Computer Science
Technical Report

Computing the moments of k-bounded pseudo-Boolean functions over Hamming spheres of arbitrary radius in polynomial time

Andrew M. Sutton, L. Darrell Whitley, Adele E. Howe

June 24, 2010

Technical Report CS-10-104

Computer Science Department
Colorado State University
Fort Collins, CO 80523-1873

Phone: (970) 491-5792 Fax: (970) 491-2466
WWW: http://www.cs.colostate.edu

Computing the moments of k-bounded pseudo-Boolean functions over Hamming spheres of arbitrary radius in polynomial time

Andrew M. Sutton, L. Darrell Whitley, and Adele E. Howe

Department of Computer Science
Colorado State University, Fort Collins, CO USA
email: {sutton,whitley,howe}@cs.colostate.edu

June 24, 2010

Abstract

We show that given a k-bounded pseudo-Boolean function f, we can always compute the c^{th} moment of f over regions of arbitrary radius in Hamming space in polynomial time using algebraic information from the adjacency structure (where k and c are constants). This result has implications for evolutionary algorithms and local search algorithms because information about promising regions of the search space can be efficiently retrieved, even if the cardinality of the region is exponential in the problem size. Finally, we use our results to introduce a method of efficiently calculating the expected fitness of mutations for genetic algorithms.

1 Introduction

The class of k-bounded pseudo-Boolean functions (i.e., real valued functions over binary strings that are epistatically bounded by a constant k) plays an important role in evolutionary computation, combinatorial optimization, biophysics, and machine learning. These functions appear as objective functions in a number of well-studied combinatorial optimization problems over the set of binary strings, e.g., MAX-k-SAT problems, NK-landscapes, spin models, and several graph optimization problems such as MAX-CUT.

In the fields of evolutionary computation and local search, a *landscape* is a mathematical formalism which can be expressed as a tuple $(\mathcal{X}, \mathcal{N}, f)$. Here $\mathcal{X}$ is a finite set of *configurations* that make up the domain of a real-valued *objective function* f. $\mathcal{N}$ is a *neighborhood structure* which is a geometric, topological, or algebraic structure on $\mathcal{X}$ that imposes connectivity on its elements. In evolutionary computation, f is often called the *fitness function*, and the configurations that make up $\mathcal{X}$ are called *genotypes*. In this case, the landscape is often referred to as a *fitness landscape*.

We study landscapes where the configuration set is the set $\mathcal{X} = \{0,1\}^n$ (i.e., all binary strings of length n), the fitness function f is any k-bounded pseudo-Boolean function, and the neighborhood $\mathcal{N}$ is given by the standard Hamming operator.

We will show that every k-bounded pseudo-Boolean function has a sparse representation in an eigenbasis of the Hamming adjacency matrix and use this to derive polynomial time computations of the moments of the fitness function over regions of the landscape. We prove that any k-bounded pseudo-Boolean function f can be written as a linear combination of a bounded number of eigenfunctions of the Hamming neighborhood structure, each of which is polynomially computable. We first show that, for such functions, higher powers of f can also be written as sums over eigenfunctions of the neighborhood structure, each again polynomially computable. This allows us to compute the moments of f over the neighborhood of any point without explicitly examining any of the neighbors.

We then recursively generalize the neighborhood structure and show that eigenfunctions over the Hamming neighborhood are also eigenfunctions over radius-r Hamming spheres, i.e., sets of points that lie at Hamming distance r. We show that the c^{th} moment of any k-bounded pseudo-Boolean function f can be computed in polynomial time over any sphere of arbitrary radius and any Hamming ball (i.e., union of spheres) of arbitrary radius.

One immediate consequence is that central moments of f (such as the mean, variance, skewness, and kurtosis) can be computed in polynomial time over specialized regions of the landscape. This result is significant since the cardinality of such regions can be exponential in the problem size. For example, a radius $n/2$ Hamming sphere contains $\Omega(2^{n/2})$ unique states and any Hamming ball of radius $O(n)$ has $O(2^n)$ unique states. If the epistasis of f is bounded by a constant k, our approach has a time complexity of $O(n^{c^2 k})$ to calculate the exact value for the c^{th} moment of f over any sphere or ball.

The information provided by these calculations allows for a better characterization of the distribution of codomain values of a function over *localized* regions of the landscape. Such information may be useful for the design and analysis of heuristic search algorithms. Currently, the only way to characterize the distribution of a function over localized regions is either by exhaustive enumeration of the region, or estimation via direct sampling.

Finally, we present an application of the results developed in this paper and introduce an efficient algorithm for computing the expected fitness of mutations for a binary genetic algorithm using a k-bounded fitness function and a fixed mutation rate.

1.1 Background

Landscape analysis has been a useful tool to study the characteristics of the state space explored by search algorithms and evolutionary processes [12, 19, 20, 13, 1, 17, 16]. In this paper, we will study the expansion of functions in an eigenbasis of the landscape neighborhood structure. A number of combinatorial optimization domains were first observed by Grover [6] to be related by a difference equation to the neighborhood structure imposed by natural search operators. Further analysis by Stadler [20, 21] has produced a number of important results that capture many characteristics of landscapes. The method used in this paper also relies on representing functions in the orthogonal Walsh basis [23]. Walsh analysis was first introduced to the evolutionary computation community by Holland and Bethke [11, 2] and later developed by Goldberg [5].

This paper generalizes the work of Heckendorn, Rana, and Whitley [8]. Using a Walsh decomposition, they compute summary statistics (e.g., central moments such as the mean, variance, skewness, and kurtosis) over the entire search space for MAX-3-SAT and all k-bounded pseudo-Boolean functions, which they call *embedded landscapes*. In this paper we will give a method for calculating the exact summary statistics for *subsets* of the landscape corresponding to Hamming spheres and volumes of arbitrary radius.

Unless P=NP, NP-complete problems require in the worst case superpolynomial time to solve exactly. Problems such as MAX-k-SAT are therefore often "heuristically solved" in practice using local search methods. Information that is relevant to evolutionary systems and local search algorithms can often be computed in polynomial time for some classic problems that are NP-complete. The moment calculations that are presented in this paper are directly applicable to the topology of the search space that is explored by local search algorithms used to solve MAX-k-SAT problems.

Artificial evolutionary models, and in particular Holland's *genetic algorithm* and Kauffman's *NK-Landscapes*, also assume that the moments of different regions of the search space are important. In both systems, chromosomes are represented as simple binary strings with 0 and 1 alleles. These artificial bit chromosomes are decoded as simple haploid structures. From this perspective, genetic algorithms are often viewed as optimization procedures or search methods acting on pseudo-Boolean functions.

The method presented in this paper relies on direct knowledge of the Walsh coefficients. If the Walsh coefficients are unknown, but f is still epistatically bounded by a constant k, the Walsh coefficients can be efficiently retrieved deterministically in $O(n^k)$ time [14], or stochastically with negligible error in $O(n^2 \log n)$ time [10]. If there are m nonzero Walsh coefficients, Choi et al. [3] present an $O(m \log n)$ adaptive randomized algorithm for finding all of them with high probability.

1.2 Preliminaries

We first give a brief introduction of the notation and concepts used in this paper. A pseudo-Boolean function is a function $f : \{0,1\}^n \to \mathbb{R}$ that maps strings over a binary alphabet into the real numbers. We say a pseudo-Boolean function is *k-bounded* if it can be expressed as the sum of subfunctions that each depend on at most k bits (where k is a constant). Let $x, y \in \{0,1\}^n$ be two points in the domain of f. We define their inner product as

$$\langle x, y \rangle = \sum_{b=1}^{n} x[b]y[b],$$

where $x[b] \in \{0,1\}$ denotes the b^{th} element of x. The Hamming distance $\mathcal{D}(x,y)$ between x and y is the number of positions in which x and y differ. We can write

$$\mathcal{D}(x,y) = \langle x \oplus y, x \oplus y \rangle,$$

where $\oplus$ denotes componentwise exclusive or. The Hamming neighborhood of a point $x \in \{0,1\}^n$ is the set $N(x)$ of all points y such that $\mathcal{D}(x,y) = 1$. The points in $\{0,1\}^n$ along with the Hamming neighborhood form a distance transitive graph on 2^n vertices which is typically referred to as a *hypercube* graph. We can define the *adjacency matrix* as

$$\mathbf{A}_{xy} = \begin{cases} 1 & \text{if } y \in N(x) \\ 0 & \text{otherwise.} \end{cases}$$

Note that $\mathbf{A}$ is a $2^n \times 2^n$ matrix. Consider a pseudo-Boolean function

$$f : \{0,1\}^n \to \mathbb{R}.$$

We will write the c^{th} power of f as f^c where

$$f^c : \{0,1\}^n \to \mathbb{R}; \qquad f^c(x) = (f(x))^c.$$

The c^{th} moment of a discrete random variable Z can be written as

$$\mu_c = \sum_z z^c \Pr\{Z = z\},$$

where $\Pr\{Z = z\}$ is the probability mass function of Z.

