A lower-bound result on the power of a genetic algorithm

Kihong Park *
park@cs.bu.edu

BU-CS-94-009
July 31, 1993

Computer Science Department
Boston University
Boston, MA 02215

Abstract

This paper presents a lower-bound result on the computational power of a genetic algorithm in the context of combinatorial optimization. We describe a new genetic algorithm, the merged genetic algorithm, and prove that for the class of monotonic functions, the algorithm finds the optimal solution, and does so with an exponential convergence rate. The analysis pertains to the ideal behavior of the algorithm where the main task reduces to showing convergence of probability distributions over the search space of combinatorial structures to the optimal one. We take exponential convergence to be indicative of efficient solvability for the sample-bounded algorithm, although a sampling theory is needed to better relate the limit behavior to actual behavior. The paper concludes with a discussion of some immediate problems that lie ahead.

*I would like to thank Dr. Peter Gacs and Marcus Peneido for helpful discussions. A summary paper appears in *Proc. 5th International Conference on Genetic Algorithms*.

1 Introduction

This paper presents a lower-bound result on the computational power of genetic algorithms in the context of combinatorial optimization. We introduce a new genetic algorithm, the merged genetic algorithm, and show that for monotonic functions, the algorithm finds the optimal solution, and does so fast. In particular, we prove that a probability measure which completely characterizes the state of the genetic algorithm, converges to the limiting distribution encoding the optimal solution with an exponential convergence rate. The algorithm combines reproduction and cross-over operations in a novel, but simple way, which makes its behavior amenable to rigorous mathematical analysis. A previous result of [7] has shown that the n-bit MAX-SUM function, $x_1 + x_2 + \ldots + x_n$, lies within the reach of a standard genetic algorithm that uses n-point cross-over. Our result extends the class of functions efficiently solvable by a genetic algorithm (albeit nonstandard) to monotonic functions which, in some sense, are the simplest representant of the "building-block hypothesis" [3].

In technical considerations, the following points can be made regarding the difficulty of dealing with various aspects of genetic algorithms:

1. *Mutation alone.* Mutation by itself is easy to analyze. It corresponds to doing a random walk in the space of all combinatorial structures via local transitions. Although every point gets visited eventually (assuming the space is bounded), it is too expensive.

2. *Reproduction alone.* Reproduction has the effect of increasing the average fitness of a population to the fitness level of the fittest element in the initial population. Reproduction by itself cannot be viewed as doing any meaningful search.

3. *Reproduction + mutation.* Although less trivial than 1 and 2, the analysis is straightforward. Basically, it corresponds to "zooming in" via reproduction coupled with localized search through random perturbations.

4. *Crossover alone.* By no means trivial. The mixing properties of cross-over alone can be fruitfully analyzed using Markov chain techniques [4, 2, 6]. A recent result of [6] shows that under weak restrictions on the cross-over operator (symmetricity and aperiodicity), its stationary distribution is unique and easily characterizable.

5. *Reproduction + cross-over.* This seems to be the most interesting case. When cross-over is combined with reproduction, tracking the behavior of a population becomes a

difficult task because reproduction tends to disrupt the homogenizing effect of cross-over. In [7], a clever scheme of "jumping" across representations has made the analysis of the MAX-SUM function tractable. In this paper, a different approach is taken, whose main advantage lies in the "merging" of reproduction and cross-over in a new algorithm, whose behavior is more easily analyzable.

6. *Reproduction + cross-over + mutation.* Although this represents the most general situation, it can be viewed as case 5 plus perturbation. As such, it does not pose any new problems.

This paper is organized as follows. First, we give a brief overview of the probabilistic approach employed here and some background motivation. Second, we define monotonic functions and describe the merged genetic algorithm with its associated equation of motion. Third, we analyze the ideal behavior of the system by showing convergence to the solution distribution with an exponential rate. We conclude with remarks on future directions.

2 Background

Let $\mathcal{S}$ be a finite set of combinatorial structures. For example, $\mathcal{S}$ may be the set of all inputs to a n-variable Boolean function, the set of all subgraphs of some graph G, or the set of all hands in a card game such as Poker. A *population* $\mathcal{H}$ of size N is a multi-set consisting of N elements from $\mathcal{S}$. Let $\mathcal{A}$ be the set of all populations of size N. Cross-over and reproduction are probabilistic algorithms that map $\mathcal{A}$ into itself, and as such, they induce a stochastic process on $\mathcal{A}$. A *fitness function* (or objective function) f is a mapping $f : \mathcal{S} \to \mathbb{R}$. The basic goal in most applications is to find an element $x^* \in \mathcal{S}$ that optimizes f. Since $\mathcal{S}$ is usually a prohibitively large set, the strategy employed by genetic algorithms is to take a small sample (i.e. population) of size $N \ll |\mathcal{S}|$, and by generating successive populations based upon the two main operators[1] hopefully end up producing a population that contains elements that are close to optimal with respect to f.

