Stationary Densities and the Stochastic Approximation of a Certain Class of Random Algorithms

Israel Ncube[1], S. A. Campbell[2] and E. R. Vrscay[3]

[1]Toronto Western Research Institute
University Health Network
399 Bathurst Street, MP 13–308
Toronto, Ontario, Canada, M5T 2S8

[2,3]Department of Applied Mathematics
University of Waterloo
Waterloo, Ontario, Canada, N2L 3G1

Abstract. The problem of existence and uniqueness of so-called stationary densities for a certain class of non-autonomous discrete random neural network-type learning algorithms is studied. We develop and expound on the idea of a suitable complete metric space of distributions on which to carry out this type of investigation. However, because of the general level of difficulty associated with analysis on the space of distributions, we perform most of our investigations on the usual space, namely the space of normalised positive L^1 functions. This is not a handicap per se, as one can get useful information about the nature of the stationary densities, even if working on such a relatively small space. Fundamentally, the paper is a systematic investigation of the contractivity, or lack thereof, of the well-known *Frobenius-Perron operator* generated by these algorithms. It is shown that, for certain linear deterministic algorithms, the corresponding Frobenius-Perron operator is contractive over our metric space of distributions. Further, analysis in L^1 and computer simulations suggest that the operator generated by linear random algorithms and nonlinear deterministic and random algorithms is contractive in this metric space as well. However, analytically proving the latter is problematic. The primary motivation of our investigation is the promising use of stationary densities in understanding the convergence properties of the random algorithms, especially in the case when the so-called *associated ordinary differential equation* has multiple locally asymptotically stable equilibria, and no other stable structures.

Key words. neural networks, learning, stationary densities, stochastic approximation.
AMS subject classification. 92B20, 34K20, 34K15.

[1]To whom correspondence should be addressed.
email: incube@duchatelet.math.uwaterloo.ca

1 Introduction

We consider non-autonomous iterative random algorithms of the form

$$\mathbf{x}_{n+1} = \mathbf{x}_n + \gamma_n h[\mathbf{x}_n, \varphi_n] \ , \quad n = 0, 1, 2, \cdots \ , \tag{1}$$

where the $\mathbf{x}_n$'s and φ_n's are in $\mathbb{R}^m$, $\{\gamma_n\}$ is a sequence of positive decreasing-to-zero real numbers such that $\sum_n \gamma_n = \infty$, $h \in C[R^m \times \mathbb{R}^m, \mathbb{R}^m]$, and $\{\varphi_n\}$ is a sequence of random variables that are distributed according to some given law. To ensure satisfactory convergence of the algorithm, γ_n is typically taken to be $(n+1)^{-\alpha}$, where $\alpha < 1$ (see [2, 9, 10, 11] for details). Throughout the paper, we shall consider γ_n to be of this form.

We may rewrite (1) in the following way:

$$\begin{aligned} \mathbf{x}_{n+1} &= \mathbf{x}_n + \gamma_n \overline{h}(\mathbf{x}_n) + \gamma_n [h(\mathbf{x}_n, \varphi_n) - \overline{h}(\mathbf{x}_n)] \\ &= \mathbf{x}_n + \gamma_n \overline{h}(\mathbf{x}_n) + \gamma_n \xi_n \\ &= S_n(\mathbf{x}_n) + \gamma_n \xi_n \ , \end{aligned} \tag{2}$$

where

$$\overline{h}(\mathbf{z}) = E[h(\mathbf{z}, \varphi_n) \mid \varphi_0, \cdots, \varphi_{n-1}] \ , \tag{3}$$

for $\mathbf{z} \in \mathbb{R}^m$ and where E symbolises the statistical expectation with respect to φ_n.

The subject of stochastic approximation is concerned with characterisation of the long term behaviour of iterative random algorithms such as (1) or (2). A myriad of different approaches to this problem have been proposed [1-10].

A particular issue of topical interest may be posed by the following question: *Does the algorithm in (2) converge to a unique fixed point, independent of initialisation?* Ljung [1] and Kushner *et al.* [2] addressed this question by considering the (associated) ordinary differential equation (ODE)

$$\frac{d\mathbf{z}}{dt} = \overline{h}[\mathbf{z}(t)] \ , \quad \mathbf{z} \in \mathbb{R}^m \ . \tag{4}$$

Their famous theorem asserts that (2) converges, with probability one (wp1), to a stable equilibrium of (4) if the latter has exactly one stable equilibrium point and no other stable structures.

However, if the ODE (4) has multiple stable equilibria (with or without other stable structures), very little is known about the behaviour of (2). One of the few papers on this subject is that of Fort and Pagès [3]. They established and proved a theorem which permitted them to relate the convergence of solutions of the associated ODE to that of the sequence of iterates generated by the random algorithm, in the case that the ODE has no *pseudocycles* (these include bona fide periodic orbits as well as isolated equilibria). Their approach is primarily a development of the original Kushner-Clark theorem [2]. The proof of their theorem amounts to proving that the sequence of iterates generated by the algorithm has only one limiting point in the set of all equilibria of the associated ODE. Furthermore, they showed that if one of the elements of the above set is a saddle point, then the algorithm will not converge to it.

The primary focus of the present endeavour is the investigation of existence (and uniqueness), or lack thereof, of stationary densities associated with the sequence $\{\mathbf{x}_n\}$ generated by random algorithms of the form (2), especially when the ODE (4) has two stable equilibria. We study the evolution of densities under the action of the Frobenius-Perron operator corresponding to the map S_n, for both the nonlinear algorithm (2) as well as its linearised analogue

$$\mathbf{x}_{n+1} = \alpha_n \mathbf{x}_n + \gamma_n \xi_n \ , \tag{5}$$

where $\alpha_n = S_n'(\hat{\mathbf{x}}) = 1 + \gamma_n \overline{h}'(\hat{\mathbf{x}})$ and $\hat{\mathbf{x}}$ is a stable equilibrium point of (4). Eventually, we hope to use these stationary densities, if they exist, to compute probabilities of convergence of (2) to any one of the stable equilibria of (4).

In fact, some work has been done on equations related to (5). Consider the deterministic linearised one-dimensional algorithm

$$\mathbf{x}_{n+1} = (1 - \lambda n^{-\alpha})\mathbf{x}_n \ , \tag{6}$$

where $\lambda > 0$, and $\alpha \in (0,1)$. Here λ can be identified with $-\bar{h}'(\hat{x})$. We may rewrite (6) in closed form as

$$\mathbf{x}_n = \mathbf{x}_1 \prod_{k=1}^{n-1} (1 - \lambda k^{-\alpha}) \ .$$

It is then straightforward to show that the long term behaviour of the sequence $\{\mathbf{x}_n\}$ is characterised by

$$\begin{aligned} \mathbf{x}_n \quad &\sim \quad \mathbf{x}_1 e^{-\frac{\lambda}{1-\alpha} n^{1-\alpha}} \\ &\rightarrow \quad 0 \ \text{ as } \ n \rightarrow \infty \ . \end{aligned}$$

Let us now perturb (6) by a deterministic forcing term, i.e.

$$\mathbf{x}_{n+1} = (1 - \lambda n^{-\alpha})\mathbf{x}_n + n^{-\alpha}\mathbf{r}_n \ , \tag{7}$$

where $\mathbf{r}_n$ is the forcing term. Chung [3] proved that if $\mathbf{r}_n \to 0$ faster than $n^{-\epsilon}$, where $\epsilon > 0$, then $\mathbf{x}_n \to 0$ as $n \to \infty$.

Fabian [4] considered the special case $\alpha = 1$ of a scalar version of (7), and proved that $x_n \to 0$ as $n \to \infty$ if and only if $n^{-1} \sum_{j=1}^{n} r_j \to 0$ as $n \to \infty$. His result is consistent with intuition, since $n^{-1} \sum_{j=1}^{n} r_j$ may be interpreted as the mean of the sequence $\{r_k\}_{k=1}^{n}$. In other words, he demonstrated that if the mean of $\{r_k\}_{k=1}^{n}$ converges to zero, then the iterates x_n will likewise converge to zero as $n \to \infty$.

Our ultimate goal is to examine the long term behaviour of the sequence $\{\mathbf{x}_n\}$ when $\{\mathbf{r}_n\}$ is replaced by a sequence of random variables $\{\xi_n\}$ with some prescribed probability density function. In this case, it is not guaranteed that $\xi_n \to 0$ as $n \to \infty$. Nevertheless, note that if $\{\xi_n\}$ is a compactly supported sequence of random variables, and

$$\mathbf{x}_{n+1} = (1 - \lambda n^{-\alpha})\mathbf{x}_n + n^{-\gamma}\xi_n \ , \quad \gamma > \alpha \ ,$$

then Chung's result [3] guarantees that $\mathbf{x}_n \to 0$ as $n \to \infty$.

As with any random process (see [15], [17]), it is not possible to obtain meaningful convergence results by simply tracking individual trajectories, as can be done for ergodic transformations. Instead, more meaningful conclusions may be drawn based on an examination of the evolution of densities of $\{\mathbf{x}_n\}$ [5]. The utility of this approach lies in the fact that it takes into account *all* possible initial states. An initial probability density function, f_0, is defined over all the possible initial states of the algorithm. Then, the idea is to determine how this prescribed density of initial states evolves over time. Ultimately, one hopes to find a (limiting) stationary density f_* for the sequence $\{f_n\}$ of densities of $\{\mathbf{x}_n\}$.

The paper is organised as follows. Section 2 develops and articulates the pertinent mathematical framework. Sections 3 and 4 are detailed studies of linearised and nonlinear non-autonomous algorithms, respectively. In Section 5, a specific example is considered. Section 6 is the conclusion.

