Elementary Landscape Decomposition of the 0-1 Unconstrained Quadratic Optimization

Francisco Chicano · Enrique Alba

Abstract Landscapes' theory provides a formal framework in which combinatorial optimization problems can be theoretically characterized as a sum of a especial kind of landscape called elementary landscape. The elementary landscape decomposition of a combinatorial optimization problem is a useful tool for understanding the problem. Such decomposition provides an additional knowledge on the problem that can be exploited to explain the behavior of some existing algorithms when they are applied to the problem or to create new search methods for the problem. In this paper we analyze the 0-1 Unconstrained Quadratic Optimization from the point of view of landscapes' theory. We prove that the problem can be written as the sum of two elementary components and we give the exact expressions for these components. We use the landscape decomposition to compute autocorrelation measures of the problem, and show some practical applications of the decomposition.

Keywords Fitness landscapes, unconstrained quadratic optimization, landscapes theory

1 Introduction

Landscapes' theory focus on the analysis of the objective function of optimization problems (Stadler 1995). This theory has applications not only in evolutionary computation (Whitley et al 2008) but also in Chemistry (Stadler 1996), Biology (Weinberger 1990) and Physics (García-Pelayo and Stadler 1997). One of the objectives of this theory is to better understand the structure of an optimization problem. The knowledge obtained from the objective function using this theory could be used to improve the

Francisco Chicano
E.T.S. Ingeniería Informática
University of Málaga
E-mail: chicano@lcc.uma.es

Enrique Alba
E.T.S. Ingeniería Informática
University of Málaga
E-mail: eat@lcc.uma.es

performance of heuristic and metaheuristic algorithms by proposing new operators, search strategies, or even new values for tuning their parameters.

A *landscape* for a combinatorial optimization problem is a triple (X, N, f), where $f : X \to \mathbb{R}$ is the objective function and the *neighborhood* function N maps a solution $x \in X$ to the set of neighboring solutions. If $y \in N(x)$ then y is a neighbor of x. There exists a especial kind of landscape which is of particular interest due to their properties. They are the *elementary landscapes*, characterized by the *Grover's wave equation*:

$$\operatorname*{avg}_{y \in N(x)}\{f(y)\} = f(x) + \frac{k}{d}\left(\bar{f} - f(x)\right)$$

where d is the size of the neighborhood, $|N(x)|$, which we assume the same for all the solutions in the search space, $\bar{f}$ is the average solution evaluation over the entire search space, and k is a characteristic constant. 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 $f(x)$; we denote this average by $\operatorname{avg}\{f(y)\}_{y \in N(x)}$, calculated as:

$$\operatorname*{avg}_{y \in N(x)}\{f(y)\} = \frac{1}{|N(x)|} \sum_{y \in N(x)} f(y)$$

Assuming $f(x) \neq \bar{f}$ then

$$f(x) < \min\left\{\operatorname*{avg}_{y \in N(x)}\{f(y)\}, \bar{f}\right\} \vee f(x) > \max\left\{\operatorname*{avg}_{y \in N(x)}\{f(y)\}, \bar{f}\right\}$$

This means that all maxima are greater than $\bar{f}$ and all minima are less than $\bar{f}$ (Codenotti and Margara 1992).

A landscape (X, N, f) is not always elementary, since the wave equation could not hold. However, even in this case it is possible to characterize the function f as the sum of elementary landscapes. In particular, if the neighborhood N is symmetric then it is possible to find an orthogonal basis composed of elementary real functions. Then, every regular function (elementary or not) can be written as the sum of a set of elementary landscapes. The process of decomposing a landscape into its elementary components is what we call *elementary landscape decomposition*.

In the case of functions defined over a binary space with the bit-flip neighborhood, called pseudoboolean functions, the elementary landscape decomposition can be easily obtained with the help of the Walsh functions (Walsh 1923). Sutton et al (2011) present the foundations on this topic and provide valuable formulas and insights on the theoretical aspects of the decomposition of pseudoboolean functions that can be applied to the 0-1 UQO. In particular, they show how it is possible to compute summary statistics over the entire search space in an efficient way using the landscape decomposition.

The decomposition is useful from the theoretical and practical points of view. In theory, the landscape decomposition of a problem can be used to compute the exact expression for the autocorrelation functions, the autocorrelation coefficient, and the autocorrelation length (Angel and Zissimopoulos 2000a). The previous measures characterize an objective function in a way that allows one to estimate the performance of a local search method. Angel and Zissimopoulos (2000b) have studied the correlation between the performance of a local search and the autocorrelation coefficient. Also a correlation has been noticed between the autocorrelation length and the expected number of local optima of a problem (García-Pelayo and Stadler 1997).

Sutton et al (2009) provides closed-form formulas that show how the autocorrelation measures can be computed from the elementary landscape decomposition of any pseudoboolean function. In their work, they focus on the MAX-k-SAT problem, but the general results can also be applied to the 0-1 UQO problem, since it is a kind of pseudoboolean function. From these general results, in this paper we will be able to give a particular expression for the autocorrelation measures in the case of the 0-1 UQO.

In practice, the landscape decomposition together with Grover's wave equation can be used to compute the average value of the objective function in the neighborhood, which can be used as a base for new operators or algorithms. In this sense, the work by Sutton et al (2010) provides a new search strategy to escape from plateaus guided by the average value of the objective function evaluated not only in the neighborhood of a solution but in the second-order, thrid-order, and, in general, n-order neighborhood, called Hamming spheres around a solution. Their proposed approach is applied to the MAX-k-SAT problem but can be generalized to other pseudoboolean functions. Another example of practical application of the elementary landscape decomposition is that of Lu et al (2010), who use the grover's wave equation in the core of a new search method based on hill climbing.

The 0-1 Unconstrained Quadratic Optimization (UQO) is a relevant optimization problem which has been extensively dealt with in the literature (Glover et al 2002). However, to the best of our knowledge, this problem has not been analyzed before from the point of view of landscapes' theory. In this work we find the landscape decomposition of UQO and revise some practical and theoretical applications of the decomposition. In particular, we present a selection strategy based on the average value of the objective function in the neighborhood of a solution. The strategy is viable only if the landscape decomposition is known. We also give a closed-form formula for the autocorrelation length and the autocorrelation coefficient that can be computed in polynomial time. The previous parameters are a measure of the ruggedness of a landscape.

