Normal forms approach to diffusion near hyperbolic equilibria

Sergio Angel Almada Monter and Yuri Bakhtin
School of Mathematics, Georgia Tech, Atlanta GA 30332-0160, USA
E-mail: salmada3@math.gatech.edu and bakhtin@math.gatech.edu

Received 15 June 2010, in final form 2 February 2011
Published 19 May 2011
Online at stacks.iop.org/Non/24/1883

Abstract
We consider the exit problem for small white noise perturbation of a smooth dynamical system on the plane in the neighbourhood of a hyperbolic critical point. We show that if the distribution of the initial condition has a scaling limit then the exit distribution and exit time also have a joint scaling limit as the noise intensity goes to zero. The limiting law is computed explicitly. The result completes the theory of noisy heteroclinic networks in two dimensions. The analysis is based on normal forms theory.

Mathematics Subject Classification: 60H10, 34C20, 34E10

1. Introduction
Small stochastic perturbations of continuous deterministic dynamical systems have been studied intensively for several decades. One of the greatest achievements in the area is the celebrated Freidlin–Wentzell (FW) theory that allows us to explain long-term behaviour of systems with several metastable states at the level of large deviation estimates [6].

An interesting situation where one can prove more precise estimates than those provided via FW quasi-potential approach was considered by Kifer [9]. He studied the exit problem for small noise perturbations of a deterministic system in a neighbourhood of a hyperbolic fixed point (or saddle) in $\mathbb{R}^d$ assuming that the starting point for the diffusion belongs to the stable manifold of the fixed point. Kifer showed that as the noise level $\epsilon$ decays to 0, the diffusion tends to exit along the invariant manifold associated with the leading eigenvalue $\lambda_+^+$ of the linearization of the system even in the presence of other unstable directions. He also found that the random exit time $\tau_\epsilon$ is asymptotic in probability to $\lambda_+^+ \ln \epsilon^{-1}$.

When studying noisy perturbations of systems with heteroclinic networks, i.e. multiple saddle points connected by heteroclinic orbits, Bakhtin [2, 3], realized that to understand the vanishing noise behaviour of the system, one has to extend Kifer’s work and analyse (i) the
limiting distribution of the approximation error $\tau_\epsilon - \lambda_\epsilon^{-1} \ln \epsilon^{-1}$; (ii) the limiting scaling laws of the exit distribution for the neighbourhood of each saddle. In fact, the exit distribution for the first saddle point serves as the entrance distribution for the next saddle point, so that the peculiarities of the exit distribution can significantly influence the further evolution of the system.

The detailed analysis of scaling limits for distributional Poincaré maps near saddle points carried out in [3] resulted in a complete theory for noisy heteroclinic networks. This theory explains interesting non-Markovian limit effects and the emerging patterns in the winnerless competition in the process of sequential decision making (here, we are using the terminology from [12] where applications of heteroclinic networks to neural dynamics are considered). The main result is that under the logarithmic time scaling the diffusion process converges in distribution in a special topology to a precisely described limiting process that jumps between the saddles along the heteroclinic connections.

The core result that was applied in [3] iteratively for sequences of saddle points connected to one another is a lemma that computes the asymptotic scaling of the exit distribution for a neighbourhood of a saddle point given the scaling of the entrance distribution. The proof of that lemma was based on a coordinate change conjugating the driving drift vector field to a linear vector field. Although this method and the lemma based on it apply in a fairly generic situation where the so called no-resonance condition holds, there are interesting cases such as Hamiltonian dynamics where the smooth linearization is not possible due to resonances. In these cases, the system remains nonlinear even under the optimal smooth change of coordinates, but it has a certain special structure that can be studied using the classical theory of normal forms (see, e.g., [4, 7, 11]).

In this paper, we extend the key lemma of [3] to cover the resonant cases and, in fact, to the complete generality in the case $d = 2$. Our approach is based on normal forms that have particularly nice structure in the two-dimensional case. We believe that the main result of this paper can be extended to higher dimensions.

An important consequence of our result is that in two dimensions the no-resonance restriction is completely removed from the theory of noisy heteroclinic networks developed in [3], so that the theory applies to any heteroclinic networks generated by smooth vector fields on the plane. It also provides a generalization of [1, 9] in two dimensions.

The structure of the paper is the following. In section 2 we introduce the setting. In section 3 we state the main theorem and split the proof into several parts. In section 4 we introduce a simplifying change of coordinates in a small neighbourhood of the saddle point. The analysis of the transformed process in section 5 is based upon two results. Their proofs are given in sections 6 and 7. An example that exemplifies all the phenomena established in this paper is presented in section 8. In section 9 some possible research directions are presented.

2. Setting

Let us consider a $C^\infty$-smooth vector field $b$ on $\mathbb{R}^2$ and a $C^2$-smooth matrix valued function $\sigma : \mathbb{R}^2 \to \mathbb{R}^{2 \times 2}$. Let $W$ be a standard two-dimensional Wiener process. In order to ensure that the stochastic Itô equation

$$dX_\epsilon = b(X_\epsilon) \, dt + \epsilon \sigma(X_\epsilon) \, dW$$

has a unique global strong solution, our first assumption is that both $b$ and $\sigma$ are Lipschitz and bounded, i.e. there is a constant $L > 0$ such that

$$|\sigma(x) - \sigma(y)| \vee |b(x) - b(y)| \leq L|x - y|, \quad x, y \in \mathbb{R}^2,$n

$$|\sigma(x)| \vee |b(x)| \leq L, \quad x \in \mathbb{R}^2,
where $| \cdot |$ denotes the Euclidean norm for vectors and Hilbert–Schmidt norm for matrices. These conditions can be weakened, but we prefer this setting to avoid multiple localization procedures throughout the text. For a general background on stochastic differential equations see, for example, [8].

We shall denote by $S = (S^t)_{t \in \mathbb{R}}$ the flow generated by $b$:

$$\frac{d}{dt} S^t x = b(S^t x), \quad S^0 x = x.$$ 

Let $V$ be a domain in $\mathbb{R}^2$ with piecewise $C^2$ boundary. We assume that the origin 0 belongs to $V$ and it is a unique fixed point for $S$ in $\bar{V}$, or, equivalently, a unique critical point for $b$ in $\bar{V}$. Therefore, $b(x) = Ax + Q(x)$, where $A = Db(0)$ and $Q$ is the nonlinear part of the vector field satisfying $|Q(x)| = O(|x|^2)$, $x \to 0$.

We assume that 0 is a hyperbolic critical point, i.e. the matrix $A$ has two eigenvalues $\lambda_+$ and $-\lambda_-$ satisfying $-\lambda_- < 0 < \lambda_+$. Without loss of generality, we suppose that the canonical vectors are the eigenvectors for the matrix, so that $A = \text{diag}(\lambda_+, -\lambda_-)$.

According to the Hadamard–Perron theorem (see, e.g., [11, section 2.7]), the curves $W^s$ and $W^u$ defined via

$W^s = \{ x \in \mathbb{R}^2 : |S^t x| \to 0 \text{ as } t \to \infty \}$

and

$W^u = \{ x \in \mathbb{R}^2 : |S^t x| \to 0 \text{ as } t \to -\infty \}$

are smooth, invariant under $S$ and tangent to $e_2$ and, respectively, to $e_1$ at 0. The curve $W^s$ is called the stable manifold of 0, and $W^u$ is called the unstable manifold of 0.

We assume that $W^u$ intersects $\partial V$ transversally at points $q_+$ and $q_-$ such that the segment of $W^u$ connecting $q_-$ and $q_+$ lies entirely inside $V$ and contains 0.

We fix a point $x_0 \in W^s \cap V$ and equip (1) with the initial condition

$X_\epsilon(0) = x_0 + \epsilon^\alpha \xi_\epsilon, \quad \epsilon > 0$,  

(2)

where $\alpha \in (0, 1]$ is fixed, and $(\xi_\epsilon)_{\epsilon > 0}$ is a family of random vectors independent of $W$, such that for some random vector $\xi_0$, $\xi_\epsilon \to \xi_0$ as $\epsilon \to 0$ in distribution.

If $\alpha \neq 1$, then we require that the term $\epsilon^\alpha \xi_\epsilon$ in (2) to be non-trivial in directions transversal to the flow:

$P[|\xi_0 \parallel b(x_0)| = 0] = 0$,  

(3)

where $\parallel$ denotes collinearity of two vectors. This condition can be weakened, but then one has to take into account smaller order corrections induced by $\epsilon^\alpha \xi_\epsilon$ in directions that are not tangential to $b(x_0)$. Our approach applies to these situations as well, but the analysis is more tedious and we restrict ourselves to condition (3).

We are studying the exit problem for the diffusion process $X_\epsilon$ in $V$. We are interested in the asymptotic distribution of the random point of exit of $X_\epsilon$ from $V$ given by $X_\epsilon(\tau^V_\epsilon)$, where $\tau^V_\epsilon$ is the stopping time defined by

$\tau^V_\epsilon = \tau^V_\epsilon(x_0) = \inf\{ t > 0 : X_\epsilon(t) \in \partial V \}$.  


3. Main result

The main result of this paper is the following:

**Theorem 1.** In the setting described above, there is a family of random vectors \((\phi_\epsilon)_{\epsilon > 0}\), a family of random variables \((\psi_\epsilon)_{\epsilon > 0}\), and a number

\[
\beta = \begin{cases} 
1, & \alpha \lambda_- \geq \lambda_+ \\
\frac{\alpha \lambda_-}{\lambda_+}, & \alpha \lambda_- < \lambda_+ + \alpha \lambda_-
\end{cases}
\]

such that

\[
X_\epsilon(t_\epsilon^V) = q_{\text{sgn}(\psi_\epsilon)} + \epsilon^\beta \phi_\epsilon.
\]

The random vector

\[
\Theta_\epsilon = \left(\psi_\epsilon, \phi_\epsilon, t_\epsilon^V + \frac{\alpha}{\lambda_+} \ln \epsilon\right)
\]

converges in distribution as \(\epsilon \to 0\).

The distribution of \(\psi_\epsilon, \phi_\epsilon\), and the distributional limit of \(\Theta_\epsilon\) will be described precisely.

To gain more intuition about the consequences of this theorem, refer to section 8.

The first version of this theorem is the main result in [9], where it is proven, under deterministic initial condition, that the exit occurs along either \(q_+\) or \(q_-\) each with probability \(1/2\). In [9] it is also shown that \(t_\epsilon^V / |\log \epsilon| \to \lambda_+^{-1}\) in probability as \(\epsilon \to 0\). The first improvement to this result was given in [5], which computes the asymptotic distribution of the sum \(t_\epsilon^V + \lambda_+^{-1} \ln \epsilon\) in the two-dimensional situation. Theorem 1 is stated in [3] with the restriction of not allowing resonances. We remove this restriction and hence extend the domain of applicability of the theory about heteroclinic networks presented in [3].

The proof of theorem 1 has essentially three parts involving the analysis of diffusion (i) along \(W^s\), (ii) in a small neighbourhood of the origin, (iii) along \(W^u\). We split the analysis by partitioning the domain \(V\) into overlapping subdomains as shown in figures 1–3. We will show that the exit distribution from stage (i) satisfies the conditions of the initial condition for stage (ii). Likewise, the exit distribution from stage (ii) satisfies the initial condition from stage (iii) and hence the exit distribution will be the exit distribution obtained from stage (iii).

