Ergod. Th. & Dynam. Sys. (1998), **18**, 53–87
Printed in the United Kingdom

Recursive algorithms, urn processes and chaining number of chain recurrent sets

MICHEL BENAÏM
Department of Mathematics, Université Paul Sabatier, Toulouse, France

(*Received* 26 *January* 1996 *and accepted in revised form* 10 *May* 1996)

Abstract. This paper investigates the dynamical properties of a class of urn processes and recursive stochastic algorithms with constant gain which arise frequently in control, pattern recognition, learning theory, and elsewhere.

It is shown that, under suitable conditions, invariant measures of the process tend to concentrate on the Birkhoff center of irreducible (i.e. chain transitive) attractors of some vector field $F : \mathbb{R}^d \to \mathbb{R}^d$ obtained by averaging. Applications are given to simple situations including the cases where F is Axiom A or Morse–Smale, F is gradient-like, F is a planar vector field, F has finitely many alpha and omega limit sets.

1. *Introduction*

This paper considers a family of discrete time stochastic processes $\{X_n^\epsilon\}_{n\in\mathbb{N}}$, $\epsilon > 0$, living in $\mathbb{R}^d$ which are defined in the following way.

Let $\mathcal{X} = \{1, \ldots, m\}$ be a finite state space called the *space of control parameters*. For each $x \in \mathbb{R}^d$ we assume that we are given the following:

- a discrete time Markov chain on $\mathcal{X}$ represented by a $m \times m$ transition matrix $K(x) = \{K_{i,j}(x)\}_{i,j\in\mathcal{X}}$ satisfying

$$K_{i,j}(x) \geq 0; \quad \sum_{j=1}^{m} K_{i,j}(x) = 1;$$

- a family $\{\mu_x^1, \ldots, \mu_x^m\}$ of m probability measures on $T_x\mathbb{R}^d = \mathbb{R}^d$.

Let ϵ denote a (small) positive real parameter called the *gain parameter*. We consider a Markov process $\{(X_n^\epsilon, \Theta_n^\epsilon)\}_{n\in\mathbb{N}}$ defined on a probability space $\{\Omega, \mathcal{F}, \mathbf{P}\}$ taking values in $\mathbb{R}^d \times \mathcal{X}$ whose transition kernel is given by

$$\mathbf{P}((X_{n+1}^\epsilon, \Theta_{n+1}^\epsilon) \in A \times \{j\} \mid X_n^\epsilon = x, \Theta_n^\epsilon = i) = K_{i,j}(x)\mu_x^j\left(\frac{-x + A}{\epsilon}\right) \tag{1}$$

for every $i, j \in \mathcal{X}, x \in \mathbb{R}^d$ and every Borel set $A \subset \mathbb{R}^d$, where $(-x + A)/\epsilon = \{(-x + f)/\epsilon : f \in A\}$.

An equivalent and more concrete formulation is the following. For each $x \in \mathbb{R}^d$ and each $\theta \in \mathcal{X}$ let $\{\mathbf{f}_n(x, \theta)\}_{n \in \mathbb{N}}$ denote a sequence of independent random vectors taking values in $T_x\mathbb{R}^d = \mathbb{R}^d$ having μ_x^θ as the probability distribution. Then the process $\{X_n^\epsilon, \Theta_n^\epsilon\}_{n \geq 0}$ is defined by

$$\begin{cases} \mathbf{P}(\Theta_{n+1}^\epsilon = j \mid X_n^\epsilon = x, \Theta_n^\epsilon = i) = K_{i,j}(x) \\ X_{n+1}^\epsilon - X_n^\epsilon = \epsilon \mathbf{f}_{n+1}(X_n^\epsilon, \Theta_{n+1}^\epsilon). \end{cases} \tag{2}$$

Recursive processes described by (1) or (2) arise frequently in applications in pattern recognition, system identification, learning theory, economic modeling and elsewhere under the name *stochastic approximations and adaptive algorithms*. While this theory has been the focus of much attention over the last four decades in the engineering and probabilistic literature (see, e.g., [**7**, **11**, **25**, **26**]), little attention has been paid to dynamical systems issues. In this paper we continue the program initiated in [**3**, **5**, **6**] which is to investigate the asymptotic behavior of stochastic approximation processes with the point of view of abstract dynamical system theory. Our previous work was devoted to stochastic approximation processes with decreasing gain†. It was shown that under mild assumptions the limit sets of sample paths of stochastic approximation processes are almost surely connected internally chain recurrent sets for the dynamics of a suitable average vector field $F : \mathbb{R}^d \to \mathbb{R}^d$. This property was used in [**5**] to characterize the limiting behavior of a class of urn processes for which F is Morse–Smale.

This paper is concerned with processes with constant gain, that is $\epsilon > 0$. A first step is made toward describing the topological and dynamical structure of the support of invariant measures of (1) in the limit $\epsilon \to 0$. An important source of inspiration for the present paper are the works of David Ruelle [**29**] and Yuri Kifer [**24**] on random perturbations of dynamical systems. Ruelle [**29**] shows that, under suitable conditions, discrete time dynamical systems with small bounded random perturbations asymptotically live on irreducible (i.e. chain transitives) attractors. Kifer [**24**] relying on Freidlin and Weintzell theory, considers the case of discrete time dynamical systems with unbounded random perturbation which satisfy a large deviation property. This paper presents similar results for stochastic approximation processes. One of our main purposes is to describe the long term behavior of (1) by using only the dynamics of F (and not the particular properties of the data $K(x)$ and $\mu_x^1, \ldots, \mu_x^m$). To achieve this goal we introduce in §2 the notion of *chaining number* of a chain transitive set‡. Roughly speaking the chaining number of a chain transitive set Γ measures the growth rate (as $\delta \to 0$) of the minimal number of (δ, T) pseudo-orbits required to connect any given pair of points in Γ. Our main assumption will be that some components of the limit points set of F§ have their chaining number bounded by 2.

It turns out that in some simple cases including Axiom A or Morse–Smale systems, gradient-like systems, planar systems with finitely many equilibria, vector fields with finitely many alpha and omega limit sets, etc, this assumption is easily seen to be satisfied.

† That is $\epsilon = \epsilon_n$ with $\sum_n \epsilon_n = \infty$ and $\sum_n \epsilon_n^{1+\delta} < \infty$ for some $\delta > 0$. For an introduction and a recent survey of this theory see, e.g., [**12**].

‡ A chain transitive set is defined here as a subset of a chain transitive component of F (see §2 and 5).

§ The closure of all alpha and omega limit sets of F.

The key ingredients for our proofs are a large deviation principle for (1) due to Dupuis [**13**] (see also [**26**]) combined with some properties of chain recurrent flows, and some ideas and techniques introduced by Kifer [**24**].

Example 1.1. Generalized urn processes. To give a simple example, we describe here an urn process in the spirit of the *generalized urn processes* which have been considered by Hill *et al* [**18**], Arthur *et al* [**2**], Pemantle [**27**], Benaïm and Hirsch [**5**] among others.

Let

$$\Delta^d = \left\{ v \in \mathbb{R}^{d+1} : v_i \geq 0, \sum v_i = 1 \right\}$$

be the *unit d-simplex*. Let $p^+ : \Delta^d \to \Delta^d$ and $p^- : \Delta^d \to \Delta^d$ be two functions such that p^- leaves invariant the faces of the simplex (i.e. $x_i = 0$ implies $p_i^-(x) = 0$). Throughout we will call such a pair (p^+, p^-) an *urn function*.

Now, consider an urn which contains $N > 0$ balls of colors $1, \ldots, d+1$. Set

$$\epsilon = \frac{1}{N}.$$

At each time step a ball is randomly chosen in the urn and is replaced by a new ball according to the following process.

Let $X^\epsilon_{n,i}$ be the proportion of balls having color i at time n and denote by $X^\epsilon_n = (X^\epsilon_{n,1}, \ldots, X^\epsilon_{n,d+1})$ the vector of proportions. The color of the ball removed from the urn at time $n+1$ is chosen to be i with probability $p_i^-(X^\epsilon_n)$ and the color of the ball which is added is chosen to be j with probability $p_j^+(X^\epsilon_n)$. This defines a stochastic process $\{X^\epsilon_n\}_{n\geq 0}$ living in Δ^d.

Let

$$\mathcal{X} = \{1, \ldots, d+1\} \times \{1, \ldots, d+1\},$$

and let e_i denote the ith vector of the canonical basis of $\mathbb{R}^{d+1}$. Let $E^d \subset \mathbb{R}^{d+1}$ be the affine subspace spanned by Δ^d. By translating coordinates, we assume that E^d is a linear subspace and Δ^d is a neighborhood in E^d of the origin. For any $x \in E^d$ let $r(x)$ denote the nearest point of Δ^d to x. The map $r : E^d \to \Delta^d$ is easily proved to be Lipschitz.

For each $x \in E^d$, define the Markov transition matrix $K(x)$ as

$$K_{(i_0,j_0),(i,j)}(x) = p_j^+(r(x))p_i^-(r(x)), \quad (i_0, j_0) \in \mathcal{X}, (i, j) \in \mathcal{X} \tag{3}$$

and define the probability measures

$$\mu_x^{i,j} = \delta_{e_j - e_i}, \quad (i, j) \in \mathcal{X}, \tag{4}$$

where δ_f stands for the dirac measure at f. Identifying E^d with $\mathbb{R}^d$, formulae (3), (4) and (1) define a process in $\mathbb{R}^d \times \mathcal{X}$ whose first component coincides with $\{X^\epsilon_n\}_{n\in\mathbb{N}}$ on Δ^d.

Outline of contents. The organisation of the paper is as follows. §2 introduces the notation and some definitions. It briefly reviews the notion of chain recurrence and introduces the chaining number of a chain transitive set. The main results are presented in §3 and some applications are developed in §4. The properties of chain recurrence needed for the proofs are given in §5. The large deviation principle for (1) is presented

https://doi.org/10.1017/S0143385798097557 Published online by Cambridge University Press

in §6 together with some dynamical implications of this principle based on a result of Nitecki and Shub extended by Akin [**1**]. The final estimates needed for Theorem 3.7 are given in §7.

2. *Basic definitions and main hypotheses*

The notation, definitions and hypotheses introduced in this section will retain their validity throughout the paper. The Euclidean norm on $\mathbb{R}^d$ is denoted $\| \cdot \|$, and $\langle \cdot , \cdot \rangle$ is the associated inner product. The closure, interior and boundary of a set $A \subset \mathbb{R}^d$ are respectively denoted $\mathrm{clos}(A)$ (or $\overline{A}$), $\mathrm{int}(A)$ and ∂A. The complementary set of A is $A^c = \mathbb{R}^d \setminus A$. The δ neighborhood of A is $N_\delta(A) = \{x \in \mathbb{R}^d : d(x, A) < \delta\}$. The open ball of radius $r > 0$ centered at the origin is $B_r = \{x \in \mathbb{R}^d : \|x\| < r\}$ and the open ball of radius $r > 0$ and center a is $B_r(a) = B(a, r) = \{x \in \mathbb{R}^d : \|x - a\| < r\}$.

Consider the process introduced in §1 and characterized by (1) or (2). $\mathbf{P}_{x,\theta}$ (respectively $\mathbf{E}_{x,\theta}$) denotes the probability (respectively expectation) given that $X_0^\epsilon = x$ and $\Theta_0^\epsilon = \theta$. For $x \in \mathbb{R}^d$ we let

$$\hat{f}_i(x) = \int_{\mathbb{R}^d} f \mu_x^i(df) \in \mathbb{R}^d, \quad i = 1, \dots, m$$

denote the average of μ_x^i and we let $\Theta^x = \{\Theta_n^x\}_{n \geq 0}$ denote the homogeneous Markov chain on $\mathcal{X}$ whose transition matrix is $K(x)$.

A set $\mathcal{X}' \subset \mathcal{X}$ is called a *recurrence class* for Θ^x if for all $i, j \in \mathcal{X}'$ there exists an integer $k = k(i, j)$ and a finite set $\{i_1, \dots, i_k\} \subset \mathcal{X}'$ such that $i_1 = i, i_k = j$ and $K_{i_l, i_{l+1}}(x) > 0, l = 1, \dots, k - 1$. If k can be chosen independently of i, j then $\mathcal{X}'$ is called *aperiodic*. The Markov chain Θ^x is called *indecomposable* if it admits a single recurrence class. It is called *irreducible* if it is indecomposable with $\mathcal{X}$ as recurrence class.

We suppose the following.

HYPOTHESIS 2.1.
(i) *The maps $x \to \hat{f}_i(x), i = 1, \dots, m$, are locally Lipschitz and bounded.*
(ii) *The map $x \to K(x)$ is locally Lipschitz.*
(iii) *For each $x \in \mathbb{R}^d$ the chain Θ^x is indecomposable.*

It is well known (see, e.g., [**22**]) that condition 2.1(iii) implies the existence of a unique invariant probability measure $\pi(x) = (\pi_1(x), \dots, \pi_m(x))$ for Θ^x defined by

$$\sum_{i=1}^{m} \pi_i(x) K_{i,j}(x) = \pi_j(x), \quad j \in \mathcal{X}. \tag{5}$$

Define the *average vector field*:

$$F : \mathbb{R}^d \to \mathbb{R}^d$$
$$F(x) = \sum_{i=1}^{m} \pi_i(x) \hat{f}_i(x).$$

LEMMA 2.2. *The average vector field is locally Lipschitz and bounded (i.e.* $\sup_x \|F(x)\| < \infty$*).*

https://doi.org/10.1017/S0143385798097557 Published online by Cambridge University Press

Proof. Since Θ^x is indecomposable, 1 is a simple characteristic root of $K(x)$ (see e.g., [**22**]). Therefore, a straightforward application of the implicit function theorem shows that $\pi(x)$ depends smoothly on the coefficients $K_{i,j}(x), i, j \in \mathcal{X}$, and the result follows from the definition of F and assumptions 2.1(i) and 2.1(ii). □

The preceding lemma implies that F is *completely integrable*. This means that F generates a *flow*

$$\Phi : \mathbb{R} \times \mathbb{R}^d \mapsto \mathbb{R}^d$$
$$(t, x) \to \Phi_t(x)$$

defined by

$$\Phi_0 = \text{Identity}, \quad \frac{d\Phi_t(x)}{dt} = F(\Phi_t(x)).$$

The vector field F is called *dissipative* if there is a ball $B \subset \mathbb{R}^d$ (of finite radius) with the property that for every compact $K \subset \mathbb{R}^d$ there exists $T > 0$ such that $\Phi_t(K) \subset B$ for all $t \geq T$. Dissipation implies the existence of a compact invariant set $X \subset \mathbb{R}^d$ called a *global attractor*, which uniformly attracts each compact set of initial values.

HYPOTHESIS 2.3. *The average vector field $F : \mathbb{R}^d \to \mathbb{R}^d$ is dissipative.*

Notice that, with this hypothesis, we can always suppose (by multiplying F by a smooth positive function which goes to zero as $\|x\| \to \infty$) that the flow induced by F is defined on the d-sphere $M = S^d = \mathbb{R}^d \cup \{\infty\}$, where the point at infinity is a source.

We conclude this section with a few standard definitions of dynamical systems that will be useful throughout. A set $\Gamma \subset \mathbb{R}^d$ is said to be *invariant* (respectively *positively invariant*) under the flow Φ if for all $t \in \mathbb{R}$, $\Phi_t(\Gamma) \subset \Gamma$ (respectively, for all $t \geq 0$). In this case we let $\Phi|\Gamma$ denote the restricted flow (respectively semi-flow).

For $x \in \mathbb{R}^d$, $\gamma(x) = \{\Phi_t(x) : t \in \mathbb{R}\}, \gamma^+(x) = \{\Phi_t(x) : t \in \mathbb{R}_+\}$ and $\gamma^-(x) = \{\Phi_t(x) : t \in \mathbb{R}^-\}$ respectively denote the *orbit*, the *forward orbit* and the *backward orbit* of x.

The *omega limit set* of $x \in \mathbb{R}^d$, denoted by $\omega(x)$, is the set of $p \in \mathbb{R}^d$ such that $\lim_{k\to\infty} \Phi_{t_k}(x) = p$ for some sequence $t_k > 0$ with $\lim_{k\to\infty} t_k = +\infty$. Since by Hypothesis 2.3 the forward trajectory $\gamma^+(x)$ has compact closure, $\omega(x)$ is a nonempty compact invariant connected set. The *alpha limit set* $\alpha(x)$ of x is defined as the omega limit set of x for the reversed flow $\{\Phi_{-t}\}_{t\geq 0}$. $\alpha(x)$ is either a nonempty compact invariant connected subset of $\mathbb{R}^d$ or the point ∞. If $K \subset \mathbb{R}^d$ is a compact invariant set for Φ we let $L(\Phi|K)$ denote the closure of all alpha and omega limit sets of trajectories in K. That is,

$$L(\Phi|K) = \text{clos}\left(\bigcup_{x\in K} \omega(x) \cup \alpha(x)\right).$$

The *Birkhoff center* of Φ, denoted $\mathcal{B}(\Phi)$, is the closure of the set of $x \in \mathbb{R}^d$ such that $x \in \omega(x)$. This is a compact invariant set which contains equilibria and periodic orbits.

