Probab. Theory Relat. Fields (2011) 150:1–42
DOI 10.1007/s00440-010-0264-0

Noisy heteroclinic networks

Yuri Bakhtin

Received: 23 December 2007 / Revised: 15 January 2010 / Published online: 9 February 2010

Abstract We consider a white noise perturbation of dynamics in the neighborhood of a heteroclinic network. We show that under the logarithmic time rescaling the diffusion converges in distribution in a special topology to a piecewise constant process that jumps between saddle points along the heteroclinic orbits of the network. We also obtain precise asymptotics for the exit measure for a domain containing the starting point of the diffusion.

Mathematics Subject Classification (2000) 60J60 · 34E10 · 60F17

1 Introduction

In this note, we study small noise perturbations of a smooth continuous time dynamical system in the neighborhood of its heteroclinic network.

The deterministic dynamics is defined on $\mathbb{R}^d$ as the flow $(S^t)_{t\in\mathbb{R}}$ generated by a smooth vector field $b : \mathbb{R}^d \to \mathbb{R}^d$, i.e. $S^t x_0$ is the solution of the initial-value problem

$$\begin{aligned} \dot{x}(t) &= b(x(t)), \\ x(0) &= x_0. \end{aligned} \tag{1.1}$$

We assume that the vector field b generates a heteroclinic network, that is a set of isolated critical points connected by heteroclinic orbits of the flow S. Heteroclinic orbits arise naturally in systems with symmetries. Moreover, they are often robust under perturbations of the system preserving the symmetries, see the survey [9] and references therein, for numerous examples and a discussion of mechanisms of robustness.

Yu. Bakhtin (✉)
School of Mathematics, Georgia Tech, Atlanta, GA 30332-0160, USA
e-mail: bakhtin@math.gatech.edu

We consider the system (1.1) perturbed by uniformly elliptic noise:

$$\begin{aligned} dX_\varepsilon(t) &= b(X_\varepsilon(t))dt + \varepsilon\sigma(X_\varepsilon(t))dW(t), \\ X_\varepsilon(0) &= x_0, \end{aligned} \tag{1.2}$$

where W is a standard d-dimensional Wiener process, $\sigma(x)$ is a nondegenerate matrix of diffusion coefficients for every x, and $\varepsilon > 0$ is a small number. The initial point x_0 is assumed to belong to one of the heteroclinic orbits.

Our principal result on the vanishing noise intensity asymptotics can be informally stated as follows.

Theorem *Under some technical nondegeneracy assumptions, as $\varepsilon \to 0$, the process $(X_\varepsilon(t\ln(\varepsilon^{-1})))_{t\geq 0}$ converges in distribution in an appropriate topology to a process that spends all the time on the set of critical points and jumps instantaneously between them along the heteroclinic trajectories.*

In fact, we shall provide much more detailed information on the limiting process and describe its distribution precisely. Thus, our result provides a unified and mathematically rigorous background for the existing phenomenological studies, see e.g. [1]. In particular, we shall see that in many cases the limiting process is not Markov. The precise description of the limiting process allows to obtain asymptotics of the exit distribution for domains containing the starting point. This asymptotic result is of a different kind than the one provided by the classical Freidlin–Wentzell (FW) theory, see [5]. In fact, it allows to compute precisely the limiting probabilities of specific exit points that are indistinguishable from the point of view of the FW quasi-potential.

To prove the main result, we have to trace the evolution of the process along the heteroclinic orbits and in the neighborhood of hyperbolic critical points. The latter was studied in [2,8], where it was demonstrated that in the vanishing diffusion limit, of all possible directions in the unstable manifold the system chooses to evolve along the invariant curve associated to the highest eigenvalue of the linearization of the system at the critical point, and the asymptotics for the exit time was obtained. However, these results are not sufficient for the derivation of our main theorem, and more detailed analysis is required.

The paper is organized as follows. In Sect. 2, we study a simple example of a heteroclinic network. In Sect. 3, we give non-rigorous analysis of the behavior of the diffusion in the neighborhood of one saddle point. The general setting and the main result on weak convergence are given in Sect. 5. To state the result we need to define in what sense the convergence is understood. Therefore, we begin our exposition in Sect. 4 with a brief description of the relevant metric space, postponing the proofs of all technical statements concerning the metric space till Sect. 12. In Sect. 6, we state a result on the exit asymptotics and derive it from the main result of Sect. 5. In Sect. 7, we give the statement of the central technical lemma and use it to prove the main result. The proof of the central lemma is split into two parts. They are given in Sect. 8 and Sect. 9, respectively. Proofs of some auxiliary statements used in Sect. 8, are given in Sect. 11. Section 10 is devoted to an informal discussion of implications of our results and their extensions.

2 A simple example

Here we recall a simple example of a heteroclinic network described in [9]. If $a_1 < 0$, the deterministic cubic system defined by the drift of the following stochastic system

$$\begin{aligned} dX_{\varepsilon,1} &= X_{1,\varepsilon}(1 + a_1 X_{1,\varepsilon}^2 + a_2 X_{\varepsilon,2}^2 + a_3 X_{\varepsilon,3}^2)dt + \varepsilon dW_1, \\ dX_{\varepsilon,2} &= X_{2,\varepsilon}(1 + a_1 X_{\varepsilon,2}^2 + a_2 X_{\varepsilon,3}^2 + a_3 X_{\varepsilon,1}^2)dt + \varepsilon dW_2, \\ dX_{\varepsilon,3} &= X_{3,\varepsilon}(1 + a_1 X_{\varepsilon,3}^2 + a_2 X_{\varepsilon,1}^2 + a_3 X_{\varepsilon,2}^2)dt + \varepsilon dW_3 \end{aligned} \tag{2.1}$$

has six critical points

$$\begin{aligned} z_1^{\pm} &= (\pm\sqrt{-1/a_1}, 0, 0), \\ z_2^{\pm} &= (0, \pm\sqrt{-1/a_1}, 0), \\ z_3^{\pm} &= (0, 0, \pm\sqrt{-1/a_1}). \end{aligned}$$

The matrix of the linearization of the system at z_1^+ is given by

$$\mathrm{diag}\left(-2, \frac{a_1 - a_3}{a_1}, \frac{a_1 - a_2}{a_1}\right),$$

and the linearizations at all other critical points are can be obtained from it using the symmetries of the system. We see that if

$$a_3 > a_1 \quad \text{and} \quad a_2 < a_1, \tag{2.2}$$

then all the critical points are saddles with one unstable direction corresponding to the eigenvalue $(a_1 - a_3)/a_1$. It is shown in [9] that system (2.1) admits 12 orbits connecting $z_1^{\pm}$ to $z_2^{\pm}$, $z_2^{\pm}$ to $z_3^{\pm}$, and $z_3^{\pm}$ to $z_1^{\pm}$. Each of these orbits lies entirely in one of the coordinate planes.

Let us equip (2.1) with an initial condition x_0 on one of the 12 heteroclinic connections, say, on the one connecting z_1^+ to z_2^+ denoted by $z_1^+ \to z_2^+$.

The theory developed in this paper allows to describe precisely the limiting behavior of the process as $\varepsilon \to 0$. Namely, the process X_ε will stay close to the heteroclinic network, moving mostly along the heteroclinic connections between the saddle points. At each saddle point it spends a time of order $\ln(\varepsilon^{-1})$. More precisely, the process Z_ε defined by $Z_\varepsilon(t) = X_\varepsilon(t \ln(\varepsilon^{-1}))$ converges to a process that jumps from x_0 to z_2^+ instantaneously along the heteroclinic connection $z_1^+ \to z_2^+$, sits at z_2^+ for some time, then chooses one of the orbits $z_2^+ \to z_3^+$ and $z_2^+ \to z_3^-$, and jumps *along that orbit* instantaneously, spends some time in the endpoint of that orbit and then chooses a new outgoing orbit to follow, etc. However, the details of the process depend crucially on the eigenvalues of the linearization at the critical point.

If at each saddle point the contraction is stronger than expansion, i.e. if

$$\frac{a_1 - a_2}{a_1} < -\frac{a_1 - a_3}{a_1},$$

which is equivalent to $a_2 + a_3 < 2a_1$, then the system exhibits loss of memory and the sequence of saddle points visited by the limiting process is a standard random walk on the directed graph formed by the network, i.e. a Markov chain on saddle points that at each point chooses to jump along one of the two possible outgoing connections with probability $1/2$.

If the expansion is stronger than the contraction, i.e. $a_2 + a_3 > 2a_1$, then at z_2^+, the first saddle point visited, the process still chooses each of the two possible next heteroclinic connections with probability $1/2$. However, if it chooses $z_2^+ \to z_3^+$, then it will cycle through z_1^+, z_2^+, z_3^+ and never visit any other critical points. If it chooses $z_2^+ \to z_3^-$, then it will cycle through z_1^+, z_2^+, z_3^- and never visit any other critical points. This situation is strongly non-Markovian, because the choice the system makes at z_2^+ at any time is determined by its choice during the first visit to that saddle point.

The case where $a_2 + a_3 = 2a_1$ combines certain features of the situations described above. The limiting random saddle point sequence explores all heteroclinic connections, but it makes asymmetric choices determined by its the history, also producing non-Markov dynamics.

Besides the probability structure of limiting saddle point sequences, our main result also provides a description of the random times spent by the process at each of the saddles it visits.

3 Non-rigorous analysis of a linear system

In this section we consider the diffusion near a saddle point in the simplest possible case:

$$
\begin{aligned}
dX_{\varepsilon,1} &= \lambda_1 X_{\varepsilon,1} dt + \varepsilon dW_1,\\
dX_{\varepsilon,2} &= \lambda_2 X_{\varepsilon,2} dt + \varepsilon dW_2
\end{aligned}
$$

with initial conditions $X_{\varepsilon,1} = 0$, $X_{\varepsilon,2} = 1$. Here $\lambda_1 > 0 > \lambda_2$, and W_1, W_2 are i.i.d. standard Wiener processes.

Let us study the exit distribution of X_ε for the strip $\{(x_1, x_2) : |x_1| \le 1\}$. In other words, we are interested in the distribution of $X_\varepsilon(\tau_\varepsilon)$, where $\tau_\varepsilon = \inf\{t : |X_{\varepsilon,1}(t)| = 1\}$. Duhamel's principle implies

$$
X_{\varepsilon,1}(t) = \varepsilon e^{\lambda_1 t} \int_0^t e^{-\lambda_1 s} dW_1(s).
$$

The integral in the r.h.s. converges a.s. to a centered Gaussian r.v. N_1, so that for large t,

$$
|X_{\varepsilon,1}(t)| \approx \varepsilon e^{\lambda_1 t} N_1.
$$

Therefore, for the exit time τ_ε, we have to solve

$$1 \approx \varepsilon e^{\lambda_1 t}|N_1|,$$

so that

$$\tau_\varepsilon \approx \frac{1}{\lambda_1}\ln(\varepsilon|N_1|)^{-1}.$$

So, the time spent by the diffusion in the neighborhood of the saddle point is about $\lambda^{-1}\ln(\varepsilon^{-1})$.

On the other hand,

$$X_{\varepsilon,2}(t) = e^{\lambda_2 t} + \varepsilon\int_0^t e^{\lambda_2(t-s)}dW_2(s).$$

and the integral in the r.h.s. converges in distribution to N_2, a centered Gaussian r.v. Plugging the expression for τ_ε into this relation, we get

$$X_{\varepsilon,2}(\tau_\varepsilon) \approx \varepsilon^{-\lambda_2/\lambda_1}|N_1|^{-\lambda_2/\lambda_1} + \varepsilon N_2 \approx \begin{cases} \varepsilon N_2, & \lambda_2 < -\lambda_1, \\ \varepsilon^{-\lambda_2/\lambda_1}|N_1|^{-\lambda_2/\lambda_1}, & \lambda_2 > -\lambda_1, \\ \varepsilon(|N_1| + N_2), & \lambda_2 = -\lambda_1. \end{cases}$$

Therefore, when the contraction is stronger than expansion ($\lambda_2 < -\lambda_1$), the limiting exit distribution is centered Gaussian. In the opposite case ($\lambda_2 > -\lambda_1$), the limiting exit distribution is strongly asymmetric and concentrated on the positive semiline reflecting the fact that the initial condition $X_{\varepsilon,2}(0) = 1$ was positive and presenting a strong memory effect. In the intermediate case ($-\lambda_2 = \lambda_1$), the limit is the distribution of the sum of a symmetric r.v. N_2 and a positive r.v. $|N_1|$, and the resulting asymmetry also serves as a basis for a certain memory effect.

In general, the asymmetry in the exit distribution means that at the next visited saddle point the choices of the exit direction will not be symmetric, thus leading to non-Markovian dynamics. Notice that three types of behavior for the linear system that we just derived correspond to the three types of cycling through the saddle points in the example of Sect. 2.

One of the main goals of this paper is the precise mathematical meaning of the approximate identities of this section and their generalizations to multiplicative white noise perturbations of nonlinear dynamics in higher dimensions.

4 Convergence of graphs in space-time

The main result of this paper states the convergence of a family of continuous processes to a process with jumps. This kind of convergence is impossible in the traditionally used Skorokhod topology on the space D of processes with left and right limits, since the set of all continuous functions is closed in the Skorokhod topology, see [3].

In this section we replace D by another extension of the space of continuous functions. This extension allows to describe not only the fact of an instantaneous jump from one point to another, but also the curve along which the jump is made. We shall also introduce an appropriate topology to characterize the convergence in this new space.

It is interesting that in his classical paper [11] Skorokhod introduced several topologies for trajectories with jumps. Only one of them is widely known as the Skorokhod topology now. However, the construction that we are going to describe here did not appear either in [11], or anywhere else, to the best of our knowledge, at least in the literature on stochastic processes.

We consider all continuous functions ("paths")

$$\gamma : [0, 1] \to [0, \infty) \times \mathbb{R}^d$$

such that $\gamma^0(s)$ is nondecreasing in s. Here (and often in this paper) we use superscripts to denote coordinates: $\gamma = (\gamma^0, \gamma^1, \ldots, \gamma^d)$.

We say that two paths γ_1 and γ_2 are equivalent, and write $\gamma_1 \sim \gamma_2$ if there is a path γ^* and nondecreasing surjective functions $\lambda_1, \lambda_2 : [0, 1] \to [0, 1]$ with $\gamma_1(s) = \gamma^* \circ \lambda_1(s)$ and $\gamma_2(s) = \gamma^* \circ \lambda_2(s)$ for all $s \in [0, 1]$, where $\circ$ means the composition of two functions. (These are essentially reparametrizations of the path γ^* except that we allow λ_1, λ_2 to be not strictly monotone.) In Sect. 12 we shall prove the following statement:

Lemma 4.1 *The relation $\sim$ on paths is a well-defined equivalence relation.*

Any non-empty class of equivalent paths will be called a curve. We denote the set of all curves by $\mathbb{X}$. Our choice of the equivalence relation ensures that each curve is a closed set in sup-norm (see Sect. 12), and we shall be able to introduce a metric on $\mathbb{X}$ induced by the sup-norm.

Since each curve in $\mathbb{X}$ is nondecreasing in the zeroth coordinate which plays the role of time, it can be thought of as the graph of a function from $[0, T]$ to $\mathbb{R}^d$ for some nonnegative T. However, any value $t \in [0, T]$ can be attained by a path's "time" coordinate for a whole interval of values of the variable parametrizing the curve, thus defining a curve in $\{t\} \times \mathbb{R}^d$, which is interpreted as the curve along which the jump at time t is made.

We would like to introduce a distance in $\mathbb{X}$ that would be sensitive to the geometry of jump curves, but not to their parametrization. So, for two curves $\Gamma_1, \Gamma_2 \in \mathbb{X}$, we denote

$$\rho(\Gamma_1, \Gamma_2) = \inf_{\gamma_1 \in \Gamma_1, \gamma_2 \in \Gamma_2} \sup_{s \in [0,1]} |\gamma_1(s) - \gamma_2(s)|, \tag{4.1}$$

where $|\cdot|$ denotes the Euclidean norm in $[0, T] \times \mathbb{R}^d$.