The organization of the paper is as follows. In Section 2 we present the foundations of landscapes' theory required in this paper with the aim of making the work self-contained. Section 3 presents the 0-1 Unconstrained Quadratic Optimization and its landscape decomposition is shown in Section 4. We revise some practical and theoretical applications of the decomposition in Section 5 and finish the paper with some conclusions and future work in Section 6.

2 Background on Landscapes' Theory

In this section we present some fundamental results on landscapes' theory. After defining the concept of elementary landscape, we present Grover's wave equation in Theorem 1 and we prove in Theorem 2 that any landscape with a symmetric neighborhood can be written as the sum of elementary landscapes.

Let X be a finite set of solutions, $f : X \to \mathbb{R}$ be a real-valued function defined on X and $N : X \to \mathcal{P}(X)$ the neighborhood operator. The pair (X, N) is called *configuration space* and can be represented using a graph $G(X, E)$ in which X is the set of vertices and a directed edge (x, y) exists in E if $y \in N(x)$ (Biyikoglu et al 2007). We can represent the neighborhood operator by its adjacency matrix

$$A_{xy} = \begin{cases} 1 \text{ if } y \in N(x) \\ 0 \text{ otherwise} \end{cases}$$

The degree matrix D is defined as the diagonal matrix

$$D_{xy} = \begin{cases} |N(x)| & \text{if } x = y \\ 0 & \text{otherwise} \end{cases}$$

Any discrete function over the set of candidate solutions, e.g. f, can be characterized as a vector in $\mathbb{R}^{|X|}$. Using the graph representation of the configuration space, any function f can be interpreted as a labeling of the nodes in the graph, where the label of node x is $f(x)$. Any $|X| \times |X|$ matrix can be interpreted as a linear map that acts on vectors in $\mathbb{R}^{|X|}$. The Laplacian matrix of a neighborhood operator is defined as

$$\Delta = A - D$$

The Laplacian matrix acts on function f as follows

$$\Delta\, f = \begin{pmatrix} \sum_{y \in N(x_1)} (f(y) - f(x_1)) \\ \sum_{y \in N(x_2)} (f(y) - f(x_2)) \\ \vdots \\ \sum_{y \in N(x_{|X|})} \left(f(y) - f(x_{|X|}) \right) \end{pmatrix}$$

The component x of this matrix-vector product can thus be written as:

$$(\Delta\, f)(x) = \sum_{y \in N(x)} (f(y) - f(x)) \tag{1}$$

In this paper, we will restrict our attention to *regular neighborhoods*, where $|N(x)| = d > 0$ for a constant d, for all $x \in X$. When a neighborhood is regular, $\Delta = A - dI$. We say that a neighborhood is *symmetric* if $y \in N(x)$ implies $x \in N(y)$. Stadler (1995) defines the class of elementary landscapes where the function f is an eigenvector (or eigenfunction) of the Laplacian up to an additive constant. Formally, a function f is elementary if there exists a constant b and an eigenvalue λ of Δ such that $\Delta(f - b) = \lambda(f - b)$. This means that every elementary function, f, can be written as the sum of an eigenfunction of Δ, g, and a constant b, i.e., $f = g + b$.

Taking into account basic results of linear algebra, it is not difficult to prove that if f is elementary with eigenvalue λ, $af + b$ is also elementary with the same eigenvalue λ. If f is a constant function, i.e., $f(x) = b\ \forall x \in X$ for a constant b, then $\Delta\, f = 0$ and f is eigenfunction of Δ with eigenvalue $\lambda = 0$. But one of the most interesting properties of elementary landscapes is the so-called Grover's wave equation (Grover 1992), which we present without proof in the next theorem.

Theorem 1 (Grover's wave equation) *Let (X, N, f) be a landscape where the neighborhood, N, is regular and symmetric. Then, f is elementary if and only if the following expression holds*

$$\underset{y \in N(x)}{\mathrm{avg}} \{f(y)\} = f(x) + \frac{k}{d} \left(\bar{f} - f(x) \right) \quad \forall x \in X \tag{2}$$

where k is the additive inverse of the eigenvalue of f, that is, $k = -\lambda$.

From Grover's wave equation we conclude that, in an elementary landscape, there exists a linear relationship between the average of the function in the neighborhood of a solution and the value of the function in that solution. To finish this section, let us prove that any objective function can be written as the sum of elementary landscapes when the neighborhood is symmetric.

Theorem 2 (Elementary landscape decomposition) *Let (X, N, f) be a landscape where the neighborhood, N, is symmetric. Then, there exist n elementary landscapes with $1 \leq n \leq |X|$ such that f can be written as the sum of all of them.*

Proof From linear algebra we know that if a square real matrix Δ of size $|X|$ is symmetric, then there exists an orthogonal basis of the vectorial space $\mathbb{R}^{|X|}$ that is composed of eigenvectors of Δ. Then, we can write every vector of $\mathbb{R}^{|X|}$ as the weighted sum of the vectors in the orthogonal basis. If we translate these concepts into the landscapes language, this means that for any symmetric neighborhood N it is possible to find an orthogonal basis composed of elementary functions. Then, any function f can be written as the weighted sum of a set of elementary landscapes.

3 The 0-1 Unconstrained Quadratic Optimization (UQO)

Using the previous concepts and theorems we will analyze the elementary nature of the UQO problem in Section 4. Let us first precise the problem itself.

The 0-1 Unconstrained Quadratic Optimization (UQO) consists in finding a binary vector $x \in \{0,1\}^n$ to minimize the function:

$$f(x) = \sum_{i,j=1}^{n} q_{ij} x_i x_j \tag{3}$$

where q_{ij} are constants that depend on the problem instance. This problem is NP-hard (Garey and Johnson 1979) and many other combinatorial optimization problems can be formulated as special cases of UQO. Some examples are determining maximum cliques, maximum cuts, maximum vertex packing, minimum coverings, maximum independent sets, and maximum independent weighted sets (Glover et al 2002).

