



OpenAIR@RGU

The Open Access Institutional Repository at Robert Gordon University

<http://openair.rgu.ac.uk>

This is an author produced version of a paper published in

Proceedings of the 2013 UK Workshop on Computational Intelligence
(UKCI), 2013. (ISBN 9781479915668)

This version may not include final proof corrections and does not include
published layout or pagination.

Citation Details

Citation for the version of the work held in 'OpenAIR@RGU':

CHRISTIE, L. A., LONIE, D. P., and McCALL, J. A. W., 2013. Partial structure learning by subset Walsh transform. Available from *OpenAIR@RGU*. [online]. Available from: <http://openair.rgu.ac.uk>

Citation for the publisher's version:

CHRISTIE, L. A., LONIE, D. P., and McCALL, J. A. W., 2013. Partial structure learning by subset Walsh transform. In: *Proceedings of the 2013 UK Workshop on Computational Intelligence (UKCI)*, 2013. IEEE Press, pp. 128-135.

Copyright

Items in 'OpenAIR@RGU', Robert Gordon University Open Access Institutional Repository, are protected by copyright and intellectual property law. If you believe that any material held in 'OpenAIR@RGU' infringes copyright, please contact openair-help@rgu.ac.uk with details. The item will be removed from the repository while the claim is investigated.

(c) 2013 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other users, including reprinting/ republishing this material for advertising or promotional purposes, creating new collective works for resale or redistribution to servers or lists, or reuse of any copyrighted components of this work in other works.

Partial Structure Learning by Subset Walsh Transform

Lee A. Christie
IDEAS Research Institute
Robert Gordon University
Aberdeen, United Kingdom
l.a.christie4@rgu.ac.uk

David P. Lonie
School of Computing Science and Digital Media
Robert Gordon University
Aberdeen, United Kingdom
d.p.lonie@rgu.ac.uk

John A. W. McCall
IDEAS Research Institute
Robert Gordon University
Aberdeen, United Kingdom
j.mccall@rgu.ac.uk

Abstract—Estimation of distribution algorithms (EDAs) use structure learning to build a statistical model of good solutions discovered so far, in an effort to discover better solutions. The non-zero coefficients of the Walsh transform produce a hypergraph representation of structure of a binary fitness function; however, computation of all Walsh coefficients requires exhaustive evaluation of the search space. In this paper, we propose a stochastic method of determining Walsh coefficients for hyperedges contained within the selected subset of the variables (complete local structure). This method also detects parts of hyperedges which cut the boundary of the selected variable set (partial structure), which may be used to incrementally build an approximation of the problem hypergraph.

I. INTRODUCTION

The relationship between random variables in function optimisation algorithms is fundamentally relevant to the performance of an algorithm. Genetic algorithms are said to implicitly learn the structure, or linkage, of a problem by exploration of the search space. However, recombination operators which are blind to structure may cause the GA to converge more slowly or fail to converge on optimal solutions. This is caused when important linkage between variables is disrupted by the recombination. Estimation of distribution algorithms (EDAs) make learning explicit by building a *model*, which may be sampled to produce higher fitness solutions with higher probability than it produces lower fitness solutions.

The representation of problem structure is a vital consideration in structure learning. Early EDAs such as PBIL [1] and UMDA [2] use a univariate model i.e. representing only the marginal probabilities of variables in the solution. The field of EDAs developed to multivariate models, modelling the joint probabilities, such as Bayesian networks [3] and Markov random field models [4].

Closely-related to the concept of a Markov random field model, the Walsh coefficients of a binary fitness function define a hypergraph representation of the structure of a binary function. There is a fast method of computing the Walsh coefficients given known fitness values for every solution in the search space [5] however, this requirement of 2^ℓ function evaluations for an ℓ -bit problem means that any method of optimisation which requires completely determining the structure of a problem can perform no better than exhaustive search.

The problem of computing the Walsh coefficients without

exhaustive evaluation of the search space is impossible to solve without loss of generality. This is because *every* solution in the search space affects the values of the Walsh coefficients. Algorithms have been developed based on the limited probing method of [6] based on the assumption that the solution is additively decomposable and some genetic algorithms have been developed which take account of structure learning in their recombination operators. [7] [8] [9]

In this paper, we introduce a stochastic procedure, the subset Walsh transform. The subset Walsh transform of a chosen subset of variables of k , sampled n times, using $n2^k$ function evaluations will give exact Walsh coefficients for any combinations which are complete parts of the problem structure and reveals indications as to which combinations provide partial structure, which may be completed by including variables not sampled in this instance. Hence, interactions of order $> k$ are indicated by the subset Walsh transform, which may be used to direct a search process towards discovering these larger interactions.

In remainder of this paper: section II describes the Walsh transform; section III gives a description of how the Walsh transform describes problem structure and introduces the benchmark functions we will use; section IV demonstrates the calculation of Walsh coefficients; section V gives a description of the subset Walsh transform; section VI analyses the theory behind the subset Walsh transform with respect to the benchmark functions chosen; section VII shows the result of applying the subset Walsh transform to the identified benchmark functions; and section VIII concludes the paper.

II. WALSH TRANSFORM OF BINARY FUNCTIONS

A problem is said to be *univariate* if there are no interactions between the variables. Given the univariate Walsh function:

$$W_i(x) = \begin{cases} 1 & \text{if } x_i = 1 \\ -1 & \text{if } x_i = 0 \end{cases} \quad (1)$$

Univariate binary functions may always be rewritten in the form:

$$f(x) = \alpha_\emptyset + \alpha_0 W_0(x) + \dots + \alpha_{\ell-1} W_{\ell-1}(x) \quad (2)$$

In this form (the Walsh expansion), it can be shown that the independent *constant term* (α_\emptyset) is equal to the arithmetic mean fitness of all candidates in the search space.