Theorem 4.1 (1) *The function ρ defined above is a metric on $\mathbb{X}$.*
(2) *The space $(\mathbb{X}, \rho)$ is Polish (i.e. complete and separable).*

We postpone the proof of this statement to Sect. 12.

Naturally, any continuous function $f : [0, T] \to \mathbb{R}^d$ defines a path γ_f by

$$\gamma_f(t) = \Big(tT, f^1(tT), f^2(tT), \dots, f^d(tT)\Big), \quad t \in [0, 1], \tag{4.2}$$

and a curve Γ_f that is the equivalence class of γ_f.

The following result shows that the convergence of continuous functions in sup-norm is consistent with convergence of the associated curves in metric ρ.

Lemma 4.2 *Let $(f_n)_{n\in\mathbb{N}}$ and g be continuous functions on $[0, T]$ for some $T > 0$. Then*

$$\sup_{t\in[0,T]} |f_n(t) - g(t)| \to 0, \quad n \to \infty$$

is necessary and sufficient for

$$\rho(\Gamma_{f_n}, \Gamma_g) \to 0, \quad n \to \infty.$$

The proof of this lemma is also given in Sect. 12. In fact, it can be extended to describe the convergence of graphs of functions with varying domains.

We shall need the following notion in the statement of our main result.

An element Γ of $\mathbb{X}$ is called *piecewise constant* if there is a path $\gamma \in \Gamma$, a number $k \in \mathbb{N}$ and families of numbers

$$0 = s_0 \le s_1 \le \cdots \le s_{2k} \le s_{2k+1} = 1,$$
$$0 = t_0 \le t_1 \le \cdots \le t_{k-1} \le t_k,$$

and points

$$y_1, \dots, y_k \in \mathbb{R}^d,$$

such that $(\gamma^1(s), \dots, \gamma^d(s)) = y_j$ for $s \in [s_{2j-1}, s_{2j}]$, $j = 1, \dots, k$, and $\gamma^0(s) = t_j$ for $s \in [s_{2j}, s_{2j+1}]$, $j = 0, \dots, k$. A piecewise constant Γ describes a particle that sits at point y_j between times t_{j-1} and t_j, and at time t_j jumps *along the path* $\gamma_j = (\gamma^1, \dots, \gamma^d)|_{[s_{2j}, s_{2j+1}]}$. It is natural to identify Γ with a sequence of points and jumps, and we write

$$\Gamma = (\gamma_0, y_1, \Delta t_1, \gamma_1, y_2, \Delta t_2, \gamma_2, \dots, y_k, \Delta t_k, \gamma_k), \tag{4.3}$$

where $\Delta t_j = t_j - t_{j-1}$ denotes the time spent by the particle at point y_j.

5 The setting and the main weak convergence result

In this section we describe the setting and state the main result. The conditions of the setting and possible generalizations are discussed in Sect. 10.3.

We assume that the vector field $b : \mathbb{R}^d \to \mathbb{R}^d$ is C^2-smooth, and the $d \times d$-matrix valued function σ is also C^2. We assume that for each x_0, the flow $S^t x_0$ associated to the system (1.1) is well-defined for all $t \in \mathbb{R}$ (including negative values of t). We assume that b admits a heteroclinic network of a special kind that we proceed to describe.

We suppose that there is a finite or countable set of points $(z_i)_{i\in\mathcal{C}}$, where $\mathcal{C} = \mathbb{N}$ or $\mathcal{C} = \{1, \dots, N\}$ for some $N \geq 1$, with the following properties.

(i) For each $i \in \mathcal{C}$, there is a neighborhood U_i of z_i and a $d \times d$ matrix A_i such that

$$b(x) = A_i(x - z_i) + Q_i(x), \quad x \in U_i,$$

where $|Q_i(x)| \leq C_i|x - z_i|^2$, for a constant C_i and every $x \in U_i$. In particular, z_i is a critical point for b, since $b(z_i) = 0$. Moreover, we require that S^t is conjugated on U_i to a linear dynamical system $\dot{y} = A_i y$ by a C^2-diffeomorphism f_i satisfying $f_i(z_i) = 0$. This means that for any $x_0 \in U_i$, there is $t_0 = t_0(x_0) > 0$ such that for all $t \in (-t_0, t_0)$,

$$\frac{d}{dt} f_i(S^t x_0) = A_i f_i(S^t x_0).$$

(ii) For each $i \in \mathcal{C}$, the eigenvalues $\lambda_{i,1}, \dots, \lambda_{i,d}$ of A_i are real and simple, we also assume that there is an integer ν_i with $2 \leq \nu_i \leq d$ such that

$$\lambda_{i,1} > \dots > \lambda_{i,\nu_i-1} > 0 > \lambda_{i,\nu_i} > \dots > \lambda_{i,d}. \tag{5.1}$$

These requirements mean, in particular, that each z_i is a hyperbolic fixed point (saddle) for the dynamics. The Hartman–Grobman theorem (see Theorem 6.3.1 in [7]) guarantees the existence of a homeomorphism conjugating the flow generated by vector field b to linear dynamics. Our condition (i) imposes a stronger requirement for this conjugation to be C^2. This requirement is still often satisfied as follows from the Sternberg linearization theorem for hyperbolic fixed points with no resonances, see Theorem 6.6.6 in [7]. In particular, the cubic system of Sect. 2 is C^∞-conjugated to a linear system at each saddle point for typical values of its parameters.

We also want to make several assumptions on orbits of the flow connecting these saddle points to each other. First, we denote by $v_{i,1}, \dots, v_{i,d}$ the unit eigenvectors associated with the eigenvalues $\lambda_{i,1}, \dots, \lambda_{i,d}$. The hyperbolicity (and, even more straightforwardly, the conjugation to a linear flow) implies that for every $i \in \mathcal{C}$ there is a $d - \nu + 1$-dimensional C^2-manifold $\mathcal{W}_i^s$ containing z_i such that $\lim_{t\to+\infty} S^t x = z_i$ for every $x \in \mathcal{W}_i^s$ (i.e. $\mathcal{W}_i^s$ is the stable manifold associated to z_i.) For each i the unstable manifold of z_i is also well-defined.

However, it is known, see [2,8], that if the initial data for the stochastic flow are close to the stable manifold then after passing the saddle z_i the solution evolves mostly along the invariant manifold associated to the highest eigenvalue of A_i. So, what we need is the curve $\gamma_i \in C^2$ containing z_i, tangent to $v_{i,1}$ at z_i, and invariant under the flow. Of course, the intersection of γ_i with U_i is well-defined and coincides with $f_i^{-1}(\mathrm{span}(v_{i,1}) \cap f_i(U_i))$.

For each $i \in \mathcal{C}$ we denote $g_i = f_i^{-1}$ and set

$$q_i^{\pm} = g_i(\pm R_i v_{i,1}),$$

where the numbers $(R_i)_{i\in\mathcal{C}}$ are chosen so that

$$\tilde{U}_i = \left\{ x : \max_{k=1,\dots,d} |P^{i,k} f_i(x)| \le R_i \right\} \subset U_i,$$

and these sets are mutually disjoint. Here, for any $y \in \mathbb{R}^d$, the number $P^{i,k}y$ is defined by

$$y = \sum_{k=1}^{d} (P^{i,k}y) v_{i,k},$$

and denotes the kth coordinate of y in the coordinate system defined by $v_{i,1}, \dots, v_{i,d}$.

We denote the orbits of $q_i^{\pm}$ by $\gamma_i^{\pm}$, and assume that for each $i \in \mathcal{C}$, there are numbers $n^{\pm}(i) \in \mathcal{C}$ such that

$$\lim_{t\to\infty} S^t q_i^{\pm} = z_{n^{\pm}(i)}.$$

This means that the curves $\gamma_i^{\pm}$ are heteroclinic orbits connecting the saddle point z_i to saddle points $z_{n^{\pm}(i)}$. We do not prohibit these orbits to be homoclinic and connect z_i to itself, i.e. the situations where $n^{\pm}(i) = i$ are allowed.

For any $i \in \mathcal{C}$ we define

$$h_i^{\pm} = \inf\{t : S^t q_i^{\pm} \in \tilde{U}_{n^{\pm}(i)}\}, \tag{5.2}$$

$$x_i^{\pm} = S^{h_i^{\pm}} q_i^{\pm}. \tag{5.3}$$

so that $h_i^{\pm}$ is the time it takes to travel from $q_i^{\pm}$ to the neighborhood of the next saddle, and $x_i^{\pm}$ is the point of entrance to that neighborhood.

Our first nondegeneracy assumption is that for all $i \in \mathcal{C}$,

$$P^{n^{\pm}(i), \nu_{n^{\pm}(i)}} f_{n^{\pm}(i)}(x_i^{\pm}) \neq 0, \tag{5.4}$$

which means that each heteroclinic orbit $\gamma_i^{\pm}$ has a nontrivial component in the direction of the $v_{n^{\pm}(i),\nu_{n^{\pm}(i)}}$ as it approaches $z_{n^{\pm}(i)}$. Although this condition holds true for all systems of interest (e.g. the system considered in Sect. 2), it is easy to adapt our reasoning to the situations where other components of the projection of $f_{n^{\pm}(i)}(x_i^{\pm})$ on the stable directions dominate.

We shall also need a nondegeneracy condition on the linearization of (1.1) along $\gamma_i^{\pm}$. For each $x \in \mathbb{R}^d$ we consider the fundamental matrix $\Phi_x(\cdot)$ solving the equation in variations along the orbit $(S^t x)_{t\geq 0}$:

$$\begin{aligned} \frac{d}{dt}\Phi_x(t) &= Db(S^t x)\Phi_x(t), \quad t \geq 0, \\ \Phi_x(0) &= I. \end{aligned}$$

For all i, j we denote

$$\bar{v}_{i,j}^{\pm} = \Phi_{q_i^{\pm}}(h_i^{\pm})(Df_i(q_i^{\pm}))^{-1} v_{i,j}.$$

The technical nondegeneracy assumption on Φ that we need is:

$$P^{n^{\pm}(i),k} Df_{n^{\pm}(i)}(x_i^{\pm})\bar{v}_{i,j}^{\pm} \neq 0, \quad i \in \mathcal{C}, \quad j \in \{2, \nu_i\},\ k = \begin{cases} 1, 2, & \nu_{n^{\pm}(i)} > 2, \\ 1, & \nu_{n^{\pm}(i)} = 1. \end{cases} \tag{5.5}$$

Again, we work with this condition since it holds true for any system of interest, but it is easy to adapt our reasoning to the situations where this condition is violated.

To formulate our main theorem we need a notion of an *entrance-exit map* describing the limiting behavior of X_ε in the neighborhood of a saddle points, namely, the asymptotics of the random entrance-exit Poincaré map as $\varepsilon \to 0$.

We denote the set of all probability Borel measures on $\mathbb{R}^d$ by $\mathcal{P}(\mathbb{R}^d)$.

Let us denote by In_i the set of all triples (x, α, μ) where

(1) $x \in \tilde{U}_i \cap \mathcal{W}_i^s$ satisfies $P_i^{\nu_i} f_i(x) \neq 0$;
(2) $\alpha \in (0, 1]$;
(3) $\mu \in \mathcal{P}(\mathbb{R}^d)$ with

$$\begin{aligned} \mu\{\phi :\ P_i^1 Df_i(x)\phi \neq 0\} &= 1, \quad \text{if } \alpha < 1, \\ \mu\{\phi :\ P_i^2 Df_i(x)\phi \neq 0\} &= 1, \quad \text{if } \alpha < 1 \text{ and } \nu_i > 2. \end{aligned}$$

This set will be used to describe the initial condition for Eq. (1.2): $X_\varepsilon(0) = x + \varepsilon^\alpha \phi_\varepsilon$, where the distribution of ϕ_ε weakly converges to μ as $\varepsilon \to 0$.

We also define

$$\mathrm{Out} = \{(t, p, x, \beta, F) :\ t \in (0, \infty),\ p \in [0, 1],\ x \in \mathbb{R}^d,\ \beta \in (0, 1], F \in \mathcal{P}(\mathbb{R}^d)\}$$

and

$$\begin{aligned} \widehat{\mathrm{Out}}_i = \{&((t_-, p_-, x_-, \beta_-, F_-), (t_+, p_+, x_+, \beta_+, F_+)) \in \mathrm{Out}^2 :\\ &t_- = t_+,\ x_{\pm} = x_i^{\pm},\ p_- + p_+ = 1,\ \beta_- = \beta_+\}. \end{aligned}$$

Here, the numbers $p_\pm$ define the limiting probabilities of choosing each of the two branches of the invariant curve associated with the highest eigenvalue of the linearization at the saddle point; $x_\pm$ are points on these orbits that serve as entrance points to neighborhoods of the next saddle points; $t_\pm$ are the times it takes to reach these points under the proper (logarithmic) renormalization; β is the scaling exponent so that the exit distribution (serving as the entrance distribution to the next saddle's neighborhood) takes the form $x_\pm + \varepsilon^\beta \psi_\varepsilon$, where the distribution of ψ_ε converges to F_+ or F_- depending on which of the two branches was chosen.

It is possible (see Lemma 7.1) to give a precise description of the asymptotic behavior of the diffusion in the neighborhood of each saddle point in terms of an appropriate entrance-exit map, i.e. a map that for each saddle point computes a description of the exit parameters in terms of the entrance parameters:

$$\Psi_i : \mathrm{In}_i \to \widehat{\mathrm{Out}}_i, \quad i \in \mathcal{C}.$$

For an entrance-exit map $(\Psi_i)_{i\in\mathcal{C}} = (\Psi_{i,-}, \Psi_{i,+})_{i\in\mathcal{C}}$ we shall denote its components by $t_i = t_{i,\pm}$, $p_{i,\pm}$, $x_{i,\pm}$, $\beta_i = \beta_{i,\pm}$, $F_{\pm,i}$.

Suppose x_0 belongs to one of heteroclinic orbits of the network. A sequence $\mathbf{z} = (\theta_0, z_{i_1}, \theta_1, z_{i_2}, \theta_2 \dots, \theta_{k-1}, z_{i_k}, \theta_k)$ is called *admissible* for x_0 if

(1) θ_0 is the positive orbit of x_0 with $\lim_{t\to\infty} S^t x_0 = z_{i_1}$;
(2) for each $j \in \{1, \dots, k\}$, $\theta_j = \gamma_{i_j}^+$ or $\theta_j = \gamma_{i_j}^-$;
(3) for each $j \in \{1, \dots, k-1\}$,

$$i_{j+1} = \begin{cases} n^+(i_j), & \theta_j = \gamma_{i_j}^+, \\ n^-(i_j), & \theta_j = \gamma_{i_j}^-. \end{cases}$$

The number $k = k(\mathbf{z})$ is called the length of $\mathbf{z}$.

Our main limit theorem uses entrance-exit maps to assign limiting probabilities to admissible sequences. Let us proceed to describe this procedure.

