Complex Systems **5** (1991) 31–44

Punctuated Equilibria in Genetic Search*

Michael D. Vose†
The University of Tennessee, Computer Science Department, 107 Ayres Hall, Knoxville, TN 37996-1301

Gunar E. Liepins‡
Oak Ridge National Laboratory, MS 6360 Bldg. 6025, PO Box 2008, Oak Ridge, TN 37831-6360

Abstract. We introduce a formalization of a simple genetic algorithm. Mathematically, two matrices F and M determine selection and recombination operators. Fixed points and their stability for these operators are investigated in terms of the eigenvalues of the associated matrices. We apply our results to one-point crossover with mutation to illustrate how the interaction between the focusing operator (selection) and the dispersion operator (recombination) results in the punctuated equilibrium frequently observed in genetic search.

1. Introduction

Designed to search irregular, poorly understood spaces, Genetic Algorithms (GAs) are general purpose algorithms developed by Holland [7] and based on ideas of Bledsoe [3] and others. Inspired by the example of population genetics, genetic search proceeds over a number of generations. The criterion of "survival of the fittest" provides evolutionary pressure for populations to develop increasingly fit individuals. Although there are many variants, the basic mechanism of a GA consists of:

1. evaluation of individual fitness and formation of a gene pool.
2. mutation and crossover.

Individuals resulting from these operations form the members of the next generation, and the process is iterated until the system ceases to improve.

*This research was supported by the National Science Foundation (IRI-8917545).
†Phone: (615) 974-5067; Electronic mail address: vose@cs.utk.edu
‡Phone: (615) 574-6640; Electronic mail address: gxl@msr.epm.ornl.gov

Fixed-length binary strings are typically the members of the population. They are selected (with replacement) for the gene pool with probability proportional to their relative fitness, which is determined by the objective function. There, they are recombined by mutation and crossover. Mutation corresponds to flipping the bits of an individual with some small probability (the mutation rate). The simplest implementation of crossover selects two "parents" from the pool and, after choosing the same random position within each string, exchanges their tails. Crossover is typically performed with some probability (the crossover rate), and parents are otherwise cloned. This recombination cycle repeats, contributing one of the resulting "offspring" each time until the next generation is full.

While this description may suffice for successful application of the genetic paradigm, it is not particularly amenable to mathematical analysis, nor does it illuminate the punctuated equilibrium (alternation between generations of relatively stable populations and periods of sudden rapid evolution) frequently observed in genetic search.

Holland [8] has addressed punctuated equilibrium from the perspective of hyperplane transforms (schemata analysis). In contrast, we develop a rigorous mathematical formalism for a simple GA and model genetic search directly.

We model GAs geometrically in sections 2 and 3 as dynamical systems in a high-dimensional Euclidean space. In section 4, we develop the basic structure of the model and establish preliminary results regarding fixed points and their stability. In section 5 we apply our qualitative results to one-point crossover with mutation to illustrate the phenomenon of punctuated equilibrium. We assume a background in mathematics including algebra (at the level of Ref. [6]) and calculus (at the level of Ref. [10]).

2. Preliminary considerations

Let Ω be the set of all length-ℓ binary strings, and let $N = 2^\ell$. Thinking of elements of Ω as binary numbers, we identify Ω with the interval of integers $[0, N-1]$. We also regard Ω as the product group

$$\mathcal{Z}_2 \times \cdots \times \mathcal{Z}_2$$

where $\mathcal{Z}_2$ denotes the additive group of integers modulo 2. The group operation $\oplus$ acts on integers in $[0, N-1]$ via these identifications, and we use $\otimes$ to represent component-wise multiplication. Hence, $\oplus$ is *exclusive-or* on integers and $\otimes$ is *logical-and*.

The tth generation of the genetic algorithm is modeled by a vector $s^t \in \mathcal{R}^N$, where the ith component of s^t is the probability that i is selected for the gene pool. Populations excluding members of Ω are modeled by vectors s^t having corresponding coordinates zero.

Let $p^t \in \mathcal{R}^N$ be a vector with ith component equal to the proportion of i in the tth generation, and let $r_{i,j}(k)$ be the probability that k results from the recombination process based on parents i and j.

