

# Exact Computation of the Fitness-Distance Correlation for Pseudoboolean Functions with One Global Optimum

Francisco Chicano and Enrique Alba\*

University of Málaga, Spain,  
{chicano,alba}@lcc.uma.es

**Abstract.** Landscape theory provides a formal framework in which combinatorial optimization problems can be theoretically characterized as a sum of a special kind of landscapes called elementary landscapes. The decomposition of the objective function of a problem into its elementary components can be exploited to compute summary statistics. We present closed-form expressions for the fitness-distance correlation (FDC) based on the elementary landscape decomposition of the problems defined over binary strings in which the objective function has one global optimum. We present some theoretical results that raise some doubts on using FDC as a measure of problem difficulty.

**Keywords:** Landscape theory, fitness landscapes, fitness-distance correlation

## 1 Introduction

The theory of landscapes focuses on the analysis of the structure of the search space that is induced by the combined influences of the objective function of the optimization problem and the neighborhood operator [16]. In the field of combinatorial optimization, this theory has been previously used to characterize optimization problems [8], improve search algorithms [12], and obtain global statistics of the problems [20].

A *landscape* for a combinatorial optimization problem (COP) is a triple  $(X, N, f)$ , where  $X$  is the set of *tentative solutions* of the COP,  $f : X \mapsto \mathbb{R}$  defines the objective or *fitness function* and  $N$  is the *neighborhood operator*. There exists a special kind of landscapes, called *elementary landscapes* (EL), which are of particular interest due to their properties [22]. We define and analyze the elementary landscapes in Section 2, but we can advance that they are characterized by the *Grover's wave equation*:

$$\text{avg}\{f(y)\}_{y \in N(x)} = \frac{1}{d} \sum_{y \in N(x)} f(y) = f(x) + \frac{\lambda}{d} (\bar{f} - f(x)), \quad (1)$$

---

\* Research partially funded by the Spanish Ministry of Science and Innovation and FEDER under contracts TIN2008-06491-C04-01 and TIN2011-28194 and the Andalusian Government under contract P07-TIC-03044.

where  $d$  is the size of the neighborhood,  $|N(x)|$ , which we assume is the same for all the solutions in the search space,  $\bar{f}$  is the average solution evaluation over the entire search space and  $\lambda$  is a characteristic constant. For a given problem instance whose objective function is elementary, the values  $\bar{f}$  and  $\lambda$  can be easily computed in an efficient way, usually from the problem data. Thus, the wave equation makes it possible to compute the average value of the fitness function  $f$  evaluated over all of the neighbors of  $x$  using only the value  $f(x)$ , without actually evaluating any of the neighbors.

When the landscape is not elementary it is always possible to write the objective function as a sum of elementary components, called *elementary landscape decomposition* (ELD) of a problem [6]. In the case of binary strings with length  $n$  under the one-change neighborhood, the number of elementary components is at most  $n$ . Then, Grover’s wave equation can be applied to each elementary component and all the results are summed to give the average fitness in the neighborhood of a solution. Furthermore, for some problems the average cannot be limited to the neighborhood of a solution, but it can be extended to the second-order neighbors (neighbors of neighbors), third-order neighbors, and, in general, to any arbitrary region around a given solution, including the whole search space. Sutton *et al.* [19] show how to compute the averages over *spheres* and *balls* of arbitrary radius around a given solution in polynomial time using the elementary landscape decomposition of pseudoboolean functions.

Landscape theory has been proven to be quite effective computing summary statistics of the optimization problem. Measures like the autocorrelation length and the autocorrelation coefficient can be efficiently computed using the ELD of a problem [8]. Recently, Chicano and Alba [5] and Sutton and Whitley [18] have shown how the expected value of the fitness of a mutated individual can be exactly computed using the ELD. In short, landscape theory can be applied to any COP and thus is generally beneficial for the whole community in discrete optimization, representing a general and usable formalism in practice.

The main contribution of the present work is an exact expression for the Fitness-Distance Correlation (FDC) of COPs defined over a set of binary strings (pseudoboolean functions) having one global optimum. This expression is based on the ELD of the problem. We also analyze the expression in order to discuss the usefulness of the FDC as a difficulty measure for a problem.

The remainder of the paper is organized as follows. In Section 2 we present the mathematical tools required to understand the rest of the paper. Section 3 presents the exact expression for the FDC and other theoretical results, while Section 4 validates FDC in practice with this theoretical background. Finally, we summarize our findings and future work in Section 5.

## 2 Background

In this section we present some fundamental results of landscape theory. We will only focus on the relevant information required to understand the rest of the paper. The interested reader can deepen on this topic in [15].

