JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS: Vol. 62, No. 1, JULY 1989

Simulated Annealing with Noisy or Imprecise Energy Measurements[1]

S. B. GELFAND[2] AND S. K. MITTER[3]

Communicated by R. Conti

Abstract. The annealing algorithm (Ref. 1) is modified to allow for noisy or imprecise measurements of the energy cost function. This is important when the energy cannot be measured exactly or when it is computationally expensive to do so. Under suitable conditions on the noise/imprecision, it is shown that the modified algorithm exhibits the same convergence in probability to the globally minimum energy states as the annealing algorithm (Ref. 2). Since the annealing algorithm will typically enter and exit the minimum energy states infinitely often with probability one, the minimum energy state visited by the annealing algorithm is usually tracked. The effect of using noisy or imprecise energy measurements on tracking the minimum energy state visited by the modified algorithms is examined.

Key Words. Simulated annealing, combinatorial optimization, noisy measurements, Markov chains.

1. Introduction

Motivated by hard combinatorial optimization problems such as arise in computer design and operations research, Kirkpatrick, Gelatt, and Vecchi (Ref. 1) and, independently, Cerny (Ref. 3) have proposed a random optimization algorithm called simulated annealing. The annealing algorithm stands in contrast to heuristic methods based on iterative improvement in which only decreases in the cost function are allowed at each iteration. In the annealing algorithm, increases in the cost function are allowed with

[1] The research reported here has been supported under Contracts AFOSR-85-0227, DAAG-29-84-K-0005, and DAAL-03-86-K-0171 and a Purdue Research Initiation Grant.

[2] Assistant Professor, Computer Vision and Image Processing Laboratory, School of Electrical Engineering, Purdue University, West Lafayette, Indiana.

[3] Professor, Center for Intelligent Control Systems and Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts.

0022-3239/89/0700-0049$06.00/0

certain probability. This probability is slowly decreased to zero. Simulated annealing is based on an analogy to a physical system which is first melted and then cooled or "annealed" into a low energy state. In this analogy, the cost of the optimization problem is identified with the energy of an imaginary physical system; see Ref. 1. The annealing algorithm has been applied with mixed success to a variety of difficult problems (Refs. 4–7). In addition, the annealing algorithm has sparked considerable theoretical interest, and investigations into its convergence have generated fundamentally new results in the theory of nonstationary Markov chains; see Refs. 2 and 8–10, and see Ref. 11 for a review.

The annealing algorithm may be described as follows. Let Σ be a finite set, and let $U(\cdot)$ be a real-valued function on Σ, the cost or energy function. The goal is to find an element of Σ which minimizes or nearly minimizes $U(\cdot)$. Let $\{T_k\}$ be a sequence of positive numbers, the temperature schedule. T_k will tend to zero at a suitably slow rate. Let $Q=[q_{ij}]$ be a $\Sigma\times\Sigma$ stochastic matrix. Typically, Q is irreducible and may also satisfy a reversibility condition such as

$$q_{ij}=q_{ji}, \qquad \text{for all } i,j\in\Sigma.$$

The annealing algorithm consists of simulating a random process $\{X_k\}$ which takes values in Σ and whose successive values are determined in the following manner. Suppose that $X_k=i$. Then, select a candidate state j with probability q_{ij}. If $U(j)-U(i)\leq 0$, set $X_{k+1}=j$; if $U(j)-U(i)>0$, set $X_{k+1}=j$ with probability $\exp\{-[U(j)-U(i)]/T_k\}$; otherwise, set $X_{k+1}=i$. It is seen that $\{X_k\}$ is in fact a nonstationary Markov chain with one-step transition probabilities

$$P\{X_{k+1}=j\,|\,X_k=i\}$$

$$=\begin{cases} q_{ij}\ \exp\{-[U(j)-U(i)]/T_k\}, & \text{if } U(j)-U(i)>0,\\ q_{ij}, & \text{if } U(j)-U(i)\leq 0, \end{cases} \tag{1}$$

for all $i,j\in\Sigma$ with $j\neq i$.[4] We call $\{X_k\}$ the annealing chain. Note that $T_k>0$ implies that the annealing chain can with positive probability make transitions to higher energy states and so escape from local minima of the energy function. Note also that since $T_k\to 0$, the probability of the annealing chain making a transition to a higher energy state tends to zero. Intuitively, if T_k is decreased to zero at a suitably slow rate, then the annealing chain eventually spends most of its time among and hopefully converges in an appropriate probabilistic sense to the minimum energy states.

[4] This also specifies $P\{X_{k+1}=i\,|\,X_k=i\}$ when $P\{X_k=i\}>0$.

