

## MARKOV CHAINS WITH RARE TRANSITIONS AND SIMULATED ANNEALING<sup>†</sup>

JOHN N. TSITSIKLIS

*Massachusetts Institute of Technology*

We consider "approximately stationary" Markov chains in which the entries of the one-step transition probability matrix are known to be of different orders of magnitude and whose structure (that is, the orders of magnitude of the transition probabilities) does not change with time. For such Markov chains we present a method for generating order of magnitude estimates for the  $t$ -step transition probabilities, for any  $t$ . We then notice that algorithms of the simulated annealing type may be represented by Markov chains which are approximately stationary over fairly long time intervals. Using our results we obtain a characterization of the convergent "cooling" schedules for the most general class of algorithms of the simulated annealing type.

**1. Introduction.** The main objective of this paper is the characterization of the cooling schedules under which a simulated annealing algorithm converges to a set of desired states, such as the set where some cost function is minimized, thus generalizing the results of Hajek [9]. The method we follow is based on the observation that in simulated annealing algorithms the "temperature" remains approximately constant for sufficiently long times. For this reason, we may exploit bounds and estimates which are valid for singularly perturbed, approximately stationary Markov chains and obtain interesting conclusions for simulated annealing algorithms. In our development we derive certain results on approximately stationary singularly perturbed Markov chains which seem to be of independent interest.

The structure of the paper is the following. In §2 we assume that we are dealing with a discrete time Markov chain in which each of the one-step transition probabilities is roughly proportional to a certain power of  $\epsilon$ , where  $\epsilon$  is a small parameter. We then present an algorithm, consisting of the solution of certain shortest path problems and some graph theoretic manipulations, which provides estimates for the transition probabilities of the Markov chain for any time between 0 and  $1/\epsilon$ . Then, in §3, we indicate how the procedure of §2 may be applied recursively to produce similar estimates on the transition probabilities for all times. In §4 we use the results of §3 to characterize the convergence of simulated annealing algorithms.

**2. Markov chains parametrized by a small parameter.** In this section we derive order of magnitude estimates on the transition probabilities of a nonstationary discrete time Markov chain. Our results are based on the assumption that such order of magnitude information is available on the one-step transition probabilities of the Markov chain.

\*Received January 2, 1986; revised September 2, 1987.

*AMS 1980 subject classification.* Primary: 60J10; Secondary: 93E25.

*IAOR 1973 subject classification.* Main: Stochastic processes.

*OR/MS Index 1978 subject classification.* Primary: 568 Probability/Markov chains.

*Key words.* Simulated annealing, Markov chains,  $n$ -step transition probabilities, estimation.

<sup>†</sup>Research supported by the Army Research Office under contract DAAG-29-84-K-0005.

We start with some notation. We use  $\mathcal{N}$  and  $\mathcal{N}_0$  to denote the positive and the nonnegative integers, respectively. We also let  $\mathcal{U}$  denote the set of functions  $f: (0, 1) \rightarrow (0, \infty)$  such that for every  $n \in \mathcal{N}_0$  there exists some  $c_n > 0$  such that  $f(\epsilon) \leq c_n \epsilon^n, \forall \epsilon \in (0, 1)$ . In particular, if  $f$  is bounded, membership in  $\mathcal{U}$  is determined by its behavior near zero. Notice that  $\mathcal{U}$  has the property that  $f(\epsilon)/\epsilon^n \in \mathcal{U}, \forall f \in \mathcal{U}, \forall n \in \mathcal{N}$ . Also notice that  $c^{1/\epsilon} \in \mathcal{U}$ , for any  $c \in (0, 1)$ .

We consider a (generally nonstationary) finite state, discrete time Markov chain  $X = \{X(t): t \in \mathcal{N}_0\}$  with state space  $S = \{1, \dots, N\}$ . We assume that  $X(t)$  is also defined for noninteger times by  $X(t) = X(\lfloor t \rfloor)$ , where  $\lfloor t \rfloor$  denotes the integer part of  $t$ . For any  $t \geq 0$  we let  $q_{ij}(t) = P(X(t+1) = j | X(t) = i)$  and  $p_{ij}(t) = P(X(t) = j | X(0) = i)$ . We assume that some structural information is available on this Markov chain. More precisely, let there be given a collection  $\mathcal{A} = \{\alpha_{ij}; 1 \leq i, j \leq N\}$  of elements of  $\mathcal{N}_0 \cup \{\infty\}$ . Let  $f \in \mathcal{U}$  and let  $C_1, C_2$  be positive constants. Suppose that for some  $\epsilon \in (0, 1)$  and for every  $i, j \in S$  we have

$$(2.1) \quad C_1 \epsilon^{\alpha_{ij}} \leq q_{ij}(t) \leq C_2 \epsilon^{\alpha_{ij}}, \quad \forall t \geq 0, \quad \text{if } \alpha_{ij} < \infty,$$

$$(2.2) \quad 0 \leq q_{ij}(t) \leq f(\epsilon), \quad \forall t \geq 0, \quad \text{if } \alpha_{ij} = \infty.$$

We introduce the convention  $\epsilon^\infty = 0, \forall \epsilon \in (0, 1)$ . Then equations (2.1), (2.2) imply that

$$C_1 \epsilon^{\alpha_{ij}} \leq q_{ij}(t) \leq C_2 \epsilon^{\alpha_{ij}} + f(\epsilon), \quad \forall i, j \in S, \quad \forall t \geq 0.$$

We make the natural assumption that for each  $i$  there exists some  $j$  such that  $\alpha_{ij} = 0$ . (Otherwise no Markov chain could satisfy (2.1) and (2.2) for  $\epsilon$  small enough.) We call  $\mathcal{A}$  the *structure* of the Markov chain  $X$ . We will now assume that  $\mathcal{A}, C_1, C_2, f$  are fixed and we denote by  $\mathcal{M}_\epsilon(\mathcal{A}, C_1, C_2, f)$  the set of all Markov chains  $X$  for which (2.1) and (2.2) hold. (Occasionally we use the shorter notation  $\mathcal{M}_\epsilon$ , provided that no confusion may arise.) Notice that if we start with some  $X \in \mathcal{M}_\epsilon$  and shift the time origin we still obtain a Markov chain in  $\mathcal{M}_\epsilon$ , a fact that we will repeatedly use without comment.

As an example, consider the four state stationary Markov chain of Figure 1, for which

$$\mathcal{A} = \begin{bmatrix} 0 & \infty & \infty & 3 \\ \infty & 0 & 1 & \infty \\ 0 & \infty & \infty & \infty \\ \infty & 0 & \infty & \infty \end{bmatrix}.$$

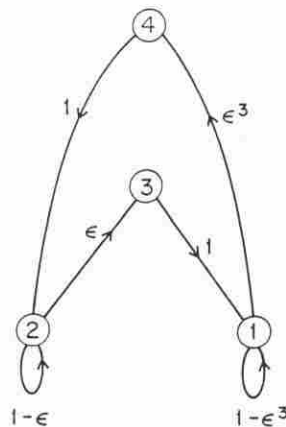


FIGURE 1. An Example.

We define a *path* from  $i$  to  $j$  to be a sequence  $(i_1, \dots, i_m)$  of (not necessarily distinct) states such that  $i_1 = i$ ,  $i_m = j$  and  $m \geq 2$ . The *length* of such a path is defined as  $\sum_{k=1}^{m-1} \alpha_{i_k i_{k+1}}$ . We now provide a classification of the states in  $S$ . In this classification only zero length paths are of interest. A state  $i$  is called *transient* if there exist a state  $j$  and a zero length path from  $i$  to  $j$  but no zero length path from  $j$  to  $i$ ; otherwise,  $i$  is called *recurrent*. This coincides with the usual definition if  $\epsilon = 0$  and the Markov chain is stationary. Let  $TR$ ,  $R$  denote the sets of transient and recurrent states, respectively. For any  $i \in R$ , we let  $R_i$  be the set of all  $j$  such that there exists a zero length path from  $i$  to  $j$ .

In our example,  $R = \{1, 2\}$ ,  $R_1 = \{1\}$ ,  $R_2 = \{2\}$ ,  $TR = \{3, 4\}$ .

- LEMMA 2.1. (i) We have  $i \in R$  and  $j \in R_i$  if and only if  $j \in R$  and  $i \in R_j$ .  
 (ii) If  $i \in R$ ,  $j \notin R_i$ , then  $\alpha_{ij} \geq 1$ .  
 (iii) If  $i \in R$ ,  $j \in TR$ , then  $\alpha_{ij} \geq 1$ .

PROOF. If  $i \in R$  and  $j \in R_i$  but  $j \notin R$ , then there exist a zero length path from  $i$  to  $j$  and a zero length path from  $j$  to some  $k$  but no zero length path from  $k$  to  $j$ . It follows that there exists a zero length path from  $i$  to  $k$  but no zero length path from  $k$  to  $i$ , which contradicts the assumption  $i \in R$ . Thus,  $j \in R$ . The fact that  $i \in R_j$  is then obvious, which proves part (i). The remaining parts hold because if  $i \in R$  and  $\alpha_{ij} = 0$ , then  $j \in R_i$  (by definition) and  $j \in R$  (part (i)), which implies  $j \notin TR$ . ■

We introduce the following assumption on  $\mathcal{A}$ :

*Assumption TRI.* For any  $i, j, k$ , such that at least one of them belongs to  $R$ , we have

$$(2.3) \quad \alpha_{ik} \leq \alpha_{ij} + \alpha_{jk}.$$

This assumption is made for convenience because it leads to some simplification of the proofs. It is not satisfied by the example of Figure 1 and will be removed at the end of this section. We collect a few useful consequences of Assumption TRI.

LEMMA 2.2. Under Assumption TRI:

- (i) If  $i \in R$  and  $\alpha_{ij} = 0$ , then  $j \in R$ ,  $j \in R_i$  and  $\alpha_{ji} = 0$ .  
 (ii) If  $i \in R$ , then  $\alpha_{ii} = 0$  and  $i \in R_i$ .  
 (iii) If  $i \in R$ ,  $j \in R_i$ , then  $\alpha_{ij} = 0$ .  
 (iv) If  $i \in TR$ , then there exists some  $j \in R$  such that  $\alpha_{ij} = 0$ . In particular,  $P(X(t+1) \in TR | X(t) = i) \leq 1 - C_1, \forall i \in TR, \forall t \geq 0$ .

PROOF. (i) If  $i \in R$  and  $\alpha_{ij} = 0$ , then there exists a zero length path from  $j$  to  $i$ , because of the definition of  $R$ . We then apply Assumption TRI along this path to obtain  $\alpha_{ji} = 0$ . Also,  $j \in R_i$ , by the definition of  $R_i$ , and  $j \in R$ , by Lemma 2.1(i).

(ii) Let  $i \in R$ . By assumption, there exists some  $j$  such that  $\alpha_{ij} = 0$ . By part (i) of the lemma, we also have  $\alpha_{ji} = 0$ . Using Assumption TRI, we obtain  $\alpha_{ii} \leq \alpha_{ij} + \alpha_{ji} = 0$ . That  $i \in R_i$  follows from part (i).

(iii) If  $i \in R$ ,  $j \in R_i$ , then there exists a zero length path from  $i$  to  $j$ . Using Assumption TRI along this path we obtain  $\alpha_{ij} = 0$ .

