

# Quasi-linear equations with a small diffusion term and the evolution of hierarchies of cycles

Leonid Koralov\*, Lucas Tcheuko†

## Abstract

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  $\varepsilon^2$  at the diffusion term. The solution to a 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  $\lambda$ . We describe the evolution of the hierarchies with respect to  $\lambda$  in order to gain information on the limiting behavior of the solution of the PDE.

2000 Mathematics Subject Classification Numbers: 60F10, 35K55.

Keywords: Large Deviations, Metastable State, Metastable Distribution.

## 1 Introduction

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

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

$$u^\varepsilon(0, x) = g(x), \quad x \in \mathbb{R}^d. \quad (2)$$

Equations with diffusion coefficients that depend on particle concentration  $u^\varepsilon$  arise naturally in many applications, in particular in population genetics. Note that  $b$  is assumed not to depend on  $u^\varepsilon$  here. Any dependence of  $b$  on  $u^\varepsilon$  leads to additional levels of complexity, and will be a subject of our future work.

---

\*Dept. of Mathematics, University of Maryland, College Park, MD 20742, koralov@math.umd.edu

†Dept. of Mathematics, University of Maryland, College Park, MD 20742, lucast@umd.edu

We assume that the coefficients of equation (1) are Lipschitz continuous and bounded; the matrix  $(a_{ij}(x, u))$  is assumed to be uniformly positive definite. Under these conditions, problem (1)-(2) has a unique solution for any continuous bounded  $g(x)$  (see, for instance, [8]).

We'll be interested in the asymptotics of the solution  $u^\varepsilon$  to the problem (1)-(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  $u^\varepsilon$ . Then the Cauchy problem takes the form

$$\frac{\partial u^\varepsilon(t, x)}{\partial t} = \frac{\varepsilon^2}{2} \sum_{i,j=1}^d a_{ij}(x) \frac{\partial^2 u^\varepsilon(t, x)}{\partial x_i \partial x_j} + b(x) \cdot \nabla_x u^\varepsilon(t, x), \quad x \in \mathbb{R}^d, \quad t > 0, \quad (3)$$

$$u^\varepsilon(0, x) = g(x), \quad x \in \mathbb{R}^d. \quad (4)$$

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

$$dX_t^{x,\varepsilon} = b(X_t^{x,\varepsilon})dt + \varepsilon \sigma(X_t^{x,\varepsilon})dW_t, \quad X_0^{x,\varepsilon} = x \in \mathbb{R}^d, \quad (5)$$

where  $\sigma$  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  $|x|$ , while  $(b(x), x - O) \leq -c|x - O|^2$  for all  $x$  in a sufficiently small neighborhood of  $O$ . Then it is easy to check that

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

for any neighborhood  $U$  of the equilibrium  $O$ . Taking into account that the solution  $u^\varepsilon$  of (3)-(4) can be written in the form  $u^\varepsilon(t, x) = \mathbb{E}g(X_t^{x,\varepsilon})$  and the continuity of  $g$ , we conclude that

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

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 equilibria  $O_1, \dots, O_r$ . Let  $D_i$  be the basin of  $O_i$ ,  $1 \leq i \leq r$ , and assume that the set  $\mathbb{R}^d \setminus (D_1 \cup \dots \cup D_r)$  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  $(\varepsilon, t)$  approaches  $(0, \infty)$  and is now determined by the transitions of  $X_t^{x,\varepsilon}$  between the attractors  $O_1, \dots, O_r$ . 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  $x \in D_1 \cup \dots \cup D_r$  and each  $\lambda \in (0, \infty) \setminus \Lambda$ , one equilibrium  $O_{M(x,\lambda)}$  is defined such that the measures  $\mu^\varepsilon(\Gamma) = \mathbb{P}(X_{T^\varepsilon(\lambda)}^{x,\varepsilon} \in \Gamma)$  converge weakly to the  $\delta$ -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  $u^\varepsilon(t, x) = \mathbb{E}g(X_t^{x,\varepsilon})$  it follows that

$$\lim_{\varepsilon \downarrow 0} u^\varepsilon(T^\varepsilon(\lambda), x) = g(O_{M(x,\lambda)}). \quad (6)$$

It is worth noting that for all sufficiently large  $\lambda$ , 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  $\lambda$ .

The theory of metastability (of sublimiting distributions) was developed in [1] (see also [3], [6], [9]). 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  $\varepsilon \downarrow 0$  ([6]):

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

for absolutely continuous  $\varphi$ ,  $S_{0,T}(\varphi) = +\infty$  for  $\varphi$  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  $O_1, \dots, O_r$  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  $\delta$ -neighborhood of  $O_i$  will enter a  $\delta$ -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  $O_i, \mathcal{N}(O_i), \mathcal{N}^2(O_i) = \mathcal{N}(\mathcal{N}(O_i)), \dots, \mathcal{N}^n(O_i), \dots$  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  $O_i, \mathcal{N}(O_i), \dots, \mathcal{N}^{n-1}(O_i)$ . 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  $C_1^1, \dots, C_{m_1}^1$ . 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  $\delta$ -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  $\delta$ -neighborhood of  $\mathcal{N}(C)$ , provided that the process  $X_t^{x,\varepsilon}$  starts in a  $\delta$ -neighborhood of one of the equilibriums that belong to  $C$ . We'll say that the process exits  $C$  through the equilibrium  $O$  and will use the notation  $O = \mathcal{E}(C)$  in this situation.

It has been shown (see [6]) 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  $\delta$ -neighborhood of  $\mathcal{N}(C)$ . Then there is  $\lambda(C)$  such that  $\varepsilon^2 \ln \tau^{x,\varepsilon} \rightarrow \lambda(C)$  almost surely as  $\varepsilon \downarrow 0$ , provided that  $x$  belongs to a  $\delta$ -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  $O_1, \dots, O_r$ . 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 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  $\lambda_e$  can be expressed through linear combinations of  $V_{ij}$ .

Let us return to the non-linear problem (1)-(2). The first thing to observe is that representation (6) is not immediately available now - metastable states need to be replaced by metastable distributions (see [5]), whose dependence on  $\lambda$  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  $u^\varepsilon$ ). 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 (1)-(2) is defined by the following system on the unknown family of processes and unknown function  $u^\varepsilon$  (see [2], Ch. 5):

$$dX_s^{t,x,\varepsilon} = b(X_s^{t,x,\varepsilon})ds + \varepsilon\sigma(X_s^{t,x,\varepsilon}, u^\varepsilon(t-s, X_s^{t,x,\varepsilon}))dW_s, \quad s \leq t, \quad X_0^{t,x,\varepsilon} = x, \quad (7)$$

$$u^\varepsilon(t, x) = \text{Eg}(X_t^{t,x,\varepsilon}), \quad (8)$$

where the entries  $\sigma_{ij}$  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 (7)-(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  $u^\varepsilon$  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  $u^\varepsilon$ . Consider, however, a time interval  $[T^\varepsilon(\lambda - \delta), T^\varepsilon(\lambda)]$ , where  $\delta$  is small. As will be seen,  $u^\varepsilon$  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  $\lambda_\varepsilon(z)$ . This is the main idea that will allow us to study the evolution in  $\lambda$  of the limit of  $u^\varepsilon(T^\varepsilon(\lambda), x)$ .

In Section 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 3 we review the case of two equilibrium points. This has been considered in an earlier paper ([5]), 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 4 and 5.

## 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

$$k|\xi|^2 \leq \sum_{i,j=1}^d \alpha_{ij}(x)\xi_i\xi_j \leq K|\xi|^2, \quad x \in \mathbb{R}^d, \quad \xi \in \mathbb{R}^d. \quad (9)$$

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

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

$$(b(x), x - O_i) \leq -c|x - O_i|^2 \quad (10)$$

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

$$dX_t^{x,\varepsilon} = b(X_t^{x,\varepsilon})dt + \varepsilon\sigma(X_t^{x,\varepsilon})dW_t, \quad X_0^{x,\varepsilon} = x, \quad (11)$$

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  $\varphi$  defined on  $[0, T]$ ,  $\varphi_0 = x$ , and  $S_{0,T}^\alpha(\varphi) = \infty$  if  $\varphi$  is not absolutely continuous or if  $\varphi_0 \neq x$  (see [6]).

