Geometric singular perturbation theory for stochastic differential equations

Nils Berglund and Barbara Gentz

Abstract

We consider slow–fast systems of differential equations, in which both the slow and fast variables are perturbed by additive noise. When the deterministic system admits a uniformly asymptotically stable slow manifold, we show that the sample paths of the stochastic system are concentrated in a neighbourhood of the slow manifold, which we construct explicitly. Depending on the dynamics of the reduced system, the results cover time spans which can be exponentially long in the noise intensity squared (that is, up to Kramers' time). We give exponentially small upper and lower bounds on the probability of exceptional paths. If the slow manifold contains bifurcation points, we show similar concentration properties for the fast variables corresponding to non-bifurcating modes. We also give conditions under which the system can be approximated by a lower-dimensional one, in which the fast variables contain only bifurcating modes.

Date. March 30, 2002.
2000 *Mathematical Subject Classification.* 37H20, 34E15 (primary), 60H10 (secondary)
Keywords and phrases. Singular perturbations, slow–fast systems, invariant manifolds, dynamic bifurcations, stochastic differential equations, first-exit times, concentration of measure.

1 Introduction

Systems involving two well-separated timescales are often described by slow–fast differential equations of the form

$$\begin{aligned} \varepsilon\dot{x} &= f(x,y,\varepsilon), \\ \dot{y} &= g(x,y,\varepsilon), \end{aligned} \tag{1.1}$$

where ε is a small parameter. Since $\dot{x}$ can be much larger than $\dot{y}$, x is called the *fast variable* and y is called the *slow variable*. Such equations occur, for instance, in climatology, with the slow variables describing the state of the oceans, and the fast variables the state of the atmosphere. In physics, slow–fast equations model in particular systems containing heavy particles (e. g. nuclei) and light particles (e. g. electrons). Another example, taken from ecology, would be the dynamics of a predator–prey system in which the rates of reproduction of predator and prey are very different.

The system (1.1) behaves singularly in the limit $\varepsilon \to 0$. In fact, the results depend on the way this limit is performed. If we simply set ε to zero in (1.1), we obtain the algebraic–differential system

$$\begin{aligned} 0 &= f(x,y,0), \\ \dot{y} &= g(x,y,0). \end{aligned} \tag{1.2}$$

Assume there exists a differentiable manifold with equation $x = x^\star(y)$ on which $f = 0$. Then $x = x^\star(y)$ is called a *slow manifold*, and the dynamics on it is described by the *reduced equation*

$$\dot{y} = g(x^\star(y), y, 0). \tag{1.3}$$

Another way to analyze the limit $\varepsilon \to 0$ is to scale time by a factor $1/\varepsilon$, so that the slow–fast system (1.1) becomes

$$\begin{aligned} x' &= f(x,y,\varepsilon), \\ y' &= \varepsilon g(x,y,\varepsilon). \end{aligned} \tag{1.4}$$

In the limit $\varepsilon \to 0$, we obtain the so-called *associated system*

$$\begin{aligned} x' &= f(x,y,0), \\ y' &= 0, \end{aligned} \tag{1.5}$$

in which y plays the rôle of a parameter. The slow manifold $x = x^\star(y)$ consists of equilibrium points of (1.5), and (1.4) can be viewed as a perturbation of (1.5) with slowly drifting parameter y.

Under certain conditions, both the reduced equation (1.3) and the associated system (1.5) give good approximations of the initial slow–fast system (1.1), but on different timescales. Assume for instance that for each y, $x^\star(y)$ is an asymptotically stable equilibrium of the associated system (1.5). Then solutions of (1.1) starting in a neighbourhood of the slow manifold will approach $x^\star(y)$ in a time of order $\varepsilon|\log\varepsilon|$. During this time interval they are well approximated by solutions of (1.5). This first phase of the motion is sometimes called the *boundary-layer* behaviour. For larger times, solutions of (1.1) remain in an ε-neighbourhood of the slow manifold, and are thus well approximated by solutions of the reduced equation (1.3). This result was first proved by Gradšteĭn [15] and Tihonov [26].

Fenichel [11] has given results allowing for a geometrical description of these phenomena in terms of invariant manifolds. He showed, in particular, the existence of an invariant manifold

$$x = \bar{x}(y,\varepsilon), \qquad \text{with } \bar{x}(y,\varepsilon) = x^\star(y) + \mathcal{O}(\varepsilon), \tag{1.6}$$

for sufficiently small ε, whenever $x^\star(y)$ is a family of hyperbolic equilibria of the associated system (1.5). The dynamics on this invariant manifold is given by the equation

$$\dot{y} = g(\bar{x}(y,\varepsilon), y, \varepsilon), \tag{1.7}$$

which can be treated by methods of regular perturbation theory, and reduces to (1.3) in the limit $\varepsilon \to 0$. In fact, Fenichel's results are more general. For instance, if $x^\star(y)$ is a saddle, they also show the existence of invariant manifolds associated with the stable and unstable manifolds of $x^\star(y)$. See [17] for a review.

New, interesting phenomena arise when the dynamics of (1.7) causes y to approach a bifurcation point of (1.5). For instance, the passage through a saddle–node bifurcation, corresponding to a fold of the slow manifold, produces a jump to some other region in phase space, which can cause relaxation oscillations and hysteresis phenomena (see in particular [24] and [16], as well as [21] for an overview). Transcritical and pitchfork bifurcations generically lead to a smoother transition to another equilibrium [20, 19], while the passage through a Hopf bifurcation is accompanied by the delayed appearance of oscillations [22, 23]. There exist many more recent studies of what has become known as the field of dynamic bifurcations, see for instance [4].

In many situations, low-dimensional ordinary differential equations of the form $\dot{x} = f(x)$ are not sufficient to describe the dynamics of the system under study. The effect of unknown degrees of freedom is often modelled by noise, leading to a stochastic differential equation (SDE) of the form

$$\mathrm{d}x_t = f(x_t)\,\mathrm{d}t + \sigma F(x_t)\,\mathrm{d}W_t, \tag{1.8}$$

where σ is a small parameter, and W_t denotes a standard, generally vector-valued Brownian motion. On short timescales, the main effect of the noise term $\sigma F(x_t)\,\mathrm{d}W_t$ is to cause solutions to fluctuate around their deterministic counterpart, but the probability of large deviations is very small (of the order $\mathrm{e}^{-const/\sigma^2}$). On longer timescales, however, the noise term can induce transitions to other regions of phase space.

The best understood situation is the one where f admits an asymptotically stable equilibrium point $x^\star$. The first-exit time $\tau(\omega)$ of the sample path $x_t(\omega)$ from a neighbourhood of $x^\star$ is a random variable, the characterization of which is the object of the *exit problem.* If f derives from a potential U (i.e., $f = -\nabla U$) of which $x^\star$ is a local minimum, the asymptotic behaviour of the typical first-exit time for $\sigma \ll 1$ has been long known by physicists: it is of order e^{2H/σ^2}, where H is the height of the lowest potential barrier separating $x^\star$ from other potential wells. A theory of large deviations generalizing this result to quite a large class of SDEs has been developed by Freidlin and Wentzell [14]. More detailed information on the asymptotics of the expected first-exit time, and on the distribution of τ, has been obtained, see in particular [2, 12, 9].

The more difficult problem of the dynamics near a saddle point has been studied in [18] and in [10]. The situation where f depends on a parameter and undergoes bifurcations has not yet been studied in that much detail. An approach based on the notion of random attractors [25, 1, 8] gives information on the limit $t \to \infty$, when the system has reached a stationary state. Note, however, that the time needed to reach this regime, in which (in the gradient case) x_t is most likely to be found near the deepest potential well, may be very long if the wells are separated by barriers substantially higher than σ^2. The dynamics on intermediate timescales, known as the *metastable regime*, is not yet well understood in the presence of bifurcations.

In this work, we are interested in the effect of additive noise on slow–fast systems of the form (1.1). Such systems have been studied before in [13], using techniques from large deviation theory to describe the limit $\sigma \to 0$. Here we use different methods to give a more precise description of the regime of small, but finite noise intensity, our main goal being to estimate quantitatively the noise-induced spreading of typical paths, as well as the probability of exceptional paths. We will consider situations in which both the slow and fast variables are affected by noise, with noise intensities taking into account the difference between the timescales. In (1.8), the diffusive nature of the Brownian motion causes paths to spread like $\sigma\sqrt{t}$. In the case of the slow–fast system (1.1), we shall choose the following scaling of the noise intensities:

$$\begin{aligned} \mathrm{d}x_t &= \frac{1}{\varepsilon} f(x_t, y_t, \varepsilon)\,\mathrm{d}t + \frac{\sigma}{\sqrt{\varepsilon}} F(x_t, y_t, \varepsilon)\,\mathrm{d}W_t, \\ \mathrm{d}y_t &= g(x_t, y_t, \varepsilon)\,\mathrm{d}t + \sigma' G(x_t, y_t, \varepsilon)\,\mathrm{d}W_t. \end{aligned} \tag{1.9}$$

In this way, σ^2 and $(\sigma')^2$ both measure the ratio between the rate of diffusion squared and the speed of drift, respectively, for the fast and slow variable. We consider general finitedimensional $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$, while W_t denotes a k-dimensional standard Brownian motion. Accordingly, F and G are matrix-valued functions of respective dimensions $n \times k$ and $m \times k$. We consider ε, σ and σ' as small parameters, and think of σ and σ' as functions of ε. We limit the analysis to situations where σ' does not dominate σ, i.e., we assume $\sigma' = \rho\sigma$ where ρ may depend on ε but is uniformly bounded above in ε.

We first consider the case where the deterministic slow–fast system (1.1) admits an asymptotically stable slow manifold $x^\star(y)$. Our first main result, Theorem 2.4, states that the sample paths of (1.9) are concentrated in a "layer" surrounding the adiabatic manifold $\bar{x}(y,\varepsilon)$, of the form

$$\mathcal{B}(h) = \bigl\{(x,y) \colon \bigl\langle (x - \bar{x}(y,\varepsilon)), \overline{X}(y,\varepsilon)^{-1}(x - \bar{x}(y,\varepsilon))\bigr\rangle < h^2\bigr\} \tag{1.10}$$

up to time t, with a probability behaving roughly like $(t^2/\varepsilon)\,\mathrm{e}^{-h^2/2\sigma^2}$ as long as the paths do not reach the vicinity of a bifurcation point. The matrix $\overline{X}(y,\varepsilon)$, defining the elliptical cross-section of the layer, is itself a solution of a slow–fast system, and depends only on the values of F and $\partial_x f$ on the slow manifold. In particular, $\overline{X}(y,0)$ is a solution of the Liapunov equation

$$A^\star(y)X + XA^\star(y)^T + F(x^\star(y),y,0)F(x^\star(y),y,0)^T = 0, \tag{1.11}$$

where $A^\star(y) = \partial_x f(x^\star(y),y,0)$. For instance, if f derives from a potential U, $-A^\star$ is the Hessian matrix of U at its minimum, and $\mathcal{B}(h)$ is more elongated in those directions in which the curvature of U is smallest.

Theorem 2.5 gives a more detailed description of the dynamics inside $\mathcal{B}(h)$, by showing that paths (x_t, y_t) are concentrated in a neighbourhood of the deterministic solution $(x_t^{\det}, y_t^{\det})$ at least up to times of order 1. The spreading in the y-direction grows at a rate corresponding to the finite-time Lyapunov exponents of the deterministic solution.

Next we turn to situations where the deterministic solution approaches a bifurcation point of the associated system. In this case, the adiabatic manifold $\bar{x}(y,\varepsilon)$ is not defined in general. However, by splitting x into a stable direction x^- and a bifurcating direction z, one can define a (centre) manifold $x^- = \bar{x}^-(z,y,\varepsilon)$ which is locally invariant under the deterministic flow. Theorem 2.7 shows that paths of the stochastic system are concentrated in a neighbourhood of $\bar{x}^-(z,y,\varepsilon)$. The size of this neighbourhood again depends on noise and linearized drift term in the stable x^--direction.

In order to make use of previous results on the passage through bifurcation points for one-dimensional fast variables, such as [7, 5, 6], it is necessary to control the deviation between solutions of the full system (1.9), and the reduced stochastic system obtained by setting x^- equal to $\bar{x}^-(z,y,\varepsilon)$. Theorem 2.8 provides such an estimate under certain assumptions on the dynamics of the reduced system.

We present the detailed results in Section 2, Subsection 2.2 containing a summary of results on deterministic slow–fast systems, while Subsection 2.3 is dedicated to the random case with a stable slow manifold and Subsection 2.4 to the case of bifurcations. Sections 3 to 5 contain the proofs of these results.

Acknowledgements

B. G. thanks the Forschungsinstitut für Mathematik at the ETH Zürich and its director Professor Marc Burger for kind hospitality.

2 Results

2.1 Preliminaries

Let $\mathcal{D}$ be an open subset of $\mathbb{R}^n \times \mathbb{R}^m$ and ε_0 a constant. We consider slow–fast stochastic differential equations of the form

$$
\begin{aligned}
\mathrm{d}x_t &= \frac{1}{\varepsilon} f(x_t, y_t, \varepsilon)\,\mathrm{d}t + \frac{\sigma}{\sqrt{\varepsilon}} F(x_t, y_t, \varepsilon)\,\mathrm{d}W_t, \\
\mathrm{d}y_t &= g(x_t, y_t, \varepsilon)\,\mathrm{d}t + \sigma' G(x_t, y_t, \varepsilon)\,\mathrm{d}W_t,
\end{aligned}
\tag{2.1}
$$

with drift coefficients $f \in \mathcal{C}^2(\mathcal{D} \times [0, \varepsilon_0), \mathbb{R}^n)$ and $g \in \mathcal{C}^2(\mathcal{D} \times [0, \varepsilon_0), \mathbb{R}^m)$, and diffusion coefficients $F \in \mathcal{C}^1(\mathcal{D} \times [0, \varepsilon_0), \mathbb{R}^{n \times k})$ and $G \in \mathcal{C}^1(\mathcal{D} \times [0, \varepsilon_0), \mathbb{R}^{m \times k})$.

We require that f, g, and all their derivatives up to order 2 are uniformly bounded in norm in $\mathcal{D} \times [0, \varepsilon_0)$, and similarly for F, G and their derivatives. We also assume that f and g satisfy the usual (local) Lipschitz and bounded-growth conditions which guarantee existence and pathwise uniqueness of a strong solution $\{(x_t, y_t)\}_{t \geqslant t_0}$ of (2.1).

The stochastic process $\{W_t\}_{t \geqslant 0}$ is a standard k-dimensional Brownian motion on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Initial conditions (x_0, y_0) are always assumed to be square-integrable with respect to $\mathbb{P}$ and independent of $\{W_t\}_{t \geqslant 0}$. Our assumptions on f and g guarantee the existence of a continuous version of $\{(x_t, y_t)\}_{t \geqslant 0}$. Therefore we may assume that the paths $\omega \mapsto (x_t(\omega), y_t(\omega))$ are continuous for $\mathbb{P}$-almost all $\omega \in \Omega$.

We introduce the notation $\mathbb{P}^{t_0,(x_0,y_0)}$ for the law of the process $\{(x_t, y_t)\}_{t \geqslant t_0}$, starting in (x_0, y_0) at time t_0, and use $\mathbb{E}^{t_0,(x_0,y_0)}$ to denote expectations with respect to $\mathbb{P}^{t_0,(x_0,y_0)}$. Note that the stochastic process $\{(x_t, y_t)\}_{t \geqslant t_0}$ is a time-homogeneous Markov process. Let $\mathcal{A} \subset \mathcal{D}$ be Borel-measurable. Assuming $(x_0, y_0) \in \mathcal{A}$, we denote by

$$
\tau_{\mathcal{A}} = \inf\bigl\{t \geqslant 0 \colon (x_t, y_t) \notin \mathcal{A}\bigr\}
\tag{2.2}
$$

the first-exit time of (x_t, y_t) from $\mathcal{A}$. Note that $\tau_{\mathcal{A}}$ is a stopping time with respect to the filtration of $(\Omega, \mathcal{F}, \mathbb{P})$ generated by the Brownian motion $\{W_t\}_{t \geqslant 0}$.

Throughout this work, we use the following notations:

- Let a, b be real numbers. We denote by $\lceil a \rceil$, $a \wedge b$ and $a \vee b$, respectively, the smallest integer greater than or equal to a, the minimum of a and b, and the maximum of a and b.
- By $g(u) = \mathcal{O}(u)$ we indicate that there exist $\delta > 0$ and $K > 0$ such that $g(u) \leqslant Ku$ for all $u \in [0, \delta]$, where δ and K of course do not depend on ε, σ or σ'.
- We use $\|x\|$ to denote the Euclidean norm of $x \in \mathbb{R}^d$, while $\|A\|$ stands for the corresponding operator norm of a matrix $A \in \mathbb{R}^{d_1 \times d_2}$. If $A(t)$ is a matrix-valued function defined for t in an interval I, we denote by $\|A\|_I$ the supremum of $\|A(t)\|$ over $t \in I$, and often we write $\|A\|_\infty$ if the interval is evident from the context.
- For a given set B, we denote by 1_B the indicator function on B, defined by $1_B(x) = 1$, if $x \in B$, and $1_B(x) = 0$, otherwise.
- If $\mathbb{R}^n \times \mathbb{R}^m \ni (x, y) \mapsto f(x, y) \in \mathbb{R}^d$ is differentiable, we write $\partial_x f(x, y)$ and $\partial_y f(x, y)$ to denote the Jacobian matrices of $x \mapsto f(x, y)$ and $y \mapsto f(x, y)$, respectively.

2.2 Deterministic stable case

We start by recalling a few properties of deterministic slow–fast systems of the form

$$\begin{aligned} \varepsilon\dot{x} &= f(x,y,\varepsilon), \\ \dot{y} &= g(x,y,\varepsilon). \end{aligned} \tag{2.3}$$

Definition 2.1. *Let $\mathcal{D}_0 \subset \mathbb{R}^m$ and assume that there exists a (continuous) function $x^\star : \mathcal{D}_0 \to \mathbb{R}^n$ such that*

- *$(x^\star(y), y) \in \mathcal{D}$ for all $y \in \mathcal{D}_0$,*
- *$f(x^\star(y), y, 0) = 0$ for all $y \in \mathcal{D}_0$.*

Then the set $\{(x,y) \colon x = x^\star(y), y \in \mathcal{D}_0\}$ is called a slow manifold *of the system* (2.3).

Let $A^\star(y) = \partial_x f(x^\star(y), y, 0)$. The slow manifold is called

- hyperbolic *if all eigenvalues of $A^\star(y)$ have nonzero real parts for all $y \in \mathcal{D}_0$;*
- uniformly hyperbolic *if all eigenvalues of $A^\star(y)$ have real parts uniformly bounded away from zero (for $y \in \mathcal{D}_0$);*
- asymptotically stable *if all eigenvalues of $A^\star(y)$ have negative real parts for all $y \in \mathcal{D}_0$;*
- uniformly asymptotically stable *if all eigenvalues of $A^\star(y)$ have negative real parts, uniformly bounded away from zero for $y \in \mathcal{D}_0$.*

Gradšteĭn [15] and Tihonov [26] have shown that if $x^\star$ represents a uniformly hyperbolic slow manifold of (2.3), then the system (2.3) admits particular solutions which remain in a neighbourhood of order ε of the slow manifold. If, moreover, the slow manifold is asymptotically stable, then the solutions starting in a neighbourhood of order 1 of the slow manifold converge exponentially fast in t/ε to an ε-neighbourhood of the slow manifold.

Fenichel [11] has given extensions of this result based on a geometrical approach. If (2.3) admits a hyperbolic slow manifold, then there exists, for sufficiently small ε, an invariant manifold

$$y = \bar{x}(y,\varepsilon) = x^\star(y) + \mathcal{O}(\varepsilon), \qquad y \in \mathcal{D}_0. \tag{2.4}$$

Here invariant means that if $y_0 \in \mathcal{D}_0$ and $x_0 = \bar{x}(y_0,\varepsilon)$, then $x_t = \bar{x}(y_t,\varepsilon)$ as long as t is such that $y_s \in \mathcal{D}_0$ for all $s \leqslant t$. We will call the set $\{(\bar{x}(y,\varepsilon), y) \colon y \in \mathcal{D}_0\}$ an *adiabatic manifold.* It is easy to see from (2.3) that $\bar{x}(y,\varepsilon)$ must satisfy the PDE

$$\varepsilon \partial_y \bar{x}(y,\varepsilon) g(\bar{x}(y,\varepsilon), y, \varepsilon) = f(\bar{x}(y,\varepsilon), y, \varepsilon). \tag{2.5}$$

The local existence of the adiabatic manifold follows directly from the centre manifold theorem. Indeed, we can rewrite System (2.3) in the form

$$\begin{aligned} x' &= f(x,y,\varepsilon), \\ y' &= \varepsilon g(x,y,\varepsilon), \\ \varepsilon' &= 0, \end{aligned} \tag{2.6}$$

where prime denotes derivation with respect to the fast time t/ε. Any point of the form $(x^\star(y), y, 0)$ with $y \in \mathcal{D}_0$ is an equilibrium point of (2.6). The linearization of (2.6) around such a point admits 0 as eigenvalue of multiplicity $m+1$, the n other eigenvalues being those of $A^\star(y)$, which are bounded away from the imaginary axis. The centre manifold theorem implies the existence of a local invariant manifold $x = \bar{x}(y,\varepsilon)$. Fenichel's result shows that this manifold actually exists for all $y \in \mathcal{D}_0$.

Being a centre manifold, the adiabatic manifold is not necessarily unique (though in the present case, $\bar{x}(y,0) = x^\star(y)$ is uniquely defined). Nevertheless, $\bar{x}(y,\varepsilon)$ has a unique Taylor series in y and ε, which can be obtained by solving (2.5) order by order. The dynamics on the adiabatic manifold is described by the so-called *reduced equation*

$$\dot{y} = g(\bar{x}(y,\varepsilon), y, \varepsilon) = g(x^\star(y), y, 0) + \mathcal{O}(\varepsilon). \tag{2.7}$$

If $x^\star(y)$ is uniformly asymptotically stable, $\bar{x}(y,\varepsilon)$ is locally attractive and thus any solution of (2.3) starting sufficiently close to $\bar{x}(y,\varepsilon)$ converges exponentially fast to a solution of (2.7).

2.3 Random stable case

We turn now to the random slow–fast system given by the stochastic differential equation

$$\begin{aligned}
\mathrm{d}x_t &= \frac{1}{\varepsilon} f(x_t, y_t, \varepsilon)\,\mathrm{d}t + \frac{\sigma}{\sqrt{\varepsilon}} F(x_t, y_t, \varepsilon)\,\mathrm{d}W_t, \\
\mathrm{d}y_t &= g(x_t, y_t, \varepsilon)\,\mathrm{d}t + \sigma' G(x_t, y_t, \varepsilon)\,\mathrm{d}W_t,
\end{aligned} \tag{2.8}$$

where we will assume the following.

Assumption 2.2. For $\sigma = \sigma' = 0$, System (2.8) admits a uniformly hyperbolic, asymptotically stable slow manifold $x = x^\star(y)$, $y \in \mathcal{D}_0$.

By Fenichel's theorem, there exists an adiabatic manifold $x = \bar{x}(y,\varepsilon)$ with $\bar{x}(y,0) = x^\star(y)$, $y \in \mathcal{D}_0$. We fix a particular solution $(x_t^{\mathrm{det}}, y_t^{\mathrm{det}}) = (\bar{x}(y_t^{\mathrm{det}}, \varepsilon), y_t^{\mathrm{det}})$ of the deterministic system. (That is, y_t^{det} satisfies the reduced equation (2.7).) We want to describe the noise-induced deviations of the sample paths $(x_t, y_t)_{t \geqslant 0}$ of (2.8) from the adiabatic manifold.

