

# Tails of exit times from unstable equilibria on the line

Yuri Bakhtin, Zsolt Pajor-Gyulai

*Courant Institute of Mathematical Sciences  
New York University  
New York, NY, USA*

January 9, 2020

## Abstract

For a one-dimensional smooth vector field in a neighborhood of an unstable equilibrium, we consider the associated dynamics perturbed by small noise. We give a revealing elementary proof of a result proved earlier using heavy machinery from Malliavin calculus. In particular, we obtain precise vanishing noise asymptotics for the tail of the exit time and for the exit distribution conditioned on atypically long exits.

## 1 Introduction

In a recent paper [BPG19], we studied tails of diffusion exit times from neighborhoods of unstable critical points, in the limit of vanishing noise, in one dimension. The typical exit time  $\kappa_\varepsilon$  in this setting (see the detailed description of the setting below) is of the order of  $\frac{1}{\lambda} \log \frac{1}{\varepsilon}$ , where  $\lambda > 0$  is the local expansion coefficient of the linearization of the system near the critical point, and  $\varepsilon \downarrow 0$  is the noise magnitude. The main result of [BPG19] is that the following polynomial asymptotics holds for a class of initial conditions near the critical point:

$$\mathbb{P} \left( \kappa_\varepsilon > \frac{\alpha}{\lambda} \log \frac{1}{\varepsilon} \right) = c\varepsilon^{\alpha-1}(1 + o(1)), \quad \varepsilon \downarrow 0, \quad (1.1)$$

for  $\alpha > 1$ , with an explicit dependence of the factor  $c$  on the initial condition and the parameters of the model, see (1.7) below for details.

This result is a part of an ongoing effort to understand the long-term properties of multi-dimensional diffusions in the context of noisy heteroclinic networks, including the limiting behavior of invariant distributions associated with such systems in the vanishing noise limit. The typical behavior in such settings is understood for time scales logarithmic in  $\varepsilon^{-1}$ , see [Bak10], [Bak11], [AB11]. To see what happens in the long run though, one has to quantify rare events responsible for transitions that are atypical at the logarithmic time scale. Our work in progress shows that these rare events play a crucial role in the long-term dynamics near noisy heteroclinic networks. Moreover, we argue that they occur exactly due to atypically long stays near unstable critical points. The resulting picture is similar to that of metastability but with polynomial transition rates in place of exponential ones. We give more details on this picture in Section 6 while here we only reiterate that the result of the form (1.1) and its ramifications will be crucial

for that program. However, the technique we used in [BPG19] to analyze densities of auxiliary random variables, was based on heavy tools from Malliavin calculus. That approach somewhat obscures the reason why this result is true and does not seem to be tractable when applied to the study of the analogous exit problem in the neighborhood of a hyperbolic saddle in  $\mathbb{R}^d$ ,  $d > 1$ , i.e. when both attracting and repelling directions are present.

In the present note, our goal is to give a new proof of this result that (a) is based on a more precise description of the dynamics at small scales, (b) uses more elementary tools of stochastic calculus, and (c) has a strong potential to be applicable in higher dimensions. In fact, we prove a slightly more general result on probabilities of the form  $\mathbf{P}(\kappa_\varepsilon > \frac{\alpha}{\lambda} \log \frac{1}{\varepsilon} + t)$ ,  $\alpha > 1$ , for all  $t \in \mathbb{R}$  instead of  $t = 0$  considered in [BPG19]. It turns out that, asymptotically, the dependence on  $t$  is exponential, which implies that for any  $T \in \mathbb{R}$ ,  $\kappa_\varepsilon - \frac{\alpha}{\lambda} \log \frac{1}{\varepsilon} - T$  conditioned on  $\kappa_\varepsilon - \frac{\alpha}{\lambda} \log \frac{1}{\varepsilon} > T$  converges in distribution to an exponential random variable. This phenomenon is a manifestation of loss of memory in the system under conditioning and it is consistent with the fact that  $\kappa_\varepsilon - \frac{1}{\lambda} \log \frac{1}{\varepsilon}$  converges in distribution to a random variable with exponentially decaying right tails, see e.g. [Bak10].

An important ingredient in this note is a conditional equidistribution result (Lemma 3.2) that states that the distribution of the diffusion, conditioned on no exit from a small interval, converges to the uniform distribution. Thus our new approach is closer in the spirit to the one based on quasi-stationary distributions, see [CV16]. However, the existing general theory does not provide answers for us since in our situation both the system and the domain depend on  $\varepsilon$ . Moreover, the time scales we are interested in are too short for the  $t \rightarrow \infty$  limit to be a good approximation while taking  $\varepsilon \downarrow 0$ .

Let us be more precise now. We consider the family of stochastic differential equations

$$dX_\varepsilon(t) = b(X_\varepsilon(t)) dt + \varepsilon \sigma(X_\varepsilon(t)) dW(t), \quad (1.2)$$

on a bounded interval  $\mathcal{I} = [q_-, q_+] \subseteq \mathbb{R}$  with origin in its interior. The drift is given by a vector field  $b \in \mathcal{C}^2(\mathbb{R})$  and the random perturbation is given via a standard Brownian motion  $W$  with respect to a filtration  $(\mathcal{F}_t)_{t \geq 0}$  defined on some probability space  $(\Omega, \mathcal{F}, \mathbf{P})$  under the usual conditions. The noise magnitude is given by a small parameter  $\varepsilon > 0$  in front of the diffusion coefficient  $\sigma$ , which is assumed to be Lipschitz and satisfy  $\sigma(0) > 0$ . Although we are interested only in the evolution within  $\mathcal{I}$ , we can assume that  $b$  and  $\sigma$  are globally Lipschitz without changing the setting.

Standard results on stochastic differential equations (see, e.g., [KS91, Chapter 5]) imply that for any starting location  $X^\varepsilon(0) \in \mathcal{I}$ , the equation (1.2) has a unique strong solution up to the exit time from  $\mathcal{I}$ ,

$$\tau_{\mathcal{I}}^\varepsilon = \inf\{t \geq 0 : X_\varepsilon(t) \in \partial\mathcal{I}\}.$$

Let  $(S^t)_{t \in \mathbb{R}}$  be the flow generated by the vector field  $b$ , i.e.,  $x(t) = S^t x_0$  is the solution of the autonomous ordinary differential equation

$$\dot{x}(t) = b(x(t)), \quad x(0) = x_0 \in \mathbb{R}$$

(we recall that solving this equation for negative times is equivalent to solving  $\dot{y}(t) = -b(y(t))$  for  $y(t) = x(-t)$ ). We assume that there is a unique repelling zero of the vector field  $b$  on  $\mathbb{R}$ ,

which, without loss of generality, we place at the origin. In other words, we assume that  $b(0) = 0$  and, for some  $\lambda > 0$  and  $\eta \in \mathcal{C}^2(\mathcal{I})$ ,

$$b(x) = \lambda x + \eta(x)|x|^2, \quad x \in \mathcal{I}. \quad (1.3)$$

Note that since the origin is the only zero of  $b$  in the closed interval  $\mathcal{I}$ , this assumption implies that for all  $x \neq 0$ , there is a uniquely defined finite time  $T(x)$  such that  $S^{T(x)} \in \partial\mathcal{I}$ .

Under the condition (1.3), the map  $f : \mathcal{I} \rightarrow \mathbb{R}$  defined by

$$f(x) = \lim_{t \rightarrow \infty} e^{\lambda t} S^{-t} x = x - \int_0^\infty e^{\lambda s} \eta(S^{-s} x) |S^{-s} x|^2 ds \quad (1.4)$$

is an order preserving  $\mathcal{C}^2$ -diffeomorphism (see [Eiz84]). In particular,  $f(q_-) < 0 < f(q_+)$ . This map linearizes the flow  $(S^t)$  (see (2.1)) and helps to state the main result concisely, see (1.7).

Under the above assumptions, a version of the following theorem was proved in [BPG19]. In its statement and throughout the paper we use

$$\psi(t, x) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}. \quad (1.5)$$

**Theorem 1.1.** *Consider  $X_\varepsilon$  defined by (1.2) with initial condition  $X_\varepsilon(0) = \varepsilon x$  and let  $K(\varepsilon)$  be any function that satisfies*

$$\lim_{\varepsilon \downarrow 0} \varepsilon^\gamma K(\varepsilon) = 0, \quad \forall \gamma > 0. \quad (1.6)$$

*Then, for all  $\alpha > 1$  and all  $t \in \mathbb{R}$ ,*

$$\lim_{\varepsilon \downarrow 0} \sup_{|x| \leq K(\varepsilon)} \left| \varepsilon^{-(\alpha-1)} \mathbb{P} \left( \tau_{\mathcal{I}}^\varepsilon > \frac{\alpha}{\lambda} \log \varepsilon^{-1} + t; X_\varepsilon(\tau_{\mathcal{I}}^\varepsilon) = q_\pm \right) - e^{-\lambda t} |f(q_\pm)| \psi_0(x) \right| = 0, \quad (1.7)$$

where

$$\psi_0(x) = \psi \left( \frac{\sigma^2(0)}{2\lambda}, x \right) = \sqrt{\frac{\lambda}{\pi}} \frac{e^{-\lambda \left( \frac{x}{\sigma(0)} \right)^2}}{\sigma(0)}.$$

*In particular, for any  $T \in \mathbb{R}$ ,*

$$\begin{aligned} \text{LAW} \left[ \left( \tau_{\mathcal{I}}^\varepsilon - \frac{\alpha}{\lambda} \log \varepsilon^{-1} - T, X_\varepsilon(\tau_{\mathcal{I}}^\varepsilon) \right) \middle| \tau_{\mathcal{I}}^\varepsilon > \frac{\alpha}{\lambda} \log \varepsilon^{-1} + T \right] \\ \Rightarrow \exp_\lambda \otimes \left( \frac{|f(q_-)|}{|f(q_-)| + |f(q_+)|} \delta_{q_-} + \frac{|f(q_+)|}{|f(q_-)| + |f(q_+)|} \delta_{q_+} \right), \end{aligned}$$

where  $\Rightarrow$  stands for weak convergence, and  $\exp_\lambda$  is the exponential distribution with rate  $\lambda > 0$ , i.e.,  $\exp_\lambda[t, \infty) = e^{-\lambda t}$  for  $t \geq 0$ .

**Remark 1.1.** We say that a function satisfying (1.6) *grows at most subpolynomially* at 0. We say that a function  $c(\varepsilon)$  *decays at most subpolynomially* at 0 if  $1/c(\varepsilon)$  grows subpolynomially at 0. For brevity, we will usually omit the reference to 0 and simply say *grows/decays subpolynomially* even when the function might not actually grow.

**Remark 1.2.** The theorem is stated for initial conditions that are at most of the order of  $\varepsilon$  away from the origin up to a subpolynomial factor. The case of initial conditions of the order of  $\varepsilon^\beta$  for  $\beta < 1$  is less interesting since then the tails of exit times decay as stretched exponentials of  $\varepsilon^{-1}$  instead of the power decay given by (1.7) (see Proposition 2.1).

**Remark 1.3.** In the statement of Theorem 1.1 and in the sequel, we adopt the usual convention that each relation involving  $\pm$  and  $\mp$  stands for two relations, one with all top signs and one with all bottom signs.

