The Dynamics of a Genetic Algorithm under Stabilizing Selection

L. Magnus Rattray*
Computer Science Department, University of Manchester, Oxford Road
Manchester M13 9PL, U.K.

September 12, 1995

Abstract

A formalism recently introduced by Prügel-Bennett and Shapiro uses the methods of statistical mechanics to model the dynamics of genetic algorithms. To be of more general interest this formalism must be able to describe problems other than the test cases they consider. In this paper, the technique is applied to the subset sum problem, which is a combinatorial optimization problem with a strongly non-linear energy (fitness) function and many local minima under single spin flip dynamics. It is a problem which exhibits an interesting dynamics, reminiscent of stabilizing selection in population biology. The dynamics are solved under certain simplifying assumptions and are reduced to a set of difference equations for a small number of relevant quantities. The quantities used are the population's cumulants, which describe its shape, and the mean correlation within the population, which measures the microscopic similarity of population members. Including the mean correlation allows a better description of the population than the cumulants alone would provide and represents a new and important extension of the technique. The formalism includes finite population effects and describes problems of realistic size. The theory is shown to agree closely to simulations of a real genetic algorithm and the mean best energy is accurately predicted.

1 Introduction

Genetic Algorithms (GAs) are general-purpose search techniques which are loosely based on natural selection [1, 2]. A population of solutions evolve under the influence of genetic operators, which are roughly analogous to the processes at work in biological populations. GAs are growing in popularity and are being used in a large variety of problem domains (see [3] for examples). It is the existence of a population of solutions being processed in parallel which makes the GA different from other stochastic search techniques, such as simulated annealing. It is thought that the population allows the GA to find mutually useful features in distinct solutions, which it may then recombine to create better solutions. Although there is a large body of theoretical work on GAs, the established theory does not yet provide a complete picture.

Prügel-Bennett and Shapiro (PBS) have introduced a formalism for modeling the dynamics of the GA, using methods from statistical mechanics [4, 5, 6]. They considered two closely related toy problems; the random field paramagnet and the Ising spin chain, for which the dynamics can be solved exactly [5]. In this paper their formalism is generalized to a harder combinatorial optimization problem, subset sum. Although strictly NP-complete, this is a problem which is quasi-polynomial and in most cases can be solved with traditional algorithms in polynomial time [7]. This is still a toy problem, since stochastic methods will not perform as

*Internet address: rattraym@cs.man.ac.uk.

well as these algorithms, but there are closely related strong NP-complete problems, such as bin-packing, to which GAs have been applied effectively [8]. The solution of the dynamics for subset sum may be useful in understanding these harder cases.

In the simple GA considered here, solutions to a problem are coded as binary strings and each string is assigned an energy, or negative fitness, through some mapping function. A population of such strings is generated at random and GA operators act on it in sequence, over a number of generations, in order to find solutions of low energy (high fitness). The three most common operators are selection, crossover and mutation and these are the only operators considered in this work. Under selection, new population members are selected with replacement by some probabilistic method weighted towards the fittest. Selection requires the duplication of population members in order to keep the population size constant. Crossover mixes pairs of population members, creating offspring that may lie far from either parent in Hamming space. Under mutation, bits are flipped at random within the population. The GA has many tunable parameters, such as the population size, selection strength and mutation rate and a good choice of these parameters is often crucial to the GA's performance.

A full description of the GA's dynamics is very difficult. The number of possible population realizations is astronomical for a typical problem and crossover introduces a strongly non-linear interaction within the population. Following the formalism due to PBS [5], the population is modeled by a small number of macroscopic quantities and anything not trivially related to these quantities is retrieved through a maximum entropy method. This reduces the dynamics to a small number of difference equations describing the average effect of each operator on each relevant quantity. The formalism includes finite population effects and describes problems of realistic size. Describing the dynamics in this manner lends insight into which are the most important features of each operator and can be used to make optimal parameter choices.

The subset sum problem exhibits a very interesting dynamics. The optimal solution lies in a dense region of solution space, so that the population seems to stabilize around it. This form of selective pressure is known to population biologists as stabilizing selection and is very different from the problems considered by PBS [5], where selection is directional and pushes the population towards the optimal solution, which lies in a sparse region of the solution space. Because of the non-linearity within the energy, the relationship between the string coding and energy is more complex than in the paramagnet and spin chain considered by PBS [5]. The population's cumulants (which describe it's shape) are not sufficient to describe the state of the population accurately and another degree of freedom is required; the mean correlation within the population, which is a measure of the similarity between population members.

2 The algorithm

2.1 Subset sum

Posed as a question, the subset sum problem asks whether a set of numbers has a subset which exactly sums to a goal value [7]. Posed as an optimization problem, one wishes to find the subset which comes as close as possible to the goal value. Let the set of possibilities be $\{w_1, w_2 \ldots, w_N\}$, chosen from some arbitrary distribution. In this paper, the w_i come from a uniform distribution of reals over the interval $[0, 1]$, although the theoretical results are valid for any distribution with a well defined variance in the large N limit. If G is the goal, then one wishes to minimize,

$$|\sum_{i=1}^{N} x_i w_i - G| \qquad \text{where} \quad x_i \in \{0, 1\}$$

One possible choice of energy, or negative fitness, is the squared deviation,

$$E = \frac{(h - G)^2}{N} \qquad \text{where} \quad h = \sum_{i=1}^{N} x_i w_i \tag{1}$$

so that the aim of the GA would be to minimize this energy by making a particular choice of the x_i. There are other possible choices of energy function, but this choice is analytically advantageous as it is continuous with respect to h. The factor of $1/N$ is chosen to make the energy of order N. Here, h is the field value associated with energy E.

2.2 The genetic algorithm

A simple GA is considered. A random population of solutions is created, in this case binary strings of the form $\{x_1, x_2 \ldots, x_N\}$ where the alleles, x_i, are as defined above. The size of this population, P, remains fixed. Under selection, population members are chosen by some process weighted towards the individuals with the lowest energy. Although Baker (or deterministic) selection is generally thought to be more effective [9], we choose to use roulette wheel selection, since this makes the problem more amenable to analysis, allowing an accurate model for finite population effects. The population is then divided into pairs of parents for crossover. Under uniform crossover, the child's alleles are chosen independently and randomly from each parent, perhaps favoring alleles from one parent. Under mutation, alleles are flipped at random throughout the population with some probability, p_m.

Figure 1 shows the evolution of a simple GA, averaged over a number of runs using the energy defined in equation (1). The evolution is depicted in terms of the distribution of field variables, $p(h)$, which are defined in equation (1) above, since these allow a better description of the GA dynamics than the energy distribution. Clearly the field distribution will become centered around G as the GA evolves. Two stages in the evolution can be identified:

1. A directional stage, where the population's mean moves towards G.

2. A convergent stage, where the population's variance is reduced as population members close in on the optimum value. Here, selection tends to stabilize the population's mean around G.

As the distribution converges, the best population member will usually improve, although it may fluctuate between generations. Eventually the population may reach equilibrium, where the mutation rate keeps the population from converging further. Notice that the field distribution is always close to a Gaussian, while the energy distribution would clearly become unsymmetrical as the population evolves. In choosing to model the field distribution, one avoids the problem of dealing with a strongly non-Gaussian distribution.

