A Normed Space of Genetic Operators with Applications to Scalability Issues

Jonathan E. Rowe   J.E.Rowe@cs.bham.ac.uk
School of Computer Science, University of Birmingham, Birmingham B15 2TT, UK

Abstract
We define an abstract normed vector space where the genetic operators are elements. This is used to define the *disturbance* of the generational operator $\mathcal{G}$ as the distance between the crossover and mutation operator (combined) and the identity. This quantity appears in a bound on the variance of fixed-point populations, and in a bound on the force $\|\mathbf{v} - \mathcal{G}(\mathbf{v})\|$ that applies to the optimal population $\mathbf{v}$. When analyzed for the case of fixed-length binary strings, a connection is shown between these measures and the size of the search space. Guides for parameter settings are given, if population convergence is required as the string length tends to infinity.

Keywords
Normed spaces, dynamical systems model, infinite population, genetic operators, scalability, population variance.

1 Introduction

It is well known that a fixed-point population of the simple genetic algorithm (comprising proportional selection, crossover, and mutation) may not necessarily be uniform (consist of copies of a single individual) in the infinite population limit. In fact, if mutation is non-zero, then the fixed-point population will definitely be mixed (see Eigen et al. (1989), Vose and Wright (1994), and van Nimwegen et al. (1997)). Intuitively, an increase in the size of the search space might well increase the spread of the population, since there will be a greater number of possible individuals. A major concern of this paper is to analyze the effects of the GA parameter settings on this phenomenon. In particular, a means of setting the mutation and crossover rates will be given that guarantees the convergence of the fixed-point population in the limit as the size of the search space increases.

It is also known that an actual finite population only approximates the flow described by the infinite population model, since there are only a finite number of possible populations. In particular, if the infinite population model predicts a small change from one generation to the next, then an actual population might remain fixed at what has been called a "metastable" state (see Vose and Liepins (1991), Vose (1999a), van Nimwegen et al. (1999), and Rowe (1998)). A second concern of this paper will be to look at whether or not populations will stall in this way in regions close to the optimal solution. Again, the parameter settings proposed will guarantee that the optimal solution will increasingly become a stalling region as the size of the search space increases. This does not guarantee that the optimal solution will be found. Rather, it guarantees that if the genetic algorithm should happen to produce a population mostly comprising copies of the optimal solution, then it is likely to settle on this population for some time.

In this paper, we view genetic operators as elements of an abstract vector (or linear) space. This enables the application of simple results from functional analysis (see Maddox (1988) for an introduction). We will be concerned with the "distance" of the operator representing crossover and mutation from the identity operator. This quantity gives a measure of how much these operators are disturbing the population, and it appears in the bounds that are derived for the variance of the fitnesses within a fixed-point population and for the force that applies to a population consisting entirely of copies of the optimal solution. Much of the paper is concerned with setting up the appropriate abstract spaces for the different genetic operators and then deriving these bounds. This is done within a general setting. The application to fixed-length binary strings then gives rise to the results concerning the effects of increasing search space size.

2 Definitions

Suppose we have a finite search space Ω of size n. To simplify notation, we identify the elements of Ω with the integers $\{0, 1, \ldots, n-1\}$. The dynamical systems model for the simple genetic algorithm (SGA) (Vose, 1999a; Vose and Wright, 1994, 1995) represents populations as vectors $\mathbf{p} = (p_0, \ldots, p_{n-1})$, so that $\mathbf{p} \in \Re^n$. In this vector, p_j is the proportion of the population taken up by the individual $j \in \Omega$. For any population, $\sum_{j=0}^{n-1} p_j = 1$. To record this property, we will make use of the function $h : \Re^n \to \Re$ defined by

$$h(\mathbf{x}) = \sum_{j=0}^{n-1} x_j$$

for any $\mathbf{x} \in \Re^n$. This is an example of a *bounded linear functional* on $\Re^n$. A linear functional on a vector space is a function that maps elements of the vector space to values in the underlying scalar field (in our case, the real numbers) that is also linear. That h is bounded is a consequence of our vector space being finite-dimensional. The set of bounded linear functionals on a vector space X is called the *dual space*, denoted by $X^\dagger$. It is important to realize that the set $X^\dagger$ can itself be considered a vector space over the same scalar field as X.

We can now write the set of all possible populations as:

$$\Lambda = \left\{ \mathbf{x} = (x_0, \ldots, x_{n-1}) : h(\mathbf{x}) = 1, x_j \geq 0 \right\},$$

which is known as the *simplex*.

Given a fitness function $f : \Omega \to \Re^+$, the effect of proportional selection on a population is given by a diagonal matrix

$$S_{k,k} = f(k),$$

where $f(k)$ is the fitness of individual k. The selection operator $\mathcal{F}$ is then

$$\mathcal{F}(\mathbf{p}) = \frac{S\mathbf{p}}{h(S\mathbf{p})}$$

for all $\mathbf{p} \in \Lambda$. This operator is also well defined for the whole of $\Re^n$, except for those points $\mathbf{x} \in \Re^n$ where $h(S\mathbf{x}) = 0$. When $\mathcal{F}$ acts on a point in the simplex, it gives the expected next population after the application of proportional selection. Note that Λ is invariant with respect to $\mathcal{F}$. The mean fitness of a population $\mathbf{p}$ is given by $\mathbf{Mean}[\mathbf{p}] = h(S\mathbf{p})$.

The effects of applying mutation and crossover to a population can also be represented by operators. We denote the mutation operator by $\mathcal{U} : \Re^n \to \Re^n$ and the crossover operator by $\mathcal{C} : \Re^n \to \Re^n$. We look at these in more detail in the next two sections. For now, we note the following result due to Vose and Wright (1995). For mutation and crossover operators in standard use,

$$\mathcal{C} \circ \mathcal{U} = \mathcal{U} \circ \mathcal{C},$$

where $\circ$ denotes operator composition. That is, the order in which mutation and crossover are performed is irrelevant; the result is the same. We denote the two operators acting together by $\mathcal{M}$. The full effect of one generation of the SGA on a population $\mathbf{p}$ is given by

$$\mathcal{G}(\mathbf{p}) = \mathcal{M} \circ \mathcal{F}(\mathbf{p}).$$

The remainder of the paper proceeds as follows. First, the mutation operator is shown to be linear, representable by a matrix U. The set of bounded linear operators on a space may be given a *norm*. This is calculated for U. Second, crossover is examined from the point of view of the space of *bilinear* operators. The linear space of *quadratic* operators is defined, of which the pure crossover operator is a member. Again, a norm is established on this space. It is then shown how linear operators that are invariant on Λ are equivalent to certain quadratic operators, and that therefore, under this equivalence, they are also members of this space. We can, at this stage, talk sensibly about the distances between operators, sequences of operators, and the convergence of such sequences. We will use this normed space to derive some bounds on two properties of the SGA: the variance of the fitness of a fixed-point population, and the stability of the population comprising copies of the optimum solution. In the case where the search space comprises fixed-length binary strings, these bounds will be seen to depend on the size of the search space. We will then be able to address scalability issues, particularly with regard to the appropriate choice of operator as the size of the search space grows.