The brief outline of our approach to the proof of this theorem is as follows. It is convenient to work in coordinates given by the function  $f$  defined in (1.4) where the drift is linear. We study the dynamics of the linear process in two separate phases: (1) in a neighborhood of the critical point of radius  $\varepsilon^\beta$  for  $\beta \in (0, 1)$ ; (2) between leaving this small neighborhood and reaching the boundary of  $f(\mathcal{I})$ .

In the second stage, the drift dominates the noise, and the process closely follows the corresponding deterministic trajectory one obtains by setting  $\varepsilon = 0$ . The outcome of the first stage, i.e., the exit from  $[-\varepsilon^\beta, \varepsilon^\beta]$ , is determined though by a delicate interplay between the noise and the drift in an even smaller neighborhood of the origin ( $\beta$  can be chosen arbitrarily close to one). We study this regime by introducing an auxiliary process  $Z_\varepsilon(t)$  with constant diffusion coefficient approximating  $Y_\varepsilon(t) = f(X_\varepsilon(t))$  pathwise at least over time intervals that are not too large and for which Theorem 1.1 is easier to establish. Since  $Z_\varepsilon(t)$  and  $Y_\varepsilon(t)$  do not, in general, stay close on longer timescales, we introduce an iterative scheme to tackle this problem. Namely, we split the longer time interval into shorter ones and show that a useful approximation result, which holds under conditioning on the process not having exited the spatial interval, can be applied sequentially.

The plan of the paper is as follows. In Section 2, we perform the aforementioned change of variables to linearize the drift and prove Theorem 1.1 using an intermediate result on the exit from a small neighborhood of the origin. In Section 3, we introduce an auxiliary process, which is fully linear and thus allows us to derive certain properties of the exit problem through explicit calculations. In Section 4, we prove an approximation result which allows us to transfer these properties from the fully linear process to the case where only the drift is linear as long as the timescales involved are not too large. Finally, in Section 5, we use an iterative scheme to lift this limitation thereby finishing the proof of the intermediate result. In Section 6, we explain how the result of this paper fits our program on long-term behavior of diffusions near heteroclinic networks.

**Acknowledgment.** Yuri Bakhtin gratefully acknowledges partial support from NSF via grant DMS-1811444.

## 2 Proof of Theorem 1.1

As outlined above, we study the system first in a small neighborhood of the origin and then after the process has escaped this small neighborhood.

Let us start with the first part. The diffeomorphism  $f : \mathcal{I} \rightarrow \mathbb{R}$  introduced in (1.4) and its inverse  $g = f^{-1}$  provide a conjugation between the flow  $(S^t)$  and a linear flow:

$$f(S^t x) = e^{\lambda t} f(x), \quad \text{or} \quad f'(x)b(x) = \lambda f(x). \quad (2.1)$$

Note that the integrand in (1.4) is quadratic when  $x$  is close to zero and thus we have  $f(0) = 0$  and  $f'(0) = 1$ . Outside of  $\mathcal{I}$ , we define  $f$  so that  $f'$  and  $f''$  are bounded.

Let  $Y_\varepsilon(t) = f(X_\varepsilon(t))$  for times prior to the escape from  $\mathcal{I}$ . Itô's formula and (2.1) then imply that this process satisfies the stochastic differential equation

$$dY_\varepsilon(t) = \lambda Y_\varepsilon(t)dt + \varepsilon \tilde{\sigma}(Y_\varepsilon(t))dW(t) + \frac{\varepsilon^2}{2}h(Y_\varepsilon(t))dt \quad (2.2)$$

for  $t < \tau_{\mathcal{I}}^\varepsilon$ , where  $\tilde{\sigma}(y) = f'(g(y))\sigma(g(y))$  and  $h(y) = f''(g(y))\sigma^2(g(y))$ . Due to boundedness of  $f'$  and  $f''$ ,  $\tilde{\sigma}$  and  $h$  are also bounded.

By Duhamel's formula,  $Y_\varepsilon$  satisfies the integral equation

$$Y_\varepsilon(t) = e^{\lambda t} \left( Y_\varepsilon(0) + \varepsilon U_\varepsilon(t) + \varepsilon^2 V_\varepsilon(t) \right), \quad (2.3)$$

where

$$U_\varepsilon(t) = \int_0^t e^{-\lambda s} \tilde{\sigma}(Y_\varepsilon(s))dW(s), \quad V_\varepsilon(t) = \frac{1}{2} \int_0^t e^{-\lambda s} h(Y_\varepsilon(s))ds.$$

Due to our conventions on  $f', f''$  outside of  $\mathcal{I}$ , the processes  $U_\varepsilon(t)$  and  $V_\varepsilon(t)$  are defined for all  $t \geq 0$ . Moreover, the boundedness of  $h$  immediately implies the boundedness of  $V_\varepsilon(t)$ :

$$\|V_\varepsilon(\cdot)\|_\infty \leq \frac{\|h\|_\infty}{2\lambda}, \quad (2.4)$$

where  $\|\cdot\|_\infty$  is the sup-norm on  $[0, \infty)$ . The boundedness of  $\tilde{\sigma}$  yields a similar conclusion about the quadratic variation of  $U_\varepsilon(t)$ . Hence, the existence of constants  $c_1, c_2, N_0 > 0$  such that

$$\mathbf{P} \left( \sup_{t \geq 0} |U_\varepsilon(t \wedge \tau_{\mathcal{I}}^\varepsilon)| \geq N \right) \leq c_1 e^{-c_2 N^2}, \quad N \geq N_0, \quad (2.5)$$

is implied by the following exponential martingale inequality (see, e.g., Problem 12.10 in [Bas11]):

**Lemma 2.1.** *Let  $M(t)$  be a centered martingale with quadratic variation process  $\langle M \rangle_t$ . Then*

$$\mathbf{P} \left( \sup_{t \geq 0} |M(t)| \geq a; \langle M \rangle_\infty \leq b \right) \leq 2e^{-\frac{a^2}{2b}}.$$

Let us take  $\beta \in (0, 1)$  and set  $\mathcal{V} = g([- \varepsilon^\beta, \varepsilon^\beta]) \subseteq \mathcal{I}$ . The following result describes the tail behavior of  $\tau_{\mathcal{V}}^\varepsilon$ , the exit time from  $\mathcal{V}$ . In particular, it says that, in the  $\varepsilon \downarrow 0$  asymptotics, the choice of the exit direction is distributed symmetrically independently of the exit time.

**Theorem 2.1.** *Let  $Y_\varepsilon(0) = \varepsilon y$ , where  $|y| \leq K(\varepsilon)$  with  $K(\varepsilon)$  growing subpolynomially at 0. Then, for all  $\alpha > 1$ ,  $C \in \mathbb{R}$ , and any function  $c(\varepsilon)$  satisfying  $\lim_{\varepsilon \rightarrow 0} c(\varepsilon) = 0$ , there is  $\beta_0 \in (0, 1)$  such that for  $\beta \in (\beta_0, 1)$ , we have*

$$\lim_{\varepsilon \downarrow 0} \sup_{|y| \leq K(\varepsilon)} \left| \varepsilon^{-(\alpha-1)} \mathbf{P} \left( \tau_{\mathcal{V}}^\varepsilon > \frac{\alpha - \beta}{\lambda} \log \varepsilon^{-1} - C + c(\varepsilon); Y_\varepsilon(\tau_{\mathcal{V}}^\varepsilon) = \pm \varepsilon^\beta \right) - e^{\lambda C} \psi_0(y) \right| = 0. \quad (2.6)$$

We give the proof of Theorem 2.1 in Section 5.

After exit from  $\mathcal{V}$ , the deterministic dynamics dominates the evolution, which means that the exit time will be close to

$$T_\varepsilon^\pm := T\left(g\left(\pm\varepsilon^\beta\right)\right) = \frac{\beta}{\lambda} \log \varepsilon^{-1} + \frac{1}{\lambda} \log |f(q_\pm)|, \quad (2.7)$$

the time it takes for  $X_0(t)$  to exit  $\mathcal{I}$  starting at  $g(\pm\varepsilon^\beta)$ . This is captured by the following standard large deviation estimates.

**Proposition 2.1.** *Let  $X_\varepsilon(0) = g(\pm\varepsilon^\beta)$ . Then for every  $\beta' \in (0, \beta)$  and subpolynomially decaying function  $c(\varepsilon) > 0$ , we have*

$$\mathbb{P}\left(|\tau_\mathcal{I}^\varepsilon - T_\varepsilon^\pm| > c(\varepsilon)\right) = o\left(e^{-\frac{1}{\varepsilon^{2(1-\beta')}}}\right) \quad (2.8)$$

and

$$\mathbb{P}\left(X_\varepsilon(\tau_\mathcal{I}^\varepsilon) = q_\pm\right) \geq 1 - o\left(e^{-\frac{1}{\varepsilon^{2(1-\beta')}}}\right) \quad (2.9)$$

PROOF: We start by showing that with overwhelming probability the exit happens through the endpoint that is on the same side as the starting point. Indeed, (2.3), (2.4), and (2.5) imply

$$\begin{aligned} \mathbb{P}\left(X_\varepsilon(\tau_\mathcal{I}^\varepsilon) = q_+\right) &= \mathbb{P}\left(Y_\varepsilon(\tau_\mathcal{I}^\varepsilon) = f(q_+)\right) = \mathbb{P}\left(\varepsilon^\beta < |\varepsilon U_\varepsilon(\tau_\mathcal{I}^\varepsilon) + \varepsilon^2 V_\varepsilon(\tau_\mathcal{I}^\varepsilon)|\right) \\ &\leq \mathbb{P}\left(\varepsilon^\beta < \varepsilon \sup_{t>0} |U_\varepsilon(t \wedge \tau_\mathcal{I}^\varepsilon)| + \varepsilon^2 \frac{\|h\|_\infty}{2\lambda}\right) \\ &\leq \mathbb{P}\left(\varepsilon^{-(1-\beta)} - \varepsilon \frac{\|h\|_\infty}{2\lambda} < \sup_{t>0} |U_\varepsilon(t \wedge \tau_\mathcal{I}^\varepsilon)|\right) = o\left(e^{-\frac{1}{\varepsilon^{2(1-\beta')}}}\right), \end{aligned}$$

and (2.9) follows.