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

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

$$d\tilde{X}_t^{x,\varepsilon} = b(\tilde{X}_t^{x,\varepsilon})dt + \varepsilon\tilde{\sigma}^\varepsilon(t, \tilde{X}_t^{x,\varepsilon})dW_t, \quad (12)$$

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

$$\sup_{(t,x) \in \mathbb{R}^+ \times \mathbb{R}^d} |\tilde{\alpha}_{ij}^\varepsilon(t, x) - \alpha_{ij}(x)| \leq \varkappa, \quad (13)$$

where  $\varkappa$  is small. The reason to introduce the process  $\tilde{X}_t^{x,\varepsilon}$  is that we would like to study the behavior of the process  $X_s^{t,x,\varepsilon}$  given by (7)-(8) on a time interval where the variable  $u^\varepsilon$  found inside the diffusion coefficient of (7) does not change much. Since a-priori we don't know much about the behavior of the diffusion coefficients in (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

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

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

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

**Lemma 2.1.** *Suppose that  $b$  is fixed,  $\alpha$  is Lipschitz continuous with Lipschitz constant  $L$ ,  $\tilde{\alpha}^\varepsilon$  is continuous in  $(t, x)$  and Lipschitz continuous in  $x$ , and*

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

$$k|\xi|^2 \leq \sum_{i,j=1}^d \tilde{\alpha}_{ij}^\varepsilon(t, x)\xi_i\xi_j \leq K|\xi|^2 \quad \text{for } (t, x) \in \mathbb{R}^+ \times D, \quad \xi \in \mathbb{R}^d. \quad (14)$$

For each  $\delta > 0$  there are  $\varkappa > 0$  and a function  $\rho(\varepsilon)$  (that depend on  $\alpha$  and  $\tilde{\alpha}$  through  $L, k$  and  $K$ ) such that  $\lim_{\varepsilon \downarrow 0} \rho(\varepsilon) = 0$  and

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

provided that

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

This lemma was proved in [5]. The main idea was to show that  $S^\alpha$  serves a purpose similar to the action functional for the process  $\tilde{X}_t^{x,\varepsilon}$ , even though the diffusion coefficients for the process are time-dependent.

The next simple lemma does not require the proximity of  $\tilde{\alpha}^\varepsilon$  to  $\alpha$ , but only the boundedness of the entries of  $\tilde{\alpha}^\varepsilon$ . It can be proved by standard arguments from large deviation theory (compare with chapter 3 of [6]).

**Lemma 2.2.** *Suppose that  $b$  is fixed and  $\tilde{\alpha}^\varepsilon$  is continuous in  $(t, x)$  and Lipschitz continuous in  $x$  and satisfies (14). For any compact  $M \subset D$ , there is  $v_0 > 0$  which depends on  $\tilde{\alpha}^\varepsilon$  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*

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

The next lemma implies that the solution of (1)-(2) is nearly constant inside each of the domains  $D_i^\delta = \{x \in D_i : \text{dist}(x, \partial D_i) \geq \delta, |x| \leq 1/\delta\}$ ,  $\delta > 0$ ,  $1 \leq i \leq r$ , for  $\varepsilon$  small enough.

**Lemma 2.3.** *Let  $u^\varepsilon$  be the solution of (1)-(2). For every positive  $\lambda_0$  and  $\delta$  there is a positive  $\varepsilon_0$  such that*

$$|u^\varepsilon(T^\varepsilon(\lambda), x) - u^\varepsilon(T^\varepsilon(\lambda), O_i)| \leq \delta \tag{15}$$

whenever  $x \in D_i^\delta$ ,  $\varepsilon \leq \varepsilon_0$  and  $\lambda \geq \lambda_0$ .

For a proof of this lemma we refer the reader to [4], 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

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

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

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 ([7]).

### 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  $\lambda_e, \lambda_h : [g_{\min}, g_{\max}] \rightarrow \mathbb{R}$  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  $u^\varepsilon$  in the diffusion coefficient. An example with the graphs of these functions is shown in Figure 1.

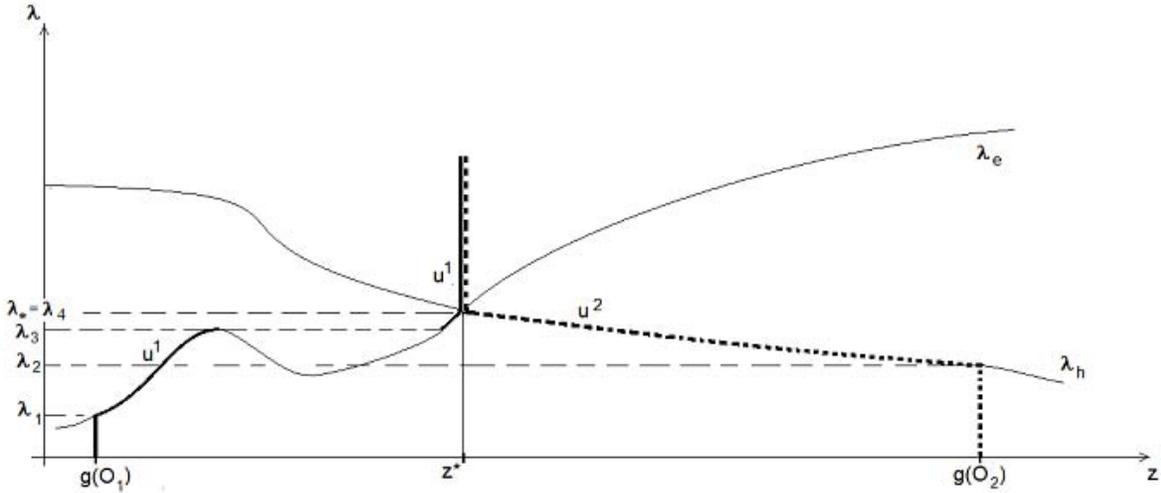


Figure 1: The case of two equilibrium points

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:

$$z^1(\lambda) = \begin{cases} g(O_1), & 0 < \lambda < \lambda_1, \\ \min\{g(O_2), \min\{z : z \in [g(O_1), g(O_2)], \lambda_e(z) = \lambda\}\}, & \lambda \geq \lambda_1, \end{cases} \quad (16)$$

$$z^2(\lambda) = \begin{cases} g(O_2), & 0 < \lambda < \lambda_2, \\ \max\{g(O_1), \max\{z : z \in [g(O_1), g(O_2)], \lambda_h(z) = \lambda\}\}, & \lambda \geq \lambda_2. \end{cases} \quad (17)$$

Let  $\lambda_* = \inf\{\lambda : z^1(\lambda) \geq z^2(\lambda)\}$ . Assume that the functions  $z^1$  and  $z^2$  are continuous at  $\lambda_*$ . Let  $z^* = z^1(\lambda_*) = z^2(\lambda_*)$ . Let  $u^1(\lambda) = \min(z^1(\lambda), z^*)$  and  $u^2(\lambda) = \max(z^2(\lambda), z^*)$ . On Figure 1, 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 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*

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

*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 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  $\lambda$ . By drawing the vertical segment through the point  $(g(O_1), 0)$  in the  $(z, \lambda)$  plane till the intersection with the graph of  $\lambda_e$ , we locate the time scale  $\lambda_1$ . After the time scale  $\lambda_1$ , the function  $u^1$  is just the inverse of the function  $\lambda_e$  for a certain range of values of  $\lambda$ . (The time scale  $\lambda_2$  plays the same role for the function  $u^2$ .) Then, after  $\lambda_*$ , which corresponds to the intersection of the graphs of  $\lambda_e$  and  $\lambda_h$ , 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  $\lambda_e$ .

In the general case (more than two equilibria) 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  $\lambda_e$ , others will correspond to intersection points between  $\lambda_e$  and  $\lambda_h$ , yet others will correspond to local maxima of  $\lambda_e$  for some  $e$  (and there will be other special time scales that are not found in this example).