An *invariant measure* for Φ is a Borel probability measure μ on $\mathbb{R}^d$ such that $\mu(\Phi_t(A)) = \mu(A)$ for every $t \in \mathbb{R}$ and every Borel set $A \subset \mathbb{R}^d$. By the Poincaré recurrence theorem, the support of any invariant measure is contained in $\mathcal{B}(\Phi)$.

Chain recurrence. A deterministic notion of recurrence for Φ well suited to analyse the behavior of (1) is Conley's *chain recurrence*, which we now briefly introduce. A more detailed discussion of this concept will be given later (see §5).

Consider a flow Φ on a compact metric space (X, d). (Here we can think of X as the global attractor of F or as the d-sphere obtained by putting a source at the infinity.) Let T and δ be positive numbers. A (δ, T)*-pseudo-orbit* from $p \in X$ to $q \in X$ is a finite sequence of partial trajectories

$$\{\Phi_t(x_i) : 0 \leq t \leq t_i\}; \quad i = 1, \dots, k-1; t_i \geq T$$

such that

$$\begin{aligned} d(x_1, p) &< \delta, \\ d(\Phi_{t_j}(x_j), x_{j+1}) &< \delta, j = 1, \dots, k-1; \\ x_k &= q. \end{aligned}$$

In this case we write $p \hookrightarrow_{T,\delta} q$. When this holds for all $T > 0, \delta > 0$, we write $p \hookrightarrow q$. If $A \subset X$ and $B \subset X$ are two subsets such that $p \hookrightarrow q$ for all $p \in A, q \in B$ we write $A \hookrightarrow B$.

A point $x \in X$ is *chain recurrent* if $x \hookrightarrow x$. We let $\mathcal{R}(\Phi)$ denote the set of chain recurrent points for Φ. This is a closed invariant set which contains equilibria, periodic orbits, alpha and omega limit points and, more generally, nonwandering points of Φ.

A subset $L \subset X$ is *internally chain recurrent* provided L is a nonempty compact invariant set of which every point is chain recurrent for the restricted flow $\Phi|L$ (i.e. $\mathcal{R}(\Phi|L) = L$).

For example, Conley [**10**] shows that $\mathcal{R}(\Phi)$ is internally chain recurrent and also that alpha or omega limit sets are connected internally chain recurrent sets. Under some suitable conditions, limit sets of solutions to stochastic approximation processes with decreasing gain [**3**] or, more generally, limit sets of *asymptotic pseudo-trajectories* [**6**] satisfy the same property.

For $(x, y) \in \mathcal{R}(\Phi)$ x is said to be equivalent to y (written $x \sim y$) if $x \hookrightarrow y$ and $y \hookrightarrow x$. It is clear from the definition of $\mathcal{R}(\Phi)$ that $\sim$ is an equivalence relation. An equivalence class is called a *chain transitive* component. By a result of Conley [**10**], chain transitive components are connected components of $\mathcal{R}(\Phi)$.

Chaining number. A nonempty set $L \subset X$ is called *chain transitive* if it is contained in a chain transitive component of Φ. It is called *internally chain transitive* if it is compact invariant and chain transitive for the restricted flow $\Phi|L$. This is equivalent to saying that L is connected and internally chain recurrent. For instance, connected components of $\mathcal{R}(\Phi)$ are internally chain transitive sets.

Let $L \subset X$ be a chain transitive set. Let s, δ and T be positive numbers. For $p \in L, q \in L$ define

$$J_\delta^s(p, q, T) = \inf \sum_i d(\Phi_{t_i}(y_i), y_{i+1})^s,$$

where the infimum is taken over all (δ, T) pseudo-orbits going from p to q. Since L is chain transitive, $J_\delta^s(p, q, T) < \infty$.

https://doi.org/10.1017/S0143385798097557 Published online by Cambridge University Press

Let

$$J^s_\delta(L) = \sup_{(p,q)\in L; T>0} J^s_\delta(p,q,T).$$

It is easily seen from the definition that $J^s_\delta(L)$ increases when δ decreases. Therefore, the limit

$$J^s(L) = \lim_{\delta\to 0} J^s_\delta(L) = \sup_{\delta>0} J^s_\delta(L)$$

exists in $\mathbb{R}_+ \cup \{\infty\}$. Let $t > s > 0$. If $\{\{\Phi_t(y_i) : 0 \le t \le t_i\}; i = 0, \ldots, k-1\}$ is a (δ, T) pseudo-orbit from p to q we have

$$\sum_i d(\Phi_{t_i}(y_i), y_{i+1})^s \ge \delta^{s-t} \sum_i d(\Phi_{t_i}(y_i), y_{i+1})^t \ge \delta^{s-t} J^t_\delta(p,q,T).$$

Therefore,

$$J^s_\delta(p,q,T) \ge \delta^{s-t} J^t_\delta(p,q,T)$$

and consequently,

$$J^s_\delta(L) \ge \delta^{s-t} J^t_\delta(L).$$

This last relation justifies the following definition.

Definition 2.4. The *chaining number* of Φ in L is the number $J(\Phi, L)$ uniquely defined as

$$J(\Phi, L) = \inf\{s > 0 : J^s(L) = 0\} = \sup\{s > 0 : J^s(L) = \infty\}.$$

We now mention some elementary properties of $J(\Phi, L)$ whose proofs are left to the reader.

(i) Two flows Φ and Ψ defined on metric spaces X and Y are *topologically equivalent* if there exists an homeomorphism $h : X \to Y$ which take orbits of Φ to orbits of Ψ preserving time orientation.
If L is chain transitive for Φ, $h(L)$ is chain transitive for Ψ. If, furthermore, h is Lipschitz continuous, it is easy to verify that $J(\Phi, L) = J(\Psi, h(L))$.

(ii) A consequence of (i) is that $J(\Phi, L)$ is unchanged if the metric d on X is replaced by a metric d' equivalent to d (i.e. $C_1 d \le d' \le C_2 d$ for some constants $0 < C_1 \le C_2$).

(iii) Let $L \subset X$ be a compact invariant set (not necessarily chain recurrent). Let $\{B_{i,\delta}\}_{i\in I}$ be a finite (i.e. card$(I) < \infty$) family of balls of radius $\delta > 0$ such that $\{\bigcup_{t\ge 0} \Phi_t(B_{i,\delta})\}_{i\in I}$ covers L and let $N(\delta, L)$ be the minimum number of such balls. Define

$$\mathrm{Dim}_\Phi(L) = \liminf_{\delta\to 0} \frac{\log(N(\delta, L))}{\log(1/\delta)}.$$

If $\Phi = \mathrm{Id}$ is the identity flow, then

$$\mathrm{Dim}_{\mathrm{Id}}(L) = \mathrm{Dim}_{\mathrm{b}}(L),$$

where $\mathrm{Dim}_{\mathrm{b}}(L)$ denotes the *box dimension* of L. The following inequality gives some rough bounds on $J(\Phi, L)$ when L is *internally* chain transitive:

$$J(\Phi, L) \le J(\Phi|L, L) \le \mathrm{Dim}_\Phi(L) \le \mathrm{Dim}_{\mathrm{b}}(L) \le \mathrm{Dim}_{\mathrm{b}}(X).$$

We remark that when X is a subset of a d-dimensional manifold, then $\mathrm{Dim}_{\mathrm{b}}(X) \le d$.

3. *Main results*

An averaging lemma. In order to analyse the behavior of $\{X_n^\epsilon\}_{n\geq 0}$ in terms of the behavior of the flow Φ it is convenient to introduce the interpolated process $x^\epsilon : \mathbb{R}_+ \to \mathbb{R}^d$ defined by:

(i) $x^\epsilon(n\epsilon) = X_n^\epsilon$;

(ii) x^ϵ is affine on $[n\epsilon, (n+1)\epsilon]$.

The following proposition is a standard averaging result for stochastic approximation processes. It will be deduced from a more general theorem proved in [**7**].

PROPOSITION 3.1. *Let $K \subset \mathbb{R}^d$ be a compact set, $T \geq 0$ and $\delta > 0$. Then*

$$\lim_{\epsilon\to 0} \mathbf{P}_{x,\theta}\left(\sup_{0\leq t\leq T} \|x^\epsilon(t) - \Phi_t(x)\| \geq \delta\right) = 0$$

uniformly in $x \in K$ and $\theta \in \mathcal{X}$.

Proof. Let Π_x denote the transition probability defined on $\mathbb{R}^d \times \mathcal{X}$ by

$$\Pi_x((f, i), A \times \{j\}) = K_{i,j}(x)\mu_x^j(A) \tag{6}$$

for every Borel set $A \subset \mathbb{R}^d$, $i, j \in \mathcal{X}$, $f \in \mathbb{R}^d$. Let $\mathcal{F}_n$ denote the σ-field generated by $(\Theta_0, X_0), \ldots, (\Theta_n, X_n)$. Equation (1) can be rewritten as

$$X_{n+1}^\epsilon - X_n^\epsilon = \epsilon f_{n+1},$$

where

$$\mathbf{P}(\Theta_{n+1} = j, f_{n+1} \in A|\mathcal{F}_n) = \Pi_{X_n^\epsilon}((f_n, \Theta_n,); \{j\} \times A).$$

We are exactly in the situation considered in [**7**, Part II, chapter I] and we will deduce Proposition 3.1 from their averaging theorem (Theorem 9, p. 232). This theorem is proved under five assumptions (assumptions A1–A5, chapter 1). Assumptions A1–A3 and A5 are easy consequences of the boundness Hypothesis 2.1(i). The only nontrivial assumption (A4) is the existence of a function $W_x(\cdot)$ defined on $\mathbb{R}^d \times \mathcal{X}$ taking values in $\mathbb{R}^d$ sufficiently regular in x such that

$$W_x(f, i) - \Pi_x.W_x(f, i) = f - F(x), \tag{7}$$

where

$$\Pi_x.W_x(i, f) = \int_{\mathcal{X}\times\mathbb{R}^d} \Pi_x((i, f), dv)W_x(v).$$

Let $\mathbb{R}^{\mathcal{X}}$ denote the vector space of functions $h : \mathcal{X} \to \mathbb{R}$, $i \to h_i$. It is useful to consider $\mathbb{R}^{\mathcal{X}}$ as an Euclidean vector space with the norm $\|\cdot\|_{\pi(x)}$ defined by

$$\|h\|_{\pi(x)}^2 = \sum_i h_i^2 \pi_i(x). \tag{8}$$

We see $K(x)$ as an operator on $\mathbb{R}^{\mathcal{X}}$ defined by

$$(K(x)\cdot h)_i = \sum_{j\in\mathcal{X}} K_{i,j}(x)h_j.$$

Since $\pi(x)$ is an invariant measure (equation (5)) the vector space

$$\mathbf{1}^{\perp} = \left\{ h \in \mathbb{R}^{\mathcal{X}} : \sum_{i \in \mathcal{X}} \pi_i(x) h_i = 0 \right\}$$

is invariant by $K(x)$ and because Θ^x is indecomposable $\mathrm{Id} - K(x)$ induces a linear isomorphism on $\mathbf{1}^{\perp}$:

$$\begin{aligned} L(x) &: \mathbf{1}^{\perp} \to \mathbf{1}^{\perp}, \\ h &\to h - K(x)h. \end{aligned} \tag{9}$$

Therefore it is possible to define a map

$$\begin{aligned} \hat{g}(x) &: \mathcal{X} \to \mathbb{R}^d, \\ i &\to \hat{g}_i(x) \end{aligned}$$

uniquely given by

$$\langle \hat{g}(x), \alpha \rangle = L(x)^{-1} \langle (\hat{f}(x) - F(x)), \alpha \rangle \tag{10}$$

for all $\alpha \in \mathbb{R}^d$, where

$$\begin{aligned} \langle \hat{g}(x), \alpha \rangle &: \mathcal{X} \to \mathbb{R}, \\ i &\to (\langle \hat{g}(x), \alpha \rangle)_i = \langle \hat{g}_i(x), \alpha \rangle \end{aligned} \tag{11}$$

and

$$\begin{aligned} \langle (\hat{f}(x) - F(x)), \alpha \rangle &: \mathcal{X} \to \mathbb{R}, \\ i &\to (\langle (\hat{f}(x) - F(x)), \alpha \rangle)_i = \langle (\hat{f}_i(x) - F(x)), \alpha \rangle. \end{aligned} \tag{12}$$

Now set

$$W_x(i, f) = f - \hat{f}_i(x) + \hat{g}_i(x).$$

It is easy to check that:
(i) W_x satisfies (7);
(ii) W_x is Lipschitz in x uniformly in (f, i).
Therefore the assumptions of [**7**, Theorem 9, p. 232]) are fulfilled and the proposition follows from their general result. □

Recall that a sequence $\{\mu_\epsilon\}_{\epsilon > 0}$ of Borel probability measures on $\mathbb{R}^d$ is said to converge *weakly* toward a probability measure μ, as $\epsilon \to 0$, if

$$\lim_{\epsilon \to 0} \int_{\mathbb{R}^d} g(x) \mu_\epsilon(dx) = \int_{\mathbb{R}^d} g(x) \mu(dx)$$

for every bounded continuous function $g : \mathbb{R}^d \to \mathbb{R}$. We now derive the following consequence from Proposition 3.1.

COROLLARY 3.2. *Let ν^ϵ be an invariant probability measure of the Markov process $(X_n^\epsilon, \Theta_n^\epsilon)$ and let $\nu^{1,\epsilon}$ denote the marginal probability measure defined by $\nu^{1,\epsilon}(A) = \nu^\epsilon(A \times \mathcal{X})$. Suppose that the family $\{\nu^{1,\epsilon}\}_{\epsilon > 0}$ is tight. Then:*

https://doi.org/10.1017/S0143385798097557 Published online by Cambridge University Press

(i) *Any limit point of* $\{\nu^{1,\epsilon}\}_{\epsilon>0}$ *for the topology of weak convergence is an invariant measure of* Φ.

(ii) *Let* ν^1 *be a limit point of* $\{\nu^{1,\epsilon}\}_{\epsilon>0}$ *and let* $\text{supp}(\nu^1)$ *denote the support of* ν^1. *Then* $\text{supp}(\nu^1)$ *is a compact invariant set which verifies*

$$\text{supp}(\nu^1) = \mathcal{B}(\Phi|\,\text{supp}(\nu^1)) = \mathcal{R}(\Phi|\,\text{supp}(\nu^1)).$$

(iii) *Let* $K \subset \mathbb{R}^d$ *be a compact set disjoint from* $\mathcal{B}(\Phi)$. *Then,*

$$\lim_{\epsilon\to 0} \nu^{1,\epsilon}(K) = 0.$$

Proof. (i) Let $g : \mathbb{R}^d \to \mathbb{R}$ be a continuous bounded function. Fix $T > 0$ and let $T(\epsilon) = \epsilon[T/\epsilon]$, where [ ] denotes the integer part. By invariance,

$$\int_{\mathbb{R}^d\times\mathcal{X}} \mathbf{E}_{x,\theta}[g(x^\epsilon(T(\epsilon)))]\nu^\epsilon(dx\otimes d\theta) = \int_{\mathbb{R}^d} g(x)\nu^{1,\epsilon}(dx). \tag{13}$$

Let $\delta > 0$. By the tightness of $\{\nu^{1,\epsilon}\}_{\epsilon>0}$ there exists a compact set $K \subset \mathbb{R}^d$ such that $\nu^{1,\epsilon}(K) \geq 1 - \delta/\|g\|$ for all $\epsilon > 0$. Thus,

$$\left|\int_{K\times\mathcal{X}} \mathbf{E}_{x,\theta}[g(x^\epsilon(T(\epsilon)))]\nu^\epsilon(dx\otimes d\theta) - \int_{\mathbb{R}^d\times\mathcal{X}} \mathbf{E}_{x,\theta}[g(x^\epsilon(T(\epsilon)))]\nu^\epsilon(dx\otimes d\theta)\right| \leq \delta \tag{14}$$

uniformly in $\epsilon > 0$.

Let ν^1 be a limit point of $\{\nu^{1,\epsilon}\}_{\epsilon>0}$ for the weak topology. By uniform continuity of g on a compact neighborhood of $\Phi_T(K)$ there exists $\alpha > 0$ such that $d(u, \Phi_T(x)) < \alpha/2$ implies $d(g(u), g(\Phi_T(x))) < \delta$. Thus (for ϵ small enough),

$$|\mathbf{E}_{x,\theta}[g(x^\epsilon(T(\epsilon))) - g(\Phi_T(x))]| \leq \delta + 2\|g\|\mathbf{P}_{x,\theta}\{d(x^\epsilon(T(\epsilon)), \Phi_{T(\epsilon)}(x)) \geq \alpha\} \tag{15}$$

for all $x \in K$. Using Proposition 3.1 together with inequalities (14) and (15) gives

$$\left|\int_{\mathbb{R}^d} g(\Phi_t(x))\nu^1(dx) - \int_{\mathbb{R}^d} g(x)\nu^1(dx)\right| \leq 2\delta.$$

Since δ and g are arbitrary, this shows that ν^1 is an invariant measure of the flow Φ.

