REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** 9 Jan 97	**3. REPORT TYPE AND DATES COVERED**

4. TITLE AND SUBTITLE Generalized Hill Climbing Algorithms For Discreter Optimization Problems	**5. FUNDING NUMBERS**
6. AUTHOR(S) Alan W. Johnson	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Virginia Polytechnic Institute and State University	**8. PERFORMING ORGANIZATION REPORT NUMBER** 96-45D
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) DEPARTMENT OF THE AIR FORCE AFIT/CIA 2950 P STREET WPAFB OH 45433-7765	**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION AVAILABILITY STATEMENT Unlimited	**12b. DISTRIBUTION CODE**

13. ABSTRACT *(Maximum 200 words)*

14. SUBJECT TERMS	**15. NUMBER OF PAGES** 128
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION OF ABSTRACT**	**20. LIMITATION OF ABSTRACT**

GENERALIZED HILL CLIMBING ALGORITHMS FOR DISCRETE OPTIMIZATION PROBLEMS

by

Alan W Johnson

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

INDUSTRIAL AND SYSTEMS ENGINEERING

APPROVED:

S.H. Jacobson, Chair

Donald Allison

D. C. S. Allison

Ben S Blanchard

B. S. Blanchard

Subhash C Sarin

S. C. Sarin

H. D. Sherali

October 1, 1996

Blacksburg, Virginia

Key Words: combinatorial optimization, discrete optimization, convergence, heuristics, hill climbing algorithms, simulated annealing

GENERALIZED HILL CLIMBING ALGORITHMS FOR DISCRETE OPTIMIZATION PROBLEMS

by

Alan W Johnson

Sheldon H. Jacobson, Chair

Industrial and Systems Engineering

(ABSTRACT)

Generalized hill climbing (GHC) algorithms are introduced, as a tool to address difficult discrete optimization problems. Particular formulations of GHC algorithms include simulated annealing (SA), local search, and threshold accepting (TA), among others. A proof of convergence of GHC algorithms is presented, that relaxes the sufficient conditions for the most general proof of convergence for stochastic search algorithms in the literature (Anily and Federgruen [1987]).

Proofs of convergence for SA are based on the concept that deteriorating (hill climbing) transitions between neighboring solutions are accepted by comparing a deterministic function of both the solution change cost and a temperature parameter to a uniform (0,1) random variable. GHC algorithms represent a more general model, whereby deteriorating moves are accepted according to a general random variable.

Computational results are reported that illustrate relationships that exist between the GHC algorithm's finite-time performance on three problems, and the general random variable formulations used. The dissertation concludes with suggestions for further research.

CHAPTER 1: INTRODUCTION

1.1 Motivation

Many discrete optimization (minimization) problems belong to a class of problems that are difficult to solve, i.e., the class of NP-complete problems (Garey and Johnson [1979, pg 13]). There is no known polynomial-time algorithm that can solve any problem in this class. One classical example is the traveling salesman problem (Garey and Johnson [1979, pg 4]). Given a list of J nodes, the problem is to find a Hamiltonian circuit of minimum cost. (Note that a solution is an ordering of the J nodes, with total cost equal to the sum of the costs of the corresponding $(J+1)$ arcs connecting the nodes. The total number of solutions is of order $(J-1)!$.)

Since the NP-complete class of problems contains many examples of practical interest, heuristic methods have been developed that efficiently find near-optimal solutions. Sangiovanni-Vincentelli [1991] separates heuristic 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. Constructive methods tend to exploit specific features of the problem to be solved and are therefore difficult to generalize, while iterative methods are more flexible. Iterative methods all share the same basic structure: starting from an initial solution, a sequence of solutions are generated until a termination criterion is satisfied. Iterative algorithms are specified by the rules for generating and accepting new solutions, and by termination criteria. *Greedy* iterative algorithms select only those solutions whose costs are less than or equal to the cost of the incumbent solution, and therefore generally become trapped in local optima. One particular greedy iterative algorithm is local search (Papadimitriou and Steiglitz [1982]). *Stochastic hill climbing* algorithms have the ability to probabilistically accept candidate solutions with higher cost than that of the incumbent solution, in an effort to escape local optima.

1.2 Research Goals

This research explores how generalizing the solution acceptance function model of stochastic hill climbing algorithms can improve their performance on hard discrete optimization problems. A frequently used stochastic hill climbing algorithm for discrete optimization is simulated annealing (SA) (Eglese [1990]). SA exploits the analogy of discrete optimization to the physical annealing of crystalline solids, in which a solid is cooled very slowly from some elevated temperature and thereby allowed to relax toward its low energy states. The appeal of SA derives from its guarantee of asymptotic convergence to a global extremum.

A key feature of stochastic hill climbing algorithms is their potential to escape local optima. For example, proofs of convergence of SA are based on the concept that deteriorating (hill climbing) transitions between solutions are probabilistically accepted by comparing a deterministic function of both the solution change cost and a temperature parameter to a uniform $(0,1)$ random variable. This research examines a more general acceptance probability model, titled *generalized hill climbing* (GHC), where deteriorating moves are accepted according to a general random variable.

One limitation of SA is that traditional SA convergence theory fixes the random variable as an exponential function. Anily and Federgruen [1987] present an SA convergence theory that addresses more general acceptance probability functions, but their theory requires a restrictive sufficient condition that is very difficult to verify; furthermore, they do not provide computational results to address whether different acceptance probability functions would affect SA's finite-time performance (in terms of solution quality versus algorithm execution time). In essence, no unifying convergence theory has been developed that provides sufficient conditions on any acceptance probability distribution function, for stochastic hill climbing algorithms to achieve asymptotic convergence to a global extremum.

Another limitation of SA is that the convergence behavior is asymptotic. Thus global optimality is obtained only after an infinite number of algorithm iterations. Although the asymptotic behavior of SA has been extensively studied, it is the finite-time behavior that interests practitioners (Romeo and Sangiovanni-Vincentelli [1991], Strenski and Kirkpatrick [1991], and Tovey [1988]).

The contributions from this research focus on two areas: first, a new method of proving convergence of stochastic hill climbing algorithms is presented, that relaxes the sufficient conditions found in the literature. This result creates a large body of convergent stochastic hill climbing algorithms where only SA existed previously. Second, tests of the performance of selected probability distributions for the general random variable on specific problems are conducted. These tests empirically study which probability distributions enhance the GHC algorithm's finite-time performance (in terms of solution quality versus algorithm execution time) on these problems.

1.3 The Generalized Hill Climbing Algorithm

Define a discrete optimization minimization problem as a three-tuple $(\Omega, \mathcal{N}, c)$ where:

1. Ω is a finite solution space composed of $card(\Omega)$ elements, where $card(\Omega)$ is the cardinality of Ω,
2. $\mathcal{N}: \Omega \rightarrow 2^{card(\Omega)}$ is a neighborhood function of Ω,
3. $c: \Omega \rightarrow \Re^{+}$ is a non-negative objective function.

Generalized Hill Climbing (GHC) algorithms (depicted in pseudo-code in Figure 1.1) are initialized with solution $i \in \Omega$, having objective function value c_i. The total number of

```
Initialization: specify $(\Omega, \mathcal{N}, c)$, and select initial solution $i \in \Omega$
*While* stopping criterion not met:
    Set the outer loop counter $k = 0$
    *While* iteration $k \neq K$:
        Set the inner loop counter $m = 0$
        *While* $m \neq M$:
                Generate $j \in \mathcal{N}(i)$ according to probability $g_{i,j}(k)$
                Calculate the change in objective function value $\Delta_{i,j} = c_j - c_i$
                accept solution $j$ $(i \Leftarrow j)$ if $R_k(i,j) \geq \Delta_{i,j}$
                $m \Leftarrow m + 1$
        $k \Leftarrow k + 1$
```

Figure 1.1. The Generalized Hill Climbing algorithm.

outer loop iterations *K*; the total number of inner loop iterations *M*; a nonnegative random variable $R_k(i,j)$ such that $i,j \in \Omega, j \in \mathcal{N}(i)$, and $k \in K$; and a stopping criterion must all be specified.

1.4 Research Questions

Two research questions are investigated.

1) Asymptotic Optimality: *Given (at iteration k) a solution* $i \in \Omega$ *and a neighbor* $j \in \mathcal{N}(i)$, *where solution j is generated with probability* $g_{i,j}(k)$, *and the transition from i to j occurs if*

$$R_k(i,j) \geq \Delta_{i,j}$$

what are sufficient conditions on $R_k(i,j)$ *such that*

$$\lim_{k \to \infty} \Pr(\textit{solution } j \in \{\textit{set of globally optimal solutions of } \Omega\}) = 1 \text{ ?} \qquad (1.1)$$

2) Finite-Time Performance: *Is there a connection between selected probability distribution functions for* $R_k(i,j)$ *and the finite-time performance (in terms of solution quality versus algorithm execution time) of the GHC algorithm on a specified set of problems?* (1.2)

CHAPTER 2: LITERATURE REVIEW

Chapter 2 reviews the probabilistic hill climbing algorithm literature. Section 2.1 addresses convergence properties and finite-time behavior of the simulated annealing algorithm. Section 2.2 discusses the threshold accepting algorithm, while Section 2.3 addresses probabilistic tabu search, the noising method, and genetic algorithms.

2.1 Simulated Annealing

2.1.1 Overview

Simulated annealing (SA) is a local search algorithm with the capability to escape from local optima. It is often used to solve nonconvex discrete optimization problems. Four recent survey articles that provide a good overview of SA's theoretical development and application are Eglese [1990], Fleischer [1995], Koulamas et al. [1994], and Romeo and Sangiovanni-Vincentelli [1991]. Aarts and Korst [1989] and Laarhoven and Aarts [1987] devote entire books to the subject. Fox [1995] shows that selected methods of improving the finite-time performance of SA do not detract from its asymptotic convergence properties. The SA algorithm is depicted in Figure 2.1.

This section reviews the basic SA algorithm, its convergence properties, and its finite-time behavior. Note that although a substantial literature exists for the application of SA to problems with continuous variables, this review focuses only on discrete problems.

```
Initialization: specify $(\Omega, \mathcal{N}, c)$, a temperature parameter $t_k$, $k = 0,1,\ldots$,
and select an initial solution $i \in \Omega$
*While* stopping criterion not met:
    Set the outer loop counter $k = 0$
    *While* iteration $k \neq K$:
        Set the inner loop counter $m = 0$
        *While* $m \neq M$:
            Generate $j \in \mathcal{N}(i)$ according to probability $g_{i,j}(k)$
            Calculate the change in objective function value $\Delta_{i,j} = c_j - c_i$
            If $\Delta_{i,j} \leq 0$, then accept solution $j$ $(i \Leftarrow j)$
            Else, generate a random number $U \sim U(0,1)$
                If $U \leq \exp\left(-\Delta_{i,j} / t_k\right)$, then $(i \Leftarrow j)$
                Else, $(i \Leftarrow i)$
            $m \Leftarrow m + 1$
        $k \Leftarrow k + 1$
```

Figure 2.1. The Simulated Annealing algorithm.

SA is so named because of its analogy to physical annealing with solids, in which a crystalline solid is heated and then allowed to cool very slowly until it achieves its most regular possible crystal lattice configuration (i.e., its minimum lattice energy state), and thus is free of crystal defects. SA establishes the connection between this type of thermodynamic behavior and the search for the global minimum of an objective function in a discrete optimization problem; and further, it provides an algorithmic means for exploiting the connection. SA is based on the Metropolis acceptance criterion (Metropolis et al. [1953]) which models how a thermodynamic system moves from its current solution (state) to a candidate solution, in which the energy content is being minimized. The probability of making such a move is

$$\Pr\{\text{Accept candidate } j \text{ as next solution}\} = \begin{cases} \exp(-\Delta_{i,j} / t_k) & \Delta_{i,j} > 0 \\ 1 & \Delta_{i,j} \leq 0 \end{cases} \tag{2.1}$$

where Ω is a finite solution space, $i, j \in \Omega$ are the current and candidate solutions of the system, respectively, and t_k is the temperature parameter at (outer loop) iteration k, such that

$$t_k > 0 \quad \text{for all } k \text{ and } \lim_{k \to \infty} t_k = 0. \tag{2.2}$$

Let $\Delta_{i,j} \equiv c_j - c_i$, where c_i and c_j denote the energies associated with solutions i and j, respectively. The candidate solution j is chosen at random from among the set of *neighbors* of solution i, defined by $\mathcal{N}(i)$, and becomes the current solution, based on the acceptance probability in (2.1). This acceptance probability is the basic element of the search mechanism in SA. If the probability of generating a candidate solution j from the neighbors of solution i is $g_{i,j}(k)$, where

$$\sum_{j \in \mathcal{N}(i)} g_{i,j}(k) = 1, \quad \textit{for all } i \in \Omega,\ k = 1,2,... \tag{2.3}$$

then a nonnegative square stochastic matrix $P(k)$ can be defined with transition probabilities

$$P_{i,j}(k) = \begin{cases} g_{i,j}(k)\exp(-\Delta_{i,j} / t_k) & j \in \mathcal{N}(i),\, j \neq i \\ 0 & j \notin \mathcal{N}(i),\, j \neq i \\ 1 - \sum\limits_{\substack{l \in \mathcal{N}(i) \\ l \neq i}} P_{i,l}(k) & j = i \end{cases} \tag{2.4}$$

for all solutions $i \in \Omega$, and all $k = 1,2,\ldots$. These transition probabilities define an inhomogeneous Markov chain (Romeo and Sangiovanni-Vincentelli [1991]). Note that boldface type indicates matrix/vector notation, and all vectors are row vectors, unless otherwise indicated.

The key characteristic of SA is that it provides a means of escaping from local optima by allowing *hill climbing* moves (i.e., moves which may worsen the objective function value). As the temperature parameter t_k is decreased to zero, hill climbing moves occur less frequently, and the solution distribution associated with the inhomogeneous Markov chain converges to a form in which all the probability is concentrated on the set of globally optimal solutions.

2.1.2 Homogeneous Markov Chain Theory

Convergence proofs are grouped into two approaches: *homogeneous* and *inhomogeneous*. The homogeneous Markov chain approach (Aarts and Laarhoven [1985], Faigle and Kern [1991], Granville et al. [1994], Lundy and Mees [1986], Mitra et al. [1986], and Rossier et al. [1986]) assumes that each temperature t_k is held constant for a sufficient number of iterations *m* for the stochastic matrix *P*(*k*) to approach its stationary probability distribution $\pi(k)$. (Note that in the interest of simplifying notation as much as possible, the inner loop index *m* is suppressed. However, the reader should interpret the index *k* as the double index *k,m*, where a sequence of $m = 1,2,\ldots,M$ SA iterations occur for each fixed *k*.) The existence of stationary distributions for each temperature is based on the following theorem. (Note: to ensure that Theorem 2.1 is consistent with the SA algorithm depicted in Figure 2.1, without loss of generality, let t_k be a function only of each outer loop iteration *k*, and let the respective number of inner loop iterations *M* and outer loop iterations *K* each approach infinity).

Theorem 2.1: *Let $P_{i,j}(k)$ be the probability of going from solution i to solution j in one step at iteration k, and let $P_{i,j}^{(m)}(k)$ be the probability of going from solution i to solution j in m steps. If a Markov chain is irreducible and aperiodic with finitely many solutions, then $\lim_{m\to\infty} P_{i,j}^{(m)}(k) = \pi_j(k)$ exists for all $i, j \in \Omega$ and iterations k . Furthermore, $\pi_j(k)$ is the unique strictly positive solution of*

$$\pi_j(k) = \sum_{i\in\Omega} \pi_i(k) P_{i,j}(k) \text{, for all } j \in \Omega, \tag{2.5}$$

and

$$\sum_{j\in\Omega} \pi_j(k) = 1. \tag{2.6}$$

Proof: Define Cinlar's [1975, pg 153] result as a function of each iteration *k*, and the result follows. ■

Key requirements for the existence of stationary distributions and for convergence of the sequence of $\pi(k)$ vectors include:

a) transition matrix *irreducibility* (for every finite iteration *k*, the transition matrix assigns a path of nonzero probability between any two solutions $i, j \in \Omega$),

b) *aperiodicity* (starting at solution $j \in \Omega$, it is possible to return to *j* in any number of inner-loop iterations *m*),

c) a stationary transition matrix probability distribution (which, when *k* goes to infinity, is nonzero only at globally optimal solutions).

Note that all SA proofs of convergence in the literature that are based on homogeneous Markov chain theory, either explicitly or implicitly require the sufficient condition of *reversibility* (also called *detailed balance*) (Ross [1993, pg 172]), defined as

$$\pi_i(k)P_{i,j}(k) = \pi_j(k)P_{j,i}(k), \quad \textit{for all } i, j \in \Omega, \textit{ and all iterations } k. \tag{2.7}$$

The reversibility condition is sufficient for a unique solution to exist for (2.5) and (2.6) at each iteration *k*. Ross [1993, pg 177] shows that a necessary condition for reversibility is *multiplicativity* (i.e., for any three solutions $h, i, j \in \Omega$ such that $c_h \le c_i \le c_j$, and for all iterations *k*,

$$a_{h,j}(k, \Delta_{h,j}) = a_{h,i}(k, \Delta_{h,i}) a_{i,j}(k, \Delta_{i,j}) \tag{2.8}$$

where $a_{h,i}(k, \Delta_{h,i})$ is the probability of accepting the transition from solution *h* to solution *i* at iteration *k*). Reversibility is enforced by assuming conditions of symmetry on the solution generation probabilities and by either directly expressing the acceptance probability as the exponential form, or by requiring the multiplicative condition (2.8).

The homogeneous proofs of convergence in the literature (implicitly or explicitly) require condition (2.8) for the acceptance function, and then address the sufficient conditions on the solution generation matrix. For example, the original homogeneous proofs of convergence (Aarts and Laarhoven [1985] and Lundy and Mees [1986]) require the multiplicative condition for the acceptance function, and then assume that the solution generation function is symmetric and constant for all iterations *k*. Rossier et al. [1986] partition the solution space into blocks composed of neighboring solutions of equal objective function value, and then require only that the solution generation probabilities be symmetric between the blocks. They express the acceptance function as a ratio of the stationary distribution probabilities (discussed in Section 2.1.3). Faigle and Schrader [1988] and Faigle and Kern [1991] use a graph theoretic approach to relax the solution generation function symmetry condition. However, they require that the solution acceptance probability function satisfies (2.8).

Granville et al. [1994] propose an SA procedure for filtering binary images, where the acceptance function is based on the probability of the current solution, instead of the change in objective function value as in basic SA. Their probability function for accepting a candidate solution at (outer loop) iteration *k* is based on the ratio of the stationary probability of the incumbent solution from iteration $k-1$, versus the stationary probability of an initial solution (which is based on a maximum likelihood estimate). Their acceptance probability is

$$\xi(k) = (q\pi_i(0) / \pi_j(k-1))^{\phi(k)}$$

where $q = \inf_{i\in\Omega} \pi_i / \sup_{j\in\Omega} \pi_j$ (q must also be estimated), and $\phi(k)$ is a slowly increasing function. Therefore, the probability of a solution transition does not consider the objective function value of the candidate solution. They provide a proof of asymptotic convergence of this approach, but note that their proof methodology does not show that the set of globally optimal solutions are asymptotically uniformly distributed.

2.1.3 Origins of the Simulated Annealing Homogeneous Theory

SA and its homogeneous convergence theory are based on the work of Metropolis et al. [1953], who address problems in equilibrium statistical mechanics (Hammersley and Handscomb [1964, pg 117]). To see the relationship, consider a system in thermal equilibrium with its surroundings, in solution (state) S with energy $F(S)$. The probability density in phase space of the point representing S is proportional to

$$\exp(-\mathcal{B}F(S)), \tag{2.9}$$

where $\mathcal{B} = (zt)^{-1}$, z is Boltzmann's constant, and t is the absolute temperature of the surroundings. Therefore the proportion of time that the system spends in solution S is proportional to (2.9) (Hammersley and Handscomb [1964]), and so the equilibrium probability density for all $S \in \Omega$ is

$$\pi_S = \frac{\exp(-\mathcal{B}F(S))}{\int \exp(-\mathcal{B}F(S))dS}. \tag{2.10}$$

The expectation of any valid solution function *f(S)* is thus

$$E[f] = \frac{\int f(S)\exp(-\mathcal{B}F(S))dS}{\int \exp(-\mathcal{B}F(S))dS}. \tag{2.11}$$

The problem is that for many solution functions, (2.11) cannot be evaluated analytically. Hammersley and Handscomb [1964] note that one could theoretically use naive Monte Carlo techniques to estimate the value of the two integrals in (2.11), but this fails in practice because the exponential factor means that the significant portion of the integrals are concentrated in a very small region of the phase space Ω. The remedy is to use *importance sampling* (Bratley, Fox, and Schrage [1987, chapter 2]), by generating solutions with probability density (2.10). This approach would also seem to fail, because of the integral in the denominator of (2.10). However, Metropolis et al. [1953] solve this problem by first discretizing the solution space, such that the integrals in (2.10) and (2.11) are replaced by summations over discrete solutions $j \in \Omega$. They then construct an irreducible aperiodic Markov chain with transition probabilities $P_{i,j}$ such that

$$\pi_j = \sum_{i \in \Omega} \pi_i P_{i,j}, \quad \text{for all } j \in \Omega, \tag{2.12}$$

where

$$\pi_j = \frac{\exp(-\beta F(j))}{\sum_{i\in\Omega}\exp(-\beta F(i))}, \quad \text{for all } j \in \Omega. \tag{2.13}$$

Note that to compute the equilibrium distribution π, the denominator of (2.13) (a normalizing constant) does not need to be calculated. Instead, the *ratios* π_j / π_i need only be computed and a transition matrix ***P*** defined that satisfies (2.12). Metropolis et al. [1953] accomplish this by defining ***P*** as the product of symmetric solution generation probabilities $g_{i,j}$, and the equilibrium ratios π_j / π_i as

$$P_{i,j} = \begin{cases} g_{i,j}\pi_j / \pi_i & \textit{if } \ \pi_j / \pi_i < 1,\ j \neq i \\ g_{i,j} & \textit{if } \ \pi_j / \pi_i \geq 1,\ j \neq i \\ g_{i,i} + \sum\limits_{\substack{b\in\Omega, \\ \pi_b < \pi_i}} g_{i,b}\big(1 - (\pi_b / \pi_i)\big) & \textit{if } \ j = i \end{cases} \tag{2.14}$$

where

$$g_{i,j} \geq 0, \ \sum_{j\in\Omega} g_{i,j} = 1, \text{ and } g_{i,j} = g_{j,i} \quad \text{for all } i,j \in \Omega. \tag{2.15}$$

The use of stationary probability ratios to define the solution acceptance probabilities, combined with symmetric solution generation probabilities, enable Metropolis et al. [1953] to use the reversibility condition (2.7) to show that (2.14) and (2.15) satisfy (2.12).

2.1.4 Limitations of the Reversibility Condition

Homogeneous proofs of convergence for SA become more difficult when the reversibility condition is not met. Note that *existence* of a unique stationary distribution for each iteration k is easily shown by specifying that each transition matrix $\boldsymbol{P}(k)$ be irreducible and aperiodic. On the other hand, it becomes very difficult to derive an *explicit closed-form expression* for each stationary distribution $\pi(k)$ that remains analytically tractable as the problem's solution space becomes large. One can no longer use (2.7) to describe each stationary distribution, because in general, the multiplicative condition is not met. Instead, one must directly solve the system of equations (2.5) and (2.6). For example, Davis [1991] attempts to obtain a closed-form expression for $\pi(k)$ by using Cramer's rule and rewriting (2.5) and (2.6) as

$$\pi(k)(\boldsymbol{I} - \boldsymbol{P}(k)) = \boldsymbol{0} \tag{2.16}$$

and

$$\pi(k)\boldsymbol{1}^T = 1, \tag{2.17}$$

where boldface type indicates vector/matrix notation, $\boldsymbol{I}$ is the identity matrix, and $\boldsymbol{1}^T$ is a column vector of ones. Note that the $card(\Omega) \times card(\Omega)$ transition matrix $\boldsymbol{P}(k)$ associated with (2.16) is of rank $(card(\Omega) - 1)$ (Cinlar [1975, pg 154). Hence by deleting any one equation from (2.16), and substituting (2.17), the result is the set of $card(\Omega)$ linearly independent equations

$$\pi(k)\underline{(\boldsymbol{I} - \boldsymbol{P}(k))} = \boldsymbol{e} \tag{2.18}$$

where the square matrix $\underline{(\boldsymbol{I} - \boldsymbol{P}(k))}$ is obtained by substituting the i^{th} column of the matrix $(\boldsymbol{I} - \boldsymbol{P}(k))$ for a column vector of ones. The vector $\boldsymbol{e}$ is a row vector of zeroes, except for a one in the i^{th} position. Since $\underline{(\boldsymbol{I} - \boldsymbol{P}(k))}$ is of full rank, then its determinant (written as $\det\underline{(\boldsymbol{I} - \boldsymbol{P}(k))}$), is nonzero. Let $\underline{(\boldsymbol{I} - \boldsymbol{P}(k))^j}$ be the same matrix as $\underline{(\boldsymbol{I} - \boldsymbol{P}(k))}$

except that the elements of the j^{th} row of $\underline{(I - P(k))}$ are replaced by the vector $\boldsymbol{e}$. Therefore, for all iterations k,

$$\pi_j(k) = \frac{\det\left(I - P(k)\right)^j}{\det\underline{\left(I - P(k)\right)}}, \quad \text{for all } j \in \Omega. \tag{2.19}$$

Davis attempts to solve (2.19) for each $j \in \Omega$ via a multivariate Taylor series expansion of each determinant, but is not able to derive a tractable result.

