The Annals of Applied Probability
1998, Vol. 8, No. 3, 896–916

ON THE PROBLEM OF EXIT FROM CYCLES FOR SIMULATED ANNEALING PROCESSES— A BACKWARD EQUATION APPROACH[1]

BY TZUU-SHUH CHIANG AND YUNSHYONG CHOW

Academia Sinica

For a simulated annealing process X_t on S with transition rates $q_{ij}(t) = p_{ij}\exp(-(U(i, j))/T(t))$ where $i, j \in S$ and $T(t) \downarrow 0$ in a suitable way, we study the exit distribution $P_{t,i}(X_\tau = a)$ and mean exit time $E_{t,i}(\tau)$ of X_t from a cycle c as $t \to \infty$. A cycle is a particular subset of S whose precise definition will be given in Section 1. Here τ is the exit time of the process from c containing i and a is an arbitrary state not in c. We consider the differential (backward) equation of $P_{t,i}(X_\tau = a)$ and $E_{t,i}(\tau)$ and show that $\lim_{t\to\infty} P_{t,i}(X_\tau = a)/\exp(-(U(c, a) - V(c))/T(t))$ and $\lim_{t\to\infty} E_{t,i}(\tau)/\exp(V(c)/T(t))$ exist and are independent of $i \in c$. The constant $U(c, a)$ is usually referred to as the cost from c to a and $V(c)$, $(\leq U(c, a))$ is the minimal cost coming out of c. We also obtain estimates of $|P_{t,i}(X_\tau = a) - P_{t,j}(X_\tau = a)|$ and $|E_{t,i}(\tau) - E_{t,j}(\tau)|$ for $i \neq j$ as $t \to \infty$. As an application, we shall show that similar results hold for the family of Markov processes with transition rates $q_{ij}^{\varepsilon} = p_{ij}\exp(-U(i, j)/\varepsilon)$ where $\varepsilon > 0$ is small.

1. Introduction. On a finite set $S = \{1, 2, \ldots, M\}$, consider a (time) inhomogeneous Markov chain X_t with the following transition rates:

$$q_{ij}(t) = \begin{cases} p_{ij}\exp\left(-\dfrac{U(i, j)}{T(t)}\right), & \text{for } j \neq i, \\ -\displaystyle\sum_{k\neq i} q_{ik}(t), & \text{for } j = i, \end{cases} \tag{1.1}$$

where $P = (p_{ij})$ is an irreducible neighborhood choosing matrix with nonnegative entries, $U(i, j)$ is an arbitrary (cost) function from $S \times S$ to $[0, \infty]$ and $T(t) > 0$ is a suitable temperature function converging to 0. Originally, $U(i, j) = [U(j) - U(i)]^+$ in (1.1) and such a chain is called a simulated annealing process ([7], [8], [9] and [11]). We shall, however, abuse the name and call any Markov process of the form (1.1) a simulated annealing process. For convenience, we always assume $p_{ij} = 0$ if and only if $U(i, j) = \infty$. We refer readers to [7], [11] and [16] for some of their applications and motivation.

In [3] and [4], the asymptotic behavior of processes (1.1) was obtained by solving the associated forward equation. Indeed, if $F_i(t) = P(X_t = i)$ and $\lambda(t) = \exp(-1/T(t))$, then the forward equation of X_t takes the following

Received February 1996; revised January 1997.
[1]Supported in part by the National Science Council, Republic of China.
AMS 1991 *subject classifications*. Primary 60J27, 60J99; secondary 15A18, 15A51, 90B40.
Key words and phrases. Simulated annealing process, backward equations, cycles.

form:

$$\dot{F}(t) = Q^T(t)F(t), \tag{1.2}$$

where $Q^T(t)$ is the transpose of $(q_{ij}(t))$ in (1.1), and there are constants $\beta_i > 0$ and a certain function h such that

$$\lim_{t\to\infty} F_i(t)/\lambda(t)^{h(i)} = \beta_i \tag{1.3}$$

for any initial distribution X_0. The function h depends only on U and $h(i) = U(i) - \min U$ in the potential case, that is, $U(i, j) = [U(j) - U(i)]^+$ for some potential function U. For a more probabilistic approach, see [1], [9] and [10]. In this paper we shall study, by solving backward equations, the behavior of (1.1) before the limit (1.3) is reached for low temperatures. This method is new and powerful in the sense that it can solve both the exit distribution and mean exit time problems simultaneously. Moreover, it yields finer estimates than the probabilistic methods ([2] and [13]). To describe our method, let c be a subset of S and "a" an (absorbing) state not in c. Later, we shall consider only cycles which are in some sense "nice" subsets of S and its definition will be given below. Let $x_i(t) = P_{t,i}(X_\tau = a)$ and $y_i(t) = E_{t,i}(\tau)$ where τ in the first exit time from c, $P_{t,i}$ is the distribution and $E_{t,i}$ is the expectation of a simulated annealing process starting at time t from $i \in c$. The differential (backward) equations associated with $x_i(t)$ and $y_i(t)$ take the following forms, respectively:

$$\dot{x}_i(t) = -\sum_{j\in c} q_{ij}(t)x_j(t) - q_{ia}(t), \qquad i \in c, \tag{1.4}$$

$$\dot{y}_i(t) = -\sum_{j\in c} q_{ij}(t)y_j(t) - 1, \qquad i \in c. \tag{1.4$'$}$$

In matrix notation, we have

$$\dot{x}_c(t) = -Q_c(t)x_c(t) - Q_{c,a}(t),$$
$$\dot{y}_c(t) = -Q_c(t)y_c(t) - \mathbf{1},$$

where x_c and y_c are the column vectors of x_i and y_i, $i \in c$, Q_c is the submatrix of Q restricted to c and $Q_{c,a}$ is the column vector of q_{ia}, $i \in c$. In Sections 2 and 3, we shall use (1.4) and (1.4)$'$ to establish results of the following type: there exist some positive constants α, α', θ, θ' and γ, such that for any $i, j \in c$,

$$\lim_{t\to\infty} \frac{x_i(t)}{\lambda^\alpha(t)} = \theta > 0, \qquad |x_i(t) - x_j(t)| = O(\lambda^{\alpha+\gamma}(t))$$

and

$$\lim_{t\to\infty} \frac{y_i(t)}{\lambda^{-\alpha'}(t)} = \theta' > 0, \qquad |y_i(t) - y_j(t)| = O(\lambda^{-\alpha'+\gamma}(t)).$$

To precisely describe our results, we shall briefly recall the definition of cycles and state some technical assumptions on (1.1). Please see [3] and [9] for their origins and necessity. Let U be a cost function on S as in (1.1). Let $V(i) = \min_{j\neq i} U(i, j)$. We start with such a triplet $(S^0, U^0, V^0) = (S, U, V)$.

For two different states $i, j \in S^0$, we say that $i \to j$ if there exists a path $i = i_0, i_1, \ldots, i_n = j$ such that $U(i_k, i_{k+1}) = V(i_k)$ for each $0 \le k \le n-1$. A state $i \in S^0$ is called minimal if $j \to i$ whenever $i \to j$. It follows that if i is minimal and $i \to j$, then j is also minimal and $j \to i$. Two states i and j in S^0 are said to be equivalent ($i \leftrightarrow j$) if either $i = j$ or if $i \to j$ and $j \to i$. It is easy to see that "$\leftrightarrow$" is an equivalence relation and if an equivalence class has one minimal state, all the states in this class are minimal. We shall call equivalence classes that consist of minimal states a (nontrivial) cycle. A state which is not minimal will also be called a (trivial) cycle. Hence a cycle is either an equivalence class under "$\leftrightarrow$" consisting of all minimal states or only nominal and consisting of only one nonminimal state. Let S^1 be the collection of all the cycles thus formed. We next define $U^1(\cdot,\cdot)$ and V^1 on S^1 as follows. For c and $\tilde{c}$ in S^1, let

$$d^1(c) = \max_{i \in c} V^0(i), \tag{1.5}$$

$$U^1(c, \tilde{c}) = d^1(c) + \min_{\substack{i \in c \\ j \in \tilde{c}}} (U^0(i, j) - V^0(i)) \quad \text{and} \quad V^1(c) = \min_{c' \ne c} U^1(c, c'). \tag{1.6}$$

Finally, let

$$J^1(c) = V^1(c) - d^1(c). \tag{1.7}$$

We thus have a new triplet (S^1, U^1, V^1) and (S^n, U^n, V^n) can then be defined inductively until S^{N+1} becomes a singleton for some N. An example is given to illustrate the process to form (S^k, U^k), $k = 1, 2, \ldots, N$ in the Appendix. Elements in S^n shall be called nth order cycles in the sequel but the 0th order cycles will still be called "states." In theory, elements in S^n, $n \ge 1$, can be equivalence classes of equivalence classes and the statement that a state $i \in c^n$ where $c^n \in S^n$ does not necessarily make sense. However, we shall abuse the notation and say that a state i belongs to c^n if there are kth order cycles $c^k \in S^k$, $k = 1, \ldots, n-1$ such that $i \in c^1 \in \cdots \in c^n$. Such a chain of cycles is unique if it exists. The functions d^n and J^n for nth order cycles in S^n are similar: For $c \in S^n$,

$$d^n(c) = \max_{\tilde{c} \in c} (V^{n-1}(\tilde{c})) \tag{1.8}$$

and

$$J^n(c) = V^n(c) - d^n(c). \tag{1.9}$$

For convenience, we inductively define the cost between a cycle and a state: for $c \in S^n$ and $i \in S$, $i \notin c$,

$$U^n(c, i) = d^n(c) + \min_{\tilde{c} \in c} (U^{n-1}(\tilde{c}, i) - V^{n-1}(\tilde{c})).$$

A technical condition we shall always assume throughout the paper on $\lambda(t) = \exp(-1/T(t))$ is the following:

$$\dot{\lambda}(t)/\lambda(t) = o(\lambda(t)^{d^{N+1}(S^{N+1})}) \quad \text{and} \quad \lambda(t) \to 0 \text{ as } t \to \infty. \tag{1.10}$$

(See [4] or [8] for a reason.)

The results can now be described as follows.