Much of the theoretical interest in the annealing algorithm has focused on setting conditions on the temperature schedule such that the annealing chain converges in probability to the set of minimum energy states, i.e., setting conditions on $\{T_k\}$ such that

$$\lim_{k\to\infty} P\{X_k \in S\} = 1,$$

where

$$S = \{i \in \Sigma\colon U(i) \leq U(j), \forall j \in \Sigma\}.$$

Under a reversibility condition on Q, Hajek (Ref. 2) has given a characterization of monotone decreasing temperature schedules which obtain convergence in probability; Tsitsiklis (Refs. 10 and 11) later removed the reversibility condition (see Theorem 3.1).

In this paper, we consider modifications of the annealing algorithm to allow for noisy (i.e., with random error) or imprecise (i.e., with deterministic error) measurements of the energy differences which are used in selecting successive states. This is important when the energy differences cannot be computed exactly or when it is simply too costly to do so. Grover (Ref. 12) has applied such a modified algorithm to a circuit design problem and achieved significant reductions in computational load with comparable quality solutions. Here, we rigorously describe and analyze these modified algorithms. Our approach involves formulating the modified algorithms in such a way that they also involve simulating Markov chains. We then show that, under suitable conditions on the noise/imprecision and temperature schedule, the one-step transition probabilities of the modified chains and annealing chain are asymptotically equivalent; and, using results from (Ref. 10), we obtain that the modified chains converge in probability to the minimum energy states if and only if the annealing chain does. Since in general the annealing chain will only converge in probability to the minimum energy states, it will enter and exit the minimum energy states infinitely often with probability one. Hence, in applying the annealing algorithm, one usually keeps track of the minimum energy states visited up to the current time; this may be done recursively, since the energy differences are computed at each iteration. We examine the effect of using noisy or imprecise measurements of the energy differences on tracking the minimum energy state visited by the modified algorithms.

This paper is organized as follows. In Section 2, we describe the annealing algorithm modified for noisy or imprecise energy measurements. In Section 3, after reviewing a result from Ref. 10, we analyze the convergence in probability of the modified algorithms. In Section 4, we examine the problem of tracking the minimum energy state visited by the modified algorithms. In Section 5, we conclude with a brief discussion.

2. Modification of the Annealing Algorithm

We first describe the annealing algorithm modified for noisy measurements of the energy differences used to select successive states (by noisy, we mean with random error). The annealing algorithm with noisy measurements consists of simulating a random process $\{Y_k\}$ which takes values in Σ. The successive values of $\{Y_k\}$ are obtained in the same fashion as the annealing chain $\{X_k\}$ (see Section 1), except that, at each time k, the energy difference $U(j)-U(i)$ between the candidate state j and the current state i is replaced by $U(j)-U(i)+W_k$, where W_k is a real-valued random variable. More precisely, we define $\{Y_k\}$ as follows. Given that Y_1 is defined, let W_1 be a real-valued random variable with

$$P\{W_1 \leq \lambda \mid Y_1\} = F_1(\lambda), \qquad \forall \lambda \in \mathbb{R}.$$

Given that $Y_1, \ldots, Y_k$, $W_1, \ldots, W_k$ have been defined, let Y_{k+1} be a Σ-valued random variable with

$$P\{Y_{k+1}=j \mid Y_1, \ldots, Y_{k-1}, Y_k=i, W_1, \ldots, W_{k-1}, W_k=\lambda\}$$

$$= \begin{cases} q_{ij} \exp\{-[U(j)-U(i)+\lambda]/T_k\}, & \text{if } U(j)-U(i)+\lambda > 0, \\ q_{ij}, & \text{if } U(j)-U(i)+\lambda \leq 0, \end{cases} \tag{2}$$

for all $i, j \in \Sigma$ with $j \neq i$ and all $\lambda \in \mathbb{R}$; and let W_{k+1} be a real-valued random variable with

$$P\{W_{k+1} \leq \lambda \mid Y_1, \ldots, Y_{k+1}, W_1, \ldots, W_k\} = F_{k+1}(\lambda), \qquad \forall \lambda \in \mathbb{R}. \tag{3}$$

Proceeding in this way, we inductively define a sequence of random variables $\{Y_k, W_k\}$.

