Complex Systems **5** (1991) 45–61

Polynomials, Basis Sets, and Deceptiveness in Genetic Algorithms*

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

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

Abstract. The degree to which the genetic optimization process is transparent is in part determined by the form of the objective function. We develop two forms from first principles: polynomial forms and basis sets. We characterize three function classes that are fully easy for the genetic algorithm in terms of the polynomial representation. We generate functions of varying degrees of deceptiveness in terms of the representation provided by basis sets. We further show the relationship between these representations and the more standard Walsh polynomials.

1. Introduction

Many significant optimization problems are defined over finite spaces. These include virtually all combinatorial optimization problems, many of which are extremely difficult. Genetic algorithms (GAs) are one general class of techniques proposed to solve such problems. Although GAs have been successful in a variety of applications—including the design of turbine blades [14], communication networks [6, 8], VLSI design [5], and stack filter design [7]—their behavior is still not fully understood. Furthermore, the class of functions for which genetic algorithms are suited has not been well characterized.

Our work builds on and helps unify work of Bethke [3], Goldberg [9, 10], Bridges and Goldberg [4], Holland [11], Liepins and Vose [13], and Battle and

*This research was supported by the Air Force Office of Scientific Research and the Office of Naval Research (F 49620-90-C-0033), and by the National Science Foundation (IRI-8917545).

†Phone: (615) 574-6640; Electronic mail address: `gxl@msr.epm.ornl.gov`
‡Phone: (615) 974-5067; Electronic mail address: `vose@cs.utk.edu`

© 1991 Complex Systems Publications, Inc.

Vose [2]. At the suggestion of Barto, Bethke [3] used Walsh functions to begin characterizing functions as easy or difficult for genetic optimization. His premise and that of work following his was that easily optimized functions have properly aligned schemata utilities. Although he did not obtain a complete characterization, he developed sufficient conditions for easy functions and partial conditions for hard functions. However, because his conditions are stated in the transform space, they are not intuitive and are difficult to understand.

Goldberg [10] reviewed Walsh transforms and constructed a fully deceptive function of order 3. Holland [11], recognizing that Walsh transform analysis was a viable approach to characterizing the difficulty of problems, introduced the hyperplane transform as a tool for simplifying the analysis. Bridges and Goldberg [4] introduced non-uniform Walsh transforms to analyze the evolution of the (implicit) estimates of schemata utilities on the basis of the current population. Liepins and Vose [13] constructed fully deceptive functions of arbitrary order, and then exhibited representational change operators that rendered them fully easy. Battle and Vose [2] noted that this change of representation could be reinterpreted as search through M-schemata space induced by M-crossovers.

This paper continues the theme of these previous investigations while viewing deception from the perspective offered by alternate representations. Three function classes that are fully easy for the genetic algorithm are characterized in terms of polynomial forms. Polynomial forms are investigated in the context of M-schemata analysis, and the relationship to Walsh polynomials is developed. Basis sets are introduced and functions of various degrees of deceptiveness are generated in terms of them.

2. Deception

The essence of deceptiveness stems from the observation that, in some sense, schemata represent the direction of genetic search. It follows from the Schema Theorem that the number of instances of a schema is expected to increase in the next generation if it is of above average utility and is not disrupted by crossover. Therefore, such schemata indicate the area within the search space that the GA explores, and 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*.

Although many of our techniques extend to any finite cardinality alphabet, we limit our attention to functions defined over the collection Ω of length ℓ binary strings. We regard Ω as the vector space

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

where $\mathcal{Z}_2$ denotes finite field of integers modulo 2. The additive group operation $\oplus$ is equivalent to component-wise "exclusive-or." We think of vectors as column vectors, but will for notational simplicity often display elements of Ω as binary numbers, thus identifying Ω with the interval of integers $[0, N-1]$.

It is also convenient to sometimes regard a binary vector v as an incidence vector, that is, as representing the set of subscripts for which $v_i = 1$.

Definition 1. *Let M be an ℓ-by-ℓ matrix over $\mathcal{Z}_2$, and let $\mathcal{L}_j M$ be the linear span of those columns of M represented by the incidence vector j.*

Holland schemata correspond to those subsets of Ω that can be represented as $k \oplus \mathcal{L}_j I$ for some k, where I is the identity matrix. For example, the schema $*11$ has the form

$$*11 \;=\; \{011, 111\} \;=\; 111 \oplus \mathcal{L}_{100} I$$

Those subsets $\{k \oplus \mathcal{L}_j M\}_{j,k \in \Omega}$ of Ω where M is some fixed invertible matrix are referred to as M-schemata, or simply schemata. These generalizations of Holland schemata were introduced and analyzed by Battle and Vose [2] to interpret the representational transformations of Liepins and Vose [13].

