The Annals of Probability
1989, Vol. 17, No. 4, 1483–1502

A LIMIT THEOREM FOR A CLASS OF INHOMOGENEOUS MARKOV PROCESSES[1]

BY TZUU-SHUH CHIANG AND YUNSHYONG CHOW

Academia Sinica

Let $\{X(t)\colon t \in R^+ \text{ or } I^+\}$ be an (aperiodic) irreducible Markov process with a finite state space S and transition rate $q_{ij}(t) = p(i, j)(\lambda(t))^{U(i,j)}$, where $0 \le U(i, j) \le \infty$ and $\lambda(t)$ is some suitable rate function with $\lim_{t\to\infty}\lambda(t) = 0$. We shall show in this article that there are constants $h(i) \ge 0$ and $\beta_i > 0$ such that independent of $X(0)$, $\lim_{t\to\infty} P(X(t) = i) \div (\lambda(t))^{h(i)} = \beta_i$ for each $i \in S$. The height function h is determined by $(p(i, j))$ and $(U(i, j))$. In particular, a limit distribution exists and concentrates on $\underline{S} = \{i \in S\colon h(i) = 0\}$.

1. Introduction. Let $\{X(t)\colon t \in R^+ \text{ or } I^+\}$ be a continuous time or discrete time irreducible Markov process with a finite state space $S = \{1, 2, \dots, a\}$ and transition rate $(q_{ij}(t))$. We assume it is aperiodic (see Remark 1, Section 5) in the discrete time case. In the case $(q_{ij}(t)) = (q_{ij})$, $\{X(t)\}$ becomes homogeneous and it is well-known [Chung (1967)] that $X(t)$ has a limit distribution independent of $X(0)$. The main purpose of this article is to show that a similar result holds for the inhomogeneous Markov processes whose transition rates are of the type

$$q_{ij}(t) = \begin{cases} p(i, j)(\lambda(t))^{U(i,j)} & \text{if } j \ne i, \\ 1_{\{\text{discrete time}\}} - \sum_{k \ne i} q_{ik}(t) & \text{if } j = i, \end{cases} \tag{1.1}$$

where $\lambda(t)$ is a suitable rate function with $\lim_{t\to\infty}\lambda(t) = 0$, $P = (p(i, j))$ is a matrix with $p(i, j) \ge 0$ for $i \ne j$, and $U\colon S \times S \to [0, \infty]$ is a "cost" function which measures the degree of "reachability" from one state to another. Without loss of generality we may assume $U(i, i) = 0$ and $U(i, j) = \infty$ iff $p(i, j) = 0$. Note that it is clear from (1.1) that $p(i, i)$ is irrelevant and can be arbitrarily defined. We require $\{X(t)\}$ be irreducible in the sense that any state j can be reached from any other state i, that is, there exists a sequence of states $i_0 = i, i_1, \dots, i_n = j$ such that $p(i_k, i_{k+1}) > 0$ [or equivalently, $U(i_k, i_{k+1}) < \infty$] for each $0 \le k < n$. In this case both U and P are also called irreducible. (See Remark 2, Section 5.)

Markov processes of the above-mentioned type appeared in problems of simulated annealing [Geman and Geman (1984) and Kirkpatrick, Gebatt and Vecchi (1983)], where $U(i, j) = (u(j) - u(i))^+$ is determined by a potential function u and $T(t) = (-\log\lambda(t))^{-1}$ is called the "temperature" at time t. In this special case it is shown in Chiang and Chow (1988) that under some natural

Received December 1987; revised October 1988.
[1]Partially supported by the National Science Council, Republic of China.
AMS 1980 *subject classifications*. Primary 60J27; secondary 60F05, 60F10.
Key words and phrases. Forward equations, Perron–Frobenius theorem, cycle method, inhomogeneous Markov process, convergence rate.

conditions on P and λ, there exist positive constants $\beta_1, \beta_2, \dots, \beta_a$ such that

$$(1.2) \qquad \lim_{t\to\infty} P(X(t) = i)/(\lambda(t))^{u(i)-\min u} = \beta_i \quad \text{for each } i \in S.$$

In particular, for the global minimum set $\underline{S} = \{i: u(i) = \min u\}$, one has [Gidas (1985) and Hajek (1988)]

$$(1.3) \qquad \lim_{t\to\infty} P(X(t) \in \underline{S}) = 1.$$

For a general cost function U, the situation becomes much more complicated. One first has to define a function in place of $u(i) - \min u$ that appeared in (1.2) and then impose suitable conditions on $\lambda(t)$ so that the limit theorem (1.2) holds for the general case as well. We employ a generalized cycle method (see Remark 3, Section 5) in this article to tackle these two problems. After having found the correct quantities, we then show that some of the techniques in Chiang and Chow (1988) can be pushed through to yield all the necessary estimates for (1.2). The concept now is basically very simple and deals with the balance of mass flow among different cycles. We, however, think it is worthwhile because it is more natural and general than the ones existing in the literature. (See also Remarks 5 and 6, Section 5.)

For the convenience of discussion, we assume all the cost functions to be *integer-valued*. It should be apparent from the discussion that this is only for technical reasons.

We start by introducing some definition and notation. Let $\tilde{S}$ be a finite set and $\tilde{U}$ a cost function on $\tilde{S}$. For any two states $i, j \in \tilde{S}$, we say that $i \geq j$ (relative to $\tilde{U}$) if there exist $i_0 = i, i_1, \dots, i_m = j$ such that

$$\tilde{U}(i_v, i_{v+1}) = \min_{i_v \neq z \in \tilde{S}} \tilde{U}(i_v, z) \quad \text{for each } v,$$

and $i \geq j$ at level k if, in addition, $\tilde{U}(i_v, i_{v+1}) \leq k$ for each v. A state i is called minimal at level k if $j \geq i$ at level k whenever $i \geq j$ at level k. Two different states $i, j \in \tilde{S}$ are said to be equivalent at level k ($i \sim_k j$) if (i) i is minimal at level k and (ii) $i \geq j$ and $j \geq i$ both at level k. We always assume $i \sim_k i$. An equivalent class of $\tilde{S}$ under the equivalence relation "$\sim_k$" is called a kth-order cycle of $\tilde{S}$ (relative to $\tilde{U}$), and a nontrivial kth-order cycle of $\tilde{S}$ if it has more than one element. (See Remarks 3 and 4, Section 5.)

For $S = \{1, 2, \dots, a\}$ and U given in (1.1), we can successively define (S^n, U^n, V^n), $n = 0, 1, 2, \dots$, as follows: Let $(S^0, U^0) = (S, U)$ and $V^0(i) = \min_{i \neq j \in S^0} U(i, j)$ for each $i \in S^0$. Having defined S^{n-1}, U^{n-1} and V^{n-1}, let $S^n = \{(n-1)\text{th-order cycles of } S^{n-1} \text{ (relative to } U^{n-1})\}$ and for any $C^n = \{C_i^{n-1}: C_i^{n-1} \in C^n\}$, $\tilde{C}^n = \{\tilde{C}_j^{n-1}: \tilde{C}_j^{n-1} \in \tilde{C}^n\}$ in S^n, let

$$
\begin{aligned}
& d_{n-1}(C^n) = \max_i V^{n-1}(C_i^{n-1}), \\
& R(C^n) = \left\{C_i^{n-1} \in C^n: V^{n-1}(C_i^{n-1}) = d_{n-1}(C^n)\right\}, \\
(1.4) \quad & U^n(C^n, \tilde{C}^n) = d_{n-1}(C^n) + \min_{i,j}\left\{U^{n-1}(C_i^{n-1}, \tilde{C}_j^{n-1}) - V^{n-1}(C_i^{n-1})\right\}, \\
& V^n(C^n) = \min_{C^n \neq \tilde{C}^n \in S^n} U^n(C^n, \tilde{C}^n).
\end{aligned}
$$

Note that $U^n(C^n, C^n) = 0$ and U^n is irreducible on S^n by the irreducibility assumption on U. We say that $(C_i^{n-1}, \tilde{C}_j^{n-1}) \in \mathrm{Ix}(C^n, \tilde{C}^n)$, if (1.4) is attained at $(C_i^{n-1}, \tilde{C}_j^{n-1})$. Otherwise,

$$U^{n-1}\big(C_i^{n-1}, \tilde{C}_j^{n-1}\big) + d_{n-1}(C^n) - V^{n-1}\big(C_i^{n-1}\big) \geq U^n(C^n, \tilde{C}^n) + 1. \tag{1.5}$$

The double arrow "$C^n \Rightarrow \tilde{C}^n$" means $U^n(C^n, \tilde{C}^n) = V^n(C^n)$. Hence,

$$U^n(C^n, \tilde{C}^n) \geq V^n(C^n) + 1 \quad \text{if } C^n \not\Rightarrow \tilde{C}^n. \tag{1.6}$$

In the sequel, the symbol C^n or $\tilde{C}^n$ will always denote an element in S^n.