3 The formalism

Following PBS [5], the dynamics are described in terms of each operator's average effect on the distribution's cumulants, which describe the population's shape, and the results for each operator are presented in the following three sections. The third and fourth cumulant expressions for crossover and mutation are not presented here, as they are equivalent to the paramagnet result in Ref. [5] under a simple change in variables. Unlike the paramagnet or the spin chain, for this problem the mean correlation within the population cannot be deduced from the population's variance and must be treated independently. The correlation expressions are derived in section 7 and represent an important new contribution to the formalism introduced by PBS [5]. Once expressions are derived for each operator's effect on the cumulants and the mean correlation, they can be iterated in sequence in order to simulate the GA's evolution.

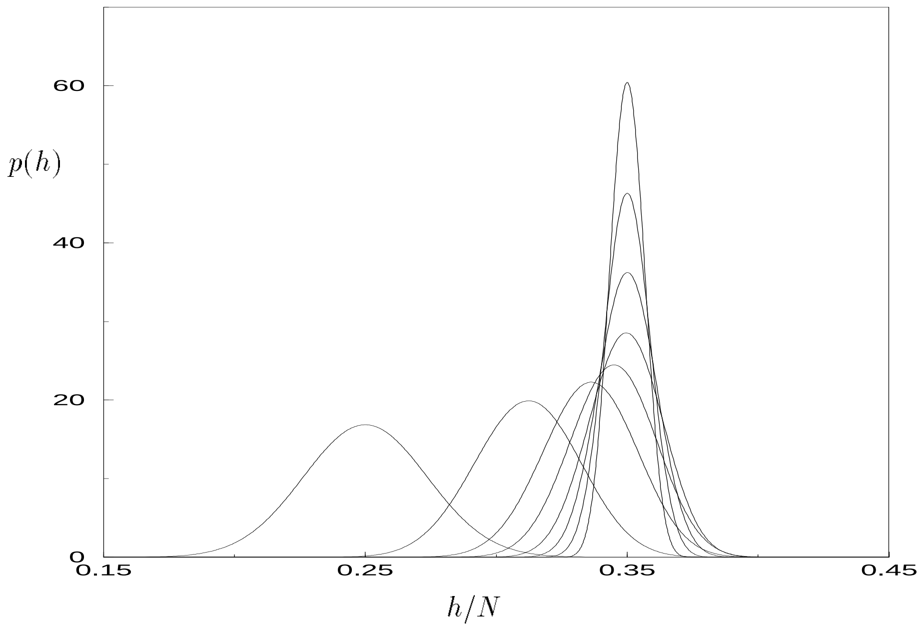

Figure 1: Evolution of a simple GA under the energy in equation (1) averaged over 1000 runs. The population's field distribution is shown at $0, 10, 20, 30, 50, 80, 110$ and 140 generations. The goal value, G/N, was 0.35.

3.1 The population's cumulants

The nth cumulant will be denoted κ_n, with κ_1 and κ_2 defined as the mean and variance respectively, while higher cumulants are a measure of the higher moments digression from those of a Gaussian. The third and fourth cumulants are related to the population's skewness and kurtosis respectively. The cumulants of the population can be generated from the logarithm of a partition function, Z,

$$\kappa_n = \frac{\partial^n}{\partial \gamma^n} \log Z|_{\gamma=0} \quad \text{where} \quad Z = \sum_{\alpha=1}^{P} e^{\gamma h_\alpha} \tag{2}$$

The first two cumulants are,

$$\kappa_1 = \langle h_\alpha \rangle_\alpha = \sum_{i=1}^{N} w_i \langle x_i^\alpha \rangle_\alpha \tag{3}$$

$$\begin{aligned} \kappa_2 &= \langle h_\alpha^2 \rangle_\alpha - \langle h_\alpha \rangle_\alpha^2 \;=\; (1 - \frac{1}{P})(\langle h_\alpha^2 \rangle_\alpha - \langle h_\alpha h_\beta \rangle_{\alpha \neq \beta}) \\ &= (1 - \frac{1}{P}) \left(\sum_{i=1}^{N} w_i^2 (\langle x_i^\alpha \rangle_\alpha - \langle x_i^\alpha x_i^\beta \rangle_{\alpha \neq \beta}) + \sum_{i=1}^{N} \sum_{j \neq i} w_i w_j (\langle x_i^\alpha x_j^\alpha \rangle_\alpha - \langle x_i^\alpha x_j^\beta \rangle_{\alpha \neq \beta}) \right) \end{aligned} \tag{4}$$

where α and β index population members and P is the population size. The averages over the population are defined as,

$$\langle x^\alpha \rangle_\alpha = \frac{1}{P} \sum_{\alpha=1}^{P} x^\alpha \qquad \langle x^\alpha x^\beta \rangle_{\alpha \neq \beta} = \frac{1}{P(P-1)} \sum_{\alpha=1}^{P} \sum_{\beta \neq \alpha} x^\alpha x^\beta$$

The $(1-1/P)$ factor in equation (4) is the finite population correction to the infinite population's variance. Notice that $x_i^2 = x_i$, since $x_i \in \{0, 1\}$.

An initial, random distribution will be close to Gaussian, which has $\kappa_n = 0$ for $n > 2$, although there are $O(N)$ corrections to the higher even cumulants due to the finite string length and $O(1/P)$ corrections due to the finite population. The first two cumulants of the initial distribution are,

$$\kappa_1^i = \sum_{i=1}^{N} \frac{w_i}{2} \tag{5}$$

$$\kappa_2^i = (1 - \frac{1}{P}) \sum_{i=1}^{N} \frac{w_i^2}{4} \tag{6}$$

3.2 The correlation

An important quantity is the weighted correlation between two population members, α and β, which is defined,

$$Q_{\alpha\beta} = \frac{1}{4N} \sum_{i=1}^{N} w_i^2 (2x_i^\alpha - 1)(2x_i^\beta - 1) \tag{7}$$

This is a simple measure of the correlation between different strings within the population and is not a function of energy. It differs from the usual measure used in statistical mechanics because of the w_i^2 factor. Knowing the correlation is important in determining the effect of crossover. The average correlation within the population is Q, defined,

$$Q = \langle Q_{\alpha\beta} \rangle_{\alpha \neq \beta} = \frac{1}{N} \sum_{i=1}^{N} \frac{w_i^2}{4} - \frac{1}{N} \sum_{i=1}^{N} w_i^2 (\langle x_i^\alpha \rangle_\alpha - \langle x_i^\alpha x_i^\beta \rangle_{\alpha \neq \beta}) \tag{8}$$

