Polynomial Time Summary Statistics for Two General Classes of Functions

Robert B. Heckendorn
Department of Computer Science
University of Idaho
Moscow, ID 83844-1010 USA
heckendo@cs.uidaho.edu

Abstract

In previous work we showed, by Walsh analysis, that summary statistics such as mean, variance, skew, and higher order statistics can be computed in polynomial time for embedded landscapes. We then used these statistics to study the epistatic structure of MAXSAT problems. These results were dependent on two facts: these functions have a polynomial number of nonzero Walsh coefficients and the coefficients can be computed in polynomial time. It has since been shown that for any arbitrary function in which the number of epistatically interacting bits is bounded above by k the nonzero Walsh coefficients are also polynomial in number and can be computed in polynomial time. This extends the applicability of our earlier results.

In this paper, I extend these results further to include hyperplane statistics. These statistics can help us understand the hyperplane structure of sparsely epistatic functions as well as functions of bounded epistasis.

1 INTRODUCTION

Summary statistics (mean, variance, skew, ...) tell us about the location, spread, symmetry, etc of a distribution of numbers. When applied to fitness function they give us statistical information on the expected value of a random sample, as in an initial population, and how much we can expect elements in our sample, on average, to deviate in both directions from the mean. Furthermore, it is widely believed [Whitley et al., 1995, Holland, 1975] that GAs work by a form of implicit parallelism. That is, all hyperplanes in a given population compete in parallel for representation in succeeding populations. The effectiveness of this competition is limited by several factors including the ability of a sample of elements from the hyperplane to represent the true fitness of the entire hyperplane. Summary statistics about a specific hyperplane would be useful in determining how accurately a sampling of elements from a hyperplane may statistical "stand in" for that hyperplane in a population.

Clearly, directly computing the **summary statistics** (mean, variance, skew, ...) for arbitrary fitness functions would require exponential time relative to the size of the domain in bits. The same could be said for low order hyperplanes. In this paper, I show that for two large and useful subclasses of functions, this computation can be done in polynomial time by using Walsh analysis. Specifically, I show that three kinds of statistical moments can be computed in polynomial time for any function in which there are at most a polynomially bounded number of nonzero Walsh coefficients. Previous work has shown that these Walsh coefficients can be computed in polynomial time for both embedded landscapes, which include NK-landscapes and MAXSAT problems, and for functions of bounded epistasis.

An **embedded landscape** over L bits, $f : \mathcal{B}^L \rightarrow \mathbb{R}$, models important broad classes of combinatorial and constraint satisfaction problems. It can be expressed as the sum of P subfunctions, g_i:

$$f(x) = \sum_{j=1}^{P} g_j(pack(x, m_j)).$$

The $pack$ function uses a bit mask, m_j, to extract the subset of bits from x indicated by the `1`'s in the mask. The extracted bits form the arguments to the subfunctions g_j. If $bc(z)$ is a function that returns the number of `1`'s in its argument, z, then the subfunctions are defined over a smaller set of $bc(m_j)$ bits, $g_j : \mathcal{B}^{bc(m_j)} \rightarrow \mathbb{R}$. There are no restrictions on the number of subfunctions, P. The g_i generally have lower dimensional domains than f and are hence considered to be **embedded** in higher dimensional space. For example, there often exists a k, $k \ll L$, such that $k \geq \max_{j=1..P} bc(m_j)$. This occurs in MAXSAT problems where the clause size is limited to k variables, which is far fewer than the total number of variables available.

For a fixed value of k, we have shown that all of the Walsh coefficients of an embedded landscape can be computed in polynomial time relative to 2^k, provided the subfunction masks are known in advance. Since k is fixed independently of L, the Walsh coefficients can be computed in polynomial time relative to the function size in number of bits. We also showed that only a polynomial number of Walsh coefficients were nonzero.

Epistatically bounded functions are functions in which the number of epistatically interacting bits is bound above by k. Technically, this class of functions can be modeled by embedded landscapes in which $bc(m_j) \leq k, \forall j$. However, the number of subfunctions necessary to model an arbitrary L bit function whose epistatic interactions are bound by k bits is $\binom{L}{k}$ where most of the values of the subfunction provide redundant epistatic information. This makes embedded landscapes impractical for the most general version of this class of function.