A notion of height is required to describe our results. Since U is an irreducible cost function on S, there exists a smallest number N such that $\{C^N\colon C^N \in S^N\}$ forms a single Nth-order cycle under U^N, that is,

$$|S^{N+1}| = 1 \quad \text{but} \quad |S^N| \geq 2.$$

For each $i \in S$, there exist uniquely $C^n \in S^n$, $0 \leq n \leq N+1$, such that $i = C^0 \in C^1 \in \cdots \in C^{N+1}$. Define the kth-order height of C^j as

$$h_k(C^j) = \begin{cases} \sum_{n=j}^{k} \big(d_n(C^{n+1}) - V^n(C^n)\big) & \text{if } j \leq k, \\ 0 & \text{if } j > k. \end{cases} \tag{1.7}$$

In particular, the (overall) height of a state $i \in S$ is given by

$$h(i) = h_N(i) = \sum_{n=0}^{N} \big(d_n(C^{n+1}) - V^n(C^n)\big). \tag{1.8}$$

Note that

$$h_n(C^n) = \begin{cases} 0 & \text{if } |C^{n+1}| = 1, \\ n - V^n(C^n) & \text{if } C^{n+1} \text{ is nontrivial}, \end{cases} \tag{1.9}$$

and represents the height of C^n relative to $R(C^{n+1})$. The constant $h_k(C^j)$ has a similar interpretation. Define the set $\underline{S}$ as

$$\underline{S} = \{i \in S\colon h(i) = 0\}. \tag{1.10}$$

We shall show that $\underline{S}$ is the candidate for the "global minimum set" and $h(i)$ is the "height" of state i relative to $\underline{S}$. In accordance with our intuition, it is not hard to show under the weak reversibility condition [Hajek (1988)] that $h(i) = u(i) - \min u$ if $U(i, j) = (u(j) - u(i))^+$. Hence $\underline{S}$ defined above is the same as the one given in (1.3).

Let

$$\delta = \max_{i \notin \underline{S}} \min_{j \in \underline{S}} V^k(C^k), \qquad \Gamma = \max_{i \in S} \min_{j \in \underline{S},\, j \neq i} V^k(C^k), \tag{1.11}$$

where k is the smallest integer such that $i = C^0 \in C^1 \in \cdots \in C^k \in C^{k+1} \in \cdots \in C^{N+1}$ and $j = \tilde{C}^0 \in \tilde{C}^1 \in \tilde{C}^2 \in \cdots \in \tilde{C}^k \in C^{k+1}$. Roughly speaking, δ is the minimum cost required for any nonglobal minimum state to reach $\underline{S}$, and Γ is the minimum cost for both above and any two states in $\underline{S}$ to reach each other. Thus $\delta = \Gamma$ if $|\underline{S}| = 1$. We remark that δ is related to the convergence rate to 0 of the second eigenvalue of $(q_{ij}(t))$ as $t \to \infty$. See Chiang and Chow (1987) and Ventcel (1972) for details.

In the following, we introduce three kinds of regularity condition on λ. In each case E is a nonnegative parameter.

$$(\text{A.1}; E) \qquad \int_0^{\infty} (\lambda(t))^E \, dt \left[\text{or } \sum_{t=0}^{\infty} (\lambda(t))^E\right] = \infty.$$

$$(\text{A.2}; E) \qquad \lambda'(t) \; [\text{or } \lambda(t+1) - \lambda(t)] = o\big((\lambda(t))^{E+1}\big).$$

$$(\text{A.3}; E) \qquad \frac{\lambda'(t)}{\lambda(t)} \left[\text{or } \frac{\lambda(t+1) - \lambda(t)}{\lambda(t)}\right] = \begin{cases} O\big((\lambda(t))^E\big) & \text{for } E > 0, \\ \text{unrestricted} & \text{for } E = 0. \end{cases}$$

The main result of this article can be stated as follows.

THEOREM 1.1. *Let* $\{X(t)\colon\ t \in R^+ \text{ or } I^+\}$ *be an (aperiodic) irreducible Markov process with state space S and transition rate* $(q_{ij}(t))$ *given in* (1.1). *Let* $h(i)$, $\underline{S}$, δ *and* Γ *be given as in* (1.1)–(1.11). *Assume* (A.1; Γ) *and* (A.2; δ) *hold. Then there exist positive constants* β_i, $i \in S$, *such that*

$$(1.12) \qquad \lim_{t \to \infty} P(X(t) = i)/(\lambda(t))^{h(i)} = \beta_i \quad \textit{for each } i \in S.$$

In particular, (1.3) *holds.*

Our method of proof is to consider the Kolmogorov forward equations associated with $X(t)$, that is, for each $i \in S$,

$$(1.13) \qquad F_i'(t) \; [\text{or } F_i(t+1)] = \sum_{j \in S} q_{ji}(t) F_j(t),$$

where $F_i(t) = P(X(t) = i)$ and $(q_{ij}(t))$ is given as in (1.1). It will be apparent from the proof that the limiting constants β_j can be obtained through solving systems of linear equations in $(p(i, j))$.

If one is only interested in (1.3), the assumptions in Theorem 1.1 can be weakened as follows.

THEOREM 1.2. (i) (1.3) *holds under the assumptions* (A.1; δ) *and* (A.3; δ).
(ii) *Assume* (A.1; δ) *and* (A.2; δ) *hold. Let* $b = \min_{i \notin \underline{S}} h(i)$. *Then*

$$P\big(X(t) \in \underline{S}\big) = 1 + O\big((\lambda(t))^b\big) \quad \textit{as } t \to \infty.$$

Theorems 1.1 and 1.2 will be proved in Section 3, after we obtain some preliminary results in Section 2. Two examples will be given in Section 4 to demonstrate the basic ideas behind the proofs. Finally, we make some remarks in Section 5.

2. Some preliminary estimates. As mentioned in the Introduction, we intend to analyze the system of differential (or difference) equations in (1.13).

Obviously,

$$\sum_{i \in S} F_i(t) = 1 \text{ and } F_i(t) = O(1) \quad \text{for } t \geq 0 \text{ and } i \in S.$$

The ultimate estimate (1.12) for each state cannot be obtained in one step, but through successive improvements of order $O(\lambda(t)), O(\lambda^2(t)), \ldots$ and so forth. It is clear that all terms of order λ^{k+1} or higher on the right-hand side of (1.13) can be neglected, if we only want to claim that $F_i(t) = O(\lambda^k(t))$. The following lemma plays an important role in our approach.

LEMMA 2.1. *Let $f(t)$ be a complex-valued function and α a complex number with* $\operatorname{Re} \alpha > 0$. *Suppose*

$$f'(t)\ [\text{or } f(t+1) - f(t)] = -\alpha\lambda^E(t) f(t) + \Delta\lambda^F(t), \tag{2.1}$$

where $\Delta = o(1)$ or $O(\lambda(t))$. Then

$$f(t) = \Delta\lambda^{F-E}(t) \quad \text{as } t \to \infty, \tag{2.2}$$

if

$$\begin{cases} \text{(i) } (\text{A.1}; E) \text{ and } (\text{A.2}; E) \text{ hold, or} \\ \text{(ii) } \Delta = o(1),\ F = E \text{ and } (\text{A.1}; E) \text{ holds.} \end{cases} \tag{2.3}$$

Lemma 2.1 can be used to yield some estimates for states in a cycle.

LEMMA 2.2. *Let $\tilde{S} \subseteq \tilde{T}$ be finite sets and $\tilde{U}$: $\tilde{T} \times \tilde{T} \to [0, \infty]$ be a cost function on $\tilde{T}$. Suppose that* (2.3) *holds and for each $i \in \tilde{S}$,*

$$f_i'(t)\ [\text{or } f_i(t+1)] = \sum_{j \in \tilde{S}} \tilde{q}_{ji}(t) f_j(t) + \Delta\lambda^E(t), \tag{2.4}$$

where E is a nonnegative integer, $\Delta = o(1)$ or $O(\lambda(t))$, $0 \leq f_i(t) \leq 1$ for $i \in \tilde{S}$, and $(\tilde{q}_{ij}(t))$ is of the type given in (1.1) *with U and $(p(i,j))$ there replaced by $\tilde{U}$ and $(\tilde{p}(i,j))$, respectively.*

(i) *If $\tilde{S}$ does not contain any nontrivial cycle of order $\leq E$, then for each $i \in \tilde{S}$ and $t \to \infty$,*

$$f_i'(t)\ [\text{or } f_i(t+1) - f_i(t)] = \Delta\lambda^E(t), \tag{2.5}$$

$$(\lambda(t))^{\tilde{V}(i)} f_i(t) = \Delta\lambda^E(t), \quad \text{where } \tilde{V}(i) = \min_{j \neq i} \tilde{U}(i,j). \tag{2.6}$$

(ii) *If $\tilde{S}$ is an Eth-order nontrivial cycle and contains no nontrivial cycle of less order, then there exist positive constants $\theta(i)$, $i \in \tilde{S}$, such that* (2.5) *holds and*

$$(\lambda(t))^{\tilde{V}(i)} f_i(t) = \theta(i)\lambda^E(t)\left(\sum_{j \in \tilde{S}} f_j(t) + \Delta\right). \tag{2.7}$$

Formula (2.7) means that within an error of order Δ, states in $\tilde{S}$ are comparable in probability and can be merged into a single one. If it is known in addition that $\sum_{j \in \tilde{S}} f_j = O(\lambda^K)$, then it is reasonable to expect that the estimation error in (2.7) can be improved.

