An Analysis of Genetic Algorithms Using Statistical Mechanics*

Adam Prügel-Bennett and Jonathan L. Shapiro
Department of Computer Science, University of Manchester, Manchester, M13 9PL, UK.
(September 17, 1994)

A formalism is developed for studying Genetic Algorithms by considering the evolution of the distribution of fitness in the population. The effects of selection on the population are problem independent. The formalism predicts the optimal amount of selection. Crossover is solved for a model problem — finding low energy states of the one dimensional Ising spin glass. The theory is found to be in good agreement with simulations.

PACS 0550, 0250

Genetic Algorithms (GAs) are search techniques for finding good solutions to hard problems [1,2]. They have been applied to problems as diverse as the Travelling Salesman Problem, and the design of efficient aerofoils. Instead of making changes to a single solution, a population of solutions is evolved. Improvements are made by combining good solutions to produce (possibly) better ones.

To understand how GAs work, and thus to optimize their performance, it is necessary to understand their dynamics. Although a Genetic Algorithm can be described as a Markov Chain and thus solved formally [3], the effects of finite population make the transition probabilities very complicated. Consequently, this formulation has not yielded a predictive description of genetic dynamics. Another approach would be to exploit the well-known relationship between stochastic dynamics and the statistical mechanics of disordered systems. This is the subject of this paper. Statistical mechanics has already been applied to the study of other genetic dynamics (e.g. [4,5]). However, in these previous studies the selection of the individual to reproduce is random, there is no notion of *fitness*, whereas in GAs an individual reproduces with a probability determined by its fitness. In addition, the mechanisms of genetic mixing in the previous works were simpler than the crossover of the GA.

In this letter we show that statistical mechanics can be used to predict the evolution of a GA. We show that selection can be understood in terms of Derrida's Random Energy Model [6]. To elucidate other aspects we study a toy problem — finding low lying states of the one dimensional Ising spin glass. Here we just consider the two most important operations, selection and crossover. The techniques we have used can be readily extended to other problems and other GA operators. A fuller discussion will be given elsewhere [7].

The techniques developed here are useful in extending simple genetic and population models to include selection and crossover. In addition, they could have practical benefits to those applying GAs in optimization. This formulation predicts the evolution of the GA in terms of the amount of selection and other operators; this knowledge could help find the optimal values of selection and crossover. As an example, many investigators have found that increasing the degree of selection as the population evolves considerably improves performance [8]. The statistical mechanics formulation predicts the optimal value of selection in terms of properties of distribution of fitness, and does show that this optimum should increase during evolution.

By studying a problem with well defined statistical properties the techniques of statistical mechanics can be used to calculate the behaviour of a typical sample. The toy problem we examine is that of finding low lying states of a one dimensional spin glass with random nearest-neighbour couplings J_i drawn from a Gaussian distribution with zero mean and unit variance. The energy for a configuration of spins, $\vec{S} = (S_1, S_2, \ldots, S_{N+1})$ is

$$E(\vec{S}) = -\sum_{i=1}^{N} J_i S_i S_{i+1}. \tag{1}$$

The ground state energy is $E_{min} = -\sum_i |J_i|$. Although this problem is trivial to solve, it is nevertheless interesting in that is has an exponential number of local minima under single spin flip dynamics (typically $2^{N/3}$). Its thermodynamic properties have been extensively studied, see for example [9,10].

In its most basic form GAs work as follows: a populations of solutions, each represented by a string, is generated at random; the fitness of the solutions are measured and the fitter solutions are selected for; these fit solutions are combined in pairs to form new solutions; a random mutation can be applied; the fitness of the new solutions are tested and the selection, crossover and mutation is repeated; the algorithm finishes when the fitness of the population stops improving. Many modification to this basic scheme are possible. For example, members of one generation can be transferred to the next generation without any modifications. These enhancements can be studied within this formalism.

