ANNALES DE L'I. H. P., SECTION B

M. F. BARNSLEY

S. G. DEMKO

J. H. ELTON

J. S. GERONIMO

Invariant measures for Markov processes arising from iterated function systems with place-dependent probabilities

Annales de l'I. H. P., section B, tome 24, nº 3 (1988), p. 367-394

<http://www.numdam.org/item?id=AIHPB_1988__24_3_367_0>

Ann. Inst. Henri Poincaré,

Vol. 24, n° 3, 1988, p. 367-394. *Probabilités et Statistiques*

Invariant measures for Markov processes arising from iterated function systems with place-dependent probabilities

by

M. F. BARNSLEY (*), S. G. DEMKO, J. H. ELTON (*) and J. S. GERONIMO (**)

School of Mathematics,
Georgia Institute of Technology,
Atlanta, GA 30332

ABSTRACT. – Consider a discrete-time Markov process on a locally compact metric space X obtained by randomly iterating Lipschitz maps $w_1, \ldots, w_N$; the probability $p_i(x)$ of choosing map w_i at each step is allowed to depend on current position x. Assume sets of finite diameter in X are relatively compact.

It is shown that if the maps are *average-contractive*, i. e.,

$$\sum_{i=1}^{N} p_i(x) \log \frac{d(w_i x, w_i y)}{d(x, y)} < 0$$

uniformly in x and y, and if the p_i's are bounded away from zero and satisfy a Dini-type continuity condition (weaker than Hölder-continuity), then the process converges in distribution to a unique invariant measure.

Also discussed are Perron-Frobenius theory and primitive weakly almost-periodic Markov operators, discontinuous maps, Julia sets, and running dynamical systems backwards.

Key words : Invariant measures, discrete-time Markov processes, Markov operators.

(*) Supported in part by DARPA "Image Compression Project" O.N.R. Contact #N00014-86-C-0446.

(**) Supported in part by N.S.F. Grant #DMS-8620079.

RÉSUMÉ. — On considère un processus de Markov en temps discret sur un espace métrique localement compact obtenu par itération aléatoirement des cartes Lipschitzian $w_1, w_2, \ldots, w_N$. La probabilité $p_i(x)$ qu'on choisisse la carte w_i peut dépendre sur la position courante. On suppose que les ensembles de diamètre fini sont relativement compact.

On démontre que si les cartes sont moyen-contractunt, i. e.,

$$\sum_{i=1}^{N} p_i(x) \log \frac{d(w_i x, w_i y)}{d(x, y)} < 0$$

uniformément dans x et y, et si les p_i restent supérieur de zéro et satisfont une condition de continuité Dini, puis le processus converge en distribution à une mesure unique.

Aussi on discute la théorie de Perron-Frobenius et les opérateurs de Markov vaguement presque-périodique, les cartes discontinues, et les ensembles de Julia.

TABLE OF CONTENTS

1. INTRODUCTION

We study certain discrete-time Markov processes defined on locally compact metric spaces. These processes arise naturally when iterating functions and have been studied in connection with population dynamics [OM], learning models ([K], [DF]), Julia sets ([BD], [BGM]), fractals ([H], [BD]), and computer graphics ([DS], [DHN]). The set-up in this paper is as follows: (X, d) is a metric space in which sets of finite diameter are relatively compact, so that X is locally compact and σ-compact; $\{w_i\}_{i=1}^{N}$ are Borel measurable functions from X into X; and $\{p_i\}_{i=1}^{N}$ is a non-negative Borel measurable partition of unity on X. For a given $x \in X$ and

Borel subset $B \subseteqq X$ the probability of transfer from x to B is defined by

$$P(x, B) = \sum_{i=1}^{N} p_i(x)\,\delta_{w_i(x)}(B) = \sum_{i=1}^{N} p_i(x)\,1_B(w_i(x)) \qquad (1.1)$$

where δ_z denotes the Dirac measure concentrated at z and 1_B denotes the indicator function of B. Closely connected with P is the Markov operator T defined for real-valued functions on X by

$$(T f)(x) = \int f(y)\,P(x, dy) = \sum_{i=1}^{N} p_i(x)\,f(w_i(x)). \qquad (1.2)$$

The space of continuous real-valued functions on X with compact support will be denoted by $C_c(x)$. The space of bounded real-valued continuous functions on X is C(X) and the space of finite signed Borel measures on X is M(X). $\mathscr{P}(X) \subset M(X)$ are the probability measures. C(X) carries the supremum norm. If X is compact, then M(X) is the dual of C(X). In general, however, the dual of C(X) is the space of finitely additive regular finite signed Borel measures on X (*see* [DuS]).

The operator $V : M(X) \to M(X)$ defined by

$$V\nu(B) = \int P(x, B)\,d\nu(x) = \sum_{i=1}^{N} \int_{w_i^{-1}(B)} p_i(x)\,d\nu(x) \qquad (1.3)$$

describes how a probability distribution ν on X is transformed in one step of the process. It is easy to see that V is the adjoint T^* restricted to M(X). A Borel probability measure μ is said to be *invariant* or *stationary* if $V\mu = \mu$. It is said to be *attractive* if for all $\nu \in \mathscr{P}(X)$,

$$\int f\,d(V^n \nu) \to \int f\,du$$

for all $f \in C(X)$; that is, $V^n \nu$ converges to μ in distribution.

We now state these things in probabilistic notation, which we use later. For $\nu \in \mathscr{P}(X)$, let $\{Z_n^\nu;\ n = 0, 1, \ldots\}$ be the Markov stochastic process (unique up to distribution) having initial distribution ν and transition probability as above. For $x \in X$, let $\{Z_n^x\}$ be the process with initial distribution concentrated at x; that is, $Z_n^x = Z_n^{\delta_x}$. Thus $T f(x) = E f(Z_1^x)$, and $V\nu(B) = P(Z_1^\nu \in B)$; μ is attractive if $E f(Z_n^\nu) \to \int f\,d\mu$, $\forall \nu \in \mathscr{P}(X)$, $\forall f \in C(X)$.

Let $\Omega = N^{\mathbb{N}} = \{\underline{i} = (i_1, i_2, \ldots) : 1 \leqq i_j \leqq N\}$. For $x \in X$, let P_x be the probability measure defined on "thin cylinders" of Ω by

$$P_x\{\underline{i} : j\text{th coordinate of } \underline{i} \text{ is } i_j,\ j = 1, \ldots, n\} = p_{i_1}(x)\,p_{i_2}(w_{i_1} x) \ldots p_{i_n}(w_{i_{n-1}} \ldots w_{i_1} x).$$

Then it is clear that the process $\{Z_n^x\}$ may be realized on (Ω, P_x) as

$$Z_n^x(\underline{i}) = w_{i_n} \dots w_{i_1} x. \tag{1.4}$$

This is the interpretation of the process as a random walk in X, using an N-sided die (which may depend on position) to determine which map is to be used to carry us to the next position.

If T maps C(X) into C(X) (i. e., the process is a *Feller process*), then $V^n \nu$ converges in distribution to $\mu \Leftrightarrow T^{*n} \nu \overset{w^*}{\to} \mu$. In this case, it is easily seen that an attractive probability measure is necessarily invariant: since T^* is $w^* - w^*$ continuous, $T^{*n} \nu \overset{w^*}{\to} \mu$, $\forall \nu \in \mathcal{P}(X)$

$$\Rightarrow \quad T^*(T^{*n} \nu) \overset{w^*}{\to} T^* \mu \quad \Rightarrow \quad T^* \mu = \mu.$$

This will be the case when the w_i's are continuous, but at times we consider discontinuous w_i's.

If X is *compact* and T maps C(X) into C(X), then $\mathcal{P}(X)$ is w^*-compact in M(X) and T^* is $w^* - w^*$ continuous, so the existence of an invariant probability follows from a fixed-point theorem. However, we shall be concerned in Section 2 with non-compact X, so $\mathcal{P}(X)$ is not w^*-closed in M(X).

The main focus of this paper is on the existence and attractiveness of invariant measures in the case that the p_i's are non-constant and the w_i's not contractions and X is not necessarily compact. Constant p_i's and contractive w_i's have been treated by several authors ([DF], [H], [DS], [BD]), and have been used with some success to make computer pictures of certain natural objects, ([DS], [DHN]). Variable p_i's in (1.1) were studied by Onicescu and Mihoc [OM] in 1935 who made a connection with the spread of tuberculosis, by Doeblin and Fortet [DoF] in 1937, by Ionescu Tulcea and Marinescu ([ITM], [IT]) in 1948 and 1959, and by Karlin [K] in 1953. Karlin's paper was motivated by work of Bush and Mosteller on learning models, *cf.* [BM]. More recently, Bessis, Geronimo, and Moussa [BGM] studied orthogonal polynomials and the associated Jacobi matrices (or discrete Schrödinger operators) connected with invariant measures for chains of the form (1.1) where the w_i's and p_i's satisfy

$$\sum_{i=1}^{N} \frac{p_i(x)}{z - w_i(x)} = \frac{Q(z)}{S(z) - x}$$

where S is a monic polynomial of degree N having a real Julia set J, the w_i's are the inverse branches of S and Q is a monic polynomial of degree $N-1$ constrained to insure that $p_i(x) > 0$ for $x \in J$ and $i = 1, 2, \dots, N$. The typically singular continuous nature of the resulting invariant measure and the close connection with fractal structures in the physical sciences was a major motivation for this work.

The following "pure types" results for unique invariant measures for our processes follows quickly from more general results of Dubins and Freedman (Lemma 2.2 and Theorem 2.5 of [DF]).

