MATHEMATICS OF OPERATIONS RESEARCH
Vol. 13, No. 2, May 1988
Printed in U.S.A.

COOLING SCHEDULES FOR OPTIMAL ANNEALING*†

BRUCE HAJEK

University of Illinois at Champaign-Urbana

A Monte Carlo optimization technique called "simulated annealing" is a descent algorithm modified by random ascent moves in order to escape local minima which are not global minima. The level of randomization is determined by a control parameter T, called temperature, which tends to zero according to a deterministic "cooling schedule". We give a simple necessary and sufficient condition on the cooling schedule for the algorithm state to converge in probability to the set of globally minimum cost states. In the special case that the cooling schedule has parametric form $T(t) = c/\log(1 + t)$, the condition for convergence is that c be greater than or equal to the depth, suitably defined, of the deepest local minimum which is not a global minimum state.

1. Introduction. Suppose that a function V defined on some finite set $\mathcal{S}$ is to be minimized. We assume that for each state x in $\mathcal{S}$ that there is a set $N(x)$, with $N(x) \subset \mathcal{S}$, which we call the set of neighbors of x. Typically the sets $N(x)$ are small subsets of $\mathcal{S}$. In addition, we suppose that there is a transition probability matrix R such that $R(x, y) > 0$ if and only if $y \in N(x)$.

Let $T_1, T_2, \ldots$ be a sequence of strictly positive numbers such that

$$T_1 \geqslant T_2 \geqslant \ldots \quad \text{and} \tag{1.1}$$

$$\lim_{k\to\infty} T_k = 0. \tag{1.2}$$

Consider the following sequential algorithm for constructing a sequence of states $X_0, X_1, \ldots$. An initial state X_0 is chosen. Given that $X_k = x$, a potential next state Y_k is chosen from $N(x)$ with probability distribution $P[Y_k = y | X_k = x] = R(x, y)$. Then we set

$$X_{k+1} = \begin{cases} Y_k & \text{with probability } p_k, \\ X_k & \text{otherwise,} \end{cases} \quad \text{where}$$

$$p_k = \exp\left[\frac{-[V(Y_k) - V(x)]^+}{T_k}\right].$$

This specifies how the sequence $X_1, X_2, \ldots$ is chosen. Let $\mathcal{S}^*$ denote the set of states in $\mathcal{S}$ at which V attains its minimum value. We are interested in determining

*Received March 13, 1985; revised December 19, 1986.
AMS 1980 subject classification. Primary: 65C05; Secondary: 60J27.
IAOR 1973 subject classification. Main: Probabilistic Programming; Cross References: Simulation.
OR/MS Index 1978 subject classification. Primary: 663 Stochastic methods of optimization; Secondary: 761 Simulation.
Key words. Simulated annealing, Randomized optimization, Monte Carlo methods of optimization.
†This work was supported by the Office of Naval Research under Contract N000-14-82-K-0359 and the National Science Foundation under contract NSF-ECS-83-52030.

0364-765X/88/1302/0311$01.25

whether

$$\lim_{k\to\infty} P[X_k \in \mathcal{S}^*] = 1. \tag{1.3}$$

The random process $X = (X_k : k \geq 0)$ produced by the algorithm is a discrete time Markov chain. The one-step transition probability matrix at step k is

$$P_k(x, y) = P[X_{k+1} = y | X_k = x]$$

$$= \begin{cases} 0 & \text{if } y \notin N(x) \text{ and } y \neq x, \\ R(x, y)\exp(-[V(y) - V(x)]^+/T_k) & \text{if } y \in N(x) \text{ and } y \neq x, \\ 1 - \sum_{z \neq x} P_k(x, z) & \text{if } y = x. \end{cases}$$

We will motivate the choice of transition probabilities in the algorithm by briefly considering the algorithm under three simplifying assumptions. A state y is *reachable* from state x if $x = y$ or if there is a sequence of states $x = x_0, x_1, \ldots, x_p = y$ for some $p \geq 1$ such that $x_{k+1} \in N(x_k)$ for $0 \leq k < p$. The first assumption is that $(\mathcal{S}, N)$ is irreducible, by which we mean that given any two states x and y, y is reachable from x. The second assumption is that T_k is equal to T, for some constant $T > 0$. The third assumption is that R is reversible, which, by definition, means that there is a probability distribution α on $\mathcal{S}$, necessarily the equilibrium distribution for R, such that $\alpha(x)R(x, y) = \alpha(y)R(y, x)$ for all x, y in $\mathcal{S}$. A simple example for which the third assumption is valid is the case that

$$R(x, y) = \begin{cases} \dfrac{1}{|N(x)|} & \text{if } y \in N(x), \\ 0 & \text{otherwise,} \end{cases}$$

and the neighbor system is symmetric in the sense that $x \in N(y)$ if and only if $y \in N(x)$ for each pair of states x, y.

By the assumption that $T_k = T$ for all k, the Markov chain X has a stationary one-step transition probability matrix P, $P = P_k$ for all k. It then easily follows from the reversibility assumption on R that, if we define a probability distribution π_T on $\mathcal{S}$ by

$$\pi_T(x) = \alpha(x)\exp\left(-\frac{V(x)}{T}\right)\Big/ Z_T, \quad \text{where} \quad Z_T = \sum_x \alpha(x)\exp\left[-\frac{V(x)}{T}\right],$$

then P is reversible with equilibrium distribution π_T.

By the assumption that $(\mathcal{S}, N)$ is irreducible (and the fact that P is aperiodic if $\mathcal{S}^* \neq \mathcal{S}$ since $P(i, i) > 0$ for some i in that case), the Markov ergodic convergence theorem [6] implies that

$$\lim_{k\to\infty} P[X_k \in \mathcal{S}^*] = \sum_{x \in \mathcal{S}^*} \pi_T(x). \tag{1.4}$$

Examination of π_T soon yields that the right-hand side of (1.4) can be made arbitrarily close to one by choosing T small. Thus,

$$\lim_{T\to 0}\left[\lim_{k\to\infty,\, T_k = T} P[X_k \in \mathcal{S}^*]\right] = 1.$$

The idea of the simulated annealing algorithm is to try to achieve (1.3) by letting T_k tend to zero as k (time) tends to infinity.

We now return to the original case that the sequence (T_k) is nonincreasing and has limit zero. We will not require that R be reversible. Instead, a much weaker assumption will be made with the help of the following definition. We say that state y is reachable at height E from state x if $x = y$ and $V(x) \leqslant E$, or if there is a sequence of states $x = x_0, x_1, \ldots, x_p = x$ for some $p \geqslant 1$ such that $x_{k+1} \in N(x_k)$ for $0 \leqslant k < p$ and $V(x_k) \leqslant E$ for $0 \leqslant k \leqslant p$. We will assume that $(\mathscr{S}, V, N)$ has the following property.

Property WR (*Weak reversibility*): For any real number E and any two states x and y, x is reachable at height E from y if and only if y is reachable at height E from x.

State x is said to be a *local minimum* if no state y with $V(y) < V(x)$ is reachable from x at height $V(x)$. We define the *depth* of a local minimum x to be plus infinity if x is a global minimum. Otherwise, the depth of x is the smallest number E, $E > 0$, such that some state y with $V(y) < V(x)$ can be reached from x at height $V(x) + E$. These definitions are illustrated in Figure 1.1.

We define a *cup* for $(\mathscr{S}, V, N)$ to be a set C of states such that for some number E, the following is true: For every $x \in C$, $C = \{y\colon\ y$ can be reached at height E from $x\}$. For example, by Property *WR*, if $E \geqslant V(x)$ then the set of states reachable from x at height E is a cup. Given a cup C, define $\underline{V}(C) = \min\{V(x)\colon\ x \in C\}$ and $\overline{V}(C) = \min\{V(y)\colon\ y \notin C$ and $y \in N(x)$ for some x in $C\}$. The set defining $\overline{V}(C)$ is empty if and only if $C = \mathscr{S}$, and we set $\overline{V}(\mathscr{S}) = +\infty$. We call the subset B of C defined by $B = \{x \in C\colon V(x) = \underline{V}(C)\}$ the *bottom* of the cup, and we call the number $d(C)$ defined by $d(C) = \overline{V}(C) - \underline{V}(C)$ the *depth* of the cup. These definitions are

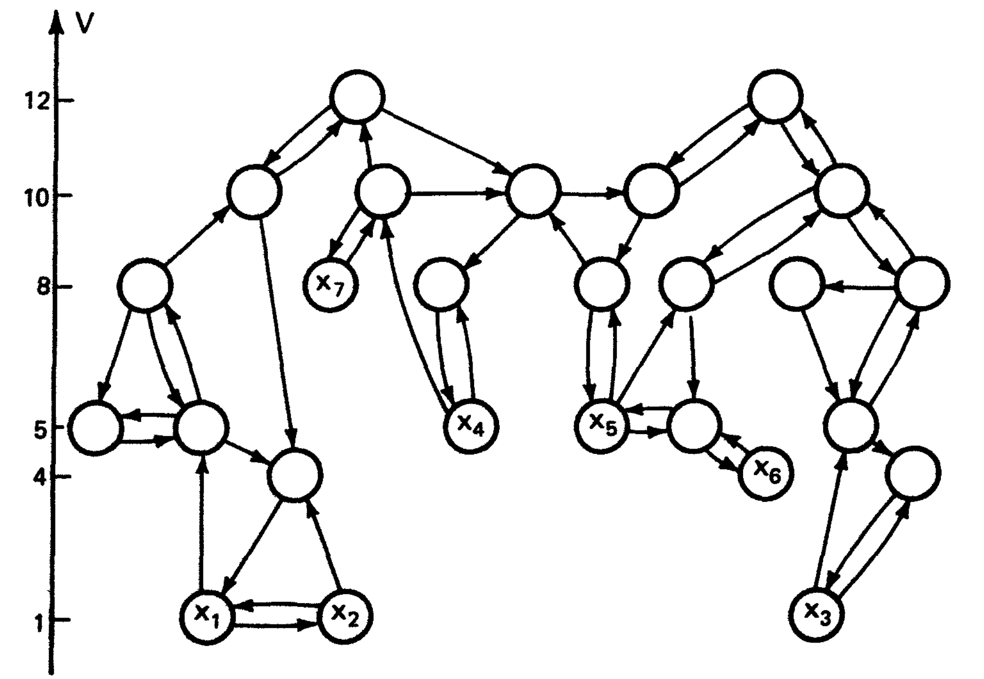

FIGURE 1.1. The graph pictured arises from a triplet $(\mathscr{S}, V, N)$. Nodes correspond to elements in $\mathscr{S}$. $V(x)$ for x in $\mathscr{S}$ is indicated by the scale at left. Arcs in the graph represent ordered pairs of states (x, y) such that $x \in N(y)$. Property *WR* is satisfied for the example shown.

