A Convergence Analysis of Generalized Hill Climbing Algorithms

Kelly A. Sullivan

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

Sheldon H. Jacobson, Chair
John E. Kobza
Joel A. Nachlas
Hanif Sherali
Keying Ye

April 12, 1999
Blacksburg, Virginia

Keywords: combinatorial optimization, convergence, discrete optimization, heuristics, hill climbing algorithms, Markov chain, simulated annealing, local search

A Convergence Analysis of Generalized Hill Climbing Algorithms

Kelly A. Sullivan

(ABSTRACT)

Generalized hill climbing (GHC) algorithms provide a unifying framework for describing several discrete optimization problem local search heuristics, including simulated annealing and tabu search. A necessary and a sufficient convergence condition for GHC algorithms are presented.

The convergence conditions presented in this dissertation are based upon a new iteration classification scheme for GHC algorithms. The convergence theory for particular formulations of GHC algorithms is presented and the implications discussed. Examples are provided to illustrate the relationship between the new convergence conditions and previously existing convergence conditions in the literature. The contributions of the necessary and the sufficient convergence conditions for GHC algorithms are discussed and future research endeavors are suggested.

This work received support from the Air Force Office of Scientific Research (F49620-95-1-0124, F49620-98-1-0111, F49620-98-1-0432).

Acknowledgments

I would first like to thank my advisor, Dr. Sheldon H. Jacobson, for his guidance and patience over the last several years. His knowledge, support, and advice were invaluable in my endeavor to complete this dissertation. I would also like to acknowledge the other members of my committee, Dr. John E. Kobza, Dr. Joel A. Nachlas, Dr. Hanif Sherali, and Dr. Keying Ye, as well as, Dr. Patrick C. Koelling for their encouragement and assistance.

I would like to thank Dr. Neal Glassman of the Air Force Office of Scientific Research (AFOSR) for sponsoring the research in this dissertation. I would also like to thank Dr. Steven LeClair, Dr. James Malas, and Dr. W. Garth Frazier of the Materials Directorate, Wright Laboratories, Wright-Patterson AFB, Ohio, and Mr. Enrique Medina of Austral Engineering and Software, Inc., Athens, Ohio, for the opportunity to participate in the AFOSR graduate student summer research fellowship program and their feedback on this line of research.

I would like to express my gratitude to Major Alan W. Johnson, Ph.D., for his unknowing guidance and mentoring. Even after he graduated, he provided guidance through his exceptional dissertation, which was the best reference and example any student could want.

I would also like to thank my parents, William and Ann Taylor, for their love and moral support. I especially want to thank my mother for instilling the importance and value of an education within me.

Finally, my deepest gratitude to Brian for his constant understanding, encouragement, and love throughout my graduate studies.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Discrete optimization (minimization) problems model many real-world systems. This creates the need to analyze and solve such problems quickly and efficiently. Discrete optimization problems can be defined by a countably finite set of solutions, as well as an objective function value assigned to each solution. Solving a discrete optimization problem requires finding solutions that globally optimize the objective function.

Many discrete optimization problems are NP-hard [43], hence a polynomial time algorithm does not exists that can solve such problems, unless P=NP. One method of finding global optima of a discrete optimization problem is complete enumeration over the entire solution space. However, if the solution space is large, then limited computing time and the associated computational costs cause complete enumeration to be highly inefficient. A significant amount of research attention continues to be focused on developing new heuristics to obtain approximate or near-optimal solutions to NP-hard discrete optimization problems in a reasonable amount of computing time.

Local search heuristics have enjoyed wide success and popularity in addressing discrete op-

timization problems, though there are associated disadvantages. For example, many local search heuristics are unable to state how close a particular solution is to a global optimum. Furthermore, such heuristics are often problem-specific, focusing on exploiting particular problem characteristics [62, 77]. In an attempt to resolve these issues, new, more effective and flexible local search heuristics are being developed to address NP-hard discrete optimization problems [3, 93]. Moreover, theoretical results are needed to determine the effectiveness of and provide performance measures for local search heuristics when applied to a specific discrete optimization problem. Such theoretical results include convergence conditions, initial parameter settings, recommended stopping criteria, and measures for the most effective local search heuristic given a fixed computing budget.

1.2 Research Goals

Several local search heuristics fall under the category of *hill climbing algorithms*. Such algorithms strive to escape local optima by accepting inferior solutions in an effort to reach a global optimum. Typically, restrictions are placed on the acceptance of an inferior solution to deter these algorithms from randomly searching the entire solution space [93].

Hill climbing algorithms that address intractable discrete optimization problems can be modeled using the *generalized hill climbing algorithm* framework [71, 72, 73]. The generalized hill climbing (GHC) algorithm framework encompasses many local search heuristics, including simulated annealing (SA) [76] and tabu search (TS) [46, 51], two widely applied hill climbing algorithms. SA is based upon an analogous relationship between discrete optimization and the physical annealing process. In the annealing process the energy of a system changes according to a cooling schedule until it converges to a steady (i.e., frozen) state. TS is an intelligent search technique guided by adaptive or flexible memory structures. One such memory structure is the *tabu list*, which records recently visited solutions as unacceptable for a specified number of iterations. The GHC algorithm allows visits to inferior solutions

according to a random (non-negative) hill climbing variable. The choice of this hill climbing random variable defines the acceptance probability function, which subsequently uniquely defines the GHC algorithm.

The popularity of SA has inspired several questions on the convergence of the algorithm. Much of the existing convergence theory of SA fixes the hill climbing variable as an exponential random variable. A convergence theory that addresses a more general hill climbing variable is introduced by Anily and Federgruen [5], but their sufficient condition for convergence is restrictive and can be difficult to verify. Furthermore, Anily and Federgruen do not present computational results comparing the performance of various hill climbing variables [68].

Johnson [71] and Johnson and Jacobson [73] provide sufficient convergence conditions for a general hill climbing variable. Their convergence result relaxes the previous sufficient conditions in the literature (predominantly in the SA literature). They also provide computational results that evaluate the finite-time performance of the GHC algorithm with different hill climbing variables.

The existing convergence theory of GHC algorithms provides sufficient convergence conditions for the algorithm when applied to a particular discrete optimization problem. The purpose of this research is to develop a new convergence theory for GHC algorithms that provides an alternative to the existing convergence theory.

1.3 Generalized Hill Climbing Algorithms

To describe the GHC algorithm framework, several definitions are needed. Define the *solution space*, Ω, to be the set of all solutions for a discrete optimization (minimization) problem, where $|\Omega|$ is assumed to be finite. Define a non-negative *objective function* $c : \Omega \rightarrow \Re^+$ that assigns a real value to each element of the solution space. The *neighborhood function* $\eta : \Omega \rightarrow 2^{\Omega}$, where $\eta(i) \subset \Omega$ for all $i \in \Omega$, is an important component of the GHC algorithm.

The neighborhood function determines adjacent solutions for all solutions in the solution space and a probability generating function determines a particular neighboring solution to be generated at each iteration.

GHC algorithms (depicted in pseudo-code in Figure 1.1) are initialized with a solution $i \in \Omega$, having objective function value c_i. The total number of outer loop iterations K, the total number of inner loop iterations $N(k)$, $k = 1, 2, \ldots, K$, and a set of hill climbing random variables $R_k(i, j)$, $i, j \in \Omega$, $j \in \eta(i)$ must all be specified.

The candidate solution j is chosen (generated) from among the set of neighbors of the current solution i, defined by $\eta(i)$. The probability of generating a candidate solution j among the neighbors of solution i at outer loop iteration k is g_{ij}^k, where

$$\sum_{j \in \eta(i)} g_{ij}^k = 1, \quad for \;\; i \in \Omega, \;\; k = 1, 2, \ldots, K. \tag{1.1}$$

The candidate solution j becomes the current solution according to the hill climbing random variable, which is uniquely defined by $R_k(i, j)$, $i, j \in \Omega$, $j \in \eta(i)$, $k = 1, \ldots, K$.

Initialization: specify the neighborhood function η and select an initial solution $i \in \Omega$
    Set the outer loop counter $k = 1$
    While iteration $k \leq K$:
        Set the inner loop counter $n = 1$
        While $n \leq N(k)$:
            Generate $j \in \eta(i)$ according to probability g_{ij}^k
            Calculate the change in objective function value $\Delta_{ij} = c_j - c_i$
            Accept solution j $(i \Leftarrow j)$ if $R_k(i, j) \geq \Delta_{ij}$
            $n \Leftarrow n + 1$
        $k \Leftarrow k + 1$

Figure 1.1: The Generalized Hill Climbing Algorithm

1.4 Research Questions

Two research questions are investigated:

1. What are necessary/sufficient conditions on the transition probabilities between solutions of the solution space Ω, such that a GHC algorithm will converge in probability to the set of globally optimal solutions?

2. How do these necessary/sufficient convergence conditions for GHC algorithms compare to existing convergence conditions in the literature?

Chapter 2

Literature Review

This chapter presents an overview of the literature pertaining to local search heuristics. The particular heuristics discussed are simulated annealing, threshold accepting, the noising method, tabu search, genetic algorithms, and generalized hill climbing algorithms. The history and development of each heuristic are described; their theoretical results and applications are also discussed.

2.1 Simulated Annealing

2.1.1 History and Development

Simulated annealing (SA) is motivated by an algorithm from statistical thermodynamics developed by Metropolis et al. [82] that simulates the cooling of material in a heat bath – known as *annealing*. Annealing is a thermal process that finds low energy states of solids. The annealing process initially melts a crystalline solid and then reduces the temperature slowly, spending a great deal of time at temperatures close to the freezing point. When a solid melts to a liquid, the particles are randomly arranged. When the liquid cools back into a solid state, the structural properties of the cooled solid depend upon the rate of cooling.

If the liquid cools too quickly, the resulting crystal contains imperfections. On the other hand, a slow cooling schedule allows the solid to obtain its ground state, where the particles are arranged in a lattice that achieves a minimum energy configuration. The Metropolis algorithm simulates the change in energy of a system under a particular cooling schedule, until convergence to a steady (i.e., frozen) state occurs.

The laws of thermodynamics state that the probability of an increase in energy of magnitude Δ at a specific temperature t is $p(\Delta) = exp(\frac{-\Delta}{\beta t})$ where β is the Boltzmann constant. At iteration k, the Metropolis algorithm generates a new solution j through a small displacement of a particle in the present solution i and calculates the consequent energy change, Δ_{ij}. The new solution j is accepted as the current solution according to the following probability, known as the *Metropolis criterion*,

$$P\{Accept\ \ solution\ \ j\ \ as\ \ new\ \ solution\} = \begin{cases} \exp(\frac{-\Delta_{ij}}{t_k}) & \Delta_{ij} > 0 \\ \\ 1 & \Delta_{ij} \leq 0 \end{cases} \tag{2.1}$$

where t_k is the temperature parameter at iteration k, such that

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

The algorithm follows a carefully designed cooling schedule of temperatures. At each temperature, a large number of iterations is carried out such that the solid can reach thermal equilibrium.

Approximately thirty years after this initial formulation, the Metropolis algorithm was modified and applied to discrete optimization problems by Kirkpatrick et al. [76] and, independently, by Cerny [16]. They draw analogies between the physical cooling process and discrete optimization problems, as described in Table 2.1. They show that a discrete optimization algorithm can be created by randomly searching the neighborhood of the current solution for a new solution via a neighborhood function and computing the change in the objective function. An inferior solution is accepted according to the probability given in (2.1). The

Table 2.1: Analogies Between Physical Cooling Process and Discrete Optimization Problems

Thermodynamic Simulation	Discrete Optimization
System States	Feasible Solutions
Energy	Cost
Change of State	Neighboring Solution
Temperature	Control Parameter
Frozen State	Heuristic Solution

temperature parameter and the amount of increase in the objective function influence the extent to which the algorithm hill climbs. When the temperature is high, the algorithm is likely to accept all solutions, hence hill climb, while a low temperature only allows the acceptance of better quality solutions.

2.1.2 Theoretical Results

Numerous theoretical convergence results of SA have been published since Kirkpatrick [76] proposed the application of the annealing algorithm to discrete optimization problems. The SA algorithm can be modeled using Markov chain analysis. Theoretical results exist for the algorithm modeled as either a homogeneous Markov chain or an inhomogeneous Markov chain. These results show that generic and problem-specific decisions within the algorithm need to be made in order to guarantee convergence. In addition, there has been much research on the statistical behavior of the algorithm, including such issues as determining suitable objective functions, neighborhood functions, and parameters for the cooling schedule [93].

Homogeneous Markov Chain Models

When the temperature parameter t_k is kept constant for a sufficient number of iterations, the probability of moving from one solution to another at iteration k fixed may be represented

using a transition matrix, P^k, where the transition probability of moving from solution i to solution j, P^k_{ij}, depends only on i and j, for all $i, j \in \Omega$. This representation corresponds to a homogeneous Markov chain and is used by Aarts and van Laarhoven [2], Duque-Anton [34], Faigle and Kern [36], Faigle and Schrader [38], Granville et al. [55], Lundy and Mees [81], and Schuur [100].

This homogeneous Markov chain model representation has a unique stationary probability π^k_i at iteration k, for all $i \in \Omega$. All of the SA proofs of convergence based on the homogeneous Markov chain require the sufficient condition of *weak reversibility* (also known as *detailed balance*) [95], defined as

$$\pi^k_i P^k_{ij} = \pi^k_j P^k_{ji}, \quad for \ \ all \ \ i, j, \in \Omega, \ \ and \ \ all \ \ iterations \ \ k. \tag{2.3}$$

A necessary condition for reversibility is multiplicativity [95], defined as

$$a^k_{hj}(\Delta_{hj}) = a^k_{hi}(\Delta_{hi}) a^k_{ij}(\Delta_{ij}), \quad for \ \ all \ \ iterations \ \ k, \tag{2.4}$$

where $a^k_{hi}(\Delta_{hi})$ is the probability of accepting the transition from solution h to solution i at iteration k. Reversibility (2.3) is guaranteed when

1. the transition matrix P^k is symmetric, and
2. the acceptance probability function is expressed in an exponential form or the multiplicative condition (2.4) is satisfied.

The multiplicative condition (2.4) is required for all the homogeneous proofs of convergence in the literature. Both Aarts and van Laarhoven [2] and Lundy and Mees [81] present proofs of convergence for the SA algorithm that require a symmetric transition matrix P^k and the multiplicative condition for the acceptance function. Rossier et al. [96] represent the acceptance function as a ratio of the stationary probabilities. They also partition the solution space into blocks and require that the transition probabilities be symmetric among blocks. Faigle and Shrader [38] and Faigle and Kern [36] describe a graph theoretic approach

that relaxes the transition probability condition of symmetry, though condition (2.4) is still required.

Granville et al. [55] apply the SA algorithm for the filtering of binary images. They use an acceptance function based on the probability of the current solution, rather than on the objective function of the candidate solution. A proof of asymptotic convergence is presented, but the proof does not show that the set of globally optimal solutions are asymptotically uniformly distributed. Duque-Anton [34] uses a homogeneous Markov chain representation to obtain an efficient method to construct neighborhood functions and transition probabilities for SA. His method assigns equivalent configurations to the same equivalence class and the algorithm searches these equivalence classes, including the optimal configuration equivalence class. Schuur [100] presents a description for the class of acceptance functions that yields detailed balance for any symmetric transition matrix P^k.

Inhomogeneous Markov Chain Models

If the temperature parameter t_k is not kept constant, but rather reduced after a certain number of iterations, the SA algorithm can be modeled as a sequence of homogeneous Markov chains of finite length, or as an inhomogeneous Markov chain, where the probabilities of the transition matrix, P^k, are dependent on the number of iterations already executed. The inhomogeneous Markov chain approach is used by Anily and Federgruen [5], Belisle [9], Borkar [12], Chiang and Chow [20, 21], Connors and Kumar [24], Gidas [45], Hajek [58], Mitra et al. [83], and Trouve [110].

Mitra et al. [83] present a convergence proof that requires the conditions of *weak* and *strong ergodicity* [66, 101]. They also find a bound on the distance between the actual solution probability distribution and the optimal solution probability distribution after a finite number of iterations. Anily and Federgruen [5] implement more general acceptance probability functions into the SA algorithm and present a proof of convergence. The general acceptance functions must allow the acceptance of an inferior solution with a positive probability and

must be bounded and asymptotically monotone. Hajek [58] provides necessary and sufficient conditions for convergence. He derives a cooling schedule, dependent on the shape of the objective function over the neighborhood function, given by

$$t_k = \frac{c}{log(1+k)} \tag{2.5}$$

where k is the iteration number. He shows that if c is greater than or equal to the depth of the deepest local minimum which is not a global minimum, then asymptotic convergence is guaranteed.

Connors and Kumar [24] employ a concept called *orders of recurrence* to substantiate Hajek's necessary and sufficient conditions. Their SA inhomogeneous Markov chain converges in a *Cesaro* sense to the set of solutions having the largest recurrence orders. Similar to the theory of Connors and Kumar [24], Borkar [12] proves that a SA inhomogeneous Markov chain converges in a *Cesaro* sense by redefining the recurrence orders "pathwise" and exploiting a convergence/oscillation dichotomy result for martingales. Belisle [9] introduces an adaptive, rather than deterministic, cooling schedule and presents a convergence result consistent with that of Hajek [58]. Trouve [110] uses a inhomogeneous Markov chain approach to decompose the solution space into cycles. He proposes an algorithm that computes a cycle decomposition that may be useful when studying the exact asymptotic behavior of SA on small state spaces.

Generic Decisions

The generic decisions for the SA algorithm involve the cooling schedule and the stopping criterion. The cooling schedule consists of an initial temperature, t_0, a temperature function that determines how to reduce the temperature, and a stopping criterion. It is believed that if schedules cool over the same range of temperatures at approximately the same rate, then the choice of cooling schedules does not greatly effect the performance of the SA algorithm. Moreover, the best parameters for a cooling schedule are determined through extensive experimentation when first applying SA [74, 75, 93].

Lundy and Mees [81] suggest a cooling schedule where the algorithm executes only one iteration at each temperature, but reduces the temperature very slowly. At iteration k, their temperature reduction function takes the form

$$t_k = \frac{t_{k-1}}{1 + \beta t_{k-1}} \tag{2.6}$$

where t_0 is defined as the initial temperature and β is a suitably small value.

Hajek [58] develops a cooling schedule (2.5) that reduces the temperature at such a slow rate that the schedule is not feasible in practice. Also, Hajek's cooling schedule has not proven to be very useful since the depth of local optima is difficult to estimate.

Aarts and Korst [1] develop a cooling schedule that guarantees that the final distribution of the system will be sufficiently close to the stationary distribution (i.e., quasi-equilibrium). After starting with an initial temperature t_0, the temperature is reduced at iteration k according to the formula

$$t_k = \frac{t_{k-1}}{1 + \frac{t_{k-1} \ln(1+\Delta)}{3\sigma_{t_{k-1}}}} \tag{2.7}$$

where $\sigma_{t_{k-1}}$ is the standard deviation of the objective function values at temperature t_{k-1}. This particular cooling schedule is slower than cooling schedules actually implemented in practice, hence leads to large run times.

Although SA theory suggests that the temperature should decrease to zero before the algorithm stops, as the temperature approaches zero, the small probability of accepting an uphill move is indistinguishable from zero. Hence, there is no need to decrease the temperature to zero before stopping the algorithm. It is best to stop the algorithm when the probability that the algorithm will escape from the current solution to a superior solution is small. In order to find a solution that is within ϵ of the global optimum with probability θ, Lundy and Mees [81] recommend stopping the algorithm when

$$t_k \leq \frac{\epsilon}{\frac{\ln[(|\Omega|-1)}{\theta}}, \tag{2.8}$$

where Ω is the solution space. Of course, the simplest stopping criterion is to stop the algorithm after a specified number of iterations are executed. This simple rule needs to be

carefully fine-tuned against other parameters, such as the temperature parameter, so that the SA algorithm converges to the global optimum at the fastest rate possible.

Problem-Specific Decisions

When applying SA to a particular discrete optimization problem, there exists problem-specific decisions concerning how to define the neighborhood function and the objective function. Furthermore, the solution space for any given problem must be clearly defined. Once the solution space is set, the neighborhood function and the objective function are defined such that the SA algorithm can be implemented.

It has been suggested that the size of the neighborhood influences the efficiency of the SA algorithm [84]. There are several different views regarding the optimal size of a neighborhood. Cheh et al. [19] demonstrate that a small neighborhood size is beneficial for some problems. However, Yao [114] suggests that a large neighborhood size improves the performance of SA. Goldstein and Waterman [53] state that a neighborhood size that is too small or too large can hinder the performance of the algorithm. A neighborhood size that is too small can lead to slow (or even no) convergence. On the other hand, a neighborhood size that is too large can lead to premature convergence. Given Hajek's [58] result, it is not surprising that the best neighborhood function imposes a "smooth" topology over the solution space rather than a "spiky" topology, where there are many deep local minima. In an effort to resolve these neighborhood function issues, Fleischer [40] and Fleischer and Jacobson [41] model SA as a *Markov information source*. By applying information theoretic concepts, they show that the neighborhood function can affect the information rate or level of total uncertainty associated with SA. Fleischer [40] shows that as the level of entropy of the associated Markov chain increases, the finite-time performance of the SA algorithm improves.

It is important to choose the neighborhood function and the objective function so that the value of a solution can be easily computed. Every objective function is unique to the problem being analyzed. The objective function is an important factor in the effectiveness of the SA

algorithm, but a suitable objective function is not always obvious. When the solution space does not consist entirely of feasible solutions, a penalty function can be integrated into the objective function; a penalty function is useful when comparing two infeasible solutions to determine which one is closer to feasibility.

2.1.3 Applications

The SA algorithm has been applied to numerous intractable discrete optimization problems. Examples of these include flowshop sequencing [90], DNA mapping [53], and image processing [44]. The SA algorithm has also been applied to a variety of classical discrete optimization problems. Johnson et al. [74, 75] address the graph partitioning, graph coloring, number partitioning, and traveling salesman problems using SA. They apply various cooling schedules and neighborhood functions, where the results are compared with results from other heuristic techniques. Several other classical problems are investigated by Connolly [25] (quadratic assignment problem) and Dowsland [31] (the Steiner problem).

The existing applications of SA in a variety of areas demonstrate the robustness of the algorithm. In general, the SA algorithm is easy to implement, can be applied to most discrete optimization problems, and usually provides acceptable solutions. The main drawback of the SA algorithm is the excessively long run time needed for convergence to global optima.

2.2 Threshold Accepting

2.2.1 History and Development

Threshold accepting (TA), proposed by Dueck and Scheuer [33] and independently by Moscato and Fontanari [85], follows the same structure as SA; the acceptance criteria for new solutions is what distinguishes the two algorithms. TA adopts a simpler deterministic acceptance

criteria, defined by a threshold parameter. The probability of accepting a transition to an inferior solution j from the current solution i at iteration k is given by

$$P\{Accept\ \ solution\ \ j\ \ as\ \ new\ \ solution\} = \begin{cases} 0, & if\ \ \Delta_{ij} > Q_k \\ 1, & if\ \ \Delta_{ij} \leq Q_k \end{cases} \tag{2.9}$$

where Q_k is the value of the threshold parameter at iteration k.

