ABSTRACT

Title of dissertation:	ASYMPTOTIC PROBLEMS FOR STOCHASTIC PROCESSES AND CORRESPONDING PARTIAL DIFFERENTIAL EQUATIONS
	Lucas Tcheuko, Doctor of Philosophy, 2014
Dissertation directed by:	Professor Mark Freidlin & Professor Leonid Koralov Department of Mathematics

We consider asymptotic problems for diffusion processes that rely on large deviations.

In Chapter 2, we study the long time behavior (at times of order $\exp(\lambda/\varepsilon^2)$) of solutions to quasi-linear parabolic equations with a small parameter ε^2 at the diffusion term. The solution to a partial differential equation (PDE) can be expressed in terms of diffusion processes, whose coefficients, in turn, depend on the unknown solution. The notion of a hierarchy of cycles for diffusion processes was introduced by Freidlin and Wentzell and applied to the study of the corresponding linear equations. In the quasi-linear case, it is not a single hierarchy that corresponds to an equation, but rather a family of hierarchies that depend on the time scale λ. We describe the evolution of the hierarchies with respect to λ in order to gain information on the limiting behavior of the solution of the PDE.

In chapter 3, we study the asymptotic behavior of diffusion processes, with

a small diffusion term. This process is constrained to move within some bounded domain D with instantaneous reflection on hitting the boundary ∂D of D. Such processes have applications in asymptotic questions related to linear parabolic PDEs with the Neumann boundary condition. Similar problems were previously studied in [1] by Anderson and Orey. We expand on Anderson's and Orey's work by considering different equilbra in the interior of D, similarly to the problem studied in chapter 2. However, some equilibra also appear on the boundary ∂D. We use the results of Anderson and Orey together with the work of Freidlin and Wentzell to investigate the invariant measure of the process and describe the transitions of the process between the attractors. The knowledge of the invariant measure of the process and the transition rates (in the logarithmic scale) allow us to study the long term behavior of the solution to the corresponding linear parabolic PDE as the diffusion parameter goes to zero.

ASYMPTOTIC PROBLEMS FOR STOCHASTIC PROCESSES AND CORRESPONDING PARTIAL DIFFERENTIAL EQUATIONS

by

Lucas Tcheuko

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2014

Advisory Committee:
Professor. Mark Freidlin, Co-chair
Professor Leonid Koralov, Co-chair
Professor Sandra Cerrai
Professor Dimitry Dolgopyat
Professor Alexander Barg

Acknowledgments

Foremost, I would like to express my sincere gratitude to both of my my advisors Prof. Mark Freidlin and Prof. Leonid Koralov for the continuous support of my Ph.D study and research, for their patience, motivation and immense knowledge. Their guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D study.

I would like to thank Haydeé Hidalgo, Professors Paul smith and Abram Kagan for their support. My admission to the graduate program was only made possible thanks to the relentless help of Haydeé and Prof. Smith; Prof. Abram Kagan was very helpful in making sure that I secure the teacher assistantship.

I wish to thank the members of my committee, Prof. Sandra Cerrai, Prof. Dmitry Dolgopyat, and Pof. Alexander Barg.

I would also like to thank my colleague Wenqing Hu for his friendship, his contribution on our research project, and for invaluable discussions on various topics of probability theory.

Last but not the least I would like to thank my friends and family for all their support.

Table of Contents

Chapter 1: Introduction

There is a close relationship between the theory of second-order partial differential equations and Markov processes with continuous trajectories. One of the earlier formulations of this relationship can be found in the work of Kolmogorov [2] where the parabolic equations for transition probability were written down. It was established that the mean value of some functionals of the trajectories of diffusion processes are the solutions of the boundary value problems for the corresponding differential equations. It was possible to draw conclusions on Markov processes based on the behavior of the solutions of differential equations. These equations also appear in the physics literature [3]. Later progress on direct probabilistic methods for examining Markov processes made it possible to construct and study them independently of the differential equations. The construction and analysis of the trajectories of the corresponding diffusion processes via direct probabilistic methods, enable the construction of the solutions to differential equations and the study of properties of those solutions.

1.1 Stochastic Differential Equations

We consider the dynamical system

$$\dot{x}_t = b(x_t), \qquad x_0 = x \in \mathbb{R}^d. \tag{1.1}$$

Its stochastic perturbation is defined as

$$dX_t^{x,\varepsilon} = b(X_t^\varepsilon)dt + \varepsilon\sigma(X_t^{x,\varepsilon})dW_t; \quad X_0^{x,\varepsilon} = x \in \mathbb{R}^d, \tag{1.2}$$

where $\varepsilon > 0$ is a small parameter, $\sigma(x) = (\sigma_j^i(x)), i = 1, \ldots, d; \ j = 1, \ldots, \ell; \ W_t$ is a $\ell-$dimensional Wiener process, i.e. , its component are independent and satisfy the following properties:

1. W_t^i are continuous functions of t almost everywhere.

2. $W_0^i = 0$ almost everywhere.

3. $0 < t_1 < t_2 \implies W_{t_2}^i - W_{t_1}^i$ is a Gaussian random variable with zero mean and variance $t_2 - t_1$.

4. Random variables $W_{t_0}^i, W_{t_1}^i - W_{t_0}^i, \ldots, W_{t_k}^i - W_{t_{k-1}}^i$ are independent for every $k \geq 1$ and $0 = t_0 \leq t_1, \ldots, \leq t_k$.

We also assume that $|\sigma_{ij}(x) - \sigma_{ij}(y)| \leq K|x - y|$, and $|b^i(x) - b^i(y)| \leq K|x - y|$. Under these assumptions on b and σ, the solution of (1.2) is defined, and if ε is small, then on any final time interval T, the trajectories of the process $X_t^{x,\varepsilon}$ defined in (1.2) are close to the corresponding non-perturbed trajectory defined in (1.1) with probability close to 1. It is proved in [4] that for all $T > 0$ and for all $\delta > 0$ we have

$$\lim_{\varepsilon \to 0} P\left\{ \sup_{0 \leq \ t \leq T} |X_t^{x,\varepsilon} - x_t| > \delta \right\} = 0. \tag{1.3}$$

2

It follows from (1.3) that if a subset A of the space $\mathbf{C}[0,T]$ of continuous functions on the interval from $[0,T]$ to $\mathbb{R}^d$ contains a $\delta-$neighborhood of x_t in this space, then the main contribution to the probability $P\{X^{x,\varepsilon}_\cdot \in A\}$ is given by this $\delta-$neighborhood; the remaining part of the event $\{X^{x,\varepsilon}_\cdot \in A\}$ has small probability. It is possible to show that this probability is exponentially small as $\varepsilon \downarrow 0$. As will be discussed below, under certain assumptions on A, there exists a function $\varphi \in A$ such that the principal part of the probability measure of A is concentrated near φ. In particular, we just saw that if A contains the trajectory of the dynamical system $x_\cdot$ then $\varphi = x_t, t \in [0,T]$.

1.2 Action Functional

Let (a_{ij}) be the matrix whose elements satisfy $a_{ij}(x) = \sum_{k=1}^{\ell} \sigma^i_k(x)\sigma^j_k(x)$ and (a^{ij}) be the the inverse of the matrix (a_{ij}) The action functional of the process $X^{x,\varepsilon}_t$ is defined as

$$S_{0T}(\varphi) = \frac{1}{2}\int_0^T \sum_{i,j=1}^d a^{ij}(\varphi_t)(\dot{\varphi}^i_t - b_i(\varphi_t))(\dot{\varphi}^j_t - b_j(\varphi_t))dt, \quad T \geq 0, \ \varphi \in C([0,T], \mathbb{R}^d)$$

for absolutely continuous φ and $S_{0T}(\varphi) = \infty$ for φ that are not absolutely continuous.

The action functional of φ helps one to estimate the probability of the process $X^{x,\varepsilon}_t$ being in the neighborhood of φ.

Define $\rho_{0T}(\varphi, \psi) = \sup_{0 \leq t \leq T}(\varphi_t, \psi_t); \ \varphi \in C([0,T]), \psi \in C([0,T])$.

It is proved in [4] that for a compact set K, and any δ and $\gamma > 0$ there exists

3

a positive ε_0 such that for all $\varepsilon < \varepsilon_0$

$$P\{\rho_{0T}(X^{x,\varepsilon}, \varphi) < \delta\} \geq \exp\{-\varepsilon^{-2}[S_{0T}(\varphi) + \gamma]\} \qquad (1.4)$$

$\varphi \in C([0,T], \varphi_0 = x$, provided that $x \in K$.

It is also proved in [4] that for $s_0 > 0$

$$P\{\rho_{0T}(X^{x,\varepsilon}, \Phi(s)) \geq \delta\} \leq \exp\{-\varepsilon^{-2}[s - \gamma]\}, \qquad (1.5)$$

where $\Phi(s) = \{\varphi : \varphi_0 = x, S_{0T}(\varphi) \leq s\}$, $s \leq s_0$, and

$$\rho_{0T}(\varphi, \Phi(s)) = \inf_{\phi \in \Phi(s)} \rho_{0T}(\varphi, \psi), \ x \in K.$$

The inequality (1.4) gives a lower estimate of the probability that the process $X_t^{x,\varepsilon}$ passes through a $\delta-$ball about φ; while the inequality (1.5) gives the upper estimate of the probability that a trajectory of the process $X_t^{x,\varepsilon}$ moves far from the set of functions with small action functional.

Assume now that the vector field b has r asymptotically stable equilibrium points $O_1, O_2, \ldots, O_r$; that is, for every neighborhood $\mathcal{V}_i$ of O_i, $i \in 1 \ldots r$, there exists a smaller neighborhood $\mathcal{U}_i \subseteq \mathcal{V}_i$ such that the trajectories of the dynamical system (1.1) starting in $\mathcal{U}_i$ converge to O_i without leaving $\mathcal{V}_i$. For $x \in D_i$, the basin of attraction of O_i, the trajectories of the non-perturbed system (1.1) that start at x enter $\mathcal{U}_i$ in finite time. It follows from (1.3) that the solution $X_t^{x,\varepsilon}$ of (1.2) is close to the corresponding non-perturbed system with probability close to 1. Therefore, with high probability $X_t^{x,\varepsilon}$ enters a $\mathcal{U}_i$ before leaving D_i. It is also clear that for any fixed time the probability that the process $X_t^{x,\varepsilon}$ that starts in D_i stays in D_i with

probability close to 1. Let $\tau_i^\varepsilon = \min\{t : X_t^{x,\varepsilon} \in \partial D_i\}$, for $x \in D_i$ be the first time when $X_t^{x,\varepsilon}$ reaches the boundary of D_i. This time is of particular interest as it helps study the transition of the process $X_t^{x,\varepsilon}$ between different basins of attraction. By studying these different transitions, we can determine the long time behavior of the process $X_t^{x,\varepsilon}$ as ε goes to 0.

The quasi-potential is defined as

$$V(x,y) = \inf_{T,\varphi}\{S_{0,T}(\varphi) : \varphi \in C([0,T], \mathbb{R}^d), \varphi(0) = x, \varphi(T) = y\}, \quad x, y \in \mathbb{R}^d.$$

As the action functional of a function φ helps one to estimate the probability of the process $X_t^{x,\varepsilon}$ being in some some tube around φ, the quasi-potential will be helpful in estimating the time it takes to go from x to a small neighborhood of y. In particular, if x and y are different equilibrium points, the transition from a small neighborhood of $x = O_i$ to a small neighborhood of $y = O_j$ is explained in details in chapter 3 and is quite complex. If $x \in D_i$ and $y = O_i$ then $V(x,y) = 0$. It is proved in [4] that $\varepsilon^2 \ln \tau_i^\varepsilon$ converges in probability as $\varepsilon \downarrow 0$ to $V_0^i = \min_{y \in \partial D_i} V(O_i, y)$

1.3 Linear equations

The statements made at the end of the previous section imply various results for PDE's with a small parameter at the second order derivatives.

There is a connection between the stochastic differential equation (1.2) and the differential operator L^ε defined as

$L^\varepsilon = \frac{\varepsilon^2}{2} \sum_{i,j=1}^d a_{ij}(x)\frac{\partial^2}{\partial x_i \partial x_j} + b(x) \cdot \nabla_x, \quad x \in \mathbb{R}^d,$

where $a_{ij}(x)$ as in defined in section (1.2) and $b(x)$ is the same as in equation (1.1)

and (1.2). If a function u^ε belongs to the domain of the operator L^ε and $X_t^{x,\varepsilon}$ is the same as in the stochastic differential equations (1.2), then applying the Itô's formula to $u^\varepsilon(X_t^{x,\varepsilon})$ yields

$$u^\varepsilon(X_t^{x,\varepsilon}) - u^\varepsilon(x) = \int_0^t (\nabla u^\varepsilon(X_s^{x,\varepsilon}), \sigma(X_s^{x,\varepsilon})dW_s) + \int_0^t L^\varepsilon u^\varepsilon(X_s^{x,\varepsilon})ds \qquad (1.6)$$

where ∇u^ε is the gradient function of u^ε, and $(\nabla u^\varepsilon, \sigma dW)$ is the scalar product. Using properties of the stochastic integral and the assumptions on b and σ, it is proved that

$$\lim_{t\downarrow 0} \frac{1}{t}(Eu^\varepsilon(X_t^{x,\varepsilon}) - u^\varepsilon(x)) = \lim_{t\downarrow 0} \frac{1}{t}\int_0^t L^\varepsilon u^\varepsilon(X_s^{x,\varepsilon})ds = L^\varepsilon u^\varepsilon(x). \qquad (1.7)$$

Equation (1.7) is equivalent to the Feynman Kac formula which states that $Eg(X_t^{x,\varepsilon})$ is a solution to the following Cauchy problem:

$$\frac{\partial u^\varepsilon(t,x)}{\partial t} = \frac{\varepsilon^2}{2}\sum_{i,j=1}^d a_{ij}(x)\frac{\partial^2 u^\varepsilon(t,x)}{\partial x_i \partial x_j} + b(x)\cdot\nabla_x u^\varepsilon(t,x), \quad x\in\mathbb{R}^d, \ t>0, \qquad (1.8)$$

$$u^\varepsilon(0,x) = g(x), \ x\in\mathbb{R}^d. \qquad (1.9)$$

The main focus of this work is the study of (a) more general quasi-linear-equastions, and (b) the Neumann problem for parabolic equations. First, however, let us briefly discuss the long time behavior of the solution to the Cauchy problem (1.8)-(1.9), while b is assumed as in section (1.2) to have several equilibria. Let the function $t : \mathbb{R}^+ \to \mathbb{R}$ be such that $\ln(t(\varepsilon)) \sim \lambda/\varepsilon^2$ as $\varepsilon \downarrow 0$ with $\lambda > 0$. It is proved in [4] that $\lim_{\varepsilon\downarrow 0} X_{t(\varepsilon)}^{x,\varepsilon} = O_i$, for $\lambda < V_0^i$, and $x \in D_i$,

6

where V_0^i is defined as in section (1.2). It follows from the Feynman Kac formula and the continuity of g that

$$\lim_{\varepsilon \downarrow 0} u^\varepsilon(t(\varepsilon), x) = g(O_i), \text{ for } x \in D_i \text{ and } \lambda < V_0^i. \qquad (1.10)$$

Define $V_{ij} = V(O_i, O_j)$; if b has exactly two equilibria O_i and O_j, then the difficulty of passage from D_i to D_j is explained solely by V_{ij}. If we also assume without loss of generality that $V_{ij} < V_{ji}$ then for $V_{ij} < \lambda < V_{ji}$ The process $X_t^{x,\varepsilon}$ that starts in $x \in D_i$ has enough time to transition from D_i to D_j, but for $x \in D_j$, there is not sufficient time for the process to transition to D_i. It follows that $\lim_{\varepsilon \downarrow} u^\varepsilon\left(x, \exp(V_{ij}/\varepsilon^2)\right) = g(O_j)$. for those values of λ and $x \in D_i \cup D_j$. This limit does not change for $\lambda > V_{ji}$. However, it is important to notice that at those time scales the process $X_t^{x,\varepsilon}$ will make make many transitions between D_i and D_j, and will be found in D_j with probability close to 1. If b has more than two equilibria, then the knowledge of V_{ij} alone is not enough to determine the difficulty of transition from D_i to D_j. For a general case of multiple equilibria, the time scale at which the transition from D_i to D_j occurs is explained more precisely in chapter 3. The knowledge of different time scales at which the transitions occur between different equilibria allows one to describe the asymptotic of u^ε.

1.3.1 Cauchy problem with boundary Neumann condition

Let $D \subset \mathbb{R}^d$ be a bounded domain with smooth boundary ∂D. We consider the Cauchy problem with the Neumann boundary condition and a small parameter

$\varepsilon > 0$:

$$\begin{cases} \dfrac{\partial u^\varepsilon}{\partial t} = \mathcal{L}^\varepsilon u^\varepsilon \equiv \dfrac{\varepsilon^2}{2} \sum_{i,j=1}^{d} a_{ij}(x) \dfrac{\partial^2 u^\varepsilon}{\partial x_i \partial x_j} + \sum_{i=1}^{d} b^i(x) \dfrac{\partial u^\varepsilon}{\partial x_i} \ , \\[4mm] u^\varepsilon(0,x) = g(x) \ , & x \in D \cup \partial D \ ; \\[4mm] \dfrac{\partial u^\varepsilon}{\partial \gamma}(t,x) = 0 \ , & x \in \partial D \ , \ t \geq 0 \ . \end{cases} \tag{1.11}$$

where $\gamma(x)$ the conormal, is the vector such that $< \gamma(x), T(x) >_{a^{-1}} = 0$ for $T(x)$ tangential to the boundary. The scalar product with respect to the matrix $a^{-1} = a^{ij}$ is defined as $< u,v >_{a^{-1}} = \displaystyle\sum_{i,j=1}^{d} a^{ij}(x) u(x) v(x)$. It is proved that to the Cauchy problem with Neumann boundary condition (1.11) there corresponds the stochastic process

$$dX_t^{x,\varepsilon} = b(X_t^{x,\varepsilon}) dt + \varepsilon \sigma(X_t^{x,\varepsilon}) dW_t + \mathbf{1}_{\partial D}(X_t^{x,\varepsilon}) \gamma(X_t^{x,\varepsilon}) d\xi_t^{x,\varepsilon} \ , \ X_0^{x,\varepsilon} = x \ , \ \xi_0^{x,\varepsilon} = 0,$$

$$\tag{1.12}$$

where $\xi_t^{x,\varepsilon}$, the local time at the boundary, is a continuous process which only grows when $X_t^{x,\varepsilon}$ is on the boundary. The long time behavior of the solution to problem (1.11), will be considered in chapter 3 and will be shown to similar to the long time behavior of problem (1.8) and (1.9).

1.3.2 Quasi-linear equations

In chapter 2 we consider the quasi-linear problem

$$\frac{\partial u^\varepsilon(t,x)}{\partial t} = \frac{\varepsilon^2}{2} \sum_{i,j=1}^{d} a_{ij}(x, u^\varepsilon) \frac{\partial^2 u^\varepsilon(t,x)}{\partial x_i \partial x_j} + b(x) \cdot \nabla_x u^\varepsilon(t,x), \quad x \in \mathbb{R}^d, \quad t > 0, \tag{1.13}$$

$$u^\varepsilon(0,x) = g(x), \ x \in \mathbb{R}^d. \tag{1.14}$$

We still assume that b has multiple equilibria. This problem is equivalent to the system

$$dX_s^{t,x,\varepsilon} = b(X_s^{t,x,\varepsilon})ds + \varepsilon\sigma(X_s^{t,x,\varepsilon}, u^\varepsilon(t-s, X_s^{t,x,\varepsilon}))dW_s, \quad s \leq t, X_0^{t,x,\varepsilon} = x ,$$

$$u^\varepsilon(t,x) = \mathbb{E}g(X_t^{t,x,\varepsilon}),$$

$$(1.15)$$

with unknown $X_\cdot^{t,x,\varepsilon}$ and u^ε. The process $X_s^{t,x,\varepsilon}$ can be viewed as a nonlinear stochastic perturbation of the dynamical system (1.1). Though the process $X_s^{t,x,\varepsilon}$ is defined implicitly because of the dependence of its diffusion coefficient σ on the unknown PDE u^ε, the inequality (1.3) still holds as long as b and σ satisfy the Lipschitz condition. This implies that for $x \in D_i$, $X_s^{t,x,\varepsilon}$ is close to O_i if ε is small and is large (but not exponentially large as $\varepsilon \downarrow 0$). This could also be seen by looking at the action functional, which is still zero along the trajectories of the unperturbed system. The process $X_s^{t,x,\varepsilon}$ stays around a very small neighborhood of O_i for a very long time before eventually leaving as a result of large deviations. Intuitively, during this time, the process is close to the following one

$$dX_s^{t,x,\varepsilon} = b(X_s^{t,x,\varepsilon})ds + \varepsilon\sigma(X_s^{t,x,\varepsilon}, g(O_i))dW_s, \ s \leq t, X_0^{t,x,\varepsilon} = x, x \in D_i . \quad (1.16)$$

Thus it can be defined autonomously of u^ε, while $u^\varepsilon(t,x) = \mathbb{E}g(X_t^{t,x,\varepsilon})$.