States x_1, x_2 and x_3 are global minimum. States x_4, x_6 and x_7 are local minima of depths 5.0, 6.0, and 2.0, respectively. State x_5 is not a local minimum. State x_2 is reachable at height 1.0 from x_1 and states x_3 is reachable at height 12.0 from x_1.

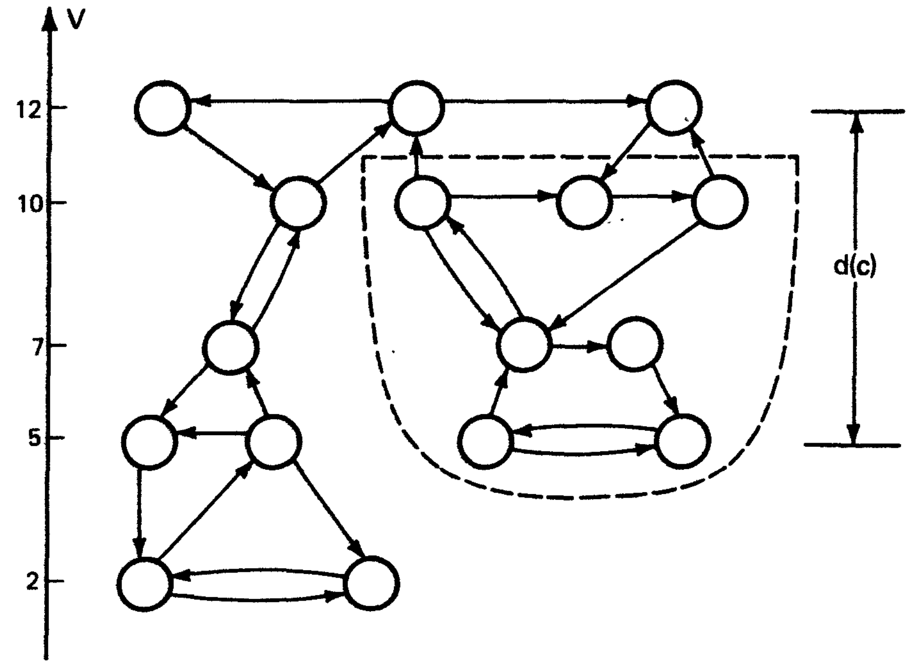

FIGURE 1.2. A cup C is enclosed with dashed lines. $\underline{V}(C) = 5$, $\overline{V}(C) = 12$, $d(C) = 7$ and the bottom B of C contains two states.

illustrated in Figure 1.2. Note that a local minimum of depth d is an element of the bottom of some cup of depth d.

THEOREM 1. *Assume that* $(\mathscr{S}, V, N)$ *is irreducible and satisfies WR, and that* (1.1) *and* (1.2) *hold.*

(a) *For any state x that is not a local minimum,* $\lim_{k\to\infty} P[X_k = x] = 0$.

(b) *Suppose that the set of states B is the bottom of a cup of depth d and that the states in B are local minima of depth d. Then* $\lim_{k\to\infty} P[X_k \in B] = 0$ *if and only if* $\sum_{k=1}^{\infty} \exp(-d/T_k) = +\infty$.

(c) (Consequence of (a) and (b)) *Let* d^* *be the maximum of the depths of all states which are local but not global minima. Let* $\mathscr{S}^*$ *denote the set of global minima. Then*

$$\lim_{k\to\infty} P[X_k \in \mathscr{S}^*] = 1 \tag{1.5}$$

if and only if

$$\sum_{k=1}^{\infty} \exp(-d^*/T_k) = +\infty. \tag{1.6}$$

REMARKS. (1) If T_k assumes the parametric form

$$T_k = \frac{c}{\log(k+1)} \tag{1.7}$$

then condition (1.6), and hence also condition (1.5), is true if and only if $c \geq d^*$. This result is consistent with the work of Geman and Geman [2]. They considered a model which is nearly a special case of the model used here, and they proved that condition (1.5) holds if (T_k) satisfies equation (1.7) for a sufficiently large constant c. They gave a value of c which is sufficient for convergence and which is substantially larger than d^*.

Gidas [4] also addressed the convergence properties of the annealing algorithm. The Markov chains that he considered are more general than those that we consider. He required little more than the condition that the one-step transition probability matrices P_k converge as k tends to infinity. In the special case of annealing processes, he gave a value of c (actually, c here corresponds to $1/C_0$ in Gidas' notation) which he conjectured is the smallest such that Equation (1.7) leads to Equation (1.5). His constant is different from the constant d^* defined here. Gidas also considered interesting convergence questions for functionals of the Markov chains.

Geman and Hwang [3] showed that in the analogous case of nonstationary diffusion processes that a schedule of the form (1.7) is sufficient for convergence to the global minima if c is no smaller than the difference between the maximum and minimum value of V. We conjecture that the smallest constant is given by the obvious analogue of the constant d^* that we defined here, although the proof that we give here does not readily carry over to the diffusion case.

(2) Some information on how quickly convergence occurs in Equation (1.5) can be gleaned from Theorems 3 and 4 in §§3 and 4 respectively, and from the proofs in [2-4].

(3) It would be interesting to know the behavior of $\min_{n \le k} V(X_n)$ rather than the behavior of $V(X_k)$.

(4) It would be interesting to study convergence properties when the schedule (T_k) is random and depends on the algorithm states.

(5) If x is a local minimum of depth d and no other local minimum y with $V(y) = V(x)$ can be reached from x at height $V(x) + d$, then there is a cup of depth d such that x is the only state in the bottom of the cup. Thus, under the conditions of Theorem 1, $\lim_{k \to \infty} P[X_k = x] = 0$ if and only if (1.6) is true. We conjecture that this last statement is true for any local minimum of depth d.

EXAMPLES. (1) The examples in Figure 1.1 and Figure 1.2, $d^* = 6$ and $d^* = 7$, respectively.

(2) Cerny [1] and Kirkpatrick et al. [7] independently introduced simulated annealing. They and, by now, many others have applied the simulated annealing algorithm to difficult combinatorial problems. We will briefly consider the maximum matching problem, even though this problem is easy in the sense that efficient polynomial-time algorithms are known for solving it.

Consider an undirected graph. A matching M is a subset of edges of the graph such that no two edges in M have a node in common. We take the state space $\mathcal{S}$ in the annealing setup to be the set of all matchings, and we let $V(M)$ be negative one times the number of edges in M. The maximum matching problem is then to find M in $\mathcal{S}$ to minimize $V(M)$. Suppose that the state of the annealing algorithm at the beginning of the kth iteration is M. Suppose an edge of the graph is then chosen at random, with all edges being equally likely. If the edge is not in M and can be added to M, then the next state is obtained by adding the edge to M. If the edge is in M, then the next state is obtained by removing the edge from M with probability $\exp(-1/T_k)$. Otherwise, the next state is again M. This corresponds to $R(M, M') = 1/L$ if M' can be obtained from M by adding an edge to or subtracting an edge from M, where L is the number of edges of the graph; and $R(M, M') = 0$ for other M' distinct from M.

It is well known that if a matching M does not have maximum cardinality then edges can be alternately subtracted and added until two can be added at once. Equivalently, d^* is at most one (it is zero for some graphs), no matter how large the graph is.

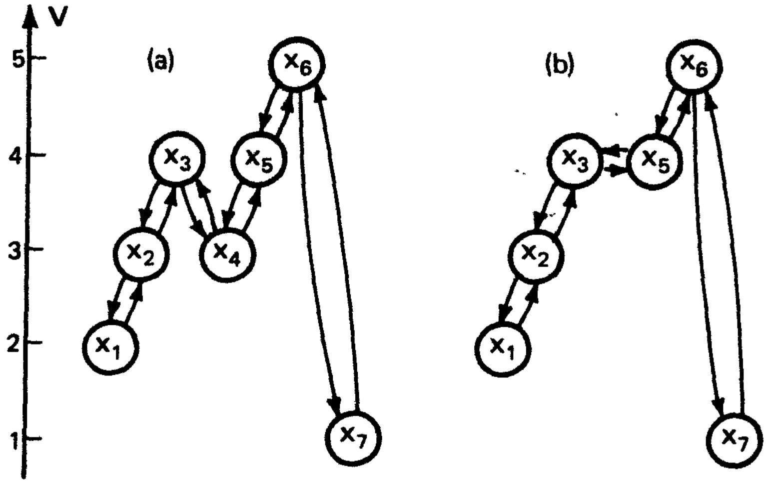

FIGURE 1.3. Diagram (a) arises from a triple $(\mathscr{S}, V, N)$ and diagram (b) is obtained by "filling-in" the cup $\{x_4\}$.

