SOME EXACT RESULTS FROM A COARSE GRAINED FORMULATION OF GENETIC DYNAMICS

C. R. Stephens
NNCP, Instituto de Ciencias Nucleares,
UNAM, Circuito Exterior, A.Postal 70-543
México D.F. 04510
e-mail: stephens@nuclecu.unam.mx

Abstract

We extend a recently developed exact schema based, or coarse grained, formulation of genetic dynamics [8, 9, 10] and its associated exact Schema theorem to an arbitrary selection scheme and a general crossover operator. We show that the intuitive"building block" interpretation of the former is preserved leading to hierarchical formal solutions of the equations that upon iteration lead to new results for the limiting distribution of a population in the case of 1-point crossover and "weak" selection, where we define quantitatively "weak". We also derive an exact, analytic form for the population distribution as a function of time for a flat landscape and 1-point crossover.

1 Introduction

Developing a better qualitative and quantitative understanding of the theory of genetic dynamics, and thereby of Genetic Algorithms (GAs), remains a challenging problem by any measure. The benefits are almost too obvious to mention. However, it is worth emphasizing the scale of the task. Almost all known theoretical results are for systems with a small number of loci and even in the case of very simple landscapes very little is known as far as explicit, rather than formal or numerical, results is concerned. In particular we are unaware of any systematic schemes for developing approximate solutions such as perturbative methods.

Passing beyond canonical results, such as Holland's Schema theorem [1], which is associated with an inequality for the dynamics rather than an equality, various exact evolution equations have been derived previously: Goldberg and Bridges [2] wrote down exact equations for two-bit problems. These equations allowed for an explicit analysis of string gains and losses. Whitley and Crabbe [3] also presented an algorithm for generating evolution equations for larger problems that was equivalent to the earlier equation of [2]. Although exact these equations are extremely unwieldy and it is difficult to infer general conclusions from their analysis.

Another related approach is that of Vose and collaborators [4] that treats GA evolution as a Markov chain. Noteworthy, and less familiar in the GA community, is analogous work in mathematical evolution theory [5] (see also the significant article of Altenberg [6] that has a foot in both camps). This formulation of GA dynamics appears to be so removed from elements such as the Schema theorem and Building Block hypothesis that the latter have very much fell out of favour with aficionados of this dynamical formulation. Also, being fundamental, microscopic equations they do not lend themselves easily to a treatment of schemata. It is important to note that this is not just an academic point: genetic dynamics with a large number of degrees of freedom is so complex that a formalism that treats the effective degrees of freedom at a more macroscopic level is certainly required. Such a formalism could be postulated directly, such as in the effective theory of Shapiro and collaborators [7], or better, derived from the underlying microscopic dynamics.

Evolution equations that offer a very intuitive interpretation, illuminate the content of the Schema theorem and the Building Block hypothesis, naturally coarse grain from string equations to schema equations, yield an interpolation between the microscopic and the macroscopic and that offer new exact results or simpler proofs of known results have been recently derived [8, 9, 10]. Originally developed for a canonical GA the basic elements have also recently been extended to Genetic Programming (GP) by Poli and coworkers [11, 12] who have showed the utility of the formalism by deriving several interesting new results.

In this paper we generalize the formalism of [8, 9, 10], to arbitrary selection schemes and a general crossover operator showing that the basic advantages of the former are preserved. We see that evolution can be fruitfully viewed as the hierarchical construction of more and more complex building blocks. The usefulness of crossover in aiding search is governed by linkage disequilibrium coefficients associated with selection probabilities rather than gene frequencies. We consider the limit distributions of a population showing that in the long time limit an arbitrary schema reaches Robbins proportions but with a coefficient different than one. We also derive an exact, analytic form for the dynamics of a GA for a flat landscape and 1-point crossover.

2 Microscopic String Evolution Equations

In this section we recapitulate the results of [8, 9, 10], taking the opportunity to show their explicit generalization to a general crossover operator and arbitrary selection. We start with an exact evolution equation that relates string proportions, $P(C_i,t) = n(C_i,t)/n$, for a genotype C_i consisting of binary alleles, at time t to those at $t+1$, where $n(C_i,t)$ is the number of strings C_i in the population at time t and n is the population size. In the limit $n \to \infty$ the $P(C_i,t)$ give the probability distribution for the population dynamics and satisfy

$$P(C_i,t+1) = \mathcal{P}(C_i)P_c(C_i,t)+\sum_{\not C_i}\mathcal{P}(C_j\to C_i)P_c(C_j,t) \quad (1)$$

where, for an infinite population, $P_c(C_i,t)$ is the expected proportion of strings of type C_i after selection and crossover.

The effective mutation coefficients are: $\mathcal{P}(C_i)$ which is the probability that string i remains unmutated, and $\mathcal{P}(C_j\to C_i)$, the probability that string j is mutated into string i. In the limit where the mutation probability, p_m, is uniform, $\mathcal{P}(C_i) = (1-p_m)^N$ and $\mathcal{P}(C_j\to C_i) = p_m^{d^H(i,j)}(1-p_m)^{N-d^H(i,j)}$, where $d^H(i,j)$ is the Hamming distance between the strings C_i and C_j. Note that for a finite population the left hand side of (1) is the expected proportion of genotype C_i to be found at $t+1$ while any $P(C_i,t)$ on the right hand side are to be considered as the actual proportions found at t.