LEMMA 2.3. *In addition to the assumptions in Lemma 2.2, we assume that* $\sum_{j \in \tilde{S}} f_j(t) = O(\lambda^K(t))$ *and for each* $i \in \tilde{S}$,

$$f_i'(t)\ [\text{or } f_i(t+1)] = \sum_{j \in \tilde{S}} \tilde{q}_{ji}(t) f_j(t) + \Delta\lambda^{E+K}(t). \tag{2.8}$$

If $\tilde{S}$ *is an Eth-order nontrivial cycle and contains no nontrivial cycle of less order, then for each* $i \in \tilde{S}$,

$$\begin{cases} f_i'(t)\ [\text{or } f_i(t+1) - f_i(t)] = \Delta\lambda^{E+K}(t), \\ (\lambda(t))^{\tilde{V}(i)} f_i(t) = \theta(i)\lambda^E(t)\left(\sum_{j \in \tilde{S}} f_j(t) + \Delta\lambda^K(t)\right), \end{cases} \tag{2.9}$$

where the constants $\theta(i)$*'s are the same as those in* (2.7).

As states in S are merged into states in S^n, we always encounter matrices of a special type called transition rate matrices.

DEFINITION 2.4. An $m \times m$ matrix $A = (a_{ij})$ is called a transition rate matrix if

$$\sum_{i=1}^{m} a_{ij} \le 0 \quad \text{for each } j \tag{i}$$

and

$$a_{ij} \ge 0 \quad \text{for all } i \ne j. \tag{ii}$$

The next lemma is essentially a consequence of the Perron–Frobenius theorem [Seneta (1973)] and we omit its proof.

LEMMA 2.5 [Chiang and Chow (1988)]. *Let* $A = (a_{ij})$ *be a transition rate matrix of order m. If* $A^{-1} = (b_{ij})$ *exists, then*:

(i) *All the eigenvalues of A have negative real parts.*

(ii) $b_{ii} \le (\min_i a_{ii})^{-1}$ *and* $b_{ii} \le b_{ij} \le 0$ *for all* i, j.

(iii) $b_{ij} < 0$ *if and only if i is reachable from j, that is, there exist* $i_0 = i$, $i_1, \dots, i_k = j$ *such that* $a_{i_n, i_{n+1}} > 0$ *for each* $0 \le n < k$.

On the other hand, if A is noninvertible but irreducible, then:

(iv) 0 *is an eigenvalue with multiplicity one and all other eigenvalues of A have negative real parts.*

(v) $\sum_{i=1}^{m} a_{ij} = 0$ *for each* $1 \le j \le m$.

(vi) *For any proper subset B of* $\{1, 2, \dots, m\}$, *the principal minor* $A_B = (a_{ij};\ i, j \in B)$ *is an invertible transition rate matrix.*

PROOF OF LEMMA 2.1. The continuous time case was treated in Chiang and Chow (1988) by using the l'Hospital rule. In the following we shall sketch a proof based on the integration-by-parts formula. It is the latter approach that can be adopted, via the Abel partial summation formula, for the discrete time case.

For the sake of convenience, we assume $\Delta = O(\lambda)$ and (2.3)(i) holds. The other cases can be treated similarly. Let $g(t) = \exp[\int_a^t \alpha\lambda^E(s)\,ds]$. A simple computation shows $(f(t)g(t))' = g(t)O(\lambda^{F+1}(t))$. Thus

$$f(t) = \left[f(a) + \int_a^t g(s)O(\lambda^{F+1}(s))\,ds\right]\Big/ g(t). \tag{2.10}$$

By using (2.3)(i),

$$\lim_{t\to\infty} \lambda^L(t)|g(t)| = \infty \tag{2.11}$$

holds for any $L \geq 0$. From (2.10) and (2.11) it suffices to show that

$$G(t) \leq c\lambda^{F+1-E}(t), \tag{2.12}$$

where $G(t) = (\operatorname{Re}\alpha)\int_a^t |g(s)|\lambda^{F+1}(s)\,ds/|g(t)|$. Write $\lambda^{F+1} = \lambda^E\lambda^{F+1-E}$ and use the integration-by-parts formula. It follows from (2.3)(i) and (2.11) that

$$\begin{aligned} G(t) &\leq \lambda^{F+1-E}(t) - (F+1-E)\int_a^t |g(s)|\lambda^{F+1-E}(s)(\lambda'/\lambda)\,ds/|g(t)| \\ &\leq \lambda^{F+1-E}(t) + \int_a^t |g(s)|\,\big|o(\lambda^{F+1}(s))\big|\,ds/|g(t)| \\ &= \lambda^{F+1-E}(t) + \varepsilon G(t) \quad \text{if } a \text{ is large enough.} \end{aligned}$$

Thus, $G(t) \leq \lambda^{F+1-E}(t)/(1-\varepsilon)$ for t large enough. This verifies (2.12) and then (2.2).

In the discrete time case it is easy to show from (2.1) that

$$|f(t+1)| \leq |f(a)|S(t,t) + c\sum_{k=a}^{t} \lambda^{F+1}(k)S(t,k),$$

where $S(t,k) = \prod_{j=k+1}^t |1 - \alpha\lambda^E(j)|$. Since $\operatorname{Re}\alpha > 0$ and $\lim \lambda(t) = 0$, $|1-\alpha\lambda^E(t)| \leq 1 - (\operatorname{Re}\alpha/2)\lambda^E(t)$ holds for t large, say $t \geq a$. Let $H(t) = \prod_{k=a}^t [1 - (\operatorname{Re}\alpha/2)\lambda^E(k)]^{-1}$. Then

$$|f(t+1)| \leq c\left[1 + \sum_{k=a}^{t} H(k)\lambda^{F+1}(k)\right]\Big/ H(t).$$

This is the analog of (2.10) for the discrete time case. It can be checked similarly that (2.11) holds with $g(t)$ replaced by $H(t)$. Since

$$H(k)\lambda^E(k)/H(t) = (2/(\operatorname{Re}\alpha))(H(k) - H(k-1))/H(t),$$

the Abel partial summation formula can be applied to do the "integration by parts." Equation (2.2) can be verified the same way as in the continuous time case. The details are omitted. □

PROOF OF LEMMA 2.2. We consider only the case $t \in R^+$ and $\Delta = O(\lambda)$. The other cases can be proved similarly.

The lemma is proved by using induction on E. For the sake of convenience we omit the symbol " $\sim$ " throughout the proof. We first introduce some notation. For any subset $A \subseteq S$ and any two different nonnegative integers m, n, let

$f_A = (f_i(t);\ i \in A)^{\mathrm{T}}$ is a column vector,
$f_A' = (f_i'(t);\ i \in A)^{\mathrm{T}}$,
$A_n = \{i \in S;\ V(i) = n\}$,
$Q_{m,n} = (1_{\{U(i,j)=m\}} p(i,j);\ i \in A_m,\ j \in A_n)^{\mathrm{T}}$ is an $|A_n| \times |A_m|$ matrix,
$Q_n = (q_n(i,j);\ i, j \in A_n)^{\mathrm{T}}$ is an $|A_n| \times |A_n|$ transition rate matrix, where

$$q_n(i,j) = \begin{cases} 1_{\{U(i,j)=n\}} p(i,j) & \text{if } j \neq i, \\ -\displaystyle\sum_{k \in S,\, U(i,k)=n} p(i,k) & \text{if } j = i, \end{cases}$$

$Q_{n,A} = (q_n(i,j);\ i, j \in A_n \cap A)^{\mathrm{T}}$.

Step 1. $E = 0$. It is enough to show that (2.5)–(2.7) hold for states in A_0. Once it is done, (2.5)–(2.6) can be checked easily for any other state. This is because $V(i) \geq 1$ for $i \notin A_0$. After collecting all terms of order $O(\lambda)$ or higher, the differential equations in (2.4) for $i \in A_0$ take the matrix form

$$f_{A_0}' = Q_0 f_{A_0} + O(\lambda). \tag{2.13}$$

In case (i) Q_0 is an invertible transition rate matrix. By Lemma 2.5 all its eigenvalues have negative real parts. Let α be an eigenvalue of Q_0 with u_1 as its corresponding left eigenvector. Multiplying (2.13) from the left by u_1,

$$(u_1 f_{A_0})' = \alpha(u_1 f_{A_0}) + O(\lambda). \tag{2.14}$$

By Lemma 2.1 and (2.14), we have

$$u_1 f_{A_0} = O(\lambda), \qquad u_1 f_{A_0}' = O(\lambda). \tag{2.15}$$

If $u_2 Q_0 = \alpha u_2 + u_1$ then, with the help of (2.15),

$$(u_2 f_{A_0})' = \alpha(u_2 f_{A_0}) + (u_1 f_{A_0}) + O(\lambda) = \alpha(u_2 f_{A_0}) + O(\lambda).$$

By the same reason (2.15) holds with u_1 replaced by u_2. Repeating the same procedure and using Jordan's decomposition theorem, we can get a basis $\{u_1, u_2, \dots, u_{|A_0|}\}$ of $\mathbb{C}^{|A_0|}$ such that (2.15) holds for each u_i. Thus

$$f_{A_0} = O(\lambda), \qquad f_{A_0}' = O(\lambda). \tag{2.16}$$

This proves (2.5) and (2.6) for case (i) and $E = 0$.

