Properties of Symmetric Fitness Functions

Sung-Soon Choi
School of Computer Science and Engineering
Seoul National University
Sillim-dong, Gwanak-gu,
Seoul, 151-744 Korea
sschoi@soar.snu.ac.kr

Yung-Keun Kwon
School of Computer Science and Engineering
Seoul National University
Sillim-dong, Gwanak-gu,
Seoul, 151-744 Korea
kwon@soar.snu.ac.kr

Byung-Ro Moon
School of Computer Science and Engineering
Seoul National University
Sillim-dong, Gwanak-gu,
Seoul, 151-744 Korea
moon@soar.snu.ac.kr

ABSTRACT

The properties of symmetric fitness functions are investigated. We show that a well-known encoding scheme inducing symmetric functions has the non-synonymous property and the search spaces obtained from symmetric functions have the zero-correlation structures. The Walsh analysis reveals the properties of symmetric functions related to additive separability, problem difficulty measures and so on. Our results support the claim of other researchers that the search spaces with symmetry induce relatively difficult problems. The results also present some limitations of existing problem difficulty measures for symmetric fitness functions.

Categories and Subject Descriptors

F.2.2 [**Nonnumerical Algorithms and Problems**]: Computations on discrete structures

General Terms

Theory

Keywords

Symmetric fitness functions, encoding scheme, search space analysis, problem difficulty measures, Walsh analysis

1. INTRODUCTION

Let Σ be a set of characters or an alphabet. Denoted by $x[i]$ is the i^{th} character of x for a string $x \in \Sigma^n$. For $x \in \Sigma^n$ and a permutation π on Σ, $\pi(x)$ is defined to be the string in Σ^n satisfying $\pi(x)[i] = \pi(x[i])$. A function $f : \Sigma^n \to \mathbb{R}$ is *symmetric* if $f(x) = f(\pi(x))$ for any $x \in \Sigma^n$ and any permutation π on Σ. Let $\Pi(x)$ be the set of strings obtained by a permutation on x:

$$\Pi(x) = \{\pi(x) | \pi \text{ is a permutation on } \Sigma\}.$$

GECCO'06, July 8–12, 2006, Seattle, Washington, USA.

So, a symmetric function f is invariant on the set $\Pi(x)$ for any $x \in \Sigma^n$. We call $\Pi(x)$ the *symmetric class* of x.

Symmetric fitness functions have been discussed in a number of papers in the area of evolutionary computation. Some of them were designed for studying the dynamics of evolutionary algorithms on search spaces with symmetry. Examples appeared in the two-max problem [1] [2], the H-IFF problem [3], and the connected clusters problem [4]. Some others were produced in encoding a class of combinatorial optimization problems. The problems are called *grouping problems* [5] and commonly concerned with partitioning a given item set into mutually disjoint subsets. In the problems, the k-ary encoding, in which k subsets are represented by the integers from 0 to $k-1$, has been generally used. Examples of the problems include the max-cut problem [6], graph partitioning [7] [8] [9], graph coloring, bin packing, and workshop layouting [5]. Besides the above, some Ising problems [4] [10], which originated from statistical physics, induce symmetric fitness functions, too.

There have been studies for the search spaces with symmetry. Throughout a series of papers [11] [12] [2] [4], Van Hoyweghen *et al.* provided the results that the search spaces induced by symmetric fitness functions cause a *synchronization problem*. The synchronization problem refers that an evolutionary algorithm is stuck in a local optimum because *noninferior building blocks* of different optima cannot be combined to improve, and so leads to the slow genetic search. For alleviating the problem, they proposed some solutions.

There were some results indicating that traditional crossover operators may lose the power of exploitation in the search spaces induced by symmetric fitness functions. In order to alleviate the problems, a few approaches were proposed including adaptive crossovers that recombine parents in terms of phenotypes [7] [13] [14] and *normalization* methods that transform one parent to another genotype to be consistent with the other parent and recombine the parents using traditional crossovers [8] [9] [15] [16]. Recently, it was shown that a well-devised crossover operator outperforms a mutation operator in searching optima on the H-IFF problem [17] and the Ising problem on the ring [10], both of which induce symmetric fitness functions.

In this paper, we consider the properties of symmetric fitness functions. The properties that we investigate are concerned with encoding schemes, the correlation structures of search spaces, the Walsh analysis of fitness functions, and statistical measures of problem difficulty. Rigorous analysis from a number of viewpoints gives new insights into the symmetric functions. Our results support the previous empirical results that the search spaces with symmetry induce relatively difficult problems.

This paper is organized as follows. In Section 2, we present the non-synonymous property of the encoding schemes in which a phenotype is represented by the strings of a symmetric class. In Section 3, correlation structures of landscapes induced by symmetric functions are investigated. We analyze symmetric functions using Walsh transform and provide their properties including those related to additive separability in Section 4. In Section 5, some results of symmetric functions are presented in terms of a few statistical measures of problem difficulty. Finally, we make our conclusions in Section 6.

2. MAXIMALLY NON-SYNONYMOUS PROPERTY

The notion of synonymous/non-synonymous encodings was first introduced by Rothlauf and Goldberg [18]. According to them, a redundant encoding is *synonymous* if the genotypes that are assigned to the same phenotype are similar to each other and is *non-synonymous* otherwise. They showed that for synonymously redundant encodings GA performance does not change and that non-synonymously redundant encodings do harm to the GA performance. In this section, we consider the encodings in which a phenotype is represented by the strings of a symmetric class. As stated above, such encodings have been generally used in grouping problems. We show that the encodings have the maximally non-synonymous property.

Given the phenotype space $\mathcal{S}$ for a problem, a redundant encoding is a surjective genotype-phenotype map $\varphi : \mathcal{G} \to \mathcal{S}$, where $\mathcal{G}$ represents the genotype space and $|\mathcal{G}| > |\mathcal{S}|$. For a phenotype $s \in \mathcal{S}$, we define $\mathcal{H}(s) \subseteq \mathcal{G}$ by

$$\mathcal{H}(s) = \{x \in \mathcal{G} | \varphi(x) = s\},$$

which is the set of genotypes that share the phenotype s. We call $\mathcal{H}(s)$ the *coset* (or *neutral set*) corresponding to s.