Notice that the second part of this quantity appears in the expression for the variance, equation (4), so that an increase in correlation corresponds to a proportional and opposite change in the variance if all other terms remain fixed. A completely uncorrelated population has $Q = 0$ while at maximum correlation, where each population member is identical, $Q = Q_{max}$ where,

$$Q_{max} = \frac{1}{N} \sum_{i=1}^{N} \frac{w_i^2}{4} \tag{9}$$

4 Selection

Roulette wheel selection is used, where the new population is chosen from the old with replacement and the probability of choosing a population member is equal to it's Boltzmann weight [10, 4]. This form of fitness scaling keeps the distribution very close to a Gaussian distribution and since the other genetic operators tend to return the population to a Gaussian it is a natural choice for this problem. The probability of selection for population member α with energy E_α is,

$$p_\alpha = \frac{e^{-\beta E_\alpha}}{\sum^P e^{-\beta E_\alpha}} \tag{10}$$

where β is the selection strength, which determines the difference in selection probability for solutions of different energy. If energy is a function of the field value, h, then one can generalize the calculation due to PBS [4] to calculate the effect of selection on the field distribution's cumulants. Define a partition function for selection,

$$Z_s = \sum_{\alpha=1}^{P} e^{-\beta E(h_\alpha) + \gamma h_\alpha} \tag{11}$$

The logarithm of this quantity is the generating function for the field distribution's cumulants after selection.

$$\kappa_n^s = \frac{\partial^n}{\partial \gamma^n} \log Z_s|_{\gamma=0} \tag{12}$$

Following PBS one makes the approximation that population members are independently drawn from a continuous distribution, $p(h)$, so that the average of $\log Z_s$ is,

$$\langle \log Z_s \rangle_{p(h)} = \left(\prod_{\alpha=1}^{P} \int_{-\infty}^{\infty} p(h_\alpha) dh_\alpha \right) \log Z_s \tag{13}$$

This can be transformed into a double integral which can be calculated numerically or approximated analytically in the limit of small β (see appendix B). In the simulations presented in section 8 the integration is done numerically, but the small β expansion shows the relevant contributions for each cumulant. Expanding equation (B.8) for the first three cumulants to first order one finds,

$$\begin{aligned}
\kappa_1^s &= \kappa_1 + \beta_s(1 - \frac{1}{P}) \left(2(G - \kappa_1) - \frac{\kappa_3}{\kappa_2} \right) + O(\beta_s^2) && (14)\\
\kappa_2^s &= (1 - \frac{1}{P})\kappa_2 - 2\beta_s(1 - \frac{3}{P}) \left(\kappa_2 - (G - \kappa_1)\frac{\kappa_3}{\kappa_2} \right) + O(\beta_s^2) && (15)\\
\kappa_3^s &= (1 - \frac{3}{P})\kappa_3 - 6\beta_s \left((1 - \frac{8}{P})\kappa_3 + \frac{2}{P}(G - \kappa_1)\kappa_2 \right) + O(\beta_s^2) && (16)
\end{aligned}$$

where $\beta_s = \beta\kappa_2/N$ is a scaled selection parameter. Notice that the variance is reduced even for zero selection strength because of finite population sampling effects. The third cumulant starts at zero (since the distribution is initially symmetrical) and becomes negative, indicating a skewed population. This is solely a finite population effect and is due to the sparseness of the population at the edge of the distribution, important during the directional stage of the dynamics. The major consequence of the negative third cumulant is an accelerated reduction in variance under further selection (the cumulants are $O(N)$, so the term in equation (14) involving the third cumulant is small). Although a narrow population is not necessarily bad in this problem, the reduction in variance due to the third cumulant is due only to increased correlations within the population. In section 7 it is shown explicitly that the higher cumulants introduced by selection increase the accumulation of correlations under selection, causing the GA to converge faster.

After a few generations, the mean becomes arbitrarily close to G and the magnitude of the third cumulant is reduced as the population becomes more symmetrical. Ignoring higher cumulants, the ratio of the second cumulant after and before selection at this stage is,

$$\frac{\kappa_2^s}{\kappa_2} = \frac{1}{1 + 2\beta_s} - \frac{1}{P(1 + 4\beta_s)^{\frac{3}{2}}} \tag{17}$$

One can keep this ratio fixed by scaling β so that β_s is kept constant. This requires an increased selection strength as the GA converges, a method recommended in the GA literature to avoid premature convergence to non-optimal solutions [2]. Although the skewness is reduced, selection still increases the magnitude of the negative fourth cumulant, related to the kurtosis, which increases the energy (reduces the fitness) of the best population member on average (see appendix D).

5 Crossover

The crossover and mutation calculations for the field distribution are equivalent to the paramagnet considered by PBS [5] under a simple change of variables, so only an outline of the derivation and results for the first two cumulants will be presented here. Consider two population members, α and β. They have associated field values,

$$h_\alpha = \sum_{i=1}^{N} x_i^\alpha w_i \qquad h_\beta = \sum_{i=1}^{N} x_i^\beta w_i \tag{18}$$

Then the field value of one child after crossover will be,

$$h_c = \sum_{i=1}^{N} (C_i x_i^\alpha + (1 - C_i) x_i^\beta) w_i \quad \text{where } C_i = \begin{cases} 1 & \text{with probability } a \\ 0 & \text{with probability } 1 - a \end{cases} \tag{19}$$

The parameter a is the probability associated with choosing alleles from parent α, with $a = 1/2$ the most common choice for uniform crossover. Reducing a reduces the degree of disruption and the size of a typical step made by crossover. After averaging over the C_i variables, the expressions for the first two cumulants, from equations (3) and (4), become,

$$\kappa_1^c = \kappa_1 \tag{20}$$

$$\kappa_2^c = \kappa_2 + 2a(1-a)\left((1 - \frac{1}{P}) N (Q_{max} - Q) - \kappa_2\right) \tag{21}$$

where Q and Q_{max} are defined in equations (8) and (9). The higher cumulants are reduced towards their natural value, with the constraint that the allele frequency at each site remains fixed within the population. For the higher cumulants a maximum entropy ansatz is used to calculate any terms not trivially related to the population's cumulants or mean correlation [5]. Crossover reduces the contribution to the second cumulant from the different site terms, so that the fixed point of the second cumulant under crossover is given by equation (4) without the $i \neq j$ contribution.

6 Mutation

During mutation, bits are flipped throughout the population with probability p_m. The resultant field after mutation of population member α is,

$$h_\alpha^m = \sum_{i=1}^{N} ((1 - M_i^\alpha) x_i^\alpha + M_i^\alpha (1 - x_i^\alpha)) w_i \quad \text{where } M_i^\alpha = \begin{cases} 1 & \text{with probability } p_m \\ 0 & \text{with probability } 1 - p_m \end{cases} \tag{22}$$

Averaging over the M_i^α variables in the first two cumulants leads to,

$$\kappa_1^m = \kappa_1 + 2p_m(\kappa_1^i - \kappa_1) \tag{23}$$

$$\kappa_2^m = \kappa_2 + 4p_m(1 - p_m)(\kappa_2^i - \kappa_2) \tag{24}$$

where κ_1^i and κ_2^i are the mean and variance of the initial, random population, defined in equations (5) and (6), which is the fixed point of the population under mutation. The higher cumulants are also reduced in magnitude [5]. The mutation probability, p_m, determines the rate at which the population approaches a random population. Sometimes the mutation rate is annealed while the GA evolves, since a high mutation rate seems to be most beneficial at the beginning of the search. In this problem, mutation was not found to be particularly beneficial, since any shift of the population's mean away from G reduces the number of good solutions considerably and this seems to outweigh the benefits of increased diversity.