In case (ii) Q_0 is noninvertible but irreducible. It follows from Lemma 2.5 that all eigenvalues of Q_{A_0} have negative real parts except 0, which is an eigenvalue with multiplicity 1 and having $e = (1, 1, \dots, 1)$ as its left eigenvector. By the same technique (2.14) becomes $ef_{A_0}' = O(\lambda)$ and we can get generalized eigenvectors $\{u_2, u_3, \dots, u_{|A_0|}\}$ such that (2.15) holds for each u_i, $2 \leq i \leq |A_0|$. Since $\{e, u_2, \dots, u_{|A_0|}\}$ forms a basis of $\mathbb{C}^{|A_0|}$, (2.5) is proved. For the remaining part, we fix a state $i_0 \in A_0$ and let $B = A_0 \setminus \{i_0\}$. The following linear equation can be obtained from (2.13) and (2.5):

$$Q_{0,B} f_B = -\left(p(i_0, i);\, i \in B\right)^{\mathrm{T}} f_{i_0} + O(\lambda).$$

By Lemma 2.5(vi) and (iii) $Q_{0,B}$ has an inverse and all elements of $(-Q_{0,B})^{-1}$ are nonnegative and the diagonal ones are positive. Since

$$f_B = (-Q_{0,B})^{-1}\left(p(i_0, i);\, i \in B\right)^{\mathrm{T}} f_{i_0} + O(\lambda),$$

it is clear that there exist constants $\theta(i, i_0) \geq 0$ such that $f_i = \theta(i_0, i) f_{i_0} + O(\lambda)$ and by Lemma 2.5(iii), $\theta(i_0, i) > 0$ if $p(i_0, i) > 0$. Since S is irreducible and i_0 is arbitrary, (2.7) follows easily. This completes the proof for $E = 0$.

Step 2. Case (i) and $E \geq 1$. Without loss of generality we may assume S is connected. Certainly $O(\lambda^{E+1})$ is $O(\lambda^E)$. By the induction hypothesis,

$$f_i' = O(\lambda^E) \quad \text{and} \quad \lambda^{V(i)} f_i = O(\lambda^E). \tag{2.17}$$

Since $\lambda^{V(i)} f_i = O(\lambda^{E+1})$ for $V(i) \geq E + 1$, it follows from (1.6), (2.4) and (2.17) that

$$f_{A_i}' = \lambda^i Q_i f_{A_i} + \sum_{k=0,\, k \neq i}^{E} \lambda^k Q_{ki} f_{A_k} + O(\lambda^{E+1}), \qquad i \geq 0. \tag{2.18}$$

We claim that for each $0 \leq i \leq E$, there exist an invertible transition rate matrix Q_i^* and nonnegative matrices Q_{ki}^*, $k > i$, such that

$$f_{A_i}' = \lambda^i Q_i^* f_{A_i} + \sum_{k=i+1}^{E} \lambda^k Q_{ki}^* f_{A_k} + O(\lambda^{E+1}) = O(\lambda^{E+1}). \tag{2.19}$$

Suppose temporarily that (2.19) holds. Then

$$\lambda^i f_{A_i} = O(\lambda^{E+1}) \quad \text{for each } 0 \leq i \leq E. \tag{2.20}$$

Letting $i = E$ in (2.19), we obtain (2.20) for $i = E$. Applying this result to (2.19) with $i = E - 1$, we obtain (2.20) for $i = E - 1$. Thus (2.20) can be proved by repeating the same argument.

Equation (2.6) follows immediately from (2.20). Note that (2.6) holds automatically for any state i with $V(i) \geq E + 1$. As to (2.5), that it holds for any state i with $V(i) \geq E + 1$ follows from (2.18) and (2.6). The remaining cases are already shown in (2.19). Thus Step 2 is finished, except that we have to check (2.19).

Equation (2.19) can be shown by using induction on i. First observe that each Q_i is an invertible transition rate matrix. This is because S does not contain any nontrivial cycle of order $\leq E$. Thus Lemma 2.5 is applicable. We start from $i = 0$. Let α, u_1 be the same as in (2.14). Define

$$g = u_1 f_{A_0} + h, \quad \text{where } h = \alpha^{-1} u_1 \left(\sum_{k=1}^{E} \lambda^k Q_{k0} f_{A_k} \right). \tag{2.21}$$

Under the assumption (A.2; E), $\lambda'/\lambda = O(\lambda^E) = O(\lambda)$. By (2.17),

$$h' = O(\lambda^{E+1}). \tag{2.22}$$

Multiplying (2.18) from the left by u_1 and using (2.22),

$$g' = \alpha g + O(\lambda^{E+1}).$$

As in (2.15) we obtain that both $g, g' = O(\lambda^{E+1})$. Combining with (2.21) and (2.22),

$$(2.23) \qquad (u_1 f_{A_0})' = u_1 \left[Q_0 f_{A_0} + \sum_{k=1}^{E} \lambda^k Q_{k0} f_{A_k} \right] + O(\lambda^{E+1}) = O(\lambda^{E+1}).$$

By the same technique used in the proof of (2.16), we can obtain a basis $\{u_1, u_2, \dots, u_{|A_0|}\}$ of $\mathbb{C}^{|A_0|}$ such that (2.23) holds for each u_i. This verifies (2.19) for $i = 0$ with $Q_0^* = Q_0$ and $Q_{k0}^* = Q_{k0}$. Suppose that (2.19) holds for all $i = 0, 1, \dots, j-1$. Then for each $i < j$,

$$(2.24) \qquad \lambda^i f_{A_i} = \sum_{k=i+1}^{E} \lambda^k (-Q_i^*)^{-1} Q_{ki}^* f_{A_k} + O(\lambda^{E+1}).$$

That is, $\lambda^i f_{A_i}$ can be expressed, within an error of $O(\lambda^{E+1})$, as a sum of higher-order ones. Applying (2.24) to (2.18) and making some rearrangements, we have

$$f'_{A_j} = \lambda^j Q_j^* f_{A_j} + \sum_{k=j+1}^{E} \lambda^k Q_{kj}^* f_{A_k} + O(\lambda^{E+1})$$

for some matrices Q_j^* and Q_{kj}^*. That each Q_{kj}^* is nonnegative is clear from applying Lemma 2.5(ii) to Q_i^*, $0 \le i < j$. The following facts about Q_j^* can be shown:

(i) Q_j^* is a transition rate matrix.

(ii) For any two different states k, r in A_j we have $(Q_j^*)_{k,r} > 0$ iff $k \ge r$ at level j, that is, state r can be reached from state k through states in $\bigcup_{k=0}^{j} A_k$.

(iii) Q_j^* is invertible unless $\{i\colon U(r,i) = j \text{ for some } r \in A_j\} \cap (\bigcup_{k>j} A_k) = \varnothing$. In that case all column sums of Q_j^* are 0.

The details can be found in the proof of Lemma 2.5 in Chiang and Chow (1988) and are omitted.

Since by assumption S does not contain any nontrivial cycle of order $\le E$, Q_j^* is invertible. By the same argument we can obtain $f'_{A_j} = O(\lambda^{E+1})$. This verifies (2.19) by induction.

Step 3. Case (ii) and $E \ge 1$. We claim that (2.17) holds for each $i \in S$. Since $S \setminus A_E$ does not contain any nontrivial cycle of order $\le E - 1$ and $\lambda^{U(j,k)} f_j = O(\lambda^E)$ for each $j \in A_E$, Lemma 2.2(i) can be applied to $S \setminus A_E$. Thus (2.17) holds for each $i \in S \setminus A_E$. Applying this result to the equation for f_{A_E} in (2.18), we obtain immediately that $f'_{A_E} = O(\lambda^E)$. The claim is verified.

The same argument in Step 2 can be repeated to show that (2.19) holds, except that now Q_E^* in

$$f'_{A_E} = \lambda^E Q_E^* f_{A_E} + O(\lambda^{E+1})$$

is not invertible but irreducible. As in the case $E = 0$, we can find a basis $\{e = (1, 1, \dots, 1), u_2, \dots, u_{|A_E|}\}$, which consists of generalized left-eigenvectors of Q_E^*, such that

$$\text{all } ef'_{A_E},\ u_i f'_{A_E} \text{ and } u_i f_{A_E} \text{ are } O(\lambda^{E+1}),$$

and then there are positive constants $\theta(i)$, $i \in A_E$, such that for each $i \in A_E$,

$$\lambda^{V(i)} f_i = \theta(i)\lambda^E\big(ef_{A_E} + O(\lambda)\big). \tag{2.25}$$

Applying (2.25) to (2.24) with $i = E - 1$, we can obtain positive constants $\theta(i)$, $i \in A_{E-1}$, such that (2.25) holds for each state i in A_{E-1}. Repeat the same argument to obtain (2.25) successively for states in $A_{E-2}, \dots, A_1$ and A_0. This proves (2.7) if one notes that $\sum_{j \in S} f_j = ef_{A_E} + O(\lambda)$ by (2.17). □

Proof of Lemma 2.3. As in the proof of Lemma 2.2, we consider only the case $t \in R^+$ and $\Delta = O(\lambda)$. We assume for convenience that K is an integer.