This problem has applications in social psychology (Harary 1953), financial analysis (McBride and Yormack 1980), computer aided design (Krarup and Pruzan 1978), traffic management (Gallo et al 1980), machine scheduling (Alidaee et al 1994), cellular radio channel allocation (Chartaire and Sutter 1994), and molecular conformation (Philips and Rosen 1994) among others.

Due to its importance the problem has been solved using exact methods (Helmberg and Rendl 1998) and metaheuristics like Tabu Search (Liu et al 2005), Scatter Search (Amini et al 1999), GRASP (Palubeckis and Tomkevièius. 2002), Simulated Annealing (Katayama and Narihisa 2001) and Evolutionary Algorithms (Merz and Katayama 2004).

4 Landscape Decomposition of UQO

The goal of this section is to find the elementary components of the objective function of UQO. First, we need to completely define the landscape. In the previous section we

defined the search space (binary strings of length n) and the objective function, therefore we only need to define the neighborhood N. We will use the bit-flip neighborhood. That is, we say that two solutions x and y are neighboring if there exists a binary digit i in which they differ $x_i \neq y_i$ and for all the other binary digits both solutions have the same value $x_k = y_k$ for all $k \neq i$. Given one arbitrary solution, the number of neighbors of the solution is n, each one consisting in flipping the value of one binary digit. This neighborhood is symmetric and it is the more natural one for all the problems based on binary representation. In particular, the bit-flip neighborhood has commonly been used in the UQO literature (Merz and Freisleben 2002).

Once we have completely defined the landscape, our goal in this section is to find the elementary components of the objective function. That is, we want to find several elementary functions that summed together equals Eq. (3), the objective function of the UQO. As we previously stated in Theorem 2, for a symmetric neighborhood N there exists an orthonormal basis composed of elementary landscapes. Let us denote this basis with $\phi_{\lambda,i}$ where λ is the eigenvalue of the vector (function) and i is an index to distinguish the different vectors with the same eigenvalue. Then, a Fourier expansion of function f is

$$f = \sum_{\lambda} \sum_{i} a_{\lambda,i} \phi_{\lambda,i}$$

where the values $a_{\lambda,i} = \langle \phi_{\lambda,i}, f \rangle$ are the Fourier coefficients. Using this Fourier expansion it is possible to compute the landscape decomposition by summing the terms with the same eigenvalue. Each elementary component can be computed as:

$$f_\lambda = \sum_{i} a_{\lambda,i} \phi_{\lambda,i} \tag{4}$$

A especial case is that of f_0, the elementary landscape with $\lambda = 0$. The function f_0 is the constant value $\bar{f}$, and it can be added to any of the other elementary components without affecting its elementariness.

In the case of the binary strings with the bit-flip neighborhood an orthogonal basis is that of the Walsh functions (Sutton et al 2009). Thus, we can use these functions to provide a landscape decomposition for UQO. This kind of analysis is similar to the decomposition of Sutton et al (2010) for the MAX-k-SAT problem. In fact, Heckendorn et al (1998) proved that a k-bounded function can always be expressed as a sum of k elementary landscapes. A function is said to be k-bounded if it can be expressed as the sum of subfunctions that each depend on at most k bits. According to (3), the UQO problem is a 2-bounded function, which means that it can be expressed as the sum of two elementary landscapes. In the following we are going to use the Walsh functions to give a closed-form formula for these elementary landscapes. Before giving the details we present the Walsh functions and some of their properties.

Given the space of binary strings of length n, $\{0,1\}^n$, a (non-normalized) Walsh function with parameter $w \in \{0,1\}^n$ is defined as:

$$\psi_w(x) = \prod_{i=1}^{n} (-1)^{w_i x_i} = (-1)^{\sum_{i=1}^{n} w_i x_i} \tag{5}$$

It is not difficult to prove that $\psi_w \cdot \psi_v = \psi_{w+v}$ where $w+v$ is the bitwise sum in $\mathbb{Z}_2$ of w and v. Another useful property of Walsh functions is $\psi_w^2 = \psi_w \cdot \psi_w = \psi_{2w} = \psi_0 = 1$.

We define the *order* of a Walsh function ψ_w as the value $\sum_{i=1}^{n} w_i$, that is, the number of ones in w.

The main advantage of using Walsh functions for our goals is that they are elementary functions in the bit-flip neighborhood. In particular, a Walsh function with order p is elementary with characteristic constant $k = 2p$ (Stadler 1995). Furthermore, 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 function.

In this paper we only need Walsh functions of order 1 and 2, so we will introduce a especial notation for them. We denote with ψ_i the first-order Walsh function whose parameter has one only 1 in the i-th position and we denote with ψ_{ij} the second-order Walsh function whose parameter has two 1s in the i-th and j-th positions. Using this new notation, we can write $\psi_i \cdot \psi_j = \psi_{ij}$ if $i \neq j$ and $\psi_i^2 = 1$. These equalities will be useful later in this section.

The first step for the landscape decomposition is to rewrite Eq. (3) using the Walsh functions. We must take into account the equality $\psi_i(x) = (-1)^{x_i} = 1 - 2x_i$ to write:

$$f = \sum_{i,j=1}^{n} q_{ij} \frac{(1 - \psi_i)(1 - \psi_j)}{4} \tag{6}$$

and computing the product we get

$$
\begin{aligned}
f &= \sum_{i,j=1}^{n} q_{ij} \frac{1 - \psi_i - \psi_j + \psi_i \psi_j}{4} \\
&= \frac{1}{4} \left[\sum_{i,j=1}^{n} q_{ij} - \sum_{i,j=1}^{n} q_{ij}(\psi_i + \psi_j) + \sum_{i,j=1}^{n} q_{ij} \psi_i \psi_j \right] \\
&= \frac{1}{4} \left[\left(\sum_{i,j=1}^{n} q_{ij} + \sum_{i=1}^{n} q_{ii} \right) - \sum_{i=1}^{n} v_i \psi_i + \sum_{\substack{i,j=1 \\ i \neq j}}^{n} q_{ij} \psi_{ij} \right]
\end{aligned}
\tag{7}
$$