(3) Consider $(\mathscr{S}, V, N)$ giving rise to Figure 1.3. Note that $d^* = 3$. Now suppose that (T_k) satisfies (1.1), (1.2),

$$\sum_{k=1}^{\infty} \exp(-2/T_k) = +\infty, \quad \text{and} \tag{1.8}$$

$$\sum_{k=1}^{\infty} \exp(-3/T_k) < +\infty. \tag{1.9}$$

Since x_4 is reachable at height $V(x_1) + 2$ and since (1.8) is assumed, one might (correctly) guess that if the process starts at x_1 then it will eventually reach x_4 with probability one. By similar reasoning, one might then (incorrectly) guess that the process must eventually reach x_7. However, by Theorem 1 and (1.9), $\lim_{k\to\infty} P[X_k = x_7] \neq 1$. What happens is that if k is large so that T_k is small and if the process is in state x_4 at time k, then it is much more likely that the process hits state x_1 before it hits (if ever) state x_6. We think of the cup consisting of state x_4 alone as being "filled-in" (see Figure 1.3), so that to get from x_1 to x_6, the process has to climb up three levels. Roughly speaking, the small depression in V at x_4 does not allow the process to always make it up three levels by going up two at a time and "resting" in between. This would not be true if condition *WR* were violated by, for example, setting the probability of jumping from x_4 to x_3 to zero.

This paper is organized as follows. In §2 we state Theorem 2, which is a generalization of Theorem 1. The rest of the paper is devoted to proving Theorem 2. The theorem will first be proved under the assumption that $(\mathscr{S}, V, N)$ has the "continuous increase" property, which is formulated in §2. In §§3 and 4 we state and prove Theorems 3 and 4, which describe how the process exits from a cup. The proof of Theorem 2, including how to remove the continuous increase assumption, is presented in §5. There is an interplay in §5 between the necessity and sufficiency of condition (1.6).

2. Generalization of Theorem 1. It will be helpful to generalize Theorem 1 before proving it. We assume throughout the paper that $(\mathscr{S}, V, N)$ is irreducible and has the property *WR* defined in §1. Throughout the remainder of the paper $(\lambda_t\colon t \geq 0)$ will be assumed to satisfy $0 < \lambda_t < 1$, λ_t is nonincreasing in t, and $\lim_{t\to\infty} \lambda_t = 0$. Suppose that $(X_t\colon t \geqslant 0)$ is defined by $X_t = Y_k$ for $U_k \leqslant t < U_{k+1}$ where $((U_k, Y_k)\colon k \geqslant 0)$ is a homogeneous Markov random process with state space $[0, +\infty) \times \mathscr{S}$ and one-step transition probabilities

$$P\left[U_{k+1} \geqslant u, Y_{k+1} = y \middle| (U_i, Y_i)\colon 0 \leqslant i \leqslant k\right] = \int_u^\infty Q(Y_k, y, t, \lambda_t)\Phi(dt, U_k, Y_k),$$

where for some strictly positive constants D, a, c_1 and c_2:

A.1 $\Phi(\cdot, s, x)$ is a probability distribution function with $\Phi(s-, s, x) = 0$, $\int_0^\infty (t-s)\Phi(dt, s, x) \leqslant D$, and $E[\min\{k\colon U_k \geqslant t + a\} | U_0 = t,\ Y_0 = x] \leqslant 1/a$ for all $t \geqslant 0$, $x \in \mathscr{S}$.

A.2 For $t \geqslant 0$ and $0 \leqslant \lambda \leqslant 1$, $Q(\cdot, \cdot, t, \lambda)$ is a probability transition matrix with

$$c_1 \lambda^{(V(y) - V(x))^+} \leqslant Q(x, y, t, \lambda) \leqslant c_2 \lambda^{(V(y) - V(x))^+} \quad \text{if } y \in N(x)$$

$$Q(x, y, t, \lambda) = 0 \quad \text{if } y \notin N(x) \bigcup \{x\}.$$

Note that Y_{k+1} is conditionally independent of $(U_0, Y_0, \dots, U_{k-1}, Y_{k-1}, U_k)$ given (U_{k+1}, Y_k).

For example, if $(T_k\colon k \geqslant 0)$ and R are as in §1 and "I_A" denotes the indicator variable of a set A, we can let $\lambda_t = \exp(-1/T_{[t]})$, $\Phi(t, s, x) = I_{\{t \geqslant s+1\}}$ and $Q(x, y, t, \lambda) = R(x, y)\lambda_t^{-(V(y) - V(x))^+}$ for $x \neq y$. Then $(X_t\colon t \geqslant 0)$ sampled at integer times has the same transition probabilities as $(X_k\colon k \geqslant 0)$ does in §1. When λ, Φ and Q are chosen as in this example, Theorem 2 below reduces to Theorem 1.

If instead we let $\Phi(t, s, x) = I_{\{t \geqslant s\}}(1 - \exp(-(t-s)))$, then X is a continuous-time analogue of the process in §1.

THEOREM 2. (a) *For any state x that is not a local minimum,* $\lim_{t \to \infty} P[X_t = x] = 0$.

(b) *Suppose that B is the bottom of a cup C of depth d and that the states in B are local minima of depth d. Then* $\lim_{t \to \infty} P[X_t \in B] = 0$ *if and only if* $\int_0^\infty \lambda_t^d\, dt = +\infty$.

(c) (Consequence of (a) and (b)). *Let d^* be the maximum of the depths of all states which are local but not global minima. Let $\mathscr{S}^*$ denote the set of global minima. Then* $\lim_{t \to \infty} P[X_t \in \mathscr{S}^*] = 1$ *if and only if* $\int_0^\infty \lambda_t^{d^*}\, dt = +\infty$.

Let $E_0, E_1, \dots, E_q$ denote the possible values of $V(x)$ as x varies over $\mathscr{S}$, ordered so that $E_0 < E_1 < \cdots < E_q$. We will first prove Theorem 2 under the assumption that $(\mathscr{S}, V, N)$ has the following property:

Continuous Increase Property: Given any two states x and y, if $V(x) = E_i$ and $V(y) = E_j$ and $j \geqslant i + 2$, then $y \notin N(x)$.

The $(\mathscr{S}, V, N)$ in Figure 1.1 does not have the continuous increase property, while the one in Figure 1.2 does.

3. How cups runneth over.

3.1. *Theorem statement and initial steps of proof.* Suppose that C is a cup, and let m and k be the integers such that E_m is the minimum and E_k is the maximum value of $V(x)$, for x in C. We assume that $k < q$ (otherwise C is all of $\mathscr{S}$), so that $0 \leqslant m \leqslant k < q$. Let $F = \{y\colon y \notin C \text{ and } y \in N(x) \text{ for some } x \in C\}$. By the continuous increase property, $V(x) = E_{k+1}$ for all $x \in F$. Define u and g by $u = E_{k+1} - E_k$ and $g = E_k - E_m$. Then $d = u + g$, where d is the depth of C. Let $\Theta = (\mathscr{S}, V, N, D, a, c_1, c_2)$ where D, a, c_1 and c_2 are the constants appearing in Assumptions A.1 and A.2.

Theorem 3. *There exist $\epsilon > 0$ and $\bar{\lambda} > 0$ depending only on Θ and C so that for every time $t_o \geq 0$, every $x_o \in C$, every $y_o \in F$ and every $(\lambda_t\colon t \geq 0)$ such that $\lambda_{t_o} \leq \bar{\lambda}$ and $\int_{t_o}^{\infty} \lambda_s^d \, ds = +\infty$ the following conditions hold*:

$$E\left[\int_{t_o}^{\tau(C)} \lambda_s^d \, ds \,|\, (U_0, Y_0) = (t_o, x_o)\right] \leq \frac{1}{\epsilon} \tag{a}$$

where $\tau(C) = U_W$ *and* $W = \min\{k \geq 0\colon Y_k \in F\}$;

$$P\left[Y_W = y_o | (U_0, Y_0) = (t_o, x_o)\right] \geq \epsilon. \tag{b}$$

The theorem will be proved by induction once we establish it for a cup C with the aid of the following hypothesis.

Induction Hypothesis: If C' is any cup such that the depth of C' is strictly smaller than the depth of C, then Theorem 3 is true for C'.

In the remainder of this subsection, we briefly exploit the induction hypothesis. The two properties in Theorem 3, (a) and (b), that we need to establish for C are then proven in the next two subsections, respectively. The set of states C can be partitioned into (disjoint) sets $A, C_1, \ldots, C_p$ such that $C_1, \ldots, C_p$ are caps, $A = \{x \in C\colon V(x) = E_k\}$ and $\max\{V(x)\colon x \in C_i\} = E_{k-1}$. This is illustrated in Figure 3.1. In the special case that $m = k$, the partition consists of the single set A. In general, two elements x and y of C with $V(x) < E_k$ and $V(y) < E_k$ are in the same partition set C_i if and only if x can be reached from y at height E_{k-1}.

