Small Random Perturbations of Dynamical Systems: Exponential Loss of Memory of the Initial Condition

Fabio Martinelli[1] and Elisabetta Scoppola[2]

[1] Dipartimento di Matematica, Universita' "La Sapienza", Pz. A. Moro 2, I-00185 Roma, Italy
[2] Dipartimento di Fisica, Universita' "La Sapienza", Pz. A. Moro 2, I-00185 Roma, Italy

Abstract. We consider Markov processes arising from small random perturbations of non-chaotic dynamical systems. Under rather general conditions we prove that, with large probability, the distance between two arbitrary paths starting close to a same attractor of the unperturbed system decreases exponentially fast in time. The case of paths starting in different basins of attraction is also considered as well as some applications to the analysis of the invariant measure and to elliptic problems with small parameter in front to the second derivatives. The proof is based on a multiscale analysis of the typical trajectories of the Markov process; this analysis is done using techniques involved in the proof of Anderson localization for disordered quantum systems.

Section 0. Introduction

This paper is concerned with the study of the dependence on the initial condition of some Markov processes X_t^ε which arise from small random perturbations of dynamical systems. These kind of processes arise more and more frequently in different areas of natural sciences like theoretical physics, statistical mechanics, chemistry, ergodic theory and ecology (see e.g. [1–5]) and their mathematical theory has been developed first by Ventzel and Freidlin in their basic work [6] and by Kiefer [7]. Following their ideas Gora analyzed the discrete case [8].

More precisely we will treat two different examples:

a) the discrete case, $t \in N$ in which the Markov chain X_t^ε is obtained by randomly composing two maps T and T^ε from a compact differentiable manifold M, with T^ε very close to the identity as $\varepsilon \to 0$.

b) the continuous case when the process X_t^ε is a diffusion process in R^n, solution of an ordinary Ito equation:

$$dX_t^\varepsilon = b(X_t^\varepsilon)dt + \varepsilon dw_t. \tag{0.1}$$

The precise hypotheses on T and T^ε are stated in Sects. 1 and 2. Basically we assume that the deterministic dynamical systems $x \to T(x)$ or $dx_t = b(x_t)dt$ have only a finite number of asymptotically stable fixed points or periodic orbits with

negative Lyapunov exponent. A typical example is given by $b(x) = -\nabla U(x)$ with $U(x) = x^4/4 - x^2/2$. A critical discussion of our hypotheses is contained in the first paragraph of Sect. 1 and mainly in Sect. 2.

The general problem we want to address is the analysis of the behaviour of the random map $x \to X_t^\varepsilon(x)$ for any large enough t, where $X_t^\varepsilon(x)$ is the process at time t starting at x.

In particular we ask whether with large probability the process loses memory of the initial condition x path by path at least when x is in a small neighborhood C of the attractors of the unperturbed system. By this we mean that the map $x \to X_t^\varepsilon(x)$ becomes almost constant on the set C as time increases and ε is small but uniform in t.

Another alternative way to look at this problem is to ask whether two arbitrary initial conditions x and y close to the same attractor will eventually glue together under the *same* random evolution.

This problem is, to our knowledge, completely new and any answer to the above question is clearly relevant in all those problems of probability theory in which the loss of memory of the initial condition plays an important role, e.g. the proof of the exponential distribution of the exist time of the process from a completely attracted domain (see [9, 10 and especially 11] where an infinite dimensional case is also discussed), the detailed study of the invariant measure of the process (see Sect. 1, Theorem 1.2) and the behaviour of the solution of elliptic boundary value problems with small parameter in front of the highest derivatives (see Sect. 3). Moreover these kind of questions seem to be relevant also in statistical mechanics of disordered systems, e.g. in the study of the dynamics of some spin-glass models and neural-networks (see [12]). However our results do not yet cover these last examples.

Clearly our problem belongs to the more general context of the study of stochastic flows; we refer the reader to the book by Kiefer [13]. In particular for the flows considered in this paper the general theory of the characteristic exponents (Lyapunov exponent) has been developed in [14, 15] (see also [16]) and some concrete examples were provided with the negative largest Lyapunov exponent. The questions that we ask are somehow related to the problems discussed in these last papers but our point of view and the techniques that we adopt are completely different.

Our main result can be formulated as follows. Let C_i be a sufficiently small neighborhood of the attractor x_i (in the discrete case x_i can be replaced by an attracting periodic orbit). Then there are positive constants m, k, t_0 such that for all sufficiently small ε,

$$P\left(\sup_{x,y \in C_i} \frac{\text{dist}(X_t^\varepsilon(x), X_t^\varepsilon(y))}{\text{dist}(x,y)} \leq e^{-mt} \quad \text{for all } t > t_0\right) > 1 - \exp\left(-\frac{k}{\varepsilon^2}\right). \quad (0.2)$$

In order to understand which problems are behind a result like the one above let us first consider a trivial example in one dimension,

$$dX_t^\varepsilon = -mX_t^\varepsilon dt + \varepsilon dw_t, \quad (0.3)$$

where $m > 0$ and w_t is the standard Brownian motion. If we denote by $\phi(t) = X_t^\varepsilon(x) -$

$X_t^\varepsilon(y)$, then we get

$$d\phi(t) = -m\phi(t)dt; \quad \phi(0) = x - y,$$

that is

$$\phi(t) = (x - y)e^{-mt}.$$

Thus in this case the result on the glueing of the paths is trivially true. However if instead of the simple harmonic oscillator we consider a double well situation e.g.:

$$b(x) = -\nabla U(x), \quad U(x) = x^4/4 - x^2/2,$$

then the previous argument does not work anymore. The reason is that in this new situation, between the two attractors $x = +/-1$, we have an unstable fixed point at $x = 0$. Around this point the two paths $X_t^\varepsilon(x), X_t^\varepsilon(y)$ tend to separate exponentially fast and in principle the effect can destroy the glueing produced by the attractors. However it follows from the theory of Ventzel and Freidlin that the typical time spent by the process X_t^ε around $x = 0$ during a jump from one attractor to the other is much smaller than the time spent around the attractors. Therefore one can conceive that the glueing is still taking place.

In the above example it is also interesting to see whether the glueing of the paths occur only for initial conditions x and y close to the same attractor or also for x and y belonging to different basins of attraction. This is what we call "global glueing", and it is proved in several general situations. It is important however to observe that global glueing may take place only because of the noise in the system and that it can be seen only after a time scale exponentially large in ε.

It is important to outline two basic features of our result:

  i) the supremum over the initial conditions is inside the probability,
  ii) the event considered belongs to the σ-algebra at infinity.

These are the main qualitative differences between our result and the known results in the theory of small random perturbations of dynamical systems, and are also the main sources of technical problems.

Taking the supremum over the initial conditions inside the probability may cause problems as it clearly emerges from the following example. Coming back to our double well potential let us consider initial conditions close to the top of the barrier. It follows from the Ventzel and Freidlin theory that

$$\sup_{|x| < \delta} P_x(|X_t^\varepsilon| \leq 1/2) \to 0 \quad \text{as} \quad \varepsilon \to 0$$

if $t > c\ln(1/\varepsilon)$.

However, since with large probability the paths starting to the left of the origin fall in the minimum $x = -1$, while those starting to the right fall in $x = 1$, by the continuity of $X_t^\varepsilon(x)$ with respect to x, for any $t \approx c\ln(1/\varepsilon)$ and most paths of the brownian motion εw_s, there exists an initial condition $x_0(t), |x_0| < \delta$, depending on $\varepsilon w_s, s \leq t$, such that $|X_t^\varepsilon(x_0(t))| < 1/2$. This implies

$$P\left(\sup_{|x| < \delta} |X_t^\varepsilon| \leq 1/2\right) \to 1 \quad \text{as} \quad \varepsilon \to 0.$$

However the main problems come from considering an infinite time interval. In fact in our example in an infinite amount of time fast jumps occur as well as anomalous long jumps.

Overcoming these difficulties has required a rather detailed analysis of the structure of the typical paths of the process X_t^ε. This analysis has been carried out inductively on a sequence of rapidly increasing time scales. From a probabilistic point of view we found the problem was very similar to the problem of Anderson localization at high disorder solved in [17, 18]. Therefore we borrowed the whole strategy and some important geometric constructions directly from the above papers and we organized the proof in such a way that the main probabilistic estimate becomes identical to an estimate which can be found in [17, 18, 20].

It is very interesting that ideas and techniques developed in a completely different area like statistical mechanics and quantum theory of disordered systems can be applied successfully to this kind of problem. Our approach to the problem of loss of memory gives a rather detailed analysis of the typical paths of the process X_t^ε, and the mechanism leading to the glueing emerges rather clearly. We believe that our results will contribute to simplify and to improve several results in the theory of small random perturbations of dynamical systems.

The paper is organized into three sections, one for the discrete case, the second for the continuous case and the third one with some application to elliptic problems. In turn Sects. one and two are divided into three paragraphs containing respectively the definition of the model, the statement of the main results and their proof. At the end of Sects. 1 and 3 there are two appendices A, B with some technical estimates.

Section 1. The Discrete Case

1. The Model

In this first section we examine the long time behaviour of a Markov chain obtained by randomly composing two maps T and T^ε from a connected compact differential manifold M of dimension n, with a metric d, into itself. Our main assumption is that the attractors of the map T are only fixed points and (or) periodic orbits and that T^ε is close to the identity for ε small enough. More precisely let T, T^ε be differentiable maps from M into itself such that:

hp1   i) $0 < d(T^\varepsilon(x), x) \leqq \varepsilon$ $\forall x \in M$,
         ii) T^ε is "*minimal*" i.e.

$$\bigcup_{n=1}^{\infty} (T^\varepsilon)^n(x) \text{ is dense in } M \quad \forall x \in M.$$

In order to state the assumptions on the map T we have to introduce, following the work of Gora [8], an equivalence relation in M as follows. Let for any path $x \equiv (x_0, x_1, \ldots, x_n, x_{n+1}, \ldots)$; with $x_i \in M \ \forall i$,

$$I_{ON}(x) = \sum_{i=1}^{N} d(T(x_{i-1}), x_i). \tag{1.1}$$

The quantity $I_{ON}(x)$ gives a measure of the way of x "against the current" of the map T. We also set

$$D(x, y) = \inf_{N_x} \{I_{ON}(x); x_0 = x, x_N = y\}.$$

It turns out that $D(x, y)$ is a continuous function of both variables. Then one defines an equivalence relation between two points x and y by

$$x \approx y \Leftrightarrow x = y \quad \text{or} \quad D(x, y) = D(y, x) = 0. \tag{1.2}$$

In the sequel an equivalence class with more than one point or a fixed point will be called a *basic set*. A basic set K is said to be an attractor if $\forall x \in K$, $\forall y \notin K$ $D(x, y) > 0$. With these notations we can now state:

hp2 *Any attractor basic set consists of only a finite number of points and the number of equivalence classes with more than one point is finite, as well as the number of fixed points of T. Furthermore any ω-limit set of the deterministic flow generated by T is contained in a basic set.*

Our next two assumptions will regard the structure of the maps T and T^ε in a neighborhood of the basic sets. In order to state them we first fix some notations. Let K be a basic set, and let $B(x, \delta)$ denote the ball of radius δ around x. Then we set:

$$\mathbf{K}_\delta = \bigcup_{x \in \mathbf{K}} B(x, \delta).$$

hp3 i) *There exists a constant $a_0 > 0$ such that for any ε sufficiently small and small enough $\delta > 0$ and arbitrary non-attractor K the following holds: there exists a positive integer $n = n(\delta) < a_0 \delta / \varepsilon$ such that $(T^\varepsilon)^n(x) \notin \mathbf{K}_\delta$ for any $x \in \mathbf{K}_\delta$.*
   ii) *For any non-attractor K and ε sufficiently small there exists $\delta > 0$ such that for any $\delta' < \delta$ the following holds: if $x \in \mathbf{K}_\delta$ then the T-trajectory of the point $(T^\varepsilon)^n(x)$ never visits $\mathbf{K}_{\delta'}$ where n is as in i).*

The above assumptions are taken from the work by Gora and are crucial in order to prove that the invariant measure of the Markov chain that we will construct is concentrated over the attractors of the map T.

We also remark that it is possible to find examples of maps T such that hp3 is never verified for any choice of the maps T^ε. In fact hp3 implies in particular that for any non-attractor basic set K and any $x \in K$ there exists $y \notin K$ with $d(x, y)$ arbitrarily small.

We now specify the structure of the maps T and T^ε around the attractors.

Let for any continuous map $F: M \to M$,

$$L(F, x) = \lim_{r \downarrow 0} \sup_{y \in B(x, r)} \frac{d(F(x), F(y))}{d(x, y)}.$$

If F_1 and F_2 are both differentiable we have

$$L(F_2 \circ F_1, x) \leqq L(F_2, F_1(x)) L(F_1, x). \tag{1.3}$$

hp4 *There exists a neighborhood A of the union of the attractors, denoted by K_a and a number $\beta < 1$ such that*

i) $\sup\limits_{x \in A} L(T^{n(x)}, x) \leq \beta$,

where $n(x) = \#\{y \in K(x)\}$ with $K(x)$ the attractor such that $\min\limits_{y \in K(x)} D(x, y) = 0$.

ii) *There exist $d > 0$ and two neighborhoods of $K_a C_0$, C with $C_0 \subset C \subset A$, $T(A) \subset C_0$*

and

$\min(\operatorname{dist}(\partial C_0, \partial A), \operatorname{dist}(\partial C_0, \partial C)) = d.$

iii) $\sup\limits_{x \in A} L(I - T^\varepsilon, x) \leq \beta\varepsilon$ *with* $0 \leq \beta < +\infty.$

We now introduce the Markov chain $\{X_n^\varepsilon\}_{n \geq 0}$ that will be analyzed in the subsequent paragraphs by specifying its transition probabilities. For any set $A \subset M$ and any $x \in M$ we set

$$P(x, A) = \operatorname{Prob}\{X_0^\varepsilon = x; X_1^\varepsilon \in A\} = p\chi_A(T(x)) + (1 - p)\chi_A(T^\varepsilon(x)), \qquad (1.4)$$

where $\chi_A(\ )$ denotes the indicator function of the set $A \subset M$ and p is a number between 0 and 1. In other words the story (or "path") of the chain $\{X_n^\varepsilon\}_{n \geq 0}$ is obtained by applying successively to its starting point x the map T or T^ε with probability p and $1\text{-}p$ respectively.

For notational convenience it is useful to encode the paths of X^ε with strings of 0's and 1's. We denote by $\omega_L = \{\omega_L(i)\}_{i = 1...L}$ a string of length L of 0's and 1's and we associate to it the map:

$$T_{\omega_L} = T_{\omega_L(L)} \circ \cdots \circ T_{\omega_L(1)}, \qquad (1.5)$$

with the convention that $T_0 \equiv T$ and $T_1 \equiv T^\varepsilon$. The probability of a string ω_L is denoted by $p(\omega_L)$ and it is equal to

$$p(\omega_L) = p^{\#\{0\text{'s in }\omega_L\}} \cdot (1 - p)^{\#\{1\text{'s in }\omega_L\}}. \qquad (1.6)$$

For such a Markov chain Gora studied the existence and uniqueness of the invariant measure denoted by μ^ε, and in particular its behavior as ε goes to zero. Under conditions more general than ours (in particular he allows for attractors with an infinite number of points) he showed that, as one would expect, any invariant measure μ^ε is asymptotically concentrated as $\varepsilon \to 0$ on the attractors of T in the sense that

$$\lim_{\varepsilon \to 0} \mu^\varepsilon(F) = 0 \qquad (1.7)$$

$\forall F$ disjoint from the attractors.

Under some extra hypotheses similar to our hp4 but slightly more restrictive he also proved uniqueness of the invariant measure. In the next paragraph we will state our main results concerning the "loss of memory" of the process X^ε and their consequence on the regularity of the invariant measure.

From now on we will assume that hp1 ... hp4 holds.

2. Main Results

Let C be the neighborhood of the union of the attractors and $\beta < 1$ be as in hp4.

Theorem 1.1. *For any*

$$m < m_0 = \frac{p \ln(1/\beta)}{\sup_x n(x)},$$

there exists $\varepsilon_0 > 0$ and $\gamma > 0$ such that for any $\varepsilon < \varepsilon_0$ and for any $L > n$ with $n = 0(1/\varepsilon)$ we have:

a) $P\left(T_{\omega_L}(C) \subset C \text{ and } \sup_{x \in C} L(T_{\omega_L}, x) \leq e^{-mL} \right) \geq (1 - e^{-\gamma/\varepsilon})$,

b) $P\left(\sup_{x \in C} L(T_{\omega_L}, x) \leq e^{-mL} \forall L \gg n \right) \geq 1 - e^{-\gamma/\varepsilon}$.

In b) $L \gg n$ means $L > \text{const } n$ for a sufficiently large constant independent of ε.

