ERGODICITY IN PARAMETRIC NONSTATIONARY MARKOV CHAINS: AN APPLICATION TO SIMULATED ANNEALING METHODS

SHOSHANA ANILY

University of British Columbia, Vancouver, British Columbia

AWI FEDERGRUEN

Columbia University, New York, New York

(Received May 1985; revision received May, November 1986; accepted November 1986)

A nonstationary Markov chain is *weakly ergodic* if the dependence of the state distribution on the starting state vanishes as time tends to infinity. A chain is *strongly ergodic* if it is weakly ergodic and converges in distribution. In this paper we show that the two ergodicity concepts are equivalent for finite chains under rather general (and widely verifiable) conditions. We discuss applications to probabilistic analyses of general search methods for combinatorial optimization problems (simulated annealing).

Innumerable practical problems in areas such as inventory management, queueing systems, maintenance, and manpower planning have been modeled as Markov chains. In these discrete-time probability models, the probability distribution of the state of the system at a given stage may depend on the system's state at the previous stage. Most of these models use *stationary* Markov chains, so that the transition probabilities are time-homogeneous; this situation prevails in spite of the fact that important nonstationarities arise in many of the underlying real-world systems. The popularity of the stationary Markov chain model is explained by the fact that its long-run or steady-state behavior is easily characterized and diagnosed. To date, our understanding of the steady-state behavior of *nonstationary chains* is considerably more limited.

This paper provides tools for characterizing the long-run behavior of finite, nonstationary Markov chains in which the time-dependent transition probabilities converge to a limiting matrix. We apply our results to the analysis of simulated annealing methods, which represent a general class of solution methods for combinatorial optimization problems based on randomized local neighborhood searches. Unlike the more traditional deterministic search methods (heuristics), this approach follows a random path in the solution space that allows for occasional changes that worsen the solution, so as to avoid becoming trapped in local optima. In the course of the algorithm, the method varies the probabilities of accepting specific switches; the process may be described as a nonstationary Markov chain.

A finite nonstationary Markov chain is described by a sequence of transition matrices $\{P(k)\}_{k=1}^{\infty}$ defined on a common state space $\{1, \ldots, N\}$. In period k, the system moves from state i to state j with probability $P(k)_{ij}$.

A *stationary* chain with transition matrix P is called *ergodic* if $\lim_{n\to\infty} (P^n_{ij} - P^n_{lj}) = 0$ for all $i, l, j \in \{1, \ldots, N\}$, i.e., if the effect of the starting state vanishes as time progresses. A stationary chain is *ergodic* if and only if it is *aperiodic* and has a *single subchain*, in which case it satisfies a stronger convergence result: a unique steady-state distribution π exists with $\lim_{n\to\infty} P^n_{ij} = \pi_j$ for all $i, j \in \{1, \ldots, N\}$.

For *nonstationary* chains, the effect of the starting state may vanish, while the products $\{P(1) \cdots P(k)\}_{k=1}^{\infty}$ fail to converge. Thus we need to make an essential distinction between two types of ergodicity.

We write $P^{(m,k)}$ for the product $P(m) \cdots P(k)$:

A nonstationary Markov chain is *weakly ergodic* if, for each m, a sequence of row vectors $\pi(m, k)$ exists such that $\lim_{k\to\infty} [P^{(m,k)}_{ij} - \pi(m, k)_j] = 0$ for all $i, j \in \{1, \ldots, N\}$.

A nonstationary Markov chain is *strongly ergodic* if a steady state distribution π exists with $\lim_{k\to\infty} P^{(m,k)}_{ij} = \pi_j$ for all $m \geq 1$ and all $i, j \in \{1, \ldots, N\}$.

Subject classification: 568 nonstationary Markov chains, weak and strong ergodicity, applications to simulated annealing methods.

0030-364X/87/3506-0867 $01.25

In this paper we provide general conditions under which these two ergodicity concepts are equivalent, and we argue that these conditions will be satisfied in all but the most contrived situations. As a side result, we find that, under these conditions, π arises as the limit of steady-state distributions associated with $\{P(k)\}_{k=1}^{\infty}$. The equivalency result implies a significant simplification of strong ergodicity tests. Previous tests for strong ergodicity usually required establishing weak ergodicity as well as a convergence condition with respect to a sequence $\{\pi(k)\}_{k=1}^{\infty}$ of left eigenvectors of the matrices $\{P(k)\}_{k=1}^{\infty}$. The latter condition is often the most intricate part of the ergodicity test, especially in the common case in which these eigenvectors cannot be obtained in closed form. (Sufficient conditions that establish strong ergodicity directly, without proving weak ergodicity, are rather restrictive; see, for example, Isaacson and Madsen 1976, Chapter V.)

After establishing preliminary results in Section 1, in Section 2 we derive equivalence conditions. In Section 3 we apply our results to the analysis of simulated annealing methods. Ideally, the sequence of solutions generated by such a method converges with probability one to the set of global optima. To verify this property, one first needs to establish convergence in distribution of the sequence of generated solutions; in view of the prior observations, this approach reduces to verifying strong ergodicity of the associated nonstationary Markov chain.

Ergodicity concepts for nonstationary Markov chains were first introduced by Dobrushin (1956), Hajnal (1956, 1958) and Mott (1957). Subsequent results were obtained by a variety of researchers, including Iosifescu and Theodorescu (1969), Paz (1963, 1971), Griffeath (1975) and Madsen and Isaacson (1973). Isaacson and Madsen (1976, Chapter V) give a comprehensive survey of most known results up to 1976. The study of backward products of convergent sequences of stochastic matrices is closely related; see Federgruen (1981) and the references therein.

