The Annals of Probability
1983, Vol. 11, No. 2, 403–413

MINIMIZATION ALGORITHMS AND RANDOM WALK ON THE d-CUBE[1]

BY DAVID ALDOUS

University of California at Berkeley

Consider the number of steps needed by algorithms to locate the minimum of functions defined on the d-cube, where the functions are known to have no local minima except the global minimum. Regard this as a game: one player chooses a function, trying to make the number of steps needed large, while the other player chooses an algorithm, trying to make this number small. It is proved that the value of this game is approximately of order $2^{d/2}$ steps as $d \to \infty$. The key idea is that the hitting times of the random walk provide a random function for which no algorithm can locate the minimum within $2^{d(1/2-\varepsilon)}$ steps.

1. Introduction. An important practical problem is to find algorithms to locate the minimum of functions on R^d. As an abstraction of the "dimensionality" part of this problem, Tovey (1981) studied algorithms to locate the minimum of functions defined on the vertices of the unit cube in d dimensions (the *d-cube*). Let $I = \{i, j, k, \cdots\}$ be the vertices of the d-cube. Let $\mathscr{F}$ be the set of functions $f\colon I \to R$ taking distinct values. For brevity, say "evaluate i" for "evaluate the value of f at vertex i." An *algorithm* **a** is a rule specifying which vertex to evaluate next, depending on the vertices already evaluated. Mathematically, an algorithm can be described as a collection $(j_1;\, a_n,\, n \geq 1)$ where j_1 is an initial vertex and a_n a function $I^n \times R^n \to I$. When the algorithm is applied to f, it evaluates the series $j_1, j_2, \cdots$ of vertices, where

$$j_{n+1} = a_n(j_1, \cdots, j_n, f(j_1), \cdots, f(j_n)),$$

and terminates at vertex j_N, for some $N = N(f, \mathbf{a})$. Since we are concerned with minimization algorithms, we require

$$f(j_N) = \min_I f = f_{(1)}, \quad \text{for each } f, \tag{1.1}$$

where $f_{(1)}, f_{(2)}, \cdots, f_{(2^d)}$ is the increasing ordering of $\{f(i)\colon i \in I\}$. We are interested in algorithms which make small the number $N(f, \mathbf{a})$ of vertices which have to be evaluated (the *number of steps*). If no restrictions are placed on f, then to ensure (1.1), an algorithm must evaluate every vertex, so $N(f, \mathbf{a})$ cannot be smaller than 2^d. But if we restrict f to some subclass $\mathscr{F}_0$ of $\mathscr{F}$, it may be possible to do better. Following Tovey (1981) we shall consider the class $\mathscr{F}_0$ of *local-global* functions. That is, functions f such that for each vertex j,

either $f(j) = \min_I f$,
or there exists a neighbor k of j such that $f(k) < f(j)$.

For this class there is an obvious algorithm, the *optimal-adjacency* algorithm $\mathbf{a}^*$ which can be summarised as follows.

1. Let j be the initial vertex j_1.
2. Evaluate the neighbors $k_1, \cdots, k_d$ of j. If $f(j) < \min f(k_n)$, terminate. Otherwise proceed to 3.
3. Let k be the neighbor for which $f(k) = \min f(k_n)$. Set $j = k$ and go to 2.

Received January 1982.

[1] Research supported by National Science Foundation grant MCS80-02698.

AMS 1980 *subject classifications*. Primary 60J15; secondary 68C25.

Key words and phrases. Minimization algorithms, computational complexity, random walk, d-dimensional cube.

For future reference note

$$\text{if} \quad f(j_1) = f_{(m)} \quad \text{then} \quad N(f, \mathbf{a}^*) \le md + 1. \tag{1.2}$$

How good is this algorithm? Tovey (1981) observed that the worst case is very bad:

$$\sup_{\mathscr{F}_0} N(f, \mathbf{a}^*)/2^{d(1-\varepsilon)} \to \infty \quad \text{as} \quad d \to \infty; \qquad \varepsilon > 0.$$

But his simulations with several specific distributions μ on $\mathscr{F}_0$ suggest that $E_\mu N(f, \mathbf{a}^*)$ is of order d^2 for these specific distributions, so that in practice this algorithm seems good.

Our purpose is to present a theoretical result, adopting the viewpoint of game theory. Imagine one player choosing f at random from some distribution μ on $\mathscr{F}_0$, and another player choosing $\mathbf{a}$ at random from some distribution ν on the set $\mathscr{A}$ of algorithms. The first player seeks to maximize $N(f, \mathbf{a})$, the second to minimize it. Now $\mathscr{F}$ and $\mathscr{A}$ are essentially finite, because only the ordering $(f_{(n)})$ matters, so the fundamental theorem of game theory asserts that the game has a value

$$v = \sup_\mu \inf_\nu E_{\mu,\nu} N(f, \mathbf{a}) = \inf_\nu \sup_\mu E_{\mu,\nu} N(f, \mathbf{a}).$$

We shall prove that v is roughly of order $2^{d/2}$. Precisely

THEOREM 1.3. $\dfrac{\log(v)}{d\log(2)} \to \dfrac{1}{2}$ *as* $d \to \infty$.

Informally, one might say that locating the minimum of a local-global function requires exponential rather than polynomial time. Such problems are studied in computational complexity theory, but the techniques used there do not seem to give explicit results like Theorem 1.3.