For N sufficiently large, the ideal behavior of such algorithms can be described in a probabilistic setting as follows. Let $\mathcal{M}$ be the set of all probability measures on $\mathcal{S}$. Then a genetic algorithm induces a map $h : \mathcal{M} \to \mathcal{M}$. This is over two successive populations. The

[1]The effect of mutation will be ignored in this presentation due to its secondary nature and easy characterization.

iterative process leads to a dynamical system where the central problem lies in determining the asymptotic behavior of $(h^i(p))$ for $p \in \mathcal{M}$. The optimization problem[2] can be expressed as

$$\max_{p \in \mathcal{M}} \sum_{x \in \mathcal{S}} f(x)p(x).$$

The "solution distribution," p^*, is then given by $p^*(x) = 1$ if $x = x^*$, and $p^*(x) = 0$, $x \neq x^*$ where x^* satisfies $f(x^*) = \max_{x \in \mathcal{S}} f(x)$. Informally, a genetic algorithm can be said to *find* the optimal solution if $h^i(p) \to p^*$ as $i \to \infty$. We can say a genetic algorithm *solves* a problem if it finds the optimal solution to all instances of the problem set. Let T_r and T_c denote the reproduction and cross-over operators, respectively. The most popular form of T_r and T_c in terms of their ideal behavior is given by

$$p'(y) = T_r(p)|_y = \frac{f(y)p(y)}{\sum_{x \in \mathcal{S}} f(x)p(x)}, \qquad p'(z) = T_c(p)|_z = \sum_{x,y \in \mathcal{S}} P_r\{z|x,y\}p(x)p(y).$$

T_c need not be limited to being quadratic and other forms abound. It is clear that $\lim_{i \to \infty} T_r^i(p) = p^*$. Hence T_r "finds" the optimal solution. Nevertheless, since the sample size N needs to be of the same order of magnitude as $|\mathcal{S}|$ to adequately emulate T_r, we cannot say that T_r is an efficient procedure. Formalizations of efficient solvability and a corresponding sampling theory can be found in [5].

For cross-over operator T_c, $P_r\{z|x,y\}$ is a fixed quantity, hence T_c induces a stationary stochastic process. It is not hard to see that by lifting the system to the product space $\mathcal{S} \times \mathcal{S}$ and suitably extending T_c, we get a Markov chain over $\mathcal{S} \times \mathcal{S}$. Thus Markov chain techniques can be applied to this larger space to obtain characterizations, which are then projected back to $\mathcal{S}$ to yield statements about the original system. In [6], this method was employed to show that under weak restrictions on T_c (symmetricity and aperiodicity), its stationary distribution is unique and easily characterizable via linear invariants. The composite map $h = T_c T_r$ can be expressed as

$$p'(z) = h(p)|_z = \sum_{x,y \in \mathcal{S}} \frac{f(x)f(y)}{(\sum_{u \in \mathcal{S}} f(u)p(u))^2} P_r\{z|x,y\}p(x)p(y).$$

Unlike before, $T_c T_r$ is a more complicated map due to inheriting nonstationarity from T_r ($p(u)$ in the denominator), and mixing ($P_r\{z|x,y\}$) from T_c. The analysis of the dynamics of $(T_c T_r)^i$ under various assumptions on f is of extreme interest. Intuitively, for x, $y \in \mathcal{S}$

[2] Without loss of generality, let us only consider maximization problems.

with $f(x)$, $f(y) > \sum_{u \in \mathcal{S}} f(u)p(u)$, the transition probability $P_r\{z|x, y\}$ will receive a "boost" whereas for elements with fitness values below the mean, the opposite occurs. This subject is under investigation and will be reported elsewhere. In this paper, we impose additional structure on $\mathcal{S}$, which in conjunction with the monotonicity assumption on f, allows us to prove the lower-bound results.

3 The merged genetic algorithm

Let X be a finite set, and let the elements of X be indexed by $\{1, 2, \ldots, m\}$. Let $f(x_1, x_2, \ldots, x_n)$ be a n-variable fitness function $f : X^n \to \mathbb{R}$. In our earlier notation, $\mathcal{S} = X^n$. Let $\mathcal{M}$ be the set of probability measures on X^n. A probability measure $p \in \mathcal{M}$ is *fully supported* if $p(x) > 0$, $\forall x \in X^n$. We begin with a definition of monotonicity.

