PRODUCTS OF RANDOM MATRICES AND
COMPUTER IMAGE GENERATION

BY

PERSI DIACONIS AND MEHRDAD SHAHSHAHANI

TECHNICAL REPORT NO. 228
NOVEMBER 1984

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT MCS80-24649

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
PRODUCTS OF RANDOM MATRICES AND
COMPUTER IMAGE GENERATION

BY
Persi Diaconis and Mehrdad Shahshahani

TECHNICAL REPORT NO. 228
November 1984

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT MCS80-24649

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
PRODUCTS OF RANDOM MATRICES AND
COMPUTER IMAGE GENERATION

Persi Diaconis
Department of Statistics
Stanford University
Stanford, California

Mehrdad Shahshahani
Boeing Aerospace
Seattle, Washington

§1 Introduction.

Fractals as models for certain natural phenomena were defined and popularized by Mandelbrot ([M]). Subsequently, Mandelbrot and others produced realistic pictures of mountains, landscapes, clouds etc. using fractal methods. In [H], Hutchinson proposed a beautiful mathematical framework for the study of fractals (see §2). We noticed that this method can be used for computer generation of pictures of certain natural objects while achieving substantial data compression. For example, the poplar tree below was generated by storing only 52 numbers. The question naturally arose as to whether the fixed-point method of Hutchinson can be used in a systematic fashion for computer image generation and data encoding. It became clear that further theoretical development is necessary to apply this procedure in a systematic manner. In this paper we give a brief account of some of our results in this direction.
We conclude this discussion with a brief description of the algorithm that produced Figure 1. The algorithm uses as input n affine transformations \( \{S_1, \ldots, S_n\} \). Here \( S_i x = A_i x + c_i \) with \( x \) and \( c_i \in \mathbb{R}^2 \) and \( A_i \) a 2x2 matrix with \( |A_i x| < |x| \) for all \( x \neq 0 \). Such a transformation has a unique fixed point; from \( S_i x = x \) we have \( x = (I - A_i)^{-1} c_i \). The first stage of the algorithm plots
the n fixed points. The second stage plots the fixed points of all \( n^2 \) products \( S_i S_j \). At stage \( k \), all \( n^k \) fixed points of products of length \( k \) are plotted. The algorithm terminates when sufficiently many points have been plotted.

For the picture in Figure 1, the leaf is based on \( n = 4 \), \( k = 7 \) and for the stem we have \( n = 3 \), \( k = 7 \).

§2 Fractals.

We first review the relevant aspects of the mathematical framework proposed by Hutchinson [H] for the study of fractals. Let \( \Delta = \{ S_1, \ldots, S_n \} \) be a finite set of contracting affine transformations of \( \mathbb{R}^d \). Then every transformation \( S_{i_1} \ldots S_{i_k} \) has a unique fixed point which we denote by \( F_{i_1 \ldots i_k} \). The fractal object associated to \( \Delta \) is

\[
F[\Delta] = \text{closure} \{ F_{i_1 \ldots i_k} \mid \text{all } k \text{ and all } i_1, \ldots, i_k \}.
\]

This is the "limiting picture" produced by the algorithm of Section 1.

THEOREM 2.1 - '(H).

(a) \( F[\Delta] \) is the unique compact subset \( K \) of \( \mathbb{R}^d \) with the property

\[
K = S_1(K) \cup \cdots \cup S_n(K).
\]

Let

\[
\Omega = \Pi\{1, \ldots, n\}
\]

be the set of all mappings of \( \mathbb{N} \) into the finite set \( \{1, \ldots, n\} \) equipped with the product topology.

(b) For an element \( w = (i_1, i_2, \ldots) \in \Omega \) and any \( x \in \mathbb{R}^d \),

\[
\lim_{k \to \infty} S_{i_1} S_{i_2} \ldots S_{i_k}(x)
\]
exists and is independent of \( x \). We denote by \( \pi(w) \) this limiting value, and so we have a map

\[
\pi: \Omega \rightarrow \mathbb{R}^d .
\]

(c) The map \( \pi \) is continuous and

\[
\text{Im} \; \pi = F[\Delta] .
\]

It is not difficult to see

Corollary - \( F[\Delta] \) is a perfect set unless all \( S_j \)'s have the same fixed point \( F \), in which case \( F[\Delta] = \{ F \} \).

It is clear that understanding the structure of the set \( F[\Delta] \) is a fundamental problem in this area. There is an important special case where we can say more about \( F[\Delta] \).

Condition SC - We say condition SC (strongly contracting) holds if there is an open set \( U \subset \mathbb{R}^d \) and disjoint compact neighborhoods \( V_1, \ldots, V_n \) of \( F_1, \ldots, F_n \) such that

\[
S_j(U) \subset V_j, \quad \bigcup_{j=1}^{n} V_j \subset U
\]

THEOREM 2.2 - Assume (SC) holds.