As an example we consider a GA to find the low lying states of a spin glass chain. The chain is represented by the vector of spins, $\vec{S}^{\alpha}$, where $\alpha = 1, \ldots P$ labels the different members of the population. The vectors of spins could be the strings used

*Published in *Phys. Rev. Lett.* **72** 1305–1309 (1994).

by the GA; under this representation the system has many local minima. Of course, other representations could be used. For example the spin variables could have been replaced by a new set of spin variables $\tau_i^\alpha = S_i^\alpha S_{i+1}^\alpha$, in which case the problem becomes much simpler, having no local minima. Alternatively we could have put the spins in a random order which would have made the problem considerably harder for the GA, as crossover would break many bonds. Thus the effect of different representations can be explored within this model. Here we will consider only the first of these representations.

The initial population is generated by choosing at random each spin to be $+1$ or -1. The energy, $E^\alpha = E(\vec{S}^\alpha)$, (or negative fitness) for each member of the population is calculated from equation (1). A new population of P members is selected by choosing members from the old population with a probability depending on their fitness. A number of different selection probabilities have been proposed. Here we weight each member with it "Boltzmann" probability

$$p^\alpha = \frac{\mathrm{e}^{-\beta E^\alpha}}{Z}, \qquad Z = \sum_{\alpha=1}^{P} \mathrm{e}^{-\beta E^\alpha}, \tag{2}$$

where β controls the amount of selection. In crossover a pair of solutions are mixed together to form a new solution. We have used "simple" or "single-point" crossover in which the strings are divided at a randomly chosen bond L and the subchains after that point are swapped. Since good solutions correspond to many bonds being satisfying, it is natural to use a crossover which preserves as many bonds as possible. Had we used uniformed crossover, in which each spin is chosen from either of the parents at random, then so many bonds would be broken that the child would have a very small chances of inheriting the fitness of its parents. In mutation randomly chosen spins are flipped. For the spin glass chain mutation turns out not to be so important. We will see why this is when we consider crossover in more detail.

Our approach to understanding GAs is to examine the change in the distribution of energies in the population, $\rho_t(E)$, at each generation t. For a finite population the energy distribution is a sum of delta functions which will depend on the particular choice of the couplings and on the randomly chosen initial population. We will therefore consider the statistical properties of the distribution, the mean, variance and higher cumulants (which are simply related to the moments). A typical evolution is illustrated in Fig. . The distribution $\rho_t(E)$, averaged over 100 samples, is shown for $t = 0$, 10, 20, 30 and 40. Initially the distribution is nearly Gaussian with zero mean and variance N. There are small correction to the Gaussian distribution which are easily calculated: the largest correction is to the fourth cumulant, which is equal to $-6N(1 + (N - 7)/P)$. As the population evolves its average energy decreases while the width of the distribution narrows — this narrowing is commonly referred to as convergence. Convergence is the inevitable consequence of selection acting on a finite population, since the fitter individuals will be selected more than once and children sharing the same parent are correlated. The distribution are also slightly skewed by selection, although this is hard to see by eye. This skewness together with higher cumulants reduces the efficiency of the search. In terms of the cumulants the role of crossover can be viewed as reducing the skewness.

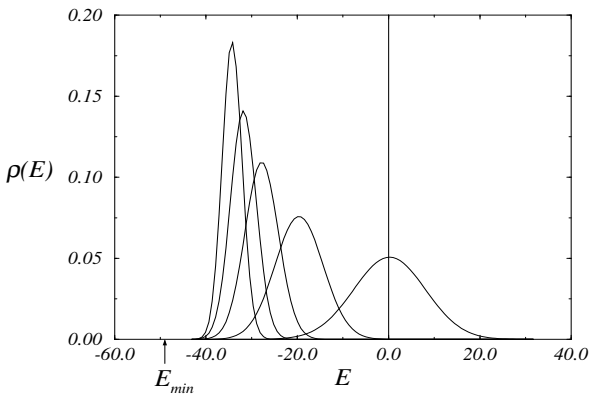

FIG. 1: The distribution of energy is shown after 0, 10, 20, 30 and 40 steps of the GA applied to the 1-d spin glass. The curves were created by averaging over 1 000 samples for $N = 63$, $P = 50$ and $\beta = 0.05$.