THEOREM 1.1. *Let X_t be a simulated annealing process on S satisfying* (1.10). *For any states i, j in an nth order cycle $c \in S^n$, $0 \le n \le N$, and $a \notin c$, we have the following*:

(i) $\lim_{t\to\infty} P_{t,i}(X_\tau = a)/\lambda^{\alpha}(t) = \theta$ *where θ is a positive constant independent of i*;

(ii) $|P_{t,i}(X_\tau = a) - P_{t,j}(X_\tau = a)| = O(\lambda^{\gamma(i,j)+\alpha}(t))$ *as $t \to \infty$.*

Here τ is the exit time from c, $\alpha = U^n(c,a) - V^n(c)$ and $\gamma(i,j) = \sum_{r=k(i,j)}^{n} J^r(c^r)$ where $c^{k(i,j)} \in S^{k(i,j)}$ is the cycle of the lowest order that contains both i and j and $c^{k(i,j)} \in c^{k(i,j)+1} \in \cdots \in c^n = c$.

REMARK. The above statement (i) precisely describes the exit distribution from cycles when t is large. It is also obvious that from (i) with overwhelming probability, the process will exit to those states $a \in S$ where $U^n(c,a) = V^n(c)$.

THEOREM 1.1′. *Let X_t, c, τ, γ and α be as in Theorem* 1.1. *Then for any $i, j \in c$, we have the following*:

(i) $\lim_{t\to\infty} E_{t,i}(\tau)/\lambda^{-V^n(c)}(t) = \delta$ *where δ is a positive constant independent of i*;

(ii) $|E_{t,i}(\tau) - E_{t,j}(\tau)| = O(\lambda^{-V^n(c)+\gamma(i,j)}(t))$ *as $t \to \infty$.*

If c is an arbitrary set of S, our method is still valid but the expression of h will be much more complicated and it will also depend on the starting state i. We shall only concentrate on the cycle case. Our analysis actually asserts that the first-order approximation of the asymptotic behaviors of $x_i(t)$ and $y_i(t)$ can be obtained by first equating $\dot{x}_{c(t)}$ and $\dot{y}_{c(t)}$ to 0 in (1.4) and (1.4)′, respectively, and then by solving the systems of linear (variable coefficient) inhomogeneous equations: for $i \in c$ and $a \notin c$,

$$0 = -\sum_{j\in c} q_{ij}(t)x_j(t) - q_{ia}(t) \quad \text{as } t \to \infty, \tag{1.11}$$

$$0 = -\sum_{j\in c} q_{ij}(t)y_j(t) - 1 \quad \text{as } t \to \infty. \tag{1.11}'$$

Equations (1.11) and (1.11)′ are actually the equations corresponding to the same problem for a family of homogeneous Markov chains X_t^ε with transition rates

$$q_{ij}^\varepsilon = \begin{cases} p_{ij}\exp\left(-\dfrac{U(i,j)}{\varepsilon}\right), & \text{for } j \ne i, \\ -\sum_{k\ne i} q_{ik}^\varepsilon, & \text{for } j = i, \end{cases} \tag{1.12}$$

where $\varepsilon > 0$ is fixed. Indeed, let $x_i^\varepsilon = P_i^\varepsilon(X_\tau = a)$ and $y_i^\varepsilon = E_i^\varepsilon(\tau)$. Then we have

$$0 = -\sum_{j\in c} q_{ij}^\varepsilon x_j^\varepsilon - q_{ia}^\varepsilon, \tag{1.13}$$

$$0 = -\sum_{j\in c} q_{ij}^\varepsilon y_j^\varepsilon - 1. \tag{1.13$'$}$$

Theorems 1.1 and 1.1′ can then be translated to processes (1.12) without any change.

THEOREM 1.2. *Let X_t^ε be the family of Markov chains with transition rates* (1.12). *For any states i, j in an nth order cycle $c \in S^n$, $0 \le n \le N$, and $a \notin c$, we have*:

(i) $$\lim_{\varepsilon\to 0} P_i^\varepsilon(X_\tau = a)/\exp(-\alpha/\varepsilon) = \theta > 0;$$

(ii) $$|P_i^\varepsilon(X_\tau = a) - P_j^\varepsilon(X_\tau = a)| = O(\exp[-(\gamma(i,j)+\alpha)/\varepsilon]) \text{ as } \varepsilon \to 0.$$

Here $\varepsilon, \alpha, \theta$ and γ are defined in Theorem 1.1.

THEOREM 1.2′. *Let X_t^ε, c and τ be as in Theorem* 1.2. *Then for any $i, j \in c$, we have the following*:

(i) $\lim_{\varepsilon\to 0} E_i^\varepsilon(\tau)/\exp(V^n(c)/\varepsilon) = \delta$ *where δ is the same as in Theorem* 1.1′;

(ii) $$|E_i^\varepsilon(\tau) - E_j^\varepsilon(\tau)| = O(\exp(V^n(c) - \gamma(i,j))/\varepsilon) \quad \text{as } \varepsilon \to 0.$$

Processes of the form (1.12) have been studied extensively in [6]. In the context of metastability, this form was studied more recently in [12] and [14] for the case where U is the positive part of the Hamiltonian difference of stochastic Ising models with a small external field. Let $H(\sigma) = -\frac{1}{2}\sum_{|x-y|=1}\sigma(x)\sigma(y) - (h/2)\sum_x \sigma(x)$ where σ is a configuration on $\Lambda_N = \{1, \dots, N^2\}$, $\sigma(x) = 1$ or -1 for $x \in \Lambda_N$ and h is assumed positive. At a fixed temperature T, the Gibbs state is given by $\mu_T(\sigma) = (1/z)\exp(-H(\sigma)/T)$. For any fixed N and $T \to 0$, μ_T concentrates its mass on the configuration with all positive spins, which will be denoted by $+\underline{1}$. The Metropolis algorithm in this set-up is a continuous time Markov chain on the space of all configurations where the transition is only possible between configurations which differ at only one site:

$$q(\eta, \eta^x) = \exp -\left(\frac{H(\eta^x) - H(\eta)}{T}\right)^+,$$

where

$$\eta^x(y) = \begin{cases} \eta(y), & \text{if } x \ne y, \\ -\eta(y), & \text{if } x = y. \end{cases}$$

Obviously, the Metropolis algorithm is reversible with respect to μ_T, and starting from any configuration (particularly $-\underline{1}$) will thus result in a neighborhood of $+\underline{1}$ for T small. The physical interest in this phenomenon is the behavior of the process from $-\underline{1}$ to $+\underline{1}$. It was shown in [12] that for small h and T, if the Metropolis algorithm starts from $-\underline{1}$, it stays close to $-\underline{1}$ for a long time while small rectangles of +1's appear and disappear at various places in the lattice. Then all of a sudden a big rectangle (critical droplet) appears and, in a relatively short period of time, nucleates to the configuration with all +1's. The precise size of the critical droplet and time needed to form a critical droplet were also calculated in [12]. In the language of cycles, we shall describe the process as follows.

Let S^0 be the space of all configurations and $U(\eta, \eta^x) = (H(\eta^x) - H(\eta))^+$. Let $-\underline{1} = c_-^0 \in c_-^1 \in \cdots \in c_-^k$ and $+\underline{1} = c^0 + 0 \in c_+^1 \in \cdots \in c_+^k$ be the unique sequences of cycles containing -1 and $+1$ respectively such that c_-^k, c_+^k will be the first cycles to be included in the same cycle in S^{k+1}. Starting from $-\underline{1}$, the process will spend some time in c_-^k whose expected value is given in Theorem 1.2'. A critical droplet then corresponds to a state in c_+^k and its exact shape is predicted by Theorem 1.2. The fact that the time the process stays close to $+\underline{1}$ will be much longer than $-\underline{1}$ is be predicted by the inequality $V^k(c_+^k) > V^k(c_-^k)$. Actually, their ratio is $\exp((V^k(c_+^k) - V^k(c_-^k))/T)$, which is not contained in [12]. Moreover, the appearance of small droplet of $+1$'s corresponds to the lower order cycles $c_-^1, c_-^2, \ldots$ and the lengths of their duration are also predicted in Theorem 1.2.

We remark that solving the backward equations (1.4) and $(1.4)'$ is very different from the forward equation (1.2). First, (1.4) is not an initial value problem and we can only study its positive bounded solutions. Equation $(1.4)'$ is similar to (1.4) but it does not have bounded positive solutions. Actually, an a priori estimate of $E_{t,i}(\tau)$ is necessary to distinguish $E_{t,i}(\tau)$ from other solutions of $(1.4)'$. This estimate will be pointed out in Section 2. Once this is done, the asymptotic behavior of $(1.4)'$ can be obtained similarly to (1.4). Our method is complicated but, we think, worthwhile because it can treat two different problems simultaneously and yield precise estimates.

2. Case I. First order cycles. All of our techniques in this paper can actually be traced to a simple observation of a first-order ordinary differential equation. This is the content of the following three lemmas and will be used repeatedly throughout.

Let $\lambda(t)$ be a positive function converging to 0 with $\lambda'(t)/\lambda(t) = o(\lambda^k(t))$ as $t \to \infty$. The following lemma will be useful for (1.4).

LEMMA 2.1. *Let $f(t)$ be a bounded complex function. If $f'(t) = (b\lambda^k(t) + o(\lambda^k(t)))f(t) - c\lambda^{k+d}(t) + o(\lambda^{k+d}(t))$ where $\operatorname{Re} b > 0$, then $f(t) = -c\lambda^d(t)/b + o(\lambda^d(t))$ as $t \to \infty$.*

REMARK. The lemma implies that we can equate $f'(t)$ to 0 and solve an algebraic equation to obtain a first-order approximation of $f(t)$.

PROOF. Multiplying both sides of the equation by an integration factor, we have

$$f(t)\cdot\exp\int_{t_0}^{t}(-b\lambda^k+o(\lambda^k))=f(t_0)+\int_{t_0}^{t}(-c\lambda^{k+d}+o(\lambda^{k+d}))\exp\int_{t_0}^{u}(-b\lambda^k+o(\lambda^k)).$$

Under the assumptions on λ it is easy to show that $\int_{t_0}^{\infty}\lambda^k(s)\,ds=\infty$ and for any $L>0$, $\exp\int_{t_0}^{t}(\operatorname{Re}b)\lambda^k(s)\,ds\geq\lambda^{-L}(t)$ holds for t large. Since f is bounded, the right-hand side of the above equation must converge to 0 as $t\to\infty$. If $c\neq 0$, then by l'Hospital's rule,

$$\begin{aligned}&\lim_{t\to\infty}\frac{f(t)}{\frac{c}{b}\lambda^d(t)}\\&\quad=\lim_{t\to\infty}\left[f(t)\exp\int_{t_0}^{t}(-b\lambda^k+o(\lambda^k))\right]\Big/\left[\frac{c}{b}\lambda^d(t)\exp\int_{t_0}^{t}(-b\lambda^k+o(\lambda^k))\right]\\&\quad=\lim_{t\to\infty}\left[-c\lambda^{k+d}+o(\lambda^{k+d})\right]\Big/\left[(cd/b)\lambda^d o(\lambda^k)-c\lambda^{d+k}+o(\lambda^{d+k})\right]=1.\end{aligned}$$

Hence $f(t)=c\lambda^d(t)/b+o(\lambda^d(t))$. The case $c=0$ can be treated similarly. □

Since solutions of (1.4)′ are never bounded, we need the following two lemmas in place of Lemma 2.1 for (1.4)′.