*Proof of Theorem 3.1.* Let us show that if  $z^1$  is continuous at  $\lambda$ , then

$$\limsup_{\varepsilon \downarrow 0} \sup_{x \in D_1^\delta} u^\varepsilon(T^\varepsilon(\lambda), x) \leq z^1(\lambda). \quad (18)$$

Similarly, if  $z^2$  is continuous at  $\lambda$ , then

$$\liminf_{\varepsilon \downarrow 0} \inf_{x \in D_2^\delta} u^\varepsilon(T^\varepsilon(\lambda), x) \geq z^2(\lambda). \quad (19)$$

Due to Lemma 2.3, in order to prove (18), it is sufficient to show that

$$\limsup_{\varepsilon \downarrow 0} u^\varepsilon(T^\varepsilon(\lambda), O_1) \leq z^1(\lambda), \quad (20)$$

Note that by Lemma 2.2 and (8) there is a positive  $v_0$  such that for every  $0 < \delta < v_0$  there is  $\varepsilon_0 > 0$  such that

$$|u^\varepsilon(T^\varepsilon(\lambda), x) - g(O_i)| \leq \delta \quad (21)$$

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

If (20) fails for a certain value of  $\lambda$ , then due to continuity of the functions  $u^\varepsilon(t, O_i)$  in  $t$ , it follows from (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

$$u^{\varepsilon_n}(t, O_1) \in [z^1(\lambda) + \delta'', z^1(\lambda) + \delta'] \quad \text{for } t \in [T^{\varepsilon_n}(\mu_n), T^{\varepsilon_n}(\lambda_n)]. \quad (22)$$

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

$$u^{\varepsilon_n}(T^{\varepsilon_n}(\lambda_n), O_1) = \mathbb{E} u^{\varepsilon_n} \left( T^{\varepsilon_n}(\mu_n), X_{T^{\varepsilon_n}(\lambda_n) - T^{\varepsilon_n}(\mu_n)}^{T^{\varepsilon_n}(\lambda_n), O_1, \varepsilon_n} \right). \quad (23)$$

Since  $z^1$  is continuous at  $\lambda$ , there are arbitrarily small  $\delta' > 0$  such that  $\lambda_\varepsilon(z^1(\lambda) + \delta') > \lambda_\varepsilon(z^1(\lambda)) = \lambda$ . Since  $\lambda_n \leq \lambda$ , a process starting at  $O_1$  and satisfying (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 2.1, 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  $D_1^\delta$ , where  $\delta$  is sufficiently small. Therefore, due to (22) and Lemma 2.3, we can choose  $\delta''$  sufficiently close to  $\delta'$  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  $\delta'$  and  $\delta''$  thus fixed, we let  $\varepsilon_n \downarrow 0$  in (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 (20) holds, which in turn implies that (18) holds. The proof of (19) is completely similar.

Note that the arguments used to prove (20) also lead to the following statement: for each  $\lambda_0 > 0$

$$\limsup_{\varepsilon \downarrow 0} \sup_{\lambda' \in [\lambda_0, \lambda]} u^\varepsilon(T^\varepsilon(\lambda'), O_1) \leq \lim_{\lambda' \downarrow \lambda} z^1(\lambda'), \quad (24)$$

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

$$\liminf_{\varepsilon \downarrow 0} \inf_{\lambda' \in [\lambda_0, \lambda]} u^\varepsilon(T^\varepsilon(\lambda'), O_2) \geq \lim_{\lambda' \downarrow \lambda} z^2(\lambda'). \quad (25)$$

Let us show that if  $z^1$  is continuous at  $\lambda$ , then

$$\liminf_{\varepsilon \downarrow 0} \inf_{x \in D_1^\delta} u^\varepsilon(T^\varepsilon(\lambda), x) \geq \min(z^1(\lambda), \lim_{\lambda' \downarrow \lambda} z^2(\lambda')). \quad (26)$$

Similarly, if  $z^2$  is continuous at  $\lambda$ , then

$$\limsup_{\varepsilon \downarrow 0} \sup_{x \in D_2^\delta} u^\varepsilon(T^\varepsilon(\lambda), x) \leq \max(z^2(\lambda), \lim_{\lambda' \downarrow \lambda} z^1(\lambda')). \quad (27)$$

Due to Lemma 2.3, in order to prove (26), it is sufficient to show that

$$\liminf_{\varepsilon \downarrow 0} u^\varepsilon(T^\varepsilon(\lambda), O_1) \geq \min(z^1(\lambda), \lim_{\lambda' \downarrow \lambda} z^2(\lambda')). \quad (28)$$

If (28) fails, then for each  $\lambda_0 > 0$  there is  $\delta' > 0$  and a sequence  $\varepsilon_n \downarrow 0$  such that

$$u^{\varepsilon_n}(T^{\varepsilon_n}(\lambda), O_1) < z^1(\lambda) - \delta'. \quad (29)$$

$$u^{\varepsilon_n}(T^{\varepsilon_n}(\lambda), O_1) < \inf_{\lambda' \in [\lambda_0, \lambda]} u^{\varepsilon_n}(T^{\varepsilon_n}(\lambda'), O_2) - \delta'. \quad (30)$$

These two inequalities can not hold at the same time as follows from Lemma 3.11 of [4], where an analogue of (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 (30) the proof goes through without major modifications. We have thus justified (26), and (27) is absolutely similar.

Note that (18), (19), (26), and (27) imply the statement of the theorem for  $0 < \lambda < \lambda_*$ . Expressing the solution at time  $T^\varepsilon(\lambda)$  in terms of the solution at an earlier time  $T^\varepsilon(\lambda')$  (similarly to (23)), we see that if

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

then

$$\begin{aligned} \liminf_{\varepsilon \downarrow 0} \inf_{x \in D_1^\delta} u^\varepsilon(T^\varepsilon(\lambda'), x) &\leq \liminf_{\varepsilon \downarrow 0} \inf_{x \in D_1^\delta \cup D_2^\delta} u^\varepsilon(T^\varepsilon(\lambda), x) \leq \\ &\leq \limsup_{\varepsilon \downarrow 0} \sup_{x \in D_1^\delta \cup D_2^\delta} u^\varepsilon(T^\varepsilon(\lambda), x) \leq \limsup_{\varepsilon \downarrow 0} \sup_{x \in D_2^\delta} u^\varepsilon(T^\varepsilon(\lambda'), x). \end{aligned}$$

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

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

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

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

## 4 General case: preliminary considerations, notations and assumptions

We assume that there are  $r$  equilibrium points  $O_1, \dots, O_r$ . Recall that their domains of attraction are denoted by  $D_1, \dots, D_r$ . As follows from Lemma 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  $\Lambda = \{\lambda_0, \dots, \lambda_N\}$  such that there are limits

$$u^i(\lambda) = \lim_{\varepsilon \downarrow 0} u^\varepsilon(T^\varepsilon(\lambda), x), \quad x \in D_i, \quad \lambda \in (0, \infty) \setminus \Lambda. \quad (31)$$

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  $\lambda$  for  $\lambda > \lambda_N$ , that is the solution of the PDE tends to a constant in the appropriate time scale.

While constructing the set  $\Lambda$  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 (7) by formally setting  $u^\varepsilon$  to be identically equal to  $z$ . We drop the superscript corresponding to the initial time (since the coefficients are now time-independent) 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

$$dY_t^{z,x,\varepsilon} = b(Y_t^{z,x,\varepsilon})dt + \varepsilon \sigma(Y_t^{z,x,\varepsilon}, z)dW_t, \quad Y_0^{z,x,\varepsilon} = x, \quad (32)$$

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’re 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 (31) also belong to this interval.

**Assumption  $A_1$ .** We assume that there is a finite set of points  $Z = \{z_1, \dots, z_K\} \subseteq (g_{\min}, g_{\max})$  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  $I_1 = (g_{\min}, z_1)$ ,  $I_2 = (z_1, z_2)$ ,  $\dots$ ,  $I_{K+1} = (z_K, g_{\max})$ .

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 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.