In the same notations of Theorem 1.1 we also have:

Corollary 1.1. *Let C_i be a connected component of C and let γ, n be as in Theorem 1.1. Then:*

$$P\left(\sup_{x, y \in C_i} d(T_{\omega_L}(x), T_{\omega_L}(y)) \leq e^{-mL} \forall L \gg n \right) \geq 1 - e^{-\gamma/\varepsilon}.$$

As a direct consequence of Theorem 1.1 we obtain the next result about the regularity of the invariant measures μ_ε of the chain X^ε. Let ε_0 and m be as in Theorem 1.1. Then:

Theorem 1.2. *For any $\varepsilon < \varepsilon_0$ there exists a unique invariant measure μ_ε of the Markov chain $\{X_n^\varepsilon\}$. Furthermore if $2e^{-m} < 1$ then μ_ε is singular continuous.*

Before giving the proof of the above results let us make some comments. The first theorem may appear at first sight rather simple since it just says that the maps T_{ω_L} for most of the trajectories ω_L behave like a power proportional to L of the map T, at least as far as the contracting properties around the attractors is concerned. In fact by (1.3) we have

$$L(T_{\omega_L}, x) \leq \prod_{j=1...L} L(T_{\omega(j)}, T_{\omega(j-1)}(x)). \tag{1.8}$$

Using the fact proved in Lemma 1.5, that the process X^ε spends a large fraction of the time L in the set C it is possible to show that

$$\sup_{x \in M} P(L(T_{\omega_L}, x) \leq e^{-mL}) \geq 1 - e^{-cL} \tag{1.9}$$

for some constant $C > 0$ and L large enough. However in Theorem 1.1 the supremum over $x \in C$ is taken inside the probability; in other words the starting point x is not fixed but it can vary in the set C depending on the given trajectory ω_L. As it is well known, taking the supremum from outside to inside the probability may affect considerably the probabilistic estimate.

In particular in the general case we cannot replace in Theorem 1.1. the $\sup_{x \in C}$ with $\sup_{x \in M}$. As an example consider a differential map T from S^1 onto S^1 satisfying our hypotheses, and let T_ε be the rotation by ε. Then T_{ω_L} will be onto for any ω_L, and thus $\sup_{x \in S^1} L(T_{\omega_L}, x) \geqq 1$, while estimate (1.9) remains true. This and related topics will be discussed more completely in the next section when we will analyze the case of stochastic differential equations.

The corollary expresses in a precise way the glueing of two trajectories with different initial conditions but equal noise. This also gives the loss of memory of the process X^ε at least for initial conditions x, y in the same connected component of the set C. This means for example that, in order to have the glueing, if both x and y are in the basin of attraction of the same periodic orbit, then they must also be "in phase." It should be emphasized that the fact that two trajectories with suitable initial conditions eventually glue together with large probability is much more than the simple fact that they spend most of their time close to the same attractor. In particular they will make rate jumps to visit other attractors but always staying closer and closer together. It would be rather nice to remove the restriction on the initial conditions x and y of being in the same C_i. This is certainly possible if one is able to show that two arbitrary points $x, y \in C$ visit at some future time t the *same* connected component C_i of C. This is of course not true in general; just consider the case of a map T on S_1 with stable fixed points at $\theta = \pi/2, 3\pi/2$, and unstable fixed points at $\theta = 0, \pi$, symmetric under reflection around the South–North and East–West axes. As usual T_ε will denote the rotation by ε. Thus by construction

$$d(T_{\omega_L}(\pi/2), T_{\omega_L}(3\pi/2)) = \pi$$

for any ω_L.

In the discrete case it would be possible to make hypotheses on T and T_ε to avoid situations like the above, but we preferred to discuss this problem in the continuous case where it can be solve in a more natural way.

3. Proofs of the Results

We start with the proofs of Corollary 1.1 and Theorem 1.2.

Proof of Corollary 1.1. Let $x, y \in C_i$ and let ω_L be fixed. Then by the definition of $L(T_{\omega_L}, x)$ we easily obtain

$$d(T_{\omega_L}(x), T_{\omega_L}(y)) \leqq \sup_{x' \in C_i} L(T_{\omega_L}, x') d(x, y).$$

The result follows from Theorem 1.1.

Proof of Theorem 1.2. To prove the uniqueness of the invariant measure μ_ε we observe, following Gora, that it is sufficient to show that

$$\lim_{L \to \infty} \frac{1}{L} \int d\mu^\varepsilon(x) E\{\log S(\omega_L, x)\} < 0, \tag{1.10}$$

where $E\{\ \}$ denotes the expectation value over the trajectories ω_L and

$$S(\omega_L, x) = \prod_{j=1\ldots L} L(T_{\omega_{L(j)}}, x_{j-1}) \tag{1.11}$$

with the convention that

$$x_j = T_{\omega(j)}(x). \tag{1.12}$$

Because of (1.3) one has:

$$L(T_{\omega_L}, x) \leqq S(\omega_L, x), \tag{1.13}$$

and, as we will see in the proof of Theorem 1.1, that

$$P\left(\sup_{x \in C} S(\omega_L, x) \leqq e^{-mL}\right) \geqq 1 - e^{-\gamma/\varepsilon} \tag{1.14}$$

$\forall L \geqq n$, $n = 0(1/\varepsilon)$. Therefore (1.10) follows from (1.14) and the result of Gora:

$$\lim_{\varepsilon \to 0} \mu_\varepsilon(C) = 1. \tag{1.15}$$

The proof of the second part of the theorem will follow very closely the proof of the analogous result given in [19] for the random composition of two rational maps on the real line. Let Ω_L be the following set of trajectories ω_L:

$$\Omega_L = \left\{\omega_L; \sup_{x \in C} L(T_{\omega L}, x) \leqq e^{-mL}\right\}$$

and let

$$C_{\Omega_L} = \bigcup_{\omega_L \in \Omega_L} T_{\omega_L}(C),$$

$$\Sigma = \bigcap_{k \geqq 1} \bigcup_{L \geqq k} C_{\Omega_L}.$$

By construction the Lebesgue measure of Σ, denoted by $|\Sigma|$, is estimated by

$$|\Sigma| \leqq \lim_{k \to \infty} \sum_{L \geqq k} |C_{\Omega_L}|. \tag{1.16}$$

By the same argument used in the proof of Corollary 1.1 we get

$$|C_{\Omega_L}| \leqq \text{const}\, |C| e^{-mL}{}_{\#}\{\omega_L \in \Omega_L\} \leqq \text{const}\, e^{-mL} 2^L. \tag{1.17}$$

Thus if $2e^{-m} < 1$ we obtain

$$|\Sigma| = 0.$$

In order to estimate the μ_ε-measure of Σ we use the equation expressing the invariance of the measure μ_ε,

$$\mu_\varepsilon(B) = p\mu_\varepsilon(T^{-1}(B)) + (1-p)\mu_\varepsilon((T^\varepsilon)^{-1}(B)), \tag{1.18}$$

valid for any measurable set B in M.

We have

$$\mu^\varepsilon(B) = \lim_{k \to \infty} \mu_\varepsilon\left(\bigcup_{L \geq k} C_{\Omega_L}\right) \geq \lim_{k \to \infty} \mu_\varepsilon(C_{\Omega_k})$$

$$\geq \lim_{k \to \infty} \sum_{\omega_L \in \Omega_k} p(\omega_k)\mu_\varepsilon(C) \geq \mu_\varepsilon(C)(1 - \exp(-\gamma)), \tag{1.19}$$

with γ as in Theorem 1.1. In the inequalities we have used the result of Theorem 1.1 together with Eq. (1.18) applied to the set C_{Ω_K}. Thus, again by the result of Gora, $\mu_\varepsilon(\Sigma) > 0$ for ε small enough, and thus μ_ε has at least a singular component. However, by taking

$$\Sigma = \bigcup_L \bigcup_{\omega_L} T_{\omega_L}(\Sigma),$$

one gets immediately, $\mu_\varepsilon(\Sigma) = 1 \ \forall \varepsilon$ small enough, while $|\Sigma| = 0$.

Proof of Theorem 1.1. The proof relies on a rather detailed study of the structure of the typical path of the Markov chain $\{X_n^\varepsilon\}$. Using deterministic arguments we will construct a set of paths ω_L such that the corresponding T_{ω_L} satisfies the conditions of the theorem, and afterwards we will show that these paths have a very large probability. The above strategy together with all the technical details is borrowed rather directly from the rigorous work on the Anderson localization done by Fröhlich, Martinelli, Scoppola and Spencer [17] (see also [18] for a complete and detailed review). Of course the techniques employed in the proof of the Anderson localization must be modified and adapted to the present case; however this requires little work and afterwards the two proofs will be almost identical. Therefore we decided to omit all those parts of the proof (mainly probabilistic) which can be found without problems in the reference quoted above.

Before going into the technical details let us explain in simple words the idea behind the proof. As shown by Gora typical paths of the chain spend a very long time in the neighborhood C of the attractors and make rare jumps among different attractors. When two trajectories (with the same noise) stay in the same connected component C_i of C they tend to glue together, because of hp4, while during the "jumps" they may diverge exponentially fast since $L(T,x)$ can be greater than 1 outside A. However if the "jump" is "fast," i.e. occurs on a time scale shorter than the typical time spent in C, then the divergence of the two trajectories outside C is largely compensated by the glueing inside C and, as a result, the two trajectories will glue together exponentially fast. Of course in a very long time L all kind of "jumps," fast or slow, occur and one has to classify them according to their time scale. This simple argument should justify the multiscale analysis of the typical paths that we will now discuss.

For $a > 0$, let $l(a, \varepsilon) = [\exp(a/\varepsilon)]$, and let for $\omega \in \{0, 1\}^N$ and $j \in N$,

$$\omega^j = \{\omega((j-1)l), \omega((j-1)l+1), \ldots, \omega(jl)\}.$$

The composition of maps T, T^ε associated to ω^j will be also denoted by T_j whenever this will not cause confusion.

We next introduce the set $S_0 \equiv S_0(\omega)$ of singular blocks:

$$S_0 = \{j \in N; {}_\#\{t; (j-1)l \leq t \leq jl, \omega^j(t) = 0\} < \alpha p l \tag{1.20}$$

or there exists a string of consecutive 1's in ω^j of length greater than $d/2\varepsilon$} with $\alpha < 1$.

It is easy to check (see Appendix A) that if $j \notin S_0$ then

$$T_j(A) \subset C$$

and

$$\sup_{x \in C} L(T_j, x) \leqq e^{-ml} \quad \text{with} \quad m < \alpha m_0, \quad m_0 = \frac{p \log 1/\beta}{\sup_x n(x)},$$

provided ε is small enough.

As a preliminary but important step we estimate the probability that a given site $j \in N$ belongs to S_0:

Lemma 1.1. *There exist constants a_0, ε_0 and γ such that*

$$P(j \in S_0) \leqq e^{-\gamma/\varepsilon} \quad \forall j \in N,$$

provided $a < a_0$ and $\varepsilon < \varepsilon_0$.

The easy proof of the lemma is given in Appendix A. From the lemma it follows that for a typical trajectory the set S_0 will consist mainly of small clusters very well isolated one from the other. It is also clear that, if Λ is an interval in N of length L such that $\Lambda \cap S_0 = \varnothing$, then

$$T_\Lambda = T_{j+L} \circ \cdots \circ T_{j+1} \circ T_j$$

satisfies:

   i) $T_\Lambda(C) \subset C$,
   ii) $\sup_{x \in C} L(T_\Lambda, x) \leqq e^{-mlL}$,   (1.21)

with m as in (1.20).

In order to study maps T_Λ with $\Lambda \cap S_0 \neq \varnothing$ we introduce a hierarchy of singular sets $S_{k+1} \subset S_k \subset \cdots \subset S_0$. Let $d_k = \exp(\lambda(5/4)^k)$, $\lambda > 1$. We set

$$S_{k+1} = S_k \backslash S_k{}^g,$$

where $S_k{}^g = U C_k{}^\alpha$ is the maximal union of clusters of sites in S_k (not necessarily connected) satisfying:

   i) diam $C_k{}^\alpha \leqq d_k$,
   ii) $\text{dist}(S_k \backslash C_k{}^\alpha, C_k{}^\alpha) \geqq 2d_{k+1}$,
   iii) There exists an interval $\bar{C}_k{}^\alpha$ such that:

$$(n_0 - 1)d_k \leqq \text{dist}(\partial \bar{C}_k{}^\alpha, C_k{}^\alpha) \leqq n_0 d_k, \quad n_0 \in N,$$
$$C_k{}^\alpha \subset \bar{C}_k{}^\alpha, \quad T_{\bar{C}_k}\alpha(C) \subset C$$
$$\bar{C}_k{}^\alpha \text{ is } (k-1) \text{ admissible where:} \qquad\qquad (1.22)$$

Definition. A set $\Lambda \subset N$ such that $\partial \Lambda \cap \bar{C}_j{}^\alpha = \varnothing \, \forall \alpha, \forall j \leqq k$ is said to be k-admissible.

In the course of the proof the two parameters λ and n_0, appearing in the definition of the times scales and of $C_k{}^\alpha$ respectively, will be chosen in a suitable way.

Proposition 1.1. *Let Λ be an interval in N of length L such that*

   i) $\Lambda \cap S_{k+1} = \varnothing$,
   ii) $L \geqq d_{k+1}/5$,
   iii) Λ *is k-admissible.*

Then

   a) $T_\Lambda(C) \subset C$,
   b) $\sup\limits_{x \in C} L(T_L, x) < \exp(-m_{k+1} Ll)$,

where $m_{k+1} = m_k - \text{const } d_k/d_{k+1}$ for a suitable constant independent of λ and k.

Proof. The proof is by induction. Let F_k denote the assertion of the proposition of step k. We already know that F_0 is true. We will show that $F_k \Rightarrow F_{k+1}$. This is trivially true if $\Lambda \cap S_k = \varnothing$. Thus let us assume that $\Lambda \cap S_k^g \neq \varnothing$. This is clearly the only possibility left since by hypothesis $\Lambda \cap S_{k+1} = \varnothing$. It is also clear that we can restrict ourselves to the case:

$$\Lambda \cap S_k^g = C_k^\alpha. \tag{1.23}$$

In fact if $\Lambda \cap S_k^g = C_k^{\alpha_1} \cup C_k^{\alpha_2} \cup \cdots \cup C_k^{\alpha_m}$, then we can always find m intervals $\Lambda_1, \ldots, \Lambda_m$ satisfying the hypotheses i), ii), iii) of F_{k+1}, and such that

$$\Lambda = \bigcup_{i=1}^m \Lambda_i, \quad \Lambda_i \cap S_k^g = C_k^{\alpha_i}.$$

If F_{k+1} holds for each Λ_i then it obviously extends to Λ because of (1.3). The existence of the sets Λ_i is guaranteed by (1.22) and by the following lemma whose proof can be found in [20]:

Lemma 1.2. *If λ is taken large enough then for any $k \geqq 0$ given two intervals $A \subset B$ in N such that $\text{dist}(\partial A, \partial B) \geqq 10 d_k$ there exists a k-admissible set A_0 such that*

$$A \subset A_0 \subset B.$$

Thus we now examine the case (1.23).

For notational convenience we call 1 the left boundary of Λ, L its right boundary and by a (b) the left (right) boundary of $\bar{C}_k^a$. Then the set $A \equiv [1, a-1]$, $B \equiv [b+1, L]$ are $(k-1)$-admissible, and thus by construction are such that

$$T_A(C) \subset C, \tag{1.24}$$

and the same for B. In fact, since A and B are $(k-1)$-admissible, and $A \cap S_k = \phi$, $B \cap S_k = \phi$, they contain only intervals not in S_0 separated by intervals $C_i^\alpha, i < k$, and thus, by (1.22), they map C into C.

Furthermore, by hypothesis, $T_{\bar{C}_k^a}(C) \subset C$, and thus

$$T_A(C) \subset C. \tag{1.25}$$

The quantity $L(T_\Lambda, x)$ is estimated as follows:

$$\sup_{x \in C} L(T_\Lambda, x) \leqq \sup_{x \in C} L(T_A, x) \sup_{x \in C} L(T_{\bar{C}_k} \alpha, x) \sup_{x \in C} L(T_B, x). \tag{1.26}$$

If diam $A < d_k/5$, then diam $B > L\text{-}(n_0 + (1/5))d_k) \geqq L(1 - 5(n_0 + (1/5))(d_k/d_{k+1})) > (d_k/5)$ if λ is large enough independent of k. We used here that $L > d_{k+1}/5$. Thus, by applying F_k to B we obtain

$$\sup_{x \in C} L(T_A, x) \leqq \exp\{ -m_k Ll + m_k 5(n_0 + 1/5)Lld_k/d_{k+1} + Vl((d_k/5) + n_0 d_k)\}$$

$$\leqq e^{-m_{k+1}Ll}, \tag{1.27}$$

where $V = \log \sup\left(\sup_x L(T, x), \sup_x L(T^\varepsilon, x) \right)$ and $m_{k+1} \equiv m_k - (m_k 5(n_0 + 1/5) + 5V(n_0 + 1/5))d_k/d_{k+1}$. The induction step is proved.

Remark. For any given $0 < \alpha < 1$ we can find λ so large that $m_k > \alpha m \; \forall k$.

The next step in our analysis is to compute the probability that a given set Λ intersects S_k.