The lemma is proved by using induction on K. The case $K = 0$ is established in Lemma 2.2(ii). Suppose that the lemma holds for $K - 1$. Since $O(\lambda^{K+1}) = O(\lambda^K)$, (2.9) holds for $K - 1$ by induction hypothesis. Using the assumption $\sum_{i \in S} f_i = O(\lambda^K)$, we then obtain

$$f_i',\ \lambda^{V(i)} f_i = O(\lambda^{E+K}) \quad \text{for each } i \in S. \tag{2.26}$$

Observe that $U(i, j) \ge V(i) + 1$ if $i \nRightarrow j$, that is, $U(i, j) \ne V(i)$. It follows from (2.26) and (2.8) that (2.18) holds with $O(\lambda^{E+1})$ replaced by $O(\lambda^{E+K+1})$. Now repeat the same procedure used in the proof of Lemma 2.2(ii) and (2.9) for K can be obtained. □

3. Proof of Theorems 1.1 and 1.2. Recall that C^n or $\tilde{C}^n$ represents a typical state in S^n. For each $C^0 \in S^0 = S$, there is a unique chain $C^0 \in C^1 \in C^2 \in \cdots \in C^{N+1}$, where $C^j \in S^j$ for $0 \le j \le N + 1$. Let

$$0 \le m_1 < m_2 < \cdots < m_L \le N$$

be those indices $j \le N$ such that $|C^{j+1}| > 1$. Certainly L and all m_j exist by the irreducibility assumption and depend on C^0. Let $\mathrm{OS}(C^n) = \{i \in S$: there exist $C^j \in S^j$ such that $i = C^0 \in C^1 \in \cdots \in C^n\}$ be the offsprings of C^n. Define $F_{C^n}(t) = P(X(t) \in \mathrm{OS}(C^n))$. It is clear that

$$F_{C^n} = \sum_{i \in \mathrm{OS}(C^n)} F_i = \sum_{C_i^{n-1} \in C^n} F_{C_i^{n-1}}.$$

We first prove a lemma.

Lemma 3.1. *Let $n \ge 1$ be fixed. The following results hold as long as Lemmas 2.2 and 2.3 are applicable to S.*

(i) *Either $m_1 \ge n$ or there is a largest $m_r \le n - 1$. In the former case,*

$$\lambda^{V^0(C^0)} F_{C^0} = \lambda^{V^1(C^1)} F_{C^1} = \cdots = \lambda^{V^{n-1}(C^{n-1})} F_{C^{n-1}} = O(\lambda^n). \tag{3.1}$$

In the latter case, there exist positive constants $\theta(C^{m_1}), \theta(C^{m_2}), \dots, \theta(C^{m_r})$ such *that*

$$(3.2)\qquad \lambda^{V^j(C^j)}F_{C^j} = \begin{cases} \lambda^{V^{m_r+1}(C^{m_r+1})}F_{C^{m_r+1}} = O(\lambda^n) \\ \qquad \text{if } m_r < j \le n-1, \\ \lambda^{V^{m_k}(C^{m_k})}F_{C^{m_k}} = \theta(C^{m_k})\lambda^{m_k}\left(F_{C^{m_k+1}} + \Delta\lambda^{H_n(C^{m_k+1})}\right) \\ \qquad \text{if } m_{k-1} < j \le m_k \text{ and } 1 \le k \le r, \end{cases}$$

where, in general, $H_n(C^j) = h_{n-1}(C^j) + (n - V^n(C^n))^+$. *In both cases,*

$$(3.3)\qquad F_{C^j} = O(\lambda^{H_n(C^j)}) \quad \text{for } j \le n-1,$$

$$(3.4)\qquad \lambda^{U(i,j)}F_i = \Delta\lambda^{n-1} \quad \text{for each } (i,j) \in (\mathrm{OS}(C^n), S \setminus \mathrm{OS}(C^n)),$$

$$(3.5)\qquad \lambda^{V^n(C^n)}F_{C^n} = \Delta\lambda^{n-1}.$$

(ii) *An irreducible* $|S^n| \times |S^n|$ *matrix* $P_n = (p(C^n, \tilde{C}^n))$ *with nonnegative off-diagonal elements can be constructed recursively from the matrix* $(p(i,j))$ *given in* (1.1), *such that for each* $C^n \in S^n$,

$$(3.6)\qquad F'_{C^n}(t)\ [\text{or } F_{C^n}(t+1)] = \sum_{\tilde{C}^n} q_{\tilde{C}^n, C^n}(t)F_{\tilde{C}^n}(t) + \Delta\lambda^n(t),$$

where

$$q_{C^n,\tilde{C}^n}(t) = \begin{cases} p(C^n,\tilde{C}^n)\lambda^{U^n(C^n,\tilde{C}^n)}(t) & \text{if } \tilde{C}^n \ne C^n, \\ 1_{\{\text{discrete time}\}} - \sum_{\tilde{C}^n \ne C^n} q_{C^n,\tilde{C}^n}(t) & \text{if } \tilde{C}^n = C^n. \end{cases}$$

PROOF. We consider only the case $t \in R^+$ and $\Delta = O(\lambda)$. The other cases can be treated similarly.

We start by noting that $F_{C^j} = F_{C^{j+1}}$, $V^j(C^j) = V^{j+1}(C^{j+1})$ if $|C^{j+1}| = 1$. Moreover, by (1.7),

$$h_{n-1}(C^j) = \begin{cases} 0 & \text{if } m_1 \ge n \text{ or } m_r < j, \\ \sum_{j \le m_k \le n-1} (m_k - V^{m_k}(C^{m_k})) & \text{if } j \le m_r. \end{cases}$$

The lemma is proved by using induction on n. Take $n = 1$. Write $S^0 = (\bigcup_{k=1}^M C_k^1) \cup (\bigcup_{l=1}^{\tilde{M}} \tilde{C}_l^1)$, where all C_k^1's and $\tilde{C}_l^1$'s are zeroth-order cycles except that $|C_k^1| > 1$ and $|\tilde{C}_l^1| = 1$. Since $V^1(C_k^1) \ge 1$ it is clear that (3.3)–(3.5) hold for $n = 1$ and $C^0 \in C_k^1$. Thus in the differential equation (1.1) for $i = C^0 \in \tilde{C}_l^1$, the contributions from states in $\bigcup_1^M C_k^1$ are all $O(\lambda)$. Applying Lemma 2.2(i) with $\tilde{S} = S \setminus \bigcup_1^M C_k^1$ and $E = 0$, we obtain

$$(3.7)\qquad \lambda^{U(i,j)}F_i \le \lambda^{V^0(C^0)}F_{C^0} = O(\lambda).$$

This proves (3.1) and also completes the proof of (3.3)–(3.5) for $n = 1$. Using (3.7) and then applying Lemma 2.2(ii) with $\tilde{S} = C_k^1$ and $E = 0$, we obtain positive constants $\theta(i)$, $i \in C_k^1$, such that

$$F_i = \theta(i)\big(F_{C_k^1} + O(\lambda)\big), \tag{3.8}$$

which is (3.2) for $n = 1$.

It remains to show (3.6). As the notation suggests, C_k^1 will be treated as a single state in S^1. This is achieved by using (3.8) to merge, within an error of order $O(\lambda)$, together the states in C_k^1. For example, the contributions from C_k^1 to $\tilde{C}_l^1 = \{j\}$ can be written as

$$\begin{aligned}\sum_{i \in C_k^1} p(i,j)\lambda^{U(i,j)} F_i &= \left[\sum_{i \in C_k^1} \theta(i) p(i,j) \lambda^{U(i,j)}\right]\big(F_{C_k^1} + O(\lambda)\big)\\ &= \left\{\left[\sum_{(i,j)\in \mathrm{Ix}(C_k^1, \tilde{C}_l^1)} \theta(i)p(i,j)\right]\lambda^{U^1(C_k^1,\tilde{C}_l^1)}(1 + O(\lambda))\right\}\\ &\quad \times \big(F_{C_k^1} + O(\lambda)\big),\end{aligned} \tag{3.9}$$

because by (1.5), the uncounted terms in the summation are at least one order higher. Similarly, the contributions from C_k^1 to a different nontrivial C_l^1 satisfy

$$\begin{aligned}&\sum_{i \in C_k^1,\, j \in C_l^1} p(i,j)\lambda^{U(i,j)} F_i\\ &\quad = \left\{\left[\sum_{(i,j)\in \mathrm{Ix}(C_k^1, C_l^1)} \theta(i)p(i,j)\right]\lambda^{U^1(C_k^1, C_l^1)}(1 + O(\lambda))\right\}\big(F_{C_k^1} + O(\lambda)\big)\end{aligned}$$

and the contributions from $\tilde{C}_l^1 = \{i\}$ to a different zeroth-order cycle C^1 satisfy

$$\sum_{j \in C^1} p(i,j)\lambda^{U(i,j)} F_i = \left\{\left[\sum_{(i,j)\in \mathrm{Ix}(\tilde{C}_l^1, C^1)} p(i,j)\right]\lambda^{U^1(\tilde{C}_l^1, C^1)}(1 + O(\lambda))\right\} F_{\tilde{C}_l^1}.$$

Since $U^1(C_k^1, \cdot\,) \geq 1$ and (3.5) holds, (3.6) follows by letting

$$p(C^1, \tilde{C}^1) = \sum_{(i,j)\in \mathrm{Ix}(C^1, \tilde{C}^1)} \theta(i) p(i,j).$$

Here and throughout the proof, we take the convention that $\theta(C^j) = 1$ if $C^j \in C^{j+1}$ and $|C^{j+1}| = 1$. Note that the diagonal elements of matrix P_1, as those of $(p(i,j))$, are unimportant and can be arbitrarily defined. This remark also applies to all subsequent P_k, $k \geq 2$. This proves the lemma for $n = 1$.