The annealing concept seems to have been introduced in statistical mechanics by Metropolis et al. (1953). Recently, Kirkpatrick, Gelatt and Vecchi (1983) and Cerny (1985) introduced the method as a general solution approach for discrete optimization problems. Their observations, reinforced by articles in the popular press—see, for example, *The Economist* (1984)—led to several successful applications to a variety of problem areas; see Aragon et al. (1987) for a review. Lundy and Mees (1985) and Romeo and Sangiovanni-Vincentelli (1984) were the first to analyze the annealing method as a Markov chain defined on the solution space. Their treatment is, however, restricted to the case in which the probabilities of accepting specific switches remain constant throughout the algorithm (i.e., the process is described by a stationary chain). Geman and Geman (1984), Mitra, Romeo and Sangiovanni-Vincentelli (1986), Hajek (1985), Gidas (1985) and Gelfand and Mitter (1985) analyze nonstationary implementations for the special case of exponential acceptance probabilities (see Equation 8). (Gidas discusses a few closely related cases as well.) Our results are used in Anily and Federgruen (1985) to provide a full probabilistic analysis for general implementations using general acceptance probabilities.

1. Preliminaries

Following convention, we define the ergodic coefficient of a stochastic matrix P, denoted by $\alpha(P)$, by $\alpha(P) = min_{i,k} \sum_j \min(P_{ij}, P_{kj})$. Weak ergodicity of a nonstationary chain $\{P(k)\}_{k=1}^{\infty}$ is most easily verified by inspecting the ergodic coefficients of blocks of consecutive matrices in the chain.

Lemma 1. (see Theorem V.3.2. in Isaacson and Madsen). *Let $\{P(k)\}_{k=1}^{\infty}$ be a nonstationary Markov chain. This chain is weakly ergodic if and only if for some subdivision of $P(1)P(2)P(3)\cdots$ into blocks of matrices* $[P(1)P(2)\cdots P(n_1)]\cdot[P(n_1+1)P(n_1+2)\cdots P(n_2)]\cdots[P(n_j+1)P(n_j+2)\cdots P(n_{j+1})]\cdots$

$$\sum_{j=0}^{\infty} \alpha(P^{(n_j+1,n_{j+1})}) = \infty \quad \textit{where } n_0 = 0. \tag{1}$$

Note that, in verifying (1), we need only identify a sequence of lower bounds $\underline{\alpha}(P^{(n_j+1,n_{j+1})})$ for which $\sum_{j=0}^{\infty} \underline{\alpha}(P^{(n_j+1,n_{j+1})}) = \infty$.

As pointed out in the introduction, weak ergodicity may fail to imply strong ergodicity.

Example 1. (see Example V.4.1 in Isaacson and Madsen). Let

$$P(2n-1) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

and

$$P(2n) = \begin{bmatrix} 0 & 1 \\ 1-\dfrac{1}{2n} & \dfrac{1}{2n} \end{bmatrix} \text{for } n = 1, 2, 3, \ldots.$$

Note that $\alpha(P^{(2n-1,2n)}) = 1/2n$ so that $\{P(k)\}_{k=1}^{\infty}$ is weakly ergodic. Note also by induction that

$P_{22}^{(1,2k)} = 1$ for all $k \geq 1$, while

$$P_{11}^{(1,2k)} = \prod_{l=1}^{k}\left(1 - \frac{1}{2l}\right).$$

Thus

$$\lim_{k\to\infty} P_{11}^{(1,2k)} = 0 \quad \text{as} \quad \sum_{k=1}^{\infty} \frac{1}{2k} = \infty.$$

Thus,

$$\lim_{k\to\infty} \{P(1)P(2)\cdots P(2k-1)P(2k)\} = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix},$$

but

$$\lim_{k\to\infty} P(1)P(2)\cdots P(2k-1)P(2k)P(2k+1)$$

$$= \left\{\lim_{k\to\infty} P(1)P(2)\cdots P(2k)\right\} \lim_{k\to\infty} P(2k+1)$$

$$= \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix},$$

thus showing that $\{P(k)\}_{k=1}^{\infty}$ fails to be strongly ergodic. Of course it contains two converging subsequences.

The preceding example might seem to suggest that the lack of strong ergodicity arises due to the periodicity of $\lim_{n\to\infty} P(n)$. The following example (adapted from example V.5.1 in Isaacson and Madsen) shows, however, that the equivalence between weak and strong ergodicity may fail even when $\lim_{n\to\infty} P(n)$ is aperiodic.

Example 2. Let

$$S(n) = \begin{bmatrix} 1 & 0 \\ \frac{1}{n} & 1 - \frac{1}{n} \end{bmatrix} \text{ and } T(n) = \begin{bmatrix} 1 - \frac{1}{n} & \frac{1}{n} \\ 0 & 1 \end{bmatrix}.$$

For each $n \geq 1$, choose $P(n)$ to be either $S(n)$ or $T(n)$. Note that

$$\lim_{n\to\infty} P(n) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

and

$$\alpha(S(n)) = \alpha(T(n)) = \frac{1}{n},$$

so that each such constructed nonstationary chain $\{P(n)\}_{n=1}^{\infty}$ is weakly ergodic (apply Lemma 1). By appropriate choices of $P(n)$, the products $\{P^{(1,n)}\}_{n=1}^{\infty}$ can, however, be made into an oscillating sequence. In particular, $P_{11}^{(1,n)}$ can be made less than ¼ and greater than ¾ infinitely often.

