DECEPTIVENESS AND GENETIC ALGORITHM DYNAMICS[†]

Gunar E. Liepins

Oak Ridge National Laboratory
MS 6360 Bldg. 6025
PO Box 2008
Oak Ridge, TN 37831-6360
gxl@msr.epm.ornl.gov

Michael D. Vose

The University of Tennessee
C. S. Dept., 107 Ayres Hall
Knoxville, TN 37996-1301
(615) 974-5067 vose@cs.utk.edu

ABSTRACT We address deceptiveness, one of at least four reasons genetic algorithms can fail to converge to function optima. We construct fully deceptive functions and other functions of intermediate deceptiveness. For the fully deceptive functions of our construction, we generate linear transformations that induce changes of representation to render the functions fully easy. We further model genetic algorithm selection and recombination as the interleaving of linear and quadratic operators. Spectral analysis of the underlying matrices allows us to draw preliminary conclusions about fixed points and their stability. We also obtain an explicit formula relating the nonuniform Walsh transform to the dynamics of genetic search.

INTRODUCTION

Designed to search irregular, poorly understood spaces, GAs are general purpose algorithms (akin to simulated annealing in this sense) developed by Holland (1975) and based on ideas of Bledsoe (1961) and others. Holland's hopes were to develop powerful, broadly applicable techniques, to provide a means to attack problems resistant to other known methods. Inspired by the example of population genetics, genetic search proceeds over a number of generations. The criteria 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. Recombination and mutation.

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

Fixed length binary strings are typically the members (genes) of the population. They contribute to the gene pool in proportion to their relative fitness (determined by the objective function). There, they are mutated and recombined by 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. The resulting "offspring" form the subsequent population. A thorough introduction and overview of GAs is provided in Goldberg (1989a), and public domain code is available from Grefenstette (1984).

The most obvious factors affecting performance are the parameter settings for population size, crossover rate, and mutation rate. Grefenstette (1986) has considered the use of meta-level GAs for determining parameter settings, but perhaps the most systematic study of these parameters was undertaken by DeJong (1975). For a survey of these and related research, see Liepins and Hilliard (1989). Less obvious factors related to performance involve the estimation of schemata utilities. In order to explain the difficulties that arise, we first sketch the basics of schemata analysis.

A schema (Holland, 1975) describes a subset of strings with similarities at certain string positions. For example, consider strings of length 3 over the alphabet $\{0,1\}$. The two strings

$$011$$

$$111$$

are similar in the sense that they are identical when the first position is ignored. Regarding $*$ as a symbol which may be instanciated to either 0 or 1, these two strings may therefore be represented by

$$*11$$

Strings over the alphabet $\{*,0,1\}$ represent schemata, and play a central role in analyzing GAs.

Let P be a finite population drawn from some universe Ω of length n binary strings. Let f be a real value fitness (objective) function

$$f : \Omega \longrightarrow \mathcal{R}$$

For any schema H, define the utility of H with respect to P as[1]

$$f_P(H) \;=\; \frac{1}{|H \cap P|} \sum_{p \in P \cap H} f(p)$$

and define the utility of H as $f_\Omega(H)$. Regarding P as changing under the influence of a genetic algorithm, let P_t denote the generation under consideration. The Schema Theorem is the inequality

$$\mathcal{E}\, |H \cap P_{t+1}| \;\geq\; |H \cap P_t|\, \frac{f_P(H)}{f_P(\Omega)}\, [1 - \alpha(H,t) - \beta(H,t)]$$

---

[1] P is a multiset, and an element of $P \cap H$ is regarded as having the multiplicity it had in P.

where $\mathcal{E}$ is an expectation operator, and α and β approximate the probabilities that an instance of H will be destroyed by crossover or mutation (respectively). The functions α and β are usually taken to be constants estimated in terms of properties provided by the concrete representation of H. A proof of this inequality is given in Goldberg (1989a).