(ii) By (i) ν^1 is an invariant measure of Φ. Let H denote its support. H is a closed invariant set. Clearly, the measure μ defined by $\mu(A) = \nu^1(A\cap H)$ for every Borel set $A \subset \mathbb{R}^d$ is an invariant probability measure for Φ. Thus, by the Poincaré recurrence theorem $\mu(x : x \notin \omega(x)) = 0$. This implies that $\nu^1(\mathcal{B}(\Phi|H)) = 1$. It follows that $H \subset \mathcal{B}(\Phi|H) \subset \mathcal{R}(\Phi|H) \subset H$.

(iii) Since by tightness, $\{\nu^{1,\epsilon}\}_{\epsilon>0}$ is relatively compact, it suffices to show that $\nu^1(K) = 0$ for every limit point ν^1 of $\{\nu^{1,\epsilon}\}_{\epsilon>0}$. Let $g : \mathbb{R}^d \to [0, 1]$ be a smooth function which is one on a neighborhood of K and zero on a neighborhood of $\mathcal{B}(\Phi)$. We have

$$\nu^{1,\epsilon}(K) \leq \int_{\mathbb{R}^d} g(x)\nu^{1,\epsilon}(dx).$$

Thus $\limsup_{\epsilon\to 0} \nu^{1,\epsilon}(K) \leq \int_{\mathbb{R}^d} g(x)\nu^1(dx) = 0$, where the right-hand equality follows from Lemma 3.1 and the Poincaré recurrence theorem. □

COROLLARY 3.3. *Consider the urn process of Example 1.1, where we use the same notation. Assume:*
(i) *the urn functions p^+ and p^- are Lipschitz continuous;*
(ii) $p^+(\Delta^d) \subset \mathrm{int}(\Delta^d)$, $p^-(\mathrm{int}(\Delta^d)) \subset \mathrm{int}(\Delta^d)$.
Let $F : E^d \to E^d$ be the vector field defined by

$$F(x) = p^+(r(x)) - p^-(r(x)). \tag{16}$$

Then:
(a) *F is dissipative with a global attractor $X \subset \mathrm{int}(\Delta^d)$.*
(b) *Let $K \subset \Delta^d$ be a compact subset disjoint from $\mathcal{B}(\Phi)$, where Φ is the flow induced by the vector field (16). Then*

$$\lim_{\epsilon\to 0}\lim_{n\to\infty} \mathbf{P}(X_n^\epsilon \in K) = 0.$$

Proof. First observe that Hypothesis 2.1 is obviously satisfied and the Markov chain Θ^x is indecomposable with recurrence class

$$\mathcal{X}' = \{(i, j) \in \mathcal{X} : r_i(x) > 0\}.$$

(a) A simple computation shows that the average vector field associated to the process is given by (16). Let $V : E^d \to \mathbb{R}_+$ be the map defined by $V(x) = d(x, \Delta^d) = \|x - r(x)\|$. By the convexity of Δ^d, we get

$$V((1-t)x + t(x + F(x))) \le (1-t)V(x) + tV(x + F(x)).$$

Thus

$$\frac{V(x + tF(x)) - V(x)}{t} \le V(x + F(x)) - V(x).$$

Now as the map V is convex and continuous, it admits a right derivative. Therefore, letting t go to zero in the last inequality gives

$$\left.\frac{dV(\Phi_t(x))}{dt}\right|_{t=0} \le V(x + F(x)) - V(x).$$

Let $x \notin \Delta^d$. Since $r(x) + F(x) \in \mathrm{int}(\Delta^d)$, $V(x + F(x)) = d(x + F(x), \Delta^d) < d(x + F(x), r(x) + F(x)) = \|x - r(x)\| = d(x, \Delta^d)$. Thus

$$\left.\frac{dV(\Phi_t(x))}{dt}\right|_{t=0} < 0$$

for $x \notin \Delta^d$ and

$$\left.\frac{dV(\Phi_t(x))}{dt}\right|_{t=0} = 0$$

for $x \in \Delta^d$. This implies that Δ^d contains a global attractor $X \subset \Delta^d$. Since, furthermore, $F(x)$ points inward Δ^d whenever $x \in \partial\Delta^d$, X must be contained in $\mathrm{int}(\Delta^d)$.

(b) For each value of $\epsilon = 1/N$ the process $\{X_n^\epsilon\}_{n\in\mathbb{N}}$ is a finite Markov chain on the lattice $\{v = (v_1, \dots, v_{d+1}) \in \Delta^d : Nv_i \in \mathbb{N}, i = 1, \dots, d+1\}$. Assumption (ii) implies that this chain is irreducible and aperiodic. Therefore it admits a unique invariant measure μ^ϵ and $\lim_{n\to\infty} \mathbf{P}(X_n^\epsilon \in K) = \mu^\epsilon(K)$. The corollary now follows from Corollary 3.2(ii). □

https://doi.org/10.1017/S0143385798097557 Published online by Cambridge University Press

Main results. Let $X \subset \mathbb{R}^d$ denote the global attractor of Φ (see Hypothesis 2.3). In this section we make the following additional assumption.

HYPOTHESIS 3.4. *There exists an open ball B_R which contains X such that for all $x \in B_R$ the following properties hold:*

(i) *the Markov chain Θ^x is irreducible and aperiodic;*

(ii) *there exist m probability measures $\mu^1, \dots, \mu^m$ with bounded support, and m density functions $p_x^1(f), \dots, p_x^m(f)$ such that*

$$\mu_x^i(df) = p_x^i(f)\mu^i(df), \quad i = 1, \dots, m;$$

(iii) *the functions $p_x^i(f)$ are locally Lipschitz in $x \in B_R$ and uniformly in $f \in \mathbb{R}^d$;*

(iv) *for every compact set $K \subset B_R$ there are numbers $0 < a \leq A < \infty$ such that $a \leq p_x^i(f) \leq A$ for all $x \in K$, $f \in \mathbb{R}^d$.*

The main result of this paper (Theorem 3.7 below) shows how Corollary 3.2 can be notably sharpened provided that the process (1) satisfies a certain nondegeneracy condition that we now introduce.

Introduce the $m+1$ quadratic forms

$$Q_x^i : \mathbb{R}^d \to \mathbb{R}_+, \quad i = 0, \dots, m$$

defined by

$$Q_x^0(\alpha) = \|\langle \hat{g}(x), \alpha\rangle\|^2_{\pi(x)} - \|\langle K(x)\hat{g}(x), \alpha\rangle\|^2_{\pi(x)}$$

and

$$Q_x^i(\alpha) = \int_{\mathbb{R}^d} \langle (f - \hat{f}_i(x)), \alpha\rangle^2 \mu_x^i(df), \quad i = 1, \dots, m,$$

where $\hat{g}(x)$ and $\langle \hat{g}(x), \alpha\rangle$ are defined by (10) and (11), and $\|\cdot\|_{\pi(x)}$ is the norm on $\mathbb{R}^{\mathcal{X}}$ given by (8).

Observe that the quadratic forms Q_x^i are nonnegative. This is obvious for $i \geq 1$ and for Q_x^0 this is an easy consequence of Jensen's inequality.

Definition 3.5. Let

$$Q_x = Q_x^0 + \sum_{i \in \mathcal{X}} \pi_i(x) Q_x^i. \tag{17}$$

We will say that the process (1) is *nondegenerate* at $x \in X$ if Q_x is nondegenerate, that is $Q_x(\alpha) = 0$ if and only if $\alpha = 0$.

Some simple sufficient conditions ensuring that a point x is nondegenerate will be given below (see Remark 3.10).

Since the Birkhoff center of Φ may contain 'unstable' sets, it is natural to think that invariant measures of (1) tend to concentrate on 'stable' subsets of $\mathcal{B}(\Phi)$ provided that unstable sets consist of nondegenerate points. We now make this point precise.

A nonempty compact invariant set $A \subset \mathbb{R}^d$ is an *attractor* if A has an open neighborhood W in $\mathbb{R}^d$ such that

$$\lim_{t \to \infty} d(\Phi_t(x), A) = 0$$

uniformly in $x \in W$. An attractor is said to be *irreducible* if it has no proper attractor.

For $x \in \mathcal{R}(\Phi)$, let C_x denote the connected component (or, equivalently, the chain transitive component) of $\mathcal{R}(\Phi)$ which contains x. The relation $\hookrightarrow$ induces on $\mathcal{R}(\Phi)/\sim$ a partial ordering (written $\succ$) defined as follows: $C_x \succ C_y$ if $x \hookrightarrow y$. Following Ruelle [**29**], a minimal component for $\succ$ is called a *quasi-attractor*.

Let C be a connected component of $\mathcal{R}(\Phi)$. Define

$$C^+ = \{x \in \mathbb{R}^d : \exists y \in \mathcal{R}(\Phi) \setminus C \text{ and } C \hookrightarrow y \hookrightarrow x\}.$$

An invariant set Γ is said to be *isolated* if it is the maximal invariant set in some neighborhood N of itself. The following proposition summarizes some properties of irreducible attractors in relation with chain recurrence. It will be proved in §5.

PROPOSITION 3.6.
(i) *Let C be a nonempty subset of X. The following statements are equivalent:*
  (a) *C is an irreducible attractor;*
  (b) *C is an isolated quasi-attractor;*
  (c) *C is a connected component of $\mathcal{R}(\Phi)$ isolated and $C^+ = \emptyset$.*
(ii) *Assume C is an isolated connected component of $\mathcal{R}(\Phi)$ which is not an attractor. Then C^+ is a nonempty compact invariant set. If, furthermore, C^+ is isolated, then C^+ is an attractor.*

We now state the main result of the paper.

THEOREM 3.7. *Let C be a connected component of $\mathcal{R}(\Phi)$ which is not an attractor. Suppose that:*
(i) *C and C^+ are isolated;*
(ii) *$J(\Phi, L(\Phi|C)) < 2$;*
(iii) *for every $x \in L(\Phi|C)$ the process (1) is nondegenerate at x.*
Let ν^ϵ be an invariant probability measure of the Markov process $\{X_n^\epsilon, \Theta_n^\epsilon\}$. Then there exists a neighborhood V of C such that

$$\lim_{\epsilon \to 0} \nu^\epsilon(V \times \mathcal{X}) = 0.$$

COROLLARY 3.8. *Suppose that $\mathcal{R}(\Phi)$ admits a finite number of connected components and that for each component which is not an attractor, assumptions* (ii) *and* (iii) *of Theorem 3.7 hold. Suppose that the family $\{\nu^{1,\epsilon} = \nu^\epsilon(\cdot \times \mathcal{X})\}_{\epsilon>0}$ is tight. Let ν^1 be a limit point (for the topology of weak convergence) of $\{\nu^{1,\epsilon}\}_{\epsilon>0}$. Then, the support of ν^1 is contained in the Birkhoff center of irreducible attractors.*

Proof. It suffices to verify that C^+ is isolated when C is not an attractor. □

Remark 3.9. Suppose that the limit point set decomposes as

$$L(\Phi) = \bigcup_{i=1}^{k} \Gamma_i,$$

where $\Gamma_1, \ldots, \Gamma_k$ are compact invariant disjoint internally chain transitive sets such that $J(\Phi, \Phi|\Gamma_i) < 2$. Then it follows easily from Theorem 6.7 that $J(\Phi, L(\Phi|C)) < 2$.

Remark 3.10. The nondegeneracy condition on x depends on the measures μ_x^i and the kernel $K(x)$. In many cases it is easy to verify. Here are several such cases.

(i) Let

$$\mathrm{Var}_x^i = \int_{\mathbb{R}^d} (f - \hat{f}_i(x))(f - \hat{f}_i(x))^{\mathrm{T}} \mu_x^i(df)$$

denote the *covariance matrix* of μ_x^i, where $^{\mathrm{T}}$ denotes the transpose of a matrix. Suppose that for some $i \in \mathcal{X}$, Var_x^i is nondegenerate (i.e. invertible) and $\Pi_i(x) > 0$. Then Q_x^i, hence Q_x is definite positive.

(ii) Let

$$\hat{g}(x)_\alpha : \mathcal{X} \to \mathbb{R}^d,$$
$$i \to \langle \hat{g}_i(x), \alpha \rangle,$$

where $\hat{g}(x)$ is given by (10). Another condition is obtained by noticing that

$$Q_x^0(\alpha) = \langle (\mathrm{Id} - K^*(x)K(x))\hat{g}(x)_\alpha, \hat{g}(x)_\alpha \rangle_{\pi(x)},$$

where $\langle\, , \rangle_{\pi(x)}$ is the inner product on $\mathbb{R}^{\mathcal{X}}$ associated to $\|\cdot\|_{\pi(x)}$ and $K^*(x)$ denotes the adjoint of $K(x)$ for the Euclidean structure induced by $\|\cdot\|_{\pi(x)}$.

Therefore, if the Markov chain associated to the kernel $K^*(x)K(x)$ is irreducible we get the estimate

$$Q_x^0(\alpha) \geq (1 - \lambda)\|\hat{g}(x)_\alpha\|^2_{\pi(x)}, \tag{18}$$

where $0 \leq 1 - \lambda < 1$ is the second largest eigenvalue of $K^*(x)K(x)$. Suppose now that the vectors $\hat{f}_1(x) - F(x), \ldots, \hat{f}_m(x) - F(x)$ span $T_x\mathbb{R}^d = \mathbb{R}^d$. Then it is easily seen from (10) that $\hat{g}_1(x), \ldots, \hat{g}_m(x)$ also span $\mathbb{R}^d$ and inequality (18) implies that Q_x^0 is irreducible.

(iii) In some simple cases the irreducibility of $K^*(x)K(x)$ can be deduced from the transition graph of $K(x)$ and the fact that $K(x)$ is irreducible and aperiodic. This is the case, for example, when $K_{i,i}(x) > 0$ for all $i \in \mathcal{X}$ or when $K_{i,j}(x) > 0 \Rightarrow K_{j,i}(x) > 0$ for all $(i, j) \in \mathcal{X}$. However, it is easy to construct irreducible aperiodic Markov transition matrices K for which K^*K is not irreducible. The following 3×3 matrix gives such an example:

$$K = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

COROLLARY 3.11. *Consider the urn process of Example 1.1 under assumptions* (i) *and* (ii) *of Corollary 3.3. Let C be a connected component of $\mathcal{R}(\Phi)$, where Φ is the flow induced by the vector field* (16). *If C satisfies assumptions* (i) *and* (ii) *of Theorem 3.7, then*

$$\lim_{\epsilon \to 0} \lim_{n \to \infty} \mathbf{P}(X_n^\epsilon \in C) = 0.$$

Proof. We first remark that assumptions (i) and (ii) of Corollary 3.3 imply that the Markov chain $K(x)$ given by (3) is irreducible and aperiodic for all $x \in \mathrm{int}(\Delta)$. Since the vector field (16) has a global attractor in $\mathrm{int}(\Delta)$ (see Corollary 3.3), condition (i) of Hypothesis 3.4 is satisfied. Conditions (ii)–(iv) follow from the definition of the measures (4).

Remark 3.10(iii) above and the definition of $K(x)$ (equation (3)) show that the process is nondegenerate at every point $x \in \text{int}(\Delta)$. The corollary now follows from Theorem 3.7 and the fact that $\{X_n^\epsilon\}_{n\in\mathbb{N}}$ is a finite irreducible aperiodic Markov chain as shown in the proof of Corollary 3.3. □

4. *Examples*

Axiom A systems. If the flow induced by the average vector field F is Axiom A and has no cycles [**30**] or, equivalently, if $\mathcal{R}(\Phi)$ is hyperbolic [**16**], then $\mathcal{R}(\Phi)$ decomposes as $\mathcal{R}(\Phi) = \bigcup_{i=1}^k C_i$, where the C_i are distinct compact invariant connected topologically transitive sets†. It is readily seen that $J(\Phi, C_i) = 0$. Thus Corollary 3.8 applies. In this case irreducible attractors are the C_i's for which $W^u(C_i) \subset C_i$, where $W^u(C_i)$ denotes the unstable manifold of C_i. The Birkhoff center of C_i is C_i.

A particular simple case is the case where F is a *Morse–Smale* vector field. This means that each C_i is a hyperbolic equilibrium or periodic orbit. Dynamics of urn processes with decreasing gain associated to a Morse–Smale vector field have been extensively studied by Benaïm and Hirsch [**5**]. For the urn process described in Example 1.1 we get the following result.

COROLLARY 4.1. *Consider the urn process of Example 1.1 under assumptions* (i) *and* (ii) *of Corollary 3.3. Suppose that the vector field F given by* (16) *is Morse–Smale. Let $p_1, \dots, p_r$ denote the asymptotically stable equilibria of F and let $\gamma_1, \dots, \gamma_l$ denote the asymptotically stable nonstationary periodic orbits of F. Then:*

(a)

$$\lim_{\epsilon\to 0}\lim_{n\to\infty} \mathbf{P}\bigg(X_n^\epsilon \in \bigcup_{i=1}^r \{p_i\} \bigcup_{j=1}^l \gamma_j\bigg) = 1.$$

(b) *Let ν^1 be a limit point of the (unique) invariant measure of the process $\{X_n^\epsilon\}_{n\geq 0}$. Then ν^1 is a convex combination of the measures*

$$\delta_{p_i}(\cdot), \quad i = 1, \dots, r$$

and

$$\frac{1}{T_i}\int_0^{T_i} \delta_{\Phi_t(x_i)}(\cdot)\,dt, \quad i = 1, \dots, l,$$

where T_i is the period of γ_i and x_i is any point in γ_i.