where we have used the properties $\psi_i \psi_j = \psi_{ij}$ if $i \neq j$, $\psi_i^2 = 1$, and we have introduced a new vector v defined by:

$$v_i = \sum_{j=1}^{n} (q_{ij} + q_{ji}) \tag{8}$$

Equation (7) clearly shows the landscape decomposition of the objective function f. We can distinguish three elementary terms in (7). These terms can be defined as functions:

$$f_0 = \frac{1}{4}\left(\sum_{i,j=1}^{n} q_{ij} + \sum_{i=1}^{n} q_{ii}\right) \tag{9}$$

$$f_2 = -\frac{1}{4}\sum_{i=1}^{n} v_i \psi_i \tag{10}$$

$$f_4 = \frac{1}{4}\sum_{\substack{i,\,j\,=\,1 \\ i\,\neq\,j}}^{n} q_{ij}\psi_{ij} \tag{11}$$

The first function, f_0, is a constant and, thus, is an elementary component with $k = 0$. The second function, f_2, is a weighted sum of first-order Walsh functions. The first-order Walsh functions are elementary with $k = 2$ and the linear combination of elementary functions with the same characteristic constant k is also elementary with constant k. Thus, f_2 is an elementary component with $k = 2$. Finally, the third function, f_4, is a linear combination of second-order Walsh functions, which are elementary with $k = 4$. Then, f_4 is an elementary component with $k = 4$.

We can observe that the objective function f is the sum of the three previous components, that is, $f = f_0 + f_2 + f_4$. As we saw in Section 2, if we sum a constant to an elementary function we get another elementary function, that is, the constants do not change the elementariness of a function. We can use this property of the elementary landscapes to sum f_0 (which is a constant) to f_2 or f_4 and reduce in this way the number of elementary components. Furthermore, it is usual to discard the constant component (if any) when counting the number of elementary components of a landscape. Thus, we can say that the UQO is a sum of two elementary components: $f_0 + f_2$ and f_4. To finish this section we express f_2 and f_4 as functions of x instead of the Walsh functions.

$$f_2(x) = -\frac{1}{4}\sum_{i=1}^{n} v_i(1 - 2x_i) \tag{12}$$

$$f_4(x) = \frac{1}{4}\sum_{\substack{i,\,j\,=\,1 \\ i\,\neq\,j}}^{n} q_{ij}(1 - 2|x_i - x_j|) \tag{13}$$

5 Applications of the Landscape Decomposition

In this section we revise some applications of the landscape decomposition of UQO. First, we present a selection operator that uses the average fitness value in the neighborhood of a solution to select the individuals. Second, we give closed-form formulas for the autocorrelation coefficient and the autocorrelation length of UQO.

5.1 Selection Operators

The landscape decomposition of UQO allows one to compute the average value of the objective function in the neighborhood of any solution without the need of evaluating

all the neighboring solutions. According to the Grover's wave equation (2), this average value for UQO is

$$
\begin{aligned}
\operatorname*{avg}_{y \in N(x)}\{f(y)\} &= \operatorname*{avg}_{y \in N(x)}\{f_0(y)\} + \operatorname*{avg}_{y \in N(x)}\{f_2(y)\} + \operatorname*{avg}_{y \in N(x)}\{f_4(y)\} \\
&= f_0(x) + \left(f_2(x) + \frac{2}{n} \left(\overline{f_2} - f_2(x) \right) \right) + \left(f_4(x) + \frac{4}{n} \left(\overline{f_4} - f_4(x) \right) \right) \\
&= f_0(x) + f_2(x) + f_4(x) - \frac{2}{n} f_2(x) - \frac{4}{n} f_4(x) \\
&= f(x) - \frac{2}{n} f_2(x) - \frac{4}{n} f_4(x)
\end{aligned}
\tag{14}
$$

where we took into account that $\overline{f_2}$ and $\overline{f_4}$ in the whole solution space is zero because the average of the non-null order Walsh functions is zero.

Equation (14) only requires the evaluation in x of f, f_2, and f_4 to compute the average value of f in the n neighbors of x, but we can make the computation even more efficient. The evaluation of f_4 is computationally more expensive than the one of f_2 because of the double sum (see Eq. (11)). If we take into account that $f = f_0 + f_2 + f_4$ we can replace f_4 by $f - f_0 - f_2$ in (14) to get:

$$
\operatorname*{avg}_{y \in N(x)}\{f(y)\} = \left(1 - \frac{4}{n} \right) f(x) + \frac{2}{n} f_2(x) + \frac{4}{n} f_0
\tag{15}
$$

which is more efficient once we have computed $f(x)$, since it only requires the computation of simple sums of order $O(n)$. Equation (15) can be used as a base for new operators that exploit the neighborhood average. For example, it is possible to design new selection operators that select individuals according to the average fitness value in the neighborhood (see Fig. 1). In elementary landscapes, selecting one individual according to its fitness value is the same as selecting one individual according to the average in the neighborhood, since $\operatorname{avg}\{f(y)\}_{y \in N(x)} = a f(x) + b$. However, this is not the case when f is the sum of several elementary landscapes with different characteristic constants. In Fig. 1 we show a situation in which the traditional fitness-based selection strategy would select the solution at the right (assuming minimization) while this does not seem to be the most appropriate selection. An average-based selection strategy would prefer the solution at the left, since the neighbors of the left solution are promising.
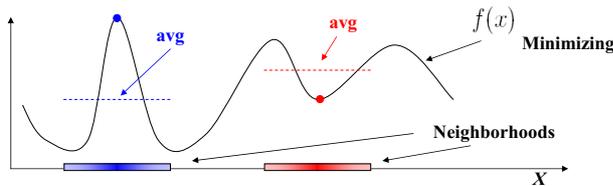

Fig. 1 An average-based selection strategy would be more appropriate than a fitness-based one in the above scenario.