In some sense, schemata represent the direction of the genetic search. It follows from the schema theorem that the number of instances of a schema H for which $f_P(H) > f_P(\Omega)$ is expected to increase in the next generation when $\alpha(H,t)$ and $\beta(H,t)$ are small. Therefore, such schemata indicate the area within Ω which the GA explores. Hence it is important that, at some stage, these schema contain the object of search. Problems for which this is not true are called deceptive. We provide formal definitions of deceptiveness later in the paper.

The utility of H with respect to P may be though of as an estimate of the utility of H. Holland's results (1975) regarding allocation of trials to k-armed bandit problems suggest that a GA optimally allocates its resources so long as schemata utilities are correctly estimated. Other factors influencing GA performance are the encoding of the search domain into bit strings and their manipulation by the GA. Therefore, GAs may fail to locate a function optima for several reasons which include:

1. The chosen embedding (i.e., choice of domain) is inappropriate.

2. The problem is not deceptive, but schemata utilities cannot be reliably estimated because sampling error is too large.

3. Schemata utilities can be reliably estimated, but crossover destroys individuals which represent schemata of high utility.

4. The problem is deceptive.

The first failure mode has been partially addressed by Shaefer (1985, 1987) who has incorporated dynamically changing embeddings into his ARGOT code.

The second failure mode is virtually unstudied, although some results in Goldberg (1988a) might be extended to shed some light on this issue.

The third failure mode is partially addressed by the schema theorem; adjusting crossover rate may help overcome this problem. Representational changes can also be useful. Early studies of inversion (a permutation applied to string positions) were done by Bagley (1967), Cavicchio (1970) and Frantz (1972). Holland (1975) discusses inversion as a basic genetic operator. More recently, Whitley (1987) has reported encouraging results with the use of inversion, and Goldberg and Bridges (1990) have considered reordering operators to prevent schemata disruption.

The fourth failure mode has been studied from the perspective of Walsh transforms by Bethke (1981) and Goldberg (1988, 1989b). Although the Walsh transform approach allows the construction of deceptive problems, the analysis is performed in the transform space, and lacks some degree of intuitive accessibility. Holland (1989) investigated a (computationally equivalent) hyperplane transform which deals with schemata directly. Bridges and Goldberg (1990) have shown the computational equivalence of the hyperplane transform and the Walsh transform, and extend schemata analysis to population estimates.

3

Our work follows and extends the themes developed in Liepins and Vose (1990a) and Vose and Liepins (1990). We directly construct fully deceptive functions whenever the chromosome length n is greater than two. For these same fully deceptive functions, we generate linear transformations that render the functions fully easy. We introduce the concept of basis sets and use it construct various classes of problems of intermediate deceptiveness.

The second half of our paper goes beyond static schemata analysis and models the (expected) genetic algorithm process as a dynamic system. What emerges from this model is a clearer understanding of selection as a "focusing operator" and recombination as a "dispersion operator", and a decomposition of genetic search which makes explicit the transitions between populations, between selection probabilities, and between nonuniform Walsh transforms from one generation to the next.

EMBEDDING, REPRESENTATION AND DECEPTIVENESS

A useful point of departure for understanding difficulties with genetic optimization is a commutative diagram (figure 1) that makes explicit the steps involved in their use. Frequently, the function f to be optimized is defined on a real vector space S. Since genetic search explores a finite space of bit strings, some finite subset D of S needs to be selected for investigation. Thereafter, D is regarded as the domain of f and is mapped into bit strings by an invertible map i. We refer to the injection d which maps D into S as the *embedding*, and refer to i as the *representation*. The following diagram induces functions f_d and f_i defined on D and Ω by commutativity.
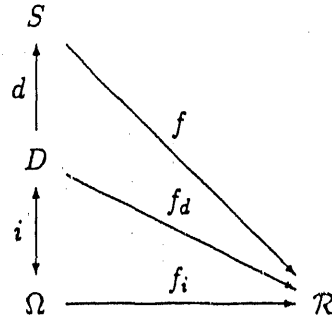

Figure 1: Commutative Diagram for Genetic Optimization

Let $s^* \in S$ be an optimum of f (i.e., a point which maximizes f). The objective of genetic optimization is to determine a point $x^* \in \Omega$ such that the difference $|f(s^*) - f_i(x^*)|$ is acceptably small.