Setting $\alpha_i = g(O_i)$, the action functional corresponding to this process is defined as

$$S_{0,T}^{\alpha_i}(\varphi) = \frac{1}{2}\int_0^T \sum_{i,j=1}^d a^{ij}(\varphi_t, \alpha_i)(\dot{\varphi}_t^i - b_i(\varphi_t))(\dot{\varphi}_t^j - b_j(\varphi_t))dt, \ \ T \geq 0, \ \varphi \in C([0,T], D_i)$$

9

for absolutely continuous φ defined on $[0, T]$ with values in D_i, and $\varphi_0 = x$. $S_{0,T}^{\alpha_i}(\varphi) = \infty$ if φ is not absolutely continuous or if $\varphi_0 \neq x$. We define

$$V^{\alpha_i}(x, y) = \inf_{T, \varphi}\{S_{0,T}^{\alpha_i}(\varphi) : \varphi \in C([0, T], D_i), \varphi(0) = x, \varphi(T) = y\}, \quad x, y \in D_i,$$

and $V_0^{\alpha_i} = \min_{y \in \partial D_i} V^{\alpha_i}(O_i, y)$. Similarly to (1.10), we obtain

$$\lim_{\varepsilon \downarrow 0} u^\varepsilon(t(\varepsilon), x) = g(O_i), \quad \text{for } x \in D_i, \text{ and } \lambda < V_0^{\alpha_i}.$$

Thus $u^\varepsilon(t(\varepsilon), x)$ is nearly constant as a function of $x \in D_i$ if $\lambda < V_0^{\alpha_i}$. It turns out that the constant does not depend on the time scale λ either, provided that $\lambda < V_0^{\alpha_i}$. We will show that, the solution to the system (1.13)-(1.14) is nearly constant on D_i for every time scale $t(\varepsilon) \sim \exp(\lambda/\varepsilon^2)$ although the constant will depend on λ, in general. In chapter 2, we provide in the description of the evolution of the solution of the system (1.13)- (1.14) together with the corresponding process at time scales of order $\lambda > V_0^{\alpha_i}$.

Chapter 2: Quasi-linear equations with a small diffusion term and the evolution of hierarchies of cycles

2.1 Introduction

Consider the Cauchy problem for the quasi-linear equation with a small parameter at the second order term

$$\frac{\partial u^\varepsilon(t,x)}{\partial t} = \frac{\varepsilon^2}{2} \sum_{i,j=1}^d a_{ij}(x,u^\varepsilon)\frac{\partial^2 u^\varepsilon(t,x)}{\partial x_i \partial x_j} + b(x) \cdot \nabla_x u^\varepsilon(t,x), \quad x \in \mathbb{R}^d, \quad t > 0, \quad (2.1)$$

$$u^\varepsilon(0,x) = g(x), \ x \in \mathbb{R}^d. \tag{2.2}$$

Equations with diffusion coefficients that depend on particle concentration u^ε arise naturally in many applications, in particular in population genetics. The situation when the drift b depends on both x and u^ε, with certain additional assumptions, can also be considered, but we assume here that b depends only on x for the sake of simplicity.

We assume that the coefficients of equation (2.1) are Lipschitz continuous and bounded; the matrix $(a_{ij}(x,u))$ is assumed to be uniformly positive definite. Under these conditions, problem (2.1)-(2.2) has a unique solution for any continuous bounded $g(x)$ (see, for instance, [5]).

We'll be interested in the asymptotics of the solution u^ε to the problem (2.1)-

(2.2) at times of order $\exp(\lambda/\varepsilon^2)$ for $\lambda > 0$ and $\varepsilon \downarrow 0$. Before we study the quasi-linear equation, it is helpful to discuss the linear case, i.e., when a_{ij} do not depend on u^ε. Then the Cauchy problem takes the form

$$\frac{\partial u^\varepsilon(t, x)}{\partial t} = \frac{\varepsilon^2}{2} \sum_{i,j=1}^{d} a_{ij}(x) \frac{\partial^2 u^\varepsilon(t, x)}{\partial x_i \partial x_j} + b(x) \cdot \nabla_x u^\varepsilon(t, x), \quad x \in \mathbb{R}^d, \ t > 0, \qquad (2.3)$$

$$u^\varepsilon(0, x) = g(x), \ x \in \mathbb{R}^d. \qquad (2.4)$$

Let $X_t^{x,\varepsilon}$ be the corresponding family of diffusion processes, namely

$$dX_t^{x,\varepsilon} = b(X_t^{x,\varepsilon})dt + \varepsilon\sigma(X_t^{x,\varepsilon})dW_t, \quad X_0^{x,\varepsilon} = x \in \mathbb{R}^d, \qquad (2.5)$$

where σ is assumed to be Lipschitz continuous and to satisfy $a(x) = (a_{ij}(x)) = \sigma(x)\sigma^*(x)$.

Suppose for a moment that the vector field b has just one asymptotically stable equilibrium point O such that all the points get attracted to O and $(b(x), x - O) \leq -c|x - O|$ for some positive constant c and all sufficiently large $|x|$. Then it is easy to check that

$$\lim_{(\varepsilon,t)\to(0,\infty)} \mathrm{P}(X_t^{x,\varepsilon} \in U) = 1$$

for any neighborhood U of the equilibrium O. Taking into account that the solution u^ε of (2.3)-(2.4) can be written in the form $u^\varepsilon(t, x) = \mathrm{E}g(X_t^{x,\varepsilon})$ and the continuity of g, we conclude that

$$\lim_{(\varepsilon,t)\to(0,\infty)} u^\varepsilon(t, x) = g(O).$$

The situation becomes more complicated if the dynamical system $\dot{x}(t) = b(x(t))$ has more than one asymptotically stable attractor. Assume, for brevity, that all the stable attractors are equilibriums $O_1, ..., O_r$. Let D_i be the basin of O_i,

12

$1 \le i \le r$, and assume that the set $\mathbb{R}^d \setminus (D_1 \cup ... \cup D_r)$ belongs to a finite union of surfaces of dimension $d-1$. The long time behavior of $X_t^{x,\varepsilon}$ and $u^\varepsilon(t,x)$ depends on the way in which (ε, t) approaches $(0, \infty)$ and is now determined by the transitions of $X_t^{x,\varepsilon}$ between the attractors $O_1, ..., O_r$. More precisely, let $T^\varepsilon(\lambda) = \exp(\lambda/\varepsilon^2)$. In the generic case, there is a finite set $\Lambda \subset (0, \infty)$ such that for each $x \in D_1 \cup ... \cup D_r$ and each $\lambda \in (0, \infty) \setminus \Lambda$, one equilibrium $O_{M(x,\lambda)}$ is defined such that the measures $\mu^\varepsilon(\Gamma) = \mathrm{P}(X_{T^\varepsilon(\lambda)}^{x,\varepsilon} \in \Gamma)$ converge weakly to the δ-measure concentrated at $O_{M(x,\lambda)}$. The state $O_{M(x,\lambda)}$ is called the metastable state for the initial point x and the time scale $T^\varepsilon(\lambda)$. From the representation $u^\varepsilon(t,x) = \mathrm{E}g(X_t^{x,\varepsilon})$ it follows that

$$\lim_{\varepsilon \downarrow 0} u^\varepsilon(T^\varepsilon(\lambda), x) = g(O_{M(x,\lambda)}). \tag{2.6}$$

It is worth noting that for all sufficiently large λ, the metastable state $O_{M(x,\lambda)}$ does not depend on x. Therefore, the solution $u^\varepsilon(T^\varepsilon(\lambda), x)$ tends to a constant as $\varepsilon \downarrow 0$ for all sufficiently large λ.

The theory of metastability (of sublimiting distributions) was developed in [6] (see also [7], [4], [8]). The notion of a hierarchy of cycles, which is discussed below, was introduced there. Let $S_{0,T}(\varphi)$ be the action functional for the family $X_t^{x,\varepsilon}$ in $C([0,T], \mathbb{R}^d)$ as $\varepsilon \downarrow 0$ ([4]):

$$S_{0,T}(\varphi) = \frac{1}{2} \int_0^T \sum_{i,j=1}^d a^{ij}(\varphi_t)(\dot{\varphi}_t^i - b_i(\varphi_t))(\dot{\varphi}_t^j - b_j(\varphi_t))dt, \quad T \ge 0, \ \varphi \in C([0,T], \mathbb{R}^d)$$

for absolutely continuous φ, $S_{0,T}(\varphi) = +\infty$ for φ that are not absolutely continuous. Here a^{ij} are the elements of the inverse matrix, that is $a^{ij} = (a^{-1})_{ij}$. The quasi-potential is defined as

$$V(x,y) = \inf_{T,\varphi}\{S_{0,T}(\varphi) : \varphi \in C([0,T], \mathbb{R}^d), \varphi(0) = x, \varphi(T) = y\}, \quad x, y \in \mathbb{R}^d.$$

Note that while the term "quasi-potential" is normally applied to the function V of the variable y with x being a fixed equilibrium point, we use the same term for the function of two variables. The hierarchy of cycles is determined by the numbers

$$V_{ij} = V(O_i, O_j), \quad 1 \leq i, j \leq r.$$

The equilibriums $O_1, ..., O_r$ are the cycles of rank zero. In the generic case, for each O_i there exists a unique "next" equilibrium $O_l = \mathcal{N}(O_i)$ defined by $V_{il} = \min_{k:k\neq i} V_{ik}$. For each sufficiently small $\delta > 0$, with probability close to one as $\varepsilon \downarrow 0$, the process $X_t^{x,\varepsilon}$ that starts in a δ-neighborhood of O_i will enter a δ-neighborhood of $\mathcal{N}(O_i)$ before visiting the basins of any of the equilibriums other than O_i and $\mathcal{N}(O_i)$. The time before the process enters the neighborhood of $O_l = \mathcal{N}(O_i)$ is logarithmically equivalent to $\exp(V_{il}/\varepsilon^2)$. If the sequence $O_i, \mathcal{N}(O_i), \mathcal{N}^2(O_i) = \mathcal{N}(\mathcal{N}(O_i)), ..., \mathcal{N}^n(O_i), ...$ is periodic, that is $\mathcal{N}^n(O_i) = O_i$ for some n, then a cycle of rank one appears. It contains the cycles of rank zero $O_i, \mathcal{N}(O_i), ..., \mathcal{N}^{n-1}(O_i)$. If $\mathcal{N}^n(O_i) \neq O_i$ for any $n \geq 1$, we say that O_i forms a cycle of rank one. The entire set of equilibriums is decomposed into cycles of rank one, which will be denoted by $C_1^1, ..., C_{m_1}^1$. Note that some of the cycles of rank one may consist of one cycle of rank zero.

Next, the transitions between cycles of rank one can be considered. Namely, in the generic case, for each cycle C_i^1 there is a different cycle $\mathcal{N}(C_i^1)$ of rank one determined by V_{ij}, $1 \leq i, j \leq r$, with the following property: if the process starts at one of the equilibrium points in C_i^1, then, with probability close to one as $\varepsilon \downarrow 0$, it will enter a δ-neighborhood of one of the equilibrium points inside the cycle $\mathcal{N}(C_i^1)$

before visiting basins of any of the equilibriums outside C_i^1 and $\mathcal{N}(C_i^1)$. This leads to the decomposition of the set of cycles or rank one into cycles of rank two. This procedure can be continued inductively until we arrive at a single cycle of finite rank R that contains all the equilibrium points.

Let C be a cycle of rank less than R. In the generic case, the process goes from C to $\mathcal{N}(C)$ in the following fashion: there is an equilibrium point $O \in C$ such that, with probability close to one as $\varepsilon \downarrow 0$, the basin of attraction of O is the last one (among the basins of equilibriums that belong to C) visited by the process before the process reaches a δ-neighborhood of $\mathcal{N}(C)$, provided that the process $X_t^{x,\varepsilon}$ starts in a δ-neighborhood of one of the equilibriums that belong to C. We'll say that the process exits C though the equilibrium O and will use the notation $O = \mathcal{E}(C)$ in this situation.

It has been shown (see [4]) that for each cycle C within the hierarchy, the transition from C to $\mathcal{N}(C)$ happens at an exponential time scale. More precisely, let $\tau^{x,\varepsilon}$ be the first time when the process $X_t^{x,\varepsilon}$ enters a δ-neighborhood of $\mathcal{N}(C)$. Then there is $\lambda(C)$ such that $\varepsilon^2 \ln \tau^{x,\varepsilon} \to \lambda(C)$ almost surely as $\varepsilon \downarrow 0$, provided that x belongs to a δ-neighborhood of one of the equilibriums within C.

Now we can introduce a directed labeled graph G associated with the above hierarchy of cycles. The vertices of the graph are the equilibrium points $O_1, ..., O_r$. We'll say that G contains an edge e leading from O_i to O_j if there is a cycle C within the hierarchy such that $O_i = \mathcal{E}(C)$ and $O_j = \mathcal{N}(C)$. We'll use the notation $e = \mathcal{R}(C)$ to indicate that e is the edge leading out of C. With each edge e of the graph, we'll associate a transition time scale $\lambda_e = \lambda(C)$. The meaning is that it

15

takes the process time $\exp(\lambda_e/\varepsilon^2)$ (approximately, up to a sub-exponential factor) to make a transition along the edge e, i.e., from $\mathcal{E}(C)$ to $\mathcal{N}(C)$. The transition time scales λ_e can be expressed through linear combinations of V_{ij}.

Let us return to the non-linear problem (2.1)-(2.2). The first thing to observe is that representation (2.6) is not immediately available now - metastable states need to be replaced by metastable distributions (see [9]), whose dependence on λ needs to be explored. The family of diffusion processes corresponding to a nonlinear problem is also more complicated due to the dependence of the coefficients on time (through the unknown function u^ε). Namely, taking into account the representation of the solution of the (linear) Cauchy problem as the expected value of an appropriate functional of the process, the family corresponding to the problem (2.1)-(2.2) is defined by the following system on the unknown family of processes and unknown function u^ε (see [10], Ch. 5):

$$dX_s^{t,x,\varepsilon} = b(X_s^{t,x,\varepsilon})ds + \varepsilon\sigma(X_s^{t,x,\varepsilon}, u^\varepsilon(t-s, X_s^{t,x,\varepsilon}))dW_s, \quad s \le t, \quad X_0^{t,x,\varepsilon} = x, \quad (2.7)$$

$$u^\varepsilon(t, x) = \mathrm{E}g(X_t^{t,x,\varepsilon}), \quad (2.8)$$

where the entries σ_{ij} of the matrix $\sigma(x, u)$ are Lipschitz continuous and $\sigma\sigma^* = a$. Under the above assumptions on the coefficients and the function g, the solution of the system (2.7)-(2.8) exists and is unique.

While the transition scales were determined by the time-independent coefficients in the linear case, now we will consider a family of transition scales $\lambda_e(z)$. These are obtained by formally replacing the unknown function u^ε by a constant z in the second argument of the diffusion coefficient in the equation (which, formally,

gives us a linear equation). The motivation is, roughly speaking, that in each time scale $T^\varepsilon(\lambda)$ the solution is close to a constant in each of the domains D_i.

We see that now the transition scales evolve in time due their dependence on the (unknown) solution u^ε. Consider, however, a time interval $[T^\varepsilon(\lambda - \delta), T^\varepsilon(\lambda)]$, where δ is small. As will be seen, u^ε typically does not change much in time on this time interval, and the large deviation theory still applies without drastic modifications, which allows us to express the limit of $u^\varepsilon(T^\varepsilon(\lambda), x)$, as $\varepsilon \downarrow 0$, in terms of the limit of $u^\varepsilon(T^\varepsilon(\lambda - \delta), x)$ and the functions $\lambda_e(z)$. This is the main idea that will allow us to study the evolution in λ of the limit of $u^\varepsilon(T^\varepsilon(\lambda), x)$.

In Section 2.2 we discuss some a priori estimates for processes with time-dependent coefficients, assuming that the coefficients can be obtained via a small perturbation of time independent functions. These estimates will be needed for the processes corresponding to the nonlinear problem. In Section 2.3 we review the case of two equilibrium points. This has been considered in an earlier paper ([9]), and we include it here so that to illustrate the technique in the simplest case, when the hierarchy of cycles does not change. The main result, including an inductive description of the changes in the hierarchies of cycles, is stated and proved in Sections 2.4 and 2.5.

2.2 Diffusion processes with time dependent coefficients

Let $\alpha(x)$ be a symmetric $d \times d$ matrix whose elements $\alpha_{ij}(x)$ are Lipschitz continuous with Lipschitz constant L and satisfy

$$k|\xi|^2 \leq \sum_{i,j=1}^{d} \alpha_{ij}(x)\xi_i\xi_j \leq K|\xi|^2, \quad x \in \mathbb{R}^d, \quad \xi \in \mathbb{R}^d. \tag{2.9}$$

Let α^{ij} be the elements of the inverse matrix, that is $\alpha^{ij} = (\alpha^{-1})_{ij}$, and σ be a square matrix such that $\alpha = \sigma\sigma^*$. We choose σ in such a way that σ_{ij} are also Lipschitz continuous.

We assume that all the attractors of the bounded Lipschitz continuous vector field b are equilibriums $O_1, ..., O_r$. Assume that their domains of attraction $D_1, ..., D_r$ are such that the set $\mathbb{R}^d \setminus (D_1 \cup ... \cup D_r)$ belongs to a finite union of surfaces of dimension $d - 1$. We also assume that there are $r > 0$ and $c > 0$ such that

$$(b(x), x - O_i) \leq -c|x - O_i|^2 \tag{2.10}$$

whenever x is in the r-neighborhood of O_i, $1 \leq i \leq r$.

Let $S_{0,T}^{\alpha}$ be the normalized action functional for the family of processes $X_t^{x,\varepsilon}$ satisfying

$$dX_t^{x,\varepsilon} = b(X_t^{x,\varepsilon})dt + \varepsilon\sigma(X_t^{x,\varepsilon})dW_t, \quad X_0^{x,\varepsilon} = x, \tag{2.11}$$

where b is a bounded Lipschitz continuous vector field on $\mathbb{R}^d$. Thus

$$S_{0,T}^{\alpha}(\varphi) = \frac{1}{2} \int_0^T \sum_{i,j=1}^{d} \alpha^{ij}(\varphi_t)(\dot{\varphi}_t^i - b_i(\varphi_t))(\dot{\varphi}_t^j - b_j(\varphi_t))dt$$

for absolutely continuous φ defined on $[0,T]$, $\varphi_0 = x$, and $S_{0,T}^{\alpha}(\varphi) = \infty$ if φ is not absolutely continuous or if $\varphi_0 \neq x$ (see [4]).

We'll be interested in the long-time behavior of processes whose diffusion coefficients are time-dependent, but are close to functions that do not depend on time.

Let $\widetilde{\alpha}^\varepsilon(t,x)$ be a uniformly positive definite symmetric $d \times d$ matrix whose elements $\widetilde{\alpha}^\varepsilon_{ij}$ are continuous in (t,x) and Lipschitz continuous in x. Let $\widetilde{\sigma}^\varepsilon$ be a square matrix such that $\widetilde{\alpha}^\varepsilon = \widetilde{\sigma}^\varepsilon(\widetilde{\sigma}^\varepsilon)^*$. We choose $\widetilde{\sigma}^\varepsilon$ in such a way that $\widetilde{\sigma}^\varepsilon_{ij}$ are also continuous in (t,x) and Lipschitz continuous in x.

Let $\widetilde{X}^{x,\varepsilon}_t$ satisfy $\widetilde{X}^{x,\varepsilon}_0 = x$ and

$$d\widetilde{X}^{x,\varepsilon}_t = b(\widetilde{X}^{x,\varepsilon}_t)dt + \varepsilon\widetilde{\sigma}^\varepsilon(t, \widetilde{X}^{x,\varepsilon}_t)dW_t, \tag{2.12}$$

where b is the same as above. The law of this process depends on $\widetilde{\sigma}^\varepsilon$ only through $\widetilde{\alpha}^\varepsilon = \widetilde{\sigma}^\varepsilon(\widetilde{\sigma}^\varepsilon)^*$. We will assume that the diffusion coefficients for the process $\widetilde{X}^{x,\varepsilon}_t$ are close to those of $X^{x,\varepsilon}_t$. Namely, let us assume that

$$\sup_{(t,x)\in\mathbb{R}^+\times\mathbb{R}^d} |\widetilde{\alpha}^\varepsilon_{ij}(t,x) - \alpha_{ij}(x)| \leq \varkappa, \tag{2.13}$$

where $\varkappa$ is small. The reason to introduce the process $\widetilde{X}^{x,\varepsilon}_t$ is that we would like to study the behavior of the process $X^{t,x,\varepsilon}_s$ given by (2.7)-(2.8) on a time interval where the variable u^ε found inside the diffusion coefficient of (2.7) does not change much. Since a-priori we don't know much about the behavior of the diffusion coefficients in (2.7) (other than that they don't significantly change in time on a certain time interval), it is convenient to consider a generic process whose diffusion coefficients are close to functions that don't depend on time.

Given a domain D and $\delta > 0$, we define

$$D^\delta = \{x \in D : \operatorname{dist}(x, \partial D) \geq \delta,\ |x| \leq 1/\delta\}.$$

Let x_0 be an asymptotically stable equilibrium of b and D be a domain attracted to

x_0. Let

$$v = \inf_{T,\varphi}\{S_{0,T}^\alpha(\varphi) : \varphi \in C([0,T],\overline{D}), \varphi(0) = x_0, \varphi(T) \in \partial D\}.$$

Lemma 2.2.1. *Suppose that b is fixed, α is Lipschitz continuous with Lipschitz*

constant L, $\widetilde{\alpha}^\varepsilon$ is continuous in (t,x) and Lipschitz continuous in x, and

$$k|\xi|^2 \leq \sum_{i,j=1}^{d} \alpha_{ij}(x)\xi_i\xi_j \leq K|\xi|^2 \quad for \quad x \in D, \quad \xi \in \mathbb{R}^d,$$

$$k|\xi|^2 \leq \sum_{i,j=1}^{d} \widetilde{\alpha}_{ij}^\varepsilon(t,x)\xi_i\xi_j \leq K|\xi|^2 \quad for \quad (t,x) \in \mathbb{R}^+ \times D, \quad \xi \in \mathbb{R}^d. \qquad (2.14)$$

For each $\delta > 0$ there are $\varkappa > 0$ and a function $\rho(\varepsilon)$ (that depend on α and $\widetilde{\alpha}$ through

L, k and K) such that $\lim_{\varepsilon \downarrow 0} \rho(\varepsilon) = 0$ and

$$\sup_{(t,x)\in[T^\varepsilon(\delta),T^\varepsilon(v-\delta)]\times D^\varkappa} \mathrm{P}(|\widetilde{X}_t^{x,\varepsilon} - x_0| < \delta, \ \widetilde{X}_s^{x,\varepsilon} \in D \ \ for \ \ s \leq t) \geq 1 - \rho(\varepsilon),$$

provided that

$$\sup_{(t,x)\in\mathbb{R}^+\times D^\varkappa} |\widetilde{\alpha}_{ij}^\varepsilon(t,x) - \alpha_{ij}(x)| \leq \varkappa.$$

This lemma was proved in [9]. The main idea was to show that S^α serves

a purpose similar to the action functional for the process $\widetilde{X}_t^{x,\varepsilon}$, even though the

diffusion coefficients for the process are time-dependent.