The depth of C_i for each i is at most g, which is smaller than the depth, d, of C. Therefore, we can apply the induction hypothesis to C_i for each i. We thus get ϵ_i and $\bar{\lambda}_i$ for each i, and by taking the smallest of these we get ϵ and $\bar{\lambda}$ depending only on Θ

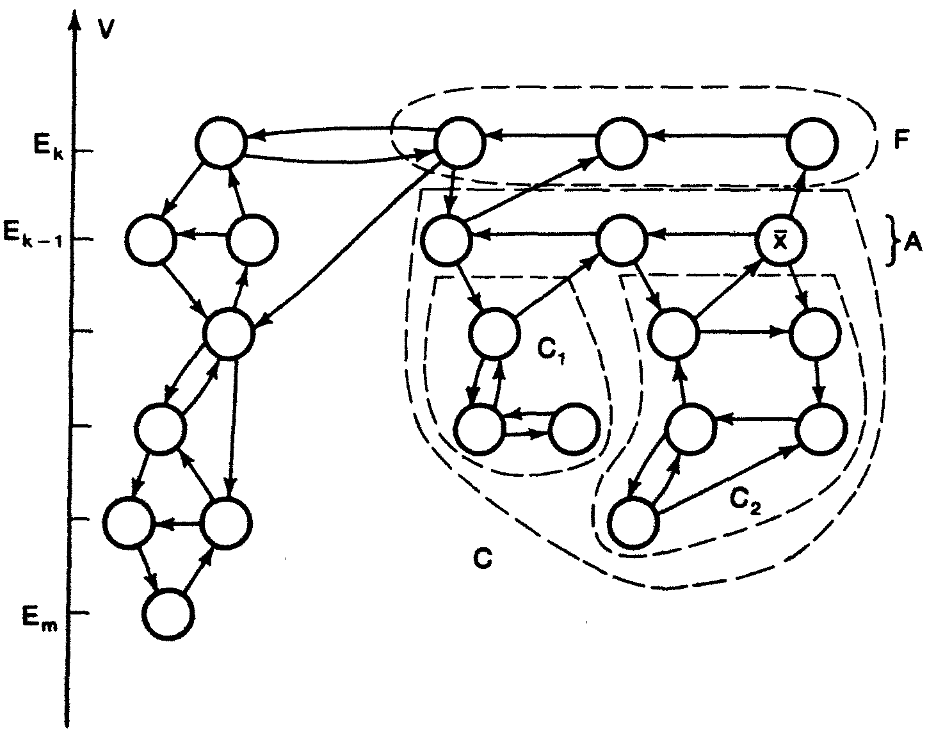

Figure 3.1. The partition of a cup C into a top layer A and cups from which the top layer can be directly reached. A state $\bar{x}$ in A is indicated from which F can be directly reached.

and C such that if $\lambda_{t_o} \leqslant \bar{\lambda}$ and $x_o \in C_i$ for some i then

$$E\left[\int_{t_0}^{\tau(C_i)} \lambda_s^g \, ds | (U_0, Y_0) = (t_o, x_o)\right] \leqslant \frac{1}{\epsilon}$$

and for every y such that $y \in A \cap N(x)$ for some $x \in C_i$

$$P[\text{hit } y \text{ upon first jump out of } C_i | (U_0, Y_0) = (t_o, x_o)] \geqslant \epsilon.$$

3.2. *Proof of Theorem* 3, *Part a*. In this subsection we prove part a of Theorem 3 under the induction hypothesis and its consequences described in §3.1. Let $A^0 = \{x \in A: N(x) \cap F \neq \varnothing\}$.

LEMMA 3.2.1. *Let* $K = (\min\{k > 0: Y_k \in A^0\} \wedge W) + 1$. *There is a constant* D_1 *depending only on* Θ *and* C *such that for every* $x_o \in C$ *and every* $t_o > 0$ *with* $\lambda_{t_o} \leqslant \bar{\lambda}$,

$$E\left[\int_{t_0}^{U_K} \lambda_t^d I_{\{t \leqslant U_W\}} \, dt | (U_0, Y_0) = (t_o, x_o)\right] \leqslant D_1 \lambda_{t_o}^u. \tag{3.1}$$

PROOF. Let $J_0 = 0$, let $J_{i+1} = \min\{k > J_i: Y_k \in A\} \wedge W$ for $i > 0$ and consider the discrete-time random process $\tilde{Y} = (\tilde{Y}_k)$ where $\tilde{Y}_k = Y_{J_k}$. The induction hypothesis described in §3.1 and the fact that Y first visits a state in A upon exiting any of the C_i, implies that $\tilde{Y}$ has the following properties:

- Any state in A is reachable from any other state in A (not necessarily in one step).
- There exists $\epsilon > 0$ depending only on Θ and C so that, for x and y in A such that a direct transition of $\tilde{Y}$ is possible, $P[\tilde{Y}_{k+1} = y | \mathscr{F}_k] \geqslant \epsilon$ on the event $\{\tilde{Y}_k = x\}$, where $\mathscr{F}_k$ is the σ-algebra generated by $(U_i: 0 \leqslant i \leqslant J_k + 1)$ and $(Y_i: 0 \leqslant i \leqslant J_k)$.
- There exists a constant D_2 depending only on Θ and C such that

$$E\left[\int_{U_{J_k+1}}^{U_{J_{k+1}+1}} \lambda_t^g I_{\{t \leqslant U_W\}} \, dt | \mathscr{F}_k\right] \leqslant D_2.$$

Let $M = \min\{j > 0: \tilde{Y}_j \in A^0 \cup F\}$. The first two properties stated imply that $E[M | (U_0, Y_0) = (t_o, x_o)] \leqslant D_3$ for some constant D_3 depending only on Θ and C. The third property implies that

$$\left(M \wedge k - \frac{1}{D_2} \int_{t_o}^{U_{J_k}+1} \lambda_t^g I_{\{t \leqslant U_W\}} \, dt, \mathscr{F}_k\right)_{k \geqslant 0}$$

is a submartingale, so that the kth term has nonnegative expectation for each k. Letting k tend to infinity and using the fact $M = J_K + 1$, this yields

$$E\left[\int_{t_o}^{U_K} \lambda_t^g I_{\{t \leqslant U_W\}} \, dt | U_0 = t_o, Y_0 = x_o\right] \leqslant D_2 D_3.$$

Since $\lambda_s^g \geqslant \lambda_s^d / \lambda_{t_o}^u$ for $s \geqslant t_o$, this implies Equation (3.1) for $D_1 = D_2 D_3$. ■

Let $c_3 = c_2 \max\{|N(x) \cap F|: x \in A\}$, $K_0 = 0$,

$$K_{i+1} = [\min\{k > K_i: Y_k \in A^0\} \wedge W] + 1 \quad \text{for } i \geqslant 1, \quad \text{and}$$

$$\Phi(j, k) = \int_{U_j}^{U_k} \lambda_s^d I_{\{s \leqslant U_W\}} \, ds.$$

LEMMA 3.2.2. *For* $1 \leqslant i < j < +\infty$

$$E\left[\Phi(K_i, K_j)|U_{K_i}, Y_{K_i-1}\right] \leqslant D_1/c_3. \tag{3.2}$$

PROOF. We prove the Lemma by backwards induction on i. Inequality (3.2) is true if $i = j$. Suppose for some i, j with $1 \leqslant i < j < +\infty$ that (3.2) is true with $i+1$ in place of i. Then Lemma 3.2.1 and the fact that $\Phi(K_i, K_j) = \Phi(K_i, K_{j+1}) + \Phi(K_{i+1}, K_j)$ imply that (writing $\nu(k)$ for $\lambda^u_{U_k}$)

$$E\left[\Phi(K_i, K_j)|U_{K_i}, Y_{K_i}\right] \leqslant D_1\nu(K_i) + D_1/c_3.$$

Then using the fact that $\Phi(K_i, K_j) = 0$ unless $W > K_i$, or equivalently, unless $Y_{K_i} \in C$, and Assumption A.2, we find

$$\begin{aligned} E\left[\Phi(K_i, K_j)|U_{K_i}, Y_{K_i-1}\right] &= E\left[E\left[\Phi(K_i, K_j)|U_{K_i}, Y_{K_i}, Y_{K_i-1}\right]I_{\{Y_{K_i} \in C\}}|U_{K_i}, Y_{K_i-1}\right] \\ &\leqslant \left(D_1\nu(K_i) + D_1/c_3\right)P\left[Y_{K_i} \in C|U_{K_i}, Y_{K_i-1}\right] \\ &\leqslant \left(D_1\nu(K_i) + D_1/c_3\right)(1 - c_3\nu(K_i)) \leqslant D_1/c_3, \end{aligned}$$

which completes the proof of the lemma. ■

We now complete the proof of Theorem 3, part a. By the monotone convergence theorem we obtain by letting $i = 0$ and letting j tend to infinity in (3.2) that $E[\Phi(K_1, W)|U_{K_1}, Y_{K_1-1}] \leqslant D_1/c_3$ and by Lemma 3.2.1, $E[\Phi(0, K_1)|(U_0, Y_0) = (t_o, x_o)] \leqslant D_1$. Thus

$$E\left[\Phi(0, W)|(U_0, Y_0) = (t_o, x_o)\right] \leqslant D_1(1 + c_3)/c_3$$

which implies part a of Theorem 3.