To prove (2.8), let us introduce

$$U_\varepsilon = U_\varepsilon\left((T_\varepsilon^\pm + c(\varepsilon)) \wedge \tau_\mathcal{I}^\varepsilon\right), \quad V_\varepsilon = V_\varepsilon\left((T_\varepsilon^\pm + c(\varepsilon)) \wedge \tau_\mathcal{I}^\varepsilon\right),$$

and note that (2.3) implies

$$\begin{aligned} \mathbb{P}\left(\tau_\mathcal{I}^\varepsilon > T_\varepsilon^\pm + c(\varepsilon)\right) &\leq \mathbb{P}\left(f(q_-) < e^{\lambda(T_\varepsilon^\pm + c(\varepsilon))} \left(\pm\varepsilon^\beta + \varepsilon U_\varepsilon + \varepsilon^2 V_\varepsilon\right) < f(q_+)\right) \\ &= \mathbb{P}\left(\varepsilon^\beta e^{-\lambda c(\varepsilon)} \frac{f(q_-)}{|f(q_\pm)|} < \pm\varepsilon^\beta + \varepsilon U_\varepsilon(t) + \varepsilon^2 V_\varepsilon(t) < \varepsilon^\beta e^{-\lambda c(\varepsilon)} \frac{f(q_+)}{|f(q_\pm)|}\right) \\ &\leq \mathbb{P}\left(|U_\varepsilon| > \varepsilon^{-(1-\beta)} \left(1 - e^{-\lambda c(\varepsilon)}\right) - \varepsilon \frac{\|h\|_\infty}{2\lambda}\right) + o\left(e^{-\frac{1}{\varepsilon^{2(1-\beta')}}}\right) \\ &= o\left(e^{-\frac{1}{\varepsilon^{2(1-\beta')}}}\right), \end{aligned}$$

where we used (2.7), (2.9), (2.4), (2.5), and the subpolynomial decay of  $c(\varepsilon)$ . Similarly,

$$\begin{aligned}
\mathbb{P}(\tau_{\mathcal{I}}^{\varepsilon} < T_{\varepsilon}^{\pm} - c(\varepsilon)) &\leq \mathbb{P}\left(\sup_{t < T_{\varepsilon}^{\pm} - c(\varepsilon)} e^{\lambda(t - c(\varepsilon))} \left| \pm \varepsilon^{\beta} + \varepsilon U_{\varepsilon}(t) + \varepsilon^2 V_{\varepsilon}(t) \right| > |f(q_{\pm})| \right) + o\left(e^{-\frac{1}{\varepsilon^{2(1-\beta')}}}\right) \\
&\leq \mathbb{P}\left(e^{\lambda(T_{\varepsilon}^{\pm} - c(\varepsilon))} \left( \varepsilon^{\beta} + \varepsilon \sup_{t < T_{\varepsilon}^{\pm}} |U_{\varepsilon}(t)| + \varepsilon^2 \frac{\|h\|_{\infty}}{2\lambda} \right) > |f(q_{\pm})| \right) + o\left(e^{-\frac{1}{\varepsilon^{2(1-\beta')}}}\right) \\
&= \mathbb{P}\left(\varepsilon^{\beta} + \varepsilon \sup_{t < T_{\varepsilon}^{\pm}} |U_{\varepsilon}(t)| + \varepsilon^2 \frac{\|h\|_{\infty}}{2\lambda} \geq \varepsilon^{\beta} e^{\lambda c(\varepsilon)} \right) + o\left(e^{-\frac{1}{\varepsilon^{2(1-\beta')}}}\right) \\
&\leq \mathbb{P}\left(\sup_{t > 0} |U_{\varepsilon}(t \wedge \tau_{\mathcal{I}}^{\varepsilon})| \geq \varepsilon^{-(1-\beta)} \left( e^{\lambda c(\varepsilon)} - 1 \right) - \varepsilon \frac{\|h\|_{\infty}}{2\lambda} \right) + o\left(e^{-\frac{1}{\varepsilon^{2(1-\beta')}}}\right) \\
&= o\left(e^{-\frac{1}{\varepsilon^{2(1-\beta')}}}\right).
\end{aligned}$$

□

*Proof of Theorem 1.1.* We have

$$Y_{\varepsilon}(0) = f(X_{\varepsilon}(0)) = f(\varepsilon x) = \varepsilon x + \mathcal{O}(\varepsilon^2 x^2) = \varepsilon(x + \mathcal{O}(\varepsilon x^2))$$

and

$$|x + \mathcal{O}(\varepsilon x^2)| \leq K(\varepsilon) + \mathcal{O}(\varepsilon(K(\varepsilon))^2) \leq 2K(\varepsilon)$$

for small enough  $\varepsilon$ , so the right-hand side is a function subpolynomially growing at 0. Let us define  $\theta_{\varepsilon}^{\pm} = T_{\varepsilon}^{\pm} - (\tau_{\mathcal{I}}^{\varepsilon} - \tau_{\mathcal{V}}^{\varepsilon})$ . Due to (2.9),

$$\mathbb{P}\left(X_{\varepsilon}(\tau_{\mathcal{I}}^{\varepsilon}) = q_{\pm}; \tau_{\mathcal{I}}^{\varepsilon} > \frac{\alpha}{\lambda} \log \varepsilon^{-1} + t\right) = \mathbb{P}\left(Y_{\varepsilon}(\tau_{\mathcal{V}}^{\varepsilon}) = \varepsilon^{\beta}; \tau_{\mathcal{V}}^{\varepsilon} > \frac{\alpha}{\lambda} \log \varepsilon^{-1} + t - T_{\varepsilon}^{\pm} + \theta_{\varepsilon}^{\pm}\right) + o(\varepsilon^{\alpha-1}),$$

where the error term (along with all subsequent error terms) is uniform in the starting points  $|x| \leq K(\varepsilon)$ . Note that the strong Markov property implies the conditional independence of  $\theta_{\varepsilon}^{\pm}$  and  $\tau_{\mathcal{V}}^{\varepsilon}$  given  $Y_{\varepsilon}(\tau_{\mathcal{V}}^{\varepsilon}) = \varepsilon^{\beta}$ . This, along with (2.8), allows us to give upper and lower estimates of the first term on the right-hand side:

$$\begin{aligned}
&\mathbb{P}\left(Y_{\varepsilon}(\tau_{\mathcal{V}}^{\varepsilon}) = \pm \varepsilon^{\beta}; \tau_{\mathcal{V}}^{\varepsilon} > \frac{\alpha}{\lambda} \log \varepsilon^{-1} + t - T_{\varepsilon}^{\pm} + c(\varepsilon)\right) + o(\varepsilon^{\alpha-1}) \\
&\leq \mathbb{P}\left(Y_{\varepsilon}(\tau_{\mathcal{V}}^{\varepsilon}) = \varepsilon^{\beta}; \tau_{\mathcal{V}}^{\varepsilon} > \frac{\alpha}{\lambda} \log \varepsilon^{-1} + t - T_{\varepsilon}^{\pm} + \theta_{\varepsilon}^{\pm}\right) \\
&\leq \mathbb{P}\left(Y_{\varepsilon}(\tau_{\mathcal{V}}^{\varepsilon}) = \pm \varepsilon^{\beta}; \tau_{\mathcal{V}}^{\varepsilon} > \frac{\alpha}{\lambda} \log \varepsilon^{-1} + t - T_{\varepsilon}^{\pm} - c(\varepsilon)\right) + o(\varepsilon^{\alpha-1}).
\end{aligned}$$

where  $c(\varepsilon)$  is an arbitrary positive function that decays subpolynomially. Now we may apply Theorem 2.1 with  $C = -t + \lambda^{-1} \log |f(q_{\pm})|$  to both sides and conclude the proof. □

### 3 Linear system with additive noise

In this section, we introduce an auxiliary process, which is a simpler special case of (2.2). Namely, we consider

$$dZ_{\varepsilon}(t) = \lambda Z_{\varepsilon}(t)dt + \varepsilon \sigma_0 dW(t), \quad Z_{\varepsilon}(0) = \varepsilon z, \quad (3.1)$$

where  $\sigma_0 = \tilde{\sigma}(0) = \sigma(0) > 0$ ,  $|z| < K(\varepsilon)$ , and  $K(\varepsilon)$  grows subpolynomially at 0.

We will need a precise description of the exit of  $Z_\varepsilon$  from  $\mathcal{V}_Z = [-\varepsilon^\beta(1 + \delta_\varepsilon), \varepsilon^\beta(1 + \delta_\varepsilon)]$  for  $\beta \in (0, 1)$  and any  $\delta_\varepsilon > 0$  satisfying  $\delta_\varepsilon \downarrow 0$  as  $\varepsilon \downarrow 0$ . Let us introduce a stopping time

$$\tau_\varepsilon = \inf\{t > 0 : |Z_\varepsilon(t)| = \varepsilon^\beta(1 + \delta_\varepsilon)\},$$

and, for a constant  $C$  and a subpolynomially decaying at 0 function  $c(\varepsilon)$ , a deterministic time

$$t_\varepsilon = \frac{\alpha - \beta}{\lambda} \log \varepsilon^{-1} - C + c(\varepsilon). \quad (3.2)$$

We will often use  $C_\varepsilon = C - c(\varepsilon)$ . The first result of this section is a version of Theorem 2.1 for the process  $Z_\varepsilon$  with stronger control of the dependence on the initial point.

**Lemma 3.1.** *For any  $\alpha > 1$  and any subpolynomially decaying function  $c(\varepsilon)$  in the definition of  $t_\varepsilon$ , there is  $c > 0$  such that*

$$\lim_{\varepsilon \downarrow 0} \sup_{|z| \leq K(\varepsilon)} e^{cz^2} \left| \varepsilon^{-(\alpha-1)} \mathbf{P}(\tau_\varepsilon > t_\varepsilon) - 2e^{\lambda C} \psi_0(z) \right| = 0.$$

PROOF: Duhamel's formula gives an explicit solution to (3.1):

$$Z_\varepsilon(t) = \varepsilon e^{\lambda t} (z + \sigma_0 N(t)), \quad (3.3)$$

where  $N(t) = \int_0^t e^{-\lambda s} dW(s)$ . Plugging in  $\tau_\varepsilon$  for  $t$ , we obtain

$$(1 + \delta_\varepsilon) \varepsilon^\beta = |Z_\varepsilon(\tau_\varepsilon)| = \varepsilon e^{\lambda \tau_\varepsilon} |z + \sigma_0 N(\tau_\varepsilon)|, \quad (3.4)$$

which is equivalent to

$$\tau_\varepsilon = \frac{1 - \beta}{\lambda} \log \varepsilon^{-1} - \frac{1}{\lambda} \log |z + \sigma_0 N(\tau_\varepsilon)| + o(1).$$

Therefore,

$$\{\tau_\varepsilon > t_\varepsilon\} = \left\{ |z + \sigma_0 N(\tau_\varepsilon)| < e^{\lambda C_\varepsilon} \varepsilon^{\alpha-1} (1 + o(1)) \right\}.$$

By the martingale convergence theorem,  $N(t)$  converges to a random variable  $N_\infty$  as  $t \rightarrow \infty$  almost surely and in  $L^1$ . We claim that, in addition, there are  $c_1, c_2 > 0$  such that for any  $L > 0$ ,

$$\mathbf{P} \left( |N(\tau_\varepsilon) - N_\infty| > L \varepsilon^{\alpha-\beta}; \tau_\varepsilon > t_\varepsilon \right) \leq c_1 e^{-c_2 L^2}. \quad (3.5)$$

Indeed, Lemma 2.1 implies

$$\mathbf{P} \left( \sup_{t \geq t_\varepsilon} |N(t) - N_\infty| \geq L \varepsilon^{\alpha-\beta} \right) \leq 2 \mathbf{P} \left( \sup_{t \geq t_\varepsilon} |N(t) - N(t_\varepsilon)| \geq L \varepsilon^{(\alpha-\beta)}/2 \right) \leq c_1 e^{-c_2 L^2 \varepsilon^{2(\alpha-\beta)} e^{2\lambda t_\varepsilon}}$$

for some  $c_1, c_2 > 0$ , which proves the claim since  $\varepsilon^{\alpha-\beta} e^{\lambda t_\varepsilon} = e^{-\lambda C_\varepsilon}$ .