The proof of the upper bound is very easy. It suffices to exhibit a random algorithm (distribution ν, say, on $\mathscr{A}$) such that

$$E_\nu N(f, \mathbf{a}) \le Cd2^{d/2}, \qquad f \in \mathscr{F}_0, \tag{1.4}$$

where C is a constant not depending on f, d. Set $M = [2^{d/2}]$ and consider the following random algorithm.

1. Pick M vertices $J_1, \cdots, J_M$ at random (uniformly, independently) from I. Let J be a vertex for which $f(J) = \min f(J_n)$.
2. Follow the optimal-adjacency algorithm from J.

Fix $f \in \mathscr{F}_0$. For $m \ge 1$,

$$\begin{aligned} P_\nu(N(f, \mathbf{a}) > M + md + 1) &\le P(f(J) > f_{(m)}) \quad \text{by (1.2)} \\ &= P(f(J_n) > f_{(m)}, 1 \le n \le M) \\ &= (1 - m2^{-d})^M. \end{aligned} \tag{1.5}$$

Now

$$E_\nu N(f, \mathbf{a}) = \textstyle\sum_{n\ge0} P(N(f, \mathbf{a}) > n) \le M + 1 + d \sum_{m\ge0} (1 - m2^{-d})^M \quad \text{by (1.5)}$$

and calculus gives (1.4).

The proof of the lower bound involves the random walk, whose basic properties are developed in Section 2. The reader may wish first to glance at Section 3, which outlines how the random walk is used to obtain the lower bound.

2. The random walk. Each vertex i of the d-cube can be represented as a d-tuple $(i_1, \cdots, i_d)$ of 0's and 1's. Let $|i - j| = \sum |i_r - j_r|$, so vertices i, j are *neighbors* if $|i - j| = 1$. Let π be the uniform distribution on I; $\pi(i) \equiv 2^{-d}$. By the *random walk* $(X_t)_{t\ge0}$ on I we mean the the continuous-time Markov chain with Q-matrix

$$Q(i,j) = 1/d; \qquad |i-j| = 1$$
$$= 0; \quad \text{other} \quad i \neq j.$$

Thus X holds at vertex i for an exponential (mean 1) time, then jumps to a neighboring vertex selected uniformly from the d neighbors of i. Unless otherwise specified, take the initial distribution X_0 to be uniform on I. Write $P_i(\cdot)$, $E_i(\cdot)$ for probability, expectation given $X_0 = i$. Write $p_{i,j}(t) = P_i(X_t = j)$.

The process (X_t) may be written coordinatewise as $(X_t^1, \cdots, X_t^d)$. It is easy to verify that the coordinate processes (X_t^r) are independent Markov chains on $\{0, 1\}$ with Q-matrix $Q(0, 1) = Q(1, 0) = 1/d$. It is therefore elementary to obtain the formula

$$p_{i,j}(t) = 2^{-d}(1 + e^{-2t/d})^{d-m}(1 - e^{-2t/d})^m, \quad \text{where} \quad m = |i-j|. \tag{2.1}$$

The simplicity of this formula is the reason for using the continuous time random walk instead of the discrete time version. Note that the uniform distribution π is the stationary distribution, and $p_{i,j}(t) \to 2^{-d} = \pi(j)$ as $t \to \infty$.

Let H_i be the first *hitting* time of vertex i:

$$H_i = \inf\{t \geq 0 \colon X_t = i\}.$$

Let H_i^+ be the first *entry* time of i:

$$H_i^+ = \min\{t > 0 \colon X_t = i, X_{t-} \neq i\}.$$

So $H_i^+ = H_i$ unless $X_0 = i$. Let L_i be the first *exit* time from i:

$$L_i = \inf\{t > H_i \colon X_t \neq i\}.$$

For a non-empty subset A of I let

$$H_A = \min\nolimits_A H_i; \qquad H_A^+ = \min\nolimits_A H_i^+; \qquad L_A = \min\nolimits_A L_i.$$

Note that L_A is the first time X exits some vertex of A, *not* the first time X exits A.

Fix $t > 0$. Call two processes $(Y_u)_{0 \leq u \leq t}$, $(Z_u)_{0 \leq u \leq t}$ *time-reversals* of each other if (Y_u) has the same distribution as (Z_{t-u}). It is not hard to verify that the following pairs are time-reversals of each other. Throughout, (X_u) is the random walk.

(2.2)
(a) (X_u) with X_0 uniform, conditioned on $X_t = i$;
(b) (X_u) with $X_0 = i$.

(2.3)
(a) (X_u) with X_0 uniform, conditioned on $L_i = t$;
(b) (X_u) with $X_0 = i$, conditioned on $H_i^+ > t$.

(2.4)
(a) (X_u) with X_0 uniform, conditioned on $L_A = t$, $X(L_A-) = i$;
(b) (X_u) with $X_0 = i$, conditioned on $H_A^+ > t$.

Think of the event "$L_A = t$, $X(L_A-) = i$" as meaning "X hits A at vertex i, and first exits i at time t." Of course events like this have probability zero: strictly, our assertions should be formulated in terms of regular conditional distributions.

We shall need estimates for the transition probabilities $p_{i,j}(t)$ and the distribution of H_i. The lemmas below are sufficient for our needs (Lemma 2.8 could alternatively be derived by transform techniques, which give an explicit though complicated expression for the distribution of H_i).

