2003

Bistability of the needle function in the presence of truncation selection

Greg Cripe
The University of Montana

The University of

Montana

Permission is granted by the author to reproduce this material in its entirety, provided that this material is used for scholarly purposes and is properly cited in published works and reports.

Please check "Yes" or "No" and provide signature

Yes, I grant permission _____

No, I do not grant permission _____

Author's Signature: _____

Date: _____ 7/27/03 _____

Any copying for commercial purposes or financial gain may be undertaken only with the author's explicit consent.

8/98

Bistability of the Needle Function in the Presence of Truncation
Selection

by

Greg Cripe

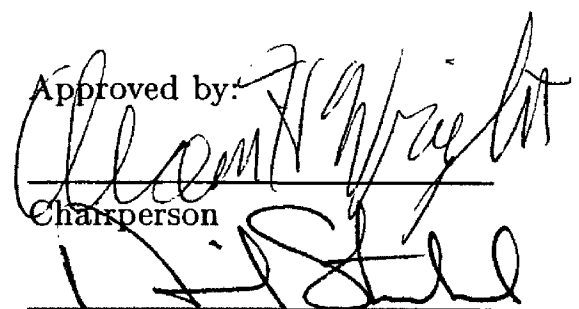presented in partial fulfillment of the requirements

for the degree of

Master of Science in Computer Science

The University of Montana

July 2003

Approved by:

_____
Chairperson

_____
Dean, Graduate School

_____7-31-03_____
Date

UMI Number: EP38986

UMI

Dissertation Publishing

UMI EP38986

ProQuest

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

Cripe, Greg M.S., July 2003   Computer Science

Bistability of the Needle Function in the Presence of Truncation Selection

Director: Alden Wright

It is possible for a genetic algorithm (GA) to have two stable fixed points on a single-peak fitness landscape. This paper studies a gene pool GA model with mutation and the truncation selection operator. For the needle function, the fixed point equation reduces to a single variable and the stability of the fixed points can be explicitly determined. We show that bistability can occur for certain parameter values of the model. When bistability does occur, it may happen that the GA will spend long periods of time near the non-optimal fixed point.

Contents

Introduction

This paper uses the gene pool GA model to investigate bistability. Bistability occurs when the GA has two stable fixed points even though the fitness function has a single peak. It is only known to occur when the GA includes recombination, mutation and selection. When there is bistability, depending on the initial distribution of the population, the infinite population GA model may converge to a distribution which has a very low probability for the optimum string.

Bistability was first discovered by Boerlijst, Bonhoeffer, and Nowak [1] in the context of viral quasi-species and the AIDS virus. Other papers on bistability include Ochoa and Harvey [4], Wright, Rowe, Stephens,and Poli [10], and Wright [9]. The last citation has a more complete review of the literature in the area.

For the needle fitness function, Wright, Rowe, Stephens and Poli [10] show how the gene pool GA model can be reduced to a polynomial equation in a single variable. The authors show that bistability can occur for certain parameter values.

This paper closely follows Wright et al [10]. We extend their work by showing that bistability also exists when truncation selection is used instead of proportional selection. The proof of existence of bistability is substantially different in the case of truncation selection. The equation for bistability is derived and is analyzed to find the fixed points and determine their stabilities. The regions of the parameter space in which bistability can occur are described. Results for the gene pool recombination are compared with a finite simulation of the two-parent recombination GA. As in

Wright et al, we show the gene pool GA model to be a reasonable approximation for the corresponding two parent GA.

Genetic Algorithms

A genetic algorithm (GA) searches for a solution by an approach that is loosely based on simulated evolution. GAs search a space of hypotheses to identify the best hypothesis. The best hypothesis is defined to be the one that optimizes a numerical measure of the problem. This measure is referred to as the fitness.

In general, the search space is too large to perform an exhaustive search. The GA guides the search process by starting with a collection of hypotheses, called a population. The GA then iteratively generates successive populations until some stopping criteria is met. Most GAs use a combination of three operators during this iterative process.

Selection : On each iteration all of the current members of the population, called individuals, are evaluated using the fitness function. Based on their fitness, a certain number of individuals are selected as the basis for creating new individuals in the population. The selection may be probabilistic or deterministic. Common selection schemes include proportional, tournament, truncation, and ranking. Part or all of the current generation will be replaced with the newly created individuals. There are many choices for replacement schemes.

Mutation : With some low probability, the genotype of an individual is slightly

altered. For example, if the representation of the hypothesis is a binary string, the genotype is mutated by flipping some randomly chosen bits.

Crossover : Typically, the crossover operator produces two offspring, called children, from two selected parents by copying portions of the genotype from each parent. This process is often referred to as two parent recombination (TPR).

Common crossover schemes for binary strings, in order from the least mixing of bits to the highest recombination of bits, include one-point crossover, two-point crossover and uniform crossover.

Gene pool recombination (GPR) is an alternative to TPR used in genetic algorithms. An individual created by gene pool recombination is chosen from the distribution determined by the whole population rather than from two parents. In other words, gene pool recombination produces an even greater recombination of bits than uniform crossover since the "parents" are the entire population instead of only two individuals. Gene pool recombination was introduced to the evolutionary computation community by Syswerda [5].

In order to simplify the computations performed in this paper, we use an infinite population model to describe the random processes. An infinite population model characterizes the GA as the population size goes to infinity. Instead of an actual distribution as in the finite model, we now consider the limiting behavior of the distribution as the population size goes to infinity.

The infinite population GPR model can serve as an approximation for the finite

population GA. If the current population is p, we choose the next population by sampling from the probability distribution resulting from applying one iteration of the infinite gene pool recombination model GA to p. If the population is of size N, we simply make N independent samples from this distribution.

For further reference, the reader should consult Vose [6]. Vose discusses his exact Markov chain model and describes how the infinite population model relates to the finite population model in an exact model.

Notation

The search space, Ω, is the set of all binary length strings of length ℓ. A population (a multi-set of Ω) is represented as a population vector of length 2^ℓ indexed by Ω. For example, if $\ell = 2$ and the population as a multi-set is $\{00,01,01,11\}$, then the corresponding population vector is $[\frac{1}{4}, \frac{1}{2}, 0, \frac{1}{4}]$. Each component of the population vector is a relative frequency and is therefore a natural representation for infinite-population models. All population vectors are contained in the simplex $\Lambda = \{x : \sum_{j \in \Omega} x_j = 1 \text{ and } x_j \geq 0 \text{ for all } j\}$.

