Convergence with probability one of stochastic approximation algorithms whose average is cooperative

Article *in* Nonlinearity · February 2000

DOI: 10.1088/0951-7715/13/3/305

1 author:

Michel Benaim
Université de Neuchâtel

Convergence with probability one of stochastic approximation algorithms whose average is cooperative

2000 Nonlinearity 13 601

(http://iopscience.iop.org/0951-7715/13/3/305)

Nonlinearity **13** (2000) 601–616. Printed in the UK PII: S0951-7715(00)01965-4

Convergence with probability one of stochastic approximation algorithms whose average is cooperative

Michel Benaïm
Departement de Mathématiques, Université Cergy-Pontoise, France

Received 18 February 1999, in final form 29 September 1999
Recommended by L Bunimovich

Abstract. We consider a stochastic approximation process

$$x_{n+1} - x_n = \gamma_{n+1}(F(x_n) + U_{n+1})$$

where $F : \mathbb{R}^m \to \mathbb{R}^m$ is a C^2 irreducible cooperative dissipative vector field, $\{\gamma_n\}_{n\geqslant 0}$ is a sequence of positive numbers decreasing to 0 and $\{U_n\}_{n\geqslant 0}$ a sequence of uniformly bounded $\mathbb{R}^m$ martingale differences. We show that under certain conditions on $\{\gamma_n\}$ and $\{U_n\}$ the sequence $\{x_n\}_{n\geqslant 0}$ converges with probability one toward the equilibria set of the vector field F.

AMS classification scheme numbers: 62L20, 34Fxx, 34Cxx

1. Introduction

We consider a *stochastic approximation (SA) algorithm* defined on $\mathbb{R}^m$ by

$$x_{n+1} - x_n = \gamma_{n+1}(F(x_n) + U_{n+1}) \qquad n \geqslant 0 \tag{1}$$

where $\{\gamma_n\}$ is a sequence of non-negative step sizes,

$$F : \mathbb{R}^m \to \mathbb{R}^m$$

$$u = (u^1, \ldots, u^m) \to (F_1(u), \ldots, F_m(u))$$

is a smooth vector field and $\{U_n\}_{n\geqslant 0}$ is a sequence of $\mathbb{R}^m$-valued random variables.

Throughout the paper the standing assumptions are hypotheses 1.1 and 1.2 below.

Hypothesis 1.1. *F is a C^2, cooperative irreducible and dissipative vector field.*

Cooperative means

$$\frac{\partial F_i}{\partial u^j}(u) \geqslant 0 \qquad \text{for all} \quad u \in \mathbb{R}^m \quad \text{and} \quad i \neq j$$

and *irreducible* means that the Jacobian matrix $DF(x)$ is irreducible for all $x \in \mathbb{R}^m$.

The (local) *solution flow* induced by F is denoted by $\Phi = \{\Phi_t\}$ where $t \to \Phi_t(x)$ is the solution to $\dot{y} = F(y)$ with initial condition $y(0) = x$. The assumption that F is *dissipative* means that there exists a ball $B \subset \mathbb{R}^m$ with the property that for every compact set $K \subset \mathbb{R}^m$ there exists $T > 0$ such that $\Phi_t(K) \subset B$ for all $t \geqslant T$. The set

$$\Lambda = \bigcap_{t\geqslant 0} \Phi_t(B)$$

0951-7715/00/030601+16$30.00

is a compact invariant set called the *global attractor* of F.

Invariant and empirical occupation measures of SA associated with a cooperative vector field have been considered recently by Benaïm and Hirsch (1999a) for processes with *constant step sizes* (i.e. $\gamma_n = \gamma$).

The purpose of this paper is to study the limiting behaviour of (1) for processes with *decreasing step sizes.* This is motivated by the study of certain adaptive processes arising in game theory (see, e.g., Benaïm and Hirsch 1999b) and neural networks (Fort and Pages 1995, Sadeghi 1998, Benaïm *et al* 1998).

Our next assumption is classical in the stochastic approximation literature (see, e.g., Kushner and Yin 1997, Duflo 1996, 1997, Benaïm 1999).

Here and throughout, the Euclidean norm of $x \in \mathbb{R}^m$ is denoted by $\|x\|$; the open ball about x of radius δ is denoted by $B_\delta(x)$.

Hypothesis 1.2. *(a) F is Lipschitz and bounded on a neighbourhood of the set* $\{x_n : n \geqslant 0\}$. *That is, there exists $\delta > 0$ such that*

$$\sup\left\{\|F(x)\| : x \in \bigcup_{n\geqslant 0} B_\delta(x_n)\right\}$$

and

$$\sup\left\{\frac{\|F(x) - F(y)\|}{\|x-y\|} : x, y \in \bigcup_{n\geqslant 0} B_\delta(x_n),\ x \neq y\right\}$$

are almost certainly finite.

(b)
$$\sum_n \gamma_n = \infty.$$

(c) For all $c > 0$,

$$\sum_n \mathrm{e}^{-c/\gamma_n} < \infty.$$

(d) $\{U_n\}$ is a bounded martingale difference sequence.

Condition (d) means that the $\{U_n\}$ are defined on a probability space $(\Omega, \mathcal{F}, P)$ and that there is a non-decreasing sequence of the sub-sigma field $\{\mathcal{F}_n\}$ such that

- U_n is $\mathcal{F}_n$ measurable and $E(U_{n+1}|\mathcal{F}_n) = 0$, where $E(\cdot|\mathcal{F}_n)$ stands for the conditional expectation with respect to $\mathcal{F}_n$.
- There exists a constant $\Gamma > 0$ such that $\sup_n \|U_n\| \leqslant \Gamma$ almost certainly.

Our main result asserts that (under suitable conditions on F) sample paths of (1) converge with probability one toward equilibria of F (theorem 2.1). Under a milder assumption on F, we can estimate the probability of convergence toward a sink (theorem 2.5). This is obtained by combining several recent results both from the theory of stochastic approximation and the theory of cooperative dynamical systems.

General results are stated in section 2 and proved in section 3. Section 4 improves the main results for certain classes of cooperative vector fields.

2. Statement of the main results

Let $\mathcal{E}$ denote the *equilibrium set* of F; that is

$$\mathcal{E} = \{p \in \mathbb{R}^m \ : F(p) = 0\}.$$

We let $\lambda_1(p) \geqslant \cdots \geqslant \lambda_m(p)$ denote the real parts of the eigenvalues of the Jacobian matrix $DF(p)$.

As usual an equilibrium p is said to be *linearly unstable* if $\lambda_1(p) > 0$. We let $\mathcal{E}_{lu}$ denote the set of linearly unstable equilibria.

A point $p \in \mathcal{E}$ is a *sink* if $\lambda_1(p) < 0$.

A *non-degenerate arc of equilibria* $J \subset \mathcal{E}$ is the image of the closed unit interval under an injective C^1 immersion $h : [0, 1] \to \mathbb{R}^m$ such that $J = h([0, 1]) \subset \mathcal{E}$. We say that J is *ordered* if $h'(t)$ has positive entries for all $t \in [0, 1]$. A *degenerate arc of equilibria* is an equilibrium.

We let $\mathcal{M}_{erg}(\Phi)$ denote the set of Borel probability measures which are invariant and ergodic for Φ. For $\mu \in \mathcal{M}_{erg}(\Phi)$ we let $\lambda_1(\mu) \geqslant \lambda_2(\mu) \ldots$ denote the Lyapounov exponents of Φ with respect to μ.