The next simple lemma does not require the proximity of $\widetilde{\alpha}^\varepsilon$ to α, but only

the boundedness of the entries of $\widetilde{\alpha}^\varepsilon$. It can be proved by standard arguments from

large deviation theory (compare with chapter 3 of [4]).

Lemma 2.2.2. *Suppose that b is fixed and $\widetilde{\alpha}^\varepsilon$ is continuous in (t,x) and Lipschitz*

continuous in x and satisfies (2.14). For any compact $M \subset D$, there is $v_0 > 0$ which

20

depends on $\widetilde{\alpha}^{\varepsilon}$ only through K such that for each $\delta \in (0, v_0)$ there is a function $\rho(\varepsilon)$

such that $\lim_{\varepsilon \downarrow 0} \rho(\varepsilon) = 0$ and

$$\sup_{(t,x) \in [T^{\varepsilon}(\delta), T^{\varepsilon}(v_0)] \times M} \mathrm{P}(|\widetilde{X}_t^{x,\varepsilon} - x_0| < \delta, \ \widetilde{X}_s^{x,\varepsilon} \in D \ \text{ for } \ s \le t) \ge 1 - \rho(\varepsilon).$$

The next lemma implies that the solution of (2.1)-(2.2) is nearly constant

inside each of the domains $D_i^{\delta} = \{x \in D_i : \mathrm{dist}(x, \partial D_i) \ge \delta, \ |x| \le 1/\delta\}$, $\delta > 0$,

$1 \le i \le r$, for ε small enough.

Lemma 2.2.3. *Let u^{ε} be the solution of (2.1)-(2.2). For every positive λ_0 and δ*

there is a positive ε_0 such that

$$|u^{\varepsilon}(T^{\varepsilon}(\lambda), x) - u^{\varepsilon}(T^{\varepsilon}(\lambda), O_i)| \le \delta \tag{2.15}$$

whenever $x \in D_i^{\delta}$, $\varepsilon \le \varepsilon_0$ and $\lambda \ge \lambda_0$.

For a proof of this lemma we refer the reader to [11], where the same statement

was proved in the case of a single domain. The main idea is to express the solution

at time $T^{\varepsilon}(\lambda)$ in terms of the solution at a slightly earlier time $T^{\varepsilon}(\lambda) - t$ as follows

$$u^{\varepsilon}(T^{\varepsilon}(\lambda), x) = \mathrm{E}u^{\varepsilon}\left(T^{\varepsilon}(\lambda) - t, X_t^{T^{\varepsilon}(\lambda), x, \varepsilon}\right),$$

$$u^{\varepsilon}(T^{\varepsilon}(\lambda), O_i) = \mathrm{E}u^{\varepsilon}\left(T^{\varepsilon}(\lambda) - t, X_t^{T^{\varepsilon}(\lambda), O_i, \varepsilon}\right).$$

If t is chosen appropriately, then both $X_t^{T^{\varepsilon}(\lambda), x, \varepsilon}$ and $X_t^{T^{\varepsilon}(\lambda), O_i, \varepsilon}$ belong to a small

neighborhood of O_i with overwhelming probability. Therefore, the right hand sides

in the expressions above are very close due to a priori estimates on solutions of

quasi-linear PDEs ([12]).

2.3 The case of two equilibrium points

In this section we assume that there are two asymptotically stable equilibrium points $O_1, O_2 \in \mathbb{R}^d$. Let $D_1 \subset \mathbb{R}^d$ be the set of points in $\mathbb{R}^d$ that are attracted to O_1 and $D_2 \subset \mathbb{R}^d$ the set of points attracted to O_2. We assume that $D_1 \cup D_2 \in \mathbb{R}^d \setminus S$, where S is a $(d-1)$-dimensional manifold. Note that in the case of two equilibrium points, the hierarchy of cycles is always the same: O_1 and O_2 are cycles of rank zero, and there is one cycle of rank one which contains both O_1 and O_2.

Let $g_{\min} = \inf_{x \in \mathbb{R}^d} g(x)$ and $g_{\max} = \sup_{x \in \mathbb{R}^d} g(x)$. Let G be the graph with two vertices O_1 and O_2 and two directed edges: e going from O_1 to O_2 and h going from O_2 to O_1. Define the functions $\lambda_e, \lambda_h : [g_{\min}, g_{\max}] \to \mathbb{R}$ via

$$\lambda_e(z) = V_{O_1,O_2}^{a(\cdot,z)}, \quad \lambda_h(z) = V_{O_2,O_1}^{a(\cdot,z)},$$

where in the right hand side we have quasi-potentials for the linear problem obtained by inserting constant z instead of the variable u^ε in the diffusion coefficient. An example with the graphs of these functions is shown in Figure 2.1.

Without loss of generality we may assume that $g(O_1) \leq g(O_2)$. Let $\lambda_1 = \lambda_e(g(O_1))$ and $\lambda_2 = \lambda_h(g(O_2))$. In order to formulate the results on the asymptotics of $u^\varepsilon(T^\varepsilon(\lambda), x)$, we need the functions $z^1(\lambda)$ and $z^2(\lambda)$, $\lambda > 0$, defined as follows:

$$z^1(\lambda) = \begin{cases} g(O_1), & 0 < \lambda < \lambda_1, \\ \min\{g(O_2), \min\{z : z \in [g(O_1), g(O_2)], \lambda_e(z) = \lambda\}\}, & \lambda \geq \lambda_1, \end{cases}$$
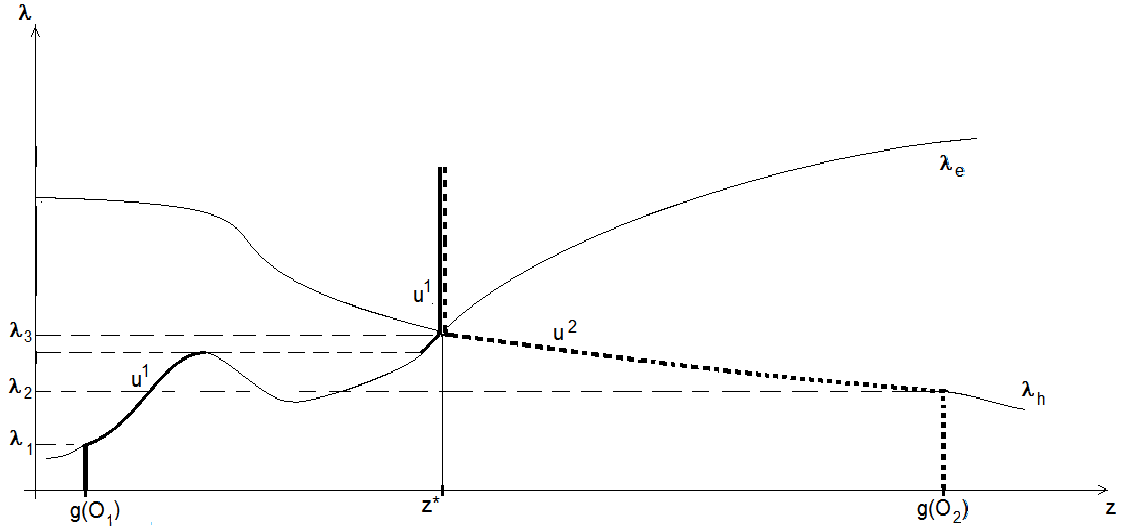
$$(2.16)$$

22

Figure 2.1: The case of two equilibrium points

$$
z^2(\lambda) = \begin{cases} g(O_2), & 0 < \lambda < \lambda_2, \\ \max\{g(O_1), \max\{z : z \in [g(O_1), g(O_2)], \lambda_h(z) = \lambda\}\}, & \lambda \geq \lambda_2. \end{cases}
$$
(2.17)

Let $\lambda_3 = \inf\{\lambda : z^1(\lambda) \geq z^2(\lambda)\}$. Assume that the functions z^1 and z^2 are continuous at λ_3. Let $z^* = z^1(\lambda_3) = z^2(\lambda_3)$. Let $u^1(\lambda) = \min(z^1(\lambda), z^*)$ and $u^2(\lambda) = \max(z^2(\lambda), z^*)$. On Figure 2, the graphs of u^1 and u^2 are denoted by the thick and the dotted lines, respectively.

The asymptotics of $u^\varepsilon(T^\varepsilon(\lambda), x)$ is described by the following theorem.

Theorem 2.3.1. *Let the above assumptions be satisfied. Suppose that the function $u^1(\lambda)$ is continuous at a point $\lambda \in (0, \infty)$. Then for every $\delta > 0$ the following limit*

$$
\lim_{\varepsilon \downarrow 0} u^\varepsilon(T^\varepsilon(\lambda), x) = u^1(\lambda)
$$

23

is uniform in $x \in D_1^\delta$. Suppose that the function $u^2(\lambda)$ is continuous at a point

$\lambda \in (0, \infty)$. Then for every $\delta > 0$ the following limit

$$\lim_{\varepsilon \downarrow 0} u^\varepsilon(T^\varepsilon(\lambda), x) = u^2(\lambda)$$

is uniform in $x \in D_2^\delta$.

Before we proceed with the proof of the theorem, let us briefly discuss Figure 2.1. Observe that there are several "special" time scales, where the behavior of the functions $u^1(\lambda)$ and $u^2(\lambda)$ changes qualitatively. Namely, these functions are constants (equal to $g(O_1)$ and $g(O_2)$, respectively) for small values of λ. By drawing the vertical segment through the point $(g(O_1), 0)$ in the (z, λ) plane till the intersection with the graph of λ_e, we locate the time scale λ_1. After the time scale λ_1, the function u^1 is just the inverse of the function λ_e for a certain range of values of λ. (The time scale λ_2 plays the same role for the function u^2.) Then, after λ_3, which corresponds to the intersection of the graphs of λ_e and λ_h, the functions $u^1(\lambda)$ and $u^2(\lambda)$ become constant again. Note that there is also a point of discontinuity for $u^1(\lambda)$, which corresponds to a local maximum of λ_e.

In the general case (more than two equilibriums) we'll encounter several type of special time scales. In particular, some will correspond to the intersections of vertical lines going through $(g(O_i), 0)$ with the graphs of λ_e, others will correspond to intersection points between λ_e and λ_h, yet others will correspond to local maxima of λ_e for some e (and there will be other special time scales that are not found in this example).

Proof of Theorem 2.3.1. Let us show that if z^1 is continuous at λ, then

$$\limsup_{\varepsilon \downarrow 0} \sup_{x \in D_1^\delta} u^\varepsilon(T^\varepsilon(\lambda), x) \leq z^1(\lambda). \tag{2.18}$$

Similarly, if z^2 is continuous at λ, then

$$\liminf_{\varepsilon \downarrow 0} \inf_{x \in D_2^\delta} u^\varepsilon(T^\varepsilon(\lambda), x) \geq z^2(\lambda). \tag{2.19}$$

Due to Lemma 2.2.3, in order to prove (2.18), it is sufficient to show that

$$\limsup_{\varepsilon \downarrow 0} u^\varepsilon(T^\varepsilon(\lambda), O_1) \leq z^1(\lambda), \tag{2.20}$$

Note that by Lemma 2.2.2 and (2.8) there is a positive v_0 such that for every $0 < \delta < v_0$ there is $\varepsilon_0 > 0$ such that

$$|u^\varepsilon(T^\varepsilon(\lambda), x) - g(O_i)| \leq \delta \tag{2.21}$$

whenever $x \in D_i^\delta$, $0 < \varepsilon \leq \varepsilon_0$ and $\delta \leq \lambda \leq v_0$.

If (2.20) fails for a certain value of λ, then due to continuity of the functions $u^\varepsilon(t, O_i)$ in t, it follows from (2.21) that for an arbitrarily small $\delta' > 0$ there are sequences $\varepsilon_n \downarrow 0$ and $\lambda_n \in [\delta', \lambda]$ such that

$$u^{\varepsilon_n}(t, O_1) \leq z^1(\lambda) + \delta', \quad T^{\varepsilon_n}(\delta') \leq t \leq T^{\varepsilon_n}(\lambda_n)$$

and

$$u^{\varepsilon_n}(T^{\varepsilon_n}(\lambda_n), O_1) = z^1(\lambda) + \delta'.$$

Take $\delta'' \in (0, \delta')$ which will be specified later. Due to the continuity of $u^{\varepsilon_n}(t, O_1)$ in t, we can find a sequence $\mu_n \in [\delta', \lambda_n)$ such that

$$u^{\varepsilon_n}(T^{\varepsilon_n}(\mu_n), O_1) = z^1(\lambda) + \delta''$$

and

$$u^{\varepsilon_n}(t, O_1) \in [z^1(\lambda) + \delta'', z^1(\lambda) + \delta'] \quad \text{for} \quad t \in [T^{\varepsilon_n}(\mu_n), T^{\varepsilon_n}(\lambda_n)]. \tag{2.22}$$

We can express $u^{\varepsilon_n}(T^{\varepsilon_n}(\lambda_n), O_1)$ in terms of the process $X_s^{T^{\varepsilon_n}(\lambda_n), O_1, \varepsilon}$ and the solution at the earlier time $T^{\varepsilon_n}(\mu_n)$ as follows

$$u^{\varepsilon_n}(T^{\varepsilon_n}(\lambda_n), O_1) = \mathrm{E} u^{\varepsilon_n} \left(T^{\varepsilon_n}(\mu_n), X_{T^{\varepsilon_n}(\lambda_n) - T^{\varepsilon_n}(\mu_n)}^{T^{\varepsilon_n}(\lambda_n), O_1, \varepsilon_n} \right). \tag{2.23}$$

Since z^1 is continuous at λ, there are arbitrarily small $\delta' > 0$ such that $\lambda_e(z^1(\lambda) + \delta') > \lambda_e(z^1(\lambda)) = \lambda$. Since $\lambda_n \leq \lambda$, a process starting at O_1 and satisfying (2.11) with

$$\sigma\sigma^*(x) = a(x, u^{\varepsilon_n}(T^{\varepsilon_n}(\lambda_n), O_1)) = a(x, z^1(\lambda) + \delta'))$$

will be in an arbitrarily small neighborhood of O_1 at time $T^{\varepsilon_n}(\lambda_n) - T^{\varepsilon_n}(\mu_n)$ with probability which tends to one when $\varepsilon_n \downarrow 0$. By Lemma 2.2.1, this remains true if the constant $u^{\varepsilon_n}(T^{\varepsilon_n}(\lambda_n), O_1)$ is replaced by a function which is sufficiently close to this constant in D_1^δ, where δ is sufficiently small. Therefore, due to (2.22) and Lemma 2.2.3, we can choose δ'' sufficiently close to δ' so that $X_{T^{\varepsilon_n}(\lambda_n) - T^{\varepsilon_n}(\mu_n)}^{T^{\varepsilon_n}(\lambda_n), O_1, \varepsilon_n}$ will be in a small neighborhood of O_1 with probability which tends to one when $\varepsilon_n \downarrow 0$. With δ' and δ'' thus fixed, we let $\varepsilon_n \downarrow 0$ in (2.23). The left hand side is equal to $z^1(\lambda) + \delta'$, while the right hand side tends to $z^1(\lambda) + \delta''$. This leads to a contradiction which proves that (2.20) holds, which in turn implies that (2.18) holds. The proof of (2.19) is completely similar.

Note that the arguments used to prove (2.20) also lead to the following state-

ment: for each $\lambda_0 > 0$

$$\limsup_{\varepsilon \downarrow 0} \sup_{\lambda' \in [\lambda_0, \lambda]} u^\varepsilon(T^\varepsilon(\lambda'), O_1) \leq \lim_{\lambda' \downarrow \lambda} z^1(\lambda'), \tag{2.24}$$

now without assuming that z^1 is continuous at λ. Similarly, for each $\lambda_0 > 0$

$$\liminf_{\varepsilon \downarrow 0} \inf_{\lambda' \in [\lambda_0, \lambda]} u^\varepsilon(T^\varepsilon(\lambda'), O_2) \geq \lim_{\lambda' \downarrow \lambda} z^2(\lambda'). \tag{2.25}$$

Let us show that if z^1 is continuous at λ, then

$$\liminf_{\varepsilon \downarrow 0} \inf_{x \in D_1^\delta} u^\varepsilon(T^\varepsilon(\lambda), x) \geq \min(z^1(\lambda), \lim_{\lambda' \downarrow \lambda} z^2(\lambda')). \tag{2.26}$$

Similarly, if z^2 is continuous at λ, then

$$\limsup_{\varepsilon \downarrow 0} \sup_{x \in D_2^\delta} u^\varepsilon(T^\varepsilon(\lambda), x) \leq \max(z^2(\lambda), \lim_{\lambda' \downarrow \lambda} z^1(\lambda')). \tag{2.27}$$

Due to Lemma 2.2.3, in order to prove (2.26), it is sufficient to show that

$$\liminf_{\varepsilon \downarrow 0} u^\varepsilon(T^\varepsilon(\lambda), O_1) \geq \min(z^1(\lambda), \lim_{\lambda' \downarrow \lambda} z^2(\lambda')). \tag{2.28}$$

If (2.28) fails, then for each $\lambda_0 > 0$ there is $\delta' > 0$ and a sequence $\varepsilon_n \downarrow 0$ such that

$$u^{\varepsilon_n}(T^{\varepsilon_n}(\lambda), O_1) < z^1(\lambda) - \delta'. \tag{2.29}$$

$$u^{\varepsilon_n}(T^{\varepsilon_n}(\lambda), O_1) < \inf_{\lambda' \in [\lambda_0, \lambda]} u^{\varepsilon_n}(T^{\varepsilon_n}(\lambda'), O_2) - \delta'. \tag{2.30}$$

These two inequalities can not hold at the same time as follows from Lemma 3.11 of [11], where an analogue of (2.29) is ruled out for the case of the initial-boundary value problem with one equilibrium point inside the domain. Now the boundary condition is replaced by the presence of the second equilibrium point, but due to (2.30) the proof goes through without major modifications. We have thus justified (2.26), and (2.27) is absolutely similar.

Note that (2.18), (2.19), (2.26), and (2.27) imply the statement of the theorem for $0 < \lambda < \lambda_3$. Expressing the solution at time $T^\varepsilon(\lambda)$ in terms of the solution at an earlier time $T^\varepsilon(\lambda')$ (similarly to (2.23)), we see that if

$$\liminf_{\varepsilon \downarrow 0} \inf_{x \in D_1^\delta} u^\varepsilon(T^\varepsilon(\lambda'), x) \le \limsup_{\varepsilon \downarrow 0} \sup_{x \in D_2^\delta} u^\varepsilon(T^\varepsilon(\lambda'), x),$$

then

$$\liminf_{\varepsilon \downarrow 0} \inf_{x \in D_1^\delta} u^\varepsilon(T^\varepsilon(\lambda'), x) \le \liminf_{\varepsilon \downarrow 0} \inf_{x \in D_1^\delta \cup D_2^\delta} u^\varepsilon(T^\varepsilon(\lambda), x) \le$$

$$\le \limsup_{\varepsilon \downarrow 0} \sup_{x \in D_1^\delta \cup D_2^\delta} u^\varepsilon(T^\varepsilon(\lambda), x) \le \limsup_{\varepsilon \downarrow 0} \sup_{x \in D_2^\delta} u^\varepsilon(T^\varepsilon(\lambda'), x).$$

As follows from the definition of the functions $u^1(\lambda)$ and $u^2(\lambda)$, this allows us to extend the result to $\lambda \ge \lambda_3$. $\qquad \square$

Remark. If $\lambda > \lambda_3$, then $u^1(\lambda) = u^2(\lambda) = z^*$. It is possible to show that the limit

$$\lim_{\varepsilon \downarrow 0} u^\varepsilon(T^\varepsilon(\lambda), x) = z^*$$

is uniform in $(x, \lambda) \in B_{1/\delta} \times [\overline{\lambda}, \infty)$ for each $\overline{\lambda} > \lambda_3$, where $B_{1/\delta}$ is the ball of radius $1/\delta$ centered at the origin. Therefore, for each $\delta > 0$ and $\overline{\lambda} > \lambda_3$ there is $\varepsilon_0 > 0$ such that

$$|u^\varepsilon(t, x) - z^*| \le \delta$$

whenever $\varepsilon \in (0, \varepsilon_0)$, $x \in B_{1/\delta}$ and $t \ge T^\varepsilon(\overline{\lambda})$.