Definition 1 *A function $f : X^n \to \mathbb{R}$ is monotonic if $\forall x_i$, $i \in \{1, \ldots, n\}$, $\exists$ total order $\prec_i$ on X such that $a \prec_i b \Longrightarrow f(x_1, \ldots, x_{i-1}, a, x_{i+1}, \ldots, x_n) < f(x_1, \ldots, x_{i-1}, b, x_{i+1}, \ldots, x_n)$.*

Thus, for example, all multinomials in n variables with positive coefficients are monotonic. This includes the MAX-SUM function as a special case. Intuitively, monotonic functions should be good examplars of the "building-block hypothesis" [3] since the components of the function are maximally independent. Later we will exploit this property to show that this is indeed the case. In the context of combinatorial optimization, an n-argument function from the natural numbers to the reals is clearly sufficient to represent all "interesting" optimization problems, certainly the class NP. For instance, with $x_i \in \{0, 1\}$, f may encode SAT, an NP-complete problem. Finding the largest clique in a graph, whose decision problem is also NP-complete, can be encoded an a n-argument function where n corresponds to the number of vertices in the graph. Unless $P = NP$, these considerations show that a genetic algorithm which operates on f cannot be expected to perform miracles. Even for function computation problems such as MAX-Clique, recent results in complexity theory have shown that approximating the size of the maximum clique is as hard as computing its exact value [1]. Hence attempting to prove most general results on the power of GA's will be as daunting as proving $P \neq NP$.

Let us view each x_i $(i = 1, 2, \ldots, n)$ as a random variable, and define a discrete probability measure p_i for each x_i. Let p_{ij} denote $p_{ij} \equiv P_r\{x_i = x_{ij}\}$. Let t be a discrete time index. We shall see shortly that $p_{ij}(t)$ will be treated as a function of time. Let p be the product measure $p = p_1 \times p_2 \times \ldots \times p_n$. At any time instant $t \geq 0$, the state of the merged genetic algorithm (m-GA) will be completely specified by $p(t)$. First, we give the description of the m-GA.

---

m-GA:

begin

  repeat

    $S_{ij} := 0, \forall i, j.$

    $p := p'.$

    for N times **do**

      generate $v := a_1 a_2 \ldots a_n$ randomly from p.

      $S_{ia_i} := S_{ia_i} + f(a_1, a_2, \ldots, a_n), \forall i.$

    endfor

    $p'_{ij} := S_{ij} / \sum_{k=1}^{m} S_{ik}, \forall i, j.$

  until $||p' - p|| < \epsilon$

end

The mechanics of the algorithm is easy to understand. At each generation, N samples from p are generated, and each p_{ij} is updated by the normalized weight of the sampled f values. Mixing is achieved by generating each component a_i independently with probability p_{ia_i}. Biased selection is done the same way as in standard reproduction. Let $z = (z_1, z_2, \ldots, z_{n-1}) \in X^{n-1}$. Denote by $f_{ij}(z) \equiv f(z_1, \ldots, z_{i-1}, j, z_i, \ldots, z_{n-1})$ and $\Pi_i(z) \equiv p_{1z_1} p_{2z_2} \ldots p_{i-1 z_{i-1}} p_{i+1 z_i} \ldots p_{n z_{n-1}}$. The equation of motion governing the ideal behavior of the algorithm when N is unbounded is given by

$$p_{ij}(t+1) = p_{ij}(t) \frac{\sum_{z \in X^{n-1}} f_{ij}(z) \Pi_i(z)}{\sum_{k=1}^{m} p_{ik}(t) \sum_{z \in X^{n-1}} f_{ik}(z) \Pi_i(z)} \tag{1}$$

The next section deals with the asymptotic analysis of this equation. First, we prove that the above map is continuous (which is obvious for p fully supported) which is needed for a later proof. The stronger property of local Lipschitz continuity is needed in the context of a sampling theory [5], and is stated here for completeness.

Proposition 1 *Let* $||p-q|| \leq \frac{1}{2m}$. *Then* $||h(p)-h(q)|| < K||p-q||$ *where* $K = K(n,m)$.

Proof Let $p' = h(p)$. It suffices to show that for all i,j, $|p'_{ij} - q'_{ij}| < K(n,m)||p-q||$. Let $\delta = ||p-q||$. *Case 1*: using equation 1,

$$|p'_{ij} - q'_{ij}| = \left| \frac{\sum_{z\in X^{n-1}} f_{ij}(z) p_{1z_1}\dots p_{ij}\dots p_{nz_{n-1}}}{\sum_{k=1}^{m}\sum_{z\in X^{n-1}} f_{ik}(z) p_{1z_1}\dots p_{ik}\dots p_{nz_{n-1}}} - \frac{\sum_{z\in X^{n-1}} f_{ij}(z) q_{1z_1}\dots q_{ij}\dots q_{nz_{n-1}}}{\sum_{k=1}^{m}\sum_{z\in X^{n-1}} f_{ik}(z) q_{1z_1}\dots q_{ik}\dots q_{nz_{n-1}}} \right|$$
$$\leq \frac{\sum_{z\in X^{n-1}} f_{ij}(z) p_{ij}\Pi_i(z)}{\sum_{k=1}^{m}\sum_{z\in X^{n-1}} f_{ik}(z) p_{ik}\Pi_i(z)} - \frac{\sum_{z\in X^{n-1}} f_{ij}(z)(p_{1z_1}-\delta)\dots(p_{nz_{n-1}}-\delta)}{\sum_{k=1}^{m}\sum_{z\in X^{n-1}} f_{ik}(z)(p_{1z_1}+\delta)\dots(p_{nz_{n-1}}+\delta)}$$