With each admissible sequence $\mathbf{z}$ we associate the following sequence:

$$\zeta(\mathbf{z}) = ((\tilde{x}_0, \alpha_0, \mu_0), (t_1, p_1, \tilde{x}_1, \alpha_1, \mu_1), \dots, (t_k, p_k, \tilde{x}_k, \alpha_k, \mu_k)). \tag{5.6}$$

Here $\tilde{x}_0 = S^{\tilde{t}(x_0)} x_0$, $\alpha_0 = 1$, and

$$\mu_0 = \mathrm{Law}\left(\Phi_{x_0}(\tilde{t}(x_0)) \int_0^{\tilde{t}(x_0)} \Phi_{x_0}^{-1}(s)\sigma(S^s x_0) dW(s) \right), \tag{5.7}$$

where

$$\tilde{t}(x_0) = \inf\{t \geq 0 : \ S^t x_0 \in \tilde{U}_{i_1}\} + 1. \tag{5.8}$$

We add 1 in the r.h.s. so that the distribution μ_0 is nondegenerate (and the maps $\Psi_{i_1,\pm}(x_0, \alpha_0, \mu_0)$ are well-defined) even if $x_0 \in \tilde{U}_{i_1}$. All other entries in (5.6) are obtained according to the following recursive procedure. For each j,

$$(t_j, p_j, \tilde{x}_j, \alpha_j, \mu_j) = \begin{cases} \Psi_{i_j,+}(\tilde{x}_{j-1}, \alpha_{j-1}, \mu_{j-1}), & \theta_j = \gamma_{i_j}^+ \\ \Psi_{i_j,-}(\tilde{x}_{j-1}, \alpha_{j-1}, \mu_{j-1}), & \theta_j = \gamma_{i_j}^-. \end{cases} \tag{5.9}$$

The numbers $t_1 = t_1(\mathbf{z}), \ldots, t_k = t_k(\mathbf{z})$ defined above play the role of time, and the admissible sequence $\mathbf{z}$ can be identified with a piecewise constant trajectory $\Gamma(\mathbf{z}) \in \mathbb{X}$:

$$\Gamma(\mathbf{z}) = (\theta_0, z_{i_1}, t_1, \theta_1, z_{i_2}, t_2, \theta_2, \ldots, z_{i_k}, t_k, \theta_k).$$

The numbers $p_1 = p_1(\mathbf{z}), \ldots, p_k = p_k(\mathbf{z})$ defined in (5.6) play the roles of conditional probabilities, and we denote

$$\pi(\mathbf{z}) = p_1(\mathbf{z}) p_2(\mathbf{z}) \ldots p_k(\mathbf{z}). \tag{5.10}$$

The set of all admissible sequences for x_0 has the structure of a binary tree. The natural partial order on it is determined by inclusion. We say that a set L of admissible sequences for x_0 is *free* if no two sequences in L are comparable with respect to this partial order. If additionally, for any sequence not from L, it is comparable to one of the sequences from L, the set L is called *complete*. In the language of graph theory, a complete set is a section of the binary tree.

It is clear that for any free set L, $\pi(L) \le 1$, where $\pi(L) = \sum_{\mathbf{z}\in L} \pi(\mathbf{z})$. A free set L is called *conservative* if $\pi(L) = 1$. Every complete set is finite and conservative.

We are ready to state our main result now.

Theorem 5.1 *Under the conditions stated above there is an entrance-exit map Ψ with the following property.*

Let x_0 belong to one of the heteroclinic orbits of the network. For each $\varepsilon > 0$ define a stochastic process Z_ε by

$$Z_\varepsilon(t) = X_\varepsilon(t \ln(\varepsilon^{-1})), \quad t \ge 0, \tag{5.11}$$

where X_ε is the strong solution of (1.1) with initial condition $X_\varepsilon(0) = x_0$.

For any conservative set L of x_0-admissible sequences, there is a family of stopping times $(T_\varepsilon)_{\varepsilon>0}$ such that the distribution of the graph $\Gamma_{Z_\varepsilon(t), t\le T_\varepsilon}$ converges weakly in $(\mathbb{X}, \rho)$ to the measure $M_{x_0,L}$ concentrated on the set

$$\{\Gamma(\mathbf{z}) \;:\; \mathbf{z} \in L\},$$

and satisfying

$$M_{x_0,L}\{\Gamma(\mathbf{z})\} = \pi(\mathbf{z}), \quad \mathbf{z} \in L, \tag{5.12}$$

where $\pi(\mathbf{z})$ is defined via Ψ in (5.10).

For any entrance-exit map the family of conservative sets includes all finite complete sets, so that the content of Theorem 5.1 is nontrivial.

Importantly, we actually construct the desired entrance-exit map Ψ in the proof. This allows to study the details of the limiting process in Sect. 10. At this point let us just mention that the sequence of saddles visited by the limiting process can be Markov or non-Markov depending on the eigenvalues of the linearizations at the saddle points. We discuss this memory effect and some other implications and possible extensions of Theorem 5.1 in Sect. 10.

6 Exit measure asymptotics

We shall now apply Theorem 5.1 to the exit problem along a heteroclinic network, and formulate a theorem that, in a sense, gives more precise information on the exit distribution than the FW theory. We assume that there is a domain $D \subset \mathbb{R}^d$ with piecewise smooth boundary such that $x_0 \in D$. The FW theory implies that, as $\varepsilon \to 0$, the exit measure for the process X_ε started at x_0 concentrates at points $y \in \partial D$ that provide the minimum value of the so called quasi-potential $V(x_0, y)$. Since for all the points that are reachable from x_0 along the heteroclinic network, the quasi-potential equals 0, we conclude that in the case of heteroclinic networks, the exit measure asymptotically concentrates at the boundary points that can be reached from x_0 along the heteroclinic network. However, this approach does not allow to distinguish between the exit points while ours allows to determine an exact limiting probability for each exit point.

We take a point x_0 on one of the heteroclinic orbits of the network and denote by $L(x_0, D)$ the set of all the admissible sequences $\mathbf{z}$ (of any length k) for x_0 such that the last curve θ_k of the sequence intersects ∂D transversally at a point $q(\mathbf{z})$ (if there are several points of intersection we take the first one with respect to the natural order on θ_k), and θ_j does not intersect ∂D for all $j < k$. Let

$$\tau_\varepsilon(x_0, D) = \inf\{t : X_\varepsilon(t) \in \partial D\}.$$

The distribution of $X_\varepsilon(\tau_\varepsilon(x_0, D))$ is concentrated on ∂D.

For each $\mathbf{z} \in L(x_0, D)$, one can define $\pi(\mathbf{z})$ via (5.10).

Theorem 6.1 *For the setting described above, if the set $L(x_0, D)$ is conservative, then the distribution of $X(\tau_\varepsilon(x_0, D))$ converges weakly, as $\varepsilon \to 0$, to*

$$P_{x_0,D} = \sum_{\mathbf{z}\in L(x_0,D)} \pi(\mathbf{z})\delta_{q(\mathbf{z})}.$$

Proof Let us use Theorem 5.1 to choose the times $(T_\varepsilon)_{\varepsilon>0}$ providing convergence of the distribution of $\Gamma_{Z_\varepsilon(t),t\le T_\varepsilon}$ to $M_{x_0,L(x_0,D)}$. The theorem follows since $X_\varepsilon(\tau_\varepsilon(x_0, D))$ is a functional of $\Gamma_{Z_\varepsilon(t),t\le T_\varepsilon}$, continuous on the support of $M_{x_0,L(x_0,D)}$. □

Remark 6.1 Notice that for different sequences $\mathbf{z}$ and $\mathbf{z}'$ it is possible to have $q(\mathbf{z}) = q(\mathbf{z}')$ so that an exit point can accumulate its limiting probability from a variety of admissible sequences.

Remark 6.2 The behavior of the system up to $\tau_\varepsilon(x_0, D)$ is entirely determined by the drift and diffusion coefficients inside D. Therefore, there is an obvious generalization of this theorem for heteroclinic networks in a domain, where one requires the invariant manifolds associated to the highest eigenvalue at a critical point to connect that critical point either to another critical point, or to a point on ∂D. An advantage of that theorem is that one does not have to specify the (irrelevant) coefficients of (1.2) outside of D. We omit the precise formulation for brevity.

Remark 6.3 In the case of nonconservative $L(x_0, D)$, the limit theorem is harder to formulate. In the limit, the exit happens along the sequences belonging to $L = L(x_0, D)$ with positive probability $\pi(L) < 1$. With probability $1 - \pi(L)$, the exit happens in a more complicated way (and in a longer than logarithmic time) depending on the details of the driving vector field.

7 Proof of Theorem 5.1

We begin with the central lemma that we need in the proof. It has a lengthy statement, and after formulating it, we also give a brief informal explanation.

For each $i \in \mathcal{C}$ we introduce U_i^+ and U_i^- via

$$U_i^{\pm} = \{x \in U : P^{i,1} f_i(x) = \pm R_i\}. \tag{7.1}$$

Lemma 7.1 *For each $i \in \mathcal{C}$, there is a map*

$$\Psi_i = ((t_{i,-}, p_{i,-}, x_{i,-}, \beta_{i,-}, F_{i,-}), (t_{i,+}, p_{i,+}, x_{i,+}, \beta_{i,+}, F_{i,+})) : \mathrm{In}_i \to \widehat{\mathrm{Out}_i}$$

with the following property.

Take any $(x, \alpha, \mu) \in \mathrm{In}_i$ and any family of distributions $(\mu_\varepsilon)_{\varepsilon>0}$ in $\mathcal{P}(\mathbb{R}^d)$ with $\mu_\varepsilon \Rightarrow \mu$ as $\varepsilon \to 0$. For each $\varepsilon > 0$, consider the solution X_ε of (1.2) with initial condition

$$X_\varepsilon(0) = x + \varepsilon^\alpha \phi_\varepsilon, \tag{7.2}$$

where

$$\mathrm{Law}(\phi_\varepsilon) = \mu_\varepsilon, \quad \varepsilon > 0, \tag{7.3}$$

and define a stopping time

$$T_{\mathrm{out},\varepsilon} = \inf\{t \geq 0 : X_\varepsilon(t) \in U_i^+ \cup U_i^-\}, \quad \varepsilon > 0,$$

and two events

$$A_{i,\pm,\varepsilon} = \{X_\varepsilon(T_{\mathrm{out},\varepsilon}) \in U_i^{\pm}\}, \quad \varepsilon > 0.$$

Then

(1) *As* $\varepsilon \to 0$,

$$\frac{T_{\mathrm{out},\varepsilon}}{\ln(\varepsilon^{-1})} \xrightarrow{\mathsf{P}} t_{i,\pm}(x,\alpha,\mu).$$

(2) *As* $\varepsilon \to 0$,

$$\mathsf{P}(A_{i,\pm,\varepsilon}) \to p_{i,\pm}(x,\alpha,\mu).$$

(3) *There is a family of random vectors* $(\psi_{i,\varepsilon})_{\varepsilon>0}$ *such that*
(a) *for every* $\varepsilon > 0$, *on* $A_{i,\pm,\varepsilon}$

$$X_\varepsilon(T_{\mathrm{out},\varepsilon} + h_i^\pm) = x_{i,\pm}(x,\alpha,\mu) + \varepsilon^{\beta_{i,\pm}(x,\alpha,\mu)}\psi_{i,\varepsilon},$$

where $h_i^\pm$ *was defined in* (5.2)*;*
(b) *as* $\varepsilon \to 0$,

$$\mathrm{Law}(\psi_{i,\varepsilon}|A_{i,\pm,\varepsilon}) \Rightarrow F_{i,\pm}(x,\alpha,\mu);$$

(c)

$$F_{i,\pm}(x,\alpha,\mu)\left\{\psi : P^{n^\pm(i),k} Df_{n^\pm(i)}(x_{i,\pm})\psi \neq 0\right\} = 1, \quad k = \begin{cases} 1,2, & \nu > 2,\\ 1, & \nu = 2.\end{cases}$$

(4) *For any* $r > 0$, *there is* $T(r)$ *such that, as* $\varepsilon \to 0$,

$$\mathsf{P}\left\{\sup_{0\le t\le T(r)} |X_\varepsilon(t) - S^t x| \ge r\right\} \to 0,$$
$$\mathsf{P}\left\{\sup_{T(r)\le t\le T_{\mathrm{out},\varepsilon}-T(r)} |X_\varepsilon(t) - z_i| \ge r\right\} \to 0,$$
$$\mathsf{P}\left(A_{i,\pm,\varepsilon} \cap \left\{\sup_{T_{\mathrm{out},\varepsilon}-T(r)\le t\le T_{\mathrm{out},\varepsilon}+h_i^\pm} |X_\varepsilon(t) - S^{t-T_{\mathrm{out},\varepsilon}} q_{i,\pm}| \ge r\right\}\right) \to 0.$$

(5) *For any* $r > 0$, *as* $\varepsilon \to 0$,

$$\mathsf{P}\left(A_{i,\pm,\varepsilon} \cap \left\{\sup_{0\le t\le(\ln\varepsilon^{-1})^{1/2}} |X_\varepsilon(T_{\mathrm{out},\varepsilon} + h_i^\pm + t) - S^t x_{i,\pm}| \ge r\right\}\right) \to 0.$$

Part (1) of the lemma describes the asymptotic behavior of exit times. Part (2) determines the limiting probabilities of choosing each of the two outgoing heteroclinic orbits. Part (3) describes the entrance distribution for the next visited saddle point (it takes $T_{\mathrm{out},\varepsilon} + h_i^\pm$ to reach its neighborhood); parts (a) and (b) give the asymptotic scaling law, and part (c) ensures that we can apply this lemma at the next saddle, too, which gives rise to the iteration scheme (see the definition of In_i). Part (4) formalizes the fact that for small ε, with high probability, the diffusion trajectory first closely follows the deterministic trajectory $S^t x$, then spends some time in a small neighborhood of the saddle point, and then follows closely one of the outgoing heteroclinic connections until it reaches the neighborhood of the next saddle point. Part (5) shows that after reaching the neighborhood of the next saddle point the trajectory continues to follow the same heteroclinic connection for a sublogarithmic time.

The proof of this Lemma will be given in Sects. 8 and 9. The solution X_ε spends most of the time in the neighborhood of the saddle points and in beetween it travels from one saddle point to another along a heteroclinic connection. We split the analysis in two parts accordingly. In Sect. 8 we describe the behavior of the system in the neighborhood of the saddle point assuming that the initial data is given by (7.2). In Sect. 9 we describe the motion between neighborhoods of two saddle points and finish the proof of Lemma 7.1.

The rest of this section is devoted to the derivation of our main result from Lemma 7.1.

Proof of Theorem 5.1. We have to show that the map Ψ constructed in Lemma 7.1 satisfies the statement of the theorem.

For any sequence $\mathbf{z} = (\theta_0, z_{i_1}, \theta_1, z_{i_2}, \theta_2 \ldots, \theta_{k-1}, z_{i_k}, \theta_k)$ and any ε we define a sequence of stopping times in the following way. First, we set

$$\tau_{\varepsilon,\mathbf{z},1,\mathrm{in}} = \tilde{t}(x_0),$$

where $\tilde{t}(x_0)$ was defined in (5.8). Then, for $j = 1, \ldots, k$, we recall that $U^{\pm}_{i_j}$ was defined in (7.1) and define recursively

$$\tau_{\varepsilon,\mathbf{z},j,\mathrm{out}} = \begin{cases} \inf\{t \ge \tau_{\varepsilon,\mathbf{z},j,\mathrm{in}} : X_\varepsilon(t) \in U^+_{i_j}\}, & \theta_j = \gamma^+_{i_j}, \\ \inf\{t \ge \tau_{\varepsilon,\mathbf{z},j,\mathrm{in}} : X_\varepsilon(t) \in U^-_{i_j}\}, & \theta_j = \gamma^-_{i_j}, \end{cases}$$

$$\tau_{\varepsilon,\mathbf{z},j+1,\mathrm{in}} = \begin{cases} \tau_{\varepsilon,\mathbf{z},j,\mathrm{out}} + h^+_{i_j}, & \theta_j = \gamma^+_{i_j}, \\ \tau_{\varepsilon,\mathbf{z},j,\mathrm{out}} + h^-_{i_j}, & \theta_j = \gamma^-_{i_j}. \end{cases}$$

Less formally, for each j, the process X_ε leaves the neighborhood of z_{i_j} at time $\tau_{\varepsilon,\mathbf{z},j,\mathrm{out}}$ and travels for time $t^{\pm}_{i_j}$ along one of the two heteroclinic connections $\theta_j = \gamma^{\pm}_{i_j}$ emerging from z_{i_j}. At time $\tau_{\varepsilon,\mathbf{z},j+1,\mathrm{in}}$ the solution is in the neighborhood of the next saddle point of the sequence, close to $x^{\pm}_{i_j}$.