2.4 General case: preliminary considerations, notations and assumptions

We assume that there are r equilibrium points $O_1, ..., O_r$. Recall that their domains of attraction are denoted by $D_1, ..., D_r$. As follows from Lemma 2.2.3, for a fixed value of $\lambda > 0$ and $\varepsilon \downarrow 0$, the solution $u^\varepsilon(T^\varepsilon(\lambda), x)$ is nearly constant inside each of the domains D_i. We'll prove that under certain assumptions there is a finite set $\Lambda = \{\lambda_0, ..., \lambda_N\}$ such that there are limits

$$u^i(\lambda) = \lim_{\varepsilon \downarrow 0} u^\varepsilon(T^\varepsilon(\lambda), x), \quad x \in D_i, \quad \lambda \in (0, \infty) \setminus \Lambda. \tag{2.31}$$

The functions u^i are continuous on $(0, \infty) \setminus \Lambda$ and are determined by the coefficients of the equation through the quasi-potential. They will be described inductively: first we'll explain how to define each $u^i(\lambda)$ on the interval $(0, \lambda_1)$, then, given the values of $\lim_{\lambda \uparrow \lambda_1} u^i(\lambda)$ for all i, we define $\lim_{\lambda \downarrow \lambda_1} u^i(\lambda)$, then define $u^i(\lambda)$ for $\lambda \in (\lambda_1, \lambda_2)$, etc. As a by-product of our construction, we'll see that $u^i(\lambda)$ does not depend on either i or λ for $\lambda > \lambda_N$, that is the solution of the PDE tends to a constant in the appropriate time scale.

While constructing the set Λ of "special" time scales, we'll make certain assumptions on the coefficients of the equation. These assumptions will be satisfied in a wide range of situations.

For a fixed value of $z \in \mathbb{R}$, we consider the family of processes obtained from (2.7) by formally setting u^ε to be identically equal to z. We drop the superscript corresponding to the initial time (since the coefficients are now time-independent)

29

and insert the superscript z to indicate the dependence of the coefficients on the choice of the constant, denoting the resulting process by $Y_t^{z,x,\varepsilon}$. Thus

$$dY_t^{z,x,\varepsilon} = b(Y_t^{z,x,\varepsilon})dt + \varepsilon\sigma(Y_t^{z,x,\varepsilon}, z)dW_t, \quad Y_0^{z,x,\varepsilon} = x, \qquad (2.32)$$

We can construct the hierarchy of cycles corresponding to $Y_t^{z,x,\varepsilon}$, provided that the "next" equilibrium point can be defined uniquely for each of the cycles. Observe that different values of z may result, in principle, in different hierarchies of cycles. We'll in interested in the values of z that belong to the interval $[g_{\min}, g_{\max}]$, where $g_{\min} = \inf_{x \in \mathbb{R}^d} g(x)$ and $g_{\max} = \sup_{x \in \mathbb{R}^d} g(x)$, since it will be seen that the values of the limits in (2.31) also belong to this interval. In the generic situation, there will be a finite set of points $Z = \{z_1, ..., z_K\} \subseteq (g_{\min}, g_{\max})$ such that the notion of the "next" equilibrium is defined uniquely for $z \in [g_{\min}, g_{\max}] \setminus Z$ and the hierarchy doesn't change on each of the segments $I_1 = (g_{\min}, z_1), I_2 = (z_1, z_2), ..., I_{K+1} = (z_K, g_{\max})$.

The hierarchy corresponding to a given value of $z \in I_k$ will be denoted by H_k. Recall that H_k can be viewed as a directed graph. For an edge $e \in H_k$, we define $C_k^-(e)$ to be the cycle such that e connects $\mathcal{E}_k(C_k^-(e))$ with $\mathcal{N}_k(C_k^-(e))$. The notation $\mathcal{E}(C)$ and $\mathcal{N}(C)$ has been introduced in Section 2.1. Here we use an additional subscript k to stress that the hierarchy corresponds to $z \in I_k$.

We define $C_k^+(e)$ to be the smallest cycle that contains e. Let $\lambda_e(z)$ be the transition scale along the edge $e \in H_k$ for the processes $Y_t^{z,x,\varepsilon}$.

Let us spell out some of the assumptions on the functions $\lambda_e(z)$. Since $\lambda_e(z)$ can be expressed in terms of $a_{ij}(\cdot, z)$ and $b(\cdot)$, these are essentially assumptions on the coefficients of the equation. We assume that if $\lambda_e(z)$ is defined on I_k, then it

can be continued to the end-points of the segment in such a way that it is a smooth function of z for $z \in \overline{I}_k$, where $\overline{I}_k$ is the closure of I_k. We'll assume that each λ_e has only a finite number of critical points (i.e., where the derivative is zero) and that these points do not coincide with the end-points of I_k. We do not, however, exclude the possibility that λ_e is defined on both I_k and I_{k+1}, and when continued to the common end point of the two segments $d\lambda_e(z)/dz$ has two different values (possibly of opposite signs).

Later, as we consider specific cases, we'll introduce additional assumptions.

Changes in the Hierarchy.

It is important to understand how the quantities $\lambda_e(z)$ behave and how the hierarchy changes when z passes through a point $z_k \in Z$. What causes a change in the hierarchy, is that the notion of the "next" cycle, which is uniquely defined for z immediately to the left of z_k, is not unique for $z = z_k$ for one of the cycles. We assume that for each z_k only one cycle gets thus affected (possibly leading to creation or destruction of another cycle, as will be discussed below). Let $e \in H_k$ be the edge that connects $\mathcal{E}_k(C_k^-(e))$ with $\mathcal{N}_k(C_k^-(e))$. Let $g \in H_{k+1}$, $g \neq e$, be the edge that connects $\mathcal{E}_{k+1}(C_{k+1}^-(g))$ with $\mathcal{N}_{k+1}(C_{k+1}^-(g))$, where $C_k^-(e) = C_{k+1}^-(g)$. In other words, it is the cycle $C_k^-(e) = C_{k+1}^-(g)$ that gets affected by the change in the notion of the next equilibrium when z passes through z_k.

We'll assume that if an edge $h \neq e, g$ is represented in the hierarchy H_k (or H_{k+1}), then it is also represented in H_{k+1} (or H_k). We also assume that $\lambda_h(z_k)$ is distinct from $\lim_{z \uparrow z_k} \lambda_e(z) = \lim_{z \downarrow z_k} \lambda_g(z)$ and that $\lambda_{h_1}(z_k) \neq \lambda_{h_2}(z_k)$ for h_1 and h_2

that are different from e and g.

We distinguish three ways in which a change in the hierarchy can take place.

(a) Assume that $\mathcal{N}_{k+1}(C_{k+1}^-(g)) \in C_k^+(e)$. In this case, the only difference between H_k and H_{k+1} is that H_{k+1} contains the extra edge g. For H_{k+1}, the edge e connects $C_{k+1}^+(g)$ with $\mathcal{N}_{k+1}(C_{k+1}^+(g))$. We'll say that the hierarchy undergoes the restructuring of the first kind at z_k in this case. (See Figure 2.2.)
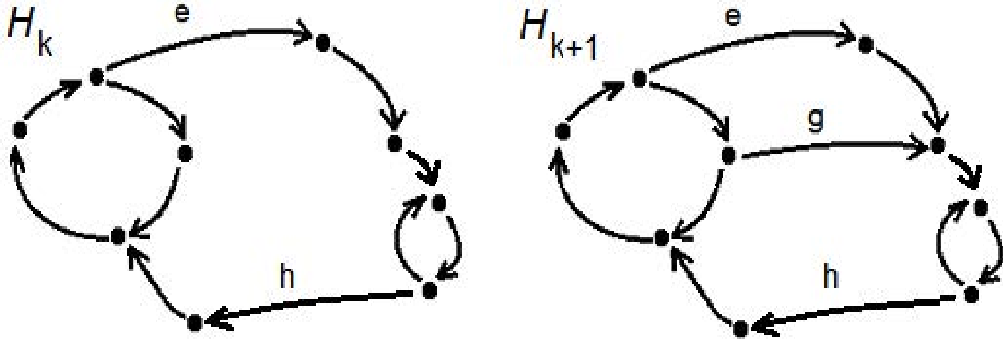

Figure 2.2: Restructuring of the first kind - the hierarchies H_k and H_{k+1}.

(b) Another possibility is that $\mathcal{N}_{k+1}(C_{k+1}^-(g)) \notin C_k^+(e)$, but $\mathcal{N}_k(C_k^-(e)) \in C_{k+1}^+(g)$. In this case, the only difference between H_k and H_{k+1} is that H_{k+1} does not contain the edge e. For H_k, the edge g connects $C_k^+(e)$ with $\mathcal{N}_k(C_k^+(e))$. We'll say that the hierarchy undergoes the restructuring of the second kind at z_k in this case. (See Figure 2.3.)

(c) Finally, it is possible that $\mathcal{N}_{k+1}(C_{k+1}^-(g)) \notin C_k^+(e)$ and $\mathcal{N}_k(C_k^-(e)) \notin C_{k+1}^+(g)$. In this case, H_k and H_{k+1} contain the same edges, but the notion of the next cycle changes, i.e., $\mathcal{N}_k(C_k^-(e)) \neq \mathcal{N}_{k+1}(C_{k+1}^-(g))$. For H_k, g connects $C_k^+(e)$ with $\mathcal{N}_k(C_k^+(e))$. For H_{k+1}, e connects $C_{k+1}^+(g)$ with $\mathcal{N}_{k+1}(C_{k+1}^+(g))$. We'll say that
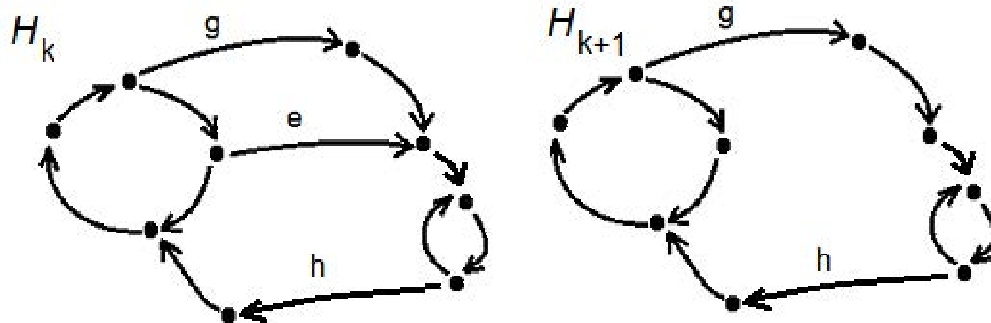
Figure 2.3: Restructuring of the second kind - the hierarchies H_k and H_{k+1}.

the hierarchy undergoes the restructuring of the third kind at z_k in this case. (See Figure 2.4.)

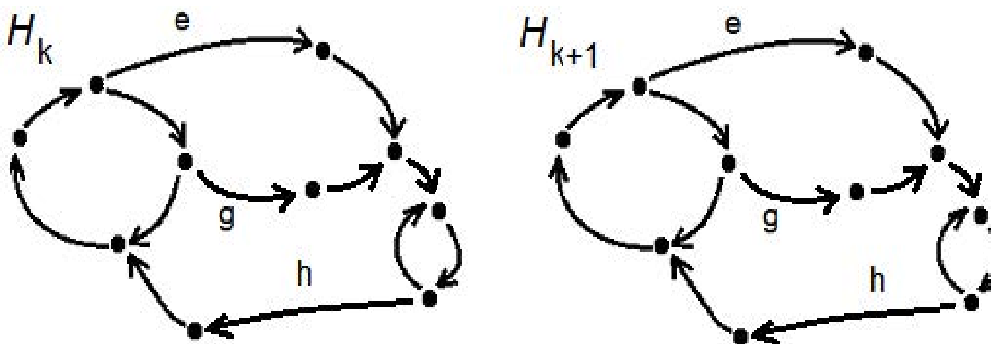Figure 2.4: Restructuring of the third kind - the hierarchies H_k and H_{k+1}.

Observe a certain symmetry: a restructuring of the first (second) kind can be viewed as a restructuring of the second (first) kind, respectively, as we go from H_{k+1} to H_k. A restructuring of the third kind remains a restructuring of the third kind as we go from H_{k+1} to H_k.

Special time scales.

Now let us describe the set Λ of special time scales and the corresponding set $\Pi \subset \mathbb{R} \times \mathbb{R}^+$ of points in the (z, λ) space. We define

$$\Lambda_0 = \{0\}, \quad \Pi_0 = \{(g(O_1), 0), ..., (g(O_r), 0)\}.$$

We'll assume that all the $g(O_i)$ are distinct.

Let $z \in I_k$. We'll say that z is a maximum point if there is $e \in H_k$ such that the function $\lambda_e(\cdot)$ has a local maximum at z. Let's denote the set of maximum points by M. We'll assume that M is finite and that no two functions λ_{e_1}, λ_{e_2} have a local maximum at the same point $z \in I_k$ for distinct e_1, e_2. Define

$$\Lambda_1 = \{\lambda_e(z) : z \in M, \ \lambda_e \text{ has a local maximum at } z\},$$

$$\Pi_1 = \{(z, \lambda_e(z)) : z \in M, \ \lambda_e \text{ has a local maximum at } z\}.$$

We assume that the number of elements in Λ_1 is the same as in M, i.e., no two maximum points result in the same value of λ_e.

We'll say that $z \in I_k$ is an intersection point if there are $e, g \in H_k$ such that $\lambda_e(z) = \lambda_g(z)$. We'll assume that the set of intersection points is finite and that each intersection point arises as an intersection of exactly two functions (i.e., there are no triple intersections). Let's denote the set of intersection points by N. Define

$$\Lambda_2 = \{\lambda_e(z) : z \in N, \ \lambda_e(z) = \lambda_g(z) \text{ for } e \neq g\},$$

$$\Pi_2 = \{(z, \lambda_e(z)) : z \in N, \ \lambda_e(z) = \lambda_g(z) \text{ for } e \neq g\}.$$

We assume that the number of elements in Λ_2 is the same as in N, i.e., no two intersection points result in the same value of λ_e.

Define

$$\Lambda_3 = \bigcup_{k=1}^{K} \Lambda_3^k = \left(\bigcup_{k=1}^{K} \{\lim_{z \uparrow z_k} \lambda_e(z), e \in H_k\} \right) \bigcup \left(\bigcup_{k=1}^{K} \{\lim_{z \downarrow z_k} \lambda_e(z), e \in H_{k+1}\} \right),$$

i.e., Λ_3 is determined by the intersections of the graphs of λ_e for all e with the vertical lines given by $z = z_k$ for all k. We'll assume that $\Lambda_3^{k_1}$ and $\Lambda_3^{k_2}$ do not intersect for distinct k_1 and k_2. Define

$$\Pi_3 = \bigcup_{k=1}^{K} \{(z_k, \lambda), \lambda \in \Lambda_3^k\}.$$

We'll assume that $\Lambda_0, ..., \Lambda_3$ are disjoint.

The cascade of special time scales.

So far, we operated with the hierarchies H_k corresponding to the process (2.32) for different values of z. The motivation is that processes of this type come up if we consider the value of the function u^ε in (2.7) to be fixed and identically equal to z. The full picture is somewhat more complicated. For each value of λ, the solution is approximately constant in each of the domains D_i, yet the constants may be different from one domain to another. Moreover, at a given timescale λ, the process might or might not be able to exit a domain D_i (or a larger cycle) in time $\exp(\lambda/\varepsilon^2)$. At each time scale, we'll be interested in whether the process can or can not exit a cycle, and below we'll classify individual cycles within the hierarchy accordingly. The time scales where this classification may change form the "special" set that includes the union of $\Lambda_1, ..., \Lambda_3$. Besides the sets $\Lambda_1, ..., \Lambda_3$, there are other "special" time scales that we'll now define inductively.

Let $\overline{\Lambda} = \Lambda_0 \cup ... \cup \Lambda_3$ and $\overline{\overline{\Pi}} = \Pi_0 \cup ... \cup \Pi_3$. Let $p = (z, 0) \in \Pi_0$. We'll say that p is a point of the first generation. For each point p of the first generation, we consider the ray l that goes vertically upward in the (z, λ) plane starting at p. All the intersection points of l with the graphs of λ_e (for all e) are considered points of the second generation. We assume that the vertical rays (here and in the construction below) don't intersect any of the graphs at a point where the tangent to the graph is horizontal.

Let $\widetilde{l}$ be the horizontal ray in the (z, λ) plane that starts at a point of the second generation (intersection of l with the graph of λ_e) and goes in the direction where the function λ_e decreases. All the other intersection points of the ray $\widetilde{l}$ with the graphs of λ_g (for all g, where λ_g is defined on the closure of I_k for some k) are considered the points of the third generation. We continue the process inductively by alternating between the vertical rays and the horizontal rays in the direction where the corresponding λ_e decreases. The set of points of all generations obtained in this way will be called the cascade of points generated by p. We assume that none of the points of the cascade belongs to $\overline{\Lambda}$, other than the point of the first generation. Moreover, we assume that the same point of the cascade can't be obtained by following two distinct sequences of horizontal and vertical rays originating at p.

The definition of a cascade starting at $p \in \Pi_2$ is the same. The cascade starting at $p \in \Pi_1 \cup \Pi_3$ is defined similarly but starting with the union of two horizontal rays. We'll assume that the cascades generated by distinct points $p \in \overline{\overline{\Pi}}$ do not contain common points. The union of cascades will be denoted by Π. The projection of Π on the λ axis will be denoted by Λ.

36

The set of points of Π that were obtained by intersecting the horizontal rays with the graphs of λ_e will be denoted by Π_h, the set of points of Π that were obtained by intersecting the vertical rays with the graphs of λ_e will be denoted by Π_v. Thus $\Pi = \overline{\overline{\Pi}} \cup \Pi_h \cup \Pi_v$.

Lemma 2.4.1. *Under the above assumptions on the functions λ_e, the number of points in Π is finite.*

Proof. We assumed that the number of cascades (which is the number of points in $\overline{\overline{\Pi}}$) was finite. Each cascade may be viewed as a tree-like graph rooted at a point of the first generation. The edges of the graph are formed by connecting points of k-th generation with the corresponding points of $k+1$-st generation via horizontal or vertical segments. If a vertical ray emanating from a point of k-th generation does not cross any of the graphs of λ_e, then the ray itself is also considered a semi-infinite edge of the graph.

By a path we'll mean a sequence of edges connecting a point of the first generation with a corresponding point of the second generation, then with a point of the third generation, etc.; the last edge of the path may be semi-infinite.

Let us show that each path terminates with a semi-infinite edge (i.e., there can't be paths with infinite number of edges) and that the number of paths is finite. This will imply that the number of points in a cascade is finite since each point of a cascade belongs to a path.

Since the functions λ_e are continuous on $\overline{I}_k$ (each function is defined on a union of some of the segments), there is a constant K such that $\sup \lambda_e < K$ for all e (the

supremum is taken over the domain where λ_e is defined). For a point $p \in \Pi_2 \cup \Pi_3$,

let $S_p(\delta)$ denote the square in the (z, λ) plane with side δ and centered at p. For

each $p = (z', \lambda') \in \Pi_2 \cup \Pi_3$ we can take $\delta > 0$ such that:

(a) $\lambda_e(z)$ is monotonic on $[z' - \delta/2, z']$ whenever e is such that $\lim_{z \uparrow z'} \lambda_e(z) = \lambda'$,

(b) $\lambda_e(z) \notin [\lambda' - \delta/2, \lambda' + \delta/2]$ for $z \in [z' - \delta/2, z']$ whenever e is such that
$\lim_{z \uparrow z'} \lambda_e(z) \neq \lambda'$,

(c) $\lambda_e(z)$ is monotonic on $[z', z' + \delta/2]$ whenever e is such that $\lim_{z \downarrow z'} \lambda_e(z) = \lambda'$,

(d) $\lambda_e(z) \notin [\lambda' - \delta/2, \lambda' + \delta/2]$ for $z \in [z', z' + \delta/2]$ whenever e is such that
$\lim_{z \downarrow z'} \lambda_e(z) \neq \lambda'$,

(e) $S_p(\delta)$ does not contain other points of $\Pi_2 \cup \Pi_3$.

Since the number of points in $\Pi_2 \cup \Pi_3$ is finite, we can take the same δ for

all $p \in \Pi_2 \cup \Pi_3$. Moreover, since all the points of $\Pi_2 \cup \Pi_3$ are at different heights

(different values of λ), we can choose δ sufficiently small so that $[\lambda' - \delta/2, \lambda' + \delta/2]$

does not overlap with $[\lambda'' - \delta/2, \lambda'' + \delta/2]$, where λ' and λ'' are the λ-coordinates of

$p', p'' \in \Pi_2 \cup \Pi_3$.

The purpose of this construction was to cut out small neighborhoods of all

the points of $\Pi_2 \cup \Pi_3$. Since all the functions λ_e are continuous and intersect only

at the points of $\Pi_2 \cup \Pi_3$, there is a positive constant κ such that a vertical edge

starting outside $\cup_p S_p(\delta)$ is longer than κ. Also note that a path that enters $S_p(\delta)$

will leave it after going along at most two vertical edges due to the monotonicity of

the functions λ_e that was a part of our construction (see Figure 2.5).

The next time the path returns to $S_p(\delta)$ (if it returns at all), it will be higher by

at least κ than when it left $S_p(\delta)$. Therefore, each path will have at most $2(1 + \delta/\kappa)$
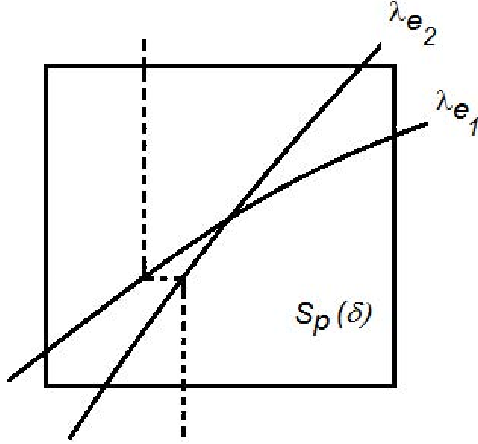
Figure 2.5: A path passing through $S_p(\delta)$ - the dotted line represents a part of the path.

vertical segments that intersect with $S_p(\delta)$. Let n be the number of points in $\Pi_2 \cup \Pi_3$. Then a path can have at most $2(1 + \delta/\kappa)n$ vertical segments that intersect with $\cup_p S_p(\delta)$. All the other vertical segments are longer than κ. Since $\sup \lambda_e < K$ for all e, there are no vertical segments that start above the level $\lambda = K$. Therefore the number of vertical segments in a path is bounded from above by $2(1 + \delta/\kappa)n + K/\kappa$.