2 Mathematical framework

2.1 The space $L^1_D(X, \Lambda, \mu)$

Conventionally, most authors (see [12], for example) formulate the analysis of the sequence of densities $\{f_n\}_{n=1}^{\infty}$ in the space L^1_D of probability density functions. However, as will be shown in Section 2.3, this space is not always appropriate. For completeness, we give a brief description of L^1_D. Let (X, Λ, μ) be a measure space, where Λ is a σ-algebra in X, and where μ is a measure on Λ. Consider the set Γ of all real-valued functions f which are absolutely integrable over X, i.e.

$$\int_X \mid f \mid d\mu < \infty \ ,$$

where, in most practical applications, μ is Lebesgue measure. Note that Γ is a linear space, since every finite linear combination of integrable functions is integrable. Now define the subset $\Gamma_0 \subset \Gamma$ by $\Gamma_0 = \{f \in \Gamma \mid f = 0 \ a.e.\}$. Then the real valued functional p defined below is a norm on the factor space Γ/Γ_0, where Γ/Γ_0 is denoted by $L^1(X, \Lambda, \mu)$:

$$p : L^1 \to \mathbb{R} \ , \ \ p(f) = \int_X \mid f \mid d\mu \ , \text{ for all } f \in L^1 \ . \tag{8}$$

It is usual to denote $p(f)$ by $\|f\|_1$. $L^1(X, \Lambda, \mu)$ is a metric space with the metric given by

$$\rho(f, g) = \|f - g\|_1 \ , \text{ for all } f, g \in L^1(X, \Lambda, \mu) \ . \tag{9}$$

The "conventional" space on which densities are defined is denoted by $L^1_D(X, \Lambda, \mu) \subset L^1(X, \Lambda, \mu)$, and defined by

$$L^1_D(X, \Lambda, \mu) = \left\{ f \mid f \geq 0 \text{ and } \int_X f d\mu = 1 \right\} \ . \tag{10}$$

It is well-known that the space L^1_D is complete with respect to the L^1 norm.

Definition 2.1 *Any function $f \in L^1_D(X, \Lambda, \mu)$ is called a density.*

Remark 2.1 *Under some conditions, a sequence of densities $\{f_n\}_{n=1}^{\infty}$ which does not converge in the space $L^1_D(X, \Lambda, \mu)$ can converge in an appropriate larger space. This necessitates the definition of the* space of all objects to which one can associate a distribution*, that includes $L^1_D(X, \Lambda, \mu)$. We will return to this issue in subsections 2.3 and 2.4.*

2.2 The Frobenius-Perron operator

Suppose that we have a non-singular, measurable transformation $S : X \to X$ on a measure space (X, Λ, μ). For our purposes, S shall be defined as the deterministic part of the algorithm under consideration, and may or may not be explicitly dependent on time. For example, in equation (5), it is given by $S(x_n) = \alpha_n x_n$, where S explicitly depends on time. In what follows, the notion of the Frobenius-Perron operator for deterministic *autonomous* algorithms is developed. The generalisation of the results to deterministic *non-autonomous* algorithms is straightforward, as is demonstrated in Section 3.1. In that case (non-autonomous), the corresponding operator is dependent on time. So, instead of getting one operator for the transformation S as is the case with autonomous algorithms, one obtains a sequence of operators, $\{P_n \ , \quad n \in \mathbb{N}\}$, where P_n corresponds to the transformation $S_n(\cdot)$. Let $f \in L^1_D(X, \Lambda, \mu)$ be an arbitrary density. The Frobenius-Perron operator, $P : L^1_D \to L^1_D$, describes the evolution of f induced by S. In other words, if f defines the distribution of initial conditions, i.e. points $x_0 \in X$, then Pf gives the distribution of points $x_1 = Sx_0$. Define the action of P on f as follows

$$\int_A Pf(x)\mu(dx) = \int_{S^{-1}(A)} f(x)\mu(dx) \ , \text{ for } A \in \Lambda \ . \tag{11}$$

This relationship uniquely defines P (see [12] for details). From (11), it may be shown that P has the following properties.

1. P is a linear operator. That is,

$$P(\lambda_1 f_1 + \lambda_2 f_2) = \lambda_1 P f_1 + \lambda_2 P f_2 \ , \tag{12}$$

for all $f_1, f_2 \in L^1, \ \lambda_1, \lambda_2 \in \mathbb{R}$.

2. For $f \in L^1$,

$$Pf \geq 0 \text{ if } f \geq 0 \ \text{ on } \ X. \tag{13}$$

3.

$$\int_X Pf(x)\mu(dx) = \int_X f(x)\mu(dx) \ . \tag{14}$$

4. If $S_n = S \circ \overset{n\ times}{\cdots\cdots} \circ S$ and P_n is the Frobenius-Perron operator corresponding to S_n, then $P_n = P^n$, where P is the Frobenius-Perron operator corresponding to S.

5. In the special case when $X = \mathbb{R}$, choosing $A = [0, x]$, we have that

$$Pf(x) = \frac{d}{dx} \int_{S^{-1}([0,x])} f(u)du \ .$$

Note that (13) and (14) imply that Pf is a density.

We now digress for a moment to introduce the *Koopman operator* (see [12] for details), which is adjoint to the Frobenius-Perron operator.

Definition 2.2 *Let (X, Λ, μ) be a measure space, $S : X \to X$ a non-singular transformation, and $f \in L^\infty$. The operator $U : L^\infty \to L^\infty$ defined by*

$$Uf(x) = f(S(x))$$

is called the Koopman operator induced by S.

As a result of the non-singularity of S, the operator U is well-defined since $f_1(x) = f_2(x)$ *a.e.* implies that $f_1(S(x)) = f_2(S(x))$ *a.e.*. Listed below are some essential properties of U:

1. $U(\lambda_1 f_1 + \lambda_2 f_2) = \lambda_1 U f_1 + \lambda_2 U f_2$ for all $f_1, f_2 \in L^\infty$, $\lambda_1, \lambda_2 \in \mathbb{R}$;

2. For every $f \in L^\infty$,
$$\|Uf\|_{L^\infty} \leq \|f\|_{L^\infty} \ .$$
We note that [12] refers to any operator satisfying this property (e.g. P) as *contractive*, which is not the usual definition of contractive operators.

3. For every $f_1 \in L^1_D$, $f_2 \in L^\infty$,
$$\langle Pf_1, f_2 \rangle \ = \ \langle f_1, Uf_2 \rangle \ ,$$
so that U is adjoint to the Frobenius-Perron operator.

Definition 2.3 *Any function $f \in L^1_D(X, \Lambda, \mu)$ that satisfies $Pf = f$ is called a stationary density of P.*

To illustrate the computation of the Frobenius-Perron operator for *autonomous* deterministic algorithms, consider the map

$$x_{n+1} = \alpha x_n \ , \ \ \alpha \in (0, 1) \ .$$

Here $S(x_n) = \alpha x_n$, and $X = \mathbb{R}$. For any $f \in L^1_D$, the associated Frobenius-Perron operator is

$$\begin{aligned} Pf(x) &= \frac{d}{dx}\int_{S^{-1}([a,x])} f(u)du \\ &= \frac{d}{dx}\int_{\frac{a}{\alpha}}^{\frac{x}{\alpha}} f(u)du \\ &= \frac{1}{\alpha}f(\frac{x}{\alpha}) , \end{aligned}$$

where $[a, x] \subset \mathbb{R}$. In light of this, we obtain

$$\begin{aligned} f_n(x) &\stackrel{def}{=} P^n f(x) \\ &= \frac{1}{\alpha^n}f(\frac{x}{\alpha^n}) . \end{aligned}$$

2.3 An example showing the inadequacy of L^1_D

To motivate the need for a broader space of densities that encompasses L^1_D, consider the iteration of a uniform probability density function (pdf)

$$f(x) = \begin{cases} 1 & , \ x \in (-\frac{1}{2}, \frac{1}{2}) \\ 0 & , \ \text{elsewhere} \end{cases} \tag{15}$$

under the Frobenius-Perron operator defined above. The action of P on f generates a sequence of densities, $\{f_n\}$, defined by

$$f_n(x) \stackrel{def}{=} P^n f(x) = \begin{cases} \frac{1}{\alpha^n} & , \ x \in (-\frac{\alpha^n}{2}, \frac{\alpha^n}{2}) \\ 0 & , \ \text{elsewhere} . \end{cases} \tag{16}$$

Suppose that $m < n$, and consider the densities f_m and f_n. As illustrated in Fig 1, we wish to compute the L^1 distance between f_m and f_n. We obtain

$$\begin{aligned} \|f_n - f_m\|_1 &= \|P^n f(x) - P^m f(x)\|_1 \\ &= 2(1 - \alpha^{n-m}) \\ &< 2(1 - \alpha^n) \to 2 \ \text{as} \ n \to \infty . \end{aligned} \tag{17}$$

Thus, $\{f_n\}$ is *not* Cauchy in L^1_D. However, the algorithm $x_{n+1} = \alpha x_n$, $\alpha \in (0, 1)$, represents the iteration of the contraction mapping $f(x) = \alpha x$,

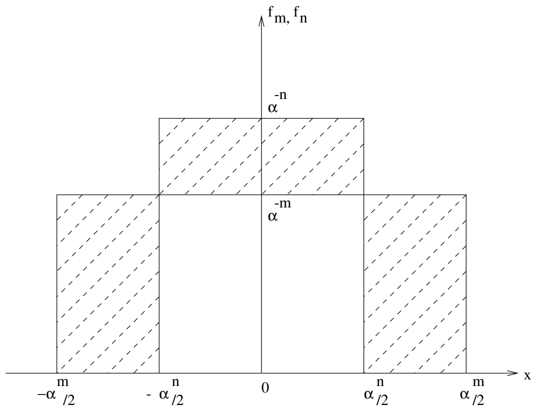

Figure 1: Iteration of a uniform pdf f under P associated with the map $S(x) = \alpha x$, $\alpha \in (0,1)$.

which has a unique fixed point $\bar{x} = 0$. Thus, we might expect the Frobenius-Perron operator to possess a unique invariant density which is *not* a function but rather the unit mass measure $\bar{\mu}$ with support at $x = 0$, i.e. the "Dirac distribution function", $\delta(x)$. Thus, L^1_D is not an appropriate space for the analysis of evolution of densities under the action of this Frobenius-Perron operator. What is needed is to view these densities as special cases of distributions.

In view of the above example, we now formally introduce the space of normalised positive distributions, denoted by $\overline{D}'_+(X)$. The motivation for the work in the next section comes from a paper of Forte and Vrscay [16]. Essentially, they look at a similar space, but with more general distributions. In this paper, we focus on a subset of their space, the space of normalised positive distributions.

2.4 A suitable metric space of distributions

Let X be a compact and connected subset of $\mathbb{R}$. Distributions are defined as continuous linear functionals over a suitable space of test functions, denoted as $D(X)$. The space of distributions on X, denoted by $D'(X)$, is the set of all continuous linear functionals on $D(X)$. In other words, the set of $F : D(X) \to \mathbb{R}$ such that

1. For any sequence of test functions $\{\psi_\nu(t)\}_{\nu=1}^{\infty}$ that converges in $D(X)$

to $\psi(t)$, the sequence of numbers $\{F(\psi_\nu)\}_{\nu=1}^{\infty}$ converges to the real number $F(\psi)$ in the usual sense [18].

