Logarithmic Convergence of Random Heuristic Search

Michael D. Vose
Computer Science Department, 107 Ayres Hall
The University of Tennessee
Knoxville, TN 37996-1301
e-mail: vose@cs.utk.edu

Abstract
This paper speaks to the inherent emergent behavior of genetic search. For completeness and generality, a class of stochastic search algorithms, *random heuristic search*, is reviewed. A general convergence theorem for this class is then proved. Since the simple genetic algorithm (GA) is an instance of random heuristic search, a corollary is a result concerning GAs and time to convergence.
Keywords
Convergence, genetic algorithm, time to convergence, random heuristic approach.

1. Introduction

Typically, a paper dealing with time to convergence behavior might begin with a discussion of convergence, or at least define what is meant by "time to convergence" early on. This paper is not organized along those lines for two reasons. First, when speaking of convergence it is relevant to know *what* it is that is in some sense "converging." Second, when sorting out what "convergence" is supposed to mean, difficulties arise with naive notions of convergence for the reason that convergence might not take place or might correspond to an infinite amount of time.

Therefore, this paper begins with a review of the type of algorithm being considered so as to both approach the question of "what is converging" and to also provide the appropriate context in which to arrive at a reasonable definition of time to convergence. Following that, a brief discussion of the difficulties surrounding time to convergence will explain and introduce the definition of the convergence behavior that this paper addresses.

The demonstration of the main result is contained in Section 3, and the final section is devoted to concluding remarks. The reader is referred to Akin (1993), Loomis and Sternberg (1968), Royden (1968), and Belitskii and Lyubich (1988) for background material on the stable manifold theorem, calculus, real analysis, and matrix norms. The notation used in this paper is standard; these sources are also recommended for definitions.

1.1 Random Heuristic Search

This review is included for completeness. A more thorough account can be found in Vose and Wright (1993) and Vose (in press).

An instance of *random heuristic search* can be thought of as an initial collection of elements P_0 chosen from the search space Ω of length ℓ binary strings, together with some transition rule τ, which from P_i will produce another collection P_{i+1}. In general, τ will be iterated to produce a sequence of collections

$$P_0 \xrightarrow{\tau} P_1 \xrightarrow{\tau} P_2 \xrightarrow{\tau} \cdots$$

The beginning collection P_0 is referred to as the *initial population*; the first population (or *generation*) is P_1; the second generation is P_2; and so on. Populations are generated successively until some stopping criteria is reached, at which point it is hoped that the object of search was encountered along the way.

The algorithms comprising random heuristic search are further constrained by what transition rules are allowed. Obtaining a good representation for populations is the first step toward characterizing admissible τ. Define the *simplex* to be the set

$$\Lambda = \{\langle x_0, \ldots, x_{2^\ell - 1}\rangle : x_j \in \Re, \quad x_j \geq 0, \quad \Sigma x_j = 1\}$$

Integers in the interval $[0, 2^\ell)$ are identified with elements of Ω through their binary representation. An element p of Λ corresponds to a population according to the following rule for defining its components:

p_i = the proportion of i contained in the population

Populations are bags. The cardinality of each generation is a constant r called the *population size*. Hence, the proportional representation given by p unambiguously determines a population once r is known. The vector p is referred to as a *population vector* (or *descriptor*).

Given the current population P, the next population $Q = \tau(P)$ cannot be predicted with certainty because τ is stochastic; Q results from r independent, identically distributed random choices. Let $\mathcal{G} : \Lambda \to \Lambda$ be a *heuristic function* that, given the current population vector p, produces a vector whose ith component is the probability that i is the result of a random choice. That is, $\mathcal{G}(p)$ is that probability vector which specifies the distribution from which the aggregate of r choices forms the next generation Q.

In terms of search, P is the starting configuration with corresponding descriptor p, and $\mathcal{G}(p)$ is the *bias* according to which the search space is to be explored. The result $Q = \tau(P)$ of that exploration invokes a new bias $\mathcal{G}(q)$, and the cycle repeats (here q is the descriptor of Q).

