The Simple Genetic Algorithm and the Walsh Transform: part II, The Inverse

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

Alden H. Wright[1]
Computer Science Dept.
The University of Montana
Missoula, MT 59812-1008
wright@cs.umt.edu

ABSTRACT

This paper continues the development, begun in part I, of the relationship between the simple genetic algorithm and the Walsh transform. The mixing scheme (comprised of crossover and mutation) is essentially "triangularized" when expressed in terms of the Walsh basis. This leads to a formulation of the inverse of the expected next generation operator. The fixed points of the mixing scheme are also determined, and a formula is obtained giving the fixed point corresponding to any starting population. Geiringer's theorem follows from these results in the special case corresponding to zero mutation.

1 Introduction

The formalism used in this paper to model the simple genetic algorithm is that of random heuristic search with heuristic $\mathcal{G}$, though the focus is on the case where the search space Ω consists of c-ary strings of length ℓ (see Vose & Wright, 1994, and Vose, 1996, the most comprehensive account is in Vose, 1998).

As explained in the first paper in this sequence (see Vose & Wright, 1998), it is the *Fourier* transform, not the Walsh transform, that is appropriate in the general cardinality case (i.e., when $c > 2$). However, when $c = 2$ the Fourier transform *is* the Walsh transform; by working with the Fourier transform in the body of this paper we therefore implicitly deal with the Walsh transform while simultaneously providing a framework that extends directly to higher cardinality alphabets. This paper explicitly deals with the Walsh transform by focusing on the binary case ($c = 2$) in the examples and the concrete results. The notation and several of the abstract results, however, will be stated in greater generality (for arbitrary c) to make plain how the analysis extends.

The reader is referred to part I (Vose & Wright, 1998) for relevant context, technical details, notation, and definitions. However, to orient the reader, the introduction does include a brief overview of selected mathematical objects, connecting them with their intended interpretations.

Part I demonstrated how mixing (mutation and crossover) is simplified by the Fourier transform. The operator $\mathcal{M}$ which encorporates the effects of mixing is referred to as the *mixing scheme* and is defined

[1]Part of this work was done while the second author was visiting the Computer Science Department of the University of Tennessee.

in terms of a *mixing matrix* M and a group of permutation matrices $\{\sigma_j\}$,

$$\mathcal{M}(x) \;= <\ldots, (\sigma_i x)^T M \sigma_i x, \ldots>$$

If p represents the current population ($p_i =$ the proportion of string i in the population), the mixing scheme satisfies

$$\mathcal{M}(p)_i \quad = \quad \Pr\{\ i \text{ results from mixing parents uniformly selected from } p\ \}$$

The operator $\mathcal{F}$ which encorporates the effects of selection is referred to as the *selection scheme* and satisfies

$$\mathcal{F}(p)_i \quad = \quad \Pr\{\ i \text{ is selected as a parent from } p\ \}$$

The composition of the selection and mixing schemes yeilds the simple genetic algorithm's *heuristic* function $\mathcal{G} = \mathcal{M} \circ \mathcal{F}$ which satisfies

$$\mathcal{G}(p)_i \quad = \quad \Pr\{\ i \text{ is contained in the next generation given current population } p\ \}$$

An alternate interpretation is that $\mathcal{G}(p)$ is the expected next generation, given current population p. As explained in (Vose, 1996) and in (Vose, 1998), the heuristic is a key component of the finite population model – a Markov chain – having transition probabilities given by the matrix

$$Q_{x,x'} \quad = \quad r! \prod_j \frac{\mathcal{G}(x)_j^{rx'_j}}{(rx'_j)!}$$

where r is the population size, and where the state space is the set of population vectors corresponding to populations of size r.

The Fourier matrix is defined by

$$W_{i,j} \quad = \quad c^{-\ell/2} e^{2\pi\sqrt{-1}\, i^T j}$$

and the Fourier transform $\widehat{v}$ of v is defined as

$$\begin{array}{rl} W x^C & \text{if } v \text{ is a column vector } x \\ W A^C W^C & \text{if } v \text{ is a matrix } A \\ y^C W^C & \text{if } v \text{ is a row vector } y \end{array}$$

where superscript C denotes complex conjugate. In the binary case ($c = 2$), all objects are real, the conjugation may be dispensed with, and the Walsh transform results. By using the columns of W as the basis $\mathcal{B}$ of the coordinate system in which to express how $\mathcal{M}$ transforms, $\mathcal{M}$ is put into a particularly managable form. Let the standard basis be $\{e_0, ..., e_{n-1}\}$ where e_j is the jth column of the identity matrix, and let

$$x \quad = \quad \sum x_j\, e_j$$

The kth component of $\mathcal{M}(x)$ with respect to the basis $\mathcal{B} = \{\widehat{e_0}, ..., \widehat{e_{n-1}}\}$ is

$$\sqrt{n} \sum_i \widehat{x}_i^{\,C} \widehat{x}_{k\ominus i}^{\,C} \widehat{M}_{i,i\ominus k}$$

where $\oplus$ is componentwise addition modulo c, $\ominus$ is componentwise subtraction modulo c, and $\otimes$ is componentwise multiplication modulo c. We begin in the next section from this point, making use of the notation, definitions, formulas, and results of part I (see Vose & Wright, 1998).

The first main result is obtained in Section 2. There a formula for $\mathcal{M}^{-1}$ is obtained. A consequence is that, for the simple genetic algorithm, the function $\mathcal{G}$ which maps the current population to the expected next generation is invertible.

Section 3 is concerned with characterizing, for arbitrary initial population p, the limiting behavior of iterated mixing

$$p,\ \mathcal{M}(p),\ \mathcal{M}^2(p),\ \ldots$$

This corresponds to the trajectory of expected next generations of a simple GA under uniform selection, crossover, and mutation. In the zero mutation case, our results include Geiringer's theorem.

2 The Inverse GA

The inverse of the function $\mathcal{G}$ which produces the expected next generation is obtained in this section (see Juliany & Vose, 1994 for a general introduction to $\mathcal{G}$ in the binary case). This provides the only known method for simulating a genetic algorithm backwards in time. Specifically, $\mathcal{G}^{-1}$ applied to a population p yeilds the unique population q which has the property that the expected next generation from q is p. As the population size increases, stochastic fluctuations die out, so the expectation can be dispensed with. That is, $\mathcal{G}(q)$ actually is the next generation from q and $\mathcal{G}^{-1}(p)$ actually is the generation previous to p, in the infinite population case.