In order to define norms on the different spaces of operators, we need a base norm on the linear space $\Re^n$. Rather than using the usual Euclidean norm, we will use the absolute norm,

$$\|\mathbf{x}\| = \sum_{j=0}^{n-1} |x_j|,$$

which has some nice properties in this context. For example,

$$\mathbf{p} \in \Lambda \Rightarrow \|\mathbf{p}\| = h(\mathbf{p}) = 1,$$

and therefore, the simplex makes up part of the unit sphere, as defined by this norm. Similarly,

$$\mathbf{p} \in \Lambda \Rightarrow \|S\mathbf{p}\| = h(S\mathbf{p}) = \mathbf{Mean}[\mathbf{p}].$$

3 Mutation is a Linear Operator

Mutation is an operator that works by changing (or mutating), with a certain probability, one individual into another. The operator $\mathcal{U}$ is, therefore, given by a matrix U, where $U_{i,j}$ is the probability that $j \in \Omega$ is mutated into $i \in \Omega$.

It should be noted that any linear operator from $\Re^n \to \Re^n$ can be represented uniquely by a corresponding $n \times n$ matrix. It is often convenient to blur the distinction between a linear operator and the matrix to which it corresponds. In particular, since

U is a finite-dimensional matrix, we can immediately deduce that it is a bounded (and therefore continuous) linear operator on $\Re^n$.

The set of all bounded linear operators on a normed vector space X is denoted $\mathcal{B}(X)$ and is itself a vector space under the following definitions. If $A, B \in \mathcal{B}(X)$ and λ, μ are scalars, then define

$$(\lambda A + \mu B)(\mathbf{x}) = \lambda(A\mathbf{x}) + \mu(B\mathbf{x})$$

for any $\mathbf{x} \in X$. The space $\mathcal{B}(X)$ can also be given a norm. For any $A \in \mathcal{B}(X)$, we define

$$\|A\|_L = \sup_{\|\mathbf{x}\|=1} \|A\mathbf{x}\| \tag{1}$$

(the subscript L indicating a space of *linear* operators). It follows that $\|A\mathbf{x}\| \leq \|A\|_L \|\mathbf{x}\|$ for any $\mathbf{x} \in X$ and $A \in \mathcal{B}(X)$, and moreover, $\|A\|_L$ is the smallest constant for which this is true for all $\mathbf{x}$.

We are particularly interested in linear operators that are invariant on Λ (that map Λ into itself). For such operators, we have the following result.

THEOREM 1: *Let $A : \Lambda \to \Lambda$ be linear. Then $\|A\|_L = 1$.*

PROOF: Consider A as an $n \times n$ matrix. Let $\mathbf{e}_j$ be the vector with a 1 in position j and zeros elsewhere. The ith component of $A\mathbf{e}_j$ is $A_{i,j}$. But since $\mathbf{e}_j \in \Lambda$, we know that $A\mathbf{e}_j \in \Lambda$. Therefore,

- $A_{i,j} \geq 0$ for all i, j
- $\sum_i A_{i,j} = 1$ for all j

Thus all the entries of the matrix A are non-negative, and the columns of A sum to 1. Given that we are using the absolute norm on $\Re^n$, it is a standard result (Kantorovich and Akilov, 1982) that, for *any* matrix A,

$$\|A\|_L = \max_j \sum_{i=0}^{n-1} |A_{i,j}|. \tag{2}$$

The result follows. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

COROLLARY 1: *Any mutation matrix $U \in \mathcal{B}(\Re^n)$ has norm 1.*

This means that, for any $\mathbf{x} \in \Re^n$, $\|U\mathbf{x}\| \leq \|\mathbf{x}\|$. In fact, mutation is norm-preserving:

$$\|U\mathbf{x}\| = \|\mathbf{x}\|$$

for all $\mathbf{x} \in \Re^n$. Since U is invariant on the simplex, we have $\mathbf{p} \in \Lambda \Rightarrow \|U\mathbf{p}\| = \|\mathbf{p}\| = 1$. Notice that this holds for any finite search space Ω and for any means of defining mutation. In particular, it is not restricted to a search space comprising binary strings with bitwise mutation.

THEOREM 2: *The set of linear operators in $\mathcal{B}(\Re^n)$ that are invariant on Λ is convex.*

PROOF: Let $A, B \in \mathcal{B}(\Re^n)$ be invariant on Λ, and let $\lambda \in [0, 1]$. Then, for any $\mathbf{x} \in \Lambda$, let

$$\mathbf{y} = (\lambda A + (1 - \lambda)B)(\mathbf{x}) = \lambda A\mathbf{x} + (1 - \lambda)B\mathbf{x}$$

Then

$$
\begin{aligned}
h(\mathbf{y}) &= h(\lambda A\mathbf{x} + (1-\lambda)B\mathbf{x}) \\
&= \lambda h(A\mathbf{x}) + (1-\lambda)h(B\mathbf{x}) \\
&= \lambda + 1 - \lambda \\
&= 1
\end{aligned}
$$

And for all $k = 0, \ldots n-1$,

$$
\begin{aligned}
\mathbf{y}_k &= (\lambda A\mathbf{x} + (1-\lambda)B\mathbf{x})_k \\
&= \lambda(A\mathbf{x})_k + (1-\lambda)(B\mathbf{x})_k \\
&\geq 0
\end{aligned}
$$

Therefore, $\lambda A + (1-\lambda)B$ is invariant on Λ. $\qquad\square$

The selection matrix S, defined above, is also a bounded linear operator (though not necessarily invariant on Λ). That is, $S \in \mathcal{B}(\Re^n)$. Its norm is

$$
\|S\|_L = \max_k S_{k,k} = \max_k \{f(k)\},
$$

since the fitness function f is non-negative. $\|S\|$ is therefore equal to the maximum fitness (i.e., the global optimum).

Recall that the space of bounded linear functionals on a vector space is itself a vector space, called the dual space. If the underlying space has a norm, then a norm can be defined for the dual.

$$
\|g\| = \sup_{\|\mathbf{x}\|=1} |g(\mathbf{x})|
$$

for any $g \in X^\dagger$. In our case, $X = \Re^n$ and $h \in \Re^{n\dagger}$. It is easy to show that $\|h\| = 1$.