The success of genetic optimization depends on both the embedding and the representation. For real valued optimization problems, it may be important that the embedding is centered near a function optimum and is of sufficient resolution to reflect the functions variation. One can attempt to achieve this by dynamically changing the embedding. Shaefer (1985, 1987) implements this idea in his ARGOT code by varying the parameters of an affine mapping which defines the embedding.

The role of representation is perhaps not as clear. Its importance can be explained in terms of the "building block hypothesis." This hypothesis asserts that GA search proceeds not from individual

chromosome to individual chromosome, but rather from high utility schemata with few fixed bits to high utility schemata with many fixed bits. If in the chosen representation the function optima does not lie in the schemata estimated to be of high utility, genetic search may be mislead.

The first published study of deceptiveness was undertaken by Bethke (1981). His analysis made use of the Walsh functions as a basis for the set of real valued functions on Ω. Define the j th Walsh function w_j for $j = <j_1, \ldots, j_n> \in \Omega$ by

$$w_j(k) = (-1)^{j \cdot k}$$

where

$$j \cdot k = \sum_{i=1}^{n} j_i k_i$$

Given a function $f : \Omega \to \mathcal{R}$, define the j th Walsh coefficient $\hat{f}_j$ by

$$\hat{f}_j = 2^{-n/2} \sum_{k \in \Omega} f(k) w_j(k)$$

The inversion formula

$$f(k) = 2^{-n/2} \sum_{j \in \Omega} \hat{f}_j w_j(k)$$

is a special case of the relationship between a function defined on a group and the set of group characters (Rudin, 1967). The utility of the schema H with respect to the function f can be expressed in terms of the Walsh coefficients:

$$f_\Omega(H) = \frac{1}{|H|} \sum_{p \in H} f(p) = |H|^{-1} 2^{-n/2} \sum_{p \in H} \sum_{j \in \Omega} \hat{f}_j w_j(p) = |H|^{-1} 2^{-n/2} \sum_{j \in \Omega} \hat{f}_j \sum_{p \in H} w_j(p)$$

Let the order $o(j)$ of a bit string j be the number of 1's in j, and let the order $o(H)$ of a schema H be the number of fixed positions in H.

Bethke's results rest on the observations that:

1. $f(k) = 2^{-n/2} \left(\sum_{j \cdot k \ is \ even} \hat{f}_j - \sum_{j \cdot k \ is \ odd} \hat{f}_j \right)$.

2. $\sum_{p \in H} w_j(p) = |\{x \in H : x \cdot j \ is \ even\}| - |\{x \in H : x \cdot j \ is \ odd\}|$

3. If j contains a 1 in a position where H contains a $*$, then $\hat{f}_j$ does not influence $f_\Omega(H)$.

4. If $o(j) > o(H)$, then $\hat{f}_j$ does not influence $f_\Omega(H)$.

Let H and H' be two competing schemata such that for all fixed positions i, $H'_i = 0 \Rightarrow H_i = 0$. It follows from these observations that if $\hat{f}_j = 0$ when $1 < o(j) \leq o(H)$ and if $\hat{f}_j > 0$ when $o(j) = 1$, then $f_\Omega(H) > f_\Omega(H')$.