THEOREM 1.2. — *(a) If* $V\mu=\mu$ *with* $\mu\geq 0$, *then the discrete and the continuous parts of* μ *are also fixed points of* V. *In particular, if there is a unique invariant probability measure, then it is either discrete or continuous.*

(b) Suppose m is a probability measure on X *such that* $m(A)=0$ *implies* $m(w_i(A))=0$ *for* $1\leq i\leq N$. *If* μ *is invariant for* (1.1), *then so are the absolutely continuous part and the singular part (with respect to m). Consequently, if* μ *is the unique invariant measure for* (1.1), *then it is of pure type.*

Note that *(b)* applies if the w_i's are Lipschitz (as in Section 2A) and m is Lebesgue measure on $\mathbb{R}^n$.

The paper is organized as follows. In Section 2A we prove that there is an attractive invariant measure under the assumptions that the moduli of continuity of the p_i's satisfy Dini's condition and that the w_i's are Lipschitz continuous and satisfy an average contractivity condition between any two points. This average contractivity condition was first formulated in [BE] and does not require any of the w_i's to be contractions. In Section 2B we relax the continuity condition on the w_i's and restrict attention to $X=[0,1]$. Section 3 considers Perron-Frobenius theory in the case of compact X; a strong ergodic theorem is proved for primitive weakly almost periodic Markov operators, which includes an important class of operators of type (1.2). Section 4 is concerned with examples and applications.

All four authors of this paper were involved in formulating and motivating these problems, and in discovering the right background. Section 2A is largely the work of John Elton, with assistance from Jeff Geronimo; section 2B is largely due to Jeff Geronimo; section 3 is largely due to Steve Demko, and section 4 is due to Michael Barnsley.

2A. LIPSCHITZ MAPS ON LOCALLY COMPACT METRIC SPACES

Let w_i, $i=1, \ldots, N$ be Lipschitz functions from X into X, with $d(w_i x, w_i y)\leq s_i d(x, y)$ for all x, y in X. Let $s=\max\{s_i : i=1, \ldots, N\}$.

Let $p_i : X\rightarrow[0,1]$ be continuous, $i=1, \ldots, N$, with $p_i(x)\geq 0$ and $\sum_{i=1}^{N} p_i(x)=1$ for all x in X.

Thus the operator T in (1.2) takes $C(X)\rightarrow C(X)$ (but does not in general take $C_c(X)\rightarrow C_c(X)$).

The case of variable p_i's, when $X=[0,1]$ and the w_i's are affine *contractions*, was considered by Karlin [K]. In Section 6, p. 749, he states that an attractive invariant measure exists when the p_i's are strictly positive and merely *continuous*. However, there seems to be an error in his proof: he assumes that the p_i's are differentiable in his proof (Lemma on page 750 and the application of it on page 751). But $\|p_i-q_i\|_\infty<\delta$ does *not* imply $|T_p^n f(x)-T_q^n f(x)|<\varepsilon$ uniformly in n, so it does no good to uniformly approximate a continuous p_i by a differentiable one.

We prove that it is sufficient that the moduli of uniform continuity φ_i of the p_i's satisfy "Dini's condition", i. e., $\varphi_i(t)/t$ is integrable over $(0, \delta)$ for some $\delta>0$. This is a little more than continuity, but includes the case when the p_i's are in Lip_α for some $\alpha>0$. We do not require X to be compact or any of the w_i's to be contractions, but require only an average contractivity condition between points, as in [BE]. These conditions together will guarantee that $\{T^n f)$ is equicontinuous for $f\in C_c(X)$. We do not require that the p_i's be bounded away from zero, but do require that the probability of contraction between any two points be bounded away from zero in order to get an attractive invariant measure.

We now state the main results of this section.

THEOREM 2.1. – *Suppose there exists* $r<1$ *and* $q>0$ *such that*

$$\|d(Z_1^x, Z_1^y)\|_q \leqq r\, d(x, y),$$

with the norm taken in $L^q(\Omega, P_x)$ [*recall* Z_1^x *was defined by* (1.3)]. *That is,*

$$\sum_{i=1}^{N} p_i(x)\, d^q(w_i x, w_i y) \leqq r^q d^q(x, y), \qquad \forall x, y \text{ in } X;$$

assume that the modulus of uniform continuity of each p_i *satisfies Dini's condition, and that there exists* $\delta>0$ *such that*

$$\sum_{i\,:\, d(w_i x,\ w_i y)\leqq r\, d(x,\ y)} p_i(x)\, p_i(y) \geqq \delta^2, \qquad \forall x, y \text{ in } X$$

[*for example, if* $p_i(x)\geqq\delta$ *for all i,* $\forall x\in X$]. *Then there is an attractive* (*hence unique*) *invariant probability measure for the Markov process described above. For every* $f\in C_c(X)$, $T^n f$ *converges to a constant uniformly on sets of finite diameter.*

Remark. – An ergodic theorem for the *orbits* of such processes has recently been proved by one of the authors [E].

Remark. – A special case when the average contractivity condition in Theorem 2.1 hilds is if

$$\sum_{i=1}^{N} p_i(x)\, s_i^q \leqq r^q, \qquad \forall x\in X.$$

The following gives a seemingly weaker average contractivity hypothesis.

COROLLARY 2.2. — *Assume that there exists* $r_1 < 1$ *such that*

$$E_x(\log d(Z_1^x, Z_1^y)) \leqq \log r_1 d(x, y);$$

that is,

$$\prod_{i=1}^{N} d(w_i x, w_i y)^{p_i(x)} \leqq r_1 d(x, y), \qquad \forall x, y \text{ in } X.$$

$\left(\textit{For example, if } \prod_{i=1}^{N} s^{p_i(x)} \leqq r_1, \ \forall x \textit{ in } X.\right)$ *Assume that the modulus of uniform continuity of each* p_i *satisfies Dini's condition, and that there exists* $\delta > 0$ *such that* $p_i(x) \geqq \delta$ *for each i,* $\forall x$ *in x. Then in fact the hypotheses of Theorem* 2.1 *hold for some* $q > 0$ *and some* $r < 1$, *so the conclusions hold also.*

Remark. — It is not required that there is contraction on the average between any two points; it is sufficient that before some fixed number of iterations (independent of x and y), there is contraction on the average. That is, there exists $r < 1$ and an integer n_0 such that for all $x \neq y$ in X, $\exists n \leqq n_0$ such that

$$E_x(\log d(Z_n^x, Z_n^y)) \leqq \log r d(x, y).$$

COROLLARY 3.3. — *Suppose the* w_i *are* affine *maps on* $\mathbb{R}^n$ *(any norm) with linear part* a_i *(i.e.,* $w_i x = a_i x + b_i$*). Suppose the moduli of uniform continuity of the* p_i*'s satisfy Dini's condition, and* $p_i(x) \geqq \delta > 0$ *for all x. Then the conclusions of Theorem* 2.1 *hold if the following holds:*

There exists $r < 1$ *and an integer* m_0 *such that for all* $x \in \mathbb{R}^n$ *and for all* $y \in \mathbb{R}^n$ *with* $\|y\| = 1$, *there exists* $m \leqq m_0$ *such that*

$$E_x(\log \| a_{i_m} \dots a_{i_1} y \| \leqq \log r;$$

that is

$$\sum_{i_1 \dots i_m} \sum p_{i_1}(x) p_{i_2}(w_{i_1} x) \dots p_{i_m}(w_{i_{m-1}} \dots w_{i_1} x) \log \| a_{i_m} \dots a_{i_1} y \| \leqq \log r.$$

In case all the w_i's are *strict* contractions (that is, $s < 1$), one can give a much quicker proof than the one we give in Lemma 2.5 that $\{T^n f\}$ is equicontinuous. But we still seem to need the Dini's condition on the moduli of continuity of the p_i's even in this case; we do not, however, have an example to show that continuity alone is insufficient.

Our argument in the proof of Theorem 2.1 uses a modification of Karlin's clever decomposition of $T^m f(x) - T^m f(y)$ based on the idea of recurrence. We use a conditional expectation argument together with Chebyshev's inequality to guarantee the occurrence of a desired event in finite time; it is not quite a recurrent event in our case.

We also give the following conditions for the existence of various moments of the invariant measure; the stronger the average contractivity condition, the faster the drop-off rate of the measure.

THEOREM 2.4. – *Let* $0<q<\infty$ *and let* $x_0 \in X$. *Let*

$$C=\max\{d(w_i x_0, w_0), i=1, \ldots, N\}.$$

Assume $\exists r_q<1$ *such that the hypotheses of Thorem* 2.1 *are satisfied, with* $r=r_q$. *Then*

$$\int d^q(x, x_0)\, d\mu(x) \leqq B_q,$$

where

$$B_q=\begin{cases} C^q/(1-r_q^q) & \text{if } 0<q\leqq 1 \\ C^q/(1-r_q)^q & \text{if } 1\leqq q<\infty \end{cases}$$

and μ *is the invariant probability measure which exists by Theorem* 2.1.

Proof of Theorem 2.1 *and Corollaries.* – By the modulus of uniform continuity φ of a uniformly continuous function p on X we mean the function

$$\varphi(t)=\sup\{|p(x)-p(y)| : d(x, y)\leqq t\}, \qquad t\geqq 0.$$

We say "φ satisfies Dini's condition" if $\varphi(t)/t$ is integrable on $(0, \alpha)$ for some $\alpha>0$.

LEMMA 2.5. – *Let* φ_i *be the modulus of uniform continuity of* p_i, $i=1, \ldots, N$, *and assume that each* φ_i *satisfies Dini's condition. Assume* $\exists q>0$ *and* $r<1$ *such that* $\forall x, y \in X$,

$$\| d(Z_1^x, Z_1^y) \|_q \leqq r\, d(x, y).$$

Then $\forall f \in C_c(x)$, $\{T^n f\}$ *is equi-uniformly continuous.*

Proof. – Note that since φ_i is a modulus of uniform continuity, it is (*a*) continuous, (*b*) non-decreasing, (*c*) $\varphi_i(0)=0$, and (*d*) subadditive (*see* [Lor], p. 43).