The last point of the sequence $\mathbf{z}$ is special. We define now

$$T_{\varepsilon,\mathbf{z}} = \frac{\tau_{\varepsilon,\mathbf{z},k+1,\mathrm{in}} + (\ln \varepsilon^{-1})^{1/2}}{\ln(\varepsilon^{-1})}$$

and

$$T_\varepsilon = \varepsilon^{-1} \wedge \inf_{z \in L} T_{\varepsilon,\mathbf{z}}.$$

The ε^{-1} term is introduced to make T_ε finite (to take into account the improbable case where X_ε does not evolve along any $\mathbf{z} \in L$) so that the curve $\Gamma_{Z_\varepsilon(t), t \le T_\varepsilon}$ is well-defined.

According to Theorem 2.1 in [3], it suffices to show that for any open set $G \in \mathbb{X}$,

$$\liminf_{\varepsilon \to 0} \mathsf{P}\{\Gamma_{Z_\varepsilon(t), t \le T_\varepsilon} \in G\} \ge M_{x_0, L}(G),$$

where the probability measure $M_{x_0,L}$ is defined by (5.12). Since the measure $M_{x_0,L}$ is discrete, it is sufficient to check that for any $\mathbf{z} \in L$ and any open set $G \subset \mathbb{X}$ containing $\Gamma(\mathbf{z})$,

$$\lim_{\varepsilon \to 0} \mathsf{P}\{\Gamma_{Z_\varepsilon(t), t \le T_\varepsilon} \in G\} \ge \pi(\mathbf{z}). \tag{7.4}$$

We start with the linearization along the orbit $S^t x_0$. It follows from [4] (see also Lemma 9.2) that

$$X_\varepsilon(\tilde{t}(x_0)) = \tilde{x}_0 + \varepsilon \phi_\varepsilon,$$

with $\mathrm{Law}(\phi_\varepsilon) \Rightarrow \mu_0$, where μ_0 was defined in (5.7).

Now the strong Markov property allows to apply Lemma 7.1 iteratively along the saddle points of sequence $\mathbf{z}$. We can complete the proof by reparametrizing the heteroclinic connections of $\mathbf{z}$ appropriately and applying the following proximity criterion to derive (7.4).

Lemma 7.2 *Suppose that δ is a positive number and a path γ defines a piecewise constant curve Γ as given by (4.3). Suppose a continuous function f and nondecreasing nonnegative number sequences $(r_m)_{m=0}^{2k+1}$ and $(s'_m)_{m=0}^{2k+1}$ satisfy the following properties:*

(1) $t_j - \delta \le r_{2j} \le t_j \le r_{2j+1} \le t_j + \delta, \quad j = 0, \dots, k$;
(2) $|f(r) - y_j| \le \delta$ *for* $j = 1, \dots, k$ *and* $r \in [r_{2j-1}, r_{2j}]$;
(3) $s_{2j} \le s'_{2j} \le s'_{2j+1} \le s_{2j+1}, \quad j = 0, \dots, k$;
(4) *for each* $j = 0, \dots, k$, *there is a nondecreasing bijection*

$$\lambda_j : [r_{2j}, r_{2j+1}] \to [s'_{2j}, s'_{2j+1}]$$

such that

$$|\gamma(\lambda_j(r)) - (r, f^1(r), \dots, f^d(r))| < \delta, \quad r \in [r_{2j}, r_{2j+1}];$$

(5) *for each* $j = 0, \dots, k$,

$$|\gamma(s) - \gamma(s_{2j})| \le \delta, \qquad s \in [s_{2j}, s'_{2j}],$$
$$|\gamma(s) - \gamma(s_{2j+1})| \le \delta, \quad s \in [s'_{2j+1}, s_{2j+1}].$$

Then $\rho(\Gamma, \Gamma_f) \le 3\delta$.

The proof of Lemma 7.2 is given in Sect. 12 □

8 The system in the neighborhood of a saddle point

In this section we fix $i \in \mathcal{C}$ and consider a saddle point z_i. We recall our assumption that the dynamics generated by b in a small neighborhood U_i of z_i is conjugated by a C^2-diffeomorphism f_i to that generated by a linear vector field generated by a matrix A_i in a neighborhood of the origin. In this section we often denote A_i, f_i, etc., by A, f, etc., omitting the dependence on i. In particular, the flow generated by the linearized vector field is given by e^{tA} and sometimes will be denoted by S_A^t.

We recall that the eigenvalues $\lambda_1, \ldots, \lambda_d$ of A are real and simple, and there is a number $\nu \geq 2$ such that

$$\lambda_1 > \cdots > \lambda_{\nu-1} > 0 > \lambda_\nu > \cdots > \lambda_d.$$

We denote the associated eigenvectors by $v_1, \ldots, v_d$, and introduce the coordinates $u^1, \ldots, u^d$ of a vector u via $u = \sum_{k=1}^d u^k v_k$. We define

$$L = \operatorname{span}\{v_2, \ldots, v_d\}, \quad L^- = \operatorname{span}\{v_\nu, \ldots, v_d\},$$

and denote by Π_L the projection on L along v_1.

We begin the analysis with the derivation of the Itô equation for $Y_\varepsilon(t) = f(X_\varepsilon(t))$. The Itô formula gives:

$$\begin{aligned} dY_\varepsilon^j(t) &= \sum_{k=1}^d \partial_k f^j(X_\varepsilon(t))dX_\varepsilon^k + \frac{\varepsilon^2}{2}\sum_{k,l=1}^d (\sigma\sigma^*)^{kl}(X_\varepsilon(t))\partial_{kl} f^j(X_\varepsilon(t))dt \\ &= \sum_{k=1}^d \partial_k f^j(X_\varepsilon(t))b^k(X_\varepsilon(t))dt + \varepsilon\sum_{k=1}^d \partial_k f^j(X_\varepsilon(t))\sigma_m^k(X_\varepsilon(t))dW^m(t) \\ &\quad + \frac{\varepsilon^2}{2}\sum_{k,l=1}^d (\sigma\sigma^*)^{kl}(X_\varepsilon(t))\partial_{kl} f^j(X_\varepsilon(t))dt \end{aligned}$$

where σ^* denotes the transpose of σ, and $(\sigma\sigma^*)^{kl}(x)$ denotes $(\sigma(x)\sigma^*(x)))^{kl}$. Since the pushforward of the vector field b under f at a point y is given by Ay, we have

$$Ay = Df(g(y))b(g(y)),$$

where $g = f^{-1}$, and the equation above rewrites as

$$\begin{aligned} dY_\varepsilon^j(t) &= \sum_{k=1}^d A_k^j Y_\varepsilon^k(t)dt + \varepsilon\sum_{k=1}^d \partial_k f^j(g(Y_\varepsilon(t)))\sigma_m^k(g(Y_\varepsilon(t)))dW^m(t) \\ &\quad + \frac{\varepsilon^2}{2}\sum_{k,l=1}^d (\sigma\sigma^*)^{kl}(g(Y_\varepsilon(t)))\partial_{kl} f^j(g(Y_\varepsilon(t)))dt. \end{aligned}$$

We rewrite the last identity as

$$dY_\varepsilon(t) = AY_\varepsilon(t)dt + \varepsilon B(Y_\varepsilon(t))dW(t) + \varepsilon^2 C(Y_\varepsilon(t))dt, \tag{8.1}$$

or, equivalently,

$$\begin{aligned} dY_\varepsilon^j(t) &= A_k^j Y_\varepsilon^k(t)dt + \varepsilon B^j(Y_\varepsilon(t))dW(t) + \varepsilon^2 C^j(Y_\varepsilon(t))dt \\ &= A_k^j Y_\varepsilon^k(t)dt + \varepsilon \sum_{k=1}^{d} B_k^j(Y_\varepsilon(t))dW^k(t) + \varepsilon^2 C^j(Y_\varepsilon(t))dt. \end{aligned}$$

We see that B and C are continuous and bounded in the neighborhood $f(U)$ of the origin, and B is nodegenerate.

It is also clear that if we assume (7.2), then

$$Y_\varepsilon(0) = y_0 + \varepsilon^\alpha \xi_\varepsilon, \tag{8.2}$$

where $y_0 = f(x)$, and ξ_ε converges, as $\varepsilon \to 0$, in distribution to $\xi_0 = DF(x_0)\phi_0$, with $\mathrm{Law}(\phi_0) = \mu$, see (7.3).

Recall that for the saddle point $z = z_i$, two neighborhoods $\tilde{U} \subset U$ are defined. Define $V = f(U)$ and $\tilde{V} = f(\tilde{U})$, so that

$$\tilde{V} = \{y \in \mathbb{R}^d : |y^j| \le R, j = 1, \dots, d\} \subset V,$$

(we use the notation $R = R_i$, $\tilde{U} = \tilde{U}_i$, $U = U_i$ in this section). Then $y_0 \in \tilde{V} \cap L^-$. In particular, $y_0^k = 0$ for all $k < \nu$.

In the remainder of this section we study the system (8.1) with initial data given by (8.2). The solution is actually defined up to a stopping time $t_{V,\varepsilon}$ at which the solution hits ∂V. Let us define another stopping time

$$t_\varepsilon = \inf\{t \ge 0 : |Y_\varepsilon^1(t)| = R\} \wedge t_{V,\varepsilon}.$$

As we shall see later,

$$\mathsf{P}\{t_\varepsilon < t_{V,\varepsilon}\} \to 1, \quad \varepsilon \to 0,$$

and thus it makes sense to study the asymptotic behavior of t_ε and $Y(t_\varepsilon)$.

To state our main result on system (8.1), (8.2), see Lemma 8.1 below, we have to introduce a certain multidimensional distribution that is easier to describe in terms of random variables defined on some sufficiently rich probability space.

We start with the random vector ξ_0 introduced after (8.2). Then, on the same probability space we define a $(\nu-1)$-dimensional centered Gaussian vector $(N_0^1, \dots, N_0^{\nu-1})$ with covariance

$$\mathsf{E} N_0^k N_0^j = \int_0^\infty e^{-(\lambda_k+\lambda_j)s} (BB^*)^{kj} \left(S_A^s y_0\right) ds, \quad k, j < \nu, \tag{8.3}$$

independent of ξ_0. Next, we define $(\varkappa^1, \dots, \varkappa^{\nu-1})$ by

$$\varkappa^k = \xi_0^k + N_0^k \mathbf{1}_{\alpha=1}, \quad k < \nu. \tag{8.4}$$

Finally, on the same probability space we define a random vector $(\bar{N}_0^\nu, \dots, \bar{N}_0^d)$. Conditioned on each of the two events $\{\mathrm{sgn}(\varkappa^1) = \pm 1\}$, it is a centered Gaussian vector with

$$\mathsf{E}\bar{N}_0^k \bar{N}_0^j = \int_{-\infty}^{0} e^{-(\lambda_k+\lambda_j)s} (BB^*)^{kj} \left(S_A^s(\pm R v_1)\right) ds, \quad k, j \geq \nu,$$

and independent of $(|\varkappa^1|, \varkappa^2, \dots, \varkappa^{\nu-1})$.

Lemma 8.1 *Suppose the following nondegeneracy conditions are satisfied*:

$$\begin{aligned} &y_0^\nu \neq 0, \\ &\mathsf{P}\{\xi_0^1 \neq 0\} = 1, \quad \textit{if}\ \alpha < 1, \\ &\mathsf{P}\{\xi_0^2 \neq 0\} = 1, \quad \textit{if}\ \alpha < 1 \textit{ and } \nu > 2. \end{aligned}$$

Then there exists a number β, and a random vector (y', ξ', ζ) such that the random vector

$$\left(Y_\varepsilon^1(t_\varepsilon),\ \varepsilon^{-\beta} \Pi_L Y_\varepsilon(t_\varepsilon),\ t_\varepsilon - \frac{\alpha}{\lambda_1} \ln \frac{1}{\varepsilon} \right)$$

converges in distribution to (y', ξ', ζ). More precisely,

$$\beta = \begin{cases} 1, & \nu = 2 \ \text{ and } \ -\alpha\lambda_\nu \geq \lambda_1, \\ -\alpha\frac{\lambda_\nu}{\lambda_1}, & \nu = 2 \ \text{ and } \ -\alpha\lambda_\nu < \lambda_1, \\ \alpha\left(1 - \frac{\lambda_1}{\lambda_2}\right), & \nu > 2 \ \text{ and } \ -\lambda_\nu \geq \lambda_1 - \lambda_2, \\ -\alpha\frac{\lambda_\nu}{\lambda_1}, & \nu > 2 \ \text{ and } \ -\lambda_\nu < \lambda_1 - \lambda_2, \end{cases}$$

$$y' = \mathrm{sgn}(\varkappa^1) R,$$

and

$$\xi' = \begin{cases} N, & \nu = 2 \ \text{ and } \ -\alpha\lambda_\nu > \lambda_1, \\ N + \eta_-, & \nu = 2 \ \text{ and } \ -\alpha\lambda_\nu = \lambda_1, \\ \eta_-, & \nu = 2 \ \text{ and } \ -\alpha\lambda_\nu < \lambda_1, \\ \eta_+, & \nu > 2 \ \text{ and } \ -\lambda_\nu > \lambda_1 - \lambda_2, \\ \eta_+ + \eta_-, & \nu > 2 \ \text{ and } \ -\lambda_\nu = \lambda_1 - \lambda_2, \\ \eta_-, & \nu > 2 \ \text{ and } \ -\lambda_\nu < \lambda_1 - \lambda_2, \end{cases} \tag{8.5}$$

with

$$\eta_- = R^{\frac{\lambda_\nu}{\lambda_1}} |\varkappa^1|^{-\frac{\lambda_\nu}{\lambda_1}} y_0^\nu v_\nu,$$
$$\eta_+ = R^{\frac{\lambda_2}{\lambda_1}} |\varkappa^1|^{-\frac{\lambda_2}{\lambda_1}} \varkappa^2 v_2.$$

and

$$N = \sum_{k=\nu}^{d} \bar{N}_0^k v_k.$$

Remark 8.1 Even if the nondegeneracy assumptions do not hold, a version of this lemma still holds true. This will be obvious from the proof, and, for brevity, we omit a variety of related results on these degenerate situations.

Remark 8.2 We see that of all random variables N, η_-, η_+, involved in the description of the limit, conditioned on $\mathrm{sgn}(\varkappa^1)$, only N does not depend in any way on the initial distribution data given by y_0, α, and ξ_0. This guarantees the Markovian loss of memory for the case $[\nu = 2; -\alpha\lambda_\nu > \lambda_1]$, and potentially leads to non-Markov situations in all the other cases, see Sect. 10 for further discussion.

The proof consists of two parts. The first part provides the analysis of the evolution of Y_ε mostly along the stable manifold. The second part is mostly responsible for the motion along the unstable manifold of the origin.

