A Solvable Model Of A Hard Optimisation Problem

A. Rogers and A. Prügel-Bennett

October 15, 1999

Abstract

The dynamics of a genetic algorithm (GA) on a model of a hard optimisation problem are analysed using a formalism which describes the changing fitness distribution of the GA population under ranking selection, uniform crossover and mutation. The time to solve the optimisation problem is calculated in a closed form expression which enables the effect of the various GA parameters — population size, mutation rate and selection scheme — to be understood.

1 Introduction

Much of the theoretical analysis of genetic algorithms has necessarily been performed on simple simple fitness landscapes such as one-max [4, 12, 6, 7] and the royal road functions of Mitchell, Holland and Forest [3]. Whilst giving some insight into the behavior of GA, they are not representative of the types of problems to which GA are often applied. GA are shown to perform poorly on these problems compared to other techniques such as simple hill climbing and little knowledge regarding the setting of the various GA parameters is gained.

The simple GA models studied usually omit the crossover operator in their analysis. Empirical comparisons however suggest that the performance of a GA is qualitatively different with and without crossover [5]. Thus drawing general conclusions from the analysis is often misleading.

In this paper we analyse a previously published model of a hard optimisation problem [10] — the 'Basin with a Barrier' fitness landscape. This model has some of the features of a real world problem whilst still being amenable to analysis. It is one of the few test problems where a GA is shown to have a distinct advantage over other conventional search methods such as simulated annealing [8].

The dynamics of a generational GA using ranking selection [1], uniform crossover [11] and mutation are calculated accurately on this model by describing the effect that each genetic operator has on several macroscopic variables which describe the population distribution. Closed form expressions can be derived for the end point of evolution and the number of function evaluations required to solve the problem. In this paper, we use these results to study the effect that the various GA parameters have on how the problem space is searched and show how the optimal mutation rate is obtained. It is hoped that by studying this problem, some insight into real world problems is gained.

2 The Basin with a Barrier

Discussions of what characterises a hard optimisation problem often result in a debate regarding the validity of certain measures of problem difficulty. Despite this debate, 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 fitness barrier from even fitter 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.

The 'Basin with a Barrier' attempts to model these types of landscapes [10]. It has some of the features of a hard optimisation 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 'magnetisation', M, of the string

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

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

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

In describing the landscape we must also consider the entropy of the system, S. Entropy describes the number of possible states which the system may occupy. Clearly there are only two states with the maximum or minimum magnetisation whilst the majority of possible states represent magnetisations somewhere in between.

The entropy of the system, 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.

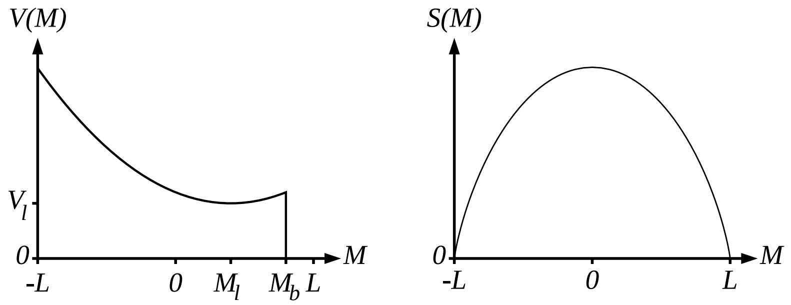

Figure 1: Diagram of potential and entropy for the 'Basin with a Barrier' fitness landscape.

We propose that whilst being a toy problem, this model holds some of the features seen in combinatorial optimisation problems. 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 one another in problem space but they may be separated by non-local moves.

3 Modelling the GA

The GA we consider is a finite population generational GA using ranking selection [1] with either roulette wheel or stochastic universal sampling (SUS) [2]. Mutation is applied with a fixed small probability that each spin mutates at each generation and uniform crossover [11] is applied between pairs.