The first part is based on a theorem borrowed from [3, lemma 9.2]. To state the theorem, we need to introduce \(\Phi_\epsilon(t)\) as the linearization of \(S\) along the orbit of \(x \in \mathbb{R}^2\), i.e., we define \(\Phi_\epsilon(t)\) to be the solution to the matrix ODE

\[
\frac{d}{dt} \Phi_\epsilon(t) = A(t) \Phi_\epsilon(t), \quad \Phi_\epsilon(0) = I,
\]

where \(A(t) = Db(S'x)\). The theorem reads as

**Theorem 2.** Let \(x \in \mathbb{R}^2\) and \((\xi_\epsilon)_{\epsilon > 0}\) be a family of random vectors independent of \(W\) and convergent in distribution, as \(\epsilon \to 0\), to \(\xi_0\). Suppose \(\alpha \in (0, 1]\) and let \(X_\epsilon\) be the solution of the stochastic differential equation (SDE) (1) with initial condition \(X_\epsilon(0) = x + \epsilon^\alpha \xi_\epsilon\). Then, for every \(T > 0\), the following representation holds true:

\[
X_\epsilon(T) = S^T x + \epsilon^\alpha \xi_\epsilon, \quad \epsilon > 0,
\]

where

\[
\xi_\epsilon \xrightarrow{\text{law}} \xi_0, \quad \epsilon \to 0,
\]

with

\[
\xi_0 = \Phi_\epsilon(T) \xi_0 + 1_{[\alpha=1]} N.
\]
Normal forms approach to diffusion near hyperbolic equilibria

Figure 1. Analysis of the diffusion $X_{\epsilon}$ along $\mathcal{W}^\prime$.

Figure 2. Analysis of the diffusion $X_{\epsilon}$ in a small neighbourhood of the origin.

\[ N = \Phi_x(T) \int_0^T \Phi_x(s)^{-1} \sigma(S^T x) \, dW(s). \]

If $\alpha = 1$ or assumption (3) holds, then $P[\tilde{\xi}_0 \parallel b(S^T x)] = 0$.

The second part of the analysis is the core of the paper. Theorem 3 below describes the behaviour of the process in a small neighbourhood $U$ of the origin. Note that since $x_0 \in \mathcal{W}^\prime$, one can choose $T$ large enough to ensure that $S^T x_0 \in \mathcal{W}^\prime \cap U$. Therefore, the conditions of the following result are met if we use the terminal distribution of theorem 2 (applied to the initial data given by (2)) as the initial distribution.
Theorem 3. There are two neighbourhoods of the origin $U \subset U' \subset V$, two positive numbers $\delta < \delta'$, and $C^2$ diffeomorphism $f : U' \to (-\delta', \delta')^2$, such that $f(U) = (-\delta, \delta)^2$ and the following property holds:

Suppose $x \in \mathcal{W}^D \cap U$, and $(\xi_\epsilon)_{\epsilon > 0}$ is a family of random variables independent of $W$ and convergent in distribution, as $\epsilon \to 0$, to $\xi_0$, where $\xi_0$ satisfies (3) with respect to $x$. Assume that $\alpha \in (0, 1]$ and that $X_\epsilon$ solves (1) with initial condition

$$X_\epsilon(0) = x + \epsilon^a \xi_\epsilon,$$

where $\xi_\epsilon$ satisfies condition (3) with respect to $x$.

There is also a family of random vectors $(\phi'_\epsilon)_{\epsilon > 0}$ and a family of random variables $(\psi'_\epsilon)_{\epsilon > 0}$, such that

$$X_\epsilon(\tau U_\epsilon) = g(\text{sgn}(\psi'_\epsilon) \delta e_1) + \epsilon^\beta \phi'_\epsilon,$$

where $g = f^{-1}$, $\beta$ is defined in (4), and the random vector

$$\Theta'_\epsilon = \left( \psi'_\epsilon, \phi'_\epsilon, \tau U_\epsilon + \frac{\alpha}{\lambda_+} \ln \epsilon \right)$$

converges in distribution as $\epsilon \to 0$.

The notation for $\Theta'_\epsilon$ and its components is chosen to match the notation involved in the statement of theorem 1. Random elements $\psi'_\epsilon, \phi'_\epsilon$ and the distributional limit of $\Theta'_\epsilon$ will be described precisely, see (27). Obviously, the symmetry or asymmetry in the limiting distribution of $\psi'_\epsilon$ results in the symmetric or asymmetric choice of exit direction so that the exits in the positive and negative directions are equiprobable or not. On the other hand, the limiting distribution of $\phi'_\epsilon$ which determines the asymptotics of the exit point can also be symmetric or asymmetric which results in the corresponding features of the random choice of the exit direction at the next saddle point visited by the diffusion. See section 8 for more insight.

In section 5 we prove theorem 3 using the approach based on normal forms.

The last part of the analysis is devoted to the exit from $V$ along $\mathcal{W}^D$. We need the following statement which is a specific case of the main result of [10].
Theorem 4. In the setting of theorem 2, assume additionally that (i) \( q = S^T x \in \partial V \); (ii) there is no \( t \in [0, T) \) with \( S' x \in \partial V \); (iii) \( b(q) \) is transversal (i.e. not tangent) to \( \partial V \) at \( q \). Then

\[
\tau^V_\epsilon \xrightarrow{P} T, \quad \epsilon \to 0
\]

and

\[
\epsilon^{-a} (X_\epsilon (\tau^V_\epsilon) - q) \xrightarrow{\text{Law}} \pi_{\tilde{\xi}_0}^F, \quad \epsilon \to 0,
\]

where \( \pi \) denotes the projection along \( b(q) \) onto the tangent line to \( \partial V \) at \( q \).

Now theorem 1 follows from the consecutive application of theorems 2–4 and with the help of the strong Markov property. In fact, in this chain of theorems, the conclusion of theorem 2 ensures that the conditions of theorem 3 hold, and the conclusion of the latter ensures that the conditions of theorem 4 hold. Note that the total time needed to exit \( V \) equals the sum of times described in the three theorems. Note also that at each step we can compute the limiting initial and terminal distributions explicitly. Theorems 2 and 4 contain the respective formulas in their formulations, and the explicit limiting distribution for \( \Theta_\epsilon^1 \) of theorem 3 is computed in (27).

4. Simplifying change of coordinates

In this section we start analysing the diffusion in the neighbourhood of the saddle point. The first step is to find a smooth coordinate change that would simplify the system. This can be done with the help of the theory of normal forms.

Let \( g \) be a \( C^\infty \) diffeomorphism of a neighbourhood of the origin with inverse \( f \). When \( X_\epsilon \) is close to the origin and belongs to the image of that neighbourhood under \( g \), we can use It\'o’s formula to see that

\[
X_\epsilon(t) = X_\epsilon(0) + \int_0^t b(X_\epsilon(s)) \, ds + \int_0^t \sigma(X_\epsilon(s)) \, dW(s).
\]

Normal form theory asserts (see [4, 7]) that for any pair of integers \( R \geq 1 \) and \( k \geq 1 \), there are two neighbourhoods of the origin \( \Omega_f \) and \( \Omega_k \) and a \( C^k \)-diffeomorphism \( f : \Omega_f \to \Omega_k \) with inverse \( g : \Omega_k \to \Omega_f \) such that

\[
(Dg(y))^{-1} b(g(y)) = Ay + P(y) + \mathcal{R}(y), \quad y \in \Omega_k,
\]

where \( P \) is a polynomial containing only resonant monomials of order at most \( R \) and \( \mathcal{R}(\xi) = O(|\xi|^{R+1}) \). If \( \lambda \) is non-resonant, then \( f \) can be chosen so that both \( P \) and \( \mathcal{R} \) in (9) are
identically zero. Moreover, due to [7, theorem 3, section 2], if $\lambda$ is one-resonant then $f$ can be chosen so that $R$ in (9) is identically zero. More precisely, if $\lambda$ is a one-resonant pair, then for any pair of integers $R \geq 1$ and $k \geq 1$, there are two neighbourhoods of the origin $\Omega_{f}$ and $\Omega_{g}$ and a $C^{k}$-diffeomorphism $f : \Omega_{f} \to \Omega_{g}$ with inverse $g : \Omega_{g} \to \Omega_{f}$ such that

$$(Dg(y))^{-1}b(g(y)) = Ay + P(y), \quad y \in \Omega_{g},$$

where $P$ is a polynomial that contains only resonant monomials.

Note that $(\lambda_{+}, -\lambda_{-})$ is either non-resonant or one-resonant (resonant cases that are not one-resonant are possible in higher dimensions where pairs of eigenvalues get replaced by vectors of eigenvalues). The non-resonant case (in any dimension) was studied in [3]. In this paper, we extend the analysis of [3] to the resonant case, i.e. the one-resonant case, given that we are working in two dimensions.

To find all resonant monomials of a given order $r \geq 2$, we have to find all the integer solutions to the two $2 \times 2$ systems of equations:

$$\alpha_{1}\lambda_{+} - \alpha_{2}\lambda_{-} = \pm\lambda_{\pm},$$
$$\alpha_{1} + \alpha_{2} = r.$$

Therefore, the power multi-indices of a resonant monomial of order $r$ has to coincide with one of the following:

$$(\alpha_{1}^{+}(r), \alpha_{2}^{+}(r)) = \frac{1}{\lambda_{+} + \lambda_{-}}(\lambda_{+} + r\lambda_{-}, (r - 1)\lambda_{+}),$$
$$(\alpha_{1}^{-}(r), \alpha_{2}^{-}(r)) = \frac{1}{\lambda_{+} + \lambda_{-}}((r - 1)\lambda_{+}, r\lambda_{+} + \lambda_{-}).$$

Let us make some elementary observations on integer solutions of these equations for $r \geq 2$.

1. None of the solution indices can be 0. Moreover, neither $\alpha_{1}^{+}(r)$ nor $\alpha_{2}^{-}(r)$ can be equal to 1.
2. As functions of $r$, $\alpha_{i}^{±}(r)$ are increasing.
3. Expressions (11) and (12) cannot be an integer for $r = 2$.
4. The term $P = (P_{1}, P_{2})$ in (10) satisfies $P_{1}(y) = O(|y_{1}||y_{2}|)$ and $P_{2}(y) = O(|y_{1}|y_{2}^{2})$.

This observation is a consequence of observations 1 and 3 since they imply that resonant multi-indices have to satisfy $\alpha^{+}(r) \geq (2, 1)$ and $\alpha^{-}(r) \geq (1, 2)$ coordinatewise.

5. If at least one of the coordinates $y_{1}$ and $y_{2}$ is zero, then $P(y_{1}, y_{2}) = 0$. This is a direct consequence of the previous observation.

Given all these considerations, the main theorem of this section is a simple consequence of [7].

**Theorem 5.** In the setting described in section 2, there is a number $\delta' > 0$, a neighbourhood of the origin $U'$, and a $C^{2}$-diffeomorphism $f : U' \to (-\delta', \delta')$ with inverse $g : (-\delta', \delta')^{2} \to U'$ such that the following property holds.

If $X_{\epsilon}(0) \in U$, then the stochastic process $Y_{\epsilon} = (Y_{\epsilon,1}, Y_{\epsilon,2})$ given by

$$\dot{Y}_{\epsilon}(t) = f(X_{\epsilon}(t \wedge t_{\epsilon}^{U}))$$

satisfies the following system of SDEs up to $t_{\epsilon}^{U}$:

$$dY_{\epsilon,1} = (\lambda_{+}Y_{\epsilon,1} + H_{1}(Y_{\epsilon}, \epsilon)) \, dt + \epsilon\tilde{\sigma}_{1}(Y_{\epsilon}) \, dW$$
$$dY_{\epsilon,2} = (-\lambda_{-}Y_{\epsilon,2} + H_{2}(Y_{\epsilon}, \epsilon)) \, dt + \epsilon\tilde{\sigma}_{2}(Y_{\epsilon}) \, dW,$$

where $\tilde{\sigma}_{i} : (-\delta', \delta')^{2} \to \mathbb{R}$ are $C^{1}$ functions for $i = 1, 2$. The functions $H_{i}$ are given by $H_{i} = \tilde{H}_{i} + \epsilon^{2}\Psi_{i}$, where $\Psi_{i} : (-\delta', \delta')^{2} \to \mathbb{R}^{2}$ are continuous bounded functions, and
\( \hat{H}_i : (-\delta', \delta')^2 \times [0, \infty) \) are polynomials, so that for some constant \( K_1 > 0 \) and for any \( y \in (-\delta', \delta')^2 \),
\[
|\hat{H}_1(y)| \leq K_1|y_1|^{\omega_1} |y_2|^{\omega_2}, \\
|\hat{H}_2(y)| \leq K_1|y_1|^{\omega_1} |y_2|^{\omega_2}.
\]

Here, the integer numbers \( \omega_i, i = 1, 2 \), are such that \((\omega_1^+, \omega_2^+)\) is of the form (11) for some choice of \( r = r_1 \geq 3 \), and \((\omega_1^-, \omega_2^-)\) is of the form (12) for some choice \( r = r_2 \geq 3 \). In particular,
\[
|H_1(y, \epsilon)| \leq K_1 y_1^2 |y_2| + K_2 \epsilon^2, \\
|H_2(y, \epsilon)| \leq K_1 |y_1|^2 y_2^2 + K_2 \epsilon^2,
\]
for some constants \( K_1 > 0 \) and \( K_2 > 0 \).

