General Cardinality Genetic Algorithms

Gary J. Koehler
Decision and Information Sciences
Warrington College of Business
Administration, 351 BUS
University of Florida
Gainesville, FL 32611
koehler@ufl.edu

Siddhartha Bhattacharyya
Information and Decision Sciences
College of Business Administration
University of Illinois at Chicago
601 South Morgan Street (M/C 294)
Chicago, IL 60607
sidb@uic.edu

Michael D. Vose
Computer Science Department
The University of Tennessee
107 Ayres Hall
Knoxville, TN 37996-1301
vose@cs.utk.edu

Abstract
A complete generalization of the Vose genetic algorithm model from the binary to higher cardinality case is provided. Boolean AND and EXCLUSIVE-OR operators are replaced by multiplication and addition over rings of integers. Walsh matrices are generalized with finite Fourier transforms for higher cardinality usage. Comparison of results to the binary case are provided.
Keywords
Genetic algorithms, general cardinality, Markov chain, Fourier transform.

1. Introduction

In 1990, Vose presented an infinite-sized population model for simple binary genetic algorithms (GAs) that gives their exact expected behavior over time (Vose, 1990). Similar models were later provided by T. Davis (1991) and Vose and Liepins (1991). Nix and Vose (1992) extended this model to the finite-population case. Vose (1993, 1997, in press) provided important generalizations and obtained significant further results. Working with colleagues, Vose also considered several special cases and related issues (Juliany & Vose, 1994; Vose & Wright, 1994, 1995). Others also contributed to knowledge about these GA models, including Koehler (1994), Aytug and Koehler (1996), and Suzuki (1993). However, all these papers restrict GA strings to those formed by binary components.

Practitioners have had mixed results with different encodings. Some (see L. Davis, 1991) have often noted better GA performance when higher cardinality alphabets were used; others have reported better performance with binary encodings (see Shaffer, 1984). Antonisse (1989) gives a theoretical argument that suggests that higher cardinality alphabets may be preferred to binary strings. Similarly, Reeves and Wright (1995) argue from an experimental design approach that binary encodings would not make an epistatic higher cardinality encoding any less so. Finally, higher cardinality encodings often provide a more natural correspondence with the real world.

Determining for a given GA whether a higher cardinality representation is better than a binary encoding can, in theory, be answered by using GA models that capture the exact expected behavior of both. De Jong (1995) and Spears and De Jong (1997) have shown how the exact model for the binary GA can give important insights to transient GA behavior. These same approaches could be used to study the impact of different cardinalities if a general cardinality model were available.

Some attempts have been made to generalize the Vose models to higher cardinality encodings. Bhattacharyya and Koehler (1994) considered GA strings composed of components drawn from 2^v cardinality alphabets and extended the Vose model to cover this case. Their method of analysis follows the binary case fairly closely, as might be expected. To date, no complete generalization of the Vose model to higher cardinalities has been accomplished. This paper presents such a generalization.

In this paper we provide a complete generalization of the simple GA Vose models to strings having components drawn from alphabets of cardinality c (where c is a whole number greater than 1). The tools of analysis depart from those used in the binary and 2^v cases. At first glance, the extension from binary to cardinality c seems obvious: Use strings having cardinality c components and replace Boolean AND and EXCLUSIVE-OR operators by multiplication and addition over rings of integers. However, some properties that are true for the binary representation fail for higher cardinalities, requiring a different approach (e.g., see Appendix, Lemma A1, part 1). Furthermore, analysis of these higher cardinality models requires generalization of some familiar tools. For example, Walsh matrices are generalized with Fourier transforms.

In Section 2 we provide our notation. In Section 3 a generalization of Walsh matrices is presented, and important properties are derived (Theorem 1). These properties are used throughout the remainder of the paper. The simple GA Vose model (infinite population size) is extended from binary to cardinality c in Section 4. As an aside, we note that this model can be trivially extended to the finite population model, as was done by Nix and Vose (1992) for the simple binary GA. The behavior of the infinite population case has a lot to do with finite population dynamics. Although this may appear esoteric, it is fundamental to a complete understanding of the dynamical behavior of these systems.

The general cardinality model presented in Section 4 is for the most general form of the simple GA—a form using mutation and crossover masks. This form allows for the various specialized operators used in practice. However, since most practitioners use simple cases of these masks, we show how the masks can be specialized to mutation rates and crossover rates (Remarks 1 and 2).

A true understanding of the behavior of the general cardinality model requires knowledge of its asymptotic properties. Fixed points of the GA mapping function are asymptotically stable if the spectral radius of the differential at the fixed point is less than 1. In Section 6 we derive an explicit formula for the spectrum of the twist of the mixing matrix (Theorem 4) that directly leads to the spectrum of the GA mapping at fixed points in the zero mutation case. It is interesting to note that this result is identical to the binary case, which makes the applications given in Vose and Wright (1995) directly applicable to the general cardinality case. We also show that, in general, the spectral radius of the differential of the mixing operator at the fixed point is bounded by 1 (Corollary 1). Section 5 provides the properties of the GA mixing matrix and its twist that are needed in Section 6.

In Section 7 we consider an important special case of GA search—the case where mutation and one-point crossover are defined by rates. This is important for practitioners because they often use rates rather than masks. It is also important because the spectral properties discovered in Section 6 can be evaluated for an important class of problems (Theorem 7). In particular, exact conditions are derived (Corollary 2) that relate stability of fixed points under mixing alone (i.e., uniform fitness) to mutation rates and alphabet cardinality by allowing one to determine whether the spectral radius of the differential at a fixed point of the GA mapping function is less than 1.

We conclude with a summary and some future directions and highlight important differences between the binary and general cardinality case.

2. Notation

This section introduces notation that will be used in the following sections. The notational framework described is a generalization from the binary to the multicardinality case. Although similar in many respects, the foundations presented are more than mere extensions of the binary case; they are, among several possibilities, the proper generalizations that enable the subsequent theory to unfold.

Let Z denote the set of integers, Z^+ the set of positive integers, and, for integer $c \geq 2$, let Z_c denote the set $\{0, \ldots, c-1\}$ of integers modulo c. For $a, b \in Z_c$, define the commutative operators $\oplus$ and $\otimes$ as

$$a \oplus b = (a + b) \bmod c \qquad \text{and } a \otimes b = (ab) \bmod c$$

Note that multiplication $\otimes$ takes precedence over $\oplus$ and that $\otimes$ distributes over $\oplus$. Denote by Z_c the set[1] of integers modulo c. The unique additive inverse of $x \in Z_c$ will be denoted by $-x$. To simplify notation, we use $y \ominus x$ for $y \oplus (-x)$. Define the unique complement k of $k \in Z_c$ by $k \oplus \bar{k} = 1$. Note that for any $j, k \in Z_c$,

$$j = j \otimes k \oplus j \otimes \bar{k} \tag{L1}$$

Two useful properties of Z_c under $\oplus$ and $\otimes$ are summarized below (see Appendix).

LEMMA A1: *Let $j, k, p, q \in Z_c$. Then,*

1. $p \otimes k = j \otimes k$ *and* $q \otimes \bar{k} = j \otimes \bar{k} \Rightarrow p \otimes k \oplus q \otimes \bar{k} = j$.
2. $p \otimes k = j \otimes k$ *and* $q \otimes \bar{k} = j \otimes \bar{k} \Leftarrow p \otimes k \oplus q \otimes \bar{k} = j$ *and* $k \in \Omega_2$.

A string s is a finite sequence of elements from Z_c and is displayed as

$$s_{\ell-1}s_{\ell-2}\cdots s_0$$

The s_i terms are called digits of s, and ℓ is its length. Let $\ell \in Z^+$ be fixed. The remainder of this paper suppresses dependence on ℓ, since the string length does not change. Let Ω be the set of all such possible strings when c is understood, or as Ω_c otherwise.