Explicitly $P_c(C_i,t)$ is given by

$$\begin{aligned}P_c(C_i,t) &= P'(C_i,t)\\ &-\sum_{m=1}^{2^N}p_c(m)\sum_{C_j\neq C_i}\mathcal{C}^{(1)}_{C_iC_j}(m)P'(C_i,t)P'(C_j,t)\\ &+\sum_{m=1}^{2^N}p_c(m)\sum_{C_j\neq C_i}\sum_{C_l\neq C_i}\mathcal{C}^{(2)}_{C_jC_l}(m)P'(C_j,t)P'(C_l,t) \quad (2)\end{aligned}$$

where $\sum_{m=1}^{2^N}$ is the sum over all possible crossover masks $m \in \mathcal{M}$, $\mathcal{M}$ being the space of masks, and $p_c(m)$ is the probability to implement the mask m. $P'(C_i,t)$ is the probability that genotype C_i is selected and the coefficients $\mathcal{C}^{(1)}_{C_iC_j}(m)$ and $\mathcal{C}^{(2)}_{C_jC_l}(m)$, represent the probabilities that, given that C_i was one of the parents, it is destroyed by the crossover process, and the probability that given that neither parent was C_i it is created by the crossover process.

In order to write these probabilities more explicitly we denote the set of alleles inherited by an offspring from parent C_j as $\mathcal{S}$ and the alleles inherited from parent C_l, i.e. the set $C_l - \mathcal{S}$, by $\mathcal{C}$. Naturally, $\mathcal{S}$ and $\mathcal{C}$ both depend on the particular crossover mask chosen. Then,

$$\mathcal{C}^{(1)}_{C_iC_j}(m) = \theta(d^H_{\mathcal{S}}(i,j))\theta(d^H_{\mathcal{C}}(i,j)) \quad (3)$$

and

$$\begin{aligned}\mathcal{C}^{(2)}_{C_jC_l}(m) = \frac{1}{2}[&\delta(d^H_{\mathcal{S}}(i,j))\delta(d^H_{\mathcal{C}}(i,l))\\ &+\delta(d^H_{\mathcal{C}}(i,j))\delta(d^H_{\mathcal{S}}(i,l))] \quad (4)\end{aligned}$$

where $d^H_{\mathcal{S}}(i,j)$ is the Hamming distance between the strings C_i and C_j measured only over the set $\mathcal{S}$, with the other arguments in (3) and (4) being similarly defined. $\theta(x) = 1$ for $x > 0$ and is 0 for $x = 0$, whilst $\delta(x) = 0 \ \ \forall x \neq 0$ and $\delta(0) = 1$. The equations (1) and (2) yield an exact expression for the expectation values, $n(C_i,t)$, and in the limit $n \to \infty$ yield the correct probability distribution governing the GA evolution for arbitrary selection, mutation and crossover. It takes into account exactly the effects of destruction and construction of strings and, at least at the formal level, is either a generalization of or is equivalent to other exact formulations of GA dynamics.

As an explicit example, consider 1-point crossover. In this case there are only $N-1$ pertinent crossover masks labelled by the crossover point k. Then, $p_c(m) = p_c/N-1$ for $m = k$, $k \in [1, N-1]$ and $p_c(m) = 0$ otherwise. Also, $\mathcal{S} = L$ and $\mathcal{C} = R$ (or vice versa) where L and R refer to the parts of the string to the left and right of the crossover point respectively. For 2-point crossover there are ${}^{N-1}C_2$ non-zero masks labelled by two crossover points, k_1 and k_2. Then $p_c(m) = p_c/{}^{N-1}C_2$ for $m \in \{k_1, k_2\}$ with $k_1,\ k_2 \in [1, N-1]$

and $k_1 > k_2$. In this case $\mathcal{S}$ represents the parts of the string outside k_1 and k_2 and $\mathcal{C}$ the part between them (or vice versa). As a final example, for uniform crossover $p_c(m) = p_c\mu^{N_S}(1-\mu)^{N-N_S}/2^N$ where μ is the probability that a given allele is inherited from parent c_j, $\mathcal{S}$ being the set of alleles inherited from c_j.

Computationally, the above representation is very redundant. The matrix $\mathcal{C}^{(2)}_{c_j c_l}(m)$ has dimensionality $(2^N-1)\times(2^N-1)$ but only $(2^{N_S}-1)\times(2^{N_C}-1)$ matrix elements are non-zero where N_S is the number of alleles of c_i inherited from one parent and N_c the number from the other. As a consequence, and also for other reasons, iterating the dynamics is exceedingly difficult. The equations (1) and (2) offer little in terms of intuitive understanding of the evolutionary dynamics and, arguably, not a great deal in terms of formal results, though Vose and coworkers have made important contributions in this area. Intuitive elements of GA theory such as the Building block hypothesis seem to play no role here and are not apparent in any degree. Additionally, a more general formulation giving the dynamics of schemata is relatively unnatural in this formulation.

3 Coarse Grained Evolution Equations

Far greater progress can be made by changing basis from a microscopic string basis to a coarse-grained basis associated with schemata. Such a basis emerges very naturally anyway from equation (2).