Kargupta et al. [Kargupta and Park, 1999] have shown that for epistatically bounded functions, all the Walsh coefficients can be computed in polynomial time relative to 2^k. Again, this can be considered polynomial time in cases where the epistatic bound k is a fixed constant independent of L. As in the case of embedded landscapes, there are a polynomial number of nonzero Walsh coefficients.

Kargupta's algorithm uses hyperplane averages, that is, the average value of the function for all values in the hyperplane (see the notation section), to probe a function's epistatic structure. Using this technique the algorithm does not need to know *a priori* the epistatic structure of the function. However, in order to perform in polynomial time, the algorithm makes the statistically reasonable assumption that if there exists any nonzero Walsh coefficient in the hyperplane, the hyperplane average will be nonzero. For many functions with essentially random Walsh coefficients or for functions whose Walsh coefficients are all of one sign this algorithm can be guaranteed to work in polynomial time.

Soraya Rana suggested that the r^{th} moments of embedded landscapes might also be computed in polynomial time. Subsequently, we showed [Heckendorn et al., 1999a] that this was indeed the case. We observed that MAXSAT problems ($k \geq 2$), which are members of the set of NP-complete problems, were also embedded landscapes. This meant that since we can compute all of the Walsh coefficients, as well as the summary statistics for an NP-complete problem in polynomial time, that either NP is P or knowing the Walsh coefficients and summary statistics is insufficient to discover the optimum of the problem in polynomial time [Heckendorn et al., 1999a]. This is an important result on the limits of the usefulness of epistatic information.

Our proofs hinged only on the facts that, for the functions of interest, there were at most a polynomial number of nonzero Walsh coefficients and that their values could be discovered in polynomial time. With the introduction of Kargupta's algorithm our results extend to all epistatically bounded functions.

In the next section, I will briefly review the terminology of embedded landscapes and the notation of Walsh analysis. I will then recapitulate the proof the r^{th} moments of embedded landscapes that we presented in [Heckendorn et al., 1999a]. Using this theorem as a model, I provide theorems showing how to compute two kinds of statistical moments for hyperplanes polynomial time.

2 WALSH ANALYSIS AND NOTATION

The **Walsh transform** is the analog to the discrete Fourier transform but, for functions whose domain is a bit string. Every real valued function f over an L-bit string, $f : \mathcal{B}^L \rightarrow \mathbb{R}$, can be expressed as a weighted sum of a set of 2^L orthogonal functions called **Walsh functions**.

$$f(x) = \sum_{j=0}^{2^L-1} w_j \psi_j(x) \qquad (1)$$

where the Walsh Functions are denoted $\psi_j : \mathcal{B}^L \rightarrow \{-1, 1\}$. The Walsh functions play the role that sine and cosine play in the Fourier transform. The weights $w_j \in \mathbb{R}$ are called **Walsh coefficients**. The indices of both Walsh functions and coefficients may be expressed as either binary or the numerical equivalent.

The j^{th} Walsh function can be defined:

$$\psi_j(x) = (-1)^{bc(j \wedge x)}$$

where $j, x \in \mathcal{B}^L$. Thus, if $bc(j \wedge x)$ is odd, then $\psi_j(x) = -1$ and if $bc(j \wedge x)$ is even, then $\psi_j(x) = 1$. The j^{th} Walsh function looks at the parity of the bits selected by j. Hence, there are 2^L Walsh functions.

An important property of Walsh coefficients is that w_j measures the contribution to the evaluation function by the interaction of the bits indicated by the positions of the 1's in j. Thus, w_{0001} measures the linear contribution to the evaluation function associated with bit position 0, while w_{0101} measures the nonlinear (multiple bit) interaction between the bits in positions 0 and 2, and so on. Therefore, Equation 1 says that any function over L bit space can be represented as a weighted sum of all possible 2^L bit interaction functions ψ_j. This nonlinearity is an important feature in determining problem difficulty for genetic algorithms [Goldberg, 1989a, Goldberg, 1989b, Reeves and Wright, 1995].