(iv) Given  $i \in TR$ , we define a finite sequence  $(i_1, \dots, i_n)$  of distinct states as follows. Let  $i_1 = i$ . Having chosen  $i_m$ , if  $i_m \in R$  let  $n = m$  and stop. Otherwise, choose  $i_{m+1}$  so that there exists a zero length path from  $i_m$  to  $i_{m+1}$  but no such path from  $i_{m+1}$  to  $i_m$ . (Such an  $i_{m+1}$  exists, by the definition of  $TR$  and must be different from  $i_1, \dots, i_m$  because otherwise there would be a zero length path from  $i_m$  to  $i_{m-1}$ .) Since the state space is finite, the termination condition must be met eventually. Thus

there exist some  $j \in R$  and a zero length path from  $i$  to  $j$ . We then use (inductively) Assumption TRI along this path, to conclude that  $\alpha_{ij} = 0$ . The last statement of the Lemma is then an immediate consequence of (2.1). ■

Let us point out that the sets  $R_i$  determine a partition of  $R$  into disjoint classes which is analogous to the usual partition of recurrent states into ergodic classes for stationary Markov chains.

We need a preliminary result which provides order of magnitude estimates on the probability that a state  $j \in R$  is the first recurrent state to be visited, starting from a transient state  $i$ . We use the notation  $T = \min\{t \geq 0: X(t) \in R\}$ .

**PROPOSITION 2.1.** *For any  $\mathcal{A}$ ,  $C_1$ ,  $C_2$ ,  $f$ , there exist  $F > 0$  and  $g \in \mathcal{U}$  such that for any  $\epsilon \in (0, 1)$ ,  $X \in \mathcal{M}_\epsilon(\mathcal{A}, C_1, C_2, f)$ , we have:*

(i) *If  $i \in TR$ ,  $j \in R$ , then*

$$(2.4) \quad C_1 \epsilon^{\alpha_i} \leq P(X(T) = j | X(0) = i) \leq F \epsilon^{\alpha_i} + g(\epsilon).$$

(ii) *If  $i \in TR$ ,  $j \in TR$ , then*

$$P(X(t) = j, T > t | X(0) = i) \leq (1 - C_1)^{t-1} (NC_2 t \epsilon)^{\min\{1, \alpha_i\}}, \quad \forall t \geq 1.$$

**PROOF.** Let us fix  $\mathcal{A}$ ,  $C_1$ ,  $C_2$ ,  $f$ .

(i) Let  $j \in R$ . We define, for  $\alpha \in \mathcal{N}_0 \cup \{\infty\}$ ,  $S_\alpha = \{i \in TR: \alpha_{ij} = \alpha\}$  and  $Q_\alpha = \{i \in TR: \alpha_{ij} \geq \alpha\}$ . We then define  $p_{\alpha, \epsilon} = \sup_{X \in \mathcal{M}_\epsilon} \max_{i \in Q_\alpha} P(X(T) = j | X(0) = i)$ . We first prove, by induction on  $\alpha$ , that for any  $\alpha < \infty$  there exists some  $F_\alpha > 0$  such that  $p_{\alpha, \epsilon} \leq F_\alpha \epsilon^\alpha$ ,  $\forall \epsilon \in (0, 1)$ . This is clearly true for  $\alpha = 0$ . Suppose it is true for all  $\alpha$  less than some positive integer  $\beta$ . Let  $i \in Q_\beta$  and  $X \in \mathcal{M}_\epsilon$ . Notice that for any state  $k$  we have  $\alpha_{ik} + \alpha_{kj} \geq \alpha_{ij} \geq \beta$ , because of Assumption TRI. Using (2.1) and the induction hypothesis we obtain

$$\begin{aligned} P(X(T) = j | X(0) = i) & \leq \sum_{\alpha=0}^{\beta-1} \sum_{k \in S_\alpha} P(X(T) = j | X(1) = k) P(X(1) = k | X(0) = i) \\ & \quad + P(X(1) \in Q_\beta | X(0) = i) \max_{l \in Q_\beta} P(X(T) = j | X(1) = l) \\ & \quad + P(X(1) = j | X(0) = i) \\ & \leq \sum_{\alpha=0}^{\beta-1} \sum_{k \in S_\alpha} F_\alpha \epsilon^{\alpha_k} C_2 \epsilon^{\alpha_{ik}} + (1 - C_1) p_{\beta, \epsilon} + C_2 \epsilon^\beta \\ & \leq \left[ N \max_{\alpha < \beta} \{F_\alpha\} C_2 + C_2 \right] \epsilon^\beta + (1 - C_1) p_{\beta, \epsilon}. \end{aligned}$$

Taking the supremum of the left-hand side over all  $i \in Q_\beta$  and all  $X \in \mathcal{M}_\epsilon$ , we obtain, for some constant  $F_\beta$ ,

$$p_{\beta, \epsilon} \leq F_\beta \epsilon^\beta + (1 - C_1) p_{\beta, \epsilon}$$

from which it follows that the induction hypothesis is also true for  $\beta$ .

Finally, we assume that  $i \in S_\infty$ . Notice that  $\alpha_{ik} = \infty, \forall k \notin S_\infty$ , because otherwise  $\alpha_{ij} \leq \alpha_{ik} + \alpha_{kj} < \infty$ , contradicting the assumption  $i \in S_\infty$ . Now,

$$\begin{aligned} P(X(T) = j | X(0) = i) & \\ & \leq P(X(1) \in TR, X(1) \notin S_\infty | X(0) = i) + P(X(1) = j | X(0) = i) \\ & \quad + P(X(1) \in S_\infty | X(0) = i) p_{\infty, \epsilon} \\ & \leq Nf(\epsilon) + (1 - C_1) p_{\infty, \epsilon}. \end{aligned}$$

Thus  $p_{\infty, \epsilon} \leq (N/C_1)f(\epsilon), \forall \epsilon > 0$ . This completes the proof of the second inequality in (2.4), with  $F = \max F_\beta$  (where the maximum is over all  $\beta$  such that  $S_\beta$  is nonempty) and with  $g = Nf/C_1$ . The first inequality is a trivial consequence of (2.1).

(ii) If  $\alpha_{ij} = 0$ , the result follows because

$$P(X(t) = j, T > t | X(0) = i) \leq P(T > t | X(0) = i) \leq (1 - C_1)^t.$$

So, suppose that  $\alpha_{ij} \geq 1$ . Let  $A = \{i\} \cup \{k \in TR: \alpha_{ik} = 0\}$  and  $B = \{l \in TR: l \neq i, \alpha_{il} \geq 1\}$ . We claim that  $\alpha_{kl} \geq 1, \forall k \in A, \forall l \in B$ . Indeed, if  $k = i$ , this follows from the definition of  $B$ ; if  $k \neq i$ , then  $1 \leq \alpha_{il} \leq \alpha_{ik} + \alpha_{kl} = \alpha_{kl}$ , which proves the claim. Therefore,

$$P(X(\tau + 1) \in B | X(\tau) \in A) \leq NC_2\epsilon, \quad \forall \tau \geq 0.$$

Notice that  $i \in A, j \in B$ ; thus,

$$\begin{aligned} P(X(t) = j, T > t | X(0) = i) & \leq \sum_{r=0}^{t-1} P(T > t, X(\tau) \in A, X(\tau + 1) \in B | X(0) = i) \\ & \leq \sum_{\tau=0}^{t-1} P(T > \tau | X(0) = i) P(X(\tau + 1) \in B | X(\tau) \in A) \\ & \quad \times P(T > t | X(\tau + 1) \in B) \\ & \leq \sum_{\tau=0}^{t-1} (1 - C_1)^\tau (NC_2\epsilon) (1 - C_1)^{t-\tau-1} \\ & = (1 - C_1)^{t-1} (NC_2\epsilon t), \end{aligned}$$

which completes the proof. ■

Proposition 2.1 is false if Assumption TRI is removed.

The main result of this section is based on the following algorithm, which provides important structural information on the long-run behavior of Markov chains in  $\mathcal{M}_c$ .

*Algorithm 1.* (Input:  $\mathcal{A} = \{\alpha_{ij}: 1 \leq i, j \leq N\}$ ; Output:  $V = \{V(i, j): 1 \leq i, j \leq N\}$ , sets  $R \subset S, TR \subset S$  and for each  $i \in R$  a set  $R_i \subset R$ .)

1. Given  $\mathcal{A}$ , determine  $R, TR$  and the classes  $R_i$  using the definition given earlier.
2. Let  $c_{ij} = \alpha_{ij} - 1$ , if  $i \in R, j \in R, j \notin R_i$ ; let  $c_{ij} = \alpha_{ij}$ , otherwise. (Notice that  $c_{ij} > 0$  always holds.)
3. Solve the shortest path problem from any origin  $i \in R$  to any destination  $j \in R$ , with respect to the link lengths  $c_{ij}$  and subject to the constraint that any intermediate

state on a path must be an element of  $R$ . Let  $V(i, j)$  be the length of such a shortest path. Notice that the  $V(i, j)$ 's satisfy  $V(i, i) = 0, \forall i \in R$  and

$$(2.5) \quad V(i, j) = \min_{k \in R} \{V(i, k) + c_{kj}\}, \quad \forall i, j \in R.$$

4. If  $i \in R, j \in TR$ , let

$$(2.6) \quad V(i, j) = \min_{k \in R} \{V(i, k) + c_{kj}\} = \min_{k \in R} \{V(i, k) + \alpha_{kj}\}.$$

5. If  $i \in TR$ , let

$$(2.7) \quad V(i, j) = \min_{k \in R} \{c_{ik} + V(k, j)\} = \min_{k \in R} \{\alpha_{ik} + V(k, j)\}.$$

Notice that the output  $V(i, j)$  of the above algorithm is equal to the length (with respect to the  $c_{ij}$ 's) of a shortest path from  $i$  to  $j$  subject to the constraint that all states on the path belong to  $R$ , except possibly for the first and the last one. We continue with a few elementary observations on this algorithm:

- PROPOSITION 2.2.** (i)  $V(i, j) \geq 0, \forall i, j$ .  
 (ii)  $V(i, j) \geq 1, \forall i, \forall j \in TR$ .  
 (iii)  $V(i, j) \leq V(i, k) + V(k, j), \forall i, j, k$ .  
 (iv) If  $j \in R$  and  $j' \in R_j$ , then  $V(i, j) = V(i, j'), \forall i$ . Also, if  $i \in R$  and  $i' \in R_i$ , then  $V(i, j) = V(i', j), \forall j$ .

