*Probability in the Engineering and Informational Sciences*, 2, 1988, 387–414. Printed in the U.S.A.

# MIXING MARKOV CHAINS AND THEIR IMAGES

MICHAEL F. BARNSLEY
*School of Mathematics*
*Georgia Institute of Technology*
*Atlanta, Georgia 30332*

MARC A. BERGER
*Department of Theoretical Mathematics*
*Weizmann Institute of Science*
*Rehovot 76100, Israel*

*and*

*Department of Mathematics*
*Carnegie-Mellon University*
*Pittsburgh, Pennsylvania 15213*

H. METÉ SONER
*Department of Mathematics*
*Carnegie-Mellon University*
*Pittsburgh, Pennsylvania 15213*

Recently, orbits of two-dimensional Markov chains have been used to generate computer images. These chains evolve according to products of i.i.d. affine maps. We deal with mixing models, whereby one mixes together several of these Markov chains, so as to create a mixed image. These mixtures involve starting one Markov chain off at the stationary distribution of another, and then running it for a geometrically distributed number of steps. We use this to analyze various mixing scenarios.

Research supported by DARPA Contract No. N00014-86-C-0446 (M.F. Barnsley) and by AFOSR Grant No. AFOSR-87-0137 (M.A. Berger).

 0269-9648/88 $5.00 + .00

## 1. INTRODUCTION

Let $T$: $\mathbb{R}^d \to \mathbb{R}^d$ be a (nonrandom) strictly contractive map. Then it has a unique fixed point, and starting at any $x_0$ the iterates $x_n = T^n x_0$ converge to this fixed point. In [6] and [7], a stochastic analogue of this is examined, where $\mathfrak{J}$: $\mathbb{R}^d \to \mathbb{R}^d$ is a *random* affine map, $\mathfrak{J}$: $x \mapsto \mathfrak{A}x + b$. Here $\mathfrak{A} = \mathfrak{A}(\mathfrak{J})$ and $b = b(\mathfrak{J})$ denote the random matrix and random vector parts of $\mathfrak{J}$, respectively. Theorem 2 of [6] asserts that if $\mathbb{E}\log^+\|\mathfrak{A}\| < \infty$, $\mathbb{E}\log^+|b| < \infty$, and if the random matrix $\mathfrak{A}$ has a negative Lyapunov exponent, then $\mathfrak{J}$ has a unique fixed distribution $X$ satisfying $\mathfrak{J} * X = X$ (equality in distribution). Furthermore, if $X_0$ is any finite random variable and if $\{\mathfrak{J}_n\}$ is an independent and identically distributed (i.i.d.) sequence of random affine maps distributed like $\mathfrak{J}$, then the Markov chain $X_n = \mathfrak{J}_n \cdots \mathfrak{J}_1 X_0$ converges in distribution to this fixed distribution. This means that the chain $\{X_n\}$ is asymptotically stationary.

This result is the basis of a popular probabilistic algorithm for image generation. The algorithm runs as follows. Let $T_1, \ldots, T_K$: $\mathbb{R}^2 \to \mathbb{R}^2$ be nonrandom affine maps, and assign to them respective weights $p_1, \ldots, p_K$ with $p_i > 0$, $\Sigma p_i = 1$. Initialize $X_0$ and at stage $n + 1$, when $X_0, \ldots, X_n$ have already been determined, randomly pick one of the maps $T_i$ according to the weights $p_i$. Say the chosen map is $T_k$. Apply this map to $X_n$, thereby obtaining the next point $X_{n+1} = T_k X_n$. Once $n$ is sufficiently large, plot the successive points $X_n$. In this way, one obtains an image from the orbit of the Markov chain $\{X_n\}$. To generate a color image, record the frequencies and color the points correspondingly according to the frequencies that $\{X_n\}$ visits them. Figure 1 is a flowchart of this algorithm, and Figure 2 is a schematic used to represent it. For a discussion of this application, see [1–5], [8], [9], and [11]. Observe that the *random* affine map $\mathfrak{J}$ here is the one with atomic distribution $\mathbb{P}(\mathfrak{J} = T_i) = p_i$, $1 \le i \le K$ [12].

The proof in [6] of the result mentioned above, that the chain $X_n = \mathfrak{J}_n \cdots \mathfrak{J}_1 X_0$ converges in distribution to $X$, is based on a reversibility phenomenon. Define $\hat{X}_n = \mathfrak{J}_1 \cdots \mathfrak{J}_n X_0$, with successive maps $\mathfrak{J}_n$ applied from the inside rather than from the outside. Then $\{\hat{X}_n\}$ is not a Markov process, but it has two important properties.

1. For each fixed $n$, $X_n$ and $\hat{X}_n$ have identical distributions. (Note that the *processes* $\{X_n\}$ and $\{\hat{X}_n\}$ have different distributions, in general, owing to the noncommutativity of $d \times d$ matrices.)

2. For any constant initial value $X_0 = x$, $\{\hat{X}_n\}$ converges a.s. to a limit random variable $X$. To see why this is so, observe that

$$\begin{aligned}\hat{X}_{n+1} - \hat{X}_n &= \mathfrak{J}_1 \cdots \mathfrak{J}_n \mathfrak{J}_{n+1} X_0 - \mathfrak{J}_1 \cdots \mathfrak{J}_n X_0 \\ &= \mathfrak{A}_1 \cdots \mathfrak{A}_n (\mathfrak{J}_{n+1} X_0 - X_0),\end{aligned}$$

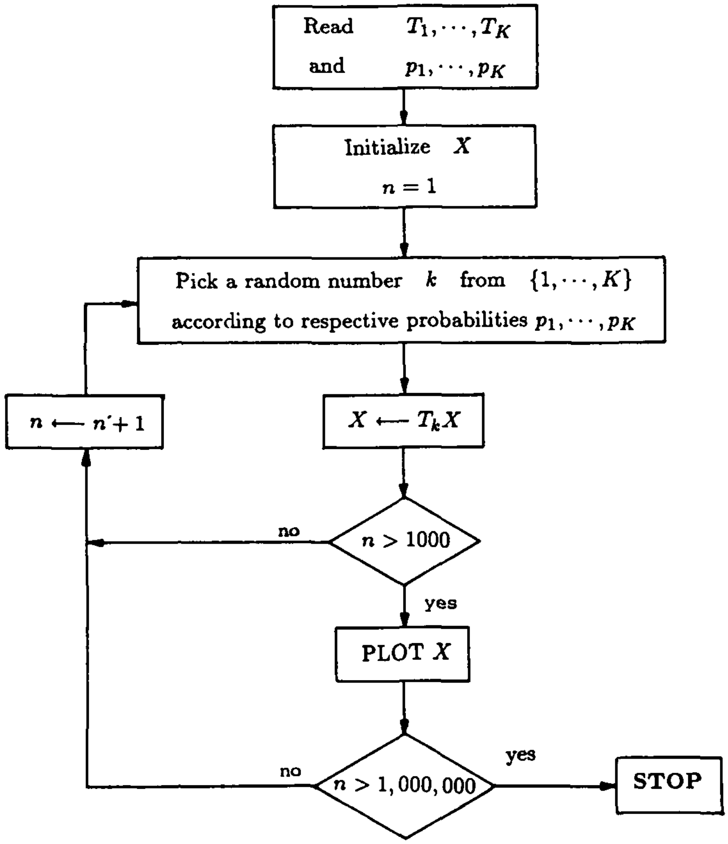



**FIGURE 1.** Basic image generation algorithm. To obtain a color image, increase the frequency count of the pixel that $X$ belongs to every time the PLOT $X$ command is executed. Use a color map to convert frequencies to colors.

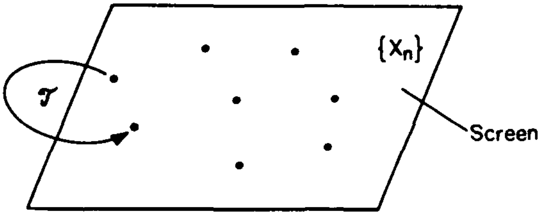



**FIGURE 2.** Schematic of the basic algorithm.

where $\mathfrak{A}_i = \mathfrak{A}(\mathfrak{I}_i)$ is the (random) matrix part of $\mathfrak{I}_i$. Since the Lyapunov exponent of $\mathfrak{A}$ is negative, $\|\mathfrak{A}_1 \cdots \mathfrak{A}_n\|$ behaves like $r^n$ for large $n$, where $r < 1$. If in addition $\mathbb{E} \log^+|\mathfrak{I}X_0 - X_0| < \infty$ (which is the case if $X_0 = x$) then

$$\lim_{n\to\infty} \frac{1}{n} \log^+|\mathfrak{I}_{n+1}X_0 - X_0| = 0,$$

and thus $|\hat{X}_{n+1} - \hat{X}_n|$ also decays geometrically for large $n$. Refer to [6] for details. We shall see, in Sections 2 and 3 below, the occurrence of similar reversibility phenomena for the mixed chains as well.

As mentioned above, in the image generation application $\mathfrak{I}$ has a discrete distribution $\mathbb{P}(\mathfrak{I} = T_i) = p_i$, $1 \le i \le K$. In this case, the stationary condition $\mathfrak{I} * X = X$ becomes

$$\nu(B) = \sum_{i=1}^{K} p_i \nu(T_i^{-1}B) \tag{1}$$

for any Borel set $B \subset \mathbb{R}^d$, where $\nu$ is the distribution of $X$. Let $C = \text{supp}(\nu)$ be the support of $\nu$. Then applying Eq. (1) to $B = C$ and to $B = \cup_{i=1}^{K} \overline{T_i(C)}$, we arrive at the conclusion

$$C = \bigcup_{i=1}^{K} \overline{T_i(C)}. \tag{2}$$

This shows that the maps $T_i$ induce a covering of $C$ by sets $T_i(C)$ which are each affinely similar to $C$. This is a sort of puzzle where the sets $T_i(C)$ are the pieces, but it is different from standard puzzles in that the pieces are allowed to overlap. That is, there is no claim of disjointness on the right-hand side of Eq. (2). For a discussion of the uniqueness of the set $C$ satisfying Eq. (2), see [6, Section 2]. The formula given by Eq. (2) plays a fundamental role in the theory of iterated affine maps. It was first enunciated by Hutchinson [10].

We are concerned in this paper with mixing models, whereby one mixes together the individual Markov chains generated by the random maps $\mathfrak{I}$ and $\mathfrak{S}$. Corresponding to them, there are algorithms (Figures 3,4 and 5,6 below) which produce an image with the textures of the original images infused together. The basic result underlying the mixing is the following.

