UC Berkeley

UC Berkeley Previously Published Works

Title

Stochastic perturbations of dynamical systems

Permalink

<https://escholarship.org/uc/item/2ng4z38t>

Journal

Lecture in pure and applied mathematics, 176

Author

Hirsch, MW

Publication Date

1996

Peer reviewed

11/13

Ariel

Odyssey Enabled

Document Delivery Article

Duke University Libraries Perkins Library Document
Delivery

Location:

Call #: Perkins/Bostock Library |
Stacks | QA1 .L4386

Trans. #: 944234

Journal Title: Lecture notes in pure and
applied mathematics.

Borrower: GZM

Lending String:

MNU,COD,*NDD,UMC,MUU,NOC,VA@,IAY,SOI

Volume: 176

Issue:

Patron:

Date: 1996

DATE: Wednesday, November 12, 2014

Pages: 79-91

Maxcost: 55.00IFM

Article Title: Stochastic perturbations of
dynamical systems

Shipping Address:

Article Author: Hirsch, Morris W.

ILL Borrowing-University of WI Madison
Room B106F Memorial Library

Imprint: New York, M. Dekker. 1971

728 State Street

Madison Wisconsin 53706-1494 United Stat

ILL Number: 137043011

FAX: 608 262-4649

Ariel: ariलगzm.library.wisc.edu

Odyssey: 216.54.119.76

Perkins Library

Duke University

Document Delivery Services

Box 90183

Durham, NC 27708-0183

NDD

Phone: 919-660-5891

Fax: 919-660-5964

Ariel: ariel.lib.duke.edu

If there are problems with this Ariel/Fax document:

Missing Page(s): _____

Edge(s) Cut Off: _____

Unable/Difficult to Read: _____

For replacement pages, please contact within 48 hours.

NOTICE: THIS MATERIAL MAY
BE PROTECTED BY
COPYRIGHT LAW
(TITLE 17 U.S. CODE)

THANK YOU FOR USING OUR SERVICES.

Stochastic Perturbations of Dynamical Systems

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

INTRODUCTION

This is a report of recent joint work with Michel Benaïm of the University of Toulouse. Detailed proofs of the theorems on urn models will appear in Benaïm and Hirsch, (in press). The material on stochastic Newton's method is new; proofs will be given elsewhere.

Let $F : \mathbf{R}^m \rightarrow \mathbf{R}^m$ be a smooth map and $\Delta \subset \mathbf{R}^m$ 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

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

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

- The sequence $\{\gamma_i\}$ is decreasing and $\sum_i \gamma_i = +\infty$. Such a sequence will be called a *decreasing gain sequence*.
- 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) $E(U_{n+1} | \mathcal{F}_n) = 0$.
- There exists $K > 0$ such that $\|U_n\| \leq K$ for all $n \geq 0$.

Throughout this paper $\|\cdot\|$ denotes the Euclidean norm on $\mathbf{R}^m$ and $\langle \cdot, \cdot \rangle$ 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

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

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 work 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 asymptotic behavior of urn models associated to a Morse-Smale vector field.

The key to our results are recent papers by R. Pemantle (1990) and M. Benaïm (1993). Pemantle showed that under reasonable assumptions, the probability that sample paths converge to an unstable equilibrium of F is zero. Benaïm 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 . Recently we have proved that the restriction of Φ to $R(F)$ is chain recurrent.

We have extended 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 $\Delta^m \subset \mathbf{R}^{m+1}$ is the set

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

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

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

An urn initially (i. e. at time $n = 0$) contains $n_0 \geq 1$ balls of colors $1, \dots, m+1$. 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 $x_n = (x_{n,1}, \dots, x_{n,m+1})$. 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 $f : \Delta^m \rightarrow \Delta^m$.

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* $f: \Delta^m \rightarrow \Delta^m$. 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 one 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 added to the urn at time $n+1$, given the value of the proportion vector x_n , is a $\{0, 1\}$ -valued 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

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

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 identify the affine space $\{v \in \mathbf{R}^{m+1} : \sum_{j=1}^{m+1} v_j = 1\}$ with the linear subspace E^m by parallel translation, and also with $\mathbf{R}^m$ by any convenient affine isometry. Under the latter identification, process (3) takes exactly the form (1), where $F: \mathbf{R}^m \rightarrow \mathbf{R}^m$ denotes any map which equals $-\text{Id} + f$ on Δ^m , and $\gamma_{n+1} = \frac{1}{n_0 + n + 1}$.

