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

Riccardo Poli

Department of Computer Science
University of Essex, UK
rpoli@essex.ac.uk

Christopher R. Stephens

Instituto de Ciencias Nucleares
UNAM, Mexico
stephens@nuclecu.unam.mx

Alden H. Wright

Computer Science Department
University of Montana, USA
wright@cs.umt.edu

Jonathan E. Rowe

School of Computer Science
The University of Birmingham, UK
j.e.rowe@cs.bham.ac.uk

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 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 clarifies what it means for a population of variable length strings to present a statistically independent distribution of primitives and indicates the presence of a mixing process that invariably pushes the population to one such distribution.

1 INTRODUCTION

Most search algorithms present biases in the way they sample the search space. For example, an algorithm will sample certain areas of the search space sooner than another, or will ignore altogether some areas which the other samples, or maybe it will allocate more samples to certain areas, perhaps resampling some points, while the other algorithm does

not, and so on. Naturally, the search bias of an algorithm is the result of the interaction of the biases of all its components. In the case of GAs these components are selection, crossover and mutation. Knowing the exact direction and intensity of the biases exerted by these components at any particular point in population space (which we could imagine as different force fields) is important because a GA follows cyclically the selection, crossover and mutation biases which, therefore, fully determine the resulting trajectory in population space.

Nowadays we have several kinds of mathematical models which can help us understand the trajectories followed by a GA or a GP system in population space in a variety of conditions and for a variety of operators. If we can assume infinitely large populations (or at least population sizes which are big enough, and numbers of generations which are small enough, that we can ignore sampling noise) then we can treat GAs and GP systems as deterministic systems (in fact dynamical systems) and we have exact models to predict their trajectories in population space (Vose 1999; Stephens and Waelbroeck 1997; Poli 2001).

So, at least in theory, we could integrate numerically the models' equations to draw our force maps and understand the strength, direction and fixed points of selection, crossover and mutation for a variety of systems. Unfortunately, in practice, except in specific cases, GAs and GP systems have a huge number of degrees of freedom and are described by unmanageably large systems of difference equations which defeat any numerical integration approach. Also, even if we could numerically integrate the equations of motion for the system, all we could do is to determine where the GA would go in a finite number of cases and for a finite number of generations.

Fortunately, in some cases the equations of motion for GAs and GP systems can be studied mathematically. Particularly manageable is the case in which infinite populations are assumed and the repeated iteration of only one operator at a time is considered since, in these cases, the evolution equations simplify sufficiently to make their study possible. For example, it may become possible to study mathematically the fixed points of such equations. These are interesting because they describe in a succinct and easily understandable way where each operator is trying to push the population.

Over the years a large body of evidence has been gathered in this way for GAs operating on fixed length strings (often borrowing from population genetics). For example, we know that homologous crossovers (such as uniform crossover or one-point crossover) try to push the GA to populations where the alleles in the initial population are perfectly mixed — a result which was wonderfully obtained by Hilda Geiringer almost 60 years ago (Geiringer 1944). For this reason we call the subset of population space containing all perfectly mixed populations the *Geiringer manifold*. We also know that in general selection tries to push the GA to populations containing only copies of the best string in the population (assuming that only one type of individual has the highest fitness in the initial generation). Importantly, all homogeneous populations belong also to the Geiringer manifold for homologous crossovers (because with these operators when one crosses an individual with itself, one gets that individual back again). Finally, we know that point mutation pushes the population to a particular point in population space where all possible strings are equally represented in the population (Vose 1999) and that this point also belongs to the Geiringer manifold for homologous crossovers. In addition, we have quite clear characterisations of the behaviour of each combination of two operators.

This bias information, although incomplete, may be very valuable. It is important to note, however, that, while for the case of GAs acting on linear fixed-length representations we already know a lot about the biases of the operators from various sources, in the case of GP or GAs acting on linear, but variable-length, strings, we know very little about these biases. (Naturally, even in these cases we know how selection is going to affect the population, since selection is representation independent.) For example, before the work described in this paper and in (Poli, Stephens, Wright, and Rowe 2002) it was entirely unclear what it meant for a population of variable length strings or a population of trees to be perfectly mixed (or in linkage equilibrium) with respect to a given crossover operator. Likewise we still have very little knowledge as to the fixed points for a population of variable size and shape entities repeatedly undergoing variable-size sub-structure mutation.

In recent research we have started using some exact new schema-theoretic models of GP and variable length GAs with the objective of understanding the biases of different crossover and mutation operators in terms of the way they affect the size distribution of the structures undergoing evolution (Poli and McPhee 2001a; McPhee and Poli 2001; McPhee, Poli, and Rowe 2001). More recently we have started focusing on the biases of the operators with respect to the alleles or primitives in the representation finding fixed points for the allele/primitive distribution in the case of linear, but variable-length representations, undergoing subtree crossover (Poli, Rowe, Stephens, and Wright 2002) and homologous crossover (Poli, Stephens, Wright, and Rowe 2002). The first type of operator is important because it is widely used (e.g. in a popular form of linear GP called Grammatical Evolution (O'Neill and Ryan 2001)). The second class of operators is also important because it generalises the notion of crossover used in fixed-length GAs. Unfortunately, until now, for each operator we were only able to provide a family of fixed points (the Geiringer manifold) and to show empirically that populations seemed to invariably converge towards it. In this paper we make further progress, by generalising Geiringer's proof of convergence to homologous crossover acting on linear variable length structures and clarifying for which homologous operators our family of fixed points is a global attractor for the system. In the future, we hope to be able to do the same for subtree crossover.