**Assumption  $A_2$ .** 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  $z \in \bar{I}_k$ , where  $\bar{I}_k$  is the closure of  $I_k$ . We’ll assume that each  $\lambda_e$  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  $\lambda_e$  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.

**Assumption  $A_3$ .** 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  $z_k$ . Since we assumed that only one cycle is affected by the change in the hierarchy at  $z_k$ , 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  $H_k$ ).

**Assumption  $A_4$ .** We 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) Consider the case when  $\mathcal{N}_{k+1}(C_{k+1}^-(g)) \in C_k^+(e)$ . 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.)

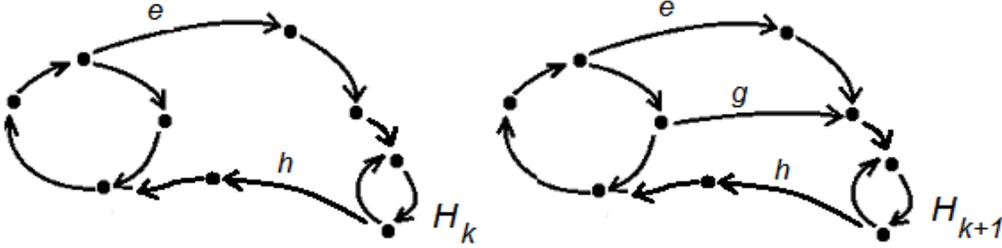


Figure 2: Restructuring of the first kind - the hierarchies  $H_k$  and  $H_{k+1}$ .

(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 3.)

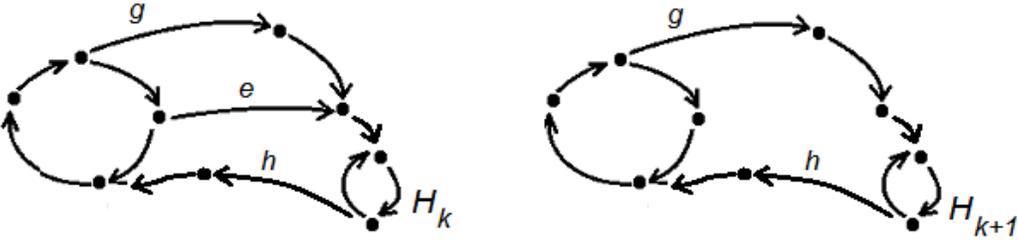


Figure 3: Restructuring of the second kind - the hierarchies  $H_k$  and  $H_{k+1}$ .

(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 the hierarchy undergoes the restructuring of the third kind at  $z_k$  in this case. (See Figure 4.)

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  $H_k$ .

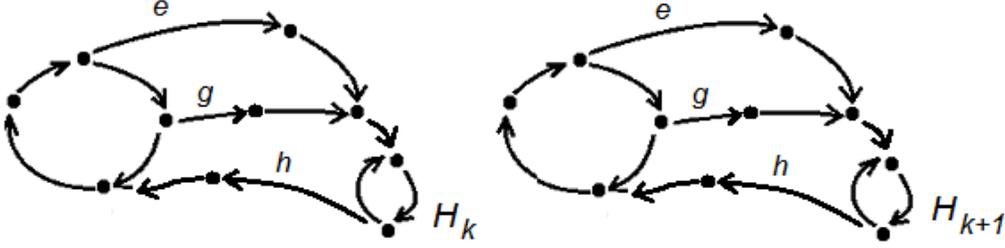


Figure 4: Restructuring of the third kind - the hierarchies  $H_k$  and  $H_{k+1}$ .

### Special time scales.

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

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

**Assumption  $B_1$ .** 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  $M$ . Define

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

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

**Assumption  $B_2$ .** We'll assume that  $M$  is finite and that no two functions  $\lambda_{e_1}, \lambda_{e_2}$  have a local maximum at the same point  $z \in I_k$  for distinct  $e_1, e_2$ . We also assume that the number of elements in  $\Lambda_1$  is the same as in  $M$ , i.e., no two maximum points result in the same value of  $\lambda_e$ .

We'll say that  $z \in I_k$  is an intersection point if there are  $e, g \in H_k$  such that  $\lambda_e(z) = \lambda_g(z)$ . Let's denote the set of intersection points by  $N$ . Define

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

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

**Assumption  $B_3$ .** 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). We also assume that the number of elements in  $\Lambda_2$  is the same as in  $N$ , i.e., no two intersection points result in the same value of  $\lambda_e$ .

Define

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

i.e.,  $\Lambda_3$  is determined by the intersections of the graphs of  $\lambda_e$  for all  $e$  with the vertical lines given by  $z = z_k$  for all  $k$ .

**Assumption  $B_4$ .** We'll assume that  $\Lambda_3^{k_1}$  and  $\Lambda_3^{k_2}$  don't intersect for distinct  $k_1$  and  $k_2$ . Define

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

**Assumption  $B_5$ .** We'll assume that  $\Lambda_0, \dots, \Lambda_3$  are disjoint.

### The cascade of special time scales.

So far, we operated with the hierarchies  $H_k$  corresponding to the process (32) for different values of  $z$ . The motivation is that processes of this type come up if we consider the value of the function  $u^\varepsilon$  in (7) to be fixed and identically equal to  $z$ . The full picture is somewhat more complicated. For each value of  $\lambda$ , 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  $\lambda$ , 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  $\Lambda_1, \dots, \Lambda_3$ . Besides the sets  $\Lambda_1, \dots, \Lambda_3$ , there are other "special" time scales that we'll now define inductively.

Let  $\bar{\Lambda} = \Lambda_0 \cup \dots \cup \Lambda_3$  and  $\bar{\Pi} = \Pi_0 \cup \dots \cup \Pi_3$ . 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  $(z, \lambda)$  plane starting at  $p$ . All the intersection points of  $l$  with the graphs of  $\lambda_e$  (for all  $e$ ) are considered points of the second generation.

**Assumption  $B_6$ .** 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  $\tilde{l}$  be the horizontal ray in the  $(z, \lambda)$  plane that starts at a point of the second generation (intersection of  $l$  with the graph of  $\lambda_e$ ) and goes in the direction where the function  $\lambda_e$  decreases. All the other intersection points of the ray  $\tilde{l}$  with the graphs of  $\lambda_g$  (for all  $g$ , where  $\lambda_g$  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  $\lambda_e$  decreases (the latter is defined due to Assumption  $B_6$ ). The set of points of all generations obtained in this way will be called the cascade of points generated by  $p$ .

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.

**Assumption  $B_7$ .** We assume that a point of a cascade can't be obtained by following two distinct sequences of horizontal and vertical rays originating at the same point. Moreover, let  $p_1$  and  $p_2$  belong to two cascades or to the same cascade and let  $\lambda_1$  and  $\lambda_2$  be their projections on the  $\lambda$ -axis. We assume that if  $p_1$  and  $p_2$  are obtained as a result of

intersecting vertical rays from the same cascade with the graphs (say,  $\lambda_e$  and  $\lambda_g$ ), then  $\lambda_1$  and  $\lambda_2$  are distinct. We also assume that if  $p_1$  and  $p_2$  belong to different cascades, then  $\lambda_1$  and  $\lambda_2$  are distinct unless  $\lambda_1 = \lambda_2 = 0$ .

The union of cascades will be denoted by  $\Pi$ . The projection of  $\Pi$  on the  $\lambda$  axis will be denoted by  $\Lambda$ . The set of points of  $\Pi$  that were obtained by intersecting the horizontal rays with the graphs of  $\lambda_e$  will be denoted by  $\Pi_h$ , the set of points of  $\Pi$  that were obtained by intersecting the vertical rays with the graphs of  $\lambda_e$  will be denoted by  $\Pi_v$ . Thus  $\Pi = \bar{\Pi} \cup \Pi_h \cup \Pi_v$ .