Let  $(X, N, f)$  be a landscape, where  $X$  is a finite set of candidate solutions,  $f : X \rightarrow \mathbb{R}$  is a real-valued function defined on  $X$  and  $N : X \rightarrow 2^X$  is the

neighborhood operator. The pair  $(X, N)$  is called *configuration space* and induces a graph in which  $X$  is the set of nodes and an arc between  $(x, y)$  exists if  $y \in N(x)$ . The adjacency and degree matrices of the neighborhood  $N$  are:

$$A_{xy} = \begin{cases} 1 & \text{if } y \in N(x), \\ 0 & \text{otherwise;} \end{cases} \quad D_{xy} = \begin{cases} |N(x)| & \text{if } x = y, \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

We restrict our attention to regular neighborhoods, where  $|N(x)| = d > 0$  for a constant  $d$ , for all  $x \in X$ . Then, the degree matrix is  $D = dI$ , where  $I$  is the identity matrix. The Laplacian matrix  $\Delta$  associated to the neighborhood is defined by  $\Delta = A - D$ . In the case of regular neighborhoods it is  $\Delta = A - dI$ . Any discrete function,  $f$ , defined over the set of candidate solutions can be characterized as a vector in  $\mathbb{R}^{|X|}$ . Any  $|X| \times |X|$  matrix can be interpreted as a linear map that acts on vectors in  $\mathbb{R}^{|X|}$ . For example, the adjacency matrix  $A$  acts on function  $f$  as follows

$$A f = \begin{pmatrix} \sum_{y \in N(x_1)} f(y) \\ \sum_{y \in N(x_2)} f(y) \\ \vdots \\ \sum_{y \in N(x_{|X|})} f(y) \end{pmatrix}, \quad (A f)(x) = \sum_{y \in N(x)} f(y). \quad (3)$$

Thus, the component  $x$  of  $(A f)$  is the sum of the function values of all the neighbors of  $x$ . Stadler defines the class of *elementary landscapes* where the function  $f$  is an eigenvector (or eigenfunction) of the Laplacian up to an additive constant [16].

**Definition 1.** *Let  $(X, N, f)$  be a landscape and  $\Delta$  the Laplacian matrix of the configuration space. The landscape is said to be elementary if there exists a constant  $b$ , which we call offset, and an eigenvalue  $\lambda$  of  $-\Delta$  such that  $(-\Delta)(f - b) = \lambda(f - b)$ . When the neighborhood is clear from the context we also say that  $f$  is elementary.*

We use  $-\Delta$  instead of  $\Delta$  in the definition to avoid negative eigenvalues, since  $\Delta$  is negative semidefinite. In connected neighborhoods, where the graph related to the configuration space  $(X, N)$  is connected, the offset  $b$  is the average value of the function over the whole search space:  $b = \bar{f}$ . Taking into account basic results of linear algebra, it can be proved that if  $f$  is elementary with eigenvalue  $\lambda$ ,  $af + b$  is also elementary with the same eigenvalue  $\lambda$ . Furthermore, in regular neighborhoods, if  $g$  is an eigenfunction of  $-\Delta$  with eigenvalue  $\lambda$  then  $g$  is also an eigenfunction of  $A$ , the adjacency matrix, with eigenvalue  $d - \lambda$ . The average value of the fitness function in the neighborhood of a solution can be computed using the expression  $\text{avg}\{f(y)\}_{y \in N(x)} = \frac{1}{d}(A f)(x)$ . If  $f$  is an elementary function with eigenvalue  $\lambda$ , then the average is computed as:

$$\begin{aligned} \text{avg}\{f(y)\}_{y \in N(x)} &= \text{avg}_{y \in N(x)} \{f(y) - \bar{f}\} + \bar{f} = \frac{1}{d}(A(f - \bar{f}))(x) + \bar{f} \\ &= \frac{d - \lambda}{d}(f(x) - \bar{f}) + \bar{f} = f(x) + \frac{\lambda}{d}(\bar{f} - f(x)), \end{aligned}$$

and we get Grover’s wave equation. In the previous expression we used the fact that  $f - \bar{f}$  is an eigenfunction of  $A$  with eigenvalue  $d - \lambda$ .

The previous definitions are general concepts of landscape theory. Let us focus now on the binary strings with the one-change neighborhood, which is the representation and the neighborhood we use in the next section to compute the fitness-distance correlation. In this case the solution set  $X$  is the set of all binary strings of size  $n$ . Two solutions  $x$  and  $y$  are neighboring if one can be obtained from the other by flipping a bit, that is, if the Hamming distance between the solutions, denoted with  $\mathcal{H}(x, y)$ , is 1.