The technique we use to model the GA is that developed by Prügel-Bennett , Shapiro and Rattray. Several macroscopic variables are used to describe the magnetisation distribution of the population. In this analysis, we make a

Gaussian approximation to the distribution and describe it by the first two cumulants — the mean of the distribution, K_1, and the variance of the distribution, K_2 — and a measure of the degree of replication of individuals within the population, C. By combining these variables, we can also calculate the correlation of the population, q. These parameters describe the ensemble average of many finite populations and thus predict the mean behavior of a finite population.

We make a number of approximations as we wish to capture the full dynamics of the GA but not at the cost of an over complex model. The complete analysis including the derivation of the results used here and justification underlining the approximations made are published elsewhere [8]. In this paper we use the results of this analysis and use them to study the effect of changing the various GA parameters.

As each genetic operator — selection, mutation and crossover — acts on the population, we obtain a set of coupled equations which describe the changing state of the macroscopic variables — K_1, K_2 and C — in terms of the GA parameters — the selection strength, MAX, the mutation rate, γ, the population size, P, the string length, L and a measure of the rate of genetic drift inherent in the selection scheme, r. By iterating these, we can describe the full dynamics of the GA.

Selection

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

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

Mutation

$$
\begin{aligned}
K_1 &\to \Gamma K_1 \\
K_2 &\to \Gamma^2 K_2 + L\left(1-\Gamma^2\right) \\
C &\to \frac{\Gamma^2\left(K_1^2 - L^2\right)}{\Gamma^2 K_1^2 - L^2} C \quad \text{where} \quad \Gamma = 1 - 2\gamma
\end{aligned}
$$

Crossover

$$
\begin{aligned}
K_1 &\to K_1 \\
K_2 &\to \frac{K_2}{2} + \frac{L}{2}(1-q) \quad \text{where} \quad q = C + (1-C)\frac{K_1^2}{L^2}
\end{aligned} \tag{3}
$$

Besides giving an accurate description of the dynamics, these expressions provide an intuitive idea of how the GA is searching the fitness landscape.

- Selection acts to focus the population onto areas of improved fitness and thus increases the mean magnetisation. In doing so however it reduces the variance of the population leading to a smaller area of the landscape being sampled and increasing the correlation of the population.

- Mutation will generate new population members around the selected area but will act to push the population back to the maximum entropy state thereby increasing the variance and decreasing the correlation and mean.

- Crossover does not effect the mean but forces the variance towards a *natural* value defined by the correlation of the population, q. This acts to restore variance to the population lost through selection.

By considering an initial population whose spins 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 2 show the theory predictions compared to simulation results from repeated runs of a real GA using roulette wheel selection and stochastic universal sampling. The simulation data is averaged over 10 000 runs and uses the parameters $L = 48$, $\gamma = 1/48$, $P = 100$, $\mathrm{MAX} = 1.4$ and $M_l = L/2$. The figures show very good agreement between theory and simulations.