Anily and Federgruen [1987] use perturbation analysis techniques (see e.g., Meyer [1980]) to prove convergence of a particular stochastic hill-climbing algorithm, by bounding the deviations of the sequence of stationary distributions of the particular case against the sequence of known stationary distributions corresponding to the SA algorithm. In general, this convergence proof approach is only useful for a restrictive class of SA algorithms, because the transition matrix condition number grows explosively as the number of iterations k becomes large; this issue is addressed in Chapter 4.

Overall, the difficulty of explicitly expressing the stationary distributions for large problem solution spaces, combined with problems of bounding the transition matrix condition number for large k, suggest that in general it is very difficult to prove asymptotic convergence of the SA algorithm by treating (2.5) and (2.6) as a linear algebra problem.

2.1.5 Convergence Rates for the Homogeneous Approach

Lundy and Mees [1986] note that for each fixed outer loop k, convergence to the solution equilibrium probability distribution vector $\pi(k)$ (in terms of the Euclidean distance between $P^{(m)}(k)$ and $\pi(k)$, as $m \rightarrow \infty$) is geometric since the solution space is finite, and the convergence factor is given by the second largest eigenvalue of the transition matrix $P(k)$. This result is based on a standard convergence theorem for irreducible, aperiodic homogeneous Markov chains (see e.g., Cinlar [1975, pg 378]). Note that the large solution space precludes practical calculation of the eigenvalues. Lundy and Mees [1986] conjecture that when the temperature t_k is near zero, the second largest eigenvalue will be close to one for problems with local optima, and thus convergence to the equilibrium distribution will be very slow. (Recall that the dominant eigenvalue for $P(k)$ is one, with algebraic multiplicity one (Isaacson and Madsen [1976, pg 125]).) Lundy and Mees [1986] use their conjecture to justify why SA should be initiated with a relatively high temperature. For an overview of current methods for assessing nonasymptotic rates of convergence for general homogeneous Markov chains, see Rosenthal [1995].

Overall, the assumption of stationarity for each iteration k limits practical application of homogeneous theory--Romeo and Sangiovanni-Vincentelli [1991] show that if equilibrium (for a Markov chain that satisfies the reversibility condition) is reached in a finite number of steps, then it is achieved in one step. Thus, Romeo and Sangiovanni-Vincentelli [1991] conjecture that there is essentially no hope for the most-used versions of SA to reach equilibrium in a finite number of iterations.

2.1.6 Inhomogenous Markov Chain Theory

The second convergence approach (Anily and Federgruen [1987], Borkar [1992], Connors and Kumar [1989], Gidas [1985], Hajek [1988], and Mitra et al. [1986]), is based on inhomogeneous Markov chain theory. In this approach, the Markov chain need not reach a stationary distribution (e.g., the SA inner loop need not be infinitely long) for each k. On the other hand, an infinite sequence of (outer loop) iterations k must still be examined, with the condition that the temperature parameter t_k cools sufficiently slowly. The proof given by Mitra et al. [1986] is based on satisfying the inhomogeneous Markov chain conditions of *weak* and *strong ergodicity* (Isaacson and Madsen [1976], Seneta [1981]). The proof requires four conditions:

a) The inhomogeneous SA Markov chain must be weakly ergodic (i.e., dependence on the initial solution vanishes in the limit).

b) An eigenvector $\pi(k)$ with eigenvalue one must exist such that (2.5) and (2.6) hold for every iteration k.

c) The Markov chain must be strongly ergodic (i.e., the Markov chain must be weakly ergodic and the sequence of eigenvectors $\pi(k)$ must converge to a limiting form), e.g.,

$$\sum_{k=0}^{\infty} \left\| \pi(k) - \pi(k+1) \right\| < \infty .$$

d) The sequence of eigenvectors must converge to a form where all probability mass is concentrated on the set of globally optimal solutions $opt \in \Omega^{opt} \subset \Omega$, such that $c_{opt} \equiv \min_{i \in \Omega}\{c_i\}$. Hence,

$$\lim_{k \to \infty} \pi(k) = \pi^{opt} .$$

(Note that weak and strong ergodicity are equivalent for homogeneous Markov chain theory.)

Mitra et al. [1986] satisfy condition a) (weak ergodicity) by first forming a lower bound on the probability of reaching any solution from any locally minimal solution, and then showing that this bound does not approach zero too quickly. For example, they define the lower bound for the SA transition probabilities (2.4) as

$$P_{i,j}^{(m)}(k) \geq w^m \exp\left(-m \Delta_L / t_{km-1}\right) ,$$

where m is the number of transitions needed to reach any solution from any solution of non-maximal objective function value, $w > 0$ is a lower bound on the one-step solution generation probabilities, and Δ_L is the maximum one-step increase in objective function value between any two solutions. Mitra et al. [1986] show the Markov chain is weakly ergodic if

$$\sum_{k=k_0}^{\infty} w^m \exp\left(-m \Delta_L / t_{km-1}\right) = \infty . \tag{2.20}$$

Therefore, weak ergodicity is obtained if the temperature t_k is reduced sufficiently slowly to satisfy (2.20). In general, the (infinite) sequence of temperatures $\{t_k\}$, $k = 1,2,\ldots$ needs to satisfy

$$t_k \geq \frac{\beta}{\log(k)} \tag{2.21}$$

where $\lim\limits_{k\to\infty} t_k = 0$, β is a problem-dependent constant, and k is the number of iterations.

Mitra et al. [1986] show that conditions b), c), and d) are satisfied by using the homogeneous Markov chain theory developed for the transition probabilities (2.4), and assuming a condition of symmetry on the solution generation function.

Romeo and Sangiovanni-Vincentelli [1991] comment that while the logarithmic cooling schedule (2.21) is a sufficient condition for the convergence, there are only a few values of β which make the logarithmic rule also necessary. Furthermore, there exists a unique choice of β which makes the logarithmic rule both necessary and sufficient for the convergence of SA to the set of global optima. Hajek [1988] was the first to show that the logarithmic cooling schedule (2.21) is both necessary and sufficient, by developing the tightest bound for the parameter β. He defines β to be greater than or equal to the depth of the deepest local minimum which is not a global minimum, under a *weak reversibility* assumption. Hajek defines a Markov chain to be weakly reversible if, for any pair of solutions $i, j \in \Omega$ and for any nonnegative real number *h*, *i* is reachable at height *h* from *j* if and only if *j* is reachable at height *h* from *i*. Note that Hajek does not attempt to satisfy the conditions of weak and strong ergodicity, but instead uses a more general probabilistic approach to develop a lower bound on the probability of escaping local, but not global optima. Hajek's necessary and sufficient conditions are substantiated by Connors and Kumar [1989], who base their convergence proof on a concept they call *orders of recurrence*,

$$B_i \equiv \sup\left\{ x \geq 0 : \sum_{k=0}^{\infty} \exp\left(-x / t_k\right)\pi_i(k) = \infty \right\}, \text{ for all } i \in \Omega .$$

They show that these orders of recurrence quantify the asymptotic behavior of each solution's occupation probability. Their key result is that the SA inhomogeneous Markov chain converges in a *Cesaro* sense to the set of solutions having the largest recurrence orders. Borkar [1992] improves on Connors and Kumar's [1989] convergence result by using a convergence/oscillation dichotomy result for martingales. Tsitsiklis [1989] uses bounds and estimates for singularly perturbed, approximately stationary Markov chains to develop a convergence theory that subsumes Hajek's [1988] condition of weak reversibility. (Note that Tsitsiklis [1989] defines $N(h) \subset \Omega$ as the set of all local minima (in terms of objective function value) of depth $h+1$ or more. Hence β is the smallest h such that all local (but not global) minima have depth h or less. Tsitsiklis [1989] conjectures that without some form of reversibility, there will exist no h such that the global optima are contained in the set of local optima.) Note that Chiang and Chow [1988], [1994], Borkar [1992], Connors and Kumar [1989], Hajek [1988], and Mitra et al. [1986] all require (either explicitly or implicitly) the multiplicative condition (2.8) for their proofs of convergence.

Anily and Federgruen [1987] present a proof of convergence of SA with more general acceptance probability functions. Using inhomogeneous Markov chain theory, they prove convergence under the following necessary and sufficient conditions:

a) The acceptance probability function must, for any iteration k, allow any hill climbing transition to occur with positive probability.

b) The acceptance probability function must be bounded and asymptotically monotone, with limit zero for hill climbing solution transitions.

c) In the limit, the stationary probability distribution must have zero probability mass for every non-globally optimal solution.

d) The probability of escaping from any locally (but not globally) optimal solution must not approach zero too quickly.

Anily and Federgruen [1987] use condition c) to relax the acceptance function multiplicative condition (2.8). However, in practice condition c) would be very difficult to check without assuming that (2.8) holds, for the reasons discussed in Section 2.1.4. Condition d) provides the necessary condition for the rate that the probability of hill climbing transitions approaches zero. Condition d) is expressed quantitatively as follows: let t_k be specified by (2.2), and define the minimum one-step acceptance probability as

$$\underline{a}(t_k) = \min_{\substack{i\in\Omega,\\ j\in\mathcal{N}(i)}} a_{i,j}(t_k).$$

Define the set of local optima $L \subset \Omega$ such that $i \in L \Rightarrow c_i \leq c_j$ for all $j \in (\mathcal{N}(i) \cap \Omega)$, and let

$$\overline{a}(t_k) = \max_{\substack{i\in L,\\ j\in\mathcal{N}(i)\backslash L}} a_{i,j}(t_k).$$

Finally, let any solution $j \in \Omega$ be reachable from any solution $i \in \Omega$ in q transitions or less. Then if (non-globally) locally optimal solutions exist, and

$$\sum_{k=1}^{\infty} \left(\underline{a}(t_k)\right)^q = \infty, \tag{2.22}$$

and conditions a), b), and c) hold, then the SA algorithm will asymptotically converge to the set of global optima, with probability one. However, if (non-globally) locally optimal solutions exist and

$$\sum_{k=1}^{\infty} \overline{a}(t_k) < \infty, \tag{2.23}$$

then the probability of each solution is asymptotically dependent upon the initial solution, and therefore the SA algorithm will not always converge to the set of global optima with probability one.

The inhomogeneous proof concept is stronger than the homogeneous approach in that it provides necessary conditions for the rate of convergence, but its asymptotic nature suggests that practical implementation is infeasible. Romeo and Sangiovanni-Vincentelli

[1991] note that "there is no reason to believe that truncating the logarithmic temperature sequence would yield a good configuration, since the tail of the sequence is the essential ingredient in the proof." In addition, the logarithmic cooling schedule dictates a very *slow* rate of convergence. Hence most recent work has focused on methods of improving SA's finite-time behavior and modifying or blending the algorithm with other search methods such as genetic algorithms (Liepins and Hilliard [1989]), tabu search (Glover [1994]), or both (Fox [1993]).

2.1.7 Finite-Time Behavior

Mitra et al. [1986] were the first to present a bound on the distance of the actual solution probability distribution from the optimal distribution after a finite number of iterations. They found that for a large number of iterations k, the L_1-norm of the difference of the current probability distribution from the optimal distribution is $O\left(1/k^{\min(b,d)}\right)$, where b and d are values characteristic of the problem instance. The values b and d are themselves functions of several parameters that describe the solution space topology, including the connectivity of the graph of the Markov chain, the maximum one-step hill climb between neighboring solutions, the cooling schedule, and on other parameters that are themselves functions of both solution objective function values and connectivity between solutions.

Implementation Issues: Implementation of SA follows two paths--that of specifying problem-specific choices (neighborhood, objective function, and constraints), and that of specifying generic choices (generation and acceptance probability functions, and the cooling schedule) (Eglese [1990]). The principal shortcoming of SA is that it often requires extensive computer time. Implementation work generally strives to retain SA's asymptotic convergence character, but at reduced computer run-time. The methods discussed here are mostly heuristic.

Problem-Specific Choices:

Neighborhoods: A key problem-specific choice concerns the neighborhood definition. SA's efficiency appears to be influenced by the neighborhood structure used (Moscato [1993]). The choice of neighborhood serves to enforce a topology--Eglese [1990] reports that "a neighborhood structure which imposes a 'smooth' topology where the local minima are shallow is preferred to a 'bumpy' topology where there are many deep local minima." Solla, Sorkin and White [1986] report similar conclusions, as does Fleischer and Jacobson [1996]. This makes intuitive sense, and it supports Hajek's results, which show that asymptotic convergence to the set of global optima depends on the depth of the local minima.

Another factor is neighborhood size. No theoretical results are available, other than the necessity of reachability (in a finite number of steps) from any solution to any other solution. Cheh et al. [1991] report that small neighborhoods are best, while Ogbu and Smith [1990] provide evidence that larger neighborhoods perform better than smaller neighborhoods. Goldstein and Waterman [1988] conjecture that if the neighborhood size is small compared to the total solution space cardinality, then the Markov chain cannot move around the solution space fast enough to find the minimum in a reasonable time. On the other hand, a very large neighborhood has the algorithm merely sampling randomly from a large portion of the solution space, and thus unable to focus on specific areas of the solution space. It is reasonable to believe that neighborhood size is heavily problem-specific, such that problems whose topology smoothness is relatively insensitive to different neighborhood definitions, may benefit from larger neighborhood sizes.

Fleischer [1994] and Fleischer and Jacobson [1996] use information theoretic concepts to show that the neighborhood structure can affect the *information rate* or total uncertainty associated with SA. Fleischer [1994] shows that SA tends to perform better as the entropy level of the associated Markov chain increases, and thus conjectures that an entropy measure could be useful for predicting when SA would perform well on a given problem. However, efficient ways of estimating the entropy are needed to make the concept practical.

A final issue on neighborhood definition addresses the solution space itself. Chardaire et al. [1995] propose a method for addressing 0-1 optimization problems, in which the solution space is progressively reduced by fixing the value of *strongly persistent* variables (which have the same value in all optimal solutions). They isolate the persistent variables during SA's execution by periodically estimating the expectation of the random variable (a vector of binary elements) that describes the current solution, and fixing the value of those elements in the random variable that meet threshold criteria.

Objective Functions: Another problem-specific choice involves the objective function specification. Stern [1992] recommends a heuristic temperature-dependent penalty function as a substitute for the actual objective function for problems where low cost solutions have neighbors of much higher cost, or in cases of degeneracy (i.e., large neighborhoods of solutions of equal, but high costs). The original objective function surfaces, as the penalty and the temperature are gradually reduced to zero. One speed-up technique is to evaluate only the difference in objective functions, $\Delta_{i,j}$, instead of calculating both c_i and c_j. Tovey [1988] suggests several methods of approximating $\Delta_{i,j}$ by using surrogate functions (faster to evaluate than $\Delta_{i,j}$, but not as accurate) probabilistically for cases when evaluation of $\Delta_{i,j}$ is expensive. He calls this technique the *surrogate function swindle*.

Straub et al. [1995] improve SA's performance on problems in chemical physics by using the classical density distribution instead of the molecular dynamics of single point particles to describe the potential energy landscape. Ma and Straub [1994] report that this has the effect of smoothing the energy landscape by reducing the number and depth of local minima.

Yan and Mukai [1992] consider the case when a closed-form formula for the objective function is not available. They use a probabilistic simulation to generate a sample objective function value for an input solution, and then accept the solution if the sample objective function value falls within a predetermined bound. They provide a proof of asymptotic convergence by extrapolating the convergence proofs for SA, and analyze the rate of convergence.

SA Generic Choices:

Generation Probability Functions: Generation probability functions are usually chosen as a uniform distribution with probabilities proportional to the size of the neighborhood. The generation probability function is usually not temperature-dependent. Fox [1993] suggests that instead of *blindly* generating neighbors uniformly, adopt an *intelligent* generation mechanism that modifies the neighborhood and its probability distribution to accommodate search intensification or diversification, in the spirit of the tabu search heuristic. Fox notes that SA convergence theory does not preclude this idea. Tovey [1988] suggests an approach with a similar effect--he calls it the *neighborhood prejudice swindle*.

Acceptance Probability Functions: The literature shows considerable experimentation with acceptance probability functions for hill climbing transitions. The most popular is the exponential form (2.1).

Ogbu and Smith [1990] consider replacing the basic SA acceptance function $a_{i,j}(k, \Delta_{i,j})$ with a geometrically decreasing form that is independent of the change in objective function value. They adopt a *probabilistic-exhaustive* heuristic technique in which randomly chosen neighbors of a solution are examined and all solutions that are accepted are noted, but only the last solution accepted becomes the new incumbent. Their hope is that this scheme will explore a more broad area of the solution space of a problem. Their acceptance probability function is defined for all solutions $i,j \in \Omega$ and for k = 1,2,..., K as

$$a_{i,j}(k) = a(k) = \begin{cases} a(1)x^{k-1} & \text{if } c_j > c_i \\ 1 & \text{otherwise} \end{cases}$$

where $a(1)$ is the initial acceptance probability value, x (<1) is the reducing factor, and K is the number of stages (equivalent to a temperature cooling schedule). They experimented with this method (and a neighborhood of large cardinality) on a permutation flowshop problem, and report that their approach found comparable solutions to the basic SA algorithm in one-third the time.

Cooling Schedules: An SA cooling schedule is defined by an initial temperature, a schedule for reducing temperature, and a stopping criterion. Romeo and Sangiovanni-Vincentelli [1991] note that an effective schedule is essential to reducing the amount of time required by the algorithm. Therefore much of the literature (Cardoso et al. [1994], Fox and Heine [1993]) is devoted to this topic.

Homogeneous SA convergence theory is often used to design good cooling schedules: start with an initial temperature t_0 for which a good approximation of the stationary distribution $\pi(t_0)$ is quickly reached. Reduce t_0 by an amount $\delta(t)$ small enough such that $\pi(t_0)$ is a good starting point to approximate $\pi(t_0 - \delta(t))$. Fix the temperature at a constant value during the iterations needed for the solution distribution to approximate $\pi(t_0 - \delta(t))$. Repeat the above process of cooling and iterating until no further improvement seems possible (Romeo and Sangiovanni-Vincentelli [1991]).

Cooling schedules are grouped into two classes: *static* schedules, which must be completely specified before the algorithm begins; and *adaptive* schedules, which adjust the temperature's rate of decrease from information obtained during the algorithm's execution. Cooling schedules are almost always heuristic; they seek to balance moderate execution time with SA's dependence on asymptotic behavior.

Strenski and Kirkpatrick [1991] present an exact (non-heuristic) characterization of finite-length annealing schedules. They consider extremely small problems that represent features (local optima and smooth/hilly topologies), and solve for solution probabilities after a finite number of iterations. They find that optimal cooling schedules are *not* monotone decreasing in temperature. Another result is that, for their test problem (a white noise surface), geometric and linear annealing schedules perform better than inverse logarithmic schedules, when sufficient computing effort is allowed. Their experiments do not show measurable performance differences between linear and geometric schedules. They also find that geometric schedules don't suffer seriously when initial temperatures are set too high. Their technique does not allow simulating adaptive schedules, but they do check some of the underlying assumptions. Their results show that the even the most robust adaptive schedule "produces annealing trajectories which are never in equilibrium" (Strenski and Kirkpatrick [1991]). However, they also conclude that the transition acceptance rate is not sensitive to the degree of closeness to an equilibrium distribution.

Christoph and Hoffmann [1993] also attempt to characterize optimal annealing schedules. They derive a relationship between a finite sequence of optimal temperature values and the number of iterations at each respective temperature for several small test problems to reach optimality (minimal mean final energy). They find that this *scaling behavior* is of the form

$$x_n = a_n v^{-b_n} \tag{2.24}$$

where a and b are scaling coefficients, $x_n = e^{-1/t_k}$ is referred to as temperature, v is the number of iterations at temperature x_n, and n is the number of times the temperature x_n is reduced. Their approach is to solve for the coefficients a and b based on known temperature and iteration parameter values for an optimal schedule for a few total annealing steps, and then use (2.24) to predict the optimal annealing schedule for intermediate times. They make no suggestions on how to efficiently solve for the necessary optimal schedules for a (typically large) problem instance.

Romeo and Sangiovanni-Vincentelli [1991] conclude that the theoretical results obtained thus far have not been able to explain why SA is so successful even when wild static annealing schedule heuristics are used. They conjecture that the neighborhood and the corresponding topology of the objective function are responsible for the behavior of the algorithm.

2.2 Threshold Accepting

Questioning the very need for a randomized acceptance function, Dueck and Scheuer [1990], and independently, Moscato and Fontanari [1990] propose the concept of threshold accepting (TA), where the acceptance function is defined as

$$a_{i,j}(k,\Delta_{i,j}) = \begin{cases} 1 & \textit{if } Q_k \geq \Delta_{i,j} \\ 0 & \textit{otherwise} \end{cases}$$

where Q_k is the threshold value at iteration *k*. Q_k is usually defined as a deterministic, nonincreasing step function. Dueck and Scheuer [1990] report dramatic improvements in traveling salesman problem solution quality and algorithm run-time over basic SA. Moscato and Fontanari [1990] report more conservative results--they conjecture that SA's probabilistic acceptance function does not play a major role in the search for near-optimal solutions.

Althofer and Koschnick [1991] develop a convergence theory for threshold accepting based on the concept that SA belongs to the convex hull of TA. Their idea is that (for a finite Q_k threshold sequence) there can exist only finitely many TA transition matrices; but SA can have infinitely many transition matrices because of the real-valued nature of the temperature at each iteration. However, every SA transition matrix for a given problem can be represented as a convex combination of the finitely many TA transition matrices. Althofer and Koschnick [1991] are unable to prove that TA will asymptotically reach a global minimum, but they do prove the existence of threshold schedules that provide convergence to within an ε-neighborhood of the optimal solutions.

Hu et al. [1995] modify TA to include a non-monotonic, self-tuning threshold schedule in the hope of improving finite-time performance. They allow the threshold Q_k to change dynamically up or down, based on the perceived likelihood of being near a local minimum. This is accomplished by using a principle they call *dwindling expectation*--when the algorithm fails to move to neighboring solutions, the threshold Q_k is gradually increased, in the hope of eventually escaping a local optimum. Conversely, when solution transitions are successful, the threshold is reduced, in order to explore local optima. Their experimental results on two traveling salesman problems showed that their algorithm outperformed TA in terms of finding good solutions earlier in the optimization process.

TA's greatest values are its ease of implementation and its generally faster execution time than SA, due to the reduced computational effort from avoiding acceptance probability calculations and generation of random numbers (Moscato and Fontanari [1990]). Compared to SA, relatively few TA applications are reported in the literature (Lin et al. [1995], Scheermesser and Bryngdahl [1995], and Nissen and Paul [1995]).

2.3 Other Stochastic Approaches

2.3.1 Probabilistic Tabu Search

Tabu search (Glover [1994]) is a general framework for a variety of iterative local search strategies for discrete optimization. Tabu search (TS) uses the concept of *memory* by controlling the algorithm's execution via a dynamic list of forbidden moves. This allows the TS algorithm to intensify or diversify its search of a given $(\Omega, \mathcal{N}, c)$ problem's solution space as necessary, in an effort to avoid entrapment in local optima. Note that a criticism of SA is that it is completely memoryless, i.e., SA disregards all historical information gathered during the algorithm's execution. On the other hand, no proofs of convergence exist in the literature, for the general TS algorithm proposed by Glover. Therefore, Faigle and Kern [1992] propose a particular TS algorithm they call *probabilistic tabu search,* which attempts to capitalize on both the asymptotic optimality of SA and the memory feature of TS. In probabilistic TS, the probabilities of generating and accepting each candidate solution are set as functions of both a temperature parameter (as in SA) and information gained in previous iterations. Faigle and Kern are then able to prove asymptotic convergence of their particular TS algorithm, by using methods developed for SA (Faigle and Kern [1991]).

2.3.2 Random Noise

Charon and Hudry [1993] advocate a simple descent algorithm they call the *Noising Method.* The algorithm first perturbs the solution space by adding random noise to the problem's objective function values. The noise gradually reduced to zero during the algorithm's execution, allowing the original problem structure to reappear. Charon and Hudry provide computational results, but do not prove that the algorithm will asymptotically converge to the set of globally optimal solutions.

Storer et al. [1992] propose an optimization strategy for sequencing problems, by integrating fast, problem-specific heuristics with local search. Their key contribution is to base the definition of the search neighborhood on a heuristic problem pair (*h, p*), where *h* is a fast, known, problem-specific heuristic and *p* represents the problem data. By perturbing the heuristic, the problem, or both, a neighborhood of solutions is developed. This neighborhood then forms the basis for local search. Their hope is that the perturbations will cluster good solutions close together, thus making it easier to perform local search.

2.3.3 Genetic Algorithms

Genetic algorithms (Liepins and Hilliard [1989]) emulate the evolutionary behavior of biological systems. They generate a sequence of populations of candidate solutions to the underlying optimization problem by using a set of genetically inspired stochastic solution transition operators to transform each population of candidate solutions into a descendent population. The three most popular transition operators are reproduction, cross-over, and mutation (Davis [1991]). Davis [1991] and Rudolph [1994] attempt to use homogeneous finite Markov chain techniques to prove convergence of genetic algorithms, but are unable to develop a theory comparable in scope to that of SA.

CHAPTER 3: GENERALIZED HILL CLIMBING CONVERGENCE BASED ON REVERSIBLE MARKOV CHAIN THEORY

Chapter 3 provides a proof of convergence, and examples, for a class of GHC algorithms. When the objective function value of a globally optimal solution is known (and the goal is to identify an associated optimal solution), then the solution acceptance probability for all $i, j \in \Omega$ can be expressed as a ratio of acceptance probabilities involving the globally optimal solution. This enables the proof of convergence to use reversible homogeneous Markov chain theory.