4   Pure Crossover is a Quadratic Operator

In this section, we start by considering a pure crossover operator. Given a population distribution $\mathbf{p} \in \Lambda$ from which the first parent is drawn and $\mathbf{q} \in \Lambda$ for the second parent, let $\omega(\mathbf{p}, \mathbf{q}) \in \Lambda$ be the population distribution resulting from crossing over random parents from $\mathbf{p}$ and $\mathbf{q}$.

Let $\mathbf{e}_i$ be the vector with a 1 at position i and zeros elsewhere. Thus $\mathbf{e}_i$ represents a population comprised entirely of copies of i. Then the kth component of $\omega(\mathbf{e}_i, \mathbf{e}_j)$ is equal to the probability that crossing i and j will produce k. Therefore,

$$
\omega(\mathbf{p}, \mathbf{q})_k = \sum_{i,j} p_i q_j \omega(\mathbf{e}_i, \mathbf{e}_j)_k,
$$

and thus,

$$
\begin{aligned}
\omega(\mathbf{p}, \mathbf{q}) &= \sum_k \left(\sum_{i,j} p_i q_j \omega(\mathbf{e}_i, \mathbf{e}_j)_k \right) \mathbf{e}_k \\
&= \sum_{i,j} p_i q_j \sum_k \omega(\mathbf{e}_i, \mathbf{e}_j)_k \mathbf{e}_k \\
&= \sum_{i,j} p_i q_j \omega(\mathbf{e}_i, \mathbf{e}_j).
\end{aligned}
$$

$\omega(\mathbf{p}, \mathbf{q})$ is thus completely determined by the n^2 vectors $\omega(\mathbf{e}_i, \mathbf{e}_j)$, and this definition can be naturally extended over the whole of $\Re^n$. It is an example of a *bilinear* operator. That is, if $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \Re^n$, and λ, μ are scalars, then

$$\omega(\lambda\mathbf{x} + \mu\mathbf{y}, \mathbf{z}) = \lambda\omega(\mathbf{x}, \mathbf{z}) + \mu\omega(\mathbf{y}, \mathbf{z}),$$

and

$$\omega(\mathbf{z}, \lambda\mathbf{x} + \mu\mathbf{y}) = \lambda\omega(\mathbf{z}, \mathbf{x}) + \mu\omega(\mathbf{z}, \mathbf{y}).$$

An operator must also be bounded to be bilinear, but this condition always holds in finite dimensions.

The set of bilinear operators can itself be made into a normed linear space (Kantorovich and Akilov, 1982), but we will be concerned with a different, though related, space. This is because the crossover operator draws both parents from the same distribution, not two different ones. Thus, given a bilinear operator ω, we can define a related operator $C : \Re^n \to \Re^n$ by

$$C(\mathbf{x}) = \omega(\mathbf{x}, \mathbf{x}) = \sum_{i,j} x_i x_j \omega(\mathbf{e}_i, \mathbf{e}_j).$$

We will call such an operator *quadratic* and denote the set of quadratic operators on a linear space X by $\mathcal{Q}(X)$.

THEOREM 3: *Let $A, B \in \mathcal{Q}(X)$, and let λ, μ be scalars. Then, defining*

$$(\lambda A + \mu B)(\mathbf{x}) = \lambda(A(\mathbf{x})) + \mu(B(\mathbf{x}))$$

for all $\mathbf{x} \in X$ makes $\mathcal{Q}(X)$ a vector space over the field of scalars.

Notice that a quadratic operator is non-linear, since for example, $\omega(\lambda\mathbf{x}) = \lambda^2\omega(\mathbf{x})$ (a fact pointed out for crossover in Wright and Bidwell (1996)). A norm can be defined on $\mathcal{Q}(X)$ as follows:

$$\|A\|_Q = \sup_{\|\mathbf{x}\|=1} \|A(\mathbf{x})\| \tag{3}$$

for any $A \in \mathcal{Q}(X)$, the subscript Q referring to the fact that we have a space of *quadratic* operators. It is straightforward to show that

$$\|A(\mathbf{x})\| \leq \|A\|_Q \|\mathbf{x}\|^2. \tag{4}$$

THEOREM 4: *Let C be the crossover operator defined above. Then $\|C\|_Q = 1$.*

PROOF: Let $\|\mathbf{x}\| = 1$. Then,

$$\begin{aligned}
\|C(\mathbf{x})\| &= \|\sum_{i,j} x_i x_j \omega(\mathbf{e}_i, \mathbf{e}_j)\| \\
&= \sum_{i,j} |x_i x_j| \|\omega(\mathbf{e}_i, \mathbf{e}_j)\|.
\end{aligned}$$

Now, since $\omega(\mathbf{e}_i, \mathbf{e}_j) \in \Lambda$, we have $\|\omega(\mathbf{e}_i, \mathbf{e}_j)\| = 1$, and so

$$\|C(\mathbf{x})\| = \sum_{i,j} |x_i x_j|$$

$$\begin{aligned} &= \sum_i |x_i| \sum_j |x_j| \\ &= \|\mathbf{x}\|^2 \\ &= 1 \end{aligned}$$

$\square$

It is also important to note that the composition of a quadratic operator with a linear operator is also quadratic.

THEOREM 5: *If $Q \in \mathcal{Q}(X)$, and $A \in \mathcal{B}(X)$, then*

$$A \circ Q \in \mathcal{Q}(X),$$

and

$$Q \circ A \in \mathcal{Q}(X).$$

PROOF: Since $Q \in \mathcal{Q}(X)$, there must be a bilinear operator B_Q such that $Q(\mathbf{x}) = B_Q(\mathbf{x}, \mathbf{x})$ for all $\mathbf{x} \in X$. Let $V : X \times X \to X$ be defined by

$$V(\mathbf{x}, \mathbf{y}) = B_Q(A\mathbf{x}, A\mathbf{y})$$

for any $\mathbf{x}, \mathbf{y} \in X$. Then, if $\mathbf{z} \in X$, and λ, μ are scalars, then

$$\begin{aligned} V(\lambda\mathbf{x} + \mu\mathbf{y}, \mathbf{z}) &= B_Q(A(\lambda\mathbf{x} + \mu\mathbf{y}), A\mathbf{z}) \\ &= B_Q(\lambda A\mathbf{x} + \mu A\mathbf{y}, A\mathbf{z}) \\ &= \lambda B_Q(A\mathbf{x}, A\mathbf{z}) + \mu B_Q(A\mathbf{y}, A\mathbf{z}) \\ &= \lambda V(\mathbf{x}, \mathbf{z}) + \mu V(\mathbf{y}, \mathbf{z}). \end{aligned}$$

A similar argument shows $V(\mathbf{x}, \lambda\mathbf{y} + \mu\mathbf{z}) = \lambda V(\mathbf{x}, \mathbf{y}) + \mu V(\mathbf{x}, \mathbf{z})$, and so V is a bilinear operator. Clearly, $(Q \circ A)(\mathbf{x}) = V(\mathbf{x}, \mathbf{x})$ for all $\mathbf{x} \in X$ and so $Q \circ A$ is quadratic.