Lemma 1. *Let $\mathcal{E}$ denote expectation, then*

$$\mathcal{E}\, p_k^{t+1} = \sum_{i,j} s_i^t\, s_j^t\, r_{i,j}(k)$$

Proof. The expected proportion of k in the next generation is computed by summing over all possible ways of producing k. If k results from reproduction based on parents i and j, then i is selected for reproduction with probability s_i^t, j is selected for reproduction with probability s_j^t, and k is the result of recombination with probability $r_{i,j}(k)$. ∎

Taking the limit as population size $\to \infty$, the law of large numbers gives $p_k^{t+1} \to \mathcal{E}\, p_k^{t+1}$. Thus Lemma 1 can be used to determine how the probability vector s^t changes from one generation to the next in a GA with infinite population. But first, we note an important property of $r_{i,j}(k)$:

Lemma 2. *If recombination is a combination of mutation and crossover, then*

$$r_{i,j}(k \oplus l) = r_{i\oplus k, j\oplus k}(l)$$

Proof. Let $C(i,j)$ represent the possible results of crossing i and j. Note that $k \oplus l \in C(i,j)$ iff $k \in C(i \oplus l, j \oplus l)$. Let $X(i)$ represent the result of mutating i, for some fixed mutation. Note that $k \oplus l = X(i)$ iff $k = X(i \oplus l)$. Since recombination is a combination of operations that commute with group translation, the result follows. ∎

Let F be the nonnegative diagonal matrix with i,ith entry $f(i)$, where f is the objective function, and let M be the matrix with i,jth entry $m_{i,j} = r_{i,j}(0)$. Define permutations σ_j on $\mathcal{R}^N$ by

$$\sigma_j \langle s_0, \ldots, s_{N-1}\rangle^T = \langle s_{j\oplus 0}, \ldots, s_{j\oplus(N-1)}\rangle^T$$

where vectors are regarded as columns, and T denotes transpose. Define the operator $\mathcal{M}$ by

$$\mathcal{M}(s) = \langle (\sigma_0\, s)^T M \sigma_0\, s, \ldots, (\sigma_{N-1}\, s)^T M \sigma_{N-1}\, s\rangle^T$$

Let $\sim$ be the equivalence relation on $\mathcal{R}^N$ defined by $x \sim y$ if and only if there exists $\lambda > 0$ such that $x = \lambda y$.

Theorem 1. *Let $\mathcal{E}$ denote expectation, then $\mathcal{E}\, s^{t+1} \sim F\mathcal{M}(s^t)$.*

Proof.

$$\begin{aligned}
\mathcal{E}\, p_k^{t+1} &= \sum_{i,j} s_i^t\, s_j^t\, r_{ij}(k) \\
&= \sum_{i,j} s_i^t\, s_j^t\, r_{i\oplus k, j\oplus k}(0) \\
&= \sum_{i\oplus k, j\oplus k} s_{i\oplus k}^t\, s_{j\oplus k}^t\, r_{ij}(0) \\
&= (\sigma_k\, s)^T M \sigma_k\, s
\end{aligned}$$

Since $s^{t+1} \sim Fp^{t+1}$ (the probability of selection is proportional to relative fitness), the result follows. ∎

The (expected) behavior of a simple GA is therefore determined by two matrices: fitness information appropriate for selection is contained in F, while M encodes mixing information appropriate for recombination. Moreover, the relation

$$s^{t+1} \sim F\mathcal{M}(s^t)$$

is an exact representation of the limiting behavior as population size $\to \infty$.

The matrix M has many special properties, the most obvious of which are:

Theorem 2. *The matrix M is nonnegative and symmetric, and for all i,j satisfies*

$$1 = \sum_k m_{i\oplus k,\, j\oplus k}$$

Proof. M is nonnegative since its entries are probabilities, and is symmetric since $m_{i,j} = r_{i,j}(0)$ and the results of recombination depend on the unordered set of parents. Moreover,

$$1 = \sum_k r_{i,j}(k) = \sum_k r_{i\oplus k,\, j\oplus k}(0) = \sum_k m_{i\oplus k,\, j\oplus k}$$ ∎

A more subtle property is that conjugation by the Walsh matrix W triangulates the *twist* M_* of M, where the i,jth entry of M_* is $m_{i\oplus j,i}$. We define the Walsh matrix $W = (w_{i,j})$ by

$$w_{i,j} = \prod_{k=1}^{\ell} r_{k\,(\lfloor i2^{1-k} \rfloor \operatorname{mod} 2)}(j)$$

where the Rademacher functions $r_i(x)$ are given by

$$r_i(x) = 1 - 2\left(\left\lfloor \frac{x2^i}{N} \right\rfloor \bmod 2\right)$$

In fact, conjugation by W transforms the positive[1] matrix M into a sparse symmetric matrix C having nonzero entries only where the row (or column) is smaller than the gcd of the column (row) and N. The proof of this result is outlined in the appendix.