Using Itô's formula it is easy to verify that Duhamel's principle holds:

$$Y_\varepsilon(t) = e^{At} Y_\varepsilon(0) + \varepsilon e^{At} \int_0^t e^{-As} B(Y_\varepsilon(s)) dW(s) + \varepsilon^2 e^{At} \int_0^t e^{-As} C(Y_\varepsilon(s)) ds,$$

or, equivalently,

$$Y_\varepsilon^k(t) = e^{\lambda_k t} \left(y_0^k + \varepsilon^\alpha \xi_\varepsilon^k + \varepsilon \int_0^t e^{-\lambda_k s} B^k(Y_\varepsilon(s)) dW(s) + \varepsilon^2 \int_0^t e^{-\lambda_k s} C^k(Y_\varepsilon(s)) ds \right). \tag{8.6}$$

We start with a study of the outcome of the evolution of Y_ε along the stable manifold. We fix a number $\bar{\alpha} \in (0, \alpha)$. Our first goal is to analyze the distribution of $Y_\varepsilon(\tau_\varepsilon)$, where

$$\tau_\varepsilon = \min\{\tau_\varepsilon^k,\ k = 1, \ldots, d\},$$

and for every $k = 1, \ldots, d$,

$$\tau_\varepsilon^k = \inf\left\{t : \left|Y_\varepsilon^k(t) - \left(S_A^t y_0\right)^k\right| = \varepsilon^{\bar{\alpha}}\right\}.$$

Lemma 8.2

$$\sup_{t\ge 0}\left|Y_\varepsilon(\tau_\varepsilon\wedge t)-\left(S_A^{t\wedge\tau_\varepsilon}y_0\right)\right|\xrightarrow{\mathsf{P}}0,\quad \varepsilon\to\infty.$$

Proof This statement is obvious due to

$$\sup_{t\ge 0}\left|Y_\varepsilon(\tau_\varepsilon\wedge t)-\left(S_A^{t\wedge\tau_\varepsilon}y_0\right)\right|\le d\varepsilon^{\alpha},\tag{8.7}$$

which follows from the definition of τ_ε. □

Lemma 8.3

$$\tau_\varepsilon\xrightarrow{\mathsf{P}}\infty,\quad \varepsilon\to\infty.$$

A proof based on (8.6) is given in Sect. 11.

Lemma 8.4

$$\mathsf{P}\left\{\tau_\varepsilon=\tau_\varepsilon^1\right\}\to 1,\quad \varepsilon\to\infty.$$

A sketch of a proof is also given in Sect. 11.

Let us now take a closer look at the stochastic integral term in the expression (8.6). For $k<\nu$ we introduce

$$N_\varepsilon^k(t)=\int_0^t e^{-\lambda_k s}B^k(Y_\varepsilon(s))dW(s)=\sum_{m=1}^d\int_0^t e^{-\lambda_k s}B_m^k(Y_\varepsilon(s))dW^m(s)\tag{8.8}$$

and

$$M^k(t)=\int_0^t e^{-\lambda_k s}B^k(S_A^s y_0)dW(s)=\sum_{m=1}^d\int_0^t e^{-\lambda_k s}B_m^k(S_A^s y_0)dW^m(s).$$

A straightforward application based on BDG inequalities (see e.g. [6, Theorem 3.28, Chapter 3]), local Lipschitzness of B, and (8.7) implies that, as $\varepsilon\to 0$,

$$\sup_{t\le\tau_\varepsilon}\left|N_\varepsilon^k(t)-M^k(t)\right|=\sup_{t\le\tau_\varepsilon}\left|\int_0^t e^{-\lambda_k s}(B^k(Y_\varepsilon(s))-B^k(S_A^s y_0))dW(s)\right|\xrightarrow{\mathsf{P}}0.$$

Lemma 8.3 implies that the terminal value $N_\varepsilon(\tau_\varepsilon)$ converges to $M(\infty)$. Computing $\mathsf{E}M^k(\infty)M^j(\infty)$, we see that $\mathrm{Law}(M(\infty))=\mathrm{Law}(N_0)$, where N_0 is a centered Gaussian vector defined in (8.3). Therefore,

$$N_\varepsilon(\tau_\varepsilon)\xrightarrow{\mathrm{Law}}N_0,\quad \varepsilon\to 0.$$

We also notice that

$$\sup_{t\le\tau_\varepsilon}\left|\varepsilon^2\int\limits_0^t e^{-\lambda_k s}C^k(Y_\varepsilon(s))ds\right| = o_{\mathsf{P}}(\varepsilon),$$

where we use the notation $\phi(\varepsilon) = o_{\mathsf{P}}(\psi(\varepsilon))$ for any families of random variables $\phi(\varepsilon), \psi(\varepsilon), \varepsilon > 0$ such that $\frac{\phi(\varepsilon)}{\psi(\varepsilon)} \stackrel{\mathsf{P}}{\to} 0$ as $\varepsilon \to 0$.

Therefore, for $k < \nu$ we have

$$Y_\varepsilon^k(\tau_\varepsilon) = e^{\lambda_k\tau_\varepsilon}(\varepsilon^\alpha\xi_\varepsilon^k + \varepsilon N_\varepsilon^k(\tau_\varepsilon) + o_{\mathsf{P}}(\varepsilon)). \tag{8.9}$$

We denote

$$\varkappa_\varepsilon^k = Y_\varepsilon^k(\tau_\varepsilon)e^{-\lambda_k\tau_\varepsilon}\varepsilon^{-\alpha},$$

so that

$$\varkappa_\varepsilon^k = \xi_\varepsilon^k + \varepsilon^{1-\alpha}N_\varepsilon^k(\tau_\varepsilon) + o_{\mathsf{P}}(\varepsilon^{1-\alpha}),$$

and rewrite (8.9) as

$$Y_\varepsilon^k(\tau_\varepsilon) = e^{\lambda_k\tau_\varepsilon}\varepsilon^\alpha\varkappa_\varepsilon^k. \tag{8.10}$$

Lemma 8.4 implies that

$$\mathsf{P}\left\{|Y_\varepsilon^1(\tau_\varepsilon)| = \varepsilon^{\bar\alpha}\right\} \to 1. \tag{8.11}$$

So, taking logarithms of

$$e^{\lambda_1\tau_\varepsilon}\varepsilon^\alpha\left|\varkappa_\varepsilon^1\right| = \varepsilon^{\bar\alpha}$$

we see that with probability approaching 1, as $\varepsilon \to 0$,

$$\tau_\varepsilon - \frac{\bar\alpha-\alpha}{\lambda_1}\ln\varepsilon = -\frac{\ln\left|\varkappa_\varepsilon^1\right|}{\lambda_1} \tag{8.12}$$

and

$$Y_\varepsilon^1(\tau_\varepsilon)\varepsilon^{-\bar\alpha} = \operatorname{sgn}\left(\varkappa_\varepsilon^1\right), \tag{8.13}$$

Plugging (8.12) into (8.10), we obtain the following result:

Lemma 8.5 *For* $1 < k < \nu$,

$$Y_\varepsilon^k(\tau_\varepsilon) = \varepsilon^{\frac{\lambda_k}{\lambda_1}(\bar\alpha-\alpha)+\alpha} \varkappa_\varepsilon^k \left|\varkappa_\varepsilon^1\right|^{-\frac{\lambda_k}{\lambda_1}}.$$

This lemma describes the asymptotics of $Y_\varepsilon^k(\tau_\varepsilon), k < \nu$ very precisely since due to (8.12) we have the following obvious statement:

Lemma 8.6

$$\left(\xi_\varepsilon, \tau_\varepsilon - \frac{\bar\alpha - \alpha}{\lambda_1}\ln\varepsilon, \varkappa_\varepsilon^1, \ldots, \varkappa_\varepsilon^{\nu-1}\right) \overset{\text{Law}}{\to} \left(\xi_0, -\frac{\ln|\varkappa^1|}{\lambda_1}, \varkappa^1, \ldots, \varkappa^{\nu-1}\right),$$

where the random variables in the r.h.s. were defined before the statement of Lemma 8.1.

We shall now consider $k \ge \nu$. For these values of k, we denote

$$R_\varepsilon^k(t) = e^{\lambda_k t}\int_0^t e^{-\lambda_k s} B^k(Y_\varepsilon(s))dW(s).$$

Let us take any numbers $T \in \mathbb{N}$ and $p > 0$, and use BDG inequalities to write

$$\begin{aligned} \mathsf{P}\left\{\sup_{s\in[0,T]}|R_\varepsilon^k(t\wedge\tau_\varepsilon)| > \varepsilon^{-p}\right\} &\le \sum_{n=0}^{T-1}\mathsf{P}\left\{\sup_{t\in[n,n+1]}\left|R_\varepsilon^k(t\wedge\tau_\varepsilon)\right| > \varepsilon^{-p}\right\} \\ &\le \sum_{n=0}^{T-1}\mathsf{P}\left\{\sup_{t\le n+1}\left|\int_0^{t\wedge\tau_\varepsilon} e^{-\lambda_k s}B^k(Y_\varepsilon(s))dW(s)\right| \ge e^{-\lambda_k n}\varepsilon^{-p}\right\} \le KT\varepsilon^{-2p}, \end{aligned} \quad (8.14)$$

for some K.

Since the term

$$e^{\lambda_k t}\int_0^t e^{-\lambda_k s} C^k(Y_\varepsilon(s))ds$$

is bounded on $t \le \tau_\varepsilon$, we see that (8.6), (8.12) and (8.14) imply the following result:

Lemma 8.7 *For all* $k \ge \nu$ *and any* $p > 0$,

$$Y_\varepsilon^k(\tau_\varepsilon) = \varepsilon^{\frac{\lambda_k}{\lambda_1}(\bar\alpha-\alpha)}\left|\varkappa_\varepsilon^1\right|^{-\frac{\lambda_k}{\lambda_1}}\left(y_0^k + \varepsilon^\alpha\xi_\varepsilon^k\right) + o_{\mathsf{P}}(\varepsilon^{1-p}).$$

Lemmas 8.5, 8.6, and 8.7 provide all the necessary information on the behavior of Y_ε up to τ_ε. We now turn to the second part of the proof, the analysis of the evolution of Y_ε after τ_ε.

Let $\bar{Y}_\varepsilon(t) = Y_\varepsilon(\tau_\varepsilon + t)$. We shall study $\bar{Y}_\varepsilon$ conditioned on $\bar{Y}^1(0) = Y^1_\varepsilon(\tau_\varepsilon) = \pm\varepsilon^{\bar{\alpha}}$. We consider only the case of $+\varepsilon^{\bar{\alpha}}$, since the analysis of the other case is entirely the same. First, we rewrite (8.6) for $\bar{Y}_\varepsilon$:

$$\bar{Y}^k_\varepsilon(t) = e^{\lambda_k t}\left(Y^k_\varepsilon(\tau_\varepsilon) + \varepsilon\int_0^t e^{-\lambda_k s}B^k(\bar{Y}_\varepsilon(s))dW(s) + \varepsilon^2\int_0^t e^{-\lambda_k s}C^k(\bar{Y}_\varepsilon(s))ds\right). \tag{8.15}$$

Let $\bar{\tau}_\varepsilon = \inf\{t : |\bar{Y}^1_\varepsilon| = R\}$. We can rewrite (8.15) for $k = 1$, $t = \bar{\tau}_\varepsilon$ as

$$\bar{Y}^1_\varepsilon(\bar{\tau}_\varepsilon) = e^{\lambda_1\bar{\tau}_\varepsilon}\varepsilon^{\bar{\alpha}}(1 + \eta_\varepsilon), \tag{8.16}$$

with

$$\eta_\varepsilon = \varepsilon^{1-\bar{\alpha}}\int_0^{\bar{\tau}_\varepsilon} e^{-\lambda_1 s}B^1(\bar{Y}_\varepsilon(s))dW(s) + \varepsilon^{2-\alpha}\int_0^{\bar{\tau}_\varepsilon} e^{-\lambda_1 s}C^1(\bar{Y}_\varepsilon(s))ds \xrightarrow{\mathsf{P}} 0.$$

The last relation is obvious if B and C are bounded. In the general case it follows from a localization argument.

Relation (8.16) implies

$$\bar{\tau}_\varepsilon = -\frac{\bar{\alpha}}{\lambda_1}\ln\varepsilon + \frac{1}{\lambda_1}\ln\frac{R}{1+\eta_\varepsilon}. \tag{8.17}$$

Plugging this into (8.15) and applying Lemmas 8.5 and 8.7 we can prove the following statement:

Lemma 8.8

$$\sup_{t\le\bar{\tau}_\varepsilon}|\bar{Y}_\varepsilon(t) - S^t_A(\varepsilon^{\bar{\alpha}}v_1)| \xrightarrow{\mathsf{P}} 0. \tag{8.18}$$

However, we need a more detailed information on $\bar{Y}_\varepsilon(\bar{\tau}_\varepsilon)$. To that end, we analyze $\bar{Y}^k_\varepsilon(\bar{\tau}_\varepsilon)$ separately for $2 \le k < \nu$ and $\nu \le k \le d$.

For $2 \le k < \nu$, plugging (8.17) into (8.15), using Lemma 8.5 for the first term and Lemma 8.8 to estimate the integral terms, we see that

$$\begin{aligned}\bar{Y}^k_\varepsilon(\bar{\tau}_\varepsilon) &= \varepsilon^{-\frac{\lambda_k}{\lambda_1}\bar{\alpha}}R^{\frac{\lambda_k}{\lambda_1}}(1+\eta_\varepsilon)^{-\frac{\lambda_k}{\lambda_1}}\left(\varepsilon^{\frac{\lambda_k}{\lambda_1}(\bar{\alpha}-\alpha)+\alpha}\varkappa^k_\varepsilon\left|\varkappa^1_\varepsilon\right|^{-\frac{\lambda_k}{\lambda_1}} + o_{\mathsf{P}}\left(\varepsilon^{\frac{\lambda_k}{\lambda_1}(\bar{\alpha}-\alpha)+\alpha}\right)\right)\\ &= \varepsilon^{\alpha\left(1-\frac{\lambda_k}{\lambda_1}\right)}R^{\frac{\lambda_k}{\lambda_1}}\varkappa^k_\varepsilon|\varkappa^1_\varepsilon|^{-\frac{\lambda_k}{\lambda_1}}(1+o_{\mathsf{P}}(1)), \quad 2 \le k < \nu. \end{aligned}\tag{8.19}$$

For $k \ge \nu$ we denote

$$\bar{N}_\varepsilon^k = e^{\lambda_k \bar{\tau}_\varepsilon} \int_0^{\bar{\tau}_\varepsilon} e^{-\lambda_k s} B^k(Y_\varepsilon(s)) dW(s)$$

Lemma 8.9 *As $\varepsilon \to 0$,*

$$(\bar{N}_\varepsilon^\nu, \ldots, \bar{N}_\varepsilon^d) \stackrel{\text{Law}}{\to} (\bar{N}_0^\nu, \ldots, \bar{N}_0^d),$$

where $(\bar{N}_0^\nu, \ldots, \bar{N}_0^d)$ is the centered Gaussian vector defined before the statement of Lemma 8.1.

Proof Denoting

$$\tau'_\varepsilon = -\frac{\bar{\alpha}}{\lambda_1} \ln \varepsilon + \frac{1}{\lambda_1} \ln R \to \infty,$$

we obtain

$$\tau'_\varepsilon \to \infty, \quad \varepsilon \to 0$$

and

$$\bar{\tau}_\varepsilon - \tau'_\varepsilon \stackrel{\mathsf{P}}{\to} 0, \quad \varepsilon \to 0.$$

Therefore,

$$\bar{N}_\varepsilon^k - e^{\lambda_k \tau'_\varepsilon} \int_0^{\tau'_\varepsilon} e^{-\lambda_k s} B^k(Y_\varepsilon(s)) dW(s) \stackrel{\mathsf{P}}{\to} 0. \tag{8.20}$$

It follows from Lemma 8.8 that

$$e^{\lambda_k \tau'_\varepsilon} \int_0^{\tau'_\varepsilon} e^{-\lambda_k s} B^k(Y_\varepsilon(s)) dW(s) - \hat{N}_\varepsilon^k \stackrel{\mathsf{P}}{\to} 0, \tag{8.21}$$

where

$$\hat{N}_\varepsilon^k = e^{\lambda_k \tau'_\varepsilon} \int_0^{\tau'_\varepsilon} e^{-\lambda_k s} B^k(S_A^s \varepsilon^{\bar{\alpha}} v_1) dW(s), \quad k \ge \nu.$$

We see that $(\hat{N}_\varepsilon^\nu, \ldots, \hat{N}_\varepsilon^d)$ is a centered Gaussian vector with

$$\begin{aligned}\mathsf{E}\hat{N}_\varepsilon^j \hat{N}_\varepsilon^k &= \int_0^{\tau'_\varepsilon} e^{(\lambda_k+\lambda_j)(\tau'_\varepsilon - s)} (BB^*)^{kj} (S_A^s \varepsilon^{\bar{\alpha}} v_1) ds \\ &= \int_{-\tau'_\varepsilon}^0 e^{-(\lambda_k+\lambda_j)r} (BB^*)^{kj} (S_A^{\tau'_\varepsilon + r} \varepsilon^{\bar{\alpha}} v_1) dr \\ &\to \int_{-\infty}^0 e^{-(\lambda_k+\lambda_j)r} (BB^*)^{kj} \left(S_A^r R v_1\right) dr, \quad \varepsilon \to 0, \end{aligned} \tag{8.22}$$

where the second indentity follows from the change of variables $s - \tau'_\varepsilon = r$, and the convergence in the last line is implied by the uniform convergence $S_A^{\tau'_\varepsilon + r} \varepsilon^{\bar{\alpha}} v_1 \to S_A^r R v_1$, $r \le 0$.