**Lemma 4.1.** *If Assumptions  $A_1$ - $A_4$  and  $B_1$ - $B_7$  hold, then the number of points in  $\Pi$  is finite.*

*Proof.* Since the number of equilibrium points  $O_1, \dots, O_r$  is finite, and due to Assumptions  $A_1$ ,  $B_2$ , and  $B_3$ , the number of points in  $\bar{\Pi}$  is finite. It is equal to the the number of cascades, which is then also 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  $\lambda_e$ , 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, by Assumption  $A_2$ , the functions  $\lambda_e$  are continuous on  $\bar{I}_k$  (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  $\lambda_e$  is defined). For a point  $p \in \Pi_2 \cup \Pi_3$ , let  $S_p(\delta)$  denote the square in the  $(z, \lambda)$  plane with side  $\delta$  and centered at  $p$ . By Assumptions  $A_4$  and  $B_3$ , for each  $p = (z', \lambda') \in \Pi_2 \cup \Pi_3$  there are at most two edges for which  $\lim_{z \uparrow z'} \lambda_e(z) = \lambda'$  and at most two edges for which  $\lim_{z \downarrow z'} \lambda_e(z) = \lambda'$ . By Assumption  $A_2$ , 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  $\delta$  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  $\lambda$ ),

as follows from Assumptions  $B_3, B_4$ , and  $B_5$ , we can choose  $\delta$  sufficiently small so that  $[\lambda' - \delta/2, \lambda' + \delta/2]$  does not overlap with  $[\lambda'' - \delta/2, \lambda'' + \delta/2]$ , where  $\lambda'$  and  $\lambda''$  are the  $\lambda$ -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  $\lambda_e$  are continuous and intersect only at the points of  $\Pi_2 \cup \Pi_3$ , there is a positive constant  $\kappa$  such that a vertical edge starting outside  $\cup_p S_p(\delta)$  is longer than  $\kappa$ . 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  $\lambda_e$  that was a part of our construction (see Figure 5).

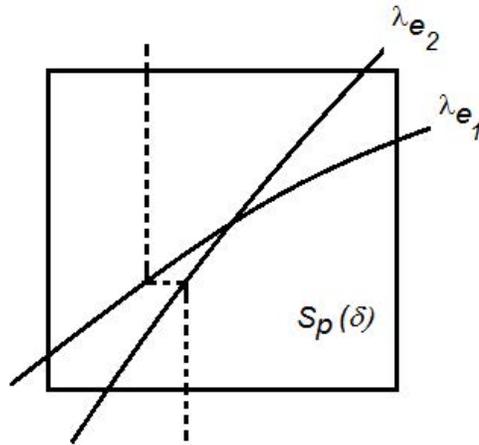


Figure 5: A path passing through  $S_p(\delta)$  - the dotted line represents a part of the path.

The next time the path returns to  $S_p(\delta)$  (if it returns at all), it will be higher by at least  $\kappa$  than when it left  $S_p(\delta)$ . Therefore, each path will have at most  $2(1 + \delta/\kappa)$  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  $\kappa$ . 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  $\lambda_e$  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  $\lambda_e$ , 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  $\lambda_e$ , 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.  $\square$

## 5 Inductive construction and the main result

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

**Theorem 5.1.** *Under the above assumptions ( $A_1$ - $A_4$  and  $B_1$ - $B_7$ ), for each  $\lambda \in (0, \infty) \setminus \Lambda$  there are limits*

$$u^i(\lambda) = \lim_{\varepsilon \downarrow 0} u^\varepsilon(T^\varepsilon(\lambda), x), \quad x \in D_i, \quad (33)$$

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  $(\lambda_N, \infty)$ .

The construction below generalizes the situation with two equilibrium points considered in Section 3. Our emphasis is on providing the algorithm for constructing the functions  $u^i$  found in the left hand side of (33). The fact that the limits in (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,  $G_0, G_1, \dots, G_N$  associated to the segments  $(0, \lambda_1), (\lambda_1, \lambda_2), \dots, (\lambda_N, \infty)$ , and a sequence of directed labeled graphs  $G_1^*, \dots, G_N^*$  associated to the points  $\lambda_1, \dots, \lambda_N$ . They will be constructed inductively:  $G_n^*$ ,  $n = 1, \dots, N$ , 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  $O_1, \dots, O_r$ . Each of the directed edges will carry a label – active, engaged (to the right or to the left), or conflicted, as will be described below.

Let  $\lambda \in (\lambda_n, \lambda_{n+1})$ ,  $0 \leq n \leq N$ , where we put  $\lambda_0 = 0$  and  $\lambda_{N+1} = \infty$ . Assume that  $u^i(\lambda)$ ,  $1 \leq i \leq r$ , have been defined. Consider the hierarchy of cycles corresponding to the linear equation (7) with the second argument in  $\sigma$  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  $\lambda_e$  and  $\lambda_g$  are the transition scales.

We define  $G_n$  to contain those edges of  $\overline{G}_n$  for which  $\lambda_e \leq \lambda$ . We'll see that the labeling of the edges (which will be done via an inductive procedure below) has the

following properties: an edge  $e \in G_n$  is active if  $e = \mathcal{R}(C)$  for some  $C$  (the edge leading out of  $C$ ),  $\mathcal{N}(C)$  is defined uniquely and satisfies  $\lambda_e < \lambda$ . Edges  $e$  and  $g$  are conflicted if  $e = g = \mathcal{R}(C)$  (i.e., the notion of next is not uniquely defined) and  $\lambda_e = \lambda_g < \lambda$ . An edge  $e$  is engaged to the right (or 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  $u^i(\lambda) < u^j(\lambda)$  (or  $u^i(\lambda) > u^j(\lambda)$ , respectively).

Consider an example with two equilibria corresponding to Figure 1. In that example, there are intervals  $(\lambda_0, \lambda_1)$ , ...,  $(\lambda_4, \infty)$  with the corresponding graphs  $G_0, \dots, G_4$ .  $G_0$  does not contain edges.  $G_1$  contains an edge  $e$  leading from  $O_1$  to  $O_2$ , which is engaged to the right.  $G_2$  and  $G_3$  in addition to that contain an edge  $g$  leading from  $O_2$  to  $O_1$ , which is engaged to the left. Finally, both edges  $e$  and  $g$  are active for  $G_4$ .

We'll separate each graph  $G_n$  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. In particular, there are no edges leading out of a sleeping cluster to a different 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. An example in Figure 6 shows a hypothetical graph with edges of different types. The active edges are denoted by solid lines, conflicted edges by dotted lines, and engaged edges by dashed lines.

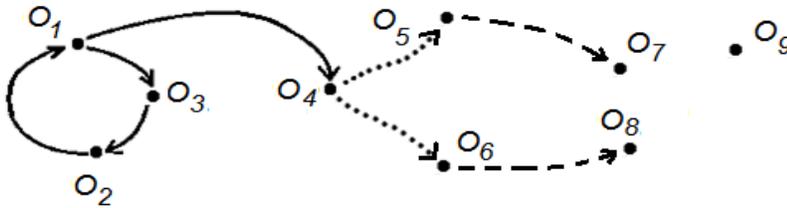


Figure 6: In this graph, vertices  $O_1, O_2, O_3,$  and  $O_4$  form a conflicted cluster,  $O_5$  and  $O_6$  are separate engaged clusters, while  $O_7, O_8,$  and  $O_9$  are separate sleeping clusters.

**Remark 5.2.** *Based on the inductive construction below, 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.*

We'll see that for a sleeping cluster  $u^i(\lambda)$  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  $p \in (\Pi_0 \cup \Pi_2 \cup \Pi_h) \cap \{(z, \lambda) : \lambda < \lambda_n\}$  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  $\lambda$  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, (31) holds for  $\lambda \in (0, \delta)$  with the right hand side equal to  $g(O_i)$  for some positive  $\delta$ , which explains our definition of  $u_+^i(\lambda_0)$ . We put  $G_0$  to be the graph with the vertices at  $O_1, \dots, O_r$  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  $z \in (u_+^i(\lambda_n), u_+^i(\lambda_n) + \delta)$  for some positive  $\delta$ . 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  $\delta'$  can be chosen. By the definition of the sets  $\Lambda_1$  and  $\Lambda_3$ , the function  $\lambda_e(z)$  will keep increasing at least until it reaches the value  $\lambda_{n+1}$ . 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 the 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  $\lambda_n$  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  $\Gamma_1, \Gamma_2, \dots, \Gamma_M$ . 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  $\Gamma_{m_1}$  and  $\Gamma_{m_2}$ , 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  $\Gamma_{m_2}$ , while  $g$  leads to a vertex in  $\Gamma_{m_1}$ . Then both  $e$  and  $g$  become active for  $G_n^*$ , i.e., the clusters  $\Gamma_{m_1}$  and  $\Gamma_{m_2}$  merge forming a sleeping cluster (See Figure 7).