Proposition 1.2. *There exists $\varepsilon_0 > 0$ and for any $\varepsilon < \varepsilon_0$ there exists $\eta(\varepsilon)$, with $\eta(\varepsilon) \uparrow + \infty$ as $\varepsilon \downarrow 0$ such that*

$$P(\Lambda \cap S_k \neq \varnothing) \leqq |\Lambda| P(1 \in S_0)^{1/2}/d_k^{\eta},$$

where

$$|\Lambda| = \#\{j \in \Lambda\}.$$

Before discussing the proof of the above result let us complete the proof of Theorem 1.1.

a) Let us fix $m < m_0$ and let us choose λ so large that $m_\infty > (m + m_0)/2$. We write $L = lL_0 + L_1, L_0, L_1 \in N, L_1 < l = [\exp(a/\varepsilon)]$.

We distinguish between two cases:

   i) $L_1 \leqq d/\varepsilon$,
   ii) $L_1 > d/\varepsilon$.

In the first case by assumption $L_0 \geqq 1$. Let $\Lambda \subset N$ be the set:

$$\Lambda = \{1, 2, \ldots, L_0\}.$$

We define $k(L)$ as the smallest integer such that

$$L_0 < d_{k(L)+1}, \tag{1.28}$$

and we set

$$\Omega_L = \{\omega_L; \Lambda \cap S_{k(L)}(\omega_L) = \varnothing \text{ and } \Lambda \text{ is } k(L)\text{-admissible}\}.$$

From Proposition 1.2 we obtain:

$$P(\Omega_L) \geqq 1 - P(1 \in S_0)^{1/2}[L_0]^{-\eta(\varepsilon)(4/5)+1} - P(\Lambda \text{ is not } k\text{-admissible for some } k). \tag{1.29}$$

The third term in the right-hand side of (1.29) is clearly bounded by

$$2P\left(1 \in \bigcup_{j=0}^{\infty} \bigcup_{\alpha} \bar{C}_j^{\alpha}\right) \leqq 2 \sum_{k=0}^{\infty} P([1, n_0 d_k] \cap S_k \neq \emptyset)$$

$$\leqq 2P(1 \in S_0)^{1/2} \sum_{k=0}^{\infty} n_0 d_k / d_k^{\eta(\varepsilon)} \leqq \text{const } P(1 \in S_0)^{1/2} \quad (1.30)$$

if $\eta(\varepsilon) > 1$. In conclusion, with probability greater than

$$1 - P(1 \in S_0)^{1/2} = 1 - e^{-\gamma/2\varepsilon},$$

the set Λ satisfies all the hypothes of $F_{k(L)}$ and thus:

  i) $T_\Lambda(C) \subset C$,
  ii) $\sup_{x \in C} L(T_\Lambda, x) \leqq e^{-m_\infty L_0 l}$.

Therefore we have

$$\sup_{x \in C} L(^T\omega_L, x) \leqq e^{-m_\infty L_0 l + V L_1} < e^{-mL}$$

for ε sufficiently small.

In order to show that $^T\omega_L^{(C) \subset C}$ we observe that, as shown in the proof of Lemma 1.1. in the Appendix, $\text{dist}(T_\Lambda(C), C_0) < d/2$ and therefore if between lL_0 and L there exists no strings of consecutive 1's of length greater than $d/2\varepsilon$, then $^T\omega_L^{(C) \subset C}$. This last event occurs with probability greater than $1 - \exp(-\gamma/\varepsilon)$.

In the second case, that is $L_1 > d/\varepsilon$, we proceed as above but we impose also that between $L_0 l$ and L there are no strings of consecutive 1's of length greater than $d/2\varepsilon$ and that $\#\{t; L_0 l \leqq t \leqq L, \omega(t) = 0\} \geqq \alpha p l$. This last event occurs with probability greater than $1 - \exp(-\gamma/\varepsilon)$. As observed in Remark 1 of Appendix A the above condition ensures that the block $[lL_0, L]$ has exactly the same properties i), ii) as the blocks not in S_0, thus proving part a) of the theorem also in this case.
b) Using Proposition 1.1 we get:

$$P\left(\sup_{x \in C} L(^T\omega_L, x) \leqq e^{-mL} \forall L \gg n\right)$$

$$> P(1 \text{ is } k\text{-admissible } \forall k \geqq 0 \text{ and } \Lambda \cap S_{k(L)-1} = \emptyset \ \forall L > Nl).$$

$$P\left(\sup_{x \in C} L(^T\omega_L, x) \leqq e^{-mL} \forall L, Nl \geqq L \gg n\right), \quad (1.31)$$

where Λ is as in a) with $L_0 = [L/l]$ and N is so large that $k(L) \geqq 1$.

In fact, using Lemma 1.2, if 1 is k-admissible for any $k \geqq 0$ and $\Lambda \cap S_{k(L)-1} = \emptyset$, then we can always find a site $j \in \Lambda$ with

$$L_0 - 10 d_{k(L)-1} \leqq j \leqq L_0$$

such that

$$\Lambda' \equiv [1, 2, \ldots, j] \text{ is } (k(L) - 1)\text{-admissible and } \Lambda' \cap S_{k(L)-1} = \emptyset.$$

Thus, by Proposition 1.1,

$$\sup_{x \in C} L(T_{\omega_L}, x) \leqq \sup_{x \in C} L(T_{A'}, x) \exp \{ V d_{k(L)-1} | \}$$

$$\leqq \exp \{ - m_\infty L + V(1 + d_{k(L)-1})l \} \leqq \exp \{ - m_\infty L + \text{const } l L_0^{4/5} \}$$

$$\leqq e^{-mL} \quad \text{if } L_0 \geqq N \text{ is large enough.}$$

The first factor appearing in the right-hand side of (1.31) is estimated as (1.30) and it is larger than $1 - e^{-\gamma/\varepsilon}$ if ε is sufficiently small. The second factor of (1.31) is estimated as in the proof of Lemma 1.1. in the Appendix and it is greater than $1 - e^{-\gamma'/\varepsilon}$ for some $\gamma' > 0$.

This concludes the proof of the theorem.

It remains to discuss the probabilistic estimate given in Proposition 1.2. As the reader can easily check, the proof of this estimate coincides with the proof of an analogous result, stated with exactly the same notations, given in [20] (see Sect. 5), provided one could estimate the probability that a given interval A satisfying the first two geometrical conditions (1.22) on scale k violates the third one. As we will see, the above event implies that, under the random evolution associated with A, some point among the attractors spends an unusually long time outside a neighborhood of the attractors. The probability of this last event will be then estimated in Appendix A by a rather standard construction in the theory of Markov processes.

For the reader already familiar with the works on Anderson localization [17, 18] the above problem is the exact counterpart in our context of the estimate of the probability of having an energy resonance in the random hamiltonian. We now define precisely the event which will be analyzed.

Let A be an interval $[a, b]$ in N of length $L \leqq d_k$, and let Ω_A be the event:

$$\Omega_A \equiv \{ \omega; A \cap S_0 \neq \varnothing, A \text{ satisfies } (1.22), \text{i)}, \text{ii) on scale } k \text{ but not iii).} \} \quad (1.32)$$

Lemma 1.3.

$$P(\Omega_A) \leqq d_k P(1 \in S_0)^{1/4} \exp \{ - q d_k \}$$

for some constant $q > 0$ independent of k, ε provided n_0 is large enough and ε is sufficiently small independent of k.

To prove the lemma we show that the event Ω_A implies another event $\Omega'_A, \Omega'_A \supset \Omega_A$, whose probability is easier to estimate. In order to describe the new event Ω'_A we first fix some notations. Let for $x \in K_a$ (the union of all the attractors) $\tau(x, t, T)$ be the time spent between times t, T by the process X^ε starting at x at time t outside the set C, let for any $0 < \rho < l$,

$$\Omega'_{A,x,t} = \{ \omega; A \cap S_0 \neq \varnothing; \tau(x, lt, (b + d_k)l) \geqq \rho d_k l \},$$

where b is the rightmost site of A and let

$$\Omega'_A = \bigcup_{x \in K_a} \quad \bigcup_{t \in \mathbf{N}, a - n_0 d_k \leqq t \leqq a - (n_0 - 1)d_k} \Omega'_{A,x,t}. \quad (1.33)$$

Then there exist $\lambda_0, \tilde{n}_0, \rho_0$ such that if $\lambda > \lambda_0, n_0 > \tilde{n}_0, \rho < \rho_0$, then

$$\Omega_A \subset \Omega'_A \quad \forall L. \quad (1.34)$$

To prove (1.34) let us fix $\omega \in \Omega_\Lambda$. Using (1.22) i), ii), this implies that for any $(k-1)$-admissible set $\Lambda \subset N$, $\Lambda \supset \Lambda$, $\Lambda = (a', b')$ with $a - n_0 d_k \leqq a' \leqq a - (n_0 - 1)d_k$, $b + (n_0 - 1)d_k \leqq b' \leqq b + n_0 d_k$, there exists a point $x \in C$ such that

$$T_\Lambda(x) \notin C.$$

Actually we have much more. In fact for a fixed a' as above we can find $a - 10d_{k-1} < a'' < a$ such that $[a', a'']$ is $(k-1)$-admissible and $[a', a''] \cap S_k = \varnothing$. Thus if λ is so large that $(n_0 - 1)d_k - 10d_{k-1} > (1/5)d_k$, then we can apply Proposition 1.1 to the set $[a', a'']$ to get that

$$\text{dist}(T_{[a',a'']}(C), T_{[a',a'']}(K_a)) \leqq \exp\{-ml((n_0 - 1)d_k - 10d_{k-1})\}. \qquad (1.35)$$

The above distance can only increase by a factor $e^{V(d_k + 10d_{k-1})}$ under the action of $T_{[a'',b]}$, and therefore if n_0 is large enough we get that

$$\text{dist}(T_{[a',b]}(C), T_{[a',b]}(K_a)) \leqq \exp\{-\text{const } n_0 ld_k\}. \qquad (1.36)$$

Thus, if for a $(k-1)$-admissible $b', b \leqq b' \leqq b + d_k$, $\text{dist}(T_{[a',b]}(K_a), C_0) \leqq d/2$, then $T_{[a',b']}(C) \subset C$ if ε is small enough and n_0 is large enough.

This inequality shows the role played by the thick left layer (a', a) which we assumed has empty intersection with S_k. The layer allows us to replace the dynamics of an uncountable set of points (the set C) with the dynamics of only a finite number of points (the set K_a). It is now easy to show that, if $T_{[a',b']}(C) \subset C$ for some b' as above, then the same holds for any $(k-1)$-admissible $b'', b \leqq b'' \leqq b + n_0 d_k$. Thus we have shown that $\omega \in \Omega_\Lambda$ implies that for some $(k-1)$-admissible a', with $a - n_0 d_k \leqq a' \leqq a - (n_0 - 1)d_k$, any $(k-1)$-admissible b with $b \leqq b' \leqq b + d_k$ there exists $x \in K_a$ such that $\text{dist}(T_{[a',b']}(x), C_0) > d/2$. In other words:

$$\sum_{x \in K_a} \tau(x, a'l, (b + d_k)l) \geqq \alpha pl \#\{b'; b' \text{ is } (k-1)\text{-admissible}, b \leqq b' \leqq b + d_k\}, \qquad (1.37)$$

where α is as in (1.20) and $p = \text{Prob}(\omega(1) = 0)$. The reason for the factor αpl in the right-hand side of (1.37) is the following: let b' be $(k-1)$-admissible; since $b' \notin S_k$ in particular $b' \notin S_0$, and therefore if for some $t \in N$, $(b'-1)l \leqq t \leqq b'l$ the process X^ε happens to be in C and $X^\varepsilon_{t+1} = T(X^\varepsilon_t)$, then $X^\varepsilon_{t+1} \in C_0$, and therefore $\text{dist}(X^\varepsilon_{b'l}, C_0) \leqq d/2$ if ε is small enough. Thus, for the process X^ε to have a distance from C_0 greater than $d/2$ at time $b'l$ implies that $X^\varepsilon_t \notin C$ for all times t such that

$$(b'-1)l \leqq t \leqq b'l \quad \text{and} \quad T_{\omega(t)} = T.$$

From (1.20) it follows that the cardinality of such set of times is at least αpl.

We next estimate the number appearing in the right-hand side of (1.37):

Lemma 1.4. *For any $L \in N$, $L \geqq d_k$,*

$$\#\{j \in N; j \leqq L; j \text{ is } (k-1)\text{-admissible}\} \geqq (1 - \text{const } d_0/d_1)L$$

for a suitable constant independent of λ.

The proof is given in Appendix A.

Using the lemma and (1.37) we get that $\omega \in \Omega_\Lambda$ implies that there exists $x \in K_a$ such that

$$\tau(x, a'l, (b + d_k)l) \geqq \alpha pl(1 - \text{const } d_0/d_1)d_k/N \qquad (1.38)$$

for some $d - n_0 d_k \leq a' \leq a - (n_0 - 1)d_k$, where $N = \#\{x \in K_a\}$. Thus (1.34) is proved if $\rho_0 = \alpha p(1 - \text{const } d_0/d_1)/N$. It remains to estimate $P(\Omega'_A)$. We have:

$$P(\Omega'_A) \leq \sum_{a - n_0 d_k \leq a' \leq a - (n_0 - 1)d_k} \sum_{x \in K_a} \mathbf{P}(\tau(x, la', (b + d_k)l) > \rho l d_k)$$

$$\leq d_k N \sup_{x \in K_a} \mathbf{P}_x(\tau(x, 0, (n_0 + 2)d_k l) > d_k l \rho). \tag{1.39}$$

To complete the estimate we need the following last result provided in Appendix A:

Lemma 1.5. *For any* $0 < \rho < 1$ *and* $a' > 0$, *there exist* $K(\rho)$ *and* ε_0 *such that for any* $\varepsilon < \varepsilon_0$:

$$\sup_{x \in K_a} P_x(\tau(x, 0, T) > \rho T) \leq \exp\{- K(\rho) T \exp(- a'/\varepsilon)\}.$$

Thus using the Lemma and (1.39) we get

$$P(\Omega_A) \leq d_k l N \exp\{- K(\rho) d_k l \exp(- a'/\varepsilon)\}. \tag{1.40}$$

On the other hand,

$$P(\Omega_A) \leq |A| P(1 \in S_0), \tag{1.41}$$

and thus by taking the geometric mean between (1.40), (1.41) we get

$$P(\Omega_A) \leq d_k l [P(1 \in S_0)]^{1/2} [N \exp\{- K(\rho) d_k l \exp(- a'/\varepsilon)\}]^{1/2}.$$

Using Lemma 1.1, if we take $a' < a < \gamma/4$, we get the result with a q which actually diverges as $\varepsilon \to 0$.

Appendix A

Proof of Lemma 1.1. By means of standard large deviations estimates for i.i.d. r.v. we have for $\alpha > 1$,

$$P(\#\{t < l; \omega^1(t) = 0\} < \alpha pl) < \exp(- kl)$$

if ε is sufficiently small. Furthermore

$$P\{\text{there exists a string of consecutive 1's in } \omega^j \text{ of length greater than } d/2\varepsilon\}$$
$$< l(1 - p)^{d/2\varepsilon}.$$

Thus if $a < - \ln(1 - p)d/2$ we have the result.

We next show that if $j \notin S_0$ then:

$$T_j(C) \subset C \tag{A.1}$$

and

$$\sup_{x \in C} L(T_j, x) \leq e^{- ml} \tag{A.2}$$

for any $m < \alpha m_0$ if ε is small enough. By stationarity we can consider only $j = 1$. Let $x \in C$ and let $1 \leq t \leq l$ be the first time such $\omega_l(t) = 0$. Since $\text{dist}(T^\varepsilon(x), \varepsilon x) \leq \varepsilon$ and $t < d/(2\varepsilon)$, by assumption, we get that $x^\varepsilon_{t-1} \in A$ and therefore $x^\varepsilon_t \in C_0$ since $T(A) \subset C_0$. By iterating this argument we get that $\text{dist}(x^\varepsilon_l, C_0) < d/2$ i.e. $x^\varepsilon_l \in C$. (A.1) is proved.

To prove (A.2) we observe that because of (A.1) for any $x \in C$ the integer $n(x)$ of hp4 stays constant under the random evolution, i.e.

$$n(x_t^\varepsilon) = n(x) \quad \forall 1 \leqq t \leqq l.$$

This is because the point x is not able to leave the basin of attraction of the attractor $K(x)$. For notational convenience we set $n(x) = \underline{n}$. Next we devide the string ω_l into N substrings ω_l^i of length L_i with the property that the first $N - 1$ ones contain exactly $\underline{n}$ 0's, where $N = [\#\{0\text{'s in } \omega_l\}/\underline{n}]$. We also denote by N_- the number of substrings of length smaller than $1/\sqrt{\varepsilon}$. Using the hp4 we have that if $L_i < 1/\sqrt{\varepsilon}$,

$$\sup_{x \in C} L(T_{\omega_l} i, x) \leqq \beta + O(\sqrt{\varepsilon}).$$

If $l_i > 1/\sqrt{\varepsilon}$ and using the fact that in any case $L_i < d\underline{n}/2\varepsilon$, then

$$\sup_{x \in C} L(T_{\omega_l} i, x) \leqq \text{const.}$$

Therefore

$$\sup_{x \in C} L(T_{\omega_l}, x) \leqq \exp\{N_- \ln(\beta + O(\sqrt{\varepsilon})) + (N - N_- + 1)\text{const}\}.$$

Since $N - N_- < 1\sqrt{\varepsilon}$ and $N > (\alpha pl/\underline{n}) - 1$, we obtain (A.2).