To see this, consider first the destruction term. For a given crossover mask the matrix (3) restricts the sum over the strings c_j to those that differ from c_i in at least one bit both in $\mathcal{S}$ and in $\mathcal{C}$. One can convert the sum over c_j into an unrestricted sum by subtracting off those c_j that have $d^H_S(i,j)=0$ and/or $d^H_c(i,j)=0$. We then use the fact that $\sum_{C_j} P'(c_j,t)=1$ to gain a tremendous simplification of the destruction term. Similarly one may write unrestricted sums for the construction term by subtracting off explicitly those terms that have been added. The unrestricted construction term then becomes

$$\sum_{c_j\supset c_i^S}\sum_{c_l\supset c_i^c} P'(c_j,t)P'(c_l,t) \qquad (5)$$

where c_i^S is the part of c_i inherited from c_j, i.e. the alleles $\mathcal{S}$, while c_i^c is the part inherited from c_l, i.e. the alleles $\mathcal{C}$. Obviously, both c_i^S and c_i^c are schemata. Noting that the extra terms originating from the conversion of the restricted sums to unrestricted sums in the destruction and construction terms exactly cancel one can finally rewrite $P_c(c_i,t)$ as

$$P_c(c_i,t) = P'(c_i,t)(1-p_c) + \sum_{m=1}^{2^N} p_c(m)P'(c_i^S(m),t)P'(c_i^c(m),t) \qquad (6)$$

where $p_c = \sum_{m=1}^{2^N} p_c(m)$ is the probability to implement crossover (irrespective of the mask) and

$$P'(c_i^S(m),t) = \sum_{c_j\supset c_i^S} P'(c_j,t) \qquad (7)$$

and similarly for $P'(c_i^c(m),t)$. In the absence of mutation $P_c(c_i,t) = P(c_i,t+1)$. It is important to note here that in this form the evolution equation shows that crossover explicitly introduces the idea of a schema and the consequent notion of a coarse graining relating a string c_i to its building blocks, c_i^S and c_i^c, which are schemata of order N_S and $N_c = N - N_S$ respectively. Notice by going to this coarse-grained basis that the potentially $(2^N-1)\times 2^N$ destruction terms, (2^N-1) being the number of terms in the restricted string sum and 2^N being the number of possible crossover masks, have been reduced to only one term and that for a given crossover mask the $(2^N-1)^2$ construction terms have been converted into one single crossover term $P'(c_i^S(m),t)P'(c_i^c(m),t)$. Actually, this somewhat overstates the case as the schema averages contain (2^N_S-1) and (2^N_c-1) terms respectively. However, at the very least the change in basis has explicitly eliminated the zero elements of $\mathcal{C}^{(2)}_{c_j c_l}(m)$.

So let's recapitulate how the three different operators - mutation, crossover and selection - enter in this exact dynamics. First of all, selection is embodied in the factor P' which is a "black box" for the purposes of the dynamics. That is to say, the highly simplified and elegant form of the dynamical equation (6) is independent of the type of selection used and is a direct consequence of the form invariance of P' under a coarse graining. In other words the relationship between $P'(c_i,t)$ at the string level and $P'(c_i^S,t)$ at the coarse grained level is strictly linear, i.e. $P'(c_i^S,t) = \sum_{C_i\supset C_i^S} P'(c_i,t)$. However, if one wishes to know how the functional dependence on P dynamically changes then one must implement a particular selection scheme. For example, with proportional selection $P'(c_i,t) = (f(c_i)/\overline{f}(t))P(c_i,t)$ where $f(c_i)$ is the fitness of string c_i and $\overline{f}(t)$ is the average population fitness. It is easy to demonstrate in this case that $P'(c_i^S,t) = (f(c_i^S,t)/\overline{f}(t))P(c_i^S,t)$, where $f(c_i^S,t)$ is the fitness of the schema c_i^S, which demonstrates the form invariance of proportional selection. Thus, the form invariance in this case also holds at the level of the Ps not just the P's. This will not generally be the case. For example with a "non-linear" selection scheme where $P'(C_i,t) = a(C_i,t)P(C_i,t) + b(C_i,t)P^2(C_i,t)$ there is no natural form invariance at the level of the Ps as $P'(C_i^s,t) \neq a(C_i^s,t)P(C_i^s,t) + b(C_i^s,t)P^2(C_i^s,t)$. We also see that the elegance of the equation (6) is independent of the type of crossover implemented. The details of the crossover mechanism enter in $p_c(m)$ and also determine which alleles are included in $\mathcal{S}$.

To iterate the equation it is clear that we need to know the dynamics of the schemata C_i^s and C_i^c. Thus we need to obtain an equivalent equation for an arbitrary schema, ξ of order N_2 and defining length l. To do this we need to sum with $\sum_{C_i\supset\xi}$ on both sides of the equation (1). This can simply be done to obtain [8, 9, 10]

$$P(\xi,t+1) = \mathcal{P}(\xi)P_c(\xi,t) + \sum_{\not\xi_i} \mathcal{P}(\not\xi_i\to\xi)P_c(\not\xi_i,t) \quad (8)$$

where the sum is over all schemata, $\not\xi_i$, that differ by at least one bit from ξ in one of the N_2 defining bits of ξ. In other words any schema competing with ξ and belonging to the same partition. $P_c(\xi,t) = \prod_{k=1}^{N_2}(1-p_m(k))$ is the probability that ξ remains unmutated and $P_c(\not\xi_i,t) = \prod_{k\in\{\xi-\not\xi_i\}} p_m(k) \prod_{k\in\{\xi-\not\xi_i\}_c}(1-p_m(k))$ is the probability that the schema $\not\xi_i$ mutate to the schema ξ. $P_c(\xi,t) = \sum_{C_i\supset\xi} P_c(C_i,t)$ is the probability of finding the schema ξ after selection and crossover. Note the form invariance of the equation. i.e. (8) has exactly the same form as (1). To complete the transformation to schema dynamics we need the schema analog of (6). This also can be obtained by acting with $\sum_{C_i\supset\xi}$ on both sides of the equation. One obtains