Several modifications to the TA algorithm exist. Freisleben and Schulte [42] present a parallel adaptive TA algorithm. They partition the traveling salesman problem into subproblems and use an adaptive TA algorithm to solve the subproblems in parallel. Dueck [32] introduces the great deluge algorithm and the record-to-record travel algorithm with principles taken from the general TA algorithm. Both of these modified algorithms differ from the original TA algorithm in their acceptance criteria for inferior solutions. Dueck concludes that the new algorithms perform as well as the original TA algorithm. A TA strategy called Old Bachelor Acceptance (OBA) is proposed by Hu et al. [65]. The OBA approach differs from the original TA algorithm in three ways:

1. the method is specifically motivated by the practicality of a prescribed time constraint
2. the threshold schedule is self-tuning
3. the threshold schedule is non-monotone and negative threshold values are permitted.

Under this formulation, the original TA algorithm becomes a special case of the OBA method. They conclude that the OBA approach to discrete optimization problems may outperform other hill climbing algorithms under computing time constraints. Lin et al. [80] present an adaptive TA algorithm that uses recent search performance to guide the algorithm's future performance. Their computational results for three scheduling problems show that TA algorithms perform as good (or sometimes better) than SA with respect to solution quality and average computational time. Nissen and Paul [86] introduce a threshold function into the TA algorithm, similar to the cooling schedule of SA, and analyze the performance.

2.2.2 Theoretical Results

Despite the similarities between SA and TA, research concerning the theory of TA is quite scarce. Dueck and Scheuer [33] claim that their TA algorithm produces superior results to those of SA with respect to the run time and the number of "new configuration choice steps." A convergence result is derived by Althofer and Koschnick [4]. They demonstrate that, in some sense, each execution of SA lies within the convex hull of a set of TA executions and conclude that TA is provably good; this convergence result is weaker than convergence results for SA. Jacobson and Yücesan [69] prove that if $Q_k \to 0$ as $k \to +\infty$, then TA does not converge to the set of global optima.

2.2.3 Applications

The TA algorithm has been quite successful in applications due to the ease of implementation and faster execution time (compared to SA). However, the number of applications which have used TA is small compared to SA. Scheermesser and Bryngdahl [99] present an application of the TA algorithm to a digital halftoning problem. The TA algorithm is successfully applied by Chipman and Winker [22] to solve a problem concerning econometric models and by Lidia and Carr [78] to solve a magnet sorting problem. Lin et al. [80] apply the TA algorithm to three scheduling problems, while Nissen and Paul [86] apply their modified TA algorithm to the quadratic assignment problem. Winker and Fang [113] use TA to evaluate the discrepancy of a given set of points.

2.3 The Noising Method

2.3.1 History and Development

Charon and Hudry [17] propose the noising method based on a simple descent algorithm. To implement the noising method, the solution space is altered by adding noise to the problem's objective function values. The amount of noise added decreases as the algorithm proceeds. At the last iteration, no noise is added and the final solution reported is the best solution obtained during the execution of the algorithm.

2.3.2 Theoretical Results

There has been no published research regarding the theory of the noising method. Charon and Hudry [17] postulate that the algorithm performs as good as (or even better than) SA, but there are no theoretical results to substantiate their claim.

2.3.3 Applications

Published applications of the noising method are as scarce as the theoretical results. There exists literature on the algorithm applied to the clique partitioning problem [17] and to the problem of constructing covering nodes [18]. Charon and Hudry [17] have results, yet to be published, on the application of the noising method to the traveling salesman problem and a specific voting theory problem. Sudhakar and Murthy [104] apply a modified noising method to the graph partitioning problem. Their results show that for this particular graph partitioning problem, the modified noising method outperforms the original noising method and SA, with regard to run time and the quality of solutions obtained.

2.4 Tabu Search

2.4.1 History and Development

Tabu search (TS) is introduced and described by Glover [46] and Glover and Laguna [51], and independently by Hansen [59] . The algorithm can be described as an intelligent search technique guided by adaptive or flexible memory structures. Similar to SA and TA, inferior solutions to the current solution may be accepted when applying TS; but, unlike SA and TA, TS uses historical information gathered through memory structures [46] to guide the algorithm to global optima. The memory structures allow TS to intensify or diversify the search (when necessary) in an effort to escape local optima.

The short-term memory structure in the algorithm is the *tabu list*. The tabu list, of length T, characterizes the solutions visited in the last T iterations and classifies these solutions as tabu. The purpose of the tabu list is to prevent the algorithm from cycling back to a recently visited solution. After a new solution is generated, the tabu status is evaluated. If the new solution is not tabu, it is accepted as the current solution, even if it is an inferior solution. On the other hand, a tabu solution is only accepted if it satisfies the predefined aspiration criteria. A common aspiration criterion used for overriding the tabu status consists of allowing a tabu solution to become the current solution if it is the best solution to date. There are also long-term memory structures that can be incorporated into TS [49]. For example, the frequency of previously visited solutions can be used to drive TS into regions not previously visited. There are many other suggestions of how to refine the TS algorithm and exploit its flexible memory structures [48, 49].

Several authors have modified the TS algorithm. Glover [47] introduces a probabilistic acceptance function into the TS algorithm. Battiti and Tecchiolli [7] propose a reactive TS, which reacts to the occurrence of cycles by adjusting the size of the tabu list throughout the search. Initially the size of the tabu list is set to one. If the algorithm repeatedly appears to be visiting previously visited solutions (i.e., cycles), then the tabu list increases in length,

hence promoting diversification of the search. Furthermore, if no cycles are observed for a sufficient period of time, then the tabu list is decreased, hence promoting intensification of the search. Glover [50] presents a hybrid of SA, TA and TS, termed tabu thresholding. Glover's tabu thresholding algorithm consists of two phases, the *Improving Phase* and the *Mixed Phase*. The Improving Phase applies a local search method to the solution space, resulting in a local optimum. Conversely, the Mixed Phase allows the acceptance of inferior solutions. The algorithm moves back and forth between the two phases. The number of iterations the algorithm spends in the Mixed Phase is determined by a random or deterministic threshold schedule. Computational results for tabu thresholding are not reported.

2.4.2 Theoretical Results

Similar to TA and the noising method, very little published theoretical work has been reported for TS. Faigle and Kern [37] provide a convergence result for a probabilistic version of TS. Probabilistic TS incorporates the acceptance function of SA into the TS framework. Faigle and Kern [37] prove that probabilistic TS converges asymptotically by using methodology employed in a previous SA proof of convergence [36]. Similarly, Tian et al. [109] introduce the acceptance criteria of SA into the framework of TS. They show that this new stochastic TS converges asymptotically to the global optima, where the rate of convergence is faster than that of SA.

2.4.3 Applications

TS has been applied to numerous discrete optimization problems, with most of these applications occurring over the past ten years. The popularity of TS can be attributed to its efficiency and flexibility. Widmer and Hertz (1989) were amongst the first to implement TS to address scheduling problems. Scheduling continues to provide a fruitful area for the application of TS. Other applications explored by TS include problems in flow shop sequenc-

ing [106], transportation [89], layout and circuit design [107], probabilistic logic and expert systems [70, 60], telecommunications [88], graphs [61], and neural networks [29].

2.5 Genetic Algorithms

2.5.1 History and Development

Genetic algorithms (GA) are an adaptive heuristic search technique based on evolutionary concepts of natural selection and genetics. The first concrete research on GA was published by Holland [63] and contains much of the initial theory. GA were first proposed by Holland and his colleagues at the University of Michigan during the 1960's and 1970's. Initially, the research focused on the area of artificial intelligence and function optimization. Only recently has the research turned toward operations research applications, such as discrete optimization problems. Within the past several years, GA emerged as a very versatile and efficient method to address discrete optimization problems; this development has resulted in an increasing amount of research devoted to the application, practice, and theory of GA.

GA seek optimal solutions to complex problems by combining sections of existing solutions, similar to how offspring are genetically reproduced. Genetic operators, such as *crossover* and *mutation*, are performed on the solutions. Crossover consists of exchanging sections of existing solutions. Mutation consists of randomly modifying an existing solution (e.g., a permutation of its sections). At each iteration, GA maintain a population of current solutions (corresponding to parents) whose objective function values are known. Solutions from this population are chosen to crossover (corresponding to mating) with each other. This crossover operation is performed randomly and/or based on the objective function values of the current solutions. The crossover operation produces new solutions (corresponding to offspring); after applying mutation, the resulting solutions become the current population. This process is repeated until a stopping criterion is satisfied. Several types of crossover and

mutation operators exist; the choice of these operators significantly impacts the performance of the algorithm [26, 93].

2.5.2 Theoretical Results

Much of the current GA theory is based on the initial theory presented by Holland [63]; his fundamental theorem is the *Schema Theorem*, where the term *schema* refers to the similar beneficial parts among solutions. The schema theorem states that individual solutions with good, low order schema should be evaluated and allowed to crossover in an exponentially increasing number of successive populations. Schaffer [98] analyzes the schema theorem and how it is affected by the solutions chosen to crossover. Bridges and Goldberg [13] extend the Schema Theorem by calculating the expected number of a given schema in the solution space.

It is important for GA to converge to global optima at an *acceptable rate*. Research conducted on the convergence of GA has focused on one of the following two issues (Patnaik and Srinivas [103]):

1. finite versus infinite populations
2. homogeneous versus inhomogeneous convergence.

The current focus of GA convergence theory research is on the parameters of the algorithm, such as population size and mutation probabilities, and how these parameters should be set such that the algorithm converges to a global optimum at a rate of convergence that is greater than a predetermined bound.

Grefenstette [56] proposes optimal control parameters that enable GA to balance the exploitation of previously sampled regions and the exploration of new regions in the solution space. DeJong and Spears [30] perform a similar analysis on the control parameters, but base their results on a large population size (100 solutions versus the 30 solution population

of Grefenstette); due to the larger population, they report different optimal control parameters. A convergence analysis of GA is presented by Eiben et al. [35]; the authors use a Markov chain analysis to obtain a unifying theory for GA and SA. Davis [27] and Davis and Principe [28] present a convergence theory based on the convergence of SA. Nix and Vose [87] and Vose [111] demonstrate how their defined GA can be modeled as a Markov chain and Liepins [79] uses a homogeneous Markov chain model to prove global convergence for a specific type of GA. Horn [64] applies finite Markov chain analysis for GA with niching. A finite Markov chain analysis is also used by Rudolph [97] to analyze the convergence of canonical GA. Barrios et al. [6] develop a convergence theory using a Walsh expansion of the objective function. Vose [112] presents a convergence theory of GA that concerns a logarithmic convergence rate. Thierens and Goldberg [108] model the convergence of different GA selection schemes by applying concepts of normal distribution theory. Cerf [14, 15] employs concepts from Freidlin-Wentzell theory (i.e., the study of random perturbations of dynamical systems) to show asymptotic convergence for a modified GA. By modeling GA as a random perturbation of a simple selection scheme, Cerf proves convergence to the global optimum in terms of the population size. Suzuki [105] uses a Markov chain analysis to model GA and computes a convergence rate in terms of the mutation probability for GA that employs the modified elitist strategy (i.e., the solution with the best objective value of the current population remains in the next population). Since the inception of GA by Holland, the quantity and quality of research conducted in the field has been increasing. Much of the existing GA convergence theory is directed towards particular algorithm formulations. Therefore, the main issues in the investigation of GA are the need for theoretical results regarding convergence of a general GA formulation and how the algorithm executes.

2.5.3 Applications

The number of successful applications of GA in the literature suggests that GA are a powerful and robust optimization technique, since the crossover and mutation operators enable GA

to quickly and effectively search the solution space. GA have been successfully applied to numerous discrete optimization problems, such as the traveling salesman problem [57], sequencing and scheduling problems [8, 92], graph coloring [26], knapsack problem [39, 52], set covering problem [94], and the bin packing problem [91]. As reported by Srinivas and Patnaik [103], the versatility of GA allows for applications in other diverse areas, such as music generation, machine learning, and genetic synthesis.

2.6 Generalized Hill Climbing Algorithms

2.6.1 History and Development

The general acceptance probability model, generalized hill climbing (GHC) algorithms, is introduced by Johnson [71] and Johnson and Jacobson [72, 73]. They propose a hill climbing algorithm that describes several local search heuristics, based on the choice of the hill climbing random variable. SA, TA, TS, Monte Carlo search, local search, and Weibull accepting [71] are all particular GHC algorithms, defined uniquely through the hill climbing random variable. Pseudo-code for GHC algorithms is presented in Figure 1.1.

The GHC algorithm accepts a transition at outer loop iteration k from the current solution i to an inferior solution j according to the following probability

$$P\{Accept\ \ solution\ \ j\ \ as\ \ new\ \ solution\} = \begin{cases} 1 & R_k(i,j) \geq \Delta_{ij} \\ \\ 0 & otherwise \end{cases} \tag{2.10}$$

where $R_k(i,j)$ is the value of the hill climbing random variable for $i, j \in \Omega$, $j \in \eta(i)$ at outer loop iteration k. Using this representation (2.10), if $R_k(i,j) = -t_k \ln(u)$, $i, j \in \Omega$, $j \in \eta(i)$, where t_k is a temperature parameter and u is a $U(0,1)$ random variable, the resulting GHC algorithm is SA (see (2.1)). If $R_k(i,j) = Q_k$, $i, j \in \Omega$, $j \in \eta(i)$, where Q_k is a non-negative real constant, the resulting GHC algorithm is TA (see (2.9)). If $R_k(i,j) = +\infty$, $i, j \in \Omega$,

$j \in \eta(i) \equiv \Omega$, the resulting GHC algorithm is Monte Carlo search (i.e., all new solutions are accepted, even if the new solution is inferior). If $R_k(i,j) = 0$, $i,j \in \Omega$, $j \in \eta(i)$, the resulting GHC algorithm is local search (i.e., only improving solutions are accepted). If $R_k(i,j) = t_k(-\ln(u))^{\frac{1}{\alpha}}$, $i,j \in \Omega$, $j \in \eta(i)$, where t_k is a temperature parameter, u is a $U(0,1)$ random variable and $\alpha > 0$ is a shape parameter, the resulting GHC algorithm is Weibull accepting, since $R_k(i,j)$ is distributed Weibull with mean $t_k\Gamma(1+(\frac{1}{\alpha}))$. If $R_k(i,j) = +\infty$, $i,j \in \Omega$, $j \in \eta(i)$, for $j \notin \mathcal{L}$, and $R_k(i,j) = -\infty$, for $j \in \mathcal{L}$, where $\mathcal{L}$ is a tabu list of solutions, the resulting GHC algorithms is TS. In fact, defining $R_k(i,j)$, $i,j \in \Omega$, $j \in \eta(i)$, to be any random variable results in a unique GHC algorithm.

2.6.2 Theoretical Results

Theoretical results regarding the convergence of GHC algorithms are presented by Johnson [71] and Johnson and Jacobson [72, 73]. Johnson and Jacobson [72] present a convergence result for a particular class of GHC algorithms. The particular class of GHC algorithms requires that the globally optimal objective function value be known. Johnson [71] and Johnson and Jacobson [73] introduce the notion of paths, defined as a sequence of solutions between global optimum and/or local optimum, where the intermediate solutions in the sequence are not global or local optima. Two paths can either be *equivalent* to each other or *distinct* from each other. Two paths are equivalent to each other if all of the solutions visited along both paths are identical and if the solutions are visited in the same order. A path that is not equivalent to any other path is said to be distinct. From these definitions, path probabilities are constructed. Using a homogeneous Markov chain representation without reversibility (detailed balance), Johnson [71] and Johnson and Jacobson [73] present sufficient convergence conditions for GHC algorithms. These sufficient conditions relax the most general proof for convergence currently in the literature ([5]) since their theoretical results do not require reversibility and exponentially distributed acceptance functions, often required in past proofs of convergence. According to Johnson [71] and Johnson and Jacobson [73],

SA and the noising method may converge if defined correctly. However, local search and TA do not meet the sufficient conditions for convergence (though this does not imply that local search and TA do not converge to the set of globally optimal solutions).

2.6.3 Applications

Though there are only a few applications under the GHC framework reported, specific GHC formulations (i.e., SA, TA, TS) have been applied numerous times to a variety of discrete optimization problems. Johnson [71] applies GHC algorithms to a flexible assembly system design problem, a generic configuration space problem, and an Air Force manufacturing process design problem. Jacobson et al. [67, 68] present more extensive results of GHC algorithms applied to the same Air Force manufacturing process design problem.

Chapter 3

Background and Definitions

The first two sections of this chapter contain definitions, notation, and lemmas from matrix theory and Markov chain theory that are applied in subsequent chapters. The third section provides definitions needed to discuss GHC algorithms in greater depth and to introduce a new iteration classification scheme. The final section describes the convergence results for GHC algorithms in Johnson [71] and Johnson and Jacobson [73].

3.1 Matrix Definitions and Notation

To define *irreducibility* and *aperiodicity* for a matrix A, suppose $A = [a_{ij}]$ is an $n \times n$ stochastic non-negative matrix, where $A^{(x)} = [a_{ij}^{(x)}]$. The entry $a_{ij}^{(x)}$ represents the probability of moving from state i to state j after x transitions.

State j is said to be *accessible* from state i (i.e., $i \rightarrow j$) if $a_{ij}^{(x)} > 0$ for some $x \geq 0$. Two states i and j that are accessible from each other are said to *communicate* (i.e., $i \leftrightarrow j$). The matrix A is *irreducible* if for all $i, j = 1, \ldots, n$, there exists a positive integer x such that $a_{ij}^{(x)} = P\{i \rightarrow j \ \ in \ \ x \ \ transitions\} > 0$ (i.e., all states communicate with each other).

A state i is said to have *period* d if $a_{ii}^{(c)} = 0$ whenever c is not divisible by d and d is

the greatest integer with this property. A state with period 1 is said to be *aperiodic*. An irreducible matrix with all states aperiodic is said to be an aperiodic matrix. A non-negative $n \times n$ matrix A is said to be *reducible* if its rows and columns can be rearranged simultaneously so that the resulting matrix has the form

$$A = \begin{pmatrix} A_{11} & A_{12} \\ \\ 0 & A_{22} \end{pmatrix} \tag{3.1}$$

where A_{11} and A_{22} are square matrices. Note that the eigenvalues of A are the eigenvalues of A_{11} and A_{22} [54].

Notation for matrices is now introduced. Suppose $A = [a_{ij}]$ and $B = [b_{ij}]$ are matrices of order $(m \times n)$. Then

$$A \geq B \;\; if \;\; a_{ij} \geq b_{ij} \;\; for \;\; all \;\; i = 1, \ldots, m, \;\; j = 1, \ldots, n,$$

$$A > B \;\; if \;\; a_{ij} > b_{ij} \;\; for \;\; all \;\; i = 1, \ldots, m, \;\; j = 1, \ldots, n.$$

If $\omega_1, \omega_2, \ldots, \omega_n$ are the eigenvalues of A (an $n \times n$ matrix), then

$$\rho(A) = \max_i |\omega_i|, \;\; for \;\; i = 1, \ldots, n,$$

is called the *spectral radius* of A.

3.2 Matrix Lemmas

Berman and Plemmons [10] and Graham [54] present proofs of the following lemmas.

Lemma 3.1 *If $A \geq 0$ is irreducible and $B \geq 0$, then $A+B$ is irreducible.*

Lemma 3.2 *If $0 \leq B \leq A$ then $\rho(B) \leq \rho(A)$.*

Lemma 3.3 *If $0 \leq B \leq A$, $A \neq B$, and $A + B$ is irreducible, then $\rho(B) < \rho(A)$.*

Lemma 3.4 *If* $\rho(A) < 1$*, then*

$$(I - A)^{-1} = \sum_{q=1}^{+\infty} (A)^{q-1}.$$

Lemma 3.5 *Consider a matrix* $A = [a_{ij}]$ *of order* $(n \times n)$*. If* $a_{ii} > 0$ *for all* $i = 1, \ldots, n$*, then* A *is an aperiodic matrix.*

3.3 Definitions for GHC Algorithms

The solution space of a discrete optimization problem, Ω, can be decomposed into three mutually exclusive sets:

1. $G = \{G_1, G_2, \ldots, G_\gamma\}$, the set of γ global optima,
2. $L = \{L_1, L_2, \ldots, L_\lambda\}$, the set of λ local (but not global) optima,
3. $H = \{H_1, H_2, \ldots, H_\phi\}$, the set of ϕ hill solutions,

where $\Omega = G \cup L \cup H$, $G \cap L = G \cap H = H \cap L = \emptyset$. Moreover, the neighborhood function η is defined such that for all $g \in G$ and all $l \in L$,

$$\eta(g) \backslash g \subseteq H \quad \textit{and} \quad \eta(l) \backslash l \subseteq H. \tag{3.2}$$

Hence, (3.2) states that the neighborhoods of local and global optima contain only hill solutions. The iterations of a GHC algorithm can be classified as either micro or macro iterations. A *micro iteration* is an iteration that moves the algorithm from the current solution either to an immediate neighbor or back to itself. A *macro iteration* moves the algorithm from a global optimum or a local optimum to any global optimum or any local optimum, passing only through elements of H.

The iterations of a GHC algorithm can be modeled as a sequence of homogeneous discrete-time Markov chains. This occurs when, for a fixed outer loop iteration k of the GHC algorithm, R_k is kept constant for $N(k)$ inner loop iterations before being changed. Note that if $N(k)$ represents the number of visits to elements of H between visits to elements of $L \cup G$, at outer loop iteration k, then the outer loop iterations can be viewed as macro iterations, and the inner loop iterations can be viewed as micro iterations.

For a fixed macro iteration $k \in Z^+$, the micro transition probability from the current solution i to a new solution j is defined as

$$P_{ij}^k = \begin{cases} g_{ij}^k Pr\{R_k(i,j) \geq \Delta_{ij}\} & for\ \ all\ \ i \in \Omega, j \in \eta(i), j \neq i \\ 1 - \sum_{l \in \eta(i), l \neq i} P_{il}^k & j = i \\ 0 & otherwise \end{cases} \tag{3.3}$$

where g_{ij}^k satisfies the condition in (1.1) and Δ_{ij} is as defined in Figure 1.1. For a macro iteration k (fixed), define the *micro transition matrix* P_m^k as

$$P_m^k = \begin{pmatrix} P_{GG}^k & P_{GL}^k & P_{GH}^k \\ P_{LG}^k & P_{LL}^k & P_{LH}^k \\ P_{HG}^k & P_{HL}^k & P_{HH}^k \end{pmatrix} \tag{3.4}$$

where the entries of P_m^k represent the micro transition probabilities, as defined in (3.3), between all solutions in the solution space. Without loss of generality, assume that the micro transition matrix P_m^k is irreducible, hence all states communicate for a finite macro iteration k. The irreducibility of P_m^k together with (1.1) and (3.3) guarantee that P_m^k is aperiodic, since for any two solutions $i, j \in \Omega,\ j \in \eta(i)$, such that $c_i < c_j$, the micro transition probability $P_{ij}^k > 0$ for some macro iteration k, which is a sufficient criterion for aperiodicity [23]. Note that for any two solutions $i, j \in \Omega,\ j \in \eta(i)$, such that $c_i \geq c_j$,

the micro transition probability $P_{ij}^k = g_{ij}^k > 0$ (i.e., improving solutions when generated are always accepted).

