ELSEVIER
Discrete Applied Mathematics 119 (2002) 37–57
DISCRETE APPLIED MATHEMATICS

On the convergence of generalized hill climbing algorithms ☆

A.W. Johnson[a,]*, S.H. Jacobson[b]

[a] Department of Mathematical Sciences, United States Military Academy, West Point, NY 10996, USA
[b] Department of Mechanical and Industrial Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA

Received 8 July 1999; received in revised form 11 April 2000; accepted 17 September 2000

Abstract

Generalized hill climbing (GHC) algorithms provide a general local search strategy to address intractable discrete optimization problems. GHC algorithms include as special cases stochastic local search algorithms such as simulated annealing and the noising method, among others. In this paper, a proof of convergence of GHC algorithms is presented, that relaxes the sufficient conditions for the most general convergence proof for stochastic local search algorithms in the literature. Note that classical convergence proofs for stochastic local search algorithms require either that an exponential distribution be used to model the acceptance of candidate solutions along a search trajectory, or that the Markov chain model of the algorithm must be reversible. The proof in this paper removes these limitations, by introducing a new path concept between global and local optima. Convergence is based on the asymptotic behavior of path probabilities between local and global optima. Examples are given to illustrate the convergence conditions. Implications of this result are also discussed. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Simulated annealing; Threshold accepting; Hill climbing algorithms; Discrete optimization; Convergence; Local search

1. Introduction

Many discrete optimization (minimization) problems belong to a class of problems that are difficult to solve (i.e., the class of NP-hard problems [8]). There are no known polynomial-time algorithms that can solve any problem in this class. Therefore, heuristic methods have been developed that efficiently find near-optimal solutions.

☆ This research is supported in part by the Air Force Office of Scientific Research (F49620-98-1-0111) and the National Science Foundation (DMI-9907980). The authors wish to thank Dr. Neal Glassman, Dr. W. Garth Frazier, and Dr. James Malas for their support of this research, as well as two anonymous referees for their insightful comments that have lead to a significantly improved manuscript.

* Corresponding author. Tel.: +1-845-938-7142.

PII: S0166-218X(01)00264-5

Sangiovanni-Vincentelli [16] separates these methods into two conceptual classes: a class that computes the best solution constructively, starting from raw data, and a class that iteratively improves upon an existing solution.

Iterative algorithms are specified by the rules for generating and accepting new solutions, and by termination criteria. *Local search* [17] is a simple iterative algorithm that uses the concepts of a *neighborhood* and an *objective function*. Given a current solution, its neighborhood is the set of all solutions that can be generated by single transformation of some aspect of it, while the objective function assesses the cost of each solution. For example, the objective of the traveling salesman problem is to find a Hamiltonian circuit among J nodes that minimizes the sum of the weights of the arcs connecting the nodes. A current solution is a Hamiltonian circuit, and a neighborhood can be defined as the set of all Hamiltonian circuits that are produced by pair-wise node exchanges in the current circuit. The objective function is the sum of the arc weights of each circuit.

After a neighborhood and an objective function are defined, local search proceeds as follows: given a current solution, a candidate solution is selected from its neighbors. If the candidate has a lesser objective function value than the current solution, then the candidate becomes the new current solution; otherwise the candidate is rejected. The process is repeated until no neighbor has a lesser objective function value than the current solution. At this point, the algorithm has reached a local minimum with respect to the neighborhood definition, and the algorithm is halted. The principal shortcoming of local search is that the algorithm cannot guarantee that the local minimum is also a global minimum.

Stochastic search algorithms are local search algorithms that probabilistically accept hill climbing solutions (e.g., solutions of higher objective function value than the current solution), in the hope of escaping local optima, so that a global optimum can eventually be reached. For example, *simulated annealing* [7,14], is based on the concept that hill climbing transitions between solutions are probabilistically accepted by comparing a deterministic function (of the increase in solution value and of a control parameter) to a uniform (0,1) random variable. The *noising method* [4], randomly perturbs each solution and then performs local search, using the objective function values of the perturbed data. The process is repeated, while the perturbations are gradually reduced, allowing the original problem structure to reappear. *Threshold accepting* [6] is a local search algorithm, with each candidate solution accepted as the new current solution if the objective function value change is less than a specified threshold (typically defined as a deterministic step function that approaches zero as the algorithm progresses).

Johnson [12] presents a general acceptance probability model, termed *generalized hill climbing* (GHC) algorithms. GHC algorithms include as special cases local search, simulated annealing, the noising method, and threshold accepting. The principal contribution of this paper is a convergence proof that relaxes the sufficient conditions of the most general proof of convergence for stochastic local search algorithms in the literature. Note that classical convergence proofs for stochastic local

```
While stopping criterion not met:
    Set the outer loop counter k = 0
    While iteration k ≠ K :
        Set the inner loop counter m = 0
        While m ≠ M :
            Generate j ∈ η(i) with probability g_{i,j}(k)
            Calculate the objective function value change Δ_{i,j}
            Accept solution j (i⇐j) if R_k(i,j) ≥ Δ_{i,j}
            m ⇐ m+1
        k ⇐ k+1
```

Fig. 1. The generalized hill climbing algorithm.

search algorithms require either that an exponential distribution be used to model the acceptance of candidate solutions along a search trajectory, or that the Markov chain model of the algorithm must be reversible. The proof in this paper removes these limitations, by introducing a new path concept between global and local optima. Convergence is based on the asymptotic behavior of path probabilities between local and global optima.

The paper is organized as follows: in Section 2, notation is explained and the GHC algorithm framework is defined. Section 3 presents the proof of convergence for the GHC algorithm. Section 4 presents some illustrative examples. Section 5 discusses implications and provides concluding comments. An appendix contains all the proofs of the results.

2. Definitions and notation

Define a discrete optimization minimization problem as a two-tuple (Ω, c) where
1. Ω is a finite space composed of (Ω, c) solutions,
2. $c : \Omega \to \mathfrak{R}^+$ is an non-negative objective function.

Define a *neighborhood function* $\eta : \Omega \to 2^\Omega$, which provides connections between the elements of Ω. Define $\Delta_{ij} = c_j - c_i$ to be the change in objective function value between two distinct solutions $i, j \in \Omega$. Define $G \subset \Omega$ to be the set of globally optimal solutions, with objective function value $c_{\text{opt}} = \min_{i \in \Omega} \{c_i\}$. Define $L \subset \Omega \setminus G$ to be the set of locally (but not globally) optimal solutions (i.e., $i \in L$ if $\Delta_{i,j} \geqslant 0$ for all $i \in \Omega \setminus G$ and $j \in \eta(i)$). Finally, define $H = \Omega \setminus (L \cup G)$ to be the set of all other solutions in Ω.

A GHC algorithm is initialized with a solution $\mathrm{i} \in \Omega$ having objective function value c_i. The total number of outer loop iterations K, the total number of inner loop iterations M, the *solution generation probabilities* $g_{i,j}(k)$, non-negative random variables $R_k(i, j)$, and a stopping criterion must all be specified. The GHC algorithm is depicted in pseudocode in Fig. 1.

