Séminaire de probabilités (Strasbourg)

Michel Benaïm

Dynamics of stochastic approximation algorithms

Séminaire de probabilités (Strasbourg), tome 33 (1999), p. 1-68

<http://www.numdam.org/item?id=SPS_1999__33__1_0>

Dynamics of Stochastic Approximation Algorithms

Michel Benaïm

Abstract

These notes were written for a D.E.A course given at *Ecole Normale Supérieure* de Cachan during the 1996-97 and 1997-98 academic years and at *University Toulouse III* during the 1997-98 academic year. Their aim is to introduce the reader to the dynamical system aspects of the theory of stochastic approximations.

Contents

1 Introduction

Stochastic approximation algorithms are discrete time stochastic processes whose general form can be written as

$$x_{n+1} - x_n = \gamma_{n+1} V_{n+1} \tag{1}$$

where x_n takes its values in some euclidean space, V_{n+1} is a random variable and $\gamma_n > 0$ is a "small" step-size.

Typically x_n represents the parameter of a system which is adapted over time and $V_{n+1} = f(x_n, \xi_{n+1})$. At each time step the system receives a new information ξ_{n+1} that causes x_n to be updated according to a *rule* or *algorithm* characterized by the function f. Depending on the context f can be a function designed by a user so that some goal (estimation, identification, ...) is achieved, or a model of adaptive behavior.

The theory of *stochastic approximations* was born in the early 50s through the works of Robbins and Monro (1951) and Kiefer and Wolfowitz (1952) and has been extensively used in problems of signal processing, adaptive control (Ljung, 1986; Ljung and Sőderstrőm, 1983; Kushner and Yin, 1997) and recursive estimation (Nevelson and Khaminski, 1974). With the renewed and increased interest in the *learning paradigm* for artificial and natural systems, the theory has found new challenging applications in a variety of domains such as neural networks (White, 1992; Fort and Pages, 1994) or game theory (Fudenberg and Levine, 1998).

To analyse the long term behavior of (1), it is often convenient to rewrite the noise term as

$$V_{n+1} = F(x_n) + U_{n+1} \tag{2}$$

where $F : \mathbb{R}^m \to \mathbb{R}^m$ is a deterministic vector field obtained by suitable averaging. The examples given in Section 2 will illustrate this procedure. A natural approach to the asymptotic behavior of the sequences $\{x_n\}$ is then, to consider them as approximations to solutions of the ordinary differential equation (ODE)

$$\frac{dx}{dt} = F(x). \tag{3}$$

One can think of (1) as a kind of Cauchy-Euler approximation scheme for numerically solving (3) with step size γ_n. It is natural to expect that, owing to the fact that γ_n is small, the noise washes out and that the asymptotic behavior of $\{x_n\}$ is closely related to the asymptotic behavior of the ODE. This method called the *ODE method* was introduced by Ljung (1977) and extensively studied thereafter. It has inspired a number of important works, such as the book by Kushner and Clark (1978), numerous articles by Kushner and coworkers, and more recently the books by Benveniste, Métivier and Priouret (1990), Duflo (1996) and Kushner and Yin (1997).

However, until recently, most works in this direction have assumed the simplest dynamics for F (for example that F is the negative of the gradient of a cost function), and little attention has been payed to dynamical system issues.

The aim of this set of notes is to show how dynamical system ideas can be fully integrated with probabilistic techniques to provide a rigorous foundation to the ODE method beyond gradients or other dynamically simple systems. However it is not intended to be a comprehensive presentation of the theory of stochastic approximations. It is principally focused on the *almost sure dynamics of stochastic approximation processes with decreasing step sizes.* Questions of weak convergence, large deviation, or rate of convergence, are not considered here. The assumptions on the "noise" process are chosen for simplicity and clarity of the presentation.

These notes are partially based on a DEA course given at Ecole Normale Supérieure de Cachan during the 1996-1997 and 1997-1998 academic years and at University Paul Sabatier during the 1997-1998 academic year. I would like to especially thank Robert Azencott for asking me to teach this course and Michel Ledoux for inviting me to write these notes for *Le Séminaire de Probabilités.*

An important part of the material presented here results from a collaboration with Morris W. Hirsch and it is a pleasure to acknowledge the fundamental influence of Moe on this work. I have also greatly benefited from numerous discussions with Marie Duflo over the last months which have notably influenced the presentation of these notes. Finally I would like to thank Odile Brandière, Philippe Carmona, Laurent Miclo, Gilles Pages and Sebastian Schreiber for valuable insights and informations.

Although most of the material presented here has already been published, some results appear here for the first time and several points have been improved.

1.1 Outline of Contents

These notes are organized as follows.

Section 2 presents simple motivating examples of stochastic approximation processes.

Section 3 introduces the notion of asymptotic pseudotrajectories for a semiflow. This a purely deterministic notion due to Benaïm and Hirsch (1996) which arises in many dynamical settings but turns out to be very well suited to stochastic approximations.

In section 4 classical results on stochastic approximations are (re)formulated in the language of asymptotic pseudotrajectories. Attention is restricted to the classical situation where U_{n+1} (see (2)) is a sequence of martingale differences. It is shown that, under suitable conditions, the continuous time process obtained by a convenient interpolation of $\{x_n\}$ is almost surely an asymptotic pseudotrajectory of the semiflow induced by the associated ODE. This section owes much to the ideas and techniques developped by Kushner and his co-workers (Kushner and Clark, 1978; Kushner and Yin, 1997), Métivier and Priouret (1987) (see also Benveniste, Métivier and Priouret (1990)) and Duflo (1990, 1996, 1997). The case of certain diffusions and jump processes is also considered.

Section 5 characterizes the limit sets of asymptotic pseudotrajectories. It begins with a comprehensive introduction to chain-recurrence and chain-transitivity. Several properties of chain transitive sets are formulated. The main result of the section establishes that limit sets of precompact asymptotic pseudotrajectories are internally chain-transitive. This theorem was originally proved in (Benaïm, 1996) but I have chosen to present here the proof of Benaïm and Hirsch (1996). I find this proof conceptually attractive and it is somehow more directly related to the original ideas of Kushner and Clark (1978).

Section 6 applies the abstract results of section 5 in various situations. It is shown how assumptions on the deterministic dynamics can help to identify the possible limit sets of stochastic approximation processes with a great deal of generality. This section generalizes and unifies many of the results which appear in the literature on stochastic approximation.

Section 7 establishes simple sufficient conditions ensuring that a given attractor of the ODE has a positive probability to host the limit set of the stochastic approximation process. It also provides lower bound estimates of this probability. This section is based on unpublished works by Duflo (1997) and myself.

Section 8 considers the question of *shadowing.* The main result of the section asserts that when the step size of the algorithm goes to zero at a suitable rate (depending on the expansion rate of the ODE) trajectories of (1) are almost surely asymptotic to forward trajectories of (3). This section represents a synthesis of the works of Hirsch (1994), Benaïm (1996), Benaïm and Hirsch (1996) and Duflo (1996) on the question of shadowing. Several properties and estimates of the expansion rate due to Hirsch (1994) and Schreiber (1997) are presented. In particular, Schreiber's ergodic characterization of the expansion rate is proved.

Section 9 pursues the qualitative analysis of section 7. The focus is on the behavior of stochastic approximation processes near "unstable" sets. The centerpiece of this section is a theorem which shows that stochastic approximation processes have zero probability to converge toward certain repelling sets including linearly unstable equilibria and periodic orbits as well as normally hyperbolic manifolds. For unstable equilibria this problem has often been considered in the literature but, to my knowledge, only the works by Pemantle (1990) and Brandière and Duflo (1996) are fully satisfactory. I have chosen here to follow Pemantle's arguments. The geometric part contains new ideas which allow to cover the general case of normally hyperbolic manifolds but the probability part owes much to Pemantle.

Section 10 introduces the notion of a stochastic process being a *weak asymptotic pseudotrajectory* for a semiflow and analyzes properties of its empirical occupation measures. This is motivated by the fact that any stochastic approximation process with decreasing step size is a weak asymptotic pseudotrajectory of the associated ODE regardless of the rate at which $\gamma_n \to 0$.

2 Some Examples

2.1 Stochastic Gradients and Learning Processes

Let $\{\xi_i\}_{i\geq 1}$, $\xi_i \in E$ be a sequence of independent identically distributed random inputs to a system and let $x_n \in \mathbb{R}^m$ denote a parameter to be updated, $n \geq 0$. We suppose the updating to be defined by a given map $f : \mathbb{R}^m \times E \to \mathbb{R}^m$, and the following stochastic algorithm:

$$x_{n+1} - x_n = \gamma_{n+1} f(x_n, \xi_{n+1}). \tag{4}$$

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

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

and set

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

It is clear that this algorithm has the form given by [(1) (2)]. Such processes are classical models of adaptive algorithms.

A situation often encountered in "machine learning" or "neural networks" is the following: Let I and O be euclidean spaces and $M : \mathbb{R}^m \times I \to O$ a smooth function representing a system (e.g a neural network). Given an *input* $y \in I$ and a parameter x the system produces the *output* $M(x, y)$.

Let $\{\xi_n\} = \{(y_n, o_n)\}$ be a sequence of i.i.d random variables representing the *training set* of M. Usually the law μ of ξ_n is unknown but many samples of ξ_n are available. The goal of learning is to adapt the parameter x so that the output $M(x, y_n)$ gives a good approximation of the *desired output* o_n. Let $e : O \times O \to \mathbb{R}_+$ be a smooth error function. For example $e(o, o') = ||o - o'||^2$. Then a basic training procedure for M is given by (4) where

$$f(x,\xi) = -\frac{\partial}{\partial x} e(M(x,y),o), \text{ and } \xi = (y,o).$$

Assuming that derivation and expectation commute, the associated ODE is the gradient ODE given by

$$F(x) = -\nabla C(x) \text{ with } C(x) = \int e(M(x,y),o)d\mu(y,o).$$

2.2 Polya's Urns and Reinforced Random Walks

The *unit m-simplex* $\Delta^m \subset \mathbb{R}^{m+1}$ is the set

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

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

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

An urn initially (i. e. ,at time $n = 0$) contains $n_0 > 0$ 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 $f : \Delta^m \to \Delta^m$.

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) used this kind of model to describe competing technologies in economics.

An urn model is determined by the initial *urn composition* (x_0, n_0) and the *urn function* $f : \Delta^m \to \Delta^m$. We assume that the initial composition (x_0, n_0) is fixed one for all. The σ-field $\mathcal{F}_n$ is the field generated by the random variables $x_0, \ldots, x_n$. One easily verifies that the equation

$$x_{n+1} - x_n = \frac{1}{n_0 + n + 1}(-x_n + f(x_n) + U_{n+1}) \tag{5}$$

defines random variables $\{U_n\}$ that satisfy $\mathsf{E}(U_{n+1}|\mathcal{F}_n) = 0$. We can identify the affine space $\{v \in \mathbb{R}^{m+1} : \sum_{j=1}^{m+1} v_j = 1\}$ with E^m by parallel translation, and also with $\mathbb{R}^m$ by any convenient affine isometry. Under the latter identification, we see that process (5) has exactly the form of [(1) (2)], taking F to be any map which equals $-\mathrm{Id} + f$ on Δ^m, and setting $\gamma_n = \dfrac{1}{n_0 + n}$.

Observe that f being arbitrary, the dynamics of $F = -Id + f$ can be arbitrarily complicated.

The next example is a generalization of urn processes which I call *Generalized Vertex Reinforced Random Walks* after Diaconis and Pemantle. These are non-Markovian discrete time stochastic processes living on a finite state space for which the transition probabilities at each step are influenced by the proportion of time each state has been visited.

Let $\mathcal{M}_{m+1}(\mathbb{R})$ denote the space of real $(m+1) \times (m+1)$ matrices and let $M : E^m \to \mathcal{M}_{m+1}(\mathbb{R})$ be a smooth map such that for all $v \in \Delta^m$, $M(v) = \{M_{i,j}(v)\}$ is Markov transition matrix. Given a point $x_0 \in Int(\Delta^m)$, a vertex $y \in \{1, \ldots, m+1\}$ and a positive integer $n_0 \in \mathbb{N}$, consider a stochastic process $\{(Y_n, (S_1(n), \ldots, S_{m+1}(n))\}_{n \geq 0}$ defined on $\{1, \ldots, m+1\} \times \mathbb{R}_+^{m+1}$ by

- $S_i(0) = n_0 x_{0,i}$, $Y_0 = y$.
- $S_i(n) = S_i(0) + \sum_{k=1}^{n} \delta_{Y_k, i}$, $n \geq 0$.
- $P(Y_{n+1} = j|\mathcal{F}_n) = M_{Y_n, j}(x_n)$

where $\mathcal{F}_n$ denotes the σ-field generated by $\{Y_j : 0 \le j \le n\}$ and $x_n = \frac{S(n)}{n+n_0}$ is the empirical occupation measure of $\{Y_n\}$.

Suppose that for each $v \in \Delta^m$ the Markov chain $M(v)$ is indecomposable (i.e has a unique recurrence class), then by a standard result of Markov chains theory, $M(v)$ has a unique invariant probability measure $f(v) \in \Delta^m$. As for Polya's urns, equation (5) defines a sequence of random variables $\{U_n\}$. Here the $\{U_n\}$ are no longer martingale differences but the ODE governing the long term behavior of $\{x_n\}$ is still given by the vector field $F(x) = -x + f(x)$ (see Benaïm, 1997).

The original idea of these processes is due to Diaconis who introduced the process defined by

$$M_{i,j}(v) = \frac{R_{i,j}v_j}{\sum_k R_{i,k}v_k}$$

with $R_{i,j} > 0$. For this process called a *Vertex Reinforced Random Walk* the probability of transition to site j increases each time j is visited. The long term behavior of $\{x_n\}$ has been analyzed by Pemantle (1992) for $R_{ij} = R_{ji}$ and by Benaïm (1997) in the non-symmetric case. With a non-symmetric R the ODE may have nonconvergent dynamics and the behavior of the process becomes highly complicated (Benaïm, 1997).

2.3 Stochastic Fictitious Play in Game Theory

Our last example is an adaptive learning process introduced by Fudenberg and Kreps (1993) for repeated games of incomplete information called *stochastic fictitious play.* It belongs to a flourishing literature which develops the explanation that equilibria in games may arise as the result of learning rather than from rationalistic analysis. For more details and economics motivation we refer the reader to the recent book by Fudenberg and Levine (1998).

For notational convenience we restrict attention to a two-players and two-strategies game. The players are labeled $i = 1, 2$ and the set of strategies is denoted $\{0, 1\}$.

Let $\{\xi_n\}_{n\ge1}$ be a sequence of identically distributed random variables describing the *states of nature.* The *payoff* to player i at time n is a function $U^i(.,\xi_n) : \{0,1\}^2 \to \mathbb{R}$. We extend $U^i(.,\xi_n)$ to a function $U^i(.,\xi_n) : [0,1]^2 \to \mathbb{R}$ defined by $U^i(x^1, x^2, \xi_n) =$

$$x^1[x^2U^i(1,1,\xi_n)+(1-x^2)U^i(1,0,\xi_n)]+(1-x^1)[x^2U^i(0,1,\xi_n)+(1-x^2)U^i(0,0,\xi_n)].$$

Consider now the repeated play of the game. At round n player i chooses an action $s_n^i \in \{0,1\}$ independently of the other player. As a result of these choices player i receives the payoff $U^i(s_n^1, s_n^2, \xi_n)$. The basic assumption is that $U^i(.,\xi_n)$ is known to player i at time n but the strategy chosen by her opponent is not. At the end of the round, both players observe the strategies played.

Fictitious play produces the following adaptive process: At time $n+1$ player 1 (respectively 2) knowing her own payoff function $U^1(.,\xi_{n+1})$ and the strategies played by her opponent up to time n computes and plays the action which maximizes her expected payoff under the assumption that her opponent will play an action whose probability distribution is given by historical frequency of past plays. That is

$$s_{n+1}^1 = Argmax_{s\in\{0,1\}} U^1(s, x_n^2, \xi_{n+1})$$

where

$$x_n^i = \frac{1}{n}\sum_{k=1}^{n} s_k^i.$$

A simple computation shows that the vector of empirical frequencies $x_n = (x_n^1, x_n^2)$ satisfies a recursion of type [(1),(2)] with $\gamma_n = \frac{1}{n}$, $E(U_{n+1}|\mathcal{F}_n) = 0$ and F is the vector field given by

$$F(x^1, x^2) = (-x^1 + h^1(x^2), -x^2 + h^2(x^1)) \tag{6}$$

where

$$h^1(x^2) = P(U^1(1, x^2, \xi) > U^1(0, x^2, \xi)),$$
$$h^2(x^1) = P(U^2(x^1, 1, \xi) > U^2(x^1, 0, \xi)).$$

The mathematical analysis of stochastic fictitious play has been recently conducted by (Benaïm and Hirsch, 1994) and (Kaniovski and Young, 1995). We will give in section 6.4 (see Example 6.16) a simple argument ensuring the convergence of the process.

3 Asymptotic Pseudotrajectories

A *semiflow* Φ on a metric space (M, d) is a continuous map

$$\Phi : \mathbb{R}_+ \times M \to M,$$
$$(t, x) \mapsto \Phi(t, x) = \Phi_t(x)$$

such that

$$\Phi_0 = \text{Identity},\ \Phi_{t+s} = \Phi_t \circ \Phi_s$$

for all $(t, s) \in \mathbb{R}_+ \times \mathbb{R}_+$. Replacing $\mathbb{R}_+$ by $\mathbb{R}$ defines a *flow*.

A continuous function $X : \mathbb{R}_+ \to M$ is an *asymptotic pseudotrajectory* for Φ if

$$\lim_{t\to\infty} \sup_{0\le h\le T} d(X(t+h), \Phi_h(X(t))) = 0$$

for any $T > 0$. Thus for each fixed $T > 0$, the curve

$$[0, T] \to M : h \mapsto X(t+h)$$

shadows the Φ-trajectory of the point $X(t)$ over the interval $[0, T]$ with arbitrary accuracy for sufficiently large t. By abuse of language we call X *precompact* if its image has compact closure in M.

The notion of asymptotic pseudotrajectories has been introduced in Benaïm and Hirsch (1996) and is particularly useful for analyzing the long term behavior of stochastic approximation processes.

3.1 Characterization of Asymptotic Pseudotrajectories

Let $C^0(\mathbb{R}, M)$ denote the space of continuous M-valued functions $\mathbb{R} \to M$ endowed with the topology of uniform convergence on compact intervals. If $X : \mathbb{R}_+ \to M$ is a continuous function, we consider X as an element of $C^0(\mathbb{R}, M)$ by setting $X(t) = X(0)$ for $t < 0$. The space $C^0(\mathbb{R}, M)$ is metrizable. Indeed, a distance is given by: for all $f, g \in C^0(\mathbb{R}, M)$,

$$d(f,g) = \sum_{k \in \mathbb{N}} \frac{1}{2^k} \min(1, d_k(f,g))$$

where $d_k(f,g) = \sup_{t \in [-k,k]} d(f(t), g(t))$.

The *translation flow* $\Theta : C^0(\mathbb{R}, M) \times \mathbb{R} \to C^0(\mathbb{R}, M)$ is the flow defined by:

$$\Theta^t(X)(s) = X(t+s).$$

Let Φ be a flow or a semiflow on M. For each $p \in M$, the trajectory $\Phi^p : t \to \Phi_t(p)$ is an element of $C^0(\mathbb{R}, M)$ (with the convention that $\Phi^p(t) = p$ if $t < 0$ and Φ is for a semiflow). The set of all such Φ^p defines a subspace $\mathbf{S}_\Phi \subset C^0(\mathbb{R}, M)$.

It is easy to see that the map $H : M \to \mathbf{S}_\Phi$ defined by $H(p) = \Phi^p$ is an homeomorphism which conjugates $\Theta|S_\Phi$ (Θ restricted to S_Φ) and Φ. That is

$$\Theta^t \circ H = H \circ \Phi_t$$

where $t \in \mathbb{R}$ if Φ is a flow and $t \geq 0$ if Φ is a semiflow. This makes $\mathbf{S}_\Phi$ a closed set invariant under Θ. Define the retraction $\hat{\Phi} : C^0(\mathbb{R}, M) \to \mathbf{S}_\Phi$ as

$$\hat{\Phi}(X) = H(X(0)) = \Phi^{X(0)}$$

Lemma 3.1 *A continuous function $X : \mathbb{R}_+ \to M$ is an asymptotic pseudotrajectory of Φ if and only if:*

$$\lim_{t \to \infty} d(\Theta^t(X), \hat{\Phi} \circ \Theta^t(X)) = 0.$$

Proof Follows from definitions. **QED**

Roughly speaking, this means that an asymptotic pseudotrajectory of Φ is a point of $C^0(\mathbb{R}_+, M)$ whose forward trajectory under Θ is attracted by $\mathbf{S}_\Phi$. We also have the following result:

Theorem 3.2 *Let $X : \mathbb{R}_+ \to M$ be continuous function whose image has compact closure in M. Consider the following assertions*

(i) *X is an asymptotic pseudotrajectory of Φ*

(ii) *X is uniformly continuous and every limit point[1] of $\{\Theta^t(X)\}$ is in $\mathbf{S}_\Phi$ (i.e a fixed point of $\hat{\Phi}$).*

[1]By a limit point of $\{\Theta^t(X)\}$ we mean the limit in $C^0(\mathbb{R}, M)$ of a convergent sequence $\Theta^{t_k}(X), t_k \to \infty$.

(iii) *The sequence* $\{\Theta^t(X)\}_{t\geq 0}$ *is relatively compact in* $C^0(\mathbb{R}, M)$.

Then (i) and (ii) are equivalent and imply (iii).

Proof Suppose that assertion (*i*) holds. Let K denote the closure of $\{X(t) : t \geq 0\}$. Let $\epsilon > 0$. By continuity of the flow and compactness of K there exists $a > 0$ such that $d(\Phi_s(x), x) < \epsilon/2$ for all $|s| \leq a$ uniformly in $x \in K$. Therefore $d(\Phi_s(X(t)), X(t)) < \epsilon/2$ for all $t \geq 0$, $|s| \leq a$.

Since X is an asymptotic pseudotrajectory of Φ, there exists $t_0 > 0$ such that $d(\Phi_s(X(t)), X(t+s)) < \epsilon/2$ for all $t > t_0$, $|s| \leq a$. It follows that $d(X(t+s), X(t)) < \epsilon$ for all $t > t_0$, $|s| \leq a$. This proves uniform continuity of X. On the other hand, Lemma 3.1 shows that any limit point of $\{\Theta^t(X)\}$ is a fixed point of $\hat{\Phi}$. This proves that (*i*) implies (*ii*).

Suppose now that (*ii*) holds. Since $\{X(t) : t \geq 0\}$ is relatively compact and X is uniformly continuous, $\{\Theta^t(X)\}_{t\geq 0}$ is equicontinuous and for each $s \geq 0$ $\{\Theta^t(X)(s)\}_{t\geq 0}$ is relatively compact in M. Hence by the Ascoli Theorem (see e.g Munkres 1975, Theorem 6.1), $\{\Theta^t(X)\}$ is relatively compact in $C^0(\mathbb{R}, M)$.