Let $X \subseteq \{0,1\}^n$ be a set of points. Moments of f over X are defined as moments of a random variable that assumes the value of f evaluated at a point drawn uniformly at random from X. Since each element of X is drawn with equal probability, the probability mass function is $\frac{1}{|X|}$ and we can define

$$\mu_c(X) = \frac{1}{|X|} \sum_{x \in X} f^c(x) \tag{1}$$

to be the c^{th} moment of f over the set X. For any nonempty set X, it should be clear that $\mu_0(X) = 1$. The first moment, $\mu_1(X)$, is the average value of the function f evaluated over each point in X. The variance of f (the second *central moment*) over the set X can be written as

$$\sigma^2 = \mu_2(X) - \mu_1(X)^2.$$

In general, the c^{th} central moment of f over the subset X can be computed as

$$\sum_{i=0}^{c} \binom{c}{i} (-1)^{c-i} \mu_i(X) \mu_1(X)^{c-i}.$$

Higher central moments correspond to statistical quantities such as *skewness* and *kurtosis* which further characterize the shape of the distribution of the random variable in question.

We identify the elements of $\{0,1\}^n$ with the integers $0, 1, \ldots, 2^n - 1$ in the natural way. Specifically, each bitstring is the n-digit binary representation of a unique integer in the interval $[0, 2^n)$. For example, the integer 0 corresponds to the point $(000\ldots0)$, and the integer 2^a (where $0 \leq a \leq n-1$) corresponds to the point x where

$$x[b] = \begin{cases} 1 & \text{if } a = b \\ 0 & \text{otherwise.} \end{cases}$$

Note that by identifying binary strings with integers in this way, we can characterize a function f as an element of a real vector space of dimension 2^n by associating each state $x \in \{0,1\}^n$ with a standard basis function e_x. Thus $f(x)$ can be written as

$$f(x) = e_x^\top f,$$

This characterization will be useful for studying landscapes algebraically. In particular, $2^n \times 2^n$ matrices can be considered as linear operators on f.

For example the matrix-vector product $\mathbf{A}f$ also lies in $\mathbb{R}^{2^n}$ and can in turn be seen as a discrete function over $\{0,1\}^n$. In particular, the x^{th} element of this vector is the inner product of f with the row of $\mathbf{A}$ corresponding to x.

$$\begin{aligned} \mathbf{A}f(x) &= e_x^\top \mathbf{A} f \\ &= \sum_{y \in \{0,1\}^n} \mathbf{A}_{xy} f(y) \\ &= \sum_{y \in N(x)} f(y). \end{aligned} \tag{2}$$

So we can view

$$\mathbf{A} : \mathbb{R}^{2^n} \to \mathbb{R}^{2^n}$$

as a linear map acting on f. The image of f under $\mathbf{A}$,

$$\mathbf{A}f : \{0,1\}^n \to \mathbb{R},$$

can be treated as a function that gives the sum of f evaluated over the neighborhood of each state.

The Hamming sphere of radius r around a point x is defined as the set

$$S^{(r)}(x) = \{y \in \{0,1\}^n : \mathcal{D}(x,y) = r\}.$$

Of course the Hamming sphere is a straightforward generalization of the Hamming neighborhood since

$$N(x) = S^{(1)}(x).$$

The Hamming ball of radius r around a point x consists of the union of all Hamming spheres of radius less than or equal to r around x:

$$B^{(r)}(x) = \{y \in \{0,1\}^n : \mathcal{D}(x,y) \leq r\}.$$

2 Eigenfunctions of the neighborhood structure

In order to study the relationship of a k-bounded pseudo-Boolean function to the Hamming neighborhood structure, we will utilize a decomposition of the function into eigenfunctions of the neighborhood structure. We say a pseudo-Boolean function g is an *eigenfunction* of the Hamming neighborhood structure if and only if

$$\mathbf{A}g = \lambda g$$

for a scalar λ. That is, we have $\forall x \in \{0,1\}^n$,

$$\begin{aligned}\mathbf{A}g(x) &= e_x^\top \mathbf{A}g \\ &= e_x^\top \lambda g \\ &= \lambda g(x).\end{aligned}$$

In other words, the image of x under the function $\mathbf{A}g$ can be compactly represented by the image of x under g multiplied by a constant λ. Functions with this property (up to an additive constant) together with the neighborhood structure are exactly the so-called *elementary landscapes* of Stadler et al. [20].

For $0 \leq i \leq 2^n - 1$, the i^{th} Walsh function is defined as

$$\psi_i(x) = (-1)^{\langle i,x\rangle},$$

where the inner product is taken over x and the length-n binary string representation of i. The *order* of the i^{th} Walsh function is $\langle i,i\rangle$, that is, the number of ones in the length-n binary string representation of i. The following simple identity will become useful.

$$\begin{aligned}\psi_i(x)\psi_j(x) &= (-1)^{\langle i,x\rangle + \langle j,x\rangle} \\ &= (-1)^{\langle i\oplus j,x\rangle} = \psi_{i\oplus j}(x). \qquad (3)\end{aligned}$$

In this paper, we will rely on the fact that the Walsh functions form an orthogonal basis of eigenfunctions of the Hamming neighborhood matrix $\mathbf{A}$. This is partially captured by the following Lemma.

Lemma 1 (Walsh Eigenfunction Lemma)**.** *The i^{th} Walsh function is an eigenfunction of $\mathbf{A}$:*

$$\mathbf{A}\psi_i = (n - 2\langle i,i\rangle)\psi_i.$$

Proof. Let $x \in \{0,1\}^n$ be arbitrary. We have

$$\begin{aligned}\mathbf{A}\psi_i(x) &= \sum_{y\in N(x)} \psi_i(y) && \text{by (2);} \\ &= \sum_{y\in N(x)} (-1)^{\langle i,y\rangle}.\end{aligned}$$

For each $y \in N(x)$, because x and y differ by a single bit, there exists a unique $0 \leq a \leq 2^n - 1$ for which $x \oplus y = 2^a$. Thus we can make the following case distinction. Let $\wedge$ denote componentwise conjunction in the binary representation. If $i \wedge 2^a = 0$ (i.e., $\langle i, (x \oplus y)\rangle = 0$) then $\langle i, y\rangle = \langle i, x\rangle$ and $\psi_i(y) = \psi_i(x)$. On the other hand, if $i \wedge 2^a = 2^a$ then $|\langle i, y\rangle - \langle i, x\rangle| = 1$ and $(-1)^{\langle i,y\rangle} = -(-1)^{\langle i,x\rangle}$, or equivalently, $\psi_i(y) = -\psi_i(x)$.

Since each Hamming neighbor differs from x in each of the n possible bit positions, there are $n - \langle i, i\rangle$ elements y of $N(x)$ that satisfy the first condition and $\langle i, i\rangle$ that satisfy the second. Hence

$$\begin{aligned}\sum_{y \in N(x)} \psi_i(y) &= \left(\left(n - \langle i, i\rangle\right) \psi_i(x) - \langle i, i\rangle \psi_i(x)\right) \\ &= \left(n - 2\langle i, i\rangle\right) \psi_i(x).\end{aligned}$$

Since we chose x arbitrarily, the property holds for any basis function e_x and we have the general equation

$$\mathbf{A}\psi_i = \left(n - 2\langle i, i\rangle\right) \psi_i,$$

and ψ_i is an eigenfunction of $\mathbf{A}$. □

From Lemma 1 it is easy to show the following is true. Summing over all i with $\langle i, i\rangle = p$, we have

$$\mathbf{A}\left(\sum_{i:\langle i,i\rangle = p} a_i \psi_i\right) = (n - 2p) \sum_{i:\langle i,i\rangle = p} a_i \psi_i. \tag{4}$$

where a_i is an arbitrary coefficient. Therefore the Walsh functions of a particular order p form the basis of an eigenspace corresponding to eigenvalue $n - 2p$. Thus any function that can be expressed as a linear combination of Walsh functions of a given order p is also an eigenfunction of $\mathbf{A}$. This is critical because it supports analysis by decomposition.