5. Proof of theorem 3

In this section we derive theorem 3 from several auxiliary statements. Their proofs are postponed to later sections.

Theorem 5 allows us to work with the process \( Y_\epsilon = f(X_\epsilon) \) instead of \( X_\epsilon \) up to the time that \( Y_\epsilon \) leaves \((-\delta', \delta')^2\).

If we take \( \delta \in (0, \delta') \), then for the initial conditions considered in theorem 3 and given in (5),
\[
P[X_\epsilon(0) \in U'] \to 1, \quad \epsilon \to 0,
\]
i.e.
\[
P[Y_\epsilon(0) \in (-\delta', \delta')^2] \to 1, \quad \epsilon \to 0.
\]

Moreover, denoting \( f(x) \) by \( y = (0, y_2) \) we can write
\[
Y_\epsilon(0) = y + \epsilon^2 \chi = (\epsilon^2 \chi_1, y_2 + \epsilon^2 \chi_2), \quad \epsilon > 0,
\]
where \( \chi = (\chi_1, \chi_2) \) is a random vector convergent in distribution to \( \chi_0 = (\chi_{0,1}, \chi_{0,2}) = Df(x) \xi_0 \). Due to the hypothesis in theorem 3, we note that the distribution of \( \chi_{0,1} \) has no atom at 0.

Let us take any \( \lambda < 1 \) such that
\[
1 - \frac{\lambda_+}{\lambda_-} < p < \frac{\lambda_-}{\lambda_+ + \lambda_-}, \quad (15)
\]
and define the following stopping time:
\[
\hat{\tau}_c = \inf \{ t : |Y_{\epsilon,1}(t)| = \epsilon^{ap} \}.
\]

Up to time \( \hat{\tau}_c \), the process \( X_\epsilon \) mostly evolves along the stable manifold \( W^s \). After \( \hat{\tau}_c \), it evolves mostly along the unstable manifold \( W^u \). Process \( Y_\epsilon \) evolves accordingly, along the images of \( W^s \) and \( W^u \) coinciding with the coordinate axes.

Let us introduce random variables \( \eta^\pm \) via
\[
\eta^+_\epsilon = \epsilon^{-\frac{\lambda_+}{\lambda_-}} Y_{\epsilon,1}(\hat{\tau}_c), \\
\eta^-_\epsilon = \epsilon^{-\frac{1-p\mu_-}{\mu_-}} Y_{\epsilon,2}(\hat{\tau}_c).
\]

Also we define the distribution of random vector \((\eta^+_0, \eta^-_0)\) via
\[
\eta^+_0 = \chi_{0,1} + 1_{\{\mu_1=1\}} N^+, \\
\eta^-_0 = |\eta^+_0|^\frac{\lambda_-}{\lambda_+} y_2.
\]

\( N^+ \) is a standard normal random variable.
where
\[ N^* = \int_0^\infty e^{-\lambda s} \sigma_1(0, e^{-\lambda s} y_2) \, dW \] (17)
is independent of \( \chi_{0,1} \).

**Lemma 6.** If the first inequality in (15) holds, then
\[ P\{Y_{\epsilon,1}(\hat{\tau}_\epsilon) = e^{\alpha p \text{sgn} \eta^+_\epsilon}\} \to 1, \quad \epsilon \to 0. \] (18)
and
\[ \left( \eta^+_{\epsilon}, \eta^-_{\epsilon}, \hat{\tau}_\epsilon + \frac{\alpha}{\lambda_+} (1 - p) \log \epsilon \right) \xrightarrow{\text{Law}} \left( \eta^+_0, \eta^-_0, -\frac{1}{\lambda_+} \log |\eta^+_0| \right), \quad \epsilon \to 0. \] (19)

We prove this lemma in section 6. Along with the strong Markov property, it allows us to reduce the study of the evolution of \( Y_\epsilon \) after \( \hat{\tau}_\epsilon \) to studying the solution of system (13)–(14) with initial condition
\[ Y_\epsilon(0) = (e^{\alpha p \text{sgn} \eta^+_\epsilon}, e^{(1 - p) \lambda_-/\lambda_+} \eta^-_\epsilon), \] (20)
where
\[ (\eta^+_\epsilon, \eta^-_\epsilon) \xrightarrow{\text{Law}} (\eta^+_0, \eta^-_0), \quad \epsilon \to 0. \] (21)

We denote
\[ \tau_\epsilon = \tau_\epsilon(\delta) = \inf \{ t \geq 0 : |Y_{\epsilon,1}(t)| = \delta \}. \] (22)

Our next goal is to describe the behaviour of \( Y(\tau_\epsilon) \). To that end, we introduce a random variable \( \theta \) via
\[ \theta \xrightarrow{\text{Law}} \begin{cases} N, & \alpha \lambda_- > \lambda_+, \\ \left( \frac{|\eta^+_0|}{\delta} \right)^{\lambda_-/\lambda_+} y_2 + N, & \alpha \lambda_- = \lambda_+, \\ \left( \frac{|\eta^+_0|}{\delta} \right)^{\lambda_-/\lambda_+} y_2, & \alpha \lambda_- < \lambda_+. \end{cases} \] (23)
where the distribution of \( N \) conditioned on \( \eta^+_0 \), on \( \{\text{sgn} \eta^+_0 = \pm 1\} \) is centred Gaussian with variance
\[ \sigma_{\pm} = \int_{-\infty}^0 e^{2\lambda s} |\tilde{\sigma}_2(\pm \delta e^{\lambda s}, 0)|^2 \, ds. \]

Let us also recall that \( \beta \) is defined in (4).

**Lemma 7.** Consider the solution to system (13)–(14) equipped with initial conditions (20) satisfying (21). If the second inequality in (15) holds, then
\[ P\{|Y_{\epsilon,1}(\tau_\epsilon)| = \delta\} \to 1, \quad \epsilon \to 0, \] (24)
\[ \tau_\epsilon + \frac{\alpha p}{\lambda_+} \log \epsilon \to \frac{1}{\lambda_-} \log \delta, \] (25)
\[ e^{-\beta Y_{\epsilon,2}(\tau_\epsilon)} \xrightarrow{\text{Law}} \theta. \] (26)

Moreover, if \( \beta < 1 \), then the convergence in probability also holds.

A proof of this lemma is given in section 7.

Now theorem 3 follows from lemmas 6 and 7. In fact, the strong Markov property and (18) imply
\[ P\{\tau_\epsilon^U = \hat{\tau}_\epsilon + \tau_\epsilon(\delta)\} \to 1, \quad \epsilon \to 0, \]
so that the asymptotics for \( \tau_{U}^{\epsilon} \) is defined by that of \( \hat{\tau}_{\epsilon}^{+} \) and \( \tau_{\epsilon}(\delta) \). It is also clear that one can set

\[
\psi'_{\epsilon} = \eta_{0}^{+}, \quad \phi'_{\epsilon} = Dg(\text{sgn}(\eta_{0}^{+})\delta e_{1})Y_{\epsilon}(\tau_{\epsilon}),
\]

so that the limiting distribution of \( \Theta'_{\epsilon} \) is given by

\[
\left( \eta_{0}^{+}, \; Dg(\text{sgn}(\eta_{0}^{+})\delta e_{1})(\theta_{2})^{\dagger}, \; \frac{1}{\lambda^{+}} \log \frac{\delta}{|\eta_{0}^{+}|} \right),
\]

(27)

where random variables \( \eta_{0}^{+} \) and \( \theta \) are defined in (16) and (23).

6. Proof of lemma 6

In this section we shall prove lemma 6 using several auxiliary lemmas. We start with some terminology.

Definition 1. Given a family \((\xi_{\epsilon})_{\epsilon > 0}\) of random variables or random vectors and a function \( h : (0, \infty) \rightarrow (0, \infty) \) we say that \( \xi_{\epsilon} = O_p(h(\epsilon)) \) if for some \( \epsilon_0 > 0 \) distributions of \((\xi_{\epsilon}/h(\epsilon))_{0 < \epsilon < \epsilon_0}\) form a tight family, i.e. for any \( \delta > 0 \) there is a constant \( K_{\delta} > 0 \) such that

\[
P\{|\xi_{\epsilon}| > K_{\delta}h(\epsilon)\} < \delta, \quad 0 < \epsilon < \epsilon_0.
\]

Definition 2. A family of random variables or random vectors \((\xi_{\epsilon})_{\epsilon > 0}\) is called slowly growing as \( \epsilon \to 0 \) (or just slowly growing) if \( \xi_{\epsilon} = O_P(\epsilon^{-r}) \) for all \( r > 0 \).

Our first lemma estimates the martingale component of the solution of SDEs (13) and (14).

Let us define

\[
S_{+}^{\epsilon}(T) = \sup_{t \leq T} \left| \int_{0}^{t} e^{-\lambda s} \tilde{\sigma}_{1}(Y_{\epsilon}(s)) dW(s) \right|, \quad T > 0,
\]

\[
S_{-}^{\epsilon}(T) = \sup_{t \leq T} \left| \int_{0}^{t} e^{-\lambda(t-s)} \tilde{\sigma}_{2}(Y_{\epsilon}(s)) dW(s) \right|, \quad T > 0.
\]

Lemma 8. Suppose \((\tau_{\epsilon})_{\epsilon > 0}\) is a family of stopping times (w.r.t. the natural filtration of \( W \)). Then

\[
S_{+}^{\epsilon}(\tau_{\epsilon}) = O_P(1).
\]

If additionally \((\tau_{\epsilon})_{\epsilon > 0}\) is slowly growing, then \( S_{-}^{\epsilon}(\tau_{\epsilon}) \) is also slowly growing.

Proof. Both estimates are elementary. The first one is an easy consequence of the martingale property of the stochastic integral involved in the definition of \( S_{+}^{\epsilon} \), and the BDG inequality (see [8, theorem 3.3.28]). As for the second one, we note that the stochastic integral in the definition of \( S_{-}^{\epsilon} \) behaves essentially like an Ornstein–Uhlenbeck process, and similar bounds apply.

Lemma 9. Suppose \( Y_{\epsilon} \) is the solution of equations (13)–(14) with initial conditions given by

\[
Y_{\epsilon,1}(0) = e^{\alpha} \chi_{\epsilon,1} \quad \text{and} \quad Y_{\epsilon,2}(0) = y_{2} + e^{\alpha} \chi_{\epsilon,2},
\]

(28)

where distributions of random variables \((\chi_{\epsilon,1})_{\epsilon > 0}\) and \((\chi_{\epsilon,2})_{\epsilon > 0}\) form tight families. Let us fix any \( R > 0 \) and denote \( l_{\epsilon} = t_{\epsilon}^{\epsilon} \wedge \left( -\frac{2}{\lambda \epsilon} \log \epsilon + R \right) \) for \( \epsilon > 0 \). Then

\[
\sup_{t \leq l_{\epsilon}} e^{-\lambda t}|Y_{\epsilon,1}(t)| = O_P(e^{\alpha}),
\]

and the family

\[
\left( e^{-\alpha} \sup_{t \leq l_{\epsilon}} |Y_{\epsilon,2}(t) - e^{-\lambda t}(y_{2} + e^{\alpha} \chi_{\epsilon,2})| \right)_{\epsilon > 0}
\]

is slowly growing.