Let the order $o(j)$ of j be the number of 1s in j, and let the order $o(s)$ of a schema s be the codimension of s. In other words, if $s = k \oplus \mathcal{L}_j M$, then $o(s) = \ell - o(j)$ (for Holland schema, this reduces to the number of fixed positions). Two different schemata s and s' are said to be competing if they are translates of each other, that is, if $s = k \oplus s'$ for some k.

Definition 2. *Let f be a real-valued function on Ω, and let s be a schema. The utility $u_f(s)$ of s with respect to f is*

$$u_f(s) = \frac{1}{|s|} \sum_{k \in s} f(k)$$

Definition 3. *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 s and s' are competing schemata of order not greater than m,*

$$x \in s \implies u_f(s) > u_f(s')$$

Bethke [3] approached the analysis of deceptiveness by expressing Holland schemata utilities in terms of Walsh transforms. Let $1 < d = 2b < n-1$ be the desired order of deceptiveness. Bethke's construction showed the existence of a constant $c_d < 0$ such that the function f defined below in terms of its Walsh coefficients has maximum at $x^* = \vec{1}$ and is deceptive of order d:

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

His construction begs several related questions. Do functions exist that are deceptive of all orders $d < n$? Do functions exist that 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.

Although Bethke's construction leads to deceptive functions, they need not be the most difficult ones. In the sense that all schemata lead diametrically away from the optimal, fully deceptive functions are maximally deceptive.

Definition 4. *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^* (i.e., $\vec{1} \oplus x^*$). The function f is fully deceptive iff, whenever s and s' are competing schemata of order less than n,*

$$x^c \in s \implies u_f(s) > u_f(s')$$

The counterparts to fully deceptive functions are fully easy functions.

Definition 5. *Let f be a real-valued function on Ω with unique optimum at x^*. Then f is fully easy iff, whenever s and s' are competing schemata,*

$$x^* \in s \Longrightarrow u_f(s) > u_f(s')$$

The first construction of a fully deceptive function was given by Goldberg [10] for $\ell = 3$. Liepins and Vose [13] later gave a construction of a fully deceptive function f for string lengths $\ell > 2$:

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

Amazingly, this class of fully deceptive functions f can be transformed into fully easy functions g via $g = f \circ M$, where M is the linear transformation of Ω with matrix

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

Battle and Vose [2] have explained this result by the observation that a function may be fully deceptive with respect to Holland schemata and fully easy with respect to M-schemata for appropriate M. Moreover, they show how the choice of schemata that direct genetic search can be made using a suitable linear transformation. We therefore consider deceptiveness in the context of M-schemata in this paper.

3. Polynomial forms

A simple inductive argument proves that any function f over Ω may be expressed uniquely in the form

$$f(x) = \sum_{N \subset \{1,\ldots,\ell\}} \alpha_N \prod_{n \in N} e_n^T x$$

where the vector e_n contains 1 in the nth column and 0 elsewhere, T denotes transpose, and the α_N are coefficients. Regarding the vector x as having components $x_1, \ldots, x_\ell$, we may view f as a polynomial in the variables $x_1, \ldots, x_\ell$.

The proof is little more than the observation that

$$\begin{aligned} f(x) &= f(x_1, \ldots, x_\ell) \\ &= x_\ell f(x_1, \ldots, x_{\ell-1}, 1) + (1 - x_\ell) f(x_1, \ldots, x_{\ell-1}, 0) \end{aligned}$$

Since $f(x_1, \ldots, x_{\ell-1}, 1)$ and $f(x_1, \ldots, x_{\ell-1}, 0)$ have fewer variables than $f(x_1, \ldots, x_\ell)$, this reduction can be applied recursively to develop the required representation.

For example, the fitness function for the fully deceptive problem of Liepins and Vose [13] for $\ell = 3$ has the representation

$$f(x) = \tfrac{5}{6} - \tfrac{1}{2}x_1 - \tfrac{1}{2}x_2 - \tfrac{1}{2}x_3 + \tfrac{1}{6}x_1x_2 + \tfrac{1}{6}x_1x_3 + \tfrac{1}{6}x_2x_3 + \tfrac{7}{6}x_1x_2x_3$$

In order to relate polynomials to (Holland) schemata utilities, it is convenient to regard Ω as a probability space. Let $d\nu$ be a normalized counting measure on $\mathcal{Z}_2$, and let $d\mu$ denote the product measure $d\mu = d\mu_\ell = d\nu \times \cdots \times d\nu$, ℓ factors.