One relevant set of eigenvectors of the Laplacian in the binary representation is that of Walsh functions (4). Furthermore, the Walsh functions form an orthogonal basis of eigenvectors in the configuration space. Thus, they have been used to find the elementary landscape decomposition of problems with a binary representation like the MAXSAT [14]. Given the space of binary strings of length  $n$ ,  $\mathbb{B}^n$ , a (non-normalized) Walsh function with parameter  $w \in \mathbb{B}^n$  is defined as:

$$\psi_w(x) = \prod_{i=1}^n (-1)^{w_i x_i} = (-1)^{\sum_{i=1}^n w_i x_i}. \quad (4)$$

Two useful properties of Walsh functions are  $\psi_w \cdot \psi_v = \psi_{w+v}$  where  $w + v$  is the bitwise sum in  $\mathbb{Z}_2$  of  $w$  and  $v$ ; and  $\psi_w^2 = \psi_w \cdot \psi_w = \psi_{2w} = \psi_0 = 1$ . We define the *order* of a Walsh function  $\psi_w$  as the value  $\langle w|w \rangle = \sum_{i=1}^n w_i$ , that is, the number of ones in  $w$ . A Walsh function with order  $p$  is elementary with eigenvalue  $\lambda = 2p$  [16]. The average value of a Walsh function of order  $p > 0$  is zero, that is,  $\overline{\psi_w} = 0$  if  $w$  has at least one 1. The only Walsh function of order  $p = 0$  is  $\psi_0 = 1$ , which is a constant.

Since the Walsh functions form an orthogonal basis of  $\mathbb{R}^{2^n}$ , any arbitrary pseudoboolean function can be written as a weighted sum of Walsh functions in the following way:  $f = \sum_{w \in \mathbb{B}^n} a_w \psi_w$ , where the values  $a_w$  are called Walsh coefficients. We can group together the Walsh functions having the same order to find the elementary landscape decomposition of the function. That is:  $f_{[p]} = \sum_{\substack{w \in \mathbb{B}^n \\ \langle w|w \rangle = p}} a_w \psi_w$ , where each  $f_{[p]}$  is an eigenvector of the Laplacian with eigenvalue  $2p$ , also called order- $p$  elementary component of  $f$ . The function  $f$  can be written as a sum of the  $n + 1$  elementary components, that is:  $f = \sum_{p=0}^n f_{[p]}$ .

We define the sphere of radius  $k$  around a solution  $x$  as the set of all solutions lying at Hamming distance  $k$  from  $x$  [19]. In analogy to the adjacency matrix we define the sphere matrices of radius  $k$  as  $S_{xy}^{(k)} = 1$  if  $\mathcal{H}(x, y) = k$  and  $S_{xy}^{(k)} = 0$  otherwise.

The sphere matrix of radius one is the adjacency matrix of the one-change neighborhood,  $A$ , and the sphere matrix of radius zero is the identity matrix,  $I$ . Each sphere matrix  $S^{(k)}$  can be written as a polynomial in  $A$  (the adjacency matrix) [5]. Then, each eigenvector of  $A$  is an eigenvector of  $S^{(k)}$ , with a different eigenvalue. As a consequence, the eigenvectors of the Laplacian matrix  $\Delta$  are eigenvectors of the sphere matrices  $S^{(k)}$ . Furthermore, an order- $p$  function (having eigenvalue  $2p$  for  $-\Delta$ ) is eigenvector of the sphere matrix  $S^{(k)}$  with eigen-

value  $\mathcal{K}_{k,p}^{(n)}$ , which is the  $(k, p)$  element of the  $n$ -th Krawtchouk matrix (see [19] for details). Krawtchouk matrices can be defined with the equation:

$$\mathcal{K}_{k,p}^{(n)} = \sum_{l=0}^n (-1)^l \binom{n-p}{k-l} \binom{p}{l}, \quad (5)$$

where  $n \geq 0$ ,  $0 \leq k, p \leq n$  and we consider that  $\binom{a}{b} = 0$  when  $b < 0$  or  $b > a$ .