Because $\mathcal{G}$ is defined as the composition

$$\mathcal{M} \circ \mathcal{F}$$

it is invertible provided both $\mathcal{M}$ and $\mathcal{F}$ are. In both part I and in this paper, the selection scheme is proportional, which has the consequence that inverting $\mathcal{F}$ is particularly easy (more general selection schemes have been inverted however, including ranking selection and tournament selection, see Vose, 1998). The challenging part is to invert the mixing scheme. This will be shown possible as a consequence of the following theorem which follows from theorem 4.3 of (Vose & Wright, 1998) and the discussion preceeding it:

Theorem 2.1 *The k th component of $y = \mathcal{M}(x)$ with respect to the basis $\mathcal{B} = \{\widehat{e_0}, ..., \widehat{e_{n-1}}\}$ is*

$$\widehat{y}_k^C \;=\; \sqrt{n}\sum_i \widehat{x}_i^C \widehat{x}_{k\ominus i}^C \widehat{M}_{i,i\ominus k}$$

Before proceeding to invert $\mathcal{M}$, an example of computing $\widehat{y}$ via the formula given by theorem 2.1 will be given. Begin with the population {01, 02, 10, 12, 22} where $\ell = 2$ and $c = 3$. This population has corresponding population vector

$$x \;=\; <0, \tfrac{1}{5}, \tfrac{1}{5}, \tfrac{1}{5}, 0, \tfrac{1}{5}, 0, 0, \tfrac{1}{5}>$$

The conjugate Fourier transform of x is

$$\widehat{x}^C \;=\; <\tfrac{1}{3}, \tfrac{-1+\sqrt{-3}}{15}, \tfrac{-1-\sqrt{-3}}{15}, \tfrac{1-\sqrt{-3}}{30}, \tfrac{-1-\sqrt{-3}}{15}, \tfrac{-1}{15}, \tfrac{1+\sqrt{-3}}{30}, \tfrac{-1}{15}, \tfrac{-1+\sqrt{-3}}{15}>$$

The mutation vector

$$\mu \;=\; <\tfrac{49}{64}, \tfrac{7}{128}, \tfrac{7}{128}, \tfrac{7}{128}, \tfrac{1}{256}, \tfrac{1}{256}, \tfrac{7}{128}, \tfrac{1}{256}, \tfrac{1}{256}>$$

corresponds to a mutation rate of 1/8, and the crossover vector

$$\chi \quad = \quad <\tfrac{1}{4}, \tfrac{1}{4}, 0, \tfrac{1}{4}, \tfrac{1}{4}, 0, 0, 0, 0>$$

corresponds to a one-point crossover rate of 1/2 (or, equivalently, to a uniform crossover rate of 1). The mixing matrix M corresponding to these mutation and crossover vectors is

$$M \quad = \quad \begin{bmatrix} \frac{49}{64} & \frac{105}{256} & \frac{105}{256} & \frac{105}{256} & \frac{225}{1024} & \frac{225}{1024} & \frac{105}{256} & \frac{225}{1024} & \frac{225}{1024} \\ \frac{105}{256} & \frac{7}{128} & \frac{7}{128} & \frac{225}{1024} & \frac{15}{512} & \frac{15}{512} & \frac{225}{1024} & \frac{15}{512} & \frac{15}{512} \\ \frac{105}{256} & \frac{7}{128} & \frac{7}{128} & \frac{225}{1024} & \frac{15}{512} & \frac{15}{512} & \frac{225}{1024} & \frac{15}{512} & \frac{15}{512} \\ \frac{105}{256} & \frac{225}{1024} & \frac{225}{1024} & \frac{7}{128} & \frac{15}{512} & \frac{15}{512} & \frac{7}{128} & \frac{15}{512} & \frac{15}{512} \\ \frac{225}{1024} & \frac{15}{512} & \frac{15}{512} & \frac{15}{512} & \frac{1}{256} & \frac{1}{256} & \frac{15}{512} & \frac{1}{256} & \frac{1}{256} \\ \frac{225}{1024} & \frac{15}{512} & \frac{15}{512} & \frac{15}{512} & \frac{1}{256} & \frac{1}{256} & \frac{15}{512} & \frac{1}{256} & \frac{1}{256} \\ \frac{105}{256} & \frac{225}{1024} & \frac{225}{1024} & \frac{7}{128} & \frac{15}{512} & \frac{15}{512} & \frac{7}{128} & \frac{15}{512} & \frac{15}{512} \\ \frac{225}{1024} & \frac{15}{512} & \frac{15}{512} & \frac{15}{512} & \frac{1}{256} & \frac{1}{256} & \frac{15}{512} & \frac{1}{256} & \frac{1}{256} \\ \frac{225}{1024} & \frac{15}{512} & \frac{15}{512} & \frac{15}{512} & \frac{1}{256} & \frac{1}{256} & \frac{15}{512} & \frac{1}{256} & \frac{1}{256} \end{bmatrix}$$

The Fourier transform of the mixing matrix is

$$\widehat{M} \quad = \quad \begin{bmatrix} 1 & \frac{13}{32} & \frac{13}{32} & \frac{13}{32} & \frac{169}{1024} & \frac{169}{1024} & \frac{13}{32} & \frac{169}{1024} & \frac{169}{1024} \\ \frac{13}{32} & 0 & 0 & \frac{169}{1024} & 0 & 0 & \frac{169}{1024} & 0 & 0 \\ \frac{13}{32} & 0 & 0 & \frac{169}{1024} & 0 & 0 & \frac{169}{1024} & 0 & 0 \\ \frac{13}{32} & \frac{169}{1024} & \frac{169}{1024} & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{169}{1024} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{169}{1024} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{13}{32} & \frac{169}{1024} & \frac{169}{1024} & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{169}{1024} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{169}{1024} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Computing $\widehat{y}^C$ via theorem 2.1 shows it to be

$$<\tfrac{1}{3}, \tfrac{-13+13\sqrt{-3}}{240}, \tfrac{-13-13\sqrt{-3}}{240}, \tfrac{13-13\sqrt{-3}}{480}, \tfrac{-169-169\sqrt{-3}}{9600}, \tfrac{-1183}{38400}, \tfrac{13+13\sqrt{-3}}{480}, \tfrac{-1183}{38400}, \tfrac{-169+169\sqrt{-3}}{9600}>$$