LEMMA 2.1′. *Let $f(t)$ be a complex function satisfying $f(t)=O(\lambda^{-N}(t))$ for some $N>0$. If $f'(t)=(b\lambda^k(t)+o(\lambda^k(t)))f(t)-1+O(\lambda)$ where $\operatorname{Re}b>0$, then $f(t)=\lambda^{-k}(t)/b+o(\lambda^{-k}(t))$.*

PROOF. It is routine as in Lemma 2.1 to show

$$f(t)=\exp\int_{t_0}^{t}b\lambda^k(1+o(1))\left\{f(t_0)+\int_{t_0}^{t}(-1+O(\lambda))\exp\int_{t_0}^{u}-b\lambda^k(1+o(1))\right\}.$$

Because of the a priori estimate $f=O(\lambda^{-N})$, the right-hand side of the above equation can be of order λ^{-N} only if the sum inside the parenthesis converges to 0 as $t\to\infty$. Hence $f(t_0)=-\int_{t_0}^{\infty}(-1+O(\lambda))\exp\int_{t_0}^{u}-b\lambda^k(1+O(1))$ and then

$$f(t)\exp\int_{t_0}^{t}-b\lambda^k(1+o(1))=-\int_{t}^{\infty}(-1+O(\lambda))\exp\int_{t_0}^{u}-b\lambda^k(1+o(1)).$$

The conclusion follows by applying the l'Hospital's rule as in Lemma 2.1. □

LEMMA 2.1″. *Let f be in Lemma 2.1′ and satisfy $\dot f(t)=(b\lambda^k(t)+o(\lambda^k(t)))f(t)+O(\lambda)$. Then $f=O(\lambda^{-k+1})$.*

The proof is similar to that of Lemma 2.1′ and will be omitted. The a priori estimate $f=O(\lambda^{-N})$ in Lemmas 2.1′ and 2.1″ is not trivial. Actually, Lemmas 2.1′ and 2.2″ are false without such a condition. This, however, does not pose a problem to our case. It is known ([4], Theorem 0.2) that $P_{t,i}(\tau>s)\leq$

$\exp\int_t^s(-\delta\lambda^{V^n(c)}+O(\lambda^{V^n(c)+1}))$ for some $\delta>0$ and s large. In fact, by adopting the method used in [5], there is a constant $N>0$ such that the estimate above holds for s and t large with $s\geq t+\lambda^{-N}(t)$. Hence for t large,

$$\begin{aligned}E_{t,i}(\tau)&=\int_t^\infty P_{t,i}(\tau>s)\,ds\\&\leq\int_t^{t+\lambda^{-N}(t)}1+\int_{t+\lambda^{-N}(t)}^\infty\exp\int_t^s(-\delta\lambda^{V^n(c)}+O(\lambda^{V^n(c)+1}))\\&\leq\lambda^{-N}(t)+\int_t^\infty\exp\int_t^s(-\delta\lambda^{V^n(c)}+O(\lambda^{V^n(c)+1}))\\&\leq\lambda^{-N}(t)+2\lambda^{-V^n(c)}(t)/\delta,\end{aligned}$$

where an integration by parts and (1.10) have been used in the last inequality.

Let c be a first-order cycle and $a\in S$ but $a\notin c$. Let $d^1(c)=m$, $V^1(c)=m+r$ and $U^1(c,a)=m+r+s$ for some $r\geq 1$ and $s\geq 0$. For two disjoint sets A,B contained in c, let $Q_{A,B}(t)$ be the transition between A, and B, that is, $Q_{A,B}(t)=(q_{ij}(t))_{i\in A,\,j\in B}$. We define a diagonal matrix $D_{A,B}(t)$ of size $|A|\times|A|$ whose (i,i) position is $-\sum_{j\in B}q_{ij}$. Defined in this way, the row sums of $Q_{A,B}(t)$ and $D_{A,B^c}(t)$ are equal but opposite in sign. Hence $D_{A,B^c}(t)\mathbf{1}+Q_{A,B}(t)\mathbf{1}=0$ where $\mathbf{1}$ is the column vector of suitable size whose entries are all 1's. Similarly, let $Q_A(t)$ be the transitions among the states in A, that is, $Q_A(t)=(q_{ij}^A(t))_{i,\,j\in A}$. Here $q_{ij}^A=q_{ij}$ if $i\neq j$ but on the diagonal, $q_{ii}^A(t)=-\sum_{j\neq i,\,j\in A}q_{ij}(t)$ which is different from $q_{ii}(t)$ defined in (1.1). $Q_A(t)$ is defined in such a way that the row sums of $Q_A(t)$ are 0.

If we use $\dot{x}_A$ for the column vector $(\dot{x}_i;\ i\in A)$ and let $c(k)=\{i\in c;\ V^1(c)=k\}$, then the master equation (1.4) and (1.4)′ can be grouped in the following forms, respectively:

$$\begin{aligned}\dot{x}_{c(k)}=&-Q_{c(k)}x_{c(k)}-D_{c(k),\,c\setminus c(k)}x_{c(k)}-Q_{c(k),c\setminus c(k)}x_{c(k)}\\&-D_{c(k),\,S\setminus c}x_{c(k)}-Q_{c(k),\,a}\quad\text{for } 0\leq k\leq m,\end{aligned}\tag{2.1}$$

$$\begin{aligned}\dot{y}_{c(k)}=&-Q_{c(k)}y_{c(k)}-D_{c(k),\,c\setminus c(k)}y_{c(k)}-Q_{c(k),\,c\setminus c(k)}y_{c(k)}\\&-D_{c(k),\,S\setminus c}y_{c(k)}-\mathbf{1}\quad\text{for } 0\leq k\leq m.\end{aligned}\tag{2.1$'$}$$

To make (2.1) shorter and easier to handle, we shall abbreviate $c(k)$ and $c\setminus c(k)$ to k and $\hat{k}$, respectively, when there is no confusion. Since $Q_{c(k)}(t)$ and $D_{c(k)}(t)$ and $D_{c(k),\,c\setminus c(k)}(t)$ are polynomials in $\lambda(t)$ with lowest term degree k, we can write $-Q_{c(k)}(t)-D_{c(k),c\setminus c(k)}(t)=-Q_k(t)-D_{k,\,\hat{k}}(t)=M_k(1+O(\lambda(t)))\lambda^k(t)$. Similarly, $-Q_{c(k),\,c\setminus c(k)}(t)=N_k(1+O(\lambda(t)))\lambda^k(t)$, $-D_{c(k),\,S\setminus c}(t)=-R_k(1+O(\lambda(t)))\lambda^{k+r}(t)$ and $-Q_{c(k),\,a}(t)=S_k(1+O(\lambda(t)))\lambda^{k+r+s}(t)$ where r,s are defined above. Hence we arrive at the following: for $0\leq k\leq m$,

$$\begin{aligned}\dot{x}_k=&M_k(1+O(\lambda))\lambda^k x_k+N_k(1+O(\lambda))\lambda^k x_{\hat{k}}\\&+R_k(1+O(\lambda))\lambda^{k+r}x_k+S_k(1+O(\lambda))\lambda^{k+r+s},\end{aligned}\tag{2.2}$$

$$\dot{y}_k=M_k(1+O(\lambda))\lambda^k y_k+N_k(1+O(\lambda))\lambda^k y_k+R_k(1+O(\lambda))\lambda^{k+r}y_k-\mathbf{1},\tag{2.2$'$}$$

where M_k, N_k, R_k and S_k are constant matrices satisfying the following characteristic properties:

(2.3) $M_k(1 + O(\lambda(t))$ is the negative of a transition rate matrix,

that is, its off-diagonal elements are negative and all its row sums are positive for each t. Note that this implies M_k is also the negative of a transition rate matrix:

(2.4) $N_k(1 + O(\lambda))$ has all negative entries and $M_k(1 + O(\lambda))\mathbf{1} + N_k(1 + O(\lambda))\mathbf{1} = 0$,

(2.5) $R_k(1+O(\lambda))$ is a positive, diagonal matrix and $S_k(1+O(\lambda))$ is a negative column vector such that $R_k(1+O(\lambda))\mathbf{1}+S_k(1+O(\lambda))\lambda^s \geq 0$.

The reason that we prefer (2.2) rather than (1.4) is that we can eliminate $c(0), \ldots, c(m-1)$ successfully from (2.2). By first showing $\dot{x}_0$ is $O(\lambda^{r+s+1})$, we then can express x_0 in terms of x_k's, $k \geq 1$. Substituting this expression into (2.2) for $k \geq 1$, we thus obtain a system similar to (2.2) [i.e., satisfying (2.3)–(2.5) but without x_0]. Repeating this elimination process until there is only one equation left, Lemma 2.1 then yields the desired result. In the process showing that $\dot{x}_k = O(\lambda^{k+r+s+1})$, we also have to show that all the terms involving $O(\lambda)$ in (2.2) do not really matter and can be absorbed into an error term $O(\lambda^{k+r+s+1})$. We want to remark that (2.2)′ can be treated similarly. Actually, the exact same proof for x_k can be used for y_k except that one uses Lemmas 2.1′ and 2.1″ instead of Lemma 2.1 and keeps in mind the a priori estimate $y_k = O(\lambda^{-m-r})$ in (2.2)′ compared with $x_k = O(\lambda^0)$ in (2.2). Hence for the sake of brevity, we shall only work on (2.2) and omit the obvious translation to (2.2)′. This scheme will be carried out in Steps 1–3 and Lemmas 2.2–2.4.

