, $\mathbb{P}_y\left(\lim_{t\to\infty} X^\varepsilon_t = 0\right) = 0$.

(Hd) The points of Γ are isolated in $\mathbb{R}^d$.

(He) For any $y \in \overline{G} \cap \Gamma$, g is $\mathcal{C}^2$ at (y, y) and $H_{1,1}g(y, y) + H_{1,2}g(y, y)$ is invertible.

(Hf) All the trajectories of the deterministic system (5.1) with initial value in ∂G converges to 0 as $t \to \infty$.

Assumption (Ha) states that the domain G is an *attracting* domain for (5.1). If Assumption (Hb) fails, all points of ∂G are equally unlikely on the large deviations scale. We have given in Theorem 3.6 (sections 3.4) conditions under which (Hc) holds. Assumption (Hd) is required in the large deviation Theorem 4.1. We have already encountered an assumption similar to (He) in Propositions 3.7 and 4.5. It allows to control the non-degeneracy of $a(x)$ near $\overline{G} \cap \Gamma$. Finally, Assumption (Hf) prevents situations where ∂G is the characteristic boundary of the domain of attraction of 0. This last assumption is needed only for Point (b) of the following result. Note that when (Hf) is true, $\overline{G} \cap \Gamma = \{0\}$

Theorem 5.1
(a) *Assume (H) and (Ha–e). Then, for all $x \in G \setminus \{0\}$ and $\delta > 0$,*

$$\lim_{\varepsilon\to 0} \mathbb{P}_x(\tau^\varepsilon > e^{(\bar{V}-\delta)/\varepsilon}) = 1. \tag{5.2}$$

(b) *Assume (H) and (Ha–f). If N is a closed subset of ∂G and if $\inf_{z\in N} V(0,z) > \bar{V}$, then for any $x \in G \setminus \{0\}$,*

$$\lim_{\varepsilon\to 0} \mathbb{P}_x(X^\varepsilon_{\tau^\varepsilon} \in N) = 0. \tag{5.3}$$

In particular, if there exists $z^* \in \partial G$ such that $V(0,z^*) < V(0,z)$ for all $z \in \partial G\setminus\{z^*\}$, then, for any $\delta > 0$ and $x \in G \setminus \{0\}$,

$$\lim_{\varepsilon\to 0} \mathbb{P}_x(\|X^\varepsilon_{\tau^\varepsilon} - z^*\| < \delta) = 1. \tag{5.4}$$

5.2 Example

We consider again the example of Sections 2.4 and 3.5. We remind that it satisfies the assumptions of Theorem 4.1, that $\Gamma = \{-e_1, 0, e_1\}$ and that $-e_1$ and e_1 are stable equilibria of the canonical equation, with respective basins of attraction $(-\infty, 0)\times\mathbb{R}^{d-1}$ and $(0,\infty)\times\mathbb{R}^{d-1}$. We will assume that X^ε_0 belongs to the basin of attraction of $-e_1$.

In order to apply Theorem 5.1, we need to assume $d \geq 2$ and to take $G \subset (-\infty, 0)\times\mathbb{R}^{d-1}$ for Assumption (Ha) to hold. It is easy to prove that Assumption (Hb) is satisfied, for example by constructing a function ψ on $[0, T]$ linking $-e_1$ to 0 in $[-e_1, 0]\times\{0\}^{d-1}$ having a similar behaviour near $-e_1$ and 0 as the function (4.8) in the proof of Propositon 4.5. Then $I_{T-2\delta,x}(\psi_\delta)$ is finite and bounded as a function of $\delta > 0$, where ψ_δ is the restriction of ψ on the time interval $[\delta, T - \delta]$. Conditions ensuring Assumption (Hc) have been obtained in Section 3.5. Assumption (Hd) holds and Assumption (He) has been proved in Section 3.5.

Then, Theorem 5.1 (a) gives a lower bound on the first time τ^ε of exit from G (which is a.s. finite).

Now, observe that, by (2.11), for all $x \in \mathbb{R}^d$,

$$a^{-1}(x) = \frac{\sqrt{2\pi}}{\|\nabla_1 g(x,x)\|}\left(\mathrm{Id} - \frac{1}{2}\,\frac{\nabla_1 g(x,x)\nabla_1 g(x,x)^*}{\|\nabla_1 g(x,x)\|^2}\right).$$