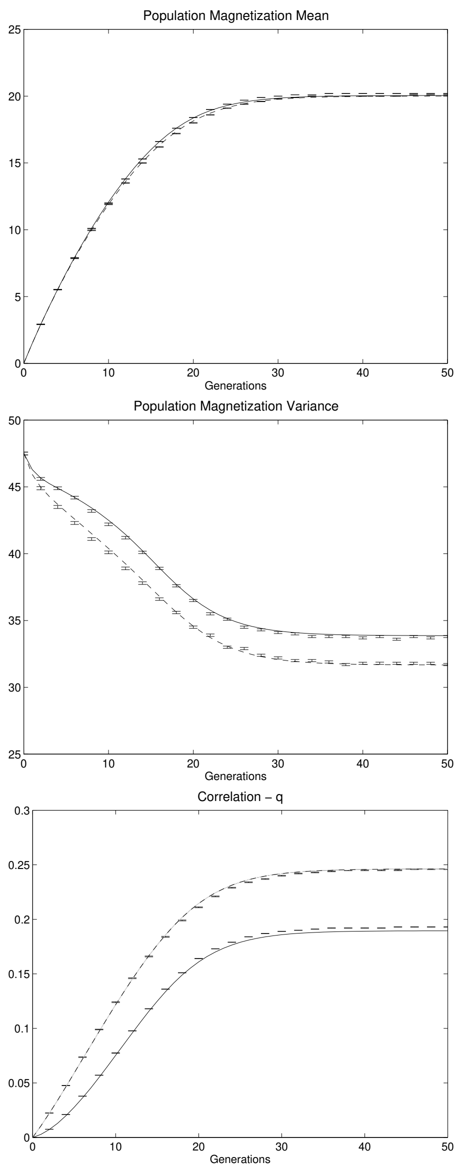

Figure 2: Comparison of theoretical and simulation results for the first two cumulants and correlation for a GA with ranking selection, mutation and crossover. Roulette wheel selection (dashed line) and SUS (solid line) are shown. Parameters used were $L = 48$, $\gamma = 1/48$, $P = 100$, $\mathrm{MAX}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

4 Equilibrium Point

The conflicting forces of selection, mutation and crossover result in the population evolving to an equilibrium state where the magnetisation mean, variance and correlation no longer change from generation to generation. Although at the macroscopic level the population is in equilibrium, the GA is continually generating new population members and sampling the area of problem space described by the macroscopic variables. We can solve these expressions as a set of simultaneous equations by simply making the demand that after all the three operator — selection, mutation and crossover — the macroscopic variables remain unchanged. Doing this numerically is easy but in order to derive closed form expressions for the equilibrium point, several approximations are made. In the case when the equilibrium distribution has a significant overlap with the local minimum, a simple approximation can be made for the error term [1]in the selection term

$$\mathrm{erf}\,(x) \approx x \qquad \text{where} \quad x = \frac{M_l - K_1}{\sqrt{K_2}}.$$

The response of the magnetisation mean 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

$$K_1^* \approx \frac{(\mathrm{MAX}-1)\,\Gamma M_l}{\sqrt{\pi}\,(1-\Gamma)+(\mathrm{MAX}-1)\,\Gamma}. \tag{4}$$

This approximation holds in most cases except when the mutation rate is so high and the selection strength so weak that the population does not move far from the initial position. 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

$$C^* = \frac{x\,(1-r)}{1-xr} \tag{5}$$

where

$$x = \frac{\Gamma^2\left(K_1^{*2}-L^2\right)}{\Gamma^2 K_1^{*2}-L^2}. \tag{6}$$

The equilibrium correlation is thus given by

$$q^* \approx C^* + (1-C^*)\,\frac{K_1^{*2}}{L^2}. \tag{7}$$

[1]Where erf(x) represents the standard error function.

To find the equilibrium variance, we again assume the equilibrium distribution overlaps the local minimum and when can approximate the response to selection of the variance given in Eq. 3 as

$$K_2 \to r \left[1 - \frac{2\left(\mathrm{MAX} - 1\right)}{\pi}\right] K_2 \tag{8}$$

Solving the remaining expressions as a set of simultaneous equations thus gives the equilibrium variance as

$$K_2^* \approx \frac{L\left(1-\Gamma^2\right) + L\left(1-q^*\right)}{2 - \Gamma^2 r \left[1 - \frac{2(\mathrm{MAX}-1)}{\pi}\right]}. \tag{9}$$

5 First Passage Time

The picture which emerges from this analysis is of a population evolving towards the local minimum and reaching an equilibrium point close to it. How close it is to the minimum depends on the mutation rate and selection strength. The variance of the population at this equilibrium point depends on the various GA parameters in a rather complicated way.

We are interested in how long it takes for mutation and crossover to generate an individual which is at the extremes of the distribution and is in the global minimum. Clearly this will be related to the population size and mean and variance of the population magnetisation distribution at its equilibrium point.