The limit set of the sequence $\{x_n\}$, denoted by $L(\{x_n\})$, is the set of points $p \in \mathbb{R}^m$ such that $\lim_{k\to\infty} x_{n_k} = p$ for some sequence $n_k \to \infty$. We write $L(\{x_n\}) = \{\infty\}$ if $\lim_{n\to\infty} \|x_n\| = \infty$.

Convergence with probability one

Given a compact set $K \subset \mathbb{R}^m$ we say that process (1) is *non-degenerate* at K if there exists a neighbourhood $\mathcal{N}$ of K and $b > 0$ such that for all unit vectors $v \in \mathbb{R}^m$

$$E(\sup(\langle U_{n+1}, v\rangle, 0)|\mathcal{F}_n,\ x_n \in \mathcal{N}) \geqslant b.$$

Theorem 2.1. *Assume:*

(i) The process (1) is non-degenerate at Λ (the global attractor of F).
(ii) There exists $\frac{1}{2} < \alpha \leqslant 1$ such that

$$\sup\{(1+\alpha)\lambda_2(\mu) - \lambda_1(\mu) :\ \mu \in \mathcal{M}_{erg}(\Phi)\} < 0.$$

Then with probability one either $L(\{x_n\}) = \infty$; or $L(\{x_n\})$ is an ordered arc (possibly degenerate) contained in $\mathcal{E} \setminus \mathcal{E}_{lu}$. If F is real analytic, then with probability one either $L(\{x_n\}) = \infty$; or $L(\{x_n\}) = p \in \mathcal{E} \setminus \mathcal{E}_{lu}$.

Remark 2.2. The verification of condition (ii) in theorem 2.1 requires the computation of the quantity $\alpha(\Phi, r) = \sup_\mu(r\lambda_2(\mu) - \lambda_1(\mu))$ where the supremum is taken over all ergodic measures. Estimates of $\alpha(\Phi, r)$ are given in Benaïm (1997), in terms of the entries of DF.

Conjecture 2.3. *The conclusion of theorem 2.1 holds whether or not condition (ii) is satisfied.*

Convergence with positive probability

In addition to hypothesis (1.1) we shall assume in this section that F satisfies the following hypothesis.

Hypothesis 2.4. *(a) F is (globally) Lipschitz and bounded.*
(b) All equilibria are simple:

$$\forall p \in \mathcal{E}\ \mathrm{Det}(DF(p)) \neq 0.$$

We let Leb denote the Lebesgue measure on $\mathbb{R}^m$. We also write $Leb(A) = \int_A \mathrm{d}x$ for any Borel set $A \subset \mathbb{R}^m$.

Theorem 2.5. *Let $M \subset \mathbb{R}^m$ be a compact neighbourhood of Λ positively invariant under Φ. For all $\epsilon > 0$ there exists a compact set $M_\epsilon \subset M$, positively invariant under Φ, and $\delta = \delta(\epsilon) > 0$ such that*

(a) $Leb(M \setminus M_\epsilon) \leqslant \epsilon$.
(b) $M_{\epsilon_2} \subset M_{\epsilon_1}$ *for* $\epsilon_2 \leqslant \epsilon_1$, $\bigcup_{\epsilon>0} M_\epsilon = M$.
(c) Let L denote the Lipschitz constant of F, $B = \mathrm{e}^{-2L}/8m\Gamma^2$ and

$$k_0 = \inf \left\{ k \in \mathbb{N} : \sup_{j \geqslant k} \gamma_j \leqslant \tfrac{1}{2} B\delta^2 \right\}.$$

Then for all $k \geqslant k_0$

$$P(L(\{x_n\}) \textit{ is a sink } |\mathcal{F}_k) \geqslant \left[1 - 2m \sum_{j=k}^{\infty} \gamma_j \exp\left(\frac{-B\delta^2}{\gamma_j} \right) \right] P(\exists m \geqslant k : x_m \in M_\epsilon | \mathcal{F}_k).$$

Corollary 2.6. *Let M_ϵ, δ and k_0 be as in theorem 2.5. Then*

(a) $\forall k \geqslant k_0$

$$P(L(\{x_n\}) \textit{ is a sink } | x_k \in M_\epsilon) \geqslant \left[1 - 2m \sum_{j=k}^{\infty} \gamma_j \exp\left(\frac{-B\delta^2}{\gamma_j} \right) \right].$$

(b)
$$\lim_{k\to\infty} P(L(\{x_n\}) \textit{ is a sink } | x_k \in M_\epsilon) = 1.$$

Suppose now that the sequence $\{\gamma_n\}$ depends on a parameter $\theta \in \mathbb{R}^p$. That is $\gamma_n = \gamma_n^\theta$. Suppose, furthermore, that

(i) $\gamma_n^\theta \leqslant \gamma_n^{\theta'}$ for $\theta \leqslant \theta'$,
(ii) $\lim_{\theta\to 0}(\sup_{n\in\mathbb{N}} \gamma_n^\theta) = 0$.

This situation occurs, for example, in certain problems of game theory for which $\theta \in \mathbb{R}_+$ and

$$\gamma_n^\theta = \frac{1}{n + 1/\theta}.$$

Let $\{x_n^\theta\}$ denote the sequence defined by (1) with γ_n^θ instead of γ_n.

Corollary 2.7. *Let M_ϵ be as in theorem 2.5. Then*

$$\lim_{\theta\to 0} P(L(\{x_n^\theta\}) \textit{ is a sink } | x_0^\theta \in M_\epsilon) = 1.$$

Proof. Let $\theta_0 \in \mathbb{R}^p$. For $\theta \leqslant \theta_0$

$$\sum_{j\geqslant 0} \gamma_j^\theta \exp\left(\frac{-b\delta^2}{\gamma_j^\theta} \right) \leqslant \sup_j \gamma_j^\theta \sum_{j\geqslant 0} \exp\left(\frac{-b\delta^2}{\gamma_j^{\theta_0}} \right).$$

Thus, corollary 2.6 implies the result. □

Corollary 2.8. *Let μ be a probability measure on $\mathbb{R}^m$ absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^m$. Suppose*

(i) x_0^θ is a random variable having μ as probability distribution.
(ii) The Radon–Nikodym derivative $\mathrm{d}\mu/\mathrm{d}x$ is in $L^p(\mathrm{d}x)$ for some $p \geqslant 1$.

Then

$$\lim_{\theta\to 0} P(L(\{x_n^\theta\}) \textit{ is a sink}) = 1.$$

Proof. Let $\eta > 0$. Choose a compact set $M \subset \mathbb{R}^m$ positively invariant under Φ such that $\mu(M) \geqslant 1 - \eta$.

