SIMULATED ANNEALING TYPE MARKOV CHAINS AND THEIR ORDER BALANCE EQUATIONS [1]

D. P. Connors and P. R. Kumar [2]

Abstract

We consider generalized simulated-annealing type Markov chains where the transition probabilities are proportional to powers of a vanishing small parameter. One can associate with each state an "order of recurrence" which quantifies the asymptotic behavior of the state occupation probability. These orders of recurrence satisfy a fundamental balance equation across every edge-cut in the graph of the Markov chain. Moreover, the Markov chain converges in a Cesaro-sense to the set of states having the largest recurrence orders. These results convert the analytic problem of determining the asymptotic properties of the time-inhomogeneous stochastic process into a purely algebraic problem of solving the balance equations to determine the recurrence orders.

We provide graph theoretic algorithms to determine the solutions of the balance equations. By applying these results to the problem of optimization by simulation algorithm, we show that the sum of the recurrence order and the cost is a constant for all states in a certain connected set, whenever a "weak-reversibility" condition is satisfied. This allows us to obtain the necessary and sufficient condition for the optimization algorithm to hit the global minimum with probability one.

Keywords: Simulated Annealing, Optimization, Markov Chains.

---

[1] This research has been supported in part by JSEP Contract No. N00014-84-C-0149, USARO Contract No. DAAL-03-88-K-0046, and by AFOSR Contract No. AFOSR-88-0181.

[2] (Please address all correspondence to the second author.) Dept. of Electrical and Computer Engineering and the Coordinated Science Laboratory, University of Illinois, 1101 W. Springfield Ave., Urbana, Illinois 61801.

1 INTRODUCTION

We consider finite state Markov chains $\{x(t)\}$ with transition probabilities of the type,

$$p_{ij}(t) = c_{ij}\epsilon(t)^{V_{ij}},$$

where $\epsilon(t)$ is a small parameter converging to zero. In a previous paper [1] we have shown that if one defines "orders of recurrence" by (more precise definitions are given in Section 2),

$$\beta_i := \sup\{c \geq 0 : \sum_{t=0}^{\infty} \epsilon(t)^c \pi_i(t) = +\infty\},$$

then

(i) these recurrence orders satisfy a balance equation, $\max_{i\in A, j\in A^c}(\beta_i - V_{ij}) = \max_{i\in A, j\in A^c}(\beta_j - V_{ji})$, for every subset A, and

(ii) the Markov process converges to the set of states with the largest orders of recurrence.

This provides a novel approach to analyzing the asymptotic behavior of such time-inhomogeneous Markov processes. Specifically, one uses (i) to solve the balance equations, and then (ii) provides the limiting behavior. Moreover the orders of recurrence also provide information about the rates of convergence of the state occupation probabilities. This approach via recurrence orders therefore converts the analytic problem of determining the asymptotic behavior of the time-inhomogeneous process into a purely algebraic problem of solving the balance equations.

A significant motivation for studying such Markov chains lies in the fact that in the method of optimization by simulated annealing, if $\{W_i\}$ is the cost function whose minimum is sought, then one obtains a Markov chain with,

$$p_{ij}(t) = c_{ij}\epsilon(t)^{\max(0, W_j - W_i)}$$

Thus simulated annealing is a special case where the powers V_{ij} satisfy,

$$V_{ij} := \max(0, W_j - W_i),$$

for some $\{W_i\}$.

In order to pursue the above approach to analyzing such time-inhomogeneous Markov chains, it is necessary to be able to solve the balance equations.

However, there can be non-unique solutions to the balance equations. We present graph-theoretic circulation based algorithms to obtain *a* solution, as well as *all* solutions, to the balance equations. We show by an example the interesting phenomenon that such non-uniqueness can arise when the asymptotic properties of the Markov process, and the recurrence orders, depend not just on the *exponents* V_{ij}, but also on the *proportionality constants* c_{ij}.

By applying these results to the Markov chain arising from the method of optimization by simulated annealing when the "weak reversibility" condition of Hajek [1] holds, we show that the sum of the recurrence order and the cost is a constant on sets connected by recurrent arcs. This allows us to obtain the necessary and sufficient condition for the optimization algorithm to hit the global minimum with probability one. Our necessity result is a stronger sample path result than is found in [1] or [2].

Background

Tsitsiklis [2] has also investigated Markov chains with transition probabilities proportional to powers of a small time-varying parameter. His analysis was based on observing that due to the slow variation of $\{\epsilon(t)\}$, one can employ bounds on the state occupation probabilities for *stationary* Markov chains, where $\epsilon(t)$ is held constant, to obtain bounds for the time-inhomogeneous case. His approach is quite different from ours.

Based on an analogy to the physical process of annealing, the sequence $\epsilon(t)$ is called the "cooling schedule," and just as in the physical analogy it plays a key role in determining asymptotic behavior. It was shown by Geman and Geman [3], Mitra, Romeo and Sangiovanni-Vincentelli [4], and Gidas [5], that simulated annealing converges in probability to a minimum of the optimization problem provided $\sum_{t=0}^{\infty} \epsilon(t)^p = +\infty$ for large enough p. Hajek [1] has determined the necessary and sufficient conditions on the value of p for the algorithm to converge *in probability* to the minimum when a "weak reversibility" assumption is satisfied.

2 ORDERS OF RECURRENCE AND BALANCE EQUATIONS

Consider a Markov chain over a finite state space X whose transition probabilities are proportional to *powers* of a vanishing time varying parameter $\epsilon(t)$; that is, the transition probabilities $p_{ij}(t) := \Pr(x(t+1) = j | x(t) = i)$

are given by,

$$p_{ij}(t) = c_{ij}\epsilon(t)^{V_{ij}}, \qquad \text{for all } i,j \in X, i \neq j, \text{ and } t \in \mathcal{Z}^+, \text{ and} \tag{1}$$
$$p_{ii}(t) = 1 - \sum_{j \neq i} p_{ij}(t).$$

where

$$0 \leq V_{ij} \leq +\infty, \quad \text{for all } i,j \in X, i \neq j, \tag{2}$$
$$c_{ij} \geq 0, \qquad \text{for all } i,j \in X, i \neq j, \text{ and } \sum_j c_{ij} = 1 \text{ for all } i \tag{3}$$

Regarding the small parameter $\{\epsilon(t)\}$, we will assume that,

$$0 < \epsilon(t) < 1, \qquad \text{for all } t \in \mathcal{Z}^+ \tag{4}$$
$$\exists M < \infty \text{ such that } \epsilon(t) \leq M\epsilon(s) \text{ whenever } t \geq s, \text{ and} \tag{5}$$
$$\textstyle\sum_{t=1}^{\infty} \epsilon(t)^p < \infty \qquad \text{for some } p \in [1, +\infty) \tag{6}$$

In what follows we will assume that in (1–3) we have

$$c_{ij} = 0 \Longleftrightarrow V_{ij} = +\infty$$

which is clearly without any loss of generality. We shall denote by N_i the set of all states j with $c_{ij} > 0$. Finally, we will assume that the Markov chain is "connected", i.e. for every $i, j \in X$, there exists a path $i = i_0, \ldots, i_p = j$, with $i_l \in N_{i_{l-1}}$ for $1 \leq l \leq p$.

Let $\pi_i(t) := \Pr(x(t) = i)$ be the probability distribution of $x(t)$, and $\pi_{ij}(t) := \Pr(x(t) = i, x(t+1) = j)$ be the probability of a *transition* from state i to j at time t.

The following example motivates the notion of "orders of recurrence" introduced in [1].

Example 1 Suppose, for a certain Markov chain (with more than 2 states!), we have

$$\pi_1(t) = 1/t^{1/3}, \quad \pi_2(t) = 1/t^{2/3} \quad \text{and} \quad \epsilon(t) = 1/t^{1/3}.$$

Then note that $\sum_{t=0}^{\infty} \epsilon(t)^c \pi_1(t)$ is *finite* if $c > \beta_1 := 2$ and $+\infty$ if $c \leq \beta_1$. Similarly, $\sum_{t=0}^{\infty} \epsilon(t)^c \pi_2(t)$ is *finite* if $c > \beta_2 := 1$ and $+\infty$ if $c \leq \beta_2$. Now $\pi_1(t)$ converges to zero more slowly than $\pi_2(t)$ and it is easy to see that this information is also captured by the demarcation points β_1 and β_2, which thus provide a measure by which to rank the rates at which $\pi_1(t)$ and $\pi_2(t)$ converge to zero. ■

Motivated by this we define the *recurrence orders* for the states and transitions of the Markov process, as follows.

Definition 1 *The order of recurrence of a state $i \in X$, denoted β_i, is*

$$\beta_i := \begin{cases} -\infty & \text{if } \sum_{t=0}^{\infty} \pi_i(t) < +\infty, \\ p^- & \text{if } p = \sup\{c \geq 0 : \sum_{t=0}^{\infty} \epsilon(t)^c \pi_i(t) = +\infty\} \text{ and } \sum_{t=0}^{\infty} \epsilon(t)^p \pi_i(t) < +\infty, \\ p & \text{if } p = \max\{c \geq 0 : \sum_{t=0}^{\infty} \epsilon(t)^c \pi_i(t) = +\infty\}. \end{cases}$$

We say a state i is *transient* if $\beta_i = -\infty$; otherwise we say the state is *recurrent.*

In a similar manner we define the *order of recurrence of the transition from i to j.*

Definition 2 *The order of recurrence of the transition from state i to j, denoted β_{ij}, is*

$$\beta_{ij} := \begin{cases} -\infty & \text{if } \sum_{t=0}^{\infty} \pi_{ij}(t) < +\infty \\ p^- & \text{if } p = \sup\{c \geq 0 : \sum_{t=0}^{\infty} \epsilon(t)^c \pi_{ij}(t) = +\infty\} \text{ and } \sum_{t=0}^{\infty} \epsilon(t)^p \pi_{ij}(t) < +\infty, \\ p & \text{if } p = \max\{c \geq 0 : \sum_{t=0}^{\infty} \epsilon(t)^c \pi_{ij}(t) = +\infty\}. \end{cases}$$

Again, we say the transition from i to j is *transient* if $\beta_{ij} = -\infty$; otherwise we say the transition is *recurrent.*

It is also convenient to define ρ, the *order of cooling* of $\{\epsilon(t)\}$, as follows.