**PROOF.** Part (i) follows from the shortest path interpretation and the nonnegativity of the  $c_{ij}$ 's. Part (ii) follows from (2.6) and the fact (Lemma 2.1(iii)) that  $\alpha_{kj} \geq 1$ , whenever  $k \in R$  and  $j \in TR$ . Part (iii) is clearly true for  $k \in R$ , due to the shortest path interpretation. So, assume that  $k \in TR$ . Let us take shortest paths from  $i$  to  $k$  (of length  $V(i, k)$ ) and from  $k$  to  $j$  (of length  $V(k, j)$ ) and concatenate them. This produces a path from  $i$  to  $j$ , of length  $V(i, k) + V(k, j)$ , such that all intermediate states, except from  $k$ , belong to  $R$ . If  $k_1$  and  $k_2$  are the predecessor and the successor, respectively, of  $k$  in this path, we have  $k_1 \in R, k_2 \in R$  and we can use (2.3) to conclude that  $c_{k_1 k} + c_{k k_2} \geq c_{k_1 k_2}$  which shows that  $k$  may be eliminated from this path, to produce a path from  $i$  to  $j$ , with all intermediate states belonging to  $R$ , and with length less than or equal to  $V(i, k) + V(k, j)$ , as desired. Finally, part (iv) follows from the shortest path interpretation and the fact that  $c_{kl} = \alpha_{kl} = 0$ , whenever  $k \in R$  and  $l \in R_k$ , which is a straightforward consequence of Assumption TRI. ■

We notice that, as a consequence of part (iv) of the proposition, the algorithm need not be carried out for all states. It suffices to consider transient states and one representative from each class  $R_i$ .

The following proposition provides the interpretation of the  $V(i, j)$ 's.

**PROPOSITION 2.3.** For any  $\mathcal{A}, C_1, C_2, f$ , there exist positive constants  $G_1, G_2, G_3, G_4$ , with  $G_4 < 1$ , and some  $g \in \mathcal{U}$  such that, for any  $\epsilon > 0$ , for any  $X \in \mathcal{M}_\epsilon(\mathcal{A}, C_1, C_2, f)$  and any states  $i, j$  we have

$$(2.8) \quad G_1(\epsilon(t - N))^N \epsilon^{V(i, j)} \leq p_{ij}(t) \leq G_2 \epsilon^{V(i, j)} + \chi_{ij} G_3 G_4^t (t\epsilon)^{\min\{1, \alpha_{ij}\}} + g(\epsilon),$$

$$\forall t \in [N + 1, 1/\epsilon],$$

where  $\chi_{ij} = 1$ , if  $i \in TR, j \in TR$ , and  $\chi_{ij} = 0$ , otherwise. (The upper bound in (2.8) is

also true for  $t \in [1, N + 1]$ .) In particular, there exist  $G_1 > 0$ ,  $G_2 > 0$ ,  $g \in \mathcal{U}$  such that for all states  $i, j$ ,

$$(2.9) \quad G_1 \epsilon^{V(i,j)} \leq p_{ij} \left( \frac{1}{\epsilon} \right) \leq G_2 \epsilon^{V(i,j)} + g(\epsilon).$$

PROOF. Let  $\mathcal{A}$ ,  $C_1$ ,  $C_2$ ,  $f$  be fixed throughout the proof. Let  $i \in R$ . Suppose that  $\epsilon \geq 1/(2NC_2)$ . For any  $j \in R_i$  we have  $\alpha_{ji} = 0$  and

$$P(X(t+1) \in R_i | X(t) \in R_i) \geq P(X(t+1) = i | X(t) \in R_i) \geq \min_{j \in R_i} q_{ij}(t) \geq C_1.$$

Thus, if  $0 \leq s \leq t \leq 1/\epsilon$ , then

$$P(X(t) \in R_i | X(s) \in R) \geq C_1^{2NC_2}, \quad \forall X \in \mathcal{M}_\epsilon, \quad \forall \epsilon \in (1/(2NC_2), 1).$$

Suppose now that  $\epsilon < 1/(2NC_2)$ . Notice that for any  $k \in R_i$ ,  $j \notin R_i$  we have  $q_{kj}(t) \leq C_2 \epsilon$ ,  $\forall t$ . It follows that  $P(X(t+1) \notin R_i | X(t) \in R_i) \leq NC_2 \epsilon$  and, for  $0 \leq s \leq t \leq 1/\epsilon$ ,

$$P(X(t) \in R_i | X(s) \in R_i) \geq (1 - NC_2 \epsilon)^{1/\epsilon} \geq \left(\frac{1}{2}\right)^{2NC_2},$$

$$\forall X \in \mathcal{M}_\epsilon, \quad \forall \epsilon \in (0, 1/(2NC_2)).$$

We conclude that there exists some  $F_1 \geq 0$  such that

$$(2.10) \quad P(X(t) \in R_i | X(s) \in R_i) \geq F_1, \quad 0 \leq s \leq t \leq 1/\epsilon,$$

$$\forall \epsilon > 0, \quad \forall X \in \mathcal{M}_\epsilon, \quad \forall i \in R.$$

We now start the proof of the lower bound in (2.8). If  $V(i, j) = \infty$ , there is nothing to prove, so we will be assuming that  $V(i, j) < \infty$ . We first assume that  $i \in R$  and  $j \in R$ . Then there exists a sequence  $i = i_1, i_2, \dots, i_n = j$  of elements of  $R$  (with  $n \leq N$ ), such that  $\sum_{k=1}^{n-1} c_{i_k i_{k+1}} = V(i, j)$ . Let  $k \in \{1, \dots, n\}$  and suppose that there exists some  $F_k > 0$  such that, for all  $\epsilon > 0$  and for all  $X \in \mathcal{M}_\epsilon$ ,

$$(2.11) \quad P(X(t) \in R_{i_k} | X(0) = i) \geq F_k (\epsilon(t - k + 1))^{k-1} \epsilon^{\sum_{l=1}^{k-1} c_{i_l i_{l+1}}},$$

$$\forall t \in [k - 1, 1/\epsilon].$$

We then have, for  $t \in [k, 1/\epsilon]$ ,

$$(2.12) \quad P(X(t) \in R_{i_{k+1}} | X(0) = i)$$

$$\geq \sum_{s=0}^{t-1} P(X(t) \in R_{i_{k+1}} | X(s+1) \in R_{i_{k+1}})$$

$$\times P(X(s+1) \in R_{i_{k+1}} | X(s) \in R_{i_k}) P(X(s) \in R_{i_k} | X(0) = i)$$

$$\geq \sum_{s=k}^{t-1} F_1 C_1 \epsilon^{\alpha_{i_k i_{k+1}}} (F_k (\epsilon(s - k + 1))^{k-1} \epsilon^{\sum_{l=1}^{k-1} c_{i_l i_{l+1}}})$$

$$\geq (F_k F_1 C_1) \epsilon^{\sum_{l=1}^k c_{i_l i_{l+1}}} \epsilon \sum_{s=k}^{t-1} (\epsilon(s - k + 1))^{k-1}.$$

Clearly, there exists a constant  $F'_k$  such that

$$\sum_{s=k}^{t-1} (s-k+1)^{k-1} \geq F'_k (t-k)^k, \quad \forall t \geq k.$$

Inequality (2.10) shows that (2.11) holds for  $k=1$ . We have thus proved by induction on  $k$  that (2.11) holds for all  $k$ . Notice that

$$\begin{aligned} p_{ij}(t) &= P(X(t) = j | X(0) = i) \\ &\geq P(X(t) = j | X(t-1) \in R_j) P(X(t-1) \in R_j | X(0) = i) \\ &\geq C_1 P(X(t-1) \in R_j | X(0) = i), \end{aligned}$$

from which the left-hand side of (2.8) follows for  $t \geq N$  and for  $i \in R$  and  $j \in R$ .

Suppose now that  $i \in R$ ,  $j \in TR$  and let  $k \in R$  be such that  $V(i, j) = V(i, k) + \alpha_{kj}$ . If  $\alpha_{kj} = \infty$ , then  $V(i, j) = \infty$  and there is nothing to prove. So, assume that  $\alpha_{kj} < \infty$ . Then, for  $t \geq N+1$ ,

$$\begin{aligned} P(X(t) = j | X(0) = i) &\geq P(X(t) = j | X(t-1) = k) P(X(t-1) = k | X(0) = i) \\ &\geq C_1 \epsilon^{\alpha_{kj}} P(X(t-1) = k | X(0) = i). \end{aligned}$$

Given that we have already proved the lower bound for  $p_{ik}(t)$ , the desired result for  $p_{ij}(t)$  follows.

Finally, let  $i \in TR$  and  $t \geq N+1$ . The result follows similarly by choosing  $k \in R$  so that  $\alpha_{ik} + V(k, j) = V(i, j)$  and using the inequality

$$P(X(t) = j | X(0) = i) \geq P(X(1) = k | X(0) = i) P(X(t) = j | X(1) = k).$$

We now turn to the proof of the upper bound in (2.8). Let  $i \in R$  be fixed. We define  $E_\alpha = \{j \in R: V(i, j) = \alpha\}$ ,  $T_\alpha = \{j \in TR: V(i, j) = \alpha\}$ ,  $E_{\leq \alpha} = \bigcup_{\beta \leq \alpha} E_\beta$ . We also define similarly  $E_{\geq \alpha}$ ,  $T_{\leq \alpha}$ ,  $T_{\geq \alpha}$ . We will prove by induction that for any  $\alpha < \infty$  the following statements hold:

( $SE_\alpha$ ): There exists some  $G_\alpha > 0$  such that  $\forall \epsilon \in (0, 1)$ ,  $\forall X \in \mathcal{M}_\epsilon$ ,  $\forall j \in E_{\geq \alpha}$  and  $\forall t \leq 1/\epsilon$  we have  $p_{ij}(t) \leq G_\alpha \epsilon^\alpha$ .

( $ST_\alpha$ ): There exists some  $G'_\alpha > 0$ , such that  $\forall \epsilon \in (0, 1)$ ,  $\forall X \in \mathcal{M}_\epsilon$ ,  $\forall j \in T_{\geq \alpha}$  and  $\forall t \leq 1/\epsilon$  we have  $p_{ij}(t) \leq G'_\alpha \epsilon^\alpha$ .

Statements  $SE_0$  and  $ST_0$  are trivially true, with  $G_0 = G'_0 = 1$ . We now prove  $ST_1$ . (Notice that  $T_{\geq 1} = TR$ .) Now,

$$\begin{aligned} (2.13) \quad P(X(t+1) \in TR | X(0) = i) \\ &\leq P(X(t+1) \in TR | X(t) \in TR) P(X(t) \in TR | X(0) = i) \\ &\quad + P(X(t+1) \in TR | X(t) \in R) \\ &\leq (1 - C_1) P(X(t) \in TR | X(0) = i) + NC_2 \epsilon. \end{aligned}$$

Since  $i \in R$ ,  $P(X(0) \in TR | X(0) = i) = 0$  and (2.13) implies  $P(X(t) \in TR | X(0) = i) \leq (NC_2 \epsilon) / C_1$ ,  $\forall t \geq 0$ , which proves  $ST_1$ .



Now let  $\alpha$  be some positive integer and assume that statements  $SE_{\beta-1}$  and  $ST_{\beta}$  are true, for all  $\beta \leq \alpha$ . We will prove that  $SE_{\alpha}$  and  $ST_{\alpha+1}$  are also true. We first need the following lemma.

**LEMMA 2.3.** *If  $j \in J = E_{\leq(\alpha-1)} \cup T_{\leq\alpha}$  and  $k \in K = E_{>\alpha} \cup T_{>(\alpha+1)}$ , then  $V(i, j) + \alpha_{jk} \geq \alpha + 1$ .*

**PROOF.** (i) If  $j \in E_{\leq(\alpha-1)}$ ,  $k \in E_{>\alpha}$ , then  $V(i, j) + \alpha_{jk} = V(i, j) + c_{jk} + 1 \geq V(i, k) + 1 \geq \alpha + 1$ .