It is interesting to note that the weak and strong ergodicity concepts of a nonstationary chain $\{P(n)\}_{n=1}^{\infty}$ are equivalent if $P = \lim_{n\to\infty} P(n)$ exists, has a single subchain (as in Example 1), and is aperiodic (as in Example 2).

Lemma 2 (see Theorem V.4.5 in Isaacson and Madsen). *Let $\{P(n)\}_{n=1}^{\infty}$ be a nonstationary Markov chain. If $P(n) \to P$ as $n \to \infty$ where P is ergodic, then the chain is strongly ergodic.*

The nonstationary chains that arise in the analysis of annealing methods (see Section 3) usually have limit matrices with multiple ergodic sets, and Lemma 2 is thus inapplicable. Alternative tests for strong ergodicity require verification of a convergence condition with respect to a sequence $\{\pi(k)\}_{k=1}^{\infty}$ of left eigenvectors of the matrices $\{P(k)\}_{k=1}^{\infty}$.

For $x \in R^N$, let $|x| = \sum_{j=1}^{N} |x_j|$.

Lemma 3 (see the proof of Theorem V.4.3. of Isaacson and Madsen). *Let $\{P(n)\}_{n=1}^{\infty}$ be a weakly ergodic nonstationary Markov chain. If there exists a corresponding sequence of left eigenvectors*

$\{\pi(n)\}_{n=1}^{\infty}$

satisfying

$$\sum_{n=1}^{\infty} |\pi(n) - \pi(n+1)| < \infty, \tag{2}$$

then the chain is strongly ergodic, with $\lim_{n\to\infty} P_{ij}^{(m,n)} = \lim_{n\to\infty} \pi(n)_j$, *for all* $m \geq 1$.

As such, it appears that Lemma 3 can be applied only when a sequence $\{\pi(n)\}_{n=1}^{\infty}$ is known in closed form. In the next section we show that condition (2) holds in all but the most contrived situations. (Note that, in Examples 1 and 2, the entries of the matrices $P(n)$ are allowed to vary erratically with the period index n.) In addition, we give easily verifiable conditions for (2) that do not depend on a closed form representation of a sequence $\{\pi(n)\}_{n=1}^{\infty}$.

2. Weak and Strong Ergodicity: Equivalency Conditions

In this section we assume that $P(n)$ is available as a closed-form function of n. We first rephrase Lemma 3 in terms of continuous extensions of nonstationary Markov chains. In particular, we show that condition (2) is equivalent to the existence of a vector function $\bar{\pi}(c)$ of bounded variation on (0, 1] with the property that, for some sequence $\{c_n\}_{n=1}^{\infty} \downarrow 0$, $\bar{\pi}(c_n)$ is a left eigenvector of $\bar{P}(c_n) = P(n)$ for all $n \geq 1$.

To invoke the "bounded variation" property, we need to extend the sequence $\{P(n)\}_{n=1}^{\infty}$ to a matrix function of a variable defined on a bounded interval. This approach leads to the following somewhat unusual definition of extensions of sequences:

Definition 1. Let $\{a(n)\}_{n=1}^{\infty}$ be a sequence with $a(n) \in R^m$ for some $m \geq 1$. The (vector) function $\bar{a}(c)$: $(0, 1] \to R^m$ is an extension of the sequence if $\bar{a}(c_n) = a(n)$ for some sequence $\{c_n\}_{n=1}^{\infty}$, with $\lim_{n\to\infty} c_n = 0$.

As an example, for a given positive vector $\alpha \in R^m$, let the ith component of $a(n)$ be specified as $a(n)_i = n^{-\alpha_i}$ for $i = 1, \ldots, m$. The function $\bar{a}(\cdot)$ defined by $\bar{a}(c)_i = c^{\alpha_i}$ is an extension of the sequence. (Take $c_n = n^{-1}$ for $n \geq 1$.) We also recall the following standard definition:

Definition 2 (see, for example, p. 98 in Royden 1968). A real-valued function f defined on the interval (0, 1] is of bounded variation if

$$\sup\left\{\sum_{i=1}^{\infty} |f(x_i) - f(x_{i-1})| \,\Big|\, x_i < x_{i-1} < \cdots < x_1 = 1 \text{ and } \lim_{i\to\infty} x_i = 0\right\} < \infty. \quad (3)$$

Theorem 1. *Let $\{P(n)\}_{n=1}^{\infty}$ be a weakly ergodic nonstationary Markov chain and let $\bar{P}(c)$ be an extension corresponding to a sequence $\{c_n\}_{n=1}^{\infty} \downarrow 0$.*

(a) *Let $\bar{\pi}(c)$ be a (vector) function of bounded variation in c, with $\bar{\pi}(c_n)$ a left eigenvector of $\bar{P}(c_n) = P(n)$ for all $n \geq 1$. Then $\{P(n)\}_{n=1}^{\infty}$ is strongly ergodic and $\lim_{n\to\infty} P_{ij}^{(m,n)} = \lim_{c\downarrow 0} \bar{\pi}(c)_j$ for all $1 \leq i, j \leq N$ and all $m \geq 1$.*