Therefore $\lim_{t\to\infty} d(\Theta^t(X), \hat{\Phi}(\Theta^t(X)) = 0$ which by Lemma 3.1 implies (*i*). The above discussion also shows that (*ii*) implies (*iii*). **QED**

Remark 3.3 Let $D(\mathbb{R}, M)$ be the space of functions which are right continuous and have left-hand limits (*càd làg* functions). The definition of asymptotic pseudotrajectories can be extended to elements of $D(\mathbb{R}, M)$. Since the convergence of a sequence $\{f_n\} \in D$ toward a continuous function f is equivalent to the uniform convergence of $\{f_n\}$ toward f on compact intervals, Lemma 3.1 continues to hold and Theorem 3.2 remains valid provided that we replace the statement that X is uniformly continuous by the weaker statement:

$\forall \epsilon > 0$ there exists $a > 0$ such that

$$\limsup_{t\to\infty} \sup_{|s|\leq a} d(X(t+s), X(t)) \leq \epsilon.$$

4 Asymptotic Pseudotrajectories and Stochastic Approximation Processes

4.1 Notation and Preliminary Result

Let $F : \mathbb{R}^m \to \mathbb{R}^m$ be a continuous map. Consider here a discrete time process $\{x_n\}_{n\in\mathbb{N}}$ living in $\mathbb{R}^m$ (an algorithm) whose general form can be written as

$$x_{n+1} - x_n = \gamma_{n+1}(F(x_n) + U_{n+1}) \tag{7}$$

where

- $\{\gamma_n\}_{n\geq 1}$ is a given sequence of nonnegative numbers such that

$$\sum_k \gamma_k = \infty, \ \lim_{n\to\infty} \gamma_n = 0.$$

- $U_n \in \mathbb{R}^m$ are (deterministic or random) perturbations.

Formula (7) can be considered to be a perturbed version of a variable step-size Cauchy-Euler approximation scheme for numerically solving $dx/dt = F(x)$:

$$y_{k+1} - y_k = \gamma_{k+1} F(y_k).$$

It is thus natural to compare the behavior of a sample path $\{x_k\}$ with trajectories of the flow induced by the vector field F. To this end we set

$$\tau_0 = 0 \text{ and } \tau_n = \sum_{i=1}^{n} \gamma_i \text{ for } n \geq 1,$$

and define the continuous time *affine* and *piecewise constant interpolated processes* $X, \overline{X} : \mathbb{R}_+ \to \mathbb{R}^m$ by

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

for all $n \in \mathbb{N}$ and $0 \leq s < \gamma_{n+1}$. The "inverse" of $n \to \tau_n$ is the map $m : \mathbb{R}_+ \to \mathbb{N}$ defined by

$$m(t) = \sup\{k \geq 0 \, : t \geq \tau_k\} \tag{8}$$

let $\overline{U}, \overline{\gamma} : \mathbb{R}_+ \to \mathbb{R}^m$ denote the continuous time processes defined by

$$\overline{U}(\tau_n + s) = U_{n+1}, \overline{\gamma}(\tau_n + s) = \gamma_{n+1}$$

for all $n \in \mathbb{N}$, $0 \leq s < \gamma_{n+1}$. Using this notation (7) can be rewritten as

$$X(t) - X(0) = \int_0^t [F(\overline{X}(s)) + \overline{U}(s)]ds \tag{9}$$

The vector field F is said to be *globally integrable* if it has unique integral curves. For instance a bounded locally Lipschitz vector field is always globally integrable. We then have

Proposition 4.1 *Let F be a continuous globally integrable vector field. Assume that*

A1 *For all $T > 0$*

$$\lim_{n\to\infty} \sup\{||\sum_{i=n}^{k-1} \gamma_{i+1} U_{i+1}|| \, : k = n+1, \ldots, m(\tau_n + T)\} = 0.$$

or equivalently

$$\lim_{t\to\infty} \Delta(t, T) = 0$$

with

$$\Delta(t, T) = \sup_{0 \leq h \leq T} || \int_t^{t+h} \overline{U}(s)ds||. \tag{10}$$

A2 $\sup_n ||x_n|| < \infty$, *or*

A2' *F is Lipschitz and bounded on a neighborhood of* $\{x_n : n \geq 0\}$.

Then the interpolated process X is an asymptotic pseudotrajectory of the flow ϕ *induced by F. Furthermore, under assumption* **A2'**, *for* $t \geq 0$ *large enough we have the estimate*

$$\sup_{0 \leq h \leq T} ||X(t+h) - \Phi_h(X(t))|| \leq C(T)[\Delta(t-1, T+1) + \sup_{t \leq s \leq t+T} (\overline{\gamma}(s))] \quad (11)$$

where $C(T)$ *is a constant depending only on T and F.*

Proof By continuity of F and assumption **A2** there exists $K > 0$ such that $||F(X(t))|| \leq K$ for all $t \geq 0$. Thus (9) and **A1** imply

$$\limsup_{t \to \infty} \sup_{0 \leq h \leq T} ||X(t+h) - X(t)|| \leq KT.$$

Hence X is uniformly continuous.

On the other hand, a simple computation shows that

$$\Theta^t(X) = L_F(\Theta^t(X)) + A_t + B_t \quad (12)$$

where $L_F : C^0(\mathbb{R}, \mathbb{R}^m) \to C^0(\mathbb{R}, \mathbb{R}^m)$ is the continuous function defined as

$$L_F(X)(s) = X(0) + \int_0^s F(X(u))du.$$

and

$$A_t(s) = \int_t^{t+s} [F(\overline{X}(u)) - F(X(u))]du,$$

$$B_t(s) = \int_t^{t+s} \overline{U}(u)du.$$

By assumption **A1**, $\lim_{t \to \infty} B_t = 0$ in $C^0(\mathbb{R}, \mathbb{R}^m)$.

For any $T > 0$ and $t \leq u \leq t+T$, (9) implies

$$||X(u) - \overline{X}(u)|| = || \int_{\tau_{m(u)}}^u F(\overline{X}(s)) + \overline{U}(s)ds||$$

$$\leq K\overline{\gamma}(u) + || \int_{\tau_{m(u)}}^u \overline{U}(s)ds||$$

For t large enough $\overline{\gamma}(u) < 1$, therefore

$$|| \int_{\tau_{m(u)}}^u \overline{U}(s)ds|| \leq || \int_{t-1}^{\tau_{m(u)}} \overline{U}(s)ds|| + || \int_{t-1}^u \overline{U}(s)ds|| \leq 2\Delta(t-1, T+1).$$

Thus

$$\sup_{t \leq u \leq t+T} ||X(u) - \overline{X}(u)|| \leq 2\Delta(t-1, T+1) + \sup_{t \leq u \leq t+T} K\overline{\gamma}(u).$$

Under assumption **A2** F is uniformly continuous on a neighborhood of $\{x_n\}$, therefore $\lim_{t\to\infty} A_t = 0$ in $C^0(\mathbb{R}, \mathbb{R}^m)$.

Let X^* denote a limit point of $\{\Theta^t(X)\}$. Then

$$X^* = L_F(X^*).$$

By uniqueness of integral curves, this implies

$$X^* = \hat{\Phi}(X^*)$$

Therefore Theorem 3.2 shows that X is an asymptotic pseudotrajectory of Φ.

To prove the estimate in case F is Lipschitz with Lipschitz constant L observe that for $0 \leq s \leq T$

$$||A_t(s)|| \leq LT(2\Delta(t-1, T+1) + \sup_{t\leq s\leq t+T} \overline{\gamma}(s)K),$$

$$||B_t(s)|| \leq \Delta(t-1, T+1),$$
$$\Delta(t, T) \leq 2\Delta(t-1, T+1)$$

and by equation (12)

$$||X(t+s) - \Phi_s(X(t))|| \leq L\int_0^s ||X(t+u) - \Phi_u(X(t))||du + ||A_t(s)|| + ||B_t(s)||.$$

Then use Gronwall's inequality. **QED**

4.2 Robbins-Monro Algorithms

In application of Proposition 4.1 to stochastic approximation algorithms one usually tries to verify assumption **A1** by use of maximal inequalities and martingale techniques.

To illustrate this idea let us consider here the simplest case of stochastic approximation algorithms.

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\{\mathcal{F}_n\}$ a nondecreasing sequence of sub-σ-algebras of $\mathcal{F}$. We say that a stochastic process $\{x_n\}$ given by (7) satisfies the *Robbins-Monro* or *Martingale difference Noise* (Kushner and Yin, 1997) condition if

(i) $\{\gamma_n\}$ is a deterministic sequence.

(ii) $\{U_n\}$ is *adapted:* U_n is measurable with respect to $\mathcal{F}_n$ for each $n \geq 0$.

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

The next proposition is a particular case of a general theorem due to Métivier and Priouret (1987). The proof contains several inequalities that will be used later.

Proposition 4.2 *Let $\{x_n\}$ given by (7) be a Robbins-Monro algorithm. Suppose that for some $q \geq 2$*

$$\sup_n E(||U_{n+1}||^q) < \infty$$

and

$$\sum_n \gamma_n^{1+q/2} < \infty.$$

Then assumption **A1** *of proposition 4.1 holds with probability 1.*

Proof For any $t \geq 0$ Burkholder's inequality (see e.g Stroock, 1993) implies

$$E\{\sup_{n<k\leq m(\tau_n+T)} ||\sum_{i=n}^{k-1} \gamma_{i+1} U_{i+1}||^q\} \leq C_q E\{[\sum_{i=n}^{m(\tau_n+T)-1} \gamma_{i+1}^2 ||U_{i+1}||^2]^{q/2}\} \quad (13)$$

for some universal constant $C_q > 0$.

To go further we need the following inequality:

For any $\alpha_i \geq 0$, $\beta_i \in \mathbb{R}$, $u > 1$ and $0 < \delta < 1$

$$(\sum_i |\alpha_i \beta_i|)^u \leq (\sum_i \alpha_i^{\delta u/(u-1)})^{u-1} \sum_i \alpha_i^{(1-\delta)u} |\beta_i|^u \quad (14)$$

The proof of (14) is a consequence of the familiar Hőlder inequality

$$\sum_i x_i y_i \leq (\sum_i x_i^u)^{1/u} (\sum_i y_i^{u/(u-1)})^{(u-1)/u}$$

obtained with $x_i = \alpha_i^{1-\delta}|\beta_i|$ and $y_i = \alpha_i^\delta$.

Suppose $q > 2$. We now apply (14) with $u = q/2$, $\delta = (q-2)/2q$, $\alpha_i = \gamma_{i+1}^2$ and $\beta_i = ||U_{i+1}||^2$. Hence (13) yields

$$E(\sup_{n<k\leq m(\tau_n+T)} ||\sum_{i=n}^{k-1} \gamma_{i+1} U_{i+1}||^q) \leq C_q E((\sum_{i=n}^{m(\tau_n+T)-1} \gamma_{i+1})^{q/2-1} \sum_{i=n}^{m(\tau_n+T)-1} \gamma_{i+1}^{1+q/2} ||U_{i+1}||^q$$

$$\leq C_q T^{q/2-1} E(\sum_{i=n}^{m(\tau_n+T)-1} \gamma_{i+1}^{1+q/2} ||U_{i+1}||^q) \quad (15)$$

$$\leq C(q,T) \sum_{i=n}^{m(\tau_n+T)-1} \gamma_{i+1}^{1+q/2} \leq C(q,T) \int_{\tau_n}^{\tau_n+T} \overline{\gamma}^{q/2}(s) ds$$

for some constant $C(q,T) > 0$.

From the preceding inequality we get that

$$E(\Delta(t,T)^q) \leq C(q,T) \int_t^{t+T} \overline{\gamma}^{q/2}(s) ds \quad (16)$$

where $\Delta(t,T)$ is as in (10). If $q = 2$ inequality (16) follows directly from (13). Hence for $q \geq 2$

$$\sum_{k\geq 0} E(\Delta(kT,T)^q) \leq C(q,T)\int_0^\infty \overline{\gamma}^{q/2}(s)ds = \sum \gamma_{i+1}^{1+q/2} < \infty. \qquad (17)$$

By the Borel-Cantelli Lemma this proves that

$$\lim_{k\to\infty} \Delta(kT,T) = 0$$

with probability one. On the other hand for $kT \leq t < (k+1)T$

$$\Delta(t,T) \leq 2\Delta(kT,T) + \Delta((k+1)T,T).$$

Hence assumption **A1** is satisfied **QED**

Remark 4.3 Suppose that $\{\gamma_n\}$ is a sequence of random variables such that γ_{n+1} is $\mathcal{F}_n$ measurable. Then the conclusion or corollary 4.2 remains valid provided that we strengthen the assumption on $\{U_n\}$ to

$$\sup_n E(||U_{n+1}||^q|\mathcal{F}_n) \leq C$$

for some deterministic constant $C < \infty$, and replace the assumption on $\{\gamma_n\}$ by

$$E(\sum_n \gamma_n^{1+q/2}) < \infty.$$

The sequence $\{U_n\}$ is said to be *subgaussian* if there exists a positive number Γ such that for all $\theta \in \mathbb{R}^m$

$$E(\exp(\langle\theta, U_{n+1}\rangle|\mathcal{F}_n)) \leq \exp(\frac{\Gamma}{2}||\theta||^2).$$

This is for instance the case if $||U_n||$ is bounded by $\sqrt{\Gamma}$.

The following result follows from Duflo (1997) (see also Kushner and Yin (1997) or Benaïm and Hirsch (1996))

Proposition 4.4 *Let $\{x_n\}$ given by (7) be a Robbins-Monro algorithm. Suppose $\{U_n\}$ is subgaussian and $\{\gamma_n\}$ is a deterministic sequence such that*

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

for each $c > 0$. Then assumption **A1** *of proposition 4.1 is satisfied with probability 1. Therefore if* **A2** *and* **A2'** *hold almost surely and F has unique integral curves the interpolated process X is almost surely an asymptotic pseudotrajectory of the flow*

Proof Let

$$Z_n(\theta) = \exp[\sum_{i=1}^{n}\langle\theta, \gamma_i U_i\rangle - \frac{\Gamma}{2}\sum_{i=1}^{n}\gamma_i^2||\theta||^2],\ n \geq 1.$$

By the assumption on $\{U_n\}$, $\{Z_n(\theta)\}$ is a supermartingale. Thus for any $\beta > 0$

$$P(\sup_{n<k\leq m(\tau_n+T)}\langle\theta, \sum_{i=n}^{k-1}\gamma_{i+1}U_{i+1}\rangle \geq \beta)$$

$$\leq P(\sup_{n<k\leq m(\tau_n+T)} Z_k(\theta) \geq Z_n(\theta)\exp(\beta - \frac{\Gamma}{2}||\theta||^2\sum_{i=n}^{m(\tau_n+T)-1}\gamma_{i+1}^2)$$

$$\leq \exp(\frac{\Gamma}{2}||\theta||^2\sum_{i=n}^{m(\tau_n+T)-1}\gamma_{i+1}^2 - \beta).$$

Let $e_1, \ldots, e_m$ be the canonical basis of $\mathbb{R}^m$, $\alpha > 0$ and $e \in \{e_1, \ldots, e_m\} \cup \{-e_1, \ldots, -e_m\}$.

Set

$$R = \alpha/(\Gamma\sum_{i=n}^{m(\tau_n+T)-1}\gamma_{i+1}^2),$$

$\beta = R\alpha$ and $\theta = Re$. Then

$$P(\sup_{n<k\leq m(\tau_n+T)}\langle e, \sum_{i=n}^{k-1}\gamma_{i+1}U_{i+1}\rangle \geq \alpha) = P(\sup_{n<k\leq m(\tau_n+T)}\langle\theta, \sum_{i=n}^{k-1}\gamma_{i+1}U_{i+1}\rangle \geq \beta)$$

$$\leq exp(\frac{-\alpha^2}{2\Gamma\sum_{i=n}^{m(\tau_n+T)-1}\gamma_{i+1}^2}).$$

It follows that

$$P(\Delta(t, T) \geq \alpha) \leq C\exp(\frac{-\alpha^2}{C'\int_t^{t+T}\overline{\gamma}(s)ds}) \leq C\exp(\frac{-\alpha^2}{C'T\overline{\gamma}(u)}) \tag{18}$$

for some $t \leq u \leq t + T$ and some positive constants C, C' depending on m (the dimension of $\mathbb{R}^m$) and Γ. The end of the proof is now exactly as in proposition 4.2. **QED**

Remark 4.5 Propositions 4.2 and 4.4 assume a Robbins Monro type algorithm. However it is not hard to verify that the conclusions of these propositions continue to hold if $\{x_n\}$ satisfies the more general recursion

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

where U_n is a martingale difference noise and $\lim_{n\to\infty} b_n = 0$ almost surely.

4.3 Continuous Time Processes

The technique used in the proof of corollary 4.4 can be easily adapted to analyse a class of continuous time stochastic processes which include certain diffusion and jump processes.

Let $\epsilon : \mathbb{R}_+ \to \mathbb{R}_+^*$ be a continuous non-increasing function. Consider the families of operators $\{L_t^d\}_{t\geq 0}$, $\{L_t^j\}_{t\geq 0}$ and $\{L_t\}$ acting on C^2 functions $f : \mathbb{R}^m \to \mathbb{R}$ according to the formulas

$$L_t^d f(x) = \sum_{i=1}^m G_i(x)\frac{\partial f}{\partial x_i}(x) + \frac{\epsilon(t)}{2}\sum_{i,j} a_{i,j}(x)\frac{\partial^2 f}{\partial x_i \partial x_j}(x) \tag{19}$$

$$L_t^j f(x) = \frac{1}{\epsilon(t)}\int_{\mathbb{R}^m}(f(x+\epsilon(t)v) - f(x))\mu_x(dv) \tag{20}$$

and

$$L_t = L_t^d + L_t^j \tag{21}$$

where

(i) G is a bounded continuous vector field on $\mathbb{R}^m$,

(ii) $a = (a_{ij})$ is a $m \times m$ matrix-valued continuous bounded function such that $a(x)$ is symmetric and nonnegative definite for each $x \in \mathbb{R}^m$.

(iii) $\{\mu_x\}_{x\in\mathbb{R}^m}$ is a family of positive measures on $\mathbb{R}^m$ such that

- $x \to \mu_x(A)$ is measurable for each Borel set $A \subset \mathbb{R}^m$.
- The support of μ_x is contained in a compact set independent of x.

Under these assumptions there exists a nonhomogeneous Markov process $X = \{X(t) : t \geq 0\}$ with sample paths in $D(\mathbb{R}, \mathbb{R}^m)$ (the space of càd lag functions) and initial condition $X(0) = x_0 \in \mathbb{R}^m$ which solves the martingale problem for $\{L_t\}$ (Ethier and Kurtz, 1975; Stroock and Varadhan 1997). That is for each C^∞ function $f : \mathbb{R}^m \to \mathbb{R}$ with compact support,

$$f(X_t) - \int_0^t L_s f(X_s)ds$$

is a martingale with respect to $\mathcal{F}_t = \sigma\{X(s) : s \leq t\}$.

Define the vector field

$$F(x) = G(x) + \int_{\mathbb{R}^m} v\mu_(x)(dv) \tag{22}$$

Proposition 4.6 *Suppose that*

(i) *F is a continuous globally integrable vector field*

(ii)

$$\int_0^\infty exp(-\frac{c}{\epsilon(t)})dt < \infty$$

for all $c > 0$

(iii) $P(\sup_t ||X(t)|| < \infty) = 1$ *or* F *is Lipschitz.*

Then X *is almost surely an asymptotic pseudotrajectory of the flow induced by* F*. Furthermore when* F *is Lipschitz we have the estimate: There exist constant* $C, C(T) > 0$ *such that for all* $\alpha > 0$

$$P(\sup_{0\leq h\leq T} ||X(t+h) - \Phi_t(X(t))|| \geq \alpha) \leq C\exp(-\alpha^2 C(T)/\epsilon(t)).$$

Proof Set $\Delta(t,T) = \sup_{0\leq h\leq T} ||X(t+h) - X(t) - \int_t^{t+h} F(X(s))ds||$

Let $f(x) = exp\langle\theta, x - x_0\rangle$ and $\tau_n = \inf\{t \geq 0\ \overline{X[0,t]} \cap B(0,n)^c \neq \emptyset\}$ where $B(0,n) = \{x \in \mathbb{R}^m\ ||x|| < n\}$. Since the measure μ_x has uniformly bounded support there exists $r > 0$ such that $f(X(t\wedge\tau_n)) = f_n(X(t\wedge\tau_n))$ where f_n is a C^∞ function with compact support which equals f on $B(0, n+r)$. Since $f_n(X(t)) - \int_0^t L_s f_n(X(s))ds$ is a martingale and τ_n a stopping time $f(X(t\wedge\tau_n)) - \int_0^{t\wedge\tau_n} L_s f(X(s))ds$ is a martingale. Hence

$$f(X(t\wedge\tau_n)exp[-\int_0^{t\wedge\tau_n} \frac{L_s f(X(s))}{f(X(s))}ds]$$

is a martingale (Ethier and Kurtz, 1975).

Let $g(u) = e^u - u - 1$. Then

$$\frac{L_s f(x)}{f(x)} = \langle F(x), \theta\rangle + \frac{\epsilon(s)}{2}\langle\theta, a(x)\theta\rangle + \frac{1}{\epsilon(s)}\int_{\mathbb{R}^m} g(\epsilon(s)\langle\theta, v\rangle)\mu_x(dv).$$

Now using the facts that $g(|u|) \leq g(u)$, g is non-decreasing on $\mathbb{R}^+$, $g(u) = u^2/2 + o(u)$ and the boundedness assumptions on x and the support of μ_x it is not hard to verify that there exist a constant $\Gamma > 0$ and $t_0 > 0$ such that for $s \geq t_0$

$$\frac{L_s f(x)}{f(x)} - \langle F(x), \theta\rangle \leq \Gamma||\theta||^2\epsilon(s).$$

Therefore

$$\exp[\langle\theta, X(t\wedge\tau_n) - X(t_0\wedge\tau_n) - \int_{t_0\wedge\tau_n}^{t\wedge\tau_n} F(X(s))\rangle ds - \int_{t_0\wedge\tau_n}^{t\wedge\tau_n} \Gamma||\theta||^2\epsilon(s)ds]$$

is an $\mathcal{F}_t$ supermartingale for $t \geq t_0$. As is Proposition 4.4 we obtain

$$P(\sup_{0\leq h\leq T} ||X((t+h)\wedge\tau_n) - X(t\wedge\tau_n) - \int_{t\wedge\tau_n}^{(t+h)\wedge\tau_n} F(X(s))|| \geq \alpha) \leq C\exp(-\frac{\alpha^2}{C'T\epsilon(t)})$$

for $t \geq t_0$ and by Fatou's lemma we conclude that

$$P(\Delta(t,T) \geq \alpha) \leq C \exp(-\frac{\alpha^2}{C'T\epsilon(t)}).$$

The rest of the proof is now exactly as in Proposition 4.4. Details are left to the reader. **QED**

5 Limit Sets of Asymptotic Pseudotrajectories

5.1 Chain Recurrence and Attractors

In this section we introduce some basic terminology and a few results from (topological) dynamics that will be useful to understand the behavior of asymptotic pseudotrajectories and stochastic approximation processes. In particular we introduce the notion of *chain recurrence* and emphasize its relation with the notion of *attractors* (thanks to Moe Hirsch who taught me the importance of this relation). The material of this section is fairly standard to dynamicists and can be found in numerous places. However since the students for which these notes have been written (as well as the typical reader of the *Séminaire*) may not be familiar with these notions we have tried to give a self contained and comprehensive presentation.