Definition 3 *The order of the cooling schedule $\{\epsilon(t)\}$, denoted ρ, is defined as,*

$$\rho := \begin{cases} -\infty & \text{if } \sum_{t=0}^{\infty} \epsilon(t) < +\infty, \\ p^- & \text{if } p = \sup\{c \geq 0 : \sum_{t=0}^{\infty} \epsilon(t)^c = +\infty\} \text{ and } \sum_{t=0}^{\infty} \epsilon(t)^p < +\infty, \\ p & \text{if } p = \max\{c \geq 0 : \sum_{t=0}^{\infty} \epsilon(t)^c = +\infty\}. \end{cases}$$

The relationship between β_i, β_{ij} and ρ is given in the following Lemma 1. It will be convenient in the sequel to define the operation "$\ominus$" as follows:

$$a \ominus b := \begin{cases} -\infty & \text{if } a < b \\ a - b & \text{if } a \geq b. \end{cases}$$

Lemma 1 *β_{ij} and β_i are related by*

$$\beta_{ij} = \beta_i \ominus V_{ij}, \qquad \textit{for all } i, j \in X, \tag{7}$$

while ρ and β_i are related by,

$$\max_{i \in X} \beta_i = \rho. \tag{8}$$

Proof: If $j \notin N_i$ then it immediately follows that $\beta_{ij} = -\infty$. If $j \in N_i$ then applying the Chapman-Kolmogorov equation,

$$\begin{aligned} \pi_{ij}(t) &= \pi_i(t) p_{ij}(t) \\ &= c_{ij}\epsilon(t)^{V_{ij}} \pi_i(t), \end{aligned}$$

gives the first assertion. Similarly, since

$$\sum_{t=0}^{\infty} \epsilon(t)^p = \sum_{i \in X} \sum_{t=0}^{\infty} \epsilon(t)^p \pi_i(t),$$

the second assertion also follows. ■

Knowledge of the β_i's provides useful information about the asymptotic properties of $\{x(t)\}$. The following Theorem shows that the time-inhomogeneous Markov chain converges in a *Cesaro sense* to the set of states having the largest orders of recurrence.

Theorem 1 *Let $\mathcal{M}$ be the set of states with the largest orders of recurrence,*

$$\mathcal{M} := \{i \in X : \beta_i = \rho\}.$$

Then

$$\limsup_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} \Pr(x(t) \in \mathcal{M}) = 1. \tag{9}$$

Proof: Let us first consider the set $\overline{\mathcal{M}}$ defined by,

$$\overline{\mathcal{M}} := \begin{cases} \mathcal{M} & \text{if } \rho = 0, -\infty \text{ or } p^- \text{ for some } p \in \mathcal{R},\, p > 0, \\ \mathcal{M} \cup \{i \in X : \beta_i = p^-\} & \text{if } \rho = p \text{ for some } p \in \mathcal{R},\, p > 0. \end{cases}$$

Note that if $\rho = p$, then $\overline{\mathcal{M}}$ may be slightly larger than $\mathcal{M}$ since it includes states, if any, whose recurrence orders are p^-; otherwise it is the same as $\mathcal{M}$. We will first show that,

$$\limsup_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} \Pr(x(t) \in \overline{\mathcal{M}}) = 1. \tag{10}$$

Consider first the case $\rho > 0$. Clearly $\rho = p$ or p^- for some $p \in \mathcal{R}$, where $p > 0$. Let

$$Q = \{q \in \mathcal{R} : \text{ for some } i \in \overline{\mathcal{M}}^c, \beta_i = q \text{ or } q^-\}.$$

Let $\theta = \inf_{q\in Q}(p - q)$, where $\inf \phi = +\infty$. Let

$$\gamma = \begin{cases} \theta & \text{if } \theta < +\infty \\ p & \text{if } \theta = +\infty. \end{cases}$$

Consider the states in $\overline{\mathcal{M}}^c$ and observe that for sufficiently small $\delta > 0$,

$$\sum_{t=0}^{\infty} \Pr(x(t) \in \overline{\mathcal{M}}^c)\epsilon(t)^{p-\gamma+\delta} < +\infty,$$

since the state space is finite. An application of Kronecker's Lemma (see Chung [6]) gives

$$\lim_{N\to\infty} \epsilon(N)^{p-\gamma+\delta} \sum_{t=1}^{N} \Pr(x(t) \in \overline{\mathcal{M}}^c) = 0;$$

that is,

$$\lim_{N\to\infty} (N\epsilon(N)^{p-\gamma+\delta}) \frac{1}{N} \sum_{t=1}^{N} \Pr(x(t) \in \overline{\mathcal{M}}^c) = 0. \tag{11}$$

Now we claim that

$$\limsup_{N\to\infty} N\epsilon(N)^{p-\gamma+\delta} > 0. \tag{12}$$

Suppose not. Then,

$$\lim_{N\to\infty} N\epsilon(N)^{p-\gamma+\delta} = 0,$$

and so

$$\lim_{N\to\infty} \frac{1/N}{\epsilon(N)^{p-\gamma+\delta}} = +\infty.$$

In particular, we have

$$\lim_{N\to\infty}\left(\frac{1/N}{\epsilon(N)^{p-\gamma+\delta}}\right)^{\frac{p-\delta}{p-\gamma+\delta}} = +\infty,$$

implying that

$$\lim_{N\to\infty}\frac{(1/N)^{\frac{p-\delta}{p-\gamma+\delta}}}{\epsilon(N)^{p-\delta}} = +\infty.$$

However, since $\sum_{t=0}^{\infty}\epsilon(t)^{p-\delta} = +\infty$, this would imply that

$$\sum_{N=1}^{\infty}\left(\frac{1}{N}\right)^{\frac{p-\delta}{p-\gamma+\delta}} = +\infty, \text{ for all small } \delta > 0,$$

which is false. Hence, (12) holds and from (11) we deduce that

$$\liminf_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\Pr(x(t)\in\overline{\mathcal{M}}^c) = 0. \tag{13}$$

But since

$$\frac{1}{N}\sum_{t=1}^{N}\Pr(x(t)\in\overline{\mathcal{M}}) + \frac{1}{N}\sum_{t=1}^{N}\Pr(x(t)\in\overline{\mathcal{M}}^c) = 1,$$

the result (10) follows.

Now turn to the case $\rho = 0$. Then clearly, $\sum_{t=0}^{\infty}\Pr(x(t)\in\overline{\mathcal{M}}^c) < +\infty$, and so (13) is again true and the result (10) follows.

If $\rho = -\infty$, the result (10) is trivial.

In order to proceed from (10) to (9), it is clearly sufficient to show that in the case $\rho = p$ for some $p \in \mathcal{R}$, $p > 0$,

$$\lim_{t\to\infty}\Pr(x(t)\in\{i : \beta_i = p^-\}) = 0.$$

This involves some results on the structure of the recurrence orders and is demonstrated in Lemma 5. ■

Thus, knowledge of the recurrence orders $\{\beta_i\}$ provides knowledge about the asymptotic properties of the time-inhomogeneous Markov chain. In fact, as the reader may see from Example 1, the recurrence orders also provide information about the rates of convergence.

Our goal therefore is to determine the recurrence orders, and critical to that will be the following result established in [1], which shows that there is a fundamental balance of recurrence orders across every edge-cut in the graph of the Markov chain.

Theorem 2 (Order Balance)

$$\max_{i \in A, j \in A^c} \beta_{ij} = \max_{i \in A, j \in A^c} \beta_{ji}, \qquad \textit{for every } A \subseteq X. \tag{14}$$

Equivalently, using the "$\ominus$" notation and (7),

$$\max_{i \in A, j \in A^c} \beta_i \ominus V_{ij} = \max_{i \in A, j \in A^c} \beta_j \ominus V_{ji}, \qquad \textit{for every } A \subseteq X. \tag{15}$$

Proof: We sketch the proof; see [7] for the precise proof. Choose $A \subseteq X$ and note that if $\{\tau(n)\}_{n\geq 1}$ is the sequence of random times at which the process moves from A to A^c, while $\{\sigma(n)\}_{n\geq 1}$ is the sequence of random times at which the process moves from A^c back to A, then one has,

$$\tau(n) < \sigma(n) < \tau(n+1),$$

where we have assumed, without loss of generality, that $x(0) \in A$ to give $\tau(1) < \sigma(1)$. Using this it follows from (5) that,

$$\begin{aligned}
\sum_{t=0}^{+\infty} \epsilon(t)^c I(x(t) \in A^c, x(t+1) \in A) &= \sum_{n=1}^{+\infty} \epsilon(\sigma(n))^c \\
&\leq M^c \sum_{n=1}^{+\infty} \epsilon(\tau(n))^c \\
&= M^c \sum_{t=0}^{+\infty} \epsilon(t)^c I(x(t) \in A, x(t+1) \in A^c) \\
&= M^c \sum_{n=1}^{+\infty} \epsilon(\tau(n+1))^c + M^c \epsilon(\tau(1))^c \\
&\leq M^{2c} \sum_{n=1}^{+\infty} \epsilon(\sigma(n))^c + M^{2c}\epsilon(0)^c \\
&= M^{2c} \sum_{t=0}^{+\infty} \epsilon(t)^c I(x(t) \in A^c, x(t+1) \in A) + M^{2c}\epsilon(0)^c.
\end{aligned}$$

By taking expected values and using the Monotone Convergence Theorem, it follows that,

$$\sum_{t=0}^{\infty} \epsilon(t)^c \sum_{i \in A, j \in A^c} \pi_{ij}(t) < +\infty \Longleftrightarrow \sum_{t=0}^{\infty} \epsilon(t)^c \sum_{i \in A^c, j \in A} \pi_{ij}(t) < +\infty.$$

Hence both sides above converge or diverge together. Now if c is so large that every term on the LHS with $i \in A, j \in A^c$ converges, then clearly c is also so large that every term on the RHS converges. Thus,

$$c > \max_{i \in A, j \in A^c} \beta_{ij} \Longleftrightarrow c > \max_{i \in A, j \in A^c} \beta_{ji}.$$

Likewise if c is small enough so that some term on the LHS diverges, then c is also small enough so that some term on the RHS diverges, and so

$$c \leq \max_{i \in A, j \in A^c} \beta_{ij} \Longleftrightarrow c \leq \max_{i \in A, j \in A^c} \beta_{ji}.$$

■

Note that through Theorems 1 and 2 we have converted the problem of determining the asymptotic properties of the time-inhomogeneous Markov chain into an *algebraic* problem of solving the balance equations (14). Note that (14) provides a maximum of $2^{|X|}$ equations, one for each edge-cut.