Now let $W : X \times X \to X$ be defined by

$$W(\mathbf{x}, \mathbf{y}) = A(B_Q(\mathbf{x}, \mathbf{y})),$$

then

$$\begin{aligned} W(\lambda\mathbf{x} + \mu\mathbf{y}, \mathbf{z}) &= A(B_Q(\lambda\mathbf{x} + \mu\mathbf{y}, \mathbf{z})) \\ &= A(\lambda B_Q(\mathbf{x}, \mathbf{z}) + \mu B_Q(\mathbf{y}, \mathbf{z})) \\ &= \lambda A(B_Q(\mathbf{x}, \mathbf{z})) + \mu A(B_Q(\mathbf{y}, \mathbf{z})) \\ &= \lambda W(\mathbf{x}, \mathbf{z}) + \mu W(\mathbf{y}, \mathbf{z}). \end{aligned}$$

Similarly, it can be shown that $W(\mathbf{x}, \lambda\mathbf{y} + \mu\mathbf{z}) = \lambda W(\mathbf{x}, \mathbf{y}) + \mu W(\mathbf{x}, \mathbf{z})$, and so W is bilinear. Clearly, $(A \circ Q)(\mathbf{x}) = W(\mathbf{x}, \mathbf{x})$ for all $\mathbf{x} \in X$, and so $A \circ Q$ is quadratic. $\square$

5 A Normed Space for Genetic Operators

The pure crossover operator C, defined in the last section, represents the effect of applying crossover to a population vector. In practice, there is usually a probability called

the *crossover rate* with which crossover is to be carried out. If the crossover rate is χ, then the effect of this mixed crossover operator is

$$\mathcal{C}(\mathbf{x}) = \chi C(\mathbf{x}) + (1 - \chi)\mathbf{x}.$$

This suggests that $\mathcal{C}$ is a linear combination of the quadratic operator C and the linear identity operator I. However, we can simplify things by making the following observations.[1]

THEOREM 6: *Define a quadratic operator $\mathcal{I} \in \mathcal{Q}(\Re^n)$ by setting*

$$\omega(\mathbf{e}_i, \mathbf{e}_j) = \frac{\mathbf{e}_i + \mathbf{e}_j}{2}$$

so that

$$\mathcal{I}(\mathbf{x}) = \sum_{i,j} x_i x_j \frac{\mathbf{e}_i + \mathbf{e}_j}{2}$$

for any $\mathbf{x} \in \Re^n$. Then for all $\mathbf{p} \in \Lambda$,

$$\mathcal{I}(\mathbf{p}) = \mathbf{p}.$$

PROOF: For any $\mathbf{x} \in \Re^n$,

$$
\begin{aligned}
\mathcal{I}(\mathbf{x}) &= \sum_{i,j} x_i x_j \frac{\mathbf{e}_i + \mathbf{e}_j}{2} \\
&= \frac{1}{2}\sum_{i,j} x_i x_j \mathbf{e}_i + \frac{1}{2}\sum_{i,j} x_i x_j \mathbf{e}_j \\
&= \frac{1}{2}\sum_{i} x_i \mathbf{e}_i \sum_{j} x_j + \frac{1}{2}\sum_{j} x_j \mathbf{e}_j \sum_{i} x_i \\
&= \frac{1}{2}\mathbf{x}\sum_{j} x_j + \frac{1}{2}\mathbf{x}\sum_{i} x_i \\
&= \mathbf{x}\sum_{i} x_i
\end{aligned}
$$

Therefore, for all $\mathbf{p} \in \Lambda$,

$$
\begin{aligned}
\mathcal{I}(\mathbf{p}) &= \mathbf{p}\sum_{i} p_i \\
&= \mathbf{p}
\end{aligned}
$$

Thus the quadratic operator $\mathcal{I}$ is the identity on the simplex. □

COROLLARY 2: *Let $A : \Re^n \to \Re^n$ be a linear operator. Then $A \circ \mathcal{I} \in \mathcal{Q}(\Re^n)$, and*

$$(A \circ \mathcal{I})(\mathbf{p}) = A\mathbf{p}$$

for all $\mathbf{p} \in \Lambda$. Moreover,

$$\|A \circ \mathcal{I}\|_Q \leq \|A\|_L$$

---

[1]Thanks to Michael Vose for these observations.

using the appropriate norms on each side of the equation.

PROOF: That $A \circ \mathcal{I} \in \mathcal{Q}(\Re^n)$ follows from Theorem 5.

$$
\begin{aligned}
\|A \circ \mathcal{I}\|_Q &= \sup_{\|\mathbf{x}\|=1} \|A \circ \mathcal{I}(\mathbf{x})\| \\[2mm]
&= \sup_{\|\mathbf{x}\|=1} \left\| A \left(\mathbf{x} \sum_i x_i \right) \right\| \\[2mm]
&= \sup_{\|\mathbf{x}\|=1} \left| \sum_i x_i \right| \|A\mathbf{x}\| \\[2mm]
&\leq \sup_{\|\mathbf{x}\|=1} \left(\sum_i |x_i| \right) \|A\mathbf{x}\| \\[2mm]
&= \sup_{\|\mathbf{x}\|=1} \|\mathbf{x}\| \|A\mathbf{x}\| \\[2mm]
&= \sup_{\|\mathbf{x}\|=1} \|A\mathbf{x}\| \\[2mm]
&= \|A\|_L
\end{aligned}
$$

$\square$

It can be seen then that any linear operator A is equivalent to a quadratic operator $A \circ \mathcal{I}$ for all points in the simplex. We will therefore identify linear operators with their quadratic equivalents. For example, the identity matrix I is identified with $\mathcal{I}$. Similarly, we will identify the mutation matrix U with the quadratic operator $U \circ \mathcal{I}$. When there is no ambiguity, we will refer to both of these operators as being U. Under this identification, we see that mutation $U \in \mathcal{Q}(\Re^n)$, and the mixed crossover operator $\mathcal{C}$ can be written

$$
\mathcal{C} = \chi C + (1 - \chi)I \in \mathcal{Q}(\Re^n).
$$

Notice, moreover, that if we have any linear operators A and B, then

$$
(A + B) \circ \mathcal{I} = A \circ \mathcal{I} + B \circ \mathcal{I}.
$$

Hereafter, we assume that all operators are in the normed space $\mathcal{Q}(\Re^n)$ unless stated otherwise. Consequently, we write the norm $\|\cdot\|_Q$ for this space as simply $\|\cdot\|$. We retain the notation $\|\cdot\|_L$ to refer to the norm on the space of linear operators where necessary.