2. $F(c_1\psi_1 + c_2\psi_2) = c_1 F(\psi_1) + c_2 F(\psi_2)$, $c_1, c_2 \in \mathbb{R}, \psi_1, \psi_2 \in D(X)$.

The space $D'(X)$ includes the following as special cases:

a. Functions $f \in L^1$, for which the corresponding distributions are given by

$$F(\psi) = \int_X f(x)\psi(x)dx \ , \text{ for all } \psi \in D(X) \ .$$

b. The Dirac distribution, $\delta(x-a)$, which may be defined in the distributional sense as follows: For a point $a \in X$, $F(\psi) = \psi(a)$, for all $\psi \in D(X)$. This is often symbolically expressed as

$$F(\psi) = \int_X \psi(x)\delta(x-a)dx \ .$$

In this paper, we are interested in the space of *normalised positive distributions*, denoted as $\overline{D}'_+(X)$, and defined by

$$\overline{D}'_+(X) = \{F \in D'(X) | \ F(1) = 1 \ , \ \text{ and } \ F(\psi) \geq 0 \ \ \forall \psi \in C_+^\infty(X)\} \ , \quad (18)$$

where $C_+^\infty = \{\psi \in C^\infty | \ \psi(x) \geq 0 \ , \ x \in X\}$. The physical motivation for the assumptions $F(\psi) \geq 0$ and $F(1) = 1$ is that we wish to interpret the values of the distributions as probability measures, as is done for integrals of density functions in L_D^1. In fact, note that $L_D^1(X) \subset \overline{D}'_+(X)$.

In the following analysis involving the space $\overline{D}'_+(X)$, we shall restrict our test functions to a subset of C_+^∞, namely, positive C^∞ functions that are Lip_1 on X, viz. $D(X) \stackrel{def}{=} Lip_1^+(X)$, where

$$Lip_1^+(X) = \{\psi \in Lip_1(X) | \ \psi(x) \geq 0 \ , \ \text{ for all } \ x \in X\} \ , \quad (19)$$

and where

$$Lip_1(X) = \{\psi : X \to \mathbb{R} \, | \ |\psi(x_1) - \psi(x_2)| \leq d(x_1, x_2) \ , \ \text{ for all } \ x_1, x_2 \in X\} \ .$$

The following property is very important in formulating a representation theory for distributions in $\overline{D}'_+(X)$.

Theorem 2.1 *For any distribution $F \in \overline{D}'_+(X)$, there exists a sequence of test functions $f_n \in Lip_1^+(X)$, $n = 1, 2, \cdots$, such that for all $\psi \in Lip_1^+(X)$,*

$$\begin{aligned} \lim_{n\to\infty} F_n(\psi) &= \lim_{n\to\infty} \int_X f_n(x)\psi(x)dx \\ &\stackrel{def}{=} F(\psi) \, . \end{aligned}$$

This result is a rather simple specialisation of a theorem for the case $F \in D'(X)$, which was stated in [16]. By recourse to the above result, it will be convenient to express the distribution $F \in D'_+(X)$ symbolically as

$$F(\psi) = \int_X f(x)\psi(x)dx \, ,$$

even though there may not exist a pointwise function $f(x)$ which defines F (e.g. the Dirac distribution). *For notational convenience, given $f \in L^1_D$, we will write "$f \in \overline{D}'_+$" meaning that one can associate a distribution $F \in D'_+(X)$ to f.* (In the same way, we can write "$\delta \in D'_+(X)$", where δ is the Dirac delta function.)

In [16], a metric was introduced over the space $D'(X)$. Following this treatment, we introduce a metric over the space $\overline{D}'_+(X)$:

$$d_{\overline{D}'_+}(f,g) = sup_{\psi \in Lip_1^+(X)} \left\{ | \int_X (f-g)(x)\psi(x)dx | \right\} \, , \text{ for all } f,g \in \overline{D}'_+(X) \, . \tag{20}$$

A major difference is the use of Lip_1^+ test functions in this metric, as opposed to test functions inside the unit C^∞ ball used in [16]. Our restriction to normalised positive distributions permits the use of Lip_1^+ functions, as we now show.

Given two test functions $\psi_1(x)$, $\psi_2(x)$ such that $\psi_1(x) = \psi_2(x) + c$, where $c \in \mathbb{R}$ and $f, g \in \overline{D}'_+(X)$, then

$$\int_X f(x)\psi_2(x)dx - \int_X g(x)\psi_2(x)dx = \int_X f(x)\psi_1(x)dx - \int_X g(x)\psi_1(x)dx \, .$$

In other words, the metric will not be affected by translations in the test functions. This allows us to use Lip_1 functions as is done for probability measures [7].

Theorem 2.2 *The metric space $(\overline{D}'_+(X), d_{\overline{D}'_+})$ is complete.*

Proof
Let $\{f_n\}_{n=1}^{\infty}$ be a Cauchy sequence in $(\overline{D}'_+(X), d_{\overline{D}'_+})$. In other words, for any $\epsilon > 0$, there exists an $N(\epsilon)$ such that $d_{\overline{D}'_+}(f_n, f_m) < \epsilon$, for all $n, m > N(\epsilon)$. From the definition of $d_{\overline{D}'_+}$ in (20), it follows that for any fixed $\psi \in Lip_1^+(X)$, the sequence of real numbers $\{t_n(\psi)\}_{n=1}^{\infty}$, where

$$t_n(\psi) = \int_X f_n(x)\psi(x)dx \ ,$$

is a Cauchy sequence on $\mathbb{R}$. The latter is true since, for any $\tilde{\psi} \in Lip_1^+(X)$, we have that $|t_n(\tilde{\psi}) - t_m(\tilde{\psi})| \leq sup_{\psi \in Lip_1^+(X)} |t_n(\psi) - t_m(\psi)| = d_{\overline{D}'_+}(f_n, f_m) < \epsilon$, for all $n, m > N(\epsilon)$. Let $\bar{t}(\psi)$ denote the limit of this sequence. Note that $\bar{t}(\psi) \geq 0$ for each $\psi \in Lip_1^+(X)$, since $\{t_n(\psi)\}_{n=1}^{\infty}$ is non-negative. By setting $F(\psi) = \bar{t}(\psi)$, we define a continuous linear functional F on $Lip_1^+(X)$. Furthermore, since $t_n(1) = 1$, it follows that $\bar{t}(1) = 1$. Therefore $F(1) = 1$, implying that $F \in \overline{D}'_+(X)$. This procedure can be easily extended to all C_+^{∞} test functions on X by noting (via the Mean Value Theorem) that $M^{-1} \times \psi \in Lip_1^+(X)$, where $M = \|\psi'\|_{\infty}$. This completes the proof. □

We now illustrate the use of the metric space $(\overline{D}'_+(X), d_{\overline{D}'_+})$ in the investigation of the Frobenius-Perron operator P corresponding to the simple linear map $S(x) = \alpha x$, $\alpha \in (0,1)$. Note that P will now have to be a mapping from $\overline{D}'_+(X)$ to itself. We proceed in a manner analogous to that described in Section 2.2, with particular reference to the example considered in that section. Now, for any $f \in \overline{D}'_+(X)$, the distribution $q = Pf$ is defined by the linear functional

$$\begin{aligned} Q(\psi) &= \int_X q(x)\psi(x)dx \\ &= \int_X (Pf)(x)\psi(x)dx \\ &= \alpha^{-1} \int_{X_\alpha} f(\frac{x}{\alpha})\psi(x)dx \\ &= \int_X f(y)\psi(\alpha y)dy \ , \end{aligned} \tag{21}$$

where $\psi \in D(X)$. In the penultimate line, $X_\alpha \overset{def}{=} \{y \in X \mid y = S(x),\ x \in X\}$.

Theorem 2.3 *P is contractive in $(\overline{D}'_+(X), d_{\overline{D}'_+})$.*

Proof
Suppose that $\psi \in Lip_1^+(X)$, and define $\tilde{\psi}(y) = \alpha^{-1}\psi(\alpha y)$. Then

$$\begin{aligned} |\tilde{\psi}(x) - \tilde{\psi}(y)| &= \alpha^{-1}|\psi(\alpha x) - \psi(\alpha y)| \\ &\leq |x - y| \, , \end{aligned}$$

which implies that $\tilde{\psi} \in Lip_1^+(X)$. In addition, define

$$\tilde{L} = \{\tilde{\psi} \in Lip_1^+(X)| \;\; \tilde{\psi}(x) = \alpha^{-1}\psi(\alpha x) \, , \;\; \text{for some} \;\; \psi \in Lip_1^+(X)\} \, .$$

Then, (20) yields

$$\begin{aligned} d_{\overline{D}'_+}(Pf, Pg) &= sup_{\psi \in Lip_1^+(X)}\{| \int_X [f - g](y)\psi(\alpha y)dy|\} \\ &= sup_{\tilde{\psi} \in \tilde{L}}\{\alpha| \int_X [f - g](y)\tilde{\psi}(y)dy|\} \\ &\leq \alpha \sup_{\psi \in Lip_1^+(X)} \{| \int_X [f - g](y)\psi(y)dy|\} \, , \;\; \text{since} \;\; \tilde{L} \subset Lip_1^+(X) \\ &= \alpha d_{\overline{D}'_+}(f, g) \, , \end{aligned} \tag{22}$$

which gives the desired result. □

This result resolves the difficulty encountered in Section 2.3. By Banach's Fixed Point Theorem, there exists a unique fixed point of the operator P in the metric space $(\overline{D}'_+(X), d_{\overline{D}'_+})$. Therefore our sequence of densities converges when viewed in the space of distributions. Together, theorems 2.2 and 2.3 imply that there exists a *unique* $f_* \in \overline{D}'_+(X)$ such that

1. $Pf_* = f_*$, and
2. $d_{\overline{D}'_+}(P^n f, f_*) \to 0$, as $n \to \infty$ for any $f \in \overline{D}'_+(X)$.

The last statement follows from the observation that

$$d_{\overline{D}'_+}(P^n f, f_*) = d_{\overline{D}'_+}(P(P^{n-1}f), P(P^{n-1}f_*)) \leq \alpha^n d_{\overline{D}'_+}(f, f_*) \to 0 \, , \;\; \text{as} \;\; n \to \infty \, .$$

We now show that $f_* = \delta(x)$, the Dirac delta "function". Let F be the Dirac distribution, i.e.

$$\begin{aligned} F(\psi) &= \int_X \delta(x)\psi(x)dx \\ &= \psi(0) \, . \end{aligned}$$

Then, from (21), Q, the distribution associated with the Frobenius-Perron operator, is

$$\begin{aligned} Q(\psi) &= \int_X \delta(y)\psi(\alpha y)dy \\ &= \psi(0) , \end{aligned}$$

which gives the desired result.