(ii) If  $j \in E_{\leq(\alpha-1)}$ ,  $k \in T_{>(\alpha+1)}$ , then  $V(i, j) + \alpha_{jk} = V(i, j) + c_{jk} \geq V(i, k) \geq \alpha + 1$ .

(iii) If  $j \in T_{\leq\alpha}$ ,  $k \in E_{>\alpha}$ , let  $l \in R$  be such that  $V(i, l) + \alpha_{lj} = V(i, j)$ . Suppose that  $l \in R_k$ . Then  $V(i, l) = V(i, k) \geq \alpha$  and  $V(i, j) = V(i, l) + \alpha_{lj} \geq \alpha + 1$ , which contradicts the assumption  $j \in T_{\leq\alpha}$ . We thus assume that  $l \notin R_k$ . Then  $V(i, j) + \alpha_{jk} = V(i, l) + \alpha_{lj} + \alpha_{jk} \geq V(i, l) + \alpha_{lk} = V(i, l) + c_{lk} + 1 \geq V(i, k) + 1 \geq \alpha + 1$ .

(iv) If  $j \in T_{\leq\alpha}$ ,  $k \in T_{>(\alpha+1)}$ , let  $l \in R$  be such that  $V(i, l) + \alpha_{lj} = V(i, j)$ . Then  $V(i, j) + \alpha_{jk} = V(i, l) + \alpha_{lj} + \alpha_{jk} \geq V(i, l) + \alpha_{lk} \geq V(i, k) \geq \alpha + 1$ . ■

Let  $j \in J$ . If  $j \in E_{\leq\alpha-1}$ , then  $V(i, j) \leq \alpha - 1$ , by definition. Also,  $j \in E_{V(i, j)}$ , and by the induction hypothesis  $p_{ij}(t) \leq G\epsilon^{V(i, j)}$ , where  $G = \max\{G_{\beta-1}, G'_{\beta}; \beta \leq \alpha\}$ . A similar argument gives the same conclusion if  $j \in T_{\leq\alpha}$ . Now, if  $k \in K$ , then

$$P(X(t+1) = k | X(t) = j)P(X(t) = j | X(0) = i) \leq C_2\epsilon^{\alpha_{jk}}G\epsilon^{V(i, j)} \leq C_2G\epsilon^{\alpha+1},$$

where the last inequality follows from Lemma 2.3.

Notice that  $V(i, i) \leq \alpha_i = 0$ , since  $i \in R$ . Thus,  $i \in E_0$  which implies that  $i \in J$ . Therefore,

$$\begin{aligned} P(X(t) \in K | X(0) = i) &\leq \sum_{\tau=0}^{t-1} \sum_{k \in K, j \in J} P(X(\tau+1) = k | X(\tau) = j) \\ &\quad \times P(X(\tau) = j | X(0) = i) \\ &\leq \frac{1}{\epsilon} C_2 G \epsilon^{\alpha+1} N^2, \quad \forall t \in [1, 1/\epsilon], \end{aligned}$$

which proves  $SE_{\alpha}$ . Finally,

$$\begin{aligned} P(X(t+1) \in T_{>\alpha+1} | X(0) = i) &\leq (1 - C_1)P(X(t) \in T_{>\alpha+1} | X(0) = i) \\ &\quad + NG_{\alpha}\epsilon^{\alpha}C_2\epsilon + N^2C_2G\epsilon^{\alpha+1} \end{aligned}$$

which shows that

$$P(X(t) \in T_{>\alpha+1} | X(0) = i) \leq (1/C_1)(NG_{\alpha}C_2 + N^2C_2G)\epsilon^{\alpha+1}, \quad \forall t \in [1, 1/\epsilon].$$

This proves  $ST_{\alpha}$  and completes the induction.

We have thus completed the proof of the upper bound in (2.8) for the case where  $i \in R$  and  $V(i, j) < \infty$ . The proof for the case  $i \in R$  and  $V(i, j) = \infty$  is very simple and is omitted.

We now assume that  $i \in TR$  and  $j \in TR$ . Let  $T$  be the random time of Proposition 2.1. Then, for some  $F > 0$ ,  $G > 0$ ,  $g, g', g'' \in \mathcal{U}$ , we have

$$\begin{aligned} p_{ij}(t) &\leq \sum_{k \in R} P(X(t) = j | X(T) = k, T \leq t) P(X(T) = k, T \leq t | X(0) = i) \\ &\quad + P(X(t) = j, T > t | X(0) = i) \\ &\leq \sum_{k \in R} [G\epsilon^{V(k,j)} + g(\epsilon)] [F\epsilon^{\alpha_k} + g'(\epsilon)] + (1 - C_1)^{t-1} (NC_2 t \epsilon)^{\min\{1, \alpha_{ij}\}} \\ &\leq (NG)F\epsilon^{V(i,j)} + g''(\epsilon) + (NC_2)(1 - C_1)^{t-1} (t\epsilon)^{\min\{1, \alpha_{ij}\}}, \quad \forall t \in [1, 1/\epsilon], \end{aligned}$$

which has the same form as the right-hand side of (2.8). (In the above inequalities we use the easily verifiable fact  $V(k, j) + \alpha_{ik} \geq V(i, j)$ .)

Finally, we consider the case where  $i \in TR$  and  $j \in R$ . The argument is identical as in the preceding case except that the term  $P(X(t) = j, T > t | X(0) = i)$  is now equal to zero. ■

We point out that the bounds (2.8) are also applicable to  $P(X(t+s) = j | X(s) = i)$ , for every  $s \geq 0$ , because  $\mathcal{M}_\epsilon$  is invariant under translations of the time origin.

Notice that the gap between the upper and lower bounds is a multiplicative factor independent of  $\epsilon$ , when  $t = 1/\epsilon$ . For smaller times the bounds are much further apart. It is not hard to close this gap, although we do not need to do this for our purposes.

The remainder of this section is devoted to showing that Assumption TRI is not an essential restriction. Roughly speaking, we will establish that our results are applicable to any Markov chain which is aperiodic in the fastest time scale.

Let there be given a set  $\mathcal{A} = \{\alpha_{ij}; 1 \leq i, j \leq N\}$  of elements of  $\mathcal{N}_0 \cup \{\infty\}$ , not necessarily satisfying (2.3). For any  $i, j \in S$ ,  $n \in \mathcal{N}$ , we define  $\beta_{ij}^1 = \alpha_{ij}$  and  $\beta_{ij}^{n+1} = \min_k \{\beta_{ik}^n + \alpha_{kj}\}$ . That is,  $\beta_{ij}^n$  equals the distance of a shortest path (with respect to the link lengths  $\alpha_{ij}$ ) from  $i$  to  $j$  which has exactly  $n$  hops. Also, let  $\beta_{ij} = \min_{n \in \mathcal{N}} \beta_{ij}^n$ . Notice that for every  $i, j, k, m, n$  we have  $\beta_{ik}^m + \beta_{kj}^n \geq \beta_{ij}^{m+n}$ . We introduce the following assumption on  $\mathcal{A}$ .

*Assumption AP.* There exists some positive integer  $M$  with the following property: if  $i \in R$  or  $j \in R$ , then  $\beta_{ij}^n = \beta_{ij}$ ,  $\forall n \geq M$ .

For any Markov chain whose structure is described by  $\mathcal{A}$ , meaning that the estimates (2.1), (2.2) are valid, assumption AP amounts to the following: if we substitute 0 for  $\epsilon$ , and decompose the resulting Markov chain into ergodic classes, in the usual manner, then each of the noncommunicating classes of recurrent states is aperiodic.

It can be shown that if  $\mathcal{A}$  satisfies Assumption AP, then  $M$  can be chosen to be smaller than  $N^2$ . (This is related to the fact that the ‘‘index of primitivity’’ of any primitive nonnegative matrix is bounded above by  $N^2 - 2N + 2$ ; for more details, see Chapter 2 of [13].)

Given  $\mathcal{A}$ , some positive constants  $C_1, C_2$ , some  $f \in \mathcal{U}$  and some  $\epsilon > 0$ , consider the set  $\mathcal{M}_\epsilon(\mathcal{A}, C_1, C_2, f)$ . Let  $Q$  be some positive integer. For any  $X \in \mathcal{M}_\epsilon(\mathcal{A}, C_1, C_2, f)$ , let us define  $X^Q$  to be the discrete time Markov chain obtained by sampling  $X$  every  $Q$  time units. Finally, let  $\mathcal{B}^Q = \{\beta_{ij}^Q; 1 \leq i, j \leq N\}$ . The following proposition establishes that the coefficients  $\beta_{ij}^Q$  describe the structure of the sampled Markov chain  $X^Q$ .

**PROPOSITION 2.4.** *For any  $\mathcal{A}, C_1, C_2, f \in \mathcal{U}$ , and for any  $Q \in \mathcal{N}$ , there exist some positive  $C'_1, C'_2$  and some  $f' \in \mathcal{U}$  such that  $\{X^Q: X \in \mathcal{M}_\epsilon(\mathcal{A}, C_1, C_2, f)\}$  is a subset of  $\mathcal{M}_\epsilon(\mathcal{B}^Q, C'_1, C'_2, f')$ .*

PROOF. The result is immediate from the fact that  $P(X^Q(t+1) = j | X^Q(t) = i)$  equals the sum, over all paths from  $i$  to  $j$  with exactly  $Q$  hops, of the probability that  $X(t)$  follows any such path. The desired conclusion follows with  $C'_1 = C_1^Q$ ,  $C'_2 = C_2^Q N^Q$  and  $f'(\epsilon) = N^Q f(\epsilon)$ . (The factor  $N^Q$  arises because there are less than  $N^Q$  such paths.) ■

The main reason, however, for introducing  $X^Q$  is the following.

PROPOSITION 2.5. *Suppose that  $\mathcal{A}$  satisfies Assumption AP and that  $Q \geq M + N$ , where  $M$  is the constant in that assumption. Then  $\mathcal{B}^Q$  satisfies Assumption TRI.*

PROOF. Let us fix some  $i, j, k$  and suppose that  $i \in R$ . Let  $n$  be such that  $\beta_{jk}^n = \beta_{jk}$ . Without loss of generality we may assume that  $n \leq N$ . Furthermore,  $\beta_{jk}^Q \geq \beta_{jk} = \beta_{jk}^n$ . Therefore, using Assumption AP and the inequalities  $Q \geq M$ ,  $Q - n \geq M$ , we have  $\beta_{ij}^Q + \beta_{jk}^Q \geq \beta_{ij}^Q + \beta_{jk}^n = \beta_{ij}^{Q-n} + \beta_{jk}^n \geq \beta_{ik}^Q$ , which is the desired inequality. The proof for the cases  $j \in R$  or  $k \in R$  is identical and is omitted. ■

As a consequence of Propositions 2.4 and 2.5, Proposition 2.3 becomes applicable to an appropriately sampled version of a given Markov chain, assuming Condition AP. We notice that Proposition 2.3 will provide us with estimates of the transition probabilities only for those times which are integer multiples of  $Q$ . However, it is easy to show that the same estimates are also valid for intermediate times as well.

With a more elaborate choice of  $Q$ , the conclusions of Proposition 2.5 are valid even if Assumption AP fails to hold. However, in this case, the corresponding conclusions of Proposition 2.3 will only be valid for the sampled chain  $X^Q$  and not, in general, for the original Markov chain. Rather, the order of magnitude of  $p_{ij}(t)$  will vary periodically with  $t$ .

