Lecture 3

The Genetic Operators.

Unlike the case of selection the **genetic operators**, mutation and selection, depend on the model or problem being studied. We will restrict our attention to using the energy of simple "spin systems" as fitness functions (these models have be studied extensively in reference [17, 18, 21]). This approach has also been applied to problems with more complicated fitness functions, namely the subset-sum problem [28] and generalization in a binary perceptron [22]. We begin this section by considering the simpler of the two genetic operators mutation.

3.1 Mutation.

3.1.1 Mutation for the Ideal Paramagnet

We consider first the simplest fitness function, the ideal paramagnet,

$$F_\alpha = \sum_{i=1}^{N} S_i^\alpha \tag{3.1.1}$$

with $S_i^\alpha \in \{-1, 1\}$. We will assume that each gene is mutated with a probability γ

$$S_i^\alpha \to S_i'^\alpha = \begin{cases} -S_i^\alpha & \text{with probability } \gamma \\ S_i^\alpha & \text{with probability } 1-\gamma. \end{cases} \tag{3.1.2}$$

Averaging over the two possible outcomes we find

$$\langle S_i'^\alpha \rangle_{mut} = \Gamma S_i^\alpha, \qquad \Gamma \equiv 1 - 2\gamma. \tag{3.1.3}$$

Using this result it is straightforward to calculate what happens to the average cumulants after applying the mutation operator. The fitness after mutation of a sequences $\vec{S}^\alpha$ is, on average,

$$\langle F'_\alpha \rangle_{mut} = \left\langle \sum_{i=1}^{N} S_i'^\alpha \right\rangle_{mut} = \Gamma \sum_{i=1}^{N} S_i^\alpha = \Gamma F_\alpha. \tag{3.1.4}$$

The first cumulant is

$$\bar{K}_1^m = \left\langle\!\!\left\langle \left\langle \langle F'_\alpha \rangle_{mut} \right\rangle_{\mathcal{P}} \right\rangle\!\!\right\rangle = \Gamma \langle\!\langle K_1 \rangle\!\rangle = \Gamma \bar{K}_1 \tag{3.1.5}$$

where we have commuted the average over the members of the population and the average over all possible mutations. The higher cumulants are only a little more complicated

$$\begin{aligned} \left\langle (F'_\alpha)^2 \right\rangle_{mut} &= \left\langle \sum_{i \neq j} S_i'^\alpha S_j'^\alpha + \sum_i (S_i'^\alpha)^2 \right\rangle_{mut} \\ &= \Gamma^2 \sum_{i \neq j} S_i^\alpha \, S_j^\alpha + \sum_i 1 = \Gamma^2 \, F_\alpha^2 + (1 - \Gamma^2)\, N \end{aligned} \tag{3.1.6}$$

and we have used $(S_i'^\alpha)^2 = 1$. The second cumulant is

$$K_2 = \left\langle F_\alpha^2 \right\rangle_{\mathcal{P}} - \left\langle F_\alpha \right\rangle_{\mathcal{P}}^2.$$

Using this result together with equations (3.1.4) and (3.1.6) we find

$$\bar{K}_2^m = \Gamma^2 \, \bar{K}_2 + (1 - \Gamma^2)\, N \tag{3.1.7}$$

The higher cumulants are

$$\bar{K}_3^m = \Gamma^3 \, \bar{K}_3 - 2\,\Gamma\,(1 - \Gamma^2)\, \bar{K}_1 \tag{3.1.8}$$

$$\bar{K}_4^m = \Gamma^4 \, \bar{K}_4 - 8\,\Gamma^2\,(1 - \Gamma^2)\, \bar{K}_2 - 2\,(1 - \Gamma^2)\,(1 - 3\,\Gamma^2)\, N. \tag{3.1.9}$$

The next two orders are given in reference [17]. The calculation of these terms is an exercise in combinatorics — it can be facilitated by using graphs; this is discussed in the appendices of reference [21]. The effect of mutation is schematically illustrated in figure 3.1.

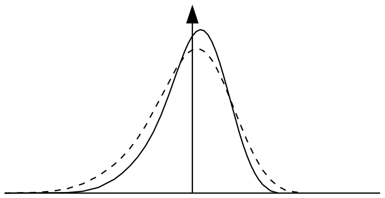

Figure 3.1: Schematic illustration of the effect of mutation on the distribution of fitnesses. The solid curve shows the distribution before mutation while the dashed curve shows the distribution after mutation. The average fitness is reduced while the width of the distribution is increased and the skewness reduce — these changes are all quite small.

The covariances can be found similarly. $V_{1,1}$ is defined as

$$V_{1,1} = \left\langle\!\!\left\langle K_1^2 \right\rangle\!\!\right\rangle - \left\langle\!\!\left\langle K_1 \right\rangle\!\!\right\rangle^2 = \left\langle\!\!\left\langle \frac{1}{P}\left(\sum_{\alpha=1}^{P} F_\alpha\right)^2 \right\rangle\!\!\right\rangle - \bar{K}_1^2 \tag{3.1.10}$$

so that

$$V_{1,1}^m = \Gamma^2 V_{1,1} \tag{3.1.11}$$

and similarly $V^m_{n,m} = \Gamma^{n+m} V_{n,m}$. This simple result arises because, for mutation on an infinite population, there are no fluctuations. If we had used the cumulants κ_n for the finite population distribution as ensemble variables then we would have found additional terms contributing to $v_{n,m}$ as we would expect. These results together with the selection results were used to calculate the evolution of an asexual population in a multiplicative fitness landscape shown in figure 2.2.

The effect of mutation on the correlation is also easy to calculate

$$q^m = \left\langle \left\langle \frac{1}{N} \sum_i S_i'^\alpha S_i'^\beta \right\rangle_{mut} \right\rangle_{\mathcal{P}} = \Gamma^2 q. \tag{3.1.12}$$

These equations are exact for all mutation rates. For this problem, selection only depends on the cumulants and not on the details of the sequence. This is not typical, but we might have anticipated that this would be the case here as the change in fitness due to mutation depends only on the fitness of the gene sequence and not on the sequence itself.