Without loss of generality for macro iteration k (fixed), assume that the probability the GHC algorithm remains at a global or local optima is positive:

$$\begin{aligned} P_{G_iG_i}^k > 0 \;\; for \;\; all \;\; i = 1, \ldots, \gamma, \;\; and \\ P_{L_jL_j}^k > 0 \;\; for \;\; all \;\; j = 1, \ldots, \lambda. \end{aligned} \tag{3.5}$$

By definition, the rows of the micro transition matrix P_m^k sum to one. Therefore,

$$\sum_{j=1}^{\gamma} P_{G_qG_j}^k + \sum_{j=1}^{\lambda} P_{G_qL_j}^k + \sum_{j=1}^{\phi} P_{G_qH_j}^k = 1, \tag{3.6}$$

$$\sum_{j=1}^{\gamma} P_{L_rG_j}^k + \sum_{j=1}^{\lambda} P_{L_rL_j}^k + \sum_{j=1}^{\phi} P_{L_rH_j}^k = 1, \tag{3.7}$$

and

$$\sum_{j=1}^{\gamma} P_{H_sG_j}^k + \sum_{j=1}^{\lambda} P_{H_sL_j}^k + \sum_{j=1}^{\phi} P_{H_sH_j}^k = 1, \tag{3.8}$$

for $q = 1, \ldots, \gamma$, $r = 1, \ldots, \lambda$, and $s = 1, \ldots, \phi$.

3.4 Existing GHC Convergence Theory

This section presents the existing convergence theory for GHC algorithms. This convergence theory, presented by Johnson [71] and Johnson and Jacobson [73], contains the most general convergence results in the literature for GHC algorithms. The following convergence theory is referenced when discussing the illustrative examples in Chapter 7.

To present this convergence theory, several definitions are needed. The notion of a *path* is defined to understand how a GHC algorithm searches the solution space Ω. A *path* from i to j, denoted as $i \Rightarrow j$, 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 outer

loop iterations k. The *path probability* between any two solutions $i, j \in L \cup G$ is defined as $P^k(i \Rightarrow j)$.

Recall that the GHC algorithm is composed of an outer loop, indexed on k, and an inner loop, indexed on n. Define Π_i^k as the stationary probability vector for all solutions $i \in \Omega$, for each k, as $N(k)$ approaches infinity. Define δ_i^k as the vector of stationary probabilities δ_i^k for all $i \in L \cup G$ such that

$$\delta_i^k \equiv \frac{\Pi_i^k}{\sum_{i \in L \cup G} \Pi_i^k}.$$

Theorem 3.1 provides sufficient conditions for the convergence of a GHC algorithm to the set $L \cup G$, as k approaches infinity.

Theorem 3.1 *Let (Ω, c) denote an instance of a discrete optimization problem. For a neighborhood function η, let the generation probabilities g_{ij}^k satisfy (1.1) and the conditions*

(a) *for all $i, j \in \Omega$ and all outer loop iterations k, there exists an integer $d \geq 1$ and a corresponding sequence of solution $l_0, l_1, l_2, \ldots, l_d \in \Omega$, with $l_0 = i, l_d = j$, and $g_{ij}^k > 0,\ m = 0, 1, \ldots, d - 1$.*

(b) $\lim_{k \to +\infty} g_{ij}^k > 0$ *for all $i, j, \in \Omega, j \in \eta(i)$.*

Moreover, let the acceptance probabilities satisfy

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

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

Then

$$\lim_{k \to +\infty} \Pi_i^k = 0 \ \ for \ \ all \ \ i \in H.$$

Proof:

See Johnson and Jacobson [73].

$\diamondsuit$

The following definitions are used to provide the sufficient convergence conditions in Theorem 3.2. The path of minimum positive probability between any local (but not global) optimum and any (local or global) optimum is defined to be

$$P^k(Min\_Path) \equiv \{P^k(j \Rightarrow i) | j \in L, i \in (L \cup G), \ \ and \ \ P^k(j \Rightarrow i) > 0 \ \ for \ \ all \ \ k\}.$$

The path of maximum positive probability between any global optimum and any local (but not global) optimum is defined to be

$$P^k(Max\_Path) \equiv \max\{P^k(i \Rightarrow j) | i \in G, j \in L\}.$$

The maximal product of locally (but not globally) optimal solution equilibrium probabilities and their associated path probabilities to other local (but not global) optima is defined as

$$P^k(Max\_Prod) \equiv \max\{\delta_j^k P^k(j \Rightarrow q) | j, q \in L, q \neq j, q \notin \eta(j)\}.$$

Theorem 3.2 provides sufficient conditions for the convergence of a GHC algorithm to the set G as k approaches infinity.

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

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

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

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

then

$$\lim_{k \to +\infty} \delta_j^k = 0 \ \ for \ \ all \ \ j \in L.$$

Proof:

See Johnson and Jacobson [73].

♢

Chapter 4

General Convergence Results

Chapter 4 provides necessary/sufficient convergence conditions for GHC algorithms. Section 4.1 expands on the concepts of the iteration classification scheme introduced in Section 3.2 (i.e., micro and macro iterations) for a GHC algorithm; this classification scheme provides the foundation upon which the convergence conditions are developed. Section 4.2 presents the necessary/sufficient convergence conditions. The implications of these convergence conditions are discussed in Section 4.3.

4.1 Properties of the Iteration Classification Scheme

The following lemma proves that certain micro transition probabilities have value zero. The values of these particular micro transition probabilities play an integral role in subsequent lemmas and proofs.

Lemma 4.1 *For macro iteration k (fixed), the micro transition probability from a global optimum to another global optimum, $P^k_{G_iG_j}$, $i, j = 1, \ldots, \gamma$, $i \neq j$, or to any local optimum, $P^k_{G_iL_r}$,$i = 1, \ldots, \gamma$, $r = 1, \ldots, \lambda$, is zero. Similarly, the micro transition probability from a local optimum to another local optimum, $P^k_{L_rL_s}$, $r, s = 1, \ldots, \lambda$, $r \neq s$, or to any global*

optimum, $P^k_{L_rG_i}$, $i = 1, \ldots, \gamma$, $r = 1, \ldots, \lambda$, *is zero.*

Proof:

The result follows from (3.2).

$\diamondsuit$

Lemma 4.2 presents a result for the matrix of micro transition probabilities between hill solutions, P^k_{HH}, that is used to prove Lemma 4.3.

Lemma 4.2 *For macro iteration k (fixed), consider the irreducible and aperiodic micro transition matrix* P^k_m *(3.4). The matrix* P^k_{HH}, *a submatrix of* P^k_m, *satisfies*

$$(I - P^k_{HH})^{-1} = \sum_{q=1}^{+\infty} (P^k_{HH})^{q-1} \tag{4.1}$$

Proof:

Since the micro transition matrix P^k_m is a stochastic matrix (i.e., P^k_m is a non-negative matrix and all of the rows of P^k_m sum to one) the spectral radius of P^k_m is one ([10]) (i.e., $\rho(P^k_m) = 1$).

Define the $(\gamma + \lambda + \phi) \times (\gamma + \lambda + \phi)$ matrix P^k_h as

$$P^k_h = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & P^k_{HH} \end{pmatrix}. \tag{4.2}$$

Since $P^k_m \geq 0$ is an irreducible matrix (hence $P^k_h \geq 0$), $P^k_m + P^k_h$ is an irreducible matrix (by Lemma 3.1).

Therefore,

$$P^k_m = \begin{pmatrix} P^k_{GG} & P^k_{GL} & P^k_{GH} \\ P^k_{LG} & P^k_{LL} & P^k_{LH} \\ P^k_{HG} & P^k_{HL} & P^k_{HH} \end{pmatrix} \geq \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & P^k_{HH} \end{pmatrix} = P^k_h.$$

Moreover, $P_h^k \neq P_m^k$, since if not (i.e., $P_h^k = P_m^k$), then

$$P_m^k = \begin{pmatrix} P_{GG}^k & P_{GL}^k & P_{GH}^k \\ P_{LG}^k & P_{LL}^k & P_{LH}^k \\ P_{HG}^k & P_{HL}^k & P_{HH}^k \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & P_{HH}^k \end{pmatrix},$$

and the rows of the micro transition matrix P_m^k no longer sum to one. This means that P_m^k cannot be a transition matrix of a Markov chain, which is a contradiction. Therefore, $P_h^k \neq P_m^k$.

In addition, since $0 \leq P_h^k \leq P_m^k$, $P_h^k \neq P_m^k$, and $P_h^k + P_m^k$ is an irreducible matrix, then by Lemma 3.3

$$\rho(P_h^k) < \rho(P_m^k) = 1. \tag{4.3}$$

The matrix P_h^k can be rewritten in the matrix form depicted in (3.1), where $A_{11} = 0$ and $A_{22} = P_{HH}^k$. Therefore, the matrix P_h^k is a reducible matrix, where the eigenvalues of P_h^k are the eigenvalues of the zero matrix (A_{11}) and P_{HH}^k. Therefore,

$$\rho(P_h^k) = \max\{\rho(0), \rho(P_{HH}^k)\}.$$

Since $0 \leq P_{HH}^k$, then by Lemma 3.2,

$$\rho(0) \leq \rho(P_{HH}^k),$$

hence

$$\rho(P_{HH}^k) = \rho(P_h^k) < 1.$$

The result follows from Lemma 3.4.

$\diamondsuit$

For macro iteration k (fixed), the *macro transition matrix*, P_M^k, can now be defined such that its entries represent the probability of a GHC algorithm moving from a global optimum or a local optimum to any global optimum or any local optimum. Lemma 4.3 provides a closed form expression for the macro transition matrix P_M^k.

Lemma 4.3 *For macro iteration k (fixed), the $(\gamma+\lambda)\times(\gamma+\lambda)$ macro transition matrix P_M^k is given by*

$$P_M^k = \begin{pmatrix} M^k_{G_1G_1} & \dots & M^k_{G_1G_\gamma} & M^k_{G_1L_1} & \dots & M^k_{G_1L_\lambda} \\ \vdots & & \vdots & \vdots & & \vdots \\ M^k_{G_\gamma G_1} & \dots & M^k_{G_\gamma G_\gamma} & M^k_{G_\gamma L_1} & \dots & M^k_{G_\gamma L_\lambda} \\ M^k_{L_1G_1} & \dots & M^k_{L_1G_\gamma} & M^k_{L_1L_1} & \dots & M^k_{L_1L_\lambda} \\ \vdots & & \vdots & \vdots & & \vdots \\ M^k_{L_\lambda G_1} & \dots & M^k_{L_\lambda G_\gamma} & M^k_{L_\lambda L_1} & \dots & M^k_{L_\lambda L_\lambda} \end{pmatrix} \tag{4.4}$$

where

$$M^k_{G_iL_r} = P^k_{G_iH}\mathcal{T}^k P^k_{HL_r},$$

$$M^k_{L_rG_i} = P^k_{L_rH}\mathcal{T}^k P^k_{HG_i},$$

$$M^k_{L_rL_r} = P^k_{L_rH}\mathcal{T}^k P^k_{HL_r} + P^k_{L_rL_r},$$

$$M^k_{L_rL_s} = P^k_{L_rH}\mathcal{T}^k P^k_{HL_s}, \;\; r \neq s,$$

$$M^k_{G_iG_i} = P^k_{G_iH}\mathcal{T}^k P^k_{HG_i} + P^k_{G_iG_i},$$

$$M^k_{G_iG_j} = P^k_{G_iH}\mathcal{T}^k P^k_{HG_j}, \;\; i \neq j,$$

for $i,j = 1,\dots,\gamma$ and $r,s = 1,\dots,\lambda$, where

$$\mathcal{T}^k = (I - P^k_{HH})^{-1}.$$

Proof:

For macro iteration k (fixed), the probability of transitioning from a global optimum G_i, $i = 1,\dots,\gamma$, to a local optimum L_r, $r = 1,\dots,\lambda$, through q hill solutions, $q = 1,2,\dots$, can be expressed as

$$P^k_{G_iH}(P^k_{HH})^{q-1}P^k_{HL_r}.$$

Therefore, the probability of transitioning from a global optimum G_i to a local optimum L_r via at least one hill solution (i.e., the probability of a macro iteration) is

$$\sum_{q=1}^{+\infty} P_{G_iH}^k (P_{HH}^k)^{q-1} P_{HL_r}^k.$$

To calculate the probability of a macro transition, $M_{G_iL_r}^k$, between a global optimum G_i and a local optimum L_r, the probability of moving from a global optimum G_i to a local optimum L_r without moving through a hill (i.e., a micro iteration), $P_{G_iL_r}^k$, must also be included.

Therefore, the probability of a macro transition between a global optimum G_i and a local optimum L_r is

$$M_{G_iL_r}^k = \sum_{q=1}^{+\infty} P_{G_iH}^k (P_{HH}^k)^{q-1} P_{HL_r}^k + P_{G_iL_r}^k.$$

The following macro transition probabilities similarly follow:

$$M_{L_rG_i}^k = \sum_{q=1}^{+\infty} P_{L_rH}^k (P_{HH}^k)^{q-1} P_{HG_i}^k + P_{L_rG_i}^k,$$

$$M_{G_iG_j}^k = \sum_{q=1}^{+\infty} P_{G_iH}^k (P_{HH}^k)^{q-1} P_{HG_j}^k + P_{G_iG_j}^k,$$

$$M_{L_rL_s}^k = \sum_{q=1}^{+\infty} P_{L_rH}^k (P_{HH}^k)^{q-1} P_{HL_s}^k + P_{L_rL_s}^k,$$

for $i, j = 1, \ldots, \gamma$ and $r, s = 1, \ldots, \lambda$.

Let

$$\mathcal{T}^k = (I - P_{HH}^k)^{-1}. \tag{4.5}$$

Lastly from Lemma 4.1 and Lemma 4.2, the macro transition probabilities can be rewritten as

$$M_{G_iL_r}^k = P_{G_iH}^k \mathcal{T}^k P_{HL_r}^k,$$

$$M_{L_rG_i}^k = P_{L_rH}^k \mathcal{T}^k P_{HG_i}^k,$$

$$M_{L_rL_r}^k = P_{L_rH}^k \mathcal{T}^k P_{HL_r}^k + P_{L_rL_r}^k,$$

$$M^k_{L_rL_s} = P^k_{L_rH}\mathcal{T}^k P^k_{HL_s}, \;\; r \neq s,$$

$$M^k_{G_iG_i} = P^k_{G_iH}\mathcal{T}^k P^k_{HG_i} + P^k_{G_iG_i},$$

$$M^k_{G_iG_j} = P^k_{G_iH}\mathcal{T}^k P^k_{HG_j}, \;\; i \neq j,$$

for $i, j = 1, \ldots, \gamma$ and $r, s = 1, \ldots, \lambda$.

$\diamondsuit$

Lemma 4.4 proves that the macro transition matrix P^k_M is a stochastic matrix.

Lemma 4.4 *For macro iteration k (fixed), consider the macro transition matrix P^k_M in (4.4). Then*

$$\sum_{j=1}^{\gamma} M^k_{G_iG_j} + \sum_{j=1}^{\lambda} M^k_{G_iL_j} = 1 \tag{4.6}$$

and

$$\sum_{j=1}^{\gamma} M^k_{L_rG_j} + \sum_{j=1}^{\lambda} M^k_{L_rL_j} = 1 \tag{4.7}$$

for $i = 1, \ldots, \gamma$ and $r = 1, \ldots, \lambda$. Therefore, all the rows of the macro transition matrix P^k_M sum to one.

Proof:

The sum of the rows in P^k_M (4.4) can be expressed as

$$\sum_{j=1}^{\gamma} M^k_{G_iG_j} + \sum_{j=1}^{\lambda} M^k_{G_iL_j} = \sum_{j=1}^{\gamma} P^k_{G_iH}\mathcal{T}^k P^k_{HG_j} + \sum_{j=1}^{\lambda} P^k_{G_iH}\mathcal{T}^k P^k_{HL_j} + P^k_{G_iG_i} \tag{4.8}$$

and

$$\sum_{j=1}^{\gamma} M^k_{L_rG_j} + \sum_{j=1}^{\lambda} M^k_{L_rL_j} = \sum_{j=1}^{\gamma} P^k_{L_rH}\mathcal{T}^k P^k_{HG_j} + \sum_{j=1}^{\lambda} P^k_{L_rH}\mathcal{T}^k P^k_{HL_j} + P^k_{L_rL_r} \tag{4.9}$$

for $i = 1, \ldots, \gamma$ and $r = 1, \ldots, \lambda$. First, (4.8) can be rewritten as

$$P^k_{G_iH}\mathcal{T}^k \sum_{j=1}^{\gamma} P^k_{HG_j} + P^k_{G_iH}\mathcal{T}^k \sum_{j=1}^{\lambda} P^k_{HL_j} + P^k_{G_iG_i}$$

$$= P^k_{G_iH}\mathcal{T}^k(\sum_{j=1}^{\gamma} P^k_{HG_j} + \sum_{j=1}^{\lambda} P^k_{HL_j}) + P^k_{G_iG_i}$$

$$= P^k_{G_iH}(I - P^k_{HH})^{-1} \begin{pmatrix} \sum_{j=1}^{\gamma} P^k_{H_1G_j} + \sum_{j=1}^{\lambda} P^k_{H_1L_j} \\ \sum_{j=1}^{\gamma} P^k_{H_2G_j} + \sum_{j=1}^{\lambda} P^k_{H_2L_j} \\ \vdots \\ \sum_{j=1}^{\gamma} P^k_{H_\phi G_j} + \sum_{j=1}^{\lambda} P^k_{H_\phi L_j} \end{pmatrix} + P^k_{G_iG_i}.$$

From (3.8), this becomes

$$P^k_{G_iH}(I - P^k_{HH})^{-1} \begin{pmatrix} 1 - \sum_{j=1}^{\phi} P^k_{H_1H_j} \\ 1 - \sum_{j=1}^{\phi} P^k_{H_2H_j} \\ \vdots \\ 1 - \sum_{j=1}^{\phi} P^k_{H_\phi H_j} \end{pmatrix} + P^k_{G_iG_i}$$

$$= P^k_{G_iH}(I - P^k_{HH})^{-1}(I - P^k_{HH}) \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} + P^k_{G_iG_i} = P^k_{G_iH} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} + P^k_{G_iG_i}$$

$$= \sum_{j=1}^{\phi} P^k_{G_iH_j} + P^k_{G_iG_i}. \tag{4.10}$$

Since $P^k_{G_iL_j} = 0$, $j = 1, \ldots, \lambda$, and $P^k_{G_iG_j} = 0$, $i \neq j$, $j = 1, \ldots, \gamma$ (Lemma 4.1), then from (3.6), (4.10) equals one. (4.9) follows in a similar manner.

$\diamondsuit$

Lemma 4.5 provides bounds on the number of macro iterations executed within a fixed number of micro iterations. This result is used in Lemma 4.6 to show that the macro transition matrix P^k_M is irreducible.

Lemma 4.5 *Consider a GHC algorithm execution consisting of a sequence of micro iterations with macro iterations imbedded within these micro iterations. For macro iteration k (fixed), let* $\alpha, \beta \in L \cup G$ *with* $P_{\alpha\beta}^{k\ (z)} > 0$*, for* $z > 0$ *micro iterations. If* D *is the number of macro iterations executed within the given* z *micro iterations, then*

$$\begin{cases} 1 \leq D \leq z, & \alpha = \beta \\ \\ 1 \leq D \leq z-1, & \alpha \neq \beta \end{cases} \tag{4.11}$$

Proof:

If $\alpha = \beta$, then since the GHC algorithm needs to hill climb in order to leave α, it is possible that the algorithm never leaves the solution α during all z micro iterations. Therefore, the maximum number of macro iterations executed during the z micro iterations is z (hence $D \leq z$).

It is also possible that the first iteration takes the algorithm to a hill solution and the algorithm continues to visit hill solutions until iteration z, when the algorithm returns to solution $\alpha = \beta$. Therefore, the minimum number of macro iterations executed during the z micro iterations is 1, hence $D \geq 1$. Therefore, $1 \leq D \leq z \ \ for \ \ \alpha = \beta$.

If $\alpha \neq \beta$, then for the GHC algorithm to move from α to β after z iterations, at least one hill solution needs to be visited, since α and β cannot be neighbors. Therefore, the maximum number of macro iterations executed during the z micro iterations is $z-1$, hence $D \leq z-1$.

Similar to the case where $\alpha = \beta$, the minimum number of macro iterations executed during the z micro iterations is 1, hence $D \geq 1$. Therefore, $1 \leq D \leq z-1 \ \ for \ \ \alpha \neq \beta$.

$\diamondsuit$

Lemma 4.6 proves that the macro transition matrix P_M^k is irreducible. This result is used in a subsequent proof.

Lemma 4.6 *For macro iteration k (fixed), the macro transition matrix P_M^k (4.4) is irreducible.*

Proof:

For macro iteration k (fixed), let $\alpha, \beta \in L \cup G$. Since the micro transition matrix P_m^k is irreducible and aperiodic, there exists a positive integer z such that

$$P_{\alpha\beta}^{k\ (z)} > 0$$

where z represents micro iterations. In other words,

$$P^k\{\alpha \to \beta \ in \ z \ micro \ iterations\} > 0.$$

Suppose $\alpha = \beta$. From Lemma 4.5, by conditioning on the number of macro iterations executed during the z micro iterations,

$$P^k\{\alpha \to \beta \ in \ z \ micro \ iterations\} =$$

$$\sum_{y=1}^{z} P^k\{\alpha \to \beta \ in \ z \ micro \ iterations | \alpha \to \beta \ in \ y \ macro \ iterations\} *$$

$$P^k\{\alpha \to \beta \ in \ y \ macro \ iterations\}$$

Therefore, there exists a positive integer $1 \leq y \leq z$ such that

$$P^k\{\alpha \to \beta \ in \ z \ micro \ iterations | \alpha \to \beta \ in \ y \ macro \ iterations\} *$$

$$P^k\{\alpha \to \beta \ in \ y \ macro \ iterations\} > 0.$$

This means that for all $\alpha, \beta \in L \cup G$, $\alpha = \beta$, there exists a positive integer $1 \leq y \leq z$ such that

$$P^k\{\alpha \to \beta \ in \ y \ macro \ iterations\} > 0,$$

Suppose $\alpha \neq \beta$. Similar to the above case, for all $\alpha, \beta \in L \cup G$ and $\alpha \neq \beta$ there exists a positive integer $1 \leq y \leq z - 1$ such that $P^k\{\alpha \to \beta \ in \ y \ macro \ iterations\} > 0$.

Therefore, the macro transition matrix P_M^k is irreducible.

$\diamondsuit$

Lemma 4.7 establishes the aperiodicity of the macro transition matrix. This result, in conjunction with Lemma 4.6, is used in subsequent proofs.