Moreover, it follows from (2.10) that $b(x)\cdot x \le -C\|b(x)\|\|x\|$ for some constant C and for $\|x\|$ bigger than some $R_0 > 0$. Therefore, for any ψ linking $-e_1$ to z such that $\|z\| = R \ge R_0$,

$$\begin{aligned}
I_{T,x}(\psi) &\ge C\int_0^T \left|\dot\psi(t)\cdot\frac{\psi(t)}{\|\psi(t)\|} - b(\psi(t))\cdot\frac{\psi(t)}{\|\psi(t)\|}\right|^2 \frac{1}{\|b(\psi(t))\|}\mathbf{1}_{\{\dot\psi(t)\cdot\psi(t)>0,\ \|\psi(t)\|\ge R_0\}}dt \\
&\ge C\int_0^T \dot\psi(t)\cdot\frac{\psi(t)}{\|\psi(t)\|}\mathbf{1}_{\{\dot\psi(t)\cdot\psi(t)>0,\ \|\psi(t)\|\ge R_0\}}dt \ge C(R-R_0),
\end{aligned}$$

where the constant C may change from line to line. This quantity is larger than $\bar V$ if R is big enough. Therefore, taking $G = [(-\infty,0)\times\mathbb{R}^{d-1}]\cap B(0,R)$ for R large enough in Theorem 5.1 (a), it is clear from the proof of this result that $X^\varepsilon_{\tau^\varepsilon} \in \{0\}\times\mathbb{R}^{d-1}$. Hence, by symmetry of the model, the process X^ε has a probability 1/2 to enter $B(e_1,\eta)$ before $B(-e_1,\eta)$ after time τ^ε. In this case, we say that there is a change of basin of attraction for the process, or in biological words, that there is an evolutionary jump to a new punctuated equilibrium. Next a new change of basin of attraction can occur. It is then clear that Theorem 5.1 (a) gives a lower bound for the time scale of punctuated evolution.

In order to apply Theorem 5.1 (b), G must satisfy Assumption (Hf), which requires that $\overline{G}\subset(-\infty,0)\times\mathbb{R}^{d-1}$ in our example. This type of results is useful in the case where there are several other basins of attractions, in order to decide which one is the next visited. Note that, to this end, one would need to take for G the whole basin of attraction, which contradicts Assumption (Hf). This problem is solved for non-degenerate diffusions in [24, Ch. 6], where the asymptotic chain of visits of basins of attractions is described in terms of the quasi-potential V. Our degenerate case is not covered by these results and the extension is not trivial. However, since we were able to prove Theorem 5.1, we conjecture that our method could be extended to describe the chain of visits of basins of attraction in general models and to provide lower estimates on the sequence of first times of visit. Note that, in our example, the sequence of punctuated equilibria visited by the evolutionary process is trivial, since there are only two of them.

5.3 Proof of Theorem 5.1

The proof of results like Theorem 5.1 is classically guided by the heuristics that, as $\varepsilon\to 0$, X^ε wanders around 0 for an exponentially long time, during which its chance of hitting a closed set $N\subset\partial G$ is determined by $\inf_{z\in N} V(0,z)$. Any excursion off the stable point 0 has an overwhelmingly high probability of being pulled back near 0, and it is not the time spent near any part of ∂G that matters but the *a priori* chance for a direct, fast exit due to a rare segment in the Brownian motion's path.

Usually, such results also include an upper bound for τ^ε. We are not able to obtain such a result because of the singularity of the process X^ε at 0. Because the matrix $a(x)$

is 0 at $x = 0$, the time spent by the process near 0 before hitting $S(\rho)$ is not uniformly bounded (in probability) with respect to the initial condition (actually, it is even infinite when $X_0^\varepsilon = 0$).

For this reason, the proof of a similar result in Dembo and Zeitouni [13] (Thm. 5.7.11 and Cor. 5.7.16) cannot be directly adapted to our situation. Below, we are only going to detail the steps that must be modified. In particular, Theorem 5.1 (a) will be obtained exactly as in [13], whereas Point (b) has to be obtained without using any upper bound on τ^ε.

We are going to use four lemmas. The first one gives estimates on continuity of $V(x, \cdot, t)$ around 0 and ∂G.

Lemma 5.2 *Assume (H), (Hd) and (He). For any $\delta > 0$, there exists $\rho > 0$ small enough such that*

$$\sup_{(x,y)\in(B(\rho)\setminus\{0\})\times B(\rho)} \inf_{t\in[0,1]} V(x, y, t) < \delta \tag{5.5}$$

and

$$\sup_{\{(x,y)\in(\mathbb{R}^d\setminus\Gamma)\times\mathbb{R}^d,\ \inf_{z\in\partial G}(\|y-z\|+\|x-z\|)\leq\rho\}} \inf_{t\in[0,1]} V(x, y, t) < \delta. \tag{5.6}$$

For the next lemmas, we define

$$\sigma_\rho := \inf\{t \geq 0 : X^\varepsilon \in B(\rho) \cup \partial G\}.$$

The second lemma gives a uniform lower bound on the probability of an exit from G starting from a small sphere around 0 before hitting an even smaller sphere.