7 Correlations

The expressions for selection and mutation only require knowledge of the distribution's cumulants, but the expression for the effect of crossover on the second cumulant requires the estimation of the correlation, Q, defined in equation (8). This is because crossover involves the interaction of population members, unlike the other two operators. In the paramagnet and spin chain, considered by PBS [5], the mean correlation can be deduced from the population's variance, but this is not possible in general. In order to deal with harder problems it will be necessary to evolve the mean correlation explicitly. This is fairly natural for crossover and mutation, but for selection one must make some assumption about the relationship between a population member's energy and it's microstate - the bit string coding. To deal with this issue a maximum entropy method is used, extending the ansatz presented by PBS [5].

Crossover does not change the mean correlation, since the allele frequency at each site within the population remains fixed. Using the same method outlined in section 6, one finds the correlation after mutation to be,

$$Q_m = (1-2p_m)^2 Q \tag{25}$$

This is reasonable, since the fixed point under mutation is clearly an uncorrelated distribution.

After selection, the expectation value for the correlation is,

$$\begin{aligned} Q_s &= \sum_{\alpha=1}^{P} p_\alpha^2 Q_{max} + \sum_{\alpha=1}^{P} \sum_{\beta\neq\alpha} p_\alpha p_\beta Q_{\alpha\beta} \\ &= \Delta Q_{dup} + (1-\frac{1}{P})Q_{nat} \end{aligned} \tag{26}$$

where p_α is the probability of selecting member α, defined in equation (10) and $Q_{\alpha\beta}$ is the correlation between population members α and β, defined in equation (7). The first term is the effect of the duplication of population members under selection, where Q_{max} is the correlation between identical population members, defined in equation (9). The second term is due to the natural increase in correlation as population members become fitter. Consider each term independently.

7.1 Duplication term

As in the selection calculation, one makes the assumption that population members are independently drawn from a continuous distribution. Then one can average over each population member in the first term of equation (26) and since the contribution from each term in the sum is equal, only the first term is required,

$$\begin{aligned} \Delta Q_{dup} &= PQ_{max}\left(\prod_{\alpha=1}^{P}\int_{-\infty}^{\infty} p(h_\alpha)dh_\alpha\right)\frac{e^{-2\beta E(h_1)}}{(\sum_\alpha e^{-\beta E(h_\alpha)})^2} \\ &= PQ_{max}\left(\prod_{\alpha=1}^{P}\int_{-\infty}^{\infty} p(h_\alpha)dh_\alpha\right) e^{-2\beta E(h_1)}\int_0^\infty dt\; te^{-t\sum_\alpha e^{-\beta E(h_\alpha)}} \end{aligned} \tag{27}$$

The integration introduces an integrand which factors, so that each average can be done independently,

$$\Delta Q_{dup} = PQ_{max}\int_0^\infty dt\; tf(t)g^{P-1}(t) \tag{28}$$

where,

$$\begin{aligned} f(t) &= \int_{-\infty}^{\infty} dh\, p(h)\exp(-2\beta(G-h)^2/N - te^{-\beta(G-h)^2/N}) \\ g(t) &= \int_{-\infty}^{\infty} dh\, p(h)\exp(-te^{-\beta(G-h)^2/N}) \end{aligned}$$

This expression can be expanded in the small β limit by the same method used for the selection calculation in appendix B. This turns out to be very accurate for reasonable selection strengths and to leading order one finds,

$$\Delta Q_{dup} = \frac{\hat{\rho}(2\beta) P Q_{max}}{(\hat{\rho}(3\beta)/\hat{\rho}(2\beta) + (P-1)\hat{\rho}(\beta))^2} \tag{29}$$

where,

$$\hat{\rho}(n\beta) = \int_{-\infty}^{\infty} dh\, p(h) \exp(-n\beta(G-h)^2/N) \tag{30}$$

This expression can be calculated using the cumulant expansion for $p(h)$ described in appendix A. To second order in β and excluding terms $O(1/\sqrt{N})$ and less one finds,

$$\Delta Q_{dup} \simeq \frac{Q_{max}}{P}\left[1 + 2\beta_s^2(1-\frac{4}{P})\left(1 + \frac{2(G-\kappa_1)^2}{\kappa_2} - \frac{2(G-\kappa_1)\kappa_3}{\kappa_2^2}\right)\right] \tag{31}$$

where β_s is the scaled selection parameter defined in section 4. Initially, the negative third cumulant introduced by selection tends to increase the accumulation of correlations due to duplication.

7.2 Natural increase term

The second term in equation (26) is more tricky, since it involves the relationship between energy and correlation. Ignoring the lost duplication terms and treating α and β as independent, one can average over each population member,

$$Q_{nat} = P^2 \langle p_\alpha p_\beta Q_{\alpha\beta} \rangle_{p(h_\alpha)p(h_\beta)} \tag{32}$$

Strictly speaking the lost duplication terms mean that α and β are not independent, but for reasonable selection strengths this approximation is valid. To calculate Q_{nat}, one can estimate the probability of $Q_{\alpha\beta}$ given h_α and h_β before selection, $p(Q_{\alpha\beta}|h_\alpha, h_\beta)$, and average this over the distribution after selection, $p_s(h)$,

$$\begin{aligned} Q_{nat} &= \int_{-\infty}^{\infty} dQ_{\alpha\beta}\, dh_\alpha dh_\beta\, Q_{\alpha\beta}\, p_s(h_\alpha) p_s(h_\beta) p(Q_{\alpha\beta}|h_\alpha, h_\beta) \\ &= \lim_{t\to 0} \frac{1}{N}\frac{d}{dt} \log\left(\int_{-\infty}^{\infty} dh_\alpha dh_\beta\;\; p_s(h_\alpha) p_s(h_\beta) \hat{\rho}(-itN|h_\alpha, h_\beta)\right) \end{aligned} \tag{33}$$

where $\hat{\rho}(-itN|h_\alpha, h_\beta)$ is the Fourier transform of $p(Q_{\alpha\beta}|h_\alpha, h_\beta)$,

$$\hat{\rho}(-itN|h_\alpha, h_\beta) = \int_{-\infty}^{\infty} dQ_{\alpha\beta}\, p(Q_{\alpha\beta}|h_\alpha, h_\beta)\, e^{tNQ_{\alpha\beta}} \tag{34}$$

A conditional probability for correlations, $p(Q_{\alpha\beta}|h_\alpha, h_\beta)$, can be defined if the population's alleles are chosen from some distribution,

$$\begin{aligned} p(Q_{\alpha\beta}|h_\alpha, h_\beta) &= \frac{p(Q_{\alpha\beta}, h_\alpha, h_\beta)}{p(h_\alpha, h_\beta)} \\ &= \frac{\langle \delta(Q_{\alpha\beta} - \frac{1}{4N}\sum w_i^2 (2x_i^\alpha - 1)(2x_i^\beta - 1))\, \delta(h_\alpha - \sum w_i x_i^\alpha)\, \delta(h_\beta - \sum w_i x_i^\beta) \rangle}{\langle \delta(h_\alpha - \sum w_i x_i^\alpha)\, \delta(h_\beta - \sum w_i x_i^\beta) \rangle} \end{aligned} \tag{35}$$

where the angled brackets denote averages over x_i^α and x_i^β. The alleles at each site are assumed to be distributed according to,

$$p(x_i) = X_i \delta(x_i - 1) + (1 - X_i)\delta(x_i) \tag{36}$$

where $X_i = \langle x_i^\alpha \rangle_\alpha$ and is defined by,

$$2X_i - 1 = \tanh(w_i(z + y\eta_i)) \tag{37}$$

The parameters y and z are chosen so that the population has the correct mean correlation and first cumulant. The noise term, η_i, is drawn at each site from a Gaussian distribution with unit variance. This is equivalent to assuming a maximum entropy distribution for the alleles, where y and z are Lagrange multipliers chosen to enforce constraints on two known macroscopics (see appendix C, equation (C.8)).