For a multivariate function, the Walsh function on a *clique* (mutually-connected subset of variables) v is defined as the product of the univariate Walsh functions of the elements, and hence, the domain of the multivariate Walsh function is $\{-1, 1\}$. The multivariate Walsh function is defined as:

$$W_K(x) = \prod_{i \in K} W_i(x) \quad (3)$$

Any binary function may be rewritten in the form:

$$f(x) = \sum_{K \subseteq \mathcal{L}} \alpha_K W_K(x) \quad (4)$$

where $\mathcal{L} = \{0, \dots, \ell - 1\}$

The first term in this expansion is $\alpha_\emptyset W_\emptyset(x)$, and as $W_\emptyset(x)$ is equal to 1, this term reduces to α_\emptyset , the arithmetic mean of all candidates in the search space. This *constant term* is not a part of the intrinsic structure of the problem, as it simply specifies an offset which shifts the fitness landscape up the fitness axis, and does not affect the relationship between candidates in the search space. Without this term, the total or mean fitness of all candidates would be zero.

This univariate Walsh transform is the basis for the univariate variant of the Distribution Estimation Using Markov random fields with direct sampling algorithm (DEUM_d [10], which estimates the univariate Walsh coefficients of a population of candidates using least-squares estimation on a system of linear equations constructed from the population. The DEUM family of EDAs has been expanded to include a multivariate form [4] and several approaches have been applied to structure learning for a multivariate DEUM.

III. WALSH EXPANSION AND PROBLEM STRUCTURE

In this section we show the connection between the Walsh expansion of a binary function and its structure. We give 4 examples of binary fitness functions: the Ones function (univariate structure), the 1D-Checkerboard function (bivariate neighbor structure), the OddZeros (simple multivariate structure), and the 4-Trap (highly-interconnected multivariate structure). These four functions will be used again later in the paper in section VII when the Subset Walsh Transform is applied to these examples.

When expressed in the form of a Walsh expansion, the structure of the problem is revealed by the non-zero coefficients. For instance, a univariate problem will contain some or all non-zero coefficients of order ≤ 1 , a bivariate problem will contain some or all non-zero coefficients of order ≤ 2 . The non-zero coefficients may be regarded as edge weights on a hypergraph. This hypergraph is a representation of the problem structure.

A. Ones Function

A classic example of a univariate problem is the *Ones* function; this is simply a sum of all the bits in the bit string. The *Ones* function is given by:

$$f^{ones}(x) = \sum_{i=0}^{\ell-1} x_i \quad (5)$$

There are no interactions between the variables in this problem. The Walsh expansion for the *Ones* function has non-zero Walsh coefficients for only the univariate cliques. The Walsh expansion of the *1D Checkerboard* function is given by:

$$f^{ones}(x) = \frac{\ell}{2} + \frac{1}{2}W_0(x) + \dots + \frac{1}{2}W_{\ell-1}(x) \quad (6)$$

The structure of a univariate problem such as the *Ones* function is represented by assigning a point in the hypergraph to each non-zero Walsh coefficient. The resulting structure is a disconnected set of points shown in figure 1:

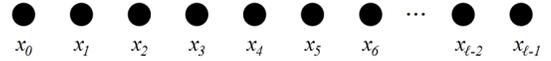


Fig. 1. Structure of the *Ones* function as a hypergraph.

B. 1D-Checkerboard Function

The *1D-Checkerboard* function is a sum of unmatched neighboring pairs. A global optimum is reached if the bit string is an alternating sequence of 0s and 1s. The two optima are the inverses of one another as there is no preference for whether the sequence begins with a 0 or a 1. The *1D-Checkerboard* function is given by:

$$f_1^{check}(x) = \sum_{i=0}^{\ell-2} g(x_i, x_{i+1}) \quad (7)$$

$$\text{where } g(y, z) = \begin{cases} 0 & \text{if } y = z \\ 1 & \text{if } y \neq z \end{cases}$$

As only the comparison of neighboring bits is significant, and there is no preference for the independent value of any bit, only bivariate cliques of neighboring variables have non-zero Walsh coefficients. The Walsh expansion of the *1D Checkerboard* function is given by:

$$f_1^{check}(x) = \frac{\ell-1}{2} - \frac{1}{2}W_{\{0,1\}}(x) - \frac{1}{2}W_{\{1,2\}}(x) - \dots - \frac{1}{2}W_{\{\ell-2, \ell-1\}}(x) \quad (8)$$

The structure of a univariate problem such as the *1D-Checkerboard* function is represented by assigning a line in the hypergraph to each non-zero Walsh coefficient. The resulting structure is a chain of neighboring interactions shown in figure 2 (the univariate terms are *not* present, this is indicated by unfilled circles on the diagram):

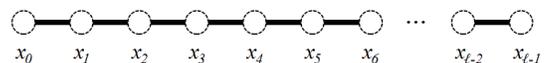


Fig. 2. Structure of the *1D-Checkerboard* function as a hypergraph.