Also, it is reasonable to conjecture that the invariant measure of $\{X_n^\epsilon\}$ generically converges as $\epsilon \to 0$ toward a single invariant measure. More precisely, we have the following.

CONJECTURE 4.2. *Let $U(d) \subset C^1(\Delta^d) \times C^1(\Delta^d)$ denote the set of C^1 urn functions and let $MS(d) \subset U(d)$ denote the subset of $U(d)$ consisting of functions (p^+, p^-) such that the vector field $F = p^+ - p^-$ is Morse–Smale. Then:*

† A topologically transitive set is a set with a dense orbit.

https://doi.org/10.1017/S0143385798097557 Published online by Cambridge University Press

(a) *there exists a subset $\mathcal{U} \subset MS(d)$ open and dense for the C^1 topology such that if $(p^+, p^-) \in \mathcal{U}$ the invariant measure of $\{X_n^\epsilon\}$ converges (for topology of weak convergence) as ϵ goes to zero toward a measure ν^1 which is either a dirac measure δ_{p_i} at a stable equilibrium p_i of F or the invariant measure $(1/T_i)\int_0^{T_i} \delta_{\Phi_t(x_i)}(\cdot)\,dt$ associated to a stable periodic γ_i;*

(b) *When $d = 3$, (3 color urn processes) $\mathcal{U}$ is dense in $U(d)$.*

Simple flows. The flow Φ is called *simple* provided that there are finitely many alpha and omega limit points for Φ, necessarily constituting the set $\mathrm{Eq}(\Phi)$ of equilibria.

A set $\Gamma \subset \mathbb{R}^d$ is called an *orbit chain* if Γ can be expressed as the union

$$\Gamma = \{p_1, \dots, p_k\} \bigcup \gamma_1 \bigcup \cdots \bigcup \gamma_{k-1}$$

of equilibria $\{p_1, \dots, p_k\}$ and nonsingular orbits $\{\gamma_1, \dots, \gamma_{k-1}\}$ connecting them. That is,

$$p_i = \alpha(\gamma_i), \quad p_{i+1} = \omega(\gamma_i).$$

The orbit chain is called *cyclic* if $p_1 = p_k$.

Given two points $u, v \in \mathbb{R}^d$, we write $u \rightsquigarrow v$ if either $\Phi_t(u) = v$ for some $t \geq 0$ or if there exists an orbit chain Γ such that $u \in \mathrm{clos}(\gamma_i)$, $v \in \mathrm{clos}(\gamma_j)$, and $i < j$.

Suppose that Φ is a simple flow. Let $e \in \mathrm{Eq}(\Phi)$. We say that e is *weakly unstable* if there exists some equilibrium $f \in \mathrm{Eq}(\Phi)$ such that $e \rightsquigarrow f$ and $f \not\rightsquigarrow e$. An equilibrium which is not weakly unstable is called *weakly stable*. Observe that an equilibrium can be weakly stable without being asymptotically (or even Lyapounov) stable.

THEOREM 4.3. *Assume:*

(i) *Φ is a simple flow;*

(ii) *the process (1) is nondegenerate at every weakly unstable equilibrium;*

(iii) *$\{\nu^{1,\epsilon}\}_{\epsilon>0}$ is tight.*

Let $\mathrm{Eq}^s(\Phi) \subset \mathrm{Eq}(\Phi)$ denote the set of weakly stable equilibria for Φ. Then

$$\lim_{\epsilon\to\infty} \nu^{1,\epsilon}(\mathrm{Eq}^s(\Phi)) = 1.$$

Expressed differently: if ν^1 is a limit point of $\{\nu^{1,\epsilon}\}_{\epsilon>0}$, then the support of ν^1 is contained in the set of weakly stable equilibria.

Proof. Let C be a chain transitive component of $\mathcal{R}(\Phi)$. By the compactness of C and finiteness of $\mathrm{Eq}(\Phi)$ it is possible to find a compact neighborhood N of C such that $N \cap \mathrm{Eq}(\Phi) = C \cap \mathrm{Eq}(\Phi)$. Let $x \in N$. Suppose that $\gamma(x) \subset N$. Then there exist equilibria $e = \alpha(x)$ and $f = \omega(x)$ which are in C. Since C is chain transitive it follows that $\gamma(x) \subset C$. This shows that C is isolated. Similarly, C^+ is isolated (or empty).

Suppose now that C is not an attractor. Proposition 3.6(ii) implies that $C^+ \neq \emptyset$. Let $x \in C$ and $y \in C^+$. By Theorem 3.1 of [**4**] the relation $x \hookrightarrow y$ is equivalent for a simple flow to the relation $x \rightsquigarrow y$. It follows that equilibria of C are weakly unstable and the result follows from Corollary 3.2, Theorem 3.7 and the trivial fact that the chaining number of an equilibrium is zero. □

https://doi.org/10.1017/S0143385798097557 Published online by Cambridge University Press

Remark 4.4. Theorem 4.3 admits the following obvious extension. Suppose that Φ admits finitely many alpha and omega limit sets denoted $\Gamma_1, \ldots, \Gamma_k$. Let $\widetilde{\mathbb{R}}^d$ denote the space obtaining by collapsing each Γ_i to a point and let $\widetilde{\Phi}$ denote the quotient flow. Suppose that $\widetilde{\Phi}$ is a simple flow. Call the limit set Γ_i *weakly stable* (respectively weakly unstable) if Γ_i is weakly stable (unstable) for $\widetilde{\Phi}$. If the process (1) is nondegenerate at every point x belonging to a weakly unstable limit set then the support of any limit point of $\{\nu^{1,\epsilon}\}_{\epsilon>0}$ is contained in the weakly stable limit sets.

Gradient-like systems. In applications of stochastic recursive algorithms to problems of engineering, control, learning theory and elsewhere, it is common to consider average vector fields which admit a strict Lyapounov function. A function $V : \mathbb{R}^d \to \mathbb{R}$ is called a *strict Lyapounov function* for the flow Φ if V is a continuous nonnegative function which strictly decreases along nonconstant forward orbits of Φ.

If F admits a strict Lyapounov function and isolated equilibria then it is easily seen that Φ is a simple flow which has no cyclic orbit chains and Theorem 4.3 obviously applies. In this case weakly stable equilibria are equilibria $e \in \mathrm{Eq}(\Phi)$ for which $\{x \in \mathbb{R}^d : \alpha(x) = e\} = \{e\}$.

Planar systems. In this section we suppose that $F : \mathbb{R}^2 \to \mathbb{R}^2$ is a planar vector field with isolated equilibria. We let P denote the union of all nonstationary periodic orbits of Φ, and $\mathrm{Eq}(\Phi)$ the equilibrium set.

An invariant set R for Φ is *strongly chain recurrent* if every point of R belongs to a cyclic orbit chain (as defined in the preceding subsection on simple flows) or a periodic orbit. It is *strongly chain transitive* provided every pair of points belong to a common cyclic chain or periodic orbit. If these cyclic chains and periodic orbits can be found *in* R, then R is called *internally* strongly chain recurrent or chain transitive.

THEOREM 4.5. *Assume:*
(i) *F is a planar vector field with isolated equilibria;*
(ii) *$\{\nu^1_\epsilon\}_{\epsilon>0}$ is tight.*
Let ν^1 be a limit point of $\{\nu^1_\epsilon\}_{\epsilon>0}$ and let R be a connected component of the support of ν^1. Then:
(a) *R is compact invariant and internally strongly chain recurrent;*
(b) *each component of $R \cap P$ which is not an isolated periodic orbit is homeomorphic to an annulus and each component of $R \setminus P$ is internally strongly chain transitive.*
Suppose, furthermore, that:
(iii) *P has finitely many components;*
(iv) *the process (1) is nondegenerate at every point $x \in P \cup \mathrm{Eq}(\Phi)$.*
If $x \in \mathbb{R}^2 \setminus R$ is such that $\alpha(x) \cap R \neq \emptyset$, then there exists a cyclic orbit chain Γ which contains x and has nonempty intersection with R.

Proof. Let H denote the support of ν^1. By Corollary 3.2, H is internally chain recurrent for Φ, and assertions (a) and (b) follow from Theorems 1.4 and 1.5 in [**4**] which characterize internally chain recurrent set of planar flows.

https://doi.org/10.1017/S0143385798097557 Published online by Cambridge University Press

Now suppose that hypotheses (iii) and (iv) hold. Let C be any chain transitive component of Φ. It is always possible to find a compact neighborhood N of C such that $N \cap \mathrm{Eq}(\Phi) = C \cap \mathrm{Eq}(\Phi)$ and $N \cap P = C \cap P$ because $\mathrm{Eq}(\Phi)$ is finite and P has finitely many components. Let $\gamma(x) \subset N$. By the Poincaré–Bendixson theorem, $\omega(x)$ must contain an equilibrium or a periodic orbit and similarly for $\alpha(x)$. Thus both $\alpha(x)$ and $\omega(x)$ meet C and since C is a chain transitive component it follows that $x \in C$. This shows that C is isolated.

Now let C denote the chain transitive component which contains R. By Theorem 1.4 of [**5**], C is internally strongly chain recurrent. Thus $J(\Phi, C) \leq 1$ because the chaining number of a cyclic orbit chain is zero and the chaining number of an annulus is 1. Thus, Theorem 3.7 implies that C is an attractor. Let $x \in \mathbb{R}^2$ be such that $\alpha(x) \cap R \neq \emptyset$. Let $p \in \alpha(x) \cap R$ and let $q \in \omega(x)$. One has $p \hookrightarrow x \hookrightarrow q$. Because C is an attractor C^+ is empty (see Proposition 3.6). Thus, $q \in C$ and therefore $x \in C$. Because C is internally strongly chain transitive x and p belong to a cyclic orbit chain of periodic orbit contained in C. In the second situation, the periodic orbit must lie in R by the invariance of R. □

Discussion: decreasing gain versus constant gain. Consider a process described by (2) such that $X_n^\epsilon \in \mathbb{R}^2$. Suppose that the process is nondegenerate for all $x \in \mathbb{R}^2$ and that the average vector field F has a phase portrait as shown in figure 1.

The curve '∞' is the unique irreducible attractor of F and its Birkhoff center is the equilibrium point B. Therefore, it follows from the results stated §3 that $\nu^{1,\epsilon}$ converges weakly toward δ_B as $\epsilon \to 0$.

Now consider a version of this process with a decreasing gain; that is

$$\begin{cases} \mathbf{P}(\Theta_{n+1} = j \mid X_n = x, \Theta_n = i) = K_{i,j}(x) \\ X_{n+1} - X_n = \epsilon_{n+1} f_{n+1}(X_n^\epsilon, \Theta_{n+1}^\epsilon), \end{cases}$$

where

$$\epsilon_n \geq 0, \quad \sum \epsilon_n = \infty$$

and

$$\sum \epsilon_n^{1+\delta} < \infty$$

for some $\delta > 0$. Suppose for simplicity that the sequences f_n and Θ_n are such that $\{X_n\}$ remains bounded with probability one. According to a theorem proved in [**3**] the limit set $L(X_n)$ of $\{X_n\}$ is (with probability one) a connected internally chain recurrent set (see §5) of the flow induced by F. On the other hand, under mild natural assumptions (see Pemantle [**27**] or Brandiere and Duflo [**9**]) this limit set cannot be (with positive probability) one of the points A, B or C because these points are linearly unstable equilibria. It follows that $L(X_n)$ is (with probability one) either the ∞ or one of the two handles which compose the ∞. It is interesting to notice that this behavior is quite different from the behavior of the process with fixed ϵ.

5. *Chain recurrence and attractors*

The aim of this section is to prove Proposition 3.6. For a general theory of chain recurrence we refer the reader to [**1, 8, 10, 15, 21**]. Relations with hyperbolicity are

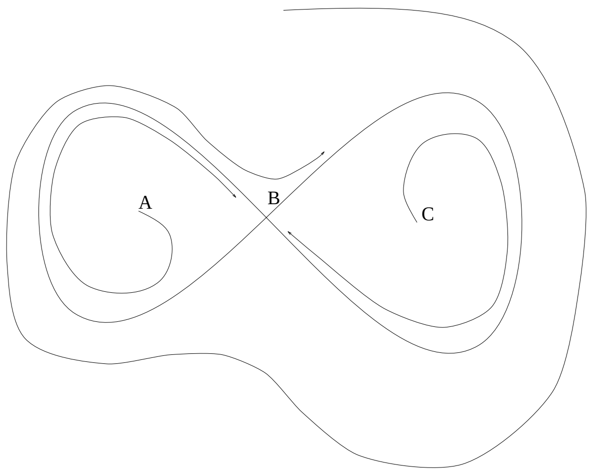

FIGURE 1.

discussed in [**8**, **16**, **28**]. Planar systems are considered in [**4**, **20**].

PROPOSITION 5.1. *Let $A \subset X$ be an attractor. A is irreducible if and only if A is a connected component of $\mathcal{R}(\Phi)$.*

Proof. Let Ψ be a flow on a compact metric space Y. It is proved in [**10**] that $\mathcal{R}(\Psi) = \cap\{A' \cup (Y \setminus B(A'))\}$ where the intersection is taken over all attractors A' in Y.

Suppose now that A is a connected component of $\mathcal{R}(\Phi)$. Set $Y = A$ and $\Psi = \Phi|A$. Since connected components of $\mathcal{R}(\Phi)$ are internally chain recurrent sets it follows that $\mathcal{R}(\Psi) = A = Y$. Thus, if $A' \subset A$ is an attractor A is contained in $A' \cup (A \setminus B(A'))$ and since A is connected, it follows that $A = A'$. The converse statement is proved by similar arguments. □

Let Γ be an isolated set. If N is a compact neighborhood of Γ in which Γ is the maximal invariant set, then N is called an *isolating neighborhood*.

LEMMA 5.2. *Let C be a connected component of $\mathcal{R}(\Phi)$. C is isolated if and only if there exists a compact neighborhood N of C such that $N \cap \mathcal{R}(\Phi) = C$.*

Proof. Assume N is a compact neighborhood of C such that $N \cap \mathcal{R}(\Phi) = C$. Let $\Lambda \subset N$ be an invariant set and let $x \in \Lambda$. We have $\alpha(x) \subset \Lambda$ and $\omega(x) \subset \Lambda$. Since $\alpha(x)$ and $\omega(x)$ are internally chain recurrent, we have $\alpha(x) \subset N \cap \mathcal{R}(\Phi) = C$ and $\omega(x) \subset N \cap \mathcal{R}(\Phi) = C$. From this it is easily seen that $C \hookrightarrow x \hookrightarrow C$. Therefore $x \in C$ proving that C is isolated.

Conversely, assume C is isolated. Let N_0 be an isolating neighborhood of C. Assume that every neighborhood N of C contains a chain recurrent point which is not in C. Then there exists a sequence $x_n \in \mathcal{R}(\Phi)$ such that $x_n \notin C$, and $d(x_n, C) \to 0$. With C being isolated, with isolating neighborhood N_0, there exists $y_n \in \gamma(x_n)$ and $y_n \notin N_0$. Since x_n

is chain recurrent, $\gamma(x_n)$ belongs to the connected component of $\mathcal{R}(\Phi)$ which contains x_n. Thus, $x_n \hookrightarrow y_n \hookrightarrow x_n$. The relation $\hookrightarrow$ being closed, we get $x \hookrightarrow y \hookrightarrow x$ for some $x \in C$ and $y \in X|N_0$. This yields to the contradiction that $y \in C \cap X|N_0$. Therefore there must be some neighborhood N such that $N \cap \mathcal{R}(\Phi) = C$. □

The proof of the next lemma is analogous to the proof of Lemma 5.2.

LEMMA 5.3. *C^+ is isolated if only if there exits a compact neighborhood N of C^+ such that $N \cap \mathcal{R}(\Phi) = C^+ \cap \mathcal{R}(\Phi)$.*

We now proceed to the proof of Proposition 3.6(i). The implications (a) $\Rightarrow$ (b) $\Rightarrow$ (c) are straightforward. To obtain (c) $\Rightarrow$ (a) we shall use the following useful lemma proved in [**10**].

LEMMA 5.4. *Let $N \subset X$ be a compact set. Let $A \subset X$ be the maximal invariant set contained in N. If A is nonempty and A is* not *an attractor, then there exists $p \in \partial N$ such that $\gamma_-(p) \subset N$ and $\alpha(p)$ is a nonempty subset of A.*

Assume (c) holds and suppose (a) is false. Let N be an isolating neighborhood of C. By Lemma 5.4 there exists $p \in \partial N$ such $\alpha(p)$ is a nonempty subset of C. Since $C^+ = \emptyset$ we must have $\omega(p) \subset C$. Thus $C \hookrightarrow p \hookrightarrow C$. This implies $p \in C$ which is contradictory.