A string can be represented as the unique integer

$$s \equiv \sum_{i=0}^{\ell-1} s_i c^i$$

1 Under $\oplus$ and $\otimes$, Z_c is a commutative ring and is a field if c is prime.

We choose to think of Ω equivalently and interchangeably as a set of strings, as the set of integer representations of these strings, or by vectors of length ℓ, where we interpret $s \in \Omega$ as

$$s = \begin{bmatrix} s_0 \\ s_1 \\ \vdots \\ s_{\ell-1} \end{bmatrix}$$

Operations $\oplus$ and $\otimes$ are extended to these vectors by applying them coordinate-wise. Operations $\oplus$ and $\otimes$ are extended to strings by component-wise application.[2] Suppose $s \in \Omega_c$ and $p \in \Omega_2$, then $s \otimes p$ will be interpreted as an element of Ω_c, since $\Omega_2 \subseteq \Omega_c$.

Let $\mathbf{0}$ be a vector of zeros, $\mathbf{1}$ a vector of 1's and x^{T} the transpose of vector x; b_i is the ith column of the $\ell \times \ell$ identity matrix, where indexing of rows and columns begins with zero.

EXAMPLE: For $c = 2$ (binary alphabet) and $\ell = 3$, we have

$\Omega = \{000, 001, 010, 011, 100, 101, 110, 111\}$ in explicit string representation,

$\Omega = \{0, 1, 2, 3, 4, 5, 6, 7\}$ in integer representation, and

$$\Omega = \left\{ \begin{bmatrix}0\\0\\0\end{bmatrix}, \begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}1\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}, \begin{bmatrix}1\\0\\1\end{bmatrix}, \begin{bmatrix}0\\1\\1\end{bmatrix}, \begin{bmatrix}1\\1\\1\end{bmatrix} \right\} \text{ in vector notation.}$$

For $s \in \Omega$, let $n(s)$ be the number of nonzero digits of s. When $c = 2$, $n(s) = s^{\mathrm{T}}\mathbf{1}$. For s complex, $|s|$ will refer to the modulus of s. We define width(s) as zero if s is zero and as the difference between the position of the highest nonzero position of s and the lowest nonzero position of s otherwise. That is,

$$\text{width}(s) = \begin{cases} 0, & s = 0 \\ \max\{i : b_i^{\mathrm{T}} s > 0\} - \min\{i : b_i^{\mathrm{T}} s > 0\} & \text{otherwise} \end{cases}$$

Let $\mathcal{R}$ be the set of real numbers and define Λ_n as the simplex

$$\Lambda_n = \{x \in \mathcal{R}^n : x^{\mathrm{T}}\mathbf{1} = 1, \quad x \geq \mathbf{0}\}$$

where n is the cardinality of Ω (i.e., c^ℓ or 2^ℓ).

We use the indicator function

$$[p] = \begin{cases} 1 & \text{if } p \text{ is true or } p \text{ is nonzero} \\ 0 & \text{if } p \text{ is false or zero} \end{cases}$$

depending on the context of usage. Define the permutation matrix $(P_k)_{s,t} \equiv [k \oplus s = t]$. With

2 Ω is also a commutative ring.

premultiplication, P_k permutes rows by

$$\text{row } i \mapsto \text{ row } i \ominus k$$

whereas postmultiplication has the effect column $j \mapsto$ column $j \oplus k$. Note that $P_s P_t = P_{s \oplus t}$ and $P_k^{-1} = P_k^H$, where for any matrix A, A^H is its conjugate transpose. The conjugate of x is denoted as $\bar{x}$.

For any square matrix A, denote its spectrum by $\sigma(A)$ and its spectral radius as $\rho(A)$. For any differentiable map $\mathcal{H}$, let $d\mathcal{H}_x$ be the differential of $\mathcal{H}$ at x.

For each $j \in \Omega$, we have a positive fitness measure $f(j)$. Let F be the diagonal matrix having $F_{j,j} = f(j)$, where F is called the fitness matrix. Finally, let $f \equiv F\mathbf{1}$.

In the next section we introduce finite Fourier matrices (denoted by W). To simplify notation with their usage in Sections 5, 6, 7 and 8, define the transform $\hat{\mathbf{x}}$ of a vector x as Wx and the transform $A^\wedge$ of a matrix A as WAW^H.

3. The Finite Fourier Transform

Historically, Walsh matrices have played a useful role in the study of GAs. We also find them useful in simplifying many expressions, except that their definition requires generalization to make them suitable for a general cardinality (as opposed to binary) alphabet. What emerges as the appropriate extension is the finite Fourier transform. To streamline notation, let $e(x) = e^{2\pi x\sqrt{-1}}$. Define the finite Fourier matrix by

$$W_{r,s} = c^{-\ell/2} e(r^{\mathrm{T}} s/c), \qquad r, s \in \Omega$$

Below are several useful properties of these matrices, which are analogous to properties of Walsh matrices. In fact, choosing $c = 2$ specializes the Fourier matrix to a Walsh matrix.

THEOREM 1 (properties of Walsh matrices):

1. $W = W^{\mathrm{T}}$.
2. $W_{p,r\oplus s} = c^{\ell/2} W_{r,p} W_{p,s}$.
3. *For any* $r, s \in \Omega$ *and* $j \in \{0, 1, \ldots, \ell - 1\}$,

$$\sum_{y=0}^{c-1} W_{r,s+yb_j} = \begin{cases} cW_{r,s} & \text{if } r_j = 0 \\ 0 & \text{otherwise} \end{cases}$$

4. $W\mathbf{1} = c^{\ell/2} b_0$.
5. $W^H W = WW^H = I$.
6. $WP_k W^H = c^{\ell/2} \operatorname{diag}(Wb_{-k})$.

PROOF: Properties 1 to 5 are well-known properties of finite Fourier transforms and follow easily from the definitions and basic complex arithmetic. The exception is property 6, which shows that the change of basis corresponding to conjugation by W simultaneously diagonalizes the P_k matrices. The demonstration of that follows:

6. Substituting identities gives

$$\begin{aligned}(WP_kW^H)_{i,j} &= \sum_x \sum_y W_{i,x}(P_k)_{x,y} W_{-j,y} \\ &= \sum_x \sum_y W_{i,x} W_{-j,y}[x \oplus k = y] = \sum_x W_{i,x} W_{x \oplus k,-j}\end{aligned}$$

By properties 1 and 2,

$$W_{i,x} W_{x\oplus k,-j} = c^{\ell/2} W_{i,x} W_{x,-j} W_{k,-j} = W_{x,i\ominus j} W_{k,-j}$$

Thus, again using property 1, we get

$$(WP_kW)_{i,j} = W_{k,-j} \sum_x W_{i\ominus j,x} = W_{k,-j}(W\mathbf{1})_{i\ominus j} = W_{j,-k}\, c^{\ell/2}[i \ominus j = 0]$$

which is the desired result (the last equality follows from property 4). □

The importance of the Walsh transform, in the binary case, and the Fourier transform, in the general cardinality case, can be seen with the results that can be obtained by their use. They both unravel complex expressions, as will be shown. Similarly, the binary model gains importance through the analysis that it supports. As will be shown in the following sections, much of that analysis carries over to the general cardinality case by way of the Fourier transform.

4. Mutation and Crossover

We now begin (to be completed in Section 7) an extension to the simple binary GA Vose model to cardinality c. This model can be extended to the finite population model, as was done by Nix and Vose (1992) for the simple binary GA and by Vose (1995) in the general context of random heuristic search. The general cardinality model presented below is for a general form of the simple GA—a form using mutation and crossover masks. This form allows for various specialized operators used in practice. Remarks 1 and 2 show how the masks can be specialized to mutation rates and crossover rates, as required by many practitioners.

A mutation of $y \in \Omega$ is $y \oplus m$, where mask $m \in \Omega$ is randomly selected under the distribution $\mu \in \Lambda_{c^\ell}$ (the probability of selecting m is μ_m). The mutation distribution is called positive if $\mu > 0$.