As we are describing the magnetisation distribution as a Gaussian, the probability of finding any one population member with a magnetisation less than the barrier, M_b, is simply given by

$$p = \Phi\left(x\right) \qquad \text{where} \qquad x = \frac{M_b - K_1^*}{\sqrt{K_2^*}} \tag{10}$$

and $\Phi\left(x\right)$ represents integration of a unit Gaussian from negative infinity to x. We can approximate this expression to

$$p \approx 1 - \frac{\mathrm{e}^{-x^2/2}}{x\sqrt{2\pi}}. \tag{11}$$

The probability of finding one member above the barrier and thus in the global minimum in any generation is $1 - p^P$ and since p is small this can be approximated to $1 - Pp$. Thus the expected time in terms of function evaluations, n, is given by $P/(1 - Pp)$. Using the above result gives

$$n \approx x\sqrt{2\pi}\mathrm{e}^{x^2/2}. \tag{12}$$

The most significant factor here is the exponential dependence on the equilibrium mean and variance

$$n \propto \mathrm{e}^{\frac{(M_l - K_1)^2}{2K_2}}. \tag{13}$$

The second interesting point is the absence of any population size dependence. Once the population has reached the equilibrium point, the number of evaluations required to find a solution is independent of the population size.

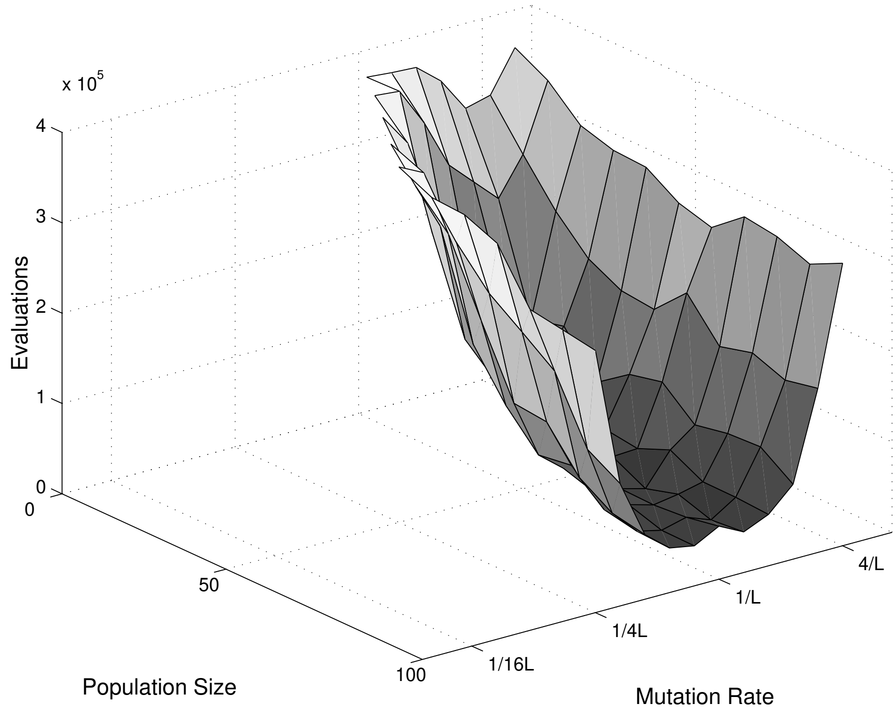

Figure 3: Simulation results to solve 'Basin with a Barrier' using a GA with stochastic universal sampling, mutation and uniform crossover. Parameters used were MAX $= 2$, $L = 48$, $M_l = L/2$ and $M_b = 7L/8$.

6 Simulation Results

The time to solve a typical size 'Basin with a Barrier' problem was found for a range of population sizes and mutation rates. The results are shown in figure 3 and are the results of averaging over 100 runs.

The results show a very clear optimal mutation rate with some population size dependence particularly when the population is small.

7 Theoretical Analysis

Having derived expressions for the time to solve the problem and understanding how each of the genetic operators influence the evolution of the population, we can consider what effect GA parameters such as mutation rate and population size have on search. Figure 4 shows the results of solving the analytical solutions for the time to solve the problem simulated in figure 3. The theoretical analysis very accurately predicts the time required and shows the same influence of mutation rate and population size.