(ii) From assertion (i) we see that C^+ is nonempty. C^+ is clearly invariant. Let $x_n \in C^+$ be a convergent sequence and $x = \lim_{n\to\infty} y_n$. To prove the compactness of C^+ it suffices to show that $x \in C^+$. Let $y_n \in \mathcal{R}(\Phi) \setminus C$ such that $C \hookrightarrow y_n \hookrightarrow x_n$. Let y be a limit point of $\{y_n\}$. Since the chain recurrent set and the relation $\hookrightarrow$ are closed, $C \hookrightarrow y \hookrightarrow x$ and $y \in \mathcal{R}(\Phi)$. On the other hand, Lemma 5.2 shows that the sequence y_n can be chosen outside an open neighborhood of C. Therefore, y is not in C. This proves that $y \in C^+$ and the compactness of C^+ follows.

Suppose now that C^+ is isolated. Let N be an isolating neighborhood of C^+. By Lemma 5.4, if C^+ is not an attractor then there exists $p \in \partial N$ with $\alpha(p) \subset C^+$. Thus $C \hookrightarrow C^+ \hookrightarrow p \hookrightarrow \omega(p)$. Since $\omega(p) \subset \mathcal{R}(\Phi)$ this implies $\omega(p) \subset C^+$. Thus $p \in C^+$ which is contradictory. □

6. *Large deviation properties, separating sets and invariant decompositions*

The first part of this section presents a large deviation principle for system (1) which follows mainly from Dupuis [**13**]. For an introduction to the theory of large deviations for stochastic approximation processes and for other references we also refer the reader to chapter 11 of Kushner [**26**]. The next subsection is devoted to some dynamical implications of this principle.

Large deviations properties for (1): Dupuis's theorem. Without loss of generality we will assume in this section that Hypothesis 3.4 holds for all $x \in \mathbb{R}^d$. Let $C[0, T]$ denote the set of continuous functions $h : [0, T] \to \mathbb{R}^d$ endowed with the topology of uniform convergence induced by the uniform norm: $\|h\| = \sup_{0\le t\le T} \|h(t)\|$. Let $C_x[0, T] \subset C[0, T]$ denote the subset of functions for which $h(0) = x$.

Let $\{\Theta_n^x, f_n^x\}$ denote the Markov chain on $\mathcal{X} \times \mathbb{R}^d$ whose transition kernel Π_x is given by (6). Let

$$H_n(x, \alpha, \theta, f) = \frac{1}{n} \log \mathbf{E}\Bigg[\exp\Bigg(\Bigg\langle \alpha, \sum_{i=1}^{n} f_i^x \Bigg\rangle\Bigg) \mid \Theta_0^x = \theta, f_0^x = f \Bigg].$$

By the Markov property

$$H_n(x, \alpha, \theta, f) = \frac{1}{n} \log(K^n(x, \alpha).1)(\theta) = \frac{1}{n} \log\Bigg(\sum_{j=1}^{m} K_{\theta,j}^n(x, \alpha) \Bigg),$$

where $K(x, \alpha)$ is the matrix given by

$$K_{i,j}(x, \alpha) = K_{i,j}(x) \int_{\mathbb{R}^d} e^{\langle \alpha, f \rangle} \mu_x^j(df).$$

It follows from Hypothesis 3.4(i), (ii) and (iv) that $K(x, \alpha)$ is eventually positive (i.e. $K^n(x, \alpha)$ has positive entries for some $n \in \mathbb{N}$). Thus, by the Perron–Frobenius theorem $K(x, \alpha)$ has a positive eigenvalue $\rho(x, \alpha)$ which is isolated, positive and greater in absolute values than any other eigenvalue. Moreover, letting

$$H(x, \alpha) = \log(\rho(x, \alpha))$$

gives

$$H(x, \alpha) = \lim_{n \to \infty} H_n(x, \alpha, \theta, f).$$

The function $\alpha \to H(x, \alpha)$ is convex as a limit of convex functions. Since $\rho(x, \alpha)$ is an isolated characteristic root, the implicit function theorem implies that $H(x, \alpha)$ is smooth (even analytic) in α and continuous in x. Let

$$L(x, \beta) = \sup_{\alpha \in \mathbb{R}^d} \langle \alpha, \beta \rangle - H(x, \alpha)$$

be the Legendre–Fenchel transform of $H(x, .)$. By standard results of convex analysis (see, e.g., [**14**, Th. VII 2.1] or [**13**, Lemma 2.1]), $\beta \to L(x, \beta)$ is nonnegative, lower semicontinuous and essentially strictly convex.

Let $h \in C_x[0, T]$. Define

$$\mathcal{L}_{x,T}(h) = \int_0^T L(h(t), h'(t))\, dt$$

if h is absolutely continuous and set

$$\mathcal{L}_{x,T}(h) = \infty$$

otherwise. For convenience, we summarize some properties of $\mathcal{L}$ in the next lemma.

LEMMA 6.1.

(i) *Let* $\kappa > 0$ *be such that* $\operatorname{supp}(\mu^i) \subset B_\kappa = \{x \in \mathbb{R}^d : \|x\| < \kappa\}$ *for all* $i \in \mathcal{X}$. *Then* $L(x, \beta) = \infty$ *when* $\|\beta\| > \kappa$.

(ii) $h \to \mathcal{L}_{h(0),T}(h)$ *is lower semicontinuous on* $C[0, T]$.

https://doi.org/10.1017/S0143385798097557 Published online by Cambridge University Press

(iii) *Let* $K \subset \mathbb{R}^d$ *be a compact set. Then* $\{h \in C[0,T] : h(0) \in K, \mathcal{L}_{h(0),T}(h) \le s\}$ *is compact for every* $s \ge 0$.
(iv) $\mathcal{L}_{h(0),T}(h) = 0$ *if and only if* $h(t) = \Phi_t(h(0))$ *for all* $0 \le t \le T$.

Proof. (i) follows easily from the definition of $L(x,\beta)$ because $H(x,\alpha)$ is bounded by $\kappa\|\alpha\|$. (ii) is proved in Dupuis [**13**, Lemma A1]. (iii) If $\mathcal{L}_{h(0),T}(h) \le s$, part (i) implies that h is Lipschitz continuous with constant κ on $[0,T]$. Thus, the set $\{h \in C[0,T] : h(0) \in K, \mathcal{L}_{h(0),T}(h) \le s\}$ is bounded equicontinuous, and compactness follows from Ascoli's theorem. (iv) If $h(t) = \Phi_t(x), 0 \le t \le T$, then $\mathcal{L}_{x,T}(h) = 0$. Conversely, suppose that $\mathcal{L}_{x,T}(h) = 0$ and $h(0) = x$. Then, $L(h(t), h'(t)) = 0$ for almost every $0 \le t \le T$. By the smoothness of $H(x,\alpha)$ in α and Theorem II.6.3 of Ellis [**14**], $\beta = \lim_{n\to\infty}(1/n)\sum_{i=1}^n f_i^x = F(x)$ is the unique solution of the equation $L(x,\beta) = 0$. Thus, $h'(t) = F(h(t))$ for almost every $0 \le t \le T$. By Lipschitz continuity of F this implies that $h(t) = \Phi_t(x), 0 \le t \le T$. □

The following theorem follows from section 4 (Lemma 4.1) of Dupuis [**13**].

THEOREM 6.2. [**13**] *Let* $K \subset \mathbb{R}^d$ *be a compact set. For any Borel set* $A \subset C_x[0,T]$, *the following estimates hold uniformly in* $x \in K, \theta \in \mathcal{X}$:
(i) $\liminf_{\epsilon\to 0} \epsilon \log \mathbf{P}_{x,\theta}(x^\epsilon(\cdot) \in A) \ge -\inf\{\mathcal{L}_{x,T}(h) : h \in \mathrm{int}(A)\}$;
(ii) $\limsup_{\epsilon\to 0} \epsilon \log \mathbf{P}_{x,\theta}(x^\epsilon(\cdot) \in A) \le -\inf\{\mathcal{L}_{x,T}(h) : h \in \mathrm{clos}(A)\}$.

We conclude this section with two lemmas that will be useful in the following.
For $A \subset \mathbb{R}^d$ and $B \subset \mathbb{R}^d$ define

$$\mathcal{L}(A,B) = \inf_{T>0} \inf\{\mathcal{L}_{h(0),T}(h) : h \in C([0,T]), h(0) \in A, h(T) \in B\}.$$

LEMMA 6.3. *Let* $K \subset \mathbb{R}^d$ *be a compact set. Then for* $r > 0$ *large enough,* $\mathcal{L}(K, B_r^c) > 0$.

Proof. It follows from Hypothesis 2.3 and standard results on attractors and Lyapounov functions (see, e.g., Wilson [**31**]) that there exists a smooth compact d-dimensional manifold with boundary $M_0 \subset \mathbb{R}^d$ such that:
(i) $K \subset \mathrm{int}(M_0)$;
(ii) $\Phi_t(M_0) \subset \mathrm{int}(M_0)$ for all $t > 0$;
(iii) for all $x \in \mathbb{R}^d, \omega(x) \in \mathrm{int}(M_0)$.

Choose $r > r_0 > 0$ large enough so that $M_0 \subset B_{r_0}$. Let $T > 0$ and $\psi \in C[0,T]$ be such that $\psi(0) \in K, \psi(T) \in B_r^c$. Define $t_0 = \sup\{T \ge t \ge 0 : \psi(t) \in \partial M_0\}$ and $t_1 = \inf\{t \ge t_0 : \|\psi(t)\| \ge r\}$.

If $\mathcal{L}_{\psi(0),T}(\psi) = \infty$ there is nothing to prove, so we can assume $\mathcal{L}_{\psi(0),T}(\psi) < \infty$. By Lemma 6.1(i), $\|\psi'(s)\| \le \kappa$ for almost all $0 \le s \le T$. Thus, $r \le (t_1 - t_0)\kappa + r_0$.

Set $T' = (r - r_0)/\kappa$,

$$\mathcal{H} = \{h \in C[0,T'] : h(0) \in \partial M_0; h([0,T']) \subset \overline{B_r} \setminus \mathrm{int}(M_0)\}$$

and

$$\alpha = \inf\{\mathcal{L}_{h(0),T'}(h) : h \in \mathcal{H}\}.$$

https://doi.org/10.1017/S0143385798097557 Published online by Cambridge University Press

We have

$$\mathcal{L}_{\psi(0),T}(\psi) \geq \int_{t_0}^{t_0+T'} L(\psi, \psi') \geq \alpha.$$

Therefore, to conclude it suffices to show that $\alpha > 0$. Suppose the contrary. By Lemma 6.1(iii) the set $\mathcal{H}' = \{h \in \mathcal{H} : \mathcal{L}_{h(0),T'}(h) \leq 1\}$ is compact in $C[0, T']$. It follows that there is some sequence $h_n \in \mathcal{H}'$ for which $\lim_{n\to\infty} h_n = h \in \mathcal{H}'$ and $\liminf_{n\to\infty} \mathcal{L}_{h(0),T'}(h_n) = 0$. By lower-semi continuity this gives $\mathcal{L}_{h(0),T'}(h) = 0$. But Lemma 6.1(iv) leads to the impossible conclusion that $h(T') = \Phi'_T(h(0)) \in \overline{B_r} \setminus \mathrm{int}(M_0)$ with $h(0) \in M_0$ and $T' > 0$. □

The next lemma is similar to Lemma 5.3 of Kifer [**24**].

LEMMA 6.4. *Let $K \subset \mathbb{R}^d$ be a compact set which contains no omega limit points. Let*

$$\tau(K, T) = \inf\{\mathcal{L}_{\psi(0),T}(\psi) : \psi \in C([0, T]), \psi([0, T]) \subset K\}.$$

Then $\lim_{T\to\infty} \tau(K, T) = \infty$.

Proof. Since K contains no omega limit point, any forward orbit with initial condition in K spends all but a finite amount of time outside K. Furthermore, by compactness, this amount of time is bounded by some constant $T_K > 0$. Let $\alpha = \inf\{\mathcal{L}_{h(0),2T_K}(h) : h \in C([0, 2T_K]), h([0, 2T_K]) \subset K\}$. A proof similar to that of Lemma 6.3 shows that $\alpha > 0$. To conclude we note that if $T \geq 2nT_K, n \in \mathbb{N}$, and $h([0, 2nT_K]) \subset K$ then $\mathcal{L}_{h(0),T}(h) \geq n\alpha$. □

Large deviations properties, separating sets and invariant decompositions. The purpose of this section is to compare the relations

$$\text{‘}x \hookrightarrow y\text{’}$$

(there exists a pseudo-orbit from x to y) and

$$\text{‘}\mathcal{L}(x, y) = 0\text{’}$$

(the cost of large deviation from x to y is zero). The main result of the section (Theorem 6.7) will allow us to describe (under some suitable conditions) the set of y such that $\mathcal{L}(x, y) = 0$ by using only the nature of the limit points set $L(\Phi)$.

PROPOSITION 6.5. *Let $K \subset \mathbb{R}^d$ and $K' \subset \mathbb{R}^d$ be two compact sets. Suppose $\mathcal{L}(K, K') = 0$. Then one of the two following properties hold:*

(i) *there exists $x \in K$ such that $\gamma^+(x) \cap K' \neq \emptyset$;*

(ii) *there exist $x \in K$, $y \in K'$ such that $x \hookrightarrow y$.*

Proof. Suppose $\mathcal{L}(K, K') = 0$. Then for all $n \in \mathbb{N}^*$ there exists $T_n > 0, h_n \in C([0, T_n]), h_n(0) \in K, h_n(T_n) \in K'$ and $\mathcal{L}_{h_n(0),T_n}(h_n) \leq 1/n$.

First, suppose that $\{T_n\}$ is bounded. By taking a subsequence we may assume that $T_n \to T$ as $n \to \infty$ for some $T > 0$. Set $T_\epsilon = T - \epsilon$. For n large enough, $T_n \geq T_\epsilon$. Thus, $\mathcal{L}_{h_n(0),T_\epsilon}(h_n) \leq \mathcal{L}_{h_n(0),T_n}(h_n) \leq 1/n$. By Lemma 6.1(iii) we can extract from $\{h_n\}$

https://doi.org/10.1017/S0143385798097557 Published online by Cambridge University Press

a subsequence $\{h_{n_i^\epsilon}\}$ which converges uniformly on $[0, T_\epsilon]$ toward some function h_ϵ. Assertions (ii) and (iv) of Lemma 6.1 imply that $\Phi_{T_\epsilon}(h_\epsilon(0)) = h_\epsilon(T_\epsilon)$. On the other hand,

$$\begin{aligned} |h_{n_i^\epsilon}(T_{n_i^\epsilon}) - h_\epsilon(T_\epsilon)| &\leq |h_{n_i^\epsilon}(T_\epsilon) - h_\epsilon(T_\epsilon)| + |h_{n_i^\epsilon}(T_\epsilon) - h_{n_i^\epsilon}(T_{n_i^\epsilon})| \\ &\leq |h_{n_i^\epsilon}(T_\epsilon) - h_\epsilon(T_\epsilon)| + \kappa\epsilon \end{aligned}$$

because of Lemma 6.1(i) and the fact that $\mathcal{L}_{h_n(0),T_\epsilon}(h_n)$ is finite. Thus, for n_i^ϵ large enough this gives

$$|h_{n_i^\epsilon}(T_\epsilon) - \Phi_{T_\epsilon}(h_{\epsilon(0)})| \leq \epsilon(1+\kappa).$$

Since $h_n(T_n) \in K'$ and $h_\epsilon(0) \in K$, this gives by compactness a point $x \in K$ for which $\Phi_T(x) \in K'$.

Suppose now that $\{T_n\}$ is unbounded. By taking a subsequence we may assume that $T_n \to \infty$ as $n \to \infty$. By Lemma 6.3 we may also assume that $h_n(\mathbb{R}_+) \subset K_1$, where K_1 is some compact set.