3.1 Generalized Hill Climbing Class Description

Consider a class of GHC algorithms where, for all $i,j \in \Omega$ and $opt \in \Omega^{opt}$, the solution acceptance probabilities are defined as

$$\Pr\left(R_k(i,j) \geq \Delta_{i,j}\right) \equiv \min\left\{1, \frac{\Pr\left(\widetilde{R}_k(opt,j) \geq \Delta_{opt,j}\right)}{\Pr\left(\widetilde{R}_k(opt,i) \geq \Delta_{opt,i}\right)}\right\}, \tag{3.1}$$

where $\widetilde{R}_k(opt,i)$ and $\widetilde{R}_k(opt,j)$ are random variables. The elements of the transition matrix $\boldsymbol{P}(k)$ are defined as

$$P_{i,j}(k) = \begin{cases} g_{i,j}(k)\Pr\left(R_k(i,j) \geq \Delta_{i,j}\right) & \textit{for all } i \in \Omega,\ j \in \mathcal{N}(i),\ j \neq i \\ 1 - \sum\limits_{\substack{l \in \mathcal{N}(i) \\ l \neq i}} P_{i,l}(k) & j = i \\ 0 & \textit{otherwise}, \end{cases} \tag{3.2}$$

and the solution generation probabilities $g_{i,j}(k)$ satisfy

$$\sum_{j\in\mathcal{N}(i)} g_{i,j}(k) = 1\,. \tag{3.3}$$

Note that from (3.1), this class of GHC algorithms requires that the globally optimal objective function value, c_{opt}, be known.

3.2 Proof of Convergence

Theorem 3.1 provides sufficient conditions on this class of GHC algorithms for the existence of a unique stationary distribution for each iteration *k*. Theorem 3.2 provides sufficient conditions for the sequence of stationary distributions $\pi(k)$ to converge to the set of globally optimal solutions.

Theorem 3.1: *Let* $(\Omega, \mathcal{N}, c)$ *denote an instance of a discrete optimization problem. Let each* GHC *transition probability* $P_{i,j}(k)$ *be defined by* (3.2), *where the acceptance probabilities are defined by* (3.1). *Let the generation probabilities* $g_{i,j}(k)$ *satisfy* (3.3) *and the conditions*

(a) *for all* $i, j \in \Omega$ *and all iterations k, there exists an integer* $d \geq 1$ *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* $g_{l_n, l_{n+1}}(k) > 0,\ n = 0, 1, \ldots, d-1$,

(b) *for all* $i, j \in \Omega$, $g_{i,j}(k) = g_{j,i}(k)$, *and* $g_{i,j}(k)$ *is independent of k* *(i.e.,* $g_{i,j}(k) = g_{j,i}(k) = g_{i,j} = g_{j,i}$*).*

Moreover, let the acceptance probabilities satisfy

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

Then the stationary distribution for all $i \in \Omega$, *and for each outer loop iteration k, is*

$$\pi_i(k) = \frac{\Pr\left(\tilde{R}_k(opt,i) \geq \Delta_{opt,i}\right)}{\sum\limits_{n \in \Omega} \Pr\left(\tilde{R}_k(opt,n) \geq \Delta_{opt,n}\right)} \tag{3.4}$$

Proof : The irreducibility and aperiodicity conditions of Theorem 2.1 are first shown to be satisfied. Then (3.1), (3.2), and (3.4) are shown to satisfy the reversibility condition (2.7).

Irreducibility: From condition (a), the matrix corresponding to the solution generation probabilities must assign a path of positive probability between any two solutions $i, j \in \Omega$. Furthermore, condition (c) requires that the acceptance probability for hill climbing moves between all solutions be strictly positive for all finite iterations *k*. Therefore, conditions (a) and (c) together imply that all solutions communicate, which is an equivalent definition of irreducibility (Ross [1993, pg. 144]).

Aperiodicity (From Aarts and Korst [1989, pg. 39]): Let $i, j \in \Omega$ with $c_i < c_j$ and $g_{i,j} > 0$. By condition (a), such a pair always exists provided $\Omega \neq \Omega^{opt}$. Without loss of generality, assume that for at least one solution $n \in \mathcal{N}(i)$, $\Pr\left(R_k(i,n) \geq \Delta_{i,n}\right) < 1$ (without this assumption, then it is possible that for some iteration *k*, all hill climbing transitions would be accepted with probability one, which represents pure Monte Carlo search (Hammersley and Handscomb [1964]). Note that the opposite extreme would occur for iterations *k* if no hill climbing transitions are accepted, which represents local search (Papadimitriou and Steiglitz [1982]); hill climbing search algorithms are designed to operate between these two extremes). Therefore

$$
\begin{aligned}
P_{i,i}(k) &= 1 - \sum_{n\in\Omega,\ n\neq i} g_{i,n} \Pr\left(R_k(i,n) \geq \Delta_{i,n}\right) \\
&= 1 - g_{i,j} \Pr\left(R_k(i,j) \geq \Delta_{i,j}\right) - \sum_{n\in\Omega,\ n\neq i,j} g_{i,n} \Pr\left(R_k(i,n) \geq \Delta_{i,n}\right) \\
&> 1 - g_{i,j} - \sum_{n\in\Omega,\ n\neq i,j} g_{i,n} \\
&= 1 - \sum_{n\in\Omega,\ n\neq i} g_{i,n} \\
&= 0
\end{aligned}
$$

which is the criterion for aperiodicity (Cinlar [1975, pg. 126]).

To complete the proof, reversibility (2.7) is shown to be satisfied. Substituting (3.1), (3.2), and (3.4) into (2.7) leads to

$$
\frac{\Pr\left(\tilde{R}_k(opt,i) \geq \Delta_{opt,i}\right)}{\sum_{n\in\Omega} \Pr\left(\tilde{R}_k(opt,n) \geq \Delta_{opt,n}\right)} P_{i,j}(k) = \frac{\Pr\left(\tilde{R}_k(opt,j) \geq \Delta_{opt,j}\right)}{\sum_{n\in\Omega} \Pr\left(\tilde{R}_k(opt,n) \geq \Delta_{opt,n}\right)} P_{j,i}(k). \tag{3.5}
$$

Consider any two solutions $i,j \in \Omega,\ i \neq j$, and examine reversibility under the following cases:

Case 1: If $\Pr\left(\tilde{R}_k(opt,j) \geq \Delta_{opt,j}\right) < \Pr\left(\tilde{R}_k(opt,i) \geq \Delta_{opt,i}\right)$, then (3.5) can be rewritten as

$$
\frac{\Pr\left(\tilde{R}_k(opt,i) \geq \Delta_{opt,i}\right)}{\sum_{n\in\Omega} \Pr\left(\tilde{R}_k(opt,n) \geq \Delta_{opt,n}\right)} g_{i,j}(k) \frac{\Pr\left(\tilde{R}_k(opt,j) \geq \Delta_{opt,j}\right)}{\Pr\left(\tilde{R}_k(opt,i) \geq \Delta_{opt,i}\right)} = \frac{\Pr\left(\tilde{R}_k(opt,j) \geq \Delta_{opt,j}\right)}{\sum_{n\in\Omega} \Pr\left(\tilde{R}_k(opt,n) \geq \Delta_{opt,n}\right)} g_{j,i}(k)
$$

where reversibility holds, from condition (b).

Case 2: If $\Pr\left(\tilde{R}_k(opt,j) \geq \Delta_{opt,j}\right) = \Pr\left(\tilde{R}_k(opt,i) \geq \Delta_{opt,i}\right)$, then (3.5) can be rewritten as

$$\frac{\Pr\left(\tilde{R}_k(opt,i) \ge \Delta_{opt,i}\right)}{\sum_{n\in\Omega}\Pr\left(\tilde{R}_k(opt,n) \ge \Delta_{opt,n}\right)} g_{i,j}(k) = \frac{\Pr\left(\tilde{R}_k(opt,j) \ge \Delta_{opt,j}\right)}{\sum_{n\in\Omega}\Pr\left(\tilde{R}_k(opt,n) \ge \Delta_{opt,n}\right)} g_{j,i}(k).$$

By condition (b), reversibility again holds.

Case 3: If $\Pr\left(\tilde{R}_k(opt,j) \ge \Delta_{opt,j}\right) > \Pr\left(\tilde{R}_k(opt,i) \ge \Delta_{opt,i}\right)$, then by the *min* function in (3.1), this reduces to Case 2. ■

Theorem 3.2 proves convergence of the stationary distributions $\pi(k)$ to the set of globally optimal solutions.

Theorem 3.2: *Under the conditions and assumptions of Theorem 3.1, if, for all $i,j \in \Omega$ and for all iterations k,*

$$c_i < c_j \Rightarrow \lim_{k\to\infty} \Pr\left(R_k(i,j) \ge \Delta_{i,j}\right) = 0 \qquad (3.6)$$

then, as k approaches infinity, the unique stationary distributions $\pi(k)$ of Theorem 3.1 approach the limiting form

$$\lim_{k\to\infty} \pi_i(k) = \begin{cases} \left(card\left(\Omega^{opt}\right)\right)^{-1} & if\ \ i \in \Omega^{opt} \\ 0 & otherwise. \end{cases} \qquad (3.7)$$

Proof: The proof, based on a result by Aarts and Korst [1989, pg. 18], examines whether a limit exists for the sequence of stationary distributions as k approaches infinity. Taking the limit of (3.4) for all $i \in \Omega$ leads to

$$\lim_{k\to\infty} \pi_i(k) = \lim_{k\to\infty} \frac{\Pr\left(\widetilde{R}_k(opt,i) \geq \Delta_{opt,i}\right)}{\sum_{j\in\Omega} \Pr\left(\widetilde{R}_k(opt,j) \geq \Delta_{opt,j}\right)} \tag{3.8}$$

For any $opt \in \Omega^{opt}$, (3.8) reduces to

$$\lim_{k\to\infty} \pi_{opt}(k) = \lim_{k\to\infty} \frac{1}{\sum_{j\in\Omega^{opt}} (1)}$$

$$= \left(card\left(\Omega^{opt}\right)\right)^{-1} \tag{3.9}$$

For all nonoptimal solutions $n \in \Omega \backslash \Omega^{opt}$, from (3.6), (3.8) reduces to

$$\lim_{k\to\infty} \pi_n(k) = \lim_{k\to\infty} \frac{0}{\sum_{j\in\Omega^{opt}} (1)}$$

$$= 0 \tag{3.10}$$

Together, (3.9) and (3.10) satisfy (3.7), which completes the proof. ■

3.3 Illustrative Examples

Four illustrative examples (Johnson and Jacobson [1996]) are presented to show how the class of GHC algorithms defined in Section 3.1 offers flexibility in defining the solution acceptance random variable.

3.3.1 Generalized Hill Climbing Acceptance Using an Exponential Random Variable

Set $\widetilde{R}_k(opt,n) \equiv -t_k \ln(U)$ for all $opt \in \Omega^{opt}, n \in \Omega$ and for all iterations k, where t_k is the temperature parameter (2.2) and U is distributed $U(0,1)$. Then $\Pr\left(\widetilde{R}_k(opt,n) \geq \Delta_{opt,n}\right) = \exp(-\Delta_{opt,n} / t_k)$, and therefore (3.1) can be expressed as

$$\Pr\left(R_k(i,j) \geq \Delta_{i,j}\right) = \min\left\{1, \frac{\exp(-\Delta_{opt,j} / t_k)}{\exp(-\Delta_{opt,i} / t_k)}\right\}$$
$$= \min\left\{1, \exp(-\Delta_{i,j} / t_k)\right\},$$

which is the SA formulation (Laarhoven and Aarts [1987, pg. 20]). Note that for all $i,j \in \Omega$ and for all $t_k > 0$, $\exp(-\Delta_{i,j} / t_k) > 0$, hence condition (c) is satisfied. Therefore if conditions (3.3), (a), and (b) on the solution generation probabilities are met, then Theorem 3.1 applies. Finally, observe that $\lim_{k \to \infty}\left[\min\left\{1, \exp(-\Delta_{i,j} / t_k)\right\}\right] = 0$ if $\Delta_{i,j} > 0$, and 1 otherwise, hence condition (3.6) is satisfied and so Theorem 3.2 also applies. Note that the optimal objective function value c_{opt} does *not* need to be known to calculate (3.1).

3.3.2 Generalized Hill Climbing Acceptance Using a (Continuous-Valued) Geometric Random Variable

Set $\widetilde{R}_k(opt,n) \equiv \dfrac{\ln(U)}{\ln(1-p_k)}$, for all $opt \in \Omega^{opt}, n \in \Omega$ and for all iterations *k*, where p_k is a probability parameter such that $0 < p_k < 1$ and $\lim_{k \to \infty} p_k = 1$, and *U* is distributed $U(0,1)$. Then $\Pr\left(\widetilde{R}_k(opt,n) \geq \Delta_{opt,n}\right) = (1-p_k)^{\Delta_{opt,n}}$, and therefore (3.1) can be expressed as

$$\Pr\left(R_k(i,j) \geq \Delta_{i,j}\right) = \min\left\{1, \frac{(1-p_k)^{\Delta_{opt,j}}}{(1-p_k)^{\Delta_{opt,i}}}\right\}$$
$$= \min\left\{1, (1-p_k)^{\Delta_{i,j}}\right\}.$$

(This geometric formulation is continuous-valued because $\Delta_{i,j} \in \Re$.) Observe that for all $i,j \in \Omega$ and for all $p_k < 1$, $(1-p_k)^{\Delta_{i,j}} > 0$, hence condition (c) is satisfied. Therefore if conditions (3.3), (a), and (b) on the solution generation probabilities are met, then Theorem 3.1 applies. Also, note that $\lim_{k\to\infty}\left[\min\left\{1, (1-p_k)^{\Delta_{i,j}}\right\}\right] = 0$ if $\Delta_{i,j} > 0$, and 1 otherwise, hence condition (3.6) is satisfied and so Theorem 3.2 applies. Again, the value of c_{opt} is not needed to calculate (3.1).

3.3.3 Generalized Hill Climbing Acceptance Using a Weibull Random Variable

Set $\widetilde{R}_k(opt,n) \equiv t_k(-\ln(U))^{1/\alpha}$ for all $opt \in \Omega^{opt}, n \in \Omega$ and for all iterations k, where $\alpha > 0$ is a shape parameter, t_k is the temperature parameter described in (2.2), and U is distributed $U(0,1)$. Then $\Pr\left(\widetilde{R}_k(opt,n) \geq \Delta_{opt,n}\right) = \exp\left(-(\Delta_{opt,n} / t_k)^{\alpha}\right)$, and therefore (3.1) can be written as

$$\Pr\left(R_k(i,j) \geq \Delta_{i,j}\right) = \min\left\{1, \frac{\exp\left(-(\Delta_{opt,j} / t_k)^{\alpha}\right)}{\exp\left(-(\Delta_{opt,i} / t_k)^{\alpha}\right)}\right\}$$

$$= \min\left\{1, \exp\left(\frac{1}{t_k^{\alpha}}\left[\Delta_{opt,i}^{\alpha} - \Delta_{opt,j}^{\alpha}\right]\right)\right\}. \qquad (3.11)$$

Observe that for all $i,j \in \Omega$ and for all $t_k > 0$, $\exp\left(\frac{1}{t_k^{\alpha}}\left[\Delta_{opt,i}^{\alpha} - \Delta_{opt,j}^{\alpha}\right]\right) > 0$, hence condition (c) is satisfied. Thus if conditions (3.3), (a), and (b) on the solution generation probabilities are met, then Theorem 3.1 applies. Also, note that, for all k, $\exp\left(-(\Delta_{opt,j} / t_k)^{\alpha}\right) < \exp\left(-(\Delta_{opt,i} / t_k)^{\alpha}\right)$ implies $c_j > c_i$, and so

$\lim\limits_{k\to\infty}\left[\min\left\{1,\ \exp\left(\frac{1}{t_k^{\alpha}}\left[\Delta_{opt,i}^{\alpha}-\Delta_{opt,j}^{\alpha}\right]\right)\right\}\right]=0$ if $i,j\notin\Omega^{opt}$, and 1 otherwise, hence condition (3.6) is satisfied and Theorem 3.2 applies. Finally, note that the Weibull random variable formulation requires c_{opt} to be known, and so the acceptance probability distribution for $R_k(i,j)$ cannot be written explicitly in terms only of solutions i and j. However, its distribution function can be expressed in closed form, as seen by (3.11), and as shown for the general case in Section 3.3.4.

3.3.4 The Generalized Hill Climbing Acceptance Distribution for a General Random Variable

Any acceptance probability distribution for $\widetilde{R}_k(opt,n)$ that can be described *explicitly*, enables the acceptance probability distribution for $R_k(i,j)$ to be expressed *implicitly*. To see this, rewrite (3.1) as

$$\begin{aligned}\Pr\left(R_k(i,j)\le\Delta_{i,j}\right)&=1-\Pr\left(R_k(i,j)\ge\Delta_{i,j}\right)\\&=1-\frac{\Pr\left(\widetilde{R}_k(opt,j)\ge\Delta_{opt,j}\right)}{\Pr\left(\widetilde{R}_k(opt,i)\ge\Delta_{opt,i}\right)}\\&=\frac{\Pr\left(\widetilde{R}_k(opt,i)\ge\Delta_{opt,i}\right)-\Pr\left(\widetilde{R}_k(opt,j)\ge\Delta_{opt,j}\right)}{\Pr\left(\widetilde{R}_k(opt,i)\ge\Delta_{opt,i}\right)}.\end{aligned}\tag{3.12}$$

Therefore, (3.12) provides a general expression for the distribution function for $R_k(i,j)$ in terms of the distribution functions for $\widetilde{R}_k(opt,i)$ and $\widetilde{R}_k(opt,j)$. If (3.12) is invertible, then a closed form expression for $R_k(i,j)$ can be obtained.

CHAPTER 4: GENERALIZED HILL CLIMBING CONVERGENCE WITHOUT REVERSIBILITY

Chapter 4 provides a proof of GHC algorithm convergence based on homogeneous Markov chain theory that does not require the sufficient condition of reversibility (2.7). The proof methodology is based on a perturbation theory developed for problems in linear algebra. Convergence is shown by bounding the sequence of stationary distributions associated with a general GHC algorithm, with a sequence of stationary distributions associated with a GHC algorithm with known convergence properties (e.g., see Section 3.3.1). The key benefits of this approach are that the reversibility condition is not required and the objective function value of a globally optimal solution is not needed. One limitation of this approach is that as *k* approaches infinity, the transition matrix of the general GHC algorithm must converge to the transition matrix of the convergent GHC algorithm at a rate fast enough to control the bound.

4.1 Generalized Hill Climbing Class Description

Consider a class of GHC algorithms where, for all $i,j \in \Omega$, the solution acceptance probabilities are required only to satisfy the conditions

$$\Pr\left(R_k(i,j) \geq \Delta_{i,j}\right) > 0 \textit{ for all iterations } k. \tag{4.1}$$

and

$$c_i < c_j \Rightarrow \lim_{k\to\infty} \Pr\left(R_k(i,j) \geq \Delta_{i,j}\right) = 0 \tag{4.2}$$

Conditions (4.1) and (4.2), when combined with the generation probability conditions of Theorem 4.1, ensure that all neighboring solutions communicate for all iterations *k*, and that communication is gradually lost as *k* approaches infinity.

4.2 Proof of Convergence

Define $\boldsymbol{P}^R(k)$ and $\pi^R(k)$ to be the respective transition matrix and stationary distribution for an arbitrary GHC algorithm, for each iteration *k*. Let $\boldsymbol{P}^E(k)$ and $\pi^E(k)$ be the respective transition matrix and stationary distribution for an instance of the GHC algorithm that is known to converge to the set of globally optimal solutions, as *k* approaches infinity. Theorem 4.1 provides sufficient conditions on the arbitrary GHC algorithm transition matrix $\boldsymbol{P}^R(k)$ to ensure the existence of a unique stationary distribution $\pi^R(k)$ for each finite iteration *k*. Theorem 4.2 proves the existence of a generalized matrix inverse, and Theorem 4.3 formulates a bound for $\pi^R(k)$ in terms of the difference between $\boldsymbol{P}^E(k)$ and $\boldsymbol{P}^R(k)$, and a generalized inverse corresponding to $\boldsymbol{P}^E(k)$. Theorem 4.4 proves asymptotic convergence to the set of optimal solutions under a condition on the bound.

Theorem 4.1: *Let* $(\Omega, \mathcal{N}, c)$ *denote an instance of a discrete optimization problem. Let the one-step transition probabilities associated with the* GHC *algorithm be defined as*

$$P_{i,j}^{R}(k)=\begin{cases} g_{i,j}(k)\Pr\left(R_k(i,j)\geq \Delta_{i,j}\right) & j\in\mathcal{N}(i),\, j\neq i \\ 0 & j\notin\mathcal{N}(i),\, j\neq i \\ 1-\sum\limits_{\substack{l\in\mathcal{N}(i)\\ l\neq i}} P_{i,l}(k) & j=i. \end{cases} \tag{4.3}$$

Let the acceptance probabilities satisfy (4.1). Let the generation probabilities $g_{i,j}(k)$ *satisfy*:

(a) *For all k and for all* $i,j\in\Omega$ *there exists an integer* $d\geq 1$ *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* $i,j\in\Omega$, $g_{i,j}(k)>0\Rightarrow g_{j,i}(k)>0$.

(c) *Each generation probability* $g_{i,j}(k)$ *is independent of* k (e.g., $g_{i,j}(k)=g_{i,j}$ *for all* $i,j\in\Omega$ *and all iterations k*).

Then for each iteration k, there exists a unique stationary distribution $\pi^R(k)$.

Proof: The irreducibility and aperiodicity conditions of Theorem 2.1 are satisfied, as shown in the proof of Theorem 3.1. ■

Conditions (4.1), (a), (b), and (c) ensure that a path exists between any two solutions in Ω, for all iterations *k*. Condition (c) also implies that all transition probabililities are asymptotically monotone, which is needed for the proof of Theorem 4.5.

Note that although proving *existence* of the unique stationary distribution $\pi^R(k)$ for the transition probabilities (4.3) is straightforward, it is very difficult to derive an analytically tractable *explicit formulation* of the stationary distribution, for the reasons discussed in Chapter 2. Therefore, convergence to the set of optimal solutions is shown in terms of a bound on the stationary probabilities $\pi^R(k)$, as k approaches infinity.

To achieve this, the *group inverse*, $A^\#(k)$, is introduced. Let $\{P^E(k)\}$, $k = 1,2,\ldots$, be the sequence of transition matrices of irreducible, aperiodic Markov chains whose corresponding stationary distributions $\{\pi^E(k)\}$, $k = 1,2,\ldots$, converge as $k \to \infty$ to the form where all probability mass is concentrated on the set of optimal solutions Ω^{opt}. Meyer [1975] defines $A^\#(k)$ as follows:

first let

$$A(k) \equiv I - P^E(k), \tag{4.4}$$

where I is an $|\Omega| \times |\Omega|$ identity matrix, and let

$$\Pi(k) \equiv \begin{bmatrix} \pi^E(k) \\ \vdots \\ \pi^E(k) \end{bmatrix} \tag{4.5}$$

be the square stochastic matrix composed of identical rows of the stationary probability vector $\pi^E(k)$ corresponding to $P^E(k)$. The group inverse is then defined as

$$A^\#(k) \equiv \left(I - P^E(k) + \Pi(k)\right)^{-1} - \Pi(k) \tag{4.6}$$

(Meyer [1975]). Theorem 4.2 guarantees that $A^{\#}(k)$ exists, for all *k*.

Theorem 4.2: *For every transition matrix P(k), the generalized inverse matrix* $A^{\#}(k)$ *exists, where* $A(k)$ *is defined by* (4.4). *Furthermore, if P(k) is irreducible and aperiodic, then* $A^{\#}(k)=\sum_{m=0}^{\infty}\left(P^{(m)}(k)-\Pi(k)\right)$, *where* $P^{(m)}(k)$ *refers to the m-step transition probability matrix as defined in Theorem 2.1, and* $\Pi(k)$ *is defined by* (4.5).

Proof: Modify Meyer's [1975] result by making each matrix a function of k. ■

Define the vector one-norm $\|x\|_1=\sum_{j\in\Omega}|x_j|$, with corresponding matrix one-norm $\|X\|_1=\max_{i\in\Omega}\sum_{j\in\Omega}|X_{i,j}|$. Theorem 4.3 provides a bound for $\pi^R(k)$ in terms of $\left(P^E(k)-P^R(k)\right)$ and $A^{\#}(k)$, for each iteration *k*.

Theorem 4.3: *For each iteration k, let* $P^E(k)$ *be the transition matrix of an irreducible, aperiodic homogeneous Markov chain, let* $P^R(k)$ *be defined as in Theorem 4.1, and let* $A^{\#}(k)$ *be defined by* (4.6). *Then*

$$\left\|\pi^E(k)-\pi^R(k)\right\|_1\leq\left\|\left(P^E(k)-P^R(k)\right)A^{\#}(k)\right\|_1. \tag{4.7}$$

Proof: Modify Meyer's [1980] Theorem 4.2 by making each vector and matrix a function of k. ■

From Theorem 4.3, to prove that $\lim_{k\to\infty}\left\|\pi^E(k)-\pi^R(k)\right\|_1=0$, it suffices to show that $\left\|\left(P^E(k)-P^R(k)\right)A^{\#}(k)\right\|_1$ approaches zero, for k sufficiently large. Theorem 4.4

represents a special case of the GHC algorithm where $\left\|\left(\boldsymbol{P}^{E}(k)-\boldsymbol{P}^{R}(k)\right)A^{\#}(k)\right\|_{1}$ can be bounded by an arbitrarily small number.