7.1 Population Size

As seen in the analysis of the first passage time, the size of the population does not enter our expressions in Eq. 12 for the time required for our population in equilibrium to produce a solution in the global minimum. The population size does however affect this time indirectly by changing the final end point distribution of the population. With small populations the stochastic nature of the selection operator becomes significant and must be accounted for to accurately describe the dynamics and end point distribution.

Population size enters our model as an extra factor, r, caused by the stochastic nature of the selection scheme. It is less than one and tends to one as the population size goes to infinity. It causes a further reduction in population variance than that predicted for an infinite population and also occurs as a factor increasing population correlation at each selection step. A small population thus experiences a faster reduction in population variance and reaches an equilibrium magnetisation with less variance. This clearly has a strong effect on the time to solve the problem.

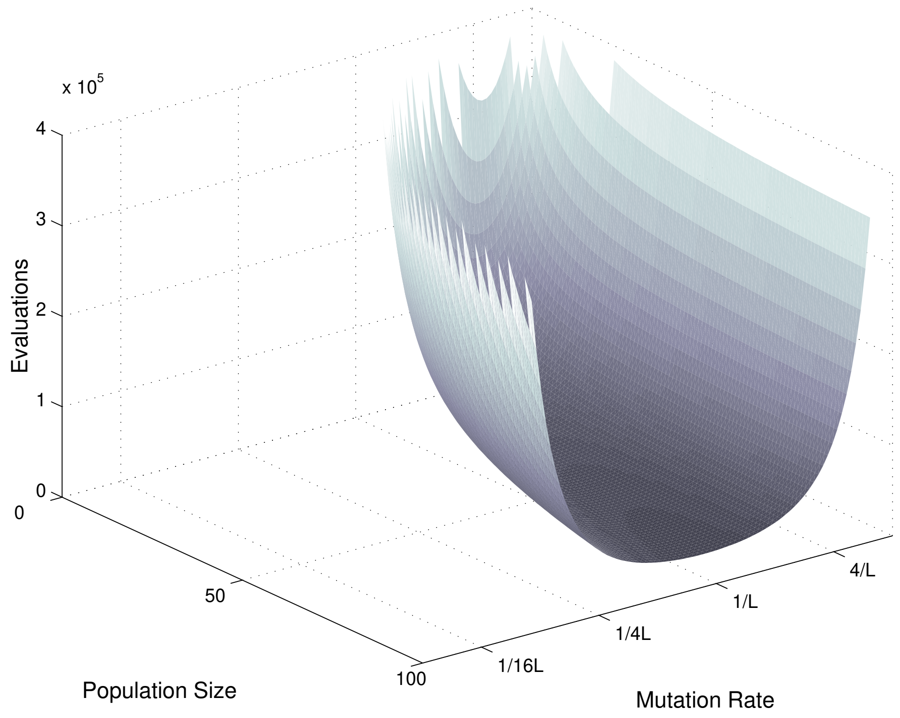

Figure 4: Theoretical results to solve 'Basin with a Barrier' using a GA with stochastic universal sampling, mutation and uniform crossover. Parameters used were MAX $= 2$, $L = 48$, $M_l = L/2$ and $M_b = 7L/8$.

What signifies a small population can be found by calculating how r changes with population size. The factor is given by

$$r = \frac{P - \langle n^2 \rangle}{P - 1}. \tag{14}$$

The term $\langle n^2 \rangle$ is the expected square of the number of times any population member is selected and is a function of the population size, selection strength and the selection scheme used — roulette wheel or stochastic universal sampling. It has been calculated in another paper [8] and has been used by the authors in a comparison of various selection schemes [9]. Figure 5 shows the results of these against population size.