Let us fix  $\gamma \in (\alpha - 1, \alpha - \beta)$  and write

$$\mathbf{P}(\tau_\varepsilon > t_\varepsilon) = H_1(z, \varepsilon) + H_2(z, \varepsilon),$$



where

$$\begin{aligned} H_1(z, \varepsilon) &= \mathbb{P} \left( |z + \sigma_0 N_\infty + \sigma_0 (N(\tau_\varepsilon) - N_\infty)| < e^{\lambda C_\varepsilon} \varepsilon^{\alpha-1} (1 + o(1)); |N(\tau_\varepsilon) - N_\infty| \leq \varepsilon^\gamma \right), \\ H_2(z, \varepsilon) &= \mathbb{P} \left( |z + \sigma_0 N(\tau_\varepsilon)| < e^{\lambda C_\varepsilon} \varepsilon^{\alpha-1} (1 + o(1)); |N(\tau_\varepsilon) - N_\infty| > \varepsilon^\gamma \right). \end{aligned}$$

The random variables  $N(\tau_\varepsilon)$  and  $N(\tau_\varepsilon) - N_\infty$  are independent due to the strong Markov property. So the Gaussian tail of the maximum of the Brownian motion and (3.5) imply

$$H_2(z, \varepsilon) \leq c_1 e^{-c_2 z^2} o(\varepsilon^{\alpha-1}).$$

The desired asymptotics of  $H_1(z, \varepsilon)$  follows from the explicit form of the Gaussian density of the random variable  $z + \sigma_0 N_\infty$ .  $\square$

The next result is based on the fact that the distribution of  $Z_\varepsilon(t_\varepsilon)$  conditioned on non-exit is approximately uniform over  $[-\varepsilon^\beta(1 + \delta_\varepsilon), \varepsilon^\beta(1 + \delta_\varepsilon)]$ . In fact, this stronger statement is proved as an intermediate step. We first note that the density of any absolutely continuous random variable conditioned on a positive probability event is well-defined.

**Lemma 3.2.** *If  $f_\varepsilon^c(u)$  is the probability density of  $Z_\varepsilon(t_\varepsilon)$  conditioned on  $\{\tau_\varepsilon > t_\varepsilon\}$ , then for any  $\delta > 0$*

$$\lim_{\varepsilon \downarrow 0} \sup_{|u| \leq (1-\delta)\varepsilon^\beta} \sup_{|z| \leq K(\varepsilon)} \left| \varepsilon^\beta f_\varepsilon^c(u) - \frac{1}{2} \right| = 0. \quad (3.6)$$

Moreover, for any integrable function  $h$  with exponentially decaying tails and any subpolynomially growing function  $K(\varepsilon)$ ,

$$\lim_{\varepsilon \downarrow 0} \sup_{|z| \leq K(\varepsilon)} \left| \varepsilon^{-(1-\beta)} \mathbb{E} [h(\varepsilon^{-1} Z_\varepsilon(t_\varepsilon)) | \tau_\varepsilon > t_\varepsilon] - \frac{1}{2} \int_{-\infty}^{\infty} h(y) dy \right| = 0. \quad (3.7)$$

PROOF: By (3.3) and the Dambis–Dubins–Schwartz theorem (see, e.g., [KS91, Section 3.4B]),

$$Z_\varepsilon(t) = \varepsilon e^{\lambda t} \left( z + \sigma_0 \int_0^t e^{-\lambda s} dW(s) \right) = \varepsilon e^{\lambda t} (z + B(r(t))), \quad (3.8)$$

where  $r(t) = \sigma_0^2 (1 - e^{-2\lambda t}) / (2\lambda)$ , and  $B$  is an auxiliary standard Brownian motion. Since

$$r(t_\varepsilon) = \sigma_0^2 \frac{1 - \varepsilon^{2(\alpha-\beta)} e^{2\lambda C_\varepsilon}}{2\lambda},$$

we have

$$Z_\varepsilon(t_\varepsilon) = \varepsilon^{1-\alpha+\beta} e^{-\lambda C_\varepsilon} \left( z + B \left( \sigma_0^2 \frac{1 - \varepsilon^{2(\alpha-\beta)} e^{2\lambda C_\varepsilon}}{2\lambda} \right) \right).$$

This is a Gaussian random variable. Its density at a point  $u \in \mathbb{R}$  is given by

$$p_\varepsilon(u) = \frac{\sqrt{\lambda} e^{\lambda C_\varepsilon}}{\sqrt{\pi} \sigma_0 \varepsilon^{1-\alpha+\beta} (1 + o(1))} \exp \left\{ - \frac{(u \varepsilon^{\alpha-\beta-1} e^{\lambda C_\varepsilon} - z)^2}{\frac{\sigma_0^2}{\lambda} (1 + o(1))} \right\}.$$

If  $|u| \leq \varepsilon^\beta$ , then  $|u\varepsilon^{\alpha-1-\beta}| \leq \varepsilon^{\alpha-1}$ , so

$$p_\varepsilon(u) = \varepsilon^{\alpha-\beta-1} e^{\lambda C} \psi_0(z)(1 + o(1)), \quad (3.9)$$

uniformly over  $u$  and  $z$  satisfying  $|u| \leq \varepsilon^\beta$ ,  $|z| \leq K(\varepsilon)$ .

Therefore, (3.6) will follow from

$$\lim_{\varepsilon \downarrow 0} \sup_{|u| \leq (1-\delta)\varepsilon^\beta} \sup_{|z| \leq K(\varepsilon)} \left| \frac{f_\varepsilon(z, u)}{\varepsilon^{\alpha-\beta-1} e^{\lambda C} \psi_0(z)} - 1 \right| = 0, \quad (3.10)$$

where  $f_\varepsilon(z, u)$  is the sub-probability density of  $Z_\varepsilon(t_\varepsilon)$  on the event  $\{\tau_\varepsilon > t_\varepsilon\}$ . For this, it suffices to see that

$$\sup_{|u| \leq (1-\delta)\varepsilon^\beta} \sup_{|z| \leq K(\varepsilon)} \frac{g_\varepsilon(z, u)}{\psi_0(z)}, \quad (3.11)$$

decays exponentially fast as  $\varepsilon \downarrow 0$ , where  $g_\varepsilon(z, u)$  is the sub-probability density of  $Z_\varepsilon(t_\varepsilon)$  on the event  $\{\tau_\varepsilon \leq t_\varepsilon\}$ . Given that  $Z_\varepsilon(\tau_\varepsilon) = \varepsilon^\beta(1 + \delta_\varepsilon)$  (similarly for  $-\varepsilon^\beta(1 + \delta_\varepsilon)$ ), we have

$$Z_\varepsilon(t_\varepsilon) = e^{\lambda(t_\varepsilon - \tau_\varepsilon)} \left( \varepsilon^\beta(1 + \delta_\varepsilon) + \varepsilon\sigma_0 \int_0^{t_\varepsilon - \tau_\varepsilon} e^{-\lambda s} dW(s) \right).$$

and thus

$$g_\varepsilon(z, u) = \int_0^{t_\varepsilon} \mathbb{P}\{\tau_\varepsilon \in [t, t + dt)\} G(\varepsilon, t, u), \quad (3.12)$$

where

$$G(\varepsilon, t, u) = \psi \left( \varepsilon^2 \sigma_0^2 \frac{w^2 - 1}{2\lambda}, w\varepsilon^\beta(1 + \delta_\varepsilon) - u \right) = \frac{1}{\sqrt{2\pi\varepsilon^2 \sigma_0^2 \frac{w^2 - 1}{2\lambda}}} e^{-\frac{\lambda(w\varepsilon^\beta(1 + \delta_\varepsilon) - u)^2}{\varepsilon^2 \sigma_0^2 (w^2 - 1)}},$$

with  $\psi(\cdot, \cdot)$  that was introduced in (1.5) and  $w = e^{\lambda(t_\varepsilon - t)}$ .

We claim that there is  $c > 0$  such that

$$\sup\{G(\varepsilon, t, u) : 0 \leq t \leq t_\varepsilon, |u| \leq (1 - \delta)\varepsilon^\beta\} = \mathcal{O}\left(e^{-c/\varepsilon^{2(1-\beta)}}\right). \quad (3.13)$$

This, along with (3.12) and the fact that  $\psi_0(z) > \lambda^{1/2}\pi^{-1/2}\sigma_0^{-1}e^{-\lambda K^2(\varepsilon)/\sigma_0^2}$  for all  $\varepsilon$  and  $z$  satisfying  $|z| < K(\varepsilon)$ , will imply the desired exponential decay in (3.11).

Let us fix any  $w_0$  and find  $c_0 > 0$  such that

$$\lambda(w\varepsilon^\beta(1 + \delta_\varepsilon) - (1 - \delta)\varepsilon^\beta)^2 / (\sigma_0^2(w^2 - 1)) \geq c_0\varepsilon^{2\beta}$$

for  $w > w_0$  and all  $\varepsilon > 0$ . Then there is a constant  $c_0$  such that for all  $|u| \leq (1 - \delta)\varepsilon^\beta$  and  $\varepsilon > 0$ ,

$$G(\varepsilon, t, u) \leq \frac{1}{\sqrt{2\pi\varepsilon^2 \sigma_0^2 \frac{w_0^2 - 1}{2\lambda}}} e^{-c_0/\varepsilon^{2(1-\beta)}}, \quad w > w_0. \quad (3.14)$$

If  $1 \leq w \leq w_0$ , then  $(w\varepsilon^\beta(1 + \delta_\varepsilon) - u)^2 \geq \Delta := \varepsilon^{2\beta}\delta^2$ . So, denoting  $D = \varepsilon^2\sigma_0^2(w^2 - 1)/(2\lambda)$ , we obtain

$$G(\varepsilon, t, u) \leq \frac{1}{\sqrt{2\pi D}} e^{-\frac{\Delta}{2D}} = \psi(D, \Delta), \quad 1 \leq w \leq w_0. \quad (3.15)$$

The restriction on  $w$  implies  $0 \leq D \leq c_1 \varepsilon^2$  for some  $c_1$ . To maximize  $\psi(D, \Delta)$ , we compute

$$\partial_D \ln \psi(D, \Delta) = -\frac{1}{2D} + \frac{\Delta}{2D^2},$$

so  $\psi(D, \Delta)$  grows in  $D \in [0, \Delta]$ , and we obtain from (3.15):

$$G(\varepsilon, t, x) \leq \frac{1}{\sqrt{2\pi c_1 \varepsilon^2}} e^{-\frac{\Delta}{2c_1 \varepsilon^2}}. \quad (3.16)$$

Combining (3.14) and (3.16), we obtain (3.13) and hence (3.11). Thus, (3.6) is proved.