Suppose that we are given a redundant encoding $\varphi : \mathcal{G} \to \mathcal{S}$ and a distance metric d defined on $\mathcal{G}$. We assume that d is the Hamming distance metric following the convention. The sum of the distances between genotypes is decomposed into

$$\sum_{x,y \in \mathcal{G}} d(x,y) = \mathfrak{D}(\varphi) + \mathfrak{D}'(\varphi),$$

where

$$\mathfrak{D}(\varphi) = \sum_{s \in \mathcal{S}} \sum_{x,y \in \mathcal{H}(s)} d(x,y)$$

and

$$\mathfrak{D}'(\varphi) = \sum_{s \neq t \in \mathcal{S}} \sum_{x \in \mathcal{H}(s)} \sum_{y \in \mathcal{H}(t)} d(x,y).$$

$\mathfrak{D}(\varphi)$ is the sum of the distances among the genotypes belonging to the same coset and $\mathfrak{D}'(\varphi)$ is the sum of the distances between the genotypes belonging to different cosets. Since $\mathfrak{D}(\varphi)$ indicates how closely the genotypes in cosets are located in the genotype space for the encoding φ, an encoding φ may be regarded as synonymous if the value of $\mathfrak{D}(\varphi)$ is relatively small (equivalently, the value of $\mathfrak{D}'(\varphi)$ is relatively large since $\mathfrak{D}(\varphi) + \mathfrak{D}'(\varphi)$ is constant for a fixed $\mathcal{G}$) and non-synonymous otherwise. A measure equivalent to $\mathfrak{D}(\varphi)$ was mentioned in [18].

We first investigate which encoding has a maximally non-synonymous property.

THEOREM 1. *Suppose that a phenotype space $\mathcal{S}$ is encoded by the genotype space Σ^n, where the coset sizes of phenotypes are pre-defined. If an encoding φ satisfies that*

$$|\{x \in \mathcal{H}(s) | x[i] = \mathtt{a}\}| = |\{x \in \mathcal{H}(s) | x[i] = \mathtt{b}\}| \qquad (1)$$

for all $s \in \mathcal{S}$, $i \in \{0, 1, \ldots, n-1\}$, and $\mathtt{a}, \mathtt{b} \in \Sigma$, then φ maximizes $\mathfrak{D}(\varphi)$.

PROOF. Note that

$$\begin{aligned} \mathfrak{D}(\varphi) &= \sum_{s \in \mathcal{S}} \sum_{x,y \in \mathcal{H}(s)} d(x,y) \\ &= \sum_{s \in \mathcal{S}} \sum_{x,y \in \mathcal{H}(s)} \sum_{i=1}^{n} 1(x[i] \neq y[i]) \\ &= \sum_{s \in \mathcal{S}} \sum_{i=1}^{n} \sum_{x,y \in \mathcal{H}(s)} 1(x[i] \neq y[i]), \end{aligned}$$

where $1(\cdot)$ means the indicator function. Now we consider the value of $\sum_{x,y \in \mathcal{H}(s)} \delta(x[i], y[i])$ for a fixed $s \in \mathcal{S}$ and i. Let $t_\mathtt{a} = |\{x \in \mathcal{H}(s) | x[i] = \mathtt{a}\}|$ for $\mathtt{a} \in \Sigma$. Then,

$$\begin{aligned} \sum_{x,y \in \mathcal{H}(s)} \delta(x[i], y[i]) &= \sum_{\mathtt{a} \neq \mathtt{b} \in \Sigma} t_\mathtt{a} t_\mathtt{b} \\ &= \sum_{\mathtt{a},\mathtt{b} \in \Sigma} t_\mathtt{a} t_\mathtt{b} - \sum_{\mathtt{a} \in \Sigma} t_\mathtt{a}^2 \\ &= \left(\sum_{\mathtt{a} \in \Sigma} t_\mathtt{a} \right)^2 - \sum_{\mathtt{a} \in \Sigma} t_\mathtt{a}^2. \end{aligned}$$

Since $\sum_{\mathtt{a} \in \Sigma} t_\mathtt{a} = |\mathcal{H}(s)|$, $\sum_{x,y \in \mathcal{H}(s)} \delta(x[i], y[i])$ is maximized if and only if $\sum_{\mathtt{a} \in \Sigma} t_\mathtt{a}^2$ is minimized. By the Lagrange multiplier method, $\sum_{\mathtt{a} \in \Sigma} t_\mathtt{a}^2$ subject to $\sum_{\mathtt{a} \in \Sigma} t_\mathtt{a} = |\mathcal{H}(s)|$ is minimized when $t_\mathtt{a} = t_\mathtt{b}$ for all $\mathtt{a}, \mathtt{b} \in \Sigma$. This completes the proof. □

THEOREM 2. *Suppose that a phenotype space $\mathcal{S}$ is encoded by the genotype space Σ^n, where the coset sizes of phenotypes are defined to be multiples of $|\Sigma|$. If an encoding φ maximizes $\mathfrak{D}(\varphi)$, then it satisfies the condition (1).*

PROOF. From Theorem 1, it is enough to show that there exists an encoding satisfying the condition (1). Let π be a permutation on Σ such that $\pi^i(\mathtt{a}) \neq \mathtt{a}$ for all $1 \leq i \leq |\Sigma| - 1$ and all $\mathtt{a} \in \Sigma$. We partition the genotype space Σ^n into the orbits generated by the permutation π. Since $|\mathcal{H}(s)|$ is a multiple of $|\Sigma|$ for each $s \in \mathcal{S}$ and the size of each orbit is $|\Sigma|$, there is an encoding φ in which the genotypes in each orbit belong to the same coset. The encoding φ satisfies the condition (1) and the proof completes. □

Inspired by the above theorems, we call an encoding satisfying the condition (1) a *maximally non-synonymous encoding*. Given a maximally non-synonymous encoding, if we choose a genotype from a fixed coset uniformly at random, the allele in each gene follows the uniform distribution on the alphabet.

COROLLARY 1. *Suppose that each phenotype is represented by the strings of a symmetric class under an encoding φ. Then, φ is maximally non-synonymous.*

The corollary implies that the k-ary encoding for grouping problems is maximally non-synonymous. Combining the corollary with the result of Rothlauf and Goldberg [18], it is partially answered why traditional crossovers struggle in grouping problems.