Proof. The tightness property implies that without loss of generality we can assume that \(|\chi_{\epsilon,1}|, |\chi_{\epsilon,2}| < C\) for some constant \( C > 0 \) and every \( \epsilon > 0 \).
and stopping times

\[ P[S_{q}^{+}(l_{e}) > c] < \gamma / 2 \]

where \( q \) is an arbitrary number satisfying \( 0 < q < \alpha \). Let us introduce a constant \( K = (3e) \vee C \) and stopping times

\[
\begin{align*}
\beta_+ &= \inf \{ t \geq 0 : e^{-\lambda_+ t} \vert Y_{e,1}(t) \vert \geq 2Ke^\alpha \}, \\
\beta_- &= \inf \{ t \geq 0 : \vert Y_{e,2}(t) - e^{-\lambda_- t} (y_{2} + e^\alpha Y_{e,2}) \vert \geq 2Ke^{-\alpha q} \}, \\
\beta &= \beta_+ \wedge \beta_- \wedge l_{e}.
\end{align*}
\]

We start with an estimate for \( Y_{e,1} \). Duhamel’s principle for (13), theorem 5 and lemma 8 imply that the estimate

\[
\sup_{t \leq \beta} e^{-\lambda_+ t} \vert Y_{e,1}(t) \vert \leq \epsilon_0 K + K_1 \int_{0}^{\beta} e^{-\lambda_+ s} \vert Y_{e,1}(s) \vert \, ds + K_2 e^{\frac{3}{2} + \epsilon_0 S_{q}^{+}(\beta)} \leq \epsilon_0 K + K_1 \int_{0}^{\beta} e^{-\lambda_+ s} \vert Y_{e,1}(s) \vert \, ds + K_2 e^{\frac{3}{2} + \epsilon} + \frac{1}{3} (29) \]

holds with probability at least \( 1 - \gamma / 2 \). We analyse each term in the rhs of equation (29).

Let us start with the integral in (29). For \( s \leq \beta \), we see that

\[
Y_{e,1}(s)^2 \vert Y_{e,2}(s) \vert \leq 4K^2 e^{2\alpha} e^{2\alpha s} \left( (Y_{e,2}(s) - e^{-\lambda_- t} (y_{2} + e^\alpha Y_{e,2})) + e^{-\lambda_- t} \vert y_{2} + e^\alpha Y_{e,2} \vert \right) \leq 8K^3 e^{2\alpha q} e^{2\alpha s} + 4K^2 e^{2\alpha} e^{2\alpha (2\alpha + \lambda_+)} \vert y_{2} \vert + \epsilon^2 C.
\]

Therefore,

\[
K_1 \int_{0}^{\beta} e^{-\lambda_+ s} \vert Y_{e,1}(s) \vert \, ds \leq 8K^3 K_1 e^{\lambda_+ R} e^{2\alpha q} + 4K_1 K^2 e^{2\alpha} \vert y_{2} \vert + \epsilon^2 C \int_{0}^{\beta} e^{(\lambda_+ - \lambda_-) t} \, ds
\]

for all \( \epsilon > 0 \) small enough. Note that this is a rough estimate, the constants on the rhs are not optimal but sufficient for our purposes. This also applies to some other estimates in this proof.

Let us estimate the integral on the rhs of (30). When \( \lambda_+ > \lambda_- \), the integral is bounded by

\[
\frac{1}{\lambda_+ - \lambda_-} e^{(\lambda_+ - \lambda_-) \beta} \leq \frac{e^{(\lambda_+ - \lambda_-) R}}{\lambda_+ - \lambda_-} e^{-\alpha + \lambda_+ / \lambda_+}.
\]

if \( \lambda_- < \lambda_+ \), then the integral on the rhs of (30) is bounded by \( (\lambda_+ - \lambda_-)^{-1} \); if \( \lambda_+ = \lambda_- \), then the integral is bounded by \( 2\alpha \lambda_+^{-1} \log \epsilon \). Hence, for some constant \( K_{\lambda_+\lambda_-} > 0 \) and \( \epsilon > 0 \) small enough,

\[
K_1 \int_{0}^{\beta} e^{-\lambda_+ s} \vert Y_{e,1}(s) \vert \, ds \leq K e^{\alpha / 2} + K_{\lambda_+\lambda_-} e^{2\alpha (1 - \lambda_+ / \lambda_+)} \vert \log \epsilon \vert \leq K e^{\alpha / 6}.
\]

Also, for \( \epsilon > 0 \) small enough,

\[
K_2 e^{2 / \lambda_+ + \epsilon K / 3} < K e^{\alpha / 2}.
\]

From (29), (31) and (32) we obtain that for all \( \epsilon > 0 \) small enough, the event

\[
A = \left\{ \sup_{t \leq \beta} e^{-\lambda_+ t} \vert Y_{e,1}(t) \vert \leq 5K e^{\alpha / 3} \right\}
\]

is such that \( P(A) > 1 - \gamma / 2 \).
Let us now consider $Y_{\epsilon,2}(t)$ and denote

$$Z_{\epsilon}(t) = Y_{\epsilon,2}(t) - e^{-\lambda_{-}t}(y_2 + e^{\alpha} x_{\epsilon,2}).$$

Duhamel’s principle for $Y_{\epsilon,2}$, the definition of $\beta$, theorem 5 and lemma 8 imply that the inequalities

$$\sup_{t \leq \beta} |Z_{\epsilon}(t)| \leq K_1 \sup_{t \leq \beta} \int_0^t e^{-\lambda_{-}(t-s)} |Y_{\epsilon,1}(s)|^a \ |Y_{\epsilon,2}(s)|^b \ ds + K_2 \epsilon^{2}/\lambda_{-} + \epsilon S_{\epsilon}(\beta)$$

$$\leq K_1 \sup_{t \leq \beta} \int_0^t e^{-\lambda_{-}(t-s)} |Y_{\epsilon,1}(s)|^a \ |Y_{\epsilon,2}(s)|^b \ ds$$

$$+ e^{2\beta} (K_2 \epsilon^{2-a+1}/\lambda_{-} + \epsilon^{1-a+1} S_{\epsilon}(\beta))$$

$$\leq 2^{a_0} e^{a_0 \beta} K^{a_0} \sup_{t \leq \beta} e^{\lambda_{-}t} \int_0^t e^{(\lambda_{-} \alpha_{\epsilon})/(\alpha_{\epsilon}) |Y_{\epsilon,2}(s)|^b} \ ds + e^{a_0} K/2 \quad (33)$$

hold with probability at least $1 - \gamma/2$ and for all $\epsilon > 0$ small enough. We analyse the integral term in (33). Note that, from the definition of $\beta$, and the inequality $(a + b)' \leq 2^{a_0} (a + b')$, we have that for any $t \leq \beta$ and any $\epsilon > 0$ small enough,

$$|Y_{\epsilon,2}(t)|^b_2 \leq 2^{a_0-1} Z_{\epsilon}(t)^b_2 + e^{-\alpha_{\epsilon} \lambda_{-}t} |y_2 + e^{\alpha} x_{\epsilon,2}|^b_2$$

$$\leq 2^{a_0-1} K^{a_0} e^{2(\alpha_{\epsilon} \lambda_{-} - 1)} |x_{\epsilon,2}|^b_2$$

$$\leq e^{2(\alpha_{\epsilon} \lambda_{-} - 1)} (2 K^{a_0} + e^{a_0} |x_{\epsilon,2}|^b_2) + 2^{(\alpha_{\epsilon} \lambda_{-} - 1)} e^{-\alpha_{\epsilon} \lambda_{-}t} |y_2|^{b_2}.$$

Hence there is a constant $K_{a_0} > 0$ such that

$$|Y_{\epsilon,2}(t)|^b_2 \leq e^{a_0 (\alpha_{\epsilon} \lambda_{-})} K_{a_0} + K_{a_0} e^{-\alpha_{\epsilon} \lambda_{-}t}, \quad t \leq \beta.$$

Using the last inequality, the definition of $\beta$, and the fact $\alpha_{\epsilon} \lambda_{-} = (\alpha_{\epsilon} - 1) \lambda_{-} = 0$ from theorem 5, we obtain

$$e^{a_0 t} e^{-\lambda_{-}t} \int_0^t e^{(\lambda_{-} \alpha_{\epsilon} \lambda_{-})t} |Y_{\epsilon,2}(s)|^b_2 \ ds$$

$$\leq e^{a_0 (\alpha_{\epsilon} \lambda_{-})} e^{\lambda_{-} \alpha_{\epsilon} \beta} K_{a_0} e^{-\alpha_{\epsilon} \lambda_{-}t} \lambda_{-} + \alpha_{\epsilon} \lambda_{+} + K_{a_0} e^{a_0 \beta} \int_0^t e^{(\alpha_{\epsilon} \lambda_{-} - (\alpha_{\epsilon} \lambda_{-} - 1) \lambda_{-})t} |x_{\epsilon,2}|^b_2 \ ds$$

$$\leq e^{a_0 (\alpha_{\epsilon} \lambda_{-})} K_{a_0} e^{a_0 \beta} \lambda_{+} + \alpha_{\epsilon} \lambda_{+} + K_{a_0} e^{a_0 \beta} \lambda_{+}. \quad (34)$$

Again, from theorem 5 we know that $\alpha_{\epsilon} \lambda_{+} \geq 1$ and $\alpha_{\epsilon} \lambda_{-} \geq 2$ which together with (34) imply that for all $\epsilon > 0$ small enough

$$2^{a_0} e^{a_0 \beta} K^{a_0} \sup_{t \leq \beta} e^{\lambda_{-}t} \int_0^t e^{(\lambda_{-} \alpha_{\epsilon} \lambda_{-})t} |Y_{\epsilon,2}(s)|^b_2 \ ds \leq K e^{a_0 t}/6. \quad (35)$$

Using (35) and (33) we conclude that the event

$$B = \left\{ \sup_{t \leq \beta} |Y_{\epsilon,2}(t) - e^{-\lambda_{-}t}(y_2 + e^{\alpha} x_{\epsilon,2})| \leq 2K e^{a_0 t}/3 \right\}$$

is such that $P(B) \geq 1 - \gamma/2$, for all $\epsilon > 0$ small enough.
The proof will be complete once we show that $\beta = l_\epsilon$ with probability at least $1 - \gamma$.

The latter is a consequence of the following chain of inequalities that hold for all $\epsilon > 0$ small enough:

\[
P(\beta_+ \wedge \beta_- \leq l_\epsilon \mid A \cap B) + P(A^c) + P(B^c) \\
\leq P(\beta_+ \wedge \beta_- \leq l_\epsilon \mid A \cap B) + \gamma \\
\leq P(\beta_+ \leq \beta_- \wedge l_\epsilon \mid A) + P(\beta_- \leq \beta_+ \wedge l_\epsilon \mid B) + \gamma \\
= P(2 \leq 5/3) + P(2 \leq 2/3) + \gamma = \gamma.
\]

□

Let us now analyse the evolution of the process $Y_\epsilon$ up to time $\hat{\tau}_\epsilon \wedge \tau^U_\epsilon$. We start with an application of Duhamel’s principle:

\[
Y_{\epsilon,1}(t) = e^{\lambda t} Y_{\epsilon,1}(0) + \int_0^t e^{\lambda(t-s)} H_1(Y_\epsilon(s), \epsilon) \, ds + \epsilon e^{\lambda t} N_0^+(t),
\]

\[
Y_{\epsilon,2}(t) = e^{-\lambda t} Y_{\epsilon,2}(0) + \int_0^t e^{-\lambda(t-s)} H_2(Y_\epsilon(s), \epsilon) \, ds + \epsilon N_0^-(t),
\]

where $N_0^\pm(t)$ are defined by

\[
N_0^+(t) = \int_0^t e^{-\lambda s} \tilde{\sigma}_1(Y_\epsilon(s)) \, dW(s),
\]

\[
N_0^-(t) = \int_0^t e^{-\lambda(t-s)} \tilde{\sigma}_2(Y_\epsilon(s)) \, dW(s).
\]

Lemma 10.

\[
\sup_{t \leq \hat{\tau}_\epsilon} |Y_{\epsilon,2}(t) - e^{-\lambda t} y_2| = O_P(\epsilon^\alpha).
\]

Proof. Duhamel’s principle, theorem 5, and the definition of $\hat{\tau}_\epsilon$ imply that for some $K > 0$,

\[
|Y_{\epsilon,2}(t) - e^{-\lambda t} y_2| \leq e^{\alpha t} |X_{\epsilon,2}| \int_0^t e^{-\lambda(t-s)} \left( K_1 |Y_{\epsilon,1}(s)| Y_{\epsilon,2}^2(s) + K_2 \epsilon \right) \, ds + \epsilon S_\epsilon^-(t) \\
\leq e^{\alpha t} |X_{\epsilon,2}| + K \epsilon + \epsilon^{2-p} S_\epsilon^-(\hat{\tau}_\epsilon)
\]

for any $t \in (0, \hat{\tau}_\epsilon)$. The result follows since by lemma 8 the rhs is $O_P(\epsilon^{2-p})$.