The average-based selection strategy could be especially interesting in the case of plateaus, as shown in Fig. 2. In this case, the average-based selection strategy could

distinguish between two solutions with the same fitness value, taking into account the neighborhood of the solutions.
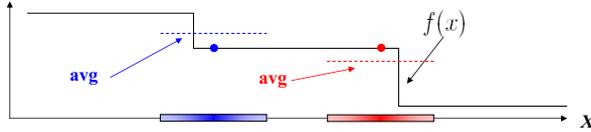

Fig. 2 In plateaus the average-based selection strategy could distinguish between two solutions with the same fitness value.

Sutton et al (2010) did work previously on this idea and they applied it to the MAX-k-SAT problem. In their work, they computed the averages of the objective function not only in the neighborhood of a solution, but in second-order neighbors, third-order neighbors, and, in general, in any sphere of arbitrary radius around the solution itself. In (Sutton et al 2011) they also extend this computation to balls (union of spheres up to a given radius) of arbitrary size. Their empirical results in (Sutton et al 2010) show that guiding the search using the average values over the spheres helps the algorithm to escape more quickly from plateaus.

In the following section we are going to present some theoretical applications of the landscape decomposition of UQO related to the concept of autocorrelation.

5.2 Autocorrelation of the 0-1 UQO

The *autocorrelation coefficient* ξ of a problem is a parameter proposed by Angel and Zissimopoulos (1998) that gives a measure of its ruggedness. This coefficient is related to the number of local optima of a landscape and the performance that local search methods have on the problem. The autocorrelation length conjecture (see below) claims that the higher the value of ξ the lower the number of local optima and, as a consequence, the better could be the performance of a local search method. The definition of ξ is based on the autocorrelation function ρ proposed by Weinberger (1990), which is defined as

$$\rho(s) = 1 - \frac{\text{avg}\{(f(x) - f(y))^2\}_{\{x,y|d(x,y)=s\}}}{\text{avg}\{(f(x) - f(y))^2\}_{x,y}} \tag{16}$$

where the numerator is the average value of $(f(x) - f(y))^2$ over all solution pairs x and y that are at distance s, and the denominator is the average value of $(f(x) - f(y))^2$ over all solution pairs x and y. We say that two solutions are at distance s if there exists a sequence of solutions $x = x_0, x_2, \ldots, x_s = y$ such that $x_{i+1} \in N(x_i)$. The autocorrelation coefficient is then defined as $\xi = \frac{1}{1-\rho(1)}$.

It is possible to define another autocorrelation function. Let us consider a random walk $\{x_0, x_1, \ldots\}$ on the solution space such that $x_{i+1} \in N(x_i)$. The autocorrelation function r is defined as:

$$r(s) = \frac{\text{avg}\{f(x_t)f(x_{t+s})\}_{x_0,t} - \text{avg}\{f(x_t)\}_{x_0,t}^2}{\text{avg}\{f(x_t)^2\}_{x_0,t} - \text{avg}\{f(x_t)\}_{x_0,t}^2} \tag{17}$$

where the averages are computed over all the starting solutions x_0 and all the solutions in the sequence. The autocorrelation functions ρ and r are different in general, but they have the same value for $s = 1$, that is, $\rho(1) = r(1)$ (Angel and Zissimopoulos 2000a).

Stadler (1995) proved that if $f = \sum_i a_i \phi_i$ is a Fourier expansion of f in a landscape, then the autocorrelation function of f is given by

$$r(s) = \sum_{i \neq 0} \frac{a_i^2}{\sum_{j \neq 0} a_j^2} \left(1 - \frac{k_i}{d} \right)^s \tag{18}$$

where k_i is the characteristic constant associated to the elementary function ϕ_i. In particular, for an elementary landscape $r(s) = (1 - k/d)^s$, and the autocorrelation coefficient is $\xi = d/k$. For a general landscape (elementary or not) we have the following result

$$\rho(1) = r(1) = \frac{\sum_{i \neq 0} a_i^2 \left(1 - \frac{k_i}{d} \right)}{\sum_{j \neq 0} a_j^2} = 1 - \frac{\sum_{i \neq 0} a_i^2 \frac{k_i}{d}}{\sum_{j \neq 0} a_j^2}$$

and the autocorrelation coefficient can be computed as

$$\xi = \frac{d \sum_{j \neq 0} a_j^2}{\sum_{i \neq 0} a_i^2 k_i} \tag{19}$$

The sum of the squared Fourier coefficients a_j^2 associated to the same characteristic constant k_i is $|X|(\overline{f_i^2} - \overline{f_i}^2)$, where f_i is the sum of all the elementary components $a_i \phi_i$ with the same characteristic constant k_i and the overline represents the average over the entire search space X. The sum of the squared Fourier coefficients a_j^2 with $j \neq 0$ is $|X|(\overline{f^2} - \overline{f}^2)$. Introducing these two expressions in (19) we can write:

$$\xi = \frac{d|X|(\overline{f^2} - \overline{f}^2)}{\sum_{i \neq 0} |X|(\overline{f_i^2} - \overline{f_i}^2) k_i} = \left(\frac{\sum_{i \neq 0} |X|(\overline{f_i^2} - \overline{f_i}^2) k_i}{d|X|(\overline{f^2} - \overline{f}^2)} \right)^{-1}$$
$$= \left(\frac{\sum_{i \neq 0} (\overline{f_i^2} - \overline{f_i}^2) k_i}{d(\overline{f^2} - \overline{f}^2)} \right)^{-1} = \left(\sum_{i \neq 0} W_i \frac{k_i}{d} \right)^{-1} \tag{20}$$

where the values W_i are defined as

$$W_i = \frac{\overline{f_i^2} - \overline{f_i}^2}{\overline{f^2} - \overline{f}^2} \tag{21}$$