For $i \in \Omega$, $j \in \eta(i)$, and all k, the one-step transition probability $P_{i,j}(k)$ of accepting the neighboring solution as the new current solution is expressed as

$$P_{i,j}(k) = \begin{cases} g_{i,j}(k)\Pr(R_k(i,j) \geqslant \Delta_{i,j}) & \text{for all } i \in \Omega,\ j \in \eta(i),\ j \neq i, \\ 1 - \displaystyle\sum_{\substack{q \in \eta(i), \\ q \neq i}} P_{i,q}(k) & j = i, \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

Note that $\Pr(R_k(i,j) \geqslant \Delta_{i,j})$ defines the *solution acceptance probability*. Moreover, the generation probabilities $g_{i,j}(k)$ must be non-negative, and satisfy

$$\sum_{j \in \eta(i)} g_{i,j}(k) = 1. \quad (2)$$

The GHC algorithm can be modeled as an inhomogeneous Markov chain, or as a sequence of K homogeneous Markov chains, where each chain is of length M, and each state is a solution in Ω. When certain conditions (as described in Section 3) are placed on the transition matrix associated with each homogeneous Markov chain, then as M approaches infinity, each associated Markov chain approaches its unique equilibrium distribution $\pi(k)$. Additional conditions (also described in Section 3) on $R_k(i,j)$ ensure that as K approaches infinity, the sequence of equilibrium distributions converges to a form where all the probability mass is concentrated on the set of globally optimal solutions.

GHC algorithms traverse the solution space Ω in search of a globally optimal solution. To understand this process, the concept of a *path* between solutions must be defined.

Definition 2.1. A *path* from i to j, depicted as $i \to j$, for all $i,j \in L \cup G$ and all k, is a sequence of solutions $l_0, l_1, \ldots, l_d \in \Omega$ with $l_0 = i$, $l_d = j$, $l_1, l_2, \ldots, l_{d-1} \in H$, and $g_{l_m,l_{m+1}}(k) > 0$ for $m = 0, 1, \ldots, d-1$.

Note that a local or global optimum cannot be an intermediate solution on any path. However, the GHC algorithm can move from i to j via an intermediate solution $l \in L \cup G$, but the trajectory would not be defined as a path. A path can be *equivalent* to some other path, or *distinct*. These concepts are formally defined.

Definition 2.2. A path between solutions $i,j \in L \cup G$ is said to be *equivalent* to another path between i and j, if

(a) all the solutions visited along both paths are identical;

(b) the order in which each solution is visited along both paths is identical.

If a path between solutions $i,j \in L \cup G$ is not equivalent to any other path between i and j, then the path is said to be *distinct*.

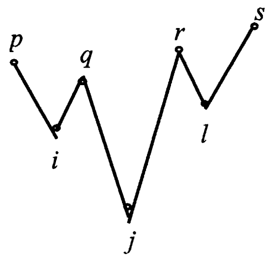

Fig. 2. Illustrative example.

Using these definitions, the probability of transitioning via a path between solutions $i,j \in L \cup G$ can be defined. Suppose that the distinct paths from i to j are labeled $s = 1,2,\ldots,S(i,j)$. Note that $S(i,j)$ can be infinite if one or more solutions in H can be visited infinitely often along a path. Let $P_k(i \overset{s}{\rightarrow} j)$ be the probability of transitioning along the sth (distinct) path between $i,j \in L \cup G$ at iteration k. Therefore, $P_k(i \overset{s}{\rightarrow} j)$ is the product of all one-step transition probabilities between adjacent solutions along the sth path; e.g., for $i,j \in \Omega$, the sth path $i, l_1, \ldots, l_{d-1}, j \in \Omega$ occurs with probability

$$P_k(i \overset{s}{\rightarrow} j) = \prod_{m=0}^{d-1} P_{l_m, l_{m+1}}(k).$$

Note that path distinctness is sufficient for the probability of the union of all distinct paths between $i,j \in \Omega$ to be equal to the sum of the probabilities of the paths [5, p. 3].

Definition 2.3. The *path* probability between any two solutions $i,j \in L \cup G$ is

$$P_k(i \overset{s}{\rightarrow} j) \equiv \begin{cases} \displaystyle\sum_{s=1}^{S(i,j)} P_k(i \rightarrow j), & i \neq j, \\ 1 - \displaystyle\sum_{t \in (L \cup G) \setminus \{i\}} P_k(i \rightarrow j), & i = j, \\ 0, & \text{otherwise}. \end{cases} \tag{3}$$

Note that $P_k(i \rightarrow j)$ can also be viewed as the transition probability from i to j at iteration k. The example depicted in Fig. 2 illustrates how the path probabilities (3) are defined, where $i, l \in L$, $j \in G$, and $p,q,r,s \in H$. The neighborhood structure (indicating all positive one-step transition probabilities) is indicated by the lines connecting the nodes. Therefore, $P_k(i \rightarrow j) > 0$, since i and j are separated only by q. Similarly, $P_k(l \rightarrow j) > 0$, since l and j are separated only by r. However, the only way to reach solution l from solution i is to pass through the global optimum j, and so $P_k(i \rightarrow l) = 0$ from the "otherwise" case of (3).

Recall that the GHC algorithm is composed of an outer loop, indexed on k, and an inner loop, indexed on m. Furthermore, when the sufficient conditions of Theorem 3.1 (see Section 3) are satisfied, then $\pi(k)$ is the equilibrium (long-run) probability vector for all solutions $i \in \Omega$, for each k, as M approaches infinity. Definition 2.4 introduces $\delta(k)$ as the vector of equilibrium probabilities $\delta_i(k)$ for all solutions $i \in L \cup G$. Each probability $\delta_i(k)$ is obtained by scaling the corresponding equilibrium probability $\pi_i(k)$ over the total equilibrium probability $\omega(k)$ of all solutions in $L \cup G$.

Definition 2.4. Let $\omega(k) \equiv \sum_{i \in L \cup G} \pi_i(k)$. Define the equilibrium probability

$$\delta_i(k) \equiv \pi_i(k)/\omega(k) \quad \text{for all } i \in L \cup G \text{ and all } k.$$

Note that $\delta_i(k) \geqslant \pi_i(k)$ for all k, since $\omega(k) \leqslant 1$.

Definitions 2.3 and 2.4 allow the primary convergence proof (Theorem 3.2) to focus only on the sets of local and global optima, which will be shown to be the only solutions of significance for a GHC algorithm. This is done purely for theoretical development. In practice, a GHC algorithm can visit any element in Ω, including elements in H.

3. Proof of convergence

Theorem 3.1 provides the sufficient conditions for a unique equilibrium distribution $\pi(k)$ to exist for each iteration k. Corollary 3.1 shows that the GHC algorithm converges to the set of solutions $L \cup G$ as k approaches infinity. Theorem 3.2 provides the additional sufficient conditions for the GHC algorithm to converge to the set of globally optimal solutions G, as k approaches infinity.