3.3. *Proof of Theorem* 3, *Part b*. We prove Theorem 3, (b) under the induction hypothesis stated in §3.1, thereby completing the inductive proof of Theorem 3. We will continue to refer to $\{A, C_1, \ldots, C_p\}$ and F, which were defined in the first two subsections.

Fix $t_o \geqslant 0$, $x_o \in C$ and $y_o \in F$. Choose a state $\bar{x} \in A$ so that $y_o \in N(\bar{x})$. Throughout this subsection we will assume that $(U_0, Y_0) = (t_o, x_o)$ and $\lambda_{t_o} \leqslant \bar{\lambda}$. Let $(J_k, \tilde{Y}_k, \mathscr{F}_k)_{k \geqslant 0}$ be defined as in the proof of Lemma 3.2.1. Then let

$$L^* = \min\left\{l \geqslant 0 \colon \tilde{Y}_l \in F\right\}, \quad L_0 = 0 \quad \text{and}$$

$$L_{i+1} = \min\left\{l > L_i \colon \tilde{Y}_l = \bar{x}\right\} \wedge L^* \quad \text{for } i \geqslant 0.$$

Note that $J_{L^*} = W$ and $Y_W = \tilde{Y}_{L^*}$. Define events $(H_i\colon\ i \geqslant 0)$ and $(G_i\colon\ i \geqslant 1)$ by $H_0 = \{J_{L_1} = W\}$, $G_i = \{W = J_{L_i} + 1\}$, and $H_i = \{J_{L_i} + 1 < W = J_{L_{i+1}}\}$ for $i \geqslant 1$.

LEMMA 3.3.2. *The following inequalities hold for some positive constants D_4 and D_5 depending only on Θ and C*

$$P[H_0] \leqslant D_4\lambda^u_{t_o}, \tag{3.3}$$

$$P[H_i] \leqslant P[G_i]D_5 \quad \text{for } i \geqslant 1. \tag{3.4}$$

PROOF. Let $Z_k = I_{\{\tilde{Y}_k \in F\}}$. Assumption A.1 implies that (writing $\nu(n)$ for $\lambda^u_{U_n}$):

$$\left[Z_k - \sum_{j=0}^{(k \wedge L^*)-1} c_3 \nu(J_j + 1), \mathscr{F}_k\right]_{k \geq 0}$$

is a supermartingale. Since Z_k is equal to $I_{\{J_k = W\}}$, H_i is true if and only if $Z_{L_{i+1}} - Z_{L_i+1} = 1$. Using this fact, the optional sampling theorem for supermartingales and the fact that λ is nonincreasing, we obtain

$$P\left[H_i | \mathscr{F}_{L_i}\right] = E\left[Z_{L_{i+1}} - Z_{L_i+1} | \mathscr{F}_{L_i}\right] \leq c_3 \nu(J_{L_i} + 1) E\left[(L_{i+1} - L_i - 1)^+ | \mathscr{F}_{L_i}\right].$$

Now $(L_{i+1} - L_i - 1)^+$ is the number of times $\tilde{Y}$ visits states in $A - \{\bar{x}\}$ before visiting $\{\bar{x}\} \cup F$. By the argument in Lemma 3.2.1, we thus conclude that $E[(L_{i+1} - L_i - 1)^+ | \mathscr{F}_{L_i}] \leq D_6$, where D_6 depends only on Θ and C. Hence,

$$P\left[H_i | \mathscr{F}_{L_i}\right] \leq c_3 D_6 \nu(J_{L_i} + 1). \tag{3.5}$$

If $i = 0$, this implies that $P[H_0] \leq c_3 D_6 \lambda^u_{t_o}$. For $i \geq 1$, Assumption A.2 implies that

$$P\left[G_i | \mathscr{F}_{L_i}\right] \geq c_1 \nu(J_{L_i} + 1) I_{\{\tilde{Y}_{L_i} = \bar{x}\}}. \tag{3.6}$$

Since $P[H_i | \mathscr{F}_{L_i}] = 0$ off the event $\{\tilde{Y}_{L_i} = \bar{x}\}$, we deduce from (3.5) and (3.6) that $P[H_i | \mathscr{F}_{L_i}] \leq D_5 P[G_i | \mathscr{F}_{L_i}]$ for $D_5 = D_6 c_3 / c_1$, which yields $P[H_i] \leq D_5 P[G_i]$. Lemma 3.3.2 is proved. ■

Now

$$P\left[\tilde{Y}_{L^*-1} = \bar{x}\right] = \sum_{i=1}^{\infty} P[G_i] \geq \sum_{i=1}^{\infty} P[H_i]/D_5 = \left(1 - P[H_0] - P\left[\tilde{Y}_{L^*-1} = \bar{x}\right]\right)/D_5$$

which by (3.3) implies that $P[\tilde{Y}_{L^*-1} = \bar{x}] \geq (1 - D_4 \lambda^u_{t_0})/(1 + D_5)$. Since $P[Y_W = y_o | \tilde{Y}_{L^*-1} = \bar{x}] \geq c_1/c_3$ we thus infer that

$$P[Y_W = y_o] \geq \frac{c_1}{c_3}\left[\frac{1 - D_4 \lambda^u_{t_o}}{1 + D_5}\right]$$

which implies part b of Theorem 3. The proof of Theorem 3 by induction is complete. ■

4. Upper bound on cup exit probability. We continue to assume that irreducibility, *WR*, A.1, A.2 and the continuous increase assumption hold.

THEOREM 4. *There exist $\lambda_o > 0$ and $\Gamma > 0$ depending only on Θ such that the following is true. If C is a cup with depth d and bottom B, $\lambda_t \leq \lambda_o$, $x_o \in B$ and $r \geq t$ then*

$$P\left[X_s \in C \text{ for } t \leq s \leq r | (U_0, Y_0) = (t, x_o)\right] \geq \exp\left(-\Gamma\left(\lambda^d_t + \int_t^r \lambda^d_s \, ds\right)\right).$$

Since there are only finitely many cups, it suffices to prove the theorem for a particular cup C. Let F, m and k be as defined in the beginning of §3. We can assume that $k < q$, since otherwise the theorem is trivial, and for ease of notation we assume that $m = 0$. Hence $d = E_{k+1} - E_0$.

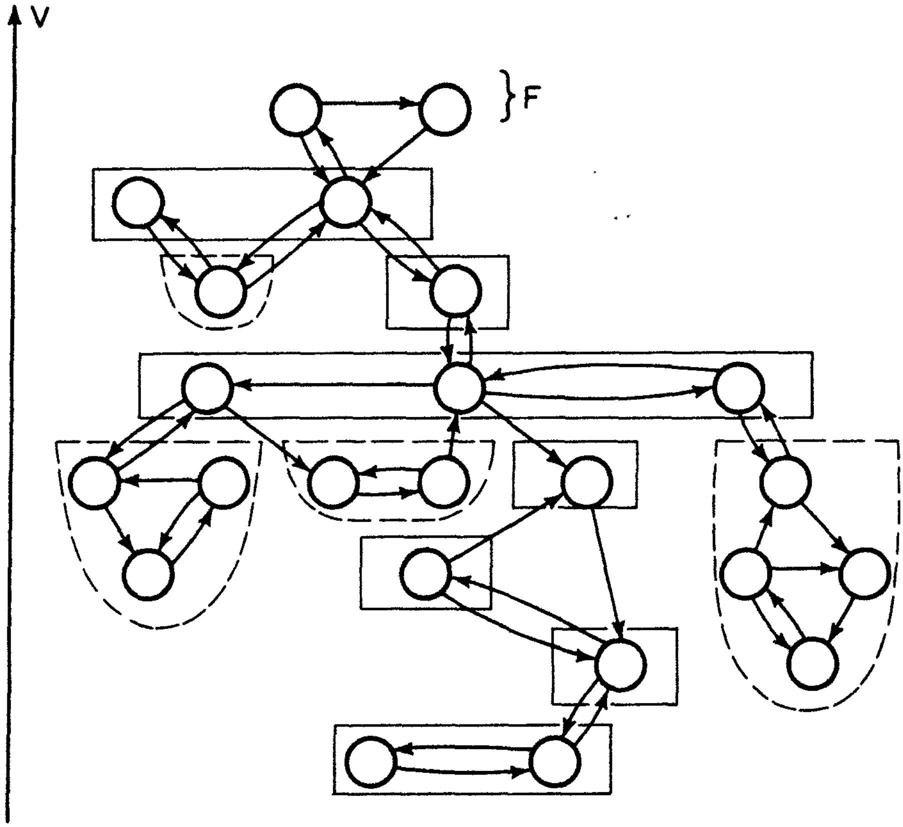

FIGURE 4.1. A cup C and the set F of states directly above C are shown. The cup is partitioned into a set of smaller cups, each of which is encircled with dashed lines, and a set J. The set J is itself partitioned into subsets, each of which is enclosed by a rectangle and consists of states in J with a common value of V.

Given x in C, let $h(x)$ denote the smallest height at which a state in B can be reached from x. There is a partition $\{J, C_1, \ldots, C_N\}$ of C such that $J = \{x \in C\colon h(x) = V(x)\}$ and each C_i is a cup. In fact, if $x \in C$ and $h(x) > V(x)$, then $x \in C_i$ for some i and C_i is the set of all states that can be reached from x at height strictly less than $h(x)$. This is illustrated in Figure 4.1. The partition has the property that Y cannot jump directly from one of the cups of the partition to another without visiting a state in J. Define L_{k+1} by $L_{k+1} = F$ and let $\Delta_j = E_{j+1} - E_j$ for $0 \leqslant j \leqslant k$.