The details of the computation of $\widehat{y}_7^C$ are given by

$$\begin{aligned} \widehat{y}_7^C &= 3\left(\widehat{x}_0^C \widehat{x}_7^C \widehat{M}_{0,5} + \widehat{x}_1^C \widehat{x}_6^C \widehat{M}_{1,3} + \widehat{x}_6^C \widehat{x}_1^C \widehat{M}_{6,2} + \widehat{x}_7^C \widehat{x}_0^C \widehat{M}_{7,0}\right) \\ &= 3\left(-\tfrac{169}{46080} - \tfrac{169}{115200} - \tfrac{169}{115200} - \tfrac{169}{46080}\right) \\ &= -\tfrac{1183}{38400}. \end{aligned}$$

Theorem 2.1 effectively "triangulates" the equations which relate one generation to the next. The key observation is that $\widehat{M}_{i,i\ominus k} = 0$ unless $i^T(i \ominus k) = 0$ (this follows from theorem 3.2 of Vose & Wright, 1998). The condition $i^T(i \ominus k) = 0$ means that at every component where i differs from k, it must be zero. Therefore, the nonzero terms in the summation above correspond to elements i of Ω_k which agree with k at every nonzero component. In particular, the k th component of $\mathcal{M}(x)$ with respect to the basis $\mathcal{B}$ is determined by objects subscripted by elements of Ω_k. The "triangularization" is obtained by ordering the k th components by $\#k$.

The proof of the next theorem will illustrate this principle more formally. Let Ω'_k denote $\Omega_k \setminus \{0, k\}$. It is only for notational convenience that the summation in theorem 2.2 below is taken to be over Ω'_k. As noted above, nonzero terms actually correspond to objects subscripted by special elements of Ω_k (those agreeing with k at nonzero components). In particular, if $i \in \Omega'_k$ and k is nonzero, then neither i nor $k \ominus i$ can be k, and therefore both $\#i < \#k$ and $\#(k \ominus i) < \#k$ provided $i^T(i \ominus k) = 0$.

Theorem 2.2 *If M^* is invertible, then $\mathcal{M}$ has an inverse on Λ given by the map $y \mapsto x$ which is defined by the following recursion:*

$$\widehat{x}_k^C = \begin{cases} \widehat{y}_k^C & \text{if } k = 0 \\ (2\widehat{M}_{k,0})^{-1}(\widehat{y}_k^C - \sqrt{n} \sum_{i \in \Omega'_k} \widehat{x}_i^C \widehat{x}_{k \ominus i}^C \widehat{M}_{i, i \ominus k}) & \text{if } k > 0 \end{cases}$$

Sketch of proof: If $x \in \Lambda$, the vector $y = \mathcal{M}(x)$ satisfies

$$\widehat{y}_k^C = \begin{cases} \widehat{x}_k^C & \text{if } k = 0 \\ 2\widehat{M}_{k,0}\, \widehat{x}_k^C + \sqrt{n} \sum_{i \in \Omega'_k} \widehat{x}_i^C \widehat{x}_{k \ominus i}^C \widehat{M}_{i, i \ominus k} & \text{if } k > 0 \end{cases}$$

This relationship follows directly from theorem 2.1, the fact that $\widehat{M}_{0,-k} = \widehat{M}_{k,0}$ (since M is symmetric), and the observation that $\widehat{x}_0 = 1/\sqrt{n}$ for all $x \in \Lambda$. Solving for $\widehat{x}_k^C$ gives the recursion displayed in theorem 2.2. As observed in the discussion preceeding theorem 2.2, if $k > 0$ then elements $i \in \Omega'_k$ corresponding to nonzero terms in the sum (i.e., satisfying $i^T(i \ominus k) = 0$) are such that both $\#i < \#k$ and $\#(k \ominus i) < \#k$. Therefore, the recursion terminates at $\widehat{x}_0^C$. This recursion, and hence the inverse, is well defined provided division by zero is avoided, i.e., provided $\widehat{M}_{k,0} \neq 0$. Theorem 3.12 of (Vose & Wright, 1998) shows the 0 th column of $\widehat{M}$ is the spectrum of M^*. Hence the invertibility of $\mathcal{M}$ is equivalent to the invertibility of M^*. □

Suppose crossover can be described in terms of masks – such as one-point, n-point, or uniform crossover (see Vose & Wright, 1994 for the technical definition of crossover masks). If the crossover rate is less than 1, it follows that the crossover distribution is such that $\chi_0 > 0$. The following corollary relates this conclusion to the invertibility of $\mathcal{M}$.

Corollary 2.3 *If mutation is given by a mutation rate $0 < \mu < 1$ satisfying $\mu \neq 1 - 1/c$, and if crossover is such that $\chi_0 > 0$, then $\mathcal{M}^{-1}$ exists.*

Sketch of proof: According to theorem 2.2 and theorem 3.12 of (Vose & Wright, 1998), it suffices that $\widehat{M}_{k,0}$ is nonzero for all k. Since $\widehat{\mu}_0 = 1/\sqrt{n}$, it follows from theorem 3.2 of (Vose & Wright, 1998) that

$$\widehat{M}_{k,0} = \frac{\sqrt{n}}{2}\, \widehat{\mu}_{-k} \sum_j (\chi_j + \chi_{\bar{j}})\, [k \otimes j = 0]$$

The summation above is positive given $\chi_0 > 0$. Hence it suffices that $\widehat{\mu}_k$ is nonzero for all k. Note that

$$\begin{aligned}\sqrt{n}\,\widehat{\mu}_k &= \sum_j e(j^T k)(\mu/(c-1))^{\#j}\,(1-\mu)^{\ell-\#j} \\ &= (1-\mu)^\ell \sum_j (\mu/((1-\mu)(c-1)))^{\#j}\,e(j^T k)\end{aligned}$$