3.1.2 Mutation for Other Fitness Functions

The Random Field Paramagnet

The calculation presented above readily generalizes to other fitness functions but with some important differences. To illustrate this we consider the random field paramagnet

$$F_\alpha = \sum_{i=1}^N h_i S_i^\alpha \tag{3.1.13}$$

where $S_i^\alpha \in \{-1, 1\}$ and the h_i's are random. Following the argument we used above

$$\begin{aligned}
\langle F'_\alpha \rangle_{mut} &= \left\langle \sum_{i=1}^N h_i S_i'^\alpha \right\rangle_{mut} = \Gamma \sum_{i=1}^N h_i S_i^\alpha = \Gamma F_\alpha \\
\left\langle (F'_\alpha)^2 \right\rangle_{mut} &= \Gamma^2 F_\alpha^2 + (1-\Gamma^2) \sum_{i=1}^N h_i^2 \\
\left\langle (F'_\alpha)^3 \right\rangle_{mut} &= \Gamma^3 F_\alpha^2 + 3\Gamma (1-\Gamma^2) F_\alpha \sum_{i=1}^N h_i^2 - 2\Gamma (1-\Gamma^2) \sum_{i=1}^N h_i^3 S_i^\alpha
\end{aligned}$$

etc. Using these results we find

$$\bar{K}_1^m = \Gamma \bar{K}_1 \tag{3.1.14}$$

$$\bar{K}_2^m = \Gamma^2 \bar{K}_2 + (1-\Gamma^2) \sum_{i=1}^N h_i^2 \tag{3.1.15}$$

$$\bar{K}_3^m = \Gamma^3 \bar{K}_3 - 2\Gamma (1-\Gamma^2) \sum_{i=1}^N h_i^3 \langle S_i^\alpha \rangle_{\mathcal{P}} \tag{3.1.16}$$

etc. The fourth order term is given in reference [21]. These results differ form the constant field case in that the fields now appear explicitly. The change in the second cumulant

depends on $\sum_i h_i^2$ so we must use information about the statistic of the fields. The most significant difference to the constant field paramagnetic reveals itself first at third order. This is the appearance of the term

$$\sum_{i=1}^N h_i^3 \langle S_i^\alpha \rangle_{\mathcal{P}} .$$

It arises because mutation will depend on whether a few large fields are satisfied (i. e. $h_i S_i^\alpha > 0$) or whether many small fields are satisfied. The quantity

$$s_i = \langle S_i^\alpha \rangle_{\mathcal{P}} \tag{3.1.17}$$

is a measure of average **gene satisfaction** and will depend on the fields h_i. To calculate the effect of mutation on the higher cumulants we need to know about the distribution of s_i as a function of h_i. This information cannot be inferred *a priori* from the distribution of fitnesses. In the next section we will use a **maximum entropy ansatz** to estimate it.

The Spin-Glass Chain

We briefly consider mutation in the spin-glass chain as this involves spatial correlation and thus introduces some new terms. The fitness function for the spin-glass chain is defined by

$$F_\alpha = \sum_{i=1}^N J_i S_i^\alpha S_{i+1}^\alpha \tag{3.1.18}$$

where again the alleles take on Ising values, the J_i's are random variables, and where we have set the external fields to zero. This problem is interesting because there are many (typically $2^{N/3}$) local minima to single mutations (see references [31–34] for a discussion of its thermodynamic properties). Following the now familiar procedure, we find

$$\langle F'_\alpha \rangle_{mut} = \left\langle \sum_{i=1}^N J_i S_i'^\alpha S_{i+1}'^\alpha \right\rangle_{mut} = \Gamma^2 F_\alpha,$$

and thus $\bar{K}_1^m = \Gamma^2 \bar{K}_1$. The average fitness is reduced by a factor Γ^2 rather than Γ because mutating an allele breaks two bonds. Continuing to second order we must now separate out two terms: the $i = j$ term (mutations occurring at the same site) and the $i = j \pm 1$ term (mutations occurring on neighbouring sites)

$$\left\langle (F'_\alpha)^2 \right\rangle_{mut} = \Gamma^4 F_\alpha^2 + (1-\Gamma^4) \sum_{i=1}^N J_i^2 + 2\Gamma^2 (1-\Gamma^2) \sum_{i=1}^N J_i J_{i+1} S_i^\alpha S_{i+2}^\alpha$$

from which we find

$$\bar{K}_2^m = \Gamma^4 \bar{K}_2 + (1-\Gamma^4) \sum_{i=1}^N J_i^2 + 2\Gamma^2 (1-\Gamma^2) \sum_{i=1}^N J_i J_{i+1} \left\langle S_i^\alpha S_{i+2}^\alpha \right\rangle_{\mathcal{P}}$$

This last term introduces spatial correlations (i. e. the probability of neighbouring bond both being satisfied). Ignoring this spatial correlation the new term is approximately proportional to K_1^2/N. A fuller discussion of this calculations is given in reference [21, section 3.2 and appendix D].

3.2 Maximum Entropy Method

We are faced with a common problem in statistical mechanics: we know some macroscopic quantities, **the macro-state**, of a system but the dynamics depends on the **micro-states** of the system. In our case we know the fitness of a gene sequence but we need to know about the distribution of alleles as a function of the fields h_i

$$P(s_i|h_i), \qquad s_i = \langle S_i^\alpha \rangle_{\mathcal{P}}\,.$$

The usual approach is to use a **maximum entropy ansatz**; that is we assume that there is sufficient **microscopic chaos** to ensure that all micro-states consistent with the known macro-state are equally probable (or at least there is no systematic bias). This is an assumption about the underlying microscopic dynamics which is seldom known *a priori* and is usually be justified *a posteriori* by the quality of the agreement achieved. (A discussion of the role of maximum entropy in statistical mechanics is given in section II of Ma's book [35].)

We will restrict our attention to the random field paramagnet. To estimate the distribution $P(s_i|h_i)$ we will average over all configurations of the alleles, S_i^α, in the population subject to the two constraints

$$\begin{aligned} K_1 &= \left\langle \sum_{i=1}^N h_i\, S_i^\alpha \right\rangle_{\mathcal{P}} = \sum_{i=1}^N h_i\, \langle S_i^\alpha \rangle_{\mathcal{P}} = \sum_{i=1}^N h_i\, s_i \\ q &= \left\langle \frac{1}{N} \sum_{i=1}^N h_i\, S_i^\alpha S_i^\beta \right\rangle_{\mathcal{P}} = \frac{1}{N} \sum_{i=1}^N \langle S_i^\alpha \rangle_{\mathcal{P}}^2 = \frac{1}{N} \sum_{i=1}^N s_i^2. \end{aligned} \tag{3.2.1}$$

(There are other constraints we could impose – these are not related to the cumulants but to the founder effect correlations. These correction terms have not, to date, been explored.) To impose these two constraints we introduce Lagrange multipliers x and z so that the probability for a particular set of s_i's is

$$P(\{s_i\}|\{h_i\}) = \frac{1}{Z} \left(\prod_i \mathcal{N}(s_i) \right) \mathrm{e}^{\displaystyle zP \sum_i h_i s_i + P^2 x^2 \sum_i s_i^2/2} \tag{3.2.2}$$

where Z is a normalization factor and the numerical factors multiplying the Lagrange multipliers are chosen for convenience. The Lagrange multipliers x and z must be chosen so that the constraints (3.2.1) are satisfied. The term $\mathcal{N}(s_i)$ is the number of ways that the alleles can be chosen so that $\langle S_i^\alpha \rangle_{\mathcal{P}} = s_i$,

$$\mathcal{N}(s_i) = \operatorname*{Tr}_{\{S_i^\alpha\}} \delta\left(P s_i - \sum_{\alpha=1}^P S_i^\alpha \right) = \begin{pmatrix} P \\ P(1+s_i)/2 \end{pmatrix}. \tag{3.2.3}$$