3 Densities of Linear Algorithms

The previous section has given us a complete description of the long term behaviour of the linear map

$$x_{n+1} = \alpha x_n \ , \ \alpha \in (0,1) \ , \tag{23}$$

where $S(x_n) \stackrel{def}{=} \alpha x_n$. The main thrust of this section is to investigate the long term behaviour of sequence of densities $\{f_n\}$ for two generalisations and noise-driven versions of (23). The analysis in this section suggests that the sequence $\{f_n\}$ does converge to elements outside of L^1_D, and hence the need for a broader space $\overline{D}'_+(X)$ that includes L^1_D. However, in the rest of this paper, we shall formulate the analysis of $\{f_n\}$ in L^1_D, following the treatment of Mackey and Lasota [12]. Analysis in the space of distributions $\overline{D}'_+(X)$ of algorithms more complicated than (23) is difficult and beyond the scope of this paper. Besides, it appears that the space L^1_D is sufficient for at least a preliminary study of random processes: one does get some indications (albeit heuristic) that, in the limit as $n \to \infty$, the sequence of densities is concentrated in arbitrarily small neighbourhoods (as is illustrated on page 11 of [12]).

The most general linear algorithm that we will consider is of the form

$$x_{n+1} = \alpha_n x_n + \gamma_n \xi_n \ , \tag{24}$$

where $\{\xi_n\}$ is a sequence of *i.i.d.* random variables. Following Section 1, $\alpha_n = 1 + \gamma_n \bar{h}'(\hat{x}) < 1$, where $\hat{x}$ is a stable equilibrium point of the ODE

$$\dot{x} = \bar{h}(x) \ .$$

The parameter γ_n is called the gain of the algorithm (also referred to as the *learning parameter* in neural networks literature) and has the properties mentioned in Section 1, viz.

$$\gamma_n \to 0 \text{ as } n \to \infty , \text{ and } \sum_n \gamma_n = \infty .$$

Below, we study the following linearised (see Section 1 for details) variations of algorithm (2):

i. $x_{n+1} = \alpha_n x_n$, and

ii. $x_{n+1} = \alpha_n x_n + \gamma_n \xi_n$.

3.1 Algorithm (i)

Applying result 5 in Section 2.2, it can be shown that the n^{th} iterate of the density $f(x)$ is given by

$$P_n f(x) = \alpha_n^{-1} f(\alpha_n^{-1} x) , \tag{25}$$

where P_n is the n^{th} Frobenius-Perron operator. Note that P_n acting on f contracts it in the x direction and expands it in the y direction. In other words, the RHS of (25) gets more "spiked" around the origin, as $n \to \infty$. Each α_n gives a different operator, leading to a sequence of operators $\{P_1, P_2, ..., P_n\}$. Now, iterating these operators, (25) becomes

$$P_n P_{n-1} ... P_1 f(x) = \frac{1}{\alpha_n ... \alpha_1} f\left(\frac{x}{\alpha_n ... \alpha_1}\right) . \tag{26}$$

Consider an arbitrary bounded interval $[-A, A] \subset \mathbb{R}$, where $A > 0$. Then

$$\int_{-A}^{A} P_n P_{n-1} ... P_1 f(x) dx = \int_{-\frac{A}{\alpha_n ... \alpha_1}}^{\frac{A}{\alpha_n ... \alpha_1}} f(x) dx , \tag{27}$$

which approaches $\int_{-\infty}^{\infty} f(x)dx = 1$ as $n \to \infty$. In other words, for any $A > 0$, the RHS of (27) can be made as close to unity as desired by taking n large enough. This suggests that the sequence of densities $\{f_n\}_{n=1}^{\infty}$ converges to the Dirac delta function $\delta(x)$, where the convergence is understood in the sense of distributions. We illustrate this in the following way: Suppose $f(x) = f_0 > 0$ for $x \in [-\frac{1}{2f_0}, \frac{1}{2f_0}]$, and zero elsewhere. Then

$$P_n \ldots P_1 f(x) = \begin{cases} 0\ , \text{ for } x \notin [-\frac{\alpha_1 \ldots \alpha_n}{2f_0}, \frac{\alpha_1 \ldots \alpha_n}{2f_0}]\ , \\ \frac{f_0}{\alpha_n \ldots \alpha_1}\ , \text{ for } x \in [-\frac{\alpha_1 \ldots \alpha_n}{2f_0}, \frac{\alpha_1 \ldots \alpha_n}{2f_0}]\ . \end{cases}$$

For any $A > \frac{\alpha_1 \cdots \alpha_n}{2f_0}$, the integral in (27) has value 1. This is obvious since the P_n map normalised distributions to normalised distributions. Since $\frac{f_0}{\alpha_n \ldots \alpha_1} \to \infty$ as $n \to \infty$, this suggests that $\{f_n\}_{n=1}^{\infty}$ converges to $\delta(x)$. To close, note that (i) may be re-expressed as

$$x_{n+1} = x_n + \gamma_n \bar{h}'(\hat{x}) x_n\ ,$$

giving the associated ODE

$$\dot{z} = z\bar{h}'(\hat{x})\ , \tag{28}$$

where $\bar{h}'(\hat{x}) < 0$. We will return to this ODE in Section 4, where, in addition, we shall give an interpretation of the above results.

3.2 Algorithm (ii)

It is difficult to determine the Frobenius-Perron operator for algorithm (ii) directly from the definition. Instead, we will follow the procedure used in [6] and [8], which uses the Laplace transform to study this operator. Let $F_n(p)$ be the Laplace transform of the pdf of the x_n, i.e.

$$\begin{aligned} F_n(p) &= (\mathcal{L} f_n)(p) \\ &= \int_0^{\infty} f_n(x) e^{-px} dx \\ &= \int_0^{\infty} P_n f(x) e^{-px} dx \\ &= (\mathcal{L} P_n f)(p)\ , \end{aligned}$$

and $G(p)$ be the Laplace transform of the pdf of the ξ_n. Then, $F_n(\alpha_n p)$ is the transform of the pdf of the $\alpha_n x_n$, and $G(\gamma_n p)$ is the transform of the pdf of the $\gamma_n \xi_n$. Following [6] and [8], we then have that

$$F_{n+1}(p) = F_n(\alpha_n p) G(\gamma_n p)\ . \tag{29}$$

Now, (29) may be expressed in closed form as

$$F_n(p) = F_0\left(p\prod_{j=0}^{n-1}\alpha_j\right)\prod_{k=0}^{n-1}G\left(\gamma_k p\prod_{i=k+1}^{n-1}\alpha_i\right) \ . \tag{30}$$

Assume that the pdf of the x_0 is given by the Dirac delta function

$$f_{x_0}(z) = \delta(z - x_0) \ .$$

Then the pdf of the $\prod_{j=0}^{n-1}\alpha_j x_0$ is

$$\left(\prod_{j=0}^{n-1}\alpha_j\right)^{-1}\delta\left(\frac{z - x_0\prod_{j=0}^{n-1}\alpha_j}{\prod_{j=0}^{n-1}\alpha_j}\right) \ . \tag{31}$$

The Laplace transform of (31) is then

$$\int_0^\infty e^{-px}\left(\prod_{j=0}^{n-1}\alpha_j\right)^{-1}\delta\left(\frac{x}{\prod_{j=0}^{n-1}\alpha_j} - x_0\right)dx = e^{-px_0\prod_{j=0}^{n-1}\alpha_j} \ . \tag{32}$$

Equations (30) and (32) give

$$F_0\left(p\prod_{j=0}^{n-1}\alpha_j\right) = e^{-px_0\prod_{j=0}^{n-1}\alpha_j} \ . \tag{33}$$

Whence, (30) reduces to

$$F_n(p) = e^{-px_0\prod_{j=0}^{n-1}\alpha_j}\prod_{k=0}^{n-1}G\left(\gamma_k p\prod_{i=k+1}^{n-1}\alpha_i\right) \ . \tag{34}$$

We shall choose a pdf of $\gamma_n\xi_n$ such that the product in (34) is easily computed. Following [6], one possible form, *which allows Gaussian and other more general distributions*, has Laplace transform

$$G(p) = e^{\beta p^\eta} \ , \ \beta \in \mathbb{R} \ , \ \beta < 0 \ , \tag{35}$$

which yields

$$G\left(\gamma_k p\prod_{i=k+1}^{n-1}\alpha_i\right) = e^{\{\beta\gamma_k^\eta p^\eta\left(\prod_{i=k+1}^{n-1}\alpha_i\right)^\eta\}} \ .$$

Therefore, assuming that (35) holds, we get

$$\prod_{k=0}^{n-1} G\left(\gamma_k p \prod_{i=k+1}^{n-1} \alpha_i\right) = \exp\{\beta p^\eta \sum_{k=0}^{n-1} \gamma_k^\eta \left(\prod_{i=k+1}^{n-1} \alpha_i\right)^\eta\} . \tag{36}$$

Substituting (36) back into (34) yields

$$F_n(p) = \exp\{-px_0 \prod_{j=0}^{n-1} \alpha_j + \beta p^\eta \sum_{k=0}^{n-1} \gamma_k^\eta \left(\prod_{i=k+1}^{n-1} \alpha_i\right)^\eta\} . \tag{37}$$

The density of x_n is thus given by

$$f_n(x) = \mathcal{L}^{-1}\{\exp\{-px_0 \prod_{j=0}^{n-1} \alpha_j + \beta p^\eta \sum_{k=0}^{n-1} \gamma_k^\eta \left(\prod_{i=k+1}^{n-1} \alpha_i\right)^\eta\}\} , \tag{38}$$

where $\mathcal{L}^{-1}$ denotes the inverse Laplace transform. For $\eta = 2$, equation (38) becomes

$$\begin{aligned} f_n(x) &= 0.5 \left\{ \frac{\pi}{\beta \sum_{k=0}^{n-1} \gamma_k^2 \left(\prod_{i=k+1}^{n-1} \alpha_i\right)^2} \right\}^{\frac{1}{2}} U\left(x - x_0 \prod_{j=0}^{n-1} \alpha_j\right) \times \\ &\quad \exp\left\{ -\frac{\left(x - x_0 \prod_{j=0}^{n-1} \alpha_j\right)^2}{\beta \sum_{k=0}^{n-1} \gamma_k^2 \left(\prod_{i=k+1}^{n-1} \alpha_i\right)^2} \right\} , \end{aligned} \tag{39}$$

where U is the Heaviside step function. It is straightforward to show that

$$\prod_{j=0}^{n-1} \alpha_j \to 0 \ \text{ as } \ n \to \infty . \tag{40}$$

Let us now examine the quantities in the denominator of (39), viz.