It turns out to be convenient to use the transformation

$$\begin{aligned}
x_t &= \bar{x}(y_t^{\mathrm{det}} + \eta_t, \varepsilon) + \xi_t, \\
y_t &= y_t^{\mathrm{det}} + \eta_t,
\end{aligned} \tag{2.9}$$

which yields a system of the form

$$\begin{aligned}
\mathrm{d}\xi_t &= \frac{1}{\varepsilon} \hat{f}(\xi_t, \eta_t, t, \varepsilon)\,\mathrm{d}t + \frac{\sigma}{\sqrt{\varepsilon}} \widehat{F}(\xi_t, \eta_t, t, \varepsilon)\,\mathrm{d}W_t, \\
\mathrm{d}\eta_t &= \hat{g}(\xi_t, \eta_t, t, \varepsilon)\,\mathrm{d}t + \sigma' \widehat{G}(\xi_t, \eta_t, t, \varepsilon)\,\mathrm{d}W_t,
\end{aligned} \tag{2.10}$$

where the new drift and diffusion coefficients are given by

$$\begin{aligned}
\hat{f}(\xi,\eta,t,\varepsilon) &= f(\bar{x}(y_t^{\mathrm{det}} + \eta, \varepsilon) + \xi, y_t^{\mathrm{det}} + \eta, \varepsilon) \\
&\qquad - \varepsilon \partial_y \bar{x}(y_t^{\mathrm{det}} + \eta, \varepsilon) g(\bar{x}(y_t^{\mathrm{det}} + \eta, \varepsilon) + \xi, y_t^{\mathrm{det}} + \eta, \varepsilon), \\
\widehat{F}(\xi,\eta,t,\varepsilon) &= F(\bar{x}(y_t^{\mathrm{det}} + \eta, \varepsilon) + \xi, y_t^{\mathrm{det}} + \eta, \varepsilon) \\
&\qquad - \rho\sqrt{\varepsilon} \partial_y \bar{x}(y_t^{\mathrm{det}} + \eta, \varepsilon) G(\bar{x}(y_t^{\mathrm{det}} + \eta, \varepsilon) + \xi, y_t^{\mathrm{det}} + \eta, \varepsilon), \\
\hat{g}(\xi,\eta,t,\varepsilon) &= g(\bar{x}(y_t^{\mathrm{det}} + \eta, \varepsilon) + \xi, y_t^{\mathrm{det}} + \eta, \varepsilon) - g(\bar{x}(y_t^{\mathrm{det}}, \varepsilon), y_t^{\mathrm{det}}, \varepsilon), \\
\widehat{G}(\xi,\eta,t,\varepsilon) &= G(\bar{x}(y_t^{\mathrm{det}} + \eta, \varepsilon) + \xi, y_t^{\mathrm{det}} + \eta, \varepsilon).
\end{aligned} \tag{2.11}$$

Note that because of the property (2.5) of the adiabatic manifold, $\hat{f}(0,0,t,\varepsilon) = 0$. We introduce the notation

$$\begin{aligned} A(y_t^{\det},\varepsilon) &= \partial_\xi \hat{f}(0,0,t,\varepsilon) \\ &= \partial_x f(\bar{x}(y_t^{\det},\varepsilon), y_t^{\det},\varepsilon) - \varepsilon \partial_y \bar{x}(y_t^{\det},\varepsilon)\partial_x g(\bar{x}(y_t^{\det},\varepsilon), y_t^{\det},\varepsilon) \end{aligned} \tag{2.12}$$

for the linearization of $\hat{f}$ at the origin. Note that for $\varepsilon = 0$, we have $A(y_t^{\det},0) = \partial_x f(\bar{x}(y_t^{\det},0), y_t^{\det},0) = A^\star(y_t^{\det})$, so that by Assumption 2.2, the eigenvalues of $A(y_t^{\det},\varepsilon)$ have negative real parts for sufficiently small ε.

One of the basic ideas of our approach is to compare the solutions of (2.10) with those of the "linear approximation"

$$\begin{aligned} \mathrm{d}\xi_t^0 &= \frac{1}{\varepsilon} A(y_t^{\det},\varepsilon)\xi_t^0 \,\mathrm{d}t + \frac{\sigma}{\sqrt{\varepsilon}} F_0(y_t^{\det},\varepsilon)\,\mathrm{d}W_t, \\ \mathrm{d}y_t^{\det} &= g(\bar{x}(y_t^{\det},\varepsilon), y_t^{\det},\varepsilon)\,\mathrm{d}t, \end{aligned} \tag{2.13}$$

where $F_0(y_t^{\det},\varepsilon) = \widehat{F}(0,0,t,\varepsilon)$. Note that the definition of the adiabatic manifold implies $F_0(y,0) = F(x^\star(y),y,0)$. For fixed t, ξ_t^0 is a Gaussian random variable with covariance matrix

$$\operatorname{Cov}(\xi_t^0) = \frac{\sigma^2}{\varepsilon}\int_0^t U(t,s)F_0(y_s^{\det},\varepsilon)F_0(y_s^{\det},\varepsilon)^T U(t,s)^T \,\mathrm{d}s, \tag{2.14}$$

where $U(t,s)$ denotes the principal solution of the homogeneous system $\varepsilon\dot{\xi} = A(y_t^{\det},\varepsilon)\xi$.

We now observe that $\sigma^{-2}\operatorname{Cov}(\xi_t^0)$ is the X-variable of a particular solution of the slow–fast system

$$\begin{aligned} \varepsilon\dot{X} &= A(y,\varepsilon)X + XA(y,\varepsilon)^T + F_0(y,\varepsilon)F_0(y,\varepsilon)^T, \\ \dot{y} &= g(\bar{x}(y,\varepsilon),y,\varepsilon). \end{aligned} \tag{2.15}$$

This system admits a slow manifold $X = X^\star(y)$, given by the Liapunov equation

$$A^\star(y)X^\star(y) + X^\star(y)A^\star(y)^T + F_0(y,0)F_0(y,0)^T = 0, \tag{2.16}$$

which is known [3] to admit the (unique) solution

$$X^\star(y) = \int_0^\infty \mathrm{e}^{sA^\star(y)}\, F_0(y,0)F_0(y,0)^T \,\mathrm{e}^{sA^\star(y)^T}\,\mathrm{d}s. \tag{2.17}$$

Moreover, the eigenvalues of the operator $X \mapsto AX + XA^T$ are exactly $a_i + a_j$, $1 \leqslant i,j \leqslant n$, where a_i are the eigenvalues of A. Thus the slow manifold $X = X^\star(y)$ is uniformly asymptotically stable (for small enough ε), so that Fenichel's theorem shows the existence of an adiabatic manifold

$$X = \overline{X}(y,\varepsilon) = X^\star(y) + \mathcal{O}(\varepsilon). \tag{2.18}$$

Note that $\overline{X}(y_t^{\det},\varepsilon)$ is uniquely determined by the "initial" value $\overline{X}(y_0^{\det},\varepsilon)$ via the relation

$$\overline{X}(y_t^{\det},\varepsilon) = U(t)\biggl[\overline{X}(y_0^{\det},\varepsilon) + \frac{1}{\varepsilon}\int_0^t U(s)^{-1}F_0(y_s^{\det},\varepsilon)F_0(y_s^{\det},\varepsilon)^T U(s)^{-T}\,\mathrm{d}s\biggr]U(t)^T, \tag{2.19}$$

where $U(t) = U(t,0)$ and $U(s)^{-T} = [U(s)^{-1}]^T$.

We now introduce the set

$$\mathcal{B}(h) = \left\{(x,y)\colon y \in \mathcal{D}_0, \left\langle (x-\bar{x}(y,\varepsilon)), \overline{X}(y,\varepsilon)^{-1}(x-\bar{x}(y,\varepsilon))\right\rangle < h^2\right\}, \tag{2.20}$$

assuming that $\overline{X}(y,\varepsilon)$ is invertible for all $y \in \mathcal{D}_0$. The set $\mathcal{B}(h)$ is a "layer" around the adiabatic manifold $x = \bar{x}(y,\varepsilon)$, with ellipsoidal cross-section determined by $\overline{X}(y,\varepsilon)$. For fixed t, the solution ξ_t^0 of the linear approximation (2.13) is concentrated (in density) in the cross-section of $\mathcal{B}(\sigma)$ taken at y_t. Our first main result (Theorem 2.4 below) gives conditions under which the whole sample path (x_t, y_t) of the original equation (2.8) is likely to remain in such a set $\mathcal{B}(h)$. By

$$\tau_{\mathcal{B}(h)} = \inf\{t \geqslant 0\colon (x_t,y_t) \notin \mathcal{B}(h)\} \tag{2.21}$$

we denote the first-exit time of the sample path (x_t,y_t) from $\mathcal{B}(h)$.

Remark 2.3. Fix y for the moment. If $X^\star(y)^{-1}$ is bounded, then $\overline{X}(y,\varepsilon)^{-1}$ is bounded for sufficiently small ε. A sufficient condition for $X^\star(y)^{-1}$ to be bounded is that the symmetric matrix $F_0(y,0)F_0(y,0)^T$ be positive definite. This condition is, however, by no means necessary. In fact, $X^\star(y)$ is singular if and only if there exists a vector $x \neq 0$ such that

$$F_0(y,0)^T \,\mathrm{e}^{sA^\star(y)^T} x = 0 \qquad \forall s \geqslant 0, \tag{2.22}$$

which occurs if and only if

$$F_0(y,0)^T (A^\star(y)^T)^k x = 0 \qquad \forall k = 0,1,2,\dots \tag{2.23}$$

i.e., when the kernel of $F_0(y,0)^T$ is invariant under $A^\star(y)^T$.

Theorem 2.4. *Assume that $\|\overline{X}(y,\varepsilon)\|$ and $\|\overline{X}(y,\varepsilon)^{-1}\|$ are uniformly bounded in $\mathcal{D}_0$. Choose a deterministic initial condition $y_0 \in \mathcal{D}_0$, $x_0 = \bar{x}(y_0,\varepsilon)$, and let*

$$\tau_{\mathcal{D}_0} = \inf\{s > 0\colon y_s \notin \mathcal{D}_0\}. \tag{2.24}$$

Then there exist constants $\varepsilon_0, \Delta_0, h_0 > 0$ (independent of the chosen initial condition y_0) such that for all $\varepsilon \leqslant \varepsilon_0$, $\Delta \leqslant \Delta_0$, $h \leqslant h_0$, and all $0 < \gamma < 1/2$, the following assertions hold.

(a) The upper bound: *For all $t > 0$,*

$$\mathbb{P}^{0,(x_0,y_0)}\{\tau_{\mathcal{B}(h)} < t \wedge \tau_{\mathcal{D}_0}\} \leqslant C^+_{n,m,\gamma,\Delta}(t,\varepsilon)\frac{h^2}{\sigma^2}\,\mathrm{e}^{-\kappa^+ h^2/\sigma^2}, \tag{2.25}$$

where

$$\kappa^+ = \gamma\left[1 - \mathcal{O}(h) - \mathcal{O}(\Delta) - \mathcal{O}\left(\mathrm{e}^{-const/\varepsilon}/(1-2\gamma)\right)\right] \tag{2.26}$$

and

$$C^+_{n,m,\gamma,\Delta}(t,\varepsilon) = \frac{(1+t)^2}{\Delta\varepsilon}\left[(1-2\gamma)^{-n} + \mathrm{e}^{n/4} + \mathrm{e}^{m/4}\right]. \tag{2.27}$$

(b) The lower bound: *There exists $t_0 > 0$ of order 1 such for all $t > 0$,*

$$\mathbb{P}^{0,(x_0,y_0)}\{\tau_{\mathcal{B}(h)} < t\} \geqslant C^-_{n,m}(t,\varepsilon,h,\sigma)\,\mathrm{e}^{-\kappa^- h^2/\sigma^2}, \tag{2.28}$$

where

$$\kappa^- = \frac{1}{2}\left[1 + \mathcal{O}(h) + \mathcal{O}\left(\mathrm{e}^{-const\,(t\wedge t_0)/\varepsilon}\right)\right] \tag{2.29}$$

and

$$C^-_{n,m}(t,\varepsilon,h,\sigma) = 1 - \left(\mathrm{e}^{n/4} + 4\,\mathrm{e}^{m/4}\right)\mathrm{e}^{-\kappa^- h^2/(2\sigma^2)}. \tag{2.30}$$

(c) General initial conditions: *There exist $\delta_0 > 0$ and a time t_1 of order $\varepsilon|\log h|$ such that for all $\delta \leqslant \delta_0$, all initial conditions (x_0, y_0) which satisfy $y_0 \in \mathcal{D}_0$ as well as $\langle \xi_0, \overline{X}(y_0)^{-1}\xi_0\rangle < \delta^2$, and all $t > t_1$,*

$$\mathbb{P}^{0,(\xi_0,0)}\Big\{ \sup_{t_1 \leqslant s \leqslant t\wedge\tau_{\mathcal{D}_0}} \langle \xi_s, \overline{X}(y_s)^{-1}\xi_s\rangle \geqslant h^2 \Big\} \leqslant \mathcal{C}^+_{n,m,\gamma,\Delta}(t,\varepsilon)\frac{h^2}{\sigma^2}\,\mathrm{e}^{-\kappa^+ h^2/\sigma^2}, \qquad (2.31)$$

where $\mathcal{C}^+_{n,m,\gamma,\Delta}(t,\varepsilon)$ is the same prefactor as in (2.25), *and*

$$\kappa^+ = \gamma\big[1 - \mathcal{O}(h) - \mathcal{O}(\Delta) - \mathcal{O}\big(\delta\,\mathrm{e}^{-const\,(t_1\wedge 1)/\varepsilon}/(1-2\gamma)\big)\big]. \qquad (2.32)$$

Unless explicitly stated, the error terms in the exponents κ^+ and κ^- are uniform in t.

Estimate (2.25) shows that for $h \gg \sigma$, paths starting in $\mathcal{B}(h)$ are far more likely to leave this set through the "border" $\{y \in \partial\mathcal{D}_0, \langle\xi, \overline{X}(y,\varepsilon)^{-1}\xi\rangle < h^2\}$ than through the "sides" $\{y \in \operatorname{int}\mathcal{D}_0, \langle\xi, \overline{X}(y,\varepsilon)^{-1}\xi\rangle = h^2\}$, unless we wait for time spans exponentially long in h^2/σ^2. Below we discuss how to characterize $\tau_{\mathcal{D}_0}$ more precisely, using information on the reduced dynamics on the adiabatic manifold. If, for instance, all deterministic solutions starting in $\mathcal{D}_0$ remain in this set, $\tau_{\mathcal{D}_0}$ will typically be very large.

The upper bound (2.25) has been designed to yield the best possible exponent κ^+, while the prefactor $\mathcal{C}^+_{n,m,\gamma,\Delta}$ is certainly not optimal. Note that an estimate with the same exponent, but with a smaller prefactor holds for the probability that the endpoint (x_t, y_t) does not lie in $\mathcal{B}(h)$, cf. Corollary 3.10. The parameters Δ and γ can be chosen arbitrarily within their intervals of definition. Taking Δ small and γ close to $1/2$ improves the exponent while increasing the prefactor. A convenient choice is to take Δ and $1/2-\gamma$ of order h or ε. The kind of time-dependence of $\mathcal{C}$ is probably not optimal, but the fact that $\mathcal{C}$ increases with time is to be expected, since it reflects the fact that the probability of observing paths making excursions away from the adiabatic manifold increases with time. As for the dependence of the prefactor on the dimensions n and m, it is due to the fact that the tails of standard Gaussian random variables show their typical decay only outside a ball of radius scaling with the square-root of the dimension.

The upper bound (2.25) and lower bound (2.28) together show that the exponential rate of decay of the probability to leave the set $\mathcal{B}(h)$ before time t behaves like $h^2/(2\sigma^2)$ in the limit of σ, ε and h going to zero, as one would expect from other approaches, based for instance on the theory of large deviations. The bounds hold, however, in a full neighbourhood of $\sigma = \varepsilon = h = 0$.

Finally, Estimate (2.31) allows to extend these results to all initial conditions in a neighbourhood of order 1 of the adiabatic manifold. The only difference is that we have to wait for a time of order $\varepsilon|\log h|$ before the path is likely to have reached the set $\mathcal{B}(h)$. After this time, typical paths behave as if they had started on the adiabatic manifold.

We remark in passing that the assumption that $\|\overline{X}(y,\varepsilon)^{-1}\|$ is uniformly bounded in $\mathcal{D}_0$ excludes purely multiplicative noise.

The behaviour of typical paths depends essentially on the dynamics of the reduced deterministic system (2.7). In fact, in the proof of Theorem 2.4, we use the fact that y_t does not differ too much from $y_t^{\det}$ on timescales of order 1 (see Lemma 3.4). There are thus two main possibilities to be considered:

- either the reduced flow is such that $y_t^{\det}$ reaches the boundary of $\mathcal{D}_0$ in a time of order 1 (for instance, $y_t^{\det}$ may approach a bifurcation set of the slow manifold); then y_t is likely to leave $\mathcal{D}_0$ as well;

- or the reduced flow is such that $y_t^{\det}$ remains in $\mathcal{D}_0$ for all times $t \geqslant 0$; in that case, paths can only leave $\mathcal{B}(h)$ due to the influence of noise, which we expect to be unlikely on subexponential timescales.

We will discuss the first situation in more detail in Subsection 2.4. In both situations, it is desirable to have a more precise description of the deviation η_t of the slow variable y_t from its deterministic counterpart $y_t^{\det}$, in order to achieve a better control of the first-exit time $\tau_{\mathcal{D}_0}$.

The following coupled system gives a better approximation of the dynamics of (2.10) than the system (2.13):

$$\begin{aligned} \mathrm{d}\xi_t^0 &= \frac{1}{\varepsilon} A(y_t^{\det},\varepsilon)\xi_t^0 \,\mathrm{d}t + \frac{\sigma}{\sqrt{\varepsilon}} F_0(y_t^{\det},\varepsilon)\,\mathrm{d}W_t, \\ \mathrm{d}\eta_t^0 &= \big[B(y_t^{\det},\varepsilon)\eta_t^0 + C(y_t^{\det},\varepsilon)\xi_t^0\big]\,\mathrm{d}t + \sigma' G_0(y_t^{\det},\varepsilon)\,\mathrm{d}W_t, \end{aligned} \tag{2.33}$$

where $G_0(y_t^{\det},\varepsilon) = \widehat{G}(0,0,t,\varepsilon) = G(\bar{x}(y_t^{\det},\varepsilon), y_t^{\det},\varepsilon)$ and the Jacobian matrices B and C are given by

$$\begin{aligned} B(y_t^{\det},\varepsilon) &= \partial_\eta \hat{g}(0,0,t,\varepsilon) \\ &= C(y_t^{\det},\varepsilon)\partial_y \bar{x}(y_t^{\det},\varepsilon) + \partial_y g(\bar{x}(y_t^{\det},\varepsilon), y_t^{\det},\varepsilon), \end{aligned} \tag{2.34}$$

$$\begin{aligned} C(y_t^{\det},\varepsilon) &= \partial_\xi \hat{g}(0,0,t,\varepsilon) \\ &= \partial_x g(\bar{x}(y_t^{\det},\varepsilon), y_t^{\det},\varepsilon). \end{aligned} \tag{2.35}$$

The coupled system (2.33) can be written in compact form as

$$\mathrm{d}\zeta_t^0 = \mathcal{A}(y_t^{\det},\varepsilon)\zeta_t^0\,\mathrm{d}t + \sigma\mathcal{F}_0(y_t^{\det},\varepsilon)\,\mathrm{d}W_t, \tag{2.36}$$

where $(\zeta^0)^T = ((\xi^0)^T, (\eta^0)^T)$ and

$$\mathcal{A}(y_t^{\det},\varepsilon) = \begin{pmatrix} \frac{1}{\varepsilon}A(y_t^{\det},\varepsilon) & 0 \\ C(y_t^{\det},\varepsilon) & B(y_t^{\det},\varepsilon) \end{pmatrix}, \qquad \mathcal{F}_0(y_t^{\det},\varepsilon) = \begin{pmatrix} \frac{1}{\sqrt{\varepsilon}}F_0(y_t^{\det},\varepsilon) \\ \rho G_0(y_t^{\det},\varepsilon) \end{pmatrix}. \tag{2.37}$$

The solution of the linear SDE (2.36) is given by

$$\zeta_t^0 = \mathcal{U}(t)\zeta_0 + \sigma\int_0^t \mathcal{U}(t,s)\mathcal{F}_0(y_s^{\det},\varepsilon)\,\mathrm{d}W_s, \tag{2.38}$$

where $\mathcal{U}(t,s)$ denotes the principal solution of the homogeneous system $\dot{\zeta} = \mathcal{A}(y_t^{\det},\varepsilon)\zeta$. It can be written in the form

$$\mathcal{U}(t,s) = \begin{pmatrix} U(t,s) & 0 \\ S(t,s) & V(t,s) \end{pmatrix}, \tag{2.39}$$

where $U(t,s)$ and $V(t,s)$ denote, respectively, the fundamental solutions of $\varepsilon\dot{\xi} = A(y_t^{\det},\varepsilon)\xi$ and $\dot{\eta} = B(y_t^{\det},\varepsilon)\eta$, while

$$S(t,s) = \int_s^t V(t,u)C(y_u^{\det},\varepsilon)U(u,s)\,\mathrm{d}u. \tag{2.40}$$

The Gaussian process ζ_t^0 has a covariance matrix of the form

$$\begin{aligned}\operatorname{Cov}(\zeta_t^0) &= \sigma^2 \int_0^t \mathcal{U}(t,s)\mathcal{F}_0(y_s^{\det},\varepsilon)\mathcal{F}_0(y_s^{\det},\varepsilon)^T\mathcal{U}(t,s)^T \,\mathrm{d}s \\ &= \sigma^2 \begin{pmatrix} X(t) & Z(t) \\ Z(t)^T & Y(t) \end{pmatrix}. \end{aligned} \tag{2.41}$$

The matrices $X(t) \in \mathbb{R}^{n\times n}$, $Y(t) \in \mathbb{R}^{m\times m}$ and $Z(t) \in \mathbb{R}^{n\times m}$ are a particular solution of the following slow–fast system, which generalizes (2.15):

$$\begin{aligned}
\varepsilon\dot{X} &= A(y,\varepsilon)X + XA(y,\varepsilon)^T + F_0(y,\varepsilon)F_0(y,\varepsilon)^T, \\
\varepsilon\dot{Z} &= A(y,\varepsilon)Z + \varepsilon ZB(y,\varepsilon)^T + \varepsilon XC(y,\varepsilon)^T + \sqrt{\varepsilon}\rho F_0(y,\varepsilon)G_0(y,\varepsilon)^T, \\
\dot{Y} &= B(y,\varepsilon)Y + YB(y,\varepsilon)^T + C(y,\varepsilon)Z + Z^TC(y,\varepsilon)^T + \rho^2 G_0(y,\varepsilon)G_0(y,\varepsilon)^T, \\
\dot{y} &= g(\bar{x}(y,\varepsilon),y,\varepsilon).
\end{aligned} \tag{2.42}$$

This system admits a slow manifold given by

$$\begin{aligned}
X &= X^\star(y), \\
Z &= Z^\star(y,\varepsilon) = -\sqrt{\varepsilon}\rho A(y,\varepsilon)^{-1}F_0(y,\varepsilon)G_0(y,\varepsilon)^T + \mathcal{O}(\varepsilon),
\end{aligned} \tag{2.43}$$

where $X^\star(y)$ is given by (2.17). It is straightforward to check that this manifold is uniformly asymptotically stable for sufficiently small ε, so that Fenichel's theorem yields the existence of an adiabatic manifold $X = \overline{X}(y,\varepsilon)$, $Z = \overline{Z}(y,\varepsilon)$, at a distance of order ε from the slow manifold. This manifold attracts nearby solutions of (2.42) exponentially fast, and thus asymptotically, the expectations of $\xi_t^0(\xi_t^0)^T$ and $\xi_t^0(\eta_t^0)^T$ will be close, respectively, to $\sigma^2\overline{X}(y_t^{\det},\varepsilon)$ and $\sigma^2\overline{Z}(y_t^{\det},\varepsilon)$.

In general, the matrix $Y(t)$ cannot be expected to approach some asymptotic value depending only on $y_t^{\det}$ and ε. In fact, if B has eigenvalues with positive real parts, $\|Y(t)\|$ can grow exponentially fast. In order to measure this growth, we introduce the functions

$$\chi^{(1)}(t) = \sup_{0\leqslant s\leqslant t} \int_0^s \Bigl(\sup_{u\leqslant v\leqslant s} \|V(s,v)\|\Bigr)\,\mathrm{d}u, \tag{2.44}$$

$$\chi^{(2)}(t) = \sup_{0\leqslant s\leqslant t} \int_0^s \Bigl(\sup_{u\leqslant v\leqslant s} \|V(s,v)\|^2\Bigr)\,\mathrm{d}u. \tag{2.45}$$

The solution of (2.42) with initial condition $Y(0) = Y_0$ satisfies

$$\begin{aligned}
Y(t;Y_0) &= V(t)Y_0V(t)^T \\
&\quad + \rho^2 \int_0^t V(t,s)G_0(y_t^{\det},\varepsilon)G_0(y_t^{\det},\varepsilon)^T V(t,s)^T\,\mathrm{d}s + \mathcal{O}((\varepsilon+\rho\sqrt{\varepsilon})\chi^{(2)}(t)).
\end{aligned} \tag{2.46}$$

We thus define an "asymptotic" covariance matrix $\overline{\mathcal{Z}}(t) = \overline{\mathcal{Z}}(t;Y_0,\varepsilon)$ by

$$\overline{\mathcal{Z}}(t;Y_0,\varepsilon) = \begin{pmatrix} \overline{X}(y_t^{\det},\varepsilon) & \overline{Z}(y_t^{\det},\varepsilon) \\ \overline{Z}(y_t^{\det},\varepsilon)^T & Y(t;Y_0) \end{pmatrix}, \tag{2.47}$$

and use $\overline{\mathcal{Z}}(t)^{-1}$ to characterize the ellipsoidal region in which $\zeta(t)$ is concentrated.