Fix $T > 0, \delta > 0$. For $n \in \mathbb{N}$, write $T_n = k(n)(T + r_n)$, where $k(n) \in \mathbb{N}$ and $0 \leq r_n < T/k(n)$. We have

$$\int_0^{T_n} L(h_n, h_n') = \sum_{i=0}^{k(n)-1} \int_{i(T+r_n)}^{(i+1)(T+r_n)} L(h_n, h_n') \leq \frac{1}{n}.$$

Set $x_{i,n} = h_n(i(T + r_n)); i = 0, \dots, k(n) - 1$. We claim that for n large enough, $d(\Phi_{T+r_n}(x_{i,n}), x_{i+1,n}) \leq \delta$. This gives a (δ, T) pseudo-orbit from K to K' and part (ii) of the lemma is proved. To prove the claim, suppose that to the contrary there exist subsequences $n_j \to \infty$ and $i_j \leq k(n_j) - 1$ for which

$$d(\Phi_{T+r_{n_j}}(x_{i_j,n_j}), x_{i_j+1,n_j}) > \delta.$$

Set $g_{n_j}(u) = h_{n_j}(u + i_j(T + r_{n_j}))$. We have $x_{i_j,n_j} = g_{n_j}(0) \in K_1$ and

$$\mathcal{L}_{g_{n_j}(0),T}(g_{n_j}) \leq \int_{i_j(T+r_{n_j})}^{(i_j+1)(T+r_{n_j})} L(h_{n_j}, h_{n_j}') \leq \frac{1}{n_j}.$$

Thus, by Lemma 6.1(iii) and (ii) we may assume that g_{n_j} converges uniformly on $[0, T]$ toward some function g for which $\mathcal{L}_{g(0),T}(g) = 0$. By Lemma 6.1(iv), this implies that $d(\Phi_T(x_{i_j,n_j}), g_{n_j}(T)) \to 0$ as $n_j \to \infty$. Since g_{n_j} is Lipschitz with constant κ and $t \to \Phi_t(x)$ is Lipschitz continuous on $[0, T]$ uniformly in $x \in K_1$ it follows easily that

$$\lim_{n_j \to \infty} d(\Phi_{T+r_{n_j}}(x_{i_j,n_j}), x_{i_j+1,n_j}) = 0.$$

Hence, we get a contradiction. □

The next result gives a partial converse to Proposition 6.5. Recall that

$$L(\Phi) = \text{clos}\left(\bigcup_{x \in \mathbb{R}^d} \omega(x) \cup \alpha(x)\right).$$

https://doi.org/10.1017/S0143385798097557 Published online by Cambridge University Press

LEMMA 6.6. *Suppose that the process (1) is nondegenerate at every point $x \in L(\Phi)$. Let $y \in \mathbb{R}^d$ be such that the backward trajectory of y, $\gamma^-(y)$, is bounded. Let*

$$F_y = \{z \in \mathbb{R}^d : \forall t \geq 0, \mathcal{L}(\Phi_t(z), \alpha(y)) = 0\}.$$

Suppose that $\text{clos}(F_y) \cap L(\Phi)$ *is open in* $L(\Phi)$*. Then* $x \hookrightarrow y$ *implies* $\mathcal{L}(x, \alpha(y)) = 0$.

Proof.

Claim.
(i) The set $\text{clos}(F_y)$ is a nonempty closed invariant set.
(ii) Let $x \in \mathbb{R}^d$. If $\omega(x) \cap \text{clos}(F_y) \neq \emptyset$, then $x \in F_y$.
The set F_y contains $\alpha(y)$. Thus it is nonempty. It is closed by definition and it is clearly invariant. Let $z \in \omega(x) \cap \text{clos}(F_y)$. There exist sequences $t_n \to \infty$ and $z_n \in F_y$ such that $p_n = \Phi_{t_n}(x) \to z$ and $z_n \to z$ as $n \to \infty$. Let $\Theta : \mathbb{R} \to [0, 1]$ be a C^∞ function which is 1 on $(\infty, 0]$ and 0 on $[1, \infty)$. Define

$$h_n(s) = \Theta(s)\Phi_s(p_n) + (1 - \Theta(s))\Phi_s(z_n).$$

The function h_n is a smooth interpolation between $s \to \Phi_s(p_n)$ and $s \to \Phi_s(z_n)$ with the property that $h_n(s) = \Phi_s(z_n)$ for $s \geq 1$ and $h_n(s) = \Phi_s(p_n)$ for $s \leq 0$. Since $p_n \to z$ and $z_n \to z$ it follows that $h_n(s)$ converges to $\Phi(s, z)$ for the uniform C^1 topology on $[0, 1]$. Thus, by uniform continuity of F on a neighborhood of $\overline{\gamma(z)}$, this implies that $\lim_{n\to\infty} \|h_n'(s) - F(h_n(s))\| = 0$ uniformly for $0 \leq s \leq 1$. Now, Lemma 7.5 implies that $\mathcal{L}_{p_n,1}(h_n) \to 0$ as $n \to \infty$. Since $\mathcal{L}_{p_n,1}(h_n) = \mathcal{L}_{x,t_n+1}(h_n)$ and $h_n(t_n + 1) = z_n \in F_y$ it follows that $x \in F_y$. This proves the claim.

Let $x \in \mathbb{R}^d$ be such that $x \hookrightarrow y$. The set of z such that $x \hookrightarrow z$ is closed and invariant and thus contains $\alpha(y)$. Suppose $x \notin F_y$. Then, by the claim, $\omega(x) \cap F_y = \emptyset$. Therefore we are in the situation where:
(i) $\text{clos}(F_y)$ is a nonempty closed invariant set;
(ii) $\text{clos}(F_y) \cap L(\Phi)$ is open and closed in $L(\Phi)$;
(iii) $\omega(x) \cap F_y = \emptyset$;
(iv) $x \hookrightarrow \alpha(y) \subset F_y \subset \text{clos}(F_y)$.
Properties (i) and (ii) mean that $\text{clos}(F_y)$ *separates* $L(\Phi)$. Then by a lemma due to Nitecki and Shub, and extended by Akin [**1**, Lemma 13, p 86], there exists $u \in \mathbb{R}^d$ such that $x \hookrightarrow u$, $\omega(u) \cap \text{clos}(F_y) \neq \emptyset$ and $\alpha(u) \cap \text{clos}(F_y) = \emptyset$. But, according to the claim, the property $\omega(u) \cap \text{clos}(F_y) \neq \emptyset$ implies $u \in F_y \subset \text{clos}(F_y)$ and since $\text{clos}(F_y)$ is closed and invariant, it follows that $\alpha(u) \subset \text{clos}(F_y)$ which is contradictory. Therefore $x \in F_y$. □

We now give the main result of this section.

THEOREM 6.7. *Suppose that:*
(i) *the limit points set decomposes as*

$$L(\Phi) = \bigcup_{i=1}^{k} \Gamma_i,$$

where $\Gamma_1, \ldots, \Gamma_k$ *are compact invariant disjoint internally chain transitive sets such that* $J(\Phi|\Gamma_i, \Gamma_i) < 2$;

https://doi.org/10.1017/S0143385798097557 Published online by Cambridge University Press

(ii) *the process is nondegenerate at every* $x \in L(\Phi)$.
Then the following relations are equivalent:
(a) $\mathcal{L}(x, y) = 0$*;*
(b) $x \hookrightarrow y$ *or* $y = \Phi_t(x)$ *for some* $t \geq 0$*;*
(c) *there exist an integer* $2 \leq j \leq k$*, points* $x_1, \dots, x_k$ *and a permutation* $\sigma : \{1, \dots, k\} \to \{1, \dots, k\}$ *such that*

$$x_1 = x, \omega(x) \subset \Gamma_{\sigma(1)}; x_k = y, \alpha(y) \subset \Gamma_{\sigma(k-1)};$$
$$\alpha(x_j) \subset \Gamma_{\sigma(j-1)}, \omega(x_j) \subset \Gamma_{\sigma(j)}; j = 2, \dots, k-1.$$

Proof. The implication (a) $\Rightarrow$ (b) follows from Lemma 6.5. Conversely, if $y \in \gamma^+(x)$ then $\mathcal{L}(x, y) = 0$. If $x \hookrightarrow y$ let F_y be as in Lemma 6.6. Choose $i \in \{1, \dots, k\}$ such that $\alpha(y) \subset \Gamma_i$. Then $\mathrm{clos}(F_y) \cap \Gamma_i \neq \emptyset$ and it follows easily from Lemma 7.6 (§7) that $\Gamma_i \subset \mathrm{clos}(F_y)$ and consequently $\mathrm{clos}(F_y) \cap L(\Phi)$ is open in $L(\Phi)$. Thus, $\mathcal{L}(x, \alpha(y)) = 0$ by Lemma 6.6. Lemma 7.6 again implies that $\mathcal{L}(x, y) = 0$. To obtain the equivalence between (b) and (c) consider the quotient space $\widetilde{\mathbb{R}^d}$ obtained by collapsing each Γ_i to a point and let $\widetilde{\Phi}$ denote the quotient flow. Clearly, $\widetilde{\Phi}$ is a simple flow†. As already noticed in the proof of Theorem 4.3, the relation $x \hookrightarrow y$ and $x \rightsquigarrow y$ are equivalents for a simple flow. This implies the equivalence of (b) and (c). □

7. *Proof of Theorem 3.7*

The induced chain. In this subsection we state a proposition (Proposition 7.1) whose proof is postponed to the next subsection and we show how Theorem 3.7 can be deduced from this proposition. The argument follows the lines of the exposition in Kifer [**24**] or Freidlin and Wentzell [**17**].

Assume that the hypotheses of Theorem 3.7 hold. Set $\Lambda_1 = C, \Lambda_2 = C^+$ and $\Lambda_3 = \{x \in \mathcal{R}(\Phi) : x \notin C \cup C^+\}$. It follows from compactness of $\mathcal{R}(\Phi)$, Lemmas 5.2 and 5.3 that the Λ_i's are compact (invariant) disjoint sets. Therefore, for $r > 0$ small enough, the sets $U_i = N_r(\Lambda_i), i = 1, 2, 3$, are disjoint neighborhoods. Let $V_i \subset U_i$ be an open neighborhood of Λ_i. Set $V = V_1 \cup V_2 \cup V_3$ and define the *induced chain* $(\tilde{X}_n^\epsilon, \tilde{\Theta}_n^\epsilon)$ on $V \times \mathcal{X}$ whose transition probabilities $\tilde{P}_{x,\theta}^\epsilon(B \times \Sigma)$ are given as

$$\tilde{P}_{x,\theta}^\epsilon(B \times \Sigma) = \mathbf{P}(X_{T_V}^\epsilon \in B; \Theta_{T_V}^\epsilon \in \Sigma \mid X_0^\epsilon = x; \Theta_0^\epsilon = \theta),$$

where $T_V = \inf\{n > 0; X_n^\epsilon \in V\}$, B is a Borel subset of V and $\Sigma \subset \mathcal{X}$.

The neighborhoods V_1, V_2, V_3 can be chosen in such a way that have the following.

PROPOSITION 7.1. *There exist constants* $b > a > 0$ *and a sequence of integers* $\{n(\epsilon)\}_{\epsilon>0}$ *such that:*
(a) $\liminf_{\epsilon\to 0} \epsilon \log(\tilde{P}_{x,\theta}^\epsilon(n(\epsilon), V_2 \times \mathcal{X})) \geq -a$ *uniformly in* $x \in V_1, y \in \mathcal{X}$*;*
(b) $\limsup_{\epsilon\to 0} \epsilon \log(\tilde{P}_{x,\theta}^\epsilon(n(\epsilon), V_j \times \mathcal{X})) \leq -b$ *uniformly in* $x \in V_i, y \in \mathcal{X}$*, for all* $(i, j) \in \{(2, 1), (1, 3), (2, 3)\}$.
Here $(\tilde{P}_{x,\theta}^\epsilon(n, .))$ *denotes the transition kernel of the induced chain in* n *steps.*

† Simple flows were defined in §4.

Now, let ν^ϵ be an invariant probability measure of $\{X_n^\epsilon, \Theta_n^\epsilon\}$. By Proposition 5.3 of Kifer [**24**], if $\nu^\epsilon(V \times \mathcal{X}) > 0$ the probability measure $\tilde{\nu}^\epsilon$ defined on $V \times \mathcal{X}$ by

$$\tilde{\nu}^\epsilon(B \times \Sigma) = \frac{\nu^\epsilon(B \times \Sigma)}{\nu^\epsilon(V \times \mathcal{X})}$$

for any Borel set $B \subset V, \Sigma \subset \mathcal{X}$, is an invariant probability measure of the induced chain. Suppose $\nu^\epsilon(V \times \mathcal{X}) > 0$ (otherwise there is nothing to prove). Then define for $i \neq j$,

$$P_{i,j}^\epsilon = \frac{1}{\tilde{\nu}_\epsilon(V_i \times E)} \int_{V_i \times \mathcal{X}} \tilde{P}_{x,\theta}^\epsilon(n(\epsilon), V_j \times \mathcal{X}) \tilde{\nu}_\epsilon(dx \otimes d\theta)$$

if $\nu^\epsilon(V_i \times \mathcal{X}) \neq 0$ and $P_{i,j}^\epsilon = 0$ otherwise. Set $P_{i,i}^\epsilon = 1 - \sum_{j \neq i} P_{i,j}^\epsilon$. Since $\tilde{\nu}^\epsilon$ is an invariant measure of the induced chain, the vector μ^ϵ defined by $\mu_i^\epsilon = \nu^\epsilon(V_i \times \mathcal{X}), i = 1, 2, 3$, is an invariant probability measure of the Markov chain defined on $\{1, 2, 3\}$ by the transition matrix $P^\epsilon = (P_{i,j}^\epsilon)_{i,j=1,2,3}$.

LEMMA 7.2. *Let P^ϵ be a 3×3 Markov transition matrix such that:*
(a) $P_{1,2}^\epsilon \geq e^{-a'/\epsilon}$;
(b) $P_{i,j}^\epsilon \leq e^{-b'/\epsilon}$ *for all* $(i, j) \in \{(2, 1), (1, 3), (2, 3)\}$,
where $b' > a' > 0$. Let μ^ϵ be an invariant probability vector of P^ϵ. Then $\lim_{\epsilon \to 0} \mu_1^\epsilon = 0$.

Proof. We have

$$\mu_1^\epsilon(P_{1,2}^\epsilon + P_{1,3}^\epsilon) = \mu_2^\epsilon P_{2,1}^\epsilon + \mu_3^\epsilon P_{3,1}^\epsilon \tag{19}$$

$$\mu_3^\epsilon(P_{3,1}^\epsilon + P_{3,2}^\epsilon) = \mu_1^\epsilon P_{1,3}^\epsilon + \mu_2^\epsilon P_{2,3}^\epsilon. \tag{20}$$

If $P_{3,1}^\epsilon = 0$, then (19) implies $\mu_1^\epsilon \leq P_{2,1}^\epsilon / P_{1,2}^\epsilon \leq e^{-(b'-a')/\epsilon}$ and then $\mu_1^\epsilon \to 0$ as $\epsilon \to 0$. If $P_{3,1}^\epsilon \neq 0$, then (20) implies $\mu_3^\epsilon \leq 2e^{-b'/\epsilon}/(P_{3,1}^\epsilon + P_{3,2}^\epsilon) \leq 2e^{-b'/\epsilon}/P_{3,1}^\epsilon$. Using this last inequality in (19) gives $\mu_1^\epsilon \leq (P_{2,1}^\epsilon + 2e^{-b'/\epsilon})/P_{1,2}^\epsilon \leq 3e^{-(b'-a')/\epsilon}$ and then $\mu_1^\epsilon \to 0$ as $\epsilon \to 0$. □

Theorem 3.7 now follows from the preceding discussion, Proposition 7.1 and Lemma 7.2.

Proof of Proposition 7.1.

LEMMA 7.3. *Let C be a connected component of $\mathcal{R}(\Phi)$, and U a neighborhood of C. There exist $\delta > 0, T > 0$ such that for all $x \in C, y \in C$ every (δ, T) pseudo-orbit from x to y belongs to U.*

Proof. Assume the contrary. Then there is an open neighborhood U of C such that for any $\delta > 0, T > 0$ one can find $x_{\delta,T} \in C, y_{\delta,T} \in C$ and $p_{\delta,T} \notin U$ with $x_{\delta,T} \hookrightarrow_{T,\delta} p_{\delta,T} \hookrightarrow_{T,\delta} y_{\delta,T}$. Letting $\delta \to 0$ and $T \to \infty$ we get, by compactness, $x \in C, y \in C$ and $p \in C \setminus U$ with $x \hookrightarrow p \hookrightarrow y$. This implies $p \in C$. This is contradictory to $p \notin U$. □

If $\{\Phi_t(x_i) : 0 \leq t \leq t_i\}, i = 1, \dots, k-1$, is a (δ, T) pseudo-orbit, we define its *length* as the number $l = \sum_{i=1}^{k-1} t_i$.

https://doi.org/10.1017/S0143385798097557 Published online by Cambridge University Press

LEMMA 7.4. *Let C be a connected component of $\mathcal{R}(\Phi)$ and $\Gamma \subset C$ a nonempty compact invariant set. Let $s > J(\Phi, \Gamma)$ and $\eta > 0$. Then for any $\delta > 0, T > 0$ there exists a neighborhood of Γ, $V \subset N_\delta(\Gamma)$ and a positive constant $l(\delta, T) > T$ such that:*

(i) *For all $x \in V, y \in V$ it is possible to find a (δ, T) pseudo-orbit from x to y $\{\Phi_t(x_i) : 0 < t \le t_i\}, i = 1, \dots, k-1$, which satisfies:*
   (a) $\sum_i d(\Phi_{t_i}(x_i), x_{i+1})^s \le \eta$;
   (b) *The length of the pseudo-orbit is bounded by $l(\delta, T)$.*

(ii) *If, furthermore, $L(\Phi|C) \subset \Gamma$ then the pseudo-orbit from $x \in V$ to $y \in V$ can be chosen such that $x_i \in N_\delta(\Gamma)$ for all i.*

Proof. (i) Using Lemma 7.3 choose $0 < \delta' < \delta/2$ and $T > 0$ such that any (δ', T) pseudo-orbit from $x \in C$ to $y \in C$ remains in $N_{\delta/2}(C)$.