THEOREM 1: *Let $\{X_n\}$ be a Markov chain with initial distribution $\pi_0$ and transition probabilities $P(x,dy)$. Let $0 < p \le 1$ and let $\sigma$ be a random variable independent of $\{X_n\}$ with distribution $\mathbb{P}(\sigma = k) = pq^k$, $k \ge 0$. Then the distribution $\pi$ satisfies*

$$\pi(dy) = q \int P(x,dy)\pi(dx) + p\pi(dy) \tag{3}$$

*if and only if $\pi$ is the distribution of $X_\sigma$.*

PROOF: Sufficiency. Simply observe that

$$\mathbb{P}(X_\sigma \in B) = q\mathbb{P}(X_{\sigma+1} \in B) + p\mathbb{P}(X_0 \in B).$$

Necessity. Iterate, obtaining

$$\pi(B) = \sum_{k=0}^{n-1} pq^k \mathbb{P}(X_k \in B) + q^n \int P^{(n)}(x,B)\pi(dx),$$

where $P^{(n)}$ is the $n$-step transition probability. ■

## 2. A SIMPLE MIXING EXAMPLE

Let $\{\mathfrak{I}_n\}$ and $\{\mathcal{S}_n\}$ be i.i.d. sequences of random affine maps, each sequence independent of the other, and let $U$ be a fixed (nonrandom) affine map. We want to mix the individual Markov chains generated by the $\mathfrak{I}$ and $\mathcal{S}$ sequences. The $X$ process will be generated exactly as before, i.e.,

$$X_{n+1} = \mathfrak{I}_{n+1} X_n, \qquad n \geq 0, \tag{4}$$

but the $Y$ process is to have a selection mechanism,

$$Y_{n+1} = \begin{cases} UX_n & \text{with probability } p > 0, \\ \mathcal{S}_{n+1} Y_n & \text{with probability } q = 1 - p. \end{cases} \tag{5}$$

The choice in Eq. (5) is to be independent of $\{\mathfrak{I}_n\}$ and $\{\mathcal{S}_n\}$. It can be modeled as an i.i.d. sequence $\{I_n\}$ of Bernoulli 0,1 random variables, with $\mathbb{P}(I = 0) = p$. Then

$$Y_{n+1} = \begin{cases} UX_n & \text{if } I_{n+1} = 0, \\ \mathcal{S}_{n+1} Y_n & \text{if } I_{n+1} = 1. \end{cases} \tag{6}$$

This corresponds to the algorithm and schematic in Figures 3 and 4.

The pair process $\{(X_n, Y_n)\}$ is a Markov chain on $\mathbb{R}^d \times \mathbb{R}^d$. We are concerned with its asymptotic behavior. Let $\sigma$ be a random variable independent of $\{\mathfrak{I}_n\}$ and $\{\mathcal{S}_n\}$ with distribution $\mathbb{P}(\sigma = k) = pq^k$, $k \geq 0$.

THEOREM 2: *Assume that* $\mathbb{E}\log^+\|\mathcal{A}\| < \infty$ *and* $\mathbb{E}\log^+|b| < \infty$, *and that the Lyapunov exponent of* $\mathcal{A}$ *is negative, where* $\mathcal{A} = \mathcal{A}(\mathfrak{I})$ *and* $b = b(\mathfrak{I})$. *Then the chain* $\{(X_n, Y_n)\}$ *defined in Eqs. (4) and (6) is asymptotically stationary, and the limiting distribution of* $(X_n, Y_n)$ *is* $(\mathfrak{I}_1 \cdots \mathfrak{I}_{\sigma+1} * X, \mathcal{S}_1 \cdots \mathcal{S}_\sigma U * X)$, *where* $X$ *is the unique fixed distribution of* $\mathfrak{I}$, $\mathfrak{I} * X = X$. *(If* $\sigma = 0$ *then* $\mathcal{S}_1 \cdots \mathcal{S}_\sigma$ *is interpreted as the identity.)*

PROOF: Define $I_0 = 0$ and set $N(n) = \max(k \leq n\colon I_k = 0)$. Observe that

$$Y_n = \begin{cases} \mathcal{S}_n \cdots \mathcal{S}_1 Y_0 & \text{if } N(n) = 0, \\ \mathcal{S}_n \cdots \mathcal{S}_{N(n)+1} U X_{N(n)-1} & \text{if } N(n) > 0. \end{cases} \tag{7}$$

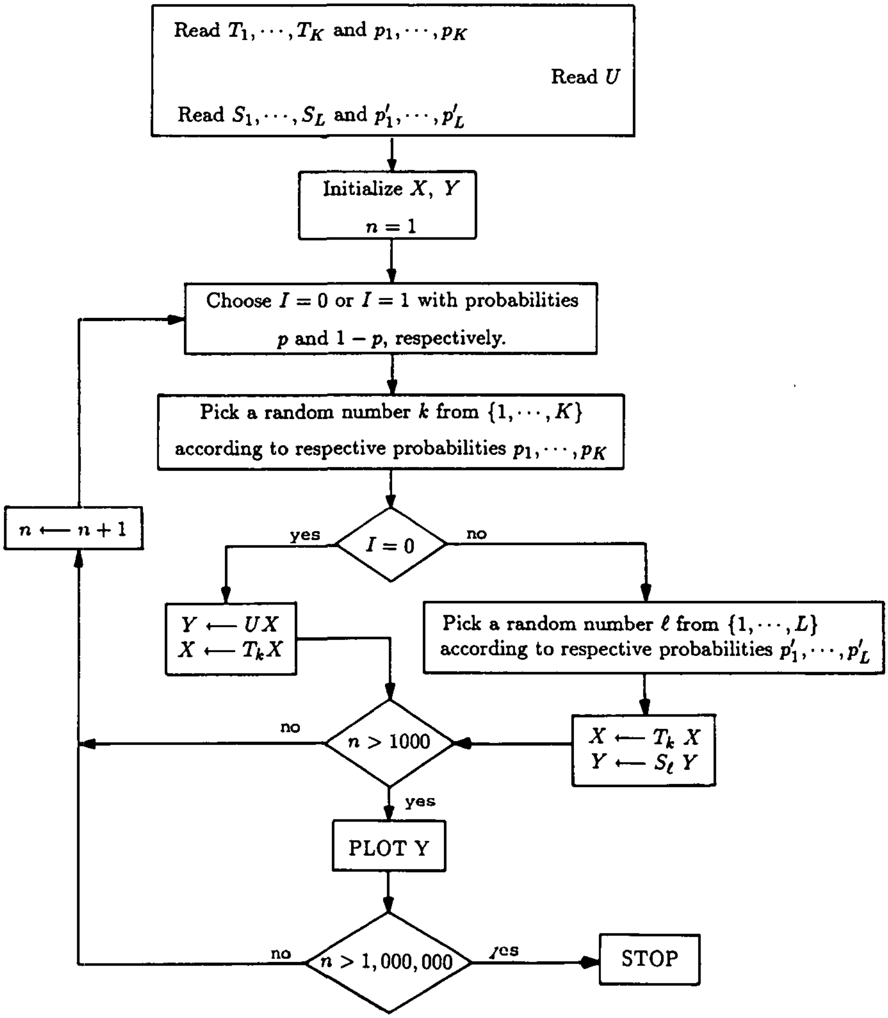



**FIGURE 3.** Simple mixing algorithm. This algorithm mixes the individual images which are generated from the mappings $T_i$ and $S_i$.

Since the $I$ process is reversible, $N(n)$ and $\max(0, n - \sigma)$ have identical distributions. (This can also be checked directly by computation, as in the proof of Lemma 4 in Section 3 below.) Let $\{\mathfrak{J}'_n\}$ be another i.i.d. sequence distributed like $\{\mathfrak{J}_n\}$, but independent of $\{\mathfrak{J}_n\}$ and $\{\mathfrak{S}_n\}$; and let $\{\hat{X}_n\}$ be the corresponding reversed process discussed in Section 1, $\hat{X}_n = \mathfrak{J}'_1 \cdots \mathfrak{J}'_n X_0$. As mentioned above it follows from [6, Section 1] that as long as $\mathbb{E} \log^+ |\mathfrak{J}X_0 - X_0| < \infty$ (in

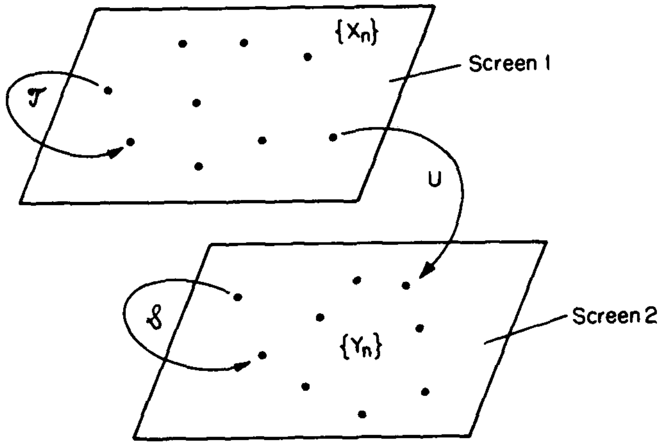



**FIGURE 4.** Schematic of the simple mixing algorithm. Only screen 2 gets plotted.

particular, this holds if $X_0 = x$), that $\hat{X}_n$ converges a.s. to a limit random variable $X$, whose distribution is fixed under $\mathfrak{I}$; i.e., $\mathfrak{I} * X = X$.

Next define

$$\tilde{X}_n = \begin{cases} \mathfrak{I}_1 \cdots \mathfrak{I}_n X_0 & \text{if } \sigma \geq n, \\ \mathfrak{I}_1 \cdots \mathfrak{I}_{\sigma+1} \hat{X}_{n-\sigma-1} & \text{if } \sigma < n, \end{cases}$$

$$\tilde{Y}_n = \begin{cases} \mathcal{S}_1 \cdots \mathcal{S}_n Y_0 & \text{if } \sigma \geq n, \\ \mathcal{S}_1 \cdots \mathcal{S}_\sigma U \hat{X}_{n-\sigma-1} & \text{if } \sigma < n. \end{cases}$$

It can be checked using Eq. (7) that for any fixed $n$, the pairs $(X_n, Y_n)$ and $(\tilde{X}_n, \tilde{Y}_n)$ have identical distributions. Since

$$\lim_{n\to\infty} (\tilde{X}_n, \tilde{Y}_n) = (\mathfrak{I}_1 \cdots \mathfrak{I}_{\sigma+1} X,\ \mathcal{S}_1 \cdots \mathcal{S}_\sigma U X) \text{ a.s.}$$

our result follows. ■

Since the limiting distribution of $Y_n$ is $\mathcal{S}_1 \cdots \mathcal{S}_\sigma U * X$, we are put into the setting of Theorem 1. Indeed, let $\{Z_n\}$ be the Markov chain evolving as $Z_{n+1} = \mathcal{S}_{n+1} Z_n$, with initial distribution $Z_0 = UX$. Then $Y = Z_\sigma$, and thus by Eq. (3)

$$\mathbb{P}(Y \in B) = q\mathbb{P}(\mathcal{S} Y \in B) + p\mathbb{P}(UX \in B), \tag{8}$$

where $\mathcal{S}$ and $Y$ are independent. In case $\mathcal{S}$ has the atomic distribution $\mathbb{P}(\mathcal{S} = S_i) = p_i'$, $1 \le i \le L$, then Eq. (8) takes the form

$$\nu_Y(B) = q \sum_{i=1}^{L} p_i' \nu_Y(S_i^{-1}B) + p\nu_X(U^{-1}B), \tag{9}$$

where $\nu_X$ and $\nu_Y$ are the distributions of $X$ and $Y$, respectively. Let $C_X$ and $C_Y$ be the respective supports of $\nu_X$ and $\nu_Y$. Then applying Eq. (9) to $B = C_Y$ and to $B = \cup_{i=1}^{L} \overline{S_i(C_Y)} \cup \overline{U(C_X)}$ leads to

$$C_Y = \bigcup_{i=1}^{L} \overline{S_i(C_Y)} \cup \overline{U(C_X)}.$$

This shows in what sense $Y$ consists of a mixture of $X$ and the chain generated from $\mathcal{S}$. The set $C_Y$ is now covered by $L$ pieces $S_i(C_Y)$ which are affinely similar to $C_Y$, and one piece which is affinely similar to $C_X$.