(b) *Let $\{\pi(n)\}_{n=1}^{\infty}$ be a sequence of left eigenvectors of $\{P(n)\}_{n=1}^{\infty}$. If $\sum_{n=1}^{\infty} |\pi(n+1) - \pi(n)| < \infty$, then some extension $\bar{\pi}(c)$ of $\{\pi(n)\}_{n=1}^{\infty}$ is of bounded variation.*

Proof. (a) Immediate from the definition of bounded variation and Lemma 3. (b) Consider the step function $\bar{\pi}(c) = \pi(\lfloor c^{-1} \rfloor)$ where $\lfloor x \rfloor$ is the largest integer smaller than or equal to x. Consider a decreasing sequence $\{c_i\}_{i=1}^{\infty}$ with $\lim_{i\to\infty} c_i = 0$.

$$\sum_{i=2}^{\infty} |\bar{\pi}(c_i) - \bar{\pi}(c_{i-1})| = \sum_{i=2}^{\infty} |\pi(\lfloor c_i^{-1} \rfloor) - \pi(\lfloor c_{i-1}^{-1} \rfloor)|.$$

Let $\{n_i\}_{i=1}^{\infty}$ be the collection of *distinct* integers in $\{\lfloor c_i^{-1} \rfloor\}_{i=1}^{\infty}$ and note that $\{n_i\}_{i=1}^{\infty}$ is a subsequence of the positive integers. Hence,

$$\begin{aligned}
&\sum_{i=2}^{\infty} |\bar{\pi}(c_i) - \bar{\pi}(c_{i-1})| \\
&= \sum_{i=2}^{\infty} |\pi(n_i) - \pi(n_{i-1})| \\
&\leq \sum_{i=2}^{\infty} \{|\pi(n_i) - \pi(n_i - 1)| \\
&\quad + |\pi(n_i - 1) - \pi(n_i - 2)| \\
&\quad + \cdots + |\pi(n_{i-1} + 1) - \pi(n_{i-1})|\} \\
&\leq \sum_{n=2}^{\infty} |\pi(n) - \pi(n-1)| < \infty.
\end{aligned}$$

This result verifies (2) for all $\pi(\cdot)_j$ for $j = 1, \ldots, N$.

The remainder of this section is devoted to identifying simply verifiable sufficient conditions under which a function $\bar{\pi}(c)$ exists that is of bounded variation and for which $\bar{\pi}(c_n)P(n) = \bar{\pi}(c_n)$, for some $\{c_n\}_{n=1}^{\infty} \downarrow 0$.

Example 3 shows, unfortunately, that it is *insufficient* to require that some extension $\bar{P}(c)$ of $\{P(n)\}_{n=1}^{\infty}$ is of bounded variation.

Example 3. Consider the nonstationary Markov chain

$$P(n) = \begin{bmatrix} 1 - e^{-n} & e^{-n} \\ e^{-n}\sin^2\left(\frac{n\pi}{2}\right) & 1 - e^{-n}\sin^2\left(\frac{n\pi}{2}\right) \end{bmatrix}.$$

Observe that the extension

$$\bar{P}(c) = \begin{bmatrix} 1 - e^{-1/c} & e^{-1/c} \\ e^{-1/c}\sin^2\left(\frac{\pi}{2c}\right) & 1 - e^{-1/c}\sin^2\left(\frac{\pi}{2c}\right) \end{bmatrix}$$

is of bounded variation on (0, 1]. (Note that $d\bar{P}(c)/dc$ exists and is continuous on (0, 1] with $\lim_{c\downarrow 0} (d\bar{P}(c)/dc = 0$. Thus $\bar{P}(c)$ is of bounded variation; see, for example, Lemma 6, p. 101, in Royden.) The unique left eigenvector $\pi(n)$ of $P(n)$ is obtained as the solution of the following system of equations:

$$\pi(n)_1 + \pi(n)_2 = 1,$$

$$(1 - e^{-n})\pi(n)_1 + \left(e^{-n}\sin^2\left(\frac{n\pi}{2}\right)\right)\pi(n)_2 = \pi(n)_1$$

Hence $\{\pi(n)_2\}_{n=1}^{\infty} = \{[1 + \sin^2(n\pi/2)]^{-1}\}_{n=1}^{\infty}$ alternates between 1 and ½ and condition (2), that $\sum_{n=1}^{\infty} |\pi(n) - \pi(n+1)| < \infty$, fails to hold. Observe that this example is all the more remarkable as each of the entries of the matrix $P(n)$ is an extremely smooth function of n. In fact each of the entries is expandable as a Taylor series in n with infinite convergence radius.

The problem, in this example, arises due to the fact that $P(n)_{21}$ and $P(n)_{22}$, and hence $\pi(n)_2$, have infinitely many local optima in n. This connection is explained by the following lemma.

Lemma 4. *Let $\bar{\pi}(c)$ be an extension of a sequence $\{\pi(n)\}_{n=1}^{\infty}$ that is continuously differentiable on* (0, 1]. *If $\bar{\pi}(c)$ is asymptotically monotone (i.e., each component has finitely many local optima) then $\bar{\pi}(c)$ is of bounded variation on* (0, 1].

Proof. There exists a constant $c^* > 0$ such that $\bar{\pi}(c)$ is monotone for $c < c^*$. In view of (3), it suffices to show that $\bar{\pi}(c)$ is of bounded variation on $[c^*, 1]$. The latter follows from $\bar{\pi}(\cdot)$ being continuously differentiable on $[c^*, 1]$ (see Lemma 6, p. 101, in Royden).