Clearly the curve approaches unity — the infinite population limit — very quickly. For small population sizes however, the deviation from unity is large. This feature explains the relative lack of dependence of population

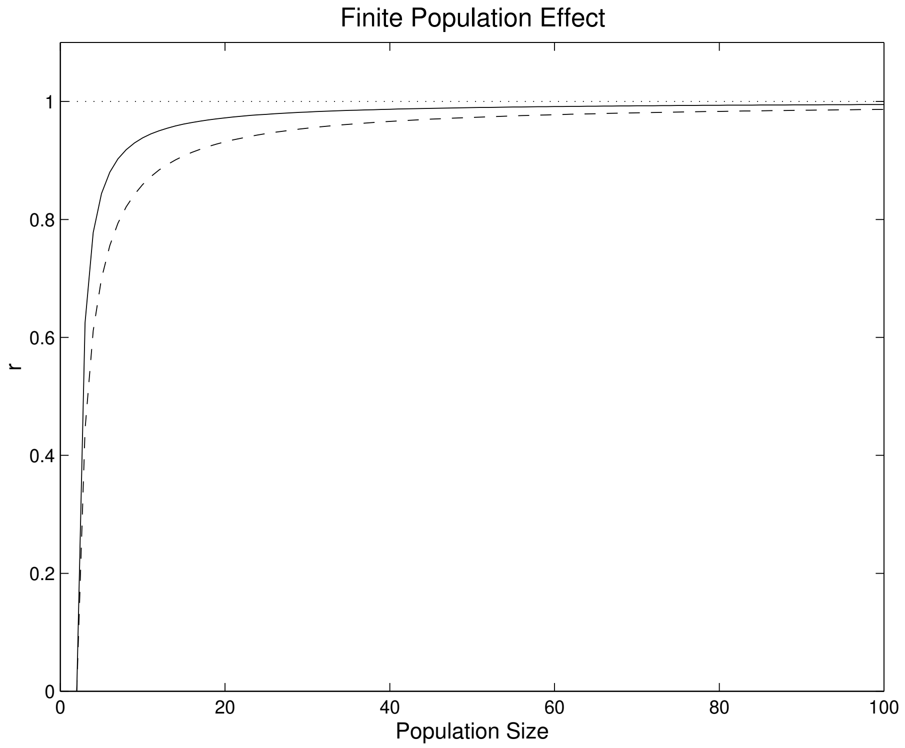

Figure 5: Finite population effect for roulette wheel (dashed line) and stochastic universal sampling (solid line).

size on the time to solve the problem for large population sizes and the very rapid decline in performance as the population size becomes small.

In this analysis, we have assumed cases where the initial dynamics are comparatively short in comparison to the time spent at equilibrium, whilst waiting for the mutation and crossover operators to generate a solution in the global minimum. If we consider a more realistic problem consisting of a cascade of barriers, it is clear that this initial dynamic phase favors a smaller population as it requires less function evaluations to move the population to its new equilibrium point.

This suggests an optimum population size which is a balance between the need to maintain the speed which the population can move around the landscape whilst not be so small that the area of the landscape being searched is significantly reduced by finite population effects.

7.2 Selection Scheme

Stochastic universal sampling [2] was suggested by Baker as an alternative to roulette wheel selection. Baker noted that whilst on average any individual is expected to be selected a certain number of times depending on its fitness within the population, 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 — and what we are calculating in this formalism when we calculate r.

Ranking selection allows us to calculate r for both stochastic universal sampling and roulette wheel selection [8]. In figure 5 both selection schemes are plotted. For roulette wheel selection, the finite population effects are much larger over the entire range of population sizes.

In figure 2 showing the evolving dynamics of the GA, stochastic universal sampling reaches an equilibrium distribution with greater variance and less correlation than the equivalent roulette wheel case. This is a direct cause of the reduction in finite population effects achieved through using stochastic universal sampling. In the test problem shown in this paper, this difference in variance leads to a halving of the number of function evaluations required to solve the problem.

7.3 Mutation Rate

Perhaps the most significant feature of the surfaces plotted is the strong dependence on mutation rate, with extremely poor performance outside the optimal range. Understanding this feature involves the interplay of all the effects previously discussed.

Mutation has been shown to increase the variance of the final population equilibrium distribution but also move the mean of the distribution away from the global minimum back towards the maximum entropy state. The second effect is most significant and mutation has a detrimental effect on performance. Increasing mutation rate increases the number of function evaluations required to solve the problem.