3. CORRELATION STRUCTURE

3.1 Correlation Measures

It is crucial to understand the structure of the search space landscape for predicting the performance of a given algorithm. Ruggedness is one of the most important features to characterize the structure of a landscape. It is believed that, the more rugged the landscape is, the worse the performance of an evolutionary algorithm is. Ruggedness of a search space landscape is strongly related to correlation between fitness and distance: A smooth landscape induces a high correlation between fitness and distance while a rugged landscape does a low correlation. For this reason, many measures for estimating the ruggedness of a landscape are based on the correlation. Examples include *autocorrelation* [19], *correlation length* [20], *fitness distance correlation* (FDC) [21], and the measures by Boese *et al.* [22].

FDC proposed by Jones and Forrest [21] have been widely used for predicting problem difficulty in evolutionary computation area [23] [24]. The FDC coefficient is defined by

$$\rho_{\mathrm{FDC}} = \mathrm{Corr}[f_{\mathrm{opt}}, d_{\mathrm{opt}}] = \frac{\mathrm{Cov}[f_{\mathrm{opt}}, d_{\mathrm{opt}}]}{\sigma[f_{\mathrm{opt}}] \cdot \sigma[d_{\mathrm{opt}}]},$$

where f_{opt} and d_{opt} are random variables representing the fitness difference and distance between a sampled genotype and its nearest optimum, respectively. For a minimization problem, a value of $\rho_{\mathrm{FDC}} = 1.0$ means that fitness and distance are perfectly correlated and evolutionary search will be successful. A value of ρ_{FDC} close to zero indicates that fitness and distance are not linearly correlated and supports that evolutionary search may fail for the problem. Boese *et al.* [22] proposed another correlation-based measure, ρ_{avg}. It is defined by

$$\rho_{\mathrm{avg}} = \mathrm{Corr}[f_{\mathrm{opt}}, d_{\mathrm{avg}}] = \frac{\mathrm{Cov}[f_{\mathrm{opt}}, d_{\mathrm{avg}}]}{\sigma[f_{\mathrm{opt}}] \cdot \sigma[d_{\mathrm{avg}}]},$$

where f_{opt} is a random variable representing the fitness difference between a sampled genotype and an optimum and d_{avg} is a random variable representing the average distance from the sampled genotype to others. They applied the measure to local optima space for a few combinatorial optimization problems to show that a globally convex structure (also known as "big valley") appears in the landscapes of the problems.

For a genotype x, we consider a correlation measure ρ_x as follows.

$$\rho_x = \mathrm{Corr}[f_x, d_x] = \frac{\mathrm{Cov}[f_x, d_x]}{\sigma[f_x] \cdot \sigma[d_x]},$$

where f_x and d_x are random variables representing the fitness difference and distance between a sampled genotype and x, respectively. The measure ρ_x reflects the correlation structure between fitness and distance toward x. If x is the unique optimum in the genotype space, the two measures ρ_{FDC} and ρ_x are equivalent. In multimodal landscapes, however, nearest optima for different genotypes may be different, which generally induces a difference between ρ_{FDC} and ρ_x with an optimum x.

A typical evolutionary search may be considered as a process in which a population of genotypes are evolved to converge toward an optimal (or near-optimal) genotype. On the other hand, more than one optimal genotype are generally reflected into the value of ρ_{FDC} in multimodal landscapes. In the landscapes induced by symmetric fitness functions, a number of optimal genotypes are located distantly from one another as seen in the last section, so the measure ρ_{FDC} has a potential to mis-estimate the evolvability of an evolutionary algorithm. For that reason, we use the measure ρ_x with an optimum x (instead of ρ_{FDC}) for analyzing the landscapes induced by symmetric fitness functions.

3.2 Zero Correlation for Symmetric Functions

Now we investigate the properties of the landscapes induced by symmetric fitness functions. For two subsets A and B of Σ^n, we denote by $\bar{d}(A, B)$ the expected distance between the genotypes chosen uniformly at random from A and B.

PROPOSITION 1. *For any subset A of Σ^n and any $x \in \Sigma^n$,*

$$\bar{d}(A, \Pi(x)) = \left(\frac{|\Sigma| - 1}{|\Sigma|}\right) n.$$

PROOF. Note that

$$\begin{aligned} \bar{d}(A, \Pi(x)) &= \frac{1}{|A| \cdot |\Pi(x)|} \sum_{a \in A} \sum_{y \in \Pi(x)} d(a, y) \\ &= \frac{1}{|A| \cdot |\Pi(x)|} \sum_{a \in A} \sum_{y \in \Pi(x)} \sum_{i=1}^{n} 1(a[i] \neq y[i]) \\ &= \frac{1}{|A| \cdot |\Pi(x)|} \sum_{i=1}^{n} \sum_{a \in A} \sum_{y \in \Pi(x)} 1(a[i] \neq y[i]). \end{aligned}$$

Since $\sum_{y \in \Pi(x)} 1(a[i] \neq y[i]) = \left(\frac{|\Sigma|-1}{|\Sigma|}\right) |\Pi(x)|$ for any $a \in A$,

$$\begin{aligned} \bar{d}(A, \Pi(x)) &= \frac{1}{|A| \cdot |\Pi(x)|} \sum_{i=1}^{n} \sum_{a \in A} \left(\frac{|\Sigma| - 1}{|\Sigma|}\right) |\Pi(x)| \\ &= \left(\frac{|\Sigma| - 1}{|\Sigma|}\right) n. \end{aligned}$$

□

The proposition means that a symmetric class is at the same distance from any set of genotypes on average. It implies that fitnesses and distances are not linearly correlated in the landscapes induced by symmetric fitness functions.