LEMMA 2.5. (a) *For* $\varepsilon < 1$ *and* $t \geq (d/2)\log(2d/\varepsilon)$, $|p_{i,j}(t) - 2^{-d}| \leq \varepsilon 2^{-d}$.
(b) $\int_0^d p_{i,i}(t)\,dt \to 1$ *as* $d \to \infty$.
(c) $\int_0^{2^{d/2}} p_{i,i}(t)\,dt \to 1$ *as* $d \to \infty$.
(d) $P_i(X_t = i$ *for some* $L_i < t \leq 2^{d/2}) \to 0$ *as* $d \to \infty$.

PROOF. Assertion (a) is a straightforward consequence of (2.1). The integral in (b) may be written as

$$\int_0^\infty \left\{\frac{1+e^{-2t/d}}{2}\right\}^d 1_{(t\le d)}\,dt.$$

The integrand converges pointwise to e^{-t} and is dominated by e^{-at}, where a > 0 is chosen so that $e^{ax} + e^{(a-2)x} \le 2$ on $0 \le x \le 1$. Use the dominated convergence theorem. For (c),

$$\int_d^{2^{d/2}} p_{i,i}(t)\,dt \le 2^{d/2}\sup_{t\ge d} p_{i,i}(t) = 2^{d/2}\left\{\frac{1+e^{-2}}{2}\right\}^d \to 0.$$

Assertion (d) follows easily from (c) and the fact that each sojourn at i has expected duration 1.

Define

$$t_1 = \frac{d}{2}\log(2d2^d). \tag{2.6}$$

Roughly, t_1 is of order d^2. By (2.5)(a),

$$|p_{i,j}(u) - 2^{-d}| \le 2^{-2d}, \qquad u \ge t_1. \tag{2.7}$$

So conditional on X_0, the distribution of X at any time $u \ge t_1$ is almost uniform; informally, events separated by time t_1 or more are almost independent.

Let *time* $(t \le t_0: X_t = j)$ be the random variable measuring the duration of time before t_0 for which X is at j.

LEMMA 2.8. *There exist constants* $\varepsilon_1(d), \varepsilon_2(d) \to 0$ *as* $d \to \infty$ *such that*
(a) $P(H_i \le u) \ge u2^{-d}(1 - \varepsilon_1(d)),\ u \le 2^{d/2}$.
(b) $P(H_i \le u) \le u2^{-d}(1 + \varepsilon_2(d)),\ u \ge t_1$.

PROOF. (a) By considering the first hit on i,

$$E_\pi \operatorname{time}(t \le u: X_t = i) \le P_\pi(H_i \le u)\cdot E_i \operatorname{time}(t \le u; X_t = i) = P_\pi(H_i \le u)\cdot\int_0^u p_{i,i}(t)\,dt.$$

By symmetry, E_π time$(t \le u: X_t = i)$ does not depend on i, so must equal $u2^{-d}$. Apply (2.5)(c).

(b) Again by considering the first hit on i,

$$E_\pi \operatorname{time}(t \le u + d: X_t = i) \ge P_\pi(H_i \le u)\cdot E_i \operatorname{time}(t \le d: X_t = i) = P_\pi(H_i \le u)\cdot\int_0^d p_{i,i}(t)\,dt.$$

Again E time$(t \le u + d: X_t = i) = (u + d)2^{-d}$ by symmetry. Rearrange, and apply (2.5)(b) and (2.6).

3. The random walk and the lower bound for v. As in Section 2, (X_t) is the random walk with uniform initial distribution. The collection (H_i) of first hitting times may be regarded as a random function $H: I \to [0, \infty)$. For each ω and each $j \in I$:

either $H_j(\omega) = 0$ (that is, $X_0(\omega) = j$);

or $X_{H_j(\omega)-}$ is a neighboring vertex (k, say) of j, and $H_k(\omega) < H_j(\omega)$.

Thus each $H(\omega)$ is a local-global function. For technical reasons it is convenient to use the first exit times (L_i) instead of the first hitting times. Again, we may regard the collection (L_i) as a random local-global function L. Of course L attains its minimum at the vertex X_0. We shall prove that no algorithm can locate this minimum in much less than $2^{d/2}$ steps.

Precisely:

$$\frac{\inf_{\mathbf{a}} EN(L, \mathbf{a})}{2^{d(1/2-\varepsilon)}} \to \infty \quad \text{as} \quad d \to \infty; \qquad \varepsilon > 0, \tag{3.1}$$

and this will establish the lower bound in Theorem 1.3.

The details of the argument are rather intricate; here are the underlying ideas. First we argue that there is no loss of generality in assuming that the first vertex j_1 evaluated has $L_{j_1} \simeq 2^{d/2}$. Now suppose the algorithm has evaluated L at vertices $j_1, \cdots, j_n$. Let $A = \{j_1, \cdots, j_n\}$ and let i_0 be the element of A for which $L_{i_0} = \min L_{j_r}$. The algorithm selects some vertex j to evaluate next: suppose we can prove that for any j

(*) $$P(L_j < L_{i_0} - t_1) \quad \text{is at most order } 2^{-d/2}$$

where t_1 was defined at (2.6). Since $\min_I L$ is almost zero, (*) and induction show that $N(L, \mathbf{a})$ is likely to be at least $2^{d/2}/t_1$, giving the result. Why does (*) hold? By time reversal, the probability in (*) is at most

(**) $$P_{i_0}(X \text{ visits } j \text{ during } [t_1, 2^{d/2}] \,|\, X \text{ does not re-enter } A).$$

Without the conditioning, this probability would indeed be of order $2^{-d/2}$, by Lemma 2.8(b). To handle the conditioning, we need regularity properties of A to ensure that the conditioning does not affect much the behavior of X after time t_1; and then we need to ensure that these regularity properties are carried forward in the induction. This makes the proof technically complicated.