The main and original reference for this section is Conley (1978). The books by Shub (1987) and Robinson (1995) also contain most of the material here.

Basic Notions of Recurrence

Let Φ be a flow or semiflow on the metric space (M,d). We let $\mathbb{T} = \mathbb{R}_+$ if Φ is a semiflow and $\mathbb{T} = \mathbb{R}$ if Φ is a flow.

A subset $A \subset M$ is said *positively invariant* if $\Phi_t(A) \subset A$ for all $t \geq 0$. It is said *invariant* if $\Phi_t(A) = A$ for all $t \in \mathbb{T}$.

A point $p \in M$ is an *equilibrium* if $\Phi_t(p) = p$ for all t. When M is a manifold and Φ is the flow induced by a vector field F, equilibria coincide with zeros of F. A point $p \in M$ is a periodic point of period $T > 0$ if $\Phi_T(p) = p$ for some $T > 0$ and $\Phi_t(p) \neq p$ for $0 < t < T$.

The *forward orbit* of $x \in M$ is the set $\gamma^+(x) = \{\Phi_t(x) : t \geq 0\}$ and the orbit of x is $\gamma(x) = \{\Phi_t(x) : t \in \mathbb{T}\}$. A point $p \in M$ is an *omega limit point* of x if $p = \lim_{t_k \to \infty} \Phi_{t_k}(x)$ for some sequence $t_k \to \infty$. The *omega limit set* of x denoted $\omega(x)$ is the set of omega limit points of x.

If $\gamma^+(x)$ has compact closure, $\omega(x)$ is a compact connected invariant set (It is a good warm up exercise for the reader unfamiliar with these notions) and $\overline{\gamma^+(x)} = \gamma^+(x) \cup \omega(x)$.

If Φ is a flow the *alpha limit set* of x is defined as the omega limit set of x for the reversed flow $\Psi = \{\Psi_t\}$ with $\Psi_t = \Phi_{-t}$.

Further we set $Eq(\Phi)$ the set of equilibria, $Per(\Phi)$ the closure of the set of periodic orbits $\mathcal{L}_+(\Phi) = \overline{\bigcup_{x\in M}\omega(x)}, \mathcal{L}_-(\Phi) = \overline{\bigcup_{x\in M}\alpha(x)}$ and

$$\mathcal{L}(\Phi) = \mathcal{L}_+(\Phi) \cup \mathcal{L}_-(\Phi).$$

Chain Recurrence and Attractors

Equilibria, periodic and omega limit points are clearly "recurrent" points. In general, we may say that a point is recurrent if it somehow returns near where it was under time evolution.

A notion of recurrence related to slightly perturbed orbits and well suited to analyse stochastic approximation processes is the notion of *chain recurrence* introduced by Bowen (1975) and Conley (1978).

Let $\delta > 0$, $T > 0$. A *(δ, T)-pseudo-orbit* from $a \in M$ to $b \in M$ is a finite sequence of partial trajectories

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

such that

$$\begin{aligned} d(y_0, a) &< \delta, \\ d(\Phi_{t_j}(y_j), y_{j+1}) &< \delta,\ j = 0, \ldots, k-1; \\ y_k &= b. \end{aligned}$$

We write $(\Phi : a \hookrightarrow_{\delta,T} b)$ (or simply $a \hookrightarrow_{\delta,T} b$ when there is no confusion on Φ) if there exists a *(δ, T)-pseudo-orbit* from a to b. We write $a \hookrightarrow b$ if $a \hookrightarrow_{\delta,T} b$ for every $\delta > 0$, $T > 0$. If $a \hookrightarrow a$ then a is a *chain recurrent* point. If every point of M is chain recurrent then Φ is a chain recurrent semiflow (or flow).

If $a \hookrightarrow b$ for all $a, b \in M$ we say the flow Φ is *chain transitive.*

We denote by $R(\Phi)$ the set of chain recurrent points for Φ. It is easy to verify (again a good warm up exercise) that $R(\Phi)$ is a closed, positively invariant set and that

$$Eq(\Phi) \subset Per(\Phi) \subset \mathcal{L}(\Phi) \subset R(\Phi.)$$

We will see below that $R(\Phi)$ is always invariant when it is compact (Theorem 5.5).

Let $\Lambda \subset M$ be a nonempty invariant set. Φ is called *chain recurrent on* Λ if every point $p \in \Lambda$ is a chain recurrent point for $\Phi|\Lambda$, the restriction of Φ to Λ. In other words, $\Lambda = R(\Phi|\Lambda)$.

A compact invariant set on which Φ is chain recurrent (or chain transitive) is called an *internally chain recurrent* (or *internally chain transitive*) set.

Example 5.1 Consider the flow on the unit circle $S^1 = \mathbb{R}/2\pi\mathbb{Z}$ induced by the differential equation

$$\frac{d\theta}{dt} = f(\theta)$$

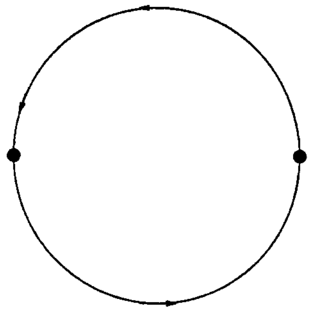

Figure 1: $\dot{\theta} = f(\theta)$

where f is a 2π-periodic smooth nonnegative function such that

$$f^{-1}(0) = \{k\pi : k \in \mathbf{Z}\}.$$

We have

$$Eq(\Phi) = \{0, \pi\} = L_+(\Phi) = L(\Phi)$$

and

$$R(\Phi) = S^1.$$

Internally chain recurrent sets are $\{0\}, \{\pi\}$ and S^1. Remark that the set $X = [0, \pi]$ is a compact invariant set consisting of chain recurrent points. However, X is not internally chain recurrent.

A subset $A \subset M$ is an *attractor* for Φ provided:

(i) A is nonempty, compact and invariant ($\Phi_t A = A$); and

(ii) A has a neighborhood $W \subset M$ such that $\text{dist}(\Phi_t x, A) \to 0$ as $t \to \infty$ uniformly in $x \in W$.

The neighborhood W is usually called a *fundamental neighborhood* of A. The *basin* of A is the positively invariant open set comprising all points x such that $\text{dist}(\Phi_t x, A) \to 0$ as $t \to \infty$. If $A \neq M$ then A is called a *proper* attractor. A *global attractor* is an attractor whose basin is all the space M. An equilibrium (= stationary point) which is an attractor is called *asymptotically stable.*

The following Lemma due to Conley (1978) is quite useful.

Lemma 5.2 *Let $U \subset M$ be an open set with compact closure. Suppose that $\Phi_T(\overline{U}) \subset U$ for some $T > 0$. Then there exists an attractor $A \subset U$ whose basin contains $\overline{U}$.*

Proof By compactness of $\Phi_T(\overline{U})$ there exists an open set V such that $\Phi_T(\overline{U}) \subset V \subset \overline{V} \subset U$. By continuity of the flow there exists $\epsilon > 0$ such that $\Phi_t(\overline{U}) \subset V$ for $T-\epsilon \leq t \leq T+\epsilon$. Let $t_0 = T(T+1)/\epsilon$. For $t > t_0$ write $t = k(T+r/k)$ with $k \in \mathbb{N}$ and $0 \leq r/k < \epsilon$. Therefore for all $x \in \overline{U}$ $\Phi_t(x) = \Phi_{T+r/k} \circ \ldots \circ \Phi_{T+r/k}(x) \in V$.

Then $A_t = \overline{\cup_{s\geq t}\Phi_s(U)} \subset \overline{V} \subset U$. and $A = \cap_{t\geq 0} A_t \subset U$. It is now easy to verify that A is an attractor. Details are left to the reader. **QED**

The following proposition originally due to Bowen (1975) makes precise the relation between the different notions we have introduced.

Proposition 5.3 *Let $\Lambda \subset M$. The following assertions are equivalent*

(i) *Λ is internally chain-transitive*

(ii) *Λ is connected and internally chain-recurrent*

(iii) *Λ is a compact invariant set and $\Phi|\Lambda$ admits no proper attractor.*

Proof $(i) \Rightarrow (ii)$ is easy and left to the reader. $(ii) \Rightarrow (iii)$. Let $A \subset \Lambda$ be a nonempty attractor. To prove that $\Lambda = A$ it suffices to show that A is open and closed in Λ. Let W be an open (in Λ) fundamental neighborhood of A. We claim that $W = A$. Suppose to the contrary that there exists $p \in W \setminus A$. Let $U_\delta = \{x \in \Lambda : d(x, A) \leq \delta\}$. Choose δ small enough so that $U_\delta \subset U_{2\delta} \subset W$ and $B(p,\delta) \subset W \setminus U_{2\delta}$. For T large enough and $t \geq T$ $\Phi_t(W) \subset U_\delta$. Therefore it is impossible to have $p \hookrightarrow_{\delta,T} p$. A contradiction.

$(iii) \Rightarrow (i)$. Let $x \in \Lambda, \delta > 0, T > 0$ and $V = \{y \in \Lambda : (\Phi|\Lambda : x \hookrightarrow_{\delta,T} y)\}$. The set V is open (by definition) and satisfies $\Phi_T(\overline{V}) \subset V$. It then follows from Lemma 5.2 that V contains an attractor but since there are no proper attractors $V = \Lambda$. Since this is true for all $x \in \Lambda, \delta > 0$ and $T > 0$ it follows that Λ is internally chain transitive. **QED**

Corollary 5.4 *If an internally chain transitive set K meets the basin of an attractor A, it is contained in A.*

Proof By compactness, $K \cap A$ is nonempty, hence an attractor for the $\Phi|K$ Since $\Phi|K$ has no proper attractors, being chain transitive, it follows that $K \subset A$. **QED**

The following theorem was proved by Conley (1978) for flows but the proof given here is adapted from a proof given by Robinson (1977) for diffeomorphims.

Theorem 5.5 *If M is compact then $R(\Phi)$ is internally chain recurrent.*

Proof First observe that $R(\Phi)$ is obviously a compact subset of M. Let $p \in R(\Phi)$. For $n \in \mathbb{N}$ and $T > 0$ there exist points $p = p_0^n, \ldots, p_{k_n}^n$ in M and times $t_1, \ldots, t_{k_n}$ with $t_i > T$ such that $p_0^n = p$, $d(\Phi_{t_i}(p_i^n), p_{i+1}^n) < 1/n$ for $i = 0, \ldots k_n - 1$ and $d(p_{k_n}^n, p) < 1/n$. Further we can always assume (by adding points to the sequence) that $t_i \leq 2T$ and since $C_n = \{p_0^n, \ldots, p_{k_n}^n\}$ is a compact subset of M we can also assume (by replacing C_n by a subsequence C_{n_j}) that $\{C_n\}$ converges toward some compact set C for the Hausdorff topology[2]. By

[2] If A and B are closed subsets of M the Hausdorff distance $D(A,B)$ is defined as $D(A,B) = \inf\{\epsilon > 0\, A \subset U_\epsilon(B) \text{and} B \subset U_\epsilon(A)$. This distance makes the space of closed subsets of M a compact space (se e.g Munkres exercise 7 page 279).

construction $C \subset R(\Phi)$ and $p \in C$. Fix $\epsilon > 0$. By uniform continuity of $\Phi : [0, 2T] \times M \to M$ there exists $0 < \delta < \epsilon/3$ so that $d(a, b) \leq \delta \Rightarrow d(\Phi_t(a), \Phi_t(b)) \leq \epsilon$ for all $0 \leq t \leq 2T$. Now for n large enough $C \subset U_\delta(C_n)$ and $C_n \subset U_\delta(C)$. Therefore there exist points $q_i^n \in C$ such that $d(q_i^n, p_i^n) \leq \delta$. It follows that

$$d(\Phi_{t_i}(q_i^n), q_{i+1}^n) \leq d(\Phi_{t_i}(p_i^n), \Phi_{t_i}(q_i^n)) + d(\Phi_{t_i}(p_i^n), p_{i+1}^n) + d(p_{i+1}^n, q_{i+1}^n) \leq \epsilon.$$

Thus we have constructed an (ϵ, T) pseudo orbit from p to p which lies entirely in $C \subset R(\Phi)$. To conclude the proof it remains to show that $R(\Phi)$ is invariant. It is clearly positively invariant. Let p and C_n as above. By extracting convergent subsequences from $\{t_{k_n-1}\}$ and $\{p_{k_n-1}\}$ we obtain points $\tau \in [T, 2T]$ and $p* \in R(\Phi)$ such that $\Phi_\tau(p*) = p$. Hence $p \in \Phi_t(R(\Phi))$ for all $0 \leq t \leq T$ and since p is arbitrary $R(\Phi) \subset \Phi_t(R(\Phi))$ for all $0 \leq t \leq T$. By the semiflow property this implies $R(\Phi) \subset \Phi_t(R(\Phi))$ for all $t \geq 0$. **QED**

Corollary 5.6 *Let $x \in M$ (non-necessarily compact). If $\overline{\gamma(x)^+}$ is compact then $\omega(x)$ is internally chain transitive.*

Proof Let $T = [0, 1] \times \overline{\gamma^+(x)}$ and Ψ the semiflow on T defined by $\Psi_t(u, y) = (e^{-t}u, \Phi_t(y))$. Clearly $\{0\} \times \omega(x)$ is a global attractor for Ψ and points of $\{0\} \times \omega(x)$ are chain recurrent for Ψ. Therefore $R(\Psi) = \{0\} \times \omega(x)$. By Theorem 5.5 $R(\Psi | R(\Psi)) = R(\Psi)$. This implies $R(\Phi | \omega(x)) = \omega(x)$ and $\omega(x)$ being connected it is internally chain transitive by Proposition 5.3. **QED**

5.2 The Limit Set Theorem

Let $X : \mathbb{R}_+ \to M$ be an asymptotic pseudotrajectory of a semiflow Φ. The *limit set* $L(X)$ of X, defined in analogy to the omega limit set of a trajectory, is the set of limits of convergent sequences $X(t_k), t_k \to \infty$. That is

$$L(X) = \bigcap_{t \geq 0} \overline{X([t, \infty))}.$$

Theorem 5.7

(i) *Let X be a precompact asymptotic pseudotrajectory of Φ. Then $L(X)$ internally chain transitive.*

(ii) *Let $L \subset M$ be an internally chain transitive set, and assume M is locally path connected. Then there exists an asymptotic pseudotrajectory X such that $L(X) = L$.*

Proof We only give the proof of (i). We refer the reader to Benaïm and Hirsch (1996) for a proof of (ii) and further results. Since $\{X(t) : t \geq 0\}$ is relatively compact, Theorem 3.2 shows that $\{\Theta^t(X) : t \in \mathbb{R}\}$ is relatively compact in $C^0(\mathbb{R}, M)$ and $\lim_{t\to\infty} d(\Theta^t(X), S_\Phi) = 0$. Therefore by Corollary 5.6 the omega limit set of X for Θ, denoted by $\omega_\Theta(X)$, is internally chain transitive for the semiflow $\theta | S_\Phi$.

The homeomorphism $H : M \to \mathbf{S}_\Phi$, defined by $H(x)(t) = \Phi_t(x)$ conjugates $\Theta|\mathbf{S}_\Phi$ and Φ:

$$(\Theta^t|\mathbf{S}_\Phi) \circ H = H \circ \Phi_t$$

where $t \geq 0$ for a semiflow Φ, and $t \in \mathbb{R}$ for a flow. Since the property of being chain transitive is (obviously) preserved by conjugacy it suffices to verify that

$$H(L(X)) = \omega_\Theta(X)$$

to prove assertion (i). Let $p \in L(X)$. Then $p = \lim_{t_k \to \infty} X(t_k)$. By relative compactness of $\{\Theta^t(X)\}_{t\geq 0}$ we can always suppose that $\{\Theta^{t_k}(X)\}$ converges toward some point $Y \in C^0(\mathbb{R}, M)$. By lemma 3.1 $Y = \hat{\Phi}(Y) = H(Y(0)) = H(p)$. This shows that $H(L(X)) \subset \omega_\Theta(X)$. The proof of the converse inclusion is similar. **QED**

Remark 5.8 Our proof of Theorem 5.7 follows from Benaïm and Hirsch (1996). It has the nice interpretation that the limit set $L(X)$ can be seen as an omega limit set for an extension of the flow to some larger space. A more direct proof in the spirit of Theorem 5.5 can be found in Benaïm (1996) (see also Duflo 1996).

6 Dynamics of Asymptotic Pseudotrajectories

Theorem 5.7 and its applications in later sections show the importance of understanding the dynamics and topology of internally chain recurrent sets (which in most dynamical settings are the same as limit sets of asymptotic pseudotrajectories). Many of the results which appear in the literature on stochastic approximation can be easily deduced (and generalized) from properties of chain recurrent sets. While there is no general structure theory for internally chain recurrent sets, much can be said about many common situations. Several useful results are presented in this section. The main source of this section are the papers (Benaïm, 1996) and Benaïm and Hirsch (1996) but some results have been improved. In particular we give an elementary proof of the convergence of stochastic gradient algorithms with possibly infinitely many equilibria. Several results by Fort and Pages (1996) are similar to those of this section.

We continue to assume that $X : \mathbb{R}_+ \to M$ is an asymptotic pseudotrajectory for a flow or semiflow Φ in a metric space M. Remark that we do not *a priori* assume that X is precompact.

6.1 Simple Flows, Cyclic Orbit Chains

A flow on M is called *simple* if it has only a finite set of alpha and omega limit points (necessarily consisting of equilibria). This property is inherited by the restriction of Φ to invariant sets.

A subset $\Gamma \subset M$ is a *orbit chain* for Φ provided that for some natural number $k \geq 2$, Γ can be expressed as the union

$$\Gamma = \{e_1, \ldots, e_k\} \bigcup \gamma_1 \bigcup \cdots \bigcup \gamma_{k-1}$$

of equilibria $\{e_1, \ldots, e_k\}$ and nonsingular orbits $\gamma_1, \ldots, \gamma_{k-1}$ connecting them: this means that γ_i has alpha limit set $\{e_i\}$ and omega limit set $\{e_{i+1}\}$. Neither the equilibria nor the orbits of the orbit chain are required to be distinct. If $e_1 = e_k$, Γ is called a *cyclic orbit chain.* A homoclinic loop is an example of a cyclic orbit chain.

Concerning cyclic orbit chains, Benaïm and Hirsch (1995a, Theorem 3.1) noted the following useful consequence of the important Akin-Nitecki-Shub Lemma (Akin 1993).

Proposition 6.1 *Let $L \subset M$ be an internally chain recurrent set. If $\Phi|L$ is a simple flow, then every non-stationary point of L belongs to a cyclic orbit chain in L.*

From Theorem 5.7 we thus get:

Corollary 6.2 *Assume that X is precompact and $\Phi|L(X)$ is a simple flow. Then every point of $L(X)$ is an equilibrium or belongs to a cyclic orbit chain in $L(X)$.*

Corollary 6.3 *Assume $L \subset M$ is an internally chain recurrent set such that*

$$\mathcal{L}(\Phi|L) \subset \Lambda = \bigcup_{j=1}^{n} \Lambda_j$$

where $\Lambda_1, \ldots, \Lambda_n$ are compact invariant subsets of L. Then for every point $p \in L$ either $p \in \Lambda$ or there exists a finite sequence $x_1, \ldots, x_k \in L \setminus \Lambda$ and indices $i_1, \ldots, i_k$ such that

(i) $\{i_1, \ldots, i_{k-1}\} \subset \{1, \ldots, n\}$ *and* $i_k = i_1$

(ii) $\alpha(x_{i_l}) \subset \Lambda_{i_l}$, $\omega(x_{i_l}) \subset \Lambda_{i_{l+1}}$ *for* $l = 1, \ldots, k-1$.

In particular if there is no cycle among the Λ_j then $L \subset \Lambda$.

Proof Let $\hat{L}$ be the topological quotient space obtained by collapsing each Λ_i to a point. Let π denote the quotient map $\pi : L \to \hat{L}$. We claim that $\hat{L}$ is metrizable. By the Urysohn metrization Theorem, it suffices to verify that $\hat{L}$ is a regular space with a countable basis.

We first construct a countable basis $\{\hat{U}_n(\hat{x})\}_{n\geq 1}$ at each point $\hat{x} = \pi(x) \in \hat{L}$ as follows. If $x \notin \Lambda$ choose $0 < d_x < d(x, \Lambda)$ and set $\hat{U}_n(\hat{x}) = \pi(B(x, \frac{d_x}{n}))$.

Let $0 < \epsilon < \inf_{i \neq j} d(\Lambda_i, \Lambda_j)$. For $x \in \Lambda_i$ set $\hat{U}_n(\hat{x}) = \{\pi(U_{\epsilon/n}(\Lambda_i)\}$.

Using this basis it is immediate to verify that $\hat{L}$ is Hausdorff, and since it is compact (by continuity of π) it is a regular space. Now let $\{x_i\}_{i\geq 1}$ be a countable dense set in L. The family $\{\hat{U}_n(\hat{x}_i)\}_{n,i\geq 1}$ is a countable basis of $\hat{L}$.

The flow Φ induces a flow $\hat{\Phi}$ on $\hat{L}$ defined by $\hat{\Phi} \circ \pi = \pi \circ \Phi$ which has simple dynamics, and the Λ_j as equilibria.

Let $x \in L$. It is clear, by definition of chain recurrence and uniform continuity of π, that $\pi(x)$ is chain recurrent for $\hat{\Phi}$. Hence $\hat{L}$ is internally chain recurrent and the result follows from Proposition 6.1. **QED**

6.2 Lyapounov Functions and Stochastic Gradients

Let $\Lambda \subset M$ be a compact invariant set of the semiflow Φ. A continuous function $V : M \to \mathbb{R}$ is called a *Lyapounov function* for Λ if the function $t \in \mathbb{R}_+ \to V(\Phi_t(x))$ is constant for $x \in \Lambda$ and strictly decreasing for $x \in M \setminus \Lambda$. If Λ equals the equilibria set $Eq(\Phi)$, V is called a *strict Lyapounov function* and Φ a *gradientlike system.*

Proposition 6.4 *Let $\Lambda \subset M$ be a compact invariant set and $V : M \to \mathbb{R}$ a Lyapounov function for Λ. Assume that $V(\Lambda) \subset \mathbb{R}$ has empty interior. Then every internally chain transitive set L is contained in Λ and $V|L$ is constant.*

Proof Let $L \subset M$ be an internally chain transitive set. Let $v* = \inf\{V(x) : x \in L\}$. We claim that $L \cap \Lambda \neq \emptyset$ and

$$v* = \inf\{V(x) \ : x \in L \cap \Lambda\}.$$

Let $x \in L$. The function $t \to V(\Phi_t(x))$ being non-increasing and bounded the limit $V_\infty(x) = \lim_{t\to\infty} V(\Phi_t(x))$ exists. Therefore $V(p) = V_\infty(x) \leq V(x)$ for all $p \in \omega(x)$. By invariance of $\omega(p)$, V is constant along trajectories in $\omega(x)$. Hence $\omega(x) \subset \Lambda$. This proves the claim.

