Geiringer theorems: from population genetics to computational intelligence, memory evolutive systems and Hebbian learning.

Mitavskiy, Boris; Tuci, E; Cannings, Christopher; Rowe, Jonathan; He, Jun

[Link to publication on Research at Birmingham portal](#)

A Schema-theory-based Extension of Geiringer's Theorem for Linear GP and Variable-length GAs under Homologous Crossover

Abstract

In this paper we study, using a schema-theoretic approach, the search biases produced by GP homologous crossovers when applied to linear representations, such as those used in linear GP or in variable length GAs. The study naturally leads to generalisations of Geiringer's theorem and of the notion of linkage equilibrium, which, until now, were applicable only to fixed-length representations. This indicates the presence of a mixing process which pushes the population towards a statistically independent distributions of primitives.

1 Introduction

Schemata are sets of points in a search space sharing some syntactic feature. Typically schema theorems are descriptions of how the proportion of members of the population belonging to a schema vary over time. If $\alpha(H, t)$ denotes the probability that at time t a newly created individual samples (or matches) the schema H, which we term the *total transmission probability* of H, then an exact schema theorem for a generational system is simply

$$E[\Phi(H, t+1)] = \alpha(H, t), \tag{1}$$

where $\Phi(H, t+1)$ is the proportion of individuals sampling H at generation $t+1$ and $E[\cdot]$ is the expectation operator. Holland's (Holland, 1975) and other (e.g. (Poli & Langdon, 1998b)) worst-case-scenario schema theories normally provide a lower bound for $\alpha(H, t)$ or, equivalently, for $E[\Phi(H, t+1)]$. Only recently schema theorems which provide the exact value for $\alpha(H, t)$ have become available for fixed-length GAs with one-point crossover and mutation (Stephens & Waelbroeck, 1997; Stephens & Waelbroeck, 1999) and other homologous crossovers (Stephens, 2001). Even more recent is the development of exact schema theorems for variable-length GAs, linear GP and tree-based GP. These now cover a variety of crossover and mutation operators including one-point crossover (Poli, 2000b; Poli, 2000a; Poli, 2001b; Langdon & Poli, 2002), standard and other subtree-swapping crossovers (Poli, 2001a; Poli & McPhee, 2001b; McPhee & Poli, 2001; Langdon & Poli, 2002), different types of subtree mutation and headless chicken crossover (Poli & McPhee, 2001a; McPhee *et al.*, 2001a), and, finally, the class of homologous crossovers (Poli & McPhee, 2001c). These are a set of operators, including GP one-point crossover (Poli & Langdon, 1997) and GP uniform crossover (Poli & Langdon, 1998a), where the offspring are created preserving the position of the genetic material taken from the parents.

Exact schema theorems provide probabilistic models (the schema evolution equations) of a GA or a GP system that can be used to understand the system and study its behaviour over one, or, under certain assumptions (e.g. infinite populations), multiple generations. Exact models allow, for example, a formal study of the biases of the operators. This can be done either through simulation (i.e., by integrating the equations) or through mathematical analysis. In recent research we have started using the schema equations with the objective of studying and understanding the *length* biases of different crossover and mutation operators (Poli & McPhee, 2001b; McPhee & Poli, 2001; McPhee *et al.*, 2001b; Rowe & McPhee, 2001). These studies have shed some light on an important emergent phenomen in GP known as bloat (McPhee & Miller, 1995; Soule *et al.*, 1996; Langdon *et al.*, 1999). More recently we have started focusing on the biases of the operators with respect to the alleles or primitives in the representation (Poli *et al.*, 2002a; Poli *et al.*, 2002b) finding fixed points for the allele/primitive distribution in the case of linear, but variable-length representations, undergoing subree crossover and homologous crossover. Particularly relevant for the work reported in this paper is (Poli *et al.*, 2002b) where we hypothesised the possibility of extending Geiringer theorem to variable length structures, but we were only able provide a family of fixed points and to show empirically that populations seemed to converge towards it. In this paper we complete the job, by generalising the Geiringer's proof of convergence and clarifying for which homologus operators the family of fixed points is a global attractor for the system.

The paper is organised as follows. We provide some background information on the GP schema theory for homologous crossover and Geiringer's theorem in Sections 2. Then we provide a summary of the main results in (Poli *et al.*, 2002b) since they are the starting point for this paper. So, we simplify the schema theory for the case of linear, but variable-length, structures and use it to study the size evolution in linear systems in Section 3. We provide a fixed point for the allele distribution in the infinite population hypothesis and give the proof of convergence in in Section 4. We discuss this result and we draw some conclusions in Section 5.

2 Background

Syntactically a GP schema is a tree composed of functions from the set $\mathcal{F} \cup \{=\}$ and terminals from the set $\mathcal{T} \cup \{=\}$, where $\mathcal{F}$ and $\mathcal{T}$ are the function and terminal sets used in a GP run. The primitive $=$ is a "don't care" symbol which stands for a *single* terminal or function. A schema H represents the set of all programs having the same shape as H and the same non-$=$ nodes as H. Particularly important for the GP schema theory are schemata containing "don't care" symbols only, since they represent all the programs of a particular shape. Let $G_1, G_2, \cdots$ be an enumeration of such shape-representing schemata.

In GP homologous crossovers the offspring are created by exchanging genetic material (nodes and subtrees) taken from the same position in the parents trees (as formally described below). To account for the possible structural diversity of the two parents, the selection of the nodes and the

roots of the subtrees to swap are constrained to belong to the *common region*. This is the largest rooted region where the two parent trees have the same topology. Formally, the common region between two generic trees h_1 and h_2 is the set

$$C(h_1, h_2) = \{(d, i) | \mathcal{C}(d, i, h_1, h_2)\},$$

where (d, i) is a pair of coordinates in a Cartesian node reference system (see below). The predicate $\mathcal{C}(d, i, h_1, h_2)$ is true if $(d, i) = (0, 0)$ (i.e., if (d, i) is the root node). It also true if $A(d-1, i', h_1) = A(d-1, i', h_2) \neq 0$ and $\mathcal{C}(d-1, i', h_1, h_2)$ is true, where $A(d, i, h)$ returns the arity of the node at coordinates (d, i) in h, $i' = \lfloor i/a_{\max} \rfloor$, $a_{\max}$ is the maximum arity of the functions in the function set, and $\lfloor \cdot \rfloor$ is the integer-part function. The predicate is false otherwise. The notion of common region can be applied to schemata, too. In the following we will denote the common region between programs of shape G_j and programs of shape G_k with $C(G_j, G_k)$.

The Cartesian node reference system used in the definition of common region is obtained by considering an ideal infinite tree consisting entirely of nodes of arity $a_{\max}$ (Poli, 2001*a*; Poli & McPhee, 2001*c*). This maximal tree would include 1 node of arity $a_{\max}$ at depth 0, $a_{\max}$ nodes of arity $a_{\max}$ at depth 1, $(a_{\max})^2$ nodes of arity $a_{\max}$ at depth 2, and generally $(a_{\max})^d$ nodes at depth d. Then one could imagine organising the nodes in the tree into layers of increasing depth and assigning an index to each node in a layer. The layer number d and the index i can then be used to define a Cartesian coordinate system. Clearly, one can also use this reference system to locate the nodes of non-maximal trees. This is possible because a non-maximal tree can always be described using a subset of the nodes and links in the maximal tree.

To complete our formal description of the class of GP homologous crossovers, we need to extend to GP the notions of crossover masks and recombination distributions used in genetics (Geiringer, 1944) and in the GA literature (Booker, 1992; Altenberg, 1995; Spears, 2000). For any given common region c we can define a set of *GP crossover masks*, χ_c, which contains all different trees with the same size and shape as the common region which can be built with nodes labelled 0 and 1 (Poli & McPhee, 2001*c*; Poli *et al.*, 2001). Each crossover mask rapresents one of the ways in which one could generate an offspring through crossover: nodes of the offspring corresponding to internal 1's in the mask will be taken from the first parent, nodes corresponding to internal 0's from the second parent, subtrees of the first parent whose root corresponds to leaves labelled with a 1 in the mask will be transferred to the same position in the offspring, and, finally, subtrees of the second parent whose root corresponds to leaves labelled with a 0 in the mask will be transferred to the same position in the offspring. The *GP recombination distribution* p_l^c gives the probability that, for a given common region c, crossover mask l will be chosen from the set χ_c. Each GP homologous crossover is characterised by a different recombination distribution. For example, the recombination distribution for GP uniform crossover with 50% probability of exchanging nodes is $p_i^c = (0.5)^{N(c)}$, where $N(c)$ is the size of common region c. Since the size and shape of the common region can be inferred from the mask l, in the following we will often omit the superscript c from p_l^c.

Finally, before we introduce the exact schema equation for GP homologous crossover developed in (Poli & McPhee, 2001*c*) we need to define the notion of hyperschema. A *GP hyperschema* is a rooted tree composed of internal nodes from $\mathcal{F} \cup \{=\}$ and leaves from $\mathcal{T} \cup \{=, \#\}$. Again, = is a "don't care" symbols which stands for exactly one node, while # stands for any valid subtree. In the theory we use (functions returning) hyperschemata to represent the characteristics the two parents must have in order for them to produce instances of a particular schema of interest.