To prove (3.7), we write

$$\mathbb{E} [h(\varepsilon^{-1} Z_\varepsilon(t_\varepsilon)) | \tau_\varepsilon > t_\varepsilon] = \int_{|u| < (1-\delta)\varepsilon^\beta} h(\varepsilon^{-1}u) f_\varepsilon^c(u) du + \int_{(1-\delta)\varepsilon^\beta \leq |u| \leq (1+\delta_\varepsilon)\varepsilon^\beta} h(\varepsilon^{-1}u) f_\varepsilon^c(u) du$$

and notice that the first term on the right-hand side equals

$$\begin{aligned} \int_{-(1-\delta)\varepsilon^\beta}^{(1-\delta)\varepsilon^\beta} h(\varepsilon^{-1}x) f_\varepsilon^c(x) dx &= \frac{1}{2\varepsilon^\beta} (1 + o(1)) \int_{-(1-\delta)\varepsilon^\beta}^{(1-\delta)\varepsilon^\beta} h(\varepsilon^{-1}x) dx \\ &= \frac{1}{2} \varepsilon^{1-\beta} (1 + o(1)) \int_{-(1-\delta)\varepsilon^{-(1-\beta)}}^{(1-\delta)\varepsilon^{-(1-\beta)}} h(x) dx = \frac{1}{2} \varepsilon^{1-\beta} \int_{-\infty}^{\infty} h(x) dx + o(\varepsilon^{1-\beta}), \end{aligned}$$

while the second term decays much faster than  $\varepsilon^{1-\beta}$  due to the decay assumption on  $h$ .  $\square$

#### 4 Proof of Theorem 2.1 for $\alpha \in (1, 1 + \beta)$

We start by studying the deviations  $\Delta_\varepsilon(t) = Y_\varepsilon(t) - Z_\varepsilon(t)$  as long as both processes  $Y_\varepsilon(t)$  and  $Z_\varepsilon(t)$  are close to the origin. Let us fix  $\beta \in (0, 1)$  and introduce the stopping times

$$\bar{\tau}_\varepsilon = \inf\{t \geq 0 : |Y_\varepsilon(t)| = \varepsilon^\beta\}, \quad \bar{\bar{\tau}}_\varepsilon = \inf\{t \geq 0 : |Y_\varepsilon(t)| = 2\varepsilon^\beta\}.$$

**Lemma 4.1.** *Suppose  $\alpha \in (1, 1 + \beta)$ ,  $\beta' \in (\alpha - 1, \beta)$ ,  $L(\varepsilon) > 0$  is a bounded function, and*

$$t'_\varepsilon = \frac{\alpha - \beta}{\lambda} \log(L(\varepsilon)\varepsilon^{-1}). \quad (4.1)$$

*Then for sufficiently small  $\varepsilon > 0$ , we have*

$$\mathbb{P} \left( \sup_{0 \leq t \leq t'_\varepsilon \wedge \bar{\bar{\tau}}_\varepsilon} |\Delta_\varepsilon(t)| > \varepsilon^{\beta+\beta'-(\alpha-1)} \right) \leq 2e^{-\frac{c}{\varepsilon^{2(\beta-\beta')}}}, \quad (4.2)$$

*provided  $Y_\varepsilon(0) = Z_\varepsilon(0) = \varepsilon y$ , for every  $y \in (-2\varepsilon^\beta, 2\varepsilon^\beta)$ .*

PROOF: Recalling (2.3), we obtain

$$\Delta_\varepsilon(t) = \varepsilon e^{\lambda t} \left( I_\varepsilon^{(1)}(t) + I_\varepsilon^{(2)}(t) \right), \quad I_\varepsilon^{(1)}(t) = \int_0^t e^{-\lambda s} (\tilde{\sigma}(Y_\varepsilon(s)) - \sigma_0) dW(s), \quad I_\varepsilon^{(2)} = \mathcal{O}(\varepsilon).$$

Clearly,  $I_\varepsilon^{(1)}$  is a martingale satisfying  $\langle I_\varepsilon^{(1)} \rangle_t = \mathcal{O}(\varepsilon^{2\beta})$  for  $t \leq \bar{\tau}_\varepsilon$ , so for any  $c_0 > 0$ , there is  $c > 0$  such that

$$\mathbf{P} \left( \sup_{s \leq \bar{\tau}_\varepsilon} |I_\varepsilon^{(1)}| > c_0 \varepsilon^{\beta'} \right) \leq 2e^{-c \frac{\varepsilon^{2\beta'}}{\varepsilon^{2\beta}}} \leq 2e^{-c \frac{1}{\varepsilon^{2(\beta-\beta')}}},$$

by Lemma 2.1. Therefore

$$\mathbf{P} \left( \sup_{s \leq \bar{\tau}_\varepsilon} |I_\varepsilon^{(1)} + I_\varepsilon^{(2)}| > 2c_0 \varepsilon^{\beta'} \right) \leq 2e^{-c \frac{1}{\varepsilon^{2(\beta-\beta')}}},$$

and on the complementary event we have

$$\sup_{0 \leq t \leq t'_\varepsilon \wedge \bar{\tau}_\varepsilon} |\Delta_\varepsilon(t)| < 2c_0 \varepsilon L^{\alpha-\beta}(\varepsilon) \varepsilon^{-(\alpha-\beta)} \varepsilon^{\beta'} < \varepsilon^{\beta+\beta'-(\alpha-1)}$$

if  $c_0$  is chosen sufficiently small, which finishes the proof.  $\square$

Based on this approximation result and the calculation for  $Z_\varepsilon$  in the previous section, the following theorem proves Theorem 2.1 for  $\alpha$  not too large.

**Theorem 4.1.** *Let  $\alpha \in (1, 1 + \beta)$  and let  $t_\varepsilon$  be as in (3.2). Then there is  $c > 0$  such that*

$$\lim_{\varepsilon \downarrow 0} \sup_{|y| \leq K(\varepsilon)} e^{cy^2} \left| \varepsilon^{-(\alpha-1)} \mathbf{P} \left( \bar{\tau}_\varepsilon > t_\varepsilon; Y_\varepsilon(\bar{\tau}_\varepsilon) = \pm \varepsilon^\beta \right) - e^{\lambda C} \psi_0(y) \right| = 0. \quad (4.3)$$

PROOF: We start with an upper bound on  $\mathbf{P}(\bar{\tau}_\varepsilon > t_\varepsilon)$  in terms of  $Z_\varepsilon$ . Let us fix any  $\beta'' \in (\beta, 2\beta - (\alpha - 1))$ , so that  $\beta' = \beta'' - \beta + \alpha - 1 \in (\alpha - 1, \beta)$ , which will allow us to apply Lemma 4.1 several times in this proof. Let us take any family of events  $(B_\varepsilon)_{\varepsilon > 0}$  and estimate

$$\begin{aligned} \mathbf{P}(\bar{\tau}_\varepsilon > t_\varepsilon; B_\varepsilon) &= I_1(\varepsilon) + I_2(\varepsilon) \\ &= \mathbf{P} \left( \bar{\tau}_\varepsilon > t_\varepsilon; \sup_{0 \leq t \leq t_\varepsilon \wedge \bar{\tau}_\varepsilon} |\Delta_\varepsilon(t)| < \varepsilon^{\beta''}; B_\varepsilon \right) + \mathbf{P} \left( \bar{\tau}_\varepsilon > t_\varepsilon; \sup_{0 \leq t \leq t_\varepsilon \wedge \bar{\tau}_\varepsilon} |\Delta_\varepsilon(t)| \geq \varepsilon^{\beta''}; B_\varepsilon \right). \end{aligned}$$

Note that  $t_\varepsilon$  is of the form (4.1) with  $L(\varepsilon) = e^{-\frac{(\alpha-\beta)C_\varepsilon}{\lambda}}$  and thus (4.2) implies

$$I_2(\varepsilon) = o_{\text{exp}}(1), \quad (4.4)$$

where, for any  $\gamma > 0$ , we use  $o_{\text{exp}}(1)$  as a shorthand for  $o(e^{-\varepsilon^{-\gamma}})$ . Also

$$I_1(\varepsilon) = \mathbf{P} \left( \bar{\tau}_\varepsilon > t_\varepsilon; \sup_{0 \leq t \leq t_\varepsilon} |\Delta_\varepsilon(t)| < \varepsilon^{\beta''}; B_\varepsilon \right).$$

We need to approximate this in terms of the exit time of  $Z_\varepsilon$  instead of  $\bar{\tau}_\varepsilon$ . We do not have control over the difference of these two times in general as we can only control the difference of the processes until  $t_\varepsilon$ . Instead, we are going to set a different threshold for  $Z_\varepsilon$  to reach. Let  $\gamma \in (\beta, \beta'')$ ,  $l_\varepsilon^1 = \varepsilon^\beta$  and  $l_\varepsilon^2 = l_\varepsilon^1 + \varepsilon^\gamma$ . This implies

$$\mathbf{P} \left( \bar{\tau}_\varepsilon > t_\varepsilon; \tau_\varepsilon^Z \leq t_\varepsilon; \sup_{0 \leq t \leq t_\varepsilon} |\Delta_\varepsilon(t)| \leq \varepsilon^{\beta''}; B_\varepsilon \right) = 0,$$

where  $\tau_\varepsilon^Z$  is the exit time from  $[-l_\varepsilon^2, l_\varepsilon^2]$  for  $Z_\varepsilon$ , and thus

$$I_1(\varepsilon) = \mathbf{P} \left( \bar{\tau}_\varepsilon > t_\varepsilon; \tau_\varepsilon^Z > t_\varepsilon; \sup_{0 \leq t \leq t_\varepsilon} |\Delta_\varepsilon(t)| \leq \varepsilon^{\beta''}; B_\varepsilon \right) \leq \mathbf{P} (\tau_\varepsilon^Z > t_\varepsilon; B_\varepsilon).$$

Combining this with (4.4), we obtain

$$\mathbf{P} (\bar{\tau}_\varepsilon > t_\varepsilon; B_\varepsilon) \leq \mathbf{P} (\tau_\varepsilon^Z > t_\varepsilon; B_\varepsilon) + o_{\text{exp}}(1). \quad (4.5)$$

Next, we set  $l_\varepsilon^3 = l_\varepsilon^1 - \varepsilon^\gamma$  and define  $\eta_\varepsilon^Z$  to be the exit time of  $Z_\varepsilon$  from  $[-l_\varepsilon^3, l_\varepsilon^3]$ . Lemma 4.1 and the fact that  $\{\bar{\tau}_\varepsilon \leq t_\varepsilon; \sup_{0 \leq t \leq t_\varepsilon \wedge \bar{\tau}_\varepsilon} |\Delta_\varepsilon(t)| \leq \varepsilon^{\beta''}\} \subset \{\eta_\varepsilon^Z \leq t_\varepsilon\}$ , imply

$$\begin{aligned} \mathbf{P}(\bar{\tau}_\varepsilon > t_\varepsilon; B_\varepsilon) &\geq \mathbf{P} \left( \bar{\tau}_\varepsilon > t_\varepsilon; \sup_{0 \leq t \leq t_\varepsilon \wedge \bar{\tau}_\varepsilon} |\Delta_\varepsilon(t)| \leq \varepsilon^{\beta''}; B_\varepsilon \right) \\ &= \mathbf{P}(B_\varepsilon) + o_{\text{exp}}(1) - \mathbf{P} \left( \bar{\tau}_\varepsilon \leq t_\varepsilon; \sup_{0 \leq t \leq t_\varepsilon \wedge \bar{\tau}_\varepsilon} |\Delta_\varepsilon(t)| \leq \varepsilon^{\beta''}; B_\varepsilon \right) \\ &\geq \mathbf{P}(B_\varepsilon) + o_{\text{exp}}(1) - \mathbf{P}(\eta_\varepsilon^Z \leq t_\varepsilon; B_\varepsilon) = \mathbf{P}(\eta_\varepsilon^Z > t_\varepsilon; B_\varepsilon) + o_{\text{exp}}(1). \end{aligned} \quad (4.6)$$

Combining (4.5) and (4.6), we obtain

$$\mathbf{P}(\eta_\varepsilon^Z > t_\varepsilon; B_\varepsilon) - o_{\text{exp}}(1) \leq \mathbf{P}(\bar{\tau}_\varepsilon > t_\varepsilon; B_\varepsilon) \leq \mathbf{P}(\tau_\varepsilon^Z > t_\varepsilon; B_\varepsilon) + o_{\text{exp}}(1). \quad (4.7)$$