REMARK 1 (mutation rates): For binary GAs, one often specifies a mutation rate: $0 \leq \mu < 0.5$. A corresponding mutation distribution that yields the same results is given by

$$\mu_m \equiv \mu^{\mathbf{1}^T m}(1-\mu)^{\ell - \mathbf{1}^T m}$$

(see Vose, in press). Here, $\mathbf{1}^T m$ is the number of bits that will be mutated (i.e., flipped to their complement), and $\ell - \mathbf{1}^T m$ is the number of bits that do not get mutated. For $c > 2$, we can generalize this idea to a mutation rate μ and get

$$\mu_m \equiv \left(\frac{\mu}{c-1}\right)^{n(m)} (1-\mu)^{\ell - n(m)}$$

where $n(m)$ is the number of nonzero digits of m.

Children resulting from crossing over $y, z \in \Omega$ are

$$y \otimes m \oplus \bar{m} \otimes z \quad \text{(child 1)} \qquad \text{and} \qquad z \otimes m \oplus \bar{m} \otimes y \quad \text{(child 2)}$$

where mask $m \in \Omega_2$ is randomly selected under the distribution $\chi \in \Lambda_{2^\ell}$ (the probability of selecting m is χ_m). Only one child is kept.

Note that the underlying mechanism—mutation and crossover masks—is carried over to the general cardinality case intact. This is an important point because it allows the basic structure of the binary model to be preserved.

REMARK 2 (one-point crossover rates): For binary GAs, a crossover rate $\chi \in [0, 1]$ is often used. A corresponding crossover distribution that yields the same results for one-point crossover is given by

$$\chi_p \equiv \begin{cases} \dfrac{\chi}{\ell - 1} & \text{if } \; p = 2^b - 1 \quad \text{for some} \quad b \in (0, \ell) \\ 1 - \chi & \text{if } \; p = 0 \\ 0 & \text{otherwise} \end{cases}$$

(see Vose, in press). This result carries to the general cardinality case without modification.

The probability of a child $w \in \Omega$ resulting from the mutation of parents $y, z \in \Omega$ followed by crossover is

$$s_{y,z}(w) \equiv 0.5 \sum_{p \in \Omega_2} \sum_{j,k \in \Omega} \mu_j \mu_k (\chi_p + \chi_{\bar{p}})[(y \oplus j) \otimes p \oplus \bar{p} \otimes (z \oplus k) = w]$$

The probability of a child $w \in \Omega$ resulting from crossover of parents $y, z \in \Omega$ followed by mutation is

$$r_{y,z}(w) \equiv 0.5 \sum_{p \in \Omega_2} \sum_{j \in \Omega} \mu_j (\chi_p + \chi_{\bar{p}})[y \otimes p \oplus \bar{p} \otimes z \oplus j = w]$$

These forms match those for the binary case, except in minor details. This facilitates the analysis in Section 5 and makes straightforward the transfer of the following results to the general cardinality case. Vose (in press) showed for the binary case that these two probabilities are equal for arbitrary χ if μ is independent. Vose called μ independent if

$$\mu_j = \sum_{k \otimes p = 0} \mu_{j \oplus p} \sum_{\bar{k} \otimes p = 0} \mu_{j \oplus p}$$

for all j and k. If, instead, we define μ as independent with the same expression but with k limited to $k \in \Omega_2$, then his proofs carry through (see the Appendix) to give

THEOREM A2: *If μ is independent, then*

$$s_{y,z}(w) = r_{y,z}(w).$$

For the binary case, Vose (in press) showed that

$$r_{y,z}(w) = r_{y\oplus w,z\oplus w}(\mathbf{0}).$$

This relationship is of central importance to the model because it collapses a three-dimensional problem (the degrees of freedom are parent y, parent z and child w) to a two-dimensional case (the parents may vary, but the child is fixed at 0). Not only is a simplified model obtained, but the action of the group operation $\oplus$ is woven into the result, setting the stage for later application of the Fourier transform. The proof in the more general case is given below.

THEOREM 2: *For $w, y, z \in \Omega$, $r_{y,z}(w) = r_{y\ominus w,z\ominus w}(\mathbf{0})$.*

PROOF: It suffices to show the equivalence of

$$y \otimes p \oplus \bar{p} \otimes z \oplus j = w \qquad \text{and } (y \ominus w) \otimes p \oplus \bar{p} \otimes (z \ominus w) \oplus j = \mathbf{0}$$

Note that

$$y \otimes p \oplus \bar{p} \otimes z \oplus j = w$$

implies

$$y \otimes p \oplus \bar{p} \otimes z \oplus j = w \otimes p \oplus w \otimes \bar{p}$$

by Equation L1, giving

$$y \otimes p \oplus \bar{p} \otimes z \oplus j \oplus (-(w \otimes p \oplus w \otimes \bar{p})) = \mathbf{0}$$

After rearrangement we get the equivalent statement

$$((y \ominus w) \otimes p \oplus \bar{p} \otimes (z \ominus w)) \oplus j = \mathbf{0}$$ □

As a final note to this section, when mutation is zero, the probability of child 0 resulting from crossover of parents $y, z \in \Omega$ reduces to

$$s_{y,z}(0) = r_{y,z}(0) = 0.5 \sum_{p\in\Omega_2} (\chi_p + \chi_{\bar{p}})[y \otimes p \oplus \bar{p} \otimes z = 0]$$

5. Properties of M and M^*

This section demonstrates how the generalized model of mixing (mutation and crossover) supports the same kind of analysis as does the binary model. In fact, the statements of the lemmas and theorems are almost identical to those of the binary case. All aspects of mixing are captured by the $r_{i,j}(\mathbf{0})$ values. Let M and M^* be c^ℓ by c^ℓ matrices defined by

$$M_{i,j} = r_{i,j}(\mathbf{0}) \qquad \text{and} \qquad M^*_{y,z} = M_{y\ominus z,y}$$

Theorem 3 below provides various properties of M and M^*. In particular, the Fourier transform is used to unravel the intricacies of the mixing matrix M, showing the transform of its twist M^* to be triangular for every choice of mutation and crossover (as given by masks). The Fourier transform of M^* is denoted by $M^{*\wedge}$.

The following lemma will prove useful.

LEMMA 1: *$A^{H\wedge} = A^{\wedge H}$ and $A^{H\wedge *} = A^{*\wedge H}$.*

PROOF: The first identity follows from $(WAW^H)^H = W^{HH}A^HW^H = WA^HW^H$. For the second part,

$$(A^{H\wedge *})_{i,j} = (A^{H\wedge})_{i\ominus j,i}$$

Expanding this last term and taking the conjugate gives

$$\text{conj}\left(\sum_{u,v} W_{i\ominus j}\bar{A}_{v,u}W_{-i,v}\right) = \sum_{u,v} W_{j\ominus i}A_{v,u}W_{i,v}$$

On the other hand,

$$\text{conj}(A^{*\wedge H})_{i,j} = (A^{*\wedge})_{j,i} = (WA^*W^H)_{j,i} = \sum_{u,v} W_{j,u}A^*_{u,v}W_{-i,v} = \sum_{u,v} W_{j,u}A_{u\ominus v,u}W_{-i,v}$$

Making a change of variables $v \mapsto u \ominus v$, this sum becomes

$$\sum_{u,v} W_{j,u}A_{v,u}W_{-i,u\ominus v} = c^{\ell/2}\sum_{u,v} W_{j,u}A_{v,u}W_{-i,-v}W_{u,-i} = \sum_{u,v} W_{j\ominus i,u}A_{v,u}W_{i,v} \qquad \square$$

THEOREM 3 (properties of M and M^*):

1. *M is symmetric.*
2. $M^{\wedge}_{a,b} \neq 0 \Rightarrow a^{\mathrm{T}}b = 0$.
3. $M^{*\wedge}_{a,b} \neq 0 \Rightarrow b^{\mathrm{T}}(b \ominus a) = 0$.
4. *$M^{*\wedge}$ is lower triangular.*
5. $\mathbf{1}^{\mathrm{T}}M^* = \mathbf{1}^{\mathrm{T}}$.
6. *If mutation is zero, then $M = M^{\wedge}$.*
7. *If mutation is zero, then M^* is upper triangular.*