The Walsh matrix is symmetric and orthogonal, and the rows are group characters:

$$w_{i\oplus j,k} = w_{i,k} w_{j,k}$$

For an introduction to Walsh functions and their properties see Harmuth [9]. For previous applications of Walsh functions to GAs see Bethke [2] and Goldberg [4, 5].

[1]When mutation is nonzero.

Theorem 3. *The matrix WM_*W is lower triangular.*

Proof. In view of the preceding remarks, we may represent M as WCW so that the i,jth entry of the matrix in question is

$$\sum_{k_3,k_4} w_{i,k_3} \sum_{k_1,k_2} w_{k_3 \oplus k_4,k_1} c_{k_1,k_2} w_{k_2,k_3} w_{k_4,j}$$

Using symmetry, the group character property, and rearranging gives

$$\sum_{k_1,k_2,k_3} w_{i,k_3} w_{k_1 \oplus k_2,k_3} c_{k_1,k_2} \sum_{k_4} w_{k_1,k_4} w_{j,k_4}$$

By orthogonality, the inner sum is zero unless $k_1 = j$. Hence we may simplify (modulo some multiplicative constant) to

$$\sum_{k_2} c_{j,k_2} \sum_{k_3} w_{i,k_3} w_{j \oplus k_2,k_3}$$

Orthogonality forces $k_2 = i \oplus j$, and we simplify as before to obtain $c_{j,i\oplus j}$. It therefore suffices to show

$$j > i \Longrightarrow (j \geq \gcd(i \oplus j, N)) \wedge (i \oplus j \geq \gcd(j, N))$$

This follows by a simple induction on ℓ. ∎

3. Formalization

Definition 1. Simple genetic search *corresponds to the operator $\mathcal{G} = F \circ \mathcal{M}$, where F is the fitness matrix and M is any mixing matrix satisfying Theorems 2 and 3. An initial population is modeled by a point $s^0 \in \mathcal{R}^N$, and the transition between generations is determined by $s^{t+1} \sim \mathcal{G}(s^t)$.*

This formalization generalizes the recombination induced by mutation and one-point crossover, and regards GAs with finite populations as approximations to the ideal of simple genetic search. The generality of our model allows n-point or uniform crossover, or any other mixing operator that commutes with group translation (see Lemma 2) and whose associated matrix satisfies Theorem 3.

One natural geometric interpretation of simple genetic search is to regard F and $\mathcal{M}$ as maps from $\mathcal{S}$—the nonnegative points (i.e., points with nonnegative coordinates) of the unit sphere in $\mathcal{R}^N$—to $\mathcal{S}$ (since apart from the origin, each equivalence class of $\sim$ has a unique member of norm 1). An initial population then corresponds to a point on $\mathcal{S}$, the progression from one generation to the next is given by the iterations of $\mathcal{G}$, and convergence (of the GA) corresponds to a fixed point of $\mathcal{G}$.

4. Basic properties

Regarding F as a map on $\mathcal{S}$, its fixed points correspond to the eigenvectors of F, which are the unit basis vectors $u_0, \ldots, u_{N-1}$. (Here u_j differs from the zero vector only in that the jth component of u_j is 1.) If $f(i) = f(j)$, then by passing to a quotient space (moding out by the subspace generated by u_i and u_j), the subspace corresponding to i and j is collapsed to a single dimension. Hence we may assume that f is injective by considering a suitable homomorphic image.

Theorem 4. *The basin of attraction of the fixed point u_j (of F) is given by the intersection of $\mathcal{S}$ with the (solid) ellipsoid*

$$\sum_i \left(s_i \frac{f(i)}{f(j)} \right)^2 < 1$$

Proof. Let $s \in \mathcal{S}$. The cosine of the angle between s and u_j is given by the dot product $s \cdot u_j$, and the cosine of the angle between Fs and u_j is given by $Fs/\|Fs\| \cdot u_j$. Hence, the angle between s and u_j is decreased by F when

$$s_j < \frac{s_j f(j)}{\|Fs\|}$$

which is equivalent to the statement of the theorem. ∎

Only the fixed points corresponding to the maximal value of the objective function f are in the interior of their basins of attraction. Hence *all other fixed points are unstable.* This follows from the observation that when $f(j)$ is maximal, *no* point of $\mathcal{S}$ moves away from u_j since

$$\sum_i \left(s_i \frac{f(i)}{f(j)} \right)^2 \leq \sum_i s_i^2 = 1$$

Intuitively, Theorem 4 is not surprising. Selection is a focusing operator that moves the population toward one containing the maximally fit individuals that are initially present.