Consider the Fourier transform of $p(Q_{\alpha\beta}|h_\alpha, h_\beta)$ with respect to $Q_{\alpha\beta}$, since this appears in the appropriate generating function (see equation (33)),

$$\hat{\rho}(-itN|h_\alpha, h_\beta) = \frac{\hat{\rho}(-itN, h_\alpha, h_\beta)}{\hat{\rho}(0, h_\alpha, h_\beta)} \tag{38}$$

Writing the delta functions as integrals (ignoring multiplicative constants) and noticing that one of the integrals is removed by the Fourier transform, one finds,

$$\hat{\rho}(-itN, h_\alpha, h_\beta) = \left\langle \int_{-i\infty}^{i\infty} dk_\alpha dk_\beta \exp[F(k_\alpha, k_\beta, t)] \right\rangle_{\{x_i^\alpha, x_i^\beta\}} \tag{39}$$

where,

$$F(k_\alpha, k_\beta, t) = k_\alpha h_\alpha + k_\beta h_\beta - \sum_{i=1}^{N} \left[k_\alpha w_i x_i^\alpha + k_\beta w_i x_i^\beta - \frac{t w_i^2}{4}(2x_i^\alpha - 1)(2x_i^\beta - 1) \right] \tag{40}$$

Each site decouples and the average over each site can be done independently by integrating over the probability distribution defined in equation (36). After averaging one finds,

$$\hat{\rho}(-itN, h_\alpha, h_\beta) = \int_{-i\infty}^{i\infty} dk_\alpha dk_\beta \exp[G(k_\alpha, k_\beta, t)] \tag{41}$$

where,

$$\begin{aligned} G(k_\alpha, k_\beta, t) &= k_\alpha h_\alpha + k_\beta h_\beta + t\sum_{i=1}^{N} \frac{w_i^2}{4} \\ &+ \sum_{i=1}^{N} \left[\log(X_i^2 e^{-w_i(k_\alpha + k_\beta)} + X_i(1 - X_i)(e^{-k_\beta w_i - t w_i^2/2} + e^{-k_\alpha w_i - t w_i^2/2}) + (1 - X_i)^2) \right] \end{aligned} \tag{42}$$

Since the exponent is $O(N)$, this integral can be computed for large N by the saddle point method (see, for example, [11]),

$$\hat{\rho}(-itN, h_\alpha, h_\beta) = \exp[G(k_\alpha^*(t), k_\beta^*(t), t)] \tag{43}$$

where the saddle point equations fix the values of k_α^* and k_β^*,

$$\frac{\partial G}{\partial k_\alpha}(k_\alpha^*(t), k_\beta^*(t), t) = 0 \tag{44}$$

$$\frac{\partial G}{\partial k_\beta}(k_\alpha^*(t), k_\beta^*(t), t) = 0 \tag{45}$$

Define $\hat{\rho}(-itN)$, whose logarithm is the generating function for Q_{nat} (see equation (33)),

$$\begin{aligned} \hat{\rho}(-itN) &= \int_{-\infty}^{\infty} dh_\alpha dh_\beta \, p_s(h_\alpha) p_s(h_\beta) \hat{\rho}(-itN|h_\alpha, h_\beta) \\ &= \int_{-\infty}^{\infty} dh_\alpha dh_\beta \, p_s(h_\alpha) p_s(h_\beta) e^{G(k_\alpha^*(t), k_\beta^*(t), t) - G(k_\alpha^*(0), k_\beta^*(0), 0)} \end{aligned} \tag{46}$$

The probability distributions can be expressed in terms of their cumulant expansions, but since the higher cumulants do not make any contribution in the large N limit one can consider Gaussian distributions without loss of generality. Then, after again applying the saddle point method, one finds (ignoring multiplicative constants),

$$\hat{\rho}(-itN) = \exp[H(t)] \tag{47}$$

where,

$$H(t) = G(k_\alpha^*(t), k_\beta^*(t), t) - G(k_\alpha^*(0), k_\beta^*(0), 0) - \frac{(h_\alpha^* - \kappa_1^s)^2}{2\kappa_2^s} - \frac{(h_\beta^* - \kappa_1^s)^2}{2\kappa_2^s} \tag{48}$$

Here, κ_1^s and κ_2^s are the first two cumulants after selection and the saddle point equations fix h_α^* and h_β^*,

$$\frac{\partial H}{\partial h_\alpha}(h_\alpha^*, h_\beta^*) = k_\alpha^*(t) - k_\alpha^*(0) - \frac{h_\alpha^* - \kappa_1^s}{\kappa_2^s} = 0 \tag{49}$$

$$\frac{\partial H}{\partial h_\beta}(h_\alpha^*, h_\beta^*) = k_\beta^*(t) - k_\beta^*(0) - \frac{h_\beta^* - \kappa_1^s}{\kappa_2^s} = 0 \tag{50}$$

As $t \to 0$ the saddle point equations (44), (45), (49) and (50) are satisfied by,

$$k_\alpha^*(0) = k_\beta^*(0) = k \tag{51}$$

$$h_\alpha^* = h_\beta^* = \kappa_1^s \tag{52}$$

which are related through the equation,

$$\kappa_1^s = \sum_{i=1}^{N} \overline{\frac{w_i(X_i e^{-kw_i})}{1 - X_i(1 - e^{-kw_i})}} \tag{53}$$

where X_i is defined in equation (37) and the bar denotes an average over the Gaussian noise variable, η_i. The natural increase contribution to the mean correlation after selection can be generated from the the Fourier transform defined in equation (46),

$$\begin{aligned} Q_{nat} &= \lim_{t\to 0} \frac{d}{dt}\left(\frac{\log \hat{\rho}(-itN)}{N}\right) \\ &= \frac{1}{N}\sum_{i=1}^{N} \frac{w_i^2}{4} \overline{\left(\frac{(1 - X_i(1 + e^{-kw_i}))}{(1 - X_i(1 - e^{-kw_i}))}\right)^2} \end{aligned} \tag{54}$$

Again, the bar denotes an average over the noise in X_i. Notice that for $k = 0$ one retrieves the original values of κ_1 and Q given in appendix C. For small changes in the mean one can expand equations (53) and (54) around $k = 0$ to obtain,

$$Q_{nat} = Q - \frac{k}{N}\sum_{i=1}^{N} w_i^3 \overline{X_i(1 - X_i)(2X_i - 1)} + O(k^2) \tag{55}$$

$$k \simeq \frac{\kappa_1 - \kappa_1^s}{N(Q_{max} - Q)} \tag{56}$$

Notice that when X_i is greater than 1/2 on average, an increase in the population's mean results in an increased correlation. This is as expected, because the population is moving into a region of lower entropy. The expressions can be averaged numerically over the Gaussian noise and the weights, which are distributed uniformly over the interval $[0, 1]$. The small k approximation is a good starting point for the numerical root finding required to determine k from equation (53) and then to calculate Q_{nat} from equation (54).