This, Lemma 3.1, and our choice of  $l_\varepsilon^i$ ,  $i = 1, 2, 3$  imply

$$\lim_{\varepsilon \downarrow 0} \sup_{|y| \leq K(\varepsilon)} e^{cy^2} \left| \varepsilon^{-(\alpha-1)} \mathbf{P}(\bar{\tau}_\varepsilon > t_\varepsilon) - 2e^{\lambda C} \psi_0(y) \right| = 0. \quad (4.8)$$

To finish the proof, we will need

$$\lim_{\varepsilon \downarrow 0} \sup_{\varepsilon^{\beta''} \leq |u| \leq \varepsilon^\beta} \left| \mathbf{P} \left( Y_\varepsilon(\bar{\tau}_\varepsilon) = \varepsilon^\beta | Y_\varepsilon(t_\varepsilon) = u \right) - \mathbf{1}_{u>0} \right| = 0, \quad (4.9)$$

which holds since, due to (2.3),  $Y_\varepsilon(\bar{\tau}_\varepsilon) = \varepsilon^\beta$  is equivalent to  $\varepsilon^{-1}u + U_\varepsilon(\bar{\tau}_\varepsilon) + \varepsilon V_\varepsilon(\bar{\tau}_\varepsilon) > 0$ , and so

$$\left| \mathbf{P} \left( Y_\varepsilon(\bar{\tau}_\varepsilon) = \varepsilon^\beta | Y_\varepsilon(t_\varepsilon) = u \right) - \mathbf{1}_{u>0} \right| \leq \mathbf{P} \left( |U_\varepsilon(\bar{\tau}_\varepsilon)| \geq \varepsilon^{-(1-\beta'')} - \varepsilon V(\bar{\tau}_\varepsilon) \right) \rightarrow 0, \quad \varepsilon \downarrow 0,$$

due to the boundedness of  $V(\bar{\tau}_\varepsilon)$  and (2.5).

Using (4.9), we can write

$$\begin{aligned} \mathbf{P} \left( Y(\bar{\tau}_\varepsilon) = \varepsilon^\beta | \bar{\tau}_\varepsilon > t_\varepsilon \right) &= \mathbf{E} \left[ \mathbf{P} \left( Y_\varepsilon(\bar{\tau}_\varepsilon) = \varepsilon^\beta | Y_\varepsilon(t_\varepsilon) \right) | \bar{\tau}_\varepsilon > t_\varepsilon \right] \\ &= \mathbf{E} \left[ \mathbf{P} \left( Y_\varepsilon(\bar{\tau}_\varepsilon) = \varepsilon^\beta | Y_\varepsilon(t_\varepsilon) \right); |Y_\varepsilon(t_\varepsilon)| < \varepsilon^{\beta''} | \bar{\tau}_\varepsilon > t_\varepsilon \right] + \mathbf{P} \left( Y_\varepsilon(t_\varepsilon) > \varepsilon^{\beta''} | \bar{\tau}_\varepsilon > t_\varepsilon \right) + o(1) \\ &= A_1 + A_2 + o(1). \end{aligned} \quad (4.10)$$

Due to (4.2),

$$\begin{aligned} A_1 &= \mathbf{P} \left( |Y_\varepsilon(t_\varepsilon)| < \varepsilon^{\beta''}; \sup_{0 \leq t \leq t_\varepsilon} |\Delta_\varepsilon(t)| \leq \varepsilon^{\beta''} | \bar{\tau}_\varepsilon > t_\varepsilon \right) + o(1) \\ &\leq \mathbf{P} \left( |Z_\varepsilon(t_\varepsilon)| < 2\varepsilon^{\beta''} | \bar{\tau}_\varepsilon > t_\varepsilon \right) + o(1) \rightarrow 0. \end{aligned} \quad (4.11)$$

In the last convergence, we used (3.9) to compute

$$\mathbf{P}(|Z_\varepsilon(t_\varepsilon)| < 2\varepsilon^{\beta''}) = 2\varepsilon^{\beta''} \varepsilon^{1-\alpha+\beta} e^{\lambda C} \psi_0(y)(1 + o(1)),$$

and we used (4.7) with  $B_\varepsilon \equiv \Omega$ , along with Lemma 3.1 to compute  $\mathbf{P}(\bar{\tau}_\varepsilon > t_\varepsilon) = 2e^{\lambda C} \psi_0(y) \varepsilon^{\alpha-1}(1 + o(1))$ , so  $\mathbf{P}(|Z_\varepsilon(t_\varepsilon)| < 2\varepsilon^{\beta''})/\mathbf{P}(\bar{\tau}_\varepsilon > t_\varepsilon) \rightarrow 0$  follows from our assumptions on  $\alpha, \beta, \beta''$ .

Also due to (4.2),

$$\mathbf{P}\left(Z_\varepsilon(t_\varepsilon) > 2\varepsilon^{\beta''} | \bar{\tau}_\varepsilon > t_\varepsilon\right) + o(1) \leq A_2 \leq \mathbf{P}(Z_\varepsilon(t_\varepsilon) > 0 | \bar{\tau}_\varepsilon > t_\varepsilon) + o(1).$$

Using (4.7) with  $B_\varepsilon = \{Z_\varepsilon(t_\varepsilon) > 2\varepsilon^{\beta''}\}$ ,  $B_\varepsilon = \{Z_\varepsilon(t_\varepsilon) > 0\}$ , and  $B_\varepsilon = \Omega$ , we can switch conditioning to that in terms of  $Z_\varepsilon$ :

$$\mathbf{P}\left(Z_\varepsilon(t_\varepsilon) > 2\varepsilon^{\beta''} | \eta_\varepsilon^Z > t_\varepsilon\right) + o(1) \leq A_2 \leq \mathbf{P}(Z_\varepsilon(t_\varepsilon) > 0 | \tau_\varepsilon^Z > t_\varepsilon) + o(1),$$

where both the left and the right hand side converge to  $1/2$  as  $\varepsilon \downarrow 0$  due to Lemma 3.2. Combining this with (4.10), (4.11), noticing that all the  $o(1)$  terms in these estimates are independent of the starting point  $y$ , and using (4.8) we obtain (4.3), which completes the proof.  $\square$

## 5 Extension to arbitrary timescales

The goal of this section is to extend Theorem 4.1 for arbitrary  $\alpha > 1$  and thus prove Theorem 2.1. We set  $\theta = \alpha - \beta$ ,  $L(\varepsilon) = e^{-\frac{\lambda}{\theta} C_\varepsilon}$ , and

$$t_\varepsilon = \frac{\theta}{\lambda} \log(L(\varepsilon)\varepsilon^{-1}) = \frac{\theta}{\lambda} \log \varepsilon^{-1} - C_\varepsilon.$$

When  $\theta \in (1 - \beta, 1)$ , Theorem 4.1 applies and there is nothing new to prove. Here we study the case  $\theta \geq 1$ . Up to this point the only restriction on  $\beta$  was  $\beta \in (0, 1)$ . Let us now set  $N = [\theta] + 1 \geq 2$ ,  $\beta_0 = \frac{1}{2} \left(1 + \frac{\theta}{N}\right)$  and assume  $\beta \in (\beta_0, 1)$  throughout this section. We have

$$\theta < N < \frac{\theta}{1 - \beta}. \quad (5.1)$$

We also define  $t'_\varepsilon = t_\varepsilon/N$  and  $t_{\varepsilon,k} = kt'_\varepsilon$ ,  $k = 0, 1, \dots, N$ . Our plan is to track  $Y_{\varepsilon,k} = Y_\varepsilon(t_{\varepsilon,k})$ ,  $k = 0, 1, \dots, N$ , using the results of the previous section on the short intervals  $[t_{\varepsilon,k}, t_{\varepsilon,k+1}]$ .

The first step is the following lemma which establishes that the process needs to stay close to the origin to delay the exit.

**Lemma 5.1.** *We have*

$$\max_{k=0, \dots, N-1} \mathbf{P}\left(\sup_{t \leq t_{\varepsilon,k}} |Y_\varepsilon(t)| > \varepsilon K(\varepsilon); \bar{\tau}_\varepsilon > t_{\varepsilon,k+1}\right) \leq C_1 e^{-C_2 K^2(\varepsilon)} \quad (5.2)$$

for some  $C_1, C_2 > 0$ . In particular,

$$\max_{k=0, \dots, N-1} \mathbf{P}\left(\max_{u=0, \dots, k} |Y_{\varepsilon,u}| > \varepsilon K(\varepsilon); \bar{\tau}_\varepsilon > t_{\varepsilon,k+1}\right) \leq C_1 e^{-C_2 K^2(\varepsilon)}.$$

PROOF: Using the strong Markov property and applying Duhamel's principle (2.3) and (2.4) to the initial condition  $y$  with  $|y| > \varepsilon K(\varepsilon)$ , we reduce the lemma to the estimate

$$\begin{aligned} & \mathbf{P} \left( e^{\lambda t'_\varepsilon} \inf_{t \leq t'_\varepsilon} |y + \varepsilon U_\varepsilon(t) + \varepsilon^2 V_\varepsilon(t)| < \varepsilon^\beta \right) \\ & \leq \mathbf{P} \left( \sup_{t \leq t'_\varepsilon} |U_\varepsilon(t)| > K(\varepsilon) - \varepsilon^{\beta + \frac{\theta}{N} - 1 - \frac{\theta}{N} \log(L(\varepsilon))} - \varepsilon \frac{\|h\|_\infty}{2\lambda} \right), \end{aligned}$$

and the desired inequality follows by (2.5) since  $\beta + \frac{\theta}{N} - 1 > 0$  due to (5.1).  $\square$

We now collect some results needed for our iteration scheme.

**Lemma 5.2.** *Let  $Y_\varepsilon(0) = \varepsilon y$ . Then there is  $c > 0$  such that*

$$\lim_{\varepsilon \downarrow 0} \sup_{|y| \leq K(\varepsilon)} e^{cy^2} \left| \varepsilon^{-(\frac{\theta}{N} + \beta - 1)} \mathbf{P} \left( Y_\varepsilon(\bar{\tau}_\varepsilon) = \pm \varepsilon^\beta; \bar{\tau}_\varepsilon > t'_\varepsilon \right) - e^{\frac{\lambda C}{N}} \psi_0(y) \right| = 0. \quad (5.3)$$

Moreover, for any Lipschitz function  $h$  on  $\mathbb{R}$ , exponentially decaying at  $\infty$ , we have

$$\lim_{\varepsilon \downarrow 0} \sup_{|y| \leq K(\varepsilon)} e^{cy^2} \left| \varepsilon^{-\frac{\theta}{N}} \mathbf{E} [h(\varepsilon^{-1} Y_{\varepsilon,1}); \bar{\tau}_\varepsilon > t'_\varepsilon] - e^{\frac{\lambda C}{N}} \psi_0(y) \int_{-\infty}^{\infty} h(y) dy \right| = 0. \quad (5.4)$$

PROOF: The first claim follows from Theorem 4.1 with  $C/N$  in place of  $C$  and  $\alpha = \beta + \theta/N$ . Note that this value of  $\alpha$  belongs to  $(1, 1 + \beta)$  due to (5.1).