Lemma 5.3 *Assume (H) and (Ha–e). Then*

$$\lim_{\rho\to 0} \liminf_{\varepsilon\to 0} \varepsilon \ln \inf_{y\in S(2\rho)} \mathbb{P}_y(X^\varepsilon_{\sigma_\rho} \in \partial G) \geq -\bar{V}.$$

The following upper bound relates the quasi-potential $V(0, \cdot)$ with the probability that an excursion starting from a small sphere around 0 hits a given subset of ∂G before hitting an even smaller sphere.

Lemma 5.4 *Assume (H) and (Ha–f). For any closed set $N \subset \partial G$,*

$$\lim_{\rho\to 0} \limsup_{\varepsilon\to 0} \varepsilon \ln \sup_{y\in S(2\rho)} \mathbb{P}_y(X^\varepsilon_{\sigma_\rho} \in N) \leq -\inf_{z\in N} V(0, z)$$

The last lemma is used to extend the previous upper bound to any initial condition $x \in G$.

Lemma 5.5 *Assume (H) and (Ha). For every $\rho > 0$ such that $B(\rho) \subset G$ and all $x \in G$,*

$$\lim_{\varepsilon\to 0} \mathbb{P}_x(X^\varepsilon_{\sigma_\rho} \in B(\rho)) = 1.$$

The statements of Lemmas 5.2, 5.4 and 5.5 are the same as Lemmas 5.7.8, 5.7.21 and 5.7.22 of [13], respectively. Among them, Lemmas 5.4 and 5.5 can be deduced from Corollary 4.4 exactly as in [13], so we omit their proof. Because of the degeneracy of X^ε at 0, Lemma 5.2 must be proved with a different method. Finally, Lemma 5.3 replaces

Lemma 5.7.18 of [13] and is very different since it gives no upper control on τ^ε. This lemma and the proof of Theorem 5.1 (b) are the new part of our proof.

Theorem 5.1 (a) can be proved exactly as the corresponding inequalities in Theorem 5.7.11 and Corollary 5.7.16 of [13]. It makes use of our Lemmas 5.2, 5.4 and 5.5, and of Lemmas 5.7.19 and 5.7.23 of [13], which can be proved exactly as therein. One simply must take care that x belongs to $G \setminus \{0\}$ instead of G. Let us omit this proof.

We first give the proof of Theorem 5.1 (b) and next those of Lemmas 5.2 and 5.3.

Proof of Theorem 5.1 (b) Let $\rho > 0$ be small enough to have $B(2\rho) \subset G$ (the precise choice of ρ will be specified later). Let $\theta_0 = 0$ and for $m = 0, 1, \dots$ define the stopping times

$$\begin{aligned} \tau_m &= \inf\{t \geq \theta_m : X^\varepsilon_t \in B(\rho) \cup \partial G\}, \\ \theta_{m+1} &= \inf\{t > \tau_m : X^\varepsilon_t \in S(2\rho)\}, \end{aligned} \tag{5.7}$$

with the convention that $\theta_{m+1} = \infty$ if $X^\varepsilon_{\tau_m} \in \partial G$. Each interval $[\tau_m, \tau_{m+1}]$ represents one significant excursion off $B(\rho)$. Note that, necessarily, $\tau^\varepsilon = \tau_m$ for some integer m.

First, Assumption (Hc) implies that $\theta_{m+1} < \infty$ as soon as $X^\varepsilon_{\tau_m} \in B(\rho)$. This can be proved as follows.

On the one hand, Assumption (Hc) implies that, for all $x \in S(\rho)$,

$$\lim_{\alpha \to 0} \mathbb{P}_x(\limsup_{t \to +\infty} \|X^\varepsilon_t\| \geq \alpha) = 1. \tag{5.8}$$

On the other hand, for any $\alpha > 0$, X^ε is a diffusion with bounded drift part and uniformly non-degenerate diffusion part in $B(2\rho) \cap \Gamma_{\alpha/2}$. Therefore, X^ε has a uniformly positive probability to reach $S(2\rho)$ before $S(\alpha/2)$ starting from any point of $S(\alpha)$. Hence, by the strong Markov property of Proposition 3.3, for all $x \in S(\rho)$,

$$\mathbb{P}_x(\theta_1 < \infty \mid \limsup_{t \to +\infty} \|X^\varepsilon_t\| \geq \alpha) = 1.$$

Combining this with (5.8) we have that $\mathbb{P}_x(\theta_1 < \infty) = 1$ for all $x \in S(\rho)$, which implies the required result.