Theorem 4.4: *For each iteration k, let the GHC algorithm one-step transition matrix* $\boldsymbol{P}^{R}(k)$ *be defined in Theorem 4.1, and let the GHC algorithm acceptance probability* $\Pr\left(R_{k}(i,j) \geq \Delta_{i,j}\right)$ *satisfy* (4.2). *Let* $\boldsymbol{P}^{E}(k)$ *be the transition matrix of an irreducible aperiodic Markov chain whose solution acceptance probabilities satisfy* (4.1) *and* (4.2), *and whose generation probabilities are identical to the solution generation probabilities of* $\boldsymbol{P}^{R}(k)$. *Assume that all probability mass of the stationary distributions* $\pi^{E}(k)$ *converge to the set of optimal solutions* $\Omega^{opt} \subset \Omega$ *as* $k \to \infty$. *Let* $A(k)$ *be defined by* (4.4), *and assume that for every* $Z \in \Re^{+}$, *there exists an iteration* $k_{0}(Z)$ *such that*

$$\left\|\left(\boldsymbol{P}^{E}(k)-\boldsymbol{P}^{R}(k)\right)A^{\#}(k)\right\|_{1} < Z, \tag{4.8}$$

for all $k \geq k_{0}(Z)$. *Then, by making Z arbitrarily close to zero, there exists some* $k_{0}(Z)$ *such that for all* $k \geq k_{0}(Z)$, *the stationary distributions* $\pi^{R}(k)$ *of Theorem 4.1 converge as* $k \to \infty$ *to the form where all probability mass is concentrated on the set of optimal solutions* Ω^{opt}.

Proof: For any $Z > 0$ (arbitrarily close to zero), there exists some $k_{0}(Z)$ such that (4.8) holds for all $k \geq k_{0}(Z)$. Use this and (4.7) to obtain

$$\left\|\pi^{E}(k)-\pi^{R}(k)\right\|_{1} < Z,$$

for all $k \geq k_{0}(Z)$.

Therefore,

$$\lim_{k\to\infty}\left\|\pi^E(k)-\pi^R(k)\right\|_1 \le \lim_{k\to\infty}(Z) = Z,$$

and so

$$\lim_{Z\to 0}\left(\lim_{k\to\infty}\left\|\pi^E(k)-\pi^R(k)\right\|_1\right) \le \lim_{Z\to 0}(Z) = 0,$$

which implies that

$$\lim_{k\to\infty}\left\|\pi^E(k)-\pi^R(k)\right\|_1 = 0, \text{ as } Z\to 0. \qquad\blacksquare$$

4.3 Implications of Theorem 4.4

The principal contribution of Theorem 4.4 is that it implies that matrix perturbation techniques can only be used to prove convergence of stochastic hill climbing algorithms under very restrictive conditions. This is illustrated by Theorems 4.5 and 4.6.

4.3.1 Norm of the Inverse Matrix

The key assumption of Theorem 4.4 is that $\left\|\left(P^E(k)-P^R(k)\right)A^{\#}(k)\right\|_1$ can be bounded arbitrarily close to zero, for k sufficiently large. In general, as k approaches infinity, $\left\|A^{\#}(k)\right\|_1$ becomes either unbounded or indeterminate. (See Theorem 4.5.)

To show this, each element $\left|A^{\#}_{i,j}(k)\right|$ is first expressed in terms of its cofactor and the determinant of $A(k)$, by using (4.6) and Rabenstein [1975, pg 128], to obtain

$$\left|A_{i,j}{}^{\#}(k)\right| = \left|\frac{\tilde{Q}_{j,i}(k)}{\det Q(k)} - \pi_j(k)\right|, \tag{4.9}$$

where

$$Q(k) \equiv I - P^E(k) + \Pi(k), \tag{4.10}$$

and $\widetilde{Q}_{j,i}(k)$ is the cofactor of the element $Q_{j,i}(k)$. The matrix $Q(k)$ is depicted in Figure 4.1. Lemma 4.1 states a result needed for Theorem 4.5.

Lemma 4.1 (Protter and Morrey, [1991, pg 53]: *Let* $\left|A^{\#}_{i,j}(k)\right|$ *be expressed by* (4.9) *for all* $i,j \in \Omega$ *and all iterations k, and let* $Q(k)$ *be defined by* (4.10). *If*

(a) $\lim\limits_{k\to\infty} \widetilde{Q}_{j,i}(k) = Z \in \Re,\ Z \neq 0$ *or* $\lim\limits_{k\to\infty} \widetilde{Q}_{j,i}(k) = \infty$, *and*

(b) $\det Q(k) \neq 0$ *for all iterations k, and*

(c) $\lim\limits_{k\to\infty} \det Q(k) = 0$,

then

$$\lim_{k\to\infty}\left|\frac{\widetilde{Q}_{j,i}(k)}{\det Q(k)}\right| = \infty .$$

Theorem 4.5 shows that the generalized inverse matrix norm is asymptotically unbounded or indeterminate.

$$\begin{bmatrix} \sum\limits_{j\neq 1} P^{E}_{1,j}(k) + \pi^{E}_{1}(k) & -P^{E}_{1,2}(k) + \pi^{E}_{2}(k) & .. & -P^{E}_{1,card(\Omega)}(k) + \pi^{E}_{card(\Omega)}(k) \\ -P^{E}_{2,1}(k) + \pi^{E}_{1}(k) & \sum\limits_{j\neq 2} P^{E}_{2,j}(k) + \pi^{E}_{2}(k) & .. & -P^{E}_{2,card(\Omega)}(k) + \pi^{E}_{card(\Omega)}(k) \\ \vdots & \vdots & .. & \vdots \\ \vdots & \vdots & .. & \vdots \\ -P^{E}_{card(\Omega),1}(k) + \pi^{E}_{1}(k) & -P^{E}_{card(\Omega),2}(k) + \pi^{E}_{2}(k) & .. & \sum\limits_{j\neq card(\Omega)} P^{E}_{card(\Omega),j}(k) + \pi^{E}_{card(\Omega)}(k) \end{bmatrix}$$

Figure 4.1. The matrix $Q(k)$. The elements in each row and column are ordered in terms of increasing objective function value.

Theorem 4.5: *Let* $A(k) \equiv \boldsymbol{I} - \boldsymbol{P}^E(k)$, *where* $\boldsymbol{P}^E(k)$ *is defined in Theorem 4.4. Then for any instance of a discrete optimization problem* $(\Omega, \mathcal{N}, c)$ *that contains at least one locally, but not globally optimal solution,*

$$\lim_{k\to\infty} \left\| A^{\#}(k) \right\|_1$$

is either unbounded or indeterminate.

Proof: The proof first shows that conditions (b) and (c) of Lemma 4.1 are satisfied, and concludes by examining the two possible cases for Lemma 4.1(a), as k approaches infinity.

To show that Lemma 4.1(b) holds, assume that $\det Q(k) = 0$ for some iteration k_0. Then

$$(Q(k_0))^{-1} = \left(I - P^E(k_0) + \Pi(k_0)\right)^{-1}$$

does not exist, and hence (4.6) implies that $A^{\#}(k_0)$ must not exist. But Theorem 4.2 states that the inverse $A^{\#}(k)$ exists for all iterations k, and so $\det Q(k) \neq 0$, for all k.

To show that Lemma 4.1(c) holds, assume without loss of generality that the objective function value of each solution is unique. (If not, then each solution's objective function value could be perturbed by some epsilon amount.) Then, order the elements in each row and column of the transition matrix $P(k)$ in terms of increasing objective function value. Hence $Q(k)$ approaches a lower triangular form as k increases, and thus for all $Q(k)$ terms below the main diagonal,

$$\lim_{k\to\infty}\left[-P^E_{i,j}(k)+\pi^E_j(k)\right]=\lim_{k\to\infty}\left[-g_{i,j}+\pi^E_j(k)\right]$$
$$=\begin{cases}-g_{i,j}+1 & \textit{if}\ \ j=1,\\ -g_{i,j} & \textit{otherwise.}\end{cases}$$

For all terms above the main diagonal,

$$\lim_{k\to\infty}\left[-P^E_{i,j}(k)+\pi^E_j(k)\right]=\lim_{k\to\infty}\left[-g_{i,j}\Pr\left(R_k(i,j)\geq\Delta_{i,j}\right)+\pi^E_j(k)\right]$$
$$=0,$$

and so $\det Q(k)$ approaches the product of its diagonal values as k approaches infinity. For any *locally*, but not *globally* optimal solution $i\in\Omega$, there is no solution $j\in\mathcal{N}(i)$ such that $c_j<c_i$, and so the limiting form of the corresponding diagonal element $Q_{i,i}(k)$ is

$$\begin{aligned}\lim_{k\to\infty}Q_{i,i}(k)&=\lim_{k\to\infty}\left[I_{i,i}-P^E_{i,i}(k)+\pi^E_i(k)\right]\\ &=\lim_{k\to\infty}\left[1-\left(1-\sum_{\substack{j\in\Omega,\\ j\neq i}}P^E_{i,j}(k)\right)+\pi^E_i(k)\right]\\ &=\lim_{k\to\infty}\left[\sum_{\substack{j\in\Omega,\\ j<i}}P^E_{i,j}(k)+\sum_{\substack{j\in\Omega,\\ j>i}}P^E_{i,j}(k)+\pi^E_i(k)\right]\\ &=0+\lim_{k\to\infty}\sum_{\substack{j\in\Omega,\\ j>i}}P^E_{i,j}(k)+\lim_{k\to\infty}\pi^E_i(k)\\ &=0.\end{aligned}\tag{4.11}$$

(Note that all summands are bounded and nonnegative, hence the limit of the sum is equal to the sum of the limits.) Thus for any local (but not global) optimum,

$$\lim_{k\to\infty}Q_{i,i}(k)=0\Rightarrow\ \lim_{k\to\infty}\left|\det Q(k)\right|=0,$$

and so Lemma 4.1(c) is satisfied.

Finally, consider the two possible cases for each cofactor $\tilde{Q}_{j,i}(k)$, $j,i \in \Omega$.

Case 1, $\lim_{k\to\infty} \tilde{Q}_{j,i}(k) \neq 0$: Recall that $\left|\tilde{Q}_{j,i}(k)\right|$ is defined as the absolute value of the determinant of the $(card(\Omega)-1)\times(card(\Omega)-1)$ matrix formed by deleting the j^{th} row and i^{th} column of the matrix $Q(k)$. Furthermore, since $Q(k) = \boldsymbol{I} - \boldsymbol{P}^E(k) + \Pi(k)$, and since each element of $\boldsymbol{P}^E(k)$ (and therefore $\Pi(k)$) is asymptotically monotone from (4.2) and Theorem 4.1(c), then $\lim_{k\to\infty} \tilde{Q}_{j,i}(k)$ must also be asymptotically monotone, for all $j,i \in \Omega$. Therefore, Lemma 4.1(a) is satisfied and so

$$\lim_{k\to\infty} \left|A_{i,j}{}^{\#}(k)\right| = \infty .$$

Case 2, $\lim_{k\to\infty} \tilde{Q}_{j,i}(k) = 0$: Since $\lim_{k\to\infty} \det Q(k) = 0$, then (4.9) approaches an indeterminate $0/0$ form as k approaches infinity.

Therefore from Case 1 and Case 2, $\lim_{k\to\infty} \left\|A^{\#}(k)\right\|_1$ is either unbounded or indeterminate. ■

Note that (4.8) implies that $\left\|\boldsymbol{P}^E(k) - \boldsymbol{P}^R(k)\right\|_1$ must control the convergence of $\left\|\left(\boldsymbol{P}^E(k) - \boldsymbol{P}^R(k)\right)A^{\#}(k)\right\|_1$. This in turn suggests that, with respect to k, the convergence rate of $\boldsymbol{P}^R(k)$ to $\boldsymbol{P}^E(k)$ must dominate the rate that $Q(k)$ approaches a lower triangular form. Equivalently, $\boldsymbol{P}^R(k)$ must converge to $\boldsymbol{P}^E(k)$ faster than the rate that (at least one) eigenvalue of $Q(k)$ converges to zero.

Note also that as k increases to infinity, the diagonal elements of $Q(k)$ provide a measure of the topology of the $(\Omega, \mathcal{N}, c)$ instance. This is illustrated in Lemma 4.2.

Lemma 4.2: *Suppose that for an instance* $(\Omega, \mathcal{N}, c)$,

$$\lim_{k\to\infty} Q_{i,i}(k) > 0 \text{ for every } i \in \Omega, \tag{4.12}$$

then every optimal solution must be globally optimal.

Proof (by contradiction): Assume that at least one optimal solution $j \in \Omega$ is a local, but not global, optimum. Then (4.11) holds, which contradicts (4.12), hence j must be a global optimum. ■

Note that if (4.12) holds, then a simple local search algorithm would eventually find a global optimum, with probability one. However, the solution space must be enumerated to verify that (4.12) does hold.

Finally, (4.11) implies that an $(\Omega, \mathcal{N}, c)$ instance containing multiple local optima would cause $\det Q(k)$ to decrease relatively rapidly to zero as k increases. However, the determinant would be very difficult to compute for problems with large solution space cardinalities.

4.3.2 Markov Chain Condition

The condition number of a square matrix is a measure of the distance of the matrix (in terms of a given norm) from a singular matrix (Golub and Van Loan 1989]). The degree of closeness is measured by the reciprocal of the condition number. Meyer [1980] defines the condition of the transition matrix corresponding to an irreducible, aperiodic Markov chain (indexed on each iteration *k*) as

$$cond(k) = \|A(k)\|_1 \|A^{\#}(k)\|_1,$$

where $A(k)$ is defined by (4.4), and $A^{\#}(k)$ is defined by (4.6).

One way to see that the transition matrix $A(k)$ becomes ill-conditioned, as k becomes large, is to note that for each row of $A(k)$ corresponding to a local optimum, every off-diagonal element approaches zero. This is true because every move from a local optimum is made with a hill-climbing transition probability, which must approach zero in the limit (from (4.2)). Furthermore, each diagonal term $A_{j,j}(k)$ (for a local optimum $j \in \Omega$, such that $c_i > c_j$ for all $i \in (\Omega \cap \mathcal{N}(j))$) must approach zero in the limit, since

$$A_{j,j}(k) = 1 - P_{j,j}(k) = 1 - \left(1 - \sum_{\substack{i\in\Omega,\\ i\neq j}} P_{j,i}(k)\right) = \sum_{\substack{i\in\Omega,\\ c_i>c_j}} g_{j,i}(k)\Pr\left(R_k(j,i) \geq \Delta_{j,i}\right),$$

which has limit zero as k approaches infinity. Therefore, $A(k)$ approaches a singular matrix as k approaches infinity.

Another perspective of the conditioning problem is based on a result by Ipsen and Meyer [1994]. They define an irreducible, aperiodic Markov chain (e.g., corresponding to a transition matrix $P^E(k)$), as *absolutely stable* if there is a *small* constant κ such that

$$\left|\pi_j^E(k) - \pi_j^R(k)\right| \leq \kappa \left\|P^E(k) - P^R(k)\right\|_1, \text{ for all } j \in \Omega \text{ and all iterations } k. \quad (4.13)$$

(Ipsen and Meyer note that the degree of *smallness* is problem-specific.) They then show how their definition of stability can be used to assess the condition of a Markov chain transition matrix. Theorem 4.6 provides this relationship.

Theorem 4.6: *For an n-state irreducible Markov chain, the chain is absolutely stable if and only if all entries of the group inverse* $A^{\#}(k)$ *are small, where smallness is as defined in Ipsen and Meyer [1994].*

Proof: Define the group inverse of Ipsen and Meyer's [1994, Theorem 5.2] as a function of k. ■

Theorem 4.5 shows that for optimization problems with multiple local (but not global) optima, the inverse matrix norm $\left\|A^{\#}(k)\right\|_1$ becomes either unbounded or indeterminate in the limit. Theorems 4.5 and 4.6 together suggest that the Markov chain corresponding to the matrix $A(k)$ becomes absolutely unstable as k approaches infinity, and so in general, a small constant κ does not exist that satisfies (4.13). Hence the bounds (4.7) and (4.13) are useful only for proving the convergence of a limited class of GHC algorithms.

4.4 Illustrative Example

The following example, adapted from Anily and Federgruen [1987], demonstrates how the rate that the two transition matrices converge must dominate the matrix condition number growth.

Let $P^E(k)$ be defined by (3.2) and example (3.3.1), where t_k is defined in (2.2) and $g_{i,j}(k)$ satisfies conditions (a) and (b) of Theorem 3.1. Note that Theorem 3.1 and example (3.3.1) give the result

$$\pi_i^E(k) = \frac{\exp(-\Delta_{opt,i}/t_k)}{\sum_{n\in\Omega}\exp(-\Delta_{opt,n}/t_k)} \text{ for all } i \in \Omega \text{, and all } k.$$

Define $P^R(k)$ by (4.3), and let the solution generation probabilities corresponding to $P^R(k)$ be identical to those of $P^E(k)$. Furthermore, let

$$\begin{aligned} P_{i,j}^R(k) &= g_{i,j}(k)\Pr\left(R_k(i,j) \geq \Delta_{i,j}\right) \\ &= g_{i,j}(k)\min\left\{1, \exp(-\Delta_{i,j}/t_k) + (t_k)^2\exp\left[-\Delta_{i,j}/(t_k)^\gamma\right]\right\}, \end{aligned} \qquad (4.14)$$

for all $i,j \in \Omega$, such that $i \neq j$ and any $\gamma > 1$. Note that the term

$$(t_k)^2\exp\left[-\Delta_{i,j}/(t_k)^\gamma\right]$$

dominates (4.14) for k small, while the SA term dominates for k large.

Consider the four-solution problem depicted in Figure 4.2. Note that solution p is optimal, and that $\lim_{k\to\infty}\pi_p^E(k) = 1$. The lines connecting the solutions indicate the neighborhood structure. The matrix $P^E(k) - P^R(k)$ is computed (see Figure 4.3). The matrix $A^\#(k)$ is then calculated. $N(t_k) = 1 + \exp(-1/t_k) + 2\exp(-2/t_k)$ is a normalizing value for the stationary distribution $\pi(k)$. The matrix

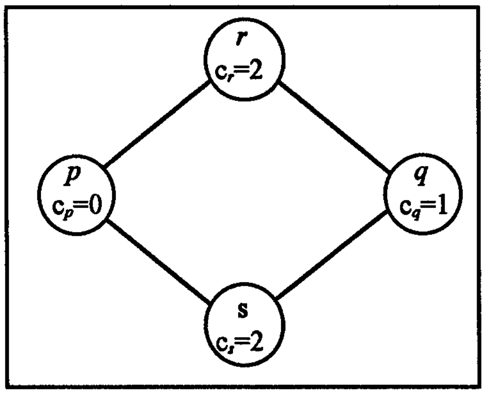

Figure 4.2. Sample Problem. The neighborhood structure is shown by the lines.

$$\begin{bmatrix} (t_k)^2 \exp\left(-2/(t_k)^\gamma\right) & 0 & -0.5(t_k)^2 \exp\left(-2/(t_k)^\gamma\right) & -0.5(t_k)^2 \exp\left(-2/(t_k)^\gamma\right) \\ 0 & (t_k)^2 \exp\left(-1/(t_k)^\gamma\right) & -0.5(t_k)^2 \exp\left(-1/(t_k)^\gamma\right) & -0.5(t_k)^2 \exp\left(-1/(t_k)^\gamma\right) \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Figure 4.3. The matrix $\boldsymbol{P}^E(k) - \boldsymbol{P}^R(k)$.

The rows and columns are arranged in order p, q, r, s.

$$\begin{bmatrix} \exp\left(\frac{-2}{t_k}\right)+\frac{1}{N(t_k)} & \frac{\exp\left(\frac{-1}{t_k}\right)}{N(t_k)} & -0.5\exp\left(\frac{-2}{t_k}\right)+\frac{\exp\left(\frac{-2}{t_k}\right)}{N(t_k)} & -0.5\exp\left(\frac{-2}{t_k}\right)+\frac{\exp\left(\frac{-2}{t_k}\right)}{N(t_k)} \\ \frac{1}{N(t_k)} & \exp\left(\frac{-1}{t_k}\right)+\frac{\exp\left(\frac{-1}{t_k}\right)}{N(t_k)} & -0.5\exp\left(\frac{-1}{t_k}\right)+\frac{\exp\left(\frac{-2}{t_k}\right)}{N(t_k)} & -0.5\exp\left(\frac{-1}{t_k}\right)+\frac{\exp\left(\frac{-2}{t_k}\right)}{N(t_k)} \\ -0.5+\frac{1}{N(t_k)} & -0.5+\frac{\exp\left(\frac{-1}{t_k}\right)}{N(t_k)} & 1+\frac{\exp\left(\frac{-2}{t_k}\right)}{N(t_k)} & \frac{\exp\left(\frac{-2}{t_k}\right)}{N(t_k)} \\ -0.5+\frac{1}{N(t_k)} & -0.5+\frac{\exp\left(\frac{-1}{t_k}\right)}{N(t_k)} & \frac{\exp\left(\frac{-2}{t_k}\right)}{N(t_k)} & 1+\frac{\exp\left(\frac{-2}{t_k}\right)}{N(t_k)} \end{bmatrix}$$

Figure 4.4. The matrix $Q(k)$. Solutions are in the same order as in Figure 4.3.

$Q(k) = \boldsymbol{I} - \boldsymbol{P}^E(k) + \Pi(k)$ is depicted in Figure 4.4. The determinant of $Q(k)$ is (using Mathematica)

$$\begin{aligned} \det Q(k) = {} & \frac{2\exp(-5/t_k)}{N(t_k)} + \frac{2\exp(-4/t_k)}{N(t_k)} \\ & + \frac{2.5\exp(-3/t_k)}{N(t_k)} + \frac{\exp(-2/t_k)}{N(t_k)} + \frac{0.5\exp(-1/t_k)}{N(t_k)}, \\ \geq {} & \frac{0.5\exp(-1/t_k)}{N(t_k)}. \end{aligned} \tag{4.15}$$

(Note that $\lim_{k\to\infty} \det Q(k) = 0$, as noted by (4.11).) The matrix $A^\#(k)$ is shown element-wise as

$$A^{\#}_{1,1}(k)=\left[1.5\exp\left(\frac{-1}{t_k}\right)+\frac{2\exp\left(\frac{-3}{t_k}\right)}{N(t_k)}+\frac{2\exp\left(\frac{-2}{t_k}\right)}{N(t_k)}+\frac{\exp\left(\frac{-1}{t_k}\right)}{N(t_k)}\right]\frac{1}{\det Q(k)}-\frac{1}{N(t_k)},$$

$$A^{\#}_{1,2}(k)=\left[0.5\exp\left(\frac{-2}{t_k}\right)-\frac{\exp\left(\frac{-3}{t_k}\right)}{N(t_k)}-\frac{\exp\left(\frac{-2}{t_k}\right)}{N(t_k)}-\frac{\exp\left(\frac{-1}{t_k}\right)}{N(t_k)}\right]\frac{1}{\det Q(k)}-\frac{\exp\left(\frac{-1}{t_k}\right)}{N(t_k)},$$

$$A^{\#}_{1,3}(k)=\left[0.5\exp\left(\frac{-3}{t_k}\right)-\frac{0.5\exp\left(\frac{-3}{t_k}\right)}{N(t_k)}-\frac{0.5\exp\left(\frac{-2}{t_k}\right)}{N(t_k)}\right]\frac{1}{\det Q(k)}-\frac{\exp\left(\frac{-2}{t_k}\right)}{N(t_k)},$$

$$A^{\#}_{1,4}(k)=\left[0.5\exp\left(\frac{-3}{t_k}\right)-\frac{0.5\exp\left(\frac{-3}{t_k}\right)}{N(t_k)}-\frac{0.5\exp\left(\frac{-2}{t_k}\right)}{N(t_k)}\right]\frac{1}{\det Q(k)}-\frac{\exp\left(\frac{-2}{t_k}\right)}{N(t_k)},$$

$$A^{\#}_{2,1}(k)=\left[0.5\exp\left(\frac{-1}{t_k}\right)-\frac{1}{N(t_k)}-\frac{\exp\left(\frac{-2}{t_k}\right)}{N(t_k)}-\frac{\exp\left(\frac{-1}{t_k}\right)}{N(t_k)}\right]\frac{1}{\det Q(k)}-\frac{1}{N(t_k)},$$

$$A^{\#}_{2,2}(k)=\left[0.5\exp\left(\frac{-2}{t_k}\right)+\frac{1}{N(t_k)}+\frac{2\exp\left(\frac{-4}{t_k}\right)}{N(t_k)}+\frac{2\exp\left(\frac{-2}{t_k}\right)}{N(t_k)}\right]\frac{1}{\det Q(k)}-\frac{\exp\left(\frac{-1}{t_k}\right)}{N(t_k)},$$

$$A^{\#}_{2,3}(k)=\left[0.5\exp\left(\frac{-3}{t_k}\right)-\frac{\exp\left(\frac{-4}{t_k}\right)}{N(t_k)}-\frac{0.5\exp\left(\frac{-2}{t_k}\right)}{N(t_k)}+\frac{0.5\exp\left(\frac{-1}{t_k}\right)}{N(t_k)}\right]\frac{1}{\det Q(k)}-\frac{\exp\left(\frac{-2}{t_k}\right)}{N(t_k)},$$

$$A^{\#}_{2,4}(k)=\left[0.5\exp\left(\frac{-3}{t_k}\right)-\frac{\exp\left(\frac{-4}{t_k}\right)}{N(t_k)}-\frac{0.5\exp\left(\frac{-2}{t_k}\right)}{N(t_k)}+\frac{0.5\exp\left(\frac{-1}{t_k}\right)}{N(t_k)}\right]\frac{1}{\det Q(k)}-\frac{\exp\left(\frac{-2}{t_k}\right)}{N(t_k)},$$

$$A^{\#}_{3,1}(k)=\left[0.5\exp\left(\frac{-1}{t_k}\right)-\frac{0.5}{N(t_k)}-\frac{0.5\exp\left(\frac{-1}{t_k}\right)}{N(t_k)}\right]\frac{1}{\det Q(k)}-\frac{1}{N(t_k)},$$