$$P_c(\xi,t) = (1 - p_c\frac{N_{\mathcal{M}_r}(l,N_2)}{N_{\mathcal{M}}})P'(\xi,t)$$
$$+ \sum_{m\in\mathcal{M}_r(l,N_2)} p_c(m)P'(\xi^s(m),t)P'(\xi^c(m),t) \quad (9)$$

where ξ^s represents the part of the schema ξ inherited from the first parent and ξ^c that part inherited from the second. $N_{\mathcal{M}_r}(l,N_2)$ is the number of crossover masks that affect ξ, $\mathcal{M}_r$ being the set of such masks. $N_{\mathcal{M}}$ is the total number of masks with $p_c(m)\neq 0$. Obviously these quantities depend on the type of crossover implemented and on properties of the schema such as defining length.

Once again for the purposes of illustration we may consider some specific examples. For 1-point crossover: $N_{\mathcal{M}} = N-1$, $N_{\mathcal{M}_r} = l-1$, $\sum_{m\in\mathcal{M}_c} p_c(m) \to p_c/N-1\sum_{k=1}^{l-1}$, where k is the crossover point, and $\xi^s = \xi^L(k)$ the part of ξ to the left of the crossover point and $\xi^c = \xi^R(k)$ is the part to the right. Substituting into (9) one finds the results of [8, 9, 10]. For m-point crossover: $N_{\mathcal{M}} = {}^{N-1}C_m$, $N_{\mathcal{M}-r} = ({}^{N-1}C_m - {}^{N-l}C_m)$ and $\sum_{m\in\mathcal{M}_r} p_c(m) \to p_c \sum_{k_1>k_2>\ldots>k_m}$, where $k_1,\ldots,k_m$ are the m crossover points.

4 Schema Theorem

With the exact evolution equations in hand we can extend the exact Schema theorem of [8, 9, 10] to a general crossover operator and arbitrary selection scheme. Once again we state it through the concept of effective fitness [8, 9, 10]

Exact "coarse grained" Schema theorem

$$P(\xi,t+1) = \frac{f_{\rm eff}(\xi,t)}{\bar{f}(t)}P(\xi,t) \quad (10)$$

where

$$f_{\rm eff}(\xi,t) = \Big(\mathcal{P}(\xi)P_c(\xi,t)$$
$$+ \sum_{\not\xi_i} \mathcal{P}(\not\xi_i\to\xi)P_c(\not\xi_i,t)\Big)\frac{\bar{f}(t)}{P(\xi,t)} \quad (11)$$

and $P_c(\xi,t)$ is given by equation (9). The interpretation of this equation is as for its analog for 1-point crossover and proportional selection [8, 9, 10]: that schemata of higher than average effective fitness are allocated an "exponentially" increasing number of trials over time. An intuitive version of the Building Block hypothesis is part and parcel of this Schema theorem: (9) explicitly shows the competition between schema destruction and construction and shows how a schema is constructed from its building blocks ξ^s and ξ^c which in their turn are composed of building blocks $\xi^{s,s}$ and ξ^{sc} and $\xi^{c,s}$ and ξ^{cc} respectively, which in their turn... The particular building blocks depend of course on the particular mask. Note that the crossover dependent part of (9) for a given crossover mask can be written as

$$\Delta(\xi,m) = P'(\xi,t) - P'(\xi^s(m),t)P'(\xi^c(m),t) \quad (12)$$

which is a selection weighted linkage disequilibrium coefficient and, other than in a flat landscape, is a more relevant quantity than the standard linkage disequilibrium coefficient which is associated with the Ps rather than the P's. If $\Delta > 0$ then crossover has a negative effect, i.e. schema destruction dominates while if $\Delta < 0$ crossover is positive. Thus, whether the selection probabilities for the building blocks of a given schema are correlated or anticorrelated determines the nature of crossover. It is important to note that the hierarchical structure inherent in the iterative solution of (6) relating a string or schema to its more coarse grained antecedents terminates after N_2 steps when one arrives at the maximally coarse grained effective degrees of freedom - 1-schemata. These play a priviliged role as they cannot be destroyed by crossover and in the continuous time limit satisfy $P(\xi^{(i)},t) = \exp\int_0^t((f(\xi,t')/\bar{f}(t)) - 1)dt'$.