C. Odd-Zeros Function

Before examining the structure of a highly-interconnected problem, we introduce the *Odd-Zeros* problem. For a given instance (a set of cliques), the *Odd-Zeros* problem assigns a score of +1 for each of these cliques which contain an odd number of zeros in the bitstring, and a score of -1 otherwise. We have defined this problem as such so that it is defined naturally by its Walsh expansion. The definition and Walsh expansion of the *Odd-Zero* function is given by:

$$f_I^{\text{oddzeros}}(x) = - \sum_{K \in I} W_K(x) \quad (9)$$

To show the structure of the *Odd-Zeros* problem we must define an instance. The instance we will use is given by:

$$I = \{ \{0, 4, 5\}, \{5, 6, 7\}, \{0, 1, 8, 9, 10\}, \{2, 10, 11\}, \{12, 13, 14, 15, 16\}, \{17, 18\}, \{19\} \} \quad (10)$$

The instance above has been selected to be asymmetric with heterogeneous clique sizes, as this will serve to illustrate the behaviour of the subset Walsh transform on such an irregular structure. The specific Walsh expansion of this instance is:

$$f_I^{\text{oddzeros}}(x) = - W_{\{0,4,5\}}(x) - W_{\{5,6,7\}}(x) - W_{\{0,1,8,9,10\}}(x) - W_{\{2,10,11\}}(x) - W_{\{12,13,14,15,16\}}(x) - W_{\{17,18\}}(x) - W_{19}(x) \quad (11)$$

The structure of this problem only contains one part per clique in the instance I . The size of the clique is the dimensionality of the part. Each n -dimensional clique contributes one $(n - 1)$ -dimensional edge to the problem hypergraph. The structure of this instance is shown in figure 3:

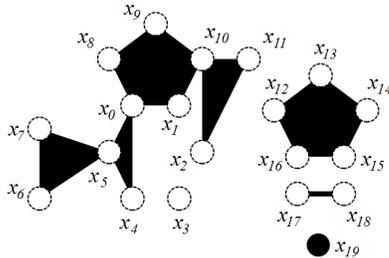


Fig. 3. Structure of an instance of *Odd-Zeros* function as a hypergraph.

D. 4-Trap Function

The *4-Trap* is an instance of the k -trap function which was designed to be *deceptive* [11]. The function has an optimum

at $x = \{1, 1, \dots, 1\}$, however, the gradient of the fitness landscape leads search towards $x = \{0, 0, \dots, 0\}$.

$$f_4^{\text{trap}}(x) = \sum_{j=0}^{(\ell/4)-1} g(x_{4i} + x_{4i+1} + x_{4i+2} + x_{4i+3}) \quad (12)$$

$$\text{where } g(u) = \begin{cases} 3 - u & \text{if } u < 4 \\ 4 & \text{if } u = 4 \end{cases}$$

The Walsh decomposition of the *4-Trap* is:

$$f_4^{\text{trap}}(x) = \sum_{j=0}^{(\ell/4)-1} g(x, 4j) \quad (13)$$

$$\begin{aligned} \text{where } g(x, i) = & \frac{21}{16} - \frac{3}{16} W_{\{i\}}(x) - \frac{3}{16} W_{\{i+1\}}(x) \\ & - \frac{3}{16} W_{\{i+2\}}(x) - \frac{3}{16} W_{\{i+3\}}(x) \\ & + \frac{5}{16} W_{\{i,i+1\}}(x) + \frac{5}{16} W_{\{i,i+2\}}(x) \\ & + \frac{5}{16} W_{\{i+1,i+2\}}(x) + \frac{5}{16} W_{\{i,i+3\}}(x) \\ & + \frac{5}{16} W_{\{i+1,i+3\}}(x) + \frac{5}{16} W_{\{i+2,i+3\}}(x) \\ & + \frac{5}{16} W_{\{i,i+1,i+2\}}(x) + \frac{5}{16} W_{\{i,i+1,i+3\}}(x) \\ & + \frac{5}{16} W_{\{i,i+2,i+3\}}(x) + \frac{5}{16} W_{\{i+1,i+2,i+3\}}(x) \\ & + \frac{5}{16} W_{\{i,i+1,i+2,i+3\}}(x) \end{aligned}$$

The structure of the *4-Trap* function is split up in individuals traps, each represented by four points, six lines, four triangles, and one tetrahedron to the hypergraph representing all of the possible combinations of the variables in this trap. There are no connection between variables in one trap with another. This is shown in figure 4.

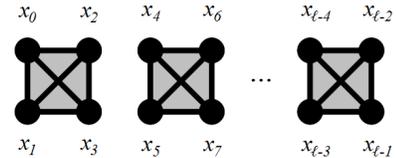


Fig. 4. Structure of *4-Trap* function as a hypergraph.

IV. THE HADAMARD MATRIX AND CALCULATION OF WALSH COEFFICIENTS

On construction of an exhaustive list of fitness evaluations \vec{f} given in Equation 14

$$\vec{f} = \begin{bmatrix} f(1, 1, 1, \dots, 1) \\ f(0, 1, 1, \dots, 1) \\ f(1, 0, 1, \dots, 1) \\ f(0, 0, 1, \dots, 1) \\ \vdots \\ f(0, 0, 0, \dots, 0) \end{bmatrix} \quad (14)$$

it can be shown that the resulting system of linear equations for the complete multivariate Walsh transform (equation 4) of fitness function of length ℓ exhibits a pattern which is best described by the H_ℓ , the Hadamard matrix, which will be used in this section to determine the exact Walsh coefficients given a complete evaluation of a search space.

Sylvester's construction [12] of the Hadamard matrix is a recursive construction. The subscripts we have chosen such that H_ℓ is a 2^ℓ -by- 2^ℓ matrix as H_ℓ is the instance used to perform the Walsh expansion on a length ℓ function.

$$H_0 = [1] \quad (15)$$

$$H_1 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad (16)$$

$$H_\ell = \begin{bmatrix} H_{\ell-1} & H_{\ell-1} \\ H_{\ell-1} & -H_{\ell-1} \end{bmatrix} = H_1 \otimes H_{\ell-1} \quad (17)$$

For Walsh coefficients in the order

$$\vec{\alpha} = \begin{bmatrix} \alpha_{\emptyset} \\ \alpha_{\{0\}} \\ \alpha_{\{1\}} \\ \alpha_{\{0,1\}} \\ \alpha_{\{2\}} \\ \alpha_{\{0,2\}} \\ \alpha_{\{1,2\}} \\ \alpha_{\{0,1,2\}} \\ \vdots \\ \alpha_{\{0,\dots,\ell-1\}} \end{bmatrix} \quad (18)$$