THEOREM 3. *For any symmetric fitness function and any $x \in \Sigma^n$, $\rho_x = 0$ and $\rho_{\mathrm{avg}} = 0$.*

PROOF. To prove that $\rho_x = 0$, it is enough to show that $\mathrm{Cov}[f_x, d_x] = \mathrm{E}[f_x d_x] - \mathrm{E}[f_x] \cdot \mathrm{E}[d_x] = 0$. We denote by $\Delta f(x, y)$ the fitness difference between x and y for $y \in \Sigma^n$. Then

$$\begin{aligned} \mathrm{E}[f_x d_x] &= \sum_{\Pi} \Pr[y \in \Pi] \cdot \mathrm{E}[\Delta f(x, y) d(x, y) | y \in \Pi] \\ &= \sum_{\Pi} \Pr[y \in \Pi] \cdot \Delta f(x, y) \cdot \mathrm{E}[d(x, y) | y \in \Pi], \end{aligned}$$

where the summation is over symmetric classes Π's. From Proposition 1, $\mathrm{E}[d(x, y) | y \in \Pi] = \bar{d}(\{x\}, \Pi) = \left(\frac{|\Sigma|-1}{|\Sigma|}\right) n$, which is invariant over symmetric classes. So,

$$\begin{aligned}
\mathrm{E}[f_x d_x] &= \left(\frac{|\Sigma|-1}{|\Sigma|}\right) n \cdot \left(\sum_{\Pi} \Pr[y \in \Pi] \cdot \Delta f(x,y)\right) \\
&= \left(\frac{|\Sigma|-1}{|\Sigma|}\right) n \cdot \mathrm{E}[f_x].
\end{aligned}$$

Since

$$\begin{aligned}
\mathrm{E}[d_x] &= \sum_{\Pi} \Pr[y \in \Pi] \cdot \mathrm{E}[d(x,y) | y \in \Pi] \\
&= \left(\frac{|\Sigma|-1}{|\Sigma|}\right) n \cdot \left(\sum_{\Pi} \Pr[y \in \Pi]\right) \\
&= \left(\frac{|\Sigma|-1}{|\Sigma|}\right) n,
\end{aligned}$$

it follows that $\mathrm{E}[f_x d_x] = \mathrm{E}[f_x] \cdot \mathrm{E}[d_x]$. We omit the proof of $\rho_{\mathrm{avg}} = 0$, which is derived in a similar way. □

The landscape induced by a symmetric fitness function does not have a globally convex structure. It is not linearly correlated between fitness and distance toward any optimum. So, it is not expected that an evolutionary algorithm for a symmetric fitness function shows high-performance.

4. WALSH ANALYSIS

In this section, we restrict our attention to the fitness functions defined on $\{0,1\}^n$. These functions are called *pseudo-Boolean*.

4.1 Walsh Transform

In the area of evolutionary computation, many interesting results were obtained by Walsh analysis that is based on Walsh transform. Walsh transform is a Fourier transform for pseudo-Boolean functions in which a pseudo-Boolean function is represented as a linear combination of Walsh functions [25].

Let $[n]$ be the set of the integers 1 through n. A binary string $x \in \{0,1\}^n$ may be viewed as a subset of $[n]$ consisting of the positions in which x has non-zero values. The Walsh function corresponding to a subset m of $[n]$, $\psi_m(x) : \{0,1\}^n \to \mathbb{R}$, is a pseudo-Boolean function defined as

$$\psi_m(x) = (-1)^{|m \cap x|}.$$

We call m and $|m|$ the support set and the order of ψ_m, respectively. (The support set and the order of a pseudo-Boolean function are defined in the below.) If we define an inner product of two pseudo-Boolean functions f and g as

$$\langle f, g \rangle = \sum_{x \in \{0,1\}^n} \frac{f(x) \cdot g(x)}{2^n},$$

the set of Walsh functions, $\{\psi_m | m \subseteq [n]\}$, becomes an orthonormal basis of the space of pseudo-Boolean functions. Hence, a pseudo-Boolean function f can be represented as

$$f = \sum_{m \subseteq [n]} \hat{f}(m) \cdot \psi_m,$$

where $\hat{f}(m) = \langle f, \psi_m \rangle$ is called the Walsh coefficient corresponding to m. We refer to [26] for surveys of the properties of Walsh functions and Walsh transform.

For two subsets s and t of $[n]$, we denote by $s \triangle t$ the symmetric difference of s and t: $s \triangle t = (s \setminus t) \cup (t \setminus s)$. Note that $x \triangle [n]$ means the bitwise complement of x and a symmetric function f satisfies the property that $f(x) = f(x \triangle [n])$ for all $x \in \{0,1\}^n$.

THEOREM 4. *Suppose that f is symmetric. Then, $\hat{f}(m) = 0$ for $m \subseteq [n]$ with odd $|m|$.*

PROOF. Since $f(x) = f(x \triangle [n])$ and $\psi_m(x) = -\psi_m(x \triangle [n])$ from the definition of the Walsh function,

$$f(x) \cdot \psi_m(x) + f(x \triangle [n]) \cdot \psi_m(x \triangle [n]) = 0$$

and so

$$\begin{aligned}
\hat{f}(m) &= \langle f, \psi_m \rangle \\
&= \sum_{x \in \{0,1\}^n} \frac{f(x) \cdot \psi_m(x)}{2^n} \\
&= \frac{1}{2} \sum_{x \in \{0,1\}^n} \frac{f(x) \cdot \psi_m(x) + f(x \triangle [n]) \cdot \psi_m(x \triangle [n])}{2^n} \\
&= 0.
\end{aligned}$$

□

THEOREM 5. *Suppose that $\hat{f}(m) = 0$ for all $m \subseteq [n]$ of odd $|m|$. Then, f is symmetric.*

PROOF. Let

$$f = \sum_{m \subseteq [n] \,:\, |m| \text{ is even}} \hat{f}(m) \cdot \psi_m.$$

Since $\psi_m(x) = \psi_m(x \triangle [n])$ for all $m \subseteq [n]$ with even $|m|$,

$$\begin{aligned}
f(x \triangle [n]) &= \sum_{m \subseteq [n] \,:\, |m| \text{ is even}} \hat{f}(m) \cdot \psi_m(x \triangle [n]) \\
&= \sum_{m \subseteq [n] \,:\, |m| \text{ is even}} \hat{f}(m) \cdot \psi_m(x) \\
&= f(x).
\end{aligned}$$

□

From the theorems, we have

COROLLARY 2. *A pseudo-Boolean function f is symmetric if and only if it can be represented as a linear combination of Walsh functions with even orders.*

4.2 2-Bounded Symmetric Functions

For a pseudo-Boolean function f, the *order* of f is the maximum order of Walsh functions that have non-zero Walsh coefficients in the Walsh transform of f. A function f is called *k-bounded* if the order of f is less than or equal to k. In this section, we investigate the properties of the 2-bounded symmetric functions.