Definition 1: Let f be a function with global optima at $\{x^*, \ldots\}$. Then f is deceptive of order m iff there exists $x \notin \{x^*, \ldots\}$ such that when H and H' are competing schemata of order not greater than m,

$$x \in H \implies f_\Omega(H) > f_\Omega(H')$$

We are now ready to provide Bethke's construction. Let $2 \leq d = 2b < n$ be the desired order of deceptiveness. The previous discussion implies the existence of a constant $c_d < 0$ (depending on d) such that the function f defined below in terms of its Walsh coefficients has maximum at $x^* = <1,\ldots,1>$ and is deceptive of order d (lower order schema containing $<0,\ldots,0>$ have greater utility):

$$\hat{f}_j = \begin{cases} 1 & \text{if } o(j) = 1 \\ c_d & \text{if } o(j) = d+1 \\ 0 & \text{otherwise} \end{cases}$$

This construction begs several related questions. Do functions exist which are deceptive of all orders $d < n$? Do functions exist which are deceptive of order $d < n-1$, but whose schemata are correctly aligned thereafter? The combinatorics of the Walsh transform analysis quickly become unwieldy, and these questions are better answered in other ways.

Definition 2: Let f be a real valued function on Ω with unique global maximum at x^*, and let x^c be the binary complement of x^*. The function f is fully deceptive iff whenever H and H' are competing schemata of order less than n,

$$x^c \in H \implies f_\Omega(H) > f_\Omega(H')$$

The opposite of fully deceptive functions are fully easy functions.

Definition 3: Let f be a real valued function on Ω with unique optimum at x^*. Then f is fully easy iff whenever H and H' are competing schemata,

$$x^* \in H \implies f_\Omega(H) > f_\Omega(H')$$

Goldberg (1989) constructed an example of a fully deceptive function for $n = 3$. Liepins and Vose (1990) showed that all fully deceptive functions can be expressed as the sum of a fully easy function and a spike function at the optimum point. Furthermore, they constructed fully deceptive functions of all orders $d > 2$ (for string lengths $n > 2$):

$$f(x) = \begin{cases} 1 - 2^{-n} & \text{if } o(x) = 0 \\ 1 - (1 + o(x))/n & \text{if } 0 < o(x) < n \\ 1 & \text{if } o(x) = n \end{cases}$$

Liepins and Vose reported that this class of fully deceptive functions could be transformed into fully easy functions through the transformation $g(k) = f \circ M(k)$, where M is an invertible linear map over Z_2. Regarding binary strings as column vectors, their $n \times n$ matrix M is

$$m_{ij} = \begin{cases} 0 \text{ for } i = j \neq n \\ 1 \text{ otherwise} \end{cases}$$

Holland (1989) noted that neither the original function nor the Walsh transform readily promote schemata analysis. Instead, he proposed the hyperplane transform which depends directly on schemata utilities. Holland's hyperplane transforms are related to a general method of representation based on what we call basis sets. Let $X = \{x_1, \ldots, x_n\}$ be a finite space. To each subset s of X corresponds an incidence vector v defined by

$$v_i = \begin{cases} 1 & \text{if } x_i \in s \\ 0 & \text{otherwise} \end{cases}$$

6

Definition 4: A collection S of subsets of X is a basis for X iff the associated incidence vectors form a basis for for $\mathcal{R}^n$.

Lemma 1: The collection of all schema containing $<1\ldots1>$ is a basis for Ω.

Proof: For any schema s, replace the "$*$"s in s with "0"s and interpret the result as a binary integer. This defines a map h from those schema containing $<1\ldots1>$ to the set of integers $\{0,\ldots,2^n-1\}$. Form a matrix having as rows the incidence vectors corresponding to schema of S, and order the rows according to increasing values of h. This matrix is upper diagonal and hence of full rank.

Theorem 1: Let $x \in \Omega$. The collection of all schema containing x is a basis for Ω.

