The Dynamics of a Genetic Algorithm on a Model Hard Optimization Problem

Alex Rogers
Adam Prügel-Bennett
Image, Speech and Intelligent Systems Research Group
Department of Electronics and Computer Science
University of Southampton
Highfield, Southampton, SO17 1BJ
England

A model of a hard optimization problem suggested in the literature is considered. The dynamics of a genetic algorithm (GA) using ranking selection, mutation and uniform crossover are completely modeled on this problem. These results are general and are valid for any symmetrical concave function of unitation. Full finite population effects are taken into account allowing a novel analytical comparison of roulette wheel and stochastic universal sampling. Closed form expressions are derived for the equilibrium population distribution of this model. The first passage time to move from a local to a global minimum in a two potential well landscape is calculated. A comparison is made with a stochastic hill climber and a GA without crossover. The GA with crossover is shown to perform orders of magnitude faster giving some insights into the nature of GA search and the crossover operator on these types of problem.

1. Introduction

Much of the theoretical modeling of genetic algorithms (GA) has necessarily been performed on simple models of the GA on a simple fitness landscape. Examples of these are functions of unitation such as one-max [8, 16, 10, 11] and the royal road functions of Mitchell, Holland and Forest [7]. Whilst giving some insights into the behavior of the GA, they are not representative of the sort of optimization problems to which GA are often applied. GA are shown to perform poorly on these model landscapes compared to other optimization techniques such as stochastic hill climbing, whilst there remains a body of empirical evidence that GA often outperform or at least perform comparatively well with other optimization methods, on real-world problems.

In order to gain some insight into the performance of GA on real world optimization problems, it is necessary to consider harder problem spaces. Characterizing the hardness of a problem has been an active area of research and has led to measures of problem difficulty such as

Fitness Distance Correlation [6] and Epistasis Variance [4]. There is however much debate as to the interpretation of these measures and their applicability to real world fitness landscapes.

Despite these arguments, there is some consensus as to what features make a problem space hard to search. There may be many local minima. These local minima may be separated by a potential barrier from better solutions, resulting in the need for non-local search steps. If these local minima occupy the majority of the search space it may take a long time to generate the moves necessary to fall into the basin of attraction of the global minimum.

In this paper we analyze a previously published model of a hard optimization problem [14] which addresses the last two instances. The problem is known as the 'Basin with a Barrier' fitness landscape and is a function of unitation. It has some of the features of a hard optimization problem but is still amenable to analysis.

The landscape consists of a large local minimum separated from the global minimum by a potential barrier. We consider a series of L spins whose value may be 1 or -1 and consider the total magnetization, M, of the string

$$M = \sum_{i=1}^{L} S_i. \tag{1}$$

The potential or fitness, which we are trying to minimize, is a function of this magnetization and is given generically as

$$V(M) = \begin{cases} (M - M_l)^2 + V_l & \text{if} \quad M \leq M_b \\ 0 & \text{if} \quad M > M_b. \end{cases} \tag{2}$$

The choice of a quadratic function is arbitrary, as is the lack of parameter controlling the 'slope' around the minimum. The use of ranking selection as opposed to some form of fitness proportional selection makes these factors redundant. Thus the only important parameters are the positions of the local minimum and the barrier — M_l and M_b — and the resulting analysis will hold directly for any symmetric concave function.

The entropy of the system, S, is such that the number of states in the global minimum is much smaller than that in the local minimum whilst the maximum entropy state is some distance from both the local and global minima. Figure 1 shows the landscape schematically.

We propose that whilst being a toy problem, this model holds some of the features seen in combinatorial optimization problems such as the traveling salesman. We expect random search to produce poor solutions most of the time and we expect these poor solutions to occupy the majority of the problem space. We expect good solutions to be near

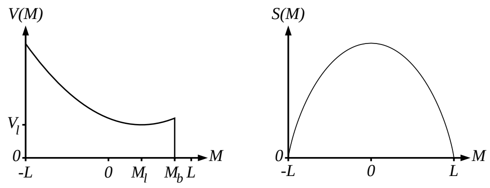


Figure 1. Diagram of potential and entropy for the 'Basin with a Barrier' fitness landscape.

one another in problem space but they may be separated by non-local moves.

Shapiro and Prügel-Bennett [14] analyzed this landscape using a statistical mechanics formalism of a GA under Boltzmann selection and uniform crossover. They showed that in the infinite population limit there is a phase in which the population will move from the local minimum to the global minimum from any initial configuration. The analysis suggests that this can occur orders of magnitude faster than a stochastic hill climber can find the global minimum. In order to obtain qualitative results, much of the complexities of modeling the dynamics of the GA were omitted and the fit between theoretical and simulation results was poor.

In this paper we extend the analysis and consider the aspect not touched on by the earlier paper — the time required for a GA to find one solution in the global minimum — the first passage time. We extend the formalism used in the earlier paper to ranking selection as opposed to Boltzmann selection. Apart from its more common use in real GA, the use of ranking selection leads to significant gains. We can solve the dynamics of ranking selection on the one-max fitness landscape by considering just the mean and variance of the population distribution. As the ranking of population members for selection removes the need to know the exact structure of the population and the fitness landscape, we can extend this analysis of selection to any symmetrical concave function. By considering finite population effects on selection, we are able to analytically compare and model the two most common selection strategies — roulette wheel and stochastic universal sampling (SUS) or Baker selection [2]. Finally by considering the correlation of the population through selection, the effect of crossover can be accurately modeled.

The result of this is an accurate model of the dynamics of the GA on any symmetrical concave function of unitation. We are able to use this to predict the first passage time to find a solution in the global minimum of the 'Basin with a Barrier' fitness landscape. This time is shown to be orders of magnitude faster than that for a stochastic hill climber and that of the GA without crossover. This finding gives some insight into the search power of crossover.

In section two we present the model GA and in section three discus the formalism used to model the dynamics. Section four details the effect of ranking selection including finite population effects and a comparison between roulette wheel selection and SUS. Section five details the modeling of the other genetic operators — mutation and crossover. Section six reviews the accuracy of the model compared to simulation results and section seven discusses the calculation of first passage time for the GA and a stochastic hill climber on the 'Basin with a Barrier' problem. The results are discussed in section eight. In appendices A,B and C, we present, in more detail the analysis of selection, mutation and crossover operators.

2. The Model Genetic Algorithm

The model genetic algorithm consist of a population of P individuals. Each individual consists of a string of L spins whose value may be 1 or -1. The magnetization of any individual μ is given by the sum of its spins

$$M_\mu = \sum_{i=1}^{L} S_i^\mu \quad \text{where} \quad S_i^\mu = \{-1, 1\}. \tag{3}$$

The potential which we are trying to minimize is given by the generic expression from section one. The population is initialized with random strings.

We consider a generational GA and use ranking selection [1] to select the mating pool from the initial population. The mutation operator is then applied whereby each spin has a small probability of mutation

$$S_i^\mu \to -S_i^\mu \quad \text{with probability } \gamma. \tag{4}$$

We then apply the crossover operator to the mating pool to produce the next generation. Population members are randomly paired and uniform crossover [15] applied whereby spins are randomly drawn from each parent

$$S_i^\mu = \chi_i^\mu S_i^\alpha + (1 - \chi_i^\mu)\, S_i^\beta \tag{5}$$

where

$$\chi_i^\mu = \begin{cases} 1 & \text{with probability } 1/2 \\ 0 & \text{with probability } 1/2. \end{cases} \tag{6}$$

The complementary offspring is also created and the pair replace their parents.

3. The Statistical Physics Formalism

Rather than model the exact structure of the population, we use a statistical physics technique whereby we model several macroscopic variables of the ensemble average of an infinite number of finite populations. In this way we model the mean behavior of an evolving finite population. The formalism used here was first developed by Prügel-Bennett and Shapiro in an analysis of a GA using Boltzmann selection [10] [11]. The macroscopic variables used are the cumulants of the population magnetization distribution.

During the course of the paper we make a number of approximations as we wish to capture the full dynamics of the GA but not at the expense of an over complex model. We make an approximation to the exact distribution by only calculating a finite number of cumulants. With Boltzmann selection it was found that at least four cumulants were needed to accurately model the dynamics of the GA due to finite population effects. When modeling ranking selection with mutation and uniform crossover we find that the higher order cumulants are sufficiently small that we may use just two cumulants — the mean and variance of the distribution — to accurately model the dynamics. Thus we assume a Gaussian magnetization distribution. This reduction in the number of macroscopic variables is significant as it enables closed form expressions to be derived for the evolution endpoints. For a finite population the first two cumulants are given by

$$\begin{aligned} \kappa_1 &= \langle M \rangle \\ \kappa_2 &= \langle M^2 \rangle - \langle M \rangle^2 \end{aligned} \tag{7}$$

where $\langle \dots \rangle$ represents the average over the population. The cumulants of the ensemble we denote as K_1 and K_2. Due to well known finite sampling effect, the cumulants of any particular finite population, of size P, drawn from the ensemble will differ slightly from those of the ensemble. The two terms are related by the expressions

$$\begin{aligned} \kappa_1 &= K_1 \\ \kappa_2 &= \left(1 - \frac{1}{P}\right) K_2. \end{aligned} \tag{8}$$

We only consider mean behavior and do not consider the fluctuations from run to run. Prügel-Bennett carried out a complete calculation of these fluctuations in a previous paper by calculating the covariance of each of the cumulants [9]. The effects of these were shown to be small.