Ignoring the $(1-\mu)^\ell$ factor, this has the form

$$\sum_j x^{\#j}\,e(j^T k) \;=\; \sum_j x^{\sum \#j_i}\,e(\textstyle\sum j_i k_i) \;=\; \prod_i \sum_{j_i} x^{\#j_i}\,e(j_i k_i)$$

Note that if $k_i = 0$, then the inner sum is nonzero. Assume therefore that $k_i > 0$, and note that in this case the sum is

$$1 + x(\sum_{j_i} e(j_i k_i) \;-\; 1) \;=\; 1 - x$$

This is nonzero provided $x \neq 1$, which is equivalent to $\mu \neq 1 - 1/c$. □

Since $\mathcal{G} = \mathcal{M} \circ \mathcal{F}$, corollary 2.3 may be brought to bear on the invertibility of $\mathcal{G}$. Starting with the equation $y = \mathcal{F}(x) = Fx/\mathbf{1}^T Fx$ corresponding to proportional selection, it follows that y is simply Fx when magnitude is ignored. Therefore the inverse to $\mathcal{F}$ on Λ is

$$x \;\mapsto\; F^{-1}x \,/\, \mathbf{1}^T F^{-1}x$$

Appealing to corollary 2.3 and the paragraph before it, the inverse of $\mathcal{G}$ exists and is given by

$$\mathcal{G}^{-1} \;=\; \mathcal{F}^{-1} \circ \mathcal{M}^{-1}$$

provided that the selection scheme is proportional, the crossover rate is less than 1, and the mutation rate $0 < \mu < 1$ satisfies $\mu \neq 1 - 1/c$.

Although it is beyond the scope of this paper, Vose has obtained formalisms for selection schemes beyond proportional selection. In particular, ranking and tournament selection can be dealt with as well (see Vose, 1998). Moreover, $\mathcal{F}^{-1}$ is known to exist in both of those cases and an algorithm to compute it has been obtained. Therefore, the invertibility of $\mathcal{G}$ is known in a wider context than that corresponding to proportional selection.

The invertibility of $\mathcal{G}$ is an important mathematical fact which contributes to the body of knowledge known about the simple genetic algorithm, allowing the genetic algorithm to be run "backwards", in some sense, from one generation to the previous. The exploration of this possibility – evolution backwards in time – was initiated in the paper "The Genetic Algorithm Fractal" for the two bit binary case (see Juliany & Vose, 1994). Interestingly, fractal structures were discovered near a point which appears to generically be the negative time limit point of evolution, i.e.,

$$\lim_{t\to\infty} \mathcal{G}^{-t}(p)$$

For a cursory discussion of $\mathcal{G}^{-1}$ from this perspective, as well as color figures depecting the fractal structures discovered, see (Juliany & Vose, 1994).

3 Recombination Limits

This section considers the effect of iterated mixing on an arbitrary initial population x. The recombination equations given by theorem 2.1 allow the fixed points of $\mathcal{M}$ to be determined explicitly. In fact, the triangular form can be used to also determine when the limit of the sequence

$$x,\ \mathcal{M}(x),\ \mathcal{M}^2(x),\ \ldots$$

exists. An important preliminary result is

Lemma 3.1 *If $z > 0$ then $|\widehat{M}_{0,z}| \leq 1/2$. When equality holds and $z = x \ominus y$ for nonzero x and y, then $\widehat{M}_{x,y} = 0$.*

Sketch of proof: By theorem 3.2 of (Vose & Wright, 1998),

$$\widehat{M}_{-u,v} \;=\; [u^T v = 0]\,\frac{n}{2}\,\widehat{\mu}_u\,\widehat{\mu}_v \sum_k (\chi_k + \chi_{\overline{k}})\,[u \otimes \overline{k} = v \otimes k = 0]$$

Since $|[u^T v = 0]\,\frac{n}{2}\,\widehat{\mu}_u\,\widehat{\mu}_v| \leq 1/2$, and since we are interested in the case $u = 0$ and $v = z$, the first part of the lemma is satisfied if

$$1 \;\geq\; \sum_k (\chi_k + \chi_{\overline{k}})\,[z \otimes k = 0]$$

But $k \in \Omega_{\overline{z}}$ in the sum above, so the terms indexed by k and $\overline{k}$ are disjoint (since $z > 0$). Hence the sum is bounded by the sum over all χ_j, which is 1. The second part of the lemma therefore requires every nonzero χ_j to occur in the sum above so that its value is 1. Moreover, the conclusion of the second part is trivially satisfied unless $x^T y = 0$, and in that case $\Omega_{\overline{z}} = \Omega_{\overline{x \ominus y}} = \Omega_{\overline{x} \otimes \overline{y}}$. We may assume therefore that

$$1 \;=\; \sum_k (\chi_k + \chi_{\overline{k}})\,[x \otimes k = y \otimes k = 0]$$

Thus the proof would be complete (i.e., $\widehat{M}_{x,y} = 0$) if this implied that

$$0 \;=\; \sum_k (\chi_k + \chi_{\overline{k}})\,[x \otimes \overline{k} = y \otimes k = 0]$$

Since $\chi_k > 0 \Longrightarrow x \otimes k = y \otimes k = 0$ (because every nonzero χ_j occurs in the sum equaling 1), it follows that $\chi_k > 0 \Longrightarrow x \otimes \overline{k} \neq 0$ (because $x > 0$). Moreover if $k = \overline{j}$, then $y \otimes \overline{j} = 0$ and hence $y \otimes j \neq 0$ (because $y > 0$). Hence nonzero χ_k do not occur in the sum above (which corresponds to $\widehat{M}_{x,y}$). The same observation applies to $\chi_{\overline{k}}$ (i.e., nonzero $\chi_{\overline{k}}$ do not occur in the sum). Since the sum contains only zero terms, it must be zero. □

Corollary 3.2 *If $z > 0$ then $|\widehat{M}_{z,0}| \leq 1/2$. When equality holds and $z = x \ominus y$ for nonzero x and y, then $\widehat{M}_{x,y} = 0$.*