The 2^L Walsh coefficients can be computed by a Walsh transform:

$$w_j = \frac{1}{2^L} \sum_{x=0}^{2^L-1} f(x)\psi_j(x) \tag{2}$$

A **hyperplane** is a subset of values (or bit strings) from the domain that share a common set of fixed bits. A hyperplane can be represented by one of the 3^L strings of 0's, 1's and *'s where the 0's and 1's are in the **fixed bit positions** and the *'s represent either a 0 or a 1 in the **variable bit positions**. An example hyperplane h for strings in $\mathcal{B}^7$ might be **1101* which contains eight strings two of which are 1111011 and 1011010. A hyperplane with C fixed bit positions represents a hyperplane of **order** C and defines a set of $2^{(L-C)}$ strings where all possible replacements of the *'s have been defined. For hyperplane h the order of h is denoted by $o(h)$ and the number of strings in h is denoted $|h|$.

Two important functions α and β can be defined on a hyperplane by the bit by bit mappings below [Goldberg, 1989c]:

$$\alpha(h)[i] = \begin{cases} 0 & h[i] = * \\ 1 & h[i] = 0 \text{ or } 1 \end{cases}$$

$$\beta(h)[i] = \begin{cases} 0 & h[i] = * \text{ or } 0 \\ 1 & h[i] = 1 \end{cases}$$

The α returns a mask that identifies the fixed bit positions in the hyperplane. β returns a mask that identifies the bit positions that are set to 1. For the hyperplane h = **1101* : $\alpha(h)$ = 0011110 and $\beta(h)$ = 0011010.

Finally, there is a notation for bit containment [Rana et al., 1998]: $i \subseteq j$ where $i, j \in \mathcal{B}^L$ reads as i **is contained in** j. That is, wherever there is a 1 in i there is a 1 in j or, said another way, this bitwise logical statement is true: $i \wedge \overline{j} = 0$.

3 SUMMARY STATISTICS TECHNIQUES

In this section, I show how summary statistics such as skew and kurtosis can be also be computed from the Walsh coefficients by using a general formula for computing the r^{th} moment for any function where all the nonzero Walsh coefficients are known.

Theorem 1 Moment about Function Mean

The r^{th} moment, denoted μ_r, for a fitness function, $f : \mathcal{B}^L \rightarrow \mathbb{R}$ whose Walsh coefficients are w_j is

$$\mu_r = \sum_{a_1 \oplus a_2 \oplus \ldots \oplus a_r = 0} w_{a_1} w_{a_2} \ldots w_{a_r}, \qquad a_i \neq 0 \; \forall \; i$$

where $\oplus$ is the EXCLUSIVE-OR operator.

Proof:

Given the mean, μ, the formula used to compute the r^{th} moment, denoted μ_r, for a discrete random variable X is:

$$\mu_r \;=\; E[(X-\mu)^r] \;=\; \sum_{x \in X} (x-\mu)^r p(x)$$

where $p(x)$ is the probability of x occurring [Mendenhall, 1967]. We can consider $p(x) = \frac{1}{2^L}$ since we are enumerating a function over all L bit binary strings. The function then becomes:

$$\mu_r = \sum_{x \in X} \frac{(x-\mu)^r}{2^L}$$

If X represents a real valued function over an L bit domain then:

$$\mu_r = \frac{1}{2^L} \sum_{x=0}^{2^L-1} (f(x)-\mu)^r$$

We can substitute for f with the linear Walsh representation of f from Equation 1:

$$\mu_r = \frac{1}{2^L} \sum_{x=0}^{2^L-1} \left(\sum_{i=0}^{2^L-1} w_i \psi_i(x) - \mu \right)^r$$

Since $\psi_0(x) = 1 \; \forall x$, we see from Equation 2 that Walsh coefficient w_0 is the mean of all fitnesses. Therefore

$$\mu_r = \frac{1}{2^L} \sum_{x=0}^{2^L-1} \left(\sum_{i=1}^{2^L-1} w_i \psi_i(x) \right)^r$$