Proof: Let σ be the matrix corresponding to that permutation which sends the j th component of a binary vector to the $j \oplus x$ th position, where $x \in \Omega$ is fixed and $\oplus$ denotes componentwise exclusive-or. If v is the incidence vector for a schemata s, it follows that σv is the incidence vector for the schemata $x \oplus s$. Therefore the incidence vectors associated with a translation (by x in Ω) of a basis are obtained by mapping the incidence vectors associated with that basis by σ. Since permutation matrices are invertible, they preserve linear independence, and since x was arbitrary, it can translate the basis of Lemma 1 to any point of Ω.

Theorem 2: Let S be a basis for X, and f a real valued function defined on S. There exists a unique function $g : X \to \mathcal{R}$ such that

$$s \in S \implies \frac{1}{|S|} \sum_{x \in S} g(x) = f(s)$$

Proof: let M be the matrix having as rows the incidence vectors corresponding to elements of S, let g be the column vector of required function values, and let f be the column vector of given values. The condition to be satisfied is

$$DMg = f$$

where D is a diagonal matrix containing $|x|^{-1}$ for $x \in S$. Hence g is uniquely determined by $g = M^{-1}D^{-1}f$

Definition 5: Let x be a point in Ω. A schemata path at x is a nested sequence of schemata $H_0 \subset \cdots \subset H_n$ containing x such that $o(H_i) = i$

Let us now return to the concept of deceptiveness. Intuitively, deceptiveness occurs whenever a "good path" leads to a "bad point" or a "bad path" leads to a "good point".

Definition 6: Let $x \in \Omega$ and let J be a schemata path at x. Then f is decreasing at x along J of order (a,b) iff whenever H and H' are two schemata in J,

$$a \le o(H) < o(H') \le b \implies f_\Omega(H) > f_\Omega(H')$$

If $a = 0$, we shall use the term "decreasing along J of order b". The definitions for increasing along J are defined analogously.

We can now turn our attention to the existence of classes of functions of intermediate deceptiveness.

We assume that the functions of interest have a unique optimal which without loss of generality is at $< 0, \ldots, 0 >$. We prove each of the following classes are nonempty:

C1. Functions with several schemata paths at the optimal; some of them increasing of order n, and others decreasing of order $n - 1$.

C2. Functions all of whose schemata paths at the optimal are increasing for some order $d < n - 1$ and decreasing thereafter (except at order n).

C3. Functions all of whose schemata paths at the optimal are decreasing for some order $d < n - 1$ and increasing thereafter.

These classes are interesting because real problems could presumably have some paths that are deceptive and other paths that aren't, or could have some regions of deceptiveness either preceded or followed by regions which are nondeceptive. Intuitively, one might expect that the density of non-deceptive paths or the depth of deceptiveness is related to whether a GA discovers an optimum.

The proof that these classes are nonempty follows from the observation that each is defined in terms of schemata paths at the single point $< 0, \ldots, 0 >$. By Theorem 1, the collection of all schema at a point forms a basis, hence the schema involved in the definitions of these classes are linearly independent. It follows from Theorem 2 that assigning arbitrary values to any set of linearly independent schema will induce a fitness function consistent with the given utilities.

GAs AS DYNAMIC SYSTEMS

In this section we sketch our development of GAs as dynamic systems. Since many of our results have cumbersome proofs, we will report on our progress in this section and refer the reader to Vose and Liepins (1990) for technical details.

We represent recombination (1-point, 2-point, or uniform crossover followed by mutation) as a quadratic operator determined by a fixed non-negative symmetric matrix M in conjunction with a group of permutations on chromosome strings. The matrix M has special properties, many of which result from the commutativity of crossover and mutation with group translation. We formalize selection as multiplication by a diagonal matrix F. Spectral analysis of M and F allows us to draw preliminary conclusions about fixed points and their stability. We also obtain an explicit formula relating the nonuniform Walsh transform to the dynamics of genetic search.

Let Ω be the set of all length n binary strings, and let $N = 2^n$. 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

$$Z_2 \times \ldots \times Z_2$$

where 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 componentwise multiplication.

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

8

modeled by vectors s^t having corresponding coordinates zero.