To study the evolution of the energy distribution we calculate the effect of selection and crossover on an arbitrary distribution. The full evolution of the GA is easily calculated by iterating

$$\rho_t(E) \overset{\text{selection}}{\longrightarrow} \rho_t^s(E) \overset{\text{crossover}}{\longrightarrow} \rho_t^{sc}(E) = \rho_{t+1}(E), \tag{3}$$

starting from the initial population $\rho_0(E)$. Other operators, such as mutation could also be included into this sequence.

Selection depends only on the energy distribution $\rho(E)$ and not on the details of the particular problem. Thus the effect of selection is universal. To calculate this effect it is convenient to consider $\log(Z)$ as a function of β, where Z is the sum of Boltzmann weights, since $\log(Z)$ is a generating function for the cumulants after selection

$$\kappa_n^s = (-1)^n \frac{\partial^n}{\partial \beta^n} \log(Z). \tag{4}$$

Furthermore, $\log(Z)$ self-averages — that is its average value equals a typical value — so we can consider its averaged value $\langle \log(Z) \rangle_\rho$, where $\langle \cdots \rangle$ denotes averaging over the energy distribution $\rho(E)$.

The Random Energy Model (REM) proposed by Derrida [6] consists of a set of random energy levels which are populated with a probability proportional to a Boltzmann factor. Our form of selection is precisely the REM, except that the energy levels come from a distribution that changes during evolution. To evaluate $\langle \log(Z) \rangle_\rho$ we use Derrida's trick of writing the logarithm in term of an integral representation, so that

$$\langle \log(Z) \rangle_\rho = \int_0^\infty \frac{\mathrm{e}^{-t} - \langle \mathrm{e}^{-tZ} \rangle_\rho}{t} \mathrm{d}t. \tag{5}$$

Since the partition function involves a sum over independent energy levels, the average over each energy level decouples. Thus evaluating $\langle \log(Z) \rangle_\rho$ reduces to performing a two-dimensional integral, which can be performed numerically. However, by making a small selection (high temperature) expansion the integral can be done analytically. To do this we expand $\langle \exp(-tZ) \rangle_\rho$ in powers of $t(1-e^{-\beta E})$. Ignoring terms of order $1/P^2$ and using

$$\left\langle e^{-tE} \right\rangle_\rho = \int_{-\infty}^{\infty} \rho(E)\, e^{-tE} dE = \tilde{\rho}(it) \tag{6}$$

where $\tilde{\rho}(t)$ is the Fourier transform of $\rho(E)$, we find

$$\langle \log(Z) \rangle_\rho = \log\big(\tilde{\rho}(i\beta)\big) - \frac{\tilde{\rho}(2i\beta)}{2P\big(\tilde{\rho}(i\beta)\big)^2}. \tag{7}$$

To turn this into a relationship in terms of the cumulants we use the standard expansion for the Fourier transform of a probability distribution

$$\log\big(\tilde{\rho}(t)\big) = \sum_{n=1}^{\infty} \kappa_n \frac{(it)^n}{n!}. \tag{8}$$

Substituting this into equation (7) and using equation (4) we can obtain a relationship for the cumulants after selection in term of the cumulants before selection.

To leading order the first cumulant becomes $\kappa_1^s = \kappa_1 - \beta\kappa_2 + \cdots$. Thus the mean energy is shifted by an amount proportional to the selection parameter β times the variance, κ_2. The variance is changed by an amount $\kappa_2^s = (1-1/P)\kappa_2 - \beta\kappa_3 + \cdots$. In Fig. we show the rate of convergence, κ_2^s/κ_2, versus the (scaled) selection parameter $\beta\sqrt{\kappa_2/2\log(P)}$, for populations of size 2^5, 2^{10} and 2^{20}, starting from a Gaussian distribution (i.e. $\kappa_n = 0$, $n > 2$). The solid lines are calculated by numerical integration using Gaussian quadrature. The dashed lines show the small β expansion. Note that the abscissa is equal to β^2 times the specific heat for the REM. The ordinate, to first order in β, is proportional to the shift in the mean energy of the distribution.