The matrix form of the system of linear equations given by the Walsh transform is

$$H_\ell \vec{\alpha} = \vec{f} \quad (19)$$

As the Hadamard matrix is symmetric, $H_\ell^\top = H_\ell$, and a general property for the Hadamard matrix is that $H_\ell H_\ell^\top = 2^\ell I_{2^\ell}$ where I_{2^ℓ} is the identity matrix of size $2^\ell \times 2^\ell$ hence, the inverse $H_\ell^{-1} = \frac{1}{2^\ell} H_\ell$

The calculation of the Walsh coefficients may therefore be rewritten to be solved by simple matrix multiplication:

$$\vec{\alpha} = \frac{1}{2^\ell} H_\ell \vec{f} \quad (20)$$

The above method is the one used in this paper, there also exists other methods of computing the Walsh coefficients, such as the *fast Walsh transform* [5] which are faster, however, the same number of fitness evaluations must be made, which is considered to be the limiting factor, rather than the computational complexity of the coefficient calculation.

V. SUBSET WALSH TRANSFORM

Obtaining the complete Walsh transform of a fitness function requires evaluation of every valid candidate and is therefore of complexity $O(2^\ell)$. In this section we introduce a method of using the Walsh transform to obtain the alpha parameters involving a selected subset of parameters k . We shall also use k in some contexts to represent the size of the set k . The complexity of the subset Walsh transform is $O(n2^k)$. We call this method *subset Walsh transform*.

The subset Walsh transform performs n samplings of the vector $\vec{\alpha}$ (the Walsh coefficients).

In the complete Walsh transform, all 2^ℓ possible ℓ -bit strings are evaluated by the fitness function. In one sampling

of the subset Walsh transform we use all 2^k possible k -bit strings (in the order defined by the rows of B in Eqn. 21) to populate the elements of the ℓ -bit string specified by the indices in the set k . The remaining $(\ell - k)$ bits are taken from a randomly generated $(\ell - k)$ -bit string, which is reused for all 2^k evaluations in this sample.

$$B = \begin{matrix} & \overbrace{\hspace{2cm}}^{2^k\text{-by-}k} \\ \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 0 & 1 & 1 & \dots & 1 \\ 1 & 0 & 1 & \dots & 1 \\ 0 & 0 & 1 & \dots & 1 \\ & & & \vdots & \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix} \end{matrix} \quad (21)$$

The sampling process may be repeated and once n samplings of the Walsh coefficients have been produced, the mean (μ_α) and standard deviation (σ_α) of each Walsh coefficient is determined. The order of the elements in the resulting column matrices μ_α and σ_α are interpreted in the order described by Eqn. 22

$$\text{for } k = \{k_0, k_1, \dots, k_{k-1}\} \quad (22)$$

$$\vec{\alpha} = \begin{bmatrix} \alpha_{\emptyset} \\ \alpha_{\{k_0\}} \\ \alpha_{\{k_1\}} \\ \alpha_{\{k_0, k_1\}} \\ \alpha_{\{k_2\}} \\ \vdots \\ \alpha_k \end{bmatrix}$$

The complete subset Walsh transform pseudocode is given below in Algorithm 1:

- 1) **define** A as 2^k -by- n matrix
- 2) **define** μ as length 2^k column
- 3) **define** σ as length 2^k column
- 4) **for** $j \leftarrow 0$ **to** $n - 1$
 - a) **define** $r \leftarrow$ row of $(\ell - k)$ random bits
 - b) **define** f as length 2^k column
 - c) **for** $i \leftarrow 0$ **to** $2^k - 1$
 - i) **define** $s \leftarrow$ row i of B
 - ii) **define** x as length ℓ bitstring
 - iii) **for** indices in k , populate elements of x with elements of s left-to-right
 - iv) **for** indices not in k , populate elements of x with elements of r left-to-right
 - v) $f_i \leftarrow$ evaluate (x)
 - d) column $A_j \leftarrow \frac{1}{2^k} H_k f$
- 5) **for** $i \leftarrow 0$ **to** $2^k - 1$
 - a) $\mu_i \leftarrow$ mean of row A_i
 - b) $\sigma_i \leftarrow$ stdev of row A_i
- 6) **return** columns μ and σ

Algorithm 1. The subset Walsh transform where ℓ = the problem length, k = the selected subset of variables or the size of this set, and H_k = the Hadamard matrix of size 2^k -by- 2^k . B is given in Eqn. 21.

VI. THEORETICAL ANALYSIS OF THE SUBSET WALSH TRANSFORM

In this section we shall consider the theoretical basis of the Subset Walsh Transform described in this paper. First consider a partition of the set of ℓ random variables $X = \{X_0, X_1, \dots, X_{\ell-1}\}$ into $S = \{X_0, X_1, \dots, X_{k-1}\}$ and $R = \{X_k, X_{k+1}, \dots, X_{\ell-1}\}$. Samples of these sets of variables are represented by x , x_s , and x_r respectively. Note that the consecutive ordering of the random variables is purely for convenience and would not restrict the argument. On X let $f(x)$ be the function we wish to optimise.

The central idea is that the 2^k variables in S can be exhaustively sampled using the Hadamard matrix method described in section IV, whilst variables from R are randomly sampled. And evaluations of the function f at points sampled in this way allows estimates to be made of the Walsh coefficients.