Sketch of proof: Since M is symmetric, $\widehat{M}_{z,0} = \widehat{M}_{0,-z}$. Moreover, $z > 0 \Longrightarrow -z > 0$. Hence appealing to lemma 3.1 establishes the first part of the proof. Since M is real, $M^\wedge = M^{H\wedge} = M^{\wedge H}$, thus $\widehat{M}_{x,y} = \widehat{M}_{y,x}^C$. If $|\widehat{M}_{z,0}| = 1/2$ and $z = x \ominus y$, then $|\widehat{M}_{0,-z}| = 1/2$ and $-z = y \ominus x$. Appealing to lemma 3.1 once more yeilds

$$0 = \widehat{M}_{y,x} = \widehat{M}_{y,x}^C = \widehat{M}_{x,y}$$

□

Next recall that $y = \mathcal{M}(x)$ satisfies

$$\widehat{y}_k^C = \begin{cases} \widehat{x}_k^C & \text{if } k = 0 \\ 2\widehat{M}_{k,0}\,\widehat{x}_k^C + \sqrt{n}\sum_{i\in\Omega'_k} \widehat{x}_i^C \widehat{x}_{k\ominus i}^C \widehat{M}_{i,i\ominus k} & \text{if } k > 0 \end{cases}$$

which was noted in the proof of theorem 2.2. This can be put in the abbreviated form

$$\widehat{y}_k = \begin{cases} \widehat{x}_k & \text{if } k = 0 \\ \alpha_k \widehat{x}_k + \beta_k & \text{if } k > 0 \end{cases}$$

It is easy to see that $\lim_j \mathcal{M}^j(x)$ exists by considering k th components in order of increasing $\#k$ as follows. But first note when $\#k = 1$ the sum β_k is empty (because then $\Omega_k = \{0, k\}$), hence zero. Also, corollary 3.2 implies that $|\alpha_k| \leq 1$ and that $|\alpha_k| = 1 \Longrightarrow \beta_k = 0$ provided $\#k > 1$ (because of the factors $\widehat{M}_{i,i\ominus k}$ in the terms of β_k). Moreover, if $|\alpha_k| = 1$, then α_k must be a c th root of unity, which can be seen as follows. From theorem 3.2 of (Vose & Wright, 1998) it follows that

$$\alpha_k^C = \sqrt{n}\,\widehat{\mu}_k \sum_j (\chi_j + \chi_{\bar{j}})\,[k \otimes j = 0]$$

Thus α_k^C points in the direction of $\widehat{\mu}_k$ and can have – by corollary 3.2 – modulus one only if there is no cancellation in the sum

$$\widehat{\mu}_k = \sqrt{n}\sum_i \mu_i\, e(i^T k)$$

Thus cancellation is avoided only when every nonzero μ_i occurs as a coefficient of the same c th root of unity.

With these preliminary observations out of the way, we now turn to the behavior of the sequence x, $\mathcal{M}(x)$, $\mathcal{M}^2(x)$, ... Clearly, when $\#k = 0$, the k th component of $\lim_j \mathcal{M}^j(x)$ has already converged since then $\widehat{y}_0 = \widehat{x}_0$.

Next, for $\#k = 1$, if $|\alpha_k| < 1$, the k th component converges to 0 because the relation between successive iterations is $\widehat{y}_k = \alpha_k \widehat{x}_k$. When $\alpha_k = 1$ then $\widehat{y}_k = \widehat{x}_k$ so the k th component has converged. If α_k is any other c th root of unity, then the k th component is periodic, with period some divisor of c. In that case considering $\mathcal{M}^c$ in place of $\mathcal{M}$ restores convergence since then $\widehat{y}_k = \alpha_k^c \widehat{x}_k = \widehat{x}_k$.

In the general case of $\#k > 1$, the sum β_k may be treated as having already converged because it involves components subscripted by j with $\#j < \#k$. If components are periodic for some values of j, considering $\mathcal{M}^c$ in place of $\mathcal{M}$ restores convergence. If $\alpha_k = 1$ then $\widehat{y}_k = \widehat{x}_k$ (since $\beta_k = 0$) so the k th component has already converged. If α_k is any other c th root of unity, then the k th component is periodic having period some divisor of c for $\mathcal{M}$, but it has already converged for $\mathcal{M}^c$.

When $|\alpha_k| < 1$, the kth component converges to $\beta_k/(1-\alpha_k)$ since the relation between successive iterations is $\widehat{y}_k = \alpha_k \widehat{x}_k + \beta_k$.

This discussion is summarized by the following theorem.

Theorem 3.3 *The function $x \mapsto y$ defined recursively, via Walsh coordinates, by*

$$\widehat{y}_k^C = \begin{cases} \widehat{x}_k^C & \text{if } k = 0 \text{ or } |\widehat{M}_{k,0}| = 1/2 \\ \sqrt{n}(1-2\widehat{M}_{k,0})^{-1} \sum\limits_{i \in \Omega'_k} \widehat{y}_i^C \widehat{y}_{k \ominus i}^C \widehat{M}_{i,i\ominus k} & \text{otherwise} \end{cases}$$

produces the fixed point $y = \lim_j \mathcal{M}^{cj}(x)$ of $\mathcal{M}^c$. If $|\widehat{M}_{k,0}| = 1/2 \Longrightarrow \widehat{M}_{k,0} = 1/2$, then y is also the fixed point $\lim_j \mathcal{M}^j(x)$ *of $\mathcal{M}$. Otherwise, $\mathcal{M}^j(x)$ converges to a periodic orbit having period a divisor d of c and elements $\{\mathcal{M}^i(y) : 0 \le i < d\}$.*

As an example, consider the population vector x (given earlier) whose conjugate Fourier transform is

$$\widehat{x}^C = < \tfrac{1}{3}, \tfrac{-1+\sqrt{-3}}{15}, \tfrac{-1-\sqrt{-3}}{15}, \tfrac{1-\sqrt{-3}}{30}, \tfrac{-1-\sqrt{-3}}{15}, \tfrac{-1}{15}, \tfrac{1+\sqrt{-3}}{30}, \tfrac{-1}{15}, \tfrac{-1+\sqrt{-3}}{15} >$$