When exploiting the symmetries in the process of selection of the parent programs (Poli *et al.*, 2002*b*), the exact schema equations for GP with homologous crossover can be written as

$$\alpha(H, t) = (1 - p_{\text{xo}})p(H, t) + p_{\text{xo}}\alpha_{\text{xo}}(H, t) \tag{2}$$

where:

$$\alpha_{\text{xo}}(H, t) = \sum_j \sum_k \sum_{l \in \chi'_{C(G_j, G_k)}} (p_l + p_{\bar{l}}) \, p(\Gamma(H, l) \cap G_j, t)p(\Gamma(H, \bar{l}) \cap G_k, t); \tag{3}$$

p_{xo} is the crossover probability; $p(H, t)$ is the selection probability of the schema H; $\bar{l}$ is the complement of the GP crossover mask l (the complement of a mask is a tree with the same structure but with the 0's and 1's swapped); $\chi'_{C(G_j, G_k)}$ and $\bar{\chi}'_{C(G_j, G_k)}$ are two non-overlapping sets such that $\chi'_{C(G_j, G_k)} \cup \bar{\chi}'_{C(G_j, G_k)} = \chi_{C(G_j, G_k)}$ and for each mask $x \in \chi'_{C(G_j, G_k)}$, there is a mask $y \in \bar{\chi}'_{C(G_j, G_k)}$ such that $y = \bar{x}$, and vice versa; $\Gamma(H, l)$ is defined to be the empty set if l contains any node not in H. Otherwise it is the hyperschema obtained by replacing certain nodes in H with either = or # nodes:

- If a node in H corresponds to (i.e., has the same coordinates as) a non-leaf node in l that is labelled with a 0, then that node in H is replaced with a =.
- If a node in H corresponds to a leaf node in l that is labelled with a 0, then it is replaced with a #.
- All other nodes in H are left unchanged.

As discussed in (Poli & McPhee, 2001*c*), it is possible to show that, in the absence of mutation, Equations 2 and 3 generalise and refine a variety of earlier GA and GP schema theory results (Holland, 1975; Whitley, 1994; Poli & Langdon, 1997; Poli & Langdon, 1998*b*; Stephens & Waelbroeck, 1997; Stephens & Waelbroeck, 1999; Poli, 2000*b*; Poli, 2000*a*; Stephens, 2001).

Finally, we briefly introduce Geiringer's theorem (Geiringer, 1944), an important result with implications both for natural population genetics and evolutionary algorithms (Booker, 1992; Booker *et al.*, 2000; Spears, 2000). Geiringer's theorem indicates that, in a population of fixed-length chromosomes repeatedly undergoing crossover (in the absence of mutation and selective pressure), the probability of finding a generic string $h_1 h_2 \cdots h_N$ approaches a limit distribution which is only dependent on the distribution of the alleles h_1, h_2, etc. in the initial generation. More

precisely, if $\Phi(h_1 h_2 \cdots h_N, t)$ is the proportion of individuals of type $h_1 h_2 \cdots h_N$ at generation t and $\Phi(h_i, t)$ is the proportion of individuals carrying allele h_i then

$$\lim_{t \to \infty} \Phi(h_1 h_2 \cdots h_N, t) = \prod_{i=1}^{N} \Phi(h_i, 0). \tag{4}$$

If one interprets $\Phi(h_1 h_2 \cdots h_N, t)$ as a probability distribution of the possible strings in the population, we can interpret Equation 4 as saying that such a distribution is converging towards independence. This result is valid for all homologous crossover operators which allow any two loci to be separated by recombination. Strictly speaking the result is valid only for infinite populations. When, at a particular generation t, the frequency of any string in a population $\Phi(h_1 h_2 \cdots h_N, t)$ equals $\prod_{i=1}^{N} \Phi(h_i, t)$, the population is said to be in *linkage equilibrium* or *Robbins' proportions*.

It is trivial to generalise Geiringer's theorem to obtain the expected fixed-point proportion of a generic linear fixed-length GA schema H for a population undergoing crossover only:

$$\lim_{t \to \infty} \Phi(H, t) = \prod_{i \in \Delta(H)} \Phi(*^{i-1} h_i *^{N-i}, 0), \tag{5}$$

where $\Delta(H)$ is the set of indices of the defining symbols in H, h_i is one such defining symbols and we used the power notation x^y to mean x repeated y times. (Note that $\Phi(*^{i-1} h_i *^{N-i}, t)$ coincides with the frequency of allele h_i, $\Phi(h_i, t)$.)

3   Exact Schema Theory and Evolution of Size for Linear Structures

When only unary functions are used in tree-based GP, schemata (and programs) can only take the form $(h_1(h_2(h_3....(h_{N-1} h_N)....)))$ where $h_i \in \mathcal{F} \cup \{=\}$ for $1 \leq i < N$, and $h_N \in \mathcal{T} \cup \{=\}$. Therefore, they can be written unambiguously as strings of symbols of the form $H = h_1 h_2 h_3 .... h_{N-1} h_N$. It should be noted that these strings of symbols do not have to be necessarily interpreted as programs. If one uses a special terminal set $\mathcal{T}$ including only one terminal, say EOR (for End Of Representation), which will be ignored when the representation is interpreted, then strings of the form $h_1 h_2 h_3 .... h_{N-1} h_N$ can be interpreted as chromosomes of length $N - 1$ (since h_N can only be EOR). So, if $\mathcal{F} = \{0, 1\}$, where 0 and 1 are "unary functions",[1] our GP system will explore the space of variable length binary strings. If instead $\mathcal{F}$ includes the "unary functions" $\{$ADD R0 R1, MUL R0 R1, $...\}$, then our tree-based GP system explores the same search space as a machine-code GP system with the same primitive set (Nordin & Banzhaf, 1995; Nordin, 1997). So, in general our specialisation of Equation 3 will be valid for variable-length GAs and linear GP.[2]

In the specialisation to the linear case we replace the "don't care" symbol "=" with the more standard symbol "*". Also, as we did previously, we represent repeated symbols in a string using the power notation. Since in this case all trees are linear, the space of program shapes can be enumerated by $\{G_n\}$ where G_n is $*^n$ for $n > 0$ and the common region between shapes G_j and G_k is simply the shorter of the two schemata, i.e. $C(G_j, G_k) = G_{j \downarrow k} = *^{j \downarrow k}$ where the operator $\downarrow$ returns the minimum of its two arguments. Therefore, the set of crossover masks in the common region, $\chi_{C(G_j, G_k)} = \chi_{*^{j \downarrow k}}$, can be identified with the set $\{0, 1\}^{j \downarrow k}$. Then using the notation l_i to indicate the i-th element of bitmask l and the operator

$$a \bullet b = \begin{cases} a & \text{if } b = 1 \\ * & \text{otherwise,} \end{cases}$$

it is easy to prove (Poli *et al.*, 2002*b*) the following:

Theorem 1 *If* $\chi'_{*^{N \downarrow k}} = \{0, 1\}^{j \downarrow k - 1} \times \{1\}$, *then*

$$\alpha_{\mathrm{xo}}(h_1 ... h_N, t) = \sum_{k > 0} \sum_{l \in \chi'_{*^{N \downarrow k}}} (p_l + p_{\bar{l}}) \, p((h_1 \bullet l_1) \cdots (h_{N \downarrow k - 1} \bullet l_{N \downarrow k - 1}) h_{N \downarrow k} \cdots h_N, t) p((h_1 \bullet \bar{l}_1) \cdots (h_{N \downarrow k - 1} \bullet \bar{l}_{N \downarrow k - 1}) *^{k - N \downarrow k + 1}, t). \tag{6}$$

Equation 6 can be used to study, among other things, the evolution of size in linear GP/GA systems (Poli *et al.*, 2002*b*). This is because it can be specialised to describe the transmission probability of schemata of the form $*^N$ obtaining:

Theorem 2
$$\alpha_{\mathrm{xo}}(*^N, t) = p(*^N, t). \tag{7}$$

This result indicates that in linear representations length evolves under homologous crossovers as if selection only was acting. So, homologous crossovers are totally unbiased with respect to program length. The lack of length bias of homologous crossovers is made particularly clear by the following:

Corollary 3 *For a flat landscape, an infinite population and any* $t > 0$, $\Phi(*^N, t) = \Phi(*^N, 0)$.

This shows that when a homologous crossover alone is acting, any initial distribution of lengths is a fixed point length distribution for the system.

---

[1] The emphasis here is to indicate that we are uninterested in the output of these functions. We are simply interested in their topological organisation within the individual.

[2] An alternative way of transforming a tree-based GP system into a variable length GA or a linear GP system is to use whatever function set $\mathcal{F}$ is appropriate for the task and then a terminal set $\mathcal{T}$ which includes "arity-zero" versions of the functions in $\mathcal{F}$.

4 Extension of Geiringer's Theorem to Variable-size Linear Representations

The extension of Geiringer's theorem to linear, variable-length structures and homologous GP crossover requires two steps: (a) proving that, in the absence of mutation and of selective pressure and for an infinite population, a distribution $\Phi(h_1 h_2 \cdots h_N, t)$, where the alleles/primitives can be considered independent stochastic variables, is a fixed point, and (b) showing that the system indeed moves towards that fixed point. We will do this in the next two subsections.

4.1 Geiringer Manifold

The proof for the following result is available in (Poli *et al.*, 2002*b*).