We can now expand the exponential creating a set of r indices a_j where $a_j \in \mathcal{B}^L$:

$$\mu_r = \frac{1}{2^L} \sum_{x=0}^{2^L-1} \left(\sum_{a_1=1}^{2^L-1} w_{a_1} \psi_{a_1}(x) \right) \left(\sum_{a_2=1}^{2^L-1} w_{a_2} \psi_{a_2}(x) \right) \ldots$$

$$\ldots \left(\sum_{a_r=1}^{2^L-1} w_{a_r} \psi_{a_r}(x) \right)$$

Since the Walsh coefficients do not depend on x, the formula can be rewritten as:

$$\mu_r = \frac{1}{2^L} \sum_{a_1=1}^{2^L-1} \sum_{a_2=1}^{2^L-1} \cdots \sum_{a_r=1}^{2^L-1} w_{a_1} w_{a_2} \ldots$$

$$\ldots w_{a_r} \sum_{x=0}^{2^L-1} \psi_{a_1}(x) \psi_{a_2}(x) \ldots \psi_{a_r}(x)$$

Using the fact that for arbitrary p and q: $\psi_p(x)\psi_q(x) = \psi_{p \oplus q}(x)$:

$$\mu_r = \frac{1}{2^L} \sum_{a_1=1}^{2^L-1} \sum_{a_2=1}^{2^L-1} \cdots \sum_{a_r=1}^{2^L-1} w_{a_1} w_{a_2} \ldots$$

$$\ldots w_{a_r} \sum_{x=0}^{2^L-1} \psi_{a_1 \oplus a_2 \oplus .. \oplus a_r}(x)$$

Now using the fact that:

$$\sum_{x=0}^{2^L-1} \psi_i(x) = \begin{cases} 0 & \text{if } i \neq 0 \\ 2^L & \text{if } i = 0 \end{cases}$$

we see that **only when** $a_1 \oplus a_2 \oplus \ldots \oplus a_r = 0$ is the inner sum nonzero. Therefore,

$$\mu_r = \frac{1}{2^L} \sum_{a_1 \oplus a_2 \oplus \ldots \oplus a_r = 0} w_{a_1} w_{a_2} \ldots w_{a_r} 2^L, \qquad a_i \neq 0 \; \forall \; i$$

$$= \sum_{a_1 \oplus a_2 \oplus \ldots \oplus a_r = 0} w_{a_1} w_{a_2} \ldots w_{a_r}, \qquad a_i \neq 0 \; \forall \; i \qquad (3)$$

□

To summarize, given the set of nonzero Walsh coefficients, we can compute the r^{th} moment for the fitness distribution using products of the Walsh coefficients such that the EXCLUSIVE-OR of the indices is zero.

This formula allows us to compute the variance, skew and kurtosis for any fitness distribution provided we are given the Walsh coefficients.

$$\text{variance} = \mu_2 = \sigma^2 \qquad \text{skew} = \frac{\mu_3}{\sigma^3} \qquad \text{kurtosis} = \frac{\mu_4}{\sigma^4}$$

For example, since $a_1 \oplus a_2 = 0$ if and only if $a_1 = a_2$ then the variance for any function can be computed

$$\sum_{i=1}^{2^L-1} w_i w_i$$

Of course, the computation of the moment around the mean, if done directly, would take $\mathcal{O}(2^{Lr})$ time. However, in the case of functions with only a polynomial number of nonzero Walsh coefficients, the nonzero coefficients are easily enumerated. Only the nonzero coefficients need be considered in the moment calculations.

In the case of embedded landscapes, the Walsh coefficients are computed for each subfunction in $\mathcal{O}(k2^k)$ using a fast Walsh transform [Goldberg, 1989a]. So for P functions the Walsh coefficients can be computed in $\mathcal{O}(Pk2^k)$ where P is bounded by $\binom{L}{k}$ for L bit functions [Heckendorn et al., 1999b]. Even though the computation time is bounded above by $\mathcal{O}(\binom{L}{k}k2^k)$, this is quite practical when compared to the alternative provided by a straight Fast Walsh transform of $\mathcal{O}(L2^L)$. For example, if $k = 3$ then the execution time for the embedded landscape approach is $\mathcal{O}(L^3)$.