One of the strengths of the present coarse grained formulation is that, as we shall see, much can be deduced simply by inspection of the basic formulas. We shall first of all put the basic equation into a yet more elegant form. We introduce a 2^N-dimensional population vector, $\mathbf{P}(t)$, whose elements are $P(c_i,t)$, $i = 1, ..., 2^N$. Equation (1) can then be written in the form

$$\mathbf{P}(t+1) = \overline{\mathbf{W}}\mathbf{P}_c(t) \tag{13}$$

where the mutation matrix $\overline{\mathbf{W}}$ is real, symmetric and time independent and has elements $\overline{\mathbf{W}}_{ij} = p_m^{d^H(i,j)}(1-p)^{N-d^H(i,j)}$, where for simplicity we restrict now to the case of uniform mutation (the generalization to non-uniform mutation is straightforward). Restricting attention to the case of selection schemes linear in $P(c_i,t)$, $\mathbf{P}_c(t)$ can be written as

$$\mathbf{P}_c(t) = \overline{\mathbf{F}}(t)\mathbf{P}(t) + \sum_{m=1}^{2^N} p_c(m)\mathbf{j}(m,t) \tag{14}$$

where the selection - crossover destruction matrix, $\overline{\mathbf{F}}(t)$, is diagonal, and takes into account selection and the destructive component of crossover. Explicitly, for proportional selection $\overline{\mathbf{F}}_{ii}(t) = (f(c_i)/\bar{f}(t))(1-p_c)$. Finally, the "source" matrix is given by $\mathbf{j}(m,t) = P'(c_i^s(m),t)P'(c_i^c(m),t)$. Defining the selection-crossover destruction-mutation matrix $\overline{\mathbf{W}}_s(t) = \overline{\mathbf{W}}\overline{\mathbf{F}}(t)$ we have

$$\mathbf{P}(t+1) = \overline{\mathbf{W}}_s(t)\mathbf{P}(t) + \sum_{m=1}^{2^N} p_c(m)\overline{\mathbf{W}}\mathbf{j}(m,t) \tag{15}$$

The interpretation of this equation is that $\mathbf{j}(m,t)$ is a source which creates strings (or schemata) by bringing building blocks together while the first term on the right hand side tells us how the strings themselves are propagated into the next generation, the destructive effect of crossover renormalizing the fitness of the strings.

5 Formal Solutions of the Basic Equations

Needless to say solutions of these dynamical equations are hard to come by. They represent, for binary alleles, 2^N coupled non-linear difference equations, or in the continuous time limit - differential equations. As shown in [8, 9, 10], compared to a representation based on (2), even a formal solution of (6) in the absence of mutation and for 1-point crossover and proportional selection yields much valuable qualitative information, such as a simple proof of Geiringer's theorem [13] and an extension of it to the weak selection regime. Here we extend this formal solution to the case of general crossover and mutation and for any selection scheme linear in $P(c_i,t)$. The iterated solution of (15) is

$$\mathbf{P}(t) = \prod_{n=0}^{t-1} \overline{\mathbf{W}}_s(n)\mathbf{P}(0)$$
$$+ \sum_{m=1}^{2^N} p_c(m) \sum_{n=0}^{t-1} \prod_{i=n+1}^{t-1} \overline{\mathbf{W}}_s(i)\overline{\mathbf{W}}\,\mathbf{j}(m,n) \tag{16}$$

The mutation matrix $\overline{\mathbf{W}}$ being real and symmetric is diagonalizable. $\overline{\mathbf{W}}_s(t)$ is not generically symmetric and therefore not diagonalizable. This solution actually holds true for arbitrary schemata. The only changes are that the vectors are of dimension 2^{N_2}, the matrices of dimension $2^{N_2} \times 2^{N_2}$, the sum over masks for the construction terms is only over the set $\mathcal{M}_r$ and that the building blocks in $j(m,t)$ are those of the schema rather than the entire string.

The interpretation of (16) follows naturally from that of (15). Considering first the case without mutation, the first term on the right hand side gives us the probability for propagating a string or schema from $t = 0$ to t without being destroyed by crossover. In other words $\prod_{n=0}^{t-1} \overline{\mathbf{W}}_s(n)$ is the Greens function or propagator for $\mathbf{P}$. In the second term, $\widehat{\mathbf{j}}(m,n)$, each element is associated with the creation of a string or schema at time n via the juxtaposition of two building blocks associated with a mask m. The factor $\prod_{i=n}^{t-1} \overline{\mathbf{W}}_s(i)$ is the probability to propagate the resultant string or schema without crossover destruction from its creation at time n to t. The sum over masks and n is simply the sum over all possible creation events in the dynamics. This formulation lends itself very naturally to a diagrammatic representation that will be discussed elsewhere. The role of mutation is to mix the strings or schemata created in the aforementioned process.

So, what can be deduced from (16)? First of all let's consider the case of a flat fitness landscape with no mutation; then $\widehat{\mathbf{P}}(t) = \mathbf{P}(t)$ and $\overline{\mathbf{W}}$ is the unit matrix, $\mathbf{1}$, and $\overline{\mathbf{W}}_s = (1-p_c)\mathbf{1}$. Then, for an arbitrary schema ξ

$$\mathbf{P}(t) = (1 - p_c\frac{N_{\mathcal{M}_r}(l,N_2)}{N_{\mathcal{M}}})^t\mathbf{P}(0)$$

$$+(1-p_c\frac{N_{\mathcal{M}_r(l,N_2)}}{N_{\mathcal{M}}})^t\sum_{m\in\mathcal{M}_r(l,N_2)}p_c(m)$$
$$\times\sum_{n=0}^{t-1}(1-p_c\frac{N_{\mathcal{M}_r(l,N_2)}}{N_{\mathcal{M}}})^{-(n+1)}\mathbf{j}(m,n) \qquad (17)$$