We close this section by pointing out that there is nothing special about the coefficients  $\alpha_{ij}$  being integer. For example, if the  $\alpha_{ij}$  are rationals we could introduce another small parameter  $\delta$  (to replace  $\epsilon$ ) and another set of integer coefficients  $\beta_{ij}$ , so that  $\delta^{\beta_{ij}} = \epsilon^{\alpha_{ij}}$ . Even if the  $\alpha_{ij}$ 's are not rational, neither are their ratios rational, the proof of Proposition 2.3 remains valid, as long as  $\min\{\alpha_{ij} : \alpha_{ij} \neq 0\}$ . This can be always achieved by redefining the small parameter  $\epsilon$ .

Returning to the example of Figure 1, we notice that Assumption AP is satisfied and an easy calculation yields

$$\mathcal{B}^Q = \begin{bmatrix} 0 & 3 & 4 & 3 \\ 1 & 0 & 1 & 4 \\ 0 & 3 & 4 & 3 \\ 1 & 0 & 1 & 4 \end{bmatrix}, \quad \forall Q \geq 4.$$

By applying Algorithm I to  $\mathcal{B}^Q$  we obtain

$$V = \begin{bmatrix} 0 & 2 & 3 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 2 & 3 & 3 \\ 0 & 0 & 1 & 3 \end{bmatrix}.$$

**3. Determining the structure at successively slower time scales.** Proposition 2.3 allows us to determine the structure of a Markov chain  $X \in \mathcal{M}_\epsilon$  in the first of the slow time scales, that is for times of the order of  $1/\epsilon$ . We now notice that the transition probabilities  $P(X(1/\epsilon) = j | X(0) = i)$  satisfy (2.1), (2.2) (with a new choice of  $C_1, C_2, f$ ) provided that we replace  $\alpha_{ij}$  by  $V(i, j)$ . Moreover, due to part (iii) of Proposition

2.2, the coefficients  $V(i, j)$  satisfy the triangle inequality (2.3) and, therefore, Proposition 2.3 becomes applicable once more. This yields estimates for the transition probabilities  $P(X(1/\epsilon^2) = j | x(0) = i)$ . This procedure may be repeated to yield estimates for  $P(X(1/\epsilon^d) = j | x(0) = i)$ , for any positive integer  $d$ . To summarize, we have the following algorithm:

*Algorithm II.* (Input:  $\mathcal{A} = \{\alpha_{ij}; 1 \leq i, j \leq N\}$ , satisfying Assumption TRI; Output: for each  $d \in \mathcal{N}_0$ , a collection  $V^d = \{V^d(i, j); 1 \leq i, j \leq N\}$ , a subset  $R^d$  of the state space and for each  $i \in R^d$  a set  $R_i^d \subset R^d$ .)

1. Let  $V^0(i, j) = \alpha_{ij}, \forall i, j$ .
2. Let  $V^d$  be the input to Algorithm I; let  $V^{d+1}, R^d, TR^d, R_i^d$  be the outputs returned by Algorithm I.

Notice that  $R^d$  is the set of all states  $i$  with the following property: if for some  $j$  we have  $V^d(i, j) = 0$  then we must also have  $V^d(j, i) = 0$ . Also, for any  $i \in R^d$ , we have  $R_i^d = \{j \in R^d; V^d(i, j) = 0\}$ . The remarks preceding Algorithm II lead to the next proposition. (Notice that when we use Proposition 2.3 to obtain estimates for  $t \approx 1/\epsilon^d$ , the unit of time becomes  $1/\epsilon^{d-1}$ . For this reason, the variable  $t$  in Proposition 2.3 must be replaced by  $t\epsilon^{d-1}$ .)

**PROPOSITION 3.1.** *Given some  $\mathcal{A}$  satisfying Assumption TRI and some  $d \in \mathcal{N}$ , let  $V^d(i, j), R^d$ , be the collection of integers and the subset returned by Algorithm II. Then, for any positive constants  $C_1, C_2$  and for any  $f \in \mathcal{U}$ , there exist positive constants  $D_1, D_2, D_3, D_4$ , with  $D_4 < 1$ , and  $g \in \mathcal{U}$ , such that, for any  $\epsilon \in (0, 1)$  and for any Markov chain  $X \in \mathcal{M}_\epsilon(\mathcal{A}, C_1 C_2, f)$  we have<sup>1</sup>*

(3.1)

$$\begin{aligned} D_1(\epsilon(\lfloor \epsilon^{d-1} t \rfloor - N))^N \epsilon^{V^d(i, j)} &\leq p_{ij}(t) \\ &\leq D_2 \epsilon^{V^d(i, j)} + \chi_{ij}^{d-1} D_3 D_4^{\epsilon^{d-1}} (t \epsilon^d)^{\min\{V^{d-1}(i, j), 1\}} + g(\epsilon), \\ &\quad \forall t \in [(N+1)/\epsilon^{d-1}, 1/\epsilon^d], \end{aligned}$$

where  $\chi_{ij}^{d-1} = 1$ , if  $i \in TR^{d-1}, j \in TR^{d-1}$ , and  $\chi_{ij}^{d-1} = 0$ , otherwise. (The upper bound in (3.1) is also valid for  $t \in [1/\epsilon^{d-1}, (N+1)/\epsilon^{d-1}]$ .) In particular, there exist  $D_1, D_2 > 0, g \in \mathcal{U}$  such that

$$(3.2) \quad D_1 \epsilon^{V^d(i, j)} \leq p_{ij} \left( \frac{1}{\epsilon^d} \right) \leq D_2 \epsilon^{V^d(i, j)} + g(\epsilon).$$

**PROOF.** (Outline) For  $d = 1$ , the result is a restatement of Proposition 2.3. Assume that the result is true for  $d - 1$ . If  $t$  is an integer multiple of  $1/\epsilon^{d-1}$ , the result for  $d$  follows from Proposition 2.3, using  $1/\epsilon^{d-1}$  as the time unit. A simple way of obtaining the same bound for general  $t$  is the following. We restrict, without loss of generality, to the case where  $\epsilon \leq 1/2$ , because for  $\epsilon > 1/2$  the result is trivially true. Given some  $t \in [(N+1)/\epsilon^{d-1}, 1/\epsilon^d]$  which is not an integer multiple of  $1/\epsilon^{d-1}$ , let  $k$  be an integer such that  $k/\epsilon^{d-1} < t < (k+1)/\epsilon^{d-1}, k \in [N+1, 1/\epsilon]$ . Let  $\epsilon_0$  be such that  $t = k/\epsilon_0^{d-1}$ . Notice that  $1 < (\epsilon/\epsilon_0) < [(k+1)/k]^{1/(d-1)} \leq 2$ . Since  $X \in \mathcal{M}_\epsilon(\mathcal{A}, C_1, C_2, f)$ , it follows that  $X \in \mathcal{M}_{\epsilon_0}(\mathcal{A}, 2^{-B} C_1, 2^B C_2, g)$ , where  $B = \max\{\alpha_{ij}; \alpha_{ij} < \infty\}$  and where  $g \in \mathcal{U}$  is suitably defined. Since  $\epsilon_0 < \epsilon$ , we see that

<sup>1</sup>The notation  $\lfloor a \rfloor$  stands for the largest integer bounded above by  $a$ .

$k \in [N + 1, 1/\epsilon_0]$ . To summarize, we have  $t \in [(N + 1)/\epsilon_0^{d-1}, 1/\epsilon_0^d]$  and  $t$  is an integer multiple of  $1/\epsilon_0^{d-1}$  and this reduces to the case for which the result has been already proved. We conclude that (3.1) holds, with  $\epsilon_0$  in the place of  $\epsilon$ . We then rewrite (3.1), in terms of  $\epsilon$ , as follows. We have  $\lfloor \epsilon_0^{d-1}t \rfloor = k = \lfloor \epsilon^{d-1}t \rfloor$  and this term is of no concern. For the other occurrences of  $\epsilon_0$ , we use the inequalities  $\epsilon/2 \leq \epsilon_0$ ,  $\epsilon_0 < \epsilon$ , whichever is appropriate at each place. We then see that (3.1) still holds, provided that some of the constants are redefined. ■

We continue with a few remarks on the quantities computed by Algorithm II.

- PROPOSITION 3.2. (i) For any  $d, i, j, k$ , we have  $V^d(i, j) \leq V^d(i, k) + V^d(k, j)$ .  
 (ii) For any  $d$ , we have  $R^{d+1} \subset R^d$ .  
 (iii)  $V^d(i, j) + V^c(j, k) \geq V^{\max\{c, d\}}(i, k)$ ,  $\forall i, j, k, c, d$ .

PROOF. (i) This is an immediate consequence of part (iii) of Proposition 2.2.

(ii) Suppose that  $i \in R^{d+1}$ . Then,  $V^{d+1}(i, i) = 0$ . Using part (ii) of Proposition 2.2, we conclude that  $i \notin TR^d$ , or, equivalently,  $i \in R^d$ .

(iii) Using Proposition 3.1 twice, there exist constants  $D'_1, D'_2$  such that

$$D'_1 \epsilon^{V^d(i, j) + V^c(j, k)} \leq P \left( X \left( \frac{1}{2\epsilon^d} + \frac{1}{2\epsilon^c} \right) = k | X(0) = i \right) \leq D'_2 \epsilon^{V^{\max\{c, d\}}(i, k)}.$$

Moreover, this inequality is true for all  $X \in \mathcal{M}_\epsilon$  and for all  $\epsilon > 0$ . Letting  $\epsilon$  be arbitrarily small, we conclude that the claimed result holds. ■

As a corollary of Proposition 3.2 we obtain that some of the upper bounds of Proposition 3.1 are true even for times smaller than  $1/\epsilon^{d-1}$ :

COROLLARY 3.1. For any  $\mathcal{A}, C_1, C_2, f$  and any nonnegative integer  $d$ , there exist some  $h \in \mathcal{U}$  and some  $C > 0$  such that, for all  $\epsilon \in (0, 1)$ , all  $X \in \mathcal{M}_\epsilon(\mathcal{A}, C_1, C_2, f)$  we have:

(i) If  $i \in R^d$  or  $j \in R^d$ , then

$$(3.3) \quad p_{ij}(t) \leq C\epsilon^{V^d(i, j)} + h(\epsilon), \quad \forall t \in [1, 1/\epsilon^d];$$

(ii) If  $V^c(i, j) \geq 1, \forall c \leq d$ , then

$$(3.4) \quad p_{ij}(t) \leq C\epsilon + h(\epsilon), \quad \forall t \in [1, 1/\epsilon^d].$$

PROOF. (i) Suppose that  $i \in R^d$ . Given  $t$ , let us choose some  $c \leq d$  such that  $t \in [1/\epsilon^{c-1}, 1/\epsilon^c]$ . Since  $i \in R^d$ , Proposition 3.2(ii) shows that  $i \in R^c, \forall c \leq d$ , and therefore  $\chi_{ij}^{c-1} = 0$ . Then the upper bound (3.1) yields  $p_{ij}(t) \leq G\epsilon^{V^c(i, j)} + h(\epsilon)$ , for some  $h \in \mathcal{U}$ . Using Proposition 3.2(iii), we obtain  $V^c(i, j) + V^d(i, i) \geq V^d(i, j)$ . Since  $i \in R^d$ , we have  $V^d(i, i) = 0$  and therefore

$$p_{ij}(t) \leq G\epsilon^{V^c(i, j)} + h(\epsilon) \leq G\epsilon^{V^d(i, j)} + h(\epsilon),$$

as desired. The proof for the case  $j \in R^d$  is identical.