Alternatively, one may view the problem in Example 3 as arising because ratios of pairs of entries of the matrices $\{P(n)\}_{n=1}^{\infty}$ fail to be of bounded variation (even though the entries themselves are).

We now focus on identifying easily verifiable conditions under which bounded variation of an extension $\bar{\pi}(c)$ can be proven. These conditions restrict the choice of $\{\bar{P}(c)_{ij}: 1 \leq i, j \leq N\}$ to specific classes of functions while imposing a regularity assumption with respect to the chain structure of the matrices $\bar{P}(c)$, $0 < c \leq 1$.

Definition 3. A class $F \subset C^1$ of functions defined on (0, 1] is a closed class of asymptotically monotone functions (*CAM*) if

(a) $f \in F \Rightarrow f' \in F$ and $-f \in F$;
(b) $f, g \in F \Rightarrow (f + g)$ and $(f \cdot g) \in F$; and
(c) all $f \in F$ change signs finitely often on [0, 1].

Definition 4. A class F of functions defined on (0, 1] is a rationally closed class of bounded variation (*RCBV*) if

(a) $f \in F \Rightarrow f$ is of bounded variation on (0, 1];
(b) $f \in F \Rightarrow -f \in F$;
(c) $f, g \in F \Rightarrow (f + g)$ and $(f \cdot g) \in F$; and
(d) $f, g \in F$ with (f/g) bounded on (0, 1] $\Rightarrow$ f/g is of bounded variation.

Definition 5. A nonstationary Markov chain $\{P(n)\}_{n=1}^{\infty}$ is said to have the regular extension $\bar{P}(\cdot)$ if a real number $c^* > 0$ exists such that the collection of subchains of $\bar{P}(c)$ is identical for all $c < c^*$. (4)

A sufficient (though sometimes too restrictive: see Example 1) condition for regularity arises when we replace (4) by the assumption that $\{(i, j) \mid \bar{P}(c)_{ij} > 0\}$ is identical for all $c < c^*$. (5)

Theorem 2. (Main result.) *Let $\{P(n)\}_{n=1}^{\infty}$ be a weakly ergodic nonstationary Markov chain and $\bar{P}(c)$ a regular extension such that all entry functions $\bar{P}(c)_{ij}$ $(1 \leq i, j \leq N)$ belong to*

(i) *a closed class of asymptotically monotone functions, or*
(ii) *a rationally closed class of functions of bounded variation.*

Then $\{P(n)\}_{n=1}^{\infty}$ is strongly ergodic. Moreover, for n sufficiently large, each $P(n)$ has a unique steady-state distribution $\pi(n)$ with $\lim_{n\to\infty} \pi(n) = \pi^*$ and $\lim_{n\to\infty} P_{ij}^{(m,n)} = \pi_j^*$ for all $1 \leq i, j \leq N$ and $m \geq 1$.

Proof. We first show that, for all c sufficiently small, $\bar{P}(c)$ has a unique subchain. In view of the regularity of $\bar{P}(\cdot)$, the only alternative is the existence of an integer m and *two* sets of states, R_1 and R_2, such that both R_1 and R_2 are subchains of $P(n)$, for all $n \geq m$. But then $\lim_{n\to\infty} \sum_{j\in R_1} P_{ij}^{(m,n)} = 1$ for $i \in R_1$, *and* $\lim_{n\to\infty} \sum_{j\in R_1} P_{ij}^{(m,n)} = 0$ for $i \in R_2$, contradicting weak ergodicity.

Thus, for all c sufficiently small, let $\bar{\pi}(c)$ be the unique steady-state distribution of $\bar{P}(c)$. Note that $\bar{\pi}(c)$ may be obtained as the unique solution of the following system of equations:

$$\bar{\pi}(c)_i = \sum_j \bar{P}(c)_{ji}\bar{\pi}(c)_j, \quad i = 1, \ldots, N-1$$

$$\sum_{i=1}^{N} \bar{\pi}(c)_i = 1.$$

Thus,

$$\bar{\pi}(c) = \begin{bmatrix} 1-\bar{P}(c)_{11} & \cdots & -\bar{P}(c)_{N-1,1} & -\bar{P}(c)_{N,1} \\ \vdots & & & \\ -\bar{P}(c)_{1,N-1} & & 1-\bar{P}(c)_{N-1,N-1} & -\bar{P}(c)_{N,N-1} \\ 1 & & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}.$$

In view of Cramer's rule, each component of $\bar{\pi}(c)$ may thus be written in the form

$$\bar{\pi}(c)_i = \frac{Q_1(c)}{Q_2(c)},$$

where both Q_1 and Q_2 are finite sums (and differences) of finite products of the functions $\{\bar{P}(c)_{ij};\ 1 \leq i, j \leq N\}$. If the latter are chosen from a RCBV class, all components of $\bar{\pi}(c)$ are of bounded variation, and the theorem follows from Lemma 3. If the functions $\{\bar{P}(c)_{ij};\ 1 \le i, j \le N\}$ are chosen from a CAM class of functions, the derivatives $\bar{\pi}'(c)_i$ $(i = 1, \ldots, N)$ may be written as

$$\bar{\pi}'(c)_i = \frac{Q_1'(c)Q_2(c) - Q_1(c)Q_2'(c)}{\{Q_2(c)\}^2},$$

where the numerator is, again, the difference of two finite sums of finite products of the functions $\{\bar{P}(c)_{ij};\ 1 \le i, j \le N\}$. Thus, $\bar{\pi}'(c)_i$ $(1 \le i \le N)$ changes signs finitely often on (0, 1], and $\bar{\pi}(c)_i$ is asymptotically monotone and of bounded variation. The theorem follows again by invoking Lemma 3.