A schema is a type of subset of Ω denoted by a string from the alphabet $\{0,1,*\}$. The symbol * represents a don't care symbol. For example, the schema denoted by 1*0 represents the subset $\{ 100, 110 \}$. The frequency of the schema 1*0 is the sum of the frequencies of the strings $\{ 100, 110 \}$.

The number of ones in a binary string j is denoted by $\#j$. We define $\mathcal{L}$ to be the

set of strings j such that $\#j = 1$. For a given string k, we define $\mathcal{L}_k$ to be the subset of $\mathcal{L}$ such that for $j \in \mathcal{L}_k$ the bitwise AND of j and k, denoted $j \otimes k$, is not the all zero string.

For $k \in \ell$, $x_0^{(k)}$ is the frequency of the schema which has a zero in the bit associated with k. Likewise $x_1^{(k)}$ is the frequency of the schema with a one in the bit associated with k. For example with $\ell = 3$, $x_1^{010} = x_{010} + x_{011} + x_{110} + x_{111}$ and $x_0^{010} = x_{000} + x_{001} + x_{100} + x_{101}$.

Linkage Equilibrium and the Walsh basis

The Walsh matrix W is a 2^ℓ by 2^ℓ matrix defined by $W_{i,j} = (-1)^{\#(i \otimes j)}$. The Walsh matrix is symmetric and $W^{-1} = 2^{-\ell}W$. If x is a population vector, then the Walsh transform of x is Wx and is denoted $\hat{x}$. Note that this definition is slightly different than the Walsh transform given by Vose [6]. If A is a 2^ℓ by 2^ℓ matrix, then WAW is the Walsh transform of A and is denoted by $\widehat{A}$.

A population is said to be in linkage equilibrium if the frequency of each string is the product of the marginal distributions (1-schema averages) corresponding to each allele of each locus. To calculate the frequency of any string for a population at linkage equilibrium, simply take the product of the frequencies of the allele at each bit. For example, if $\ell = 3$, then x_{101}, the frequency of the string 101, is the product of the frequencies of the schemas 1**, *0*, **1.

For the infinite population model, GPR transforms any population into a linkage

equilibrium population. Geiringer's [2] theorem relates GPR to linkage equilibrium by showing that linkage equilibrium, in the infinite population model, is the limit of repeated applications of two-parent recombination.

Using a linkage equilibrium assumption simplifies the model by allowing us to track only the ℓ allele frequencies instead of all the 2^ℓ strings.

The following four lemmas are found in Wright, Rowe, Poli and Stephens [10].

Lemma 1. *For any population vector x and any $k \in \mathcal{L}$,*

$$x_0^{(k)} = \frac{1}{2}(1 + \hat{x}_k) \ and \ x_1^{(k)} = \frac{1}{2}(1 - \hat{x}_k).$$

Proof. The lemma can be proved by comparing the definition of $x_j^{(k)}$ with $\hat{x}_k$. $\square$

Lemma 2. *If population x is in linkage equilibrium, then for any $k \in \Omega$*

$$\hat{x}_k = \prod_{j \in \mathcal{L}_k} \hat{x}_j.$$

A similar result is proved as Theorem 10.9 of Vose [6] and Theorem 3.5 of Vose and Wright [7].

Lemma 3. *If population x is in linkage equilibrium, then*

$$x_0 = 2^{-\ell} \prod_{j \in \mathcal{L}} (1 + \hat{x}_j).$$

Proof. This simply rewrites Lemma 1 using the definition of linkage equilibrium. $\square$

Lemma 4. *If $x \in \Lambda$, then $\mid \hat{x}_k \mid \leq 1$ for all $k \in \Omega$.*

Proof. The simplex Λ is the convex hull of the standard basis vectors of $\mathbb{R}^N$. The Walsh transform of these vectors form the columns of the Walsh matrix W whose entries are ± 1. Therefore $\hat{x}_k$ is a convex combination of 1 and -1. $\square$

The Gene Pool Model

We describe the infinite population gene pool GA. The string length is assumed to be ℓ.

1. Choose a random population x.

2. Calculate the order-1 schemas $x_0^{(k)}$ for each $k \in \mathcal{L}$

3. Construct a new linkage equilibrium population with the same order-1 schema averages $x_0^{(k)}$.

4. Apply selection to this population.

5. Apply mutation to the resulting population.

6. Return to step 2 if termination criteria is not met.

The reader should note that step 3 represents gene pool recombination.

Gene pool recombination and Mutation

Define the function $\mathcal{M} : \Lambda \to \Lambda$ by

$$\widehat{\mathcal{M}(x)}_k = \begin{cases} \hat{x}_k & \text{if } k \in \mathcal{L} \cup \{0\} \\ \\ \prod_{j \in \mathcal{L}_k} \hat{x}_j & \text{otherwise.} \end{cases}$$

Lemma 0.2 shows that $\mathcal{M}$ models GPR and therefore $\widehat{\mathcal{M}(x)}$ is in linkage equilibrium.

Let $\mathcal{U}$ represent the mutation operator. The following lemma, taken from Wright,

Rowe, Poli and Stephens [10], shows how to compute mutation in the Walsh basis.

Lemma 0.5. $\widehat{\mathcal{U}(x)}_k = (1 - 2\mu)\hat{x}_k$ *for* $k \in \mathcal{L}$.

Truncation Selection Method

In this paper we will use the truncation selection method. Truncation selection is a

rank-based elitist method. If the population size is of finite size N and the truncation

fraction is $0 < T < 1$, we define a selection heuristic on the current population p as

follows. Without loss of generality, we assume that the search space is ordered so that

$f_0 < f_1 < \ldots < f_{2^\ell - 1}$, where f_j denotes the fitness of individual j. The truncation

selection heuristic $\mathcal{F}$ is defined in Wright and Agapie [8] by

$$\mathcal{F}(p)_k = \begin{cases} 0 & \text{if } T < \sum_{k<j} p_j \\ \\ \frac{T - \sum_{k<j} p_j}{T} & \text{if } \sum_{k<j} p_j \leq T < \sum_{k\leq j} p_j \\ \\ \frac{p_k}{T} & \text{if } \sum_{k\leq j} p_j \leq T. \end{cases}$$

If the population is finite of size N, truncation selection is performed by taking N independent samples from the above distribution.