We pointed out earlier that the order of composition of $\mathcal{C}$ and $\mathcal{U}$ does not matter for typical implementations of crossover and mutation. We now give a condition for this to be true within our framework.

THEOREM 7: *Let $\mathcal{U}$ be a linear operator and ω be a bilinear operator on $\Re^n$. Let $C : \Re^n \to \Re^n$ be defined by*

$$
C(\mathbf{x}) = \omega(\mathbf{x}, \mathbf{x}).
$$

Given a scalar χ, define

$$
\mathcal{C}(\mathbf{x}) = \chi C(\mathbf{x}) + (1 - \chi)\mathbf{x}.
$$

If for all $i, j = 1, \dots, n$,

$$
\mathcal{U}\omega(\mathbf{e}_i, \mathbf{e}_j) = \omega(\mathcal{U}\mathbf{e}_i, \mathcal{U}\mathbf{e}_j),
$$

then

$$C \circ \mathcal{U} = \mathcal{U} \circ C.$$

PROOF: Let the linear operator $\mathcal{U}$ be given by a matrix U. We will show that under the conditions stated, $U \circ C = C \circ U$, from which the result follows straightforwardly.

$$C(\mathbf{x}) = \omega(\mathbf{x}, \mathbf{x}) = \sum_{i,j} x_i x_j \omega(\mathbf{e}_i, \mathbf{e}_j)$$

Therefore,

$$
\begin{aligned}
C \circ U(\mathbf{x}) &= \omega(U\mathbf{x}, U\mathbf{x}) \\
&= \omega(U(\sum_i x_i \mathbf{e}_i), U(\sum_j x_j \mathbf{e}_j)) \\
&= \omega(\sum_i x_i U \mathbf{e}_i, \sum_j x_j U \mathbf{e}_j) \\
&= \sum_i x_i \omega(U \mathbf{e}_i, \sum_j x_j U \mathbf{e}_j) \\
&= \sum_i x_i \sum_j x_j \omega(U \mathbf{e}_i, U \mathbf{e}_j) \\
&= \sum_{i,j} x_i x_j \omega(U \mathbf{e}_i, U \mathbf{e}_j) \\
&= \sum_{i,j} x_i x_j U \omega(\mathbf{e}_i, \mathbf{e}_j) \\
&= U \sum_{i,j} x_i x_j \omega(\mathbf{e}_i, \mathbf{e}_j) \\
&= U \circ C(\mathbf{x})
\end{aligned}
$$

$\square$

The condition in the theorem is that, given two parent individuals, the probability of producing a given child individual by first applying crossover and then mutation is the same as by first applying mutation (to the parents) and then crossover. This holds for a wide range of crossover and mutation definitions, including all those in standard practice operating on binary strings (Vose and Wright, 1995).

The next two sections provide some applications of our new, normed space to developing some bounds that help describe the behavior of the SGA. For a search space consisting of fixed-length binary strings, these bounds will depend on the size of the search space (via the string length ℓ) and will therefore provide some insight into the scalability of certain choices of operator.

6   Fitness Variance of Fixed-point Populations

Our first application of the new normed space $\mathcal{Q}(\Re^n)$ is to find a bound on the variance of the fitness in a fixed-point population. A population $\mathbf{p} \in \Lambda$ is a fixed-point if

$$\mathcal{G}(\mathbf{p}) = \mathbf{p}.$$

Given the representations of selection, mutation, and crossover defined above, the generational operator $\mathcal{G}$ is

$$\mathcal{G}(\mathbf{p}) = \frac{(\mathcal{M} \circ S)(\mathbf{p})}{h(S\mathbf{p})^2},$$

where $\mathcal{M} = \mathcal{C} \circ \mathcal{U} = \mathcal{U} \circ \mathcal{C}$. So at a fixed-point,

$$(\mathcal{M} \circ S)(\mathbf{p}) = h(S\mathbf{p})^2 \mathbf{p}.$$

Recall that the mean fitness of a population $\mathbf{p}$ is given by

$$\mathbf{Mean}[\mathbf{p}] = h(S\mathbf{p}).$$

Similarly, the variance of the fitness of a population is

$$\mathbf{Var}[\mathbf{p}] = h(S^2\mathbf{p}) - h(S\mathbf{p})^2.$$

When $\mathbf{p}$ is a fixed-point, we have the following result.

THEOREM 8: *Let $\mathbf{p} \in \Lambda$ be a fixed-point of $\mathcal{G}$, then*

$$\mathbf{Var}[\mathbf{p}] \leq \max_k \{f(k)\} \mathbf{Mean}[\mathbf{p}] \|I - \mathcal{M}\|.$$

PROOF: Let $\lambda = \|S\mathbf{p}\| = h(S\mathbf{p})$ be the mean fitness of the fixed-point $\mathbf{p}$. Now define a linear operator

$$Fx = \frac{1}{\lambda} Sx.$$

Notice that $F\mathbf{p} = \mathcal{F}(\mathbf{p})$. Then,

$$h(F\mathbf{p}) = \frac{h(S\mathbf{p})}{\lambda} = 1,$$

and

$$h(F^2\mathbf{p}) = h(F(\frac{1}{\lambda}S\mathbf{p})) = h\left(\frac{S^2\mathbf{p}}{\lambda^2}\right) = \frac{h(S^2\mathbf{p})}{h(S\mathbf{p})^2}.$$