From Corollary 2, a 2-bounded symmetric function f is represented as a linear combination of the Walsh functions of order zero and two:

$$f = \sum_{m \,:\, |m| = 0 \text{ or } 2} \hat{f}(m) \cdot \psi_m.$$

For a set m with $|m| = 2$, let the two bit positions in m be i_m and j_m. Then, the value of $\psi_m(x)$ is -1 if $x[i_m] \neq x[j_m]$ and 1 otherwise, and so

$$\psi_m(x) = 1 - 2 \cdot 1(x[i_m] \neq x[j_m]).$$

Using the fact that $\psi_\emptyset(x) = 1$ for all $x \in \{0,1\}^n$,

$$\begin{aligned} f(x) &= \hat{f}(\emptyset) + \sum_{m\,:\,|m|=2} \hat{f}(m)\cdot(1 - 2\cdot 1(x[i_m] \neq x[j_m])) \\ &= \sum_{m\,:\,|m|=0 \text{ or } 2} \hat{f}(m) \\ &\quad + \sum_{m\,:\,|m|=2} -2\cdot\hat{f}(m)\cdot 1(x[i_m]\neq x[j_m]) \\ &= D_f + C_f(x), \qquad (2) \end{aligned}$$

where $D_f = \sum_{m\,:\,|m|=0 \text{ or } 2} \hat{f}(m)$ and $C_f(x) = \sum_{m\,:\,|m|=2} -2\cdot \hat{f}(m)\cdot 1(x[i_m]\neq x[j_m])$.

Given a 2-bounded symmetric function f, consider a weighted graph G as follows. Set a vertex for each bit position. For each Walsh function ψ_m such that $\hat{f}(m) \neq 0$, set the edge connecting the two positions in m and assign the weight $-2\cdot\hat{f}(m)$ on it. A binary string $x \in \{0,1\}$ may be viewed as a bipartition of bit positions in which a bit position i belongs to the "0" group if $x[i] = 0$ and it belongs to the "1" group otherwise. When we partition the bit positions into the two groups as indicated in x, $C_f(x)$ is the sum of weights of edges whose endpoints belong to different groups, which is called the *cut size* of a bipartition x. Equation (2) means that the problem of maximizing f is equivalent to the max-cut problem on G, which is to find a bipartition of the vertices of G maximizing the cut size.

The NP-hardness of the max-cut problem [27] implies that the problem of maximizing 2-bounded symmetric functions is also NP-hard. Approximation algorithms for the max-cut problem may be applied to the problem of maximizing 2-bounded symmetric functions but their approximabilities depend on the values of Walsh coefficients of a given function. For example, if $D_f \geq 0$ and the Walsh coefficients of order two are all non-positive, the algorithm of Goemans and Williamson [28] guarantees the 0.879-approximation ratio. Since the max-cut problem is MAX-SNP-hard [29] [30], there does not exist a polynomial time approximation scheme for the problem of maximizing 2-bounded symmetric functions unless P = NP.

4.3 Additive Separability

For a pseudo-Boolean function f, the set of bit positions that affect f is called the *support set* of f and denoted by $\mathfrak{s}(f)$. A function $f : \{0,1\}^n \to \mathbb{R}$ is *additively separable* into g and h if f can be represented as $g + h$ and $\mathfrak{s}(g)$ and $\mathfrak{s}(h)$ are disjoint. When the support sets of g and h are included in the subsets s and t of $[n]$, respectively, we also say that f is additively separable into s and t. For a subset s of $[n]$, the set of binary strings over the bit positions in s is denoted by B_s. If the support set of a function f is included in a subset s of $[n]$, the f *restricted to* s means the function that is defined on B_s and inherits the value of f.

Proposition 2. *A pseudo-Boolean function $f : \{0,1\}^n \to \mathbb{R}$ is additively separable into the subsets s and t of $[n]$ if and only if $\hat{f}(m) = 0$ for all $m \subseteq [n]$ such that $m \cap s \neq \emptyset$ and $m \cap t \neq \emptyset$.*

Proof. Suppose that f is additively separable into the subsets s and t of $[n]$. Then, we may write $f = g + h$, where $\mathfrak{s}(g) \subseteq s$ and $\mathfrak{s}(h) \subseteq t$. Since $\mathfrak{s}(g) \subseteq s$ ($\mathfrak{s}(h) \subseteq t$, resp.), g (h, resp.) may be represented as a linear combination of ψ_m's with $m \subseteq s$ ($m \subseteq t$, resp.). So, f can be represented as a linear combination of ψ_m's with $m \subseteq s$ or $m \subseteq t$. From the fact that ψ_m's constitute an orthonormal basis of the space of pseudo-Boolean functions, f is uniquely represented as a linear combination of ψ_m's and so the result follows.

Now consider the opposite side. Suppose that $\hat{f}(m) = 0$ for all $m \subseteq [n]$ such that $m \cap s \neq \emptyset$ and $m \cap t \neq \emptyset$. Then, we may decompose f as

$$f = \sum_{m\subseteq s} \hat{f}(m)\cdot\psi_m + \sum_{m\subseteq t}\hat{f}(m)\cdot\psi_m.$$

Letting $g = \sum_{m\subseteq s}\hat{f}(m)\cdot\psi_m$ and $h = \sum_{m\subseteq t}\hat{f}(m)\cdot\psi_m$, we have that $f = g + h$ with $\mathfrak{s}(g) \subseteq s$ and $\mathfrak{s}(h) \subseteq t$. □

Theorem 6. *Suppose that f is symmetric. If f is additively separable into the subsets s and $[n] \setminus s$ of $[n]$, $f(x) = f(x \triangle s)$ for all $x \in \{0,1\}^n$.*

Proof. From Proposition 2, f can be represented as

$$f = \sum_{m\subseteq s\,:\,even\,|m|}\hat{f}(m)\cdot\psi_m + \sum_{m\subseteq[n]\setminus s\,:\,even\,|m|}\hat{f}(m)\cdot\psi_m.$$

If we denote

$$g = \sum_{m\subseteq s\,:\,even\,|m|}\hat{f}(m)\cdot\psi_m$$

and

$$h = \sum_{m\subseteq[n]\setminus s\,:\,even\,|m|}\hat{f}(m)\cdot\psi_m,$$

we see that $g(x \triangle s) = g(x)$ and $h(x \triangle s) = h(x)$ for all $x \in \{0,1\}^n$ and so $f(x \triangle s) = f(x)$ for all $x \in \{0,1\}^n$. □

The converse of Theorem 6 does not hold for arbitrary symmetric functions but it does hold for 2-bounded symmetric functions.