In the case that two or more distinct individuals have the same fitness, we consider them to be members of one fitness class. The above selection heuristic is modified as followed. After the distribution is calculated the fitness classes with more than one member are split into separate classes by dividing their share of the distribution proportionately.

Needle Function

The NEEDLE (needle in the haystack) fitness function assigns a fitness greater than one to the string of all zeros and assigns a fitness of one to all other strings.

Let $z = z(t)$ denote the current population, and let $x = \mathcal{M}(z)$. Then $x_0 = 2^{-\ell} \prod_{j \in \mathcal{L}} (1 + \hat{z}_j)$ is the probability of the all-zeros string in x. Let $y = \mathcal{F}(x)$ denote the population produced after truncation selection is applied to x. The all-zeros string, y_0, will then appear with frequency $\min(1, \frac{x_0}{T})$, where T is the truncation fraction selected from x. Any non-zero string, y_j, with $j > 0$ will appear with frequency $y_j = x_j \left(\frac{1-y_0}{1-x_0} \right) = \left(\frac{x_j}{1-x_0} \right) (1 - y_0)$. For $k \in \mathcal{L}$, we see $y_1^{(k)} = x_1^{(k)} \left(\frac{1-y_0}{1-x_0} \right)$.

We now define G, the operator that represents steps 3,4 and 5 in our model.

$$G(z)_k = \mathcal{U} \circ \mathcal{F} \circ \mathcal{M}(z)_k$$

Then

$$\widehat{G(z)}_k = (1 - 2\mu)\hat{y}_k \qquad\qquad \text{by Lemma 0.5}$$

$$= (1 - 2\mu)(1 - 2y_1^{(k)}) \qquad\qquad \text{by Lemma 0.1}$$

$$= (1 - 2\mu)\left(1 - 2x_1^{(k)}\left(\frac{1 - y_0}{1 - x_0}\right)\right)$$

$$= (1 - 2\mu)\left(1 - (1 - \hat{z}_k)\left(\frac{1 - y_0}{1 - x_0}\right)\right).$$

If $y_0 = 1$ then $\widehat{G(z)}_k = 1 - 2\mu$. This will be a fixed point if $x_0 > T$, that is ,

$$x_0 = 2^{-\ell}\prod_{j\in\mathcal{L}}(1 + \hat{z}_j)$$

$$= 2^{-\ell}\prod_{j\in\mathcal{L}}(1 + 1 - 2\mu)$$

$$= (1 - \mu)^\ell$$

$$> T.$$

If $y_0 \neq 1$ then

$$\widehat{G(z)}_k = (1 - 2\mu)\left(1 - (1 - \hat{z}_k)\left(\frac{1 - \frac{x_0}{T}}{(1 - x_0)}\right)\right)$$

$$= (1 - 2\mu)\left(1 - (1 - \hat{z}_k)\left(\frac{T - \prod_{j\in\mathcal{L}}(1 + \hat{z}_j)/2}{T(1 - \prod_{j\in\mathcal{L}}(1 + \hat{z}_j)/2)}\right)\right).$$

In this case, the recurrence equation leads to the fixed point equations:

$$\hat{z}_k = \frac{(1 - 2\mu)\left(1 - \left(\frac{T - \prod_{j\in\mathcal{L}}(1+\hat{z}_j)/2}{T(1-\prod_{j\in\mathcal{L}}(1+\hat{z}_j)/2)}\right)\right)}{1 - (1 - 2\mu)\left(\frac{T - \prod_{j\in\mathcal{L}}(1+\hat{z}_j)/2}{T(1-\prod_{j\in\mathcal{L}}(1+\hat{z}_j)/2)}\right)}$$

The right hand side of this equation is the same for all k. Therefore, if $\hat{z}$ is a fixed point of G, then $\hat{z}$ is symmetric in the sense that all $\hat{z}_k$ for $k \in \mathcal{L}$ are equal.

If the GA model is started with a symmetric population, symmetry will be maintained. In this case, the recurrence equation can be written in terms of the variable w,

$$\widehat{G(w)} = \begin{cases} (1 - 2\mu) & \text{if } T < 2^{-\ell}(1 + w)^{\ell} \ ; \\ \\ (1 - 2\mu)\left(1 - (1 - w)\frac{T - 2^{-\ell}(1+w)^{\ell}}{T(1 - 2^{-l}(1+w)^{\ell})}\right) & \text{otherwise.} \end{cases}$$

The fixed points occur when $\widehat{G(w)} = w$, equivalently when

$$w = \begin{cases} (1 - 2\mu) & \text{if } T < 2^{-\ell}(1 + w)^{\ell} \ ; \\ \\ (1 - 2\mu)\left(1 - (1 - w)\frac{T - 2^{-\ell}(1+w)^{\ell}}{T(1 - 2^{-l}(1+w)^{\ell})}\right) & \text{otherwise.} \end{cases} \tag{1}$$

Solving (1) numerically using $\ell=8$, $\mu=.1$, and $T=.4$, we find $w = .02869$ and $w = .7222$. A third equilibrium occurs at $w = 1 - 2\mu = .8$. A graph of $G(w) = w$ is shown is figure 1. We see the three fixed points; the first and third fixed points are stable and the middle fixed point is unstable.

We now investigate the stability of the fixed points. If $g : \mathbb{R}^N \to \mathbb{R}^N$ is any differentiable function, the differential of g is the $N \times N$ matrix defined by

$$(dg_x)_{i,j} = \frac{\partial g_i(x)}{\partial x_j}.$$

It is well known for a discrete model that if all of the eigenvalues of dg_x, have modulus less than 1, then x is an asymptotically stable fixed point of g. We will show that the maximum modulus eigenvalue of $d\widehat{G}$ is equal to the derivative of the single-variable function $\widehat{G}$ defined in equation 1. Therefore the fixed points can not only be found by solving a single-variable polynomial but additionally their stability can be determined by taking a single variable derivative.