Suppose the lemma holds for n. We now show that it holds for $n + 1$. As in case $n = 1$, we decompose $S^n = (\bigcup_1^M C_k^{n+1}) \cup (\bigcup_1^{\tilde{M}} \tilde{C}_l^{n+1})$ into a union of nth-order cycles with $|C_k^{n+1}| > 1$ and $|\tilde{C}_l^{n+1}| = 1$. By (3.5) for n and (1.6),

$$\lambda^{U^n(C^n, \tilde{C}^n)} F_{C^n} = O(\lambda^{n+1}) \quad \text{if } (C^n, \tilde{C}^n) \in \big(C_k^{n+1}, S^n \setminus C_k^{n+1}\big). \tag{3.10}$$

Then (3.5) for $n + 1$ and $C^{n+1} = C_k^{n+1}$ can be proved by using (1.4). (3.4) for $n + 1$ and $C^{n+1} = C_k^{n+1}$ follows from the same argument by using (3.4) for n. (3.3) for $n + 1$ and $C^j \in \cdots \in C^{n+1} = C_k^{n+1}$ automatically holds, because (3.5) holds for n and

$$H_{n+1}(C^j) = H_n(C^j) \quad \text{for } j < n. \tag{3.11}$$

Consider the differential equations in (3.6) for $C^n \in S^n \setminus \bigcup_1^M C_k^{n+1}$. Using (3.10) and then applying Lemma 2.2(i) with $\tilde{S} = S^n \setminus \bigcup_1^M C_k^{n+1}$ and $E = n$, we obtain

$$\lambda^{V^{n+1}(\tilde{C}_l^{n+1})} F_{\tilde{C}_l^{n+1}} = \lambda^{V^n(C^n)} F_{C^n} = O(\lambda^{n+1}) \quad \text{if } \{C^n\} = \tilde{C}_l^{n+1}, \tag{3.12}$$

which completes the proof of (3.5) for $n + 1$. Note that (3.12) improves (3.5) for n. In view of (3.1) or (3.2) for n, it is not hard to check (3.3) and (3.4) for $n + 1$ and $C^j \in \cdots \in C^{n+1} = \tilde{C}_l^{n+1}$. Note that in this case,

$$H_{n+1}(C^j) = 1 + H_n(C^j) \quad \text{for } j < n, \tag{3.13}$$

iff $V^n(C^n) \le n$. In fact the same argument also shows that if $|C^{n+1}| = 1$, $V^n(C^n) \le n$ and $m_r \le n$ exists, then for $k \le r$,

$$\lambda^{U^{m_k}(\tilde{C}^{m_k}, C^{m_k})} F_{\tilde{C}^{m_k}} = \begin{cases} O(\lambda^{m_k+1+H_n(C^{m_k+1})}) & \text{if } \tilde{C}^{m_k} \in C^{m_k+1} \text{ and } \tilde{C}^{m_k} \Rightarrow C^{m_k}, \\ O(\lambda^{m_k+2+H_n(C^{m_k+1})}) & \text{otherwise.} \end{cases}$$

Or, quite equivalently, for $j = C^0 \in C^1 \in \cdots \in C^{m_k+1}$ and $i = \tilde{C}^0 \in \tilde{C}^1 \in \cdots \in \tilde{C}^{m_k}$,

$$\lambda^{U(i,j)} F_i = \begin{cases} O(\lambda^{m_k+1+H_n(C^{m_k+1})}) & \text{if } \tilde{C}^{m_k} \in C^{m_k+1} \text{ and} \\ & (\tilde{C}^{l-1}, C^{l-1}) \in \mathrm{Ix}(\tilde{C}^l, C^l) \\ & \text{for } 1 \le l \le m_k, \\ O(\lambda^{m_k+2+H_n(C^{m_k+1})}) & \text{otherwise.} \end{cases} \tag{3.14}$$

In summary, we have shown so far that (3.3)–(3.5) hold for $n + 1$.

Next, consider the differential equations in (3.6) for $C^n \in C_k^{n+1}$. (3.10) and (3.12) imply that the contributions from any $\tilde{C}^n \notin C_k^{n+1}$ are $O(\lambda^{n+1})$. Applying Lemma 2.2(ii) with $\tilde{S} = C_k^{n+1}$, $\tilde{U} = U^n$ and $E = n$, we obtain positive constants $\theta(C^n)$, $C^n \in C_k^{n+1}$, such that

$$\lambda^{V^n(C^n)} F_{C^n} = \theta(C^n) \lambda^n \big(F_{C_k^{n+1}} + O(\lambda) \big). \tag{3.15}$$

We now prove (3.1) and (3.2) for $n + 1$. By (3.12) and (3.15) it suffices to consider $j < n$. If $m_1 \ge n + 1$, then (3.1) follows from (3.12). So we may assume there is a largest $m_r \le n$ and want to show (3.2).

Case 1. $m_r = n$. That is, $C^n \in C_k^{n+1}$ for some k. Since $d_n(C_k^{n+1}) = n$, it is easy to check (3.11) and we are done.

Case 2. $m_r < n$. If $V^{m_r+1}(C^{m_r+1}) = V^n(C^n) \ge n + 1$, then we are done by (3.11). Suppose $V^n(C^n) \le n$. Then (3.13) holds. By (3.3), which is just proved,

$$F_{C^{m_k+1}} = O(\lambda^{H_{n+1}(C^{m_k+1})}) = O(\lambda^{1+H_n(C^{m_k+1})}), \qquad 1 \le k \le r.$$

Note that the estimate above is the same as the error estimate in (3.2) for n. It is reasonable to expect that the error estimate can be improved by order $O(\lambda)$.

Since $F'_{C^{m_k}} = \sum_{i \in \mathrm{OS}(C^{m_k})} F'_i$, it follows from (3.14) that for each $C^{m_k} \in C^{m_k+1}$,

$$F'_{C^{m_k}} = \sum_{\tilde{C}^{m_k} \in C^{m_k+1}} q_{\tilde{C}^{m_k}, C^{m_k}}(t) F_{\tilde{C}^{m_k}} + O\big(\lambda^{m_k+2+H_n(C^{m_k+1})}\big).$$

Applying Lemma 2.3 with $\tilde{S} = C^{m_k+1}$, $\tilde{U} = U^{m_k}$, $E = m_k$ and $K = H_{n+1}(C^{m_k+1})$, we obtain

$$\lambda^{V^{m_k}(C^{m_k})} F_{C^{m_k}} = \theta(C^{m_k}) \lambda^{m_k} \Big(F_{C^{m_k+1}} + O\big(\lambda^{1+H_{n+1}(C^{m_k+1})}\big)\Big).$$

This proves (3.2) for $n+1$.

Equations (3.4) and (3.5) for $n+1$ imply that any contribution from a different $\tilde{C}^{n+1}$ to C^{n+1} is $O(\lambda^{n+1})$, and thus by (1.6), is $O(\lambda^{n+2})$ if $\tilde{C}^{n+1} \nRightarrow C^{n+1}$. As did in (3.9), we can obtain (3.6) for $n+1$ by neglecting all terms of order $O(\lambda^{n+2})$. Note that we have

$$p(C^{n+1}, \tilde{C}^{n+1}) = \sum_{(C^n, \tilde{C}^n) \in \mathrm{Ix}(C^{n+1}, \tilde{C}^{n+1})} \theta(C^n) p(C^n, \tilde{C}^n)$$

with the convention that $\theta(C^n) = 1$ if $C^n \in C^{n+1}$ and $|C^{n+1}| = 1$.