3 THE MODIFIED BALANCE EQUATIONS

Note that if $(\beta_1, \beta_2, \ldots, \beta_{|X|})$ satisfy (15), then $(\beta_1 - a, \beta_2 - a, \ldots, \beta_{|X|} - a)$ also satisfy (15) for every a, i.e. the solution set is *translation invariant.* Thus the equation (8) which fixes the *maximum* of the β_i's needs to also be taken into account.

However, (15,8) can together *still* possess non-unique solutions for sufficiently small values of ρ. In this section, we will show how one can obtain *one* solution to (15,8); in the next section we show how to obtain *all* solutions.

In cases where there is an unique solution to the order balance equations, the algorithm of this section gives an $O(|X|^3)$ algorithm for determining it, compared to the algorithm of Section 4 for obtaining all solutions (in the non-unique case), which is exponential in $|X|$. Also, the results of this section are used in the analysis of the simulated annealing algorithm in Section 5.

It is convenient to consider the following "modified" balance equations, which as we show in the sequel, always possess a unique solution. Given

$\rho \geq 0$ and $V_{ij} \geq 0$ for $i, j = 1, \ldots, |X|$ with $i \neq j$, consider the problem of determining $\lambda := (\lambda_1, \ldots, \lambda_{|X|})$ such that

$$\max_{i \in A, j \in A^c} \lambda_i - V_{ij} = \max_{i \in A, j \in A^c} \lambda_j - V_{ji}, \qquad \text{for every } A \subseteq \{1, \ldots, |X|\}, \quad (16)$$

and

$$\max_i \lambda_i = \rho. \quad (17)$$

We call (16, 17) the "modified" balance equations. Observe that (16) differs from (15) in that the operation "$-$" is used in place of "$\ominus$." Also, the λ's can be negative in (16).

We have introduced the modified balance equations in order to avoid the difficulties in handling $-\infty$ that occur under the "$\ominus$" operation.

Theorem 3 (Properties of Order Balance and Modified Balance Equations)

1. *If λ satisfies the modified balance equations for a given ρ and V, then β defined by,*

$$\beta_i := \lambda_i \ominus 0 \quad (18)$$

satisfies the order balance equations (15,8) for the given ρ and V.

2. *For every given ρ and V, there exists a unique solution λ to the modified balance equations. Moreover, the solutions for different values of ρ are translates of each other.*

3. *Whenever ρ is large enough, there exists a unique solution to the order balance equations (15,8). These unique solutions are all translates of the solutions for the modified balance equations.*

Proof: Suppose that for a fixed ρ and V, there exist *two distinct* solutions β and $\hat{\beta}$ to the order balance equations. Define

$$A := \{k \in X : \hat{\beta}_k \leq \beta_k\}.$$

Then we claim that,

$$\max_{i \in A, j \in A^c} \beta_i \ominus V_{ij} = \max_{i \in A, j \in A^c} \beta_j \ominus V_{ji} = -\infty$$

and

$$\max_{i \in A, j \in A^c} \hat{\beta}_i \ominus V_{ij} = \max_{i \in A, j \in A^c} \hat{\beta}_j \ominus V_{ji} = -\infty.$$

We need only consider the case where $A \neq \phi$ and $A \neq X$ (otherwise the claim is trivially true), and let us suppose to the contrary that both expressions are nonnegative. Then

$$\begin{aligned}
\max_{i\in A, j\in A^c} \hat{\beta}_i \ominus V_{ij} &= \max_{i\in A, j\in A^c} \hat{\beta}_j \ominus V_{ji} > \\
\max_{i\in A, j\in A^c} \beta_j \ominus V_{ji} &= \max_{i\in A, j\in A^c} \beta_i \ominus V_{ij} \geq \\
\max_{i\in A, j\in A^c} \hat{\beta}_i \ominus V_{ij},
\end{aligned}$$

which is a contradiction. The other two cases follow similarly, and so the claim is true. This shows that solutions to the order balance equations do not differ arbitrarily; specifically, all the arcs that separate A from A^c are transient.

Hence in particular, whenever we can show that,

$$\beta_i \ominus V_{ij} \geq 0 \text{ for all } i, j, \text{ with } i \neq j, \text{ and } V_{ij} < +\infty, \tag{19}$$

then there can only exist *one* solution to the order balance equations for the given (ρ, V).

Now we show that this is indeed the case when ρ is large, which will prove the first part of the assertion 3) above. Specifically, suppose now that $\rho \geq 2 \sum_{i,j:V_{ij}<+\infty} V_{ij}$.

Let $i_0 \in X$ be a state with $\beta_{i_0} = \rho$. For arbitrary $s \in X$, let $(i^* = i_0, i_1, \ldots, i_p = s)$ be a path from i^* to s such that $V_{i_{k-1},i_k} < +\infty$ for $k = 1, \ldots, p$ and $i_k \neq i_m$ for $k \neq m$. Let $l(i) = \arg\min_j V_{ij}$. With $A = \{i_k\}$ and applying the Order Balance Theorem 2, it is easy to see that

$$\beta_{i_{k-1}} \ominus V_{i_{k-1},i_k} \leq \beta_{i_k} \ominus V_{i_k,l(i_k)} = \max_{j\neq i_k}(\beta_{i_k} \ominus V_{i_k,j}). \tag{20}$$

To prove that $\beta_s \geq \max_{i,j:V_{ij}} V_{ij} < +\infty$, it is sufficient to show that for $k = 1, \ldots, p$, along the path from i^* to s,

$$\beta_{i_k} \geq \beta_{i_0} - V_{i_0,i_1} + V_{i_1,l(i_1)} - V_{i_1,i_2} + V_{i_2,l(i_2)} - \cdots - V_{i_{k-1},i_k} + V_{i_k,l(i_k)}, \tag{21}$$

since $\beta_{i_0} = \rho \geq 2 \sum_{i,j:V_{ij}<+\infty} V_{ij}$.

We prove (21) by induction. For $k = 1$, from (20) we see that

$$\beta_{i_0} \ominus V_{i_0,i_1} \leq \beta_{i_1} \ominus V_{i_1,l(i_1)}. \tag{22}$$

Clearly, the LHS of (22) is nonnegative, implying that the RHS is also nonnegative. Thus, we can replace "$\ominus$" with "$-$" giving

$$\beta_{i_1} \geq \beta_{i_0} - V_{i_0,i_1} + V_{i_1,l(i_1)}. \tag{23}$$

Now assume (21) holds for $k-1$. From (20) we have

$$\beta_{i_{k-1}} \ominus V_{i_{k-1},i_k} \leq \beta_{i_k} \ominus V_{i_k,l(i_k)}. \tag{24}$$

The LHS of (24) is nonnegative and so

$$\begin{aligned} \beta_{i_k} &\geq \beta_{i_{k-1}} - V_{i_{k-1},i_k} + V_{i_k,l(i_k)} \\ &\geq \beta_{i_0} - V_{i_0,i_1}, + V_{i_1,l(i_1)} - V_{i_1,i_2} + V_{i_2,l(i_2)} - \cdots - V_{i_{k-1},i_k} + V_{i_k,l(i_k)}, \end{aligned}$$

which completes the induction proof. This proves (19), and therefore there exists a unique solution whenever ρ is large enough, which is the first half of assertion 3) above.

Moreover, for the large enough ρ specified earlier, due to (19), we have $\beta_i \ominus V_{ij} = \beta_i - V_{ij}$. Hence $\{\beta_i\}$ itself satisfies the modified balance equations. In fact this solution to the modified balance equations is unique, since if λ is any other solution, then one can prove in a fashion similar to the above, that $\lambda_i \geq V_{ij}$ for all $j \in N_i$, thus yielding that $\lambda_i \ominus V_{ij} = \lambda_i - V_{ij}$, which in turn proves that λ is yet another solution to the order balance equations, which is a contradiction.

Hence, at least for large enough values of ρ we have proved the existence of a unique solution to the modified balance equations. However, it is easy to see that if λ satisfies the modified equations for a given (ρ, V), then $\lambda - \delta$ satisfies the modified balance equations for $(\rho - \delta, V)$, thus proving the existence of a unique solution to the modified balance equations for *all* (ρ, V). This proves the assertion 2) as well as the second half of the assertion 3) above.

Now we turn to the proof of assertion 1) above. Let A be arbitrary, and let $\{\lambda_i\}$ be the solution of the modified balance equations, and define $\beta_i := \lambda_i \ominus 0$. Suppose

$$\max_{i \in A, j \in A^c} \lambda_i - V_{ij} < 0.$$

Then by (16) we also have

$$\max_{i \in A, j \in A^c} \lambda_j - V_{ji} < 0.$$

However, then for each $i \in A$ and $j \in A^c$,

$$\beta_i \leq \lambda_i < V_{ij} \qquad \text{and} \qquad \beta_j \leq \lambda_j < V_{ji}.$$

Hence,

$$\beta_i \ominus V_{ij} = -\infty \qquad \text{and} \qquad \beta_j \ominus V_{ji} = -\infty,$$

and so

$$\max_{i \in A, j \in A^c} \beta_i \ominus V_{ij} = \max_{i \in A, j \in A^c} \beta_j \ominus V_{ji},$$

thus satisfying the original order balance equations. If, however,

$$\max_{i \in A, j \in A^c} \lambda_i - V_{ij} = \delta \geq 0,$$

then by (16)

$$\max_{i \in A, j \in A^c} \lambda_j - V_{ji} = \delta \geq 0.$$

Suppose that $(i_1, j_1) \in A \times A^c$ and $(i_2, j_2) \in A^c \times A$ are such that

$$\lambda_{i_1} - V_{i_1, j_1} = \lambda_{j_2} - V_{j_2, i_2} = \delta.$$

Then since

$$\lambda_{i_1} = V_{i_1, j_1} + \delta \geq 0 \qquad \text{and} \qquad \lambda_{j_2} = V_{j_2, i_2} + \delta \geq 0$$

we have

$$\beta_{i_1} = \lambda_{i_1} \qquad \text{and} \qquad \beta_{j_2} = \lambda_{j_2},$$

and so

$$\beta_{i_1} - V_{i_1, j_1} = \beta_{j_2} - V_{j_2, i_2}.$$

Also, since $\lambda_k \geq \beta_k$, we have

$$\begin{aligned} \max_{i \in A, j \in A^c} \beta_i \ominus V_{i,j} &\leq \max_{i \in A, j \in A^c} \lambda_i \ominus V_{ij} \\ &\leq \max_{i \in A, j \in A^c} \lambda_i - V_{ij} \\ &= \lambda_{i_1} - V_{i_1, j_1} \\ &= \beta_{i_1} - V_{i_1, j_1} \\ &= \beta_{i_1} \ominus V_{i_1, j_1}. \end{aligned}$$

Similarly, $\max_{i \in A, j \in A^c} \beta_j \ominus V_{ji} = \beta_{j_2} \ominus V_{j_2, i_2}$, and so

$$\max_{i \in A, j \in A^c} \beta_i \ominus V_{ij} = \max_{i \in A, j \in A^c} \beta_j \ominus V_{ji}.$$

This proves the assertion 1) and the Theorem. ■

Remark 1 It is interesting to note that the existence of a solution to the modified balance equations has been proved by relying on the existence of a solution to the order balance equations, which in turn is guaranteed by the probabilistic arguments of Theorem 2. A separate independent constructive proof of existence, which does not use probabilistic arguments can be found in [8]. ■

We now give an algorithm for determining the *unique* solution to the modified balance equations. An illustrative example is convenient.