For the 0-1 UQO we get the following expression for the autocorrelation coefficient

$$\xi = \left(W_2 \frac{2}{n} + W_4 \frac{4}{n} \right)^{-1} \tag{22}$$

In the following we will reduce the computational complexity required to compute such a measure. Let us simplify (22) to make the computation more efficient. First, we take into account that, according to the definition of the W_i constants, it must hold

that $W_2 + W_4 = 1$. Then, replacing W_4 by $1 - W_2$ in (22) and making some calculations we get:

$$\xi = \frac{n}{2(2 - W_2)} \tag{23}$$

which depends only on W_2. According to (21), W_2 can be in turn computed as:

$$W_2 = \frac{\overline{f_2^2}}{\overline{f^2} - \overline{f}^2} \tag{24}$$

where we took into account that $\overline{f_2} = 0$. Since $\overline{f_4} = 0$, we have $\overline{f} = f_0$ and the only terms that we need to compute in (24) are $\overline{f_2^2}$ and $\overline{f^2}$. Using (10) we can write:

$$\sum_{x \in X} f_2^2 = \sum_{x \in X} \left(-\frac{1}{4} \sum_{i=1}^{n} v_i \psi_i \right)^2 = \frac{1}{16} \sum_{x \in X} \sum_{i,j=1}^{n} v_i v_j \psi_i \psi_j = \frac{1}{16} \sum_{i,j=1}^{n} v_i v_j \left(\sum_{x \in X} \psi_i \psi_j \right) \tag{25}$$

The inner sum in the last member can be simplified if we take into account the properties of the Walsh functions. In particular, since $\psi_i^2 = 1$, if $i = j$ then $\sum_{x \in X} \psi_i \psi_j = |X|$. On the other hand, $\psi_i \psi_j = \psi_{ij}$ when $i \neq j$ and $\overline{\psi_{ij}} = 0$. This implies $\sum_{x \in X} \psi_i \psi_j = 0$ when $i \neq j$. All the previous considerations allows us to write (25) as

$$\sum_{x \in X} f_2^2 = \frac{|X|}{16} \sum_{i=1}^{n} v_i^2 \tag{26}$$

and the value of $\overline{f_2^2}$ is:

$$\overline{f_2^2} = \frac{1}{|X|} \sum_{x \in X} f_2^2 = \frac{1}{16} \sum_{i=1}^{n} v_i^2 \tag{27}$$

The previous expression can be computed in $O(n)$. In order to compute $\overline{f^2}$ we use (3) and write:

$$\sum_{x \in X} f^2 = \sum_{x \in X} \left(\sum_{i,j=1}^{n} q_{ij} x_i x_j \right)^2 = \sum_{x \in X} \sum_{i,j=1}^{n} \sum_{i',j'=1}^{n} q_{ij} q_{i'j'} x_i x_j x_{i'} x_{j'} =$$

$$= \sum_{i,j=1}^{n} \sum_{i',j'=1}^{n} q_{ij} q_{i'j'} \left(\sum_{x \in X} x_i x_j x_{i'} x_{j'} \right) = \sum_{i,j=1}^{n} \sum_{i',j'=1}^{n} q_{ij} q_{i'j'} t(\{i, j, i', j'\}) \tag{28}$$

where we introduced a new function $t : \mathcal{P}([1, n]) \rightarrow \mathbb{N}$ defined by:

$$t(S) = \sum_{x \in X} \prod_{s \in S} x_s \tag{29}$$

Although we have now an expression for $\sum_{x \in X} f^2$, this expression is not practical, since it includes a factor, t, that requires a summation over all the elements of X. In the following we are going to simplify the expression of t to make the computation feasible.

We can observe that t is, in fact, a counting function. It is counting the number of elements in X that fulfill a given condition. Let us rewrite the definition of t as:

$$t(S) = \sum_{x \in X} True \left(\bigwedge_{s \in S} x_s = 1 \right) \tag{30}$$

where $True$ is a function that maps a Boolean value to $\{0, 1\}$. It is 1 if the Boolean expression is true and 0 if it is false. The function t counts the number of solutions (elements in X) that fulfills the condition $\bigwedge_{s \in S} x_s = 1$, that is, in all the positions of the solution contained in S the solution takes value 1. Using this characterization of t we can find an alternative expression for efficiently computing its value. The number of solutions fulfilling the previous condition is $2^{n-|S|}$ and then $t(S) = 2^{n-|S|}$. Using this definition we can rewrite (28) in the following way:

$$\sum_{x \in X} f^2 = \sum_{i,j=1}^{n} \sum_{i',j'=1}^{n} q_{ij} q_{i'j'} 2^{n - |\{i,j,i',j'\}|} \tag{31}$$

and $\overline{f^2}$ is:

$$\overline{f^2} = \sum_{i,j=1}^{n} \sum_{i',j'=1}^{n} \frac{q_{ij} q_{i'j'}}{2^{|\{i,j,i',j'\}|}} \tag{32}$$

The previous expression can be computed in polynomial time $O(n^4)$. With the values computed by (32), (27) and (9), and using Eq. (24) it is possible to calculate W_2. With W_2 it is possible to compute ξ using (23).

With the help of (18) and the constants W_2 and W_4, we can also express the autocorrelation function $r(s)$ of the UQO as:

$$r(s) = W_2 \left(1 - \frac{2}{n} \right)^s + W_4 \left(1 - \frac{4}{n} \right)^s \tag{33}$$

Sutton et al (2009) showed how the autocorrelation function $r(s)$ can be computed for all the pseudoboolean functions using the bit-flip neighborhood. The previous expression is just a particular case of the Sutton's expression.