The interested reader can deepen on Krawtchouk matrices in [7], here we only highlight their properties relevant to our mathematical derivations. One important property of the Krawtchouk matrices is:

$$(1+x)^{n-p}(1-x)^p = \sum_{k=0}^n x^k \mathcal{K}_{k,p}^{(n)}. \quad (6)$$

**Proposition 1.** *The following identity for the Krawtchouk matrices holds:*

$$\sum_{k=0}^n k \mathcal{K}_{k,p}^{(n)} = \begin{cases} n2^{n-1} & \text{if } p = 0, \\ -2^{n-1} & \text{if } p = 1, \\ 0 & \text{if } p > 1. \end{cases} \quad (7)$$

*Proof.* Taking the derivative of the two sides of (6) we have:

$$(n-p)(1+x)^{n-p-1}(1-x)^p - p(1+x)^{n-p}(1-x)^{p-1} = \sum_{k=1}^n x^{k-1} k \mathcal{K}_{k,p}^{(n)}. \quad (8)$$

If we set  $x = 1$ , the right hand side is the left hand side of (7). In the left hand side of (8) we can distinguish three cases:

- Case  $p = 0$ : the derivative polynomial evaluated in  $x = 1$  is  $n2^{n-1}$ .
- Case  $p = 1$ : the derivative polynomial evaluated in  $x = 1$  is  $-2^{n-1}$ .
- Case  $p > 1$ : the derivative polynomial evaluated in  $x = 1$  is 0.

□

Each component  $f_{[p]}$  of the elementary landscape decomposition of  $f$  is an eigenvector of the sphere matrix of radius  $k$  with eigenvalue  $\mathcal{K}_{k,p}^{(n)}$ . Thus, we can compute the sum of the fitness value in a sphere of radius  $k$  around  $x$  as:

$$\sum_{\substack{y \in \mathbb{B}^n \\ \mathcal{H}(x,y)=k}} f(y) = \sum_{p=0}^n \mathcal{K}_{k,p}^{(n)} f_{[p]}(x) \quad (9)$$

### 3 Fitness-Distance Correlation

The Fitness-Distance Correlation (FDC) is a measure introduced by Jones and Forrest [10] to measure problem difficulty. Given all the solutions in the search space, it computes the correlation coefficient between the fitness values of these solutions and the Hamming distances of the solutions to their nearest global optimum.

**Definition 2.** Given a function  $f : \mathbb{B}^n \mapsto \mathbb{R}$  the fitness-distance correlation for  $f$  is defined as

$$r = \frac{\text{Cov}_{fd}}{\sigma_f \sigma_d}, \quad (10)$$

where  $\text{Cov}_{fd}$  is the covariance of the fitness values and the distances of the solutions to their nearest global optimum,  $\sigma_f$  is the standard deviation of the fitness values in the search space and  $\sigma_d$  is the standard deviation of the distances to the nearest global optimum in the search space. Formally:

$$\begin{aligned} \text{Cov}_{fd} &= \frac{1}{2^n} \sum_{x \in \mathbb{B}^n} (f(x) - \bar{f})(d(x) - \bar{d}), \\ \bar{f} &= \frac{1}{2^n} \sum_{x \in \mathbb{B}^n} f(x), \quad \sigma_f = \sqrt{\frac{1}{2^n} \sum_{x \in \mathbb{B}^n} (f(x) - \bar{f})^2}, \\ \bar{d} &= \frac{1}{2^n} \sum_{x \in \mathbb{B}^n} d(x), \quad \sigma_d = \sqrt{\frac{1}{2^n} \sum_{x \in \mathbb{B}^n} (d(x) - \bar{d})^2}, \end{aligned} \quad (11)$$

where the function  $d(x)$  is the Hamming distance between  $x$  and its nearest global optimum.

The FDC  $r$  is a value between  $-1$  and  $1$ . The lower the absolute value of  $r$ , the more difficult the optimization problem is supposed to be. The exact computation of the FDC using the previous definition requires the evaluation of the complete search space. It is required to determine the global optima to define  $d(x)$  and compute the statistics for  $d$  and  $f$ . If the objective function  $f$  is a constant function, then the FDC is not well-defined, since  $\sigma_f = 0$ .

In the following we will focus on the case in which there exists one only global optimum  $x^*$  and we know the elementary landscape decomposition of  $f$ . The following lemma provides an expression for  $\bar{d}$  and  $\sigma_d$  in this case.

**Lemma 1.** Given an optimization problem defined over  $\mathbb{B}^n$ , if there is only one global optimum  $x^*$ , then the distance function  $d(x)$  defined in Definition 2 is the Hamming distance between  $x$  and  $x^*$  and its average and standard deviation in the whole search space are given by

$$\bar{d} = \frac{n}{2}, \quad \sigma_d = \frac{\sqrt{n}}{2}. \quad (12)$$

*Proof.* Since there is only one global optimum, the function  $d(x)$  is defined as  $d(x) = \mathcal{H}(x, x^*)$ . Given an integer number  $0 \leq k \leq n$ , the number of solutions at distance  $k$  from  $x^*$  is  $\binom{n}{k}$ . Then we can compute the two first raw moments

of  $d(x)$  over the search space as:

$$\begin{aligned}\alpha_1 = \bar{d} &= \frac{1}{2^n} \sum_{k=0}^n \binom{n}{k} k = \frac{n2^{n-1}}{2^n} = \frac{n}{2}, \\ \alpha_2 = \overline{d^2} &= \frac{1}{2^n} \sum_{k=0}^n \binom{n}{k} k^2 = \frac{n(n+1)2^{n-2}}{2^n} = \frac{n(n+1)}{4}.\end{aligned}$$

