Logarithmic convergence of random heuristic search

Michael D. Vose

Computer Science Dept., 107 Ayres Hall
The University of Tennessee
Knoxville, TN 37996-1301

ABSTRACT

A general class of stochastic search algorithms, *random heuristic search*, is reviewed. A general convergence theorem for this class is proved. Since the simple genetic algorithm is an instance of random heuristic search, a corollary is a result concerning GAs and logarithmic time to convergence.

2. INTRODUCTION

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

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

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 towards characterizing admissible τ. Define the *simplex* to be the set

$$\Lambda \;=\; \{< x_0, ..., x_{2^\ell - 1} > \,:\, x_j \in \Re,\ x_j \geq 0,\ \Sigma\, x_j = 1\}$$

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

$$p_i \;=\; \text{the proportion of } i \text{ contained in the population}$$

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*).

0-8194-1628-2/94/$6.00

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 $\mathcal{G} : \Lambda \rightarrow \Lambda$ be a function which given the current population vector p produces a vector whose i th 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 *heuristic* according to which the search space is to be explored. The result $Q = \tau(P)$ of that exploration invokes a new heuristic $\mathcal{G}(q)$ 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 used and the expected next generation? In [2] it was 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.*

It was also shown in [2] that the simple genetic algorithm is an instance of random heuristic search. Thus the results of this paper are applicable to the simple GA.

Random heuristic search algorithms are classified according to the behavior of their heuristic functions. An instance of random heuristic search is *focused* if $\mathcal{G}$ is continuously differentiable and for every $p \in \Lambda$ the sequence

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

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

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

Hence such points x satisfy $\mathcal{G}(x) = x$ and are called *fixed points* of $\mathcal{G}$.

It is a simple and instructive exercise to construct examples of random heuristic search which 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 following section.

In the case of the simple genetic algorithm, $\mathcal{G}$ is typically well behaved and a few classes of fitness functions are known to produce focused heuristics (some with restrictions on mutation and others without)[1]. The *fundamental conjecture of simple genetic algorithms* is that every fitness function produces a focused heuristic with arbitrary crossover and mutation[2]. In that case the results of this paper would imply logarithmic convergence is typical for the infinite population simple GA when fixed points are hyperbolic.

[1]Strictly speaking, τ produces the next generation from the current, while $\mathcal{G}$ produces the *representation* of the next generation from the *representation* of the current. This distinction is conveniently blurred.

[2]Provided the mutation rate is in the interval $[0, \frac{1}{2}]$.

3. LOGARITHMIC CONVERGENCE

A fixed point x is *hyperbolic* if the differential $d\mathcal{G}_x$ of $\mathcal{G}$ at x has no eigenvalues on the unit circle.

The precise definition of logarithmic time to convergence faces several obstacles. Perhaps the most obvious is that the time to convergence may depend upon 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.

One way around this difficulty is to let a probability density ρ be given over Λ and for $A \subset \Lambda$ to define the probability that the initial population is contained in A as

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

where λ is Lebesgue measure. Hence any single sequence of initial populations along which the time to convergence diverges would constitute a set of zero probability. On the other hand, the situation is not totally trivialized since the density ρ is to be arbitrary.

The task then becomes to show that for every 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 time to convergence is logarithmic.

The next difficulty is that, typically, converging populations will never get to the population towards which they are converging. It is therefore natural to let $0 < \delta < 1$ denote how close the populations are required to get to the limit, and then to require that they do so, within $O(-\log \delta)$ generations.

To summarize, *logarithmic convergence* is defined to mean: 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 $0 < \delta < 1$.

In the introduction, it was stated that when $\mathcal{G}$ 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 ment by *well behaved* is that if C has measure zero, then so does the set $\mathcal{G}^{-1}(C)$. When $\mathcal{G}$ has a local inverse which is continuously differentiable, then $\mathcal{G}$ is well behaved since $\overline{C}$ is compact (it is a subset of Λ) and the formula

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

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 genetic algorithm, $\lambda(B) = 0$ is typically true [1].

Finally, since all norms are equivalent (the dimension of Λ is finite), this paper proves logarithmic convergence only for a suitablely chosen norm. This avoids trivialities which would otherwise complicate matters.