Theorem 1. *For any polynomial f, the utility of any Holland schema is obtained by integrating f after evaluating at the fixed positions of the schema.*

Proof. Let the Holland schema be $s = k \oplus \mathcal{L}_j I$, and let $f_j(k, x)$ be that function of those variables indexed by the incidence vector j that is obtained from f by instanciating the other variables x_i with k_i. Note that $f_j(k, x)$ corresponds to f after evaluating at the fixed positions of the schema s, and may be regarded as a function over $\mathcal{L}_j I$, which represents the variable positions of s. We have

$$\begin{aligned} \frac{1}{|s|} \sum_{x \in s} f(x) &= 2^{-o(j)} \sum_{x \in \mathcal{L}_j I} f(k \oplus x) \\ &= 2^{-\ell} 2^{\ell - o(j)} \sum_{x \in \mathcal{L}_j I} f_j(k, x) \\ &= \int_{(\mathcal{L}_j I)^\perp} \int_{\mathcal{L}_j I} f_j(k, x) \, d\mu_{o(j)} \, d\mu_{\ell - o(j)} \\ &= \int f_j(k, x) \, d\mu \end{aligned}$$

■

Theorem 2. *The utility of any Holland schema is obtained by evaluating f at the schema where $*$ counts as $\frac{1}{2}$.*

Proof. This follows from Theorem 1 and the observation that, if we think of $x_n(\omega) = e_n^T \omega$ as random variables on Ω, then they are independent. Hence for any $N \subset \{1, \ldots, \ell\}$

$$\int \prod_{n \in N} x_n \, d\mu = \prod_{n \in N} \left(\int x_n \, d\nu \right) = \tfrac{1}{2}^{|N|}$$

Therefore, integrating $f_j(k,x)$ has the affect of evaluating nonfixed positions with $\frac{1}{2}$. ∎

For example, the utility of $*10$ for our fully deceptive problem is given by

$$f(\tfrac{1}{2},1,0) = \tfrac{5}{6} - \tfrac{1}{2}\tfrac{1}{2} - \tfrac{1}{2}1 - \tfrac{1}{2}0 + \tfrac{1}{6}\tfrac{1}{2}1 + \tfrac{1}{6}\tfrac{1}{2}0 + \tfrac{1}{6}10 + \tfrac{7}{6}\tfrac{1}{2}10 = \tfrac{1}{6}$$

3.1 Easy functions

Some classes of easy functions are trivial to construct. Constant functions are one such example. What is substantially more difficult is a characterization of all easy (or fully easy) functions. Liepins and Vose [13] proved full easiness for linear, monotone, and linearly dominated polynomial functions. These results are summarized below with new proofs that illustrate the power of Theorem 2. To be thorough, we should note that, in some sense, the class of linearly dominated polynomial functions was anticipated by Bethke [3]. His Theorem 3.4.1 effectively specifies a linear dominance condition in the transform space.

Linear polynomials are defined as those with no cross terms:

$$f = a_o + a_1x_1 + \cdots + a_nx_n$$

Since linearity is invariant and schemata utilities are permuted by any transformation of representation effected by translation (modulo 2), we may assume that f is maximal at $\vec{0}$. Suppose that, in some set of competing schemata, the schema containing $\vec{0}$ did not have maximal utility. Since utilities are computed by *evaluation* (where $*$ counts as $\frac{1}{2}$), it follows that changing 0 to 1 in some position increases f. Hence some term has a positive coefficient. Therefore, f could not be maximal at $\vec{0}$, which is a contradiction.

Monotone polynomials are defined as those having all coefficients of like sign. If we assume that $\vec{0}$ is a maximum and that in some set of competing schemata the schema containing $\vec{0}$ did not have maximal utility, then we conclude as before that some term has a positive coefficient. Since all coefficients have like sign, this implies that all coefficients are positive. Therefore, f could not be maximal at $\vec{0}$, which is a contradiction. The case where $\vec{1}$ is maximal is analogous.

Let f be a polynomial, and let L be the linear part of f (those terms involving at most one variable). Linearly dominated polynomials are defined as those for which

$$\left|\frac{\partial}{\partial x_j}L\right| > \left|\frac{\partial}{\partial x_j}(f-L)\right|$$

Let f be linearly dominated. Suppose that in some set of competing schemata, s contains a maximum m and does not have maximal utility. Note that L is maximized at m, since the condition on partial derivatives would otherwise imply that f is not maximized at m. Also, s is non-maximal when utilities are computed with respect to L, since changing a 0 to a 1 increases (decreases) f iff L is increased (decreased). We have therefore reduced the problem to a situation that has previously been shown inconsistent: a linear polynomial (L) maximized at a point (m) that is contained in a non-maximal schema (s).