Proposition 1. *If g is an eigenfunction of $\mathbf{A}$ that lies in the eigenspace spanned by a set of Walsh functions of the same order p then the first moment of g over the neighborhood of an arbitrary point x is*

$$\mu_1(N(x)) = \left(1 - \frac{2p}{n}\right) g(x).$$

Proof.

$$\begin{aligned}\mu_1(N(x)) &= \frac{1}{n} \sum_{y \in N(x)} g(y) \\ &= \frac{1}{n} \mathbf{A} g(x) && \text{by (2);} \\ &= \left(1 - \frac{2p}{n}\right) g(x) && \text{by (4).}\end{aligned}$$

□

This means that for such an eigenfunction, the first moment of g over the entire Hamming neighborhood of x is directly proportional to the image of x under g. So if $N(x) = \{y_1, y_2, \ldots\}$ is the set of points that compose the Hamming neighborhood of an arbitrarily selected point x, we can immediately compute the mean of their images under g, i.e., the mean of $\{g(y_1), g(y_2), \ldots\}$, without enumerating any of the elements of $N(x)$.

2.1 Decompositions of k-bounded pseudo-Boolean functions

Eigenfunctions of the Hamming neighborhood structure are a very restricted class of functions and thus their properties may not seem immediately clear. The power of these functions however comes from the fact that we can represent *arbitrary* pseudo-Boolean functions as linear combinations of component eigenfunctions of the neighborhood structure. Any linear map applied to a pseudo-Boolean function can be represented as a sum of the images of its component eigenfunctions under that map.

The Walsh basis is *functionally complete* over $\{0,1\}^n$ [23], that is, any arbitrary pseudo-Boolean function $f : \{0,1\}^n \to \mathbb{R}$ can be written as a linear combination of at most 2^n orthogonal Walsh functions

$$f(x) = \sum_{i=0}^{2^n-1} w_i \psi_i(x),$$

where w_i is a scalar called the i^{th} *Walsh coefficient.* We can group each term by its order

$$f(x) = \sum_{p=0}^{n} \varphi_{[p]}(x), \tag{5}$$

where $\varphi_{[p]}$ is an eigenfunction of order p defined as

$$\varphi_{[p]}(x) = \sum_{i:\langle i,i\rangle = p} w_i \psi_i(x). \tag{6}$$

Hence $\varphi_{[p]}$ is a linear combination of Walsh functions of order p. In other words, $\varphi_{[p]}$ is a component eigenfunction of f that lies in the eigenspace of $\mathbf{A}$ corresponding to eigenvalue $n - 2p$. Since there are $\binom{n}{p}$ orthogonal Walsh functions of a given order p, $\varphi_{[p]}$ contains at most $\binom{n}{p}$ terms.

We now prove some simple bounds on the order of non-zero Walsh coefficients which will later be used in our main theorems. This is critical to demonstrating the tractability of these computations.

Lemma 2 (Heckendorn, Rana, and Whitley [9]). *Let f be a k-bounded pseudo-Boolean function on $\{0,1\}^n$. For any length-n binary string i,*

$$w_i \neq 0 \implies \langle i,i\rangle \leq k,$$

where w_i is the i^{th} coefficient in the decomposition of f.

Proof. Since f is k-bounded it can be expressed as a sum of subfunctions f_j that each depend on at most k bits. Denote as $w_i^{(f_j)}$ the i^{th} Walsh coefficient on the j^{th} subfunction. Since the Walsh transform is linear, the i^{th} Walsh coefficient of f is the sum of the i^{th} Walsh coefficients of the subfunctions, i.e.,

$$w_i = \sum_j w_i^{(f_j)}.$$

Since any f_j depends on at most k bits, if $\langle i, i\rangle > k$ then $\forall j, w_i^{(f_j)} = 0$. Thus $\langle i, i\rangle > k \implies w_i = 0$ which gives the contrapositive. □

Lemma 2 generalizes easily to collections of binary strings and the corresponding coefficients.

Lemma 3. *Let f be a k-bounded pseudo-Boolean function on $\{0,1\}^n$. Let $\mathcal{I}$ be a set of length-n binary strings. Consider the string*

$$\oplus(\mathcal{I}) = \bigoplus_{i\in\mathcal{I}} i.$$

We have

$$\forall i \in \mathcal{I}, w_i \neq 0 \implies \langle \oplus(\mathcal{I}), \oplus(\mathcal{I})\rangle \leq |\mathcal{I}|k,$$

where w_i is the i^{th} coefficient in the decomposition of f.

Proof. By induction on $|\mathcal{I}|$. In the base case we have $|\mathcal{I}| = 1$ which is proved by Lemma 2.

Let $\mathcal{J}$ be a set of length-n binary strings with $|\mathcal{J}| \geq 1$. Consider the string

$$\oplus(\mathcal{J}) = \bigoplus_{j\in\mathcal{J}} j.$$

By the inductive hypothesis assume for strings of length n

$$w_j \neq 0 \quad \forall j \in \mathcal{J} \implies \langle \oplus(\mathcal{J}), \oplus(\mathcal{J})\rangle \leq |\mathcal{J}|k.$$

Now consider a string h such that $w_h \neq 0$ in the Walsh decomposition of f. By Lemma 2 we know $\langle h, h\rangle \leq k$. Let

$$\mathcal{I} = \mathcal{J} \cup h.$$

We are interested in the string

$$\begin{aligned}\oplus(\mathcal{I}) &= \bigoplus_{i\in\mathcal{I}} i \\ &= \bigoplus_{i\in\mathcal{J}} i \oplus h \\ &= \oplus(\mathcal{J}) \oplus h.\end{aligned}$$

But the order of h is bounded by k and the order of $\bigoplus(\mathcal{J})$ is bounded by $|\mathcal{J}|k$ so we have

$$\begin{aligned}\langle \bigoplus(\mathcal{I}), \bigoplus(\mathcal{I})\rangle &= \langle \bigoplus(\mathcal{J}) \oplus h, \bigoplus(\mathcal{J}) \oplus h\rangle \\ &\leq |\mathcal{J}|k + k \\ &= (|\mathcal{J}| + 1)k \\ &= |\mathcal{I}|k.\end{aligned}$$

□

We are now ready to prove theorems about the decomposition of k-bounded pseudo-Boolean functions (and their powers) into eigenfunctions. Lemma 2 constrains the order of nonzero coefficients in the Walsh representation of a k-bounded pseudo-Boolean function. This means we can write any such function as a linear combination of exactly those Walsh functions with nonzero coefficients. This is captured by the following theorem.

Theorem 1 (Decomposition Theorem)**.** *Every k-bounded pseudo-Boolean function f can be written as a linear combination of $k + 1$ eigenfunctions of $\mathbf{A}$.*

Proof. We can write f in the Walsh representation

$$f(x) = \sum_i w_i \psi_i(x).$$

By the contraposition of Lemma 2, w_i is zero for all $\langle i, i\rangle > k$ so we may write

$$\begin{aligned}f(x) &= \sum_{i:\langle i,i\rangle \leq k} w_i \psi_i(x) \\ &= \sum_{p=0}^{k} \varphi_{[p]}(x),\end{aligned}$$

where $\varphi_{[p]}$ is defined as in (6). Each $\varphi_{[p]}$ is a linear combination of at most $\binom{n}{p}$ Walsh functions of order p and is thus an eigenfunction of $\mathbf{A}$ corresponding to eigenvalue $n - 2p$. □

We are thus taking advantage of the fact that f has a sparse representation in the Walsh basis. It is important to see that, for any f bounded epistatically by k, there are at most $k+1$ eigenfunctions $\varphi_{[p]}$, each of which consists of a linear combination of at most $\binom{n}{p}$ terms. Since $p \leq k$, the number of terms in the linear combination is bounded by a polynomial in n of degree at most k.

In order to compute higher moments of f, it will be necessary to work with higher powers of f. If f can be written as a linear combination of Walsh functions, then clearly f^c can be written as a degree c polynomial in the Walsh functions.