4. We now start the formal argument for (3.1). In this section we show how (3.1) can be reduced to Lemmas 4.17 and 4.19, which are assertions about the behavior of the random walk, not involving any algorithm.

Let $0 < \varepsilon < 1/4$ be fixed throughout. We make two conventions. First, each assertion will hold "for d sufficiently large," that is for d larger than some $d_0(\varepsilon)$ depending only on ε and the assertion. Second, subsets A of I are assumed to satisfy

$$1 \le \#A \le 2^{d(1/2-2\varepsilon)}. \tag{4.1}$$

Recall definition (2.6) of t_1. By (2.7)

$$p_{i,j}(u) \le 2^{-d}(1 + 2^{-d}); \qquad u \ge t_1. \tag{4.2}$$

Also, by (2.6),

$$t_1(1 + \#A) \le 2^{d(1/2-\varepsilon)}. \tag{4.3}$$

(Here we use the conventions: A satisfies (4.1), and (4.3) holds for d sufficiently large.)

DEFINITION 4.4. Let $A \subset I$, $i \in A$, $t > 0$. Say (i, A, t) is *regular* if
(a) $2^{d/2} \ge t \ge 2^{d(1/2-\varepsilon)} - t_1\#A$
(b) $s(i, A, u) \le 2 \cdot 2^{-d}$, where

$$u = t - 2^{d(1/2-\varepsilon)} + t_1(1 + \#A) \tag{4.5}$$

and where s is a certain function to be specified later (5.1).

Note that if (a) holds, then by (4.3)

$$t_1 \le u \le t. \tag{4.6}$$

DEFINITION 4.7. Let $A \subset I$ and let f be a local-global function on I. Say (f, A) is *good* if $(i, A, f(i))$ is regular, where $i \in A$ is the vertex for which

$$f(i) = \min_A f. \tag{4.8}$$

Let $\mathbf{a}$ be an algorithm as in Section 1. When $\mathbf{a}$ is applied to the random local-global function L, it evaluates a sequence of vertices $J_1, J_2, \cdots$ of the form

(4.9) $$J_1 = j_1, \quad J_{n+1} = a_n(J_1, \cdots, J_n, L_{J_1}, \cdots, L_{J_n});$$

the algorithm terminates after $N = N(L, \mathbf{a})$ steps, and $J_{N(L,\mathbf{a})} = X_0$. For notational convenience, set $J_n = J_N$, $n \geq N$. We shall prove

(4.10) $$EN(L, \mathbf{a}) \geq \tfrac{1}{2}2^{d(1/2-2\varepsilon)}(1 - o(1)) \quad \text{as} \quad d \to \infty,$$

which establishes (3.1).

LEMMA 4.11. *Let J_1 be a random vertex, dependent on the random walk (X_t). If (4.10) holds for algorithms $\hat{\mathbf{a}}$ with initial vertex J_1, then it holds for all $\mathbf{a}$ with deterministic initial vertex.*

PROOF. Given any algorithm $\mathbf{a}$, we can define an algorithm $\hat{\mathbf{a}}$ which has random initial vertex J_1 but then ignores the information given by J_1, L_{J_1}, and proceeds as $\mathbf{a}$. Then $N(L, \hat{\mathbf{a}}) = N(L, \mathbf{a}) + 1$.

In other words, in proving the lower bound (4.10) we may specify the initial vertex J_1 any way we choose. We shall take J_1 to be the vertex visited immediately before the vertex occupied by X_t at time $2^{d(1/2-\varepsilon)}$;

(4.12) $$J_1 = X(H_{J_0}-), \quad \text{where} \quad J_0 = X(2^{d(1/2-\varepsilon)}).$$

The algorithm then evaluates $J_2, J_3, \cdots$ of the form (4.9). Let A_n^* be the random set $\{J_1, \cdots, J_n\}$. We shall prove

(4.13) $$P((L, A_1^*) \text{ is good}) \to 1 \quad \text{as} \quad d \to \infty;$$

(4.14) $$P((L, A_{n+1}^*) \text{ is not good} \mid (L, A_n^*) \text{ is good}) \leq 6 \cdot 2^{-d/2}$$

(this is a formalisation of the idea of (*), Section 3). Let us show how to deduce (4.10) from these assertions. Set $n_0 = [2^{d(1/2-2\varepsilon)}]$. By (4.14)

(4.15) $$\begin{aligned} P((L, A_{n_0}^*) \text{ is good}) &\geq (1 - 6 \cdot 2^{-d/2})^{n_0} P((L, A_1^*) \text{ is good}) \\ &\to 1 \quad \text{as} \quad d \to \infty \quad \text{by} \quad (4.13). \end{aligned}$$

On the set where $(L, A_{n_0}^*)$ is good, (4.4) (a) implies

$$\begin{aligned} L_{A_{n_0}^*} &\geq 2^{d(1/2-\varepsilon)} - t_1 n_0 \\ &\geq \tfrac{1}{2} \cdot 2^{d(1/2-\varepsilon)}, \quad \text{for } d \text{ sufficiently large.} \end{aligned}$$

On the set where $N(L, \mathbf{a}) \leq n_0$, we have $L_{A_{n_0}^*} = L_{X_0}$, and so $X_t = X_0$ on $0 \leq t < L_{A_{n_0}^*}$. So

(4.16) $$\begin{aligned} P((L, A_{n_0}^*) \text{ is good}, N(L, \mathbf{a}) \leq n_0) &\leq P(X_t = X_0 \quad \text{on} \quad 0 \leq t < \tfrac{1}{2} \cdot 2^{d(1/2-\varepsilon)}) \\ &\to 0 \quad \text{as} \quad d \to \infty. \end{aligned}$$

Now (4.15) and (4.16) imply $P(N(L, \mathbf{a}) \geq n_0) \to 1$ as $d \to \infty$, which establishes (4.10).