$$A^{\#}_{3,2}(k)=\left[0.5\exp\left(\frac{-2}{t_k}\right)+\frac{0.5}{N(t_k)}-\frac{\exp\left(\frac{-3}{t_k}\right)}{N(t_k)}-\frac{0.5\exp\left(\frac{-1}{t_k}\right)}{N(t_k)}\right]\frac{1}{\det Q(k)}-\frac{\exp\left(\frac{-1}{t_k}\right)}{N(t_k)},$$

$$\begin{aligned}A^{\#}_{3,3}(k)=\Bigg[&0.5\exp\left(\frac{-3}{t_k}\right)+\frac{\exp\left(\frac{-5}{t_k}\right)}{N(t_k)}+\frac{\exp\left(\frac{-4}{t_k}\right)}{N(t_k)}\\&+\frac{1.75\exp\left(\frac{-3}{t_k}\right)}{N(t_k)}+\frac{0.5\exp\left(\frac{-2}{t_k}\right)}{N(t_k)}+\frac{0.75\exp\left(\frac{-1}{t_k}\right)}{N(t_k)}\Bigg]\frac{1}{\det Q(k)}-\frac{\exp\left(\frac{-2}{t_k}\right)}{N(t_k)},\end{aligned}$$

$$A^{\#}_{3,4}(k)=\left[0.5\exp\left(\frac{-3}{t_k}\right)-\frac{\exp\left(\frac{-5}{t_k}\right)}{N(t_k)}-\frac{\exp\left(\frac{-4}{t_k}\right)}{N(t_k)}\right.$$

$$\left.-\frac{0.75\exp\left(\frac{-3}{t_k}\right)}{N(t_k)}-\frac{0.5\exp\left(\frac{-2}{t_k}\right)}{N(t_k)}+\frac{0.25\exp\left(\frac{-1}{t_k}\right)}{N(t_k)}\right]\frac{1}{\det Q(k)}-\frac{\exp\left(\frac{-2}{t_k}\right)}{N(t_k)},$$

$$A^{\#}_{4,1}(k)=\left[0.5\exp\left(\frac{-1}{t_k}\right)-\frac{0.5}{N(t_k)}-\frac{0.5\exp\left(\frac{-1}{t_k}\right)}{N(t_k)}\right]\frac{1}{\det Q(k)}-\frac{1}{N(t_k)},$$

$$A^{\#}_{4,2}(k)=\left[0.5\exp\left(\frac{-2}{t_k}\right)+\frac{0.5}{N(t_k)}-\frac{\exp\left(\frac{-3}{t_k}\right)}{N(t_k)}-\frac{0.5\exp\left(\frac{-1}{t_k}\right)}{N(t_k)}\right]\frac{1}{\det Q(k)}-\frac{\exp\left(\frac{-1}{t_k}\right)}{N(t_k)},$$

$$A^{\#}_{4,3}(k)=\left[0.5\exp\left(\frac{-3}{t_k}\right)-\frac{\exp\left(\frac{-5}{t_k}\right)}{N(t_k)}-\frac{\exp\left(\frac{-4}{t_k}\right)}{N(t_k)}\right.$$

$$\left.-\frac{0.75\exp\left(\frac{-3}{t_k}\right)}{N(t_k)}-\frac{0.5\exp\left(\frac{-2}{t_k}\right)}{N(t_k)}+\frac{0.25\exp\left(\frac{-1}{t_k}\right)}{N(t_k)}\right]\frac{1}{\det Q(k)}-\frac{\exp\left(\frac{-2}{t_k}\right)}{N(t_k)},$$

$$A^{\#}_{4,4}(k)=\left[0.5\exp\left(\frac{-3}{t_k}\right)+\frac{\exp\left(\frac{-5}{t_k}\right)}{N(t_k)}+\frac{\exp\left(\frac{-4}{t_k}\right)}{N(t_k)}\right.$$

$$\left.+\frac{1.75\exp\left(\frac{-3}{t_k}\right)}{N(t_k)}+\frac{0.5\exp\left(\frac{-2}{t_k}\right)}{N(t_k)}+\frac{0.75\exp\left(\frac{-1}{t_k}\right)}{N(t_k)}\right]\frac{1}{\det Q(k)}-\frac{\exp\left(\frac{-2}{t_k}\right)}{N(t_k)}.$$

Note that

$$\left\|\left(P^{E}(k)-P^{R}(k)\right)A^{\#}(k)\right\|_1\leq\left\|\left(P^{E}(k)-P^{R}(k)\right)\right\|_1\left\|A^{\#}(k)\right\|_1$$

(Golub and Van Loan [1989]. Furthermore, for the example depicted in Figure 4.2,

$$\left\|\left(P^{E}(k)-P^{R}(k)\right)\right\|_1=2\left(t_k\right)^2\exp\left(-1/\left(t_k\right)^{\gamma}\right)\quad\text{for all } k. \tag{4.16}$$

Using the lower bound (4.15) for $\det Q(k)$, there exists some k_0 such that for all $k>k_0$ and all $i,j\in\Omega$,

$$\left|A^{\#}_{i,j}(k)\right|\leq\left(\frac{2}{N(t_k)}\right)\left(\frac{1}{\frac{0.5\exp\left(-1/t_k\right)}{N(t_k)}}\right),$$

and so (accounting for the four solutions in Ω), the desired upper bound is

$$\left\|A^{\#}(k)\right\|_1\leq 4\left(4\exp(1/t_k)\right). \tag{4.17}$$

Therefore, (4.16) and (4.17) lead to

$$\left\|\left(P^{E}(k)-P^{R}(k)\right)\right\|_1\left\|A^{\#}(k)\right\|_1\leq 2\left(t_k\right)^2\exp\left(-1/\left(t_k\right)^{\gamma}\right)16\exp\left(1/t_k\right)$$

$$= 32(t_k)^2 \exp\left(1/t_k - 1/(t_k)^\gamma\right). \qquad (4.18)$$

Finally, taking the limit of (4.18) as k approaches infinity, and using condition (2.2) leads to

$$\lim_{t_k \to 0}\left[32(t_k)^2 \exp\left(1/t_k - 1/(t_k)^\gamma\right)\right] = 0,$$

and so

$$\lim_{k\to\infty} \pi_p^R(k) = 1,$$

which is the desired result.

Note that if $\gamma < 1$, then

$$\lim_{k\to\infty} \left\|\left(\boldsymbol{P}^E(k) - \boldsymbol{P}^R(k)\right)\right\|_1 \left\|A^\#(k)\right\|_1 = \infty,$$

and so condition (4.8) is not satisfied. Therefore if $\gamma < 1$, then Theorem 4.4 could not be used to prove convergence for this example.

CHAPTER 5: GENERAL PROOF OF CONVERGENCE FOR THE GENERALIZED HILL CLIMBING ALGORITHM

Chapter 5 provides a convergence proof for GHC algorithms, based on sufficient conditions for asymptotic transition probabilities between local and global minima. The principal contribution of this proof is the relaxation of the Markov chain reversibility condition, without requiring the special conditions on the GHC algorithm that are needed for the convergence proof in Chapter 4.

5.1 Definitions and Notation

Define $G \subset \Omega$ to be the set of globally optimal solutions (i.e., $i \in G$ if $c_i \leq c_j$ for all $j \in \Omega$). Define $L \subset \Omega \setminus G$ to be the set of locally, but not globally, optimal solutions (i.e., $i \in L$ if $c_i \leq c_j$ for all $j \in \mathcal{N}(i)$). Finally, define $H = \Omega \setminus (L \cup G)$ to be the set of all other solutions in Ω.

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 5.1: A *path* from *i* to *j*, for all $i, j \in (L \cup G)$, 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 $m = 0, 1, \ldots, d-1$, and for all iterations *k*.

Note that a local or global optimum cannot be an intermediate solution on any path. (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.) Two paths can be either *equivalent*, or *distinct*. These concepts are formally defined.

Definition 5.2: Two paths between solutions $i, j \in (L \cup G)$ are said to be *equivalent*, 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*.

Using these definitions, the probability of transitioning between solutions $i, j \in (L \cup G)$ can be defined. In particular, $P_k\left(i \overset{n}{\rightarrow} j\right)$ is the probability of transitioning along the *n*th (distinct) path between $i, j \in (L \cup G)$, at iteration *k*. $P_k\left(i \overset{n}{\rightarrow} j\right)$ is equal to the product of all one-step transition probabilities between adjacent solutions along the *n*th path; e.g., for $i, j \in \Omega$, the *n*th path $i, l_1, \ldots, l_{d-1}, j \in \Omega$ occurs with probability

$$P_k\left(i \overset{n}{\rightarrow} j\right) = \prod_{m=0}^{d-1} P_{l_m, l_{m+1}}(k) \ . \tag{5.1}$$

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 (Cinlar [1975, pg 3]), and so

$$P_k(i \to j) \equiv \begin{cases} \displaystyle\sum_{n=1}^{\substack{\textit{all distinct} \\ \textit{paths between} \\ i,j \in (L \cup G)}} P_k\left(i \overset{n}{\to} j\right) & \textit{for all } i,j \in (L \cup G),\ i \neq j, \\ 1 - \displaystyle\sum_{\substack{s \in (L \cup G) \\ s \neq i}} P_k(i \to s) & i = j, \\ 0 & \textit{otherwise.} \end{cases} \tag{5.2}$$

The example depicted in Figure 5.1 illustrates how the path probabilities (5.2) are defined. Note that $i,l \in L$, $j \in G$, and $p,q,r,s \in H$. The neighborhood structure is indicated by the lines between nodes. Hence $P_k(i \to j) > 0$, since *i* and *j* are separated only by q. Similarly, $P_k(l \to j) > 0$, 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 $P_k(i \to l) = 0$.

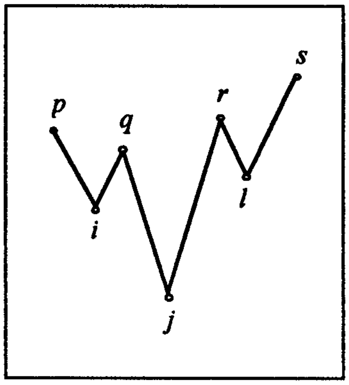

Figure 5.1. Example problem.

Recall that the GHC algorithm is composed of an outer loop, indexed on k, and an inner loop, indexed on m. Furthermore, $\pi(k)$ is the equilibrium (long-run) probability vector for all solutions $i \in \Omega$, for each k, as $m \rightarrow \infty$.

Definition 5.3: Let $\omega(k) \equiv \sum_{i \in (L \cup G)} \pi_i(k)$, and define the equilibrium probability

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

Let $\delta(k)$ be the vector of probabilities $\delta_i(k)$, for all solutions $i \in (L \cup G)$. Note that $\delta_i(k)$ is obtained by scaling the corresponding equilibrium probability $\pi_i(k)$ only for the solutions $i \in (L \cup G)$. Furthermore, $\delta_i(k) \geq \pi_i(k)$ for all k, since $\omega(k) \leq 1$.

Finally, note that (5.1) and definition 5.3 allow the convergence proof (Theorem 5.2) to focus only on the set of local and global optima, which are really the only solutions of interest for a GHC algorithm.

5.2 Proof of Convergence

Theorem 5.1 provides sufficient conditions for the GHC algorithm to converge to the set of solutions $(L \cup G)$ as k approaches infinity. Theorem 5.2 provides additional sufficient conditions for the GHC algorithm to converge to the set of globally optimal solutions $G \subset \Omega$, as k approaches infinity. Note that Theorems 5.1 and 5.2 do not use the Markov chain model to prove convergence of the GHC algorithm. However, the property of conditional independence continues to hold: for all solutions $j \in \Omega$, the probability of transitioning to the next solution is dependent only on the current solution.

Theorem 5.1: *Let $(\Omega, \mathcal{N}, c)$ denote an instance of a discrete optimization problem. Let each* GHC *transition probability* $P_{i,j}(k)$ *be defined by* (3.2). *Let the generation probabilities* $g_{i,j}(k)$ *satisfy* (3.3) *and the conditions*

(a) *for all $i, j \in \Omega$ and all iterations k, there exists an integer $d \geq 1$ 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) $\lim_{k \to \infty} g_{i,j}(k) > 0$ *for all $i, j \in \Omega,\ j \in \mathcal{N}(i)$.*

Moreover, let the acceptance probabilities satisfy

(c) $\Pr\left(R_k(i,j) \geq \Delta_{i,j}\right) > 0$ *for all $i, j \in \Omega$ and all iterations k,*

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

Then

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

Proof: (First, recall that conditions (a) and (c) are sufficient for a unique distribution $\pi(k)$ to exist for each k, as proven in Theorem 3.1.) The proof is by contradiction. Assume that all solutions in H have strictly positive probability in the limit, and order the $h = card(H)$ solutions in H such that for all $i, j \in H$, $i < j$ implies $c_i < c_j$. (Without loss of generality, let the objective function value of each solution in H be unique). Let solution number one correspond to the solution in H with the smallest objective function value, solution two correspond to the next smallest, and so on, with solution h

corresponding to the solution in H with the largest objective function value. Then, the corresponding equilibrium probabilities can be expressed in terms of h equations, using the law of total probability (Cinlar [1975, pg15]).

The proof first shows that the equilibrium probability $\lim_{k\to\infty}\pi_h(k)$ is zero for the h^{th} solution, and then establishes the equilibrium probabilities of the remaining $(h-1)$ solutions to be zero, using backward substitution.

First, express the equilibrium probability vector $\pi(k)$ in terms of the law of total probability, by conditioning on every solution $j \in \Omega$, to obtain

$$\pi_i(k) = \sum_{j\in\Omega} \pi_j(k) P_{j,i}(k), \text{ for all } i \in \Omega \text{ and all } k. \tag{5.3}$$

Next, consider the set of solutions $H \subset \Omega$, and assume that

$$\lim_{k\to\infty} \pi_i(k) = \varepsilon_i > 0, \text{ for all } i \in H. \tag{5.4}$$

For any solution $i \in H$, the equilibrium probability $\pi_i(k)$ is expressed in terms of (5.3) 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 $\pi_i(k)$ terms on the left-hand side to obtain

$$\pi_i(k)\left(1 - P_{i,i}(k)\right) = \sum_{j\in G} \pi_j(k) P_{j,i}(k) + \sum_{j\in L} \pi_j(k) P_{j,i}(k) + \sum_{\substack{j\in H\\ j\neq i}} \pi_j(k) P_{j,i}(k). \tag{5.5}$$

Note that

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

(e.g., $\left(1 - P_{i,i}(k)\right)$ is the probability that the process does not remain at solution $i \in H$ in the next transition). Hence, (5.5) can be expressed as

$$\pi_i(k)\left(\sum_{\substack{j\in\Omega,\\ i\neq j,\\ c_i\geq c_j}} P_{i,j}(k) + \sum_{\substack{j\in\Omega,\\ c_i<c_j}} P_{i,j}(k)\right) = \sum_{j\in G}\pi_j(k)P_{j,i}(k) + \sum_{j\in L}\pi_j(k)P_{j,i}(k) + \sum_{\substack{j\in H\\ j\neq i}}\pi_j(k)P_{j,i}(k).$$

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

$$\begin{aligned}\pi_i(k)\left(\sum_{\substack{j\in\Omega,\\ i\neq j,\\ c_i\geq c_j}} P_{i,j}(k) + \sum_{\substack{j\in\Omega,\\ c_i<c_j}} P_{i,j}(k)\right) &= \sum_{j\in G}\pi_j(k)P_{j,i}(k) + \sum_{j\in L}\pi_j(k)P_{j,i}(k) \\ &\quad + \sum_{\substack{j\in H,\\ j\neq i,\\ c_j\geq c_i}}\pi_j(k)P_{j,i}(k) + \sum_{\substack{j\in H,\\ c_j<c_i}}\pi_j(k)P_{j,i}(k).\end{aligned}$$

Note that since all sums are over a finite number of bounded, nonnegative elements, then the limit of each term exists, and the limit of the sums is equal to the sum of the limits (Protter and Morrey [1991, pg 37]). Therefore,

$$\begin{aligned}&\lim_{k\to\infty}\left(\pi_i(k)\sum_{\substack{j\in\Omega,\\ i\neq j,\\ c_i\geq c_j}} P_{i,j}(k)\right) + \lim_{k\to\infty}\left(\pi_i(k)\sum_{\substack{j\in\Omega,\\ c_i<c_j}} P_{i,j}(k)\right) \\ &\qquad = \lim_{k\to\infty}\left(\sum_{j\in G}\pi_j(k)P_{j,i}(k)\right) + \lim_{k\to\infty}\left(\sum_{j\in L}\pi_j(k)P_{j,i}(k)\right) \qquad (5.6) \\ &\qquad + \lim_{k\to\infty}\left(\sum_{\substack{j\in H,\\ j\neq i,\\ c_j\geq c_i}}\pi_j(k)P_{j,i}(k)\right) + \lim_{k\to\infty}\left(\sum_{\substack{j\in H,\\ c_j<c_i}}\pi_j(k)P_{j,i}(k)\right).\end{aligned}$$

From (3.2) and condition (d), 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 summation includes only deteriorating (hill climbing) transition probabilities. Therefore in the limit, (5.6) simplifies to

$$\lim_{k\to\infty}\left(\pi_i(k)\sum_{\substack{j\in\Omega,\\ i\neq j,\\ c_i\geq c_j}}P_{i,j}(k)\right)=\lim_{k\to\infty}\left(\sum_{\substack{j\in H,\\ j\neq i,\\ c_j\geq c_i}}\pi_j(k)P_{j,i}(k)\right),\text{ for all } i\in H. \qquad (5.7)$$

Consider (5.7) for the particular solution $i=h\in H$ (recall that h is the solution of maximal objective function value, e.g., $c_h>c_j$ for all $j\in H\setminus\{h\}$, since all solutions are arranged in order of increasing objective function value). The right-hand side of (5.7) is zero because there are no solutions $j\in H$ of cost greater than that of solution h.

Recall that $\lim_{k\to\infty}\pi_h(k)=\varepsilon_h>0$, from assumption (5.4). Furthermore, since h is by definition not a local minimum, then there must exist at least one solution $l\in\Omega$, $l\in\mathcal{N}(h)$, such that $c_h\geq c_l$. Therefore, from (3.2) and condition (b), the left-hand side of (5.7) is

$$\begin{aligned}\lim_{k\to\infty}\left(\pi_h(k)\sum_{\substack{j\in\Omega,\\ h\neq j,\\ c_h\geq c_j}}P_{h,j}(k)\right)&=\varepsilon_h\lim_{k\to\infty}\sum_{\substack{j\in\Omega,\\ j\in\mathcal{N}(h),\\ c_h\geq c_j}}g_{h,j}(k)\\ &\geq\varepsilon_h\lim_{k\to\infty}g_{h,l}(k)\\ &>0,\end{aligned}$$

which is a contradiction, since the right-hand side of (5.7) has limit zero. Thus

$$\lim_{k\to\infty}\pi_h(k)=0\,. \tag{5.8}$$

Now express (5.7) in terms of the particular solutions h and $(h-1)\in H$ to obtain

$$\lim_{k\to\infty}\left(\pi_{(h-1)}(k) \sum_{\substack{j\in\Omega,\\ (h-1)\neq j,\\ c_{(h-1)}\geq c_j}} P_{(h-1),j}(k)\right) = \lim_{k\to\infty}\pi_h(k)P_{h,(h-1)}(k)\,. \tag{5.9}$$

The right-hand side of (5.9) is equal to zero, from (5.8). Now consider the left-hand side of (5.9). Since the particular element $(h-1)\in H$ is the solution of maximal objective function value in the set $H\setminus\{h\}$, hence it is not a local minimum, then there must exist at least one solution $l\in\Omega$, $l\in\mathcal{N}(h-1)$, such that $c_{(h-1)}\geq c_l$. Therefore, from (3.2), condition (b), and (5.4),

$$\begin{aligned}\lim_{k\to\infty}\left(\pi_{(h-1)}(k) \sum_{\substack{j\in\Omega,\\ (h-1)\neq j,\\ c_{(h-1)}\geq c_j}} P_{(h-1),j}(k)\right) &= \varepsilon_{(h-1)}\lim_{k\to\infty}\sum_{\substack{j\in\Omega,\\ j\in\mathcal{N}(h-1),\\ c_{(h-1)}\geq c_j}} g_{(h-1),j}(k)\\ &\geq \varepsilon_{(h-1)}\lim_{k\to\infty} g_{(h-1),l}(k)\\ &> 0,\end{aligned}$$

which is a contradiction, because the right-hand side of (5.9) has limit zero. Thus

$$\lim_{k\to\infty}\pi_{(h-1)}(k)=0\,.$$

Continuing in this manner, let

$$\lim_{k\to\infty}\pi_j(k)=0\,,\ \text{for all } j=2,3,\ldots,h,\quad j\in H, \tag{5.10}$$

and replace (5.4) with the assumption that

$$\lim_{k\to\infty} \pi_1(k) = \varepsilon_1 > 0\text{, for solution } 1 \in H. \tag{5.11}$$

Note that element $1 \in H$ is the solution of minimum objective function value in the set H. Rewrite (5.7) in terms of the particular solution $1 \in H$ to obtain

$$\lim_{k\to\infty}\left(\pi_1(k) \sum_{\substack{j\in\Omega,\\ 1\neq j,\\ c_1\geq c_j}} P_{1,j}(k) \right) = \lim_{k\to\infty}\left(\sum_{\substack{j\in H,\\ j\neq 1,\\ c_j\geq c_1}} \pi_j(k) P_{j,1}(k) \right). \tag{5.12}$$

The right-hand side of (5.12) is equal to zero, from (5.10). Now consider the left-hand side of (5.12). Since the particular solution $1 \in H$ can not also be in the set $(L \cup G)$, hence it is not a local minimum, and so there must exist at least one solution $l \in \Omega$, $l \in \mathcal{N}(1)$, such that $c_1 \geq c_l$. Therefore, from (3.2), condition (b), and (5.11),

$$\begin{aligned}
\lim_{k\to\infty}\left(\pi_1(k) \sum_{\substack{j\in\Omega,\\ 1\neq j,\\ c_1\geq c_j}} P_{1,j}(k) \right) &= \varepsilon_1 \lim_{k\to\infty} \sum_{\substack{j\in\Omega,\\ j\in\mathcal{N}(1),\\ c_1\geq c_j}} g_{1,j}(k) \\
&\geq \varepsilon_1 \lim_{k\to\infty} g_{1,l}(k) \\
&> 0,
\end{aligned}$$

which is a contradiction, because the right-hand side of (5.12) has limit zero. Thus

$$\lim_{k\to\infty} \pi_1(k) = 0.$$

Therefore,

$$\lim_{k\to\infty} \pi_i(k) = 0\text{, for all } i \in H. \qquad \blacksquare$$

Corollary 5.1 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 $k \to \infty$.

Corollary 5.1: *For all* $i \in (L \cup G)$,

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

Proof: Taking the limit of (2.6) as $k \to \infty$ leads to

$$\begin{aligned}
1 &= \lim_{k\to\infty}\left(\sum_{i\in H}\pi_i(k) + \sum_{i\in(L\cup G)}\pi_i(k)\right) \\
&= \sum_{i\in H}\lim_{k\to\infty}\pi_i(k) + \sum_{i\in(L\cup G)}\lim_{k\to\infty}\pi_i(k) \\
&= 0 + \sum_{i\in(L\cup G)}\lim_{k\to\infty}\pi_i(k) \\
&= \lim_{k\to\infty}\sum_{i\in(L\cup G)}\pi_i(k) \\
&= \lim_{k\to\infty}\omega(k).
\end{aligned}$$

Since $\lim_{k\to\infty}\omega(k) = 1$ and since $0 \le \lim_{k\to\infty}\pi_i(k) \le 1$ for all $i \in \Omega$, hence the limit of the quotient $\pi_i(k)/\omega(k)$ is equal to the quotient of the limits (Protter and Morrey [1991, pg 39], and so

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

Theorem 5.2 provides sufficient conditions for the equilibrium probability of all local (but not global) optima to approach zero, as *k* approaches infinity.

Theorem 5.2: *Under the conditions and assumptions of Theorem 5.1, if*

(e) $\sum_{k=1}^{\infty} P_k(j \to i) = {}^{+}\infty$ *for all* $j \in L,\ i \in \left(L \cup G\right)$,

such that $P_k(j \to i) > 0$ *for all k,*

(f) $\sum_{k=1}^{\infty} P_k(i \to j) < {}^{+}\infty$ *for all* $i \in G,\ j \in L$,

(g) $\sum_{k=1}^{\infty}\delta_j(k)P_k(j \to q) < {}^{+}\infty$ *for all* $j,q \in L,\ q \neq j,\ q \notin \mathcal{N}(j)$,

then

$$\lim_{k\to\infty}\delta_j(k) = 0 \textit{ for all } j \in L. \tag{5.13}$$

Proof (by contradiction): First, each element of $\delta(k)$ is expressed in terms of the law of total probability, using the path probabilities (5.2). Hence for each iteration *k*,

$$\delta_j(k) = \sum_{i\in G}\delta_i(k)P_k(i \to j) + \sum_{i\in L}\delta_i(k)P_k(i \to j) \text{ for all } j \in L \cup G. \tag{5.14}$$

Next, assume there exists some $j \in L$ and an iteration k_0 such that for all $k \geq k_0$, $\delta_j(k) \geq \varepsilon > 0$. 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

$$\sum_{k=1}^{\infty}\delta_j(k)\left(1 - P_k(j \to j)\right) = \sum_{i\in G}\sum_{k=1}^{\infty}\delta_i(k)P_k(i \to j) + \sum_{\substack{i\in L,\, k=1 \\ i\neq j}}\sum^{\infty}\delta_i(k)P_k(i \to j). \tag{5.15}$$

Note that $\left(1 - P_k(j \to j)\right)$ is the probability that, given the process is in solution $j \in L$, the process transitions to any solution $i \in \left(L \cup G\right)$ except solution *j*. (Note that since all solutions communicate (from Theorem 5.1), a path must exist 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,

$$P_k(j \to q) \leq \left(1 - P_k(j \to j)\right),$$

and so (5.15) becomes

$$\sum_{k=1}^{\infty} \delta_j(k) P_k(j \to q) \leq \sum_{i \in G} \sum_{k=1}^{\infty} \delta_i(k) P_k(i \to j) + \sum_{\substack{i \in L,\\ i \neq j}} \sum_{k=1}^{\infty} \delta_i(k) P_k(i \to j). \qquad (5.16)$$

Since $\delta_i(k) \leq 1$ for all $i \in \Omega$ and all k, then (5.16) can be rewritten as

$$\sum_{k=1}^{k_0} \delta_j(k) P_k(j \to q) + \sum_{k=k_0+1}^{\infty} \varepsilon P_k(j \to q) \leq \sum_{i \in G} \sum_{k=1}^{\infty} P_k(i \to j) + \sum_{\substack{i \in L,\\ i \neq j}} \sum_{k=1}^{\infty} \delta_i(k) P_k(i \to j). \qquad (5.17)$$

For the left-hand side of (5.17), condition (e) leads to

$$\sum_{k=k_0+1}^{\infty} \varepsilon P_k(j \to q) = \infty. \qquad (5.18)$$

Now consider the right-hand side of (5.17). Conditions (f) and (g) and the fact that Ω is finite leads to

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

which contradicts (5.18). Therefore there cannot exist any iteration k_0 such that (5.18) 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,

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

and so

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

5.3 Implications of Theorems 5.1 and 5.2

Theorems 5.1 and 5.2, and Corollary 5.1, together prove that under certain conditions, the set of globally optimal solutions G must occur with probability one as k approaches infinity, since the equilibrium probabilities for any solution in $H \cup L$ approaches zero. 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 than other global optima.