To understand this more clearly, we will first present the case for $c = 2$. In other words, we will show that if f is a k-bounded pseudo-Boolean function, f^2 can be written as a second degree polynomial in the Walsh functions. Using the identity in (3) and Lemma 3, we can carry the bounds on the order of nonzero Walsh coefficients over to this case. This leads to the following theorem.

Theorem 2 (Square Decomposition Theorem)**.** *Every k-bounded pseudo-Boolean taken to the second power can be written as a sum of $2k+1$ eigenfunctions of $\mathbf{A}$.*

Proof. Again, writing f^2 in the Walsh representation we have

$$\begin{aligned} f^2(x) &= \left(\sum_i w_i \psi_i(x)\right)^2 \\ &= \sum_{i,j} w_i w_j \psi_i(x) \psi_j(x) \\ &= \sum_{i,j} w_i w_j \psi_{i \oplus j}(x) && \text{by (3).} \end{aligned}$$

By Lemma 2, the order of each i and j are bounded by k so we can first define the following set of strings,

$$\mathcal{Q} = \{i \in \{0,1\}^n : \langle i, i \rangle \leq k\}, \tag{7}$$

then take the sum over the Cartesian square of $\mathcal{Q}$.

$$f^2(x) = \sum_{(i,j) \in \mathcal{Q} \times \mathcal{Q}} w_i w_j \psi_{i \oplus j}(x).$$

We can again group together all the terms by Walsh function order,

$$\varphi_{[p]}(x) = \sum_{(i,j) \in \mathcal{Q} \times \mathcal{Q} : \langle i \oplus j, i \oplus j \rangle = p} w_i w_j \psi_{i \oplus j}(x). \tag{8}$$

Obviously, the term corresponding to pair (i,j) is only non-zero if both w_i and w_j are non-zero. Hence, by Lemma 3, the order of $i \oplus j$ is at most $2k$ so we can write f^2 as a linear combination of each $\varphi_{[p]}$ as defined in Equation (8):

$$f^2(x) = \sum_{p=0}^{2k} \varphi_{[p]}(x),$$

and each $\varphi_{[p]}$ is an eigenfunction of $\mathbf{A}$ corresponding to eigenvalue $n - 2p$. □

Generalizing this decomposition to higher powers of f is now simply an exercise of writing f^c as a degree c polynomial in the component eigenfunctions and carrying the order bounds into this case in a similar manner to the above theorem.

Theorem 3 (General Power Decomposition Theorem)**.** *Every k-bounded pseudo-Boolean function taken to the c^{th} power can be written as a sum of $ck+1$ eigenfunctions of $\mathbf{A}$.*

Proof. To simplify notation, we define the set

$$\mathcal{Q}^c = \underbrace{\mathcal{Q} \times \mathcal{Q} \times \cdots \times \mathcal{Q}}_{c},$$

where $\mathcal{Q}$ is as defined in (7). An element of $\mathcal{Q}^c$ is hence a c-tuple of bitstrings q_i:

$$q = (q_1, q_2, \dots, q_c).$$

We write f^c in the Walsh representation.

$$\begin{aligned} f^c(x) &= \left(\sum_i w_i \psi_i(x) \right)^c \\ &= \sum_{q \in \mathcal{Q}^c} \left(\prod_{i=1}^{c} w_{q_i} \right) \left(\prod_{i=1}^{c} \psi_{q_i}(x) \right). \end{aligned}$$

Letting $\oplus(q) = \bigoplus_{i=1}^{c} q_i$,

$$= \sum_{q \in \mathcal{Q}^c} \left(\prod_{i=1}^{c} w_{q_i} \right) \psi_{\oplus(q)}(x) \qquad \text{by (3).}$$

Finally, grouping the terms by Walsh function order,

$$\varphi_{[p]}(x) = \sum_{q \in \mathcal{Q}^c : \langle \oplus(q), \oplus(q) \rangle = p} \left(\prod_{i=1}^{c} w_{q_i} \right) \psi_{\oplus(q)}(x). \tag{9}$$

The term corresponding to the c-tuple q is only non-zero if all the w_{q_i} are nonzero. Hence, by Lemma 3, the order of $\oplus(q)$ is at most ck so we can write f^c as a linear combination of each $\varphi_{[p]}$ as defined in Equation (9):

$$f^c(x) = \sum_{p=0}^{ck} \varphi_{[p]}(x), \tag{10}$$

and each $\varphi_{[p]}$ is an eigenfunction of $\mathbf{A}$ corresponding to eigenvalue $n - 2p$. □

Finally, we would like to bound the number of individual terms involved in the linear combination $\varphi_{[p]}$. This bound will become useful later when we analyze the complexity of this approach.

Lemma 4. *Each $\varphi_{[p]}$ defined in Equation (9) above is a linear combination of at most $O(n^{pc})$ individual terms.*

Proof. Let $\mathcal{Q} = \{i : \langle i, i \rangle \leq k\}$. Clearly $|\mathcal{Q}| = O\left(\binom{n}{p}\right)$ is a polynomial of degree p in n. The number of terms in each $\varphi_{[p]}$ is $|\mathcal{Q}^c| = |\mathcal{Q}|^c$ which is a polynomial of degree pc in n. Hence the number of Walsh coefficients involved in the linear combination defined by $\varphi_{[p]}$ is $O(n^{pc})$. □

3 Computing the moments in polynomial time

In this section we use the decomposition of an arbitrary k-bounded pseudo-Boolean function into eigenfunctions of the Hamming neighborhood structure to compute the c^{th} moment of f over regions in polynomial time. Perhaps it is most illustrative to begin by giving a general formula for computing moments over the immediate Hamming neighborhood.

3.1 Hamming neighborhoods

We have already seen that the first moment of f over the Hamming neighborhood can be calculated using the following expression.

$$\frac{1}{n}\mathbf{A}f.$$

Specifically, $\frac{1}{n}\mathbf{A}f(x)$ gives the first moment of f over the Hamming neighbors of x. Since f can be written as a linear combination of a bounded number of component functions, the first moment of f can be characterized as the sum of images of these components under the linear map. Furthermore, since each of these components are eigenfunctions of that map, the calculation of their images under the map reduces to multiplication by a scalar.

It is useful to note that the c^{th} moment of f over the neighborhood is equal to the first moment of f^c over the neighborhood. Since we have seen that f^c is representable by a bounded number of component functions, we can extend the above reasoning to f^c.

To compute the c^{th} moment of f over the neighborhood $N(x)$ of an arbitrary point x, we simply use the fact that f^c can be decomposed into $ck+1$ eigenfunctions of $\mathbf{A}$.

$$\begin{aligned}
\mu_c(N(x)) &= \frac{1}{n}\sum_{y\in N(x)} f(y)^c \\
&= \frac{1}{n}\mathbf{A}f^c(x) && \text{by (2);} \\
&= \frac{1}{n}\mathbf{A}\sum_{p=0}^{ck}\varphi_{[p]}(x) && \text{by (10);} \\
&= \sum_{p=0}^{ck}\left(1-\frac{2p}{n}\right)\varphi_{[p]}(x). && (11)
\end{aligned}$$

Assuming c and k are constants, the calculation of $\mu_c(N(x))$ is bounded by the number of individual terms in the series in (11). There are $(ck+1)$ component eigenfunctions $\varphi_{[p]}$ in the series. By Lemma 4, each of these eigenfunctions is a linear combination of no more than $O(n^{pc})$ individual terms. Since p is bounded by ck and there is only one order zero term, there are at most $ck \times \mathit{poly}(n)+1$ individual terms where $\mathit{poly}(n)$ is $O(n^{c^2k})$. We point out that this bound is relatively non-tight.

3.2 Hamming balls of arbitrary radius

We have characterized the c^{th} moment of f over the neighborhood as a function given by the image of f^c under the linear operator $\mathbf{A}$. We have also taken into account the fact that f^c is decomposable into a bounded number of eigenfunctions of $\mathbf{A}$ to show that the moments over the neighborhood of any given point can be computed in polynomial time. Of course, the members of the Hamming neighborhood itself can be enumerated in linear time so this approach offers no computational advantage in this case (though it may prove useful when analyzing the expected value of the Walsh coefficients over a problem distribution as in [22]).