Now, since $\mathbf{p}$ is a fixed-point, $\mathbf{p} = \mathcal{G}(\mathbf{p}) = \mathcal{M} \circ \mathcal{F}(\mathbf{p}) = \mathcal{M} \circ F\mathbf{p}$. Therefore,

$$F\mathbf{p} = F \circ \mathcal{M} \circ F\mathbf{p},$$

which implies

$$h(F \circ \mathcal{M} \circ F\mathbf{p}) = h(F\mathbf{p}) = 1.$$

It follows that

$$
\begin{aligned}
\frac{\mathbf{Var}[\mathbf{p}]}{\lambda^2} &= h(F^2\mathbf{p}) - 1 \\
&= h(F^2\mathbf{p}) - h(F \circ \mathcal{M} \circ F\mathbf{p}) \\
&= h(F^2\mathbf{p} - F \circ \mathcal{M} \circ F\mathbf{p}) \\
&= h(F \circ (F\mathbf{p} - \mathcal{M} \circ F\mathbf{p})) \\
&= h(F \circ (I - \mathcal{M}) \circ F\mathbf{p}) \\
&\leq \|h\| \|F\|_L \|(I - \mathcal{M}) \circ F\mathbf{p}\|
\end{aligned}
$$

using the appropriate norms for each operator. Now we use Equation 4 to give:

$$\frac{\mathbf{Var}[\mathbf{p}]}{\lambda^2} \leq \|h\|\|F\|_L\|I - \mathcal{M}\|\|F\mathbf{p}\|^2.$$

Now $\|h\| = 1$, $\|F\|_L = \|S\|_L/\lambda$, and $\|F\mathbf{p}\| = 1$, so

$$\mathbf{Var}[\mathbf{p}] \leq \lambda\|S\|_L\|I - \mathcal{M}\|,$$

and substituting $\lambda = \mathbf{Mean}[\mathbf{p}]$ and $\|S\|_L = \max_k\{f(k)\}$ gives

$$\mathbf{Var}[\mathbf{p}] \leq \max_k\{f(k)\}\mathbf{Mean}[\mathbf{p}]\|I - \mathcal{M}\|$$

as required. □

It should be noted that this theorem sums up, within a completely general framework, the well-known convergence result that as the mutation rate and crossover rate tend to zero, so the fixed-point tends to a population that is uniform with respect to fitness values (Cerf, 1998; Rudolph, 1997).

7 Stability of the Optimum Population

The second application considers what happens to a population comprised entirely of copies of the optimum. We are interested in how far this population moves into the simplex upon application of the generational operator $\mathcal{G}$. For any population $\mathbf{p}$, the distance $\|\mathbf{p} - \mathcal{G}(\mathbf{p})\|$ is called the *force* on $\mathbf{p}$. When the force is small, it is possible for finite populations to stall there, forming a metastable state. This metaphor of "force" was introduced by Michael Vose at ICGA'91. The principle of metastability is described in Vose and Liepins (1991) and Vose (1999b). For examples of its use in explaining GA dynamics for certain fitness functions, see van Nimwegen et al. (1997, 1999) and Rowe (1998).

For the population consisting entirely of the optimum, we have:

THEOREM 9: *Let $\mathbf{v} \in \Lambda$ be the population comprised entirely of copies of the optimum. Then*

$$\|\mathbf{v} - \mathcal{G}(\mathbf{v})\| \leq \|I - \mathcal{M}\|.$$

PROOF: It is known that the population $\mathbf{v}$ is a fixed point of the SGA when there is no crossover or mutation. That is:

$$\frac{S\mathbf{v}}{h(S\mathbf{v})} = \mathbf{v}.$$

Therefore,

$$\|\mathbf{v} - \mathcal{G}(\mathbf{v})\| = \|\frac{S\mathbf{v}}{h(S\mathbf{v})} - \frac{(\mathcal{M} \circ S)(\mathbf{v})}{h(S\mathbf{v})^2}\|.$$

The force on $\mathbf{v}$ is invariant with respect to scaling the fitness function. Therefore we will assume (without loss of generality) that fitness is scaled such that the maximum fitness value is 1. This means that $\|S\mathbf{v}\| = h(S\mathbf{v}) = 1$, and so

$$
\begin{aligned}
\|\mathbf{v} - \mathcal{G}(\mathbf{v})\| &= \|S\mathbf{v} - (\mathcal{M} \circ S)(\mathbf{v})\| \\
&= \|(I - \mathcal{M}) \circ S\mathbf{v}\| \\
&\leq \|I - \mathcal{M}\|\|S\mathbf{v}\|^2 \\
&= \|I - \mathcal{M}\|.
\end{aligned}
$$

□

8 Scalability of Operators

The factor $\|I - \mathcal{M}\|$ appearing in the bounds derived above is the distance of the combined crossover and mutation operator $\mathcal{M}$ from the identity operator. That is, it is a measure of how much crossover and mutation disturb the state of the population. Theorem 8 may be interpreted as saying that the smaller the disturbance due to crossover and mutation, the less the fitness variance of the fixed-point population. Similarly, Theorem 9 says that the less the disturbance, the more likely a finite population will form a metastable state near to the optimum. This means that, should the GA create a population comprising mostly copies of the optimum, then it is likely to settle at that population for some time. It should be understood that this in no way guarantees that such a population will actually be found in any run.

It is important, therefore, to study the quantity $\|I - \mathcal{M}\|$, which we will call the *disturbance* of the generational operator $\mathcal{G}$, and denote $\mathbf{D}(\mathcal{G})$. We will see shortly that when the search space comprises binary strings in the usual way, the disturbance of $\mathcal{G}$ depends on the size of the search space. We will use this connection to explore some issues in the scalability of the genetic operators. First, we derive an upper bound for the disturbance.

THEOREM 10: *If* $\mathcal{M} = \mathcal{C} \circ \mathcal{U}$, *and* $\mathcal{C} = \chi C + (1 - \chi)I$ *with* $0 \leq \chi \leq 1$, *then*

$$\mathbf{D}(\mathcal{G}) \leq \|I - (1 - \chi)U\| + \chi.$$

PROOF:

$$
\begin{aligned}
\mathbf{D}(\mathcal{G}) &= \|I - \mathcal{M}\| \\
&= \|I - \mathcal{C} \circ U\| \\
&= \|I - \chi(C \circ U) - (1 - \chi)U\| \\
&= \|(I - (1 - \chi)U) - \chi(C \circ U)\| \\
&\leq \|(I - (1 - \chi)U)\| + \| - \chi(C \circ U)\| \\
&= \|(I - (1 - \chi)U)\| + \chi\|C \circ U\|
\end{aligned}
$$

From Equation 3,

$$\|C \circ U\| = \sup_{\|\mathbf{x}\|=1} \|C \circ U(\mathbf{x})\|,$$

and we know that U is norm-preserving, so $\|\mathbf{x}\| = 1 \Rightarrow \|U(\mathbf{x})\| = 1$. Thus

$$\|C \circ U\| = \|C\| = 1$$

by Theorem 4. Combining these results gives

$$
\begin{aligned}
\mathbf{D}(\mathcal{G}) &\leq \|(I - (1 - \chi)U)\| + \chi\|C \circ U\| \\
&\leq \|(I - (1 - \chi)U)\| + \chi.
\end{aligned}
$$

□

COROLLARY 3: *For a crossover only algorithm (i.e., zero mutation),*

$$\mathbf{D}(\mathcal{G}) \leq 2\chi.$$

The corollary results from substituting $U = I$ in Theorem 10. We also have:

COROLLARY 4: *For a mutation only algorithm (i.e., zero crossover),*

$$\mathbf{D}(\mathcal{G}) = \|I - U\| \leq \|I - U\|_L.$$

Corollary 3 is interesting with regard to a conjecture of Vose and Wright (1994) that the only asymptotically stable fixed-points of $\mathcal{G}$, when there is no mutation, are vertices of the simplex, when the search space is the usual set of fixed-length binary strings. At a vertex, the fitness variance is zero. Experiments by Wright and Bidwell (1996) suggest the existence of unstable fixed-points within the simplex, also with zero fitness variance. Our corollary, however, leaves open the possibility of fixed-points (not necessarily stable) with non-zero fitness variance, at least in this general setting.