yields an upper bound in one direction with the other extreme obtained if we switch the signs in the numerator and denominator for δ (case 2). Note, although we refrain from using a separate notation, if $p_{kl} - \delta < 0$, then $p_{kl} - \delta = 0$. Similarly, if $p_{kl} + \delta > 1$ then $p_{kl} + \delta = 1$. Next,

$$\begin{aligned}(p_{1z_1}-\delta)\dots(p_{nz_{n-1}}-\delta) \;&=\; p_{ij}\Pi_i(z) - [\delta(p_{2z_2}p_{3z_3}\dots p_{nz_{n-1}} + p_{1z_1}p_{3z_3}\dots p_{nz_{n-1}} + \dots \\ &\quad + p_{1z_1}\dots p_{n-1z_{n-2}}) - \delta^2(p_{3z_3}p_{4z_4}\dots p_{nz_{n-1}} + \dots + p_{1z_1}\dots p_{n-2z_{n-3}}) + \dots + (-\delta)^n] \\ &\geq p_{ij}\Pi_i(z) - [\delta(p_{2z_2}p_{3z_3}\dots p_{nz_{n-1}} + p_{1z_1}p_{3z_3}\dots p_{nz_{n-1}} + \dots + p_{1z_1}\dots p_{n-1z_{n-2}}) \\ &\quad + \delta^2(p_{3z_3}p_{4z_4}\dots p_{nz_{n-1}} + \dots + p_{1z_1}\dots p_{n-2z_{n-3}}) + \dots + \delta^n] \;=\; p_{ij}\Pi_i(z) - D_{\delta,p}\end{aligned}$$

where $D_{\delta,p}$ denotes the sum in the square brackets. After combining and canceling terms, the upper bound can be written as

$$\begin{aligned}|p'_{ij} - q'_{ij}| \;&\leq\; \frac{\sum_z f_{ij}(z)p_{ij}\Pi_i(z)\ \sum_k\sum_z f_{ik}(z)D_{\delta,p} \;+\; \sum_z f_{ij}(z)D_{\delta,p}\ \sum_k\sum_z f_{ik}(z)p_{ik}\Pi_i(z)}{\sum_k\sum_z f_{ik}(z)p_{ik}\Pi_i(z)\ \sum_k\sum_z f_{ik}(z)(p_{1z_1}+\delta)\dots(p_{nz_{n-1}}+\delta)} \\ &<\; \frac{f_{max}\ (\sum_k\sum_z f_{ik}(z)D_{\delta,p} \;+\; \sum_z f_{ij}(z)D_{\delta,p})}{f^2_{min}} \;\leq\; \frac{2f^2_{max}\binom{n}{n/2}(\delta+\delta^2+\dots+\delta^n)}{f^2_{min}} \\ &<\; \delta\,\frac{2f^2_{max}\binom{n}{n/2}n}{f^2_{min}} \;=\; ||p-q||\,K_1(n)\end{aligned}$$

where $f_{max} = \max_{x\in X^n} f(x)$, $f_{min} = \min_{x\in X^n} f(x)$, $f_{max} \geq \sum_k\sum_z f_{ik}(z)p_{ik}\Pi_i(z)$ (the opposite holds for f_{min}). In the second inequality, the summation was taken inside $D_{\delta,p}$, and note that the factor of $\delta^{n/2}$ is largest and can be bounded by $\binom{n}{n/2}f_{max}$. Since $K_1(n)$ is independent of p, q, this yields the Lipschitz constant for the first case. *Case 2*:

$$
\begin{aligned}
|p'_{ij} - q'_{ij}| &\leq \frac{\sum_{z\in X^{n-1}} f_{ij}(z)(p_{1z_1}+\delta)\dots(p_{nz_{n-1}}+\delta)}{\sum_{k=1}^{m}\sum_{z\in X^{n-1}} f_{ik}(z)(p_{1z_1}-\delta)\dots(p_{nz_{n-1}}-\delta)} - \frac{\sum_{z\in X^{n-1}} f_{ij}(z)p_{ij}\Pi_i(z)}{\sum_{k=1}^{m}\sum_{z\in X^{n-1}} f_{ik}(z)p_{ik}\Pi_i(z)} \\
&= \frac{\sum_z f_{ij}(z)D_{\delta,p}\ \sum_k\sum_z f_{ik}(z)p_{ik}\Pi_i(z) \ + \ \sum_z f_{ij}(z)p_{ij}\Pi_i(z)\ \sum_k\sum_z f_{ik}(z)D_{\delta,p}}{\sum_k\sum_z f_{ik}(z)p_{ik}\Pi_i(z)\ \sum_k\sum_z f_{ik}(z)(p_{1z_1}-\delta)\dots(p_{nz_{n-1}}-\delta)} \\
&< \frac{\delta\, 2 f_{max}^2 \binom{n}{n/2} n}{f_{min}^2 \sum_k\sum_z (p_{1z_1}-\delta)\dots(p_{nz_{n-1}}-\delta)} \ \leq \ \delta\, \frac{2 f_{max}^2 \binom{n}{n/2} n}{f_{min}^2 (\frac{1}{m}-\delta)^n} \\
&< \delta\, \frac{2 f_{max}^2 \binom{n}{n/2} n}{f_{min}^2 (\frac{1}{2m})^n} \ = \ ||p-q||\, K_2(n,m).
\end{aligned}
$$