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:

$$x_{n+1} = (1 - \gamma_{n+1})x_n + \gamma_{n+1}s_{n+1}. \quad (4)$$

EXAMPLE 0.2 (Stochastic Approximation) Let $\{\xi_i\}_{i \geq 1}$ be a sequence of independent identically distributed $\mathbf{R}^d$ -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 $h: \mathbf{R}^m \times \mathbf{R}^d \rightarrow \mathbf{R}^m$, 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, Métivier and Priouret (1990)).

1 CHAIN RECURRENCE AND NONCONVERGENCE TO UNSTABLE EQUILIBRIA

Throughout the paper $F : \mathbf{R}^m \rightarrow \mathbf{R}^m$ 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 doesn't 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

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

defined by

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

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

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

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 .¹ To describe the asymptotic behavior of (1) we consider the *limit set* of any sequence $\{y_n\}_{n \geq 0}$ in $\mathbf{R}^m$, denoted by $L(\{y_n\})$. It is defined, as usual, as the set of $p \in \mathbf{R}^m$ such that $\lim_{k \rightarrow \infty} y_{n_k} = p$ for some sequence $n_k \rightarrow \infty$.

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

THEOREM 1.1 (Benaïm, 1993) *Suppose given sequences $\{u_n\}$, $\{b_n\}$ in $\mathbf{R}^m$, 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) $\lim_{n \rightarrow \infty} b_n = 0$.
- (iii) For each $T > 0$,

$$\limsup_{n \rightarrow \infty} \left\{ \sum_{i=n+1}^k \gamma_i \|u_i\| : k \in \mathbf{N}, 0 \leq \tau_k - \tau_n \leq T \right\} = 0$$

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

¹ 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 $\theta \in \mathbf{R}/2\pi\mathbf{R}$, 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.

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

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

we recall that $(\{x_n : \Omega \rightarrow \mathbf{R}^m\}_{n \geq 0})$ 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).

The following probabilistic consequence can be derived from Theorem 1.1:

COROLLARY 1.2 Assume there exists $\delta \geq 1$ such that the decreasing gain sequence $\{\gamma_i\}_{i \geq 1}$ 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) $L(\{x_n\}) \subset R(F) \cap \Delta$.

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 $\mathbf{R}^k \rightarrow \mathbf{R}^m$ where $0 \leq k < m$. Consequently $W_s(p)$ has measure 0 in $\mathbf{R}^m$. 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) $\frac{A}{n^\mu} \leq \gamma_n \leq \frac{B}{n^\mu}$ where $0 < A \leq B$, $\frac{1}{2} < \mu \leq 1$.
- (iii) There exists $b > 0$ such that for all unit vector $\Theta \in \mathbb{R}^m$:

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

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

It can be shown that Pemantle's theorem remains true for $0 < \mu \leq 1$.

In Theorem 2.1, 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 $C(\Gamma)$ doesn't 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 neighborhood B such that

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

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 obtain 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) $\frac{A}{n^\mu} \leq \gamma_n \leq \frac{B}{n^\mu}$ where $0 < A \leq B$, $0 < \mu \leq 1$.
- (iii) There exists $b > 0$ such that for all unit vector $\Theta \in \mathbb{R}^m$:

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

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 ($r \geq 1$) 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

$$L(F) = Per(F) = \Gamma_0 \cup \dots \cup \Gamma_l$$

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 $R(F) = L(F) = Per(F)$.²

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

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

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

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

²For example, this can be deduced from a filtration for $L(F)$, the existence of which follows from Section 7 of (Pugh and Shub, 1970).

- (b) If conditions (i) through (iii) of Theorem 2.1 are satisfied, then $p(\Gamma_i) > 0 \Rightarrow \Gamma_i$ 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

$$\lim_{F' \rightarrow F} p(\Gamma'_i) = p(\Gamma_i).$$

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:

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 $\Theta \in \mathbf{R}^m$:

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

Then for process (5), $\lim_{n \rightarrow \infty} x_n = p$ 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. $\square$.

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 $\mathbf{R}^m$.