Lemma 4.7 *For macro iteration k (fixed), the macro transition matrix P_M^k (4.4) is aperiodic.*

Proof:

Recall from (3.5) that the micro transition probabilities $P^k_{G_iG_i} > 0$, $i = 1, \ldots, \gamma$, and $P^k_{L_jL_j} > 0$, $j = 1, \ldots, \lambda$. Then

$$M^k_{G_iG_i} = P^k_{G_iH}\mathcal{T}^k P^k_{HG_i} + P^k_{G_iG_i} > 0, \quad i = 1, \ldots, \gamma$$

and

$$M^k_{L_jL_j} = P^k_{L_jH}\mathcal{T}^k P^k_{HL_j} + P^k_{L_jL_j} > 0, \quad j = 1, \ldots, \lambda.$$

From Lemma 3.5, the macro transition matrix P_M^k is aperiodic.

$\diamondsuit$

Lemma 4.8 establishes a property of the macro transition probabilities that is used in subsequent proofs.

Lemma 4.8 *For macro iteration k (fixed) and a macro transition matrix P_M^k,*

$$\max_{i=1,\ldots,\gamma}\{\sum_{j=1}^{\lambda} M^k_{G_iL_j}\} > 0 \ \textit{and} \tag{4.12}$$

$$\max_{i=1,\ldots,\lambda}\{\sum_{j=1}^{\gamma} M^k_{L_iG_j}\} > 0. \tag{4.13}$$

Proof:

Suppose that

$$\max_{i=1,\ldots,\gamma}\{\sum_{j=1}^{\lambda} M^k_{G_iL_j}\} = 0.$$

Then $M^k_{G_iL_j} = 0$ for all $i = 1,\ldots,\gamma$ and $j = 1,\ldots,\lambda$, hence P^k_M can be written as

$$P^k_M = \begin{pmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{pmatrix}, \tag{4.14}$$

where

$$A_{11} = \begin{pmatrix} M^k_{G_1G_1} & \ldots & M^k_{G_1G_\gamma} \\ \vdots & & \vdots \\ M^k_{G_\gamma G_1} & \ldots & M^k_{G_\gamma G_\gamma} \end{pmatrix},$$

$$A_{21} = \begin{pmatrix} M^k_{L_1G_1} & \ldots & M^k_{L_1G_\gamma} \\ \vdots & & \vdots \\ M^k_{L_\lambda G_1} & \ldots & M^k_{L_\lambda G_\gamma} \end{pmatrix}, \text{ and } A_{22} = \begin{pmatrix} M^k_{L_1L_1} & \ldots & M^k_{L_1L_\lambda} \\ \vdots & & \vdots \\ M^k_{L_\lambda L_1} & \ldots & M^k_{L_\lambda L_\lambda} \end{pmatrix}.$$

Rearranging the rows and columns of (4.14) so that the resulting matrix has the form

$$\begin{pmatrix} A_{22} & A_{21} \\ 0 & A_{11} \end{pmatrix},$$

shows that the macro transition matrix P^k_M is reducible (see (3.1)). This contradicts Lemma 4.6.

Therefore,

$$\max_{i=1,\ldots,\gamma}\{\sum_{j=1}^{\lambda} M^k_{G_iL_j}\} > 0.$$

Similarly,

$$\max_{i=1,\ldots,\lambda}\{\sum_{j=1}^{\gamma} M^k_{L_iG_j}\} > 0.$$

$\diamondsuit$

4.2 Main Results

For macro iteration k (fixed), both the irreducibility and aperiodicity of the macro transition matrix P_M^k guarantee the existence of unique stationary probabilities for the macro transition matrix [71]. For the macro transition matrix, define the stationary probability of a global optimum as $\Pi_{G_i}^k$, $i = 1, \ldots, \gamma$, and the stationary probability of a local optimum as $\Pi_{L_j}^k$, $j = 1, \ldots, \lambda$. The behavior of the stationary probabilities of the global optima and local optima for macro iteration k (fixed) provide a means to measure the convergence of a GHC algorithm. Lemma 4.9 provides a lower bound and an upper bound on the sum of the stationary probabilities of the local optima.

Lemma 4.9 *Consider a GHC algorithm with macro transition matrix P_M^k (Lemma 4.3) for macro iteration k (fixed). Then a lower bound for the sum of the stationary probabilities of the local optima is*

$$\sum_{i=1}^{\lambda} \Pi_{L_i}^k \geq \frac{\min_{i=1,\ldots,\gamma}\{\sum_{j=1}^{\lambda} P_{G_iH}^k \mathcal{T}^k P_{HL_j}^k\}}{\min_{i=1,\ldots,\gamma}\{\sum_{j=1}^{\lambda} P_{G_iH}^k \mathcal{T}^k P_{HL_j}^k\} + \max_{i=1,\ldots,\lambda}\{\sum_{j=1}^{\gamma} P_{L_iH}^k \mathcal{T}^k P_{HG_j}^k\}}. \tag{4.15}$$

An upper bound for this sum is

$$\sum_{i=1}^{\lambda} \Pi_{L_i}^k \leq \frac{\max_{i=1,\ldots,\gamma}\{\sum_{j=1}^{\lambda} P_{G_iH}^k \mathcal{T}^k P_{HL_j}^k\}}{\max_{i=1,\ldots,\gamma}\{\sum_{j=1}^{\lambda} P_{G_iH}^k \mathcal{T}^k P_{HL_j}^k\} + \min_{i=1,\ldots,\lambda}\{\sum_{j=1}^{\gamma} P_{L_iH}^k \mathcal{T}^k P_{HG_j}^k\}}. \tag{4.16}$$

Proof:

For macro iteration k (fixed), the stationary probabilities of the global optima and the local optima satisfy the equations

$$\Pi^k = \Pi^k P_M^k, \tag{4.17}$$

$$\sum_{i=1}^{\gamma} \Pi_{G_i}^k + \sum_{j=1}^{\lambda} \Pi_{L_j}^k = 1, \tag{4.18}$$

where

$$\Pi^k = [\Pi_{G_1}^k, \ldots, \Pi_{G_\gamma}^k, \Pi_{L_1}^k, \ldots, \Pi_{L_\lambda}^k] \geq \mathbf{0}.$$

These equations can be written as

$$\begin{aligned}
\Pi^k_{G_i} &= \textstyle\sum_{j=1}^{\gamma} \Pi^k_{G_j} M^k_{G_j G_i} + \sum_{j=1}^{\lambda} \Pi^k_{L_j} M^k_{L_j G_i}, \quad i = 1, \ldots, \gamma \\
\Pi^k_{L_j} &= \textstyle\sum_{i=1}^{\gamma} \Pi^k_{G_i} M^k_{G_i L_j} + \sum_{i=1}^{\lambda} \Pi^k_{L_i} M^k_{L_i L_j}, \quad j = 1, \ldots, \lambda \quad . \\
\textstyle\sum_{i=1}^{\gamma} \Pi^k_{G_i} &= 1 - \textstyle\sum_{j=1}^{\lambda} \Pi^k_{L_j}
\end{aligned}$$

Since the rows of the macro transition matrix, P^k_M, sum to one (Lemma 4.4), then summing the stationary probability over all local optimum leads to

$$\sum_{j=1}^{\lambda} \Pi^k_{L_j} = \sum_{j=1}^{\lambda} \sum_{i=1}^{\gamma} \Pi^k_{G_i} M^k_{G_i L_j} + \sum_{j=1}^{\lambda} \sum_{i=1}^{\lambda} \Pi^k_{L_i} M^k_{L_i L_j}$$

$$= \sum_{i=1}^{\gamma} (\Pi^k_{G_i} \sum_{j=1}^{\lambda} M^k_{G_i L_j}) + \sum_{i=1}^{\lambda} (\Pi^k_{L_i} \sum_{j=1}^{\lambda} M^k_{L_i L_j})$$

$$= \sum_{i=1}^{\gamma} (\Pi^k_{G_i} \sum_{j=1}^{\lambda} M^k_{G_i L_j}) + \sum_{i=1}^{\lambda} [\Pi^k_{L_i} (1 - \sum_{j=1}^{\gamma} M^k_{L_i G_j})]$$

$$= \sum_{i=1}^{\gamma} (\Pi^k_{G_i} \sum_{j=1}^{\lambda} M^k_{G_i L_j}) + \sum_{i=1}^{\lambda} \Pi^k_{L_i} - \sum_{i=1}^{\lambda} (\Pi^k_{L_i} \sum_{j=1}^{\gamma} M^k_{L_i G_j}).$$

Therefore,

$$0 = \sum_{i=1}^{\gamma} (\Pi^k_{G_i} \sum_{j=1}^{\lambda} M^k_{G_i L_j}) - \sum_{i=1}^{\lambda} (\Pi^k_{L_i} \sum_{j=1}^{\gamma} M^k_{L_i G_j}). \tag{4.19}$$

If detailed balance (2.3) is required for the macro transition probabilities, then (4.19) follows. However, (4.19) does not imply detailed balance. Therefore, (4.19) is a relaxation of the detailed balance requirement of the macro transition probabilities (i.e., it is not necessary to require the macro transition probabilities satisfy detailed balance).

Equation (4.19) is used to obtain the lower bound and upper bound for the sum of stationary probabilities of the local optima, $\sum_{i=1}^{\lambda} \Pi_{L_i}^k$.

Lower Bound:

Since

$$\sum_{j=1}^{\lambda} M_{G_r L_j}^k \geq \min_{i=1,\ldots,\gamma} \{\sum_{j=1}^{\lambda} M_{G_i L_j}^k\}$$

for all $r = 1, \ldots, \gamma$, then using (4.19)

$$0 = \sum_{i=1}^{\gamma} (\Pi_{G_i}^k \sum_{j=1}^{\lambda} M_{G_i L_j}^k) - \sum_{i=1}^{\lambda} (\Pi_{L_i}^k \sum_{j=1}^{\gamma} M_{L_i G_j}^k) \geq [\min_{i=1,\ldots,\gamma} \{\sum_{j=1}^{\lambda} M_{G_i L_j}^k\} \sum_{i=1}^{\gamma} \Pi_{G_i}^k] - \sum_{i=1}^{\lambda} (\Pi_{L_i}^k \sum_{j=1}^{\gamma} M_{L_i G_j}^k).$$

This inequality can be rewritten as

$$0 \geq [\min_{i=1,\ldots,\gamma} \{\sum_{j=1}^{\lambda} M_{G_i L_j}^k\}(1 - \sum_{j=1}^{\lambda} \Pi_{L_j}^k)] - \sum_{i=1}^{\lambda} (\Pi_{L_i}^k \sum_{j=1}^{\gamma} M_{L_i G_j}^k).$$

Therefore,

$$\begin{aligned}
\min_{i=1,\ldots,\gamma} \{\sum_{j=1}^{\lambda} M_{G_i L_j}^k\} &\leq [\min_{i=1,\ldots,\gamma} \{\sum_{j=1}^{\lambda} M_{G_i L_j}^k\} \sum_{j=1}^{\lambda} \Pi_{L_j}^k] + \sum_{i=1}^{\lambda} (\Pi_{L_i}^k \sum_{j=1}^{\gamma} M_{L_i G_j}^k) \\
&\leq [\min_{i=1,\ldots,\gamma} \{\sum_{j=1}^{\lambda} M_{G_i L_j}^k\} \sum_{j=1}^{\lambda} \Pi_{L_j}^k] + [\max_{i=1,\ldots,\lambda} \{\sum_{j=1}^{\gamma} M_{L_i G_j}^k\} \sum_{i=1}^{\lambda} \Pi_{L_i}^k] \\
&= \left[\sum_{i=1}^{\lambda} \Pi_{L_i}^k\right] \left[\min_{i=1,\ldots,\gamma} \{\sum_{j=1}^{\lambda} M_{G_i L_j}^k\} + \max_{i=1,\ldots,\lambda} \{\sum_{j=1}^{\gamma} M_{L_i G_j}^k\}\right],
\end{aligned}$$

hence,

$$\sum_{i=1}^{\lambda} \Pi_{L_i}^k \geq \frac{\min_{i=1,\ldots,\gamma} \{\sum_{j=1}^{\lambda} M_{G_i L_j}^k\}}{\min_{i=1,\ldots,\gamma} \{\sum_{j=1}^{\lambda} M_{G_i L_j}^k\} + \max_{i=1,\ldots,\lambda} \{\sum_{j=1}^{\gamma} M_{L_i G_j}^k\}}.$$

By substitution,

$$\sum_{i=1}^{\lambda} \Pi_{L_i}^k \geq \frac{\min_{i=1,\ldots,\gamma} \{\sum_{j=1}^{\lambda} P_{G_i H}^k \mathcal{T}^k P_{H L_j}^k\}}{\min_{i=1,\ldots,\gamma} \{\sum_{j=1}^{\lambda} P_{G_i H}^k \mathcal{T}^k P_{H L_j}^k\} + \max_{i=1,\ldots,\lambda} \{\sum_{j=1}^{\gamma} P_{L_i H}^k \mathcal{T}^k P_{H G_j}^k\}}.$$

Upper Bound:

Since

$$\sum_{j=1}^{\lambda} M_{G_r L_j}^k \leq \max_{i=1,\ldots,\gamma} \{\sum_{j=1}^{\lambda} M_{G_i L_j}^k\}$$

for all $r = 1, \ldots, \gamma$, then using (4.19)

$$0 = \sum_{i=1}^{\gamma}(\Pi_{G_i}^k \sum_{j=1}^{\lambda} M_{G_iL_j}^k) - \sum_{i=1}^{\lambda}(\Pi_{L_i}^k \sum_{j=1}^{\gamma} M_{L_iG_j}^k) \leq [\max_{i=1,\ldots,\gamma}\{\sum_{j=1}^{\lambda} M_{G_iL_j}^k\} \sum_{i=1}^{\gamma} \Pi_{G_i}^k] - \sum_{i=1}^{\lambda}(\Pi_{L_i}^k \sum_{j=1}^{\gamma} M_{L_iG_j}^k).$$

This inequality can be rewritten as

$$0 \leq [\max_{i=1,\ldots,\gamma}\{\sum_{j=1}^{\lambda} M_{G_iL_j}^k\}(1 - \sum_{j=1}^{\lambda} \Pi_{L_j}^k)] - \sum_{i=1}^{\lambda}(\Pi_{L_i}^k \sum_{j=1}^{\gamma} M_{L_iG_j}^k).$$

Therefore,

$$\begin{aligned}
\max_{i=1,\ldots,\gamma}\{\sum_{j=1}^{\lambda} M_{G_iL_j}^k\} &\geq [\max_{i=1,\ldots,\gamma}\{\sum_{j=1}^{\lambda} M_{G_iL_j}^k\} \sum_{j=1}^{\lambda} \Pi_{L_j}^k] + \sum_{i=1}^{\lambda}(\Pi_{L_i}^k \sum_{j=1}^{\gamma} M_{L_iG_j}^k) \\
&\geq [\max_{i=1,\ldots,\gamma}\{\sum_{j=1}^{\lambda} M_{G_iL_j}^k\} \sum_{j=1}^{\lambda} \Pi_{L_j}^k] + [\min_{i=1,\ldots,\lambda}\{\sum_{j=1}^{\gamma} M_{L_iG_j}^k\} \sum_{i=1}^{\lambda} \Pi_{L_i}^k] \\
&= \left[\sum_{i=1}^{\lambda} \Pi_{L_i}^k\right] \left[\max_{i=1,\ldots,\gamma}\{\sum_{j=1}^{\lambda} M_{G_iL_j}^k\} + \min_{i=1,\ldots,\lambda}\{\sum_{j=1}^{\gamma} M_{L_iG_j}^k\}\right],
\end{aligned}$$

hence,

$$\sum_{i=1}^{\lambda} \Pi_{L_i}^k \leq \frac{\max_{i=1,\ldots,\gamma}\{\sum_{j=1}^{\lambda} M_{G_iL_j}^k\}}{\max_{i=1,\ldots,\gamma}\{\sum_{j=1}^{\lambda} M_{G_iL_j}^k\} + \min_{i=1,\ldots,\lambda}\{\sum_{j=1}^{\gamma} M_{L_iG_j}^k\}}.$$

By substitution,

$$\sum_{i=1}^{\lambda} \Pi_{L_i}^k \leq \frac{\max_{i=1,\ldots,\gamma}\{\sum_{j=1}^{\lambda} P_{G_iH}^k \mathcal{T}^k P_{HL_j}^k\}}{\max_{i=1,\ldots,\gamma}\{\sum_{j=1}^{\lambda} P_{G_iH}^k \mathcal{T}^k P_{HL_j}^k\} + \min_{i=1,\ldots,\lambda}\{\sum_{j=1}^{\gamma} P_{L_iH}^k \mathcal{T}^k P_{HG_j}^k\}}.$$

$\diamondsuit$

To guarantee the convergence of a GHC algorithm to a global optimum, criteria for convergence need to be established. The condition that the sum of the global stationary probabilities converges to one and each local stationary probability converges to zero as macro iteration $k \to +\infty$ is a sufficient condition for the convergence in probability of a GHC algorithms to a global optimum.

Theorem 4.1 provides necessary/sufficient conditions for the convergence in probability of a GHC algorithm.

Theorem 4.1

Under the conditions and assumptions of Lemma 4.9,

A) *If*

$$\lim_{k\to+\infty} \Pi_{L_i}^k = 0,\ i = 1, \ldots, \lambda,$$

then

$$\lim_{k\to+\infty} \left[\frac{\min_{i=1,\ldots,\gamma}\{\sum_{j=1}^{\lambda} P_{G_iH}^k \mathcal{T}^k P_{HL_j}^k\}}{\max_{i=1,\ldots,\lambda}\{\sum_{j=1}^{\gamma} P_{L_iH}^k \mathcal{T}^k P_{HG_j}^k\}} \right] = 0.$$

B) *If*

$$\lim_{k\to+\infty} \left[\frac{\max_{i=1,\ldots,\gamma}\{\sum_{j=1}^{\lambda} P_{G_iH}^k \mathcal{T}^k P_{HL_j}^k\}}{\max_{i=1,\ldots,\gamma}\{\sum_{j=1}^{\lambda} P_{G_iH}^k \mathcal{T}^k P_{HL_j}^k\} + \min_{i=1,\ldots,\lambda}\{\sum_{j=1}^{\gamma} P_{L_iH}^k \mathcal{T}^k P_{HG_j}^k\}} \right] = 0$$

then

$$\lim_{k\to+\infty} \Pi_{L_i}^k = 0,\ i = 1, \ldots, \lambda.$$

Proof:

A) Assume that $\lim_{k\to+\infty} \Pi_{L_i}^k = 0$ for $i = 1, \ldots, \lambda$.

Then, for all $0 < \epsilon_i < \frac{1}{\lambda}$, there exists a positive integer K_i such that $\Pi_{L_i}^k \leq \epsilon_i$ for all $k \geq K_i$ and $i = 1, \ldots, \lambda$. By summing over all i,

$$\sum_{i=1}^{\lambda} \Pi_{L_i}^k \leq \epsilon'$$

for all $k \geq K$, where $K = \max_{i=1,\ldots,\lambda}\{K_i\}$, $\epsilon' = \sum_{i=1}^{\lambda} \epsilon_i$, and $0 < \epsilon' < 1$.

Using the lower bound (4.15) for $\sum_{i=1}^{\lambda} \Pi_{L_i}^k$,

$$\frac{\min_{i=1,\ldots,\gamma}\{\sum_{j=1}^{\lambda} P_{G_iH}^k \mathcal{T}^k P_{HL_j}^k\}}{\min_{i=1,\ldots,\gamma}\{\sum_{j=1}^{\lambda} P_{G_iH}^k \mathcal{T}^k P_{HL_j}^k\} + \max_{i=1,\ldots,\lambda}\{\sum_{j=1}^{\gamma} P_{L_iH}^k \mathcal{T}^k P_{HG_j}^k\}} \leq \sum_{i=1}^{\lambda} \Pi_{L_i}^k \leq \epsilon',$$

for all $k \geq K$.

After inverting both sides of this inequality, for all $k \geq K$,

$$\frac{\min_{i=1,\ldots,\gamma}\{\sum_{j=1}^{\lambda} P_{G_iH}^k \mathcal{T}^k P_{HL_j}^k\} + \max_{i=1,\ldots,\lambda}\{\sum_{j=1}^{\gamma} P_{L_iH}^k \mathcal{T}^k P_{HG_j}^k\}}{\min_{i=1,\ldots,\gamma}\{\sum_{j=1}^{\lambda} P_{G_iH}^k \mathcal{T}^k P_{HL_j}^k\}} \geq \frac{1}{\epsilon'},$$

$$\frac{\max_{i=1,\ldots,\lambda}\{\sum_{j=1}^{\gamma} P_{L_iH}^k \mathcal{T}^k P_{HG_j}^k\}}{\min_{i=1,\ldots,\gamma}\{\sum_{j=1}^{\lambda} P_{G_iH}^k \mathcal{T}^k P_{HL_j}^k\}} + 1 \geq \frac{1}{\epsilon'},$$

$$\frac{\max_{i=1,\ldots,\lambda}\{\sum_{j=1}^{\gamma} P_{L_iH}^k \mathcal{T}^k P_{HG_j}^k\}}{\min_{i=1,\ldots,\gamma}\{\sum_{j=1}^{\lambda} P_{G_iH}^k \mathcal{T}^k P_{HL_j}^k\}} \geq \frac{1}{\epsilon'} - 1,$$

$$\frac{\max_{i=1,\ldots,\lambda}\{\sum_{j=1}^{\gamma} P_{L_iH}^k \mathcal{T}^k P_{HG_j}^k\}}{\min_{i=1,\ldots,\gamma}\{\sum_{j=1}^{\lambda} P_{G_iH}^k \mathcal{T}^k P_{HL_j}^k\}} \geq \frac{1-\epsilon'}{\epsilon'}.$$

Since $\max_{i=1,\ldots,\lambda}\{\sum_{j=1}^{\gamma} P_{L_iH}^k \mathcal{T}^k P_{HG_j}^k\} > 0$ (see (4.13)) it is valid to invert both sides of this inequality, hence

$$\frac{\min_{i=1,\ldots,\gamma}\{\sum_{j=1}^{\lambda} P_{G_iH}^k \mathcal{T}^k P_{HL_j}^k\}}{\max_{i=1,\ldots,\lambda}\{\sum_{j=1}^{\gamma} P_{L_iH}^k \mathcal{T}^k P_{HG_j}^k\}} \leq \frac{\epsilon'}{1-\epsilon'},$$

where $0 < \epsilon' < 1$.

If

$$\delta = \frac{\epsilon'}{1-\epsilon'},$$

then for all $\delta > 0$ there exists an $0 < \epsilon' < 1$, hence there exists a positive integer K such that

$$\frac{\min_{i=1,\ldots,\gamma}\{\sum_{j=1}^{\lambda} P_{G_iH}^k \mathcal{T}^k P_{HL_j}^k\}}{\max_{i=1,\ldots,\lambda}\{\sum_{j=1}^{\gamma} P_{L_iH}^k \mathcal{T}^k P_{HG_j}^k\}} \leq \delta$$

for all $k \geq K$.