Let us now turn to the usual search space for genetic algorithms: the set of binary strings of length ℓ. Mutation is implemented bitwise, with an individual bit changing with probability μ, referred to as the *mutation rate*.

THEOREM 11: *Suppose the search space comprises bit strings of length ℓ. That is, we represent each element of $\Omega = \{0, 1, \ldots, 2^\ell - 1\}$ by its binary representation. Mutation is implemented bitwise with mutation rate μ. Crossover is given by a suitable quadratic operator and applied with rate χ. The disturbance of $\mathcal{G}$ is then*

$$\mathbf{D}(\mathcal{G}) \leq 2 - 2(1 - \chi)(1 - \mu)^\ell.$$

PROOF: From Theorem 10, we have

$$\|I - \mathcal{M}\| \leq \|I - (1 - \chi)U\| + \chi \leq \|I - (1 - \chi)U\|_L + \chi.$$

The (linear) norm of the matrix $I - (1 - \chi)U$ can be found from Equation 2, noting that the diagonal entries of U are equal to $(1 - \mu)^\ell$, and that $\|U\| = 1$. The result follows directly. □

Theorem 11 gives us a link between the disturbance of $\mathcal{G}$ and the size of the search space, via the string length ℓ. We have the following results for some potential choices of mutation rate.

COROLLARY 5: *If the mutation rate is fixed as ℓ varies, then*

$$\lim_{\ell \to \infty} \mathbf{D}(\mathcal{G}) \leq 2.$$

If the mutation rate is $\mu = 1/\ell$, then

$$\lim_{\ell \to \infty} \mathbf{D}(\mathcal{G}) \leq 2 - \frac{2}{e}(1 - \chi).$$

If the mutation rate is $\mu = \kappa/\ell$ for some fixed $0 < \kappa < \ell$, then

$$\lim_{\ell \to \infty} \mathbf{D}(\mathcal{G}) \leq 2 - 2e^{-\kappa}(1 - \chi).$$

If the mutation rate is $\mu = 1/\ell^2$, then

$$\lim_{\ell \to \infty} \mathbf{D}(\mathcal{G}) \leq 2\chi.$$

This corollary indicates that if a converged (uniform) population is to be guaranteed in the limit as the string length increases (assuming that the fitness function stays bounded), then the mutation rate must decrease faster than $1/\ell$. If the Vose-Wright conjecture is true, then the setting of a fixed non-zero crossover rate χ may well not interfere with the convergence to a uniform population. However, in order to *guarantee* this, one can ensure that $\chi \to 0$ as $\ell \to \infty$.

It should be noted that the change in search space size does not influence the distance between population vectors in the simplex so as to counterbalance these convergences (see Theorem 3.1 of Vose (1999b)).

9   Illustrative Experiments

First, we give an example that illustrates that the fixed-point population need not be uniform in the limit as $\ell \to \infty$, when mutation is set at $1/\ell$. The fitness function is the *multiplicative* function (Prügel-Bennett, 1999; Woodcock and Higgs, 1996). Each 1 in the bitstring counts 1 point, and each 0 counts s, where $0 < s < 1$. The fitness of the whole string is given by multiplying these scores together. The optimum is thus the string containing all ones, which has a fitness of 1. Using the results presented in Woodcock and Higgs (1996), we can calculate[2] that, at the infinite population fixed point $\mathbf{p}$:

$$\mathbf{Mean}[\mathbf{p}] = m^{\ell},$$

and

$$\mathbf{Var}[\mathbf{p}] = (s^2 a(1-a) + m^2)^{\ell} - m^{2\ell},$$

where

$$m = 1 - as,$$

and

$$a = \frac{1}{2}\left(1 - \frac{1}{\ell} + \frac{2}{s\ell} - \sqrt{(1 - \frac{1}{\ell} + \frac{2}{s\ell})^2 - \frac{4}{s\ell}}\right).$$

Setting $s = 1/2$ and letting $\ell \to \infty$ gives[3]

$$\lim_{\ell \to \infty} \mathbf{Mean}[\mathbf{p}] = \frac{1}{e},$$

and

$$\lim_{\ell \to \infty} \mathbf{Var}[\mathbf{p}] = \frac{1}{e^{3/2}} - \frac{1}{e^2}.$$

That is, the variance tends to approximately 0.088. Some experiments were run to check this, tracking a population of 500 over 100 generations. The fitness variances of the final populations (at generation 100) for a number of string lengths is plotted in Figure 1 along with the theoretical curve for the infinite population case.

Second, to illustrate the decrease in fitness variance with increasing string length when $\mu = 1/\ell^2$, two simple experiments were conducted. Both experiments used the

---

[2]Thanks to Nick Longford of De Montfort University for help with this.
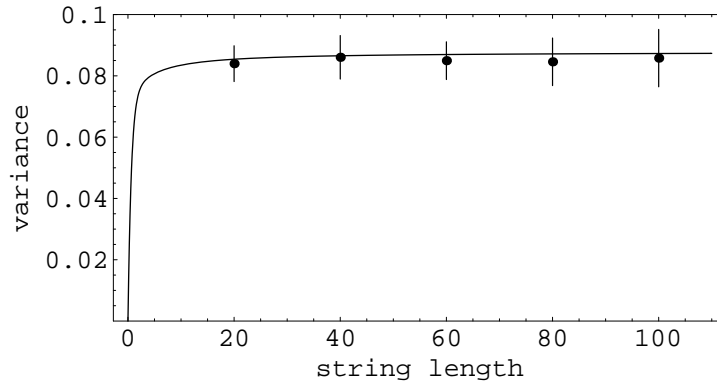[3]Calculations performed using Mathematica 3.0 ©Wolfram Research.