Let $M_\epsilon \subset M$ be as in theorem 2.5. By the Holder inequality

$$|\mu(M) - \mu(M_\epsilon)| = \left| \int_{M\setminus M_\epsilon} \frac{\mathrm{d}\mu}{\mathrm{d}x}(x)\,\mathrm{d}x \right| \leqslant \left\| \frac{\mathrm{d}\mu}{\mathrm{d}x} \right\|_p Leb(M \setminus M_\epsilon)^{1-1/p}$$

where

$$\left\| \frac{\mathrm{d}\mu}{\mathrm{d}x}\,\mathrm{d}x \right\|_p = \left(\int_{\mathbb{R}^m} \left| \frac{\mathrm{d}\mu}{\mathrm{d}x} \right|^p \right)^{1/p}.$$

Thus by theorem 2.5(a)

$$|\mu(M) - \mu(M_\epsilon| \leqslant \left\| \frac{\mathrm{d}\mu}{\mathrm{d}x} \right\|_p \epsilon^{(1-1/p)}.$$

Therefore,

$$\begin{aligned} P(L(\{x_n^\theta\}) \text{ is a sink}) &= P(L(\{x_n^\theta\}) \text{ is a sink } |x_0^\theta \in M_\epsilon)\mu(M_\epsilon) \\ &\geqslant P(L(\{x_n^\theta\}) \text{ is a sink } |x_0^\theta \in M_\epsilon)\left(1 - \eta - \left\| \frac{\mathrm{d}\mu}{\mathrm{d}x} \right\|_p \epsilon^{(1-1/p)}\right). \end{aligned}$$

Using corollary 2.7, we obtain

$$\liminf_{\theta\to 0} P(L(\{x_n^\theta\}) \text{ is a sink}) \geqslant \left(1 - \eta - \left\| \frac{\mathrm{d}\mu}{\mathrm{d}x} \right\|_p \epsilon^{1-1/p}\right).$$

Since η and ϵ are arbitrary, this proves the result. □

3. Proof of the results

The general idea of the proof is the following. On one hand, the limit sets of stochastic approximation processes are known to be *internally chain recurrent* with probability one (Benaïm 1996). On the other hand, an internally chain-recurrent set for a strongly monotonic flow either consists of equilibria or is contained in a smooth unordered hypersurface (Hirsch 1999). As such a hypersurface is 'repelling' in a positive direction, we use a recent result (Benaïm 1999) adapted from Pemantle (1990), to prove that the process $\{x_n\}$ converges toward the equilibria set.

Background and preliminary results

Let Ψ be a flow on a metric space X. A point $p \subset X$ is called *chain recurrent* for Ψ (Conley 1978) if for all $\epsilon, T > 0$ there exist points $x_0, x_1, \ldots, x_n$ in X and times $t_i > T$ such that $x_0 = x_n = x$ and

$$d(\Phi_{t_i}(x_i), x_{i+1}) < \epsilon \qquad i = 0, n-1.$$

The flow Ψ is called a chain-recurrent flow if every point $p \in X$ is chain recurrent.

A set $L \subset \mathbb{R}^m$ is said to be *internally chain recurrent* for Φ if it is a non-empty compact connected invariant set such that the restricted flow $\Phi|L$ is a chain-recurrent flow. We let $\mathcal{R}(\Phi)$ denote the set of chain-recurrent points for Φ. As the flow Φ is dissipative, it follows from Conley (1978) that $\mathcal{R}(\Phi)$ is internally chain recurrent for Φ.

The following result follows from proposition 4.4 and corollary 6.11 of Benaïm (1999) based on Benaïm (1996).

Theorem 3.1. *Suppose F is dissipative (not necessarily cooperative) and hypothesis 1.2 holds. Then with probability one either $L(\{x_n\}) = \infty$ or $L(\{x_n\})$ is internally chain recurrent.*

For cooperative vector fields, internally chain-recurrent sets enjoy strong topological properties that we now describe.

The vector order in $\mathbb{R}^m$ is written as $x \geqslant y$ with the meaning that $x_i \geqslant y_i$ for all i. If $x \geqslant y$ and $x \neq y$ we write $x > y$. If $x_i > y_i$ for all i, then we write $x \gg y$. As the vector field F is cooperative and irreducible, the flow Φ has *positive derivatives* (Hirsch 1985, Smith 1995). That is

$$D\Phi_t(x) \gg 0$$

for all $x \in \mathbb{R}^m$ and $t > 0$. In particular, Φ is *strongly monotonic*:

$$x > y \quad \Rightarrow \quad \Phi_t(x) \gg \Phi_t(y)$$

for $t > 0$.

A set $A \subset \mathbb{R}^m$ is said to be *unordered* if no two of its points are related by $>$.

An equilibrium p is said to be *asymptotically stable from below* if there exists $x \ll p$ such that $\lim_{t\to\infty} \Phi_t(x) = p$. We let $\mathcal{E}_{asb} \subset \mathcal{E}$ denote the set of equilibria which are asymptotically stable from below. By strong monotonicity if $p \in \mathcal{E}_{asb}$ then there is a non-empty open set of points whose forward trajectories converge toward p from below. It follows that $\mathcal{E}_{asb}$ is a countable set.

Given $p \in \mathcal{E}_{asb}$ let $V(p) = \{x \in \mathbb{R}^m : \omega(x) \geqslant p\}$. The following proposition summarizes some results by Hirsch (1988), Tereščák (1996) and others.

Proposition 3.2. *There exists a unique equilibrium $p_* \in \mathcal{E}_{asb}$ such that $V(p_*) = \mathbb{R}^m$. If $p \in \mathcal{E}_{asb} \setminus \{p_*\}$ then*

(a) *$S_p = \partial V(p)$ is a closed invariant unordered hypersurface.*

(b) *Let $\pi : \mathbb{R}^m \to E$ denote the orthogonal projection onto the hyperplane perpendicular to a vector $e \gg 0$. Then $\pi|S_p$ maps S_p homeomorphically onto E and the map $(\pi|S_p)^{-1}$ is C^1.*

(c) *Let $K \subset S_p$ be a compact invariant set. For each $x \in K$ there exist a unit vector $b(x) \gg 0$ and a (continuous) splitting*

$$\mathbb{R}^m = T_x S_p \oplus \mathbb{R} b(x)$$

invariant by $D\Phi_t$, such that

$$\|D\Phi_t(x)w\| \leqslant C\mathrm{e}^{-\eta t}\|D\Phi_t(x)b(x)\|$$

for every unit vector $w \in T_x S_p$. Here C and η are positive constants which can be chosen independent of K.

Equivalently: for every Φ-invariant and ergodic probability measure μ with support in K,

$$\lambda_2(\mu) - \lambda_1(\mu) < -\eta \tag{2}$$

where $\lambda_1(\mu)$ and $\lambda_2(\mu)$ denote, respectively, the largest and the second largest Lyapounov exponents of Φ with respect to μ.

(d) *Let K be as in (c). Suppose, furthermore, that K is an attractor for $\Phi|S_p$ and that there exist $\alpha > 0$ and $\eta' > 0$ (depending on K) such that the inequality (2) can be strengthened to*

$$(1+\alpha)\lambda_2(\mu) - \lambda_1(\mu) < -\eta'. \tag{3}$$

Then the basin of attraction of K in S_p is a $C^{1+\alpha}$ manifold.

Proof. Let $p*$ be an infimum (for the vector ordering) of the set $\mathcal{E}$. That is $p \leqslant p*$ and $p \in \mathcal{E}$ imply $p = p*$. Let $x \in \mathbb{R}^m$. Since almost every trajectory has an omega limit set consisting of equilibria (Hirsch 1985) there exists y smaller than x and $p*$ such that the omega limit set of y consists of equilibria. Therefore, by monotonicity $\omega(y) = p* \leqslant \omega(x)$.