The second claim is a direct consequence of the first one and

$$\lim_{\varepsilon \downarrow 0} \sup_{|y| \leq K(\varepsilon)} \left| \varepsilon^{-(1-\beta)} \mathbf{E} [h(\varepsilon^{-1} Y_{\varepsilon,1}) | \bar{\tau}_\varepsilon > t'_\varepsilon] - \frac{1}{2} \int_{-\infty}^{\infty} h(y) dy \right| = 0. \quad (5.5)$$

Once again, to prove this, we would like to use the result for the linear process. However, the estimate (4.2) is insufficient when applied directly. Instead, let us note that

$$\varepsilon^{-1} Y_{\varepsilon,1} - \varepsilon^{-1} Z_{\varepsilon,1} = e^{\lambda t'_\varepsilon} \int_0^{t'_\varepsilon} e^{-\lambda t} (\sigma(Y_\varepsilon(t)) - \sigma_0) dW(t) = \varepsilon^{-\theta/N} I_\varepsilon(t'_\varepsilon),$$

where  $I_\varepsilon(t'_\varepsilon) = e^{-\lambda C_\varepsilon/N} \int_0^{t'_\varepsilon} e^{-\lambda t} (\sigma(Y_\varepsilon(t)) - \sigma_0) dW(t)$  and choose any

$$\beta' \in \left( \frac{1}{2} \left( 1 + \frac{\theta}{2} \right), \beta \right).$$

The exponential martingale inequality (Lemma 2.1) and the Lipschitz continuity of  $\sigma$  imply

$$\mathbf{P} \left( \sup_{s \leq \bar{\tau}_\varepsilon} |I_\varepsilon| > \varepsilon^{\beta'} \right) = o_{\exp}(1),$$

and thus

$$\mathbf{P} \left( \varepsilon^{-(1-\beta)} |\varepsilon^{-1} Y_{\varepsilon,1} - \varepsilon^{-1} X_{\varepsilon,1}| > \varepsilon^{\beta' - \theta/N - 1 + \beta} | \bar{\tau}_\varepsilon > t'_\varepsilon \right) = o_{\exp}(1).$$

Note that  $\beta' - \theta/N - 1 + \beta > 0$  due to the choice of  $\beta$  and  $\beta'$ . Now we can use the last display and the Lipschitz continuity of  $h$  to obtain

$$\varepsilon^{-(1-\beta)} \mathbf{E} \left[ |h(\varepsilon^{-1} Y_{\varepsilon,1}) - h(\varepsilon^{-1} Z_{\varepsilon,1})| \mid \bar{\tau}_\varepsilon > t'_\varepsilon \right] \leq \|h\|_{Lip} \varepsilon^{\beta' - \theta/N - 1 + \beta} + o_{\exp}(1) \rightarrow 0$$

as  $\varepsilon \downarrow 0$ . This and (3.7) imply (5.5), which completes the proof of the lemma.  $\square$

Finally, the next Theorem implies Theorem 2.1 with  $\beta_0 = \frac{1}{2} \left(1 + \frac{\theta}{N}\right)$ .

**Theorem 5.1.** *There is  $c > 0$  such that*

$$\lim_{\varepsilon \downarrow 0} \sup_{|y| \leq K(\varepsilon)} e^{cy^2} \left| \varepsilon^{-(\alpha-1)} \mathbf{P}(\bar{\tau}_\varepsilon > t_\varepsilon) - 2e^{\lambda C} \psi_0(y) \right| = 0, \quad (5.6)$$

and

$$\lim_{\varepsilon \downarrow 0} \sup_{|y| \leq K(\varepsilon)} \left| \mathbf{P} \left( Y_\varepsilon(\bar{\tau}_\varepsilon) = \pm \varepsilon^\beta \mid \bar{\tau}_\varepsilon > t_\varepsilon \right) - \frac{1}{2} \right| = 0. \quad (5.7)$$

PROOF: We will use  $o_K(1)$  to denote any function that decays faster than any power of  $\varepsilon$  as  $\varepsilon \downarrow 0$ . It suffices to prove the theorem in the case where the function  $K(\varepsilon)$  grows fast enough as  $\varepsilon \rightarrow 0$  to guarantee that the right-hand side of (5.2) is  $o_K(1)$ . To see that the theorem will then follow in full generality, we just notice that enlarging the set of initial conditions  $y$  from  $\{|y| \leq K(\varepsilon)\}$  to  $\{|y| \leq K(\varepsilon) \vee |\log \varepsilon|\}$  reduces the situation to that special case.

We will prove by induction that for every  $k = 1, \dots, N$ , there is  $c > 0$  such that

$$\lim_{\varepsilon \downarrow 0} \sup_{|y| \leq K(\varepsilon)} e^{cy^2} \left| \varepsilon^{-(k\frac{\theta}{N} + \beta - 1)} \mathbf{P}_y(\bar{\tau}_\varepsilon > t_{\varepsilon,k}) - 2e^{\frac{\lambda k C}{N}} \psi_0(y) \right| = 0, \quad (5.8)$$

where we explicitly indicate the dependence on the starting point  $Y_\varepsilon(0) = \varepsilon y$  as a subscript in  $\mathbf{P}_y$  for clarity. The case  $k = N$  is the desired result (5.6). The base of induction, the case  $k = 1$ , is the first claim of Lemma 5.2. Let us make the induction step assuming that (5.8) holds for some  $k$ .

Lemma 5.1 and the Markov Property allows us to write

$$\begin{aligned} \mathbf{P}_y(\bar{\tau}_\varepsilon > t_{\varepsilon,k+1}) &= \mathbf{P}_y(\bar{\tau}_\varepsilon > t_{\varepsilon,k+1}; |Y_{\varepsilon,1}| \leq \varepsilon K(\varepsilon)) + o_K(1) \\ &= \int_{-K(\varepsilon)}^{K(\varepsilon)} \mathbf{P}_{y'}(\bar{\tau}_\varepsilon > t_{\varepsilon,k}) \mathbf{P}_y(Y_{\varepsilon,1} \in \varepsilon dy'; \bar{\tau}_\varepsilon > t'_\varepsilon) + o_K(1) \end{aligned} \quad (5.9)$$

Using the induction hypothesis (5.8), we obtain

$$\mathbf{P}_{y'}(\bar{\tau}_\varepsilon > t_{\varepsilon,k}) = 2\varepsilon^{k\frac{\theta}{N} + \beta - 1} e^{\frac{\lambda k C}{N}} \psi_0(y') + e^{-cy'^2} o\left(\varepsilon^{k\frac{\theta}{N} + \beta - 1}\right),$$

where the error term is uniform over  $|y'| \leq K(\varepsilon)$ . This means that the first term on the right hand side of (5.9) can be written as

$$\begin{aligned} &\int_{-K(\varepsilon)}^{K(\varepsilon)} \mathbf{P}_y(\bar{\tau}_\varepsilon > t_{\varepsilon,k+1} | Y_{\varepsilon,1} = \varepsilon y') \mathbf{P}_y(Y_{\varepsilon,1} \in \varepsilon dy'; \bar{\tau}_\varepsilon > t'_\varepsilon) \\ &= 2\varepsilon^{k\frac{\theta}{N} + \beta - 1} e^{\frac{\lambda k C}{N}} \mathbf{E}_y[\psi_0(\varepsilon^{-1} Y_{\varepsilon,1}); |Y_{\varepsilon,1}| \leq \varepsilon K(\varepsilon); \bar{\tau}_\varepsilon > t'_\varepsilon] + H(\varepsilon, y), \end{aligned} \quad (5.10)$$



where the error term satisfies

$$\sup_{|y| \leq K(\varepsilon)} |H(\varepsilon, y)| \leq \sup_{|y| \leq K(\varepsilon)} \mathbb{E}_y \left( e^{-c(\varepsilon^{-1}Y_{\varepsilon,1})^2}; \bar{\tau}_\varepsilon > t'_\varepsilon \right) \cdot o \left( \varepsilon^{k\frac{\theta}{N} + \beta - 1} \right) = o \left( \varepsilon^{(k+1)\frac{\theta}{N} + \beta - 1} \right) \quad (5.11)$$

where we used (5.4) in the last step with  $h(y) = e^{-cy^2}$ . The main term on the right-hand side of (5.10) can be estimated using Lemma 5.1 and (5.4) with  $h(y) = \psi_0(y)$  (so  $\int_{\mathbb{R}} h(y) dy = 1$ ):

$$\begin{aligned} \mathbb{E}_y [\psi_0(\varepsilon^{-1}Y_{\varepsilon,1}); |Y_{\varepsilon,1}| \leq \varepsilon K(\varepsilon); \bar{\tau}_\varepsilon > t'_\varepsilon] &= \mathbb{E}_y [\psi_0(\varepsilon^{-1}Y_{\varepsilon,1}); \bar{\tau}_\varepsilon > t'_\varepsilon] + o_K(1) \\ &= \varepsilon^{\frac{\theta}{N}} e^{\frac{\lambda C}{N}} \psi_0(y) + o \left( \varepsilon^{\frac{\theta}{N}} \right), \end{aligned} \quad (5.12)$$

and this expansion holds uniformly over  $|y| \leq K(\varepsilon)$ .

Putting together (5.9), (5.10), (5.11), and (5.12) yields, for sufficiently small  $c' > 0$ ,

$$\lim_{\varepsilon \downarrow 0} \sup_{|y| \leq K(\varepsilon)} e^{c'y^2} \left| \varepsilon^{-((k+1)\frac{\theta}{N} + \beta - 1)} \mathbb{P}_y(\bar{\tau}_\varepsilon > t_{\varepsilon, k+1}) - 2e^{\frac{\lambda(k+1)C}{N}} \psi_0(y) \right| = 0.$$

This completes the induction step and finishes the proof of (5.6).

To prove (5.7), note first that (5.2) and the strong Markov property implies

$$\begin{aligned} \mathbb{P} \left( Y_\varepsilon(\bar{\tau}_\varepsilon) = \pm \varepsilon^\beta \middle| \bar{\tau}_\varepsilon > t_\varepsilon \right) &= \\ &= \int_{-K(\varepsilon)}^{K(\varepsilon)} \mathbb{P}_y \left( Y_\varepsilon(\bar{\tau}_\varepsilon) = \pm \varepsilon^\beta \middle| \bar{\tau}_\varepsilon > t_{\varepsilon,1} \right) \mathbb{P}(Y_{\varepsilon, N-1} \in \varepsilon dy | \bar{\tau}_\varepsilon > t_{\varepsilon, N-1}) + o(1). \end{aligned}$$

Using (5.3), the integrand can be written as

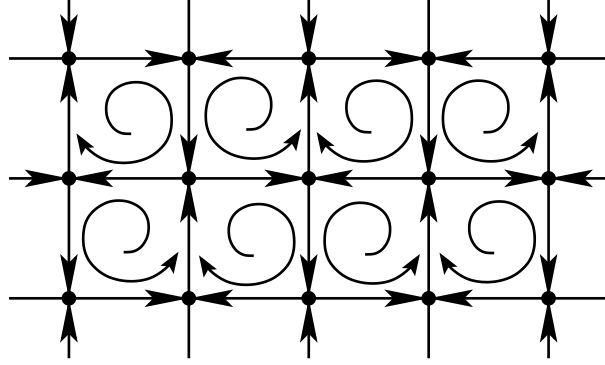
$$\mathbb{P}_y \left( Y_\varepsilon(\bar{\tau}_\varepsilon) = \pm \varepsilon^\beta \middle| \bar{\tau}_\varepsilon > t_{\varepsilon,1} \right) = \frac{\varepsilon^{\frac{\theta}{N} + \beta - 1} e^{\lambda C} \psi_0(y) + o \left( \varepsilon^{\frac{\theta}{N} + \beta - 1} \right)}{2\varepsilon^{\frac{\theta}{N} + \beta - 1} e^{\lambda C} \psi_0(y) + o \left( \varepsilon^{\frac{\theta}{N} + \beta - 1} \right)} = \frac{1}{2} + o(1),$$

where the error terms are uniform in  $y$  and thus another application of (5.2) finishes the proof of (5.7).  $\square$

## 6 Rare transitions in heteroclinic networks

In this section, we discuss, briefly and nonrigorously, the questions that lead us to study the tails of exit times in detail. These questions originate in the long-term behavior of diffusions near heteroclinic networks in the vanishing noise limit. A heteroclinic network is a feature of the phase portrait associated with a vector field composed of multiple hyperbolic critical points (“saddles”) connected to each other by heteroclinic orbits, see an example of a phase portrait with a heteroclinic network for a cellular flow on Figure 6.