3.2 Linear transformations of polynomial forms

Since M-schemata are the images under the linear transformation M of Holland schemata [2], calculation of M-schemata utilities for f coincide with calculation of Holland schemata utilities for $f \circ M$. Therefore, in principle one could apply Theorem 2 to the polynomial representation of $f \circ M$. However, if we are solely interested in the transformation of schemata utilities, this can be computed otherwise. Let

$$g(x) = f(M(x)) = \sum_{N \subset \{1,\ldots,\ell\}} \alpha_N \prod_{n \in N} \langle v_n x \rangle$$

where the nth row of M is v_n, and angle brackets denote reduction modulo 2. Applying Theorem 1, the utility $u_g(s)$ of the schema $s = k \oplus \mathcal{L}_j I$ is

$$\sum_{N \subset \{1,\ldots,\ell\}} \alpha_N \int \prod_{n \in N} \langle id_j(k, v_n x) \rangle$$

where id is the identity function. Observe that, for *any* function h, an application of Fubini's theorem [1] yields

$$\int \langle x_i + \cdots \rangle \, h(x_{i+1}, \ldots) \, d\mu = \tfrac{1}{2} \int h(x_{i+1}, \ldots) \, d\mu$$

Next, notice that any product of modulo 2 sums is either 0 or may be put in the form

$$\langle x_i + \cdots \rangle \, h(x_{i+1}, \ldots)$$

where $h(x_{i+1}, \ldots)$ is also a product of modulo 2 sums and is itself in this form; simply equate each sum to 1 and reduce (in Ω) the resulting system to triangular form. This corresponds to iteratively applying the reduction

$$\langle x + y \rangle \, \langle x + z \rangle \quad = \quad \langle x + y \rangle \, \langle 1 + y + z \rangle$$

In other words, the integral of a product of modulo 2 sums is 0 if the system that results from equating each sum to 1 is inconsistent, and is $\frac{1}{2}$ to the number of nontrivial factors otherwise.

For example, consider our previous example f and the schema $s = M(*10)$ where

$$M \;=\; \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

The utility $u_f(s)$ is

$$\int \tfrac{5}{6} - \tfrac{1}{2}\langle 1+0\rangle - \tfrac{1}{2}\langle x_1+0\rangle - \tfrac{1}{2}\langle x_1+1+0\rangle + \tfrac{1}{6}\langle 1+0\rangle\,\langle x_1+0\rangle +$$

$$+ \tfrac{1}{6}\langle 1+0\rangle\,\langle x_1+1+0\rangle + \tfrac{1}{6}\langle x_1+0\rangle\,\langle x_1+1+0\rangle +$$

$$+ \tfrac{7}{6}\langle 1+0\rangle\,\langle x_1+0\rangle\,\langle x_1+1+0\rangle\; d\mu$$

which simplifies to

$$\tfrac{5}{6} - \tfrac{1}{2} - \tfrac{1}{4} - \tfrac{1}{4} + \tfrac{1}{12} + \tfrac{1}{12} + 0 + 0 = 0$$

3.3 Walsh transforms

Since there are differing accounts of Walsh transforms, we include a brief introduction to fix notation.

The set of homomorphisms from Ω into the unit circle in the complex plane forms a group under multiplication, called the character group. Since any element x of Ω is self-inverse ($x \oplus x = 0$), the range of any group character lies in the set $\{1, -1\}$. Moreover, a character is uniquely determined by the subgroup it maps into 1. Define w_i by $w_i(x_1 \cdots x_\ell) = (-1)^{x_i}$, and for j in the power set J of $\{1, \ldots, \ell\}$, define $w_j(x) = \Pi_{i \in j} w_i(x)$.

Since the collection of 2^ℓ functions $\{w_j\}_{j \in J}$ are distinct group characters, a theorem of Artin (see Ref. [12]) implies they are linearly independent and therefore form a basis for the space of functions over Ω. In fact, this basis corresponds to the family of Walsh functions defined by $w_j(k) = (-1)^{j \cdot k}$, where the integers j and k ($0 \le j, k \le \ell - 1$) are identified with elements of Ω through their binary representation, and $j \cdot k$ is the inner product of the bit strings. The correspondence is through the identification of an incidence vector j with a set $j \in J$.

The Walsh functions have several useful properties. For all $i, j, k \in \Omega$,

1. $w_i(j) = w_j(i)$
2. $w_i(k)\, w_j(k) = w_{i \oplus j}(k)$
3. $\sum_{k \in \Omega} w_j(k)\, w_i(k) = \begin{cases} 0 & \text{if } i \ne j \\ 2^\ell & \text{if } i = j \end{cases}$

The first and second assertions follow from the representation $w_i(k) = (-1)^{i \cdot k}$. The third follows from the second and the observation that $i \oplus j = 0$ iff $i = j$. Moreover, if $i \ne j$, then

$$|\{k \in \Omega : k \cdot (i \oplus j) \text{ is odd}\}| = |\{k \in \Omega : k \cdot (i \oplus j) \text{ is even}\}|$$

Definition 6. *The Walsh transform $\hat{f}$ of a function $f : \Omega \to R$ is*

$$\hat{f}(j) = 2^{-\ell} \sum_{k \in \Omega} f(k)\, w_j(k)$$

Let $\mathcal{X}_N(x_1 \cdots x_\ell) = \prod_{n \in N} x_n$. The Walsh transform of $\mathcal{X}_N$ is given by:

Lemma 1.

$$\hat{\mathcal{X}}_N(j) = \begin{cases} (-1)^{o(j)}\, 2^{-|N|} & \text{if the incidence vector } j \text{ represents a subset of } N \\ 0 & \text{otherwise} \end{cases}$$

Proof. The terms of

$$\sum_{k \in \Omega} (-1)^{j \cdot k} \prod_{n \in N} k_n$$

are nonzero iff those bits of k indexed by N are all 1, and there are $2^{\ell - |N|}$ such terms. If the incidence vector j represents a subset of N, then all terms are $(-1)^{o(j)}$. If $j_i = 1$ and $i \notin N$, then these term are partitioned by the value of k_i; half are 1 and half are -1. ∎

Definition 7. *The Walsh polynomial $\tilde{f}$ of the polynomial*

$$f(x) = \sum_{N \subset \{1,\ldots,\ell\}} \alpha_N \prod_{n \in N} e_n^T x \quad \text{is} \quad \tilde{f}(y) = \sum_{N \subset \{1,\ldots,\ell\}} \beta_N \prod_{n \in N} e_n^T y$$

where $f(x) = \tilde{f}(y)$ and $y = \vec{1} - 2x$.

Theorem 3. *The Walsh coefficient $\hat{f}(j)$ is equal to that coefficient β_J of $\tilde{f}$ having a subscript corresponding to the incidence vector j. That is,*

$$\tilde{f}(y_1, \ldots, y_\ell) = \sum_j \hat{f}(j) \prod_n y_n^{j_n}$$

Proof. Since the Walsh transform is linear, it suffices to consider the function $\mathcal{X}_N$. Since $x = \frac{1}{2}(\vec{1} - y)$, we have

$$\tilde{\mathcal{X}}_N(y) = \prod_{n \in N} \left\{ \tfrac{1}{2}(1 - y_n) \right\} = \sum_{J \subset N} (-1)^{|J|}\, 2^{-|N|} \prod_{j \in J} y_j$$

Appealing to Lemma 1 finishes the proof. ∎

Definition 7 occurs in Goldberg [10], where the representation provided by Theorem 3 is used to *define* Walsh coefficients. Note that the Walsh polynomial may also be obtained directly (as was the polynomial f) via the reduction

$$\begin{aligned} \tilde{f}(y) &= \tilde{f}(y_1, \ldots, y_\ell) \\ &= \tfrac{1}{2}(1 - y_\ell)\tilde{f}(y_1, \ldots, y_{\ell-1}, -1) + \tfrac{1}{2}(1 + y_\ell)\tilde{f}(y_1, \ldots, y_{\ell-1}, 1) \end{aligned}$$

Walsh polynomials have several desirable properties:

1. Fitness values may be obtained by evaluation ($f(x) = \tilde{f}(y)$).

2. Schema utilities may be obtained by evaluation (Theorem 2 and $f(x) = \tilde{f}(y)$).
3. Walsh coefficients may be obtained by inspection (Theorem 3).
4. Walsh polynomials transform nicely under linear operators on Ω.

This last point is clarified by Theorem 4 below. As previously noted, M-schemata utilities of f can be reduced to Holland utilities of $f \circ M$. It is therefore of interest to compute the transformation of Walsh polynomials under linear maps. Since Walsh polynomials have Walsh coefficients, it suffices to consider how the Walsh transform reacts with linear operators.

Lemma 2. *Let M be a linear operator on Ω. Then* $\widehat{(f \circ M)}(j) = \hat{f}((M^{-1})^T j)$.

Proof. This follows from the observation that

$$\begin{aligned}\sum_{k\in\Omega} f(M\,k)\,(-1)^{j\cdot k} &= \sum_{k\in\Omega} f(k)\,(-1)^{j\cdot M^{-1}k} \\ &= \sum_{k\in\Omega} f(k)\,(-1)^{((M^{-1})^T j)\cdot k}\end{aligned}$$

■

Theorem 4. *If M is an invertible linear operator on Ω, then the Walsh polynomial corresponding to $f \circ M$ is*

$$\sum_j \hat{f}(j) \prod_n y_n^{(M^T j)_n}$$

Proof. By Theorem 3 and Lemma 2, the Walsh polynomial corresponding to $f \circ M$ is

$$\begin{aligned}\sum_j \widehat{(f \circ M)}(j) \prod_n y_n^{j_n} &= \sum_j \hat{f}((M^{-1})^T j) \prod_n y_n^{j_n} \\ &= \sum_j \hat{f}(j) \prod_n y_n^{(((M^{-1})^T)^{-1} j)_n}\end{aligned}$$

The proof is finished by noting that $(M^{-1})^T = (M^T)^{-1}$, and that reducing the exponent of y_n modulo 2 has no effect since $y_n \in \{1, -1\}$. ■