When we calculate the effect of selection on a finite population we must also consider the stochastic nature of the selection operator. This was previously done exactly for Boltzmann selection however for ranking selection we develop an approximation which is shown to be justified *a posteriori* .

The effect of mutation is calculated in section 5.1 in terms of the change in the first two cumulants. This analysis involves no approximation and is exact.

Calculating the effect of crossover involves a measure of the correlation or similarity of the strings. In previous work [10] [11], weak selection has been considered and a maximum entropy argument used to approximate the correlation from the mean and variance of the population magnetization distribution. When selection is strong however this leads to errors. As selection increases fitness through duplication of population members, these duplicates have a significant effect on the correlation of the population. In this paper we use a approach similar to that developed by Rattray [12] and model the degree of duplication within the population as another macroscopic variable. We then use a maximum entropy argument to combine the degree of duplication with the mean magnetization of the population to calculate a more accurate measure of the population correlation.

4. Selection

Ranking selection [1] is a commonly used form of GA selection and was suggested by Baker as a scheme to limit the speed with which a single highly fit individual would take over the population. The expected number of times that the population member of rank i will be represented in the next generation is controlled by the parameter MAX and is given by

$$E_i = \mathrm{MAX} - 2\,(\mathrm{MAX} - 1)\,\frac{i-1}{P-1}. \tag{9}$$

The fittest population member is expected to be represented MAX times and the least fit $(2 - \mathrm{MAX})$ times. MAX may take any value between one and two. Ranking selection has several nice features which make it amenable to analysis. The ranking of the population effectively makes a piecewise linear approximation of the fitness function. Thus although illustrated as a quadratic function, the analysis presented here holds for any symmetrical basin. Also by putting upper and lower limits on the expected number of times any individual may be selected, we can calculate and compare roulette wheel and stochastic universal sampling or Baker selection.

The effect of ranking selection on a population is found by first considering the infinite population case.

4.1 Infinite Population Model

We first consider the simpler case of the one-max landscape and an infinite population size. Whilst mutation and crossover operate on the magnetization distribution we must consider the fitenss function when calculating the effect of selection. In the case of one-max the ranking of an individual with magnetization M is solely related to its position within the population. The expression for the expected number of occurrences of any population member given in equation 9 can be rewritten in terms of the population magnetization distribution. We can write this as

$$E_M = (2 - \mathrm{MAX}) + 2\,(\mathrm{MAX} - 1)\int_{-\infty}^{M} \rho\,(M')\,\mathrm{d}M', \qquad (10)$$

where $\rho\,(M')$ describes the magnetization distribution. The first and second moments of the population distribution after selection are found by integrating the weighting over the distribution. We make the approximation that the distribution is a Gaussian with mean, K_1 and variance, K_2. Appendix A illustrates the calculation of these integrals. The resulting moments are

$$\begin{aligned}
\langle M \rangle &= \int_{-\infty}^{\infty} M E_M \rho\,(M)\,\mathrm{d}M \\
&= K_1 + (\mathrm{MAX} - 1)\sqrt{\frac{K_2}{\pi}} \\
\langle M^2 \rangle &= \int_{-\infty}^{\infty} M^2 E_M \rho\,(M)\,\mathrm{d}M \\
&= K_2 + K_1^2 + 2\,(\mathrm{MAX} - 1)\,K_1\sqrt{\frac{K_2}{\pi}}. \qquad (11)
\end{aligned}$$

Thus using the expressions in equation 7, the cumulants after selection are given by

$$\begin{aligned}
\langle K_1 \rangle_s &= K_1 + (\mathrm{MAX} - 1)\sqrt{\frac{K_2}{\pi}} \\
\langle K_2 \rangle_s &= \left[1 - \frac{(\mathrm{MAX} - 1)^2}{\pi}\right] K_2 \qquad (12)
\end{aligned}$$

where $\langle \ldots \rangle_s$ represents the average over all ways of performing selection. Clearly the variance is decreased by a factor determined by the selection pressure whilst the mean increases by a factor dependent on $\sqrt{K_2}$ — a measure of the width of the distribution. The analysis of ranking selection for an infinite population when MAX = 2 is identical to that of tournament selection which has previously been performed by Blickle and Thiele [3]. Changing the parameter MAX is equivalent to introducing a probabilistic element into tournament selection.

Extending this to the basin landscape is straightforward since ranking selection makes a piecewise linear approximation of the fitness function. Thus the results we derive here apply to any symmetrical concave fitness landscape with a minimum at M_l. The weighting of a population member whose magnetization is M is related to not only its position within the population magnetization distribution but also the position of this minimum, M_l and can be written as

$$E_M = \begin{cases} \mathrm{MAX} - 2\,(\mathrm{MAX} - 1)\displaystyle\int_{M}^{2M_l - M} \rho\,(M')\,\mathrm{d}M' & \text{when } M \leq M_l \\ \mathrm{MAX} - 2\,(\mathrm{MAX} - 1)\displaystyle\int_{2M_l - M}^{M} \rho\,(M')\,\mathrm{d}M' & \text{when } M \geq M_l \end{cases} \tag{13}$$

The calculation of the cumulants after selection is similar to that carried out above but involves rather more algebra. Again this is illustrated in appendix A. The resulting expressions are[1]

$$\langle K_1 \rangle_s = K_1 + (\mathrm{MAX} - 1)\sqrt{\frac{K_2}{\pi}}\mathrm{erf}\left(\frac{M_l - K_1}{\sqrt{K_2}}\right)$$

$$\begin{aligned} \langle K_2 \rangle_s = \Bigg[1 - \frac{2\,(\mathrm{MAX}-1)}{\pi}\exp\left(-\frac{(M_l - K_1)^2}{K_2}\right) \\ - \frac{(\mathrm{MAX}-1)^2}{\pi}\mathrm{erf}^2\left(\frac{M_l - K_1}{\sqrt{K_2}}\right)\Bigg] K_2. \end{aligned} \tag{14}$$

Clearly when M_l is sufficiently large that little of the distribution falls over the minimum, the expressions above simplify to those presented earlier for one-max.

4.2 Finite Population Effects

The above expressions describe the infinite population response to selection. A real GA however has a finite population and the stochastic nature of the selection operator leads to a deviation from the infinite population response. Whilst the first cumulant is unaffected by finite population effects, the effects on the second cumulant is significant and gives rise to genetic drift and performance differences between roulette wheel selection and stochastic universal sampling (SUS) or Baker selection [2].

To calculate directly the effect of selection on a finite population is not possible. Instead we make an approximation and consider that selection operates on an infinite population from which we sample a finite population. We consider the change in variance to be due to the

[1] Here $\mathrm{erf}\,(x)$ represents the standard error function. See appendix A.

product of the infinite population result and a factor which describes the stochastic nature of the selection scheme. We find this factor by considering sampling from a finite population independent of magnetization. In this way the factor which we calculate is a measure of the genetic drift inherent in the sampling scheme.

If we consider a finite population, the second cumulant, κ_2, is given by

$$\kappa_2 = \frac{1}{P}\sum_{\alpha=1}^{P} M_\alpha^2 - \left(\frac{1}{P}\sum_{\alpha=1}^{P} M_\alpha\right)^2. \tag{15}$$

Separating out terms which are not independent gives

$$\kappa_2 = \sum_{\alpha=1}^{P}\left(\frac{1}{P} - \frac{1}{P^2}\right) M_\alpha^2 - \frac{1}{P^2}\sum_{\alpha\neq\beta} M_\alpha M_\beta. \tag{16}$$

We now apply some sampling scheme to this population and draw from it a new population of P individuals. In this new population there are now n_α copies of population member M_α and the second cumulant is given by

$$\kappa_2' = \frac{1}{P}\sum_{\alpha=1}^{P} n_\alpha M_\alpha^2 - \left(\frac{1}{P}\sum_{\alpha=1}^{P} n_\alpha M_\alpha\right)^2. \tag{17}$$

We can separate out terms which are independent to give

$$\kappa_2' = \sum_{\alpha=1}^{P}\left(\frac{n_\alpha}{P} - \frac{n_\alpha^2}{P^2}\right) M_\alpha^2 - \sum_{\alpha\neq\beta} \frac{n_\alpha n_\beta}{P^2} M_\alpha M_\beta. \tag{18}$$