## 3. MORE COUPLING

This time we want to let $\{X_n\}$ and $\{Y_n\}$ each have switching mechanisms. So let

$$X_{n+1} = \begin{cases} VY_n & \text{with probability } p_X > 0, \\ \mathfrak{J}_{n+1}X_n & \text{with probability } q_X = 1 - p_X, \end{cases} \tag{10}$$

$$Y_{n+1} = \begin{cases} UX_n & \text{with probability } p_Y > 0, \\ \mathcal{S}_{n+1}Y_n & \text{with probability } q_Y = 1 - p_Y. \end{cases} \tag{11}$$

Assume $p_X + p_Y < 2$. (This is needed below in Section 4.) The setting is like that above: $\{\mathfrak{J}_n\}$ and $\{\mathcal{S}_n\}$ are i.i.d. sequences of affine maps, independent of one another, while $U$ and $V$ are fixed (nonrandom) affine maps. The choices in the $\{X_n\}$ and $\{Y_n\}$ chains are to be made independently of $\{\mathfrak{J}_n\}$ and $\{\mathcal{S}_n\}$, and independently of one another. This corresponds then to the algorithm and schematic shown in Figures 5 and 6. Let $\sigma^X$ and $\sigma^Y$ be random variables independent of $\{\mathfrak{J}_n\}$ and $\{\mathcal{S}_n\}$ and independent of each other, with distributions $\mathbb{P}(\sigma^X = k) = p_X q_X^k$ and $\mathbb{P}(\sigma^Y = k) = p_Y q_Y^k$, $k \ge 0$. Define

$$\mathfrak{R} = \mathfrak{J}_1 \cdots \mathfrak{J}_{\sigma^X} V \mathcal{S}_1 \cdots \mathcal{S}_{\sigma^Y} U. \tag{12}$$

THEOREM 3: *Assume that* $\mathbb{E} \log^+ \|\mathfrak{A}(\mathfrak{J})\|$, $\mathbb{E} \log^+ \|\mathfrak{A}(\mathcal{S})\|$, $\mathbb{E} \log^+ |\mathrm{b}(\mathfrak{J})|$, *and* $\mathbb{E} \log^+ |b(\mathcal{S})|$ *are all finite, and that the Lyapunov exponent of* $\mathfrak{A}(\mathfrak{R})$ *is negative. Then for any constant initial values* $X_0 = x$, $Y_0 = y$ *the random variable* $X_n$ *defined in Eq. (10) converges in distribution to the unique fixed distribution,* $X$, *of* $\mathfrak{R}$, $\mathfrak{R} * X = X$; *and the random variable* $Y_n$ *defined in Eq. (11) converges in distribution to* $\mathcal{S}_1 \cdots \mathcal{S}_{\sigma^Y} U * X$.

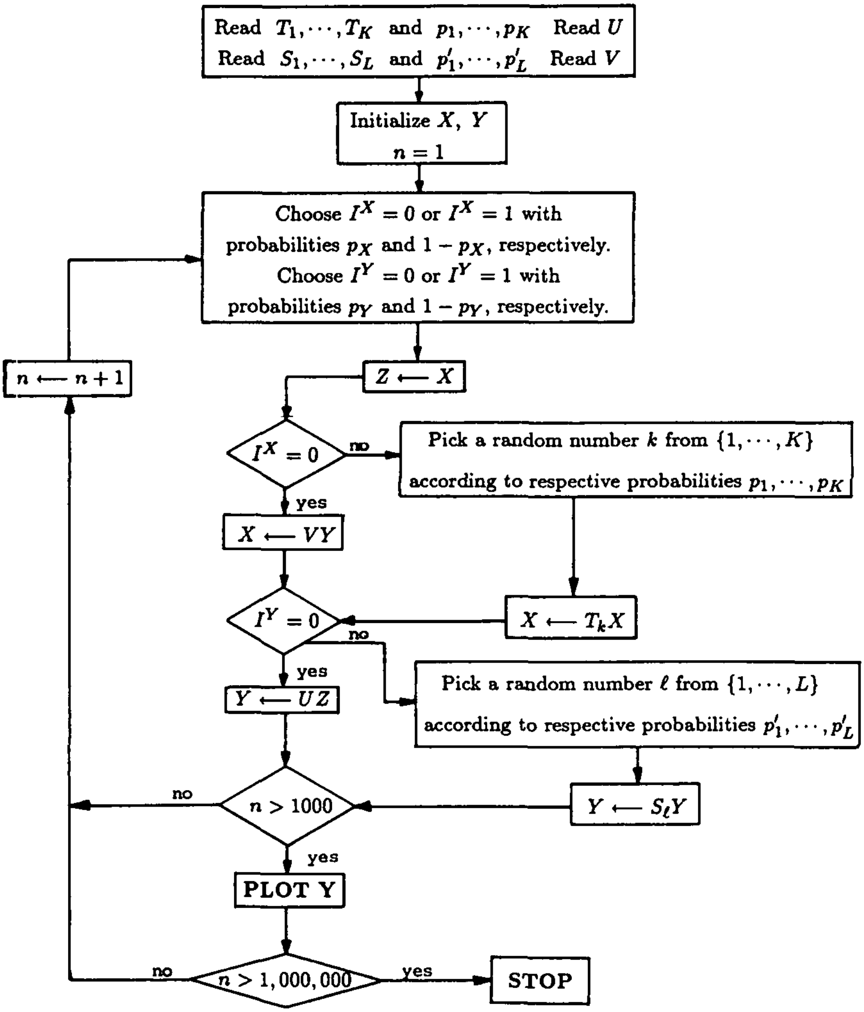



**FIGURE 5.** Two-screen mixing algorithm. This algorithm mixes the individual images which are generated from the mappings $T_i$ and $S_i$.

PROOF: Begin as above by introducing i.i.d. Bernoulli 0,1 sequences $\{I_n^X\}$ and $\{I_n^Y\}$ with $\mathbb{P}(I^X=0)=p_X$ and $\mathbb{P}(I^Y=0)=p_Y$. These sequences are to be independent of $X_0, Y_0$, $\{\mathfrak{Z}_n\}$, $\{\mathfrak{S}_n\}$ and of one another. Accordingly,

$$X_{n+1}=\begin{cases} VY_n & \text{if } I_{n+1}^X=0, \\ \mathfrak{Z}_{n+1}X_n & \text{if } I_{n+1}^X=1, \end{cases} \tag{13}$$

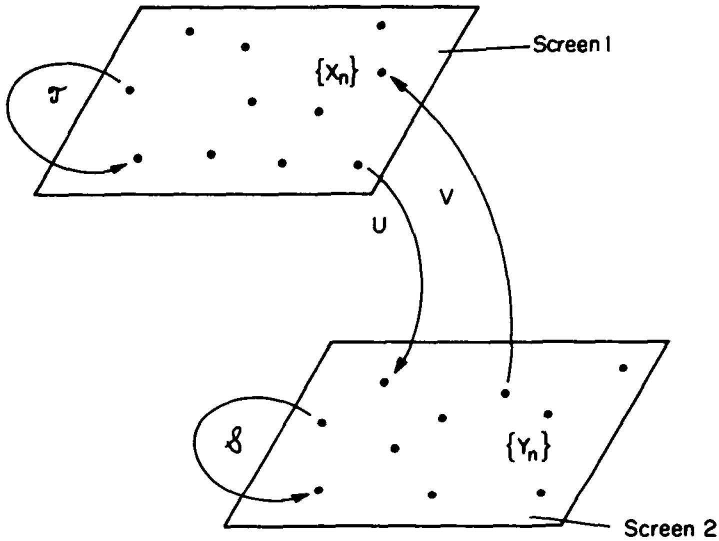



**FIGURE 6.** Schematic of the two-screen mixing algorithm. Only screen 2 gets plotted.

$$Y_{n+1} = \begin{cases} UX_n & \text{if } I_{n+1}^Y = 0, \\ \mathcal{S}_{n+1} Y_n & \text{if } I_{n+1}^Y = 1. \end{cases} \quad \textbf{(14)}$$

Define $I_0^X = I_0^Y = 0$ and set $N_X(n) = \max(k \le n\colon I_k^X = 0)$, and $N_Y(n) = \max(k \le n\colon I_k^Y = 0)$. Recursively define times $\theta_0^n = n$,

$$\theta_{2i-1}^n = N_X(\theta_{2i-2}^n) - 1, \qquad \theta_{2i}^n = N_Y(\theta_{2i-1}^n) - 1, \qquad i \ge 1,$$

until such $i = t(n) + 1$ when $\theta_i^n = -1$. From then on, define $\theta_j^n = -1, j > t(n)$. Set

$$\mathcal{R}_i^n = \mathfrak{T}_{\theta_{2i-2}^n} \cdots \mathfrak{T}_{\theta_{2i-1}^n + 2} V \mathcal{S}_{\theta_{2i-1}^n} \cdots \mathcal{S}_{\theta_{2i}^n + 2} U, \quad \textbf{(15)}$$

for $1 \le i \le s(n) = \left[\dfrac{t(n)}{2}\right]$. Then

$$X_{\theta_{2i-2}^n} = \mathcal{R}_i^n X_{\theta_{2i}^n}, \qquad 1 \le i \le s(n), \quad \textbf{(16)}$$

and

$$X_{\theta_{2s(n)}^n} = \begin{cases} \mathfrak{T}_{\theta_{2s(n)}^n} \cdots \mathfrak{T}_1 X_0 & t(n) \text{ even}, \\ \mathfrak{T}_{\theta_{2s(n)}^n} \cdots \mathfrak{T}_{\theta_{2s(n)+1}^n + 2} V \mathcal{S}_{\theta_{2s(n)+1}^n} \cdots \mathcal{S}_1 Y_0 & t(n) \text{ odd}. \end{cases} \quad \textbf{(17)}$$

See the illustration in Figure 7.

Let $\{\sigma_i^X\}$ and $\{\sigma_i^Y\}$ be i.i.d. sequences independent of one another, distributed like $\sigma^X$ and $\sigma^Y$, respectively. Set $M_0 = 1$ and define recursively

$$M_{2i-1} = M_{2i-2} + \sigma_i^X + 1, \qquad M_{2i} = M_{2i-1} + \sigma_i^Y + 1, \qquad i \geq 1.$$

Let $\hat{t}(n) = \max(i \geq 0: M_i \leq n + 1)$, $\hat{s}(n) = \left[\frac{\hat{t}(n)}{2}\right]$. Now define

$$\hat{\mathfrak{R}}_i = \mathfrak{Z}_{M_{2i-2}} \cdots \mathfrak{Z}_{M_{2i-1}-2} V \mathfrak{S}_{M_{2i-1}} \cdots \mathfrak{S}_{M_{2i}-2} U, \qquad i \geq 1, \tag{18}$$

and let

$$\hat{X}_n = \hat{\mathfrak{R}}_1 \cdots \hat{\mathfrak{R}}_{\hat{s}(n)} E_n, \qquad n \geq 1, \tag{19}$$

where

$$E_n = \begin{cases} \mathfrak{Z}_{M_{2\hat{s}(n)}} \cdots \mathfrak{Z}_n X_0 & \hat{t}(n) \text{ even}, \\ \mathfrak{Z}_{M_{2\hat{s}(n)}} \cdots \mathfrak{Z}_{M_{2\hat{s}(n)+1}-2} V \mathfrak{S}_{M_{2\hat{s}(n)+1}} \cdots \mathfrak{S}_n Y_0 & \hat{t}(n) \text{ odd}. \end{cases} \tag{20}$$

The maps $\{\hat{\mathfrak{R}}_i\}$ are i.i.d., each distributed like $\mathfrak{R}$.