Regarding $\mathcal{M}$ as a map on $\mathcal{S}$, the set $\mathcal{M}_{\text{fixed}}$ of fixed points of $\mathcal{M}$ is more difficult to analyze; it can range from all of $\mathcal{S}$ to the single point $v = \langle N^{-1/2}, \ldots, N^{-1/2} \rangle$. Moreover, intermediate behavior is possible; matrices corresponding to crossover can have surfaces of fixed points. In order to investigate $\mathcal{M}_{\text{fixed}}$ further, we need to consider the differential $\mathcal{D}_{\mathcal{M}}(x)$ of $\mathcal{M}$ at x. We need to be careful, because the differential is changed by regarding $\mathcal{M}$ as a map from $\mathcal{S}$ to $\mathcal{S}$. We therefore interpret $\mathcal{M}$ strictly (i.e., as originally defined on $\mathcal{R}^N$) in what follows.

Lemma 3. *The i,jth component of $\mathcal{D}_{\mathcal{M}}(x)$ is $2 \sum_k m_{i \oplus j, k}\, x_{i \oplus k}$.*

Proof. The calculation of $\mathcal{D}_{\mathcal{M}}$ follows from taking partial derivatives and using the symmetry of M to simplify the resulting Jacobian. ∎

A fixed point x of a map W is *stable in the sense of Lyapunov* if for any neighborhood N_1 of x there is a neighborhood N_2 of x such that the trajectory of every point of N_2 lies in N_1. A point is *asymptotically stable* if it is stable and all trajectories from some neighborhood of x converge to x. We will need the following discrete analogue from Lyapunov's theory of stability (see Ref. [1]):

Lemma 4. *Suppose that x is a fixed point of a map W and that the spectrum of the differential $\mathcal{D}_W(x)$ is contained in the open unit disk. Then x is asymptotically stable.*

Let the sum of the coordinates of $s \in \mathcal{R}^N$ be denoted by $|s|$, and let

$$\Lambda = \{x \in \mathcal{R}^N : x \text{ is nonnegative and } |x| = 1\}$$

Note that for all k,

$$I - vv^T = \sigma_k^{-1}(I - vv^T)\,\sigma_k$$

where the σ_k are regarded as permutation matrices. (Recall that $v = \langle N^{-1/2}, \ldots, N^{-1/2}\rangle$ and, interpretating the σ_k as a permutation matrices, $\sigma_k = \sigma_k^{-1} = \sigma_k^T$.) Next observe that

$$\mathcal{D}_{\mathcal{M}}(x) = 2\sum_k \sigma_k^{-1} M_*\, \sigma_k\, x_k$$

Since the column sums of M_* are constant (Theorem 2), as is also the case for $\mathcal{D}_{\mathcal{M}}(x)$ (Lemma 3), it follows from the Perron–Frobenius theory (see Ref. [11]) that, when M_* is positive, v is the unique positive eigenvector for both M_*^T and $\mathcal{D}_{\mathcal{M}}(x)^T$. Moreover, since the corresponding eigenvalues are simple and maximal, this discussion leads to a sufficient condition for a fixed point to be an attractor.

Theorem 5. *Let $x \in \mathcal{M}_{\text{fixed}}$. If the matrix M is positive, then x is asymptotically stable whenever the second largest eigenvalue of M_* is less than $1/2$.*

Proof. According to Lemma 4, it suffices to check the spectrum of the differential of $\mathcal{M}$. Since Λ is mapped into itself by $\mathcal{M}$, it suffices to consider the action of $\mathcal{M}$ restricted to Λ. The kernel of the projection $I - vv^T$ is normal to Λ, hence the spectral radius in question is $\varrho = \rho\,(\mathcal{D}_{\mathcal{M}}(x)(I - vv^T))$. Because a matrix and its adjoint share the same spectrum, the previous discussion shows

$$\begin{aligned}
\varrho &= \rho((I - vv^T)\,\mathcal{D}_{\mathcal{M}}(x)^T) \\
&= 2\rho\left(\sum_k \sigma_k^{-1}(I - vv^T)\, M_*{}^T \sigma_k\, x_k\right)
\end{aligned}$$