The paper is organised as follows. We provide some background information on Geiringer's theorem in Section 2. Then we give a summary of the main results reported in (Poli, Stephens, Wright, and Rowe 2002) since they are the starting point for this paper. In Section 3 we describe homologous crossover and report an exact schema theorem for the case of linear, but variable-length, structures. We provide a family of fixed points for the allele distribution in the infinite population limit and give the proof of convergence in Section 4. We give an example in Section 5. Finally, we discuss our results and we draw some conclusions in Section 6.

2 GEIRINGER'S THEOREM FOR FIXED-LENGTH GENETIC ALGORITHMS

Let us start by introducing Geiringer's theorem (Geiringer 1944), which, thanks to its reinterpretation within the context of evolutionary algorithms (Booker 1992; Stephens and Waelbroeck 1997; Booker, Fogel, Whitley, Angeline, and Eiben 2000; Vose and Wright 1998; Spears 2000), provides a description of the bias exerted by crossover in a GA.

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 in locus/position i then

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

If one interprets $\Phi(h_1 h_2 \cdots h_N, t)$ as the probability distribution of the possible strings in the population, we can interpret Equation 1 as saying that such a distribution is converging towards independence. This result is valid for all homologous crossover operators that 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 fixed-point proportion of a generic linear fixed-length GA schema H for a population undergoing crossover only:

$$\lim_{t \rightarrow \infty} \Phi(H, t) = \prod_{i \in \Delta(H)} \Phi(*^{i-1} h_i *^{N-i}, 0), \quad (2)$$

where $*$ is a schema's don't care symbol which stands for any primitive/allele, $\Delta(H)$ is the set of indices of the defining symbols in H , h_i is one such defining symbols and we use the power notation x^y to mean x repeated y times. (Note that $\Phi(*^{i-1} h_i *^{N-i}, t) = \Phi(h_i, t)$.)

As an example, let us consider a population of binary strings of length 2, having the following initial proportions:

$$\Phi(00, 0) = \frac{1}{3}, \quad \Phi(01, 0) = \frac{1}{3}, \quad \Phi(10, 0) = 0, \quad \Phi(11, 0) = \frac{1}{3}.$$

To these correspond the following allele frequencies:

$$\Phi(*0, 0) = \frac{1}{3}, \quad \Phi(*1, 0) = \frac{2}{3}, \quad \Phi(0*, 0) = \frac{2}{3}, \quad \Phi(1*, 0) = \frac{1}{3}.$$

Then by applying the theorem we can calculate:

$$\begin{aligned} \lim_{t \rightarrow \infty} \Phi(00, t) &= \Phi(*0, 0) \times \Phi(0*, 0) = \frac{2}{9} \\ \lim_{t \rightarrow \infty} \Phi(01, t) &= \Phi(*1, 0) \times \Phi(0*, 0) = \frac{4}{9} \\ \lim_{t \rightarrow \infty} \Phi(10, t) &= \Phi(*0, 0) \times \Phi(1*, 0) = \frac{1}{9} \\ \lim_{t \rightarrow \infty} \Phi(11, t) &= \Phi(*1, 0) \times \Phi(1*, 0) = \frac{2}{9} \end{aligned}$$

which clearly shows that initially the system was not in equilibrium.

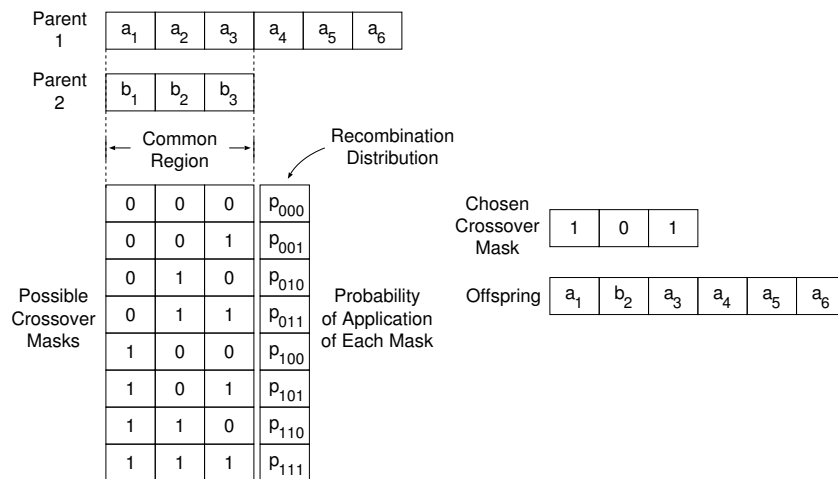


Figure 1 Homologous crossover on linear variable-length structures.

3 HOMOLOGOUS CROSSOVER AND EXACT SCHEMA THEOREM FOR LINEAR VARIABLE-LENGTH STRUCTURES

We call *homologous* those crossover operators where the offspring are created preserving the position of the genetic material taken from the parents. The one-point, two-point and uniform crossovers so often used in GAs operating on fixed length strings are specific instances in this class of operators, but the notion of homologous crossover has been recently generalised to deal with variable size strings and trees (Poli and McPhee 2001b).