LEMMA 4.1. *There is a constant c depending only on* Θ *and C such that the following is true. Let* $1 \leqslant j \leqslant k$, $x \in L_j$, $t \geqslant 0$ *and*

$$\zeta = \min\{m\colon Y_m \in L_0 \cup \cdots \cup L_{j-1} \cup L_{j+1}\}.$$

Then

$$P\left[Y_\zeta \in L_{j+1} | (U_0, Y_0) = (t, x)\right] \leqslant c\lambda_t^{\Delta_j}.$$

PROOF OF LEMMA 4.1. There exist $\bar{x} \in L_j$ and $y \in L_0 \cup \cdots \cup L_{j-1}$ such that $y \in N(\bar{x})$, and such that starting from x, Y can reach $\bar{x}$ without entering L_{j+1}. Hence, the expected cardinality of $\{n\colon 1 \leqslant n \leqslant \zeta$ and $U_n \in L_j\}$ is bounded by a constant depending only on Θ and C. Hence, the lemma can be deduced from Assumption A.2 and a martingale argument similar to that used in the proofs of Lemmas 3.2.1 and 3.2.2. ∎

Define $\Delta = \min\{\Delta_i | 0 \leqslant i \leqslant k\}$ and choose λ_o with $0 < \lambda_o < 1$ so small that $c\lambda_o^\Delta \leqslant 0.5$, where c is given by Lemma 4.1.

LEMMA 4.2. *Consider a discrete time Markov chain Z with state space $\{0, 1, \ldots, k+1\}$ and one-step transition probabilities*

$$p_{i,j} = \begin{cases} c\lambda^{\Delta_i} & \text{if } j = i+1, \\ 1 - c\lambda^{\Delta_i} & \text{if } j = i-1, \\ 0 & \text{otherwise}, \end{cases}$$

for $0 < i \leqslant k$. Then, for $\lambda \leqslant \lambda_o$

$$P[Z \text{ hits } k+1 \text{ before it hits } 0 | Z_0 = 1] \leqslant (2c)^k \lambda^{E_{k+1} - E_1}. \tag{4.1}$$

PROOF. Define x_i for $1 \leqslant i \leqslant k+1$ by

$$x_i = P[Z \text{ hits } k+1 \text{ before it hits } i-1 | Z_0 = i].$$

The left-hand side of (4.1), which we wish to bound, is x_1. Now, considering the case $Z_0 = i$ and conditioning on the value of Z_1, we obtain the relation

$$x_i = c\lambda^{\Delta_i}\{x_{i+1} - (1 - x_{i+1})x_i\} \quad \text{or} \quad x_i = \frac{c\lambda^{\Delta_i} x_{i+1}}{1 - c\lambda^{\Delta_i}(1 - x_{i+1})}.$$

Because $c\lambda_o^\Delta \leqslant 0.5$, the denominator on the right-hand side is at least 0.5, so that $x_i \leqslant 2c\lambda^{\Delta_i} x_{i+1}$, $1 \leqslant i \leqslant k$. Since $x_{k+1} = 1$, this implies the lemma. ■

LEMMA 4.3. *Suppose $\lambda_t \leqslant \lambda_o$ and $x \in L_1$. Then*

$$P[Y \text{ visits } F \text{ before visiting } B | (Y_0, U_0) = (t, x)] \leqslant (2c)^k \lambda_t^{E_{k+1} - E_1}.$$

PROOF. Define $R_0 = t$ and

$$R_{j+1} = \min\{n \geqslant R_j \colon Y_n \in J \text{ and } V(Y_n) \neq V(Y_{R_j})\}.$$

Lemma 4.1 implies that the process $(V(Y_{R_j}))_{j \geqslant 0}$ is stochastically bounded [5] above by the Markov chain E_Z, and thus has a smaller chance of hitting E_{k+1} before hitting E_0 than does E_Z. The lemma is thus a consequence of Lemma 4.2. ■

PROOF OF THEOREM 4. We will assume

$$(U_0, Y_0) = (t, x_o), \quad \text{where } x_o \in B,\ \lambda_t \leqslant \lambda_o,\ W = \min\{j \colon Y_j \in F\},$$

$$S_0 = 0, \quad S_{i+1} = \min\{j > S_i \colon Y_j \in B\} \wedge W \quad \text{for } i \geqslant 0,\ N^* = \min\{i \colon Y_{S_i} \in F\}$$

$$K_0^* = 0, \quad K_{n+1}^* = \min\{i > K_n^* \colon U_{S_i} \geqslant U_{K_n^*} + a\} \quad \text{for } n \geqslant 0.$$

Let $\mathscr{G}_i$ denote the σ-algebra generated by (U_j: $0 \leqslant j \leqslant S_i + 1$) and Y_j: $0 \leqslant j \leqslant S_i$). Then by Lemma 4.3 and Assumption A.2

$$P[N^* = i + 1 | \mathscr{G}_i] \leqslant \bar{c}\lambda^d_{U_{S_i}+1} \quad \text{for } i \geqslant 0,$$

where $\bar{c} = (2c)^k c_2 |L_1|$. Hence

$$\left[I_{\{i \geq N^*\}} - \sum_{j=0}^{(i \wedge N^*)-1} \bar{c}\lambda^d_{U_{S_j}+1}, \mathscr{G}_i \right]_{i \geq 0} \tag{4.2}$$

is a submartingale.

Since (1) $\{N^* > K_n^*\} \in \mathscr{G}_{K^*}$, (2) the process in (4.2) is a submartingale and λ is nonincreasing, (3) Assumption A.1 implies that $E[K_{n+1}^* - K_n^* | \mathscr{G}_{K_n^*}] \leq 1/a$, and (4) $\{U_{K_n^*} \geq t + na\}$ on the event $\{N^* > K_n^*\}$, we have

$$\begin{aligned} P\left[N^* > K_{n+1}^* | \mathscr{G}_{K_n^*}\right] &= \left(1 - E\left[I_{\{K_{n+1}^* \geq N^*\}} - I_{\{K_n^* \geq N^*\}} | \mathscr{G}_{K_n^*}\right]\right) I_{\{N^* > K_n^*\}} \\ &\geq \left(1 - E\left[(K_{n+1}^* - K_n^*)\bar{c}\lambda_{U_{K_n^*}}{}^d | \mathscr{G}_{K_n^*}\right]\right) I_{\{N^* > K_n^*\}} \\ &\geq \left(1 - \bar{c}\lambda_{U_{K_n^*}}{}^d / a\right) I_{\{N^* > K_n^*\}} \geq \left(1 - \bar{c}\lambda^d_{t+na}/a\right) I_{\{N^* > K_n^*\}}. \end{aligned}$$

Taking expectations, we get $P[N^* > K_{n+1}^*] \geq (1 - \bar{c}\lambda^d_{t+na}/a) P[N^* > K_n^*]$, so by induction on n,

$$P[N^* > K_{n+1}^*] \geq \prod_{j=0}^{n} \left(1 - \bar{c}\lambda^d_{t+ja}/a\right).$$

Thus, since $\{X_s \in C \text{ for } t \leq s \leq r\} \subset \{N^* > K^*_{[(r-t)/a]+1}\}$, if we suppose that λ_t is so small that $\bar{c}\lambda^d_t/a \leq x$, where x is the positive solution to $1 - x = \exp(-2x)$, then

$$\begin{aligned} P[X_s \in C \text{ for } t \leq s \leq r] &\leq \prod_{j=0}^{[(r-t)/a]} \left(1 - \bar{c}\lambda^d_{t+ja}/a\right) \leq \exp\left(-(2\bar{c}/a) \sum_{j=0}^{[(r-t)/a]} \lambda^d_{t+ja}\right) \\ &\leq \exp\left(-(2\bar{c}/a^2)\left\{\lambda^d_t + \int_t^r \lambda^d_s \, ds\right\}\right). \end{aligned}$$

This implies Theorem 4. ■

5. Proof of Theorem 2.

5.1. *Towards sufficiency of the integral condition.* Let $E \geq 0$ and let
$W_E = \{x | x \text{ is a local minimum of depth strictly larger than } E\}$,
$R_E = \{x | x \text{ is reachable from } y \text{ at height } V(y) + E \text{ for some } y \in W_E\}$.
We remind the reader that condition *WR* is in effect. The purpose of this subsection is to prove the following lemma, which will be used in the next section.

LEMMA 5.1. *If*

$$\int_0^\infty \lambda_t^E \, dt = +\infty, \tag{5.1}$$

then

$$\lim_{t \to \infty} P[X_t \in R_E] = 1. \tag{5.2}$$

The proof will be presented after the following lemma is proved. The lemma is illustrated in Figure 5.1.