By continuity of V and compactness of $L \cap \Lambda$, $v* \in V(L \cap \Lambda)$. Since $V(\Lambda)$ has empty interior there exists a sequence $\{v_n\}_{n\geq 1}$, $v_n \in \mathbb{R} \setminus V(\Lambda)$ decreasing to $v*$. For $n \geq 1$ let $L_n = \{x \in L \ : V(x) < v_n\}$. Because V is a Lyapounov function for Λ $\Phi_t(\overline{L_n}) \subset L_n$ for any $t > 0$. Hence by Lemma 5.2 and Proposition 5.3 $L = L_n$. Then $L = \cap_{n\geq 1} L_n = \{x \in L \ : V(x) = v*\}$. This implies $L = \Lambda$ and $V(L) = \{v*\}$. **QED**

Remark 6.5 The following example shows that the assumption that $V(\Lambda)$ has empty interior is essential in Proposition 6.4.

Consider the flow on the unit circle $S^1 = \mathbb{R}/2\pi\mathbb{R}$ induced by the differential equation $\frac{d\theta}{dt} = f(\theta)$ where f is a 2π periodic smooth nonnegative function such that $f^{-1}(0) = \{[k\pi, k(\pi+1)] \ : \ k \in 2\mathbb{Z}\}$. Then S^1 is clearly internally chain transitive. However any 2π periodic smooth nonnegative function $V : S^1 \to \mathbb{R}$ strictly increasing on $]0, \pi[$ and strictly decreasing on $]\pi, 2\pi[$ is a strict Lyapounov function.

Corollary 6.6 *Assume that X is precompact, Φ admits a strict Lyapounov function, and that there are countably many equilibria in $L(X)$. Then $X(t)$ converges to an equilibrium as $t \to \infty$.*

The following corollary is particularly useful in applications since it provides a general convergence result for *stochastic gradient* algorithms.

Corollary 6.7 *Assume M is a smooth C^r Riemannian manifold of dimension $m \geq 1$, $V : M \to \mathbb{R}$ a C^r map and F the gradient vector field*

$$F(x) = -\nabla V(x).$$

Assume

(i) *F induces a global flow* Φ

(ii) *X is a precompact asymptotic pseudotrajectory of* Φ

(iii) $r \geq m$

Then $L(X)$ *consists of equilibria and* $V(X(t))$ *converges as* $t \to \infty$.

Proof Let $\Lambda = Eq(\Phi)$. By Sard's theorem (Hirsch, 1976; chapter 3) $V(\Lambda)$ has Lebesgue measure zero in $\mathbb{R}$ and the result follows from Proposition 6.4 applied with the strict Lyapounov function V. **QED**

6.3 Attractors

Let $X : \mathbb{R} \to M$ be an asymptotic pseudotrajectory of Φ. For any $T > 0$ define

$$d_X(T) = \sup_{k \in \mathbb{N}} d(\Phi_T(X(kT)), X(kT + T)). \tag{23}$$

If a point $x \in M$ belongs to the basin of attraction of an attractor $A \subset M$ then $\Phi_t(x) \to A$ as $t \to \infty$. The next lemma shows that the same is true for an asymptotic pseudotrajectory X provided that $d_X(T)$ is small enough and M is locally compact. This simple lemma will appear to be very useful in the next section.

Lemma 6.8 *Assume* M *is locally compact. Let* $A \subset M$ *be an attractor with basin* $B(A)$ *and let* $K \subset B(A)$ *be a nonempty compact set. There exist numbers* $T > 0, \delta > 0$ *depending only on* K *such that:*

If X *is an asymptotic pseudotrajectory with* $X(0) \in K$ *and* $d_X(T) < \delta$, *then* $L(X) \subset A$.

Proof Choose an open set W with compact closure such that $A \cup K \subset W \subset \overline{W} \subset B(A)$ and choose $\delta > 0$ such that $U_{2\delta}(A)$ (the 2δ neighborhood of A) is contained in W. Since A is an attractor there exists $T > 0$ such that $\Phi_T(W) \subset U_\delta(A)$. Now, if $X(0) \in K$ and $d_X(T) < \delta$ we have $\Phi_T(X(0)) \in U_\delta(A)$ and $d(X(T), \Phi_T(X(0)) < \delta$. Thus $X(T) \in U_{2\delta}(A) \subset W$. By induction it follows that $X(kT) \in W$ for all $k \in \mathbb{N}$. Thus, by compactness, $L(X) \cap \overline{W} \neq \emptyset$ and $L(X)$ is compact as a subset of $\Phi([0,T] \times \overline{W})$. Since points in $L(X) \cap \overline{W}$ are attracted by A and $L(X)$ is invariant, $L(X) \cap A \neq \emptyset$. The conclusion now follows from Proposition 5.3 and Theorem 5.7. **QED**

Below we assume that M is locally compact.

Theorem 6.9 *Let* e *be a asymptotically stable equilibrium with basin of attraction* W *and* $K \subset W$ *a compact set. If* $X(t_k) \in K$ *for some sequence* $t_k \to \infty$, *then* $\lim_{t\to\infty} X(t) = e$.

In the context of stochastic approximations this result was proved by Kushner and Clark (1978). It is an easy consequence of Theorem 5.7 because the only chain recurrent point in the basin of e is e.

More generally we have:

Theorem 6.10 *Let A be an attractor with basin W and $K \subset W$ a compact set. If $X(t_k) \in K$ for some sequence $t_k \to \infty$, then $L(X) \subset A$.*

Proof Follows from Theorem 5.7 and Lemma 6.8. **QED**

Corollary 6.11 *Suppose M is noncompact but locally compact and that Φ is* dissipative *meaning that there exists a global attractor for Φ. Let $M \cup \{\infty\}$ denote the one-point compactification of M. Then either $L(X)$ is an internally chain transitive subset of M or* $\lim_{t\to\infty} X(t) = \infty$.

When applied to stochastic approximation processes such as those described in section 4 Propositions 4.2 and 4.4 under the assumption that F is bounded and Lipschitz, Corollary 6.11 implies that with probability one either $X(t) \to \infty$ or $L(X)$ is internally chain transitive for the flow induced by F.

6.4 Planar Systems

The following result of Benaïm and Hirsch (1994) goes far towards describing the dynamics of internally chain recurrent sets for planar flows with isolated equilibria:

Theorem 6.12 *Assume Φ is a flow defined on $\mathbb{R}^2$ with isolated equilibria. Let L be an internally chain recurrent set. Then for any $p \in L$ one of the following holds:*

(i) p is an equilibrium.

(ii) p is periodic (i.e $\Phi_T(p) = p$ for some $T > 0$).

(iii) There exists a cyclic orbit chain $\Gamma \subset L$ which contains p.

Notice that this rules out trajectories in L which spiral toward a periodic orbit, or even toward a cyclic orbit chain.

In view of Theorem 5.7 we obtain:

Corollary 6.13 *Let Φ be a flow in $\mathbf{R}^2$ with isolated equilibria. If X is a bounded asymptotic pseudotrajectory of Φ then $L(X)$ is a connected union of equilibria, periodic orbits and cyclic orbit chains of Φ.*

The following corollary can be seen as a Poincaré-Bendixson result for asymptotic pseudotrajectories:

Corollary 6.14 *Let Φ be a flow defined on $\mathbf{R}^2$, $K \subset \mathbf{R}^2$ a compact subset without equilibria, X an asymptotic pseudotrajectory of Φ. If there exists $T > 0$ such that $X(t) \in K$ for $t \geq T$, then $L(X)$ is either a periodic orbit or a cylinder of periodic orbits.*

Of course if $X(t)$ is an actual trajectory of Φ, the Poincaré-Bendixson theorem precludes a cylinder of periodic orbits. But this can easily occur for an asymptotic pseudotrajectory.

The next result extends Dulac's criterion for convergence in planar flows having negative divergence:

Theorem 6.15 *Let Φ be a flow in an open set in the plane, and assume that Φ_t decreases area for $t > 0$. Then:*

(a) *$L(X)$ is a connected set of equilibria which is nowhere dense and which does not separate the plane.*

(b) *If Φ has at most countably many stationary points, than $L(X)$ consists of a single stationary point.*

Proof The proof is contained in that of Theorem 1.6 of (Benaïm and Hirsch 1994); here is a sketch. The assumption that Φ decreases area implies that no invariant continuum can separate the plane. A generalization of the Poincaré-Bendixson theorem (Hirsch and Pugh, 1988) shows that an internally chain recurrent continuum (such as $L(X)$) which does not separate the plane consists entirely of stationary points. Simple topological arguments complete the proof. **QED**

Example 6.16 Consider the learning process described in section 2.3. Assume that the probability law of ξ_n is such that functions h^1, h^2 are smooth. Then the divergence of the vector field (6) at every point (x^1, x^2) is

$$Trace(DF(x^1, x^2)) = -2.$$

This implies that Φ_t decreases area for $t > 0$. Since the interpolated process of $\{x_n\}$ is almost surely an asymptotic pseudotrajectory of Φ (use Proposition 4.4), the results of Theorem 6.15 apply almost surely to the limit set of the sequence $\{x_n\}$.

For more details and examples of nonconvergence with more that two players see (Benaïm and Hirsch, 1994; Fudenberg and Levine, 1998).

7 Convergence with positive probability toward an attractor

Throughout this section X is a continuous time stochastic process defined on some probability space $(\Omega, \mathcal{F}, P)$ with continuous (or *càd làg*) paths taking value in M.

We suppose that $X(.)$ is adapted to a non-decreasing sequence of sub-σ algebras $\{\mathcal{F}_t : t \geq 0\}$ and that for all $\delta > 0$ and $T > 0$

$$P(\sup_{s \geq t}[\sup_{0 \leq h \leq T} d(X(s+h), \Phi_h(X(s)))] \geq \delta | \mathcal{F}_t) \leq w(t, \delta, T) \tag{24}$$

for some function $w : \mathbb{R}^3_+ \to \mathbb{R}_+$ such that

$$\lim_{t \to \infty} w(t, \delta, T) \downarrow 0.$$

A sufficient condition for (24) is that

$$P(\sup_{0\leq h\leq T} d(X(t+h),\Phi_h(X(t))) \geq \delta|\mathcal{F}_t) \leq \int_t^{t+T} r(t,\delta,T)dt \qquad (25)$$

for some function $r:\mathbb{R}_+^3 \to \mathbb{R}_+$ such that

$$\int_0^\infty r(t,\delta,T)dt < \infty.$$

This last condition is satisfied by most examples of stochastic approximation processes (see section (4) and section (7.2) below).

Our goal is to give simple conditions ensuring that X converges with positive probability toward a given attractor. We develop here some ideas which originally appeared in Benaïm (1997) and Duflo (1997).

7.1 Attainable Sets

A point $p \in M$ is said to be *attainable* by X if for each $t > 0$ and every open neighborhood U of p

$$P(\exists s \geq t:\ X(s) \in U) > 0.$$

Lemma 7.1 *The set $Att(X)$ of attainable points by X is closed, positively invariant under Φ and contains almost surely $L(X)$.*

Proof A point p lies in $M\setminus Att(X)$ if there exists a neighborhood U of p and $t > 0$ such that

$$P(\forall s \geq t:\ X(s) \notin U) = 1.$$

Hence it is clear that $M\setminus Att(X)$ is an open set almost surely disjoint from $L(X)$.

It remains to prove that given any $p \in Att(X)$ and $T > 0$, $\Phi_T(p) \in Att(X)$. Fix $\epsilon > 0$. By continuity of Φ_T there exists $\alpha > 0$ such that $\Phi_T(B_\alpha(p)) \subset B_{\epsilon/2}(\Phi_T(p))$. Since $p \in Att(X)$ and X is continuous there exists a sequence of rational numbers $\{s_k\}_{k\in\mathbb{N}}$ with $s_k \to \infty$ such that $P(X(s_k) \in B_\alpha(p)) > 0$.

Choose k large enough so that $P(d(\Phi_T(X(s_k)), X(s_k+T)) \geq \epsilon/2)|\mathcal{F}_{s_k}) < 1/2$. Hence $P(X(s_k+T) \in B_\epsilon(\Phi_T(p))) > 0$. This proves that $\Phi_T(p) \in Att(X)$. **QED**

Example 7.2 Let X be the interpolated process associated to the urn process $\{x_n\}$ described in section 2.2. Suppose that the urn function f maps Δ^m into $Int(\Delta^m)$. Then it is not hard to verify that every point of Δ^m is attainable. The proof is left to the reader.

Theorem 7.3 *Suppose M is locally compact. Let $A \subset M$ be an attractor for Φ with basin of attraction $B(A)$. If $Att(X) \cap B(A) \neq \emptyset$ then*

$$P(L(X) \subset A) > 0.$$

Furthermore, if $U \subset M$ is an open set relatively compact with $\overline{U} \subset B(A)$ there exist numbers $T, \delta > 0$ (depending on U) so that

$$P(L(X) \subset A) \geq (1 - w(t, \delta, T))P(\exists s \geq t : X(s) \in U).$$

Proof Let U be an open set such that $K = \overline{U}$ is a compact subset of $B(A)$. To the compact set K we can associate the numbers $T > 0, \delta > 0$ given by Lemma (6.8).

Let $t > 0$ sufficiently large so that $w(t, \delta, T) < 1$. For $n \in \mathbb{N}$ and $k \in \mathbb{N}$ set $t_n(k) = \frac{k}{2^n}$ and let

$$\tau_n = \inf_{k \in \mathbb{N}}\{t_n(k) : X(t_n(k)) \in U \text{ and } t_n(k) \geq t\}.$$

By Lemma (6.8)

$$\{\tau_n < \infty\} \cap \{\sup_{s \geq \tau_n} d(X(s+T), \Phi_T(X(s)) \leq \delta\} \subset \{L(X) \subset A\}.$$

Hence

$$P(L(X) \subset A) \geq \sum_{k \geq [2^n t]+1} E[P(\sup_{s \geq t_n(k)} d(X(s+T), \Phi_T(X(s))) \leq \delta | \mathcal{F}_{t_n(k)}) \mathbf{1}_{\tau_n = t_n(k)}]$$

$$\geq \sum_{k \geq [2^n t]+1} (1 - w(t_n(k), \delta, T))P(\tau_n = t_n(k)) \geq (1 - w(t, \delta, T))P(\tau_n < \infty).$$

Since $\lim_{n \to \infty} P(\tau_n < \infty) = P(\exists s \geq t : X(s) \in U)$ we obtain

$$P(L(X) \subset A) \geq (1 - w(t, \delta, T))P(\exists s \geq t : X(s) \in U).$$

Now, to prove that $P(L(X) \subset A) > 0$ it suffices to choose for U a neighborhood of a point $p \in Att(X) \cap B(A)$. **QED**

7.2 Examples

Proposition 7.4 *Let $F : \mathbb{R}^m \to \mathbb{R}^m$ be a Lipschitz vector field. Consider the diffusion process*

$$dX = F(X)dt + \sqrt{\epsilon(t)}dB_t$$

where ϵ is a positive decreasing function such that for all $c > 0$

$$\int_0^\infty exp(-\frac{c}{\epsilon(t)})dt < \infty.$$

Then

(i) *For each attractor $A \subset \mathbb{R}^m$ of F the event*

$$\Omega_A = \{\lim_{t \to \infty} d(X(t), A) = 0\} = \{L(X) \subset A\}$$

has positive probability and for each open set U relatively compact with $\overline{U} \subset B(A)$

$$P(\Omega_A) \geq P(\exists s \geq t : X(s) \in U)(1 - \int_t^\infty C \exp(-\delta^2 C(T)/\epsilon(s))ds$$

with δ and T given by Lemma 6.8 and $C, C(T)$ are positive constant (depending on F.)

(ii) *On Ω_A $L(X)$ is almost surely internally chain transitive.*

(iii) *If F is a dissipative vector field with global attractor Λ*

$$P(\Omega_\Lambda) = 1 - P(\lim_{t\to\infty} ||X(t)|| = \infty) > 0.$$

Proof (i) follows from the fact that the law of $X(t)$ has positive density with respect to the Lebesgue measure. Hence $Att(X) = \mathbb{R}^m$ and Theorem 7.3 applies. The lower bound for $P(\Omega_A)$ follows from Theorem 7.3 combined with Proposition 4.6, (iii). Statement (iii) follows from Theorems 7.3 and 6.11. **QED**

Similarly we have

Proposition 7.5 *Let $F : \mathbb{R}^m \to \mathbb{R}^m$ be a Lipschitz bounded vector field. Consider a Robbins-Monro algorithm (7) satisfying the assumptions of Proposition 4.2 or 4.4. Then*

(i) *For each attractor $A \subset \mathbb{R}^m$ whose basin has nonempty intersection with $Att(X)$ the event*

$$\Omega_A = \{\lim_{t\to\infty} d(X(t), A) = 0\} = \{L(X) \subset A\}$$

has positive probability and for each open set U relatively compact such that $\overline{U} \subset B(A)$

$$P(\Omega_A) \geq P(\exists s \geq t : X(s) \in U))(1 - \int_t^\infty r(\delta, T, s)ds)$$

where

$$r(\delta, T, s) = Cexp(\frac{-\delta^2 C(T)}{\overline{\gamma}(s)})$$

if $\{U_n\}$ is subgaussian (Proposition 4.4) and

$$r(\delta, T, s) = C'(T, q)(\frac{\sqrt{\overline{\gamma}(s)}}{\delta})^q$$

under the weaker assumptions given by Proposition 4.2 with δ and T given by Lemma 6.8. Here $C, C(T), C'(T, q)$ denote positive constants.

(ii) *On Ω_A $L(X)$ is almost surely internally chain transitive.*

(iii) *If F is a dissipative vector field with global attractor Λ*

$$P(\Omega_\Lambda) = (1 - P(\lim_{t\to\infty} ||X(t)|| = \infty)) > 0.$$

7.3 Stabilization

Most of the results given in the preceding sections assume a precompact asymptotic pseudotrajectory X for a semiflow Φ. Actually when X is not precompact the long term behavior of X usually presents little interest (See Corollary 6.11).

For stochastic approximation processes there are several *stability* conditions which ensure that the paths of the process are almost surely bounded. Such conditions can be found in numerous places such as Nevelson and Khasminskii (1976), Benveniste et al (1990), Delyon (1996), Duflo (1996, 1997), Fort and Pages (1996), Kushner and Yin (1997), to name just a few. We present here a theorem due to Kushner and Yin (1997, Theorem 4.3).

Theorem 7.6 *Let*

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

be a Robbins-Monro algorithm (section 4.2). Suppose that there exists a C^2 *function* $V : \mathbb{R}^m \to \mathbb{R}_+$ *with bounded second derivatives and a nonnegative function* $k : \mathbb{R}^m \to \mathbb{R}_+$ *such that:*

(i) $\langle \nabla V(x), F(x) \rangle \leq -k(x)$.

(ii) $\lim_{||x|| \to \infty} V(x) = +\infty$.

(iii) *There are positive constants* K, R *such that*

$$E(||U_{n+1}||^2 + ||F(x_n)||^2 | \mathcal{F}_n) \leq Kk(x_n)$$

when $||x_n|| \geq R$ *and*

$$E(\sum_n \gamma_{n+1}^2 (||U_{n+1}||^2 + ||F(x_n)||^2) 1_{||x_n|| \leq R}) < \infty.$$

(iv) $E(k(x_n)) < \infty$ *if* $V(x_n) < \infty$ *and* $E(V(x_0)) < \infty$.

Then $\limsup_{n \to \infty} ||x_n|| < \infty$ *with probability one.*

Proof A second order Taylor expansion and boundedness of the second derivative of V implies the existence of some constant $K_1 > 0$ such that

$$E(V(x_{n+1}) - V(x_n) | \mathcal{F}_n) \leq -\gamma_{n+1} k(x_n) + \gamma_{n+1}^2 K_1 E(||U_{n+1}||^2 + ||F(x_n)||^2) | \mathcal{F}_n).$$

The hypotheses then imply that $E(V(x_n)) < \infty$ for all n. Let

$$W_n = K_1 E(\sum_{i \geq n} \gamma_{i+1}^2 (||U_{i+1}||^2 + ||F(x_i)||^2) 1_{||x_i|| \leq R}) | \mathcal{F}_n)$$

and $V_n = V(x_n) + W_n$. V_n is nonnegative and

$$E(V_{n+1} - V_n | \mathcal{F}_n) \leq -k(x_n)\gamma_{n+1} + K_1 K k(x_n) \gamma_{n+1}^2.$$

Since $\gamma_n \to 0$ there exists $n_0 \geq 0$ such that $E(V_{n+1} - V_n|\mathcal{F}_n) \leq 0$ for $n \geq n_0$. Since $V_n \geq 0$ and $E(V_{n_0}) < \infty$ the supermartingale convergence theorem implies that $\{V_n\}$ converges with probability one toward some nonnegative L^1 random variable V. Since assumption (iii) implies that $W_n \to 0$ with probability one, $V(x_n) \to V$ with probability one. By assumption (ii) we then must have $\limsup_{n\to\infty} V(x_n) < \infty$ **QED**

8 Shadowing Properties

In this section we consider the following question:

Given a stochastic approximation process such as (7) (or more generally an asymptotic pseudotrajectory for a flow Φ) does there exist a point x such that the omega limit set of the trajectory $\{\Phi_t(x) : t \geq 0\}$ is $L(X)$?

The answer is generally negative and $L(X)$ can be an arbitrary chain transitive set. However it is useful to understand what kind of conditions ensure a positive answer to this question. A case of particular interest in applications is given by the following problem: Assume that each $\Phi-$ trajectory converges toward an equilibrium. Does X converge also toward an equilibrium ?

The material presented in this section is based on the works of Hirsch (1994), Benaïm (1996), Benaïm and Hirsch (1996), Duflo (1996) and Schreiber (1997).

We begin by a illustrative example borrowed from Benaïm (1996) and Duflo (1996).

Example 8.1 Consider the Robbins Monro algorithm given in polar coordinates (ρ, θ) by the system

$$\rho_{n+1} - \rho_n = \gamma_{n+1}\rho_n(h(\rho_n^2) + g(\rho_n)\xi_{n+1}),$$

$$\theta_{n+1} - \theta_n = \gamma_{n+1}((\rho_n sin(\theta_n))^2 + \xi_{n+1}).$$

where $\{\xi_n\}$ is a sequence of $i.i.d$ random variables uniformly distributed on $[-1, 1]$ and $\{\gamma_n\}$ satisfies the condition of Proposition 4.4. The function h is a smooth function such that $h(u) = 1 - u$ for $0 \leq u \leq 4$ and $-3 \leq h(u) \leq -4$ for $u \geq 4$, $g(\rho) = \mathbf{1}_{\{1/2 \leq \rho \leq 2\}}$, and $\gamma_n < 1/4$ for all n. These choices ensure that the algorithm is well defined (i.e $\rho_0 \geq 0$ implies $\rho_n \geq 0$ for all $n \geq 0$).