$$T_n \stackrel{def}{=} \beta \sum_{k=0}^{n-1} \gamma_k^2 \left(\prod_{i=k+1}^{n-1} \alpha_i\right)^2 . \tag{41}$$

It may be shown that

$$\left(\prod_{i=k+1}^{n-1} \alpha_i\right)^2 \sim \exp\left\{\frac{2b}{1-\alpha}[(n-1)^{1-\alpha} - (k+1)^{1-\alpha}]\right\} ,$$

where we have set $\gamma_i \stackrel{def}{=} i^{-\alpha}$, $\alpha \in (0,1)$, and $b \stackrel{def}{=} \bar{h}'(\hat{x}) < 0$. Consequently, we have that

$$\begin{aligned} T_n &\sim \beta \sum_{k=0}^{n-1} \gamma_k^2 \exp\left\{ \frac{2b}{1-\alpha}[(n-1)^{1-\alpha} - (k+1)^{1-\alpha}] \right\} \\ &= \beta \exp\left\{ \frac{2b}{1-\alpha}(n-1)^{1-\alpha} \right\} \sum_{k=0}^{n-1} \gamma_k^2 \exp\left\{ -\frac{2b}{1-\alpha}(k+1)^{1-\alpha} \right\} . \end{aligned}$$

The leading behaviour of the partial sum on the RHS of the preceding equation is found by examining the corresponding integral

$$I_n \stackrel{def}{=} \int_0^{n-1} (x+1)^{-2\alpha} \exp\left\{ -\frac{2b}{1-\alpha}(x+1)^{1-\alpha} \right\} dx ,$$

where, without loss of generality, we have set $\gamma_k \stackrel{def}{=} (k+1)^{-\alpha}$. Integrating I_n by parts yields

$$I_n \sim -\frac{n^{-\alpha}}{2b} \exp\left\{ -\frac{2b}{1-\alpha} n^{1-\alpha} \right\} ,$$

which implies that

$$\begin{aligned} T_n &\sim \beta \exp\left\{ \frac{2b}{1-\alpha}(n-1)^{1-\alpha} \right\} \times \left\{ -\frac{n^{-\alpha}}{2b} \exp\left(-\frac{2b}{1-\alpha} n^{1-\alpha} \right) \right\} \\ &= -\frac{\beta}{2b} n^{-\alpha} \longrightarrow 0 \quad \text{as} \quad n \to \infty . \end{aligned} \tag{42}$$

We may express (39) as

$$f_n(x) = \begin{cases} 0.5 \times \left\{ \frac{\pi}{T_n} \right\}^{\frac{1}{2}} \times \exp\left\{ -\frac{\left(x - x_0 \prod_{j=0}^{n-1} \alpha_j \right)^2}{T_n} \right\} & , \; x > x_0 \prod_{j=0}^{n-1} \alpha_j \\ 0.5 \times \{ \frac{\pi}{T_n} \}^{0.5} & , \; x = x_0 \prod_{j=0}^{n-1} \alpha_j \\ 0 & , \; x < x_0 \prod_{j=0}^{n-1} \alpha_j \end{cases}$$

Furthermore, for $x > x_0 \prod_{j=0}^{n-1} \alpha_j$, equation (42) implies that

$$\lim_{n \to \infty} f_n(x) = 0 .$$

On the other hand, when $x = x_0 \prod_{j=0}^{n-1} \alpha_j$, we have that

$$\lim_{n\to\infty} f_n(x) = \infty \ .$$

The above analysis suggests that $\{f_n\}_{n=1}^{\infty}$ converges, in the sense of distributions, to $\delta(x)$.

3.3 Closing remarks

Algorithm (ii) may be expressed in the form

$$x_{n+1} = x_n + \gamma_n \bar{h}'(\hat{x}) x_n + \gamma_n \xi_n \ ,$$

giving the associated ODE

$$\dot{z} = z\bar{h}'(\hat{x}) \ ,$$

where $\bar{h}'(\hat{x}) < 0$. It is straightforward to show that the ODE solution is given by

$$z(t) = z(0)e^{t\bar{h}'(\hat{x})} \to 0 \ , \quad \text{as} \ \ t \to \infty \ ,$$

thus $\hat{z} = 0$ is stable equilibrium of this ODE. Therefore from the results in Sections 3.1-3.2, we conclude that the dynamical behaviours of the ODE solutions and the sequence of densities are in accord for both algorithms (i) and (ii). For these algorithms, the stationary density of $\{x_n\}_{n=1}^{\infty}$ is concentrated about $\hat{z} = 0$, the stable fixed point of the associated ODE. Note that the presence of the random perturbation in algorithm (ii) seems to have little effect on the long term behaviour.

We have derived stationary densities of linearised algorithms of the general form given by algorithm (ii). Note that the stationary densities so derived are *localised*, about $\hat{x}$. To understand the full stationary density of $\{x_n\}_{n=1}^{\infty}$, one needs to piece together the information from the various local stationary densities which correspond to different stable equilibria of the associated ODE. In algorithm (ii), if we assume that ξ_n is characterised by a pdf of the form

$$G(p) = e^{\beta p^2} \ , \quad \beta \in \mathbb{R} \ ,$$

then we obtain $\delta(x)$ as the local stationary density. However, for the full stationary density, it is possible that different masses may be associated with each of the stable equilibria. The problem that remains is to quantify masses that are associated with each localised $\delta(x)$ distribution. If achieved, this would give an answer to the question: *To which one of the stable equilibria is the algorithm most likely to converge?* One way to approach this problem is through analysis of the Frobenius-Perron operator of the nonlinear algorithm directly. The next section will provide a theoretical framework for this analysis. Section 5 will show how this analysis may be applied to a specific algorithm, with the aid of numerical computations.

4 Densities of Nonlinear Algorithms

Consider an algorithm in the form

$$x_{n+1} = S_n(x_n) + \gamma_n \xi_n \; , \tag{43}$$

where S explicitly depends on n (refer to equation (2) to see an example of such an S). As usual, the noise amplitude is damped and $\{\xi_n\}$ is an *i.i.d.* sequence of random variables with a common density g. We suppose that the Frobenius-Perron operator associated with $S_n(\cdot)$ is P_n. In addition, we denote the noise term by

$$z_n = \gamma_n \xi_n \; . \tag{44}$$

The density of z_n is given by

$$G_n(x) = \frac{1}{\gamma_n} g(\frac{x}{\gamma_n}) \; . \tag{45}$$

Suppose that $f_n \in L^1_D(X)$ is the density of x_n. From (43), we see that x_{n+1} is the sum of two independent random variables. Note that $S_n(x_n)$ and z_n are independent since in calculating $x_1, ..., x_n$, we only need $\xi_0, ..., \xi_{n-1}$. Let $w : \mathbb{R} \to \mathbb{R}$ be an arbitrary, bounded, measurable function. Then, the mathematical expectation of $w(x_{n+1})$ is

$$E[w(x_{n+1})] = \int_{\mathbb{R}} w(x) f_{n+1}(x) dx \; . \tag{46}$$

Furthermore, using (43) and the fact that the joint density of (x_n, z_n) is $\frac{1}{\gamma_n} f_n(x) g(\frac{z}{\gamma_n})$, we have that

$$E[w(x_{n+1})] = \frac{1}{\gamma_n} \int_{\mathbb{R}} \int_{\mathbb{R}} w(S_n(y) + z) f_n(y) g(\frac{z}{\gamma_n}) dy dz \; . \tag{47}$$

Now using the change of variables $x = S_n(y) + z$, $y = y$, the integral in (47) is transformed to

$$E[w(x_{n+1})] = \frac{1}{\gamma_n} \int_{\mathbb{R}} \int_{\mathbb{R}} w(x) f_n(y) g(\frac{x - S_n(y)}{\gamma_n}) dx dy \ . \tag{48}$$

Equating (46) and (48) yields

$$\int_{\mathbb{R}} w(x) f_{n+1}(x) = \int_{\mathbb{R}} w(x) \left\{ \frac{1}{\gamma_n} \int_{\mathbb{R}} f_n(y) g(\frac{x - S_n(y)}{\gamma_n}) dy \right\} dx \ ,$$

which gives

$$f_{n+1}(x) = \frac{1}{\gamma_n} \int_{\mathbb{R}} f_n(y) g(\frac{x - S_n(y)}{\gamma_n}) dy \ . \tag{49}$$

From (49), define the operator $\overline{P}_n : L^1_D \to L^1_D$ by

$$\overline{P}_n f(x) = \frac{1}{\gamma_n} \int_{\mathbb{R}} f(y) g(\frac{x - S_n(y)}{\gamma_n}) dy \ , \tag{50}$$

for $f \in L^1_D$. Suppose that $S_n(\cdot)$ is non-singular. Therefore the Frobenius-Perron and Koopman operators, P_n and U_n respectively, corresponding to $S_n(\cdot)$ exist. Furthermore, let

$$h_{n,x}(y) \stackrel{def}{=} g(\frac{x - y}{\gamma_n}) \ . \tag{51}$$

Then (50) and (51) yield

$$\begin{aligned} \overline{P}_n f(x) &= \frac{1}{\gamma_n} \int_{\mathbb{R}} f(y) h_{n,x}(S_n(y)) dy \\ &= \gamma_n^{-1} \langle f, U_n h_{n,x} \rangle \ , \text{ since } h_{n,x}(S_n(y)) = U_n h_{n,x}(y) \\ &= \gamma_n^{-1} \langle P_n f, h_{n,x} \rangle \ , \text{ since } P_n \text{ and } U_n \text{ are adjoint operators} \\ &= \frac{1}{\gamma_n} \int_{\mathbb{R}} g(\frac{x - y}{\gamma_n}) P_n f(y) dy \ . \end{aligned} \tag{52}$$

Using the change of variable $\frac{x-y}{\gamma_n} = t$, equation (52) becomes

$$\overline{P}_n f(x) = \int_{\mathbb{R}} g(t) P_n f(x - \gamma_n t) dt \ . \tag{53}$$

Now, suppose that P_n is such that

$$P_n \overset{\|\ \|_1}{\to} P \text{ as } n \to \infty \ , \tag{54}$$

where P is a limiting operator. With this in mind, the following issue arises: *Does it follow that* $\overline{P}_n \overset{\|\ \|_1}{\to} P$ *as* $n \to \infty$*?* This issue is addressed by Theorem 4.1 below which is proved with the aid of the following lemma [12].