(ii) We choose  $c$  as in part (i) and we follow the same steps. The bounds of Proposition 3.1 now yield, for some constants  $G, G', G''$ , with  $G'' < 1$  and some

$h \in \mathcal{U}$ ,

$$\begin{aligned} p_{ij}(t) &\leq G\epsilon^{V^{c-1}(i,j)} + G'(G'')^{t\epsilon^{c-1}}(\epsilon^c t)^{\min\{1, V^{c-1}(i,j)\}} + h(\epsilon) \\ &\leq G\epsilon + \epsilon G' \left[ (G'')^{t\epsilon^{c-1}} (t\epsilon^{c-1}) \right] + h(\epsilon) \\ &\leq \epsilon \left[ G' \max_{a \geq 1} \{ a (G'')^a \} + G \right] + h(\epsilon), \end{aligned}$$

which is of the desired form. ■

For the sampled version of the example of Figure 1,  $V^1$  has been already computed in §2. We see that  $R^1 = \{1\}$  and therefore  $R^d = \{1\}$ ,  $\forall d \geq 1$ .

We continue with a few remarks on the applicability and usefulness of Algorithms I and II. We observe in Algorithm I that in order to determine  $V(i, j)$  for  $i \in R$  and  $j \in R$ , we only need to know the coefficients  $\alpha_{ij}$  for  $i$  and  $j$  belonging to  $R$ . This has the following implication for Algorithm II: in order to compute the coefficients  $\{V^{d+1}(i, j) : i, j \in R^d\}$ , we only need to know the coefficients  $\{V^d(i, j) : i, j \in R^d\}$ . Since  $R^{d+1} \subset R^d$ , it follows that the coefficients  $\{V^{d+1}(i, j) : i, j \in R^{d+1}\}$  may be computed from the coefficients  $\{V^d(i, j) : i, j \in R^d\}$ . Thus, if we are only interested in determining which states are recurrent for each time scale (as well as in determining the corresponding ergodic decomposition) we may eliminate, at each stage of Algorithm II, the states which have been found to be transient, that is the elements of  $TR^d$ . This observation, together with the fact that we only need to carry out the algorithm for just one representative from each class  $R_i^d$ , should result in a substantial amount of savings, were the algorithm to be implemented.

Naturally, Algorithm II is applicable to the appropriately sampled versions of Markov chains satisfying Assumption AP. Then, inequalities (3.1) and (3.2) are valid for times which are integer multiples of the sampling period  $Q$ . Moreover, a simple argument shows that these inequalities are valid for intermediate times as well. Also notice that sampling needs be carried out only once. Even if  $\mathcal{A}$  satisfies AP but not TRI, still the coefficients  $V^d$ ,  $d > 0$ , will automatically satisfy Assumption TRI and no sampling is required at subsequent stages of Algorithm II.

We compare Algorithm II and Proposition 3.1 to the results available in the literature. There has been a substantial amount of research on singularly perturbed stationary Markov chains [2, 3, 4, 5, 12]. Typical results obtain exact asymptotic expressions for the transition probabilities, as a small parameter  $\epsilon$  converges to zero. These asymptotic expressions are obtained recursively, by proceeding from one time scale to the next, similarly with Algorithm II. Each step in this recursion involves the solution of systems of linear equations and, possibly, the evaluation of the pseudoinverse of some matrices [2], which may be computationally demanding, especially if we are dealing with large scale systems. However, we may conceive of situations in which we are not so much interested in knowing the values of the transition probabilities, but rather we want to know which events are likely to occur (over a certain time interval) and which events have asymptotically negligible probability (as  $\epsilon$  goes to zero). For the latter case, a nonnumerical, graph-theoretic, method is more natural. This is accomplished by Algorithm II. Such a method can be also extracted from the results of [12].

**4. Cooling schedules for simulated annealing.** In simulated annealing [7, 10] we are given a set  $S = \{1, \dots, N\}$  of states together with a cost function  $J : S \mapsto \mathcal{N}$  to be minimized. (Our restriction that  $J$  takes integer values can be relaxed.) The algorithm jumps randomly from one state to another and forms a Markov chain with the

following transition probabilities:

$$(4.1) \quad P(X(t+1) = j | X(t) = i) \\ = Q(i, j) \exp[\min\{0, -(J(j) - J(i))/T(t)\}], \quad \text{if } j \neq i,$$

$$(4.2) \quad P(X(t+1) = i | X(t) = i) = 1 - \sum_{j \neq i} P(X(t+1) = j | X(t) = i),$$

where the kernel  $Q(i, j)$  is nonnegative and satisfies  $\sum_j Q(i, j) = 1$ , and  $T(t) > 0$  is the "temperature" at time  $t$ . It is known that if  $T(t)$  decreases to zero slowly enough, then  $X(t)$  converges (in probability) to the set at which  $J$  is minimized [1, 6-9, 11]. We are interested in determining how slowly  $T(t)$  must converge to zero, so that convergence to the minimizing states is obtained.

We formulate the problem in a slightly more general manner, as follows. Suppose that we are given a stochastic matrix  $P^\epsilon$  (whose  $ij$ th entry is denoted by  $p_{ij}^\epsilon$ ) parameterized by a positive parameter  $\epsilon$  and assume that there exist positive constants  $C_1, C_2$  and a collection  $\mathcal{A} = \{\alpha_{ij} : 1 \leq i, j \leq N\}$  such that  $\alpha_{ij} \in \mathcal{N}_0 \cup \{\infty\}$ ,  $\forall i, j$  and such that  $p_{ij}^\epsilon = 0$ , whenever  $\alpha_{ij} = \infty$ , and  $C_1 \epsilon^{\alpha_{ij}} \leq p_{ij}^\epsilon \leq C_2 \epsilon^{\alpha_{ij}}$ ,  $\forall \epsilon \in (0, 1]$ , whenever  $\alpha_{ij} < \infty$ . Finally, we are given a monotonically nonincreasing function (cooling schedule)  $\epsilon : \mathcal{N}_0 \rightarrow (0, 1)$ . We are interested in the Markov chain  $X(t)$  with transition probabilities given by  $P(X(t+1) = j | X(t) = i) = p_{ij}^{\epsilon(t)}$ .

Clearly, the simulated annealing algorithm is of the type described in the preceding paragraph, provided that we identify  $\epsilon(t)$  with  $e^{-1/T(t)}$  and provided that we define  $\alpha_{ij} = \infty$ , if  $Q(i, j) = 0$ ,  $i \neq j$ , and  $\alpha_{ij} = \max\{0, J(j) - J(i)\}$ , if  $Q(i, j) \neq 0$ ,  $i \neq j$ . The coefficient  $\alpha_{ii}$  may be specified similarly.

We now return to our general formulation. We thus assume that  $\mathcal{A}, C_1, C_2$  are given, together with the schedule  $\{\epsilon(t)\}$ . We assume that  $\mathcal{A}$  satisfies Assumption TRI and we define, for any  $d \in \mathcal{N}_0$ , the quantities  $V^d(i, j)$  and the sets  $R^d$  by means of Algorithm II of §3. Our main result is the following.

PROPOSITION 4.1. *Assume that for some integer  $d \geq 0$ ,*

$$(4.3) \quad \sum_{t=0}^{\infty} \epsilon^d(t) = \infty,$$

$$(4.4) \quad \sum_{t=0}^{\infty} \epsilon^{d+1}(t) < \infty.$$

Then

- (i)  $\lim_{t \rightarrow \infty} P(X(t) \in R^d | X(0) = i) = 1, \forall i$ .
- (ii) For any  $i \in R^d$ ,  $\limsup_{t \rightarrow \infty} P(X(t) = i | X(0) = i) > 0$ .

PROOF. The main idea of the proof is to partition  $[0, \infty)$  into a set of disjoint time intervals  $[t_k, t_{k+1})$  such that  $X(t)$  is approximately stationary during each such interval, in the sense of §2, and then use the estimates available for such Markov chains. To simplify notation, if  $t$  is not an integer, we define  $\epsilon(t)$  to be equal to  $\epsilon(\lfloor t \rfloor)$ .

The proof for the case  $d = 0$  is rather easy and is omitted. We present the comparatively harder proof for the case  $d \geq 1$ .



We start with the proof of part (i) of the proposition. We define  $t_0 = 0$  and

$$(4.5) \quad t_{k+1} = t_k + \frac{1}{\epsilon^{d-1}(t_k)}, \quad \text{if } \epsilon\left(t_k + \frac{1}{\epsilon^{d-1}(t_k)}\right) \geq \frac{1}{2}\epsilon(t_k),$$

$$(4.6) \quad t_{k+1} = \min\{t: \epsilon(t) \leq \frac{1}{2}\epsilon(t_k)\}, \quad \text{otherwise.}$$

We define  $A_L$  (respectively,  $A_S$ ) as the set of all  $k$ 's such that  $t_{k+1}$  is defined by (4.5) (respectively, (4.6)).

LEMMA 4.1. *The sequence  $\{\epsilon(t_k)\}$  has the following properties:*

$$(4.7) \quad \frac{1}{2}\epsilon(t_k) \leq \epsilon(t) \leq \epsilon(t_k), \quad \forall t \in [t_k, t_{k+1}),$$

$$(4.8) \quad \sum_{k \in A_L} \epsilon(t_k) = \infty,$$

$$(4.9) \quad \sum_{k=0}^{\infty} \epsilon^2(t_k) < \infty.$$

Let  $f(k, l)$  be the cardinality of  $A_L \cap \{l, \dots, k-1\}$ , for  $k \geq l$ . Then, for any  $C \in (0, 1)$ ,

$$(4.10) \quad \sum_{k=0}^{\infty} \sum_{l=0}^k (1-C)^{f(k,l)} \epsilon(t_k) \epsilon(t_l) < \infty,$$

$$(4.11) \quad \lim_{k \rightarrow \infty} \sum_{l=0}^k (1-C)^{f(k,l)} \epsilon(t_l) = 0.$$

PROOF. Inequality (4.7) is an immediate consequence of (4.5), (4.6).

We notice that for any  $k \in A_S$ ,  $k' \in A_S$ , with  $k' > k$ , we have  $\epsilon(t_{k'}) \leq (1/2)\epsilon(t_k)$ . Hence,

$$(4.12) \quad \sum_{k \in A_S} \epsilon(t_k) \leq \epsilon(0) \sum_{k=0}^{\infty} 2^{-k} < \infty.$$

Furthermore,

$$\begin{aligned} \sum_{k \in A_L} \epsilon(t_k) &= \sum_{k \in A_L} \epsilon^d(t_k) [t_{k+1} - t_k] \\ &= \sum_{k=0}^{\infty} \epsilon^d(t_k) [t_{k+1} - t_k] - \sum_{k \in A_S} \epsilon^d(t_k) [t_{k+1} - t_k] \\ &\geq \sum_{t=0}^{\infty} \epsilon^d(t) - \sum_{k \in A_S} \epsilon(t_k) = \infty, \end{aligned}$$

which proves (4.8).

