Understanding Elementary Landscapes

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

Department of Computer Science
Colorado State University
Fort Collins, CO 80523
{whitley,sutton,howe}@cs.colostate.edu

ABSTRACT

The *landscape* formalism unites a finite candidate solution set to a neighborhood topology and an objective function. This construct can be used to model the behavior of local search on combinatorial optimization problems. A landscape is *elementary* when it possesses a unique property that results in a relative smoothness and decomposability to its structure.

In this paper we explain elementary landscapes in terms of the expected value of solution components which are transformed in the process of moving from an incumbent solution to a neighboring solution. We introduce new results about the properties of elementary landscapes and discuss the practical implications for search algorithms.

Categories and Subject Descriptors

I.2.8 [**Artificial Intelligence**]: Problem Solving, Control Methods, and Search

General Terms

Theory

Keywords

Combinatorial Optimization, Local Search

1. INTRODUCTION

Grover [9] originally made the observation that the behavior of local search on some NP-Hard problems could be modeled using a difference equation: a discrete analogue of the continuous *wave equation* that describes wave propagation in physics. This discrete wave equation can be expressed as an eigendecomposition of the neighborhood transition matrix under a particular local search operator for a certain class of problems.

Stadler [12] named this class of problems "elementary landscapes" and energetically explored properties of these

landscapes. Except for several notable examples, little research and exploration has been conducted for some time on this class. There is also a perception that elementary landscapes are in fact quite complex, or at least that the tools needed to understand elementary landscapes are recondite.

This paper has two goals. One goal is to explain and understand elementary landscapes from a simpler perspective. In practice, the wave equation discovered by Grover is a consequence of the decomposability of *components* that make up a solution. In many cases, constraints in the relationships among these components within the neighborhood of an incumbent solution force Grover's equation to hold.

The other goal of this paper is to show that this simplified view permits new insights and intuitions about elementary landscapes. We illustrate this by introducing several new proofs about properties of elementary landscapes.

The remainder of the paper is organized as follows. In the next section, we explain the relationship between objective function components that hold on many well-studied problems that cause the wave equation to be obeyed. We introduce this relationship in terms of *expected value* of the evaluation of solutions selected uniformly at random from the neighborhood of a fixed point. We derive the wave equation for symmetric TSP using 2-opt.

In Section 3, we discuss some scalability implications of the wave equation and show that the expected change in value can always be bounded. In Section 4 we prove some properties about plateaus and local optima on elementary landscapes and discuss the implications for search algorithms. In Section 5 we conclude the paper.

2. ELEMENTARY LANDSCAPES

We introduce the wave equation in terms of the *expected value* of the objective function evaluation at a solution drawn randomly from the local search neighborhood.

For a given instance to a combinatorial problem, we have a set of all possible solutions X which we will refer to as the set of "candidate solutions." The nature of the candidate solution set is dependent on the problem to solve. We define the *objective function* f as a map $f : X \mapsto \mathbb{R}$ and without loss of generality assume we must find the element of X that minimizes f.

We define a *neighborhood operator* as a function N that maps elements of X to elements of its power set $N : X \mapsto \mathcal{P}(X)$. The neighborhood operator is determined by the heuristic search method in question. We will only examine neighborhood operators that apply to search processes that keep a single candidate solution. Population based searches

that allow *recombination*, i.e. sharing partial components of solutions, admit more complex structures [14].

In this paper we shall concern ourselves only with neighborhoods that are *symmetric* (i.e. $y \in N(x) \iff x \in N(y)$) and *regular* (i.e. $d = |N(x)|$ for all x and some constant d). Barnes et al. [2] have extended the notion of elementary landscapes to non-symmetric and non-regular neighborhoods.

The *landscape* for a combinatorial problem instance is defined by a triple (X, N, f): its candidate solution set, the neighborhood operator which imposes a connective structure on the candidate solution points, and the objective function that assigns a value to each point.

This landscape formalism was originally introduced by Wright [16] in the context of evolutionary dynamics but can be extended to any function over a discrete domain with a corresponding "neighborhood": the context in which we are interested here. A landscape is elementary when the objective function f is an eigenfunction of the Laplacian of the graph induced by the neighborhood operator [9, 12].