As we are considering sampling independently of the individuals magnetization, n_α is independent of M_α and we may average separately over all ways of performing the sampling. This gives the result

$$\kappa_2' = \left(\frac{\langle n\rangle}{P} - \frac{\langle n^2\rangle}{P^2}\right)\sum_{\alpha=1}^{P} M_\alpha^2 - \frac{\langle n_\alpha n_\beta\rangle}{P^2}\sum_{\alpha\neq\beta} M_\alpha M_\beta. \tag{19}$$

Our sampling scheme keeps the population size constant so $\langle n\rangle = 1$. We can use this to derive the following simple identity

$$\left(\sum_{\alpha=1}^{P} n_\alpha\right)^2 = P^2 = \sum_{\alpha=1}^{P} n_\alpha^2 + \sum_{\alpha\neq\beta} n_\alpha n_\beta \tag{20}$$

and thus

$$\langle n_\alpha n_\beta\rangle = \frac{P - \langle n^2\rangle}{P-1}. \tag{21}$$

Using this result gives the expression

$$\kappa_2' = \frac{P - \langle n^2 \rangle}{P-1} \left[\sum_{\alpha=1}^{P} \left(\frac{1}{P} - \frac{1}{P^2} \right) M_\alpha^2 - \frac{1}{P^2} \sum_{\alpha \neq \beta} M_\alpha M_\beta \right]. \tag{22}$$

The expression in the square brackets is simply the variance of the initial population and thus

$$\kappa_2' = \frac{P - \langle n^2 \rangle}{P-1} \kappa_2. \tag{23}$$

This factor describes the loss in variance when sampling is performed on a finite population independently of magnetization and is simply dependent on $\langle n^2 \rangle$ — a measure of the variance in the number of times any population member is selected. This is a general result and has been used by the authors in another paper as a basis for the comparison of genetic drift in a range of selection schemes [13].

We can combine this result with the infinite population result to give the variance after selection of a finite population

$$\langle K_2 \rangle_s = \frac{P - \langle n^2 \rangle}{P-1} \left[1 - \frac{2\,(\mathrm{MAX}-1)}{\pi} \exp\left(-\frac{(M_l - K_1)^2}{K_2} \right) - \frac{(\mathrm{MAX}-1)^2}{\pi} \mathrm{erf}^2 \left(\frac{M_l - K_1}{\sqrt{K_2}} \right) \right] K_2 \tag{24}$$

4.3 Calculating $\langle n^2 \rangle$

The finite population effect of the selection scheme calculated above was shown to be entirely dependent on the value of $\langle n^2 \rangle$ - a measure of the variance in the number of times any population member is selected. Baker [2] noted that whilst any individual with rank i is expected to occur E_i times after selection, the stochastic nature of roulette wheel selection allows anywhere between 0 and P copies to be selected. This is the source of convergence of a finite population due to stochastic effects — genetic drift.

Baker proposed stochastic universal sampling (SUS) as a selection scheme which limits the range of possible occurrences to either $\lfloor E_i \rfloor$ (E_i rounded down to the nearest integer) or $\lceil E_i \rceil$ (E_i rounded up to the nearest integer). Whilst no arguments were made as to the virtue of doing this in the original paper, it is generally understood that the use of SUS reduces the effects of genetic drift. Intuitively we can see that limiting the range of possible occurrences will reduce $\langle n^2 \rangle$ and hence reduce the loss in variance through stochastic effects. However using the expression derived above we can calculate this difference and formally compare them.

4.3.1 Roulette Wheel Selection

Selecting using roulette wheel selection gives rise to a binomial distribution in which we make m trials with a probability of success p. The standard result for a binomial distribution is

$$\langle n^2 \rangle = m\,(m-1)\,p^2 + mp. \tag{25}$$

Assuming independence between each population member, $\langle n^2 \rangle$ can be found by averaging over the weighting for each rank. If the probability of success is E_i/P and we make P trials, the resulting expression is

$$\langle n^2 \rangle = \frac{1}{P}\sum_{i=1}^{P}\left[P\,(P-1)\,\frac{E_i^2}{P^2} + P\frac{E_i}{P}\right]. \tag{26}$$

Using the expression for E_i given in equation 9 and performing the summation gives

$$\langle n^2 \rangle = 3 + \frac{(P+1)\left(\mathrm{MAX}^2 - 2\mathrm{MAX} - 2\right)}{3P}. \tag{27}$$

4.3.2 Stochastic Universal Sampling

In stochastic universal sampling, for any value of E_i we have either $\lfloor E_i \rfloor$ or $\lceil E_i \rceil$ copies after selection. The probabilities of either are given by

$$n_i = \begin{cases} \lceil E_i \rceil & \text{with probability } E_i - \lfloor E_i \rfloor \\ \lfloor E_i \rfloor & \text{with probability } \lceil E_i \rceil - E_i. \end{cases} \tag{28}$$

Assuming independence between individuals we can find $\langle n^2 \rangle$ by calculating the distribution averaged over each ranking

$$\langle n^2 \rangle = \frac{1}{P}\sum_{i=1}^{P}\lceil E_i \rceil^2\left(E_i - \lfloor E_i \rfloor\right) + \lfloor E_i \rfloor^2\left(\lceil E_i \rceil - E_i\right). \tag{29}$$

Calculating this for most selection schemes in far from trivial as the exact population fitness distribution is required. However for ranking selection E_i and thus $\lfloor E_i \rfloor$ and $\lceil E_i \rceil$ are known independently of the population structure. For $i \le P/2$, $\lfloor E_i \rfloor = 1$ and $\lceil E_i \rceil = 2$ whilst when $i > P/2$, $\lfloor E_i \rfloor = 0$ and $\lceil E_i \rceil = 1$. Applying these ranges gives

$$\langle n^2 \rangle = \frac{1}{P}\sum_{i=1}^{P/2} 4\,(E_i - 1) + (2 - E_i) + \frac{1}{P}\sum_{i=\frac{P}{2}+1}^{P} E_i. \tag{30}$$

Again using the expression for E_i given in equation 9 and performing the summation gives

$$\langle n^2 \rangle = \mathrm{MAX} - \frac{(\mathrm{MAX}-1)}{2}\frac{P-2}{P-1}. \tag{31}$$

4.4 Discussion

For any particular value of MAX, $\langle n^2 \rangle$ for SUS is less than that of roulette wheel selection, showing that the variance in the number of times any population member is selected is less. This results in a smaller reduction of population magnetization variance at each selection step and thus slower convergence of the population through stochastic effects. Under neutral selection where MAX $= 1$, SUS shows no loss in population variance whilst under roulette wheel selection variance still decreases at each generation by a factor of $(1 - 1/P)$ — the rate of genetic drift. This analysis confirms the generally held beliefs about SUS and roulette wheel selection.

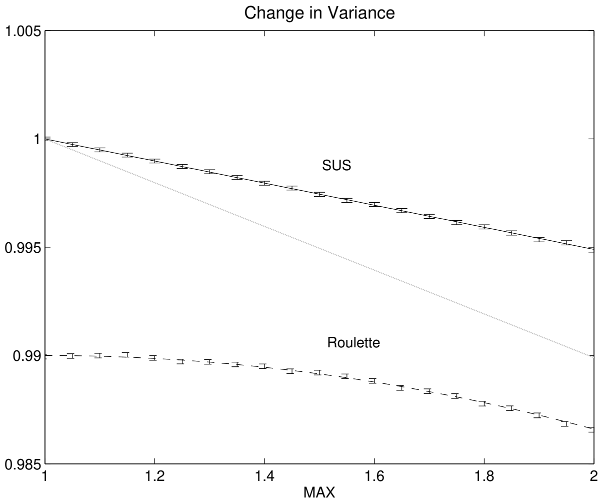


Figure 2. Comparison of theoretical and simulation results for $\frac{P-\langle n^2 \rangle}{P-1}$ after roulette wheel selection (dashed line) and SUS (solid line). Population size P=100.

The results of this analysis of sampling are exact and are general for any fitness function. Figure 2 shows the exact analytical calculation for the loss of variance after sampling a population against simulation results. In this case the population size is 100. Although the difference is apparently small, after many generations, these effect accumulate and it is this difference which gives rise to a significant difference in performance of the two selection schemes.