Lemma 4.1 *For every* $f \in L^1$, $I \subseteq \mathbb{R}$ *bounded or not,*

$$\lim_{h \to 0} \int_I |f(x+h) - f(x)| dx = 0 \ .$$

The proof of this result may be found in [12].

Theorem 4.1 *For the system defined by (43),*

$$lim_{n\to\infty} \|\overline{P}_n f - P_n f\|_1 = 0 \ , \ \text{for all } f \in L^1_D(X) \ ,$$

where P_n *is the Frobenius-Perron operator corresponding to* $S_n(\cdot)$ *and* $\overline{P}_n$ *is the operator corresponding to (43).*

Proof
Write

$$\overline{P}_n f(x) - P_n f(x) = \int_{\mathbb{R}} g(t)\{P_n f(x - \gamma_n t) - P_n f(x)\} dt \ .$$

Pick an arbitrarily small $\delta > 0$. Since g is an integrable function on $\mathbb{R}$, there must exist an $r > 0$ such that

$$\int_{|y| \geq r} g(y) dy \leq \frac{\delta}{4} \ . \tag{$*$}$$

To compute an upper bound for $\|\overline{P}_n f - P_n f\|_1$, we proceed as follows.

$$\begin{aligned} \|\overline{P}_n f - P_n f\|_1 &\leq \int_{\mathbb{R}} \int_{\mathbb{R}} g(y) |P_n f(x - \gamma_n y) - P_n f(x)| dx dy \\ &\overset{def}{=} I_1 + I_2 \ , \end{aligned}$$

where

$$I_1 \overset{def}{=} \int_{\mathbb{R}} \int_{|y| \leq r} g(y) |P_n f(x - \gamma_n y) - P_n f(x)| dy dx \ ,$$

and
$$I_2 \stackrel{def}{=} \int_{\mathbb{R}} \int_{|y|\geq r} g(y)|P_nf(x-\gamma_n y) - P_nf(x)|dydx \ .$$
We consider each of these integrals in turn. First, since P_nf is integrable, Lemma 4.1 implies that there exists an $N > 0$ such that
$$\int_{\mathbb{R}} |P_nf(x-\gamma_n y) - P_nf(x)|dx < \frac{\delta}{2} \ , \text{ for all } n \geq N \text{ and for } |y| \leq r \ .$$
Hence
$$I_1 \leq \frac{\delta}{2}\int_{|y|\leq r} g(y)dy \leq \frac{\delta}{2}\int_{\mathbb{R}} g(y)dy = \frac{\delta}{2} \ . \tag{55}$$
For I_2, we use the triangle inequality to write
$$I_2 \leq \int_{\mathbb{R}} \int_{|y|\geq r} g(y)P_nf(x)dydx + \int_{\mathbb{R}} \int_{|y|\geq r} g(y)P_nf(x-\gamma_n y)dydx \ . \tag{56}$$
Using the change of variables $v = y$ and $z = x - \gamma_n y$, one obtains
$$\begin{aligned} \int_{\mathbb{R}} \int_{|y|\geq r} g(y)P_nf(x-\gamma_n y)dydx &= \int_{\mathbb{R}} \int_{|v|\geq r} g(v)P_nf(z)dvdz \\ &= \int_{|v|\geq r} g(v)dv \times \int_{\mathbb{R}} P_nf(z)dz \\ &= \int_{|v|\geq r} g(v)dv \ \ (\text{since } \int_{\mathbb{R}} P_nf(z)dz = 1) \\ &\leq \frac{\delta}{4} \ (\text{by assumption } (*) \text{ above}) \ , \qquad (57) \end{aligned}$$
and
$$\begin{aligned} \int_{\mathbb{R}} \int_{|y|\geq r} g(y)P_nf(x)dydx &= \int_{|y|\geq r} g(y)dy \times \int_{\mathbb{R}} P_nf(x)dx \\ &\leq \frac{\delta}{4} \ . \end{aligned} \tag{58}$$
Hence, combining (56), (57) and (58) gives
$$I_2 \leq \frac{\delta}{2} \ . \tag{59}$$
Therefore, (55) and (59) imply that
$$\lim_{n\to\infty} \|\overline{P}_nf - P_nf\|_1 = 0 \ .$$
□

A Corollary of the preceding theorem is stated and proved below.

Corollary 4.1 *Suppose that* $S_n(\cdot)$ *and* g *are given and that we have a sequence* $\{f_n\}$ *generated by* $f_{n+1} = \overline{P}_n f_n$ *and such that* $f_n \to f_*$ *as* $n \to \infty$*. In other words, given an arbitrary* $\epsilon > 0$*, there exists an* N *such that*

$$\|f_n - f_*\|_1 < \frac{\epsilon}{2}\ , \ \textit{for all}\ n > N\ . \tag{60}$$

Furthermore, suppose that P_n*, the Frobenius-Perron operator associated with* $S_n(\cdot)$*, is such that* $P_n \to P$ *as* $n \to \infty$*, where* P *is a limiting operator. Then* f_* *is a stationary density for* P*, viz.* $Pf_* = f_*$*.*

Proof of Corollary
Write

$$\overline{P}_n f_* = f_{n+1} + \overline{P}_n (f_* - f_n)\ .$$

This gives

$$\begin{aligned}\|\overline{P}_n f_* - f_*\|_1 &\leq \|f_{n+1} - f_*\|_1 + \|\overline{P}_n (f_* - f_n)\|_1 \\ &\leq \|f_{n+1} - f_*\|_1 + \|f_* - f_n\|_1 \\ &\leq \epsilon\ , \text{ for all } n > N\ ,\end{aligned}$$

which implies that

$$\|\overline{P}_n f_* - f_*\|_1 \to 0 \text{ as } n \to \infty. \tag{61}$$

Now, from (54) and Theorem 4.1, we get that

$$\|P_n f_* - P f_*\|_1 \to 0 \ \text{ and } \ \|\overline{P}_n f_* - P_n f_*\|_1 \to 0 \text{ as } n \to \infty\ .$$

This leads to

$$\|\overline{P}_n f_* - P f_*\|_1 \to 0 \text{ as } n \to \infty\ ,$$

which, in combination with (61), gives $Pf_* = f_*$. □

To summarise, we have shown how to calculate the Frobenius-Perron operator $\overline{P}_n$ of the nonlinear, nonautonomous random algorithm (43), see equation (50). We further showed that if P_n, the Frobenius-Perron operator of the nonautonomous map $S_n(\cdot)$, converges to P as $n \to \infty$, then

1. $\overline{P}_n$ also converges to P as $n \to \infty$
2. the limit of a convergent sequence of densities generated by $\overline{P}_n$ is a stationary density of P

Unfortunately, this result tells us nothing about a specific f_*. In fact, in the particular case given by (43), f_* is not unique since P is the Frobenius-Perron operator associated with the identity map I.

In the next section we will show how to apply this theory to a specific example of algorithm (43). As the analysis is not tractable, we will use numerical computations to study the convergence of the densities generated by the Frobenius-Perron operator.

5 Particular example for which the associated ODE has multiple equilibria

5.1 Introduction

Consider the deterministic non-autonomous algorithm

$$x_{n+1} = x_n + \gamma_n \bar{h}(x_n) \stackrel{def}{=} S_n(x_n) \ , \tag{62}$$

where we assume that

$$\bar{h}(x_n) = -x_n(x_n^2 + 2x_n + 0.5) \ .$$

The associated ODE is thus given by

$$\dot{x} = -x(x^2 + 2x + 0.5) \stackrel{def}{=} g(x) \ .$$

This ODE has two stable equilibria, $x_1^* = 0$ and $x_3^* = -1 - \frac{1}{2}\sqrt{2} \approx -1.707$, and a single unstable equilibrium, viz. $x_2^* = -1 + \frac{1}{2}\sqrt{2} \approx -0.293$. The multipliers of the stable equilibria are given by $g'(0) = -0.5$ and $g'(-1.707) = -1 - \sqrt{2} \approx -2.414$, implying that -1.707 is the relatively more attractive of the two. The graph depicting (62) is shown in Fig 2.

The following question arises: *to which one of x_1^* or x_3^* is $\{x_n\}$ likely to converge as $n \to \infty$, given an arbitrary starting value x_0?*
Note that with the given $\bar{h}$, we have that $S_n : [0, x_3^*] \to [0, x_3^*]$. Thus, we focus on the dynamics inside the box outlined in Fig 2. The Frobenius-Perron operator associated with the transformation S_n is formulated below. Pick an interval $[x, 0] \subset [-1.707, 0]$ so that the counter-image of $[x, 0]$ under

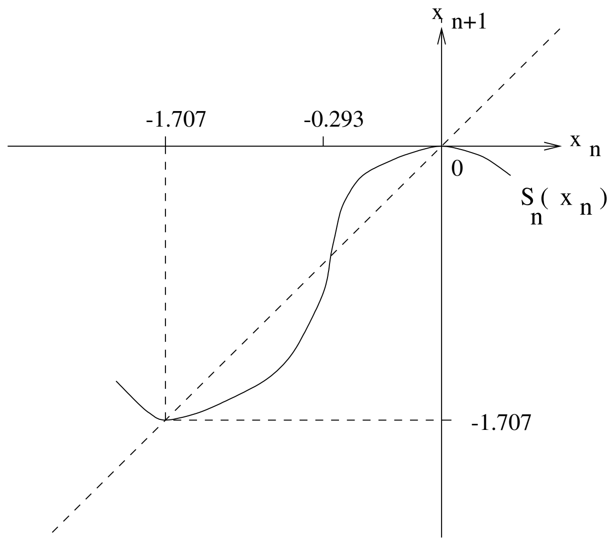

Figure 2: x_{n+1} with a cubic $\bar{h}$

S_n is given by

$$S_n^{-1}([x,0]) = [g_{1,n}(x),0] \cup [g_{2,n}(x), -1+(2\gamma_n)^{-0.5}\sqrt{2+\gamma_n}] \cup [g_{3,n}(x), -1-(2\gamma_n)^{-0.5}\sqrt{2+\gamma_n}] \;, \tag{63}$$

where $g_{i,n}(x)$, $i = 1...3$ are the three solutions of the equation $x = S_n(y)$. Following [12], the Frobenius-Perron operator corresponding to S_n is defined by

$$P_n f(x) = \frac{d}{dx}\int_{S_n^{-1}([x,0])} f(u)du = \sum_{j=1}^{3} f(g_{j,n}(x))g'_{j,n}(x) \;. \tag{64}$$