PROOF:

1. This follows directly from the definitions of $s_{y,z}(w)$ and $r_{y,z}(w)$.
2. Assume $(WMW^H)_{a,b} \neq 0$. Then,

$$0 \neq \sum_{y,z} W_{a,y}M_{y,z}W_{z,-b}$$

so

$$0 \neq \sum_{y,z} W_{a,y}W_{z,-b}\, 0.5 \sum_{p\in\Omega_2}\sum_{j\in\Omega} \mu_j(\chi_p + \chi_{\bar{p}})[(y \otimes p \oplus \bar{p} \otimes z) \oplus j = \mathbf{0}]$$

and then

$$0 \neq \sum_{y,z} W_{a,y} W_{z,-b} 0.5 \sum_{p \in \Omega_2} \sum_{\substack{j \in \Omega \\ y \otimes p \oplus \bar{p} \otimes z \oplus j = 0}} \mu_j (\chi_p + \chi_{\bar{p}})$$

Using Lemma A1 (see Appendix), parts 1 and 2, on the right side, gives an equivalent representation:

$$0 \neq 0.5 \sum_{p \in \Omega_2} (\chi_p + \chi_{\bar{p}}) \sum_{j \in \Omega} \mu_j \left(\sum_{y \otimes p = -j \otimes p} W_{a,y} \sum_{z \otimes \bar{p} = -j \otimes \bar{p}} W_{z,-b} \right)$$

This implies

$$0 \neq \sum_{y \otimes p = -j \otimes p} W_{a,y} \sum_{z \otimes \bar{p} = -j \otimes \bar{p}} W_{z,-b}$$

for some **p** and **j**. The desired result follows by noting that the varying digits of **y** are where the digits of **p** are zero and that the varying digits of **z** are where the digits of **p** are a 1. By Theorem 1, part 3, this implies that **a** must have zeros where **p** has zeros and that $-\mathbf{b}$ must have zeros where **p** has 1's. Thus, $a_i = 0$, or $b_i = 0$ for each i. (This, in turn, implies $a^{\mathrm{T}} b = 0$.)

3. Assume $M^{*\wedge}_{a,b} \neq 0$. Then,

$$0 \neq M^{*\wedge}_{a,b} = \mathrm{conj}(M^{*\wedge H}_{b,a}) = \mathrm{conj}(M^{H\wedge *}_{b,a}) = \mathrm{conj}(M^{\wedge H}_{b \ominus a, b}) = M^{\wedge}_{b, b \ominus a}$$

Now apply part 2.

4. Suppose to the contrary that $b > a$, and $(WM^* W^H)_{a,b} \neq 0$. By part 3, the latter hypothesis implies $b_i = 0$, or $a_i = b_i$ for each i. But $b > a$ implies for some i that $b_i > a_i \geq 0$, which is a contradiction.

5. By Theorem 2,

$$(\mathbf{1}^{\mathrm{T}} M^*)_k = \sum_j M^*_{j,k} = \sum_j M_{j \ominus k, j} = \sum_j r_{-k \oplus j, 0 \oplus j}(\mathbf{0}) = \sum_j r_{-k,0}(-\mathbf{j}) = 1$$

6. (This result for the binary case was first noticed by Alden Wright.) If mutation is zero, then

$$\begin{aligned} M^{\wedge}_{a,b} &= \sum_{y,z} W_{a,y} W_{z,-b} 0.5 \sum_{p \in \Omega_2} (\chi_p + \chi_{\bar{p}})[y \otimes p \oplus \bar{p} \otimes z = \mathbf{0}] \\ &= 0.5 \sum_{p \in \Omega_2} (\chi_p + \chi_{\bar{p}}) \sum_{y,z} W_{a,y} W_{z,-b} [y \otimes p \oplus \bar{p} \otimes z = \mathbf{0}] \end{aligned}$$

By Lemma A1 (see Appendix),

$$M^{\wedge}_{a,b} = 0.5 \sum_{p \in \Omega_2} (\chi_p + \chi_{\bar{p}}) \sum_{y,z} W_{a,y} W_{z,-b} [y \otimes p = \mathbf{0}][\bar{p} \otimes z = \mathbf{0}]$$

so

$$M^{\wedge}_{a,b} = 0.5 \sum_{p \in \Omega_2} (\chi_p + \chi_{\bar{p}}) \sum_y W_{a,y}[y \otimes p = \mathbf{0}] \sum_z W_{z,-b}[\bar{p} \otimes z = \mathbf{0}]$$

From Theorem 1, part 3, and the definition of W, we get

$$\begin{aligned}
\sum_y W_{a,y}[y \otimes p = \mathbf{0}] &= c^{\ell - \mathbf{1}^T p} c^{-\ell/2} [a \otimes \bar{p} = \mathbf{0}] \\
\sum_z W_{z,-b}[\bar{p} \otimes z = \mathbf{0}] &= c^{\ell - \mathbf{1}^T \bar{p}} c^{-\ell/2} [b \otimes p = \mathbf{0}]
\end{aligned}$$

which, using Lemma A1, gives

$$\begin{aligned}
M^{\wedge}_{a,b} &= 0.5 \sum_{p \in \Omega_2} (\chi_p + \chi_{\bar{p}})[a \otimes \bar{p} \oplus b \otimes p = \mathbf{0}] \\
&= 0.5 \sum_{p \in \Omega_2} (\chi_p + \chi_{\bar{p}})[b \otimes p \oplus \bar{p} \otimes a = \mathbf{0}] \\
&= M_{b,a} = M_{a,b}
\end{aligned}$$

7. $M = M^H$, since M is symmetric and real. Now, since $M = M^{\wedge}$ (from part 6), we get $M = M^{H\wedge}$, and so $M^* = M^{H\wedge *} = M^{*\wedge H}$ (the last equality by Lemma 1). But $M^{*\wedge}$ is lower triangular (by part 4). Hence, $M^{*\wedge H} = M^*$ is upper triangular. □

The significance of being able to triangularize M^* is that access to its spectrum is gained. As will be shown in the following section, this leads to a formula for the spectrum of the mixing operator $\mathcal{M}$ and, ultimately, to knowledge concerning the transition operator $\mathcal{G}$, which maps a population to the expected next population.

A true understanding of the behavior of the general cardinality model requires knowledge of its asymptotic properties (although a discussion of why and how is beyond the scope of this paper). Fixed points of $\mathcal{G}$ are asymptotically stable if the spectral radius of the differential at the fixed point is less than 1. The following section investigates the differential at fixed points for the zero mutation case.

6. Spectral Properties of $\mathcal{G}()$, $\mathcal{M}()$, and M^*

In this section we derive an explicit formula for the spectrum of the twist of the mixing matrix (Theorem 4) that directly leads to the spectrum of the mixing operator $\mathcal{M}$ (Lemma 2 and Theorem 5). The derivation parallels that given by Vose (in press). Fixed points are asymptotically stable if the spectral radius of the differential $d\mathcal{M}_x$ at the fixed point is less than 1. We show below that, in general, the spectral radius of the differential at the fixed point x is bounded by 1 (Corollary 1). In the zero mutation case, the spectrum of the differential $d\mathcal{G}_x$ at vertex fixed points is also determined. Surprisingly, it turns out to be independent of the cardinality c of the representation's alphabet.