(This is just equal to the binomial coefficient). The **entropy** of the state s_i is

$$S(s_i) = \log\left(\mathcal{N}(s_i)\right) \approx -\frac{P}{2} \log\left(1 - s_i^2\right) + \frac{P s_i}{2} \log\left(\frac{1-s_i}{1+s_i} \right) \tag{3.2.4}$$

where we have used Stirling's approximation. The product over sites in equation (3.2.2) decouples

$$P(\{s_i\}|\{h_i\}) = \prod_{i=1}^N P(s_i).$$

We can use the familiar properties of Gaussians to decouple the quadratic term $P^2 x^2 s_i^2/2$ to yield

$$P(s_i) = \frac{1}{Z} \frac{\mathrm{d}\eta_i}{\sqrt{2\pi}} \mathrm{e}^{\displaystyle -\eta_i^2/2 + P\, G(s_i, h_i, \eta_i)} \tag{3.2.5}$$

where

$$G(s_i, h_i, \eta_i) = S(s_i)/P + z h_i s_i + x \eta_i s_i. \tag{3.2.6}$$

For large P the term $\exp(P\,G)$ will be a very sharply peaked function of s_i so that the distribution $P(s_i)$ will be dominated by the maximum of G which is given by

$$\frac{\partial G}{\partial s_i} = \frac{\partial s(s_i)}{\partial s_i} + z h_i + x \eta_i = -\operatorname{atanh}(s_i) + z h_i + x \eta_i = 0 \tag{3.2.7}$$

and thus

$$s_i = \langle\!\langle S_i \rangle\!\rangle = \tanh(z h_i + x \eta_i) \tag{3.2.8}$$

where η_i is a Gaussian distributed variable. The Lagrange multipliers x and z are chosen to satisfy the two constraints

$$K_1 = \sum_i h_i\, \overline{\tanh(z h_i + x\eta)} \tag{3.2.9}$$

$$q = \sum_i \overline{\tanh^2(z h_i + x\eta)} \tag{3.2.10}$$

where the bar denotes an integration over the Gaussian variable η. Thus the average field satisfaction s_i will be a random variable between -1 and 1 whose average and fluctuations depend of K_1 and q. The quantity needed to calculate K_3^m in equation (3.1.16) is

$$\sum_i h_i^3 s_i = \sum_i h_i^3 \overline{\tanh(z h_i + x\eta)}.$$

In deriving these results we used a large P approximation in three places: firstly we used Lagrange multipliers which only ensures that the constraints are satisfied on average (we could have evoked the ensemble to justify this approximation); secondly we used Stirling's approximation; and finally we used a saddle point type approximation. *A priori* there appears to be no reason for these approximations to be malicious (this is not the case in selection where the $1/P$ corrections accumulate at each generation). We can test the maximum entropy ansatz independently by looking at the distribution $P(s_i|h_i)$. In figure 3.2 we show the first two moments, $\bar{s}_i$ and $\bar{s_i^2}$, of this distribution as a function of the field strength h_i measured from simulations and computed from the theory. The curves where obtained for the spin-glass chain after 25 generations. The form of the curves are what we would intuitively expect: the alleles associated with small fields have a good chance of being unsatisfied while those for the large fields are almost all satisfied. In addition the fluctuations are much larger for small fields. The cause of the discrepancy between the simulations and theory is anyone's guess. The analysis certainly needs to be repeated and should be done on the random field paramagnet where everything is simpler.

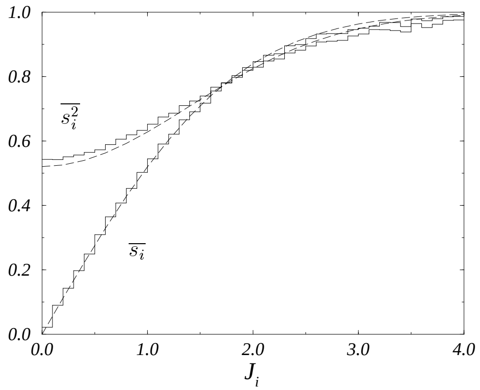

Figure 3.2: The curves show the average of s_i and s_i^2 measured from simulations (solid lines) and the theoretical prediction (dashed curve) plotted against the coupling strength h_i. The parameters are the same as those used in figure 2.

3.3 Crossover or Sexual Recombination.

The most interesting evolution operator is sexual recombination or crossover. This opens up a new way to explore the sequence space, which is not possible except with a population. We being this section by outlining different strategies before we discuss how crossover can be calculated.

3.3.1 Outline

There are many different ways of performing crossover both in Genetic Algorithms and in biology. We will again use the jargon of the GA community as this usually derives from English rather than Latin. The most popular form of crossover is pairwise crossover in which two **parents** $\vec{S}^\alpha$ and $\vec{S}^\beta$ are crossed together to form a child $\vec{S}^\mu$. (We will use the indices $\alpha, \beta, \ldots$ to refer to the parent gene sequences and $\mu, \nu, \ldots$ to refer to their children.) Pairwise crossover can be expressed symbolically by

$$S_i^\mu = \chi_i^\mu \, S_i^\alpha + (1 - \chi_i^\mu) \, S_i^\beta \tag{3.3.1}$$

where $\chi_i^\mu \in \{0, 1\}$. In Genetic Algorithm it is a frequent practice to take the complementary child

$$S_i^\nu = (1 - \chi_i^\mu) \, S_i^\alpha + \chi_i^\mu \, S_i^\beta. \tag{3.3.2}$$

which ensures that both the population size and the alleles in the population remain unchanged.

There are different methods for performing crossover depending on how the χ_i^μ's are chosen. The classic method is **single-point crossover** where

$$\chi_i^\mu = \begin{cases} 1 & \text{if } i \le A_\mu \\ 0 & \text{if } i > A_\mu \end{cases} \tag{3.3.3}$$

and $A_\mu \in \{1, 2, \cdots, N\}$ is usually chosen uniformly. This type of crossover is clearly useful in cases where the fitness function has spatial correlation, as in the three colouring problem or the spin-glass chain. A second popular crossover technique is uniform crossover

$$\chi_i^\mu = \begin{cases} 1 & \text{with probability } a \\ 0 & \text{with probability } 1 - a \end{cases} \tag{3.3.4}$$

This is the method of choice for problems with no spatial component such as the paramagnet, the subset-sum problem [28], or generalization in a binary perceptron [22]. The **bias** a controls the amount of mixing. Optimal mixing is achieved when $a = 1/2$, but this is also likely to be the most disruptive and consequently the most costly in terms of the average loss in fitness. In practical GA applications this may be an important parameter to control.