LEMMA 4: *For each fixed n, $X_n$ and $\hat{X}_n$ have the same distribution.*

PROOF: By comparing Eqs. (15)–(17) with Eqs. (18)–(20), we see that it suffices to show that for each fixed $n$ the sequences $(\theta_i^n: i \geq 0)$ and $(\max(-1, n + 1 - M_i): i \geq 0)$ are identically distributed. To this end, observe that if $k_1 > k_2 > \cdots k_l \geq 0$, $l$ odd, then (refer to Figure 8)

$$\begin{aligned} \mathbb{P}(\theta_i^n = k_1, \ldots, \theta_l^n = k_l), \\ &= \mathbb{P}(N_X(n) = k_1 + 1, N_Y(k_1) = k_2 + 1, \ldots, N_X(k_{l-1}) = k_l + 1), \\ &= p_X q_X^{n-k_1-1} p_Y q_Y^{k_1-k_2-1} \cdots p_X q_X^{k_{l-1}-k_l-1}; \end{aligned}$$

and likewise

$$\begin{aligned} \mathbb{P}(n + 1 - M_1 = k_1, \ldots, n + 1 - M_l = k_l), \\ &= \mathbb{P}(\sigma_1^X = n - k_1 - 1, \sigma_1^Y = k_1 - k_2 - 1, \ldots, \\ &\qquad \sigma_{(l+1)/2}^X = k_{l-1} - k_l - 1), \\ &= p_X q_X^{n-k_1-1} p_Y q_Y^{k_1-k_2-1} \cdots p_X q_X^{k_{l-1}-k_l-1}. \end{aligned}$$

Similar calculations apply when $l$ is even, or when some of the $k_i$ are $-1$. ■

Continuing with the proof of Theorem 3, it follows from the definition of $\hat{s}(n)$ that

$$2\hat{s}(n) = \sum_{i=1}^{\hat{s}(n)} (\sigma_i^X + \sigma_i^Y) \leq n < 2 + 2\hat{s}(n) + \sum_{i=1}^{\hat{s}(n)+1} (\sigma_i^X + \sigma_i^Y).$$

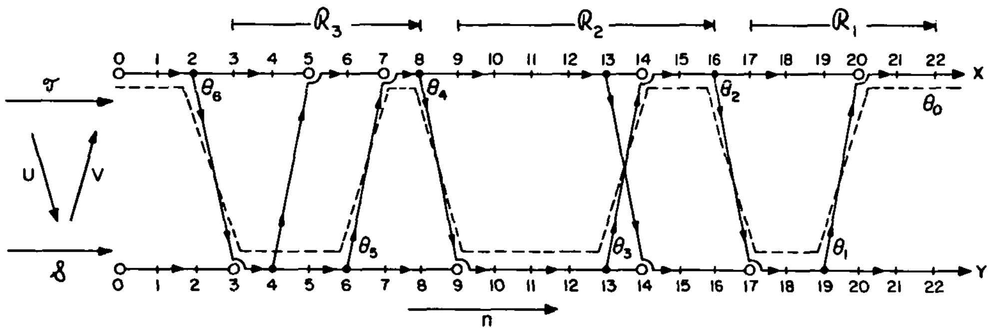
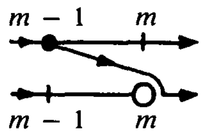



FIGURE 7. Tracing $X_n$ all the way back to the initial conditions $X_0$, $Y_0$. The convention here is that




means that $X_m = T_m X_{m-1}$ and $Y_m = UX_{m-1}$. For this illustration $n = 22$ and $\theta_1 = 19$, $\theta_2 = 16$, $\theta_3 = 13$, $\theta_4 = 8$, $\theta_5 = 6$, $\theta_6 = 2$, $\theta_7 = -1$. Thus, $t = 6$, $s = 3$, $\mathcal{R}_1 = \mathfrak{T}_{22}\mathfrak{T}_{21} V \mathcal{S}_{19}\mathcal{S}_{18} U$, $\mathcal{R}_2 = \mathfrak{T}_{16}\mathfrak{T}_{15} V \mathcal{S}_{13}\mathcal{S}_{12}\mathcal{S}_{11}\mathcal{S}_{10} U$, and $\mathcal{R}_3 = \mathfrak{T}_8 V \mathcal{S}_6\mathcal{S}_5\mathcal{S}_4 U$. Observe that $X_{22} = \mathcal{R}_1 X_{16}$, $X_{16} = \mathcal{R}_2 X_8$, $X_8 = \mathcal{R}_3 X_2$, and $X_2 = \mathfrak{T}_2\mathfrak{T}_1 X_0$.

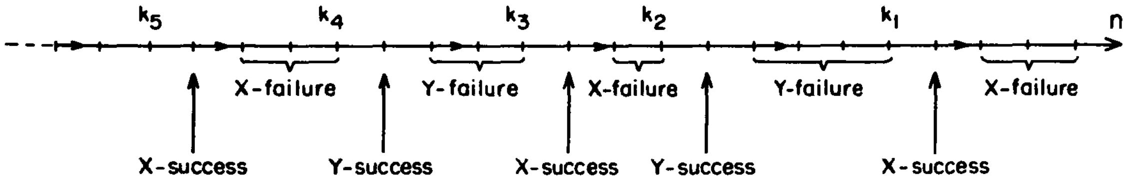



FIGURE 8. Success/failure interpretation of the event $\{\theta_1^n = k_1, \ldots, \theta_l^n = k_l\}$. Respective probabilities for $X$ and $Y$ success are $p_X$ and $p_Y$.

By the law of large numbers, then,

$$\lim_{n\to\infty} \frac{n}{\hat{s}(n)} = 2 + \frac{q_X}{p_X} + \frac{q_Y}{p_Y}, \qquad \text{a.s.}$$

Thus, if $\hat{\mathfrak{A}}_i = \mathfrak{A}(\hat{\mathfrak{R}}_i)$ then

$$\lim_{n\to\infty} \frac{1}{n} \log\| \hat{\mathfrak{A}}_1 \cdots \hat{\mathfrak{A}}_{\hat{s}(n)} \|$$

$$= \left(2 + \frac{q_X}{p_X} + \frac{q_Y}{p_Y}\right)^{-1} \lim_{n\to\infty} \frac{1}{n} \log\| \hat{\mathfrak{A}}_1 \cdots \hat{\mathfrak{A}}_n \| < 0 \qquad \text{a.s.,} \tag{21}$$

the last step by virtue of our hypothesis on $\mathfrak{A}(\mathfrak{R})$.

Consider the differences

$$\hat{X}_{n+1} - \hat{X}_n = \hat{\mathfrak{A}}_1 \cdots \hat{\mathfrak{A}}_{\hat{s}(n)} F_n, \tag{22}$$

where

$$F_n = \begin{cases} \mathfrak{J}_{M_{2\hat{s}(n)}} \cdots \mathfrak{J}_n (\mathfrak{J}_{n+1} X_0 - X_0) & \hat{\imath}(n+1) = \hat{\imath}(n) = \text{even}, \\ \mathfrak{J}_{M_{2\hat{s}(n)}} \cdots \mathfrak{J}_{M_{2\hat{s}(n)+1}-2} V \mathfrak{S}_{M_{2\hat{s}(n)+1}} \cdots & \\ \qquad \mathfrak{S}_n (\mathfrak{S}_{n+1} Y_0 - Y_0) & \hat{\imath}(n+1) = \hat{\imath}(n) = \text{odd}, \\ \mathfrak{J}_{M_{2\hat{s}(n)}} \cdots \mathfrak{J}_n (V Y_0 - X_0) & \hat{\imath}(n+1) = \hat{\imath}(n) + 1 = \text{odd}, \\ \mathfrak{J}_{M_{2\hat{s}(n)}} \cdots \mathfrak{J}_{M_{2\hat{s}(n)+1}-2} V \mathfrak{S}_{M_{2\hat{s}(n)+1}} \cdots & \\ \qquad \mathfrak{S}_n (U X_0 - Y_0) & \hat{\imath}(n+1) = \hat{\imath}(n) + 1 = \text{even}. \end{cases}$$

Accordingly, set

$$\begin{aligned} \log G_n = {} & \log|\mathfrak{J}_1 \cdots \mathfrak{J}_{n+1-M_{2\hat{s}(n)}} (\mathfrak{J}' X_0 - X_0)| \\ & + \log|\mathfrak{J}_1 \cdots \mathfrak{J}_{M_{2\hat{s}(n)+1} - M_{2\hat{s}(n)} - 1} V \mathfrak{S}_1 \cdots \mathfrak{S}_{n+1-M_{2\hat{s}(n)+1}} (\mathfrak{S}' Y_0 - Y_0)| \\ & + \log|\mathfrak{J}_1 \cdots \mathfrak{J}_{n+1-M_{2\hat{s}(n)}} (V Y_0 - X_0)| \\ & + \log|\mathfrak{J}_1 \cdots \mathfrak{J}_{M_{2\hat{s}(n)+1} - M_{2\hat{s}(n)} - 1} V \mathfrak{S}_1 \cdots \mathfrak{S}_{n+1-M_{2\hat{s}(n)+1}} (U X_0 - Y_0)|, \end{aligned}$$

where $\mathfrak{J}'$ and $\mathfrak{S}'$ are independent of $\{\mathfrak{J}_n\}$ and $\{\mathfrak{S}_n\}$. Observe that for any $\alpha > 0$

$$\mathbb{P}(\log|F_n| \geq \alpha) \leq \mathbb{P}(\log G_n \geq \alpha). \tag{23}$$

Set

$$\beta_i = \max(\log^+ \|\mathfrak{A}(\mathfrak{J}_i)\|, \log^+ \|\mathfrak{A}(\mathfrak{S}_i)\|, \log^+ \|\mathfrak{A}(U)\|).$$

Then

$$\log G_n \leq 4 \sum_{i=1}^{n+1-M_{2\hat{s}(n)}} \beta_i + \mathcal{K} \leq 4 \sum_{i=1}^{\sigma^X_{\hat{s}(n)} + \sigma^Y_{\hat{s}(n)}} \beta_i + \mathcal{K},$$

where

$\mathcal{K} = \log|\mathfrak{I}'X_0 - X_0| + \log|\mathcal{S}'Y_0 - Y_0| + \log|VY_0 - X_0| + \log|UX_0 - Y_0|.$