As in Vose (1993), define the functions $\mathcal{G}$ and $\mathcal{M}$ by

$$\mathcal{G}(\mathbf{x}) \equiv \mathcal{M}(Fx/f^T x)$$

where

$$\mathcal{M}(x)_k \equiv x^{\mathrm{T}} P_k^{\mathrm{T}} M P_k x$$

Let $\mathbf{x} \in \Lambda_{c^\ell}$ represent the distribution of members of a population. Then, $\mathcal{G}(\mathbf{x})$ is the expected proportion of members in the next population. For simple GAs using mutation rates less than 0.5, a fundamental conjecture of GAs is that $\mathcal{G}^n(\mathbf{x})$ converges as $n \to \infty$ for every $\mathbf{x} \in \Lambda_{2^\ell}$. We are interested in exploring the general cardinality case with $\mathbf{x} \in \Lambda_{c^\ell}$. A fixed point x of $\mathcal{G}$ is asymptotically stable if $\rho(d\mathcal{G}_x) < 1$. Vose and Liepins (1991) and Vose and Wright (1995) have shown that

$$d\mathcal{G}_x = \frac{1}{f^{\mathrm{T}}x} d\mathcal{M}_{Fx/f^{\mathrm{T}}x} F \left(I - \frac{xf^{\mathrm{T}}}{f^{\mathrm{T}}x} \right)$$

This relationship, as well as the following result, are unaffected by the cardinality of the components used to form matrix M.

LEMMA 2 (Vose, in press): *If* $\mathbf{x}$ *is a fixed point of* $\mathcal{G}$, *then* $\sigma(d\mathcal{G}_x)$ *is equivalent to*

$$\frac{1}{\mathbf{1}^{\mathrm{T}}Fx} \sigma(d\mathcal{M}_{Fx/f^{\mathrm{T}}x}F)$$

except that an eigenvalue of $\sigma(d\mathcal{M}_{Fx/f^{\mathrm{T}}x}F)$ *equal to 2 should be replaced by zero.*

Thus, knowledge of $\sigma(d\mathcal{M}_z)$ is critical to our understanding of the dynamical system component of GAs. It is easy to show that

$$d\mathcal{M}_z = 2 \sum_j z_j P_j^{\mathrm{T}} M^* P_j$$

Hence,

$$\sigma(d\mathcal{M}_{Fx/f^{\mathrm{T}}x}\mathbf{F}) = \frac{2}{f^{\mathrm{T}}x} \sigma \left(\sum_j f_j x_j P_j^{\mathrm{T}} M^* P_j F \right)$$

For the special case when x is a vertex of the simplex, say, $x = b_j$, this reduces to

$$2\sigma(P_j^{\mathrm{T}} M^* P_j F) = 2\sigma(M^* P_j F P_j^H)$$

since P_j is real, and $P_j^{-1} = P_j^H$. The following lemma will be useful.

LEMMA 3: $P_j F P_j^H = \mathrm{diag}(P_j f)$.

PROOF:

$$\begin{aligned}(P_j F P_j^H)_{a,b} &= \sum_{u,v} (P_j)_{a,u} [u = v] f_u (P_j)_{b,v} = \sum_{u,v} [j \oplus a = u][u = v] f_u [j \oplus b = v] \\ &= \sum_u [j \oplus a = u] f_u [j \oplus b = u] = [a = b] f_{j \oplus a}\end{aligned}$$

□

The spectra $\sigma(d\mathcal{M}_z)$ and $\sigma(M^*)$ are related (as shown below) and are relevant to the spectra $\sigma(d\mathcal{G})$. We first consider $\sigma(M^*)$.

LEMMA 4: $\sigma(M^*) = \{M^{*\wedge}_{i,i} : 0 \le i < c^\ell\}$.

PROOF: The spectrum is unchanged by a change in basis. The rest follows, since WM^*W^H is lower triangular (Theorem 3). □

The following two lemmas will be used to compute the spectrum of M^*.

LEMMA 5: $M^{*\wedge}_{a,a} = c^{-\ell/2}(WM\mathbf{1})_a$.

PROOF: By part 2 of Theorem 1, the left-hand side is

$$\sum_y \sum_z W_{a,y} M^*_{y,z} W_{z,-a} = \sum_y \sum_z c^{-\ell/2} W_{a,y\ominus z} M_{y \ominus z, y}$$

Making the change of variables $b = y \ominus z$ yields

$$c^{-\ell/2} \sum_b W_{a,b} \sum_z M_{b,b\oplus z} = c^{-\ell/2} \sum_b W_{a,b}(M\mathbf{1})_b = c^{-\ell/2}(WM\mathbf{1})_a \qquad \square$$

LEMMA 6:

$$(M\mathbf{1})_a = 0.5 \sum_{p \in \Omega_2} c^{\mathbf{1}^T p}(\chi_p + \chi_{\bar{p}}) \sum_{j \in \Omega} \mu_j [a \otimes p = -p \otimes j]$$

PROOF:

$$\begin{aligned}
(M\mathbf{1})_a &= \sum_x r_{a,x}(0) \\
&= 0.5 \sum_{p\in\Omega_2} (\chi_p + \chi_{\bar{p}}) \sum_{j\in\Omega} \mu_j \sum_x [a \otimes p \oplus \bar{p} \otimes x \oplus j = 0] \\
&= 0.5 \sum_{p\in\Omega_2} (\chi_p + \chi_{\bar{p}}) \sum_{j\in\Omega} \mu_j \sum_x [\bar{p} \otimes x = -\bar{p} \otimes j][a \otimes p = -p \otimes j] \\
&= 0.5 \sum_{p\in\Omega_2} (\chi_p + \chi_{\bar{p}}) \sum_{j\in\Omega} \mu_j [a \otimes p = -p \otimes j] \sum_x [\bar{p} \otimes x = -\bar{p} \otimes j]
\end{aligned}$$

But the innermost sum is $c^{\mathbf{1}^T p}$, since the only position of x that can vary corresponds to positions of p that are 1. Hence,

$$(M\mathbf{1})_a = 0.5 \sum_{p \in \Omega_2} c^{\mathbf{1}^T p}(\chi_p + \chi_{\bar{p}}) \sum_{j \in \Omega} \mu_j [a \otimes p = -p \otimes j] \qquad \square$$

THEOREM 4:

$$\sigma(M^*) = \left\{ M^{*\wedge}_{a,a} = 0.5 c^{\ell/2} \sum_{p\in\Omega_2} [\bar{p} \otimes a = 0](\chi_p + \chi_{\bar{p}}) \sum_{j\in\Omega} \mu_j W_{a,j}, \quad 0 \le a < c^\ell \right\}$$

PROOF: By Lemmas 4, 5, and 6, the left-hand side has typical elements

$$\begin{aligned}
M^{\wedge}_{a,a} &= c^{-\ell/2}(WM\mathbf{1})_a = c^{-\ell/2} \sum_x W_{a,x}(M\mathbf{1})_x \\
&= 0.5 c^{-\ell/2} \sum_{p\in\Omega_2} c^{\mathbf{1}^T p}(\chi_p + \chi_{\bar{p}}) \sum_{j\in\Omega} \mu_j \sum_x W_{a,x}[x \otimes p = -p \otimes j]
\end{aligned}$$

Consider the innermost sum. By Theorem 1, part 2, it factors as

$$c^{-\ell/2} \prod_{i=0}^{\ell-1} \left(c^{\ell/2} \sum_{x_i=0}^{c-1} W_{a,x_i b_i} [x \otimes p = -p \otimes j] \right)$$

Using Theorem 1, parts 2 and 3, and noting that the only varying digits of x occur where p has zeros, gives this last expression as

$$W_{a,-p\otimes j} c^{\ell - 1^T p} [\bar{p} \otimes a = 0]$$

Thus,

$$\begin{aligned} M^{*\wedge}_{a,a} &= 0.5c^{-\ell/2} \sum_{p \in \Omega_2} c^{1^T p} (\chi_p + \chi_{\bar{p}}) \sum_{j \in \Omega} \mu_j W_{a,-p\otimes j} c^{\ell - 1^T p} [\bar{p} \otimes a = 0] \\ &= 0.5c^{\ell/2} \sum_{p \in \Omega_2} [\bar{p} \otimes a = 0] (\chi_p + \chi_{\bar{p}}) \sum_{j \in \Omega} \mu_j W_{-a,p\otimes j} \end{aligned}$$

The observation that $\bar{p} \otimes a = 0 \Rightarrow a = p \otimes a$ allows $W_{-a,p\otimes j}$ to be simplified to $W_{-a,j}$ in the expression above. The proof is completed with the change of variable $a \mapsto -a$ (note that $\bar{p} \otimes a = 0 \Leftrightarrow \bar{p} \otimes (-a) = 0$). □

The following result provides bounds (via Lemma 4) on the moduli of eigenvalues of M^*.