Second, fix a closed set $N \subset G$ such that $\bar{V}_N := \inf_{z \in N} V(0, z) > \bar{V}$. Assume $\bar{V}_N < \infty$ (otherwise, $\bar{V}_N$ may be replaced by any arbitrary large constant in the proof below). Fix $\eta > 0$ such that $\eta < (\bar{V}_N - \bar{V})/3$. Applying Lemmas 5.3 and 5.4, we fix $\rho > 0$ and $\varepsilon_0 > 0$ such that

$$\inf_{y \in S(2\rho)} \mathbb{P}_y(X^\varepsilon_{\sigma_\rho} \in \partial G) \geq e^{-(\bar{V} + \eta)/\varepsilon}, \quad \forall \varepsilon \leq \varepsilon_0 \tag{5.9}$$

and

$$\sup_{y \in S(2\rho)} \mathbb{P}_y(X^\varepsilon_{\sigma_\rho} \in N) \leq e^{-(\bar{V}_N - \eta)/\varepsilon}, \quad \forall \varepsilon \leq \varepsilon_0.$$

Fix $y \in B(\rho)$. For any $l \geq 1$, we have

$$\mathbb{P}_y(X^\varepsilon_{\tau^\varepsilon} \in N) \leq \mathbb{P}_y(\tau^\varepsilon > \tau_l) + \sum_{m=1}^{l} \mathbb{P}_y(\tau^\varepsilon = \tau_m \text{ and } X^\varepsilon_{\tau^\varepsilon} \in N). \tag{5.10}$$

The second term can be bounded as follows: for $m \geq 1$, $y \in B(\rho)$ and $\varepsilon \leq \varepsilon_0$, it follows from the strong Markov property that

$$\begin{aligned}\mathbb{P}_y(\tau^\varepsilon = \tau_m \text{ and } X^\varepsilon_{\tau^\varepsilon} \in N) &= \mathbb{P}_y(\tau^\varepsilon > \tau_{m-1})\mathbb{P}_y(X^\varepsilon_{\tau_m} \in N \mid \tau^\varepsilon > \tau_{m-1}) \\ &= \mathbb{P}_y(\tau^\varepsilon > \tau_{m-1})\mathbb{E}_y[\mathbb{P}_{X^\varepsilon_{\theta_m}}(X^\varepsilon_{\sigma_\rho} \in N) \mid \tau^\varepsilon > \tau_{m-1}] \\ &\leq \sup_{x \in S(2\rho)} \mathbb{P}_x(X^\varepsilon_{\sigma_\rho} \in N) \leq e^{-(\bar{V}_N - \eta)/\varepsilon}.\end{aligned}$$

Concerning the first term of the right-hand side of (5.10), for any $l \geq 1$ and $y \in B(\rho)$,

$$\mathbb{P}_y(\tau^\varepsilon > \tau_l) = \mathbb{E}_y[\mathbb{P}_{X^\varepsilon_{\theta_1}}(\tau^\varepsilon > \tau_{l-1})] \leq \sup_{x \in S(2\rho)} \mathbb{P}_x(\tau^\varepsilon > \tau_{l-1}),$$

and, for any $x \in S(2\rho)$ and $k \geq 1$,

$$\begin{aligned}\mathbb{P}_x(\tau^\varepsilon > \tau_k) &= [1 - \mathbb{P}_x(\tau^\varepsilon = \tau_k \mid \tau^\varepsilon > \tau_{k-1})]\mathbb{P}_x(\tau^\varepsilon > \tau_{k-1}) \\ &= [1 - \mathbb{E}_x[\mathbb{P}_{X^\varepsilon_{\theta_k}}(X^\varepsilon_{\sigma_\rho} \in \partial G) \mid \tau^\varepsilon > \tau_{k-1}]]\mathbb{P}_x(\tau^\varepsilon > \tau_{k-1}) \\ &\leq (1-q)\mathbb{P}_x(\tau^\varepsilon > \tau_{k-1}),\end{aligned}$$

where $q := \inf_{y \in S(2\rho)} \mathbb{P}_y(X^\varepsilon_{\sigma_\rho} \in \partial G) \geq e^{-(\bar{V}+\eta)/\varepsilon}$ by (5.9). Therefore,

$$\sup_{y \in S(2\rho)} \mathbb{P}_x(\tau^\varepsilon > \tau_k) \leq (1-q)^k.$$

Putting together these estimates in (5.10), we obtain that, for all $y \in B(\rho)$ and $\varepsilon \leq \varepsilon_0$

$$\mathbb{P}_y(X^\varepsilon_{\tau^\varepsilon} \in N) \leq \left(1 - e^{-\frac{\bar{V}+\eta}{\varepsilon}}\right)^{l-1} + l e^{-\frac{\bar{V}_N - \eta}{\varepsilon}}.$$