Thus, since the $\beta_i$ are independent of the $\sigma_i^X$ and $\sigma_i^Y$,

$$\mathbb{P}(\log G_n \geq \alpha) \leq \mathbb{P}\left(4 \sum_{i=1}^{\sigma^X_{\hat{s}(n)}+\sigma^Y_{\hat{s}(n)}} \beta_i + \mathcal{K} \geq \alpha\right) = \mathbb{P}\left(4 \sum_{i=1}^{\sigma^X+\sigma^Y} \beta_i + \mathcal{K} \geq \alpha\right). \tag{24}$$

It follows from our hypothesis that

$$\mathbb{E} \sum_{i=1}^{\sigma^X+\sigma^Y} \beta_i = \mathbb{E}(\sigma^X + \sigma^Y)\mathbb{E}\beta_1 < \infty,$$

and if $X_0 = x$, $Y_0 = y$, then also $\mathbb{E}\mathcal{K} < \infty$. Thus, by the Borel–Cantelli lemma, using Eqs. (23) and (24),

$$\lim_{n\to\infty} \frac{1}{n} \log^+|F_n| = 0, \qquad \text{a.s.}$$

Using this together with Eqs. (21) and (22), we see that $\hat{X}_n \to X$ a.s., where $X$ is a random variable whose distribution is fixed under $\mathcal{R}$. Using Lemma 4, it follows that $X_n$ converges in distribution to $X$. From the arguments in Section 2, it now follows also that $Y_n$ converges in distribution to $\mathcal{S}_1 \cdots \mathcal{S}_{\sigma^Y} U * X$. ■

There is a heuristic way of obtaining the limiting distributions $X$ and $Y$ for $X_n$ and $Y_n$, respectively. According to Theorem 1, these limits must satisfy

$$X = \mathfrak{I}_1 \cdots \mathfrak{I}_{\sigma^X} V * Y, \qquad Y = \mathcal{S}_1 \cdots \mathcal{S}_{\sigma^Y} U * X.$$

These can be "solved simultaneously" to yield $X = \mathcal{R} * X$, where $\mathcal{R}$ is given by Eq. (12).

Suppose $\mathfrak{I}$ and $\mathcal{S}$ have the respective atomic distributions $\mathbb{P}(\mathfrak{I} = T_i) = p_i$, $1 \leq i \leq K$ and $\mathbb{P}(\mathcal{S} = S_i) = p_i'$, $1 \leq i \leq L$. Let $C_X$ and $C_Y$ be the supports of $X$ and $Y$, respectively. One can argue as in Section 2 and show that

$$C_X = \bigcup_{i=1}^{K} \overline{T_i(C_X)} \cup \overline{V(C_Y)},$$

$$C_Y = \bigcup_{i=1}^{L} \overline{S_i(C_Y)} \cup \overline{U(C_X)}.$$

These equations clearly show the additional coupling that is involved.

## 4. JOINT ASYMPTOTIC STATIONARITY

In this section, we show that the Markov chain $\{(X_n, Y_n)\}$, constructed above in Eqs. (13) and (14), is in fact (jointly) asymptotically stationary, so that the joint distribution of $(X_n, Y_n)$ converges weakly to a limiting distribution which does not depend on the initial condition $(X_0, Y_0)$. Our previous result, Theorem 3, established the convergence in distribution of each of the individual *marginals* $X_n$ and $Y_n$ to respective limits $X$ and $Y$ which do not depend on the constant initial values $X_0 = x$, $Y_0 = y$. We are thus led to investigate special conditions under which convergence of the individual marginals suffices to infer joint convergence.

THEOREM 5: *Let $X_n$ and $Y_n$, $n \geq 1$, be random variables which converge in distribution to respective limits $X$ and $Y$. Let $\tau$ be a nonnegative integer-valued random variable which is finite a.s., and let $Z_n, n \geq 1$, be a random variable having the property that*

$$Z_n = \begin{cases} \phi(X_{n-\tau}, E_\tau) & \text{on } \{\tau < n\} \cap A, \\ \phi(Y_{n-\tau}, F_\tau) & \text{on } \{\tau < n\} \cap A^c. \end{cases}$$

*The random variables $\{E_n, F_n \colon n \geq 0\}$, the event $A$, and $\tau$ are all assumed to be independent of $\{X_n, Y_n\}$; and $\phi$ is assumed to be a continuous function. Then $Z_n$ converges in distribution to $\phi(X,E_\tau)\chi_A + \phi(Y,F_\tau)\chi_{A^c}$ (where $\chi_A$ denotes the indicator function of the set $A$).*

PROOF:

Let $f$ be any bounded continuous function. For each fixed $n$

$$\mathbb{E}f(Z_n) = \sum_{k=0}^{n-1} \left[ \int_{\{\tau=k\}\cap A} f(\phi(X_{n-k}, E_k)) + \int_{\{\tau=k\}\cap A^c} f(\phi(Y_{n-k}, F_k)) \right] + \int_{\{\tau \geq n\}} f(Z_n). \tag{25}$$

For each fixed $k$

$$\lim_{n\to\infty} \int_{\{\tau=k\}\cap A} f(\phi(X_{n-k}, E_k)) = \int_{\{\tau=k\}\cap A} f(\phi(X, E_k)),$$

$$\lim_{n\to\infty} \int_{\{\tau=k\}\cap A^c} f(\phi(Y_{n-k}, F_k)) = \int_{\{\tau=k\}\cap A^c} f(\phi(Y, F_k)).$$

Thus, we can pass to the limit in Eq. (25) as $n \to \infty$. ∎

We want to use this result with $Z_n$ having the same distribution as the pair $(X_n, Y_n)$ for our Markov chain constructed above in Eqs. (13) and (14). Using the sequences $\{\sigma_i^X\}$ and $\{\sigma_i^Y\}$ defined above in Section 3, set $\alpha = \min(i \geq 1\colon \sigma_i^X \neq \sigma_i^Y)$. Observe that $\alpha < \infty$ a.s., since $p_X + p_Y \in (0,2)$. Define

$$\tau = \sum_{i=1}^{\alpha-1} (\sigma_i^X + 1) + \sigma_\alpha^X \wedge \sigma_\alpha^Y,$$

and let $A$ be the event $A = \{\sigma_\alpha^Y < \sigma_\alpha^X\}$.

To construct the appropriate random variables $E_n$ and $F_n$, we introduce an auxiliary Markov chain. For this purpose, let $\{\mathfrak{Z}'_N: N \geq 0\}$ and $\{\mathcal{S}'_N: N \geq 0\}$ be i.i.d. sequences distributed like $\mathfrak{Z}$ and $\mathcal{S}$, respectively, but independent of our original sequences $\{\mathfrak{Z}_n\}$, $\{\mathcal{S}_n\}$, $\{I_n^X\}$, and $\{I_n^Y\}$, and also of one another. Let $\{I_N^{XY}\}$ be an i.i.d. Bernoulli 0,1 sequence, independent of all six sequences mentioned just above, with

$$\mathbb{P}(I^{XY}=0) = \mathbb{P}(I^X = 0 | I^X = I^Y) = \frac{p_X p_Y}{p_X p_Y + q_X q_Y}. \tag{26}$$

Define the Markov chain $\{(X'_N, Y'_N)\}$ similar in spirit to $\{(X_n, Y_n)\}$, but with *simultaneous mixing*, so that the selection mechanisms in Eqs. (10) and (11) operate simultaneously. We want the crossover times depicted in Figure 7 to all coincide. Precisely,

$$X'_{N+1} = \begin{cases} VY'_N & \text{if } I_{N+1}^{XY} = 0, \\ \mathfrak{Z}'_{N+1} X'_N & \text{if } I_{N+1}^{XY} = 1, \end{cases} \tag{27}$$

$$Y'_{N+1} = \begin{cases} UX'_N & \text{if } I_{N+1}^{XY} = 0, \\ \mathcal{S}'_{N+1} Y'_N & \text{if } I_{N+1}^{XY} = 1. \end{cases} \tag{28}$$

Let $\{(\bar{X}_N, \bar{Y}_N): N \geq 0\}$ denote the chain of Eq. (27) and (28) with initial condition

$$\bar{X}_0 = \mathfrak{Z}'_0 X, \qquad \bar{Y}_0 = UX,$$

and similarly let $\{(\bar{\bar{X}}_N, \bar{\bar{Y}}_N): N \geq 0\}$ denote the same chain of Eqs. (27) and (28) but with the different initial condition

$$\bar{\bar{X}}_0 = VY, \qquad \bar{\bar{Y}}_0 = \mathcal{S}'_0 Y.$$

THEOREM 6: *Under the same assumptions as in Theorem 3 above, the Markov chain $\{(X_n, Y_n)\}$ constructed above in Eqs. (13) and (14) is asymptotically stationary. It has the stationary distribution*

$$(\bar{X}_\tau, \bar{Y}_\tau)\chi_A + (\bar{\bar{X}}_\tau, \bar{\bar{Y}}_\tau)\chi_{A^c}.$$

PROOF: For each fixed $n$, let $\{(\bar{X}_N^n, \bar{Y}_N^n): N \geq 0\}$ be the chain of Eqs. (27) and (28) with initial condition

$$\bar{X}_0^n = \begin{cases} \mathfrak{Z}'_0 X_{n-\tau-1} & \text{if } \tau < n, \\ X_0 & \text{if } \tau \geq n, \end{cases}$$

$$\bar{Y}_0^n = \begin{cases} UX_{n-\tau-1} & \text{if } \tau < n, \\ Y_0 & \text{if } \tau \geq n. \end{cases}$$

Similarly, let $\{(\bar{\bar{X}}_N^n, \bar{\bar{Y}}_N^n): N \geq 0\}$ be the same chain of Eqs. (27) and (28) but with initial condition

$$\bar{\bar{X}}_0^n = \begin{cases} VY_{n-\tau-1} & \text{if } \tau < n, \\ X_0 & \text{if } \tau \geq n, \end{cases}$$

$$\bar{\bar{Y}}_0^n = \begin{cases} S_0' Y_{n-\tau-1} & \text{if } \tau < n, \\ Y_0 & \text{if } \tau \geq n. \end{cases}$$

LEMMA 7: *For each fixed n, the pairs* $(X_n, Y_n)$ *and*

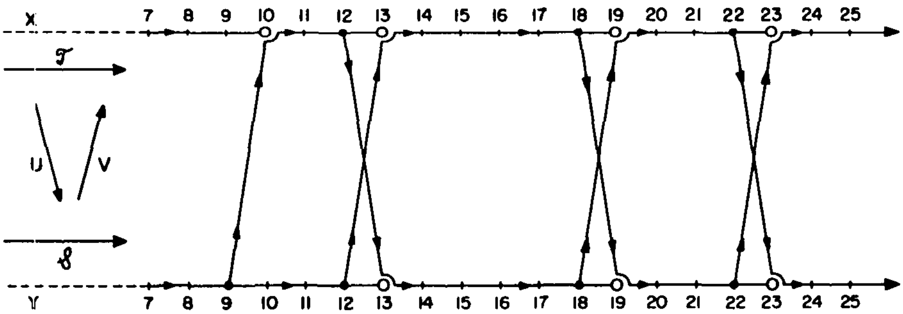
$$(\tilde{X}_n, \tilde{Y}_n) = (\bar{X}^n_{\tau \wedge n}, \bar{Y}^n_{\tau \wedge n})\chi_A + (\bar{\bar{X}}^n_{\tau \wedge n}, \bar{\bar{Y}}^n_{\tau \wedge n})\chi_{A^c}$$