This property can be expressed more concretely as an equation relating the *expected value* of the objective function at a uniformly drawn random element from a candidate solution's set of neighbors. Throughout this paper, we will denote x as some fixed but arbitrary element of X and y as a element of $N(x)$ drawn *uniformly at random*. We will also denote as $\bar{f}$ the mean value of f over all solutions in X. On an elementary landscape we have

$$\mathbb{E}[f(y)] = f(x) + \frac{k}{d}(\bar{f} - f(x)) \qquad (1)$$

for some k which is fixed for the entire landscape. Since y is drawn uniformly at random, we also have

$$\mathbb{E}[f(y)] = \frac{1}{d} \sum_{z \in N(x)} f(z) \qquad (2)$$

In other words, the expected value of the objective function evaluation of a neighbor y is always equal to the average value of the evaluation over all solutions z in the neighborhood.

2.1 Intracomponents and intercomponents

Several interesting consequences arise from Equation (1). First, we investigate what relationships exist in practice that allow this wave equation to hold.

In most practical elementary landscapes studied, the objective function for a particular candidate solution can be written as a linear combination of a subset of a collection of *components*. Examples of solution components are *edge weights* in graph problems such as TSP or Min-Cut partitioning, or *collisions* in constraint satisfaction problems such as graph coloring. In such problems, a candidate solution $x \in X$ specifies this subset for inclusion in the sum. Let C be a set of real valued components ($|C|$ is polynomial in problem size). Then there exists $C_x \subset C$ such that

$$f(x) = \sum_{c \in C_x} c$$

We refer to the set C_x as the *intracomponents* of a solution x, and the set $C - C_x$ as the *intercomponents* of x.

When a local search algorithm "moves" from an incumbent solution x to a neighboring solution $y \in N(x)$, an exchange of these components is made. In particular, a subset of the intracomponents $c_{out} \subset C_x$ is removed and a subset of the intercomponents $c_{in} \subset C - C_x$ is added. In other words

$$f(y) = f(x) - \sum_{c \in c_{out}} c + \sum_{c \in c_{in}} c$$

If we fix x arbitrary and let y be a uniform random "move" or "mutation" obtained by one application of the neighborhood operator acting on x, we can compute an *expected value* for the objective function $f(y)$ at y.

$$\mathbb{E}[f(y)] = \mathbb{E}\left[f(x) - \sum_{c \in c_{out}} c + \sum_{c \in c_{in}} c\right]$$

$$= f(x) - \mathbb{E}\left[\sum_{c \in c_{out}} c\right] + \mathbb{E}\left[\sum_{c \in c_{in}} c\right]$$

We compute the expected value by uniformly sampling all the intracomponents for potential removal and all the intercomponents for potential addition. A landscape thus satisfies Equation (1) when the following relationship holds for all $x \in X$.

$$\mathbb{E}\left[\sum_{c \in c_{in}} c\right] - \mathbb{E}\left[\sum_{c \in c_{out}} c\right] = \frac{k}{d}(\bar{f} - f(x))$$

The existence of components that satisfy this relationship is a *sufficient but not necessary* condition for a landscape to be elementary.

2.2 Case study: TSP under 2-opt

A number of NP-hard optimization problems have been shown to satisfy Grover's wave equation. As with local search, the move operator is critical. The following is an inexhaustive list of elementary landscapes.

1. The symmetric Traveling Salesman Problem under 2-exchange [9], 2-opt and 3-exchange, [5].

2. The antisymmetric Traveling Salesman Problem under 2-opt and 2-exchange [13] and the weakly-symmetric Traveling Salesman Problem [11], and variants of the multiple Traveling Salesman Problem [6].

3. The Min-Cut Graph Partitioning problem with a simple exchange of two vertices across the cut [9].

4. Graph Coloring with the neighborhood being a change in color at any vertex in the graph [9].

5. Weight Partitioning over N objects. Each object is assigned a sign $S_i \in (-1, 1)$ with the neighborhood being a change in sign [9].