It is now easy to show that the total number of paths is also finite. Since there is a finite number of functions λ_e and each is piece-wise monotonic, a given horizontal edge can be followed by only a finite number of different vertical edges (not exceeding the number of monotonicity intervals for all the functions λ_e, which will be denoted by L). A given vertical edge can be followed by only a finite number of different horizontal edges (not exceeding the number of different functions λ_e, which is bounded from above by L). Therefore, the total number of paths is bounded from above by $L^{2[2(1+\delta/\kappa)n+K/\kappa]}$, which completes the proof. $\qquad \square$

2.5 Inductive construction and the main result

Recall that $\Lambda = \{\lambda_0, ..., \lambda_N\}$ is the set of special time scales constructed in the previous section. We list them in the increasing order: $0 = \lambda_0 < ... < \lambda_N$. Our main result is the following.

Theorem 2.5.1. *For each $\lambda \in (0, \infty) \setminus \Lambda$ there are limits*

$$u^i(\lambda) = \lim_{\varepsilon \downarrow 0} u^\varepsilon(T^\varepsilon(\lambda), x), \quad x \in D_i, \tag{2.33}$$

where the functions u^i are continuous on $(0, \infty) \setminus \Lambda$.

First, we'll identify the limits for $\lambda \in (0, \lambda_1)$. In particular, we'll get the limits $\lim_{\lambda \uparrow \lambda_1} u^i(\lambda)$, which will be denoted by $u^i_-(\lambda_1)$. Using those, we'll determine $u^i_+(\lambda_1) = \lim_{\lambda \downarrow \lambda_1} u^i(\lambda)$, which, in turn, will allow us to define the limits $u^i(\lambda)$ for $\lambda \in (\lambda_1, \lambda_2)$. The procedure will then continue by induction till we reach the interval (λ_N, ∞).

It will be clear from our construction that $u^i(\lambda)$ do not depend on λ for $\lambda \in (\lambda_N, \infty)$. Let us denote these values by $\overline{u}^i$. Fix a value $\overline{\lambda} > \lambda_N$, and observe that on the interval $[T^\varepsilon(\overline{\lambda}), \infty)$ the function u^ε satisfies the equation that is almost linear, i.e., the diffusion matrix in (2.1) is close to $a(x, \overline{u}^i)$ for $x \in D_i$. We can consider equation (2.1) on the interval $[T^\varepsilon(\overline{\lambda}), \infty)$ with initial data that is close to $\overline{u}^i$ on D_i.

It is not difficult to show that, as in the linear case, the solution tends to a constant in the time scale $T^\varepsilon(\lambda)$ if λ is sufficiently large. This implies that $\overline{u}^i$ do not depend on i. Therefore, we have the following corollary.

Corollary 2.5.2. *There is a constant $\bar{u}$ such that for each $\lambda > \lambda_N$ and each $x \in \mathbb{R}^d$ there is the limit*

$$\lim_{\varepsilon \downarrow 0} u^\varepsilon(T^\varepsilon(\lambda), x) = \bar{u}.$$

Strictly speaking, the fact that $u^i(\lambda)$ do not depend on i implies the statement for $x \in \cup_i D_i$, but then it is not difficult to extend it to $x \in \mathbb{R}^d \setminus \cup_i D_i$ since the process $X_s^{t,x,\varepsilon}$ starting at such x will reach the interior of one of the domains D_i in sub-exponential time.

The construction below generalizes the situation with two equilibrium points considered in Section 2.3. Our emphasis is on providing the algorithm for constructing the functions u^i found in the left hand side of (2.33). The fact that the limits in (2.33) coincide with the constructed functions can be justified similarly to the case of two equilibrium points (with additional notational complications due to the induction), and so this part of the proof will not be repeated.

Preliminary description of the graphs G_n and G_n^*.

As a part of the inductive construction, we'll need to introduce a sequence of directed labeled graphs, $G_0, G_1, ..., G_N$ associated to the segments $(0, \lambda_1)$, $(\lambda_1, \lambda_2), ...,$ (λ_N, ∞), and a sequence of directed labeled graphs $G_1^*, ..., G_N^*$ associated to the points $\lambda_1, ..., \lambda_N$. They will be constructed inductively: G_n^*, $n = 1, ..., N$, will be determined by G_{n-1} and the values of $u_-^i(\lambda_n)$, while G_n will be determined by G_n^* and the values of $u_-^i(\lambda_n)$.

The vertices of the graphs are the equilibrium points $O_1, ..., O_r$. The directed edges are labeled as either active, engaged (to the right or to the left), or conflicted. We'll separate each graph into classes of equivalence (clusters) by saying that O_i and O_j are in the same equivalence class if they are connected by a chain of active edges to the same equilibrium (each equilibrium is considered to be connected to itself by an empty chain of edges). As a part of the inductive construction, we'll see that each graph has only three types of clusters:

(a) The only edges leading out of the points of a cluster are the active edges. In this case we'll say that this is a sleeping cluster.

(b) There is one equilibrium O_i within a cluster with an engaged edge e leading out of O_i. This edge leads to a point outside the cluster. The rest of the edges leading out of the points of the cluster are active. The vertex O_i can be reached from any of the points of the cluster by following a chain of active edges. In this case we'll say that this is an engaged cluster.

(c) There is a cycle C within a cluster such that there are two conflicted edges leading out of C. These edges lead to two distinct points outside the cluster. The rest of the edges leading out of the points of the cluster are active. The cycle C can be reached from any of the points of the cluster by following a chain of active edges. In this case we'll say that we have a conflicted cluster.

We'll say that a vertex O_i is subordinated to O_j if we can reach O_j from O_i by following a chain of active edges.

Let us briefly discuss the properties of the graphs G_n (which will follow from our inductive construction). Let $\lambda \in (\lambda_n, \lambda_{n+1})$, $0 \le n \le N$, where we put $\lambda_0 = 0$

42

and $\lambda_{N+1} = \infty$. Assume that $u^i(\lambda)$ has been defined. Consider the hierarchy of cycles corresponding to the linear equation (2.7) with the second argument in σ replaced by $u^i(\lambda)$ for $x \in D_i$. We can view this hierarchy as a directed graph, which will be denoted by $\overline{G}_n$. It will be seen that $\overline{G}_n$ does not depend on the choice of $\lambda \in (\lambda_n, \lambda_{n+1})$. The values of $u^i(\lambda)$ may in general be different for different i, and therefore the hierarchy $\overline{G}_n$ may be different from either of the hierarchies H_k constructed earlier. The notion of a next equilibrium for some of the cycles of the hierarchy $\overline{G}_n$ may be not correctly defined. Namely, it may happen that for a given cycle C there are two edges e and g connecting C with the equilibria O_i and O_j such that $\lambda_e = \lambda_g$, where λ_e and λ_g are the transition scales.

We'll say that an edge $e \in \overline{G}_n$ is active if $e = \mathcal{R}(C)$ for some C (the edge leading out of C), $\mathcal{N}(C)$ is defined uniquely and satisfies $\lambda_e < \lambda$. We'll say that e and g are conflicted if $e = g = \mathcal{R}(C)$ (i.e., the notion of next is not uniquely defined) and $\lambda_e = \lambda_g < \lambda$. We'll say that e is engaged to the right (engaged to the left) if $e = \mathcal{R}(C)$, $\mathcal{N}(C)$ is uniquely defined, $\lambda_e = \lambda$, and e connects O_i to O_j with $u^i(\lambda) < u^j(\lambda)$ ($u^i(\lambda) > u^j(\lambda)$). We'll say that e is passive if $\lambda_e > \lambda$. We'll see that each of the edges belongs to one of these four types and the type does not depend on $\lambda \in (\lambda_n, \lambda_{n+1})$. We define G_n to contain those edges of $\overline{G}_n$ that are either active, conflicted, or engaged (i.e., we dispose of passive edges).

We'll see that for $\lambda \in (\lambda_n, \lambda_{n+1})$ the values $u^i(\lambda)$ do not depend on i for all O_i within a given cluster. For a sleeping cluster, we'll see that $u^i(\lambda)$ do not depend on $\lambda \in (\lambda_n, \lambda_{n+1})$ for O_i that belong to the cluster. Moreover, for a sleeping cluster $u^i(\lambda)$ is equal to the projection of some point $p \in (\Pi_0 \cup \Pi_2 \cup \Pi_h) \cap \{(z, \lambda) : \lambda < \lambda_n\}$

43

onto the z-axis.

For an engaged cluster, there is one equilibrium O_i within a cluster with an engaged edge e leading out of O_i. We'll see that $\lambda_e(u^i(\lambda)) = \lambda$, while $u^j(\lambda) = u^i(\lambda)$ for the other points O_j within the cluster.

For a conflicted cluster, there is a cycle C within a cluster such that there are two conflicted edges e and g leading out of C to some vertices O_{j_1} and O_{j_2}, respectively. We'll see that the values of $u^i(\lambda)$ do not depend on λ for O_i that are within the cluster. These values are equal to $z_k \in Z$ for some k, while $u^{j_1}(\lambda) < z_k < u^{j_2}(\lambda)$ for $\lambda \in (\lambda_n, \lambda_{n+1})$. Moreover, C is a cycle in H_k for which the notion of "next" changes at $z = z_k$.

Inductive construction.

Now we formulate the basis of the induction process. We put $u_+^i(\lambda_0) = g(O_i)$. By Lemma 2.2.2, (2.31) holds for $\lambda \in (0, \delta)$ with the right hand side equal to $g(O_i)$ for some positive δ, which explains our definition of $u_+^i(\lambda_0)$. We put G_0 to be the graph with the vertices at $O_1, ..., O_r$ and with no edges. It is clear that this trivial graph has all the properties that we claimed should hold for the graphs G_n.

The inductive step will consist of three parts.

Part I. First, we define $u^i(\lambda)$ for $\lambda \in (\lambda_n, \lambda_{n+1})$ while assuming that we know G_n and $u_+^i(\lambda_n)$ for all i. Namely, consider first O_i such that there is an engaged edge e between O_i and O_j with $\mathcal{E}(C) = O_i$ and $\mathcal{N}(C) = O_j$. Assume that e is engaged to the right. It will follow from the inductive construction of G_n that $\lambda_e(z)$ is defined for z immediately to the right of $u_+^i(\lambda_n)$ and is strictly increasing for

44

$z \in (u^i_+(\lambda_n), u^i_+(\lambda_n) + \delta)$ for some positive δ. We can therefore define $u^i(\lambda)$ for $\lambda \in (\lambda_n, \lambda_n + \delta')$ for some $\delta' > 0$ as the inverse function to $\lambda_e(z)$, $z \in (u^i_+(\lambda_n), u^i_+(\lambda_n) + \delta)$.

Let's examine how large δ' can be chosen. By the definition of the sets Λ_1 and Λ_3, the function $\lambda_e(z)$ will keep increasing till it reaches the value λ_{n+1}. This means that $u^i(\lambda)$ can be defined for $\lambda \in (\lambda_n, \lambda_{n+1})$. The case when e is engaged to the left is only different in that we need to look at values of $\lambda_e(z)$ for z to the left of $u^i_+(\lambda_n)$.

If there is no engaged edge starting at O_i but O_i is subordinated to O_j with an engaged edge leading out of O_j, then we define $u^i(\lambda) = u^j(\lambda)$ for $\lambda \in (\lambda_n, \lambda_{n+1})$, where the right hand side has been defined in the previous step. When we discussed the structure of the clusters, we noted that there is at most one such O_j for a given O_i, so this is a correct definition. If O_i does not have an engaged edge starting from it and is not subordinated to any O_j with an engaged edge starting at O_j, then we define $u^i(\lambda) = u^i_+(\lambda)$ for $\lambda \in (\lambda_n, \lambda_{n+1})$. To sum up this construction, the values of $u^i(\lambda)$ are defined to be the same within a cluster. They depend on $\lambda \in (\lambda_n, \lambda_{n+1})$ if the cluster has a vertex with an engaged edge coming out of it, otherwise the values $u^i(\lambda)$ are constant in $\lambda \in (\lambda_n, \lambda_{n+1})$ and are equal to $u^i_+(\lambda_n)$.

Finally, we define $u^i_-(\lambda_{n+1}) = \lim_{\lambda \uparrow \lambda_{n+1}} u^i(\lambda)$ for all i. The first part of our inductive construction is complete.

Part II. Now let's define G^*_n while assuming that we know G_{n-1} and $u^i_-(\lambda_n)$. The description will depend on the type of point λ_n is (i.e., whether it belongs to $\Lambda_1, \Lambda_2, \Lambda_3$, or is obtained as a part of the cascade) and what type of an edge is affected.

Recall that G_{n-1} can be separated into clusters, which will be called $\Gamma_1, \Gamma_2, ..., \Gamma_M$.

Each sleeping cluster has no engaged edges leading out of it, each engaged cluster has one engaged edge leading out of it, each conflicted cluster has two conflicted edges leading out of it.

Case 1. Suppose that $\lambda_n \in \Lambda_1$. Then we define $G_n^* = G_{n-1}$.

Case 2. Now consider the case $\lambda_n \in \Lambda_2$. This means that there is $z \in I_k$ for some k and $e, g \in H_k$ such that $\lambda_e(z) = \lambda_g(z)$, with $e \neq g$. Different scenarios are possible:

1) Suppose that e and g are engaged edges leading out of engaged clusters (say Γ_{m_1} and Γ_{m_2}, respectively) such that $u_-^i(\lambda_n) = z$ for $O_i \in \Gamma_{m_1} \cup \Gamma_{m_2}$. Suppose also that e leads to a vertex in Γ_{m_2}, while g leads to a vertex in Γ_{m_1}. Then both e and g become active for G_n^* (i.e., the clusters Γ_{m_1} and Γ_{m_2} merge forming a sleeping cluster).

2) Suppose that everything is as above, except that e leads to a vertex in Γ_{m_2}, while g leads to a vertex outside Γ_{m_1}. In this case only e becomes active, while g stays engaged (i.e., the clusters merge forming an engaged cluster).

3) Suppose that e is engaged leading out of an engaged cluster Γ_m, while g is an active edge within Γ_m leading out of O_j. Suppose that $u_-^i(\lambda_n) = z$ for $O_i \in \Gamma_m$. In this case g becomes engaged, while the other edges are unaffected (i.e., the cluster Γ_m falls apart in two engaged clusters). It becomes engaged to the right if λ_g is increasing at z and engaged to the left if λ_g is decreasing at z.

4) If neither of the scenarios 1)-3) takes place, then we define $G_n^* = G_{n-1}$.

Case 3. Now consider the case $\lambda_n \in \Lambda_3$. This means that there is a cycle C in the hierarchy H_k such that the notion of "next" is not defined uniquely for it

at $z = z_k$ and $(z_k, \lambda_n) \in \Pi_3$. Let us assume that $u^i_-(\lambda_n) = z_k$ for the equilibria O_i within C, otherwise we define $G^*_n = G_{n-1}$. Again, we go through different scenarios in all of which we assume that C is the cycle for which the notion of "next" changes at z_k.

1) Suppose that C was a part of an engaged cluster Γ and e is the edge leading out of C for H_k, i.e., $R_k(C) = e$. Let g be the edge leading out of C for H_{k+1} to an equilibrium O_j, i.e., $R_{k+1}(C) = g$, $N_{k+1}(C) = O_j$. Suppose that Γ is engaged to the right for $\lambda \in (\lambda_{n-1}, \lambda_n)$ (the case of Γ engaged to the left is treated similarly).

If $u^j_-(\lambda_n) < z_k$ (which implies that $O_j \notin \Gamma$), then the cycle C together with all the vertices subordinated to it forms a new cluster that is conflicted for G^*_n. The remaining part of the cluster Γ stays engaged to the right. The difference between G_{n-1} and G^*_n is that the edge g is present in G^*_n (the edges e and g are conflicted), while only e is present in G_{n-1} (either engaged or active).

If $u^j_-(\lambda_n) > z_k$ (which implies that $O_j \notin \Gamma$), then a new cluster forms consisting of C and all the vertices subordinated to it. The cluster is engaged to the right via the edge g. The remaining part of the cluster Γ stays engaged to the right, while the edge e disappears.

Suppose that $u^j_-(\lambda_n) = z_k$. In this case O_j must belong to Γ as follows from the earlier assumptions on the function λ_e. If g was not represented in G_{n-1}, then it gets added to the cluster Γ, which remains engaged to the right for G^*_n (g remains in the cluster if it was represented in G_{n-1}). Whether e remains represented in G^*_n depends on whether z_k corresponds to the restructuring of the first, second, or third kind. Namely, in the cases of restructuring of the first and third kind, e remains in

Γ if and only if O_j is subordinated to C in G_{n-1}. In the case of restructuring of the second kind, e is not represented in G_n^*.

2) Suppose that C was a part of a conflicted cluster Γ. Suppose that there are two conflicted edges e and g leading out of C to some vertices O_{j_1} and O_{j_2}, respectively, such that $u^{j_1}(\lambda) < z_k < u^{j_2}(\lambda)$ for $\lambda \in (\lambda_{n-1}, \lambda_n)$. Suppose that O_{j_1} belongs to a cluster Γ_{m_1} and O_{j_2} belongs to a cluster Γ_{m_2} for the graph G_{n-1}. Suppose that Γ_{m_1} is engaged to the right via an edge h and $u_-^i(\lambda_n) = z_k$ for $O_i \in \Gamma_{m_1}$ (the case when Γ_{m_2} is engaged to the left and $u_-^i(\lambda_n) = z_k$ for $O_i \in \Gamma_{m_2}$ is treated similarly).

In this case the clusters Γ_{m_1} and Γ from the graph G_{n-1} become one cluster for the graph G_n^*. If h leads to a vertex within the cluster Γ, then e and h becomes active edges within the larger cluster $\Gamma \cup \Gamma_{m_1}$. The cluster $\Gamma \cup \Gamma_{m_1}$ becomes engaged to the right via the edge g in the graph G_n^*. If h leads to a vertex outside the cluster Γ, then it stays an engaged edge within the larger cluster, g disappears, and e becomes an active edge within $\Gamma \cup \Gamma_{m_1}$.

Suppose now that $u_-^i(\lambda_n) < z_k$ for $O_i \in \Gamma_{m_1}$ and $u_-^i(\lambda_n) > z_k$ for $O_i \in \Gamma_{m_2}$. Suppose also that there is a cluster Γ_{m_3} in G_{n-1} that is engaged via an edge h to a vertex in Γ and is such that $u_-^i(\lambda_n) = z_k$ for $O_i \in \Gamma_{m_3}$. In this case the clusters Γ_{m_3} and Γ from the graph G_{n-1} become one cluster for the graph G_n^*. The edge h becomes active and the cluster $\Gamma \cup \Gamma_{m_1}$ is conflicted in G_n^*.

3) If neither of the scenarios 1)-2) takes place, then we define $G_n^* = G_{n-1}$.

Case 4. Finally consider the case when λ_n was obtained as a part of a cascade. Namely, assume that there is $z \in I_k$ such that (z, λ_n) is obtained as the intersection

of a vertical ray with the graph of $\lambda_e(z)$, where e is an edge represented in H_k.

Suppose that G_{n-1} contains a sleeping cluster Γ such that $u^i_-(\lambda_n) = z$ for $O_i \in \Gamma$.

1) Suppose that the edge e (represented in H_k, but not in G_{n-1}) leads from a vertex in Γ to a vertex O_j. If $u^j_-(\lambda_n) = z$ (in which case O_j belongs to Γ), then e is added as an active edge, and Γ with this extra edge remains a sleeping cluster within G^*_n. If $u^j_-(\lambda_n) < z$ (or $u^j_-(\lambda_n) > z$), then the edge e is added as an engaged edge and Γ becomes engaged to the left (right) in G^*_n.

2) Suppose that G_{n-1} contains a cluster Γ_m distinct from Γ such that $u^i_-(\lambda_n) = z$ for $O_i \in \Gamma_m$. Suppose that Γ_m is engaged to a vertex in Γ via an edge h. Then Γ_m and Γ from the graph G_{n-1} become one cluster for the graph G^*_n. The edge h becomes active and the cluster $\Gamma \cup \Gamma_{m_1}$ is sleeping in G^*_n.

3) If neither of the scenarios 1)-2) takes place, then we define $G^*_n = G_{n-1}$.

Part III. Now let's define G_n and $u^i_+(\lambda_n)$ while assuming that we know G^*_n and $u^i_-(\lambda_n)$. Let Γ be a cluster within G^*_n such that $u^i_-(\lambda_n) = z^*$ for $O_i \in \Gamma$. Suppose that Γ is engaged via an edge e.

From our construction above it follows that G^*_n contains at most once such cluster Γ with the following property: either Γ is engaged to the right and $\lambda_e(z) < \lambda_n$ for $z \in (z^*, z^* + \delta)$ for some positive δ or Γ is engaged to the left and $\lambda_e(z) < \lambda_n$ for $z \in (z^* - \delta, z^*)$ for some positive δ. In this case we'll say that Γ is a "special" cluster. If G^*_n does not contain a special cluster, then we define $G_n = G^*_n$ and $u^i_+(\lambda_n) = u^i_-(\lambda_n)$ for all i.