*have identical distributions.*

PROOF: The main idea here is that we trace the chain $\{(X_n, Y_n)\}$ backwards from stage $n$ to arrive at a "common parent," or antecedent $X_{\xi(n)}$ or $Y_{\xi(n)}$, from which $X_n$ and $Y_n$ both stem. Then from stage $\xi(n) + 1$ on we let $\{(X_k, Y_k)\}$ re-evolve forward in time, but with simultaneous, or coalesced mixing transitions. For this purpose, define

$$\xi(n) = \begin{cases} \max(k \leq n: I_k^X + I_k^Y = 1) - 1 & \text{if such } k \text{ exists}, \\ -1 & \text{otherwise}. \end{cases}$$

Then whenever $\xi(n) \geq 0$, both $X_n$ and $Y_n$ have a common antecedent at stage $\xi(n)$. This is illustrated in Figure 9.




$$X_{25} = T_{25}T_{24}VS_{22}S_{21}S_{20}UT_{18}T_{17}T_{16}T_{15}T_{14}VS_{12}S_{11}S_{10}Y_9$$

$$Y_{25} = S_{25}S_{24}UT_{22}T_{21}T_{20}VS_{18}S_{17}S_{16}S_{15}S_{14}UT_{12}T_{11}VY_9$$

**FIGURE 9.** Tracing back to a common parent. Here $\xi(25) = 9$, and $X_{25}$, $Y_{25}$ can both be traced back to $Y_9$, as indicated. The crossovers after stage 10 are all simultaneous.

The connection between the original chain $\{(X_n, Y_n)\}$ and the auxiliary chain $\{(X'_N, Y'_N)\}$ can be formulated in terms of $\xi$. The basic relationship is

$$\mathbb{P}_{xy}((X_n, Y_n) \in C \mid \xi(n) = -1) = \mathbb{P}_{xy}((X'_n, Y'_n) \in C), \qquad n \geq 1. \tag{29}$$

The point is that although $\xi(n)$ is *not* a (Markov) stopping time, nevertheless the conditioned process $\{(X_k, Y_k): 0 \leq k \leq n\}$, given that $\xi(n) = -1$, is still Markov—in fact it is simply the simultaneous process $\{(X'_k, Y'_k): 0 \leq k \leq n\}$. The reason for this is that our original process $\{(X_n, Y_n)\}$ defined in Eqs. (13) and (14) is really *compound* Markov. It has the effect of the random affine products and the effect of the Bernoulli crossover mechanisms. Knowing that $\xi(n) = -1$ only allows us to foresee something about the switching mechanisms (namely, that they coalesce), but tells nothing about the random affine maps. Thus, even though $\xi(n)$ is predictive, the resulting conditioned process is still Markov. This is what Eq. (29) is saying. It can be verified by conditioning on the $I^X$ and $I^Y$ processes, and using Eq. (26).

From the definition of $\xi(n)$ and the time homogeneity of our Markov process, we also have the relationships

$$\begin{aligned}\mathbb{P}_{xy}&((X_n, Y_n) \in C \mid \xi(n) = 0)\\ &= \mathbb{E}_{xy}[\mathbb{P}_{X_1 Y_1}((X_{n-1}, Y_{n-1}) \in C \mid \xi(n-1) = -1) \mid \xi(1) = 0]\end{aligned} \tag{30}$$

and

$$\begin{aligned}\mathbb{P}_{xy}((X_1, Y_1) \in C \mid \xi(1) = 0) &= \mathbb{P}((\mathfrak{J}x, Ux) \in C)\mathbb{P}(I^Y = 0 \mid I^X \neq I^Y)\\ &\quad + \mathbb{P}((Vy, \mathfrak{S}y) \in C)\mathbb{P}(I^Y = 1 \mid I^X \neq I^Y).\end{aligned} \tag{31}$$

Combining these, we infer that for any $0 \leq k < n$

$$\begin{aligned}\mathbb{P}_{xy}&((X_n, Y_n) \in C \mid \xi(n) = k)\\ &= \mathbb{E}_{xy}[\mathbb{P}_{X_k Y_k}((X_{n-k}, Y_{n-k}) \in C \mid \xi(n-k) = 0) \mid \xi(k+1) = k]\\ &\qquad \text{(by the Markov property)},\\ &= \mathbb{E}_{xy}[\mathbb{P}_{X_{k+1} Y_{k+1}}((X_{n-k-1}, Y_{n-k-1}) \in C \mid \xi(n-k-1) = -1)\\ &\qquad \mid \xi(k+1) = k] \qquad \text{(by Eq. (30))},\\ &= \mathbb{E}_{xy}[\mathbb{P}_{X_{k+1} Y_{k+1}}((X'_{n-k-1}, Y'_{n-k-1} \in C) \mid \xi(k+1) = k]\\ &\qquad \text{(by Eq. (29))},\\ &= \mathbb{E}_{xy}\mathbb{E}_{X_k Y_k}[\mathbb{P}_{X_1 Y_1}((X'_{n-k-1}, Y'_{n-k-1}) \in C) \mid \xi(1) = 0]\\ &\qquad \text{(by the Markov property)},\\ &= \mathbb{E}_{xy}\mathbb{P}_{\mathfrak{J}X_k UX_k}((X'_{n-k-1}, Y'_{n-k-1}) \in C)\mathbb{P}(I^Y = 0 \mid I^X \neq I^Y)\\ &\quad + \mathbb{E}_{xy}\mathbb{P}_{VY_k \mathfrak{S}Y_k}((X'_{n-k-1}, Y'_{n-k-1}) \in C)\mathbb{P}(I^Y = 1 \mid I^X \neq I^Y)\\ &\qquad \text{(by Eq. (31))}.\end{aligned} \tag{32}$$

Likewise, we have the series of computations

$$\begin{aligned}
&\mathbb{P}_{xy}((\tilde{X}_n, \tilde{Y}_n) \in C \mid \tau = n-k-1) \\
&\qquad = \mathbb{P}_{xy}((\bar{X}^n_{n-k-1}, \bar{Y}^n_{n-k-1}) \in C, A \mid \tau = n-k-1) \\
&\qquad\quad + \mathbb{P}_{xy}((\bar{\bar{X}}^n_{n-k-1}, \bar{\bar{Y}}^n_{n-k-1}) \in C, A^c \mid \tau = n-k-1), \qquad (33) \\
&\qquad = \mathbb{E}_{xy}\mathbb{P}_{\mathfrak{I}X_k UX_k}((X'_{n-k-1}, Y'_{n-k-1}) \in C)\mathbb{P}(A \mid \tau = n-k-1) \\
&\qquad\quad + \mathbb{E}_{xy}\mathbb{P}_{VY_k\mathcal{S}Y_k}((X'_{n-k-1}, Y'_{n-k-1} \in C)\mathbb{P}(A^c \mid \tau = n-k-1).
\end{aligned}$$

Similarly, for $\tau \geq n$

$$\begin{aligned}
&\mathbb{P}_{xy}((\tilde{X}_n, \tilde{Y}_n) \in C \mid \tau \geq n) \\
&\quad = \mathbb{P}_{xy}((\bar{X}^n_n, \bar{Y}^n_n) \in C, A \mid \tau \geq n) + \mathbb{P}_{xy}((\bar{\bar{X}}^n_n, \bar{\bar{Y}}^n_n) \in C, A^c \mid \tau \geq n), \\
&\quad = \mathbb{P}_{xy}((X'_n, Y'_n) \in C, A \mid \tau \geq n) + \mathbb{P}_{xy}((X'_n, Y'_n) \in C, A^c \mid \tau \geq n), \qquad (34) \\
&\quad = \mathbb{P}_{xy}((X'_n, Y'_n) \in C).
\end{aligned}$$

Observe now that for $0 \leq k < n$

$$\mathbb{P}(A \mid \tau = n-k-1) = \mathbb{P}(I^Y = 0 \mid I^X \neq I^Y).$$

From this we conclude, using Eqs. (32) and (33), that

$$\begin{aligned}
&\mathbb{P}_{xy}((X_n, Y_n) \in C \mid \xi(n) = k) \\
&\qquad = \mathbb{P}_{xy}((\tilde{X}_n, \tilde{Y}_n) \in C \mid \tau = n-k-1), \qquad 0 \leq k < n.
\end{aligned}$$

In addition, using Eqs. (29) and (34), it follows that

$$\mathbb{P}_{xy}((X_n, Y_n) \in C \mid \xi(n) = -1) = \mathbb{P}_{xy}((\tilde{X}_n, \tilde{Y}_n) \in C \mid \tau \geq n).$$

Next observe that $\xi(n)$ and $n - (\tau \wedge n) - 1$ have identical distributions. This is argued by a success/failure analysis, as in the proof of Lemma 4 above. (See Figure 8 there.) From this we conclude, by conditioning on $\xi(n)$ and $\tau$, that

$$\mathbb{P}_{xy}((X_n, Y_n) \in C) = \mathbb{P}_{xy}((\tilde{X}_n, \tilde{Y}_n) \in C). \qquad \blacksquare$$

Continuing with the proof of Theorem 6, we want to apply Theorem 5 to the random variable $Z_n = (\tilde{X}_n, \tilde{Y}_n)$. Observe that if $\tau < n$, then

$$\bar{X}^n_\tau = E'_\tau X_{n-\tau-1}, \qquad \bar{Y}^n_\tau = E''_\tau X_{n-\tau-1},$$

where $E'_k$ and $E''_k$ are $(k+1)$-fold products from among the affine maps $\{\mathfrak{I}'_N\}$, $\{\mathcal{S}'_N\}$, $U$, and $V$. Similarly, if $\tau < n$, then

$$\bar{\bar{X}}^n_\tau = F'_\tau Y_{n-\tau-1}, \qquad \bar{\bar{Y}}^n_\tau = F''_\tau Y_{n-\tau-1},$$

where $F_k'$ and $F_k''$ are likewise $(k+1)$-fold products from among these same affine maps. We can apply Theorem 5 now, with $\phi$ being affine in $x$:

$$\phi(x,E) = \phi(x,(E',E'')) = (E'x,E''x).$$

This leads to our desired conclusion. ■

## 5. GENERAL MIXING

It should be clear from the arguments in Sections 2–4 that it is straightforward indeed to extend the mixing models so that the crossover maps $U$ and $V$ themselves be random, independent of everything else. Thus, Eqs. (13) and (14) now become

$$X_{n+1} = \begin{cases} \mathcal{V}_{n+1} Y_n & \text{if } I_{n+1}^X = 0, \\ \mathfrak{T}_{n+1} X_n & \text{if } I_{n+1}^X = 1, \end{cases}$$

$$Y_{n+1} = \begin{cases} \mathcal{U}_{n+1} X_n & \text{if } I_{n+1}^Y = 0, \\ \mathcal{S}_{n+1} Y_n & \text{if } I_{n+1}^Y = 1. \end{cases}$$