Lemma 0.6. *Let* $P = 2^{-\ell} \prod_{i \in \mathcal{L}}(1 + \hat{z}_i)$, $P_j = 2^{-\ell} \prod_{i \in \mathcal{L} \backslash \{j\}}(1 + \hat{z}_i)$. *Then for* $j \neq k$

and $x_0 < T$,

$$\frac{\partial \widehat{G(z)}_k}{\partial \hat{z}_j} = \frac{(1 - 2\mu)(1 - \hat{z}_k)\left(\frac{P_j}{T} - P_j\right)}{(1 - P)^2}$$

$$\frac{\partial \widehat{G(z)}_k}{\partial \hat{z}_k} = \frac{(1 - 2\mu)(1 - \hat{z}_k)\left(\frac{P_k}{T} - P_k\right)}{(1 - P)^2} + \frac{(1 - 2\mu)(1 - \frac{P}{T})}{1 - P}.$$

Furthermore both partial derivatives are nonnegative for $0 \leq \mu \leq 1/2$ *and* $\hat{z}$ *in the*

simplex.

Proof. Note that $\frac{\partial P}{\partial \hat{x}_j} = P_j$ and that $x_0 = P < T$. The computation of the partial

derivatives is straightforward. For $\hat{z}$ in the simplex, $-1 \leq \hat{z}_i \leq 1$, showing P and P_j

are nonnegative. □

Lemma 0.7. *Let* $P = 2^{-\ell}(1 + w)^{\ell}$, $P' = 2^{-\ell}\ell(1 + w)^{\ell-1}$. *Then the derivative of the*

symmetric $\mathcal{G}$ *is given by*

$$\frac{d\widehat{G}}{dw} = \begin{cases} 0 & \text{if } T < 2^{-\ell}(1 + w)^{\ell} \ ; \\ \frac{(1-2\mu)(1-w)\left(\frac{P'}{T}-P'\right)}{(1-P)^2} + \frac{(1-2\mu)(1-\frac{P}{T})}{1-P} & \text{otherwise.} \end{cases} \tag{2}$$

The following lemma comes from Wright, Rowe, Poli and Stephens [10].

Lemma 0.8. *Let* A *be an* $\ell \times \ell$ *matrix where all of the diagonal entries are equal to*

d *and all of the off-diagonal entries are equal to* e. *Then the eigenvalues of* A *are*

$d + (\ell-1)e$ *with multiplicity 1 and* $d-e$ *with multiplicity* $\ell-1$.

Theorem 0.9. *At a symmetric point* $\hat{z}$ *where* $\hat{z}_k = w$ *for all* $k \in \mathcal{L}$,

$$\frac{d\hat{G}}{dw} = \frac{\partial \widehat{G(z)}_k}{\partial \hat{z}_k} + (\ell - 1)\frac{\partial \widehat{G(z)}_k}{\partial \hat{z}_j}$$

Thus the largest modulus eigenvalue of $G_{\hat{z}}$ is equal to $\frac{dG}{dw}$.

Proof. When $x_0 > T$ all of the derivatives are zero. The theorem follows by comparing the formulas in Lemma 0.6 and Lemma 0.7. Lemma 0.6 shows that the entries of $G_{\hat{z}}$ are nonnegative, and Lemma 0.8 shows that its largest eigenvalue is the one with multiplicity 1 given by the right hand side of the formula in Theorem 0.9. $\square$

Since the largest modulus eigenvalue of $\widehat{dG(z)}$ is equal to the derivative of the single variable function G, the stability of the fixed points in the cube $[-1, 1]^\ell$ is the same as the stability of the fixed points in the one variable space of symmetric populations.

We have shown that we can find the fixed points of the model by solving a single variable polynomial of degree $\ell + 1$ and furthermore, the stability of the fixed points can be determined from this equation.

Explorations of Parameter Space

We begin to explore the parameter space by finding a relationship between T, and w, when w is a fixed point. Solving equation 1 for T, in the case $T > 2^{-\ell}(1 + w)^\ell$, we find

$$T(w) = \frac{-(1 - 2\mu)(1 - w)}{(w - 1 + 2\mu - \frac{2\mu w}{x_0})}. \tag{3}$$

That is, we define $T(w)$ to be the value of T for which w is a fixed point.

Lemma 0.10. $x_0 = 2^{-\ell}(1 + w)^\ell < T(w)$ *for* $0 < w < 1 - 2u$ *and* $x_0 > T(w)$ *for* $1 > w > 1 - 2\mu$.

Proof.

$$T(w) - x_0 = \frac{x_0(1 - w - 2u + 2uw) + x_0^2(w - 1 + 2u + 2uw)}{-wx_0 + x_0 - 2ux_0 + 2uw}.$$

If $0 < w < 1 - 2u$, the denominator,

$$x_0(1 - w) - 2u(x_0 - w) > x_0 2u - 2ux_0 + 2uw = 2uw > 0.$$

$(1 - 2u) > w$ implies $(1 - w - 2u) > 0$. The numerator is therefore also greater than

zero since $x_0 < 1$.

For $w > (1 - 2\mu)$, $\frac{dx_0}{dw} > 2^\ell > \frac{dT}{dw}$. $\qquad\qquad\qquad\qquad\qquad$ $\square$

Lemma 0.11. $T(w) > 0$ *for* $0 < w < 1 - 2u$.

Proof. The numerator and the denominator are both negative for w in this range. $\quad\square$

Lemmas 0.10 and 0.11 show that the fixed points can occur in the region $0 < w <$

$1 - 2\mu$.

Recall that if $T < (1 - \mu)^\ell$, then one fixed point occurs at $w = 1 - 2\mu$. In order for

bistability to exist, there must be two additional fixed points, both less than $1 - 2\mu$.

These are solutions to equation 3. Figure 2 shows the plots of $T(w)$ for various values

of μ. Bistability occurs for a fixed value of T if a horizontal line drawn at height

T intersects the curve three times. The plot of $T(w)$ ends with a vertical line at

$w = 1 - 2\mu$.