In GAs acting on variable length strings or in linear GP systems, homologous crossovers first align the parents at one end, and identify the set of positions which are present in both parents (see top of Figure 1). This set is called the *common region*. The offspring is then created by copying the alleles in the common region from either parent. When the last allele in the common region is taken from the longer parent, then also all other alleles in that same parent to the right of the common region are copied. A more precise description of the class of homologous crossovers is given below. This requires extending the notions of crossover masks and recombination distributions used in genetics (Geiringer 1944) and in the GA literature (Booker 1992; Altenberg 1995; Spears 2000).

Let us start by noting that the common region contains as many loci as the shorter parent. Since these loci can be represented with consecutive integers from 1 to say c , we can just use the last of them, c , as a representation for the entire set. So, a common region containing loci 1, 2, and 3 could just be represented using the number 3. Therefore, the common region between parent strings of length j and k is simply $\min(j, k)$. For any given common region c we can then define a set of *crossover masks*, $\chi_c = \{0, 1\}^c$, which contains all different bit strings of size c (Poli and McPhee 2001b; Poli, Rowe, and McPhee 2001). Each crossover mask represents one way in which one could generate an offspring through

crossover. Except for the last bit in the mask, a 1 in the mask at a given locus means that, in the offspring, that locus will have to be filled with the corresponding allele in the first parent. A 0 in the mask at a given locus means that that locus will have to be filled with an allele from the second parent. The last bit in the mask has a similar meaning, but this time if the allele to be transferred to form the offspring belongs to the longer parent, then also the bits to its right are transferred (see Figure 1). With the definition of crossover mask in hand, we can now introduce the notion of recombination distribution.

A *recombination distribution* p_l^c gives the probability that, for a given common region c , crossover mask l will be chosen from the set \mathcal{X}_c (see Figure 1). Each variable-length homologous crossover is characterised by a different recombination distribution. For example, the recombination distribution for GP uniform crossover with 50% probability of exchanging primitives is $p_i^c = (0.5)^c$, where c is the size of the common region. Since the size 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 . It is important to note that, although normally the implementation of a specific type of homologous crossover will not be based on the notion of crossover mask and recombination distribution, the procedure described above could always be used to obtain a probabilistically equivalent implementation.

In (Poli, Stephens, Wright, and Rowe 2002) we derived the following exact schema theorem for homologous crossovers acting on variable length strings:

Theorem 1 *In a variable length GA or linear GP system in the presence of homologous crossover applied with 100% probability:*

$$E[\Phi(h_1 \dots h_N, t + 1)] = \sum_{k > 0} \sum_{l \in \mathcal{X}'_{N \downarrow k}} (p_l + p_{\bar{l}}) \times \quad (3)$$

$$p((h_1 \bullet l_1) \cdots (h_{N \downarrow k-1} \bullet l_{N \downarrow k-1}) h_{N \downarrow k} \cdots h_N, t) p((h_1 \bullet \bar{l}_1) \cdots (h_{N \downarrow k-1} \bullet \bar{l}_{N \downarrow k-1}) *^{k-N \downarrow k+1}, t)$$

where: a) $h_1 \dots h_N$ is a generic string or schema of length N , b) $\Phi(H, t)$ is the proportion of individuals in a schema H , c) $p(H, t)$ is the selection probability of the schema H , d) the operator \downarrow returns the minimum of its two arguments and it has higher precedence than $+$ and $-$, e) $\mathcal{X}'_{N \downarrow k} = \{0, 1\}^{N \downarrow k-1} \times \{1\}$ is the set of all crossover masks of length $N \downarrow k$ ending in 1, f) \bar{l} is the complement of crossover mask l , g) the notation l_i indicates the i -th element of bitmask l , and h) if a is a string (of length ≥ 1) then $a \bullet b = \begin{cases} a & \text{if } b = 1 \\ * & \text{otherwise.} \end{cases}$

Equation 3 can be used to study, among other things, the evolution of size in linear GP/GA systems (Poli, Stephens, Wright, and Rowe 2002). This is because it can be specialised to describe the future expected proportion of schemata of the form $*^N$. For example, it can be used to show that when a homologous crossover alone is acting, any initial distribution of lengths is a fixed-point length distribution for the system, i.e. $\forall t \geq 0, \Phi(*^N, t) = \Phi(*^N, 0)$.

4 EXTENSION OF GEIRINGER'S THEOREM TO VARIABLE-SIZE LINEAR REPRESENTATIONS

The extension of Geiringer's theorem to linear, variable-length structures and homologous 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 of the following result is available in (Poli, Stephens, Wright, and Rowe 2002).

Theorem 2 *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*

$$\Phi(h_1 h_2 \cdots h_N, t) = \Phi(*^{N-1} h_N, 0) \prod_{i=1}^{N-1} \frac{\Phi(*^{i-1} h_i \#, 0)}{\Phi(*^i \#, 0)}, \quad (4)$$

where $\#$ is a don't care symbol which stands for any string of length ≥ 1 and so $\frac{\Phi(*^{i-1} h_i \#, 0)}{\Phi(*^i \#, 0)}$ is the relative proportion of chromosomes carrying allele h_i in locus i out of all the chromosomes having such a locus at generation 0.¹

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. The class of these fixed points forms the Geiringer manifold for homologous crossovers.

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.² 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.