□

As a simple corollary of this lemma, the first statement in theorem 6 follows:

Corollary 11. As $\epsilon \to 0$,

\[
P(\tau^U_\epsilon < \hat{\tau}_\epsilon) \to 0.
\]

In particular, (18) holds true.

Lemma 12. Let

\[
N_0^+(t) = \int_0^t e^{-\lambda s} \tilde{\sigma}_1(0, e^{-\lambda s} y_2) \, dW.
\]

Then

\[
\sup_{t \leq \hat{\tau}_\epsilon} |N_0^+(t) - N_0^+(t)| \overset{L^2}{\to} 0, \quad \epsilon \to 0.
\]
Proof. BDG inequality implies that for some constants $C_1, C_2 > 0$,
\[
E \sup_{t \leq \tilde{\tau}_\epsilon} |N^\epsilon_s(t) - N^0_s(t)|^2 \leq C_1 E \int_0^\tilde{\tau}_\epsilon e^{-2s_\epsilon} \tilde{\sigma}_1(Y_{\epsilon,1}(s), Y_{\epsilon,2}(s)) - (0, e^{-\lambda s_\epsilon} y_2)|^2 \, ds
\]
\[
\leq C_2 E \sup_{t \leq \tilde{\tau}_\epsilon} |\tilde{\sigma}_1(Y_{\epsilon,1}(s), Y_{\epsilon,2}(s)) - \tilde{\sigma}_1(0, e^{-\lambda s_\epsilon} y_2)|^2. 
\] (39)

From lemma 10 and the definition of $\tilde{\tau}_\epsilon$, it follows that
\[
\sup_{t \leq \tilde{\tau}_\epsilon} |\langle Y_{\epsilon,1}(t), Y_{\epsilon,2}(t) \rangle - (0, e^{-\lambda_\epsilon \tau t} y_2)| = O_P(\epsilon^{\alpha p}). 
\] (40)

The desired convergence follows now from (39), (40), and the boundedness and Lipschitzness of $\tilde{\sigma}_1$.

We are now in position to give the first rough asymptotics for the time $\tilde{\tau}_\epsilon$. From now on we restrict ourselves to the event $\{\epsilon^anders \mid \hat{\tau}_\epsilon \}$ since due to corollary 11 its probability is arbitrarily high.

Lemma 13. As $\epsilon \to 0$,
\[
P \left\{ \tilde{\tau}_\epsilon > \frac{\alpha}{\lambda_\epsilon} \log \epsilon \right\} \to 0.
\]

Proof. Let $u_\epsilon$ be the solution to the following SDE:
\[
du_\epsilon(t) = \lambda_u u_\epsilon(t) \, dt + \epsilon \tilde{\sigma}_1(Y_\epsilon(t)) \, dW(t),
\]
\[
u_\epsilon(0) = e^u \chi_{\epsilon,1}.
\]

Let us take $\delta_0 \in (0, 1)$ to be specified later and consider the following stopping time
\[
\tilde{\tau}_\epsilon = \inf \left\{ t : |u_\epsilon(t)| = \epsilon^{\alpha \delta_0} \right\}.
\]

Duhamel’s principle for $u_\epsilon$ writes as
\[
u_\epsilon(t) = e^{\alpha \delta_0 t} \chi_{\epsilon,1} + e^{\alpha \delta_0 t} N^\epsilon_s(t)
\]
\[
= e^{\alpha \delta_0 t} \eta_\epsilon(t),
\]
with
\[
\eta_\epsilon(t) = \chi_{\epsilon,1} + e^{1-\alpha \delta_0} N^\epsilon_s(t).
\] (41)

Hence, the definition of $\tilde{\tau}_\epsilon$ implies $\epsilon^{\alpha \delta_0} = e^{\alpha \delta_0} \tilde{\sigma}_1(\eta_\epsilon(\tilde{\tau}_\epsilon))$, so that
\[
\tilde{\tau}_\epsilon = -\frac{\alpha}{\lambda_\epsilon} (1 - \delta_0) \log \epsilon - \frac{1}{\lambda_\epsilon} \log |\eta_\epsilon(\tilde{\tau}_\epsilon)|.
\]

Due to (41) and lemma 12, the distributions of $\frac{1}{\alpha \delta_0} \log |\eta_\epsilon(\tilde{\tau}_\epsilon)|$ form a tight family. Therefore,
\[
\lim_{\epsilon \to 0} P \left\{ \tilde{\tau}_\epsilon > -(1 - \delta_0) \frac{\alpha}{\lambda_\epsilon} \log \epsilon \right\} = 0.
\] (42)

This fact allows us to use lemma 9 to estimate $Y_\epsilon$ up to $\tilde{\tau}_\epsilon \wedge \hat{\tau}_\epsilon$. From (36), the difference $\Delta_\epsilon = Y_{\epsilon,1} - u_\epsilon$ is given by
\[
\Delta_\epsilon(t) = e^{\lambda_\epsilon t} \int_0^t e^{-\lambda_\epsilon s} H_1(Y_\epsilon(s), \epsilon) \, ds.
\]

We can use (42) to justify the application of lemma 9 up to time $\tilde{\tau}_\epsilon \wedge \hat{\tau}_\epsilon$. Then, we combine theorem 5, lemma 9, and the definition of $\tilde{\tau}_\epsilon$ to see that
\[
\sup_{t \leq \tilde{\tau}_\epsilon \wedge \hat{\tau}_\epsilon} e^{-\lambda_\epsilon t} |H_1(Y_\epsilon(t), \epsilon)| \leq K_1 \sup_{t \leq \tilde{\tau}_\epsilon \wedge \hat{\tau}_\epsilon} \left( \left( e^{-\lambda_\epsilon t} |Y_{\epsilon,1}(t)| \right) |Y_{\epsilon,1}(t)| \cdot |Y_{\epsilon,2}(t)| \right) + K_2 \epsilon^2
\]
\[
= O_P(\epsilon^{\alpha \delta_0}).
\]
and 
\[ e^{\lambda_{\epsilon} \wedge \tilde{\tau}_\epsilon} = O_P \left( e^{-\alpha \left( 1 - \delta_0 \right)} \right). \]

These two estimates together with (42) imply 
\[ \sup_{t \leq \hat{\tau}_\epsilon \wedge \tilde{\tau}_\epsilon} |\Delta_e(t)| = O_P \left( \epsilon^{\alpha(\delta_0 + 2)} |\log \epsilon| \right). \]

On the one hand, (42) implies 
\[ P \left( \left\{ \hat{\tau}_\epsilon > -\frac{\alpha}{\lambda_e} \log \epsilon \right\} \cap \left\{ \hat{\tau}_\epsilon \leq \tilde{\tau}_\epsilon \right\} \right) \rightarrow 0. \]

On the other hand, if \( \hat{\tau}_\epsilon > \tilde{\tau}_\epsilon \) then 
\[ |Y_{\epsilon,1}(\tilde{\tau}_\epsilon)| = \left| e^{\alpha \delta_0} + O_P(\epsilon^{\alpha(\delta_0 + 2)} |\log \epsilon|) \right| \]
and 
\[ |Y_{\epsilon,1}(\tilde{\tau}_\epsilon)| < \epsilon^{\alpha p}. \]

These relations contradict each other for sufficiently small \( \epsilon \) if we choose \( \delta_0 < p \). So, this choice of \( \delta_0 \) guarantees that 
\[ P \left\{ \hat{\tau}_\epsilon > \tilde{\tau}_\epsilon \right\} \rightarrow 0 \]
implies the result. \( \square \)

**Proof of lemma 6.** Recall that we work on the high probability event \( \{\hat{\tau}_\epsilon < \tau_{U\epsilon}\} \). Hence, for each \( \epsilon > 0 \), we have the identity 
\[ \epsilon^{\alpha p} = e^{\alpha e^{\lambda_{\epsilon} \wedge \tilde{\tau}_\epsilon} |\eta_\epsilon^+|}. \]

Solving for \( \hat{\tau}_\epsilon \) and then plugging it back into \( Y_{\epsilon,1} \), we obtain 
\[ \hat{\tau}_\epsilon = -\frac{\alpha}{\lambda_e} (1 - p) \log \epsilon - \frac{1}{\lambda_e} \log |\eta_\epsilon^+|, \tag{43} \]
\[ Y_{\epsilon,1}(\hat{\tau}_\epsilon) = \epsilon^{\alpha p} \text{sgn}(\eta_\epsilon^+). \]

Using this information we are in a position to obtain the asymptotic behaviour of the random variables \( \eta_{\epsilon}^\pm \). First, from relation (36) we obtain 
\[ \eta_\epsilon^+ = \chi_{\epsilon,1} + e^{-\alpha} \int_0^{\hat{\tau}_\epsilon} e^{-\lambda_e s} H_1(Y_\epsilon(s), \epsilon) \, ds + \epsilon^{1-\alpha} N_\epsilon^+(\hat{\tau}_\epsilon). \tag{44} \]

Using (43) in (37) we obtain 
\[ \eta_\epsilon^- = |\eta_\epsilon^+|^{\lambda_e/\lambda_e} (Y_2 + e^{\alpha \chi_{\epsilon,1}} + |\eta_\epsilon^+|^{\lambda_e/\lambda_e} \int_0^{\hat{\tau}_\epsilon} e^{1-\delta} H_2(Y_\epsilon(s), \epsilon) \, ds + \epsilon^{1-\alpha(1-p)\lambda_e/\lambda_e} N_\epsilon^-(\hat{\tau}_\epsilon). \tag{45} \]

The main part of the proof is based on representations (43)–(45).

Lemma 13 allows us to use the estimates established in lemma 9 up to time \( \hat{\tau}_\epsilon \). In particular, now we can conclude that the family 
\[ \left( e^{-\alpha} \sup_{t \leq \hat{\tau}_\epsilon} |Y_{\epsilon,2}(t) - e^{\lambda_e s} Y_2| \right)_{\epsilon > 0} \]
is slowly growing thus improving lemma 10.

To obtain the desired convergence for \( \eta_\epsilon^+ \), we analyse the rhs of (44) term by term. The convergence of the first term was one of our assumptions. For the second one, we need to
estimate $H_1(Y_\epsilon, \epsilon)$. Using lemma 9, the boundedness of $Y_{\epsilon,2}$ and the definition of $t_\epsilon$, we see that
\[
\sup_{t \leq t_\epsilon} e^{-\lambda_2 t} Y_{\epsilon,2}^2 (t) = \mathcal{O}(e^{q_{\epsilon} t}).
\] (47)

This estimate and theorem 5 imply that
\[
e^{-\alpha} \int_0^{t_\epsilon} e^{-\lambda_2 s} H_1(Y_\epsilon(s), \epsilon) \, ds \leq K_\epsilon e^{-\alpha} \int_0^{t_\epsilon} e^{-\lambda_2 s} Y_{\epsilon,1}^2 (s) |Y_{\epsilon,2}(s)| \, ds + K_2 \epsilon^{2-\alpha}
\]
\[
= \mathcal{O}(\epsilon^{\alpha q} |\log \epsilon|).
\]

Let us estimate the third term in (44). We can use the last estimate along with (44) and lemma 12 to conclude that the distributions of positive part of $\lambda_\tau^{-1} \log |\eta_\tau^\epsilon|$ form a tight family. Therefore, (43) implies that
\[
\hat{t}_\epsilon \overset{P}{\to} \infty, \quad \epsilon \to 0.
\]

Combined with Itô isometry and lemma 12, this implies
\[
N_\epsilon^\tau (\hat{t}_\epsilon) \overset{L^2}{\to} N^\tau, \quad \epsilon \to 0,
\]
which completes the analysis of $\eta_\tau^\epsilon$ and, due to (43), of $t_\epsilon$.