Also, $\varphi_i(t)\leqq 1$, $\forall t$ since $|p_i(x)-p_i(y)|\leqq 1$, $\forall x, y$. Let

$$\varphi_0(t)\begin{cases} t, & 0\leqq t\leqq 1 \\ 1, & t>1. \end{cases}$$

It is easy to see that φ_0 satisfies (*a*)$-$(*d*) also.

Let $\varphi^*=\varphi_0 \vee \varphi_1 \vee \ldots \vee \varphi_N$. Then it is easy to see that φ^* satisfies (*a*)$-$(*d*) (we denote $\max\{t, u\}=t \vee u$).

Bet B the region *under* the graph of φ^* on $[0, \infty)$ and let φ be the upper boundary curve of the *convex hull* of B. Note $\varphi(t)=1$ for $t\geqq 1$, and φ is a concave function. It follows from Theorem 8 in ([Lor], p. 45) applied to the region under the graph of φ^* for $0\leqq t\leqq 1$ that φ satisfies

$(a)-(d)$ and in addition $\varphi^* \leqq \varphi \leqq 2\varphi^*$. Thus φ satisfies Dini's condition also.

Let $f \in C_c(X)$, $\|f\| \leqq 1$, and also assume $f \in \mathrm{Lip}_1$, so

$$|f(x)-f(y)| \leqq C\, d(x, y),\ \forall x, y \in X.$$

We may take $C \geqq 2$.

Without loss of generality, we may assume $q \leqq 1$ in the hypothesis of the lemma.

Define

$$\beta^*(t) = \frac{N \vee C}{1-r^q} \int_0^{tr^{-q}} \frac{\varphi(u)}{u}\, du.$$

This is finite since φ is Dini. Then $\beta^*(0)=0$, and β^* is continuous and strictly increasing. Also, β^* is a *concave* function: compute

$$\frac{d}{dt}\beta^*(t) = \frac{N \vee C}{1-r^q} \frac{\varphi(tr^{-q})}{t}, \qquad \forall t,$$

and this is non-increasing since φ is concave, by a simple argument.

Let $\gamma(t)=t^q$. Fix $x, y \in X$. Let $h(i)=d(w_i x, w_i y)$. Then the second assumption in the hypothesis can be expressed as

$$\int \gamma \circ h\, dp \leqq \gamma(r\, d(x, y)),$$

where p is the discrete probability measure $p\{i\}=p_i(x)$.

Let $\beta=\beta^* \circ \gamma$. Then $\beta \circ \gamma^{-1} = \beta^*$ which is concave and strictly increasing, so by Jensen's inequality

$$\int (\beta \circ \gamma^{-1}) \circ (\gamma \circ h)\, dp \leqq (\beta \circ \gamma^{-1})\left(\int \gamma \circ h\, dp\right),$$

i. e.,

$$\int \beta \circ h\, dp \leqq (\beta \circ \gamma^{-1})\left(\int \gamma \circ h\, dp\right).$$

But $\beta \circ \gamma^{-1}$ is increasing, so

$$\int \beta \circ h\, dp \leqq (\beta \circ \gamma^{-1})(\gamma(r\, d(x, y))) = \beta(r\, d(x, y)).$$

That is,

$$\sum_{i=1}^{N} p_i(x)\, \beta(d(w_i x, w_i y)) \leqq \beta(r\, d(x, y)), \qquad \forall x, y \in X. \tag{2.1}$$

Next, observe

$$\beta(t)-\beta(rt)=\beta^*(t^q)-\beta^*(r^q t^q)$$

$$\geqq \frac{N}{1-r^q}\,\frac{\varphi(t^q)}{(t/r)^q}[(t/r)^q-t^q]=N\,\varphi(t^q)\geqq N\,\varphi(t). \quad (2.2)$$

Now $\beta(t)\geqq C\int_0^{t^q} du \geqq Ct$ for $0\leqq t\leqq 1$, and $\beta(t)\geqq C\geqq 2$ for $t\geqq 1$, so we conclude: β is a modulus of uniform continuity for f.

Induction hypothesis. — β is a modulus of uniform continuity for $T^{m-1} f$. Then

$$\begin{aligned}|T(T^{m-1} f)(x)-T(T^{m-1} f)(y)| &\leqq \sum_{i=1}^{N} p_i(x)\,|T^{m-1} f(w_i x)-T^{m-1} f(w_i y)| \\ &\quad+\sum_{i=1}^{N} |p_i(x)-p_i(y)|\,.\,|T^{m-1} f(y)| \\ &\leqq \sum_{i=1}^{N} p_i(x)\,\beta(d(w_i x, w_i y))+N\,\varphi(d(x, y)),\end{aligned}$$

by the induction hypothesis. Thus using (2.1), we have

$$\begin{aligned}|(T^m f)(x)-(T^m f)(y)| &\leqq \beta(r\,d(x, y))+N\,\varphi(d(x, y)) \\ &\leqq \beta(d(x, y)), \quad \text{from} \quad (2.2)\end{aligned}$$

Thus β is a modulus of uniform continuity for $T^m f$ also, so this is true for all m.

Finally, since $\mathrm{Lip}_1(X)\cap C_c(X)$ is dense in $C_c(X)$, the result follows by a 3ε argument. □

LEMMA 2.6. — *Let* $p_i(x)\geqq\delta>0$, $\forall x$, $i=1, \ldots, N$, *and suppose* $\exists r_1<1$ *such that*

$$E_x(\log d(Z_1^x, Z_1^y))\leqq \log r_1\, d(x, y).$$

Then for any r *with* $r_1<r<1$, $\exists q_0>0$ *such that* $0<q\leqq q_0\Rightarrow$

$$\|d(Z_1^x, Z_1^y)\|_q\leqq r\,d(x, y).$$

Proof. — It is a standard fact that if $f\in L_1^+(\mu)$ then

$$\lim_{q\downarrow 0}\|f\|_q=\exp\int \log f\,d\mu \quad (2.3)$$

(*see* [HS], p. 201). It is easy to see that for $0<a<b$ this convergence is uniform for $a \leqq f \leqq b$. Now let

$$f_{x,y}(i)=\frac{d(w_i x, w_i y)}{d(x, y)} \vee \left(\frac{r_1}{s}\right)^{1/\delta}.$$

Thus $\left(\frac{r_1}{s}\right)^{1/\delta} \leqq f_{x,y}(i) \leqq s$, $\forall i$, $\forall x, y \in X$. One easily shows from the hypothesis that

$$E_x(\log f_{x,y}) \leqq \log r_1,$$

and the lemma now follows, using (2.3).

The next lemma contains the probabilistic reasoning.

LEMMA 2.7. — *Assume that the hypotheses of Theorem* 2.1 *hold. Let* $f \in C_c(X)$. *Then* $\forall x, y$ *in* X, $\lim_{m\to\infty} |T^m f(x) - T^m f(y)| = 0$, *and the convergence is uniform on sets of finite diameter in* X.

Proof. — Let S be a set in X, of diameter $M<\infty$. Then there exists $C<\infty$ such that $d(x, w_i x) \leqq C$ for all $x \in S$, all i.

Fix $x, y \in S$. Let $\Omega' = \{\underline{i}' = (i'_1, i'_2, \ldots)\}$ be another copy of Ω, and let

$$\Omega^* = \Omega \times \Omega' = \{\underline{i}^* = (i_1, i'_1, i_2, i'_2, \ldots)\}.$$

Let $P^* = P_x \times P_y$, a probability measure on Ω^*. Thus if we define on Ω^*,

$$Z_n^x(\underline{i}^*) = w_{i_n} \ldots w_{i_1} x \qquad \text{and} \qquad \tilde{Z}_n^y(\underline{i}^*) = w_{i'_n} \ldots w_{i'_1} y,$$

then Z_n^x and $\tilde{Z}_n^y$ are independent copies of the Markov process with initial distributions concentrated on x and y.

Let $\alpha>0$. Let

$$G_n = \{\underline{i}^* : d(Z_n^x, \tilde{Z}_n^y) \leqq \alpha, \text{ and } d(Z_j^x, \tilde{Z}_j^y) > \alpha \text{ for } j<n\}$$

[we suppress the $\underline{i}^*$ in $Z_n^x(\underline{i}^*)$, etc.], so that the G_n are disjoint. Let $B_m = \sim \bigcup_{n=1}^{m} G_n$. The key thing is to show that $P^*(B_m) \to 0$ *as* $m \to \infty$, which we shall do in the sublemma which follows. Then the following decomposition will conclude the proof; this idea (in a simpler setting and different notation) is due to Karlin. The idea is to stop the two independent processes when they are sufficiently close (which happens with probability one). For $f \in C_c(X)$,

$$
\begin{aligned}
T^m f(x) - T^m f(y) &= E f(Z_m^x) - E f(\tilde{Z}_m^y) \\
&= \sum_{n=1}^{m} E[1_{G_n}(f(Z_m^x) - f(\tilde{Z}_m^y))] + E[1_{B_m}(f(Z_m^x) - f(\tilde{Z}_m^y))] \\
&= \sum_{n=1}^{m} E[1_{G_n}[E(f(Z_m^x) \mid \Sigma_n) - E(f(\tilde{Z}_m^y) \mid \Sigma_n)]] \\
&\qquad + E[1_{B_m}(f(Z_m^x) - f(\tilde{Z}_m^y))] \\
&= \sum_{n=1}^{m} E[1_{G_n}(T^{m-n} f(Z_n^x) - T^{m-n} f(\tilde{Z}_n^y))] \\
&\qquad + E[1_{B_m}(f(Z_m^x) - f(\tilde{Z}_m^y))], \quad (2.4)
\end{aligned}
$$

where Σ_n above is the σ-algebra in Ω^* generated by $Z_1^x, \ldots, Z_n^x, \tilde{Z}_1^y, \ldots, \tilde{Z}_n^y$, and the expectation means with respect to the measure P^*. The last equation in the above decomposition follows by writing out, for example,

$$
E(f(Z_m^x) \mid \Sigma_n) = \sum_{i_{n+1}} \ldots \sum_{i_m} p_{i_{n+1}}(Z_n^x) \times \ldots \times p_{i_m}(w_{i_{m-1}} \ldots w_{i_{n+1}} Z_n^x) f(w_{i_m} \ldots w_{i_{n+1}} Z_n^x) = T^{m-n} f(Z_n^x).
$$

We shall now conclude the proof under the assumption that $\forall \alpha > 0$, $P^*(B_m) \to 0$. Let $\varepsilon > 0$, and choose, by Lemma 2.5, $\alpha > 0$ such that $d(u, v) \leqq \alpha \Rightarrow |T^n f(u) - T^n f(v)| < \varepsilon$ for all n. Then from (2.4),

$$
|T^m f(x) - T^m f(y)| \leqq \sum_{n=1}^{m} E[1_{G_n} \varepsilon] + E[1_{B_m} 2 \| f \|] \leqq \varepsilon + 2 \| f \| P^*(B_m),
$$

since the G_n are disjoint. Thus the proof of the Lemma will be complete when we prove the following:

SUBLEMMA. — *For* $\alpha > 0$, $P^*(B_m) \to 0$ *as* $m \to \infty$. *The convergence is uniform over* x, y *in* S.