Theorem 2.5. *Assume that $\|\overline{X}(y_s^{\det},\varepsilon)\|$ and $\|\overline{X}(y_s^{\det},\varepsilon)^{-1})\|$ are uniformly bounded for $0 \leqslant s \leqslant t$ and that Y_0 has been chosen in such a way that $\|Y(s)^{-1}\| = \mathcal{O}(1/(\rho^2+\varepsilon))$ for $0 \leqslant s \leqslant t$. Fix an initial condition (x_0,y_0) with $y_0 \in \mathcal{D}_0$ and $x_0 = \bar{x}(y_0,\varepsilon)$, and let t be such that $y_s^{\det} \in \mathcal{D}_0$ for all $s \leqslant t$. Define*

$$R(t) = \|\overline{\mathcal{Z}}\|_{[0,t]}\Big[1 + \Big(1 + \|Y^{-1}\|_{[0,t]}^{1/2}\Big)\chi^{(1)}(t) + \chi^{(2)}(t)\Big]. \tag{2.48}$$

There exists a constant $h_0 > 0$, independent of y_0 and t, such that for all $h \leqslant h_0 R(t)^{-1}$,

$$\mathbb{P}^{0,(0,0)}\Big\{\sup_{0\leqslant s\leqslant t\wedge\tau_{\mathcal{D}_0}} \big\langle\zeta_u, \overline{\mathcal{Z}}(u)^{-1}\zeta_u\big\rangle \geqslant h^2\Big\} \leqslant \mathcal{C}_{n+m,\gamma,\Delta}(t,\varepsilon)\,\mathrm{e}^{-\kappa h^2/\sigma^2}, \tag{2.49}$$

with

$$\mathcal{C}_{n+m,\gamma,\Delta}(t,\varepsilon) = \Big\lceil\frac{t}{\Delta\varepsilon}\Big\rceil\Big[\Big(\frac{1}{1-2\gamma}\Big)^{(n+m)/2} + \mathrm{e}^{(n+m)/4}\Big], \tag{2.50}$$

$$\kappa = \gamma\Big[1 - \mathcal{O}\big(\varepsilon + \Delta + hR(t)\big)\Big]. \tag{2.51}$$

Let us first consider timescales of order 1. Then the functions $\|\overline{\mathcal{Z}}\|_{[0,t]}$, $\chi^{(1)}(t)$ and $\chi^{(2)}(t)$ are at most of order 1, and $\|Y(t)^{-1}\|$ remains of the same order as $\|Y_0^{-1}\|$. The probability (2.49) becomes small as soon as $h \gg \sigma$. Because of the restriction $h \leqslant h_0R(t)^{-1}$, the result is useful provided $\|Y^{-1}\|_{[0,t]} \ll \sigma^{-2}$. In order to obtain the optimal concentration result, we have to choose Y_0 according to two opposed criteria. On the one hand, we would like to choose Y_0 as small as possible, so that the set $\big\langle\zeta_u, \overline{\mathcal{Z}}(u)^{-1}\zeta_u\big\rangle < h^2$ is small. On the other hand, $\|Y_0^{-1}\|$ must not exceed certain bounds for Theorem 2.5 to be valid. Thus we require that

$$Y_0 > \big[\sigma^2 \vee (\rho^2+\varepsilon)\big]\,\mathbb{1}_m. \tag{2.52}$$

Because of the Gaussian decay of the probability (2.49) in σ/h, we can interpret the theorem by saying that the typical spreading of paths in the y-direction is of order $\sigma(\rho+\sqrt{\varepsilon})$ if $\sigma < \rho+\sqrt{\varepsilon}$ and of order σ^2 if $\sigma > \rho+\sqrt{\varepsilon}$.

The term ρ is clearly due to the intensity $\sigma' = \rho\sigma$ of the noise acting on the slow variable. It prevails if $\rho > \sigma \vee \sqrt{\varepsilon}$. The term $\sqrt{\varepsilon}$ is due to the linear part of the coupling between slow and fast variables, while the behaviour in σ^2 observed when $\sigma > \rho+\sqrt{\varepsilon}$ can be traced back to the *nonlinear* coupling between slow and fast variables.

For longer timescales, the condition $h \leqslant h_0R(t)^{-1}$ obliges us to take a larger Y_0, while $Y(t)$ typically grows with time. If the matrix B always has eigenvalues with positive real parts (or, more precisely, if the largest Lyapunov exponent is positive), this growth is exponential in time, so that the spreading of paths along the adiabatic manifold will reach order 1 in a time of order $\log|\sigma \vee (\rho^2+\varepsilon)|$.

Remark 2.6. Consider the *reduced stochastic system*

$$\mathrm{d}y_t^0 = g(\bar{x}(y_t^0,\varepsilon),y_t^0,\varepsilon)\,\mathrm{d}t + \sigma' G(\bar{x}(y_t^0,\varepsilon),y_t^0,\varepsilon)\,\mathrm{d}W_t \tag{2.53}$$

obtained by setting x equal to $\bar{x}(y,\varepsilon)$ in (2.8). One may wonder whether y_t^0 gives a better approximation of y_t than $y_t^{\det}$. In fact, one can show that

$$
\begin{aligned}
\mathbb{P}^{0,(0,0)}\Bigl\{\sup_{0\leqslant s\leqslant t\wedge\tau_{\mathcal{D}_0}} \|y_s^0 - y_s^{\det}\| \geqslant h_1\Bigr\} &\leqslant c\Bigl(1+t\Bigr)\,\mathrm{e}^{m/4}\exp\Bigl\{-\frac{\kappa_1 h_1^2(1-\mathcal{O}(h_1\chi^{(1)}(t)))}{(\sigma')^2\chi^{(2)}(t)}\Bigr\},\\
\mathbb{P}^{0,(0,0)}\Bigl\{\sup_{0\leqslant s\leqslant t\wedge\tau_{\mathcal{B}(h)}} \|y_s - y_s^{0}\| \geqslant h\Bigr\} &\leqslant c\Bigl(1+t\Bigr)\,\mathrm{e}^{m/4}\exp\Bigl\{-\frac{\kappa_1 h_1^2(1-\mathcal{O}(h_1\chi^{(1)}(t)))}{(\sigma')^2\chi^{(2)}(t)}\Bigr\}\\
&\quad + c\Bigl(1+\frac{t}{\varepsilon}\Bigr)\,\mathrm{e}^{m/4}\exp\Bigl\{-\frac{\kappa_2 h^2(1-\mathcal{O}((h+h_1)\chi^{(1)}(t)))}{[(\sigma')^2h^2+\sigma^2\varepsilon]\chi^{(2)}(t)}\Bigr\}
\end{aligned}
\tag{2.54}
$$

holds for all h, h_1 up to order $\chi^{(1)}(t)^{-1}$ and some positive constants c, κ_1, κ_2. (The proofs can be adapted from the proof of Lemma 3.4). This shows that the typical spreading of y_t^0 around $y_t^{\det}$ is of order $\sigma'\chi^{(2)}(t)^{1/2} = \rho\sigma\chi^{(2)}(t)^{1/2}$, while the typical deviation of paths y_t^0 of the reduced system from paths y_t of the original system is of order $\sigma\sqrt{\varepsilon}\chi^{(2)}(t)^{1/2}$. Thus for $\rho > \sqrt{\varepsilon}$, the reduced stochastic system gives a better approximation of the dynamics than the deterministic one.

If B has no eigenvalue with positive real part, the spreading of paths will grow more slowly. As an important particular case, let us consider the situation where $y_t^{\det}$ is an asymptotically stable periodic orbit with period T, entirely contained in $\mathcal{D}_0$ (and not too close to its boundary). Then all coefficients in (2.33) depend periodically on time, and, in particular, Floquet's theorem allows us to write

$$
V(t) = P(t)\,\mathrm{e}^{\Lambda t},
\tag{2.55}
$$

where $P(t)$ is a T-periodic matrix. The asymptotic stability of the orbit means that all eigenvalues but one of the monodromy matrix Λ have strictly negative real parts, the last eigenvalue, which corresponds to translations along the orbit, being 0. In that case, $\chi^{(1)}(t)$ and $\chi^{(2)}(t)$ grow only linearly with time, so that the spreading of paths in the y-direction remains small on timescales of order $1/(\sigma \vee (\rho^2+\varepsilon))$.

In fact, we even expect this spreading to occur mainly along the periodic orbit, while the paths remain confined to a neighbourhood of the orbit on subexponential timescales. To see that this is true, we can use a new set of variables in the neighbourhood of the orbit. In order not to introduce too many new notations, we will replace y by (y, z), where $y \in \mathbb{R}^{m-1}$ describes the degrees of freedom transversal to the orbit, and $z \in \mathbb{R}$ parametrizes the motion along the orbit. In fact, we can use an equal-time parametrization of the orbit, so that $\dot{z} = 1$ on the orbit, i.e., we have $z_t^{\det} = t \pmod T$. The SDE takes the form

$$
\begin{aligned}
\mathrm{d}x_t &= \frac{1}{\varepsilon} f(x_t, y_t, z_t, \varepsilon)\,\mathrm{d}t + \frac{\sigma}{\sqrt{\varepsilon}} F(x_t, y_t, z_t, \varepsilon)\,\mathrm{d}W_t,\\
\mathrm{d}y_t &= g(x_t, y_t, z_t, \varepsilon)\,\mathrm{d}t + \sigma' G(x_t, y_t, z_t, \varepsilon)\,\mathrm{d}W_t,\\
\mathrm{d}z_t &= \bigl[1 + h(x_t, y_t, z_t, \varepsilon)\bigr]\,\mathrm{d}t + \sigma' H(x_t, y_t, z_t, \varepsilon)\,\mathrm{d}W_t,
\end{aligned}
\tag{2.56}
$$

where $h = \mathcal{O}(\|y_t\|^2 + \|x_t - x_t^{\det}\|^2)$ and $\partial_y g(x_t^{\det}, 0, z_t^{\det}, \varepsilon)$ has eigenvalues with negative real parts, uniformly bounded away from zero. As linear approximation of the dynamics of $(\xi_t, \eta_t) = (x_t - x_t^{\rm det}, y_t - y_t^{\rm det}) = (x_t - x_t^{\rm det}, y_t)$ we take

$$\begin{aligned}
\mathrm{d}\xi_t^0 &= \frac{1}{\varepsilon} A(z_t^{\rm det}, \varepsilon)\xi_t^0 \,\mathrm{d}t + \frac{\sigma}{\sqrt{\varepsilon}} F_0(z_t^{\rm det}, \varepsilon)\,\mathrm{d}W_t, \\
\mathrm{d}\eta_t^0 &= \big[B(z_t^{\rm det}, \varepsilon)\eta_t^0 + C(z_t^{\rm det}, \varepsilon)\xi_t^0\big]\,\mathrm{d}t + \sigma' G_0(z_t^{\rm det}, \varepsilon)\,\mathrm{d}W_t, \\
\mathrm{d}z_t^0 &= \mathrm{d}t + \sigma' H_0(z_t^{\rm det}, \varepsilon)\,\mathrm{d}W_t,
\end{aligned} \tag{2.57}$$

which depends periodically on time. One can again compute the covariance matrix of the Gaussian process $(\xi_t^0, \eta_t^0, z_t^0)$ as a function of the principal solutions U and V associated with A and B. In particular, the covariance matrix $Y(t)$ of η_t^0 still obeys the ODE

$$\dot{Y} = B(y,\varepsilon)Y + YB(y,\varepsilon)^T + C(y,\varepsilon)\bar{Z} + \bar{Z}^T C(y,\varepsilon)^T + \rho^2 G_0(y,\varepsilon)G_0(y,\varepsilon)^T. \tag{2.58}$$

This is now a linear, inhomogeneous ODE with time-periodic coefficients. It is well known that such a system admits a unique periodic solution $Y_t^{\rm per}$, which is of order $\rho^2 + \varepsilon$ since $\bar{Z}$ is of order $\rho\sqrt{\varepsilon} + \varepsilon$ and $\rho^2 G_0 G_0^T$ is of order ρ^2. We can thus define an asymptotic covariance matrix $\bar{\mathcal{Z}}(t)$ of (ξ_t^0, η_t^0), which depends periodically on time. If $\zeta_t = (\xi_t, \eta_t)$, Theorem 2.5 shows that on timescales of order 1 (at least), the paths ζ_t are concentrated in a set of the form $\langle \zeta_t, \bar{\mathcal{Z}}(t)^{-1}\zeta_t\rangle < h^2$, while z_t remains h-close to $z_t^{\rm det}$.

On longer timescales, the distribution of paths will be smeared out along the periodic orbit. However, the same line of reasoning as in Section 3.2, based on a comparison with different deterministic solutions on successive time intervals of order 1, can be used to show that ζ_t remains concentrated in the set $\langle \zeta_t, \bar{\mathcal{Z}}(t)^{-1}\zeta_t\rangle < h^2$ up to exponentially long timescales.

2.4 Bifurcations

In the previous section, we have assumed that the slow manifold $x = x^\star(y)$ is uniformly asymptotically stable for $y \in \mathcal{D}_0$. We consider now the situation arising when the reduced deterministic flow causes $y_t^{\rm det}$ to leave $\mathcal{D}_0$, and to approach a bifurcation point of the slow manifold.

We call $(\hat{x}, \hat{y})$ a bifurcation point of the deterministic system

$$\begin{aligned}
\varepsilon\dot{x} &= f(x,y,\varepsilon), \\
\dot{y} &= g(x,y,\varepsilon),
\end{aligned} \tag{2.59}$$

if $f(\hat{x}, \hat{y}, 0) = 0$ and $\partial_x f(\hat{x}, \hat{y}, 0)$ has q eigenvalues on the imaginary axis, $q \in \{1, \dots, n\}$. We consider here the situation where $q < n$ and the other $n - q$ eigenvalues have strictly negative real parts.

The most generic cases are the saddle–node bifurcation (where $q = 1$), corresponding to a fold in the slow manifold, and the Hopf bifurcation (where $q = 2$), in which the slow manifold changes stability, while absorbing or expelling a family of periodic orbits. In these two cases, the set of bifurcation values $\hat{y}$ typically forms a codimension-1 submanifold of $\mathbb{R}^m$.

The dynamics of the deterministic slow–fast system (2.59) in a neighbourhood of the bifurcation point $(\hat{x}, \hat{y})$ can again be analyzed by a centre-manifold reduction. Introduce coordinates (x^-, z) in $\mathbb{R}^n$, with $x^- \in \mathbb{R}^{n-q}$ and $z \in \mathbb{R}^q$, in which the matrix $\partial_x f(\hat{x}, \hat{y}, 0)$ becomes block-diagonal, with a block $A^- \in \mathbb{R}^{(n-q)\times(n-q)}$ having eigenvalues in the left half-plane, and a block $A^0 \in \mathbb{R}^{q\times q}$ having eigenvalues on the imaginary axis. On the fast timescale t/ε, (2.59) can be rewritten as

$$
\begin{aligned}
(x^-)' &= f^-(x^-, z, y, \varepsilon), \\
z' &= f^0(x^-, z, y, \varepsilon), \\
y' &= \varepsilon g(x^-, z, y, \varepsilon), \\
\varepsilon' &= 0,
\end{aligned}
\tag{2.60}
$$

which admits $(\hat{x}^-, \hat{z}, \hat{y}, 0)$ as an equilibrium point. The linearization at this point has $q + m + 1$ eigenvalues on the imaginary axis (counting multiplicity), which correspond to the directions z, y and ε. In other words, z has become a slow variable near the bifurcation point.

The centre manifold theorem implies the existence, for sufficiently small ε and (z, y) in a neighbourhood $\mathcal{N}$ of $(\hat{z}, \hat{y})$, of a locally attracting invariant manifold $x^- = \bar{x}^-(z, y, \varepsilon)$, with $\bar{x}^-(\hat{z}, \hat{y}, 0) = \hat{x}$. $\bar{x}^-$ plays the same rôle the adiabatic manifold played in the stable case, and the dynamics on $\bar{x}^-$ is governed by the reduced equation

$$
\begin{aligned}
\varepsilon \dot{z} &= f^0(\bar{x}^-(z, y, \varepsilon), z, y, \varepsilon), \\
\dot{y} &= g(\bar{x}^-(z, y, \varepsilon), z, y, \varepsilon).
\end{aligned}
\tag{2.61}
$$

The function $\bar{x}^-(z, y, \varepsilon)$ solves the PDE

$$
\begin{aligned}
f^-(\bar{x}^-(z, y, \varepsilon), z, y, \varepsilon) = {}& \partial_z \bar{x}^-(z, y, \varepsilon) f^0(\bar{x}^-(z, y, \varepsilon), z, y, \varepsilon) \\
&+ \varepsilon \partial_y \bar{x}^-(z, y, \varepsilon) g(\bar{x}^-(z, y, \varepsilon), z, y, \varepsilon).
\end{aligned}
\tag{2.62}
$$

Let us now turn to random perturbations of the slow–fast system (1.1). In the variables (x^-, z, y), (2.1) can be written as

$$
\begin{aligned}
\mathrm{d}x_t^- &= \frac{1}{\varepsilon} f^-(x_t^-, z_t, y_t, \varepsilon)\,\mathrm{d}t + \frac{\sigma}{\sqrt{\varepsilon}} F^-(x_t^-, z_t, y_t, \varepsilon)\,\mathrm{d}W_t, \\
\mathrm{d}z_t &= \frac{1}{\varepsilon} f^0(x_t^-, z_t, y_t, \varepsilon)\,\mathrm{d}t + \frac{\sigma}{\sqrt{\varepsilon}} F^0(x_t^-, z_t, y_t, \varepsilon)\,\mathrm{d}W_t, \\
\mathrm{d}y_t &= g(x_t^-, z_t, y_t, \varepsilon)\,\mathrm{d}t + \sigma' G(x_t^-, z_t, y_t, \varepsilon)\,\mathrm{d}W_t.
\end{aligned}
\tag{2.63}
$$

The noise-induced deviation of x_t^- from the adiabatic manifold is described by the variable $\xi_t^- = x_t^- - \bar{x}^-(z_t, y_t, \varepsilon)$, which obeys an SDE of the form

$$
\mathrm{d}\xi_t^- = \frac{1}{\varepsilon} \hat{f}^-(\xi_t^-, z_t, y_t, \varepsilon)\,\mathrm{d}t + \frac{\sigma}{\sqrt{\varepsilon}} \widehat{F}^-(\xi_t^-, z_t, y_t, \varepsilon)\,\mathrm{d}W_t,
\tag{2.64}
$$

with, in particular,

$$
\begin{aligned}
\hat{f}^-(\xi^-, z, y, \varepsilon) = {}& f^-(\bar{x}^-(z, y, \varepsilon) + \xi^-, z, y, \varepsilon) \\
&- \partial_z \bar{x}^-(z, y, \varepsilon) f^0(\bar{x}^-(z, y, \varepsilon) + \xi^-, z, y, \varepsilon) \\
&- \varepsilon \partial_y \bar{x}^-(z, y, \varepsilon) g(\bar{x}^-(z, y, \varepsilon) + \xi^-, z, y, \varepsilon).
\end{aligned}
\tag{2.65}
$$

Note that (2.62) implies that $\hat{f}^-(0, z, y, \varepsilon) = 0$. We further define the matrix

$$
\begin{aligned}
A^-(z, y, \varepsilon) = \partial_\xi \hat{f}^-(0, z, y, \varepsilon) = {}& \partial_x f^-(\bar{x}^-(z, y, \varepsilon), z, y, \varepsilon) \\
&- \partial_z \bar{x}^-(z, y, \varepsilon) \partial_x f^0(\bar{x}^-(z, y, \varepsilon), z, y, \varepsilon) \\
&- \varepsilon \partial_y \bar{x}^-(z, y, \varepsilon) \partial_x g(\bar{x}^-(z, y, \varepsilon), z, y, \varepsilon).
\end{aligned}
\tag{2.66}
$$

Since $A^-(\hat{z},\hat{y},0) = A^-$, the eigenvalues of $A^-(z,y,\varepsilon)$ have uniformly negative real parts, provided we take the neighbourhood $\mathcal{N}$ and ε small enough.

Consider now the "linear approximation"

$$\begin{aligned}
\mathrm{d}\xi^0_t &= \frac{1}{\varepsilon} A^-(z^{\det}_t, y^{\det}_t, \varepsilon)\xi^0_t \,\mathrm{d}t + \frac{\sigma}{\sqrt{\varepsilon}} F^-_0(z^{\det}_t, y^{\det}_t, \varepsilon)\,\mathrm{d}W_t, \\
\mathrm{d}z^{\det}_t &= \frac{1}{\varepsilon} f^0(\bar{x}^-(z^{\det}, y^{\det}, \varepsilon), z^{\det}_t, y^{\det}_t, \varepsilon)\,\mathrm{d}t, \qquad (2.67)\\
\mathrm{d}y^{\det}_t &= g(\bar{x}^-(z^{\det}, y^{\det}, \varepsilon), z^{\det}_t, y^{\det}_t, \varepsilon)\,\mathrm{d}t
\end{aligned}$$

of (2.63)–(2.64), where $F^-_0(z,y,\varepsilon) = F^-(\bar{x}^-(z,y,\varepsilon),z,y,\varepsilon)$. Its solution ξ^0_t has a Gaussian distribution with covariance matrix

$$\operatorname{Cov}(\xi^0_t) = \frac{\sigma^2}{\varepsilon}\int_0^t U^-(t,s)F^-_0(z^{\det}_s, y^{\det}_s, \varepsilon)F^-_0(z^{\det}_s, y^{\det}_s, \varepsilon)^T U^-(t,s)^T\,\mathrm{d}s, \qquad (2.68)$$

where U^- is the fundamental solution of $\varepsilon\dot{\xi}^0 = A^-\xi^0$. Note that $\sigma^{-2}\operatorname{Cov}(\xi^0_t)$ is the X^--variable of a particular solution of the slow–fast system

$$\begin{aligned}
\varepsilon\dot{X}^- &= A^-(z,y,\varepsilon)X^- + X^-A^-(z,y,\varepsilon)^T + F^-_0(z,y,\varepsilon)F^-_0(z,y,\varepsilon)^T, \\
\varepsilon\dot{z} &= f^0(\bar{x}^-(z,y,\varepsilon),z,y,\varepsilon), \qquad (2.69)\\
\dot{y} &= g(\bar{x}^-(z,y,\varepsilon),z,y,\varepsilon),
\end{aligned}$$

which admits an invariant manifold $X^- = \overline{X}^-(z,y,\varepsilon)$ for $(z,y) \in \mathcal{N}$. We thus expect the paths to be concentrated in a set

$$\mathcal{B}^-(h) = \bigl\{(x^-,z,y)\colon (z,y)\in\mathcal{N}, \bigl\langle x^- - \bar{x}^-(z,y,\varepsilon), \overline{X}^-(z,y,\varepsilon)^{-1}(x^- - \bar{x}^-(z,y,\varepsilon))\bigr\rangle < h^2\bigr\}. \qquad (2.70)$$

The following theorem shows that this is indeed the case, as long as (z_t,y_t) remains in $\mathcal{N}$.