Pairwise crossover is certainly not necessary; any number of parents may be used to create a new child. The case when the alleles of a child is chosen from any parent in the population is known as **bit-simulated crossover**, although it is tempting to call this **orgy crossover**. This strategy really only makes sense in problems where there is no cost involved in crossover. Unfortunately these problems do not have any local minima so are not of great interest. Nevertheless Baum, Boneh and Garrett [36] have proved the rather intriguing result that in one class of problems with no local minima (a version of one's counting where the alleles take on values $\{1, 2, \ldots, L\}$, the fitness being the number of 1's) a GA using bit-simulated crossover could out perform a hill climbing.

Because crossover can be disruptive and cause an average loss in fitness a common practice is to keep some fraction of the population intact to prevent good members being lost. The folklore is that a good crossing fraction is 60%. This will be problem independent and would be worth investigating within this framework.

3.3.2 The Interface Energy

There is often a cost associated with crossover which arises when the fitness of the alleles depend on each other. We will refer to this as the **interface energy**. Nature tries to minimize this cost by allowing crossover to occur only as certain restriction sites. It would be hopeless to allow uniform crossover on base pairs between two alleles coding for the same gene. Thus protein sequences evolve mostly through mutation (at least, this is my understanding).

The interface energy will depend on the fitness of the parents, the number of "bonds" broken and the correlation between the parents. We can illustrate this by considering the spin-glass chain. The fitness of the child μ produced by single-point crossover is

$$\begin{aligned} F_\mu &= \sum_{i=1}^{A-1} J_i S_i^\alpha S_{i+1}^\alpha + J_A S_A^\alpha S_{A+1}^\beta + \sum_{i=A+1}^{N} J_i S_i^\beta S_{i+1}^\beta \\ &= \sum_{i=1}^{A} J_i S_i^\alpha S_{i+1}^\alpha + \sum_{i=A+1}^{N} J_i S_i^\beta S_{i+1}^\beta + \Delta_\mu \end{aligned} \tag{3.3.5}$$

where Δ_μ is the interface energy given by

$$\Delta_\mu = J_A S_A^\alpha (S_{A+1}^\beta - S_{A+1}^\alpha). \tag{3.3.6}$$

Averaging over all possible crossover sites

$$\langle \Delta_\mu \rangle = \frac{1}{N} \sum_{A=1}^{N} J_A \left\langle S_A^\alpha (S_{A+1}^\beta - S_{A+1}^\alpha) \right\rangle_{\mathcal{P}}. \tag{3.3.7}$$

The second term is equal to K_1/N while the first term we can write as

$$\frac{1}{N} \sum_{A=1}^{N} J_A \left\langle S_A^\alpha S_{A+1}^\alpha S_{A+1}^\alpha S_{A+1}^\beta \right\rangle_{\mathcal{P}} \tag{3.3.8}$$

where we have used the identity $(S_{A+1}^\alpha)^2 = 1$. However,

$$\frac{1}{P} \sum_{\beta \neq \alpha} S_{A+1}^\alpha S_{A+1}^\beta \approx q \tag{3.3.9}$$

where q is the average correlation between gene sequences (we ignore fluctuations depending on α and A). Thus

$$\langle \Delta_\mu \rangle = -\frac{K_1(1-q)}{N} \tag{3.3.10}$$

We can obtain corrections to higher order cumulants similarly. The change in the average fitness due to the interface energy is, as advertised, proportional to the average fitness K_1 and $(1-q)$. For the spin-glass chain only a single bond is broken; for problems such as the graph colouring problem where many "bonds" will be broken the interface energy will be proportional to the number of broken bonds. For non-spatial problems there may also be an average cost associated with crossover although there will not be a clearly defined interface.

It may be possible to reduce the interface energy by tailoring the crossover to the problem in question. For example, in graph colouring one could try to "relax" the interface to reduce the interface cost. Alternatively one could try to permute the colours in one of the parents to reduce this cost. I do not know how common this practice is nor has it been studied systematically within the framework.

3.3.3 The Shuffling Problem

Before discussing the mixing effect of crossover we consider a simplified problem, namely **"The Shuffling Problem"**. This is where there is only crossover with no selection or mutation. The problem is to understand how the initial genes are shuffled by the crossover operator. We will denote the initial generation by a Latin index $\vec{S}^a$. To measure the degree of shuffling it is useful to define the **ancestor function** $\llbracket \cdots \rrbracket_a$

$$\llbracket S_i^\alpha(t) \rrbracket_a = \begin{cases} 1 & \text{if } S_i^\alpha(t) \text{ originated from } a \\ 0 & \text{otherwise} \end{cases} \tag{3.3.11}$$

We denote the number of alleles in gene sequence $\vec{S}^\alpha(t)$ originating in gene sequence $\vec{S}^a(0)$ by the **ancestry variable**

$$K_a^\alpha(t) = \sum_{i=1}^{N} \llbracket S_i^\alpha(t) \rrbracket_a. \tag{3.3.12}$$

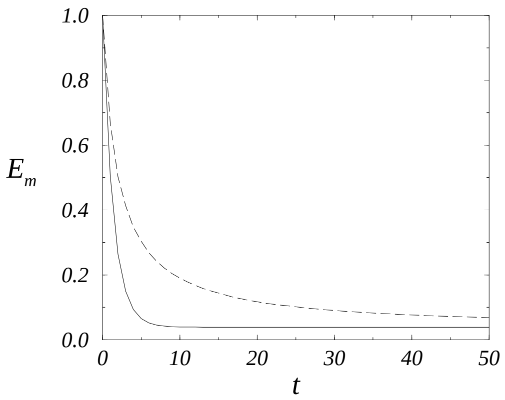

Figure 3.3: The curves show the time evolution of the mixing energy for shuffling using uniform crossover (solid line) and single-point crossover (dashed line). The simulations where for $P = 32$ and $N = 128$ and show the average over 100 runs. The errors are of the same magnitude as the line width. Taken from [29].

Finally we can define a **mixing energy** (density) E_m measuring how well mixed the genes sequences are

$$E_m(t) = \frac{1}{N^2 P} \sum_{a=1}^{P} \sum_{\alpha=1}^{P} \left(K_a^\alpha(t) \right)^2. \tag{3.3.13}$$

Initially $E_m(0) = 1$ while when the gene sequences are totally random the average mixing energy is $(N+P-1)/NP$. In figure 3.3 we show the mixing energy for uniform crossover (solid curve) and single-point crossover (dashed curve). The mixing produced by uniform crossover is substantially faster than that for single-point crossover.