Using these moments we can compute the standard deviation as  $\sqrt{\alpha_2 - \alpha_1^2}$ , which yields:

$$\sigma_d = \sqrt{\frac{n(n+1)}{4} - \frac{n^2}{4}} = \sqrt{\frac{n}{4}} = \frac{\sqrt{n}}{2}. \quad (13)$$

□

Now we are ready to prove the main result of this work.

**Theorem 1.** *Let  $f$  be an objective function whose elementary landscape decomposition is  $f = \sum_{p=0}^n f_{[p]}$ , where  $f_{[0]}$  is the constant function  $f_{[0]}(x) = \bar{f}$  and each  $f_{[p]}$  with  $p > 0$  is an order- $p$  elementary function with zero offset. If there exists only one global optimum in the search space  $x^*$ , the FDC can be exactly computed as:*

$$r = \frac{-f_{[1]}(x^*)}{\sigma_f \sqrt{n}}. \quad (14)$$

*Proof.* Let us expand the covariance as

$$\begin{aligned}Cov_{fd} &= \frac{1}{2^n} \sum_{x \in \mathbb{B}^n} f(x)d(x) - \bar{f} \bar{d} = \frac{1}{2^n} \sum_{k=0}^n k \sum_{\substack{x \in \mathbb{B}^n \\ \mathcal{H}(x, x^*)=k}} f(x) - \bar{f} \frac{n}{2} \\ &= \frac{1}{2^n} \sum_{k=0}^n k \sum_{\substack{x \in \mathbb{B}^n \\ \mathcal{H}(x, x^*)=k}} \sum_{p=0}^n f_{[p]}(x) - f_{[0]} \frac{n}{2} = \frac{1}{2^n} \sum_{k=0}^n k \sum_{p=0}^n \mathcal{K}_{k,p}^{(n)} f_{[p]}(x^*) - f_{[0]} \frac{n}{2} \\ &= \sum_{p=0}^n \left( \frac{1}{2^n} \sum_{k=0}^n k \mathcal{K}_{k,p}^{(n)} \right) f_{[p]}(x^*) - f_{[0]} \frac{n}{2} = \frac{n}{2} f_{[0]} - \frac{1}{2} f_{[1]}(x^*) - f_{[0]} \frac{n}{2} \\ &= -\frac{1}{2} f_{[1]}(x^*),\end{aligned} \quad (15)$$

where we used the result in Proposition 1. Substituting in (10) we obtain (14). □

The previous theorem shows that the only thing we need to know on the global optimum is the value of the first elementary component. With this information we can exactly compute the FDC. Some problems for which we know the elementary landscape decomposition based on the numerical data defining a problem instance are MAX-SAT, 0-1 Unconstrained Quadratic Optimization

(UQO), the Subset Sum problem (SS), the NK-landscapes, etc. For all of them we could provide expressions for their FDC.

The result of the previous theorem starts an interesting discussion. Some works on landscape analysis claim that the ruggedness of a landscape is related to its hardness [2]. The *autocorrelation coefficient*  $\xi$  and the *autocorrelation length*  $\ell$  of a problem are two measures of the ruggedness of the problem proposed to characterize an objective function in a way that allows one to estimate the performance of a local search method: the lower their value the higher their ruggedness. Angel and Zissimopoulos [1] have studied the relationship between the performance of a local search and the autocorrelation coefficient. Also a relationship has been noticed between the autocorrelation length and the expected number of local optima of a problem [8]. Furthermore, the *autocorrelation length conjecture* [17] claims that the higher the value of  $\xi$  and  $\ell$ , the lower the number of local optima and, as a consequence, the better could be the performance of a local search method. In summary, empirical and theoretical results support the hypothesis that a rugged landscape is more difficult than a problem with a smooth landscape.

In the case of the elementary functions defined over binary strings, the functions with higher order are more rugged than the ones with lower order. The order-1 elementary landscapes are the smoothest landscapes and, in fact, they can always be solved in polynomial time. Following this chain of reasoning, in a general landscape, the elementary components with order  $p > 1$  are the ones that make the problem difficult. However, from Theorem 1 we observe that only the order-1 elementary component of a function  $f$  is taken into account in the computation of the FDC. This fact poses some doubts on the value of the FDC as a measure of difficulty of a problem, since FDC is shown to neglect the rest of information captured in the higher order components. This is true under the assumption that one single global optimum exists in the search space. We defer to future work the analysis of the general case. The doubts on FDC as being a difficulty indicator have also been raised by other authors. Two examples are the work by Tomassini et al. [21] focused on genetic programming and the one by Bierwirth et al. [3] based on the Job Shop Scheduling.