Theorem 3.1. *Let* (Ω, c) *denote an instance of a discrete optimization problem with neighborhood function* η. *Let the GHC transition probabilities* $P_{i,j}(k)$ *be defined by* (1). *Assume that the generation probabilities* $g_{i,j}(k)$ *satisfy*

(a) *for all* $i, j \in \Omega$ *and all iterations* k, *there exists an integer* $d \geqslant 1$ *and a corresponding sequence of solutions* $l_0, l_1, l_2, \ldots, l_d \in \Omega$, *with* $l_0 = i$, $l_d = j$, *and* $g_{l_m, l_{m+1}}(k) > 0$, $m = 0, 1, \ldots, d-1$,

(b) *for all* $i, j \in \Omega$, $j \in \eta(i)$, $\lim_{k\to\infty} g_{i,j}(k)$ *exists and is strictly positive.*

Moreover, assume that the acceptance probabilities satisfy

(c) $\Pr(R_k(i,j) \geqslant \Delta_{i,j}) > 0$ *for all* $i \in \Omega$, $j \in \eta(i)$, *and all iterations* k.

(d) $c_i < c_j \Rightarrow \lim_{k\to\infty} \Pr(R_k(i,j) \geqslant \Delta_{i,j}) = 0$.

Then $\pi(k)$ *exists for each* k.

Proof. See Appendix 1.

Corollary 3.1 show that the equilibrium probability of all non-optimal solutions (i.e., solutions that are neither local nor global optima) approach zero as k approaches infinity.

Corollary 3.1. *If* $\lim_{k\to\infty} \pi(k)$ *exists, then under the conditions and assumptions of Theorem* 3.1,

$$\lim_{k\to\infty} \pi_i(k) = 0 \quad \text{for all } i \in H. \tag{4}$$

Proof. See Appendix 2.

Corollary 3.2 shows that for all local and global optima $i \in L \cup G$, the probabilities $\delta_i(k)$ approach the equilibrium probabilities $\pi_i(k)$ in value, as k approaches infinity.

Corollary 3.2. *If* $\lim_{k\to\infty} \pi(k)$ *exists, then under the conditions and assumptions of Theorem* 3.1, *for all* $i \in L \cup G$,

$$\lim_{k\to\infty} \delta_i(k) = \lim_{k\to\infty} \pi_i(k).$$

Proof. See Appendix 3.

The following definitions of minimal and maximal path probabilities are used to obtain the sufficient conditions presented in Theorem 3.2.

Definition 3.1. The minimum positive path probability between any local (but not global) optimum and any (local or global) optimum at iteration k is

$$P_k(Min\_Path) \equiv \min\{P_k(j \to i) \mid j \in L,\ i \in L \cup G \text{ and } P_k(j \to i) > 0\}.$$

Definition 3.2. The maximum path probability between any global optimum and any local (but not global) optimum at iteration k is

$$P_k(Max\_Path) \equiv \max\{P_k(i \to j) \mid i \in G,\ j \in L\}.$$

Definition 3.3. The maximal product of locally (but not globally) optimal solution equilibrium distribution probabilities and their associated path probabilities to other local (but not global) optima at iteration k is

$$P_k(Max\_Prod) \equiv \max\{\delta_j(k) P_k(j \to q) \mid j, q \in L,\ q \neq j\}.$$

Theorem 3.2 provides sufficient conditions for the equilibrium probability of all local (but not global) optima to approach zero, as k approaches infinity.

Theorem 3.2. *Under the conditions and assumptions of Theorem* 3.1 *and Corollary* 3.2, *if*

(e) $\sum_{k=1}^{\infty} P_k(Min\_Path) = {}^{+}\infty$,

(f) $\sum_{k=1}^{\infty} P_k(Max\_Path) <^{+} \infty$,

(g) $\sum_{k=1}^{\infty} P_k(Max\_Prod) <^{+} \infty$,

then

$$\lim_{k\to\infty} \delta_j(k) = 0 \quad \textit{for all } j \in L. \tag{5}$$

Proof. See Appendix 4.

The assumption that $\lim_{k\to\infty} \pi(k)$ exists is not a severe restriction, and is satisfied by any GHC algorithm (that also meets conditions (a)–(d) whose hill climbing probabilities become monotone nonincreasing as K approaches infinity. This asymptotic monotone behavior is satisfied by the hill climbing acceptance functions typically used in practice. This assumption does preclude, for example, pathological acceptance function formulations that could lead to oscillating $\pi(k)$ vector values for k even versus k odd. Moreover, in practice, it is difficult to find GHC formulations that satisfy both (e) and (f) unless $P_k(Max\_Path) < P_k(Min\_Path)$. Examples 4.1 and 4.3 illustrate two different approaches of formulating algorithms that satisfy this inequality.

Note that Anily and Federgruen [3] provide the most general sufficient conditions in the literature. However, while condition (g) requires that the equilibrium probability distribution $\pi(k)$ be explicitly known only for the set of local and global optima, Anily and Federgruen's [3] convergence theorem requires that the equilibrium distribution $\pi(k)$ be known for *all* solutions in Ω. Hence Theorem 3.2 is a relaxation of the Anily and Federgruen [3] result, in that Theorem 3.2 requires equilibrium distribution information for only a (presumably small) subset of the solution space. Note also that other convergence theorems (e.g. [9]) use either the exponential acceptance function formulation or reversible Markov chain theory to ensure that the solution probability distributions (e.g., each $\pi(k)$) are explicitly known. Therefore, Theorem 3.2 allows the use of acceptance functions that previously were unable to be proven convergent. Section 4 provides several examples.

Corollary 3.3 shows that all probability mass is concentrated on the set of optimal solutions G, as k approaches infinity.

Corollary 3.3. *Under the assumptions of Theorem* 3.2, $\lim_{k\to\infty}(\sum_{i\in G}\pi_i(k)) = 1$.

Proof. See Appendix 5.

4. Illustrative examples

Section 4.1 illustrates how a GHC acceptance function, based on a rational function of k, satisfies the conditions of Theorems 3.1–3.3. Section 4.2 shows that the threshold accepting algorithm does not satisfy the sufficient conditions of Theorem 3.1. Additional GHC acceptance formulations are discussed in Sections 4.3 and 4.4.

4.1. Generalized hill climbing acceptance as a rational function of k

Consider the eight-solution example depicted in Fig. 3, where $G\{p\}$, $L = \{q_1, q_2, q_3\}$, and $H = \{r_1, r_2, r_3, r_4\}$, and the neighborhood structure is shown by the lines connecting the nodes. Let each one-step transition probability be defined (for all $k \geqslant 2$) as in Fig. 4. Note that the rows and columns are arranged in order $p, q_1, q_2, q_3, r_1, r_2, r_3, r_4$. Also, $g_{i,j}(k) \equiv 1/2$ for all $i \in Q$, $j \in \eta(i)$, and zero otherwise, for all k. Note that if any two nodes are not connected by a line, then the two nodes are not neighbors and the generation probability between these nodes is zero. Finally, $R_k(p, r_j) \equiv \Delta_{p,r_j}/(k^2U)$ for $p, r_j \in \Omega$, $j = 1,2,3,4$ and $R_k(q_i, r_j) \equiv \Delta_{q_i,r_j}/(kUj)$ for all $i = 1,2,3$ and $j = 1,2,3,4$,