Here $\{\mathcal{V}_n\}$ and $\{\mathcal{U}_n\}$ are i.i.d. sequences of affine maps, independent of $\{\mathfrak{T}_n\}$, $\{\mathcal{S}_n\}$, $\{I_n^X\}$, and $\{I_n^Y\}$ and of one another. The conclusions of Theorems 3 and 6 remain valid if we add the hypotheses that $\mathbb{E}\log^+\|\mathfrak{A}(\mathcal{V})\|$, $\mathbb{E}\log^+\|\mathfrak{A}(\mathcal{U})\|$, $\mathbb{E}\log^+|b(\mathcal{V})|$ and $\mathbb{E}\log^+|b(\mathcal{U})|$ also all be finite. If $\mathcal{V}$ and $\mathcal{U}$ are atomic distributions with atoms at $V_i$ and $U_i$, respectively, then the support equations for the limiting distributions $X$ and $Y$ become

$$C_X = \bigcup \overline{T_i(C_X)} \cup \bigcup \overline{V_i(C_Y)},$$

$$C_Y = \bigcup \overline{S_i(C_Y)} \cup \bigcup \overline{U_i(C_X)}.$$

We can further generalize our model so as to allow for more than two screens, or processes. Some typical models are schematized in Figures 10 and 11. To formulate an $N$-screen setup, we introduce an $N \times N$ transition probability matrix $Q$ and $N^2$ i.i.d. sequences $\{\mathfrak{T}_n(i,j): n \geq 1\}$ for $1 \leq i, j \leq N$, all mutually independent. We also introduce $N$ i.i.d. sequences $\{I_n(i): n \geq 1\}$ for $1 \leq i \leq N$, with $\mathbb{P}(I_n(i) = j) = Q_{ij}$ for our switching mechanisms. These sequences $\{I_n(i)\}$ are to be independent of the sequences $\{\mathfrak{T}_n(i,j)\}$ and also of one another.

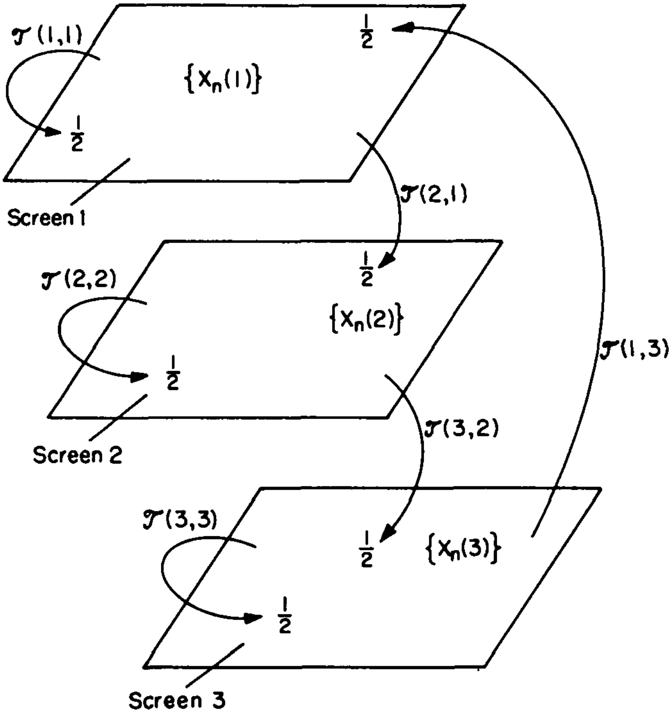



**FIGURE 10.** 3-screen mixing schematic. The switching is governed by

$$Q = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

Using these ingredients, we can specify the evolution of our mixed chain $\{(X_n(1), \ldots, X_n(N))\}$ in $\underbrace{\mathbb{R}^d \times \cdots \times \mathbb{R}^d}_{N \text{ times}}$ as

$$X_{n+1}(i) = \mathfrak{Z}_{n+1}(i, I_{n+1}(i)) X_n(I_{n+1}(i)), \qquad n \geq 0. \tag{35}$$

In Section 2, our model corresponds to

$$Q = \begin{pmatrix} 1 & 0 \\ p & q \end{pmatrix},$$

and in Section 3, it corresponds to

$$Q = \begin{pmatrix} q_X & p_X \\ p_Y & q_Y \end{pmatrix}.$$

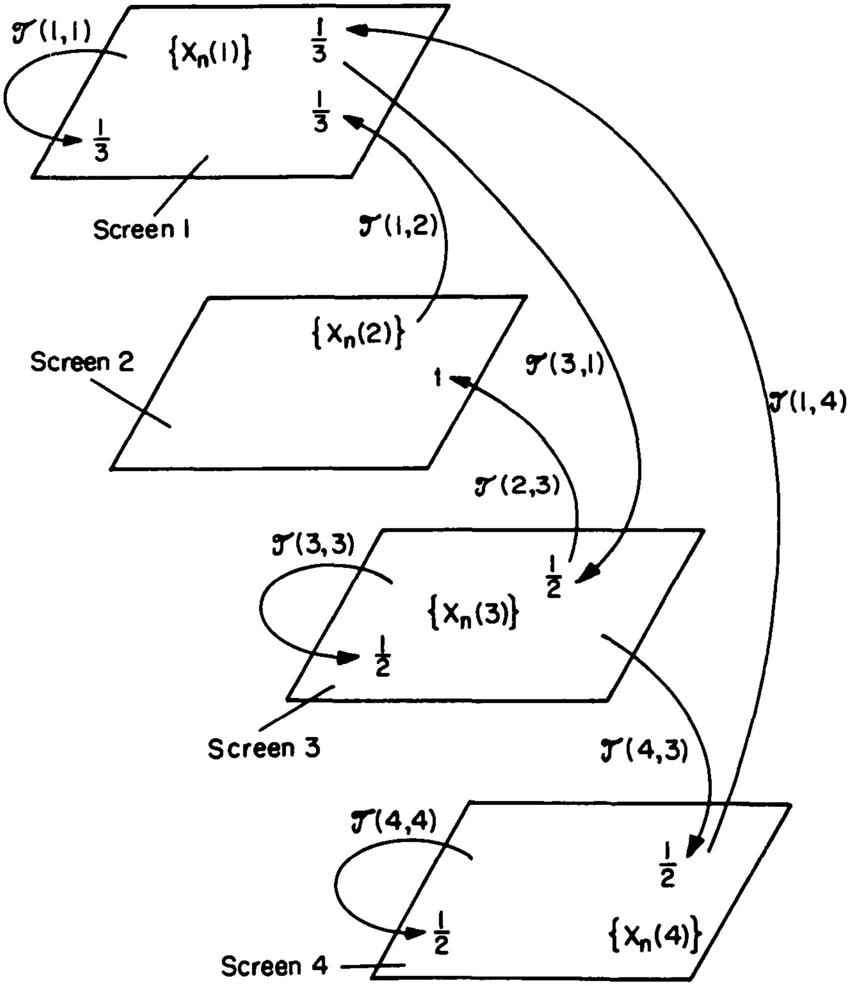



**FIGURE 11.** 4-screen mixing schematic. The switching is governed by

$$Q = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

Suppose the marginals $X_n(i)$ converge in distribution to limits $X(i)$. Then the invariance condition becomes

$$\nu_i(B) = \sum_j Q_{ij}\nu_j(\mathfrak{I}^{-1}(i,j)B), \qquad 1 \le i \le N, \tag{36}$$

where $\nu_i$ denotes the distribution of $X(i)$. In particular, if the $\mathfrak{I}(i,j)$ are discrete with atoms $T_k(i,j)$, then the supports $C_i$ of $\nu_i$ will satisfy the coupled set equations

$$C_i = \bigcup_{j:Q_{ij}>0} \bigcup_k \overline{T_k(i,j)(C_j)}. \tag{37}$$

The results of Sections 2 and 3 lead us to investigate the influence of the Markov chain $\{\Gamma_n: n \geq 0\}$ on $\{1,\ldots,N\}$ with transition probability matrix $Q$. If we start at stage $n$ and trace the evolution of $\{X_k(i)\}$ *backwards* in time, then the screen crossovers follow this chain $\{\Gamma_k\}$, with initial condition $\Gamma_0 = i$. Thus, the reversed process jumps from screen to screen in a Markovian fashion. This is the key point—*reversibility*. Analysis of this general mixing model will be presented in a subsequent work.

## 6. ILLUSTRATIONS

Figures 12 and 13 illustrate the supports of the stationary distribution for two (unmixed) Markov chains arising from products of i.i.d. affine maps. The ran-

$$T_1: \begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} .8x + .1 \\ .8y + .04 \end{bmatrix} \qquad p_1 = \frac{1}{4}$$

$$T_2: \begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} .5x + .25 \\ .5y + .4 \end{bmatrix} \qquad p_2 = \frac{1}{4}$$

$$T_3: \begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} .355x - .366y + .266 \\ .355x + .355y + .078 \end{bmatrix} \qquad p_3 = \frac{1}{4}$$

$$T_4: \begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} .355x + .355y + .378 \\ -.355x + .355y + .434 \end{bmatrix} \qquad p_4 = \frac{1}{4}$$

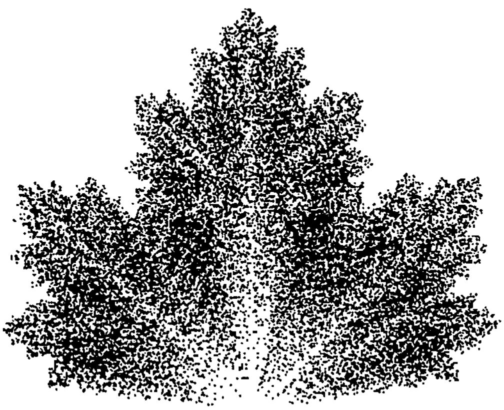


**FIGURE 12.** This image is generated from the four mappings above. The window here is $0 \leq x, y \leq 1$.

$$T_1 : \begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} .856x + .0414y + .07 \\ -.0205x + .858y + .147 \end{bmatrix} \qquad p_1 = .7$$

$$T_2 : \begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} .244x - .385y + .393 \\ .176x + .224y + .102 \end{bmatrix} \qquad p_2 = .1$$

$$T_3 : \begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} -.144x + .39y + .527 \\ .181x + .259y - .014 \end{bmatrix} \qquad p_3 = .1$$

$$T_4 : \begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} .486 \\ .031x + .216y + .05 \end{bmatrix} \qquad p_4 = .1$$

**Figure 13.** This image is generated from the four mappings above. The window here is $0 \le x, y \le 1$.

dom maps for the chains are discrete, each having four atoms, as indicated in each figure. This corresponds then to the basic algorithm depicted in Figure 1. Observe, in particular, the self-similarity of the fern—each branch is a copy of the whole fern. The leaf has a similar structure, which can be detected by studying its boundary.

Figure 14 illustrates a mixed image constructed from the fern and the leaf. The schematic for the mixing model is given in Figure 15. Observe that unlike Figures 12 and 13, the mixed image no longer exhibits self-similarity. The branch is no longer simply a copy of the whole tree—it has leaves, but not its own branches, growing from it.