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 $f : \Delta^m \rightarrow \Delta^m$ is C^1 .*
- (ii) *$f(\Delta^m) \subset \text{Int}(\Delta^m)$.*
- (iii) *The vector field $F = -\text{Id} + f$ has an attractor $\Lambda \subset \text{Int}(\Delta^m)$.*

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

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

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 $f : \Delta^m \rightarrow \Delta^m$ is C^2 .*
- (ii) *$f(\Delta^m) \subset \text{Int}(\Delta^m)$.*
- (iii) *There exists a Morse-Smale vector field F on $\mathbb{R}^m$ such that $F|_{\Delta^m} = -\text{Id} + f$*

Then:

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

$$(b) p(\Gamma_i) > 0 \Leftrightarrow \Gamma_i \text{ is linearly stable.}$$

If f is not required to satisfy (ii), one can construct urn functions so that the vector field $-\text{Id} + f$ 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 $f(\Delta^m) \subset \text{Int}(\Delta^m)$ is not very restrictive, as the following result shows.

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

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 $g(\Delta^m) \subset \text{Int}(\Delta^m)$. $\square$

REMARK 2.9 Most of our results on urn processes remain valid if the assumption $f(\Delta^m) \subset \text{Int}(\Delta^m)$ is replaced by the weaker hypothesis:

- (i) If $z \in \text{Int}(\Delta^m)$ then $f(z) \in \text{Int}(\Delta^m)$;
- (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 $\chi^r(\Delta^2)$.³

Define the map $T : \mathcal{A} \rightarrow \chi^r(\Delta^2)$ by $T(f) = -Id + f$. 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.*

3 STOCHASTIC NEWTON'S METHOD

Consider the problem of numerically approximating a root of an equation $f(z) = 0$, where $f : \mathbf{R}^n \rightarrow \mathbf{R}^n$ is a given smooth map. A standard approach is the *Deterministic Newton's Method with step size $\gamma > 0$* :

$$x_{n+1} = x_n - \gamma Df(x_n)^{-1} f(x_n) \quad (6)$$

starting from an initial guess x_0 . This deterministic process has the following geometric interpretation. Define the *Newton vector field*:

$$N(x) = -DF(x)^{-1} f(x),$$

defined where the Jacobian determinant $J(X) = \text{Det}(Df(x))$ is nonzero. It is clear that the map f carries trajectories of N into trajectories of the vector field $y \mapsto -y$, because a computation shows

$$\frac{d}{dt} f(x(t)) = -f(x(t)).$$

Therefore if $x(t)$ is a trajectory of N defined for $0 \leq t \leq T$, then

$$f(x(t)) = e^{-t} f(x(0)).$$

³Peixoto'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.

Process (6) is just a Cauchy-Euler approximation to a solution curve of $dx/dt = N(X)$ with initial value $x(0) = x_0$.

A basic problem arises because the Newton vector field $N(x)$ is not defined at *critical points* x of f , that is, points where the Jacobian determinant $J(x) = \text{Det} Df(x)$ is 0. There are various refinements of Newton's method designed to solve this difficulty; see Hirsch and Smale (1979), Shub and Smale (1986). Either the step size can be made variable, or the Newton field can be redefined near critical points, or both.

Another approach is to make small random perturbations of trajectories of the Newton field. Consider the simple case of a single complex polynomial map

$$p: \mathbb{C} \rightarrow \mathbb{C}, p(z) = z^d + a_{d-1}z^{d-1} + \cdots + a_0$$

with complex coefficients a_k . In this case the Newton field can be written in the notation of complex functions as $N(z) = -f(z)/f'(z)$. We regularize the Newton field by multiplying it by the nonnegative scalar function $|f'(z)|^2$. The resulting field is smooth everywhere, and has the same orbits (reparameterized) outside the set of critical points (roots of $f'(z)$). The *regularized Newton field* is thus

$$M(Z) = |p'(z)|^2 N(z) = -p(z)\overline{p'(z)}.$$

A Cauchy-Euler approximation to solving

$$\frac{dz}{dt} = M(z)$$

yields the deterministic process

$$z_{n+1} = z_n + \frac{1}{n+1} [-p(z_n)\overline{p'(z_n)}].$$

Critical points still make trouble, however, as it can happen that the sequence $\{z_n\}$ converges to a critical point. To avoid this we add noise, obtaining a class of *stochastic Newton's methods*:

$$z_{n+1} = z_n + \frac{1}{n+1} [-p(z_n)\overline{p'(z_n)} + s_n W_{n+1}] \quad (7)$$

where $\{W_n\}$ is a sequence of independent identically distributed mean zero random variables, each taking only the values $\{1, -1, i, -i\}$, with equal probability. We also use random variables $s_n \in \{0, 1\}$ as switches, turning on the noise W_{n+1} if $|f'(z_n)|$ is small. Precisely:

$$s_n = \begin{cases} 1 & \text{if } |f'(z_n)| \leq \frac{1}{n+1}, \\ 0 & \text{otherwise.} \end{cases}$$