Perhaps the first and most natural question concerning random heuristic search is: What connection is there between the heuristic function used and the expected next generation? In Vose and Wright (1993) it is shown that if p is the current population vector and $\mathcal{G}$ is the heuristic, then the expected next population vector is $\mathcal{G}(p)$. According to the law of large numbers, if the next generation's population vector q were obtained as the result of an infinite sample from the distribution described by $\mathcal{G}(p)$, then q would match the expectation; hence, $q = \mathcal{G}(p)$. Because this corresponds to random heuristic search with an infinite population, the algorithm resulting from $\tau = \mathcal{G}$[1] is called the *infinite population algorithm*.

Identifying a population with the population vector that represents it, the possible populations of size r correspond to a set S_r of points in the simplex Λ. An exact model of random heuristic search is provided by the following Markov chain (Nix & Vose, 1991). The states are given by elements of S_r, and the transition probabilities for $x, x' \in S_r$ are given by the matrix Q, where

$$Q_{x,x'} = r! \prod_{j=0}^{2^\ell - 1} \frac{\{\mathcal{G}(x)_j\}^{rx'_j}}{(rx'_j)!}$$

1 Strictly speaking, τ produces the next generation from the current, whereas $\mathcal{G}$ produces the *representation* of the expected next generation from the representation of the current. This distinction—between a population and the vector that represents it—is conveniently blurred.

An instance of random heuristic search is *ergodic* if the finite population model (the Markov chain defined above) is ergodic for every population size. This is always the case for a simple genetic algorithm (GA), given positive mutation (since the components of $\mathcal{G}$ are positive in that case), and has the consequence that the GA will visit every state infinitely often.

Observe that since the representation of a population by a population vector x is proportional, x may represent populations of unspecified size. As $r \to \infty$, the state space of the Markov chain becomes dense in Λ, and the limit of the transition behavior of the Markov chain is a discrete dynamical system on Λ with transition operator $\mathcal{G}$. Thus, both the finite and infinite population models are viewed as operating in Λ.

Many generational GAs currently in use fit into the framework of random heuristic search, provided that only one child of the recombination/mutation process is kept. If the following outline is followed, then an instance of random heuristic search is obtained:

1. Generate an initial population x containing r elements from Ω.
2. Produce a "child" for the next generation using any stochastic function of x.
3. If the next generation is incomplete, repeat step 2.
4. Replace x by the new generation just formed and go to step 2.

The heuristic corresponding to this instance of random heuristic search is determined by the component equations

$$\mathcal{G}(x)_i = \text{Prob}\{i \text{ results from step 2, given the population } x\}$$

As may be imagined, the behavior of random heuristic search depends largely on properties of the heuristic function $\mathcal{G}$. An instance of random heuristic search is *focused* if $\mathcal{G}$ is continuously differentiable and for every $p \in \Lambda$ the sequence of iterates

$$p, \quad \mathcal{G}(p), \quad \mathcal{G}(\mathcal{G}(p)), \quad \ldots$$

converges. In this case, $\mathcal{G}$ is also said to be focused. In terms of search, this condition means that following the path of expectations will lead to some state x. By the continuity of $\mathcal{G}$,

$$\mathcal{G}(\lim_{l\to\infty} \mathcal{G}^l(p)) = \lim_{l\to\infty} \mathcal{G}^{l+1}(p) = x$$

Hence, such points x satisfy $\mathcal{G}(x) = x$ and are called *fixed points* of $\mathcal{G}$. A fixed point x is *hyperbolic* if no eigenvalue of the differential $d\mathcal{G}_x$ of $\mathcal{G}$ at x has modulus 1.

It is a simple and instructive exercise to construct examples of random heuristic search that are not focused. However, when an instance is focused and the fixed points are hyperbolic, then the infinite population algorithm converges in logarithmic time, provided that $\mathcal{G}$ is well behaved. What this means precisely and how it is proved is the subject of the next section.

1.2 Logarithmic Convergence

The precise definition of logarithmic time to convergence faces several obstacles. The most obvious is that ergodic random heuristic search does not converge, since it corresponds to an ergodic Markov chain. Because genetic search is typically conducted with some nonzero level of mutation, it follows that convergence, strictly speaking, does not typically take place for GAs.

The naive definition of convergence as time to discover the optimal is generally useless as well. The "no free lunch theorem" (Wolpert & Macready, 1995) implies that, even with an elitest strategy and aggregating (or collapsing) all populations containing the optimal into an absorbing state, time to convergence is in general no better than that achieved by enumeration. The underlying problem here is that the metric of how good random heuristic search is at *function optimization* is in general worthless to gauge inherent emergent behavior.