Now, obviously $\lim_{t\to\infty}(1-p_c\frac{N_{\mathcal{M}_r(l,N_2)}}{N_{\mathcal{M}}})^t = 0$ hence $\mathbf{P}(t)\to 0$ as $t\to\infty$ unless the summation over time leads to a cancellation of this damping factor. Given that the building block constituents of $\mathbf{j}(m,n)$ are associated with damping factors $(1-p_c\frac{N_{\mathcal{M}_r}(l',N_2')}{N_{\mathcal{M}}})^t$ and $(1-p_c\frac{N_{\mathcal{M}_r}(l-l',N_2-N_2')}{N_{\mathcal{M}}})^t$ this can only occur if there is no damping of the consituent building blocks and this only happens if they are 1-schemata as $N_{\mathcal{M}_r}(1,1)=0$, i.e. you can't cut a 1-schema. Given that it's a flat landscape $P(\xi^{(i)},t)=P(\xi^{(i)},0)$ where $P(\xi^{(i)},t)$ is the probability of finding the 1-schema $\xi^{(i)}$ corresponding to the bit i, hence the only term of $\mathbf{j}(m,t)$ which gives a non-zero contribution when integrated is $\prod_{i=1}^{N_2}P(\xi^{(i)},0)$. Thus the limit distribution is

$$P^*(\xi)=\lim_{t\to\infty}P(\xi,t)=\prod_{i=1}^{N_2}P(\xi^{(i)},0) \qquad (18)$$

which is Geiringer's theorem for a general crossover operator.

The only role the type of crossover is playing here is how fast the transient corrections to the limit distribution die out. The damping is controlled by $N_{\mathcal{M}_r}(l,N_2)$, hence the bigger it is the faster the corresponding transient dies out. For example, for entire strings (or schemata of defining length N) for m and m'-point crossover, where $m>m'$

$$\frac{N^{(m)}_{\mathcal{M}_r}(l,N_2)}{N^{(m')}_{\mathcal{M}_r}(l,N_2)}=\left(\frac{1-\frac{\Gamma(N-l+1)\Gamma(N-m)}{\Gamma(N-l-m+1)\Gamma(N)}}{1-\frac{\Gamma(N-l+1)\Gamma(N-m')}{\Gamma(N-l-m'+1)\Gamma(N)}}\right)>1 \qquad (19)$$

where $\Gamma(N)=(N-1)!$. Hence, we can quantify exactly how quickly m-point crossover transients die off relative to m'-point crossover transients.

We can take this further and generalize Geiringer's theorem to non-flat fitness landscapes for selection schemes linear in $P(C_i,t)$ such as proportional selection and for 1-point crossover and without mutation. Without loss of generality we write $P'(\xi,t)=(1+\epsilon\delta(\xi,t))$ where ϵ will serve as a control parameter to determine how "weak" selection is, $\epsilon\delta(\xi,t)$ being the selection pressure for the schema ξ. Weak selection will imply $\epsilon|\delta(\xi,t)|<1\ \forall\xi,\ t$. A 1-schema evolves as

$$P(\xi^{(i)},t)=\mathrm{e}^{\epsilon\Delta(\xi^{(i)},t,t')}P(\xi^{(i)},t') \qquad (20)$$

where $\Delta(\xi^{(i)},t,t')=\int_{t'}^{t}\delta(\xi^{(i)},t'')dt''$ is the integrated selection pressure for the 1-schema.

Theorem

The limit distribution, $P(\xi)^*$, of $P(\xi,t)$ as $t\to\infty$ for "weak" selection and no mutation and 1-point crossover in the continuous time limit is

$$P(\xi)^*=\mathcal{A}(\xi,N,l,p_c)\prod_{i=1}^{N_2}P(\xi^{(i)},0) \qquad (21)$$

where the selection dependent amplitude $\mathcal{A}(\xi,N,l,p_c)$ is given by

$$\mathcal{A}(\xi,N,l,p_c)=$$
$$\lim_{t\to\infty}\int_0^{\frac{p_c(l-1)t}{N-1}}\mathrm{e}^{-\tau}\mathrm{e}^{\left(1-\frac{p_c(l-1)}{N-1}\right)\Delta(\xi,t,t-\frac{N-1}{p_c(l-1)})}\times$$
$$\prod_{i=1}^{N_2}(1+\delta(\xi^{(i)},t-\tfrac{N-1}{p_c(l-1)}))\mathrm{e}^{\sum_{i=1}^{N_2}\Delta(\xi^{(i)},t-\frac{N-1}{p_c(l-1)},0)}d\tau \qquad (22)$$

Note that for a flat landscape $\mathcal{A}\to 1$. Due to space limitations here a proof of this theorem will be presented elsewhere, though we can outline the logic. The key to the theorem is deciding just how weak selection must be in order that all higher order schema transients die out. It is for this reason that a tuning parameter was introduced. For a given weak selection landscape and values of l, N and p_c there exists a critical value of ϵ, $\epsilon_{cr}(\xi,N,l,p_c)$, below which all non-1-schema transients die out as $t\to\infty$. One examines the growth or decay rate of a schema and all its constituent building blocks. Growth is caused by selection and decay by crossover. By examining the dominant growth mode, other than that associated with pure 1-schemata, one can tune ϵ such that it becomes a decay mode. All other growing modes are subdominant and therefore also decay. Hence, the only non-zero mode in the limit $t\to\infty$ is the 1-schema mode of (21). The difference with the result for a flat landscape is that in this case the effective fitness landscape for 1-schemata is not flat. As an example, if one considers a weak counting ones landscape, i.e. where $f(C_i)=1+\epsilon n_1(C_i)$, $n_1(C_i)$ being the number of ones on the string, then $\epsilon_{cr}\sim p_c/N^3$ for strings.