Since conjugation by W diagonalizes both $(I - vv^T)$ and the σ_k (this follows from the orthogonality of the Walsh functions), and also triangulates ${M_*}^T$ (Theorem 3), there is a basis in which every term of the sum is lower triangular. Because the spectral radius is invariant under change of basis and is subadditive over lower triangular matrices, we have

$$\begin{aligned} \varrho &\leq \sum_k \rho\,(\sigma_k^{-1}(I - vv^T)\,{M_*}^T \sigma_k\, x_k) \\ &= 2\,|x|\,\rho\,((I - vv^T)\,{M_*}^T) \end{aligned}$$

It remains to show that postmultiplication of ${M_*}^T$ by $I - vv^T$ sends the maximal eigenvalue of ${M_*}^T$ to zero and otherwise leaves the spectrum alone. Let ${M_*}^T v = \alpha v$ and suppose that ${M_*}^T w = \beta w$. Let $w = w' + w''$, where $w' \perp v$ and $w'' \parallel v$. Then

$$(I - vv^T)\,{M_*}^T w' = (I - vv^T)\,(\beta w - \alpha w'') = \beta w'$$

Hence every eigenvalue of ${M_*}^T$ is an eigenvalue of $(I - vv^T){M_*}^T$ with the possible exception of α. Conversely, suppose that $(I - vv^T)\,{M_*}^T w = \beta w$. It follows that

$${M_*}^T w = \beta w + vv^T {M_*}^T w = \beta w + \gamma v$$

for some scalar γ. Therefore,

$${M_*}^T \left(w + \frac{\gamma}{\beta - \alpha} v \right) = \beta \left(w + \frac{\gamma}{\beta - \alpha} v \right)$$

if $\beta \neq \alpha$. Hence every eigenvalue of $(I - vv^T)\,{M_*}^T$ is an eigenvalue of ${M_*}^T$ with the possible exception of α. Finally, suppose that $\beta = \alpha$, so that w is an eigenvector for the matrix $(I - vv^T)\,{M_*}^T$ corresponding to α. Then extending the set $\{v, w\}$ to a basis for representing ${M_*}^T$ yields

$${M_*}^T = \begin{pmatrix} \alpha & 0 & 0 & \cdots \\ \gamma & \alpha & 0 & \cdots \\ \vdots & \vdots & \vdots & \end{pmatrix}$$

which contradicts the simplicity of the maximal eigenvalue α. ∎

Although $\mathcal{M}_{\text{fixed}}$ can vary drastically, there is a group of symmetries that acts on it.

Theorem 6. *For all j, and for every mixing matrix M, $\mathcal{M}(\sigma_j\, x) = \sigma_j \mathcal{M}(x)$. In particular, we have $\sigma_j \mathcal{M}_{\text{fixed}} = \mathcal{M}_{\text{fixed}}$, and $v \in \mathcal{M}_{\text{fixed}}$.*

Proof.

$$\begin{aligned} \sigma_j \mathcal{M}(x) &= \sigma_j \langle (\sigma_0\, x)^T M \sigma_0\, x, \ldots, (\sigma_{N-1}\, x)^T M \sigma_{N-1}\, x \rangle^T \\ &= \langle (\sigma_{j+0}\, x)^T M \sigma_{j+0}\, x, \ldots, (\sigma_{j+N-1}\, x)^T M \sigma_{j+N-1}\, x \rangle^T \\ &= \langle (\sigma_0 \sigma_j\, x)^T M \sigma_0 \sigma_j\, x, \ldots, (\sigma_{N-1} \sigma_j\, x)^T M \sigma_{N-1} \sigma_j\, x \rangle^T \\ &= \mathcal{M}(\sigma_j\, x) \end{aligned}$$

Since $x = \mathcal{M}(x) \Rightarrow \sigma_j x = \sigma_j \mathcal{M}(x) = \mathcal{M}(\sigma_j x)$, it follows that $\sigma_j \mathcal{M}_{\text{fixed}} = \mathcal{M}_{\text{fixed}}$. Since $\sigma_j \mathcal{M}(v) = \mathcal{M}(\sigma_j v) = \mathcal{M}(v)$, it follows that $\mathcal{M}(v)$ is fixed by each σ_j and must therefore have equal components (i.e., $\mathcal{M}(v) \sim v$). ∎

Since the σ_j are isometries of $\mathcal{S}$, we have

$$\|\mathcal{M}(x) - \mathcal{M}(y)\| = \|\sigma_j\,(\mathcal{M}(x) - \mathcal{M}(y))\| = \|\mathcal{M}(\sigma_j\, x) - \mathcal{M}(\sigma_j\, y)\|$$

Hence the dynamical system on $\mathcal{S}$ corresponding to $\mathcal{M}$ *looks the same at every member of the population*—any neighborhood of u_j is mapped by $\sigma_{j \oplus i}$ to a neighborhood of u_i.