The transition probabilities exhibited in the previous section for random heuristic search (i.e., $Q_{x,x'}$) may be interpreted as indicating that the transition from population x to the next generation is given by $\mathcal{G}(x)$ plus multinomially distributed "noise." If the perturbations effected by this noise are not too great, then the initial transient of random heuristic search is characterized by moving toward and spending time in the vicinity of that fixed point $\omega(x)$ to which the underlying dynamical system $\mathcal{G}(x), \mathcal{G}^2(x), \mathcal{G}^3(x), \ldots$ converges.

In fact, the scenario above is provably correct as the population size grows, since the magnitude of the noise decreases with increasing population size. It is therefore natural to consider the time to convergence of $\mathcal{G}(x), \mathcal{G}^2(x), \mathcal{G}^3(x), \ldots$ as an indication of the "settling time" of the initial transient, that is, of how long it might take for random heuristic search to move into the vicinity of $\omega(x)$, assuming that the multinomially distributed "noise" is not too great.

Even after accepting the above concept as an interesting one to pursue, several problems remain. Whereas considering the time to convergence of $\mathcal{G}(x), \mathcal{G}^2(x), \mathcal{G}^3(x), \ldots$ would eliminate stochasticity, it trivializes the matter. In the case of the simple GA, Vose and Wright (1997) have proved that $\mathcal{G}$ is typically invertible, and thus, strictly speaking, the time to convergence is typically zero or infinite according to whether x is a fixed point.

The essential point made above was that random heuristic search, under the influence of the underlying dynamical system corresponding to $\mathcal{G}$, may temporarily explore the *vicinity* of $\omega(x)$. This being the case, *approaching* $\omega(x)$ is what matters, and if the concept to be pursued is how the *signal* component provided by $\mathcal{G}$ (as opposed to the multinomially distributed noise component) relates to this issue, then the most straightforward way to capture the essential idea is to determine, for every δ, the time taken by $\mathcal{G}(x), \mathcal{G}^2(x), \mathcal{G}^3(x), \ldots$ to come within δ of $\omega(x)$. So as to streamline exposition, the time referred to in the last sentence—which obviously depends on x and δ—will be referred to as "time to convergence."

Difficulties remain. Perhaps the most obvious is that the time to convergence depends on the initial population, and there is nothing to prevent the existence of a sequence of initial populations along which the time to convergence diverges to infinity. For example, consider any instance of focused random heuristic search such that u and v are distinct attracting fixed points, and let $s(t) = tu + (1 - t)v$. Let t^* be the supremum of $t \in [0, 1]$ such that $\omega(s(t)) = v$. If the time to convergence to v were bounded, say by k, then by the uniform continuity of $\mathcal{G}^k$ (it is continuous, and Λ is compact), it follows that $\mathcal{G}^k(s(t^*))$ is mapped within δ of v and hence converges to v (for suitably small δ), since v is an attractor. But this contradicts the previous statement that t^* was the supremum, because the same continuity argument would imply that an open neighborhood of t^* converges to v. Therefore, in general, the time to convergence cannot be uniformly bounded.

However, the possibility remains that time to convergence could be uniformly bounded for "most" initial populations. Let a probability density ρ be given over Λ, and for any $A \subset \Lambda$, define the probability that the initial population is contained in A as

$$\int_A \rho \, d\lambda$$

where λ is Lebesgue measure. A natural definition of "most" is a set of probability at least $1 - \varepsilon$ for small ε.

A position has now been reached where a reasonable definition can be formulated: *Logarithmic convergence* of the infinite population algorithm is defined to mean that for every probability density ρ and every $\varepsilon > 0$, there exists a set A of probability at least $1 - \varepsilon$ such that if the initial population p is in A, then the number of generations k required for $\|\mathcal{G}^k(p) - \omega(p)\| < \delta$ is $O(-\log \delta)$, where $\omega(p)$ denotes $\lim_k \mathcal{G}^k(p)$, and $0 < \delta < 1$.

This paper proves logarithmic convergence only for a suitably chosen norm. The restriction that $0 < \delta < 1$ in the definition of logarithmic convergence is made to streamline the proof given in the next section. Since all norms are equivalent (the dimension of Λ is finite), there is no loss of generality, and the norm $\|\cdot\|$ may be chosen such that the diameter of Λ is less than a number $\alpha < 1$.