We have now reduced the proof of Theorem 1.3 to the proof of (4.13) and (4.14). Next, we translate these assertions into assertions about the random walk. For (4.13) this is straightforward from Definitions 4.7 and 4.4; it suffices to prove

LEMMA 4.17.
(a) $P(2^{d/2} \geq L_{J_1} \geq 2^{d(1/2-\varepsilon)} - t_1) \to 1$ *as* $d \to \infty$;
(b) $s(j, \{j\}, u) \leq 2 \cdot 2^{-d}$ *for* $j \in I$; $t_1 \leq u \leq 2^{d/2}$.

We shall prove this in Section 5. Translating (4.14) requires more effort. For $i \in A$ define

(4.18) $$\tilde{P}(\cdot \mid i, A, t) = P(\cdot \mid X(L_A-) = i, L_A = t).$$

In Section 6 we shall prove

LEMMA 4.19. *Suppose* (i, A, t) *is regular*, $j \notin A$. *Then*
(a) $\tilde{P}((i, A \cup \{j\}, t)$ *is not regular*, $L_j \geq L_A | i, A, t) \leq 2 \cdot 2^{-d/2}$
(b) $\tilde{P}((j, A \cup \{j\}, L_j)$ *is not regular*, $L_j < L_A | i, A, t) \leq 4 \cdot 2^{-d/2}$.

These remain true for $j \in A$, since then $L_j \geq L_A$ and $(i, A \cup \{j\}, t) = (i, A, t)$ is regular.

The rest of this section is devoted to proving (4.14) from Lemma 4.19. We need the technical lemma below. For $A \subset I$ define the post-L_A-σ-field $\mathscr{L}_A = \sigma(L_A; X(L_A-); X_u, u \geq L_A)$. Plainly

$$(4.20) \qquad L_j \in \mathscr{L}_A, \qquad X_{L_j} \in \mathscr{L}_A; \qquad j \in A$$

$$(4.21) \qquad \mathscr{L}_B \subset \mathscr{L}_A, \qquad B \subset A.$$

LEMMA 4.22. $\{A_n^* = A, J_{n+1} = j\} \in \mathscr{L}_A, n \geq 1$.

PROOF. We shall prove that for any $n \geq 1, A = \{j_1, \cdots, j_n\}, t_1, \cdots, t_n$,

$$(4.23) \qquad \{A_n^* = A; J_1 = j_1, \cdots, J_n = j_n; L_{j_1} \geq t_1, \cdots, L_{j_n} \geq t_n\} \in \mathscr{L}_A.$$

This is true for $n = 1$, since $\{A_1^* = \{j_1\}, J_1 = j_1, L_{j_1} \geq t\} = \{t \leq L_{j_1} < 2^{d(1/2-\varepsilon)}; X(u) = X(L_{j_1})$ for $L_{j_1} \leq u \leq 2^{d(1/2-\varepsilon)}\} \in \mathscr{L}_{j_1}$. Suppose inductively that (4.23) holds for n. Then by (4.9),

$$(4.24) \qquad \{A_n^* = A; J_1 = j_1, \cdots, J_{n+1} = j_{n+1}; L_{j_1} \geq t_1, \cdots, L_{j_n} \geq t_n\} \in L_A \subset L_{A \cup \{j_{n+1}\}}.$$

Since $\{L_{j_{n+1}} \geq t_{n+1}\} \in L_{A \cup \{j_{n+1}\}}$ by (4.20) and (4.21), and since $A_{n+1}^* = A_n^* \cup \{J_{n+1}\}$, we see that (4.23) holds for $n + 1$. So by induction (4.23) is true for all n; so (4.24) is true for all n, and the lemma follows.

Now let $\mathscr{F}_A$ be the strict pre-L_A σ-field $\sigma(X_u: u < L_A)$. By a variant of the strong Markov property,

$$(4.25) \qquad P(B | \mathscr{L}_A) = P(B | X(L_A-), L_A); \qquad B \in \mathscr{F}_A$$
$$= \tilde{P}(B | X(L_A-), A, L_A) \quad \text{in notation of (4.18).}$$

We can now prove (4.14). By definition 4.7, $(L, A \cup \{j\})$ is good if

$$\text{either} \quad L_j \geq L_A \quad \text{and} \quad (i, A \cup \{j\}, t) \quad \text{is regular}$$
$$\text{or} \quad L_j < L_A \quad \text{and} \quad (j, A \cup \{j\}, L_j) \quad \text{is regular.}$$

Thus by Lemma 4.19,

$$(4.26) \qquad \tilde{P}((L, A \cup \{j\}) \text{ is not good} | i, A, t) \leq 6 \cdot 2^{-d/2} \quad \text{if} \quad (i, A, t) \quad \text{regular.}$$

Now

$$\{(L, A \cup \{j\}) \text{ is not good}\} \in \sigma(L_{A \cup \{j\}}, X(L_{A \cup \{j\}}-)) \quad \text{by definition}$$
$$\subset \mathscr{F}_{A \cup \{j\}} \subset \mathscr{F}_A.$$