The next section gives the proof of logarithmic convergence by proceeding through a series of steps, each reducing the problem to a simpler one. It is assumed throughout the rest of this paper that $\mathcal{G}$ is focused, well behaved, and has hyperbolic fixed points.

4. DEMONSTRATION

Proposition 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,

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

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 a sequence of indices for which $\eta = \lim_j (x_{i_j} - x)/\|x_{i_j} - x\|$. Note that

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

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

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

which contracts the hypothesis that the fixed point x is hyperbolic.

A fixed point x is *stable* if the spectrum of the differential $d\mathcal{G}_x$ at x is less than 1. A fixed point x is *unstable* if the spectrum of $d\mathcal{G}_x$ is greater than 1. The *basin of attraction* of x is the set

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

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.

Proposition 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 $\lambda(\mathcal{U}) = 0$. Since there are countably many $\mathcal{B}_x$ (proposition 1), and since

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

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 translates of the graph within a small neighborhood of x (move along any unstable direction). Since λ is translation invariant, it follows that $\lambda(U) = 0$.

By proposition 2, attention can be focused on $\mathcal{S}$ since the complement has probability zero.

Proposition 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

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

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 that

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

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

Given ρ and given $\varepsilon > 0$, let A be the compact subset of $\mathcal{S}$ which exists by proposition 3. The next proposition shows it is sufficient to consider local convergence behavior.

Proposition 4: If for every $x \in A$ there exists an integer N_x such that for $0 < \delta < 1$

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

then the infinite population algorithm converges in logarithmic time.

Sketch of proof: Without loss of generality, the N_x 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 $p \in A$ 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 the x_j 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 are 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 which is mapped into U by $\mathcal{G}^{(k)}$. Hence if $t \geq k$, then

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

since the norm may be chosen so that the diameter of Λ is α. This inequality will be referred to as $(*)$. Given δ, choose t such that $\alpha^{1+t-k} \leq \delta < \alpha^{t-k}$, and let $N = -k/\log \alpha$. Note that

$$-N \log \delta \quad \geq \quad (k/\log \alpha)(1 + t - k) \log \alpha \quad \geq \quad t$$

Combining this with $(*)$ yields

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

which contradicts that the x_j enter V (since the N_{x_j} are unbounded).

According to proposition 4, all that remains is:

Proposition 5: For every $x \in A$ there exists an integer N_x such that for $0 < \delta < 1$

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

It was shown in the proof of proposition 4 that there exists a neighborhood V of x such that

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

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 $N_x = N$.

5. CONCLUSION

Random heuristic search need not be limited to binary alphabets, fixed length strings, or constant population sizes. Neither is the proof of logarithmic convergence given in this paper valid only for a generalization of random heuristic search.

On the other hand, the simple genetic algorithm is the object which I have chosen to study, and it seems to me that the framework in which logarithmic convergence has been discussed in this paper is most appropriate for that purpose; it forms a reasonably general context in which the simple genetic algorithm may be better understood.

What this paper has to say about the infinite population simple genetic algorithm is this: if fixed points are hyperbolic and $\mathcal{G}$ is focused and well behaved, then convergence is logarithmic. Each of these conditions will be briefly commented on below.

Hyperbolic fixed points

It is conjectured that an arbitrarily small perturbation of fitness will insure this. It seems fitness would have to be contrived to violate hyberbolicity; every randomly generated example has been hyperbolic.

$\mathcal{G}$ is focused

This is the fundamental conjecture of the simple genetic algorithm. It is known to be true when fitness is linear (and zero mutation) [2] or when fitness is close to constant (and mutation in the range $(0, \frac{1}{2})$) [1]. Moreover, it is empirically true for injective fitness functions (no counterexample has ever been found).

$\mathcal{G}$ is well behaved

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 [1].

6. ACKNOWLEDGEMENTS

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

7. REFERENCES

[1] M. D. Vose, The Simple Genetic Algorithm: Foundations and Theory, To be published by MIT Press.

[2] M. D. Vose & A. H. Wright (1993), "Simple Genetic Algorithms With Linear Fitness", Submitted to *Evolutionary Computation.*