Figure 3 shows the progression from three fixed points to one fixed point when T

is increased. A bifurcation occurs when T is approximately .43. When $T = .4$ there

are three fixed points, one near zero, one at approximately $w = .5$, and one at the critical value of $w = 1 - 2\mu = .8$. When $T = .43$ the middle fixed point merges with the fixed point at $w = .8$. When T is further increased to .45, the last fixed point has disappeared completely, leaving only the fixed point near $w = 0$.

To aid in the understanding of $T(w)$ we differentiate equation 3 and set the result equal to zero. This gives

$$0 = -\ell w^2 + (\ell + x_0 - 1)w + (x_0 - 1). \tag{4}$$

The reader should note that since equation 4 is independent of μ, the minimums in figure 2 all occur at the same value of w.

Our next task is to show the conditions under which T has a local minimum in the region $0 < w < 1 - 2\mu$.

Lemma 0.12. *For equation 4, $w = 1$ is a root of multiplicity at least 2.*

Proof. Substituting in $w = 1$, equivalently $x_0 = 1$, into the equation, we see that $w = 1$ is a root. Differentiating the equation gives

$$-2\ell w + \ell + 2^{-\ell}(1 + w)^\ell + w 2^{-\ell}\ell(1 + w)^{\ell-1} - 1 + 2^{-\ell}\ell 2^{\ell-1} = 0.$$

Substituting in $w = 1$ gives the desired result. $\qquad\square$

Lemma 0.13. $\frac{dT}{dw} = 0$ *has exactly one solution between 0 and $1 - 2\mu$ when*

$$1 - \ell\mu(1 - 2\mu) - \mu - (1 - \mu)^{\ell+1} < 0.$$

Proof. The right hand side of the equation 4 can be written as

$$-\ell w^2 + \ell w + 2^{-\ell}\left(w + \ell w^2 + w\sum_{k=2}^{\ell}\binom{\ell}{k}w^k\right) - w$$

$$+ 2^{-\ell}\left(1 + \ell w + \frac{\ell(\ell-1)}{2}w^2 + \sum_{k=3}^{\ell}\binom{\ell}{k}w^k\right) - 1$$

$$= 2^{-\ell}w\sum_{k=2}^{\ell}\binom{\ell}{k}w^k + 2^{-\ell}\sum_{k=3}^{\ell}\binom{\ell}{k}w^k$$

$$+ \left(2^{-\ell}\ell - \ell + 2^{-\ell}\frac{\ell(\ell-1)}{2}\right)w^2$$

$$+ (\ell + 2^{-\ell} - 1 + 2^{-\ell}\ell)w + 2^{-\ell} - 1.$$

The w^k terms, $k > 2$, have positive coefficients. The coefficient of the w^2 term is negative. This can be seen, for example, by plotting the coefficient with ℓ as the independent variable. The coefficient of the w term is positive and the constant term is negative. By Descartes rule of signs there are three or one positive roots of the equation. We have already seen that two of the roots are $w = 1$. Therefore there is another positive root. We now show that this root is less than $1 - 2\mu$.

$$\frac{dT}{dw}(0) = \frac{2\mu(1 - 2^{\ell})}{1 - 2\mu}$$

which is less than zero when $\mu < \frac{1}{2}$.

$$\frac{dT}{dw}(1 - 2\mu) = \frac{-1}{2}\frac{(2\ell\mu^2 - \ell\mu + 1 - \mu)(1 - \mu)^{\ell} - (1 - \mu)^{2\ell+1}}{\mu(1 - 2\mu)(1 - \mu)}$$

which is greater than zero when $1 - \ell\mu(1 - 2\mu) - \mu - (1 - \mu)^{\ell+1} < 0$.

In this case, a solution to $\frac{dT}{dw} = 0$ and likewise a root to equation 4, exists for

$0 < w < 1 - 2\mu$ $\square$

Let w_c be the critical point of T between 0 and $1 - 2\mu$. Then $T(w_c)$ is a minimum

since the first derivative of T passes from negative to positive. Since w_c is the only

critical point, it must be a global minimum in the interval $0 < w < 1 - 2\mu$.

For the parameter values in figure 2, we calculate that $w_c = .1758$. If $.1758 <$

$1 - 2\mu$, or $\mu > .4121$ then T will not have an interior local minimum. We check that

the hypothesis in Lemma 0.13 is not satisfied. $1 - \ell(1 - 2\mu) - \mu - (1 - \mu)^{\ell+1} > 0$

when $\mu > .4121$. We also should note that the minimum string length that satisfies

the hypothesis is $\ell = 4$.

Lemma 0.14. *In the case* $T < (1 - \mu)^{\ell}$, *if* $w_c < 1 - 2\mu$, *equivalently if*

$$1 - \ell\mu(1 - 2\mu) - \mu - (1 - \mu)^{\ell+1} < 0,$$

then there exists a value of T *that gives bistability.*

Proof. By Lemma 0.13 there exists horizontal lines that will cross the graph of $T(w)$

more than once. Each place of intersection represents a fixed point. Another fixed

point of G exists at $w = 1 - 2\mu$. $\square$

We also note that bistability exists if $T(w_c) < T < (1 - \mu)^{\ell}$.

It remains to determine the stability of the fixed points.

Lemma 0.15. $G(w) > 0$ *for* $0 \leq w \leq 1$.

Proof. If $T > x_0$

$$G(0) = (1 - 2u)\left(1 - \frac{T(1 - 2^{-\ell}) - T + 2^{-\ell}}{T(1 - 2^{-\ell})}\right)$$

$$= (1 - 2u)\left(1 - \frac{2^{-\ell}(1 - T)}{T(1 - 2^{-\ell})}\right)$$

$$> (1 - 2u)\left(1 - \frac{2^{-\ell}(1 - 2^{-\ell})}{2^{-\ell}(1 - 2^{-\ell})}\right)$$

$$= 0.$$

If $T \leq x_0$ then $G(0) = 1 - 2\mu$. By Lemma 0.6, $\frac{d\widehat{G}}{dw} > 0$, and the conclusion follows. $\square$

Theorem 0.16. *If $0 < \mu < 1/2$, then there can be at most three fixed points for $\widehat{G}$. When there are three fixed points, they are stable, unstable and stable when ordered by w values. If there are two fixed points, then a small perturbation of either T or μ can give one fixed point.*

Proof. Since $\frac{dG}{dw} > 0$ then at a fixed point where the graph crosses from above to below the slope must be less than one and therefore must be stable. A fixed point where the graph crosses from below to above must have slope greater than one and is unstable.