(a) \( \pi \) is a homeomorphism onto \( F[\Delta] \).

(b) Suppose furthermore \( S_j \)'s are similitudes (so \( S_j = r_jA_j + c_j \) with \( A_j \) orthogonal) with contraction rates \( r_j \) respectively. Then

Hausdorff dimension \( F[\Delta] = D \)

where \( D \) is the unique real number such that

\[
\sum r_j^D = 1 ,
\]

((b) is proved in [H].)
Let \((X,d)\) be a compact metric space and \(\varepsilon > 0\). Define

\[N(\varepsilon) = \text{smallest number of balls of radius} \ \varepsilon \ \text{covering} \ X\]

\[M(\varepsilon) = \text{largest integer} \ m \ \text{such that there are} \ m \ \text{points} \ x_1, \ldots, x_m \in X \]

\[\text{with} \ d(x_i, x_j) \geq \varepsilon \ \text{for} \ i \neq j.\]

The limits

\[\lim_{{\varepsilon \to 0}} \frac{\log M(\varepsilon)}{-\log \varepsilon} = C, \quad \lim_{{\varepsilon \to 0}} \frac{\log N(\varepsilon)}{-\log \varepsilon} = H,\]

if they exist, are called \textit{capacity} and \textit{metric entropy} of \(X\). Let \(\rho_1\) be the operator norm of \(S_1\), and \(\eta_1\) be the largest number such that

\[|A_1 x| \geq \eta_1 |x|\]

for all \(x \in \mathbb{R}^d\). Set

\[\rho = \max(\rho_1, \ldots, \rho_n),\]

\[\eta = \min(\eta_1, \ldots, \eta_n).\]

For the metric space \(F[\Delta]\) with induced metric from \(\mathbb{R}^d\) we show

\textbf{Proposition 2.3.}

(a)

\[\lim_{-\log \varepsilon} \frac{\log N(\varepsilon)}{-\log \varepsilon} < \frac{\log n}{-\log \rho}.\]

(b) If condition (SC) holds, then

\[\frac{\log n}{-\log \eta} < \lim_{-\log \varepsilon} \frac{\log M(2\varepsilon)}{-\log \varepsilon}.\]

Furthermore, if \(S_j\)'s are contractions with the same contraction rate \(\rho\), then the capacity and metric entropy of \(F[\Delta]\) exist and are equal to the Hausdorff dimension.
$\log n \\
-\log \rho$.

§3 The Associated Markov Chain.

Let $p_1, \ldots, p_n$ be positive numbers with $\sum p_j = 1$, and $P$ be the probability measure on the affine group $\Lambda(d) = \mathbb{R}^d$, $\text{GL}(d, \mathbb{R})$ (semi-direct product) supported on $\Delta = \{S_1, \ldots, S_n\}$ and assigning mass $p_j$ to $S_j$. Consider the Markov chain $\{X_n\}_{n=0,1,\ldots}$ on $\mathbb{R}^d$ determined by the starting state $X_0 = x_0$ and

$$X_{n+1} = A_{n+1} X_n + c_{n+1} \quad n \geq 0$$

with $(c_n, A_n)$ i.i.d. from $P$. The basic properties of this Markov are summarized in

THEOREM 3.1. (See [DB] and [H]).

(a) The Markov chain $\{X_n\}_{n=0,1,\ldots}$ has a unique stationary distribution represented by the sum

$$S = c_1 + A_1 c_2 + A_1 A_2 c_3 + \ldots$$

with $(c_n, A_n)$ i.i.d. from $P$.

(b) The law of $\mu$ is of pure type. It is continuous unless all $S_j$'s have the same fixed point $F$, in which case $S$ is atomic concentrated at $F$.

(c) The law of $S$ satisfies

$$E[f(S)] = \int E[f(AS+c)] P(dc, dA)$$

for every Borel function $f$.

(d) $\text{Supp}(\mu) = F[\Delta]$.

The stationary distribution $\mu$ is a mathematical representation of the picture we see on the monitor. To make this relationship precise, let $E_k$ denote
the empirical measure of the set of fixed points of all words of length \( k \) (counted with multiplicities). Then it is not difficult to show

**Proposition 3.2.**

(a) \( \mathbb{E}_k \) converges to \( \mu \) in the weak star topology.

(b) \( \bigcup_{j=1}^{k} \text{Supp } E_j \) converges to \( \text{supp}(\mu) \) in the Hausdorff metric, and \( \text{supp}(\mu) \) varies continuously with \( (c_i, A_i) \).