Theorem 4 *A fixed point distribution for the proportion of a linear, variable-length schema $h_1 h_2 \cdots h_N$ under homologous crossover for an infinite population on a flat fitness landscape in the absence of mutation is*

$$\lim_{t \to \infty} \Phi(h_1 h_2 \cdots h_N, t) = \Phi(*^{N-1} h_N, 0) \prod_{i=1}^{N-1} \frac{\Phi(*^{i-1} h_i \#, 0)}{\Phi(*^i \#, 0)}, \tag{8}$$

where $\Phi(*^{i-1} h_i \#, 0) = \sum_{n > 0} \Phi(*^{i-1} h_i *^n, 0)$ and $\Phi(*^i \#, 0) = \sum_{n > 0} \Phi(*^{i+n}, 0)$.

So, depending on the initial length distribution and the initial allele distribution different fixed points exist for the system. These, however, are all characterised by the fact that they represent statistically independent allele distributions. We will call the class of these fixed points the *Geiringer manifold*.

4.2 Convergence Towards the Geiringer Manifold

The proof of convergence we propose is effectively a generalisation of Geiringer's original proof. However, while in the fixed length case this proof is rather compact and simple to follow, in the variable length case things are significantly more complicated. Both proofs are proofs by induction on the order of the building blocks for a schema or a string. Firstly, we prove that the order-1 building blocks have a unique fixed point and their frequencies converge to that fixed point.[3] Then, we prove that if building blocks of order up to o converge to a unique fixed point, then so must do building blocks of order $o + 1$. The proof is completed by noting that in the previous section we have provided a fixed point, which therefore must be the one to which the system is converging.

4.2.1 Base Step

We now need to introduce a few definitions which are necessary to denote the building blocks of a generic schema, the probabilities that these will be affected by particular crossover masks, and to characterise the freedom of movement of alleles.

Definition 5 *Given a generic set loci $\{i, j, \ldots v\}$ in structures of length N occupied by alleles $h_i, h_j, \ldots, h_v$, respectively, where $1 \le i < j < \ldots < v \le N$, we define the following schemata*

$$
\begin{aligned}
H_i^N &= *^{i-1} h_i *^{N-i} \\
H_{i,j}^N &= H_i^N \cap H_j^N = *^{i-1} h_i *^{j-i-1} h_j *^{N-j} \\
&\vdots \\
H_{i,j,\ldots,v}^N &= H_i^N \cap H_j^N \cap \ldots \cap H_v^N = *^{i-1} h_i *^{j-i-1} h_j * \ldots * h_v *^{N-v}
\end{aligned}
$$

Definition 6 *Let $b_i, b_j, \ldots, b_v$ be arbitrary binary digits and let $l_i, l_j, \ldots, l_v$ be the elements of crossover mask l at positions $i, j, \ldots, v$, respectively. We define*

$$
\begin{aligned}
\mathcal{P}_c(l_i = b_i) &= \sum_{\substack{l \in \chi'_{*c} \\ l_i = b_i}} (p_l + p_{\bar{l}}) \\
\mathcal{P}_c(l_i = b_i, l_j = b_j) &= \sum_{\substack{l \in \chi'_{*c} \\ l_i = b_i, l_j = b_j}} (p_l + p_{\bar{l}}) \\
&\vdots \\
\mathcal{P}_c(l_i = b_i, l_j = b_j, \ldots, l_v = b_v) &= \sum_{\substack{l \in \chi'_{*c} \\ l_i = b_i, l_j = b_j, \ldots, l_v = b_v}} (p_l + p_{\bar{l}})
\end{aligned}
$$

---

[3] Note that in the fixed length case on a flat landscape there is no evolution for the order-1 building blocks (in an infinite population), while in the variable length case alleles can be tranferred between length classes, and so, order-1 building blocks have a more complex dynamics.

Definition 7 *A GP homologous crossover operator acting on variable length linear structures is* allele transferring *if* $0 < \mathcal{P}_c(l_i = b_i) < 1$ *for all* $c > 0$ *and for all* $i \in \{1, \ldots, c-1\}$.

Note that $\mathcal{P}_c(l_i = 1) = 1 - \mathcal{P}_c(l_i = 0)$ and so if $0 < \mathcal{P}_c(l_i = 1) < 1$ then also $0 < \mathcal{P}_c(l_i = 0) < 1$. In an allele transferring homologous crossover, there is always a chance that an allele in a particular locus of a string of a particular length be transferred to the same locus of a string of a different length. This is because, for any given c and i there must be a crossover mask $\hat{l} \in \chi'_{*c}$, having non-zero probability $p_{\hat{l}}$ of being chosen, such that $\hat{l}_i = 0$ and $\hat{l}_c = 1$ and/or a crossover mask $\tilde{l} \in \chi'_{*c}$, having non-zero probability $p_{\tilde{l}}$ of being chosen, such that $\tilde{l}_i = 1$ and $\tilde{l}_c = 0$. Note that we cannot require that $\mathcal{P}_c(l_c = 1) < 1$, since by definition $\chi'_{*c} = \{0,1\}^{c-1} \times \{1\}$, and so $\mathcal{P}_c(l_c = 1) = 1$. That is, the terminals can never be transferred to strings of different length.

Lemma 8 *Let us consider an infinite population of linear, variable-length structures with an initial length distribution such that* $\Phi(*^n, 0) > 0$ *for* $n = 1, \ldots N_m$ *and* $\Phi(*^n, 0) = 0$ *for* $n > N_m$. *If the population undergoes repeated crossover with 100% probability with an allele-transferring homologous crossover on a flat fitness landscape,* $\Phi(H_i^N, t)$ *asymptotically approaches the fixed point provided in Equation 8, i.e.*

$$\lim_{t \to \infty} \Phi(H_i^N, t) = \Phi(*^N, 0) \frac{\Phi(*^{i-1} h_i \#, 0)}{\Phi(*^i \#, 0)},$$

for any $i < N$, *while* $\Phi(H_N^N, t) = \Phi(H_N^N, 0)$ *for any* $t > 0$.

Proof: We start by specialising Equations 6 and 2:

$$
\begin{aligned}
\Phi(H_i^N, t+1) &= \sum_k \delta(i < N \downarrow k) \sum_{l \in \chi'_{*N \downarrow k}} (p_l + p_{\bar{l}})\, \Phi(*^{i-1}(h_i \bullet l_i) *^{N-i}, t) \Phi(*^{i-1}(h_i \bullet \bar{l}_i) *^{k-i}, t) \\
&+ \sum_k \delta(i \geq N \downarrow k) \sum_{l \in \chi'_{*N \downarrow k}} (p_l + p_{\bar{l}})\, \Phi(H_i^N, t) \Phi(*^k, t),
\end{aligned}
\tag{9}
$$

where $\delta(x) = 1$ if x is true, $\delta(x) = 0$ otherwise. So, if $i = N$ then,

$$\Phi(H_N^N, t+1) = \sum_k \sum_{l \in \chi'_{*N \downarrow k}} (p_l + p_{\bar{l}})\, \Phi(H_N^N, t) \Phi(*^k, t) = \Phi(H_N^N, t) \sum_k \Phi(*^k, t) \overbrace{\sum_{l \in \chi'_{*N \downarrow k}} (p_l + p_{\bar{l}})}^{=1} = \Phi(H_N^N, t) \underbrace{\sum_k \Phi(*^k, t)}_{=1} = \Phi(H_N^N, t)$$

which proves the second part of the lemma.

We can now go back to Equation 9 and consider the case $i < N$. In general $\delta(i < N \downarrow k) = \delta(i < N)\delta(i < k)$ and $\delta(i \geq N \downarrow k) = \delta(k > N)\delta(i \geq N) + \delta(k \leq N)\delta(i \geq k)$. However, since we know that $i < N$, in Equation 9 we replace $\delta(i < N \downarrow k)$ with $\delta(k > i)$ and $\delta(i \geq N \downarrow k)$ with $\delta(k \leq N)\delta(i \geq k) = \delta(k \leq i < N) = \delta(k \leq i)$ obtaining

$$
\begin{aligned}
\Phi(H_i^N, t+1) &= \sum_{k > i} \sum_{l \in \chi'_{*N \downarrow k}} (p_l + p_{\bar{l}})\, \Phi(*^{i-1}(h_i \bullet l_i) *^{N-i}, t) \Phi(*^{i-1}(h_i \bullet \bar{l}_i) *^{k-i}, t) + \sum_{k \leq i} \sum_{l \in \chi'_{*N \downarrow k}} (p_l + p_{\bar{l}})\, \Phi(H_i^N, t) \Phi(*^k, t) \\
&= \sum_{k > i} \sum_{\substack{l \in \chi'_{*N \downarrow k} \\ l_i = 0}} (p_l + p_{\bar{l}})\, \Phi(*^N, t) \Phi(H_i^k, t) + \sum_{k > i} \sum_{\substack{l \in \chi'_{*N \downarrow k} \\ l_i = 1}} (p_l + p_{\bar{l}})\, \Phi(H_i^N, t) \Phi(*^k, t) \\
&+ \Phi(H_i^N, t) \sum_{k \leq i} \Phi(*^k, t) \underbrace{\sum_{l \in \chi'_{*N \downarrow k}} (p_l + p_{\bar{l}})}_{=1} \\
&= \sum_{k > i} \Phi(H_i^k, t)\left(\mathcal{P}_{N \downarrow k}(l_i = 0)\Phi(*^N, t)\right) + \Phi(H_i^N, t)\left(\sum_{k > i}\mathcal{P}_{N \downarrow k}(l_i = 1)\Phi(*^k, t) + \sum_{k \leq i}\Phi(*^k, t)\right) \\
&= \sum_{k > i, k \neq N} \Phi(H_i^k, t)\left(\mathcal{P}_{N \downarrow k}(l_i = 0)\Phi(*^N, t)\right) + \Phi(H_i^N, t)\left(\sum_{k > i, k \neq N}\mathcal{P}_{N \downarrow k}(l_i = 1)\Phi(*^k, t) + \sum_{k \leq i}\Phi(*^k, t) + \Phi(*^N, t)\right).
\end{aligned}
$$