Proof. — We cannot use recurrence, as Karlin did. We look at blocks of geometrically increasing length. Assume W.L.O.G. that $q \leqq 1$. Now

$$
\begin{aligned}
E(d^q(Z_{n+1}^x, x) \mid Z_n^x) &= \sum_{j=1}^{N} d^q(w_j Z_n^x, x) p_j(Z_n^x) \\
&\leqq \sum_{j=1}^{N} [d^q(w_j Z_n^x, w_j x) + d^q(w_j x, x)] p_j(Z_n^x) \\
&\leqq r^q d^q(Z_n^x, x) + C^q,
\end{aligned}
$$

using the hypothesis of the theorem. Thus if $n_2 > n_1$,

$$
\begin{aligned}
E(d^q(Z_{n_2}^x, x) \mid Z_{n_1}^x) &= E[E(d^q(Z_{n_2}^x, x) \mid Z_{n_2-1}^x) \mid Z_{n_1}^x] \\
&\leqq C^q + r^q E(d^q(Z_{n_2-1}^x, x) \mid Z_{n_1}^x).
\end{aligned}
$$

Repeating this argument, we are led to

$$E(d^q(Z^x_{n_2}, x) \mid Z^x_{n_1}) \leqq \frac{C^q}{1-r^q} + r^{(n_2-n_1)q} d^q(Z^x_{n_1}, x). \tag{2.5}$$

Now,

$$\begin{aligned} d^q(Z^x_{n_1}, x) &= d^q(w_{i_n} Z^x_{n_1-1}, x) \\ &\leqq d^q(w_{i_n} Z^x_{n_1-1}, w_{i_n} x) + d^q(w_{i_n} x, x) \\ &\leqq s^q d^q(Z^x_{n_1-1}, x) + C^q, \end{aligned}$$

and repeating this, assuming W.L.O.G. $s^q > 2$, we get from (2.5)

$$E(d^q(Z^x_{n_2}, x) \mid Z^x_{n_1}) \leqq C^q \left[\frac{1}{1-r^q} + r^{(n_2-n_1)q} s^{n_1 q} \right].$$

Let $\gamma = \dfrac{\log(s/r)}{\log(1/r)}$, and assume $n_2 \geqq \gamma n_1$; then

$$E(d^q(Z^x_{n_2}, x) \mid Z^x_{n_1}) \leqq \frac{2C^q}{1-r^q} = \frac{\lambda^q}{2}, \quad \text{say.}$$

By Chebyshev's inequality,

$$P^*(d(Z^x_{n_2}, x) > \lambda \mid Z^x_{n_1}) \leqq \frac{\lambda^q/2}{\lambda^q} = 1/2.$$

A similar result holds for the process $\{\tilde{Z}^y_n\}$, and since they are independent, $P^*(d(Z^x_{n_2}, x) \leqq \lambda$ and $d(\tilde{Z}^y_{n_2}, y) \leqq \lambda \mid Z^x_{n_1}, \tilde{Z}^y_{n_2}) \geqq 1/4$.

Recall $Z^x_n = w_{i_n} Z^x_{n-1}$, so that

$$P^*(i_n = i'_n \text{ and } d(Z^x_n, \tilde{Z}^y_n) \leqq r\, d(Z^x_{n-1}, \tilde{Z}^y_{n-1}) \mid Z^x_{n-1}, \tilde{Z}^y_{n-1}) \geqq \delta^2.$$

Thus if $n_2 \geqq \gamma n_1 + k$, then by the Markov property

$P^*(i_{n_2-j} = i'_{n_2-j}$ and

$d(Z^x_{n_2-j}, \tilde{Z}^y_{n_2-j}) \leqq r\, d(Z^x_{n_2-j-1}, \tilde{Z}^y_{n_2-j-1})$ for $j=0, \ldots, k-1$

$$\text{and } d(Z^x_{n_2-k}, x) \leqq \lambda \text{ and } d(\tilde{Z}^y_{n_2-k}, y) \leqq \lambda \mid Z^x_{n_1}, \tilde{Z}^y_{n_1}) \geqq \frac{1}{4}\delta^{2k}.$$

Now by the triangle inequality, the above yields

$$P^*(d(Z^x_{n_2}, \tilde{Z}^y_{n_2}) \leqq r^k(2\lambda + M) \mid Z^x_{n_1}, \tilde{Z}^y_{n_1}) \geqq \frac{1}{4}\delta^{2k}.$$

Take k so large that $r^k(2\lambda + M) < \alpha$. Thus

$$P^*(d(Z^x_{n_2}, \tilde{Z}^y_{n_2}) > \alpha \mid Z^x_{n_1}, \tilde{Z}^y_{n_1}) \leqq 1 - \frac{1}{4}\delta^{2k}.$$

Finally, let $n_1, n_2, \ldots$ be a sequence of integers such that $n_{j+1} \geqq \gamma n_j + k$, $\forall j$. Then by the above, and the Markov property,

$$P^*(d(Z^x_{n_j}, \tilde{Z}^y_{n_j}) > \alpha, j=1, \ldots, l) \leqq \left(1 - \frac{1}{4}\delta^{2k}\right)^{l-1},$$

so $P^*(B_m) \leqq \left(1 - \frac{1}{4}\delta^{2k}\right)^{l-1}$ if $m \geqq n_l$. Thus $P^*(B_m) \to 0$ as $m \to \infty$, and the convergence is uniform over x, y in S since γ does not depend on x or y. This completes the proof of the sublemma and the lemma. □

Proof of Theorem 2.1. — The same conditional expectation argument used in the proof of the sublemma above shows that for any x,

$$E(d^q(Z^x_n, x) \leqq C^q(1-r^q)^{-1}.$$

Let $\varepsilon > 0$. By the above and Chebyshev's inequality,

$$P^*(d(Z^x_n, x) > \rho) < \varepsilon$$

for some $\rho < \infty$. Let K be the closed ball of radius ρ centered at x. Then for any $f \in C(X)$, for all n,

$$|T^n f(x)| = |E f(Z^x_n)| \leqq \varepsilon \| f \| + \| f \|_K \tag{2.6}$$

where $\| f \|_K = \sup \{ |f(x)| : x \in K \}$.

Now let $f \in C_c(X)$. By Ascoli's theorem, since $\{T^n f\}$ is equicontinuous, for some subsequence $T^{n_j} f$ converges uniformly on compact sets to what must be a constant function, say c, by Lemma 2.7. Note K is compact, so for sufficiently large j, $\| T^{n_j} f - c \|_K < \varepsilon$. Thus for $n \geqq n_j$,

$$\begin{aligned} |(T^n f - c)(x)| &= |T^{n-n_j}(T^{n_j} f - c)(x)| \\ &\leqq \varepsilon(\| f \| + c) + \| T^{n_j} f - c \|_K < \varepsilon(\| f \| + c + 1) \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, this shows $T^n f(x) \to c$, hence $T^n f(y) \to c$ for all y, by Lemma 2.7.

We may now define, for $f \in C_c(X)$, $L(f) = \lim_{n \to \infty} T^n f(x)$, which is clearly a bounded positive linear, functional so by Riesz's theorem, $\lim_{n \to \infty} T^n f(x) = \int f \, d\mu$ for some positive Borel measure. A standard argument using (2.6) and Urysohn's lemma shows that μ is a probability measure. Finally, let ν be any probability measure. Then for $f \in C_c(X)$, $T^{*n} \nu(f) = \int T^n f \, d\nu \to \int \left(\int f \, d\mu \right) d\nu = \int f \, d\mu$, by Lebesgue's theorem. Since μ is a probability measure, this holds for any $f \in C(X)$ as well, so $T^{*n} \nu \xrightarrow{w^*} \mu$, as desired. □

Proof of Corollary 2.2. This follows immediately from Lemma 2.6 and Theorem 2.1.

Proof of Corollary 2.3. This follows by easy modifications of the proofs of Lemmas 2.5, 2.6, and 2.7.