Let $B_{\delta/2}(p_k), k = 1, \dots, n$, be a finite covering of Γ by balls of center $p_i \in \Gamma$ and radius $\delta/2$. Since Γ is chain transitive $p_k \hookrightarrow p_m$ for $k, m = 1, \dots, n$. Therefore, the definition of $J(\Phi, \Gamma)$ and the assumption $s > J(\Phi, \Gamma)$ imply the existence of a (δ', T) pseudo-orbit from p_k to p_m with the property that $\sum_i d(\Phi_{t_i}(x_i), x_{i+1})^s \le \eta$. Let $l_{k,m}$ denote the length of this pseudo-orbit and let $l(\delta, T) = \sup\{l_{k,m} : k, m = 1, \dots, n\}$. Set $V = \bigcup_{k=1}^n B_{\delta/2}(p_k)$. For $x \in V, y \in V$ there exist k, m such that $x \in B_{\delta/2}(p_k)$ and $y \in B_{\delta/2}(p_m)$. The (δ', T) pseudo-orbit from p_k to p_m gives a (δ, T) pseudo-orbit from x to y which satisfies assertions (a) and (b).

(ii) Now suppose $L(\Phi) \subset \Gamma$. Let $\delta > 0$. Since any alpha or omega limit point of $\Phi|C$ is in Γ there exists $T(\delta) > 0$ such that for all $x \in C \setminus (N_{\delta/2}(\Gamma))$ and $|T| > T(\delta)$, $\Phi_T(x) \in N_{\delta/2}(\Gamma)$.

Let $T > T(\delta)$. By Lipschitz continuity of the flow on $C \times [-T, T]$ there exists some constant $k(T)$ such that

$$d(\Phi_s(x), \Phi_s(y)) \le k(T)\, d(x, y)$$

for $|s| \le T$.

Set $\delta_1 = \delta/2k(T)$ and $\eta_1 = \eta/(k(T)^s \delta^s)$. Part (i) of the lemma gives a neighborhood $V \subset N_{\delta_1}(\Gamma) \subset N_\delta(\Gamma)$ and a $(\delta_1, 2T)$ pseudo-orbit $\{\Phi_t(x_i) : 0 \le t \le t_i\}, i = 1, \dots, k-1, t_i > 2T$ from $x \in V$ to $y \in V$ having a length bounded by $l(\delta_1, 2T)$ and such that $\sum_i d(\Phi_{t_i}(x_i), x_{i+1})^s \le \eta_1$.

Let $J = \{j \in \{1, \dots, k-1\} : x_j \notin N_\delta(\Gamma)\}$. For $j \notin J$ set $y_j = x_j$ and for $j \in J$ set $y_j = \Phi_{-T}(x_j)$. The y_j are now all in $N_\delta(\Gamma) \cup N_\delta(\Gamma')$.

Notice that for $j \in J$, $\Phi_T(x_j)$ and $\Phi_{-T}(x_j)$ are in $N_{\delta/2}(\Gamma) \cup N_{\delta/2}(\Gamma')$. Thus x_{j+1} and x_{j-1} are in $N_\delta(\Gamma) \cup N_\delta(\Gamma')$. Therefore, when $j \in J$, $j-1$ and $j+1$ are not in J. Using this remark, define a new sequence of time as follows. If $j \in J$ set $t'_j = t_j + T$ and $t_{j-1} = t_{j-1} - T$. If $j \notin J$ and $j+1 \notin J$ set $t'_j = t_j$.

This gives a (δ, T) pseudo-orbit $\{\Phi_t(y_i) : 0 \le t \le t'_i\}, i = 1, \dots, k-1, t'_i > T$ having length less that $l(\delta_1, 2T)$ and such that

$$\sum_i d(\Phi_{t'_i}(y_i), y_{i+1})^s \le k(T)^s \delta^s \eta_1 \le \eta.$$

This concludes the proof. □

https://doi.org/10.1017/S0143385798097557 Published online by Cambridge University Press

The next lemma gives the key estimate to prove Proposition 7.1.

LEMMA 7.5. *Let $K \subset \mathbb{R}^d$ be a compact set. Assume that (1) is nondegenerate at each point $x \in K$. Then there exist a compact neighborhood N of K, $r > 0$ and a positive constant C such that*

$$L(x, \beta) \leq C\|\beta - F(x)\|^2$$

for all $x \in N$ and $\beta \in B_r(F(x)) = B(F(x), r)$.

Proof. Let $H_x : \mathbb{R}^d \to \mathbb{R}$ denote the map given by $H_x(\alpha) = H(x, \alpha) = \log(\rho(x, \alpha))$ and let $D^k H_x$ denote the kth derivative of H_x. As usual we identify $DH_x(\alpha)$ with a vector of $\mathbb{R}^d$ and $D^2 H_x(\alpha)$ with a $d \times d$ symmetric matrix. As already noticed, $\rho(x, \alpha)$ is positive analytic in α and continuous in x. Similarity, $D^k H_x(\alpha)$ is continuous in x and analytic in α.

Our first goal is to compute $DH_x(0)$ and $D^2 H_x(0)$. Fix $x \in \mathbb{R}^d$ and $\alpha \in \mathbb{R}^d$ and define the maps $T(z) = K(x, z\alpha)$ and $\lambda(z) = \rho(x, z\alpha)$, where z denote a real (or possibly complex) number. Clearly

$$T(z) = T + \sum_{k \geq 1} z^k T^{(k)}$$

with $T = T(0) = K(x)$ and $T^{(k)} = K(x)D^{(k)}/k!$, where $D^{(k)}$ denotes the diagonal matrix whose generic entry is

$$D_{i,i}^{(k)} = \int (\langle \alpha, f \rangle)^k \mu_x^i(df).$$

If we write the second-order expansion of $\lambda(z)$ in the neighborhood of $z = 0$ in the form

$$\lambda(z) = 1 + z\lambda^{(1)} + z^2 \lambda^{(2)} + o(z^2)$$

we get that

$$\log(\lambda(z)) = z\lambda^{(1)} + z^2(\lambda^{(2)} - \tfrac{1}{2}(\lambda^{(1)})^2) + o(z^2)$$

and therefore

$$DH_x(0).\alpha = \lambda^{(1)}$$

and

$$\langle D^2 H_x(0)\alpha, \alpha \rangle = 2\lambda^{(2)} - (\lambda^{(1)})^2.$$

To compute $\lambda^{(1)}$ and $\lambda^{(2)}$ we will use the perturbation theory of linear operators [**23**]. Recall that $\langle\,,\rangle_{\pi(x)}$ denotes the inner product on $\mathbb{R}^{\mathcal{X}}$ defined by $\langle f, g \rangle_{\pi(x)} = \sum_{i \in \mathcal{X}} f_i g_i \pi_i(x)$ while $\langle\,,\rangle$ is the usual inner product on $\mathbb{R}^d$. The eigenvector space of $T = K(x)$ associated to the eigenvalue $\lambda(0) = 1$ is

$$\mathbb{R}.\mathbf{1} = \{h \in \mathbb{R}^{\mathcal{X}} : h_1 = \cdots = h_d\}$$

and the associated eigenprojector is $P : \mathbb{R}^{\mathcal{X}} \to \mathbb{R}^{\mathcal{X}}$ defined by

$$Ph = \langle h, \mathbf{1} \rangle_{\pi(x)} \mathbf{1} = \left(\sum_{i \in \mathcal{X}} \pi_i(x) h_i \right) \mathbf{1}.$$

https://doi.org/10.1017/S0143385798097557 Published online by Cambridge University Press

Define the operator $S : \mathbb{R}^{\mathcal{X}} \to \mathbf{1}^{\perp} \subset \mathbb{R}^{\mathcal{X}}$ by

$$Sh = L(x)^{-1}(Ph - h),$$

where $L(x)$ is the operator given by (9).

It follows from [**23**, p. 79] that

$$\begin{aligned} \lambda^{(1)} &= \mathrm{Tr}(T^{(1)}P) \\ \lambda^{(2)} &= \mathrm{Tr}(T^{(2)}P - T^{(1)}ST^{(1)}P). \end{aligned}$$

Now it is easy to check that $\mathrm{Tr}(AP) = \langle A\mathbf{1}, \mathbf{1}\rangle_{\pi(x)}$ and $\langle Kh, \mathbf{1}\rangle_{\pi(x)} = \langle h, \mathbf{1}\rangle_{\pi(x)}$. Using these facts we get that $\lambda^{(1)} = \langle K(x)D^{(1)}\mathbf{1}, \mathbf{1}\rangle_{\pi(x)} = \langle D^{(1)}\mathbf{1}, \mathbf{1}\rangle_{\pi(x)} = \langle \alpha, F(x)\rangle$. Thus

$$DH_x(0) = F(x).$$

Similarly, we have

$$\begin{aligned} \lambda^{(2)} &= \left\langle \frac{D^{(2)}}{2}\mathbf{1}, \mathbf{1} \right\rangle_{\pi(x)} - \langle D^{(1)}SK(x)D^{(1)}\mathbf{1}, \mathbf{1}\rangle_{\pi(x)} \\ &= \left\langle \frac{D^{(2)}}{2}\mathbf{1}, \mathbf{1} \right\rangle_{\pi(x)} - \left\langle K(x)SD^{(1)}\mathbf{1}, D^{(1)}\mathbf{1}\right\rangle_{\pi(x)} \end{aligned}$$

because $D^{(1)}$ is self-adjoint for $\langle\,,\rangle_{\pi(x)}$ and $K(x)$ commutes with S. Since $K(x)S$ maps $\mathbb{R}^{\mathcal{X}}$ into $\mathbf{1}^{\perp}$ we can rewrite $\lambda^{(2)}$ as

$$\lambda^{(2)} = \left\langle \frac{D^{(2)}}{2}\mathbf{1}, \mathbf{1} \right\rangle_{\pi(x)} - \langle K(x)SD^{(1)}\mathbf{1}, D^{(1)} - PD^{(1)}\mathbf{1}\rangle_{\pi(x)}.$$

Now, we use the fact that

$$D^{(1)} - PD^{(1)} = \langle \hat{f}(x) - F(x), \alpha\rangle = (\hat{f}(x) - F(x))_{\alpha},$$

where we use the notation (11) and (12). This gives

$$SD^{(1)}\mathbf{1} = -\hat{g}(x)_{\alpha}.$$

Thus, a simple computation gives

$$\lambda^{(2)} = \left\langle \frac{D^{(2)}}{2}\mathbf{1}, \mathbf{1} \right\rangle_{\pi(x)} - \tfrac{1}{2}\|\langle \hat{f}(x) - F(x), \alpha\rangle\|^2_{\pi(x)} + \tfrac{1}{2}Q^0_x(\alpha).$$

Finally, we obtain that

$$\langle D^2H_x(0)\alpha, \alpha\rangle = \langle D^{(2)}\mathbf{1}, \mathbf{1}\rangle_{\pi(x)} - \|\langle \hat{f}(x) - F(x), \alpha\rangle\|^2_{\pi(x)} - \langle \alpha, F(x)\rangle^2 + Q^0_x(\alpha) = Q_x(\alpha).$$

Our next goal is to establish the inequality stated in the lemma. By continuity of $x \to Q_x$ there exists a compact neighborhood N of K such that Q_x is nondegenerate at every point $x \in N$. Thus the eigenvalues of Q_x are uniformly bounded below by some positive number λ. Equivalently,

$$\|[D^2H_x(0)]^{-1}\| \leq 1/\lambda \tag{21}$$

for all $x \in N$, where $\|[D^2_xH(0)]^{-1}\|$ is the norm operator.

https://doi.org/10.1017/S0143385798097557 Published online by Cambridge University Press

For $x \in N$ set $\psi_x(\alpha) = DH_x(\alpha)$. The derivative $D\Psi_x(0) = D^2H_x(0)$ is invertible. On the other hand, the mean value theorem and the continuity of $D\Psi_x(\alpha)$ in (x, α) easily imply the existence of numbers $\eta > 0$ and $0 < s < 1$ such that

$$\|D\Psi_x(\alpha) - D\Psi_x(0)\| \leq \lambda(1-s)$$

for any $\alpha \in \operatorname{clos}(B_\eta)$.

Set $r = \lambda s \eta$. A refinement of the inverse function theorem (see, e.g., Hirsch and Pugh [**19**]) shows that:

(i) $\Psi_x(\operatorname{clos}(B_\eta)) \supset \operatorname{clos}(B(\Psi_x(0), r)) = \operatorname{clos}(B(F(x), r))$;

(ii) there exists open sets U_x, V_x such that $U_x \subset \operatorname{clos}(B_\eta)$ and $V_x \supset \operatorname{clos} B(F(x), r)$ such that Ψ_x induces a diffeomorphism from U_x onto V_x whose inverse mapping will be denoted by Ψ_x^{-1} for convenience.

Given $x \in N$ and $\beta \in B(F(x), r)$ set $\alpha = \psi_x^{-1}(\beta)$. By a property of the Legendre–Fenchel transform (see, e.g., Ellis [**14**, Theorem VII.2.1]),

$$L(x, \beta) = \langle \alpha, \Psi_x(\alpha)\rangle - H(x, \alpha). \tag{22}$$

Now the Taylor formula gives

$$\|\Psi_x(\alpha) - F(x) - D^2H_x(0).\alpha\| \leq \frac{c_1}{2}\|\alpha\|^2 \tag{23}$$

and

$$\|H_x(\alpha) - \langle F(x), \alpha\rangle - \tfrac{1}{2}\langle D^2H_x(0).\alpha, \alpha\rangle\| \leq \frac{c_1}{6}\|\alpha\|^3, \tag{24}$$

where $c_1 = \sup_{x\in N, \alpha\in \operatorname{clos}(B_\eta)} \|D^3H_x(\alpha)\|$. Thus (22)–(24) give

$$\|L(x, \beta) - \tfrac{1}{2}Q_x(\alpha)\| \leq c_1\|\alpha\|^3. \tag{25}$$

Equation (23) implies

$$\|[D^2H_x(0)]^{-1}(\beta - F(x)) - \alpha\| \leq \frac{c_1}{2\lambda}\|\alpha\|^2 = c_2\|\alpha\|^2, \tag{26}$$

with $c_2 = c_1/2\lambda$. Also,

$$\|\alpha\| = \|\Psi_x^{-1}(\beta) - \Psi_x^{-1}(F(x))\| \leq c_3\|\beta - F(x)\|, \tag{27}$$

with $c_3 = \sup_{x\in N, \alpha\in \operatorname{clos}(B_\eta)} \|D\Psi_x^{-1}(\alpha)\|$. Finally, by putting (25)–(27) together we get that

$$\|L(x, \beta) - \tfrac{1}{2}\langle [D_x^2H(0)]^{-1}(\beta - F(x)), \beta - F(x)\rangle\| \leq c_4\|\beta - F(x)\|^2$$

for some $c_4 > 0$ depending on c_1, c_2, c_3 and λ. Thus

$$\|L(x, \beta)\| \leq C\|\beta - F(x)\|^2,$$

where $C = 1/(2\lambda + c_4)$ is independent of $x \in N$ and $\beta \in B(F(x), r)$. □

LEMMA 7.6. *Let C be a connected component of $\mathcal{R}(\Phi)$. Assume that $J(\Phi, L(\Phi|C)) < 2$ and that the process (1) is nondegenerate at each $x \in L(\Phi|C)$. Then for any $\eta > 0$ and any neighborhood U of C there exists a neighborhood $W \subset U$ of C and a positive number T_1 with the property that for all $x, y \in W$ one can find a C^1 function $\psi : [0, l] \to U, l \leq T_1$ which satisfies:*

https://doi.org/10.1017/S0143385798097557 Published online by Cambridge University Press

(a) $\Psi(0) = x$; $\Psi(l) = y$;

(b) $\mathcal{L}_{x,l}(\psi) \leq \eta$.

Proof. Set $\Gamma = L(\Phi|C)$. Choose $0 < \alpha < 1$ such that $2 > 2 - \alpha > J(\Phi, \Gamma)$ and set $s = 2 - \alpha$. Choose $\delta > 0$ small enough so that $N_\delta(\Gamma) \subset U$ and for all $x \in N_\delta(\Gamma)$ the form Q_x is nondegenerate.

Fix $T > 1$, and then apply Lemma 7.4 to Γ. This gives us a neighborhood of Γ, $V \subset N_\delta(\Gamma)$ and a positive number $T_0 = l(\delta, T)$. Now, since $\Gamma = L(\Phi|C)$ there exists an open neighborhood of C, $W \subset \mathrm{clos}(W) \subset U$ such that for every $x \in W$, $\gamma^+(x) \cap V \neq \emptyset$ and $\gamma^-(x) \cap \Phi_T(V) \neq \emptyset$. Therefore it suffices to prove the lemma with $x \in V$ and $y \in \Phi_T(V)$.

By Lemma 7.4 there exists a (δ, T) pseudo-orbit from x to $\Phi_{-T}(y)$:

$$\{\Phi_t(x_i) : 0 \leq t \leq t_i\}, \quad i = 1, \dots, k-1, t_i > T, x_k = \Phi_{-T}(y),$$

having a length less than T_0 and such that $\sum_i d(\Phi_{t_i}(x_i), x_{i+1})^s \leq \eta$. To the sequence $\{x_i : i = 1, \dots, k\}$ we add the points $x_0 = x$ and $x_{k+1} = \Phi_T(x_k) = y$. Then, define numbers

$$t_0 = t_{k+1} = T \quad \text{and} \quad \tau_i = \sum_{k=0}^{i} t_k$$

for $i = 0, \dots, k+1$. Set $T_1 = T_0 + 2T$.