An equation of this form can be written for all order-one building blocks H_i^N involving locus i and allele h_i, i.e. for $N = i+1, i+2, \ldots, N_m$, where N_m is the length of the longest program in the initial population (homologous crossovers can never generate programs longer than that, since if $\Phi(*^N, 0) = 0$ for a particular N, then $\Phi(*^N, t) = 0$ for any $t > 0$). This leads to a system of $N_m - i$ linear homogeneous difference equations in the $N_m - i$ variables $\Phi(H_i^{i+1}, t), \ldots, \Phi(H_i^{N_m}, t)$. Since $\Phi(*^k, t) = \Phi(*^k, 0)$ for any $t > 0$ (see Corollary 3), the system has the form

$$x(t+1) = Ax(t), \tag{10}$$

where $x(t) = \left[\Phi(H_i^{i+1}, t), \Phi(H_i^{i+2}, t), \ldots, \Phi(H_i^{N_m}, t) \right]^T$ and A is a matrix $A = (a_{uv})$ the diagonal elements of which are given by

$$a_{uu} = \sum_{k>i, k \neq i+u} \mathcal{P}_{i+u \downarrow k}(l_i = 1) \Phi(*^k, 0) + \sum_{k \leq i} \Phi(*^k, 0) + \Phi(*^{i+u}, 0).$$

The off-diagonal elements of A are

$$a_{uv} = \mathcal{P}_{i+u \downarrow i+v}(l_i = 0) \Phi(*^{i+u}, 0).$$

Let us highlight some properties of the matrix A. Firstly, because $\Phi(*^n, 0) > 0$ for $0 < n \leq N_m$ and the crossover operator is allele-preserving by hypothesis, then A is a positive matrix, i.e. $a_{uv} > 0$ for all u, v. Let us now prove that A is also a stochastic matrix, i.e. $\sum_u a_{uv} = 1$:

$$\sum_u a_{uv} = a_{vv} + \sum_{u \neq v} a_{uv}$$

$$= \sum_{k>i, k \neq i+v} \mathcal{P}_{i+v \downarrow k}(l_i = 1) \Phi(*^k, 0) + \sum_{k \leq i} \Phi(*^k, 0) + \Phi(*^{i+v}, 0) + \sum_{u>0, u \neq v} \mathcal{P}_{i+u \downarrow i+v}(l_i = 0) \Phi(*^{i+u}, 0)$$

$$= \sum_{k>i, k \neq i+v} \mathcal{P}_{k \downarrow i+v}(l_i = 1) \Phi(*^k, 0) + \sum_{k \leq i} \Phi(*^k, 0) + \Phi(*^{i+v}, 0) + \sum_{k>i, k \neq i+v} \mathcal{P}_{k \downarrow i+v}(l_i = 0) \Phi(*^k, 0)$$

where we changed the summation variable from u to $k = u + i$ in the last summation and we used the symmetry relation $\mathcal{P}_{a \downarrow b}(\ldots) = \mathcal{P}_{b \downarrow a}(\ldots)$. So,

$$\sum_u a_{uv} = \sum_{k>i, k \neq i+v} \left(\mathcal{P}_{k \downarrow i+v}(l_i = 1) + \mathcal{P}_{k \downarrow i+v}(l_i = 0) \right) \Phi(*^k, 0) + \sum_{k \leq i} \Phi(*^k, 0) + \Phi(*^{i+v}, 0)$$

$$= \sum_{k>i, k \neq i+v} \Phi(*^k, 0) + \sum_{k \leq i} \Phi(*^k, 0) + \Phi(*^{i+v}, 0)$$

$$= \sum_k \Phi(*^k, 0)$$

$$= 1$$

Therefore, A is a primitive stochastic matrix and the Perron-Frobenius theorem (see for example (Davis & Principe, 1993)) guarantees that A has an eigenvalue $r = 1$ and that all other eigenvalues λ have magnitude smaller than 1. So, if $\mathbf{a}$ is the eigenvector of A associated to the eigenvalue $r = 1$

$$\lim_{t \to \infty} x(t) = \lim_{t \to \infty} A x(t-1) = \lim_{t \to \infty} A^t x(0) = \left(\lim_{t \to \infty} A^t \right) x(0) = A^\infty x(0) = \mathbf{a} \, \mathbf{1}^T x(0)$$

where $\mathbf{1}$ is a vector the components of which are all 1. As a result,

$$\lim_{t \to \infty} x(t) = \mathbf{a} \sum_{n>i} \Phi(H_i^n, 0) = \mathbf{a} \, \Phi(*^{i-1} h_i \#, 0). \tag{11}$$

So, a unique fixed point for $x(t)$ exists. Using Theorem 4 it is easy to show that a fixed point for $x(t)$ is

$$\mathbf{x} = \frac{\Phi(*^{i-1} h_i \#, 0)}{\Phi(*^i \#, 0)} \left[\Phi(*^{i+1}, 0), \Phi(*^{i+2}, 0), \ldots, \Phi(*^{N_m}, 0) \right]^T,$$

which therefore must be a unique global attractor for the system.[4] $\qquad \square$

Lemma 9 *Given a system of linear difference equations of the form*

$$x(t+1) = A x(t) + b(t) \tag{12}$$

with $A = (a_{ij})$ a square matrix such that $a_{ij} \geq 0$ and $\sum_i a_{ij} < 1$ and $b(t)$ a vector such that $\lim_{t \to \infty} b(t)$ exists, then the system has a unique, global attractor.

Proof: A theorem in (Bellman, 1960, p. 288) states that the system of linear equations

$$x = A x + b, \tag{13}$$

where $a_{ij} \geq 0$ for all i and j and $\sum_j a_{ij} < 1$ for all j, has a unique solution. If we take the limit for $t \to \infty$ of both sides of Equation 12, we obtain an equation exactly like Equation 13 (note, we know that the limit for $b(t)$ exists). So, the theorem in (Bellman, 1960) tells us that a fixed point for Equation 12 exists and is unique. $\qquad \square$

---

[4]Incidentally, this allows us to see that $\mathbf{a} = \dfrac{1}{\Phi(*^i \#, 0)} \left[\Phi(*^{i+1}, 0), \Phi(*^{i+2}, 0), \ldots, \Phi(*^{N_m}, 0) \right]^T$.

4.2.2 Induction Step and Theorem

Note that the previous lemma does not tell us what the fixed point is. However, if we could guess one, the lemma tells us that that fixed point is the only one and that $x(t)$ approaches it. This is exactly how we will prove that the fixed point in Theorem 4 is a unique global attractor for the system. In order to do that, however, we need to restrict our attention only to "well-behaved" crossover operators:

Definition 10 *A GP homologous crossover operator acting on variable length linear structures is* not fully linked *if* $0 < \mathcal{P}_c(l_i = b, l_j = b, \ldots, l_v = b) < 1$ *for any choice of* $b \in \{0, 1\}$ *and* $c > 0$, $\mathcal{P}_c(l_i = 0, l_j = 0, \ldots, l_v = 0) + \mathcal{P}_c(l_i = 1, l_j = 1, \ldots, l_v = 1) < 1$ *for any choice of* $c > 0$ *and for any choice and number of loci* $i, j, \ldots, v$ *such that* $0 < i < j < \ldots < v < c$.

Note that a crossover operator which is not fully linked is also allele transferring. We are now ready to prove the following

Theorem 11 *For a not fully linked GP homologous crossover operator, the fixed point for the schema evolution equations provided in Theorem 4 is unique global attractor for the system.*

Proof: We will proceed by induction on the number of defining nodes in the schema. Lemma 8 proves the base step, i.e. shows that order 1 schemata converge to the fixed point provided in Theorem 4. As an induction hypothesis we will assume that schemata of order $1, 2, \ldots, o$ converge to the fixed points provided in Theorem 4 and will show that this is also true for schemata of order $o + 1$.