Example 2 Let $\rho = 5$ and

$$V = [V_{ij}] = \begin{pmatrix} \star & 4 & 3 & 1 \\ 6 & \star & 3 & 7 \\ 6 & 2 & \star & 4 \\ 2 & 6 & 5 & \star \end{pmatrix}.$$

Our goal is to determine $\lambda = (\lambda_1, \ldots, \lambda_4)$ which satisfies (16, 17). We shall refer to $\lambda_i - V_{ij}$ as the λ-*flow* along the arc (i, j). Consider first the modified balance equation for the edge cut $A = \{i\}$,

$$\max_{j \neq i} \lambda_i - V_{ij} = \max_{j \neq i} \lambda_j - V_{ji}. \tag{25}$$

Observe that the LHS of (25) can be written as

$$\lambda_i - \min_{j \neq i} V_{ij},$$

and so the arc of *maximum* λ-*flow* out of $A = \{i\}$ is the arc $(i, l(i))$ where

$$l(i) = \arg\min_{j \neq i} V_{ij}.$$

(Note that $l(i)$ may not be unique.)

We now construct the directed graph $G_1 = (V_1, E_1)$, with $V_1 = \{\{1\}, \ldots, \{4\}\}$ and $(i, j) \in E_1$ if $j = l(i)$. See Figure 1.

Note that G_1 has two directed cycles $\{1\} \longrightarrow \{4\} \longrightarrow \{1\}$ and $\{2\} \longrightarrow \{3\} \longrightarrow \{2\}$. Let us examine the λ-flows on the directed cycle $\{1\} \longrightarrow \{4\} \longrightarrow \{1\}$. Since $\lambda_1 - V_{14}$ is the maximum λ-flow out of $\{1\}$, it is not smaller than any λ-flow into $\{1\}$, and so in particular

$$\lambda_1 - V_{14} \geq \lambda_4 - V_{41}.$$

Also, $\lambda_4 - V_{41}$ is the maximum λ-flow out of $\{4\}$ and so

$$\lambda_4 - V_{41} \geq \lambda_1 - V_{14}.$$

We thus observe that the λ-flows along the directed cycle $\{1\} \longrightarrow \{4\} \longrightarrow \{1\}$ are *equal*; that is,

$$\lambda_1 - V_{14} = \lambda_4 - V_{41},$$

and so

$$\lambda_1 - 1 = \lambda_4 - 2. \tag{26}$$

Thus, we have determined the *difference* between λ_1 and λ_4.

In exactly the same way, from the directed cycle $\{2\} \longrightarrow \{3\} \longrightarrow \{2\}$ we see that

$$\lambda_2 - 3 = \lambda_3 - 2. \tag{27}$$

thus determining the difference between λ_2 and λ_3.

At the next step of the algorithm, consider the modified balance equations for the edge cut (A, A^c) where $A = \{1,4\}$ and $A^c = \{2,3\}$. Observe that for $A = \{1,4\}$, the LHS of the modified balance equation

$$\max_{i \in A, j \in A^c} \lambda_i - V_{ij} = \max_{i \in A, j \in A^c} \lambda_j - V_{ji} \tag{28}$$

can be written as

$$\max(\lambda_1 - V_{12}, \lambda_1 - V_{13}, \lambda_4 - V_{42}, \lambda_4 - V_{43});$$

that is,

$$\max(\lambda_1 - 4, \lambda_1 - 3, \lambda_4 - 6, \lambda_4 - 5).$$

We have previously determined that $\lambda_4 - \lambda_1 = 1$, and so the maximum is achieved by $\lambda_1 - V_{13} = \lambda_1 - 3$, and the arc of maximum λ-flow out of $\{1,4\}$ is the arc $(1,3)$.

In a similar fashion, examining the RHS of the modified balance Equation (28), we determine that the maximum λ-flow out of $\{2,3\}$ is achieved by $\lambda_3 - V_{34} = \lambda_3 - 4$, and so the arc of maximum λ-flow out of $\{2,3\}$ is $(3,4)$.

We now consider the directed graph $G_2 = (V_2, E_2)$, with $V_2 = \{\{1,4\},\{2,3\}\}$ and $E_2 = \{(1,3),(3,4)\}$ shown in Figure 2. Note that E_2 is the set of the arcs of maximum λ-flow out of the edge cuts in V_2.

Observe that G_2 has a directed cycle $\{1,4\} \longrightarrow \{2,3\} \longrightarrow \{1,4\}$. Now note that $\lambda_1 - V_{13}$ is the maximum λ-flow out of $\{1,4\}$ and $\lambda_3 - V_{34}$ is the maximum λ-flow out of $\{2,3\}$ and so

$$\lambda_1 - V_{13} = \lambda_3 - V_{34};$$

that is,

$$\lambda_1 - 3 = \lambda_3 - 4. \tag{29}$$

Combining (26, 27, 29) gives

$$\lambda_1 - 3 = \lambda_2 - 5 = \lambda_3 - 4 = \lambda_4 - 4. \tag{30}$$

We now know the *pairwise differences* between *all* of the λ_i's, and so we do not need to consider any additional edge cuts. To fix the values of $\{\lambda_i\}$, we use the value of ρ to give,

$$\max_{i \in X} \lambda_i = \rho = 5$$

Since, from (30), λ_2 is the largest, we set $\lambda_2 = 5$. We thus obtain the solution to the modified balance equations as,

$$\lambda_1 = 3, \qquad \lambda_2 = 5, \qquad \lambda_3 = \lambda_4 = 4.$$

■

The principal idea used to solve the modified balance equations in Example 2 is summarized in the following lemma.

Lemma 2

1. *Given $A \subseteq X$ for which we know the pairwise differences between all the λ_i's for states in A, we can determine the arc of maximum λ-flow out of A (without knowing the λ_i's themselves).*

2. *Let $A_1, A_2, \ldots, A_p$ be a partition of X and suppose for each A_k we know all the pairwise differences between the λ_i's for all states in A_k. Let (i_k, j_k) denote the arc of maximum λ-flow out of A_k. Construct the directed graph $G = (V, E)$, with $V = \{A_1, \ldots, A_p\}$ and $E = \{(i_1, j_1), \ldots, (i_p, j_p)\}$. There exists a directed cycle on G. If $\{A_{n_1}, \ldots, A_{n_m}\}$ is the list of vertices, in order, along the directed cycle, then the λ-flow on the directed cycle is constant; that is,*

$$\lambda_{i_{n_1}} - V_{i_{n_1}, j_{n_1}} = \cdots = \lambda_{i_{n_m}} - V_{i_{n_m}, j_{n_m}},$$

and we can determine the pairwise differences between the values of the λ_i's for all the states in $\bigcup_{k=1}^{m} A_{n_k}$.

Proof:

1. Without loss of generality, suppose A is the set of states $\{1, 2, \ldots, r\}$. Let $\alpha_i := \lambda_1 - \lambda_i$. (We know the α_i's.) Then

$$\max_{i \in A, j \in A^c} \lambda_i - V_{ij} = \max_{i \in A, j \in A^c} \lambda_1 - \alpha_i - V_{ij} = \lambda_1 - \min_{i \in A, j \in A^c} (\alpha_i + V_{ij}).$$

   Thus, the arc

$$(i^*, j^*) := \arg \min_{i \in A, j \in A^c} (\alpha_i + V_{ij})$$

   is an arc of maximum λ-flow out of A.

2. The out-degree of each vertex of G is at least one, and so from elementary graph theory it follows that G has a directed cycle. Suppose

$$A_{n_1} \longrightarrow A_{n_2} \longrightarrow \cdots \longrightarrow A_{n_m} \longrightarrow A_{n_1}$$

   is such a directed cycle. Then we have the situation shown in Figure 3.

   Now (i_{n_k}, j_{n_k}) is the arc of maximum λ-flow *out* of A_{n_k}, and so the λ-flow on this arc is not less than the λ-flow of any arc *into* A_{n_k}. In particular,

$$\lambda_{i_{n_k}} - V_{i_{n_k}, j_{n_k}} \geq \lambda_{i_{n_{k-1}}, j_{n_{k-1}}} - V_{i_{n_{k-1}}, j_{n_{k-1}}}, \qquad \text{for } k = 1, \ldots, m,$$

   where, for convenience, we implicitly identify i_{n_0} with i_{n_m} and j_{n_0} with j_{n_m}. Thus,

$$\begin{aligned} \lambda_{i_{n_m}} - V_{i_{n_m}, j_{n_m}} &\geq \lambda_{i_{n_{m-1}}} - V_{i_{n_{m-1}}, j_{n_{m-1}}} \\ &\geq \lambda_{i_{n_{m-2}}} - V_{i_{n_{m-2}}, j_{n_{m-2}}} \\ &\vdots \\ &\geq \lambda_{i_{n_1}} - V_{i_{n_1}, j_{n_1}} \\ &\geq \lambda_{i_{n_m}} - V_{i_{n_m}, j_{n_m}}. \end{aligned}$$

   Therefore, the λ-flow on the directed cycle is a constant,

$$\lambda_{i_{n_1}} - V_{i_{n_1}, j_{n_1}} = \lambda_{i_{n_2}} - V_{i_{n_2}, j_{n_2}} = \cdots = \lambda_{i_{n_m}} - V_{i_{n_m}, j_{n_m}}. \qquad (31)$$

   For each A_{n_k} in the directed cycle, we know the pairwise differences between the λ_i's for states in A_{n_k}. Using (31) we can now easily determine the pairwise differences between all the λ_i's for states in $\bigcup_{k=1}^m A_{n_k}$. ■

The algorithm for solving the modified balance equations is outlined below.