It is easy to show that $\{Y_k\}$ defined as above is a Markov chain with one-step transition probabilities given by

$$P\{Y_{k+1}=j \mid Y_k=i\} = E\{P\{Y_{k+1}=j \mid Y_k, W_k\} \mid Y_k=i\}$$

$$= \underset{W_k}{E}\{P\{Y_{k+1}=j \mid Y_k=i, W_k\}\}$$

$$= \int_{\lambda > U(i)-U(j)} q_{ij} \exp\{-[U(j)-U(i)+\lambda]/T_k\}\, dF_k(\lambda)$$

$$+ q_{ij} F_k[U(i)-U(j)], \qquad \forall j \neq i. \tag{4}$$

From now on, we only consider the case where W_k is Gaussian with mean

0 and variance $\sigma_k^2 > 0$. Hence (4) can be written as

$$P\{Y_{k+1} = j \mid Y_k = i\}$$
$$= \int_{U(i)-U(j)}^{\infty} q_{ij} \exp\{-[U(j) - U(i) + \lambda]/T_k\}\, dN(0, \sigma_k^2)(\lambda)$$
$$+ q_{ij} N(0, \sigma_k^2)[-\infty, U(i) - U(j)], \qquad \forall j \neq i, \tag{5}$$

where $N(m, a)(\cdot)$ denotes one-dimensional normal measure with mean m and variance a. We refer to $\{Y_k\}$ as the annealing chain with noisy measurements.

We next describe the annealing algorithm modified for imprecise measurements of the energy differences used to select successive states (by imprecise, we mean with deterministic error). The annealing algorithm with imprecise measurements consists of simulating a random process $\{Z_k\}$ which takes values in Σ. The successive values of $\{Z_k\}$ are obtained in the same fashion as the annealing chain $\{X_k\}$ (see Section 1), except that, at each time k, the energy difference $U(j) - U(i)$ between the candidate state j and the current state i is replaced by $U(j) - U(i) + \beta_k$, where β_k is a number. It is seen that the process $\{Z_k\}$ is a Markov chain with one-step transition probabilities

$$P\{Z_{k+1} = j \mid Z_k = i\}$$
$$= \begin{cases} q_{ij} \exp\{-[U(j) - U(i) + \beta_k]/T_k\}, & \text{if } U(j) - U(i) + \beta_k > 0, \\ q_{ij}, & \text{if } U(j) - U(i) + \beta_k \leq 0, \end{cases} \tag{6}$$

for all $i, j \in \Sigma$ with $j \neq i$. We refer to $\{Z_k\}$ as the annealing chain with imprecise measurements.

3. Convergence of the Modified Algorithms

In this section, we give conditions such that the modified annealing chains converge in probability to the set of globally minimum energy states. We first state a result from Ref. 10 on the convergence of a class of nonstationary Markov chains.

Theorem 3.1. (Ref. 10). For each $\epsilon \in [0, 1)$, let $\{N_k^\epsilon\}$ be a Markov chain with state space Σ which satisfies

$$c_1 \epsilon^{\alpha(i,j)} \leq P\{N_{k+1}^\epsilon = j \mid N_k^\epsilon = i\} \leq c_2 \epsilon^{\alpha(i,j)}, \tag{7}$$

for all $i, j \in \Sigma$ with $j \neq i$, where $\alpha(i, j) \in [0, \infty]$ and c_1, c_2 are positive constants.

Suppose that $\{N_k^\epsilon\}$ is irreducible for all $\epsilon > 0$ and that the irreducible components of $\{N_k^0\}$ are aperiodic. Let $\{\epsilon_k\}$ be a sequence of numbers with $\epsilon_k \in (0, 1)$ and $\epsilon_k \downarrow 0$, and let $\{N_k\}$ be a Markov chain with state space Σ which satisfies

$$P\{N_{k+1} = j \mid N_k = i\} = P\{N_{k+1}^{\epsilon_k} = j \mid N_k^{\epsilon_k} = i\}, \qquad \forall j \neq i.$$

Let $\Lambda \subset \Sigma$. Then, there exists $\delta^* \in [0, \infty]$, depending only on $\alpha(\cdot, \cdot)$ and Λ, such that $\lim_{k\to\infty} P\{N_k \in \Lambda\} = 1$, iff

$$\sum_{k=1}^{\infty} \epsilon_k^{\delta^*} = \infty.$$

Remark 3.1. The statement of Theorem 3.1 in Ref. 10 assumes that (7) holds for all $i, j \in \Sigma$; but it is enough that (7) hold only for $j \neq i$ as stated above.