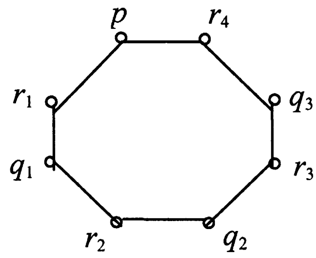

Fig. 3. Solutions and neighbors.

$$
\begin{bmatrix}
1-\frac{1}{k^2} & 0 & 0 & 0 & \frac{1}{2k^2} & 0 & 0 & \frac{1}{2k^2} \\
0 & 1-\frac{3}{4k} & 0 & 0 & \frac{1}{2k} & \frac{1}{4k} & 0 & 0 \\
0 & 0 & 1-\frac{5}{12k} & 0 & 0 & \frac{1}{4k} & \frac{1}{6k} & 0 \\
0 & 0 & 0 & 1-\frac{7}{24k} & 0 & 0 & \frac{1}{6k} & \frac{1}{8k} \\
\frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\
\frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0
\end{bmatrix}
$$

Fig. 4. The one-step transition matrix $\boldsymbol{P}(k)$, defined for $k \geqslant 2$.

where U is a random variable, distributed $U(0,1)$. Then $\Pr(R_k(p,r_j) \geqslant \Delta_{p,r_j}) = 1/k^2$, and $\Pr(R_k(q_i,r_j) \geqslant \Delta_{q_i,r_j}) = 1/(kj)$ for all $i = 1,2,3$ and $j = 1,2,3,4$, and all k. Therefore, all solutions in Ω communicate, and so conditions (a) and (b) of Theorem 3.1 are satisfied. Furthermore, all hill climbing transition probabilities from each solution in $L \cup G$ to its neighbors in H are strictly positive, and decrease monotonically with limit zero as k approaches infinity, hence conditions (c) and (d) are satisfied, and so Theorem 3.1 and Corollary 3.1 apply. The sufficient conditions of Theorem 3.2 are now addressed.

Condition (e) examines the path of minimum positive probability from the set of local optima L to solutions in $L \cup G$. Nine positive path probabilities exist:
(i) $P_k(q_1 \to q_2) = P_{q1,r2}(k)P_{r2,q2}(k) = (1/(4k))(1/2) = 1/(8k)$,
(ii) $P_k(q_1 \to p) = 1/(4k)$,

(iii) $P_k(q_1 \to q_1) = 1 - P_k(q_1 \to q_2) - P_k(q_1 \to p) = 1 - 3/(8k)$,
(iv) $P_k(q_2 \to q_1) = 1/(8k)$,
(v) $P_k(q_2 \to q_3) = 1/(12k)$,
(iv) $P_k(q_2 \to q_2) = 1 - P_k(q_2 \to q_1) - P_k(q_2 \to q_3) = 1 - 5/(24k)$,
(vii) $P_k(q_3 \to q_2) = 1/(12k)$,
(viii) $P_k(q_3 \to p) = 1/(16k)$,
(ix) $P_k(q_3 \to q_3) = 1 - P_k(q_3 \to q_2) - P_k(q_3 \to p) = 1 - 7/(48k)$.

Note that $P_k(q_2 \to p) = P_k(q_1 \to q_3) = P_k(q_3 \to q_1) = 0$, since either path must visit an intermediate solution in $L \cup G$. The probability (viii) is the minimal value of the nine positive path probabilities, and so $P_k(Min\_Path) = P_k(q_3 \to p)$. Hence

$$\sum_{k=2}^{\infty} P_k(q_3 \to p) = \sum_{k=2}^{\infty} 1/(16k) = +\infty,$$

and therefore condition (e) holds.

To address condition (f), the path probabilities from global to local optima must be examined. From the problem symmetry and neighborhood structure, only two positive path probabilities exist, and they are equal. Hence

$$\sum_{k=2}^{\infty} P_k(Max\_Path) = \sum_{k=2}^{\infty} P_k(p \to q_1) = \sum_{k=2}^{\infty} P_k(p \to q_3) = \sum_{k=2}^{\infty} 1/(2k^2)(1/2)$$
$$= \sum_{k=2}^{\infty} 1/(4k^2) < +\infty,$$

and so (f) holds.

Condition (g) requires that the equilibrium probability vector $\delta(k)$ be known for all solutions in L. Hence, solving for the equilibrium probabilities (see [5, p. 154]),

$$\pi_p(k) = \frac{60k^2}{60k^2 + 360k + 215}, \qquad \pi_{q_1}(k) = \frac{78k}{60k^2 + 360k + 215},$$

$$\pi_{q_2}(k) = \frac{114k}{60k^2 + 360k + 215}, \quad \text{and} \quad \pi_{q_3}(k) = \frac{168k}{60k^2 + 360k + 215}.$$

Therefore,

$$\omega(k) = \frac{60k^2 + 360k}{60k^2 + 360k + 215}$$

and so

$$\delta_{q_1}(k) = \frac{13}{10k + 60}, \quad \delta_{q_2}(k) = \frac{19}{10k + 60}, \quad \text{and} \quad \delta_{q_3}(k) = \frac{28}{10k + 60}.$$

The maximal path probability is

$$P_k(Max\_Prod) = \delta_{q_2}(k) P_k(q_2 \to q_1) = \left(\frac{19}{10k + 60}\right)\left(\frac{1}{8k}\right),$$

and hence for condition (g),

$$\sum_{k=1}^{\infty} P_k(Max\_Prod) = \sum_{k=1}^{\infty} \left(\frac{19}{10k+60}\right)\left(\frac{1}{8k}\right) = \sum_{k=1}^{\infty} \left(\frac{19}{80k^2+480k}\right) < \infty,$$

and so the sufficient conditions of Theorems 3.1 and 3.2 are satisfied. Therefore $\lim_{k\to\infty} \delta_p(k) = 1$. Note that

$$\lim_{k\to\infty} \delta_p(k) = \lim_{k\to\infty} \frac{k^2}{k^2+6k} = \lim_{k\to\infty} \pi_p(k) = \lim_{k\to\infty} \frac{60k^2}{60k^2+360k+215} = 1,$$

which further confirms the convergence result.

4.2. GHC formulated as threshold accepting

The threshold accepting (TA) algorithm [6] results from fixing the random variable R_k as a constant for each k. To implement the TA algorithm, define an initial threshold Q_0 such that

$$Q_0 \geqslant \max_{\substack{all\ i\in\Omega \\ j\in\eta(i)}} (c_j - c_i),$$

$$|Q_k| \leqslant Q_0 \quad \text{for all } k,$$

$$\lim_{k\to\infty} Q_k = 0. \tag{6}$$

The initial threshold Q_0 represents the minimum one-step increase in objective function value necessary for the GHC algorithm to be able to transition from any solution i to any neighboring solution j. Then the GHC acceptance probability distribution is

$$\Pr(R_k(i,j) \geqslant \Delta_{i,j}) = \mathrm{PR}(Q_k \geqslant \Delta_{i,j}) = \begin{cases} 1 & \text{if } Q_k \geqslant \Delta_{i,j} \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

Note that no proofs of TA convergence to (4) and (5) are presented in the literature [2]. Furthermore, the TA formulation (6) does not satisfy the sufficient condition (c) Theorem 3.1.