STEP 1. Treat R_k and S_k as error terms and (2.2) becomes

$$\dot{x}_k = M_k(1 + O(\lambda))\lambda^k x_k + N_k(1 + O(\lambda))\lambda^k x_{\hat{k}} + O(\lambda^{k+r}), \qquad k \geq 0. \tag{2.6}$$

Since x_k is a bounded function, we shall start the induction from the following weaker form (2.7) and trivial estimate (2.8):

$$\dot{x}_k = M_k\lambda^k x_k + N_k\lambda^k x_{\hat{k}} + O(\lambda^{k+1}), \tag{2.7}$$

$$\dot{x}_k = O(\lambda^k) \quad \text{and} \quad x_i - x_j = O(\lambda^0) \qquad \text{for } i, j \in S. \tag{2.8}$$

The next lemma shows that we can improve the error terms in (2.7) from $O(\lambda^{k+1})$ to $O(\lambda^{k+r})$ and in (2.8) from $O(\lambda^k)$, $O(\lambda^0)$ to $O(^{k+r})$, $O(\lambda^r)$, respectively.

LEMMA 2.2. *In* (2.6), *we have* $\dot{x}_0 = M_0x_0 + N_0x_{\hat{0}} + O(\lambda)$.

PROOF. By (2.7), $\dot{x}_0 = M_0 x_0 + N_0 x_{\hat{0}} + O(\lambda)$. Since $-M_0$ is a subtransition rate matrix, all the eigenvalues of M_0 have positive real parts by the Perron–Frobenius theorem [15]. Let μ be an eigenvalue of M_0 and v a corresponding left eigenvector. Multiplying (2.7) by v, we have

$$v\dot{x}_0 = \mu v x_0 + v N_0 x_{\hat{0}} + O(\lambda). \tag{2.9}$$

Let $g = vx_0 + (1/\mu)(vN_0 x_{\hat{0}})$. Since, $\dot{x}_{\hat{0}} = O(\lambda^1)$ from (2.8), we have $\dot{g} = \mu g + O(\lambda)$. Hence by Lemma 2.1, $g = O(\lambda)$ and $v\dot{x}_0 = O(\lambda)$. If w is a generalized eigenvector corresponding to μ, that is, $wM_0 = \mu w + v$ where v is an eigenvector, then

$$w\dot{x}_0 = \mu w x_0 + v x_0 + w N_0 x_{\hat{0}} + O(\lambda).$$

Let $g = wx_0 + (vx_0 + wN_0 x_{\hat{0}})/\mu + O(\lambda)$. Then $\dot{g} = \mu g + O(\lambda)$ by (2.8) and (2.9) and $g = O(\lambda)$ by Lemma 2.1. Thus $w\dot{x}_{c_0} = O(\lambda)$. Since all such v's and w's form a basis of $R^{|C_0|}$ by Jordan's theorem, we thus conclude $\dot{x}_0$ is actually $O(\lambda)$, which improves the first statement in (2.8) for $k = 0$. Solving (2.7) for $k = 0$, we have

$$x_0 = -M_0^{-1} N_0 x_{\hat{0}} + O(\lambda) = -M_0^{-1}(Q_{0,k}(0)x_k + Q_{0,\widehat{0,k}}(0)x_{\widehat{0,k}}) + O(\lambda). \tag{2.10}$$

Here $x_{\widehat{0,k}}$ is the column vector of $\{x_i,\ i \in c\backslash c_0 \cup c_k\}$, $Q_{0,k}$ and $Q_{0,\widehat{0,k}}$ are abbreviation of Q_{c_0,c_k} and $Q_{c_0,c\backslash c_0\cup c_k}$, respectively. Note that M_0^{-1} is a nonnegative matrix and $M_0\mathbf{1} + N_0\mathbf{1} = 0$ in (3.9) implies $-M_0^{-1}N_0\mathbf{1} = \mathbf{1}$, that is, $-M_0^{-1}N_0$ is a usual Markov chain transition matrix. Substituting (2.10) into (2.7) for $k \geq 1$, we have

$$\begin{aligned}
\dot{x}_k &= M_k\lambda^k x_k + Q_{k,0}(k)\lambda^k x_0 + Q_{k,\widehat{0,k}}(k)\lambda^k x_{\widehat{0,k}} + O(\lambda^{k+1}) \\
&= M_k\lambda^k x_k + Q_{k,0}(k)\lambda^k(-M_0^{-1})(Q_{0,k}(0)x_k + Q_{0,\widehat{0,k}}(0)x_{\widehat{0,k}}) \\
&\quad + Q_{k,\widehat{0,k}}(k)\lambda^k x_{\widehat{0,k}} + O(\lambda^{k+1}) \\
&= (M_k + Q_{k,0}(k)(-M_0^{-1})Q_{0,k}(0))\lambda^k x_k + (Q_{0,\widehat{0,k}}(k) \\
&\quad + (Q_{k,0}(k)(-M_0^{-1})Q^{(0)}_{0,\widehat{0,k}})\lambda^k x_{\widehat{0,k}} + O(\lambda^{k+1}) \\
&= M'_k\lambda^k x_k + N'_k\lambda^k x_{\hat{k}} + O(\lambda^{k+1}), \qquad k = 1, 2, \ldots, m.
\end{aligned}$$

Since M_0^{-1} is a positive matrix and

$$\begin{aligned}
M'_k\mathbf{1} + N'_k\mathbf{1} &= M_k\mathbf{1} + Q_{k,\widehat{0,k}}(k)\mathbf{1} + Q_{k,0}(k)(-M_0^{-1})N_0\mathbf{1} \\
&= M_k\mathbf{1} + Q_{k,\widehat{0,k}}(k)\mathbf{1} + Q_{k,0}(k)\mathbf{1} \\
&= M_k\mathbf{1} + Q_{k,\hat{k}}(k)\mathbf{1} = M_k\mathbf{1} + N_k\mathbf{1} = 0,
\end{aligned}$$

properties (2.3)–(2.5) can easily be checked for M'_k and N'_k. We thus have successfully eliminated x_0 from the system (2.2). An obvious induction then follows to establish

$$\dot{x}_k = O(\lambda^{k+1}), \qquad 0 \leq k \leq m-1 \tag{2.11}$$

and we finally obtain

$$\dot{x}_m = M_m \lambda^m x_m + O(\lambda^{m+1}), \tag{2.12}$$

where $-M_m$ is a transition rate matrix because c is a cycle. The same arguments establishing (2.10) can now be applied to show that

$$\dot{x}_m = O(\lambda^{m+1}). \tag{2.13}$$

Hence $M_m x_m = O(\lambda)$. By the Perron–Frobenius theorem, there are $v_1, v_2, \ldots, v_{|c_m|-1}$ generalized left eigenvectors corresponding to nonzero eigenvalues of M_m. Then, for each i, v_i is orthogonal to $\mathbf{1}$, which is the unique right eigenvector corresponding to eigenvalue 0. Thus $\operatorname{span}\{v_1, \ldots, v_{|c_m|-1}\} = \{\mathbf{1}\}^{\perp} = \operatorname{span}\{e_i - e_{i+1}; i = 1, \ldots, |c_m| - 1\}$. Since $v x_m = O(\lambda)$ for each $v \in \operatorname{span}\{v_1, \ldots, v_{c_m - 1}\}$, we have

$$(e_i - e_{i+1}) x_m = x_i - x_{i+1} = O(\lambda) \quad \text{for each } i \in c_m.$$

Substituting it into (2.10) from $k = m - 1$ to 0 successively yields that

$$x_i - x_j = O(\lambda) \quad \text{for all } i, j \in c. \tag{2.14}$$

Hence (2.12), (2.13) and (2.14) improve (2.8) by increasing the power of λ by 1. Again, (2.4), (2.12), (2.13) and (2.14) imply that (2.6) can be written as

$$\dot{x}_k = M_k(1 + O(\lambda))\lambda^k x_k + N_k(1 + O(\lambda))\lambda^k x_{\hat{k}} = O(\lambda^2).$$

Hence we can start over again with an error term $O(\lambda^2)$. An induction obviously follows to establish the lemma.

STEP 2. Treat S_k as the only error term and (2.2) becomes

$$\begin{aligned} \dot{x}_k = {} & M_k(1 + O(\lambda))\lambda^k x_k + N_k(1 + O(\lambda))\lambda^k x_{\hat{k}} \\ & + R_k(1 + O(\lambda))\lambda^{k+r} x_k + O(\lambda^{k+r+s}). \end{aligned} \tag{2.15}$$

By (2.4) and Lemma 2.2, we have from (2.15) the following: for $0 \le k \le m$ and $i, j \in S$,

$$\dot{x}_k = M_k \lambda^k x_k + N_k \lambda^k x_{\hat{k}} + R_k \lambda^{k+r} x_k + O(\lambda^{k+r+1}) \tag{2.16}$$

and

$$\dot{x}_k = O(\lambda^{k+r}), \qquad x_i - x_j = O(\lambda^r) \quad \text{and} \quad x_i = O(\lambda^0). \tag{2.17}$$

The next lemma shows that we can improve the error terms in (2.16) from $O(\lambda^{k+r+1})$ to $O(\lambda^{k+r+s})$ and in (2.17) from $O(\lambda^{k+r})$, $O(\lambda^r)$ and $O(\lambda^0)$ to $O(\lambda^{k+r+s})$, $O(\lambda^{r+s})$ and $O(\lambda^s)$, respectively.

LEMMA 2.3. *In* (2.15), *we have* $\dot{x}_k = O(\lambda^{k+r+s})$, $x_i - x_j = O(\lambda^{r+s})$ *and* $x_i = O(\lambda^s)$.