Another measure of ruggedness is the *autocorrelation length* (García-Pelayo and Stadler 1997) whose definition is based on the autocorrelation function r as:

$$\ell = \sum_{s=0}^{\infty} r(s) = d \sum_{i \neq 0} \frac{W_i}{k_i} \tag{34}$$

In 0-1 UQO the autocorrelation length is:

$$\ell = n \left(\frac{W_2}{2} + \frac{W_4}{4} \right) = \frac{n(1 + W_2)}{4} \tag{35}$$

We can observe in Figure 3 that both autocorrelation measures, (23) and (35), increase with W_2 and their values are between $n/4$ when $W_2 = 0$ and $n/2$ when $W_2 = 1$. However, the curve of ℓ is linear while the one of ξ is non-linear. The autocorrelation
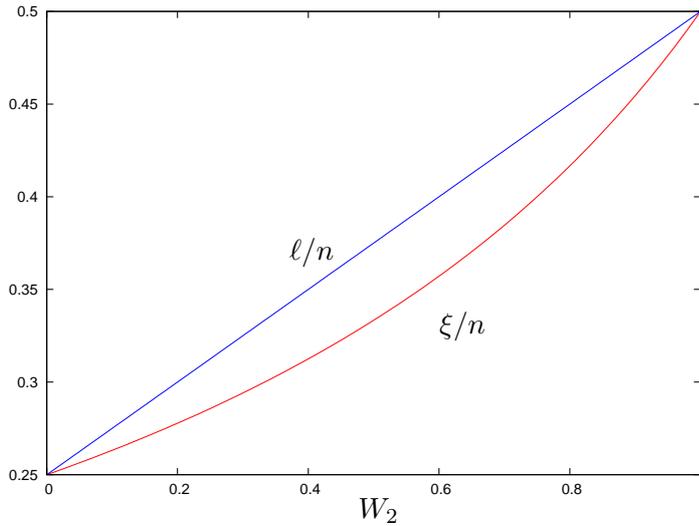
Fig. 3 Value of ξ/n and ℓ/n against W_2.

length is specially important in optimization because of the *autocorrelation length conjecture*, which claims that in many landscapes the number of local optima M can be estimated by the expression (Stadler 2002):

$$M \approx \frac{|X|}{|X(x_0, \ell)|} \tag{36}$$

where $X(x_0, \ell)$ is the set of solutions reachable from x_0 in ℓ or less local movements. The previous expression is not exact, but an approximation. It can be useful to compare the estimated number of local optima in two instances of the same problem. In effect, for a given problem in which (36) is valid, the higher the value of ℓ (or ξ) the lower the number of local optima. In a landscape with a low number of local optima, a local search strategy can *a priori* find the global optimum using less steps. This phenomenon has been empirically observed for the Quadratic Assignment Problem (QAP) by Angel and Zissimopoulos (2000b).

In our particular case, the set $X(x_0, \ell)$ is the set of solutions at Hamming distance ℓ or less from x_0. The cardinality of the previous set does not depend on the initial solution considered x_0, but it only depends on ℓ and is given by:

$$|X(x_0, \ell)| = \sum_{i=0}^{\ell} \binom{n}{i} \tag{37}$$

and the estimation for M is

$$M \approx \frac{2^n}{\sum_{i=0}^{\ell} \binom{n}{i}} \tag{38}$$

In order to check the autocorrelation length conjecture in UQO we have randomly generated 1650 UQO instances using the instance generator by Palubeckis (2000). The size of the instances varies between $n = 10$ and $n = 20$. The density δ of the instances

n	10	11	12	13	14	15
ρ	-0.5073	-0.5143	-0.5563	-0.4588	-0.4530	-0.5352
n	16	17	18	19	20	
ρ	-0.4784	-0.5625	-0.5786	-0.5163	-0.4780	

Table 1 Spearman correlation coefficient ρ for the number of local optima and the autocorrelation length.

is taken from the set $\{10, 30, 50, 70, 90\}$. For each pair n and δ, 30 random instances were generated and we computed their autocorrelation length ℓ. We also computed the number of local optima (minima) of each instance by complete enumeration of the search space.
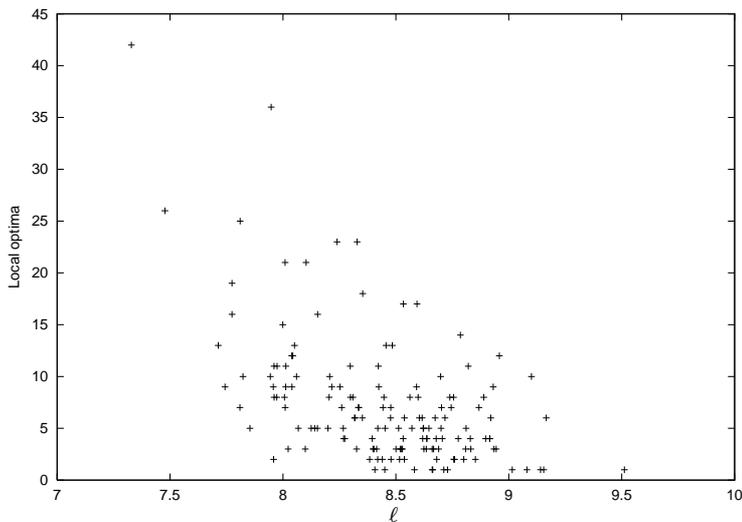

Fig. 4 Number of local optima against the autocorrelation length ℓ for instances with $n = 20$.

In Figure 4 we plot the number of local optima against the autocorrelation length ℓ for all the instances of size $n = 20$. We can observe a clear trend: as the autocorrelation length increases the number of local optima decreases. The trend is the same in all the instances with different sizes (we omit their plots). We computed the Spearman correlation coefficient for all the plots and we obtained the results shown in Table 1. 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.

6 Conclusions

In this paper we have studied the 0-1 UQO from the point of view of the landscapes theory. With the help of a Walsh analysis of the problem, we have proven that UQO can be written as a sum of two elementary landscapes. The landscape decomposition of UQO has practical and theoretical applications. From a practical point of view, this

decomposition allows one to compute the average value of the objective function in the bit-flip neighborhood of a solution x using only the value of the elementary components in x. That is, instead of evaluating all the solutions in the neighborhood of x and computing the average of the obtained values, the elementary landscape decomposition offers the same value with a single evaluation of the two elementary components of the problem in x. This means an important reduction in the computational effort. This efficient average computation has been recently used to propose new search methods and operators (as we saw in Section 5) and it can be used in the future for the same purpose. Furthermore, the decomposition can be used to provide closed-form formulas of the expected objective value of a solution after the application of a given stochastic operation. Sutton et al (2011) propose such a closed-form formula for the traditional bit-flip mutation operator.

Thanks to the elementary landscape decomposition we are able to provide an efficient (polynomial time) exact closed-form formula for the autocorrelation coefficient and the autocorrelation length of the UQO. These autocorrelation measures can be used to understand why a local search method is better than another one for the problem.