We will derive the expected value wave equation for the TSP under 2-opt and examine some of the consequences. Suppose $G(V, E)$ is a complete weighted undirected graph on n vertices (cities). Every vertex pair $v_i, v_j \in V$ has an edge $e_{i,j} \in E$. Let $w_{i,j}$ be the weight (or cost or distance) associated with edge $e_{i,j}$. The solution space X thus is comprised of all valid Hamiltonian cycles in G. In particular a given $x \in X$ specifies a permutation on $n - 1$ points $(x_1, x_2, \ldots, x_{n-1})$. Without loss of generality we fix the "home" city at n.

$$f(x) = \left(\sum_{i=1}^{n-2} w_{x_i, x_{i+1}}\right) + w_{x_{n-1}, n} + w_{n, x_1}$$
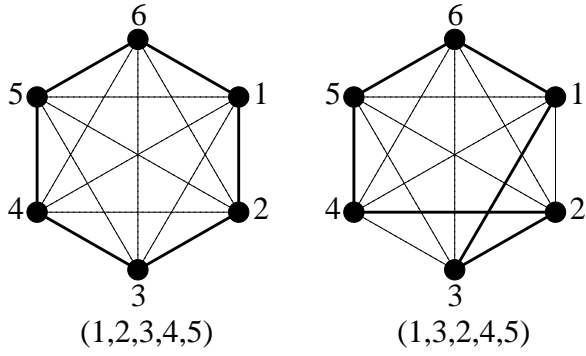
Figure 1: A candidate tour (left) and one of its 2-opt neighbors (right) obtained by deleting edges $(1,2)$ and $(3,4)$ and adding edges $(1,3)$ and $(2,4)$. The corresponding permutation representation is shown below each graph.

This means that the evaluation function is decomposable into the costs associated with every possible edge. The component set C is exactly the set of real-valued edge weights $w_{i,j}$. The intracomponents of a solution x are those edge weights used in the calculation of $f(x)$; the intercomponents are the edge weights not in the tour specified by x.

In Figure 1 a 6 city TSP is shown. The tour $x = (1\ 2\ 3\ 4\ 5\ 6)$ is shown on the left and the tour $y = (1\ 3\ 2\ 4\ 5\ 6)$ is shown on the right. The edges $\{e_{1,2},\ e_{2,3},\ e_{3,4},\ e_{4,5},\ e_{5,6},\ e_{6,1}\}$ are the intracomponents of x. All possible edges that are not intracomponents of x are by definition intercomponents of x. These are the lighter interior dashed lines in the figure.

To compute $\bar{f}$, the average over all possible solutions, we also need to know the probability of each edge occurring in any particular solution. In this case, that probability is uniform. Since a tour length is n and there are $n(n-2)/2$ possible edges it follows that

$$\bar{f} = \sum_{e_{i,j} \in E} w_{i,j} \frac{n}{n(n-1)/2}$$

and by simple algebra

$$\bar{f} = \frac{2}{n-1} \sum_{e_{i,j} \in E} w_{i,j}$$

Now, given an incumbent solution x we need to compute the expected cost over all of the neighbors of x. The deletion of intracomponents of x and the addition of intercomponents of x in effect "connect" x to all of its neighborhood.

To calculate the expected value of a neighbor of x we examine the subset of intracomponents that must be removed (c_{out}) and the subset of intercomponents that must be added (c_{in}). All intracomponents of x are uniformly removed when constructing the set of neighbors of x; all of the intercomponents of x are then uniformly added in when constructing the set of neighbors of x.

Under 2-opt, exactly two edges change in each neighbor. The construction of each neighbor of x therefore requires deleting two edges from x. Thus, components (edges) are deleted with probability $2/n$. Note that the sum of all the intracomponents is simply the tour length which is given by $f(x)$. Then the *expected value* of the sum of intracomponents

removed is

$$\mathbb{E}\left[\sum_{c \in c_{out}} c\right] = \frac{2}{n} f(x)$$

The entire cost matrix has $n(n-1)/2$ components and x has n components. This means that the number of intercomponents is given by

$$(n(n-1)/2) - n = n(n-3)/2$$

Furthermore, the sum of all the intercomponents is simply the sum of all the edges in the graph G minus the sum of all the intracomponents $f(x)$

$$\sum_{e_{i,j} \in E} w_{i,j} - f(x)$$

Since two intercomponents must be added, we have the expected value of the sum of intercomponents added

$$\mathbb{E}\left[\sum_{c \in c_{in}} c\right] = \frac{2}{n(n-3)/2}\left(\sum_{e_{i,j} \in E} w_{i,j} - f(x)\right)$$

So, fixing the incumbent solution at an arbitrary x, the expected value of a randomly selected neighbor $y \in N(x)$ can be written as

$$\mathbb{E}[f(y)] = f(x) - \mathbb{E}\left[\sum_{c \in c_{out}} c\right] + \mathbb{E}\left[\sum_{c \in c_{in}} c\right] \qquad (3)$$

$$= f(x) - \frac{2}{n} f(x) + \frac{2}{n(n-3)/2}\left(\sum_{e_{i,j} \in E} w_{i,j} - f(x)\right)$$

Because the probabilities are uniform we can look at the construction of each neighbor as an independent random event. Thus, two edges are deleted from x and two edges are inserted to create a neighbor y. In reality, the edges that are deleted determine the edges that are inserted, but this can be ignored by looking at the neighborhood as an aggregate.