Except for the assertion that $\pi(S_p) = E$ and $(\pi|S_p)^{-1}$ is C^1 the proof of (a) and (b) is similar to Hirsch (1988, theorem 2.1). Smoothness of $(\pi|S_p)^{-1}$ is proved by Tereščák (1996). To see that $\pi(S_p) = E$ let $z \in E$. Then there exist real numbers $s < t$ such that $z + se \ll p \ll z + te$. Thus $\omega(z + se) \leqslant p \leqslant \omega(z + te)$ and there exists $t \leqslant \tau \leqslant s$ such that $z + \tau e \in S_p$. Hence $\pi(z + \tau e) = z$. Assertion (c) follows from the well known *exponential separation property* and has been proved many times (see, e.g., Ruelle 1979, Tereščák 1996, Benaïm 1997). Assertion (d) is proved in Benaïm (1997). □

Given $p \in \mathcal{E}_{asb}$ we let $\mathcal{R}(\Phi|S_p)$ denote the chain-recurrent set of $\Phi|S_p$. The next result follows from Hirsch (1999).

Theorem 3.3. *Let $L \subset \mathbb{R}^m$ be an internally chain recurrent. Then either L is an ordered arc (possibly degenerate) contained in $\mathcal{E} \setminus \mathcal{E}_{lu}$; or there exists $p \in \mathcal{E}_{abs} \setminus \{p_*\}$ such that $L \subset \mathcal{R}(\Phi|S_p)$. When F is real analytic ordered arc of equilibria are degenerates.*

Proof. By theorem 1.6 of Hirsch (1999), L is either an unordered set or a C^1 ordered arc of equilibria. If L is unordered and is not an equilibrium, or L is a linearly unstable equilibrium then by corollary 3.4 (more precisely the proof of corollary 3.4) and proposition 3.5 of Benaïm and Hirsch (1999a), L lies in S_p for some $p \in \mathcal{E}_{abs}$. By a result of Jiang (1991), when F is real analytic, then it cannot have a non-degenerate ordered arc of equilibria. □

Proof of theorem 2.1

Let $p \in \mathcal{E}_{asb} \setminus \{p_*\}$ and let $\mathcal{E}_p^- \subset \mathcal{E}$ denote the set of equilibria $q \in S_p$ satisfying $\lambda_1(q) \leqslant 0$. For $q \in \mathcal{E}_p^-$, the Perron Frobenius theorem (or proposition 3.2(c)) implies that q is an attractor for the restricted flow $\Phi|S_p$ whose basin of attraction is the open subset of S_p defined as

$$B(q, \Phi|S_p) = \left\{x \in S_p : \lim_{t\to\infty} \mathrm{dist}(\Phi_t(x), q) = 0\right\}.$$

Define

$$\mathcal{R}'_p = \mathcal{R}(\Phi|S_p) \setminus \mathcal{E}_p^-.$$

It is not hard to see that

$$\mathcal{R}'_p = \mathcal{R}(\Phi|S_p) \bigcap \left\{S_p \setminus \bigcup_{q\in\mathcal{E}_p^-} B(q, \Phi|S_p)\right\}.$$

This makes $\mathcal{R}'_p$ a compact invariant subset of $\mathcal{R}(\Phi|S_p)$.

As each component of $\mathcal{R}(\Phi)$ is internally chain recurrent (Conley 1978), theorem 3.3 implies that

$$\mathcal{R}(\Phi) = \left\{ \bigcup_{p \in \mathcal{E}_{asb} \setminus p*} \mathcal{R'}_p \right\} \bigcup \mathcal{E} \setminus \mathcal{E}_{lu}. \tag{4}$$

By theorem 3.1 the sequence $\{x_n\}$ converges almost certainly toward $\mathcal{R}(\Phi)$. Therefore, to prove theorem 2.1 we have to show that

$$P\left(\lim_{n\to\infty} \mathrm{dist}\left(x_n, \bigcup_{p \in \mathcal{E}_{asb} \setminus p*} \mathcal{R}'_p \right) = 0 \right) = 0.$$

Since $\mathcal{E}_{asb}$ is countable it suffices to show that for all $p \in \mathcal{E}_{asb} \setminus p*$

$$P(\lim_{n\to\infty} \mathrm{dist}(x_n, \mathcal{R}'_p) = 0) = 0. \tag{5}$$

To prove (5) we use the following general result proved in Benaïm (1999; theorem 9.1) and improved by Tarrès (1999; theorem 3.1), which generalizes a result of Pemantle (1990).

Theorem 3.4. *Let $\{x_n\}$ be a SA process of the form (1) where F is a smooth vector field (not necessarily cooperative). Let K be a compact invariant set and S a smooth locally invariant manifold containing K. Assume that:*

(i) The tangent bundle of $\mathbb{R}^m$ restricted to K splits continuously into two sub-bundles invariant by $D\Phi_t$

$$T_K\mathbb{R}^m = T_K S \oplus L.$$

(ii) There exist positive constants C, β such that for all $x \in K$, $w \in L$ and $t \geqslant 0$

$$\|D\Phi_t(x)w\| \geqslant C\mathrm{e}^{\beta t}\|w\|.$$

(iii) There exists $\frac{1}{2} < \alpha \leqslant 1$ such that F and S are $C^{1+\alpha}$.

(iv) The process (1) is non-degenerate at K.

Then $P(\lim_{n\to\infty} d(x_n, K) = 0) = 0$.

We can now pass to the proof of (5). Set $K = \mathcal{R}'_p$ and $S = S_p$. Assumption (i) of theorem 3.4 follows from proposition 3.2(c).

By definition of K each equilibrium $q \in K$ satisfies $\lambda_1(q) > 0$. Therefore, by corollary 4.3 of Benaïm (1997),

$$\lim_{t\to\infty} \frac{1}{t} \log\left(\sup_{x\in K} \|D\Phi_t(x)b(x)\| \right) > 0.$$

This proves assumption (ii). Assumption (iii) follows from proposition 3.2(d) applied to the global attractor of $\Phi|S_p$, and assumption (iv) is part of the hypotheses. This concludes the proof of theorem 2.1.

Proof of theorem 2.5

The proof of theorem 2.5 uses a different type of argument. Let $\mathcal{E}_{ls} \subset \mathcal{E}$ denote the set of sinks (linearly stable equilibria). For each $p \in S$ let $W(p)$ denote the basin of attraction of p:

$$W(p) = \{x \in \mathbb{R}^m : \omega(x) = \{p\}\}$$

and let

$$W(\mathcal{E}_{ls}) = \cup_{p \in \mathcal{E}_{ls}} W(p).$$

Let $N \subset W(\mathcal{E}_{ls}) \cap M$ denote a fundamental neighbourhood of $\mathcal{E}_{ls}$ (N is compact and $\Phi_t(N) \subset \mathrm{Int}(N)$ for all $t > 0$). By definition of N and $W(\mathcal{E}_{ls})$

$$\bigcup_{t \geqslant 0} \Phi_{-t}(N) = W(\mathcal{E}_{ls}).$$

Thus, for all $\epsilon > 0$

$$T_\epsilon = \inf\{t > 0 : Leb(M \cap W(\mathcal{E}_{ls}) \setminus \Phi_{-t}(N)) \leqslant \epsilon\} < \infty.$$

Set $M_\epsilon = M \cap \Phi_{-T_\epsilon}(N)$.

By theorem 4.4(b) of Hirsch (1985), $Leb(\mathbb{R}^m \setminus W(\mathcal{E}_{ls})) = 0$. Hence $Leb(M \setminus M_\epsilon) \leqslant \epsilon$ proving assertion (a) of theorem 2.5. Assertion (b) follows directly from the definition of M_ϵ.

We now pass to the proof of assertion (c). Set

$$\tau_0 = 0 \qquad \text{and} \qquad \tau_n = \sum_{i=1}^{n} \gamma_i \qquad \text{for} \quad n \geqslant 1.$$

Let $X : \mathbb{R}_+ \to \mathbb{R}^m$ denote the continuous time affine interpolated process defined by

$$X(\tau_n + s) = x_n + s \frac{x_{n+1} - x_n}{\tau_{n+1} - \tau_n}$$

for all $n \in \mathbb{N}$ and $0 \leqslant s \leqslant \gamma_{n+1}$.

For $t \geqslant 0$ define

$$D_X(t) = \sup_{k \in \mathbb{N}} \|\Phi_1(X(t+k)) - X(t+k+1)\|.$$

Given $\epsilon > 0$ choose $\delta > 0$ small enough so that

$$N_\delta(\Phi_1(M_\epsilon)) \subset \mathrm{int}(M_\epsilon) \tag{6}$$

where $N_\delta(A)$ stands for the δ neighbourhood of set A.

We claim that if $X(t) \in M_\epsilon$ and $D_X(t) \leqslant \delta$ then the limit set of $\{x_n\}$ is a sink. Indeed, the assumptions $X(t) \in M_\epsilon$ and $D_X(t) \leqslant \delta$ combined with (6) imply that $X(t+k) \in M_\epsilon$ for all $k \geqslant 0$. Therefore, $L(\{x_n\}) \cap M_\epsilon \neq \emptyset$. Since $L(\{x_n\})$ is internally chain recurrent (theorem 3.1) $L(\{x_n\})$ has to be a sink.

Therefore, for all $k \geqslant 0$

$$\{\exists j \geqslant k \; x_j \in M_\epsilon \text{ and } D_X(\tau_j) \leqslant \delta\} \subset \{L(\{x_n\}) \text{ is a sink}\}. \tag{7}$$

To conclude the proof we shall now estimate the probability of the event

$$\{\exists j \geqslant k \; x_j \in M_\epsilon \text{ and } D_X(\tau_j) \leqslant \delta\}.$$

Lemma 3.5. *Let $J \in \mathbb{N} \cup \{\infty\}$ be a stopping time. For $\delta > 0$ set*

$$k_0(\delta) = \inf \left\{ k \in \mathbb{N} : \sup_{j \geqslant k} \gamma_j \leqslant B\delta^2/2 \right\}.$$

Then for δ small enough

$$P((k_0(\delta) \leqslant J < \infty) \text{ and } D_X(\tau_J) \geqslant \delta | \mathcal{F}_J) \leqslant \mathbf{1}_{\{k_0(\delta) \leqslant J < \infty\}} \int_{\tau_J}^{\infty} r(s, \delta)\, \mathrm{d}s$$

where

$$r(s, \delta) = 2m \exp\left(\frac{-B\delta^2}{\gamma_{k+1}} \right) \qquad \text{for} \quad \tau_k \leqslant s < \tau_{k+1}.$$

Proof. To shorten the notation write $k_0 = k_0(\delta)$,

$$\begin{aligned} P(D_X(\tau_J) \geqslant \delta | \mathcal{F}_J) \mathbf{1}_{\{k_0 \leqslant J < \infty\}} &= \sum_{j \geqslant k_0} P((D_X(\tau_j) \geqslant \delta) \mathbf{1}_{\{J=j\}} | \mathcal{F}_J) \\ &= \sum_{j \geqslant k_0} P((D_X(\tau_j) \geqslant \delta) \mathbf{1}_{\{J=j\}} | \mathcal{F}_j) \\ &= \sum_{j \geqslant k_0} P((D_X(\tau_j) \geqslant \delta) | \mathcal{F}_j) \mathbf{1}_{\{J=j\}}. \end{aligned}$$

Therefore, it suffices to prove that

$$P(D_X(\tau_j) \geqslant \delta | \mathcal{F}_j) \leqslant \int_{\tau_j}^{\infty} r(s, \delta)\, \mathrm{d}s \tag{8}$$

for all $j \geqslant k_0$. By definition of $D_X(\tau_j)$ we have

$$P(D_X(\tau_j) \geqslant \delta | \mathcal{F}_j) \leqslant \sum_{k \in \mathbb{N}} P(\|X(\tau_j + k + 1) - \Phi_1(X(\tau_j + k)\| \geqslant \delta | \mathcal{F}_j).$$

Let $m(t) = \inf\{j \in \mathbb{N} : \tau_j \geqslant t\}$. Then

$$P(D_X(\tau_j) \geqslant \delta | \mathcal{F}_j) \leqslant \sum_{k \in \mathbb{N}} E(P(\|X(\tau_j + k + 1) - \Phi_1(X(\tau_j + k)\| \geqslant \delta | \mathcal{F}_{m(\tau_j + k)} | \mathcal{F}_j). \tag{9}$$

We claim that for all $t \geqslant \tau_{k_0}$

$$P(\|X(t+1) - \Phi_1(X(t))\| \geqslant \delta | \mathcal{F}_{m(t)}) \leqslant \int_t^{t+1} r(s, \delta)\, \mathrm{d}s. \tag{10}$$

Clearly (10) combined with (9) proves (8).

Proof of the claim. Let $\overline{X}, \overline{U}, \overline{\gamma} : \mathbb{R}_+ \to \mathbb{R}^m$ denote the continuous time processes defined by

$$\overline{X} = x_n \qquad \overline{U}(t) = U_{n+1} \qquad \overline{\gamma}(t) = \gamma_{n+1}$$

for $\tau_n \leqslant t < \tau_{n+1}$. Then for all $0 \leqslant h \leqslant 1$ and $t \geqslant 0$

$$X(t+h) - \Phi_h(X(t)) = \int_0^h [F(\overline{X}(t+s)) - F(\Phi_s(X(t))]\, \mathrm{d}s + \int_t^{t+h} \overline{U}(s)\, \mathrm{d}s.$$

F and $\{U_n\}$ being bounded $\|X(t) - \overline{X}(t)\| = O(\gamma(t))$. Hence

$$\|X(t+h) - \Phi_h(X(t))\| \leqslant L \int_0^h \|X(t+s) - \Phi_s(X(t))\| \, ds + C \sup_{t \leqslant s \leqslant t+1} \overline{\gamma}(s) + \Delta(t)$$

where

$$\Delta(t) = \sup_{0 \leqslant h \leqslant 1} \left\| \int_t^{t+h} \overline{U}(s) \, ds \right\|$$

and C is a positive constant ($C = L(\Gamma + \sup_x \|F(x)\|)$). Thus, by Gronwall's inequality we find that

$$\sup_{0 \leqslant h \leqslant 1} \|X(t+h) - \Phi_h(X(t))\| \leqslant e^L \Big(\Delta(t) + C \sup_{t \leqslant s \leqslant t+1} \overline{\gamma}(s)\Big).$$

Since $t \geqslant \tau_{k_0}$, $\sup_{t \leqslant s \leqslant t+1} \overline{\gamma}(s) = O(\delta^2)$. Hence for δ small enough

$$C \sup_{t \leqslant s \leqslant t+1} \overline{\gamma}(s) \leqslant \delta/2.$$

Therefore,

$$P(\|X(t+1) - \Phi_1(X(t))\| \geqslant \delta | \mathcal{F}_{m(t)}) \leqslant P(\Delta(t) \geqslant \tfrac{1}{2}\delta e^{-L} | \mathcal{F}_{m(t)})$$

By a classical application of exponential martingale inequality (inequality (18) in Benaïm (1998)) we have

$$P(\Delta(t) \geqslant \alpha | \mathcal{F}_{m(t)}) \leqslant 2m \exp\left(\frac{-\alpha^2}{2\, m\Gamma^2 \int_t^{t+1} \overline{\gamma}(s) \, ds}\right).$$

Hence

$$P(\|X(t+1) - \Phi_1(X(t))\| \geqslant \delta | \mathcal{F}_{m(t)}) \leqslant \exp\left(\frac{-B\delta^2}{\int_t^{t+1} \overline{\gamma}(s) \, ds}\right)$$

As the function $x \to e^{-B\delta^2/x}$ is convex for $x \leqslant B\delta^2/2$, the Jensen inequality implies

$$P(\|X(t+1) - \Phi_1(X(t))\| \geqslant \delta | \mathcal{F}_{m(t)}) \leqslant \int_t^{t+1} \exp\left(\frac{-B\delta^2}{\overline{\gamma}(s)}\right) ds$$

for $t \geqslant \tau_{k_0}$. This concludes the proof of the claim. □

Let $k \geqslant k_0$. Define the stopping time $J = \inf\{j \geqslant k : x_j \in M_\epsilon\}$. Inclusion (7) and lemma (3.5) imply

$$\begin{aligned}
P(L(\{x_n\} \text{ is a sink } | \mathcal{F}_k) &\geqslant P(J < \infty \text{ and } D_X(\tau_J) \leqslant \delta | \mathcal{F}_k) \\
&= E(E(\mathbf{1}_{J<\infty} P(D_X(\tau_J) \leqslant \delta | \mathcal{F}_J) | \mathcal{F}_k)) \geqslant E\left(\left(1 - \int_{\tau_J}^{\infty} r(s, \delta) \, ds\right) \mathbf{1}_{J<\infty} | \mathcal{F}_k\right) \\
&\geqslant \left(1 - \int_{\tau_k}^{\infty} r(\delta, s) \, ds\right) P(J < \infty | \mathcal{F}_k) \\
&= \left(1 - 2m \sum_{i>k} \gamma_i \exp\left(\frac{-B\delta^2}{\gamma_i}\right)\right) P(\exists j \geqslant k \ : x_j \in M_\epsilon | \mathcal{F}_k).
\end{aligned}$$

This concludes the proof of the result.

4. Special systems

The purpose of this section is to answer, at least partially, conjecture 2.3 for certain classes of cooperative systems. We continue to assume hypotheses 1.1 and 1.2.

Theorem 4.1. *Suppose F is Morse–Smale and (1) is non-degenerate at Λ. Then with probability one either $L(\{x_n\}) = \{\infty\}$ or; $L\{x_n\}$ is a linearly stable equilibrium.*

Proof. Let γ be a periodic orbit of F of period $T > 0$. Since F is cooperative γ is unordered (see, e.g., theorem 3.3) and unstable (this follows, for example, from proposition 3.2(c)). The vector field F being C^2 and Morse–Smale γ is thus a hyperbolic linearly unstable periodic orbit whose local stable manifold $W^s_{loc}(\gamma)$ is C^2. Therefore, theorem 3.4 applied with $K = \gamma$ and $S = W^s_{loc}(\gamma)$ shows that $P(\lim_{n\to\infty} \mathrm{dist}(x_n, \gamma) = 0) = 0$ (cf Benaïm and Hirsch 1995b). Similarly, $P(\lim_{n\to\infty} \mathrm{dist}(x_n, e) = 0) = 0$ for every linearly unstable equilibrium (cf Pemantle 1990, Brandiere and Duflo 1996). Since for a Morse–Smale vector field the chain-recurrent set consists of finitely many hyperbolic equilibria and periodic orbits, theorem 3.1 implies the result. □

The next result has been used by Benaïm *et al* (1998), as well as Sadeghi (1998), to prove the convergence of that dimensional neural network Kohonen algorithm; and by Benaïm and Hirsch (1999b) in game theory.

Theorem 4.2. *Suppose F has one unique equilibrium p_* then with probability one either $L(\{x_n\}) = \{\infty\}$ or $L(\{x_n\}) = p_*$.*

Proof. The proof follows from theorems 3.1 and 3.3. □

Two-dimensional systems

Theorem 4.3. *Suppose $m = 2$. Then*

(i) With probability one, $L(x_n)$ is a point or a compact arc of equilibria, such an arc being C^1 and either unordered or ordered.

(ii) If (1) is non-degenerate at Λ then $L(x_n)$ is almost certainly an ordered arc (possibly degenerate) contained in $\mathcal{E} \setminus \mathcal{E}_{lu}$. If, furthermore, F is real analytic then $\{x_n\}$ converges almost certainly toward an equilibrium $p \in \mathcal{E} \setminus \mathcal{E}_{lu}$.

Proof. By theorems 3.1 and 3.3 $L(x_n)$ is either an ordered arc of equilibria or a compact connected subset of $\mathcal{R}(\Phi|S_p)$ for some $p \in \mathcal{E}_{abs}$. In the latter case, as S_p is one-dimensional, $\mathcal{R}(\Phi|S_p)$ consists of equilibria and therefore $L(x_n)$ is an unordered C^1 arc of equilibria. This proves (i).

Given $p \in \mathcal{E}_{abs} \setminus p*$. Fix $0 < \epsilon < \frac{1}{4}\eta$ where η is the constant given in proposition 3.2(c). Let $A^*_p(\epsilon)$ denote the set of equilibria $q \in S_p$ such that $\lambda_2(q) \geqslant \epsilon$. It is clear that $A^*_p(\epsilon)$ is a finite set which is a repeller for $\Phi|S_p$. Let

$$A_p(\epsilon) = \{x \in S_p \cap \Lambda : \alpha(x) \notin A^*_p(\epsilon)\}$$

denote the dual attractor of $A^*_p(\epsilon)$. Inequality (2) implies that $\lambda_1(q) \geqslant 2\lambda_2(q) + \eta/2$ for all equilibrium $q \in A_p(\epsilon)$. Since S_p is one dimensional, ergodic measures supported by S_p are Dirac measures at equilibria. Hence by proposition 3.2(d), the basin of attraction of $A_p(\epsilon)$ for $\Phi|S_p$ is a C^2 invariant manifold. Therefore,

$$P\Big(\lim_{n\to\infty} \mathrm{dist}(x_n, \mathcal{R}'_p \cap A_p(\epsilon)) = 0\Big) = 0$$

by application of theorem 3.4 exactly as in the proof of theorem 2.1. Since $A_p^*(\epsilon)$ consists of linearly unstable equilibria

$$P\left(\lim_{n\to\infty} \mathrm{dist}(x_n, A_p^*(\epsilon)) = 0\right) = 0$$

by application of Pemantle (1990), Brandiere and Duflo (1996) or theorem 3.4 applied with $K =$ a linearly unstable equilibrium and S the central stable manifold of K. Part (ii) of the proof now follows from the decomposition (4).