We choose $l = \lfloor 2e^{(\bar{V}+2\eta)/\varepsilon} \rfloor$, where $\lfloor \cdot \rfloor$ denotes the integer part function. Then, for ε small enough, $l - 1 > e^{(\bar{V}+2\eta)/\varepsilon}$ and

$$\mathbb{P}_y(X^\varepsilon_{\tau^\varepsilon} \in N) \leq \left(\left(1 - \frac{1}{u_\varepsilon}\right)^{u_\varepsilon}\right)^{e^{\eta/\varepsilon}} + 2e^{\frac{\bar{V} - \bar{V}_N + 3\eta}{\varepsilon}},$$

where $u_\varepsilon := e^{(\bar{V}+\eta)/\varepsilon}$. Since $u_\varepsilon \to +\infty$, we have $(1 - 1/u_\varepsilon)^{u_\varepsilon} \to 1/e$, and, finally,

$$\lim_{\varepsilon \to 0} \sup_{y \in B(\rho)} \mathbb{P}_y(X^\varepsilon_{\tau^\varepsilon} \in N) = 0.$$

The proof of (5.3) is now completed by combining Lemma 5.5 and the inequality

$$\mathbb{P}_x(X^\varepsilon_{\tau^\varepsilon} \in N) \leq \mathbb{P}_x(X^\varepsilon_{\sigma_\rho} \notin B(\rho)) + \sup_{y \in B(\rho)} \mathbb{P}_y(X^\varepsilon_{\tau^\varepsilon} \in N).$$

Applying (5.3) to $N = \{z \in \partial G : \|z - z^*\| \geq \delta\}$ and observing that Lemma 5.2 implies the continuity of $z \mapsto V(0, z)$ on ∂G, we easily obtain (5.4). $\square$

Proof of Lemma 5.2 (5.5) Fix $\delta, \rho > 0$, $x \in B(\rho) \setminus \{0\}$ and $y \in B(\rho)$. In order to simplify the notations, we will use the complex notation for the coordinates of points of the (two-dimensional) plane of $\mathbb{R}^d$ containing 0, x and y, and we will assume that $x = r \in \mathbb{R}$ and $y = r'e^{i\theta}$, with $0 < r \leq \rho$ and $0 \leq r' \leq \rho$. Define $\psi \in \mathcal{C}([0,1], B(\rho))$ by

$$\psi(t) = \begin{cases} (1-(3t)^2)r + (3t)^2\rho & \text{if} \quad 0 \leq t \leq 1/3 \\ \rho e^{i\theta(3t-1)} & \text{if} \quad 1/3 \leq t \leq 2/3 \\ (1-(3-3t)^2)r'e^{i\theta} + (3-3t)^2\rho e^{i\theta} & \text{if} \quad 2/3 \leq t \leq 1. \end{cases}$$

Then $\psi(0) = x$ and $\psi(1) = y$, and $\psi(t) \in B(\rho) \setminus \{0\}$ for any $t \in [0,1)$. Moreover, for $0 \leq t \leq 1/3$, $\psi(t) = r + 9t^2(\rho - r)$, so that $\|\psi(t)\| \geq 9t^2(\rho - r)$, and, similarly, for $2/3 \leq t \leq 1$, $\|\psi(t)\| \geq 9(1-t)^2(\rho - r')$. Thanks to assumption (He), a calculation similar to equation (4.9) in the proof of Proposition 4.5 gives that, with the same K, $\mathcal{N}_0$ and a_0 as therein, if $B(\rho) \subset \mathcal{N}_0$,

$$\begin{aligned} I_{1,x}(\psi) &\leq \frac{1}{2a_0}\left(\int_0^{1/3} \frac{2(18t(\rho-r))^2 + 2K^2\|\psi(t)\|^2}{\|\psi(t)\|}dt \right. \\ &\qquad + \int_{1/3}^{2/3} \frac{2(3\theta\rho)^2 + 2K^2\|\psi(t)\|^2}{\|\psi(t)\|} \\ &\qquad \left. + \int_{2/3}^{1} \frac{2(18(1-t)(\rho-r'))^2 + 2K^2\|\psi(t)\|^2}{\|\psi(t)\|}dt \right) \\ &\leq \frac{1}{2a_0}\left(\int_0^{1/3}(648(\rho-r) + 2K^2\|\psi(t)\|)dt + \int_{1/3}^{2/3}(18\theta^2 + 2K^2)\rho dt \right. \\ &\qquad \left. + \int_{2/3}^{1}(648(\rho-r') + 2K^2\|\psi(t)\|)dt \right) \\ &\leq \frac{(216 + 2K^2/3)\rho + (6\theta^2 + 2K^2/3)\rho + (216 + 2K^2/3)\rho}{2a_0}. \end{aligned}$$

Consequently, for sufficiently small $\rho > 0$ not depending on x and y, $I_{1,x}(\psi) \leq \delta/2$, which yields (5.5). □

Proof of Lemma 5.2 (5.6) Fix $\delta > 0$. Thanks to Assumption (He), using the same method as above, for any $z \in \partial G \cap \Gamma$, one can find a positive ρ_z such that

$$\sup_{(x,y)\in(B(z,\rho_z)\setminus\{0\})\times B(z,\rho_z)} \inf_{t\in[0,1]} V(x,y,t) < \delta/2, \tag{5.11}$$

where $B(z,r)$ is the closed ball centered at z with radius r

Let $\bar{\rho}_0$ be the infimum of ρ_z for $z \in \partial G \cap \Gamma$. Since G is bounded, because of Assumption (Hd), this set is finite and $\bar{\rho}_0 > 0$. Reducing $\bar{\rho}_0$ if necessary, we can assume that $B(\bar{\rho}_0) \subset G$ and that $d(\Gamma \cap (\mathbb{R}^d \setminus \overline{G}), \overline{G}) > \bar{\rho}_0$.