Remark 1. It is important to observe that in the above argument we never used in any essential way the fact that $l = \exp(a/\varepsilon)$. As the reader can easily check the only requirement on l is to be larger than c/ε for a suitable constant $c \gg d$.

Proof of Lemma 1.4. Let Λ be an interval in N of length $L \geq d_k$. We have to count how many sites in Λ do not intersect $\bar{C}_j^\alpha$ for some α and some $j \leq K - 1$. Using (1.22) we get

$$\#\{x \in \Lambda; x \in C_j^\alpha \text{ for some } \alpha \text{ and } j \leq K - 1\} \leqq \sum_{j=0}^{k-1} (2n_0 + 1)d_j \frac{L}{d_{j+1}} \leqq \text{const } L\frac{d_0}{d_1}.$$

The lemma is proved.

Proof of Lemma 1.5. The proof uses a construction which is quite common in this kind of problems (see e.g. [6] and [8]). Let $B \equiv C^c$ and set for $x \in C$:

$$\tau_1 = \inf\{t \geqq 0; X_t^\varepsilon(x) \in B\},$$
$$\sigma_1 = \inf\{t \geqq \tau_1; X_t^\varepsilon(x) \in C_0\},$$
$$\tau_i = \inf\{t \geqq \sigma_{i-1}; X_t^\varepsilon(x) \in B\}. \tag{A.6}$$

Clearly we have

$$\tau(x, 0, T) \leqq \sum_{i=1}^{v} \sigma_i - \tau_i, \tag{A.7}$$

where $v = \inf\{n; \sigma_n \geqq T\}$.

Therefore

$$P_x(\tau(x, 0, T) \geq \rho T) \leq e^{-\beta \rho T} \mathbf{E}_x \exp\left\{\beta \sum_{i=1}^{v} (\sigma_i - \tau_i)\right\}$$

$$\leq e^{-\beta \rho T} \sum_{n=1}^{T} \mathbf{P}_x(v = n)^{1/2}\left(\mathbf{E}_x \exp\left\{2\beta \sum_{i=1}^{n} (\sigma_i - \tau_i)\right\}^{1/2}\right) \quad (A8)$$

for any $\beta > 0$. Using the strong Markov property the right-hand side of (A.8) can be estimated by

$$e^{-\beta \rho T} \sum_{n=1}^{T} \mathbf{P}_x(v = n)^{1/2}\left(\sup_{x \in B} \mathbf{E}_x \exp\{2\beta(\sigma_1 - \tau_1)\}\right)^{n/2}. \quad (A.9)$$

In order to estimate this last expression we need the following results which will be proved later on.

Lemma A.1. *For any n such that* $n > 1 + 4T(1 - p)^{d/2\varepsilon}/(|\ln(1 - p)|[d/\varepsilon])$,

$$\sup_{x \in A} P_x(v = n) \leq \exp\{-n[d/\varepsilon]|\ln(1 - p)|/8\}.$$

Lemma A.2. *For any* $a > 0$ *there exist* $\varepsilon_0 > 0, \beta_0 > 0$ *such that for any* $\varepsilon < \varepsilon_0$ *and any* $\beta < \beta_0 \exp\{-a/\varepsilon\}$,

$$\sup_{x \in B} \mathbf{E}_x e^{\beta \sigma_1} \leq 2.$$

Using the Lemmas we get that the expression (A.9) is bounded by

$$\exp\{-\beta \rho T/3\}, \quad (A.10)$$

provided β is as in Lemma A.2 and ε is small enough. This concludes the proof of Lemma 1.5.

It remains to prove Lemma A.1, A.2. The first one follows easily from the trivial estimate:

$$P_x(v = n) \leq P_x (\exists n\text{-strings of consecutive 1's in } \omega_T \text{ of length greater than } [d/\varepsilon]), \quad (A.11)$$

and from standard estimates on i.i.d. random variables. To prove the second lemma we estimate:

$$\sup_{x \in B} P_x\{\sigma_1 > kt_0\}, \quad (A.12)$$

where $t_0 = \exp\{a/\varepsilon\}$ with $a > 0$ and $k \in N$. We will show that for any choice of $a > 0$ we can find ε so small that the above probability is bounded by $(1/2)^k$. In fact given $a > 0$ let $\delta' \ll \delta$ be so small that hp3 applies to any $K_{\delta'}$ in the set B and let $D = D(\delta, \delta')$ be the set $B \backslash U K_{\delta'}$. Let also $t_D(x)$ be the classical exist time from the region D starting at x and let n_0 be the number of basic sets in B. Given $x \in B$ we will now construct a string $\omega(x)$ of length of order δ'/ε such that $T_{\omega(x)}(x)$ reaches C_0. This will provide an estimate from below of the probability of $\{\sigma_1 < c\delta'/\varepsilon\}$ by means of the probability of this string $\omega(x)$.
We have to discuss two cases:

i) $x \in D$,
ii) $x \in UK_{\delta'} \cap B$.

In the first case the string $\omega(x)$ will consist of $n_0 + 1$ substrings $\omega_i(x)$ of consecutive 0's of length l_i, separated one from the other by strings of exactly n consecutive 1's where $n = n(2\delta')$ is as in hp3. The lengths l_i are given by the following rule: let x_k be the position of the process X^e at time $t_k = n(k-1) + \sum_{i \le k} l_i$. Then we set recursively

$$l_1 = t_D(x),$$
$$l_2 = t_D((T^e)^n(x_1)) \quad \text{if } x_1 \in UK_{\delta'} \cap B \quad \text{and} \quad l_2 = 0 \quad \text{if } x_1 = \in C_0,$$

.

.

$$l_i = t_D((T^e)^n(x_{i-1})) \quad \text{if } x_{i-1} \in UK_{\delta'} \cap B \quad \text{and} \quad l_i = 0 \quad \text{if } x_{i-1} \in C_0.$$

Since T is a continuous map $\sup_{x \in B} t_D(x) < t_D(\delta, \delta')$ and therefore the length of $\omega(x)$ does not exceed $n_0 t_D(\delta, \delta') + a_0 2\delta'/\varepsilon < c\delta'/\varepsilon$, if ε is small enough. It remains to prove that the point x under the action of $T_{\omega(x)}$ reaches the set C_0 before time $c\delta'/\varepsilon$. If this is not the case then, using hp3, we could find at least two basic sets K^i and K^j in B and a path $\underline{x}$ of length $N < c\delta'/\varepsilon$ going from $K_\delta{}^i$ to $K_\delta{}^j$ and from $K_\delta{}^j$ to $K_\delta{}^j$ were equivalent contrary to our hypotheses. In the second case $x \in UK_{\delta'} \cap B$ the argument is exactly the same with the only difference that we set $l_1 = 0$.

With this construction we have

$$P_x\{\sigma_1 < c\delta'/\varepsilon\} \ge P\{\omega(x)\} \ge \exp\{-c'\delta'/\varepsilon\} \tag{A.13}$$

for a suitable constant c'. Using the Markov property we get for any integer j,

$$P_x\{\sigma_1 > jc\delta'/\varepsilon\} < (1 - \exp\{-c'\delta'/\varepsilon\})^j. \tag{A.14}$$

Thus if we choose $j = k[\exp(a/\varepsilon)/(c\delta'/\varepsilon)]$ with $k \in N$ we get

$$\sup_{x \in B} P_x\{\sigma_1 > kt_0\} < (1 - \exp\{-c'\delta'/\varepsilon\})^{[\exp(a/\varepsilon)/(c\delta'/\varepsilon)]k} < (1/2)^k \tag{A.15}$$

if $c'\delta' < a$ and ε is small enough. Clearly (A.15) proves the lemma.

Section 2. The Continuous Case

1. The Model

In this section we extend the results of the previous section to the case of an ordinary Ito stochastic differential equation in $\mathbf{R}^n$ with small constant diffusion:

$$dX_t^\varepsilon = b(X_t^\varepsilon)dt + \varepsilon w_t. \tag{2.1}$$

Here w_t denotes the ordinary brownian motion in $\mathbf{R}^n$, and the drift b is assumed to be smooth and to satisfy a sequence of assumptions that we state below. For notational convenience we will denote by $X_{t,t_0}^\varepsilon(x)$ the solution at time t of the

equation:

$$X^\varepsilon_{t,t_0}(x) = x + \int_{t_0}^t ds\, b(X^\varepsilon_{s,t_0}(x)) + \varepsilon(w_t - w_{t_0}).$$

As usual $X^\varepsilon_{t,0}(x)$ will be denoted by $X^\varepsilon_t(x)$.

The first assumption on $\underline{b}$ makes sure that the process X^ε_t solution of (2.1) admits a unique smooth invariant measure μ^ε (see e.g. [22]):

hp1. *There exist $R_0 > 0$ and $a > 0$ such that if $n(x, R)$ denotes the outward normal to the surface of the sphere B_R, centered at 0 of radius R, at the point x, then*

$$\sup_{R > R_0} \sup_{x \in \partial B_R} b(x) n(x, R) < -a < 0.$$

Thus the drift b is confining and the process X^ε_t will spend most of its time in the ball B_{R_0}. We also assume that $\sup_{x \in \mathbf{R}^n} |\nabla b(x)| < K$ for some $K < 0$.

The next assumption concerns the long time behaviour of the dynamical system:

$$\frac{dX_t}{dt} = b(X_t). \tag{2.2}$$

Following Ventzel and Freidlin [6] let for any continuous function $\phi:[0, T] \to \mathbf{R}^n$:

$$I_{0,T}(\phi) = \int_0^T dt\, \| \dot\phi_t - b(\phi_t) \|^2 \tag{2.3}$$

if the integral exists and $I_{0,T}(\phi) = \infty$ if not. Here $\| \quad \|$ denotes the euclidean norm in $\mathbf{R}^n$. For $x, y \in \mathbf{R}^n$ we also let

$$V(x, y) = \inf_{\phi; \phi(0) = x, \phi(T) = y} I_{0,T}(\phi). \tag{2.4}$$

Like in the discrete case we establish an equivalence relation in $\mathbf{R}^n$,

$$x \approx y \leftrightarrow x = y \quad \text{or} \quad V(x, y) = V(y, x) = 0. \tag{2.5}$$

An equivalence class K is said to be "stable" if

$$V(x, y) > 0 \quad \forall x \in K, \quad \forall y \notin K,$$

and "unstable" if it is not stable.

hp2. *There exist finitely many compacta $K_1, \ldots, K_N$ such that*

   i) *for any two points x, y in $K_i x \approx y$,*
   ii) *if $x \in K_i$ and $y \notin K_i$ then $x \not\approx y$,*
   iii) *every $\omega - $ limit set of the system (2.2) is contained in some K_i,*
   iv) *only the first $l > N$ compacta are stable and they consist of exactly one point (fixed points).*

It is easy to see that $V(x, y)$ attains the same value V_{ij} for all $x \in K_i, y \in K_j$. In terms of the numbers V_{ij} the stability of a compactum K_i can be also stated as follows (see [6]):

$$K_i \text{ is stable iff } V_{ij} > 0 \quad \forall j \neq i. \tag{2.6}$$

Let now F_t be a family of smooth map from $\mathbf{R}^n$ to $\mathbf{R}^n$ and let

$$L(F_t, x) = \lim_{\varepsilon \to 0} \sup_{\|x-y\| < \varepsilon} \frac{\|F_t(x) - F_t(y)\|}{\|x - y\|}.$$

hp3. *There exist a number m_0 and, for each stable compactum $K_i, i = 1, \dots, l$, a neighborhood A_i such that*

$$\sup_{x \in A_i} L(X_t, x) < e^{-m_0 t} \quad \forall t > 0,$$

where $X_t(x)$ is the solution of (2.2) starting at x.

This last assumption is clearly satisfied if the Jacobian matrices $\partial b_i / dx_j$ at the stable fixed points have eigenvalues with negative real part smaller than $-m_0$. As we will discuss later, hp1, 2, 3 are not the more general ones. Remarks and generalizations can be found at the end of the next paragraph.

Using only hp1 and a weaker form of hp2 Ventzel and Freidlin [6] analyzed in great detail the long time behaviour of the process X_t^ε as $\varepsilon \to 0$. Essentially they showed that the process X_t^ε can be approximate, in some appropriate sense, by a Markov chain with state space a small neighborhood of the union of the stable compacta $K_i, i = 1, \dots, l$. They proved that the transition probabilities P_{ij} from K_i to K_j are of order

$$P_{ij} \asymp \exp(-V_{ij}/\varepsilon^2), \tag{2.8}$$

where $\asymp$ means: $\lim_{\varepsilon \to 0} -2\varepsilon^2 \ln(p_{ij}) = V_{ij}$.

Using the Markov chain they were also able to estimate the asymptotic behaviour of the invariant measure μ^ε as $\varepsilon \to 0$ and to prove that μ^ε is exponentially (in $1/\varepsilon^2$) concentrated on a small neighborhood of a subset M of the stable compacta $K_i, i = 1, \dots, l$. The subset M was determined uniquely by an algebraic procedure involving only the numbers V_{ij}, (see (2.9)). In the next paragraphs we state and prove our main results, very similar to the ones proved for the discrete case.

2. Main Results

We assume hp's 1, 2, 3. Let x_i be a stable fixed point and let

$$d_i = \sup \left\{ \delta > 0; \ \sup_{x \in B_\delta(x_i)} L(X_t, x) \leqq e^{-m_0 t} \ \forall t > 0 \right\}.$$

Using hp3 the quantity $d = \min_{1 \leqq i \leqq l} d_i$ is strictly greater than zero and we will denote with A_i the set $B_d(x_i)$ and we define

$$A = \bigcup_{i=1}^{l} A_i \quad C_i = B_{d/2}(x_i) \quad C = \bigcup_{i=1}^{l} C_i$$

$$C_{0i} = B_{d/4}(x_i) \quad C_0 = \bigcup_{i=1}^{l} C_{0i}.$$

With the above notations our results read as follows:

Theorem 2.1. *For any $m < m_0$ there exist positive constants $K, K', t_0, \varepsilon_0$, such that for any $\varepsilon < \varepsilon_0$ and any $t > t_0$:*

a)

$$P\left(X_t^\varepsilon(x) \in C \quad \forall x \in C \text{ and } \sup_{x \in C} L(X_t^\varepsilon, x) \leq e^{-mt}\right) \geq 1 - e^{-K/\varepsilon^2},$$

b)

$$P\left(\sup_{x \in C} L(X_t^\varepsilon, x) \leq e^{-mt} \, \forall t > t_0\right) \geq 1 - e^{-K'/\varepsilon^2},$$

where $X_t^\varepsilon(x)$ is the solution of 2.1 starting at x.

Let $t_i(x) = \inf\{t \geq 0; X_t(x) \in A_i\}$ and let $B_{i,T} = \{x; t_i(x) < T\}$, then

Corollary 2.1. *For any $m < m_0$ and T there exist positive constants $T_0, K_1, K_2, \varepsilon_0$, such that for any $\varepsilon < \varepsilon_0$,*

$$P\left(\sup_{x,y \in B_{i,T}} \frac{\|X_t^\varepsilon(x) - X_t^\varepsilon(y)\|}{\|x - y\|} \leq K_2 e^{-mt} \, \forall t > T_0\right) \geq 1 - e^{-K_1/\varepsilon^2}.$$

The above theorem and corollary are exactly the counterpart of Theorem 1.1, Corollary 1.1 in the discrete case.

The natural question now is the following: Is it possible to extend the result of Corollary 2.1 to any couple of arbitrary points x, y in $\bigcup_i B_{i,T}$, removing the restriction $x, y \in B_{i,T}$? In the continuous case, contrary to the discrete one, we can answer this question. In fact the main difference in the theory of small random perturbations of dynamical systems between the continuous and the discrete case is contained in the following (see also Gora [8]) probabilistic estimate valid in the continuous case:

$$P_x\left(\sup_{0 \leq t \leq T} \|X_t^\varepsilon - \phi_t\| < \delta\right) \geq \exp\left[\frac{-I_{0T}(\phi) - h}{2\varepsilon^2}\right]$$

for any continuous function ϕ with $\phi(0) = x$ and such that $\int_0^T (1 + \|\dot\phi_s\|) < \infty$. On the contrary in the discrete case it is very easy to construct simple examples in which the above estimate is false.

In order to give the precise result, we have to distinguish between the general n-dimensional case $n > 1$, and the one dimensional case. In order to discuss the general case we need to recall the basic result of [6] about the asymptotics as $\varepsilon \to 0$ of the invariant measure μ^ε of the chain X^ε. Let L be the set whose elements, denoted by letters i, j, m, n etc., are the subfix of the stable fixed points x_i, x_j, x_m etc. By assumption the cardinality of L is l. By $\{i\} - graph \ g, i \in L$, we mean a graph consisting of arrows $m \to n$ ($m \in L \backslash \{i\}$, $n \in L$, $n \neq m$) such that

1. exactly one arrow comes from any $m \in L \backslash \{i\}$,
2. there are no closed cycles in the graph.