Theorem 2.7. *Assume that $\|\overline{X}^-(z,y,\varepsilon)\|$ and $\|\overline{X}^-(z,y,\varepsilon)^{-1})\|$ are uniformly bounded in $\mathcal{N}$. Choose a deterministic initial condition $(z_0,y_0)\in\mathcal{N}$, $x^-_0 = \bar{x}^-(z_0,y_0,\varepsilon)$, and let*

$$\tau_{\mathcal{N}} = \inf\{s>0\colon (z_s,y_s)\notin\mathcal{N}\}. \qquad (2.71)$$

Then there exist constants $h_0>0$, $\Delta_0>0$ and $\nu\in(0,1]$ such that for all $h\leqslant h_0$, all $\Delta\leqslant\Delta_0$ and all $0<\gamma<1/2$,

$$\mathbb{P}^{0,(x^-_0,z_0,y_0)}\{\tau_{\mathcal{B}^-(h)} < t\wedge\tau_{\mathcal{N}}\} \leqslant C_{n,m,q,\gamma,\Delta}(t,\varepsilon)\frac{h^2}{\sigma^2}\,\mathrm{e}^{-\kappa h^2/\sigma^2}, \qquad (2.72)$$

provided $\varepsilon|\log(h(1-2\gamma))|\leqslant 1$. Here

$$\begin{aligned}
\kappa &= \gamma\bigl[1-\mathcal{O}(\Delta)-\mathcal{O}(h^\nu(1-2\gamma)^{1-\nu}|\log(h(1-2\gamma))|^{1/2})\bigr], \qquad (2.73)\\
C_{n,m,q,\gamma,\Delta}(t,\varepsilon) &= \mathit{const}\,\frac{t}{\Delta\varepsilon}\Bigl(1+\frac{t}{\varepsilon}\Bigr)\Bigl[(1-2\gamma)^{-(n-q)} + \mathrm{e}^{(n-q)/4} + \mathrm{e}^{m/4} + \mathrm{e}^{q/4}\Bigr]. \qquad (2.74)
\end{aligned}$$

The exponent ν is related to the maximal rate of divergence of solutions of the reduced system (2.61), see Subsection 5.1.

This result shows that on timescales of order 1 (and larger if, e. g., $\mathcal{N}$ is positively invariant), paths are likely to remain in a small neighbourhood of the adiabatic manifold

$x = \bar{x}^-(z,y,\varepsilon)$. The dynamics will thus be essentially governed by the behaviour of the "slow" variables z and y.

In fact, it seems plausible that the dynamics of (2.63) will be well approximated by the dynamics of the *reduced stochastic system*

$$
\begin{aligned}
\mathrm{d}z_t^0 &= \frac{1}{\varepsilon} f^0(\bar{x}^-(z_t^0,y_t^0,\varepsilon),z_t^0,y_t^0,\varepsilon)\,\mathrm{d}t + \frac{\sigma}{\sqrt{\varepsilon}} F^0(\bar{x}^-(z_t^0,y_t^0,\varepsilon),z_t^0,y_t^0,\varepsilon)\,\mathrm{d}W_t, \\
\mathrm{d}y_t^0 &= g(\bar{x}^-(z_t^0,y_t^0,\varepsilon),z_t^0,y_t^0,\varepsilon)\,\mathrm{d}t + \sigma' G(\bar{x}^-(z_t^0,y_t^0,\varepsilon),z_t^0,y_t^0,\varepsilon)\,\mathrm{d}W_t,
\end{aligned}
\tag{2.75}
$$

obtained by setting x^- equal to $\bar{x}^-(z,y,\varepsilon)$ in (2.63). This turns out to be true under certain hypotheses on the solutions of (2.75). Let us fix an initial condition $(z_0^0,y_0^0) \in \mathcal{N}$, and call $\zeta_t^0 = (z_t^0,y_t^0)$ the corresponding process. We define the (random) matrices

$$
B(\zeta_t^0,\varepsilon) = \begin{pmatrix} \partial_z f^0 & \partial_y f^0 \\ \varepsilon\partial_z g & \varepsilon\partial_y g \end{pmatrix}\bigg|_{x=\bar{x}^-(z_t^0,y_t^0,\varepsilon),z=z_t^0,y=y_t^0}, \tag{2.76}
$$

$$
C(\zeta_t^0,\varepsilon) = \begin{pmatrix} \partial_x f^0 \\ \varepsilon\partial_x g \end{pmatrix}\bigg|_{x=\bar{x}^-(z_t^0,y_t^0,\varepsilon),z=z_t^0,y=y_t^0}. \tag{2.77}
$$

Observe that $C((\hat{z},\hat{y}),0) = 0$ because of our choice of coordinates, so that $\|C(\zeta_t^0,\varepsilon)\|$ will be small in a neighbourhood of the origin. We denote, for each realization $\zeta^0(\omega)$, by $\mathcal{V}_\omega$ the principal solution of

$$
\mathrm{d}\zeta_t(\omega) = \frac{1}{\varepsilon} B(\zeta_t^0(\omega),\varepsilon)\zeta_t(\omega)\,\mathrm{d}t. \tag{2.78}
$$

(Note that we may assume that almost all realizations $\zeta^0(\omega)$ are continuous.) We need to assume the existence of deterministic functions $\vartheta(t,s)$, $\vartheta_C(t,s)$, and a stopping time $\tau \leqslant \tau_{\mathcal{B}^-(h)}$ such that

$$
\big\|\mathcal{V}_\omega(t,s)\big\| \leqslant \vartheta(t,s), \qquad \big\|\mathcal{V}_\omega(t,s)C(\zeta_s^0(\omega),\varepsilon)\big\| \leqslant \vartheta_C(t,s) \tag{2.79}
$$

hold for all $s \leqslant t \leqslant \tau(\omega)$ and (almost) all paths $(\zeta_u^0(\omega))_{u\geqslant 0}$ of (2.75). Then we define

$$
\begin{aligned}
\chi^{(i)}(t) &= \sup_{0\leqslant s\leqslant t} \frac{1}{\varepsilon}\int_0^s \vartheta(s,u)^i\,\mathrm{d}u, \\
\chi_C^{(i)}(t) &= \sup_{0\leqslant s\leqslant t} \frac{1}{\varepsilon}\int_0^s \Big(\sup_{u\leqslant v\leqslant s} \vartheta_C(s,v)^i\Big)\,\mathrm{d}u
\end{aligned}
\tag{2.80}
$$

for $i = 1,2$, and the following result holds.

Theorem 2.8. *Assume that there exist constants $\Delta,\vartheta_0 > 0$ (of order 1) such that $\vartheta(s,u) \leqslant \vartheta_0$ and $\vartheta_C(s,u) \leqslant \vartheta_0$ whenever $0 < s-u \leqslant \Delta\varepsilon$. Then there exist constants $h_0,\kappa_0 > 0$ such that for all $h \leqslant h_0[\chi^{(1)}(t) \vee \chi_C^{(1)}(t)]^{-1}$ and all initial conditions $(x_0^-,z_0^0,y_0^0) \in \mathcal{B}^-(h)$,*

$$
\begin{aligned}
&\mathbb{P}^{0,(x_0^-,z_0^0,y_0^0)}\Big\{\sup_{0\leqslant s\leqslant t\wedge\tau} \big\|(z_s,y_s)-(z_s^0,y_s^0)\big\| \geqslant h\Big\} \\
&\qquad \leqslant \mathcal{C}_{m,q}(t,\varepsilon)\exp\Big\{-\kappa_0\frac{h^2}{\sigma^2}\frac{1}{\chi_C^{(2)}(t)+h^2\chi^{(2)}(t)}\Big\},
\end{aligned}
\tag{2.81}
$$

where

$$
\mathcal{C}_{m,q}(t,\varepsilon) = \mathit{const}\Big(1+\frac{t}{\varepsilon}\Big)\,\mathrm{e}^{(m+q)/4}. \tag{2.82}
$$

This result shows that typical solutions of the reduced system (2.75) approximate solutions of the initial system (2.63) to order $\sigma\chi_C^{(2)}(t)^{1/2}$, as long as $\chi^{(1)}(t) \ll 1/\sigma$. Checking the validity of Condition (2.79) for a reasonable stopping time τ is, of course, not straightforward, but it depends only on the dynamics of the reduced system, which is usually easier to analyze.

Example 2.9. Assume the reduced equation has the form

$$
\begin{aligned}
\mathrm{d}z_t^0 &= \frac{1}{\varepsilon}\bigl[y_t^0 z_t^0 - (z_t^0)^3\bigr]\,\mathrm{d}t + \frac{\sigma}{\sqrt{\varepsilon}}\,\mathrm{d}W_t, \\
\mathrm{d}y_t^0 &= 1,
\end{aligned}
\tag{2.83}
$$

i. e., there is a pitchfork bifurcation at the origin. We choose an initial condition (z_0, y_0) with $y_0 < 0$ at time $t_0 = y_0$, to that $y_t = t$. In [7] we proved that if $\sigma \leqslant \sqrt{\varepsilon}$, the paths $\{z_s\}_{s\geqslant t_0}$ are concentrated in strip of width of order $\sigma/(|y|^{1/2} \vee \varepsilon^{1/4})$ up to time $\sqrt{\varepsilon}$.

Using for τ the first-exit time from a set of this form, one finds that $\chi_C^{(2)}(\sqrt{\varepsilon})$ is of order $\sqrt{\varepsilon} + \sigma^2/\varepsilon$. Thus the typical spreading of z_s around reduced solutions z_s^0 is of order $\sigma\varepsilon^{1/4} + \sigma^2/\sqrt{\varepsilon}$, which is smaller than the spreading of z_s^0 around a deterministic solution. Hence the reduced system provides a good approximation to the full system up to time $\sqrt{\varepsilon}$.

For larger times, however, $\chi_C^{(2)}(\sqrt{\varepsilon})$ grows like $\mathrm{e}^{t^2/\varepsilon}$ until the paths leave a neighbourhood of the unstable equilibrium $z = 0$, which typically occurs at a time of order $\sqrt{\varepsilon|\log\sigma|}$. Thus the spreading is too fast for the reduced system to provide a good approximation to the dynamics. This shows that Theorem 2.8 is not quite sufficient to reduce the problem to a one-dimensional one, and a more detailed description has to be used for the region of instability.

3 Proofs – Exit from $\mathcal{B}(h)$

In this section, we consider the SDE

$$
\begin{aligned}
\mathrm{d}x_t &= \frac{1}{\varepsilon} f(x_t, y_t, \varepsilon)\,\mathrm{d}t + \frac{\sigma}{\sqrt{\varepsilon}} F(x_t, y_t, \varepsilon)\,\mathrm{d}W_t, \\
\mathrm{d}y_t &= g(x_t, y_t, \varepsilon)\,\mathrm{d}t + \sigma' G(x_t, y_t, \varepsilon)\,\mathrm{d}W_t
\end{aligned}
\tag{3.1}
$$

under Assumption 2.2, that is, when starting near a uniformly asymptotically stable manifold. We denote by $(x_t^{\mathrm{det}}, y_t^{\mathrm{det}})$, with $x_t^{\mathrm{det}} = \bar{x}(y_t^{\mathrm{det}}, \varepsilon)$, the deterministic solution starting in $y_0^{\mathrm{det}} = y_0 \in \mathcal{D}_0$.

The transformation

$$
\begin{aligned}
x_t &= \bar{x}(y_t^{\mathrm{det}} + \eta_t, \varepsilon) + \xi_t, \\
y_t &= y_t^{\mathrm{det}} + \eta_t
\end{aligned}
\tag{3.2}
$$

yields a system of the form (2.10), which can be written, using Taylor expansions, as

$$
\begin{aligned}
\mathrm{d}\xi_t &= \frac{1}{\varepsilon}\bigl[A(y_t^{\mathrm{det}}, \varepsilon)\xi_t + b(\xi_t, \eta_t, t, \varepsilon)\bigr]\,\mathrm{d}t + \frac{\sigma}{\sqrt{\varepsilon}}\bigl[F_0(y_t^{\mathrm{det}}, \varepsilon) + F_1(\xi_t, \eta_t, t, \varepsilon)\bigr]\,\mathrm{d}W_t, \qquad (3.3) \\
\mathrm{d}\eta_t &= \bigl[C(y_t^{\mathrm{det}}, \varepsilon)\xi_t + B(y_t^{\mathrm{det}}, \varepsilon)\eta_t + c(\xi_t, \eta_t, t, \varepsilon)\bigr]\,\mathrm{d}t + \sigma'\bigl[G_0(y_t^{\mathrm{det}}, \varepsilon) + G_1(\xi_t, \eta_t, t, \varepsilon)\bigr]\,\mathrm{d}W_t.
\end{aligned}
$$

There are constants M, M_1 such that the remainder terms satisfy the bounds

$$
\begin{aligned}
\|b(\xi,\eta,t,\varepsilon)\| &\leqslant M\bigl(\|\xi\|^2+\|\xi\|\|\eta\|\bigr), \\
\|c(\xi,\eta,t,\varepsilon)\| &\leqslant M\bigl(\|\xi\|^2+\|\eta\|^2\bigr), \\
\|F_1(\xi,\eta,t,\varepsilon)\| &\leqslant M_1\bigl(\|\xi\|+\|\eta\|\bigr), \\
\|G_1(\xi,\eta,t,\varepsilon)\| &\leqslant M_1\bigl(\|\xi\|+\|\eta\|\bigr)
\end{aligned}
\tag{3.4}
$$

for all (ξ,η) in a compact set and all t such that $y_t^{\det}\in\mathcal{D}_0$.

3.1 Timescales of order 1

We first examine the behaviour of ξ_u on an interval $[s,t]$ with $\Delta=(t-s)/\varepsilon=\mathcal{O}_\varepsilon(1)$. For this purpose, we fix an initial condition $y_0\in\mathcal{D}_0$ and assume that t is chosen in such a way that $y_u^{\det}\in\mathcal{D}_0$ for all $u\leqslant t$.

To ease notations, we will not indicate the ε-dependence of $\overline{X}(y)$. We assume that $\|\overline{X}(y)\|\leqslant K_+$ and $\|\overline{X}(y)^{-1}\|\leqslant K_-$ for all $y\in\mathcal{D}_0$, and define the functions

$$
\begin{aligned}
\Psi(t) &= \frac{1}{\varepsilon}\int_0^t \bigl\|U(t,u)^T\overline{X}(y_t^{\det})^{-1}U(t,u)\bigr\|\,\mathrm{d}u, \\
\Phi(t) &= \frac{1}{\varepsilon}\int_0^t \operatorname{Tr}\bigl[U(t,u)^T\overline{X}(y_t^{\det})^{-1}U(t,u)\bigr]\,\mathrm{d}u, \\
\Theta(t) &= \frac{1}{\varepsilon}\int_0^t \|U(t,u)\|\,\mathrm{d}u,
\end{aligned}
\tag{3.5}
$$

where $U(t,u)$ again denotes the principal solution of $\varepsilon\dot{\xi}=A(y_t^{\det},\varepsilon)\xi$. Note that the stability of the adiabatic manifold implies that $\|U(t,u)\|$ is bounded by a constant times $\exp\{-K_0(t-u)/\varepsilon\}$, $K_0>0$, for all t and $u\leqslant t$. Hence $\Psi(t)$ and $\Theta(t)$ are of order 1, while $\Phi(t)$ is of order n. In particular, $\Phi(t)\leqslant n\Psi(t)$ holds for all times t.

We first concentrate on upper estimates on the probabilities and will deal with the lower bound in Corollary 3.5. Let us remark that on timescales of order 1, we may safely assume that the deviation η_s of y_s from its deterministic counterpart remains small. We fix a deterministic $h_1>0$ and define

$$
\tau_\eta=\inf\bigl\{s>0\colon \|\eta_s\|\geqslant h_1\bigr\}. \tag{3.6}
$$

Lemma 3.4 below provides an estimate on the tails of the distribution of τ_η. The following proposition estimates the probability that x_t leaves a "layer" similar to $\mathcal{B}(h)$ during the time interval $[s,t]$ despite of η_u remaining small. Note that in the proposition the "thickness of the layer" is measured at $y_u^{\det}$ instead of y_u.

Proposition 3.1. *For all $\alpha\in[0,1)$, all $\gamma\in(0,1/2)$ and all $\mu>0$,*

$$
\begin{aligned}
\sup_{\xi_0\,:\,\langle\xi_0,\overline{X}(y_0)^{-1}\xi_0\rangle\leqslant\alpha^2h^2} &\mathbb{P}^{0,(\xi_0,0)}\Bigl\{\sup_{s\leqslant u\leqslant t\wedge\tau_\eta}\bigl\langle\xi_u,\overline{X}(y_u^{\det})^{-1}\xi_u\bigr\rangle\geqslant h^2\Bigr\} \\
&\leqslant \frac{1}{(1-2\gamma)^{n/2}}\exp\Bigl\{-\gamma\frac{h^2}{\sigma^2}\Bigl[1-\alpha^2-M_0\bigl(\Delta+\mu h+(h+h_1)\Theta(t)\bigr)\Bigr]\Bigr\} \\
&\quad+\mathrm{e}^{\Phi(t)/4\Psi(t)}\exp\Bigl\{-\frac{h^2}{\sigma^2}\frac{\mu^2(1-M_0\Delta)}{8M_1^2(\sqrt{K_+}+h_1/h)^2\Psi(t)}\Bigr\}
\end{aligned}
\tag{3.7}
$$

holds for all $h<1/\mu$, with a constant M_0 depending only on the linearization A of f, K_+, K_-, M and $\|F_0\|_\infty$.

PROOF: The solution of (3.3) can be written as

$$\begin{aligned}\xi_u = U(u)\xi_0 &+ \frac{\sigma}{\sqrt{\varepsilon}}\int_0^u U(u,v)F_0(y_v^{\det},\varepsilon)\,\mathrm{d}W_v \\ &+ \frac{\sigma}{\sqrt{\varepsilon}}\int_0^u U(u,v)F_1(\xi_v,\eta_v,v,\varepsilon)\,\mathrm{d}W_v + \frac{1}{\varepsilon}\int_0^u U(u,v)b(\xi_v,\eta_v,v,\varepsilon)\,\mathrm{d}v.\end{aligned} \tag{3.8}$$

Writing $\xi_u = U(u,s)\Upsilon_u$ and defining

$$\tau_\xi = \inf\{u \geqslant 0 \colon \langle \xi_u, \overline{X}(y_u^{\det})^{-1}\xi_u\rangle \geqslant h^2\}, \tag{3.9}$$

the probability on the left-hand side of (3.7) can be rewritten as

$$P = \mathbb{P}^{0,(\xi_0,0)}\biggl\{\sup_{s\leqslant u\leqslant t\wedge\tau_\xi\wedge\tau_\eta} \|Q(u)\Upsilon_u\| \geqslant h\biggr\}, \tag{3.10}$$

where $Q(u) = Q_s(u)$ is the symmetric matrix defined by

$$Q(u)^2 = U(u,s)^T\overline{X}(y_u^{\det})^{-1}U(u,s). \tag{3.11}$$

To eliminate the u-dependence of Q in (3.10), we estimate P by

$$P \leqslant \mathbb{P}^{0,(\xi_0,0)}\biggl\{\sup_{s\leqslant u\leqslant t\wedge\tau_\xi\wedge\tau_\eta} \|Q(t)\Upsilon_u\| \geqslant H\biggr\}, \tag{3.12}$$

where

$$H = h\biggl(\sup_{s\leqslant u\leqslant t}\bigl\|Q(u)Q(t)^{-1}\bigr\|\biggr)^{-1}. \tag{3.13}$$

In order to estimate the supremum in (3.13), we use the fact that $Q(v)^{-2}$ satisfies the differential equation

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}v}Q(v)^{-2} &= \frac{1}{\varepsilon}U(s,v)\Bigl[-A(y_v^{\det})\overline{X}(y_v^{\det}) - \overline{X}(y_v^{\det})A(y_v^{\det})^T + \varepsilon\frac{\mathrm{d}}{\mathrm{d}v}\overline{X}(y_v^{\det})\Bigr]U(s,v)^T \\ &= \frac{1}{\varepsilon}U(s,v)F_0(y_v^{\det},\varepsilon)F_0(y_v^{\det},\varepsilon)^TU(s,v)^T,\end{aligned} \tag{3.14}$$

and thus

$$Q(u)^2Q(t)^{-2} = 1\!\!1 + Q(u)^2\frac{1}{\varepsilon}\int_u^t U(s,v)F_0(y_v^{\det},\varepsilon)F_0(y_v^{\det},\varepsilon)^TU(s,v)^T\,\mathrm{d}v = 1\!\!1 + \mathcal{O}(\Delta). \tag{3.15}$$

(Recall that $t-u \leqslant t-s \leqslant \varepsilon\Delta$ in this subsection, which implies $\|U(s,v)\| = 1 + \mathcal{O}(\Delta)$ and $\|Q(u)^2\| \leqslant K_-(1+\mathcal{O}(\Delta))$.) Therefore, $H = h(1-\mathcal{O}(\Delta))$.

We now split Υ_u into three parts, writing $\Upsilon_u = \Upsilon_u^0 + \Upsilon_u^1 + \Upsilon_u^2$, where

$$\begin{aligned}\Upsilon_u^0 &= U(s)\xi_0 + \frac{\sigma}{\sqrt{\varepsilon}}\int_0^u U(s,v)F_0(y_v^{\det},\varepsilon)\,\mathrm{d}W_v, \\ \Upsilon_u^1 &= \frac{\sigma}{\sqrt{\varepsilon}}\int_0^u U(s,v)F_1(\xi_v,\eta_v,v,\varepsilon)\,\mathrm{d}W_v, \\ \Upsilon_u^2 &= \frac{1}{\varepsilon}\int_0^u U(s,v)b(\xi_v,\eta_v,v,\varepsilon)\,\mathrm{d}v,\end{aligned} \tag{3.16}$$

and estimate P by the sum of the corresponding probabilities

$$\begin{aligned} P_0 &= \mathbb{P}^{0,(\xi_0,0)}\Big\{\sup_{s\leqslant u\leqslant t}\|Q(t)\Upsilon^0_u\|\geqslant H_0\Big\}, \\ P_1 &= \mathbb{P}^{0,(\xi_0,0)}\Big\{\sup_{s\leqslant u\leqslant t\wedge\tau_\xi\wedge\tau_\eta}\|Q(t)\Upsilon^1_u\|\geqslant H_1\Big\}, \\ P_2 &= \mathbb{P}^{0,(\xi_0,0)}\Big\{\sup_{s\leqslant u\leqslant t\wedge\tau_\xi\wedge\tau_\eta}\|Q(t)\Upsilon^2_u\|\geqslant H_2\Big\}, \end{aligned} \tag{3.17}$$

where H_0, H_1, H_2 satisfy $H_0 + H_1 + H_2 = H$. Note that P_2 can be estimated trivially using the fact that

$$\sup_{s\leqslant u\leqslant t\wedge\tau_\xi\wedge\tau_\eta}\|Q(t)\Upsilon^2_u\| \leqslant \sqrt{K_-}M(K_+h^2+\sqrt{K_+}hh_1)(1+\mathcal{O}(\Delta))\Theta(t) := \overline{H}_2. \tag{3.18}$$

Now, we choose

$$\begin{aligned} H_2 &= 2\overline{H}_2, \\ H_1 &= \mu hH, \\ H_0 &= H - H_1 - H_2 \end{aligned} \tag{3.19}$$

for $0 < \mu < 1/h$, and estimate the remaining probabilities P_0 and P_1 by Lemmas 3.2 and 3.3 below which completes the proof. $\square$

Lemma 3.2. *Under the hypotheses of Proposition 3.1, we have for every* $\gamma \in (0, 1/2)$,

$$P_0 = \mathbb{P}^{0,(\xi_0,0)}\Big\{\sup_{s\leqslant u\leqslant t}\|Q(t)\Upsilon^0_u\|\geqslant H_0\Big\} \leqslant \frac{1}{(1-2\gamma)^{n/2}}\exp\Big\{-\gamma\frac{H_0^2-\alpha^2h^2}{\sigma^2}\Big\}, \tag{3.20}$$

holding uniformly for all ξ_0 *such that* $\langle\xi_0, \overline{X}(y_0)^{-1}\xi_0\rangle \leqslant \alpha^2h^2$.