In the case of the more general Kargupta's algorithm for a function that is epistatically bounded by k, the number of nonzero Walsh coefficients is $\mathcal{O}\binom{L}{k}$. All of the k-order Walsh coefficients can be computed by averaging 2^k function evaluations and walsh function evaluations (assuming random Walsh coefficients) for each of the $\binom{L}{k}$ coefficients. The $(k-1)$-order Walsh coefficients can now be computed using 2^{k-1} function evaluations and subtracting away the effects of the k order k Walsh coefficients. This process can certainly be done in $\mathcal{O}(kL\binom{L}{k}2^k)$ operations, if k is known in advance.

Given that the nonzero Walsh coefficients are now identified in both classes of functions, Theorem 1 can clearly be used to compute the r^{th} moment in $\mathcal{O}(n^r)$ time, where n is the number of nonzero Walsh coefficients. Since n is polynomial in size relative to L so is n^r for fixed r independent of L. Since both the Walsh coefficient calculation and the moment computation can be carried out in polynomial time for fixed r, any summary statistics can be computed from Theorem 1 in polynomial time.

4 HYPERPLANE STATISTICS

A similar approach to that which was used in the last section can be used to calculate summary statistics for a given hyperplane. Hyperplane statistics can be used to study the distribution of hyperplane fitnesses and make statistical inferences about the effectiveness of hyperplane sampling. There are two types of moments for a hyperplane: the moment about the mean of the entire function and the moment about the mean of just the hyperplane itself. We will treat these two cases in that order.

Theorem 2 Moment of Hyperplane about the Function Mean

The r^{th} moment of the elements of hyperplane h about the mean μ for function f given the Walsh coefficients of f is:

$$\mu_r(h) = \sum_{a_1 \oplus \ldots \oplus a_r \subseteq \alpha(h)} w_{a_1} \ldots w_{a_r} \psi_{a_1 \oplus \ldots \oplus a_r}(\beta(h)), \;\; a_i \neq 0 \;\forall\; i$$

Proof:

From the definition of r^{th} moment and assuming an equal probability of selecting any domain value in the hyperplane:

$$\mu_r(h) = \frac{1}{|h|} \sum_{x \in h} (f(x) - \mu)^r$$

where μ is the mean for the entire function. We now proceed as with the earlier derivation:

$$\mu_r(h) = \frac{1}{|h|} \sum_{a_1=1}^{2^L-1} \sum_{a_2=1}^{2^L-1} \cdots \sum_{a_r=1}^{2^L-1} w_{a_1} w_{a_2} \ldots$$
$$\ldots w_{a_r} \sum_{x \in h} \psi_{a_1 \oplus a_2 \oplus .. \oplus a_r}(x)$$

Using the fact that [Heckendorn and Whitley, 1999]

$$\sum_{x \in h} \psi_j(x) = \begin{cases} 0 & \text{if } \; j \not\subseteq \alpha(h) \\ \psi_j(\beta(h))|h| & \text{if } \; j \subseteq \alpha(h) \end{cases}$$

we get:

$$|h|\mu_r(h) = \sum_{a_1 \oplus a_2 \oplus \ldots \oplus a_r \subseteq \alpha(h)} w_{a_1} \ldots w_{a_r} (\psi_{a_1 \oplus \ldots \oplus a_r}(\beta(h))|h|),$$
$$\text{where} \qquad a_i \neq 0 \;\forall\; i$$

Therefore, the r^{th} moment about the mean for the entire function over hyperplane h is:

$$\mu_r(h) = \sum_{a_1 \oplus a_2 \oplus \ldots \oplus a_r \subseteq \alpha(h)} w_{a_1} w_{a_2} \ldots w_{a_r} \psi_{a_1 \oplus a_2 \oplus \ldots \oplus a_r}(\beta(h)),$$
$$\text{where} \qquad a_i \neq 0 \;\forall\; i$$

□

Now consider the case where the mean used in the moment calculations is the mean of the hyperplane. We denote this moment for hyperplane h about the mean of h as $\widehat{\mu}_r(h)$.