So by (4.25) and (4.26)

$$P((L, A \cup \{j\}) \text{ is not good} | \mathscr{L}_A) \leq 6 \cdot 2^{-d/2} \quad \text{on} \quad \{(X(L_A-), A, L_A) \text{ is regular}\}.$$

Applying Lemma 4.22,

$$P((L, A_{n+1}^*) \text{ is not good} | \mathscr{L}_A) = P((L, A \cup \{j\}) \text{ is not good} | \mathscr{L}_A) \quad \text{on} \quad \{A_n^* = A, J_{n+1} = j\}$$
$$\leq 6 \cdot 2^{-d/2} \quad \text{on} \quad \{A_n^* = A, (X(L_A-), A, L_A) \text{ is regular}\}.$$

But this last set is $\{A_n^* = A, (L, A_n^*) \text{ is good}\}$, and (4.14) follows.

5. Here is the definition of the function s appearing in the definition (4.4) of "regular".

DEFINITION 5.1. For $A \subset I$, $i \in A$, $u \geq 0$ define

$$s(i, A, u) = \max_{j \in I} P_i(X_u = j \mid H_A^+ > u).$$

PROOF OF LEMMA 4.17. (a) $L_{J_1} \leq 2^{d/2}$ by definition. Set $t_0 = 2^{d(1/2-\varepsilon)}$.

$$P(L_{J_1} < t_0 - t_1) = P(L_{J_1} < t_0 - t_1 \mid X(t_0) = k) \quad \text{by symmetry}$$

$$= P_k(X_u = X(L_k) \text{ for some } t_1 \leq u \leq t_0) \quad \text{by time-reversal (2.2).}$$

By conditioning on $(L_k, X(L_k))$, this is at most

$$P_k(L_k > \tfrac{1}{2}t_1) + P_j(X_u = j \quad \text{for some} \quad \tfrac{1}{2}t_1 \leq u \leq t_0).$$

As $d \to \infty$, the first term tends to zero since $P_k(L_k \geq u) = e^{-u}$, and the second term tends to zero by Lemma 2.5(d).

(b) Let $t_1 \leq u \leq 2^{d/2}$. Then

$$P_j(X_u = k \mid H_j^+ > u) \leq P_j(X_u = k)/P_j(H_j^+ > u)$$

$$\leq 2^{-d}(1 + 2^{-d})/P_j(H_j^+ > 2^{d/2}) \text{ by (4.2)}$$

$$\leq 2 \cdot 2^{-d} \quad \text{for } d \text{ sufficiently large,}$$

because $P_j(H_j^+ > 2^{d/s}) \to 1$ as $d \to \infty$ by Lemma 2.5(d). So $s(j, \{j\}, u) \leq 2 \cdot 2^{-d}$ for sufficiently large d.

6. It remains to prove Lemma 4.19. The key idea is that the time-reversal (2.4) can be written, in the notation (4.18), as

$$\tilde{P}((X_u)_{0 \leq u \leq t} \in \cdot \mid i, A, t) = P_i((X_{t-u})_{0 \leq u \leq t} \in \cdot \mid H_A^+ > t). \tag{6.1}$$

Thus the assertions of Lemma 4.19 can be translated into assertions about the behavior of the random walk started at $i \in A$ and conditional on $\{H_A^+ > t\}$. We start with a sequence of lemmas concerning this conditioned random walk.

LEMMA 6.2. *$P_i((X_{t+u})_{t \geq 0} \in \cdot \mid H_A^+ > u) \leq 2^d s(i, A, u) \cdot P_\pi((X_t)_{t \geq 0} \in \cdot)$.*

PROOF.

$$P_i((X_{t+u}) \in \cdot \mid H_A^+ > u) = \textstyle\sum_k P_i(X_u = k \mid H_A^+ > u) \cdot P_k((X_t) \in \cdot)$$

$$\leq 2^d \max_k P_i(X_u = k \mid H_A^+ > u) \textstyle\sum_k 2^{-d} P_k((X_t) \in \cdot)$$

$$= 2^d s(i, A, u) P_\pi((X_t) \in \cdot).$$

LEMMA 6.3. *If $u \leq 2^{d/2}$ and $s(i, A, u) \leq 2 \cdot 2^{-d}$ then*

$$P_i(H_A^+ > 2^{d/2} \mid H_A^+ > u) \geq 1 - 2^{-\varepsilon d}.$$

PROOF.

$$P_i(H_A^+ \leq 2^{d/2} \mid H_A^+ > u) \leq 2P_\pi(H_A^+ \leq 2^{d/2} - u) \quad \text{by Lemma 6.2}$$

$$\leq 2P_\pi(H_A \leq 2^{d/2})$$

$$\leq 2\#A P_\pi(H_j \leq 2^{d/2}) \quad \text{by symmetry}$$

$$\leq 2\#A 2^{-d/2}(1 + \varepsilon_2(d)) \quad \text{by Lemma 2.8(b)}$$

$$\leq 2^{-\varepsilon d} \quad \text{for } d \text{ sufficiently large, by (4.1).}$$

The first lemma below is elementary; the second is an immediate consequence of (4.2).

LEMMA 6.4. *For events G, H, $\hat{H}$ with $H \subset \hat{H}$ and $P(H) > 0$,*

$$P(G|H) \le P(G|\hat{H})/P(H|\hat{H}).$$

LEMMA 6.5. $P_i(X_{u+t_1} = j | X_r, r \le u) \le 2^{-d}(1 + 2^{-d})$.