PROOF: For every $\widehat{\gamma} > 0$, $(\exp\{\widehat{\gamma}\|Q(t)\Upsilon^0_u\|^2\})_{u\geqslant s}$ is a positive submartingale and, therefore, Doob's submartingale inequality yields

$$P_0 = \mathbb{P}^{0,(\xi_0,0)}\Big\{\sup_{s\leqslant u\leqslant t}\mathrm{e}^{\widehat{\gamma}\|Q(t)\Upsilon^0_u\|^2}\geqslant \mathrm{e}^{\widehat{\gamma}H_0^2}\Big\} \leqslant \mathrm{e}^{-\widehat{\gamma}H_0^2}\,\mathbb{E}^{0,(\xi_0,0)}\big\{\mathrm{e}^{\widehat{\gamma}\|Q(t)\Upsilon^0_t\|^2}\big\}. \tag{3.21}$$

Now, the random variable $Q(t)\Upsilon^0_t$ is Gaussian, with expectation $E = Q(t)U(s)\xi_0$ and covariance matrix

$$\Sigma = \frac{\sigma^2}{\varepsilon}Q(t)\bigg(\int_0^t U(s,v)F_0(y^{\det}_v,\varepsilon)F_0(y^{\det}_v,\varepsilon)^TU(s,v)^T\,\mathrm{d}v\bigg)Q(t)^T. \tag{3.22}$$

Thus, using completion of squares to compute the Gaussian integral, we find

$$\mathbb{E}^{0,(\xi_0,0)}\big\{\mathrm{e}^{\widehat{\gamma}\|Q(t)\Upsilon^0_t\|^2}\big\} = \frac{\mathrm{e}^{\widehat{\gamma}\langle E,(1\!\!1-2\widehat{\gamma}\Sigma)^{-1}E\rangle}}{(\det[1\!\!1-2\widehat{\gamma}\Sigma])^{1/2}}. \tag{3.23}$$

By (2.19), we can write

$$\Sigma = \sigma^2Q(t)U(s,t)\big[\overline{X}(y^{\det}_t)-U(t)\overline{X}(y^{\det}_0)U(t)^T\big]U(s,t)^TQ(t)^T = \sigma^2\big[1\!\!1-RR^T\big], \tag{3.24}$$

where $R = Q(t)U(s)\overline{X}(y_0^{\rm det})^{1/2}$, and we have used the fact that $U(s,t)\overline{X}(y_t^{\rm det})U(s,t)^T = Q(t)^{-2}$. This shows in particular that

$$\det[\mathbb{1} - 2\widehat{\gamma}\Sigma] \geqslant (1-2\widehat{\gamma}\sigma^2)^n. \tag{3.25}$$

Moreover, since $\|RR^T\| = \|R^TR\| \in (0,1)$, we also have

$$\begin{aligned}\langle E, (\mathbb{1}-2\widehat{\gamma}\Sigma)^{-1}E\rangle &= \langle \overline{X}(y_0^{\rm det})^{-1/2}\xi_0, R^T(\mathbb{1}-2\widehat{\gamma}\Sigma)^{-1}R\overline{X}(y_0^{\rm det})^{-1/2}\xi_0\rangle \\ &\leqslant \alpha^2h^2\big\|R^T\big(\mathbb{1}-2\widehat{\gamma}\sigma^2[\mathbb{1}-RR^T]\big)^{-1}R\big\| \\ &\leqslant \alpha^2h^2\big([1-2\widehat{\gamma}\sigma^2]\|R^TR\|^{-1}+2\widehat{\gamma}\sigma^2\big)^{-1} \leqslant \alpha^2h^2\end{aligned} \tag{3.26}$$

for all ξ_0 satisfying $\langle \xi_0, \overline{X}(y_0)^{-1}\xi_0\rangle \leqslant \alpha^2h^2$. Now, (3.20) follows from (3.23) by choosing $\widehat{\gamma} = \gamma/\sigma^2$. □

Lemma 3.3. *Under the hypotheses of Proposition 3.1,*

$$P_1 = \mathbb{P}^{0,(\xi_0,0)}\Big\{\sup_{s\leqslant u\leqslant t\wedge\tau_\xi\wedge\tau_\eta}\|Q(t)\Upsilon_u^1\| \geqslant H_1\Big\} \leqslant \exp\bigg\{-\frac{\big(H_1^2-\sigma^2M_1^2(\sqrt{K_+}h+h_1)^2\Phi(t)\big)^2}{8\sigma^2M_1^2(\sqrt{K_+}h+h_1)^2H_1^2\Psi(t)}\bigg\} \tag{3.27}$$

holds uniformly for all ξ_0 such that $\langle\xi_0,\overline{X}(y_0)^{-1}\xi_0\rangle \leqslant h^2$.

Proof: Let τ denote the stopping time

$$\tau = \tau_\xi\wedge\tau_\eta\wedge\inf\{u\geqslant 0\colon \|Q(t)\Upsilon_u^1\|\geqslant H_1\}, \tag{3.28}$$

and define, for a given γ_1, the stochastic process

$$\Xi_u = \mathrm{e}^{\gamma_1\|Q(t)\Upsilon_u^1\|^2}. \tag{3.29}$$

$(\Xi_u)_u$ being a positive submartingale, another application of Doob's submartingale inequality yields

$$P_1 \leqslant \mathrm{e}^{-\gamma_1H_1^2}\,\mathbb{E}^{0,(\xi_0,0)}\big\{\Xi_{t\wedge\tau}\big\}. \tag{3.30}$$

Itô's formula (together with the fact that $(\mathrm{d}W_u)^TR^TR\,\mathrm{d}W_u = \mathrm{Tr}(R^TR)\,\mathrm{d}u$ for any matrix $R\in\mathbb{R}^{n\times k}$) shows that Ξ_u obeys the SDE

$$\mathrm{d}\Xi_u = 2\gamma_1\frac{\sigma}{\sqrt{\varepsilon}}\Xi_u(\Upsilon_u^1)^TQ(t)^2U(s,u)F_1(\xi_u,\eta_u,u,\varepsilon)\,\mathrm{d}W_u + \gamma_1\frac{\sigma^2}{\varepsilon}\Xi_u\,\mathrm{Tr}\big[R_1^TR_1+2\gamma_1R_2^TR_2\big]\,\mathrm{d}u, \tag{3.31}$$

where

$$\begin{aligned}R_1 &= Q(t)U(s,u)F_1(\xi_u,\eta_u,u,\varepsilon),\\ R_2 &= (\Upsilon_u^1)^TQ(t)^2U(s,u)F_1(\xi_u,\eta_u,u,\varepsilon).\end{aligned} \tag{3.32}$$

The first term in the trace can be estimated as

$$\begin{aligned}\mathrm{Tr}\big[R_1^TR_1\big] = \mathrm{Tr}\big[R_1R_1^T\big] &\leqslant M_1^2\big(\|\xi_u\|+\|\eta_u\|\big)^2\,\mathrm{Tr}\big[Q(t)^TU(s,u)U(s,u)^TQ(t)\big]\\ &\leqslant M_1^2\big(\|\xi_u\|+\|\eta_u\|\big)^2\,\mathrm{Tr}\big[U(t,u)^T\overline{X}(y_t^{\rm det})^{-1}U(t,u)\big],\end{aligned} \tag{3.33}$$

while the second term satisfies the bound

$$\begin{aligned}\operatorname{Tr}\bigl[R_2^T R_2\bigr] &= \bigl\|F_1(\xi_u,\eta_u,u,\varepsilon)^T U(s,u)^T Q(t)^2 \Upsilon_u^1\bigr\|^2 \\ &\leqslant M_1^2\bigl(\|\xi_u\|+\|\eta_u\|\bigr)^2 \bigl\|U(s,u)^T Q(t)\bigr\|^2 \|Q(t)\Upsilon_u^1\|^2 \\ &= M_1^2\bigl(\|\xi_u\|+\|\eta_u\|\bigr)^2 \bigl\|U(t,u)^T \overline{X}(y_t^{\det})^{-1} U(t,u)\bigr\| \|Q(t)\Upsilon_u^1\|^2. \end{aligned} \tag{3.34}$$

Using the fact that $\|\xi_u\| \leqslant \sqrt{K_+}h$, $\|\eta_u\| \leqslant h_1$ and $\|Q(t)\Upsilon_u^1\| \leqslant H_1$ hold for all $0 \leqslant u \leqslant t\wedge\tau$, we obtain

$$\begin{aligned}&\mathbb{E}^{0,(\xi_0,0)}\bigl\{\Xi_{u\wedge\tau}\bigr\} \\ &\quad\leqslant 1+\gamma_1\frac{\sigma^2}{\varepsilon} M_1^2\bigl(\sqrt{K_+}h+h_1\bigr)^2 \int_0^u \mathbb{E}^{0,(\xi_0,0)}\bigl\{\Xi_{v\wedge\tau}\bigr\} \\ &\qquad\times\Bigl[\operatorname{Tr}\bigl[U(t,v)^T\overline{X}(y_t^{\det})^{-1}U(t,v)\bigr] + 2\gamma_1 H_1^2 \bigl\|U(t,v)^T\overline{X}(y_t^{\det})^{-1}U(t,v)\bigr\|\Bigr]\,\mathrm{d}v, \end{aligned} \tag{3.35}$$

and Gronwall's inequality yields

$$\mathbb{E}^{0,(\xi_0,0)}\bigl\{\Xi_{t\wedge\tau}\bigr\} \leqslant \exp\Bigl\{\gamma_1\sigma^2 M_1^2\bigl(\sqrt{K_+}h+h_1\bigr)^2\bigl[\Phi(t)+2\gamma_1 H_1^2\Psi(t)\bigr]\Bigr\}. \tag{3.36}$$

Now, (3.30) implies

$$P_1 \leqslant \exp\Bigl\{-\gamma_1\bigl(H_1^2-\sigma^2M_1^2(\sqrt{K_+}h+h_1)^2\Phi(t)\bigr)+2\gamma_1^2\sigma^2M_1^2(\sqrt{K_+}h+h_1)^2H_1^2\Psi(t)\Bigr\}, \tag{3.37}$$

and (3.27) follows by optimizing over γ_1. □

Proposition 3.1 allows to control the first-exit time of (x_t, y_t) from $\mathcal{B}(h)$, provided $\eta_s = y_s - y_s^{\det}$ remains small. In order to complete the proof of Theorem 2.4 we need to control the tails of the distribution of τ_η. The following lemma provides a rough *a priori* estimate which is sufficient for the time being. We will provide more precise estimates in the next section.

Recall the notations $V(u,v)$ for the principal solution of $\dot\eta = B(y_u^{\det},\varepsilon)\eta$, and

$$\chi^{(1)}(t) = \sup_{0\leqslant s\leqslant t}\int_0^s\Bigl(\sup_{u\leqslant v\leqslant s}\|V(s,v)\|\Bigr)\,\mathrm{d}u, \tag{3.38}$$

$$\chi^{(2)}(t) = \sup_{0\leqslant s\leqslant t}\int_0^s\Bigl(\sup_{u\leqslant v\leqslant s}\|V(s,v)\|^2\Bigr)\,\mathrm{d}u. \tag{3.39}$$

from Subsection 2.3.

Lemma 3.4. *There exists a constant $c_\eta > 0$ such that for all choices of $t > 0$ and $h_1 > 0$ satisfying $y_s^{\det} \in \mathcal{D}_0$ for all $s \leqslant t$ and $h_1 \leqslant c_\eta\chi^{(1)}(t)^{-1}$,*

$$\begin{aligned}&\sup_{\xi_0\,:\,\langle\xi_0,\overline{X}(y_0)^{-1}\xi_0\rangle\leqslant h^2}\mathbb{P}^{0,(\xi_0,0)}\Bigl\{\sup_{0\leqslant u\leqslant t\wedge\tau_{\mathcal{B}(h)}}\|\eta_u\|\geqslant h_1\Bigr\} \\ &\quad\leqslant 2\Bigl\lceil\frac{t}{\Delta\varepsilon}\Bigr\rceil\,\mathrm{e}^{m/4}\exp\Bigl\{-\kappa_0\frac{h_1^2(1-\mathcal{O}(\Delta\varepsilon))}{\sigma^2(\rho^2+\varepsilon)\chi^{(2)}(t)}\Bigl[1-M_0'\,\chi^{(1)}(t)\,h_1\Bigl(1+K_+\frac{h^2}{h_1^2}\Bigr)\Bigr]\Bigr\}, \end{aligned} \tag{3.40}$$

where κ_0, M_0' are constants depending only on $\|\widehat{F}\|_\infty$, $\|\widehat{G}\|_\infty$, M, $\|C\|_\infty$ and U.

Proof: We first consider a time interval $[s,t]$ with $t-s=\Delta\varepsilon$. Let $u\in[s,t]$ and recall the defining SDE (3.3) for η_u. Its solution can be split into four parts, $\eta_u=\eta_u^0+\eta_u^1+\eta_u^2+\eta_u^3$, where

$$\begin{aligned}
\eta_u^0 &= \sigma' \int_0^u V(u,v)\widehat{G}(\xi_v,\eta_v,v,\varepsilon)\,\mathrm{d}W_v,\\
\eta_u^1 &= \frac{\sigma}{\sqrt{\varepsilon}} \int_0^u S(u,v)\widehat{F}(\xi_v,\eta_v,v,\varepsilon)\,\mathrm{d}W_v,\\
\eta_u^2 &= \int_0^u V(u,v)c(\xi_v,\eta_v,v,\varepsilon)\,\mathrm{d}v,\\
\eta_u^3 &= \frac{1}{\varepsilon}\int_0^u S(u,v)b(\xi_v,\eta_v,v,\varepsilon)\,\mathrm{d}v,
\end{aligned}\tag{3.41}$$

with

$$S(u,v)=\int_v^u V(u,w)C(y_w^{\det},\varepsilon)U(w,v)\,\mathrm{d}w. \tag{3.42}$$

Let $\tau=\tau_{\mathcal{B}(h)}\wedge\tau_\eta$. It follows immediately from the definitions of $\tau_{\mathcal{B}(h)}$, τ_η and the bounds (3.4) that

$$\begin{aligned}
\bigl\|\eta_{u\wedge\tau}^2\bigr\| &\leqslant M(1+\mathcal{O}(\Delta\varepsilon))\chi^{(1)}(t)(K_+h^2+h_1^2),\\
\bigl\|\eta_{u\wedge\tau}^3\bigr\| &\leqslant M'\chi^{(1)}(t)(K_+h^2+\sqrt{K_+}hh_1)
\end{aligned}\tag{3.43}$$

for all $u\in[s,t]$. Here M' depends only on M, U and $\|C\|_\infty$. Furthermore, using similar ideas as in the proof of Lemma 3.3, it is straightforward to establish for all $H_0,H_1>0$ that

$$\begin{aligned}
\mathbb{P}^{0,(\xi_0,0)}\Bigl\{\sup_{s\leqslant u\leqslant t\wedge\tau}\|\eta_u^0\|\geqslant H_0\Bigr\} &\leqslant \mathrm{e}^{m/4}\exp\Bigl\{-\frac{H_0^2(1-\mathcal{O}(\Delta\varepsilon))}{8(\sigma')^2\|\widehat{G}\|_\infty^2\chi^{(2)}(t)}\Bigr\},\\
\mathbb{P}^{0,(\xi_0,0)}\Bigl\{\sup_{s\leqslant u\leqslant t\wedge\tau}\|\eta_u^1\|\geqslant H_1\Bigr\} &\leqslant \mathrm{e}^{m/4}\exp\Bigl\{-\frac{H_1^2(1-\mathcal{O}(\Delta\varepsilon))}{8\sigma^2\varepsilon c_S\|\widehat{F}\|_\infty^2\chi^{(2)}(t)}\Bigr\},
\end{aligned}\tag{3.44}$$

where c_S is a constant depending only on S. Then the local analogue of estimate (3.40) (without the t-dependent prefactor) is obtained by taking, for instance, $H_0=H_1=\frac{1}{2}h_1-2(M+M')\chi^{(1)}(t)(K_+h^2+h_1^2)$.

It remains to extend (3.40) to a general time interval $[0,t]$ for t of order 1. For this purpose, we choose a partition $0=u_0<u_1<\dots<u_K=t$ of $[0,t]$, satisfying $u_k=k\Delta\varepsilon$ for $0\leqslant k<K=\lceil t/(\Delta\varepsilon)\rceil$. Applying the local version of (3.40) to each interval $[u_k,u_{k+1}]$ and using the monotonicity of $\chi^{(2)}(u)$, the claimed estimate follows from

$$\mathbb{P}^{0,(\xi_0,0)}\Bigl\{\sup_{0\leqslant u\leqslant t\wedge\tau_{\mathcal{B}(h)}}\|\eta_u\|\geqslant h_1\Bigr\}\leqslant\sum_{k=0}^{K-1}\mathbb{P}^{0,(\xi_0,0)}\Bigl\{\sup_{u_k\leqslant u\leqslant u_{k+1}\wedge\tau_{\mathcal{B}(h)}}\|\eta_u\|\geqslant h_1\Bigr\}. \tag{3.45}$$

□

Proposition 3.1 and Lemma 3.4 together are sufficient to prove Theorem 2.4 on a timescale of order 1. We continue to assume that $y_0\in\mathcal{D}_0$ but we will no longer assume that $y_u^{\det}\in\mathcal{D}_0$ automatically holds for all $u\leqslant t$. Instead, we will employ Lemma 3.4 to compare $y_u\in\mathcal{D}_0$ and $y_u^{\det}$, taking advantage of the fact that on timescales of order 1, η_t is likely to remain small. Note that if the uniform-hyperbolicity Assumption 2.2 holds for

$\mathcal{D}_0$, then there exists a $\delta > 0$ of order 1 such that the δ-neighbourhood $\mathcal{D}_0^+(\delta)$ also satisfies this assumption. We introduce the first-exit time $\tau_{\mathcal{D}_0}^{\text{det}}$ of the deterministic process y_u^{det} from $\mathcal{D}_0^+(\delta)$ as

$$\tau_{\mathcal{D}_0}^{\text{det}} = \inf\{u \geqslant 0 \colon y_u^{\text{det}} \notin \mathcal{D}_0^+(\delta)\} \tag{3.46}$$

and remark in passing that $\tau_{\mathcal{B}(h)} \wedge \tau_\eta \leqslant \tau_{\mathcal{D}_0}^{\text{det}}$ holds whenever $h_1 \leqslant \delta$.

Corollary 3.5. *Fix a time $t > 0$ and $h > 0$ in such a way that $h \leqslant c_1 \chi^{(1)}(t \wedge \tau_{\mathcal{D}_0}^{\text{det}})^{-1}$ for a sufficiently small constant $c_1 > 0$ and $\chi^{(2)}(t \wedge \tau_{\mathcal{D}_0}^{\text{det}}) \leqslant (\rho^2 + \varepsilon)^{-1}$. Then for any $\alpha \in [0,1)$, any $\gamma \in (0,1/2)$ and any sufficiently small Δ,*

$$C_{n,m}^-(t,\varepsilon)\,\mathrm{e}^{-\kappa^-(0)h^2/\sigma^2} \leqslant \mathbb{P}^{0,(\xi_0,0)}\bigl\{\tau_{\mathcal{B}(h)} < t\bigr\} \leqslant C_{n,m,\gamma}^+(t,\varepsilon)\,\mathrm{e}^{-\kappa^+(\alpha)h^2/\sigma^2} \tag{3.47}$$

holds uniformly for all ξ_0 satisfying $\langle \xi_0, \overline{X}(y_0)^{-1}\xi_0\rangle \leqslant \alpha^2 h^2$. Here

$$\kappa^+(\alpha) = \gamma\bigl[1 - \alpha^2 - \mathcal{O}(\Delta) - \mathcal{O}((1 + \chi^{(1)}(t \wedge \tau_{\mathcal{D}_0}^{\text{det}}))h)\bigr], \tag{3.48}$$

$$\kappa^-(0) = \frac{1}{2}\bigl[1 + \mathcal{O}(h) + \mathcal{O}\bigl(\mathrm{e}^{-K_0 t/\varepsilon}\bigr)\bigr], \tag{3.49}$$

$$C_{n,m,\gamma}^+(t,\varepsilon) = \left\lceil \frac{t}{\Delta\varepsilon}\right\rceil \left[\frac{1}{(1-2\gamma)^{n/2}} + \bigl(\mathrm{e}^{n/4} + 2\,\mathrm{e}^{m/4}\bigr)\,\mathrm{e}^{-\kappa^+(0)h^2/\sigma^2}\right], \tag{3.50}$$

$$C_{n,m}^-(t,\varepsilon) = \left(\sqrt{\frac{2}{\pi}}\frac{h}{\sigma} \wedge 1\right) - \bigl(\mathrm{e}^{n/4} + 4\,\mathrm{e}^{m/4}\bigr)\,\mathrm{e}^{-\frac{h^2}{2\sigma^2}[1 - \mathcal{O}(\mathrm{e}^{-K_0 t/\varepsilon}) - \mathcal{O}((1+\chi^{(1)}(t \wedge \tau_{\mathcal{D}_0}^{\text{det}}))h)]}. \tag{3.51}$$

Proof: We first establish the upper bound. Fix an initial condition $(\xi_0, 0)$ satisfying $\langle \xi_0, \overline{X}(y_0)^{-1}\xi_0\rangle \leqslant \alpha^2 h^2$, and observe that

$$\begin{aligned}\mathbb{P}^{0,(\xi_0,0)}\bigl\{\tau_{\mathcal{B}(h)} < t\bigr\} &\leqslant \mathbb{P}^{0,(\xi_0,0)}\bigl\{\tau_{\mathcal{B}(h)} < t \wedge \tau_\eta\bigr\} + \mathbb{P}^{0,(\xi_0,0)}\bigl\{\tau_\eta < t \wedge \tau_{\mathcal{B}(h)}\bigr\} \\ &= \mathbb{P}^{0,(\xi_0,0)}\bigl\{\tau_{\mathcal{B}(h)} < t \wedge \tau_{\mathcal{D}_0}^{\text{det}} \wedge \tau_\eta\bigr\} + \mathbb{P}^{0,(\xi_0,0)}\bigl\{\tau_\eta < t \wedge \tau_{\mathcal{D}_0}^{\text{det}} \wedge \tau_{\mathcal{B}(h)}\bigr\}.\end{aligned} \tag{3.52}$$

To estimate the first term on the right-hand side, we again introduce a partition $0 = u_0 < u_1 < \dots < u_K = t$ of the time interval $[0,t]$, defined by $u_k = k\Delta\varepsilon$ for $0 \leqslant k < K = \lceil t/(\Delta\varepsilon)\rceil$. Thus we obtain

$$\mathbb{P}^{0,(\xi_0,0)}\bigl\{\tau_{\mathcal{B}(h)} < t \wedge \tau_{\mathcal{D}_0}^{\text{det}} \wedge \tau_\eta\bigr\} \leqslant \sum_{k=1}^{K} \mathbb{P}^{0,(\xi_0,0)}\bigl\{u_{k-1} \leqslant \tau_{\mathcal{B}(h)} < u_k \wedge \tau_{\mathcal{D}_0}^{\text{det}} \wedge \tau_\eta\bigr\}. \tag{3.53}$$

Before we estimate the summands on the right-hand side of (3.53), note that by the boundedness assumption on $\|\overline{X}(y)\|$ and $\|\overline{X}^{-1}(y)\|$, we have $\overline{X}(y_u)^{-1} = \overline{X}(y_u^{\text{det}})^{-1} + \mathcal{O}(h_1)$ for $u \leqslant \tau_{\mathcal{D}_0}^{\text{det}} \wedge \tau_\eta$. Thus the bound obtained in Proposition 3.1 can also be applied to estimate first-exit times from $\mathcal{B}(h)$ itself:

$$\begin{aligned}&\mathbb{P}^{0,(\xi_0,0)}\bigl\{u_{k-1} \leqslant \tau_{\mathcal{B}(h)} < u_k \wedge \tau_{\mathcal{D}_0}^{\text{det}} \wedge \tau_\eta\bigr\} \\ &\qquad\leqslant \mathbb{P}^{0,(\xi_0,0)}\Bigl\{\sup_{u_{k-1} \leqslant u < u_k \wedge \tau_{\mathcal{D}_0}^{\text{det}} \wedge \tau_\eta} \bigl\langle \xi_u, \overline{X}(y_u^{\text{det}})^{-1}\xi_u\bigr\rangle \geqslant h^2(1 - \mathcal{O}(h_1))\Bigr\}.\end{aligned} \tag{3.54}$$

The second term on the right-hand side of (3.52) can be estimated by Lemma 3.4 with

$$\mu^2 = 8M_1^2\bigl[\sqrt{K_+} + h_1/(h(1 - \mathcal{O}(h_1)))\bigr]^2 \Psi(t \wedge \tau_{\mathcal{D}_0}^{\text{det}}) / \bigl[1 - \mathcal{O}(h_1) - M_0\Delta\bigr] \tag{3.55}$$

and $h_1 = h/\sqrt{\kappa_0}$ which are chosen in such a way that the Gaussian part of ξ_t gives the major contribution to the probability. Thus we obtain that the probability in (3.52) is bounded by

$$\left\lceil \frac{t}{\Delta\varepsilon} \right\rceil \biggl[\frac{1}{(1-2\gamma)^{n/2}} \exp\Bigl\{-\gamma\frac{h^2}{\sigma^2}\bigl[1-\alpha^2-\mathcal{O}(\Delta)-\mathcal{O}(h)\bigr]\Bigr\} + \mathrm{e}^{n/4}\,\mathrm{e}^{-h^2/\sigma^2} + 2\,\mathrm{e}^{m/4}\exp\biggl\{-\frac{h^2(1-\mathcal{O}(\chi^{(1)}(t\wedge\tau_{\mathcal{D}_0}^{\det})h)-\mathcal{O}(\Delta\varepsilon))}{\sigma^2(\rho^2+\varepsilon)\chi^{(2)}(t\wedge\tau_{\mathcal{D}_0}^{\det})}\biggr\}\biggr], \tag{3.56}$$

where we have used the fact that $\Phi(t) \leqslant n\Psi(t)$, while $\Psi(t)$ and $\Theta(t)$ are at most of order 1. This completes the proof of the upper bound in (3.47).