Now we derive Equation (1). Note that

$$\sum_{e_{i,j} \in E} w_{i,j} = \frac{n-1}{2} \bar{f}$$

and therefore rewriting Equation (3):

$$\mathbb{E}[f(y)] = f(x) - \frac{2}{n} f(x) + \frac{2}{n(n-3)/2}\left(\sum_{e_{i,j} \in E} w_{i,j} - f(x)\right)$$

$$= f(x) - \frac{2}{n} f(x) + \frac{2}{n(n-3)/2}\left(\frac{n-1}{2}\bar{f} - f(x)\right)$$

$$= f(x) - \frac{2}{n} f(x) + \frac{n-1}{n(n-3)/2}\bar{f} - \frac{2}{n(n-3)/2} f(x)$$

$$= f(x) - \frac{2(n-3)/2 + 2}{n(n-3)/2} f(x) + \frac{n-1}{n(n-3)/2}\bar{f}$$

$$= f(x) - \frac{n-1}{n(n-3)/2} f(x) + \frac{n-1}{n(n-3)/2}\bar{f}$$

$$= f(x) + \frac{n-1}{n(n-3)/2}(\bar{f} - f(x))$$

$$= f(x) + \frac{k}{d}(\bar{f} - f(x))$$

Thus we have Equation (1) with $k = n - 1$ and $d = n(n-3)/2$. Here d is, of course, by definition the neighborhood size for classic 2-opt.

Note that when $f(x) < \bar{f}$, the second term is always positive and $\mathbb{E}[f(y)]$ is always greater than $f(x)$. On the other hand, when $f(x) > \bar{f}$ the second term becomes negative and the expected value of a random neighbor is less than $f(x)$.

3. SCALABILITY

Since we can exactly compute $\mathbb{E}[f(y)]$ for a randomly selected neighbor y of x we can also ask a fundamental question about scalability. How does the difference between $f(x)$ and $\mathbb{E}[f(y)]$ scale with problem size? We can prove using simple algebra that the difference scales in a linear fashion. This is intuitive: for large problems, $f(x)$ and $\mathbb{E}[f(y)]$ are closer to each other relative to $\bar{f}$ and the change decreases linearly as a function of problem size. In any one problem, the difference between $f(x)$ and $\mathbb{E}[f(y)]$ grows smaller as we approach $\bar{f}$ and increases as we move away from $\bar{f}$.

To understand this better, assume that a TSP instance with n vertices is normalized by $\bar{f}/n$. In effect, this normalization is such that the average edge length is 1. Thus, the normalized average solution is $\frac{\bar{f}}{\bar{f}/n} = n$. This also lets us compute bounds on $f(x)$ and $\mathbb{E}[f(y)]$ relative to a normalized edge length of 1.

Assuming we are looking for the best (minimal) solutions, consider the case where

$$f(x) < \mathbb{E}[f(y)] < \bar{f}$$

To better understand the difference $\mathbb{E}[f(y)] - f(x)$ we will use a normalization of the cost matrix such that

$$\bar{f} = n = \sum_{e_{i,j} \in E} w_{i,j} \frac{n}{n(n-1)/2} \iff \frac{\sum_{e_{i,j} \in E} w_{i,j}}{n(n-1)/2} = 1$$

Or more concisely,

$$\bar{f} = n \iff \overline{w_{i,j}} = 1$$

We can then compute the difference between $\mathbb{E}[f(y)] - f(x)$ and obtain the following bound.

$$\mathbb{E}[f(y)] - f(x) = f(x) + \frac{n-1}{n(n-3)/2}(\bar{f} - f(x)) - f(x)$$
$$= \frac{n-1}{n(n-3)/2}(n - f(x))$$
$$= \frac{n-1}{n(n-3)/2}n - \frac{n-1}{n(n-3)/2}f(x)$$
$$= \frac{n-1}{(n-3)/2} - \frac{n-1}{(n-3)/2}\frac{f(x)}{n}$$
$$= 2\left(\frac{n-1}{n-3}\right)\left(1 - \frac{f(x)}{n}\right)$$