We equip \( \Omega \) with the usual cylinder measure \( \lambda \) which is the unique exchangeable probability measure on \( \Omega \) where

\[
\lambda((j_1, \ldots, j_k, \star \ldots \star)) = p_{j_1} \cdots p_{j_k}.
\]

The following theorem is a version of the law of large numbers; it implies that the points hit by the Markov chain can be approximated by \( \mu \).

**THEOREM 3.3** Assume condition (SC) holds, and \( B \subset \mathbb{R}^d \) is an open set satisfying

\[
\tilde{B} \cap F[\Delta] = B \cap F[\Delta].
\]

Let \( x \in \mathbb{R}^d \) and for \( \omega = (i_1, i_2, \ldots) \in \Omega \) let

\[
N(\omega, k) = \text{Cardinality}\{(i_1, \ldots, i_j) \mid j \leq k, S_{i_1} \cdots S_{i_j} (x) \in B\}.
\]

Then there is \( \Omega' \) of \( \lambda \)-measures 1, and independent of \( B \), such that for all \( \omega \in \Omega' \)

\[
\mu(B) = \lim_{k \to \infty} \frac{N(\omega, k)}{k}.
\]

This theorem enables us to give a dynamical systems interpretation to the Markov chain. Assume condition (SC) holds and choose \( \omega = (i_1, i_2, \ldots) \) for which the conclusion of the theorem holds. Consider the time dependent difference scheme

\[
T_j(x) = S_{i_j} (x).
\]

Then \( F[\Delta] \) is the (strange) attractor for this dynamical system. This also
gives a partial answer to a question raised by Cuckeheimer about what kind of fractals can appear as strange attractors, [Gu].

It is instructive to compare Theorem 3.3 with conclusions easily derived from the ergodic theorem. The ergodic theorem implies that for almost all \( x \) in the support of \( \mu \) and almost all sample paths, the empirical measure converges weak* to \( \mu \). Theorem 3.3 concludes this for any starting state under the hypothesis (SC). Consideration of simple examples such as \( x \to \frac{1}{3} x, \ x \to \frac{1}{3} x + \frac{2}{3} \) shows that points outside \( F[\Delta] \) may never get into \( F[\Delta] \).

Understanding the stationary measure is a fundamental problem. Clearly if (SC) holds then \( \mu \) is singular. There are a number of isolated results regarding absolute continuity of \( \mu \), for example, Erdös [E] has shown

(a) If \( d = 1, n = 2, p_1 = p_2 = \frac{1}{2} \),

\[
S_1(x) = ax+1, \quad S_2(x) = ax-1
\]

and \( a = 1/\zeta \) where \( \zeta \) is a P.V. number, then \( \mu \) is singular continuous. (There are infinity of P.V. numbers in the interval \((1,2)\).)

(b) For almost all \( a \) sufficiently close to 1, \( \mu \) is absolutely continuous. See also [G] for conditions ensuring absolute continuity of \( \mu \).

The techniques of Erdös and Garcia can be generalized to higher dimensions to yield some conditions for singularity or absolute continuity for \( \mu \). This gives a nice application: mathematics invented for a very different purpose being used for a problem in computer graphics.

§4 Inverse Fractal Problem.

For the practical problem of image generation and data encoding the inverse fractal problem is the most fundamental question. In its simplest form it may be formulated as follows:
IFPI - Given a compact set $X \subset \mathbb{R}^d$, does there exist a finite set $\Delta = \{S_1, \ldots, S_n\}$ of contracting affine transformations such that $F[\Delta]$ approximates $X$ to given accuracy, say in the Hausdorff metric?

If

$$\lim_{\varepsilon \to 0} \frac{N(\varepsilon)}{-\log \varepsilon} = H$$

exists, then it is easy to see that for every $\delta > 0$, a $\Delta$ exists such that

$$d(X, F[\Delta]) < \delta.$$ 

However, in this solution the cardinality of $\Delta$ is of the order of $\delta^{-H}$ as $\delta \to 0$, so that such a solution has small practical implication. For the special case of a convex polytope one has the following simple solution:

**Proposition 4.1.** Let $X$ be a convex polytope with $n$ vertices. Then there is a set $\Delta = \{S_1, \ldots, S_n\}$ of $n$ contracting affine transformations such that

$$X = F[\Delta].$$

In this proposition we take $S_j$'s to be of the form

$$S_j(x) = \alpha_j I x + c_j \quad \alpha_j \in \mathbb{R}, \quad c_j \in \mathbb{R}^d$$

with $\alpha_j$'s sufficiently close to 1 and $\{F_j\}$ is the set of vertices of the convex polytope $X$. Then, the polytope becomes the attractor $K$ of Theorem 2.1 which yields the desired result.