The lower bound is a consequence of the fact that the Gaussian part of ξ_t gives the major contribution to the probability in (3.47). To check this, we split the probability as follows:

$$\mathbb{P}^{0,(\xi_0,0)}\bigl\{\tau_{\mathcal{B}(h)} < t\bigr\} = \mathbb{P}^{0,(\xi_0,0)}\bigl\{\tau_{\mathcal{B}(h)} < t, \tau_\eta \geqslant t, \tilde\tau_\xi \geqslant t\bigr\} + \mathbb{P}^{0,(\xi_0,0)}\bigl\{\tau_\eta \geqslant t, \tilde\tau_\xi < t\bigr\} + \mathbb{P}^{0,(\xi_0,0)}\bigl\{\tau_{\mathcal{B}(h)} < t, \tau_\eta < t\bigr\}, \tag{3.57}$$

where

$$\tilde\tau_\xi = \inf\{u \geqslant 0\colon \langle \xi_u, \overline{X}(y_u^{\det})^{-1}\xi_u\rangle \geqslant h^2(1+\mathcal{O}(h_1))\}, \tag{3.58}$$

and the $\mathcal{O}(h_1)$-term stems from estimating $\overline{X}(y_u)^{-1}$ by $\overline{X}(y_u^{\det})^{-1}$ as in (3.54). Also note that $\tau_{\mathcal{B}(h)} \leqslant \tilde\tau_\xi$. The first term on the right-hand side of (3.57) can be estimated as in the proof of Proposition 3.1. (Note that a lower bound is obtained trivially by considering the endpoint instead of the whole path.) Instead of applying Lemma 3.2, the Gaussian contribution can be estimated below by a straightforward calculation. The non-Gaussian parts are estimated *above* as before and are of smaller order. Finally, we need an upper bound for the probability that $\tau_\eta < t \leqslant \tau_{\mathcal{B}(h)}$ which can be obtained from Lemma 3.4. □

3.2 Longer timescales

Corollary 3.5 describes the dynamics on a timescale of order 1, or even on a slightly longer timescale if $\chi^{(1)}(t)$, $\chi^{(2)}(t)$ do not grow too fast. It may happen, however, that $y_t^{\det}$ remains in $\mathcal{D}_0$ for all positive times (e. g. when $\mathcal{D}_0$ is positively invariant under the reduced deterministic flow). In such a case, one would expect the vast majority of paths to remain concentrated in $\mathcal{B}(h)$ for a rather long period of time.

The approach used in Subsection 3.1 fails to control the dynamics on timescales on which $\chi^{(i)}(t) \gg 1$, because it uses in an essential way the fact that $\eta_t = y_t - y_t^{\det}$ remains small. Our strategy in order to describe the paths on longer timescales is to compare them to different deterministic solutions on time intervals $[0,T]$, $[T,2T]$, ..., where T is a possibly large constant such that Corollary 3.5 holds on time intervals of length T, provided y_t remains in $\mathcal{D}_0$. Essential ingredients for this approach are the Markov property and the following technical lemma, which is based on integration by parts.

Lemma 3.6. *Fix constants $s_1 \leqslant s_2$ in $[0,\infty]$, and assume we are given two continuously differentiable functions*

- $\varphi : [0,\infty) \to [0,\infty)$, *which is monotonously increasing and satisfies* $\varphi(s_2) = 1$,
- $\varphi_0 : [0,\infty) \to \mathbb{R}$ *which satisfies* $\varphi_0(s) \leqslant 0$ *for all* $s \leqslant s_1$.

Let $X \geqslant 0$ be a random variable such that $\mathbb{P}\{X < s\} \geqslant \varphi_0(s)$ for all $s \geqslant 0$. Then we have, for all $t \geqslant 0$,

$$\mathbb{E}\big\{1_{[0,t)}(X)\hat{\varphi}(X)\big\} \leqslant \hat{\varphi}(t)\mathbb{P}\{X < t\} - \int_{s_1\wedge t}^{s_2\wedge t} \varphi'(s)\varphi_0(s)\,\mathrm{d}s, \tag{3.59}$$

where $\hat{\varphi}(s) = \varphi(s) \wedge 1$.

We omit the proof of this result, which is rather standard. See, for instance, [7, Lemma A.1] for a very similar result.

When applying the preceding lemma, we will also need an estimate on the probability that $\langle \xi_T, \overline{X}(y_T)^{-1}\xi_T\rangle$ exceeds h^2. Corollary 3.5 provides, of course, such an estimate, but since it applies to the whole path, it does not give optimal bounds for the endpoint. An improved bound is given by the following lemma. Recall the definition of the first-exit time $\tau_{\mathcal{D}_0}$ of y_t from $\mathcal{D}_0$ from (2.24).

Lemma 3.7. *If T and h satisfy $h \leqslant c_1\chi^{(1)}(T\wedge\tau^{\mathrm{det}}_{\mathcal{D}_0})^{-1}$ and $\chi^{(2)}(T\wedge\tau^{\mathrm{det}}_{\mathcal{D}_0}) \leqslant (\rho^2+\varepsilon)^{-1}$, we have, for every $\gamma\in(0,1/2)$,*

$$\sup_{\xi_0\,:\,\langle\xi_0,\overline{X}(y_0)^{-1}\xi_0\rangle\leqslant h^2} \mathbb{P}^{0,(\xi_0,0)}\Big\{\langle \xi_T, \overline{X}(y_T)^{-1}\xi_T\rangle \geqslant h^2, \tau_{\mathcal{D}_0} \geqslant T\Big\} \leqslant \widehat{C}_{n,m,\gamma}(T,\varepsilon)\,\mathrm{e}^{-\kappa' h^2/\sigma^2}, \tag{3.60}$$

where

$$\kappa' = \gamma\big[1-\mathcal{O}(\Delta)-\mathcal{O}(h)-\mathcal{O}\big(\mathrm{e}^{-2K_0T/\varepsilon}/(1-2\gamma)\big)\big], \tag{3.61}$$

$$\widehat{C}_{n,m,\gamma}(T,\varepsilon) = \frac{1}{(1-2\gamma)^{n/2}} + 4C^{+}_{n,m,\gamma}(T,\varepsilon)\,\mathrm{e}^{-2\kappa^+(0)h^2/\sigma^2}. \tag{3.62}$$

Proof: We decompose ξ_t as $\xi_t = \xi^0_t + \xi^1_t + \xi^2_t$, where

$$\begin{aligned}
\xi^0_t &= U(t)\xi_0 + \frac{\sigma}{\sqrt{\varepsilon}}\int_0^t U(t,u)F_0(y^{\mathrm{det}}_u,\varepsilon)\,\mathrm{d}W_u,\\
\xi^1_t &= \frac{\sigma}{\sqrt{\varepsilon}}\int_0^t U(t,u)F_1(\xi_u,\eta_u,u,\varepsilon)\,\mathrm{d}W_u,\\
\xi^2_t &= \frac{1}{\varepsilon}\int_0^t U(t,u)b(\xi_u,\eta_u,u,\varepsilon)\,\mathrm{d}u,
\end{aligned} \tag{3.63}$$

and introduce the notations $\tilde{\tau}_\xi$ and $\tilde{\tau}_\eta$ for the stopping times which are defined like τ_ξ and τ_η in (3.9) and (3.6), but with h and h_1 replaced by $2h$ and $2h_1$, respectively. The probability in (3.60) is bounded by

$$\mathbb{P}^{0,(\xi_0,0)}\Big\{\langle \xi_T, \overline{X}(y^{\mathrm{det}}_T)^{-1}\xi_T\rangle \geqslant h^2(1-\mathcal{O}(h_1)), \tilde{\tau}_\eta > T\Big\} + \mathbb{P}^{0,(\xi_0,0)}\Big\{\tilde{\tau}_\eta \leqslant T\Big\}. \tag{3.64}$$

Let $H^2 = h^2(1-\mathcal{O}(h_1))$. As in the proof of Proposition 3.1, the first term on the right-hand side can be further decomposed as

$$\begin{aligned}
&\mathbb{P}^{0,(\xi_0,0)}\Big\{\langle \xi_T, \overline{X}(y^{\mathrm{det}}_T)^{-1}\xi_T\rangle \geqslant H^2, \tilde{\tau}_\eta > T\Big\}\\
&\quad\leqslant \mathbb{P}^{0,(\xi_0,0)}\Big\{\big\|\overline{X}(y^{\mathrm{det}}_T)^{-1/2}\xi^0_T\big\| \geqslant H_0\Big\} + \mathbb{P}^{0,(\xi_0,0)}\Big\{\tilde{\tau}_\eta > T, \tilde{\tau}_\xi \leqslant T\Big\}\\
&\qquad+ \mathbb{P}^{0,(\xi_0,0)}\Big\{\big\|\overline{X}(y^{\mathrm{det}}_T)^{-1/2}\xi^1_T\big\| \geqslant H_1, \tilde{\tau}_\eta > T, \tilde{\tau}_\xi > T\Big\}\\
&\qquad+ \mathbb{P}^{0,(\xi_0,0)}\Big\{\big\|\overline{X}(y^{\mathrm{det}}_T)^{-1/2}\xi^2_T\big\| \geqslant H_2, \tilde{\tau}_\eta > T, \tilde{\tau}_\xi > T\Big\},
\end{aligned} \tag{3.65}$$

where we choose H_1, H_2 twice as large as in the proof of Proposition 3.1, while $H_0 = H - H_1 - H_2$.

The first term on the right-hand side can be estimated as in Lemma 3.2, with the difference that, the expectation of ξ_T^0 being exponentially small in T/ε, it leads only to a correction of order $\mathrm{e}^{-2K_0T/\varepsilon}/(1-2\gamma)$ in the exponent. The second and the third term can be estimated by Corollary 3.5 and Lemma 3.3, the only difference lying in a larger absolute value of the exponent, because we enlarged h and h_1. The last term vanishes by our choice of H_2. Finally, the second term on the right-hand side of (3.64) can be estimated by splitting according to the value of $\tau_{\mathcal{B}(2h)}$ and applying Lemma 3.4 and Corollary 3.5. □

We are now prepared to establish an improved estimate on the distribution of $\tau_{\mathcal{B}(h)}$. As we will restart the process $y_t^{\det}$ whenever t is a multiple of T, we need the assumptions made in the previous section to hold uniformly in the initial condition $y_0 \in \mathcal{D}_0$. Therefore we will introduce replacements for some of the notations introduced before. Note that $\chi^{(1)}(t) = \chi^{(1)}_{y_0}(t)$ and $\chi^{(2)}(t) = \chi^{(2)}_{y_0}(t)$ depend on y_0 via the principal solution V. Also $\tau_{\mathcal{D}_0}^{\det} = \tau_{\mathcal{D}_0}^{\det}(y_0)$ naturally depends on y_0. We define

$$\widehat{\chi}^{(1)}(t) = \sup_{y_0\in\mathcal{D}_0} \chi^{(1)}_{y_0}\bigl(t\wedge\tau_{\mathcal{D}_0}^{\det}(y_0)\bigr), \tag{3.66}$$

$$\widehat{\chi}^{(2)}(t) = \sup_{y_0\in\mathcal{D}_0} \chi^{(2)}_{y_0}\bigl(t\wedge\tau_{\mathcal{D}_0}^{\det}(y_0)\bigr). \tag{3.67}$$

In the same spirit, the $\chi^{(i)}(T)$-dependent $\mathcal{O}(\cdot)$-terms in the definitions of $\kappa^+(\alpha)$, κ' and the prefactors like $C^+_{n,m,\gamma}(T,\varepsilon)$ are modified.

We fix a time T of order 1 satisfying $\widehat{\chi}^{(2)}(T) \leqslant (\rho^2+\varepsilon)^{-1}$. T is chosen in such a way that whenever $h \leqslant c_1\widehat{\chi}^{(1)}(T)^{-1}$, Corollary 3.5 (and Lemma 3.7) apply. Note that larger T would be possible unless ρ is of order 1, but for larger T the constraint on h becomes more restrictive which is not desirable. Having chosen T, we define the probabilities

$$P_k(h) = \mathbb{P}^{0,(0,0)}\bigl\{\tau_{\mathcal{B}(h)} < kT\wedge\tau_{\mathcal{D}_0}\bigr\}, \tag{3.68}$$

$$Q_k(h) = \mathbb{P}^{0,(0,0)}\bigl\{\langle\xi_{kT}, \overline{X}(y_{kT})^{-1}\xi_{kT}\rangle \geqslant h^2, \tau_{\mathcal{D}_0} \geqslant kT\bigr\}. \tag{3.69}$$

Corollary 3.5 provides a bound for $P_1(h)$, and Lemma 3.7 provides a bound for $Q_1(h)$. Subsequent bounds are computed by induction, and the following proposition describes one induction step.

Proposition 3.8. *Let $\hat\kappa \leqslant \kappa^+(0)\wedge\kappa'$. Assume that for some $k\in\mathbb{N}$,*

$$P_k(h) \leqslant D_k\,\mathrm{e}^{-\hat\kappa h^2/\sigma^2}, \tag{3.70}$$

$$Q_k(h) \leqslant \widehat{D}_k\,\mathrm{e}^{-\hat\kappa h^2/\sigma^2}. \tag{3.71}$$

Then the same bounds hold for k replaced by $k+1$, provided

$$D_{k+1} \geqslant D_k + C^+_{n,m,\gamma}(T,\varepsilon)\widehat{D}_k\frac{\gamma}{\gamma-\hat\kappa}\,\mathrm{e}^{(\gamma-\hat\kappa)h^2/\sigma^2} \tag{3.72}$$

$$\widehat{D}_{k+1} \geqslant \widehat{D}_k + \widehat{C}_{n,m,\gamma}(T,\varepsilon). \tag{3.73}$$

Remark 3.9. Note that $\gamma-\hat\kappa = \mathcal{O}(h)+\mathcal{O}(\Delta)+\mathcal{O}((1+\widehat{\chi}^{(1)}(T))h)+\mathcal{O}\bigl(\mathrm{e}^{-2K_0T/\varepsilon}/(1-2\gamma)\bigr)$. In the case $\kappa^+(0)=\kappa'=\gamma$, we may either choose $\hat\kappa < \kappa^+(0)\wedge\kappa'$, or we may replace (3.72) by

$$D_{k+1} \geqslant D_k + C^+_{n,m,\gamma}(T,\varepsilon)\widehat{D}_k\biggl[1+\log\biggl(\frac{C^+_{n,m,\gamma}(T,\varepsilon)}{\widehat{D}_k}\,\mathrm{e}^{\gamma h^2/\sigma^2}\biggr)\biggr]. \tag{3.74}$$

Below we will optimize with respect to $\hat\kappa$.

Proof of Proposition 3.8. We start by establishing (3.73). The Markov property allows for the decomposition

$$\begin{aligned}
&Q_{k+1}(h)\\
&\leqslant \mathbb{P}^{0,(0,0)}\bigl\{\tau_{\mathcal{B}(h)} < kT, \tau_{\mathcal{D}_0} \geqslant kT\bigr\}\\
&+ \mathbb{E}^{0,(0,0)}\Bigl\{1_{\{\tau_{\mathcal{B}(h)}\geqslant kT\}}\mathbb{P}^{kT,(\xi_{kT},0)}\bigl\{\langle\xi_{(k+1)T}, \overline{X}(y_{(k+1)T})^{-1}\xi_{(k+1)T}\rangle \geqslant h^2, \tau_{\mathcal{D}_0} \geqslant (k+1)T\bigr\}\Bigr\}\\
&\leqslant Q_k(h) + \widehat{C}_{n,m,\gamma}(T,\varepsilon)\,\mathrm{e}^{-\hat\kappa h^2/\sigma^2}, \qquad (3.75)
\end{aligned}$$

where the initial condition $(\xi_{kT}, 0)$ indicates that at time kT, we also restart the process of the deterministic slow variables y_t^{det} in the point $y_{kT} \in \mathcal{D}_0$. In the second line, we used Lemma 3.7. This shows (3.73).

As for (3.72), we again start from a decomposition, similar to (3.75):

$$\begin{aligned}
P_{k+1}(h) &= \mathbb{P}^{0,(0,0)}\bigl\{\tau_{\mathcal{B}(h)} < kT \wedge \tau_{\mathcal{D}_0}\bigr\}\\
&\quad+ \mathbb{E}^{0,(0,0)}\Bigl\{1_{\{\tau_{\mathcal{B}(h)}\geqslant kT\}}\mathbb{P}^{kT,(\xi_{kT},0)}\bigl\{\tau_{\mathcal{B}(h)} < (k+1)T \wedge \tau_{\mathcal{D}_0}\bigr\}\Bigr\}. \qquad (3.76)
\end{aligned}$$

Corollary 3.5 allows us to estimate

$$\begin{aligned}
&P_{k+1}(h)\\
&\quad\leqslant P_k(h) + \mathbb{E}^{0,(0,0)}\Bigl\{1_{\{\langle\xi_{kT},\overline{X}(y_{kT})^{-1}\xi_{kT}\rangle\leqslant h^2\}}\bigl[\varphi\bigl(\langle\xi_{kT}, \overline{X}(y_{kT})^{-1}\xi_{kT}\rangle\bigr)\wedge 1\bigr] \Bigm| \tau_{\mathcal{D}_0} \geqslant kT\Bigr\}\\
&\qquad\qquad\times \mathbb{P}^{0,(0,0)}\bigl\{\tau_{\mathcal{D}_0} \geqslant kT\bigr\}, \qquad (3.77)
\end{aligned}$$

with

$$\varphi(s) = C^+_{n,m,\gamma}(T,\varepsilon)\,\mathrm{e}^{(\gamma-\hat\kappa)h^2/\sigma^2}\,\mathrm{e}^{-\gamma(h^2-s)/\sigma^2}. \qquad (3.78)$$

(3.71) shows that

$$\mathbb{P}^{0,(0,0)}\bigl\{\langle\xi_{kT}, \overline{X}(y_{kT})^{-1}\xi_{kT}\rangle < s \bigm| \tau_{\mathcal{D}_0} \geqslant kT\bigr\} \geqslant \varphi_k(s), \qquad (3.79)$$

where

$$\varphi_k(s) := \bigl(1 - \widehat{D}_k\,\mathrm{e}^{-\hat\kappa s/\sigma^2}\bigr)/\mathbb{P}^{0,(0,0)}\bigl\{\tau_{\mathcal{D}_0} \geqslant kT\bigr\}. \qquad (3.80)$$

The functions φ and φ_k fulfil the assumptions of Lemma 3.6 with

$$\mathrm{e}^{\gamma s_2/\sigma^2} = C^+_{n,m,\gamma}(T,\varepsilon)^{-1}\,\mathrm{e}^{\hat\kappa h^2/\sigma^2} \qquad\text{and}\qquad \mathrm{e}^{\hat\kappa s_1/\sigma^2} = \widehat{D}_k. \qquad (3.81)$$

For $h^2 \leqslant s_1$, (3.70) becomes trivial, while for $h^2 > s_1$, Lemma 3.6 shows

$$\begin{aligned}
P_{k+1}(h) &\leqslant P_k(h) - \varphi(h^2\wedge s_2)\bigl[1 - \mathbb{P}^{0,(0,0)}\bigl\{\langle\xi_{kT}, \overline{X}(y_{kT})^{-1}\xi_{kT}\rangle < h^2, \tau_{\mathcal{D}_0} \geqslant kT\bigr\}\bigr]\\
&\qquad+ \varphi(s_1) + \int_{s_1}^{s_2\wedge h^2} \varphi'(s)\widehat{D}_k\,\mathrm{e}^{-\hat\kappa s/\sigma^2}\,\mathrm{d}s\\
&\leqslant P_k(h) + C^+_{n,m,\gamma}(T,\varepsilon)\widehat{D}_k\frac{\gamma}{\gamma-\hat\kappa}\,\mathrm{e}^{(\gamma-\hat\kappa)h^2/\sigma^2}\,\mathrm{e}^{-\hat\kappa h^2/\sigma^2}. \qquad (3.82)
\end{aligned}$$

Now, (3.72) is immediate. □

Repeated application of the previous result finally leads to the following estimate, which is equivalent to the upper bound in Theorem 2.4. Note that the lower bound in Theorem 2.4 is a direct consequence of the lower bound in Corollary 3.5, so that the corollary below completes the proof of the theorem.

Corollary 3.10. *Assume that* $y_0 \in \mathcal{D}_0$, $x_0 = \bar{x}(y_0, \varepsilon)$. *Then, for every* $t > 0$, *we have*

$$\begin{aligned}
&\mathbb{P}^{0,(x_0,y_0)}\bigl\{\tau_{\mathcal{B}(h)} < t \wedge \tau_{\mathcal{D}_0}\bigr\} \\
&\qquad \leqslant C^+_{n,m,\gamma}(T,\varepsilon)\biggl[1 + \widehat{C}_{n,m,\gamma}(T,\varepsilon)\Bigl(\frac{1}{2} + \frac{t}{T}\Bigr)^2 \frac{\gamma}{2(\gamma-\hat{\kappa})}\biggr] \mathrm{e}^{-(2\hat{\kappa}-\gamma)h^2/\sigma^2}.
\end{aligned} \tag{3.83}$$

In addition, the distribution of the endpoint ξ_t *satisfies*

$$\mathbb{P}^{0,(x_0,y_0)}\bigl\{\langle \xi_t, \overline{X}(y_t)^{-1}\xi_t\rangle \geqslant h^2, \tau_{\mathcal{D}_0} \geqslant t\bigr\} \leqslant \widehat{C}_{n,m,\gamma}(T,\varepsilon)\left\lceil \frac{t}{T}\right\rceil \mathrm{e}^{-\hat{\kappa}h^2/\sigma^2}. \tag{3.84}$$

PROOF: We already know the bounds (3.70) and (3.71) to hold for $k = 1$, with $D_1 = C^+_{n,m,\gamma}(T,\varepsilon)$ and $\widehat{D}_1 = \widehat{C}_{n,m,\gamma}(T,\varepsilon)$. Now the inductive relations (3.72) and (3.73) are seen to be satisfied by

$$\begin{aligned}
\widehat{D}_k &= k\widehat{C}_{n,m,\gamma}(T,\varepsilon), \\
D_k &= C^+_{n,m,\gamma}(T,\varepsilon)\biggl[1 + \widehat{C}_{n,m,\gamma}(T,\varepsilon)\frac{\gamma}{\gamma-\hat{\kappa}}\,\mathrm{e}^{(\gamma-\hat{\kappa})h^2/\sigma^2}\sum_{j=1}^{k-1} j\biggr].
\end{aligned} \tag{3.85}$$

The conclusion follows by taking $k = \lceil t/T\rceil$ and bounding the sum by $\frac{1}{2}(t/T)(t/T+1) \leqslant \frac{1}{2}(t/T+1/2)^2$. $\square$

To complete the proof of Theorem 2.4, we first optimize our choice of $\hat{\kappa}$, taking into account the constraint $\hat{\kappa} \leqslant \kappa^+(0) \wedge \kappa'$. By doing so, we find that

$$\frac{\gamma}{2(\gamma-\hat{\kappa})}\,\mathrm{e}^{-(2\hat{\kappa}-\gamma)h^2/\sigma^2} \leqslant \frac{2h^2}{\sigma^2}\,\mathrm{e}^{-\kappa h^2/\sigma^2}, \tag{3.86}$$

where we have set

$$\kappa = \gamma\bigl[1 - \mathcal{O}(h) - \mathcal{O}(\Delta) - \mathcal{O}\bigl(\mathrm{e}^{-const/\varepsilon}/(1-2\gamma)\bigr)\bigr]. \tag{3.87}$$

Simplifying the prefactor in (3.83) finally yields

$$\mathbb{P}^{0,(x_0,y_0)}\bigl\{\tau_{\mathcal{B}(h)} < t \wedge \tau_{\mathcal{D}_0}\bigr\} \leqslant \frac{(1+t)^2}{\Delta\varepsilon}\Bigl[(1-2\gamma)^{-n} + \mathrm{e}^{n/4} + \mathrm{e}^{m/4}\Bigr]\frac{h^2}{\sigma^2}\,\mathrm{e}^{-\kappa h^2/\sigma^2}. \tag{3.88}$$

3.3 Approaching the adiabatic manifold

The following result gives a rather rough description of the behaviour of paths starting at a (sufficiently small) distance of order 1 from the adiabatic manifold. It is, however, sufficient to show that with large probability, these paths will reach the set $\mathcal{B}(h)$, for some $h > \sigma$, in a time of order $\varepsilon|\log h|$.