To illustrate, the Walsh polynomial $\tilde{f}(y)$ for our previous example $f(x)$ may be obtained through the substitution $x = \frac{1}{2}(\vec{1} - y)$, which after simplification yields

$$\tilde{f}(y) = \tfrac{17}{48} + \tfrac{1}{48}y_1 + \tfrac{1}{48}y_2 + \tfrac{1}{48}y_3 + \tfrac{3}{16}y_1y_2 + \tfrac{3}{16}y_1y_3 + \tfrac{3}{16}y_2y_3 - \tfrac{7}{48}y_1y_2y_3$$

The utility of the Holland schema $*10$ is obtained by evaluation of either f or $\tilde{f}$:

$$f(\tfrac{1}{2}, 1, 0) = \tilde{f}(0, -1, 1) = \tfrac{17}{48} - \tfrac{1}{48} + \tfrac{1}{48} - \tfrac{3}{16} = \tfrac{1}{6}$$

The Walsh polynomial $\tilde{g}$ for $g = f \circ M$ (where M is our example from Section 4.2) results from mapping the monomials of $\tilde{f}$ by M^T:

$$\begin{array}{rcl} y_1 & \longrightarrow & y_2y_3 \\ y_2 & \longrightarrow & y_1y_3 \\ y_3 & \longrightarrow & y_1y_2y_3 \\ y_1y_2 & \longrightarrow & y_1y_2 \\ y_1y_3 & \longrightarrow & y_1 \\ y_2y_3 & \longrightarrow & y_2 \\ y_1y_2y_3 & \longrightarrow & y_3 \end{array}$$

producing

$$\begin{aligned} \tilde{g}(y_1, y_2, y_3) &= \tfrac{17}{48} + \tfrac{3}{16}y_1 + \tfrac{3}{16}y_2 - \tfrac{7}{48}y_3 + \tfrac{3}{16}y_1y_2 + \tfrac{1}{48}y_1y_3 \\ &\quad + \tfrac{1}{48}y_2y_3 + \tfrac{1}{48}y_1y_2y_3 \end{aligned}$$

If s is the image under M of the Holland schema $*10$, then evaluation gives the utility

$$u_f(s) = u_g(*10) = \tilde{g}(0, -1, 1) = \tfrac{17}{48} - \tfrac{3}{16} - \tfrac{7}{48} - \tfrac{1}{48} = 0$$

4. Basis sets

Fully easy and fully difficult functions are readily generated in terms of the representation provided by polynomial forms. Constructing functions of intermediate difficulty is not as straightforward. We propose a different representation, *basis sets*, which is better suited to the analysis of intermediate difficulty.

In this section we will use N to represent the number of elements in the set Ω of length ℓ binary strings.

4.1 The uniform case

Definition 8. *A collection S of subsets of Ω is a basis iff the incidence vectors that represent the elements of S form a basis for $\mathcal{R}^N$.*

Theorem 5. *Let S be a basis, and u a real-valued function defined on S. There exists a unique function $f : \Omega \to \mathcal{R}$ such that*

$$s \in S \quad \Longrightarrow \quad \frac{1}{|s|} \sum_{x \in s} f(x) \;=\; u(s)$$

Proof. Let A be the matrix having as rows the incidence vectors corresponding to elements of S, let f be the column vector of required function values, and let u be the column vector of given values. The linear system relating function values to schemata utilities is

$$DAf \;=\; u$$

where D is a diagonal matrix containing $|s|^{-1}$ for $s \in S$. Hence f is uniquely determined by $f \;=\; A^{-1}D^{-1}u$. ∎

Definition 9. *Let x be a point in Ω. A schemata path at x is a nested sequence of schemata $x = H_n \subset \cdots \subset H_0 = \Omega$ 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 10. *Let $x \in \Omega$ and let S be a schemata path at x. Then f is increasing at x along S of order (a, b) iff, whenever H and H' are two schemata in S,*

$$a \leq o(H) < o(H') \leq b \implies u_f(H) < u_f(H')$$

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

Observe that fully deceptive functions have a unique optimal x^* and are increasing at x^c of order $n-1$ along all schemata paths. Fully easy functions have a unique optimal x^* and are increasing at x^* of order n along all schemata paths. This follows from the relation

$$u_f(H) = \tfrac{1}{2}\left\{u_f(H_0) + u_f(H_1)\right\}$$

where the H_j are any two schemata of order $1 + o(H)$ that partition H.