The following proposition enumerates a number of RCBV and CAM classes of functions. We first need the following definitions.

Definition 6. A real-valued function f is an exponential sum in $1/c$ if it is of the form $\sum_{j=1}^{n} Q_j(1/c)e^{\lambda_j/c}$, with λ_j a given real number and $Q_j(\cdot)$ a given polynomial $(j = 1, \ldots, n)$.

Definition 7. A real-valued function f is an exponential rational in $1/c$ if it is the ratio of two exponential sums in $1/c$. Note that the exponential rationals contain the rational functions as well as the exponential sums as subclasses.

Proposition 1. *The following classes of functions defined on* (0, 1] *are RCBV and CAM*:

(I) *the polynomials (RCBV and CAM)*;
(II) *the rational functions (RCBV and CAM)*;
(III) *the piecewise rationals (RCBV), and*
(IV) *the exponential rationals (CAM).*

Proof. The statements about (I)–(III) are immediate. To verify (IV), we need only to check property (*c*) in Definition 3, the other two properties being immediate. Substituting $x = c^{-1}$, we may write the derivative of a rational exponential (with respect to x) in the form

$$\frac{\sum_{j=1}^{n_1} Q_j^1(x)e^{\lambda_j^1 x}}{\{\sum_{j=1}^{n_2} Q_j^2(x)e^{\lambda_j^2 x}\}^2},$$

with $Q_j^1(x)$, $Q_j^2(x)$ given polynomials and λ_j^1, λ_j^2 $(1 \le j \le n)$ given constants. In view of Lemma 4, it thus suffices to prove that the sign of an exponential sum $\sum_{j=1}^{n} \sum_{l=1}^{n_j} a_{j,l}x^l e^{\lambda_j x}$ remains constant for x sufficiently large. Assume, without loss of generality, that $\lambda_1 \ge \cdots \ge \lambda_n$ and $a_{j,n_j} \ne 0$ $(j = 1, \ldots, n)$. It is easy to verify that

$$\lim_{x\to\infty} \operatorname{sign}\left\{\sum_{j=1}^{n} \sum_{l=1}^{n_j} a_{j,l}x^l e^{\lambda_j x}\right\} = \operatorname{sign}(a_{1,n_1}).$$

Thus, if $\{P(n)\}_{n=1}^{\infty}$ is a weakly ergodic chain with a regular extension $\bar{P}(c)$ $(c \in (0, 1])$ whose entries are chosen from any of the above classes of functions, then the chain $\{P(n)\}_{n=1}^{\infty}$ is strongly ergodic.

3. Convergence of Simulated Annealing Methods

Consider a discrete minimization problem with N feasible solutions numbered in ascending order of their objective function values $(f_i,\ i = 1, \ldots, N)$. Iterative search methods specify a neighborhood $N_i \subset \{1, \ldots, N\}$ for each solution i $(i = 1, \ldots, N)$. (Usually, $|N_i| \ll N$.) The classical deterministic methods start with an arbitrary solution, and proceed to its lowest valued neighbor, provided this switch results in a strict improvement. The methods repeat this step until no improving neighbor can be found, i.e., until they reach a local optimum. Examples of such methods are the interchange or greedy heuristics for the simple plant location problem (Cornuejols et al. 1977), and the r-opt methods or the Lin and Kernighan (1973) heuristic for the traveling salesman problem.

The classical deterministic methods encounter two major problems:

(i) the solution obtained is heavily dependent on the starting solution; and
(ii) the methods often converge to inferior local optima.

Simulated annealing methods randomize the procedure to overcome these problems, allowing for occasional switches that worsen the solution. Assume the current solution is i $(1 \le i \le N)$. A specific neighbor $j \in N_i$ is generated with probability g_{ij}. The switch (between i and j) is implemented according to a positive acceptance probability a_{ij} that depends on a control parameter c (i.e., $a_{ij} \equiv a_{ij}(c)$). In the course of the algorithm, the method decreases this control parameter to zero according to a prespecified sequence $\{c_n\}_{n=1}^{\infty}$. The acceptance probability functions satisfy the following properties:

$$a_{ij}(c) = 1 \quad \text{if } f_j \le f_i, \tag{6}$$

$$\lim_{c\downarrow 0} a_{ij}(c) = 0 \quad \text{if } f_j > f_i. \tag{7}$$

Examples of frequently used acceptance probabilities (Kirkpatrick, Gelatt and Vecchi; Lundy and Mees) are

$$a_{ij}(c) = e^{(f_i - f_j)/c}, \quad f_j > f_i. \tag{8}$$

A crucial element in designing a simulated annealing method is choosing a sequence $\{c_n\}_{n=1}^{\infty}$ to ensure convergence to the set of global minima. Let s_n be the solution generated at the nth iteration. Observe that the sequence $\{s_n\}_{n=1}^{\infty}$ is generated by a nonstationary Markov chain with state space $\{1, \ldots, N\}$, and transition probabilities

$$P_{ij}(n) = \begin{cases} g_{ij} a_{ij}(c_n), & j \in N_i, \\ 1 - \sum_{j \in N_i} g_{ij} a_{ij}(c_n), & j = i, \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

Also, assume a global minimum that can be reached from any initial solution (following a path through successive neighborhoods). Thus, $P(n)$ has a single subchain for all $n \geq 1$; let $\pi(n)$ be its unique steady-state distribution.