Proposition 3.11. *Let t satisfy the hypotheses of Corollary 3.5. Then there exist constants h_0, δ_0, c_0 and K_0 such that, for $h \leqslant h_0$, $\delta \leqslant \delta_0$, $\gamma \in (0,1/2)$ and Δ sufficiently small,*

$$\sup_{\xi_0\,:\,\langle\xi_0,\overline{X}(y_0)^{-1}\xi_0\rangle\leqslant\delta^2} \mathbb{P}^{0,(\xi_0,0)}\biggl\{\sup_{0\leqslant s\leqslant t\wedge\tau_{\mathcal{D}_0}} \frac{\langle\xi_s,\overline{X}(y_s)^{-1}\xi_s\rangle}{(h+c_0\delta\,\mathrm{e}^{-K_0 s/\varepsilon})^2}\geqslant 1\biggr\} \tag{3.89}$$
$$\leqslant \biggl\lceil\frac{t}{\Delta\varepsilon}\biggr\rceil \biggl[\frac{1}{(1-2\gamma)^{n/2}}+\bigl(\mathrm{e}^{n/4}+2\,\mathrm{e}^{m/4}\bigr)\,\mathrm{e}^{-\kappa h^2/\sigma^2}\biggr]\,\mathrm{e}^{-\kappa h^2/\sigma^2},$$

where $\kappa=\gamma[1-\mathcal{O}(h)-\mathcal{O}(\Delta)-\mathcal{O}(\delta)]$.

PROOF: We start again by considering an interval $[s,t]$ with $t-s=\Delta\varepsilon$. Let $y_0^{\det}=y_0\in\mathcal{D}_0$. Then

$$P=\mathbb{P}^{0,(\xi_0,0)}\biggl\{\sup_{s\leqslant u\leqslant t\wedge\tau_{\mathcal{D}_0}\wedge\tau_\eta} \frac{\langle\xi_u,\overline{X}(y_u)^{-1}\xi_u\rangle}{(h+c_0\delta\,\mathrm{e}^{-K_0 u/\varepsilon})^2}\geqslant 1\biggr\}$$
$$\leqslant \mathbb{P}^{0,(\xi_0,0)}\biggl\{\sup_{s\leqslant u\leqslant t\wedge\tau} \langle\xi_u,\overline{X}(y_u^{\det})^{-1}\xi_u\rangle\geqslant H^2\biggr\}, \tag{3.90}$$

where τ is a stopping time defined by

$$\tau=\tau_{\mathcal{D}_0}\wedge\tau_\eta\wedge\inf\bigl\{u\geqslant 0\colon \langle\xi_u,\overline{X}(y_u^{\det})^{-1}\xi_u\rangle\geqslant(h+c_0\delta\,\mathrm{e}^{-K_0u/\varepsilon})^2(1-\mathcal{O}(\Delta))\bigr\}, \tag{3.91}$$

and H^2 is a shorthand for $H^2=H_t^2=(h+c_0\delta\,\mathrm{e}^{-K_0t/\varepsilon})^2(1-\mathcal{O}(\Delta))$.

The probability on the right-hand side of (3.90) can be bounded, as in Proposition 3.1, by the sum $P_0+P_1+P_2$, defined in (3.17), provided $H_0+H_1+H_2=H$. Since $\|U(s)\|$ decreases like $\mathrm{e}^{-K_0s/\varepsilon}=\mathrm{e}^{K_0\Delta}\,\mathrm{e}^{-K_0t/\varepsilon}$, we have

$$P_0\leqslant\frac{1}{(1-2\gamma)^{n/2}}\exp\Bigl[-\gamma\frac{H_0^2-\mathit{const}\;\delta^2\,\mathrm{e}^{2K_0\Delta}\,\mathrm{e}^{-2K_0t/\varepsilon}}{\sigma^2}\Bigr]. \tag{3.92}$$

Following the proof of Lemma 3.3, and taking into account the new definition of τ, we further obtain that

$$P_1\leqslant\mathrm{e}^{n/4}\exp\Bigl\{-\frac{H_1^2}{\sigma^2}\frac{1}{M_1^2\,\mathit{const}\;[(h+h_1)^2\Psi(t)+c_0^2\delta^2(t/\varepsilon)\,\mathrm{e}^{-2K_0t/\varepsilon}]}\Bigr\}. \tag{3.93}$$

As for P_2, it can be estimated trivially, provided

$$H_2\geqslant\mathit{const}\;\frac{M}{K_0}\Bigl[(h^2+hh_1)\Theta(t)+c_0^2\delta^2\,\mathrm{e}^{K_0\Delta}\,\mathrm{e}^{-K_0t/\varepsilon}\Bigr]. \tag{3.94}$$

Choosing H_1 in such a way that the exponent in (3.93) equals H^2/σ^2, we obtain

$$P\leqslant\biggl(\frac{1}{(1-2\gamma)^{n/2}}+\mathrm{e}^{n/4}\,\mathrm{e}^{-H^2/(2\sigma^2)}\biggr)\exp\Bigl\{-\gamma\frac{H^2}{\sigma^2}\bigl[1-\mathcal{O}(\Delta)-\mathcal{O}(h+h_1+c_0\delta)\bigr]\Bigr\}, \tag{3.95}$$

where we choose h_1 proportional to $h+c_0\delta\,\mathrm{e}^{K_0\Delta}\,\mathrm{e}^{-K_0t/\varepsilon}$. The remainder of the proof is similar to the proofs of Lemma 3.4 and Corollary 3.5. □

The preceding lemma shows that after a time t_1 of order $\varepsilon|\log h|$, the paths are likely to have reached $\mathcal{B}(h)$. As in Lemma 3.7, an improved bound for the distribution of the endpoint ξ_{t_1} can be obtained. Repeating the arguments leading to Theorem 2.4, namely using Lemma 3.6 on integration by parts and mimicking the proof of Corollary 3.10, one can show that after time t_1, the probability of leaving $\mathcal{B}(h)$ behaves as if the process had started on the adiabatic manifold, i.e.,

$$\mathbb{P}^{0,(\xi_0,0)}\Bigl\{\sup_{t_1\leqslant s\leqslant t\wedge\tau_{\mathcal{D}_0}}\langle\xi_s,\overline{X}(y_s)^{-1}\xi_s\rangle\geqslant h^2\Bigr\}\leqslant C^+_{n,m,\gamma,\Delta}(t,\varepsilon)\frac{h^2}{\sigma^2}\,\mathrm{e}^{-\kappa^+h^2/\sigma^2}, \tag{3.96}$$

uniformly for all ξ_0 such that $\langle\xi_0,\overline{X}(y_0)^{-1}\xi_0\rangle\leqslant\delta^2$. Here $C^+_{n,m,\gamma,\Delta}(t,\varepsilon)$ is the same prefactor as in Theorem 2.4, cf. (2.27), and

$$\kappa^+=\gamma\bigl[1-\mathcal{O}(h)-\mathcal{O}(\Delta)-\mathcal{O}\bigl(\delta\,\mathrm{e}^{-const\,(t_1\wedge1)/\varepsilon}/(1-2\gamma)\bigr)\bigr]. \tag{3.97}$$

This completes our discussion of general initial conditions.

4 Proofs – Dynamics of ζ_t

In this section, we consider again the SDE

$$\begin{aligned}\mathrm{d}x_t&=\frac{1}{\varepsilon}f(x_t,y_t,\varepsilon)\,\mathrm{d}t+\frac{\sigma}{\sqrt{\varepsilon}}F(x_t,y_t,\varepsilon)\,\mathrm{d}W_t,\\ \mathrm{d}y_t&=g(x_t,y_t,\varepsilon)\,\mathrm{d}t+\sigma'G(x_t,y_t,\varepsilon)\,\mathrm{d}W_t\end{aligned} \tag{4.1}$$

under Assumption 2.2, that is, when starting near a uniformly asymptotically stable manifold. We denote by $(x_t^{\det},y_t^{\det})$, with $x_t^{\det}=\bar{x}(y_t^{\det},\varepsilon)$, the deterministic solution starting in $y_0^{\det}=y_0\in\mathcal{D}_0$. The system can be rewritten in the form (3.3), or, in compact notation, as

$$\mathrm{d}\zeta_t=\bigl[\mathcal{A}(y_t^{\det},\varepsilon)\zeta_t+\mathcal{B}(\zeta_t,t,\varepsilon)\bigr]\,\mathrm{d}t+\sigma\bigl[\mathcal{F}_0(y_t^{\det},\varepsilon)+\mathcal{F}_1(\zeta_t,t,\varepsilon)\bigr]\,\mathrm{d}W_t, \tag{4.2}$$

where $\zeta^T=(\xi^T,\eta^T)$, $\mathcal{A}$ and $\mathcal{F}_0$ have been defined in (2.37), and the components of $\mathcal{B}^T=(\varepsilon^{-1}b^T,c^T)$ and $\mathcal{F}_1^T=(\varepsilon^{-1/2}F_1^T,\rho G_1^T)$ satisfy the bounds (3.4).

The solution of (4.2) with initial condition $\zeta_0^T=(\xi_0^T,0)$ can be written in the form

$$\begin{aligned}\zeta_t={}&\mathcal{U}(t)\zeta_0+\sigma\int_0^t\mathcal{U}(t,s)\mathcal{F}_0(y_s^{\det},\varepsilon)\,\mathrm{d}W_s\\ &+\int_0^t\mathcal{U}(t,s)\mathcal{B}(\zeta_s,s,\varepsilon)\,\mathrm{d}s+\sigma\int_0^t\mathcal{U}(t,s)\mathcal{F}_1(\zeta_s,s,\varepsilon)\,\mathrm{d}W_s.\end{aligned} \tag{4.3}$$

The components of the principal solution $\mathcal{U}(t,s)$ satisfy the bounds

$$\begin{aligned}\|U(t,s)\|&\leqslant const\,\mathrm{e}^{-K_0(t-s)/\varepsilon},\\ \|S(t,s)\|&\leqslant const\|C\|_\infty\frac{\varepsilon}{K_0}\bigl(1-\mathrm{e}^{-K_0(t-s)/\varepsilon}\bigr)\sup_{s\leqslant u\leqslant t}\|V(t,u)\|.\end{aligned} \tag{4.4}$$

We want to estimate the first-exit time

$$\tau_\zeta=\inf\bigl\{u\geqslant0\colon\langle\zeta_u,\overline{\mathcal{Z}}(u)^{-1}\zeta_u\rangle\geqslant h^2\bigr\}, \tag{4.5}$$

with $\overline{\mathcal{Z}}(u)$ defined in (2.47). The inverse of $\overline{\mathcal{Z}}(u)$ is given by

$$\overline{\mathcal{Z}}^{-1} = \begin{pmatrix} (\overline{X} - \overline{Z} Y^{-1} \overline{Z}^T)^{-1} & -\overline{X}^{-1} \overline{Z} (Y - \overline{Z}^T \overline{X}^{-1} \overline{Z})^{-1} \\ - Y^{-1} \overline{Z}^T (\overline{X} - \overline{Z} Y^{-1} \overline{Z}^T)^{-1} & (Y - \overline{Z}^T \overline{X}^{-1} \overline{Z})^{-1} \end{pmatrix}. \tag{4.6}$$

Since we assume $\|\overline{X}\|_\infty$ and $\|\overline{X}^{-1}\|_\infty$ to be bounded, $\|\overline{Z}\|_\infty = \mathcal{O}(\sqrt{\varepsilon}\rho + \varepsilon)$ and $\|Y^{-1}\|_{[0,t]} = \mathcal{O}(1/(\rho^2+\varepsilon))$, we have

$$\overline{\mathcal{Z}}^{-1} = \begin{pmatrix} \mathcal{O}(1) & \mathcal{O}(1) \\ \mathcal{O}(1) & \mathcal{O}(1/(\rho^2+\varepsilon)) \end{pmatrix}. \tag{4.7}$$

As in Section 3, we start by examining the dynamics of ζ_u on an interval $[s,t]$ with $\Delta = (t-s)/\varepsilon = \mathcal{O}_\varepsilon(1)$.

The following functions will play a similar rôle as the functions Φ and Ψ, introduced in (3.5), played in Section 3:

$$\begin{aligned} \widehat{\Phi}(t) &= \int_0^t \operatorname{Tr}\bigl[\mathcal{J}(v)^T \mathcal{U}(t,v)^T \overline{\mathcal{Z}}(t)^{-1} \mathcal{U}(t,v) \mathcal{J}(v)\bigr] \, \mathrm{d}v, \\ \widehat{\Psi}(t) &= \int_0^t \bigl\| \mathcal{J}(v)^T \mathcal{U}(t,v)^T \overline{\mathcal{Z}}(t)^{-1} \mathcal{U}(t,v) \mathcal{J}(v) \bigr\| \, \mathrm{d}v, \end{aligned} \tag{4.8}$$

where

$$\mathcal{J}(v) = \frac{1}{\sqrt{2} M_1 h \|\overline{\mathcal{Z}}\|_\infty^{1/2}} \mathcal{F}_1(\zeta_v, v, \varepsilon) = \begin{pmatrix} \mathcal{O}(\frac{1}{\sqrt{\varepsilon}}) \\ \mathcal{O}(\rho) \end{pmatrix} \tag{4.9}$$

for $v \leqslant \tau_\zeta$. Using the representations (2.39) of $\mathcal{U}$ and (4.6) of $\overline{\mathcal{Z}}^{-1}$ and expanding the matrix product, one obtains the relations

$$\begin{aligned} \widehat{\Phi}(t) &\leqslant \Phi(t) + \rho^2 \int_0^t \operatorname{Tr}\bigl[V(t,v)^T Y(t)^{-1} V(t,v)\bigr] \, \mathrm{d}v + \mathcal{O}\bigl((n+m)(1+\chi^{(1)}(t) + \chi^{(2)}(t))\bigr), \\ \widehat{\Psi}(t) &\leqslant \Psi(t) + \rho^2 \int_0^t \bigl\| V(t,v)^T Y(t)^{-1} V(t,v) \bigr\| \, \mathrm{d}v + \mathcal{O}\bigl(1+\chi^{(1)}(t) + \chi^{(2)}(t)\bigr), \end{aligned} \tag{4.10}$$

valid for all $t \leqslant \tau_\zeta$. Now we are ready to establish the following analog of Proposition 3.1.

Proposition 4.1. *Fix an initial condition (x_0, y_0) with $y_0 \in \mathcal{D}_0$ and $x_0 = \bar{x}(y_0, \varepsilon)$, and let t be such that $y_u^{\det} \in \mathcal{D}_0$ for all $u \leqslant t$. Then, for all $\alpha \in [0,1]$, all $\gamma \in (0,1/2)$ and all $\mu > 0$,*

$$\begin{aligned} &\sup_{\zeta_0 = (\xi_0, 0) \,:\, \langle \xi_0, \overline{X}(y_0)^{-1} \xi_0 \rangle \leqslant \alpha^2 h^2} \mathbb{P}^{0,\zeta_0} \Bigl\{ \sup_{s \leqslant u \leqslant t \wedge \tau_{\mathcal{D}_0}} \bigl\langle \zeta_u, \overline{\mathcal{Z}}(u)^{-1} \zeta_u \bigr\rangle \geqslant h^2 \Bigr\} \\ &\quad \leqslant \frac{1}{(1-2\gamma)^{(n+m)/2}} \exp\Bigl\{ -\gamma \frac{h^2}{\sigma^2} \Bigl[1 - \alpha^2 - \mathcal{O}\bigl(\Delta + \varepsilon + \mu h + h \|\overline{\mathcal{Z}}\|_{[0,t]} \|Y^{-1}\|_{[0,t]}^{1/2} \chi^{(1)}(t)\bigr) \Bigr] \Bigr\} \\ &\qquad + \mathrm{e}^{\widehat{\Phi}(t)/4\widehat{\Psi}(t)} \exp\Bigl\{ -\frac{h^2}{\sigma^2} \frac{\mu^2 (1 - \mathcal{O}(\Delta))}{16 M_1^2 \|\overline{\mathcal{Z}}\|_{[0,t]} \widehat{\Psi}(t)} \Bigr\} \end{aligned} \tag{4.11}$$

holds for all $h < 1/\mu$.

Proof: Writing $\zeta_u = \mathcal{U}(u,s) \Upsilon_u$, we have

$$\mathbb{P}^{0,\zeta_0} \Bigl\{ \sup_{s \leqslant u \leqslant t \wedge \tau_{\mathcal{D}_0}} \bigl\langle \zeta_u, \overline{\mathcal{Z}}(u)^{-1} \zeta_u \bigr\rangle \geqslant h^2 \Bigr\} = \mathbb{P}^{0,\zeta_0} \Bigl\{ \sup_{s \leqslant u \leqslant t \wedge \tau_{\mathcal{D}_0} \wedge \tau_\zeta} \|\mathcal{Q}(u) \Upsilon_u\| \geqslant h \Bigr\} \tag{4.12}$$

where $\mathcal{Q}(u)$ is the symmetric matrix defined by

$$\mathcal{Q}(u)^2 = \mathcal{U}(u,s)^T \overline{\mathcal{Z}}(u)^{-1} \mathcal{U}(u,s). \tag{4.13}$$

As in the proof of Proposition 3.1, we want to eliminate the u-dependence of $\mathcal{Q}$ in (4.12). It turns out that the relation $\|\mathcal{Q}(u)\mathcal{Q}(t)^{-1}\| = 1 + \mathcal{O}(\Delta)$ still holds in the present situation, although the proof is less straightforward than before. We establish this result in Lemma 4.2 below.

Splitting Υ_u into the sum $\Upsilon_u = \Upsilon_u^0 + \Upsilon_u^1 + \Upsilon_u^2$, where the Υ_u^i are defined in a way analogous to (3.16), we can estimate the probability in (4.12) by the sum $P_0 + P_1 + P_2$, where

$$\begin{aligned}
P_0 &= \mathbb{P}^{0,\zeta_0}\Bigl\{ \sup_{s \leqslant u \leqslant t \wedge \tau_{\mathcal{D}_0}} \|\mathcal{Q}(t)\Upsilon_u^0\| \geqslant H_0 \Bigr\}, \\
P_1 &= \mathbb{P}^{0,\zeta_0}\Bigl\{ \sup_{s \leqslant u \leqslant t \wedge \tau_{\mathcal{D}_0} \wedge \tau_\zeta} \|\mathcal{Q}(t)\Upsilon_u^1\| \geqslant H_1 \Bigr\}, \\
P_2 &= \mathbb{P}^{0,\zeta_0}\Bigl\{ \sup_{s \leqslant u \leqslant t \wedge \tau_{\mathcal{D}_0} \wedge \tau_\zeta} \|\mathcal{Q}(t)\Upsilon_u^2\| \geqslant H_2 \Bigr\},
\end{aligned} \tag{4.14}$$

and $H_0 + H_1 + H_2 = h(1 - \mathcal{O}(\Delta))$. Following the proof of Lemma 3.2, it is straightforward to show that

$$P_0 \leqslant \frac{1}{(1-2\gamma)^{(n+m)/2}} \exp\Bigl\{ -\frac{\gamma}{\sigma^2}(H_0^2 - \alpha^2 h^2)(1 - \mathcal{O}(\varepsilon)) \Bigr\}, \tag{4.15}$$

the sole difference being the factor $\mathcal{O}(\varepsilon)$ in the exponent which stems from the fact that $\langle \zeta_0, \overline{\mathcal{Z}}(0)^{-1}\zeta_0 \rangle = \langle \xi_0, \overline{X}(0)^{-1}\xi_0 \rangle (1 + \mathcal{O}(\varepsilon))$. Furthermore, similar arguments as in the proof of Lemma 3.3 lead to the bound

$$P_1 \leqslant \exp\Bigl\{ -\frac{(H_1^2 - 2\sigma^2 M_1^2 h^2 \|\overline{\mathcal{Z}}\|_{[0,t]} \widehat{\Phi}(t))^2}{16\sigma^2 M_1^2 h^2 H_1^2 \|\overline{\mathcal{Z}}\|_{[0,t]} \widehat{\Psi}(t)} \Bigr\}. \tag{4.16}$$

Finally, the estimate

$$\begin{aligned}
\|\mathcal{Q}(t)\Upsilon_{u\wedge\tau_\zeta}^2\|^2 &\leqslant \int_0^{u\wedge\tau_\zeta} \int_0^{u\wedge\tau_\zeta} \bigl\| \mathcal{B}(\zeta_v, v, \varepsilon)^T \mathcal{U}(t,v)^T \overline{\mathcal{Z}}(t)^{-1} \mathcal{U}(t,w) \mathcal{B}(\zeta_w, w, \varepsilon) \bigr\| \,\mathrm{d}v \,\mathrm{d}w \\
&\leqslant \mathit{const}\; h^4 \|\overline{\mathcal{Z}}\|_{[0,t]}^2 \bigl(1 + \|Y^{-1}\|_{[0,t]} \chi^{(1)}(u)^2\bigr)
\end{aligned} \tag{4.17}$$

shows that $P_2 = 0$ for $H_2 \geqslant \mathcal{O}(h^2 \|\overline{\mathcal{Z}}\|_{[0,t]} \|Y^{-1}\|_{[0,t]}^{1/2} \chi^{(1)}(t))$. Hence (4.11) follows by taking $H_1 = \mu h^2 (1 - \mathcal{O}(\Delta))$. $\square$

In the proof of Proposition 4.1, we have used the following estimate.

Lemma 4.2. *For $\Delta = (t-s)/\varepsilon$ sufficiently small,*

$$\sup_{s \leqslant u \leqslant t} \|\mathcal{Q}(u)\mathcal{Q}(t)^{-1}\| = 1 + \mathcal{O}(\Delta). \tag{4.18}$$

PROOF: Using the fact that $\mathcal{Q}(v)^{-2}$ satisfies the ODE

$$\frac{\mathrm{d}}{\mathrm{d}v} \mathcal{Q}(v)^{-2} = \mathcal{U}(s,v) \mathcal{F}_0(y_v^{\det}, \varepsilon) \mathcal{F}_0(y_v^{\det}, \varepsilon)^T \mathcal{U}(s,v)^T, \tag{4.19}$$

we obtain the relation

$$\mathcal{Q}(u)^2\mathcal{Q}(t)^{-2} = 1\!\!1 + \mathcal{Q}(u)^2 \int_u^t \mathcal{U}(s,v)\mathcal{F}_0(y_v^{\det},\varepsilon)\mathcal{F}_0(y_v^{\det},\varepsilon)^T\mathcal{U}(s,v)^T\,\mathrm{d}v. \tag{4.20}$$

The definition of $\mathcal{F}_0$ and the bound (4.4) on $\|S\|$ allow us to write

$$\mathcal{U}(s,v)\mathcal{F}_0(y_v^{\det},\varepsilon)\mathcal{F}_0(y_v^{\det},\varepsilon)^T\mathcal{U}(s,v)^T = \begin{pmatrix} \mathcal{O}(1/\varepsilon) & \mathcal{O}(\Delta+\rho/\sqrt{\varepsilon}) \\ \mathcal{O}(\Delta+\rho/\sqrt{\varepsilon}) & \mathcal{O}(\Delta^2\varepsilon+\rho^2) \end{pmatrix}. \tag{4.21}$$

Using the estimate (4.7) for $\overline{\mathcal{Z}}^{-1}$ and the fact that we integrate over an interval of length $\Delta\varepsilon$, it follows that

$$\mathcal{Q}(u)^2\mathcal{Q}(t)^{-2} - 1\!\!1 = \Delta \begin{pmatrix} \mathcal{O}(1) & \mathcal{O}(\Delta\varepsilon+\rho\sqrt{\varepsilon}) \\ \mathcal{O}(1) & \mathcal{O}(\varepsilon+\rho\sqrt{\varepsilon}) \end{pmatrix}, \tag{4.22}$$

which implies (4.18). □

Now, Theorem 2.5 follows from Proposition 4.1, by taking a regular partition of $[0,t]$ with spacing $\Delta\varepsilon$ and $\mu = 4M_1\|\overline{\mathcal{Z}}\|_{[0,t]}^{1/2}\widehat{\Psi}(t)^{1/2}$. We use in particular the fact that $\widehat{\Psi}(t) = \mathcal{O}(1+\chi^{(1)}(t)+\chi^{(2)}(t))$.