¹This is because $\Phi(*^{i-1} h_i \#, 0) = \sum_{n>0} \Phi(*^{i-1} h_i *^n, 0)$ is the proportion of chromosomes carrying allele h_i in locus i and $\Phi(*^i \#, 0) = \sum_{n>0} \Phi(*^{i+n}, 0)$ is the proportion of chromosomes having locus i .

²In the fixed length case on a flat landscape there is no evolution for the order-1 building blocks (in an infinite population). In the variable length case alleles can be transferred between length classes, and so, order-1 building blocks have a more complex dynamics (Stephens, Poli, Wright, and Rowe 2002).

4.2.1 Definitions

Before we can proceed with the proof we need to introduce definitions that are necessary to denote the building blocks of a generic schema and the probabilities that these will be affected by particular crossover masks.

Definition 3 Given a generic set loci $\{i, j, \dots, v\}$ in structures of length N occupied by alleles h_i, h_j, \dots, h_v , respectively, where $1 \leq i < j < \dots < v \leq N$, we define the following schemata

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

Definition 4 Let b_i, b_j, \dots, b_v be arbitrary binary digits and let l_i, l_j, \dots, l_v be the elements of crossover mask l at positions i, j, \dots, v , respectively. We define

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

where $\mathcal{X}'_c = \{0, 1\}^{c-1} \times \{1\}$.

Finally we need to focus our attention on two classes of “well-behaved” crossover operators.

Definition 5 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 $i \in \{1, \dots, c-1\}$.

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 $\hat{l} \in \mathcal{X}'_c$, 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 $\tilde{l} \in \mathcal{X}'_c$, having non-zero probability $p_{\tilde{l}}$ of being chosen, such that $\tilde{l}_i = 1$ and $\tilde{l}_c = 0$. Note also that we cannot require that $\mathcal{P}_c(l_c = 1) < 1$, since by definition $\mathcal{X}'_c = \{0, 1\}^{c-1} \times \{1\}$, and so $\mathcal{P}_c(l_c = 1) = 1$. That is, the terminals can never be transferred to strings of different length.

Definition 6 A GP homologous crossover operator acting on variable length linear structures is not fully linked if $0 < \mathcal{P}_c(l_i = b, l_j = b, \dots, l_v = b)$ for any choice of $b \in \{0, 1\}$ and $c > 0$, $\mathcal{P}_c(l_i = 0, l_j = 0, \dots, l_v = 0) + \mathcal{P}_c(l_i = 1, l_j = 1, \dots, l_v = 1) < 1$ for any choice of $c > 0$ and for any choice and number of loci i, j, \dots, v such that $0 < i < j < \dots < v < c$.

Note that a crossover operator which is not fully linked is also allele transferring.

We are now ready to proceed with the proof of convergence.

4.2.2 Base Step

Lemma 7 Let us consider an infinite population of linear, variable-length structures with an initial length distribution such that $\Phi(*^n, 0) = 0$ 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 4, i.e.

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

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 3 for an infinite population and a flat fitness:

$$\begin{aligned} & \Phi(H_i^N, t+1) \\ &= \sum_{k>0} \delta(i < N \downarrow k) \sum_{l \in \mathcal{X}'_{N \downarrow k}} (p_l + p_{\bar{l}}) \Phi(*^{i-1}(h_i \bullet l_i) *^{N-i}, t) \Phi(*^{i-1}(h_i \bullet \bar{l}_i) *^{k-i}, t) \\ &+ \sum_{k>0} \delta(i \geq N \downarrow k) \sum_{l \in \mathcal{X}'_{N \downarrow k}} (p_l + p_{\bar{l}}) \Phi(H_i^N, t) \Phi(*^k, t), \end{aligned} \quad (5)$$

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

$$\begin{aligned} \Phi(H_N^N, t+1) &= \sum_{k>0} \sum_{l \in \mathcal{X}'_{N \downarrow k}} (p_l + p_{\bar{l}}) \Phi(H_N^N, t) \Phi(*^k, t) \\ &= \Phi(H_N^N, t) \sum_{k>0} \Phi(*^k, t) \overbrace{\sum_{l \in \mathcal{X}'_{N \downarrow k}} (p_l + p_{\bar{l}})}^{=1} = \Phi(H_N^N, t) \underbrace{\sum_{k>0} \Phi(*^k, t)}_{=1} = \Phi(H_N^N, t) \end{aligned} \quad (6)$$

which proves the second part of the lemma.

We can now go back to Equation 5 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 5 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

$$\Phi(H_i^N, t+1) = \sum_{k>i} \sum_{l \in \mathcal{X}'_{N \downarrow k}} (p_l + p_{\bar{l}}) \Phi(*^{i-1}(h_i \bullet l_i) *^{N-i}, t) \Phi(*^{i-1}(h_i \bullet \bar{l}_i) *^{k-i}, t)$$