When we consider the effect of selection dependent on magnetization value in a finite population, we are making an approximation that we can consider the finite and infinite cases separately and then combine them for the final result. Figure 3 shows a comparison of simulation results with theory in this case. A population of size P=2, 10 and 100

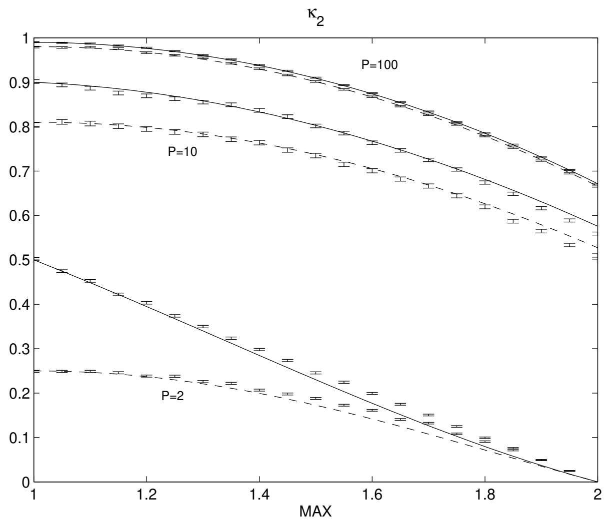


Figure 3. Comparison of theoretical and simulation results for κ_2 after roulette wheel selection (dashed line) and SUS (solid line). Population size P=2, 10 and 100. In each case, SUS is the curve with the least change in variance

was initially assigned a fitness drawn from a unit Gaussian ($K_2 = 1$) and then selection applied once. The variance of the population after selection was calculated and averaged over 1 000 runs. Both roulette wheel and SUS are shown with SUS consistently experiencing less loss in variance through selection.

The finite sampling effect described in section two is clearly seen in this figure as we are plotting κ_2 — the variance of a finite population. The variance of the initial finite population drawn from the unit Gaussian is $(1 - 1/P)$ and not unity as may be expected.

The theoretical results are in good agreement with the simulation results even at extremes of small population sizes where finite population effects are dominant. Our approximation that we can consider finite population effects to be the product of the two factors seems to be justified over the typical range of parameter values. The analysis clearly indicates the effect different selection schemes have on genetic drift and hence GA performance.

5. Genetic Operators

The search power of a GA is determined by how the selection scheme interacts with other genetic operators — mutation and crossover. To fully understand the dynamics of the GA we must model these operators.

5.1 Mutation

When the mutation operator is applied, each spin has a small probability of mutation, γ. The analysis of the effect of mutation on the cumulants of the ensemble magnetization distribution has previously been performed by Prügel-Bennett and Shapiro [10] [11]. We present the results here and, for completeness, present the derivation in the notation used here in appendix B. The first and second cumulant after mutation are

$$\begin{aligned}\langle K_1\rangle_m &= \Gamma K_1\\ \langle K_2\rangle_m &= \Gamma^2 K_2 + L\left(1-\Gamma^2\right) \quad \text{where} \quad \Gamma = 1-2\gamma\end{aligned} \tag{32}$$

and $\langle\ldots\rangle_m$ represents the average over all ways of performing mutation.

The effect of mutation can be clearly seen in these expressions. It acts to push the ensemble distribution back to the maximum entropy state, decreasing the mean and increasing the variance. In this way it acts against selection which is reducing the variance and increasing the mean magnetization.

5.2 Crossover

Like mutation, the analysis of uniform crossover has been performed previously by Prügel-Bennett and Shapiro [10] [11]. Details of the derivation are included in appendix C. The effects of crossover on the first two cumulants are

$$\begin{aligned}\langle K_1\rangle_X &= K_1\\ \langle K_2\rangle_X &= \frac{K_2}{2} + \frac{L}{2}\left(1-q\right)\end{aligned} \tag{33}$$

where q is defined as

$$q = \frac{1}{P\left(P-1\right)}\sum_{\alpha\neq\beta}\frac{1}{L}\sum_{i=1}^{L} S_i^\alpha S_i^\beta. \tag{34}$$

We call q the correlation of the population, as it describes the similarity of strings in the population. In the initial random population, the correlation is zero. In a population consisting a P identical strings, correlation is equal to one.

The first cumulant does not change under crossover. This is expected as crossover conserves the states of all the spins. Crossover does change the second cumulant however and increases the variance towards the natural variance of the population which is defined by the value of q.

The analysis of higher cumulants show that they are greatly reduced by uniform crossover. It is this feature along with ranking selection which allows us to make the Gaussian approximation and accurately model the dynamics with just two cumulants.

To fully understand the effect of crossover, we must understand how correlation evolves. It is unaffected by crossover but changes through selection and mutation. As S_i^α and S_i^β are not independent, taking the average of q directly is not trivial.

In the original formalism of Prügel-Bennett and Shapiro, a maximum entropy inference was used to derive the correlation from the mean and variance of the distribution. However this methods underestimates the correlation as spins are not randomly distributed amongst population members. Selection acts by duplicating individuals and these duplicates significantly affect the correlation of the population. In this paper we pursue a similar approach to that developed by Rattray [12] and use a measure of the degree of duplication with the population. We then use a maximum entropy argument to combine this measure with the mean magnetization to calculate the correlation.

We introduce another macroscopic variable which describes the probability that the spins at the same site in two randomly drawn population members, originate from the same ancestor in the initial population. That is

$$C = \frac{1}{P\,(P-1)} \sum_{\alpha} \sum_{\beta \neq \alpha} \frac{1}{L} \sum_{i} \left[S_i^\alpha \sim S_i^\beta \right] \tag{35}$$

where

$$\left[S_i^\alpha \sim S_i^\beta \right] = \begin{cases} 1 & \text{if both spins come from the same ancestor} \\ 0 & \text{otherwise.} \end{cases} \tag{36}$$

In a sense we a labelling every spin in the initial population and keeping track of how these are distributed in the population at a later time. If the spins originate from the same ancestor, they will be identical and contribute $+1$ to the correlation. If they originate from different ancestors, we can calculate their contribution to the correlation by using a maximum entropy argument and assuming they will be distributed in proportion to the mean magnetization of the population. Thus we are able to describe the correlation of the population in terms of our duplication variable, C, and the mean magnetization

$$q = C + (1 - C)\,\frac{K_1^2}{L^2}. \tag{37}$$

In the initial population each spin is unique, $C = 0$, and q is defined by the mean magnetization of the population. If the population consists of identical population members, $C = 1$ and the population is perfectly correlated. Crossover will conserve the spins and thus have no effect on C. Selection will act to increase C by introducing multiple copies of individuals whilst mutation will act to reduce C by introducing new labels. To fully model crossover we must calculate the effect these operators have on C.

5.2.1 Selection

If an initial population has a particular value of C, its value after selection, C', will depend on the probability that a pair drawn from the population for crossover are identical. This is related to the number of copies introduced by selection and the probability that they are drawn from the population together. If n_α is the number of copies of population member α, C' is given by

$$C' = \sum_{\alpha=1}^{P} \left[\frac{n_\alpha}{P} \frac{n_\alpha - 1}{P - 1} + C \frac{n_\alpha}{P} \left(1 - \frac{n_\alpha - 1}{P - 1} \right) \right]. \tag{38}$$

Averaging over all n_α and using the fact that population size is constant and $\langle n \rangle = 1$ gives

$$1 - \langle C \rangle_s = \frac{P - \langle n^2 \rangle}{P - 1} (1 - C) \tag{39}$$

where $\langle n^2 \rangle$ is the variance of the selection scheme as calculated in section three. Indeed the factor here is exactly the same as the factor describing the loss in population variance after selection.

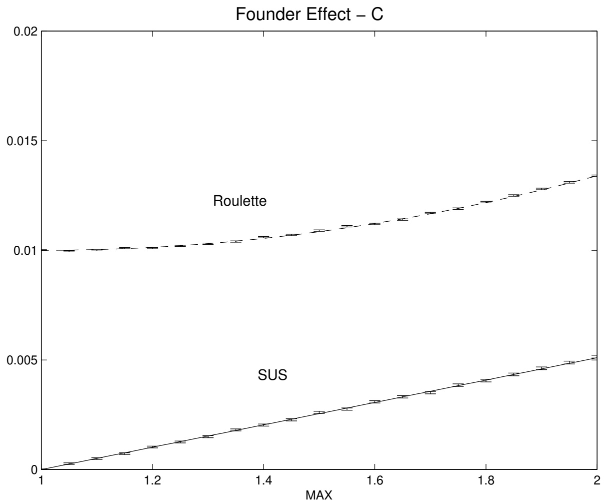


Figure 4. Comparison of theoretical and simulation results for C when roulette wheel selection (dashed line) and SUS (solid line) are used with ranking selection. Population size is 100.

Figure 4 shows simulation and theoretical results for C after one selection process. An initial population of 100 individuals were created with a random strings. Selection was applied using both SUS and roulette wheel selection and C calculated. Results were averaged over 1 000 runs. The difference between the selection schemes is clear. Roulette wheel selection shows a large increase in correlation even at very low selection strengths due to its propensity to produce more duplicates than SUS. Over time this correlation builds up within the population, decreasing the natural variance and thus reducing the effectiveness of crossover in restoring the population variance lost through selection. The GA using roulette wheel selection exhibits a smaller final population variance than one using SUS and thus searches a smaller area of the problem space.