The set of all $\{i\} - graphs$ is denoted by $G\{i\}$. Ventzel and Freidlin proved the

following: Let

$$V\{i\} = \min_{g \in G\{i\}} \sum_{(m \to n) \in g} V_{mn}$$

and let $M \subset L$ be the set of all i for which

$$V\{i\} = \min(V\{1\} \dots V\{l\}).$$

Then if F is a closed set that does not intersect $\bigcup_{i \in M} K_i$ then

$$\lim_{\varepsilon \to 0} \mu^\varepsilon(F) = 0$$

Using the above notations we have:

Proposition 2.1. *Let M consist of a single element i_0. Then for any $m < m_0$ there exist positive constants ε_0, V, c, k such that $\forall \varepsilon < \varepsilon_0$, if $t(\varepsilon) = e^{V/\varepsilon^2}$, then*

$$P\left(\forall t > t(\varepsilon), \sup_{x,y \in C} \frac{\| X_t^\varepsilon(x) - X_t^\varepsilon(y) \|}{\| x - y \|} \leq e^{-m(t - t(\varepsilon))} \right) \geq 1 - e^{-K/\varepsilon^2}.$$

In the one dimensional case we do not need the restriction $M = \{i_0\}$ and we can take the supremum over x and y in **R**:

Proposition 2.2. *Let $n = 1$ and let $m < m_0$. Then there exist positive constants k, V, ε_0, such that $\forall \varepsilon < \varepsilon_0$. If $t(\varepsilon) = e^{V/\varepsilon^2}$, then*

$$P\left(\forall t > t(\varepsilon), \sup_{x,y \in B_{R_0}} \frac{\| X_t^\varepsilon(x) - X_t^\varepsilon(y) \|}{\| x - y \|} \leq e^{-m(t - t(\varepsilon))} \right) \geq 1 - e^{-k/\varepsilon^2}.$$

Before giving the proofs of the above results let us make some comments about the two propositions. As already explained in Sect. 1 the intuitive picture of the stochastic flow for ε small is the following: in a finite time t_0, which depends only on the dynamical system (2.2), all the points in the set B_{i,R_0} fall in the set A_i and there they stay, glueing together, for a time of order (see [6])

$$T_i \sim e^{\min_{i \neq j} V_{ij}}.$$

After time T_i they make a common jump to some other stable set A_j. If it happens that the typical time T_j spent by the process X_t^ε in the set A_j is much longer than T_i, then, with large probability, the paths of X_t^ε starting in A_i and those starting in A_j will meet in A_j if $i \to j$ was the first transition from i. Therefore they will have a long time, of order $T_j - T_i$, to glue together. The condition in Proposition 2.1 assures exactly that among the set $A_i, i = 1, \dots, l$ there is only one, A_{i_0}, with the largest time T_{i_0}. Therefore all the other A_i, will have time to make a transition $i \to i_0$ (it may be very well a multistep transition) before the trajectories starting in A_{i_0} are able to leave A_{i_0}, thus producing the glueing of the paths.

In the one dimensional case it is not necessary to assume that the set M consists of a single point. In fact let $x < y$ and let K_{i_1} be the rightmost stable fixed point. Then when $X_t^\varepsilon(x)$ reaches the set A_{i_1} by the uniqueness of the solution of (2.1), the path $X_t^\varepsilon(y)$ (solution of (2.1) starting at y and subjected to the same brownian perturbation as $X_t^\varepsilon(x)$) will stay to the right of $X_t^\varepsilon(x)$, and therefore they will both

belong to the basin of attraction of A_{i_1}. At this point they will start to glue together according to Corollary 2.1.

So far, our results, Corollary 2.1 and Proposition 2.1, show that in dimension greater than one two arbitrary initial conditions x and y will eventually glue together provided they start around the same attractor or around different attractors, provided there exists a unique attractor which is the most "stable" one. Thus nothing is said about what happens to an initial condition starting e.g. close to an unstable fixed point. The reason is that in general we do not know that with large probability two arbitrary initial conditions will be eventually in the basin of attraction of the same attractor so that Theorem 2.1 could be applied. For example if $b(x) = \nabla U(x)$ with $U(x)$ a periodic function then the above requirement is certainly not satisfied if x and y are two different absolute minima of U. However for "confining" drifts b like these ones considered in this section one expects that in many cases the requirement is fulfilled.

In order to understand the problem, consider the case of two equivalent attractors x_{i_0}, x_{i_1}: to prove the glueing of paths with starting points in different basins of attraction, we have to exclude that when a large fluctuation producing tunneling leads the points in A_{i_0} to the neighborhood A_{i_1} of x_{i_1}, at the same time the same fluctuation leads the points of A_{i_1} to the neighborhood A_{i_0}.

It is easy to see that this can be the case when the basin of attraction of one stable point surrounds the other basin of attraction.

In order to prove the global glueing of paths when M consists of more than one point it is sufficient to make the assumption that there exists a stable point whose basin of attraction is so "fat" that it contains the translation of the sphere B_R with R such that any K_i belongs to B_R. More precisely:

Proposition 2.3. *If for* $R_0 > \max_i \mathrm{dist}(K_i, 0)$ *there is a stable point* x_i, *and a vector* $\vec{\xi}$ *such that*

$$B_{2R_0} + \vec{\xi} \subset B_i,$$

where $B_{2R_0} + \vec{\xi}$ *means the translation of the ball* B_{2R_0} *by the vector* $\vec{\xi}$, *then* $\forall R > 0$ *and* $\forall m < m_0$ *there exist positive constants* k, ε_0, V, *such that for any* $\varepsilon < \varepsilon_0$ *if* $T(\varepsilon) = \exp V / \varepsilon^2$, *then*

$$P\left(\forall t > T(\varepsilon), \sup_{x,y \in B_R} \frac{\| X_t^\varepsilon(x) - X_t^\varepsilon(y) \|}{\| x - y \|} \leqq e^{-m(t - T(\varepsilon))} \right) \geqq 1 - e^{-K/\varepsilon^2}.$$

The hypotheses of Proposition 2.3 are actually quite general, and thus the global glueing holds in a large class of cases.

A typical example of the problem in which the glueing cannot be proved by Proposition 2.3 is the case of two spiral basins of attraction.

It should however be observed that the time $T(\varepsilon)$ appearing in the proposition may be a very bad estimate of the actual time scale of the glueing which sensitively depends on the detailed structure of $b(x)$. This is the case for example if we apply Proposition 2.3 to the one dimensional case instead of Proposition 2.2.

We conclude this section with a critical discussion on the hypotheses behind our results and with some suggestions for further generalizations.

Remark 1. About the drift b. A close look at the proof of Theorem 2.1 shows that hp1,2 are not the most general ones. We can also treat the case of a drift term $b(x)$ with an infinite number of stable and unstable fixed points provided they do not accumulate and hp2 (iii), hp3 hold. This is the case for example of $b(x) = \nabla U(x)$, where U is a periodic function on $\mathbf{R}$ with e.g. quadratic absolute minima at the sites of the lattice $\mathbf{Z}^n$ or if $b(x)$ is the hierarchical drift introduced in [21].

In the periodic case, however, it is clear that the result of Propositions 2.1, 2.2, 2.3 cannot hold. Also the requirement of the absence of attracting limit cycles can be relaxed provided there is still one stable fixed point x_i around which the process X_t^ε spends a finite (as $\varepsilon \to 0$) fraction of its time.

A much more crucial assumption was hp3 on the structure of the drift $b(x)$ around the stable fixed points. We now examine what happens when this hypothesis is replaced by a weaker one.

Let us suppose that the origin is a stable equilibrium point for the dynamical system (2.2) and let us denote by $B(x)$ the matrix:

$$B_{ij}(x) = \partial b_i / \partial x_j.$$

We assume that there exist constants $c > 0, \delta > 0, \beta > 0, \infty > n > 0$ such that for any $|x| < 2\delta$:

$$\langle \vec{z}, B(x)\vec{z} \rangle \leq -c|z|^2 |x|^\beta. \tag{2.9}$$

A simple example of a drift $b(x)$ satisfying the above condition is the following:

$$b(x) = -\nabla U(x)$$

with

$$U(x) = \langle \vec{x}, V\vec{x} \rangle^2,$$

where V is a $n \times n$ positive definite matrix. In this case $\beta = 2$. It is clear that (2.9) violates hp3. Nevertheless we will show that for the stochastic differential equation (2.1) we can find an ε-dependent "mass" $m(\varepsilon)$ such that

$$\sup_{|x| < \delta} L(X_t^\varepsilon, x) \leq e^{-m(\varepsilon)t}$$

with large probability.

Proposition 2.4. *There exist* $t_0, \varepsilon_0, k > 0, a > 0$ *and for any* $\varepsilon < \varepsilon_0$ *there exists* $m(\varepsilon) \geq \varepsilon^{3\beta}$ *such that*

$$P\left(\forall t, t_0 \leq t \leq e^{a/\varepsilon^2}, \sup_{|x| < \delta} L(X_t^\varepsilon, x) \leq e^{-m(\varepsilon)t} \right) \geq 1 - \exp\left(-\frac{k}{\varepsilon^2} \right).$$

The reader may wonder at this point whether the fact of having a very small "mass" $m(\varepsilon)$ does not cause a problem to the rest of our analysis and in particular to the inductive procedure discussed in detail in Sect. 1. As the reader can check the total loss of "mass" at the end of the inductive procedure was of order (see (1.27)) $1/d_0^{1/4}$. Thus in order for the whole scheme to be meaningful it is necessary

that $1/d_0^{1/4} \ll m(\varepsilon)$, i.e. $d_0 \gg \varepsilon^{-8\beta}$. However it is also clear that we have to take d_0 much smaller than the typical distance between blocks in S_0 since otherwise most of the blocks in S_0 will fail to satisfy the geometric conditions (1.22) and the argument will not work again.

Fortunately in our case there is no contradiction between the two requirements on d_0 because of the exponential bound in $1/\varepsilon^2$ given in the proposition. In fact a choice of the initial scale d_0 of order $\varepsilon^{-(12\beta+1)}$ will work perfectly.

Remark 2. About the brownian motion εw_t. An important assumption we made about the random perturbation εw_t was to assume its covariance matrix to be constant. Although this case already covers many interesting problems arising in theoretical physics and other areas of natural sciences it is interesting to see what happens when one considers the more general case. We do this for a one dimensional equation:

$$dX_t^\varepsilon = b(X_t^\varepsilon)dt + \varepsilon\sigma(X_t^\varepsilon)dw_t \quad X_0^\varepsilon = x, \tag{2.10}$$

where b is as before and $\sigma \in C^2(\mathbf{R})$ with $1/\gamma > \sigma(x) > \gamma$, $|(d/dx)\sigma| < C_1$, $|(d^2/dx^2)\sigma| < C_2$, for suitable constants $\gamma > 0, C_1 > 0, C_2 > 0$.

We just prove Lemma 2.1 (see below) for the above case, since, once this is done, the rest of the discussion is the same.

Let us assume for simplicity that the origin is a stable fixed point $b(0) = 0$ with $b'(0) = -m_0$ and let $\delta > 0$ be so small that $\sup\limits_{|x| < \delta} |b'(x) + m_0| \leq \eta$, where $\eta > 0, \eta \ll m_0$.

Proposition 2.5. *For any* $T > 0$, $m < m_0$, *there exist* $\varepsilon_0 > 0$, $K > 0$ *such that*

$$P\left(\sup_{|x| < \delta/2} \left|\frac{d}{dx} X_t^\varepsilon(x)\right| < \frac{1}{\gamma} e^{-mt} \,\forall t < T\right) \geq 1 - e^{-K/\varepsilon^2}$$

$\forall \varepsilon < \varepsilon_0$.

3. Proofs of the Results

We first prove Corollary 2.1 and Propositions 2.1, 2.2, 2.3, 2.4, 2.5 and then Theorem 2.1.

Proof of Corollary 2.1. Let $T_0' = \inf\{t \geq 0; X_t(x) \in C_{0,i} \,\forall x \in B_{i,T}\}$. By assumption $T_0' < \infty$. We now want to estimate the probability that $X_{T_0'}^\varepsilon(x) \in C_i \,\forall x \in B_{i,T}$. To do this we compare the path of X_t^ε with the solution of (2.2) with the same initial condition:

$$X_t^\varepsilon(x) - X_t(x) = \varepsilon w_t + \int_0^t [b(X_s^\varepsilon(x)) - b(X_s(x))]ds. \tag{2.11}$$

If we set

$$\phi_t \equiv |X_t^\varepsilon(x) - X_t(x)| \quad t \leq T_0',$$

then

$$\phi_t \leq \sup_{t \leq T_0'} |\varepsilon w_t| + \lambda \int_0^t \phi_s ds \tag{2.12}$$

for a suitable constant λ; thus by the Gronwall inequality

$$\phi_{T_0'} \leqq \sup_{t \leqq T_0'} |\varepsilon w_t| e^{\lambda T_0'}. \tag{2.13}$$

Using (2.13) we get

$$P(X_{T_0'}^\varepsilon(x) \in C_i \; \forall x \in B_{i,T}) \geqq P\left(\sup_{t \leqq T_0'} |\varepsilon w_t| \leqq e^{-\lambda T_0'} d/4\right) \geqq 1 - \exp\left(-\frac{K'}{\varepsilon^2}\right) \tag{2.14}$$

for ε sufficiently small (see e.g. [22]).

Let finally $t \geqq T_0' + t_0 \equiv T_0$, where t_0 is the constant appearing in Theorem 2.1. Using Theorem 2.1 and (2.14), the random flow

$$x \to X_t^\varepsilon(x) = (X_{t,T_0'}^\varepsilon \circ X_{T_0'}^\varepsilon)(x)$$

is such that

$$\sup_{x \in A} L(X_{t,T_0'}^\varepsilon, x) \leqq e^{-m(t-T_0')} \leqq e^{-m(t-T_0)} \quad \forall t > T_0 \tag{2.15}$$

and $X_{T_0'}^\varepsilon(x) \in C_i \; \forall x \in B_{i,T}$ with probability bigger than $1 - \exp(-K''/\varepsilon^2)$.

By elementary estimates (2.15) shows that

$$\sup_{x,y \in B_{i,T}} \frac{\|X_t^\varepsilon(x) - X_t^\varepsilon(y)\|}{\|x-y\|} \leqq e^{-m(t-T_0)} e^{\lambda T_0'} = e^{-mt} K_2$$

with probability larger than $1 - \exp(-K_1/\varepsilon^2)$ for a suitable constant $K_1 > 0$.

Proof of Proposition 2.1. The proof is based on the following two probabilistic estimates which will be proved in Appendix B and in the course of the proof of Theorem 2.1 respectively.

Let

$$S_1(T,x,y) = \{t \leqq T; \exists i; X_t^\varepsilon(x) \in A_i, X_t^\varepsilon(y) \in A_i\}$$

and

$$S_2(T,\alpha,m) = \left\{t \leqq T; \sup_{x \in A} L(X_{\alpha,T,t}^\varepsilon(x)) \leqq e^{-m(\alpha T - t)}\right\}$$

with $\alpha > 1$.

Then for any $m \leqq m_0$ there exist positive constants V_0, V_1, K_1, K_2 such that for any $V_0 \leqq \bar V \leqq V_1$ and any ε small enough,

$$P\left(\inf_{x,y \in C} |S_1(T,x,y)| \geqq 2T/3, \forall T \geqq \bar t(\varepsilon)\right) \geqq 1 - \exp\left(-\frac{k_1}{\varepsilon^2}\right), \tag{2.16}$$

$$P(S_2(T,\alpha,m) \geqq 2T/3, \forall T \geqq \bar t(\varepsilon)) \geqq 1 - \exp\left(-\frac{k_2}{\varepsilon^2}\right), \tag{2.17}$$

where $\bar t(\varepsilon) = e^{\bar V/\varepsilon^2}$. Let us assume (2.16) and (2.17); it follows that

$$P\left(\sup_{x,y \in C} \frac{\|X_t^\varepsilon(x) - X_t^\varepsilon(y)\|}{\|x-y\|} \leqq e^{-mt} \; \forall t \geqq t(\varepsilon)\right)$$

$$\geq P\left(\forall x, y \in C, \forall t \geq t(\varepsilon) \, S_1\left(\frac{t}{\alpha}, x, y\right) \cap S_2\left(\frac{t}{\alpha}, \alpha, \frac{m\alpha + \lambda}{\alpha - 1}\right) \neq \emptyset\right), \quad (2.18)$$

where $\lambda = \sup_x L(X_t^\varepsilon, x)$, and provided α is so large that

$$\frac{m\alpha + \lambda}{\alpha - 1} < m_0.$$

In fact let $t \geq t(\varepsilon)$ be given; if

$$S_1\left(\frac{t}{\alpha}, x, y\right) \cap S_2\left(\frac{t}{\alpha}, \alpha, \frac{m\alpha + \lambda}{\alpha - 1}\right) \neq \emptyset,$$

then there exists $t_0 = t_0(x, y) \leq t/\alpha$ such that

$$X_{t_0}^\varepsilon(x) \in A_i, \quad X_{t_0}^\varepsilon(y) \in A_i$$

for some i and

$$\sup_{x \in A} L(X_{t,T}^\varepsilon(x)) \leq e^{-(m\alpha + \lambda/\alpha - 1)(t - t_0)}.$$

Thus

$$\frac{\|X_t^\varepsilon(x) - X_t^\varepsilon(y)\|}{\|x - y\|} \leq e^{\lambda t_0 - (m\alpha + \lambda/\alpha - 1)(t - t_0)} \leq e^{-mt}$$

if α is large enough and ε is sufficiently small.