Lemma 8.9 follows now from (8.20) to (8.22). □

Now, for $k \ge \nu$, Eqs. (8.15), (8.17), and Lemma 8.7 imply:

$$\begin{aligned}\bar{Y}_\varepsilon^k(\tau_\varepsilon) &= e^{\lambda_k \left(-\frac{\bar{\alpha}}{\lambda_1} \ln \varepsilon + \frac{1}{\lambda_1} \ln \frac{R}{1+\eta_\varepsilon}\right)} \left(\varepsilon^{\frac{\lambda_k}{\lambda_1}(\bar{\alpha}-\alpha)} \left|\varkappa_\varepsilon^1\right|^{-\frac{\lambda_k}{\lambda_1}} \left(y_0^k + \varepsilon^\alpha \xi_\varepsilon^k\right) + o_{\mathsf{P}}(\varepsilon^{1-p}) \right) \\ &\quad + \varepsilon \bar{N}_\varepsilon^k + o_{\mathsf{P}}(\varepsilon) \\ &= \varepsilon^{-\frac{\lambda_k}{\lambda_1}\alpha} R^{\frac{\lambda_k}{\lambda_1}} |\varkappa_\varepsilon^1|^{-\frac{\lambda_k}{\lambda_1}} y_0^k (1 + o_{\mathsf{P}}(1)) + o_{\mathsf{P}}(\varepsilon^{1-p-\frac{\lambda_k}{\lambda_1}\bar{\alpha}}) + \varepsilon \bar{N}_\varepsilon^k + o_{\mathsf{P}}(\varepsilon).\end{aligned} \tag{8.23}$$

We are now ready to finish the proof of Lemma 8.1. We notice that formulas analogous to (8.19) and (8.23) hold true if we condition on $\bar{Y}_\varepsilon^1(0) = Y_\varepsilon^1(\tau_\varepsilon) = -\varepsilon^\alpha$. Since $\bar{Y}_\varepsilon(\bar{\tau}_\varepsilon) = Y_\varepsilon(t_\varepsilon)$ and $t_\varepsilon = \tau_\varepsilon + \bar{\tau}_\varepsilon$, Lemma 8.1 is a consequence of the strong Markov property and an elementary analysis of relations (8.12), (8.17), (8.19), (8.23). The proof reduces to extracting the leading order terms in (8.19), (8.23) in each of the cases that appear in the statement of the lemma.

For example, in the case of $\nu > 2$ and $-\lambda_\nu < \lambda_1 - \lambda_2$, the greatest contribution (as $\varepsilon \to 0$) in (8.19) is of order of $\varepsilon^{\alpha(1-\frac{\lambda_2}{\lambda_1})}$ and corresponds to $k = 2$. The greatest contribution in (8.23) is of order of $\varepsilon^{-\frac{\lambda_\nu}{\lambda_1}\alpha}$ and corresponds to $k = \nu$ since one can choose p sufficiently small and neglect the asymptotic contribution of the $o_{\mathsf{P}}(\varepsilon^{1-p-\frac{\lambda_k}{\lambda_1}\bar{\alpha}})$ term. (Notice also that in this case the $o_{\mathsf{P}}(\varepsilon)$ term can also be neglected since $\varepsilon = o(\varepsilon^{-\frac{\lambda_\nu}{\lambda_1}\alpha})$ due to $\alpha \le 1$ and $-\frac{\lambda_\nu}{\lambda_1} < \frac{\lambda_1 - \lambda_2}{\lambda_1} < 1$.) Among these two contributions, $\varepsilon^{-\frac{\lambda_\nu}{\lambda_1}\alpha}$ dominates providing the desired asymptotics

$$\Pi_L Y_\varepsilon(t_\varepsilon) \sim \varepsilon^{-\frac{\lambda_\nu}{\lambda_1}\alpha} R^{\frac{\lambda_\nu}{\lambda_1}} \left|\varkappa_\varepsilon^1\right|^{-\frac{\lambda_k \nu}{\lambda_1}} y_0^\nu v_\nu.$$

The analysis of all the other cases is similar, and we omit it.

Having Lemma 8.1 at hand, it is straightforward to write down the asymptotics for the original process $X_\varepsilon(t_\varepsilon) = g(Y_\varepsilon(t_\varepsilon))$.

Lemma 8.10 *Let X_ε solve the system (1.2) with initial condition (7.2) and assume $(x, \alpha, \mu) \in \mathrm{In}_i$. Let us define $\varkappa^1$, $\varkappa^2$ via (8.2), (8.3), (8.4), and β, ξ', ζ as in Lemma 8.1.*

Then

$$X_\varepsilon(t_\varepsilon) = q_\varepsilon + \varepsilon^\beta \phi'_\varepsilon, \tag{8.24}$$

where

$$\mathsf{P}\{q_\varepsilon = g(\pm R)\} = 1, \quad \varepsilon > 0,$$

and

$$\left(q_\varepsilon, \phi'_\varepsilon,\ t_\varepsilon - \frac{\alpha}{\lambda_1} \ln \frac{1}{\varepsilon}\right) \stackrel{\text{Law}}{\to} (q, \phi', \zeta), \quad \varepsilon \to 0.$$

Here

$$q = g(\operatorname{sgn}(\varkappa^1) R),$$

and

$$\phi' = (Df(q))^{-1} \xi'.$$

Remark 8.3 The case of $\alpha = 1$ and $\phi_\varepsilon = 0$ is also covered by this lemma. This situation corresponds to the deterministic initial condition for all $\varepsilon > 0$.

9 Asymptotics along the heteroclinic orbit

In this section we consider the Eq. (1.2), equipped with the initial condition

$$X_\varepsilon(0) = x_0 + \varepsilon^\alpha \phi_\varepsilon,$$

on a finite time horizon (up to a nonrandom time T). Here $\alpha \in (0, 1]$, and $(\phi_\varepsilon)_{\varepsilon > 0}$ is a family of random vectors satisfying

$$\phi_\varepsilon \stackrel{\text{Law}}{\to} \phi_0, \quad \varepsilon \to 0,$$

for some nondegenerate random vector ϕ_0, independent of the Wiener process driving the equation.

The following statement is elementary.

Lemma 9.1 *As* $\varepsilon \to 0$*,*

$$\sup_{t\in[0,T]} |X_\varepsilon(t) - S^t x_0| \xrightarrow{\mathsf{P}} 0.$$

We are going to give more precise asymptotics in the spirit of [4]. To that end we denote

$$Y(t) = X(t) - S^t x_0,$$

and write

$$\begin{aligned} Y_\varepsilon(t) &= \varepsilon^\alpha \phi_\varepsilon + \int_0^t (b(X_\varepsilon(s) - b(S^t x_0)))ds + \varepsilon \int_0^t \sigma(X_\varepsilon(s))dW(s) \\ &= \varepsilon^\alpha \phi_\varepsilon + \int_0^t Db(S^t x_0)Y_\varepsilon(s)ds + \int_0^t Q(S^t x_0, Y_\varepsilon(s))ds + \varepsilon \int_0^t \sigma(X_\varepsilon(s))dW(s), \end{aligned}$$

where

$$|Q(x, y)| \le Cy^2, \tag{9.1}$$

for some constant C and all y with $|y| < 1$. Treating this as a linear nonhomogeneous equation we apply the Duhamel principle to see that

$$\begin{aligned} Y_\varepsilon(t) = \varepsilon^\alpha \Phi_{x_0}(t)\phi_\varepsilon + \varepsilon \Phi_{x_0}(t) \int_0^t \Phi_{x_0}^{-1}(s)\sigma(X_\varepsilon(s))dW(s) \\ + \Phi_{x_0}(t) \int_0^t \Phi_{x_0}^{-1}(s) Q(S^t x_0, Y_\varepsilon(s))ds. \end{aligned} \tag{9.2}$$

We introduce a stopping time τ_ε by

$$\tau_\varepsilon = \inf\{t : |Y_\varepsilon(t)| \ge \varepsilon^{\frac{2}{3}(1\wedge\alpha)}\}.$$

On the event $\{\tau_\varepsilon \le T\}$ we use (9.2) to write

$$Y_\varepsilon(t \wedge \tau_\varepsilon) = I_1(t, \varepsilon) + I_2(t, \varepsilon) + I_3(t, \varepsilon), \tag{9.3}$$

with

$$|I_1(t, \varepsilon)| = |\varepsilon^\alpha \Phi_{x_0}(t \wedge \tau_\varepsilon)\phi_\varepsilon| \le K_1 \varepsilon^\alpha |\phi_\varepsilon|$$

for some constant K_1,

$$|I_2(t,\varepsilon)| = \varepsilon \left| \Phi_{x_0}(t\wedge\tau_\varepsilon) \int_0^{t\wedge\tau_\varepsilon} \Phi_{x_0}^{-1}(s)\sigma(X_\varepsilon(s))dW(s) \right| = \varepsilon |N_\varepsilon(t)|,$$

where N_ε is a tight (in C with sup-norm) family of continuous processes due to Lemma 9.1, and, finally,

$$|I_3(t,\varepsilon)| \le K_3 \varepsilon^{\frac{4}{3}(1\wedge\alpha)}$$

for some constant K_3 due to (9.1). It now follows from (9.3) for $t = T$, that on $\{\tau_\varepsilon \le T\}$

$$\varepsilon^{\frac{2}{3}(1\wedge\alpha)} \le K_1 \varepsilon^\alpha |\phi_\varepsilon| + \varepsilon |N_\varepsilon(\tau_\varepsilon)| + K_3 \varepsilon^{\frac{4}{3}(1\wedge\alpha)},$$

which automatically implies

$$\mathsf{P}\{\tau_\varepsilon \le T\} \to 0, \quad \varepsilon \to 0.$$

On the complement, $\{\tau_\varepsilon > T\}$, we have

$$Y_\varepsilon(T) = I_1'(\varepsilon) + I_2'(\varepsilon) + I_3'(\varepsilon), \tag{9.4}$$

where

$$I_1'(\varepsilon) = \varepsilon^\alpha \Phi_{x_0}(T)\phi_\varepsilon, \tag{9.5}$$

$$I_2'(\varepsilon) = \varepsilon \Phi_{x_0}(T) \int_0^T \Phi_{x_0}^{-1}(s)\sigma(X_\varepsilon(s))dW(s), \tag{9.6}$$

$$|I_3'(\varepsilon)| \le K_4 \varepsilon^{\frac{4}{3}(1\wedge\alpha)} \tag{9.7}$$

Now (9.4)–(9.7) imply the following result:

Lemma 9.2 *For every* $\varepsilon > 0$

$$X_\varepsilon(T) = S^T x_0 + \varepsilon^\alpha \bar{\phi}_\varepsilon,$$

where

$$\bar{\phi}_\varepsilon \overset{\text{Law}}{\to} \bar{\phi}_0, \quad \varepsilon \to 0,$$

with

$$\bar{\phi}_0 = \Phi_{x_0}(T)\phi_0 + \mathbf{1}_{\alpha=1} N,$$

N being a Gaussian vector:

$$N = \Phi_{x_0}(T) \int_0^T \Phi_{x_0}^{-1}(s)\sigma(S^s x_0)dW(s).$$

Remark 9.1 The lemma also holds true for the case where $\alpha = 1$ and $\phi_\varepsilon = 0$. This situation corresponds to the deterministic initial condition for all $\varepsilon > 0$.

We can now finish the proof of Lemma 7.1.

Proof of Lemma 7.1 Parts 1 and 2 follow directly from Lemma 8.10. Part 3 follows from consecutive application of Lemmas 8.10 and 9.2 together with Strong Markov property and the nondegeneracy assumption (5.5). The last two statements of the Lemma are obvious in view of the analysis above, and we omit their proofs. □

Remark 9.2 Notice that not only we are able to prove that the entrance-exit map with desired properties exists, but we also can describe this map explicitly, tracing the details from Lemmas 8.1 and 9.2. For example,

$$p_{i,\pm}(x, \alpha, \mu) = \mathsf{P}\{\mathrm{sgn}(\varkappa^1) = \pm 1\},$$

where $\varkappa^1$ is the random variable constructed in Lemma 8.1 applied to the linearization about z_i.

10 Discussion

In this section we informally comment on implications of our main result and its possible extensions.

10.1 The structure of the limiting process

According to our main result, the limiting probability for the solution of the system to evolve along a sequence of heteroclinic connections is computed by a recursive procedure described in (5.6)–(5.10).

The key result for the analysis of the Markov property for the limiting process is Lemma 8.1, which gives the "exit distribution" $\mathrm{Law}(\xi')$ via the "entrance distribution" $\mathrm{Law}(\xi_0)$.

Analyzing (8.5), we see that in the case where for each $i \in \mathcal{C}$ we have $\nu_i = 2$, and $-\lambda_{i,2} > \lambda_{i,1}$ (i.e. contraction is stronger than expansion) our iteration scheme gives that at each saddle point the limiting process chooses each of the outgoing heteroclinic connections with equal probabilities, thus defining a simple random walk on the network viewed as a directed graph.

However, in general, the limiting distribution on sequences of saddles does not necessarily define a Markov chain. In fact, if the exit distribution defined in (8.5) involves η_-, then it is asymmetric. Therefore, at the next saddle point, the choice between the two heteroclinic orbits is asymmetric as well, and probabilities are generically not equal to $1/2$. In fact they may equal 0 or 1.

Let us now look at η_+, the other random variable involved in the exit distribution. If the entrance distribution is symmetric, then η_+ also has symmetric distribution. On the other hand, if the entrance distribution is asymmetric, then η_+ is also (typically) asymmetric.

Notice that there are plenty of situations where the entrance distribution is strongly asymmetric (concentrated on a semiline) and the probabilities to choose one of the two outgoing connections are 0 and 1, i.e. the choice is asymptotically deterministic (and dependent on the history of the process).

So, there are many possibilities, but roughly the limiting random walk on the saddles may be described as follows. The system starts evolving in a Markov fashion (choosing the next saddle out of two possible ones with probabilities $1/2$ independently of the history of the process) until it meets a saddle point at which the exit distribution becomes asymmetric. After that, the choice of one of the two heteroclinic connections is not Markov any more (being defined by the entrance distribution whose asymmetry is in turn defined by the history of the process.) Then, at each saddle point the following three things may happen: (i) new asymmetry is brought in due to the presence of η_- in the exit asymptotics; (ii) the asymmetry present in the entrance distribution is transferred to the exit distribution by η_+; (iii) $\nu = 2$ and the contraction is strong enough to ensure that the exit distribution does not involve η_- or η_+ thus being symmetric. In the first two cases the system remembers its past encoded in the asymmetry of $\eta_\pm$. However, in the last case, the system loses all the memory and goes back to the "Markov mode" (which is just a convenient name for this phase of the system's evolution; of course the system is not truly Markov since some information from the past is encoded in the fact that the system is in the "Markov mode" presently.)