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

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

Let $C(i,j)$ represent the possible results of crossing i and j, and note that $k \oplus \ell \in C(i,j)$ if and only if $k \in C(i \oplus \ell, j \oplus \ell)$. If $X(i)$ represents the result of mutating i for some fixed mutation, then $k \oplus \ell = X(i)$ if and only if $k = X(i \oplus \ell)$. Since recombination is a combination of operations which commute with group translation, it follows that

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

This allows recombination to be expressed via the matrix M defined by $m_{i,j} \;=\; r_{i,j}(0)$. Let F be the nonnegative diagonal matrix with ii th entry $f(i)$, where f is the objective function, and let σ_j be permutations on $\mathcal{R}^N$ given by

$$\sigma_j < y_0, \ldots, y_{N-1} >^T \;=\; < y_{j \oplus 0}, \ldots, y_{j \oplus (N-1)} >^T$$

where T denotes transpose. Define operators $\mathcal{M}$, $\mathcal{F}$, and $\mathcal{G}$ by

$$\mathcal{M}(s) \;=\; < (\sigma_0 s)^T M \sigma_0 s, \ldots, (\sigma_{N-1} s)^T M \sigma_{N-1} s >^T$$

$$\mathcal{F}(s) \;=\; Fs$$

$$\mathcal{G} \;=\; \mathcal{F} \circ \mathcal{M}$$

and let $\sim$ represent the equivalence relation on $\mathcal{R}^N$ defined by $x \sim y$ if and only if $\exists \lambda > 0$. $x = \lambda y$. It follows that

$$\mathcal{E} s^{t+1} \;\sim\; \mathcal{G}(s^t)$$

The expected behavior of a simple GA is therefore determined by matrices M and F. Fitness information appropriate for the selection operator $\mathcal{F}$ is contained in F, while M encodes mixing information appropriate for the recombination operator $\mathcal{M}$. Moreover, the relation

$$s^{t+1} \;\sim\; \mathcal{G}(s^t)$$

is an exact representation of the limiting behavior as population size $\longrightarrow \infty$. One natural geometric interpretation of this formalization is to regard $\mathcal{F}$ and $\mathcal{M}$ as maps from $\mathcal{S}$, the nonnegative points 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 $\mathcal{G}$.

The properties of the operator $\mathcal{F}$ are straightforward to analyze. Regarding $\mathcal{F}$ as a map on $\mathcal{S}$, its fixed points correspond to the eigenvectors of F, i.e., the unit basis vectors $u_0, \ldots, u_{N-1}$. If $f(i) = f(j)$, then by passing to a quotient space (moding out by the linear span of 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. The basin of attraction of the fixed point u_j is given by the intersection of S with the (solid) ellipsoid

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

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 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 1 is not surprising. Selection is a focusing operator which moves the population towards one containing only the maximally fit individuals which are initially present. The properties of $\mathcal{M}$ are less immediate.

For 1-point crossover with mutation, we explicitly calculate the mixing matrix M for crossover rate χ and mutation rate μ as

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

Here integers are to be regarded as bit vectors when occurring in $|\cdot|$, division by zero when $\mu \in \{0, 1\}$ is to be removed by continuity, and

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

Our results concerning fixed points and their stability derive from general properties of M. The most obvious of which are that M is nonnegative, symmetric, and for all i, j satisfies

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

Associated with the matrix M is a matrix M_* related to the differential of $\mathcal{M}$ whose i, j th entry is $m_{i \oplus j, i}$. We have shown that if M_* is positive and its second largest eigenvalue is less than $1/2$, then every fixed point of recombination is asymptotically stable. Several computer runs calculating the spectrum of M_* show that for one point crossover with mutation, if $0 \leq \mu \leq 0.5$ then

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

Thus, when the mutation rate is between 0 and $1/2$, the fixed population distributions under 1-point crossover followed by mutation are asymptotically stable. A conjecture of G. R. Belitskii and Yu. I. Lyubich relating to discrete dynamical systems further implies that fixed population would therefore

10

be unique. Using symmetry p.operties, we can show the uniform population (i.e., all chromosomes represented equally) is fixed by recombination. Hence the uniform population is the only fixed point of recombination. This supports our view of recombination as a "dispersion" operator.