5.2.2 Mutation

The value of C will also change with mutation, as it reduces the degree of duplication within the population. By considering the expected value of individual spins after mutation as derived in appendix A, $\langle S_i \rangle_m = \Gamma S_i$, and the definition of correlation given in equation 34, the correlation after mutation can be shown to be

$$\langle q \rangle_m = \Gamma^2 q. \tag{40}$$

Using the definition of q in equation 37 and the change in the first cumulant due to mutation, $\langle K_1 \rangle_m = \Gamma K_1$, C after mutation is given by

$$\langle C \rangle_m = \frac{\Gamma^2 \left(K_1^2 - L^2 \right)}{\Gamma^2 K_1^2 - L^2} C. \tag{41}$$

6. Modeling the Dynamics

The result of the analysis is a set of equations describing the change in each macroscopic variable with each operator — selection, mutation and crossover.

Selection

$$\langle K_1\rangle_s = K_1 + (\mathrm{MAX}-1)\sqrt{\frac{K_2}{\pi}}\mathrm{erf}\left(\frac{M_l - K_1}{\sqrt{K_2}}\right)$$

$$\begin{aligned}\langle K_2\rangle_s = r\Bigg[1 - &\frac{2\,(\mathrm{MAX}-1)}{\pi}\exp\left(-\frac{(M_l-K_1)^2}{K_2}\right)\\ &-\frac{(\mathrm{MAX}-1)^2}{\pi}\mathrm{erf}^2\left(\frac{M_l-K_1}{\sqrt{K_2}}\right)\Bigg]K_2\end{aligned}$$

$$1 - \langle C\rangle_s = r\,(1-C)$$

Mutation

$$\begin{aligned}\langle K_1\rangle_m &= \Gamma K_1\\ \langle K_2\rangle_m &= \Gamma^2 K_2 + L\left(1-\Gamma^2\right)\\ \langle C\rangle_m &= \frac{\Gamma^2\left(K_1^2 - L^2\right)}{\Gamma^2 K_1^2 - L^2}C\end{aligned}$$

Crossover

$$\begin{aligned}\langle K_1\rangle_x &= K_1\\ \langle K_2\rangle_x &= \frac{K_2}{2} + \frac{L}{2}\,(1-q)\end{aligned}$$

where

$$r = \frac{P - \langle n^2\rangle}{P-1} \quad , \quad \Gamma = 1-2\gamma \quad \text{and} \quad q = C + (1-C)\,\frac{K_1^2}{L^2}$$

By considering an initial population whose strings are assigned randomly ($K_1 = 0, K_2 = L$ and $C = 0$) we can iterate these equations to predict the dynamics of the GA. Figure 5 show the theory predictions compared to simulation results from repeated runs of a real GA using roulette wheel selection and SUS. The simulation data is averaged over 10 000 runs and uses the parameters $L = 48$, $\gamma = 1/48$, $P = 100$, MAX $= 1.4$ and $M_l = L/2$. The figures show very good agreement between theory and simulations.

The population evolves until the increase in mean magnetization due to selection is balanced by the decrease due to mutation. The loss in variance due to selection is balanced against the increase due to mutation and crossover. Selection by stochastic universal sampling leads to a final distribution with less correlation and higher population variance than roulette wheel. The mean of both distributions is the

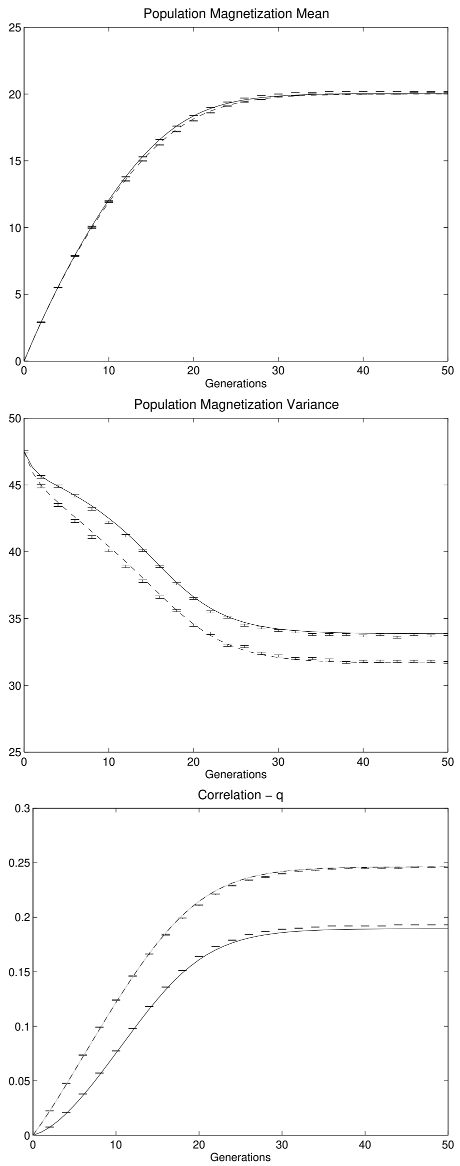


Figure 5. Comparison of theoretical and simulation results for the first two cumulants and correlation for a GA with ranking selection, mutation and crossover. Roulette wheel selection (dashed line) and SUS (solid line) are shown. Parameters used were $L = 48$, $\gamma = 1/48$, $P = 100$, $\mathrm{MAX} = 1.4$ and $M_l = L/2$.

same and is close to the minimum at M_l. Thus both distributions are centered around the same point but the larger variance in SUS means the GA is continually sampling through crossover a larger area of the problem space.

6.1 Equilibrium Distribution

To find the equilibrium distribution to which the population evolves was not possible for earlier models due to the dependence on higher order cumulants. The small set of expressions resulting from this model however, enable the equilibrium point to be solved by a set of simultaneous equations. It is not possible to do this exactly analytically but it is easy to do numerically. Figure 6 shows the simulation results and theoretical predictions for the first two cumulants of the equilibrium distribution against changing selection pressure. Simulation results are averaged over 1 000 runs.

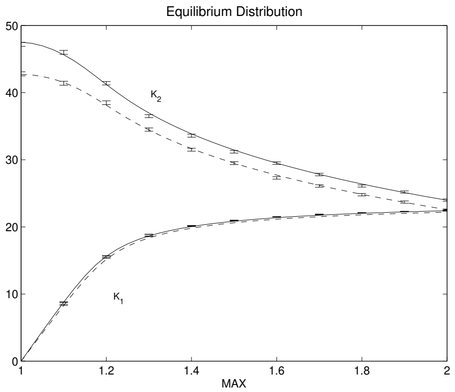


Figure 6. Comparison of theoretical numerical solutions and simulation results for equilibrium distribution. Roulette wheel selection (dashed line) and SUS (solid line) are shown. Parameters used were $L = 48$, $\gamma = 1/48$, $P = 100$ and $M_l = L/2$.

We can also consider the case when just the mutation operator is applied by omitting those macroscopic variables and expressions detailing crossover. Here the model fit is slightly poorer, as without crossover to suppress the higher cumulants, the distribution becomes skewed from a Gaussian. However when the distribution significantly overlaps the minimum, a Gaussian shape is restored and we obtain reasonable agreement without having to calculate higher order cumulant terms. Figure 7 shows the results for the same GA without the crossover operator. We see the same behavior in the mean whilst the variance of the distribution is greatly reduced. Without crossover acting to restore the

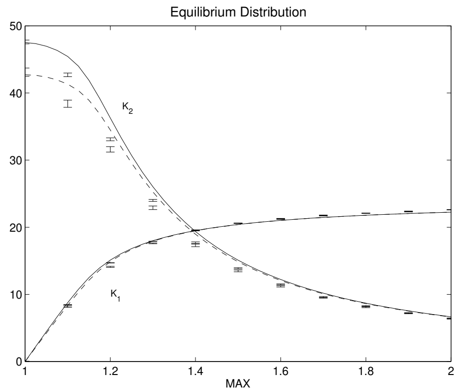


Figure 7. Comparison of theoretical numerical solutions and simulation results for equilibrium distribution without crossover. Roulette wheel selection (dashed line) and SUS (solid line) are shown. Parameters used were $L = 48$, $\gamma = 1/48$, $P = 100$ and $M_l = L/2$.

variance to its natural value, the population evolves very rapidly to a highly converged distribution and samples a very much smaller area of the problem space.