We can see from Fig. that even for $\beta = 0$ (arbitrary selection) there is an intrinsic convergence rate, which reduces the variance in the population by a factor $1 - 1/P$. This arises because, by chance, some members of the population will not be selected while other members will be selected more than once. As β increases the curve is initially flat. The improvement in the mean energy is proportional to β so it pays to increase β, since there is no extra loss of diversity in the flat region. As β increases passed some optimal value around the shoulder of the curves in Fig. , the diversity in the population decreases very rapidly with selection and the GA will be unlikely to find a good solution. Thus we see for Boltzmann selection there is an optimal choice of the selection parameter, β, and we can predict how it scales. Since the variance decreases as the population converges the degree of selection should be increased, as has already been observed [8]. So far we have ignored higher cumulants. Selection introduces a skewness into the distribution which slows down the shift in the average energy and increases the rate of convergence. Crossover reduces these higher cumulants allowing the GA to search for better solutions.

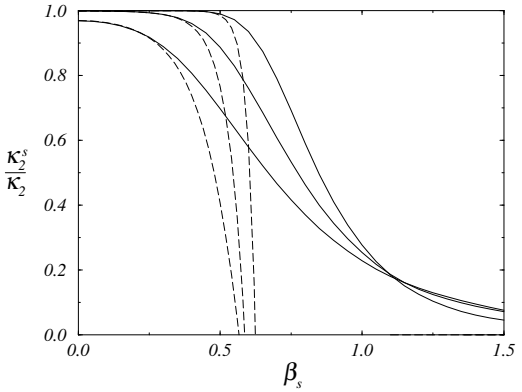

FIG. 2: The curves show the change in the variance of the energy distribution after selection starting from a Gaussian distribution for $P = 2^5$, 2^{10} and 2^{20} versus the scaled selection parameter $\beta_s = \beta(\kappa_2/2\log(P))^{1/2}$. The solid lines are calculated by numerical integration. The dashed lines show the small β expansion.

Crossover, unlike selection, is highly problem dependent. To calculate its effect we consider crossing two chain at a randomly chosen site L. The energy of each parent is equal to the sum of the energies of its two subchains $E^\alpha = E_L^\alpha + E_{N-L}^\alpha$, where

$$E_L^\alpha = -\sum_{i=1}^{L} J_i S_i^\alpha S_{i+1}^\alpha, \quad E_{N-L}^\alpha = -\sum_{i=L+1}^{N} J_i S_i^\alpha S_{i+1}^\alpha. \tag{9}$$

One children will have energy $E^{\alpha'} = E_L^\alpha + E_{N-L}^\beta + \Delta_I$, where Δ_I is the change in energy at the interface

$$\Delta_I = -J_L(S_L^\alpha S_{L+1}^\beta - S_L^\alpha S_{L+1}^\alpha). \tag{10}$$

While the energy of the other child is obtained by interchanging α and β. The sum of the energies of the two children will equal that of the two parents except for a possible correction due to the interface energy. For the spin glass chain the interface energy is very small so that the penalty for doing crossover is very low. In fact, on average crossover will cause less disruption than mutation, therefore in this model mutation does not play an important role. This would not be true for most models, for example in higher dimensional spin glasses crossover would cause a disruption of order L^{D-1}.

To calculate the effect of crossover we need to know the probability distributions for the energies of the subchains and the interface. The probability, $\rho^c(E)$, that a child will have energy E after crossover is just a product of the probabilities for the energies of the subchain and at the interface. In terms of Fourier transforms

$$\tilde{\rho}^c(t) = \tilde{p}_L(t)\, \tilde{p}_{N-L}(t)\, \tilde{p}_{\Delta_I}(t), \tag{11}$$

and the cumulants after crossover are equal to the sum of cumulants for the three probability distributions p_L, p_{N-L} and p_{Δ_I}. Calculating the energy of the subchains is in general complicated since we need to know the distribution of correlation between the chains. Rather than attempt to calculate this *ab initio* we estimate it from the energy distribution. This calculation becomes especially simple in the large N limit. Assuming the energy is spread evenly over the population then the probability of a subchain of length $L = lN$ having energy E_L is

$$\mathrm{P}(E_L|E) = \frac{\mathrm{e}^{-(E_L - lE)^2/2l(1-l)\kappa_2}}{\sqrt{2\pi l(1-l)\kappa_2}}. \qquad (12)$$