Since p is a probability measure, $\forall i$, $1 \leq i \leq n$, $\exists\, j$, $1 \leq j \leq m$ such that $p_{ij} \geq 1/m$. This establishes $\sum_k\sum_z (p_{1z_1}-\delta)\dots(p_{nz_{n-1}}-\delta) \geq (\frac{1}{m}-\delta)^n$. Using our assumption, $\delta \leq 1/2m$, yields the last inequality. Note, both $K_1(n)$ and $K_2(n,m)$ are exponential functions of n. Setting $K = K_2(n,m) > K_1(n)$ completes the proof. □

4 Analysis of convergence

In this section, we will show that m-GA finds the solution, and does so with an exponential (or geometric) convergence rate. For all $i \in \{1,\dots,n\}$, let u_i be the element such that $\forall a \in X$ and $a \neq u_i$, $a \prec_i u_i$. Let v_i be the element such that $\forall a \in X$ and $a \neq u_i$, $v_i \prec_i a$. Let $E_{ij} = \sum_{z\in X^{n-1}} f_{ij}(z)\Pi_i(z)$ and let $D_i = \sum_{k=1}^{m} p_{ik}(t) \sum_{z\in X^{n-1}} f_{ik}(z)\Pi_i(z)$.

Lemma 1 *Let f be monotonic and let $p(0)$ be fully supported. Then $\forall i \in \{1,\dots,n\}$, $p_{iv_i}(0) > p_{iv_i}(1) > p_{iv_i}(2) > \dots$, and $p_{iu_i}(0) < p_{iu_i}(1) < p_{iu_i}(2) < \dots$. Moreover, $p_{iv_i}(t) \to 0$, and $p_{iu_i}(t) \to 1$, as $t \to \infty$.*

Proof

$$
\begin{aligned}
D_i - E_{iv_i} &= \sum_{k=1}^{m} p_{ik}(t) \sum_{z\in X^{n-1}} f_{ik}(z)\Pi_i(z) \ - \sum_{z\in X^{n-1}} f_{iv_i}(z)\Pi_i(z) \\
&> \sum_{k=1}^{m} p_{ik}(t) \sum_{z\in X^{n-1}} f_{iv_i}(z)\Pi_i(z) \ - \sum_{z\in X^{n-1}} f_{iv_i}(z)\Pi_i(z) \ = \ 0.
\end{aligned}
$$

Hence $p_{iv_i}(t)$ is monotonically decreasing. A symmetric argument, where we substitute u_i for k in D_i, shows $E_{iu_i} > D_i$. Let $p^*_{iv_i} = \lim_{t\to\infty} p_{iv_i}(t)$. By compactness of $[0,1]$, $p^*_{iv_i} \in [0,1]$. Since h is continuous and 0 is the only left stationary point, $p^*_{iv_i} = 0$. A similar argument shows that $\lim_{t\to\infty} p_{iu_i}(t) = 1$. □

Note, lemma 1 implies that all other probabilities except p_{iu_i} converge to 0. The next lemma shows that they do so monotonically after an initial transient period.

Lemma 2 *Let f be monotonic and let $p(0)$ be fully supported. Then $\forall a$, $v_i \prec_i a \prec_i u_i$, $\exists t_a > 0$ such that $\forall t > t_a$, $p_{ia}(t+1) < p_{ia}(t)$.*

Proof Let $v_i \prec_i a_2 \prec_i a_3 \prec_i \ldots \prec_i a_{m-1} \prec_i u_i$ denote the ordering induced by $\prec_i$. First, consider p_{ia_2}. We need to show that eventually $D_i - E_{ia_2} > 0$, and remains so thereafter. In the following, note $p_{iv_i} + p_{ia_2} + \ldots + p_{ia_{n-1}} + p_{iu_i} = 1$.