Algorithm to solve modified balance equations

Step 1: Set $A_i^1 = \{i\}$ for $i = 1, \dots, |X|$. We call the A_i^k's *coalitions at step* k. Note that for every i, the pairwise differences between the λ-values for all states in A_i^1 are (trivially) known. Set $A^1 := \{A_1^1, A_2^1, \dots, A_{|X|}^1\}$. Let $N(1) = |A^1| =:$ the number of elements in the set $A^1 =$ number of coalitions at Step 1.

Step k: Given $A^k := \{A_1^k, A_2^k, \dots, A_{N(k)}^k\}$, where for each $A_j^k \in A^k$ the pairwise differences between all of the λ_i's for i's in A_j^k are known, construct A^{k+1} as follows. Using Lemma 2, identify the directed cycles in the graph. (There exists at least one directed cycle.) The elements of A^{k+1} consist of the directed cycles identified in the graph, and those $A_j^k \in A^k$ which are not in any directed cycle. (More precisely, if $\{A_{n_1}^k, A_{n_2}^k, \dots, A_{n_m}^k\}$ is a directed cycle, then $\bigcup_{i=1}^m A_{n_i}$ is an element of A^{k+1}). Note that for every $A_j^{k+1} \in A^{k+1}$, the pairwise differences between all of the λ_i's for i's in A_j^{k+1} are known. Furthermore, if $N(k) := |A^k|$, then $N(k+1) < N(k)$.

Last Step: Stop when $N(k) = 1$. Note that the pairwise differences between *all* λ_i's are known, and the λ satisfying the modified balance equations can be obtained by a translation by using the given value of ρ. ∎

4 AN ALGORITHM TO OBTAIN ALL SOLUTIONS OF THE ORDER BALANCE EQUATIONS

We now characterize *all* solutions to the order balance equations, and describe an algorithm for generating *all* these solutions. To do so we will use the coalitions $\{A_i^k\}$ generated by the algorithm of the preceding section. Let us call $\lambda_i - V_{ij}$ and $\beta_{ij} = \beta_i \ominus V_{ij}$ as the λ-*flow* and β-*flow*, respectively, along the arc (i, j).

Lemma 3 *1. If (i, j) is an arc of maximum λ-flow out of A_l^k, then it is also an arc of maximum β-flow out A_l^k.*

2. *If $\{A_1^k, \ldots, A_p^k\}$ is a directed cycle obtained at step k, then the β-flow along the directed cycle is a constant.*

3. *If the β-flow along the directed cycle $\{A_1^k, \ldots, A_p^k\}$ obtained at step k is $-\infty$, then the β-flow along any directed cycle obtained at step $n > k$ containing $A_l^n = \bigcup_{i=1}^p A_i^k$ as a node, is also $-\infty$.*

4. *If the β-flow along the directed cycle $\{A_1^k, \ldots, A_p^k\}$ obtained at step k is ≥ 0, then for every $i, j \in A_l^{k+1} := \bigcup_{m=1}^p A_m^k$ there exists a path $(i = i_0, i_1, \ldots, i_q = j)$ such that $i_m \in A_l^{k+1}$ and $\beta_{i_m, i_{m+1}} \geq 0$ for $0 \leq m \leq q-1$.*

Proof: We will first prove 1), 2) and 3) by induction. Consider $k = 1$. Since A_l^k is then just a singleton, say $A_l^k = \{l\}$, an arc (l, m) of maximum λ-flow out of $\{l\}$ is just one for which $V_{lm} = \min_n V_{ln}$. Clearly this is also an arc for which $\beta_l \ominus V_{lm} = \min_n \beta_l \ominus V_{ln}$. Now suppose that $\{A_1^k, \ldots, A_p^k\}$ is a directed cycle of such maximum flows. Then an application of the Order Balance Theorem to each A_l^k shows that $\beta_{12} = \beta_{23} = \ldots = \beta_{p1}$. Suppose now that $\beta_{12} = \beta_{23} = \ldots = \beta_{p1} = -\infty$. Then if (l, m) is an arc of maximum β-flow out of $\bigcup_{i=1}^p A_i^k$, clearly $\beta_{lm} \leq \beta_{l,l+1} = -\infty$. Thus the assertion is true for $k = 1$.

Now suppose that the assertion is true for $1, 2, \ldots, k-1$. Consider a coalition A_l^k. If the β-flow along some directed cycle $\{A_1^n, \ldots, A_q^n\}$ at some step $n < k$ with $A_l^k = \bigcup_{i=1}^q A_i^n$ was $-\infty$, then clearly the maximum β-flow out of A_l^k is $-\infty$, and so *any* arc out of A_l^k is an arc of maximum β-flow. On the other hand if the β-flow along the directed cycle $\{A_1^n, \ldots, A_q^n\}$ is ≥ 0, then the differences between the β_i's for states $i \in A_l^k$ are the *same* as the differences between the λ_i's, i.e.

$$\beta_i - \beta_j = \lambda_i - \lambda_j \text{ for all } i, j \in A_l^k, \tag{32}$$

and so the arc of maximum λ-flow out of from A_l^k is also an arc of maximum β-flow out from A_l^k. Moreover, if $\{A_1^k, \ldots, A_p^k\}$ is a directed cycle at step k, then an application of the Order Balance Theorem to each A_l^k shows that the β-flow along the directed cycle is a constant. Finally, if this β-flow is $-\infty$, suppose that (r, m) is a maximum flow arc out of $\bigcup_{l=1}^p A_l^k$. Suppose that $r \in A_l^k$. Then clearly $\max_{i \in A_l^k, j \in A_l^{k^c}} \beta_{ij} \geq \beta_{rm}$ and so $\beta_{rm} = -\infty$. This completes the induction and the proof.

Finally, to see 4), note first that from 1), 2) and 3), the β-flow along any directed cycles contained within A_l^{k+1} is ≥ 0. Since A_l^{k+1} is formed as the union of such directed cycles, the result follows. ∎

Motivated by 3) and 4) above, we introduce the following definition.

Definition 4 *We shall say that i is* recurrently connected *to j if there exists a path $(i = i_0, i_1, \ldots, i_q = j)$ with $\beta_{i_m, i_{m+1}} \geq 0$ for $0 \leq m \leq q-1$.*

We shall say that a set $A \subseteq X$ is a recurrently connected set *if for every $i, j \in A$ and $k \in A^c$, i is recurrently connected to j but not to k.*

From Lemma 3 it follows that recurrently connected sets are precisely those A_l^k's for which the β-flow out of A_l^k is $-\infty$, while the β-flows along the directed cycles contained within A_l^k are ≥ 0. Note also that the recurrently connected sets form a *partition* of X.

We now proceed to determine which sets are possible candidates for being recurrently connected sets. Consider a typical candidate A_l^{k+1}. Let $\mathcal{F}$ denote the β-flow on the cycle $\{A_1^k, \ldots, A_p^k\}$, where $A_l^{k+1} = \bigcup_{i=1}^p A_i^k$. Then if (i_m, j_m) is the arc of maximum flow out of A_m^k (and, by construction, into $A^k_{(m+1)\mathrm{mod}_p}$), we must have,

$$\mathcal{F} = \beta_{i_1} - V_{i_1,j_1} = \beta_{i_2} - V_{i_2,j_2} = \cdots = \beta_{i_p} - V_{i_p,j_p} \geq 0, \quad \max_{i \in A_l^{k+1}, j \notin A_l^{k+1}} \beta_i - V_{ij} < 0, \text{ and } \max_{i \in A_l^{k+1}} \beta_i \leq \rho.$$

We will now attempt to determine whether there exist $\{\beta_i : i \in A_l^{k+1}\}$ which satisfy these conditions. Note that if this is not feasible, then A_l^{k+1} cannot be a recurrently connected set.

Let (x, y) denote the arc of maximum β-flow out of A_l^{k+1}. Then $\beta_x < V_{xy}$. Fix m to be an arbitrarily chosen state from A_l^{k+1}. Then for every state $h \in A_l^{k+1}$ we know the value of $(\beta_h - \beta_m)$ from Lemma 3 above. Let us define

$$\zeta_h := \beta_h - \beta_m.$$

Then

$$\begin{aligned}
\mathcal{F} &= \beta_{i_1} - V_{i_1,j_1} \\
&= \beta_m + \zeta_{i_1} - V_{i_1,j_1} \\
&= \beta_x - \zeta_x + \zeta_{i_1} - V_{i_1,j_1} \\
&< V_{xy} - \zeta_x + \zeta_{i_1} - V_{i_1,j_1} =: M_1,
\end{aligned}$$

giving an upper bound on $\mathcal{F}$.

We must also satisfy the constraint $\max_{i \in A_l^{k+1}} \beta_i \leq \rho$, and so let

$$\theta := \arg \max_{i \in A_l^{k+1}} \zeta_i.$$

Then it is clear that $\beta_\theta \geq \max_{i \in A_l^{k+1}} \beta_i$. Thus,

$$\begin{aligned}
\rho &\geq \beta_\theta \\
&= \beta_m + \zeta_\theta \\
&= \beta_{i_1} - \zeta_{i_1} + \zeta_\theta \\
&= \beta_{i_1} - V_{i_1,j_1} + V_{i_1,j_1} - \zeta_{i_1} + \zeta_\theta \\
&= \mathcal{F} + V_{i_1,j_1} - \zeta_{i_1} + \zeta_\theta,
\end{aligned}$$

and so

$$\mathcal{F} \leq \rho - V_{i_1,j_1} + \zeta_{i_1} - \zeta_\theta =: M_2.$$

giving yet another upper bound on $\mathcal{F}$. (Note: if $A_l^{k+1} = \{i\}$, then $M_1 = \min_j V_{ij}$ and $M_2 = \rho$.)

Any choice of $\mathcal{F}$ from the interval

$$\Omega(A_l^{k+1}) := [0, M_1) \bigcap [0, M_2]$$

will allow assignments for the recurrence orders of states in A_l^{k+1} consistent with the assumption that the coalition A_l^{k+1} is a recurrently connected set, if A_l^{k+1} is not a singleton set. If $A_l^{k+1} = \{i\}$ is a singleton set, then we define $\Omega(\{i\}) := \{-\infty\} \bigcup ([0, M_1) \bigcap [0, M_2])$. Note also that $\Omega(X) = [0, M_2]$, since the M_1 upperbound is $+\infty$ due to there being no maximal flow out of X. If $\Omega(A_l^{k+1}) = \phi$ then there is no assignment, and so A_l^{k+1} is not a recurrently connected set.

We still need to determine the set of all recurrently connected sets. To do this we construct a *rooted tree* having the coalitions produced by the general procedure as nodes, and having a directed edge from coalition A_p^{k+1} to A_r^k if $A_p^{k+1} \supseteq A_r^{(k)}$. Hence, the root of the tree is X, and its leaves are the singleton sets $\{1\}, \{2\}, \ldots, \{n\}$. Let D_i be the set of the leaves of the tree which are *descendants* of the node i in the rooted tree.