Assume that G^*_n contains a special cluster and that it is engaged to the right (the case when it is engaged to the left is treated similarly). Let us project the set

49

Π on z-axis and list those of the points of the projection that are to the right of z^* in the increasing order, denoting them by $\bar{z}_1, ..., \bar{z}_L$.

We'll describe the transition from G_n^* to G_n using a two-part induction procedure, the first part being induction on $l = 1, ..., L$ (the second part of the induction will be explained below). At each step of induction we'll have a new "transitional" graph separated into clusters - the first graph will coincide with G_n^* and the last one will coincide with G_n. For each cluster Γ within a transitional graph, we'll define the value $u(\Gamma)$. The first set of values $u(\Gamma)$ is just $u_-^i(\lambda_n)$ for $O_i \in \Gamma$.

Let Γ now denote the special cluster within G_n^*. At the fist step of induction we change the value of $u(\Gamma)$ to $\bar{z}_1$. We need to examine the following scenarios to describe the transitional graph that replaces G_n^*.

1)**A previously engaged cluster Γ_{m1} attached to the special clusterΓ.** Suppose that G_n^* contains an engaged cluster Γ_{m_1} with $u(\Gamma_{m_1}) = \bar{z}_1$ such that Γ_{m_1} is engaged to a vertex in Γ via an edge h. Suppose that e leads to a vertex in a cluster Γ_{m_2} with $u(\Gamma_{m_2}) > \bar{z}_1$. Then Γ_{m_1} and Γ form one cluster in the transitional graph and the edge h becomes active in this cluster. This larger cluster is special for the transitional graph.

2)**The special cluster Γ merges with a previously engaged cluster Γ_m into a sleeping cluster.** Suppose that G_n^* contains an engaged cluster Γ_m with $u(\Gamma_m) = \bar{z}_1$ such that Γ_m is engaged to a vertex in Γ via an edge h. Suppose that e leads to a vertex in Γ_m. Then Γ_m and Γ form one sleeping cluster in the transitional graph and the edges e and h become active in this cluster. The new cluster does not have the "special" property (it is not even engaged).

50

3) **The special cluster Γ becomes engaged after attaching to an engaged cluster Γ_m; namely,** Suppose that G_n^* contains an engaged cluster Γ_m with $u(\Gamma_m) = \bar{z}_1$ such that Γ_m is engaged to a vertex outside Γ via an edge h. Suppose that e leads to a vertex in Γ_m. Then Γ_m and Γ form one engaged cluster in the transitional graph, which is engaged via the edge h. The edge e becomes active in this cluster. The new cluster does not have the "special" property.

4) **A conflicted cluster Γ_m merges with the special cluster Γ to form another special cluster** Suppose that G_n^* contains a conflicted cluster Γ_m with $u(\Gamma_m) = \bar{z}_1$ (in which case $\bar{z}_1$ coincides with one of the points z_k) such that Γ_m has conflicted edges g and h. Suppose that g leads to a vertex in Γ. Suppose that e leads to a vertex in Γ_m. Then Γ_m and Γ form one cluster in the transitional graph, the edge h becomes engaged to the right, while e and g become active in this cluster. The new cluster still has the "special" property.

5) **The special cluster Γ attaches to a conflicting cluster Γ_m; namely,** Suppose that G_n^* contains a conflicted cluster Γ_m with $u(\Gamma_m) = \bar{z}_1$ (in which case $\bar{z}_1$ coincides with one of the points z_k) such that Γ_m has conflicted edges g and h. Suppose that neither g nor h lead to a vertex in Γ. Suppose that e leads to a vertex in Γ_m. Then Γ_m and Γ form one conflicted cluster in the transitional graph, where the edge e becomes active. The new cluster does not have the "special" property.

6) **The special cluster Γ joins a sleeping cluster Γ_m; namely,** Suppose that G_n^* contains a sleeping cluster Γ_m with $u(\Gamma_m) = \bar{z}_1$. Suppose that e leads to a vertex in Γ_m. Then Γ_m and Γ form one sleeping cluster in the transitional graph and the edge e becomes active in this cluster. The new cluster does not have the

"special" property.

7) **A sub cycle Γ_m of the special cluster Γ becomes engaged; namely,**

Suppose that Γ contains a vertex O_i such that an edge g leading out of it (whether active or engaged) has the property that $\lambda_g(\bar{z}_1) = \lambda_n$. Then O_j and all the vertices of Γ subordinated to it form a new cluster that is engaged to the right via the edge g but is not special. The remaining part of Γ (if it is non-empty) forms an engaged special cluster in the transitional graph.

8) Suppose that $\bar{z}_1$ coincides with one of the points z_k. Suppose that Γ contains a cycle C that is present in both H_k and H_{k+1} for which the notion of "next" changes at z_k.

Suppose that e' is the edge leading out of C for H_k, i.e., $R_k(C) = e'$. Let g be the edge leading out of C for H_{k+1} to an equilibrium O_j, i.e., $R_{k+1}(C) = g$, $N_{k+1}(C) = O_j$.

A sub cycle of Γ becomes conflicted. If $O_j \in \Gamma_m$ for some $\Gamma_m \neq \Gamma$ with $u(\Gamma_m) < \bar{z}_1$, then the cycle C together with all the vertices subordinated to it forms a new cluster that is conflicted for the transitional graph. The remaining part of the cluster Γ stays engaged to the right. The difference between G_n^* and the transitional graph is that the edge g is present in the transitional graph (the edges e' and g are conflicted), while only e' is present in G_n^*. There is at most one special cluster in the transitional graph.

If $O_j \in \Gamma_m$ for some $\Gamma_m \neq \Gamma$ with $u(\Gamma_m) > \bar{z}_1$, then a new cluster forms consisting of C and all the vertices subordinated to it. The cluster is engaged to the right via the edge g. The remaining part of the cluster Γ stays engaged to the

52

right, while the edge e' disappears. There are now up to two special clusters in the transitional graph.

Suppose that $O_j \in \Gamma$. If g was not represented in G_n^*, then it gets added to the cluster Γ, which remains engaged to the right for the transitional graph (g remains in the cluster if it was represented in G_n^*). Whether e' remains represented in the transitional graph depends on whether z_k corresponds to the restructuring of the first, second, or third kind. Namely, in the cases of restructuring of the first and third kind, e' remains in Γ if and only if O_j is subordinated to C in G_n^*. In the case of restructuring of the second kind, e' is not represented in the transitional graph.

If neither of the scenarios described above takes place, then the transitional graph coincides with G_n^*. This completes our description of the transitional graph corresponding to $\overline{z}_1$. Now we replace G_n^* by the transitional graph. If it has no special clusters, the inductive procedure stops. If it has one special cluster, we repeat the above construction with $\overline{z}_1$ replaced by $\overline{z}_2$. If it has two special clusters, we repeat the construction applied to one of the clusters with the other special cluster temporarily fixed. In this way, we proceed from $\overline{z}_2$ to $\overline{z}_3$, etc., until we reach $\overline{z}_L$. Observe that there are no special clusters with the value $\overline{z}_L$.

The second part of the inductive construction concerns the fact that once we reach $\overline{z}_L$, the resulting graph may still have special clusters. The values $u(\Gamma)$ for those clusters belong to the set $\{\overline{z}_1, ..., \overline{z}_L\}$, i.e., are to the right of z^*. We can now apply the above construction to each of the clusters successively, thus making sure that a transitional graph only has clusters with the values of $u(\Gamma)$ in the set $\{\overline{z}_2, ..., \overline{z}_L\}$. Continuing by induction on l, we can get rid of all the special clusters.

The last transitional graph will be denoted by G_{n+1} and the values $u(\Gamma)$ for this graph give us $u_+^i(\lambda_n)$.

Now, that our inductive construction is complete, let us make an observation concerning the values of $u^i(\lambda)$ for $\lambda \in (\lambda_N, \infty)$. If the graph G_N contains an engaged cluster, then there is an equilibrium O_i within the cluster with an engaged edge e leading out of O_i such that $\lambda_e(u^i(\lambda)) = \lambda$. This implies that $\sup \lambda_e > \lambda_N$. Therefore, either λ_e has a maximum that is larger than λ_N or the graph of λ_e intersects one of the vertical lines $z = z_k$ at a point that is higher than λ_N. In either of these cases, λ_N can't be the maximum of the set Λ due to our definition of the sets Λ_1 and Λ_3. We conclude that G_N can't contain engaged clusters. For all the clusters that are sleeping or conflicted, the values $u^i(\lambda)$ do not depend on $\lambda \in (\lambda_N, \infty)$. Therefore, $u^i(\lambda) = \overline{u}^i$ for some constants $\overline{u}^i$, which justifies the discussion preceding Corollary 2.5.2. Also, by Corollary 2.5.2, the values $\overline{u}^i$ do not depend on i, which implies that G_N contains a single cluster.

Chapter 3:   Random perturbations of dynamical systems with reflecting boundary and corresponding PDE with a small parameter

3.1   Introduction

Consider the following parabolic initial-boundary value problem with a small parameter $\varepsilon > 0$:

$$
\begin{cases}
\dfrac{\partial u^{\varepsilon}}{\partial t} = \mathcal{L}^{\varepsilon} u^{\varepsilon} \equiv \dfrac{\varepsilon^2}{2} \sum_{i,j=1}^{d} a_{ij}(x) \dfrac{\partial^2 u^{\varepsilon}}{\partial x_i \partial x_j} + \sum_{i=1}^{d} b^{i}(x) \dfrac{\partial u^{\varepsilon}}{\partial x_i} \ , & \\[2mm]
u^{\varepsilon}(0,x) = g(x) \ , & x \in D \cup \partial D \ ; \\[2mm]
\dfrac{\partial u^{\varepsilon}}{\partial \gamma}(t,x) = 0 \ , & x \in \partial D \ , \ t \geq 0 \ .
\end{cases} \qquad (3.1)
$$

Here D is a d-dimensional bounded domain in $\mathbb{R}^d$ with smooth boundary ∂D. The initial condition $g(.)$ is smooth in $D \cup \partial D$. The matrix $a(x) = (a_{ij}(x))_{1 \leq i,j \leq d}$ is positive definite. The functions $a_{ij}(x)$ are smooth and uniformly bounded, with uniformly bounded derivatives. There is a constant $\theta > 0$ such that for any $\xi = (\xi^1, ..., \xi^d)$ we have $\theta^2 |\xi|^2_{\mathbb{R}^d} \leq \sum_{i,j=1}^{d} a_{ij}(x) \xi^i \xi^j \leq \theta^{-2} |\xi|^2_{\mathbb{R}^d}$. The vector field $b(x) = (b^1(x), ..., b^d(x))$ has terms which are uniformly bounded, smooth in $[D]$

(here and below $[D]$ is the closure of D in Euclidean metric), and have uniformly bounded derivatives. The vector field $\gamma(x) = (\gamma^1(x), ..., \gamma^d(x))$ is the inward *co-normal* unit vector field on ∂D with respect to the matrix $a^{-1}(x) \equiv (a^{ij}(x))_{1 \leq i,j \leq d} = (a_{ij}(x))_{1 \leq i,j \leq d}^{-1}$. That is to say, for any vector $v(x) = (v^1(x), ..., v^d(x))$ tangent to ∂D we have $(\gamma, v)_{a^{-1}(x)} \equiv \sum_{i,j=1}^{d} a^{ij}(x)\gamma^i(x)v^j(x) = 0$. (Here and below $(\gamma, v)_{a^{-1}(x)}$ is the inner product with respect to the matrix $a^{-1}(x)$. For a detailed discussion of the co-normal condition we refer to [13, Section 2.5]). We also have $|\gamma|_{\mathbb{R}^d} = 1$.

Let us assume that the vector field $b(x)$ is pointing outward to D on a connected subset $\partial_1 D$ of ∂D, and it is pointing inward on $\partial_2 D \equiv \partial D \backslash \partial_1 D$. (It is never tangent to ∂D.) Let $\bar{b}(x)$ be the field coinciding with $b(x)$ everywhere except at points of $\partial_1 D$. At these points $\bar{b}(x)$ is defined as the projection of $b(x)$ onto the direction of the boundary. Suppose the dynamical system $\dot{x}_t = \bar{b}(x_t)$ has some of its ω-limit sets on $\partial_1 D$. These ω-limit sets are points $O_1, ..., O_l$ $(l \geq 1)$.

Our goal in this paper is to describe the long-time behavior of the solution $u^\varepsilon(t, x)$ of (3.1) as $\varepsilon \to 0$ and $t \to \infty$. One can relate problem (3.1) with a certain diffusion process X_t^ε with small diffusion and reflection with respect to γ on ∂D. This process can be described as a solution of the following stochastic differential equation:

$$dX_t^{x,\varepsilon} = b(X_t^{x,\varepsilon})dt + \varepsilon\sigma(X_t^{x,\varepsilon})dW_t + \mathbf{1}_{\partial D}(X_t^{x,\varepsilon})\gamma(X_t^{x,\varepsilon})d\xi_t^{x,\varepsilon} \ , \ X_0^{x,\varepsilon} = x \ , \ \xi_0^{x,\varepsilon} = 0 \ .$$

$$(1.2)$$

56

Here $\sigma(x)$ is a $d \times d$ matrix with smooth terms (and bounded derivatives) that satisfies $\sigma(x)\sigma^T(x) = \sigma^T(x)\sigma(x) = a(x)$. The function $\mathbf{1}_{\partial D}(.)$ is the indicator function of ∂D. The processes X_t^ε and ξ_t^ε are continuous time stochastic processes, adapted to the filtration $(\mathcal{F}_t)_{t \geq 0}$. They satisfy the following assumptions with probability 1:

(1) The process $X_t^{x,\varepsilon} \in [D]$;

(2) The process $\xi_t^{x,\varepsilon}$ is non-decreasing in t and increases only at $\Delta = \{t \; ; \; X_t^{x,\varepsilon} \in \partial D\}$;

(3) The set Δ has Lebesgue measure zero.

Under these assumptions, it was proved in [1] (also see [14]) that such a pair of processes $(X_t^{x,\varepsilon}, \xi_t^{x,\varepsilon})$ exists and is unique (in the sense of probability 1). The process $\xi_t^{x,\varepsilon}$ is called *the local time* of the process $X_t^{x,\varepsilon}$ on ∂D. (We remark here that this notion of the local time for the multidimensional diffusion process extends the classical 1-dimensional local time in [15]. See [14] for a discussion based on SDE approach. For other discussions of the local time for multidimensional diffusion process we also refer to [16] and [17].) The process $X_t^{x,\varepsilon}$ is a strong Markov process in $[D]$ and it satisfies the Doeblin condition, which leads to the existence and uniqueness of an invariant measure in $[D]$.

It turns out that the solution $u^\varepsilon(t,x)$ of (3.1) can be represented as $u^\varepsilon(t,x) = \mathbf{E}_x g(X_t^{x,\varepsilon})$ (see Section 4 for details). Thus the asymptotic behavior of solution $u^\varepsilon(t,x)$ as $\varepsilon \to 0, t \to \infty$ is determined by the asymptotic behavior of the process $X_t^{x,\varepsilon}$. However, the latter can be calculated using the Freidlin-Wentzell large

deviation theory (see [4], [18]).

In Section 2 of the present paper we will give an expression of the action functional S_{0T}^+ of the process $X_t^{x,\varepsilon}$. By using the large deviation principle for the family of processes $\{X_t^{x,\varepsilon}\}_{\varepsilon>0}$ we will give a description of the asymptotic behavior of $X_t^{x,\varepsilon}$ in Section 3. Since the proof is based on the method of [4, Ch.6] and [18], we will only prove some key technical lemmas and sketch the result. In particular, we give the algorithm on the calculation of metastable states. Section 4 provides the corresponding result for problem (3.1). We point out that a related question for elliptic boundary value problems was already considered in [19] (also see [4, Section 10.3]). An example is given in Section 5.

3.2   Calculation of the action functional

In this section we give an expression of the action functional corresponding to the large deviation principle of the process $X_t^{x,\varepsilon}$. The main proofs and justifications of our results are contained in [1] (also see [4, Section 10.3]), so we just summarize the results we need.

In [1], the authors have constructed the process $(X_t^{x,\varepsilon}, \xi_t^{x,\varepsilon})$ corresponding to (1.2) by first realizing it in the space $\mathbb{R}_+^d$ using the following stochastic differential equation:

$$dY_t^\varepsilon = b(\Gamma(Y_t^\varepsilon))dt + \varepsilon\sigma(\Gamma(Y_t^\varepsilon))dW_t \ , \ Y_0^\varepsilon = x \in \mathbb{R}_+^d. \tag{3.1}$$

Here $\Gamma : \mathbf{C}_{[0,\infty)}(\mathbb{R}^d) \to \mathbf{C}_{[0,\infty)}(\mathbb{R}^d_+)$ is a functional defined by

$$\Gamma(\psi_t) \equiv (\Gamma(\psi))_t \equiv \Gamma_t(\psi) = (\psi_t^1 - 0 \wedge \inf_{0 \leq s \leq t} \psi_s^1, \psi_t^2, ..., \psi_t^d), \qquad (3.2)$$

for $\psi_t = (\psi_t^1, ..., \psi_t^d) \in \mathbf{C}_{[0,\infty)}(\mathbb{R}^d)$. It was proved in [1] that in the case of a half space $\mathbb{R}^d_+$, one can take $(X_t^\varepsilon, \xi_t^\varepsilon) = (\Gamma(Y_t^\varepsilon), (\Gamma(Y_t^\varepsilon) - Y_t^\varepsilon)^1)$.

In the general case when D is a bounded region in $\mathbb{R}^d$ with smooth boundary, one can take a finite covering of D by a set of open neighborhoods $\{\mathcal{U}_1, ..., \mathcal{U}_N\}$. Within each $\mathcal{U}_i$ $(i = 1, ..., N)$, the process can be constructed via a homeomorphism between $\mathcal{U}_i$ and $\mathbb{R}^d$, or between $\mathcal{U}_i \cap D$ and $\mathbb{R}^d_+$ (when $\mathcal{U}_i \cap \partial D \neq \emptyset$). In the latter case we use the construction of the process in half space as above. By appropriately "gluing" these pieces of the trajectories together one can construct the processes $(X_t^\varepsilon, \xi_t^\varepsilon)$. The process X_t^ε is the diffusion process with reflection in D and the process ξ_t^ε is the local time on ∂D. For details of this construction we refer to [1], [13, Section 1.6].

It was shown in [1, Section 1.2] that the corresponding action functional for the family of processes $\{X_t^\varepsilon\}_{\varepsilon > 0}$ as $\varepsilon \downarrow 0$ is given by the formula:

$$S_{0T}^+(\varphi) = \begin{cases} \dfrac{1}{2} \displaystyle\int_0^T \|\dot{\varphi}_s - b(\varphi_s) - \mathbf{1}_{\partial D}(\varphi_s)\omega(s)\gamma(\varphi_s)\|_{a^{-1}(\varphi_s)}^2 ds \,, \\ \qquad\qquad \text{for } \varphi \in \mathbf{C}_{[0,T]}([D]) \text{ absolutely continuous }, \ \varphi_0 = x \,; \\ +\infty \,, \qquad\qquad \text{for the rest of } \varphi \in \mathbf{C}_{[0,T]}([D]) \,. \end{cases}$$

$$(2.3)$$

Here

$$\omega(s) = \frac{(\dot{\varphi}_s - b(\varphi_s), \gamma(\varphi_s))_{a^{-1}(\varphi_s)}}{\|\gamma(\varphi_s)\|^2_{a^{-1}(\varphi_s)}} \vee 0 \ ,$$

and $\|v\|_{a^{-1}(x)} = (v,v)^{1/2}_{a^{-1}(x)}$, for vector $v \in \mathbb{R}^d$.

We have

$$\mathbf{1}_{\partial D}(\varphi_s)\omega(s)\gamma(\varphi_s)$$

$$= \mathbf{1}_{\partial D}(\varphi_s)\gamma(\varphi_s)\left(\frac{(\dot{\varphi}_s, \gamma(\varphi_s))_{a^{-1}(\varphi_s)}}{\|\gamma(\varphi_s)\|^2_{a^{-1}(\varphi_s)}} - \frac{(b(\varphi_s), \gamma(\varphi_s))_{a^{-1}(\varphi_s)}}{\|\gamma(\varphi_s)\|^2_{a^{-1}(\varphi_s)}}\right) \vee 0$$

$$= -\mathbf{1}_{\partial D}(\varphi_s)\frac{\gamma(\varphi_s)}{\|\gamma(\varphi_s)\|^2_{a^{-1}(\varphi_s)}}[0 \wedge (b(\varphi_s), \gamma(\varphi_s))_{a^{-1}(\varphi_s)}] \quad \text{for a.s. } s \in [0,T] \ .$$

Define

$$\bar{b}(x) = b(x) - \mathbf{1}_{\partial D}(x)\frac{\gamma(x)}{\|\gamma(x)\|^2_{a^{-1}(x)}}[0 \wedge (b(x), \gamma(x))_{a^{-1}(x)}] \ . \tag{2.4}$$

We see that $\bar{b}(x)$ is the field coinciding with $b(x)$ everywhere except at those

points of $\partial_1 D$. (Recall that $\partial_1 D$ is the part of the boundary ∂D on which $b(x)$ is

pointing outward). At these points $\bar{b}(x)$ is defined as the projection of $b(x)$ onto the

direction of the boundary. The action functional for the family of processes $\{X_t^\varepsilon\}_{\varepsilon>0}$

can now be formulated as

$$S_{0T}^+(\varphi) = \begin{cases} \dfrac{1}{2}\displaystyle\int_0^T \|\dot{\varphi}_s - \bar{b}(\varphi_s)\|^2_{a^{-1}(\varphi_s)}ds \ , \\ \qquad\qquad \text{for } \varphi \in \mathbf{C}_{[0,T]}([D]) \text{ absolutely continuous }, \varphi_0 = x \ ; \\ +\infty, \quad \text{for the rest of } \varphi \in C_{[0,T]}([D]). \end{cases} \tag{2.5}$$

The deterministic trajectory X_t^0 at which the above action functional is 0

is also calculated in [1]. It is given by the system $\dot{x}_t = \bar{b}(x_t)$, $x_0 = x$, where

$\bar{b}(x_t) = b(x)$ everywhere except at points of $\partial_1 D$, and at those points of $\partial_1 D$, $\bar{b}(x_t)$ is the projection of $b(x)$ onto the direction of the boundary.