Since $\widehat{G}(0) > 0$, when there are three fixed points, the graph must cross the diagonal from above to below, then below to above.

$\square$

We have exactly characterized the fixed points for the gene pool model on the NEEDLE fitness function. For fixed values of T and μ the location of the fixed points can be found using equation 1.

For a fixed value of u the range of values of T which give bistability can be found. For example see figure 4. The area between the two curves is the region in (μ, T) space where bistability occurs. The top curve is $T = (1 - \mu)^\ell$ and the lower curve is found by solving $\frac{dT}{dw} = 0$ to find w_c and then taking $T(w_c)$.

For a fixed value of T and ℓ the range of μ which give bistability can be found. This can be seen in figure 5. This figure shows the region in (ℓ, μ) space for which bistability occurs. As noted before, a string length of 4 or more is needed for bistability. The top curve of each pair in this figure is $\mu = 1 - T^{\frac{1}{\ell}}$ The bottom curve of each pair was found for each ℓ by solving $T(w_c) = T$ for μ for $T = .1, .3, .5, .7$.

Likewise for a fixed value of μ and ℓ the range of T which give bistability can be found. This can be seen in figure 6. The top curve of each pair is $T = (1 - \mu)^\ell$. The bottom curve of each pair is found by finding $T(w_c)$.

Empirical Results

Simulations were performed to see if the fixed points predicted by the infinite gene pool model occur in finite populations GA runs. The simulations were run using a program written by Dr. Alden Wright, J.D. Zeiler, and Jennifer Parham. The results are given in the table below. In the simulations we used truncation selection, the needle fitness function, and a string length of 15. The simulations were run for twenty generations. The average x_0 were calculated by taking the mean of the last 10 generations. The initial population was generated by using an allele probability

equal to the predicted steady state allele distribution. The populations did stay close to the predicted value of x_0 for all 20 generations. We also note that the averages are close to the predicted values even when the population size is relatively small.

μ	T	Crossover	Population Size	average x_0	Model Prediction
.1	.15	uniform	10000	.00003, .20571	.000031, .20589
.1	.15	uniform	100	.000, .235	.000031, .20589
.1	.15	genepool	10000	.00003, .20551	.000031, .20589
.1	.15	genepool	100	.000, .192	.000031, .20589
.05	.15	uniform	10000	.00003, .46335	.000031, .46329
.05	.15	uniform	100	.000, .451	.000031, .46329
.05	.15	genepool	10000	.00006, .46093	.000031, .46329
.05	.15	genepool	100	.000, .473	.000031, .46329
.05	.2	uniform	10000	.00002, .46117	.000031, .46329
.05	.2	uniform	100	.000, .445	.000031, .46329
.05	.2	genepool	10000	.00001, .46451	.000031, .46329
.05	.2	genepool	100	.000, .442	.000031, .46329

Conclusion

In this paper we have shown that an infinite population GPR model closely approximates the finite population two-parent uniform crossover GA when the fitness function exhibits a single peak. Under the recombination model, the complicated

dynamical system of the finite population GA becomes tractable. In particular, we can explicitly calculate the fixed points and determine their stability by examining a single variable polynomial function. Finite population simulations suggest that TPR and GPR produce bistability. The fixed points produced in the simulations closely match those predicted by the model.

Furthermore, we have demonstrated that the infinite population GPR model correctly predicts the presence of bistability in the finite population GA. We have derived explicit formulas that relate the parameter values under which the bistability phenomena occurs.

Figures
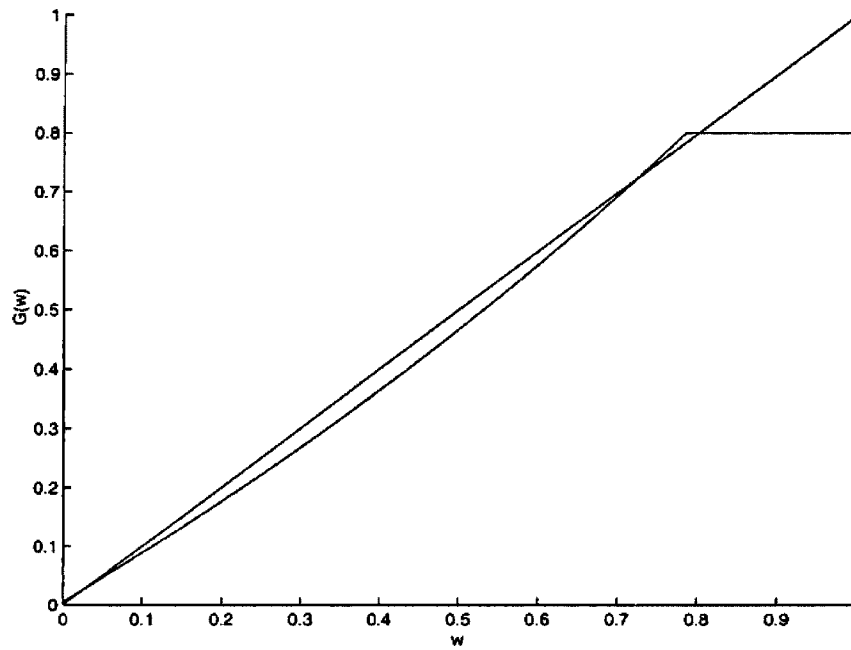
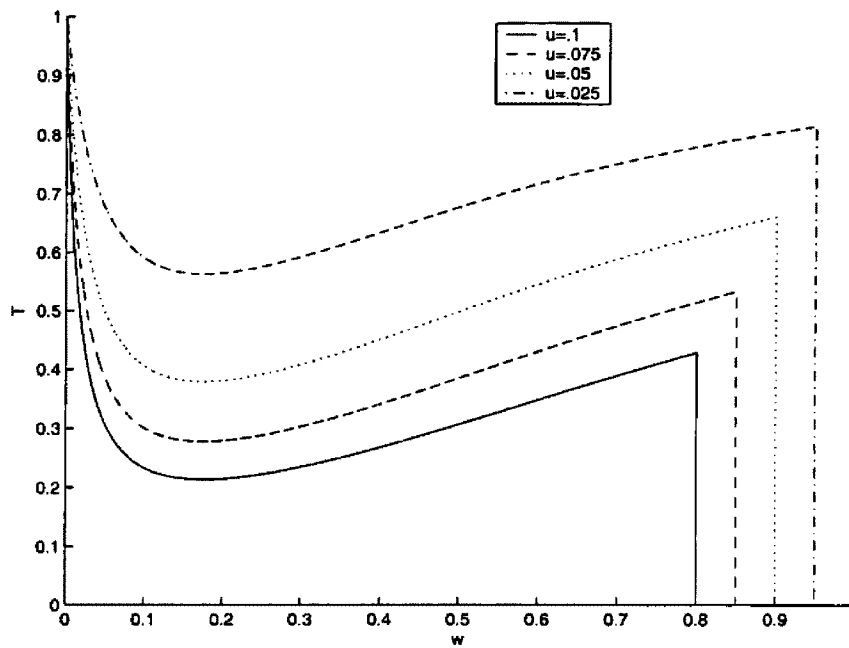Figure 1: Graph of $\widehat{G(w)}$ and 45° line for $\mu = .1$, $\ell = 8$ and $T = .4$.