We say that a set Ξ of nodes is a *proper cover* if

$$\bigcup_{A \in \Xi} D_A = X$$

and

$$D_A \bigcap D_{A'} = \phi \qquad \text{for } A \neq A'.$$

Now the algorithm to determine all the solutions of (15,8) proceeds as follows. Let a set $\Xi := \{A_1, A_2, \ldots, A_k\}$ be a proper cover. Now we will determine whether Ξ can be a set of *all* recurrently connected sets, as follows. First we determine $\Omega(A_j)$ for every $A_j \in \Xi$. If *any* of the $\Omega(A_j)$'s is empty, then the guess Ξ is *not* a feasible set of recurrently connected sets. If *every* $\Omega(A_j)$ is non-empty, then let $\overline{\mathcal{F}}_j := \sup \Omega(A_j)$. If this "sup" is not attained, then we cannot assign ρ to any state in A_j. If this "sup" is attained, then we determine for each such A_j whether with the choice of $\overline{\mathcal{F}}_j$ there is a state $i_j \in A_j$ with $\beta_{i_j} = \rho$. If no such state exists for *any* A_j, then again Ξ is *not* a feasible set of recurrently connected sets. Finally, if there exist such A_j's, then let $\mathcal{A}(\Xi)$ be the set of all such A_j's. Now, the set of all solutions corresponding to Ξ is obtained by picking, in turn, an A_j from $\mathcal{A}(\Xi)$, fixing its flow as $\overline{\mathcal{F}}$, and choosing all other $\mathcal{F}_j$'s arbitrarily from the $\Omega(A_j)$'s. By checking *every* proper cover Ξ, we thus determine all solutions to the order balance equations, as the following Theorem shows.

Theorem 4 *All solutions to the order balance equations can be generated by using the method described above.*

Proof: Suppose β satisfies the order balance equations. Then for this solution determine the set Ξ of recurrently connected sets. This set must be a proper cover. For this set Ξ, there must be some A_j with corresponding β-flow equal $\overline{\mathcal{F}}_j$. Now determine the β-flows on the recurrently connected sets. We generate this solution β when we choose Ξ as the set of recurrently connected sets, and A_j as the coalition with maximum flow equal to $\overline{\mathcal{F}}_j$, and assign the correct β-flows on the other recurrently connected sets. ■

This algorithm takes an exponential in $|X|$ number of steps, due to the necessity of checking all proper covers. However, the complexity issue is not the paramount concern here, since the problem of asymptotic analysis of the time-inhomogeneous stochastic process is not a-priori known to be a problem resolvable by a finite algorithm.

We illustrate the procedure for determining all solutions to the order balance equations.

Example 3 We construct all solutions to the order balance equations for Example 2 when $\rho = 4$. See Figure 4 for the rooted tree. We check the proper covers:

1. $\Xi = \{X\}$: $\Omega(X)$ is empty, so X cannot be a recurrently connected set.
2. $\Xi = \{\{1,4\},\{2,3\}\}$: Using the method described above we obtain

$$\beta_1 = \alpha, \qquad \beta_2 = 4, \qquad \beta_3 = 3, \qquad \beta_4 = 1 + \alpha,$$

where $1 \le \alpha < 3$.

3. $\Xi = \{\{1,4\},\{2\},\{3\}\}$: $\max_{i \in X} \beta_i < 4$, a contradiction.

4. $\Xi = \{\{1\},\{4\},\{2,3\}\}$:

$$\beta_1 = \gamma, \qquad \beta_2 = 4, \qquad \beta_3 = 3, \qquad \beta_4 = \theta,$$

where $\gamma = -\infty$ or $0 \le \gamma < 1$, and $\theta = -\infty$ or $0 \le \theta < 2$.

5. $\Xi = \{\{1\},\{2\},\{3\},\{4\}\}$: $\max_{i \in X} \beta_i < 4$, and so $\{\{1\},\{2\},\{3\},\{4\}\}$ is not a set of recurrently connected sets.

We have checked all proper covers. Hence the set of all solutions is $\{(\alpha, 4, 3, 1+\alpha) : 1 \le \alpha < 3\} \cup \{(\gamma, 4, 3, \theta) : \gamma = -\infty$ or $0 \le \gamma < 1$ and $\theta = -\infty$ or $0 \le \theta < 2\}$. ■

How can non-unique solutions to the order balance equations arise, and what is the implication of such non-uniqueness? First let us consider the case where a unique solution exists. Since such a solution is uniquely determined by the algorithm, it is clear that the recurrence orders of the states, and thus the rates of convergence of the transition probabilities, depend only on the V_{ij}'s in the transition probabilities $p_{ij}(t) = c_{ij}\epsilon(t)^{V_{ij}}$, and *not* on the proportionality constants $\{c_{ij}\}$. However, in the case of *non-unique* solutions, the following example shows that the recurrence orders may even depend on the the *proportionality constants* $\{c_{ij}\}$.

Example 4 Let $X = \{1, 2, 3\}$ and $V_{ij} = \max\{0, j - i\}$. Let $c_{13} = c_{23} = 1$, $c_{31} = 1 - \alpha$ and $c_{32} = \alpha$, where $\alpha \in (0, 1)$. Set $c_{ij} = 0$ for all other i, j. See Figure 5. Let the cooling schedule be $\epsilon(t) = 1/t$. Then the *complete* set of order balance equations obtained by using *all* edge cuts is,

$$\begin{aligned} \beta_2 \ominus V_{23} &= \beta_3 \ominus V_{32}, \\ \beta_3 \ominus V_{31} &= \beta_1 \ominus V_{13}, \\ \max(\beta_2 \ominus V_{23}, \beta_1 \ominus V_{13}) &= \max(\beta_3 \ominus V_{32}, \beta_3 \ominus V_{31}), \end{aligned}$$

with the maximum given by,

$$\max_{i \in X} \beta_i = 1.$$

The assignments

$$\beta_1 = 1, \quad \beta_2 = \gamma, \quad \beta_3 = -\infty$$

satisfy the order balance equations for *every* $\gamma \in \{-\infty\} \cup [0, 1)$. Thus any value of $\beta_2 < 1$ gives a solution of the order balance equations.

However, a calculation, which can be found in [8], shows that the correct order of recurrence of state 2 is

$$\beta_2 = \alpha.$$

Thus, the order of recurrence, and the rate of convergence of the probability $\Pr(x(t) = 2)$ to 0, depends on the proportionality constant $c_{32} = \alpha$ involved. ■

Based on the above results, we obtain the following property of the orders of recurrence of the states in a recurrently connected set.

Lemma 4 *Consider a recurrently connected set A.*

1. *If $\beta_i \in \mathcal{R}$ for some $i \in A$, then $\beta_j \in \mathcal{R}$ for* all *$j \in A$.*
2. *If for some $i \in A$, $\beta_i = p_i^-$ for some $p_i \in \mathcal{R}$, then for every $j \in A$, $\beta_j = p_j^-$ for some $p_j \in \mathcal{R}$.*

Proof: This follows immediately from (32). ■

Thus all recurrence orders in a recurrently connected set are of the same type, i.e. either they are all real numbers p_i, or they are all of the type p_i^-, or they are all $-\infty$ (see Definition 1).

This gives us the following Lemma which completes the proof of Theorem 1.

Lemma 5 *Suppose the rate of cooling is $\rho = p \in \mathcal{R}$, with $p > 0$, i.e. the maximum is achieved in Definition 3. If there is a state $i \in X$ for which $\beta_i = p^-$, then $\lim_{t\to\infty} \Pr(x(t) = i) = 0$.*

Proof: Suppose A is the recurrently connected class to which i belongs. Since all arcs *between* recurrently connected sets are transient, it follows from the Borel-Cantelli Lemma that along almost every sample path ω there can only be a *finite* number of transitions between different recurrently connected sets. Hence for almost every ω, $\{x(t, \omega)\}$ converges to some recurrently connected set. Hence the limit $\lim_{t\to\infty} \Pr(x(t) \in A)$ exists. Now we show that this limit is 0. Suppose not, i.e. suppose $\lim_{t\to\infty} \sum_{j\in A} \pi_j(t) = \delta > 0$. Then it follows that $\sum_{t=0}^{\infty} \epsilon(t)^p \sum_{j\in A} \pi_j(t) = +\infty$. Hence for some $j \in A$, $\beta_j = p$. But then by Lemma 4, $\beta_i \in \mathcal{R}$, which gives a contradiction. ■

5 WEAK REVERSIBILITY AND SIMULATED ANNEALING

We now turn our attention to the special class of Markov chains arising from the method of *optimization by simulated annealing.* Recall that the Markov chains in this class satisfy (1–6) with the special choice of

$$V_{ij} := \max\{0, W_j - W_i\}.$$

In [7] it was shown that under the "symmetric neighborhood" assumption, $c_{ij} > 0$ if and only if $c_{ji} > 0$, the orders of recurrence satisfy the following *detailed order balance*,

$$\beta_{ij} = \beta_{ji} \qquad \text{for every } i, j \in X.$$

It is easy to see that the detailed order balance above is equivalent to the sum of the order of recurrence of a state and its cost being constant on recurrently connected sets.

In this section we will show that this constancy property of the sum of the recurrence order and cost on recurrently connected sets continues to hold under the much weaker assumption of "weak reversibility" introduced by Hajek in [1].

Definition 5 *A state i is said to be reachable from state j if there is a sequence of states $j = i_0, i_1, \ldots, i_p = i$ such that $c_{i_k, i_{k+1}} > 0$ for $0 \leq k \leq p-1$.*

Definition 6 *A state i is reachable at height H from j if there is a path from j to i as in* Definition 5 *for which $W_{i_k} \leq H$ for $0 \leq k \leq p$.*

Assumption 1 (Weak Reversibility) *For any real number H and any two states i and j, i is reachable at height H from j if and only if j is reachable at height H from i.*

In what follows we assume weak reversibility.

Theorem 5 (The Potential Theorem) *Under* Assumption 1, *for every recurrently connected set A there exists a constant $\alpha(A)$ such that $\beta_i + W_i = \alpha(A)$ for every $i \in A$.*

Proof: Fix attention on a particular recurrently connected set A. Assume to the contrary that A can be partitioned into *equipotential* sets $C_1, C_2, \ldots, C_r$ such that $\beta_i + W_i = \alpha(C_k)$ for every $i \in C_k$, where the $\alpha(C_k)$'s are distinct constants. We will show that there is only one equipotential set, namely A.