We suppose given $\rho_0 > 0$. It is then not hard to verify there exist some constants $0 < k(\rho_0) < K(\rho_0)$ such that $k(\rho_0) \leq \rho_n \leq K(\rho_0)$ for all $n \geq 0$.

Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be the vectorfield defined by

$$F(x, y) = (xh(x^2 + y^2) - y^3,\ yh(x^2 + y^2) + xy^2). \tag{26}$$

Then $X_n = (x_n, y_n) = (\rho_n cos(\theta_n),\ \rho_n sin(\theta_n))$ satisfies a recursion of the form

$$X_{n+1} - X_n = \gamma_{n+1}(F(X_n) + U_{n+1}) + 0(\gamma_{n+1}^2)$$

where $\{U_n\}$ is a sequence of bounded random variables such that $E(U_{n+1}|\mathcal{F}_n) = 0$.

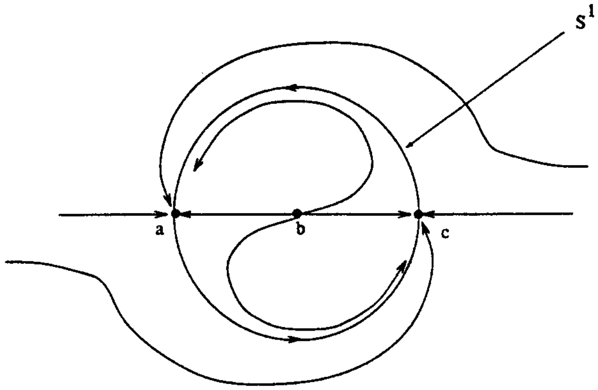

Figure 2: The phase portrait of $F(x,y) = (xh(x^2+y^2) - y^3,\ yh(x^2+y^2) + xy^2)$.

Let Φ be the flow induced by F (see figure 2). Equilibria of Φ are the points $a = (-1,0)$, $b = (0,0)$, $c = (0,1)$, and every trajectory of Φ converges toward one of these equilibria. Internally chain transitive sets are the equilibria $\{a\}, \{b\}, \{c\}$ and the unit circle $S^1 = \{\rho = 1\}$ which is a cyclic orbit chain.

Since $\{(x_n, y_n)\}$ lives in some compact set disjoint from the origin, Theorem 5.7 combined with Proposition 4.4 and Remark 4.5 imply that the limit set of $\{(x_n, y_n)\}$ is almost surely one of the sets $\{a\}, \{c\}$ or S^1. We claim that if $\sum \gamma_n^2 = \infty$ then this limit set is almost surely S^1. Suppose on the contrary that $\{(x_n, y_n)\}$ converges toward one of the points a or c. Then $\lim_{n\to\infty} d(\theta_n, \pi\mathbb{Z}) = 0$ and since $|\theta_{n+1} - \theta_n| = 0(\gamma_{n+1})$ the sequence $\{\theta_n\}$ must converges. On the other hand by the law of iterated logarithm for martingales $\limsup_{n\to\infty} \sum_{i=1}^n \gamma_i \xi_i = \infty$. Thus

$$\limsup_{n\to\infty} \theta_n \geq \theta_0 + \limsup_{n\to\infty} \sum_{i=1}^{n} \gamma_i \xi_i = \infty.$$

A contradiction.

This example shows that the limiting behavior of a stochastic approximation process can be quite different from the limiting behavior of the associated ODE. We will show later (see Example 8.16) that $\{(x_n, y_n)\}$ actually converges toward one the points a or c provided that γ_n goes to zero "fast enough".

8.1 λ-Pseudotrajectories

Let X denote an asymptotic pseudotrajectory for a semiflow Φ on the metric space M. For $T > 0$ let

$$e(X,T) = \limsup_{t\to\infty} \frac{1}{t} \log\left(\sup_{0\leq h\leq T} d(X(t+h), \Phi_h(X(t))\right)$$

and define the *asymptotic error rate of* X to be

$$e(X) = \sup_{T>0} e(X,T).$$

If $e(X) \leq \lambda < 0$ we call X a *$\lambda-$pseudotrajectory* of Φ.

The maps $\{\Phi_t\}$ are said to be Lipschitz, locally uniformly in $t > 0$ if for each $T > 0$ there exists $L(T) \geq 0$ such that

$$d(\Phi_h(x), \Phi_h(y)) \leq L(T)d(x, y)$$

for all $0 \leq h \leq T$, $x, y \in M$.

Lemma 8.2 *If the $\{\Phi_t\}$ are Lipschitz, locally uniformly in $t > 0$ then $e(X,T) = e(X)$ for all $T > 0$.*

Proof Let $T' > T > 0$. It is clear from the definition that $e(X,T) \leq e(X,T')$. Conversely, write $T' = kT + r$ with $k \in \mathbb{N}$ and $0 \leq r < T$ and set $D(X,R,t) = \sup_{0 \leq h \leq R} d(\Phi_h(X(t), X(t+h))$. Then

$$D(X,T',t) \leq D(X,(k+1)T,t) \leq \sup\{D(X,kT,T), L(T)D(X,kT,t)+D(X,T,t+kT)\}.$$

Thus

$$e(X,T') \leq e(X,(k+1)T) \leq \sup\{e(X,T), e(X,kT)\}.$$

Therefore $e(X,T') \leq e(X,(k+1)T) \leq e(X,T)$. **QED**

Main Examples

Our main example of λ-pseudotrajectories is given by stochastic approximation processes whose step sizes go to zero at a "fast" rate:

Proposition 8.3 *Let $\{x_n\}$ given by (7) be a Robbins-Monro algorithm. Suppose*

$$l(\gamma) = \limsup_{n\to\infty} \frac{\log(\gamma_n)}{\tau_n} < 0.$$

Assume that F is a Lipschitz bounded vector field and that $\{U_n\}$ satisfies the assumptions of Proposition 4.2. Then X (the interpolated process) is almost surely a $\frac{l(\gamma)}{2}$-pseudotrajectory of Φ (the flow induced by F).

Proof Set $\lambda = l(\gamma)$ and let $0 < \epsilon < -\lambda$. For t large enough $\overline{\gamma}(t) \leq e^{(\lambda+\epsilon)t}$. Therefore, with $\Delta(t,T)$ given by (10) and $k \in \mathbb{N}$ large enough, equation (16) implies

$$E(\Delta(kT,T)^q) \leq TC(q,T) \exp \frac{kT(\lambda+\epsilon)q}{2}.$$

Let $\alpha > 0$ be such that $\frac{\lambda+\epsilon}{2} + \alpha < 0$. By Markov inequality

$$P(\Delta(kT,T) \geq e^{-kT\alpha}) \leq TC(q,T) \exp(qkT(\alpha + \frac{\lambda+\epsilon}{2})).$$

By Borel Cantelli Lemma this implies that

$$\limsup_{k\to\infty} \frac{1}{k} \log(\Delta(kT,T)) \leq -T\alpha$$

almost surely. Since α can be chosen arbitrary close to $-\lambda/2$ this proves that

$$\limsup_{k\to\infty} \frac{1}{kT} \log(\Delta(kT, T)) \leq \lambda/2$$

almost surely. Since $\Delta(t, T) \leq 2\Delta(kT, T) + \Delta((k+1)T, T)$ we get that

$$\limsup_{t\to\infty} \frac{1}{t} \log(\Delta(t, T)) \leq \lambda/2$$

almost surely and we conclude the proof by using inequality (11). **QED**

Remark 8.4 If $\gamma_n = f(n)$ for some positive decreasing function with $\int_1^\infty f(s)ds = \infty$, then

$$l(\gamma) = \limsup_{x\to\infty} \frac{log(f(x))}{\int_1^x f(s)ds}.$$

For example, if

$$\gamma_n = \frac{A}{n^\alpha \log(n)^\beta}$$

then $l(\gamma) = 0$ for $0 < \alpha < 1$ and $\beta \geq 0$, $l(\gamma) = -1/A$ for $\alpha = 1$ and $\beta = 0$, and $l(\gamma) = -\infty$ for $\alpha = 1$ and $0 < \beta \leq 1$.

Similar to Proposition 8.3 is the next proposition whose proof is left to the reader:

Proposition 8.5 *Let X be the continuous time Markov process associated to the generator (21) and let*

$$\lambda = \limsup_{t\to\infty} \frac{\log(\epsilon(t))}{2t}.$$

If the vector field F given by (22) is Lipschitz continuous then X is almost surely a λ-pseudotrajectory of the flow induced by F.

Consider now the following situation. Suppose that X is a λ-pseudotrajectory whose limit set is contained in some compact positively invariant set K. Let $Y(t) \in K$ denote a point nearest to $X(t)$. It is not true in general that Y is a λ-pseudotrajectory for $\Phi|K$ but, for reasons that will be made clear later, it may be useful to know when this is true. The end of this section is devoted to this question.

Let $K \subset M$ be a compact positively invariant set for Φ and $B \subset M$ a set containing K. We say that K *attracts B exponentially at rate $\alpha < 0$* if there exists $C > 0$ such that

$$d(\phi_t(x), K) \leq Ce^{\alpha t} d(x, K)$$

for all $x \in B$ and $t \geq 0$.

Example 8.6 Suppose Φ is a C^1 flow on $\mathbb{R}^m$ and $\Gamma \subset \mathbb{R}^m$ is a periodic orbit of period $T > 0$. For any $p \in \Gamma$ let $\lambda_1 = e^{\mu_1}, \ldots, \lambda_m = e^{\mu_m}$ be the eigenvalues of $D\Phi_T(p)$ (counted with their multiplicities). The λ_i are are called the *characteristic (or Floquet) multipliers.* They are independent on p and the unity is always a Floquet multiplier (see Hartman (1964)).

Let $\alpha < 0$. If 1 has multiplicity 1 and the remaining $m-1$ Floquet multipliers are strictly inside the complex disk of center 0 and radius e^α then there exists a neighborhood B of Γ such that Γ attracts exponentially B at rate α/T. In this case Γ is called an *attracting hyperbolic periodic orbit.*

Lemma 8.7 *Let $\alpha < 0, \lambda < 0$ and $\beta = \sup(\alpha, \lambda)$. Suppose $K \subset B$ attracts exponentially B at rate α. Let X be a λ-pseudotrajectory for Φ such that $X(t) \in B$ for all $t \geq 0$ and let $Y(t) \in K$ be a point nearest to $X(t)$. Then*

(i)
$$\limsup_{t\to\infty} \frac{1}{t} \log d(X(t), Y(t)) \leq \beta.$$

(ii) *If the $\{\Phi_t\}$ are Lipschitz, locally uniformly in $t > 0$ then Y is a β–pseudotrajectory for $\Phi|K$*

Proof Choose $0 < \epsilon \leq -\beta$ and choose $T > 0$ large enough such that

$$d(\Phi_T(x), K) \leq e^{(\alpha+\epsilon)T} d(x, K)$$

for all $x \in B$. Thus there exists t_0 such that for $t \geq t_0$

$$d(X(t+T), K) \leq d(X(t+T), \Phi_T(X(t)) + d(\Phi_T(X(t)), K) \leq e^{(\lambda+\epsilon)t} + e^{(\alpha+\epsilon)T} d(X(t), K).$$

Let $v_k = d(X(kT), K)$, $\rho = e^{(\beta+\epsilon)T}$ and $k_0 = [t_0/T] + 1$. Then $v_{k+1} \leq \rho^k + \rho v_k$ for $k \geq k_0$. Hence

$$v_{k_0+m} \leq \rho^m (m\rho^{k_0-1} + v_{k_0})$$

for $m \geq 1$. It follows that

$$\limsup_{k\to\infty} \frac{\log(v_k)}{kT} \leq \frac{\log(\rho)}{T} = \beta + \epsilon.$$

Also for $kT \leq t \leq (k+1)T$ and $k \geq k_0$

$$d(X(t), K) \leq d(\Phi_{t-kT}(X(kT), X(t)) + d(\Phi_{t-kT}(X(kT)), K)$$
$$\leq e^{(\lambda+\epsilon)kT} + Ce^{\alpha(t-kT)} v_k \leq e^{(\lambda+\epsilon)kT} + Cv_k.$$

Thus

$$\limsup_{t\to\infty} \frac{log(d(X(t), K)}{t} \leq \beta + \epsilon$$

and since ϵ is arbitrary we get the desired result.

To obtain (ii) observe that

$$d(Y(t+h), \Phi_h(Y(t)) \leq d(\Phi_h(Y(t)), \Phi_h(X(t)) + d(\Phi_h(X(t)), X(t+h)) + d(X(t+h), Y(t+h)$$

Then for t large enough and $T > 0$

$$\sup_{0\leq h\leq T} d(Y(t+h), \Phi_h(Y(t)) \leq L(T)d(X(t), Y(t)) + e^{(\lambda+\epsilon)t} + \sup_{0\leq h\leq T} d(X(t+h), Y(t+h)).$$

QED

8.2 Expansion Rate and Shadowing

From now on we assume that M is a Riemannian manifold and Φ a C^1 flow on M. The norm of a tangent vector v in the Riemannian metric is denoted by $||v||$. In our applications M will be a submanifold of $\mathbb{R}^m$ positively invariant under the flow generated by a smooth vector field.

Let $K \subset M$ denote a compact positively invariant set. The *expansion constant* of Φ_t at K (Hirsch, 1994) is the number

$$EC(\Phi_t, K) = \inf_{x\in K} m(D\Phi_t(x))$$

where

$$m(D\Phi_t(x)) = \inf\{||D\Phi_t(x)v|| \, : ||v|| = 1\}$$

denote the *minimal norm* of $D\Phi_t(x)$. Observe that since Φ is a flow then

$$m(D\Phi_t(x)) = ||D\Phi_t(x)^{-1}||^{-1} = ||D\Phi_{-t}(\Phi_t(x))||^{-1}.$$

The *expansion rate* of Φ at K is defined as

$$\mathcal{E}(\Phi, K) = \lim_{t\to\infty} \frac{1}{t} \log(EC(\Phi_t, K))$$

where the limit exists by a standard subadditivity argument whose verification is left to the reader.

Remark 8.8 It is important to understand that the expansion rate of Φ at K depends on the dynamics of Φ in M *and not only* in K. As a simple example illustrating this point, consider a smooth flow in $\mathbb{R}^m$ having a non-stationary periodic orbit Γ of period $T > 0$. Then it is not hard to see that $\mathcal{E}(\Phi, \Gamma)$ equals the smallest real part of the Floquet exponents of Γ divided by T. (this easily follows from Theorem 8.12.) If we now set $M = \Gamma$ and $\Psi = \Phi|\Gamma$ then $\mathcal{E}(\Psi, \Gamma) = 0$.

We now state a shadowing result due to Benaïm and Hirsch (1996) whose proof is an (easy) adaptation of Hirsch's shadowing theorem (Hirsch,1994).

Theorem 8.9 *Let $K \subset M$ be a compact positively invariant set. Let X be a λ-pseudotrajectory for Φ. Suppose*

(a) $L(X) \subset K$.

(b) $\lambda < \min\{0, \mathcal{E}(\Phi, K)\}$.

Then

(i) *There exists $r \geq 0$ and $x \in M$ such that*

$$\limsup_{t\to\infty} \frac{1}{t} \log d(X(t), \Phi_{t+r}(x)) \leq \lambda.$$

(ii) *Let x be as in (i). Suppose $l \geq 0$, $y \in M$ are such that*

$$\limsup_{t\to\infty} \frac{1}{t} \log d(X(t), \Phi_{t+l}(y)) \leq \lambda.$$

Then x and y are on the same orbit of Φ.

Proof Since $\lambda < \mathcal{E}(\Phi, K)$ we can choose $T > 0$ large enough so that

$$\min_{x\in K} m(D\Phi_T(x)) > e^{T\lambda}.$$

Set $f = \Phi_T$, $y_k = X(kT)$ and fix ν such that $e^{\lambda T} < \nu < \min_{x\in K} m(Df(x))$. Thus for k large enough

$$d(y_{k+1}, f(y_k)) \leq \nu^k. \tag{27}$$

By continuity of Df and compactness of K there exists a neighborhood U of K such that

$$\min_{x\in U} m(Df(x)) = \mu > \nu. \tag{28}$$

Claim: There exists a neighborhood $N \subset U$ of K and $\rho^* > 0$ such that

$$B(f(x), \rho\mu) \subset f(B(x, \rho))$$

for all $x \in N$ and $\rho \leq \rho^*$.

Proof of the claim: Choose a neighborhood U' of K, and $r > 0$, small enough such that for all $y \in U'$ and $d(y, z) \leq r$ there exists a C^1 curve $\gamma_{y,z} : [0, 1] \to U$ with the properties:

(i) $\gamma_{y,z}(0) = y$, $\gamma_{y,z}(1) = z$,

(ii) $\gamma_{y,z}([0, 1]) \subset U$,

(iii) $\int_0^1 ||\gamma'_{y,z}(s)||ds = d(y, z)$.

Set $N = f^{-1}(U') \cap U$ and $\rho^* = \frac{r}{\mu}$. Let $x \in N, \rho \leq \rho^*$ and $d(z, f(x)) \leq \rho\mu$. Then

$$d(f^{-1}(z), x) = d(f^{-1}(z), f^{-1}(f(x))) \leq \int_0^1 ||Df^{-1}(\gamma_{f(x),z}(s))\gamma'_{f(x),z}(s)||ds$$

$$\leq \frac{1}{\mu} \int_0^1 ||\gamma'_{f(x),z}(s)||ds = \frac{1}{\mu} d(f(x), z) \leq \rho.$$

This proves the claim.

Since $L(X) \subset K$, $y_k \in N$ for k large enough. Let $\nu < \delta < \min\{1, \mu\}$. We claim that for k large enough

$$B(y_k, \delta^k) \subset f(B(y_{k-1}, \delta^{k-1})). \tag{29}$$

Indeed let $z \in B(y_k, \delta^k)$. Then (27) implies that for k large enough

$$d(z, f(y_{k-1})) \leq d(z, y_k) + d(f(y_{k-1}), y_k) \leq \delta^k + \nu^k$$

Since $\nu < \delta < \mu$, $\delta^k + \nu^{k-1} \leq \delta^{k-1}\mu$ for $k \geq \frac{\mu - \delta}{\log(\delta/\nu)} + 1$. Thus for k large enough, say $k \geq m$,

$$z \in B(f(y_{k-1}), \delta^{k-1}\mu) \subset f(B(y_{k-1}, \delta^{k-1})$$

where the last inclusion follows from the claim. This proves (29).

Set $B_k = B(y_k, \delta^k)$. For $n \geq m$ estimate (29) implies that

$$Q_n = \bigcap_{i \geq 0} f^{-i}(B_{n+i})$$

is a nonempty compact set. Also for $z \in Q_n$, $d(f^i(z), y_{n+i})) \leq \delta^{n+i}$. This proves statement (i) of the Theorem.

Since $\delta < \mu$ the claim shows that the diameter of $f^{-i}(B_{n+i})$ goes to zero as $i \to \infty$. This implies that $Q_n = \{z_n\}$ all $n \geq m$ where $z_n = \lim_{i \to \infty} f^{-i}(y_{n+i})$. This implies (ii). **QED**

Corollary 8.10 *Let Φ be a semiflow on M, $A \subset M$ a positively invariant submanifold of M and $K \subset A$ a compact positively invariant set. Let X be a λ-pseudotrajectory for Φ. Suppose that*

(a) $L(X) \subset K$.

(b) *There is a neighborhood of K (in M) which is attracted exponentially at rate $\alpha < 0$ by K.*

(c) *There is a C^1 flow Ψ on A such that $\Phi_t | A = \Psi_t$ for all $t \geq 0$.*

(d) $\beta = \sup(\alpha, \lambda) < \min\{0, \mathcal{E}(\Psi, K)\}$.

Then there exist $r \geq 0$ and $x \in A$ such that

$$\limsup_{t \to \infty} \frac{1}{t} \log d(X(t), \Phi_{t+r}(x)) \leq \beta.$$

Proof Let $Y(t) \in K$ be a point nearest to $X(t)$. By Lemma 8.7 (ii), Y is a β-pseudotrajectory for Ψ, and the result follows from Theorem 8.9 combined with estimate (i) of Lemma 8.7 **QED**

Example 8.11 As an illustration of Corollary 8.10 consider the diffusion on $\mathbb{R}^m$

$$dX = F(X)dt + \sqrt{\epsilon(t)}dB_t$$

where F and ϵ are as in Proposition 7.4.

Let $\Gamma \subset \mathbb{R}^m$ be an *attracting hyperbolic periodic orbit* of period $T > 0$. (see Example 8.6) such that the multipliers distinct of the unity have moduli $< e^{\alpha}$ for some $\alpha < 0$.

According to Proposition 7.4 the event $\Omega_\Gamma = \{L(X) \subset \Gamma\}$ has positive probability. If we furthermore assume that

$$\lambda = \limsup \frac{1}{2t} \log(\epsilon(t)) < 0$$

then Corollary 8.10 applied with $A = K = \Gamma$ and Proposition 8.5 imply that for almost every $\omega \in \Omega_\Gamma$ there exists $x(\omega) \in \Gamma$ such that

$$\limsup \frac{1}{t} \log(d(X(t), \Phi_t(x(\omega))) \leq \sup(\lambda, \alpha/T).$$

8.3 Properties of the Expansion Rate

This section presents several useful estimates of the expansion rate. The key result is an ergodic characterization of the expansion rate due to Schreiber (1997).

We continue to assume that Φ is a C^1 flow on a Riemannian manifold M. Since we are concerned by the behavior of Φ restricted to a compact positively invariant set we furthermore assume without loss of generality that M is compact.

In order to present Schreiber's result we need to introduce a few notions of ergodic theory. We let $\mathcal{P}(M)$ denote the space of Borel probability measures on M with the topology of weak convergence[3]. A Borel probability measure $\mu \in \mathcal{P}(M)$ is said $\Phi-$*invariant* or *invariant* under Φ if for every Borel set $A \subset M$ and every $t \in \mathbb{R}$

$$\mu(A) = \mu(\Phi_t(A)).$$

We let $\mathcal{M}(\Phi) \subset \mathcal{P}(M)$ denote the set of invariant measures. It is a nonempty compact convex subset of $\mathcal{P}(M)$ (Mañé, 1987, chapter 1). A measure $\mu \in \mathcal{M}(\Phi)$ is said $\Phi-$*ergodic* if every Φ-invariant set has measure 0 or 1.