We formulate below the large deviation principle for the family of processes $\{X_t^\varepsilon\}_{\varepsilon>0}$.

Theorem 2.1. (Large deviation principle) *For the process X_t^ε, we have*

(i) The set $\Phi(s) = \{\varphi \in \mathbf{C}_{[0,T]}([D]) : S_{0T}^+(\varphi) \leq s\}$ is compact for every $s \geq 0$;

(ii) Given $\varphi \in \mathbf{C}_{[0,T]}([D])$. For any $\delta > 0$ and any $\gamma > 0$ there exist an $\varepsilon_0 > 0$ such that for any $0 < \varepsilon < \varepsilon_0$ we have

$$\mathbf{P}\{\rho_{0T}(X^{x,\varepsilon}, \varphi) < \delta\} \geq \exp[-\varepsilon^{-2}(S_{0T}^+(\varphi) + \gamma)] , \qquad (2.6)$$

where $T > 0$ and $\rho_{0T}(.,.)$ denotes the uniform distance between functions in $\mathbf{C}_{[0,T]}([D])$;

(iii) For any $\delta, \gamma > 0$ and any $s > 0$ there exists an $\varepsilon_0 > 0$ such that for any $0 < \varepsilon < \varepsilon_0$ we have

$$\mathbf{P}\{\rho_{0T}(X^{x,\varepsilon}, \Phi(s)) \geq \delta\} \leq \exp[-\varepsilon^{-2}(s - \gamma)] , \qquad (2.7)$$

where $\rho_{0T}(\varphi, \Phi(s)) = \inf_{\psi \in \Phi(s)} \rho_{0T}(\varphi, \psi)$.

3.3 Asymptotic behavior of $X_t^{x,\varepsilon}$

3.3.1 Estimates on the time to converge to ω-limit sets on the boundary

We now begin our study of the asymptotic behavior of the process $X_t^{x,\varepsilon}$. First, since the dynamical system $\dot{x}_t = \bar{b}(x_t)$ does not have any ω - limit set within D, we shall expect that as ε is small, the trajectories of $X_t^{x,\varepsilon}$ come to the boundary $\partial_1 D$ within finite time. (Notice that at points of $\partial_1 D$ the vector field $b(x)$ is pointing outward and at points of $\partial_2 D$ it is pointing inward. Therefore the deterministic trajectory X_t^0 will not come to $\partial_2 D$.)

For any $x, y \in [D]$, we define

$$V^+(x,y) = \inf_{\varphi \in \mathbf{C}_{[0,T]}([D])} \{S_{0T}^+(\varphi), \varphi_0 = x, \varphi_T = y, \varphi_t \in D \cup \partial D, 0 \le t \le T < \infty\} .$$

Recall that the dynamical system $\dot{x}_t = \bar{b}(x_t)$ has all its ω-limit sets on $\partial_1 D$. These ω-limit sets are points $O_1, ..., O_l$ ($l \ge 1$). Let us suppose, that for any x and y in $[D]$, $x \ne y$ we have at least one of $V^+(x,y)$ and $V^+(y,x)$ being > 0.

For each $O_i, i = 1, 2, ..., l$, by an α-*neighborhood* $\mathcal{E}_\alpha(O_i)$ of O_i, we refer to the intersection of D with an open ball having center O_i and radius $\alpha > 0$. We use the symbol $\partial \mathcal{E}_\alpha(O_i)$ to mean the intersection of $[D]$ with the boundary of the open α-ball centered at O_i. We call $\partial \mathcal{E}_\alpha(O_i)$ the *boundary* of the α-neighborhood of O_i. Let us choose $\alpha > 0$ such that the α-neighborhoods $\mathcal{E}_\alpha(O_i)$ for all $O_i, i = 1, 2, ..., l$, does not intersect each other. We now prove the following:

Theorem 3.1. *There exist positive constants c and T_0 such that for all suffi-ciently small $\varepsilon > 0$ and any $x \in [D], X_0^{x,\varepsilon} = x$ we have*

$$\mathbf{P}_x\{\zeta_\alpha > T\} \le \exp[-\varepsilon^{-2}c(T - T_0)] ,$$

where $\zeta_\alpha = \inf\{t : X_t^{x,\varepsilon} \in \bigcup_{i=1}^{l}[\mathcal{E}_\alpha(O_i)]\}$.

Proof. We consider the dynamical system $\dot{x}_t = \bar{b}(x_t)$ on the whole domain $D \cup \partial D$, where vector field $\bar{b}(x)$ is defined as before. Since system $\dot{x}_t = \bar{b}(x_t)$ does not have any ω - limit set in D and $(b(x), \gamma(x))_{a^{-1}(x)}|_{\partial_2 D} > 0$, we can say that the time $T_1(x)$ that the trajectory $x_t(x)$ spends until reaching $\partial_1 D$ is finite (if $x \in \partial_1 D$, let $T_1(x) = 0$). Let $y(x)$ be the point where trajectory first hits $\partial_1 D$. Starting from $y(x)$, the time $T_2(y(x), \alpha) = T_2(x, \alpha)$ that the trajectory of system $\dot{x}_t = \bar{b}(x_t)$ on $\partial_1 D$ spend to come into $\bigcup_{i=1}^{l}[\mathcal{E}_{\frac{\alpha}{2}}(O_i)]$ is also finite (as is the same, if $y(x) \in \bigcup_{i=1}^{l}[\mathcal{E}_{\frac{\alpha}{2}}(O_i)]$, then $T_2(x, \alpha) = 0$). The function $T(x, \alpha) = T_1(x) + T_2(x, \alpha)$ is upper semi-continuous in x (i.e., for $x, x_0 \in [D]$ we have $\overline{\lim}_{x \to x_0} T(x, \alpha) \le T(x_0, \alpha)$) because $x_t(x)$ depends continuously on x. Thus there exists $T_0 = \max_{x \in [D]} T(x, \alpha) < \infty$. The set of functions in $\mathbf{C}_{[0,T_0]}([D])$ assuming their values in $[D] \backslash \left(\bigcup_{i=1}^{l} \mathcal{E}_{\frac{\alpha}{2}}(O_i) \right)$ is closed and thus $S_{0T_0}^+$ attains a minimum A on this set. Taking into account the construction of T_0 and the form of $S_{0T_0}^+$ in (2.5), we see that $A > 0$. Let $0 < \delta < \frac{\alpha}{2}$. Let $\Phi_x(A/2) = \{\varphi \in \mathbf{C}_{[0,T_0]}(D), \varphi_0 = x, S_{0T_0}^+(\varphi) \le A/2\}$. We see that trajectories for which $\zeta_\alpha > T_0$ are at a distance $\ge \delta$ from $\Phi_x(A/2)$. Thus by the part (iii) of the large deviation principle we have

$$\mathbf{P}_x\{\zeta_\alpha > T_0\} \le \exp[-\varepsilon^{-2}(A/2 - \gamma)]$$

63

for some $0 < \gamma < A/2$.

Thus by strong Markov property,

$$\begin{aligned}
\mathbf{P}_x\{\zeta_\alpha > (n+1)T_0\} &= \mathbf{E}_x[\zeta_\alpha > nT_0; \mathbf{P}_{X_{nT_0}^\varepsilon}\{\zeta_\alpha > T_0\}] \\
&\leq \mathbf{P}_x\{\zeta_\alpha > nT_0\}\exp[-\varepsilon^{-2}(A/2 - \gamma)] \ .
\end{aligned}$$

So by induction we see that

$$\begin{aligned}
\mathbf{P}_x\{\zeta_\alpha > T\} &\leq \mathbf{P}_x\{\zeta_\alpha > \left[\frac{T}{T_0}\right]T_0\} \\
&\leq \exp\{-\varepsilon^{-2}(A/2 - \gamma)\left[\frac{T}{T_0}\right]\} \\
&\leq \exp\{-\varepsilon^{-2}\left(\frac{T}{T_0} - 1\right)(A/2 - \gamma)\} \ .
\end{aligned}$$

Putting $c = \dfrac{A/2 - \gamma}{T_0}$, we get the desired statement. $\square$

3.3.2   Transition probabilities between neighborhoods of the O_i's

In this section we study the asymptotic transition probabilities between neighborhoods of the ω-limit sets $\{O_1, ..., O_l\}$. We first provide several auxiliary lemmas.

Lemma 3.1. *There exists a constant $L > 0$ such that for any $x, y \in [D]$ sufficiently close to each other, there exists a function $\varphi \in \mathbf{C}_{[0,T]}([D]), \varphi_0 = x, \varphi_T = y$, such that we have $S_{0T}^+(\varphi) < L \cdot |x - y|_{\mathbb{R}^d}$.*

Proof. Let x and y be so close to each other that they can be covered by one coordinate chart U. Let this coordinate chart correspond to a coordinate function

$u : U \to \mathbb{R}^d(\text{ or } \mathbb{R}^d_+)$. The function u is smooth with bounded derivatives. Let us take $T = |x - y|_{\mathbb{R}^d}$,

$$\varphi_t = u^{-1}\left(u(x) + \frac{t}{T}(u(y) - u(x)) \right) .$$

We have, for some constant $M > 0$,

$$
\begin{aligned}
S^+_{0T}(\varphi_s) &= \frac{1}{2} \int_0^T \|\dot{\varphi}_s - \bar{b}(\varphi_s)\|^2_{a^{-1}(\varphi_s)} ds \\
&= \frac{1}{2} \int_0^T \sum_{i,j=1}^d a_{ij}^{-1}(\varphi_s)(\dot{\varphi}^i_s - \bar{b}^i(\varphi_s))(\dot{\varphi}^j_s - \bar{b}^j(\varphi_s)) ds \\
&\leq \frac{\theta^2}{2} \int_0^T |\dot{\varphi}_s - \bar{b}(\varphi_s)|^2_{\mathbb{R}^d} ds \\
&= \frac{\theta^2}{2} \int_0^T \left| \frac{1}{T}(u^{-1})'\left(u(x) + \frac{t}{T}(u(y) - u(x)) \right) \circ (u(y) - u(x)) - \bar{b}(\varphi_s) \right|^2_{\mathbb{R}^d} ds \\
&\leq \theta^2 M \left\{ \frac{1}{T}|y - x|^2_{\mathbb{R}^d} + T \max_{x \in [D]} |\bar{b}(x)|^2_{\mathbb{R}^d} \right\} .
\end{aligned}
$$

Taking into account that $T = |y - x|_{\mathbb{R}^d}$, we are done. $\square$

Lemma 3.2. *For any $\gamma > 0$ and any compact subset $K \subseteq [D]$ there exists T_0 such that for any $x, y \in K$ there exists a function $\varphi_t, 0 \leq t \leq T, \varphi_0 = x, \varphi_T = y, T \leq T_0$ such that $S^+_{0T}(\varphi) \leq V^+(x, y) + \gamma$.*

Proof. We choose a finite δ-net $\{x_i\}$ of points in K; we connect them with curves at which the action functional assumes values differing from the infimum by less than $\frac{\delta\gamma}{2}$ and complete them with end sections using Lemma 3.1: from x to a point x_i near x and then from x_i to a point x_j near y, and from x_j to y. By choosing δ small enough we get the desired result. $\square$

We define

$$\widetilde{V}^+(O_i, O_j) = \inf_{\varphi \in \mathbf{C}_{[0,T]}([D])} \left\{ S_{0T}^+(\varphi) : \varphi_0 = O_i, \varphi_T = O_j, \varphi_t \in [D] \backslash \bigcup_{s \neq i,j} \{O_s\}, 0 < t < T \right\}.$$

A "$\widetilde{V}^+(O_i, O_j)$ version" of the above Lemma can proved similarly: one can take the curve φ in such a way that it avoids $\bigcup_{s \neq i,j} \{O_s\}$ and such that $S_{0T}(\varphi) \leq \widetilde{V}^+(O_i, O_j) + \gamma$. We omit the proof.

Let constant $\rho_0 > 0$ be small. Let constant $0 < \rho_1 < \rho_0$. We denote by C the set $D \cup \partial D$ from which we delete the ρ_0-neighborhoods of the $O_i, i = 1, 2, ..., l$; by Γ_i the boundaries of the ρ_0-neighborhoods of O_i: $\Gamma_i = \partial \mathcal{E}_{\rho_0}(O_i)$; by g_i the ρ_1-neighborhoods of the O_i, and by g the union of all the g_i.

We introduce the following random times $\tau_0 = 0, \sigma_n = \inf\{t \geq \tau_n, X_t^{x,\varepsilon} \in C\}, \tau_n = \inf\{t \geq \sigma_{n-1}, X_t^{x,\varepsilon} \in \partial g\}$. We consider the Markov chain $Z_n = X_{\tau_n}^{x,\varepsilon}$ for $n \geq 0$. We see that from $n = 1$ on $Z_n \in \partial g$. Also, $X_{\sigma_0}^{\varepsilon}$ can be any point of C, all the following $X_{\sigma_n}^{x,\varepsilon}$ belong to one of the Γ_i's. The chain never stops.

We are now ready to prove:

Theorem 3.2. *For any $\gamma > 0$ there exists $\rho_0 > 0$ (which can be chosen arbitrary small) such that for any ρ_2, $0 < \rho_2 < \rho_0$, there exists ρ_1, $0 < \rho_1 < \rho_2$ such that for all x in the ρ_2-neighborhood of $O_i(i = 1, ..., l)$ the one-step transition*

probabilities of Z_n, $Z_0 = x$ satisfy the inequality

$$\exp[-\varepsilon^{-2}(\widetilde{V}^+(O_i, O_j) + \gamma)] \leq \mathbf{P}(x, \partial g_j) \leq \exp[-\varepsilon^{-2}(\widetilde{V}^+(O_i, O_j) - \gamma)]$$

for some $0 < \varepsilon < \varepsilon_0$.

Proof. We can assume $\widetilde{V}^+(O_i, O_j) < \infty$. Set $\widetilde{V}_0^+ = \max\limits_{i,j=1,2,\ldots,l} \widetilde{V}^+(O_i, O_j)$. Choose positive ρ_0 small enough. For every pair O_i, O_j for which $\widetilde{V}^+(O_i, O_j) < \infty$, we choose a function $\varphi_t^{i,j} \in \mathbf{C}_{[0,T]}([D]), 0 \leq t \leq T = T(O_i, O_j)$, such that $\varphi_0^{i,j} = O_i, \varphi_T^{i,j} = O_j, \varphi_t^{i,j}$ does not touch $\bigcup\limits_{s \neq i,j} \{O_s\}$, and such that (by Lemma 3.2)

$$S_{0T}^+(\varphi^{i,j}) \leq \widetilde{V}^+(O_i, O_j) + 0.5\gamma .$$

We choose positive ρ_1 smaller than $\frac{\rho_0}{2}$, ρ_2 and

$$\frac{1}{2} \min\{\rho(\varphi_t^{i,j}, \bigcup\limits_{s \neq i,j} \{O_s\}) : 0 \leq t \leq T, i, j = 1, 2, \ldots l\} .$$

For every x in a ρ_2-neighborhood of O_i, we take a curve connecting x with O_i and for which the value of S^+ does not exceed 0.3γ (by Lemma 3.1). We combine this curve with the curve $\varphi_t^{i,j}$ and obtain a function $\varphi_t, 0 \leq t \leq T, \varphi_0 = x, \varphi_T = O_j$ (with a possible small change of T from $T = T(O_i, O_j)$) such that

$$S_{0T}^+(\varphi) \leq \widetilde{V}^+(O_i, O_j) + 0.8\gamma .$$

From Lemma 3.2 we choose a $T_0 \geq T$, and extend the curve φ_t to $T \leq t \leq T_0$ by using a trajectory of the dynamical system $\dot{x}_t = \bar{b}(x_t)$ on $\partial_1 D$, without changing the value of $S_{0T_0}^+(\varphi)$ from that of $S_{0T}^+(\varphi)$. We choose positive δ less than $\rho_1, \rho_0 - \rho_2$. For a trajectory of X_t^ε starting from x, passing at a distance from φ_t smaller than δ for

67

$0 \le t \le T_0$, it must intersect with Γ_i and reaches the δ-neighborhood of O_j without getting closer than δ from any of the other $O_s, s \ne i, j$. Moreover, $X_{\tau_1}^{x,\varepsilon} \in \partial g_j$, thus

$$\mathbf{P}(x, \partial g_j)$$

$$\ge \mathbf{P}_x\{\rho_{0T_0}(X^\varepsilon, \varphi) < \delta\}$$

$$\ge \exp[-\varepsilon^{-2}(S_{0T_0}^+(\varphi) + 0.1\gamma)]$$

$$> \exp[-\varepsilon^{-2}(\widetilde{V}^+(O_i, O_j) + \gamma)].$$

Now we turn to the proof of the upper estimates. For any curve $\varphi_t, 0 \le t \le T$ beginning at x, touching the δ-neighborhood of ∂g_j, not touching any of the $O_s, s \ne i, j$, we have

$$S_{0T}^+(\varphi) \ge \widetilde{V}^+(O_i, O_j) - 0.7\gamma .$$

We use Theorem 3.1 to choose T_1 such that for any $x \in [D] \setminus g$ we have $\mathbf{P}_x^\varepsilon\{\tau_1 > T_1\} \le \exp(-\varepsilon^{-2}V_0^+)$ for some $V_0^+ > 0$.

Any trajectory $X_t^{x,\varepsilon}$ beginning at x and being in ∂g_j at time τ_1 either spends time T_1 without touching ∂g or reaches ∂g_j over time T_1, in this case

$$\rho_{0T_1}(X^{x,\varepsilon}, \Phi_x(\widetilde{V}^+(O_i, O_j) - 0.7\gamma)) \ge \delta .$$

Therefore we have

$$\mathbf{P}_x^\varepsilon\{X_{\tau_1}^{x,\varepsilon} \in \partial g_j\} \quad \le \mathbf{P}_x^\varepsilon\{\tau_1 > T_1\} + \mathbf{P}_x^\varepsilon\{\rho_{0T_1}(X^{x,\varepsilon}, \Phi_x(\widetilde{V}^+(O_i, O_j) - 0.7\gamma)) \ge \delta\}$$

$$\le \exp(-\varepsilon^{-2}V_0^+) + \exp[-\varepsilon^{-2}(\widetilde{V}^+(O_i, O_j) - 0.9\gamma)]$$

$$\le \exp[-\varepsilon^{-2}(\widetilde{V}^+(O_i, O_j) - \gamma)]$$

for sufficiently small ε. $\square$

In an exactly similar way one can also formulate the estimate on transition probability based on the quantities

$$\widetilde{V}^+(x, O_j) = \inf_{\varphi \in \mathbf{C}_{[0,T]}([D])} \{S_{0T}^+(\varphi) : \varphi_0 = x, \varphi_T = O_j, \varphi_t \in [D] \setminus \bigcup_{s \neq j}\{O_s\}, 0 < t < T\} \ .$$

We have

Theorem 3.3. *For any $\gamma > 0$ there exists $\rho_0 > 0$ (which can be chosen arbitrary small) such that for any ρ_2, $0 < \rho_2 < \rho_0$, there exists ρ_1, $0 < \rho_1 < \rho_2$ such that for all x outside the ρ_2-neighborhood of $O_i(i = 1, ..., l)$ the one-step transition probabilities of Z_n, $Z_0 = x$ satisfy the inequality*

$$\exp[-\varepsilon^{-2}(\widetilde{V}^+(x, O_j) + \gamma)] \leq \mathbf{P}(x, \partial g_j) \leq \exp[-\varepsilon^{-2}(\widetilde{V}^+(x, O_j) - \gamma)]$$

for some $0 < \varepsilon < \varepsilon_0$.

3.3.3 The invariant measure of $X_t^{x,\varepsilon}$; sublimiting distribution

In this section we study the invariant measure of the process $X_t^{x,\varepsilon}$. Based on the estimates on transition probabilities given above, the proof of the asymptotic result is the same as that of [4, Ch.6] and [18]. Let us formulate and prove two more technical lemmas, after which the rest of the proof is just a study of Markov chains on graphs. The latter part will be omitted since it is the same as [4, Ch.6] and [18].

Lemma 3.3. *For $i \in \{1, 2, ..., l\}$, define*

$$\tau_{\mathcal{E}_\delta(O_i)} = \inf\{t, X_0^\varepsilon = x, X_t^{x,\varepsilon} \in \partial\mathcal{E}_\delta(O_i)\} \ .$$

For any $\gamma > 0$, there exist $\delta > 0$ such that for all sufficiently small ε and $x \in \mathcal{E}_\delta(O_i)$

we have

$$\mathbf{E}_x^\varepsilon \tau_{\mathcal{E}_\delta(O_i)} < \exp(\gamma\varepsilon^{-2}) \ .$$

Proof. Choose point $z \in D$ close to O_i. Put $\delta = \frac{|z - O_i|}{2}$. Connect x with O_i

and O_i with z with the values of S^+ not exceeding $\frac{\gamma}{4}$ and $\frac{\gamma}{2}$, the resulting function

is called $\widetilde{\varphi}_t$. The length of the time interval of $\widetilde{\varphi}_t$ is uniformly bounded by T_0 for

all $x \in G$. We extend $\widetilde{\varphi}_t$ up to T_0 by using a trajectory of $\dot{x}_t = \bar{b}(x_t)$ in $D \cup \partial D$

without making S^+ larger.