This tells us how S_n transforms a given density f into a new one $P_n f$. To see how (64) works, pick an initial uniform density $f(x) = 2/(2+\sqrt{2}) \approx 0.586$ for $x \in [-1.707, 0]$, and zero elsewhere. Then (64) becomes

$$P_n f(x) = 0.586 \sum_{j=1}^{3} g'_{j,n}(x) \;, \text{ for } x \in [-1.707, 0] \;. \tag{65}$$

Substituting this expression for $P_n f$ in place of f on the RHS of (64) yields

$$P_2P_1f(x) = 0.586\sum_{k=1}^{3}\left\{\sum_{j=1}^{3} g'_{j,1}(g_{k,2}(x))\right\} g'_{k,2}(x)\ . \tag{66}$$

Similarly, we obtain that

$$P_3P_2P_1f(x) = 0.586\sum_{l=1}^{3}\sum_{k=1}^{3}\sum_{j=1}^{3} g'_{j,1}(g_{k,2}(g_{l,3}(x)))g'_{k,2}(g_{l,3}(x))g'_{l,3}(x)\ .$$

Given an initial density $f \in L^1_D$, we seek to determine $\lim_{n\to\infty} P_n ... P_1 f$. This limit, if it exists, gives the sought-after stationary density. With this in mind, we return to the cubic equation $x = S_n(y)$, which gives

$$y^3 + 2y^2 + \frac{1}{2\gamma_n}(\gamma_n - 2)y + \frac{x}{\gamma_n} = 0\ . \tag{67}$$

We wish to find the three roots of (67), viz. $g_{1,n}(x)$, $g_{2,n}(x)$, and $g_{3,n}(x)$. We may write (67) as

$$y^3 + a_1y^2 + a_2y + a_3 = 0\ ,$$

where $a_1 = 2$, $a_2 = \frac{1}{2\gamma_n}(\gamma_n - 2)$,and $a_3 = \frac{x}{\gamma_n}$. Let $Q = \frac{1}{9}(3a_2 - a_1^2)$, $R = \frac{1}{54}(9a_1a_2 - 27a_3 - 2a_1^3)$, $S = (R + \sqrt{(Q^3 + R^2)})^{\frac{1}{3}}$, and $T = (R - \sqrt{(Q^3 + R^2)})^{\frac{1}{3}}$. Furthermore, let $D = Q^3 + R^2$ be the discriminant. Now, it may be shown that, for $x \in [-1.707, 0]$, we have that $D < 0$ if and only if

$$(1458x + 972 + 378\gamma_n)^2 < (62.5 + \frac{225}{\gamma_n} + \frac{270}{\gamma_n^2} + \frac{108}{\gamma_n^3})\ .$$

Clearly, for any fixed $x \in [-1.707, 0]$, the LHS of this inequality approaches a fixed positive real number as $n \to \infty$. On the other hand, the RHS of the inequality approaches $+\infty$ as $n \to \infty$. Therefore, there will be a "cross-over" value, γ_N say, for which the above inequality holds for all $n \geq N$. From the definitions of Q and R above, we have that

$$\frac{R}{\sqrt{-Q^3}} = \beta_n(7\gamma_n + 18 + 27x)\ ,$$

where

$$\beta_n \stackrel{def}{=} -0.5\sqrt{\frac{\gamma_n}{(2.5\gamma_n + 3)^3}} \to 0^- \text{ as } n \to \infty. \tag{68}$$

Since there exists an N such that $D < 0$ (for $x \in [-1.707, 0]$) for all $n \geq N$, it is well-known [1] that all the roots of (67) are real and distinct for all $n \geq N$. These roots are given by

$$g_{1,n}(x) \stackrel{def}{=} \frac{2}{3}\sqrt{\frac{6+5\gamma_n}{2\gamma_n}}\cos(\frac{\theta}{3}) - \frac{2}{3} , \tag{69}$$

$$g_{2,n}(x) \stackrel{def}{=} \frac{2}{3}\sqrt{\frac{6+5\gamma_n}{2\gamma_n}}\cos(\frac{\theta}{3} + \frac{2\pi}{3}) - \frac{2}{3} , \tag{70}$$

and

$$g_{3,n}(x) \stackrel{def}{=} \frac{2}{3}\sqrt{\frac{6+5\gamma_n}{2\gamma_n}}\cos(\frac{\theta}{3} + \frac{4\pi}{3}) - \frac{2}{3} , \tag{71}$$

where

$$\theta \stackrel{def}{=} \arccos\{\beta_n(7\gamma_n + 18 + 27x)\} . \tag{72}$$

It may be shown that only $g_{3,n}(x)$ maps the closed interval $[-1.707, 0]$ to itself, for all n. Hence, $g_{3,n}(x)$ is the only root we need to use in evaluating the Frobenius-Perron operator corresponding to S_n. From (64), the sought-after operator is given by

$$P_n f(x) = f(g_{3,n}(x))g'_{3,n}(x) , \tag{73}$$

for an arbitrary initial density $f \in L^1_D$. We choose an initial pdf given by: $f(x) = 0.586$ for $x \in [-1.707, 0]$, and zero elsewhere. Then (73) becomes

$$P_n f(x) = 0.586 g'_{3,n}(x) . \tag{74}$$

Hence we get the following iterates

$$P_2 P_1 f(x) = 0.586 g'_{3,1}(g_{3,2}(x))g'_{3,2}(x) ,$$

and

$$P_3 P_2 P_1 f(x) = 0.586 g'_{3,1}(g_{3,2}(g_{3,3}(x)))g'_{3,2}(g_{3,3}(x))g'_{3,3}(x) .$$

In a similar way, higher iterates of $P_n ... P_1 f(x)$ may be computed. Now, analytically determining $\lim_{n\to\infty} P_n ... P_1 f(x)$ is intractable. However, a numerical approach is feasible. Fig 3 shows $P_1 f(x)$, $P_2 P_1 f(x)$, and $P_4 P_3 P_2 P_1 f(x)$ respectively, for $x \in [-1.707, 0]$. The learning parameter used is $\gamma_n \stackrel{def}{=} (n+3)^{-0.99}$. The plots suggest that $P_n ... P_1 f(x)$ converges, in the distribu-

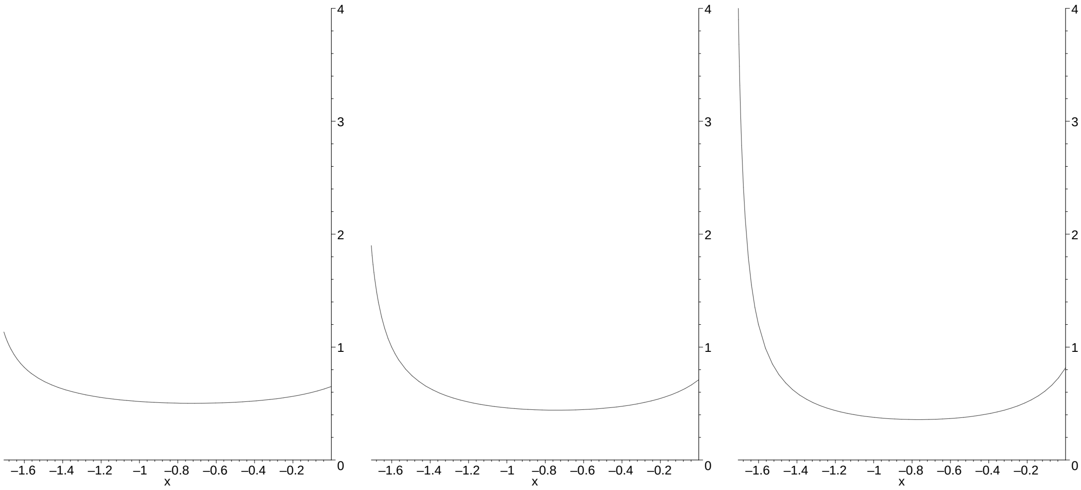

Figure 3: (a) $P_1 f(x)$ (b) $P_2 P_1 f(x)$ (c) $P_4 P_3 P_2 P_1 f(x)$

tional sense, to two Dirac distributions centred at the two stable equilibria as $n \to \infty$. It should be noted that these two Dirac distributions emerge at different rates. The one located at -1.707 has a relatively bigger mass compared to the one located at the origin. This distribution of mass is clearly related to the magnitudes of the multipliers of the two stable equilibria. However, it is not clear how one might go about actually quantifying these masses. Finally, note that these numerical results validate the idea of linearising $\bar{h}(x)$ about each stable point before deriving the local densities, as previously done.

5.2 A stationary density for the perturbed operator $\overline{P}_n$?

This section investigates the dynamical behaviour of the perturbed non-autonomous system

$$x_{n+1} = S_n(x_n) + \gamma_n \xi_n \ , \tag{75}$$

where $S_n(x_n) \stackrel{def}{=} x_n + \gamma_n \bar{h}(x_n)$ and $\bar{h}(x_n)$ is as defined in the previous example. As usual, assume that $\{\xi_n\}$ is a sequence of *i.i.d.* random variables, each with density g. The density of the random variable $\gamma_n \xi_n$ is $\gamma_n^{-1} g(\gamma_n^{-1} x)$. The

Frobenius-Perron operator associated with (75) is given by [12]

$$\overline{P}_n f(x) = \gamma_n^{-1} \int_{\mathbb{R}} f(y) g(\frac{x - S_n(y)}{\gamma_n}) dy \ , \tag{76}$$

where $f \in L^1_D$ is an arbitrary initial density of all the possible initial states of (75). From (76), we may find $\overline{P}_1 f$, $\overline{P}_2 \overline{P}_1 f$,...,$\overline{P}_n \overline{P}_{n-1} ... \overline{P}_1 f$. The issue is to characterise

$$\lim_{n \to \infty} \overline{P}_n \overline{P}_{n-1} ... \overline{P}_1 f \ , \tag{77}$$

provided that such a limit exists. Now assume that

$$g(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \ , \tag{78}$$

a mean zero and unit variance Gaussian distribution. Furthermore, assume that the initial density f is given by

$$f(x) = \begin{cases} \frac{1}{3} & , \quad x \in [-2, 1] \\ 0 & , \quad \text{elsewhere} \ . \end{cases} \tag{79}$$