The following three properties (in order of increasing strength) are desired for any annealing method:

(a) (*Asymptotic independence of starting solution.*) The dependence of the distribution of s_k with respect to the starting solution vanishes as $k \to \infty$.
(b) (*Convergence in distribution.*) s_k converges in distribution.
(c) (*Convergence to a global minimum.*) The algorithm converges to M, the set of global minima, with probability one.

Note that the first (second) property is equivalent to weak (strong) ergodicity of $\{P(n)\}_{n=1}^{\infty}$. As Section 1 pointed out, the first property is usually verified easily with the help of Lemma 1, as the following example demonstrates.

Example 4. Consider a problem with four feasible solutions, i.e., $N = 4$. Let $f_1 = 0, f_2 = 1, f_3 = f_4 = 2$. Let $N_1 = N_2 = \{3, 4\}$ and $N_3 = N_4 = \{1, 2\}$. Consider a simulated annealing method with acceptance probabilities (8) and $g_{ij} = 1/2$ for $j \in N_i$ ($i = 1, \ldots, 4$). Note that $\alpha(P_n) = e^{-2/c_n}$ for n sufficiently large. Thus, with $c = 2/\log n$, we have $\alpha(P_n) = n^{-1}$ and (1) is satisfied with the choice $n_j = j, j \geq 1$. Hence (a) holds.

The following corollary, immediate from Theorem 2, shows that an annealing method converges in distribution (property (b)) to $\pi^* =^{\text{def}} \lim_{n\to\infty} \pi(n)$ *for almost all reasonable acceptance probabilities*, provided $\{P(n)\}_{n=1}^{\infty}$ is weakly ergodic. Moreover, investigating whether the algorithm converges to a global optimum (property (c)) thus reduces to verifying whether $\sum_{i\in M} \pi_i^* = 1$. (Note that in $P = \lim_{n\to\infty} P(n)$, each local optimum may represent a subchain by itself; thus, Lemma 2 cannot be applied.)

Corollary 1. *Consider a simulated annealing method with a given neighborhood structure* $\{N_i, i = 1, \ldots, N\}$, *generation probabilities* $\{g_{ij}\}$, *and control sequence* $\{c_n\}_{n=1}^{\infty}$. *Assume* $\{P(n)\}_{n=1}^{\infty}$ *is weakly ergodic and let the acceptance probability functions be chosen from a RCBV and CAM family of functions. Then, the distribution of* s_n *converges to* $\pi^* = \lim_{n\to\infty}\pi(n)$.

Anily and Federgruen (Theorem 3) derive an upper bound on $\text{Prob}\{s_k \in M\}$ ($k \geq 1$) that may be used in formulating a stopping criterion. (Current implementations employ a variety of heuristic termination rules.) This bound also provides insight into the *transient* behavior of the chain, as well as its convergence rate.

Next, consider the exponential acceptance probability functions (8) and assume the generation probabilities (g_{ij}) are generated from a symmetric matrix $Q = (Q_{ij})$ satisfying

$$g_{ij} = \frac{Q_{ij}}{\sum_l Q_{il}}, \quad j \in N_i. \tag{10}$$

(In particular, (10) implies that the neighbor relation is symmetric, i.e., $j \in N_i \Leftrightarrow i \in N_j$; (10) includes the special case in which Q is a symmetric Boolean matrix such that $g_{ij} = 1/|N_i|$, i.e., the neighborhood relation is symmetric and the generation probabilities are uniform). In a sequel paper (Anily and Federgruen), we show the existence of two constants K_1 and K_2 such that $\{P(n)\}_{n=1}^{\infty}$ is weakly ergodic if $c_n \leq (K_1/\log n)$ but fails to be weakly ergodic if $c_n \geq (K_2/\log n)$ for n sufficiently large. It is also easy to verify that $\sum_{i\in M} \pi_i^* = 1$ (see Lundy and Mees, and Anily and Federgruen). In view of Theorem 2 or Corollary 1 we arrive at the following conclusion.

Corollary 2 (see also Anily and Federgruen, Theorem 2(i)). *Consider a simulated annealing method with generation probabilities satisfying* (10), *and acceptance probabilities* (8). *There exists a constant* K_1 *such that the distribution of* s_n *converges to* π^* *if* $\liminf_{n\to\infty}\{c_n \log n\} \geq K_1$.

As pointed out in the preceding discussion, Lemma 1's characterization of weak ergodicity is easily used to verify weak ergodicity. It suffices to identify *one* subdivision of $P(1)P(2) \cdots$ into blocks of matrices such that (1) holds. Disproving weak ergodicity through Lemma 1 is usually intractable since a violation of (1) needs to be demontrated for *all* possible subdivisions of the chain. Theorem 2, on the other hand, is often easily applied to disprove weak ergodicity. Consider, e.g., Example 4; however, with $c_n = (1/2 \log n)$ for $n \geq 1$. It is easy to verify that $\lim \inf_{n\to\infty} P_{22}^{(1,n)} \geq \pi_{l=1}^{n} (1 - 1/l^2) > 0$ since $\sum_{l=1}^{\infty} (1/l^2) < \infty$. Corollary 1 and the fact that $\lim_{n\to\infty} \pi(n) = (1, 0, 0, 0)$ implies that for this choice of $\{c_n\}_{n=1}^{\infty}$ the chain $\{P(n)\}_{n=1}^{\infty}$ fails to be weakly ergodic.