Theorem 3 Moment of Hyperplane about Hyperplane Mean

The moment of the elements of hyperplane h about the mean, $\widehat{\mu}$, for hyperplane h in terms of the Walsh coefficients of f is:

$$\widehat{\mu}_r(h) = \sum_{a_1 \oplus a_2 \oplus \ldots \oplus a_r \subseteq \alpha(h)} w_{a_1} w_{a_2} \ldots w_{a_r} \psi_{a_1 \oplus a_2 \oplus \ldots \oplus a_r}(\beta(h)),$$
$$\text{where} \qquad a_i \not\subseteq \alpha(h) \;\forall\; i$$

Proof:

Returning to the original equation for moment we get:

$$\widehat{\mu}_r(h) = \frac{1}{|h|}\sum_{x\in h}(f(x)-\widehat{\mu})^r$$

The Hyperplane Averaging theorem [Heckendorn and Whitley, 1999] states:

$$\frac{1}{|h|}\sum_{x\in h} f(x) = \sum_{j\subseteq\alpha(h)} w_j\psi_j(\beta(h))$$

Substituting the Walsh transform for the function f and using the Hyperplane Averaging theorem for $\widehat{\mu}$ we get:

$$\widehat{\mu}_r(h) = \frac{1}{|h|}\sum_{x\in h}\left(\sum_{i=0}^{2^L-1} w_i\psi_i(x) - \sum_{k\subseteq\alpha(h)} w_k\psi_k(\beta(h))\right)^r$$

The left sum in parentheses can now be broken into two parts

$$\widehat{\mu}_r(h) = \frac{1}{|h|}\sum_{x\in h}\left(\sum_{i\subseteq\alpha(h)} w_i\psi_i(x) + \sum_{j\not\subseteq\alpha(h)} w_j\psi_j(x) - \sum_{k\subseteq\alpha(h)} w_k\psi_k(\beta(h))\right)^r$$

Regrouping under the sums gives

$$\widehat{\mu}_r(h) = \frac{1}{|h|}\sum_{x\in h}\left(\sum_{i\subseteq\alpha(h)}\big(w_i\psi_i(x) - w_i\psi_i(\beta(h))\big) + \sum_{j\not\subseteq\alpha(h)} w_j\psi_j(x)\right)^r$$

Note that $x \in h$ and $i \subseteq \alpha(h)$ therefore, $\psi_i(x) = \psi_i(\beta(h))$! This means $w_i\psi_i(x) - w_i\psi_i(\beta(h))$ is zero and we get

$$\widehat{\mu}_r(h) = \frac{1}{|h|}\sum_{x\in h}\left(\sum_{j\not\subseteq\alpha(h)} w_j\psi_j(x)\right)^r$$

Proceeding with the expansion of the r^{th} power as we did in the earlier proofs:

$$\begin{aligned}
\widehat{\mu}_r(h) &= \frac{1}{|h|}\sum_{x\in h}\sum_{a_1\not\subseteq\alpha(h)}\sum_{a_2\not\subseteq\alpha(h)}\cdots \\
&\quad\cdots\sum_{a_r\not\subseteq\alpha(h)} w_{a_1}w_{a_2}\ldots w_{a_r}\psi_{a_1\oplus a_2\oplus\ldots\oplus a_r}(x) \\
&= \frac{1}{|h|}\sum_{a_1\not\subseteq\alpha(h)}\sum_{a_2\not\subseteq\alpha(h)}\cdots \\
&\quad\cdots\sum_{a_r\not\subseteq\alpha(h)} w_{a_1}w_{a_2}\ldots w_{a_r}\sum_{x\in h}\psi_{a_1\oplus a_2\oplus\ldots\oplus a_r}(x) \\
&= \frac{1}{|h|}\sum_{a_1\not\subseteq\alpha(h)}\sum_{a_2\not\subseteq\alpha(h)}\cdots \\
&\quad\cdots\sum_{a_r\not\subseteq\alpha(h)} w_{a_1}w_{a_2}\ldots w_{a_r}|h|\psi_{a_1\oplus a_2\oplus\ldots\oplus a_r}(\beta(h)), \\
&\qquad\text{with } a_1\oplus a_2\oplus\ldots\oplus a_r\subseteq\alpha(h) \\
&= \sum_{a_1\not\subseteq\alpha(h)}\sum_{a_2\not\subseteq\alpha(h)}\cdots \\
&\quad\cdots\sum_{a_r\not\subseteq\alpha(h)} w_{a_1}w_{a_2}\ldots w_{a_r}\psi_{a_1\oplus a_2\oplus\ldots\oplus a_r}(\beta(h)), \\
&\qquad\text{with } a_1\oplus a_2\oplus\ldots\oplus a_r\subseteq\alpha(h)
\end{aligned}$$