Figure 7: Two clusters consisting of  $O_1, O_2, O_3, O_4$  on the one hand, and  $O_5, O_6$  on the other hand, which are engaged in  $G_{n-1}$ , merge to form a sleeping cluster in  $G_n^*$ .

2) Suppose that everything is as above, except that  $e$  leads to a vertex in  $\Gamma_{m_2}$ , while  $g$  leads to a vertex outside  $\Gamma_{m_1}$ . In this case only  $e$  becomes active, while  $g$  stays engaged, i.e., the clusters merge forming an engaged cluster (See Figure 8).

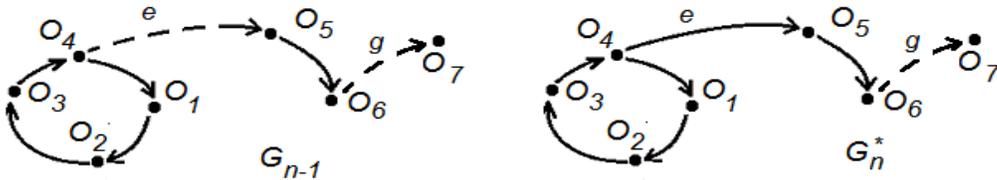


Figure 8: Two clusters consisting of  $O_1, O_2, O_3, O_4$  on the one hand, and  $O_5, O_6$  on the other hand, which are engaged in  $G_{n-1}$ , merge to form one engaged cluster in  $G_n^*$ .

3) Suppose that  $e$  is engaged leading out of an engaged cluster  $\Gamma_m$ , while  $g$  is an active edge within  $\Gamma_m$  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  $\Gamma_m$  falls apart in two engaged clusters (See Figure 9). It becomes engaged to the right if  $\lambda_g$  is increasing at  $z$  and engaged to the left if  $\lambda_g$  is decreasing at  $z$ .

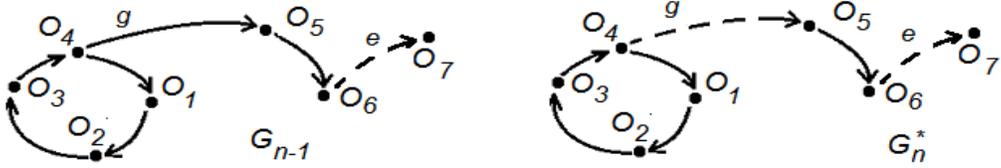


Figure 9: The cluster consisting of  $O_1, \dots, O_6$  breaks apart into two engaged clusters.

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  $\Gamma$  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  $\Gamma$  is engaged to the right for  $\lambda \in (\lambda_{n-1}, \lambda_n)$  (the case of  $\Gamma$  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  $\Gamma$  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 (See Figure 10).

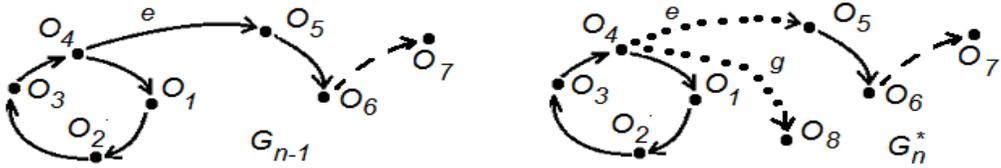


Figure 10: A part of an engaged cluster breaks away to form a conflicted cluster.

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  $\Gamma$  stays engaged to the right, while the edge  $e$  disappears (See Figure 11).

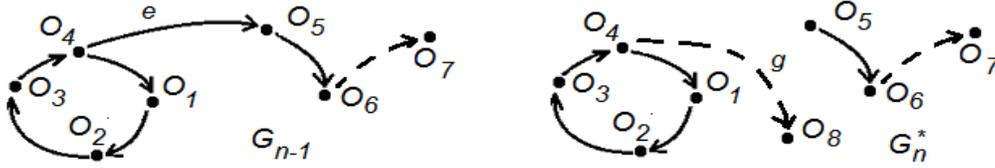


Figure 11: A part of an engaged cluster breaks away to form an engaged cluster.

Suppose that  $w_-^j(\lambda_n) = z_k$ . In this case  $O_j$  must belong to  $\Gamma$  as follows from the earlier assumptions on the function  $\lambda_e$ . If  $g$  was not represented in  $G_{n-1}$ , then it gets added to the cluster  $\Gamma$ , 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 (see Figure 12) and third (see Figure 13) kind,  $e$  remains in  $\Gamma$  if and only if  $O_j$  is subordinated to  $C$  in  $G_{n-1}$ .

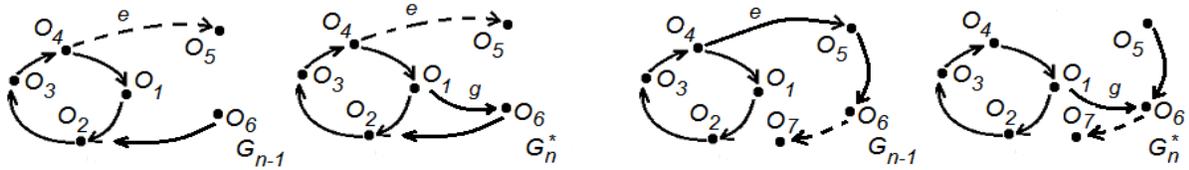


Figure 12: Restructuring of the first kind: two possible scenarios.

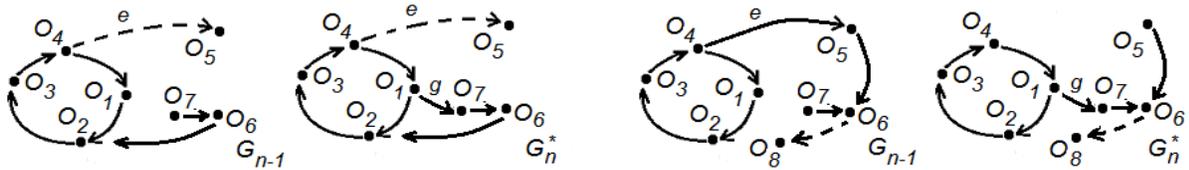


Figure 13: Restructuring of the third kind: two possible scenarios.

In the case of restructuring of the second kind,  $e$  is not represented in  $G_n^*$  (see Figure 14).

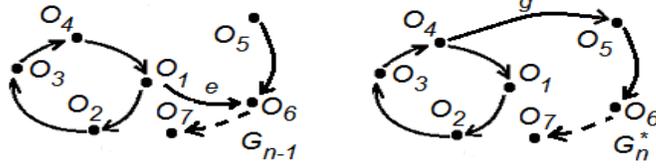


Figure 14: Restructuring of the second kind.

2) Suppose that  $C$  was a part of a conflicted cluster  $\Gamma$ . 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  $\Gamma_{m_1}$  and  $O_{j_2}$  belongs to a cluster  $\Gamma_{m_2}$  for the graph  $G_{n-1}$ . Suppose that  $\Gamma_{m_1}$  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  $\Gamma_{m_2}$  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  $\Gamma_{m_1}$  and  $\Gamma$  from the graph  $G_{n-1}$  become one cluster for the graph  $G_n^*$ . If  $h$  leads to a vertex within the cluster  $\Gamma$ , 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  $G_n^*$  (See Figure 15).



Figure 15: A conflicted cluster becomes engaged.

If  $h$  leads to a vertex outside the cluster  $\Gamma$ , then it stays an engaged edge within the larger cluster,  $g$  disappears, and  $e$  becomes an active edge within  $\Gamma \cup \Gamma_{m_1}$  (See Figure 16).