Proof of Theorem 2.4. Let δ_{x_0} be the Dirac measure at x_0. Let $f(x)=d^q(x, x_0)$. For each $M<\infty$, let $f_M=f\wedge M$, so $f_M \in C(X)$. We compute

$$(T^{*m}\delta_{x_0})(f_M)=\int T^m f_M\, d\delta_{x_0}=T^m f_M(x_0)\leqq E\, f(Z_m^{x_0}, x_0).$$

In case $0<q\leqq 1$, it was already shown in the proof of Theorem 2.1 that

$$E(d^q(Z_m^{x_0}, x_0))\leqq \frac{C^q}{1-r_q^q}=B_q.$$

Now suppose $1\leqq q<\infty$. By Minkowski's inequality, for any $n<m$, it follows that

$$E(d^q(Z_{n+1}^{x_0}, x_0)\,|\,Z_n^{x_0})\leqq \{r_q\, d(Z_n^{x_0}, x_0)+C\}^q,$$

by a similar argument to the one in the sublemma. Thus

$$\begin{aligned}[E(d^q(Z_{n+1}^{x_0}, x_0))]^{1/q}&=[E(E(d^q(Z_{n+1}^{x_0}, x_0)\,|\,Z_n^{x_0}))]^{1/q}\\ &\leqq r_q[E(d^q(Z_n^{x_0}, x_0))]^{1/q}+C\end{aligned}$$

by Minkowski again. Therefore, by induction

$$E[(d^q(Z_m^{x_0}, x_0))]^{1/q}\leqq \frac{C}{1-r_q},$$

so

$$E(d^q(Z_m^{x_0}, x_0))\leqq \frac{C^q}{(1-r_q)^q}=B_q.$$

Now let $m\to\infty$. By Theorem 2.1, $T^{*m}\delta_{x_0}\overset{w^*}{\to}\mu$, so $(T^{*m}\delta_{x_0})(f_M)\to\mu(f_M)$, so $\mu(f_M)\leqq B_q$, $\forall M<\infty$. Now let $M\uparrow\infty$, so $f_M\uparrow f$; by the monotone convergence theorem, $\mu(f)\leqq B_q$ also. □

2B. DISCONTINUOUS MAPS ON [0, 1]

We will now consider the case when the maps w_i have discontinuities; we will still assume that the p_i's are continuous. For simplicity, we discuss only the following special cases: let $X=I=[0, 1]$ and let $\Gamma=\{x_i, i\in\mathbb{N}\}$ be any countable subset of I. We shall say $f\in D(I)$ if f is continuous except on Γ where f may have discontinuities of the first kind, and is *right* continuous.

The special hypotheses imposed on the maps in Theorems 2.9 and 2.10 below to obtain existence and attractiveness of an invariant measure were motivated by analogy with the case of inverse branches of rational functions on the complex plane, i. e., Julia set theory. The application of our theory to the Julia set theory with place-dependent probabilities will appear in a future paper.

Let $\varphi \in D(I)^*$; then it is not difficult to show that

$$\varphi(f) = \int f\, d\sigma + \sum_{i=1}^{\infty} c_i (f(x_i) - f(x_i^-)) \tag{2.7}$$

for all $f \in D(I)$, where σ is a finite signed Borel measure and $\sum |c_i| < \infty$.

We shall assume that the w_i's are such that the operator T of (1.2) takes D(I) into D(I) [but not necessarily $C(I) \to C(I)$].

LEMMA 2.8. — *There exists a positive, norm-one functional* $\varphi \in D(I)^*$ *which is* invariant; *that is,* $T^* \varphi = \varphi$.

Proof. — This follows from a fixed-point theorem, as in [BD] or [DF]. □

Suppose

(*a*) For every $x \in I$ and every n, the points $w_{i_n} \ldots w_{i_1} x$, as $(i_1, \ldots, i_n)$ ranges over the n-tuples of integers between 1 and N, are all distinct, and

(*b*) There is $\delta > 0$ such that $p_j(x) \geqq \delta$ for all x and all j.

THEOREM 2.9. — *Given* (*a*) *and* (*b*), *if* $\varphi \in D(I)^*$ *is an invariant positive norm-one functional, then* φ *is a probability measure which is* continuous (*i. e., has no atoms*).

Proof. — In (2.7), note $c_i \geqq 0$ and σ is a positive measure since φ is positive.

First we show that σ is a continuous measure. Let x_0 be any point in I, and let $\varepsilon > 0$. We shall show $\sigma(\{x_0\}) \leqq \varepsilon$. Choose n so that $(1-\delta)^n < \varepsilon$. One may choose f with the following properties: f is continuous, $0 \leqq f \leqq 1$, $f(x_0) = 1$, and for all x, $T^n f(x) < \varepsilon$. To see this, note that for each $x \in I$, $\exists f_x$ continuous with $0 \leqq f_x \leqq 1$, $f_x(x_0) = 1$, and $f_x(w_{i_n} \ldots w_{i_1} x) = 0$ except for at most *one* choice of $(i_1, \ldots, i_n)$; this follows from hypothesis (*a*). Then $T^n f_x(x) \leqq (1-\delta)^n < \varepsilon$. Then by right continuity of $T^n f_x$, there is $b_x > x$ such that $T^n f_x(y) < \varepsilon$ for $x < y < b_x$. By compactness, there is a finite set F such that I is covered by $\{(x, b_x) : x \in F\}$. Let $f = \min \{f_x, x \in F\}$. Then $T^n f(y) < \varepsilon$ for all y.

Now we have

$$\sigma(\{x_0\}) \leqq \int f\, d\sigma = \varphi(f) = \varphi(T^n f) \leqq \varepsilon$$

since $\| T^n f \|_\infty \leqq \varepsilon$.

Next we shall show each $c_i = 0$. Fix i. Let $\varepsilon > 0$; we shall show $c_i \leqq \varepsilon$. By the same compactness argument as above, there is a function f such that

f is continuous except at x_i, $f(y)=0$ for $y<x_i$, $f(x_i)=1$, $0\leqq f\leqq 1$, and $T^n f(y)<\varepsilon$ for all y. Consequently

$$c_i \leqq c_i + \int f\,d\mu = \varphi(f) = \varphi(T^n f) \leqq \varepsilon. \quad \square$$

Now we shall discuss attractiveness of the invariant measure. Suppose now that

(*c*) each modulus of continuity φ_i of p_i satisfies Dini's condition (*see* the paragraph preceeding Lemma 2.5);

(*d*) there exists an integer m_0, real numbers $q>0$ and $0<r<1$ such that if $m\geqq m_0$,

$$E^{1/q}(d^q(Z_x^{m+1}, Z_y^{m+1}) \mid Z_x^m, Z_y^m) \leqq r\,d(Z_x^m, Z_y^m).$$

for all x, y.

THEOREM 2.10. — *Given* (*a*)–(*d*), *there exists an invariant probability measure* μ *such that for every probability measure* ν, $T^{*n}\nu(f) \to \mu(f)$ *for every* $f\in D(I)$; *that is,* $T^{*n}\nu \overset{w^*}{\to} \mu$.

Proof. — Using (*c*) and (*d*) and following the proof of Lemma 2.5, one finds that for all $f\in C(I)$, if $m\geqq m_0$,

$$|T^n f(Z_x^m) - T^n f(Z_y^m)| \leqq \beta(d(Z_x^m, Z_y^m))$$

for all x and y and all n. Here, β is as in the proof of Lemma 2.5. Using this, one modifies the proof of Lemma 2.7 in an obvious way (it is somewhat simpler since X is now compact) to conclude: for all $f\in C(I)$,

$$\lim_{n\to\infty} T^n f(x) - T^n f(y) = 0$$

for all x and y in I; the convergence is uniform.

Since $T^n f$ is not continuous, we cannot use Ascoli's theorem as was used in the proof of Theorem 2.1. Fix $x_0\in I$. By a diagonal argument, $T^{n_j} f(x_0)$ converges for all f in a countable dense subset of $C(I)$, and a 3ε-argument gives this for all f in $C(I)$. Let $\mu(f)=\lim T^{n_j} f(x_0)$. By the previous paragraph, $\lim T^{n_j} f(x)=\mu(f)$ for *all* x in I. This is a linear, positive, norm-one functional, so μ is a probability measure. By Theorem 2.9, there exists an invariant probability measure σ. Then for $f\in C(I)$, $\sigma(f)=\int T^{n_j} f\,d\sigma \to \int \mu(f)\,d\sigma = \mu(f)$. Thus, $\sigma=\mu$. It now follows by Lebesgue's theorem that $\int T^{n_j} f\,d\nu \to \mu(f)$ for all probability measures ν and $f\in C(I)$. Suppose for some other subsequence $\int T^{m_j} f\,d\nu \to \mu'(f)$ for all ν. Then since μ is invariant, $\mu=\mu'$. Thus for all probability measures

ν, $T^{*n}\nu(f) \to \mu(f)$ for all $f \in C(I)$. It now follows from a standard fact about convergence in distribution that since the set of discontinuities of any $f \in D(I)$ is contained in Γ which has μ-measure 0 (Γ is countable and μ is continuous),

$$T^{*n}\nu(f) \to \mu(f)$$

for all $f \in D(I)$ as well. □

3. COMPACT METRIC SPACES; PRIMITIVE AND IRREDUCIBLE OPERATORS

We now assume that X is compact and that T maps C(X) into C(X). This does not force the the w_i's to be continuous as the following examples show.

Example 3.1. – Let $X = \{z \in \mathbb{C} : |z| \leqq 1\}$ and $w_1(z) = \sqrt{z}$, $w_2(z) = -\sqrt{z}$, $p_i = 1/2$. Here, we take $w_1(z) = w_1(r e^{i\theta}) = \sqrt{r}\, e^{i\theta/2}$, $-\pi < \theta \leqq \pi$. So, the w_i's are not continuous along the negative real axis. However, if $f \in C(X)$ then Tf is also in $C(X)$, as one readily checks.

Example 3.2. – Let X be the Riemann sphere $\bar{\mathbb{C}}$ and let R be a rational function, $R : \bar{\mathbb{C}} \to \bar{\mathbb{C}}$, of degree N and let $\{w_i\}_{i=1}^{N}$ be a complete assignment of the inverse branches of R. Then, with $p_i = 1/N$, T maps C(X) into itself. More generally one can consider the iterated Riemann surfaces of [BD].

As was stated in the Introduction the existence of a stationary distribution is immediate. The next result gives a sufficient condition for unicity.

PROPOSITION 3.3. – *Suppose that for each non-zero* $f \in C(X)$ *there is a sequence of polynomials* $\{q_n\}$ *with degree of* $q_n = n$ *with coefficients summing to* 1 *such that* $\{q_n(T) f\}$ *converges boundedly pointwise to a constant* $c = c(f)$. *Then, there is unique stationary distribution* ν^*. *Furthermore, for any Borel probability measure* μ *on* X *and for* $g \in C(X)$, $\int g\, d[q_n(T^*)\mu] \to \int g\, d\nu^*$.