$$\begin{aligned}
& + \sum_{0 < k \leq i} \sum_{l \in \mathcal{X}'_{N \downarrow k}} (p_l + p_{\bar{l}}) \Phi(H_i^N, t) \Phi(*^k, t) \\
= & \sum_{k > i} \sum_{\substack{l \in \mathcal{X}'_{N \downarrow k} \\ l_i = 0}} (p_l + p_{\bar{l}}) \Phi(*^N, t) \Phi(H_i^k, t) \\
& + \sum_{k > i} \sum_{\substack{l \in \mathcal{X}'_{N \downarrow k} \\ l_i = 1}} (p_l + p_{\bar{l}}) \Phi(H_i^N, t) \Phi(*^k, t) \\
+ & \Phi(H_i^N, t) \sum_{0 < k \leq i} \Phi(*^k, t) \underbrace{\sum_{l \in \mathcal{X}'_{N \downarrow k}} (p_l + p_{\bar{l}})}_{=1} \\
= & \sum_{k > i} \Phi(H_i^k, t) \left(\mathcal{P}_{N \downarrow k}(l_i = 0) \Phi(*^N, t) \right) \\
+ & \Phi(H_i^N, t) \left(\sum_{k > i} \mathcal{P}_{N \downarrow k}(l_i = 1) \Phi(*^k, t) + \sum_{0 < k \leq i} \Phi(*^k, t) \right) \\
= & \sum_{k > i, k \neq N} \Phi(H_i^k, t) \left(\mathcal{P}_{N \downarrow k}(l_i = 0) \Phi(*^N, t) \right) \tag{7} \\
+ & \Phi(H_i^N, t) \left(\sum_{k > i, k \neq N} \mathcal{P}_{N \downarrow k}(l_i = 1) \Phi(*^k, t) + \sum_{0 < k \leq i} \Phi(*^k, t) + \Phi(*^N, t) \right).
\end{aligned}$$

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 $N = i + 1, i + 2, \dots, N_m$, 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 $\Phi(H_i^{i+1}, t), \dots, \Phi(H_i^{N_m}, t)$. Since $\Phi(*^k, t) = \Phi(*^k, 0)$ for any $t > 0$ (see end of Section 3), the system has the form

$$x(t+1) = Ax(t), \tag{8}$$

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

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

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, let us assume $\Phi(*^n, 0) > 0$ for $0 < n \leq N_m$. Then, because the crossover operator is allele-transferring by hypothesis, 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. $\sum_u a_{uv} = 1$:

$$\begin{aligned}
\sum_u a_{uv} &= a_{vv} + \sum_{u \neq v} a_{uv} \\
&= \sum_{k > i, k \neq i+v} \mathcal{P}_{(i+v)\downarrow k}(l_i = 1) \Phi(*^k, 0) + \sum_{0 < k \leq i} \Phi(*^k, 0) + \Phi(*^{i+v}, 0) \\
&\quad + \sum_{u > 0, u \neq v} \mathcal{P}_{(i+u)\downarrow(i+v)}(l_i = 0) \Phi(*^{i+u}, 0) \\
&= \sum_{k > i, k \neq i+v} \mathcal{P}_{k\downarrow(i+v)}(l_i = 1) \Phi(*^k, 0) + \sum_{0 < k \leq i} \Phi(*^k, 0) + \Phi(*^{i+v}, 0) \\
&\quad + \sum_{k > i, k \neq i+v} \mathcal{P}_{k\downarrow(i+v)}(l_i = 0) \Phi(*^k, 0)
\end{aligned}$$

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

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

Therefore, A is a primitive stochastic matrix and the Perron-Frobenius theorem (see for example (Davis and 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 $r = 1$

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

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

$$\lim_{t \rightarrow \infty} x(t) = \mathbf{a} \sum_{n > i} \Phi(H_i^n, 0) = \mathbf{a} \Phi(*^{i-1} h_i \#, 0). \tag{9}$$

So, a unique fixed point for $x(t)$ exists. Using Theorem 2 it is easy to show that a fixed point for $x(t)$ is

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

which therefore must be a unique global attractor for the system.³

Earlier we assumed that $\Phi(*^n, 0) > 0$ for $n = 1, \dots, N_m$. However, the proof is still valid even if $\Phi(*^N, 0) = 0$ for some $N \in \{1, \dots, N_m\}$. Indeed, for such an N , Equation 6 would still be valid, while Equation 7 would collapse to

$$\Phi(H_i^N, t+1) = \Phi(H_i^N, t) \left(\sum_{k>i, k \neq N} \mathcal{P}_{N \downarrow k}(l_i = 1) \Phi(*^k, t) + \sum_{0 < k \leq i} \Phi(*^k, t) + \Phi(*^N, t) \right).$$

Since the condition $\Phi(*^N, 0) = 0$ implies that $\Phi(H_i^N, 0) = 0$ for $0 < i \leq N$, then $\forall t$, $\Phi(H_i^N, t) = 0$ for $0 < i \leq N$. So, if we remove from the matrix A and the vector $x(t)$ in Equation 8 the rows/columns corresponding to the variables $\Phi(H_i^N, t)$ for every N such that $\Phi(*^N, 0) = 0$, we still have a Markov chain with primitive transition matrix, and so the proof provided above still holds. \square

4.2.3 Induction Step and Theorem

The induction step involves showing that the schema evolution equations for schemata of order higher than 1 have the form $x(t+1) = Ax(t) + b(t)$, where A and $b(t)$ satisfy the following

Lemma 8 *Given a system of linear difference equations of the form*

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

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 $\lim_{t \rightarrow \infty} b(t)$ exists, then the system has a unique, global attractor.