The *minimal center of attraction* of Φ is the set

$$MC(\Phi) = \overline{\bigcup_{\mu \in \mathcal{M}(\Phi)} supp(\mu)}.$$

The Birkhoff center of Φ is the set

$$BC(\Phi) = \overline{\{x \in M \; : x \in \omega(x)\}}.$$

By the Poincaré recurrence Theorem (Mañé, 1987, chapter 1)

$$MC(\Phi) \subset BC(\Phi). \tag{30}$$

[3] the one a functional analyst would call weak*

Let $\mu \in \mathcal{M}(\Phi)$. By the celebrated Oseledec's Theorem (see e.g Mañé 1987, chapter 11) there exists a Borel set $R \subset M$ of full measure ($\mu(R) = 1$) such that for all $x \in R$, there exist numbers $\lambda_1(x) < \ldots \lambda_{r(x)}(x)$ and a decomposition of T_xM into

$$T_xM = E_1(x) \oplus \ldots \oplus E_{r(x)}(x)$$

such that for all $v \in E_i(x)$

$$\lim \frac{1}{t} \log ||D\Phi_t(x)v|| = \lambda_i(x).$$

The maps $x \to r(x), \lambda_i(x), E_i(x)$ are measurable.

When μ is ergodic there is an integer $r(\mu)$ and real numbers $\lambda_i(\mu)$, $i = 1, \ldots, r(\mu)$, such that $r(x) = r(\mu)$ and $\lambda_i(x) = \lambda_i(\mu)$ for μ almost every x. The $\lambda_i(\mu)$ are called the *Lyapounov exponents* of μ.

Schreiber (1997) proves the following result:

Theorem 8.12 *Let $K \subset M$ be a compact positively invariant set. Then*

$$\mathcal{E}(\Phi, K) = \inf_{\mu} \lambda_1(\mu)$$

where the infimum is taken over all ergodic measures with support in K.

Proof Let $f = \Phi_1$ and let $X = \{(x, v) : x \in K, v \in T_xM, ||v|| = 1.\}$ By compactness of X and definition of $\mathcal{E}(\Phi, K)$ there exists a sequence $(x_n, v_n) \in X$ such that

$$\mathcal{E}(\Phi, K) = \lim_{n \to \infty} \frac{1}{n} \log(EC(f^n, K)) = \lim_{n \to \infty} \frac{1}{n} \log(||Df^n(x_n)v_n||).$$

Define a map $G : X \to X$ by

$$G(x, v) = (f(x), \frac{Df(x)v}{||Df(x)v||}).$$

Let $\{\theta_n\}_{n \geq 0}$ be the sequence of probability measures defined on X by

$$\theta_n = \frac{1}{n} \sum_{i=0}^{n-1} \delta_{G^i(x_n, v_n)}.$$

By compactness of X we can always suppose (by replacing θ_n by some subsequence if necessary) that the sequence θ_n converges weakly toward some probability measure θ. Continuity of G easily implies that θ is $G-$invariant.

Let $h : X \to \mathbb{R}$ be the map defined by $h(x, v) = \log(||Df(x)v||)$. By the chain rule we get

$$\sum_{i=0}^{n-1} h(G^i(x, v)) = \log(||Df^n(x)v||).$$

Therefore

$$\int_X hd\theta = \lim_{n\to\infty} \int_X hd\theta_n = \lim_{n\to\infty} \frac{1}{n} \log(||Df^n(x_n)v_n||) = \mathcal{E}(\Phi, K).$$

Now, by the ergodic decomposition theorem (see Mañé, 1987 chapter 6 Theorem 6.4)

$$\int_X hd\theta = \int_X (\int_X hd\theta_{(x,v)})d\theta$$

where $\theta_{(x,v)}$ are ergodic G-invariant probability measures. It follows that for each $\epsilon > 0$ there exists an ergodic G-invariant measure $\nu = \theta_{(x,v)}$ for some (x, v), such that $\int_X hd\nu \leq \mathcal{E}(\Phi, K) + \epsilon$. Birkhoff's ergodic theorem implies the existence of a Borel set $X' \subset X$ such that $\nu(X') = 1$ and

$$\lim_{n\to\infty} \frac{1}{n} \log(||Df^n(x)v||) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} h(G^i(x, v)) = \int_X hd\nu \leq \mathcal{E}(\Phi, K)+\epsilon \quad (31)$$

for all $(x, v) \in X'$. Let μ be the marginal probability measure defined on K by

$$\mu(B) = \nu\{(x, v) \in X : x \in B\}.$$

Clearly μ is f-invariant and ergodic. On the other hand by Oseledec's Theorem for μ almost all $x \in K$ and all $v \in T_xM$ the limit $\lim_{n\to\infty} \frac{1}{n} \log(||Df^n(x)v||)$ exists and satisfies

$$\lambda_1(\mu) \leq \lim_{n\to\infty} \frac{1}{n} \log(||Df^n(x)v||) \quad (32)$$

Therefore from (31) and (32) we deduce

$$\inf_{\mu\, ergodic} \lambda_1(\mu) \leq \mathcal{E}(\Phi, K).$$

To prove the converse inequality let μ be any ergodic measure. By Oseledec's Theorem for μ almost all $x \in K$ and all $v \in E_1(x)$:

$$\lambda_1(\mu) = \lim_{n\to\infty} \frac{1}{n} \log(||Df^n(x)v||) \geq \mathcal{E}(\Phi, K)$$

QED

Corollary 8.13

$$\mathcal{E}(\Phi, K) = \mathcal{E}(\Phi, MC(\Phi|K)) = \mathcal{E}(\Phi, BC(\Phi|K)).$$

Proof The first equality follows from Theorem 8.12 and the second from Poincaré recurrence theorem (equation 30). **QED**

Theorem 8.12 and Corollary 8.13 can be used to estimate the Expansion rate. Here are two such estimates.

Corollary 8.14 *Assume that $M = \mathbb{R}^m$, Φ is generated by a smooth vector field F and $MC(\Phi|K) \subset Eq(\Phi)$ (the equilibria set). Then*

$$\mathcal{E}(\Phi, K) = \inf\{\lambda_1(p) : p \in Eq(\Phi) \cap K\}$$

where $\lambda_1(p)$ denote the smallest real part of the eigenvalues of the jacobian matrix $DF(p)$.

Proof Under the assumption that $MC(\Phi|K) \subset Eq(\Phi)$ every ergodic measure with support in K has to be a Dirac measure at an equilibrium point. Let $\mu = \delta_p$ be such a measure. Then $\lambda_1(\mu) = \lambda_1(p)$ and the result follows from Theorem 8.12. **QED**

The following result is based on Hirsch (1994).

Corollary 8.15 *Assume $M = \mathbb{R}^m$ and Φ is generated by a smooth vector field F. Let $\beta(x)$ equals the smallest eigenvalue of the matrix $\frac{1}{2}(DF(x) + DF(x)^T)$ where T denotes matrix transpose. Then*

$$\mathcal{E}(\Phi, K) \geq \inf_{x \in MC(\Phi|K)} \beta(x).$$

Proof Let $y \in MC(\Phi|K)$. By invariance of $MC(\Phi|K)$ $\Phi_t(y) \in MC(\Phi|K)$ for all $t \in \mathbb{R}$. The variational equation along orbits of the reversed time flow Φ_{-t} gives

$$\frac{d}{dt} D\Phi_{-t}(y) = -DF(\Phi_{-t}(y)) D\Phi_{-t}(y)$$

Therefore for every nonzero vector $v \in \mathbb{R}^m$ and $t \geq 0$ we have,

$$\frac{d}{dt}||D\Phi_{-t}(y)v||^2 = -2\langle D\Phi_{-t}(y)v, DF(\Phi_{-t}(y))D\Phi_{-t}(y)v\rangle \leq -2\beta||D\Phi_{-t}(y)v||^2$$

with $\beta = \inf_{x \in MC(\Phi|K)} \beta(x)$. Therefore

$$||D\Phi_{-t}(y)|| \leq e^{-t\beta}$$

for all $y \in MC(\Phi|K)$ and $t \geq 0$. To conclude set $y = \Phi_t(x)$ for $x \in MC(\Phi|K)$ and the estimate follows from the definition of $\mathcal{E}(\Phi, K)$ combined with Corollary 8.13. **QED**

Example 8.16 Let $(x_n, y_n) \in \mathbb{R}^2$ be the Robbins Monro algorithm described in Example 8.1. It is convenient here to express the dynamics of the vector field (26) in polar coordinates. That is

$$\frac{d\rho}{dt} = \rho h(\rho^2),\ \frac{d\theta}{dt} = (\rho \sin\theta)^2.$$

Let $B_\epsilon = \{|1-\rho| \leq \epsilon\}$. For $\epsilon << 1$ and $(\rho, \theta) \in B_\epsilon$

$$\frac{d(1-\rho)^2}{dt} = -2(1-\rho)\rho(1-\rho^2) \leq -2(1-\rho)^2(2-\epsilon)(1-\epsilon).$$

Thus $|1-\rho(t)| \leq e^{-(2-\epsilon)(1-\epsilon)t}|1-\rho|$. This shows that S^1 attract B_ϵ at rate $-2+O(\epsilon)$.

To compute $\mathcal{E}(\Phi|S^1, S^1)$ we use Corollary 8.14. The dynamics on S^1 being given by $\frac{d\theta}{dt} = sin^2(\theta)$, 0 is the eigenvalue of the linearized ODE at equilibria points 0 and π. Thus $\mathcal{E}(\Phi|S^1, S^1) = 0$. Suppose now that

$$l(\gamma) = \limsup_{n\to\infty} \frac{\log(\gamma_n)}{\tau_n} < 0.$$

Then Corollary 8.10 and Proposition 8.3 imply that $\{x_n, y_n\}$ converges almost surely toward one of the points a or c of Figure 2.

9 Nonconvergence to Unstable points, Periodic orbits and Normally Hyperbolic Sets

Let $\{x_n\}$ be given by (7) and F a smooth vector field. Let $p \in \mathbb{R}^m$ 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.*

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^1 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 (7), 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 recently by Pemantle (1990) and Brandière and Duflo (1996) (under different sets of assumptions), provided the vector fields F is C^2 and the gain sequence is well behaved.

The ideas of Pemantle have been used in Benaïm and Hirsch (1995b) to tackle the case of hyperbolic unstable periodic orbits. This section is an extension of these results which covers a larger class of repelling sets. Recent works by Brandière (1996,1997) address similar questions and prove the nonconvergence of stochastic approximation processes toward certain types of repelling sets which are not considered here.

Throughout this section we assume given

- A smooth vector field $F : \mathbb{R}^m \to \mathbb{R}^m$ generating a flow $\Phi = \{\Phi_t\}$.
- A smooth $(m-d)$ dimensional (embedded) submanifold $S \subset \mathbb{R}^m$, where $d \in \{1, \ldots, m\}$.
- A nonempty compact set $\Gamma \subset S$ invariant under Φ.

We assume that S is *locally invariant*, meaning that there exists a neighborhood U of Γ (in $\mathbb{R}^m$) and a positive time t_0 such that

$$\Phi_t(U \cap S) \subset S$$

for all $|t| \leq t_0$.

We further assume that for every point $p \in \Gamma$,

$$\mathbb{R}^m = T_pS \oplus E_p^u$$

where

(i) $p \to E_p^u$ is a continuous map from Γ into the Grassman manifold $G(d, m)$ of d planes in $\mathbb{R}^m$.

(ii) $D\Phi_t(p)E_p^u = E_{\Phi_t(p)}^u$ for all $t \in \mathbb{R}, p \in E_p^u$.

(iii) There exist $\lambda > 0$ and $C > 0$ such that for all $p \in \Gamma$, $w \in E_p^u$ and $t \geq 0$

$$||D\Phi_t(p)w|| \geq Ce^{\lambda t}||w||.$$

Examples

Linearly Unstable Equilibria: Suppose $\Gamma = \{p\}$ where $p \in \mathbb{R}^m$ is a linearly unstable equilibrium of F. Then $\mathbb{R}^m = E_p^s \oplus E_p^c \oplus E_p^u$ where E_p^s, E_p^c and E_p^u are the generalized eigenspaces of $DF(p)$ corresponding to eigenvalues with real parts < 0, equal to 0 and > 0. Because p is linearly unstable, the dimension of E_p^u is at least 1.

Using stable manifold theory (see e.g Shub, (1987) or Robinson, (1995)) there exists a locally invariant manifold S tangent to $E_p^s \oplus E_p^c$ - the *center stable manifold* of p - which is C^k when F is C^k.

Since $D\Phi_t(p) = e^{tDF(p)}$ there exist $\lambda > 0$ and $C > 0$ such that $||D\Phi_t(p)w|| \geq Ce^{\lambda t}||w||$ for all $w \in E_p^u$.

Linearly Unstable Periodic Orbits: Let $\Gamma \subset \mathbb{R}^m$ be a periodic orbit. Γ is said to be hyperbolic if the unity is a multiplier with multiplicity one and the $m-1$ other multipliers have moduli different from 1. Γ is said to be linearly unstable if some multiplier has modulus strictly greater than 1.

Suppose Γ is a hyperbolic linearly unstable periodic orbit for the C^k vector field F. By hyperbolicity (see for example Shub 1987) there exist positive constants C, λ 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 $p \in \Gamma$, $t \geq 0$:

$$\| D\Phi_t(p)|_{E_p^s} \| \leq Ce^{-\lambda t}, \tag{33}$$

$$\| D\Phi_{-t}(p)|_{E_p^u} \| \leq Ce^{-\lambda t},$$

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

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

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

Each E_p^u is a linear subspace of $\mathbb{R}^m$. The map $p \mapsto E_p^u$ is a continuous map from Γ into the Grassmann manifold of linear subspaces of the appropriate dimension (it is actually C^k due to the fact that $T_\Gamma\Phi_t$ maps $p \times E_p^u$ to $\Phi_t(p) \times E^u_{\Phi_t(p)}$, and that $T_\Gamma\Phi$ is a C^k flow).

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

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

Using stable manifold theory, we take ϵ small enough so that $W_\epsilon^s(p)$ is a C^k (embedded) submanifold of $\mathbb{R}^m$, whose tangent space at p is $T_pW_\epsilon^s(p) = E_p^s$.

We set $S = \bigcup_{p\in\Gamma} W_\epsilon^s(p)$. This is the *local stable manifold of* Γ. It follows from the above that S is a C^k locally invariant submanifold, with $T_pS = E_p^s \oplus E_p^\Phi$.

Let $0 < \alpha \leq 1$. We call a map $C^{1+\alpha}$ if it is C^1 and its derivative is αHőlder. Note that C^{1+1} is weaker than C^2. A $C^{1+\alpha}$ manifold is a manifold whose transition functions may be chosen $C^{1+\alpha}$.

Theorem 9.1 *Let* $\{x_n\}$ *given by (7) be a Robbins Monro algorithm (section 4.2) Assume:*

(i) *There exists* $K > 0$ *such that* $||U_n|| \leq K$ *for all* $n \geq 0$.

(ii) $\{\gamma_n\}$ *is as in Proposition (4.4).*

(iii) *There exists a neighborhood* $\mathcal{N}(\Gamma)$ *of* Γ *and* $b > 0$ *such that for all unit vector* $v \in \mathbb{R}^m$

$$\mathbf{E}(\langle U_{n+1}, v\rangle^+ | \mathcal{F}_n) \geq b 1_{\{x_n \in \mathcal{N}(\Gamma)\}}.$$

(iv) *There exists* $0 < \alpha \leq 1$ *such that:*

(a) F *and* S *are* $C^{1+\alpha}$,

(b)

$$\lim_{n\to\infty} \frac{\gamma_{n+1}^\alpha}{\sqrt{\lim_{m\to\infty}\sum_{i=n+1}^m \gamma_i^2}} = 0.$$

Then

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

Remark 9.2 If $\frac{A}{n^\beta} \leq \gamma_n \leq \frac{B}{n^\beta}$ with $\beta > 1/2$, $0 < A \leq B$ then condition $(iv), (b)$ of Theorem 9.1 is fulfilled provided that

$$\alpha > \frac{2\beta - 1}{2}.$$

If $\sum_n \gamma_n^2 = \infty$ condition (iv) of Theorem 9.1 is always satisfied for $\alpha > 0$.

9.1 Proof of Theorem 9.1

The proof of this result relies, on one hand, upon the construction of a suitable Lyapounov function, and on the other hand on probabilistic estimates due to Pemantle (1990).

Construction of a suitable Lyapounov function

The construction given here is very similar to the construction given in Benaïm and Hirsch (1995b), but instead of defining the Lyapounov function as the distance to S in the unstable direction for an adapted Riemann metric obtained by time averaging (as in Benaïm and Hirsch (1995b)), we consider the usual distance and we then define the Lyapounov function by averaging over time.

It appears that this construction leads to much easier estimates and allows us to handle the fact that the splitting $T_\Gamma \mathbb{R}^m = T_\Gamma S \oplus E^u$ is only continuous.

step1 The first step of the construction is to replace the continuous invariant splitting $T_\Gamma \mathbb{R}^m = T_\Gamma S \oplus E^u$ by a smooth (noninvariant) splitting $T_\Gamma \mathbb{R}^m = T_\Gamma S \oplus \tilde{E}^u$ close enough to the first one to control the expansion of $D\Phi_t$ along fibers of $\tilde{E}^u$.

Choose $T > 0$ large enough so that for $f = \Phi_T$, $p \in \Gamma$, and $w \in E^u_p$:

$$||Df(p)w|| \geq 5||w||.$$

By Whitney embedding theorem (Hirsch, 1976, chapter 1) we can embed $G(d,m)$ into $\mathbb{R}^D$ for some $D \in \mathbb{N}$ large enough so that we can see $p \to E^u_p$ as a map from $\Gamma \to \mathbb{R}^D$. Thus by Tietze extension theorem (Munkres, 1975) we can extend this map to a continuous map from $\mathbb{R}^m$ into $\mathbb{R}^D$. Let ϱ denote a C^∞ retraction from a neighborhood of $G(d,m) \subset \mathbb{R}^D$ onto $G(d,m)$ whose existence follows from a classical result in differential topology (Hirsch, 1976, chapter 4). By composing the extension of $p \to E^u_p$ with ϱ we obtain a continuous map defined on a neighborhood N of Γ, taking values in $G(d,m)$ and which extends $p \to E^u_p$. To shorten notation, we keep the notation $p \in N \to E^u_p \in G(d,m)$ to denote this new map.

By replacing N by a smaller neighborhood if necessary, we can further assume that N is compact, $N \subset U$, $f(N) \subset U$ and

$$||Df(p)w|| \geq 4||w||$$

for all $p \in N$, and $w \in E^u_p$.

Now, by a standard approximation procedure, we can approximate $p \in N \to E^u_p \in G(d,m)$ by a C^∞ function from N into $\mathbb{R}^D$. Then, by composing with ϱ, we obtain a C^∞ map $p \in N \to \tilde{E}^u_p \in G(d,m)$ which can be chosen arbitrary close to $p \in N \to E^u_p$ in the C^0 topology.

For $p \in N$, let

$$P_p : T_pS \oplus E^u_p \to T_pS,$$
$$u + v \to u$$

and let

$$\tilde{P}_p : T_pS \oplus \tilde{E}^u_p \to T_pS,$$
$$u + v \to u.$$

Fix $\epsilon > 0$ small enough so for all $p \in N$, $\epsilon(||Df(p)||+4) < 1$ and $\epsilon||Df(p)||(||P_p||+||\tilde{P}_p|| + 1)) < 1$ (this choice will be clarified in the next lemma). From now on, we will assume that the map $p \in N \to \tilde{E^u}_p \in G(d,m)$ is chosen such that for all $p \in N \cap S$:

(i) $\mathbb{R}^m = T_pS \oplus \tilde{E^u}_p$

(ii) The projector $\tilde{P}_p : T_pS \oplus \tilde{E^u}_p \to T_pS$, satisfies

$$||P_p - \tilde{P}_p|| \leq \epsilon.$$

Let $\tilde{E}^u = \{(p,v) \in S \cap N \times \mathbb{R}^m : v \in \tilde{E}^u_p\}$. Since S is $C^{1+\alpha}$, $\tilde{E}^u$ is a $C^{1+\alpha}$ vector bundle over $S \cap N$. Let $H : \tilde{E}^u \to \mathbb{R}^m$ be the map defined by $H(p,v) = p + v$. It is easy to see that the tangent map of H at a point $(p,0)$ is invertible. The inverse function theorem implies that H is a local $C^{1+\alpha}$ diffeomorphism at each point of the zero section of $\tilde{E}$. Since H maps the zero section to $S \cap N$ by the diffeomorphism $(p,0) \to p$, it follows that H restricts to a $C^{1+\alpha}$ diffeomorphism $H : N_0' \to N_0$ between open neighborhoods N_0' or the zero section and $N_0 \subset N$ of $S \cap N$. We now define the maps

$$\Pi : N_0 \to S,$$
$$x \to \Pi(x) = p$$

for $H^{-1}(x) = (p,v)$ and

$$V : N_0 \to \mathbb{R}_+,$$
$$x \to ||x - \Pi(x)||.$$

Observe that for any $p \in N_0 \cap S$

$$D\Pi(p) = \tilde{P}_p.$$

Step 2 The second step consists in the construction of Lyapounov function which is zero on S and increases exponentially along trajectories outside S. This function (see Proposition 9.5) is obtained from V by some averaging procedure.

Lemma 9.3 *There exists a neighborhood of* Γ, $N_1 \subset N_0$, *and* $\rho > 1$ *such that for all* $x \in N_1$

$$V(f(x)) \geq \rho V(x).$$

To prove this Lemma we use the following estimates

Lemma 9.4 *Let* $P, \tilde{P} : \mathbb{R}^m \to \mathbb{R}^m$ *be two projectors and* $A : \mathbb{R}^m \to \mathbb{R}^m$ *a linear map. Assume there exist* $\epsilon, \alpha > 0$ *such that*

(i) $||P - \tilde{P}|| \leq \epsilon$,

(ii) $||Au|| \geq \alpha||u||$ *for all* $u \in KerP$.

Then

(i) $||Av|| \geq (\alpha(1-\epsilon) - \epsilon||A||)||v||$ *for all* $v \in Ker\tilde{P}$.

(ii) $||\tilde{P}A(Id - \tilde{P}) - PA(Id - P)|| \leq \epsilon||A||(1 + ||P|| + ||\tilde{P}||)$.

Proof Let $v \in Ker\tilde{P}$ be a unit vector. Write $v = (v - Pv) + (P - \tilde{P})v$. Thus

$$||Av|| \geq \alpha||v - Pv|| - \epsilon||A|| \geq \alpha(1-\epsilon) - \epsilon||A||.$$

This proves (i) while (ii) follows from

$||\tilde{P}A(Id-\tilde{P})-PA(Id-P)|| = ||(P-\tilde{P})A(Id-\tilde{P})-PA(P-\tilde{P})||$
$\leq \epsilon||A||(||Id-\tilde{P}||+||P||)$.

QED

Proof of Lemma 9.3
let $x \in N_0 \cap f^{-1}(N_0)$ and set $p = \Pi(x)$.

$V(f(x)) \geq ||f(x)-f(p)||-||f(p)-\Pi(f(x))|| = ||Df(p)(x-p)||$
$-||D\Pi(f(p))Df(p)(x-p)||+o(||x-p||)$.