If we replace $\alpha_j$'s by general linear transformations, then we do not know of any simple extension of Proposition 4.1. However, it is important to note that the fractal object $F[\Delta]$ changes continuously as we vary the parameters. This fact is quite significant in the practical applications of the fixed point method. Furthermore, we can use the Markov chain introduced in §3 to propose a solution to the following more precise reformulation of IFPI:

IFP2 - Given a compactly supported probability measure $\alpha$ on $\mathbb{R}^d$, does there exist a finite set $\Delta = \{S_1, \ldots, S_n\}$ of contracting affine transformations such
that the stationary measure associated to $\Delta$ (for some fixed $p_1, \ldots, p_n$) approximates $\alpha$ to given accuracy, say in a Vasershtein metric?

For simplicity of notation, we describe our proposed solution to this problem only for $d = 1$. Note that Theorem 3.1 (c) is equivalent to

$$\int f(x)d\mu(x) = \sum_{j} p_j \int f(S_j x)d\mu(x)$$

where $\mu$ is the stationary measure. Substituting

$$f(x) = x^k$$
in (*) we see that

$$1.\text{h.s.}(\ast) = m_k(\mu)$$
i.e., $k^{th}$ central moment of $\mu$. Evaluating r.h.s. $(\ast)$, setting it equal to l.h.s. and simplifying we get

$$m_k(\mu) = \frac{\sum_{j=1}^{n} p_j \sum_{r=1}^{k} \binom{k}{r} a_j^{k-r} c_j^r m_{k-r}(\mu)}{1 - \sum_{j=1}^{n} p_j a_j^k}$$

(\ast\ast)

Now given a compactly supported probability measure $\alpha$ on $\mathbb{R}$, we calculate its first $2k (k > 1)$ moments. Then in principle we can solve the system of equations (\ast\ast) to obtain $a_j$'s and $c_j$'s, so $k$ affine transformations whose stationary measure has the same first $2k$ moments as $\alpha$. It is not difficult to show that relative to the Vasershtein metric (see e.g. [BF] for definition and basic properties), the distance between $\alpha$ and $\mu$ tends to 0 as $\frac{c}{\nu^k}$.

The calculation above can be extended to higher dimensions, but the algebraic manipulations are significantly more complex. A similar method for iteration of maps had been developed earlier by M. Barnsley and Demko [B]. The applicability of this procedure to actual problems has not yet been fully tested,
§5 Some Open Problems.

Here we briefly indicate a few open problems.

(a) As noted earlier, under rather restrictive hypothesis one can establish the existence of capacity and metric entropy for the fractal $F[\Delta]$. One may conjecture that these quantities always exist for such sets. A number of numerical invariants, such as information dimension, are defined in [FOY]. Computing these invariants for $F[\Delta]$ is an interesting problem.

(b) In view of Theorem 2.2(a), if condition (SC) holds, then $F[\Delta]$ is totally disconnected. Little is known about the topological structure of $F[\Delta]$. Consider for example the case of 3 contracting affine transformations $S_1, S_2, S_3$ of $\mathbb{R}^2$. Write

$$S_j(x) = A_j x + c_j$$

and assume $S_j$'s have distinct fixed points $\{F_j\}$ not lying on the same line. If

$$A_j x = \alpha_j x$$

and

$$\Sigma \alpha_j \geq 2$$

then $F[\Delta]$ is the triangle $E$ with vertices $F_1, F_2, F_3$. If $0 < \alpha_j < 1/2$, then condition (SC) is satisfied and so $F[\Delta]$ is totally disconnected. We can show that in intermediate cases where $\alpha_j$'s are such that $\alpha_j + \alpha_k > 1$ for all $j, k$, and $\alpha_1 + \alpha_2 + \alpha_3 < 2$, $F[\Delta]$ is path-wise connected, and its first homotopy group $\pi_1(F[\Delta])$ is infinitely generated. Clearly one may conjecture many generalizations of this result.

(c) The stationary measure $\mu$ is a fundamental quantity both from the theoretical and practical standpoints. Unfortunately, we only know of some
isolated results about the nature of $\mu$. One possible approach for establishing absolute continuity of $\mu$ in some cases is as follows: Define the operator $T$ on $L^1(\mathbb{R}^d)$ by

$$T(\phi)(x) = \sum p_j(\det A_j) \phi(S_j x)$$

where $S_j x = A_j x + c_j$. Given any probability density $\phi$, the average

$$\frac{1}{m} \sum_{j=0}^{m-1} T^j(\phi) = F_m(\phi)$$

converges weakly to the stationary distribution $\mu$. If one shows that the sequence $\{F_m(\phi)\}$ is of bounded variation, then one can deduce absolute continuity of $\mu$. This idea has been successfully used in the theory of iteration of maps (see e.g. [LY]), and so one may expect it to be applicable to this case too.

Arguments based on Fourier analysis (such as in [E1] and [E2]) are another possible approach to understanding the nature of the measure $\mu$.

REFERENCES