COROLLARY 1: *The moduli of eigenvalues of M^* satisfy*

$$|M^{*\wedge}_{0,0}| = M^{*\wedge}_{0,0} = 1 \qquad \text{and} \qquad |M^{*\wedge}_{a,a}| \leq 0.5, \quad a \neq 0$$

PROOF: For $a = 0$ we get, using Lemma 4 and the proof of Theorem 4,

$$M^{*\wedge}_{0,0} = 0.5c^{\ell/2} \sum_{p \in \Omega_2} (\chi_p + \chi_{\bar{p}}) \sum_{j \in \Omega} \mu_j c^{-\ell/2} = 0.5 \sum_{p \in \Omega_2} (\chi_p + \chi_{\bar{p}}) = 1.0$$

Suppose that $a > 0$. Then, by Lemma 4 and the proof of Theorem 4,

$$\begin{aligned} |M^{*\wedge}_{a,a}| &= \left| 0.5c^{\ell/2} \sum_{p \in \Omega_2} [\bar{p} \otimes a = 0] (\chi_p + \chi_{\bar{p}}) \sum_{j \in \Omega} \mu_j W_{-a,j} \right| \\ &\leq 0.5c^{\ell/2} \sum_{p \in \Omega_2} [\bar{p} \otimes a = 0] (\chi_p + \chi_{\bar{p}}) \sum_{j \in \Omega} \mu_j c^{-\ell/2} \\ &= 0.5 \sum_{p \in \Omega_2} [\bar{p} \otimes a = 0] (\chi_p + \chi_{\bar{p}}) \end{aligned}$$

Since $a \neq 0$, $\bar{p} \otimes a = 0 \Rightarrow p \otimes a \neq 0$. Thus, no component of χ will appear in the sum $\sum_{p \in \Omega_2} (\chi_p + \chi_{\bar{p}})$ more than once. Hence,

$$|M^{*\wedge}_{a,a}| \leq 0.5 \sum_{p \in \Omega_2} \chi_p = 0.5$$

□

The next result gives the relationship between $\sigma(d\mathcal{M}_z)$ and $\sigma(M^*)$.

THEOREM 5: $\sigma(d\mathcal{M}_z) = 2\sigma(M^*)$.

PROOF: Thus,

$$d\mathcal{M}_z^\wedge = 2\sum_j \mathbf{z}_j P_j^{T\wedge} M^{*\wedge} P_j^\wedge$$

Moreover, the eigenvalues of $d\mathcal{M}_z^\wedge$ and $d\mathcal{M}_z$ are identical (since the former is the latter in a different basis). But $P_j^\wedge$ is diagonal (Theorem 1), and $M^{*\wedge}$ is lower triangular (Theorem 3). Hence, the spectrum is given by the diagonal elements. A typical diagonal element is

$$2\sum_j \mathbf{z}_j (P_j^{T\wedge})_{a,a} M_{a,a}^{*\wedge} (P_j^\wedge)_{a,a}$$

By Theorem 1, part 6, this is

$$2\sum_j \mathbf{z}_j c^\ell W_{a,j} M_{a,a}^{*\wedge} W_{a,-j} = 2M_{a,a}^{*\wedge} \sum_j \mathbf{z}_j = 2M_{a,a}^{*\wedge}$$

□

Returning to the spectrum of $d\mathcal{G}_x$ at a vertex fixed point (with positive mutation, all fixed points are in the interior of Λ, hence a vertex fixed point corresponds to the zero mutation case), suppose that $x = b_j$, and mutation is zero. By Lemma 2, $\sigma(d\mathcal{G}_x)$ is equivalent to

$$f_j^{-1}\sigma(d\mathcal{M}_{Fx/f^Tx}F)'$$

where the prime indicates that an eigenvalue of maximum modulus should be replaced by zero. The earlier discussion preceding Lemma 4 shows this to be

$$2f_j^{-1}\sigma(M^* \operatorname{diag}(P_j f))'$$

Since M^* is upper triangular (Theorem 3, part 7) and its diagonal is explicitly known, this leads directly to the following theorem.

THEOREM 6: *With zero mutation,*

$$\sigma(d\mathcal{G}_{b_j}) = \left\{ \frac{f_{i\oplus j}}{f_j} \sum_{k\in\Omega_2} (\chi_k + \chi_{\bar{k}})[k \otimes i = 0], \quad i > 0 \right\} \bigcup \{0\}$$

PROOF:

$$\begin{aligned} 2f_j^{-1}(M^* \operatorname{diag}(P_j f))_{i,i} &= 2f_j^{-1}(M^*)_{i,i}(P_j f)_i = 2f_j^{-1}M_{0,i}f_{i\oplus j} \\ &= \frac{f_{i\oplus j}}{f_j} \sum_{p\in\Omega_2} (\chi_p + \chi_{\bar{p}})[p \otimes i = 0] \end{aligned}$$

When $i = 0$, this term sums to 2 and should then be replaced with 0. □

It is interesting to note that this result is identical to the binary case. Therefore, the applications given in Vose and Wright (1995) carry over to the general cardinality case.

7. Specialization to Mutation and Crossover Rates

Mutation and crossover masks provide a nice generalization for mutation and crossover rates and allow for various specialized operators used in practice. However, many users of GAs employ only rates. In this section we specialize the results on the spectrum of M^* to GAs using mutation rates and one-point crossover rates. From Remarks 1 and 2, we have

$$\mu_m \equiv \left(\frac{\mu}{c-1}\right)^{n(m)} (1-\mu)^{\ell - n(m)}$$

and

$$\chi_p \equiv \begin{cases} \dfrac{\chi}{\ell - 1} & \text{if} \quad p = 2^b - 1 \quad \text{for some} \quad b \in (0, \ell) \\ 1 - \chi & \text{if} \quad p = 0 \\ 0 & \text{otherwise} \end{cases}$$

with $0 \leq \mu < 1 - c^{-1}$ and $0 \leq \chi \leq 1$.

The following results generalize those found in Koehler (1994) for the binary case.

THEOREM 7: *When using mutation and one-point crossover rates, the spectrum of M^* is*

$$\sigma(M^*) = \left\{ 1.0, \quad M^{*\wedge}_{a,a} = 0.5 \left(1 - \chi \frac{\text{width}(a)}{\ell - 1}\right) \left(1 - \frac{c\mu}{c-1}\right)^{n(a)} : \quad 1 \leq a < c^{\ell} \right\}$$

where, for $a > 0$,

$$\text{width}(a) = \max\{i : b_i^{\mathrm{T}} a > 0\} - \min\{i : b_i^{\mathrm{T}} a > 0\}$$

PROOF: Substitution with the mutation rate-equivalent distribution into the expression given by Theorem 4 shows a typical element of $\sigma(M^*)$ as

$$\begin{aligned} &0.5c^{\ell/2} \sum_{p \in \Omega_2} [\bar{p} \otimes a = 0](\chi_p + \chi_{\bar{p}}) \sum_{j \in \Omega} \left(\frac{\mu}{c-1}\right)^{n(j)} (1-\mu)^{\ell - n(j)} W_{a,j} \\ &\quad = 0.5c^{\ell/2}(1-\mu)^{\ell} \sum_{p \in \Omega_2} [\bar{p} \otimes a = 0](\chi_p + \chi_{\bar{p}}) \sum_{j \in \Omega} \left(\frac{\eta}{c-1}\right)^{n(j)} W_{a,j} \end{aligned}$$

where $\eta \equiv \mu/(1-\mu)$. The case for $a = 0$ was given in Corollary 1. Assume that a is nonzero. From the above it follows that

$$M^{*\wedge}_{a,a} = 0.5c^{\ell/2}(1-\mu)^{\ell} \sum_{j \in \Omega} \left(\frac{\eta}{c-1}\right)^{n(j)} W_{a,j} \sum_{p \in \Omega_2} [\bar{p} \otimes a = 0](\chi_p + \chi_{\bar{p}})$$