For each equipotential set C_i, determine an arc of maximum β-flow out of the set. From Lemma 2, there exists a directed cycle of these equipotential sets, and the β-flow along the directed cycle is constant. Moreover, from Lemma 3, since A is a recurrently connected set, these β-flows are all non-negative. Without loss of generality, label the sets along the directed cycle $C_1, C_2, \ldots, C_p$ such that the constant $\alpha(C_1)$ associated with the set C_1 is smallest. Let (i_s, j_s) be the arc of maximum β-flow out of the set C_s. By construction, $i_s \in C_s$ and $j_s \in C_{(1+s) \bmod p}$ and

$$\beta_{i_1,j_1} = \beta_{i_2,j_2} = \ldots = \beta_{i_p,j_p} \geq 0.$$

Knowing that $\beta_{i_1,j_1} \geq 0$ we consider the two cases (1) $W_{j_1} \geq W_{i_1}$, or (2) $W_{j_1} < W_{i_1}$.

If case (1) is true then since j_1 is reachable at height W_{j_1} from i, by the weak reversibility assumption there exists a path from j_1 back to i_1 which does not go through any states with costs larger than W_{j_1}. Let (k, l) be the particular arc of that path which exits C_2. Note that

$$\begin{aligned} \beta_{i_1,j_1} &= \beta_{i_2,j_2} \\ &\geq \beta_{kl}, \end{aligned}$$

because β_{i_2,j_2} is the arc of maximum β-flow out of C_2. If $\beta_{kl} \geq 0$ then $\beta_{kl} = \beta_k + W_k - W_l$. If $\beta_{kl} < 0$ then $\beta_k + W_k - W_l < 0$. In either case, since $\beta_{i_1,j_1} \geq 0$, we have that

$$\beta_{i_1,j_1} = \beta_{i_1} + W_{i_1} - W_{j_1} \geq \beta_k + W_k - W_l.$$

Now by the weak reversibility assumption, $W_{j_1} \geq W_l$, and so

$$\beta_{i_1} + W_{i_1} \geq \beta_k + W_k;$$

that is,

$$\alpha(C_1) \geq \alpha(C_2),$$

which is a contradiction.

If case (2) is true, then there is a path from j_1 to i_1 which does not pass through any states with costs larger than i_1. Again, identify the particular arc of that path which exits C_2 as (k, l). Note that

$$\begin{aligned} \beta_{i_1} &= \beta_{i_1,j_1} \\ &= \beta_{i_2,j_2} \\ &\geq \beta_{kl}. \end{aligned}$$

Using similar arguments as in case (1), since $\beta_{i_1} \geq 0$ we have $\beta_{i_1} \geq \beta_k + W_k - W_l$. Now by the weak reversibility assumption, $W_{i_1} \geq W_l$, and so $\alpha(C_1) \geq \alpha(C_2)$, which is again a contradiction.

Hence there is only one equipotential set, A. ■

Since $W_i + \beta_i = \alpha(A)$ for all $i \in A$, where A is a recurrently connected set, we obtain the following necessary and sufficient condition for simulated annealing to hit a *global minimum* with probability one from all states $i \in X$.

Let $M := \{i \in X : W_i \leq W_j \text{ for all } j \in X\}$ be the set of global minima. We now have the following definition due to [1].

Definition 7 *Let d^* be the* smallest *number with the property that for every $i \in X$ there exists a path $(i = i_0, \ldots, i_p)$ with $c_{i_k,i_{k+1}} > 0$ for $0 \leq k \leq p$ and ending in a minimizer $i_p \in M$ such that*

$$W_{i_k} - W_i \leq d^* \text{ for } k = 1, \ldots, p.$$

We shall call d^* the depth *of the minimization problem.*

Theorem 6 (Necessary and Sufficient Condition to Hit Global Minimum With Probabilit *Suppose that weak reversibility holds.*

1. *If $\sum_{t=1}^{\infty} \epsilon(t)^{d^*} = +\infty$, then for every initial condition $x(0) \in X$,*

$$\limsup_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} \Pr(x(t) \in M) = 1,$$

   and the global minimum is hit with probability one.

2. *If $\sum_{t=1}^{\infty} \epsilon(t)^{d^*} < +\infty$, then there exists an initial condition $x(0) \in X$ for which,*

$$\Pr(x(t) \in M^c \text{ for all } t \geq 1) > 0.$$

Proof: The proof is the same as in Theorem (4.6) in [7] except that if $(i = i_0, \ldots, i_p = j)$ is a path from i to j with $c_{i_k, i_{k+1}} > 0$ and $W_{i_k} - W_i \le \gamma$ for $1 \le k \le p$, then instead of using the reversed path $(j = i_p, \ldots, i_0 = i)$ given by the assumption of symmetric neighborhoods, one uses the path $(j = l_0, \ldots, l_q = i)$ with $c_{l_k, l_{k+1}} > 0$ and $W_{l_k} - W_i \le \gamma$ for $1 \le k \le q$, guaranteed by the weak reversibility assumption. ■

The same condition $\sum_{t=1}^{\infty} \epsilon(t)^{d^*} = \infty$ has been earlier shown by Hajek [1] to be necessary and sufficient for $\lim_{t \to \infty} \Pr(x(t) \in M) = 1$, i.e. for convergence *in probability*. Thus while result 1) above is weaker than his since it involves Cesaro as opposed to regular convergence, the result 2) is a stronger sample path result.

The above result has been proved earlier in [7] under the stronger assumption of *symmetric neighborhoods*, $c_{ij} > 0 \Longleftrightarrow c_{ji} > 0$. Moreover, under this assumption [7] have proved a *detailed balance* result which we can obtain as a corollary of Theorem 5, as we show below.

Corollary 1 (Detailed Balance) *Under the symmetric neighborhood assumption,*

$$\beta_{ij} = \beta_{ji} \textit{ for every } i, j \in X.$$

Proof: If i and j are not neighbors then $\beta_{ij} = \beta_{ji} = -\infty$.

If i and j are neighbors and $i \in R$ and $j \in T$, where R is the set of recurrent states and T is the set of transient states, then

$$\beta_{jk} = -\infty \qquad \text{for all } k$$

and so

$$-\infty = \max_{k \ne j} \beta_{jk} = \max_{k \ne j} \beta_{kj} \ge \beta_{ij}$$

showing that $\beta_{ij} = \beta_{ji} = -\infty$. A similar argument holds if $i \in T$ and $j \in R$.

Finally, if i and j are neighbors and $i, j \in R$, without loss of generality let us assume that $W_i \ge W_j$. Then $\beta_{ij} = \beta_i \ge 0$, and so i and j belong to a common recurrently connected set. Hence by Theorem 5, $\beta_i + W_i = \beta_j + W_j$. Since $\beta_{ij} = \beta_i$ and $\beta_{ji} = \beta_j + W_j - W_i$, it follows that $\beta_{ij} = \beta_{ji}$. ■

Note that by the above results, if the order of recurrence of even *one* state in a connected component is known, then the orders of recurrence for *all* the states belonging to the connected component are determined. However, as Example 4 shows, it is not always possible to determine the order of recurrence of even one state in a connected component from the order balance equations alone. In that example, the connected components of recurrent states are the sets $\{1\}$ and $\{2\}$, and the detailed balance equations do not determine the order of recurrence β_2 of the *single* state in the connected component $\{2\}$. The reason for this inadequacy, as mentioned earlier in Example 4, is that the orders of recurrence do depend on the the proportionality constants c_{ij} involved in the transition probabilities. In any case, the β-flows do satisfy Corollary 1.

6 Conclusions

The notion of order of recurrence provides a novel approach for analyzing the class of Markov chains whose transition probabilities are proportional to powers of a time-varying parameter $\epsilon(t)$. These recurrence orders satisfy a set of balance equations, and the Markov chain converges in a Cesaro sense to the set of states with the largest recurrence orders. We have given an algorithm for generating a solution to the order balance equations and have also provided a method for characterizing all solutions to these equations. The algebraic methods presented in this paper for solving the order balance equations are not always sufficient for determining the recurrence orders. In some situations where non-unique solutions exist, the orders of recurrence can depend on the proportionality constants involved in the transition probabilities, and not just on their orders of magnitude. This problem remains an open issue. The method of optimization by simulated annealing falls within the framework of this class of Markov chains. We have shown that if the Markov process is weakly reversible, then the sum of the recurrence order and the cost is a constant on each set of states connected by recurrent arcs. This allows us to determine the necessary and sufficient on the cooling rate for the optimization algorithm to hit a global minimum with probability one from all initial states.

References

[1] B. Hajek, "Cooling schedules for optimal annealing," *Mathematics of Operations Research*, vol. 13, pp. 311–329, 1988.

[2] J. N. Tsitsiklis, "Markov chains with rare transitions and simulated annealing," preprint, Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, August 1985. rev. November 1985.

[3] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, pp. 721–741, November 1984.

[4] D. Mitra, F. Romeo, and A. Sangiovanni-Vincentelli, "Convergence and finite-time behavior of simulated annealing," *Advances in Applied Probability*, vol. 18, pp. 747–771, 1986.

[5] B. Gidas, "Non-stationary Markov chains and convergence of the annealing algorithm," *Journal of Statistical Physics*, vol. 39, pp. 73–131, 1985.

[6] K. L. Chung, *A Course in Probability Theory.* Orlando, Florida: Academic Press, 1974.

[7] D. P. Connors and P. R. Kumar, "Balance of recurrence order in time-inhomogeneous Markov chains with application to simulated annealing," *Probability in the Engineering and Informational Sciences*, vol. 2, pp. 157–184, 1988.

[8] D. P. Connors, *Balance of Recurrence Order in Time-Inhomogeneous Markov Chains with Application to Simulated Annealing.* PhD thesis, University of Illinois, Urbana, Illinois, 1988.