Therefore,

$$\lim_{k\to+\infty}\left[\frac{\min_{i=1,\ldots,\gamma}\{\sum_{j=1}^{\lambda} P_{G_iH}^k \mathcal{T}^k P_{HL_j}^k\}}{\max_{i=1,\ldots,\lambda}\{\sum_{j=1}^{\gamma} P_{L_iH}^k \mathcal{T}^k P_{HG_j}^k\}}\right] = 0.$$

B) Assume that

$$\lim_{k\to+\infty}\left[\frac{\max_{i=1,\ldots,\gamma}\{\sum_{j=1}^{\lambda} P_{G_iH}^k \mathcal{T}^k P_{HL_j}^k\}}{\max_{i=1,\ldots,\gamma}\{\sum_{j=1}^{\lambda} P_{G_iH}^k \mathcal{T}^k P_{HL_j}^k\} + \min_{i=1,\ldots,\lambda}\{\sum_{j=1}^{\gamma} P_{L_iH}^k \mathcal{T}^k P_{HG_j}^k\}}\right] = 0.$$

Therefore, for all $\delta > 0$, there exists a positive integer K such that

$$\frac{\max_{i=1,\ldots,\gamma}\{\sum_{j=1}^{\lambda} P_{G_iH}^k \mathcal{T}^k P_{HL_j}^k\}}{\max_{i=1,\ldots,\gamma}\{\sum_{j=1}^{\lambda} P_{G_iH}^k \mathcal{T}^k P_{HL_j}^k\} + \min_{i=1,\ldots,\lambda}\{\sum_{j=1}^{\gamma} P_{L_iH}^k \mathcal{T}^k P_{HG_j}^k\}} \leq \delta$$

for all $k \geq K$.

Using the upper bound (4.16) for $\sum_{i=1}^{\lambda} \Pi_{L_i}^k$, $\sum_{i=1}^{\lambda} \Pi_{L_i}^k \leq \delta$. Therefore, $\Pi_{L_i}^k \leq \delta$ for all $i = 1, 2, \ldots, \lambda$.

This means that for all $\delta > 0$, there exists a positive integer K such that

$$\Pi_{L_i}^k \leq \delta$$

for all $k \geq K$ and $i = 1, \ldots, \lambda$, hence,

$$\lim_{k\to+\infty} \Pi_{L_i}^k = 0,\ i = 1, \ldots, \lambda.$$

$\diamondsuit$

Given a GHC algorithm applied to a solution space Ω with neighborhood function η, Corollary 4.1 provides a relaxed sufficient convergence condition that follows from Theorem 4.1if for all macro iterations k, each local optimum of the solution space is able to reach at least one global optimum by passing only through solutions of H.

Corollary 4.1

Under the conditions and assumptions of Lemma 4.9, suppose that a neighborhood function η is defined such that the macro transition probabilities from all local optima to at least one global optimum is positive. If

$$\lim_{k\to+\infty}\left[\frac{\max_{i=1,\ldots,\gamma}\{\sum_{j=1}^{\lambda} P_{G_iH}^k \mathcal{T}^k P_{HL_j}^k\}}{\min_{i=1,\ldots,\lambda}\{\sum_{j=1}^{\gamma} P_{L_iH}^k \mathcal{T}^k P_{HG_j}^k\}}\right] = 0,$$

then

$$\lim_{k\to+\infty} \Pi^k_{L_i} = 0,\ i = 1, \ldots, \lambda.$$

Proof:

From Lemma 4.8 and under the given assumptions, for all $i = 1, \ldots, \lambda$,

$$M^k_{L_iG_j} > 0, \quad for\ \ at\ \ least\ \ one\ \ j = 1, \ldots, \gamma.$$

Therefore,

$$\min_{i=1,\ldots,\lambda} \{\sum_{j=1}^{\gamma} M^k_{L_iG_j}\} > 0,$$

or equivalently,

$$\min_{i=1,\ldots,\lambda} \{\sum_{j=1}^{\gamma} P^k_{L_iH} \mathcal{T}^k P^k_{HG_j}\} > 0.$$

Therefore, for all $\delta > 0$, there exists a positive integer K such that

$$\frac{\max_{i=1,\ldots,\gamma} \{\sum_{j=1}^{\lambda} P^k_{G_iH} \mathcal{T}^k P^k_{HL_j}\}}{\min_{i=1,\ldots,\lambda} \{\sum_{j=1}^{\gamma} P^k_{L_iH} \mathcal{T}^k P^k_{HG_j}\}} \leq \frac{\delta}{1-\delta}$$

for all $k \geq K$, where $0 < \delta < 1$, hence

$$\lim_{k\to+\infty} \left[\frac{\max_{i=1,\ldots,\gamma} \{\sum_{j=1}^{\lambda} P^k_{G_iH} \mathcal{T}^k P^k_{HL_j}\}}{\min_{i=1,\ldots,\lambda} \{\sum_{j=1}^{\gamma} P^k_{L_iH} \mathcal{T}^k P^k_{HG_j}\}} \right] = 0.$$

$\diamondsuit$

From (4.18), $\sum_{i=1}^{\gamma} \Pi^k_{G_i} + \sum_{j=1}^{\lambda} \Pi^k_{L_j} = 1$ at macro iteration k (fixed). Since there exists a finite number of global and local optima, then $\lim_{k\to+\infty} \Pi^k_{L_j} = 0$ for all $j = 1, \ldots, \lambda$, if and only if $\lim_{k\to+\infty} \sum_{i=1}^{\gamma} \Pi^k_{G_i} = 1$. This result, together with Theorem 4.1, provides a necessary condition and a sufficient condition for the convergence of GHC algorithms, based on the stationary probabilities of the local and global optima.

4.3 Implications of Main Results

Given the macro transition matrix (4.4), when the sufficient convergence condition for the macro transition probabilities in Corollary 4.1 is satisfied and the assumptions of Corollary 4.1 hold, the set of global optima, G, must occur with probability one as macro iteration k approaches positive infinity (since the equilibrium probabilities for any solution in L approach zero). Intuitively, the sufficient convergence condition states that if the largest probability of escaping any global optimum to a local optimum in a macro iteration converges to zero faster than the smallest probability of escaping any local optimum to a global optimum in a macro iteration, then the GHC algorithm converges to the set of global optima. Similarly, the necessary convergence condition in Theorem 4.1 states if the GHC algorithm converges to the set of global optima, then the smallest probability of escaping any global optimum to a local optimum in a macro iteration converges to zero faster than the largest probability of escaping any local optimum to a global optimum in a macro iteration. These necessary/sufficient conditions provide an alternative approach to evaluating GHC algorithm convergence. In order to apply these conditions, a ratio of macro transition probabilities is calculated. This contrasts previous conditions that focus solely on summations of probabilities ([73]).

Chapter 5

Convergence Result Applications

Chapter 5 presents two illustrations of the convergence results in Chapter 4. Theorem 4.1 applies to a solution space of a discrete optimization problem with γ global optima, λ local optima, and ϕ hill solutions. The first case restricts the solution space to one global optimum and one local optimum (i.e., $\gamma = 1$, $\lambda = 1$). The second case restricts the solution space to one global optimum and λ local optima (i.e., $\gamma = 1$).

5.1 Case 1: $|G| = 1$ and $|L| = 1$

Consider a solution space with a neighborhood function for which there is one global optimum, $\gamma = 1$, and one local optimum, $\lambda = 1$. For this case, the lower bound and upper bound on the stationary probability of the local optimum are equal. A closed form expression for the stationary probabilities of the global and local optimum is presented in Corollary 5.1.

Corollary 5.1 *Consider a GHC algorithm with macro transition matrix P_M^k (Lemma 4.3) for macro iteration k (fixed). The stationary probabilities of the global optimum, Π_G^k, and*

local optimum, Π_L^k*, are*

$$\Pi_G^k = \frac{P_{LH}^k \mathcal{T}^k P_{HG}^k}{P_{GH}^k \mathcal{T}^k P_{HL}^k + P_{LH}^k \mathcal{T}^k P_{HG}^k} \quad \textit{and} \tag{5.1}$$

$$\Pi_L^k = \frac{P_{GH}^k \mathcal{T}^k P_{HL}^k}{P_{GH}^k \mathcal{T}^k P_{HL}^k + P_{LH}^k \mathcal{T}^k P_{HG}^k}. \tag{5.2}$$

Corollary 5.2 provides necessary/sufficient convergence conditions that follow from Theorem 4.1. Note that for this case, the necessary and the sufficient convergence conditions are equivalent.

Corollary 5.2 *Under the conditions and assumptions of Corollary 5.1,*

$$\lim_{k \to +\infty} \Pi_G^k = 1 \quad \textit{and} \quad \lim_{k \to +\infty} \Pi_L^k = 0$$

if and only if

$$\lim_{k \to +\infty} \left[\frac{P_{GH}^k \mathcal{T}^k P_{HL}^k}{P_{LH}^k \mathcal{T}^k P_{HG}^k} \right] = 0.$$

5.2 Case 2: $|G| = 1$ and $|L| = \lambda$

Consider a solution space with a neighborhood function for which there is one global optimum, $\gamma = 1$, and λ local optima. From Lemma 4.9, Corollary 5.3 provides lower and upper bounds on the sum of the stationary probabilities of the local optima.

Corollary 5.3 *Consider a GHC algorithm with macro transition matrix* P_M^k *(Lemma 4.3) for macro iteration* k *(fixed). A lower bound for the sum of the stationary probabilities of the local optima is*

$$\sum_{i=1}^{\lambda} \Pi_{L_i}^k \geq \frac{\sum_{j=1}^{\lambda} P_{GH}^k \mathcal{T}^k P_{HL_j}^k}{\sum_{j=1}^{\lambda} P_{GH}^k \mathcal{T}^k P_{HL_j}^k + \max_{i=1,\ldots,\lambda} \{P_{L_iH}^k \mathcal{T}^k P_{HG}^k\}}, \tag{5.3}$$

while an upper bound for this sum is

$$\sum_{i=1}^{\lambda} \Pi_{L_i}^k \leq \frac{\sum_{j=1}^{\lambda} P_{GH}^k \mathcal{T}^k P_{HL_j}^k}{\sum_{j=1}^{\lambda} P_{GH}^k \mathcal{T}^k P_{HL_j}^k + \min_{i=1,\ldots,\lambda}\{P_{L_iH}^k \mathcal{T}^k P_{HG}^k\}}. \tag{5.4}$$

Corollary 5.4 provides necessary/sufficient convergence conditions that follow from Theorem 4.1.

Corollary 5.4 *Under the conditions and assumptions of Corollary 5.3,*

A) *If*

$$\lim_{k\to+\infty} \Pi_{L_i}^k = 0,\ i = 1, \ldots, \lambda,$$

then

$$\lim_{k\to+\infty} \left[\frac{\sum_{j=1}^{\lambda} P_{GH}^k \mathcal{T}^k P_{HL_j}^k}{\max_{i=1,\ldots,\lambda}\{P_{L_iH}^k \mathcal{T}^k P_{HG}^k\}} \right] = 0.$$

B) *If*

$$\lim_{k\to+\infty} \left[\frac{\sum_{j=1}^{\lambda} P_{GH}^k \mathcal{T}^k P_{HL_j}^k}{\sum_{j=1}^{\lambda} P_{GH}^k \mathcal{T}^k P_{HL_j}^k + \min_{i=1,\ldots,\lambda}\{P_{L_iH}^k \mathcal{T}^k P_{HG}^k\}} \right] = 0$$

then

$$\lim_{k\to+\infty} \Pi_{L_i}^k = 0,\ i = 1, \ldots, \lambda.$$

Corollary 5.5 provides a sufficient convergence condition that follows from Corollary 4.1. Given a GHC algorithm applied to a solution space Ω with neighborhood function η, this sufficient convergence condition results if for all macro iterations k, each local optimum of the solution space is able to reach at least one global optimum by passing only through solutions of H.

Corollary 5.5 *Under the conditions and assumptions of Corollary 5.3 and Corollary 4.1,*

If

$$\lim_{k\to+\infty}\left[\frac{\sum_{j=1}^{\lambda} P_{GH}^{k}\mathcal{T}^{k}P_{HL_j}^{k}}{\min_{i=1,\ldots,\lambda}\{P_{L_iH}^{k}\mathcal{T}^{k}P_{HG}^{k}\}}\right]=0$$

then

$$\lim_{k\to+\infty}\Pi_{L_i}^{k}=0,\ i=1,\ldots,\lambda.$$

Following from Theorem 4.1 and Corollary 4.1, Corollary 5.4 and Corollary 5.5 provide necessary/sufficient convergence conditions for a GHC algorithm given that the solution space contains one global optimum and multiple local optima.

Chapter 6

Convergence Results for Particular Hill Climbing Variable Forms

Chapter 6 presents the convergence results in Chapter 4 for two particular GHC algorithm formulations. The significance of these results are also discussed. The first section restricts the solution space to eight elements and the hill climbing random variables to specific rational functions [102]. The second section restricts the hill climbing random variable to be exponential (hence SA).

6.1 Hill Climbing Variables as Rational Functions

Consider the eight element solution space in Figure 6.1, where $G = \{p_1, p_2\}$, $L = \{q_1, q_2\}$, and $H = \{r_1, r_2, r_3, r_4\}$. The neighborhood function is defined by the lines connecting the solutions. Let the micro transition probabilities be defined (for all macro iterations $k \geq 2$) from the micro transition matrix in Figure 6.2, where the rows are arranged in the order $p_1, p_2, q_1, q_2, r_1, r_2, r_3, r_4$.

Note that $g_{ij}^k = \frac{1}{2}$ for all $i \in \Omega$, $j \in \eta(i)$, and for all macro iterations $k \geq 2$. Also,

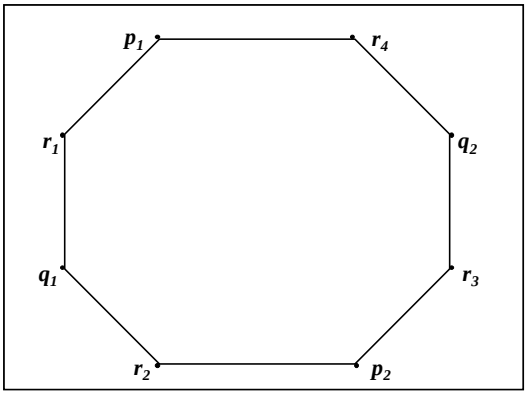

Figure 6.1: Solution Space for the Hill Climbing Variables as Rational Functions

$$P_m^k = \begin{pmatrix} 1-\frac{1}{k^{a1}} & 0 & 0 & 0 & \frac{1}{2k^{a1}} & 0 & 0 & \frac{1}{2k^{a1}} \\ 0 & 1-\frac{1}{k^{a1}} & 0 & 0 & 0 & \frac{1}{2k^{a1}} & \frac{1}{2k^{a1}} & 0 \\ 0 & 0 & 1-\frac{1}{k^{a2}} & 0 & \frac{1}{2k^{a2}} & \frac{1}{2k^{a2}} & 0 & 0 \\ 0 & 0 & 0 & 1-\frac{1}{k^{a2}} & 0 & 0 & \frac{1}{2k^{a2}} & \frac{1}{2k^{a2}} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \end{pmatrix}$$

Figure 6.2: The Micro Transition Matrix P_m^k (Hill Climbing Variables as Rational Functions)

$R_k(p_i, r_j) \equiv \frac{\Delta_{p_i r_j}}{(k^{a1}U)}$ for all $p_i, r_j \in \Omega$, $i = 1, 2$, and $j = 1, 2, 3, 4$, and $R_k(q_i, r_j) \equiv \frac{\Delta_{q_i r_j}}{(k^{a2}U)}$ for all $q_i, r_j \in \Omega$, $i = 1, 2$, and $j = 1, 2, 3, 4$, where U is distributed $U(0, 1)$. Then $Pr\{R_k(p_i, r_j) \geq \Delta_{pr_j}\} = \frac{1}{k^{a1}}$, for all $i = 1, 2$, and $j = 1, 2, 3, 4$, and all macro iterations $k \geq 2$. In addition, $Pr\{R_k(q_i, r_j) \geq \Delta_{q_i r_j}\} = \frac{1}{k^{a2}}$ for all $i = 1, 2$ and $j = 1, 2, 3, 4$, and all macro iterations $k \geq 2$. Therefore, for all macro iterations k finite, all solutions in Ω communicate, implying that the micro transition matrix, P_m^k, is irreducible. The irreducibility of P_m^k together with (1.1) and (3.3) guarantee that P_m^k is aperiodic, since for any two solutions $i, j \in \Omega$, $j \in \eta(i)$, such that $c_i < c_j$, the micro transition probability $P_{ij}^k > 0$, which is a

sufficient criterion for aperiodicity [23].

Corollary 6.1 presents necessary and sufficient convergence conditions for this GHC algorithm. These convergence conditions provide a more efficient means to establish convergence, as compared to the necessary/sufficient conditions in Theorem 4.1.

Corollary 6.1 *Given the solution space depicted in Figure 6.1 and the micro transition matrix defined in Figure 6.2,*

$$\lim_{k\to+\infty} \Pi_{q_1}^k = 0 \ \ and \ \ \lim_{k\to+\infty} \Pi_{q_2}^k = 0$$

$$if \ \ and \ \ only \ \ if \ \ (a1 - a2) > 0.$$

Proof:

A) (Necessary) From Theorem 4.1, the necessary condition for convergence of a GHC algorithm is

$$\lim_{k\to+\infty} \left[\frac{\min_{i=1,2}\{\sum_{j=1}^{2} P_{p_iH}^k \mathcal{T}^k P_{Hq_j}^k\}}{\max_{i=1,2}\{\sum_{j=1}^{2} P_{q_iH}^k \mathcal{T}^k P_{Hp_j}^k\}} \right] = 0. \tag{6.1}$$

Since $P_{HH}^k = \mathbf{0}$, the matrix $\mathcal{T}^k = \mathbf{I}$. Therefore, (6.1) reduces to

$$\lim_{k\to+\infty} \left[\frac{\min_{i=1,2}\{\sum_{j=1}^{2} P_{p_iH}^k P_{Hq_j}^k\}}{\max_{i=1,2}\{\sum_{j=1}^{2} P_{q_iH}^k P_{Hp_j}^k\}} \right] = 0.$$

For this example,

$$P_{GH}^k = \begin{pmatrix} \frac{1}{2k^{a1}} & 0 & 0 & \frac{1}{2k^{a1}} \\ 0 & \frac{1}{2k^{a1}} & \frac{1}{2k^{a1}} & 0 \end{pmatrix}, \ \ P_{LH}^k = \begin{pmatrix} \frac{1}{2k^{a2}} & \frac{1}{2k^{a2}} & 0 & 0 \\ 0 & 0 & \frac{1}{2k^{a2}} & \frac{1}{2k^{a2}} \end{pmatrix},$$

$$P_{HG}^k = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \\ 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix} \ \ and \ \ P_{HL}^k = \begin{pmatrix} \frac{1}{2} & 0 \\ \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{pmatrix}.$$

It follows that

$$\lim_{k\to+\infty}\left[\frac{\min_{i=1,2}\{\sum_{j=1}^{2}P_{p_iH}^{k}P_{Hq_j}^{k}\}}{\max_{i=1,2}\{\sum_{j=1}^{2}P_{q_iH}^{k}P_{Hp_j}^{k}\}}\right]$$

$$=\lim_{k\to+\infty}\left[\frac{\min\{\frac{1}{2k^{a1}},\frac{1}{2k^{a1}}\}}{\max\{\frac{1}{2k^{a2}},\frac{1}{2k^{a2}}\}}\right]=\lim_{k\to+\infty}\left[\frac{\frac{1}{2k^{a1}}}{\frac{1}{2k^{a2}}}\right]$$

$$=\lim_{k\to+\infty}\left[\frac{1}{k^{(a1-a2)}}\right].$$

Therefore,

$$\lim_{k\to+\infty}\left[\frac{\min_{i=1,2}\{\sum_{j=1}^{2}P_{p_iH}^{k}\mathcal{T}^{k}P_{Hq_j}^{k}\}}{\max_{i=1,2}\{\sum_{j=1}^{2}P_{q_iH}^{k}\mathcal{T}^{k}P_{Hp_j}^{k}\}}\right]=0 \;\; if \;\; and \;\; only \;\; if \;\; \lim_{k\to+\infty}\left[\frac{1}{k^{(a1-a2)}}\right]=0.$$

Lastly,

$$\lim_{k\to+\infty}\left[\frac{1}{k^{(a1-a2)}}\right]=0 \;\; if \;\; and \;\; only \;\; if \;\; (a1-a2)>0,$$

hence $(a1-a2)>0$ is a necessary convergence condition for this GHC algorithm.

B) (Sufficient) From Theorem 4.1, the sufficient condition for convergence of a GHC algorithm is

$$\lim_{k\to+\infty}\left[\frac{\max_{i=1,2}\{\sum_{j=1}^{2}P_{p_iH}^{k}\mathcal{T}^{k}P_{Hq_j}^{k}\}}{\max_{i=1,2}\{\sum_{j=1}^{2}P_{p_iH}^{k}\mathcal{T}^{k}P_{Hq_j}^{k}\}+\min_{i=1,2}\{\sum_{j=1}^{2}P_{q_iH}^{k}\mathcal{T}^{k}P_{Hp_j}^{k}\}}\right]=0. \qquad (6.2)$$

Since $P_{HH}^{k}=\mathbf{0}$, the matrix $\mathcal{T}^{k}=\mathbf{I}$. Therefore, (6.2) reduces to

$$\lim_{k\to+\infty}\left[\frac{\max_{i=1,2}\{\sum_{j=1}^{2}P_{p_iH}^{k}P_{Hq_j}^{k}\}}{\max_{i=1,2}\{\sum_{j=1}^{2}P_{p_iH}^{k}P_{Hq_j}^{k}\}+\min_{i=1,2}\{\sum_{j=1}^{2}P_{q_iH}^{k}P_{Hp_j}^{k}\}}\right]=0.$$

It follows that

$$\lim_{k\to+\infty}\left[\frac{\max_{i=1,2}\{\sum_{j=1}^{2}P_{p_iH}^{k}P_{Hq_j}^{k}\}}{\max_{i=1,2}\{\sum_{j=1}^{2}P_{p_iH}^{k}P_{Hq_j}^{k}\}+\min_{i=1,2}\{\sum_{j=1}^{2}P_{q_iH}^{k}P_{Hp_j}^{k}\}}\right]$$

$$=\lim_{k\to+\infty}\left[\frac{\max\{\frac{1}{2k^{a1}},\frac{1}{2k^{a1}}\}}{\max\{\frac{1}{2k^{a1}},\frac{1}{2k^{a1}}\}+\min\{\frac{1}{2k^{a2}},\frac{1}{2k^{a2}}\}}\right]=\lim_{k\to+\infty}\left[\frac{\frac{1}{2k^{a1}}}{\frac{1}{2k^{a1}}+\frac{1}{2k^{a2}}}\right]$$

$$= \lim_{k \to +\infty} \left[\frac{1}{1 + k^{(a1-a2)}} \right].$$

Therefore,

$$\lim_{k \to +\infty} \left[\frac{\max_{i=1,2} \{ \sum_{j=1}^{2} P_{p_i H}^k \mathcal{T}^k P_{H q_j}^k \}}{\max_{i=1,2} \{ \sum_{j=1}^{2} P_{p_i H}^k \mathcal{T}^k P_{H q_j}^k \} + \min_{i=1,2} \{ \sum_{j=1}^{2} P_{q_i H}^k \mathcal{T}^k P_{H p_j}^k \}} \right] = 0$$

$$if \ \ and \ \ only \ \ if \ \ \lim_{k \to +\infty} \left[\frac{1}{1 + k^{(a1-a2)}} \right] = 0.$$

Lastly,

$$\lim_{k \to +\infty} \left[\frac{1}{1 + k^{(a1-a2)}} \right] = 0 \ \ if \ \ and \ \ only \ \ if \ \ (a1 - a2) > 0,$$

hence $(a1 - a2) > 0$ is a sufficient convergence condition for this GHC algorithm.