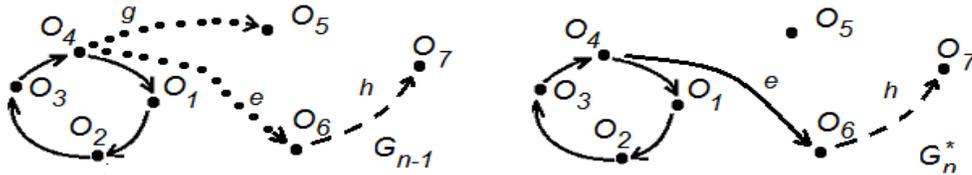


Figure 16: A conflicted cluster becomes engaged. One of the conflicted edges disappears.

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  $\Gamma_{m_3}$  in  $G_{n-1}$  that is engaged via an edge  $h$  to a vertex in  $\Gamma$  and

is such that  $u_-^i(\lambda_n) = z_k$  for  $O_i \in \Gamma_{m_3}$ . In this case the clusters  $\Gamma_{m_3}$  and  $\Gamma$  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  $G_n^*$  (see Figure 17).

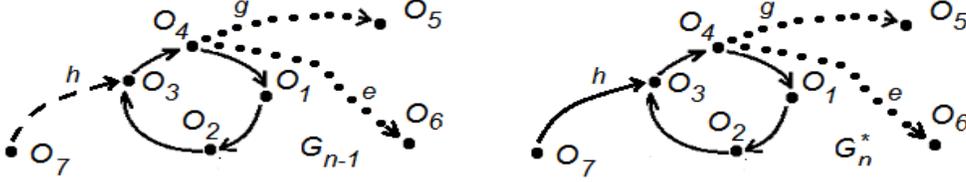


Figure 17: An engaged cluster gets attached to a conflicted one.

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  $\lambda_n$  was obtained as a part of a cascade. Namely, assume that there is  $z \in I_k$  such that  $(z, \lambda_n)$  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  $\Gamma$  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  $\Gamma$  to a vertex  $O_j$ . If  $u_-^j(\lambda_n) = z$  (in which case  $O_j$  belongs to  $\Gamma$ ), then  $e$  is added as an active edge, and  $\Gamma$  with this extra edge remains a sleeping cluster within  $G_n^*$  (see Figure 18).

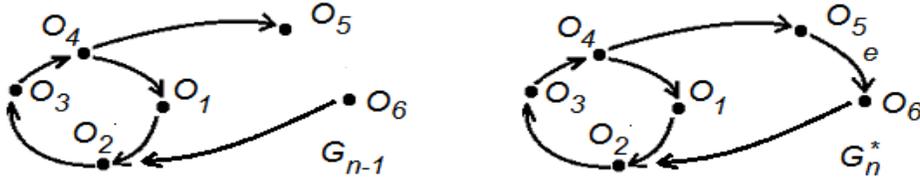


Figure 18: An active edge is added to a sleeping cluster.

If  $u_-^j(\lambda_n) < z$  (or  $u_-^j(\lambda_n) > z$ ), then the edge  $e$  is added as an engaged edge and  $\Gamma$  becomes engaged to the left (right) in  $G_n^*$  (see Figure 19).

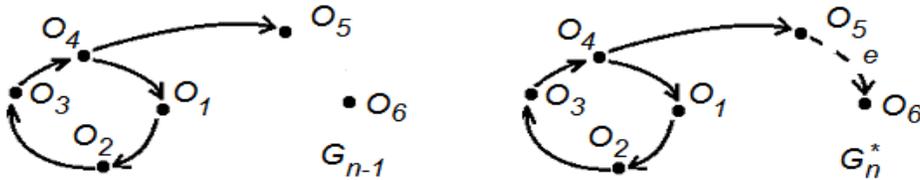


Figure 19: A sleeping cluster becomes engaged.

2) Suppose that  $G_{n-1}$  contains a cluster  $\Gamma_m$  distinct from  $\Gamma$  such that  $u_-^i(\lambda_n) = z$  for  $O_i \in \Gamma_m$ . Suppose that  $\Gamma_m$  is engaged to a vertex in  $\Gamma$  via an edge  $h$ . Then  $\Gamma_m$  and  $\Gamma$  from the graph  $G_{n-1}$  become one cluster for the graph  $G_n^*$ . The edge  $h$  becomes active and the cluster  $\Gamma \cup \Gamma_m$  is sleeping in  $G_n^*$  (see Figure 20).



Figure 20: An engaged cluster gets attached to a sleeping one.

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  $\Gamma$  be a cluster within  $G_n^*$  such that  $u_-^i(\lambda_n) = z^*$  for  $O_i \in \Gamma$ . Suppose that  $\Gamma$  is engaged via an edge  $e$ .

From our construction above it follows that  $G_n^*$  contains at most once such cluster  $\Gamma$  with the following property: either  $\Gamma$  is engaged to the right and  $\lambda_e(z) < \lambda_n$  for  $z \in (z^*, z^* + \delta)$  for some positive  $\delta$  or  $\Gamma$  is engaged to the left and  $\lambda_e(z) < \lambda_n$  for  $z \in (z^* - \delta, z^*)$  for some positive  $\delta$ . In this case we'll say that  $\Gamma$  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  $\Pi$  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  $\bar{z}_1, \dots, \bar{z}_L$ .

We'll describe the transition from  $G_n^*$  to  $G_n$  using a two-part induction procedure, the first part being induction on  $l = 1, \dots, L$  (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  $\Gamma$  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  $\Gamma$  now denote the special cluster within  $G_n^*$ . At the first 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) The special cluster  $\Gamma$  merges with a previously engaged cluster  $\Gamma_m$  to form a sleeping cluster. Namely, suppose that  $G_n^*$  contains an engaged cluster  $\Gamma_m$  with  $u(\Gamma_m) = \bar{z}_1$  such that  $\Gamma_m$  is engaged to a vertex in  $\Gamma$  via an edge  $h$ . Suppose that  $e$  leads to a vertex in  $\Gamma_m$ . Then  $\Gamma_m$  and  $\Gamma$  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). The situation can be illustrated by Figure 7, in which  $G_{n-1}$  and  $G_n^*$  need to be replaced by  $G_n^*$  and ‘the new transitional graph’, respectively. The letters above the edges also need to be adjusted to the current notations.

2) A previously engaged cluster  $\Gamma_m$  attaches to the special cluster  $\Gamma$ . Namely, suppose that  $G_n^*$  contains an engaged cluster  $\Gamma_m$  with  $u(\Gamma_m) = \bar{z}_1$  such that  $\Gamma_m$  is engaged to a vertex in  $\Gamma$  via an edge  $h$ . Suppose that  $e$  leads to a vertex in a cluster  $\Gamma_{m_1}$  with  $u(\Gamma_{m_1}) > \bar{z}_1$ . Then  $\Gamma_m$  and  $\Gamma$  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. Up to notation for the edges, the situation can be illustrated by Figure 8, in which  $G_{n-1}$  and  $G_n^*$  need to be replaced by  $G_n^*$  and ‘the new transitional graph’, respectively.

3) The special cluster  $\Gamma$  merges with an engaged cluster  $\Gamma_m$  to form an engaged cluster. Namely, suppose that  $G_n^*$  contains an engaged cluster  $\Gamma_m$  with  $u(\Gamma_m) = \bar{z}_1$  such that  $\Gamma_m$  is engaged to a vertex outside  $\Gamma$  via an edge  $h$ . Suppose that  $e$  leads to a vertex in  $\Gamma_m$ . Then  $\Gamma_m$  and  $\Gamma$  form one engaged cluster in the transitional graph, which is engaged via the edge  $h$ . The edge  $e$  becomes active in this cluster. This situation can also be illustrated by Figure 8, the difference from the previous case being that the resulting cluster now doesn’t have the ‘special’ property.