6.2 A Closed Form Solution

We can attempt an analytical solution by making the assumption that selection is strong enough for the population to approach the minimum and make the approximation to the error term in the selection expressions

$$\operatorname{erf}(x) \approx \frac{2x}{\sqrt{\pi}} \qquad \text{where} \quad x = \frac{M_l - K_1}{\sqrt{K_2}}. \tag{42}$$

The response of the first cumulant to selection simplifies significantly in this case and is no longer dependent on the variance of the distribution. We can then simply solve for the equilibrium value of K_1 to give

$$K_1^* \approx \frac{2\,(\text{MAX}-1)\,\Gamma M_l}{\pi\,(1-\Gamma) + 2\,(\text{MAX}-1)\,\Gamma}. \tag{43}$$

The lack of dependence of the equilibrium mean on the equilibrium variance seen in Figures 6 and 7 is explained. The equilibrium mean will be close to M_l. For small mutation rates, the distance from the minimum is independent of string length L and is of the same order as the number of spins per string which are expected to mutate at each generation.

$$M_l - K_1^* \approx \frac{\gamma\pi}{(\text{MAX}-1)} M_l. \tag{44}$$

The equilibrium correlation is determined by the balance of mutation and selection and the equilibrium mean. We can simply solve for the equilibrium value of C

$$C^* = \frac{x\,(1-r)}{1-xr} \tag{45}$$

where

$$x = \frac{\Gamma^2\left(K_1^{*2}-L^2\right)}{\Gamma^2 K_1^{*2}-L^2} \quad \text{and} \quad r = \frac{P-\langle n^2\rangle}{P-1}. \tag{46}$$

The equilibrium correlation is thus given by

$$q^* \approx C^* + (1-C^*)\,\frac{K_1^{*2}}{L^2}. \tag{47}$$

To find the equilibrium variance, we sacrifice some accuracy in order to derive a simple result. We again assume that the endpoint population distribution significantly overlaps the local minimum. In this case the reduction in variance due to selection approximates to

$$\langle K_2\rangle_s \approx r\left[1-\frac{2\,(\mathrm{MAX}-1)}{\pi}\right]K_2. \tag{48}$$

Thus the equilibrium variance is given by

$$K_2^* \approx \frac{L\left(1-\Gamma^2\right)+L\,(1-q^*)}{2-\Gamma^2 r\left[1-\frac{2(\mathrm{MAX}-1)}{\pi}\right]}. \tag{49}$$

For the case when no crossover is applied we may go straight to the result below using the same approximation

$$K_2^* \approx \frac{L\left(1-\Gamma^2\right)}{1-\Gamma^2 r\left[1-\frac{2(\mathrm{MAX}-1)}{\pi}\right]}. \tag{50}$$

For all but the largest mutation rates, $(1-\Gamma^2)$ is small compared to the correlation term $(1-q^*)$. Thus the correlation is significant in producing the larger equilibrium variance of the GA with crossover. For all but the smallest population sizes and lowest mutation rates, C^* is small and the correlation is thus defined by the equilibrium mean and thus the position of the minimum, M_l.

Figures 8 and 9 show the closed form results above plotted against the previously shown numerical solutions for the GA with and without crossover.

As expected we see reasonable agreement when selection is strong and the distribution is near the minimum. The accuracy is better for

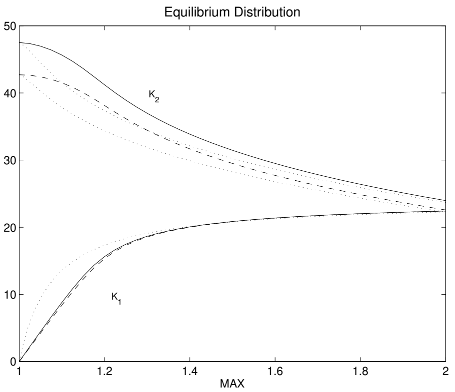


Figure 8. Comparison of numerical solution and closed form approximations for the equilibrium distribution of a GA with crossover. Curves are roulette wheel selection (dashed line), SUS (solid line) and closed form approximations (dotted line). GA parameters are as before.

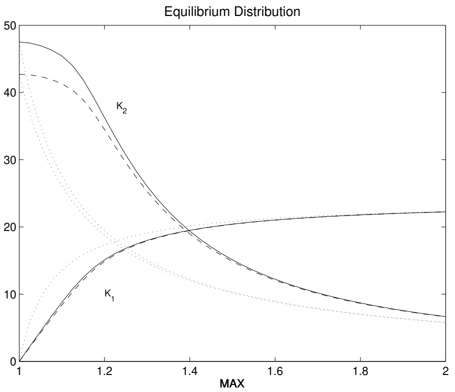


Figure 9. Comparison of numerical solution and closed form approximations for the equilibrium distribution of a GA without crossover. Curves are roulette wheel selection (dashed line), SUS (solid line) and closed form approximations (dotted line). GA parameters are as before.

the case with crossover as the larger distribution variance gives the distribution a significant overlap over the minimum which improves the accuracy of the approximations made in equations 42 and 48. The closed form results provide a clear insight into the factors determining the final shape of the distribution. In the next section we see that the equilibrium variance is of direct importance to how long it takes to move from the local to the global minimum in the 'Basin with a Barrier' problem.

7. Solving the Basin with a Barrier

From the analysis of the dynamics of a GA on a concave fitness potential, we can calculate the first passage time of the 'Basin with a Barrier' problem presented earlier. We compare the GA to a stochastic hill climber on a range of problem sizes — $L = 8$, 16, 32 and 48.

7.1 Modelling the GA

Having solved the dynamics of the GA on the local minimum, we can calculate the time for one population member to exceed the barrier — have magnetization greater than M_b — and fall into the global minimum. The fraction of the ensemble distribution less than the barrier at generation i is

$$p_i = \int_{-\infty}^{M_b} \rho\,(M)\,\mathrm{d}M. \tag{51}$$

Using the Gaussian approximation and describing the distribution in terms of its mean and variance allows us to write

$$p_i = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{M_b - K_1}{\sqrt{2K_2}}\right)\right]. \tag{52}$$

The probability that at least one member of the population of size P exists in the global minimum at each generation, is given by $1 - p_i^P$. Thus the expected number of function evaluations for one population member to exceed the barrier is given by

$$n = P\sum_{i=0}^{\infty}\left[i\left(1 - p_i^P\right)\prod_{j=0}^{i-1} p_j^P\right]. \tag{53}$$

Figure 10 shows the results of the model against simulation results on a logarithmic scale for a typical problem with $M_l = L/2$ and $M_b = 7L/8$. The positions of the barrier and local minimum were chosen to allow easy scaling over many string lengths. The GA is using SUS and crossover, and simulation results are averaged over 1 000 runs.

This simple analysis gives reasonable predictions of the first passage time. However we have only considered the mean and variance of the ensemble distribution. There will be fluctuations between runs and the mean time to solve the problem we calculate here takes no account of these fluctuations. A more detailed calculation of the size and effect of fluctuations has been performed by Prügel-Bennett [9] but is beyond the scope of this analysis.

When the time to solve the problem becomes large, the initial dynamics of the GA become less important and the mean and variance of the equilibrium distribution are most significant. At this limit we can

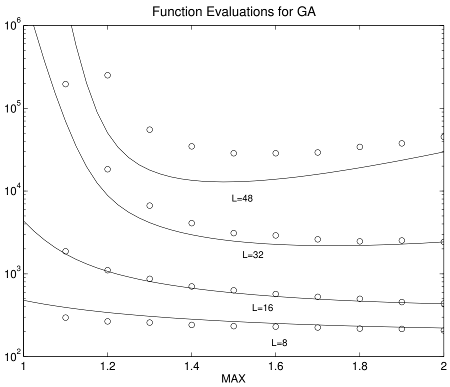


Figure 10. Comparison of theoretical and simulation first passage times for a GA with SUS to solve the 'Basin with a Barrier'. Simulations are averaged over 1 000 runs and the error bars are approximately the size of the symbols. Problem sizes are L = 8, 16, 32 and 48. GA parameters were $P = 100$ and $\gamma = 1/L$.

use a simple expansion of the error function — $\mathrm{erf}\,(x) \approx 1 - \mathrm{e}^{-x^2}/x\sqrt{\pi}$ — in equation 52 to give an analytical expression for the fraction of the distribution less than the barrier. The expansion above is valid when x is large which corresponds to the case where the barrier is some distance away from the center of magnetization distribution. The resulting expression is

$$p \approx 1 - \frac{\mathrm{e}^{-x^2/2}}{x\sqrt{2\pi}} \qquad \text{where} \qquad x = \frac{M_b - K_1^*}{\sqrt{K_2^*}}. \tag{54}$$

The probability of finding one member in the global minimum is again $1 - p^P$. As p is small, we can approximate this to $1 - Pp$ and thus the expected time in terms of function evaluations is given by $P/\left(1 - Pp\right)$. Using the above result gives

$$n \approx x\sqrt{2\pi}\mathrm{e}^{x^2/2}. \tag{55}$$

The most significant term here is the exponential. As shown earlier, the equilibrium mean magnetization, K_1^*, is constant over the cases considered - roulette wheel selection, SUS and the GA without mutation. The time to solve the problem is thus strongly dependent on the equilibrium variance

$$n \propto \mathrm{e}^{\frac{(M_b - K_1^*)^2}{2K_2^*}}. \tag{56}$$

For the case considered here, this results in a factor of approximately two in the performance through using SUS as opposed to roulette wheel selection. However the very small equilibrium variance which results when running the GA without crossover has a large influence in the performance. In the case when $L = 48$, the GA without crossover is predicted to perform many orders of magnitude worse. This is verified by experiment in that even after 10^8 function evaluations — the largest practical limit — the GA does not solve the problem.