Remark 3.2. For each $T \geq 0$, let $\{X_k^T\}$ be the constant temperature ($T_k = T$) annealing chain. Suppose that Q is irreducible. Then, $\{X_k^T\}$ is irreducible for all $T > 0$ and the irreducible components of $\{X_k^0\}$ must be aperiodic. Let

$$\epsilon = \exp(-1/T), \qquad \epsilon_k = \exp(-1/T_k),$$

and

$$\alpha(i, j) = \begin{cases} \max\{0, U(j) - U(i)\}, & \text{if } q_{ij} > 0, \\ \infty, & \text{if } q_{ij} = 0, \end{cases}$$

for all $i, j \in \Sigma$ with $j \neq i$. Then, Theorem 3.1 may be applied with

$$N_k^\epsilon = X_k^T, \qquad N_k = X_k, \qquad \Lambda = S$$

to obtain: there exists a $\delta^* \in [0, \infty]$ such that $\lim_{k\to\infty} P\{X_k \in S\} \to 1$, iff

$$\sum_{k=1}^{\infty} \exp(-\delta^*/T_k) = \infty. \tag{8}$$

If Q satisfies a certain reversibility condition, it may be shown that $\delta^* < \infty$ and has a simple interpretation as the maximum "depth," suitably defined, of all local minima of $U(\cdot)$ which are not global minima; see Ref. 2.

We next apply Theorem 3.1 to the modified annealing chains $\{Y_k\}$ and $\{Z_k\}$. We treat $\{Y_k\}$ in detail and then state the corresponding results for $\{Z_k\}$ without proof.

Proposition 3.1. Suppose that $T_k \to 0$ and $\sigma_k = o(T_k)$, as $k \to \infty$. Then,

$$P\{Y_{k+1} = j \mid Y_k = i\} \sim P\{X_{k+1} = j \mid X_k = i\}, \qquad \text{as } k \to \infty, \tag{9}$$

for all $i, j \in \Sigma$ with $j \neq i$.

Proof. Fix $i, j \in \Sigma$ with $j \neq i$ and $q_{ij} > 0$. Let

$$a_k = \int_{U(i)-U(j)}^{\infty} q_{ij} \exp\{-[U(j) - U(i) + \lambda]/T_k\}\, dN(0, \sigma_k^2)(\lambda),$$

$$b_k = q_{ij} N(0, \sigma_k^2)(-\infty, U(i) - U(j)],$$

so that (5) becomes

$$P\{Y_{k+1} = j \mid Y_k = i\} = a_k + b_k. \tag{10}$$

Since $\sigma_k = o(1)$, we have

$$\lim_{k\to\infty} a_k = 0, \qquad \text{if } U(j) - U(i) < 0, \tag{11}$$

$$\lim_{k\to\infty} b_k = q_{ij}, \qquad \text{if } U(j) - U(i) < 0. \tag{12}$$

Also,

$$\lim_{k\to\infty} b_k = q_{ij}/2, \qquad \text{if } U(j) - U(i) = 0. \tag{13}$$

We make the following claim.

Claim 3.1. The following results hold as $k \to \infty$:

$$a_k \to q_{ij}/2, \qquad \text{if } U(j) - U(i) = 0, \tag{14}$$

$$a_k \sim q_{ij} \exp(-[U(j) - U(i)]/T_k), \qquad \text{if } U(j) - U(i) > 0, \tag{15}$$

$$b_k = o(\exp\{-[U(j) - U(i)]/T_k\}), \qquad \text{if } U(j) - U(i) > 0. \tag{16}$$

Suppose that Claim 3.1 is true. Then, combining (10)-(16) gives (9) as required. It remains to prove Claim 3.1.

Proof of Claim 3.1. We have

$$a_k = q_{ij} \exp\{-[U(j) - U(i)]/T_k\} \times \int_{[U(i)-U(j)]/T_k}^{\infty} \exp(-\lambda)\, dN(0, \sigma_k^2/T_k^2)(\lambda), \tag{17}$$

after a change of variable. Observe that $\sigma_k = o(T_k)$ implies that $N(0, \sigma_k^2/T_k^2)(\cdot)$ converges weakly to the unit measure concentrated at the origin. It follows that

$$\lim_{k\to\infty} \int_{[U(i)-U(j)]/T_k}^{\infty} \exp(-\lambda)\, dN(0, \sigma_k^2/T_k^2)(\lambda)$$
$$= \begin{cases} 1/2, & \text{if } U(j)-U(i)=0, \\ 1, & \text{if } U(j)-U(i)>0. \end{cases} \tag{18}$$