By specialising Equations 6 and 2 for a generic schema $H_{i,j,n,\ldots,w,v}^N$ of order $o + 1$, it is possible to prove (see Appendix A) that they form a linear system of difference equations

$$x(t + 1) = Ax(t) + b(t)$$

where $x(t) = \left[\Phi(H_{i,\ldots,v}^v, t), \Phi(H_{i,\ldots,v}^{v+1}, t), \ldots, \Phi(H_{i,\ldots,v}^{N_m}, t) \right]^T$ and $A = (a_{qr})$ is a matrix and $b(t) = (b_q(t))$ is a time-varying vector both with indices starting from 0. The diagonal elements of A are given by

$$
\begin{aligned}
a_{qq} \quad = \quad & \sum_{k < i} \Phi(*^k, 0) \\
& + \sum_{i \leq k < j} \mathcal{P}_k(l_i = 1) \Phi(*^k, 0) \\
& + \sum_{j \leq k < n} \mathcal{P}_k(l_i = 1, l_j = 1) \Phi(*^k, 0) + \ldots \\
& + \sum_{w \leq k < v} \mathcal{P}_k(l_i = 1, l_j = 1, \ldots, l_w = 1) \Phi(*^k, 0) \\
& + \sum_{v \leq k < v+q} \mathcal{P}_k(l_i = 1, l_j = 1, \ldots, l_w = 1, l_v = 1) \Phi(*^k, 0) \\
& + \left(\mathcal{P}_{v+q}(l_i = 0, l_j = 0, \ldots, l_v = 0) + \mathcal{P}_{v+q}(l_i = 1, l_j = 1, \ldots, l_w = 1, l_v = 1) \right) \Phi(*^{v+q}, 0) \\
& + \sum_{k > v+q} \mathcal{P}_{v+q}(l_i = 1, l_j = 1, \ldots, l_w = 1, l_v = 1) \Phi(*^k, 0).
\end{aligned}
$$

The off-diagonal elements of A are

$$a_{qr} \quad = \quad \mathcal{P}_{v+r}(l_i = 0, l_j = 0, \ldots, l_v = 0) \Phi(*^{v+q}, 0)$$

if $r < q$, and

$$a_{qr} \quad = \quad \mathcal{P}_{v+q}(l_i = 0, l_j = 0, \ldots, l_v = 0) \Phi(*^{v+q}, 0)$$

if $r > q$. The elements of $b(t)$ are given by

$$
\begin{aligned}
b_q(t) \quad = \quad & \sum_{i \leq k < j} \left(\mathcal{P}_k(l_i = 0) \Phi(H_{j,n,\ldots,w,v}^{v+q}, t) \Phi(H_i^k, t) \right) \\
& + \sum_{j \leq k < n} \left(\mathcal{P}_k(l_i = 0, l_j = 0) \Phi(H_{n,\ldots,w,v}^{v+q}, t) \Phi(H_{i,j}^k, t) \right. \\
& \qquad + \mathcal{P}_k(l_i = 1, l_j = 0) \Phi(H_{i,n,\ldots,w,v}^{v+q}, t) \Phi(H_j^k, t) \\
& \qquad \left. + \mathcal{P}_k(l_i = 0, l_j = 1) \Phi(H_{j,n,\ldots,w,v}^{v+q}, t) \Phi(H_i^k, t) \right) \\
& \vdots \\
& + \sum_{w \leq k < v} \left(\mathcal{P}_k(l_i = 0, l_j = 0, \ldots, l_w = 0) \Phi(H_v^{v+q}, t) \Phi(H_{i,j,n\ldots,w}^k, t) \right)
\end{aligned}
$$

$$+\mathcal{P}_k(l_i = 1, l_j = 0, \ldots, l_w = 0)\Phi(H_{i,v}^{v+q}, t)\Phi(H_{j,n\ldots,w}^k, t)$$
$$+\ldots$$
$$+\mathcal{P}_k(l_i = 0, l_j = 1, \ldots, l_w = 1)\Phi(H_{j,n,\ldots,w,v}^{v+q}, t)\Phi(H_i^k, 0)\Big)$$
$$+ \sum_{v \leq k < v+q} \Big(\mathcal{P}_k(l_i = 1, l_j = 0, \ldots, l_v = 0)\Phi(H_i^{v+q}, 0)\Phi(H_{j,n\ldots,v}^k, t)$$
$$+\mathcal{P}_k(l_i = 0, l_j = 1, \ldots, l_v = 0)\Phi(H_j^{v+q}, t)\Phi(H_{i,n\ldots,v}^k, t)$$
$$+\ldots$$
$$+\mathcal{P}_k(l_i = 0, l_j = 1, \ldots, l_v = 1)\Phi(H_{j,n,\ldots,w,v}^{v+q}, t)\Phi(H_i^k, 0)\Big)$$
$$+ \sum_{k \geq v+q} \Big(\mathcal{P}_{v+q}(l_i = 1, l_j = 0, \ldots, l_v = 0)\Phi(H_i^{v+q}, 0)\Phi(H_{j,n\ldots,v}^k, t)$$
$$+\mathcal{P}_{v+q}(l_i = 0, l_j = 1, \ldots, l_v = 0)\Phi(H_j^{v+q}, t)\Phi(H_{i,n\ldots,v}^k, t)$$
$$+\ldots$$
$$+\mathcal{P}_{v+q}(l_i = 0, l_j = 1, \ldots, l_v = 1)\Phi(H_{j,n,\ldots,w,v}^{v+q}, t)\Phi(H_i^k, 0)\Big)$$

Let us highlight some properties of the matrix A. Firstly, because $\Phi(*^n, 0) > 0$ for $0 < n \leq N_m$ and the crossover operator is not fully linked by hypothesis, then A is a positive matrix, i.e. $a_{qr} > 0$ for all q and r. In Appendix B we prove that for $\sum_q a_{qr} < 1$ for all r.

Let us now study the time evolution of the source term $b(t)$. Each component of $b(t)$ is the (weighted) sum of products of the form $\Phi(H_S^{v+q}, t)\Phi(H_{\bar{S}}^k, t)$, where S is a proper non-empty subset of $\{i, j, n, \ldots, w, v\}$ and $\bar{S} = \{i, j, n, \ldots, w, v\}\backslash S$. So, since $|\{i, j, n, \ldots, w, v\}| = o+1$, the order of the schemata involved in the calculation of $b(t)$ is at most o. Since, by the induction hypothesis we assumed that, for any schema of order not bigger than o, $\Phi(H_{i,\ldots,w}^k, t)$ converges to a unique fixed point, then so does each of the products $\Phi(H_S^{v+q}, t)\Phi(H_{\bar{S}}^k, t)$. Therefore, $\lim_{t\to\infty} b(t)$ exists.

In conclusion, A and $b(t)$ satisfy the conditions of Lemma 9, and so a fixed point for $x(t)$ exists and is unique. Since, Theorem 4 gives us a fixed point, then that must be the only one and must be globally stable. $\qquad\square$

5 Discussion and Conclusions

Characterisations of the genetic biases of the operators (such as the ones offered in this paper and in (Poli *et al.*, 2002a)) are important because they allow the users of GP/GA systems to evaluate whether their operators provide the desired search behaviour for the system. If this is not the case, then the knowledge of the search biases of other operators allows for an informed choice for an alternative.

Here we have focused our attention on the biases of homologous crossovers with respect to length and allele distribution in a population of variable length linear structures and presented theoretical results describing the asymptotic behaviour for a GP/GA system evolving in a flat fitness landscape. Elsewhere (Poli *et al.*, 2002a) we have provided experimental evidence that firmly corroborates the theory, showing a perfect match (within experimental errors) between the predictions of the theory based on generation 0 data and the observed length and allele frequencies at later generations.

The behaviour we have observed and characterised is simple: a) homologous crossovers are totally unbiased with respect to program length, and b) crossover shuffles the alleles present in different individuals and pushes the string distribution towards locus-wise independence.

A mixing behaviour is present in most crossover operators described in the literature on fixed length GAs. It is well known that this destroys "linkage", i.e. correlations, between different allele positions in the population. In the fixed length case the asymptotic convergence towards independence described by Geiringer's theorem is the result of the decay of correlations due to the mixing effect of crossover. Because the representation and operators considered in this paper are generalisations of the corresponding fixed-length ones, it is not so surprising to see that linear GP is also moving towards an independent fixed-point string distribution. In other words, allele mixing is the reason why the right hand side of Equation 8 is a product, like the right hand side of Equation 4. We have no reason to believe that the situation would be significantly different in tree-based GP. (Note that in fixed-length GAs the decay of correlations is exponential in time and in the case of a continuous time evolution can be solved for exactly (Stephens, 2001) showing that higher order correlations decay faster (exponentially) than lower order ones. It is likely that a similar behaviour also characterises GP homologous crossovers.)

Our theoretical results were obtained for the extreme case of infinite populations and flat fitness landscapes. So, why should these be of any relevance to finite GP/GA populations and realistic landscapes? Firstly, because the biases of homologous crossovers in the absence of selection indicate the precise way in which this type of operators would *naturally* tend to explore the search space. When selection is added, the search bias will be modified by the focusing bias of selection, but, except in cases of very strong selection, many of the features of the search bias shown on a flat landscape will be retained. Secondly, because as shown in our experiments, the results obtained with real (but large) populations match very closely the infinite population theory. For smaller populations, the theory can still be used to give short term indications of the behaviour of the system.

References

Altenberg, L. (1995) The Schema Theorem and Price's Theorem. In *Foundations of Genetic Algorithms 3*, (Whitley, L. D. & Vose, M. D., eds), pp. 23–49 Morgan Kaufmann, Estes Park, Colorado, USA.

Bellman, R. (1960) *Introduction to Matrix Analysis*. McGraw-Hill, New York.

Booker, L. B. (1992) Recombination distributions for genetic algorithms. In *FOGA-92, Foundations of Genetic Algorithms*, Vail, Colorado.

Booker, L. B., Fogel, D. B., Whitley, D., Angeline, P. J. & Eiben, A. E. (2000) Recombination. In *Evolutionary Computation 1: Basic Algorithms and Operators*, (Bäck, T., Fogel, D. B. & Michalewicz, T., eds),. Institute of Physics Publishing.

Davis, T. E. & Principe, J. C. (1993) A Markov chain framework for the simple genetic algorithm. *Evolutionary Computation,* **1** (3), 269–288.

Geiringer, H. (1944) On the probability theory of linkage in Mendelian heredity. *Annals of Mathematical Statistics,* **15** (1), 25–57.

Holland, J. (1975) *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, USA.

Langdon, W. B. & Poli, R. (2002) *Foundations of Genetic Programming*. Springer.