Note that Jacobson and Yücesan [11] provide necessary convergence conditions for GHC algorithms. They use these conditions to establish that threshold accepting will not converge asymptotically to the set of globally optimal solutions if the limit equation in (6) holds.

4.3. GHC formulated as simulated annealing

Johnson and Jacobson [13] show that the GHC solution acceptance probability can be formulated as simulated annealing, by setting $R_k(i,j) \equiv -t_k \ln(U)$ for all $i,j \in \Omega$, $j \in \eta(i)$, and all k, where t_k is a cooling parameter and U is a $U(0,1)$ random variable. Then for hill climbing moves, the acceptance probability is $\Pr(R_k(i,j) \geqslant$

$\Delta_{i,j}) = \exp(-\Delta_{i,j}/t_k)$. (Recall from Fig. 1 that transitions to equal or improving solutions are accepted with probability one). Note that for all $i, j \in \Omega$ and for all $t_k > 0$, $\exp(-\Delta_{i,j}/t_k) > 0$ and so condition (c) in Theorem 3.1 is satisfied. If $\lim_{k\to\infty} t_k = 0$, then $c_i < c_j$ implies that $\lim \exp(-\Delta_{i,j}/t_k) = 0$, hence condition (d) in Theorem 3.1 is satisfied. Therefore if conditions (a) and (b) on the solution generation probabilities are met, then Theorem 3.1 applies. Aarts, Korst, and van Laarhoven [14] show that the simulated annealing stationary distribution for all $i \in \Omega$, and for each outer loop iteration k, is

$$\pi_i(k) = \frac{\exp(-c_i/t_k)}{\sum_{n\in\Omega} \exp(-c_n/t_k)}. \tag{8}$$

To show that the conditions in Theorem 3.2 can hold, first, from conditions (a)–(c), there exists some positive integer b such that $P_k(Min\_Path)$ is bounded below by a single path of length b between the Min_Path solutions $j \in L$, $i \in L \cup G$, such that

$$\begin{aligned} P_k(Min\_Path) &\geqslant P_{j,l_1}(k) P_{l_1,l_2}(k) \ldots P_{l_b,j}(k) \\ &= g_{j,l_1}(k) \exp(-\Delta_{j,l_1}/t_k) g_{l_1,l_2}(k) \exp(-\Delta_{l_1,l_2}/t_k) \ldots g_{l_b,i}(k) \\ &\quad \exp(-\Delta_{l_b,i}/t_k) \\ &= g_{j,l_1}(k) g_{l_1,l_2}(k) \ldots g_{l_b,i}(k) \exp(-\Delta_{j,i}/t_k). \end{aligned}$$

Define d^* as the depth of the deepest local minimum in the set L [9], and set $d^* = \Delta_{j,i}$. Moreover, define the cooling parameter $t_k = d^*/\ln(k+1)$. Finally, set the solution generation probabilities $g_{i,j}/(k)$ to be independent of k for all $i, j \in \Omega$ (i.e., $g_{i,j}(k) = g_{i,j}$ for all k). Then condition (e) is satisfied since

$$\begin{aligned} \sum_{k=1}^{\infty} P_k(Min\_Path) &\geqslant \sum_{k=1}^{\infty} P_{j,l_1}(k) P_{l_1,l_2}(k) \ldots P_{l_b,j}(k) \\ &= \sum_{k=1}^{\infty} g_{j,l_1}(k) g_{l_1,l_2}(k) \ldots g_{l_b,i}(k) \exp(-\Delta_{j,i}/t_k) \\ &= +\infty. \end{aligned}$$

To address condition (f), from condition (a), there exists two positive integers d and s such that $P_k(Max\_Path)$ is bounded above by a single path of length d between the Max_Path solutions $m \in G, n \in L$, such that

$$\begin{aligned} P_k(Max\_Path) &\leqslant s[P_{m,l_1}(k) P_{l_1,l_2}(k) \ldots P_{l_d,n}(k)] \\ &= s[g_{m,l_1}(k) \exp(-\Delta_{m,l_1}/t_k) g_{l_1,l_2}(k) \exp(-\Delta_{l_1,l_2}/t_k) \ldots g_{l_d,n}(k) \\ &\quad \exp(-\Delta_{l_d,n}/t_k)] \\ &= s[g_{m,l_1}(k) g_{l_1,l_2}(k) \ldots g_{l_d,n}(k) \exp(-\Delta_{m,n}/t_k)]. \end{aligned}$$

If $d^* < \Delta_{m,n}$, then

$$\sum_{k=1}^{\infty} P_k(Max\_Path) \leqslant \sum_{k=1}^{\infty} s[P_{m,l_1}(k)P_{l_1,l_2}(k)\ldots P_{l_d,n}(k)]$$
$$= s\sum_{k=1}^{\infty}[g_{m,l_1}(k)g_{l_1,l_2}(k)\ldots g_{l_d,n}(k)\exp(-\Delta_{m,n}/t_k)]$$
$$< +\infty,$$

hence condition (f) is satisfied. Note that while the requirement that $d^* < \Delta_{m,n}$ may seem restrictive, Hajek's necessary and sufficient conditions for simulated annealing require that the depth of an optimal solution be *infinite* [9].

Finally, to show that condition (g) holds, from (8) and Definition 2.4,

$$\delta_i(k) = \frac{\pi_i(k)}{\omega(k)} = \frac{\exp(-c_i/t_k)}{\sum_{n\in\Omega}\exp(-c_n/t_k)} \Big/ \sum_{m\in L\cup G}\frac{\exp(-c_m/t_k)}{\sum_{n\in\Omega}\exp(-c_n/t_k)}$$
$$= \frac{\exp(-c_i/t_k)}{\sum_{m\in L\cup G}\exp(-c_m/t_k)}.$$

From condition (a), there exists two positive integers v and x such that $P_k(Max\_Prod)$ is bounded above by a single path of length v between the Max_Prod solutions $j,q\in L$. This implies that

$$P_k(Max\_Prod) = \delta_j(k)P_k(j\to q)$$
$$\leqslant \frac{\exp(-c_j/t_k)}{\sum_{m\in L\cup G}\exp(-c_m/t_k)}(x)[P_{j,l_1}(k)P_{l_1,l_2}(k)\ldots P_{l_v,q}(k)]$$
$$= \frac{\exp(-c_j/t_k)}{\sum_{m\in L\cup G}\exp(-c_m/t_k)}(x)[g_{j,l_1}(k)\ldots g_{l_v,q}(k)\exp(-\Delta_{j,q}/t_k)]$$
$$= (x)\frac{\exp(-c_q/t_k)}{\sum_{m\in L\cup G}\exp(-c_m/t_k)}[g_{j,l_1}(k)\ldots g_{l_v,q}(k)].$$

Therefore

$$\sum_{k=1}^{\infty} P_k(Max\_Prod) \leqslant \sum_{k=1}^{\infty}(x)\frac{\exp(-c_q/t_k)}{\sum_{m\in L\cup G}\exp(-c_m/t_k)}[g_{j,l_1}(k)\ldots g_{l_v,q}(k)]$$
$$= (x)\sum_{k=1}^{\infty}\frac{1}{\sum_{m\in L\cup G}\frac{\exp(-c_m/t_k)}{\exp(-c_q/t_k)}}[g_{j,l_1}(k)\ldots g_{l_v,q}(k)]$$
$$= (x)\sum_{k=1}^{\infty}\frac{1}{\sum_{m\in L\cup G}\exp(-\Delta_{q,m}/t_k)}[g_{j,l_1}(k)\ldots g_{l_v,q}(k)]$$
$$< +\infty,$$

hence condition (g) is satisfied, and Theorem 3.2 and Corollary 3.3 apply. Finally, note that if $R_k(i,j)\equiv 0$ for all $i,j\in\Omega$, $j\in\eta(i)$, and all k, the resulting algorithm is local search, while if $R_k(i,j) \equiv +\infty$ for all $i,j \in \Omega$ and all k, and $\eta(i) = \Omega$, the resulting algorithm is Monte Carlo search.