The curve for uniform crossover can be calculated exactly, which we will do below. The curve for single-point crossover can also be calculated to extremely high accuracy, but we do not derive this here as the theory of single-point crossover is still being developed [29].

To calculate the effect of uniform shuffling we consider the ancestry variable, K_a^μ, for a child produced by uniform crossover

$$K_a^\mu = \sum_{i=1}^{N} \llbracket S_i^\mu \rrbracket_a = \sum_{i=1}^{N} \left(\chi_i^\mu \llbracket S_i^\alpha \rrbracket_a + (1-\chi_i^\mu) \llbracket S_i^\beta \rrbracket_a \right). \tag{3.3.14}$$

To calculate the mixing energy we must look at the square of this quantity

$$(K_a^\mu)^2 = \sum_{i \neq j} \llbracket S_i^\mu \rrbracket_a \llbracket S_j^\mu \rrbracket_a + \sum_i \llbracket S_i^\mu \rrbracket_a \tag{3.3.15}$$

where we have separated out the $i = j$ term and used that fact that

$$\llbracket S_i^\mu \rrbracket_a^2 = \llbracket S_i^\mu \rrbracket_a$$

which follows because it is a binary valued function. To calculate what happens on average we average over χ_i^μ, using

$$\langle \chi_i^\mu \rangle_\chi = a. \tag{3.3.16}$$

This gives

$$\left\langle (K_a^\mu)^2 \right\rangle_\chi = \sum_{i \neq j} T_i^\mu T_j^\mu + \sum_i T_i^\mu = \left(\sum_i T_i^\mu \right)^2 + \sum_i \left(T_i^\mu - \left(T_i^\mu \right)^2 \right) \tag{3.3.17}$$

where

$$T_i^\mu = a \left[\!\left[S_i^\alpha \right]\!\right]_a + (1-a) \left[\!\left[S_i^\beta \right]\!\right]_a .$$

From this we find

$$\left(T_i^\mu \right)^2 = a^2 \left[\!\left[S_i^\alpha \right]\!\right]_a + (1-a)^2 \left[\!\left[S_i^\beta \right]\!\right]_a$$

since the cross term must vanish (both S_i^α and S_i^β cannot originate from the same ancestor in shuffling). Using the definition of K_a^α

$$\left\langle (K_a^\mu)^2 \right\rangle_\chi = \left(a K_a^\alpha + (1-a) K_a^\beta \right)^2 + a(1-a) \left(K_a^\alpha + K_a^\beta \right). \tag{3.3.18}$$

The average mixing energy at the next generation is

$$\langle E_m(t+1) \rangle_\chi = \frac{1}{N^2 P} \sum_{a=1}^{P} \sum_{\mu=1}^{P} \left(K_a^\mu \right)^2 . \tag{3.3.19}$$

Substituting in equation (3.3.18) and making the replacement

$$\frac{1}{P} \sum_{\mu=1}^{P} = \frac{1}{P(P-1)} \sum_{\alpha=1}^{P} \sum_{\alpha \neq \beta}$$

that is, the children are produced from all pairs of parents, then

$$\langle E_m(t+1) \rangle_\chi = \left(a^2 + (1-a)^2 \right) E_m + \frac{a(1-a)}{P(P-1)N^2} \sum_a \sum_{\alpha \neq \beta} \left(2 K_a^\alpha K_a^\beta + K_a^\alpha + K_a^\beta \right). \tag{3.3.20}$$

Writing

$$\sum_{\alpha \neq \beta} K_a^\alpha K_a^\beta = \left(\sum_\alpha K_a^\alpha \right)^2 - \sum_\alpha (K_a^\alpha)^2$$

we get after collecting together terms

$$\langle E_m(t+1) \rangle_\chi = r E_m(t) + (1-r) E_m^\infty \tag{3.3.21}$$

where

$$r = 1 - \frac{2a(1-a)P}{P-1}, \qquad E_m^\infty = \frac{N+P-1}{NP}. \tag{3.3.22}$$

Since $\langle E_m(t+1) \rangle_\chi$ depends only linearly on $E_m(t)$ we can average over possible crossovers at each time step. Using the initial condition $E_m(0) = 1$ we find the average mixing energy at time t is

$$\langle E_m(t) \rangle_\chi = E_m^\infty - r^t \left(1 - E_m^\infty \right). \tag{3.3.23}$$

Thus for uniform crossover the population reaches the asymptotic mixing energy, E_m^∞, exponentially fast, with a rate $-\log(r)$. As we would expect, this is optimal when half the spins are crossed.

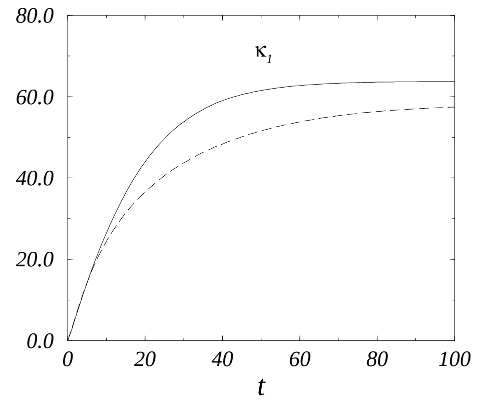

Figure 3.4: The curves show the average fitness for the ideal paramagnet problem. Both uniform crossover (solid curve) and single-point crossover (dashed curve) are shown. The simulations are for $P = 100$, $N = 64$ and $\beta = 0.05$. The results are averaged over 1000 runs. The error bars are the same thickness as the line.

3.3.4 The Mixing Effect of Crossover

We will restrict our attention to the ideal paramagnet,

$$F_\alpha = \sum_{i=1}^{N} S_i^\alpha \tag{3.3.24}$$

with $S_i^\alpha \in \{-1, 1\}$. Generalization of this calculation to the case of the random-field paramagnet is analogous to the case of mutation. Figure 3.4 illustrates the evolution of the average fitness for a Genetic Algorithm with selection and crossover only. The solid curve shows the case when uniform crossover is used and the dashed curve shows the case when single-point crossover is used. The same parameters were used in both cases. Uniform crossover clearly out performs single-point crossover on this problem (there is no interface energy). The reason is simply that uniform crossover mixes the alleles much more effectively than single-point crossover, as discussed in the previous section.

The theory of crossover is current under revision and a proper treatment of single-point crossover has still to be developed (I am working on it [29]). We will discuss just the case of uniform crossover which is far simpler. This calculation follows closely the calculation we just carried out for the shuffling problem.