This completes the proof of the lemma. □

Define $R^*(C^n) = \{i \in \mathrm{OS}(C^n)\colon C^j \in R(C^{j+1})$ for $0 \le j < n$, where $i = C^0 \in C^1 \in \cdots \in C^n\}$ to be the "root" of C^n in S. It is clear that if Lemma 3.1 holds for n, then

$$\text{(3.16)} \qquad \begin{cases} F_i \approx F_{C^n} & \text{for } i \in R^*(C^n), \\ F_i = o(1) & \text{otherwise.} \end{cases}$$

We now proceed to the proof of Theorems 1.1 and 1.2. In the following we assume $\delta > 0$. The case $\delta = 0$ can be treated more easily in the same manner. Suppose (A.1; δ) and (A.3; δ) hold. Since (A.3; δ) is stronger than (A.2; $\delta - 1$), Lemma 3.1 is applicable for $n \le \delta$ and $\Delta = O(\lambda)$. Therefore, for $C^\delta \in S^\delta$,

$$\text{(3.17)} \qquad F'_{C^\delta}(t)\ [\text{or } F_{C^\delta}(t+1)] = \sum_{\tilde{C}^\delta} q_{\tilde{C}^\delta, C^\delta}(t) F_{\tilde{C}^\delta}(t) + O\big(\lambda^{\delta+1}(t)\big).$$

We want to show that $F_i = o(1)$ if $i \notin \underline{S}$. Suppose $i \in \mathrm{OS}(C^\delta)$. If $i \notin R^*(C^\delta)$ then, since (3.16) holds for $n = \delta$, $F_i = o(1)$. Otherwise, replacing $O(\lambda^{\delta+1})$ in (3.17) by $o(\lambda^\delta)$ and applying Lemma 2.2 with $\Delta = o(1)$, we obtain $F_i \le F_{C^\delta} = o(1)$. This proves Theorem 1.2(i).

If (A.2; δ) is assumed instead of (A.3; δ), we want to show

$$\text{(3.18)} \qquad F_i = O(\lambda^{h(i)}) \quad \text{for } i \notin \underline{S},$$

which implies Theorem 1.2(ii). By (1.11), if a nontrivial cycle exists in the chain $\tilde{C}^\delta \in \tilde{C}^{\delta+1} \in \cdots \in \tilde{C}^{N+1}$, say $|\tilde{C}^n| > 1$ for some $n > \delta$, and $V^{n-1}(\tilde{C}^{n-1}) > \delta$, then $\tilde{C}^{n-1} \in R(\tilde{C}^n)$. Moreover, $R^*(\tilde{C}^{n-1}) \subseteq R^*(\tilde{C}^n) \subseteq \cdots \subseteq \underline{S}$. Thus $h_N(C^{n-1}) = 0$ for $C^{n-1} \in R(\tilde{C}^n)$. We claim that under (A.1; δ) and (A.2; δ)

$$\text{(3.19)} \qquad \text{(3.3)–(3.5) hold for } \delta + 1 \le n \le N \text{ and } \Delta = O(\lambda).$$

Once it is proved, (3.18) follows by letting $j = 0$ and $n = N$ in (3.3) and using the equality $H_N(C^0) = h_N(C^0) = h(C^0)$. It is clear from (3.17) that Lemma 3.1 now holds for $n = \delta + 1$ and $\Delta = O(\lambda)$. Thus (3.19) is true for $n = \delta + 1$ and

$$F'_{C^{\delta+1}}(t)\ [\text{or } F_{C^{\delta+1}}(t+1)] = \sum_{\tilde{C}^{\delta+1}} q_{\tilde{C}^{\delta+1}, C^{\delta+1}}(t) F_{\tilde{C}^{\delta+1}}(t) + O(\lambda^{\delta+2}(t)).$$

Write $S^{\delta+1} = (\cup_1^M C_k^{\delta+2}) \cup (\cup_1^{\tilde{M}} \tilde{C}_l^{\delta+2})$ with $|C_k^{\delta+2}| > 1$ and $|\tilde{C}_l^{\delta+2}| = 1$. By (3.19) for $n = \delta + 1$ and (1.6), (3.4) and (3.5) hold for $C_k^{\delta+2}$ and $\Delta = O(\lambda)$. Applying Lemma 2.2(i) with $\tilde{S} = \cup \tilde{C}_l^{\delta+2}$, $\tilde{U} = U^{\delta+1}$, $E = \delta + 1$ and $\Delta = O(\lambda)$, we obtain (3.4) and (3.5) for $\tilde{C}_l^{\delta+2}$. Then (3.3) for $n = \delta + 2$ can be proved by using the results just obtained and repeating the proof of Lemma 3.1(i) with $n = \delta + 2$ and S replaced by $S \setminus \cup_{k=1}^M R^*(R(C_k^{\delta+2}))$. This verifies (3.19) for $n = \delta + 2$. Note that in the argument above, we cannot apply Lemma 3.1 to S to get (3.2) for all $j \le \delta + 2$. This is no harm. Because $R^*(C^{\delta+1}) \subseteq \underline{S}$ for any $C^{\delta+1} \in R(C_k^{\delta+2})$ and $F_i = O(1)$ is the desired estimate for $i \in \underline{S}$. In a similar way, (3.19) can be verified successively for $n = \delta + 3, \delta + 4, \dots$ and N. We omit the details. This completes the proof of Theorem 1.2(ii).

Suppose (A.1; Γ) and (A.2; δ) hold. Lemma 3.1 is first applicable for $n = \Gamma + 1$ and $\Delta = O(\lambda)$, and then for $n = N + 1$ and $\Delta = o(1)$. Note that $\Gamma \le N$ by definition. Let $i = C^0 \in C^1 \in \cdots \in C^{N+1}$ and $m_1, m_2, \dots, m_L$ are those indices j such that $|C^{j+1}| > 1$. It follows from (3.2) that

$$F_i = \left(\prod_{k=1}^{L} \theta(C^{m_k}) \right) \lambda^{h(i)}(1 + o(1)).$$

This proves Theorem 1.1.

4. Examples.

EXAMPLE 1. Let $S = \{1, 2, 3, 4\}$. In Figure 1 only those $U(i, j)$'s with $U(i, j) = V(i)$ are shown. All the other $U(i, j)$'s are unimportant and can be neglected. This claim can be verified via Lemma 2.2. In fact, by using Lemma 2.2 six times, we can show that for each $i \in S$,

$$F_i'(t) = O(\lambda^7(t)) \quad \text{and} \quad \lambda^{V(i)}(t) F_i(t) = O(\lambda^7(t)), \tag{4.1}$$

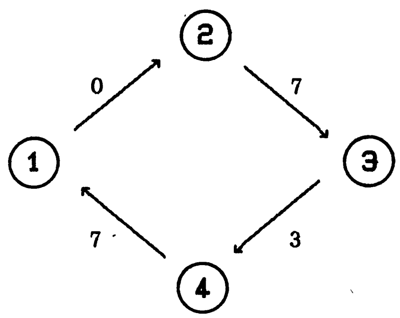

FIG. 1.

where $F_i(t) = P(X(t) = i)$. Thus

$$F_1(t) = O(\lambda^7(t)),\ F_3(t) = O(\lambda^4(t)) \text{ and } F_2(t), F_4(t) = O(1). \tag{4.2}$$

Note that in this example $\delta = 3$, $\Gamma = 7$, $h(1) = 7$, $h(3) = 4$ and $h(2) = h(4) = 0$.

Once we have the desired order estimate (4.2), Theorem 1.1 can be proved by using Lemma 2.3. We remark that (4.2) implies (4.1) by improving successively the order estimate for $F_i'(t)$ from $O(1), O(\lambda^1), \dots,$ to $O(\lambda^7)$. It follows easily from (4.1) that

$$\begin{cases} F_1'(t) = -p(1,2)F_1(t) + p(4,1)\lambda^7(t)F_4(t) + O(\lambda^8(t)) \\ \quad = O(\lambda^8(t)), \\ F_3'(t) = -p(3,4)\lambda^3(t)F_3(t) + p(2,3)\lambda^7(t)F_2(t) + O(\lambda^8(t)) \\ \quad = O(\lambda^8(t)), \end{cases} \tag{4.3}$$

which means that F_1, F_3 can be expressed, within an error of $O(\lambda^8)$, in terms of F_2 and F_4. Substituting (4.3) to the differential equations for F_2 and F_4, we have

$$\begin{pmatrix} F_2'(t) \\ F_4'(t) \end{pmatrix} = \begin{pmatrix} -p(2,3) & p(4,1) \\ p(2,3) & -p(4,1) \end{pmatrix} \lambda^7(t) \begin{pmatrix} F_2(t) \\ F_4(t) \end{pmatrix} + O(\lambda^8(t)). \tag{4.4}$$

Denote the 2×2 matrix above by A. Then $\mu = -(p(2,3) + p(4,1))$ is an eigenvalue of A with $u = (-p(2,3), p(4,1))$ as one of its corresponding left-eigenvectors. Multiplying u from the left to both sides of (4.4),

$$g'(t) = \mu\lambda^7(t)g(t) + o(\lambda^7(t)), \tag{4.5}$$

where $g(t) = u(F_2(t), F_4(t))^{\mathrm{T}}$. Note that $O(\lambda^8)$ is replaced by $o(\lambda^7)$ in (4.5), so that (2.3)(ii) holds with $E = 7$ and Lemma 2.1 can be applied to get $g(t) = o(1)$. Since $\lim(F_2(t) + F_4(t)) = 1$ by (4.2), we immediately have

$$\lim F_2(t) = p(4,1)/(p(2,3) + p(4,1)),$$
$$\lim F_4(t) = p(2,3)/(p(2,3) + p(4,1)).$$

Then by (4.3),

$$\lim F_1(t)/\lambda^7(t) = p(4,1)p(2,3)/[p(1,2)(p(2,3) + p(4,1))],$$
$$\lim F_3(t)/\lambda^4(t) = p(4,1)p(2,3)/[p(3,4)(p(2,3) + p(4,1))].$$

Since (4.2) is the key estimate for Theorem 1.1, it is interesting to give it a heuristic argument. Because all $q_{ij}(t)$'s are powers of $\lambda(t)$, it is reasonable to guess that $F_i(t) \approx \lambda^{a(i)}(t)$ as $t \to \infty$. As in the homogeneous case, we expect that, within a tolerable error, a certain "equilibrium" is reached as $t \to \infty$. Thus the amount of flow-in mass to any state i should "balance" the amount of flow-out mass from state i. [This is in fact the contents of (4.3) and (4.4).] Taking $i = 1,2,3,4$ and comparing the order of $\lambda(t)$, we obtain

$$7 + a(4) = a(1) = 7 + a(2) = 3 + a(3) = 7 + a(4).$$

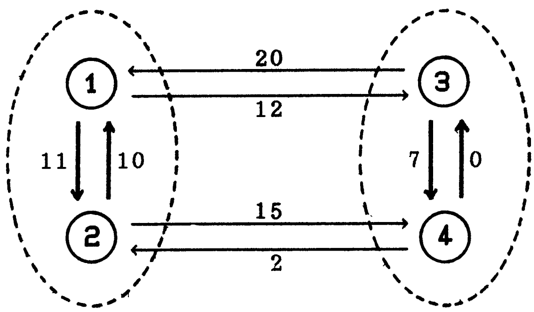

FIG. 2.