Fix x and y in $\mathbb{R}^d \setminus \bigcup_{z\in\partial G\cap\Gamma} B(z,\bar{\rho}_0)$ and assume that there exists $z \in \partial G$ with $\|x - z\| + \|y - z\| \leq \bar{\rho}_0/3$. Then $d(x,\Gamma) > 2\bar{\rho}_0/3$ and $d(y,\Gamma) > 2\bar{\rho}_0/3$. Moreover, since $\|x - y\| \leq \bar{\rho}_0/3$, the segment $[x,y]$ is included in $\Gamma_{\bar{\rho}_0/3}$.

Now, for any $t_0 > 0$, x and y such that $[x,y] \subset \Gamma_{\bar{\rho}_0/3}$, define $\psi^{(t_0)} \in \mathcal{C}([0,t_0],\mathbb{R}^d)$ by

$$\psi^{(t_0)}(t) = \left(1 - \frac{t}{t_0}\right) x + \frac{t}{t_0} y$$

for $0 \leq t \leq t_0$. Then $\psi^{(t_0)}(0) = x$ and $\psi^{(t_0)}(t_0) = y$ and $\psi^{(t_0)}(t) \in \Gamma_{\bar{\rho}_0/3}$ for all $t \in [0,t_0]$.

Since a is uniformly non-degenerate on $\Gamma_{\bar{\rho}_0/3}$, there exists a constant C bounding the eigenvalues of a^{-1} on this set. Then

$$\begin{aligned} I_{t_0,x}(\psi^{(t_0)}) &\leq \frac{C}{2} \int_0^{t_0} (\|\dot{\psi}^{(t_0)}(t)\|^2 + \|b(\psi^{(t_0)}(t))\|^2) dt \\ &\leq \frac{C}{2} \left(\frac{\|x-y\|^2}{t_0} + B^2 t_0 \right), \end{aligned}$$

where B is a bound for b on $\mathbb{R}^d$. Taking $t_0 = \|x-y\|/B$, we obtain

$$I_{\|x-y\|/B,x}(\psi^{(\|x-y\|/B)}) \leq BC\|x-y\|.$$

Therefore, there exists $\bar{\rho}_1 > 0$ such that $\inf_{t\in[0,1]} V(x,y,t) < \delta/2$ for any x and y such that $[x,y] \subset \Gamma_{\bar{\rho}_0/3}$ and $\|x-y\| \leq \bar{\rho}_1$. In view of (5.11), $\rho = \bar{\rho}_1 \wedge (\bar{\rho}_0/3)$ is an appropriate constant in (5.6). $\square$

Proof of Lemma 5.3 Fix $\eta > 0$ and let $\rho > 0$ be small enough to have $B(2\rho) \subset G$ and for Lemma 5.2 to hold with $\delta = \eta/3$ and 2ρ instead of ρ. Note that the definition of $\tilde{I}_{t,x}$ implies the inequality $\inf_{y\in S(2\rho)} V(y,z) \leq V(0,z)$ as soon as $z \notin B(2\rho)$.

Then, by (5.6) and Assumption (Hb), there exists $x \in S(2\rho)$, $z \notin \overline{G}$, $T_1 < \infty$ and $\psi \in \mathcal{C}([0,T_1],\mathbb{R}^d)$ such that $\psi(0) = x$, $\psi(T_1) = z$ and $\tilde{I}_{T_1,x}(\psi) \leq \bar{V} + \eta/3$. Moreover, by removing the beginning of the path ψ until the last time where it hits $S(2\rho)$, we can suppose that for all $t > 0$, $\psi(t) \notin B(2\rho)$.

Thanks to (5.5), for any $y \in S(2\rho)$, there exists a continuous path ψ^y of length $t_y \leq 1$ such that $\psi^y(0) = y$, $\psi^y(t_y) = x$, and $\tilde{I}_{t_y,y}(\psi^y) \leq \eta/3$. Moreover, the construction of this function in the proof of Lemma 5.2 allows us to assume that $\|\psi^y(t)\| = 2\rho$ for all $t \in [0,t_y]$. Let ϕ^y denote the path obtained by concatenating ψ^y and ψ (in that order) and extending the resulting function to be of length $T_0 = T_1 + 1$ by following (5.1) after reaching z. Since the latter path does not contribute to the rate function, we obtain that $\tilde{I}_{T_0,y}(\phi^y) \leq \bar{V} + 2\eta/3$.