Combining (17), (18) gives (14), (15). Finally, if $U(j)-U(i)>0$, then since $\sigma_k = o(T_k)$,

$$\begin{aligned} b_k &= q_{ij} N(0, \sigma_k^2)(-\infty, U(i)-U(j)] \\ &\le \exp\{-[U(j)-U(i)]^2/2\sigma_k^2\} \\ &= o(\exp\{-[U(j)-U(i)]/T_k\}), \qquad \text{as } k\to\infty, \end{aligned}$$

where we have used the standard estimate

$$N(0,1)[x,\infty) \le \exp(-x^2/2), \qquad \text{for } x \ge 0.$$

This proves (16) and, hence, Claim 3.1 and Proposition 3.1. □

Corollary 3.1. Suppose that Q is irreducible, $T_k \downarrow 0$, and $\sigma_k = o(T_k)$, as $k\to\infty$. Then,

$$\lim_{k\to\infty} P\{Y_k \in S\} = 1 \text{ iff } \lim_{k\to\infty} P\{X_k \in S\} = 1.$$

Proof. In Remark 3.2 following Theorem 3.1, we showed that Theorem 3.1 may be applied to $\{X_k\}$ to obtain that $\lim_{k\to\infty} P\{X_k \in S\} = 1$, iff (8) holds. In view of Proposition 3.1, Theorem 3.1 may also be applied to $\{Y_k\}$ to obtain that $\lim_{k\to\infty} P\{Y_k \in S\} = 1$, iff (8) holds with the same value of δ^*. □

Remark 3.3. It is not possible to assert in general that

$$P\{Y_{k+1} = i \mid Y_k = i\} \sim P\{X_{k+1} = i \mid X_k = i\}.$$

For example, if $q_{ii} = 0$ and $q_{ij} = 0$ for all $j \in \Sigma$ with $U(j)-U(i)>0$, then $P\{X_{k+1} = i \mid X_k = i\}$ is zero but $P\{Y_{k+1} = i \mid Y_k = i\}$ is strictly positive, corresponding to the positive probability of not making a transition to a state with the same or lower energy. This is why we must require only that (7) holds for $j \ne i$ in Theorem 3.1 to obtain Proposition 3.1 and, hence, Corollary 3.1.

The corresponding results for $\{Z_k\}$ are as follows.

Proposition 3.2. Suppose that $T_k \to 0$ and $\beta_k = o(T_k)$, as $k \to \infty$. Then,

$$P\{Z_{k+1} = j \mid Z_k = i\} \sim P\{X_{k+1} = j \mid X_k = i\}, \qquad \text{as } k \to \infty,$$

for all $i, j \in \Sigma$ with $j \neq i$.

Corollary 3.2. Suppose that Q is irreducible, $T_k \downarrow 0$, and $\beta_k = o(T_k)$, as $k \to \infty$. Then,

$$\lim_{k\to\infty} P\{Z_k \in S\} = 1 \text{ iff } \lim_{k\to\infty} P\{X_k \in S\} = 1.$$

4. Tracking the Minimum Energy State

As pointed out above, when implementing the annealing algorithm, one normally keeps track of the minimum energy state visited by the annealing chain up to the current time. The reason for this is that only convergence in probability of the annealing chain to the set S of minimum energy states can be guaranteed, and typically the annealing chain will enter and leave S infinitely often (with probability one). The energy differences which are used to select the successive states of the annealing chain may also be used to compute recursively the minimum energy state visited by the annealing chain. For the modified algorithms, noisy or imprecise measurements of the energy differences are used to select the successive states of the modified chains. In this section, we examine the effect of using these same noisy or imprecise measurements on computing the minimum energy state visited by the modified chains.

We introduce the following notation. For every $m \geq n$, let

$$i(n, m) = \arg\min_{n \leq k \leq m} [U(X_k) - U(X_n)], \tag{19a}$$

$$j(n, m) = \arg\min_{n \leq k \leq m} \left[U(Y_k) - U(Y_n) + \sum_{l=n}^{k-1} W_l 1_{\{Y_{l+1} \neq Y_l\}} \right], \tag{19b}$$

$$k(n, m) = \arg\min_{n \leq k \leq m} \left[U(Z_k) - U(Z_n) + \sum_{l=n}^{k-1} \beta_l 1_{\{Z_{l+1} \neq Z_l\}} \right], \tag{19c}$$

and

$$x_{n,m} = X_{i(n,m)}, \qquad y_{n,m} = Y_{j(n,m)}, \qquad z_{n,m} = Z_{k(n,m)},$$

$$x_m = x_{1,m}, \qquad y_m = y_{1,m}, \qquad z_m = z_{1,m}.$$

In words, $x_{n,m}$ is the minimum energy state visited by X_k between time n and time m, while $y_{n,m}$ and $z_{n,m}$ are estimates of the minimum energy states visited by Y_k and Z_k, respectively, between time n and time m. Note that