PROOF. The process is the same as that in Lemma 2.2. We start with $k = 0$ in (2.16): $\dot{x}_k = M_0 x_0 + N_0 x_{\hat{0}} + R_0 \lambda^r x_0 + O(\lambda^{r+1})$. For any left eigenvector v of M_0 corresponding to an eigenvalue μ, we have $(v\dot{x}_0) = \mu(vx_0) + vN_0 x_{\hat{0}} + vR_0\lambda^r x_0 + O(\lambda^{r+1})$. Let $g = \mu(vx_0 + (1/\mu)vN_0 x_{\hat{0}} + (1/\mu)vR_0\lambda^r x_0)$. Then $\dot{g} = \mu g + O(\lambda^{r+1})$ because of (2.17) and (1.10). Hence $g = O(\lambda^{r+1})$ and $v\dot{x}_0 = O(\lambda^{r+1})$. Thus by the same reason as in (2.10), we have $\dot{x}_0 = O(\lambda^{r+1})$ and thus $x_0 = -(M_0 + R_0\lambda^r)^{-1} N_0 x_{\hat{0}} + O(\lambda^{r+1})$. Since

$$-(M_0+R_0\lambda^r)^{-1} = (I+M_0^{-1}R_0\lambda^r)^{-1}(-M_0^{-1}) = (I-M_0^{-1}R_0\lambda^r)(-M_0^{-1})+O(\lambda^{2r}),$$

we have

$$\begin{aligned} x_0 &= (I + M_0^{-1}R_0\lambda^r)(-M_0^{-1})N_0 x_{\hat{0}} + O(\lambda^{r+1}) \\ &= (I - M_0^{-1}R_0\lambda^r)(-M_0^{-1})(Q_{0,k}x_k + Q_{0,\widehat{0,k}}\ x_{\widehat{0,k}}) + O(\lambda^{r+1}). \end{aligned}$$

Substituting x_0 into (2.16) for $k \geq 1$, we have

$$\begin{aligned} \dot{x}_0 &= M_k\lambda^k x_k + Q_{k,0}(k)\lambda^k x_0 + Q_{k,\widehat{0,k}}(k)\lambda^k x_{\widehat{0,k}} \\ &\quad + R_k\lambda^{k+r}x_k + O(\lambda^{k+r+1}) \\ &= M_k\lambda^k x_k + Q_{k,0}(k)\lambda^k(I - M_0^{-1}R_0\lambda^r)(-M_0^{-1}) \\ &\quad \times (Q_{0,k}(k)x_k + Q_{0,\widehat{0,k}}(k)x_{\widehat{0,k}}) \\ &\quad + Q_{k,\widehat{0,k}}(k)\lambda^k x_{\widehat{0,k}} + R_k\lambda^{k+r}x_k + O(\lambda^{k+r+1}) \\ &= M'_k\lambda^k x_k + N'_k\lambda^k x_{\widehat{0,k}} + (R_k + \widetilde{R}_k)\lambda^{k+r}x_k \\ &\quad + \bar{R}_k\lambda^{k+r}x_{\widehat{0,k}} + O(\lambda^{k+r+1}), \end{aligned} \tag{2.18}$$

where

$$\begin{aligned} M'_k &= M_k + Q_{k,0}(k)(-M_0^{-1})Q_{0,k}, \\ N'_k &= Q_{k,\widehat{0,k}}(k) + Q_{k,0}(k)(-M_0^{-1})Q_{0,\widehat{0,k}}(k), \\ \widetilde{R}_k &= Q_{k,0}(k)M_0^{-1}R_0M_0^{-1}Q_{0,k}(k) \end{aligned}$$

and

$$\bar{R}_k = Q_{k,0}(k)M_0^{-1}R_0M_0^{-1}Q_{0,\widehat{0,k}}(k).$$

Since $\widetilde{R}_k$ is a positive matrix and $x_i - x_j = O(\lambda^r)$ by the induction hypothesis, we can replace $\widetilde{R}_k$ by a positive diagonal matrix $\widetilde{D}$ whose diagonal elements are row sums of $\widetilde{R}_k$ and $\widetilde{R}_k x_k - \widetilde{D}x_k = O(\lambda^r)$. Similarly, $\bar{R}_k$ can be replaced by a positive diagonal matrix $\bar{D}$ so that $\bar{R}_k x_{\widehat{0,k}} - \bar{D}x_k = O(\lambda^r)$. If $R'_k = R_k + \widetilde{D} + \bar{D}$, then $R'_k x_k = (R_k + \widetilde{R}_k)x_k + \bar{R}x_{\widehat{0,k}} + O(\lambda^r)$. Hence

$$\dot{x}_k = M'_k\lambda^k x_k + N'_k\lambda^k x_{\widehat{0,k}} + R'_k\lambda^{k+r}x_k + O(\lambda^{k+r+1}), \tag{2.19}$$

where M'_k, N'_k and R'_k satisfy (2.3)–(2.5). An induction thus follows until $k = m$, in which case (2.19) becomes

$$\dot{x}_m = M'_m \lambda^m x_m + R'_m \lambda^{m+r} x_m + O(\lambda^{m+r+1}). \tag{2.20}$$

Let v be the unique positive unit left eigenvector of M'_m corresponding to 0. Then $(v\dot{x}_m) = \alpha\lambda^{m+r}(vx_m) + O(\lambda^{m+r+1})$, $vx_m = O(\lambda)$ and $x_m = O(\lambda)$ by Lemma 2.1. Again, $\dot{x}_m = M'_m \lambda^m x_m + O(\lambda^{k+r+1})$ from (2.20) and we obtain

$$\dot{x}_m = O(\lambda^{k+r+1}), \qquad x_i - x_j = O(\lambda^{r+1}) \quad \text{and} \quad x_i = O(\lambda) \quad \text{for } i, j \in c \tag{2.21}$$

as in the proof of Lemma 2.2. This proves the case $s = 1$ and an obvious induction follows for any s. □

We remark that the eigenvector v for M'_m in (2.20) is the ergodic distribution of M'_m since it is a transition rate matrix. Let $x_c = vx_m$. Then obviously,

$$\dot{x}_c = O(\lambda^{k+r+s}). \tag{2.22}$$

Also, we only take the leading term in the expansion of $(I + M_0^{-1} R_0 \lambda^r)^{-1}$. This can be justified by the estimate of x_i's in (2.21).

STEP 3. Consider now the full equation (2.2). Lemma 2.3 implies that

$$\dot{x}_k = M_k \lambda^k x_k + N_k \lambda^k x_{\hat{k}} + R_k \lambda^{k+r} x_k + S_k \lambda^{k+r+s} + O(\lambda^{k+r+s+1}). \tag{2.23}$$

LEMMA 2.4. *In* (2.21), *we have* $\dot{x}_k = O(\lambda^{k+r+1})$, $0 \le k \le m$, *and* $x_i = \theta \cdot \lambda^s + o(\lambda^s)$.

PROOF. Following the same method as in Lemma 2.3, we shall obtain that $\dot{x}_0 = O(\lambda^{r+s+1})$ and $x_0 = -(M_0 + R_0\lambda^r)^{-1}(N_0 x_{\hat{0}} + S_0^{r+s}) + O(\lambda^{k+r+s+1})$ in (2.21). Similarly to (2.19), we therefore have $\dot{x}_k = M'_k \lambda^k x_k + N'_k \lambda^k x_{\widehat{0,k}} + R'_k \lambda^{k+r} x_k + S'_k \lambda^{k+r+s} + O(\lambda^{k+r+s+1})$ with M'_k, N'_k, R'_k and S'_k satisfying (2.3) to (2.5). When $k = m$, we have $\dot{x}_m = M'_m \lambda^m x_m + R'_m \lambda^{m+r} x_m + S'_m \lambda^{m+r+s} + O(\lambda^{m+r+1})$. Hence $(v\dot{x}_m) = \alpha\lambda^{m+r}(vx_m) + \beta\lambda^{m+r+s} + O(\lambda^{m+r+s+1})$. Therefore $vx_m = \theta\lambda^s + o(\lambda^s)$ for some $\theta > 0$. This completes the proof. □

Combining Lemmas 2.2–2.4, we arrive at the main conclusion of this section.

THEOREM 2.5. *Let* c *be a first order cycle and* $a \notin c$. *If* τ *is the first exit time from* c, *then* $P_{t,i}(X_\tau = a) = \theta \cdot \lambda^s(t) + o(\lambda^s(t))$ *and* $|P_{t,i}(X_\tau = a) - P_{t,j}(X_\tau = a)| = O(\lambda^{r+s})$ *for* $i, j \in c$. *Here* $r = V^1(c) - d^1(c) \ge 1$ *and* $s = U^1(c, a) - V^1(c) \ge 0$.

3. Case II. Second- and higher order cycles. We shall treat second-order cycles in some detail in this section and one should easily see that higher order cycles can be treated similarly. Since we will be dealing with second-order cycles, we shall use $\mathfrak{C} \in S^2, (\mathfrak{U} = U^1)$ and $\mathfrak{V}(= V^1)$ for second-order cycles and c, U, V for first-order cycles. Let $\mathfrak{C}(k) = \{c \in \mathfrak{C};\ \mathfrak{V}(c) = k\}$ for $0 \le k \le M$ where $M = d^1(\mathfrak{C})$ $(= \max_{c\in\mathfrak{C}} \mathfrak{V}(c))$. As in Section 2, for $c \in \mathfrak{C}$, let $c(k) = \{i \in c, V(i) = k\}$ for $0 \le k \le m$ where $m = d(c)$ $(= \max_{i\in c} V(i))$. We remark that $J^1(c) = k - m > 0$. If $V^2(\mathfrak{C}) = M + R$, $R \ge 1$, and $U^2(\mathfrak{C}, a) = M+R+S$, $S \ge 0$ then the backward equation of x_i, $i \in c(k)$ takes the following form: for $0 \le k \le m = \max_{i\in c} V(i)$,

$$\begin{aligned}\dot{x}_{c(k)} = (\dot{x}_i)_{i\in c(k)} &= M_{c(k)}(1 + O(\lambda))\lambda^k x_{c(k)} + N_{c(k)}(1 + O(\lambda))\lambda^k x_{c(\hat{k})} \\ &\quad + E_{c(k)}(1 + O(\lambda))\lambda^{k+K-m} x_{c(k)} \\ &\quad + F_{c(k)}(1 + O(\lambda))\lambda^{k+K-m} x_{\mathfrak{C}\setminus c} \\ &\quad + R_{c(k)}(1 + O(\lambda))\lambda^{k+K-m+R} x_{c(k)} \\ &\quad + S_{c(k)}(1 + O(\lambda))^{k+K-m+R+S}.\end{aligned} \tag{3.1}$$

Here, as in Section 2, $M_{c(k)}$ represents transitions among $c(k)$ and $N_{c(k)}$ represents the transitions between $c(k)$ and the rest of c. $E_{c(k)}$ and $R_{c(k)}$ are diagonal matrices representing the transition rates from $c(k)$ to $\mathfrak{C}\setminus c$ and $S\setminus\mathfrak{C}$ respectively. Obviously, $c(\hat{k}) = \{i \in c,\ V(i) \ne k)\}$ and $\hat{c} = \mathfrak{C}\setminus c$. The coefficient matrices in (3.1) satisfy the following:

$$M_{c(k)}(1 + O(\lambda))\mathbf{1} + N_{c(k)}(1 + O(\lambda))\mathbf{1} = 0, \tag{3.2}$$

that is, the row sums of M_k and N_k are 0. In particular, $M_{c(k)}\mathbf{1} + N_{c(k)}\mathbf{1} = 0$,

$$E_{c(k)}(1 + O(\lambda))\mathbf{1} + F_{c(k)}(1 + O(\lambda))\mathbf{1} = 0 \quad \text{and} \quad E_{c(k)}\mathbf{1} + F_{c(k)}\mathbf{1} = 0, \tag{3.3}$$

$$R_{c(k)}(1 + O(\lambda))\mathbf{1} + S_{c(k)}(1 + O(\lambda))\lambda^s\mathbf{1} \ge 0 \tag{3.4}$$

and

(3.5) M_k is a transition rate matrix and E_k and R_k are diagonal matrices.