4) The special cluster  $\Gamma$  merges with a conflicted cluster  $\Gamma_m$  to form another special cluster. Namely, suppose that  $G_n^*$  contains a conflicted cluster  $\Gamma_m$  with  $u(\Gamma_m) = \bar{z}_1$  (in which case  $\bar{z}_1$  coincides with one of the points  $z_k$ ) such that  $\Gamma_m$  has conflicted edges  $g$  and  $h$ . Suppose that  $g$  leads to a vertex in  $\Gamma$ . Suppose that  $e$  leads to a vertex in  $\Gamma_m$ . Then  $\Gamma_m$  and  $\Gamma$  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. Here, up to notation for the edges, the situation is shown in Figure 15.

If, on the other hand,  $e$  leads to a vertex outside in  $\Gamma_m$ , then  $\Gamma_m$  and  $\Gamma$  still form one cluster in the transitional graph, the edge  $h$  disappears,  $g$  becomes active, and  $e$  remains engaged to the right. The new cluster still has the ‘special’ property. Here, up to notation for the edges, the situation is shown in Figure 16.

5) The special cluster  $\Gamma$  attaches to a conflicted cluster  $\Gamma_m$ . Namely, suppose that  $G_n^*$  contains a conflicted cluster  $\Gamma_m$  with  $u(\Gamma_m) = \bar{z}_1$  (in which case  $\bar{z}_1$  coincides with one of the points  $z_k$ ) such that  $\Gamma_m$  has conflicted edges  $g$  and  $h$ . Suppose that neither  $g$  nor  $h$  lead to a vertex in  $\Gamma$ . Suppose that  $e$  leads to a vertex in  $\Gamma_m$ . Then  $\Gamma_m$  and  $\Gamma$  form one conflicted cluster in the transitional graph, where the edge  $e$  becomes active. The new cluster does not have the ‘special’ property. Here, up to notation for the edges, the situation is shown in Figure 17.

6) The special cluster  $\Gamma$  attaches to a sleeping cluster  $\Gamma_m$ . Namely, suppose that  $G_n^*$  contains a sleeping cluster  $\Gamma_m$  with  $u(\Gamma_m) = \bar{z}_1$ . Suppose that  $e$  leads to a vertex in  $\Gamma_m$ . Then  $\Gamma_m$  and  $\Gamma$  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. Up to notation for the edges, this situation is shown in Figure 20.

7) An edge within the special cluster  $\Gamma$  becomes engaged instead of active or  $\Gamma$  itself loses the special property. Namely, suppose that  $\Gamma$  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  $\Gamma$  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  $\Gamma$  (if it is non-empty) forms an engaged special cluster in the transitional graph. Up to notation for the edges, this situation is shown in Figure 9.

8) The notion of “next” changes for the special cluster  $\Gamma$  or its sub cluster. Namely, suppose that  $\bar{z}_1$  coincides with one of the points  $z_k$ . Suppose that  $\Gamma$  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.,  $R_{k+1}(C) = g$ ,  $N_{k+1}(C) = O_j$ .

8a) A sub cluster of  $\Gamma$  becomes conflicted. Namely, 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  $\Gamma$  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. Up to notation for the edges, this situation is shown in Figure 10.

8b) A sub cluster of  $\Gamma$  breaks off to form another special cluster. Namely, 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  $\Gamma$  stays engaged to the right, while the edge  $e'$  disappears. There are now up to two special clusters in the transitional graph. Up to notation for the edges, this situation is shown in Figure 11.

8c) A sub cluster of  $\Gamma$  attaches to a different vertex within  $\Gamma$ . Namely, suppose that  $O_j \in \Gamma$ . If  $g$  was not represented in  $G_n^*$ , then it gets added to the cluster  $\Gamma$ , which remains a special cluster 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  $\Gamma$  if and only if  $O_j$  is subordinated to  $C$  in  $G_n^*$  (see Figures 12 and 13). In the case of restructuring of the second kind,  $e'$  is not represented in the transitional graph (see Figure 14, again, up to notation for the edges).

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  $\bar{z}_1$ . 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  $\bar{z}_1$  replaced by  $\bar{z}_2$ . 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  $\bar{z}_2$  to  $\bar{z}_3$ , etc., until we reach  $\bar{z}_L$ . Observe that there are no special clusters with the value  $\bar{z}_L$ .

The second part of the inductive construction concerns the fact that once we reach  $\bar{z}_L$ , the resulting graph may still have special clusters. The values  $u(\Gamma)$  for those clusters

belong to the set  $\{\bar{z}_1, \dots, \bar{z}_L\}$ , 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  $\{\bar{z}_2, \dots, \bar{z}_L\}$ . Continuing by induction on  $l$ , we can get rid of all the special clusters. The last transitional graph will be denoted by  $G_n$  and the values  $u(\Gamma)$  for this graph give us  $u_+^i(\lambda_n)$ . This concludes the inductive construction of  $G_n$  and  $u_+^i(\lambda_n)$ .  $\square$

Now, that our inductive construction is complete, let us make an observation concerning the values of  $u^i(\lambda)$  for  $\lambda \in (\lambda_N, \infty)$ .

**Theorem 5.3.** *There is a constant  $\bar{u}$  such that for each  $\lambda > \lambda_N$  and each  $x \in \mathbb{R}^d$  there is the limit*

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

*Proof.* 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  $\lambda_e$  has a local maximum that is larger than  $\lambda_N$  or the graph of  $\lambda_e$  intersects one of the vertical lines  $z = z_k$  at a point that is higher than  $\lambda_N$ . In either of these cases,  $\lambda_N$  can't be the maximum of the set  $\Lambda$  due to our definition of the sets  $\Lambda_1$  and  $\Lambda_3$ . 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,  $u^i(\lambda) = \bar{u}^i$  for some constants  $\bar{u}^i$ .

Fix a value  $\bar{\lambda} > \lambda_N$ , and observe that on the interval  $[T^\varepsilon(\bar{\lambda}), \infty)$  the function  $u^\varepsilon$  satisfies the equation that is almost linear, i.e., the diffusion matrix in (1) is close to  $a(x, \bar{u}^i)$  for  $x \in D_i$ . We can consider equation (1) on the interval  $[T^\varepsilon(\bar{\lambda}), \infty)$  with initial data that is close to  $\bar{u}^i$  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  $\lambda$  is sufficiently large. This implies that  $\bar{u}^i$  do not depend on  $i$ , that is  $\bar{u}^i = \bar{u}$  and all  $i$ .

Strictly speaking, the fact that  $u^i(\lambda)$  do not depend on  $i$  (or  $\lambda$  for  $\lambda > \lambda_N$ ) 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.  $\square$

**Acknowledgements:** While working on the paper, L. Koralov was partially supported by the Simons Fellowship in Mathematical Sciences as well as the NSF grant number 1309084.

## References

- [1] Freidlin M. I., *Sublimiting Distributions and Stabilization of Solutions of Parabolic Equations with a Small Parameter*, Soviet Math. Dokl., 235, 5, pp

- 1042–1045, 1977.
- [2] Freidlin M.I., *Functional Integration and Partial Differential Equations*, Princeton University Press, 1985.
  - [3] Freidlin M. I., *Quasi-deterministic Approximation, Metastability and Stochastic Resonance*, Physica D, 137, pp 333-352, 2000.
  - [4] Freidlin M. I., Korolov L. *Nonlinear Stochastic Perturbations of Dynamical Systems and Quasi-linear Parabolic PDE's with a Small Parameter*, Probability Theory and Related Fields (2010), 147, pp 273-301.
  - [5] Freidlin M. I., Korolov L. *Metastability for Nonlinear Random Perturbations of Dynamical Systems*, Stochastic Processes and Applications 120 (2010), no. 7, 1194-1214.
  - [6] Freidlin M. I., Wentzell A. D., *Random Perturbations of Dynamical Systems*, Springer 1998.
  - [7] Krylov N.V., *Nonlinear Elliptic and Parabolic Equations of the Second Order (Mathematics and its Applications)*, Springer 1987.
  - [8] Ladyzenskaya O.A., Ural'zeva N.N., *Linear and Quasi-Linear Parabolic Equations* (Russian), Nauka 1967.
  - [9] Oliveira E., Vares M.E. *Large Deviations and Metastability*, Cambridge University Press, 2005.