5. Punctuated equilibria

Punctuated equilibria typically characterize simple genetic search. Populations often alternate between generations of relative stability—indicating adaptation to the current environment represented by the position in the search space—and periods of sudden rapid evolution resulting in the emergence of a superior individual.

This phenomenon is explained in the context of our model by the qualitative properties of the "focusing operator" (selection) and the "diffusing operator" (recombination).

We illustrate with the example of one-point crossover with mutation. If χ is the crossover rate and μ is the mutation rate, then a simple calculation shows (see the appendix) that $m_{i,j}$ is

$$\frac{(1-\mu)^\ell}{2}\left\{\eta^{|i|}\left(1-\chi+\frac{\chi}{\ell-1}\sum_{k=1}^{\ell-1}\eta^{-\Delta_{i,j,k}}\right)+\eta^{|j|}\left(1-\chi+\frac{\chi}{\ell-1}\sum_{k=1}^{\ell-1}\eta^{\Delta_{i,j,k}}\right)\right\}$$

where $\eta = \mu/(1-\mu)$, integers are to be regarded as bit vectors when occurring in $|\cdot|$, division by zero at $\mu = 0$ and $\mu = 1$ is to be removed by continuity, and

$$\Delta_{i,j,k} = |\,(2^k - 1) \otimes i\,| - |\,(2^k - 1) \otimes j\,|$$

Several computer runs calculating the spectrum of M_* support the following:

Conjecture 1. *If* $0 < \mu < 0.5$, *then*

1. *The second largest eigenvalue of* M_* *is* $\frac{1}{2} - \mu$
2. *The third largest eigenvalue of* M_* *is* $2\left(1 - \frac{\chi}{\ell-1}\right)\left(\frac{1}{2} - \mu\right)^2$

Applying Theorem 5, we would infer from this conjecture that every fixed point of $\mathcal{M}$ is an attractor when $0 < \mu < 0.5$. When $\mu = 0$, calculations indicate that the elements of $\mathcal{M}_{\text{fixed}}$ are not isolated but form a surface, which suggests the condition of Theorem 5 may be necessary and sufficient in this case.

The following conjecture of Belitskii and Lyubich applies [1]:

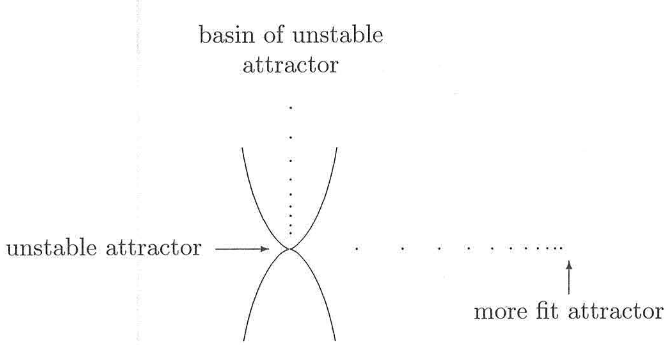
Conjecture 2. *If* $\max_{x \in X} \rho(\mathcal{D}_W(x)) < 1$, *where* X *is the compact and convex domain and codomain of* W, *then the fixed point of* W *is unique, and the sequence of iterates* $W^k(x)$ *converges to it for every choice of initial point* x.

Combining (the proof of) Theorem 5 with Theorem 6, we would infer from conjectures 1 and 2 that v is the *unique* fixed point of $\mathcal{M}$ (regarding $\mathcal{M}$ as a map on $\mathcal{S}$; Conjecture 2 would actually be applied to obtain a unique fixed point on $X = \Lambda$). Hence the dynamical system corresponding to $\mathcal{M}$ is similar to a diffusion process (when $0 < \mu < 0.5$) in that the unique fixed point v corresponds to all possible strings represented equally, all of $\mathcal{S}$ is its basin of attraction, and the dynamical system on $\mathcal{S}$ corresponding to $\mathcal{M}$ looks the same at every member of the population.

Theorem 4 shows F to have attractors corresponding to the distinct fitness values of f, all of which are unstable except the maximally fit. The geometry of the basins of attraction implies that to move out of the basin of an unstable fixed point is to move into the basin of a more fit attractor.