The main result is that if $\mathcal{G}$ is focused and its fixed points are hyperbolic, then the infinite population algorithm converges in logarithmic time, provided that $\mathcal{G}$ is well behaved. What is meant by *well behaved* is that if $C \subset \Lambda$ has measure zero, then the set $\mathcal{G}^{-1}(C)$ also has measure zero. When $\mathcal{G}$ has a local inverse that is continuously differentiable, then $\mathcal{G}$ is well behaved, since $\overline{C}$ is compact (it is a subset of Λ), and the standard change of variables formula

$$\int_{\mathcal{G}^{-1}(U)} d\lambda = \int_U |\det(d\mathcal{G}_x^{-1})|\, d\lambda(x)$$

may be used locally on C. By the inverse function theorem, a local inverse exists on the complement of the set $\mathcal{G}(B)$, where $B = \{x : \det(\mathcal{G}_x) = 0\}$. It follows that if $\lambda(B) = 0$, then $\mathcal{G}$ is well behaved. In the case of the simple GA, $\lambda(B) = 0$ is typically true (see the conclusion of this paper for further details).

The next section gives the proof of logarithmic convergence by proceeding through a series of steps, each reducing the problem to a simpler one.

2. Demonstration

It is assumed throughout the rest of this paper that $\mathcal{G}$ is focused and well behaved and has hyperbolic fixed points. The first proposition uses compactness and hyperbolicity to establish the basically finite nature of the problem.

PROPOSITION 2.1: *There are only finitely many fixed points of* $\mathcal{G}$.

SKETCH OF PROOF: Otherwise, by compactness of Λ, there would be a convergent sequence of them, say $\lim x_i = x$. Since $\mathcal{G}$ is continuous,

$$\mathcal{G}(x) = \mathcal{G}(\lim x_i) = \lim \mathcal{G}(x_i) = \lim x_i = x$$

Hence, x is also a fixed point. By compactness of the unit sphere, let η be a limit of the set

$$\left\{ \frac{x_i - x}{\|x_i - x\|} \right\}$$

and let i_j be an increasing sequence of indices for which $\eta = \lim_j (x_{i_j} - x)/\|x_{i_j} - x\|$. Note that

$$x_{i_j} = \mathcal{G}(x + (x_{i_j} - x)) = \mathcal{G}(x) + d\mathcal{G}_x(x_{i_j} - x) + o(x_{i_j} - x)$$

Subtracting x, dividing by $\|x_{i_j} - x\|$, and taking the limit as $j \to \infty$ yields

$$\eta = \lim_j d\mathcal{G}_x \left(\frac{x_{i_j} - x}{\|x_{i_j} - x\|} \right) + o\left(\frac{x_{i_j} - x}{\|x_{i_j} - x\|} \right) = d\mathcal{G}_x \eta$$

which contradicts the hypothesis that the fixed point x is hyperbolic. □

A fixed point x is *stable* if the eigenvalues of the differential $d\mathcal{G}_x$ of $\mathcal{G}$ at x have a modulus less than 1. A fixed point x is *unstable* otherwise (by the hyperbolicity assumption, some eigenvalue must have a modulus greater than 1 in this case). The *basin of attraction* of x is the set

$$\mathcal{B}_x = \{y : \lim_{k\to\infty} \mathcal{G}^k(y) = x\}$$

Let $\mathcal{S}$ be the union of $\mathcal{B}_x$ over stable x, and let $\mathcal{U}$ be the union of $\mathcal{B}_x$ over unstable x. The next result shows that attention can be focused on $\mathcal{S}$, since the complement has probability zero.

PROPOSITION 2.2: *With respect to every probability density, $\mathcal{U}$ has probability zero.*

SKETCH OF PROOF: Since probabilities are computed by integration with respect to Lebesgue measure, it suffices to show that $\lambda(\mathcal{U}) = 0$. Since there are countably many $\mathcal{B}_x$'s (Proposition 2.1), and since

$$\mathcal{B}_x = \bigcup_{k\geq 0} \mathcal{G}^{-k}(U)$$

where U is the intersection of a small neighborhood of x with the stable manifold at x, it suffices that $\lambda(U) = 0$. Since x is an unstable fixed point, the stable manifold theorem shows U to be the graph of a function over the projection to the stable subspace.