Lemma 9.4 (i) applied to $Df(p), P_p, \tilde{P}_p$ and our choice for ϵ imply

$$||Df(p)(x-p)|| \geq 3||x-p|| = 3V(x)$$

Also, by Lemma 9.4 (ii)

$||\tilde{P}_{f(p)}Df(p)(Id-\tilde{P}_p)-P_{f(p)}Df(p)(Id-P_p)|| \leq \epsilon||Df(p)||(1+||P_p||+||\tilde{P}_{f(p)}||) < 1.$

Since for $p \in \Gamma$, $P_{f(p)}Df(p)(Id - P_p) = 0$ the preceding inequality implies the existence of a neighborhood of Γ, $N_1 \subset N_0$ such that for $x \in N_1$ and $p = \Pi(x)$: $||\tilde{P}_{f(p)}Df(p)(Id - \tilde{P}_p)|| < 1$. This implies

$$||D\Pi(f(p))Df(p)(x-p)|| < ||x-p|| = V(x).$$

It follows that $V(f(x)) \geq 3V(x) - V(x) + o(V(x))$. Replacing N_1 by a smaller neighborhood gives

$$V(f(x)) \geq \rho V(x)$$

for some $\rho > 1$. **QED**

Recall that a map $\eta : \mathbb{R}^m \to \mathbb{R}$ is said to have a right derivative at point x if for all $h \in \mathbb{R}^m$ the limit

$$D\eta(x).h = \lim_{t\to 0, t>0} \frac{\eta(x+th) - \eta(x)}{t}$$

exists. If η is differentiable at x, then $D\eta(x).h = \langle\nabla\eta(x), h\rangle$ where $\nabla\eta(x) \in \mathbb{R}^m$ is the usual gradient.

Proposition 9.5 *There exists a compact neighborhood of* Γ, $\mathcal{N}(\Gamma) \subset N_1$ *and real numbers* $l > 0, \beta > 0$ *such that the map* $\eta : \mathcal{N}(\Gamma) \to \mathbb{R}$ *given by*

$$\eta(x) = \int_0^l V(\Phi_{-t}(x))dt$$

enjoys the following properties:

(i) η *is* C^r *on* $\mathcal{N}(\Gamma) \setminus S$.

(ii) *For all* $x \in \mathcal{N}(\Gamma) \cap S$, η *admits a right derivative* $D\eta(x) : \mathbb{R}^m \to \mathbb{R}^m$ *which is Lipschitz, convex and positively homogeneous.*

(iii) *If* $r \geq 1 + \alpha$ *for some* $0 < \alpha \leq 1$ *there exists* $k > 0$ *and a neighborhood* $U \subset \mathbb{R}^m$ *of* 0 *such that for all* $x \in \mathcal{N}(\Gamma)$ *and* $v \in U$

$$\eta(x+v) \geq \eta(x) + D\eta(x)v - k||v||^{1+\alpha}.$$

(iv) *There exists* $c_1 > 0$ *such that for all* $x \in \mathcal{N}(\Gamma) \setminus S$

$$||\nabla\eta(x)|| \geq c_1$$

and for all $x \in \mathcal{N}(\Gamma) \cap S$ *and* $v \in \mathbb{R}^m$

$$D\eta(x)v \geq c_1||v - D\Pi(x)v||.$$

(v) *For all* $x \in \mathcal{N}(\Gamma) \cap S$, $u \in T_xS$ *and* $v \in \mathbb{R}^m$

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

(vi) *For all* $x \in \mathcal{N}(\Gamma)$

$$D\eta(x).F(x) \geq \beta\eta(x),$$

Proof *Notation:* Given $A > 0$ we let $N^A \subset N_1$ denote a compact neighborhood of Γ such that for all $|t| \leq A$, $\Phi_t(N^A) \subset N_1$ and we let $C(A) > 0$ denote the Lipschitz constant of the map $t, x \to V(\Phi_t(x))$ restricted to $[-A, A] \times N^A$. Remark that for $|t| \leq A$ and $x \in N^A$ we have:

$$V(\Phi_t(x)) = |V(\Phi_t(x)) - V(\Phi_t(\Pi(x))| \leq C(A)V(x). \tag{34}$$

We first fix $l > 2T$ and assume that $\mathcal{N}(\Gamma) \subset N^l$. We will see below (in proving (vi)) how to choose l.

(i) is obvious.

(ii) follows from the fact that Π is C^1, and $x \to ||x||$ admits a right derivative at the origin of $\mathbb{R}^m$ given as $h \to ||h||$.

Before passing to the proof of (iii) let us compute $D\eta(x)$. For $x \in N^l$ let

$$G_t(x) = \Phi_{-t}(x) - \Pi(\Phi_{-t}(x)),$$

$$B(t,x) = DG_t(x) = [Id - D\Pi(\Phi_{-t}(x))]D\Phi_{-t}(x),$$

and for $x \in N^l \setminus S$ let

$$b(x) = \frac{x - \Pi(x)}{||x - \Pi(x)||}.$$

It is easy to verify that

$$D\eta(x).h = \int_0^l \langle B(t,x)h, b(\Phi_{-t}(x))\rangle dt \tag{35}$$

for $x \in N^l \setminus S$ and

$$D\eta(x).h = \int_0^l ||B(t,x)h||dt \tag{36}$$

for $x \in N^l \cap S$.

(*iii*) If $r \geq 1 + \alpha$, Φ and Π are C^1 with α Hölder derivatives. Hence there exits $k > 0$ such that

$$||G_t(x+u)|| - ||G_t(x)|| \geq ||G_t(x) + B(t,x)u|| - ||G_t(x)|| - k||u||^{1+\alpha}.$$

If $x \in N^l \cap S$ then $||G_t(x)|| = 0$ and the result follows from (36).

If $x \in N^l \setminus S$, convexity of the norm implies

$$||G_t(x) + B(t,x)u|| - ||G_t(x)|| \geq \frac{\langle G_t(x), B(t,x)u\rangle}{||G_t(x)||} = \langle B(t,x)u, b(\Phi_{-t}(x))\rangle$$

and the result follows from (35).

(*iv*). *Claim:* There exists $c_0 > 0$ such that $||B(t,p)b|| \geq c_0$ for all $0 \leq t \leq l$, $p \in N^l \cap S$ and unit vector $b \in \tilde{E}^u_p$.

Proof of the claim: Suppose the contrary. Then by compactness of $\{(t,p,v) : 0 \leq t \leq l, p \in N^l \cap S, b \in \tilde{E}^u_p, ||b|| = 1\}$ there exists $0 \leq t \leq l$, $p \in N^l \cap S$ and a unit vector $b \in \tilde{E}^u_p$ such that $B(t,p)b = 0$. Therefore $D\Phi_{-t}(p)b \in Ker(Id - D\Pi(\Phi_{-t}(p)) = T_{\Phi_{-t}(p)}S$. Thus $b \in D\Phi_{-t}(p)^{-1}T_{\Phi_{-t}(p)}S = T_pS$. But this is impossible because b is a unit vector in $\tilde{E}^u_p$ and $\mathbb{R}^m = T_pS \oplus \tilde{E}^u_p$. This proves the claim.

For $x \in N^l \setminus S$

$$\Phi_{-t}(x) - \Pi(\Phi_{-t}(x)) = B(t,\Pi(x))(x - \Pi(x)) + o(||x - \Pi(x)||)$$

and for any $h \in \mathbb{R}^m$ with $||h|| = 1$,

$$\langle B(t,x)h, \Phi_{-t}(x) - \Pi(\Phi_{-t}(x))\rangle = \langle B(t,\Pi(x))h, B(t,\Pi(x))(x - \Pi(x))\rangle + o(||x - \Pi(x)||).$$

Thus, if we set $h = b(x)$ we get

$$\langle B(t,x)h, b(\Phi_{-t}(x))\rangle = \frac{||B(t,\Pi(x)).b(x)||^2 + \epsilon(||x - \Pi(x)||)}{||B(t,\Pi(x)).b(x)|| + \epsilon_1(||x - \Pi(x)||)}$$

where $\lim_{u\to 0}\epsilon(u) = \lim_{u\to 0}\epsilon_1(u) = 0$. Since, according to the claim,
$||B(t,\Pi(x)).b(x)|| >$
c_0 this implies

$$\langle B(t,x)h, b(\Phi_{-t}(x))\rangle \geq c_0/2$$

for $||x-\Pi(x)||$ small enough. Formulae (35) implies that $D\eta(x)b(x) \geq c_1$ with $c_1 = lc_0/2$. Hence $||\nabla\eta(x)|| \geq c_1$.

Suppose now $x \in \mathcal{N}(\Gamma)\cap S$. It follows from *(iii)* that $B(t,x)v = B(t,x)(v - D\Pi(p)v)$ for all $v \in \mathbb{R}^m$. Now the claim together with (36) imply that $D\eta(x)v \geq c_1||v - D\Pi(x)v||$.

(vi) For $x \in N^l$ and $0 \leq t \leq l$ we can write $t = kT + r$ for $k \in \mathbb{N}$ and $0 \leq r < T$. Thus, by Lemma 9.3 and equation (34),

$$V(\Phi_t(x)) = V(f^k(\Phi_r(x)) \geq \rho^k V(\Phi_r(x)) \geq \frac{\rho^k}{C(T)}V(x) \geq C_1(T)e^{at}V(x)$$

where $C_1(T) = \dfrac{1}{\rho C(T)}$ and $a = \dfrac{\log(\rho)}{T}$.

For $s > 0$,

$$\eta(\Phi_s(x)) - \eta(x) = -\int_{l-s}^{l} V(\Phi_{-t}(x))dt + \int_{-s}^{0} V(\Phi_{-t}(x))dt \geq -V(x)\frac{e^{-al}e^{as}}{C_1(T)}s$$
$$+\int_0^s V(\Phi_{-t}(x))dt.$$

It follows that

$$\lim_{s\to 0,\, s>0} \frac{\eta(\Phi_s(x)) - \eta(x)}{s} \geq V(x)\left(1 - \frac{e^{-al}}{C_1(T)}\right)$$

It then suffices to choose l large enough so that $1 - \dfrac{e^{-al}}{C_1(T)} > 0$. With this choice of l we get that

$$D\eta(x).F(x) \geq \beta\eta(x)$$

with $\beta = \dfrac{1}{lC(l)}\left(1 - \dfrac{e^{-al}}{C_1(T)}\right) > 0$. **QED**

Probabilistic Estimates

The following lemma adapted from Pemantle (1992, Lemma 5.5) is the probabilistic key of the proof of Theorem 9.1.

Lemma 9.6 *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. Let $\{\gamma_n\}$ be a sequence of positive numbers such that $\sum_n \gamma_n^2 < \infty$ and let $\alpha_n = \sum_{i=n+1}^{\infty} \gamma_i^2$.*

Assume there exist a sequence $0 \leq \epsilon_n = o(\sqrt{\alpha_n})$, constants $a_1 > 0, a_2 > 0$ and an integer N_0 such that for all $n \geq N_0$:

(i) $|X_n| = o(\sqrt{\alpha_n})$.

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

(iii) $\mathsf{E}(S_{n+1}^2 - S_n^2|\mathcal{F}_n) \geq a_1 \gamma_{n+1}^2$.

(iv) $E(X_{n+1}^2|\mathcal{F}_n) \leq a_2 \gamma_{n+1}^2$.

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

This lemma is stated and proved in (Pemantle, 1992) for $\gamma_n = 1/n$ and but the proof adapts without difficulty to the present situation.

Proof Assume without loss of generality that $N_0 = 0$, $|X_n| \leq \sqrt{b_1 \alpha_n}$ and $\epsilon_n < \frac{1}{2}\sqrt{b_2 \alpha_n}$ where

$$2(b_1 + b_2) < a_1.$$

Given $n \in \mathbb{N}$ let T be the stopping time defined as

$$T = \inf\{i \geq n \ : S_i \geq \sqrt{b_2 \alpha_i}\}$$

Claim:

$$P(T < \infty|\mathcal{F}_n) \geq 1 - \frac{2(b_1 + b_2)}{a_1} \tag{37}$$

Proof of (37): By assumption (iii), the process $Z_k = S_k^2 - a_1 \sum_{i=0}^{k} \gamma_i^2$ is a submartingale. Therefore $\{Z_{k\wedge T}\}_{k\geq n}$ is a submartingale and for all $m \geq n$ $E(Z_{m\wedge T} - Z_n|\mathcal{F}_n) \geq 0$. Hence

$$E(S_{m\wedge T}^2 - S_n^2|\mathcal{F}_n) \geq a_1 E(\sum_{i=n+1}^{m\wedge T} \gamma_i^2|\mathcal{F}_n) \geq a_1(\sum_{i=n+1}^{m} \gamma_i^2) P(T > m|\mathcal{F}_n).$$

On the other hand

$$S_{m\wedge T}^2 - S_n^2 \leq 2(b_1 + b_2)\alpha_n$$

by definition of T and condition (i). It follows that

$$P(T > m|\mathcal{F}_n) \leq \frac{2(b_1 + b_2)\alpha_n}{a_1 \sum_{i=n+1}^{m} \gamma_i^2}.$$

Letting $m \to \infty$ proves the claim.

Now, let σ be the stopping time defined as

$$\sigma = \inf\{i \geq n \ : S_i < \frac{1}{2}\sqrt{b_2 \alpha_n}\}.$$

Claim: Let E_n be the event $E_n = \{S_n \geq \sqrt{b_2 \alpha_n}\}$. Then

$$P(\sigma = \infty|\mathcal{F}_n) 1_{E_n} \geq (\frac{b_2}{4a_2 + b_2}) 1_{E_n} \tag{38}$$

Proof of (38): The process $\{S_{i\wedge\sigma}\}_{i\geq n}$ is a submartingale. Indeed

$$E(S_{i+1\wedge\sigma} - S_{i\wedge\sigma}|\mathcal{F}_i) = 1_{\{\sigma > i\}} E(S_{i+1} - S_i|\mathcal{F}_i) \geq 1_{\{\sigma \geq i\}} 1_{\{S_i \geq \frac{1}{2}\sqrt{b_2 \alpha_i}\}} E(X_{i+1}|\mathcal{F}_i)$$

where the last term is nonnegative by condition (ii). Therefore by Doob's decomposition Lemma there exist a martingale $\{M_i\}_{i\geq n}$ and a previsible process $\{I_i\}_{i\geq n}$ such that $S_{i\wedge\sigma} = M_i + I_i$, $I_n = 0$ and $I_{i+1} \geq I_i$. The fact that $S_{i\wedge\sigma} \geq M_i$ implies

$$P(\sigma = \infty|\mathcal{F}_n) \geq P(\forall i \geq n : \; M_i \geq \frac{1}{2}\sqrt{b_2\alpha_n}|\mathcal{F}_n).$$

Thus

$$P(\sigma = \infty|\mathcal{F}_n)1_{E_n} \geq P(\forall i \geq n : \; M_i - M_n \geq -\frac{1}{2}\sqrt{b_2\alpha_n}|\mathcal{F}_n)1_{E_n} \qquad (39)$$

Our next goal is to estimate the right hand term of (39). Set $M_i' = M_i - M_n$. For $i \geq n$:

$$E(M'^2_i|\mathcal{F}_n) = \sum_{j=n}^{i-1} E((M_{j+1} - M_j)^2|\mathcal{F}_n) \leq a_2\alpha_n \qquad (40)$$

where we have used the fact that

$$E((M_{j+1} - M_j)^2|\mathcal{F}_j) = E((S_{j+1} - S_j)^2|\mathcal{F}_j) - (I_{j+1} - I_j)^2 \leq a_2\gamma_{j+1}^2$$

by condition (iv). Therefore for $s > 0$, $m \geq n$ and $t > 0$

$$P(\inf_{n\leq i\leq m} M_i' < -s|\mathcal{F}_n) \leq P(\inf_{n\leq i\leq m}(M_i' - t) < -s - t|\mathcal{F}_n) \leq P(\sup_{n\leq i\leq m}|M_i' - t| \geq s + t|\mathcal{F}_n)$$

$$\leq \frac{E(M'^2_m|\mathcal{F}_n) + t^2}{(s+t)^2} \leq \frac{a_2\alpha_n + t^2}{(s+t)^2}$$

where the last two inequalities follow from Doob's inequality combined with (40). With $s = \frac{1}{2}\sqrt{b_2\alpha_n}$ and $t = \frac{a_2\alpha_n}{s}$ we get that

$$P(\inf_{n\leq i\leq m} M_i' < -s|\mathcal{F}_n) \leq \frac{4a_2}{4a_2 + b_2}.$$

Thus

$$P(\forall i \geq n : \; M_i - M_n \geq -\frac{1}{2}\sqrt{b_2\alpha_n}|\mathcal{F}_n) \geq 1 - \frac{4a_2}{4a_2 + b_2}.$$

This proves (38).

We can now finish the proof of the Lemma. Let G denote the event that $\{S_n\}$ does not converge to zero. By definition of T and inequality (38):

$$E(1_G|\mathcal{F}_i)1_{T=i} = E(1_G|\mathcal{F}_i)1_{E_i}1_{T=i} \geq \frac{b_2}{4a_2 + b_2}1_{E_i}1_{T=i} = \frac{b_2}{4a_2 + b_2}1_{T=i}$$

for all $i \geq n$. Therefore

$$E(1_G|\mathcal{F}_n) \geq \sum_{i\geq n} E(1_G 1_{T=i}|\mathcal{F}_n) = \sum_{i\geq n} E(E(1_G|\mathcal{F}_i)1_{T=i}|\mathcal{F}_n)$$

$$\geq \frac{b_2}{4a_2+b_2} P(T<\infty|\mathcal{F}_n) \geq \frac{b_2}{4a_2+b_2}(1-\frac{2(b_1+b_2)}{a_1}) > 0$$

where the last inequality follows from (37). Since $\lim_{n\to\infty} E(1_G|\mathcal{F}_n) = 1_G$ almost surely this proves that $1_G = 1$ almost surely. **QED**

If $\sum_n \gamma_n^2 = \infty$ we use the next lemma:

Lemma 9.7 *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 and $|X_n| \leq C$. Let $\{\gamma_n\}$ be such that $\sum_i \gamma_i^2 = \infty$. Assume there exists $a_1 > 0$ and some integer N_0 such that for all $n \geq N_0$ $\mathsf{E}(S_{n+1}^2 - S_n^2|\mathcal{F}_n) \geq a_1\gamma_{n+1}^2$.*

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

Proof. As already noticed $Z_n = S_n^2 - \sum_{i=0}^{n} a_1\gamma_i^2$ is a submartingale.

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

Assume $| S_N |\leq \epsilon$ and define the stopping time $T = \inf\{k \geq N; | S_k |> \epsilon\}$. The sequence $\{(Z_{n\wedge T}, \mathcal{F}_n)\}_{n\geq N}$ is a submartingale and we have $Z_{n\wedge T} \leq (\epsilon + C)^2$. It follows from the submartingale convergence theorem that $\{Z_{n\wedge T}\}_{n\geq N}$ converges almost surely. Thus $\{\sum_{i=0}^{n\wedge T} a_1\gamma_i^2\}_{n\geq N}$ is almost surely bounded. This implies $T < \infty$ almost surely. **QED**

We now prove Theorem 9.1.

Let $N \in \mathbb{N}$. Assume $x_N \in \mathcal{N}(\Gamma)$ where $\mathcal{N}(\Gamma)$ is the neighborhood given by Proposition 9.5. Let T be the stopping time defined by

$$T = \inf\{k \geq N;\ x_n \notin \mathcal{N}(\Gamma)\}.$$

We prove Theorem 9.1 by showing that $\mathsf{P}(T < \infty) = 1$.

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:

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

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

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 suppose that $\sum \gamma_i^2 < \infty$ and verify that hypotheses (i) to (iv) of Lemma 9.6 are satisfied.

Conditions (i) and (iv). By Lipschitz continuity of η, and the boundedness of the sequences $\{F(x_n)\}$, $\{U_n\}$ we have $|X_{n+1}| = O(\gamma_{n+1}) = o(\sqrt{\alpha_n})$.

Condition (ii). Let $k' = k(||F|| + K)$ where k is given by Proposition 9.5, (iii), $||F|| = \sup\{F(x); x \in \mathcal{N}(\Gamma)\}$ and K is the uniform bound of the U_n. If $n \leq T$, using Proposition 9.5, $(ii), (iii), (v)$ and (vi) we have

$$\eta(x_{n+1}) - \eta(x_n) \geq \gamma_{n+1}\beta\eta(x_n) + \gamma_{n+1}D\eta(x_n)U_{n+1} - k'\gamma_{n+1}^{1+\alpha}. \qquad (41)$$

Thus

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

By convexity of the right derivative of η (Proposition 9.5, (ii)) and the conditional Jensen inequality we have

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

Thus

$$\mathbf{1}_{\{n\leq T\}}\mathsf{E}(X_{n+1}|\mathcal{F}_n) \geq \mathbf{1}_{\{n\leq T\}}(\gamma_{n+1}\beta\eta(x_n) - k'\gamma_{n+1}^{1+\alpha}) \qquad (42)$$

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

$$\mathbf{1}_{\{n>T\}}\mathsf{E}(X_{n+1}|\mathcal{F}_n) \geq \mathbf{1}_{\{n>T\}}\gamma_{n+1} \geq 0 \qquad (43)$$

Putting (42) and (43) together and letting $\epsilon_n = \frac{k'}{\beta}\gamma_{n+1}^{\alpha}$ proves condition (ii) of Lemma 9.6.