As $f(x)$ approaches zero relative to $\bar{f} = n$ the quantity $\mathbb{E}[f(y)] - f(x)$ is bounded by $2\left(\frac{n-1}{n-3}\right)$ (which for large n is approximately 2.0); there are of course real limits on how small $f(x)$ can be. Around $f(x) = n/2$, the normalized quantity $\mathbb{E}[f(y)] - f(x)$ is approximately 1.0. As $f(x)$ approaches the value $\bar{f} = n$ the difference between $\mathbb{E}[f(y)] - f(x)$ approaches 0.

Tables 1 and 2 report numerical calculations of $f(x)$ and $\mathbb{E}[f(y)]$ and their differences varying both $f(x)$ and n according to Equation (1).

n	$f(x)$	$\mathbb{E}[f(y)]$	$\mathbb{E}[f(y)] - f(x)$
10000	50	51.990398	1.99039811
10000	500	501.90038	1.90038011
10000	5001	5002.00000	1.00000002
10000	9000	9000.20004	0.20004001
10000	9900	9900.02000	0.02000400
10000	9990	9990.00200	0.00200040
10000	9999	9999.00020	0.00020004

Table 1: Calculations for $\mathbb{E}[f(y)]$ and $f(x) < n$ varying $f(x)$.

n	$f(x) = 0.3\bar{f}$	$\mathbb{E}[f(y)]$	$\mathbb{E}[f(y)] - f(x)$
100	30	31.428865	1.42886597
1000	300	301.40280	1.40280842
10000	3000	3001.4002	1.40028008
100000	30000	30001.400	1.40002800
1000000	300000	300001.40	1.40000280
10000000	3000000	3000001.4	1.40000028

Table 2: Calculations for $\mathbb{E}[f(y)]$ and $f(x) < n$ varying n.

To illustrate the effect using real data, we sampled solutions from symmetric instances obtained from the TSPLIB library maintained by Gerhard Reinelt at Heidelberg University[1]. On each instance we computed a normalization factor q as follows.

$$q = \frac{n(n-1)}{2\sum_{e_{i,j} \in E} w_{i,j}}$$

We define the normalized objective function as the weighted tour length sum multiplied by the normalization factor q. Clearly, the mean value of the normalized instances is n. On each normalized instance we sampled 1000 points *below* $\bar{f}$ and calculated the objective function value of the sampled point, the theoretical expected value as given by Equation (1) and exactly one uniform random sample of the neighborhood of each sampled point. We report the scatter plot of the theoretical prediction vs. the actual sampled neighbor value for five selected instances in Figure 2. Each selected instance has a different size n. In Figure 3 we report the expected *difference* vs. the sampled difference on these five instances. Figure 4 shows this difference plot for every (normalized) symmetric TSP instance in the TSPLIB with 100 points sampled per instance. The expected value of solution difference is bounded below by zero and above by 2 (in the limit as $n \to \infty$). This is marked on the plots by dotted lines.

In these experiments, we are observing a sampling process from a distribution that follows the theoretical expected value given by Equation (1). The plots show us the variance in the random variables we are sampling.

4. DISCUSSION

There are a number of interesting observations about this class of landscapes that impact local search. In our TSP case study, we saw that $0 < \frac{k}{d} = \frac{(n-1)}{n(n-1)/2} < 1$. Does this
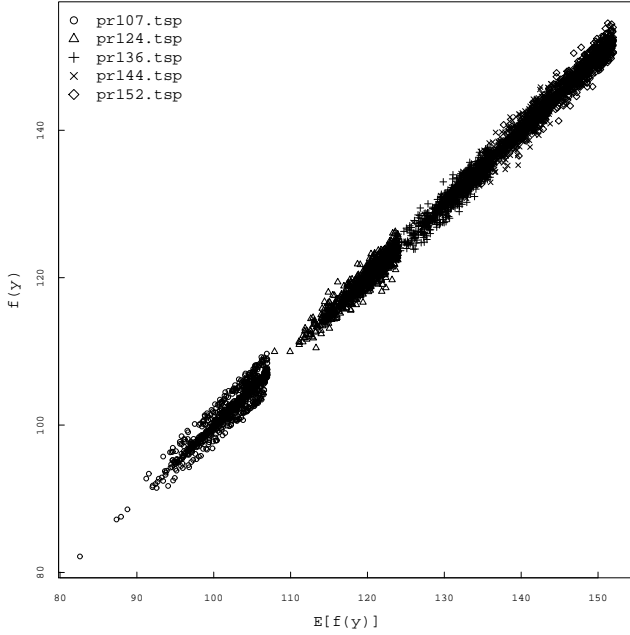