6 Explicit Solutions

To show the further utility of the present formalism we will find the general explicit solution to (6) for a flat fitness landscape, without mutation and for 1-point crossover in the continuous time limit. Lest this be thought a trivial problem it's wise to remember that it still involves the solution of 2^N coupled non-linear differential equations. To illustrate the general principles we'll consider first a three bit problem. The general structure of the equations (6) is that one builds up a solution via intermediate building blocks. The most fundamental blocks are 1-schemata as these cannot be cut and hence transform trivially under recombination (save in the case of uniform crossover). $\mathbf{P}^{(l,N_2)}(t)$ in this case is an N-dimensional vector and $\mathbf{P}^{(l,N_2)}(t) = \mathbf{P}^{(l,N_2)}(0)$, with $l = 1$ and $N_2 = 1$. Explicitly, $P(i**,t) = P(i**,0)$, $P(*j*,t) = P(*j*,0)$ and $P(**k,t) = P(**k,0)$. There are four 2-schemata per crossover point corresponding to $P(ij*,t)$ and $P(*jk,t)$, $i,\ j = 0,\ 1$. $P(ij*,t)$ satisfies

$$
\begin{aligned}
P(ij*,t) &= \mathrm{e}^{-\frac{p_c t}{2}} P(ij*,0) \\
+\frac{p_c}{2}\mathrm{e}^{-\frac{p_c t}{2}} &\int_0^t P(i**,t')P(*j*,t')\mathrm{e}^{\frac{p_c t'}{2}} dt'
\end{aligned}
\tag{23}
$$

with an analogous equation for $P(*jk,t)$. The simple solution of (23) is

$$
\begin{aligned}
P(ij*,t) &= \mathrm{e}^{-\frac{p_c t}{2}} P(ij*,0) \\
+(1-\mathrm{e}^{-\frac{p_c t}{2}}) &P(i**,0)P(*j*,0)
\end{aligned}
\tag{24}
$$

The 3-schema solution is found using (24)

$$
\begin{aligned}
P(ijk,t) &= \mathrm{e}^{-p_c t} P(ijk,0) \\
+\frac{p_c}{2}\mathrm{e}^{-p_c t} &\int_0^t \mathrm{e}^{p_c t'} \left(P(ij*,t')P(**k,t')\right. \\
&\left. +P(i**,t')P(*jk,t')\right) dt'
\end{aligned}
\tag{25}
$$

Substituting the solution (24) into (25) one finds simply

$$
\begin{aligned}
P(ijk,t) &= \mathrm{e}^{-p_c t} P(ijk,0) \\
+\mathrm{e}^{-\frac{p_c t}{2}}(1-\mathrm{e}^{-\frac{p_c t}{2}}) &\left(P(ij*,0)P(**k,0)\right. \\
&\left. + P(i**,0)P(*jk,0)\right) \\
+(1-\mathrm{e}^{-\frac{p_c t}{2}})^2 &P(i**,0)P(*j*,0)P(**k,0)
\end{aligned}
\tag{26}
$$

In the limit $t \to \infty$, $P(ijk,t) \to P(i**,0)P(*j*,0)P(**k,0)$. We see here the approach to Robbin's proportions is exponentially fast with the bigger schemata dying off quicker.

The general solution for an N-bit string is

$$
P(c_i,t) = \sum_{n=0}^{N-1} \mathrm{e}^{-\frac{n p_c t}{N-1}} (1-\mathrm{e}^{-\frac{p_c t}{N-1}})^{N-n-1} \mathcal{P}(n+1) \tag{27}
$$

where $\mathcal{P}(n+1)$ is an initial condition and represents a partition over the probabilities for finding $N-n$ building blocks at $t = 0$. For a given n there are ${}^{N-1}C_n$ such terms. Equation (26) offers a simple illustration where $N = 3$, hence $n = 0,\ 1,\ 2$. $n = 0$ corresponds to the N building block terms of which there are ${}^{N-1}C_0 = 1$ term, $P(i**,0)P(*j*,0)P(**k,0)$. $n = 1$ corresponds to the $N-1$ building block terms of which there are ${}^{N-1}C_1 = 2$ terms, $P(ij*,0)P(**k,0)$ and $P(i**,0)P(*jk,0)$. Finally, $n = 2$ corresponds to the $N-2$ building block terms of which there are ${}^{N-1}C_2 = 1$ term, $P(ijk,0)$.

Notice that (27) gives not only the asymptotic behaviour but also the complete transients. We are unaware of any similar result in the literature. It is not difficult to prove that (27) is the general solution by showing that it satisfies (6). This requires the solutions of (6) for $c^L(k)$ and $c^R(k)$, the building blocks of c_i also. From the form invariance of the equations the solution of (6) is

$$
P(c_i^L,t) = \sum_{n_L=0}^{l_L(k)-1} \mathrm{e}^{-\frac{n_L p_c t}{N-1}} (1-\mathrm{e}^{-\frac{p_c t}{N-1}})^{N-n_L-1} \mathcal{P}(n_L+1) \tag{28}
$$

where $N-n_L$ is the number of building blocks of $c_i^L(k)$ and $l_L(k)$ is the defining length of $c_i^L(k)$. Using the analogous equation for $c_i^R(k)$ and the fact that $n = n_L + n_R$ one sees that both sides of (6) have the same time dependence hence it suffices to prove the equivalence at $t = 0$. The equality at $t = 0$ hinges on the identity that $\mathcal{P}(n+1) = \sum_{n_L=0}^{n-1} \mathcal{P}(n_L+1)\mathcal{P}(n_R+1)$.