Like (2.3)–(2.5), properties (3.2)-(3.5) are characteristic of (3.1). The method for treating second-order cycles is similar to that for first-order cycles, only slightly more complicated. Our ultimate goal is to show $\dot{x}_{c(k)} = O(\lambda^{k+(K-m)+R+S+1})$ and eliminate $c(k)$, $k \le m - 1$, from c and $\mathfrak{C}(K)$, $K \le M - 1$, from $\mathfrak{C}$. This shall be done in four steps.

STEP 1. Considering first only two terms in the right-hand side of (3.1) and treating the rest as error terms, we then have an equation exactly the same as in (2.6):

$$\dot{x}_{c(k)} = M_{c(k)}(1 + O(\lambda))\lambda^k x_{c(k)} + N_{c(k)}(1 + O(\lambda))\lambda^k x_{c(\hat{k})} + O(\lambda^{k+K-m}). \tag{3.6}$$

Hence, as in Lemma (2.2), we have the following lemma.

LEMMA 3.1. *In* (3.6) *we have* $x_i - x_j = o(\lambda^{K-m})$, $\dot{x}_i = O(\lambda^{k+K-m})$ *for* $i, j \in c(k)$ *and* $\dot{x}_c = O(\lambda^k)$.

In the first step, each first-order cycle is considered separately without interaction. The estimate $\dot{x}_c = O(\lambda^K)$ is trivial from (2.12).

STEP 2. To include the terms of order λ^{k+K-m} from (3.1), we have, for $c \in \mathfrak{C}(K)$ and $d(c) = m$,

$$\begin{aligned}\dot{x}_{c(k)} &= M_{c(k)}(1+O(\lambda))\lambda^k x_{c(k)} + N_{c(k)}(1+O(\lambda))\lambda^k x_{c(\hat{k})} \\ &\quad + E_{c(k)}(1+O(\lambda))\lambda^{k+K-m} x_{c(k)} \\ &\quad + F_{c(k)}(1+O(\lambda))\lambda^{k-K-m} x_{\mathfrak{C}(\hat{c})} + O(\lambda^{k+K-m+R}).\end{aligned} \tag{3.7}$$

Here we replace states x_i, $i \in \mathfrak{C}\backslash c$ by cycles $x_{\hat{c}}$ where $i \in \hat{c}$. This can be justified by Lemma 3.1. For a first-order cycle $c \in \mathfrak{C}(K)$, we have by Lemma 3.1 the following trivial estimates for (3.7):

$$\begin{aligned}\dot{x}_{c(k)} &= M_{c(k)}\lambda^k x_{c(k)} + N_{c(k)}\lambda^k x_{c(\hat{k})} + E_{c(k)}\lambda^{k+K-m} x_{c(k)} \\ &\quad + F_{c(k)}\lambda^{k+K-m} x_{\mathfrak{C}(\hat{c})} + O(\lambda^{k+K-m+1}),\end{aligned} \tag{3.8}$$

$$\dot{x}_c = O(\lambda^K), \qquad x_c - x_{c'} = O(\lambda^0) \tag{3.9}$$

and

$$\dot{x}_{c(k)} = O(\lambda^{k+K-m}), \qquad x_i - x_j = O(\lambda^{K-m}) \quad \text{for } i, j \in c. \tag{3.10}$$

The next lemma shows that we can improve the error estimates in (3.8) from $O(\lambda^{k+K-m+1})$ to $O(\lambda^{k+K-m+R})$ and in (3.9) from $O(\lambda^K)$, $O(\lambda^0)$ to $O(\lambda^{k+R})$, $O(\lambda^R)$ and in (3.10) from $O(\lambda^{k+K-m})$, $O(\lambda^{K-m})$ to $O(\lambda^{k+K-m+R})$, $O(\lambda^{K-m+R})$, respectively.

LEMMA 3.2. *For a cycle* $c \in \mathfrak{C}(K)$ *with* $d(c) = m$, $\dot{x}_{c(k)} = O(\lambda^{k+K-m+R})$, $x_i - x_j = O(\lambda^{K-m+R})$, $\dot{x}_c = O(\lambda^{K+R})$ *and* $x_c - x_{c'} = O(\lambda^R)$.

PROOF. Since there are terms $x_{\hat{c}}$ of other cycles appearing in $\dot{x}_{c(k)}$, we cannot treat each cycle c separately. We shall start from $\mathfrak{C}(0)$.The cycles in $\mathfrak{C}(0)$ are actually states in S (trivial cycles) and their backward equation is as follows:

$$\dot{x}_{\mathfrak{C}(0)} = E_{\mathfrak{C}(0)} + F_{\mathfrak{C}(0)} x_{e(\hat{0})} + O(\lambda).$$

Here $E_{\mathfrak{C}(0)}$ is the matrix playing the same role as M_k in (2.2) and $F_{\mathfrak{C}(0)} x_{\mathfrak{C}(\hat{0})}$ is obtained from (1.4) by replacing $x_j(t)$ by $x_{\hat{c}}(t)$, $j \in \hat{c}$. Let μ be an eigenvalue of $E_{\mathfrak{C}(0)}$ and v a left eigenvector corresponding to μ. Then $v\dot{x}_{\mathfrak{C}(0)} = \mu v x_{\mathfrak{C}(0)} + vF_{\mathfrak{C}(0)} x_{\mathfrak{C}(\hat{0})} + O(\lambda)$. Let $g = vx_{\mathfrak{C}(0)} = (1/\mu) vF_{\mathfrak{C}(0)} x_{\mathfrak{C}(\hat{0})}$. We then have $\dot{g} = \mu g + O(\lambda)$ and hence $g = O(\lambda)$ by Lemma 2.1. Therefore $v\dot{x}_{\mathfrak{C}(0)} = O(\lambda)$. Proceeding

as in Lemma 2.2, we conclude $\dot{x}_{\mathfrak{C}(0)} = O(\lambda)$. Solving for $x_{\mathfrak{C}(0)}$, we have

$$\begin{aligned} x_{\mathfrak{C}(0)} &= -E_{\mathfrak{C}(0)}^{-1} F_{\mathfrak{C}(0)} x_{\mathfrak{C}(\hat{0})} + O(\lambda) \\ &= -E_{\mathfrak{C}(0)}^{-1} (F_{\mathfrak{C}(0), \mathfrak{C}\setminus c} x_{\mathfrak{C}\setminus\mathfrak{C}(0),\, c} + F_{\mathfrak{C}(0),\, c}) + O(\lambda). \end{aligned} \tag{3.11}$$

Substituting (3.11) into (3.8) for $c \in \mathfrak{C}(K)$, $K \geq 1$, we have

$$\begin{aligned} \dot{x}_{c(k)} &= M_{c(k)} \lambda^k x_{c(k)} + N_{c(k)} \lambda^k x_{c(\hat{k})} + E_{c(k)} \lambda^{k+K-m} \\ &\quad + F_{c(k),\, \mathfrak{C}(\hat{0})} \lambda^{k-K-m} x_{\mathfrak{C}(\hat{0})} F_{c(k)\mathfrak{C}(0)} \lambda^{k+K-m} x_{\mathfrak{C}(0)} + O(\lambda^{k+K-m+1}) \\ &= M_{\mathfrak{C}(k)} \lambda^k x_{c(k)} + N_{c(k)} \lambda^k x_{c(\hat{k})} + E_{c(k)} \lambda^{k+K-m} \\ &\quad + F_{c(k),\, \mathfrak{C}(\hat{0})} \lambda^{k-K-m} x_{\mathfrak{C}(\hat{0})} \\ &\quad + F_{c(k),\, \mathfrak{C}(\hat{0})} \lambda^{k-K-m} (-E_{\mathfrak{C}(0)}^{-1}) (F_{\mathfrak{C}(0),\, \mathfrak{C}\setminus c} x_{\mathfrak{C}\setminus\mathfrak{C}(0),\, c} + F_{\mathfrak{C}(0),\, c} x_c) \\ &\quad + O(\lambda^{k+K-m+1}) \\ &= M_{\mathfrak{C}(k)} \lambda^k x_{c(k)} + N_{c(k)} \lambda^k x_{c(\hat{k})} + E'_{c(k)} \lambda^{k+K-m} x_{c(k)} \\ &\quad + F'_{c(k)} \lambda^{k-K-m} x_{\mathfrak{C}(\hat{0})} + O(\lambda^{k+K-m+1}), \end{aligned} \tag{3.12}$$

where $F'_{c(k)} = F_{c(k),\, \mathfrak{C}(\hat{0})} + F_{c(k),\, \mathfrak{C}(\hat{0})}(-E_{\mathfrak{C}(0)}^{-1}) F_{\mathfrak{C}(0),\, \mathfrak{C}\setminus c}$ and $E'_{c(k)} = E_{c(k)} + \bar{E}_{c(k)}$ where $\bar{E}_{c(k)}$ is the diagonal matrix whose diagonal elements are the column vector $F_{c(k),\, \mathfrak{C}(0)}(-E_{\mathfrak{C}(0)}^{-1}) F_{\mathfrak{C}(0),\, c}$. Since $M_{c(k)}$, $N_{c(k)}$, $E'_{c(k)}$ and $F'_{c(k)}$ continue to satisfy (3.2)–(3.5), we shall drop the "primes" and still call them $E_{c(k)}$ and $F_{c(k)}$. We next consider $c \in \mathfrak{C}(1)$. From (3.12) we have

$$\dot{x}_{c(0)} = M_{c(0)} x_{c(0)} + E_{c(0)} \lambda x_{c(0)} + F_{c(0)} \lambda x_{\mathfrak{C}(\hat{0})} + O(\lambda^2).$$