Once we have established the theoretical analysis, our future work is to study the practical applications of all the theoretical results presented here. The improvement of search methods and the explanation of the behaviour of some search algorithms are just two examples of possible applications of the theoretical results. Our study serves also as a guide for new problems, in order to analyze their landscape decomposition in a similar way as we did with UQO in this paper.

References

Alidaee B, Kochenberger G, Ahmadian A (1994) 0-1 quadratic programming approach for the optimal solution of two scheduling problems. International Journal of Systems Science 25:401–408

Amini MM, Alidaee B, Kochenberger G (1999) New Ideas in Optimisation, McGraw-Hill, London, chap A Scatter Search Approach to Unconstrained Quadratic Binary Programs, pp 317–329

Angel E, Zissimopoulos V (1998) Autocorrelation coefficient for the graph bipartitioning problem. Theoretical Compuer Science 191:229–243

Angel E, Zissimopoulos V (2000a) On the classiffication of NP-complete problems in terms of their correlation coefficient. Discrete Applied Mathematics 99:261–277

Angel E, Zissimopoulos V (2000b) On the landscape ruggedness of the quadratic assignment problem. Theoretical Computer Sciences 263:159–172

Biyikoglu T, Leyold J, Stadler PF (2007) Laplacian Eigenvectors of Graphs. Lecture Notes in Mathematics, Springer-Verlag

Chartaire P, Sutter A (1994) A decomposition method for quadratic 0-1 minimization. Management Science 41(4):704–712

Codenotti B, Margara L (1992) Local properties of some NP-complete problems. Tech. Rep. TR 92-021, International Computer Science Institute, Berkeley, USA

Gallo G, Hammer P, Simeone B (1980) Quadratic knapsack problems. Mathematical Programming 12:132–149

García-Pelayo R, Stadler P (1997) Correlation length, isotropy and meta-stable states. Physica D: Nonlinear Phenomena 107(2-4):240–254

Garey MR, Johnson DS (1979) Computers and Intractability: A Guide to the Theory of NP-Completeness. W.H. Freeman

Glover F, Alidaee B, Rego C, Kochenberger G (2002) One-pass heuristics for large-scale unconstrained binary quadratic problems. European Journal of Operational Research 137(2):272–287

Grover LK (1992) Local search and the local structure of NP-complete problems. Operations Research Letters 12:235–243

Harary F (1953) On the notion of balanced of a signed graphs. Michingan Mathematical Journal 2(2):143–146

Heckendorn RB, Rana SB, Whitley LD (1998) Test function generators as embedded landscapes. In: Banzhaf W, Reeves CR (eds) FOGA, Morgan Kaufmann, pp 183–198

Helmberg C, Rendl F (1998) Solving quadratic (0,1)-problems by semidefinite programs and cutting planes. Mathematical Programming 82:291–315

Katayama K, Narihisa H (2001) Performance of simulated annealing-based heuristic for the unconstrained binary quadratic programming problem. European Journal of Operational Research 134:103–119

Krarup J, Pruzan A (1978) Computer aided layout design. Mathematical programming Study 9:75–94

Liu W, Wilkins D, Alidaee B (2005) A hybrid multi-exchange local search for unconstrained binary quadratic program. Tech. Rep. HCES-09-05, Hearin Center for Enterprise Science, University of Mississippi

Lu G, Bahsoon R, Yao X (2010) Applying elementary landscape analysis to search-based software engineering. In: Proceedings of the 2nd International Symposium on Search Based Software Engineering (SSBSE)

McBride RD, Yormack JS (1980) An implicit enumeration algorithm for quadratic integer programming. Management Science 26:282–296

Merz P, Freisleben B (2002) Greedy and local search heuristics for unconstrained binary quadratic programming. Journal of Heuristics 8(2):197–213

Merz P, Katayama K (2004) Memetic algorithms for the unconstrained binary quadratic programming problem. Biosystems 78(1-3):99–118

Palubeckis G (2000) An algorithm for construction of test cases for the quadratic assignment problem. Informatica (Lithuanian Academy of Sciences) 11(3):281–296

Palubeckis G, Tomkevièius A (2002) Grasp implementations for the uncostrained binary quadratic optimization problem. Information Technology and Control 24(3):14–20

Philips AT, Rosen JB (1994) A quadratic assignment formulation of the molecular conformation problem. Journal of Global Optimization 4:229–241

Stadler PF (1995) Toward a theory of landscapes. In: López-Peña R, Capovilla R, García-Pelayo R, Waelbroeck H, Zertruche F (eds) Complex Systems and Binary Networks, Springer-Verlag, pp 77–163

Stadler PF (1996) Landscapes and their correlation functions. Journal of Mathematical Chemistry 20:1–45

Stadler PF (2002) Biological Evolution and Statistical Physics, Springer, chap Fitness Landscapes, pp 183–204

Sutton AM, Whitley LD, Howe AE (2009) A polynomial time computation of the exact correlation structure of k-satisfiability landscapes. In: Proceedings of the 11th annual conference on Genetic and evolutionary computation, ACM, New York, NY, USA, pp 365–372

Sutton AM, Howe AE, Whitley LD (2010) Directed plateau search for MAX-k-SAT. In: Proceedings of Symposium on Combinatorial Search, Atlanta, GA, USA

Sutton AM, Whitley LD, Howe AE (2011) Computing the moments of k-bounded pseudo-boolean functions over hamming spheres of arbitrary radius in polynomial time. Theoretical Computer Science DOI 10.1016/j.tcs.2011.02.006, in press

Walsh JL (1923) A closed set of normal orthogonal functions. American Journal of Mathematics 45(1):5–24

Weinberger E (1990) Correlated and uncorrelated fitness landscapes and how to tell the difference. Biological Cybernetics 63(5):325–336

Whitley D, Sutton AM, Howe AE (2008) Understanding elementary landscapes. In: Proceedings of the 10th annual conference on Genetic and evolutionary computation, ACM, New York, NY, USA, pp 585–592