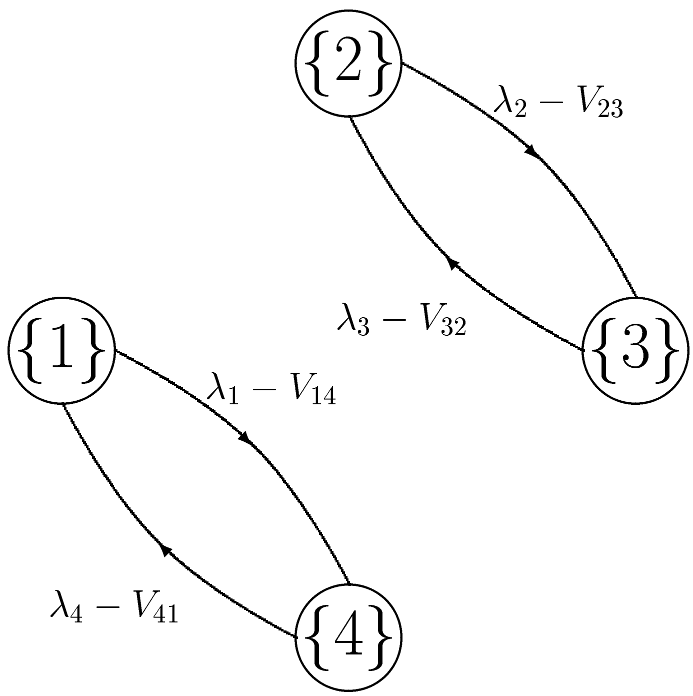

Figure 1: The graph G_1 of Example 2.

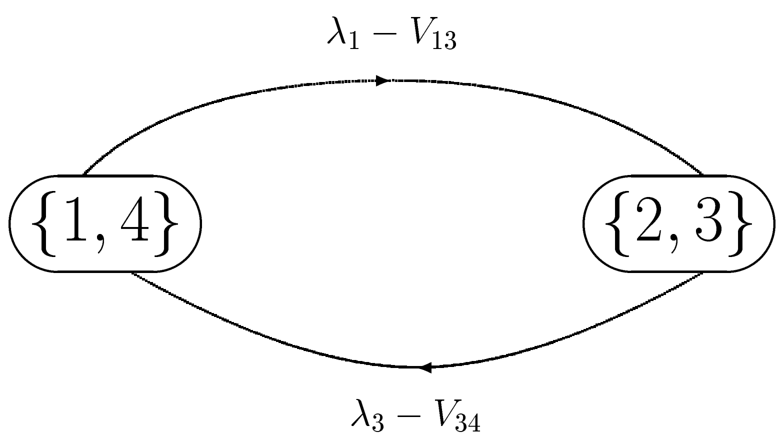

Figure 2: The graph G_2 of Example 2.

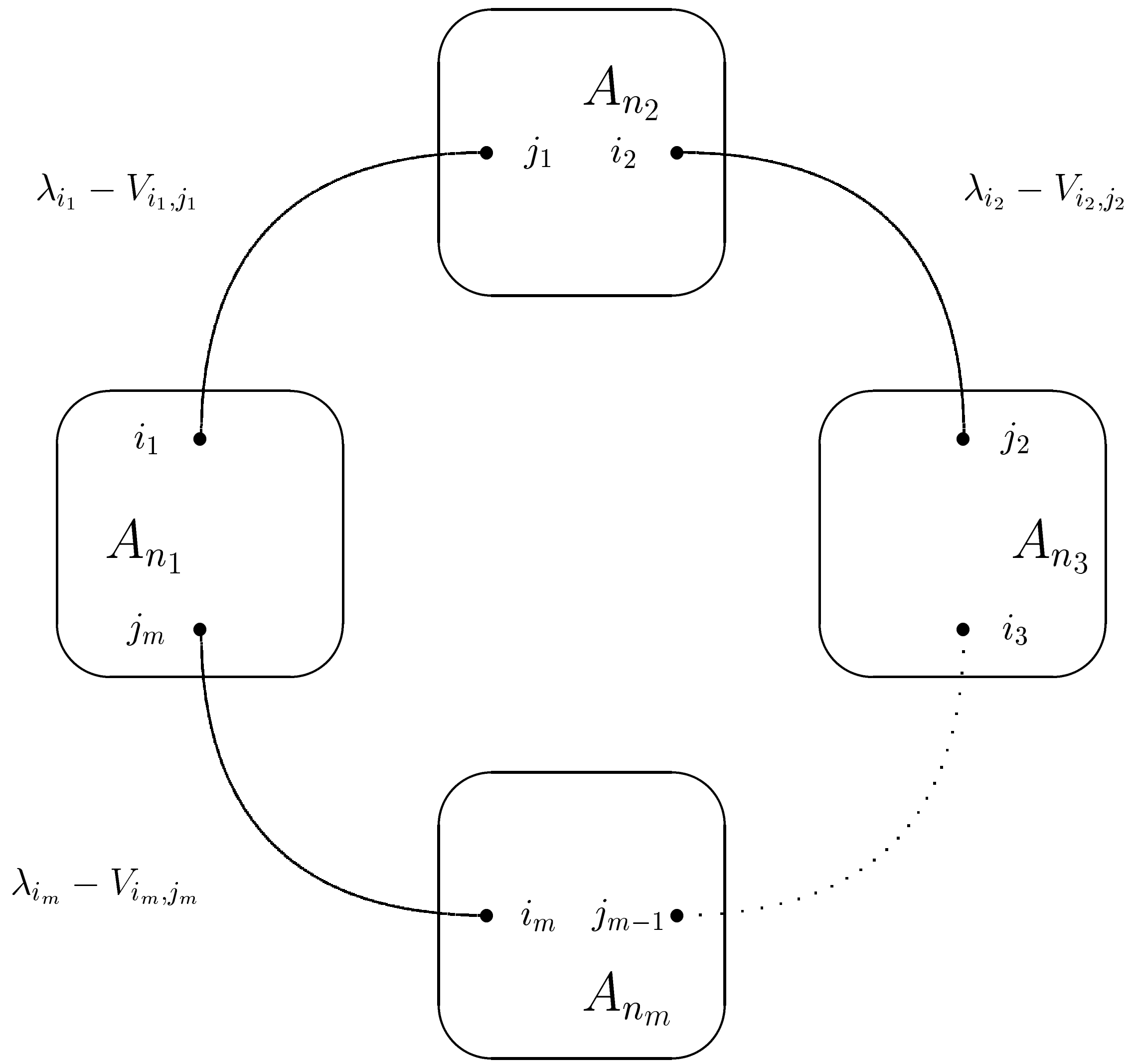

Figure 3: A directed cycle of maximum λ-flows in Lemma **??**

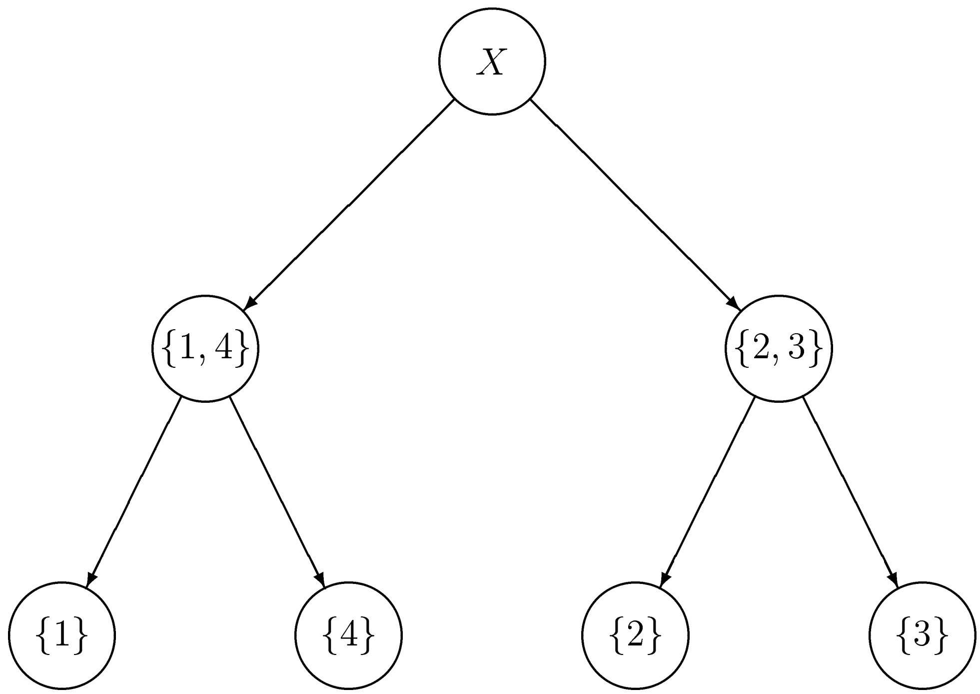

Figure 4: The rooted tree of Example 3.

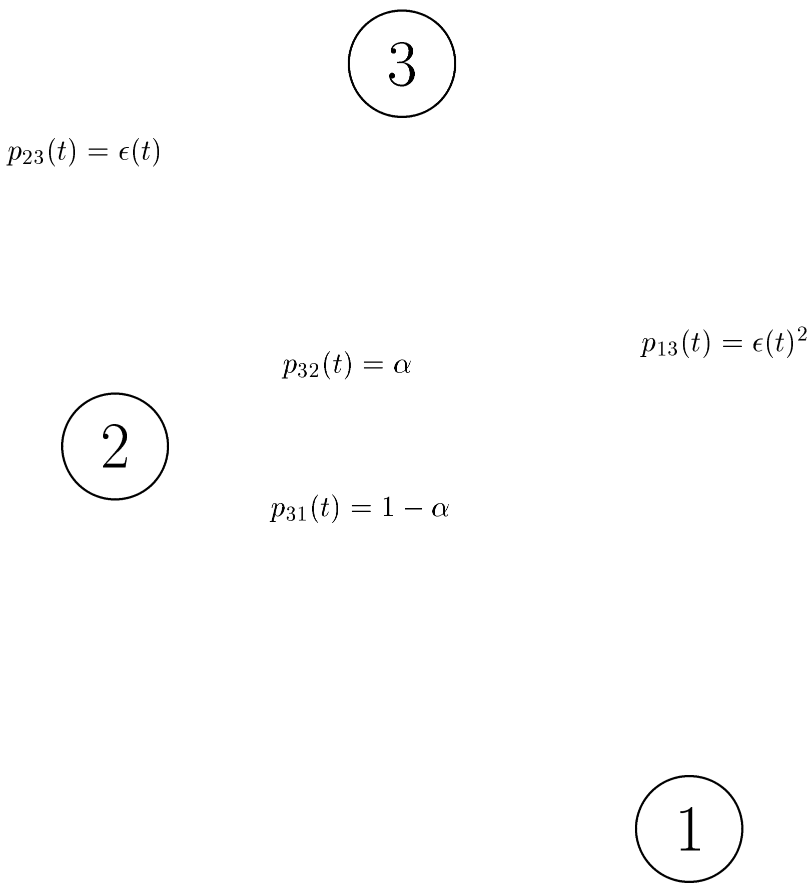

Figure 5: The Markov process of Example 4.