Therefore, the necessary/sufficient conditions in Theorem 4.1 reduce to a single necessary and sufficient condition for the GHC algorithm for this particular solution space and neighborhood function.

$\diamondsuit$

To establish convergence using the necessary and sufficient conditions in Corollary 6.1, the exponents of the rational functions that define the hill climbing random variables need to satisfy a simple inequality (i.e., $(a1 - a2) > 0$), compared to the necessary/sufficient conditions in Theorem 4.1, which require calculating the ratio of the macro transition probabilities. Furthermore, the necessary and the sufficient conditions in Corollary 6.1 are equivalent, hence the condition determines both convergence or non-convergence. Moreover, Corollary 6.1 provides a condition that is not as computationally intensive as the necessary/sufficient conditions in Theorem 4.1.

6.2 Exponential Hill Climbing Variable

This section shows that by restricting a GHC algorithm to an exponential hill climbing random variable (i.e., SA), the convergence condition in Theorem 4.1 and the sufficient condition in Corollary 4.1 are equivalent to Hajek's [58] necessary and sufficient convergence conditions. Hajek proves that the SA algorithm converges in probability to the set of global optima, G, if and only if $\sum_{k=1}^{+\infty} e^{-\frac{d^*}{t_k}} = +\infty$, where t_k is a non-increasing cooling schedule (at iteration k) approaching zero as $k \to +\infty$ and d^* is the maximum depth of all elements in L. The *depth* of an element in $L \cup G$ is defined to be the smallest difference in objective function values between the element of $L \cup G$ and a solution of H, such that another element in $L \cup G$ can be reached from this solution in H using local search. Hajek's result assumes that the depth of any element of G is positive infinity. Hence, once the SA algorithm reaches an element in G, the probability that the algorithm will escape and never return to this global optimum is zero.

Define d_{G_i} to be the depth of G_i, $i = 1, \ldots, \gamma$. Therefore,

$$d_{G_i} = \min\{d_{G_iL},\ d_{G_iG}\},\quad i = 1, \ldots, \gamma \tag{6.3}$$

where d_{G_iL} is defined as the smallest difference in objective function values between G_i and a solution of H, such that another element in L can be reached from this solution in H by means of local search; d_{G_iG} is similarly defined.

Similarly,

$$d_{L_j} = \min\{d_{L_jL},\ d_{L_jG}\},\quad j = 1, \ldots, \lambda. \tag{6.4}$$

The following corollary provides the necessary/sufficient convergence conditions for SA, using the GHC convergence conditions in Theorem 4.1 and Corollary 4.1.

Corollary 6.2 *Consider a GHC algorithm with macro transition matrix P_M^k (Lemma 4.3) for macro iteration k (fixed), and $R_k(i,j) = -t_k \ln(u)$, $i, j \in \Omega$, $j \in \eta(i)$, where u is a $U(0,1)$ random variable. Assume that t_k is a non-increasing cooling schedule at macro iteration k*

that approaches zero as $k \to +\infty$. Under the conditions and assumptions of Lemma 4.9 and Corollary 4.1,

A) *If*

$$\lim_{k\to+\infty} \Pi^k_{L_i} = 0,\ i = 1, \ldots, \lambda,$$

then

$$\lim_{k\to+\infty} \left[\frac{M^k_{G_{min}L}}{M^k_{L_{max}G}}\right] = 0.$$

B) *If*

$$\lim_{k\to+\infty} \left[\frac{M^k_{G_{max}L}}{M^k_{L_{min}G}}\right] = 0$$

then

$$\lim_{k\to+\infty} \Pi^k_{L_i} = 0,\ i = 1, \ldots, \lambda$$

where

$$M^k_{G_{min}L} = \min_{i=1,\ldots,\gamma}\{\sum_{j=1}^{\lambda} M^k_{G_iL_j}\} = \min_{i=1,\ldots,\gamma}\{\sum_{j=1}^{\lambda} P^k_{G_iH}\mathcal{T}^k P^k_{HL_j}\},$$

$$M^k_{G_{max}L} = \max_{i=1,\ldots,\gamma}\{\sum_{j=1}^{\lambda} M^k_{G_iL_j}\} = \max_{i=1,\ldots,\gamma}\{\sum_{j=1}^{\lambda} P^k_{G_iH}\mathcal{T}^k P^k_{HL_j}\},$$

$$M^k_{L_{min}G} = \min_{i=1,\ldots,\lambda}\{\sum_{j=1}^{\gamma} M^k_{L_iG_j}\} = \min_{i=1,\ldots,\lambda}\{\sum_{j=1}^{\gamma} P^k_{L_iH}\mathcal{T}^k P^k_{HG_j}\}.$$

and

$$M^k_{L_{max}G} = \max_{i=1,\ldots,\lambda}\{\sum_{j=1}^{\gamma} M^k_{L_iG_j}\} = \max_{i=1,\ldots,\lambda}\{\sum_{j=1}^{\gamma} P^k_{L_iH}\mathcal{T}^k P^k_{HG_j}\}.$$

In addition, for $i = 1, \ldots, \gamma$, $j = 1, \ldots, \lambda$

$$P^k_{G_iH} = [g^k_{G_iH_1}e^{-\frac{\Delta_{G_iH_1}}{t_k}}, \ldots, g^k_{G_iH_\phi}e^{-\frac{\Delta_{G_iH_\phi}}{t_k}}],\ \ P^k_{L_jH} = [g^k_{L_jH_1}e^{-\frac{\Delta_{L_jH_1}}{t_k}}, \ldots, g^k_{L_jH_\phi}e^{-\frac{\Delta_{L_jH_\phi}}{t_k}}],$$

$$P^k_{HG_i} = [g^k_{H_1G_i}e^{-\frac{\Delta_{H_1G_i}}{t_k}}, \ldots, g^k_{H_\phi G_i}e^{-\frac{\Delta_{H_\phi G_i}}{t_k}}]^t,\ \ \textit{and}\ \ P^k_{HL_j} = [g^k_{H_1L_j}e^{-\frac{\Delta_{H_1L_j}}{t_k}}, \ldots, g^k_{H_\phi L_j}e^{-\frac{\Delta_{H_\phi L_j}}{t_k}}]^t.$$

Proof:

Follows directly by the definition of macro transition probabilities and Theorem 4.1.

♢

The following lemma establishes the order of the macro transition probabilities defined in Corollary 6.2.

Lemma 6.1 *Consider a GHC algorithm with macro transition matrix P_M^k (Lemma 4.3) for macro iteration k (fixed), and $R_k(i,j) = -t_k \ln(u)$, $i, j \in \Omega$, $j \in \eta(i)$, where u is a $U(0,1)$ random variable. Assume that t_k is a non-increasing cooling schedule at macro iteration k, that approaches zero as $k \to +\infty$. At macro iteration k,*

$$M_{G_iL}^k = O(e^{\frac{-d_{G_iL}}{t_k}}), \quad i = 1, \ldots, \gamma, \quad \textit{and}$$

$$M_{L_jG}^k = O(e^{\frac{-d_{L_jG}}{t_k}}), \quad j = 1, \ldots, \lambda.$$

Proof:

Since the hill climbing random variable at macro iteration k is exponential with mean t_k, then for all $i = 1, \ldots, \gamma$,

$$M_{G_iL}^k = O(P\{moving \ \ from \ \ G_i \ \ to \ \ an \ \ element \ \ of \ \ L\})$$

$$= O(P\{accepting \ \ hill \ \ climbing \ \ moves \ \ out \ \ of \ \ G_i \ \ to \ \ an \ \ element \ \ of \ \ L\})$$

$$= O(e^{\frac{-d_{G_iL}}{t_k}})$$

for t_k sufficiently close to zero. Similarly,

$$M_{L_jG}^k = O(e^{\frac{-d_{L_jG}}{t_k}}), \quad for \ \ all \ \ j = 1, \ldots, \lambda$$

for t_k sufficiently close to zero.

♢

Lemma 6.2 proves the relationship between Hajek's necessary and sufficient conditions and the necessary condition in Theorem 4.1 and the sufficient condition in Corollary 4.1, given that the GHC algorithm is restricted to an exponential hill climbing random variable. Hajek's necessary condition implies the necessary condition in Theorem 4.1 when the hill climbing random variable is exponential, hence, the necessary condition generalizes Hajek's necessary condition beyond an exponential hill climbing random variable. The sufficient condition in Corollary 4.1 is equivalent to Hajek's sufficient condition given the assumptions of the corollary and the restriction of an exponential hill climbing random variable. Moreover, the sufficient condition in Corollary 4.1 generalizes Hajek's sufficient condition beyond an exponential hill climbing random variable.

Lemma 6.2 *Consider a GHC algorithm with macro transition matrix P_M^k (Lemma 4.3) for macro iteration k (fixed), and $R_k(i,j) = -t_k \ln(u)$, $i, j \in \Omega$, $j \in \eta(i)$, where u is a $U(0,1)$ random variable. Assume that t_k is a non-increasing cooling schedule at macro iteration k that approaches zero as $k \to +\infty$ and the depth of any global optima is $+\infty$. Under the conditions and assumptions of Lemma 4.9 and Corollary 4.1,*

A)

$$If \quad \sum_{k=1}^{+\infty} e^{-\frac{d^*}{t_k}} = +\infty, \quad then \quad \lim_{k \to +\infty} \left[\frac{M^k_{G_{min}L}}{M^k_{L_{max}G}} \right] = 0.$$

B)

$$\lim_{k \to +\infty} \left[\frac{M^k_{G_{max}L}}{M^k_{L_{min}G}} \right] = 0 \quad if \ and \ if \ only \ if \quad \sum_{k=1}^{+\infty} e^{-\frac{d^*}{t_k}} = +\infty.$$

Proof:

A) It follows from $\sum_{k=1}^{+\infty} e^{-\frac{d^*}{t_k}} = +\infty$ that

$$M^k_{L_{max}G} = \max_{i=1,\ldots,\lambda} \{ \sum_{j=1}^{\gamma} M^k_{L_i G_j} \} > 0 \ \ for \ all \ macro \ iterations \ k \ finite.$$

To see this, by contradiction, suppose that there exists a macro iteration K such that $M^k_{L_{max}G} = 0$ for all $k \geq K$ (note that t_k non-increasing at macro iteration k and approaching zero as $k \to +\infty$ implies $\lim_{k\to+\infty} M^k_{L_{max}G} = 0$). It then follows that $e^{-\frac{d^*}{t_k}} = 0$ for all $k \geq K$, hence $\sum_{k=1}^{+\infty} e^{-\frac{d^*}{t_k}} < +\infty$, which is a contradiction.

Therefore, from Lemma 6.1, there exists a constant $c > 0$ and a macro iteration k_0 such that for all $k \geq k_0$,

$$\frac{M^k_{G_{min}L}}{M^k_{L_{max}G}} \leq \frac{ce^{\frac{-d_{G_{min}L}}{t_k}}}{M^k_{L_{max}G}}.$$

Since $d_{G_{min}L} \geq d_{G_{min}}$, then

$$e^{\frac{-d_{G_{min}L}}{t_k}} \leq e^{\frac{-d_{G_{min}}}{t_k}},$$

hence

$$\frac{M^k_{G_{min}L}}{M^k_{L_{max}G}} \leq \frac{ce^{\frac{-d_{G_{min}}}{t_k}}}{M^k_{L_{max}G}} \quad for \;\; all \;\; k \geq k_0.$$

However, since $d_{G_{min}} = +\infty$ under Hajek's conditions, then

$$\frac{ce^{\frac{-d_{G_{min}}}{t_k}}}{M^k_{L_{max}G}} = 0 \quad for \;\; all \;\; k \geq k_0.$$

Therefore,

$$\lim_{k\to+\infty} \left[\frac{M^k_{G_{min}L}}{M^k_{L_{max}G}}\right] = 0.$$

B)

Part 1: First show $\lim_{k\to+\infty} \left[\frac{M^k_{G_{max}L}}{M^k_{L_{min}G}}\right] = 0 \quad if \quad \sum_{k=1}^{+\infty} e^{-\frac{d^*}{t_k}} = +\infty.$

It follows from $\sum_{k=1}^{+\infty} e^{-\frac{d^*}{t_k}} = +\infty$ that

$$M^k_{L_{min}G} = \min_{i=1,\ldots,\lambda} \{\sum_{j=1}^{\gamma} M^k_{L_iG_j}\} > 0 \;\; for \;\; all \;\; macro \;\; iterations \;\; k \;\; finite.$$

To see this, by contradiction, suppose that there exists a macro iteration K such that $M^k_{L_{min}G} = 0$ for all $k \geq K$ (note that t_k non-increasing at macro iteration k and approaching zero as $k \to +\infty$ implies $\lim_{k\to+\infty} M^k_{L_{min}G} = 0$). It then follows that $e^{-\frac{d^*}{t_k}} = 0$ for all $k \geq K$, hence $\sum_{k=1}^{+\infty} e^{-\frac{d^*}{t_k}} < +\infty$, which is a contradiction.

Therefore, from Lemma 6.1, there exists a constant $c > 0$ and a macro iteration k_0 such that for all $k \geq k_0$,

$$\frac{M^k_{G_{max}L}}{M^k_{L_{min}G}} \leq \frac{ce^{\frac{-d_{G_{max}L}}{t_k}}}{M^k_{L_{min}G}}.$$

Since $d_{G_{max}L} \geq d_{G_{max}}$, then

$$e^{\frac{-d_{G_{max}L}}{t_k}} \leq e^{\frac{-d_{G_{max}}}{t_k}},$$

hence

$$\frac{M^k_{G_{max}L}}{M^k_{L_{min}G}} \leq \frac{ce^{\frac{-d_{G_{max}}}{t_k}}}{M^k_{L_{min}G}} \quad for \;\; all \;\; k \geq k_0.$$

However, since $d_{G_{max}} = +\infty$ under Hajek's conditions, then

$$\frac{ce^{\frac{-d_{G_{max}}}{t_k}}}{M^k_{L_{min}G}} = 0 \;\; for \;\; all \;\; k \geq k_0.$$

Therefore,

$$\lim_{k \to +\infty} \left[\frac{M^k_{G_{max}L}}{M^k_{L_{min}G}} \right] = 0.$$

Part 2: Second show $\lim_{k\to+\infty} \left[\frac{M^k_{G_{max}L}}{M^k_{L_{min}G}}\right] = 0 \quad only \;\; if \quad \sum_{k=1}^{+\infty} e^{-\frac{d^*}{t_k}} = +\infty$.

By the definition of the macro transition probabilities $M^k_{G_{max}L}$ and $M^k_{L_{min}G}$,

$$\lim_{k \to +\infty} \left[\frac{P\{Moving \;\; from \;\; G_{max} \;\; to \;\; a \;\; solution \;\; in \;\; L \;\; at \;\; macro \;\; iteration \;\; k\}}{P\{Moving \;\; from \;\; L_{min} \;\; to \;\; a \;\; solution \;\; in \;\; G \;\; at \;\; macro \;\; iteration \;\; k\}} \right] = 0.$$

It then follows that for all $k > 0$ finite,

$$P\{Moving \;\; from \;\; L_{min} \;\; to \;\; a \;\; solution \;\; in \;\; G \;\; at \;\; macro \;\; iteration \;\; k\} > 0. \tag{6.5}$$

Suppose that

$$\sum_{k=1}^{+\infty} e^{-\frac{d^*}{t_k}} < +\infty.$$

Since $d^* = \min_{i=1,\ldots,\lambda}\{d_{L_i}\} = \min_{i=1,\ldots,\lambda}\{\min\{d_{L_iG}, d_{L_iL}\}\}$, then

$$e^{-\frac{d_{L_iG}}{t_k}} \leq e^{-\frac{d_{L_i}}{t_k}} \leq e^{-\frac{d^*}{t_k}} \;\; for \;\; i = 1, \ldots, \lambda.$$

Therefore, for any $c > 0$,

$$\sum_{k=1}^{+\infty} ce^{-\frac{d_{L_iG}}{t_k}} < +\infty \ \ for \ \ i = 1, \ldots, \lambda.$$

By Lemma 6.1 there exists a $c > 0$ and an iteration k_0 such that for all $k \geq k_0$,

$$M^k_{L_{min}G} \leq ce^{\frac{-d_{L_{min}G}}{t_k}}.$$

It then follows that

$$\sum_{k=k_0}^{+\infty} M^k_{L_{min}G} < +\infty,$$

or

$$\sum_{k=k_0}^{+\infty} P\{Moving \ \ from \ \ L_{min} \ \ to \ \ a \ \ solution \ \ in \ \ G \ \ at \ \ macro \ \ iteration \ \ k\} < +\infty.$$

By the Borel-Cantelli Lemma [11], the event of moving from L_{min} to a solution in G at a macro iteration occurs finitely often with probability one. Therefore, there exists a $k_0^{'}$ such that

$$P\{Moving \ \ from \ \ L_{min} \ \ to \ \ a \ \ solution \ \ in \ \ G \ \ at \ \ macro \ \ iteration \ \ k\} = 0 \ \ for \ \ all \ \ k \geq k_0^{'},$$

which contradicts (6.5). Therefore,

$$\sum_{k=1}^{+\infty} e^{-\frac{d^*}{t_k}} = +\infty.$$

Hence, the sufficient condition in Theorem 4.1 is equivalent to Hajek's sufficient condition when the hill climbing random variable is exponential.

$\diamondsuit$

Chapter 7

Illustrative Examples

This chapter provides four examples that illustrate how a GHC algorithm may or may not satisfy the necessary or sufficient convergence conditions in Theorem 4.1. The first example satisfies the sufficient conditions in Johnson and Jacobson [73], but does not satisfy the sufficient condition in Theorem 4.1. The second example does not satisfy the sufficient conditions in Johnson and Jacobson [73], though it satisfies the sufficient condition in Theorem 4.1. The third example satisfies the sufficient conditions in Theorem 4.1 and in Johnson and Jacobson [73]. Finally, the fourth example does not satisfy the necessary condition in Theorem 4.1 (i.e., does not converge) and hence, should not satisfy the sufficient conditions in either Theorem 4.1 or Johnson and Jacobson [73].

7.1 Example 1

The purpose of Example 1 is to illustrate that it is possible for a GHC algorithm to not satisfy the sufficient condition in Theorem 4.1, while satisfying the sufficient condition in Johnson and Jacobson [73]. This example, taken from Johnson and Jacobson [73], defines the GHC acceptance criteria as a rational function of macro iteration k.

Consider the eight element solution space depicted in Figure 7.1, where $G = \{p\}$, $L = \{q_1, q_2, q_3\}$, and $H = \{r_1, r_2, r_3, r_4\}$. The neighborhood function is defined by the lines connecting the solutions. Let the micro transition probabilities be defined (for all macro iterations $k \geq 2$) from the micro transition matrix in Figure 7.2, where the rows are arranged in the order $p, q_1, q_2, q_3, r_1, r_2, r_3, r_4$.

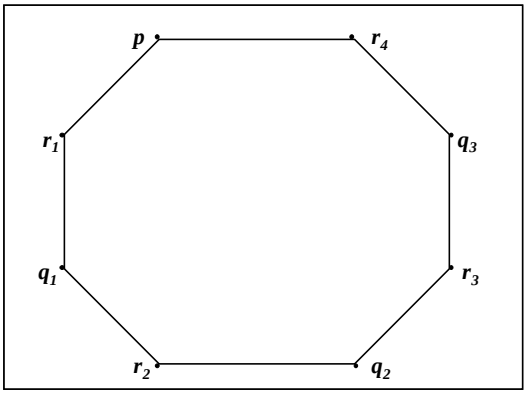

Figure 7.1: Solution Space 1

$$P_m^k = \begin{pmatrix} 1-\frac{1}{k^2} & 0 & 0 & 0 & \frac{1}{2k^2} & 0 & 0 & \frac{1}{2k^2} \\ 0 & 1-\frac{1}{k} & 0 & 0 & \frac{1}{2k} & \frac{1}{2k} & 0 & 0 \\ 0 & 0 & 1-\frac{1}{k} & 0 & 0 & \frac{1}{2k} & \frac{1}{2k} & 0 \\ 0 & 0 & 0 & 1-\frac{1}{k} & 0 & 0 & \frac{1}{2k} & \frac{1}{2k} \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \end{pmatrix}$$

Figure 7.2: The Micro Transition Matrix P_m^k (Example 1)

Note that $g_{ij}^k = \frac{1}{2}$ for all $i \in \Omega$, $j \in \eta(i)$, and for all macro iterations $k \geq 2$. Also, $R_k(p, r_j) \equiv \frac{\Delta_{pr_j}}{(k^2 U)}$ for all $p, r_j \in \Omega$, $j = 1, 2, 3, 4$, and $R_k(q_i, r_j) \equiv \frac{\Delta_{q_i r_j}}{(kU)}$ for all $q_i, r_j \in \Omega$, $i = 1, 2, 3$, and $j = 1, 2, 3, 4$, where U is distributed $U(0, 1)$. Then $Pr\{R_k(p, r_j) \geq \Delta_{pr_j}\} = \frac{1}{k^2}$,

and $Pr\{R_k(q_i, r_j) \geq \Delta_{q_i r_j}\} = \frac{1}{k}$ for all $i = 1, 2, 3$ and $j = 1, 2, 3, 4$, and all macro iterations $k \geq 2$. Therefore, for all macro iterations k finite, all solutions in Ω communicate, implying that the micro transition matrix, P_m^k, is irreducible. The irreducibility of P_m^k together with (1.1) and (3.3) guarantee that P_m^k is aperiodic, since for any two solutions $i, j \in \Omega$, $j \in \eta(i)$, such that $c_i < c_j$, the micro transition probability $P_{ij}^k > 0$, which is a sufficient criterion for aperiodicity [23].