In the last inequality we used the Gronwall inequality to bound $\|X_{t_0}^\varepsilon(x) - X_{t_0}^\varepsilon(y)\|$.

Using the estimates (2.16) (2.17), the probability appearing in (2.18) is bounded from below by

$$1 - \exp\left(-\frac{c}{\varepsilon^2}\right)$$

for a suitable constant C provided $t(\varepsilon)/\alpha \geq \bar{t}(\varepsilon)$.

This last condition is clearly true if we choose $V \leq \bar{V}$ and ε is small enough.

Proof of Proposition 2.2. We first observe that if $-R_0 < x < y < R_0$ then, as a consequence of the uniqueness of the solution of the stochastic differential equation (2.1),

$$X_t^\varepsilon(-R_0) \leq X_t^\varepsilon(x) \leq X_t^\varepsilon(y) \leq X_t^\varepsilon(R_0), \quad (2.19)$$

and therefore it is sufficient to study the probability of the event:

$$\Omega_0 = \{|X_t^\varepsilon(R_0) - X_t^\varepsilon(-R_0)| \leq 2R_0 e^{-m(t - t(\varepsilon))} \forall t \geq t(\varepsilon)\}. \quad (2.20)$$

Let i_0 be the index of the rightmost fixed point such that $i_0 \in M$ and let

$$\tau = \inf\{t \geq 0; X_t^\varepsilon(R_0) \in C_{i_0}, X_t^\varepsilon(-R_0) \in C_{i_0}\}. \quad (2.21)$$

If Ω_t denotes the event

$$\Omega_t = \left\{ \sup_{x \in C_{i_0}} L(X^\varepsilon_{t',t}, x) \leq e^{-m(t'-t)} \ \forall t' > t + t_0 \right\}, \tag{2.22}$$

t_0 as in Theorem 2.1, then using the strong Markov property we can clearly estimate the probability of Ω_0 from below by

$$\int_0^{t(\varepsilon)-t_0} dP(\tau \leq t) P(\Omega_t), \tag{2.23}$$

provided ε is small enough. Using Theorem 2.1 the above expression is bounded from below by:

$$\left(1 - \exp\left(-\frac{k}{\varepsilon^2} \right) \right) (P(\tau \leq t(\varepsilon) - t_0)). \tag{2.24}$$

In order to study $P(\tau \leq t(\varepsilon))$ we need to introduce new stopping times $\tau_{i_0}, \sigma_{i_0}, \xi_{i_0}$ as follows:

$$\tau_{i_0} = \inf \{ t > 0; X^\varepsilon_t(-R_0) \in C_{i_0} \},$$

$$\sigma_{i_0} = \inf \{ t > \tau_{i_0}; X^\varepsilon_t(R_0) \in C_{i_0} \} - \tau_{i_0},$$

$$\xi_{i_0}(x) = \inf \{ t > 0; X^\varepsilon_t(x) \in C_i \text{ for some } i \neq i_0 \}. \tag{2.25}$$

Using the Ventzel and Freidlin theory and the definition of i_0 we have that

$$\mathbf{E}(\tau_{i_0}) \leq e^{V_1/\varepsilon^2},$$

$$\mathbf{E}(\sigma_{i_0}) \leq e^{V_2/\varepsilon^2} \tag{2.26}$$

$$\inf x \in C_{i_0} P(\xi_{i_0}(x) \geq e^{V_3/\varepsilon^2}) \geq 1 - \exp\left(-\frac{k'}{\varepsilon^2} \right)$$

for any ε small enough, some positive V_1, V_2, V_3, k' with $V_2 < V_3$.

Furthermore

$$\inf_{x \in C_{i_0}} P\left(\int_0^{e^{V_3/\varepsilon^2}} dt \chi_{C_{i_0}}(X^\varepsilon_t(x)) \geq \rho e^{V_3/\varepsilon^2} \right) \geq 1 - \exp\left(-\frac{k''}{\varepsilon^2} \right) \tag{2.27}$$

for any $0 < \rho < 1$ and some constant $k'' = k''(\rho)$. Here $\chi_{C_{i_0}}(x)$ is the indicator function of the set C_{i_0}.

Let now $t(\varepsilon) = e^{V/\varepsilon^2}$ with $V > \max(V_1, V_2, V_3)$ and let $\bar{\Omega}$ be the intersection of the events:

$$\{ \tau_{i_0} < t(\varepsilon)/2; \sigma_{i_0} \leq e^{(V_2 + V_3)/2\varepsilon^2} \},$$

$$\{ \xi_{i_0}(X^\varepsilon_{\tau_{i_0}}(-R_0)) > e^{V_3/\varepsilon^2}; \xi_{i_0}(X^\varepsilon_{\tau_{i_0} + \sigma_{i_0}}(R_0)) > e^{(V_2 + V_3)2\varepsilon^2} \},$$

$$\left\{ \int_{\tau_{i_0}}^{\tau_{i_0} + e^{V_3/\varepsilon^2}} dt \chi_{C_{i_0}}(X^\varepsilon_t(-R_0)) \geq (2/3) e^{V_3/\varepsilon^2} \right\},$$

$$\left\{ \int_{\tau_{i_0} + \sigma_{i_0}}^{\tau_{i_0} + \sigma_{i_0} + e^{V_3/\varepsilon^2}} dt \chi_{C_{i_0}}(X^\varepsilon_t(R_0)) \geq (2/3) e^{V_3/\varepsilon^2} \right\}. \tag{2.28}$$

Using (2.27), (2.28) together with the strong Markov property we get that

$$P(\bar{\Omega}) > 1 - \exp\left(-\frac{k}{\varepsilon^2}\right) \qquad (2.29)$$

for some $k > 0$.

Furthermore it is easy to see that the event $\bar{\Omega}$ implies $\tau < t(\varepsilon)$ if ε is small enough. In fact by construction each path $X_t^\varepsilon(-R_0), X_t^\varepsilon(R_0)$ in the interval $(\tau_{i_0} + \sigma_{i_0}, \tau_{i_0} + e^{V_3/\varepsilon^2})$ spends in the set C_{i_0} a time bigger than:

$$(2/3)e^{V_3/\varepsilon^2} - 2\sigma_{i_0} > (1/2)(e^{V_3/\varepsilon^2} - \sigma_{i_0}) \qquad (2.30)$$

for ε small enough. Thus $\tau < \tau_{i_0} + e^{V_3/\varepsilon^2} < t(\varepsilon)$ again for ε small enough.

The proof is finished.

Proof of Proposition 2.3. Let B_{2R_0} and ξ be as in Proposition 2.3 and let $\phi_s^{t,t'}$ be the function with

$$\frac{d}{ds}\phi_s^{t,t'} = \frac{|\xi|}{(t'-t)} \quad \text{and} \quad \phi_t^{t,t'} = 0, \quad \phi_{t'}^{t,t'} = \xi.$$

For notational convenience we set $\delta t = t' - t$. We will also denote by I the value of the action integral $I_{t,t'}$ over ϕ. Let finally $T = T(R_0)$ be the smallest integer time such that after time T any point in the sphere $B_{2R_0} + \xi$ has reached under the classical flow a δ-neighborhood of the stable fixed point x_{i_0} and let $T(R, R_0)$, $R > R_0$, be the smallest integer time such that any point in the ball B_R has reached the ball $B_{R_0/2}$.

If we set by $T_g = \exp(I/\varepsilon^2)$, then we will show that

$$P(\exists t \in N; t < T_g; |X_t^\varepsilon(x) - x_{i_0}| < \delta \ \forall x \in B_{R_0}) \geq 1 - \exp\left(-\frac{k}{\varepsilon^2}\right) \qquad (2.31)$$

for a suitable constant $k > 0$.

Clearly the above estimate together with Corollary 2.1 is sufficient to prove the proposition. In order to prove (2.31) let us construct the following set of paths:

$$\Omega_n = \left\{ \{w_s\}_{0 \leq s \leq T_g}; \sup_{n - T(R,R_0) \leq s \leq n} |w(s) - w(n - T(R, R_0))| < \frac{\delta'}{\varepsilon}; \right.$$

$$\sup_{n + \delta t \leq s \leq n + \delta t + T} |w(s) - w(n)| \leq \frac{\delta'}{\varepsilon};$$

$$\left. \sup_{n \leq s \leq n + \delta t} |X_{n,s}^\varepsilon(0) - \phi(s)| \leq \delta' \right\}. \qquad (2.32)$$

Then we have that the probability in (2.31) can be estimated from below for any $R > 0$ by

$$P\left(\left\{\bigcup_{T(R,R_0) \leq n \leq T_g - 2T} \Omega_n\right\} \cap \left\{\sup_{x \in B_{R_0}} \sup_{t \leq T_g} |X_t^\varepsilon(x)| \leq R\right\}\right), \qquad (2.33)$$

provided δt is small enough and $\delta' \ll \delta$.

Let in fact $\{w_s\}_{0 \le s \le T_g}$ be in $\Omega_n \cap \left\{ \sup_{t \le T_g} |X_t^\varepsilon(x)| \le R \right\}$. Then at time $t = n$ $X_t^\varepsilon(x) \in B_{R_0}$ for any x in B_{R_0} since at time $n - T(R, R_0)$, $X_t^\varepsilon(x) \in B_R$ and between times $n - T(R, R_0)$ and n the process has followed the classical path within a distance $\exp(KT(R, R_0))\delta' < R_0/2$ if δ' is small enough. Here K is as usual the Lipschitz constant for the drift $b(x)$. At time $t' = t + \delta t$, $X_t^\varepsilon(x)$ is in $B_{2R_0} + \xi$ for any x in B_{R_0} since $\mathrm{dist}(X_{t',t}^\varepsilon(x), X_{t',t}^\varepsilon(0)) \le \exp(K\delta t)|x| \le 2R_0$ if δt is small enough. A similar argument shows that at time $t'' = n + T(r, R_0)$ the process is in a δ-neighborhood of x_{i_0}.

Using now the first of the Ventzel and Freidlin estimates we get that for a fixed n the probability of Ω_n is greater than $\exp(-(I + h)/2\varepsilon^2)$ with h small if ε is small. Therefore using the definition of T_g we get that the right-hand side of (2.31) is bounded from below by:

$$1 - \exp\left(-\frac{k}{\varepsilon^2} \right) - P\left(\left\{ \sup_{x \in B_{R_0}} \sup_{t \le T_g} |X_t^\varepsilon(x)| \ge R \right\} \right) \tag{2.34}$$

for some positive constant k. We are left with the estimate of the last term in (2.34). Let $x \in B_{R_0}$ be such that for some time $t < T_g$ the process starting at x is not in B_R. Let t_1 and t_2 be respectively the last time for which the process is still in B_{R_0} and the first time for which the process is not in B_R. Then using hp1 on the drift $b(x)$ we have that:

$$\varepsilon |(w(t_2) - w(t_1)| \ge a(t_2 - t_1) + R - R_0.$$

Using standard estimates on the brownian motion it is very easy to see that if R is chosen large enough (depending only on I and therefore on δt) the probability of this last event is of order $\exp(-k'/\varepsilon^2)$ for some positive k'. The proof is completed.

Proof of Proposition 2.4. Let $(\underline{Z}(t, \underline{x}))_i = \partial X_t^\varepsilon(t, x)/\partial x_{i_0}$, where $1 \le i_0 \le n$ will be kept fixed in the sequel. Then from Eq. (2.1) we obtain:

$$\frac{d}{dt} \underline{Z}(t, x) = B(X_t^\varepsilon(x))\underline{Z},$$

$$\underline{Z}(0, \underline{x}) = (\delta_{i,i_0}). \tag{2.35}$$

If we assume that $|X_s^\varepsilon(x)| < 2\delta \ \forall s \le t, \ \forall |x| < \delta$, then (2.35) implies, together with (2.9) that

$$\frac{1}{2} \frac{d}{dt} \langle \underline{Z}(t, \underline{x}), \underline{Z}(t, \underline{x}) \rangle \le -C |X_t^\varepsilon(x)|^\beta < \underline{Z}(t, \underline{x}), \underline{Z}(t, \underline{x}) >,$$

i.e.

$$|\underline{Z}(t, \underline{x})|^2 \le \exp\left[-2C \int_0^t ds |X_s^\varepsilon(x)|^\beta \right]. \tag{2.36}$$

Thus, in order to prove the proposition with $m(\varepsilon) = \varepsilon^{3\beta}$ it is enough to show that for a suitable choice of t_0 and a,

$$P\left(\exists t; t_0 \le t \le \exp\left(-\frac{a}{\varepsilon^2} \right); \sup_{|x| < \delta\tau_\varepsilon} (t, x) > t/2 \right) \le e^{-K/\varepsilon} \tag{2.37}$$

for some $K > 0$, where

$$\tau_\varepsilon(t, x) = \int_0^t ds\, \chi_B(X_s^\varepsilon(x))$$

with $B = \{x; |x| \leq \varepsilon^3 (1/C)^{(1/\beta)}\}$, since

$$P\left(\exists x; |x| < \delta, \exists s < \exp\left(-\frac{a}{\varepsilon^2}\right); |X_s^\varepsilon(x)| > 2\delta\right) \leq \exp\left(-\frac{\bar{K}}{\varepsilon^2}\right) \qquad (2.38)$$

(see (2.7), (2.14) if a is small enough).

The probability in (2.37) is estimated by

$$\sum_{j=t_0-1}^{\exp(-a/\varepsilon^2)+1} P\left(\sup_{|x| < \delta\tau_\varepsilon} (j, x) > j/3\right) \qquad (2.39)$$

if $t_0 > 6$. Since

$$|X_t^\varepsilon(x) - X_t^\varepsilon(y)| \leq e^{Kt}|x - y|,$$

because of the Gronwall inequality, the generic term in the sum (2.39) can be estimated from above by

$$\sum_{|x_i| < \delta, x_i \in Z^n(\varepsilon)} P(\tau_\varepsilon'(j, x_i) > j/3), \qquad (2.40)$$

where $Z^n(\varepsilon) = \varepsilon^3 e^{-kj} Z^n$, and $\tau_\varepsilon'(t, x_i)$ is defined as τ but with the ball B replaced by the ball of radius $r_\varepsilon = \varepsilon^3 (1/c)^{1/\beta} + (\sqrt{n}/2)\varepsilon^3$. So let us estimate one term of (2.40),

$$\sup_x P(\tau_\varepsilon'(t, x) > t/2) \leq e^{-\alpha t/2\varepsilon^2} \sup_x E \exp\left[\frac{\alpha \tau_\varepsilon'(t, x)}{\varepsilon^2}\right] \quad \forall \alpha > 0. \qquad (2.41)$$

Using the Markov property we get

$$\sup_x E \exp\left[\frac{\alpha \tau_\varepsilon'(t, x)}{\varepsilon^2}\right] \leq \left(\sup_x E \exp\left[\frac{\alpha \tau_\varepsilon'(\varepsilon^2, x)}{\varepsilon^2}\right]\right)^{1+t/\varepsilon^2}. \qquad (2.42)$$

Thus, since $\tau_\varepsilon'(\varepsilon^2, x) \leq \varepsilon^2$, the right-hand side of (2.42) can be estimate for α small by

$$\exp\left[(1 + t/\varepsilon^2)\frac{\alpha}{\varepsilon^2}\left(\sup_x E\tau_\varepsilon'(\varepsilon^2, x)\right) + (t/\varepsilon^2)O(\alpha^2)\right]. \qquad (2.43)$$

Now

$$\frac{1}{\varepsilon^2} E\tau_\varepsilon'(\varepsilon^2, x) = \frac{1}{\varepsilon^2}\int_0^{\varepsilon^2} ds \int_{|j| \leq r_\varepsilon} P^\varepsilon(x, y, s)\, dy, \qquad (2.44)$$

where $P^\varepsilon(x, y, s) = P_x(X_s^\varepsilon = y)$.

On the probability density $P^\varepsilon(x, y, s)$ we have the bound (see Ventzel–Freidlin [6], p. 391):

$$P^\varepsilon(x, y, s) \leq \frac{\text{const}}{(\sqrt{2\pi\varepsilon^2 s})^n},$$

which gives as one would expect, that the right-hand side of (2.44) is $O(\varepsilon^{2n})$. In conclusion, if ε is small enough and $\alpha \ll 1$ independent of ε, the right-hand side of

(2.41) is estimated from above by

$$e^{-\alpha j/3\varepsilon^2},$$

uniformly in x.

The proof is complete by plugging the above bound in (2.40) and in (2.39).