Since $z \in \mathbb{R}^d \setminus \overline{G}$, the constant $\Delta := d(z,\partial G)$ is positive. Define

$$O := \bigcup_{y\in S(2\rho)} \left\{ \psi \in \mathcal{C}([0,T_0],\mathbb{R}^d),\ \|\psi - \phi^y\|_{0,T_0} \leq \frac{\Delta \wedge \rho}{2} \right\}.$$

Observe that O is an open subset of $\mathcal{C}([0,T_0],\mathbb{R}^d)$ that contains the functions $\{\phi^y\}_{y\in S(2\rho)}$. Therefore, by Corollary 4.4,

$$\liminf_{\varepsilon\to 0} \varepsilon \ln \inf_{y\in S(2\rho)} \mathbb{P}_y(X^\varepsilon \in O) \geq -\sup_{y\in S(2\rho)} \inf_{\psi\in O} \tilde{I}_{T_0,y}(\psi) \geq -\sup_{y\in S(2\rho)} \tilde{I}_{T_0,y}(\phi^y) > -(\bar{V} + \eta).$$

If $\psi \in O$, then ψ reaches the open ball of radius $\Delta/2$ centered at z before hitting $B(\rho)$, so ψ hits ∂G before hitting $B(\rho)$. Hence, for $X_0^\varepsilon = y \in S(2\rho)$, the event $\{X^\varepsilon \in O\}$ is contained in $\{X^\varepsilon_{\sigma_\rho} \in \partial G\}$, and the proof is completed. $\square$

Acknowledgments: I would like to thank Sylvie Méléard for her constant support and many fruitful discussions about this work. I would also like to thank Régis Ferrière for useful discussions and comments on the biological motivation of this work. Finally, I would like to thank Gérard Ben Arous for having pointed out this problem and for useful comments on a preliminary draft of this work.

References

[1] de Acosta, A. A general non-convex large deviation result with applications to stochastic equations. *Probab. Theory Relat. Fields*, **118**, 483–521 (2000).

[2] Azencott, R. *Grandes déviations et applications*, St Flour Summer school VII, Lecture Notes in Math., Springer Verlag (1980).

[3] Baldi, P. and Chaleyat-Maurel, M. An extension of Ventcel-Freidlin estimates. In: *Lect. Notes Math.*, **1316**, pp. 305–327. Springer (1988).

[4] Boué, M., Dupuis, P. and Ellis, R. S. Large deviations for small noise diffusions with discontinuous statistics. *Probab. Theory Relat. Fields*, **116**, 125–149 (2000).

[5] Champagnat, N. Étude mathématique de modèles stochastiques d'évolution issus de la théorie écologique des dynamiques adaptatives. Ph.D. thesis (in English), University Paris 10 – Nanterre, France (2004).

[6] Champagnat, N. A microscopic interpretation for adaptive dynamics trait substitution sequence models. *Stoch. Proc. Appl.*, **116**(8), 1127–1160 (2006).

[7] Champagnat, N., Ferrière, R. and Ben Arous, G. The canonical equation of adaptive dynamics: A mathematical view. *Selection*, **2**, 71–81 (2001).

[8] Champagnat, N., Ferrière, R. and Méléard, S. Unifying evolutionary dynamics: from individual stochastic processes to macroscopic models via timescale separation. *Theor. Popul. Biol.*, **69**, 297–321 (2006).

[9] Champagnat, N. and Lambert, A. Evolution of discrete populations and the canonical diffusion of adaptive dynamics. *Ann. Appl. Prob.*, **17**, 102–155 (2007).

[10] Champagnat, N. and Méléard, S. Polymorphic evolution sequence and evolutionary branching. Preprint (2008).

[11] Chiang, T.-S. and Sheu, S.-J. Large deviation of diffusion processes with discontinuous drift and their occupation times. *Ann. Probab.*, **28**(1), 140–165 (2000).

[12] Demazure, M. *Bifurcations and catastrophes: Geometry of solutions to nonlinear problems.* Springer-Verlag, Berlin (2000).

[13] Dembo, A. and Zeitouni, O. *Large Deviation Techniques and Applications*, second edition, Jones & Bartlett Publishers (1993).

[14] Dercole, F., Irisson, J.-O. and Rinaldi S. Bifurcation analysis of a prey-predator coevolution model. *SIAM J. Appl. Math.* **63**, 1378–1391 (2003).

[15] Dieckmann, U. and Law, R. The dynamical theory of coevolution: a derivation from stochastic ecological processes. *J. Math. Biol.*, **34**, 579–612 (1996).