However, we note that if f has a sparse representation in the eigenbasis of *any* linear map, the above analysis holds. In this section we will see that the adjacency structure for any general radius-r Hamming sphere provides such a map.

We begin by characterizing the adjacency structure for radius-r Hamming spheres. Let x be an arbitrary but fixed point in $\{0,1\}^n$. Consider a vertex y at some distance $\mathcal{D}(x,y)$. All Hamming neighbors of y are either one vertex closer to x or one vertex further away. Define the *approaching set*

$$\alpha(x,y) = \{z \in N(y) : \mathcal{D}(x,z) = \mathcal{D}(x,y) - 1\}$$

and the *retreating set*

$$\beta(x,y) = \{z \in N(y) : \mathcal{D}(x,z) = \mathcal{D}(x,y) + 1\}.$$

Thus the approaching and retreating sets partition the neighborhood set of y and

$$\alpha(x,y) \cup \beta(x,y) = N(y). \tag{12}$$

See Figure 1 for an illustration.

The set $S^{(r)}(x)$ consists of all strings at Hamming distance r from x: those strings that differ from x in exactly r positions. Hence $|S^{(r)}(x)| = \binom{n}{r}$. Consider a state y on this sphere, that is, $\mathcal{D}(x,y) = r$. Since y differs from x in exactly r positions, there are r Hamming moves that result in some state z_1 with $\mathcal{D}(x,z_1) = r-1$. Thus we have $|\alpha(x,y)| = r$. Furthermore, there are $n-r$ Hamming moves from y that result in a state z_2 with $\mathcal{D}(x,z_2) = r+1$. Hence, $|\beta(x,y)| = n-r$.

A generalization of the adjacency matrix $\mathbf{A}$ which we will call the *sphere matrix* of radius r we define as

$$\mathbf{S}^{(r)}_{xy} = \begin{cases} 1 & \text{if } y \in S^{(r)}(x) \text{, that is, } \mathcal{D}(x,y) = r \\ 0 & \text{otherwise.} \end{cases}$$

This matrix identifies all vertex pairs in which one is contained in the radius-r sphere of the other. We construct the sphere matrix $\mathbf{S}^{(r)}$ of radius r recursively in terms of $\mathbf{A}$. In order to do so, we will first prove some useful properties about sphere matrices.

The set $\{0,1\}^n$ together with the Hamming distance function form a metric space so we have for all $x, y \in \{0,1\}^n$, $\mathcal{D}(x,y) = \mathcal{D}(y,x)$ and sphere matrices of any radius are symmetric:

$$\mathbf{S}^{(r)}_{xy} = \mathbf{S}^{(r)}_{yx}. \tag{13}$$

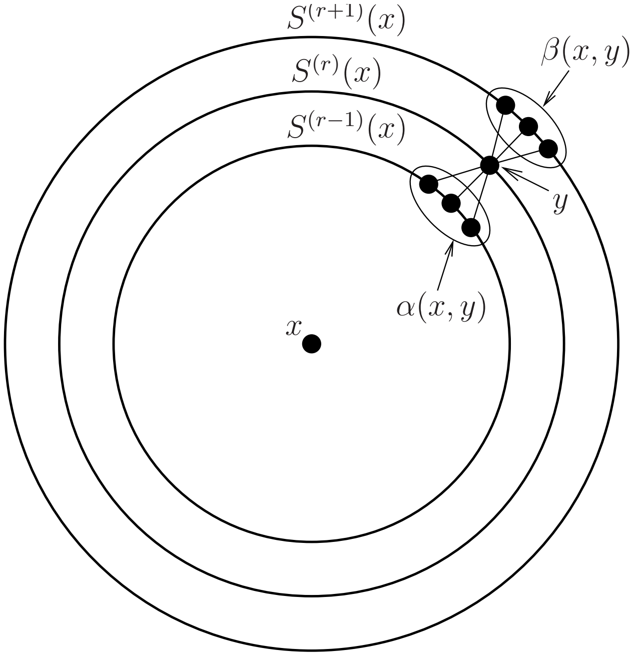

Figure 1: Illustration of approaching set $\alpha(x, y)$ and retreating set $\beta(x, y)$. For some y with $\mathcal{D}(x, y) = r$.

Given any two sphere matrices, their product is a matrix that gives the number of elements in the intersection of the spheres they represent. Formally, let $\mathbf{S}^{(r)}$ and $\mathbf{S}^{(s)}$ be sphere matrices of radius r and s respectively. The product is the matrix

$$\begin{aligned}
\left(\mathbf{S}^{(r)}\mathbf{S}^{(s)}\right)_{xy} &= \sum_z \mathbf{S}^{(r)}_{xz}\mathbf{S}^{(s)}_{zy} \\
&= \sum_z \mathbf{S}^{(r)}_{xz}\mathbf{S}^{(s)}_{yz} && \text{by (13);} \\
&= |S^{(r)}(x) \cap S^{(s)}(y)|. && (14)
\end{aligned}$$

We now characterize the particular matrix product $(\mathbf{S}^{(r-1)}\mathbf{A})$ which will be used in our recursive expression for $\mathbf{S}^{(r)}$.

Lemma 5. *Over $\{0, 1\}^n$ we have,*

$$(\mathbf{S}^{(r-1)}\mathbf{A})_{xy} = \begin{cases} r & \textit{if } y \in S^{(r)}(x) \\ n - r + 2 & \textit{if } y \in S^{(r-2)}(x) \\ 0 & \textit{otherwise.} \end{cases}$$

Proof. By (14) we have

$$(\mathbf{S}^{(r-1)}\mathbf{A})_{xy} = |S^{(r-1)}(x) \cap N(y)|,$$

since $\mathbf{A} = \mathbf{S}^{(1)}$ and $N(x) = S^{(1)}(x)$. Consider the neighbor set $N(y)$ of y. Recall from Equation (12) that the approaching and retreating sets $\alpha(x,y)$, $\beta(x,y)$ partition $N(y)$.

Suppose $y \in S^{(r)}(x)$. For all $z \in \alpha(x,y)$, $\mathcal{D}(x,z) = \mathcal{D}(x,y) - 1 = r - 1$. The neighbors of y that are in $S^{(r-1)}(x)$ are exactly the approaching set $\alpha(x,y)$. Thus we have

$$\begin{aligned}(\mathbf{S}^{(r-1)}\mathbf{A})_{xy} &= |S^{(r-1)}(x) \cap N(y)| \\ &= |\alpha(x,y)| \\ &= r.\end{aligned}$$

Now suppose $y \in S^{(r-2)}(x)$. For all $z \in \beta(x,y)$, $\mathcal{D}(x,z) = \mathcal{D}(x,y) + 1 = r - 1$. Thus the neighbors of y that are in $S^{(r-1)}(x)$ are exactly the retreating set $\beta(x,y)$. Thus we have

$$\begin{aligned}(\mathbf{S}^{(r-1)}\mathbf{A})_{xy} &= |S^{(r-1)}(x) \cap N(y)| \\ &= |\beta(x,y)| \\ &= n - r + 2.\end{aligned}$$

Finally suppose y is in neither sphere $S^{(r)}(x)$ nor $S^{(r-2)}(x)$. Then $\mathcal{D}(x,y) \neq r$ and $\mathcal{D}(x,y) \neq r - 2$. So

$$\begin{aligned}(\mathbf{S}^{(r-1)}\mathbf{A})_{xy} &= |S^{(r-1)}(x) \cap N(y)| \\ &= |(\alpha(x,y) \cup \beta(x,y)) \cap S^{(r-1)}(x)| \\ &= |\emptyset| \\ &= 0,\end{aligned}$$

since $\mathcal{D}(x,y) - 1 \neq r - 1$ and $\mathcal{D}(x,y) + 1 \neq r - 1$. □

The following lemma uses the above result to provide a matrix expression for the characteristic function of $y \in S^{(r)}(x)$. The expression involves the sphere matrices of radius $r-1$ and $r-2$. This will allow us to define $\mathbf{S}^{(r)}$ recursively in terms of lower radius sphere matrices.

Lemma 6. *Let x and y be arbitrary points in $\{0,1\}^n$. Given sphere matrices $\mathbf{S}^{(r-1)}$ and $\mathbf{S}^{(r-2)}$ we have the following identity.*

$$\frac{1}{r}\left((\mathbf{S}^{(r-1)}\mathbf{A})_{xy} - (n-r+2)\mathbf{S}^{(r-2)}_{xy}\right) = \begin{cases} 1 & \text{if } y \in S^{(r)}(x) \\ 0 & \text{otherwise.} \end{cases} \tag{15}$$

Proof. We prove this result by cases.