Figure 2: Plots of T for various μ with $\ell = 8$
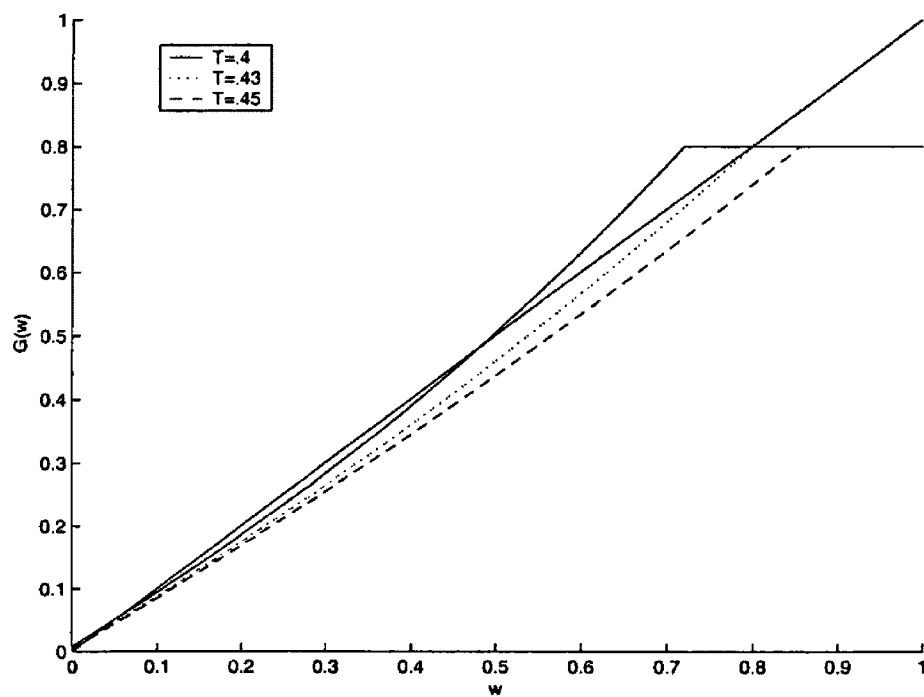

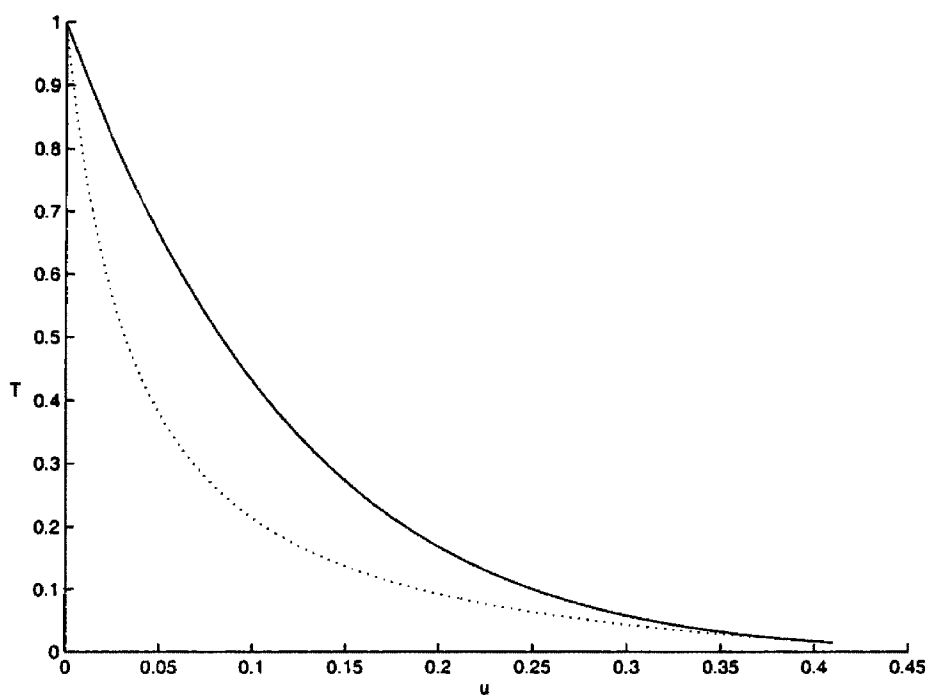Figure 3: Plots of $\widehat{G}$ for various T with $\mu = .1$ and $\ell = 8$