To prove convergence using the sufficient condition in Theorem 4.1, it must be shown that

$$\lim_{k \to +\infty} \left[\frac{\sum_{j=1}^{3} P_{GH}^k \mathcal{T}^k P_{Hq_j}^k}{\sum_{j=1}^{3} P_{GH}^k \mathcal{T}^k P_{Hq_j}^k + \min_{j=1,2,3}\{P_{q_j H}^k \mathcal{T}^k P_{HG}^k\}} \right] = 0. \tag{7.1}$$

Since $P_{HH}^k = \mathbf{0}$, the matrix $\mathcal{T}^k = \mathbf{I}$. Therefore, (7.1) reduces to

$$\lim_{k \to +\infty} \left[\frac{\sum_{j=1}^{3} P_{GH}^k P_{Hq_j}^k}{\sum_{j=1}^{3} P_{GH}^k P_{Hq_j}^k + \min_{j=1,2,3}\{P_{q_j H}^k P_{HG}^k\}} \right] = 0.$$

For this example,

$$P_{GH}^k = (\tfrac{1}{2k^2}, 0, 0, \tfrac{1}{2k^2}), \quad P_{LH}^k = \begin{pmatrix} \frac{1}{2k} & \frac{1}{2k} & 0 & 0 \\ 0 & \frac{1}{2k} & \frac{1}{2k} & 0 \\ 0 & 0 & \frac{1}{2k} & \frac{1}{2k} \end{pmatrix},$$

$$P_{HG}^k = \begin{pmatrix} \frac{1}{2} \\ 0 \\ 0 \\ \frac{1}{2} \end{pmatrix} \quad \textit{and} \quad P_{HL}^k = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} \end{pmatrix}.$$

Then

$$\lim_{k \to +\infty} \left[\frac{\sum_{j=1}^{3} P_{GH}^k P_{Hq_j}^k}{\sum_{j=1}^{3} P_{GH}^k P_{Hq_j}^k + \min_{j=1,2,3}\{P_{q_j H}^k P_{HG}^k\}} \right] = \lim_{k \to +\infty} \left[\frac{\frac{1}{4k^2} + 0 + \frac{1}{4k^2}}{\frac{1}{4k^2} + 0 + \frac{1}{4k^2} + \min\{\frac{1}{4k}, 0, \frac{1}{4k}\}} \right]$$

$$= \lim_{k \to +\infty} \left[\frac{\frac{1}{2k^2}}{\frac{1}{2k^2} + 0} \right] = 1 \neq 0.$$

Therefore, the sufficient convergence condition in Theorem 4.1 is not satisfied. However, Johnson and Jacobson [73] show that their sufficient conditions are satisfied, hence the GHC algorithm converges in probability to the set of global optima.

7.2 Example 2

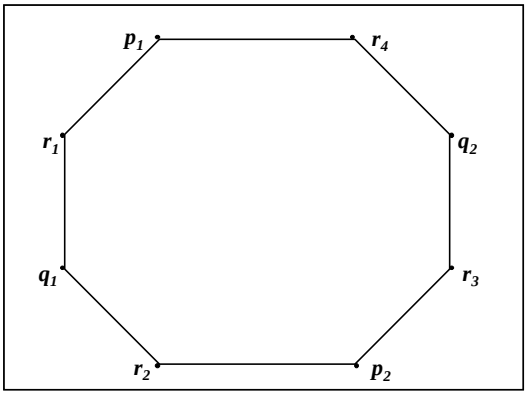

Figure 7.3: Solution Space 2

The purpose of Example 2 is to illustrate that it is possible for a GHC algorithm to satisfy the sufficient condition in Theorem 4.1, while not satisfying the sufficient condition in Johnson and Jacobson [73].

Consider the eight element solution space in Figure 7.3, where $G = \{p_1, p_2\}$, $L = \{q_1, q_2\}$, and $H = \{r_1, r_2, r_3, r_4\}$. The neighborhood function is defined by the lines connecting the solutions. Let the micro transition probabilities be defined (for all macro iterations $k \geq 2$) from the micro transition matrix in Figure 7.4, where the rows are arranged in the order $p_1, p_2, q_1, q_2, r_1, r_2, r_3, r_4$.

Note that $g_{ij}^k = \frac{1}{2}$ for all $i \in \Omega$, $j \in \eta(i)$, and for all macro iterations $k \geq 2$. Also, $R_k(p_i, r_j) \equiv \frac{\Delta_{p_i r_j}}{(kU)}$ for all $p_i, r_j \in \Omega$, $i = 1, 2$, and $j = 1, 2, 3, 4$, and $R_k(q_i, r_j) \equiv \frac{\Delta_{q_i r_j}}{(k^{\frac{1}{2}} U)}$ for all $q_i, r_j \in \Omega$, $i = 1, 2$, and $j = 1, 2, 3, 4$, where U is distributed $U(0, 1)$. Then $Pr\{R_k(p_i, r_j) \geq \Delta_{pr_j}\} = \frac{1}{k}$, for all $i = 1, 2$, and $j = 1, 2, 3, 4$, and all macro iterations

$$P_m^k = \begin{pmatrix} 1-\frac{1}{k} & 0 & 0 & 0 & \frac{1}{2k} & 0 & 0 & \frac{1}{2k} \\ 0 & 1-\frac{1}{k} & 0 & 0 & 0 & \frac{1}{2k} & \frac{1}{2k} & 0 \\ 0 & 0 & 1-\frac{1}{k^{\frac{1}{2}}} & 0 & \frac{1}{2k^{\frac{1}{2}}} & \frac{1}{2k^{\frac{1}{2}}} & 0 & 0 \\ 0 & 0 & 0 & 1-\frac{1}{k^{\frac{1}{2}}} & 0 & 0 & \frac{1}{2k^{\frac{1}{2}}} & \frac{1}{2k^{\frac{1}{2}}} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \end{pmatrix}$$

Figure 7.4: The Micro Transition Matrix P_m^k (Example 2)

$k \geq 2$. In addition, $Pr\{R_k(q_i, r_j) \geq \Delta_{q_i r_j}\} = \frac{1}{k^{\frac{1}{2}}}$ for all $i = 1, 2$ and $j = 1, 2, 3, 4$, and all macro iterations $k \geq 2$. Therefore, all solutions in Ω communicate, implying that the micro transition matrix, P_m^k, is irreducible. As with the solution space and neighborhood function for the first example, the irreducibility of P_m^k together with (1.1) and (3.3) guarantee that P_m^k is aperiodic [23].

Given the solution space and the micro transition matrix, the sufficient condition in Corollary 6.1 can be used to establish convergence. From Figure 7.4, $(a1 - a2) > 0$, since $a1 = 1$ and $a2 = \frac{1}{2}$.

Therefore, from Corollary 6.1

$$\lim_{k \to +\infty} \Pi_G^k = 1.$$

(Hence, solving for the stationary probabilities, see [23], $\Pi_{p_i}^k = \frac{k}{2k+2k^{\frac{1}{2}}+4}$, for $i = 1, 2$, and $\Pi_{q_j}^k = \frac{k^{\frac{1}{2}}}{2k+2k^{\frac{1}{2}}+4}$, for $j = 1, 2$. Note that

$$\lim_{k \to +\infty} \sum_{i=1}^{2} \Pi_{p_i}^k = \lim_{k \to +\infty} \frac{2k}{2k + 2k^{\frac{1}{2}} + 4} = 1,$$

which validates the convergence result.)

However, this example does not satisfy the sufficient conditions for convergence in Johnson

and Jacobson [73]. By the neighborhood function defined in Figure 7.3, four positive path probabilities from global to local optima exist and are equal. It then follows that

$$\sum_{k=2}^{+\infty} P^k(Max\_Path) = \sum_{k=2}^{+\infty} P^k(p_1 \Rightarrow q_1) = \sum_{k=2}^{+\infty} P^k(p_1 \Rightarrow q_3) = \sum_{k=2}^{+\infty} P^k(p_2 \Rightarrow q_1) = \sum_{k=2}^{+\infty} P^k(p_2 \Rightarrow q_3)$$

$$= \sum_{k=2}^{+\infty} P^k_{p_1 r_4} P^k_{r_4 q_3} = \sum_{k=2}^{+\infty} \frac{1}{2k} * \frac{1}{2}$$

$$= \sum_{k=2}^{+\infty} \frac{1}{4k} = +\infty,$$

violating the sufficient convergence condition in Johnson and Jacobson [73]. Therefore, Johnson and Jacobson [73] are unable to establish convergence for the GHC algorithm, though Theorem 4.1 shows that the GHC algorithm converges in probability to the set of global optima.

7.3 Example 3

The purpose of Example 3 is to illustrate that it is possible for a GHC algorithm to satisfy the sufficient condition in Theorem 4.1 and the sufficient conditions in Johnson and Jacobson [73].

Consider the eight element solution space in Figure 7.3, where $G = \{p_1, p_2\}$, $L = \{q_1, q_2\}$, and $H = \{r_1, r_2, r_3, r_4\}$. The neighborhood function is defined by the lines connecting the solutions. Let the micro transition probabilities be defined (for all macro iterations $k \geq 2$) from the micro transition matrix in Figure 7.5, where the rows are arranged in the order $p_1, p_2, q_1, q_2, r_1, r_2, r_3, r_4$.

Note that $g^k_{ij} = \frac{1}{2}$ for all $i \in \Omega$, $j \in \eta(i)$, and for all macro iterations $k \geq 2$. Also, $R_k(p_i, r_j) \equiv \frac{\Delta_{p_i r_j}}{(k^2 U)}$ for all $p_i, r_j \in \Omega$, $i = 1, 2$, and $j = 1, 2, 3, 4$, and $R_k(q_i, r_j) \equiv \frac{\Delta_{q_i r_j}}{(kU)}$ for all $q_i, r_j \in \Omega$, $i = 1, 2$, and $j = 1, 2, 3, 4$, where U is distributed $U(0, 1)$. Then $Pr\{R_k(p_i, r_j) \geq \Delta_{pr_j}\} = \frac{1}{k^2}$, for all $i = 1, 2$, and $j = 1, 2, 3, 4$, and all macro iterations $k \geq 2$. In addition, $Pr\{R_k(q_i, r_j) \geq \Delta_{q_i r_j}\} = \frac{1}{k}$ for all $i = 1, 2$ and $j = 1, 2, 3, 4$, and all

$$P_m^k = \begin{pmatrix} 1-\frac{1}{k^2} & 0 & 0 & 0 & \frac{1}{2k^2} & 0 & 0 & \frac{1}{2k^2} \\ 0 & 1-\frac{1}{k^2} & 0 & 0 & 0 & \frac{1}{2k^2} & \frac{1}{2k^2} & 0 \\ 0 & 0 & 1-\frac{1}{k} & 0 & \frac{1}{2k} & \frac{1}{2k} & 0 & 0 \\ 0 & 0 & 0 & 1-\frac{1}{k} & 0 & 0 & \frac{1}{2k} & \frac{1}{2k} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \end{pmatrix}$$

Figure 7.5: The Micro Transition Matrix P_m^k (Example 3)

macro iterations $k \geq 2$. Therefore, all solutions in Ω communicate, implying that the micro transition matrix, P_m^k, is irreducible. As with the solution space and neighborhood function for the first example, the irreducibility of P_m^k together with (1.1) and (3.3) guarantee that P_m^k is aperiodic [23].

Given the solution space and the micro transition matrix, the sufficient condition in Corollary 6.1 can be used to establish convergence. From Figure 7.5, $(a1 - a2) > 0$, since $a1 = 2$ and $a2 = 1$.

Therefore, from Corollary 6.1

$$\lim_{k \to +\infty} \Pi_G^k = 1.$$

The sufficient convergence conditions in Johnson and Jacobson [73] can also prove convergence. Since all of the solutions in Ω communicate, conditions (a) and (b) in Theorem 3.1 are satisfied. Furthermore, all hill climbing transition probabilities from each solution in $L \cup G$ to its neighbors in H are strictly positive, with limit zero as $k \to +\infty$, hence conditions (c) and (d) in Theorem 3.1 are satisfied, and so Theorem 3.1 applies. The sufficient conditions in Theorem 3.2 are now addressed.

Condition (e) examines the path of minimum positive probability from the set of local optima

L to solutions in $L \cup G$. Six positive path probabilities exist:

(i) $P^k(q_1 \Rightarrow p_1) = P^k_{q_1 r_1} P^k_{r_1 p_1} = (\frac{1}{2k})(\frac{1}{2}) = \frac{1}{4k}$

(ii) $P^k(q_1 \Rightarrow p_2) = \frac{1}{4k}$

(iii) $P^k(q_1 \Rightarrow q_1) = 1 - \frac{1}{2k}$

(iv) $P^k(q_2 \Rightarrow p_1) = \frac{1}{4k}$

(v) $P^k(q_2 \Rightarrow p_2) = \frac{1}{4k}$

(vi) $P^k(q_2 \Rightarrow p_2) = 1 - \frac{1}{2k}$

Note that $P^k(q_1 \Rightarrow q_2) = P^k(q_2 \Rightarrow q_1) = 0$, since either path must visit an intermediate solution in $L \cup G$. The minimal value of the six positive path probabilities is $\frac{1}{4k}$, and so $P^k(Min\_Path) = P^k(q_1 \Rightarrow p_1) = P^k(q_1 \Rightarrow p_2) = P^k(q_2 \Rightarrow p_1) = P^k(q_2 \Rightarrow p_2)$. Hence,

$$\sum_{k=1}^{+\infty} P^k(Min\_Path) = \sum_{k=1}^{+\infty} \frac{1}{4k} = +\infty,$$

and therefore condition (e) in Theorem 3.2 holds.

To address condition (f) in Theorem 3.2, the path probabilities from global to local optima must be examined. From the neighborhood function, only four positive path probabilities exists and they are equal. Hence,

$$\sum_{k=1}^{+\infty} P^k(Max\_Path) = \sum_{k=1}^{+\infty} P^k(p_1 \Rightarrow q_1) = \sum_{k=1}^{+\infty} P^k(p_1 \Rightarrow q_2) = \sum_{k=1}^{+\infty} P^k(p_2 \Rightarrow q_1) = \sum_{k=1}^{+\infty} P^k(p_2 \Rightarrow q_2)$$

$$= \sum_{k=1}^{+\infty} (\frac{1}{2k^2})(\frac{1}{2}) = \sum_{k=1}^{+\infty} (\frac{1}{4k^2}) < +\infty,$$

and so condition (f) holds.

Condition (g) in Theorem 3.2 requires that the vector of stationary probabilities, δ^k, be known for all solutions in L. Hence, solving for the stationary probabilities (see [23]),

$\Pi_{p_i}^k = \frac{k^2}{2k^2+2k+4}$, for $i = 1, 2$, and $\Pi_{q_j}^k = \frac{k}{2k^2+2k+4}$, for $j = 1, 2$. Therefore, $\delta_{p_i}^k = \frac{k^2}{2k^2+2k}$, for $i = 1, 2$, and $\delta_{q_j}^k = \frac{k}{2k^2+2k}$, for $j = 1, 2$. The maximal path probability is

$$P^k(Max\_Prod) = \delta_{q_1}^k P^k(q_1 \Rightarrow \ q_2) = \delta_{q_2}^k P^k(q_2 \Rightarrow q_1) = 0,$$

and hence for condition (g),

$$\sum_{k=1}^{+\infty} P^k(Max\_Prod) = 0 < +\infty.$$

Thus, the sufficient convergence conditions in Theorem 3.1 and Theorem 3.2 are satisfied. (Note that

$$\lim_{k\to+\infty} \sum_{i=1}^{2} \delta_{p_i}^k = \lim_{k\to+\infty} \frac{2k^2}{2k^2+2k} = \lim_{k\to+\infty} \sum_{i=1}^{2} \Pi_{p_i}^k = \lim_{k\to+\infty} \frac{2k^2}{2k^2+2k+4} = 1,$$

which validates the convergence result.)

This example illustrates that the sufficient conditions in Theorem 4.1 and in Johnson and Jacobson [73] prove that the GHC algorithm converges in probability to the set of global optima.

7.4 Example 4

The purpose of Example 4 is to illustrate that if a GHC algorithm does not satisfy the necessary condition in Theorem 4.1, then the algorithm does not satisfy the sufficient conditions in Theorem 4.1 and Johnson and in Jacobson [73].

Consider the eight element solution space in Figure 7.1, where $G = \{p\}$, $L = \{q_1, q_2, q_3\}$, and $H = \{r_1, r_2, r_3, r_4\}$. The neighborhood function is defined by the lines connecting the solutions. Let the micro transition probabilities be defined (for all macro iterations $k \geq 2$) from the micro transition matrix in Figure 7.6, where the rows are arranged in the order $p, q_1, q_2, q_3, r_1, r_2, r_3, r_4$.

Note that $g_{ij}^k = \frac{1}{2}$ for all $i \in \Omega$, $j \in \eta(i)$, and for all macro iterations $k \geq 2$. Also, $R_k(p, r_j) \equiv \frac{\Delta_{pr_j}}{(kU)}$ for all $p, r_j \in \Omega$, $j = 1, 2, 3, 4$, and $R_k(q_i, r_j) \equiv \frac{\Delta_{q_i r_j}}{(k^2 U)}$ for all $q_i, r_j \in \Omega$,

$$P_m^k = \begin{pmatrix} 1-\frac{1}{k} & 0 & 0 & 0 & \frac{1}{2k} & 0 & 0 & \frac{1}{2k} \\ 0 & 1-\frac{1}{k^2} & 0 & 0 & \frac{1}{2k^2} & \frac{1}{2k^2} & 0 & 0 \\ 0 & 0 & 1-\frac{1}{k^2} & 0 & 0 & \frac{1}{2k^2} & \frac{1}{2k^2} & 0 \\ 0 & 0 & 0 & 1-\frac{1}{k^2} & 0 & 0 & \frac{1}{2k^2} & \frac{1}{2k^2} \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \end{pmatrix}$$

Figure 7.6: The Micro Transition Matrix P_m^k (Example 4)

$i = 1, 2, 3$, and $j = 1, 2, 3, 4$, where U is distributed $U(0, 1)$. Then $Pr\{R_k(p, r_j) \geq \Delta_{pr_j}\} = \frac{1}{k}$, and $Pr\{R_k(q_i, r_j) \geq \Delta_{q_i r_j}\} = \frac{1}{k^2}$ for all $i = 1, 2, 3$ and $j = 1, 2, 3, 4$, and all macro iterations $k \geq 2$. Therefore, all solutions in Ω communicate, implying that the micro transition matrix, P_m^k, is irreducible. Once again, the irreducibility of P_m^k together with (1.1) and (3.3) guarantee that P_m^k is aperiodic [23].

To prove non-convergence using Theorem 4.1, it must be shown that

$$\lim_{k \to +\infty} \left[\frac{\sum_{j=1}^{3} P_{GH}^k \mathcal{T}^k P_{Hq_j}^k}{\max_{j=1,\ldots,3}\{P_{q_jH}^k \mathcal{T}^k P_{HG}^k\}} \right] \neq 0. \tag{7.2}$$

Since $P_{HH}^k = \mathbf{0}$, the matrix $\mathcal{T}^k = \mathbf{I}$. Therefore, (7.2) reduces to

$$\lim_{k \to +\infty} \left[\frac{\sum_{j=1}^{3} P_{GH}^k P_{Hq_j}^k}{\max_{j=1,\ldots,3}\{P_{q_jH}^k P_{HG}^k\}} \right] \neq 0.$$

It then follows that,

$$\lim_{k \to +\infty} \left[\frac{\sum_{j=1}^{3} P_{GH}^k P_{Hq_j}^k}{\max_{j=1,\ldots,3}\{P_{q_jH}^k P_{HG}^k\}} \right] = \lim_{k \to +\infty} \left[\frac{\frac{1}{4k} + 0 + \frac{1}{4k}}{\max\{\frac{1}{4k^2}, 0, \frac{1}{4k^2}\}} \right]$$

$$= \lim_{k \to +\infty} \left[\frac{\frac{1}{2k}}{\frac{1}{4k^2}} \right] = \lim_{k \to +\infty} 2k = +\infty \neq 0.$$

Therefore, since the necessary condition is not satisfied, the GHC algorithm does not converge to the global optimum. Moreover, this example should not satisfy any sufficient conditions for convergence. To see this, from Theorem 4.1, it must be shown that

$$\lim_{k\to+\infty}\left[\frac{\sum_{j=1}^{3} P_{GH}^{k}\mathcal{T}^{k}P_{Hq_j}^{k}}{\sum_{j=1}^{3} P_{GH}^{k}\mathcal{T}^{k}P_{Hq_j}^{k}+\min_{j=1,2,3}\{P_{q_jH}^{k}\mathcal{T}^{k}P_{HG}^{k}\}}\right]\neq 0,$$

which is true since

$$\lim_{k\to+\infty}\left[\frac{\sum_{j=1}^{3} P_{GH}^{k}\mathcal{T}^{k}P_{Hq_j}^{k}}{\sum_{j=1}^{3} P_{GH}^{k}\mathcal{T}^{k}P_{Hq_j}^{k}+\min_{j=1,2,3}\{P_{q_jH}^{k}\mathcal{T}^{k}P_{HG}^{k}\}}\right]=\lim_{k\to+\infty}\left[\frac{\frac{1}{4k}+0+\frac{1}{4k}}{\frac{1}{4k}+0+\frac{1}{4k}+\min\{\frac{1}{4k^2},0,\frac{1}{4k^2}\}}\right]$$

$$=\lim_{k\to+\infty}\left[\frac{\frac{1}{2k}}{\frac{1}{2k}+0}\right]=1\neq 0.$$

Therefore, the sufficient condition for convergence in Theorem 4.1 is not satisfied.

In addition, the GHC algorithm should not satisfy the sufficient convergence conditions in Johnson and Jacobson [73]. By the neighborhood function defined in Figure 7.1, only two positive path probabilities from global to local optima exist and these probabilities are equal. By applying the definitions of path probabilities in Johnson and Jacobson [73],

$$\sum_{k=2}^{+\infty}P^{k}(Max\_Path)=\sum_{k=2}^{+\infty}P^{k}(p\Rightarrow q_1)=\sum_{k=2}^{+\infty}P^{k}(p\Rightarrow q_3)$$

$$=\sum_{k=2}^{+\infty}P_{pr_4}^{k}P_{r_4q_3}^{k}=\sum_{k=2}^{+\infty}\frac{1}{2k}*\frac{1}{2}$$

$$=\sum_{k=2}^{+\infty}\frac{1}{4k}=+\infty,$$

which violates the sufficient conditions in Johnson and Jacobson [73].

This example illustrates how the necessary convergence condition in Theorem 4.1 can establish non-convergence of a GHC algorithm. Moreover, it demonstrates that the sufficient convergence conditions in Theorem 4.1 and in Johnson and Jacobson [73] are not satisfied when the necessary convergence condition is not satisfied.

Chapter 8

Conclusion

Many discrete optimization (minimization) problems belong to the class of NP-hard optimization problems [43], hence there does not exist a polynomial time algorithm that can solve such problems, unless P=NP. This has led researchers to develop local search heuristics to address NP-hard problems, with the hope of finding near-optimal solutions in a reasonable amount of computing time. One such heuristic approach is GHC algorithms [71, 72, 73], which include several other local search heuristics as particular formulations, including SA and TS. It is important to provide theoretical results that determine the effectiveness and performance of GHC algorithms when applied to specific problems, since many local search heuristics fall within the GHC algorithm framework.

8.1 Contributions

This dissertation presents necessary/sufficient convergence conditions for GHC algorithms. The convergence theory is based upon a new iteration classification scheme, where iterations of GHC algorithms are classified as macro or micro iterations. There are two main contributions from this research: first, the sufficient conditions for convergence, introduced in

Chapter 4, provide an alternative to the sufficient convergence conditions found in the literature. As illustrated in Chapter 7, it is possible for a GHC algorithm to satisfy the sufficient condition in Theorem 4.1, but not satisfy the sufficient conditions in Johnson and Jacobson [73], and visa-versa. Furthermore, it is possible that a GHC algorithm can either satisfy both sufficient conditions, or satisfy neither. Hence, the sufficient condition in Theorem 4.1 is not a generalization of the sufficient conditions in Johnson and Jacobson [73], but rather, another means to establish convergence of a GHC algorithm.