4.4. GHC formulated as the noising method

If $R_k(i,j) = wV_k$ for all $i,j \in \Omega$ and all k, where $w = \max\{(c_j - c_i)i,j \in \Omega\}$ and V_k are independent random variables such that $E[V_k] > 0$ for all k, and $V_k \to_P 0$ as k approaches infinity, then the resulting algorithm is the noising method. If the V_k are formulated so that the conditions of Theorem 3.2 hold, then the noising method can be designed to converge to the set of globally optimal solutions. Using a similar analysis as described for simulated annealing, this would require, for example, that the V_k have an unbounded tail (i.e., $P\{V_k > \mathrm{v}\} > 0$ for all k and for all $v \in R$). Note that since the definition of the random variables V_k are problem specific, based on how randomness is added to the different solution components, hence the resulting objective function values, each such problem must be treated on a case by case base to ensure that conditions (a)–(g) hold.

5. Implications and conclusions

This paper presents a proof of convergence for generalized hill climbing (GHC) algorithms. Implications arising from this result are discussed, and examples are presented that illustrate the GHC algorithm convergence conditions.

The principal contribution of Theorems 3.1 and 3.2 is that a large body of convergent stochastic hill climbing algorithms is created, where only simulated annealing existed previously. For example, the noising method will converge if it is defined such that it is asymptotically easier to escape from any local minimum than a global minimum. Conversely, local search and threshold accepting do not meet the sufficient conditions of Theorems 3.1 and 3.2. Though this does not prove that these algorithms do not converge, the necessary convergence conditions in Jacobson and Yücesan [11] establish such nonconvergence results.

Theorems 3.1 and 3.2, and Corollaries 3.1 and 3.2, together prove that under certain conditions the set of globally optimal solutions G must occur with probability one as k approaches infinity. However, the theorems do not show how the probability mass is asymptotically distributed among the global optima. Hence in the limit, some globally optimal solutions may occur with greater probability that other global optima. Note that conditions (e) and (g) together imply that for all solutions $j \in L$, each equilibrium probability $\delta_j(k)$ must approach zero at a minimum rate sufficient for (g) to hold, as $k \to \infty$.

Computational results for various GHC algorithms applied to a discrete manufacturing process design problem are reported by Jacobson et al. [10]. Research is in progress to determine whether the conditions of Theorem 3.2 are also necessary, and whether they can be reformulated or relaxed, in order to make them easier to verify. Research is also in progress to assess whether tailoring the probability distribution associated with the random variable $R_k(i,j)$ can optimize the GHC algorithm's finite-time performance on specific discrete optimization classes.

Appendix 1. Proof of Theorem 3.1.

See Definitions 6 and 7, and Theorem 1 of Aarts et al. [1, p. 100].

Appendix 2. Proof of Corollary 3.1.

The proof shows that the equilibrium probability $\lim_{k\to\infty} \pi_h(k)$ is zero (given that it exists), for the hth solution by establishing a contradiction based on an inductive argument.

Let H_B be all solutions in H that fail to have zero probability in the limit. Order the solutions in H_B such that for all $i,\ j \in H_B,\ i < j$ implies $c_i < c_j$. Without loss of generality, assume that the objective function value of each solution in H is unique. (Else, each solution's value could be perturbed by some epsilon amount.) So solution number one is the solution in H_B with the smallest objective function value, solution two the next smallest, and so on. Let $h \equiv card(H_B)$, so solution h has the largest objective function value. We will proceed to eliminate $i = h,\ h-1, \ldots, 1$ in that order.

First using the law of total probability [5, p. 15] and conditioning on all solutions $j \in \Omega$,

$$\pi_i(k) = \sum_{j\in\Omega} \pi_j(k)P_{j,i}(k) \quad \text{for all } i \in \Omega \text{ and all } k. \tag{9}$$

For any solution $i \in H$, the equilibrium probability $\pi_i(k)$ is expressed in terms of (9) as

$$\pi_i(k) = \sum_{j\in G} \pi_j(k)P_{j,i}(k) + \sum_{j\in L} \pi_j(k)P_{j,i}(k) + \sum_{j\in H} \pi_j(k)P_{j,i}(k).$$

Collect all the $\pi_i(k)$ terms on the left-hand side to obtain

$$\pi_i(k)(1 - P_{i,i}(k)) = \sum_{j\in G} \pi_j(k)P_{j,i}(k) + \sum_{j\in L} \pi_j(k)P_{j,i}(k) + \sum_{\substack{j\in H\\ j\neq i}} \pi_j(k)P_{j,i}(k). \tag{10}$$

Note that

$$(1 - P_{i,i}(k)) = \sum_{\substack{j\in\Omega,\\ i\neq j}} P_{i,j}(k) = \sum_{\substack{j\in\Omega,\\ i\neq j,\\ c_i \geqslant c_j}} P_{i,j}(k) + \sum_{\substack{j\in\Omega,\\ c_i < c_j}} P_{i,j}(k).$$

(e.g., $(1 - P_{i,i}(k))$ is the probability that the process does not remain at solution $i \in H$ in the next transition). Hence, (10) can be expressed as

$$\pi_i(k)\left(\sum_{\substack{j\in\Omega,\\ i\neq j,\\ c_i\geqslant c_j}} P_{i,j}(k) + \sum_{\substack{j\in\Omega,\\ c_i<c_j}} P_{i,j}(k)\right)$$

$$= \sum_{j\in G}\pi_j(k)P_{j,i}(k) + \sum_{j\in L}\pi_j(k)P_{j,i}(k) + \sum_{\substack{j\in H\\ j\neq i}}\pi_i(k)P_{j,i}(k).$$