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

$$x = Ax + b, \tag{11}$$

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 $t \rightarrow \infty$ of both sides of Equation 10, we obtain an equation exactly like Equation 11 (note, we know that the limit for $b(t)$ exists). So, the theorem in (Bellman 1960) tells us that a fixed point for Equation 10 exists and is unique. \square

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 2 is a unique global attractor for the system.

We are now ready to prove the following

Theorem 9 (Geiringer's theorem for variable-length strings) *For a not fully linked GP homologous crossover operator, the fixed point for the schema evolution equations provided in Theorem 2 is a unique global attractor for the system.*

³Incidentally, this allows us to see that $\mathbf{a} = \frac{1}{\Phi(*^i \#, 0)} [\Phi(*^{i+1}, 0), \Phi(*^{i+2}, 0), \dots, \Phi(*^{N_m}, 0)]^T$.

Proof: We will proceed by induction on the number of defining primitives in the schema. Lemma 7 proves the base step, i.e. shows that order 1 schemata converge to the fixed point provided in Theorem 2. As an induction hypothesis we will assume that schemata of order 1, 2, ..., o converge to the fixed points provided in Theorem 2 and will show that this is also true for schemata of order $o + 1$.

By specialising Equation 3 for a generic schema $H_{i,j,n,\dots,w,v}^N$ of order $o + 1$, it is possible to prove (the calculations are rather tedious and are omitted for brevity) that the schema equations form a linear system of difference equations

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

where $x(t) = [\Phi(H_{i,\dots,v}^v, t), \Phi(H_{i,\dots,v}^{v+1}, t), \dots, \Phi(H_{i,\dots,v}^{N_m}, t)]^T$ 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

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

The off-diagonal elements of A are

$$a_{qr} = \begin{cases} \mathcal{P}_{v+r}(l_i = 0, l_j = 0, \dots, l_v = 0) \Phi(*^{v+q}, 0) & \text{if } r < q, \\ \mathcal{P}_{v+q}(l_i = 0, l_j = 0, \dots, l_v = 0) \Phi(*^{v+q}, 0) & \text{if } r > q. \end{cases}$$

The elements of $b(t)$ are given by

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

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

Let us highlight some properties of the matrix A . Firstly, let us assume that $\Phi(*^n, 0) > 0$ for $0 < n \leq N_m$. Then, because the crossover operator is not fully linked by hypothesis, A is a positive matrix, i.e. $a_{qr} > 0$ for all q and r . Also, it is possible to prove (again the calculations are omitted for brevity) that $\sum_q a_{qr} < 1$ 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 $\{i, j, n, \dots, w, v\}$ and $\bar{S} = \{i, j, n, \dots, w, v\} \setminus S$. So, since $|\{i, j, n, \dots, w, v\}| = o + 1$, 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 , $\Phi(H_{i,\dots,w}^k, t)$ 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, $\lim_{t \rightarrow \infty} b(t)$ exists.

In conclusion, A and $b(t)$ satisfy the conditions of Lemma 8, and so a fixed point for $x(t)$ exists and is unique. Since, Theorem 2 gives us a fixed point, then that must be the only one and must be a global attractor for the system.

By proceeding like in the proof of Lemma 7, it is easy to show that this proof is still valid even if $\Phi(*^N, 0) = 0$ for some $N \in \{1, \dots, N_m\}$. \square

5 EXAMPLE

As an example, let us consider a population of linear computer programs constructed using the primitive set $\{x, y, \sin, \sqrt{\cdot}\}$ having the following initial proportions: $\Phi(x, 0) =$

$\Phi(y, 0) = \Phi(\sqrt{x}, 0) = \Phi(\sqrt{y}, 0) = \Phi(\sin x, 0) = \frac{1}{5}$, $\Phi(\sin y, 0) = 0$, and $\Phi(*^n, 0) = 0$ for $n \geq 3$. To apply the theorem to calculate where homologous crossover will lead this population, first we need to calculate the following primitive frequencies: $\Phi(x, 0) = \frac{1}{5}$, $\Phi(y, 0) = \frac{1}{5}$, $\Phi(*x, 0) = \frac{2}{5}$, $\Phi(*y, 0) = \frac{1}{5}$, $\Phi(\sqrt{*}, 0) = \frac{2}{5}$, and $\Phi(\sin *, 0) = \frac{1}{5}$. Then we can feed these into the theorem to obtain:

$$\begin{aligned} \lim_{t \rightarrow \infty} \Phi(x, t) &= \Phi(x, 0) = \frac{1}{5} \\ \lim_{t \rightarrow \infty} \Phi(y, t) &= \Phi(y, 0) = \frac{1}{5} \\ \lim_{t \rightarrow \infty} \Phi(\sqrt{x}, t) &= \Phi(*x, 0) \frac{\Phi(\sqrt{*}, 0)}{\Phi(**, 0)} = \frac{2}{5} \times \frac{2}{3} = \frac{4}{15} \\ \lim_{t \rightarrow \infty} \Phi(\sqrt{y}, t) &= \Phi(*y, 0) \frac{\Phi(\sqrt{*}, 0)}{\Phi(**, 0)} = \frac{1}{5} \times \frac{2}{3} = \frac{2}{15} \\ \lim_{t \rightarrow \infty} \Phi(\sin x, t) &= \Phi(*x, 0) \frac{\Phi(\sin *, 0)}{\Phi(**, 0)} = \frac{2}{5} \times \frac{1}{3} = \frac{2}{15} \\ \lim_{t \rightarrow \infty} \Phi(\sin y, t) &= \Phi(*y, 0) \frac{\Phi(\sin *, 0)}{\Phi(**, 0)} = \frac{1}{5} \times \frac{1}{3} = \frac{1}{15} \end{aligned}$$

So, clearly initially the system was not in linkage equilibrium and therefore there is evolution even if no selection pressure is present.