Lemma 1. *If $f(x) = f(x \triangle s)$ for all $x \in \{0,1\}^n$, $\hat{f}(m) = 0$ for all $m \subseteq [n]$ such that $|m \cap s|$ is odd.*

Proof. Since $f(x) = f(x \triangle s)$ and $\psi_m(x) = -\psi_m(x \triangle s)$,

$$f(x)\cdot\psi_m(x) + f(x\triangle s)\cdot\psi_m(x\triangle s) = 0$$

and so

$$\begin{aligned} \hat{f}(m) &= \langle f, \psi_m\rangle \\ &= \sum_{x\in\{0,1\}^n}\frac{f(x)\cdot\psi_m(x)}{2^n} \\ &= \frac{1}{2}\sum_{x\in\{0,1\}^n}\frac{f(x)\cdot\psi_m(x) + f(x\triangle s)\cdot\psi_m(x\triangle s)}{2^n} \\ &= 0. \end{aligned}$$

□

Theorem 7. *A 2-bounded symmetric function f is additively separable into the subsets s and $[n]\setminus s$ of $[n]$ if and only if $f(x) = f(x \triangle s)$ for all $x \in \{0,1\}^n$.*

Proof. The only-if part was proven in Theorem 6 and we prove the if part. Suppose that $f(x) = f(x \triangle s)$ for all $x \in \{0,1\}^n$. Since f is 2-bounded, it is enough to show that $\hat{f}(m) = 0$ for all $m \subseteq [n]$ such that $|m \cap s| = 1$ and $|m \cap ([n]\setminus s)| = 1$, which is proven by Lemma 1. □

THEOREM 8. *Suppose that f is symmetric and additively separable into the functions g and h. Then, g, h, the g restricted to $\mathfrak{s}(g)$, and the h restricted to $\mathfrak{s}(h)$ are symmetric.*

PROOF. From Proposition 2,

$$g = \sum_{m \subseteq \mathfrak{s}(g)\,:\,even\ |m|} \hat{f}(m) \cdot \psi_m \tag{3}$$

and

$$h = \sum_{m \subseteq \mathfrak{s}(h)\,:\,even\ |m|} \hat{f}(m) \cdot \psi_m. \tag{4}$$

Since g and h are represented as linear combinations of Walsh functions with even orders, Corollary 2 implies that g and h are symmetric. From Equations (3) and (4), it is clear that $g(x) = g(x \triangle \mathfrak{s}(g))$ and $h(x) = h(x \triangle \mathfrak{s}(h))$ for all $x \in \{0,1\}^n$, which proves that the g restricted to $\mathfrak{s}(g)$ and the h restricted to $\mathfrak{s}(h)$ are symmetric. □

COROLLARY 3. *Suppose that f is symmetric. If there exists $y \in \mathbb{R}$ such that $|\{x \in \{0,1\}^n | f(x) = y\}| = 2 \pmod 4$, f is additively inseparable.*

PROOF. Suppose that f is additively separable into s and $[n] \setminus s$ for a subset s of $[n]$. From Theorem 6, $f(x) = f(x \triangle s) = f(x \triangle ([n] \setminus s)) = f(x \triangle [n])$ for all $x \in \{0,1\}^n$ so that $|\{x \in \{0,1\}^n | f(x) = y\}| = 0 \pmod 4$ for all $y \in \mathbb{R}$. This yields a contradiction. □

Corollary 3 implies that a symmetric function that has an assignment of unique fitness up to bitwise complement is additively inseparable. So, a symmetric function that has unique optimum up to bitwise complement is additively inseparable. Many problems including the two-max problem, the H-IFF problem, and the Ising problems induce such symmetric fitness functions. Corollary 3 further implies that a symmetric function having unique fitness up to bitwise complement cannot be additively separated no matter how we permute the fitness values of assignments as long as the symmetric property is preserved. For example, permuting the fitness values in the two-max problem produces another additively inseparable function.

COROLLARY 4. *Suppose that f is symmetric. If f is additively separable into k disjoint subsets of $[n]$, f has at most 2^{n-k} different values.*

PROOF. Suppose that f is additively separable into the subsets $s_1, \ldots, s_k$ of $[n]$. From Theorem 6, we see that

$$\begin{aligned} f(x) &= f(x \triangle s_1) = f(x \triangle s_2) = \cdots \\ &= f(x \triangle s_1 \triangle s_2) = f(x \triangle s_1 \triangle s_3) = \cdots \\ &= \cdots \\ &= f(x \triangle s_1 \triangle \cdots \triangle s_k). \end{aligned}$$

Since there are at least 2^k assignments of the fitness value $f(x)$ for all $x \in \{0,1\}^n$, f has at most 2^{n-k} different values. □

We say a function f *fully separable* if it is additively separable into n disjoint subsets of $[n]$, i.e., it can be optimized at each bit position independently of the other positions.

COROLLARY 5. *If a symmetric function f is fully separable, f is constant.*

PROOF. Immediate from Corollary 4. □

5. FIRST-ORDER PROJECTION AND RELATED MEASURES

In this section, pseudo-Boolean symmetric functions are considered as in the last section. For the moments related to fitness functions, we assume the uniform distribution in which input strings are sampled uniformly at random.