5 Proofs – Bifurcations

We consider in this section the behaviour of the SDE (2.1) near a bifurcation point. The system can be written in the form

$$\begin{aligned}
\mathrm{d}\xi_t^- &= \frac{1}{\varepsilon}\hat{f}^-(\xi_t^-,z_t,y_t,\varepsilon)\,\mathrm{d}t + \frac{\sigma}{\sqrt{\varepsilon}}\widehat{F}^-(\xi_t^-,z_t,y_t,\varepsilon)\,\mathrm{d}W_t, \\
\mathrm{d}z_t &= \frac{1}{\varepsilon}\hat{f}^0(\xi_t^-,z_t,y_t,\varepsilon)\,\mathrm{d}t + \frac{\sigma}{\sqrt{\varepsilon}}\widehat{F}^0(\xi_t^-,z_t,y_t,\varepsilon)\,\mathrm{d}W_t, \\
\mathrm{d}y_t &= \hat{g}(\xi_t^-,z_t,y_t,\varepsilon)\,\mathrm{d}t + \sigma'\widehat{G}(\xi_t^-,z_t,y_t,\varepsilon)\,\mathrm{d}W_t,
\end{aligned} \tag{5.1}$$

compare (2.63) and (2.64). We consider the dynamics as long as (z_t,y_t) evolves in a neighbourhood $\mathcal{N}$ of the bifurcation point, which is sufficiently small for the adiabatic manifold to be uniformly asymptotically stable, that is, all the eigenvalues of $\partial_x\hat{f}^-(0,z,y,\varepsilon)$ have negative real parts, uniformly bounded away from zero.

5.1 Exit from $\mathcal{B}^-(h)$

Let $h_\eta, h_z \geqslant 0$. In addition to the stopping time

$$\tau_\eta = \inf\{s>0\colon \|\eta_s\| \geqslant h_\eta\}, \tag{5.2}$$

cf. (3.6), we introduce the corresponding stopping time for $z_s - z_s^{\det}$, namely,

$$\tau_z = \inf\{s>0\colon \|z_s - z_s^{\det}\| \geqslant h_z\}. \tag{5.3}$$

The following result is obtained using almost the same line of thought as in Section 3.1.

Proposition 5.1. *Let t be of order 1 at most. Then, for all initial conditions ξ_0^- such that $\langle \xi_0^-, \overline{X}^-(y_0,z_0)^{-1}\xi_0^-\rangle \leqslant \alpha^2 h^2$ with an $\alpha \in (0,1]$, all $\gamma \in (0,1/2)$, and all sufficiently small $\Delta > 0$,*

$$
\begin{aligned}
\mathbb{P}^{0,(\xi_0^-,z_0,y_0)}&\Big\{\sup_{0\leqslant s\leqslant t\wedge\tau_{\mathcal{N}}\wedge\tau_\eta\wedge\tau_z} \langle \xi_s^-, \overline{X}^-(y_s,z_s)^{-1}\xi_s^-\rangle \geqslant h^2\Big\} \\
&\leqslant \left\lceil \frac{t}{\Delta\varepsilon}\right\rceil \frac{1}{(1-2\gamma)^{(n-q)/2}} \exp\Big\{-\gamma\frac{h^2}{\sigma^2}\big[1-\alpha^2-\mathcal{O}\big(\Delta+(1+\mu)h+h_\eta+h_z\big)\big]\Big\} \\
&\quad + \left\lceil \frac{t}{\Delta\varepsilon}\right\rceil \mathrm{e}^{(n-q)/4} \exp\Big\{-\frac{h^2}{\sigma^2}\frac{\mu^2(1-\mathcal{O}(\Delta))}{\mathcal{O}((1+(h_\eta+h_z)/h)^2)}\Big\}.
\end{aligned} \tag{5.4}
$$

PROOF: The proof is similar to the proof of Corollary 3.5, the main difference being the need for the additional stopping time τ_z. Note that this results in error terms depending on $h_\eta + h_z$ instead of h_η only. □

Next, we need to control the stopping times τ_η and τ_z. Lemma 3.4 holds with minor changes, incorporating the z_t-dependent terms. We find that

Lemma 5.2. *Let ξ_0^- satisfy $\langle \xi_0^-, \overline{X}^-(y_0,z_0)^{-1}\xi_0^-\rangle \leqslant h^2$. Then*

$$
\begin{aligned}
\mathbb{P}^{0,(\xi_0,0)}&\Big\{\sup_{0\leqslant u\leqslant t\wedge\tau_{\mathcal{B}^-(h)}\wedge\tau_z} \|\eta_u\| \geqslant h_\eta\Big\} \\
&\leqslant 2\left\lceil \frac{t}{\Delta\varepsilon}\right\rceil \mathrm{e}^{m/4} \exp\Big\{-\kappa_0\frac{h_\eta^2(1-\mathcal{O}(\Delta\varepsilon))}{\sigma^2(\rho^2+\varepsilon)\chi^{(2)}(t)}\Big[1-\mathcal{O}\Big(\chi^{(1)}(t)\,h_\eta\Big(1+\frac{h^2}{h_\eta^2}+\frac{h_z^2}{h_\eta^2}\Big)\Big)\Big]\Big\}.
\end{aligned} \tag{5.5}
$$

The next result allows to control the stopping time τ_z. We define

$$
\chi_z^{(1)}(t) = \sup_{0\leqslant s\leqslant t}\int_0^s \Big(\sup_{u\leqslant v\leqslant s}\|U^0(s,v)\|\Big)\,\mathrm{d}u, \tag{5.6}
$$

$$
\chi_z^{(2)}(t) = \sup_{0\leqslant s\leqslant t}\int_0^s \Big(\sup_{u\leqslant v\leqslant s}\|U^0(s,v)\|^2\Big)\,\mathrm{d}u, \tag{5.7}
$$

where U^0 is the principal solution of $\varepsilon\dot\zeta = A^0(z_t^{\det}, y_t^{\det}, \varepsilon)\zeta$.

Lemma 5.3. *Let ξ_0^- satisfy $\langle \xi_0^-, \overline{X}^-(y_0,z_0)^{-1}\xi_0^-\rangle \leqslant h^2$. Then*

$$
\begin{aligned}
\mathbb{P}^{0,(\xi_0^-,z_0,y_0)}&\Big\{\sup_{0\leqslant s\leqslant t\wedge\tau_{\mathcal{B}^-(h)}\wedge\tau_\eta} \|z_s - z_s^{\det}\| \geqslant h_z\Big\} \\
&\leqslant 2\left\lceil \frac{t}{\Delta\varepsilon}\right\rceil \mathrm{e}^{q/4} \exp\Big\{-\kappa_0\frac{\varepsilon h_z^2(1-\mathcal{O}(\Delta\varepsilon))}{\sigma^2\chi_z^{(2)}(t)}\Big[1-\mathcal{O}\Big(\chi_z^{(1)}(t)\,h_z\Big(1+\frac{h^2}{h_z^2}+\frac{h_\eta^2}{h_z^2}\Big)\Big)\Big]\Big\}.
\end{aligned} \tag{5.8}
$$

PROOF: The proof is almost identical with the proof of Lemma 3.4 and Lemma 5.2, with σ' replaced by $\sigma/\sqrt{\varepsilon}$ and V replaced by U^0. □

PROOF OF THEOREM 2.7. We can repeat the proof of Corollary 3.10 in Section 3.2, comparing the process to different deterministic solutions on successive time intervals of length

T. The only difference lies in new values for the exponents $\kappa^+(0)$ (resulting from Proposition 5.1) and κ'. In fact, choosing h_η proportional to h, h_z proportional to $(\chi_z^{(2)}(T)/\varepsilon)^{1/2}h$ and, finally, μ^2 proportional to $1+(h_\eta+h_z)/h$, shows that

$$\mathbb{P}^{0,(\xi_0^-,z_0,y_0)}\Big\{\sup_{0\leqslant s\leqslant T\wedge\tau_{\mathcal{N}}}\langle\xi_s^-,\overline{X}^-(y_s,z_s)^{-1}\xi_s^-\rangle\geqslant h^2\Big\}\leqslant C_{n,m,q,\gamma}(T,\varepsilon)\,\mathrm{e}^{-\kappa^+(\alpha)h^2/\sigma^2}, \tag{5.9}$$

valid for all ξ_0^- satisfying $\langle\xi_0^-,\overline{X}^-(y_0,z_0)^{-1}\xi_0^-\rangle\leqslant\alpha^2h^2$ and all T of order 1 at most. Here

$$C_{n,m,q,\gamma}(T,\varepsilon)=\left\lceil\frac{T}{\Delta\varepsilon}\right\rceil\left[\frac{1}{(1-2\gamma)^{(n-q)/2}}+\mathrm{e}^{(n-q)/4}+2\,\mathrm{e}^{m/4}+2\,\mathrm{e}^{q/4}\right], \tag{5.10}$$

$$\kappa^+(\alpha)=\gamma\big[1-\alpha^2-\mathcal{O}(\Delta)-\mathcal{O}(h)-\mathcal{O}(h(\chi^{(2)}(T)\vee\chi_z^{(2)}(T))^{1/2}/\sqrt{\varepsilon})\big]. \tag{5.11}$$

Similar arguments as in the proof of Lemma 3.7 yield a bound of the form

$$\mathbb{P}^{0,(\xi_0^-,z_0,y_0)}\Big\{\langle\xi_T^-,\overline{X}^-(y_T,z_T)^{-1}\xi_T^-\rangle,\tau_{\mathcal{N}}\geqslant T\Big\}\leqslant\widehat{C}\,\mathrm{e}^{-\kappa'h^2/\sigma^2}, \tag{5.12}$$

where

$$\kappa'=\gamma\big[1-\mathcal{O}(\Delta)-\mathcal{O}(h)-\mathcal{O}(h(\chi^{(2)}(T)\vee\chi_z^{(2)}(T))^{1/2}/\sqrt{\varepsilon})-\mathcal{O}(\mathrm{e}^{-2K_0T/\varepsilon}/(1-2\gamma))\big]. \tag{5.13}$$

In order for the estimates (5.9) and (5.12) to be useful, we need to take T of order ε. However, this leads to an error term of order 1 in the exponent κ', which is due to the fact that ξ_t^- has too little time to relax to the adiabatic manifold. In order to find the best compromise, we take $T=\theta\varepsilon$ and optimize over θ. Assume we are in the worst case, when $\|U^0\|$ grows exponentially like $\mathrm{e}^{K_+t/\varepsilon}$. Then $\chi_z^{(2)}(T)$ is of the order $\varepsilon\theta\,\mathrm{e}^{2K_+\theta}$. The choice

$$\mathrm{e}^{-\theta}=\big[h(1-2\gamma)\big]^{1/(2K_0+K_+)} \tag{5.14}$$

yields an almost optimal error term of order $h^\nu(1-2\gamma)^{1-\nu}|\log(h(1-2\gamma))|^{1/2}$, with $\nu=2K_0/(2K_0+K_+)$. The smaller K_+, i. e., the slower $\chi_z^{(2)}(t)$ grows, the closer ν is to one. □

5.2 The reduced system

Given the SDE (5.1), we call

$$\begin{aligned}\mathrm{d}z_t^0&=\frac{1}{\varepsilon}\hat{f}^0(0,z_t^0,y_t^0,\varepsilon)\,\mathrm{d}t+\frac{\sigma}{\sqrt{\varepsilon}}\widehat{F}^0(0,z_t^0,y_t^0,\varepsilon)\,\mathrm{d}W_t,\\ \mathrm{d}y_t^0&=\hat{g}(0,z_t^0,y_t^0,\varepsilon)\,\mathrm{d}t+\sigma'\widehat{G}(0,z_t^0,y_t^0,\varepsilon)\,\mathrm{d}W_t\end{aligned} \tag{5.15}$$

the *reduced system* of (5.1). It is obtained by setting $\xi_t^-=0$. Let $\zeta_t^0=(z_t^0,y_t^0)$ and $\zeta_t=(z_t-z_t^0,y_t-y_t^0)$. Subtracting (5.15) from (5.1) and making a Taylor expansion of the drift coefficient, we find that (ξ_t^-,ζ_t) obeys the SDE

$$\begin{aligned}\mathrm{d}\xi_t^-&=\frac{1}{\varepsilon}\big[A^-(\zeta_t^0,\varepsilon)\xi_t^-+b(\xi_t^-,\zeta_t,\zeta_t^0,\varepsilon)\big]\,\mathrm{d}t+\frac{\sigma}{\sqrt{\varepsilon}}\widetilde{F}(\xi_t^-,\zeta_t,\zeta_t^0,\varepsilon)\,\mathrm{d}W_t,\\ \mathrm{d}\zeta_t&=\frac{1}{\varepsilon}\big[C(\zeta_t^0,\varepsilon)\xi_t^-+B(\zeta_t^0,\varepsilon)\zeta_t+c(\xi_t^-,\zeta_t,\zeta_t^0,\varepsilon)\big]\,\mathrm{d}t+\frac{\sigma}{\sqrt{\varepsilon}}\widetilde{\mathcal{G}}(\xi_t^-,\zeta_t,\zeta_t^0,\varepsilon)\,\mathrm{d}W_t,\end{aligned} \tag{5.16}$$

where $\|b\|$ and $\|c\|$ are both of order $\|\xi^-\|^2+\|\zeta\|^2$ and $\|\widetilde{\mathcal{G}}\|$ is of order $\|\xi^-\|+\|\zeta\|$, while $\|\widetilde{F}\|$ is bounded. The matrices A^-, B and C are those defined in (2.66), (2.76) and (2.77).

For a given continuous sample path $\{\zeta_t^0(\omega)\}_{t\geqslant 0}$ of (5.16), we denote by U_ω and $\mathcal{V}_\omega$ the principal solutions of $\varepsilon\dot{\xi}^- = A^-(\zeta_t^0(\omega),\varepsilon)\xi^-$ and $\varepsilon\dot{\zeta} = B(\zeta_t^0(\omega),\varepsilon)\zeta$. If we further define

$$\mathcal{S}_\omega(t,s) = \frac{1}{\varepsilon}\int_s^t \mathcal{V}_\omega(t,u)C(\zeta_u^0(\omega),\varepsilon)U_\omega(u,s)\,\mathrm{d}u, \tag{5.17}$$

we can write the solution of (5.16) as

$$\begin{aligned}
\zeta_t(\omega) = {}& \frac{\sigma}{\sqrt{\varepsilon}}\int_0^t \mathcal{V}_\omega(t,s)\widetilde{\mathcal{G}}(\xi_s^-(\omega),\zeta_s(\omega),\zeta_s^0(\omega),\varepsilon)\,\mathrm{d}W_s(\omega)\\
&+ \frac{\sigma}{\sqrt{\varepsilon}}\int_0^t \mathcal{S}_\omega(t,s)\widetilde{F}(\xi_s^-(\omega),\zeta_s(\omega),\zeta_s^0(\omega),\varepsilon)\,\mathrm{d}W_s(\omega)\\
&+ \frac{1}{\varepsilon}\int_0^t \mathcal{V}_\omega(t,s)c(\xi_s^-(\omega),\zeta_s(\omega),\zeta_s^0(\omega),\varepsilon)\,\mathrm{d}s\\
&+ \frac{1}{\varepsilon}\int_0^t \mathcal{S}_\omega(t,s)b(\xi_s^-(\omega),\zeta_s(\omega),\zeta_s^0(\omega),\varepsilon)\,\mathrm{d}s.
\end{aligned} \tag{5.18}$$

Concerning the first two summands in (5.18), note that the identities

$$\begin{aligned}
\mathcal{V}_\omega(t,s) &= \mathcal{V}_\omega(t,0)\mathcal{V}_\omega(s,0)^{-1},\\
\mathcal{S}_\omega(t,s) &= \mathcal{S}_\omega(t,0)U_\omega(s,0)^{-1} + \mathcal{V}_\omega(t,0)\mathcal{S}_\omega(s,0)^{-1}
\end{aligned} \tag{5.19}$$

allow to rewrite the stochastic integrals in such a way that the integrands are adapted with respect to the filtration generated by $\{W_s\}_{s\geqslant 0}$.

We now assume the existence of a stopping time $\tau \leqslant \tau_{\mathcal{B}^-(h)}$ and deterministic functions $\vartheta(t,s)$, $\vartheta_C(t,s)$ such that

$$\begin{aligned}
\bigl\|\mathcal{V}_\omega(t,s)\bigr\| &\leqslant \vartheta(t,s),\\
\bigl\|\mathcal{V}_\omega(t,s)C(\zeta_s^0(\omega),\varepsilon)\bigr\| &\leqslant \vartheta_C(t,s),
\end{aligned} \tag{5.20}$$

uniformly in ε, whenever $s \leqslant t \leqslant \tau(\omega)$, and define

$$\chi^{(i)}(t) = \sup_{0\leqslant s\leqslant t}\frac{1}{\varepsilon}\int_0^s \vartheta(s,u)^i\,\mathrm{d}u, \qquad i=1,2, \tag{5.21}$$

$$\chi_C^{(i)}(t) = \sup_{0\leqslant s\leqslant t}\frac{1}{\varepsilon}\int_0^s \Bigl(\sup_{u\leqslant v\leqslant s}\vartheta_C(s,v)^i\Bigr)\,\mathrm{d}u, \qquad i=1,2. \tag{5.22}$$

The following proposition establishes a local version of Theorem 2.8.

Proposition 5.4. *Let Δ be sufficiently small, fix times $s<t$ such that $t-s=\Delta\varepsilon$, and assume that there exists a constant $\vartheta_0>0$ such that $\vartheta(u,s)\leqslant\vartheta_0$ and $\vartheta_C(u,s)\leqslant\vartheta_0$, whenever $u\in[s,t]$. Then there exist constants $\kappa_0,h_0>0$ such that for all $h\leqslant h_0[\chi^{(1)}(t)\vee\chi_C^{(1)}(t)]^{-1}$,*

$$\mathbb{P}^{0,0}\Bigl\{\sup_{s\wedge\tau\leqslant u<t\wedge\tau}\|\zeta_u\|\geqslant h\Bigr\} \leqslant 2\,\mathrm{e}^{(m+q)/4}\exp\Bigl\{-\kappa_0\frac{h^2}{\sigma^2}\frac{1}{\chi_C^{(2)}(t)+h^2\chi^{(2)}(t)}\Bigr\}. \tag{5.23}$$

Proof: The proof follows along the lines of the proof of Lemma 3.4, the main difference lying in the fact that the stochastic integrals in (5.18) involve the principal solutions U_ω, $\mathcal{V}_\omega$ depending on the realization of the process. However, the existence of the deterministic bound (5.20) allows for a similar conclusion. In particular, the first and second term in (5.18) create respective contributions of the form

$$\mathrm{e}^{(m+q)/4} \exp\Big\{-\frac{H_0^2}{16\sigma^2 h^2 M_1^2 \chi^{(2)}(t)}\Big\} \tag{5.24}$$

$$\mathrm{e}^{(m+q)/4} \exp\Big\{-\frac{H_1^2}{16\sigma^2 M_1^2 \chi_C^{(2)}(t)}\Big\} \tag{5.25}$$

to the probability (5.23). The third and fourth term only cause corrections of order $h\chi^{(1)}(t)$ and $h\chi_C^{(1)}(t)$ in the exponent. □

Now Theorem 2.8 follows from Proposition 5.4 by using a partition of the interval $[0,t]$ into smaller intervals of length $\Delta\varepsilon$.

References

[1] L. Arnold. *Random Dynamical Systems.* Springer-Verlag, Berlin, 1998.

[2] R. Azencott. Petites perturbations aléatoires des systèmes dynamiques: développements asymptotiques. *Bull. Sci. Math. (2)*, 109:253–308, 1985.

[3] R. Bellman. *Introduction to Matrix Analysis.* McGraw–Hill, New York, 1960.

[4] E. Benoît, editor. *Dynamic Bifurcations*, Berlin, 1991. Springer-Verlag.

[5] N. Berglund and B. Gentz. A sample-paths approach to noise-induced synchronization: Stochastic resonance in a double-well potential. To appear in Ann. Appl. Probab. Available at `http://arXiv.org/abs/math.PR/0012267`, 2000.

[6] N. Berglund and B. Gentz. The effect of additive noise on dynamical hysteresis. *Nonlinearity*, 15(3):605–632, 2002.

[7] N. Berglund and B. Gentz. Pathwise description of dynamic pitchfork bifurcations with additive noise. *Probab. Theory Related Fields*, 122(3):341–388, 2002. DOI 10.1007/s004400100174.

[8] H. Crauel and F. Flandoli. Attractors for random dynamical systems. *Probab. Theory Related Fields*, 100(3):365–393, 1994.

[9] M. V. Day. On the exponential exit law in the small parameter exit problem. *Stochastics*, 8:297–323, 1983.

[10] M. V. Day. On the exit law from saddle points. *Stochastic Process. Appl.*, 60:287–311, 1995.

[11] N. Fenichel. Geometric singular perturbation theory for ordinary differential equations. *J. Differential Equations*, 31(1):53–98, 1979.

[12] W. H. Fleming and M. R. James. Asymptotic series and exit time probabilities. *Ann. Probab.*, 20(3):1369–1384, 1992.

[13] M. I. Freidlin. On stable oscillations and equilibriums induced by small noise. *J. Stat. Phys.*, 103:283–300, 2001.

[14] M. I. Freidlin and A. D. Wentzell. *Random Perturbations of Dynamical Systems.* Springer-Verlag, New York, second edition, 1998.

[15] I. S. Gradšteĭn. Application of A. M. Lyapunov's theory of stability to the theory of differential equations with small coefficients in the derivatives. *Mat. Sbornik N. S.*, 32(74):263–286, 1953.

[16] R. Haberman. Slowly varying jump and transition phenomena associated with algebraic bifurcation problems. *SIAM J. Appl. Math.*, 37(1):69–106, 1979.

[17] C. K. R. T. Jones. Geometric singular perturbation theory. In *Dynamical systems (Montecatini Terme, 1994)*, pages 44–118. Springer, Berlin, 1995.

[18] Y. Kifer. The exit problem for small random perturbations of dynamical systems with a hyperbolic fixed point. *Israel J. Math.*, 40(1):74–96, 1981.

[19] N. R. Lebovitz and R. J. Schaar. Exchange of stabilities in autonomous systems. *Studies in Appl. Math.*, 54(3):229–260, 1975.

[20] N. R. Lebovitz and R. J. Schaar. Exchange of stabilities in autonomous systems. II. Vertical bifurcation. *Studies in Appl. Math.*, 56(1):1–50, 1976/77.

[21] E. F. Mishchenko and N. K. Rozov. *Differential equations with small parameters and relaxation oscillations.* Plenum Press, New York, 1980. Translated from the Russian by F. M. C. Goodspeed.

[22] A. I. Neĭshtadt. Persistence of stability loss for dynamical bifurcations I. *Differential Equations*, 23:1385–1391, 1987.

[23] A. I. Neĭshtadt. Persistence of stability loss for dynamical bifurcations II. *Differential Equations*, 24:171–176, 1988.

[24] L. S. Pontryagin and L. V. Rodygin. Approximate solution of a system of ordinary differential equations involving a small parameter in the derivatives. *Soviet Math. Dokl.*, 1:237–240, 1960.

[25] B. Schmalfuß. Invariant attracting sets of nonlinear stochastic differential equations. In H. Langer and V. Nollau, editors, *Markov processes and control theory*, volume 54 of *Math. Res.*, pages 217–228, Berlin, 1989. Akademie-Verlag. Gaußig, 1988.

[26] A. N. Tihonov. Systems of differential equations containing small parameters in the derivatives. *Mat. Sbornik N. S.*, 31:575–586, 1952.

Nils Berglund
Department of Mathematics, ETH Zürich
ETH Zentrum, 8092 Zürich, Switzerland
and
FRUMAM, CPT–CNRS Luminy
Case 907, 13288 Marseille Cedex 9, France
and
PHYMAT, Université de Toulon
B. P. 132, 83957 La Garde Cedex, France
E-mail address: berglund@cpt.univ-mrs.fr

Barbara Gentz
Weierstrass Institute for Applied Analysis and Stochastics
Mohrenstraße 39, 10117 Berlin, Germany
E-mail address: gentz@wias-berlin.de