Note that there are uncountably many disjoint translations of the graph within a small neighborhood of x (move along any unstable direction). Since λ is translation invariant, it follows that $\lambda(U) = 0$. □

The following proposition shows that the set of typical initial populations may be taken to be compact.

PROPOSITION 2.3: *For every ρ and every $\varepsilon > 0$, there exists a compact subset of $\mathcal{S}$ having probability at least $1 - \varepsilon$.*

SKETCH OF PROOF: Let U_x be a small closed neighborhood of the fixed point x, and let U be the union of the U_x over stable fixed points x. Define

$$A_k = \bigcup_{0\leq j\leq k} \mathcal{G}^{-j}(U)$$

and note that the characteristic function of the set A_k converges monotonically to the characteristic function of $\mathcal{S}$. Moreover, each A_k is compact. It follows from Proposition 2.2 that

$$1 = \int_{\mathcal{S}} \rho \, d\lambda = \lim_{k\to\infty} \int_{A_k} \rho \, d\lambda$$

Therefore, a k exists for which the probability of A_k is at least $1 - \varepsilon$. □

The next result contains the heart of the matter. Given ρ and $\varepsilon > 0$, let A be a compact subset of $\mathcal{S}$, as given by Proposition 2.3. Proposition 2.4 shows that it is sufficient to consider local convergence behavior, that is, behavior at a point $x \in A$.

PROPOSITION 2.4: *If for every $x \in A$ there exists an integer N_x such that for* $0 < \delta < 1$,

$$k > -N_x \log\delta \implies \|\mathcal{G}^k(x) - \omega(x)\| < \delta$$

then the infinite population algorithm converges in logarithmic time.

SKETCH OF PROOF: Without loss of generality, the N_x's are minimal. Suppose there exists N such that $N_x < N$ for all $x \in A$. In that case the proof would be complete, since A has probability at least $1 - \varepsilon$, and for every $p \in A$ the number of generations k required for $\|\mathcal{G}^k(p) - \omega(p)\| < \delta$ would be $-N \log\delta$.

If there were no such N, then let x_j be a sequence for which N_{x_j} diverges. Since A is compact, assume that the x_j's converge to x. Let U be a small open ball with center $\omega(x)$ such that $y \in U \implies \|\mathcal{G}(y) - \omega(x)\| < \alpha\,\|y - \omega(x)\|$ for some $\alpha < 1$. Such a neighborhood exists, since the spectral radius of $d\mathcal{G}_{\omega(x)}$ is less than 1. Because the number of fixed points is finite, α may be chosen without regard to which of the stable fixed points $\omega(x)$ is. Let $k > 1$ be such that $\mathcal{G}^k(x) \in U$, and by continuity let V be an open neighborhood of x that is mapped into U by $\mathcal{G}^k$. Hence, if $t \geq k$, then

$$\sup_{v\in V} \|\mathcal{G}^t(v) - \omega(x)\| \leq \sup_{u\in U} \|u - \omega(x)\|\,\alpha^{t-k} < \alpha^{1+t-k}$$

since the norm may be chosen so that the diameter of Λ is less than α. This inequality will be referred to as $(*)$. Given δ, choose integer t such that $\alpha^{1+t-k} \leq \delta < \alpha^{t-k}$, and let $N = -2k/\log\alpha$. Consider first the case where $t > k$ and note that (in this case)

$$-N \log\delta > (2k/\log\alpha)(t - k)\log\alpha > t$$

Combining this with $(*)$ yields

$$\sup_{v\in V} \|\mathcal{G}^{\lceil -N \log\delta\rceil}(v) - \omega(x)\| \leq \sup_{v\in V} \|\mathcal{G}^t(v) - \omega(x)\| < \alpha^{1+t-k} \leq \delta$$

which contradicts that the x_j's enter V (since the N_{x_j}'s are unbounded). The remaining case is $t = k$, in which $\alpha \leq \delta$ holds. Hence,

$$\sup_{v\in V} \|\mathcal{G}^{\lceil -N \log\delta\rceil}(v) - \omega(x)\| < \delta$$

is still valid, since the diameter of Λ is less than α. □

In the preceding proof, the inequality $y \in U \implies \|\mathcal{G}(y) - \omega(x)\| < \alpha\,\|y - \omega(x)\|$ requires the choice of a suitable norm, which may depend on $\omega(x)$. Thus, the conclusion of Proposition 2.4 might appear to be valid only in the basin of attraction of $\omega(x)$. However, there are only finitely many basins (Proposition 2.1), and Proposition 2.4 is valid in each of them separately. Moreover, if

$$k > -N_x \log\delta \implies \|\mathcal{G}^k(x) - \omega(x)\| < \delta$$

for one norm, then the same is true for any other norm by simply adjusting the constant N_x.