Langdon, W. B., Soule, T., Poli, R. & Foster, J. A. (1999) The evolution of size and shape. In *Advances in Genetic Programming 3*, (Spector, L., Langdon, W. B., O'Reilly, U.-M. & Angeline, P. J., eds),. MIT Press Cambridge, MA, USA pp. 163–190.

McPhee, N. F. & Miller, J. D. (1995) Accurate replication in genetic programming. In *Genetic Algorithms: Proceedings of the Sixth International Conference (ICGA95)*, (Eshelman, L., ed.), pp. 303–309 Morgan Kaufmann, Pittsburgh, PA, USA.

McPhee, N. F. & Poli, R. (2001) A schema theory analysis of the evolution of size in genetic programming with linear representations. In *Genetic Programming, Proceedings of EuroGP 2001* LNCS Springer-Verlag, Milan.

McPhee, N. F., Poli, R. & Rowe, J. E. (2001*a*) A schema theory analysis of mutation size biases in genetic programming with linear representations. In *Proceedings of the 2001 Congress on Evolutionary Computation CEC 2001*, Seoul, Korea.

McPhee, N. F., Poli, R. & Rowe, J. E. (2001*b*) A schema theory analysis of mutation size biases in genetic programming with linear representations. In *Proceedings of the 2001 Congress on Evolutionary Computation CEC2001* pp. 1078–1085 IEEE Press, COEX, World Trade Center, 159 Samseong-dong, Gangnam-gu, Seoul, Korea.

Nordin, P. (1997). *Evolutionary Program Induction of Binary Machine Code and its Applications*. PhD thesis, der Universitat Dortmund am Fachereich Informatik.

Nordin, P. & Banzhaf, W. (1995) Complexity compression and evolution. In *Genetic Algorithms: Proceedings of the Sixth International Conference (ICGA95)*, (Eshelman, L., ed.), pp. 310–317 Morgan Kaufmann, Pittsburgh, PA, USA.

Poli, R. (2000*a*) Exact schema theorem and effective fitness for GP with one-point crossover. In *Proceedings of the Genetic and Evolutionary Computation Conference*, (Whitley, D., Goldberg, D., Cantu-Paz, E., Spector, L., Parmee, I. & Beyer, H.-G., eds), pp. 469–476 Morgan Kaufmann, Las Vegas.

Poli, R. (2000*b*) Hyperschema theory for GP with one-point crossover, building blocks, and some new results in GA theory. In *Genetic Programming, Proceedings of EuroGP 2000*, (Poli, R. & *et al.*, W. B., eds), Springer-Verlag.

Poli, R. (2001*a*) General schema theory for genetic programming with subtree-swapping crossover. In *Genetic Programming, Proceedings of EuroGP 2001* LNCS Springer-Verlag, Milan.

Poli, R. (2001*b*) Exact schema theory for genetic programming and variable-length genetic algorithms with one-point crossover. *Genetic Programming and Evolvable Machines,* **2** (2), 123–163.

Poli, R. & Langdon, W. B. (1997) A new schema theory for genetic programming with one-point crossover and point mutation. In *Genetic Programming 1997: Proceedings of the Second Annual Conference*, (Koza, J. R., Deb, K., Dorigo, M., Fogel, D. B., Garzon, M., Iba, H. & Riolo, R. L., eds), pp. 278–285 Morgan Kaufmann, Stanford University, CA, USA.

Poli, R. & Langdon, W. B. (1998*a*) On the search properties of different crossover operators in genetic programming. In *Genetic Programming 1998: Proceedings of the Third Annual Conference*, (Koza, J. R., Banzhaf, W., Chellapilla, K., Deb, K., Dorigo, M., Fogel, D. B., Garzon, M. H., Goldberg, D. E., Iba, H. & Riolo, R., eds), pp. 293–301 Morgan Kaufmann, University of Wisconsin, Madison, Wisconsin, USA.

Poli, R. & Langdon, W. B. (1998*b*) Schema theory for genetic programming with one-point crossover and point mutation. *Evolutionary Computation,* **6** (3), 231–252.

Poli, R. & McPhee, N. F. (2001*a*) Exact GP schema theory for headless chicken crossover and subtree mutation. In *Proceedings of the 2001 Congress on Evolutionary Computation CEC 2001*, Seoul, Korea.

Poli, R. & McPhee, N. F. (2001*b*) Exact schema theorems for GP with one-point and standard crossover operating on linear structures and their application to the study of the evolution of size. In *Genetic Programming, Proceedings of EuroGP 2001* LNCS Springer-Verlag, Milan.

Poli, R. & McPhee, N. F. (2001*c*) Exact schema theory for GP and variable-length GAs with homologous crossover. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2001)* Morgan Kaufmann, San Francisco, California, USA.

Poli, R., Rowe, J. E. & McPhee, N. F. (2001) Markov chain models for GP and variable-length GAs with homologous crossover. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2001)* Morgan Kaufmann, San Francisco, California, USA.

Poli, R., Rowe, J. E., Stephens, C. R. & Wright, A. H. (2002*a*) Allele diffusion in linear genetic programming and variable-length genetic algorithms with subtree crossover. In *Proceedings of EuroGP 2002*.

Poli, R., Stephens, C. R., Wright, A. H. & Rowe, J. E. (2002*b*) On the search biases of homologous crossover in linear genetic programming and variable-length genetic algorithms. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2002)* Morgan Kaufmann, New York, USA. (accepted as full paper).

Rowe, J. E. & McPhee, N. F. (2001) The effects of crossover and mutation operators on variable length linear structures. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2001)* Morgan Kaufmann, San Francisco, California, USA.

Soule, T., Foster, J. A. & Dickinson, J. (1996) Code growth in genetic programming. In *Genetic Programming 1996: Proceedings of the First Annual Conference*, (Koza, J. R., Goldberg, D. E., Fogel, D. B. & Riolo, R. L., eds), pp. 215–223 MIT Press, Stanford University, CA, USA.

Spears, W. M. (2000) Limiting distributions for mutation and recombination. In *Proceedings of the Foundations of Genetic Algorithms Workshop (FOGA 6)*, (Spears, W. M. & Martin, W., eds),, Charlottesville, VA, USA. In press.

Stephens, C. R. (2001) Some exact results from a coarse grained formulation of genetic dynamics. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2001)*, (Spector, L., Goodman, E. D., Wu, A., Langdon, W. B., Voigt, H.-M., Gen, M., Sen, S., Dorigo, M., Pezeshk, S., Garzon, M. H. & Burke, E., eds), pp. 631–638 Morgan Kaufmann, San Francisco, California, USA.

Stephens, C. R. & Waelbroeck, H. (1997) Effective degrees of freedom in genetic algorithms and the block hypothesis. In *Proceedings of the Seventh International Conference on Genetic Algorithms (ICGA97)*, (Bäck, T., ed.), pp. 34–40 Morgan Kaufmann, East Lansing.

Stephens, C. R. & Waelbroeck, H. (1999) Schemata evolution and building blocks. *Evolutionary Computation,* **7** (2), 109–124.

Whitley, L. D. (1994) A Genetic Algorithm Tutorial. *Statistics and Computing,* **4**, 65–85.

A Linearity

By specialising Equations 6 and 2 for a generic schema $H^N_{i,j,n,\ldots,w,v}$ of order $o+1$ we obtain:

$$
\begin{aligned}
\Phi(H^N_{i,j,n,\ldots,w,v},t+1) &= \sum_k \delta(N\downarrow k<i) \sum_{l\in\chi'_{*N\downarrow k}} (p_l+p_{\bar l})\,\Phi(H^N_{i,j,n,\ldots,w,v},t)\Phi(*^k,t)\\[4pt]
&+ \sum_k \delta(i\le N\downarrow k<j)\sum_{l\in\chi'_{*N\downarrow k}} \delta(l_i=0)\,(p_l+p_{\bar l})\,\Phi(H^N_{j,n,\ldots,w,v},t)\Phi(H^k_i,t)\\[4pt]
&+ \sum_k \delta(i\le N\downarrow k<j)\sum_{l\in\chi'_{*N\downarrow k}} \delta(l_i=1)\,(p_l+p_{\bar l})\,\Phi(H^N_{i,j,n,\ldots,w,v},t)\Phi(*^k,t)\\[4pt]
&+ \sum_k \delta(j\le N\downarrow k<n)\sum_{l\in\chi'_{*N\downarrow k}} \delta(l_i=0)\delta(l_j=0)\,(p_l+p_{\bar l})\,\Phi(H^N_{n,\ldots,w,v},t)\Phi(H^k_{i,j},t)\\[4pt]
&+ \sum_k \delta(j\le N\downarrow k<n)\sum_{l\in\chi'_{*N\downarrow k}} \delta(l_i=1)\delta(l_j=0)\,(p_l+p_{\bar l})\,\Phi(H^N_{i,n,\ldots,w,v},t)\Phi(H^k_j,t)\\[4pt]
&+ \sum_k \delta(j\le N\downarrow k<n)\sum_{l\in\chi'_{*N\downarrow k}} \delta(l_i=0)\delta(l_j=1)\,(p_l+p_{\bar l})\,\Phi(H^N_{j,n,\ldots,w,v},t)\Phi(H^k_i,t)\\[4pt]
&+ \sum_k \delta(j\le N\downarrow k<n)\sum_{l\in\chi'_{*N\downarrow k}} \delta(l_i=1)\delta(l_j=1)\,(p_l+p_{\bar l})\,\Phi(H^N_{i,j,n,\ldots,w,v},t)\Phi(*^k,t)\\[4pt]
&\ \ \vdots\\[4pt]
&+ \sum_k \delta(w\le N\downarrow k<v)\sum_{l\in\chi'_{*N\downarrow k}} \delta(l_i=0)\delta(l_j=0)\ldots\delta(l_w=0)\,(p_l+p_{\bar l})\,\Phi(H^N_v,t)\Phi(H^k_{i,j,n\ldots,w},t)\\[4pt]
&+ \sum_k \delta(w\le N\downarrow k<v)\sum_{l\in\chi'_{*N\downarrow k}} \delta(l_i=1)\delta(l_j=0)\ldots\delta(l_w=0)\,(p_l+p_{\bar l})\,\Phi(H^N_{i,v},t)\Phi(H^k_{j,n\ldots,w},t)\\[4pt]
&+ \ldots\\[4pt]
&+ \sum_k \delta(w\le N\downarrow k<v)\sum_{l\in\chi'_{*N\downarrow k}} \delta(l_i=1)\delta(l_j=1)\ldots\delta(l_w=1)\,(p_l+p_{\bar l})\,\Phi(H^N_{i,j,n,\ldots,w,v},t)\Phi(*^k,t)\\[4pt]
&+ \sum_k \delta(v\le N\downarrow k)\sum_{l\in\chi'_{*N\downarrow k}} \delta(l_i=0)\delta(l_j=0)\ldots\delta(l_v=0)\,(p_l+p_{\bar l})\,\Phi(*^N,t)\Phi(H^k_{i,j,n\ldots,v},t)\\[4pt]
&+ \sum_k \delta(v\le N\downarrow k)\sum_{l\in\chi'_{*N\downarrow k}} \delta(l_i=1)\delta(l_j=0)\ldots\delta(l_v=0)\,(p_l+p_{\bar l})\,\Phi(H^N_i,t)\Phi(H^k_{j,n\ldots,v},t)\\[4pt]
&+ \ldots\\[4pt]
&+ \sum_k \delta(v\le N\downarrow k)\sum_{l\in\chi'_{*N\downarrow k}} \delta(l_i=1)\delta(l_j=1)\ldots\delta(l_v=1)\,(p_l+p_{\bar l})\,\Phi(H^N_{i,j,n,\ldots,w,v},t)\Phi(*^k,t)\\[4pt]
&= \sum_k \delta(k<i)\Phi(H^N_{i,j,n,\ldots,w,v},t)\Phi(*^k,t)\\[4pt]
&+ \sum_k \delta(i\le k<j)\Big(\mathcal{P}_{N\downarrow k}(l_i=0)\Phi(H^N_{j,n,\ldots,w,v},t)\Phi(H^k_i,t)\\[4pt]
&\qquad\qquad +\mathcal{P}_{N\downarrow k}(l_i=1)\Phi(H^N_{i,j,n,\ldots,w,v},t)\Phi(*^k,t)\Big)
\end{aligned}
$$

$$+ \sum_k \delta(j \le k < n)\Big(\mathcal{P}_{N\downarrow k}(l_i = 0, l_j = 0)\Phi(H^N_{n,\ldots,w,v}, t)\Phi(H^k_{i,j}, t)$$

$$+ \mathcal{P}_{N\downarrow k}(l_i = 1, l_j = 0)\Phi(H^N_{i,n,\ldots,w,v}, t)\Phi(H^k_j, t)$$
$$+ \mathcal{P}_{N\downarrow k}(l_i = 0, l_j = 1)\Phi(H^N_{j,n,\ldots,w,v}, t)\Phi(H^k_i, t)$$
$$+ \mathcal{P}_{N\downarrow k}(l_i = 1, l_j = 1)\Phi(H^N_{i,j,n,\ldots,w,v}, t)\Phi(*^k, t)\Big)$$

$$\vdots$$

$$+ \sum_k \delta(w \le k < v)\Big(\mathcal{P}_{N\downarrow k}(l_i = 0, l_j = 0, \ldots, l_w = 0)\Phi(H^N_v, t)\Phi(H^k_{i,j,n\ldots,w}, t)$$

$$+ \mathcal{P}_{N\downarrow k}(l_i = 1, l_j = 0, \ldots, l_w = 0)\Phi(H^N_{i,v}, t)\Phi(H^k_{j,n\ldots,w}, t)$$
$$+ \ldots$$
$$+ \mathcal{P}_{N\downarrow k}(l_i = 1, l_j = 1, \ldots, l_w = 1)\Phi(H^N_{i,j,n,\ldots,w,v}, t)\Phi(*^k, t)\Big)$$

$$+ \sum_k \delta(v \le N \downarrow k)\Big(\mathcal{P}_{N\downarrow k}(l_i = 0, l_j = 0, \ldots, l_v = 0)\Phi(*^N, t)\Phi(H^k_{i,j,n\ldots,v}, t)$$

$$+ \mathcal{P}_{N\downarrow k}(l_i = 1, l_j = 0, \ldots, l_v = 0)\Phi(H^N_i, t)\Phi(H^k_{j,n\ldots,v}, t)$$
$$+ \ldots$$
$$+ \mathcal{P}_{N\downarrow k}(l_i = 1, l_j = 1, \ldots, l_v = 1)\Phi(H^N_{i,j,n,\ldots,w,v}, t)\Phi(*^k, t)\Big).$$

By noting that $\mathcal{P}_{N\downarrow k}(\ldots) = \mathcal{P}_N(\ldots)$ if $k \ge N$ and $\mathcal{P}_{N\downarrow k}(\ldots) = \mathcal{P}_k(\ldots)$ otherwise and that $\Phi(*^k, t) = \Phi(*^k, 0)$ we obtain

$$\Phi(H^N_{i,j,n,\ldots,w,v}, t+1) = \sum_{k<i} \Phi(H^N_{i,j,n,\ldots,w,v}, t)\Phi(*^k, 0)$$

$$+ \sum_{i \le k < j}\Big(\mathcal{P}_k(l_i = 0)\Phi(H^N_{j,n,\ldots,w,v}, t)\Phi(H^k_i, t) + \mathcal{P}_k(l_i = 1)\Phi(H^N_{i,j,n,\ldots,w,v}, t)\Phi(*^k, 0)\Big)$$

$$+ \sum_{j \le k < n}\Big(\mathcal{P}_k(l_i = 0, l_j = 0)\Phi(H^N_{n,\ldots,w,v}, t)\Phi(H^k_{i,j}, t)$$

$$+ \mathcal{P}_k(l_i = 1, l_j = 0)\Phi(H^N_{i,n,\ldots,w,v}, t)\Phi(H^k_j, t)$$
$$+ \mathcal{P}_k(l_i = 0, l_j = 1)\Phi(H^N_{j,n,\ldots,w,v}, t)\Phi(H^k_i, t)$$
$$+ \mathcal{P}_k(l_i = 1, l_j = 1)\Phi(H^N_{i,j,n,\ldots,w,v}, t)\Phi(*^k, 0)\Big)$$

$$\vdots$$

$$+ \sum_{w \le k < v}\Big(\mathcal{P}_k(l_i = 0, l_j = 0, \ldots, l_w = 0)\Phi(H^N_v, t)\Phi(H^k_{i,j,n\ldots,w}, t)$$

$$+ \mathcal{P}_k(l_i = 1, l_j = 0, \ldots, l_w = 0)\Phi(H^N_{i,v}, t)\Phi(H^k_{j,n\ldots,w}, t)$$
$$+ \ldots$$
$$+ \mathcal{P}_k(l_i = 1, l_j = 1, \ldots, l_w = 1)\Phi(H^N_{i,j,n,\ldots,w,v}, t)\Phi(*^k, 0)\Big)$$

$$+ \sum_{v \le k < N}\Big(\mathcal{P}_k(l_i = 0, l_j = 0, \ldots, l_v = 0)\Phi(*^N, 0)\Phi(H^k_{i,j,n\ldots,v}, t)$$

$$+ \mathcal{P}_k(l_i = 1, l_j = 0, \ldots, l_v = 0)\Phi(H^N_i, t)\Phi(H^k_{j,n\ldots,v}, t)$$
$$+ \ldots$$
$$+ \mathcal{P}_k(l_i = 1, l_j = 1, \ldots, l_v = 1)\Phi(H^N_{i,j,n,\ldots,w,v}, t)\Phi(*^k, 0)\Big)$$

$$+ \sum_{k \ge N}\Big(\mathcal{P}_N(l_i = 0, l_j = 0, \ldots, l_v = 0)\Phi(*^N, 0)\Phi(H^k_{i,j,n\ldots,v}, t)$$

$$+ \mathcal{P}_N(l_i = 1, l_j = 0, \ldots, l_v = 0)\Phi(H^N_i, t)\Phi(H^k_{j,n\ldots,v}, t)$$
$$+ \ldots$$
$$+ \mathcal{P}_N(l_i = 1, l_j = 1, \ldots, l_v = 1)\Phi(H^N_{i,j,n,\ldots,w,v}, t)\Phi(*^k, 0)\Big)$$