### 3.1 Fitness-Distance Correlation for Elementary Landscapes

If the objective function is elementary, then the expression of the exact FDC is specially simple, as the following corollary proves.

**Corollary 1.** *Let  $f$  be an elementary function of order  $p > 0$  with one only global optimum  $x^*$ , then the fitness-distance correlation can be exactly computed using the following expression:*

$$r = \begin{cases} \frac{\bar{f} - f(x^*)}{\sigma_f \sqrt{n}} & \text{if } p = 1 \\ 0 & \text{if } p > 1 \end{cases} \quad (16)$$

*Proof.* An elementary function  $f(x)$  of order  $p > 0$  can always be written as the sum of the two eigenvectors of the adjacency matrix:  $f(x) = f_{[0]} + f_{[p]}(x)$  where  $f_{[0]} = \bar{f}$ . Applying the result of Theorem 1 we obtain (16).  $\square$



The previous corollary states that only elementary landscapes with order  $p = 1$  have a nonzero FDC. Furthermore, the FDC does depend on the value of the objective function in the global optimum  $f(x^*)$  and the average value  $\bar{f}$ , but not on the solution  $x^*$  itself. We can also observe that if we are maximizing, then  $f(x^*) > \bar{f}$  and the FDC is negative, while if we are minimizing  $f(x^*) < \bar{f}$  and the FDC is positive.

Interestingly, the order-1 elementary landscapes can always be written as linear functions and they can be optimized in polynomial time. That is, if  $f$  is an order-1 elementary function then it can be written in the following way:

$$f(x) = \sum_{i=1}^n a_i x_i + b. \quad (17)$$

where  $a_i$  and  $b$  are real values. The following proposition provides the average and the standard deviation for this family of functions.

**Proposition 2.** *Let  $f$  be an order-1 elementary function, which can be written as (17). Then, the average and the standard deviation of the function values in the whole search space are:*

$$\bar{f} = b + \frac{1}{2} \sum_{i=1}^n a_i, \quad \sigma_f = \frac{1}{2} \sqrt{\sum_{i=1}^n a_i^2}. \quad (18)$$

*Proof.* Using the linearity property of the average we can write:  $\bar{f} = \sum_{i=1}^n a_i \bar{x}_i + b$ , and  $\bar{f}$  in (18) follows from the fact that  $\bar{x}_i = 1/2$ . Now we can compute the variance of  $f$  as:

$$\begin{aligned} \text{Var}[f] &= \overline{(f(x) - \bar{f})^2} = \overline{\left( \sum_{i=1}^n a_i x_i - \frac{1}{2} \sum_{i=1}^n a_i \right)^2} = \overline{\left( \sum_{i=1}^n a_i \left( x_i - \frac{1}{2} \right) \right)^2} \\ &= \sum_{i,j=1}^n a_i a_j \overline{\left( x_i - \frac{1}{2} \right) \left( x_j - \frac{1}{2} \right)} = \sum_{i,j=1}^n a_i a_j \left( \overline{x_i x_j} - \frac{1}{2} \bar{x}_i - \frac{1}{2} \bar{x}_j + \frac{1}{4} \right) \\ &= \sum_{i,j=1}^n a_i a_j \left( \overline{x_i x_j} - \frac{1}{4} \right) = \sum_{i,j=1}^n a_i a_j \left( \delta_i^j \frac{1}{4} + \frac{1}{4} - \frac{1}{4} \right) = \frac{1}{4} \sum_{i=1}^n a_i^2, \end{aligned} \quad (19)$$

where we used again  $\bar{x}_i = \bar{x}_j = 1/2$  and  $\overline{x_i x_j} = 1/4(\delta_i^j + 1)$ , being  $\delta_i^j$  the Kronecker delta. The expression for  $\sigma_f$  in (18) follows from (19).  $\square$

Using Proposition 2 we can compute the FDC for the order-1 elementary landscapes.

**Proposition 3.** *Let  $f$  be an order-1 elementary function written as (17) such that all  $a_i \neq 0$ . Then, it has one only global optimum and its FDC (assuming maximization) is:*

$$r = \frac{-\sum_{i=1}^n |a_i|}{\sqrt{n \sum_{i=1}^n a_i^2}}, \quad (20)$$

which is always in the interval  $-1 \leq r < 0$ .