A much less obvious property of M is that conjugation by the matrix $\mathcal{W}$ representing the Walsh transform triangulates M_*. Moreover, conjugation by $\mathcal{W}$ also simultaneously diagonalizes the permutation matrices σ_j to ± 1 along the diagonal, and makes the matrix M sparse.

Let Q represent the sparse matrix $\mathcal{W} M \mathcal{W}$, let D_j represent the ± 1 diagonal matrix $\mathcal{W} \sigma_j \mathcal{W}$, let $Q_j = D_j Q D_j$, and define the operator $\mathcal{H}$ by

$$\mathcal{H}(s) \; = \; < s^T Q_0\, s, \ldots, s^T Q_{N-1}\, s >^T$$

Note that $\mathcal{H}$ has relatively simple structure; it is a system of quadratic forms having sparse coefficient matrices which differ only in their signs. Moreover, $\mathcal{G}$ is representable in terms of $\mathcal{H}$:

$$\mathcal{G}(s) \; = \; F \circ \mathcal{H} \circ \mathcal{W}(s)$$

Let v^t be defined by $v^t = \mathcal{W} s^t = \mathcal{W} F p^t = $ the nonuniform Walsh transform of f at generation t. Therefore

$$s^{t+1} \sim \mathcal{G}(s^t) \implies \mathcal{W} s^{t+1} \sim \mathcal{W} \circ \mathcal{G}(s^t) \implies v^{t+1} \sim \mathcal{W} \circ \mathcal{G} \circ \mathcal{W}(v^t)$$

Hence the operator $\mathcal{W} \circ \mathcal{G} \circ \mathcal{W}$ maps the nonuniform Walsh transform from one generation to the next. Moreover we have the simple representation

$$\mathcal{W} \circ \mathcal{G} \circ \mathcal{W} \; = \; \mathcal{W} \circ F \circ \mathcal{H}$$

The operators $\mathcal{F}$, $\mathcal{H}$, and $\mathcal{W}$ may be interpreted as *selection*, *mixing* and *transform* respectively. Moreover, their interleaving models the progression of genetic search

$$\cdots \; \longrightarrow \; s^t \; \overset{\mathcal{W}}{\longrightarrow} \; v^t \; \overset{\mathcal{H}}{\longrightarrow} \; p^{t+1} \; \overset{\mathcal{F}}{\longrightarrow} \; s^{t+1} \; \longrightarrow \; \cdots$$

The transition from one generation to the next may therefore be equivalently regarded as proceeding through selection vectors, population vectors, or nonuniform Walsh transforms.

SUMMARY

In this paper we have addressed both static and dynamic properties of the genetic algorithm. The static analysis addressed problem deceptiveness. The dynamic analysis addressed the time evolution of the expected population distribution.

We began our discussion of deceptiveness by explicating four failure modes for the genetic algorithm. Next, we summarized Bethke's results regarding difficult functions. We exhibited fully deceptive functions and invertible linear transformations which transform these functions into fully easy functions.

We further introduced basis sets and used them to prove the existence of functions having various intermediate degrees of deceptiveness.

Our modeling of GAs as dynamical systems focused on the expressibility of recombination as quadratic forms in terms of a single, fixed mixing matrix M. We have discovered several special properties of this matrix and have related the spectrum of an associated matrix M_* to the stability of fixed points of recombination. Computer calculations indicate a simple relation between the spectrum of M_* and the crossover and mutation rate which leads to the conclusion that that the only stable fixed population distribution for recombination is uniform. Our model leads to a decomposition of genetic search which makes explicit the transitions between populations, between selection probabilities, and between nonuniform Walsh transforms from one generation to the next.