Epistasis variance is a statistical measure for predicting the difficulty of a given problem with respect to evolutionary algorithms [31]. Given a pseudo-Boolean function $f : \{0,1\}^n \to \mathbb{R}$, the epistasis variance of f is defined as

$$\varepsilon^2(f) = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} (f(x) - \xi(x))^2,$$

where

$$\xi(x) = \mathrm{E}[f] + \sum_{i=1}^{n} (\mathrm{E}[f(y) | y[i] = x[i]] - \mathrm{E}[f]).$$

Let P_f be the projection of f onto the function space spanned by the Walsh functions whose order is at most one. From the orthonormality of Walsh functions,

$$\mathrm{P}_f = \sum_{|m| \le 1} \hat{f}(m) \cdot \psi_m.$$

It was independently discovered in [32], [33], and [34] that ξ coincides with P_f. So, epistasis variance represents the mean square error between a given function and its approximation with Walsh functions of order one:

$$\varepsilon^2(f) = \mathrm{E}[(f - \mathrm{P}_f)^2].$$

To make epistasis variance invariant for constant multiplication, a normalized measure, which is the epistasis variance divided by the variance of the fitness function, was proposed [35] [32]. For a function f, it is defined as

$$\eta(f) = \frac{\varepsilon^2(f)}{\mathrm{Var}[f]}.$$

Since $\varepsilon^2(f) \le \mathrm{Var}[f]$, the normalized epistasis variance satisfies that $0 \le \eta(f) \le 1$. Epistasis correlation [36] is another measure based on first-order projection. It computes the correlation coefficient between f and P_f:

$$\varepsilon\rho(f) = \mathrm{Corr}[f, \mathrm{P}_f].$$

The normalized epistasis variance and the epistasis correlation measure how well a given fitness function can be approximated by its first-order projection. It is considered that if a function has the normalized epistasis variance close to one or the epistasis correlation close to zero, then evolutionary algorithms will have a hard time optimizing the function.

The following describes the relation between Walsh coefficients and moments of a pseudo-Boolean function.

LEMMA 2. *For a pseudo-Boolean function f,*

$$\mathrm{E}[f] = \hat{f}(\emptyset)$$

and

$$\mathrm{Var}[f] = \sum_{m \ne \emptyset} \hat{f}(m)^2.$$

PROOF. The balanced sum theorem (see Theorem 19 in [26]) says that $\mathrm{E}[\psi_m] = 1$ if $m = \emptyset$ and $\mathrm{E}[\psi_m] = 0$ otherwise.

By linearity of expectation,

$$E[f] = \sum_{m \subseteq [n]} E[\hat{f}(m) \cdot \psi_m] = \sum_{m \subseteq [n]} \hat{f}(m) \cdot E[\psi_m] = \hat{f}(\emptyset).$$

Next, define the norm of f as $\|f\| = \langle f, f \rangle^{\frac{1}{2}}$. Then, $E[f^2] = \|f\|^2$ and, by Parseval's identity [37],

$$\|f\|^2 = \sum_{m \subseteq [n]} \hat{f}(m)^2.$$

Since $E[f] = \hat{f}(\emptyset)$,

$$\begin{aligned} \mathrm{Var}[f] &= E[f^2] - E[f]^2 \\ &= \sum_{m \subseteq [n]} \hat{f}(m)^2 - \hat{f}(\emptyset)^2 \\ &= \sum_{m \neq \emptyset} \hat{f}(m)^2. \end{aligned}$$

□

Combining Corollary 2 and Lemma 2, we have

COROLLARY 6. *For a symmetric function f,*

$$P_f = E[f].$$

PROOF.

$$P_f = \sum_{|m| \leq 1} \hat{f}(m) \cdot \psi_m = \hat{f}(\emptyset) = E[f].$$

□

Corollary 6 indicates that the first-order projection of a symmetric function is the constant function with the expected value of the function. It draws the following.

COROLLARY 7. *For a symmetric function f,*

$$\varepsilon^2(f) = \mathrm{Var}[f] = \sum_{|m| \geq 2 \text{ and } |m| \text{ is even}} \hat{f}(m)^2.$$

PROOF.

$$\begin{aligned} \varepsilon^2(f) &= E[(f - P_f)^2] \\ &= E[(f - E[f])^2] \\ &= \mathrm{Var}[f] \\ &= \sum_{|m| \geq 2 \text{ and } |m| \text{ is even}} \hat{f}(m)^2, \end{aligned}$$

where the last equality is derived from Corollary 2 and Lemma 2. □

COROLLARY 8. *For a symmetric function f,*

$$\eta(f) = 1.$$

PROOF. Immediate from the definition of $\eta(f)$ and Corollary 7. □

COROLLARY 9. *For a symmetric function f,*

$$\varepsilon\rho(f) = 0.$$

PROOF. Since P_f is constant from Corollary 6,

$$\begin{aligned} \mathrm{Cov}[f, P_f] &= E[f \cdot P_f] - E[f] \cdot E[P_f] \\ &= E[f] \cdot P_f - E[f] \cdot P_f \\ &= 0. \end{aligned}$$

□

Corollary 8 and 9 imply that symmetric fitness functions constitute a class of the hardest problems in terms of the normalized epistasis variance and the epistasis correlation. This supports the previous empirical results that the search spaces with symmetry induce relatively difficult problems. However, it also suggests that the measures cannot discriminate the degree of problem difficulty for symmetric fitness functions. They are not useful for classifying the symmetric fitness functions in terms of problem difficulty.

6. CONCLUSION

We investigated the properties of symmetric fitness functions that are concerned with encoding schemes, the correlation structures of search spaces, the Walsh analysis of fitness functions, and statistical measures of problem difficulty. It is interesting that the properties of the functions imply the difficulty in maximizing the functions: Maximally nonsynonymous property, zero correlation structure of search spaces, additive inseparability, high epistasis variance and so on. These results partially explain why evolutionary algorithms struggle in the problems with symmetric property.

We saw that two statistical measures of problem difficulty, normalized epistasis variance and epistasis correlation, are invariant for symmetric fitness functions. It is clear that all the symmetric fitness functions are not hard to evolutionary algorithms. (For example, see [2].) Our results raise the limitations of the measures in classifying symmetric fitness functions in terms of problem difficulty. For the classification, another suitable measure must be considered.

We are currently working on investigating the properties of symmetric fitness functions further. This includes the Walsh analysis for symmetric functions over multary alphabets.

Acknowledgment

This work was supported by the Brain Korea 21 Project in 2006. This was also partly supported by grant No. (R01-2003-000-10879-0) from the Basic Research Program of the Korea Science and Engineering Foundation. The ICT at Seoul National University provided research facilities for this study.

7. REFERENCES

[1] L. J. Eshelman and J. D. Schaffer. Crossover's niche. In *Fifth International Conference on Genetic Algorithms*, pages 9–14. Morgan Kaufmann, 1993.