For Condition (iii) of Lemma 9.6, we observe that

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

If $S_n \geq \epsilon_n$, the right hand term is nonnegative by condition (ii), previously proved. If $S_n < \epsilon_n$, (42) and (43) imply $S_n\mathsf{E}(X_{n+1}|\mathcal{F}_n) \geq -\epsilon_n k'\gamma_{n+1}^{1+\alpha} = -O(\gamma_{n+1}^{1+2\alpha})$. Thus

$$\mathsf{E}(S_{n+1}^2 - S_n^2|\mathcal{F}_n) \geq \mathsf{E}(X_{n+1}^2|\mathcal{F}_n) - O(\gamma_{n+1}^{1+2\alpha}).$$

Therefore, to prove condition (iii) of Lemma 9.6, it suffices to show that

$$\mathsf{E}(X_{n+1}^2|\mathcal{F}_n) \geq b_1\gamma_{n+1}^2$$

for some $b_1 > 0$ and n large enough. From (41) we deduce

$$\mathbf{1}_{\{n\leq T\}}\left(\mathsf{E}(X_{n+1}^+|\mathcal{F}_n) - \left[\gamma_{n+1}\mathsf{E}((D\eta(x_n)U_{n+1})^+|\mathcal{F}_n) - k'\gamma_{n+1}^{1+\alpha}\right]\right) \geq 0 \qquad (44)$$

Using Proposition 9.5, (iv) and assumption (iii) of Theorem 9.1 we see that

$$\mathbf{1}_{\{n\leq T\}\cap\{x_n\notin S\}}(\mathsf{E}((D\eta(x_n)U_{n+1})^+|\mathcal{F}_n) - c_1 b) \geq 0. \qquad (45)$$

If $x_n \in S$, choose a unit vector $v_n \in \mathrm{Ker}(\mathrm{Id} - D\Pi(x_n))^{\perp}$. We have

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

Let $\mathcal{A}$ denotes the event $\mathcal{A} = \{n \leq T\} \cap \{x_n \in S\}$. By using Proposition 9.5 (*iv*), the Cauchy-Schwartz inequality and assumption (*iii*) of Theorem 9.1 we obtain

$$\begin{aligned}
\mathbf{1}_{\mathcal{A}}\mathsf{E}((D\eta(x_n)U_{n+1})^+|\mathcal{F}_n) &\geq \\
c_1\mathbf{1}_{\mathcal{A}}\mathsf{E}(||U_{n+1} - D\Pi(x_n)U_{n+1}||\,|\mathcal{F}_n) &\geq \\
c_1\mathbf{1}_{\mathcal{A}}\mathsf{E}(\langle U_{n+1} - D\Pi(x_n)U_{n+1}, v_n\rangle^+|\mathcal{F}_n) &= \\
c_1\mathbf{1}_{\mathcal{A}}\mathsf{E}(\langle U_{n+1}, v_n\rangle^+|\mathcal{F}_n) &\geq c_1 b\mathbf{1}_{\mathcal{A}}.
\end{aligned} \tag{46}$$

Putting (44), (45), (46) together and (43) give

$$\mathsf{E}(X_{n+1}^+|\mathcal{F}_n) \geq \gamma_{n+1}c_1 b - k'\gamma_{n+1}^{1+\alpha}$$

On the other hand $\mathsf{E}(X_{n+1}^2|\mathcal{F}_n) \geq \mathsf{E}(X_{n+1}^+|\mathcal{F}_n)^2$ by the Jensen inequality. It follows that $\mathsf{E}(X_{n+1}^2|\mathcal{F}_n) \geq b_1\gamma_{n+1}^2$ for $b_1 > 0$ and n large enough, as is desired.

Condition (*i*) through (*iv*) of Lemma 9.6 being satisfied, *the probability is zero that* $\{S_n\}$ *converges to zero,* according to Lemma 9.6. If $\sum \gamma_i^2 = \infty$ the proof given here also shows that conditions of Lemma 9.7 are satisfied.

Now suppose $T = \infty$. Then $\eta(x_n) = S_n$ and $\{x_n\}$ remains in $\mathcal{N}(\Gamma)$. Therefore (by Theorem 5.7) $L(\{x_n\})$ (the limit set of $\{x_n\}$) is a nonempty compact invariant subset of $\mathcal{N}(\Gamma)$, so that for all $y \in L(\{x_n\})$ and $t \in \mathbb{R}$ $\Phi_t(y) \in \mathcal{N}(\Gamma)$. By condition (*vi*) or Proposition (9.5) this implies that $\eta(\Phi_t(y)) \geq e^{\beta t}\eta(y)$ for all $t > 0$ forcing $\eta(y)$ to be zero. Thus $L(\{x_n\}) \subset S$. This implies $S_n = \eta(x_n) \to 0$. Since $\mathsf{P}(S_n \to 0) = 0$, T is almost surely finite. **QED**

10 Weak Asymptotic Pseudotrajectories

In the previous sections we have been mainly concerned with the asymptotic behavior of stochastic approximations processes with "fast" decreasing step-sizes, typically

$$\gamma_n = o(\frac{1}{\log(n)})$$

(Proposition 4.4) or

$$\gamma_n = 0(n^{-\alpha}), \alpha \leq 1$$

(Proposition 4.2).

If the step-sizes go to zero at a slower rate we cannot expect to characterize precisely the limit sets of the process[4]. However it is always possible to describe the "ergodic" or statistical behavior of the process in term of the corresponding behavior for the associated deterministic system. This is the goal of this section which is mainly based on Benaïm and Schreiber (1997). It is worth mentioning

[4]For instance, with a step-size of the order of $1/log(n)$ it is easy to construct examples for which the process never converges even though the chain recurrent set of the ODE consists of isolated equilibria.

that Fort and Pages (1997) in a recent paper largely generalize results of this section and address several interesting questions which are not considered here.

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\{\mathcal{F}_t : t \geq 0\}$ a nondecreasing family of sub-σ-algebras. Let (M, d) be a **separable** metric space equipped with its Borel σ-algebra.

A process

$$X : \mathbb{R}_+ \times \Omega \to M$$

$$(t, \omega) \to X(t, \omega)$$

is said to be a *weak asymptotic pseudotrajectory* of the semiflow Φ if

(i) It is *progressively measurable:* $X|[0,T] \times \Omega$ is $\mathcal{B}_{[0,T]} \times \mathcal{F}_T$ measurable for all $T > 0$ where $\mathcal{B}_{[0,T]}$ denotes the Borel $\sigma-$ field over $[0,T]$.

(ii)
$$\lim_{t\to\infty} P\{\sup_{0\leq h\leq T} d(X(t+h), \Phi_h(X(t)) \geq \alpha \,|\mathcal{F}_t\} = 0$$

almost surely for each $\alpha > 0$ and $T > 0$.

Recall (see section 8.3) that $\mathcal{P}(M)$ denotes the space of Borel probability measures on M with the topology of weak convergence and $\mathcal{M}(\Phi)(\subset \mathcal{P}(M))$ denotes the set of $\Phi-$invariant measures.

Let $\mu_t(\omega)$ denote the (random) occupation measure of the process :

$$\mu_t(\omega) = \frac{1}{t}\int_0^t \delta_{X(s,\omega)} ds$$

and let $\mathcal{M}(X, \omega)$ denote weak limit points of $\{\mu_t(\omega)\}$. The set $\mathcal{M}(X, \omega)$ is a (possibly empty) subset of $\mathcal{P}(M)$. However if $\{\mu_t(\omega)\}$ is tight (for example if $t \to X(t, \omega)$ is precompact) then by the Prohorov theorem $\mathcal{M}(X, \omega)$ is a nonempty compact subset of $\mathcal{P}(\mathcal{M})$.

Theorem 10.1 *Let X be a weak asymptotic pseudotrajectory of Φ. There exists a set $\tilde{\Omega} \subset \Omega$ of full measure ($P(\tilde{\Omega}) = 1$) such that for all $\omega \in \tilde{\Omega}$*

$$\mathcal{M}(X, \omega) \subset \mathcal{M}(\Phi).$$

Proof

Let $f : M \to [0, 1]$ be a uniformly continuous function and $T > 0$. For $n \geq 1$ set

$$U_n(f,T) = \int_{(n-1)T}^{nT} f(X(x(s))ds,$$

$$M_n(f,T) = \sum_{i=1}^{n} \frac{1}{i}[U_i(f,T) - E(U_i(f,T)|\mathcal{F}_{(i-1)T})],$$

and

$$N_n(f,T) = \sum_{i=2}^{n+1} \frac{1}{i}[E(U_i(f,T)|\mathcal{F}_{(i-1)T}) - E(U_i(f,T)|\mathcal{F}_{(i-2)T})].$$

The processes $\{M_n(f,T)\}_{n\geq 1}$ and $\{N_n(f,T)\}_{n\geq 1}$ are martingales with respect to the filtration $\{\mathcal{F}_{nT} : n \geq 1\}$. Since

$$\sup_n E(M_n(f,T)^2) \leq 4T^2 \sum_i \frac{1}{i^2},$$

Doob's convergence theorem implies that $\{M_n(f,T)\}_{n\geq 1}$ converges almost surely. Hence, by Kronecker lemma,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} [U_i(f,T) - E(U_i(f,T)|\mathcal{F}_{(i-1)T})] = 0 \tag{47}$$

almost surely. Similar reasoning with $\{N_n(f,T)\}$ leads to

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=2}^{n+1} [E(U_i(f,T)|\mathcal{F}_{(i-1)T}) - E(U_i(f,T)|\mathcal{F}_{(i-2)T})] = 0 \tag{48}$$

almost surely. Since $|U_i(f,T)| \leq T$, (47) implies

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=2}^{n+1} [U_i(f,T) - E(U_i(f,T)|\mathcal{F}_{(i-1)T})] = 0 \tag{49}$$

and by adding (48) and (49) we obtain

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} [U_{i+1}(f,T) - E(U_{i+1}(f,T)|\mathcal{F}_{(i-1)T})] = 0 \tag{50}$$

almost surely.

We claim that

$$\lim_{i\to\infty} E(U_{i+1}(f,T) - U_i(f\circ\Phi_T,T)|\mathcal{F}_{(i-1)T}) = 0 \tag{51}$$

almost surely. Let $\epsilon > 0$. By uniform continuity of f there exists $\alpha > 0$ such that $d(x,y) \leq \alpha$ implies $|f(x) - f(y)| \leq \epsilon$. Hence

$$|E(U_{i+1}(f,T) - U_i(f\circ\Phi_T,T)|\mathcal{F}_{(i-1)T})| \leq E\{\int_{(i-1)T}^{iT} |f(X(s+T)) - (f\circ\Phi_T)(X(s))| ds|\mathcal{F}_{(i-1)T}\}$$

$$\leq 2TP\{\sup_{(i-1)T\leq s\leq iT} |X(s+T) - \Phi_T(X(s))| \geq \alpha|\mathcal{F}_{(i-1)T}\} + T\epsilon.$$

Since X is a weak asymptotic pseudotrajectory of Φ the first term in the right of the inequality goes to zero almost surely as $i\to\infty$ and since ϵ is arbitrary, this proves the claim.

Now, write

$$U_{i+1}(f,T) - U_i(f\circ\Phi_T) = [U_{i+1}(f,T) - E(U_{i+1}(f,T)|\mathcal{F}_{(i-1)T})] +$$

$[E(U_{i+1}(f,T)|\mathcal{F}_{(i-1)T})-E(U_i(f\circ\Phi_T,T)|\mathcal{F}_{(i-1)T})]+[E(U_i(f\circ\Phi_T,T)|\mathcal{F}_{(i-1)T})$
$-U_i(f\circ\Phi_T)]$.

Then use equations (50), (51), and equation (47) with $f\circ\Phi_T$ in lieu of f. It follows that there exists a set $\Omega(f,T)\subset\Omega$ of full measure such that for all $\omega\in\Omega(f,T)$

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}U_{i+1}(f,T)-\frac{1}{n}\sum_{i=1}^{n}U_i(f\circ\Phi_T,T)=0. \tag{52}$$

Since (M,d) is a separable metric space it admits a metric $\tilde{d}$ inducing the same topology as d such that

(a) $\tilde{d}\le d$

(b) There exists a countable set $H=\{f^k\ k\in\mathbb{N}\}$ of uniformly continuous functions $f^k:(M,\tilde{d})\to[0,1]$ such that the topology of $\mathcal{P}(M)$ is induced by the metric

$$R(\mu,\nu)=\sum_k\frac{|\int_M f^k d\mu-\int_M f^k d\nu|}{2^k}$$

Statement (a) follows for example from the construction given in Lemma 3.1.4 of Stroock (1993) while (b) follows from Theorem 3.1.5 of Stroock (1993).

Let

$$\tilde{\Omega}=\bigcap_{k\in\mathbb{N},T\in\mathbb{Q}_+}\Omega(f^k,T)$$

Given $\omega\in\tilde{\Omega}$ and $\mu\in\mathcal{M}(X,\omega)$ there exists a sequence $t_j\to\infty$ (depending on ω) such that $\{\mu_{t_j}(\omega)\}$ converges weakly toward μ.

Let $n_j=[\frac{t_j}{T}]$ denote the integer part of $\frac{t_j}{T}$. Then

$$\lim_{j\to\infty}\frac{1}{n_jT}\sum_{i=0}^{n_j-1}\int_{iT}^{(i+1)T}f(X_s)ds=\int_M f(x)\mu(dx) \tag{53}$$

for any continuous and bounded function $f:M\to\mathbb{R}$. Hence by combining (53) and (52) we get that

$$\int_M(f^k\circ\Phi_T)(x)\mu(dx)=\int_M f^k(x)\mu(dx)$$

for all $f^k\in H$ and $T\in\mathbb{Q}^+$. This proves that μ is Φ_T invariant for all $T\in\mathbb{Q}^+$. By continuity of Φ this implies that μ is Φ_T invariant for all $T>0$ and since Φ is a semiflow, μ is Φ invariant. **QED**

Given $\mu\in\mathcal{P}(M)$ let $supp(\mu)$ denote the support of μ. Given a weak asymptotic pseudotrajectory X of Φ and $\omega\in\Omega$ we define the *minimal center of attraction* of $\{X(t,\omega):t\ge0\}$ as the (random) set

$$supp(X,\omega)=\overline{\bigcup_{\mu\in\mathcal{M}(X,\omega)}supp(\mu)}$$

Corollary 10.2 *Given a Borel set $A \subset M$ define*

$$\tau(\omega)(A) = \liminf_{t\to\infty} \mu_t(\omega)(A).$$

Suppose that for $P-$almost every ω $\{\mu_t(\omega)\}_{t\geq 0}$ is tight. Then for $P-$almost every ω

(i) $\tau(\omega)(supp(X,\omega)) = 1$ *and for any other closed set $A \subset M$ such that $\tau(\omega)(A) = 1$ it follows that $supp(X,\omega) \subset A$.*

(ii)

$$supp(X,\omega) \subset BC(\Phi) = \overline{\{x \in M \, : x \in \omega(x)\}}$$

Proof The proof of part (i) is an easy consequence of Theorem 10.1 and (ii) follows Theorem 10.1 and Poincaré recurrence Theorem (equation (30). **QED**

This last corollary has the interpretation that the fraction of time spends by a weak asymptotic pseudotrajectory in an arbitrary neighborhood of $BC(\Phi)$ goes to one with probability one.

10.1 Stochastic Approximation Processes with Slow Decreasing Step-Size

Consider a Robbins-Monro algorithm as described in section (4). Recall that $\overline{X} : \mathbb{R}_+ \to \mathbb{R}^m$ denotes the piecewise constant interpolated process given by $\overline{X}(t) = x_n$ for $\tau_n \leq t < \tau_{n+1}$. Set $\mathcal{F}_t = \mathcal{F}_n$ for $\tau_n \leq t < \tau_{n+1}$.

Proposition 10.3 *Let $\{x_n\}$ given by (7) be a Robbins-Monro algorithm. Assume*

(i) *F is Lipschitz on a neighborhood of $\{x_n : n \geq 0\}$,*

(ii) *x_0 is $\mathcal{F}_0$ measurable.*

(iii) $\lim_{R\to\infty}\{\sup_n E(||U_{n+1}||\mathbf{1}_{\{||U_{n+1}||\geq R\}} \, |\mathcal{F}_n)\} = 0$.

(iv) $\lim_{n\to\infty} \gamma_n = 0$.

Then $\overline{X}$ is a weak asymptotic pseudotrajectory of Φ. Hence $\overline{X}$ and X satisfy conclusion of Theorem 10.1 and Corollary 10.2.

Proof Given $R > 0$ let

$$U_{i+1}(R) = U_{i+1}\mathbf{1}_{\{|U_{i+1}|\leq R\}} - E(U_{i+1}\mathbf{1}_{\{||U_{i+1}||\leq R\}}|\mathcal{F}_i)$$

and

$$V_{i+1}(R) = U_{i+1} - U_{i+1}(R).$$

Then

$$P(\sup_{n\leq k\leq m(\tau_n+T)} ||\sum_{i=n}^{k}\gamma_{i+1}U_{i+1}|| \geq \alpha|\mathcal{F}_n) \leq$$

$$P(\sup_{n\leq k\leq m(\tau_n+T)} ||\sum_{i=n}^{k}\gamma_{i+1}U_{i+1}(R)|| \geq \alpha/2|\mathcal{F}_n)+P(\sup_{n\leq k\leq m(\tau_n+T)}$$

$$||\sum_{i=n}^{k}\gamma_{i+1}V_{i+1}(R)|| \geq \alpha/2|\mathcal{F}_n)$$

$$\leq \frac{4}{\alpha^2}C_2R^2\sum_{i=n}^{m(\tau_n+T)}\gamma_{i+1}^2 + \frac{4}{\alpha}E(\sum_{i=n}^{m(\tau_n+T)}\gamma_{i+1}E(||U_{i+1}||\mathbf{1}_{\{||U_{i+1}||>R\}}|\mathcal{F}_i)|\mathcal{F}_n)$$

where the first term in the right side of this inequality follows from inequality (16) obtained with $q = 2$ and the second term is an obvious estimate based on Markov inequality. Let $\epsilon > 0$. Assumption (iv) implies the existence of R large enough so that

$$\sup_i E(||U_{i+1}||\mathbf{1}_{\{||U_{i+1}||>R\}}|\mathcal{F}_i) \leq \epsilon.$$

Hence

$$\limsup_{n\to\infty} P(\sup_{n\leq k\leq m(\tau_n+T)} ||\sum_{i=n}^{k}\gamma_{i+1}U_{i+1}|| \geq \alpha|\mathcal{F}_n) \leq \frac{4T}{\alpha}\epsilon. \quad (54)$$

Inequality (54) combined with the estimate (11) proves the result. **QED**

References

Akin, E. (1993). *The General Topology of Dynamical Systems.* American Mathematical Society, Providence.

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

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

Benaïm, M. (1996). A dynamical systems approach to stochastic approximations. *SIAM Journal on Control and Optimization*, 34:141–176.

Benaïm, M. (1997). Vertex reinforced random walks and a conjecture of Pemantle. *The Annals of Probability*, 25:361–392.

Benaïm, M. and Hirsch, M. W. (1994). Learning processes, mixed equilibria and dynamical systems arising from repeated games. Submitted.

Benaïm, M. and Hirsch, M. W. (1995a). Chain recurrence in surface flows. *Discrete and Continuous Dynamical Systems*, 1(1):1–16.

Benaïm, M. and Hirsch, M. W. (1995b). Dynamics of morse-smale urn processes. *Ergodic Theory and Dynamical Systems*, 15:1005–1030.

Benaïm, M. and Hirsch, M. W. (1996). Asymptotic pseudotrajectories and chain recurrent flows, with applications. *J. Dynam. Differential Equations*, 8:141–176.

Benaïm, M. and Schreiber, S. J. (1997). Weak asymptotic pseudotrajectories for semiflows: Ergodic properties. Preprint.

Benveniste, A., Métivier, M., and Priouret, P. (1990). *Stochastic Approximation and Adaptive Algorithms*. Springer-Verlag, Berlin and New York.

Bowen, R. (1975). Omega limit sets of Axiom A diffeomorphisms. *J. Diff. Eq*, 18:333–339.

Brandière, O. (1996). Autour des pièges des algorithmes stochastiques. Thèse de Doctorat, Université de Marne-la-Vallée.

Brandière., O. (1997). Some pathological traps for stochastic approximation. *SIAM Journal on Control and Optimization*. To Appear.

Brandière, O. and Duflo., M. (1996). Les algorithmes stochastique contournent ils les pièges. *Annales de l'IHP*, 32:395–427.

Conley, C. C. (1978). *Isolated invariant sets and the Morse index*. CBMS Regional conference series in mathematics. American Mathematical Society, Providence.

Delyon, B. (1996). General convergence results on stochastic approximation. *IEEE trans. on automatic control*, 41:1245–1255.

Duflo, M. (1990). *Méthodes Récursives Aléatoires*. Masson. English Translation: Random Iterative Models, Springer Verlag 1997.

Duflo, M. (1996). *Algorithmes Stochastiques*. Mathématiques et Applications. Springer-Verlag.

Duflo, M. (1997). Cibles atteignables avec une probabilité positive d'après M. BENAIM. Unpublished manuscript.

Ethier, S. N. and Kurtz, T. G. (1986). *Markov Processes, Characterization and Convergence*. John Wiley and Sons, Inc.

Fort, J. C. and Pages, G. (1994). Résaux de neurones: des méthodes connexionnistes d'apprentissage. *Matapli*, 37:31–48.

Fort, J. C. and Pages, G. (1996). Convergence of stochastic algorithms: From Kushner-Clark theorem to the lyapounov functional method. *Adv. Appl. Prob*, 28:1072–1094.

Fort, J. C. and Pages, G. (1997). Stochastic algorithm with non constant step: a.s. weak convergence of empirical measures. Preprint.

Fudenberg, D. and Kreps, K. (1993). Learning mixed equilibria. *Games and Econom. Behav.*, 5:320–367.

Fudenberg, F. and Levine, D. (1998). *Theory of Learning in Games.* MIT Press, Cambridge, MA. In Press.

Hartman, P. (1964). *Ordinary Differential Equationq.* Wiley, New York.

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

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

Hirsch, M. W. (1994). Asymptotic phase, shadowing and reaction-diffusion systems. In *Differential equations, dynamical systems and control science*, volume 152 of *Lectures notes in pure and applied mathematics*, pages 87–99. Marcel Dekker, New-York.

Hirsch, M. W. and Pugh, C. C. (1988). Cohomology of chain recurrent sets. *Ergodic Theory and Dynamical Systems*, 8:73–80.

Kaniovski, Y. and Young, H. (1995). Learning dynamics in games with stochastic perturbations. *Games and Econom. Behav.*, 11:330–363.

Kiefer, J. and Wolfowitz, J. (1952). Stochastic estimation of the maximum of a regression function. *Ann. Math. Statis*, 23:462–466.

Kushner, H. J. and Clarck, C. C. (1978). *Stochastic Approximation for Constrained and Unconstrained Systems.* Springer-Verlag, Berlin and New York.

Kushner, H. J. and Yin, G. G. (1997). *Stochastic Approximation Algorithms and Applications.* Springer-Verlag, New York.

Ljung, L. (1977). Analysis of recursive stochastic algorithms. *IEEE Trans. Automat. Control.*, AC-22:551–575.

Ljung, L. (1986). *System Identification Theory for the User.* Prentice Hall, Englewood Cliffs, NJ.

Ljung, L. and Sőderstrőm, T. (1983). *Theory and Practice of Recursive Identification.* MIT Press, Cambridge, MA.

Mañé, R. (1987). *Ergodic Theory and Differentiable Dynamics.* Springer-Verlag, 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.

Munkres, J. R. (1975). *Topology a first course.* Prentice Hall.

Nevelson, M. B. and Khasminskii, R. Z. (1976). *Stochastic Approximation and Recursive Estimation.* Translation of Math. Monographs. American Mathematical Society, Providence.

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

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

Robbins, H. and Monro, S. (1951). A stochastic approximation method. *Ann. Math. Statis*, 22:400–407.

Robinson, C. (1977). Stability theorems and hyperbolicity in dynamical systems. *Rocky Journal of Mathematics*, 7:425–434.

Robinson, C. (1995). *Introduction to the Theory of Dynamical Systems.* Studies in Advances Mathematics. CRC Press, Boca Raton.

Schreiber, S. J. (1997). Expansion rates and Lyapunov exponents. *Discrete and Conts. Dynam. Sys.*, 3:433–438.

Shub, M. (1987). *Global Stability of Dynamical Systems.* Springer-Verlag, Berlin, New York, Heidelberg.

Stroock, D. W. (1993). *Probability Theory. An analytic view.* Cambridge University Press.

White, H. (1992). *Artificial Neural Networks: Approximation and Learning Theory.* Blackwell, Cambridge, Massachussets.