From (4.12) we conclude that  $\sum_{k \in A_S} \epsilon^2(t_k) < \infty$ . Also,

$$\sum_{k \in A_L} \epsilon^2(t_k) \leq \sum_{k=0}^{\infty} \epsilon^{d+1}(t_k) [t_{k+1} - t_k] \leq 2^{d+1} \sum_{t=0}^{\infty} \epsilon^{d+1}(t) < \infty,$$

which proves (4.9).



Given any  $C \in (0, 1)$ , we define a constant  $a$  by  $[2(1 - C)]^a = 3/2$ , if  $2(1 - C) > 1$ ; otherwise, we let  $a = 1$ . Let  $B = \{(k, l): k \geq l \text{ and } f(k, l) \geq a(k - l)\}$ . Then

$$\sum_{(k, l) \in B} (1 - C)^{f(k, l)} \epsilon(t_k) \epsilon(t_l) \leq \sum_{k=0}^{\infty} \sum_{l=0}^k [(1 - C)^a]^{k-l} \epsilon(t_k) \epsilon(t_l) < \infty,$$

because  $(1 - C)^a < 1$  and  $\epsilon(t_k)$  is square summable by (4.9). Now notice that  $\epsilon(t_k) \leq 2^{-(k-l)+f(k, l)} \epsilon(t_l)$ , if  $k \geq l$ . Hence,

$$\begin{aligned} \sum_{(k, l) \notin B, k \geq l} (1 - C)^{f(k, l)} \epsilon(t_k) \epsilon(t_l) &\leq \sum_{(k, l) \notin B, k \geq l} [2(1 - C)]^{f(k, l)} 2^{-(k-l)} \epsilon^2(t_l) \\ &\leq \sum_{k \geq l} (3/2)^{k-l} (1/2)^{k-l} \epsilon^2(t_l) < \infty, \end{aligned}$$

which proves (4.10). The proof of (4.11) is similar and is omitted. ■

We now define subsets  $S_0, S_1, \dots$  of the state space inductively as follows.

$$S_0 = R^d = \{i: \text{if } V^d(i, j) = 0 \text{ then } V^d(j, i) = 0\},$$

$$S_{n+1} = \{i \in R^{d-1}: i \notin S_0 \cup \dots \cup S_n \text{ and } \exists j \in S_n \text{ such that } V^{d-1}(i, j) = 1\}.$$

Also let

$$T_0 = \{i \in TR^{d-1}: \exists j \in S_0 \text{ such that } V^{d-1}(j, i) = 1\}$$

and let  $T_1$  be the complement of  $T_0$  in  $TR^{d-1}$ . Notice that  $(\bigcup_{n \geq 0} S_n) \cup T_0 \cup T_1 = \{1, \dots, N\}$ . Also, if  $i \in S_n$ ,  $n \neq 0$  and  $V^{d-1}(i, j) = 0$ , then  $j \in R_i^{d-1}$  and  $j \in S_n$ . (For a proof of this fact, if  $i \in S_n$ , then  $i \in R^{d-1}$ ; so, if  $V^{d-1}(i, j) = 0$ , then  $V^{d-1}(j, i) = 0$  and therefore  $j \in R_i^{d-1}$ . Let  $l \in S_{n-1}$  be such that  $V^{d-1}(i, l) = 1$ . Then  $V^{d-1}(j, l) = 1$ . So, either  $j \in S_n$  and we are done, or  $j \in S_0 \cup \dots \cup S_{n-1}$ . In the second case, the same argument shows that  $i \in S_0 \cup \dots \cup S_{n-1}$  which is a contradiction.)

We let  $Y(k) = X(t_k)$ . We need estimates on the transition probabilities of the  $Y(k)$  process. These are obtained by noting that, for any  $k$ , the Markov chain  $\{X(t): t \in [t_k, t_{k+1}]\}$  belongs to  $\mathcal{M}_{\epsilon(t_k)}(\mathcal{A}, 2^{-K}C_1, C_2, 0)$ , where  $K = \max\{\alpha_{ij}; \alpha_{ij} < \infty\}$ . Since  $t_{k+1} - t_k \leq 1/(\epsilon^{d-1}(t_k))$ , Corollary 3.1 may be used to obtain upper bounds. Also, for  $k \in A_L$ ,  $t_{k+1} - t_k = 1/(\epsilon^{d-1}(t_k))$  and therefore Proposition 3.1 may be used to obtain lower bounds. In more detail, we have:

LEMMA 4.2. *There are constants  $F > 0$ ,  $G > 0$ , such that, for every  $k \in \mathcal{N}_0$  we have*

$$(4.13) \quad \text{(i) If } k \in A_L, \text{ then } P(Y(k+1) \in S_n | Y(k) \in S_{n+1}) \geq F\epsilon(t_k), \forall n.$$

$$(4.14) \quad \text{(ii) } P(Y(k+1) \notin S_n | Y(k) \in S_n) \leq G\epsilon(t_k), \forall n.$$

$$(4.15) \quad \text{(iii) } P(Y(k+1) \notin S_0 \cup T_0 | Y(k) \in S_0) \leq G\epsilon^2(t_k).$$

$$(4.16) \quad \text{(iv) } P(Y(k+1) \notin S_0 \cup T_0 | Y(k) \in T_0) \leq G\epsilon(t_k).$$

$$(4.17) \quad \text{(v) } P(Y(k+1) \in T_0 | Y(k) \in S_0) \leq G\epsilon(t_k).$$

$$(4.18) \quad \text{(vi) If } k \in A_L, \text{ then } P(Y(k+1) \in S_0 | Y(k) \in T_0) \geq F.$$

$$(4.19) \quad \text{(vii) If } k \in A_L, \text{ then, for all } i,$$

$$P(Y(k+1) \in TR^{d-1} | Y(k) = i) \leq 1 - F.$$

PROOF. (i) If  $i \in S_{n+1}$ , then (by definition) there is some  $j \in S_n$  such that  $V^{d-1}(i, j) = 1$ . The result follows from the lower bound in (3.2).

(ii) Let  $i \in S_n$ ,  $j \notin S_n$ . We have shown earlier that we must have  $V^{d-1}(i, j) \geq 1$  and the result follows from (3.3).

(iii) Let  $i \in S_0$  and  $j \notin S_0 \cup T_0$ . If  $j \in S_n$ ,  $n \neq 0$ , then  $j \notin R^d$ ; hence  $V^d(i, j) \geq 1$ . Therefore, using the definition of  $V^d$ , we have  $1 \leq V^d(i, j) \leq V^{d-1}(i, j) - 1$ . Hence  $V^{d-1}(i, j) \geq 2$ . Finally, if  $j \in T_1$ , then  $V^{d-1}(i, j) \geq 2$ , because otherwise we would have  $j \in T_0$ . The result follows from (3.3).

(iv) Let  $i \in T_0$  and choose some  $l \in S_0$  such that  $V^{d-1}(l, i) = 1$  (which exists by the definition of  $T_0$ ). Suppose that  $j \notin S_0 \cup T_0$ . If  $j \in S_n$ ,  $n \neq 0$ , then  $V^{d-1}(i, j) \geq 1$ , because otherwise  $V^{d-1}(l, j) = 1$ , which contradicts the discussion in the proof of part (iii). So, for this case the result follows from (3.3). Suppose now that  $j \in T_1$ . For any  $c \leq d-1$  we must have  $V^c(i, j) \geq 1$  because otherwise (using Proposition 3.2)  $V^{d-1}(l, j) \leq V^{d-1}(l, i) + V^c(i, j) = 1$ , which contradicts the assumption  $j \in T_1$ . The result follows from (3.4).

(v) This is immediate from  $V^{d-1}(i, j) \geq 1$ ,  $\forall i \in R^{d-1}$ ,  $\forall j \in TR^{d-1}$  (Proposition 2.2, part (ii)).

(vi) Let  $i \in T_0$ . Since  $i \in TR^{d-1}$ , there exists some  $j \in R^{d-1}$  such that  $V^{d-1}(i, j) = 0$ . By the previous discussion, such a  $j$  cannot belong to  $S_n$ , for  $n \geq 1$ . The result follows from (3.2).

(vii) Similarly, for any  $i$  there exists some  $j \in R^{d-1}$  such that  $V^{d-1}(i, j) = 0$  and the result follows from (3.2). ■

Let

$$H_k = P(Y(n) \in S_0 \cup T_0, \forall n \in [0, k] | Y(0) \in S_0),$$

$$Q_k = P(Y(k) \in T_0 | Y(n) \in S_0 \cup T_0, \forall n \in [0, k-1], Y(0) \in S_0).$$

Using (4.17), (4.18), we obtain  $Q_{k+1} \leq G\epsilon(t_k) + (1 - \chi_k F)Q_k$ , where  $\chi_k = 1$  if  $k \in A_L$  and  $\chi_k = 0$ , otherwise. So,

$$Q_k \leq \frac{G}{1-F} \sum_{l=0}^k \epsilon(t_l) (1-F)^{f(k,l)}.$$

Using (4.15), (4.16),

$$(4.20) \quad H_{k+1} \geq [1 - G\epsilon(t_k)Q_k - G\epsilon^2(t_k)]H_k.$$

Now,  $\epsilon(t_k)Q_k$  is summable, by (4.10); also,  $\epsilon^2(t_k)$  is summable, by (4.9). Since  $H_0 > 0$ , it follows that there exists some  $H > 0$  such that  $H_k > H, \forall k$ . Using (4.13),

$$\begin{aligned} 1 &\geq P(Y(n) \in S_0 \cup T_0) \\ &\geq \sum_{k=0}^{n-1} P(Y(k) \notin S_0, Y(l) \in S_0 \cup T_0, \forall l \in \{k+1, \dots, n\}) \\ &\geq \sum_{k=0}^{n-2} P(Y(l) \in S_0 \cup T_0, \forall l \in \{k+2, \dots, n\} | Y(k+1) \in S_0) \\ &\quad \times P(Y(k+1) \in S_0 | Y(k) \in S_1) P(Y(k) \in S_1) \\ &\geq HF \sum_{\{k \in A_L: 0 \leq k \leq n-2\}} \epsilon(t_k) P(Y(k) \in S_1). \end{aligned}$$

So,  $\sum_{k \in A_L} \epsilon(t_k) P(Y(k) \in S_1) < \infty$ . Since  $\sum_{k \in A_S} \epsilon(t_k) < \infty$ , it follows that

$$\sum_{k=0}^{\infty} \epsilon(t_k) P(Y(k) \in S_1) < \infty.$$

We will prove by induction that for all  $n \geq 1$ ,

$$(4.21) \quad \sum_{k=0}^{\infty} \epsilon(t_k) P(Y(k) \in S_n) < \infty.$$

(We already proved the case  $n = 1$ .) Using (4.13), (4.14), we have

$$(4.22) \quad P(Y(k+1) \in S_n) \geq P(Y(k) \in S_n) - G\epsilon(t_k)P(Y(k) \in S_n) \\ + \chi_k F\epsilon(t_k)P(Y(k) \in S_{n+1}).$$