The fitness function after selection is given by

$$F_\mu = \sum_{i=1}^{N} S_i^\mu = \sum_{i=1}^{N} \left(\chi_i^\mu S_i^\alpha + (1-\chi_i^\mu) S_i^\beta \right). \tag{3.3.25}$$

Averaging over all possible crossovers (using equation (3.3.16)) we obtain

$$\left\langle F_\mu \right\rangle_\chi = \sum_{i=1}^{N} \left(a S_i^\alpha + (1-a) S_i^\beta \right) = a F_\alpha + (1-a) F_\beta \equiv \hat{F}_\mu \tag{3.3.26}$$

Averaging over all parents we obtain

$$K_1^c = K_1 \tag{3.3.27}$$

(we omit the average over the ensemble as this is trivial). The first cumulant is unchanged by crossover on average. If we had created the complementary child the first cumulant would be unchanged exactly since the all the alleles in the population would be conserved.

To calculate the change to the second cumulant we consider

$$F_\mu^2 = \sum_{i=1}^{N}\sum_{j\neq i} S_i^\mu\, S_j^\mu + \sum_{i=1}^{N} 1 \tag{3.3.28}$$

where we have separated out the term $i = j$. Averaging over all crossovers

$$\begin{aligned}
\left\langle F_\mu^2 \right\rangle_\chi &= \sum_{i=1}^{N}\sum_{j\neq i} \langle S_i^\mu\rangle_\chi \left\langle S_j^\mu\right\rangle_\chi + \sum_{i=1}^{N} 1 \\
&= \left(\sum_{i=1}^{N} \langle S_i^\mu\rangle_\chi\right)^2 - \sum_{i=1}^{N}\left(1 - \langle S_i^\mu\rangle_\chi^2\right) \\
&= \hat{F}_\mu^2 - \sum_{i=1}^{N}\left(1 - \langle S_i^\mu\rangle_\chi^2\right).
\end{aligned} \tag{3.3.29}$$

From which we find

$$K_2^c = \hat{K}_2 - \sum_{i=1}^{N}\left(1 - \left\langle \langle S_i^\mu\rangle_\chi^2 \right\rangle_{\mathcal{P}}\right). \tag{3.3.30}$$

where $\hat{K}_2$ to be the second cumulant of $\hat{F}_\mu$. To calculate $\hat{K}_n$ we use the generating function

$$\begin{aligned}
\hat{G}(z) &= \log\left\langle e^{z\hat{F}_\mu}\right\rangle_{\mathcal{P}} = \log\left\langle e^{z(aF_\alpha + (1-a)F_\beta)}\right\rangle_{\mathcal{P}} \\
&= \log\left\langle e^{zaF_\alpha}\right\rangle_{\mathcal{P}} + \log\left\langle e^{z(1-a)F_\beta}\right\rangle_{\mathcal{P}}
\end{aligned} \tag{3.3.31}$$

so that

$$\hat{K}_n = G^{(n)}(0) = \left(a^n + (1-a)^n\right) K_n. \tag{3.3.32}$$

The term $\langle S_i^\mu\rangle_\chi^2$ is equal to

$$\begin{aligned}
\langle S_i^\mu\rangle_\chi^2 &= \left(a\, S_i^\alpha + (1-a)\, S_i^\beta\right) \\
&= \left(a^2 + (1-a)^2\right) + 2a\,(1-a) S_i^\alpha\, S_i^\beta
\end{aligned} \tag{3.3.33}$$

Putting equations (3.3.32) and (3.3.33) into equation (3.3.30) we obtain

$$K_2^c = \left(a^2 + (1-a)^2\right) K_2 + 2a\,(1-a) N\,(1-q) \tag{3.3.34}$$

where q is the average string correlation defined by

$$q = \sum_{i=1}^{N}\left\langle S_i^\alpha\, S_i^\beta\right\rangle_{\mathcal{P}} = \sum_{i=1}^{N}\left\langle (S_i^\alpha)^2\right\rangle_{\mathcal{P}}. \tag{3.3.35}$$

The n^{th} cumulant result is (postulated to be) of the form

$$K_n^c = r_n K_n + (1-r_n) K_n^0 \tag{3.3.36}$$

where

$$r_n = \left(a^n + (1-a)^n\right) \tag{3.3.37}$$

and K_n^0 are the **natural cumulants** which one would obtain by a total shuffle of the alleles

$$\begin{aligned}
K_2^0 &= \sum_{i=1}^{N}\left(1 - \left\langle (S_i^\alpha)^2\right\rangle_{\mathcal{P}}\right) \\
K_3^0 &= -2\sum_{i=1}^{N}\left(\langle S_i^\alpha\rangle_{\mathcal{P}} - \left\langle (S_i^\alpha)^3\right\rangle_{\mathcal{P}}\right).
\end{aligned} \tag{3.3.38}$$

(I have no proof of equation (3.3.36) although it seems to be true, at least, up to forth order.)

The meaning of this result is that the n^{th} cumulant is moved towards its natural value at a rate $-\log(r_n)$ this is a maximum when $a = 1/2$ (half the genes are swapped) when the rate is $r_n = 2^{n-1}$. The second cumulant might will increase although this may be a small effect (this is a small effect for weak Boltzmann selection). The higher cumulants will all be reduced significantly. The effect of crossover is schematically illustrated in figure 3.5. At first sight the effect of selection might look modest but this is deceptive. Crossover re-populates the high fitness tail which are exactly the solutions of interested. Crossover dramatically increases the effectiveness of the search. Genetic Algorithms learn very quickly; their problem is that the population correlates rapidly and consequently crossover becomes ineffective.

3.3.5 Calculating Correlations

Both crossover and the maximum entropy calculation depend on the correlation between the gene sequences. We now consider how these may be calculated. This approach was first developed by Magnus Rattray [22,28]. We will take a slightly different approach still being developed [29]. To simplify the discussion we will ignore mutation and restrict our attention to the ideal paramagnet.