Since $\Sigma_1^4 F_i(t) = 1$, we have

$$\min_{1 \le i \le 4} a(i) = 0.$$

Thus $a(1) = 7$, $a(3) = 4$ and $a(2) = a(4) = 0$. This justifies (4.2).

EXAMPLE 2. Let $S = \{1, 2, 3, 4\}$. All the finite-valued $U(i, j)$'s are shown in Figure 2. A boldface arrow there means a cost with $U(i, j) = V(i)$. It is clear that states 3 and 4 will be merged into a state in $S^{(8)}$, and states 1 and 2 into a state in $S^{(12)}$. The structure of $(S^{(12)}, U^{(12)})$ is shown in Figure 3. By (1.8), $h(1) = 0$, $h(2) = 1$, $h(3) = 3$ and $h(4) = 10$. The correct order estimates

$$F_i(t) = O(\lambda^{h(i)}(t)), \qquad i \in S, \tag{4.6}$$

can be obtained through many steps. Take $i = 3, 4$ for example. Lemma 2.2(i) implies that as $t \to \infty$,

$$\lambda^{V(i)}(t) F_i(t) = O(\lambda^{E+1}(t)), \qquad i = 3, 4,$$

hold successively for $E = 0, 1, \ldots, 6$. Note that $V(3) = 7$ and $V(4) = 0$. Lemmas

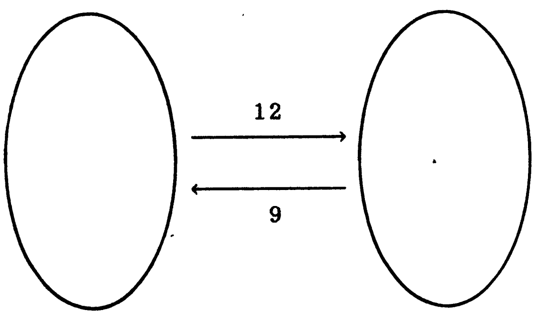

FIG. 3.

2.2 and 2.3 imply that for $i = 3, 4$,

$$\begin{cases} \lambda^9(t)(F_3(t) + F_4(t)) = O(\lambda^{E+1}(t)), \\ \lambda^{V(i)}(t)F_i(t) = \theta(i)\lambda^7(t)\left[(F_3(t) + F_4(t)) + O\left((\lambda(t))^{(E-8)^+ +1}\right)\right] \end{cases}$$

hold successively for $E = 7, 8, \dots, 11$. Now let $E = 11$ to get the estimate in (4.6) for F_3 and F_4.

The limiting constants β_i in (1.12) can be obtained in a similar way. We omit the details.

5. Remarks.

REMARK 1. The aperiodic assumption usually holds automatically. Because $\lim q_{ii}(t) > 0$ for a certain state $i \in S$, unless $\{X(t)\}$ is in fact homogeneous, which is not our concern in this article.

REMARK 2. The irreducibility assumption is for convenience only. If the process is reducible, then one can classify each state as an essential or inessential state exactly the same way as one does in the homogeneous case. We can readily obtain a limit theorem like (1.12), but the description of the β_i's will be messy and usually depends on the initial distribution $X(0)$.

REMARK 3. The concept of a cycle has appeared in Freidlin and Wentzell (1984), where a small random perturbation for a dynamical system is considered. The more general form which is used in this article appeared in some equivalent form in Hwang and Sheu (1986) and Karp (1972). Note that it is related to the minimum cost spanning tree problem in graph theory. A complete analysis of this concept is developed in Chiang and Chow (1987). Here, we only extract from it some of the necessary definitions and refer the readers to Chiang and Chow (1987) and Lawler (1976) for motivation and other applications.

REMARK 4. The reason for introducing a "cycle" is because the probabilities of the process being at two different states in a cycle are comparable, that is, $\lim_{t\to\infty} P(X(t) = i)\lambda^{V(i)}(t)/(P(X(t) = j)\lambda^{V(j)}(t))$ exists for i, j in one cycle. [The interpretation of $P(X(t) = i)$ when i is a cycle should be obvious.] This is partly the content of Lemma 2.3. The integer k that appeared in "kth-order cycle" seems cumbersome. Actually, it plays no role in the mathematics of this article and was introduced only to smooth the description of an induction argument. One can tacitly assume $k = \infty$, bearing in mind the hierarchic structure of the cycles developed at different stages [see (1.4)].

REMARK 5. A continuous analog of our problem takes the form

$$\begin{cases} dX_t = -b(X_t)\,dt + \sqrt{2T(t)}\,dW_t, \\ X_0 = x, \end{cases} \tag{5.1}$$

where $b(x)$ is required to resemble a gradient function in a strong way. The best result for the limiting distribution of (5.1) known to us is in Hwang and Sheu (1986), where a weak limit like (1.3) was obtained. We would like to see an estimate of the convergence rate (1.12) in this setup.

REMARK 6. It is clear from (1.1) that the transition rate functions $\{q_{ij}(t)\}$ considered in this article are of a special type: powers of $\lambda(t)$. In particular,

$$\lim_{t\to\infty} q_{ij}(t)/q_{kl}(t) = \begin{cases} 0 & \text{or} \\ c > 0 & \text{or} \\ \infty \end{cases} \tag{5.2}$$

holds for any i, j, k, l with $i \neq j$ and $k \neq l$. From the viewpoint of balance of the mass flow, which is explained in Section 4, it is not hard to see that Theorems 1.1 and 1.2 hold as well for general transition rate functions satisfying (5.2) and some technical conditions like (A.1; E)–(A.3; E). A more general case that $c_1\varepsilon^{U(i,j)} \leq q_{ij}(t) \leq c_2\varepsilon^{U(i,j)}$ has been considered in Tsitsiklis (1985).

REFERENCES

CHIANG, T.-S. and CHOW, Y. (1987). On Ventcel's optimal graphs. Preprint.

CHIANG, T.-S. and CHOW, Y. (1988). On the convergence rate of annealing processes. *SIAM J. Control Optim.* **26** 1455–1470.

CHUNG, K. L. (1967). *Markov Chains with Stationary Transition Probabilities*, 2nd ed. Springer, Berlin.

FREIDLIN, M. I. and WENTZELL, A. D. (1984). *Random Perturbations of Dynamical Systems.* Springer, Berlin.

GEMAN, D. and GEMAN, S. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Machine Intelligence* **6** 721–741.

GIDAS, B. (1985). Global optimization via the Langevin equation. *Proc. Twenty-Fourth Conf. Decision and Control, Fort Lauderdale, Florida* 774–778.

HAJEK, B. (1988). Cooling schedules for optimal annealing. *Math. Oper. Res.* **13** 311–329.

HWANG, C. R. and SHEU, S. J. (1986). Large time behaviors of perturbed diffusion Markov processes with applications. III. Simulated annealing. Preprint.

KARP, R. M. (1972). A simple derivation of Edmonds' algorithm for optimum branchings. *Networks* **1** 265–272.

KIRKPATRICK, S., GEBATT, C. and VECCHI, M. (1983). Optimization by simulated annealing. *Science* **220** 671–680.

LAWLER, E. L. (1976). *Combinatorial Optimization.* Holt, Rinehart and Winston, New York.

SENETA, E. (1973). *Non-Negative Matrices.* Wiley, New York.

TSITSIKLIS, J. N. (1985). Markov chains with rare transitions and simulated annealing. Preprint.

VENTCEL, A. D. (1972). On the asymptotics of eigenvalues of matrices with elements of order $\exp(-V_{ij}/2\varepsilon^2)$. *Dokl. Akad. Nauk SSSR* **202** 263–265. (In Russian; translation *Soviet Math. Dokl.* **13** 65–68.)

INSTITUTE OF MATHEMATICS
ACADEMIA SINICA
TAIPEI, TAIWAN 11529
REPUBLIC OF CHINA