Figure 1: Fitness variance of population at generation 100, for different string lengths, with multiplicative fitness function. Mutation only, $\mu = 1/\ell$. Each point is the average of 30 runs. The error bars represent one standard deviation. The solid line shows the theoretical value for the infinite population limit.
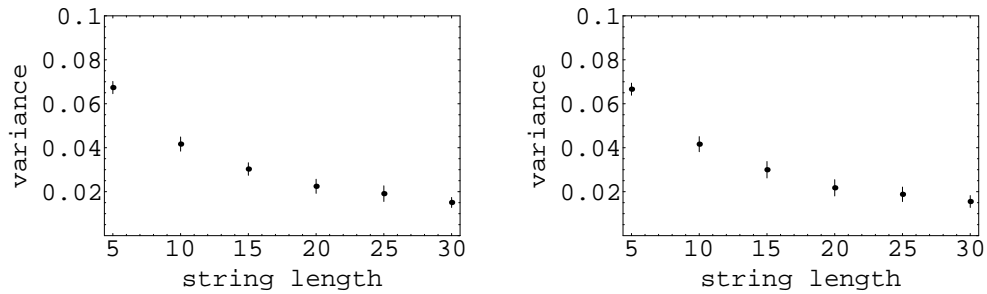

Figure 2: Fitness variance of population at generation 100, for different string lengths, with multiplicative fitness function. Left: mutation only, $\mu = 1/\ell^2$. Right: mutation and crossover, $\mu = 1/\ell^2, \chi = 1$. Each point is the average of 30 runs. The error bars represent one standard deviation.

same multiplicative function as before, with a population size of 500. In the first experiment, mutation was set to $1/\ell^2$, and there was no crossover. For six different string lengths, the GA was run for 100 generations and the fitness variance of the final population calculated. This was repeated 30 times for each string length. The results are shown in Figure 2, and it is clear that the fitness variance decreases with increasing string length. The error bars on the graphs are one standard deviation.

The second experiment was almost the same as the first, except uniform crossover was applied with $\chi = 1$. As can be seen, the application of crossover makes very little difference to the fitness variance of the final population. Of course, this is only an example, and the results might well be different for other fitness functions. However, it seems a reasonable generalization of the Vose-Wright conjecture to suggest that the fitness variance of an asymptotically stable fixed-point population is independent of the probability of the crossover rate χ. Note that the analysis followed in this paper has applied to any fixed-point, not just asymptotically stable ones.

10 Conclusions

An abstract normed vector space has been described of which the genetic operators are elements. Using this framework, a bound has been given for the fitness variance of a fixed-point population $\mathbf{p}$ (Theorem 8), namely:

$$\mathbf{Var}[\mathbf{p}] \leq \max_{k}\{f(k)\}\mathbf{Mean}[\mathbf{p}]\|I - \mathcal{M}\|.$$

Similarly, a bound has been derived for the force on the population, $\mathbf{v}$, comprising copies of the optimal solution (Theorem 9):

$$\|\mathbf{v} - \mathcal{G}(\mathbf{v})\| \leq \|I - \mathcal{M}\|.$$

The quantity $\|I - \mathcal{M}\|$ is called the disturbance of the generational operator $\mathcal{G}$, since it represents the extent to which crossover and mutation disturb a population. Analyzing this quantity in the case of fixed-length binary strings leads to a connection with the size of the search space (via the string length ℓ). In particular, assuming the fitness function remains bounded as ℓ increases, both the fitness variance of fixed-point populations and the force on the optimal population can be guaranteed to tend to zero as $\ell \to \infty$, if $\chi \to 0$ as $\ell \to \infty$, and $\mu \to 0$ faster than $1/\ell$. A counterexample has been given that shows the rate of $1/\ell$ is not sufficient. It is conjectured that the crossover rate χ actually has no effect on these quantities, though this has not been proven.

References

Cerf, R. (1998). Asymptotic convergence of genetic algorithms. *Advances in Applied Probability Journal (General Applied Probability)*, 30:521–550.

Eigen, M., McCaskill, J., and Schuster, P. (1989). The molecular quasi-species. In Prigogine, I. and Rice, S. A., editors, *Advances in Chemical Physics LXXV*, pages 149–263, John Wiley and Sons, New York, New York.

Kantorovich, L. V. and Akilov, G. P. (1982). *Functional Analysis.* Pergamon Press, Oxford, England.

Maddox, I. J. (1988). *Elements of Functional Analysis, 2nd Edition.* Cambridge University Press, Cambridge, England.

Prügel-Bennett, A. (1999). On the limit of long strings. In Banzhaf, W. and Reeves, C., editors, *Foundations of Genetic Algorithms 5*, pages 45–56, Morgan Kaufmann, San Francisco, California.

Rowe, J. E. (1998). Population fixed-points for functions of unitation. In Banzhaf, W. and Reeves, C., editors, *Foundations of Genetic Algorithms 5*. Morgan Kaufmann, San Francisco, California.

Rudolph, G. (1997). *Convergence Properties of Evolutionary Algorithms.* Kovacs, Hamburg, Germany.

van Nimwegen, E., Crutchfield, J. P., and Mitchell, M. (1997). Finite populations induce metastability in evolutionary search. *Physics Letters A*, 229(2):144–150.

van Nimwegen, E., Crutchfield, J. P., and Mitchell, M. (1999). Statistical dynamics of the Royal Road genetic algorithm. *Theoretical Computer Science*, 229(1–2):41–102.

Vose, M. D. (1999a). Random heuristic search. *Theoretical Computer Science*, 229(1–2):103–142.

Vose, M. D. (1999b). *The Simple Genetic Algorithm: Foundations and Theory.* MIT Press, Cambridge, Massachusetts.

Vose, M. D. and Liepins, G. E. (1991). Punctuated equilibria in genetic search. *Complex Systems*, 5:31–44.

Vose, M. D. and Wright, A. H. (1994). Stability of vertex fixed points and applications. In Whitley, L. D. and Vose, M. D., editors, *Foundations of Genetic Algorithms 3*, pages 103–113, Morgan Kaufmann, San Francisco, California.

Vose, M. D. and Wright, A. H. (1995). Simple genetic algorithms with linear fitness. *Evolutionary Computation*, 2(4):347–368.

Woodcock, G. and Higgs, P. G. (1996). Population evolution on a multiplicative single-peak fitness landscape. *Journal of Theoretical Biology*, 179:61–73.

Wright, A. H. and Bidwell, G. (1996). A search for counterexamples to two conjectures on the simple genetic algorithm. In Belew, R. K. and Vose, M. D., editors, *Foundations of Genetic Algorithms 4*, pages 73–81, Morgan Kaufmann, San Francisco, California.