LEMMA 6.6. *If* $s(i, A, u) \le 2 \cdot 2^{-d/2}$ *then* $s(i, A, u + t_1) \le 2 \cdot 2^{-d}$.

PROOF.

$$\begin{aligned} P_\pi(H_A \le t_1) &\le \#AP_\pi(H_j \le t_1) \\ &\le \#A2^{-d}t_1(1 + \varepsilon_2(d)) \quad \text{by Lemma 2.8(b)} \qquad (6.7)\\ &\le 2^{-d(1/2+\varepsilon)}(1 + \varepsilon_2(d)) \quad \text{by (4.3).} \end{aligned}$$

$$\begin{aligned} P_i(H_A^+ \le u + t_1 | H_A^+ > u) &\le 2^d \cdot 2 \cdot 2^{-d/2} P_\pi(H_A \le t_1) \quad \text{by Lemma 6.2} \qquad (6.8)\\ &\to 0 \quad \text{as} \quad d \to \infty \quad \text{by (6.7).} \end{aligned}$$

And by Lemma 6.5,

$$P_i(X_{u+t_1} = j | H_A^+ > u) \le 2^{-d}(1 + 2^{-d}). \qquad (6.9)$$

Finally,

$$\begin{aligned} P_i(X_{u+t_1} = j | H_A^+ > u + t_1) &\le \frac{P_i(X_{u+t_1} = j | H_A^+ > u)}{P_i(H_A^+ > u + t_1 | H_A^+ > u)} \quad \text{by Lemma 6.4} \\ &\le 2 \cdot 2^{-d} \quad \text{for } d \text{ sufficiently large, by (6.8) and (6.9).} \end{aligned}$$

LEMMA 6.10. *If* $s(i, A, u) \le 2 \cdot 2^{-d}$ *and* $P_i(H_j^+ > u | H_A^+ > u) \ge 2^{-d/2}$ *then* $s(i, A \cup \{j\}, u + t_1) \le 2 \cdot 2^{-d}$.

PROOF. By Lemma 6.4,

$$\begin{aligned} P_i(X_u = k | H_{A\cup\{j\}}^+ > u) &\le \frac{P_i(X_u = k | H_A^+ > u)}{P_i(H_{A\cup\{j\}}^+ > u | H_A^+ > u)} \\ &\le \frac{s(i, A, u)}{P_i(H_j^+ > u | H_A^+ > u)} \\ &\le 2 \cdot 2^{-d/2} \quad \text{by hypothesis.} \end{aligned}$$

So $s(i, A \cup \{j\}, u) \le 2 \cdot 2^{-d/2}$. Apply Lemma 6.6.

PROOF OF LEMMA 4.19(a). By hypothesis (i, A, t) is regular, in particular

$$s(i, A, u) \le 2 \cdot 2^{-d}, \qquad (6.11)$$

for u as at (4.5). Now $(i, A \cup \{j\}, t)$ satisfies condition (a) of Definition 4.4 because (i, A, t) satisfies that condition, so to prove regularity it suffices to verify condition (b): $s(i, A \cup \{j\}, u + t_1) \le 2 \cdot 2^{-d}$. Thus it suffices to prove

$$\begin{aligned} &\text{either (i)} \quad s(i, A \cup \{j\}, u + t_1) \le 2 \cdot 2^{-d} \\ &\text{or} \quad \text{(ii)} \quad \tilde{P}(L_j > L_A | i, A, t) \le 2 \cdot 2^{-d/2}. \end{aligned}$$

First suppose

$$P_i(H_j^+ > u | H_A^+ > u) \ge 2^{-d/2}. \qquad (6.12)$$

Then by (6.11) we can apply Lemma 6.10 and deduce that (i) holds. Now suppose (6.12) fails. Then

$$\tilde{P}(L_j > L_A | i, A, t) = P_i(H_j^+ > t | H_A^+ > t) \quad \text{by (6.1)}$$
$$\le P_i(H_j^+ > u | H_A^+ > t) \quad \text{since } u \le t \text{ by (4.6)}$$
$$\le \frac{P_i(H_j^+ > u | H_A^+ > u)}{P_i(H_A^+ > t | H_A^+ > u)} \quad \text{by Lemma 6.4}$$
$$\le 2^{-d/2}/P_i(H_A^+ > 2^{d/2} | H_A^+ > u)$$

since (6.12) fails, and $t \le 2^{d/2}$ because (i, A, t) is regular,

$$\le 2^{-d/2}/(1 - 2^{-\varepsilon d}) \quad \text{by Lemma 6.3.}$$

So (ii) holds for d sufficiently large.

One more lemma is needed before we can prove part (b) of Lemma 4.19.

LEMMA 6.13. *Suppose $s(i, A, u) \le 2 \cdot 2^{-d}$. Let $j \notin A$. Define*

$$V = \sup\{t \le u: X_t = j, X_{t-} \ne j\}$$
$$= 0 \quad \textit{if no such } t \textit{ exists.}$$

Then $P_i(V > 0, s(j, A \cup \{j\}, u + t_1 - V) > 2 \cdot 2^{-d} | H_A^+ > u) \le ½ \cdot 2^{-d/2}$.