[2] C. Van Hoyweghen, D. E. Goldberg, and B. Naudts. From twomax to the ising model: Easy and hard symmetrical problems. In *Genetic and Evolutionary Computation Conference*, pages 626–633, 2002.

[3] R. A. Watson, G. S. Hornby, and J. B. Pollack. Modeling building block interdependency. In *Parallel Problem Solving from Nature*, pages 97–106. Springer-Verlag, 1998.

[4] C. Van Hoyweghen, B. Naudts, and D. E. Goldberg. Spin-flip symmetry and synchronization. *Evolutionary Computation*, 10(4):317–344, 2002.

[5] E. Falkenauer. *Genetic Algorithms and Grouping Problems.* Wiley, 1998.

[6] S. H. Kim, Y. H. Kim, and B. R. Moon. A hybrid genetic algorithm for the max cut problem. In *Genetic*

and Evolutionary Computation Conference, pages 416–423, 2001.

[7] G. Laszewski. Intelligent structural operators for the k-way graph partitioning problem. In *Fourth International Conference on Genetic Algorithms*, pages 45–52, 1991.

[8] T. N. Bui and B. R. Moon. Genetic algorithm and graph partitioning. *IEEE Trans. on Computers*, 45(7):841–855, 1996.

[9] S. J. Kang and B. R. Moon. A hybrid genetic algorithm for multiway graph paritioning. In *Genetic and Evolutionary Computation Conference*, pages 159–166, 2000.

[10] S. Fischer and I. Wegener. The ising model on the ring: Mutation versus recombination. In *Genetic and Evolutionary Computation Conference*, pages 1113–1124, 2004.

[11] C. Van Hoyweghen and B. Naudts. Symmetry in the search space. In *Congress on Evolutionary Computation*, pages 1072–1079, 2000.

[12] C. Van Hoyweghen, D. E. Goldberg, and B. Naudts. Building block superiority, multimodality and synchronization problems. In *Genetic and Evolutionary Computation Conference*, pages 694–701, 2001.

[13] H. Mühlenbein. Parallel genetic algorithms in combinatorial optimization. In *Computer Science and Operations Research: New Developments in Their Interfaces*, pages 441–453, 1992.

[14] R. Dorne and J. K. Hao. A new genetic local search algorithm for graph coloring. In *Parallel Problem Solving from Nature*, pages 745–754. Springer-Verlag, 1998.

[15] S. S. Choi and B. R. Moon. Isomorphism, normalization, and a genetic algorithm for sorting network optimization. In *Genetic and Evolutionary Computation Conference*, pages 327–334, 2002.

[16] J. H. Kim, S. S. Choi, and B. R. Moon. Neural network normalization for genetic search. In *Genetic and Evolutionary Computation Conference*, pages 398–399, 2004.

[17] M. Dietzfelbinger, B. Naudts, C. Van Hoyweghen, and I. Wegener. The analysis of a recombinative hill-climber on H-IFF. *IEEE Trans. on Evolutionary Computation*, 7(5):417–423, 2003.

[18] F. Rothlauf and D. E. Goldberg. Redundant representations in evolutionary computation. *Evolutionary Computation*, 11(4):381–415, 2003.

[19] E. D. Weinberger. Correlated and uncorrelated fitness landscapes and how to tell the difference. *Biological Cybernetics*, 63:325–336, 1990.

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

[21] T. Jones and S. Forrest. Fitness distance correlation as a measure of problem difficulty for genetic algorithms. In *Sixth International Conference on Genetic Algorithms*, pages 184–192. Morgan Kaufmann, 1995.

[22] K. D. Boese, A. B. Kahng, and S. Muddu. A new adaptive multi-start technique for combinatorial global optimizations. *Operations Research Letters*, 16:101–113, 1994.

[23] P. Merz and B. Freisleben. Fitness landscapes, memetic algorithms and greedy operators for graph bi-partitioning. *Evolutionary Computation*, 8(1):61–91, 2000.

[24] P. Merz and B. Freisleben. Fitness landscape analysis and memetic algorithms for the quadratic assignment problem. *IEEE Trans. on Evolutionary Computation*, 4(4):337–352, 2000.

[25] J. L. Walsh. A closed set of orthogonal functions. *American Journal of Mathematics*, 55:5–24, 1923.

[26] R. B. Heckendorn and D. Whitley. Predicting epistasis directly from mathematical models. *Evolutionary Computation*, 7(1):69–101, 1999.

[27] M. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. Freeman, San Francisco, 1979.

[28] M. X. Goemans and D. P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM*, 42(6):1115–1145, 1995.

[29] C. H. Papadimitriou and M. Yannakakis. Optimization, approximation, and complexity classes. *Journal of Computer and System Sciences*, 43(3):425–440, 1991.

[30] C. H. Papadimitriou. *Computational Complexity*. Addison-Wesley, 1994.

[31] Y. Davidor. Epistasis variance: A viewpoint on ga-hardness. In *Foundations of Genetic Algorithms*, volume 1, pages 23–35. Morgan Kaufmann, 1991.

[32] C. Reeves and C. Wright. An experimental design perspective on genetic algorithms. In *Foundations of Genetic Algorithms*, volume 3, pages 7–22. Morgan Kaufmann, 1995.

[33] S. Rochet, M. Slimane, and G. Venturini. Epistasis for real encoding in genetic algorithms. In *IEEE Australian New Zealand Conference on Intelligent Information Systems*. IEEE Press, 1996.

[34] B. Naudts, D. Suys, and A. Verschoren. Epistasis as a basic concept in formal landscape analysis. In *Seventh International Conference on Genetic Algorithms*, pages 65–72. Morgan Kaufmann, 1997.

[35] M. Manela and J. A. Campbell. Harmonic analysis, epistasis and genetic algorithms. In *Parallel Problem Solving from Nature*, pages 57–64, 1992.

[36] S. Rochet, G. Venturini, M. Slimane, and E. M. El Kharoubi. A critical and empirical study of epistasis measures for predicting ga performances: A summary. In *Artificial Evolution '97*, pages 287–299, 1997.

[37] W. Rudin. *Real and Complex Analysis*. McGraw Hill, 1987.