Note that conditions (e) and (f) of Theorem 5.2 are consistent with Hajek's condition for the simulated annealing algorithm's cooling parameter β in (2.21). If the GHC algorithm is formulated as SA, then conditions (e) and (f) are satisfied if β is greater than or equal to the depth of the deepest local minimum in the set L, but less than the depth of a global minimum. Note also 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 $k \to \infty$.

Condition (g) requires that the equilibrium probability distribution $\pi(k)$ be explicitly known only for the set of local optima. On the other hand, Anily and Federgruen's [1987] convergence theorem requires that the equilibrium distribution $\pi(k)$ be known for *all* solutions in Ω--hence Theorem 5.2 is a relaxation of Anily and

Federgruen's [1987] result, in that Theorem 5.2 requires equilibrium distribution information for only a (presumably small) subset of the solution space. However, in practice this condition would still be very difficult to check, unless reversibility (2.7) is also satisfied.

5.4 Illustrative Examples

Example 5.4.1 illustrates how a GHC acceptance function, based on a polynomial function of k, satisfies the conditions of Theorems 5.1 and 5.2. Example 5.4.2 shows that the threshold accepting algorithm does not satisfy the sufficient conditions in Theorem 3.2, Theorem 4.4, and Theorem 5.1.

5.4.1 Generalized Hill Climbing Acceptance as a Polynomial Function of k

Consider the eight-solution example depicted in Figure 5.2, where $G=\{p\}$, $L=\{q_1,q_2,q_3\}$, and $H=\{r_1,r_2,r_3,r_4\}$. Let each one-step transition probability be

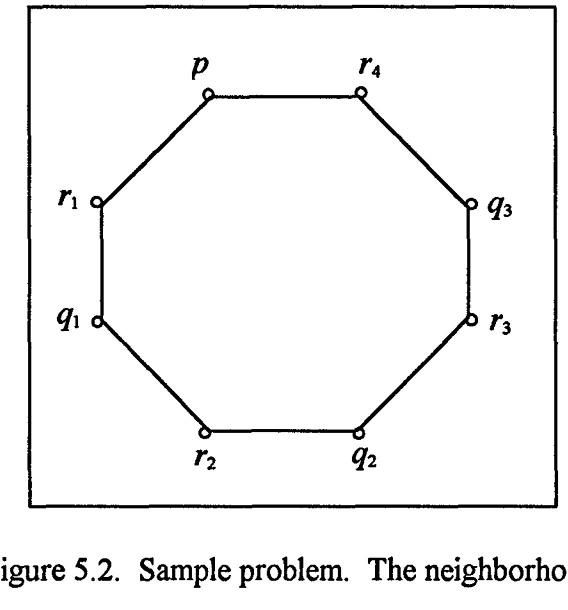

Figure 5.2. Sample problem. The neighborhood structure is shown by the lines connecting the nodes.

defined (for all $k \geq 2$) as in Figure 5.3. Note that $g_{i,j}(k) \equiv 1/2$ for all $i \in \Omega$, $j \in \mathcal{N}(i)$, and for all k. Also, $R_k(p,r_j) \equiv \Delta_{p,r_j} / (k^2 U)$ for $p, r_j \in \Omega$, $j = 1,2,3,4$, and $R_k(q_i, r_j) \equiv \Delta_{q_i,r_j} / (kU)$ for all $i = 1,2,3$ and $j = 1,2,3,4$, where U is distributed $U(0,1)$. Then $\Pr(R_k(p,r_j) \geq \Delta_{p,r_j}) = 1/k^2$, and $\Pr(R_k(q_i,r_j) \geq \Delta_{q_i,r_j}) = 1/k$ 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 5.1 are satisfied. Furthermore, all hill climbing transition probabilities

$$\boxed{\begin{bmatrix}
1-\frac{1}{k^2} & 0 & 0 & 0 & \frac{1}{2k^2} & 0 & 0 & \frac{1}{2k^2} \\
0 & 1-\frac{1}{k} & 0 & 0 & \frac{1}{2k} & \frac{1}{2k} & 0 & 0 \\
0 & 0 & 1-\frac{1}{k} & 0 & 0 & \frac{1}{2k} & \frac{1}{2k} & 0 \\
0 & 0 & 0 & 1-\frac{1}{k} & 0 & 0 & \frac{1}{2k} & \frac{1}{2k} \\
\frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\
\frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0
\end{bmatrix}}$$

Figure 5.3. The one-step transition matrix $P(k)$, defined for $k \geq 2$. The rows and columns are arranged in order $p, q_1, q_2, q_3, r_1, r_2, r_3, r_4$.

from every solution in $(L \cup G)$ to its neighbors in H are strictly positive, with limit zero as $k \to \infty$, hence conditions (c) and (d) of Theorem 5.1 are satisfied, and so Theorem 5.1 applies. The sufficient conditions of Theorem 5.2 are now addressed.

Condition (e) examines the paths of positive probability from every local optimum to all solutions in $(L \cup G)$. Nine positive path probabilities exist:

(i) $P_k(q_1 \to q_2) > 0,$

(ii) $P_k(q_1 \to p) > 0,$

(iii) $P_k(q_1 \to q_1) = 1 - P_k(q_1 \to q_2) - P_k(q_1 \to p) > 0,$

(iv) $P_k(q_2 \to q_1) > 0,$

(v) $P_k(q_2 \to q_3) > 0,$

(vi) $P_k(q_2 \to q_2) = 1 - P_k(q_2 \to q_1) - P_k(q_2 \to q_3) > 0,$

(vii) $P_k(q_3 \to q_2) > 0,$

(viii) $P_k(q_3 \to p) > 0,$

(ix) $P_k(q_3 \to q_3) = 1 - P_k(q_3 \to q_2) - P_k(q_3 \to p) > 0.$

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 $L \cup G$. Note also that from the problem's symmetry, the probabilities (i), (ii), and (iii) are respectively equal to (vii), (viii), and (ix). In addition, (i) is equal to (ii), (iv), and (v); and (iii) is equal to (vi). Hence only (i) and (iii) must be checked. For (i),

$$\begin{aligned}\sum_{k=2}^{\infty} P_k(q_1 \to q_2) &= \sum_{k=2}^{\infty}\left(P_{q_1,r_2}(k)P_{r_2,q_2}(k)\right)\\ &= \sum_{k=2}^{\infty} 1/(2k)(1/2)\\ &= \sum_{k=2}^{\infty} 1/(4k)\\ &= \infty,\end{aligned}$$

and for (iii),

$$\begin{aligned}\sum_{k=2}^{\infty} P_k(q_1 \to q_1) &= \sum_{k=2}^{\infty}\left[1-\left(P_{q_1,r_2}(k)P_{r_2,q_2}(k)\right)-\left(P_{q_1,r_1}(k)P_{r_1,p}(k)\right)\right]\\ &= \sum_{k=2}^{\infty}\left[1-1/(2k)(1/2)-1/(2k)(1/2)\right]\\ &= \sum_{k=2}^{\infty}\left[1-1/(2k)\right]\\ &= \infty.\end{aligned}$$

Therefore condition (e) holds.

Condition (f) is now addressed. From the problem symmetry,

$$\begin{aligned}\sum_{k=2}^{\infty} P_k(p \to q_1) = \sum_{k=2}^{\infty} P_k(p \to q_3) &= \sum_{k=2}^{\infty} 1/\left(2k^2\right)(1/2)\\ &= \sum_{k=2}^{\infty} 1/\left(4k^2\right)\\ &< \infty,\end{aligned}$$

and so (f) holds.

Condition (g) requires that the equilibrium probability vector $\delta(k)$ be known for all solutions in L. Hence by solving (2.5), $\pi_p(k) = \dfrac{k^2}{k^2+3k+4}$ and $\pi_{q_i}(k) = \dfrac{k}{k^2+3k+4}$, for $i = 1,2,3$. Therefore, $\omega(k) = \dfrac{k^2+3k}{k^2+3k+4}$, and so

$\delta_p(k) = \frac{k^2}{k^2+3k}$ and $\delta_{q_i}(k) = \frac{k}{k^2+3k}$ for $i = 1,2,3$. Recall that the only positive path probabilities between $q_i, q_j \in L$, $i \neq j$, are (i), (iv), (v), and (vii); furthermore, they are all equal. In addition, $\delta_{q_1}(k) = \delta_{q_2}(k) = \delta_{q_3}(k)$ for all *k*, and so it suffices to check condition (g) for only one path between solutions in *L*, e.g., for q_1 and q_2. Hence,

$$\begin{aligned}\sum_{k=1}^{\infty} \delta_{q_1}(k) P_k(q_1 \to q_2) &= \sum_{k=1}^{\infty} \left(\frac{k}{k^2+3k}\right)\left(\frac{1}{4k}\right) \\ &= \frac{1}{4}\sum_{k=1}^{\infty}\left(\frac{1}{k^2+3k}\right) \\ &< \infty,\end{aligned}$$

and so the sufficient conditions of Theorems 5.1 and 5.2 are satisfied. Therefore $\lim_{k\to\infty} \delta_p(k) = 1$. (Note that

$$\lim_{k\to\infty} \delta_p(k) = \lim_{k\to\infty} \frac{k^2}{k^2+3k} = \lim_{k\to\infty} \pi_p(k) = \lim_{k\to\infty} \frac{k^2}{k^2+3k+4} = 1,$$

which confirms the result.)

5.4.2 GHC Formulated as Threshold Accepting

The threshold accepting (TA) algorithm (Dueck and Scheuer [1990]) 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

$$Q_0 \geq \max_{\substack{all\ i\in\Omega, \\ j\in\mathcal{N}(i)}} (c_j - c_i) \tag{5.19}$$

$$|Q_k| \leq Q_0, \quad for\ all\ k \tag{5.20}$$

$$\lim_{k\to\infty} Q_k = 0. \tag{5.21}$$

(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 state i to any other state j). Then the GHC acceptance probability distribution is

$$
\begin{aligned}
\Pr(R_k \ge \Delta_{i,j}) &= \Pr(Q_k \ge \Delta_{i,j}) \\
&= \begin{cases} 1 & \textit{if } Q_k \ge \Delta_{i,j} \\ 0 & \textit{otherwise.} \end{cases}
\end{aligned}
\tag{5.22}
$$

Note that no proofs of TA convergence to (1.1) are presented in the literature (Althofer and Koschnick [1991]). Lemma 5.1 illustrates why the TA formulation does not satisfy the sufficient conditions needed for the convergence proofs of Theorems 3.2, 4.4, or 5.1.

Lemma 5.1: *Suppose the GHC algorithm acceptance probability* $\Pr(R_k \ge \Delta_{i,j})$ *is defined as the TA formulation* (5.22). *Then the transition probabilities*

$$
P_{i,j}(k) = \begin{cases} g_{i,j}(k)\Pr\, Q_k \ge \Delta_{i,j} & j \in \mathcal{N}(i),\ j \ne i, \\ 0 & j \notin \mathcal{N}(i),\ j \ne i, \\ 1 - \sum\limits_{\substack{s \in \mathcal{N}(i) \\ s \ne i}} P_{i,s}(k) & j = i, \end{cases}
\tag{5.23}
$$

do not satisfy the irreducibility condition of Theorems 3.2, 4.4, and 5.1, for the existence of a unique stationary distribution $\pi(k)$ *for each iteration k.*

Proof: Irreducibility cannot be satisfied unless all states communicate (Ross [1993, pg 144]). However, from (5.21) the TA acceptance measure Q_k has limit zero, and thus for any *locally minimal* state *i* (e.g., $c_i < c_j$ for all $j \in \mathcal{N}(i)$), there exists an iteration k_0 such that $c_j - c_i = \Delta_{i,j} > Q_k$ for all $j \in \mathcal{N}(i)$ and all $k \geq k_0$. Hence state *i* no longer communicates with any state $j \in \Omega$ (e.g., $\Pr(Q_k \geq \Delta_{i,j}) = 0$ for all $j \in \mathcal{N}(i)$ and $k \geq k_0$), and thus irreducibility fails. Without irreducibility, Theorem 2.1 cannot be used to show the existence of the sequence of stationary probability distributions necessary for the proofs of convergence of Theorems 3.2, 4.4, or 5.1. ■

CHAPTER 6: COMPUTATIONAL RESULTS

This chapter presents the methodology and computational results for experiments conducted on three discrete optimization problems. The purpose of the experiments was to illustrate the relationships between the GHC algorithm formulations and their finite-time performance on the different problems. The three problems include:

(a) a flexible assembly system design (FASD) problem (Kumar and Jacobson [1996]),

(b) a generic configuration space (Fleischer [1994]),

(c) an Air Force manufacturing process design problem, which involves optimizing machine sequencing, workpiece, and machine parameter settings (Walker [1992]).

The five GHC algorithm formulations presented include:

(i) simulated annealing, where $R_k(i,j)$ is defined as in Section 3.3.1,

(ii) threshold accepting, where $R_k(i,j)$ is defined as in Section 5.4.2,

(iii) a Weibull formulation, where $R_k(i,j)$ is defined as in Section 3.3.3,

(iv) local search, where $R_k(i,j) \equiv 0$, for all $i,j \in \Omega$ and all k, and

(v) Monte Carlo search, where $R_k(i,j) \equiv {}^{+}\infty$, for all $i,j \in \Omega$ and all k.

Tests of each $R_k(i,j)$ formulation were conducted for each problem, to determine which GHC algorithm (of the five formulations presented) reached the best solution (i.e., the solution with the lowest objective function value), for a fixed number of iterations.

The experiments were designed using guidelines proposed by Crowder, Dembo and Mulvey [1979]. All computations were performed in C on either a SUN Ultra-1 SPARCstation, using the SunOS 5.5 operating system, or a SUN 10-51 SPARCstation using the SunOS 5.3 operating system. All random numbers were generated using a uniform (0,1) pseudo-random linear congruental generator, with multiplier 16807, modulus 2147483647, and increment 0 (see Press et. al [1992, pg 278] or Law and Kelton [1991, pg 424]). The same seed, 123, was used to initiate all experiments.

6.1 Flexible Assembly System Design Problem

The FASD problem is a precedence-constrained scheduling problem, in which a set of N tasks must be completed by a set of Z processor types. Each processor type can complete a subset of the tasks, but each task can be completed by only one processor type. The goal is to determine a sequence of processors that minimizes the total required number of processors, while satisfying precedence constraints between tasks. In essence, this problem is a precedence-constrained Hamiltonian path problem, where each task constitutes a vertex, and the cost of each path is the sum of the processors (of all types) needed to complete the path. Furthermore, a path is feasible if and only if the Hamiltonian path satisfies all precedence constraints between the tasks. Kumar and Jacobson [1996] show that the FASD problem is NP-complete (Garey and Johnson [1979, pg 17]). Jacobson et. al [1996] propose a simple matrix-based, polynomial-time lower bound algorithm for the problem.

The FASD solution space Ω is composed of N-tuples of tasks, where a solution is an ordering of the N tasks. Therefore, the size of the solution space, $card(\Omega)$, is $N!$. A neighbor of a solution is defined by interchanging the positions of two tasks in the N-tuple solution. Therefore, each solution, $i \in \Omega$, has $\binom{N}{2}$ neighbors. Lastly, the objective function is defined by first penalizing all precedence relation violations, then computing the sum of all penalties, and finally adding the total number of required processors.

6.1.1 Experimental Design

Three FASD test problems were constructed: $N = 50$ tasks with $Z = 20$ processor types; $N = 100$ tasks with $Z = 20$ processor types; and $N = 150$ tasks with $Z = 20$ processor types. Each task has at most five predecessors, and exactly one processor type, for all three test problems. An experiment was specified by selecting a FASD problem, an $R_k(i, j)$ formulation, and a corresponding set of inputs describing the number of iterations and the $R_k(i, j)$ parameter's behavior, and then executing the GHC algorithm formulation on one hundred randomly generated problem instances. Each problem instance was generated using its own unique random number seed. The same seeds were used to generate the problem instances across all experiments (i.e., all experiments on each problem used the same problem instances). Furthermore, the same (feasible) initial solution was used for each instance of every experiment.

For each FASD problem instance, the solution with the minimum objective function value found by the GHC algorithm (to date) was recorded, and the solution's feasibility was noted. For each experiment, the average minimum objective function value (i.e., the average minimum number of required processors), the standard deviation, and maximum and minimum objective function values, were computed from the set of feasible (best to date) solutions obtained from the one hundred problem instances.

6.1.2 Results

The first set of experiments were conducted to determine how the five selected GHC $R_k(i,j)$ parameter choices, the problem size, and the number of iterations, affected the performance of the GHC algorithm on the FASD problem. The results are presented in Tables 6.1 - 6.5. Of the five GHC formulations, SA's performance was the least sensitive to changes in its governing parameters (e.g., its initial temperature t_0, its cooling parameter ϕ, and its inner and outer loop values M and K). The Weibull formulation experiments used the same t_0, ϕ, M, and K parameters as used for the SA formulation, hence it measured how the shape parameter α would affect the performance of SA. Two Weibull shape parameters were considered: $\alpha = 1/2$, and $\alpha = 2$. Note that the $\alpha = 1/2$ case has the effect of increasing the probabilities of SA hill climbing transitions; conversely, the $\alpha = 2$ case reduces the corresponding transition probabilities. For the FASD problem, the $\alpha = 2$ case virtually guaranteed that for each trial, the final solution would be feasible. However, the Weibull formulation did not measurably improve the mean final objective function values over those found by the SA formulation. Furthermore, the GHC algorithm $\alpha = 1/2$ Weibull formulation was unable to find as

Table 6.1. FASD results for the SA GHC algorithm, where $t_0 = 1.0$, and $t_{k+1} = \phi t_k$.

Note: cpu times are for the SUN Ultra-1, and are typical of the five GHC formulations.

Outer loop K	Inner loop M	Decrement ϕ	Mean objective function value	Standard deviation	Fraction feasible	Minimum objective function value	Maximum objective function value
(50 tasks, 20 processors)			(14 cpu minutes)				
500	200	.99	29.31	1.98	1.00	24	34
250	400	.95	29.31	1.85	1.00	24	34
100	1000	.90	29.21	1.93	0.99	25	34
100	1000	.85	29.21	2.00	1.00	24	33
			(1.5 cpu minutes)				
500	20	.99	31.37	2.18	1.00	26	36
250	40	.95	31.28	2.13	1.00	27	38
100	100	.90	31.37	2.34	1.00	24	37
100	100	.85	31.36	1.93	1.00	26	36
(100 tasks, 20 processors)			(25 cpu minutes)				
500	200	.99	47.52	3.14	0.99	37	56
250	400	.95	47.72	2.75	0.99	39	54
100	1000	.90	47.61	2.99	0.96	39	54
100	1000	.85	47.84	3.04	0.97	40	56
			(2.5 cpu minutes)				
500	20	.99	58.38	3.08	0.97	51	66
250	40	.95	57.25	3.16	0.99	51	65
100	100	.90	57.13	3.02	0.98	49	64
100	100	.85	57.37	3.41	0.98	48	64
(150 tasks, 20 processors)			(37 cpu minutes)				
500	200	.99	66.11	3.60	0.96	58	74
250	400	.95	65.77	4.13	0.96	59	79
100	1000	.90	65.71	3.80	0.97	58	77
100	1000	.85	65.88	3.93	0.94	57	75
			(4 cpu minutes)				
500	20	.99	88.90	3.71	0.96	79	96
250	40	.95	87.85	3.62	0.98	75	95
100	100	.90	87.96	3.48	0.97	81	96
100	100	.85	87.52	3.68	0.98	78	97

Table 6.2. FASD results for the TA GHC algorithm, where $Q_{k+1} = \phi Q_k$.

Outer loop K	Inner loop M	Decre-ment ϕ	Initial threshold Q_0	Mean objective function value	Standard deviation	Fraction feasible	Minimum objective function value	Maximum objective function value
(50 tasks, 20 processors)								
500	200	0.995	10	31.36	2.14	1.00	27	36
50	2000	0.90	10	29.76	1.81	1.00	25	34
500	200	0.995	100	47.29	2.15	0.21	42	50
50	2000	0.90	100	30.98	1.92	1.00	25	36
500	20	0.995	10	38.38	2.33	0.88	34	48
50	200	0.90	10	31.90	2.22	0.81	26	35
500	20	0.995	100	47.29	2.15	0.21	42	50
50	200	0.90	100	36.93	2.46	0.69	31	46
(100 tasks, 20 processors)								
500	200	0.995	10	58.19	3.38	0.98	50	67
50	2000	0.90	10	48.17	3.24	0.92	42	56
500	200	0.995	100	94.94	1.73	0.16	92	98
50	2000	0.90	100	55.26	3.36	0.93	43	64
500	20	0.995	10	82.00	5.63	0.30	72	95
50	200	0.90	10	61.42	3.54	0.33	54	69
500	20	0.995	100	94.94	1.73	0.16	92	98
50	200	0.90	100	87.14	9.06	0.14	72	98
(150 tasks, 20 processors)								
500	200	0.995	10	89.11	3.71	0.85	81	96
50	2000	0.90	10	67.46	3.51	0.78	60	76
500	200	0.995	100	142.37	2.45	0.19	136	146
50	2000	0.90	100	83.30	3.33	0.67	73	93
500	20	0.995	10	140.09	4.96	0.22	123	147
50	200	0.90	10	96.00	1.41	0.02	95	97
500	20	0.995	100	142.37	2.45	0.19	136	146
50	200	0.90	100	142.54	2.18	0.13	138	146

Table 6.3. FASD results for the Weibull GHC algorithm,

where $t_0 = 1.0$, and $t_{k+1} = \phi t_k$.

Outer loop K	Inner loop M	Decrement ϕ	Shape parameter α	Mean objective function value	Standard deviation	Fraction feasible	Minimum objective function value	Maximum objective function value
(50 tasks, 20 processors)								
250	400	0.95	0.5	29.26	2.16	0.92	24	38
250	400	0.95	2.0	29.34	1.92	1.00	25	34
100	1000	0.85	0.5	29.42	1.95	0.93	25	35
100	1000	0.85	2.0	29.28	1.80	1.00	25	33
250	40	0.95	0.5	31.00	2.23	0.73	26	35
250	40	0.95	2.0	31.40	2.02	1.00	25	36
100	100	0.85	0.5	31.25	2.16	0.79	25	36
100	100	0.85	2.0	31.33	2.01	1.00	27	36
(100 tasks, 20 processors)								
250	400	0.95	0.5	46.69	2.84	0.62	40	53
250	400	0.95	2.0	47.94	2.68	1.00	41	55
100	1000	0.85	0.5	47.48	2.92	0.64	40	53
100	1000	0.85	2.0	47.97	2.90	1.00	41	55
250	40	0.95	0.5	57.87	2.76	0.53	52	64
250	40	0.95	2.0	57.60	3.37	1.00	48	65
100	100	0.85	0.5	57.66	3.45	0.53	50	66
100	100	0.85	2.0	57.37	3.32	1.00	48	66
(150 tasks, 20 processors)								
250	400	0.95	0.5	65.09	3.61	0.43	59	73
250	400	0.95	2.0	65.88	3.69	1.00	57	76
100	1000	0.85	0.5	65.18	3.13	0.38	59	73
100	1000	0.85	2.0	65.98	3.42	1.00	58	75
250	40	0.95	0.5	89.29	3.77	0.45	83	100
250	40	0.95	2.0	86.88	4.11	1.00	78	97
100	100	0.85	0.5	88.18	4.32	0.38	81	97
100	100	0.85	2.0	86.92	3.67	1.00	78	96

Table 6.4. FASD results for the local search GHC algorithm.

Outer loop K	Inner loop M	Para-meter $R_k(i,j)$	Mean objective function value	Standard deviation	Fraction feasible	Minimum objective function value	Maximum objective function value
(50 tasks, 20 processors)							
1	100000	0	36.46	2.44	1.00	30	43
1	10000	0	36.46	2.44	1.00	30	43
(100 tasks, 20 processors)							
1	100000	0	62.42	3.43	1.00	51	71
1	10000	0	63.99	3.30	1.00	54	72
(150 tasks, 20 processors)							
1	100000	0	87.82	5.86	1.00	76	107
1	10000	0	94.04	4.46	1.00	84	107

Table 6.5. FASD results for the Monte Carlo GHC algorithm.