$\{x_{n,m}\}_{m\geq n}$ may be computed recursively from the values of the energy differences $U(X_{k+1})-U(X_k)$ which are generated in simulating $\{X_k\}$, and that $\{y_{n,m}\}_{m\geq n}$ and $\{z_{n,m}\}_{m\geq n}$ may be computed recursively from the values of the noisy/imprecise energy differences $U(Y_{k+1})-U(Y_k)+W_k$ and $U(Z_{k+1})-U(Z_k)+\beta_k$ which are generated in simulating $\{Y_k\}$ and $\{Z_k\}$, respectively. Note also that the noise/imprecision on self-transitions of $\{Y_k\}$ and $\{Z_k\}$ is ignored, since it is known when a self-transition is made.

If

$$\lim_{k\to\infty} P\{X_k\in S\}=1,$$

then

$$\lim_{n\to\infty} P\{x_k\in S, \forall k\geq n\}=1,$$

or equivalently, $x_k\in S$, for large enough k with probability one. It is also clear that this implication does not hold in general with X_k, x_k replaced by Y_k, y_k or Z_k, z_k. The problem is that large initial noise/imprecision can result in $y_k\notin S$ or $z_k\notin S$ for all k with positive probability. A less useful but still relevant result is that, if

$$\lim_{k\to\infty} P\{X_k\in S\}=1,$$

then

$$\lim_{n\to\infty} P\{x_{n,k}\in S, \forall k\geq n\}=1.$$

We show that, under suitable conditions, this implication holds with X_k, x_k replaced by Y_k, y_k or Z_k, z_k. As in Section 3, we treat $\{Y_k\}$ in detail and then give the corresponding results for $\{Z_k\}$ which require little proof.

Let

$$M_{n,k}=\sum_{l=n}^{k-1} W_l 1_{\{Y_{l+1}\neq Y_l\}}, \qquad \forall k\geq n. \tag{20}$$

Intuitively, if $P\{Y_n\in S\}$ is large and $\min_{k>n} M_{n,k}\geq 0$ with large probability, then $P\{y_{n,k}\in S, \forall k\geq n\}$ should be large. If the indicator functions in (20) were absent, then since the $\{W_k\}$ are independent, $\{M_{n,k}\}_{k\geq n}$ would be a martingale. However, it is not hard to see that the presence of the indicator functions biases $M_{n,k}$ toward negative values [see (2)]. Let $\mathcal{F}_{n,k}$ be the σ-field generated by $\{Y_n,\ldots,Y_k, W_n,\ldots,W_{k-1}\}$ for $k\geq n$. Also, let

$$P_n\{\cdot\}=P\{\cdot\mid Y_n\in S\} \qquad \text{and} \qquad E_n\{\cdot\}=E\{\cdot\mid Y_n\in S\};$$

assume that $P\{Y_n\in S\}>0$.

Lemma 4.1. $\{M_{n,k}\}_{k\geq n}$ is an $(\{\mathcal{F}_{n,k}\}_{k\geq n}, P_n)$ supermartingale.

Proof. First, observe that if $\{M_{n,k}\}_{k\geq n}$ is an $(\{\mathcal{F}_{n,k}\}_{k\geq n}, P)$ supermartingale, then clearly $E_n\{|M_{n,k}|\}<\infty$ and for $A\in\mathcal{F}_{n,k}$,

$$E_n\{M_{n,k+1}1_A\}=\frac{E\{M_{n,k+1}1_{A\cap\{Y_n\in S\}}\}}{P\{Y_n\in S\}}\leq\frac{E\{M_{n,k}1_{A\cap\{Y_n\in S\}}\}}{P\{Y_n\in S\}}$$

$$=E_n\{M_{n,k}1_A\},$$

since $\{Y_n\in S\}\in\mathcal{F}_{n,k}$, and so $\{M_{n,k}\}_{k\geq n}$ is an $(\{\mathcal{F}_{n,k}\}_{k\geq n}, P_n)$ supermartingale.