The correlation in the population is defined as

$$q(t) = \left\langle \frac{1}{N}\sum_{i=1}^{N} S_i^\alpha(t)\, S_i^\beta(t)\right\rangle_{\mathcal{P}}. \tag{3.3.39}$$

This correlation arises through two mechanisms: there is a natural correlation because the average fitness in the population and there is in addition a correlation due to the founder

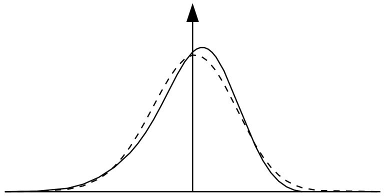

Figure 3.5: Schematic illustration of the effect of crossover on the distribution of fitnesses. The solid curve shows the distribution before crossover while the dashed curve shows the distribution after crossover. The average fitness is slightly reduced due to the interface energy while the width of the distribution is slightly increased. The most significant change is a large reduction in the skewness. The results in a re-population of the tail of high fitness.

effect. We can split up these two mechanisms. We denote by $c_2(t)$ the probability that $S_i^\alpha(t)$ and $S_i^\beta(t)$ originate from a common ancestor

$$c_2(t) \equiv \left\langle \frac{1}{N}\sum_{i=1}^{N}\frac{1}{P}\sum_{a=1}^{P} \llbracket S_i^\alpha(t) \rrbracket_a \llbracket S_i^\beta(t) \rrbracket_a \right\rangle_{\mathcal{P}}. \qquad (3.3.40)$$

If two alleles originate from the came ancestor they, $S_i^\alpha(t)\,S_i^\beta(t) = 1$. If they come from different ancestors their correlation will be that expected for random alleles with an average fitness K_1,

$$\left\langle S_i^\alpha(t)\, S_i^\beta(t) \right\rangle_{\mathcal{P}} = k_1^2(t) \qquad (3.3.41)$$

where

$$k_1(t) \equiv \frac{K_1(t)}{N}. \qquad (3.3.42)$$

We have actually made an approximation that unrelated and related alleles have the same average fitness. This may not be true. For example, related alleles may be present in the population precisely because they have above average fitness. We may try appealing to microscopic chaos to argue this effect is small, alternatively we could try to treat it more systematically. Here I will assume that it is small. The correlation is then

$$q(t) = c_2(t) + k_1^2(t)\,\left(1 - c_2(t)\right). \qquad (3.3.43)$$

We are left with the problem of determining $c_2(t)$. This is determined by the selection operator as crossover only swaps the alleles but leaves their abundance constant (and we are ignoring mutations). To calculate $c_2(t)$ we introduce the **ancestor number** L_i^a defined as the number of alleles at site i coming from ancestor $\vec{S}^a$,

$$L_i^a(t) = \sum_{\alpha=1}^{P} \llbracket S_i^\alpha(t) \rrbracket_a. \qquad (3.3.44)$$

By definition

$$\begin{aligned} c_2(t+1) &= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{P}\sum_{a=1}^{P}\frac{1}{P(P-1)}\sum_{\alpha=1}^{P}\sum_{\alpha\neq\beta} \llbracket S_i^\alpha(t) \rrbracket_a \llbracket S_i^\beta(t) \rrbracket_a \\ &= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{P^2(1-P)}\sum_{a=1}^{P}\left(\left(\sum_{\alpha=1}^{P}\llbracket S_i^\beta(t) \rrbracket_a\right)^2 - \sum_{\alpha=1}^{P}\sum_{\alpha=1}^{P}\right) \qquad (3.3.45) \\ &= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{P^2(1-P)}\sum_{a=1}^{P} L_i^a(t)(L_i^a(t)-1). \end{aligned}$$

But

$$\sum_{a=1}^{P} L_i^a(t) = P$$

since all alleles in the population originate from some ancestor. Denoting the number of times member α is selected by n_α we find that

$$\begin{aligned} \left(L_i^a(t+1)\right)^2 &= \left(\sum_{\alpha=1}^{P} n_\alpha \llbracket S_i^\alpha(t) \rrbracket_a\right)^2 \qquad (3.3.46) \\ &= \sum_{\alpha=1}^{P} n_\alpha^2 \llbracket S_i^\alpha(t) \rrbracket_a + \sum_{\alpha=1}^{P}\sum_{\alpha\neq\beta} n_\alpha\, n_\beta \llbracket S_i^\alpha(t) \rrbracket_a \llbracket S_i^\beta(t) \rrbracket_a. \qquad (3.3.47) \end{aligned}$$

Appealing once more to microscopic chaos we assume that n_α is independent of $\llbracket S_i^\alpha(t) \rrbracket_a$; that is, the number of times a member is chosen is independent of the ancestry of a particular allele. Then averaging over all possible selections

$$\left(L_i^a(t+1)\right)^2 = \left\langle n_\alpha^2 \right\rangle \sum_{\alpha=1}^{P} \llbracket S_i^\alpha(t) \rrbracket_a + \sum_{\alpha\neq\beta} n_\alpha\, n_\beta \llbracket S_i^\alpha(t) \rrbracket_a \llbracket S_i^\beta(t) \rrbracket_a \qquad (3.3.48)$$

where we have used $\langle n_\alpha \rangle = 1$ (the size of the population is conserved). Substituting this into equation (3.3.45) we find

$$c_2(t+1) = c_2(t) + \frac{\langle n_\alpha^2 \rangle - 1}{P(P-1)} \qquad (3.3.49)$$

We calculated $\langle n_\alpha^2 \rangle$ in section 2.2.4. This calculation generalizes to th higher order statistics of the founder effect need for calculating the effect of crossover on higher order cumulants.

3.4 The Benefits of Sexual Recombination.

Having developed a quantitative formalism for studying micro-evolution we may ask are there any qualitative lessons to be learned. One approach to answering this question is to simplify the formulae we have derived to obtain an approximate dynamics which can be solved in closed form. This approach was carried out in reference [21, section 6.3]. Here we will take a somewhat lighter approach and consider a model problem we have called "**The Basin with a Barrier**" [37]. The fitness function is a modification of the ideal paramagnet

$$F_\alpha = -V(m_\alpha), \qquad m_\alpha \equiv \frac{M_\alpha}{N} \equiv \frac{1}{N}\sum_{i=1}^{N} S_i^\alpha \qquad (3.4.1)$$

where the potential $V(m_\alpha)$ is of the form

$$V(m) = \begin{cases} N\left(w(m-m_l)^2/2 + v_l\right) & \text{if } m \le m_b \\ 0 & \text{if } m > m_b \end{cases} \tag{3.4.2}$$

This is shown in the left graph of figure 3.6. There is a local minima at m_l and a barrier at m_b. The problem is to find the **well** or global minima beyond the barrier. We choose this potential for computational convenience. It clearly favours Genetic Algorithms, but the conclusions hold for much smoother barriers.

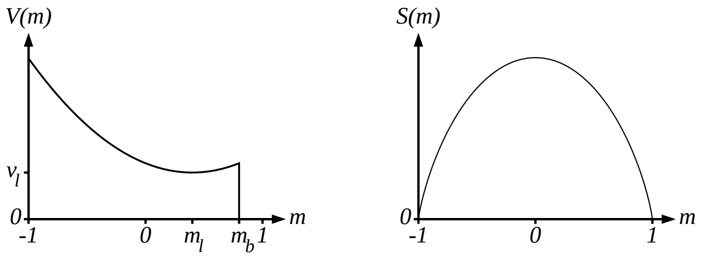

Figure 3.6: Potential and Entropy for the "Basin with a Barrier".