Finally, it is worth noting that the two existing ergodicity concepts for nonstationary chains deal with chains that asymptotically behave as "unichain" systems. A "multichain" generalization of strong ergodicity would be the existence of a matrix Π (not necessarily with constant rows) satisfying $\lim_{n\to\infty} P_{ij}^{(m,n)} = \Pi_{ij}$ for all i, j. Future work should address conditions under which this multichain ergodicity property holds.

Acknowledgment

We gratefully acknowledge numerous suggestions by the reviewers (particularly those of the Associate Editor, Martin L. Puterman) regarding the exposition of this paper.

References

ANILY, S., AND A. FEDERGRUEN. 1985. Simulated Annealing Methods with General Acceptance Probabilities. *J. Appl. Prob.* **19**.

ARAGON, C., D. JOHNSON, L. MCGEOGH AND C. SCHEVON. 1987. Optimization by Simulated Annealing: An Experimental Evaluation (in preparation).

CERNY, V. 1985. A Thermodynamical Approach to the Travelling Salesman Problem: An Efficient Simulation Algorithm. *J. Optim. Theory Appl.* **15,** 41–51.

CORNEUJOLS, G., M. FISHER AND G. NEMHAUSER. 1977. Location of Bank Accounts to Optimize Float: An Analytic Study of Exact and Approximate Solutions. *Mgmt. Sci.* **23,** 789–811.

DOBRUSHIN, R. 1956. Central Limit Theorems for Non-Stationary Markov Chains II (English translation). *Theory Prob. Its Appl.* **1,** 329–383.

Economist, The. 1984. Problem Solving: Smart Guess Work. July 28, p. 72, London.

FEDERGRUEN, A. 1981. The Rate of Convergence for Backward Products of a Convergent Sequence of Finite Markov Chains. *Stochast. Proc. Its Appl.* **11,** 187–192.

GELFAND, S., AND S. MITTER. 1985. Analysis of Simulated Annealing for Optimization. Technical Report LIDS-R-1494, Laboratory for Information and Decision Systems, MIT, Cambridge, Mass.

GEMAN, S., AND D. GEMAN. 1984. Stochastic Relaxation, Gibbs Distribution, and the Bayesian Restoration of Images. *IEEE Proc. Pattern Anal. Mach. Intell.* **6,** 721–741.

GIDAS, B. 1985. Non-Stationary Markov Chains and Convergence of the Annealing Algorithm. *J. Stat. Phys.* **39,** 73–131.

GRIFFEATH, D. 1975. Uniform Coupling of Nonhomogeneous Markov Chains. *J. Appl. Prob.* **12,** 753–778.

HAJEK, B. 1985. Cooling Schedules for Optimal Annealing (working paper; submitted to *Math. Opns. Res.*).

HAJNAL, J. 1956. The Ergodic Properties of Non-Homogeneous Finite Markov Chains. *Proc. Cambridge Philos. Soc.* **52,** 67–77.

HAJNAL, J. 1958. Weak Ergodicity in Nonhomogeneous Markov Chains. *Proc. Cambridge Philos. Soc.* **54,** 233–246.

IOSIFESCU, M., AND R. THEODORESCU. 1969. *Random Processes and Learning.* Springer-Verlag, Berlin.

ISAACSON, D., AND R. MADSEN. 1976. *Markov Chains: Theory and Applications.* John Wiley & Sons, New York.

KIRKPATRICK, S., C. GELATT, JR. AND M. VECCHI. 1983. Optimization by Simulated Annealing. *Science* **220,** 671–680.

LIN, S., AND W. KERNIGHAN. 1973. An Effective Heuristic Algorithm for the Travelling Salesman Problem. *Opns. Res.* **21,** 498–516.

LUNDY, M., AND A. MEES. 1985. Convergence of the Annealing Algorithm. *Math. Program.* **34,** 111–124.

MADSEN, R., AND D. ISAACSON. 1973. Strongly Ergodic Behavior for Nonstationary Markov Processes. *Ann. Prob.* **1,** 324–335.

METROPOLIS, W., A. ROSENBLUTH, M. ROSENBLUTH, A. TELLER AND E. TELLER. 1953. Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.* **21,** 1087–1092.

MITRA, D., F. ROMEO AND A. SANGIOVANNI-VINCENTELLI. 1986. Convergence and Finite-Time Behavior of Simulated Annealing. *J. Appl. Prob.* **18,** 741–747.

MOTT, J. 1957. Conditions for the Ergodicity of Non-Homogeneous Finite Markov Chains. *Proc. Roy. Soc. Edin.* **64,** 369–380.

PAZ, A. 1963. Graph Theoretic and Algebraic Characterizations of Some Markov Processes. *Isr. J. Math.* **3,** 109–180.

PAZ, A. 1971. *Introduction to Probabilistic Automata.* Academic Press, New York.

ROMEO, F., AND A. SANGIOVANNI-VINCENTELLI. 1984. Probabilistic Hill Climbing Algorithms: Properties and Applications. Electronics Research Laboratory, University of California, Berkeley, Calif.

ROYDEN, H. 1968. *Real Analysis,* Ed. 2. Macmillan, London.