According to Proposition 2.4, all that remains is Proposition 2.5:

PROPOSITION 2.5: *For every $x \in A$ there exists an integer N_x such that for* $0 < \delta < 1$,

$$k > -N_x \log\delta \implies \|\mathcal{G}^k(x) - \omega(x)\| < \delta$$

SKETCH OF PROOF: It was shown in the proof of Proposition 2.4 that there exists a neighborhood V of x such that

$$\sup_{v \in V} \|\mathcal{G}^{\lceil -N \log \delta \rceil}(v) - \omega(x)\| < \delta$$

where N depends on x. Moreover, the proof also makes clear that further iterations of $\mathcal{G}$ further decrease the distance to $\omega(x)$. Since $x \in V$, an acceptable choice is therefore $N_x = N$. □

This completes the proof of the following theorem:

THEOREM 2.1: *If $\mathcal{G}$ is focused and its fixed points are hyperbolic, then the infinite population algorithm converges in logarithmic time, provided that $\mathcal{G}$ is well behaved.*

3. Conclusion

Random heuristic search did not have to be defined in terms of fixed-length strings or constant population sizes. Neither is the proof of logarithmic convergence given in this paper valid only for such generalizations. On the other hand, inherent emergent behavior of the simple GA is the subject of this paper, and the framework in which logarithmic convergence has been discussed is useful for that purpose; it forms a reasonably general context in which the simple GA may be better understood.

It should be appreciated that Theorem 2.1 holds in remarkable generality. In the case of the simple GA, it is not just a statement about selection or crossover or mutation. The theorem deals with the behavior of genetic search comprising all these operators, interleaved as they are in standard practice.

In general terms, what this paper has to say about convergence of the infinite-population simple GA is this: If fixed points are hyperbolic and $\mathcal{G}$ is focused and well behaved, then the coupling between "how close" and "how long" is logarithmic. The assumptions (hyperbolic fixed points, focused, and well-behaved $\mathcal{G}$) and the implications of the logarithmic coupling for finite-population GAs will be touched on in the following two subsections.

3.1 Hypotheses

It is conjectured that an arbitrarily small perturbation of fitness will ensure the hyperbolicity hypothesis. It seems that the fitness function or selection scheme would have to be contrived to violate it; every randomly generated example that has been examined has been hyperbolic. Moreover, Eberlein and Vose[2] have recently proved a strong form of the hyperbolicity conjecture for GAs using binary fixed-length strings with proportional selection, arbitrary crossover, and positive mutation.

That $\mathcal{G}$ is focused for mutation rates less than 0.5 is a fundamental conjecture of the simple GA. The conjecture has been proved, assuming that the mutation rate is small, for fitness functions with low epistasis (Vose & Wright, 1993). Violating the constraint that the mutation rate be less than 0.5 is known to lead to counterexamples (Wright & Bidwell, 1997).

As noted in a previous section, if the set $B = \{x : \det(\mathcal{G}_x) = 0\}$ has measure zero, then $\mathcal{G}$ is well behaved. There are formulas which for arbitrary mutation and crossover decide this question. In particular, if the mutation rate is in the range $(0, 0.5)$ and the crossover rate is less than 1, then B has measure zero (Vose, in press).

---

2 This result is part of the doctoral dissertation by Mary Eberlein (C. S. Department, University of Tennessee, Knoxville, 1996).

3.2 Implications

Logarithmic convergence relates to finite population GA behavior in that if the perturbations from $\mathcal{G}$ effected by the multinomially distributed noise are not too great, then the initial transient is characterized by moving toward and spending time in the vicinity of that fixed point $\omega(x)$ to which the underlying dynamical system $\mathcal{G}(x), \mathcal{G}^2(x), \mathcal{G}^3(x), \ldots$ converges. Hence, the logarithmic coupling indicates a rapid "settling time" of the initial transient, that is, of how long it might take for random heuristic search to move into the vicinity of $\omega(x)$. Once leaving the vicinity of $\omega(x)$, a new transient phase is initiated, characterized (as described above) by moving toward and spending time in the vicinity of a (possibly different) fixed point. This scenario repeats, giving rise to punctuated equilibria.