Proof of Proposition 2.5. Let

$$y_t^\varepsilon = \int_0^{X_t^\varepsilon} dx \frac{1}{\sigma(x)}. \tag{2.45}$$

Then, by the Ito's formula, we have

$$dy_t^\varepsilon = \left(\frac{b(X_t^\varepsilon)}{\sigma(X_t^\varepsilon)} - \sigma'(X_t^\varepsilon)\varepsilon^2 \right) dt + \varepsilon dw_t,$$

which implies for $z_t \equiv dy_t^\varepsilon/dx$:

$$dz_t = \left(b'(X_t^\varepsilon)z_t - \frac{b(X_t^\varepsilon)}{\sigma(X_t^\varepsilon)}\sigma'(X_t^\varepsilon)z_t - \varepsilon^2\sigma\sigma''(X_t^\varepsilon)z_t \right) dt, \tag{2.46}$$

i.e.

$$z_t = z_0 \exp\left(\int_0^t \left(b'(X_t^\varepsilon) - \frac{b(X_t^\varepsilon)}{\sigma(X_t^\varepsilon)}\sigma'(X_t^\varepsilon) - \varepsilon^2\sigma\sigma''(X_t^\varepsilon) \right) dt' \right). \tag{2.47}$$

Thus, if $\sup_{t \leq T} \sup_{|x| < \delta/2} |X_t^\varepsilon(x)| < \delta$, then the expression appearing in the integral in (2.47) is bounded from above by

$$-m_0 + \eta + (m_0 + \eta)\delta \frac{C_1}{\gamma} + \varepsilon^2 \frac{C_2}{\gamma} < -m \tag{2.48}$$

if $m < m_0$ and $\eta, \delta, \varepsilon$ are small enough.

Formula (2.45), (2.47) and (2.48) imply that:

$$\sup_x \frac{dX_t^\varepsilon(x)}{dx} \leq \frac{1}{\gamma} e^{-mt} \quad \forall t \leq T, \tag{2.49}$$

provided

$$\sup_{|x| < \delta/2} \sup_{t \leq T} |X_t^\varepsilon(x)| < \delta. \tag{2.50}$$

The Ventzel and Freidlin estimates give that the probability of the above event is greater than $1 - e^{-K/\varepsilon^2}$ for any $T > 0$ and ε small enough for some $K = K(T, \delta)$.

Proof of Theorem 2.1. The proof is almost identical to the proof of Theorem 1.1 and therefore it will be very sketchy. Let us fix a brownian path $\{w_t\}_{t \geq 0}$ and let for $a > 0$ $l = l(a, \varepsilon) = e^{a/\varepsilon^2}$; let also $T_j \equiv [(j-1)l, jl]$ $j \in N$. We introduce the set $S_0 = S_0(\{w_t\}_{t \geq 0})$ of singular blocks T_j by

$$S_0 \left\{ j \in N; X_{jl,(j-1)l}^\varepsilon(C) \not\subset C \text{ or } \sup_{x \in C} L(X_{jl,(j-1)l}^\varepsilon, x) > e^{-ml} \right\}. \tag{2.51}$$

Lemma 2.1. *For any $m < m_0$ there exist positive constants a_0, K, ε_0 such that for any $a < a_0, \varepsilon < \varepsilon_0$,*

$$P(j \in S_0) < \exp\left(-\frac{K}{\varepsilon^2} \right).$$

The proof is given in Appendix B.

For the rest of the discussion the parameter a appearing in the scale $l(a, \varepsilon)$ will be fixed smaller than a_0. Given the set S_0 one constructs, exactly as in Sect. 1 the net of sets $S_0 \supset S_1 \supset S_2 \cdots \supset S_k$ and proves the analogous of Proposition 1.1, 1.2, Lemma 1.2, (1.23), Lemma 1.3 provided that part a) of the next lemma holds (see Lemma 1.5).

Lemma 2.2. *For any $\rho \in (0, 1)$ any $a' > 0$ and any $V > 0$ there exist constants k, k' such that for any ε small enough,*

a) $\displaystyle\sup_{x \in C} P(\tau(x, 0, T) > \rho T) \leqq \exp\left[-kT \exp\left(-\frac{a'}{\varepsilon^2} \right) \right],$

b) $\displaystyle P\left(\sup_{x \in C} \tau(x, 0, T) > \rho T \text{ for some } T > T_0 \right) \leqq \exp\left(-\frac{k'}{\varepsilon^2} \right),$

where $\tau(x, 0, T)$ is the time spent up to T by the process X_t^ε starting at x, outside the set C and $T_0 = e^{V/\varepsilon^2}$.

The proof is given in Appendix B.

Remark. The analogue of Proposition 1.1 also proves (2.17). Let in fact $T > \bar{t}(\varepsilon)$ and let $d_{k(T)}$ be such that

$$l d_{k(T)} \leqq \alpha T < d_{k(T)+1}. \tag{2.52}$$

Clearly we can always choose the parameter a in $l(a, \varepsilon)$ such that $l d_0 < \alpha T$.

Using the analogous of Proposition 1.1 we have

$$P(\{ j \in N; T_j \subset [0, [\alpha T] + 1] \} \cap S_{k(T)} = \emptyset \ \forall T > 0)$$

$$\geqq P(\{ j \in N; j \leqq d_k \} \cap S_{k-1} = \emptyset \ \forall k \geqq 1) \geqq 1 - \exp\left(-\frac{K}{\varepsilon^2} \right)$$

for some constant $K > 0$.

From Lemma 1.4 we also know that

$$\#\{ n \in N; 0 \leqq n \leqq L, n \text{ is } (j-1) - \text{admissible} \} \geqq \left(1 - \text{const} \frac{d_0}{d_1} \right) L$$

with probability one.

Thus using (2.26), (2.27),

$$P(|S_2(T, \alpha, m)| > 2T/3 \ \forall T > \bar{t}(\varepsilon))$$

$$\geqq P(\{ j \in N; T_j \subset [0, [\alpha T] + 1] \} \cap S_{k(T)} = \emptyset \ \forall T > \bar{t}(\varepsilon)) \geqq 1 - \exp\left(-\frac{K_2}{\varepsilon^2} \right),$$

provided $(1 - \text{const } d_0/d_1) > 2/3$.

Section 3. Some Applications

We discuss some applications of our results to elliptic partial differential equations with a small parameter in front of the highest derivative.

Let $\vec{b}:\mathbf{R}^n \to \mathbf{R}^n$ be as in Sect. 2 and let $D \subset \mathbf{R}^n$ be a bounded domain with a smooth boundary ∂D such that D contains in its interior exactly k, $1 \leq k \leq l$ stable fixed points of $\vec{b}$. Let $\phi:\partial D \to \mathbf{R}$ be a smooth function and let $u:D \to \mathbf{R}$ be the solution of the Dirichlet problem:

$$Lu \equiv \frac{\varepsilon^2}{2}\Delta u + \vec{b} \cdot \vec{\nabla}u = 0 \quad \forall x \in D, \quad u|_{\partial D} = \phi. \tag{3.1}$$

Let now $x_{i_1}, \ldots, x_{i_k}$ be the stable fixed points contained inside D and for $T > 0, \delta > 0$ let

$$B_{i_j} = \{x \in D; |X_t(x) - x_{i_j}| < \delta \text{ for some } t \leq T\}. \tag{3.2}$$

Then we have:

Theorem 3.1. *There exist $\delta_0 > 0$ and for any $\delta < \delta_0$, $T > 0$, there exist ε_0, K such that*

$$\sup_{1 \leq j \leq k} \sup_{x,y \in B_{i_j}} |u(x) - u(y)| \leq \exp\left(-\frac{K}{\varepsilon^2}\right) \quad \forall \varepsilon < \varepsilon_0.$$

Proof. It is well known [22] that u has the probabilistic representation

$$u(x) = E_x \phi(X^\varepsilon_{\tau_e}) \quad x \in D, \tag{3.3}$$

where $\tau_x = \inf\{t \geq 0; X^\varepsilon_t(x) \in \partial D\}$.

Let us compute $u(x) - u(y)$ for $x, y \in B_{i_j}$,

$$u(x) - u(y) = E[\phi(X^\varepsilon_{\tau_x}(x)) - \phi(X^\varepsilon_{\tau_y}(y))]. \tag{3.4}$$

Using the Ventzel and Freidlin theory, we have that:

$$\sup_{x \in B_{i_j}} P(\tau_x < 1/\varepsilon^2) \leq \exp\left(-\frac{K'}{\varepsilon^2}\right) \tag{3.5}$$

for some $K' > 0$ and using Corollary 3.1 that

$$|X^\varepsilon_{\tau_x}(x) - X^\varepsilon_{\tau_x}(y)| \leq e^{-m\tau_x}|x - y| \tag{3.6}$$

with probability greater than $1 - \exp(-K''/\varepsilon^2)$ $K'' > 0$.

Suppose now that $\tau_y \geq \tau_x$, the case $\tau_x \geq \tau_y$ being identical. Then, if $\tau_x > 1/\varepsilon^2$ and (3.6) holds then

$$\text{dist}(X^\varepsilon_{\tau_x}(y), \partial D) < \exp\left(-\frac{m}{\varepsilon^2}\right)|x - y|. \tag{3.7}$$

Note that the length scale appearing in the right-hand side of (3.7) is much smaller than the typical fluctuations of the brownian motion εw_t if $t \gg O(\exp(-2m/\varepsilon^2))$. On this time scale the drift term in the stochastic differential equation 2.1 is negligible compared to εw_t. This, together with standard arguments on BM, leads to the following easy estimate whose proof is omitted:

$$\sup_{x;\text{dist}(x,\partial D)\leq\exp(-m/\varepsilon^2)} P\left(|X^\varepsilon_{\tau_x} - x| > \exp\left(-\frac{m'}{\varepsilon^2}\right)\right) \leq \exp\left(-\frac{K'''}{\varepsilon^2}\right), \qquad (3.8)$$

for any $m' < m$, ε small enough and a suitable constant $K'''(m') > 0$.

By combining together (3.6), (3.7), (3.8) we get that

$$\sup_{x,y\in B_{ij}} P\left(|X^\varepsilon_{\tau_x} - X^\varepsilon_{\tau_y}| \leq \exp\left(-\frac{m'}{\varepsilon^2}\right)\right) > 1 - \exp\left(-\frac{K}{\varepsilon^2}\right) \qquad (3.9)$$

for any ε small enough and suitable constant $m' > 0$, $K > 0$. The above estimate together with the assumed smoothness of ϕ and (3.4) clearly gives the result.

Our result can be applied as well to the study of the distribution as $\varepsilon \to 0$ of the exit time τ^ε from an attracted domain D.

The exponential low for τ^ε has been already proved by Day [9], however, by means of our result on the exponential glueing of trajectories the exponential low can be proved in more general situations.

In [11] we will treat this problem and we will also consider the infinite dimensional case.

Appendix B

We first prove the bound (2.35) in the hypothesis of Proposition 2.1. Let i_0 be, as in Proposition 2.1, the index of the "most stable" fixed point. Let $\tau_T(x)$ be the time spent by the process $X^\varepsilon_t(x)$ up to time T in the set $C^{[i_0]} \equiv \bigcup_{i \neq i_0} C_i$. Finally let:

$$V_0 \equiv V\{i_0\} \quad (\text{see } (2.9)), \quad V_1 \equiv \min_{j \neq i_0} V_{i_0,j}. \qquad (B.1)$$

By assumption $V_1 > V_0$.

The proof is based on the following estimate:

For any $V, V_0 < V < V_1$ and any $\alpha, \alpha \in (0,1)$ there exists $K > 0$ such that

$$P\left(\sup_{x\in C} \tau_T(x) \geq \alpha T \text{ for some } T \geq T_0\right) \leq \exp\left(-\frac{K}{\varepsilon^2}\right) \qquad (B.2)$$

for any ε sufficiently small, provided $T_0 = e^{V/2\varepsilon^2}$.

In fact if B.2 holds, then we have

$$P\left(\inf_{x,y\in C} |S_1(T,x,y)| < 2T/3 \text{ for some } T \geq T_0\right)$$

$$\leq P\left(\sup_{x\in C} \tau_T(x) \geq T12 \text{ for some } T \geq T_0\right)$$

$$+ P\left(\sup_{x\in C} \tau(x,0,T) \geq T/12 \text{ for some } T \geq T_0\right), \qquad (B.3)$$

where $\tau(x,0,T)$ is the time spent up to T by the process X^ε_t starting at x outside the set C (see Lemma 2.2). In fact if we denote by $\tau_{i_0}(x,T)$ the time spent by the process $X^\varepsilon_t(x)$ up to time T in the set $C_{i_0} \subset A_{i_0}$, if $\tau(x,0,T) < T/12$ and if $\tau_T(x) < T/12$,

then

$$\tau_{i_0} = T - \tau(x, 0, T) - \tau_T(x) > T(1 - \tfrac{1}{12} - \tfrac{1}{12}) = T5/6.$$

Thus if the same bound holds also for $\tau_{i_0}(y, T)$, then we get

$$T5/3 - |S_1(T, x, y)| < \tau_{i_0}(x, T) + \tau_{i_0}(y, T) - |S_1(T, x, y)| \leq T,$$

i.e. $|S_1(T, x, y)| > T2/3$.

By using the estimate (B.2) and Lemma 2.2 we bound (B.3) by $\exp(-K_1/\varepsilon^2)$ for some suitable constant K_1 and for ε sufficiently small.

Let us now prove B.2:

$$E\chi\left(\sup_{x\in C} \tau_T(x) > \alpha T \text{ for some } T \geq T_0\right)$$

$$\leq E\Bigg[\chi\left(\sup_{x\in C} \tau_T(x) > \alpha T \text{ for some } T \geq T_0\right)$$

$$\chi\left(\sup_i \sup_{x\in C_i} \frac{\|X_T^\varepsilon(x) - X_T^\varepsilon(x_i)\|}{\|x - x_i\|} \leq e^{-mT} \,\forall\, T > T_0\right)\Bigg]$$

$$+ E\Bigg[1 - \chi\left(\sup_i \sup_{x\in C_i} \frac{\|X_T^\varepsilon(x) - X_T^\varepsilon(x_i)\|}{\|x - x_i\|} \leq e^{-mT} \,\forall\, T > T_0\right)\Bigg]$$

$$\leq \sum_i P(\tau_T(x_i) > \alpha T \text{ for some } T \geq T_0) + \exp\left(-\frac{K}{\varepsilon^2}\right). \tag{B.4}$$

For the last inequality to hold it is understood that $\tau_T(x_i)$ is evaluated for a δ-neighborhood $C_\delta^{[i_0]}$ of the original set $C^{[i_0]} = \bigcup_{i \neq i_0} C_i$. This of course has little influence on the subsequent estimates if $\delta \ll 1$.

To estimate $P(\tau_T(x_i) > \alpha T \text{ for some } T \geq T_0)$ we introduce the random times τ_j given by the time spent by the process in the set $C_\delta^{[i_0]}$ between $((j-1)T_0, jT_0]$.

The event $\tau_T > \alpha N T_0$, $N \geq 1$, implies that for at least $(\alpha/2 - \alpha)N$ intervals $((j-1)T_0, jT_0]$, $\tau_j > \alpha T_0/2$. Therefore

$$\sup_i P_{x_i}(\tau_T > \alpha N T_0)$$

$$\leq \sup_i P_{x_i}\left(\text{there exist at least } \frac{\alpha}{2-\alpha}N \text{ intervals } ((j-1)T_{0,j}, T_0]; \tau_j > \alpha T_0/2\right). \tag{B.5}$$

Using the Markov property, we estimate (B.5) by:

$$\sup_i P_{x_i}(\tau_T > \alpha N T_0) \leq \sum_{k > N\alpha'}^N \binom{N}{k}\left(\sup_x P_x(\tau_1 > \alpha T_0/2)\right)^k$$

$$\leq \left(\sup_{x\in C_\delta^{[i_0]}} P(\tau_1 > \alpha T_0/2)\right)^{N(\alpha/2 - \alpha)} 2^N. \tag{B.6}$$

We now estimate $\sup_{x\in C_\delta^{[i_0]}} P_x(\tau_1 > \alpha T_0/2)$.

Let

$$\sigma_0 \equiv \inf\{t \geq 0; X_t^\varepsilon(x) \in C_{i_0}\},$$
$$\sigma \equiv \inf\{t \geq 0; X_t^\varepsilon(x) \in C_\delta^{[i_0]}\}.$$

Then

$$\sup_{x \in C_\delta^{[i_0]}} P_x(\tau_1 > \alpha T_0/2) \leq \sup_{x \in B_r} \frac{2E_x\tau_1}{\alpha T_0} \leq \frac{2}{\alpha T_0}\left(\sup_{x \in C_\delta^{[i_0]}} E_x\sigma_0 + \sup_{x \in C_{i_0}} E_x\tau_1\right). \quad (B.7)$$

Using the Ventzel and Freidlin theory, together with hp1, one gets:

$$\lim_{\varepsilon \to 0} 2\varepsilon^2 \log \sup_{x \in C_\delta^{[i_0]}} E_x\sigma_0 = V\{i_0\}. \quad (B.8)$$

The second term in the right-hand side of (B.7) is bounded as follows:

$$\sup_{x \in C_{i_0}} E_x\tau_1 \leq \sup_{x \in C_{i_0}} P(\sigma < T_0) \sup_{x \in C_\delta^{[i_0]}} E_x\tau_1, \quad (B.9)$$

and therefore, using (B.7), we get

$$\sup_{x \in C_{i_0}} E_x\tau_1 \leq \frac{\displaystyle\sup_{x \in C_\delta^{[i_0]}} E_x\sigma_0}{1 - \displaystyle\sup_{x \in C_{i_0}} P(\sigma < T_0)}.$$

Again by the Ventzel and Freidlin theory we get that for any ε sufficiently small,

$$\sup_{x \in C_{i_0}} P(\sigma < T_0) < 1/2.$$

In conclusion for any $h > 0$ there exists $\varepsilon_0 > 0$ such that for any $\varepsilon < \varepsilon_0$,

$$\sup_{x \in C_\delta^{[i_0]}} E_x\tau_1 \leq \exp\left[-\frac{V\{i_0\} + h}{2\varepsilon^2}\right], \quad (B.10)$$

that is

$$\sup_{x \in C_\delta^{[i_0]}} P_x(\tau_1 > \alpha T_0/2) \leq \exp\left[-\frac{(V - V\{i_0\}) + 2h}{2\varepsilon^2}\right],$$

and (B.2) is proved.