[16] Dieckmann, U. and Doebeli, M. On the origin of species by sympatric speciation. *Nature*, **400**, 354–357 (1999).

[17] Diekmann, O. A beginners guide to adaptive dynamics. In: *Mathematical Modelling of Population Dynamics*, Banach Center Publ. **63**, pp. 47–86.

[18] Doss, H. and Priouret, P. Remarques sur les petites perturbations des systèmes dynamiques. In: *Séminaire de Probabilités XVII*, pp. 313–370, Lecture Notes in Mathematics, Springer (1983).

[19] Drossel, B. Biological evolution and statistical physics. *Adv. Physics*, **50**, 209–295 (2001).

[20] Dupuis, P. and Ellis, R. S. *A Weak Convergence Approach to the Theory of Large Deviations.* John Wiley & Sons (1997).

[21] Dupuis, P., Ellis, R. S. and Weiss, A. Large deviations for Markov processes with discontinuous statistics, I: general upper bounds. *Ann. Probab.*, **19**(3), 1280–1297 (1991).

[22] Eldredge, N. and Gould, S. J. Punctuated equilibria: An alternative to phyletic gradualism. In: Schopf, T. J. M. (ed.), *Models in Paleobiology*, Freeman, Cooper and Company, San Francisco, pp. 82–115 (1972).

[23] Ethier, S. N. and Kurtz, T. G. *Markov Processes, characterization and convergence.* John Wiley & Sons, New York (1986).

[24] Freidlin, M. I. and Wentzell, A. D. *Random Perturbations of Dynamical Systems.* Springer-Verlag, Berlin (1984).

[25] Geritz, S.A.H., van der Meijden, E. and Metz, J.A.J. Evolutionary dynamics of seed size and seedling competitive ability. *Theor. Pop. Biol.* **55**, 324–343 (1999).

[26] Hansen, T. F. Stabilizing selection and the comparative analysis of adaptation. *Evolution*, **51**, 1341–1351 (1997).

[27] Hofbauer, J. and Sigmund, R. Adaptive dynamics and evolutionary stability. *Applied Math. Letters*, **3**, 75–79 (1990).

[28] Karatzas, I. and Shreve, S. *Brownian Motion and Stochastic Calculus.* Springer-Verlag, New York (1988).

[29] Kisdi, E. Evolutionary branching under asymmetric competition. *J. Theor. Biol.*, **198**, 149–162 (1999).

[30] Korostelev, A. P. and Leonov, S. L. Action functional for diffusions in discontinuous media. *Probab. Theory Relat. Fields*, **94**, 317–333 (1993).

[31] Lande, R. Natural selection and random genetic drift in phenotypic evolution. *Evolution*, **30**, 314–334 (1976).

[32] Law, R., Marrow, P. and Dieckmann, U. On evolution under asymmetric competition. *Evol. Ecol.*, **11**, 485–501 (1997).

[33] Marrow, P., Dieckmann, U. and Law, R. Evolutionary dynamics of predator-prey systems: an ecological perspective. *J. Math. Biol.*, **34**, 556–578 (1996).

[34] Marrow, P., Law, R. and Cannings, C. The coevolution of predator-prey interactions: ESSs and Red Queen dynamics. *Proc. R. Soc. Lond. B*, **250**, 133–141 (1992).

[35] Metz, J. A. J., Nisbet, R. M. and Geritz, S. A. H. How should we define 'fitness' for general ecological scenarios? *Trends in Ecology and Evolution*, **7**, 198–202 (1992).

[36] Metz, J. A. J., Geritz, S. A. H., Meszéna, G., Jacobs, F. A. J. and van Heerwaarden, J. S. Adaptive Dynamics, a geometrical study of the consequences of nearly faithful reproduction. In: van Strien, S. J. & Verduyn Lunel, S. M. (ed.), *Stochastic and Spatial Structures of Dynamical Systems*, North Holland, Amsterdam, pp. 183–231 (1996).

[37] Rachev, S. T. *Probability Metrics and the Stability of Stochastic Models.* John Wiley & Sons (1991).

[38] Rand, D. A. and Wilson, H. B. Evolutionary catastrophes, punctuated equilibria and gradualism in ecosystem evolution. *Proc. Roy. Soc. London B*, **253**, 137–141 (1993).

[39] Revuz, D. and Yor, M. *Continuous Martingales and Brownian Motion.* Springer-Verlag, Berlin (1991).

[40] Roughgarden, J. A. *Theory of population genetics and evolutionary ecology: an introduction.* Macmillan, New York (1979).

[41] Serre, D. *Matrices: Theory and Applications.* Springer-Verlag (2002).

[42] Wright, S. The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proceedings of the 6th International Congress of Genetics*, **1**, 356–366 (1932).