Case 1: $y \in S^{(r)}(x)$ By Lemma 5 we have $\mathbf{S}^{(r-1)}\mathbf{A}_{xy} = r$. Furthermore, since $y \notin S^{(r-2)}(x)$ we have $\mathbf{S}^{(r-2)}_{xy} = 0$. Thus Equation (15) evaluates to

$$\begin{aligned}\frac{1}{r}\left((\mathbf{S}^{(r-1)}\mathbf{A})_{xy} - (n-r+2)\mathbf{S}^{(r-2)}_{xy}\right) &= \frac{1}{r}(r-0) \\ &= 1.\end{aligned}$$

Case 2: $y \in S^{(r-2)}(x)$ By Lemma 5 we have $\mathbf{S}^{(r-1)}\mathbf{A}_{xy} = (n-r+2)$. Since $y \in S^{(r-2)}(x)$, $\mathbf{S}^{(r-2)}_{xy} = 1$ and Equation (15) evaluates to

$$\begin{aligned} \frac{1}{r}\left((\mathbf{S}^{(r-1)}\mathbf{A})_{xy} - (n-r+2)\mathbf{S}^{(r-2)}_{xy}\right) &= \frac{1}{r}((n-r+2)-(n-r+2)) \\ &= 0. \end{aligned}$$

Case 3: $y \notin S^{(r)}(x)$ **and** $y \notin S^{(r-2)}(x)$ By Lemma 5 we have $\mathbf{S}^{(r-1)}\mathbf{A}_{xy} = 0$. Furthermore, since $y \notin S^{(r-2)}(x)$ we have $\mathbf{S}^{(r-2)}_{xy} = 0$. Thus Equation (15) evaluates to

$$\begin{aligned} \frac{1}{r}\left((\mathbf{S}^{(r-1)}\mathbf{A})_{xy} - (n-r+2)\mathbf{S}^{(r-2)}_{xy}\right) &= \frac{1}{r}(0-0) \\ &= 0. \end{aligned}$$

□

Hence, by Lemma 6 we can now define the sphere matrix recursively.

$$\mathbf{S}^{(r)} = \frac{1}{r}\left(\mathbf{S}^{(r-1)}\mathbf{A} - (n-r+2)\mathbf{S}^{(r-2)}\right). \tag{16}$$

We have the two base cases $\mathbf{S}^{(1)} = \mathbf{A}$ and $\mathbf{S}^{(0)} = \mathbf{I}$, where $\mathbf{I}$ is the $2^n \times 2^n$ identity matrix (this corresponds to the degenerate sphere $S^{(0)}(x) = \{x\}$). We now show that if f is an eigenfunction of $\mathbf{A}$ with eigenvalue λ, it is also an eigenfunction of the sphere matrix $\mathbf{S}^{(r)}$ with an eigenvalue that is a degree-r polynomial in λ.

Let f be an eigenfunction of $\mathbf{A}$. Consider the matrix-vector product $\mathbf{S}^{(r)}f$ evaluated at state x.

$$\begin{aligned} \mathbf{S}^{(r)}f(x) &= e_x^\top \mathbf{S}^{(r)} f \\ &= \sum_{y \in \{0,1\}^n} \mathbf{S}^{(r)}_{xy} f(y) \\ &= \sum_{y \in S^{(r)}(x)} f(y) \qquad (17) \end{aligned}$$

since $\mathbf{S}^{(r)}_{xy} = 1 \iff y \in S^{(r)}(x)$, otherwise it is equal to zero. Clearly, Equation (2) is the special case when $r = 1$.

It is now straightforward to show that eigenfunctions of the immediate Hamming neighborhood structure are also eigenfunctions of the radius r Hamming sphere. In particular, if f is an eigenfunction of $\mathbf{A}$ with eigenvalue λ_p, it must also be an eigenfunction of $\mathbf{S}^{(r)}$ with eigenvalue $\gamma_p^{(r)}$: a scalar that can be defined recursively using $\gamma_p^{(1)} = \lambda_p$ and $\gamma_p^{(0)} = 1$ as base cases. We capture this in the following theorem.

Theorem 4. *If f is an eigenfunction of $\mathbf{A}$ with eigenvalue λ_p, then f is an eigenfunction of $\mathbf{S}^{(r)}$ with eigenvalue $\gamma_p^{(r)}$ given by the recurrence $\gamma_p^{(r)} = \frac{1}{r}\left(\lambda_p \gamma_p^{(r-1)} - (n-r+2)\gamma_p^{(r-2)}\right)$ with $\gamma_p^{(1)} = \lambda_p$ and $\gamma_p^{(0)} = 1$.*

Proof. We proceed by induction on r. We have two base cases,

$$\begin{aligned}\mathbf{S}^{(0)}f &= \mathbf{I}f = f\\ \mathbf{S}^{(1)}f &= \mathbf{A}f = \lambda_p f.\end{aligned}$$

Thus $\gamma_p^{(0)} = 1$ and $\gamma_p^{(1)} = \lambda_p$. Suppose for induction that

$$\mathbf{S}^{(r-1)}f = \gamma_p^{(r-1)}f$$

and

$$\mathbf{S}^{(r-2)}f = \gamma_p^{(r-2)}f$$

for scalars $\gamma_p^{(r-1)}$ and $\gamma_p^{(r-2)}$. Thus,

$$\begin{aligned}\mathbf{S}^{(r)}f &= \frac{1}{r}\left(\mathbf{S}^{(r-1)}\mathbf{A} - (n-r+2)\mathbf{S}^{(r-2)}\right)f && \text{by (16);}\\ &= \frac{1}{r}\left(\lambda_p\mathbf{S}^{(r-1)}f - (n-r+2)\mathbf{S}^{(r-2)}f\right)\\ &= \frac{1}{r}\left(\lambda_p\gamma_p^{(r-1)} - (n-r+2)\gamma_p^{(r-2)}\right)f && \text{by induction,}\end{aligned}$$

so we have the recurrence

$$\gamma_p^{(r)} = \frac{1}{r}\left(\lambda_p\gamma_p^{(r-1)} - (n-r+2)\gamma_p^{(r-2)}\right). \tag{18}$$

□

Corollary 1. *The quantity $\gamma_p^{(r)}$ is a degree-r polynomial in λ_p and is computable in time linear in r.*

Proof. Clearly $\gamma_p^{(0)} = 1$ and $\gamma_p^{(1)} = \lambda_p$ are degree zero and one polynomials in λ_p respectively. Again, using induction on r it is immediately clear that the recurrence in Equation (18) describes a degree-r polynomial in λ_p. Furthermore, $\gamma_p^{(r)}$ is computable in linear time using dynamic programming starting first with $\gamma_p^{(1)} = \lambda_p$ and using the recurrence in Equation (18) to compute $\gamma_p^{(i)}$ for $i = 2, \ldots, r$. □

By Equation (18), for the eigenfunction $\varphi_{[p]}$ we have

$$\mathbf{S}^{(r)}\varphi_{[p]}(x) = \gamma_p^{(r)}\varphi_{[p]}(x),$$

where

$$\begin{aligned}\gamma_p^{(r)} &= \left(\frac{n-2p}{r}\right)\gamma_p^{(r-1)} - \left(\frac{n-r+2}{r}\right)\gamma_p^{(r-2)}\\ &\vdots\\ \gamma_p^{(1)} &= (n-2p)\\ \gamma_p^{(0)} &= 1.\end{aligned}$$

We are now ready to prove the main results of the paper. The first moment of f^c (i.e., the c^{th} moment of f) over the sphere of radius r around an arbitrary point x can be calculated using a function corresponding to the image of f^c under the linear map $\mathbf{S}^{(r)}$. Using the fact that f^c can be decomposed into a constant number of eigenfunctions of $\mathbf{S}^{(r)}$ we have the following.