Consider the last sum. Since $a \neq 0$ and $\chi_p = 0$, if $p \neq 2^b - 1$ for some $b \in (0, \ell)$ and $p \neq 0$, we get

$$\sum_{p \in \Omega_2} [\bar{p} \otimes a = 0]\chi_p = \frac{\chi}{\ell - 1} \sum_{b=1}^{\ell - 1} [(2^{\ell} - 2^b) \otimes a = 0]$$

Similarly,

$$\sum_{p\in\Omega_2} [\bar{p} \otimes a = 0]\chi_{\bar{p}} = \sum_{p\in\Omega_2} [p \otimes a = 0]\chi_p = 1 - \chi + \frac{\chi}{\ell - 1}\sum_{b=1}^{\ell-1} [(2^b - 1) \otimes a = 0]$$

Combining these two sums gives

$$1 - \chi + \frac{\chi}{\ell - 1}\sum_{b=1}^{\ell-1} ([(2^b - 1) \otimes a = 0] + [(2^\ell - 2^b) \otimes a = 0])$$

The summation is $\ell-$ width(a) $-$ 1. Substituting this result into the expression for $M^{*\wedge}_{a,a}$ and simplifying gives, for $a > 0$,

$$M^{*\wedge}_{a,a} = 0.5c^{\ell/2}(1 - \mu)^\ell \left(1 - \chi\frac{\text{width}(a)}{\ell - 1}\right)\sum_{j\in\Omega}\left(\frac{\eta}{c-1}\right)^{n(j)} W_{a,j}$$

Next consider the summation above. First note that (using Theorem 1, part 3)

$$\begin{aligned}
\sum_{y=0}^{c-1} W_{a,yb_j}\left(\frac{\eta}{c-1}\right)^{n(yb_j)} &= W_{a,0} + \left(\frac{\eta}{c-1}\right)\sum_{y=1}^{c-1} W_{a,yb_j} \\
&= W_{a,0}\left(1 - \left(\frac{\eta}{c-1}\right)\right) + W_{a,0}\left(\frac{\eta}{c-1}\right)c[a_j = 0] \\
&= \frac{W_{a,0}}{1-\mu}\left(1 - \frac{c\mu[a_j \neq 0]}{c-1}\right)
\end{aligned}$$

Therefore,

$$\begin{aligned}
\sum_{j\in\Omega}\left(\frac{\eta}{c-1}\right)^{n(j)} W_{a,j} &= \sum_{y_0=0}^{c-1}\sum_{y_1=0}^{c-1}\cdots \\
&\quad \sum_{y_{\ell-1}=0}^{c-1}\left(\frac{\eta}{c-1}\right)^{n(y_0b_0\oplus y_1b_1\oplus\cdots y_{\ell-1}b_{\ell-1})} W_{a,y_0b_0\oplus y_1b_1\oplus\cdots y_{\ell-1}b_{\ell-1}} \\
&= (1-\mu)^{-1}\left(1 - \frac{c\mu[a_{\ell-1} \neq 0]}{c-1}\right)\sum_{y_0=0}^{c-1}\sum_{y_1=0}^{c-1}\cdots \\
&\quad \sum_{y_{\ell-2}=0}^{c-1}\left(\left(\frac{\eta}{c-1}\right)^{n(y_0b_0\oplus y_1b_1\oplus\cdots y_{\ell-2}b_{\ell-2})} W_{a,y_0b_0\oplus y_1b_1\oplus\cdots y_{\ell-2}b_{\ell-2}}\right) \\
&= W_{a,0}\prod_{j=0}^{\ell-1}(1-\mu)^{-1}\left(1 - \frac{c\mu[a_j \neq 0]}{c-1}\right) \\
&= c^{-\ell/2}(1-\mu)^{-\ell}\left(1 - \frac{c\mu}{c-1}\right)^{n(a)}
\end{aligned}$$

Hence,

$$M^{*\wedge}_{a,a} = 0.5\left(1 - \chi\frac{\text{width}(a)}{\ell-1}\right)\left(1 - \frac{c\mu}{c-1}\right)^{n(a)}$$

□

Note that if $0 \le \mu \le 1 - c^{-1}$, all the eigenvalues are nonnegative.

COROLLARY 2: *If $0 \le \mu \le 1 - c^{-1}$, then the eigenvalue of M^* having the second largest modulus is*

$$0.5\left(1 - \frac{c\mu}{c-1}\right)$$

with multiplicity $(c-1)\ell$.

8. Summary and Future Directions

We presented a complete generalization of the Vose GA model from its binary formulation to a model where GA strings are composed of elements drawn from alphabets of cardinality c. The spectral properties were determined, and explicit formulas for the general case were provided. The general case was specialized to a commonly used variant employing mutation rates and crossover rates.

There are still several directions for further generalizations. Strings could be composed of components drawn from different cardinality alphabets. Here we restricted each component to cardinality c. Another direction would be generalizing components to real or complex numbers. As in the present study, these further generalizations would require reexamining the mechanisms of mutation and crossover, the tools of analysis (e.g., finite Fourier matrices), and practical implementation details.

With the generalization of this paper, empirical studies similar to those done by De Jong (1995) and Spears and De Jong (1997) on the transient behavior of GAs could be extended to also consider the role of alphabet cardinality on GA performance. These future studies could shed light on the mixed results found by others (L. Davis, 1991; Shaffer, 1984). Indeed, with the model given in this paper, these questions could, in principle, be answered directly.

Table 1 summarizes the differences between terminology and results for the general cardinality case and the binary case.

Appendix

The following useful properties are employed often:

LEMMA A1: *Let $j, k, p, q \in Z_c$. Then,*

1. $p \otimes k = j \otimes k$ *and* $q \otimes \bar{k} = j \otimes \bar{k} \Rightarrow p \otimes k \oplus q \otimes \bar{k} = j$.
2. $p \otimes k = j \otimes k$ *and* $q \otimes \bar{k} = j \otimes \bar{k} \Leftarrow p \otimes k \oplus q \otimes \bar{k} = j$ *and* $k \in \Omega_2$.

The converse of part 1 is not generally true (although it is true for $k \in \Omega_2$, as stated in part 2).

The following result follows Vose (in press) in all important details. The use of Lemma A1 is the only necessary deviation.

THEOREM A2: *If μ is independent, then $s_{y,z}(w) = r_{y,z}(w)$.*

Table 1. Summary of notation and results for the binary and general cardinality cases.