Experiments involving the subset Walsh transform procedure show, at least for selected problems, a remarkably clear-cut set of parameter estimates. To see why this should be, consider the set of functions which can be decomposed into a sum as follows:

$$f(\vec{x}) = f_S(x_0, x_1, \dots, x_{k-1}) + f_P(x_0, x_1, \dots, x_{\ell-1}) + f_R(x_k, x_{k+1}, \dots, x_{\ell-1}) \quad (23)$$

or more concisely as

$$f(x) = f_S(x_s) + f_P(x_s, x_r) + f_R(x_r) \quad (24)$$

where

- f_S depends only on the variables $(x_0, x_1, \dots, x_{k-1})$ within the subspace whose structure components are being investigated, with \vec{f}_S representing the column vector as in Eqn. 14 of all fitnesses based on the selected part of the fitness function alone;
- f_R depends only on the variables $(x_k, x_{k+1}, \dots, x_{\ell-1})$ within the complementary subspace to that whose structure components are being investigated, and which, crucially, within one sampling will be the same constant value for all 2^k evaluations, with \vec{f}_R as the column vector of fitnesses; and
- f_P represents the part of f involving partial interaction between the subspace S and the rest of the problem space, and which, within one sampling will be a random value with respect to variables $(x_0, x_1, \dots, x_{k-1})$, with \vec{f}_P as the column vector of fitnesses

Within one sampling we may therefore write:

$$f(x) = f_S(x_s) + f_P(x_s, x_r) + c \quad (25)$$

where $c = f_R(r)$ and the estimate of the structure may be decomposed as

$$\vec{\alpha} = \frac{1}{2^k} (H_k \vec{f}_S + H_k \vec{f}_P + H_k \vec{c}) = \vec{\alpha}_S + \vec{\alpha}_P + \vec{\alpha}_R \quad (26)$$

The Hadamard matrix has the property that all rows (except the first row) have equal numbers of 1s and -1 s, and so

$$\vec{\alpha}_R = \frac{1}{2^k} H_k \begin{bmatrix} c \\ c \\ \vdots \\ c \end{bmatrix} = \begin{bmatrix} \alpha_\emptyset \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (27)$$

meaning that estimates of the sub-structure α parameters (except α_\emptyset) are completely unaffected by any term in the optimised function that depends only on variables outwith the selected variables.

Further, since $\vec{\alpha}_S = \frac{1}{2^k} H_k \vec{f}_S$ is independent of x_r its contribution to $\vec{\alpha}$ is the same for all iterations. So in cases where the sub-structure under consideration, on S , fully contains a self-contained part of the entire models structure, the α parameter estimates will be exact (i.e. have no standard deviation).

The influence of $\vec{\alpha}_P = \frac{1}{2^k} H_k \vec{f}_P$ is that where the structure within selected variables S , contains some but not all of the variables forming part of the full model, the estimates of the α parameters will be expected to have non-zero standard deviation because of the influence of the random sampling of x_R .

In conclusion, provided the function can be expressed in a form of Eqn. 24, the estimation procedure we describe, will assign $\vec{\alpha}$ parameters to interactions that

- 1) have non-zero mean with zero standard deviation for Walsh coefficients which are parts of the structure completely contained within the selected variables S ;
- 2) have zero mean and zero standard deviation for Walsh coefficients which are not part of the structure; and
- 3) have non-zero standard deviation for Walsh coefficients involving parts of the structure which are contained partly in S and partly in R , with the mean being an estimation of the particular Walsh coefficient being considered.

The resulting μ , σ pairs are interpreted by Table I. Zero standard deviation indicates that the mean value is likely are correctly determined part of the structure (or non-part if zero). Where there is a non-zero standard deviation, partial structure has been detected. The indicates that a part of the structure involving the variables in these variables, plus additional variables not in k , but no additional variables in k exists. Examples of partial structure are given in section VII where the result of applying the subset Walsh transform is shown.

μ	σ	In Structure
non-zero	0	Yes
0	0	No
any	non-zero	Partial

TABLE I. STRUCTURE DETECTION

VII. RESULT OF APPLYING THE SUBSET WALSH TRANSFORM ON SELECTED VARIABLES

To illustrate the principle of the subset Walsh transform, in this section we show the result of applying the algorithm to a benchmark problem with known structure and remark on the meaning of the results. In each example the first 6 variables have been selected. Length 20 problems with a sample size of 5 have been chosen. In the final example, the sample size has been increased to 20 due to the high probability of an incorrect result for a smaller sample size. This is discussed further in the *4-Trap* subsection.

A. Ones Function

Fig 5 shows the structure of the *Ones* function with the first 6 variables selected. This selection includes 6 univariate parts of the structure. All 6 parts of the structure are completely contained within the selected variables S .

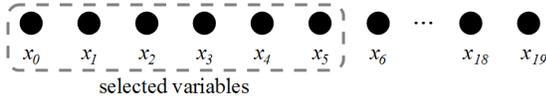


Fig. 5. Illustrating the selected variables $k = \{0, 1, 2, 3, 4, 5\}$ on the hypergraph of *Ones* function for $\ell = 20$.

Applying the subset Walsh transform method with the selected subset, the decomposition of the fitness function as in Eqn. 24 takes the form:

$$\begin{aligned} f_S(x) &= x_0 + \dots + x_5 & (28) \\ f_P(x) &= 0 \\ f_R(x) &= x_6 + \dots + x_{19} \end{aligned}$$

In this case, the fully sampled subset $\{0, 1, 2, 3, 4, 5\}$ has no interaction with the randomly sampled subset $\{6, 7, \dots, 19\}$ and $f_P(x) = 0$. Hence the contributions to the structure estimates derived from the subset Walsh transform method will be a constant $\vec{\alpha}_S, \vec{\alpha}_P = 0$ and $\vec{\alpha}_R = 0$, resulting in constant α . Hence the structure parameter estimates for the selected subset are expected to have no standard deviation and should agree with the theoretical expectations.

We ran the subset Walsh transform on this example, the result is shown in Fig 6. As all structure is inside or outside of the selection with no overlap, there is no variance on the result, hence we expect there the standard deviations vector $\vec{\sigma}$ to be zero. Repeated sampling in this case is not necessary, however, 5 samples were used to show that the result is consistent.