Which is the same as saying:

$$\widehat{\mu}_r(h) = \sum_{a_1\oplus a_2\oplus\ldots\oplus a_r\subseteq\alpha(h)} w_{a_1}w_{a_2}\ldots w_{a_r}\psi_{a_1\oplus a_2\oplus\ldots\oplus a_r}(\beta(h)),$$

$$\text{where}\qquad a_i\not\subseteq\alpha(h)\ \forall\ i$$

□

Note that when h is fixed as all $*$'s, that is, h is the whole domain then, $\alpha(h) = 0$ and the Moment of Hyperplane about Hyperplane Mean Theorem becomes the same as the first theorem in this paper.

The execution time for the statistics from the last two theorems is again polynomial in the number of nonzero Walsh coefficients. Therefore, the total execution time to compute the statistic for an embedded landscape is polynomial in the number of bits in the domain. The actual selection of the a_i's makes the computation $\mathcal{O}(n^r)$ where n is the number of nonzero Walsh coefficients.

Several observations can be made about efficiently computing the sets of a_i's for the last two theorems.

An observation can be made about the efficiently computing the a_i's for the last two theorems. The theorems require that we select a_r such that $a_1 \oplus a_2 \oplus \ldots \oplus a_r \subseteq \alpha(h)$. If A is the EXCLUSIVE-OR of the first $r-1$ values then our goal is $(A \oplus a_r) \subseteq \alpha(h)$. Therefore, $(A \oplus a_r) \wedge \overline{\alpha(h)} = 0$ so $(A \wedge \overline{\alpha(h)}) \oplus (a_r \wedge \overline{\alpha(h)}) = 0$ or $(A \wedge \overline{\alpha(h)}) = (a_r \wedge \overline{\alpha(h)})$. This says the bits in the positions selected by the `1`'s in $\overline{\alpha(h)}$ in a_r must be the same as in A. This offers some opportunities for efficiently searching through the indices of nonzero Walsh coefficients by sorting the coefficients by bits of each index that are in $\overline{\alpha(h)}$.

5 STATISTICS BY PARTITION

The domain of a function, $f : \mathcal{B}^L \to \mathbb{R}$, can be partitioned onto a set of nonintersecting hyperplanes that completely covers the domain. A **partition** is specified by a string with a `b` in the positions called **fixed bit positions** and `*`'s in the remaining positions. For example, `*b**` represents a partition that contains the two hyperplanes `*1**` and `*0**`. A partition with M `b`'s defines a set of 2^M nonintersecting hyperplanes, each composed of 2^{L-M} strings whose union is all of the strings in the domain. It is hyperplane competition within each partition and the ranking of the supporting hyperplanes in a partition that significantly influences the direction of convergence of a GA [Heckendorn et al., 1997].