In particular, if every entrance distribution involved is either symmetric or strongly asymmetric (this excludes the "rare" cases of $[\nu = 2; -\alpha\lambda_\nu = \lambda_1]$ and $[\nu > 2; -\lambda_\nu = \lambda_1 - \lambda_2]$), then the two splitting probabilities are either both equal to $1/2$, or equal 0 and 1, respectively.

It is instructional to find the limiting exit measure for the 2-dimensional example shown on Fig. 1, where the domain contains three saddle points z_1, z_2, and z_3. Let us assume that the linearizations at these points have the same eigenvalues $\lambda_1 > 0 > \lambda_2$. According to Theorem 6.1, the exit measure will weakly converge, as $\varepsilon \to 0$, to

$$p_1\delta_{y_1} + p_2\delta_{y_2} + p_3\delta_{y_3} + p_4\delta_{y_4},$$

where $y_1, \dots, y_4$ are the points on the boundary that are reachable along the network.

If $-\lambda_2 > \lambda_1$ then we have Markov evolution along the network resulting in $p_1 = 1/2$, $p_2 = 1/4$, $p_3 = p_4 = 1/8$.

If $-\lambda_2 < \lambda_1$, then the exit distribution at each saddle point is strongly asymmetric which results in deterministic motion after the first bifurcation so that $p_1 = 1/2$, $p_2 = 0$, $p_3 = 0$, $p_4 = 1/2$.

If $-\lambda_2 = \lambda_1$, then the exit distribution $N + \eta_-$ at each saddle point is asymmetric, so that the evolution is not Markov, and $p_1 = 1/2$, $0 < p_2 < 1/4$, $0 < p_3 < p_4$.

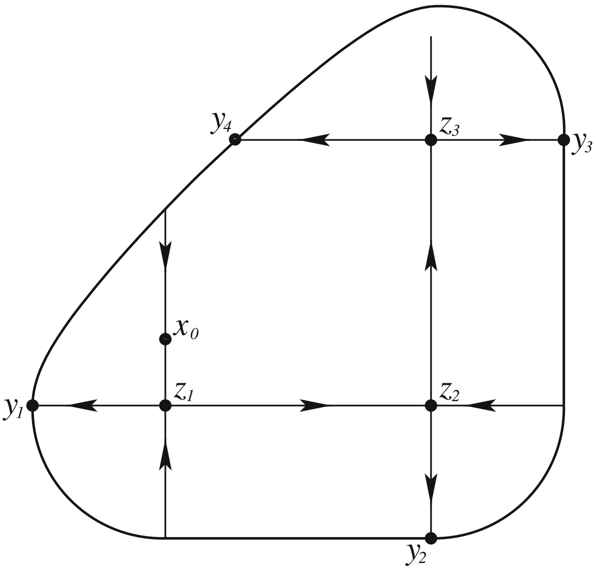

Fig. 1 A two-dimensional example with three saddle points

10.2 Dependence of the exit measure on the diffusion coefficient

Suppose that the vector field b is given. From our basic iteration procedure we see that if for some matrix-valued function σ at each saddle point the entrance distribution is either symmetric or strongly asymmetric, then for any other choice of diffusion coefficients, the situation is the same, so that the limiting process given by Theorem 5.1 and the exit measure asymptotics given by Theorem 6.1 do not depend of σ. Therefore, these theorems describe an intrinsic property of the vector field b.

10.3 Our set of assumptions and possible generalizations

We begin with an obvious and important extension of our theorem that we are going to mention without any proof. We assumed so far that every critical point in the network is a saddle with nontrivial stable and unstable manifolds. Let us see what happens if we allow situations where some of the critical points of the network are actually attracting nodes (this corresponds to $\nu_i = 1$, where ν_i is the order number of the first negative eigenvalue of the linearization at z_i). It follows from the FW theory that if the initial condition belongs to a neighborhood of such a point, then with probability approaching 1 as $\varepsilon \to 0$, the trajectory will not leave that neighborhood within the time of order $\ln(\varepsilon^{-1})$. Therefore, our main result holds true for this situation as well, with the following correction: with probability 1, for every $i \in \mathcal{C}$ with $\nu_i = 1$, if the limiting process reaches z_i, then it stays there indefinitely.

Another essential assumption that we have used is that near each saddle point the dynamics is C^2-smoothly conjugated to a linear flow. Although this assumption leaves aside a number of interesting systems (one example is 2-dimensional Hamiltonian dynamics), it is not too much restrictive. Namely, a smooth conjugation to linear dynamics is possible for saddle points satisfying so called no resonance conditions, see a discussion of local conjugation to a linear flow in [7, Section 6.6]). In most examples of heteroclinic networks in the survey [9] all the saddles satisfy the no resonance conditions for almost all parameters involved in the description of the system.

Although our proof relies entirely on the analysis of the conjugated linear systems, we believe that our results can be extended to the situation with local dynamics that has no smooth linearization. The most promising way to approach the understanding of that situation is to use normal forms (see [7, Section 6.6]).

Our assumption that all the eigenvalues of matrices A_i are real and simple can be weakened. We could have actually assumed that $\lambda_1, \lambda_2, \lambda_\nu \in \mathbb{R}$ are simple, and

$$\lambda_1 > \lambda_2 > \Re\lambda_3 > \cdots > 0 > \lambda_\nu > \cdots > \Re\lambda_{\nu+1} > \cdots .$$

Obviously, the case where the eigenvalue with greatest real part has a nonzero imaginary part leads to a completely different picture since the invariant manifold associated to λ_1 is 2-dimensional in that case. The case where λ_2 or λ_ν are not real has a lot in common with the case considered in this paper, since for small values of ε the system tends to evolve along heteroclinic connections. However, instead of a unique limit distribution for Z_ε, there may be a set of limit distributions, i.e. the attractor in the space of distributions may be nontrivial. The reason is that due to the rotation associated with the imaginary part, the exit distribution at a saddle point is not well-defined. An analogue of Lemma 8.1 holds true, but the normalization by $\varepsilon^{-\beta}$ should be replaced by a more sophisticated one involving a rotation by an angle proportional to $\ln(\varepsilon^{-1})$. Even more complicated situations arise if, instead of connecting hyperbolic fixed points, the heteroclinic orbits connect limit cycles or other invariant sets.

The last assumption we would like to discuss is the assumption that the set L is conservative. Notice that Theorem 5.1 does not necessarily describe the asymptotic behavior of the diffusion up to infinite time. Indeed, times spent at subsequently visited saddles can become smaller and smaller so that the total time accumulates to a finite value. There is no unique deterministic time horizon that serves all limiting trajectories at once, since these values may differ from sequence to sequence. If one chooses L to be conservative, then one can find a valid trajectory-dependent (therefore, random) time horizon serving all sequences from L.

11 Auxiliary lemmas

Proof of Lemma 8.3. We prove that $\tau_\varepsilon^k \stackrel{\mathsf{P}}{\to} \infty$ for every k. Due to (8.6), for any $T > 0$,

$$\mathsf{P}\{\tau_\varepsilon^k \le T\} \le \mathsf{P}(A_1(\varepsilon, T)) + \mathsf{P}(A_2(\varepsilon, T)) + \mathsf{P}(A_3(\varepsilon, T)), \tag{11.1}$$

where

$$A_1(\varepsilon, T) = \left\{ \varepsilon^{\alpha}(1 \vee e^{\lambda_k T})|\xi_\varepsilon^k| > \frac{\varepsilon^{\bar{\alpha}}}{4} \right\},$$

$$A_2(\varepsilon, T) = \left\{ \tau_\varepsilon \le T;\ \exists t < \tau_\varepsilon,\ : \varepsilon e^{\lambda_k t} \left| \int_0^t e^{-\lambda_k s} B^k(Y_\varepsilon(s)) dW(s) \right| > \frac{\varepsilon^{\bar{\alpha}}}{4} \right\},$$

$$A_3(\varepsilon, T) = \left\{ \tau_\varepsilon \le T;\ \exists t < \tau_\varepsilon : \varepsilon^2 e^{\lambda_k t} \left| \int_0^t e^{-\lambda_k s} C^k(Y_\varepsilon(s)) ds \right| > \frac{\varepsilon^{\bar{\alpha}}}{4} \right\}.$$

Now

$$\mathsf{P}(A_1(\varepsilon, T)) \le \mathsf{P}\left\{ |\xi_\varepsilon^k| > \frac{\varepsilon^{\bar{\alpha}-\alpha}}{4(1 \vee e^{\lambda_k T})} \right\} \to 0, \quad \varepsilon \to 0,$$

since $\varepsilon^{\bar{\alpha}-\alpha} \to \infty$.

To estimate $A_2(\varepsilon, T)$, we can write

$$\mathsf{P}(A_2(\varepsilon, T)) \le \mathsf{P}\left\{ \sup_{t \le \tau_\varepsilon \wedge T} |N_\varepsilon^k(t)| > \frac{\varepsilon^{\bar{\alpha}-1}}{4(1 \vee e^{\lambda_k T})} \right\} = O(\varepsilon^{2(1-\bar{\alpha})}) \to 0,$$

which follows from consecutive application of the Chebyshev inequality, BDG inequalities, and an elementary estimate on the quadratic variation of N_ε^k which was defined in (8.8).

Next, we notice that for sufficiently small ε,

$$\mathsf{P}(A_3(\varepsilon, T)) \le \mathsf{P}\left\{ (1 \vee e^{\lambda_k T}) T C^* > \frac{\varepsilon^{\bar{\alpha}-2}}{4} \right\} = 0,$$

where $C^* = \sup |C(y)|$, which completes the proof. □

Sketch of a proof of Lemma 8.4. The proof is similar to the proof of Lemma 8.3 and uses the fact that among the factors $e^{\lambda_k t}$, the one with $k = 1$ grows fastest of all in t, and t is large due to Lemma 8.3, so that the exit level $\varepsilon^{\bar{\alpha}}$ is first reached by $|Y_\varepsilon^1(t) - (S^t y_0)^1|$ with high probability. □

12 Proof of properties of the space of curves $\mathbb{X}$

Any nondecreasing continuous map of [0, 1] onto itself is called a time change. For a time change λ we define its inverse by

$$\lambda^{-1}(s) = \inf\{t : \lambda(s) \ge s\}.$$

The inverse function is nondecreasing and satisfies

$$\lambda \circ \lambda^{-1}(s) = s, \quad s \in [0, 1].$$

It is continuous on the left and has limits on the right, but it may have jumps associated with segments of constancy, and therefore it is not necessarily a time change. Nevertheless, the following lemma holds true.

Lemma 12.1 *Let λ_1 and λ_2 be two time changes. Suppose that each segment of constancy of λ_1 is contained in a segment of constancy of λ_2. Then $\lambda_2 \circ \lambda_1^{-1}$ is a time change.*

Proof The function $\lambda_2 \circ \lambda_1^{-1}$ is obviously nondecreasing with $\lambda_2 \circ \lambda_1^{-1}(0) = 0$. Let us prove that it is continuous. Take any point $s \in [0, 1]$ and let

$$[t_-, t_+] = \{t : \lambda_1(t) = s\}.$$

Since $\lambda_1^{-1}(s+) = t_+$ and $\lambda_1^{-1}(s-) = t_-$, we have

$$\lambda_2 \circ \lambda_1^{-1}(s+) = \lambda_2(t_+)$$

and

$$\lambda_2 \circ \lambda_1^{-1}(s-) = \lambda_2(t_-).$$

By the assumption of the lemma, $\lambda_2(t_+) = \lambda_2(t_-)$. Therefore,

$$\lambda_2 \circ \lambda_1^{-1}(s+) = \lambda_2 \circ \lambda_1^{-1}(s-),$$

and the continuity is proven.

If $s = 1$, then, using the same notation and reasoning we see that $t_+ = 1$, and

$$\lambda_2 \circ \lambda_1^{-1}(1-) = \lambda_2(t_-) = \lambda_2(1) = 1,$$

and the proof is complete. □

Lemma 12.2 *For any nonconstant path γ there is a path γ' with no intervals of constancy, and a time change λ so that $\gamma = \gamma' \circ \lambda$.*

Proof Let $I_1, I_2, \ldots$ be maximal nondegenerate segments of constancy of γ. By assumption, none of these segments coincides with $[0, 1]$. Therefore, there is a time change λ that has these and only these maximal nondegenerate segments of constancy (λ may be constructed analogously to the Cantor staircase). It is easy to see that $\gamma' = \gamma \circ \lambda^{-1}$ is continuous.

Let us assume that there is a nodegenerate segment $[s_-, s_+]$ and a point x, such that $\gamma'(s) = x$ for all $s \in [s_-, s_+]$. This means that $\gamma(t) = x$ for all $t \in \lambda^{-1}([s_-, s_+])$. Now notice that $\lambda^{-1}([s_-, s_+])$ consists of the left ends of all maximal (not necessarily nondegenerate) segments of constancy of λ (and thus of γ) contained in $[t_-, t_+] = [\lambda^{-1}(s_-), \lambda^{-1}(s_+)]$. So, for the left end t of any interval of constancy contained in $[t_-, t_+]$, we have $\gamma(t) = x$. Therefore, in fact, $\gamma(t) = x$ for all $t \in [t_-, t_+]$. We conclude that $[t_-, t_+]$ is contained in some segment of constancy of λ, so that

$$s_- = \lambda(t_-) = \lambda(t_+) = s_+,$$

which contradicts the nondegeneracy of $[s_-, s_+]$. This finishes the proof that $\gamma' = \gamma \circ \lambda^{-1}$ is a curve with no intervals of constancy. To see that $\gamma = \gamma' \circ \lambda$, we notice that for each s,

$$\lambda^{-1} \circ \lambda(s) = \inf\{s_0 : \lambda(s_0) = \lambda(s)\}$$

is the left-end of the segment of constancy of λ (and γ), containing s. Therefore,

$$\gamma' \circ \lambda(s) = \gamma \circ \lambda^{-1} \circ \lambda(s) = \gamma(s), \quad s \in [0, 1],$$

and the proof is complete. □

Proof of Lemma 4.1. Of all equivalence properties, only the transitivity is not quite obvious. So, we assume that $\gamma_1 \sim \gamma_2$ and $\gamma_2 \sim \gamma_3$ and prove that $\gamma_1 \sim \gamma_3$. In other words, our assumption is that there are paths γ_{12}, γ_{23} and time changes $\lambda_{12,1}$, $\lambda_{12,2}$ $\lambda_{23,2}$, $\lambda_{23,3}$ such that

$$\gamma_1 = \gamma_{12} \circ \lambda_{12,1}, \quad \gamma_2 = \gamma_{12} \circ \lambda_{12,2},$$
$$\gamma_2 = \gamma_{23} \circ \lambda_{23,2}, \quad \gamma_3 = \gamma_{23} \circ \lambda_{23,3}.$$

According to Lemma 12.2, there is a path γ' with no intervals of constancy, and a time change λ such that

$$\gamma_2 = \gamma' \circ \lambda = \gamma_{12} \circ \lambda_{12,2} = \gamma_{23} \circ \lambda_{23,2}.$$

This representation implies that segments of constancy of time changes $\lambda_{12,2}$ and $\lambda_{23,2}$ are contained in segments of constancy of λ, so that Lemma 12.1 implies that $\lambda \circ \lambda_{12,2}^{-1}$ and $\lambda \circ \lambda_{23,2}^{-1}$ are time changes. We also see that

$$\gamma_{12} = \gamma' \circ \lambda \circ \lambda_{12,2}^{-1},$$
$$\gamma_{23} = \gamma' \circ \lambda \circ \lambda_{23,2}^{-1},$$

so that

$$\gamma_1 = \gamma' \circ \lambda \circ \lambda_{12,2}^{-1} \circ \lambda_{12,1},$$
$$\gamma_3 = \gamma' \circ \lambda \circ \lambda_{23,2}^{-1} \circ \lambda_{23,2},$$

which completes the proof, since the paths γ_1, γ_3 are represented via a common underlying path γ' and time changes $\lambda \circ \lambda_{12,2}^{-1} \circ \lambda_{12,1}$ and $\lambda \circ \lambda_{23,2}^{-1} \circ \lambda_{23,2}$, respectively. □

For a curve $\Gamma \in \mathbb{X}$, we shall denote by Γ' the set of all $\gamma \in \Gamma$ with no segments of constancy. Lemma 12.2 shows that $\Gamma' \neq \emptyset$ for any curve represented by a nonconstant path.