7.2 Modelling a Single Stochastic Hill Climber

Modeling the dynamics of a single stochastic hill climber search method is straightforward. We consider that at each time step a new move is generated by allowing each spin to mutate with probability $1/L$ — equivalent to the GA mutation rate. Strictly speaking this allows global and not just local moves. However as L increases, the probability of this becomes small. We model a simple generic algorithm where steps which increase fitness are always accepted and steps which will decrease fitness are accepted with some probability, p. As there is only one barrier in this problem, there is no need to anneal this probability as is done in simulated annealing.

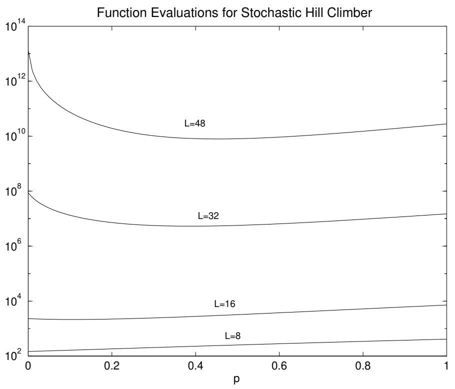


Figure 11. Theoretical first passage times for a single stochastic walker to solve the 'Basin with a Barrier'. Problem sizes are L = 8, 16, 32 and 48.

The transition times to reach the global minima are calculated directly from the transition matrix describing the probability of changing from one state to another. A randomly assigned starting string is assumed. Figure 11 shows the results for each problem size on a logarithmic scale. When $p = 0$ the hill climber performs simple steepest descent and sits at the local minima waiting for the correct mutation to jump straight into the global minimum. When $p = 1$ the hill climber performs a random walk and most time is spent in the maximum entropy area away from both the local and global minima. Between the two an optimum is reached.

As the problem size increases, the time to reach the global minimum increases rapidly. For problem sizes greater than L=16, the single walker is orders of magnitude slower at finding a solution in the global minimum than the time predicted and observed for the GA.

8. Discussion

By considering ranking selection we have been able to accurately calculate the effect of selection on any symmetrical concave function of unitation. The inclusion of finite population effects allows a framework by which the effect of roulette wheel selection and SUS may be compared analytically. The comparison confirms the empirically held belief that using SUS rather than roulette wheel selection reduces the effects of genetic drift.

Using and extending results of a previous analysis, we have been able to model the dynamics of the GA when subject to all three operators — selection, mutation and crossover. A comparison with simulation results shows very good agreement between theory and simulation.

This analysis of the dynamics allows us to look at the time to solve a model of a hard optimization problem — the 'Basin with a Barrier'. We can understand the dynamics in terms of the mean and variance of an ensemble distribution. To find solutions in the global minimum we are looking for population members in the extremes of the distribution. In terms of this search we can understand the effect of crossover and mutation operators.

In the framework of macroscopic variables, we have shown how crossover acts to restore variance to the population which is lost through selection. Without crossover the population magnetization variance is much reduced and thus the GA searches a much smaller area of the problem space.

Whilst in this framework it is natural to talk of crossover increasing the variance of the population, it is also worth considering what is happening at a microscopic level. When crossover operates on two strings which, as we would expect after a number of generations, are relatively correlated, the same spins will be present at a large number of sites. The spins which are common across the population will be those which have been shown by selection to contribute to the fitness of the individual. In a sense, individual sites act as building blocks in these fitness landscapes. Crossover will have no effect on these correlated sites. In the remaining sites however there remains much scope for crossover to combine two strings of near equal magnetization but different composition to produce a wide variety of magnetizations. This variety gives rise to the increased variance of the population after crossover and also to the search power of crossover, as it is one of these new strings which will be found to have leapt into the global minimum. By not disrupting what has already been selected for, crossover performs a non-local search centered around the current position of the population.

Unlike crossover, mutation however has no knowledge of which parts of the population members are shared by the majority of the population and thus disrupts all parts of the string. The entropy barrier which must be crossed by mutation to find a solution in the global minimum is thus much greater than that of crossover. It is this effect which makes it very difficult for mutation alone to produce an offspring which will jump directly into the global minimum. Added to this difficulty is the potential barrier which is seen by a stochastic hill climber. This makes it very unlikely that a series of mutations to produce a walk over the barrier into the global minimum. Both these effects make crossover the only viable mechanism for jumping the potential barrier into the global minimum.

Relating these effects to real problems and real search spaces is difficult. However we can use these insights gained from this work to gauge what is desirable in a crossover operator. For crossover to function as it does in this model, the fitness of the offspring produced must be correlated to the fitness of the parents. When this is not the case, crossover is essentially acting as a large scale mutation operator. If we believe the initial premise that good solutions are likely to be near one another in the search space, we expect this to be similar to random search and quite ineffectual. When chosen correctly, crossover will produce new population members biased towards areas of the problem space which selection has shown to yield fit individuals. By this this mechanism, it functions as the main search operator in the GA.

Acknowledgement

Since submitting this material, some related work has been brought to our attention by Jansen and Wegener of Dortmund University, Germany [5]. Using a similar fitness landscape known as JUMP but a different mathematical approach, they have shown crossover to be more efficient than single walkers at crossing these types of barriers.

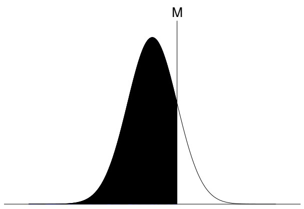
Appendix

A. Analysis of Selection

Performing the integral used in section 4.1 involves a number of standard integrals. We illustrate the calculations here by considering a unit Gaussian with zero mean and consider MAX to be 2. Extending the results to the more general cases given in section 4.1 simply involves a change of variable and some algebraic rearrangement. To shorten the algebra we introduce the notation

$$\Phi\left(M\right) = \int_{-\infty}^{M} DM \quad \text{where} \quad DM = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-M^2/2} \mathrm{d}M.$$

Considering first the one-max landscape, the weighting over the population is given by its position within the population and is thus proportional to the integral of the shaded area shown below.



Mathematically it is written as

$$E_M = 2\Phi\left(M\right).$$

Calculating the first moment after selection simply involves integrating the distribution over the weighting shown above. Calculating the first and second moments

$$\langle M \rangle = 2\int_{-\infty}^{\infty} M\Phi\left(M\right) DM = \frac{1}{\sqrt{\pi}}$$

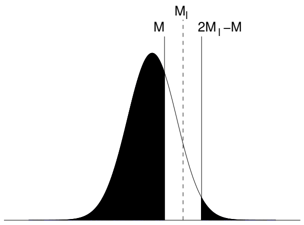
$$\langle M^2 \rangle = 2\int_{-\infty}^{\infty} M^2\Phi\left(M\right) DM = 1.$$

The first and second cumulant after selection are thus

$$K_1 = \langle M \rangle = \frac{1}{\sqrt{\pi}}$$

$$K_2 = \langle M^2 \rangle - \langle M \rangle^2 = 1 - \frac{1}{\pi}.$$

When we introduce the minimum in the potential function, the ranking of an individual now depends on its distance from the minimum. This feature makes these results applicable to any symmetrical function.



Mathematically it is more complicated than the previous case

$$E_M = \begin{cases} 2 - 2\left[\Phi\left(2M_l - M\right) - \Phi\left(M\right)\right] & \text{when } M \leq M_l \\ 2 - 2\left[\Phi\left(M\right) - \Phi\left(2M_l - M\right)\right] & \text{when } M \geq M_l \end{cases}$$

Calculating the first and second moments

$$\begin{aligned} \langle M \rangle &= 2\int_{-\infty}^{\infty} M DM - 2\int_{-\infty}^{m} M\left[\Phi\left(2M_l - M\right) - \Phi\left(M\right)\right] DM \\ &\quad - 2\int_{m}^{\infty} M\left[\Phi\left(M\right) - \Phi\left(2M_l - M\right)\right] DM \\ &= \frac{\operatorname{erf}\left(M_l\right)}{\sqrt{\pi}} \end{aligned}$$

$$\begin{aligned} \langle M^2 \rangle &= 2\int_{-\infty}^{\infty} M^2 DM - 2\int_{-\infty}^{m} M^2\left[\Phi\left(2M_l - M\right) - \Phi\left(M\right)\right] DM \\ &\quad - 2\int_{m}^{\infty} M^2\left[\Phi\left(M\right) - \Phi\left(2M_l - M\right)\right] DM \\ &= 1 - \frac{2\mathrm{e}^{-M_l^2}}{\pi}. \end{aligned}$$