The mixing matrix for zero mutation and a one-point crossover rate of 1 is

$$M = \widehat{M} = \begin{bmatrix} 1 & 1/2 & 1/2 & 1/2 & 1/4 & 1/4 & 1/2 & 1/4 & 1/4 \\ 1/2 & 0 & 0 & 1/4 & 0 & 0 & 1/4 & 0 & 0 \\ 1/2 & 0 & 0 & 1/4 & 0 & 0 & 1/4 & 0 & 0 \\ 1/2 & 1/4 & 1/4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 1/4 & 1/4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The conjugate Fourier transform of the fixed point as given by theorem 3.3 is

$$\widehat{y}^C = < \tfrac{1}{3}, \tfrac{-1+\sqrt{-3}}{15}, \tfrac{-1-\sqrt{-3}}{15}, \tfrac{1-\sqrt{-3}}{30}, \tfrac{1+\sqrt{-3}}{75}, \tfrac{-2}{75}, \tfrac{1+\sqrt{-3}}{30}, \tfrac{-2}{75}, \tfrac{1-\sqrt{-3}}{75} >$$

Combining theorem 3.3 with theorem 3.12 of (Vose & Wright, 1998), we see that the fixed points of $\mathcal{M}$ form a manifold of dimension equal to the sum of the multiplicities that cth roots of unity have as eigenvalues of $2M^*$. In particular, if the spectrum of $2M^*$ does not contain cth roots of unity (as is the case when mutation is positive), then $\mathcal{M}$ has the unique fixed point $\mathbf{1}/\sqrt{n}$. Although it is beyond the scope of this paper, Vose has proved $\mathcal{M}$ is a contraction mapping in that case (see Vose, 1998).

Next, the fixed points of $\mathcal{M}$ in the special case of zero mutation are characterized. Note in this case that $\widehat{M}_{k,0} \ge 0$ (it is a sum of non-negative terms), hence convergence to a periodic orbit is not possible, and therefore the y of theorem 3.3 is the fixed point $\lim_j \mathcal{M}^j(x)$ of $\mathcal{M}$ corresponding to x.

Lemma 3.4 *If mutation is zero, then for every k,*

$$\sum_{i \in \Omega'_k} \widehat{M}_{i,i\ominus k} \;=\; 1 - 2\widehat{M}_{k,0}$$

Sketch of proof: By corollary 3.3 of (Vose & Wright, 1998), $M = \widehat{M}$ (since mutation is zero). Note that

$$\begin{aligned}
\sum_{i \in \Omega_k} M_{i,i\ominus k} &= \sum_{i \in \Omega_k} \sum_u \frac{\chi_u + \chi_{\overline{u}}}{2} [i \otimes u \oplus \overline{u} \otimes (i \ominus k) = 0] \\
&= \sum_{i \in \Omega_k} \sum_u \frac{\chi_u + \chi_{\overline{u}}}{2} [i \ominus \overline{u} \otimes k = 0] \\
&= \sum_u \frac{\chi_u + \chi_{\overline{u}}}{2} \sum_{i \in \Omega_k} [i = \overline{u} \otimes k] \\
&= \sum_u \frac{\chi_u + \chi_{\overline{u}}}{2}
\end{aligned}$$

Therefore

$$1 \;=\; \sum_{i \in \Omega_k} M_{i,i\ominus k} \;=\; 2M_{k,0} + \sum_{i \in \Omega'_k} M_{i,i\ominus k}$$

which finishes the proof. □

Let $B_k = \{i \in \Omega_k \,:\, \#i = 1 \;\wedge\; i^T(k \ominus i) = 0\}$. The previous lemma, together with theorem 3.3, leads to the following theorem.

Theorem 3.5 *Assume zero mutation, and let y be the limit of x, $\mathcal{M}(x)$, $\mathcal{M}^2(x)$, … Suppose further that $\#k > 1 \Longrightarrow \widehat{M}_{k,0} < 1/2$. Then*

$$\widehat{y}_k = c^{(\#k-1)\ell/2} \prod_{j \in B_k} \widehat{x}_j$$

Sketch of proof: Induct on $\#k$. The base case is $\#k \leq 1$. Consider first $\#k = 0$, so $\widehat{y}_k = 1/\sqrt{n}$. Moreover,

$$c^{(\#k-1)\ell/2} \prod_{j \in B_k} \widehat{x}_j \;=\; c^{-\ell/2} \prod_{j \in \emptyset} \widehat{x}_j \;=\; 1/\sqrt{n}$$

When $\#k = 1$, appealing to lemma 3.4 gives

$$1 - 2\widehat{M}_{k,0} \;=\; \sum_{i \in \Omega'_k} \widehat{M}_{i,i\ominus k} \;=\; \sum_{i \in \emptyset} \widehat{M}_{i,i\ominus k} \;=\; 0$$

Hence $\widehat{M}_{k,0} = 1/2$, which by theorem 3.3 implies $\widehat{y}_k = \widehat{x}_k$. Moreover,

$$c^{(\#k-1)\ell/2} \prod_{j \in B_k} \widehat{x}_j \;=\; \prod_{j \in \{k\}} \widehat{x}_j \;=\; \widehat{x}_k$$

For the inductive case, assume the result for all $\widehat{y}_j$ such that $\#j < \#k$. Let $i \in \Omega'_k$ be such that $i^T(k \ominus i) = 0$. Note that $\#i < \#k$ and $\#(k \ominus i) < \#k$. By the induction hypothesis,

$$\begin{aligned}
\widehat{y}_i \widehat{y}_{k\ominus i} &= \left(c^{(\#i-1)\ell/2} \prod_{u \in B_i} \widehat{x}_u \right) \left(c^{(\#(k\ominus i)-1)\ell/2} \prod_{v \in B_{k\ominus i}} \widehat{x}_v \right) \\
&= c^{(\#i+\#(k\ominus i)-2)\ell/2} \prod_{j \in B_k} \widehat{x}_j \\
&= c^{(\#k-2)\ell/2} \prod_{j \in B_k} \widehat{x}_j
\end{aligned}$$

Applying theorem 3.3 ($\widehat{M}_{k,0} < 1/2$ by hypothesis and $\widehat{M} = M$ is real) gives

$$\widehat{y}_k = \sqrt{n}\,(1 - 2\widehat{M}_{k,0})^{-1} \sum_{i \in \Omega'_k} c^{(\#k-2)\ell/2} \widehat{M}_{i,i\ominus k} \prod_{j \in B_k} \widehat{x}_j$$