To obtain the convergence of $\eta_\tau^\epsilon$, we study (45). Combining (46), the inequality
\[
|Y_{\epsilon,2} (t)| Y_{\epsilon,2}^2 (t) \leq 2 |Y_{\epsilon,1} (t)| (|Y_{\epsilon,2} (t) - e^{-\lambda_1 t} y_2|^2 + e^{-2 \lambda_1 t} y_2^2),
\]
and the definition of $\hat{t}_\epsilon$ we see that for any $q \in (0, \alpha p)$,
\[
\sup_{t \leq \hat{t}_\epsilon} e^{-\alpha t} |Y_{\epsilon,2} (t)| Y_{\epsilon,2}^2 (t) = \mathcal{O} \left( e^{q \alpha t} y_2 + 2 \right).
\]

Hence, as a consequence of theorem 5 and (43) we have
\[
\int_0^{\hat{t}_\epsilon} e^{-\lambda_2 s} H_2(Y_\epsilon(s), \epsilon) \, ds = \mathcal{O} \left( e^{-\alpha + 1 - \alpha} + 2 \right) \mathcal{O} \left( e^{q \alpha t} y_2 + 2 \right) \mathcal{O} \left( e^{-\alpha + 1 - \alpha} + 2 \right) \mathcal{O} \left( e^{q \alpha t} y_2 + 2 \right) \mathcal{O} \left( e^{-\alpha + 1 - \alpha} + 2 \right)
\]
\[
\quad = \mathcal{O} \left( e^{q \alpha t} y_2 + 2 \right) \mathcal{O} \left( e^{-\alpha + 1 - \alpha} + 2 \right) \mathcal{O} \left( e^{q \alpha t} y_2 + 2 \right) \mathcal{O} \left( e^{-\alpha + 1 - \alpha} + 2 \right)
\]

Combining this and lemma 8 in (45) obtain
\[
\eta_\epsilon^- = |\eta_\epsilon^\tau| + \mathcal{O} (e^{\alpha^2}) + \mathcal{O} \left( e^{q \alpha t} y_2 + 2 \right) \mathcal{O} \left( e^{-\alpha + 1 - \alpha} + 2 \right) \mathcal{O} \left( e^{q \alpha t} y_2 + 2 \right) \mathcal{O} \left( e^{-\alpha + 1 - \alpha} + 2 \right)
\]
which completes the proof of lemma 6 by choosing $q$ small enough.

\[
7. \text{Proof of lemma 7}
\]

Consider the solution to system (13)–(14) equipped with initial conditions (20) satisfying (21). Let us restrict the analysis to the arbitrary high probability event
\[
\{ |\eta_\epsilon^\tau| \leq K_\pm \},
\]
for some constants $K_\pm > 0$.

**Lemma 14.** Let $p \in (0, 1)$ satisfy (15), and let $(t_\epsilon)_{\epsilon > 0}$ be a slowly growing family of stopping times. Consider $t_\gamma = t_\epsilon \wedge t_\gamma^U$, then for any $\gamma > 0$,
\[
\lim_{\epsilon \to 0} P \left( \sup_{t \leq t_\epsilon} |Y_{\epsilon,2} (t)| \leq (K_\pm + \gamma) e^{\alpha(1-p) \lambda (\gamma)} \right) = 1.
\]
Proof. Let $\gamma > 0$. We recall that $N^c_\epsilon$ is defined in (38) and introduce the process

$$M_\epsilon(t) = N^c_\epsilon(t) + \epsilon \int_0^t e^{-\lambda_\ast(t-s)} \Psi_2(Y_\epsilon(s)) \, ds,$$  

(48)

where $\Psi_2$ was introduced in theorem 5, and the stopping time

$$\beta_\epsilon = \inf \{ t : |Y_{\epsilon,2}(t)| > (K_\ast + \gamma) e^{a(t-1)\lambda_\ast/\lambda_\ast} \}.$$  

Using the fact that $Y_{\epsilon,2}$ is bounded, it is easy to see that there is a constant $K_\ast$ independent of $t$, so that for any $t \leq \beta_\epsilon \land t'$, we have

$$\int_0^t e^{-\lambda_\ast(t-s)} |Y_{\epsilon,1}(s)| Y^2_{\epsilon,2}(s) \, ds \leq K_\ast e^{2a(t-1)\lambda_\ast/\lambda_\ast}.$$  

This estimate, along with Duhamel’s principle and theorem 5 implies that for some constant $C > 0$ and any $t \leq \beta_\epsilon \land t'$,

$$|Y_{\epsilon,2}(t)| \leq e^{a(t-1)\lambda_\ast/\lambda_\ast} |Y_{\epsilon,2}| + K_1 \int_0^t e^{-\lambda_\ast(t-s)} |Y_{\epsilon,1}(s)| Y^2_{\epsilon,2}(s) \, ds + \epsilon \sup_{t \leq \beta_\epsilon} |M_\epsilon(t)|$$

$$\leq e^{a(t-1)\lambda_\ast/\lambda_\ast} K_\ast + C e^{2a(t-1)\lambda_\ast/\lambda_\ast} + \epsilon \sup_{t \leq \beta_\epsilon} |M_\epsilon(t)|.$$  

Hence, using lemma 8 to estimate $M_\epsilon$, we obtain that

$$P[\beta_\epsilon < t'] = P \left( \sup_{t \leq \beta_\epsilon \land t'} |Y_{\epsilon,2}(t)| \geq (K_\ast + \gamma) e^{a(t-1)\lambda_\ast/\lambda_\ast} \right)$$

$$\leq P \left( C e^{a(t-1)\lambda_\ast/\lambda_\ast} + \epsilon e^{1-a(t-1)\lambda_\ast/\lambda_\ast} \sup_{t \leq \beta_\epsilon} |M_\epsilon(t)| \geq \gamma \right)$$

converges to 0 as $\epsilon \to 0$ proving the lemma. \hfill $\square$

Lemma 15. Under the assumptions of lemma 14, for any $\rho \in (0, \frac{\alpha_p}{\lambda_\ast}]$, $\gamma > 0$, and $C > 0$, define $\rho_\epsilon = (-\rho \log \epsilon + C) \land t'_\epsilon$. Then, we have

$$\lim_{\epsilon \to 0} P \left( \sup_{t \leq \rho_\epsilon} |Y_{\epsilon,1}(t)| e^{-\lambda_\ast t} \leq (1 + \gamma) e^{\alpha_p} \right) = 1.$$  

Proof. Define the stopping time

$$\beta_\epsilon = \inf \{ t : |Y_{\epsilon,1}(t)| e^{-\lambda_\ast t} \geq (1 + \gamma) e^{\alpha_p} \}.$$  

As a consequence of Duhamel’s principle and theorem 5 we obtain the bound

$$\sup_{t \leq \beta_\epsilon \land \rho_\epsilon} |Y_{\epsilon,1}(t)| e^{-\lambda_\ast t} \leq e^{\alpha_p} + K_1 \int_0^{\beta_\epsilon \land \rho_\epsilon} e^{-\lambda_\ast s} Y^2_{\epsilon,2}(s) |Y_{\epsilon,2}(s)| \, ds$$

$$+ e^2 K_\ast \lambda_\ast^{-1} + \epsilon S^\ast_\epsilon(\beta_\epsilon).$$

This estimate together with lemma 14, lemma 8 and the definition of $\rho_\epsilon$ implies that for any small $\delta > 0$ we can find a constant $K > 0$, so that with probability bigger than $1 - \delta$, the inequalities

$$\sup_{t \leq \beta_\epsilon \land \rho_\epsilon} |Y_{\epsilon,1}(t)| e^{-\lambda_\ast t} \leq e^{\alpha_p} + K e^{\alpha_p + a(t-1)\lambda_\ast/\lambda_\ast} (\beta_\epsilon \land \rho_\epsilon) + K \epsilon$$

$$\leq e^{\alpha_p} (1 + 2 K \rho e^{a(t-1)\lambda_\ast/\lambda_\ast} |\log \epsilon| + K e^{1-\alpha_p}),$$
Normal forms approach to diffusion near hyperbolic equilibria

hold for all $\epsilon > 0$ small enough. Hence, for any small enough $\epsilon > 0$,

$$P \{ \beta_\epsilon < \rho_\epsilon \} = P \left\{ \sup_{t \leq \beta_\epsilon \land \rho_\epsilon} |Y_{\epsilon,1}(t)| e^{-\lambda_{\beta_\epsilon}} \geq (1 + \gamma) e^{a_\rho} \right\} \leq P \left\{ \int_{Y_{\epsilon,1}(t)} e^{-\lambda_{\rho_\epsilon}} |\log \epsilon| + K \epsilon^{1-a_\rho} \geq \gamma \right\} + \delta,$$

which implies the result. □

The following is an important consequence of lemma 14:

**Corollary 16.** With $\tau_\epsilon$ as in (22) it holds that

$$\lim_{\epsilon \to 0} P[\tau_{U,\epsilon} < \tau_\epsilon] = 0.$$ 

In particular, (24) holds.

From now on, we restrict our analysis to the high probability event $\{\tau_{U,\epsilon} \geq \tau_\epsilon\}$. Let $\theta_\epsilon^* = \epsilon^{-a_\rho} e^{-\lambda_\epsilon} Y_{\epsilon,1}(\tau_\epsilon)$. Then, (22) implies

$$\tau_\epsilon = -\frac{a_\rho}{\lambda_\epsilon} \log \epsilon + \frac{1}{\lambda_\epsilon} \log \frac{\delta}{|\theta_\epsilon^*|} \tag{49}$$

and

$$Y_{\epsilon,1}(\tau_\epsilon) = \delta \text{sgn} \theta_\epsilon^*.$$

Our analysis of these expressions will be based on the next formula which directly follows from Duhamel’s principle:

$$\theta_\epsilon^* = \text{sgn} \eta_\epsilon^* + \epsilon^{-a_\rho} \int_{\tau_\epsilon}^{\tau_{\epsilon,\eta}} e^{-\lambda_{\eta,s}} H_1(Y_{\epsilon}(s), \epsilon) \, ds + \epsilon^{1-a_\rho} N_\epsilon^*(\tau_\epsilon) \tag{50}.$$

The main term in the rhs of (50) is $\text{sgn} \eta_\epsilon^*$. We need to estimate the other two terms. Lemma 8 implies that $\epsilon^{1-a_\rho} N_\epsilon^*(\tau_\epsilon)$ converges to 0 in probability as $\epsilon \to 0$. Let us now estimate the integral term. Relations (49) and (50) imply that $(\tau_\epsilon)_{\epsilon > 0}$ is slowly growing, and we can use lemma 14 to derive

$$\sup_{t \leq \tau_\epsilon} |Y_{\epsilon,2}(t)| = O_P(\epsilon^{a(1-p)\lambda_{\rho,\lambda}}). \tag{51}$$

We can now use theorem 5 to conclude that

$$\epsilon^{-a_\rho} \sup_{t \leq \tau_\epsilon} |H_1(Y_{\epsilon}(t), \epsilon)| = O_P(\epsilon^{a(1-p)\lambda_{\rho,\lambda}} + \epsilon^{2-a_\rho}),$$

and (15) implies that the rhs converges to 0. Therefore,

$$\epsilon^{-a_\rho} \int_{\tau_\epsilon}^{\tau_{\epsilon,\eta}} e^{-\lambda_{\eta,s}} H_1(Y_{\epsilon}(s), \epsilon) \, ds \xrightarrow{P} 0.$$

The above analysis of equation (50) implies that if we define $\theta_0^* = \text{sgn} \eta_0^*$, then

$$\theta_\epsilon^* \xrightarrow{\text{Law}} \theta_0^*, \tag{52}$$

which implies (25) due to (49). It remains to prove (26).