In an infinite population, the optimum mutation rate would thus be zero. However, we are dealing with finite populations and must consider the correlation of the population caused by selection. With no mutation, the correlation of the population will increase very rapidly limiting crossovers ability to restore variance to the population. This will result in a very small equilibrium variance which searches a very small area of the problem space and thus takes a very long time to reach the global minimum.

A balance is achieved when mutation is large enough to prevent the correlation of the population but not so large as to disrupt the search.

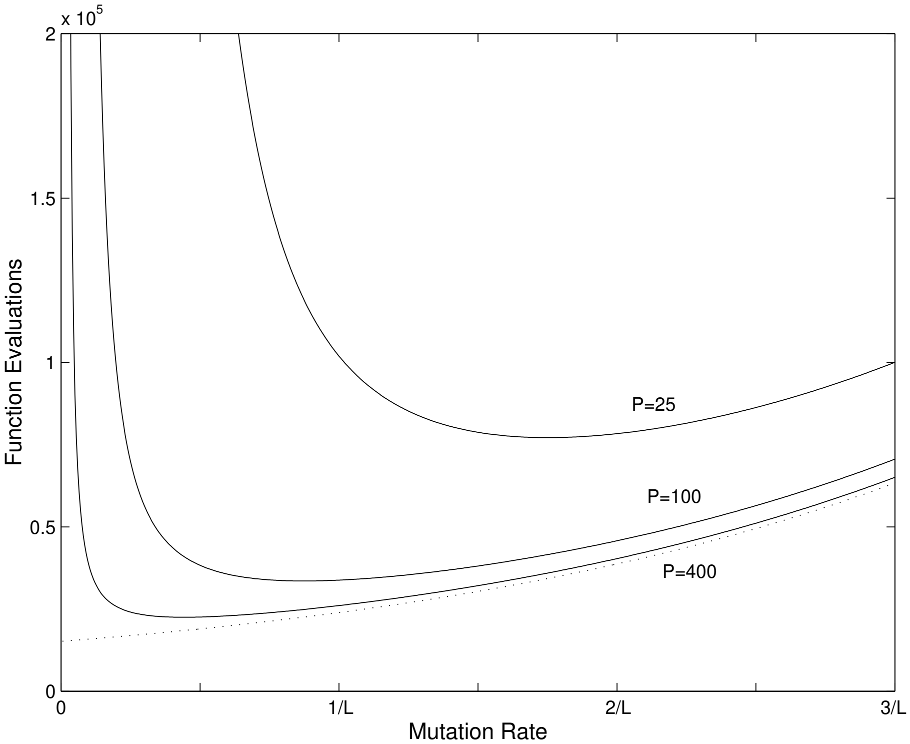

Figure 6: Theoretical results to solve 'Basin with a Barrier' with varying mutation rate at three different population sizes. $P = 25, 100$ and 400. The dotted line is the infinite population response. Parameters used were $\mathrm{MAX} = 2$, $L = 48$, $M_l = L/2$ and $M_b = 7L/8$.

Figure 6 shows theoretical results for the time to solve the 'Basin with a Barrier' with varying mutations rates. Three different population sizes are shown along with the infinite population case as a dotted line. Clearly as P increases, the finite population effects decrease and the finite populations approach the infinite case. However as the mutation rate becomes small, the correlation of the population through finite population effects becomes significant.

The optimum mutation rate is seen to be dependent on the population size. A large population suffers less correlation due to finite population effects and thus will not suffer the same correlation until the mutation rate is very small. For this particular problem we find that the optimum mutation rate is approximately proportional to $1/\sqrt{P}$.