6 DISCUSSION AND CONCLUSIONS

While for GAs operating on fixed length strings it has been known for a very long time that the effect of homologous crossovers is to shuffle the primitives present in different individuals and push the chromosome distribution towards locus-wise independence (a situation often also labelled as linkage equilibrium, Robbins' proportions, or perfectly mixed population), before the work described in this paper and in (Poli, Stephens, Wright, and Rowe 2002) it was unclear where homologous crossovers would try to push a population of variable length chromosomes. Obviously, because GP homologous crossovers and the linear representation used here are generalisations of the corresponding operators and representations used in fixed-length GAs, by continuity one would have expected that the population would move towards something similar to the manifold described by Geiringer's original theorem. However, what this generalised Geiringer manifold would look like was very unclear, although we knew it had to be a fixed point for a specialised version of the GP exact schema evolution equations (Poli and McPhee 2001b) or of the Vose-like model in (Poli, Rowe, and McPhee 2001).

By using our intuition and a few trial-and-error iterations we were able to come up with an entire family of fixed points (Equation 3) for the schema equations and we empirically demonstrated how the system seemed to always converge towards them in the work described in (Poli, Stephens, Wright, and Rowe 2002). The fact that this result nicely collapsed to Equations 1 and 2 suggested that we were on the right track to generalise Geiringer's theorem. However, a vital component was missing: a proof that the family of

fixed points we had found was a global attractor for a linear GP system or a variable length GA undergoing repeated homologous crossover. In this paper we have solved this problem and we have generalised Geiringer's original work.

This result tells us not only that when acting on variable length structures, too, homologous crossovers shuffle the primitives present in different individuals and push the chromosome distribution towards locus-wise independence, but it also clarifies what it means for a chromosome distribution to be locus-wise independent, i.e. what it means for a population of variable length strings to be in linkage equilibrium (or perfectly mixed or in Robbins' proportions). As Equation 3 clearly shows, the system is in linkage equilibrium when the probability of finding a generic string in the population equals the product of the conditional probabilities of finding each primitive of the string at the appropriate non-terminal locus (given that we restrict our attention to the strings in the population which are long enough to contain such a locus) times the probability of finding an appropriate primitive for the terminal locus within strings containing such a terminal locus.

Although the study reported in this paper was scientific in nature, characterisations of the genetic biases of the operators such as the ones offered here and in (Poli, Rowe, Stephens, and Wright 2002) may provide practical recipes for practitioners too. We discuss two examples below.

Firstly, characterisations of the genetic biases of the operators 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. For example, in (Poli, Rowe, Stephens, and Wright 2002) we studied the bias of a crossover operator (GP subtree crossover) acting on linear structures in which two crossover points (one for each parent) are independently selected and the offspring is created by taking the l.h.s. of one parent and the r.h.s of the other. This operator induces a diffusion process by which the primitives in a particular individual tend not just to be swapped with those of other individuals in the population, but also to diffuse *within* the representation of each individual. More precisely, crossover attempts to push the population towards distributions of primitives where each primitive is equally likely to be found in any position in any individual. Clearly, if prior knowledge indicates that for a specific class of problems it is expected that high quality solutions will have inhomogeneous distributions of primitives, maybe the bias provided by homologous crossovers might seem more appropriate than that of GP subtree crossover.

Secondly, when using fixed length GAs most people will normally initialise the population randomly. This generally creates a population that is very close to being in linkage equilibrium. This may be exactly what people want: since the intended behaviour of a GA is one where the selection operator (i.e. the fitness function) is to guide the choice of good areas to sample, while crossover performs the sampling, an initialisation procedure that minimises the bias of crossover in the initial, often-critical generations may be what is desirable. However, in a variable length GA, how do we know whether we are initialising our population well in this respect? To clarify the point, let us consider the simplest of cases: we initialise the population using only copies of the same individual. In fixed-length and variable-length GAs under homologous crossover, the population would be in linkage equilibrium as one would expect, but, under GP subtree crossover, it would not. Fortunately, thanks to our fixed-point results, one can now start comparing the many

initialisation algorithms described in the literature and see how close to the Geiringer manifold the distribution of strings they provide is for different classes of operators. Then, one could choose an initialisation strategy that best minimise the crossover bias (e.g. in (McPhee and Poli 2001) we found that, for variable-length linear GP systems under subtree crossover, one such initialisation strategy would be to produce programs with Gamma distributed lengths).

Finally, our results are applicable to the case of linear variable length structures. However, we have no reason to believe that the situation would be significantly different in tree-based GP. Indeed, because our extension of Geiringer's theorem is effectively the result of specialising and studying the GP schema theorem's equations, it is very likely that in the future we will be able to extend Geiringer's theorem to tree-based GP. The first step towards this further extension will be finding a good notation to express the fixed-point primitive distribution.