Process (7) can be written as:

$$z_{n+1} = z_n + \gamma_{n+1}(F(z_n) + U_{n+1})$$

with

$$\begin{aligned} \gamma_{n+1} &= 1/(n+1), \\ F(z) &= -p(z)\overline{p'(z)}, \\ U_{n+1} &= s_n W_{n+1}. \end{aligned}$$

This is a Markov process, as $E(W_{n+1}|z_n) = 0$.

When z_n is far (relative to n) from the critical set, the noise is switched off ($s_n = 0$), and then the process follows Newton trajectories with decreasing step size $1/(n+1)$. When z_n is close to a critical point the process adds a small random jump U_{n+1} to the Newton step.

It can be shown that there is a number $K > 0$ such that the disk of radius K attracts all trajectories of the modified Newton field M , and such that almost surely $|z_n| < K$. Therefore Benaïm's Theorem 1.2 is applicable, implying that almost surely the limit set of $\{z_n\}$ is a connected subset of the chain recurrent set of M . But as $|p|$ is a Liapunov function for M , the chain recurrent set of M consists of the zeroes of M , namely finite set comprising the roots and critical points of p . Therefore almost surely $\{z_n\}$ converges to a root or a critical point of p .

Of course we want $\{z_n\}$ to converge to a root, not a critical point. For this we wish to apply Pemantle's theorem 1.3. We need to know that every critical point w of p which is not a root, is a linearly unstable equilibrium of M . (One can show that every root of p , critical or not, is an asymptotically stable equilibrium of M .) A calculation shows that a critical point w is linearly unstable provided w is a simple root of p' . Hypothesis (iii) on the noise in Pemantle's theorem is guaranteed by our choice of W_{n+1} , as is hypothesis (b) about the sigma fields of process (1).

From the foregoing considerations we obtain the proof of the following result:

THEOREM 3.1 *Let $p(z)$ be complex polynomial of degree $d \geq 1$. Assume that every multiple root of $p'(z)$ is also a root of $p(z)$. Then for every $z_0 \in \mathbb{C}$, process (7) almost surely converges to a root of $p(z)$.*

BIBLIOGRAPHY

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.

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

Benaïm, M. (1993): "A dynamical system approach to stochastic approximations", preprint, University of California at Berkeley.

Benaïm, M. and Hirsch, M. (in press): Dynamics of Morse-Smale Urn Processes, *Ergodic Theory And Dynamical Systems*.

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

- Conley, C.C. (1978): *Isolated invariant sets and the Morse index*. Published for the Conference Board of the Mathematical Sciences by the American Mathematical Society, Regional conference series in mathematics, no. 38.
- Hill, B.M., Lane, D. and Sudderth, W. (1980): "A strong law for some generalized urn processes", *Annals of Probability* 8, pp 214-226.
- Hirsch, M.W., Pugh, C. (1970): "Stable manifolds for hyperbolic sets", in *Proceedings of Symposia in Pure Mathematics* 14, Global Analysis, American Mathematical Society, Providence, Rhode Island, pp 133-163.
- Hirsch, M.W. and Smale, S. (1979): "On algorithms for solving $f(x) = 0$ ", *Communications in Pure and Applied Mathematics* 55, pp 1-12.
- Kushner, H.J., Clark, D.S. (1978): *Stochastic Approximation Methods for Constrained and Unconstrained Systems*. Springer; Berlin, Heidelberg, New York.
- Métivier, M., 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.
- Palis, J. (1969): "On Morse-Smale dynamical systems", *Topology* 8, pp 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", *Annals of Probability* 18, pp 698-712.
- Peixoto, M.M. (1962): "Structural stability on two-dimensional manifolds", *Topology* 1, pp 101-120.
- Shub, M. and Smale, S. (1986): "On the existence of generally convergent algorithms", *Journal of Complexity* 2, pp 2-11
- Pugh, C. and Shub, M. (1970): "The omega stability theorem for flows", *Inventiones Math.* 11, pp 150-158.