Proof. – Suppose ν and μ are distinct probability measures which are stationary for (1). Let $g \in C(X)$ be strictly positive with $\int g\, d\nu > \int g\, d\mu$. Now,

$$\int q_n(T) g\, d\nu = \int g\, d\nu > \int g\, d\mu = \int q_n(T) g\, d\mu.$$

In the limit we get $\int c(g)\, d\nu > \int c(g)\, d\mu$ which is a contradiction. □

Perron Frobenius Theory

It is possible to extend the classical Perron-Frobenius theory for $n \times n$ stochastic matrices to Markov operators on C(X). Irreducible Markov operators have been extensively studied ([R], [JS], [J], [S], [Kr]). For our application we consider in addition primitive Markov operators, to get an ergodic theorem which applies to an important class of operators of our type (*see* Propositions 3.6 and 3.8).

First we point out that T is not weakly compact in general, even in the simplest cases. For example, consider $Tf(x)=\frac{1}{2}f\left(\frac{1}{2}x\right)+\frac{1}{2}f\left(\frac{1}{2}x+\frac{1}{2}\right)$ on $X=[0,1]$. If f_n is any bounded sequence supported on $\left[0,\frac{1}{2}\right)$ which has no weakly convergent subsequence, then $Tf_n(x)=\frac{1}{2}f_n\left(\frac{1}{2}x\right)$ has no weakly convergence subsequence, either (such sequences do exist since the unit ball of C[0,1] is not weakly compact).

Now, it is clear that the number 1 is an eigenvalue of T with the constant function being eigenvectors. There may be other eigenvectors for T, *see* [K]. If 1 was the only eigenvalue of modulus one and was a simple pole of the resolvent operator, then for any $f \in C(X)$ we would have $T^n f \to$ constant at a geometric rate independent of f. However, it can happen that 1 is not even an isolated point of the spectrum of T, even though $T^n f \to$ constant uniformly.

Example. — Let $X=\{z:|z|=1\}$ be the unit circle and let $(Tf)(z)=\frac{1}{2}f(\sqrt{z})+\frac{1}{2}f(-\sqrt{z})$. With $z=e^{it}$ define

$$f_\lambda(z)=\sum_{k=0}^{\infty}\lambda^k\cos(2^k t)$$

for $|\lambda|<1$. Since $T(\cos 2^k t)=\cos 2^{k-1}t$ for $k \geqq 1$ and $T(\cos t)=0$, we see $Tf_\lambda=\lambda f_\lambda$. So the spectrum of T is the closed unit disk. Nevertheless $T^n f \to$ constant uniformly, by the second corollary of Proposition 3.9.

As this example shows, T^n in general cannot be close to a compact operator so the uniform ergodic theorem of Kakutani doesn't apply; *see* [DuS]. But the Perron-Frobenius theory we now consider does apply to such examples. The following definitions apply to any Markov operator on C(X).

DEFINITIONS. — (1) T is *irreducible* if for every $f \in C(X)$ with $f \geqq 0$ and $f \not\equiv 0$ and every $x \in X$, there exists k such that $T^k f(x)>0$ ([R], [JS]).

Because X is compact, it is easy to see that the following definition is equivalent:

(1′) T is irreducible iff for every $f \in C(X)$ with $f \geqq 0$ and $f \not\equiv 0$, there exists k such that for all $x \in X$, $S_k f(x) > 0$, where

$$S_k f(x) = \frac{1}{k+1} \sum_{j=0}^{k} T^j f(x).$$

(2) T is *primitive* if for every $f \in C(X)$ with $f \geqq 0$ and $f \not\equiv 0$ there exists $k = k(f)$ so that for all $x \in X$, $(T^k f)(x) > 0$.

If X is finite, these definitions are equivalent to the standard definitions, see [G] for example. The next two propositions show what these conditions mean for operators of type (1.2).

PROPOSITION 3.4. — *If $p_i(x) > 0$ for all i and x, then the following are equivalent. We use standard multi-index notation: $\bar{i} = (i_1, \ldots, i_k)$ and $|\bar{i}| = k$.*

(*a*) *T is irreducible.*

(*b*) *For every closed, proper subset A of X, there is an integer $k = k(A)$ so that*

$$\bigcap_{|\bar{i}| \leq k} w_{\bar{i}}^{-1}(A) = \varnothing.$$

(*c*) *For every open, non-empty subset θ of X, there is an integer $k = k(\theta)$ so that*

$$\bigcup_{|\bar{i}| \leq k} w_{\bar{i}}^{-1}(\theta) = X.$$

Proof. — $(b) \Leftrightarrow (c)$ follows from de Morgan's Laws.

$(a) \Rightarrow (b)$. Let A be as in (*b*) and let $f \in C(X)$ satisfy $f^{-1}(0) = A$, $f \geqq 0$, $f \not\equiv 0$. Since $\sum_{n=0}^{k} (T^n f) > 0$ for some k,

$$\bigcap_{j=0}^{k} \{x : (T^j f)(x) = 0\} = \varnothing.$$

But

$$(T^j f)(x) = 0$$

if and only if $f(w_{\bar{i}}(x)) = 0$ for all $|\bar{i}| = j$ because f is non-negative and all p_i's are positive. That is, $(T^j f)(x) = 0$ if and only if $x \in \bigcap_{|\bar{i}| = j} w_{\bar{i}}^{-1}(A)$.

$(b) \Rightarrow (a)$. If for some $f \geqq 0$, $f \not\equiv 0$, for every k there is x_k with $\sum_{j=0}^{k} T^j f(x_k) = 0$, then $x_k \in \bigcap_{|\bar{i}| \leq k} w_{\bar{i}}^{-1}(f^{-1}(0))$. □

We just state an analogue for primitive T.

PROPOSITION 3.5. — *If $p_i(x) > 0$ for all i and x, then the following are equivalent.*

(*a*) T *is primitive.*

(*b*) *For every closed, proper subset* A *of* X *there is an integer* k *with*

$$\bigcap_{|\bar{i}|=k} w_{\bar{i}}^{-1}(A)=\varnothing.$$

(*c*) *For every open, non-empty subset* θ *of* X *there is an integer* k *with*

$$\bigcap_{|\bar{i}|=k} w_{\bar{i}}^{-1}(\theta)=X.$$

Now we give a natural class of primitive operators.

PROPOSITION 3.6. — *Let* X *be the Julia set for the rational map* R *and let* $w_1, \ldots, w_N$ *be the inverse branches of* R *and let*

$$(Tf)(x) := \sum_{i=1}^{N} p_i(x) f(w_i(x))$$

map C(X) *into itself. Then, if the* p_i*'s are always positive,* T *is primitive.*

Proof. — For any open, non-empty set θ, $\bigcup_{|\bar{i}|=k} w_i^{-1}(\theta)$ is just $R^{0\,k}(\theta)$. It is well known (*cf.* [B]) that if θ is open and non-empty then after a finite number of iterations $R^{0\,k}(\theta)$ contains X. □

Notice that the nature of the p_i's does not have much to do with irreducibility or primitivity.

Now we discuss convergence of the iterates of T, for any Markov operator. T is called *weakly almost periodic* [JS] if for every $f \in C(X)$, $\{T^n f\}$ has a weakly convergent subsequence, i. e., a subsequence which converges pointwise to a continuous function. It is well-known [JS], p. 1047, that for such a T, the *averages* $S_k f$ converge uniformly for each $f \in C(X)$; this follows from the mean ergodic theorem [Dus], p. 661. If in addition, T is *irreducible*, then $S_k f$ converges uniformly to a *constant* for each $f \in C(X)$, and there is a *unique* invariant measure (*see* [Kr], p. 179). This follows from the fact that for an irreducible operator, $Tf = f$ iff f is constant.

A remarkable theorem of Jamison [J], making use of de Leeuw-Glicksberg's decomposition, asserts that if T is irreducible and weakly almost periodic, then T is in fact *strongly* almost periodic; that is, $\{T^n f\}$ has a *uniformly* convergent subsequence for each $f \in C(X)$. The following lemma combined with Jamison's theorem gives the convergence result we seek for *primitive* operators.

LEMMA 3.7. — *Let* T *be primitive, and* $f \in C(X)$. *If* $\{T^n f\}$ *has a uniformly convergent subsequence, then the whole sequence* $T^n f$ *converges uniformly to a constant.*

Proof. — Suppose $T^{n_i} f \to g$ uniformly for some subsequence. Then

$$\min T^{n_i} f \to \min g = m, \quad \text{say}, \qquad \text{and} \qquad \max T^{n_i} f \to \max g = M.$$

Since T is positive, $\min h \leqq \min \mathrm{T} h \leqq \max \mathrm{T} h \leqq \max h$ for any $h \in \mathrm{C}(\mathrm{X})$, so in fact for the entire sequence,

$$\min \mathrm{T}^n f \uparrow m \qquad \text{and} \qquad \max \mathrm{T}^n f \downarrow \mathrm{M}.$$

If $m = \mathrm{M}$ we are done, so suppose $m < \mathrm{M}$. Then $g - m \geqq 0$ and $g - m \not\equiv 0$, so by primitivity, there exists k such that $\mathrm{T}^k (g-m)(x) > 0$ for all x, so $\min \mathrm{T}^k g > m$. But $\mathrm{T}^{n_i + k} f \to \mathrm{T}^k g$ uniformly, so $\min \mathrm{T}^{n_i + k} f \to \min \mathrm{T}^k g > m$, a contradiction. □

PROPOSITION 3.8. — *Let* T *be a primitive, weakly almost periodic Markov operator. Then for all* $f \in \mathrm{C}(\mathrm{X})$, $\mathrm{T}^n f \to$ *constant uniformly, so there is an* attractive *invariant measure.*

Proof. — Lemma 3.7 and Jamison's theorem. □

Remark. — Even in the matrix case, irreducibility could not be substituted for primitivity in the above, of course; consider for example $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. As far as determining conditions on the w_i's that imply the hypotheses of Proposition 3.8, we have the following easy result.