*Proof.* The global optimum  $x^*$  has 1 in all the positions  $i$  such that  $a_i > 0$  and the maximum fitness value is:

$$f(x^*) = b + \sum_{\substack{i=1 \\ a_i > 0}}^n a_i. \quad (21)$$

Using Proposition 2 we can write:

$$\bar{f} - f(x^*) = \left( b + \frac{1}{2} \sum_{i=1}^n a_i \right) - \left( b + \sum_{\substack{i=1 \\ a_i > 0}}^n a_i \right) = -\frac{1}{2} \sum_{i=1}^n |a_i|. \quad (22)$$

Replacing the previous expression and  $\sigma_f$  in (16) we prove the claimed result.  $\square$

When all the values of  $a_i$  are the same, the FDC computed with (20) is  $-1$ . This happens in particular for the Onemax problem. But if there exist different values for  $a_i$ , then we can reach any arbitrary value in  $[-1, 0)$  for  $r$ . The following theorem provides a way to do it.

**Theorem 2.** *Let  $\rho$  be an arbitrary real value in the interval  $[-1, 0)$ , then any linear function  $f(x)$  given by (17) where  $n > 1/\rho^2$ ,  $a_2 = a_3 = \dots = a_n = 1$  and  $a_1$  is*

$$a_1 = \frac{(n-1) + n|\rho|\sqrt{(1-\rho^2)(n-1)}}{n\rho^2 - 1} \quad (23)$$

*has exactly FDC  $r = \rho$ .*

*Proof.* The expression for  $a_1$  is well-defined since  $n\rho^2 > 1$ . Replacing all the  $a_i$  in (20) we get  $r = \rho$ .  $\square$

Theorem 2 provides a solid argument against the use of FDC as a measure of the difficulty of a problem. In effect, we can always build an optimization problem based on a linear function, which can be solved in polynomial time, with an FDC as near as desired to 0 (but not zero), that is, as “difficult” as desired according to the FDC. However, we have to highlight here that for a given FDC value  $\rho$  we need at least  $n > 1/\rho^2$  variables. Thus, an FDC nearer to 0 requires more variables.

## 4 FDC, Autocorrelation Length and Local Optima

The autocorrelation length  $\ell$  [8] has also been used as a measure of the difficulty of a problem. Chicano and Alba [4] found a negative correlation between  $\ell$  and the number of local optima in the 0-1 Unconstrained Quadratic Optimization problem (0-1 UQO), an NP-hard problem [9]. Kinnear [11] also studied the use of the autocorrelation measures as problem difficulty, but the results were inconclusive. In this section we investigate which of the two measures,  $\ell$  or the

absolute value of FDC, is more correlated to the number of local optima for some random instances of the 0-1 UQO. In particular, we have randomly generated 1650 UQO instances using the Palubeckis instance generator [13]. The size of the instances varies between  $n = 10$  and  $n = 20$  and the density (percentage of nonzero elements in the coefficients matrix) varies from 10 to 90 in steps of 20. For each  $n$  and density, 30 random instances were generated by randomly selecting the nonzero elements of the coefficients matrix from the interval  $[-100, 100]$ . For all the instances we computed the autocorrelation length  $\ell$ , the absolute value of the FDC  $|r|$  and the number of local optima (minima) by complete enumeration of the search space. In Table 1 we show the Spearman rank correlation coefficient between the number of local optima and  $\ell$  and  $|r|$ . The correlations are computed using all the instances with the same size  $n$ .

**Table 1.** Spearman correlation coefficient for the number of local optima against the autocorrelation length ( $\ell$ ) and the absolute value of the FDC ( $|r|$ ).

$n$	10	11	12	13	14	15
$\ell$	-0.5467	-0.5545	-0.5896	-0.4796	-0.4725	-0.5511
$ r $	-0.1407	-0.1843	-0.0787	-0.1203	-0.1944	-0.0538
$n$	16	17	18	19	20	
$\ell$	-0.4959	-0.5740	-0.5872	-0.5249	-0.4829	
$ r $	-0.1251	-0.1791	-0.1339	-0.3310	-0.0338	

We can observe a high inverse correlation (around  $-0.5$ ) between the number of local optima and the autocorrelation length, supporting the autocorrelation length conjecture. However, the correlation between the number of local optima and FDC is low, again supporting the hypothesis that FDC is not an appropriate measure of the difficulty of a problem (this time, from an experimental point of view).

## 5 Conclusion

We have applied landscape theory to exactly compute the Fitness-Distance Correlation of combinatorial optimization problems defined over sets of binary strings. The result is valid in the case in which one single global optimum exists in the landscape. We defer to future work the analysis of the general case.