In GA implementations, finite populations make low probability events occur even more infrequently; mutation and crossover do not typically produce better strings with each generation. Moreover, the emergence and growth of a string having greater fitness typically requires events less probable than does focusing a population towards a prevalent high-fitness string whose dominance is not interfered with by finite-population effects.

Therefore, a population will move under the influence of F, slowing and seeming to stabilize as it approaches a fixed point. If the fixed point is not maximally fit, then it is unstable and does not lie in the interior of its basin. The diffusion-like property of recombination may then move the population outside the basin of attraction, causing the population to experience a major change as it moves under the influence of a new attractor with greater fitness. This situation is depicted by the following diagram where the dotted path represents the trajectory of a population:



6. Summary and future research

We have formalized selection and recombination within a mathematical framework as a first step to better understand the simple genetic algorithm. We have shown recombination to be a quadratic operator determined by the matrix M and the group of permutations $\{\sigma_k\}_{0\leq k<N}$, and have shown selection to be the diagonal linear transformation F. The importance of the Walsh matrix as a natural basis that simplifies the structure of genetic search has been demonstrated (by Theorem 3 and in the appendix), and it has been shown useful in understanding fixed points and their stability (as in the proof of Theorem 5). Further explanation of the role of the Walsh matrix and the development of its relation to our model will be provided in a future paper.

A more complete understanding of selection as a focusing operator and of recombination as a dispersion operator has emerged. Their qualitative properties have been used to shed light on the phenomenon of punctuated equilibrium.

Explicit formulas for the matrices corresponding to two-point and uniform crossover can be worked out. This raises the possibility that differences in performance related to the use of different mixing operators may be clarified through their analysis. Crossover and mutation represent only a few of the recombination operators whose associated matrices possess the requisite properties. It would seem promising to investigate alternatives with simple procedural counterparts.

Another interesting possibility is the development of a *synthetic GA that does not use a population.* This may be possible by implementing an approximation to $\mathcal{G}$ via sparse matrix techniques since M is sparse in a suitable basis (see the appendix).

Appendix

The explicit formula for $m_{i,j}$ given in Section 5 is easily derived from the following considerations:

1. The probability that 0 results from parents i and j depends on the probability that mutation changes the 1s occurring in the results of crossover to 0 and leaves the other bits alone.

2. The number of 1s occurring in the results produced by crossing i and j at position k are given by $|i| - \Delta_{i,j,k}$ and $|j| + \Delta_{i,j,k}$.

3. The probability of changing a specified collection of b bits (in a length-ℓ binary string) via mutation is $(1-\mu)^{\ell-b}\mu^b$.

The fact that $C = WMW$ has nonzero entries only where the row (or column) is smaller than the gcd of the column (row) and N follows from direct calculation using the following observations:

1. If h is an arbitrary function, and $i > 0$ then

$$\sum_{k_1,k_2} w_{i,k_1} w_{k_2,j} h(|k_2|) = \sum_{k_2} w_{k_2,j} h(|k_2|) \sum_{k_1} w_{i,k_1} w_{0,k_1} = 0$$

since orthogonality implies that the inner sum is zero. Similarly, if $j > 0$ then

$$\sum_{k_1,k_2} w_{i,k_1} w_{k_2,j} h(|k_1|) = 0$$

2. Suppose i and j are greater than 0, and h is an arbitrary function. Note that

$$\begin{aligned}
&\sum_{k_1,k_2} w_{i,k_1} w_{k_2,j} \sum_{k_3=1}^{\ell-1} h(|k_2| + \Delta_{k_1,k_2,k_3}) \\
&= \sum_{k_3=1}^{\ell-1} \sum_{k_4=0}^{\ell} h(k_4) \sum_{\substack{k_1,k_2 \\ |k_2|+\Delta_{k_1,k_2,k_3}=k_4}} w_{i,k_1} w_{k_2,j} \\
&= \sum_{k_3=1}^{\ell-1} \sum_{k_4=0}^{\ell} h(k_4) \sum_{\substack{k_5=0 \\ |k_5|=k_4}}^{N-1} \sum_{k_6=0}^{2^{k_3}-1} \sum_{k_7=0}^{2^{\ell-k_3}-1} w_{i,(k_5 \bmod 2^{k_3}) \oplus (k_7 2^{k_3})} w_{k_6 \oplus (2^{k_3} \lfloor k_5 2^{-k_3} \rfloor),j} \\
&= \sum_{k_3=1}^{\ell-1} \sum_{k_4=0}^{\ell} h(k_4) \sum_{\substack{k_5=0 \\ |k_5|=k_4}}^{N-1} w_{i,k_5 \bmod 2^{k_3}} w_{2^{k_3} \lfloor k_5 2^{-k_3} \rfloor,j} \sum_{k_6=0}^{2^{k_3}-1} w_{k_6,j} \sum_{k_7=0}^{2^{\ell-k_3}-1} w_{i,k_7 2^{k_3}} \\
&= \sum_{k_3=1}^{\ell-1} \sum_{k_4=0}^{\ell} h(k_4) \sum_{\substack{k_5=0 \\ |k_5|=k_4}}^{N-1} w_{i,k_5 \bmod 2^{k_3}} w_{2^{k_3} \lfloor k_5 2^{-k_3} \rfloor,j} \\
&\qquad \times \prod_{k_8=1}^{k_3} (1 + r_{k_8}(j)) \prod_{k_9=1+k_3}^{\ell} (1 + r_{k_9}(i))
\end{aligned}$$