$$
\begin{aligned}
D_i - E_{ia_2} &= \sum_{a_2 \prec_i k} p_{ik}(t) \sum_{z \in X^{n-1}} (f_{ik}(z) - f_{ia_2}(z))\Pi_i(z) \; - \; p_{iv_i}(t) \sum_{z \in X^{n-1}} (f_{ia_2}(z) - f_{iv_i}(z))\Pi_i(z) \\
&> \sum_{a_2 \prec_i k} p_{ik}(t) \sum_{z \in X^{n-1}} (f_{ik}(z_{min}) - f_{ia_2}(z_{min}))\Pi_i(z) - p_{iv_i}(t) \sum_{z \in X^{n-1}} (f_{ia_2}(z_{max}) - f_{iv_i}(z_{max}))\Pi_i(z) \\
&> (\sum_{z \in X^{n-1}} \Pi_i(z)) \, [\sum_{a_2 \prec_i k} p_{ik}(t)(f_{ik}(z_{min}) - f_{ia_2}(z_{min})) \; - \; p_{iv_i}(t)(f_{ia_2}(z_{max}) - f_{iv_i}(z_{max}))]
\end{aligned}
$$

where z_{min} is the vector such that $f_{ik}(z_{min}) - f_{ia_2}(z_{min}) = \min_{z \in X^{n-1}} f_{ik}(z) - f_{ia_2}(z)$, and z_{max} satisfies $f_{ia_2}(z_{max}) - f_{iv_i}(z_{max}) = \max_{z \in X^{n-1}} f_{ia_2}(z) - f_{iv_i}(z)$. Since $f_{ik}(z_{min}) - f_{ia_2}(z_{min})$, $f_{ik}(z_{max}) - f_{ia_2}(z_{max}) > 0$ are fixed quantities, and by lemma 1 $p_{iv_i} \to 0$, $p_{iu_i} \to 1$, for some $t_{a_2} > 0$, $D_i - E_{ia_2} > 0$. Moreover, by monotonicity of p_{iv_i} and p_{iu_i}, $\forall t > t_{a_2}$, $p_{ia_2}(t+1) < p_{ia_2}(t)$. A simple induction on a_j, $j = 3, 4, \ldots, m-1$, completes the proof. □

This leads to the theorem stating that for monotonic f, there exists a unique invariant measure with respect to m-GA.

Theorem 1 *Let f be monotonic and let $p(0)$ be fully supported. Then $p(t) \to p^*$, $t \to \infty$.*

Proof By lemma 1, lemma 2, and the characterization of the solution distribution p^* for monotonic functions as $p^*_{ij} = 1$ if $j = u_i$, and $p^*_{ij} = 0$ otherwise, $i = 1, 2, \ldots, n$, the theorem follows directly. □

The next task is to estimate the rate of convergence. This is easy to do for v_i, the minimal element. A little consideration is needed for the other cases.

Lemma 3 *Let f be monotonic and $p(0) > 0$. Then $\forall t > 0$, $p_{iv_i}(t+1) < p_{iv_i}(t)(1-c)$, where $0 < c < 1$.*

Proof By lemma 1, $D_i > E_{iv_i}$. Denote $B_{iv_i} = D_i - E_{iv_i}$. Since $E_{iv_i}/D_i = 1 - B_{iv_i}/D_i$, to bound the rate, it suffices to find a lower bound for B_{iv_i}, $B^{lo}_{iv_i}$, and an upper bound for D_i,

D_i^{up}. First,

$$B_{iv_i} = \sum_{v_i \prec_i k} p_{ik}(t) \sum_{z \in X^{n-1}} (f_{ik}(z) - f_{iv_i}(z))\Pi_i(z) > (\sum_{z \in X^{n-1}} \Pi_i(z)) \sum_{v_i \prec_i k} p_{ik}(t)(f_{ik}(z_{min}) - f_{iv_i}(z_{min}))$$
$$> (\sum_{z \in X^{n-1}} \Pi_i(z))\, p_{iu_i}(t)\, (f_{iu_i}(z_{min}) - f_{iv_i}(z_{min})) = B_{iv_i}^{lo}$$

where z_{min} has the same interpretation as in lemma 2. Next,

$$D_i = \sum_{k=1}^{m} p_{ik}(t) \sum_{z \in X^{n-1}} f_{ik}(z)\Pi_i(z) < \sum_{k=1}^{m} p_{ik}(t) \sum_{z \in X^{n-1}} f_{ik}(z_{max})\Pi_i(z) < (\sum_{z \in X^{n-1}} \Pi_i(z)) \sum_{k=1}^{m} f_{ik}(z_{max}) = D_i^{up}$$

Thus, $$\frac{E_{iv_i}}{D_i} = 1 - \frac{B_{iv_i}}{D_i} < 1 - \frac{B_{iv_i}^{lo}}{D_i^{up}} = 1 - p_{iu_i}(t) \frac{f_{iu_i}(z_{min}) - f_{iv_i}(z_{min})}{\sum_{k=1}^{m} f_{ik}(z_{max})}$$
$$< 1 - p_{iu_i}(0) \frac{f_{iu_i}(z_{min}) - f_{iv_i}(z_{min})}{\sum_{k=1}^{m} f_{ik}(z_{max})} = 1 - c$$

where the last inequality follows from the monotonicity of $p_{iu_i}(t)$. c is a fixed, time-independent quantity with $0 < c < 1$. □

The next proposition shows that if $0 < p_{iu_i} < 0.5$, convergence to 0.5 is exponential in $1 + (1 - p_{iu_i}(0))c$.