Multiplying both sides of the above equation by the unit left eigenvector corresponding to 0 eigenvalue, we have $\dot{x}_c = E_c \lambda x_c + F_c \lambda x_{\mathfrak{C}(\hat{0})} + O(\lambda^2)$ where E_c is the sum of the diagonal of $E_{c(0)}$. By Lemma 2.1, $\dot{x}_c = O(\lambda^2) = E_c \lambda x_c + F_c \lambda x_{\mathfrak{C}(\hat{0})}$ and hence $\dot{x}_{c(0)} = M_{c(0)} x_{c(0)} + O(\lambda^2)$. Thus $x_i - x_j = O(\lambda^2)$ if $i, j \in c$ by Step 1 in Section 2. Collecting all the cycles $c \in \mathfrak{C}(1)$, we thus have $\dot{x}_{\mathfrak{C}(1)} = M_{\mathfrak{C}(1)} \lambda x_{\mathfrak{C}(1)} + N_{\mathfrak{C}(1)} \lambda x_{\mathfrak{C}(\hat{1})} + O(\lambda^2)$, which has the same form as the first-order cycle case in Section 2. Hence the induction applies and we conclude $\dot{x}_{\mathfrak{C}(M)} = M_{\mathfrak{C}(M)} \lambda^M x_{\mathfrak{C}(M)} + O(\lambda^{M+1})$. Thus $\dot{x}_{\mathfrak{C}(M)} = O(\lambda^{M+1})$ and $x_c - x_{c'} = O(\lambda^1)$ as in Lemma 2.3. This completes the proof for the first round. We then repeat the same process starting from (3.8)–(3.10) with the error term one degree higher in λ. An induction then completes the proof. □

STEP 3. To include the terms of order $\lambda^{k+K-m+R}$ in consideration from (3.1), we have

$$\begin{aligned} \dot{x}_{c(k)} &= M_{c(k)}(1 + O(\lambda)) \lambda^k x_{c(k)} + N_{c(k)}(1 + O(\lambda)) \lambda^k x_{c(\hat{k})} \\ &\quad + E_{c(k)}(1 + O(\lambda)) \lambda^{k+K-m} x_{c(k)} + F_{c(k)}(1 + O(\lambda)) \lambda^{k+K-m} \\ &\quad + R_{c(k)}(1 + O(\lambda)) \lambda^{k+K-m+R} x_{c(k)} + O(\lambda^{k+K-m+R+S}) \end{aligned} \tag{3.13}$$

where $c \in \mathfrak{C}(K)$, $d(c) = m$ and $0 \le k \le m$. By Lemma 3.2, for a first-order cycle c in (3.13), we have the following estimates:

$$\begin{aligned}\dot{x}_{c(k)} &= M_{c(k)}\lambda^k x_{c(k)} + N_{c(k)}\lambda^k x_{c(\hat{k})} + E_{c(k)}\lambda^{k+K-m} x_{c(k)} \\ &\quad + F_{c(k)}\lambda^{k+K-m} x_{\mathfrak{C}(\hat{c})} + R_{c(k)}\lambda^{k+K-m+R} x_{c(k)} + O(\lambda^{k+K-m+R+1}) \\ &= M + N + E + F + R + O(\lambda^{k+K-m+R+1}),\end{aligned} \tag{3.14}$$

$$\dot{x}_c = O(\lambda^{K+R}), \qquad x_c - x_{c'} = O(\lambda^R), \qquad x_c = O(\lambda^S) \tag{3.15}$$

and

$$\dot{x}_{c(k)} = O(\lambda^{k+K-m+R}), \qquad x_i - x_j = O(\lambda^{K-m+R}) \quad \text{for } i, j \in c. \tag{3.16}$$

We will again improve the error estimates in (3.14)–(3.16) in the following.

LEMMA 3.3. *For a cycle $c \in \mathfrak{C}(K)$ with $d(c) = m$, we have*

$$\begin{aligned}\dot{x}_c &= O(\lambda^{K+R+S}), \qquad x_c - x_{c'} = O(\lambda^{R+S}), \qquad x_c = O(\lambda^S), \\ \dot{x}_{c(k)} &= O(\lambda^{k+K-m+R+S}) \quad \textit{and} \quad x_i - x_j = O(\lambda^{K-m+R+S}).\end{aligned}$$

PROOF. Similar to that in Lemma 3.2, we start from the backward equations of $\mathfrak{C}(0)$ in (3.14),

$$\dot{x}_{\mathfrak{C}(0)} = E_{\mathfrak{C}(0)} x_{\mathfrak{C}(0)} + F_{\mathfrak{C}(0)} x_{\mathfrak{C}(\hat{0})} + R_{\mathfrak{C}(0)}\lambda^R x_{\mathfrak{C}(0)} + O(\lambda^{R+1})$$

and obtain $\dot{x}_{\mathfrak{C}(0)} = O(\lambda^{R+1})$. Solving for $x_{\mathfrak{C}(0)}$, we have

$$\begin{aligned}x_{\mathfrak{C}(0)} &= -(E_{\mathfrak{C}(0)} + R_{\mathfrak{C}(0)}\lambda^R)^{-1} F_{\mathfrak{C}(\hat{0})} x_{\mathfrak{C}(\hat{0})} + O(\lambda^{R+1}) \\ &= (I + E^{-1}_{\mathfrak{C}(0)} R_{\mathfrak{C}(0)}\lambda^R)^{-1}(-E^{-1}_{\mathfrak{C}(0)}) F_{\mathfrak{C}(0)} x_{\mathfrak{C}(\hat{0})} + O(\lambda^{R+1}) \\ &= (I - E^{-1}_{\mathfrak{C}(0)} R_{\mathfrak{C}(0)}\lambda^R)(-E^{-1}_{\mathfrak{C}(0)}) F_{\mathfrak{C}(0)} x_{\mathfrak{C}(\hat{0})} + O(\lambda^{R+1}) \\ &= (-E^{-1}_{\mathfrak{C}(0)} + E^{-1}_{\mathfrak{C}(0)} R_{\mathfrak{C}(0)}\lambda^R E^{-1}_{\mathfrak{C}(0)})\big(F_{\mathfrak{C}(0),\,\mathfrak{C}(0)\setminus c} x_{\mathfrak{C}(0)\setminus c} \\ &\qquad + F_{\mathfrak{C}(0),\,c} x_c + O(\lambda^{R+1})\big) \\ &= x_{\mathfrak{C}(0)\setminus c} + x_c + \widetilde{R}\lambda^R x_{\mathfrak{C}(0)\setminus c} + R'\lambda^R x_c + O(\lambda^{R+1}).\end{aligned} \tag{3.17}$$

Substituting (3.17) into (3.14) for $c \in \mathfrak{C}(K)$, $K \ge 1$, we have

$$\begin{aligned}\dot{x}_{c(k)} &= M + N + E + F_{c(k),\,\mathfrak{C}(0)}\lambda^{k-K-m} x_{\mathfrak{C}(0)} \\ &\quad + F_{c(k),\,\mathfrak{C}(\hat{0})}\lambda^{k+K-m} x_{\mathfrak{C}(\hat{0})} + R + O(\lambda^{k+K-m+R+1}) \\ &= M + N + E + R \\ &\quad + F_{c(k),\,\mathfrak{C}(0)}\lambda^{k-K-m}(\widetilde{F} x_{\mathfrak{C}(0)\setminus c} + \widetilde{E} x_c + \widetilde{R}\lambda^R x_{\mathfrak{C}(0)\setminus c} + R'\lambda^R x_c) \\ &\quad + O(\lambda^{k+K-m+R+1}) + F_{c(k),\,\mathfrak{C}(\hat{0})}\lambda^{k+K-m} x_{\mathfrak{C}(\hat{0})} \\ &= M + N + E' + F' + R' + O(\lambda^{k+K-m+R+1}),\end{aligned}$$

where

$$E' = E + F_{c(k),\,\mathfrak{C}(0)}\lambda^{k+K-m}\widetilde{E}x_c,$$
$$F' = F_{c(k),\,\mathfrak{C}(\hat{0})}\lambda^{k+K-m}x_{\mathfrak{C}(\hat{0})} + F_{\mathfrak{C}(k),\,\mathfrak{C}(0)}\widetilde{F}\lambda^{k+K-m}x_{\mathfrak{C}(\hat{0})}$$

and

$$R' = R + F_{c(k),\,\mathfrak{C}(0)}\lambda^{k+K-m+R}(\widetilde{R}x_{\mathfrak{C}(0)\setminus c} + R'x_c).$$

Since $x_c - x_i = O(\lambda^{K-m+R})$, for $i \in c$ by (3.15) and R' is a positive matrix, we can replace $F_{\mathfrak{C}(k),\,\mathfrak{C}(0)}R'$ by a positive diagonal matrix D' whose diagonal elements are row sums of $F_{c(k),\,\mathfrak{C}(0)}R'$ such that $F_{c(k),\,\mathfrak{C}(0)}R'x_c - D'x_{c(k)} = O(\lambda^{K-m+R})$. Similarly, $F_{c(k),\,\mathfrak{C}(0)}\widetilde{R}$ can be replaced by a positive diagonal matrix $\widetilde{D}$ with the same row sums such that $F_{c(k),\,\mathfrak{C}(0)}\widetilde{R}x_{\mathfrak{C}(0)\setminus c} - \widetilde{D}x_c = O(\lambda^R)$. Hence (3.2)–(3.5) are satisfied. Following the same procedure as in Lemma 3.2, we have

$$\dot{x}_{c(0)} = M_{c(0)}x_{c(0)} + E_{c(0)}\lambda x_{c(0)} + F_{c(0)}\lambda x_{\mathfrak{c}(\hat{c})} + R_{c(0)}\lambda^{R+1}x_{c(0)} + O(\lambda^{R+2})$$

and

$$\dot{x}_c = E_c\lambda x_c + F_c\lambda x_{\mathfrak{c}(\hat{c})} + R_c\lambda^{R+1}x_c + O(\lambda^{R+2}), \qquad c \in \mathfrak{C}(1).$$

Collecting all the cycles in $\mathfrak{C}(1)$, we have

$$\dot{x}_{\mathfrak{C}(1)} = M_{\mathfrak{C}(1)}\lambda x_{\mathfrak{C}(1)} + N_{\mathfrak{C}(1)}\lambda x_{\mathfrak{C}(\hat{1})} + R_{\mathfrak{C}(1)}\lambda^{R+1}x_{\mathfrak{C}(1)} + O(\lambda^{R+2}),$$

which has the same form as in Lemma 2.3. Thus the same induction procedures apply and this completes the proof. □