Figure 2: **Theoretical prediction of neighbor value vs. sampled neighbor value for 1000 sampled sub-mean points on five (normalized) instances in TSPLIB.**
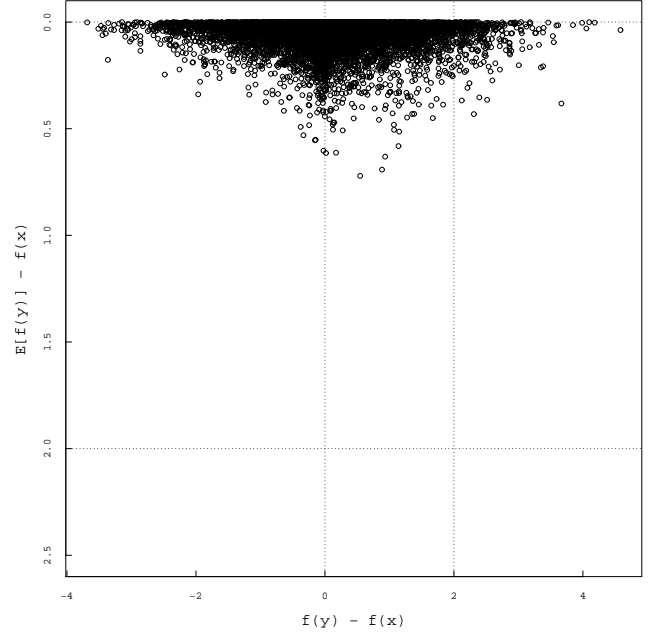

Figure 4: **Expected difference vs. sampled difference of sub-mean points and neighbors on all normalized instances in TSPLIB. Dotted lines represent bounds on expected value. Expected value is bounded above by 2 and below by 0.**
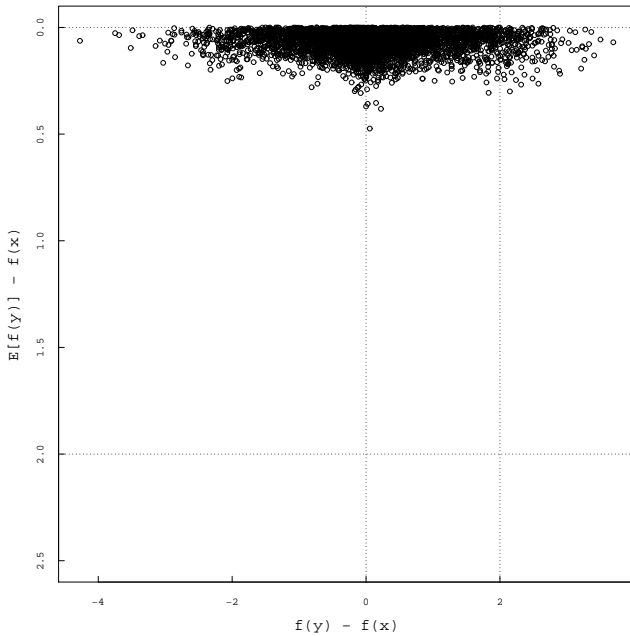

Figure 3: **Expected difference vs. sampled difference of sub-mean points and neighbors on five normalized instances. Dotted lines represent bounds on expected value. Expected value is bounded above by 2 and below by 0.**

hold in general for other problem classes? Barnes et al. [2] define two families of elementary landscapes with respect to constraints on the possible eigenvalues of the graph Laplacian. *Smooth* elementary landscapes obey the above constraint such that the value $\frac{k}{d}$ lies between 0 and 1. On the other hand, on *rugged* elementary landscapes the value of $\frac{k}{d}$ are not constrained to the unit interval, and some rather surprising results are a consequence. We are unaware of any well-studied neighborhood operator and problem class in the NP-hard set that corresponds to a rugged landscape. Thus our results apply only to smooth landscapes.