The reader is cautioned to interpret the previous paragraph carefully. Although a GA is characterized by punctuated equilibria, the "equilibria" involved are dynamically changing populations spatially located in the vicinity of certain locations in Λ. Indeed, equilibria in the form of absorbing static populations cannot exist due to the ergodicity of the Markov chain. Moreover, this paper says nothing about the duration of time spent at a dynamic equilibrium; evolutionary trajectories can move rapidly through a succession of them, thereby giving the appearance of not reaching equilibria, due to not dwelling at them. Another possibility is that the initial transient lasts for such a long time that punctuated equilibria are not observed because only one is encountered in practice.

To sort this out, one might imagine a "meta-level" Markov chain having as states the fixed points of $\mathcal{G}$ that spatially represent where in Λ the dynamic equilibria are near. The transition matrix of this meta-level chain would then describe the transitions among the dynamic equilibria. In particular, its structure would shed light on which of the two extreme behaviors mentioned in the previous paragraph—or what in between—is taking place (assuming, of course, that the noise is not overwhelming the signal). The reader is referred to Vose (1996) for the definition and characterization of the transition matrix of this meta-level Markov chain in the large population case.

It should be pointed out that the description of the inherent emergent behavior of genetic search presented in this paper does not involve global or local optima of the fitness function. This is because simple GAs of the type described in this paper (which, in fact, are commonly used for the purpose of function optimization) are not function optimizers; any honest characterization of their emergent behavior—which is the intent of this paper—could not, in general, have much to do with local/global optima of the search problem. For an abstract perspective on this issue in which the role played by representation is emphasized, see Radcliffe and Surry (1995).

The answers to applied questions concerning the suitability of genetic search for optimization will come, I believe, from a deeper understanding of when and how the inherent emergent behavior of genetic search aligns with the requirements of successful optimization. Prerequisite to this scenario is the resolution and analysis of what inherent emergent behavior genetic search exhibits.

Finally, it should be mentioned that the question of when, and to what degree, the signal component of a finite-population GA (as given by $\mathcal{G}$) is not overwhelmed by the noise component (given by the multinomial distribution) remains an open area of research.

Acknowledgments

This research was supported by the National Science Foundation IRI-9224917 and CDA-9115428.

References

Akin, E. (1993). *The general topology of dynamical systems.* Providence, RI: American Mathematical Society.

Belitskii, G. R., & Lyubich, Y. I. (1988). *Matrix norms and their applications.* Basel: Birkhäuser.

Loomis, L. H., & Sternberg, S. (1968). *Advanced calculus.* Reading, MA: Addison-Wesley.

Nix, A., & Vose, M. D. (1991). Modeling genetic algorithms with Markov chains. *Annals of Mathematics and Artificial Intelligence, 5* (1992), 79–88.

Radcliffe, N., & Surry, P. (1995). Fundamental limitations on search algorithms: Evolutionary computing in perspective. *Lecture notes in computer science, Vol. 1000* (pp. 255–291). New York: Springer-Verlag.

Royden, H. L. (1968). *Real analysis.* New York: Macmillan.

Vose, M. D. (1996). Modeling simple genetic algorithms. *Evolutionary Computation, 3*(4), 453–472.

Vose, M. D. (in press). *The simple genetic algorithm: Foundations and theory.* Cambridge, MA: MIT Press.

Vose, M. D., & Wright, A. H. (1993). Simple genetic algorithms with linear fitness. *Evolutionary Computation, 2*(4), 347–368.

Vose, M. D., & Wright, A. H. (1997). *The simple genetic algorithm and the Walsh transform: Part II, the inverse.* Manuscript submitted for publication.

Wolpert, D. H., & Macready, W. G. (1995). *No free lunch theorems for search.* (Available at www.santafe.edu/sfi/publications/95wplist.html 95-02-010)

Wright, A., & Bidwell, G. (1997). A search for counterexamples to two conjectures on the simple genetic algorithm. In R. Belew & M. D. Vose (Eds.), *Foundations of genetic algorithms IV* (pp. 73–84). San Francisco, CA: Morgan Kaufmann.