The expression for the FDC takes only into account the order-1 elementary component of the objective function, while previous work suggests that the components making a problem difficult are the higher order elementary components. This fact questions the use of FDC as a measure of difficulty of the problem. We prove that there exist polynomial time solvable problems with an FDC arbitrarily near to zero. An experimental study over random instances of the 0-1 UQO shows a low correlation between FDC and the number of local optima, supporting the hypothesis that FDC fails to capture the problem difficulty.

## References

1. Angel, E., Zissimopoulos, V.: On the landscape ruggedness of the quadratic assignment problem. *Theoretical Computer Science* 263, 159–172 (2000)
2. Barnes, J.W., Dimova, B., Dokov, S.P.: The theory of elementary landscapes. *Applied Mathematics Letters* 16, 337–343 (2003)
3. Bierwirth, C., Mattfeld, D., Watson, J.P.: Landscape regularity and random walks for the job-shop scheduling problem. In: *EvoCOP*, pp. 21–30. LNCS 3004 (2004)
4. Chicano, F., Alba, E.: Elementary landscape decomposition of the 0-1 unconstrained quadratic optimization. *Journal of Heuristics* (10.1007/s10732-011-9170-6)
5. Chicano, F., Alba, E.: Exact computation of the expectation curves of the bit-flip mutation using landscapes theory. In: *GECCO*. pp. 2027–2034 (2011)
6. Chicano, F., Whitley, L.D., Alba, E.: A methodology to find the elementary landscape decomposition of combinatorial optimization problems. *Evolutionary Computation* 19(4), 597–637 (2011)
7. Feinsilver, P., Kocik, J.: Krawtchouk polynomials and krawtchouk matrices. In: *Recent Advances in Applied Probability*, pp. 115–141. Springer US (2005)
8. García-Pelayo, R., Stadler, P.: Correlation length, isotropy and meta-stable states. *Physica D: Nonlinear Phenomena* 107(2-4), 240–254 (Sep 1997)
9. Glover, F., Alidaee, B., Rego, C., Kochenberger, G.: One-pass heuristics for large-scale unconstrained binary quadratic problems. *EJOR* 137(2), 272–287 (2002)
10. Jones, T., Forrest, S.: Fitness distance correlation as a measure of problem difficulty for genetic algorithms. In: *GECCO*. pp. 184–192. Morgan Kaufmann (1995)
11. Kinnear, K.E., J.: Fitness landscapes and difficulty in genetic programming. In: *IEEE CEC*. pp. 142–147 vol.1 (jun 1994)
12. Lu, G., Bahsoon, R., Yao, X.: Applying elementary landscape analysis to search-based software engineering. In: *Proceedings of SSBSE* (2010)
13. Palubeckis, G.: Multistart tabu search strategies for the unconstrained binary quadratic optimization problem. *Annals of Oper. Research* 131, 259–282 (2004)
14. Rana, S., Heckendorn, R.B., Whitley, D.: A tractable walsh analysis of SAT and its implications for genetic algorithms. In: *Proceedings of AAAI*. pp. 392–397 (1998)
15. Reidys, C.M., Stadler, P.F.: Combinatorial landscapes. *SIAM Review* 44(1), 3–54 (2002)
16. Stadler, P.F.: Toward a theory of landscapes. In: López-Peña, R., Capovilla, R., García-Pelayo, R., H.Waelbroeck, Zertruche, F. (eds.) *Complex Systems and Binary Networks*. pp. 77–163. Springer-Verlag (1995)
17. Stadler, P.F.: *Biological Evolution and Statistical Physics*, chap. Fitness Landscapes, pp. 183–204. Springer (2002)
18. Sutton, A.M., Whitley, D., Howe, A.E.: Mutation rates of the (1+1)-EA on pseudo-boolean functions of bounded epistasis. In: *GECCO*. pp. 973–980. ACM (2011)
19. Sutton, A.M., Whitley, L.D., Howe, A.E.: Computing the moments of k-bounded pseudo-boolean functions over hamming spheres of arbitrary radius in polynomial time. *Theoretical Computer Science* (10.1016/j.tcs.2011.02.006)
20. Sutton, A.M., Whitley, L.D., Howe, A.E.: A polynomial time computation of the exact correlation structure of k-satisfiability landscapes. In: *Proceedings of GECCO*. pp. 365–372. ACM, New York, NY, USA (2009)
21. Tomassini, M., Vanneschi, L., Collard, P., Clergue, M.: A study of fitness distance correlation as a difficulty measure in genetic programming. *Evolutionary Computation* 13(2), 213–239 (2005)
22. Whitley, D., Sutton, A.M., Howe, A.E.: Understanding elementary landscapes. In: *Proceedings of GECCO*. pp. 585–592. ACM, New York, NY, USA (2008)