STEP 4. Because of Lemma 3.3, (3.1) can now be written as follows:

$$\begin{aligned}\dot{x}_{c(k)} = {} & M_{c(k)}\lambda^k x_{c(k)} + N_{c(k)}\lambda^k x_{c(\hat{k})} + E_{c(k)}\lambda^{k+K-m}x_{c(k)} \\ & + F_{c(k)}\lambda^{k+k-m}x_{\mathfrak{C}(\hat{c})} + R_{c(k)}\lambda^{k+K-m+R}x_{c(k)} \\ & + S_{c(k)}\lambda^{k+K-m+R+S} + O(\lambda^{k+K-m+R+S+1})\end{aligned} \tag{3.18}$$

where $c \in \mathfrak{C}(k)$, $d(c) = m$ and $0 \le k \le m$. The following are from Lemma 3.3:

$$\dot{x}_{c(k)} = O(\lambda^{K+R+S}), \qquad x_c - x_{c'} = O(\lambda^{R+S}), \qquad x_c = O(\lambda^S) \tag{3.19}$$

and

$$\dot{x}_{c(k)} = O(\lambda^{k+K-m+R+S}), \qquad x_i - x_j = O(\lambda^{K-m+R+S}). \tag{3.20}$$

LEMMA 3.4. *For a cycle $c \in \mathfrak{C}$, $x_c = \theta \cdot \lambda^S + O(\lambda^{S+1})$ and $\dot{x}_c = O(\lambda^{k+R+S+1})$. Also, $\dot{x}_{c(k)} = O(\lambda^{k+K-m+R+S+1})$ and $x_i - x_j = O(\lambda^{K-m+R+S+1})$ if $i, j \in c \in \mathfrak{C}(K)$ and $d(c) = m$.*

PROOF. We start from $\mathfrak{C}(0)$,

$$\dot{x}_{\mathfrak{C}(0)} = E_{\mathfrak{C}(0)}x_{\mathfrak{C}(0)} + F_{\mathfrak{C}(0)}x_{\mathfrak{C}(\hat{0})} + R_{\mathfrak{C}(0)}\lambda^R x_{\mathfrak{C}(0)} + S_{\mathfrak{C}(0)}\lambda^{R+S} + O(\lambda^{R+S+1}).$$

As in Lemma 3.3, we have $\dot{x}_{\mathfrak{C}(0)} = O(\lambda^{R+S+1})$ and hence

$$x_{\mathfrak{C}(0)} = -(E_{\mathfrak{C}(0)} + R_{\mathfrak{C}(0)}\lambda^R)^{-1}(F_{\mathfrak{C}(0)}x_{\mathfrak{C}(\hat{0})} + S_{\mathfrak{C}(0)}\lambda^{R+S}) + O(\lambda^{R+1}).$$

Repeating the process in Lemma 3.3, we thus eliminate all states in $\mathfrak{C}(K)$, $0 \le K \le M-1$ and finally obtain the following:

$$\dot{x}_{\mathfrak{C}(M)} = M_{\mathfrak{C}(M)}\lambda^M x_{\mathfrak{C}(M)} + R_{\mathfrak{C}(M)}\lambda^{M+R}x_{\mathfrak{C}(M)} + S_{\mathfrak{C}(M)}\lambda^{M+R+S} + O(\lambda^{M+R+S+1}).$$

Multiplying by the left eigenvector of $M_{\mathfrak{C}(M)}$ corresponding to 0, we have

$$\dot{x}_{\mathfrak{C}} = \alpha\lambda^{M+R}x_{\mathfrak{C}} + \beta\lambda^{M+R+S} + O(\lambda^{M+R+S+1}).$$

Hence $x_{\mathfrak{C}} = \theta\lambda^S + O(\lambda^{S+1})$ where $\theta = \beta/\alpha$. The rest of the proof is identical to that in Lemma 2.4. □

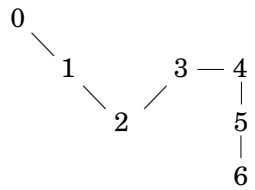
Combining Lemmas 3.1–3.4, we have the main conclusion of this section.

THEOREM 3.5. *For any states i, j in a second order cycle $\mathfrak{C}$ and a state $a \notin \mathfrak{C}$, we have $P_{t,i}(X_\tau = a) = \theta \cdot \lambda^S(t) + O(\lambda^{S+1}(t))$ for t large, where τ is the first exit time from $\mathfrak{C}$ and $S = U^2(\mathfrak{C}, a) - V^2(\mathfrak{C})$. Moreover, if i, j belong to different first order cycles then $|P_{t,i}(X_\tau = a) - P_{t,j}(X_\tau = a)| = O(\lambda^{R+S})$, where $R = V^2(\mathfrak{C}) - d^2(\mathfrak{C})$. If $i, j \in c \in \mathfrak{C}(K)$, then $|P_{t,i}(X_\tau = a) - P_{t,j}(X_\tau = a)| = O(\lambda^{(R+S+K-m)})$, where $m = d^1(c)$.*

It is clear now how the induction will go for cycles of any order. The general statement is in Theorem 1.1.

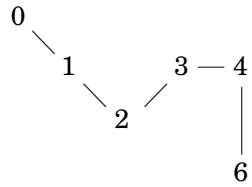
APPENDIX

We consider an example of seven points in the following graph: $S = \{0, 1, 2, 3, 4, 5, 6\}$ with $U(0,1) = U(1,2) = U(3,4) = U(3,2) = U(4,3) = u(4,5) = U(5,6) = 0, U(1,0) = U(2,1) = U(2,3) = U(5,4) = U(6,5) = 1$ and $U(i,j) = \infty$ for other i, j:



Since $\{5, 6\}$ is the only equivalence class of minimal states, it is a cycle in S^1. Other states are not minimal and each of them forms a trivial cycle. Hence $S^1 = \{\{0\}, \{1\}, \{2\}, \{3\}, \{4\}, \bar{6}\}$ where $\bar{6} = \{5, 6\}$. For brevity, we write $S^1 = \{0, 1, 2, \ldots, 4, \bar{6}\}$. An easy computation of (1.6) shows that $U^1(i,j) = U(i,j)$

for $i, j = 0, \ldots, 4$, and $U^1(4, \bar{6}) = 0, U^1(\bar{6}, 4) = 2$. The graph becomes the following:



Now, $\{1, 2, 3, 4, \bar{6}\}$ is the only equivalence class of minimal states, hence $S^2 = \{0, \bar{\bar{6}}\}$ where $\bar{\bar{6}} = \{1, 2, 3, 4, \bar{6}\}$. Again, an easy computation shows that $U^2(0, \bar{\bar{6}}) = 0$ and $U^2(\bar{\bar{6}}, 0) = 3$. The graph is the following:

Obviously S^3 is a singleton and the process stops.

The content of Theorem 1.1(i) asserts that starting from 5 or 6, with probability close to 1, the exit distribution X_τ from $\{5, 6\}$ concentrates on δ_4. This is, of course, trivial because $P_{t,5}(X_\tau = 4)$ is alway 1. However, if one adds an arrow between 3 and 5, that is, if we let $U(5, 3) = 1$ (or 2), then the cycle formation processes remain the same but $P_{t,5}(X_\tau = 3) \to \theta$ [or $P_{t,5}(X_\tau = 3)/\lambda(t) \to \theta$] for some $\theta > 0$ as $t \to \infty$.

Acknowledgment. The authors express their gratitude to the referees for their constructive suggestions on the presentation of cycles and their physical interpretation in the stochastic Ising model. They also thank the referee for pointing out an error in the definition of cycles.

REFERENCES

[1] CATONI, O. (1992). Rough large deviation estimates for simulated annealing—application to exponential schedules. *Ann. Probab.* **20** 1109–1146.

[2] CHEN, D., FENG, J. and QIAN, M. (1995). The metastability of exponentially perturbed Markov chains. *Chinese Sci. A* **25** 590–595.

[3] CHIANG, T. S. and CHOW, Y. (1989). On the asymptotic behavior of some inhomogeneous Markov processes. *Ann. Probab.* **17** 1483–1502.

[4] CHIANG, T. S. and CHOW, Y. (1994). The asymptotic behavior of simulated annelaing with absorption. *SIAM J. Control Optim.* **32** 1247–1265.

[5] CHOW, Y. and HSIEH, J. (1992). On occupation times of annealing processes. *Bull. Math. Academia Sinica* **20** 19–26.

[6] FREIDLIN, M. I. and WENTZELL, A. D. (1984). *Random Perturbations of Dynamical Systems.* Springer, New York.

[7] GEMAN, S. and GEMAN, D. (1984). Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intelligence* **6** 721–741.

[8] GIDAS, B. (1985). Global optimization via the Langevin equation. In *Proceedings 24th IEEE Conference on Decision and Control, Fort Lauderdale, FL* 774–778. IEEE, New York.
[9] HAJEK, B. (1988). Cooling schedules for optimal annealing. *Math. Oper. Res.* **13** 311–329.
[10] HWANG, C. R. and SHEU, S. J. (1992). Singular perturbed Markov chains and exact behaviors of simulated annealing process. *J. Theoret. Probab.* **5** 223–249.
[11] KIRKPATRICK, S., GELATT, C. and VECCHI, M. (1983). Optimization by simulated annealing. *Science* **220** 671–680.
[12] NEVES, E. J. and SCHONMANN, R. H. (1991). Critical droplets and metastability for a Glauber dynamics at very low temperatore. *Comm. Math. Phys.* **137** 209–230.
[13] OLIVIERI, E. and SCOPPOLA, E. (1996). Markov chains with exponentially small transition probabilities: First exit problem from a general domain II. *J. Statist. Phys.* **84** 987–1041.
[14] SCHONMANN, R. H. (1994). Slow droplet-driven relaxation of stochastic Ising models in the vicinity of the phase coexistence region. *Comm. Math. Phys.* **161** 1–49.
[15] SENETA, E. (1981). *Nonnegative Matrices and Markov Chains*, 2nd ed. Springer, New York.
[16] VAN LAARHOVEN, P. J. M. and AARTS, E. H. L. (1987). *Simulated Annealing: Theory and Applications.* Reidel, Dordrecht.

INSTITUTE OF MATHEMATICS
ACADEMIA SINICA
TAIWAN, TAIPEI 11529
E-MAIL: matsch@ccvax.sinica.edu.tw.
maychow@ccvax.sinica.edu.tw