4.1 Plateaus and local optima

The following observations were made by Codenotti and Margara [5].

1. If $f(x) < \bar{f}$ then $f(x) < \mathbb{E}[f(y)] < \bar{f}$

2. If $f(x) = \mathbb{E}[f(y)]$ then $f(x) = \bar{f} = f(y)$

3. If $f(x) > \bar{f}$ then $f(x) > \mathbb{E}[f(y)] > \bar{f}$

These results underscore Grover's [9] observation that all local minima lie below the average function value of the search space. Furthermore, the results state that certain types of *plateaus* cannot exist on elementary landscapes.

A plateau is a set P of candidate solutions in X such that for all $a, b \in P$, $f(a) = f(b)$ and there is a path $(a = x_1, x_2, \ldots, x_k = b)$ such that $x_{i+1} \in N(x_i)$. Plateaus (also known as neutral networks) are structural features that arise in many combinatorial problems [8, 10, 1, 3]. Plateaus pose an interesting challenge to local search since they preclude "gradient" information about improving movement.

As Codenotti and Margara [5] originally pointed out, on an elementary landscape, if all neighbors of a candidate solution share the evaluation of that solution, then we say the landscape is *flat*: every candidate solution belongs to the same plateau (condition 2 above). However, the wave equation imposes some constraints on plateaus in non-degenerate cases as well.

Consider a solution x that belongs to a plateau P. Then every solution on P will have the same neighborhood expected value. This is easy to see since all solutions in P have an objective function evaluation equal to $f(x)$ and obey Equation (1).

LEMMA 1. *For a plateau P on a (non-flat) elementary landscape, if $x \in P$ has only equal and disimproving neighbors, then there cannot exist a solution $z \in P$ with only equal and improving neighbors.*

PROOF. Let $x, z \in P$ for some plateau P. Suppose for contradiction that x has only (both) equal and disimproving neighbors and z has only (both) equal and improving neighbors.

Let y_x and y_z be neighbors drawn uniformly at random from the neighborhood of x and z respectively. Since $f(x) = f(z)$ (by definition of plateau), we must have $\mathbb{E}[f(y_x)] = \mathbb{E}[f(y_z)]$ since Equation (1) holds. If we assume minimization, the average value of the neighbors of x must be strictly greater than than the average value of the neighbors of z which contradicts the expected value of a uniform random selection is equal for both neighborhoods. $\square$

Note that Lemma 1 does not prevent a solution on the same plateau as x from having *both* improving and disimproving neighbors. However, in this case, the evaluation of these neighbors must have the same expected value as those in the neighborhood of x.

We define a local minimum as a solution x_{min} such that $f(y) \geq f(x_{min})$ for all $y \in N(x_{min})$. A local maximum is defined analogously. It is important to point out that this definition is distinct from other definitions of local extrema in landscapes that contain plateaus (e.g. Frank et al. [8]). Grover [9] showed that, on elementary landscapes, if x_{min} is a local minimum and x_{max} is a local maximum, then $f(x_{min}) \leq \bar{f} \leq f(x_{max})$.

We assume that, for all $x \in X$, all $f(x) > 0$. If this does not hold, the values of the objective function can be shifted by a constant value without violating Equation (1).

Let x_{min} be a local minimum. If we sample $y \in N(x_{min})$ uniformly at random, we can express the expected value of y in terms of an average over the elements of $N(x_{min})$

$$\mathbb{E}[f(y)] = \frac{1}{d}(S_{eq} + S_{gr}) \tag{4}$$

where S_{eq} is the sum of the value of equal or *neutral* moves: those $y \in N(x_{min})$ such that $f(y) = f(x_{min})$. S_{gr} is the sum of the value of the neighbors of x_{min} that are strictly greater in objective evaluation: those $y \in N(x_{min})$ such that $f(y) > f(x)$.

LEMMA 2. *Let a and b be two solutions on a plateau P with no improving neighbors. Let $n_{eq}(a)$ and $n_{eq}(b)$ denote the number of equal neighbors of a and b respectively. Without loss of generality, suppose $n_{eq}(a) \geq n_{eq}(b)$. Then, assuming minimization, the sum of all disimproving moves in the neighborhood of b is greater than or equal to the sum of all disimproving moves in the neighborhood of a.*

PROOF. We first derive an expression for the number of equal moves in an arbitrary solution x_{min}. