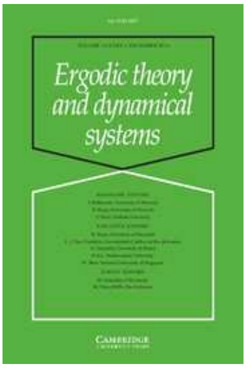

Dynamics of Morse-Smale urn processes

Michel Benaïm and Morris W. Hirsch

Ergodic Theory and Dynamical Systems / Volume 15 / Issue 06 / December 1995, pp 1005 - 1030
DOI: 10.1017/S0143385700009767, Published online: 14 October 2010

Link to this article: http://journals.cambridge.org/abstract_S0143385700009767

How to cite this article:
Michel Benaïm and Morris W. Hirsch (1995). Dynamics of Morse-Smale urn processes. Ergodic Theory and Dynamical Systems, 15, pp 1005-1030 doi:10.1017/S0143385700009767

Ergod. Th. & Dynam. Sys. (1995), **15**, 1005–1030
Printed in Great Britain

Dynamics of Morse–Smale urn processes

MICHEL BENAIM AND MORRIS W. HIRSCH
Department of Mathematics, University of California, Berkeley, CA 94720 USA

(*Received* 7 *January* 1994)

Abstract. We consider stochastic processes $\{x_n\}_{n\geq 0}$ of the form

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

where $F : \mathbb{R}^m \to \mathbb{R}^m$ is C^2, $\{\gamma_i\}_{i\geq 1}$ is a sequence of positive numbers decreasing to 0 and $\{U_i\}_{i\geq 1}$ is a sequence of uniformly bounded $\mathbb{R}^m$-valued random variables forming suitable martingale differences. We show that when the vector field F is Morse–Smale, almost surely every sample path approaches an asymptotically stable periodic orbit of the deterministic dynamical system $dy/dt = F(y)$. In the case of certain generalized urn processes we show that for each such orbit Γ, the probability of sample paths approaching Γ is positive. This gives the generic behavior of three-color urn models.

0. *Introduction*

Let $F : \mathbb{R}^m \to \mathbb{R}^m$ be a smooth map and $\Delta \subset \mathbb{R}^m$ a compact set. We consider discrete time stochastic processes $\{x_n\}_{n\geq 0}$ defined on Δ (i.e. $x_n \in \Delta$ for all $n \geq 0$) by

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

where $\{\gamma_i\}_{i\geq 1}$ is a sequence of positive numbers and $\{U_i\}_{i\geq 1}$ is a sequence of $\mathbb{R}^m$-valued random variables defined on a probability space $(\Omega, \mathcal{F}, P)$. We suppose that:

- (i) The sequence $\{\gamma_i\}$ is decreasing and $\sum_i \gamma_i = +\infty$. Such a sequence will be called a *decreasing gain sequence*.
- (ii) There exists an increasing sequence of subsigma fields $\{\mathcal{F}_n\}_{n\geq 0}$ for $(\Omega, \mathcal{F}, P)$ such that
  - (a) U_n is measurable with respect to $\mathcal{F}_n$,
  - (b) $E(U_{n+1}|\mathcal{F}_n) = 0$.
- (iii) There exists $K > 0$ such that $\| U_n \| \leq K$ for all $n \geq 0$.

Throughout this paper $\| \cdot \|$ denotes the Euclidean norm on $\mathbb{R}^m$ and $\langle , \rangle$ the associated inner product. We may denote by d the induced distance.

Processes described by (1) encompass several generalized urn models and stochastic approximation algorithms. Concerning the asymptotic behavior of these processes, the literature usually focuses on questions having the following form: does $\{x_n\}$ converge almost surely to some random variable x_∞? If so, what is the support of x_∞? What is its probability law?

A natural approach to the asymptotic behavior of the sequences $\{x_n\}$ is to consider them as approximations to trajectories of the vector field F, that is, to compare them to solutions of

$$\frac{dy}{dt} = F(y). \tag{2}$$

One can think of Equation (1) as a kind of Cauchy–Euler approximation scheme for numerically solving Equation (2), with decreasing step size γ_n and added noise U_n. It is natural to expect that, owing to assumption (b), the noise washes out in the long run, and that almost surely limit points of a sample path $\{x_n\}$ are closely related to the behavior of trajectories of Equation (2).

Until recently, most work in this direction has assumed the simplest dynamics for F, for example that F is the negative gradient of a function u. With suitable assumptions it was proved that almost surely sample paths converge to a local minimum of u.

The main purpose of this paper is to show how the asymptotic behavior of $\{x_n\}$ can be described in terms of the asymptotic behavior of the flow Φ generated by the vector field F, even in nonconvergent situations, provided the dynamics of Φ are not too complicated. In particular, we consider the asympotic behavior of urn models associated to a Morse–Smale vector field.

The key to our results are recent papers by Pemantle (1990) and Benaim (1993). Pemantle showed that under reasonable assumptions, the probability that sample paths converge to an unstable equilibrium of F is zero. Benaim showed that almost surely the limit set of a sample path $\{x_n\}$ is a compact connected invariant set of chain recurrent points of the flow of F.†

We extend Pemantle's arguments to cover unstable periodic orbits and apply our results to certain urn models in which F is a Morse–Smale vector field.

Example 0.1. (Generalized urn processes.) The *unit m-simplex* $\Delta^m \subset \mathbb{R}^{m+1}$ is the set

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

We consider Δ^m as a differentiable manifold (with corners), identifying its tangent space at any point with the linear subspace

$$E^m = \left\{ z \in \mathbb{R}^{m+1} : \sum z_j = 0 \right\}.$$

An urn initially (i.e. at time $n = 0$) contains $n_0 \geq 1$ balls of colors $1, \ldots, m+1$. At each time step a new ball is added to the urn and its color is randomly chosen as follows: let $x_{n,i}$ be the proportion of balls having color i at time n and denote by $x_n \in \Delta^m$ the vector of proportions $x_n = (x_{n,1}, \ldots, x_{n,m+1})$. The color of the ball added at time $n+1$ is chosen to be i with probability $f_i(x_n)$, where the f_i are the coordinates of a function $f : \Delta^m \to \Delta^m$.

---

† Recently we have proved that the restriction of Φ to $R(F)$ is chain recurrent.

Such processes, known as *generalized Polya urns*, have been considered by Hill, Lane and Sudderth (1980) for $m = 1$; Arthur, Ermol'ev and Kaniovskii (1983); Pemantle (1990). Arthur (1988) has used urn processes as models of competing technologies.

An urn process is determined by the initial *urn composition* (x_0, n_0) and the *urn function* $f : \Delta^m \to \Delta^m$. The process $\{x_n\}_{n\geq 0}$ is a nonstationary Markov process whose probability law we denote by $P_{(x_0,n_0)}$. We assume that the initial composition (x_0, n_0) is fixed once for all. If no confusion can arise we set $P = P_{(x_0,n_0)}$. The σ-field $\mathcal{F}_n$ is the field generated by the random variables $x_0, \dots, x_n$.

The expected number of balls of color j at time $n+1$, given the value of the proportion vector x_n, is a random variable having expected value $f_j(x_n)$. From this it is easily computed that the expected value of x_{n+1}, given the value of x_n, satisfies the equation

$$(n_0 + n + 1)E(x_{n+1}|x_n) - (n_0 + n)x_n = f(x_n).$$

Defining the random variables

$$U_{n+1} = [x_{n+1} - E(x_{n+1}\,|x_n)](n_0 + n + 1),$$

we see that

$$x_{n+1} - x_n = \frac{1}{n_0 + n + 1}(-x_n + f(x_n) + U_{n+1}) \tag{3}$$

and

$$E(U_{n+1}|\mathcal{F}_n) = E(U_{n+1}|x_n) = 0.$$

This shows that $\{x_n\}$ is a Markov process.

In discussing urn models we shall usually identify the affine space $\{v \in \mathbb{R}^{m+1} : \sum_{j=1}^{m+1} v_j = 1\}$ with the linear subspace E^m by parallel translation and also with $\mathbb{R}^m$ by any convenient affine isometry. Under the latter identification, process (3) takes exactly the form (1), where $F : \mathbb{R}^m \to \mathbb{R}^m$ denotes any map which equals $-\mathrm{Id} + f$ on Δ^m, and $\gamma_{n+1} = 1/(n_0 + n + 1)$.

An equivalent geometrical description of the urn process $\{x_n\}$ is as follows. For each n, denote by s_{n+1} the random variable whose value is one of the $m+1$ vertices $(e_1, \dots, e_{m+1})$ of Δ^m, chosen according to the probability distribution $(f_1(x_n), \dots, f_{m+1}(x_n))$. Then x_{n+1} is the convex combination:

$$x_{n+1} = (1 - \gamma_{n+1})x_n + \gamma_{n+1}\mathrm{s}_{n+1}. \tag{4}$$

Example 0.2. (Stochastic approximation.) Let $\{\xi_i\}_{i\geq 1}$ be a sequence of independent identically distributed $\mathbb{R}^d$-valued random inputs to a system and let $x_n \in \Delta$ denote a parameter to be updated, $n \geq 0$. We suppose the updating to be defined by a known bounded map $h : \mathbb{R}^m \times \mathbb{R}^d \to \mathbb{R}^m$, and the following stochastic algorithm:

$$x_{n+1} - x_n = \gamma_{n+1}h(x_n, \xi_{n+1}).$$

Let μ be the common probability law of the ξ_n. Introduce the *average vector field*

$$F(x) = \int h(x, \xi)d\mu(\xi)$$

and set

$$U_{n+1} = h(x_n, \xi_{n+1}) - F(x_n).$$

It is clear that this algorithm has the form given by (1). Such processes have been used for stochastic learning and adaptive algorithms (e.g. Kushner and Clark, 1978; Beneveniste, Metivier and Priouret, 1990)

Outline of contents. §1 reviews earlier results by Pemantle and Benaim on which the paper is based. The main results, Theorems 2.1, 2.5 and 2.7, are presented in §2. The dynamics needed for their proofs is developed in §3 and the probability in §4.

1. *Chain recurrence and nonconvergence to unstable equilibria*

Throughout the paper $F : \mathbb{R}^m \to \mathbb{R}^m$ denotes a C^r mapping, $1 \leq r \leq \infty$.

Since the stochastic process $\{x_n\}_{n\geq 0}$ takes values in the compact set Δ, the nature of F outside Δ does not affect the behavior of (1). Therefore we assume, without loss of generality, that F is bounded, that is, $\sup\{||F(x)||\} < \infty$. It follows that F is completely integrable, i.e. F generates a C^r flow

$$\begin{aligned} \Phi : \mathbb{R} \times \mathbb{R}^m &\to \mathbb{R}^m, \\ (t, x) &\mapsto \Phi_t(x) \end{aligned}$$

defined by

$$\begin{aligned} \Phi_0 &= \mathrm{Id}, \\ \frac{d\Phi_t(x)}{dt} &= F(\Phi_t(x)). \end{aligned}$$

A notion of recurrence for F well suited to analysis of the asymptotic behavior of (1) is *chain recurrence* (Conley, 1978). A point $z \in \mathbb{R}^m$ is said to be (δ, T) *recurrent* if $\delta > 0$, $T > 0$ and there exist an integer k, points y_j in $\mathbb{R}^m$ and numbers t_j, $0 \leq j \leq k-1$, such that:

$$t_j \geq T;\ ||y_0 - z|| < \delta;\ ||\Phi_{t_j}(y_j) - y_{j+1}|| < \delta,\ (j = 0, \ldots, k-1);\ y_k = z.$$

If z is (δ, T) recurrent for all $\delta > 0$, $T > 0$ then z is called *chain recurrent.*

We denote by $R(F)$ the set of chain recurrent points for F. This is a closed invariant set which contains the nonwandering set of F and, consequently, the limit sets of solution curves of F.† To describe the asymptotic behavior of (1) we consider the *limit set* of any sequence $\{y_n\}_{n\geq 0}$ in $\mathbb{R}^m$, denoted by $L(\{y_n\})$. It is defined, as usual, as the set of $p \in \mathbb{R}^m$ such that $\lim_{k\to\infty} y_{n_k} = p$ for some sequence $n_k \to \infty$.

The following result, which is purely deterministic, is proved in Benaim (1993) (which requires only that F be locally Lipschitz). The subscript n runs over the natural numbers.

---

† z is *wandering* if there is a neighborhood R of z and a positive number T such that $R \cap \Phi_t R$ is empty for all $t > T$. Otherwise z is *nonwandering*. A chain recurrent point can be wandering. For example, consider the vector field $\sin^2(\theta/2)$ on the unit circle parameterized by $\theta \in \mathbb{R}/2\pi\mathbb{R}$, whose flow is defined by $d\theta/dt = \sin^2(\theta/2)$: Every point is chain recurrent, but only the equilibrium $\theta = 0$ is nonwandering.

THEOREM 1.1. (Benaim, 1993.) *Suppose given sequences* $\{u_n\}$, $\{b_n\}$ *in* $\mathbb{R}^m$ *and a decreasing gain sequence* $\{\gamma_n\}$. *Let* $\{y_n\}$ *satisfy the recursion*

$$y_{n+1} - y_n = \gamma_{n+1}(F(y_n) + u_{n+1} + b_{n+1}).$$

Assume:

(i) $\{y_n\}$ *is bounded.*

(ii) $\lim_{n\to\infty} b_n = 0$.

(iii) *For each* $T > 0$,

$$\lim_{n\to\infty} \sup \left\{ \| \sum_{i=n+1}^{k} \gamma_i u_i \| : k \in \mathbb{N},\ 0 \le \tau_k - \tau_n \le T \right\} = 0,$$

where $\tau_n = \sum_{i=1}^{n} \gamma_i$.

Then:

(a) *the limit set* $L(\{y_n\})$ *is a nonempty, compact, connected set which is invariant under the flow* Φ *of* F; *and*

(b) $L(\{y_n\}) \subset R(F)$, *the chain recurrent set of* F.

Returning to our basic stochastic process

$$x_{n+1} - x_n = \gamma_{n+1}(F(x_n) + U_{n+1}), \tag{5}$$

we recall that $(\{x_n : \Omega \to \mathbb{R}^m\}_{n\ge 0})$ is a sequence of random variables. For any point $\omega \in \Omega$ we may consider the limit set $L(\{x_n(\omega)\})$, called a *sample limit set.* We consider it as the value at ω of the set-valued random variable $L(\{x_n\})$. By the usual abuse of language, we refer to this random variable as the *limit set* of process (5).

We now derive the following probabilistic consequence from Theorem 1.1:

COROLLARY 1.2. *Assume there exists* $\delta \ge 1$ *such that the decreasing gain sequence* $\{\gamma_i\}_{i\ge 1}$ *of (5) satisfies*

$$\sum_i \gamma_i^{1+\delta} < \infty.$$

Then almost surely sample limit sets of process (5) satisfy:

(a) $L(\{x_n\})$ *is a nonempty, compact, connected set which is invariant under the flow of* F; *and*

(b) $L(\{x_n\}) \subset R(F) \cap \Delta$.

For $\delta = 1$, this is an easy consequence of the L^2 bounded martingale convergence theorem. Indeed, we let $u_n = U_n$, $Z_n = \sum_{i=0}^{n} \gamma_i U_i$ and $b_n = 0$. The sequence $\{(Z_n, \mathcal{F}_n)\}_{n\ge 0}$ is a martingale (i.e. Z_n is $\mathcal{F}_n$ measurable and $E(Z_{n+1}|\mathcal{F}_n) = Z_n$) and we have

$$\sup_{n\ge 0} E(\| Z_n \|^2) \le K^2 \sum_{n\ge 0} \gamma_n^2 < \infty.$$

It follows that $\{Z_n\}$ converges almost surely (see e.g. (Hall and Heyde, 1980)). Thus $\lim_{n\to\infty} \sup_{j\ge 0} \| Z_{n+j} - Z_n \| = 0$. This validates assumption (iii) of Theorem 1.1.

For $\delta > 1$, $\{Z_n\}$ may be non-convergent. However the following result can be deduced from Metivier and Priouret (1987, Corollaire 11): let $\{U_n\}_{n\geq 0}$ be a sequence of $\mathcal{F}_n$ measurable random variables such that $E(U_{n+1}|\mathcal{F}_n) = 0$ and

$$\sup_{n\geq 0} E(\| U_n \|^q) < \infty$$

for some $q \geq 2$. If $\sum_{n\geq 0} \gamma_n^{1+q/2} < \infty$, then

$$\lim_{n\to\infty} \left(\sup_{\{k:0\leq \tau_k - \tau_n \leq T\}} \| \sum_{i=n+1}^{k} \gamma_i U_i \| \right) = 0$$

almost surely. Thus, letting $\delta = q/2$ proves assumption (iii) of Theorem 1.1. Since (i) and (ii) are obvious, Corollary 1.2 follows. □

Let $p \in \Delta$ be an equilibrium of F; that is $F(p) = 0$. As usual, if all eigenvalues of $DF(p)$ have nonzero real parts, p is called *hyperbolic*. If all eigenvalues of $DF(p)$ have negative real parts, p is *linearly stable*. If some eigenvalue has positive real part, p is *linearly unstable*. It is well known that if p is linearly stable then it is also *asymptotically stable*, that is, all forward trajectories starting in some neighborhood of p converge uniformly to p. It is also well known that if p is linearly unstable then it cannot be asymptotically stable; in fact there is a neighborhood of p containing no complete forward orbit other than p.

Suppose p is a hyperbolic equilibrium of F which is linearly unstable. Then the set of initial values whose forward trajectories converge to p—the *stable manifold* $W_s(p)$ of p—is the image of an injective C^r immersion $\mathbb{R}^k \to \mathbb{R}^m$ where $0 \leq k < m$. Consequently $W_s(p)$ has measure 0 in $\mathbb{R}^m$. This suggests that for the stochastic process (5), convergence of sample paths $\{x_n\}$ to p is a null event, provided the noise $\{U_n\}$ has sufficiently large components in the unstable directions at p. Such a result has been proved under mild continuity assumptions on F by Lane, Hill and Sudderth (1980) for urn models in the one-dimensional case. More recently (Pemantle, 1990) obtained the following result of this kind for the general case of (5) provided the vector fields F is C^2 and the gain sequence is well behaved.

For any real number a, let $a^+ = \max(a, 0)$.

THEOREM 1.3. (Pemantle, 1990.) *Let $p \in \Delta$ be a linearly unstable hyperbolic equilibrium of F. Assume:*

(i) *F is C^2.*

(ii) *$A/n^\mu \leq \gamma_n \leq B/n^\mu$ where $0 < A \leq B$, $\frac{1}{2} < \mu \leq 1$.*

(iii) *There exists $b > 0$ such that for all unit vector $\Theta \in \mathbb{R}^m$:*

$$E(\langle U_{n+1}, \Theta\rangle^+ | \mathcal{F}_n) \geq b.$$

Then: $P(\lim_{n\to\infty} x_n = p) = 0$.

Remark 1.4. It can be seen, by using Lemma 4.2 and the reasoning involved in the proof of Theorem 2.1, that Pemantle's theorem remains true for $0 < \mu \leq 1$. In fact the proof for $0 < \mu \leq 1/2$ is easier than for $1/2 < \mu \leq 1$. The proof of Theorem 2.1 in §4 can easily be adapted to give a proof of Pemantle's theorem.

In Theorem 2.1, proved in §4, we extend Pemantle's result: almost surely sample paths do not approach a linearly unstable hyperbolic periodic orbit.

2. *Main results*

A nonstationary periodic orbit of F is called a *cycle.*

Let $\Gamma \subset \Delta$ be a cycle of period $T > 0$. For any $p \in \Gamma$, the spectrum of $D\Phi_T(p)$ (the Jacobian matrix of Φ_T at p) can be written as $\{1\} \cup C(\Gamma)$ where $C(\Gamma)$ is the set of characteristic multipliers. If $C(\Gamma)$ does not meet the unit circle of the complex plane, Γ is called *hyperbolic.* If $C(\Gamma)$ is strictly inside the unit circle, Γ is called *linearly stable.* If $C(\Gamma)$ meets the exterior of the unit circle, i.e. if some eigenvalue of $D\Phi_T(p)$ has modulus strictly greater than one, Γ is *linearly unstable.*

A linearly stable cycle has the property of being an *attractor*: a nonempty, compact invariant set Λ having a neigborhorhood B such that

$$\lim_{t\to\infty} d(\Phi_t z,\ \Lambda) = 0$$

uniformly for $z \in B$. The union of all such B is an open neighborhood of Λ called its *basin.* When the basin is the whole state space then Λ is a *global attractor*; in this case the vector field is called *dissipative.*

Suppose Γ is a hyperbolic cycle. It is well known that Γ has a neighborhood in which Γ is the only nonempty invariant set. It therefore follows from Theorem 1.1 that if L is a sample limit set of process (5), then almost surely either $L = \Gamma$ or else $L \cap \Gamma = \emptyset$.

We prove in §4 the following extension of Theorem 1.3, concerning sample paths $\{x_n\}$ of process (5):

THEOREM 2.1. *Let* $\Gamma \subset \Delta$ *be a hyperbolic linearly unstable cycle of* F. *Assume:*

(i) F *is* C^2.

(ii) $\frac{A}{n^\mu} \le \gamma_n \le \frac{B}{n^\mu}$ *where* $0 < A \le B,\ 0 < \mu \le 1$.

(iii) *There exists* $b > 0$ *such that for all unit vector* $\Theta \in \mathbb{R}^m$:

$$E(\langle U_{n+1}, \Theta\rangle^+ | \mathcal{F}_n) \ge b.$$

Then

$$P(L(\{x_n\}) = \Gamma) = 0.$$

The preceding results can now be used to describe the global asymptotic behavior of (5) when the vector field F is Morse–Smale. Since the sample paths $\{x_n\}$ remain in Δ, we may assume that the point at infinity is a source for F, or in other words, that the flow Φ of F has a global attractor.

A C^r ($r \ge 1$) vector field on a manifold M is called *Morse–Smale* if

(i) All periodic orbits (equilibria and cycles) are hyperbolic, and all intersections of their stable and unstable manifolds are transverse.

(ii) Every alpha or omega limit set is a periodic orbit (equilibrium or cycle).

(iii) F is transverse to the boundary ∂M.

(iv) F has a global attractor.

It is known that these conditions imply that there are only finitely many periodic orbits.

Morse–Smale vector fields play an important role in the modern theory of dynamical systems in the sense that they constitute a nice class of structurally stable vector fields on compact manifolds (Palis, 1969; Palis and Smale, 1968). Furthermore, Morse–Smale vector fields on an orientable compact surface coincide with structurally stable vector fields and are generic (Peixoto, 1962).

Suppose F is a Morse–Smale vector field. Denote by $L(F)$ the union of all alpha and omega limit sets of F and by $Per(F)$ the union of all periodic orbits (equilibria and cycles). If F is Morse–Smale, $L(F)$ decomposes as $L(F) = Per(F) = \Gamma_0 \cup \cdots \cup \Gamma_l$ where the Γ_i are the distinct hyperbolic periodic orbits (perhaps equilibria). In addition, $L(F)$ has a partial order structure defined by $\Gamma_i \leq \Gamma_j$ if and only if $W^u(\Gamma_j) \cap W^s(\Gamma_i) \neq \emptyset$, where W^u and W^s denote stable and unstable manifolds (see e.g. proposition 3.2 of (Palis, 1969)). It can be shown that $R(F) = L(F) = Per(F)$.†

Corollary 1.2 implies that *if F is Morse–Smale then the limit set of process (5) is almost surely one of the* Γ_i. Denote the probability of this, for a given Γ_i, by

$$\mathbf{p}(\Gamma_i) = P(L(\{x_n\}) = \Gamma_i).$$

The following result then follows from Corollary 1.2, Pemantle's Theorem 1.3 and Theorem 2.1:

COROLLARY 2.2. *Assume:*

(i) *There exists $\delta \geq 1$ such that $\sum_{n \geq 0} \gamma_n^{1+\delta} < \infty$.*

(ii) *F is a Morse–Smale vector field.*

Let $\{\Gamma_i, i = 1, \ldots, l\}$ denote the set of periodic orbits in Δ. Then:

(a) $\sum_{i=1}^{l} \mathbf{p}(\Gamma_i) = 1$.

(b) *If conditions (i) through (iii) of Theorem 2.1 are satisfied, then $\mathbf{p}(\Gamma_i) > 0 \Rightarrow \Gamma_i$ is linearly stable.*

Remark 2.3. The structural stability of Morse–Smale systems implies that if a vector field F' is sufficiently close to F in the C^1 topology, then F' is also Morse–Smale, and there is a one-to-one correspondence $\Gamma_i \leftrightarrow \Gamma_i'$ between periodic orbits of F and of F', taking equilibria to equilibria and cycles to cycles. It is then interesting to compare limits sets of process (5) for F' and F. If the hypotheses of Corollary 2.2 are satisfied, then it is reasonable to conjecture that

$$\lim_{F' \to F} \mathbf{p}(\Gamma_i') = \mathbf{p}(\Gamma_i).$$

This would be a kind of stochastic stability for process (5) when F is Morse–Smale.

It is usually difficult to verify that a particular vector field is Morse–Smale. But it is not uncommon to deal with a vector field F admitting a *strict Liapunov function h* on the state space: this means that h is a nonnegative, continuous real-valued function which strictly decreases along nonconstant forward trajectories. Examples include many error functions for learning algorithms, and energy functions in dissipative mechanical systems.

The following result is a consequence of Theorems 1.3 and Corollary 1.2:

---

† For example, this can be deduced from a filtration for $L(F)$, the existence of which follows from §7 of (Pugh and Shub, 1970).

COROLLARY 2.4. *Assume:*

(i) *F is C^2 and dissipative, and the equilibria are hyperbolic;*

(ii) *F admits a strict Liapunov function h which has a unique local minimum at p;*

(iii)

$$\frac{A}{n^{\mu}} \leq \gamma_n \leq \frac{B}{n^{\mu}}$$

where $0 < A \leq B$, $0 < \mu \leq 1$.

(iv) *There exists $b > 0$ such that for all unit vector $\Theta \in \mathbb{R}^m$:*

$$E(\langle U_{n+1}, \Theta\rangle^{+} \,|\mathcal{F}_n) \geq b.$$

Then for process (5), $\lim_{n\to\infty} x_n = p$ almost surely.

Proof. From assumptions (i) and (ii) it follows that $R(F)$ is a finite set of hyperbolic equilibria, of which only p is linearly stable. The hypotheses of Corollary 1.2 and Theorem 1.3 hold, and those theorems imply the conclusion. ☐

Urn processes. We now consider processes (5) in the special case of *urn processes*, Equation (3). We use the notation of Corollary 2.2, setting $\Delta = \Delta^m$, and recalling the identification of the affine subspace spanned by Δ^m with $\mathbb{R}^m$.

The following result will be used in Theorem 2.7 to strengthen conclusion (*b*) of Corollary 2.2 to a double implication:

THEOREM 2.5. *Assume:*

(i) *The urn function $f : \Delta^m \to \Delta^m$ is C^1.*

(ii) $f(\Delta^m) \subset \mathrm{Int}(\Delta^m)$.

(iii) *The vector field $F = -\mathrm{Id} + f$ has an attractor $\Lambda \subset \mathrm{Int}(\Delta^m)$.*

Then for limit sets of the urn process (3):

$$P(L(\{x_n\}) \subset \Lambda) > 0.$$

The proof is given in §4

Remark 2.6. Theorem 2.5 leads to a somewhat paradoxical result. For stochastic approximation (Example 0.2), one often uses a result like Theorem 1.1 to study a stochastic process through the dynamics of the average vector field F. Frequently (e.g. in many neural learning methods) F is the negative gradient of an error function E assumed to have nondegenerate critical points, and one commonly identifies limit sets of sample paths $\{x_n\}$ with local minima of E. While this is a correct application of Theorems 1.1 and 1.3, one cannot conclude—as is sometimes assumed—that if x_0 is very close to a local minimum p, then x_n converges to p almost surely: Theorem 2.5 implies this is false in the generic situation that E is a Morse function. It may be true, however, that the probability is very small that a sample path starting near one local minimum converges to another one. This is an interesting question for further research.

For Morse–Smale urn models we have the following result:

THEOREM 2.7. *Assume:*

(i) *The urn function* $f : \Delta^m \to \Delta^m$ *is* C^2.
(ii) $f(\Delta^m) \subset \mathrm{Int}(\Delta^m)$.
(iii) *There exists a Morse–Smale vector field* F *on* $\mathbb{R}^m$ *such that* $F|_{\Delta^m} = -\mathrm{Id} + f$
Then:
(a) $\sum_{i=1}^{l} \mathbf{p}(\Gamma_i) = 1$.
(b) $\mathbf{p}(\Gamma_i) > 0 \Leftrightarrow \Gamma_i$ *is linearly stable.*

If f is not required to satisfy (ii), one can construct urn functions so that the vector field $-\mathrm{Id} + f$ has two different point attractors p, $q \in \Delta^m$, in such a way that if the urn process starts sufficiently near p, the probability of its entering the basin of q is 0. For example, take $m = 3$, p=e_1, $q = e_2$, and assume f_2 and f_3 are identically zero in a neighborhood of e_1.

On the other hand, from the point of view of global analysis the condition that $f(\Delta^m) \subset \mathrm{Int}(\Delta^m)$ is not very restrictive, as the following result shows.

Let $U^r(\Delta^m)$ denote the space of C^r urn functions $f : \Delta^m \to \Delta^m$, r being a positive integer. Let $\mathcal{A}$ be the subset of $U^r(\Delta^m)$ of functions satisfying $f(\Delta^m) \subset \mathrm{Int}(\Delta^m)$.

LEMMA 2.8. *$\mathcal{A}$ is open and dense in $U^r(\Delta^m)$.*

Proof. The openness is obvious. For the denseness, let $\{\Psi_t\}$ denote the flow of some C^∞ vector field H on Δ^m transverse to the boundary of Δ^m and pointing inward Δ^m, e.g. $H(z) = b - z$ where $b \in \Delta^m$ is the barycenter. For sufficiently small $\epsilon > 0$, any function $f \in C^r(\Delta^m, \Delta^m)$ can be approximated by $g = f \circ \Psi_\epsilon$, with $g(\Delta^m) \subset \mathrm{Int}(\Delta^m)$. □

Remark 2.9. Most of our results on urn processes remain valid if the assumption $f(\Delta^m) \subset \mathrm{Int}(\Delta^m)$ is replaced by the weaker hypothesis:
(i) If $z \in \mathrm{Int}(\Delta^m)$ then $f(z) \in \mathrm{Int}(\Delta^m)$;
(ii) If $z \in \partial(\Delta^m)$ then there exists $i \in \{1, \dots, m+1\}$ such that $z_i = 0$ and $f_i(z) \neq 0$.

Consider now the particular case of three-color urn models; thus $m = 2$.

Let $\chi^r(\Delta^2)$ be the space of C^r vector fields on Δ^2 transverse to the boundary of Δ^2. Denote by $MS(\Delta^2) \subset \chi^r(\Delta^2)$ the subspace of Morse–Smale vector fields. According to Peixoto's Theorem (1962), $MS(\Delta^2)$ is open and dense in $\chi^r(\Delta^2)$.†

Define the map $T : \mathcal{A} \to \chi^r(\Delta^2)$ by $T(f) = -\mathrm{Id} + f$. It is readily seen that $T(\mathcal{A})$ is open and T defines an homeomorphism from $\mathcal{A}$ onto $T(\mathcal{A})$. It therefore follows from Lemma 2.8 that $T^{-1}(MS(\Delta^2))$ is open and dense in $\mathcal{A}$, hence in $U^r(\Delta^2)$. Thus we have proved the following corollary of Theorem 2.7:

COROLLARY 2.10. *Let r be a positive integer ≥ 2. There exists an open dense subset $\mathcal{B} \subset U^r(\Delta^2)$ such that for any urn function $f \in \mathcal{B}$, conclusions (a) and (b) of Theorem 2.7 hold.*

---

† Peixoto's theorem is usually stated for oriented compact surfaces without boundary, but remains true for vector fields on Δ^2 provided that we restrict attention to vector fields transverse to the boundary.

3. *Geometric dynamics of hyperbolic cycles*

From now on we assume that the vector field $F : \mathbb{R}^m \to \mathbb{R}^m$ is C^r, $r \geq 2$; and that $\Gamma \subset \Delta$ denotes a hyperbolic linearly unstable cycle. Since Γ is a compact trajectory of a C^r vector field it follows that Γ is a C^{r+1} submanifold of $\mathbb{R}^m$.

In his proof that convergence of $\{x_n\}$ to a linearly unstable equilibrium p is a null event, Pemantle (1990) studied the distance $\eta(z)$ from a point z to the local stable manifold $W^s_\epsilon(p)$ of p, measured parallel to the unstable linear subspace $E^u(p)$. In our case p is replaced by a cycle Γ.

We need a similar function for Γ. Complications arise because each $p \in \Gamma$ has its own local unstable manifold $W^s_\epsilon(p)$, and its own unstable subspace $E^u(p)$, and it is not obvious how to define $\eta(z)$.

We proceed to define $\eta(z)$ in a neighborhood $\mathcal{N}(\Gamma)$. A simple way to define $\mathcal{N}(\Gamma)$ is to take all points of the form $x + v$ where, for each $p \in \Gamma$, x runs through $W^s_\epsilon(p)$ and $v \in E^u(p)$ with $||v||$ sufficiently small. But in order to obtain the needed facts about $\mathcal{N}(\Gamma)$ we use the following equivalent but more long-winded definition.

The neighborhood $\mathcal{N}(\Gamma)$ and the maps π_1, π. The flow $\{\Phi\}$ of the C^r vector field F is C^r, that is, the map $(t, x) \to \Phi_t(x)$ is C^r. Standard theory of differential equations (e.g. Hartman 1964, Chapter V, Theorem 4.1) gives us a little extra smoothness, however. First, $\Phi_t(x)$ is C^{r+1} in t (trajectories are C^{r+1} curves). Moreover the rth derivatives in x of the maps Φ_t, denoted by $D^r\Phi_t(x)$, are continuously differentiable in t.

Denote by $T_\Gamma\mathbb{R}^m = \Gamma \times \mathbb{R}^m$ the tangent vector bundle of $\mathbb{R}^m$ restricted to Γ, and note that this is a C^{r+1} manifold.

The *restricted tangent flow* $T_\Gamma\Phi$ in $T_\Gamma\mathbb{R}^m$ is defined by

$$T_\Gamma\Phi_t : T_\Gamma\mathbb{R}^m \to T_\Gamma\mathbb{R}^m, \ (p, y) \mapsto (\Phi_t(p), D\Phi_t(p)y).$$

This flow, which covers the periodic flow $\Phi|\Gamma$, maps fibres linearly to fibres. It will be important to know that *the flow $T_\Gamma\Phi$ is C^r in (t, p, y)* (and not merely C^{r-1}). The reason is that for $D\Phi_t(p)$, as a matrix-valued function of t, is the solution to the linear, nonautonomous *variational equation*

$$\begin{aligned} \frac{d}{dt}D\Phi_t(p) &= DF(\Phi_t(p))D\Phi_t(p), \\ D\Phi_0(p) &= I. \end{aligned}$$

Here p is a parameter belonging to the C^{r+1} manifold Γ. Since the $DF(x)$ is C^{r-1} in x and C^r in $p \in \Gamma$, the solution $(t, p) \mapsto D\Phi_t(p)$ to the variational equation is C^r in t and, seemingly, only C^{r-1} in p.

But locally p is an affine function of t, which is a C^{r+1} function on Γ. Thus in a neighborhood N of a point $p_0 \in \Gamma$ we can write $p = \Phi_{s(p)}p_0$ where s is a C^{r+1} function on N. Then we see that $D\Phi_t(p)$ is a composition of C^r maps, since

$$D\Phi_t(p) = D\Phi_{t+s(p)}(p_0) \circ D\Phi_{-s(p)}(p_0)^{-1}.$$

Similar considerations apply to mixed derivatives. Therefore the solution to the variational equation is C^r in (t, p).

By hyperbolicity (see for example Hirsch and Pugh (1970), Hirsch, Pugh and Shub (1977)) there exist positive constants c, α and a decomposition of $T_\Gamma \mathbb{R}^m$ as the direct sum of three vector bundles:

$$T_\Gamma \mathbb{R}^m = E^s(\Gamma) \oplus E^u(\Gamma) \oplus E^\Phi(\Gamma)$$

which is invariant under $T_\Gamma \Phi$, and such that for all $p \in \Gamma$, $t \geq 0$:

$$\| D\Phi_t(p)|_{E^s(p)} \| \leq ce^{-\alpha t}, \tag{6}$$

$$\| D\Phi_{-t}(p)|_{E^u(p)} \| \leq ce^{-\alpha t},$$

$$E^\Phi(p) = \mathrm{span}(F(p))$$

where $p \times E^u(p)$ denotes the fibre of $E^s(\Gamma)$ over p, and a similar notation applies to $E^s(p)$ and $E^\Phi(p)$.

Because Γ is linearly unstable, the dimension of $E^u(p)$ is at least 1.

Each $E^u(p)$ is a linear subspace of $\mathbb{R}^m$. The map $p \mapsto E^u(p)$ is a C^r map from Γ into the Grassmann manifold of linear subspaces of the appropriate dimension. The proof relies on the fact that $T_\Gamma \Phi_t$ maps $p \times E^u(p)$ to $\Phi_t(p) \times E^u(\Phi_t(p))$ and that $T_\Gamma \Phi$ is a C^r flow. The projection map

$$H_0 : E^u(\Gamma) \to \Gamma, \ (p, v) \mapsto p$$

is C^r of rank 1. Similar remarks apply to $E^s(\Gamma)$ and $E^\Phi(\Gamma)$.

For $p \in \Gamma$ and sufficiently small $\epsilon > 0$, the *local stable manifold of* p is defined to be the set:

$$W^s_\epsilon(p) = \{x \in \mathbb{R}^m : \forall t \geq 0 \, ||\Phi_t(x) - \Phi_t(p)|| \leq \epsilon, \text{ and } \lim_{t \to \infty} ||\Phi_t(x) - \Phi_t(p))|| = 0\}.$$

Using stable manifold theory (e.g. Hirsch and Pugh (1970), Hirsch, Pugh and Shub (1977); Hartman 1964), we take ϵ small enough so that $W^s_\epsilon(p)$ is a C^r (embedded) submanifold of $\mathbb{R}^m$, whose tangent space at p is $T_p W^s_\epsilon(p) = E^s(p)$.

It follows from the definition that:

(i) $W^s_\epsilon(p)$ is disjoint from $W^s_\epsilon(q)$ if $p \neq q$.
(ii) If $q = \Phi_t(p)$ then $W^s_\epsilon(q) = \Phi_t W^s_\epsilon(p)$.

We set $W^s_\epsilon(\Gamma) = \bigcup_{p \in \Gamma} W^s_\epsilon(p)$. This is the *local stable manifold of* Γ. It follows from the latter that

(i) $W^s_\epsilon(\Gamma)$ is a C^r submanifold, with $T_p W^s_\epsilon(\Gamma) = E^s(p) \oplus E^\Phi(p)$.
(ii) $W^s_\epsilon(\Gamma)$ is invariant under Φ_t for $t \geq 0$.
(iii) The map $\phi : W^s_\epsilon(\Gamma) \to \Gamma$ defined by $\phi(W^s_\epsilon(p)) = p$, is C^r of rank 1; in fact (ϕ, $W^s_\epsilon(\Gamma)$, Γ) is a C^r fibre bundle.

It is known that

$$W^s_\epsilon(\Gamma) = \{z \in \mathbb{R}^m : \forall t \geq 0, \, d(\Phi_t(z), \Gamma)) < \epsilon\}.$$

This implies for each $z \in \mathbb{R}^m$: either the trajectory of $\Phi_t(z)$ eventually stays at a distance $\geq \epsilon$ from Γ, or else there is a unique $p \in \Gamma$ such that $\Phi_s(z) \in W^s_\epsilon(p)$ for some $s \geq 0$, in which case $||\Phi_t(z) - \Phi_t(p)|| = O(e^{-\alpha t})$, where $\alpha > 0$ is the same as in (6).

The pullback (Hirsch 1976) of the C^r vector bundle $H_0 : E^u(\Gamma) \to \Gamma$ by the C^r map $\phi : W^s_\epsilon(\Gamma) \to \Gamma$ is a C^r vector bundle $H_1 : Y \to, W^s_\epsilon(\Gamma)$ where

$$Y = \{(p, v, x) : p \in \Gamma, x \in W^s_\epsilon(p),\ v \in E^u(p)\},$$

$$H_1((p, v, x)) = (p, x).$$

We can identify Γ with the set $\Gamma_0 \subset Y$ comprising all points $(p, 0, 0)$, $p \in \Gamma$. We can also identify $E^u(\Gamma)$ with $H_1^{-1}\Gamma_0$, equating (p, x) with $(p, 0, x)$. And we can identify $W^s_\epsilon(\Gamma)$ with the subset

$$\{(p, 0, x) : p \in \Gamma,\ x \in W^s_\epsilon(p) \subset Y\}.$$

With these identifications we observe that the two restricted flows $T_\Gamma\Phi|E^u(\Gamma)$ and $\Phi|W^s_\epsilon(\Gamma)$, which are C^r and agree on Γ, extend to a C^r flow Ψ in Y, defined by

$$\Psi_t(p, v, x) = (\Phi_t(p), D\Phi_t(p)v,\ \Phi_t(x)).$$

We define a map

$$H_2 : Y \to \mathbb{R}^m,\quad (p, v, x) \mapsto x + v.$$

Then H_2 is C^r because x and v are C^r functions of $(p, v, x) \in Y$.

The tangent map of H_2 at a point $(p, 0, 0) \in \Gamma_0$ is easily computed to be invertible. The inverse function theorem implies that H_2 is a C^r local diffeormorphism at each point of Γ_0. Since H_2 maps Γ_0 to Γ by the diffeomorphism $(p, 0, 0) \mapsto p$, it follows that H_2 restricts to a C^r diffeomorphism

$$H : \mathcal{M}(\Gamma_0) \to \mathcal{N}(\Gamma)$$

between open neighborhoods $\mathcal{M}(\Gamma_0) \subset Y$ of Γ_0 and $\mathcal{N}(\Gamma) \subset \mathbb{R}^m$ of Γ.

Now define maps

$$\pi_1 : \mathcal{N}(\Gamma) \to \Gamma,\ \ \pi : \mathcal{N}(\Gamma) \to W^s_\epsilon(\Gamma)$$

as follows. For $z = H(p, v, x) \in \mathcal{M}(\Gamma_0)$,

$$\pi_1(z) = p,\ \ \pi(z) = x.$$

For $p \in \Gamma$ define $M_\epsilon(p) = \{x + v : x \in W^s_\epsilon(p), v \in E^u(p)\}$. For $z \in \mathcal{N}(\Gamma)$ we set $E^u(z) = E^u(\pi_1(z))$.

LEMMA 3.1. *:*

(i) $\pi_1^{-1}(p) = \mathcal{N}(\Gamma) \cap M_\epsilon(p)$ *for all* $p \in \Gamma$.

(ii) $\pi^{-1}(x) = \mathcal{N}(\Gamma) \cap (x + E^u(x))$ *for all* $x \in W^s_\epsilon(\Gamma)$.

(iii) $\pi_1 \circ \pi = \pi_1$.

(iv) *For all* $z \in \mathcal{N}(\Gamma)$*:*

$$\mathbb{R}^m = \operatorname{Ker} D\pi(z) \oplus \operatorname{Ker}(\mathrm{Id} - D\pi(z)),$$

and

$$\operatorname{Ker} D\pi(z) = E^u(z).$$

(v) π_1 *and* π *are* C^r.

Proof. Follows from the definitions. □

A Riemannian metric and the function η. Fix $l > 0$ and define a Riemannian metric $||\cdot||_x$ on $\mathbb{R}^m$ by

$$\forall x, v \in \mathbb{R}^m : ||v||_x = \sqrt{\int_0^l ||D\Phi_{-t}(x)v||^2 dt}.$$

The following dynamical estimates are essential. Notice that part (iii) of the next lemma is just a continuous time version of a classical result for diffeormorphisms acting on hyperbolic sets (Hirsch and Pugh, 1970).

LEMMA 3.2. *The Riemannian metric $||\cdot||_x$ defined above satisfies the following conditions:*
(i) *For any $l > 0$ there exist constants $c_1 > 0, c_2 > 0$ such that:*

$$\forall x \in \mathcal{N}(\Gamma),\ \forall v \in \mathbb{R}^m,\ c_1||v|| \le ||v||_x \le c_2||v||.$$

(ii) *For any $l > 0$ there exists $c_3 > 0$ such that:*

$$\forall x, y \in \mathcal{N}(\Gamma),\ \forall v \in \mathbb{R}^m, |\ ||v||_x - ||v||_y\ | \le c_3||v||.$$

(iii) *For any sufficiently large $l > 0$ there exists $\beta > 0$ such that:*

$$\forall p \in \Gamma,\ \forall v \in E^u(p),\ ||D\Phi_t(p)v||_{\Phi_t(p)} \ge e^{\beta t}||v||_p,\ t > 0.$$

Proof. (i) Let

$$c_1^2 = \inf_{x \in \mathcal{N}(\Gamma)} \int_0^l \frac{1}{||D\Phi_t(x)^{-1}||^2} dt$$

and

$$c_2^2 = \sup_{x \in \mathcal{N}(\Gamma)} \int_0^l ||D\Phi_t(x)||^2 dt.$$

The estimates

$$||D\Phi_t(x)||\,||v|| \ge ||D\Phi_t(x)v|| \ge \frac{1}{||D\Phi_t(x)^{-1}||}||v||$$

imply (i).

(ii) By the Minkowsky inequality, we have

$$|\ ||v||_x - ||v||_y\ |^2 \le \int_0^l ||(D\Phi_{-t}(x) - D\Phi_{-t}(y))v||^2.$$

However, since the flow is smooth, there exists $L > 0$ such that for all $x, y \in \mathcal{N}(\Gamma)$ and all $t \in [0, l]$

$$||D\Phi_{-t}(x) - D\Phi_{-t}(y)|| \le L||x - y||.$$

Then we obtain (ii) with $c_3 = \sqrt{l}L$.

(iii) We work with the reversing flow $\Psi_t = \Phi_{-t}$. Let $p \in \Gamma$ and $v \in E^u(p)$. For $s > 0$, the chain rule implies

$$\begin{aligned} ||D\Psi_s(p)v||^2_{\Psi_s(p)} &= \int_0^l ||D\Psi_{t+s}(p)v||^2 dt = \int_s^{s+l} ||D\Psi_t(p)v||^2 dt \\ &= ||v||_p^2 + \int_l^{s+l} ||D\Psi_t(p)v||^2 dt - \int_0^s ||D\Psi_t(p)v||^2 dt \end{aligned}$$

Directions that are expanding for Φ are contracting for Ψ, and thus we have

$$||D\Psi_t(p)v||^2_{\Psi_t(p)} \leq c^2e^{-2\alpha t}||v||,$$

where α and c are the positive constants from (6). Therefore,

$$||D\Psi_s(p)v||^2_{\Psi_s(p)} \leq ||v||^2_p + sc^2e^{-2\alpha l}||v||^2 - \int_0^s ||D\Psi_t(p)v||^2 dt.$$

From this we get the inequality

$$\lim_{s\to 0,\, s>0} \frac{||D\Psi_s(p)v||^2_{\Psi_s(p)} - ||v||^2_p}{s} \leq -||v||^2(1 - c^2e^{-2\alpha l})$$

We now choose l large enough such that $(1 - c^2e^{-2\alpha l}) > 0$. Using assertion (i) of this lemma, we obtain

$$||D\Psi_s(p)v||_{\Psi_s(p)} \leq e^{-\beta s}||v||_p$$

with $\beta = (1 - c^2e^{-2\alpha s})2c_2 > 0$. As $\Psi_s = \Phi_{-s}$, this proves (iii). ☐

We now define a function η that measures the distance from the local stable manifold in the expanding direction:

$$\eta : \mathcal{N}(\Gamma) \to \mathbb{R},$$
$$\eta(x) = ||x - \pi(x)||_{\pi_1(x)}.$$

Recall that a function Ξ from a Banach space X into R is said to have a right derivative at $x \in X$ if for all $v \in X$ the limit $\lim_{t\to 0,\, t>0} \Xi(x + tv) - \Xi(x)/t$ exists. In this case we denote the limit by $D\Xi(x)v$. The resulting map $D\Xi(x) : X \to \mathbb{R}$ is called the *right derivative of* Ξ *at* x. If Ξ is differentiable at x, this is the same as the usual Frechet derivative of Ξ at x.

PROPOSITION 3.1. *The map* $\eta : \mathcal{N}(\Gamma) \to \mathbb{R}$ *enjoys following properties:*

(i) η *is* C^r *on* $\mathcal{N}(\Gamma) - W^s_\epsilon(\Gamma)$.

(ii) *For all* $x \in \mathcal{N}(\Gamma) \cap W^s_\epsilon(\Gamma)$, *the right derivative* $D\eta(x) : \mathbb{R}^m \to \mathbb{R}$ *is continuous, convex and positively homogeneous.*

(iii) *There exists* $k > 0$ *and a neighborhood* $U \subset \mathbb{R}^m$ *of* 0 *such that*

$$\forall x \in \mathcal{N}(\Gamma), \forall v \in U : \ \ \eta(x + v) \geq \eta(x) + D\eta(x)v - k||v||^2,$$

where $D\eta(x)$ *denotes the derivative at* x *when* $x \notin W^s_\epsilon(\Gamma)$, *and the right derivative at* x *when* $x \in W^s_\epsilon(\Gamma)$.

(iv) *Let* $x \in \mathcal{N}(\Gamma)$. *If* $x \notin W^s_\epsilon(\Gamma)$ *then*

$$||\nabla\eta(c)|| \geq c_1.$$

If $x \in W^s_\epsilon(\Gamma)$ *then*

$$\forall v \in \mathbb{R}^m : \ \ D\eta(x)v \geq c_1||v - D\pi(x)v||$$

(v) *For all* $x \in \mathcal{N}(\Gamma)$ *and* $v \in \mathbb{R}^m$:

$$D\eta(x)(F(x) + v) = D\eta(x)F(x) + D\eta(x)v$$

(vi) *There exists an open neighborhood $\mathcal{N}'(\Gamma) \subset \mathcal{N}(\Gamma)$ of Γ, and $\lambda > 0$, such that*

$$\forall x \in \mathcal{N}'(\Gamma), \forall v \in \mathbb{R}^m : \; D\eta(x)v \geq \lambda\eta(x).$$

Proof.

(i) is obvious.

(ii) Let $X = C^0([0,l], R)$ be the Banach space of continuous functions on $[0,l]$ with the uniform norm. Define $\Xi : X \to R$ by $\Xi(f) = \sqrt{\int_0^l \|f(t)\|^2}$. The function Ξ is convex and continuous (with Lipschitz constant $\sqrt{l}$). Therefore it admits a right derivative $D\Xi(f)$ at any point $f \in X$, which is convex continuous and positively homogeneous. We also have the standard inequality:

$$\forall h \in X : \; \Xi(f) - \Xi(f-h) \leq D\Xi(f)h \leq \Xi(f+h) - \Xi(f). \tag{7}$$

On the other hand, it is easy to verify that for $f \neq 0$ we have

$$D\Xi(f)h = \frac{\int_0^l \langle f(t), h(t)\rangle dt}{\Xi(f)}, \tag{8}$$

and

$$D\Xi(0)h = \Xi(h). \tag{9}$$

Let $g : \mathcal{N}(\Gamma) \to X$ denote the map defined by

$$g(z)(t) = D\Phi_{-t}(\pi_1(z))(z - \pi(z)).$$

Then g is C^r because $T_\Gamma\Phi, \pi$ and π_1 are C^r, and:

$$\eta(z) = \Xi(g(z)).$$

Continuing the proof of (ii), make the following estimate:

$$\begin{aligned}\frac{\eta(x+sv) - \eta(x)}{s} &= \frac{\Xi(g(x) + sDg(x)v + O(s^2\|v\|^2)) - \Xi(g(x))}{s} \\ &= \frac{\Xi(g(x) + sDg(x)v) - \Xi(g(x))}{s} + O(s\|v\|^2),\end{aligned} \tag{10}$$

where the last equality holds because Ξ is Lipschitz. It is now clear that η admits a right derivative at x given by

$$v \mapsto D\eta(x)v = D\Xi(g(x))Dg(x)v.$$

The convexity (continuity, positive homogeneity) of $D\eta(x)$ follow from the convexity (continuity, positive homogeneity) of $D\Xi(g(x))$ and the linearity of $Dg(x)$.

(iii) In equation (10) we let $s = 1$ and we use equation (7).

(iv) Fix $x \in \mathcal{N}(\Gamma) - W^s_\epsilon(\Gamma)$ and let $v_x = [x - \pi(x)]/\eta(x)$. Since $v_x \in E^u(x)$ by (ii) of Lemma 3.1, assertions (iv) and (iii) of Lemma 3.1 imply $D\pi(x)v_x = D\pi_1(x)v_x = 0$. Therefore, for all $t \in [0,l]$

$$Dg(x)v_x(t) = D\Phi_{-t}(\pi_1(x))v_x = \frac{g(x)(t)}{\eta(x)} = \frac{g(x)(t)}{\Xi(g(x))}.$$

Thus,

$$D\eta(x)v_x = \frac{D\Xi(g(x))g(x)}{\Xi(g(x))}$$

and, using (8), we obtain $D\eta(x)v_x = 1$. On the other hand, The Cauchy-Schwartz inequality and assertion (i) of Lemma 3.2 give

$$D\eta(x)v_x = \langle \nabla\eta(x), v_x\rangle \leq \|\nabla\eta(x)\|\|v_x\| \leq \|\nabla\eta(x)\|\frac{\|v_x\|_{\pi_1(x)}}{c_1} \leq \|\nabla\eta(x)\|\frac{1}{c_1}.$$

This proves $\|\nabla\eta(x)\| \geq c_1$.

Now, suppose that $x \in W^s_\epsilon(\Gamma)$. We have $x = \pi(x)$ and consequently

$$Dg(x)v(t) = D\Phi_{-t}(\pi_1(x))(v - D\pi(x)v).$$

Using (9) and Lemma 3.2, (i) we obtain:

$$D\eta(x)v = \|v - D\pi(x)v\|_{\pi_1(x)} \geq c_1\|v - D\pi(x)v\|.$$

(v) This is obvious for $x \in \mathcal{N}(\Gamma) - W^s_\epsilon(\Gamma)$ because η is differentiable at x. If $x \in W^s_\epsilon(\Gamma)$, then $x = \pi(x)$ and $D\pi(x)F(x) = F(x)$. Therefore

$$Dg(x)(F(x) + v)(t) = D\Phi_{-t}(\pi_1(x))(v - D\pi(x)v)$$

and it follows from (9) that $D\eta(x)F(x) = 0$ and $D\eta(x)(F(x) + v) = D\eta(x)v$.

(vi) Let $\mathcal{N}_r(\Gamma) = \{x \in \mathbb{R}^m : d(x, \Gamma) < r\}$. Obviously, there exists $r_0 > 0$ and $t_0 > 0$ such that for all $r \leq r_0$ and all $0 \leq t \leq t_0$, $\mathcal{N}_r(\Gamma) \subset \mathcal{N}(\Gamma)$ and $\Phi_t(\mathcal{N}_r(\Gamma)) \subset \mathcal{N}(\Gamma)$. In the following it is implicitly assumed that $x \in \mathcal{N}_{r_0}(\Gamma)$ and $t \in [0, t_0]$. We have

$$\begin{aligned}\eta(\Phi_t(x)) &= \|\Phi_t(x) - \pi(\Phi_t(x))\|_{\pi_1(\Phi_t(x))} \\ &= \|A(t,x) + B(t,x) + C(t,x)\|_{\pi_1(\Phi_t(x))},\end{aligned}$$

where

$$\begin{aligned}A(t,x) &= \Phi_t(\pi(x)) - \pi(\Phi_t(x)), \\ B(t,x) &= D\Phi_t(\pi(x))(x - \pi(x)), \\ C(t,x) &= \Phi_t(x) - \Phi_t(\pi(x)) - D\Phi_t(\pi(x))(x - \pi(x)).\end{aligned}$$

The proof decomposes in three steps.

Step 1. For any $\delta > 0$ *there exists* $r_1(\delta)$ *such that* $0 < r_1(\delta) < r_0$, *and if* $0 < r < r_1(\delta)$ *and* $\in \mathcal{N}_r(\Gamma)$, *then:*

$$\|A(t,x)\|_{\pi_1(\Phi_t(x))} \leq \delta t\eta(x).$$

Proof. Since the flow ϕ is C^2 and the second derivatives $D^2\Phi_t(x)$ are continuously differentiable in t (see beginning of §3), the same hold for A. Noting that $A(0,x) = 0$, we then have:

$$A(t,x) = A(t,\pi(x)) + \frac{\partial A}{\partial x}(t,\pi(x))(x - \pi(x)) + O(t\|x - \pi(x)\|^2).$$

On the other hand $A(t,\pi(x)) = 0$ because $W^s_\epsilon(\Gamma)$ is an invariant manifold and $\pi = \mathrm{Id}$ on $W^s_\epsilon(\Gamma) \cap \mathcal{N}(\Gamma)$. It follows that

$$\|A(t,x) - \frac{\partial A}{\partial x}(t,\pi(x))(x - \pi(x))\| = O(t\|x - \pi(x)\|^2). \tag{11}$$

It remains to estimate $\frac{\partial A}{\partial x}(t,\pi(x))(x-\pi(x))$. Since $x-\pi(x)\in E^u(x)$ $D\pi(\pi(x))(x-\pi(x))=0$ by Lemma 3.1, (iv) and by derivation of A we obtain

$$\begin{aligned}&\frac{\partial A}{\partial x}(t,\pi(x))(x-\pi(x))\\ &\quad= -D\pi[\Phi_t(\pi(x))]D\Phi_t(\pi(x))(x-\pi(x))\\ &\quad= -D\pi[\Phi_t(\pi(x))][D\Phi_t(\pi_1(x))(x-\pi(x))\\ &\qquad +D\Phi_t(\pi(x))(x-\pi(x))-D\Phi_t(\pi_1(x))(x-\pi(x))].\end{aligned}$$

By invariance of the splitting $E^s\oplus E^u\oplus E^\Phi$ under the tangent flow, the vector $v=D\Phi_t(\pi_1(x))(x-\pi(x))$ belongs to $E^u(\Phi_t(\pi_1(x))$. Thus, using Lemma 3.1, (iv) again, we get $D\pi[\Phi_t(\pi(x))]v=0$, and consequently,

$$\begin{aligned}&\left\|\frac{\partial A}{\partial x}(t,\pi(x))(x-\pi(x))\right\|\\ &\quad= O(\|D\Phi_t(\pi(x))(x-\pi(x))-D\Phi_t(\pi_1(x))(x-\pi(x))\|).\end{aligned}\tag{12}$$

Since Φ is C^2, $D^2\Phi_t(x)$ is continuously differentiable in t and $D\Phi_0(x)=\mathrm{Id}$, the following estimate holds

$$\begin{aligned}&\|D\Phi_t(\pi(x))(x-\pi(x))-D\Phi_t(\pi_1(x))(x-\pi(x))\|\\ &\quad= O(t\|\pi(x)-\pi_1(x)\|\|x-\pi(x)\|).\end{aligned}\tag{13}$$

Putting (11), (12) and (13) together, and using Lemma 3.2, (i) we see that

$$\begin{aligned}\|A(t,x)\|_{\pi_1(\Phi_t(x))} &= O\left(t\|x-\pi(x)\|_{\pi_1(x)}[\|\pi(x)-\pi_1(x)\|+\|x-\pi(x)\|]\right)\\ &= O\left(t\eta(x)[\|\pi(x)-\pi_1(x)\|+\|x-\pi(x)\|]\right)\end{aligned}$$

Since $\pi|_\Gamma=\pi_1|_\Gamma=\mathrm{Id}$, the claim is proved.

Step 2. Let β be the positive constant given by Lemma 3.2, (iii). There exists $0<r_2<r_0$ *such that if* $0<r<r_2$ *and* $x\in\mathcal{N}_r(\Gamma)$, *then:*

$$\|B(t,x)\|_{\pi_1(\Phi_t(x))}\geq\left(1+\frac{\beta}{2}t\right)\eta(x)$$

Proof. From the definition of $B(t,x)$ we have

$$B(t,x)=D\Phi_t(\pi_1(x))(x-\pi(x))+[D\Phi_t(\pi(x))(x-\pi(x))-D\Phi_t(\pi_1(x))(x-\pi(x))].$$

From this, using Lemma 3.2, (iii), estimate (13) and Lemma 3.2, (i) we deduce

$$\|B(t,x)\|_{\Phi_t(\pi_1(x))}\geq e^{\beta t}\eta(x)+O(t\eta(x)\|\pi(x)-\pi_1(x)\|).$$

Therefore, for r small enough, $x\in\mathcal{N}_r(\Gamma)$ and $t\leq t_0$ we have

$$\|B(t,x)\|_{\Phi_t(\pi_1(x))}\geq e^{\beta t}\eta(x)-\frac{\beta}{4}t\eta(x)\geq\left(1+\frac{3}{4}\beta\right)\eta(x).\tag{14}$$

Now, using Lemma 3.2, (ii) and the definition of $B(t,x)$, we see that

$$\begin{aligned} &|\ \|B(t,x)\|_{\pi_1(\Phi_t(x))} - B(t,x)\|_{\Phi_t(\pi_1(x))}\ | \\ &\quad = O\left(\|\pi_1(\Phi_t(x)) - \Phi_t(\pi_1(x))\|.\|x-\pi(x)\|\right) \\ &\quad = O\left(\|\pi_1(\Phi_t(x)) - \Phi_t(\pi_1(x))\|\eta(x)\right). \end{aligned}$$

Let $A_1(t,x) = \pi_1(\Phi_t(x)) - \Phi_t(\pi_1(x))$. Then A_1 is C^2 ; and $A_1(0,x) = A_1(t,\pi_1(x)) = 0$. Thus $A_1(t,x) = O(t\|x-\pi(x)\|)$. It follows that for r small enough and $x \in \mathcal{N}_r(\Gamma)$ we have

$$|\ \|B(t,x)\|_{\pi_1(\Phi_t(x))} - B(t,x)\|_{\Phi_t(\pi_1(x))}\ | \leq \frac{\beta}{4}\eta(x). \tag{15}$$

From (14) and (15) we deduce the claim.

Step 3: For any $\delta > 0$ there exists $r_3(\delta)$ such that $0 < r_3(\delta) < \delta$, and for $0 < r < r_3(\delta)$ and $x \in \mathcal{N}_r(\Gamma)$ we have:

$$\|C(t,x)\|_{\pi_1(\Phi_t(x))} \leq \delta t \eta(x).$$

Proof. It is clear that $C(t,x) = O(t\|x-\pi(x)\|^2)$. This proves the claim.

To conclude the proof of (vi) we choose $\delta < \beta/4$ and $r < \inf(r_1(\delta), r_2, r_3(\delta))$. Then we obtain

$$\forall x \in \mathcal{N}_r(\Gamma)\ \eta(\Phi_t(x)) \geq (1+\lambda t)\eta(x)$$

with $\lambda = \beta/2 - 2\delta$. It follows that $D\eta(x)F(x) \geq \lambda\eta(x)$ on $\mathcal{N}_r(\Gamma)$. □

4. *Proofs of main theorems*

Proof of Theorem 2.1. In this section the subscript n runs from 0 or 1 to ∞, as appropriate from the context.

The following lemma due to Pemantle (1992, Lemma 5.5) is the probabilistic key of the proof of Theorem 2.1.

LEMMA 4.1. (Pemantle, 1992.) *Let $\{S_n\}$ be a nonnegative stochastic process, $S_n = S_0 + \sum_{i=1}^n X_i$ where X_n is $\mathcal{F}_n$ measurable. Assume*

$$\frac{A}{n^\mu} \leq \gamma_n \leq \frac{B}{n^\mu}$$

where $0 < A \leq B$, $\frac{1}{2} < \mu \leq 1$. Assume there exist constants $b_1 > 0$, $b_2 > 0$, $b_3 > 0$ and an integer N_0 such that for all $n \geq N_0$:

(i) $|X_n| \leq b_1\gamma_n$.

(ii) $\mathbf{1}_{\{S_n > b_2\gamma_{n+1}\}}E(X_{n+1}|\mathcal{F}_n) \geq 0$.

(iii) $E(S_{n+1}^2 - S_n^2|\mathcal{F}_n) \geq b_3\gamma_{n+1}^2$.

Then $P(\lim_{n\to\infty} S_n = 0) = 0$.

This lemma is stated and proved in (Pemantle, 1992) for $A = 1$ and $\mu = 1$ but the proof adapts without difficulty to $A > 0$ and $\frac{1}{2} < \mu \leq 1$. The same result is proved, but not stated, in (Pemantle, 1990).

For $0 < \mu \leq 1/2$ we use the next lemma:

LEMMA 4.2. *Let $\{S_n\}$ be a stochastic process, $S_n = S_0 + \sum_{i=1}^n X_i$ where X_n is $\mathcal{F}_n$ measurable. Let $\{\gamma_n\}$ be such that $\sum_i \gamma_i^2 = \infty$. Assume there exists $b_3 > 0$ and some integer N_0 such that for all $n \geq N_0$ $E(S_{n+1}^2 - S_n^2 | \mathcal{F}_n) \geq b_3 \gamma_{n+1}^2$.*

Then $P(\lim_{n\to\infty} S_n = 0) = 0$.

Proof. Let $Z_n = S_n^2 - \sum_{i=0}^n b_3 \gamma_i^2$. We have $E(Z_{n+1} - Z_n | \mathcal{F}_n) \geq 0$. Thus, $\{(Z_n, \mathcal{F}_n)\}_{n\geq N_0}$ is a submartingale.

Suppose $P(\lim_{n\to\infty} S_n = 0) > 0$. Then for all $\epsilon > 0$ there exists $N \geq N_0$ such that $P(\bigcap_{n\geq N}\{| S_n |\leq \epsilon\}) > 0$.

Assume $| S_N | \leq \epsilon$ and define the stopping time $T = \inf\{k \geq N; | S_k | > \epsilon\}$. As usual,we let $n \wedge T = \inf(n, T)$. The sequence $\{(Z_{n\wedge T}, \mathcal{F}_n)\}_{n\geq N}$ is a submartingale and we have $Z_{n\wedge T} \leq \epsilon^2$. It follows from the submartingale convergence theorem that $\{Z_{n\wedge T}\}_{n\geq N}$ converges almost surely. Thus $\{\sum_{i=0}^{n\wedge T} b_3 \gamma_i^2\}_{n\geq N}$ is almost surely bounded. This implies $T < \infty$ almost surely. ☐

We now prove Theorem 2.1. Here we closely follow Pemantle (1990, 1992).

Choose $r > 0$ such that $\{x \in \mathcal{N}(\Gamma);\ \eta(x) < r\} \subset \mathcal{N}'(\Gamma)$ where $\mathcal{N}'(\Gamma)$ is the neighborhood given in Proposition 3.1, (vi).

Let N be an integer. Assume $\eta(x_N) < r$ and let T be the stopping time defined by

$$T = \inf\{k \geq N;\ \eta(x_k) \geq r\}.$$

We prove Theorem 2.1 by showing that $P(T < \infty) = 1$.

Without loss of generality we assume $N = 0$. (The proof is the same for any N.)

Define two sequences of random variables $\{X_n\}_{n\geq 1}$ and $\{S_n\}$ as follows:

$$X_{n+1} = [\eta(x_{n+1}) - \eta(x_n)]\mathbf{1}_{\{n\leq T\}} + \gamma_{n+1}\mathbf{1}_{\{n>T\}},$$

$$S_0 = \eta(x_0),\ S_n = S_0 + \sum_{i=1}^n X_i.$$

The process $\{S_n\}$ is clearly nonnegative. Notice that if $T = \infty$ then $X_{n+1} = \eta(x_{n+1}) - \eta(x_n)$ and S_n telescopes into $S_n = \eta(x_n)$. This will be used at the end of the proof.

We now verify that hypotheses (i), (ii) and (iii) of Lemma 4.1 are satisfied.

Condition (i) follows from the Lipschitz continuity of η and the boundness of the sequences $\{F(x_n)\}$, $\{U_n\}$.

Condition (ii). Let $k' = k(||F|| + K)$ where k is given by Proposition 3.1, (iii), $||F|| = \sup\{F(x); x \in \Delta\}$ and K is the uniform bound of the U_n. If $n \leq T$, using Proposition 3.1, (ii), (iii), (v) and (vi) we have

$$\eta(x_{n+1}) - \eta(x_n) \geq \gamma_{n+1}\lambda\eta(x_n) + \gamma_{n+1}D\eta(x_n)U_{n+1} - k'\gamma_{n+1}^2. \tag{16}$$

Thus

$$\mathbf{1}_{\{n\leq T\}}E(X_{n+1}|\mathcal{F}_n) \geq \mathbf{1}_{\{n\leq T\}}\left[(\gamma_{n+1}\lambda\eta(x_n) - k'\gamma_{n+1}^2) + \gamma_{n+1}E(D\eta(x_n)U_{n+1}|\mathcal{F}_n)\right].$$

By convexity of the right derivative of η (Proposition 3.1, (ii)) and the conditional Jensen inequalitity we have

$$E(D\eta(x_n)U_{n+1}|\mathcal{F}_n) \geq D\eta(x_n)E(U_{n+1}|\mathcal{F}_n) = 0.$$

Thus

$$\mathbf{1}_{\{n\leq T\}}E(X_{n+1}|\mathcal{F}_n) \geq \mathbf{1}_{\{n\leq T\}}(\gamma_{n+1}\lambda\eta(x_n) - k'\gamma_{n+1}^2). \tag{17}$$

If $n > T$, $X_{n+1} = \gamma_{n+1}$, so

$$\mathbf{1}_{\{n>T\}}E(X_{n+1}|\mathcal{F}_n) \geq \mathbf{1}_{\{n>T\}}\gamma_{n+1} \geq 0. \tag{18}$$

Putting (17) and (18) together and letting $b_2 = k'/\lambda$ proves condition (ii) of Lemma 4.1.

For Condition (iii) of Lemma 4.1, we observe that

$$E(S_{n+1}^2 - S_n^2|\mathcal{F}_n) = E(X_{n+1}^2|\mathcal{F}_n) + 2S_nE(X_{n+1}|\mathcal{F}_n).$$

If $S_n \geq b_2\gamma_{n+1}$, the right-hand term is nonnegative by condition (ii), previously proved. If $S_n < b_2\gamma_{n+1}$, (17) and (18) imply $S_nE(X_{n+1}|\mathcal{F}_n) \geq -b_2k'\gamma_{n+1}^3$. Thus

$$E(S_{n+1}^2 - S_n^2|\mathcal{F}_n) \geq E(X_{n+1}^2|\mathcal{F}_n) - 2b_2k'\gamma_{n+1}^3.$$

Therefore, to prove condition (iii) of Lemma 4.1, it suffices to show that

$$E(X_{n+1}^2|\mathcal{F}_n) \geq b_4\gamma_{n+1}^2$$

for some $b_4 > 0$ and n large enough. From (16) we deduce

$$\mathbf{1}_{\{n\leq T\}}\left(E(X_{n+1}^+|\mathcal{F}_n) - \left[\gamma_{n+1}E((D\eta(x_n)U_{n+1})^+|\mathcal{F}_n) - k'\gamma_{n+1}^2\right]\right) \geq 0. \tag{19}$$

Using Proposition 3.1, (iv) and assumption (iii) of Theorem 2.1 we see that

$$\mathbf{1}_{\{n\leq T\}\cap\{x_n\notin W_\epsilon^s(\Gamma)\}}(E((D\eta(x_n)U_{n+1})^+|\mathcal{F}_n) - c_1b) \geq 0. \tag{20}$$

If $x_n \in W_\epsilon^s(\Gamma)$, choose a unit vector $v_n \in \mathrm{Ker}(\mathrm{Id} - D\pi(x_n))^\perp$. By Lemma 3.1, (iv) we have

$$< U_{n+1}, v_n >=< U_{n+1} - D\pi(x_n)U_{n+1}, v > .$$

Let $\mathcal{A}$ denotes the event $\mathcal{A} = \{n \leq T\} \cap \{x_n \in W_\epsilon^s(\Gamma)\}$. By using Proposition 3.1 (iv), the Cauchy–Schwartz inequality and assumption (iii) of Theorem 2.1 we obtain

$$\begin{array}{rcl}
\mathbf{1}_{\mathcal{A}}E((D\eta(x_n)U_{n+1})^+|\mathcal{F}_n) & \geq & \\
c_1\mathbf{1}_{\mathcal{A}}E(||U_{n+1} - D\pi(x_n)U_{n+1}||\;|\mathcal{F}_n) & \geq & \\
c_1\mathbf{1}_{\mathcal{A}}E(\langle U_{n+1} - D\pi(x_n)U_{n+1}, v_n\rangle^+|\mathcal{F}_n) & = & \\
c_1\mathbf{1}_{\mathcal{A}}E(\langle U_{n+1}, v_n\rangle^+|\mathcal{F}_n) & \geq & c_1b\mathbf{1}_{\mathcal{A}}.
\end{array}$$

Putting (19), (20), (21) together and (18) give

$$E(X_{n+1}^+|\mathcal{F}_n) \geq \gamma_{n+1}c_1b - k'\gamma_{n+1}^2$$

On the other hand $E(X_{n+1}^2|\mathcal{F}_n) \geq E(X_{n+1}^+|\mathcal{F}_n)^2$ by the Jensen inequality. It follows that $E(X_{n+1}^2|\mathcal{F}_n) \geq b_4\gamma_{n+1}^2$ for $b_4 = (c_1b)^2$ and n large enough, as is desired.

Condition (i) through (iii) of Lemma 4.1 being satisfied, *the probability is zero that* $\{S_n\}$ *converges to zero,* according to Lemma 4.1 for $\frac{1}{2} < \mu < 1$ and Lemma 4.2 for $0 < \mu \leq \frac{1}{2}$.

Now suppose $T = \infty$. Then

$$\eta(x_n) = S_n$$

and $\{x_n\}$ remains in $\mathcal{N}'(\Gamma)$. Therefore by Corollary 1.2,

$$L(\{x_n\}) \subset \overline{\mathcal{N}'(\Gamma)} \cap R(F) = \Gamma.$$

Moreover $L(\{x_n\})$ is a nonempty compact invariant set, so that $L(\{x_n\}) = \Gamma$. This implies $\eta(x_n) \to 0$, and therefore $S_n \to 0$ because $S_n = \eta(x_n)$ when $T = \infty$. Since $P(S_n \to 0) = 0$, T is almost surely finite. □

Proof of Theorem 2.5. The proof is based on a coupling argument comparable with those used for $m = 1$ by Hill, Lane and Sudderth (1980), based on the following topological lemma:

LEMMA 4.3. *Let f be a C^r $(r \geq 1)$ urn function with $f(\Delta^m) \subset \mathrm{Int}(\Delta^m)$. Let $\Lambda \subset \mathrm{Int}(\Delta^m)$ be an attractor for the vector field $F = -\mathrm{Id} + f$ on $\mathbb{R}^m$, with basin of attraction B. Let $W \subset B$ be a neighborhood of Λ. Then there exist a neighborhood $V \subset W$ of Λ, an urn function $g : \Delta^m \to \Delta^m$ and a C^∞ vector field G on $\mathbb{R}^m$ such that:*

(i) $g(\Delta^m) \subset \mathrm{Int}(\Delta^m)$

(ii) $G|_{\Delta^m} = -\mathrm{Id} + g$ *and* $G|_V = F|_V$.

(iii) *The chain recurrent set* $R(G) = \Lambda \cup \Lambda_0$, *where* Λ_0 *is a finite set of hyperbolic equilibria, none of which attracting or repelling.*

(iv) *The closure of* V *is a compact* C^∞ *manifold and* $G|_{\mathbb{R}^m \setminus V}$ *is Morse–Smale.*

Proof. Wilson (1969) constructs a C^∞ function u on B which is nonnegative and has negative derivative along forward trajectories of F in $B \setminus \Lambda$. It follows that for sufficiently small $c > 0$, the set $V = \{x : u(x) \leq c\}$ is a C^∞ compact submanifold $V \subset W$ to whose boundary F is transverse, pointing to $\mathrm{Int}(V)$. We fix such a V.

The hypothesis that $f(\Delta^m) \subset \mathrm{Int}(\Delta^m)$ implies that at every boundary point $z \in \partial\Delta^m$, $F(z)$ is transverse to $\partial\Delta^m$ and points into $\mathrm{Int}(\Delta^m)$. The vector field defined by $J(z) = -z$ has the same property. We glue together F and J to obtain a vector field F^0 which agrees with J near $\partial\Delta^m$ and with F on a neighborhood of V. Explicitly, $F^0(z) = (1 - r(z))(-z) + r(z)F(z)$ where the C^∞ real-valued function r is identically 1 in a neighborhood of $\partial\Delta^m$ and identically 0 in a neighborhood of V. Replacing F by F^0, we assume that $F(z) = -z$ in a neighborhood of $\mathbb{R}^m \setminus \mathrm{Int}(\Delta^m)$.

By another glueing argument, we assume that in a neighborhood W_1 of ∂V, F has the form ∇h_0, the gradient of a C^∞ function h_0 with the following properties:

(a) ∂V is a level surface of h_0, and $h_0(\partial V) = 0$,

(b) h_0 has no critical points on ∂V,

(c) $h_0 > 0$ in $W \setminus V$.

Let h be a C^∞ function defined in a neighborhood of $\mathbb{R}^m \setminus \mathrm{Int}(V)$, agreeing with $||z||^2/2$ in some neighborhood of $\mathbb{R}^m \setminus \mathrm{Int}(\Delta^m)$ and with h_0 in a neighborhood of ∂V. Choose h to be > 0 in $\mathbb{R}^m \setminus \mathrm{Int}(V)$.

Let $\alpha > 0$ be so large that the interior of the closed ball $B_\alpha \subset \mathbb{R}^m$, of radius α and centered at the origin, contains Δ^m, and also so that $h(z) < \alpha^2/2$ for all $z \in W \cap V$.

Let $M \subset \mathbb{R}^m$ denote compact the C^∞ submanifold

$$M = B_\alpha \setminus \mathrm{Int}(V).$$

Then $h|M : M \to \mathbb{R}$ is a nonnegative C^∞ function. Its maximum value $< \alpha^2/2$ is taken on the boundary component ∂B_α; while its minimum value 0 is taken on the other boundary component ∂V. These boundary components are noncritical level surfaces of h.

There is a Morse function on M coinciding with h in a neighborhood of ∂M. Thus we may assume h is a Morse function.

We make nontrivial changes in h as follows:

(a) Applying the cancellation method of Morse (1960) and Smale (1961a), we assume that *h has no local maximum or minimum points in* $\mathrm{Int}(M)$.

(b) Applying the approximation theorem of S. Smale (1961b), we assume that *∇h is a Morse–Smale vector field.*

Next we extend the $-\nabla h$ over $\mathbb{R}^m$, setting it equal to $-z = -\nabla ||z||^2/2$ at z outside B_α, and equal to F in $\mathrm{Int}(V)$. The resulting vector field G^1 is Morse–Smale on $\mathbb{R}^m \setminus V$.

Now let $\sigma : \mathbb{R}^m \to \mathbb{R}$ be a positive C^∞ function, equal to 1 within a compact neighborhood V_1 of Λ in $\mathrm{Int}(V)$, such that $0 < \sigma(z) < 1$ for $z \in V_1 \setminus \mathrm{Int}(V)$. For each $t > 0$ define the vector field G^t on $\mathbb{R}^m$ by $G^t(z) = \sigma(z)^t G^1(z)$. The orbits of G^t are reparametrizations of those of G^1. It follows that *G^t is a dissipative Morse–Smale gradient field outside V, for all $t > 0$.*

It is easy to see that for t sufficiently large, G^t has the property that the map $g^t = \mathrm{Id} + G^t$ maps Δ^m into $\mathrm{Int}(\Delta^m)$. Fix one such t and set $g = g^t$. Now define the vector field $G = \mathrm{Id} + g$.

Outside V this vector field G is a gradient with a finite set of equilibrium, and V lies in the basin of attraction of Λ. This implies that the chain recurrent set of G consists of equilibria and Λ. This completes the proof of Lemma 4.3. ☐

Continuing the proof of Theorem 2.5, let $\Lambda \subset \mathrm{Int}(\Delta^m)$ be an attractor for F. We need to prove $\mathbf{p}(\Lambda) > 0$.

The hypothesis that $f(\Delta^m) \subset \mathrm{Int}(\Delta^m)$ implies that the flow $\{\Phi\}_t$ of vector field $F = -\mathrm{Id} + f$ takes Δ^m into $\mathrm{Int}(\Delta^m)$ for all $t > 0$.

We suppose the initial urn composition given by (x_0, n_0). Define another urn process $\{y_n\}$, with initial composition $(y_0, n_0) = (x_0, n_0)$ and urn function $g = G + \mathrm{Id}$, where g and G are the urn function and the vector field given by Lemma 4.3. Notice that G is transverse to $\partial\Delta^m$ by (i) of Lemma 4.3.

By Lemma 4.3 (iii), Corollary 1.2 and Theorem 1.3 it is clear that

$$P_{(x_0,n_0)}(\lim_{n\to\infty} d(y_n, \Lambda) = 0) = 1.$$

Thus,

$$P_{(x_0,n_0)}\left(\left\{\lim_{n\to\infty} d(y_n, \Lambda) = 0\right\} \bigcap \left\{\bigcup_{k\geq 0}\left(\bigcap_{n\geq k}\{y_n \in V\}\right)\right\}\right) = 1,$$

where V is the neighborhood given by Lemma 4.3. Therefore, there exists $k \geq 0$ such that $P_{(x_0,n_0)}(A_k) > 0$ where A_k denotes the event

$$A_k = \{\lim_{n\to\infty} d(y_n, \Lambda) = 0\} \bigcap \left\{ \bigcap_{n\geq k} \{y_n \in V\} \right\}.$$

By the Markov property, we have

$$P_{(x_0,n_0)}(A_k) = E_{(x_0,n_0)}[(P_{(y_k,n_0+k)}(A_0)] > 0,$$

where $E_{(x_0,n_0)}$ denotes the expectation for the law $P_{(x_0,n_0)}$.

Define $V^k_{(x_0,n_0)}(g)$ as the set of all possible values for y_k (the state of the process at time k) when the initial urn composition is (x_0, n_0) and the urn function g. That is, $V^k_{x_0,n_0}(g)$ is defined recursively by: $V^0_{x_0,n_0}(g) = \{x_0\}$ and

$$V^{k+1}_{x_0,n_0}(g) = \{(1 - \gamma_{k+1})y + \gamma_k e_i : \ y \in V^k_{x_0,n_0}(g),\ i \in \{1, \ldots, m+1\},\ g_i(y) > 0\},$$

where $\gamma_k = \frac{1}{n_0+k}$.

As the random variable y_k takes value in $V^k_{(x_0,n_0)}(g)$, there exists a nonempty set $S \subset V^k_{(x_0,n_0)}(g)$ such that for all $y \in S$, $P_{(y,n_0+k)}(A_0) > 0$. Since $f = g$ on V we get

$$\forall y \in S,\ P_{(y,n_0+k)}(\lim_{n\to\infty} d(x_n, \Lambda) = 0) > 0.$$

The cardinality of S being finite, there exists $\delta > 0$ such that

$$\forall y \in S,\ P_{(y,n_0+k)}(\lim_{n\to\infty} d(x_n, \Lambda) = 0) > \delta.$$

Using again the Markov property we obtain:

$$\begin{aligned} P_{(x_0,n_0)}(\lim_{n\to\infty} d(x_n, \Lambda) = 0) &= E_{(x_0,n_0)}[P_{(x_k,n_0+k)}(\lim_{n\to\infty} d(x_n, \Lambda) = 0)] \\ &\geq E_{(x_0,n_0)}[\mathbf{1}_{\{x_k\in S\}} P_{(x_k,n_0+k)}(\lim_{n\to\infty} d(x_n, \Lambda) = 0)] \\ &\geq \delta P_{(x_0,n_0)}(x_k \in S). \end{aligned}$$

Because f and g both map Δ^m into $\mathrm{Int}(\Delta^m)$, it is obvious that $V^k_{x_0,n_0}(g) = V^k_{x_0,n_0}(f)$. Therefore, $P_{(x_0,n_0)}(x_k \in S) > 0$ and consequently, $P_{(x_0,n_0)}(\lim_{n\to\infty} d(x_n, \Lambda) = 0) > 0$. $\square$

Proof of Theorem 2.7. Statement (a) follows from Corollary 2.2. As for (b), Theorem 2.5 implies that $\mathbf{p}(\Gamma_i) > 0$ when Γ_i is linearly stable.

To prove the converse, we first verify the hypotheses of Pemantle's Theorem 1.3. Conditions (i) and (ii) are obviously satisfied.

For condition (iii) we define maps Y_j as follows. For each $j \in \{1, \ldots, m+1\}$, define

$$Y_j : \Delta^m \to E^m,$$

$$x \mapsto e_j - f(x)$$

where e_j is the jth vector of the canonical basis of $\mathbb{R}^{m+1}$.

Let $\Theta \in T\Delta^m = E^m$ be a unit vector. We have

$$E(\langle U_{n+1}, \Theta\rangle^+ | x_n = x) = \sum_{i=1}^{m+1} f_i(x) \langle Y_i(x), \Theta\rangle^+.$$

By compactness of $\Delta^m \times \{\Theta \in E^m : ||\Theta|| = 1\}$, condition (iii) of Theorem 2.1 reduces to:

$$\sum_{i=1}^{m+1} f_i(x)\langle Y_i(x), \Theta\rangle^+ > 0.$$

Suppose the contrary. Note that $f_i(x) > 0$, as $f(\Delta^m) \subset \mathrm{Int}(\Delta^m)$. Thus for all $i \in 1, \ldots, m+1$

$$\langle Y_i(x), \Theta\rangle = \Theta_i - \sum_{j=1}^{m+1} f_j(x)\Theta_j \leq 0.$$

Relabel the Θ_i so that $\Theta_1 \leq \ldots \leq \Theta_{m+1}$. As $\sum_{i=1}^{m+1} \Theta_i = 0$, it must be that $\Theta_1 < \Theta_{m+1}$, otherwise Θ would be zero. Thus we get the contradiction:

$$\Theta_{m+1} \leq \sum_{j=1}^{m+1} f_j(x)\Theta_j < \Theta_{m+1} \; \Box .$$

The proof of Theorem 2.7 is completed by applying Theorem 1.3 if Γ_i is an equilibrium, or Theorem 1.1 if it is a cycle.

Acknowledgement. M. Benaïm was supported by a grant from the Centre National de la Recherche Scientifique (Programme Cogniscience)

M. Hirsch was partially supported by a grant from the National Science Foundation.

REFERENCES

Arthur, B. W. (1988). Self-reinforcing mechanisms in economics. In *The Economy as an Evolving Complex System, SFI Studies in the Sciences of Complexity*, Anderson P. W, Arrow, K. J and Pines, D. eds. Addison-Wesley, New York.

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

Benaïm, M. (1993). A dynamical system approach to stochastic approximations. *Preprint.* University of California at Berkeley. *SIAM J. Control and optimization* (1996), in press.

Beneveniste, A., Métivier, M. and Priouret, P. (1990). *Stochastic Approximations and Adaptive Algorithms.* Springer; Berlin, Heidelberg, New York. Translated from *Algorithmes adaptatifs et approximations stochastiques.* Masson; Paris (1987).

Conley, C. C. (1978). *Isolated Invariant Sets and the Morse Index*, Providence: Published for the Conference Board of the Mathematical Sciences by the American Mathematical Society, Regional Conference series in Mathematics, No. 38.

Hall, D. and Heyde, C. C. (1980). *Martingale Limit Theory and Applications.* Academic Press, New York.

Hill, B. M., Lane, D. and Sudderth, W. (1980). A strong law for some generalized urn processes. *Ann. Probab.* **8** 214–226.

Hartman, P. (1964). *Ordinary Differential Equations.* John Wiley & Sons, New York.

Hirsch, M. W. (1976). *Differential Topology.* Springer-Verlag, New York, Heidelberg, Berlin.

Hirsch, M. W. and Pugh, C. (1970). Stable manifolds for hyperbolic sets. In *Proc. Symposia in Pure Mathematics* 14, Global Analysis, American Mathematical Society, Providence, Rhode Island, pp. 133–163.

Hirsch, M. W., Pugh, C. and Shub, M. (1977). *Invariant Manifolds. Springer Lecture Notes in Mathematics, 583.* Springer-Verlag, New York, Heidelberg, Berlin.

Kushner, H. J. and Clark, D. S. (1978). *Stochastic Approximation Methods for Constrained and Unconstrained Systems..* Springer; Berlin, Heidelberg, New York.

Métivier, M. and Priouret, P. (1987). Théorèmes de convergence presque sure pour une classe d'algorithmes stochastiques à pas décroissant. *Probability Theory and Related Fields* **74** 403–428.

Morse, M. (1960). The existence of polar nondenerate functions on differentiable manifolds. *Ann. Math.* **71** 352–383.

Palis, J. (1969). On Morse–Smale dynamical systems. *Topology* **8** 385–405.

Palis, J. and Smale, S. (1968). Structural stability theorems. in *Proc. Sympo. Pure Math. 14*. Global Analysis, Chern, S. -S. and Smale, S. eds. American Mathematical Society, Providence, RI.

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

Pemantle, R. (1992). Vertex-reinforced random walk. *Probability Theory and Related Fields* **92** 117–136.

Peixoto, M. M. (1962). Structural stability on two-dimensional manifolds. *Topology* **1** 101–120.

Pugh, C. and Shub, M. (1970). The omega stability theorem for flows. *Invent. Math.* **11** 150–158.

Smale, S. (1961a). Generalized Poincaré conjecture in dimensions greater than four. *Ann. Math.* **74** 391–406.

Smale, S. (1961b). On gradient dynamical systems. *Ann. Math.* **74** 199–206.

Wilson, W. (1969). Smoothing derivatives of functions and applications. *Trans. Amer. Math. Soc.* **139** 413–428.