Theorem 5. *Fix c and k. Let f be any k-bounded pseudo-Boolean function. Let $S^{(r)}(x)$ be a sphere of radius r around an arbitrary state x. The quantity $\mu_c(S^{(r)}(x))$ (the c^{th} moment of f over the sphere) can be computed in time polynomial in n.*

Proof.

$$\begin{aligned}
\mu_c(S^{(r)}(x)) &= \frac{1}{|S^{(r)}(x)|} \sum_{y \in S^{(r)}(x)} f(y)^c \\
&= \frac{1}{|S^{(r)}(x)|} \mathbf{S}^{(r)} f^c(x) && \text{by (17);} \\
&= \frac{1}{|S^{(r)}(x)|} \mathbf{S}^{(r)} \sum_{p=0}^{ck} \varphi_{[p]}(x) && \text{by (10);} \\
&= \frac{1}{|S^{(r)}(x)|} \sum_{p=0}^{ck} \gamma_p^{(r)} \varphi_{[p]}(x) && \text{by Theorem 4,}
\end{aligned}$$

and since $\forall x \in \{0,1\}^n, |S^{(r)}(x)| = \binom{n}{r}$,

$$= \binom{n}{r}^{-1} \sum_{p=0}^{ck} \gamma_p^{(r)} \varphi_{[p]}(x).$$

The terms in the series are polynomially bounded as shown in Section 3.1 and the complexity of $\gamma_p^{(r)}$ is given by Corollary 1. □

Since a Hamming ball of radius r is a union over all spheres of radius at most r, the moment calculation can be trivially generalized to Hamming balls in the following manner.

Theorem 6. *Fix c and k. Let f be any k-bounded pseudo-Boolean function. Let $B^{(r)}(x)$ be a Hamming ball of radius r around an arbitrary state x. The quantity $\mu_c(B^{(r)}(x))$ (the c^{th} moment of f over the ball) can be computed in time polynomial in n.*

Proof.

$$\begin{aligned}\mu_c(B^{(r)}(x)) &= \frac{1}{|B^{(r)}(x)|}\sum_{s=0}^{r}\sum_{y\in S^{(s)}(x)} f(y)^c \\ &= \left(\sum_{s=0}^{r}\binom{n}{s}\right)^{-1}\sum_{s=0}^{r}\sum_{p=0}^{ck}\gamma_p^{(s)}\varphi_{[p]}(x) \\ &= \left(\sum_{s=0}^{r}\binom{n}{s}\right)^{-1}\sum_{p=0}^{ck}\varphi_{[p]}(x)\sum_{s=0}^{r}\gamma_p^{(s)}.\end{aligned}$$

As in Section 3.1, there are only $O(n^{c^2k})$ individual terms to compute in the series

$$\sum_{p=0}^{ck}\varphi_{[p]}(x).$$

Hence the calculation can be performed in time polynomial in n for a Hamming ball of any radius. □

It immediately follows that central moments of the distribution of f over Hamming regions can be computed in polynomial time in this way.

Corollary 2. *Fix c and k. Let f be any k-bounded pseudo-Boolean function. Let X be a Hamming region (sphere or ball) of some radius around an arbitrary state. The c^{th} central moment of f over X can be computed in time polynomial in n.*

Proof. This follows from the definition of central moments in terms of μ_c.

$$\sum_{i=0}^{c}\binom{c}{i}(-1)^{c-i}\mu_i(X)\mu_1(X)^{c-i}.$$

□

3.3 Algorithm to compute moments

Let us compute $\mu_c(B^{(r)}(x))$ for a function f. We first compute the nonzero Walsh coefficients of f and store them in a data structure W which is an array of (bitstring, value) pairs such that, for the j^{th} nonzero Walsh coefficient of f in some arbitrary order, $W[j] = (i, w_i)$. We shall assume that arrays are indexed from zero. Since f is k-bounded, this data structure can be constructed in polynomial time [8]. The eigenvalue $\gamma_p^{(r)}$ which is stored in an array *gamma*$[p][r]$ is computed using the recurrence in Equation (18).

We first must compute the sum of the needed coefficients over spheres of radius $s \leq r$. Let the function Tuples(c, d) return the set of all c-tuples over the index set $\{0, 1, \ldots, d-1\}$. The sum of f^c evaluated over a sphere of radius s around x can be computed as

```
SphereSum($x, s, c, W$)
 1  if $c = 0$ return 1
 2  $sum \leftarrow 0$
 3  for each $q \in$ Tuples($c$, $length[W]$)
 4      do $prod \leftarrow 1$
 5         $bits \leftarrow (000\ldots0)$
 6         for $j \leftarrow 0$ to $c - 1$
 7             do $(i, w_i) \leftarrow W[q[j]]$
 8                $bits \leftarrow bits \oplus i$
 9                $prod \leftarrow prod \times w_i$
10         $p \leftarrow \langle bits, bits \rangle$
11         $sum \leftarrow sum + prod \times gamma[p][s] \times (-1)^{\langle x, bits \rangle}$
12  return $sum$
```

Since k and c are constants, the number of tuples is a polynomial in n. Note that since multiplication and exclusive or are commutative operations, there are a large number of symmetries in the sum over all c-tuples. Thus the efficiency of the outer loop in lines 3 to 11 may be improved further using combinatorial enumeration techniques to remove these symmetries.

The ball moment $\mu_c(B^{(r)}(x))$ is computed as follows.

```
BallMoment($x, r, c, W$)
vol ← 0
$sum \leftarrow 0$
for $s \leftarrow 0$ to $r$
    do $sum \leftarrow sum +$ SphereSum($x, s, c, W$)
       $vol \leftarrow vol + \binom{n}{s}$
return $sum \,/\, vol$
```

4 Computing the expected fitness of mutations

Genetic algorithms operating on $\{0,1\}^n$ often employ some form of mutation in which each bit of a state (genotype) under consideration is flipped with some probability ρ, the so-called mutation rate. If the fitness function used is a k-bounded pseudo-Boolean function, we can also apply the decomposition presented in Section 3 to exactly compute the expected fitness of a mutated offspring.

We assume that the fitness function $f : \{0,1\}^n \to \mathbb{R}$ is k-bounded. Let $x \in \{0,1\}^n$ be the state under consideration. Mutation is a stochastic process that produces an offspring state z by changing components of x. Since the process is stochastic, we can characterize $f(z)$ as a random variable. We can calculate the expected value of this random variable as a function of $f(x)$: the fitness of the current state. In other words, we are interested in calculating the first moment of f over a ball of radius n around x, but now the sampling is no longer uniform throughout the region as it was in Equation (1). Indeed, the probability mass function of the random variable corresponding to $f(z)$ now depends on Hamming distance from x, which is captured by sphere membership.

To produce z, each bit of x is flipped with probability ρ. Thus z lies in a sphere of radius r around x with probability $\rho^r(1-\rho)^{n-r}$. The total fitness value in the sphere at radius r around x is

$$\sum_{y \in S^{(r)}(x)} f(y),$$

so the contribution to the expectation in a sphere at radius x can be obtained by multiplying this sum by the probability of the offspring lying in the sphere.

$$\rho^r(1-\rho)^{n-r} \sum_{y \in S^{(r)}(x)} f(y).$$

Since all spheres around x are disjoint, the expected fitness of the offspring of x under mutation can be computed as the sum of the expectation contributions from each sphere:

$$\begin{aligned}\sum_{r=0}^{n} \rho^r(1-\rho)^{n-r} \sum_{y \in S^{(r)}(x)} f(y) &= \sum_{r=0}^{n} \rho^r(1-\rho)^{n-r}\mathbf{S}^{(r)} f(x) && \text{by (17);}\\ &= \sum_{r=0}^{n} \rho^r(1-\rho)^{n-r} \sum_{p=0}^{k} \gamma_p^{(r)} \varphi_{[p]}(x).\end{aligned}$$

Thus the expected fitness of the offspring of x under mutation with mutation rate ρ can be computed by modifying the BallMoment computation in Section 3.3. Let W be the appropriate Walsh coefficient data structure corresponding to the fitness function f.

```
ExpectedFitness($x, \rho, W$)
sum ← 0
for $r \leftarrow 0$ to $n$
    do sum ← sum + $(\rho^r(1-\rho)^{n-r}) \times$ SphereSum($x, r, 1, W$)
return sum
```

In this case $c = 1$ so the time complexity of each call to SphereSum is $O(n^k)$. Summing over all n spheres gives a total complexity of $O(n^{k+1})$. Since the offspring can lie anywhere in $\{0,1\}^n$, a brute-force calculation of the exact expectation would require complete enumeration which has a time complexity of $\Theta(2^n)$.