## 7. A COMPARISON WITH [2]

In [2] is formulated a model whereby products of affine maps (more generally, Lipschitz maps) are generated through a Markovian index chain. Precisely, let

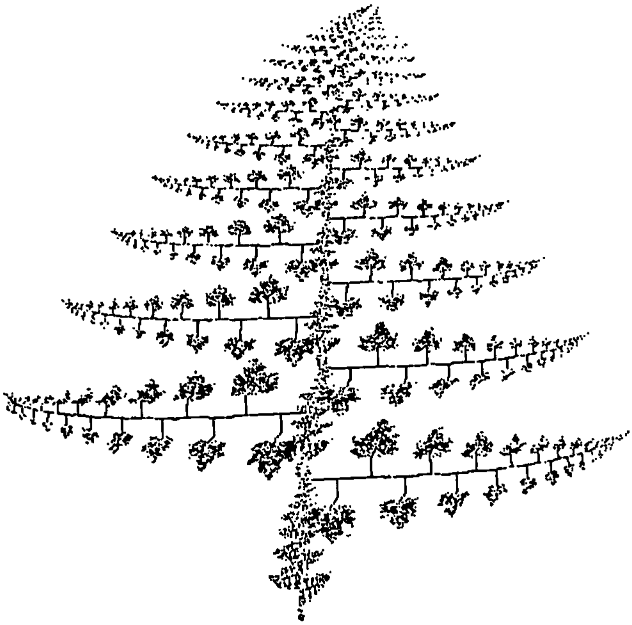


**FIGURE 14.** This mixed image combines the textures of the fern and the leaf. The branch is not a copy of the whole tree.

$T_1, \ldots, T_N$: $\mathbb{R}^m \to \mathbb{R}^m$ be affine maps and let $\{I_n: n \geq 1\}$ be a Markov chain on $\{1, \ldots, N\}$. Then the process $\{X_n\}$ on $\mathbb{R}^m$ evolves according to

$$X_{n+1} = T_{I_{n+1}} X_n.$$

This process $\{X_n\}$ is not by itself Markov, but the joint process $\{(X_n, I_n)\}$ is. We would like to show that the various $\mathbb{R}^m$ marginals of any stationary distribution for $\{(X_n, I_n)\}$ on $\mathbb{R}^m \times \{1, \ldots, N\}$ obey conditions like Eq. (36). More generally, let $\mathfrak{I}(1), \ldots, \mathfrak{I}(N)$: $\mathbb{R}^m \to \mathbb{R}^m$ be (jointly) random affine maps, and let the sequence $\{(\mathfrak{I}_n(1), \ldots, \mathfrak{I}_n(N)): n \geq 1\}$ be i.i.d., each $N$-tuple being distributed like $(\mathfrak{I}(1), \ldots, \mathfrak{I}(N))$; and independent of the chain $\{I_n\}$. As above, let $\{X_n\}$ evolve according to

$$X_{n+1} = \mathfrak{I}_{n+1}(I_{n+1}) X_n. \tag{38}$$

Let $\nu$ be any stationary distribution for $\{(X_n, I_n)\}$ on $\mathbb{R}^m \times \{1, \ldots, N\}$. It follows from the stationarity that $\pi(i) = \nu(\mathbb{R}^m \times \{i\})$ is stationary for $\{I_n\}$. Let $K = \operatorname{supp}(\pi) \subset \{1, \ldots, N\}$ and define the marginals

$$\nu_i(B) = \frac{\nu(B \times \{i\})}{\pi(i)}, \qquad i \in K,$$

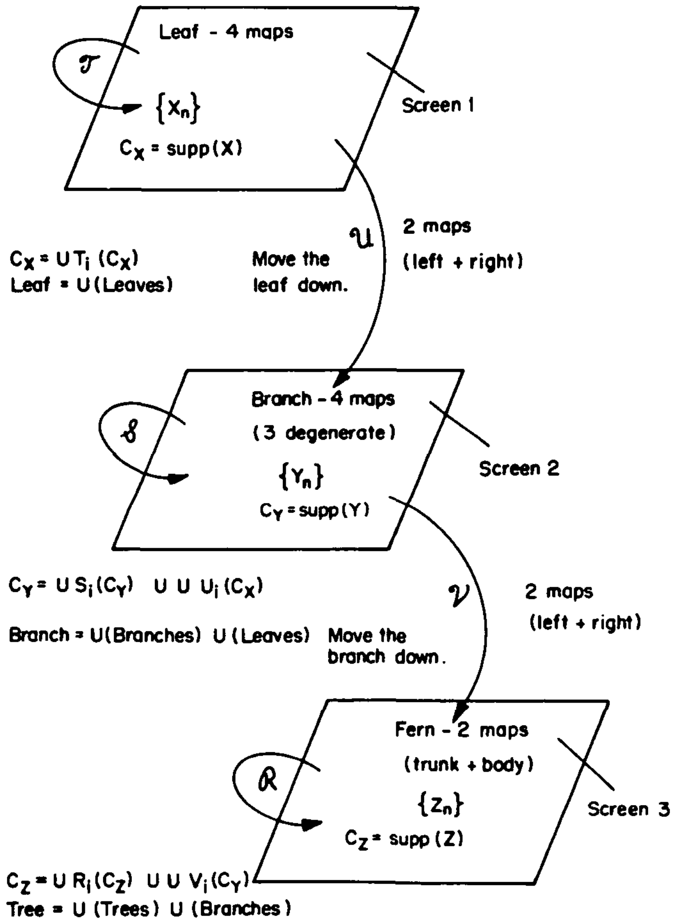



**Figure 15.** Schematic for the mixing model of Figure 14. This model involves 3 screens and 2 crossovers. Only screen 3 gets plotted.

for Borel subsets $B \subset \mathbb{R}^m$. Let $P = (P_{ij})$ be the $N \times N$ transition probability matrix for $\{I_n\}$. Then it follows again from the stationarity of $\nu$ that

$$\nu_i(B) = \mathbb{E}_i \nu_{\hat{I}_1}(\mathfrak{I}^{-1}(i)B) = \sum_{j \in K} \hat{P}_{ij} \nu_j(\mathfrak{I}^{-1}(i)B), \qquad i \in K, \tag{39}$$

where $\hat{P} = (\hat{P}_{ij})$ is the transition probability matrix on $K$ for the *reversed chain* $\{\hat{I}_n: n \geq 1\}$,

$$\hat{P}_{ij} = \frac{\pi(j)}{\pi(i)} P_{ij}, \qquad i,j \in K.$$

If the $\mathfrak{I}(i)$ are all atomic with atoms $T_k(i)$, then the supports $C_i$ of $\nu_i$ will satisfy the coupled set equations

$$C_i = \bigcup_{\substack{j \in K \\ P_{ji}>0}} \bigcup_k \overline{T_k(i)(C(j))} = \bigcup_k \overline{T_k(i)\left( \bigcup_{\substack{j \in k \\ P_{ji}>0}} C(j) \right)}, \qquad i \in K.$$

The Eq. (39) should be contrasted with Eq. (36). In particular, if we let $\mu_{ij}$ and $\mu_i$ denote the distributions of $\mathfrak{I}(i,j)$ from Section 5 and $\mathfrak{I}(i)$ here, respectively, then Eq. (36) becomes

$$\nu_i = \sum_j Q_{ij} \mu_{ij} * \nu_j \tag{40}$$

and Eq. (39) becomes

$$\nu_i = \mu_i * \mathbb{E}_i \nu_{\hat{I}_1} = \mu_i * \left( \sum_j \hat{P}_{ij} \nu_j \right). \tag{41}$$

In particular, Eq. (41) is the special case of Eq. (40) where *the chain* $\{\Gamma_n\}$ *corresponding to* $Q$ *is irreducible*, and where $\mu_{ij} = \mu_i$; i.e., *all maps leading into the same screen are the same.* The reversed chain here corresponds then to the chain $\{\Gamma_n\}$ from Section 5 above. Generating the process $\{X_n\}$ from Eq. (38) is equivalent algorithmically to generating it from Eq. (35). In the framework of Section 5, screen $i$ corresponds here to the event $I = i$. In case the chain $\{(X_n, I_n)\}$ is ergodic, we have

$$\frac{1}{n} \sum_{k=1}^{n} f(X_k) \chi_{\{I_k = i\}} \to \int f(x) \nu(dx \times \{i\}) \qquad \text{a.s.},$$

for $i \in K$, where $f$: $\mathbb{R}^m \to \mathbb{R}$ is bounded continuous. Thus, $\nu_i$ can be generated as the empirical distribution of $\{X_k\}$, but sampled only over those times $k$ when $I_k = i$. This should be contrasted with the $N$ screen simulation of Section 5, whereby the sampling is done over all times $k$.

Regarding conditions for ergodicity, the sufficient condition given in [2] for the chain $\{(X_n, I_n)\}$ to be ergodic goes as follows. The chain $\{I_n\}$ is irreducible, and if $\{I_n'\}$ is the stationary chain then the Lyapunov exponent for the stationary ergodic process $\{\mathfrak{A}(\mathfrak{I}_n(I_n')): n \geq 1\}$ is negative. For the 2-screen model of Sections 3 and 4, if $V = \mathfrak{I}$ and $U = \mathfrak{S}$, then this should be contrasted with the condition that $\mathfrak{A}(\mathfrak{R})$ have a negative Lyapunov exponent, $\mathfrak{R}$ being given by Eq. (12).

In summary, the state space for $\{\Gamma_n\}$ in Section 5 above corresponds to the $N$ screens; and the state space for $\{I_n\}$ here corresponds to the $N$ maps. In terms of the general model described at the end of [7], the process in [2] evolves as

$$X_{n+1} = g(X_n, \xi_{n+1}),$$

where $\{\xi_i\}$ is Markov (rather than just i.i.d.).

## References


1. Barnsley, M.F. & Demko, S.G. (1985). Iterated function systems and the global construction of fractals, *Proceedings of the Royal Society London* A399: 243–275.
2. Barnsley, M.F., Elton, J.H., & Hardin, D.P. (1986). Recurrent iterated function systems. Georgia Institute of Technology. Preprint.
3. Barnsley, M.F. & Sloan, A.D. (1987). Image compression. *Computer Graphics World* 10: 107–108.
4. Berger, M.A. Images generated by orbits of 2-D Markov chains. *CHANCE*, to appear.
5. Berger, M.A. Encoding images through transition probabilities. *Proceedings of the Sixth International Conference on Mathematical Modeling, Mathl. Comput. Modelling* 11(1988), 575–577.
6. Berger, M.A. & Amit, Y. (1986). Products of random affine maps. Weizmann Institute of Science. Preprint.
7. Berger, M.A. & Soner, H.M. Random walks generated by affine mappings. *J. Theor. Prob.* 1 (1988), 239–254.
8. Demko, S.G., Hodges, L., & Naylor, B. (1985). Construction of fractal objects with iterated function systems. *SIGGRAPH* 19: 271–278.
9. Diaconis, P. & Shahshahani, M. (1986). Products of random matrices and computer image generation. *Contemporary Mathematics* 50:173–182.
10. Hutchinson, J. (1981). Fractals and self-similarity. *Indiana University Journal of Mathematics* 30: 713–747.
11. Peterson, I. (1987). Packing it in. *Science News* 131: 283–285.


## Reference added in proofs


12. Barnsley, M.F. & Sloan, A.D. (1988). A better way to compress images. *Byte* 13: 215–223.