For a given partition, all of the hyperplanes in the partition will have the same α but they will all have unique a β. This means for any hyperplane in a fixed partition the hyperplane statistics for that hyperplane are computed by summing over the same set of a_i. Only the values $\beta(h)$ change. Or said another way, the difference in the hyperplane statistics for two hyperplanes h_1 and h_2, both from the same partition is the sign of the products of the Walsh coefficients that are summed. The sign being determined by:

$$\psi_{a_1 \oplus a_2 \oplus \ldots \oplus a_r}(\beta(h_i))$$

This means that the hyperplane statistics for all hyperplanes in a partition can be quickly computed without any extra multiplication of Walsh coefficients or deciding which indices, a_i, to sum over. This allows us to quickly compute comparative statistics between competing hyperplanes in a given partition.

6 CONCLUSIONS

We have shown that the summary statistics can be computed in polynomial time for not only embedded landscapes but for epistatically bounded functions by using Kargupta's algorithm. This work was further extended to show that two new summary statistics for hyperplanes can be similarly computed in polynomial time. Finally, it was observed that these statistics can be efficiently computed simultaneously for all hyperplanes in a given partition.

References

[Goldberg, 1989a] Goldberg, D. (1989a). Genetic algorithms and walsh functions: Part i, a gentle introduction. *Complex Systems*, 3:129–152.

[Goldberg, 1989b] Goldberg, D. (1989b). Genetic algorithms and walsh functions: Part ii, deception and its analysis. *Complex Systems*, 3:153–171.

[Goldberg, 1989c] Goldberg, D. (1989c). *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Publishing, Co., Reading, MA.

[Heckendorn et al., 1999a] Heckendorn, R. B., Rana, S., and Whitley, D. (1999a). Polynomial time summary statistics for a generalization of maxsat. In Banzhaf, W., Daida, J., Eiben, A. E., Garzon, M. H., Honavar, V., Jakiela, M., and Smith, R. E., editors, *Proc. of the Genetic and Evolutionary Computation Conference*, volume 1, pages 281–288, Palo Alto, CA. Morgan Kaufmann Publishers, Inc.

[Heckendorn et al., 1999b] Heckendorn, R. B., Rana, S., and Whitley, D. (1999b). Test function generators as embedded landscapes. In Bäck, T. and Banzhaf, W., editors, *FOGA - 5*, Palo Alto, CA. Morgan Kaufmann Publishers, Inc.

[Heckendorn and Whitley, 1999] Heckendorn, R. B. and Whitley, D. (1999). Walsh functions and predicting problem complexity. *Evolutionary Computation*, 7(1):69–101.

[Heckendorn et al., 1997] Heckendorn, R. B., Whitley, L. D., and Rana, S. (1997). Nonlinearity, hyperplane ranking and the simple genetic algorithm. In Belew, R. K. and Vose, M., editors, *Foundations of Genetic Algorithms - 4*, Palo Alto, CA. Morgan Kaufmann Publishers, Inc.

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

[Kargupta and Park, 1999] Kargupta, H. and Park, B. H. (1999). Fast construction of distributed and decomposed evolutionary representation. In Brave, S. and Wu, A. S., editors, *Late Breaking Papers at the 1999 Genetic and Evolutionary Computation Conference*, pages 139–148, Palo Alto, CA. Morgan Kaufmann Publishers, Inc.

[Mendenhall, 1967] Mendenhall, W. (1967). *Introduction to Probability and Statistics*. Wadsworth Publishing Company, Inc., Belmont, CA, second edition.

[Rana et al., 1998] Rana, S., Heckendorn, R. B., and Whitley, D. (1998). A tractable walsh analysis of sat and its implications for genetic algorithms. In *Proceedings of the Fifteenth National Conference on Artificial Intelligence*, pages 392–397, Menlo Park, CA. AAAI Press.

[Reeves and Wright, 1995] Reeves, C. and Wright, C. (1995). An experimental design perspective on genetic algorithms. In Whitley, D. and Vose, M., editors, *Foundations of Genetic Algorithms - 3*, pages 7–22. Morgan Kaufmann Publishers, Inc.

[Whitley et al., 1995] Whitley, D., Mathias, K., and Pyeatt, L. (1995). Hyperplane ranking in simple genetic algorithms. In Eschelman, L. J., editor, *Proc. of the 6th Int'l. Conf. on GAs*. Morgan Kaufmann Publishers, Inc.