We show that $\{M_{n,k}\}_{k\geq n}$ is an $(\{\mathcal{F}_{n,k}\}_{k\geq n}, P)$ supermartingale. Clearly, $M_{n,k}$ is $\mathcal{F}_{n,k}$ measurable and $E\{|M_{n,k}|\}<\infty$. Furthermore,

$$\begin{aligned}
&E\{M_{n,k+1}-M_{n,k}\,|\,\mathcal{F}_{n,k}\}\\
&=E\{W_k\cdot 1_{\{Y_{k+1}\neq Y_k\}}\,|\,Y_n,\ldots,Y_k,W_n,\ldots,W_{k-1}\}\\
&=E\{W_kP\{Y_{k+1}\neq Y_k\,|\,Y_n,\ldots,Y_k,W_n,\ldots,W_k\}\\
&\qquad\qquad |\,Y_n,\ldots,Y_k,W_n,\ldots,W_{k-1}\}\\
&=E\{W_kP\{Y_{k+1}\neq Y_k\,|\,Y_k,W_k\}|\,Y_n,\ldots,Y_k,W_n,\ldots,W_{k-1}\}\\
&=\underset{W_k}{E}\{W_kP\{Y_{k+1}\neq Y_k\,|\,Y_k,W_k\}\}\\
&=\int_0^\infty\lambda(P\{Y_{k+1}\neq Y_k\,|\,Y_k,W_k=\lambda\}\\
&\qquad -P\{Y_{k+1}\neq Y_k\,|\,Y_k,W_k=-\lambda\})\,dN(0,\sigma_k^2)(\lambda)\\
&\leq(1/2)E\{|W_k|\}\sup_{\lambda\geq 0}[P\{Y_{k+1}\neq Y_k\,|\,Y_k,W_k=\lambda)\\
&\qquad\qquad -P\{Y_{k+1}\neq Y_k\,|\,Y_k,W_k=-\lambda\}]\\
&\leq 0,\ \text{w.p.1.}
\end{aligned}$$

Here, the third equality follows from (2), the fourth equality from (3), and the final inequality from (2). Hence, $\{M_{n,k}\}_{k\geq n}$ is indeed an $(\{\mathcal{F}_{n,k}\}_{k\geq n}, P)$ supermartingale, and so an $(\{\mathcal{F}_{n,k}\}_{k\geq n}, P_n)$ supermartingale. □

Proposition 4.1. Suppose that $\sum_{k=1}^{\infty}\sigma_k<\infty$. Under this condition, if $\lim_{k\to\infty}P\{Y_k\in S\}=1$, then $\lim_{n\to\infty}P\{y_{n,k}\in S,\forall k\geq n\}=1$.

Proof. Let

$$\gamma=\min_{j\in\Sigma\setminus S}U(j)-\min_{i\in\Sigma}U(i). \tag{21}$$

Then, for $m \geq n$,

$$P\{y_{n,k} \in S, \forall n \leq k \leq m\}$$

$$\geq P\{Y_n \in S, \min_{n<k\leq m} [U(Y_k) - U(Y_n) + M_{n,k}] > 0\}$$

$$\geq P\{Y_n \in S, \min_{\substack{n<k\leq m \\ Y_k \in \Sigma \setminus S}} M_{n,k} > -\gamma\}$$

$$\geq P\{Y_n \in S, \min_{n<k\leq m} M_{n,k} > -\gamma\}$$

$$= P\{Y_n \in S\} P_n\{\min_{n<k\leq m} M_{n,k} > -\gamma\}. \tag{22}$$

Now, by Lemma 4.1, $\{M_{n,k}\}_{k\geq n}$ is a P_n-supermartingale. Hence, by the supermartingale inequality (Ref. 13, Theorem 35.2),

$$P_n\{\min_{n<k\leq m} M_{n,k} > -\gamma\} \geq 1 - (1/\gamma) E_n\{|M_{n,m}|\}$$

$$= 1 - (1/\gamma) E_n \left\{ \left| \sum_{k=n}^{m-1} W_k 1_{\{Y_{k+1} \neq Y_k\}} \right| \right\}$$

$$\geq 1 - (1/\gamma) \sum_{k=n}^{m-1} E_n^{1/2}\{W_k^2 1_{\{Y_{k+1} \neq Y_k\}}\}$$

$$\geq 1 - (1/\gamma) \sum_{k=n}^{m-1} \sigma_k. \tag{23}$$

Combining (22) and (23), and letting $m \to \infty$ gives

$$P\{y_{n,k} \in S, \forall k \geq n\} \geq P\{Y_n \in S\} \left(1 - (1/\gamma) \sum_{k=n}^{\infty} \sigma_k\right),$$

and so

$$\liminf_{n\to\infty} P\{y_{n,k} \in S, \forall k \geq n\} \geq \liminf_{n\to\infty} P\{Y_n \in S\},$$

and the proposition follows. □

The corresponding result for $\{Z_k\}$ is as follows.

Proposition 4.2. Suppose that $\sum_{k=1}^{\infty} |\beta_k| < \infty$. Under this condition, if $\lim_{k\to\infty} P\{Z_k \in S\} = 1$, then $\lim_{n\to\infty} P\{z_{n,k} \in S, \forall k \geq n\} = 1$.