By telescoping the inequality (4.22) and using the induction hypothesis (4.21), we see that  $\sum_{k=0}^{\infty} \chi_k \epsilon(t_k) P(Y(k) \in S_{n+1}) < \infty$ . Also,  $\sum_{k \in A_S} \epsilon(t_k) P(Y(k) \in S_{n+1}) \leq \sum_{k \in A_S} \epsilon(t_k) < \infty$  (because of (4.12)) which completes the induction step. Summing inequality (4.21), over all  $n \geq 1$ , we obtain

$$\sum_{k=0}^{\infty} \epsilon(t_k) P(Y(k) \in \bigcup_{n \geq 1} S_n) < \infty.$$

Using the fact that  $\epsilon(t_k)$  sums to infinity, we conclude that  $\liminf_{k \rightarrow \infty} P(Y(k) \in \bigcup_{n \geq 1} S_n) = 0$  and therefore,

$$\limsup_{k \rightarrow \infty} P(Y(k) \in S_0 \cup TR^{d-1}) = 1.$$

We show next that the probability of the process being in transient states goes to zero. Inequalities (4.14) and (4.19) imply

$$P(Y(k+1) \in TR^{d-1}) \leq G\epsilon(t_k) + (1 - \chi_k F)P(Y(k) \in TR^{d-1}).$$

Thus,

$$P(Y(k) \in TR^{d-1}) \leq (1 - F)^{f(k,0)} + \frac{G}{1 - F} \sum_{l=0}^k (1 - F)^{f(k,l)} \epsilon(t_l),$$

which converges to 0, as  $k$  tends to infinity, due to (4.11). We may thus conclude that  $\limsup_{k \rightarrow \infty} P(Y(k) \in S_0) = 1$ . By repeating the argument that led to (4.20) we can see that the probability that  $Y$  ever exits  $S_0 \cup T_0$ , given that  $Y(k) \in S_0$ , converges to zero, as  $k \rightarrow \infty$ . (This is a consequence of the square summability of  $\epsilon(t_k)$ .) It follows that  $\lim_{k \rightarrow \infty} P(Y(k) \in S_0) = 1$ . Finally, for any  $t \in [t_k, t_{k+1}]$  we have  $P(X(t) \in S_0) \geq P(Y(k) \in S_0) - G\epsilon(t_k)$ , which converges to 1, as  $k \rightarrow \infty$ . This completes the proof of part (i) of the proposition.

For part (ii) of the proposition, in order to avoid introducing new notation, we prove the equivalent statement that if  $\sum_{t=0}^{\infty} \epsilon^d(t) < \infty$ , then  $\limsup_{t \rightarrow \infty} P(X(t) = i | X(0) = i) > 0$ ,  $\forall i \in R^{d-1}$ . So, let  $i \in R^{d-1}$  and consider the set  $R_i^{d-1}$ . For any  $j \notin R_i^{d-1}$ , we have  $V^{d-1}(i, j) \geq 1$  and, therefore (using Corollary 3.1), there exists some  $G > 0$  such that

$$P(Y(k+1) \notin R_i^{d-1} | Y(k) \in R_i^{d-1}) \leq G\epsilon(t_k), \quad \forall k.$$

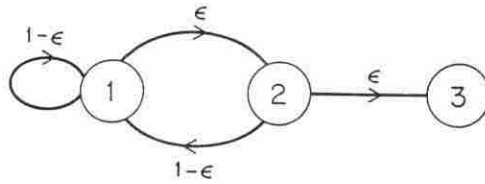


FIGURE 2. A Counterexample to Proposition 4.1.

Since we are assuming that  $\sum_{t=0}^{\infty} \epsilon^d(t) < \infty$ , it follows (as in the proof of (4.9)), that  $\sum_{k=0}^{\infty} \epsilon(t_k) < \infty$ . Consequently,

$$(4.23) \quad \inf_k P(Y(k) \in R_i^{d-1} | Y(0) = i) > 0.$$

Finally, for any  $j \in R_i^{d-1}$  we have  $V^{d-1}(j, i) = 0$ . Hence, using Proposition 3.1, there exists some  $F > 0$  such that

$$(4.24) \quad P(Y(k+1) = i | Y(k) \in R_i^{d-1}) \geq F.$$

By combining (4.23), (4.24), we obtain the desired result. ■

REMARKS. 1. With a little more effort along the lines of (4.23), (4.24) it can be shown that, under (4.4), we have  $\inf_t P(X(t) = i | X(0) = i) > 0, \forall i \in R^d$ .

2. Proposition 4.2 may be extended to the case where  $\mathcal{A}$  satisfies Assumption AP in a straightforward manner, using the method discussed in §2.

COROLLARY 4.1. *Let the transition probabilities for the simulated annealing algorithm be given by (4.1), (4.2). Consider cooling schedules of the form  $T(t) = c/\log t$ . For any initial state, the algorithm converges (in probability) to the set of global minima of  $J$  if and only if there exists some  $d$  such that the set of global minima contains  $R^d$  and  $c$  is larger than or equal to the smallest such  $d$ , to be denoted by  $d^*$ .*

PROOF. Having identified  $\exp[-1/T(t)]$  with  $\epsilon(t)$ , we see that  $\sum_{t=1}^{\infty} \epsilon^c(t) = \sum_{t=1}^{\infty} (1/t) = \infty$  and  $\sum_{t=1}^{\infty} \epsilon^{c+1}(t) = \sum_{t=1}^{\infty} (1/t^{(c+1)/c}) < \infty$ . Thus, by Proposition 4.1,  $R^c$  is the smallest set to which the algorithm converges (in probability). Thus, for convergence to the set of global minima, we need that set to contain  $R^c$ , which establishes the desired result. ■

A possibility for generalizing Proposition 4.1 arises if we allow the schedule  $\epsilon(t)$  to be nonmonotonic. In fact the proof goes through (with a minor modification in the definition of the sequence  $\{t_k\}$ ) if we assume that there exists some  $C > 0$  such that  $\epsilon(t) \leq C\epsilon(s), \forall t \geq s$ , which allows for mild nonmonotonicity. On the other hand, if  $\epsilon(t)$  is allowed to have more substantial variations, then the conclusions of Proposition 4.1 are no more true. For a simple example consider the Markov chain of Figure 2, together with the schedule  $\epsilon(t) = t^{-1/2}$ , if  $t$  is even, and  $\epsilon(t) = 1/t$ , if  $t$  is odd. For this schedule, the largest integer for which  $\sum_{t=0}^{\infty} \epsilon^d(t) = \infty$  is equal to 2. Also,  $R^2 = \{3\}$ . On the other hand,  $P(x(t) = 3 | x(0) = 1)$  does not converge to 1.

We have claimed that our result generalizes the results of [9] and we now elaborate on this claim. Hajek's result characterized  $d^*$  in an explicit manner, as the maximum depth<sup>2</sup> of local minima which are not global minima, under a "weak reversibility"

<sup>2</sup>The depth of a state  $i$  is defined as the minimum over all  $j$ , such that  $J(j) < J(i)$ , of the minimum over all paths leading from  $i$  to  $j$ , of the maximum of  $J(k) - J(i)$ , over all  $k$ 's belonging to that path; the depth of  $i$  is infinite if no such  $j$  exists.

assumption, which is equivalent to imposing certain restrictions on the structure  $\mathcal{A}$ . Our characterization is less explicit because instead of describing  $d^*$  we give an algorithm for computing it in terms of  $\mathcal{A}$ . Nevertheless, for the class of structures  $\mathcal{A}$  considered in [9], we can use our Algorithm II to show that  $R^d$  is the set of all local minima of the cost function  $J$ , of depth  $d + 1$ , or more. Hence, the  $d^*$  produced by our approach is the smallest  $d$  such that all local (but not global) minima have depth  $d$  or less, which agrees with the result of [9]. However, carrying out this argument rigorously turns out to be quite involved [14].

Let us look again at the example of Figure 1. We may think of it as being a Markov chain representing a simulated annealing algorithm, for the cost function  $J$  defined by  $J(i) = i$ . (The reversibility assumption made in [9] is violated for this example.) Notice that 1 is the unique optimal state and recall that  $R = R^0 = \{1, 2\}$ ,  $R^1 = \{1\}$ . Thus,  $d^* = 1$  is the smallest value of  $d$  for which  $R^d$  is contained in the set of optimal states. We therefore have convergence (in probability) to state 1, for every initial state, if and only if  $\sum_{t=1}^{\infty} \epsilon(t) = \infty$ . However, this example may be misleading: in general, without some form of reversibility, there will exist no  $d$  such that  $R^d$  will be contained in the set of optimal states.

**Acknowledgements.** The author is grateful to a referee for an extremely careful reading of the manuscript which caught several deficiencies and a gap in a proof and for several useful suggestions.

### References

- [1] Chiang, T.-S., Hwang, C.-R. and Sheu, S.-J. (1987). Diffusions for Global Optimization in  $R^n$ . *SIAM J. Control Optim.* **25** 737–753.
- [2] Coderch, M. (1982). Multiple Time Scale Approach to Hierarchical Aggregation of Linear Systems and Finite State Markov Processes. Ph.D. Thesis, Dept. of Electrical Engineering, M.I.T.
- [3] ———, Willisky, A. S., Sastry, S. S. and Castanon, D. A. Hierarchical Aggregation of Singularly Perturbed Finite State Markov Processes. *Stochastics* **8** 259–289.
- [4] Courtois, P. J. (1977). *Decomposability: Queuing and Computer System Applications*, Academic Press, New York.
- [5] Debecque, F. (1983). A Reduction Process for Perturbed Markov Chains. *SIAM J. Appl. Math.* **43** 2.
- [6] Gelfand, S. B., Mitter, S. K. (1985). Analysis of Simulated Annealing for Optimization. *Proc. 24th IEEE Conf. Decision and Control*, Ft. Lauderdale, FL, pp. 779–786.
- [7] Geman, S. and Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Trans. Pattern Analysis and Machine Intelligence* **6** 721–741.
- [8] Gidas, B. (1985). Non-Stationary Markov Chains and Convergence of the Annealing Algorithm. *J. Statist. Physics* **39** 73–131.
- [9] Hajek, B. (1988). Cooling Schedules for Optimal Annealing. *Math. Oper. Res.* **13** 311–329.
- [10] Kirkpatrick, S., Gelatt, C. D., Jr. and Vecchi, M. P. (1983). Optimization by Simulated Annealing. *Science* **220** 671–680.
- [11] Mitra, D., Romeo, F. and Sangiovanni-Vincentelli, A. (1985). Convergence and Finite-Time Behavior of Simulated Annealing. *Proc. 24th IEEE Conf. Decision and Control*, Ft. Lauderdale, FL, 761–767.
- [12] Rohlicek, J. R. and Willisky, A. S. (1988). The Reduction of Perturbed Markov Generators: An Algorithm Exposing the Role of Transient States. *J. Assoc. Comput. Mach.* **35** 675–696.
- [13] Varga, R. S. (1962). *Matrix Iterative Analysis*. Prentice-Hall, Englewood Cliffs, NJ.
- [14] Tsitsiklis, J. N. (1988). A Survey of Large Time Asymptotics of Simulated Annealing Algorithms. in W. Fleming, P. L. Lions (Eds.), *Stochastic Differential Systems, Stochastic Control Theory and Applications*. Springer-Verlag, New York, 583–599.