Rewrite the right-hand side in terms of hill climbing transitions for elements in set H to obtain

$$\pi_i(k)\left(\sum_{\substack{j\in\Omega,\\ i\neq j,\\ c_i\geqslant c_j}} P_{i,j}(k) + \sum_{\substack{j\in\Omega,\\ c_i<c_j}} P_{i,j}(k)\right)$$

$$= \sum_{j\in G}\pi_j(k)P_{j,i}(k) + \sum_{j\in L}\pi_j(k)P_{j,i}(k) + \sum_{\substack{j\in H,\\ j\neq i,\\ c_j\geqslant c_i}}\pi_j(k)P_{j,i}(k)$$

$$+ \sum_{\substack{j\in H,\\ c_j<c_i}}\pi_j(k)P_{j,i}(k).$$

Note that since there are only a finite number of summands, each of which is non-negative, then the limit of the sums is equal to the sum of the limits [15, p. 37]. Therefore,

$$\lim_{k\to\infty}\left(\pi_i(k)\sum_{\substack{j\in\Omega,\\ i\neq j,\\ c_i\geqslant c_j}} P_{i,j}(k)\right) + \lim_{k\to\infty}\left(\pi_i(k)\sum_{\substack{j\in\Omega,\\ c_i<c_j}} P_{i,j}(k)\right)$$

$$= \lim_{k\to\infty}\left(\sum_{j\in G}\pi_j(k)P_{j,i}(k)\right) + \lim_{k\to\infty}\left(\sum_{j\in L}\pi_j(k)P_{j,i}(k)\right)$$

$$+\lim_{k\to\infty}\left(\sum_{\substack{j\in H,\\ j\neq i,\\ c_j\geqslant c_i}}\pi_j(k)P_{j,i}(k)\right)+\lim_{k\to\infty}\left(\sum_{\substack{j\in H,\\ c_j<c_i}}\pi_j(k)P_{j,i}(k)\right). \tag{11}$$

From (1) and condition (d) in Theorem 3.1, all one-step hill climbing transition probabilities approach zero in the limit. Therefore,

$$\lim_{k\to\infty}\left(\pi_i(k)\sum_{\substack{j\in\Omega,\\ c_i<c_j}}P_{i,j}(k)\right)=\lim_{k\to\infty}\left(\sum_{j\in G}\pi_j(k)P_{j,i}(k)\right)$$
$$=\lim_{k\to\infty}\left(\sum_{j\in L}\pi_j(k)P_{j,i}(k)\right)$$
$$=\lim_{k\to\infty}\left(\sum_{\substack{j\in H,\\ c_j<c_i}}\pi_j(k)P_{j,i}(k)\right)=0,$$

since each of the summands include only hill climbing transitions. Therefore, (11) simplifies to

$$\lim_{k\to\infty}\left(\pi_i(k)\sum_{\substack{j\in\Omega,\\ i\neq j,\\ c_i\geqslant c_j}}P_{i,j}(k)\right)=\lim_{k\to\infty}\left(\sum_{\substack{j\in H,\\ j\neq i,\\ c_j\geqslant c_i}}\pi_j(k)P_{j,i}(k)\right)\quad\text{for all } i\in H. \tag{12}$$

We now proceed with the backward induction. Consider (12) for the particular solution $i=h\in H_B$ (recall that h is the solution of maximal objective function value, e.g., $c_h>c_j$ for all $j\in H_B\setminus\{h\}$, since all solutions are arranged in order of increasing objective function value). The right-hand side of (12) is zero because there are no solutions $j\in H$ of cost greater than that of solution h with nonzero limit. Recall we assume that

$$\lim_{k\to\infty}\pi_i(k)=\varepsilon_i>0\quad\text{for all } i\in H_B.$$

Furthermore, since h is by definition not a local minimum, then there must exist at least one solution $l\in\Omega$, $l\in\eta(h)$, such that $c_h\geqslant c_l$. Therefore, since $R_k(i,j)\geqslant 0$ for all $i,j \in \Omega$ and from (1) and condition (b), the left-hand side of (12) is

$$\begin{aligned}\lim_{k\to\infty}\left(\pi_h(k)\sum_{\substack{j\in\Omega,\\ h\neq j,\\ c_h\geqslant c_j}} P_{h,j}(k)\right) &= \varepsilon_h \lim_{k\to\infty}\sum_{\substack{j\in\Omega,\\ j\in\eta(h),\\ c_h\geqslant c_j}} g_{h,j}(k)\Pr(R_k(h,j)\geqslant \Delta_{h,j})\\ &= \varepsilon_h \lim_{k\to\infty}\sum_{\substack{j\in\Omega,\\ j\in\eta(h),\\ c_h\geqslant c_j}} g_{h,j}(k)(1)\\ &\geqslant \varepsilon_h \lim_{k\to\infty} g_{h,l}(k)\\ &> 0,\end{aligned}$$

which is a contradiction, since the right-hand side of (12) has limit zero. Thus

$$\lim_{k\to\infty}\pi_h(k)=0.$$

Proceed with the induction, by considering (11) for solutions $h-1, h-2,\ldots,1$. □

Appendix 3. Proof of Corollary 3.2.

Recall that the equilibrium probabilities of all solutions in Ω must sum to one. Taking the limit of this sum as k approaches infinity leads to

$$\begin{aligned}1 &= \lim_{k\to\infty}\left(\sum_{i\in H}\pi_i(k)+\sum_{i\in L\cup G}\pi_i(k)\right)\\ &= \sum_{i\in H}\lim_{k\to\infty}\pi_i(k)+\sum_{i\in L\cup G}\lim_{k\to\infty}\pi_i(k)\\ &= \lim_{k\to\infty}\sum_{i\in L\cup G}\pi_i(k)=\lim_{k\to\infty}\omega(k).\end{aligned}$$

Since $\lim_{k\to\infty}\omega(k)=1$ and $0\leqslant \lim_{k\to\infty}\pi_i(k)\leqslant 1$ for all $i\in\Omega$, hence the limit of the quotient $\pi_i(k)/\omega(k)$ is equal to the quotient of the limits [15:39], and so

$$\lim_{k\to\infty}\delta_i(k)=\lim_{k\to\infty}(\pi_i(k)/\omega(k))=\lim_{k\to\infty}\pi_i(k),\quad \text{for all } i\in L\cup G. \qquad \square$$

Appendix 4. Proof of Theorem 3.2.

(by contradiction): First, each element of $\delta(k)$ is expressed using the law of total probability, and the path probabilities (3). Hence for each iteration k,

$$\delta_j(k)=\sum_{i\in G}\delta_i(k)P_k(i\to j)+\sum_{i\in L}\delta_i(k)P_k(i\to j)\quad \text{for all } j\in L\cup G.$$

Next, assume there exists some $j \in L$ and an iteration k_1 such that for all $k \geqslant k_1$, $\delta_j(k) \geqslant \varepsilon > 0$. Summing over all iterations k leads to

$$\sum_{k=1}^{\infty} \delta_j(k) = \sum_{k=1}^{\infty} \sum_{i \in G} \delta_i(k) P_k(i \to j) + \sum_{k=1}^{\infty} \sum_{i \in L} \delta_i(k) P_k(i \to j).$$

Since Ω is finite and all summands are nonnegative, then the order of the summations can be interchanged, resulting in

$$\sum_{k=1}^{\infty} \delta_j(k) = \sum_{i \in G} \sum_{k=1}^{\infty} \delta_i(k) P_k(i \to j) + \sum_{i \in L} \sum_{k=1}^{\infty} \delta_i(k) P_k(i \to j).$$

Collecting $\delta_j(k)$ terms on the left-hand side leads to

$$\sum_{k=1}^{\infty} \delta_j(k)(1 - P_k(j \to j)) = \sum_{i \in G} \sum_{k=1}^{\infty} \delta_i(k) P_k(i \to j) + \sum_{\substack{i \in L \\ i \neq j}} \sum_{k=1}^{\infty} \delta_i(k) P_k(i \to j). \tag{13}$$

Note that $(1 - P_k(j \to j))$ is the probability that, given the process is in solution $j \in L$, the process transitions to any solution $i \in L \cup G$ except solution j. Note that since all solutions communicate (from Theorem 1), a path exists such that j can reach some $q \in L \cup G$. Therefore, two cases are possible.