B. 1D-Checkerboard Function

Fig 7 shows the structure of the *1D-Checkerboard* function with the first 6 variables selected. This selection includes 5 complete parts of the structure and one of two variables involved in the clique $\{5, 6\}$. The variable x_5 has been highlighted as we expect there to be a non-zero variance in the value of this Walsh coefficient.

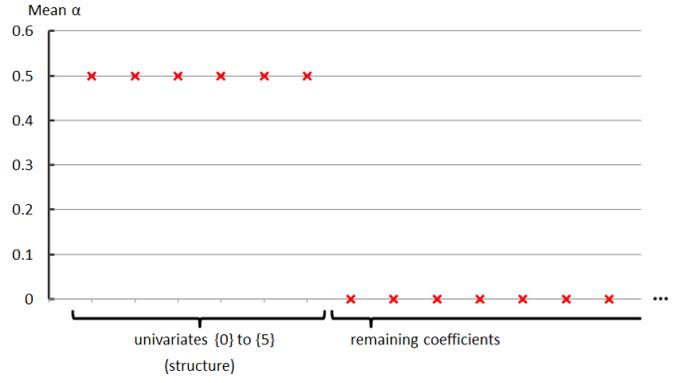


Fig. 6. Estimated Walsh coefficients in a subset Walsh transform of selected variables $k = \{0, 1, 2, 3, 4, 5\}$ of the *Ones* function for $\ell = 20$ using $n = 5$ samples. (a_0 constant term omitted).

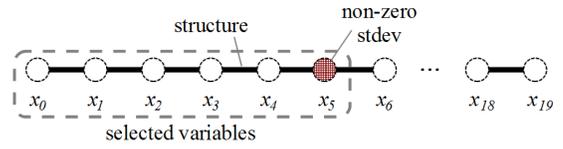


Fig. 7. Illustrating the selected variables $k = \{0, 1, 2, 3, 4, 5\}$ on the hypergraph of *1D-Checkerboard* function for $\ell = 20$.

Applying the subset Walsh transform method with the selected subset, the decomposition of the fitness function as in Eqn. 24 takes the form:

$$\begin{aligned} f_S(x) &= g(x_0, x_1) + g(x_1, x_2) + \dots + g(x_4, x_5) & (29) \\ f_P(x) &= g(x_5, x_6) \\ f_R(x) &= g(x_6, x_7) + g(x_7, x_8) + \dots + g(x_{18}, x_{19}) \end{aligned}$$

$$\text{where } g(y, z) = \begin{cases} 0 & \text{if } y = z \\ 1 & \text{if } y \neq z \end{cases}$$

In this case, the fully sampled subset $\{0, 1, 2, 3, 4, 5\}$ has interaction with the randomly sampled subset $\{6, 7, \dots, 19\}$ through $f_P = g(x_5, x_6)$. Hence the contributions to the structure estimates derived from the subset Walsh transform method will be constant $\vec{\alpha}_S, \vec{\alpha}_R = 0$ in addition to a random $\vec{\alpha}_P$ that takes one of two values depending on the value of x_5 . It can easily be checked that the non-zero parameter estimates are expected to be constant values $\alpha_{\{0,1\}} = \alpha_{\{1,2\}} = \alpha_{\{2,3\}} = \alpha_{\{3,4\}} = \alpha_{\{4,5\}} = -\frac{1}{2}$ and $\alpha_{\{5\}} = -\frac{1}{2}$ or $\frac{1}{2}$ each with probability $p = 0.5$. Hence the structure parameter estimates for the selected subset are expected to be exact except for $\alpha_{\{5\}}$ which should, for large enough samples, tend to zero and have standard deviation 0.5.

We ran the subset Walsh transform on this example, the result is shown in Fig 8. All complete structure within the selection has been detected with zero σ as before, however, as there is now a partial part of the structure, the term $\alpha_{\{5\}}$ has a non-zero standard deviation, indicating that there x_5 interacts with part of the structure outwith the selection.

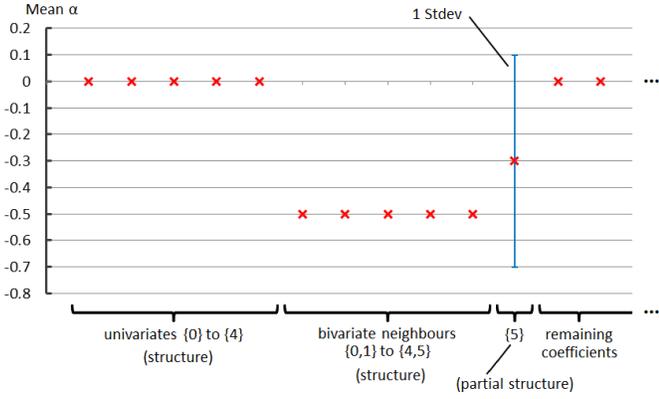


Fig. 8. Estimated Walsh coefficients in a subset Walsh transform of selected variables $k = \{0, 1, 2, 3, 4, 5\}$ of the *1D-Checkerboard* function for $\ell = 20$ using $n = 5$ samples. (a_0 constant term omitted).

C. Odd-Zeros Function

Fig 9 shows the structure of an instance of the *Odd-Zeros* function with the first 6 variables selected. This selection includes only one complete part of the structure, but includes single variables from cliques $\{5, 6, 7\}$ and $\{2, 10, 11\}$, and two variables from clique $\{0, 1, 8, 9, 10\}$. In Fig 9, this partial structure has been highlighted. Note that the univariate terms $\{0\}$ and $\{1\}$ have not been individually highlighted although they are part of the partial structure, only $\{0, 1\}$ contains the most variables possible for the clique $\{0, 1, 8, 9, 10\}$

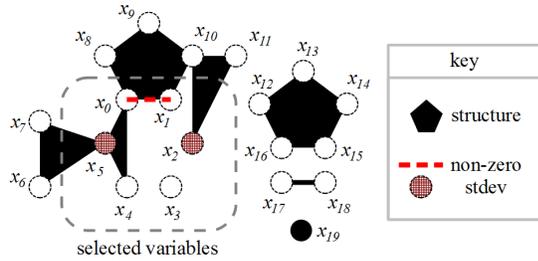


Fig. 9. Illustrating the selected variables $k = \{0, 1, 2, 3, 4, 5\}$ on the hypergraph of an instance of the *Odd-Zeros* function.