Proposition 2 *Let f be monotonic and $p(0) > 0$. Then $p_{iu_i}(t+1) > p_{iu_i}(t)(1+(1-p_{iu_i}(t))c)$ where $c > 0$.*

Proof Since $E_{iu_i} > D_i$ and $E_{iu_i}/D_i = 1 + (E_{iu_i} - D_i)/D_i$, we need to bound $E_{iu_i} - D_i$ from below and D_i from above.

$$E_{iu_i} - D_i = \sum_{k \prec_i u_i} p_{ik}(t) \sum_{z \in X^{n-1}} (f_{iu_i}(z) - f_{ik}(z))\Pi_i(z) > (\sum_{z \in X^{n-1}} \Pi_i(z)) \sum_{k \prec_i u_i} p_{ik}(t)(f_{iu_i}(z_{min}) - f_{ik}(z_{min}))$$
$$> (\sum_{z \in X^{n-1}} \Pi_i(z))\, (f_{iu_i}(z_{min}) - f_{ib}(z_{min})) \sum_{k \prec_i u_i} p_{ik}(t)$$
$$= (\sum_{z \in X^{n-1}} \Pi_i(z))\, (f_{iu_i}(z_{min}) - f_{ib}(z_{min}))\, (1 - p_{iu_i}(t))$$

where b was chosen so that $f_{iu_i}(z_{min}) - f_{ib}(z_{min}) = \min_{k \prec_i u_i} f_{iu_i}(z_{min}) - f_{ik}(z_{min})$. Thus the constant in the proposition is given by $c = (f_{iu_i}(z_{min}) - f_{ib}(z_{min}))/ \sum_{k=1}^{m} f_{ik}(z_{max})$. □

The next lemma shows the exponential convergence of all other nonoptimal probabilities which allows us to uniformly bound the optimal probability.

Lemma 4 *Let f be monotonic and $p(0) > 0$. Then $\forall a$, $v_i \prec_i a \prec_i u_i$, $p_{ia}(t+1) < p_{ia}(t)(1 - c)$, for $t > (m-2)K$ where c and K are constants depending only on i.*

Proof Let $v_i \prec_i a_2 \prec_i a_3 \prec_i \ldots \prec_i a_{m-1} \prec_i u_i$ denote the ordering induced by $\prec_i$. First, consider p_{ia_2}. Note from lemma 2 that $D_i - E_{ia_2}$ is bounded by

$$D_i - E_{ia_2} > (\sum_{z \in X^{n-1}} \Pi_i(z)) [\sum_{a_2 \prec_i k} p_{ik}(t)(f_{ik}(z_{min}) - f_{ia_2}(z_{min})) - p_{iv_i}(t)(f_{ia_2}(z_{max}) - f_{iv_i}(z_{max}))].$$

Assume $D_i - E_{ia_2} < 0$. Then for one or more k, $p_{ik}(t)(f_{ik}(z_{min}) - f_{ia_2}(z_{min})) - p_{iv_i}(f_{ia_2}(z_{max}) - f_{iv_i}(z_{max})) < 0$. Consider

$$\begin{aligned} L_{ia_2} &= p_{iu_i}(0)(f_{iu_i}(z_{min}) - f_{ia_2}(z_{min})) - p_{iv_i}(t)(f_{ia_2}(z_{max}) - f_{iv_i}(z_{max})) \\ &< p_{iu_i}(t)(f_{iu_i}(z_{min}) - f_{ia_2}(z_{min})) - p_{iv_i}(t)(f_{ia_2}(z_{max}) - f_{iv_i}(z_{max})) \\ &< \sum_{a_2 \prec_i k} p_{ik}(t)(f_{ik}(z_{min}) - f_{ia_2}(z_{min})) - p_{iv_i}(t)(f_{ia_2}(z_{max}) - f_{iv_i}(z_{max})). \end{aligned}$$

Since $p_{iv_i}(t)$ decreases exponentially with a rate at least $1 - c_v$ as given in lemma 3, $L_{ia_2} > 0$ if $t > \frac{1}{c_v} \log \frac{1}{A_{ia_2}}$, where $A_{ia_2} = p_{iu_i}(0)(f_{iu_i}(z_{min}) - f_{ia_2}(z_{min}))/(f_{ia_2}(z_{max}) - f_{iv_i}(z_{max}))$. That is, after a constant number of steps depending only on the initial condition, p_{ia_2} is assured to decrease monotonically at an exponentially rate. It is easily seen that the rate of p_{ia_2}, $1 - c_2$, can be bounded above by $1 - c_2 < 1 - L_{ia_2}/(2D_i)$ if we choose $t > \frac{1}{c_v} \log \frac{2}{A_{ia_2}}$. Since $\sum_{k=1}^{m} \sum_{z \in X^{n-1}} f_{ik}(z)$ is an upper bound of D_i, the convergence rate of p_{ia_2} is bounded by a fixed quantity independent of time. In general, for a_j,