For a continuous function $f : [0, 1] \to \mathbb{R}^{d+1}$, we use

$$|f|_\infty = \sup_{s \in [0,1]} |f(s)|.$$

Lemma 12.3 (1) *For every $\gamma_1 \in \Gamma$ and $\gamma_2 \in \Gamma'$, there is a time change λ such that $\gamma_1 = \gamma_2 \circ \lambda$. If $\gamma_1 \in \Gamma'$, then λ is a bijection.*

(2) *For every $\gamma \in \Gamma$ and every $\varepsilon > 0$, there is a $\gamma_\varepsilon \in \Gamma'$ such that*

$$|\gamma - \gamma_\varepsilon|_\infty < \varepsilon.$$

(3) *Every curve is a closed set in the sup-norm $|\cdot|_\infty$.*

(4) *The definition (4.1) is equivalent to*

$$d(\Gamma_1, \Gamma_2) = \inf_{\gamma_1 \in \Gamma_1', \gamma_2 \in \Gamma_2'} |\gamma_1 - \gamma_2|_\infty. \tag{12.1}$$

Proof Part 1. If $\gamma_1 \in \Gamma$, $\gamma_2 \in \Gamma'$, then there are time changes λ_1, λ_2 and a path γ' with $\gamma_1 = \gamma' \circ \lambda_1$ and $\gamma_2 = \gamma' \circ \lambda_2$. The map λ_2 is strictly increasing and $\gamma' \in \Gamma'$, since if any of these two conditions is violated then γ_2 has a segment of constancy. In particular, λ_2^{-1} is also a time change, and we can write $\gamma' = \gamma_2 \circ \lambda_2^{-1}$. So, $\gamma_1 = \gamma_2 \circ \lambda_2^{-1} \circ \lambda_1$, and we can set $\lambda = \lambda_2^{-1} \circ \lambda_1$. If λ is not a bijection, then γ_1 has an interval of constancy, and the proof of part 1 is complete.

To prove part 2, we use part 1 to find $\gamma' \in \Gamma'$ and a time change λ so that $\gamma = \gamma' \circ \lambda$. For any $\delta > 0$, $\lambda_\delta(s) = \delta s + (1 - \delta)\lambda(s)$ defines a time change. Notice that $|\lambda_\delta - \lambda|_\infty \leq 2\delta$ and $\gamma' \circ \lambda_\delta \in \Gamma'$. Due to the uniform continuity of γ,

$$|\gamma - \gamma' \circ \lambda_\delta|_\infty = |\gamma' \circ \lambda - \gamma' \circ \lambda_\delta|_\infty < \varepsilon,$$

for sufficiently small δ, and we are done.

Part 3. We have to prove that if $\gamma_n \in \Gamma$ for all $n \in \mathbb{N}$, and $\lim_{n\to\infty} |\gamma_n - \mu|_\infty = 0$, then $\mu \in \Gamma$. Due to part 2, it is sufficient to consider the case where $\gamma_n \in \Gamma'$ for all n. In this situation there exists a sequence of time changes λ_n and $\gamma \in \Gamma'$ such that $\gamma_n = \gamma \circ \lambda_n$ for all n. Due to Helley's Selection Theorem, see [3, Appendix II], there is a sequence $n' \to \infty$ and a nondecreasing function $\lambda_\infty : [0, 1] \to [0, 1]$ such that $\lim_{n'\to\infty} \lambda_{n'}(s) = \lambda_\infty(s)$ for every point of continuity s of λ_∞. This implies

$$\mu(s) = \lim_{n'\to\infty} \gamma_{n'}(s) = \lim_{n'\to\infty} \gamma \circ \lambda_{n'}(s) = \gamma \circ \lambda_\infty(s) \tag{12.2}$$

for all points of continuity s.

Our goal is to show that λ_∞ is actually continuous at every point in $[0, 1]$. That will imply that $\lambda_{n'}$ converges to λ_∞ uniformly, since all these functions are nondecreasing, and we shall be able to conclude that $\lambda_\infty(0) = 0$, $\lambda_\infty(1) = 1$, so that λ_∞ is a time change. Moreover, it follows that (12.2) holds for all $s \in [0, 1]$, so that $\mu = \gamma \circ \lambda_\infty$ and $\mu \in \Gamma$.

So, we assume that there is a point $s_0 \in (0, 1)$ with $\lambda_\infty(s_0-) < \lambda_\infty(s_0+)$ and we are going to show that this assumption contradicts the uniform convergence of $\gamma_{n'}$ to μ.

Take any point $r \in (\lambda_\infty(s_0-), \lambda_\infty(s_0+))$ with $\gamma(r) \neq \gamma \circ \lambda_\infty(s_0+)$ and $\gamma(r) \neq \gamma \circ \lambda_\infty(s_0-)$ (this can be done since γ has no segments of constancy), and define

$$\varepsilon = \frac{1}{2} \min\{|\gamma(r) - \gamma \circ \lambda_\infty(s_0+)|, |\gamma(r) - \gamma \circ \lambda_\infty(s_0-)|\}. \tag{12.3}$$

The continuity of γ and the existence of the right and left limits of λ_∞ at s_0 allow us to choose $\delta > 0$ so that

$$s \in (s_0, s_0 + \delta) \quad \text{implies} \quad |\gamma \circ \lambda_\infty(s) - \gamma \circ \lambda_\infty(s_0+)| < \varepsilon, \tag{12.4}$$

$$s \in (s_0 - \delta, s_0) \quad \text{implies} \quad |\gamma \circ \lambda_\infty(s) - \gamma \circ \lambda_\infty(s_0-)| < \varepsilon. \tag{12.5}$$

Since $\lambda_{n'}$ converges to λ_∞ at all its continuity points, and these are dense, we see that there is a number n_0 such that for every $n' > n_0$, there is a continuity point $s(n') \in (s_0 - \delta, s_0 + \delta)$ with $\lambda_{n'}(s(n')) = r$. For these n', if $s = s(n') \in (s_0, s_0 + \delta)$, (12.3) and (12.4) imply

$$\begin{aligned} |\gamma_{n'}(s) - \gamma \circ \lambda_\infty(s)| &= |\gamma(r) - \gamma \circ \lambda_\infty(s)| \\ &\geq |\gamma(r) - \gamma \circ \lambda_\infty(s_0+)| - |\gamma \circ \lambda_\infty(s) - \gamma \circ \lambda_\infty(s_0+)| \\ &\geq 2\varepsilon - \varepsilon = \varepsilon, \end{aligned}$$

and if $s = s(n') \in (s_0 - \delta, s_0)$, (12.3) and (12.5) imply

$$\begin{aligned} |\gamma_{n'}(s) - \gamma \circ \lambda_\infty(s)| &= |\gamma(r) - \gamma \circ \lambda_\infty(s)| \\ &\geq |\gamma(r) - \gamma \circ \lambda_\infty(s_0-)| - |\gamma \circ \lambda_\infty(s) - \gamma \circ \lambda_\infty(s_0-)| \\ &\geq 2\varepsilon - \varepsilon = \varepsilon, \end{aligned}$$

So, for all $n' > n_0$ there is $s \in (s_0 - \delta, s_0 + \delta)$ such that

$$|\gamma_{n'}(s) - \mu(s)| = |\gamma_{n'}(s) - \gamma \circ \lambda_\infty(s)| \geq \varepsilon,$$

and we obtain a contradiction with the uniform convergence of $\gamma_{n'}$ to μ. This completes the proof that there is no discontinuities of λ_∞ within $(0, 1)$. The demonstration showing that λ_∞ is continuous at the endpoints of $[0, 1]$ is similar. This completes the proof of part 3.

Part 4 follows immediately from part 2. The lemma is proven completely. □

Proof of Theorem 4.1. Part 1. We notice first that d is nonnegative and symmetric. To show the triangle inequality, for any three curves $\Gamma_1, \Gamma_2, \Gamma_3$ and $\varepsilon > 0$ we use part 4 of Lemma 12.3 to find $\gamma_{1,2} \in \Gamma_1'$, $\gamma_{2,1}, \gamma_{2,3} \in \Gamma_2'$, and $\gamma_{3,2} \in \Gamma_3'$ with

$$|\gamma_{1,2} - \gamma_{2,1}|_\infty < d(\Gamma_1, \Gamma_2) + \varepsilon$$

and

$$|\gamma_{3,2} - \gamma_{2,3}|_\infty < d(\Gamma_3, \Gamma_2) + \varepsilon.$$

Let us find an time change λ such that $\gamma_{2,3} = \gamma_{2,1} \circ \lambda$. Then

$$\begin{aligned} d(\Gamma_1, \Gamma_3) &\le |\gamma_{1,2} \circ \lambda - \gamma_{3,2}|_\infty \\ &\le |\gamma_{1,2} \circ \lambda - \gamma_{2,1} \circ \lambda|_\infty + |\gamma_{2,1} \circ \lambda - \gamma_{3,2}|_\infty \\ &\le |\gamma_{1,2} - \gamma_{2,1}|_\infty + |\gamma_{2,3} - \gamma_{3,2}|_\infty \\ &\le d(\Gamma_1, \Gamma_2) + d(\Gamma_3, \Gamma_2) + 2\varepsilon, \end{aligned}$$

and the desired inequality follows since ε is arbitrarily small.

If $d(\Gamma_1, \Gamma_2) = 0$, then there are two sequences of paths $\gamma_{1,n} \in \Gamma_1'$, $\gamma_{2,n} \in \Gamma_2'$ with $\lim_{n\to\infty} |\gamma_{1,n} - \gamma_{2,n}|_\infty = 0$. Using appropriate time changes, we see that there is $\tilde{\gamma}_1 \in \Gamma_1'$ and $\tilde{\gamma}_{2,n} \in \Gamma_2'$, $n \in \mathbb{N}$ such that $\lim_{n\to\infty} |\tilde{\gamma}_1 - \tilde{\gamma}_{2,n}| = 0$. Due to part 3 of Lemma 12.3, we see that $\tilde{\gamma}_1 \in \Gamma_2$. Therefore, $\Gamma_1 = \Gamma_2$, and we are done with part 1.

Part 2. The space $\mathbb{X}$ is separable since it inherits a dense countable set from a closed subset (of functions with nondecreasing zeroth coordinate) of the separable space $C([0, 1] \to [0, \infty) \times \mathbb{R}^d)$.

Let us now prove that $\mathbb{X}$ is complete. Suppose that $(\Gamma_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in $(\mathbb{X}, \rho)$. Using Lemma 12.3 and reparametrization, we can find an increasing number sequence $(n_k)_{k\in\mathbb{N}}$, with $n_k \in \mathbb{N}$, and a sequence of paths $(\gamma_k)_{k\in\mathbb{N}}$ with $\gamma_k \in \Gamma_{n_k}'$ such that $|\gamma_k - \gamma_{k+1}|_\infty < 2^{-k}$. This sequence of paths is Cauchy in C and, therefore, converges to a path γ_∞ which defines a path Γ_∞. Obviously, Γ_{n_k} converges to Γ_∞ as $k \to \infty$. One can now use the triangle inequality to extend this convergence to the whole sequence (Γ_n). So every Cauchy sequence is convergent, which completes the proof of Theorem 4.1. □

Proof of Lemma 4.2. We prove the necessity part of the statement since the sufficiency part is obvious. Due to Lemma 12.3, we may assume that there is a sequence of paths $\gamma_n \in \Gamma_{f_n}'$ such that

$$|\gamma_n - \gamma_g|_\infty \to 0, \quad n \to \infty, \tag{12.6}$$

where γ_g is defined in (4.2). Therefore,

$$\sup_{s\in[0,1]} \left|\gamma_n^0(s) - sT\right| \to 0, \quad n \to \infty.$$

This in turn implies that if we define a strictly increasing and continuous function $u_n(s)$ by $\gamma_n^0(u(s)) = sT$, then

$$\sup_{s\in[0,1]} |u_n(s) - sT| \to 0, \quad n \to \infty.$$

This, together with (12.6) and uniform continuity of g proves that

$$\sup_{s\in[0,1]} |f_n(sT) - g(sT)| = \sup_{s\in[0,1]} |\gamma_n(u_n(s)) - \gamma_g(s)| \to 0, \quad n \to \infty,$$

and the proof is complete. □

Proof of Lemma 7.2. It is straightforward to see that the path $\tilde{\gamma}_f$ defined by

$$\tilde{\gamma}_f(s) = \begin{cases} (\gamma^0(s), f^1(\gamma^0(s)), \ldots, f^d(\gamma^0(s))), & \gamma^0(s) \in [r_{2j-1}, r_{2j}], \quad j=1,\ldots,k \\ (r_{2j}, f^1(r_{2j}), \ldots, f^d(r_{2j})), & \gamma^0(s) \geq r_{2j}, s \leq s'_{2j}, \quad j=0, v \ldots, k \\ (\lambda_j^{-1}(s), f^1(\lambda_j^{-1}(s)), \ldots, f^d(\lambda_j^{-1}(s))), & s \in [s'_{2j}, s'_{2j+1}], \quad j=0,\ldots,k \\ (r_{2j+1}, f^1(r_{2j+1}), \ldots, f^d(r_{2j+1})), & s \geq s'_{2j+1}, \gamma^0(s) \leq r_{2j+1}, \quad j=0,\ldots,k \end{cases}$$

belongs to Γ_f and $|\gamma - \tilde{\gamma}_f|_\infty < 3\delta$. □

Acknowledgments I started to think about heteroclinic networks in the Fall of 2006, after I attended a talk on their applications to neural dynamics by Valentin Afraimovich (see his paper [10]). I am grateful to him for several stimulating discussions this work began with, as well as for his words of encouragement without which it could not have been finished. I am also grateful to NSF for their support via a CAREER grant.

References

1. Armbruster, D., Stone, E., Kirk, V.: Noisy heteroclinic networks. Chaos **13**(1), 71–86 (2003)
2. Bakhtin, Y.: Exit asymptotics for small diffusion about an unstable equilibrium. Stoch. Process. Appl. **118**(5), 839–851 (2008)
3. Billingsley, P.: Convergence of probability measures. In: Wiley Series in Probability and Statistics: Probability and Statistics, 2nd edn. Wiley, New York (1999). A Wiley-Interscience Publication
4. Blagoveščenskiĭ, J.N.: Diffusion processes depending on a small parameter. Teor. Verojatnost. i Primenen. **7**, 135–152 (1962)
5. Freidlin, M.I., Wentzell, A.D.: Random perturbations of dynamical systems. In: Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 260, 2nd edn. Springer, New York (1998). Translated from the 1979 Russian original by Joseph Szücs
6. Karatzas, I., Shreve, S.E.: Brownian motion and stochastic calculus. In: Graduate Texts in Mathematics, vol. 113. Springer, New York (1988)
7. Katok, A., Hasselblatt, B.: Introduction to the modern theory of dynamical systems. In: Encyclopedia of Mathematics and its Applications, vol. 54. Cambridge University Press, Cambridge (1995). With a supplementary chapter by Katok and Leonardo Mendoza

8. Kifer, Y.: The exit problem for small random perturbations of dynamical systems with a hyperbolic fixed point. Israel J. Math. **40**(1), 74–96 (1981)
9. Krupa, M.: Robust heteroclinic cycles. J. Nonlinear Sci. **7**(2), 129–176 (1997)
10. Rabinovich, M.I., Huerta, R., Afraimovich, V.: Dynamics of sequential decision making. Phys. Rev. Lett. **97**(18), 188103 (2006)
11. Skorohod, A.V.: Limit theorems for stochastic processes. Teor. Veroyatnost. i Primenen. **1**, 289–319 (1956)