Now we have for $x \in \mathcal{E}_\delta(O_i)$,

$$\mathbf{P}_x^\varepsilon\{\tau_{\mathcal{E}_\delta(O_i)} < T_0\} \geq \mathbf{P}_x^\varepsilon\{\rho_{0T_0}(X^{x,\varepsilon}, \widetilde{\varphi}) < \delta\} \geq \exp(-0.9\gamma\varepsilon^{-2}) \ .$$

Using the Markov property we see that

$$\mathbf{P}_x^\varepsilon\{\tau_{\mathcal{E}_\delta(O_i)} \geq nT_0\} \leq [1 - \exp(-0.9\gamma\varepsilon^{-2})]^n \ .$$

This yields

$$\mathbf{E}_x^\varepsilon \tau_{\mathcal{E}_\delta(O_i)} \leq T_0 \sum_{n=0}^{\infty} [1 - \exp(-0.9\gamma\varepsilon^{-2})]^n = T_0 \exp(0.9\gamma\varepsilon^{-2}) \ .$$

Sacrificing 0.1γ in order to get rid of T_0 we get the desired result. $\square$

Lemma 3.4. *For any $\gamma > 0$ there exist $\rho_1 > 0$ such that for all sufficiently small ε and $y \in \partial g_i$ we have*

$$\mathbf{E}_y^\varepsilon \int_0^{\sigma_0} \chi_{g_i}(X_t^{x,\varepsilon}) dt > \exp(-\gamma \varepsilon^{-2}) \ .$$

Proof. Choose ρ_1 small. We connect $y \in \partial g_i$ with O_i using a curve φ_t, extend it using the trajectory of $\dot{x}_t = \bar{b}(x_t)$ on ∂D till first exit time σ_0 from $\mathcal{E}_{\rho_0}(O_i)$, with corresponding S^+ less than 0.5γ. All the trajectories at a distance less than $\frac{\rho_1}{2}$ spends a time at least $t_0 > 0$ within g_i, uniformly for all $y \in \partial g_i$. The probability of all such trajectories is no less than $\exp(-0.9\gamma\varepsilon^{-2})$. Thus the expected value is no less than $t_0 \exp(-0.9\gamma\varepsilon^{-2})$. By sacrificing 0.1γ we can get rid of t_0. $\square$

The rest of this section is devoted to the description of the algorithm for the calculation of the invariant measure and the metastable states. The proof we shall omit here follows [4, Ch.6] and [18].

Let L be a finite set (in our case $L = \{1, 2, ..., l\}$), whose elements are denoted by letters i, j, k, m, n, etc. Let a subset W be selected in L. A graph consisting of arrows $m \to n$ ($m \in L\backslash W, n \in L, n \neq m$) is called a W-graph if it satisfies the following conditions:

(1) every point $m \in L\backslash W$ is the initial point of exactly one arrow;

(2) there are no cycles in the graph.

Intuitively, a W-graph is a graph consisting of arrows starting from each point $m \in L\backslash W$, and going along a sequence of arrows leading to some point $n \in W$.

The set of W-graphs is denoted by $G(W)$. We shall use the letter g to denote graphs.

Let $W(O_i) = \min\limits_{g \in G\{i\}} \sum_{(m \to n) \in g} \widetilde{V}^+(O_m, O_n)$. It can be proved that

$$W(O_i) = \min\limits_{g \in G\{i\}} \sum_{(m \to n) \in g} V^+(O_m, O_n) \ .$$

We have

Theorem 3.4. *Let μ^ε be the normalized invariant measure of the process X_t^ε. Then for any $\gamma > 0$ there exists $\rho_1 > 0$ such that we have*

$$\exp[-\varepsilon^{-2}(W(O_i) - \min\limits_{i} W(O_i) + \gamma)] \leq \mu^\varepsilon(g_i) \leq \exp[-\varepsilon^{-2}(W(O_i) - \min\limits_{i} W(O_i) - \gamma)]$$

for sufficiently small $\varepsilon > 0$.

We shall say that a set $N \subset [D]$ is *stable* if for any $x \in N$, $y \notin N$ we have $V^+(x,y) > 0$. One can show that for an unstable O_j $(j = 1, ..., l)$ there exist a stable O_i $(i \neq j, i = 1, ..., l)$ such that $V^+(O_i, O_j) = 0$.

Theorem 3.5. *For $x \in [D]$ set*

$$W(x) = \min[W(O_i) + V(O_i, x)] \ ,$$

where the minimum can be taken over either all of $O_1, ..., O_l$ or only stable ones. Let μ^ε be the normalized invariant measure of the process X_t^ε. Then for any $\gamma > 0$

72

there exists $\bar{\rho} > 0$ such that for any $0 < \rho < \bar{\rho}$ we have

$$\exp[-\varepsilon^{-2}(W(x) - \min_i W(O_i) + \gamma)] \le \mu^\varepsilon(\mathcal{E}_\rho(x)) \le \exp[-\varepsilon^{-2}(W(x) - \min_i W(O_i) - \gamma)]$$

for sufficiently small $\varepsilon > 0$.

Here $\mathcal{E}_\rho(x)$ is a ρ-neighborhood of x.

The above two theorems roughly say that as first $t \to \infty$ and then $\varepsilon \to 0$, the process $X_t^{x,\varepsilon}$ will be situated in one of the O_i's which minimizes the values of $W(O_i)$ (it can be calculated either via all $O_1, ..., O_l$ or only via the stable ones). In generic case, when $\min_i W(O_i)$ is attained at some unique point i, we have for any $\delta > 0$,

$$\lim_{\varepsilon \to 0} \lim_{t \to \infty} \mathbf{P}_x^\varepsilon \{|X_t^{x,\varepsilon} - O_i| > \delta\} = 0 . \tag{3.1}$$

A natural question is that how the limiting distribution behaves when we take the limit in a coordinated way, i.e. take $\varepsilon \to 0$ and $t = t(\varepsilon^{-2}) \to \infty$. This is the problem of metastability and sublimiting distributions (see [20]). Let us assume that $T = T(\varepsilon) \asymp \exp(\frac{\lambda}{\varepsilon^2})$ and we consider $\lim_{\varepsilon \to 0} \mathbf{P}_x^\varepsilon \{X_{T(\varepsilon)}^{x,\varepsilon} \in \Gamma\}$. In the generic case one can define a function $K^*(x, \lambda) \in \{1, 2, ..., l\}$ such that

$$\lim_{\varepsilon \to 0} \mathbf{P}_x^\varepsilon \{|X_{T(\varepsilon)}^{x,\varepsilon} - O_{K^*(x,\lambda)}| > \delta\} = 0 \tag{3.2}$$

for any $\delta > 0$.

The algorithm to determine $K^*(x, \lambda)$ is as follows. First we consider for each O_i (the rank 0 cycle) the "next" most probable ω-limit set $\mathcal{N}(O_i)$ that we are going

to jump to. Continuing this determination of "next" states we form the rank 1 cycle $O_i \rightarrow \mathcal{N}(O_i) \rightarrow \mathcal{N}^2(O_i) \rightarrow ... \rightarrow \mathcal{N}^{m^1_{\{i\}}}(O_i)$. We stop once we get a repetition $\mathcal{N}(\mathcal{N}^{m^1_{\{i\}}}(O_i)) = O_i$. Cycles generated by distinct initial points $i \in \{1, ..., l\}$ either do not intersect each other or coincide: in the latter case the cycle order on them is one and the same.

We continue by recurrence. Let the cycles of rank $(k-1)$ be $\pi_1^{k-1}, ..., \pi_{n_{k-1}}^{k-1}$. Starting from each $(k-1)$-cycle π_i^{k-1} one can determine the "next" most probable $(k-1)$-cycle $\pi_{\mathcal{N}(\pi_i^{k-1})}^{k-1}$ that we will first jump to. Continuing this determination we form a rank k cycle $\pi_i^{k-1} \rightarrow \pi_{\mathcal{N}(\pi_i^{k-1})}^{k-1} \rightarrow ... \rightarrow \pi_{\mathcal{N}^{m_{\pi_i^{k-1}}}(\pi_i^{k-1})}^{k-1}$. We stop once we get a repetition $\mathcal{N}(\mathcal{N}^{m_{\pi_i^{k-1}}}(\pi_i^{k-1})) = \pi_i^{k-1}$. Cycles of rank k generated by distinct cycles of rank $k-1$ either do not intersect each other or coincide.

In this way we can continue until the last cycle which is the whole of $\{O_1, ..., O_l\}$. The metastable states are determined by the timescale of the cycles that we traverse.

Let us be more precise. Starting from a cycle π, to determine the "next" cycle $\mathcal{N}(\pi)$ that we first jump to, we calculate

$$A(\pi) = \min_{g \in G(L \backslash \pi)} \sum_{(m \rightarrow n) \in g} V^+(O_m, O_n) \ . \tag{3.3}$$

Here $L = \{1, 2, ..., l\}$. The minimum of the above expression determines a $L \backslash \pi$ graph consisting of chains of arrows leading to the first state in $L \backslash \pi$ we jump to.

We put

$$C(\pi) = A(\pi) - \min_{i \in \pi} \min_{g \in G_\pi \{i\}} \sum_{(m \rightarrow n) \in g} V^+(O_m, O_n) \ . \tag{3.4}$$

74

Here $G_\pi\{i\}$ is the set of $\{i\}$-graphs restricted to π. Then the asymptotic exit time from π is of order $\asymp \exp\left(\frac{C(\pi)}{\varepsilon^2}\right)$.

Starting from $i = i(x)$ (which is the label for the first equilibrium among $O_1, ..., O_l$ that we approach in finite time, starting from x), let $\pi, \pi', ..., \pi^{(s)}$ be cycles of next to the last rank, unified into the last cycle, which exhausts $\{1, 2, ..., l\}$. If the constant λ is greater than $C(\pi), C(\pi'), ..., C(\pi^{(s)})$, then over time of order $\exp(\lambda\varepsilon^{-2})$ the process can traverse all these cycles many times (and all cycles of smaller rank inside them) and the limiting distribution is concentrated on that one of the cycles for which $C(\pi), C(\pi'), ..., C(\pi^{(s)})$ is the greatest. Within this cycle, it is concentrated on that one of the subcycles for which the corresponding constant $C(.)$ in (3.4) is the greatest possible, and so on up to points (one point in the generic case) $O_{K^*(x,\lambda)}$. This point $O_{K^*(x,\lambda)}$ is the metastable state in (3.2).

3.4   Application to PDE

The solution of (3.1) can be represented through process (1.2) by the formula $u^\varepsilon(x, t) = \mathbf{E}_x g(X_t^{x,\varepsilon})$. This is an immediate consequence of the following generalized Itô's formula:

Lemma 4.1. *Assume process* $(X_t^{x,\varepsilon}, \xi_t^{x,\varepsilon})$ *is given by (1.2),* $X_0^{x,\varepsilon} = x$. *Let* $u(x, t)$ *be of class* $\mathbf{C}^{2,1}(\mathbb{R}^d \times \mathbb{R}_+)$ *with uniform bounded derivatives up to the second order in x and up to the first order in t. Then we have*

$$u(X_t^{x,\varepsilon}, t) - u(x, 0)$$

$$= \int_0^t \left(\frac{\partial}{\partial s} + \mathcal{L}^\varepsilon \right) u(X_s^{x,\varepsilon}, s) ds + \int_0^t \nabla u(X_s^{x,\varepsilon}, s) \cdot \gamma(X_s^{x,\varepsilon}) d\xi_s^{x,\varepsilon} + \int_0^t \nabla u(X_s^{x,\varepsilon}, s) \cdot \sigma(X_s^{x,\varepsilon}) dW_s \ .$$

For a proof of this theorem see [21, section 3].

Our answer to the problem (3.1) is

Theorem 4.1. *Under all our assumptions, in generic case, for $T(\varepsilon) \asymp$*

$\exp(\frac{\lambda}{\varepsilon^2})$, *we have*

$$\lim_{\varepsilon \to 0} u^\varepsilon(x, T(\varepsilon)) = g(O_{K^*(x,\lambda)}) \ ,$$

where $K^*(x, \lambda)$ is defined as in section 3.3.

3.5 Example

Consider an example. Let the domain D be a unit disk $B(1) = \{(y_1, y_2); y_1^2 + y_2^2 < 1\}$ in $\mathbb{R}^2$. Let the smooth vector field $b_y(y_1, y_2)$ be given such that $\bar{b}_y(y_1, y_2) = (\bar{b}_{y_1}(y_1, y_2), \bar{b}_{y_2}(y_1, y_2))$ is as in Fig.1. We consider the problem
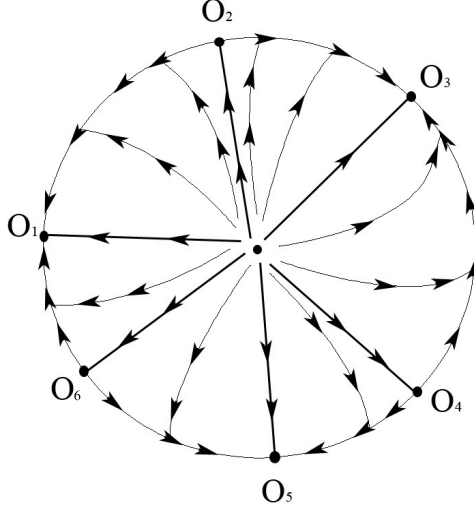
Figure 3.1: An example

$$
\begin{cases}
\dfrac{\partial u^\varepsilon(y_1, y_2, t)}{\partial t} = \dfrac{\varepsilon^2}{2} \Delta_{y_1, y_2} u^\varepsilon(t, y_1, y_2) + b_y(y_1, y_2) \cdot \nabla u^\varepsilon(t, y_1, y_2) \,, & \varepsilon > 0 \,; \\[2ex]
u^\varepsilon(0, y_1, y_2) = g(y_1, y_2) \,, & y_1^2 + y_2^2 \leq 1 \,; \\[2ex]
\dfrac{\partial u^\varepsilon}{\partial r}(t, y_1, y_2) = 0 \,, & y_1^2 + y_2^2 = 1 \,, \ t \geq 0 \,.
\end{cases}
\tag{5.1}
$$

Here $\dfrac{\partial}{\partial r}$ is the derivative with respect to the inward unit normal. The action functional takes the form

$$
S_{0T}^+(\varphi) =
\begin{cases}
\dfrac{1}{2} \displaystyle\int_0^T |\dot\varphi_s - \bar b_y(\varphi_s)|_{\mathbb{R}^2}^2 ds \,, & \varphi \in \mathbf{C}_{[0,T]}([D]) \text{ absolutely continuous} \,, \varphi_0 = x \,; \\[2ex]
+\infty \,, & \text{for the rest of } \varphi \in \mathbf{C}_{[0,T]}([D]) \,.
\end{cases}
\tag{5.2}
$$

We calculate the "quasi-potential" using (5.2)

$$
V^+(x, y) = \inf_{\varphi \in \mathbf{C}_{[0,T]}([D])} \{ S_{0T}^+(\varphi), \varphi_0 = x, \varphi_T = y, \varphi_t \in D \cup \partial D, 0 \leq t \leq T < \infty \} \,.
$$

The ω-limit sets of the dynamical system $\dot{x}_t = \bar{b}_y(x_t)$ are the zeros of the vector field $\bar{b}_y(x)$ on $\partial D = S^1$. (And also the origin but it is unstable so that we neglect it.) In Fig.1 the points O_1, O_3 and O_5 are stable equilibriums are the points O_2, O_4 and O_6 are unstable ones. We can consider only the quasi-potentials between the stable ones. Suppose we have $V^+(O_1, O_3) = 1$, $V^+(O_3, O_1) = 2$, $V^+(O_1, O_5) = 6$, $V^+(O_5, O_1) = 7$, $V^+(O_5, O_3) = 3$, $V^+(O_3, O_5) = 4$.

We are concerned with the limit $\lim_{\varepsilon \downarrow 0} u^\varepsilon(y_1, y_2, T(\varepsilon))$ for $T(\varepsilon) \asymp \exp(\frac{\lambda}{\varepsilon^2})$. Starting from the initial point (y_1, y_2), we suppose that we are attracted to O_1 first. By calculating $\min_{g \in G(L \setminus \{1\})} \sum_{(m \to n) \in g} V^+(O_m, O_n) = 1$ we see that over time $\exp(\frac{1}{\varepsilon^2})$ we are going to jump to O_3 first. We then calculate $\min_{g \in G(L \setminus \{3\})} \sum_{(m \to n) \in g} V^+(O_m, O_n) = 2$ and we see that over time $\exp(\frac{2}{\varepsilon^2})$ we will jump from O_3 back to O_1 and we form a cycle $\pi^{(1)} = \{1, 3\}$ of rank 1. We then calculate $A(\pi^{(1)}) = \min_{g \in G(L \setminus \pi^{(1)})} \sum_{(m \to n) \in g} V^+(O_m, O_n) = V^+(O_1, O_3) + V^+(O_3, O_5) = 5$ and the first state out of cycle $\pi^{(1)}$ that we are going to jump to is O_5. Within cycle $\pi^{(1)}$ we are mostly staying in O_3. We calculate $C(\pi^{(1)}) = 5 - \min_{i \in \{1,3\}} \min_{g \in G_{\{1,3\}}\{i\}} \sum_{(m \to n) \in g} V^+(O_m, O_n) = 4$. This means, that over time $\exp(\frac{4}{\varepsilon^2})$ we are jumping from O_3 to O_5. We then calculate $\min_{g \in G(L \setminus \{5\})} \sum_{(m \to n) \in g} V^+(O_m, O_n) = 3$ and we see that we are jumping from O_5 out to O_3 in time $\exp(\frac{3}{\varepsilon^2})$. This implies that within the cycle $\pi^{(2)} = \{1, 3, 5\}$ which exhausts all ω-limit sets, we are mostly staying in $\pi^{(1)}$, and within $\pi^{(1)}$ it is O_3.

Our result can be summarized as

$$\lim_{\varepsilon \downarrow 0} u^\varepsilon(T(\varepsilon), y_1, y_2) = g(O_1) \text{ for } T(\varepsilon) \asymp \exp(\frac{\lambda}{\varepsilon^2}) \text{ and } 0 < \lambda < 1 ;$$

78

$$\lim_{\varepsilon \downarrow 0} u^\varepsilon(T(\varepsilon), y_1, y_2) = g(O_3) \text{ for } T(\varepsilon) \asymp \exp(\frac{\lambda}{\varepsilon^2}) \text{ and } 1 \le \lambda \ .$$

Bibliography

[1] R . F. Anderson and S. Orey. Small random perturbations of dynamical systems with reflecting boundary. *Nagoya Math*, 60:189–216, 1976.

[2] A. Kolmogorov. Uber die analytischen methoden in der wahrscheinlichkeitsrechnung(on analytical methods in the theory of probability). *Mathematische Annalen*, 104:415–458, 1931.

[3] A. Einstein. Investigations on the theory of the brownian movement. *Dover Publications*, 1926.

[4] Mark I. Freidlin and A. D. Wentzell. *Random Perturbations of Dynamical Systems*. Springer, 1998.

[5] O. A. Ladyzenskaya and N. N. Ural'zeva. *Linear and Quasi-Linear Parabolic Equations*. Nauka, 1967.

[6] Mark I. Freidlin. Sublimiting distributions and stabilization of solutions of parabolic equations with a small parameter. *Soviet. Math. Dokl.*, 237(5):1042–1045, 1977.

[7] Mark I. Freidlin. Quasi-deterministic approximation, metastability and stochastic resonance. *Physica D*, 137:333–352, 2000.

[8] E. Oliveiri and M.E. Vares. *Large Deviations and Metastability*. Cambridge University Press, 2005.

[9] M. I. Freidlin and L. Koralov. Metastability for nonlinear random perturbations of dynamical systems. *Stochastic Processes and Applications*, 120(7), 2010.

[10] Mark I. Freidlin. *Functional Integration and Partial Differential equations,*. Princeton University Press, 1985.

[11] M.I. Freidlin and L. Koralov. Nonlinear stochastic perturbations of dynamical systems and quasi-linear parabolic pde's with a small parameter. *Probability Theory and Related Fields*, 147, 2010.

[12] N. V. Krylov. *Nonlinear Elliptic and Parabolic Equations of the Second Order.* Springer, 1987.

[13] Mark I. Freidlin. *Functional Integration and Partial Differential equations.* Princeton University Press, 1985.

[14] S. Watanabe. On stochastic differential equations for multidimensional diffusion processes with boundary, i,ii. *J. Math. Kyoto U.*, 11, 1971.

[15] K. Itô and H.P. Jr. McKean. *Diffusion Processes and their Sample Paths.* New york Academic Press Inc., 1965.

[16] K. Sato and H. Tanaka. Local times on the boundary for multidimensional reflecting diffusion. *Proc. of Japan Acad.*, 38(10), 1962.

[17] K. Sato and T. Ueno. Multidimensional diffusion and markov processes on the boundary. *Math.*, 4, 1965.

[18] M. Freidlin and A. Wentzell. On small random perturbations of dynamical systems. *Russ. Math. Surv.*, 25(1):1–56, 1970.

[19] Mark I. Freidlin and L. Zhivoglyadova. Boundary value problems with a small parameter for a diffusion process with reflection. *Russ. Math. Surv.*, 31(5):241–242, 1976.

[20] Mark I. Freiflin. Sublimiting distributions and stabilization of solutions of parabolic equations with a small parameter. *Soviet. Math. Dokl.*, 237(5):1042–1045, 1977.

[21] I. Gikman and A. Skrokhod. *The Theory of Stochastic Processes.* III. Springer, 1979.