When F is real analytic it follows from Jiang (1991) that the ordered arc of equilibria are degenerate. □

Three-dimensional systems

In this section we assume that $m = 3$. As usual an equilibrium $e \in \mathcal{E}$ is said to be *hyperbolic* if $\lambda_i(e) \neq 0$ for all $i = 1, 2, 3$. A hyperbolic equilibrium is called a *saddle* if $\lambda_1(e) > 0$ and $\lambda_3(e) < 0$.

Theorem 4.4. *Suppose*

(i) Equilibria are hyperbolics.
(ii) $\{x_n\}$ is non-degenerate at Λ
(iii) There exists $\frac{1}{2} < \alpha \leqslant 1$ such that $\lambda_1(e) > (1+\alpha)\lambda_2(e)$ at every saddle point e.

Then $\{x_n\}$ converges with probability one toward a linearly stable equilibrium.

Proof. Let $p \in \mathcal{E}_{asb} \setminus p*$. Let $\gamma \subset S_p$ be a periodic orbit of period $T > 0$ (if any) and $\lambda_1(\gamma) \geqslant \lambda_2(\gamma) \geqslant \lambda_3(\gamma)$ its Lyapounov exponents. For all $x \in \gamma$, the unity is an eigenvalue of the matrix $D\Phi_T(x)$ corresponding to the eigenvector $F(x) \in T_x S_p$. Then, one of the exponents $\lambda_2(\gamma)$ or $\lambda_3(\gamma)$ is zero. Hence, by proposition 3.2(d),

$$\lambda_1(\gamma) \geqslant \lambda_2(\gamma) + \eta \geqslant \eta$$

for some $\eta > 0$. Choose $0 < \epsilon < \eta/(1+\alpha)$, and let $A_p^*(\epsilon)$ denote the union of equilibria $q \in S_p$ and periodic orbits $\gamma \in S_p$ having the property that $\lambda_2(q) \geqslant \epsilon$, $\lambda_3(q) > 0$ and $\lambda_2(\gamma) \geqslant \epsilon$. We claim that $A_p^*(\epsilon)$ consists of finitely many equilibria and periodic orbits. Suppose for the moment that the claim is true. Then $A_p^*(\epsilon)$ is a compact repeller for $\Phi|S_p$ whose dual attractor is

$$A_p(\epsilon) = \{x \in S_p : \alpha(x) \notin A_p^*(\epsilon)\}.$$

By the Poincaré recurrence theorem and the Poincaré–Bendixson theorem it is easy to see that every ergodic invariant measure μ for Φ with support in $A_p(\epsilon)$ is either supported by an equilibrium or a periodic orbit. By definition of $A_p(\epsilon)$ and our assumptions on saddle points this implies that

$$\lambda_1(\mu) \geqslant (1+\alpha)\lambda_2(\mu) + \eta'$$

for some $\eta' > 0$. It then follows from proposition 3.2(d) that the basin of attraction of $A_p(\epsilon)$ for $\Phi|S_p$ is $C^{1+\alpha}$ and by a proof similar to theorem 2.1

$$Pr(L(x_n) \subset A_p(\epsilon) \cap \mathcal{R}'_p) = 0.$$

Since $A_p^*(\epsilon)$ consists of finitely many hyperbolic unstable equilibria and periodic orbits

$$Pr(L(x_n) \subset A_p^*(\epsilon)) = 0$$

by an argument already used in the proof of theorem 4.1. Finally, since $\mathcal{R}_p = (\mathcal{R}_p \cap A_p(\epsilon)) \cup A^*_p(\epsilon)$ we find that

$$Pr(L(x_n) \subset \mathcal{R}'_p) = 0$$

and we conclude exactly as in theorem 2.1.

Our last job is to prove the claim. Let $E \subset \mathbb{R}^m$ and $\Pi : \mathbb{R}^m \to E$ be as in proposition 3.2(b). The flow $\Phi|S_p$ is conjugate to the flow on E induced by the planar vector field $G : E \to E$ defined by $G = \Pi \circ F \circ (\Pi|S_p)^{-1}$. Assertions (c) and (d) of proposition 3.2 imply that $\Pi|S_p$ is always $C^{1+\rho}$ for some $\rho > 0$. Hence G is $C^{1+\rho}$. Therefore, the claim follows from the following lemma.

Lemma 4.5. *Let $G : \mathbb{R}^2 \to \mathbb{R}^2$ be a $C^{1+\rho}$ dissipative vector field with hyperbolic equilibria. Let $\epsilon > 0$ and let $P(\epsilon)$ denote the set of non-stationary periodic orbits whose largest Lyapounov exponent is greater than ϵ. Then $P(\epsilon)$ is finite.*

Here is a sketch of the proof. Let Φ denote the flow induced by G. Suppose that there exists a sequence $\{\gamma_n\}_{n\geqslant 0} \in P(\epsilon)$ of distinct non-stationary periodic orbits of periods $T_n > 0$. By compactness of the global attractor for G we can always suppose that γ_n converges toward a compact set C for the Haussdorf topology. As each γ_n is internally chain recurrent, it is not hard to see that C is also internally chain recurrent. It then easily follows from Benaïm and Hirsch (1995a), that C is either a periodic orbit or a *cycle of equilibria*. That is an invariant set consisting of cyclically connected saddles.

It is clear that C cannot be a periodic orbit. Since, otherwise, there would exist $T > 0$, (the period of C), $p \in C$ and $p_n \in \gamma_n$ such that $D\Phi_{T_n}(p_n) \to D\Phi_T(p)$. This would imply $C \in P(\epsilon)$, making C isolated. Therefore,

$$C = \bigcup_{i=0}^{r-1}(\{e_i\} \cup \{\Upsilon_i\})$$

where $e_i, i = 0, \ldots, r-1$ are equilibria (not necessarily distinct) and $\Upsilon_i, i = 0, \ldots, r-1$ are distinct non-stationary orbits such that $\alpha(\Upsilon_i) = e_i$ and $\omega(\Upsilon_i) = e_{i+1}$ for all $i \in \mathbb{N}$, with the convenient convention that $e_r = e_0$.

Let $\lambda_i < 0$ and $\mu_i > 0$ denote the eigenvalues of the Jacobian matrix $DF(e_i)$, $D_i = \begin{pmatrix} \lambda_i & 0 \\ 0 & \mu_i \end{pmatrix}$ and let $\alpha_i = -\lambda_i/\mu_i$.

By Hartman's C^1 planar linearization theorem† (Hartman 1960) there exist $\eta > 0$, $N_i(\eta)$ a neighbourhood of e_i and a C^1 diffeomorphism $h_i : (-\eta, \eta)^2 \to N_i(\eta)$ such that $h_i(0) = e_i$, $Dh_i(0) = Id$ and

$$h_i(\mathrm{e}^{tD_i}u) = \Phi_t((h_i(u))$$

for all $(u, t) \in (-\eta, \eta)^2 \times \mathbb{R}$ such that $\mathrm{e}^{tD_i}u \in (-\eta, \eta)^2$. This implies:

(a) The proportion of time spent by γ_n in $N_i(\eta)$ is

$$T_{i,n} = -\frac{1}{T_n}\frac{1}{\mu_i}\log\left(\frac{|y_{i,n}|}{\epsilon}\right)$$

with $y_{i,n}$ implicitly defined by $h_i(0, y_{i,n}) \in \gamma_n$.