The number of states, $\mathcal{N}(M)$, with magnetization M is given by

$$\mathcal{N}(M) = \operatorname*{Tr}_{\{S_i\}} \delta\left(M - \sum_{i=1}^{N} S_i^\alpha\right) = \binom{N}{(N-2M)/2} \tag{3.4.3}$$

The entropy is

$$S(M) = \log\left(\mathcal{N}(M)\right) \approx -N\left(\frac{1-2m}{2}\log\left(\frac{1-2m}{2}\right) + \frac{1+2m}{2}\log\left(\frac{1+2m}{2}\right)\right) \tag{3.4.4}$$

This is shown in the right graph of figure 3.6. Thus there is a large free energy barrier separating the local and global minima. The problem is to over come this barrier.

This potential is intended to capture a common feature of many hard optimization problems, namely a global minima surrounded by many local minima close to the global minima. There is often a whole hierarchy of such local minima. The further these local minima are from the global minima the higher their energy will have.

3.4.1 Stochastic Hill Climbing

In stochastic hill climbing we consider a "walker" who starts at some randomly chosen configuration and walks around the fitness landscape. Since there is only a single walker we can drop the member index α. For this problem there in no point in annealing the temperature as the walker will find the local minima very quickly. The problem is to overcome the large potential barrier through thermal fluctuations and end up in the well.

We consider a Monte Carlo dynamics in which

$$S_i \to S_i' = \begin{cases} -S_i & \text{with prob. } W(S_i, M) \\ S_i & \text{with prob. } 1 - W(S_i, M) \end{cases} \tag{3.4.5}$$

where, for $m \le m_b$,

$$W(S_i, M) = \frac{1}{2}\left(1 - \tanh\left(\frac{\beta \Delta E(S_i, M)}{2}\right)\right) \tag{3.4.6}$$

and

$$\Delta E(S_i, M) = E(M - 2\, S_i) - E(M) \tag{3.4.7}$$

This dynamics defines a Markov process (or random walk on line where the transition probabilities depend on where the walker is). In the language of Markov process the global minima is an **absorbing boundary** and we ask what is the expected time to reach this state. To calculate this we define

$$T(M) \equiv \text{expected time to reach the well starting from M.}$$

We can obtain a recursion relation for $T(M)$ by equating $T(M)$ with one plus the expected time to reach the well one time step after being at a point M. Thus

$$\begin{aligned} T(M) \quad &= \quad 1 + \frac{(N-2M)}{2N}\left[W(-1,M)\, T(M+2) + (1 - W(-1,M))\, T(M)\right] \\ &+ \frac{(N+2M)}{2N}\left[W(1,M)\, T(M-2) + (1 - W(1,M))\, T(M).\right] \end{aligned} \tag{3.4.8}$$

The expected time to reach the well one time step after being at a point M is made up of four terms

1. Probability of choosing an allele $S_i^\alpha = -1$ and accepting the mutation.
2. Probability of choosing an allele $S_i^\alpha = -1$ and rejecting the mutation.
3. Probability of choosing an allele $S_i^\alpha = 1$ and accepting the mutation.
4. Probability of choosing an allele $S_i^\alpha = 1$ and rejecting the mutation.

To solve this equation we subtracting $T(M)$ from both sides and define

$$b_M = T(M) - T(M-2). \tag{3.4.9}$$

The recursion relation (3.4.8) reduces to

$$2N - (N-2M)\, W(-1,M)\, b_{M+2} + (N+2M)\, W(1,M)\, b_M = 0. \tag{3.4.10}$$

We can solve this recursion relation subject to two boundary conditions

$$T(M) = 0, \qquad M > M_b \tag{3.4.11}$$

and

$$T(2) - T(0) = b_2 = W^{-1}(-1, 0). \tag{3.4.12}$$

To find the average time, $\langle T \rangle$, it takes to end up in the global minimum we average over all initial conditions

$$\langle T \rangle = \frac{1}{2^N} \sum_{M=0}^{M \le M_B} \binom{N}{M} T(M). \tag{3.4.13}$$

This is easy to put on a computer and solve exactly. The time taken to reach the well will depend on the "inverse temperature" β. For the case when $m_l = 0.6$ and $m_b = 0.8$ the optimal average times to reach the well for systems of size $N = 16$, 32 and 64 are $\langle T \rangle = 2389.43$, 869780 and 4.89987×10^{10}.

To calculate how these numbers scale we can evoke a thermodynamic arguement. The walker will to a good approximation thermalize in the local minima so the probability of the walker having a magnetization m_b is given by the ratio of the free energy of the barrier to the free energy at equilibruim m_{equil}

$$\frac{1}{\langle T \rangle} \approx \mathrm{e}^{F(m_b) - F(m_{equil})} \tag{3.4.14}$$

where

$$F(m) = V(m) - TS(m). \tag{3.4.15}$$

The equilibrium magnetization is given by the condition that the free energy is minimized

$$\left.\frac{\partial F(m)}{\partial m}\right|_{m=m_{equil}} = 0. \tag{3.4.16}$$

In the large N limit this becomes exact. Solving these equations numerically we find

$$\langle T \rangle \propto \mathrm{e}^{0.285781\,N}. \tag{3.4.17}$$

3.4.2 Genetic Algorithms

We can ask how a Genetic Algorithm behaves on the same problem. A GA with just mutation and no crossover (which many GA afficionados would refuse to call a GA) would perform terribly. The population would move to the local minima and get stuck. The chances of a member of the population climbing over the barrier is strongly suppressed by the presence of the other members of the population. The time for a population to overcome the barrier is likely to scale as

$$\langle T \rangle \propto \mathrm{e}^{c\,P\,N} \tag{3.4.18}$$

where c is some constant. *Populations are conservative!* They know about their absolute position through the other members of the population in contrast to the single walker who only knows the local gradient.

What about the case of a GA with crossover. Here the situation is entirely different. The population again reaches the bottom of the local minima but now it can jump the energy barrier straight into the well. This is still exponentially difficult because of the entropy barrier, but the situation is far less severe than for the single walker. For large N the time for a GA to find the well will scale something like

$$\frac{1}{\langle T \rangle} \approx \mathrm{e}^{S(m_b) - S(m_l)} \tag{3.4.19}$$

or for the same potential used above

$$\langle T \rangle \propto \mathrm{e}^{0.175319\,N}. \tag{3.4.20}$$

This is of course an optimal asymptotic performance which a real GA would not achieve because of correlations. (A more accurate estimate of the performance of a GA is given in reference [37].) Nevertheless the GA substantially outperforms stochastic hill climbing on this problem.

This potential was hand constructed and clearly favours the GA. In practice a GA outperforms stochastic hill climbing even on fairer potentials with a more rounded barrier. A cluster algorithm in which many alleles are mutated at once might well perform similarly to this ideal GA.