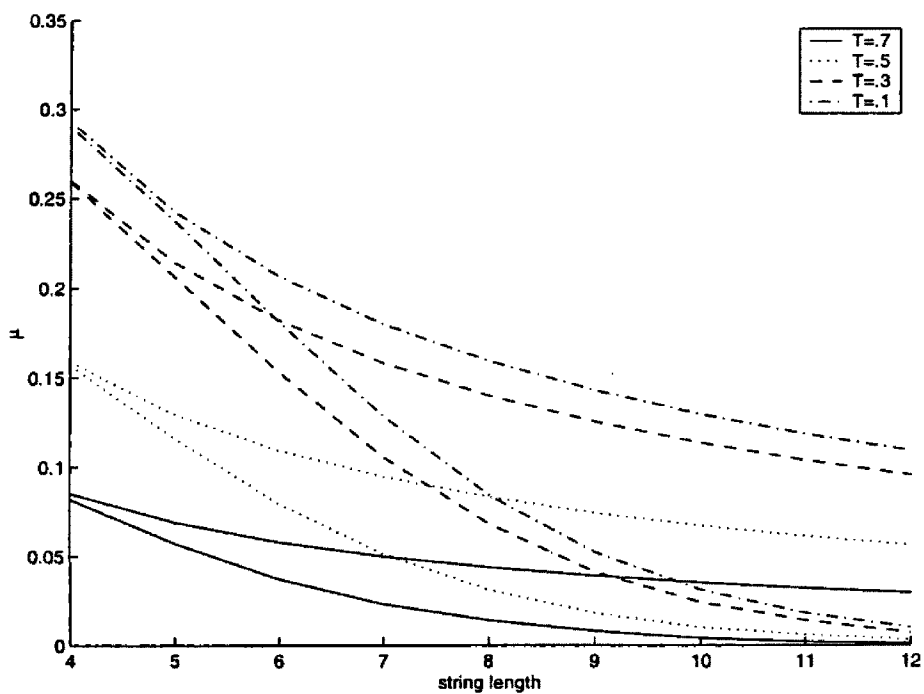Figure 4: Region of bistability in (u, T) space for $\ell = 8$

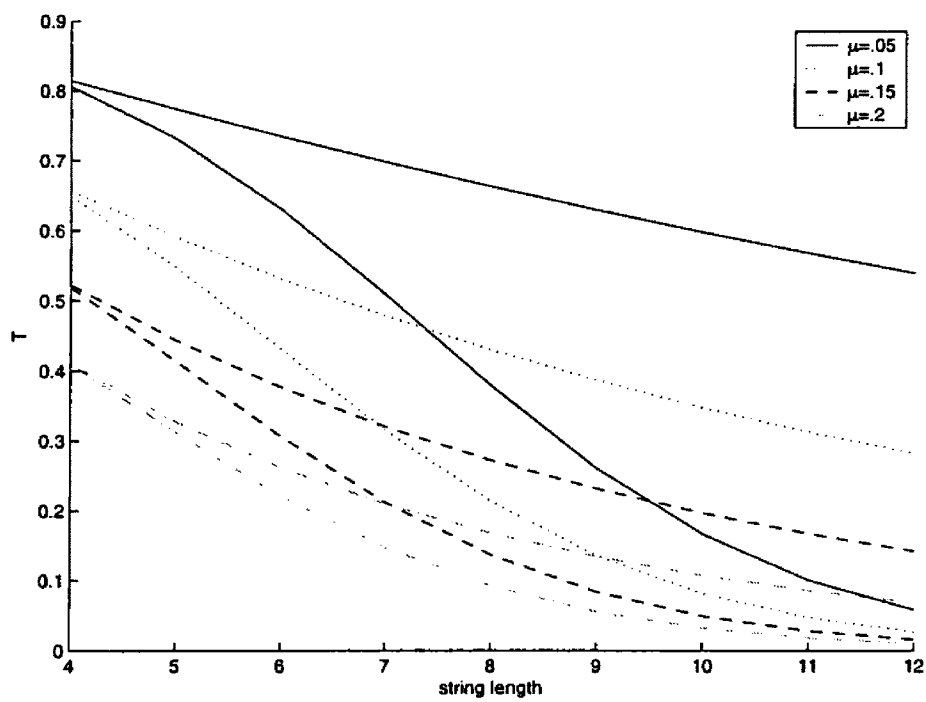Figure 5: Upper and lower critical mutation rates for different T values.

Figure 6: Region of bistability for different (ℓ, T) space.

Bibliography

[1] Boerlijst, M.C., S. Bonhoeffer, and M.A. Nowak, 1996. Viral quasi-species and recombination. *Proc. Royal Society London* **B 263**, 1577-1584.

[2] Geiringer, H., 1944, On the probability of linkage in medelian heredity. *Annals of Mathematical Statistics* **15**,25-27.

[3] Mühlenbein, H. and H.M. Voight. Genepool recombination in genetic algorithms. In I.H. Osman and J.P. Kelly (Eds), *Meta Heuristics: Theory and Applications,*Boston, London, Dordrecht, 1996. Kluwer Academic Publications. 53-62.

[4] Ochoa, G. and I. Harvey, 1997. Recombination and error thresholds in finite populations. *Foundations of Genetic Algorithms* **5** San Mateo, 245-264. Morgan Kauffman.

[5] Syswerda, G., 1993. Simulated Crossover in genetic algorithms. *Foundations of Genetic Algorithms 2,* San Mateo. Morgan Kaufmann.

[6] Vose, M.D., 1999. *The Simple Genetic Algorithm: Foundations and Theory.* Cambridge, MA: MIT Press.

[7] Vose, M.D. and A.H. Wright 1998. The simple genetic algorithm and the Walsh transform: Part II, the inverse *Evolutionary Computation*6(**3**),275-289.

[8] Wright, A.H., and Alexandru Agapie, 2001. Cyclic and Chaotic Behavior in Genetic Algorithms. In Lee Spector, Erik D. Goodman, Annie Wu, W. B. Langdon, Hans-Michael Voight, Mitsuo Gen, Sandip Sen, Marco Dorigo, Shahram Pezeshk, Max H. Garzon, Edmund Burke(Eds), *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2001)*, San Mateo: Morgan Kauffman, 718-724.

[9] Wright, A. H., J.E. Rowe, and J.R. Neil, 2002. Analysis of the simple genetic algorithm on the single-peak and double-peak landscapes. *Proceedings of the Congress on Evolutionary Computation 2002.* http://www.cs.umt.edu/u/wright

[10] Wright, A. H., J.E. Rowe, R. Poli, and C.R. Stephens, 2003. Bistability in a Gene Pool GA with Mutation. In J. E. Rowe, K. DeJong, R. Poli(Eds), *Foundations of genetic algorithms 7 (FOGA-7)*, San Mateo: Morgan Kauffman.