Acknowledgments

The authors would like to thank the anonymous reviewers for their helpful comments. CRS would like to thank the University of Birmingham for a visiting Professorship and DGAPA-PAPIIT grant IN100201. RP and CRS would like to thank the Royal Society and the University of Essex for their support. Alden Wright did this work while visiting the University of Birmingham, supported by EPSRC grant GR/R47394. RP would like to thank the members of the NEC (Natural and Evolutionary Computation) group at Essex for helpful comments and discussion.

References

- Altenberg, L. (1995, 31 July–2 August 1994). The Schema Theorem and Price's Theorem. In L. D. Whitley and M. D. Vose (Eds.), *Foundations of Genetic Algorithms 3*, Estes Park, Colorado, USA, pp. 23–49. Morgan Kaufmann.
- Bellman, R. (1960). *Introduction to Matrix Analysis*. New York: McGraw-Hill.
- Booker, L. B. (1992, 24–29 July). Recombination distributions for genetic algorithms. In *FOGA-92, Foundations of Genetic Algorithms*, Vail, Colorado.
- Booker, L. B., D. B. Fogel, D. Whitley, P. J. Angeline, and A. E. Eiben (2000). Recombination. In T. Bäck, D. B. Fogel, and T. Michalewicz (Eds.), *Evolutionary Computation 1: Basic Algorithms and Operators*, Chapter 33. Institute of Physics Publishing.
- Davis, T. E. and J. C. Principe (1993). A Markov chain framework for the simple genetic algorithm. *Evolutionary Computation* 1(3), 269–288.
- Geiringer, H. (1944, March). On the probability theory of linkage in Mendelian heredity. *Annals of Mathematical Statistics* 15(1), 25–57.
- McPhee, N. F. and R. Poli (2001, 18-20 April). A schema theory analysis of the evolution of size in genetic programming with linear representations. In *Genetic Programming, Proceedings of EuroGP 2001*, LNCS, Milan. Springer-Verlag.
- McPhee, N. F., R. Poli, and J. E. Rowe (2001, 27-30 May). A schema theory analysis of mutation size biases in genetic programming with linear representations. In *Proceedings of the 2001 Congress on Evolutionary Computation CEC2001*, COEX, World Trade Center, 159 Samseong-dong, Gangnam-gu, Seoul, Korea, pp. 1078–1085. IEEE Press.

- O'Neill, M. and C. Ryan (2001). Grammatical evolution. *IEEE Transaction on Evolutionary Computation*. Forthcomming.
- Poli, R. (2001, June). 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. and N. F. McPhee (2001a, 18-20 April). 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, Milan. Springer-Verlag.
- Poli, R. and N. F. McPhee (2001b, 7-11 July). Exact schema theory for GP and variable-length GAs with homologous crossover. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2001)*, San Francisco, California, USA. Morgan Kaufmann.
- Poli, R., J. E. Rowe, and N. F. McPhee (2001, 7-11 July). Markov chain models for GP and variable-length GAs with homologous crossover. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2001)*, San Francisco, California, USA. Morgan Kaufmann.
- Poli, R., J. E. Rowe, C. R. Stephens, and A. H. Wright (2002). Allele diffusion in linear genetic programming and variable-length genetic algorithms with subtree crossover. In *Proceedings of EuroGP 2002*.
- Poli, R., C. R. Stephens, A. H. Wright, and J. E. Rowe (2002, 9-13 July). On the search biases of homologous crossover in linear genetic programming and variable-length genetic algorithms. In W. B. Langdon, E. Cantú-Paz, K. Mathias, R. Roy, D. Davis, R. Poli, K. Balakrishnan, V. Honavar, G. Rudolph, J. Wegener, L. Bull, M. A. Potter, A. C. Schultz, J. F. Miller, E. Burke, and N. Jonoska (Eds.), *GECCO 2002: Proceedings of the Genetic and Evolutionary Computation Conference*, New York, pp. 868–876. Morgan Kaufmann Publishers.
- Spears, W. M. (2000, July). Limiting distributions for mutation and recombination. In W. M. Spears and W. Martin (Eds.), *Proceedings of the Foundations of Genetic Algorithms Workshop (FOGA 6)*, Charlottesville, VA, USA. In press.
- Stephens, C. R., R. Poli, A. H. Wright, and J. E. Rowe (2002, 9-13 July). Exact results from A coarse grained formulation of the dynamics of variable-length genetic algorithms. In W. B. Langdon, E. Cantú-Paz, K. Mathias, R. Roy, D. Davis, R. Poli, K. Balakrishnan, V. Honavar, G. Rudolph, J. Wegener, L. Bull, M. A. Potter, A. C. Schultz, J. F. Miller, E. Burke, and N. Jonoska (Eds.), *GECCO 2002: Proceedings of the Genetic and Evolutionary Computation Conference*, New York, pp. 578–585. Morgan Kaufmann Publishers.
- Stephens, C. R. and H. Waelbroeck (1997). Effective degrees of freedom in genetic algorithms and the block hypothesis. In T. Bäck (Ed.), *Proceedings of the Seventh International Conference on Genetic Algorithms (ICGA97)*, East Lansing, pp. 34–40. Morgan Kaufmann.
- Vose, M. D. (1999). *The simple genetic algorithm: Foundations and theory*. Cambridge, MA: MIT Press.
- Vose, M. D. and A. H. Wright (1998). The simple genetic algorithm and the Walsh transform: Part II, the inverse. *Evolutionary Computation* 6(3), 275–289.