Outer loop K	Inner loop M	Para-meter $R_k(i,j)$	Mean objective function value	Standard deviation	Fraction feasible	Minimum objective function value	Maximum objective function value
(50 tasks, 20 processors)							
1	100000	1e06	47.29	2.15	0.21	42	50
1	10000	1e06	47.29	2.15	0.21	42	50
(100 tasks, 20 processors)							
1	100000	1e06	94.94	1.73	0.16	92	98
1	10000	1e06	94.94	1.73	0.16	92	98
(150 tasks, 20 processors)							
1	100000	1e06	142.37	2.45	0.19	136	146
1	10000	1e06	142.37	2.45	0.19	136	145

many feasible final solutions as the corresponding SA formulation, and the feasibility percentage decreased for the larger FASD problem instances.

Of the three hill climbing GHC formulations, TA's performance was the most sensitive to changes in its parameters (e.g., its initial threshold Q_0, its reduction parameter ϕ, and its inner and outer loop values M and K). TA achieved its best results for FASD by using a low initial threshold and a fast threshold decrement rate, (i.e., by quickly converging to local search). However, TA was generally outperformed by SA and the Weibull formulation (in terms of mean final objective function values and feasibility percentages), especially for the larger FASD problem sizes. Finally, both local search and Monte Carlo search were outperformed by the three hill climbing algorithms. Monte Carlo search was the worst performer, in terms of mean objective function values and final solution feasibility percentages.

The second set of experiments were conducted to determine how the five selected $R_k(i,j)$ formulations performed over ten thousand iterations $(K = 100, M = 100)$, on the FASD $N = 50/Z = 20$, 100/20, and 150/20 test problems. The mean final objective function value found by each $R_k(i,j)$ formulation was recorded and plotted at 500-iteration increments.

The results for the FASD 50/20 problem are depicted in Figure 6.1. For SA, $t_0 = 1.0$, and the decrement was 0.85. The Weibull used the same t_0 value and decrement as SA, and set $\alpha = 0.5$. TA used $Q_0 = 10.0$, and the decrement was 0.9. The results for the FASD 100/20 and 150/20 problems are depicted in Figures 6.2 and 6.3, respectively.

For SA, $t_0 = 1.0$, and the decrement was 0.85. The Weibull used the same t_0 value and decrement as SA, and set $\alpha = 2.0$. TA used $Q_0 = 3.0$, and the decrement was 0.90.

Note that the Monte Carlo formulation was unable to locate *any* solutions with lower objective function values than the initial solution. The local search implementation did somewhat better, but was still outperformed by the hill climbing algorithms. The TA, SA, and Weibull formulations all performed about the same. The Monte Carlo formulation's poor performance suggests that the FASD problem probably contains relatively few good solutions, while the local search formulation's relatively better performance suggests that neighboring solutions are likely to have similar objective function values (e.g., neighborhoods form regions of attraction around local optima).

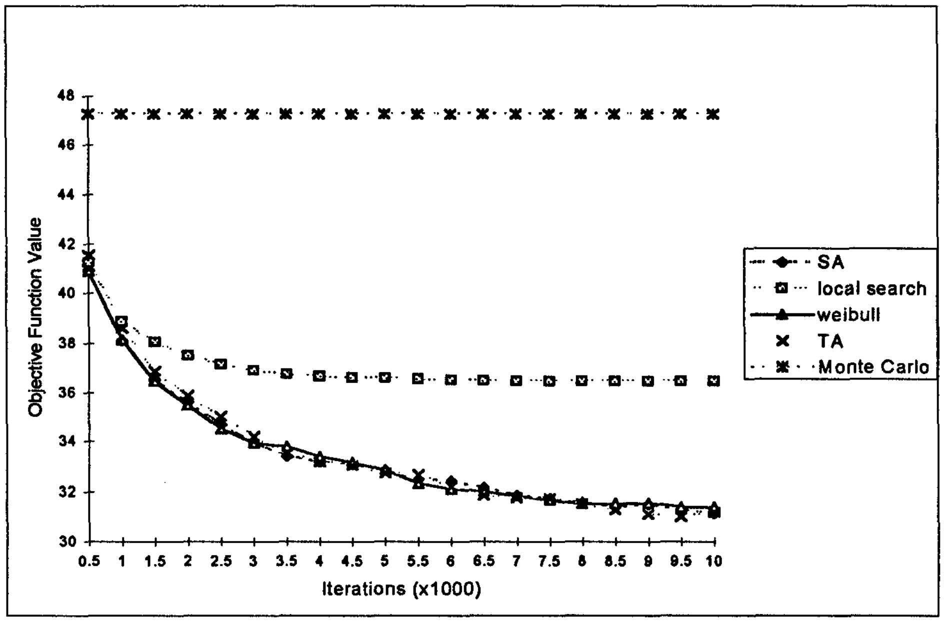

Figure 6.1 Comparison of Acceptance Formulations for the 50/20 FASD Problem.

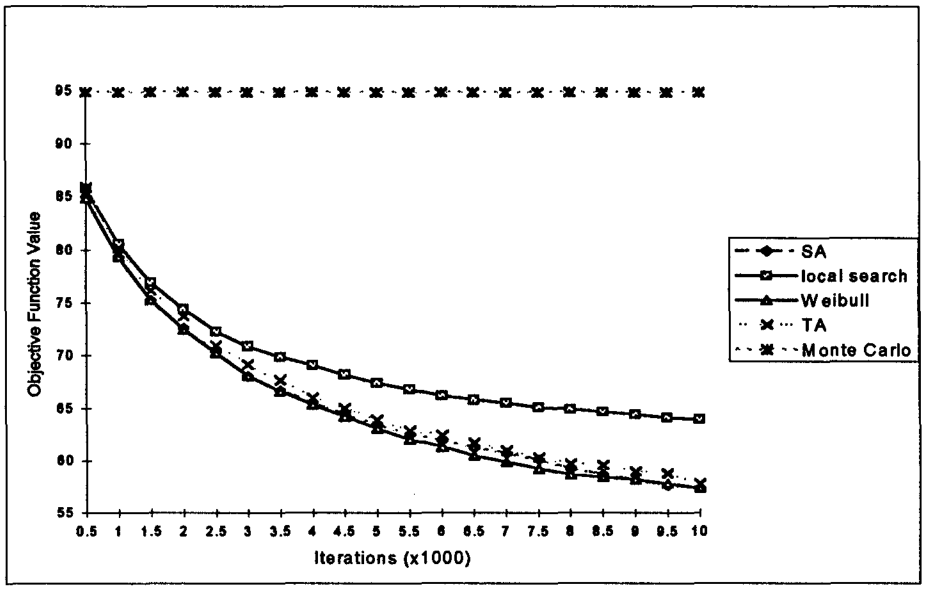

Figure 6.2 Comparison of Acceptance Formulations for the 100/20 FASD Problem.

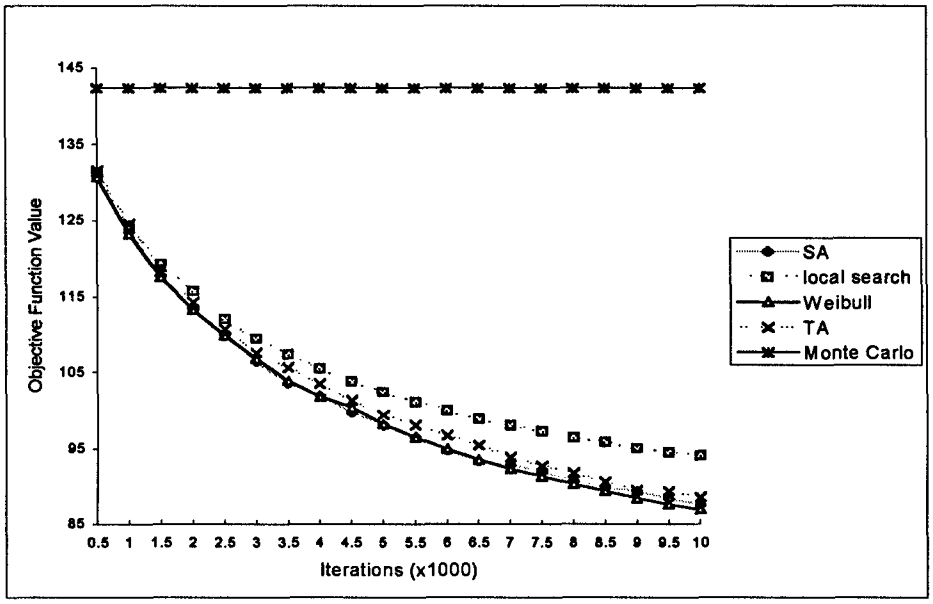

Figure 6.3 Comparison of Acceptance Formulations for the 150/20 FASD Problem.

6.2 Generic Configuration Space Problem

The generic configuration space problem (Fleischer [1994]) was selected because it allows the researcher to precisely control each test problem's number, depth, and location of local and global optima. To create a generic configuration space, a set of objective function values are randomly generated and stored in memory. A neighborhood is then defined by linking each objective function value to any specified number of other objective function values. The main disadvantage of the generic configuration space approach is that its computer memory requirement is excessive for large solution space cardinalities.

6.2.1 Experimental Design

The goal of the generic configuration space experiments was to determine whether neighborhood size and the range in objective function values (i.e., the problem's objective function topology) would affect the performance of the five GHC algorithm formulations. Two test problems were created: for the first problem, a vector of five thousand objective function values was defined, with values between zero and one hundred. For the second problem, a vector of five thousand objective function values was again defined, but with values between zero and five hundred. An experiment was defined by executing one of the five GHC algorithm formulations on one hundred randomly generated problem instances. The same initial solution (vector element) was used for each instance of each experiment. Finally, for each experiment, the mean best to date objective function value, the standard deviation, and maximum and minimum objective function values, were computed.

The experiments assessed whether changing the total number of iterations (one hundred, five hundred, or one thousand), the neighborhood size (from ten to fifty neighbors, in steps of ten), or the range in objective function values (zero to one hundred, or zero to five hundred), would affect the performance of the five GHC algorithm formulations. Finally, pilot runs were conducted for the SA, TA, and Weibull GHC formulations. The pilot runs were used to tune the input parameters of the respective GHC formulations to obtain their best possible performance for the generic configuration space problem.

6.2.2 Results

Tables 6.6 - 6.10 depict the results for the five selected GHC algorithm formulations. For the generic configuration space, the five formulations performed much differently with respect to each other than they had for the FASD problem in Section 6.1. The Monte Carlo GHC formulation (see Table 6.10) consistently found mean final objective function values that were less than or equal to those found by the four other GHC algorithm formulations. The local search formulation was the worst performer, and the TA, SA, and Weibull formulations all performed about the same.

For all five GHC formulations, larger neighborhoods and tighter objective function value ranges had the effect of improving performance. Increasing the total number of iterations improved the performance of all GHC formulations except local search, which apparently became trapped in local minima within the first 100 iterations.

Table 6.6. Results for the SA GHC algorithm, where $t_0 = 300$, and $t_{k+1} = 0.95t_k$.

Outer loop *K*	Inner loop *M*	Objective function values	Neighbor-hood size	Mean objective function value	Standard deviation	Minimum objective function value	Maximum objective function value
100	10	0-100	10	0.69	0.97	0	5
100	10	0-100	20	0.36	0.64	0	3
100	10	0-100	30	0.14	0.40	0	2
100	10	0-100	40	0.09	0.35	0	2
100	10	0-100	50	0.04	0.20	0	1
100	10	0-500	10	9.12	9.03	0	50
100	10	0-500	20	5.31	5.38	0	23
100	10	0-500	30	3.97	5.20	0	40
100	10	0-500	40	2.87	3.41	0	17
100	10	0-500	50	2.18	2.65	0	15
100	5	0-100	10	1.28	1.58	0	9
100	5	0-100	20	0.63	0.98	0	5
100	5	0-100	30	0.34	0.74	0	5
100	5	0-100	40	0.19	0.53	0	3
100	5	0-100	50	0.18	0.44	0	2
100	5	0-500	10	11.52	11.61	0	62
100	5	0-500	20	7.97	9.51	0	62
100	5	0-500	30	5.11	5.59	0	30
100	5	0-500	40	3.69	4.94	0	34
100	5	0-500	50	2.72	3.10	0	19
100	1	0-100	10	3.72	4.11	0	17
100	1	0-100	20	1.86	2.55	0	13
100	1	0-100	30	1.61	2.13	0	9
100	1	0-100	40	1.24	1.70	0	12
100	1	0-100	50	1.45	2.00	0	13
100	1	0-500	10	23.33	22.2	0	91
100	1	0-500	20	14.28	13.87	0	58
100	1	0-500	30	10.24	11.20	0	67
100	1	0-500	40	8.39	9.60	0	51
100	1	0-500	50	7.15	7.87	0	45

Table 6.7. Results for the TA GHC algorithm, where $Q_0 = 300$, and $Q_{k+1} = 0.95Q_k$.

Outer loop K	Inner loop M	Objective function values	Neighbor-hood size	Mean objective function value	Standard deviation	Minimum objective function value	Maximum objective function value
100	10	0-100	10	0.83	1.17	0	5
100	10	0-100	20	0.37	0.77	0	5
100	10	0-100	30	0.15	0.41	0	2
100	10	0-100	40	0.01	0.10	0	1
100	10	0-100	50	0.06	0.28	0	2
100	10	0-500	10	7.72	7.60	0	28
100	10	0-500	20	5.58	6.03	0	29
100	10	0-500	30	3.91	4.50	0	20
100	10	0-500	40	3.11	3.55	0	16
100	10	0-500	50	1.97	2.32	0	11
100	5	0-100	10	1.31	1.77	0	9
100	5	0-100	20	0.62	1.03	0	5
100	5	0-100	30	0.30	0.61	0	3
100	5	0-100	40	0.17	0.47	0	2
100	5	0-100	50	0.16	0.42	0	2
100	5	0-500	10	12.71	13.38	0	75
100	5	0-500	20	7.46	7.07	0	28
100	5	0-500	30	5.10	6.17	0	47
100	5	0-500	40	3.78	3.96	0	18
100	5	0-500	50	3.79	4.63	0	20
100	1	0-100	10	3.04	3.33	0	16
100	1	0-100	20	2.02	2.71	0	14
100	1	0-100	30	1.37	1.76	0	7
100	1	0-100	40	1.07	1.43	0	7
100	1	0-100	50	1.07	1.47	0	8
100	1	0-500	10	20.71	19.86	0	78
100	1	0-500	20	12.51	13.26	0	78
100	1	0-500	30	9.04	8.72	0	47
100	1	0-500	40	8.39	9.28	0	47
100	1	0-500	50	7.50	7.34	0	34

Table 6.8. Results for the Weibull GHC algorithm, where $t_0 = 300$, $t_{k+1} = 0.95t_k$, and $\alpha = 0.25$.

Outer loop K	Inner loop M	Objective function values	Neighbor-hood size	Mean objective function value	Standard deviation	Minimum objective function value	Maximum objective function value
100	10	0-100	10	0.69	1.01	0	5
100	10	0-100	20	0.41	0.70	0	3
100	10	0-100	30	0.18	0.48	0	2
100	10	0-100	40	0.10	0.36	0	2
100	10	0-100	50	0.08	0.31	0	2
100	10	0-500	10	6.57	7.55	0	53
100	10	0-500	20	3.44	3.60	0	17
100	10	0-500	30	2.59	3.23	0	20
100	10	0-500	40	1.90	2.32	0	16
100	10	0-500	50	1.50	1.91	0	10
100	5	0-100	10	1.07	1.49	0	9
100	5	0-100	20	0.46	0.74	0	3
100	5	0-100	30	0.24	0.55	0	3
100	5	0-100	40	0.20	0.47	0	2
100	5	0-100	50	0.11	0.35	0	2
100	5	0-500	10	8.89	10.38	0	62
100	5	0-500	20	5.06	6.90	0	47
100	5	0-500	30	4.14	4.44	0	23
100	5	0-500	40	3.05	3.74	0	16
100	5	0-500	50	2.45	3.16	0	15
100	1	0-100	10	2.82	3.22	0	14
100	1	0-100	20	1.85	2.63	0	13
100	1	0-100	30	1.22	1.51	0	9
100	1	0-100	40	1.21	1.82	0	10
100	1	0-100	50	1.09	1.75	0	10
100	1	0-500	10	16.78	19.37	0	141
100	1	0-500	20	10.92	11.38	0	53
100	1	0-500	30	10.03	11.33	0	58
100	1	0-500	40	7.37	7.79	0	43
100	1	0-500	50	6.64	6.97	0	36

Table 6.9. Results for the local search GHC algorithm.

Outer loop *K*	Inner loop *M*	Objective function values	Neighbor-hood size	Mean objective function value	Standard deviation	Minimum objective function value	Maximum objective function value
1	1000	0-100	10	6.28	6.22	0	27
1	1000	0-100	20	3.29	3.71	0	16
1	1000	0-100	30	2.38	2.81	0	13
1	1000	0-100	40	1.59	2.17	0	10
1	1000	0-100	50	0.95	1.25	0	6
1	1000	0-500	10	36.88	37.34	0	188
1	1000	0-500	20	21.01	23.93	0	156
1	1000	0-500	30	11.36	11.62	0	54
1	1000	0-500	40	9.37	9.11	0	50
1	1000	0-500	50	7.58	9.52	0	47
1	500	0-100	10	6.28	6.22	0	27
1	500	0-100	20	3.29	3.71	0	16
1	500	0-100	30	2.38	2.81	0	13
1	500	0-100	40	1.59	2.17	0	10
1	500	0-100	50	0.95	1.25	0	6
1	500	0-500	10	36.88	37.34	0	188
1	500	0-500	20	21.01	23.93	0	156
1	500	0-500	30	11.36	11.62	0	54
1	500	0-500	40	9.37	9.11	0	50
1	500	0-500	50	7.58	9.52	0	47
1	100	0-100	10	6.28	6.22	0	27
1	100	0-100	20	3.29	3.71	0	16
1	100	0-100	30	2.38	2.81	0	13
1	100	0-100	40	1.59	2.17	0	10
1	100	0-100	50	0.95	1.25	0	6
1	100	0-500	10	36.88	37.34	0	188
1	100	0-500	20	21.01	23.93	0	156
1	100	0-500	30	11.36	11.62	0	54
1	100	0-500	40	9.37	9.11	0	50
1	100	0-500	50	7.58	9.52	0	47

Table 6.10. Results for the Monte Carlo GHC algorithm.

Outer loop *K*	Inner loop *M*	Objective function values	Neighbor-hood size	Mean objective function value	Standard deviation	Minimum objective function value	Maximum objective function value
1	1000	0-100	10	0.37	0.72	0	3
1	1000	0-100	20	0.07	0.33	0	2
1	1000	0-100	30	0.08	0.34	0	2
1	1000	0-100	40	0.03	0.22	0	2
1	1000	0-100	50	0.01	0.10	0	1
1	1000	0-500	10	3.61	4.5	0	22
1	1000	0-500	20	1.72	2.16	0	10
1	1000	0-500	30	1.21	1.65	0	8
1	1000	0-500	40	0.93	1.3	0	6
1	1000	0-500	50	0.86	1.33	0	6
1	500	0-100	10	0.81	1.13	0	5
1	500	0-100	20	0.39	0.76	0	4
1	500	0-100	30	0.17	0.43	0	2
1	500	0-100	40	0.08	0.31	0	2
1	500	0-100	50	0.04	0.24	0	2
1	500	0-500	10	4.96	5.59	0	27
1	500	0-500	20	3.47	4.51	0	20
1	500	0-500	30	2.28	2.94	0	17
1	500	0-500	40	2.25	2.87	0	20
1	500	0-500	50	1.67	2.06	0	10
1	100	0-100	10	2.20	2.85	0	16
1	100	0-100	20	1.62	2.26	0	14
1	100	0-100	30	1.23	2.04	0	14
1	100	0-100	40	0.87	1.30	0	7
1	100	0-100	50	1.01	1.47	0	8
1	100	0-500	10	12.06	13.75	0	93
1	100	0-500	20	8.63	8.58	0	43
1	100	0-500	30	7.0	7.32	0	35
1	100	0-500	40	6.71	6.12	0	27
1	100	0-500	50	6.29	6.23	0	29

One explanation for the success of the Monte Carlo GHC formulation is that the five thousand objective function values in each problem were generated uniformly randomly between their minimum and maximum allowable values. Therefore a global maximum was as likely to be a neighbor of a global minimum as any other solution, and so the objective function topology was highly irregular, with small or nonexistent regions of attraction surrounding local minima. Another explanation is that for each experiment, the GHC algorithm had the opportunity to search a significant proportion of the solution space (between 2% and 20%, depending on the number of iterations performed). These percentages are much higher than would occur if the GHC algorithm was executed on a more typical discrete optimization problem, and thus would also favor Monte Carlo search. Research is needed to assess how the five GHC formulations would perform on a smoother objective function topology, such as a discretized version of the problem suggested by Bohachevsky, Johnson, and Stein [1986, Figure 1].

6.3 Air Force Manufacturing Process Design Problem

The research goal of this section was to examine how the five selected GHC formulations perform on an Air Force manufacturing process design problem. The Air Force wishes to develop manufacturing process design tools that optimize the selection, input parameters, and sequencing of operations for material shaping and microstructure treatment. As an initial application, the Air Force is planning to use the tools to design the manufacturing process for an aircraft turbine engine compressor rotor. Traditional production techniques use empirical rules to optimize *individual* production operations (e.g., forging, machining, and heat treatment), and sequence the individual operations based on experience or industry tradition. In contrast, the Air Force wishes to optimize the *overall* manufacturing process, in order to achieve the best balance of manufacturing cost, producibility, and final product properties.

A principal challenge to the Air Force approach is that the size of the solution space Ω can become extremely large as the number of potential process operations, operation parameters, and operation parameter values (discretized, if necessary) are increased. (The solution space cardinality of the problem used for the computations in this section is approximately 10^{20}.) In addition, Sullivan [1996] describes the difficulties of developing a measure (e.g., an objective function) that accurately captures the process stability, product properties, and total cost.

With the assistance of Wright Laboratory and Ohio University personnel, Sullivan [1996] analyzes three candidate process designs for the Air Force initial application. This research selects one of the designs--an upset forge, rough machine, and finish machine sequence, and uses the design to compare the respective performance of the five GHC algorithm formulations. The upset forge, rough machine, and finish machine models were developed and coded in C by researchers at Ohio University (Sullivan [1996]). The objective function and process feasibility criteria were developed by a team of Ohio University, Wright Laboratory, and Virginia Tech researchers. (Two causes of process infeasibility are the violation of a physical constraint (e.g., exceeding a forge press capacity), or the violation of a material property (e.g., material cracking from excessive deformation in a forging operation).) Finally, a neighborhood was defined by randomly selecting one of the process operations, then randomly selecting an operation parameter, and finally randomly perturbing the parameter's value.

6.3.1 Experimental Design

An experiment was specified by selecting a set of inputs for the forge and machine model parameters, a GHC algorithm ($R_k(i,j)$) formulation, and a corresponding set of inputs describing the number of iterations and the $R_k(i,j)$ parameter's behavior. An experiment was conducted by executing the GHC algorithm formulation on thirty problem instances. Each problem instance was generated by randomly selecting a set of inputs for the model parameters. The same seeds were used to generate the problem instances for all experiments. All experiments were replicated twice, first for ten thousand iterations (per problem instance), and then for one hundred thousand iterations (per problem instance).

For each problem instance, the solution with the minimum objective function value found by the GHC algorithm was recorded, and the solution's feasibility was noted. For each experiment, the mean minimum objective function value, the standard deviation, and maximum and minimum objective function values, were computed from the set of feasible final solutions obtained from the thirty problem instances.

6.3.2 Results

All final solutions were feasible in each experiment, hence all results are averaged over thirty trials. The SUN Ultra-1 computer required an average of 30 cpu minutes for each ten-thousand iteration experiment, and an average of 310 cpu minutes for each one-hundred-thousand iteration experiment. (This is why only thirty trials were conducted per experiment, versus one hundred trials for the problems in Sections 6.1 and 6.2.) The results are shown in Tables 6.11 - 6.14.

SA was able to find better results (lower mean objective function values) than TA for the ten-thousand iteration computations. The performance gap was most pronounced for experiments in which the $R_k(i,j)$ parameter was very slowly decreased (e.g., for the $K = 500$, $M = 20$ case of Tables 6.11 and 6.12). SA and TA performed about the same for the one hundred-thousand iteration computations. The Weibull GHC formulation (Table 6.13) provided the best overall performance for the Air Force process design problem, for the experiments in which the shape parameter alpha was strictly less than one.

Table 6.11. Results for the SA GHC algorithm, where $t_0 = 10000$, and $t_{k+1} = \phi t_k$.

Outer loop K	Inner loop M	Decrement ϕ	Mean objective function value	Standard deviation	Minimum objective function value	Maximum objective function value
500	200	.99	1645.21	8.52	1618.89	1659.44
200	500	.95	1655.78	31.56	1627.97	1745.56
200	500	.90	1700.38	44.20	1634.59	1754.67
100	1000	.85	1688.28	43.03	1636.01	1752.09
500	20	.99	1648.24	7.88	1624.55	1659.44
200	50	.95	1708.38	73.27	1638.42	1852.99
200	50	.90	1746.56	73.09	1624.55	1826.80
100	100	.85	1763.14	76.51	1648.78	1858.53

Table 6.12. Results for the TA GHC algorithm,

where $Q_0 = 10000$, and $Q_{k+1} = \phi Q_k$.