Let us consider a diffusion process solving the Itô equation (1.2) in  $\mathbb{R}^2$  (although with minor modifications the discussion below applies to higher dimensions as well) with drift  $b$  giving rise to a heteroclinic network. The typical behavior of such processes for times that are of the order of  $\log \varepsilon^{-1}$  was studied in [Bak11], [Bak10], [AB11]. Its main features depend mainly on the linearization of the drift  $b$  near the saddle points and can be described as follows. Upon



**Figure 1.** A heteroclinic network is the backbone of this phase portrait associated with a cellular flow.

reaching a small neighborhood of a saddle, the process spends a long (logarithmic in  $\varepsilon$ ) time in that neighborhood where the drift is weak and eventually exits along the unstable manifold associated with the positive eigenvalue  $\lambda$  of the linearization. This manifold is composed of two outgoing heteroclinic orbits, so the dynamics chooses one of them and follows it for a time of the order of 1 until it reaches the next saddle where it will also eventually decide between two outgoing directions, etc. This description may seem to imply the picture where the limiting (as  $\varepsilon \rightarrow 0$ ) process is essentially a random walk on the directed graph of heteroclinic connections. However, the character of the limiting process is often not Markovian and depends on the linearizations of  $b$  near the saddles.

To see what is going on, let us consider a 2-dimensional diffusion  $(X_\varepsilon, Y_\varepsilon)$  near a model saddle described by equations

$$\begin{aligned} dX_\varepsilon(t) &= \lambda X_\varepsilon(t)dt + \varepsilon dW(t), \\ dY_\varepsilon(t) &= -\mu Y_\varepsilon dt + \varepsilon dB(t), \end{aligned}$$

driven by independent standard Wiener processes  $W$  and  $B$ . Here  $\lambda > 0$  and  $\mu > 0$  can be viewed as coefficients of expansion and contraction, respectively. Assuming that  $(X_\varepsilon(0), Y_\varepsilon(0)) = (0, 1)$ , i.e., starting the process on the stable manifold of the saddle located at the origin, we are interested in the distribution of  $(X_\varepsilon(\tau_\varepsilon), Y_\varepsilon(\tau_\varepsilon))$ , where  $\tau_\varepsilon = \inf\{t \geq 0 : |X_\varepsilon(t)| = 1\}$  is the exit time from the strip  $[-1, 1] \times \mathbb{R}$ . Since

$$\begin{aligned} X_\varepsilon(t) &= \varepsilon e^{\lambda t} N(t), \\ Y_\varepsilon(t) &= e^{-\mu t} + \varepsilon M(t), \end{aligned}$$

where

$$N(t) = \int_0^t e^{-\lambda s} dW(s) \xrightarrow{d} N(\infty), \quad M(t) = \int_0^t e^{-\mu(t-s)} dB(s) \xrightarrow{d} M(\infty), \quad \text{as } t \rightarrow \infty,$$

we find that for small  $\varepsilon$ ,

$$\tau_\varepsilon \stackrel{d}{\approx} \frac{1}{\lambda} \log \frac{1}{\varepsilon} + \frac{1}{\lambda} \log \frac{1}{|N(\infty)|}, \quad (6.1)$$

and

$$Y_\varepsilon(\tau_\varepsilon) \stackrel{d}{\approx} \varepsilon^\rho |N(\infty)|^\rho + \varepsilon M(\infty), \quad (6.2)$$

where  $\rho = \mu/\lambda$ . This allows us to conclude that as  $\varepsilon \rightarrow 0$ , the distribution of the exit point  $Y_\varepsilon(\tau_\varepsilon)$  depends crucially on how  $\rho$  compares to 1. In particular, if  $\rho < 1$  (i.e., the contraction is not as strong as expansion:  $\mu < \lambda$ ) then the first term  $\varepsilon^\rho |N(\infty)|^\rho$  dominates. It is positive and scales as  $\varepsilon^\rho \gg \varepsilon$ , i.e., it is stronger than the noise magnitude  $\varepsilon$ . Therefore, in this situation, at the next saddle point the system is most likely to stay on the same side of the heteroclinic network. If  $\rho > 1$ , then the probability of choosing either of the two outgoing connections at the next saddle approaches  $1/2$  as  $\varepsilon \rightarrow 0$ .

This analysis can be extended to more general initial conditions, to nonlinear drift and diffusion coefficients, to higher dimensions, and to sequences of saddles. The result is that for each sequence of saddles one can iteratively determine the asymptotic probability of realization of each next step along that sequence and the scaling asymptotics of the associated exit distributions. This was done in [Bak11], [Bak10], and [AB11]. The result is that at the logarithmic time scales (the saddle exit times are typically logarithmic in  $\varepsilon$ , see (6.1)), certain pathways in the network are typical but many pathways are not realized due to the largely one-sided exit distributions scaling as  $\varepsilon^\alpha$  with  $\alpha < 1$  which in turn are due to insufficient contraction at saddles.

This is interesting per se and among other applications gives an explanation of the poor vocabulary of excitation patterns in neural networks and similar dynamics modeled by Lotka–Volterra type systems with small noise. However, this information is not sufficient to address questions about the time scales that are longer than logarithmic such as the limiting behavior of the invariant distribution. To answer these questions, one must quantify the probabilities of rare events corresponding to atypical exits from saddle points. This means that one needs to study probabilities of events like  $P(Y_\varepsilon(\tau_\varepsilon) \sim \varepsilon)$  for  $Y_\varepsilon(\tau_\varepsilon)$  given in (6.2). The second term in (6.2) is of the order of  $\varepsilon$ , so ignoring many technical details we reduce this question to estimating

$$P(\varepsilon^\rho |N(\infty)|^\rho \sim \varepsilon) \sim P(N(\infty) \sim \varepsilon^\gamma) \sim c\varepsilon^\gamma \quad (6.3)$$

with  $\gamma = \frac{1}{\rho} - 1$ . The last relation holds since  $N(\infty)$  has continuous Lebesgue density at 0.

This means that the probability of an atypical exit from the saddle is asymptotically polynomial in  $\varepsilon$ , of the order of  $\varepsilon^\gamma$  which in turn means that one typically has to wait for time of the order of  $\varepsilon^{-\gamma} \log \varepsilon^{-1}$  before one sees such a rare event happen. Ordering all exponents emerging in such calculations for all rare transitions:  $\gamma_1 < \gamma_2 < \dots < \gamma_N$  and introducing  $T_{k,\varepsilon} = \varepsilon^{-\gamma_k} \log \varepsilon^{-1}$ , we see that for  $t_\varepsilon$  satisfying

$$T_{k,\varepsilon} \ll t_\varepsilon \ll T_{k+1,\varepsilon},$$

transitions can be classified into admissible (that typically occur many times up to  $t_\varepsilon$ ) and rare (that typically do not occur at all up to  $t_\varepsilon$ ). As  $t_\varepsilon$  crosses a level  $T_{k,\varepsilon}$  from below, some new transitions become available. Increasing  $t_\varepsilon$  gradually from 0 to values beyond  $T_{N,\varepsilon}$  creates a hierarchical structure of merging clusters and associated time scales such that at each time scale, the system explores one cluster making no transitions between different clusters.

This picture containing the description of the limit of invariant distribution, homogenization results, etc., is similar to the Freidlin–Wentzell picture of metastability and the associated hierarchy of cycles. The important difference is that in our picture the probabilities of rare

events decay polynomially and the associated time scales grow polynomially while the large deviation estimates in the Freidlin–Wentzell theory lead to exponentially decaying probabilities and exponentially growing transition times between the metastable states.

We do not have a rigorous derivation of a general precise version of the asymptotic relation (6.3). The difficulties that emerge are related to handling nonlinearities in the drift and diffusion terms and to the fact that  $N(\infty)$  and  $M(\infty)$  are only approximations to the (mutually dependent) random variables  $N_\varepsilon(\tau_\varepsilon)$  and  $M_\varepsilon(\tau_\varepsilon)$ , where the  $\varepsilon$  subscript of  $N_\varepsilon$  and  $M_\varepsilon$  refers to the fact that for the case of non-additive noise, these stochastic processes do depend on  $\varepsilon$ . So, to realize this program, among other things we must either prove that the density of  $N_\varepsilon(\tau_\varepsilon)$  uniformly converges to the Gaussian density in a small neighborhood of zero, or to find other means to compare distribution of  $N_\varepsilon(\tau_\varepsilon)$  to the Gaussian at small scales, which requires going beyond the known weak convergence of distributions.

According to (6.1), if  $N(\infty)$  takes an atypically small value of the order of  $\varepsilon^\gamma$ , then

$$\tau_\varepsilon \approx \frac{1+\gamma}{\lambda} \log \frac{1}{\varepsilon}, \quad (6.4)$$

i.e., exit takes abnormally long time (the typical exit time corresponds to  $\gamma = 0$ ). In other words, the rare transitions determining the long-term behavior of diffusions near heteroclinic networks occur due to atypically long stays in the neighborhood of saddle points withstanding the repulsion in the unstable direction.

In the present paper (as well as in [BPG19]), we study the polynomial decay of the distribution of exit times at scales described by (6.4). We believe that the method we propose here is applicable in the multi-dimensional situation and we plan to give a rigorous treatment of it in upcoming publications.

## References

- [AB11] S. Almada and Y. Bakhtin. Normal forms approach to diffusion near hyperbolic equilibria. Nonlinearity, 24:1883–1907, 2011.
- [Bak10] Yuri Bakhtin. Small noise limit for diffusions near heteroclinic networks. Dynamical Systems, 25(3):413–431, 2010.
- [Bak11] Y. Bakhtin. Noisy heteroclinic networks. Probab. Theory Relat. Fields, 150:1–42, 2011.
- [Bas11] Richard F. Bass. Stochastic processes, volume 33 of Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2011.
- [BPG19] Yuri Bakhtin and Zsolt Pajor-Gyulai. Malliavin calculus approach to long exit times from an unstable equilibrium. Ann. Appl. Probab., 29(2):827–850, 04 2019.
- [CV16] Nicolas Champagnat and Denis Villemonais. Exponential convergence to quasi-stationary distribution and  $Q$ -process. Probability Theory and Related Fields, 164(1):243–283, Feb 2016.
- [Eiz84] Alexander Eizenberg. The exit distributions for small random perturbations of dynamical systems with a repulsive type stationary point. Stochastics, 12(3-4):251–275, 1984.

- [KS91] I. Karatzas and S.E. Shreve. Brownian Motion and Stochastic Calculus. Graduate Texts in Mathematics. Springer New York, 1991.