PROPOSITION 3.9. — *Let* $\{w_{\bar{i}}\}_{|\bar{i}|=1}^{\infty}$ *be equicontinuous, then for all* $g \in \mathrm{C}(\mathrm{X})$, $\{\mathrm{T}^n g\}$ equicontinuous if all p_i's are constants.

COROLLARY. — *Let* $\{w_i\}_{i=1}^{\mathrm{N}}$ *be non-expansive and let* T *be primitive and let the* p_i*'s be constant, then for all* $g \in \mathrm{C}(\mathrm{X})$, $\{\mathrm{T}^n g\}$ *converges uniformly to a constant.*

COROLLARY [FLM], [L]. — *Let* R *be a rational function with Julia set* X. Let $\{w_i\}_{i=1}^{\mathrm{N}}$ *be the inverse branches of* R. *Then, with* $p_i = \mathrm{N}^{-1}$ *for all* i, $\{\mathrm{T}^n f\}$ *converges uniformly to a constant for all* $f \in \mathrm{C}(\mathrm{X})$.

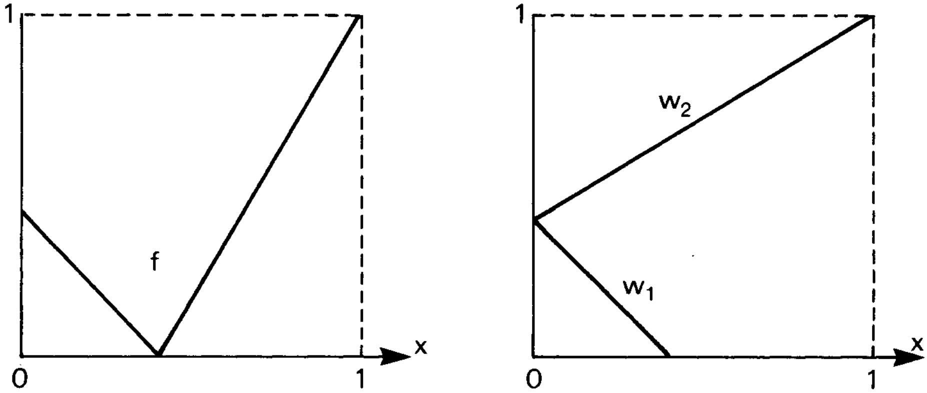
Proof. — That T is weakly almost periodic follows from Ljubich's Lemma 1, p. 359. The result now follows from Proposition 3.8. □

The main result of [FLM] actually says that $\{\mathrm{T}^{*n} \nu\}$ converges weak * to a unique measure for any starting measure and that the limit measure has support on the Julia set. But attractiveness to the Julia set is really a consequence of the basic properties of Julia sets.

4. — EXAMPLES

Let $f : [0, 1] \to [0, 1]$ be piecewise differentiable with $|f'| > 0$. Let continuous branches of the inverse of f be $\{ w_i(x) : i = 1, 2, \ldots, n \}$. Extend the domain of each w_i from its natural one, denoted D_i, to all of [0, 1], by making w_i constant on each component of $[0, 1] \setminus \mathrm{D}_i$ in such a way that $w_i : [0, 1] \to [0, 1]$ is continuous. Then w_i may have distinct left and right derivatives at the end points of D_i: for specificity we define $w_i'(x)=0$ for $x \in D_i'$ and $w_i'(x)$ is its value on its domain from left or right.

Example:



LEMMA 4.1. — *f admits an absolutely continuous Borel measure* μ *of the form*

$$\mu(B)=\int_B \rho(x)\,dx, \quad \textit{for all Borel sets } B, \text{ where } \rho \in L_1,$$

if and only if

$$\rho(x)=\sum_{\substack{i=1\\ \{i : x \in D_i\}}}^{n} \rho(w_i(x))\,|w_i'(x)| \textit{ for almost all } x \in [0, 1] \quad (1)$$

Proof. — We first show that absolute continuity implies (1). The invariance of μ means

$$\mu(B)=\mu(f^{-1}(B)) \quad \text{for all Borel sets B;}$$

whence, by changing variables,

$$\int_B \rho(x)\,dx=\sum_i \int_{w_i(b)} \rho(x)\,dx=\int_B \sum_i \rho(w_i(x))\,|w_i'(x)|\,dx,$$

and so

$$\int_B (\rho(x)-\sum_i \rho(w_i(x))\,|w_i'(x)|\,dx=0$$

which implies (1). Conversely, starting from the last inequality here we deduce the first one. □

For example, with $f(x)=4x(1-x)$ on [0, 1] we can have

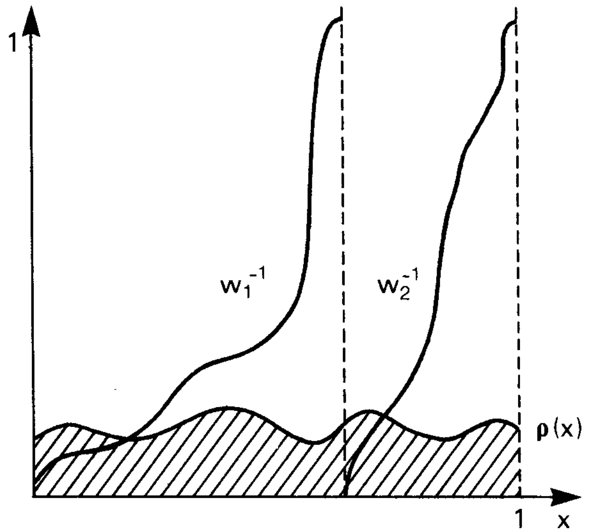
$$\rho(x)=(1/\pi\sqrt{x(1-x)});$$

with $f(x)=7x \bmod 1$ we can have $\rho(x)=1$; and with $f(x)=1/x-[1/x]$ (on [0, 1]) we can have $\rho(x)=1/(1+x)$. This last example is not strictly speaking within the framework because it is not piecewise differentiable, possessing infinitely many discontinuities, but nonetheless, we include it because all of the formalism still applies.

We are interested in what happens when such dynamical systems are "run backwards". We show that the measure $\mu = \int \rho\, dx$ is invariant for the iterated function system $\{ w_i(x) : i = 1, 2, \ldots, n \}$ with place dependent probabilities

$$p_i(x) = \rho(w_i(x)) \,|\, w_i'(x) \,|\, / \rho(x), \tag{2}$$

provided $\rho(x) > 0$. First we argue the result by recourse to a picture, then we formalize. The Figure shows $f : [0, 1]$ consisting of two continuous



components w_1^{-1} and w_2^{-1}, together with the probability density $\rho(x)$ which is invariant under f. Imagine that the dynamical system is running, and that a point has just arrived in a certain interval $[y, y + \delta y]$. How did it get there? Either it came from $[w_1(y), w_1(y) + \delta w_1(y)]$ or from $(w_2(y), w_2(y) + \delta w_2(y)]$. Consequently, if when we are at y, we send next to $w_1(y)$ with relative probability $\rho(w_1(y)) \,|\, w_1'(y) \,|$ and to $w_2(y)$ with relative probability $\rho(w_2(y) \,|\, w_2'(y) \,|$, then the probability density $\rho(x)$ should be preserved on running the system backwards. Because of (1) the correct normalization factor for the probabilities is simply $\rho(x)$, yielding (2).

Next we formalize. Recall that a measure $\tilde{\mu}$ is invariant for the i. f. s. $\{ w_i : i = 1, 2, \ldots, n \}$ with place dependent probabilities $p_i(x)$ if and only if

$$\int_0^1 h(x)\, d\tilde{\mu}(x) = \int_0^1 \sum p_i(x)\, h(w_i(x))\, d\tilde{\mu}(x) \text{ for all } h \in L_1.$$

What is the analogue of Lemma 4.1, when $\tilde{\mu}$ is absolutely continuous?

LEMMA 4.2. — *The i. f. s.* $\{w_i\}$ *with place dependent probabilities* $\{p_i\}$ *admits an absolutely continuous Borel measure* $\tilde{\mu}$ *of the form*

$$\tilde{\mu}(B)=\int_B \tilde{\rho}(x)\,dx, \quad \textit{for all Borel sets } B, \textit{ where } \tilde{\rho}\in L_1,$$

if and only if

$$\tilde{\rho}(x)=\sum_{\{i:\,x\in R_i;\,|(w_i^{-1}(x))'|\neq\infty\}} p_i(w_i^{-1}(x))\,|(w_i^{-1}(x))'|\,\tilde{\rho}(w_i^{-1}(x))$$

$$\textit{for almost all } x\in[0,1], \quad (3)$$

where R_i = *range of* w_i.

Proof. — Again this is just an exercise in change of variables. Choosing h above to be the characteristic function of B we obtain

$$\int_B \tilde{\rho}(x)\,dx=\int_0^1 \sum p_i(x)\,\chi_B(w_i(x))\,\tilde{\rho}(x)\,dx$$

$$=\int_B \sum_{\{i:\,x\in R_i;\,|(w_i^{-1}(x))|\neq\infty\}} p_i(w_i^{-1}(x))\,|(w_i^{-1}(x))'|\,\tilde{\rho}(w_i^{-1}(x))\,dx. \quad \square$$

THEOREM 4.3. — *Let* $\mu=\int \rho\,dx$ *be invariant for* f, *as in Lemma* 4.1, *with* $\rho(x)>0$ *for all* $x\in[0,1]$. *Then* μ *is invariant for the i. f. s.* $\{w_i\}$, *constructed from the branches of* f^{-1} *as above, with place dependent probabilities*

$$p_i(x)=\rho(w_i(x))\,|w_i'(x)|/\rho(x), \qquad i=1,2,\ldots,n.$$

Proof. — We verify the condition of Lemma 4.2. Substituting into the r. h. s. of (3), with $\tilde{\rho}(x)=\rho(x)$ we obtain

$$\text{r. h. s.}=\sum_{\{i:\,x\in R_i;\,(w_i^{-1}(x))'\neq\infty\}} \frac{\rho(x)\,|w_i'(w_i^{-1}(x))|}{\rho(w_i^{-1}(x))}$$

$$\times\,|(w_i^{-1}(x))'|\,\rho(w_i^{-1}(x))=\rho(x) \quad \text{for almost all } x,$$

where we use $w_i'(w_i^{-1}(x))\,(w_i^{-1}(x))'=1$ and the fact that the R_i's intersect on a set of measure zero. $\square$

An example is $f(x)=2x \bmod 1$ with $\rho(x)=1$. We find $w_1(x)=\frac{1}{2}x$ and

$$w_2(x)=\frac{1}{2}x+\frac{1}{2} \text{ with } p_1(x)=\frac{1}{2}=p_2(x).$$

Another example is $f(x)=4x(1-x)$ with $\rho(x)=1/\pi\sqrt{x(1-x)}$. A short calculation shows $w_1(x)=(1-\sqrt{1-x})/2$, $w_2(x)=(1+\sqrt{1+x})/2$ and

$p_1(x)=p_2(x)=1/2$. (More generally, this explains why on any Julia set the natural invariant measure for forward iteration measure is exactly the same as the equilibrium measure obtained by random iteration of the inverse branches, with equal weights on the branches.) For the Gauss measure associated with $f(x)=\frac{1}{x}-\left[\frac{1}{x}\right]$ we find for $n=1, 2, 3, \ldots$ $p_n(x)=(1+x)/(n+1+x)(n+x)$, $w_n(x)=1/(n+x)$, $\rho(x)=1/(1+x)$.

Here, we remark on the Lasota-Yorke theorem [LY]. If f is piecewise C^2 and $\inf|f'(x)|>1$ then there exists an absolutely continuous invariant measure $\mu=\int \rho\, dx$, with $\rho \in L_1$ and ρ is of bounded variation. This theorem does not ensure $\rho(x)>0$ for (almost) all x.

We conclude with two further examples. In the first we consider a smokestack located at the origin giving off sulphur dioxide (SO_2) at a steady rate. A simple model for the resulting time-averaged steady state two dimensional distribution of SO_2 is as follows. SO_2 spreads in two dimensions under the actions of various wind directions and speeds, which effect corresponding mappings of a gas particle at x to new locations $w_2(x)$, $w_3(x), \ldots, w_n(x)$ with probabilities $p_2(x), p_3(x), \ldots, p_n(x)$, determined by the local geography (mountain ranges, lakes, etc.). With probability $p_1(x)$ the particle at x is lost from the system by absorption through rain to the ground or other atmospheric reaction processes. Since we are looking for steady state distributions we set $w_1(x)=0$, which causes a new particle to be emitted from the smokestack whenever one is lost. Existence of an invariant measure μ in this set-up shows existence of a "steady state"; and attractiveness of μ implies stability of the steady state. Diverse random emission-diffusion-absorption models of this type can be conceived. Better models would derive from usage of continuous time random iteration processes, which remain to be investigated.

For a second example we consider a dynamical system evolving under competing force laws. Consider an autonomous system in $\mathbb{R}^2$ of the form

$$\frac{d\underline{x}}{dt}=F(\underline{x}, \sigma(\underline{x}, t)), \qquad \underline{x}\in\mathbb{R}^2$$

$$\underline{x}(0)=\underline{x}_0$$

Here $\sigma(\underline{x}, t)\in\{1,2\}$ is a random variable which at time $t=n$ takes value 1 with probability $p_1(x)$ and value 2 with probability $p_2(x)$, and which remains constant throughout the interval $[n, n+1)$, for $n=0, 1, 2, \ldots$. Both $F(\underline{x}, 1)$ and $F(\underline{x}, 2)$ are continuous maps on $\mathbb{R}^2$ into itself. Such an example might be a grandfather clock whose equations of motion are sensitive to the length of the pendulum. Some days the thermostat in the house is set on high, whilst on other days, at the whim of the owner, it is set on low. (In reality the time scales for relaxation and length scales would be small and cause the iterated function system effects to be unmeasurably small; but the clock is a good mechanism for helping envisage the process.)

If the system is observed at time n, its state then being $\underline{x}_n$ (position and velocity of pendulum), and if $w_k(\underline{x}_0)$ is the state at time 1 when $\sigma(\underline{x}_0, 0)=k$, $k\in\{1, 2\}$; then $w_k: \mathbb{R}\to\mathbb{R}$ is a diffeomorphism and the x_n's will be distributed on the attractor of the iterated function system $\{\mathbb{R}, w_k: k=1,2\}$ with place dependent probabilities $p_1(x)$ and $p_2(x)$.

REFERENCES

[B] H. Brolin, Invariant Set Under Iteration of Rational Functions, *Arkiv für Matematik*, Vol. **6**, 1965, pp. 103-144.

[BD] M. F. Barnsley and S. Demko, Iterated Functions Systems and the Global Construction of Fractals, *Proc. R. Soc. London*, Vol. **A 399**, 1985, pp. 243-275.

[BGM] D. Bessis, J. S. Geronimo and P. Moussa, Function Weighted Measures and Orthogonal Polynomials on Julia Sets, *Const. Approx.* (to appear).

[BE] M. F. Barnsley and J. Elton, Stationary Attractive Measures for a Class of Markov Chains Arising from Function Iteration, *Adv. Appl. Prob.*, March 1988 (to appear).

[BM] R. R. Bush and F. Mosteller, A Stochastic Model with Applications to Learning, *Ann. Math. Statist.*, Vol. **24**, 1953, pp. 449-585.

[DHN] S. Demko, L. Hodges and B. Naylor, Construction of Fractal Objects with Iterated Function Systems, *SIGGRAPH '85 Proceedings*, 1985, pp. 271-278.

[DF] L. Dubins and D. Freedman, Invariant Probabilities for Certain Markov Processes, *Ann. Math. Stat.*, Vol. **37**, 1966, pp. 837-848.

[DS] P. Diacones and M. Shahshahani, Products of Random Matrices and Computer Image generation, in *Random Matrices and Their Applications*, Vol. **50**, Cont. Math. A.M.S., Providence, R.I., 1986.

[DoF] W. Doeblin and R. Fortet, Sur des chaînes à liaisons complètes, *Bull. Soc. Math. de France*, Vol. **65**, 1937, pp. 132-148.

[DuS] N. Dunford and J. Schwartz, *Linear Operators*, Wiley, New York, 1958.

[DU] J. Diestel and J. J. Uhl, *Vector Measures*, American Mathematical Society, 1977.

[E] J. H. Elton, An Ergodic Theorem for Iterated Maps, *Ergod. Th. and Dyna Sys.*, Vol. **7**, 1987 (to appear).

[FLM] A. Freire, A. Lopes and R. Mañé, An Invariant Measure for Rational Maps, *Bol. Soc. Bras. Mat.*, Vol. **14**, 1983, pp. 45-62.

[G] F. R. Gantmacher, *The Theory of Matrices*, Chelsea, New York, 1977.

[H] J. Hutchinson, Fractals and Self-Similarity, *Indiana U. J. of Math.*, Vol. **30**, 1981, pp. 713-747.

[HS] E. Hewitt and K. Stromberg, *Real and Abstract Analysis*, Springer-Verlag, New York, 1969.

[IT] C. Iionescu Tulcea, On a Class of Operators Occurring in the Theory of Chains of Infinite Order, *Can. J. Math.*, Vol. **11**, 1959, pp. 112-121.

[ITM] C. Ionescu and G. Marinescu, Sur certaines chaînes à liaisons complètes, *C. R. Acad. Sci. Paris*, T. **227**, 1948, pp. 667-669.

[J] B. Jamison, Irreducible Markov Operators on C(S), *Proc. Am. Math. Soc.*, Vol. **24**, 1970, pp. 366-370.

[JS] B. JAMISON and R. SINE, Irreducible Almost Periodic Markov Operators, *J. Math. Mech.*, Vol. **18**, 1969, p. 1043-1057.

[K] S. KARLIN, *Some Random Walks Arising in Leaening Models, Pac. J. Math.*, Vol. **3**, 1953, pp. 725-756.

[Kr] U. KRENGEL, *Ergodic Theorems*, de Gruyter, Berlin, 1985.

[L] M. J. LJUBICH, Entropy Properties of Rational Endomorphisms of the Riemann Sphere, *Ergod. Th. Dynam. Sys.*, Vol. **3**, 1983, pp. 351-385.

[Lor] G. G. LORENTZ, *Approximation of Functions*, Holt, Reinhardt, and Winston, N. Y., 1966.

[LY] A. LASOTA and J. A. YORKE, On the Existence of Invariant Measures for Piecewise Monotonic Transformations, *Trans. Am. Math. Soc.*, Vol. **186**, 1973, pp. 481-488.

[N] E. NUMMELIN, *General Irreducible Markov Chains and Non-Negative Operators*, Cambridge U. Press, Cambridge, 1985.

[OM] O. ONICESCU and G. MIHOC, Sur les chaînes de variables statistiques, *Bull. Soc. Math. de France*, Vol. **59**, 1935, pp. 174-192.

[R] M. ROSENBLATT, Equicontinuous Markov Operators, *Theor. Prob. and its Appl.*, Vol. **9**, 1964, pp. 180-197 (translation of *Theor. Verojatnost. i Primenen*).

[S] R. SINE, Convergence Theorems for Weakly Almost Periodic Markov Operators, *Israel J. Math.*, Vol. **19**, 1974, pp. 246-255.

(*Manuscrit reçu le* 5 *Septembre* 1986)
révisé le 9 *Octobre* 1987.)