Second, a necessary condition for convergence is presented. The necessary convergence condition is an important contribution since it provides an alternative to the only other necessary condition in the literature [69]. Furthermore, if the GHC algorithm converges to the set of globally optimal solutions, then the necessary convergence condition for the macro transition probabilities provides a means to determine non-convergence of a GHC algorithm.

By restricting the solution space or hill climbing random variable for a GHC algorithm, the necessary/sufficient convergence conditions can be reduced. For example, the necessary/sufficient convergence conditions of Theorem 4.1 become equivalent when the solution space of a discrete optimization problem contains only one global optimum and one local optimum. This result provides a single condition to determine convergence or non-convergence of a GHC algorithm. In addition, when the GHC algorithm is restricted to SA, the necessary condition in Theorem 4.1 and the sufficient condition in Corollary 4.1, are related to the necessary and sufficient convergence conditions of Hajek [58]. Hajek's necessary condition implies the necessary condition in Theorem 4.1 when the GHC algorithm is SA. In addition, the new sufficient condition is equivalent to Hajek's sufficient condition when the hill climbing random variable is exponential and the assumptions of Corollary 4.1 hold. Moreover, the new necessary/sufficient conditions can also be viewed as a generalization of Hajek's necessary and sufficient conditions since the new convergence conditions do not require an exponential hill climbing random variable.

8.2 Future Research

Although Theorem 4.1 advances the convergence theory for GHC algorithms, there are numerous other issues that still need to be addressed. Research is in progress to determine the finite-time performance of GHC algorithms, making it possible to quantify the value of additional iterations, as well as determine optimal termination conditions when applying GHC algorithms to solve specific discrete optimization problems. Furthermore, finite-time performance results provide a measure to compare different GHC algorithm formulations with the objective of determining a priori which formulations are more likely to reach a global optimum with the least number of iterations executed. For each algorithm formulation, a finite-time performance theory would determine the iteration at which the hill climbing random variable has been sufficiently reduced such that the algorithm's output can be modeled as a homogeneous Markov chain. At this iteration of the algorithm, it may be possible to apply variance reduction techniques to improve the performance of the algorithm. When comparing various GHC algorithm formulations, the formulation that reaches the homogeneous state first will be the algorithm of choice, since such algorithms may perform superior to other formulations, in finite-time.

The performance of a GHC algorithm can also be measured by quantifing the rate at which the algorithm converges to a global optimum. Once theoretical results exist on the rate of convergence for a GHC algorithm, then it may be possible to chose a particular GHC algorithm formulation for a specific discrete optimization problem, such that the rate at which the algorithm converges to the set of global optima is maximized.

In addition, research is needed to determine necessary/sufficient convergence conditions with respect to the hill climbing random variable, $R_k(i, j)$, of the GHC algorithm. In Chapter 6, a particular hill climbing random variable (i.e., rational functions) was examined, and necessary and sufficient convergence conditions were formulated. Additional hill climbing random variables and the ensuing convergence theory (defined in terms of the hill climbing random variable) are rich areas of future investigation. The specific convergence theory for each hill

climbing random variable may provide insight into a unifying convergence theory for numerous GHC algorithm formulations. Hopefully, these conditions will provide a more practical means by which convergence or non-convergence can be established, compared to the convergence conditions in Theorem 4.1. Furthermore, these conditions should ideally provide insight into the convergence of TS, as well as other local search heuristics that fall within the GHC algorithm framework.

Bibliography

[1] Aarts, E.H.L., and Korst, J.H.M. (1989), *Simulated Annealing and Boltzmann Machines*, Wiley, Chichester.

[2] Aarts, E.H.L., and Laarhoven, P.J.M (1985), "Statistical cooling: A general approach to combinatorial optimization problems", *Phillips Journal of Research*, **40**, 193-226.

[3] Aarts, E.H.L., and Lenstra, J.K. (1997), *Local Search in Combinatorial Optimization*, Wiley and Sons, New York, New York.

[4] Althofer, I. and Koschnick, K.U. (1991), "On the convergence of threshold accepting", *Applied Mathematics and Optimization*, **24**, 183-195.

[5] Anily, S. and Federgruen, A. (1987), "Simulated annealing methods with general acceptance probabilities", *Journal of Applied Probability*, **24**, 657-667.

[6] Barrios, D., Perez Ruy-Diaz, J.A., Rios, J., and Segovia, J. (1994), "Conditions for convergence of genetic algorithms through Walsh series", *Computers and Artificial Intelligence*, **13**, 441-452.

[7] Battiti, R. and Tecchiolli, G. (1994), "The reactive tabu search", *ORSA Journal on Computing*, **6**, 126-140.

[8] Beaty, S. (1991), "Instruction scheduling using genetic algorithms", Doctoral Dissertation, Colorado State University.

[9] Belisle, C.J.P. (1992), "Convergence theorems for a class of simulated annealing algorithms on $\Re^d$", *Journal of Applied Probability*, **29**, 885-895.

[10] Berman, A. and Plemmons, R.J. (1994), *Non-negative Matrices in the Mathematical Sciences*, SIAM, Philadelphia, PA.

[11] Billingsley, P. (1995), *Probability and Measure*, 3rd Ed., Wiley, New York.

[12] Borkar, V.S. (1992), "Pathwise recurrence orders and simulated annealing", *Journal of Applied Probability*, **29**, 472-476.

[13] Bridges, C.L. and Goldberg, D.E. (1987), "An analysis of reproduction and crossover in a binary-coded genetic algorithm", *Proceedings of the 2nd International Conference on Genetic Algorithms*, 9-13.

[14] Cerf, R. (1994), "Asymptotic convergence of a genetic algorithm", *Statistics*, **319**, 271-276.

[15] Cerf, R. (1996), "An asymptotic theory for genetic algorithms", *1995 Artificial Evolution European Conference*, 37-53.

[16] Cerny, V. (1985), "Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm", *Journal of Optimisation Theory and Applications*, **45**, 41-51.

[17] Charon, I. and Hudry, O. (1993), "The noising method: A new method for combinatorial optimization", *Operations Research Letters*, **14**, 133-137.

[18] Charon, I., Hudry, O., and Lobstein, A. (1995), "Constructing covering nodes via noising", *Proceedings of the 1995 IEEE International Symposium on Information Theory*, Whistler, BC, Canada, 242.

[19] Cheh, K. M., Goldberg, J.B., and Askin, R.G. (1991), "A note on the effect of neighborhood structure in simulated annealing", *Computers and Operations Research*, **18**, 537-547.

[20] Chiang, T. and Chow, Y. (1989), "The asymptotic behavior of simulated annealing processes with absorption", *SIAM Journal of Control and Optimization*, **32**, 1247-1265.

[21] Chiang, T. and Chow, Y. (1994), "A limit theorem for a class of inhomogeneous Markov processes", *The Annals of Probability*, **17**, 1483-1502.

[22] Chipman, J.S. and Winker, P. (1995), "Optimal industrial classification by threshold accepting", *Control and Cybernetics*, **24**, 477-494.

[23] Cinlar, E. (1975), *Introduction to Stochastic Processes*, Prentice-Hall, Inc., Englewood Cliffs, New Jersy.

[24] Connors, D.P. and Kumar, P.R. (1989), "Simulated annealing type Markov chains and their order balance equations", *SIAM Journal of Control and Optimization*, **27**, 1440-1461.

[25] Connolly, D. T. (1990), "An improved annealing scheme for the QAP", *European Journal of Operational Research*, **46**, 93-100.

[26] Davis, L. (1991), *Handbook of Genetic Algorithms*, Van Nostrand Reinhold, New York.

[27] Davis, T.E. (1991), "Toward an extrapolation of the simulated annealing convergence theory onto the simple genetic algorithm", Doctoral dissertation, University of Florida.

[28] Davis, T.E. and Principe, J.C. (1991), "A simulated annealing like convergence theory for the simple genetic algorithm", *Proceedings of the Fourth Conference on Genetic Algorithms*, 174-181.

[29] de Werra, D. and Hertz, A. (1989), "Tabu search techniques: A tutorial and an application to nueral networks", *OR Spektrum*, **11**, 131-141.

[30] DeJong, K.A. and Spears, W.M. (1990), "An analysis of the interacting roles of population size and crossover in genetic algorithms", *Proceedings of the First Workshop of Parallel Problem Solving from Nature*, 38-47.

[31] Dowsland, K. A. (1991), "Hill-climbing, simulated annealing, and the Steiner problem in graphs", *Engineering Optimization*, **17**, 91-107.

[32] Dueck, G. (1993), "New optimization heuristics: The great deluge algorithm and the record-to-record travel", *Journal of Computational Physics*, **104**, 86-92.

[33] Dueck, G. and Scheuer, T. (1990), "Threshold accepting: A general purpose optimization algorithm appearing superior to simulated annealing", *Journal of Computational Physics*, **90**, 161-175.

[34] Duque-Anton, M. (1997), "Constructing efficient simulated annealing algorithms", *Discrete Applied Mathematics*, **77**, 139-159.

[35] Eiben, A.E., Aarts, E.H.L., and Van Hee, K.M. (1991), "Global convergence of genetic algorithms: A Markov chain analysis", *Parallel Problem Solving from Nature*, Schwefel, H.P. and Manner, R., Eds. Berlin and Heidelberg: Springer, 4-12.

[36] Faigle, U. and Kern, W. (1992), "Note on the convergence of simulated annealing algorithms", *SIAM Journal on Control and Optimization*, **29**, 153-159.

[37] Faigle, U. and Kern, W. (1992), "Some convergence results for probabilistic tabu search", *ORSA Journal on Computing*, **4**, 32-37.

[38] Faigle, U. and Schrader, R. (1988), "On the convergence of stationary distributions in simulated annealing algorithms", *Information Processing Letters*, **27**, 189-194.

[39] Fairley, A. (1991), "Comparison of methods of choosing the crossover point in the genetic crossover operation", Department of Computer Science, University of Liverpool.

[40] Fleischer, M. (1993), "Assessing the performance of the simulated annealing algorithm using information theory", Doctoral Dissertation, Department of Operations Research, Case Western Reserve University, Cleveland, OH.

[41] Fleischer, M.A. and Jacobson, S.H. (1999), "Information theory and the finite-time behavior of the simulated annealing algorithm: Experimental results", *INFORMS Journal on Computing* **11**, 35-43.

[42] Freisleben, B. and Schulte, M. (1992), "Combinatorial optimization with parallel adaptive threshold accepting", *Proceedings of the 1992 European Workshops on Parallel Computing*, 176-179.

[43] Garey, M. and Johnson, D.S. (1979), *Computers and Intractability: a guide to the theory of NP-completeness*, Freeman and Co., New York, NY.

[44] Geman, S. and Geman, D. (1984), "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-6**, 721-741.

[45] Gidas, B. (1985), "Nonstationary Markov chains and convergence of the annealing algorithm", *Journal of Statistical Physics*, **39**, 73-131.

[46] Glover, F. (1986), "Future paths for integer programming and links to artificial intelligence", *Computers and Operations Research*, **5**, 533-549.

[47] Glover, F. (1989), "Tabu search-Part I", *ORSA Journal on Computing*, **1**, 190-206.

[48] Glover, F. (1990), "Tabu search-Part II", *ORSA Journal on Computing*, **2**, 4-32.

[49] Glover, F. (1995), "Tabu search fundamentals and uses", Technical Report, University of Colorado, Boulder, Colorado.

[50] Glover, F. (1995), "Tabu thresholding: Improved search technique by nonmonotonic trajectories", *ORSA Journal on Computing*, **7**, 426-442.

[51] Glover, F. and Laguna, M. (1997), *Tabu Search*, Kluwer Academic Publishing, Norwell, Massachusetts.

[52] Goldberg, D.E. and Smith, R.E. (1987), "Nonstationary function optimization using genetic algorithms with dominance and diploidy", *Proceedings of the 2nd International Conference on Genetic Algorithms*, 59-68, Lawrence Erlbaum Associates, Hillsdale, New Jersy.

[53] Goldstein, L. and Waterman, M. (1989), "Neighborhood Size in the Simulated Annealing Algorithm", *American Journal of Mathematical and Management Sciences*, **8**, 409-423.

[54] Graham, A. (1987), *Non-negative Matrices and Applicable Topics in Linear Algebra*, Ellis Horwood Ltd., Chichester, England.

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

[56] Grefenstette, J.J. (1986), "Optimization of control parameters for genetic algorithms", *IEEE Transactions On Systems, Man, and Cybernetics*, **SMC-16**, 122-128.

[57] Grefenstette, J.J. (1987), "Incorporating problem-specific knowledge into genetic algorithms", in *Genetic Algorithms and Simulated Annealing*, Morgan Kauffmann, 42-60.

[58] Hajek, B. (1988), "Cooling schedules for optimal annealing", *Mathematics of Operations Research*, **13**, 311-329.

[59] Hansen, P. (1986), "The steepest ascent mildest descent heuristic for combinatorial programming", *Congress on Numerical Methods in Combinatorial Optimization*, Capri, Italy.

[60] Hansen, P., Jaumard, B., and Poggi di Aragao, M. (1992), "Mixed integer column generation algorithms and the probabilistic maximum satisfiability problem", *Proceedings of the 2nd Integer Programming and Combinatorial Optimization Conference*, Carnegie Mellon.

[61] Hertz, A. and de Werra, D. (1987), "Using tabu search techniques for graph coloring", *Computing*, **29**, 345-351.

[62] Hochbaum, D.S. and Pathria, A. (1997), "Generalized p-center problems: Complexity results and approximation algorithms", *European Journal of Operational Research*, **13**, 311-329.

[63] Holland, J.H. (1975), *Adaptation in Natural and Artificial Systems*, The University of Michigan Press, Ann Arbor, MI.

[64] Horn, J. (1993), "Finite Markov chain analysis of genetic algorithms with niching", *Proceedings of the Fifth International Conference on Genetic Algorithms*, 10-17.

[65] Hu, T.C., Kahing, A.B. and Tsao, C.W.A. (1995), "Old bachelor acceptance: A new class of non-monotone threshold accepting methods", *ORSA Journal on Computing*, **100**, 594-607.

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

[67] Jacobson, S.H., Sullivan, K.A., and Johnson, A.W. (1997), "Generalized hill climbing algorithms for discrete manufacturing process design problems using computer simulation models", *Proceedings of the 11th European Simulation Multiconference*, 473-478.

[68] Jacobson, S.H., Sullivan, K.A., and Johnson, A.W. (1998), "Discrete manufacturing process design optimization using computer simulation and generalized hill climbing algorithms", *Engineering Optimization*, **31** 247-260.

[69] Jacobson, S.H. and Yucesan, E. (1999), "A generalized hill climbing framework for studying the performance and convergence of simulated annealing and local search algorithms", Technical Report, Virginia Tech, Blacksburg, Virginia.

[70] Jaumard, B., Hansen, P., and Poggi di Aragao, M. (1992), "Column generation methods for probabilistic logic", *ORSA Journal on Computing*, **3**, 135-148.

[71] Johnson, A.W. (1996), "Generalized hill climbing algorithms for discrete optimization problems", Doctoral Dissertation, Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, VA.

[72] Johnson, A.W. and Jacobson, S.H. (1999), "A convergence result for generalized hill climbing algorithms", Technical Report, Virginia Tech, Blacksburg, VA.

[73] Johnson, A.W. and Jacobson, S.H. (1999), "On the convergence of generalized hill climbing algorithms", Technical Report, Virginia Tech, Blacksburg, VA.

[74] Johnson, D. S., Aragon, C. R., McGeoch, L. A., and Schevon, C. (1989), "Optimization by simulated annealing: An experimental evaluation; Part I, graph partitioning", *Operations Research*, **37**, 865-892.

[75] Johnson, D. S., Aragon, C. R., McGeoch, L. A., and Schevon, C. (1991), "Optimization by simulated annealing: An experimental evaluation; Part II, graph coloring and number partitioning", *Operations Research*, **39**, 378-406.

[76] Kirkpatrick, S., Gelatt, Jr., C.D., and Vecchi, M.P. (1983), "Optimization by simulated annealing", *Science*, **220**, 671-680.

[77] Kumar, A., Jacobson, S.H., and Sewell, E.C. (1999), "Computational analysis of a flexible assembly system design problem", To appear in *European Journal of Operational Research*.

[78] Lidia, S. and Carr, R. (1995), "Faster magnet sorting with a threshold acceptance algorithm", *Review of Scientific Instruments*, **66**, 1865-1867.

[79] Liepins, G.E. (1992), "On global convergence of genetic algorithms", *Proceedings of the SPIE*, **1766**, 61-65.

[80] Lin, C.K.Y, Haley, K.B., and Sparks, C. (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.

[81] Lundy, M. and Mees, A. (1986), "Convergence of an annealing algorithm", *Mathematical Programming*, **34**, 111-124.

[82] Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953), "Equation of state calculations by fast computing machines", *Journal of Chemical Physics*, **21**, 1087-1092.

[83] Mitra, D., Romeo, F., and Sangiovanni-Vincentelli, A.L. (1986), "Convergence and finite-time behavior of simulated annealing", *Advances in Applied Probability*, **18**, 747-771.

[84] 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.

[85] Moscato, P. and Fontanari, J.F. (1990), "Convergence and finite-time behavior of simulated annealing", *Advances in Applied Probability*, **18**, 747-771.

[86] Nissen, V. and Paul, H. (1995), "A modification of threshold accepting and its application to the quadratic assignment problem", *OR Spektrum*, **17**, 205-210.

[87] Nix, A. and Vose, M.D. (1992), "Modeling genetic algorithms with Markov chains", *Annals of Mathematics and Artificial Intelligence*, **5**, 27-34.

[88] Oliveira, S. and Stroud, G. (1989), "A parallel version of tabu search and the path assignment problem", *Heuristics for Combinatorial Optimization*, **4**, 1-24.

[89] Osman, I. H. (1993), "Metastrategy simulated annealing and tabu search algorithms for the vehicle routing problem", *Annals of Operations Research*, **41**.

[90] Osman, I. H. and Potts, C. N. (1989), "Simulated annealing for permutation flow-shop scheduling", *OMEGA*, **17**, 551-557.

[91] Prosser, P. (1988), "A hybrid genetic algorithm for pallet loading", *Proceedings of the 8th European Conference on Artificial Intelligence*, Pitman, London.

[92] Reeves, C.R. (1992), "A genetic algorithm approach to stochastic flowshop sequencing", *Proceedings from the IEE Colloquium on Genetic Algorithms for Control and Systems Engineering*, Digest No. 1992/106, IEE, London.

[93] Reeves, C.R. (1993), *Modern Heuristic Techniques for Combinatorial Problems*, John Wiley and Sons, New York.

[94] Richardson, J.T., Palmer, M.R., Liepins, G.E., and Hillyard, M.R. (1989), "Some guidelines for genetic algorithms with penalty functions", *Proceedings of the 3rd International Conference on Genetic Algorithms*, 191-197, Morgan Kaufmann, Los Altos, California.

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

[96] Rossier, Y., Troyon, M. and Liebling, Th.M. (1986), "Probabilistic exchange algorithms and euclidean traveling salesman problems", *OR Spektrum*, **8**, 151-164.

[97] Rudolph, G. (1994), "Convergence analysis of canonical genetic algorithms", *IEEE Transactions on Neural Networks, Special Issue on Evolutional Computing*, **5**, 96-101.

[98] Schaffer, J.D. (1987), "Some effects of selection procedures on hyperplane sampling by genetic algorithms", in *Genetic Algorithms and Simulated Annealing*, Morgan Kauffmann, Los Altos, CA, 89-103.

[99] Scheermesser, T. and Bryngdahl, O. (1994), "Threshold accepting for constrained halftoning", *Optics Communications*, **115**, 13-18.

[100] Schuur, P.C. (1997), "Classification of acceptance criteria for the simulated annealing algorithm", *Mathematics of Operations Research*, **22**, 266-275.

[101] Seneta, E. (1981), *Non-negative Matrices and Markov Chains (2nd Edition)*, Springer-Verlag, New York, NY.

[102] Simmons, G.F. (1985), *Calculus with Analytic Geometry*, McGraw-Hill, New York.

[103] Srinivas, M. and Patnaik, L.M. (1994), "Genetic algorithms: A survey", *Computer*, **27**, 17-26.

[104] Sudhakar, V. and Murthy, C.S.R. (1997), "A modified noising algorithm for the graph partitioning problem", *Integration, The VLSI Journal*, **22**, 101-113.

[105] Suzuki, J. (1995), "A Markov chain analysis on simple genetic algorithms", *IEEE Transactions on Systems, Man, and Cybernetics*, **25**, 655-659.

[106] Taillard, E. (1990), "Some efficient heuristic methods for the flow shop sequencing problem", *European Journal of Operational Research*, **47**, 65-74.

[107] Taillard, E. (1991), "Robust taboo search for the quadratic assignment probem", *Parallel Computing*, **17**, 443-455.

[108] Thierens, D. and Goldberg, D. (1994), "Convergence models of genetic algorithm selection schemes", *Parallel Problem Solving from Nature - PPSN III*, 9-14.

[109] Tian, P., Ma, J., and Fan, Z. (1997) "A stochastic tabu search strategy and its global convergence", *Proceedings from the IEEE International Conference on Systems, Man, and Cybernetics*, **4535**, 410-414.

[110] Trouve, A. (1996), "Cycle decompositions and simulated annealing", *SIAM Journal of Control and Optimization*, **34**, 966-986.

[111] Vose, M.D. (1992), "Modeling of genetic algorithms", *Foundations of Genetic Algorithms 2*, Darrell Whitley, L., ed., 63-73.

[112] Vose, M.D. (1994), "Logarithmic convergence of random heuristic search", *Proceedings of the SPIE*, **2304**, 46-52.

[113] Winker, P. and Fang, K. (1997), "Application of threshold-accepting to the evaluation of the discrepancy of a set of points", *SIAM Journal on Numerical Analysis*, **34**, 2028-2042.

[114] Yao, X. (1991), "Simulated annealing with extended neighborhood", *International Journal of Computer Mathematics*, **40**, 169-189.

Vita

Kelly A. Sullivan

Kelly A. Sullivan was born March 8, 1972 in Fort Belvoir, Virginia to John J. Sullivan and Ann W. Taylor. In May 1993, she received her Bachelor of Science in Mathematics from Virginia Polytechnic Institute and State University in Blacksburg, Virginia. In addition, she received a Master of Science in Mathematics from Virginia Polytechnic Institute and State University in May 1995. While completing her degrees, Kelly was inducted into Alpha Phi Mu and Phi Mu Epsilon, served as an officer for the Institute for Operations Research and Management Science, and was a member of Alpha Phi Omega. During the summer of 1996, she participated in the AFOSR graduate student summer research fellowship program at the Materials Directorate, Wright Laboratories, Wright-Patterson AFB, Dayton, Ohio. In December 1996, Kelly was awarded a Women and Minority Scholarship by the INFORMS College on Simulation to attend the Winter Simulation Conference. Furthermore, she co-authored a paper that received the Best Paper Award in the Industrial Simulation Track at the 1997 European Simulation Multiconference. Kelly completed her studies under the guidance of Sheldon H. Jacobson and received her Ph.D. in Industrial and Systems Engineering with a concentration in Operations Research in May 1999 from Virginia Polytechnic Institute and State University.