Proof. Let γ be given by (21). It is easy to see that

$$P\{z_{n,k} \in S, \forall k \geq n\} = P\{Z_n \in S\} \text{ if } \sum_{k=n}^{\infty} |\beta_k| < \gamma,$$

and so

$$\liminf_{n\to\infty} P\{z_{n,k} \in S, \forall k \geq n\} = \liminf_{n\to\infty} P\{Z_n \in S\},$$

and the proposition follows. □

5. Conclusions

We have considered modifications of the annealing algorithm which allow for noise or imprecision in the measurements of the energy differences which are used to select successive states. These modified algorithms like the annealing algorithm involve the simulation of nonstationary Markov chains. We showed that under suitable conditions these modified chains exhibit the same convergence in probability to the minimum energy states as the annealing chain. We also investigated the effect of using the noisy or imprecise energy differences to track the minimum energy state visited by the modified chains.

We believe that our results may be relevant to implementing the annealing algorithm in a semiparallel fashion. For example, consider the problem of updating the state of a finite lattice, each site of which has a number associated with it (this situation arises in the problem of image reconstruction from noisy observations where the sites are pixels and the numbers correspond to gray levels; cf. Ref. 4). There are many ways to update the state. It may be done asynchronously, with the sites updated sequentially in either a fixed or random order; or it may be done synchronously, with the sites updated in parallel. Our results suggest that if the state is updated sychronously but with sufficiently many asynchronous updates (as time tends to infinity and temperature tends to zero), then the same convergence to the global minima is obtained as with a purely asynchronous implementation. It is known that, in the zero-temperature algorithm, the asymptotic behavior of asynchronous and synchronous implementations is different (in the synchronous case, there may not even be convergence to a local minimum; cf. Ref. 14). Furthermore, it is not clear in the zero-temperature algorithm whether sparse asynchronous updates are sufficient for convergence to a local minimum. It seems that the randomness in the annealing algorithm is helpful in this way.

References

1. KIRKPATRICK, S., GELATT, C. D., and VECCHI, M., *Optimization by Simulated Annealing*, Science, Vol. 220, pp. 621-680, 1983.

2. HAJEK, B., *Cooling Schedules for Optimal Annealing*, Mathematics of Operations Research, Vol. 13, pp. 311–329, 1988.
3. CERNY, V., *A Thermodynamical Approach to the Travelling Statesman Problem: An Efficient Simulation Algorithm*, Journal of Optimization Theory and Applications, Vol. 45, pp. 41–51, 1985.
4. GERMAN, S., and GERMAN, D., *Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-6, pp. 721–741, 1984.
5. GOLDEN, B., and SKISCIM, C., *Using Simulated Annealing to Solve Routing and Location Problems*, Naval Research Logistics Quarterly, Vol. 33, pp. 261–279, 1986.
6. JOHNSON, D. S., ARAGON, C. R., MCGEOCH, L. A., and SCHEVON, C., *Optimization by Simulated Annealing: An Experimental Evaluation*, Preprint, 1985.
7. EL GAMAL, A., HEMACHANDRA, L., SHPERLING, I., and WEI, W., *Using Simulated Annealing to Design Good Codes*, IEEE Transactions on Information Theory, Vol. IT-33, pp. 116–123, 1987.
8. GIDAS, B., *Nonstationary Markov Chains and Convergence of the Annealing Algorithm*, Journal of Statistical Physics, Vol. 39, pp. 73–131, 1985.
9. MITRA, D., ROMEO, F., and SANGIOVANNI-VINCENTELLI, A., *Convergence and Finite-Time Behavior of Simulated Annealing*, Advances in Applied Probability, Vol. 18, pp. 747–771, 1986.
10. TSITSIKLIS, J., *Markov Chains with Rare Transitions and Simulated Annealing*, Mathematics of Operations Research, Vol. 14, pp. 70–90, 1989.
11. TSITSIKLIS, J., *A Survey of Large Time Asymptotics of Simulated Annealing Algorithms*, Massachusetts Institute of Technology, Laboratory for Information and Decision Systems, Report No. LIDS-P-1623, 1986.
12. GROVER, L., *Simulated Annealing Using Approximate Calculations*, Preprint, 1986.
13. BILLINGSLEY, P., *Probablity and Measure*, Wiley, New York, New York, 1978.
14. GOLES, E., and VICHNIAC, G., *Lyapunov Functions for Parallel Neural Networks*, Proceedings of the AIP Conference on Neural Networks for Computing, Snowbird, Utah, pp. 165–181, 1986.