Let $\Theta : \mathbb{R} \to [0, 1]$ be C^∞ function which is 1 on $(-\infty, 0]$ and zero on $[1, \infty)$. Set $l = \tau_{k+1}$ and define $\psi : [0, l] \to M$ as

$$\psi(s + \tau_i) = \Theta(s/\delta^\alpha)\Phi_{s+t_i}(x_i) + (1 - \Theta(s/\delta^\alpha)\Phi_s(x_{i+1}))$$

for $0 \leq s \leq t_{i+1}, i = 0, \dots, k$.

Since $L(x, F(x)) = 0$ and $\psi(s + \tau_i) = \Phi_s(x_{i+1})$ for $t_{i+1} \geq s \geq \delta^\alpha$, we have

$$\begin{aligned} \int_0^l L(\psi, \psi')\, ds &= \sum_{i=0}^{k} \int_{\tau_i}^{\tau_{i+1}} L(\psi, \psi') = \sum_{i=0}^{k} \int_{\tau_i}^{\tau_i + \delta^\alpha} L(\psi, \psi') \\ &\leq \sum_{i=0}^{k} \delta^\alpha \sup_{0 \leq s \leq \delta^\alpha} L(\psi(s + \tau_i), \psi'(s + \tau_i)). \end{aligned} \tag{28}$$

Therefore, to estimate $\int_0^l L(\psi, \psi')$ it suffices to estimate $\sup_{0 \leq s \leq \delta^\alpha} L(\psi(s + \tau_i), \psi'(s + \tau_i))$. This is our next goal. Let $0 \leq s \leq \delta^\alpha$. A simple computation yields

$$\begin{aligned} &\|\psi'(s + \tau_i) - F(\psi(s + \tau_i))\| \\ &\quad \leq \|\psi'(s + \tau_i) - F(\Phi_s(x_{i+1}))\| + \|F(\Phi_s(x_{i+1})) - F(\psi(s + \tau_i))\| \\ &\quad \leq \frac{1}{\delta^\alpha}\|\Theta'\|\|\Phi_s(\Phi_{t_i}(x_i)) - \Phi_s(x_{i+1})\| + \|\Theta\|\|F(\Phi_s(\Phi_{t_i}(x_i))) - F(\Phi_s(x_{i+1}))\| \\ &\quad\quad + \|\Theta\|\,\mathrm{Lip}(F)\|\Phi_s(\Phi_{t_i}(x_i)) - \Phi_s(x_{i+1})\|. \end{aligned}$$

A standard application of Gronwall's inequality shows that $x \to \Phi_s(x)$ is Lipschitz in x locally uniformly in s. Thus, from the preceding inequality, we easily obtain

$$\|\psi'(s + \tau_i) - F(\psi(s + \tau_i))\| \leq \left(\frac{A}{\delta^\alpha} + B\right) d(\Phi_{t_i}(x_i), x_{i+1}) \tag{29}$$

https://doi.org/10.1017/S0143385798097557 Published online by Cambridge University Press

for some constants $A, B > 0$.

Using inequality (29), we see that

$$\|\psi'(s+\tau_i) - F(\psi(s+\tau_i))\| = O(\delta^{1-\alpha}).$$

Therefore, it follows from Lemma 7.5 that for δ small enough

$$L(\psi(s+\tau_i), \psi'(s+\tau_i)) = O(\|\psi'(s+\tau_i) - F(\psi(s+\tau_i))\|^2) = O\left(\frac{d(\Phi_{t_i}(x_i), x_{i+1})^2}{\delta^{2\alpha}}\right).$$

Thus inequality (28) transforms (for δ small enough) into

$$\begin{aligned}\int_0^l L(\psi, \psi') &= O\left(\sum_i \frac{d(\Phi_{t_i}(x_i), x_{i+1})^2}{\delta^{\alpha}}\right)\\ &= O\left(\sum_i d(\Phi_{t_i}(x_i), x_{i+1})^{2-\alpha}\right) = O\left(\sum_i d(\Phi_{t_i}(x_i), x_{i+1})^{s}\right) = O(\eta).\end{aligned}$$

□

LEMMA 7.7. *Let $K \subset X$ be a compact invariant set and let $K' \subset X$ be a compact set disjoint from K. Assume that for all $\delta > 0$ there exists $x_\delta \in N_\delta(K)$ such that the forward trajectory $\gamma^+(x_\delta)$ meets $N_\delta(K')$. Then there exist $x \in K$ and $y \in K'$ such that $x \hookrightarrow y$.*

Proof. Let $y_\delta = \Phi_{t_\delta}(x_\delta)$ be such that $y_\delta \in N_\delta(K')$. Let x be a limit point of x_δ and let y be a limit point of y_δ (as $\delta \to 0$). Clearly, $t_\delta \to \infty$ otherwise there would exist a subsequence $t_{\delta_i} \to T$ and this would imply $y = \Phi_T(x) \in K'$. This is impossible because K is invariant disjoint from K'. The property $t_\delta \to \infty$ shows that it is possible to find a (η, T) pseudo-orbit from x to y for arbitrary small η and arbitrary large T. Thus $x \hookrightarrow y$. □

From now on, we will assume that the hypotheses of Theorem 3.7 hold. We use the notation $\Lambda_1 = C$, $\Lambda_2 = C^+$, $\Lambda_3 = \mathcal{R}(\Phi) \setminus (C \cup C^+)$, and $U_i = N_r(\Lambda_i)$, $i = 1, 2, 3$. We fix $r > 0$ small enough so that:

- $\overline{U_i}$, $i = 1, 2, 3$, are disjoint compact neighborhoods.

Let $(i, j) \in \{(2, 1), (1, 3), (2, 3)\}$. Since the relation $\hookrightarrow$ is closed we can suppose that:

- for all $y \in \overline{U_j} x \not\hookrightarrow y$;

and according to Lemma 7.7 we can also suppose that:

- for all $x \in \overline{U_i}$, $\gamma^+(x) \cap \overline{U_j} = \emptyset$.

For $i = 1, 2, 3$, $V_i \subset U_i$ denotes a open neighborhood of Λ_i and we set $V = V_1 \cup V_2 \cup V_3$.

LEMMA 7.8. *Given $\eta > 0$, there exists a neighborhood of Λ_1, $W_1 \subset U_1$ such that for any choice of $V_1 \subset W_1$, $V_2 \subset U_2$, $V_3 \subset U_3$ one has*

$$\liminf_{\epsilon \to 0} \epsilon \log(\tilde{P}^{\epsilon}_{x,\theta}(n(\epsilon), V_2 \times \mathcal{X})) \geq -\eta$$

uniformly in $x \in V_1$, $\theta \in \mathcal{X}$, where $\{n(\epsilon)\}$ is a sequence of integers depending on the choice of V.

Proof. Lemma 7.6 applied with $U = U_1$ gives a neighborhood $W_1 \subset U_1$ and a positive constant T_1. We can always suppose that W_1 is compact. Let $V_1 \subset W_1$. By Lemma 5.4 there exists $y \in \partial V_1$ with $\omega(y) \in C^+$. Therefore there exists $t_0 > 0$ such that $\Phi_t(y) \in V_2$ for all $t \geq t_0$. Let $x \in V_1$. By Lemma 7.6 there is a C^1 function $\Psi : [0, l] \to U_1, l \leq T_1$ with $\psi(0) = x, \psi(l) = y$ and $\mathcal{L}_{x,l}(\psi) \leq \eta$.

Define $\tilde{\psi}$ by $\tilde{\psi}(t) = \psi(t)$ for $0 \leq t \leq l$ and $\tilde{\psi}(t) = \Phi_{t-l}(y)$ for $l \leq l'$, where l' is chosen such that the amount of time spent by $\{\tilde{\psi}(t)\}_{0 \leq t \leq l'}$ in $V_1 \cup V_2$ equals $T_1 + 1$. Since $\tilde{\psi}$ coincides with a Φ trajectory for $t \geq l$, $\mathcal{L}_{x,l'}(\tilde{\psi}) = \mathcal{L}_{x,l}(\psi) \leq \eta$.

Now, it is easy to see that there exists $n(\epsilon) \in \mathbb{N}$ (of the order of $(T_1+1)/\epsilon$) such that $X^\epsilon_{n(\epsilon)} \in V_2$ as soon as $\|x^\epsilon(\cdot) - \tilde{\psi}\|_{[0,l']}$ is small enough. Thus, by Theorem 6.2

$$\liminf_{\epsilon \to 0} \epsilon \log(\tilde{P}^\epsilon_{x,\theta}(n(\epsilon), V_2 \times \mathcal{X})) \geq -\mathcal{L}_{x,l'}(\tilde{\psi}) \geq -\eta$$

uniformly in $x \in V_1$, $y \in \mathcal{X}$. □

LEMMA 7.9. *There exists $b > 0$ such that for any choice of $V_i \subset U_i$*

$$\limsup_{\epsilon \to 0} \epsilon \log(\tilde{P}^\epsilon_{x,\theta}(n(\epsilon), V_j \times \mathcal{X})) \leq -b$$

uniformly in $x \in V_i, \theta \in \mathcal{X}$, for all $(i, j) \in \{(2, 1), (1, 3), (2, 3)\}, n \in \mathbb{N}$.

Proof. By the choice of the neighborhoods U_i and by Lemma 6.5 there exists $\beta > 0$ such that $\mathcal{L}(\overline{U_i}, \overline{U_j}) \geq \beta$ for all $(i, j) \in \{(2, 1), (1, 3), (2, 3)\}$.

Suppose the neighborhoods V_i, $i = 1, 2, 3$, are given. Since (by assumption) the measures $\mu^i, i \in \mathcal{X}$ have bounded supports, there exist neighborhoods of Λ_i, $V_i' \subset V_i$ and $\epsilon_0 > 0$ such that if $\epsilon \leq \epsilon_0$, $x \in V_i'$ and $\|x - y\| \leq \epsilon\kappa$, then y is necessarily in V_i. Set $V' = V_1' \cup V_2' \cup V_3'$. The previous discussion proves that if $\{x^\epsilon(t) : t \geq 0\}$ meets V', then $\{X^\epsilon_n\}_{n \geq 0}$ meets V.

Let $(i, j) \in \{(2, 1), (1, 3), (2, 3)\}$ and $x \in V_i$. We have,

$$\{X^\epsilon_{T_V} \in V_j\} \subset \{x^\epsilon(t) \in V_j \text{ for some } t \leq T\} \cup \{x^\epsilon(t) \notin V' \text{ for } 1 \leq t \leq T\}.$$

Thus, by Theorem 6.2, we get

$$\limsup_{\epsilon \to 0} \epsilon \log(\mathbf{P}_{x,\theta}(X^\epsilon_{T_V} \in V_j)) \leq -\inf(\beta, \alpha(T)),$$

where $\alpha(T) = \inf\{\mathcal{L}_{h(0),T}(h) : h \in \mathcal{H}_T\}$ with $\mathcal{H}_T = \{h \in C([0, T]) : h \in C[0, T], h(0) \in \text{clos}(U_i), h([0, T]) \cap V' = \emptyset\}$.

By Lemma 6.3 there exist $r > 0, \alpha > 0$ such that $\mathcal{L}(\overline{U_i}, B_r^c) \geq \alpha$ for all $i \in \{1, 2, 3\}$. Let $\mathcal{H}_T = \mathcal{H}_{1,T} \cup \mathcal{H}_{2,T}$, where $\mathcal{H}_{1,T} = \{h \in \mathcal{H}_T : h([0, T]) \subset \text{clos}(B_r)\}$, $\mathcal{H}_{2,T} = (\mathcal{H}_T - \mathcal{H}_{1,T})$ and let $\alpha_{l,T} = \inf\{\mathcal{L}_{h(0),T}(h) : h \in \mathcal{H}_{l,T}\}$ for $l = 1, 2$.

By Lemma 6.4 it is possible to choose T such that $\alpha_{1,T} \geq \alpha$, and since $\alpha_{2,T} \geq \alpha$ we get $\alpha(T) \geq \alpha > 0$. The lemma is proved with $b = \inf\{\beta, \alpha\}$ for $n = 1$. By the Chapman–Kolmogorov formula we extend it to $n \in \mathbb{N}$. □

Putting together Lemmas 7.9 and 7.3 proves Proposition 7.1.

REFERENCES

[1] E. Akin. *The General Topology of Dynamical Systems (Graduate Studies in Mathematics vol. 1)*. American Mathematical Society, Providence, RI, 1993.

[2] B. Arthur, Y. Ermol'ev and Y. Kaniovskii. A generalized urn problem and its applications. *Cybernetics* **19** (1983), 61–71.

[3] M. Benaïm. A dynamical system approach to stochastic approximations. *Siam J. Control and Optimisation* **34**(2) (1996), 437–472.

[4] M. Benaïm and M. W. Hirsch. Chain recurrence in surface flows. *Discrete and Continuous Dynamic Systems* **1** (1995), 1–17.

[5] M. Benaïm and M. W. Hirsch. Dynamics of Morse–Smale urn processes. *Ergod. Th. & Dynam. Sys.* **15** (1995), 1005–1030.

[6] M. Benaïm and M. W. Hirsch. Asymptotic pseudo-trajectories, chain-recurrent flows and stochastic approximations. *J. Dynamics Diff. Eqns*. **8**(1) (1996), 141–176.

[7] A. Benveniste, M. Métivier and P. Priouret. *Stochastic Approximations and Adaptive Algorithms*. Springer, Berlin, Heidelberg and New York, 1990. Translated from *Algorithmes Adaptatifs et Approximations Stochastiques*. Masson, Paris, 1987.

[8] R. Bowen. ω-Limit sets for Axiom A diffeomorphisms. *J. Diff. Eqns* **18** (1975), 333–339.

[9] O. Brandiere and M. Duflo. Les algorithmes Stochastiques Contournent ils les Pieges? *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* **32**(3) (1996), 395–427.

[10] C. C. Conley. *Isolated Invariant Sets and the Morse Index (Regional conference series in mathematics, no. 38).* American Mathematical Society, 1978.

[11] M. Duflo. *Méthodes Récursives Aléatoires*. Masson, Paris, 1990.

[12] M. Duflo. *Algorithmes Stochastiques (Mathématiques et Applications 23)*. Springer, 1996.

[13] P. Dupuis. Large deviation analysis of some recursive algorithms with state depedent noise. *Annals of Probability* **16**(4) (1988), 1509–1536.

[14] R. Ellis. *Entropy, Larges Deviations and Statistical Mechanics*. Springer, New York, Berlin, Heildelberg, Tokyo, 1985.

[15] J. Franke and J. Selgrade. Abstract ω-limit sets, chain recurrent sets, and basic sets for flows. *Proc. Amer. Math. Soc.* **60** (1976), 309–316.

[16] J. Franke and J. Selgrade. Hyperbolicity and chain recurrence. *J. Diff. Eqns.* **26** (1977), 27–36.

[17] M. I. Freidlin and A. D. Wentzell. *Random Perturbations of Dynamical Systems*. Springer, New York, 1984.

[18] B. M. Hill, D. Lane and W. Sudderth. A strong law for some generalized urn processes. *Annals of Probability* **8** (1980), 214–226.

[19] M. W. Hirsch and C. C. Pugh. Stable manifolds for hyperbolic sets. *Proc. Symp. Pure Math. (Global Analysis, 14)*. American Mathematical Society, Providence, RI, pp. 133–163.

[20] M. W. Hirsch and C. C. Pugh. Cohomology of chain recurrent sets. *Ergod. Th. & Dynam. Sys.* **8** (1988), 73–80.

[21] M. Hurley. Bifurcation and chain recurrence. *Ergod. Th. & Dynam. Sys.* **3** (1983), 231–240.

[22] M. Iosifescu. *Finite Markov Chains and their Applications*. Wiley, Chichester, New York, Brisbane, Toronto, 1980.

[23] T. Kato *Perturbation Theory for Linear Operators*. Springer, Berlin, Heidelberg, New York, 1966.

[24] Y. Kifer. *Random Perturbation of Dynamical Systems*. Birkhauser, Boston, 1985.

[25] H. J. Kushner and D. S. Clark. *Stochastic Approximation Methods for Constrained and Unconstrained Systems*. Springer, Berlin, Heildelberg, New York, 1978.

[26] H. J. Kushner. *Approximation and Weak Convergence Methods for Random Processes*. The MIT Press, Cambridge, MA, 1985.

[27] R. Pemantle. Nonconvergence to unstable points in urn models and stochastic approximations. *Annals of Probability* **18** (1990), 698–712.

[28] C. Robinson. Stability theorems and hyperbolicity in dynamical systems. *Rocky J. Math.* **7** (1977), 425–434.

[29] D. Ruelle. Small random perturbations of dynamical systems and the definition of attractors. *Comm. Math. Phys.* (1981), 137–151.

[30] S. Smale. Differentiable dynamical systems. *Bull. Amer. Math. Soc.* **73** (1967), 747–817.

[31] F. W. Wilson. The structure of level surfaces of a Lyapounov function. *J. Diff. Eqns.* **3** (1967), 323–329.