Using $(1 - 2\widehat{M}_{k,0})^{-1} \sum_{i \in \Omega'_k} \widehat{M}_{i,i\ominus k} = 1$, which follows from Lemma 3.4, completes the proof. □

Corollary 3.6 *Let $u, v \in \Omega_k$ be such that $k = u \oplus v$ and $u^T v = 0$. If $\#h > 1 \Longrightarrow \widehat{M}_{h,0} < 1/2$, then $\widehat{y}_u \widehat{y}_v = \widehat{y}_k / \sqrt{n}$.*

Sketch of proof: The result follows from the proof of theorem 3.5. If $\#k \leq 1$, the result follows from the proof of the base step. When $\#k > 1$, take $i = u$, and hence $k \ominus i = v$, in the proof of the inductive step. □

A consequence of theorem 3.5 is that the space of fixed points under crossover-only mixing is homeomorphic to the ℓ-fold cartesian product of $c - 1$ dimensional simplices. A homeomorphism from

$$\{ < \alpha_0, \ldots, \alpha_{\ell-1} > : \ \alpha_u \in \Re^c \ \wedge \ (\alpha_u)_v \geq 0 \ \wedge \ \sum_v (\alpha_u)_v = 1 \}$$

into the space of fixed points is given by

$$\widehat{y}_k = c^{(\#k-1)\ell/2} \prod_{j \in B_k} \widehat{x}_j$$

where

$$\widehat{x}_j = \frac{[\,\#j = 1\,]}{\sqrt{n}} \sum_{u=0}^{\ell-1} [\, j_u > 0\,] \sum_{i=0}^{c-1} e(j_u i)\,(\alpha_u)_i$$

The space of fixed points includes every vertex of Λ, as well as the convex hull of vertices corresponding to strings that differ only at a fixed component (a $c-1$ dimensional simplex). This result is dependent on the condition that $\#k > 1 \Longrightarrow M_{k,0} < 1/2$.

Some of the previous paragraph might not seem obvious; why, for example, is the convex hull mentioned above fixed, and why is the domain of the homeomorphism the ℓ-fold cartesian product of $c - 1$

dimensional simplices? The latter fact follows from the observation that the proportion $(\alpha_u)_i$ of the population (described by) x having i in position u is

$$\begin{aligned}(\alpha_u)_i &= \sum_j [\, j^T c^u = i \,]\, x_j \\ &= \frac{1}{c} \sum_j \sum_{k=0}^{c-1} e(k(j^T c^u - i))\, x_j \\ &= \frac{1}{c} \sum_{k=0}^{c-1} e(-ki) \sum_j e(k\, j^T c^u)\, x_j \\ &= \frac{\sqrt{n}}{c} \sum_{k=0}^{c-1} e(-ki)\, \widehat{x}_{kc^u}\end{aligned}$$

Since the α_u are independent, and since each α_u varies over a $c-1$ dimensional simplex (because $(\alpha_u)_0 + \cdots + (\alpha_u)_{c-1} = 1$), these equations implicitly determine $\widehat{x}_{kc^u}$ as a function over the stated domain. Solving for $\widehat{x}_{kc^u}$ may be accomplished as follows,

$$\begin{aligned}\frac{1}{\sqrt{n}} \sum_{i=0}^{c-1} e(ki)\, (\alpha_u)_i &= \frac{1}{c} \sum_{i=0}^{c-1} e(ki) \sum_{h=0}^{c-1} e(-hi)\, \widehat{x}_{hc^u} \\ &= \frac{1}{c} \sum_{h=0}^{c-1} \widehat{x}_{hc^u} \sum_{i=0}^{c-1} e(i(k-h)) \\ &= \sum_{h=0}^{c-1} \widehat{x}_{hc^u}\, [\, k = h \,] \\ &= \widehat{x}_{kc^u}\end{aligned}$$

Moreover, theorem 3.5 determines the $\widehat{y}_u$ in terms of the $\widehat{x}_{kc^u}$. It is not difficult to see that the convex hull mentioned above (determined by vertices vertices corresponding to strings that differ only at a fixed component) is fixed. Children of such strings are identical to their parents, and $\mathcal{M}$ acts as the identity on populations comprised of such strings (alternatively, theorem 3.5 may be used directly to calculate that the edges are fixed).

We next show that theorem 3.5 implies a result of H. Geiringer (1944), who showed that, except in the case of "complete linkage", the distribution of genotypes approaches independence under repeated recombination. In other words, the limiting frequency of string k is the product of the frequencies of the components of k in the initial population. Of course, this is strictly a zero mutation result. In the general case of nonzero mutation, Geiringer's result is false and it is theorem 3.3, not Geiringer's theorem, which determines asymptotic behavior.

Two loci are *completely linked* if crossover cannot separate the alleles at these two loci. In other words, two loci are completely linked if whenever j has differing bits at the loci then $\chi_j = 0$.

Proposition 3.7 *The loci at positions u and v are completely linked if and only if $M_{c^u \oplus c^v, 0} = 1/2$.*

Sketch of proof: Suppose loci u and v are not completely linked and let $k = c^u \oplus c^v$. Then there exists a crossover mask j with $\chi_j > 0$ such that $j \otimes k \neq 0$ and $\overline{j} \otimes k \neq 0$. Note that

$$\begin{aligned} M_{k,0} &= \sum_i \frac{\chi_i + \chi_{\overline{i}}}{2} [\, i \otimes k = 0\,] \\ &= \sum_i \frac{[\, i \otimes k = 0\,] + [\, \overline{i} \otimes k = 0\,]}{2} \chi_i \end{aligned}$$

Note that χ_j has a zero coefficient, and the conditions $[\, i \otimes k = 0\,]$ and $[\, \overline{i} \otimes k = 0\,]$ are mutually exclusive. It follows that $M_{0,k} < 1/2$. Moreover, the argument is reversible. □

Corollary 3.8 *If mutation is zero and no loci are completely linked, then* $\#k > 1 \Longrightarrow \widehat{M}_{k,0} < 1/2$.