Proof of Lemma 2.1. Let $t_0 = \left[\sup_{x \in A}\{t; X_t(x) \in C_0\}\right] + 1$ and let us first prove that

$$P\left(X_{t_0}^\varepsilon(C) \subset C \text{ and } \sup_{x \in C} L(X_{t_0}^\varepsilon, x) < e^{-mt_0}\right) > 1 - \exp\left(-\frac{2K}{\varepsilon^2}\right) \quad (B.14)$$

for a suitable constant $K = K(d)$ depending only on the distance between the sets C_0, C, A. It is clear that (B.14) proves the lemma provided the constant a_0 is taken sufficiently small, e.g. $a_0 = K$.

   In fact

$$P\left(X_l^\varepsilon(C) \subset C \text{ and } \sup_{x \in C} L(X_l^\varepsilon, x) < e^{-ml}\right)$$

$$\geq 1 - \sum_{j=1}^{l/t_0}\left(1 - P(X_{jt_0,(j-1)t_0}^\varepsilon(C) \subset C \text{ and } \sup_{x \in C} L(X_{t_0}^\varepsilon, x) < e^{-mt_0})\right)$$

$$\geqq 1 - e^{a/\varepsilon^2} \exp\left(-\frac{2K}{\varepsilon^2}\right) \geqq 1 - \exp\left(-\frac{k}{\varepsilon^2}\right) \tag{B.15}$$

if $a < a_0$.

In (B.15) we have used the obvious fact that

$$L(X_t^\varepsilon, x) \leqq \prod_{j=1}^{l/t_0} L(X_{jt_0,(j-1)t_0}^\varepsilon, X_{(j-1)t_0}^\varepsilon(x)).$$

Using estimate 2.14,

$$P(X_{t_0}^\varepsilon(C) \subset C) \geqq 1 - \exp\left(-\frac{K'}{\varepsilon^2}\right) \tag{B.16}$$

for a suitable constant $K' > 0$.

In order to estimate $\sup_{x \in C} L(X_{t_0}^\varepsilon, x)$ we assume that

$$\|X_t^\varepsilon(x) - X_t(x)\| \leqq \delta \quad \forall t \in (0, t_0) \tag{B.17}$$

with $\delta \ll 1$.

For notational convenience we denote by

$$Z_{k,i}^\varepsilon(x, t) = \frac{\partial}{\partial x_i}(X_t^\varepsilon(x))_k \tag{B.18}$$

and by $Z_{k,i}(x, t)$ the analogue for the classical path X_t. Then:

$$L(X_{t_0}^\varepsilon, x) \leqq L(X_t, x) + \text{const} \sup_{k,i}|Z_{k,i}^\varepsilon(x, t_0) - Z_{k,i}(x, t_0)|. \tag{B.19}$$

Let us write

$$\phi(t) = \sup_{k,i}|Z_{k,i}^\varepsilon(x, t) - Z_{k,i}(x, t)|.$$

Then by subtracting Eq. (2.2) from (2.1) we get

$$\phi(t) \leqq \int_0^t ds\, C_1 \phi(s) + C_2 t\delta, \quad \phi(0) = 0, \tag{B.20}$$

where

$$C_1 = \sup_{x \in A} \sup_{k,i}\left|\frac{\partial b_k(\vec{x})}{\partial x_i}\right|,$$

and C_2 depends on the supremum over $x \in A$ of the second derivatives of the drift $\vec{b}$.

By the Gronwall inequality we obtain:

$$\phi(t_0) \leqq C_2 t_0 \delta e^{C_1 t_0}.$$

By hypothesis hp3,

$$L(X_{t_0}, x) \leqq e^{-m_0 t_0},$$

and therefore for any $m < m_0$ if δ is sufficiently small,

$$\sup_{x \in C} L(X_{t_0}^\varepsilon, x) \leqq e^{-m t_0}. \tag{B.21}$$

Using again estimate (2.14) the probability of (B.17) is bounded from below by

$$1 - \exp\left(-\frac{k''}{\varepsilon^2} \right) \tag{B.22}$$

for some constant $k'' > 0$.

Estimates (B.22) together with (B.16) conclude the proof.

Proof of Lemma 2.2. a) Let τ_R be the time spent up to time T outside the sphere B_R and let σ_R be the time spent up to time T in the set $B_R \backslash C$. Then we have

$$P(\tau(x, 0, T) > \alpha T) < P(\tau_R(x, 0, T) > \alpha T/2) + P(\sigma_R(x, 0, T) > \alpha T/2). \tag{B.23}$$

We estimate the two terms in (B.23) separately. We start with the first one. Let

$$\tau_1 = \inf(t > 0; X_t^\varepsilon(x) \notin B_R),$$
$$\sigma_1 = \inf(t > \tau_1; X_t^\varepsilon \in B_{R/2}),$$

$$\tau_i = \inf(t > \sigma_{i-1}; X_t^\varepsilon(x) \notin B_R).$$

Then we have (see (A.9) for the derivation)

$$P(\tau_R(x, 0, T) > \alpha T/2) < \exp(-\beta \alpha T) \sum_n \sup_{x \in B_R} (P_x(v = n))^{1/2} \left(\sup_{x \in \partial B_R} E_x(e^{2\beta \sigma_1})^{n/2} \right) \tag{B.24}$$

for any $\beta > 0$ where v is the smallest integer such that $\sigma_v > T$.

Lemma B.1. *For any $\alpha' < 1$ and any $n > \alpha' T$ and ε sufficiently small,*

$$P(v = n) < \exp\left(-\frac{kn}{\varepsilon^2} \right)$$

for some constant $k > 0$.

Lemma B.2. *There exist $\beta_0 > 0$ such that $\forall \beta < \beta_0$,*

$$\sup_{x \in \partial B_R} E_x(e^{2\beta \sigma_1}) \leq 2.$$

Clearly the two Lemmas show that there exists a constant $k = k(R)$ such that

$$P(\tau_R(x, 0, T) > \alpha T/2) < e^{-kT}. \tag{B.25}$$

Proof of Lemma B.1. If $v = n$ then we have that $\text{dist}(X_t^\varepsilon, \Phi_n) > \delta$ for some small δ, where

$$\Phi_n = (\phi; I_{0,T}(\phi) > nV(R)/2)$$

with

$$V(R) = \inf_{\phi(0) \in B_R, \phi(T) \notin B_{2R}} I_{0,T}(\phi).$$

The probability of the above event is estimated by the second of the Ventzel and Freidlin estimates by

$$\exp\left(\frac{-nV(R)((1-\alpha')/2) + \alpha' T}{\varepsilon^2} \right)$$

for any $\alpha' < 1$ and ε sufficiently small.

Proof of Lemma B.2. We estimate

$$\sup_{x \in \partial B_R} P_x(\sigma_1 > t).$$

Let

$$I = \inf_{(\phi; \phi(s) \notin B_{R/3} \,\forall s \le t \phi(0) \in \partial B_R)} I_{0,t}(\phi).$$

Clearly the event

$$\{\sigma_1 > t\}$$

implies that

$$\text{dist}(X_t^\varepsilon(x), \{\phi; I_{0,t}(\phi) < I; \phi(0) = x\}) > R/6 \quad \forall x \in \partial B_R.$$

We will show that $I \ge ct$ for a suitable constant $c > 0$. The second of the Ventzel and Freidlin estimates then implies that:

$$\sup_{x \in \partial B_{2R}} P_x(\sigma_1 > t) \le \exp\left(-\frac{kt}{\varepsilon^2}\right)$$

for some $k > 0$. This clearly proves the lemma. Thus let us compute I. For a given ϕ satisfying the above conditions we write

$$\int_0^t ds |d\phi(s)/ds - b(\phi(s))|^2 = \sum_0^v \int_{s_{i-1}}^{s_i} ds |d\phi(s)/ds - b(\phi(s))|^2 + \int_{s_v}^t ds |d\phi(s)/ds - b(\phi(s))|^2,$$

where

$$s_0 = 0; \quad s_i = \min(s > s_{i-1}; \phi(s)) \in \partial B_{(n_{i-1}+1)R} \cup \partial B_{(n_{i-1}-1)R},$$

where n_{i-1} is such that $\phi(s_{i-1}) \in \partial B_{n_{i-1}R}$, and v is such that $s_{v+1} > t > s_v$. Let $t_0 = R/a$ with a as in hp1. If now $v > t/2t_0$ then this implies that $n_i > n_{i-1}$ for at least $t/4t_0$ times since $|\phi(s)| > R/3 \,\forall s < t$; thus in this case $I > ct$ for a suitable constant $c > 0$. If $v < t/2t_0$ then, following Ventzel and Freidlin, we estimate each term in the sum by $c'[(s_i - s_{i-1})/t_0]$ for some constant $c' > 0$. Thus in this case we get:

$$I > s_v/t_0 - v + (t - s_v)/t_0 - 1 = t/2t_0 - 1.$$

The proof is finished.

We turn to the estimate of the second term in (B.23). Following the proof of Lemma 1.5, we start by showing that for any $a > 0$ and any $\beta > e^{-a/\varepsilon^2}$,

$$\sup_{x \in B_R} E_x e^{\beta \alpha_1} \le 2 \tag{B.26}$$

for ε sufficiently small where α_1 is the exist time from $B_R \backslash C$.

Let $t_0 = \exp(a/\varepsilon^2)$; Then, by the strong Markov property and the Chebyshev inequality we have

$$P(\alpha_1 > kt_0) < \left(\sup_{x \in B_R} E_x(\alpha_1)/t_0\right)^k. \tag{B.27}$$

The results of Ventzel and Freidlin show that for any $a' > 0$ there exists ε_0 such

that for any $\varepsilon < \varepsilon_0$

$$\sup_{x \in B_R} E_x(\sigma_R) < e^{a'/\varepsilon^2}.$$

Thus for ε small enough,

$$P(\alpha_1 > k t_0) < \exp\left(-\frac{a'k/2}{\varepsilon^2} \right). \qquad (B.28)$$

The second and final ingredient is to estimate the probability $P_x(v_t = n)$ where v_t counts the number of transitions between the sets $B \equiv C^c$ and C_0 (see (A.6)). For any $a > 0$ let $[0, t] = \bigcup_{i < N} [t_{i-1}, t_i]$, where $N = [t/e^{a/\varepsilon^2}] + 1$. Here $t_i = i e^{a/\varepsilon^2}$ if $i < N - 1$ and $t_N = t$.

Then we have

$$P_x(v_t = n) < \sum_{n_1, \ldots, n_N} P(n_1, \ldots, n_N), \qquad (B.29)$$

where n_i is the number of transitions in the ith interval of time with the restriction $\sum_i n_i = n$ and $P(n_1, \ldots, n_N)$ denotes the probability of having n_1 transitions in $[0, t_1], n_2$ in the second interval and so on. By the strong Markov property we can bound $P(n_1, \ldots n_N)$ by $p(n_1)p(n_2) \ldots p(n_N)$, where $p(j) = \sup_{x \in C_0} P_x(v_{t_1} = j)$. Again by the strong Markov property we get

$$p(j) \leq p(1)^j. \qquad (B.30)$$

It remains to estimate $p(1)$. As in the first part of the proof of Lemma 2.1 it is sufficient to estimate $p(1)$ for a time interval T_0 independent of ε. This last estimate is provided by the standard large deviation estimates of Ventzel and Freidlin; the final result is

$$p(1) \leq \exp\left(-\frac{k}{\varepsilon^2} \right) \qquad (B.31)$$

for some $k > 0$ and any a small enough.

Using now (B.31), (B.30) in (B.29) we get that

$$P_x(v_t = n) \leq \exp\left(-\frac{kn}{\varepsilon^2} \right) e^n c^N, \qquad (B.32)$$

where $c = \sum_n e^{-n}$. Since $N < t \exp(-a/\varepsilon^2) + 1$, (B.32) represents the analogue of the result of Lemma A.1.

By proceeding now exactly as in the discrete case we get that (B.26) together with (B.32) leads to the estimate

$$P(\sigma_R(x, 0, T) > \alpha T/2) \leq \exp\left(-k(\alpha) T \exp\left(-\frac{a}{\varepsilon^2} \right) \right)$$

for some constant $k(\alpha)$, any a sufficiently small and ε small enough.

The above estimate, together with (B.25), proves Lemma 2.2a).

b)

$$P(\exists t > T_0; \sup_{x \in C} \tau(x, 0, t) > \rho t) \leqq \sup_{x \in C} P_x(\exists t > T_0; \tau'(x, 0, t) > \rho t)$$

$$+ P\left(\sup_{x, y \in C} \text{dist}(X_t^{\varepsilon}(x), X_t^{\varepsilon}(y)) > e^{-mT_0}\right), \quad (B.29)$$

where $\tau'(x, 0, t)$ is defined as $\tau(x, 0, t)$ but computed for a slightly smaller set $C' \subset C$ with e.g. $\text{dist}(\partial C, C') \leqq \exp(-1/\varepsilon^2)$. The result now follows from Corollary 2.1 and part a) of the lemma.

References

1. Jona-Lasinio, G., Martinelli, F., Scoppola, E.: New approach to the semiclassical limit of quantum mechanics. I: Multiple tunneling in one dimension. Commun. Math. Phys. **80**, 223 (1981)
2. Cassandro, M., Galves, A., Olivieri, E., Vares, M. E.: Metastable behaviour of stochastic dynamics: A pathwise approach. J. Stat. Phys. **35**, No 5/6 (1984)
3. Ruelle, D.: Small random perturbations of dynamical systems and the definition of attractors. Commun. Math. Phys. **82**, 137 (1981)
4. Young, L. S.: Stochastic stability of hyperbolic attractors. Eng. Th. Dyn. Syst. **6**, 311 (1986)
5. Schaffer, W. M., Ellner, S., Kot, M.: Effects of noise on some dynamical models of ecology. J. Math. Biol. **24**, 479 (1986)
6. Ventzel, A., Freidlin, M. I.: On small random perturbation of dynamical systems. Usp. Math. Nauk. **25**, 3 (1970). Engl. transl. Russ. Math. Surv. **25**, 1 (1970). See also Fluctuations in dynamical systems under the action of random perturbations. Moskva: Nauka 1979
7. Kifer, Yu.: On small random perturbations of some smooth dynamical systems. Math. USSR Izvestija **8**, 1083 (1974)
8. Gora, P.: Random composing of mappings, small stochastic perturbations and attractors. Z. Wahrscheinlichtstheorie Verw. Geb. **69**, 137 (1985)
9. Day, M. V.: On the exponential exist law in the small parameter exist problem. Stochastics Verw. **8**, 297 (1983)
10. Galves, A., Olivieri, E., Vares, M. E.: Metastability for a dynamical system subject to a small random perturbation. Ann. Prob. (1987)
11. Martinelli, F., Olivieri, E., Scoppola, E.: Small random perturbations of finite and infinite dimensional dynamical systems: Unpredictability of exit times, in preparation
12. Derrida, B.: Dynamical phase transitions in non-symmetric spin glasses. J. Phys. A: Math. Gen. **20**, L721 (1987)
13. Kifer, Yu.: Ergodic theory of random transformations. Progress in probability and statistics. Boston: Birkhäuser 1986
14. Carverhill, A. P.: Flows of stochastic dynamical systems: Ergodic theory. Stochastics **14**(4), 273 (1985)
15. Carverhill, A. P., Chappel, M. J., Elworthy, K. D.: Characteristic exponents for stochastic flows. In: Stochastic processes, Mathematics and Physics, Proceedings 1984. Lecture Note in Mathematics, Vol. **1158**. Berlin, Heidelberg, New York: Springer 1986
16. Le Jan, Y.: Equilibre statistique pour les produits de diffeomorphismes aleatoires independants. Ann. Inst. H. Poincare **43**, No 1, 111 (1987)
17. Fröhlich, J., Martinelli, F., Scoppola, E., Spencer, T.: Constructive proof of localization in the Anderson tight-binding model. Commun. Math. Phys. **101**, 21 (1985)
18. Martinelli, F., Scoppola, E.: Introduction to the mathematical theory of Anderson localization. La Rivista del Nuovo Cimento **10**, serie 3, n. 10 (1987)

19. Martinelli, F., Scoppola, E.: Random composition of two rational maps: Singularity of the invariant measure. J. Stat. Phys. **50**, 516, 1021 (1988)
20. Fröhlich, J., Spencer, T.: Absence of diffusion in the Anderson tight-binding model for large disorder or low energy. Commun. Math. Phys. **88**, 151 (1983)
21. Jona-Lasinio, G., Martinelli, F., Scoppola, E.: Tunneling in one dimension: General theory, instabilities, rules of calculation, applications. In: Mathematics + Physics, Lectures on recent results, Vol. II. Streit, L. (ed.) Singapore: World Scientific 1986
22. Guikhman, I., Skorokhod, A.: Introduction a la theorie des processus aleatoires. Moscou: Editions MIR 1980