Now both terms in equation (26) can be estimated from equations (29) and (54) to give the mean correlation after selection. The agreement with simulations is very good, although there is some discrepancy during the initial evolution where the population is most skew and the maximum entropy calculation is not completely accurate. During the convergent stage of the dynamics, when $\kappa_1 = G$, the increase in correlation under selection is solely due to duplication. This calculation is an important extension of the formalism presented by PBS [5], since it might allow the method to be extended to less trivial problems. The correlation within the population provides information which cannot be derived from the statistics of the population's energies (or, in this case, field values). It is crucial to calculating the effect of crossover, since high correlations will clearly lead to less disruption, and it is an ideal measure for describing the population's convergence.

8 Simulating the dynamics

The dynamics have been described in terms of difference equations for the field distribution's cumulants and the mean correlation, Q. To simulate the GA, the difference equations for the first four cumulants and the mean correlation were iterated in sequence and compared to averaged results from a real GA. The mean, variance and correlation are shown in figure 2 as the GA evolves for a typical choice of parameter settings. Mutation was not found to be particularly useful in this problem and was not included in the simulations. The GA was stopped when the correlation was 85% of a fully correlated population. The best population member, averaged over the same runs used in figure 2, is shown in figure 3 for the later stages of the GA's evolution. The theoretical value is estimated by assuming population members are independently drawn from a distribution with the correct cumulants (see appendix D).

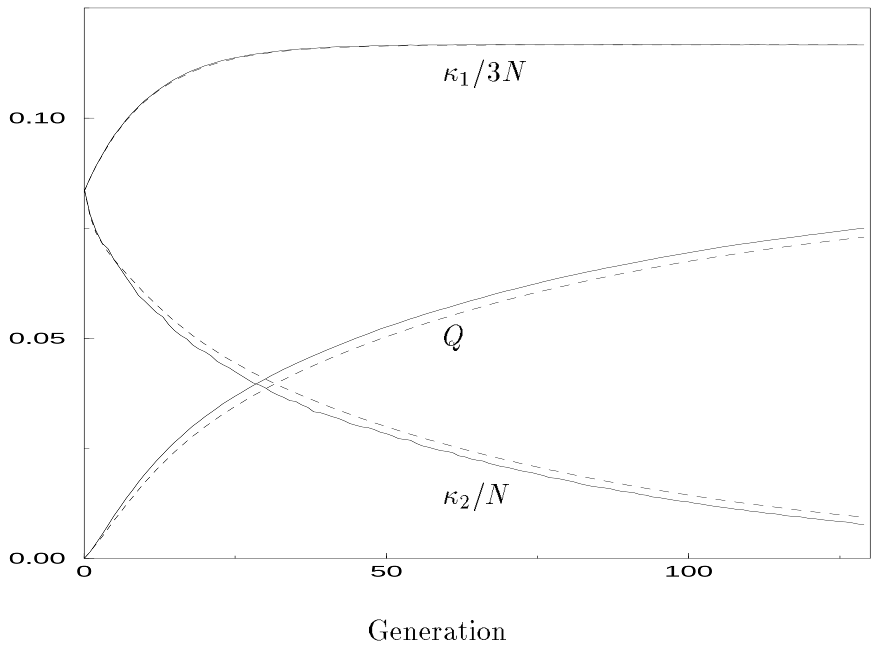

Figure 2: The theory is compared to simulations averaged over 1000 samples. The mean, variance and correlation are shown, with the dashed line showing the theory. The parameters were $\beta_s = 0.05$, Population = 70, $N = 150$, $G/N = 0.35$ and uniform crossover was used over the whole population.

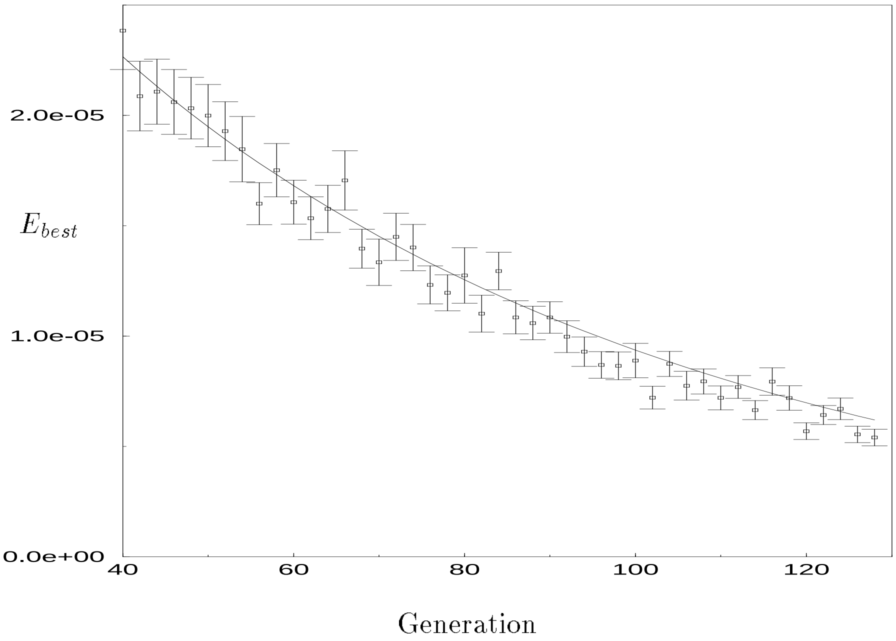

Figure 3: The best energy found in the population each generation is averaged over 1000 samples of a real GA and compared to the theory. The data is from the same runs used in figure 2. The solid curve is the value estimated from the theoretical cumulant values.

Most of the error in the curves of figure 2 is due to an initial underestimate of the increased correlation under selection. This is because the population is most skew at this point and the maximum entropy ansatz is least accurate here. The worst case in this set of simulations is an underestimate of about 5% at generation 4. After generation 15 the error is always within 0.5% and this is still during the stage of evolution where the natural increase term in equation (26) is an important contribution. If one were to model the GA starting from this point the agreement would be nearly perfect. Once the mean stabilizes at G, the increase in correlation is solely due to duplication under selection.