Duhamel’s principle along with (49) yields

$$Y_{\epsilon,2}(\tau_\epsilon) = \left(\frac{|\theta_0^*|}{\delta}\right)^{\lambda_{\rho,\lambda}} e^{a_{\lambda_{\rho,\lambda}} \eta_0^*} + \int_{0}^{\tau_\epsilon} e^{-\lambda_{\rho,\lambda}(s-\tau_\epsilon)} H_2(Y_{\epsilon}(s), \epsilon) \, ds + \epsilon N_{\epsilon}^*(\tau_\epsilon). \tag{53}$$

In order to study the convergence of $N_{\epsilon}^*(\tau_\epsilon)$ we first give a preliminary result.
Lemma 17.

\[ \sup_{t \leq \tau_e} |Y_{\epsilon,1}(t) - e^{a_p t} e^{\lambda t} \text{sgn } \eta_p^+| \xrightarrow{P} 0, \quad \epsilon \to 0. \]

**Proof.** The lemma follows from Duhamel’s principle and lemma 15. \( \square \)

The following result is essentially lemma 8.9 from [3]. It holds true in our setting since its proof is based only on the conclusion of lemma 17.

**Lemma 18.** As \( \epsilon \to 0 \),

\[ N_{\epsilon}^{-}(\tau_e) \xrightarrow{\text{Law}} N, \]

where \( N \) is the Gaussian random variable in (23).

We complete the proof of lemma 7. Recall that the process \( M_{\epsilon} \) was defined in (48) and introduce the stochastic processes

\[ R_{\epsilon}(t) = \int_0^t e^{\lambda(t-s)} \tilde{H}_2(Y_{\epsilon}(s)) \, ds. \]

(54)

Note that (53) and (49) imply

\[
Y_{\epsilon,2}(\tau_e) = e^{-\lambda \tau_e} Y_{\epsilon,2}(0) + \int_0^{\tau_e} e^{-\lambda(\tau_e-s)} H_2(Y_{\epsilon}(s), \epsilon) \, ds + \epsilon N_{\epsilon}^{-}(\tau_e) \\
= e^{-\lambda \tau_e} e^{a(1-p)\lambda/\lambda_{\epsilon}} \eta_{\epsilon}^{-} + \epsilon M_{\epsilon}(\tau_e) + R_{\epsilon}(\tau_e) \\
= \eta_{\epsilon}^{-}\left(\frac{[\theta_i^+]}{\delta}\right)^{\lambda/\lambda_{\epsilon}} e^{a(1-p)\lambda/\lambda_{\epsilon}} + \epsilon M_{\epsilon}(\tau_e) + R_{\epsilon}(\tau_e). \quad (55)
\]

Relations (21) and (52) imply

\[
\eta_{\epsilon}^{-}\left(\frac{[\theta_i^+]}{\delta}\right)^{\lambda/\lambda_{\epsilon}} \xrightarrow{\text{Law}} \left(\frac{[\eta_0^+]}{\delta}\right)^{\lambda/\lambda_{\epsilon}} \gamma_2. \quad (56)
\]

Lemma 18 and estimate (51) imply

\[ M_{\epsilon}(\tau_e) \xrightarrow{\text{Law}} N, \quad \epsilon \to 0. \quad (57)\]

Equations (56) and (57) describe the behaviour of first two terms in (55) and the proof of the lemma will be complete as soon as we show that

\[ e^{-\beta R_{\epsilon}(\tau_e)} \xrightarrow{P} 0, \quad \epsilon \to 0. \quad (58)\]

We can write the following rough estimate based on (51) and theorem 5:

\[
\sup_{t \leq \tau_e} |R_{\epsilon}(t)| = O_P(\epsilon^{a(1-p)\lambda/\lambda_{\epsilon}}). \quad (59)
\]

This is not sufficient for our purposes. We shall need a more detailed analysis instead. First, note that

\[
\sup_{t \leq \tau_e} |Y_{\epsilon,2}(t) - \epsilon M_{\epsilon}(t) - R_{\epsilon}(t)| e^{\lambda t} = e^{a(1-p)\lambda/\lambda_{\epsilon}} |\eta_{\epsilon}^{-}| = O_P(\epsilon^{a(1-p)\lambda/\lambda_{\epsilon}}). \]

Hence, for any \( \gamma > 0 \) there is a \( K_{\gamma} > 0 \) such that the event

\[
D_{\epsilon} = \left\{ \sup_{t \leq \tau_e} |Y_{\epsilon,2}(t) - \epsilon M_{\epsilon}(t) - R_{\epsilon}(t)| e^{\lambda t} < K_{\gamma} \epsilon^{a(1-p)\lambda/\lambda_{\epsilon}} \right\}
\]
has probability $P(D_2) > 1 - \gamma$ for $\epsilon > 0$ small enough. Moreover, using theorem 5 we see that for some constant $K_\beta > 0$,

$$|R_\epsilon(t)| \leq K_\beta \int_0^t e^{-\lambda_s (t-s)}Y_{\epsilon,2}(s)\,ds.$$  

Then, using the inequality $(a - b)^2 \leq 2a^2 + 2b^2$ and defining $K_\beta,\gamma = K_\beta K_\gamma$, we see that on $D_\epsilon$ for each $t \leq \tau_\epsilon$,

$$|R_\epsilon(t)| \leq K_\beta e^{-\lambda_t - t} \int_0^t (e^{\lambda_s} Y_{\epsilon,2}(s))^2 e^{-\lambda_s} \,ds$$

$$\leq 2K_\beta,\gamma e^{-\lambda_t} \int_0^t e^{-\lambda_s} e^{2\alpha_0 (1-p)\lambda_s/\lambda_t} \,ds + 2K_\beta \int_0^t e^{-\lambda_s (t-s)} |\epsilon M_\epsilon(s) + R_\epsilon(s)|^2 \,ds$$

$$\leq 2K_\beta,\gamma e^{-\lambda_t} - \lambda_t + 4K_\beta \lambda_t^2 M_{\epsilon,\infty}^2 + 4K_\beta e^{-\lambda_t} \int_0^t e^{\lambda_s} R_\epsilon(s) \,ds, \quad (60)$$

where $M_{\epsilon,\infty} = \sup_{t \leq \tau_\epsilon} |M_\epsilon(t)|$, so that (according to lemma 8) $M_{\epsilon,\infty}$ is slowly growing. Due to (59) we can find a constant $K'_\gamma > 0$ (independent of $\epsilon > 0$ and $t > 0$) so that the event

$$D'_\epsilon = D_\epsilon \cap \left\{ \sup_{t \leq \tau_\epsilon} |R_\epsilon(t)| \leq K'_\gamma e^{\alpha (1-p)\lambda_s/\lambda_t} \right\}$$

has probability $P(D'_\epsilon) > 1 - \gamma$ for all $\epsilon > 0$ small enough. Hence, multiplying both sides of (60) by $e^{\lambda_t - t}$, we see that for some constant $C_\gamma > 0$ and all $t \leq \tau_\epsilon$,

$$e^{\lambda_t - t} |R_\epsilon(t)| \mathbf{1}_{D'_\epsilon} \leq \alpha(t) + C_\gamma e^{\alpha (1-p)\lambda_s/\lambda_t} \int_0^t e^{\lambda_s - t} |R_\epsilon(s)| \mathbf{1}_{D'_\epsilon} \,ds,$$

where

$$\alpha(t) = C_\gamma e^{2\alpha_0 (1-p)\lambda_s/\lambda_t} + C_\gamma e^{2M_{\epsilon,\infty}^2/\lambda_t}.$$  

Using Gronwall’s lemma and (61) we obtain

$$\mathbf{1}_{D'_\epsilon} e^{\lambda_t - t} |R_\epsilon(t)| \leq \alpha(t) + C_\gamma e^{\alpha (1-p)\lambda_s/\lambda_t} \int_0^t \alpha(s) e^{e^{2\alpha_0 (1-p)\lambda_s/\lambda_t} (t-s)} \,ds$$

$$\leq \alpha(t) + C_\gamma e^{\alpha (1-p)\lambda_s/\lambda_t} e^{e^{2\alpha_0 (1-p)\lambda_s/\lambda_t} t}$$

$$+ \frac{C_\gamma}{\lambda_t} e^{2\alpha_0 (1-p)\lambda_s/\lambda_t} M_{\epsilon,\infty}^2 e^{e^{2\alpha_0 (1-p)\lambda_s/\lambda_t} t}.$$  

Hence,

$$\mathbf{1}_{D'_\epsilon} |R_\epsilon(t)| \leq C_\gamma e^{2\alpha_0 (1-p)\lambda_s/\lambda_t} e^{-\lambda_t - t} (1 + C_\gamma e^{\alpha (1-p)\lambda_s/\lambda_t} e^{e^{2\alpha_0 (1-p)\lambda_s/\lambda_t} t})$$

$$+ C_\gamma e^{2M_{\epsilon,\infty}^2 (1 + C_\gamma e^{\alpha (1-p)\lambda_s/\lambda_t} e^{e^{2\alpha_0 (1-p)\lambda_s/\lambda_t} t})).$$

Using (49), we obtain that for any $q > 0$,

$$\mathbf{1}_{D'_\epsilon} |R_\epsilon(\tau_\epsilon)| = O_P \left( e^{2\alpha_0 (1-p)\lambda_s/\lambda_t} e^{-\lambda_{\tau_\epsilon} - \tau_\epsilon} + e^{2M_{\epsilon,\infty}^2} \right)$$

$$= O_P \left( e^{\alpha (1-p)\lambda_s/\lambda_t + \alpha (1-p)\lambda_s/\lambda_t} + e^{2-q} \right),$$

so that (58) follows, and the proof is complete by choosing $q$ small enough.
8. Example

In this section we give an explicit example that illustrates not only the main result of this work but also the intermediate steps. We have chosen an example in which everything is explicitly computable, but rich enough to exhibit all the features considered in this paper. We will summarize all the development in theorem 19, which we will then use to give some observations. These observations apply to general two-dimensional systems.

Let \( D = [-\delta, \delta] \times [-\delta, \delta] \) for some \( \delta > 0 \). Consider \( X_\epsilon = (X_{\epsilon,1}, X_{\epsilon,2}) \), the solution to the SDE driven by the two-dimensional standard Brownian Motion \((W_1, W_2)\),

\[
\begin{align*}
\mathrm{d}X_{\epsilon,1}(t) &= \left( X_{\epsilon,1}(t) + \epsilon X_{\epsilon,1}(t) + X_{\epsilon,2}(t)^{j-1} \right) \mathrm{d}t + \epsilon \mathrm{d}W_1(t), \\
\mathrm{d}X_{\epsilon,2}(t) &= -\mu X_{\epsilon,2}(t) \mathrm{d}t + \epsilon (X_{\epsilon,1}(t) + 1) \mathrm{d}W_2(t),
\end{align*}
\]

with initial condition \((X_{\epsilon,1}(0), X_{\epsilon,2}(0)) = (0, x) + \epsilon^\alpha(1, 1)\). Here \( j \geq 2 \) is an integer, \( \alpha \in (0, 1] \), \( \mu > 0 \) and \( x \in (0, \delta) \). Let us make some observations about this equation:

- The equation is nonlinear with multiplicative noise and scaling initial condition.
- Due to (11) and (12), when \( \mu = 1 \), the nonlinear monomial in the rhs is resonant and the flow generated by the drift cannot be conjugated to a linear one.
- If \( \mu \neq 1 \), then there is a normal form conjugation that transforms this equation into a linear one.

The number \( \mu > 0 \) will be used as a control parameter to study the different types of scaling that the exit distribution of \( X_\epsilon \) from \( D \) has. In this case, \( \mu \) can be considered as the absolute value of ratio between the eigenvalues, which is an invariant of the system.