Lemma 3. *The collection S of all Holland schemata containing $\vec{0}$ is a basis.*

Proof. For $s \in S$, replace each 0 in s with 1, each $*$ in s with 0, and interpret the result as a binary integer. Using the reduction

$$\tilde{f}(y_1, \ldots, y_\ell) = y_\ell \tilde{f}(y_1, \ldots, y_{\ell-1}, 1) + (1 - y_\ell)\tilde{f}(y_1, \ldots, y_{\ell-1}, 0)$$

let $\tilde{f}$ be that polynomial that maps the binary integer representing s into the utility of s. By construction, evaluating $\tilde{f}$ correctly computes the utilities of schemata in S; hence $\tilde{f}$ may be regarded as a Walsh polynomial. Because the function $f(x)$ corresponding to the Walsh polynomial $\tilde{f}(y)$ exists and is unique (it may be computed via the substitution $y = \vec{1} - 2x$), the linear system relating function values to schemata utilities has a unique solution. Since the coefficients of that linear system are the incidence vectors corresponding to the schemata in S, it follows that S is a basis. ∎

Theorem 6. *Let $x \in \Omega$. The collection of all M-schemata containing x (for fixed M) is a basis.*

Proof. Let σ be the matrix corresponding to the permutation that sends the jth component of a binary vector to the $j \oplus x$th position, where $x \in \Omega$ is fixed. If v is the incidence vector for a Holland schema s, it follows that σv is the incidence vector for the schema $x \oplus s$. 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 3 to any point of Ω. Hence, the collection of all Holland schemata containing x is a basis. Finally, note that mapping the collection of Holland schemata by an invertible linear transformation M also induces a permutation of the components of the corresponding incidence vectors. ∎

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 that without loss of generality is at $\vec{0}$. 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 are not, or could have some regions of deceptiveness either preceded or followed by regions that are nondeceptive. Intuitively, one might expect that the density of nondeceptive 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 $\vec{0}$. By Theorem 6, the collection of all schemata at a point forms a basis, hence the schemata involved in the definitions of these classes are linearly independent. It follows from Theorem 5 that assigning arbitrary utilities to any set of linearly independent schemata will induce a fitness function consistent with the given utilities.

4.2 The non-uniform case

The previous section dealt with schemata in a uniform fashion in that utilities were calculated with respect to the entire space Ω.

In analogy with how the non-uniform Walsh transform can take account of bias introduced by non-uniform populations, the method of Basis Sets has a natural generalization that incorporates these same effects.

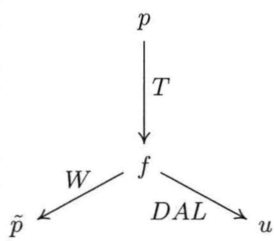
Let L be an N-by-N diagonal matrix with ith entry equal to the number of occurrences of i in the population P, let S be a basis, and let A be the matrix having as rows the incidence vectors corresponding to elements of S. Let D be a diagonal matrix containing the reciprocals of the number of instances (which may be zero) in P of schemata $s \in S$. The linear system relating the vector f of function values to the vector u of schemata utilities with respect to P is

$$DAL\,f \;=\; u$$

The rows containing division by zero correspond to schemata not represented in the population, hence these rows and the corresponding components of u are left undefined. Basis sets in the non-uniform case are analogous to the uniform case, the chief difference being that the consistency of the system $DAL\,f \;=\; u$ replaces the invertibility of the matrix A.

5. Representational relationships

The following diagram summarizes the relationships between the representations discussed in this paper.



Here f denotes the vector of function values, $f_i = f(i)$, p denotes the coefficient vector of the polynomial representation of f,

$$f(x) \;=\; \sum_j p_j \prod_n x_n^{j_n}$$

$\tilde{p}$ denotes the coefficient vector of the Walsh polynomial, $\tilde{p}_j \;=\; \hat{f}(j)$, and u denotes a vector of non-uniform schema utilities for the basis represented by an incidence matrix A and any population corresponding to consistently chosen matrices D and L. The matrix W represents the Walsh transform,

$$W_{i,j} \;=\; 2^{-\ell}\, w_i(j)$$

and T is the evaluation matrix

$$T_{i,j} \;=\; \prod_n i_n^{j_n}$$

where 0^0 is interpreted as 1, and $i, j \in \{0, \ldots, N-1\}$.

If g is obtained from f through a translation by t of the domain Ω

$$g(x) = f(x \oplus t)$$

then

$$\begin{aligned} u_g(s) &= \frac{1}{|s|}\sum_{k\in s} g(k) = \frac{1}{|s|}\sum_{k\in s} f(k\oplus t) = \frac{1}{|s|}\sum_{k\in s\oplus t} f(k) \\ &= u_f(s\oplus t) \end{aligned}$$

It follows that the relationships between the schemata utilities with respect to g are completely isomorphic to those with respect to f. From the perspective of schemata analysis, if we are interested in the basis of those schemata containing some element x of Ω, we may therefore assume $x = \vec{0}$.

Also, since M-schemata utilities for f coincide with Holland schemata utilities for $g = f \circ M$, we may as well assume that it is Holland schemata that are of interest.

Through these reductions, we may assume $A = T$. This is a consequence of the following:

Lemma 4. *The basis of all Holland schemata containing $\vec{0}$ has as incidence matrix the evaluation matrix T. Moreover, the ith row T_i of T is the incidence vector corresponding to that schema obtained from i by replacing each 1 in i with $*$.*

Proof. The matrix T is invertible since f uniquely determines p (see the diagram above). The schemata represented by T all contain $\vec{0}$, since

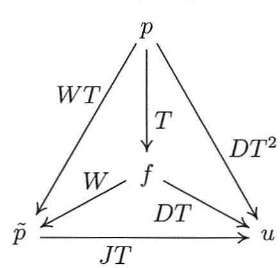
$$T_{i,0} = \prod_n i_n^{0_n} = 1$$

To see that the ith row of T represents the Holland schema obtained from i by replacing each 1 in i with $*$, note that $i_n = 0 \implies j_n = 0$ is a necessary and sufficient condition for $T_{i,j} = 1$. ∎

Let J denote the n-by-n permutation matrix that reverses order,

$$J_{i,j} = \begin{cases} 1 & \text{if } 1 + i + j = n \\ 0 & \text{otherwise} \end{cases}$$

Theorem 7. *If u represents uniform utilities for the basis of Holland schemata containing $\vec{0}$, then the following diagram commutes:*



Proof. Let $[x]^{y_1,\dots}_{z_1,\dots}$ denote x where all occurrences of $y_1, \dots$ have been replaced by $z_1, \dots$. We have

$$\begin{aligned}(DTf)_i &= u_i = u_f([i]^1_*) = \tilde{p}([[i]^1_*]^{0,1,*}_{1,-1,0}) = \tilde{p}([i]^{0,1}_{1,0}) = (T\tilde{p})_{[i]^{0,1}_{1,0}} \\ &= (JT\tilde{p})_i\end{aligned}$$

The first equality is from the diagram, the second is from Lemma 4, and the third follows from Theorem 2 and the definition of the Walsh polynomial. The fourth equality is through simplification, the fifth is by construction since T was designed to evaluate polynomials, and the last follows from the observation that

$$[x]^{0,1}_{1,0} = 2^\ell - 1 - x$$

This establishes the commutivity of the lower triangle. The commutivity of the right and left triangles is by inspection. ∎

6. Summary

We have investigated three forms for the objective function of a genetic algorithm: polynomial forms, basis sets, and Walsh polynomials.

Our motivating perspective has been deceptiveness. We defined various degrees of deceptiveness and, since representational changes can affect deceptiveness, we provided a framework that encompasses the representational change operators of Liepins and Vose [13] by means of M-schemata.

We illustrated the use of polynomial forms by using them to characterize three classes of fully easy functions. We applied basis sets to the construction of functions with specified predetermined difficulty. We demonstrated a linear interrelationship between these forms and traditional Walsh polynomials through a commutative diagram involving matrices that transform between them.

References

[1] R. G. Bartle, *The Elements of Integration* (Wiley, New York, 1966).

[2] D. Battle and M. D. Vose, "Isomorphisms of Genetic Algorithms," Workshop on the Foundations of Genetic Algorithms and Classifier Systems (Bloomington, Indiana, July 1990).

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

[4] C. L. Bridges and D. E. Goldberg, "A Note on the Non-uniform Walsh-Schema Theorem," TCGA Report No. 89004 (University of Alabama, The Clearinghouse for Genetic Algorithms, Tuscaloosa, 1989).

[5] J. P. Cohoon and W. D. Paris, "Genetic Placement," *Proceedings of the IEEE International Conference on Computer-Aided Design* (1986), 422–425.

[6] S. Coombs and L. Davis (1987), "Genetic Algorithms and Communication Link Speed Design: Constraints and Operators," *Genetic Algorithms and Their Applications: Proceedings of the Second International Conference on Genetic Algorithms* (1987), 257–260.

[7] C.-H.Chu, "A Genetic Algorithm Approach to the Configuration of Stack Filters," in Schaffer (ed.), *Genetic Algorithms* (Morgan Kaufmann, San Mateo, 1989), 219–224.

[8] L. Davis and S. Coombs (1987), "Genetic Algorithms and Communication Link Speed Design: Theoretical Considerations," *Genetic Algorithms and Their Applications: Proceedings of the Second International Conference on Genetic Algorithms* (1987), 252–256.

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

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

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

[12] S. Lang, *Algebra* (Addison Wesley, Reading, Mass., 1971).

[13] G. E. Liepins and M. D. Vose, "Representational Issues in Genetic Optimization," *Journal of Experimental and Theoretical Artificial Intelligence*, **2**(2) (1990) 4–30.

[14] D. J. Powell, S. S. Tong, and M. M. Skolnick, "EnGENEous Domain Independent Machine Learning for Design Optimization," *Proceedings of the Third International Conference on Genetic Algorithms* (1989), 151–169.