9 Conclusion

The formalism introduced by PBS has been extended to a less trivial problem with a strongly non-linear energy function. The theory shows good, although not perfect, agreement to simulations of a real GA averaged over many runs. In this problem, as in the spin chain and paramagnet considered by PBS [5], the higher cumulants are shown to increase convergence as they increase the accumulation of correlations under selection. During the later stages, the increase in correlation is shown to be almost solely due to the duplication required to maintain the population size after selection. The role of crossover seems to be to distribute the correlations more evenly in order to increase diversity, reducing the magnitude of the higher cumulants. The penalty of crossover is an increased variance, which reduces the density of solutions close to G and leads to an increase in average energy within the population. During the later stages of evolution it is shown that the ratio of the field distribution's variance before and after selection can be kept constant by scaling the selection parameter so that it is inversely proportional to variance. This problem exhibits a very different dynamics to most problems and it might be interesting to look at more realistic problems with this type of dynamics, such as bin packing and knapsack problems which have been solved by GAs in the past [8].

By evolving the mean correlation, the maximum entropy ansatz has been extended in a way which may prove important if progress is to be made on harder problems. The selection calculation has been generalized to any energy which can be expressed in terms of a field and this could be useful in examining certain toy problems, such as trap functions, which have received some attention from GA theorists (see, for example, [12]). The formalism is still limited to describing average behavior and an accurate model for fluctuations would be an improvement, although in the problems considered under this formalism the average behavior seems to express much of the most interesting behavior. To understand GA dynamics better, an array of different problem domains must be considered in order to discover generic issues and it is hoped that this work may provide a stepping stone towards that goal.

Acknowledgements

I would like to thank Adam Prügel-Bennett and Jon Shapiro for many helpful discussions and suggestions relating to this work. Thanks especially to Adam for pointing out a simplification to the calculation presented in section 7.1. This work was supported by a E.P.S.R.C. award (ref. 93315524).

A Cumulant expansion around a Gaussian

It is often necessary to describe the field distribution by a truncated cumulant expansion. The cumulants are defined for a finite sample in section 3, equation (2) and for a continuous distribution this definition can be generalized using a Fourier transform,

$$\hat{\rho}(t) = \int_{-\infty}^{\infty} p(h) e^{ith} dh \tag{A.1}$$

whose logarithm is the generating function for the cumulants,

$$\log \hat{\rho}(t) = \sum_{n=0}^{\infty} \kappa_n \frac{(it)^n}{n!} \tag{A.2}$$

One can define a function whose moments are correct given a limited number of cumulants [5],

$$p(h) = \frac{1}{\sqrt{2\pi\kappa_2}} \exp(\frac{-(h-\kappa_1)^2}{2\kappa_2}) \left(1 + \sum_{n=3}^{n_c} \frac{\kappa_n}{\kappa_2^{n/2}} u_n(\frac{h-\kappa_1}{\sqrt{\kappa_2}})\right) \tag{A.3}$$

where $u_n(x)$ are normalized Hermite polynomials and n_c is the number of cumulants used. Four cumulants are used in this work and the third and fourth Hermite polynomials are $u_3(x) = (x^3 - 3x)/3!$ and $u_4(x) = (x^4 - 6x^2 + 3)/4!$. This function is not a well defined probability distribution since it is not necessarily positive, but it does have the correct cumulants.

B Selection calculation

Following PBS [4], one can express the logarithm of the partition function as an integral,

$$\langle \log Z_s \rangle_{p(h)} = \int_0^{\infty} \frac{e^{-t} - \langle e^{-tZ_s} \rangle_{p(h)}}{t} dt \tag{B.1}$$

Now the average on the right can be taken,

$$\begin{aligned} \langle e^{-tZ_s} \rangle_{p(h)} &= \left(\prod_{\alpha=1}^{P} \int_{-\infty}^{\infty} p(h_\alpha) dh_\alpha \right) \exp(-t \sum_{\alpha=1}^{P} e^{-\beta E(h_\alpha) + \gamma h_\alpha}) \\ &= f^P(t, \beta, \gamma) \end{aligned} \tag{B.2}$$

where,

$$f(t,\beta,\gamma) = \int_{-\infty}^{\infty} p(h)\exp(-te^{-\beta E(h)+\gamma h}) \tag{B.3}$$

From equation (12), the cumulants after selection can be generated from the logarithm of the partition function,

$$\kappa_n = \lim_{\gamma\to 0}\frac{\partial^n}{\partial\gamma^n}\int_{-\infty}^{\infty}\frac{e^{-t}-f^P(t,\beta,\gamma)}{t}\,dt \tag{B.4}$$

This expression can be evaluated numerically, but the small β expansion is found to agree closely to the numerical result for any reasonable selection strength (at least for the first few cumulants). For small β one can expand $f(t,\beta,\gamma)$,

$$f(t,\beta,\gamma) \simeq 1 - t\hat{\rho}(1,\beta,\gamma) + \frac{t^2}{2}\hat{\rho}(2,\beta,\gamma) + \ldots \tag{B.5}$$

where,

$$\hat{\rho}(n,\beta,\gamma) = \int_{-\infty}^{\infty} p(h)e^{n(-\beta E(h)+\gamma h)}\,dh \tag{B.6}$$

Following PBS [5] one can exponentiate this expansion and raise to the power of P,

$$f^P(t,\beta,\gamma) = e^{-tP\hat{\rho}(1,\beta,\gamma)}\left(1+\frac{Pt^2}{2}(\hat{\rho}(2,\beta,\gamma)-\hat{\rho}^2(1,\beta,\gamma))\right) \tag{B.7}$$

Completing the integral in equation (B.4) one finds,

$$\kappa_n = \lim_{\gamma\to 0}\frac{\partial^n}{\partial\gamma^n}\left(\log(P\hat{\rho}(1,\beta,\gamma)) - \frac{1}{2P}(\frac{\hat{\rho}(2,\beta,\gamma)}{\hat{\rho}^2(1,\beta,\gamma)}-1)\right) \tag{B.8}$$

This can be calculated by representing the field distribution in terms of a cumulant expansion (see appendix A). The resulting expressions can be expanded as power series in β, which are shown to first order for the first three cumulants in equations (14) - (16) of section 4.

C Maximum entropy calculation

To calculate terms required for the determination of the higher cumulants after crossover and mutation, PBS introduce a maximum entropy calculation [5]. In this work this calculation will also be required in order to estimate the increased correlation within the population under selection. Although the calculation follows that presented by PBS closely, one of the constraints is different to that used by PBS and the weights come from a uniform distribution rather than a Gaussian.