We now reorder the terms in this equation collecting those multiplying the proportions of the schemata of order $o + 1$, i.e. those of the form $\Phi(H^k_{i,\ldots,v}, t)$, obtaining

$$
\begin{aligned}
\Phi(H^N_{i,j,n,\ldots,w,v}, t+1) \;=\; & \Phi(H^N_{i,j,n,\ldots,w,v}, t)\Bigg(\sum_{k<i} \Phi(*^k, 0) \\
& + \sum_{i\leq k<j} \mathcal{P}_k(l_i = 1)\Phi(*^k, 0) \\
& + \sum_{j\leq k\leq n} \mathcal{P}_k(l_i = 1, l_j = 1)\Phi(*^k, 0) + \ldots \\
& + \sum_{w\leq k\leq v} \mathcal{P}_k(l_i = 1, l_j = 1, \ldots, l_w = 1)\Phi(*^k, 0) \\
& + \sum_{v\leq k<N} \mathcal{P}_k(l_i = 1, l_j = 1, \ldots, l_w = 1, l_v = 1)\Phi(*^k, 0) \\
& + \Big(\mathcal{P}_N(l_i = 0, l_j = 0, \ldots, l_v = 0) + \mathcal{P}_N(l_i = 1, l_j = 1, \ldots, l_w = 1, l_v = 1)\Big)\Phi(*^N, 0) \\
& + \sum_{k>N} \mathcal{P}_N(l_i = 1, l_j = 1, \ldots, l_w = 1, l_v = 1)\Phi(*^k, 0)\Bigg) \\
+ & \sum_{v\leq k<N} \Phi(H^k_{i,j,n\ldots,v}, t)\Big(\mathcal{P}_k(l_i = 0, l_j = 0, \ldots, l_v = 0)\Phi(*^N, 0)\Big) \\
+ & \sum_{k>N} \Phi(H^k_{i,j,n\ldots,v}, t)\Big(\mathcal{P}_N(l_i = 0, l_j = 0, \ldots, l_v = 0)\Phi(*^N, 0)\Big) \\
+ & \Bigg(\sum_{i\leq k<j}\Big(\mathcal{P}_k(l_i = 0)\Phi(H^N_{j,n,\ldots,w,v}, t)\Phi(H^k_i, t)\Big) \\
& + \sum_{j\leq k<n}\Big(\mathcal{P}_k(l_i = 0, l_j = 0)\Phi(H^N_{n,\ldots,w,v}, t)\Phi(H^k_{i,j}, t) \\
& \quad + \mathcal{P}_k(l_i = 1, l_j = 0)\Phi(H^N_{i,n,\ldots,w,v}, t)\Phi(H^k_j, t) \\
& \quad + \mathcal{P}_k(l_i = 0, l_j = 1)\Phi(H^N_{j,n,\ldots,w,v}, t)\Phi(H^k_i, t)\Big) \\
& \;\vdots \\
& + \sum_{w\leq k<v}\Big(\mathcal{P}_k(l_i = 0, l_j = 0, \ldots, l_w = 0)\Phi(H^N_v, t)\Phi(H^k_{i,j,n\ldots,w}, t) \\
& \quad + \mathcal{P}_k(l_i = 1, l_j = 0, \ldots, l_w = 0)\Phi(H^N_{i,v}, t)\Phi(H^k_{j,n\ldots,w}, t) \\
& \quad + \ldots \\
& \quad + \mathcal{P}_k(l_i = 0, l_j = 1, \ldots, l_w = 1)\Phi(H^N_{j,n,\ldots,w,v}, t)\Phi(H^k_i, 0)\Big) \\
& + \sum_{v\leq k<N}\Big(\mathcal{P}_k(l_i = 1, l_j = 0, \ldots, l_v = 0)\Phi(H^N_i, 0)\Phi(H^k_{j,n\ldots,v}, t) \\
& \quad + \mathcal{P}_k(l_i = 0, l_j = 1, \ldots, l_v = 0)\Phi(H^N_j, t)\Phi(H^k_{i,n\ldots,v}, t) \\
& \quad + \ldots \\
& \quad + \mathcal{P}_k(l_i = 0, l_j = 1, \ldots, l_v = 1)\Phi(H^N_{j,n,\ldots,w,v}, t)\Phi(H^k_i, 0)\Big) \\
& + \sum_{k\geq N}\Big(\mathcal{P}_N(l_i = 1, l_j = 0, \ldots, l_v = 0)\Phi(H^N_i, 0)\Phi(H^k_{j,n\ldots,v}, t) \\
& \quad + \mathcal{P}_N(l_i = 0, l_j = 1, \ldots, l_v = 0)\Phi(H^N_j, t)\Phi(H^k_{i,n\ldots,v}, t) \\
& \quad + \ldots \\
& \quad + \mathcal{P}_N(l_i = 0, l_j = 1, \ldots, l_v = 1)\Phi(H^N_{j,n,\ldots,w,v}, t)\Phi(H^k_i, 0)\Big)\Bigg)
\end{aligned}
$$

B Limited Sum

We want to prove that $\sum_q a_{qr} < 1$ for all r.

$$
\begin{aligned}
\sum_q a_{qr} \;=\;& \sum_{q<r} a_{qr} + a_{rr} + \sum_{q>r} a_{qr} \\[4pt]
=\;& \sum_{q<r} \mathcal{P}_{v+q}(l_i = 0, l_j = 0, \ldots, l_v = 0)\Phi(*^{v+q}, 0) \\[4pt]
& + \Bigg(\sum_{k<i} \Phi(*^k, 0) \\
& \qquad + \sum_{i\le k<j} \mathcal{P}_k(l_i = 1)\Phi(*^k, 0) \\
& \qquad + \sum_{j\le k<n} \mathcal{P}_k(l_i = 1, l_j = 1)\Phi(*^k, 0) + \ldots \\
& \qquad + \sum_{w\le k<v} \mathcal{P}_k(l_i = 1, l_j = 1, \ldots, l_w = 1)\Phi(*^k, 0) \\
& \qquad + \sum_{v\le k<v+r} \mathcal{P}_k(l_i = 1, l_j = 1, \ldots, l_w = 1, l_v = 1)\Phi(*^k, 0) \\
& \qquad + \big(\mathcal{P}_{v+r}(l_i = 0, l_j = 0, \ldots, l_v = 0) + \mathcal{P}_{v+r}(l_i = 1, l_j = 1, \ldots, l_w = 1, l_v = 1) \big)\Phi(*^{v+r}, 0) \\
& \qquad + \sum_{k>v+r} \mathcal{P}_{v+r}(l_i = 1, l_j = 1, \ldots, l_w = 1, l_v = 1)\Phi(*^k, 0) \Bigg) \\[4pt]
& + \sum_{q>r} \mathcal{P}_{v+r}(l_i = 0, l_j = 0, \ldots, l_v = 0)\Phi(*^{v+q}, 0)
\end{aligned}
$$

If we change the summation variable from being q to being $k = q + v$, we obtain

$$
\begin{aligned}
\sum_q a_{qr} \;=\;& \sum_{v\le k<r+v} \mathcal{P}_k(l_i = 0, l_j = 0, \ldots, l_v = 0)\Phi(*^k, 0) \\[4pt]
& + \Bigg(\sum_{k<i} \Phi(*^k, 0) \\
& \qquad + \sum_{i\le k<j} \mathcal{P}_k(l_i = 1)\Phi(*^k, 0) \\
& \qquad + \sum_{j\le k<n} \mathcal{P}_k(l_i = 1, l_j = 1)\Phi(*^k, 0) + \ldots \\
& \qquad + \sum_{w\le k<v} \mathcal{P}_k(l_i = 1, l_j = 1, \ldots, l_w = 1)\Phi(*^k, 0) \\
& \qquad + \sum_{v\le k<v+r} \mathcal{P}_k(l_i = 1, l_j = 1, \ldots, l_w = 1, l_v = 1)\Phi(*^k, 0) \\
& \qquad + \big(\mathcal{P}_{v+r}(l_i = 0, l_j = 0, \ldots, l_v = 0) + \mathcal{P}_{v+r}(l_i = 1, l_j = 1, \ldots, l_w = 1, l_v = 1) \big)\Phi(*^{v+r}, 0) \\
& \qquad + \sum_{k>v+r} \mathcal{P}_{v+r}(l_i = 1, l_j = 1, \ldots, l_w = 1, l_v = 1)\Phi(*^k, 0) \Bigg) \\[4pt]
& + \sum_{k>v+r} \mathcal{P}_{v+r}(l_i = 0, l_j = 0, \ldots, l_v = 0)\Phi(*^k, 0) \\[4pt]
=\;& \sum_{k<i} \Phi(*^k, 0) \\[4pt]
& + \sum_{i\le k<j} \mathcal{P}_k(l_i = 1)\Phi(*^k, 0) \\[4pt]
& + \sum_{j\le k<n} \mathcal{P}_k(l_i = 1, l_j = 1)\Phi(*^k, 0) + \ldots \\[4pt]
& + \sum_{w\le k<v} \mathcal{P}_k(l_i = 1, l_j = 1, \ldots, l_w = 1)\Phi(*^k, 0)
\end{aligned}
$$

$$+ \sum_{v \leq k \leq v+r} \Big(\mathcal{P}_k(l_i = 0, l_j = 0, \ldots, l_v = 0) + \mathcal{P}_k(l_i = 1, l_j = 1, \ldots, l_w = 1, l_v = 1) \Big) \Phi(*^k, 0)$$

$$+ \sum_{k > v+r} \Big(\mathcal{P}_{v+r}(l_i = 0, l_j = 0, \ldots, l_v = 0) + \mathcal{P}_{v+r}(l_i = 1, l_j = 1, \ldots, l_w = 1, l_v = 1) \Big) \Phi(*^k, 0).$$

Because the crossover operator is not fully linked, from this equation we obtain:

$$\sum_q a_{qr} < \sum_k \Phi(*^k, 0) = 1$$

which proves that $\sum_q a_{qr} < 1$ for any value of r.