From which we find

$$\log\big(\tilde{p}_L(t)\big) = -l(l-1)\kappa_2/2 + \log\big(\tilde{\rho}(lt)\big). \qquad (13)$$

The probability that the interface bond will be broken or satisfied will, in general, depend on the energy of the two parents and their correlation. For the spin glass chain the average interface energy is

$$\langle \Delta_I \rangle = -\kappa_1\kappa_2/(2N^2). \qquad (14)$$

There are also contributions to the higher cumulants from p_{Δ_I}, but these give relatively small corrections, which are negligible in the limit $N \to \infty$. Using equations (11), (13), and (14) and averaging over the crossover point L, we find that the first cumulant is shifted by an amount $\langle \Delta_I \rangle$, the second cumulant remains unchanged, while the higher cumulants are suppressed by

$$\kappa_n^c = 2\kappa_n/(n+1) \qquad \text{for } n = 3, 4, 5. \qquad (15)$$

In Fig. we show the evolution of the first two cumulants for selection and crossover measured from simulation (solid curves) and by iterating the equations derived above starting from a Gaussian distribution and keeping the first six cumulants (dashed curves). The agreement with simulations is very good provided β and P are not too large. For larger values of these parameters the approximations used begin to break down and a more accurate treatment is necessary.

At a crude level crossover can be viewed as producing two competing effects. Firstly it reduces the higher cumulants (in particular the skewness), which means the population contains better solutions. Secondly, it produces an interface energy due to the disruption of the string which tends to reduce the average fitness of the population. By increasing the amount of mixing the higher cumulants will be further suppressed but usually at the cost of an increased interface energy. If the interface energy is large then it pays to use strong selection so that the evolution involves few crossovers. Conversely when the interface energy is small, for example by allowing crossover to occur only at bonds that do not cost any energy, then it would be better to use weaker selection.

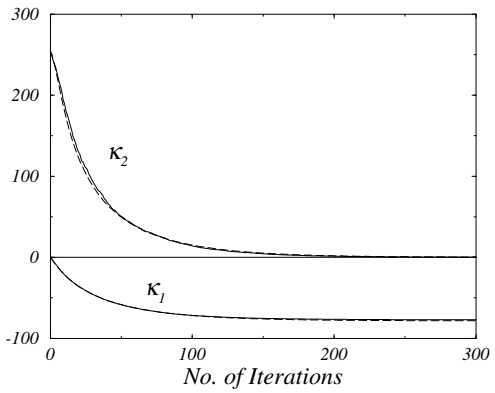

FIG. 3: Comparison of theory and simulations for selection and crossover using $\beta = 0.01$ with $P = 50$ and $N = 255$. The solid curves show the simulations averaged over 500 samples. The theory is shown by dashed curves, which are nearly obscured by the simulations.

The approach adopted here gives both a qualitative and quantitative understanding of how GAs work. This allows some strong predictions to be made about the optimal parameters and representations that should be used. This work can be easily extended to other toy problems such as the Potts spin glass which would shed light on other representational issues. Furthermore we would hope that this approach might be extendible to non-trivial problems such as finding low energy states of higher dimensional spin glasses.

We would like to thank Nick Radcliffe for initiating our interest in GAs and for his help along the way.

[1] J. H. Holland, *Adaptation in Natural and Artificial Systems* (University of Michigan Press (Ann Arbor), 1975).
[2] D. E. Goldberg, *Genetic Algorithms in Search, Optimization & Machine Learning* (Addison-Wesley (Reading, Mass), 1989).
[3] M. D. Vose and G. E. Liepins, Complex Systems **5** 31, (1991).
[4] P. G. Higgs and B. Derrida, J. Mol. Evol. **35** 454, (1992).
[5] D. S. Rokhsar, P. W. Anderson and D. L. Stein, J. Mol. Evol. **23** 119, (1986).
[6] B. Derrida, Phys. Rev. **B24**, 2613 (1984).
[7] A. Prügel-Bennett and J. L. Shapiro, in preparation. (unpublished).
[8] See, for example, reference [2] pp. 122–124.
[9] T. Li, Phys. Rev. B **24**, 6579 (1981).
[10] H. H. Chen and S. K. Ma, J. Stat. Phys. **29**, 717 (1982).