BIBLIOGRAPHY

Bagley, J. D. (1967). The Behavior of Adaptive Systems which Employ Genetic and Correlational Algorithms. (Doctoral dissertation, University of Michigan). *Dissertation Abstracts International* 28(12), 5106B. (University Microfilms No. 68-7556).

Bethke, A. D. (1981). Genetic Algorithms as Function Optimizers. (Doctoral dissertation, University of Michigan). *Dissertation Abstracts International* 41(9), 3503B. (University Microfilms No. 8106101).

Bledsoe, W. W. (1961, November). *The Use of Biological Concepts in the Analytical Study of Systems.* Paper presented at the ORSA-TIMS National meeting, San Francisco, CA.

Bridges, C. L. & D. E. Goldberg (1990)

Cavicchio, D. J. (1970). *Adaptive Search Using Simulated Evolution.* Unpublished doctorial dissertation, University of Michigan, Ann Arbor.

DeJong, K. A. (1975). An Analysis of the Behavior of a Class of Genetic Adaptive Systems. (Doctoral dissertation, University of Michigan). *Dissertation Abstracts International* 36(10), 5140B. (University Microfilms No. 76-9381).

Frantz, D. R. (1972). Non-Linearities in Genetic Adaptive Search. (Doctoral dissertation, University of Michigan). *Dissertation Abstracts International,* 33(11), 5240B–5241B. (University Microfilms No. 73-11 116).

Goldberg, D. E. (1988b). *Genetic Algorithms and Walsh Transforms: Part I, a Gentle Introduction* (TCGA Report No. 88006). Tuscaloosa: University of Alabama, The Clearinghouse for Genetic Algorithms.

Goldberg, D. E. (1989a). *Genetic Algorithms in Search, Optimization, and Machine Learning.* Addison-Wesley.

Goldberg, D. E. (1989b). *Genetic Algorithms and Walsh Functions: Part II, Deception and its Analysis* (TCGA Report No. 89001). Tuscaloosa: University of Alabama, The Clearinghouse for Genetic Algorithms.

Goldberg, D. E. & C. L. Bridges (1990). An Analysis of a Reordering Operator on a GA-Hard Problem, *Biological Cybernetics*.

Grefenstette, J. J. (1984). *A User's Guide to GENESIS* (Technical Report No. CS-84-11). Nashville: Vanderbilt University, Department of Computer Science.

Grefenstette, J. J. (1986). Optimization of Control Parameters for Genetic Algorithms. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-16(1)*, 122–128.

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

Holland, J. H. (1989). Searching Nonlinear Functions for High Values. *Applied Mathematics and Computation* (to appear).

Liepins, G. E. & M. D. Vose (1990a). Representational Issues in Genetic Optimization, *Journal of Experimental and Theoretical Artificial Intelligence*, 2(2), 4-30.

Rudin, W. R. (1967). *Fourier Analysis on Groups*. Wiley Interscience.

Shaefer, C. G. (1985). Directed Trees Method for Fitting a Potential Function. *Proceedings of an International Conference on Genetic Algorithms and Their Applications* (pp 74 –88). Hillside, NJ: Lawrence Erlbaum Associates.

Shaefer, C. G. (1987). The ARGOT Strategy: Adaptive Representation Genetic Optimizer Technique. *Genetic Algorithms and Their Applications: Proceedings of the Second International Conference on Genetic Algorithms*. 50–58.

Vose, M. D. & G. E. Liepins (1990). Punctuated Equilibria in Genetic Search, submitted to *Complex Systems*.

Whitley, I. (1987). Using Reproductive Evaluation to Improve Genetic Search and Heuristic Discovery. *Genetic Algorithms and Their Applications: Proceedings of the Second International Conference on Genetic Algorithms*. Lawrence Erlbaum Associates.

DISCLAIMER

END

DATE FILMED

11 / 07 / 90