We study the exit problem of \( X_\epsilon \) from \( D \). In this case, the stable and unstable manifolds are given by

\[
\begin{align*}
\mathcal{W}^s &= \{(x, y) \in \mathbb{R}^2 : x = 0\} \\
\mathcal{W}^u &= \{(x, y) \in \mathbb{R}^2 : y = 0\},
\end{align*}
\]

respectively. Hence, theorem 1 establishes that

\[
X_\epsilon(\tau_D^\epsilon) = \delta (\text{sgn} \eta_\epsilon, 0) + \epsilon^{\alpha+1}(0, \tilde{\theta}_\epsilon),
\]

where both random families \((\eta_\epsilon)_{\epsilon>0}\) and \((\tilde{\theta}_\epsilon)_{\epsilon>0}\) are convergent in distribution to \( \eta_0 \) and \( \theta_0 \) respectively. Let us characterize some properties of \( \eta_0 \) and \( \theta_0 \).

In this case, we can write down the solution of \( X_\epsilon \) explicitly, we combine the explicit solution with the results already proven in this text to obtain a cleaner procedure. Let

\[
\hat{\tau}_\epsilon = \inf \{ t : |X_{\epsilon,1}(t)| = \delta \},
\]

and note that corollaries 11 and 16 imply that \( P\{\tau_D^\epsilon = \hat{\tau}_\epsilon\} \to 1 \), as \( \epsilon \to 0 \). Hence, we can do all the analysis on the event \( \{\tau_D^\epsilon = \hat{\tau}_\epsilon\} \).

In order to obtain asymptotics for the time \( \hat{\tau}_\epsilon \), we first give a few observations. Itô’s formula implies that

\[
X_{\epsilon,1}(t) = \epsilon^\alpha \exp \left\{ t + \int_0^t X_{\epsilon,1}(s)^{j-1} X_{\epsilon,2}(s)^{j-1} \mathrm{d}s \right\} \\
+ \epsilon \int_0^t \exp \left\{ (t-s) + \int_s^t X_{\epsilon,1}(u)^{j-1} X_{\epsilon,2}(u)^{j-1} \mathrm{d}u \right\} \mathrm{d}W_1(s).
\]
Hence, \( X_{e,1}(t) = e^{\alpha e^{\tau_D} \eta_e} \), where we have defined

\[
\eta_e = \exp \left\{ \int_0^{\tau_D} X_{e,1}(t)^{j-1} X_{e,2}(s)^{j-1} \, ds \right\} \tag{64}
\]

and \( X_{e,1}(\tau_D) = \delta \sgn \eta_e \). In order to study the properties of \( \eta_e \), note that the definition of \( \hat{\tau}_e \) together with lemmas 6, 9, 10 and 14 imply that

\[
\int_{\hat{\tau}_e}^{0} |X_{e,1}(t)^{j-1} X_{e,2}(s)^{j-1}| \, ds \xrightarrow{P} 0, \quad \epsilon \to 0.
\]

Also, this implies that

\[
\tau_e + \alpha \log \epsilon \xrightarrow{P} \log \frac{\delta}{|\eta_0|}, \quad \epsilon \to 0,
\]

as stated in theorem 1.

The idea is to substitute (66) in \( X_{e,2} \). We use Itô’s formula and (66) to obtain that \( \theta_e = X_{e,2}(\tau_D) \) is equal to

\[
\theta_e = (x + \epsilon \alpha) e^{-\mu \tau_D} + \epsilon \int_0^{\tau_D} e^{-\mu (t_D-s)} (X_{e,1}(s)+1) \, dW_2(s)
\]

\[
= x e^{\alpha \mu} \left( \frac{|\eta_0|}{\delta} \right) + \epsilon \int_0^{\tau_D} e^{-\mu (t_D-s)} (X_{e,1}(s)+1) \, dW_2(s) + O_P(\epsilon^{\alpha (\mu+1)}).
\]

Lemmas 10 and 18 imply that the stochastic integral converges in distribution to a random variable such that conditioned on \( \{ \sgn \eta_0 = \pm 1 \} \) is a zero mean Gaussian random variable \( N = \int_{0}^{\infty} e^{-s} \, dW_1(s) \).

Hence, taking leading exponents, we obtain that

\[
\epsilon^{-(\alpha \mu+1)} \theta_e \to 1_{\{\alpha \mu \leq 1\}} X \left( \frac{|\eta_0|}{\delta} \right)^{\mu} + 1_{\{\alpha \mu \geq 1\}} \mathcal{N},
\]

in distribution, as \( \epsilon \to 0 \).

Before making some comments about the possible asymmetries that this system has, we compile a theorem:

**Theorem 19.** Consider \( X_e \), the solution to the system (62) and (63) described in this section and the domain \( D = [-\delta, \delta] \times [-\delta, \delta] \). The exit time \( \tau_D^e \) and exit distribution \( X_e(\tau_D^e) \) are given by

\[
X_e(\tau_D^e) = \delta (\sgn \eta_e, 0) + e^{\alpha \mu+1}(0, \theta'_e)
\]
and
\[ \tau^D_\epsilon = -\alpha \log \epsilon + \log \frac{\delta}{|\eta_\epsilon|}. \]

Here \((\eta_\epsilon)_{\epsilon>0}\) and \((\theta'_\epsilon)_{\epsilon>0}\) are families of random variables that converge in distribution to \(\eta_0 = 1 + \mathbf{1}_{\{\alpha=1\}} N\)

and
\[ \theta'_0 = \mathbf{1}_{\{\mu \leq 1\}} \epsilon \left( \frac{|\eta_0|}{\delta} \right) + |\eta_0| \mathbf{1}_{\{\mu > 1\}} N, \]

respectively, where \(N\) is a zero mean Gaussian random variable with variance \(1/2\) and \(N\) is, conditioned on \{\text{sgn} \(\eta_0\) = \pm 1\}, a zero mean Gaussian random variable with variance
\[ \frac{\delta^2}{2(\mu + 1)} \pm \frac{2\delta}{2\mu + 1} + \frac{1}{2\mu}. \]

Some comments are in order:

1. When \(\alpha < 1\) the perturbation in the initial condition (in this case a constant) specifies the sign of \(\eta_0\) and hence, the exit point.
2. When \(\alpha \mu \leq 1\) the limiting random variable \(\theta'_0\) is not centred at 0 creating an asymmetry.
   Two cases are in order:
   a. When \(\alpha \mu < 1\), \(\theta'_0\) has the same sign as \(x\). This case is known as strongly asymmetric.
   b. When \(\alpha \mu = 1\), the mean of \(\theta'_0\) has the same sign as \(x\). This case is known as asymmetric.
3. When \(\alpha \mu > 1\), the random variable \(\theta'_0\) is a zero mean Gaussian.

9. Extensions

In this work we have studied the exit problem for small noise diffusions. In particular, we have shown the existence of possible asymmetries in the case in which the flow generated by the drift admits a saddle point. The proof is restricted to the two-dimensional setting. Let us discuss this restriction.

Our method of proof was to transform the original equation into a very specific nonlinear equation using normal forms. Then, we obtained several estimates that intensively use the smallness of the noise and the specific form of the nonlinearity in the normal form. Let us recall the form of the nonlinearity. In our case, the nonlinearity in the normal form is given by a finite sum of resonant monomials (see section 4) of the form \((x^{a_{1}^+}, x^{a_{1}^-}, x^{a_{d}^+}, x^{a_{d}^-})\), where \((a_{i}^\pm, a_{j}^\pm) \in \mathbb{Z}^2\) satisfy the resonance relations
\[ a_{i}^+ \lambda_+ - a_{j}^\pm \lambda_- = \pm \lambda_{\pm}, \]

of some order \(r = a_{i}^+ + a_{j}^\pm \geq 2\). If we were to generalize the argument presented in this text to the \(d\)-dimensional case, we would need to take into account the particular form that the nonlinearity would have in the normal form. Indeed, there are two points to consider:

1. The resonant monomials of order \(r \geq 2\) are of the form
\[ (x_{i_{1}}^{a_{i_{1}}^+} \cdots x_{i_{d}}^{a_{i_{d}}^+}, \cdots, x_{i_{1}}^{a_{i_{1}}^-} \cdots x_{i_{d}}^{a_{i_{d}}^-}), \]

where the vector \(\alpha_i = (a_{i_{1}}, \ldots, a_{i_{d}}) \in \mathbb{Z}^d\) satisfies
\[ a_{i_{1}} \lambda_{1} + \ldots + a_{i_{d}} \lambda_{d} = \lambda_{i}, \]
\[ a_{i_{1}} + \ldots + a_{i_{d}} = r, \]

for each \(i = 1, \ldots, d\). Here \(\lambda_{1}, \ldots, \lambda_{d}\) are the eigenvalues of the matrix \(Db(0)\).
According to [7, theorem 3, section 2], the nonlinearity $N$, after being transformed by a normal form transformation of degree $R > 1$, will be of the form

$$N(x) = P(x) + Q(x),$$

where $P$ is a finite sum of resonant monomials, and $Q$ is a correction of order $|x|^{R+1}$ (as $|x| \to 0$) when the vector of eigenvalues $\lambda = (\lambda_1, ..., \lambda_d)$ is not one-resonant and identically 0 when $\lambda$ is one-resonant.

The first point implies that to obtain the exponents $\alpha_{i,j}$ more combinatorial work than the one put in section 4 is needed. Still this is not the biggest difficulty. The biggest difficulty is the lack of structure of the correction $Q$ in the case $\lambda$ is not one-resonant. Indeed, to have our techniques applicable, we require at least that whenever $\alpha_{i,k} \neq 0$ for some $k < i$, then $\alpha_{i,j} \neq 0$ for some $k < j \leq d$ (we order the eigenvalues as usual: $\Re \lambda_1 \geq ... \geq \Re \lambda_d$). There is no guarantee that a condition of this form holds in the not one-resonant case. In conclusion, a higher dimensional analogue for the saddle case can be obtained using the techniques presented in this text in the one-resonant case, but further research has to be done in the not one-resonant case.

Acknowledgments

The authors are grateful to Vadim Kaloshin for his advice on normal forms and for pointing to [7]. The work of Yuri Bakhtin is supported by NSF through a CAREER grant DMS-0742424.

References

[1] Bakhtin Y 2008 Exit asymptotics for small diffusion about an unstable equilibrium *Stochastic Process. Appl.* **118** 839–51
[2] Bakhtin Y 2010 Small noise limit for diffusions near heteroclinic networks *Dyn. Syst.* **25** 413–31
[3] Bakhtin Y Noisy heteroclinic networks *Probability Theory and Related Fields* at press also available at http://arxiv.org/abs/0712.3952
[4] Chow S-N, Li C Z and Wang D 1994 *Normal Forms and Bifurcation of Planar Vector Fields* (Cambridge: Cambridge University Press)
[5] Day M V 1995 On the exit law from saddle points *Stochastic Process. Appl.* **60** 287–311
[6] Freidlin M I and Wentzell A D 1998 *Random Perturbations of Dynamical Systems* (Grundlehren der Mathematischen Wissenschaften vol 260) (Fundamental Principles of Mathematical Sciences) 2nd edn (New York: Springer) (Translated from the 1979 Russian original by Joseph Szücs)
[7] Il’yashenko Yu S and Yakovenko S Yu 1991 Finitely smooth normal forms of local families of diffeomorphisms and vector fields *Uspekhi Mat. Nauk* **46** 3–39, 240
[8] Karatzas I and Shreve S E 1988 *Brownian Motion and Stochastic Calculus* (Graduate Texts in Mathematics vol 113) (New York: Springer)
[9] Kifer Y 1981 The exit problem for small random perturbations of dynamical systems with a hyperbolic fixed point *Israel J. Math.* **40** 74–96
[10] Monter S A A and Bakhtin Y 2011 Scaling limit for the diffusion exit problem in the levinson case *Stochastic Process. Appl.* **121** 24–37
[11] Perko L 2001 *Differential Equations and Dynamical Systems* (Texts in Applied Mathematics vol 7) 3rd edn (New York: Springer)
[12] Rabinovich M I, Huerta R, Varona P and Afraimovich V S 2008 Transient cognitive dynamics, metastability, and decision making *PLoS Comput. Biol.* **4** e1000072