Higher moments of the distribution of f under mutation can be obtained in an analogous manner.

5 Discussion

The class of k-bounded pseudo-Boolean functions plays an important role in many fields. In NK-landscape models [15], for instance, the fitness of a genotype (a string over a binary alphabet) is computed as a sum over individual k-ary gene interactions. NK-landscapes have also been employed to simulate landscapes that arise from RNA folding [4]. Another important family of k-bounded pseudo-Boolean functions are those of MAX-k-SAT problems. Local search algorithms such as variants of WALKSAT [18] have long been counted as among the state-of-the-art for solving critically constrained satisfiability problems. Since local search algorithms typically use the Hamming neighborhood as a search operator, they can be seen as exploring the landscapes described in this paper.

The results presented in this paper provide a general approach to computing moments of a k-bounded pseudo-Boolean function f over arbitrary radius regions (Hamming spheres and balls) in polynomial time. This is significant for the following reasons.

1. The calculation is exact, i.e., the moments are not approximated.

2. The calculation is computationally efficient with respect to naïve enumeration since, in general, the size of these regions is exponential in the bitstring length of the domain of f (for instance, spheres of radius $n/2$ or Hamming balls of radius $O(n)$).

Exact calculation of the moments affords opportunities not previously available for heuristic search algorithms that rely on directed sampling. The moments $\{\mu_0(X), \mu_1(X), \mu_2(X), \ldots\}$ characterize the distribution of values in the codomain of f over particular regions X of the landscape. In the context of local and genetic search, an algorithm might exploit this information by computing statistical information about unexplored regions of the landscape to determine how promising such a region might be for further exploration.

In the evolutionary computation community, the analysis of *hyperplanes* is important since the distribution of function values over hyperplanes influences the dynamics of evolutionary algorithms. By definition, a hyperplane of order m in $\{0,1\}^n$ is obtained by fixing m bit values in the neighborhood function, and allowing all other bits to vary. Since $\binom{n}{m}$ is a polynomial when m is a fixed constant, summary statistics for all hyperplanes up to m can be computed in polynomial time on k-bounded pseudo-Boolean functions [7]. This information is useful for making statistical inferences about sampling hyperplanes.

Fixing certain variables during search effectively induces a hyperplane in $\{0,1\}^n$ to which the search space becomes constrained. For example, assume we fix a binary variable to a specific variable assignment in a MAX-k-SAT problem. In effect, this transforms the objective function to a new objective function defined over the subspace (in this case a hyperplane) of $\{0,1\}^n$. These "partial" Hamming neighborhoods may arise in current heuristic search techniques for constraint satisfaction problems such as MAX-k-SAT where only certain variables are changed. Fixing a set of variables transforms an objective function f to a new objective function f', and, so long as f is k-bounded, the c^{th} moment of f' over arbitrary Hamming spheres in the induced subspace can also be computed in polynomial time.

The method we present in this paper computes the moments of the distribution of f over regions of the landscape. Hence, if we can *approximate* the distribution of f over a region by a probability density function parameterized by the known moments, this approximation may provide information about the *best values* of f in the region. Integrating this density with respect to codomain value supplies us with a cumulative distribution function over the region that could be used to estimate the probability of an optimal solution belonging to the region.

6 Conclusion

In this paper we have presented a general mechanism for computing moments of a k-bounded pseudo-Boolean function over arbitrary radius Hamming spheres. Our approach uses the fact that any epistatically bounded pseudo-Boolean function taken to a constant power can be characterized as a bounded linear combination of polynomially-computable eigenfunctions of radius-r sphere neighborhood structures.

These results hold the promise of changing the way sampling is done in evolutionary algorithms and local search algorithms since the methods provide a principled way of quickly assessing moments of the fitness in regions of the search space without doing explicit sampling. The method we have presented applies to all k-bounded pseudo-Boolean functions such as MAX-k-SAT and its variants, NK-landscapes, spin glass models, and graph optimization problems such as MAX-CUT.

Acknowledgments

This research was sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant number FA9550-08-1-0422. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon.

References

[1] Eric Angel and Vassilis Zissimopoulos. On the landscape ruggedness of the Quadratic Assignment Problem. *Theoretical Computer Science*, 263(1–2):159–172, 2001.

[2] Albert D. Bethke. *Genetic Algorithms as Function Optimizers.* PhD thesis, University of Michigan, 1980.

[3] Sung-Soon Choi, Kyomin Jung, and Jeong Han Kim. Almost tight upper bound for finding Fourier coefficients of bounded pseudo-Boolean functions. In Rocco A. Servedio and Tong Zhang, editors, *Proceedings of the 21st Conference on Learning Theory (COLT 2008)*, pages 123–134, Helsinki, Finland, July 2008. Omnipress.

[4] W. Fontana, P.F. Stadler, E.G. Bornberg-Bauer, T. Griesmacher, I.L. Hofacker, M. Tacker, P. Tarazona, E.D. Weinberger, and P. Schuster. RNA folding and combinatory landscapes. *Physical review E*, 47(3):2083–2099, 1993.

[5] David E. Goldberg. Genetic algorithms and Walsh functions. *Complex Systems*, 3:129–171, 1989.

[6] Lov K. Grover. Local search and the local structure of NP-complete problems. *Operations Research Letters*, 12:235–243, 1992.

[7] Robert B. Heckendorn. Embedded landscapes. *Evolutionary Computation*, 10(4):345–369, 2002.

[8] Robert B. Heckendorn, Soraya Rana, and Darrell Whitley. Polynomial time summary statistics for a generalization of MAXSAT. In *Genetic and Evolutionary Computation Conference (GECCO-1999)*, pages 281–288, 1999.

[9] Robert B. Heckendorn, Soraya B. Rana, and L. Darrell Whitley. Test function generators as embedded landscapes. In Wolfgang Banzhaf and Colin R. Reeves, editors, *Foundations of Genetic Algorithms (FOGA-5)*, pages 183–198. Morgan Kaufmann, 1998.

[10] Robert B. Heckendorn and Alden H. Wright. Efficient linkage discovery by limited probing. *Evolutionary Computation*, 12:517–545, 2004.

[11] John H. Holland. *Adaptation in Natural and Artificial Systems.* The University of Michigan Press, 1975.

[12] Terry Jones. *Evolutionary Algorithms, Fitness Landscapes and Search.* PhD thesis, University of New Mexico, Albuquerque, New Mexico, May 1995.

[13] Terry Jones and Stephanie Forrest. Fitness distance correlation as a measure of problem difficulty for genetic algorithms. In *International Conference on Genetic Algorithms (ICGA 95)*, 1995.

[14] Hillol Kargupta and Byung-hoon Park. Gene expression and fast construction of distributed evolutionary representation. *Evolutionary Computation*, 9(1):43–69, 2001.

[15] Stuart A. Kauffman. *The Origins of Order.* Oxford University Press, 1993.

[16] Peter Merz. Advanced fitness landscape analysis and the performance of memetic algorithms. *Evolutionary Computation*, 12(3):303–326, 2004.

[17] Christian M. Reidys and Peter F. Stadler. Combinatorial landscapes. *SIAM Review*, 44:3–54, 2002.

[18] Bart Selman, Henry Kautz, and Bram Cohen. Local search strategies for satisfiability testing. In David S. Johnson and Michael A. Trick, editors, *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, volume 26. AMS, 1996.

[19] Peter F. Stadler. Toward a theory of landscapes. In R. Lopéz-Peña, R. Capovilla, R. García-Pelayo, H. Waelbroeck, and F. Zertruche, editors, *Complex Systems and Binary Networks*, pages 77–163. Springer Verlag, 1995.

[20] Peter F. Stadler. Landscapes and their correlation functions. *Journal of Mathematical Chemistry*, 20:1–45, 1996.

[21] Peter F. Stadler. Spectral landscape theory. In J.P. Crutchfield and P. Schuster, editors, *Evolutionary Dynamics - Exploring the Interplay of Selection, Neutrality, Accident, and Function*. Oxford University Press, 2002.

[22] Andrew M. Sutton, L. Darrell Whitley, and Adele E. Howe. A polynomial time computation of the exact correlation structure of k-satisfiability landscapes. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-09)*, 2009.

[23] Joseph L. Walsh. A closed set of normal orthogonal functions. *American Journal of Mathematics*, 45(1):5–24, 1923.