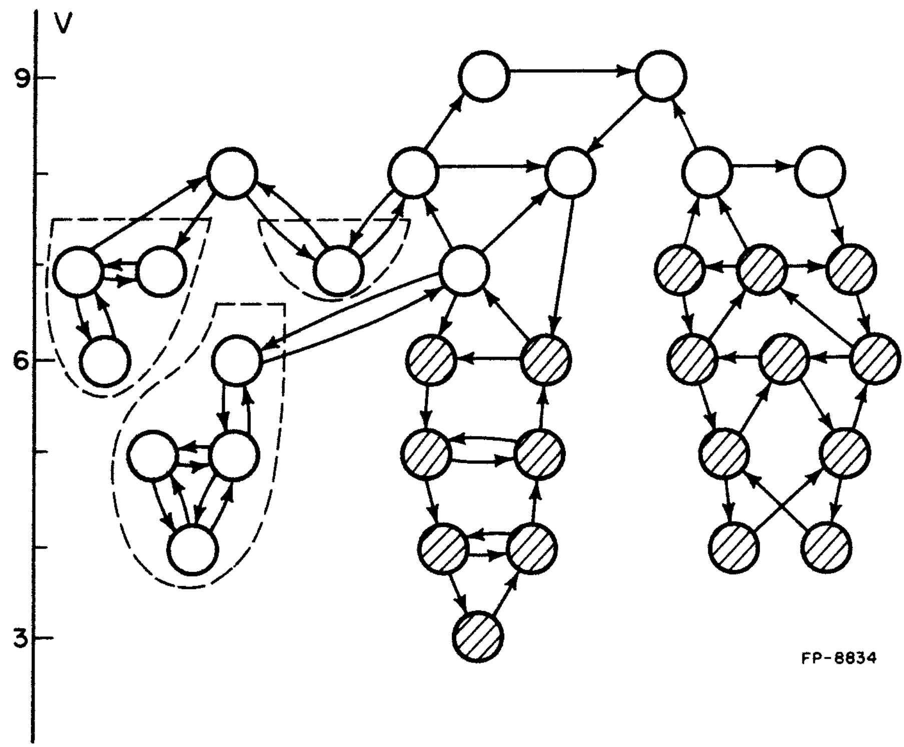

FIGURE 5.1. The partition of a set $\mathscr{S}$ into a subset J and a set of cups, corresponding to $E = 3$. The cups are enclosed by dashed lines, and all other states are in J. States in the subset R_E of J are hatched.

LEMMA 5.2. *There is a partition* $\{J, C_1, \ldots, C_N\}$ *of* $\mathscr{S}$ *such that*
(1) $J = \{x \in \mathscr{S} | W_E$ *is reachable from* x *at height* $V(x)\}$,
(2) $C_1, \ldots, C_N$ *are cups of depth at most* E,
(3) Y *cannot jump directly from one of the cups to another without visiting a state in* J.

PROOF. For $x \in \mathscr{S}$ define a cup C_x by $C_x = \{y | y$ is reachable from x at height $V(x)\}$. Fix x such that $C_x \cap J \neq \varnothing$. Then for $y \in C_x \cap J$, y is reachable from x at height $V(x)$, and W_E is reachable from y at height $V(y)$. The first of these facts implies that $V(y) \leqslant V(x)$. Then the two facts together imply that W_E is reachable from x at height $V(x)$, which is to say that $x \in J$. We have shown that if $C_x \cap J$ is nonempty, then $x \in J$. Equivalently, we have shown that $C_x \cap J = \varnothing$ if $x \notin J$.

Suppose that the depth of a cup C_x exceeds E. Let y be a state in the bottom of the cup. Then y is a local minimum of depth greater than E, so that $y \in W_E$, and hence, $y \in J$. Therefore, $c_x \cap J \neq \varnothing$, so that $x \in J$. We have just shown that if the depth of a cup C_x exceeds E, then $x \in J$. Equivalently, the depth of C_x is at most E for any $x \notin J$.

Given any x and y, either $C_x \subset C_y$, $C_y \subset C_x$ or $C_x \cap C_y = \varnothing$. Thus, the set of maximal sets from the collection $\mathscr{C} = \{C_x\colon x \notin J\}$ together with the set J forms a partition of $\mathscr{S}$ with the desired properties (1) and (2).

Choose x so that C_x is a maximal set in $\mathscr{C}$, and let y be a state not in C_x that can be reached in one step from a state in C_x. Since $V(y)$ is larger than the maximum of V over C_x (which happens to be $V(x)$), some state in C_x is reachable from y at level $V(y)$. Therefore $C_x \cap C_y$ is nonempty, so that C_x is a proper subset of C_y. Since C_x is maximal in the collection $\mathscr{C}$, it follows that C_y is not in $\mathscr{C}$, which means that $y \in J$. The partition thus has property (3) as well. ∎

PROOF OF LEMMA 5.1. Set $A_0 = 0$, $A_{i+1} = \min\{k > A_i\colon U_k \in J\}$ for $i \geqslant 0$, $\alpha = \min\{i\colon Y_{A_i} \in W_E\}$ and $\beta = \min\{i \geqslant \alpha\colon Y_{A_i} \notin R_E\}$. Since there are only finitely many cups, we can choose $\bar{\lambda} > 0$ in Theorem 3 to depend only on Θ, and we also suppose that $\bar{\lambda} < \lambda_o$ where λ_o appears in Theorem 4. By Theorem 3, part a and properties (2) and (3) in Lemma 5.2, there is a constant D_7 depending only on Θ such that if $\lambda_{t_o} \leqslant \bar{\lambda}$ then

$$E\left[\int_{U_{A_i}}^{U_{A_{i+1}}} \lambda_s^E \, ds | U_{A_i}, Y_{A_i}\right] \leqslant D_7$$

on the event $\{U_{A_i} \geqslant t_0\}$. By Theorem 3, part b and property (1) in Lemma 5.2, there exists $\epsilon > 0$ depending only on Θ such that: for any state $x \in J$ there is a sequence of distinct states $x = x_0, \dots, x_p$ such that $V(x_0) \geqslant V(x_1) \geqslant \cdots \geqslant V(x_p)$, $x_p \in W_E$ and $P[Y_{A_{i+1}} = x_{j+1} | Y_{A_i} = x_j, U_{A_i} = t] \geqslant \epsilon$ for $0 \leqslant j < p$ and $t \geqslant t_o$. This implies that there exists a constant D_8 depending only on Θ such that $E[\alpha | U_0 = t_o, Y_0 = x_o] \leqslant D_8$ for $x_o \in \mathscr{S}$ and $t \geqslant t_o$. Hence

$$E\left[\int_{t_o}^{U_{A_\alpha}} \lambda_s^E \, ds | U_0 = t_o, Y_0 = x_o\right] \leqslant D_7 D_8$$

so that

$$P\left[U_{A_\alpha} \leqslant r | U_0 = t_o, Y_0 = x_o\right] \geqslant 1 - D_7 D_8 \Big/ \int_{t_o}^{r} \lambda_s^E \, ds. \tag{5.3}$$

Note that Y_{A_α} must be in the bottom of a cup of depth at least $E + \gamma$, where $\gamma = \min\{d(C) - E\colon C \text{ is a cup with } d(C) > E\}$. Hence, by Theorem 4, and the fact (λ_t) is nonincreasing, there exists a constant Γ depending only on Θ such that for $r > t_o$,

$$P\left[U_{A_\beta} > r | U_0 = t_o, Y_0 = x_o\right] \geqslant \exp\left(-\Gamma\left(\lambda_{t_o}^{E+\gamma} + \int_{t_o}^{r} \lambda_s^{E+\gamma} \, ds\right)\right). \tag{5.4}$$

Since $X_r \in R_E$ if $U_{A_\alpha} \leqslant r < U_{A_\beta}$, we can combine (5.3) and (5.4) to yield

(5.5)

$$P[X_r \in R_E | U_0 = t_o, Y_0 = x_o] \geqslant \exp\left(-\Gamma \lambda_{t_o}^{\gamma}\left(\lambda_{t_o}^E + \int_{t_o}^{r} \lambda_s^E \, ds\right)\right) - D_7 D_8 \Big/ \int_{t_o}^{r} \lambda_s^E \, ds.$$

In view of condition (5.1), given any $\epsilon > 0$, if r is sufficiently large then there exists a large interval of time $[t_1, t_2]$ with $t_2 \leqslant r$ such that the right-hand side of (5.5) is at least $1 - \epsilon$ if $t_o \in [t_1, t_2]$. Given the process $(U_k, Y_k\colon k \geqslant 0)$ with $U_0 = 0$, we let $\rho = \min\{k\colon U_k \geqslant t_1\}$. We can choose r, and hence $t_2 - t_1$, so large that $P[U_\rho \leqslant t_2] \geqslant 1 - \epsilon$. By conditioning on (U_ρ, Y_ρ), we then obtain that $P[X_r \in R_E | U_0 = 0, Y_0 = x] \geqslant (1 - \epsilon)^2$ for all $x \in \mathscr{S}$. Lemma 5.1 is proved. ∎

5.2. *Proof of Theorem 2 under the continuous increase assumption.* In this subsection we will prove Theorem 2 under the additional assumption that the continuous increase property holds. The proof of Theorem 2 without this assumption is outlined in the next subsection.

When $E = 0$, R_E is the set of local minima and the integral condition (5.1) is always true. Lemma 5.1 thus implies part (a) of Theorem 2.

We will next prove the "if" half of part (b) of Theorem 2. Let B denote the bottom of a cup C of depth E and suppose that the states in B are local minima of depth E. We must show that $P[X_t \in B]$ has limit zero as t tends to infinity if the integral of λ^E is infinite. In view of Lemma 5.1 it is sufficient to prove that B and R_E are disjoint. For the sake of contradiction, suppose that $x \in B \cap R_E$. Then there is a state y so that the following three statements are true:

(i) x is a local minimum of depth E.

(ii) y is a local minimum of depth strictly greater than E.

(iii) x is reachable from y (and vice-versa) at height $V(y) + E$.