Define X_i, the mean value of the alleles at each site within the population,

$$X_i = \langle x_i^\alpha\rangle_\alpha = \frac{1}{P}\sum_{\alpha=1}^{P} x_i^\alpha \tag{C.1}$$

The number of ways of arranging PX_i bits from P is,

$$\Omega(X_i) = \frac{1}{2^P}\begin{pmatrix} P \\ PX_i \end{pmatrix} \tag{C.2}$$

So one can define an entropy,

$$\begin{aligned} S(X_i) &= \log[\Omega(X_i)] \\ &\sim P\log\left(\frac{1}{2(1-X_i)}\right) + PX_i\log\left(\frac{1-X_i}{X_i}\right) \end{aligned} \tag{C.3}$$

where Stirling's approximation has been used. Lagrange multipliers, z and y, are used to add constraints on the first cumulant and correlation within the population.

$$
\begin{aligned}
2zP\kappa_1 &= 2z\sum_{\alpha=1}^{P}\sum_{i=1}^{N} w_i x_i^{\alpha} = 2zP\sum_{i=1}^{N} w_i X_i && \text{(C.4)}\\
2y^2P^2Q &= 2y^2\frac{1}{4N}\sum_{\alpha=1}^{P}\sum_{\beta=1}^{P}\sum_{i=1}^{N} w_i^2(1-2x_i^{\alpha})(1-2x_i^{\beta}) &&\\
&= P^2\frac{y^2}{2N}\sum_{i=1}^{N} w_i^2(1-2X_i)^2 && \text{(C.5)}
\end{aligned}
$$

One can define a probability distribution for the $\{X_i\}$ configuration which decouples for each site,

$$
\begin{aligned}
p(\{X_i\}) &= \prod_{i=1}^{N} p(X_i) = \prod_{i=1}^{N} e^{S(X_i)+2zPw_iX_i+P^2y^2w_i^2(2X_i-1)^2/2} &&\\
p(X_i) &= \int_{-\infty}^{\infty}\frac{d\eta_i}{\sqrt{2\pi}}e^{\frac{-\eta_i^2}{2}+PG(X_i,\eta_i,w_i)} && \text{(C.6)}
\end{aligned}
$$

where

$$G(X_i,\eta_i,w_i) = S(X_i)/P + 2zw_iX_i + yw_i(2X_i-1)\eta_i \tag{C.7}$$

To find the maximally likely value of X_i, maximize G with respect to X_i. This leads to the equation,

$$(2X_i-1) = \tanh[w_i(z+y\eta_i)] \tag{C.8}$$

where η_i is drawn from a Gaussian distribution with unit variance. The constraints can be used to obtain the values of z and y,

$$
\begin{aligned}
\kappa_1 &= \sum_{i=1}^{N}\frac{w_i}{2}(1+\overline{\tanh[w_i(z+y\eta_i)]}) && \text{(C.9)}\\
Q &= \frac{1}{N}\sum_{i=1}^{N}\frac{w_i^2}{4}\overline{\tanh^2[w_i(z+y\eta_i)]} && \text{(C.10)}
\end{aligned}
$$

The bars denote averages over the noise. The averages over noise and weights have to be done numerically. The numerical root finding required to determine z and y is simplified by the fact that the functions are fairly smooth (in fact the expression for κ_1 is monotonic in z).

D Best population member

One can estimate the best population member's energy assuming that population members are independently selected from a distribution of energies, $\rho(E)$,

$$E_{best} = P\int_0^{\infty} dE\,\rho(E)E\left[\int_E^{\infty} dE'\rho(E')\right]^{P-1} \tag{D.1}$$

This is related to the field distribution, $p_h(h)$, through the transformation,

$$\rho(E) = \frac{1}{2}\sqrt{\frac{N}{E}}\left(p_h(G-\sqrt{NE}) + p_h(G+\sqrt{NE})\right) \tag{D.2}$$

Eventually, the field distribution's mean is at G and substituting this expression into equation (D.1) for the case of a Gaussian field distribution one finds,

$$E_{best} = \frac{P}{N}\sqrt{\frac{2}{\pi\kappa_2}}\int_0^\infty dx\ \mathrm{erfc}^{P-1}(x)\, x^2 \exp\left(-x^2/2\kappa_2\right) \tag{D.3}$$

One can approximate this expression by using a flat field distribution with the same height at the mean. This should not affect the value of the best population member significantly, since it will always be close to the G, where the field distribution is locally flat. To include the effect of the higher cumulants, one can use the cumulant expansion for the field distribution given in appendix A to calculate the distributions height. In this case, the height, H, is,

$$H = \frac{1}{\sqrt{2\pi\kappa_2}}\left(1 + \frac{\kappa_4}{8\kappa_2^2}\right) \tag{D.4}$$

The best population member's energy is therefore,

$$\begin{aligned} E_{best} &\simeq \frac{2HP}{N}\int_0^{\frac{1}{H}} dx\ x^2 \left(1 - 2Hx\right)^{P-1} \\ &= \frac{\pi\kappa_2}{N(P+1)(P+2)(1+\kappa_4/8\kappa_2^2)} \end{aligned} \tag{D.5}$$

This will be a lower bound, because there is a larger probability within the neighborhood of G than for a Gaussian, but it should become exact in the large P limit. This is the expression used to calculate the theoretical curve in figure 3. The assumption in writing equation (D.1) is that the population members are independent and this breaks down during the later stages of evolution, when population members become highly correlated.

References

[1] J. H. Holland "Adaptation in Natural and Artificial Systems," (Ann Arbor, The University of Michigan Press, 1975).

[2] D. E. Goldberg, "Genetic Algorithms in Search, Optimization and Machine Learning," (Addison-Wesley, Reading, MA, 1989).

[3] L. Davis (editor) "Handbook Of Genetic Algorithms," (Van Nostrand Reinhold, New York, 1991).

[4] A. Prügel-Bennett, J. L. Shapiro "An Analysis of Genetic Algorithms Using Statistical Mechanics," *Physical Review Letters*, **72**(9) 1305 (1994).

[5] A. Prügel-Bennett, J. L. Shapiro, "The Dynamics of a Genetic Algorithm for Simple Random Ising Systems," Computer Science Dept., University of Manchester, Oxford Road, Manchester M13 9PL, U.K. (1995) (submitted to *Physica D* for publication).

[6] J. L. Shapiro, A. Prügel-Bennett, L. M. Rattray, "A Statistical Mechanical Formulation of the Dynamics of Genetic Algorithms," *Lecture Notes in Computer Science*, **865**, special edition on Evolutionary Computing edited by T. C. Fogarty, (1994).

[7] M. R. Garey, D. S. Johnson, "Computers and Intractability - A Guide to the Theory of NP-Completeness," (W. H. Freeman and Co., San Francisco, 1979).

[8] B. Kroger, O. Vornberger, "Enumerative vs Genetic Optimization - Two Parallel Algorithms for the Bin Packing Problem," *Lecture Notes in Computer Science*, **594**, 330-362 (1992).

[9] J. E. Baker "Reducing bias and inefficiency in the selection algorithm," in Proceedings of the Second International Conference on Genetic Algorithms pages 14-21, edited by J. J. Grefenstette, (Lawrence Erlbaum, Hillsdale, NJ, 1987).

[10] M. De la Maza, B. Tidor "Increased Flexibility in Genetic Algorithms: The Use of Variable Boltzmann Selective Pressure to Control Propagation," Proceedings of the ORSA CSTS Conference - Computer Science and Operations Research: New Developments in their Interfaces, 425-440 (1991).

[11] J. E. Marsden, "Basic Complex Analysis" (W. H. Freeman and Co., San Francisco., 1973).

[12] K. Deb, D. E. Goldberg, "Analyzing Deception in Trap Functions," in *Foundations of Genetic Algorithms 2*, edited by L. D. Whitley (Morgan Kaufmann, San Mateo, CA, 1993).