Applying the subset Walsh transform method with the selected subset, the decomposition of the fitness function as in Eqn. 24 takes the form:

$$\begin{aligned}
 f_S(x) &= -W_{\{0,4,5\}}(x) \\
 f_P(x) &= -W_{\{5,6,7\}}(x) \\
 &\quad -W_{\{0,1,8,9,10\}}(x) \\
 &\quad -W_{\{2,10,11\}}(x) \\
 f_R(x) &= -W_{\{12,13,14,15,16\}}(x) \\
 &\quad -W_{\{17,18\}}(x) \\
 &\quad -W_{19}(x)
 \end{aligned} \tag{30}$$

In this case, the fully sampled subset $\{0, 1, 2, 3, 4, 5\}$ has multiple interaction with the randomly sampled subset $\{6, 7, \dots, 19\}$ through f_P . Hence the contributions to the

structure estimates derived from the subset Walsh transform method will be constant $\bar{\alpha}_S$, and $\bar{\alpha}_R = 0$ in addition to a random $\bar{\alpha}_P$ that take many ($2^5 = 32$) values depending on the value of x_7, x_8, x_9, x_{10} and x_{11} . Hence the structure parameter estimates involving the selected subset are expected to be exact for any values involving the subset $\{0, 4, 5\}$ while those involving the subsets $\{0, 1\}$, $\{2\}$ and $\{5\}$ should manifest themselves through partial interaction and hence have non-zero standard deviation.

We ran the subset Walsh transform on this example, the result is shown in Fig 10. The non-zero σ is of the expected cliques $\{2\}$, $\{5\}$, and $\{0, 1\}$ and *not* $\{0\}$ or $\{1\}$, suggesting that x_0 and x_1 only appear when in a clique together.

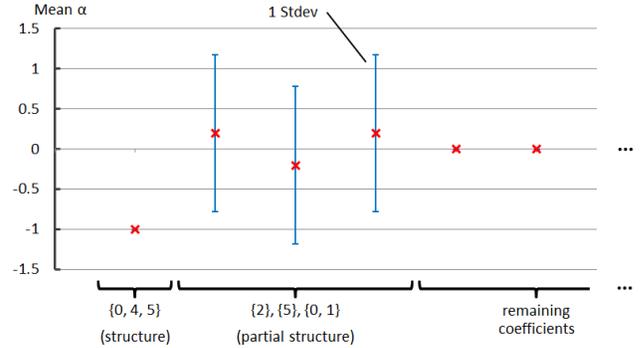


Fig. 10. Estimated Walsh coefficients in a subset Walsh transform of selected variables $k = \{0, 1, 2, 3, 4, 5\}$ of an instance of the the *Odd-Zeros* function using $n = 5$ samples. (a_0 constant term omitted)

D. 4-Trap Function

The case of the *4-Trap* function is more complex than the previous example because the cliques $\{4\}$, $\{5\}$, and $\{4, 5\}$ are complete parts of the structure, however, that are also partial structure in several other parts of structure. The clique $\{4\}$ should indicate partial structure in the cliques $\{4, 6\}$, $\{4, 7\}$, and $\{4, 6, 7\}$. The clique $\{5\}$ should indicate partial structure in the cliques $\{5, 6\}$, $\{5, 7\}$, and $\{5, 6, 7\}$. The clique $\{4, 5\}$ should indicate partial structure in the cliques $\{4, 5, 6\}$, $\{4, 5, 7\}$, and $\{4, 5, 6, 7\}$. The structure of the *4-Trap* with the first 6 variables selected is shown in Fig. 11.

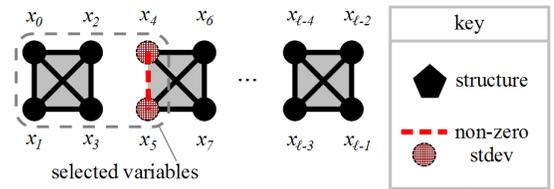


Fig. 11. Illustrating the selected variables $k = \{0, 1, 2, 3, 4, 5\}$ on the hypergraph of an instance of the *4-Trap* function.

Applying the subset Walsh transform method with the selected subset, the decomposition of the fitness function as in Eqn. 24 takes the form:

VIII. CONCLUSIONS AND FURTHER WORK

This paper has given an account of the connection between the Walsh expansion of a binary fitness function and the structure of the problem in terms of variable interactions, and presented the subset Walsh transform — a novel approach which uses partitioning of the set of variables and statistical sampling to calculate Walsh coefficients.

We have argued theoretically that the subset Walsh transform reliably produces the exact Walsh coefficient for parts of the structure which are a subset or equal to the selected variables, and for partial structure provides an estimation of the correct Walsh coefficient, with a statistical variance. Further work will explore the use of the subset Walsh transform in EDA-type search algorithms.