PROOF. Let

$$\Omega_0 = \{H_A^+ > u + t_1\} \cap \{X_t \ne j, u \le t \le u + t_1\}$$
$$\Omega_1 = \{P_i(\Omega_0^c | V, H_A^+ > u) < ⅓\}.$$

Then

$$P(\Omega_1^c | H_A^+ > u) \le 3P(\Omega_0^c | H_A^+ > u)$$
$$\le 3 \cdot 2 \cdot P_\pi(H_{A \cup \{j\}} \le t_1) \quad \text{by Lemma 6.2}$$
$$\le 6 \#(A \cup \{j\}) P_\pi(H_i \le t_1) \quad \text{by symmetry} \tag{6.14}$$
$$\le 6(1 + \#A) t_1 2^{-d}(1 + \varepsilon_2(d)) \quad \text{by Lemma 2.8(b)}$$
$$\le ½ \cdot 2^{-d/2} \quad \text{for } d \text{ sufficiently large, by (4.3).}$$

Next, since $\Omega_0 \subset \{H_A^+ > u\}$ the continuous analog of Lemma 6.4 gives

$$P_i(X(u + t_1) = k | V, \Omega_0) \le \frac{P_i(X(u + t_1) = k | V, H_A^+ > u)}{P_i(\Omega_0 | V, H_A^+ > u)}$$
$$\le \tfrac{3}{2} P_i(X(u + t_1) = k | V, H_A^+ > u) \quad \text{on } \Omega_1 \tag{6.15}$$
$$\le \tfrac{3}{2} \cdot 2^{-d}(1 + 2^{-d}) \quad \text{on } \Omega_1, \text{ by Lemma 6.5}$$
$$\le 2 \cdot 2^{-d} \quad \text{on } \Omega_1, \text{ for } d \text{ sufficiently large.}$$

Now fix $0 < v \le u$. By definition of V, Ω_0 and the Markov property at v,

$$P_i(X(u + t_1) = k | V = v, \Omega_0) = P_j(X(u + t_1 - v) = k | H_{A \cup \{j\}}^+ > u + t_1 - v). \tag{6.16}$$

So

$$s(j, A \cup \{j\}, u + t_1 - v) = \max_k P_j(X(u + t_1 - v) = k | H_{A \cup \{j\}}^+ > u + t_1 - v)$$
$$= \max P_i(X(u + t_1) = k | V = v, \Omega_0) \quad \text{by (6.16).}$$

So

$$s(j, A \cup \{j\}, u + t_1 - V) = \max P_i(X(u + t_1) = k \mid V, \Omega_0) \quad \text{on } \{V > 0\}$$
$$\leq 2 \cdot 2^{-d} \quad \text{on } \Omega_1 \cap \{V > 0\} \quad \text{by (6.15)},$$

and now the lemma follows from (6.14).

PROOF OF LEMMA 4.19(b). By hypothesis (i, A, t) is regular. Let u be as at (4.5). In order that $(j, A \cup \{j\}, L_j)$ satisfy condition (a) of (4.4) we need $L_j \geq t - u$. In order to satisfy (b) we need $s(j, A \cup \{j\}, \hat{u}) \leq 2 \cdot 2^{-d}$, for $\hat{u} = u + t_1 + L_j - t$. So it suffices to prove

(6.17) $$\tilde{P}(L_j < t - u \mid i, A, t) \leq 3 \cdot 2^{-d/2}$$

(6.18) $$\tilde{P}(S(j, A \cup \{j\}, u + t_1 + L_j - t) > 2 \cdot 2^{-d}, t - u \leq L_j < t \mid i, A, t) \leq 2^{-d/2}.$$

Say "X enters j during $[u, t]$" if there exists $r \in [u, t]$ such that $X_r = j$, $X_{r-} \neq j$. To estimate (6.17),

$$\tilde{P}(L_j < t - u \mid i, A, t) = P_i(X \text{ enters } j \text{ during } [u, t] \mid H_A^+ > t) \quad \text{by time-reversal (6.1)}$$
$$\leq \frac{P_i(X \text{ enters } j \text{ during } [u, t] \mid H_A^+ > u)}{P_i(H_A^+ > t \mid H_A^+ > u)} \quad \text{by Lemma 6.4}$$
$$\leq \frac{2P_\pi(H_j^+ \leq t - u)}{1 - 2^{-\varepsilon d}} \quad \text{by Lemmas 6.2 and 6.3, since } t \leq 2^{d/2},$$
$$\leq 2 \cdot 2^{-d/2}(1 + \varepsilon_2(d))/(1 - 2^{-\varepsilon d}) \quad \text{by Lemma 2.8(b)},$$

and so (6.17) holds for sufficiently large d.

To prove (6.18), time-reversal (6.1) shows that the probability in (6.18) equals

$$P_i(s(j, A \cup \{j\}, u + t_1 - V) > 2 \cdot 2^{-d}, V > 0, X \text{ does not enter } j \text{ during } [u, t] \mid H_A^+ > t).$$

Since $u \leq t \leq 2^{d/2}$, Lemmas 6.13 and 6.4 show this quantity is at most

$$\tfrac{1}{2} \cdot 2^{-d/2} / P_i(H_A^+ > t \mid H_A^+ > u).$$

Now
$$P_i(H_A^+ > t \mid H_A^+ > u) \geq P_i(H_A^+ > 2^{d/2} \mid H_A^+ \geq u)$$
$$\geq 1 - 2^{-\varepsilon d} \quad \text{by Lemma 6.3}$$

and this establishes (6.18).

REFERENCE

TOVEY, C. A. (1981). Polynomial local improvement algorithms in combinatorial optimization. Ph.D. thesis, Operations Research, Stanford University.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720