Here the term $\operatorname{erf}\left(M_l\right)$ refers to the standard error function which can be translated into our notation as

$$\operatorname{erf}\left(M_l\right) = 2\Phi\left(M_l\sqrt{2}\right) - 1$$

The first and second cumulant after selection are thus

$$\begin{aligned} K_1 &= \langle M \rangle = \frac{\operatorname{erf}\left(M_l\right)}{\sqrt{\pi}} \\ K_2 &= \langle M^2 \rangle - \langle M \rangle^2 = 1 - \frac{\operatorname{erf}^2\left(M_l\right)}{\pi} - \frac{2\mathrm{e}^{-M_l^2}}{\pi}. \end{aligned}$$

B. Analysis of Mutation

When the mutation operator is applied, each spin of population member μ has a small probability of mutation

$$S_i^\mu \to -S_i^\mu \quad \text{with probability } \gamma.$$

We model the effect of mutation on the ensemble magnetization distribution by first considering the effect on any individual spin. The expected value of any spin after mutation is easily shown to be

$$\langle S_i^\mu \rangle_m = \Gamma S_i^\mu \quad \text{where} \quad \Gamma = \left(1 - 2\gamma\right).$$

and $\langle \ldots \rangle_m$ represents the average over all mutations. Applying this to the expected magnetization of individual μ after mutation gives

$$\begin{aligned}\langle M_\mu \rangle_m &= \sum_{i=1}^{L} \langle S_i \rangle_m \\ &= \Gamma M_\mu.\end{aligned}$$

For second order terms the expression is more complicated as dependent terms must be collected together

$$\begin{aligned}\langle M_\mu^2 \rangle_m &= \sum_{i \neq j} \langle S_i \rangle_m \langle S_j \rangle_m + \sum_{i=1}^{L} \langle S_i^2 \rangle_m \\ &= \Gamma^2 \left(M_\mu^2 - L \right) + L \\ &= \Gamma^2 M_\mu^2 + L \left(1 - \Gamma^2 \right).\end{aligned}$$

Applying these expressions to the definitions of the cumulants and averaging over the ensemble gives

$$\begin{aligned}\langle K_1 \rangle_m &= \Gamma K_1 \\ \langle K_2 \rangle_m &= \Gamma^2 K_2 + L \left(1 - \Gamma^2 \right).\end{aligned}$$

C. Analysis of Uniform Crossover

When we apply uniform crossover [15] the spins of offspring μ are drawn from each parent, α and β, at random

$$S_i^\mu = \chi_i S_i^\alpha + (1 - \chi_i) S_i^\beta.$$

where

$$\chi_i^\mu = \begin{cases} 1 & \text{with probability } 1/2 \\ 0 & \text{with probability } 1/2. \end{cases}$$

We calculate the effect by considering each spin. The expected value of any spin from offspring μ averaged over all ways of drawing spins from each parent is simply

$$\langle S_i^\mu \rangle_x = \frac{S_i^\alpha}{2} + \frac{S_i^\beta}{2}.$$

Clearly then the expected magnetization of offspring μ produced through crossover is

$$\langle M_\mu \rangle_x = \frac{M_\alpha}{2} + \frac{M_\beta}{2}.$$

As α and β are drawn independently from the population, we may simply average over the ensemble to give the mean ensemble magnetization after crossover

$$\langle M \rangle_x = \langle M \rangle.$$

For second order terms the analysis is slightly more complicated. If we again consider offspring μ then

$$\begin{aligned}
\langle M_\mu^2 \rangle_x &= \sum_{i \neq j} \langle S_i \rangle_x \langle S_j \rangle_x + \sum_{i=1}^{L} \langle S_i^2 \rangle_x \\
&= \sum_{i \neq j} \langle S_i \rangle_x \langle S_j \rangle_x + L \\
&= \left(\sum_{i=1}^{L} \langle S_i \rangle_x \right)^2 - \sum_{i=1}^{L} \langle S_i \rangle_x^2 + L.
\end{aligned}$$

Now considering all ways of drawing the parents and the preceeding results gives

$$\begin{aligned}
\langle M_\mu^2 \rangle_x &= \left(\frac{M_\alpha}{2} + \frac{M_\beta}{2} \right)^2 - \sum_{i=1}^{L} \left(\frac{S_i^\alpha}{2} + \frac{S_i^\beta}{2} \right)^2 + L \\
&= \left(\frac{M_\alpha^2}{4} + \frac{M_\beta^2}{4} + \frac{M_\alpha M_\beta}{2} \right) - \sum_{i=1}^{L} \left(\frac{1}{2} + \frac{S_i^\alpha S_i^\beta}{2} \right) + L.
\end{aligned}$$

Averaging over all ways of drawing α and β independently from the population gives the second moment of the ensemble magnetization

$$\langle M^2 \rangle_x = \frac{\langle M^2 \rangle}{2} + \frac{\langle M \rangle^2}{2} + \frac{L}{2}(1-q)$$

where q is defined as

$$q = \frac{1}{P(P-1)} \sum_{\alpha \neq \beta} \frac{1}{L} \sum_{i=1}^{L} S_i^\alpha S_i^\beta.$$

The cumulant terms are simply found from the ensemble moments derived above

$$\begin{aligned}
\langle K_1 \rangle_x &= K_1 \\
\langle K_2 \rangle_x &= \frac{K_2}{2} + \frac{L}{2}(1-q).
\end{aligned}$$

References

[1] J. Baker. Adaptive selection methods for genetic algorithms. In J.J. Grefenstette, editor, *Proceedings of the First International Conference on Genetic Algorithms*, pages 101–111. Erlbaum, 1985.

[2] J. Baker. Reducing bias and inefficiency in the selection algorithm. In J.J. Grefenstette, editor, *Proceedings of the Second International Conference on Genetic Algorithms.* Erlbaum, 1987.

[3] T. Blickle and L. Theile. A mathematical analysis of tournament selection. In L. Eshelman, editor, *Proceedings of the Sixth International Conference on Genetic Algorithms*, pages 9–16. Erlbaum, 1995.

[4] Y. Davidor. Epistasis variance: A viewpoint on ga-hardness. In G.J.E. Rawlins, editor, *Foundations of Genetic Algorithms*, pages 23–35. Morgan Kaufmann, 1991.

[5] T. Jansen and I. Wegener. On the analysis of evolutionary algorithms — a proof that crossover really can help. In Jaroslav Nešetřil, editor, *Proceedings of the 7th Annual European Symposium on Algorithms (ESA '99)*, volume 1643 of *Lecture Notes in Computer Science*, pages 184–193, Berlin, Germany, 1999. Springer.

[6] T. Jones and S. Forrest. Fitness distance correlation as a measure of problem difficulty in genetic algorithms. In L. Eshelman, editor, *Proceedings of the Sixth International Conference on Genetic Algorithms*, pages 184–192. Morgan Kaufmann, 1995.

[7] M. Mitchell, J. Holland, and S. Forrest. When Will a Genetic Algorithm Outperform Hill Climbing? In J. Cowan, G. Tesauro, and J. Alspector, editors, *Advances in Neural Information Processing Systems*, pages 51–58, San Francisco, CA., 1994. Morgan Kauffman.

[8] A Nix and M. D. Vose. Modeling genetic algorithms with markov chains. *Annals of Mathematics and Artificial Intelligence*, 5:79–88, 1991.

[9] A. Prügel-Bennett. Modelling Evolving Populations. *J. Theor. Biol.*, 185:81–95, 1997.

[10] A. Prügel-Bennett and J. L. Shapiro. An Analysis of Genetic Algorithms Using Statistical Mechanics. *Phys. Rev. Lett.*, 72(9):1305–1309, 1994.

[11] A. Prügel-Bennett and J. L. Shapiro. The Dynamics of a Genetic Algorithm for Simple Random Ising Systems. *Physica D*, 104:75–114, 1997.

[12] M. Rattray. The Dynamics of a Genetic Algorithm under Stabilizing Selection. *Complex Systems*, 9(3):213–234, 1995.

[13] A. Rogers and A. Prügel-Bennett. Genetic Drift in Genetic Algorithm Selection Schemes. *IEEE Transactions on Evolutionary Computation*, 3(4), 1999.

[14] J. L. Shapiro and A. Prügel-Bennett. Genetic Algorithms Dynamics in Two-Well Potentials with Basins and Barriers. In R. K. Belew and M. D. Vose, editors, *Foundations of Genetic Algorithms 4*, pages 101–139, California, 1997. Morgan Kaufmann.

[15] G. Syswerda. Uniform crossover in genetic algorithms. In *Proceedings of the Third International Conference on Genetic Algorithms*. Morgan Kaufmann (San Mateo), 1989.

[16] M. D. Vose. Modelling simple genetic algorithms. In D. Whitley, editor, *Foundations of Genetic Algorithms 2*. Morgan Kaufmann (San Mateo, CA), 1992.