Now assume that neither product is zero. Then, since

$$\prod_{k_8=1}^{k_3} (1 + r_{k_8}(j)) = 2^{k_3} \prod_{k_8=1}^{k_3} \left\{ 1 - \left(\left\lfloor \frac{j 2^{k_8}}{N} \right\rfloor \bmod 2 \right) \right\}$$

and

$$\prod_{k_9=1+k_3}^{\ell} (1 + r_{k_9}(i)) = 2^{\ell-k_3} \prod_{k_9=1+k_3}^{\ell} \left\{ 1 - \left(\left\lfloor \frac{i 2^{k_9}}{N} \right\rfloor \bmod 2 \right) \right\}$$

t follows that

$$\begin{aligned}
\forall\, k_8, k_9 \,.\, 0 < k_8 \le k_3 < k_9 \le \ell \implies & \left\lfloor \frac{j 2^{k_8}}{N} \right\rfloor \bmod 2 \\
= & \left\lfloor \frac{i 2^{k_9}}{N} \right\rfloor \bmod 2 = 0
\end{aligned}$$

which implies the upper k_3 bits of j and the lower $\ell - k_3$ bits of i are zero. Hence $j < \gcd(i, N)$. Similarly, a necessary condition for the nonvanishing of

$$\sum_{k_1,k_2} w_{i,k_1} w_{k_2,j} \sum_{k_3=1}^{\ell-1} h(|k_1| - \Delta_{k_1,k_2,k_3})$$

is that $i < \gcd(j, N)$.

In fact, this argument can be modified to show that, for n-point or uniform crossover,

$$c_{i,j} > 0 \Longrightarrow i \otimes j = 0$$

Since the proof of Theorem 3 shows the i,jth entry of WM_*W is $c_{j,i\oplus j}$, the implication

$$j > i \Longrightarrow j \otimes (i \oplus j) > 0$$

shows that the matrices for n-point and uniform crossover also satisfy Theorem 3.

References

[1] G. R. Belitskii and Y. I. Lyubich, "Matrix Norms and Their Applications," *Operator Theory: Advances and Applications*, Vol. 36 (Birkhäuser, 1988).

[2] A. D. Bethke, "Genetic Algorithms as Function Optimizers" (Doctoral dissertation, University of Michigan), *Dissertation Abstracts International*, **41**(9) (1988) 3503B (University Microfilms No. 8106101).

[3] W. W. Bledsoe, "The Use of Biological Concepts in the Analytical Study of Systems," presented at the ORSA-TIMS national meeting, San Francisco, November, 1961.

[4] D. E. Goldberg, "Genetic Algorithms and Walsh Functions: Part I, a Gentle Introduction," *Complex Systems*, **3** (1989) 129–152.

[5] D. E. Goldberg, "Genetic Algorithms and Walsh Functions: Part II, Deception and its Analysis," *Complex Systems*, **3** (1989) 153–171.

[6] I. Herstein, *Topics in Algebra* (Blaisdell Publishing Company, Waltham, Mass., 1964).

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

[8] J. H. Holland, "Searching Nonlinear Functions for High Values," *Applied Mathematics and Computation*, **32** (1989) 225–274.

[9] H. F. Harmuth, *Transmission of Information by Orthogonal Functions* (Springer-Verlag, New York, 1970).

[10] L. H. Loomis and S. Sternberg, *Advanced Calculus* (Addison-Wesley, Reading, Mass., 1968).

[11] H. Minc, *Nonnegative Matrices* (Wiley, New York, 1988).