Outer loop K	Inner loop M	Decrement ϕ	Mean objective function value	Standard deviation	Minimum objective function value	Maximum objective function value
500	200	.99	1673.45	32.58	1636.01	1745.56
200	500	.95	1652.61	3.09	1636.20	1653.18
200	500	.90	1699.96	48.26	1630.35	1745.56
100	1000	.85	1707.12	46.75	1624.55	1745.56
500	20	.99	1743.24	40.30	1636.20	1819.10
200	50	.95	1757.43	59.71	1624.55	1819.10
200	50	.90	1784.11	57.14	1636.01	1852.99
100	100	.85	1768.48	67.71	1624.55	1845.64

Table 6.13. Results for the Weibull GHC algorithm, where $t_0 = 10000$, and $t_{k+1} = \phi t_k$.

Outer loop K	Inner loop M	Parameter α	Decrement ϕ	Mean objective function value	Standard deviation	Minimum objective function value	Maximum objective function value
500	200	0.25	.99	1630.32	7.42	1620.69	1647.47
500	200	0.50	.99	1630.32	7.04	1624.55	1658.41
500	200	1.25	.99	1646.65	7.85	1636.01	1661.98
500	200	0.25	.95	1648.29	23.10	1636.01	1745.56
500	200	0.50	.95	1675.45	49.58	1624.55	1790.29
500	200	1.25	.95	1712.97	56.94	1638.42	1819.10
500	20	0.25	.99	1627.94	27.04	1492.15	1661.27
500	20	0.50	.99	1643.35	7.44	1626.21	1658.82
500	20	1.25	.99	1665.02	44.44	1624.55	1843.65
500	20	1.5	.99	1669.46	44.81	1633.39	1772.38
500	20	2.0	.99	1704.08	58.02	1632.83	1772.38
500	20	4.0	.99	1701.91	53.64	1633.39	1772.38
500	20	0.25	.95	1656.82	70.44	1492.15	1828.50
500	20	0.5	.95	1731.97	70.88	1641.84	1858.44
500	20	1.25	.95	1768.00	94.50	1638.56	2012.66

Table 6.14. Results for the local search and Monte Carlo GHC algorithms.

Outer loop K	Inner loop M	Parameter $R_k(i,j)$	Mean objective function value	Standard deviation	Minimum objective function value	Maximum objective function value
(local search)						
1	100000	0	1757.29	133.88	1616.82	2241.37
1	10000	0	1796.54	110.56	1624.55	2241.37
(Monte Carlo search)						
1	100000	1e09	1855.14	65.97	1507.71	1868.41
1	10000	1e09	1852.41	65.95	1507.71	1868.41

Indeed, the $\phi = 0.99$, $\alpha = 0.25$ experiments reported the lowest average objective function value of all experiments conducted. Note that this (ϕ, α) choice represents the slowest decreasing $R_k(i,j)$ parameter schedule of all experiments conducted for the Air Force problem. Finally, neither the local search nor the Monte Carlo GHC formulations performed as well as any of the hill climbing GHC algorithms.

6.4 Summary

The FASD, the generic configuration space, and the Air Force manufacturing design problem experiments suggest that relationships do exist between the selected GHC algorithm formulations and finite-time GHC algorithm performance. The Monte Carlo formulation performed very well on the generic configuration space problem, yet performed poorly on the other two problems. For the FASD problem, TA was much more sensitive to parameter tuning than was SA. The Weibull formulation performed better for the Air Force and generic configuration space problems when the shape parameter α was strictly less than one; conversely, α values strictly greater than one worked better for the FASD problem. Additional research is necessary to more explicitly quantify how each GHC algorithm formulation performs on the three problems, and to examine additional GHC algorithm formulations, such as a gamma random variable or a binomial random variable.

CHAPTER 7: FINALE

Many discrete optimization (minimization) problems are difficult to solve, i.e., are NP-complete. There are no known polynomial-time algorithms that can solve NP-complete problems. However, since NP-complete problems contain many examples of practical interest, heuristic methods have been developed that efficiently find near-optimal solutions. Heuristic methods can be grouped 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. Stochastic hill climbing algorithms are iterative in nature, and have the ability to probabilistically accept candidate solutions with higher cost than that of the incumbent solution, in an effort to escape local optima.

7.1 Conclusions and Extensions

This dissertation introduces generalized hill climbing (GHC) algorithms, which include several existing discrete optimization search algorithms such as simulated annealing (SA), local search, and threshold accepting (TA) as particular formulations. The contributions from this research focus on two areas: first, a new method of proving asymptotic convergence of stochastic hill climbing algorithms is presented, that relaxes the sufficient conditions found in the literature. This result creates a large body of convergent stochastic hill climbing algorithms, where only SA existed previously. Second, empirical tests of the performance of various GHC formulations are conducted on specific problem classes. These tests study which probability distributions enhance the GHC algorithm's finite-time performance (in terms of solution quality versus algorithm execution time) on three selected problems, including a flexible assembly system design (FASD) problem (a Hamiltonian path problem), a generic configuration space problem, and an Air Force manufacturing process design problem.

7.1.1 Knowledge of the Optimal Objective Function Value

For some discrete optimization problems, the globally optimal objective function value is known, and the goal is to identify an associated optimal solution. Theorems 3.1 and 3.2 show that when c_{opt} is known, then *any* acceptance distribution function that satisfies (3.1), Theorem 3.1 (c), and (3.6) is sufficient for asymptotic convergence to the set of globally optimal solutions. Therefore, the class of GHC algorithms presented in Chapter 3, provides flexibility in selecting the acceptance distribution while maintaining convergence properties to the set of globally optimal solutions.

One example is the discrete event simulation problem ACCESSIBILITY discussed by Yücesan and Jacobson [1996], where the goal is to identify a sequence of events (a solution) that enables a simulation model to terminate in a known state; the objective function value of this state is defined to be zero. For problems where the optimal objective function value is not known, an estimate of c_{opt} is sufficient to execute the GHC algorithm class defined in Chapter 3. The GHC algorithm is initiated using the c_{opt} estimate, and the estimate is updated as the algorithm progresses--see Bohachevsky et al. [1986]. This approach is being used to optimize an Air Force manufacturing process design problem (Walker [1992]).

7.1.2 Convergence Using Linear Algebra

The theorems in Chapter 4 provide a proof of convergence for a general GHC algorithm, that does not require knowledge of the optimal objective function value nor the reversibility condition. The proof is based on a perturbation theory developed for problems in linear algebra. However, these theorems imply that in general, it is very difficult to prove asymptotic convergence of a stochastic hill climbing algorithm by using linear algebra techniques, unless very stringent conditions (e.g., asymptotic boundedness of the norm of a transition matrix generalized inverse, or asymptotic nonsingularity of the transition matrix), are met.

7.1.3 General Proof of Convergence

Theorem 5.2's proof of convergence of the GHC algorithm, provides the most significant theoretical contribution of this dissertation. The sufficient conditions of Theorem 5.2 represent a relaxation of the most general sufficient conditions found in the literature (Anily and Federgruen, [1987]). The principal shortcoming of Theorem 5.2 is that in practice, the conditions would be difficult to check for problems with large solution space cardinalities. Two current research issues are to determine if the conditions of Theorem 5.2 are also necessary, and if these conditions can be reformulated or further relaxed, in order to make them easier to verify. This is further discussed in Section 7.2.

7.1.4 Computational Results

Experiments on the FASD problem, the generic configuration space problem, and the Air Force manufacturing problem, suggest that relationships do exist between the selected GHC algorithm formulations and finite-time GHC algorithm performance. SA performed well on all three problems, and its performance was relatively insensitive to changes in its parameter values. The Monte Carlo formulation performed very well on the generic configuration space problem, yet performed poorly on the other two. For the FASD problem, TA was much more sensitive to parameter tuning than was SA. The Weibull formulation performed better for the Air Force and generic configuration space problems when the shape parameter α was strictly less than one; conversely, α values strictly greater than one worked better for the FASD problem. (Note that when α is equal to one, then the Weibull formulation is equivalent to SA).

Additional research is necessary to assess how the five GHC algorithm formulations perform on generic configuration spaces that exhibit more structure than the randomly generated configuration spaces used for this dissertation. This research would determine which GHC algorithm formulation and what neighborhood definition would provide the best finite-time performance on each type of generic configuration space. Additional research would then extrapolate the results to problem classes in the literature possessing characteristics similar to those of the generic configuration space problems. The goal is to provide the practitioner with specific recommendations for a GHC formulation, associated parameter values (subject to scaling), and a neighborhood definition, that would give the best finite-time performance for a given discrete optimization problem.

7.2 Additional Extensions

A key limitation to the theorems of Chapters 4 and 5 is the difficulty of checking the sufficient conditions for problems with large solution space cardinalities. One way to address this difficulty would be to develop a way to *scale* a small $(\Omega, \mathcal{N}, c)$ problem (where the conditions can be easily checked) to a similar, but much larger-sized problem. Then, an algorithm with known asymptotic convergence properties for the small problem could be presumed to exhibit similar convergence properties for the larger problem. With this approach, one could derive bounds for the matrix norms of Chapter 4, or bounds for the equilibrium probabilities $\delta_j(k)$, $j \in L$, for the Theorems of Chapter 5.

An implication from Lemma 4.2 is that, for a given GHC algorithm formulation, a relationship exists between the number of local optima and the rate that the determinant of the matrix $Q(k)$ approaches zero. If this relationship could be quantified, the result may be useful for developing a bound for the rate at which hill climbing transition probabilities must approach zero.

The theorems of Chapter 5 lend themselves to extensions in several key areas. For the first extension, recall that the Theorem 5.2 conditions (e) - (g) require that *every* path probability (for solutions in $(L \cup G)$) be checked. Instead, condition (e) can be relaxed to consider the sum (over k) of the probabilities of only the most difficult path (i.e., the path of lowest probability from any solution $j \in L$ to any solution $i \in (L \cup G)$). Similarly, conditions (f) and (g) can be relaxed to consider only the maximal path probability (i.e., the easiest path), instead of considering the probabilities of all paths.

A second extension is to derive necessary conditions for Theorem 5.2. If conditions (e) - (g) should prove to be both sufficient and necessary, then Theorem 5.2 would entirely subsume Anily and Federgruen's [1987] result.

A final extension is to relax the need to know the equilibrium probabilities $\delta_j(k)$, for all $j \in L$ and all k. However, if this relaxation is infeasible, then an analysis of conditions (e) and (g) may provide implications on the *rate* that $\delta_j(k)$ must approach zero as k grows large, for all $j \in L$. This in turn would provide a link between finite-time performance and convergence rates.

Research is needed to assess the best neighborhood size for a given problem. Equations (5.1) and (5.2) define path probabilities in terms of the product of transition probabilities between sets of neighboring solutions. Research is needed to assess whether path probabilities and neighborhood size are related, and whether a bound on the best neighborhood size could be derived from these equations and Theorem 5.2.

A final research topic is to show how the tabu search (TS) heuristic (Glover [1994]) can be formulated as a GHC algorithm, and to derive conditions under which TS will asymptotically converge to the set of optimal solutions. Note that probabilistic TS can be formulated to resemble SA (see, e.g., Faigle and Kern [1992]). Since SA is a GHC algorithm, hence the probabilistic TS formulation is a GHC algorithm. The principal challenge is to show that the most general characteristics of TS, including short term memory, long term memory, and aspiration criteria, all fall within the GHC rubric. If so, then TS could claim the same asymptotic optimality distinction enjoyed by SA.

CHAPTER 8: BIBLIOGRAPHY

Aarts, E. and J. Korst [1989], *Simulated Annealing and Boltzmann Machines,* John Wiley and Sons, Tiptree, Essex, Great Britain.

Aarts, E. and P. van Laarhoven [1985], "Statistical Cooling: A General Approach to Combinatorial Optimization Problems," *Philips Journal of Research,* 40: 193-226.

Althofer, I. and K.U. Koschnick [1991], "On the Convergence of Threshold Accepting," *Applied Mathematics and Optimization,* 24: 183-195.

Anily, S. and A. Federgruen [1987], "Simulated Annealing Methods with General Acceptance Probabilities," *Journal of Applied Probability,* 24: 657-667.

Bohachevsky, I.O., M.E. Johnson, and M.L. Stein [1986], "Generalized Simulated Annealing for Function Optimization," *Technometrics,* 28(3): 209-217.

Borkar, V.S. [1992], "Pathwise Recurrence Orders and Simulated Annealing," *Journal of Applied Probability,* 29: 472-476.

Bratley, P., B.L. Fox, and L. Schrage [1987], *A Guide to Simulation (Second Edition),* Springer-Verlag, New York.

Cardoso, M.F., R.L. Salcedo, and S.F. de Azevedo [1994], "Nonequilibrium Simulated Annealing: A Faster Approach to Combinatorial Minimization," *Industrial Engineering and Chemical Research,* 33: 1908-1918.

Chardaire, P., J.L. Lutton, and A. Sutter [1995], "Thermostatistical Persistency: A Powerful Improving Concept for Simulated Annealing Algorithms," *European Journal of Operational Research,* 86: 565-579.

Charon, I. and O. Hudry [1993], "The Noising Method: A New Method for Combinatorial Optimization," *Operations Research Letters,* 14: 133-137.

Cheh, K.M., J.B. Goldberg, and R.G. Askin [1991], "A Note on the Effect of Neighborhood Structure in Simulated Annealing," *Computers and Operations Research*, 18(6): 537-547.

Chiang, T.S. and Y. Chow [1988], "On the Convergence Rate of Annealing Processes," *SIAM Journal on Control and Optimization*, 26(6): 1455-1470.

Chiang, T.S. and Y. Chow [1994], "The Asymptotic Behavior of Simulated Annealing Processes with Absorption," *SIAM Journal on Control and Optimization*, 32(5): 1247-1265.

Christoph, M. and K.H. Hoffmann [1993], "Scaling Behavior of Optimal Simulated Annealing Schedules," *Journal of Physics (A)*, 26: 3267-3277.

Cinlar, E. [1975], *Introduction to Stochastic Processes*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

Connors, D.P. and P.R. Kumar [1989], "Simulated Annealing Type Markov Chains and Their Order Balance Equations," *SIAM Journal of Control and Optimization*, 27: 1440-1461.

Crowder, H., R.S. Dembo, and J.M. Mulvey [1979], "On Reporting Computational Experiments with Mathematical Software," *ACM Transactions on Mathematical Software*, 5(2): 193-203.

Davis, T.E. [1991], "Toward an Extrapolation of the Simulated Annealing Convergence Theory onto the Simple Genetic Algorithm," Doctoral Dissertation, Department of Electrical Engineering, University of Florida, Gainesville, FL.

Dueck, G. and T. Scheuer [1990], "Threshold Accepting: A General Purpose Optimization Algorithm Appearing Superior to Simulated Annealing," *Journal of Computational Physics*, 90: 161-175.

Eglese, R.W. [1990], "Simulated Annealing: A Tool for Operational Research," *European Journal of Operational Research*, 46: 271-281.

Faigle, U. and W. Kern [1991], "Note on the Convergence of Simulated Annealing Algorithms," *SIAM Journal on Control and Optimization*, 29(1): 153-159.

Faigle, U., and W. Kern [1992], "Some Convergence Results for Probabilistic Tabu Search," *ORSA Journal on Computing*, 4(1): 32-37.

Faigle, U. and R. Schrader [1988], "On the Convergence of Stationary Distributions in Simulated Annealing Algorithms," *Information Processing Letters*, 27: 189-194.

Fleischer, M. [1994], "Assessing the Performance of the Simulated Annealing Algorithm Using Information Theory," Doctoral Dissertation, Department of Operations Research, Case Western Reserve University, Cleveland, OH.

Fleischer, M. [1995], "Simulated Annealing: Past, Present, and Future," *Proceedings of the 1995 Winter Simulation Conference*, (ed. C Alexopoulos, K. Kang, W.R. Lilegdon, and D. Goldsman, Crystal City, VA), 155-161.

Fleischer, M. and S.H. Jacobson [1996], "Information Theory and the Simulated Annealing Algorithm: Experimental Results," Technical Report, Department of Engineering Management, Old Dominion University, Norfolk, VA.

Fox, B.L. [1993], "Integrating and Accelerating Tabu Search, Simulated Annealing, and Genetic Algorithms," *Annals of Operations Research*, 41: 47-67.

Fox, B.L. [1995], "Simulated Annealing: Folklore, Facts, and Directions," *Lecture Notes in Statistics*, 106: 17-48.

Fox, B.L. and G.W. Heine [1993], "Simulated Annealing with Overrides," Technical Report, Department of Mathematics, University of Colorado, Denver, CO.

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

Gidas, B. [1985], "Nonstationary Markov Chains and Convergence of the Annealing Algorithm," *Journal of Statistical Physics*, 39: 73-131.

Glover, F. [1994], "Tabu Search for Nonlinear and Parametric Optimization (With Links to Genetic Algorithms)," *Discrete Applied Mathematics*, 49: 231-255.

Goldstein, L. and M. Waterman [1988], "Neighborhood Size in the Simulated Annealing Algorithm," *American Journal of Mathematical and Management Sciences*, 8: 409-423.

Golub, G.H. and C.F. Van Loan [1989], *Matrix Computations (Second Edition)*, The Johns Hopkins University Press, Baltimore MD.

Granville, V., M. Krivanek, and J.P. Rasson [1994], "Simulated Annealing: A Proof of Convergence," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(6): 652-656.

Hajek, B. [1988], "Cooling Schedules for Optimal Annealing," *Mathematics of Operations Research*, 13 (2): 311-329.

Hammersley, J.M. and D.C. Handscomb [1964], *Monte Carlo Methods*, Spottiswoode, Ballantyne and Company Ltd, London, England.

Hu, T.C., A.B. Kahing, and C.W.A. Tsao [1995], "Old Bachelor Acceptance: A New Class of Non-Monotone Threshold Accepting Methods," *ORSA Journal on Computing*, 7(4): 417-425.

Ipsen, I.C.F. and C.D. Meyer [1994], "Uniform Stability of Markov Chains," *SIAM Journal on Matrix Analysis and Applications*, 15(4): 1061-1074.

Isaacson, D.L. and R.W. Madsen [1976], *Markov Chains, Theory and Applications*, John Wiley and Sons, New York, NY.

Jacobson, S.H., A.W. Johnson, K.A. Sullivan, and M.A. Fleischer [1996], "Metaheuristics for a Flexible Assembly System Design/Loading Problem," Technical Report, Department of Industrial and Systems Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA.

Johnson, A.W. and S.H. Jacobson [1996], "A Convergence Result for Generalized Hill Climbing Algorithms", Technical Report, Department of Industrial and Systems Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA.

Koulamas, C., S.R. Antony, and R. Jaen [1994], "A Survey of Simulated Annealing Applications to Operations Research Problems," *Omega, International Journal of Management Science*, 22 (1): 41-56.

Kumar, A. and S.H. Jacobson [1996], "Analysis of the Design Problem in Flexible Assembly Systems," Technical Report, Department of Industrial and Systems Engineering, Virginia Polytechnic Institute and State University, Blacksburg,VA.

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

Law, A.M. and W.D. Kelton [1991], *Simulation Modeling and Analysis* (*Second Edition*), McGraw-Hill, Incorporated, New York, NY.

Liepins, G.E. and M.R. Hilliard [1989], "Genetic Algorithms: Foundations and Applications," *Annals of Operations Research,* 21: 31-58.

Lin, C.K.Y, K.B. Haley, and C. Sparks [1995], "A Comparative Study of both Standard and Adaptive Versions of Threshold Accepting and Simulated Annealing Algorithms in Three Scheduling Problems," *European Journal of Operational Research,* 83: 330-346.

Lundy, M. and A. Mees [1986], "Convergence of an Annealing Algorithm," *Mathematical Programming*, 34: 111-124.

Ma, J. and J.E. Straub [1994], "Simulated Annealing Using the Classical Density Distribution," *Journal of Chemical Physics*, 101(1): 533-541.

Metropolis, N., A.W. Rosenbluth, M.N. Rosenbluth, A.H.Teller, and E. Teller [1953], "Equation of State Calculations by Fast Computing Machines," *Journal of Chemical Physics*, 21(6): 1087-1092.

Meyer, C.D. [1975], "The Role of the Group Generalized Inverse in the Theory of Finite Markov Chains," *SIAM Review*, 17(3): 443-464.

Meyer, C.D. [1980], "The Condition of a Finite Markov Chain and Perturbation Bounds for the Limiting Probabilities," *SIAM Journal of Algebraic and Discrete Methods*, 1(3): 273-283.

Mitra, D., F. Romeo, and A. Sangiovanni-Vincentelli [1986], "Convergence and Finite-Time Behavior of Simulated Annealing," *Advances in Applied Probability*, 18: 747-771.

Moscato, P. [1993], "An Introduction to Population Approaches for Optimization and Hierarchical Objective Functions: A Discussion on the Role of Tabu Search," *Annals of Operations Research*, 41: 85-121.

Moscato, P. and J.F. Fontanari [1990], "Stochastic Versus Deterministic Update in Simulated Annealing," *Physics Letters (A)*, 146(4): 204-208.

Nissen, V. and H. Paul [1995], "A Modification of Threshold Accepting and its Application to the Quadratic Assignment Problem," *OR Spektrum*, 17: 205-210.

Ogbu, F.A. and D.K. Smith [1990], "The Application of the Simulated Annealing Algorithm to the Solution of the *n*/*m*/$C_{\max}$ Flowshop Problem," *Computers and Operations Research*, 17 (3): 243-253.

Papadimitriou, C.H. and K. Steiglitz [1982], *Combinatorial Optimization: Algorithms and Complexity*, Prentice-Hall, Englewood Cliffs, NY.

Press, W.H., S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery [1992], *Numerical Recipes in* C: *The Art of Scientific Computing* (*Second Edition*), Cambridge University Press, New York, NY.

Protter, M.H. and C.B. Morrey [1991], *A First Course in Real Analysis* (*Second Edition*), Springer Verlag, New York, NY.

Rabenstein, A.L. [1975], *Elementary Differential Equations with Linear Algebra (Second Edition)*, Academic Press, New York, NY.

Romeo, F. and A. Sangiovanni-Vincentelli [1991], "A Theoretical Framework for Simulated Annealing," *Algorithmica*, 6: 302-345.

Rosenthal, J.S. [1995], "Convergence Rates for Markov Chains," *SIAM Review*, 37(3): 387-405.

Ross, S.M. [1993], *Introduction to Probability Models* (*Fifth Edition*), Academic Press, Inc., San Diego, CA.

Rossier, Y., M. Troyon, and Th. M. Liebling [1986], "Probabilistic Exchange Algorithms and Euclidean Traveling Salesman Problems," *OR Spektrum*, 8: 151-164.

Rudolph, G. [1994], "Convergence Analysis of Canonical Genetic Algorithms," *IEEE Transactions on Neural Networks*, 5(1): 96-101.

Sangiovanni-Vincentelli, A. [1991], "Editor's Foreward," *Algorithmica*, 6(3): 295-301.

Scheermesser, T. and O. Bryngdahl [1995], "Threshold Accepting for Constrained Halftoning," *Optics Communications*, 115: 13-18.

Seneta, E. [1981], *Non-negative Matrices and Markov Chains (Second Edition)*, Springer-Verlag, New York, NY.

Solla, S.A., G.B. Sorkin, and S.R. White [1986], "Configuration Space Analysis for Optimization Problems," *NATO ASI Series, Vol F20, Disordered Systems and Biological Organization*, 283-293.

Stern, J.M. [1992], "Simulated Annealing with a Temperature Dependent Penalty Function," *ORSA Journal on Computing*, 4 (3): 311-319.

Storer, R.H., S.D. Wu, and R.Vaccari [1992], "New Search Spaces for Sequencing Problems with Application to Job Shop Scheduling," *Management Science*, 38(10): 1495-1509.

Straub, J.E., J. Ma, and P. Amara [1995], "Simulated Annealing using Coarse Grained Classical Dynamics: Smoluchowski Dynamics in the Gaussian Density Approximation," *Journal of Chemical Physics*, 103(4): 1574-1581.

Strenski, P.N. and S. Kirkpatrick [1991], "Analysis of Finite Length Annealing Schedules," *Algorithmica*, 6: 346-366.

Sullivan, K.A. [1996], "Optimization of Multistage Manufacturing Process Simulations Using Generalized Hill Climbing Algorithms," Final Report. Graduate Student Research Program, Wright Laboratory, Materials Directorate, Wright-Patterson AFB OH.

Tovey, C.A. [1988], "Simulated Simulated Annealing," *American Journal of Mathematical and Management Sciences*, 8: 389-407.

Tsitsiklis, J.N. [1989], "Markov Chains With Rare Transitions and Simulated Annealing," *Mathematics of Operations Research*, 14(1): 70-90.

Walker, R.M. [1992], "Manufacturing Technology for Advanced Propulsion Materials," Report WL-TR-92-8077, Wright Laboratory, Air Force Materiel Command, Wright-Patterson AFB, OH.

Yan, D. and H. Mukai [1992], "Stochastic Discrete Optimization," *SIAM Journal on Control and Optimization*, 30(3): 594-612.

Yücesan, E. and S.H. Jacobson [1996], "Computational Issues for Accessibility in Discrete Event Simulation," *ACM Transactions on Modeling and Computer Simulation*, 6(1): 53-75.

VITA

Alan W. Johnson [redacted]. He received a Bachelor of Science degree in Mechanical Engineering from Montana State University in 1982. He was commissioned as a lieutenant in the United States Air Force in 1984, and spent the next four years at Wright-Patterson Air Force Base, Ohio, engaged in the research and development of imaging systems for combat aircraft. He was selected in 1988 to attend the Air Force Institute of Technology. He completed a Master of Science degree in Systems Engineering Management as a Distinguished Graduate in 1989, and was subsequently assigned to a classified program at Los Angeles Air Force Base, California. In 1993, he was selected to attend graduate school at Virginia Polytechnic Institute and State University to obtain a Ph.D in Industrial and Systems Engineering, with subsequent duty as an instructor at the Air Force Institute of Technology. Alan was promoted to major in February 1996. He is a member of the Institute for Operations Research and the Management Sciences.

[redacted]

[redacted]