In the case of zero mutation the other exact limit of the evolution equations is that of pure selection. Interestingly, the solution is only exact in the case of discrete time. With these two exact limits in hand there should be scope for obtaining perturbative solutions to the evolution equations either in powers of ϵ for weak selection or in powers of p_c for strong selection.

7 Conclusions

In this article we have generalized the formalism of [8, 9, 10] to cover arbitrary selection schemes and a general crossover operator. We have tried to emphasize the advantages of this formulation and in particular show how these advantages have concrete payoffs. We believe that our coarse grained formulation is more intuitive in its content than others, giving an exact schema theorem that contains in a very obvious manner a Building Block hypothesis. Its coarse grained hierarchical structure, both in terms of time and building block complexity, leads to a formulation wherein results such as Geiringer's theorem can be seen in such a manifestly simple way that it is essentially proof by inspection. Additionally, the way that selection, schema destruction and schema creation enter in different ways leads, using the same hierarchical structure, to a version of Geiringer's theorem for a class of "weak" selection landscapes where we were able to quantify the meaning of "weak".

The fundamental coarse grained equation show an important form invariance, i.e. after passing from schemata of a given N_2 to more coarse grained schemata of order $N_2' < N_2$, the resulting equations have exactly the same form. This is certainly not manifestly true of the fundamental string equations (2). This form invariance is an important property as it implies that if a solution can be found for one degree of coarse graining then an analogous solution can simply be written down for the more coarse grained degrees of freedom. We showed that the eficacy of the present formulation lies not only in its intutive appeal and the facility with which more general formal results can be deduced but also in how one may derive explicit analytic results, as in the case of an exact solution for the evolution of strings in the case of a flat fitness landscape and 1-point crossover.

We believe that our previously derived results and the results herein are the tip of the iceberg and that the present formalism may serve as a starting point for deriving a whole host of similar results and beyond. To name just a couple of obvious ones: finding the asympotic limiting distributions for other types of crossover and including in mutation and finding the exact dynamics for other types of crossover operator One of the most intriguing possibilities, that we have briefly alluded to, is the possibility of deriving approximate results using a systematic approximation scheme such as perturbation theory. The existence of exact solutions in the flat fitness landscape and the zero crossover limit add weight to such a supposition as does the fact that the iterative solution of the equations (6) leads to a diagrammatic series very similar to the Feynman diagrammatic series that appear in quantum field theory.

References

[1] Holland, J.H. (1975) *Adaptation in natural and artificial systems* (MIT Press, Cambridge, MA).

[2] Bridges, C.L. and Goldberg, D.E. (1987) "An Analysis of Reproduction and Crossover in a Binary-encoded Genetic Algorithm", in: *Genetic Algorithms and their Applications*, ed. John J. Grefenstette, (Lawrence Erbaum Publishers, Hillsdale, NJ) 9-14.

[3] Whitley, D., Das, R. and Crabb, C. (1992) "Tracking Primary Hyperplane Competitors in Genetic Search", *Annals of Mathematics and Artificial Intelligence* **6**, 367-388.

[4] Vose, M.D. and Nix, A. (1992) "Modelling Genetic Algorithms with Markov Chains", *Annals of Mathematics and Artificial Intelligence* **5**, 79-98.

[5] Karlin, S. (1979) "Models of Multifactorial inheritance: I, Multivariate formulations and basic convergence results", *Theoretical Population Biology* **15**, 308-355.

[6] Altenberg, L. (1995) "The Schema Theorem and Price's Theorem", *Foundations of Genetic Algorithms 3*, ed.s D. Whitley and M. Vose, 23-49 (Morgan Kaufmann, San Mateo).

[7] Prühgel Bennett, A. and Shapiro, J. (1994) "An analysis of genetic algorithms using statistical mechanics", *Phys. Rev. Lett.* **72**, 1305-1309.

[8] Stephens, C.R. and Waelbroeck, H. (1997), "Effective Degrees of Freedom in Genetic Algorithms and the Block Hypothesis", *Proceedings of the Seventh International Conference on Genetic Algorithms*, ed. T. Bäck (Morgan Kaufmann, San Mateo) 34-41.

[9] Stephens, C.R. and Waelbroeck, H. (1998) "Analysis of the Effective Degrees of Freedom in Genetic Algorithms", *Physical Review* **D57** 3251-3264.

[10] Stephens, C.R. and Waelbroeck, H. (1999) "Schemata Evolution and Building Blocks", *Evol. Comp.* **7(2)** 109-124.

[11] Poli, R. (1999) "Schema theorems without expectations", *Proceedings of GECCO99*, ed. W. Banzhaf *et al* (Morgan Kaufmann).

[12] Poli, R. (2000) "Hyperschema theory for GP with one point crossover, building blocks and some new results in GA theory", *Genetic Programming, Proceedings of Euro GP 2000*, ed. R. Poli, W. Banzhaf *et al* (Springer Verlag).

[13] Geiringer, H. (1944) "On the Probability Theory of Linkage in Mendelian Heredity", *Annals of Mathematical Statistics* **15**, 25-27.