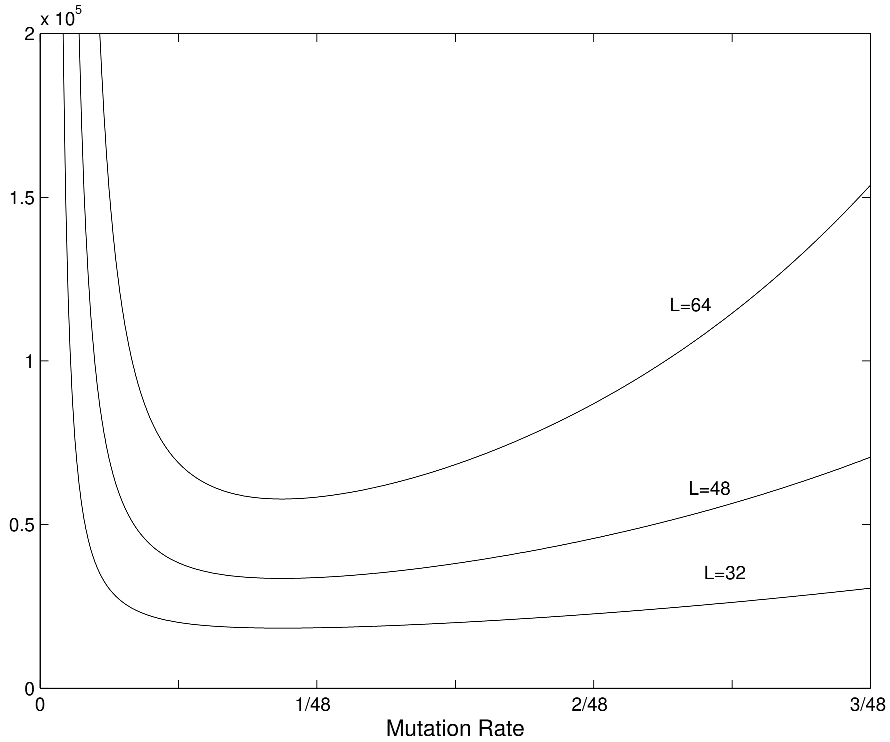

Figure 7: Theoretical results to solve 'Basin with a Barrier' with varying mutation rate at three string lengths — $L = 32, 48$ and 64. Vertical scales for $L = 32$ and $L = 64$ are multiplied and divided by ten respectively to enable easy comparison. Parameters used were MAX $= 2$, $P = 100$, $M_l = L/2$ and $M_b = 7L/8$.

As the effects discussed here are independent of the string length, we find that the optimum mutation rate is also independent of the string length. Figure 7 shows the results of varying the length of the string for various mutation rates. As expected after performing this analyse, we see no dependence on the optimum mutation rate with string length.

8 Conclusions

The 'Basin with a Barrier' problem is a caricature of a real world optimisation problem. Unlike simpler models such as one-max and royal road functions, it has local and global minima and thus shows some of the features of a hard optimisation problem.

We have been able to study the effects of all the GA parameters and say definitively how these parameters affect search on this model problem. In this case, we have seen mutation to be a disruptive force. Unlike crossover it has no knowledge of which parts of the strings are shared by many population members and thus disrupts parts of the string which are beneficial to fitness.

Crossover has been shown to be the dominant search operator on this landscape. Indeed without crossover the GA performs many orders of magnitude worse. By mixing those parts of the strings which are not identical, it is able to produce new population members without disrupting what has already been gained. However in the absence of mutation, selection very rapidly produces a highly correlated population which prevents crossover from operating. Thus we require a minimum level of mutation to overcome this correlation without disrupting the search.

At larger population sizes, the increase in correlation of the population is slower and thus the optimum mutation rate is lower. This optimum is seen to be independent of the length of the string.

Relating this work to real problems however is still some way off. However the techniques developed here have enabled a model problem to be analysed and definite statements made about the influence of parameters. In this way it represents a first step towards understanding the influences these parameters may have on real world problems.

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] 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.

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

[5] G. Ochoa, I. Harvey, and H. Buxton. On recombination and optimal mutation rates. In *Proceedings of the Genetic and Evolutionary Computation Conference 1999*, volume 1, pages 488–495, San Francisco, California, 1999. Morgan Kaufmann.

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

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

[8] A. Rogers and A. Prügel-Bennett. The Dynamics of a Genetic Algorithm on a Model Hard Optimization Problem. *Complex Systems*. To appear.

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

[10] 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.

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

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