$$\begin{aligned} L_{ia_j} &= p_{iu_i}(0)(f_{iu_i}(z_{min}) - f_{ia_j}(z_{min})) - (f_{ia_j}(z_{max}) - f_{ib}(z_{max})) \sum_{k \prec_i a_j} p_{ik}(t) \\ &< p_{iu_i}(0)(f_{iu_i}(z_{min}) - f_{ia_j}(z_{min})) - \sum_{k \prec_i a_j} p_{ik}(t)(f_{ia_j}(z_{max}) - f_{ik}(z_{max})) \\ &< \sum_{a_j \prec_i k} p_{ik}(t)(f_{ik}(z_{min}) - f_{ia_j}(z_{min})) - \sum_{k \prec_i a_j} p_{ik}(t)(f_{ia_j}(z_{max}) - f_{ik}(z_{max})) \end{aligned}$$

where b was chosen such that $f_{ia_j}(z_{max}) - f_{ib}(z_{max}) = max_{k \prec_i a_j} f_{ia_j}(z_{max}) - f_{ik}(z_{max})$. Let $1 - c_2, \ldots, 1 - c_{j-1}$ be the convergence rates of $p_{ia_2}, \ldots, p_{ia_{j-1}}$, respectively. Let $K_2, \ldots, K_{j-1}$ be the constant lower bounds consumed at previous steps to guarantee monotonic convergence. Using the same arguments as before, for $t > \frac{1}{c_b} \log \frac{1}{A_{ia_j}}$, $L_{ia_j} > 0$, where $c_b = \min\{c_v, c_2, \ldots, c_{j-1}\}$ and $A_{ia_j} = p_{iu_i}(0)(f_{iu_i}(z_{min}) - f_{ia_j}(z_{min}))/(f_{ia_j}(z_{max}) - f_{ib}(z_{max}))$. Completing the induction, we see that after at most $(m-2)K$ steps where $K = \max\{K_2, \ldots, K_{m-1}\}$, all probabilities decrease monotonically with a rate bounded above by $1 - c = 1 - \min\{c_v, c_2, \ldots, c_{m-1}\}$. □

Finally, we can state the main theorem.

Theorem 2 *Let f be monotonic and let $p(0)$ be fully supported. Then* $||p^* - p(t+1)|| < (1-c)||p^* - p(t)||$, $\forall t > (m-2)K$, *where* $0 < c < 1$ *and* $K > 0$.

Proof By lemma 4, for each $i = 1, \ldots, n$, after at most $(m-2)K(i)$ steps, where $K(i)$ depends on i, the probabilities converge monotonically with a rate at least $c(i)$ as described in the lemma. Take $K = \max\{K(1), \ldots, K(n)\}$ and $c = \min\{c(1), \ldots, c(n)\}$. Noting $p_{iu_i} = 1 - \sum_{j \neq u_i} p_{ij}$ and by the chracterization of p^*, the sup-norm inequality follows immediately. □

5 Conclusion

We have presented an analysis of the limit behavior of m-GA for efficiently solving the optimization problem for monotonic functions. Several problems still remain. First, although showing that the ideal behavior of a GA that converges to the optimal probability distribution with an exponential rate is indicative that the actual, sample-bounded GA may also fair well, such intuitive reasoning is far from sufficient. A rigorous sampling theory is needed to fill the gap. Second, tighter bounds (both lower and upper) on the power of m-GA are interesting to pursue. In particular, it would be fruitful to show that a standard GA is at least as powerful as the m-GA (in some suitable sense) which seems reasonable. In the same venue, a systematic, and quantitative characterization of the effect of k-point cross-over ($1 \leq k \leq n$) and other variational features should be manageable and illuminating. Third, analyzing the standard GA using dynamic Markov chain techniques looms as an interesting challenge. We hope this paper is a step in the right direction.

References

[1] S. Arora and S. Safra. Probabilistic checking of proofs: a new characterization of np. In *Proc. 33rd IEEE Symp. on Foundations of Computer Science*, 1992.

[2] T. Davis and J. Principe. A simulated annealing like convergence theory for the simple genetic algorithm. In *Proc. 4th International Conf. on Genetic Algorithms*, pages 174–181, 1991.

[3] D. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, 1989.

[4] J. Holland. *Adaptation in Natural and Artificial Systems*. MIT Press, 1992.

[5] K. Park. Probabilistic analysis of genetic algorithms. *Manuscript in preparation*, 1993.

[6] Y. Rabinovich, A. Sinclair, and A. Wigderson. Quadratic dynamical systems. In *Proc. 33rd IEEE Symp. on Foundations of Computer Science*, 1992.

[7] Y. Rabinovich and A. Wigderson. Analysis of a simple genetic algorithm. In *Proc. 4th International Conf. on Genetic Algorithms*, pages 215–221, 1991.