We consider two cases. First consider the case that $V(x) \leqslant V(y)$. By (i), there is a state z with $V(z) < V(x)$ that can be reached from x at height $V(x) + E$ (and hence at height $V(y) + E$). In view of (iii), this implies that the state z can also be reached from y at height $V(y) + E$. Since $V(z) < V(y)$, this contradicts (ii).

Now consider the case that $V(x) > V(y)$. By (iii), y can be reached from x at height $V(y) + E$, and this height is strictly smaller than $V(x) + E$. This contradicts (i).

We obtain a contradiction in either case, so we have proved that B and R_E are disjoint. This completes the proof of the "if" half of part (b) of Theorem 2 under the continuous increase assumption. To prove the "only if" half of part (b) we again let B denote the bottom of a cup C of depth E, we suppose that the states in B are local minima of depth E, and we assume that λ is given such that $\int_0^\infty \lambda_t^E \, dt < +\infty$. We want to prove that $P[X_t \in B]$ does not converge to zero as t tends to infinity.

Let F be the unique nonnegative number such that the integral of λ^r is infinite if $0 \leqslant r < F$ and the integral is finite if $r > F$. Then $F \leqslant E$, and we define a cup C' as follows. If $F < E$ we choose a state y in B and let $C' = \{x\colon x \text{ is reachable from } y \text{ at height } V(y) + F\}$ (in which case depth $(C') > F$) and if $F = E$ we let $C' = C$. We let B' denote the bottom of C' and then the following three facts are easy to verify: (1) If d denotes the depth of C' then the integral of λ^d is finite; (2) any state in $C' - B'$ is either not a local minimum or is a local minimum of depth strictly smaller than F, and (3) B' is a subset of B.

By the first of these facts and Theorem 4,

$$\liminf_{t \to \infty} P[X_t \in C'] > 0. \tag{5.6}$$

By the second of these facts, part (a) of Theorem 2 and the "if" half of part (b) of Theorem 2,

$$\lim_{t \to \infty} P[X_t = x] = 0 \quad \text{for } x \in C' - B'. \tag{5.7}$$

From (5.6) and (5.7) we conclude that $\liminf_{t \to \infty} P[X_t \in B'] > 0$, and from the third fact, this inequality is true with B' replaced by B. This completes the "only if" half of part (b) of Theorem 2, and hence the theorem itself, under the continuous increase assumption.

5.3. *Proof of Theorem* 2 *in general.* Suppose the conditions of Theorem 2 are satisfied, but that the continuous increase property is violated. We will construct a new system $(\hat{\mathscr{S}}, \hat{Q}, \hat{V}, \hat{\Phi})$ satisfying the conditions of Theorem 2 and (1) the continuous increase property holds so that, by what we proved so far, Theorem 2 applies to the new system, (2) $\mathscr{S} \subset \hat{\mathscr{S}}$, (3) if $x \in \mathscr{S}$, then x is a local minimum of depth d for the original system if and only if it is a local minimum of depth d for the new system, and (4) if the original and new system start in the same state, then the process $(X_t\colon t \geqslant 0)$

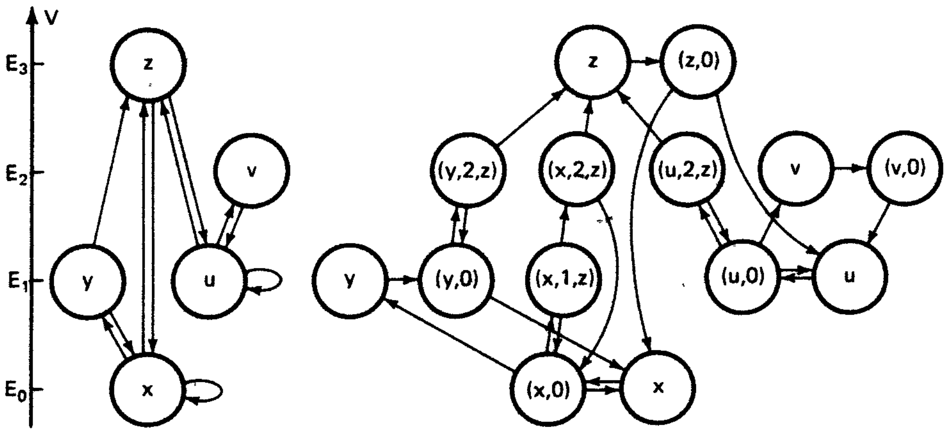

FIGURE 5.2. Example of $(\mathscr{S}, V, N)$ which violates the continuous increase assumption, and the new system.

has the same distribution for the two systems. The idea of the construction is to add new states so that the process can visit a state in each intermediate level when moving upward, and the holding times of the new states are zero. We now give the construction. An example is shown in Figure 5.2.

Let $E_0 < E_1 < \cdots < E_q$ denote the values of $V(x)$ as x ranges over S. For each x in S, define $H(x)$ so that $H(x) = i$ if $V(x) = E_i$. Define $U(x) = E_{H(x)+1}$ if $V(x) < E_q$, and $U(x) = +\infty$ if $V(x) = E_q$. In the following, x and y always denote states in $\mathscr{S}$ with $y \in N(x)$, and $s, t \geqslant 0$. Set

$$\hat{\mathscr{S}} = \bigcup_{x \in \mathscr{S}} (\{x, (x,0)\} \cup \{(x, i, y) \colon y \in N(x), V(y) > U(x), H(x) < i < H(y)\}),$$

$$\hat{V}(x) = V(x), \quad \hat{V}((x,0)) = V(x), \quad \hat{V}((x,i,y)) = E_i,$$

$$\hat{\Phi}(t,s,x) = \Phi(t,s,x), \quad \hat{\Phi}(t,s,\hat{x}) = I_{\{t \geqslant s\}} \quad \text{if } \hat{x} \in \hat{\mathscr{S}} - \mathscr{S},$$

$$\gamma(x,t,\lambda) = \left[1 + \sum_{y \colon V(y) > U(x)} Q(x,y,t,\lambda)(\lambda^{U(x)-V(y)} - 1)\right],$$

$$\hat{Q}(x,(x,0),t,\lambda) = 1,$$

$$\hat{Q}((x,0),(x,1,y),t,\lambda) = (Q(x,y,t,\lambda)\lambda^{U(x)-V(y)})/\gamma(x,t,\lambda) \quad \text{if } V(y) > U(x),$$

$$\hat{Q}((x,0),y,t,\lambda) = Q(x,y,t,\lambda)/\gamma(x,t,\lambda) \quad \text{if } V(y) \leqslant U(x),$$

$$\hat{Q}((x,i,y),(x,i+1,y),t,\lambda) = \lambda^{E_{i+1}-E_i} \quad \text{if } H(x) < i < H(y) - 1,$$

$$\hat{Q}((x,i,y),y,t,\lambda) = \lambda^{V(y)-E_i} \quad \text{if } H(x) < i = H(y) - 1,$$

$$\hat{Q}((x,i,y),y,t,\lambda) = 1 - \lambda^{E_{i+1}-E_i} \quad \text{if } H(x) < i < H(y).$$

Values of $\hat{Q}$ not specified are zero. The neighborhood structure for the new system can now be specified by $y \in \hat{N}(z)$ for $y, z \in \hat{\mathscr{S}}$ if and only if $Q(z,y,t,\lambda) > 0$ for some (equivalently for all) t, λ with $t \geqslant 0$ and $0 < \lambda < 1$. It can be checked that $\gamma(x,t,\lambda)$ is the expected number of visits the new process makes to state $(x,0)$ before reaching a state in $\mathscr{S}$, given the process starts in state x. Verification that the construction has the desired properties (1)–(4) is elementary and is left to the reader. The following is also easy to check: (5) if, for the original system, B is the bottom of a cup of depth d and the states in B are local minima of depth d, then the same is true of $\hat{B} = B \cup \{(x,0): x \in B\}$ for the new system. Theorem 2 is immediately implied by properties (1)–(5) of the construction.

Acknowledgements. The author is grateful for the comments and encouragement of the referees, associate editor and editor.

References

[1] Černy, V. (1985). A Thermodynamical Approach to the Travelling Salesman Problem: An Efficient Simulation Algorithm. *J. Optim. Theory Appl.* **45** 41–51.

[2] Geman, S. and Geman, D. (1984). Stochastic Relaxation, Gibbs Distribution, and the Bayesian Restoration of Images. *IEEE Trans. Pattern Analysis and Machine Intelligence*, **6** 721–741.

[3] Hwang, C.-R. (1984). Diffusions for Global Optimization, Preprint, Div. of Applied Math., Brown University.

[4] Gidas, B. (1985). Non-Stationary Markov Chains and Convergence of the Annealing Algorithm. *J. Statist. Phys.* **39** 73–131.

[5] Kamae, T., Krengel, V. and O'Brien, G. L. (1977). Stochastic Inequalities on Partially Ordered Spaces. *Ann. Prob.* **5** 899–912.

[6] Kemeny, J. and Snell, J. L. (1960). *Finite State Markov Chains*. D. Van Nostrand Company.

[7] Kirkpatrick, S., Gelatt, C. D. and Vecchi, M. P. (1983). Optimization by Simulated Annealing. *Science* **220** 621–680.

DEPARTMENT OF ELECTRICAL ENGINEERING AND THE COORDINATED SCIENCE LABORATORY, UNIVERSITY OF ILLINOIS, 1101 W. SPRINGFIELD AVENUE, URBANA, ILLINOIS 61801