Notation/Result	Binary Case	General Cardinality Case
$a \oplus b$	EXCLUSIVE-OR	$(a+b) \bmod c$
$a \otimes b$	AND	$(ab) \bmod c$
$a \ominus b$	$a \oplus b$	$a \oplus (-b)$
$\sum_{k=0}^{\ell-1} [s_k \neq 0]$	$s^{\mathrm{T}}\mathbf{1}$	$n(s)$
$p \otimes k = j \otimes k$		
$\Leftrightarrow$		
$q \otimes \bar{k} = j \otimes k$	Always	If $k \in \Omega_2$
$(p \otimes k) \oplus (q \otimes \bar{k}) = j$		
$W_{r,s}$	$2^{-\ell/2}(-1)^{r^{\mathrm{T}}s}$	$c^{-\ell/2}e(r^{\mathrm{T}}s/c)$
$r_{y,z}(w)$	$r_{y\oplus w,z\oplus w}(\mathbf{0})$	$r_{y\ominus w,z\ominus w}(\mathbf{0})$
$M^*_{y,z}$	$M_{y\oplus z,y}$	$M_{y\ominus z,y}$
$\sigma(M^*)$	$\{(WM^*W)_{i,i} : 0 \leq i < 2^\ell\}$	$\{(WM^*W^H)_{i,i}: 0 \leq i < c^\ell\}$
$(WM^*W^H)_{a,a} = M^{*\wedge}_{a,a}$	$2^{-\ell/2}(WM\mathbf{1})_a$	$c^{-\ell/2}(WM\mathbf{1})_a$
Mutation rates	$\mu_m \equiv \mu^{\mathbf{1}^{\mathrm{T}}m}(1-\mu)^{\ell-\mathbf{1}^{\mathrm{T}}m}$	$\mu_m \equiv \left(\frac{\mu}{c-1}\right)^{n(m)}(1-\mu)^{\ell-n(m)}$
$\sigma(M^*)$ using mutation and one-point crossover rates	$\begin{cases} 1.0 \\ 0.5\left(1-\chi\frac{\mathrm{width}(a)}{\ell-1}\right)(1-2\mu)^{\mathbf{1}^{\mathrm{T}}a} \\ : 1 \leq a < 2^\ell \end{cases}$	$\left\{\begin{matrix} 1.0 \\ 0.5\left(1-\chi\frac{\mathrm{width}(a)}{\ell-1}\right)\left(1-\frac{c\mu}{c-1}\right)^{n(a)} \\ : 1 \leq a < c^\ell \end{matrix}\right\}$
Subspectral radius for M^* using mutation and one-point crossover rates	$0.5 - 2\mu$ for $0 \leq \mu \leq 0.5$	$0.5\left(1-\frac{c\mu}{c-1}\right)$ for $0 \leq \mu \leq 1-c^{-1}$

PROOF:

$$\begin{aligned} s_{y,z}(w) &= 0.5\sum_{p\in\Omega_2}(\chi_p+\chi_{\bar{p}})\sum_{j,k\in\Omega}\mu_j\mu_k[(y\oplus j)\otimes p\oplus\bar{p}\otimes(z\oplus k)=w] \\ &= 0.5\sum_{p\in\Omega_2}(\chi_p+\chi_{\bar{p}})\sum_{j,k\in\Omega}\mu_j\mu_k[y\otimes p\oplus\bar{p}\otimes z=w\ominus(j\otimes p\oplus\bar{p}\otimes k)] \\ &= 0.5\sum_{p\in\Omega_2}(\chi_p+\chi_{\bar{p}})\sum_{h\in\Omega}[y\otimes p\oplus\bar{p}\otimes z=w\ominus h]\sum_{j\otimes p\oplus\bar{p}\otimes k=h}\mu_j\mu_k \end{aligned}$$

Now, since $p \in \Omega_2$, we have, using Lemma A1, parts 1 and 2, that the last expression is

$$0.5\sum_{p\in\Omega_2}(\chi_p+\chi_{\bar{p}})\sum_{h\in\Omega}[y\otimes p\oplus\bar{p}\otimes z=w\ominus h]\sum_{j\otimes p=h\otimes p}\mu_j\sum_{k\otimes\bar{p}=h\otimes\bar{p}}\mu_k$$

Making the change of variables $j \mapsto j \oplus h$ in the sum involving μ_j, and making the change of variable $k \mapsto k \oplus h$ in the sum involving μ_k, yields

$$0.5\sum_{p\in\Omega_2}(\chi_p+\chi_{\bar{p}})\sum_{h\in\Omega}[y\otimes p\oplus\bar{p}\otimes z=w\ominus h]\sum_{j\otimes p=0}\mu_{h\oplus j}\sum_{k\otimes\bar{p}=0}\mu_{h\oplus k}$$

But μ is independent, which gives

$$s_{y,z}(w) = 0.5 \sum_{p \in \Omega_2} (\chi_p + \chi_{\bar{p}}) \sum_{h \in \Omega} [y \otimes p \oplus \bar{p} \otimes z = w \ominus h] \mu_h$$

Thus,

$$s_{y,z}(w) = 0.5 \sum_{p \in \Omega_2} \sum_{h \in \Omega} \mu_h (\chi_p + \chi_{\bar{p}}) [y \otimes p \oplus \bar{p} \otimes z \oplus h = w] = r_{y,z}(w) \qquad \square$$

References

Antonisse, J. (1989). A new interpretation of schema notation that overturns the binary encoding constraint. In J. D. Schaffer (Ed.), *Proceedings of the Third International Conference on Genetic Algorithms* (pp. 86–97). San Mateo, CA: Morgan Kaufmann.

Aytug, H., & Koehler, G. J. (1996). Stopping criteria for finite length genetic algorithms. *ORSA Journal on Computing, 2*(8), 183–191.

Bhattacharyya, S., & Koehler, G. J. (1994). An analysis of genetic algorithms of cardinality 2^v. *Complex Systems, 8*, 227–256.

Davis, L. (1991). *Handbook of genetic algorithms.* New York: Van Nostrand Reinhold.

Davis, T. (1991). *Toward an extrapolation of the simulated annealing convergence theory onto the simple genetic algorithm.* Unpublished doctoral dissertation, Department of Decision and Information Sciences, University of Florida, Gainesville.

De Jong, K. A., Spears, W. M., & Gordon, D. F. (1995). Using Markov chains to analyze GAFOs. In L. D. Whitley and M. D. Vose (Eds.), *Foundations of genetic algorithms 3* (pp. 115–137). San Mateo, CA: Morgan Kaufmann.

Juliany, J., & Vose, M. D. (1994). The genetic algorithm fractal. *Evolutionary Computation, 2*(2), 165–180.

Koehler, G. J. (1994). A proof of the Vose-Liepins conjecture. *Annals of Mathematics and Artificial Intelligence, 10*, 409–422.

Nix, A., & Vose, M. D. (1992). Modeling genetic algorithms with Markov chains. *Annals of Mathematics and Artificial Intelligence, 5*, 79–88.

Reeves, C., & Wright, C. (1995). An experimental design perspective on genetic algorithms. In W. M. Spears & D. F. Gordon (Eds.), *Foundations of genetic algorithms 3* (pp. 85–100). San Francisco, CA: Morgan Kaufmann.

Shaffer, J. D. (1984). *Some experiments in machine learning using vector evaluated genetic algorithms.* Unpublished doctoral dissertation. Computer Science Department, Vanderbilt University, Nashville, TN.

Spears, W. M., & De Jong, K. A. (1997). Analyzing GAs using Markov models with semantically ordered and lumped states. In R. K. Belew and M. D. Vose (Eds.), *Foundations of genetic algorithms 4* (pp. 85–100). San Mateo, CA: Morgan Kaufmann.

Suzuki, J. (1993). A Markov chain analysis on a genetic algorithm. In S. Forrest (Ed.), *Proceedings of the Fifth International Conference on Genetic Algorithms* (pp. 146–153). San Mateo, CA: Morgan Kaufmann.

Vose, M. D. (May, 1990). Formalizing genetic algorithms. *Proceedings of the IEEE Workshop on Genetic Algorithms, Neural Networks, and Simulated Annealing Applied to Signal and Image Processing.* Glasgow, Scotland.

Vose, M. D. (1993). Modeling simple genetic algorithms. In L. D. Whitley (Ed.), *Foundations of genetic algorithms, 2* (pp. 63–73). San Mateo, CA: Morgan Kaufmann.

Vose, M. D. (1995). Modeling simple genetic algorithms. *Evolutionary Computation, 3*(4), 453–472.

Vose, M. D. (1997). Logarithmic convergence of random heuristic search. *Evolutionary Computation, 4*(4), 395–404.

Vose, M. D. (in press). *The simple genetic algorithm: Foundations and theory.* Cambridge, MA: MIT Press.

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

Vose, M. D., & Wright, A. H. (1994). Simple genetic algorithms with linear fitness. *Evolutionary Computation, 2*(4), 347–368.

Vose, M. D., & Wright, A. H. (1995). Stability of vertex fixed points and applications. In L. D. Whitley & M. D. Vose (Eds.), *Foundations of genetic algorithms, 3* (pp. 103–113). San Mateo, CA: Morgan Kaufmann.