Sketch of proof: Since mutation is zero, $M = \widehat{M}$. Suppose $\#k > 1$, and let $u \neq v$ be such that $\{c^u, c^v\} \subset \Omega_k$. The corollary is a consequence of corollary 3.2, proposition 3.7, and the observation that

$$\begin{aligned} M_{k,0} &= \sum_i \frac{\chi_i + \chi_{\overline{i}}}{2} [\, i \otimes k = 0\,] \\ &\leq \sum_i \frac{\chi_i + \chi_{\overline{i}}}{2} [\, i \otimes (c^u \oplus c^v) = 0\,] \\ &= M_{c^u \oplus c^v, 0} \end{aligned}$$

The inequality above follows from the implication $[i \otimes k = 0] \Longrightarrow [i \otimes (c^u \oplus c^v) = 0]$. □

Given initial population x, let $0 \leq u < \ell$, and let $j \in \Omega$. Following Geiringer, define

$$p_u(j) = \sum_k x_k [k \otimes c^u = j \otimes c^u]$$

Thus $p_u(j)$ is the proportion of strings in population x whose uth component agrees with the uth component of j. According to the discussion following corollary 3.6,

$$p_u(j) = (\alpha_u)_{j_u}$$

Moreover,

$$\begin{aligned} p_u(i\,\mathbf{1}) &= (\alpha_u)_i^C \\ &= \frac{\sqrt{n}}{c} \sum_{k=0}^{c-1} e(ki)\, \widehat{x}_{kc^u}^C \\ &= \frac{\sqrt{n}}{c} \sum_{h \in \Omega_{c^u}} e(i \mathbf{1}^T h)\, \widehat{x}_h^C \end{aligned}$$

Note that i as used above is a scalor, and therefore $i\,\mathbf{1} = < i, \ldots, i >$.

Theorem 3.9 *(Geiringer) Let mutation be zero, let* $x \in \Lambda$ *and let* $y = lim_j \mathcal{M}^j(x)$. *If no pair of loci are completely linked, then*

$$y_k = \prod_{j=0}^{\ell-1} p_j(k)$$

Sketch of proof: Beginning with the Fourier transform, and using corollary 3.6,

$$\begin{aligned}
\sqrt{n}\, y_k &= \sum_i e(k^T i)\, \widehat{y}_i^C \\
&= \sum_{u_0 \in \Omega_{c0}} \cdots \sum_{u_{\ell-1} \in \Omega_{c\ell-1}} e(k^T(u_0 \oplus \cdots \oplus u_{\ell-1}))\, \widehat{y}^C_{u_0 \oplus \cdots \oplus u_{\ell-1}} \\
&= \sum_{u_0 \in \Omega_{c0}} \cdots \sum_{u_{\ell-1} \in \Omega_{c\ell-1}} e(k^T u_0)\, \cdots\, e(k^T u_{\ell-1})\, \widehat{y}^C_{u_0} \cdots \widehat{y}^C_{u_{\ell-1}}\, \sqrt{n}^{\,\ell-1}
\end{aligned}$$

Therefore

$$y_k = \prod_{j=0}^{\ell-1} \frac{\sqrt{n}}{c} \sum_{u_j \in \Omega_{cj}} e(k^T u_j)\, \widehat{y}^C_{u_j}$$

Let $i = k_j$, then, as follows from the discussion prior to theorem 3.9,

$$\begin{aligned}
p_j(k) &= p_j(i\mathbf{1}) \\
&= \frac{\sqrt{n}}{c} \sum_{h \in \Omega_{cj}} e(i\mathbf{1}^T h)\, \widehat{x}^C_h \\
&= \frac{\sqrt{n}}{c} \sum_{u_j \in \Omega_{cj}} e(k^T u_j)\, \widehat{x}^C_{u_j}
\end{aligned}$$

The observation that $\widehat{y}_{u_j} = \widehat{x}_{u_j}$ (this was established in the base case of the proof to theorem 3.5) completes the proof. □

Continuing the previous example, the fixed point as computed by Geiringer's theorem from the population vector

$$x = <0, \tfrac{1}{5}, \tfrac{1}{5}, \tfrac{1}{5}, 0, \tfrac{1}{5}, 0, 0, \tfrac{1}{5}>$$

is

$$y = <\tfrac{2}{25}, \tfrac{2}{25}, \tfrac{6}{25}, \tfrac{2}{25}, \tfrac{2}{25}, \tfrac{6}{25}, \tfrac{1}{25}, \tfrac{1}{25}, \tfrac{3}{25}>$$

Conclusion

This paper has further developed the relationship between the simple genetic algorithm and the Walsh transform begun in part I. The major points explored were

- The triangularization of the mixing scheme.
- The inverse of the expected next generation operator.
- The fixed points of the mixing scheme.
- Geiringer's theorem in the zero mutation case.

In a future paper, the theoretical machinery developed in this series of papers will be brought to bear on analyzing the detailed behavior of orbits under mixing (i.e., $x, \mathcal{M}(x), \mathcal{M}^2(x), \ldots$) in the case where simple convergence to a fixed point does not occur. Of particular interest is the range of behavior that can be exhibited by the transient phase.

Acknowledgements

This research was supported by the National Science Foundation: IRI-8917545 and CDA-9115428. The second author was partially supported by a Montana NSF EPSCOR grant and a University of Montana University Research grant.

4 References

Geiringer, H. (1944). On the Probability of Linkage in Mendelian Heredity, *Annals of Mathematical Statistics, 15*, 25-57.

Juliany, J. & Vose, M. D. (1994). The Genetic Algorithm Fractal, *Evolutionary Computation*, *2(2)*, 165-180.

Koehler, G. (1992). A Proof of the Vose-Liepins Conjecture, *Annals of Mathematics and Artificial Intelligence 5.*

Vose, M. D. (1996). Modeling Simple Genetic Algorithms, *Evolutionary Computation, 3(4)*, 453-472

Vose, M. D. (1998). *The Simple Genetic Algorithm: Foundations and Theory*, MIT Press, In Press.

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. (1998). The Simple Genetic Algorithm and the Walsh Transform: part I, Theory, *Evolutionary Computation:* In press.