5.2.1 Computation of $\overline{P}_n \overline{P}_{n-1} ... \overline{P}_1 f$

The analytic computation of the iterates $\overline{P}_n ... \overline{P}_1 f$, for $n > 1$, is not usually possible. However, a numerical approach to this computation is feasible. Nonetheless, the latter approach is not that straightforward and easy, as it involves the evaluation of iterated integrals of increasing complexity. We now outline one possible way of performing the numerical computations of iterates of the operator $\overline{P}_n$. First, compute

$$\overline{P}_1 f(x) = \gamma_1^{-1} \int_{A_1} f(y) g(\frac{x - S_1(y)}{\gamma_1}) dy \ , \tag{80}$$

where $A_1 = [-15, 15] \subset \mathbb{R}$. Then approximate (80) by a polynomial, using the MAPLE *interp* function. This polynomial has a finite support. The choice of the size of this support is arbitrary. Denote this polynomial fit by $w_1(x)$. Next, compute

$$\overline{P}_2 f(x) = \gamma_2^{-1} \int_{A_1} f(y) g(\frac{x - S_2(y)}{\gamma_2}) dy \ ,$$

which gives

$$\overline{P}_2\overline{P}_1 f(x) \approx \gamma_2^{-1}\int_{A_1} w_1(x) g(\frac{x - S_2(y)}{\gamma_2}) dy \; . \tag{81}$$

In a similar way, compute $\overline{P}_3\overline{P}_2\overline{P}_1 f$, $\overline{P}_4\overline{P}_3\overline{P}_2\overline{P}_1 f$, etc. Even after the polynomial fits have been determined, the (symbolic) numerical integration of iterates of the perturbed operator is still relatively intensive and consumes a substantial amount of CPU time. Figs 4-5 show the first four of these iterates. The learning parameter used is $\gamma_n = (n+4)^{-0.999}$.

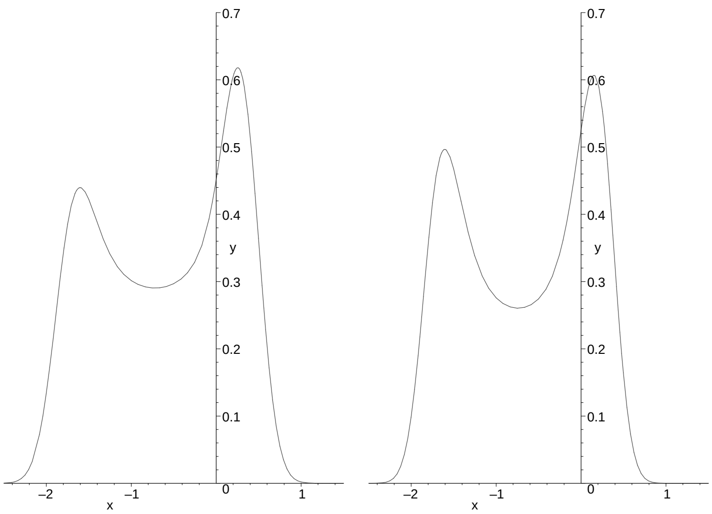

Figure 4: (a) $\overline{P}_1 f(x)$ (b) $\overline{P}_2\overline{P}_1 f(x)$

5.3 Comments on numerics

The plots in Figs 4-5 suggest that the stationary density of the sequence of perturbed operators consists of two Dirac distributions, μ_1 and μ_2 say, centred at $\bar{x}_1 = -1.707$ and $\bar{x}_2 = 0$ respectively, the two locally asymptotically stable equilibria of the associated ODE. We may express the stationary density as $\mu \overset{def}{=} \mu_1 + \mu_2$, where the component Dirac masses are

$$m_1 \overset{def}{=} \sup_{\psi \in D(X)} \left\{ \int_{-\infty}^{\infty} \psi(t)\mu_1(t - \bar{x}_1) dt \right\} = \sup_{\psi \in D(X)} \{\psi(\bar{x}_1)\} \; ,$$

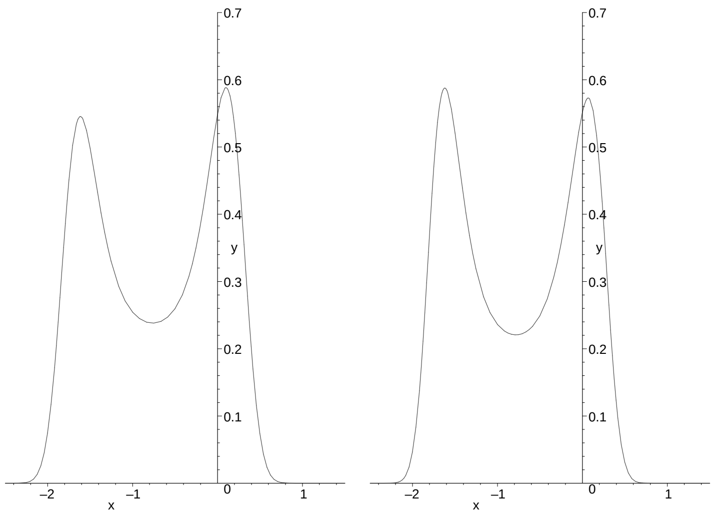

Figure 5: (a) $\overline{P}_3\overline{P}_2\overline{P}_1 f(x)$ (b) $\overline{P}_4\overline{P}_3\overline{P}_2\overline{P}_1 f(x)$

and

$$m_2 \stackrel{def}{=} \sup_{\psi \in D(X)} \left\{ \int_{-\infty}^{\infty} \psi(t)\mu_2(t - \bar{x}_2)dt \right\} = \sup_{\psi \in D(X)} \{\psi(\bar{x}_2)\} ,$$

such that $m_1 + m_2 = 1$, and where $D(X)$ is a suitable space of test functions, for example $D(X) \stackrel{def}{=} Lip_1^+(X)$. In this example, it is not immediately obvious what the values of m_1 and m_2 will be, primarily as a result of the difficulty of constructing the set of functions $\{\psi | \ \psi \in Lip_1^+(X)\}$. Also, as previously discussed in Section 2.4, analysis in the space $\overline{D}'_+(X)$ is difficult. Finally, the result in Section 5.2 is in accord with the one found in Section 5.1, thereby justifying the idea of linearising $\bar{h}$ about a stable equilibrium prior to performing a local analysis of the stationary density. The local stationary densities obtained before, via the linearisation procedure, are Dirac distributions centred at the stable equilibria of the associated system of ODEs.

6 Discussion and Conclusions

This paper has primarily been concerned with a systematic study of densities of both linear and nonlinear algorithms (see Sections 3 and 4, respectively). Our results suggest that densities of the linearised algorithms studied here do reproduce the same local characteristics as the densities of the nonlinear equations from which they are derived.

We recall that Mackey and Lasota [12] studied algorithms of the form

$$x_{n+1} = S(x_n) + \epsilon\xi_n \ , \quad 0 < \epsilon \ll 1 \ ,$$

where $S(x_n)$ is *not* explicitly dependent on n, and where $\{\xi_n\}$ is a sequence of *i.i.d.* random variables. Our work in this direction is new, because we look at more general algorithms, i.e.

$$x_{n+1} = S_n(x_n) + \gamma_n\xi_n \ ,$$

where now $S_n(x_n)$ is explicitly dependent on n, and $\{\gamma_n\}$ is a decreasing-to-zero sequence of positive real numbers such that $\sum_n \gamma_n = \infty$. Thus, our proof of Theorem 4.1 is new, and more general than that in [12]. Finally, it is important to emphasize that the numerical experiments performed in this paper were indispensable, providing insights into a difficult problem.

7 Acknowledgements

ERV and SAC acknowledge partial financial support in the form of Individual Operating Grants from the Natural Sciences and Engineering Research Council (NSERC) of Canada. SAC also acknowledges partial financial support from the Mathematics of Information Technology and Complex Systems (MITACS) programme. The research of IN was supported by the University of Waterloo's Faculty of Mathematics, the Department of Applied Mathematics as well as the grants to ERV and SAC listed above. Part of this work was carried out while IN was a Postdoctoral Fellow at York University.

8 References

[1] Abramowitz M. and Stegun I.A. (eds.) *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* Dover Publications Inc., New York (1965).

[2] A. Benveniste, M. Métivier, and P. Priouret *Adaptive Algorithms and Stochastic Approximations.* Springer-Verlag (1987).

[3] K. L. Chung *On a stochastic approximation method*, Ann. Math. Statist. (1954), **25**, 463-483.

[4] V. Fabian *On asymptotic normality in stochastic approximation*, The Annals of Mathematical Statistics (1968), Vol. 39, No. 4, 1327-1332.

[5] J. C. Fort and G. Pagès *Convergence of stochastic algorithms: From the Kushner-Clark theorem to the Lyapunov functional method*, Adv. Appl. Prob. (1996), **28**, 1072-1094.

[6] Halley J.M. and Iwasa Y. *Extinction Rate of a Population under both Demographic and Environmental Stochasticity*, Theoretical Population Biology (1998), **53**, No. 1, 1-15.

[7] Hutchinson J. *Fractals and Self-similarity*, Indiana University Journal of Mathematics (1981), **30**, 713-747.

[8] Karlin S. and Taylor H.P. *A First Course in Stochastic Processes*, 2nd ed., Academic Press, London (1975).

[9] H. J. Kushner and D. S. Clark *Stochastic Approximation Methods for Constrained and Unconstrained Systems.* Springer-Verlag, New York (1978).

[10] H. J. Kushner, and G. Yin *Stochastic Approximation Algorithms and Applications.* Springer-Verlag, New York (1997).

[11] L. Ljung *Analysis of Recursive Stochastic Algorithms*, IEEE Transactions on automatic Control (1977), **AC-22**, 551-575.

[12] C. M. Mackey and A. Lasota *Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics*, 2nd ed., Springer-Verlag (1994).

[13] I. Ncube *Stochastic Approximation of Artificial Neural Network-Type Learning Algorithms: A Dynamical Systems Approach.* Ph.D thesis, University of Waterloo, Canada (2001).

[14] H. Robbins and S. Monro *A stochastic approximation method*, Ann. Math. Statist. (1951), **22**, 400-407.

[15] Van Kampen N. G. *Stochastic Processes in Physics and Chemistry.* North-Holland (1992).

[16] Vrscay E. and Forte B. *Theory of generalised fractal transforms*, Fractal Image Encoding and Analysis (edited by Yuval Fisher), Springer-Verlag, New York (1996), 145-168.

[17] Yaglom A. M. *An Introduction to the Theory of Stationary Random Functions.* Prentice-Hall (1962).

[18] Zemanian A. H. *Distribution Theory and Transform Analysis: An Introduction to Generalised Functions, with Applications.* McGraw-Hill, New York (1965).