Case 1: Suppose the process transitions to a particular global optimum $q \in G$. Hence,

$$P_k(j \to q) \leqslant (1 - P_k(j \to j)),$$

and so (13) becomes

$$\sum_{k=1}^{\infty} \delta_j(k) P_k(j \to q) \leqslant \sum_{i \in G} \sum_{k=1}^{\infty} \delta_i(k) P_k(i \to j) + \sum_{\substack{i \in L \\ i \neq j}} \sum_{k=1}^{\infty} \delta_i(k) P_k(i \to j). \tag{14}$$

Since $\delta_i(k) \leqslant 1$ for all $i \in \Omega$ and all k, then (14) can be rewritten as

$$\sum_{k=1}^{k_1} \delta_j(k) P_k(j \to q) + \sum_{k=k_1+1}^{\infty} \varepsilon P_k(j \to q)$$
$$\leqslant \sum_{i \in G} \sum_{k=1}^{\infty} P_k(i \to j) + \sum_{\substack{i \in L \\ i \neq j}} \sum_{k=1}^{\infty} \delta_i(k) P_k(i \to j). \tag{15}$$

For the left-hand side of (15), condition (e) leads to

$$+\infty = \sum_{k=k_1+1}^{\infty} \varepsilon P_k(Min\_Path) \leqslant \sum_{k=k_1+1}^{\infty} \varepsilon P_k(j \to q). \tag{16}$$

Now consider the right-hand side of (15). Since Ω is finite, then conditions (g) and (h) lead to

$$\sum_{i\in G}\sum_{k=1}^{\infty}P_k(i\to j)+\sum_{\substack{i\in L\\ i\neq j}}\sum_{k=1}^{\infty}\delta_i(k)P_k(i\to j)$$

$$\leqslant \sum_{i\in G}\sum_{k=1}^{\infty}P_k(Max\_Path)+\sum_{\substack{i\in L,\\ i\neq j}}\sum_{k=1}^{\infty}P_k(Max\_Prod)<\infty,$$

which contradicts (16). Therefore there cannot exist any iteration k_1 such that (15) holds, and so condition (e) implies that

$$\lim_{k\to\infty}\delta_j(k)=0.$$

Case 2: Suppose the process transitions to a particular local optimum $q\in L$. Then using the same argument as Case 1,

$$\lim_{k\to\infty}\delta_j(k)=0.$$

Therefore,

$$\lim_{k\to\infty}\delta_j(k)=0 \quad \text{for all } j\in L.$$

Appendix 5. Proof of Corollary 3.3.

Express the sum of the equilibrium solution probabilities in terms of sets H, L, and G to obtain

$$\sum_{i\in\Omega}\pi_i(k)=\sum_{i\in H}\pi_i(k)+\sum_{i\in L}\pi_i(k)+\sum_{i\in G}\pi_i(k)=1.$$

Since the limit (as k approaches infinity) exists for each equilibrium probability and the solution space is finite, then the limit of the sum is equal to the sum of the limits [15, 37]. Therefore,

$$\lim_{k\to+\infty}\sum_{i\in\Omega}\pi_i(k)=\lim_{k\to+\infty}\sum_{i\in H}\pi_i(k)+\lim_{k\to+\infty}\sum_{i\in L}\pi_i(k)+\lim_{k\to+\infty}\sum_{i\in G}\pi_i(k)=1.$$

Theorems 3.1, 3.2, and Corollaries 3.1 and 3.2 lead to

$$\sum_{i\in H}\lim_{k\to+\infty}\pi_i(k)=\sum_{i\in L}\lim_{k\to+\infty}\pi_i(k)=0,$$

which implies

$$\sum_{i\in G}\lim_{k\to+\infty}\pi_i(k)=\lim_{k\to+\infty}\sum_{i\in G}\pi_i(k)=1 \qquad \square$$

References

[1] E.H.L. Aarts, J.H.M. Korst, P.J.M. van Laarhoven, in: E. Aarts, J.K. Lenstra (Eds.), Local Search in Combinatorial Optimization, John Wiley and Sons Ltd., London, England, 1997.

[2] I. Althofer, K.U. Koschnick, On the convergence of threshold accepting, Appl. Math. Optim. 24 (1991) 183–195.

[3] S. Anily, A. Federgruen, Simulated annealing methods with general acceptance probabilities, J. Appl. Probab. 24 (1987) 657–667.

[4] I. Charon, O. Hudry, The noising method: a new method for combinatorial optimization, Oper. Res. Lett. 14 (1993) 133–137.

[5] E. Cinlar, Introduction to stochastic processes, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1975.

[6] G. Dueck, T. Scheuer, Threshold accepting: a general purpose optimization algorithm appearing superior to simulated annealing, J. Comput. Phys. 90 (1990) 161–175.

[7] R.W. Eglese, Simulated annealing: a tool for operational research, European J. Oper. Res. 46 (1990) 271–281.

[8] M.R. Garey, D.S. Johnson, Computers and Intractability: A Guide to the Theory of NP-Completeness, W.H. Freeman and Company, New York, NY, 1979.

[9] B. Hajek, Cooling schedules for optimal annealing, Math. Oper. Res. 13 (2) (1988) 311–329.

[10] S.H. Jacobson, K.A. Sullivan, A.W. Johnson, Generalized hill climbing algorithms for discrete manufacturing process design problems using computer simulation models, Eng. Optim. 31 (1998) 247–260.

[11] S.H. Jacobson, E. Yücesan, On the effectiveness of generalized hill climbing algorithms, Technical Report, Department of Mechanical and Industrial Engineering, University of Illinois, Urbana, IL, 2000.

[12] A.W. Johnson, Generalized hill climbing algorithms for discrete optimization problems, Doctoral Dissertation, Department of Industrial and Systems Engineering, Virginia Polytechnic Institute and State University, Blacksburg VA, 1996.

[13] A.W. Johnson, S.H. Jacobson, A class of convergent generalized hill climbing algorithms, Appl. Math. Comput. 2000, to appear.

[14] P.J.M. van Laarhoven, E.H.L. Aarts, Simulated Annealing: Theory and Applications, D. Reidel Publishing Company, Dordrecht, Holland, 1987.

[15] M.H. Protter, C.B. Morrey, A first course in real analysis, 2nd Edition, Springer Verlag, New York, NY, 1991.

[16] A. Sangiovanni-Vincentelli, Editor's foreward, Algorithmica 6 (3) (1991) 295–301.

[17] C.A. Tovey, in: E. Aarts, J.K. Lenstra (Eds.), Local search in combinatorial optimization, Wiley, London, England, 1997.