REFERENCES

- [1] S. Baluja and R. Caruana, "Removing the genetics from the standard genetic algorithm," in *Machine Learning International Workshop then Conference*, 1995, pp. 38–46.
- [2] H. Mühlenbein and G. Paaß, "From recombination of genes to the estimation of distributions I. Binary parameters," in *Proceedings of the 4th International Conference on Parallel Problem Solving from Nature, Lecture Notes In Computer Science Vol 1141*, 1996, pp. 178–187.
- [3] M. Pelikan, D. E. Goldberg, and E. Cantú-Paz, "The Bayesian optimization Algorithm," in *Proceedings of the Conference on Genetic and Evolutionary Computation (GECCO 1999)*, 1999, pp. 525–532.
- [4] S. Shakya, A. E. I. Brownlee, J. A. W. McCall, F. Fournier, and G. Owusu, "A fully Multivariate DEUM Algorithm," in *Evolutionary Computation, 2009. (CEC '09). IEEE Congress on*, 2009, pp. 479–486.
- [5] D. E. Goldberg, "Genetic Algorithms and Walsh Functions: Part I, Gentle Introduction," *Complex Systems*, vol. 3, no. 2, pp. 129–152, 1989.
- [6] R. B. Heckendorn and A. H. Wright, "Efficient Linkage Discovery by Limited Probing," *Evolutionary Computation*, vol. 12, no. 4, pp. 517–545, 2004.
- [7] M. J. Streeter, "Upper Bounds on the Time and Space Complexity of Optimizing Additively Separable Functions," in *Proceedings of the 6th Annual Conference on Genetic and Evolutionary Computation (GECCO 2004)*. Berlin: Springer, 2004, pp. 186–197. [Online]. Available: <http://scholar.google.co.uk/scholar?cluster=5991917444417855960>
- [8] S. S. Choi, K. Jung, and J. H. Kim, "Almost Tight Upper Bound for Finding Fourier Coefficients of Bounded Pseudo-Boolean Functions," *Journal of Computer and System Sciences*, vol. 77, no. 6, pp. 1039–1053, 2011.
- [9] A. H. Wright and S. Pulavarty, "On the Convergence of an Estimation of Distribution Algorithm Based on Linkage Discovery and Factorization," in *Proceedings of the 7th Annual Conference on Genetic and Evolutionary Computation (GECCO 2005)*, 2005, pp. 695–702.
- [10] S. Shakya, J. A. W. McCall, and D. F. Brown, "Estimating the distribution in an EDA," in *Adaptive and Natural Computing Algorithms*, 2005, pp. 202–205.
- [11] K. Deb and D. E. Goldberg, "Analyzing deception in trap functions," in *FOGA2*. Palo Alto, CA: Morgan Kaufmann Publishers, Inc., 1992.
- [12] J. J. Sylvester, "Thoughts on inverse orthogonal matrices, simultaneous sign successions, and tessellated pavements in two or more colours, with applications to Newton's rule, ornamental tile-work, and the theory of numbers." *Philosophical Magazine*, vol. 34, pp. 461–475, 1867.

$$\begin{aligned}
 f_S(x) &= g(x_0 + x_1 + x_2 + x_3) \\
 f_P(x) &= g(x_4 + x_5 + x_6 + x_7) \\
 f_R(x) &= g(x_8 + x_9 + x_{10} + x_{11}) \\
 &\quad + g(x_{12} + x_{13} + x_{14} + x_{15}) \\
 &\quad + g(x_{16} + x_{17} + x_{18} + x_{19})
 \end{aligned} \tag{31}$$

where $g(u) = \begin{cases} 3 - u & \text{if } u < 4 \\ 4 & \text{if } u = 4 \end{cases}$

For each random sample x_R the vector \vec{f} has contributions from the f_P that take one of four forms (depending on the $2 \times 2 = 4$ possible combinations of the x_6 and x_7 values).

Provided the number of samples is high enough, we would expect the average structure parameter estimates associated with the sub-clique $\{4, 5\}$ to approach the true values, and for the standard deviation of those estimates to be non-zero (indicating the presence of a sub-clique).

As the structure parameter estimates associated with clique $\{0, 1, 2, 3\}$ do not depend on the random sample, it is tempting to suppose that very small sample sizes might be sufficient to expose the sub-structure, and indeed this is true for cliques entirely contained in the subset transform. Further, even with small sample sizes, the presence of such sub-cliques is usually shown via the presence of non-zero parameter estimates with non-zero standard deviation, although the values of the estimates are not exact.

However, for very small samples there is the possibility of misleading results. In the 4-trap example this has been experimentally observed as anomalous estimates of $\alpha_{\{4\}}$, $\alpha_{\{5\}}$ and $\alpha_{\{4,5\}}$ with zero standard deviation, arising when the sample happens to generate only x_6 and x_7 values that give rise to the same f_P values. To guard against this the sample size should be large enough to make the probability, of generating identical f_P values, acceptably small.

We ran the subset Walsh transform on the 4-Trap, the result is shown in Fig 12. The non-zero σ is of the expected cliques $\{4\}$, $\{5\}$, and $\{4, 5\}$. A sample size of $n = 20$ has been used in this case.

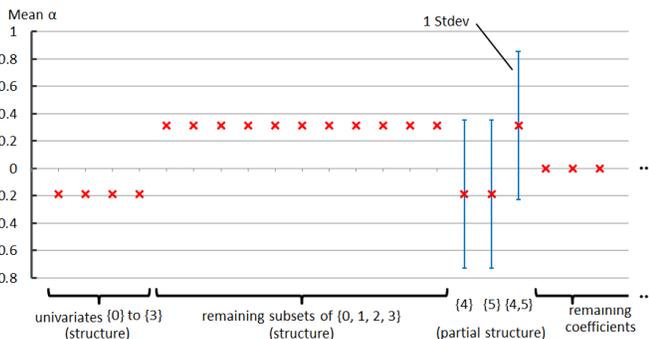


Fig. 12. Estimated Walsh coefficients in a subset Walsh transform of selected variables $k = \{0, 1, 2, 3, 4, 5\}$ of the 4-Trap function for $\ell = 20$ using $n = 20$ samples. (a_\emptyset constant term omitted).