† Hartman's original theorem is stated for C^1 maps (or vector fields) with Lipschitz derivatives but the proof easily adapts to C^1 maps with Holder derivatives

(b) $$c_1 \left| \frac{y_{i,n}}{\eta} \right|^{\alpha_i} \leqslant \left| \frac{y_{i,n+1}}{\eta} \right| \leqslant c_2 \left| \frac{y_{i,n}}{\eta} \right|^{\alpha_i}$$

where c_1 and c_2 are positive constants depending only on η.

Since $y_{r,n} = y_{0,n}$ and $\lim_{n\to\infty} |y_{0,n}| = 0$, we necessarily have

$$\alpha_0\alpha_1 \ldots \alpha_{r-1} = 1$$

(cf Hofbauer 1981). Since the proportion of time spent by γ_n in $\cup N_i(\eta)$ goes to one as $n \to \infty$ we find that

$$\lim_{n\to\infty} \frac{T_{i,n}}{T_n} = w_i$$

with

$$w_i = \frac{1}{Z} \frac{\alpha_0\alpha_1 \ldots \alpha_i}{\mu_i} \qquad i = 1, \ldots, r$$

and Z is a normalization constant defined by $\sum_{i=1}^r w_i = 1$. This has the consequence that the invariant probability measure m_n supported by γ_n converges for the weak* topology toward

$$m = \sum_{i=1}^{r} w_i \delta_{e_i}.$$

Now, using Liouville's theorem we obtain

$$\epsilon \leqslant \lim_{n\to\infty} \frac{1}{T_n} \log(\mathrm{Det}(D\Phi_{T_n}) = \lim_{n\to\infty} \frac{1}{T_n} \int \mathrm{Tr}(DF(\Phi_t(x))\,\mathrm{d}t = \lim_{n\to\infty} \int \mathrm{Tr}(DF(x))m_n(\mathrm{d}x)$$

$$= \int \mathrm{Tr}(DF(x))m(\mathrm{d}x) = \sum_{i=1}^{r} w_i \,\mathrm{Tr}(DF(e_i)) = \sum_{i=1}^{r} w_i(\lambda_i + \mu_i)$$

$$= \frac{1}{Z}(\alpha_0(1-\alpha_1) + \alpha_0\alpha_1(1-\alpha_2) + \cdots + (\alpha_0\alpha_1 \ldots \alpha_r)(1-\alpha_r))$$

$$= \frac{1}{Z}(\alpha_0 - \alpha_r) = 0.$$

We have reached a contradiction. □

To conclude this section we quote a result proved in Benaïm and Hirsch (1999b), for three-dimensional systems having negative divergence.

Theorem 4.6. *Suppose F has negative divergence:*

$$\forall u \in \mathbb{R}^m \ \mathrm{div}(F)(u) = \sum_{i=1}^{3} \frac{\partial F_i}{\partial u^i}(u) < 0.$$

Then, almost certainly $L\{x_k\}$ is a compact connected subset of $\mathcal{E}$. If, furthermore, the process (1) is non-degenerate at Λ then $L\{x_k\}$ is a compact connected subset of $\mathcal{E} \setminus \mathcal{E}_{lu}$. When F is analytic, $\mathcal{E}$ is finite.

Acknowledgments

This research was partially supported by NATO grant CRG 950857. Conversations with Morris W Hirsch have been very useful.

References

Benaïm M 1996 A dynamical system approach to stochastic approximations *SIAM J. Control Optim.* **34** 437–72

——1997 On invariant hypersurfaces of strongly monotones maps *J. Differ. Equ.* **137** 302–19

——1999 Dynamics of stochastic approximation *Le Séminaire de Probabilités* XXXII *(Springer Lecture Notes in Mathematics vol 1709)* ed J Azema *et al* (Berlin: Springer) pp 1–68

Benaïm M, Fort J C and Pages G 1998 Convergence of the one-dimensional Kohonen algorithm *Adv. Appl. Probab.* **30** 850–69

Benaïm M and Hirsch M W 1995a Chain recurrence in surface flows *Discrete Contin. Dynam. Syst.* **1** 1–16

——1995b Dynamics of Morse–Smale urn processes *Ergodic Theory Dynam. Syst.* **15** 1005–30

——1999a On stochastic approximation algorithms with constant step size whose average is cooperative *Ann. Appl. Probab.* **9** 216–41

——1999b Mixed equilibria and dynamical systems arising from fictitious play in perturbed games *Games Economic Behaviour* **29** 36–72

Brandiere O and Duflo M 1996 Les algorithmes stochastiques contournent-ils les piéges? *Ann. Inst. H Poincaré Prob. Statist.* **32** 395–427

Conley C C 1978 *Isolated Invariant Sets and the Morse Index* (*Regional Conf. Series in Mathematics* vol 38) (Providence, RI: American Mathematical Society) (published for the Conf. Board of the Mathematical Sciences)

Duflo M 1996 *Algorithmes Stochastiques (Mathématiques et Applications vol 23)* (Berlin: Springer)

——1997 Random iterative models *Applications of Mathematics* (Berlin: Springer) (translated from 1990 *Méthodes Recursives Aléatoires* (Paris: Masson))

Fort J C and Pages G 1995 On the a.s. convergence of the Kohonen algorithm with a general neighborhood function *Ann. Appl. Probab.* **4** 1177–216

Hartman P 1960 On local homeomorphism of Euclidean spaces *Bol. Soc. Mat. Mexicana* **5** 220–41

Hirsch M W 1985 Systems of differential equations that are competitive or cooperative: II. Convergence almost everywhere *SIAM J. Math. Anal.* **16** 423–39

——1988 Systems of differential equations which are competitive or cooperative: III. Competing species *Nonlinearity* **1** 51–71

——1999 Chain transitive sets for smooth strongly monotone dynamical systems. Differential equations and dynamical systems *Dynam. Discrete Contin. Impulsive Syst.* **5** 529–43

Hofbauer J 1981 A general cooperation theorem for hypercycles *Monatsh. Math.* **91** 233–40

Jiang J F 1991 Attractors for strongly monotone flows *J. Math. Anal. Appl.* **162** 210–22

Kushner H J and Yin G Y 1997 *Stochastic Approximation Algorithms and Applications* (New York: Springer)

Pemantle R 1990 Nonconvergence to unstable points in urn models and stochastic approximations *Ann. Probab.* **18** 698–712

Ruelle D 1979 Analicity properties of the characteristic exponents of random matrix products *Adv. Math.* **32** 68–80

Sadeghi A 1998 Asymptotic behavior of self-organizing maps with nonuniform stimuli distribution *Ann. Appl. Probab.* **8** 281–99

Smith H L 1995 *Monotone Dynamical Systems. An Introduction to the Theory of Competitive and Cooperative Systems (Mathematical Surveys and Monograph vol 41)* (Providence, RI: American Mathematical Society)

Tarrès P 1999 Pièges répulsifs *C. R. Acad. Sci.* at press

Tereščák I 1996 Dynamics of C^1 smooth strongly monotone discrete-time dynamical systems *Preprint*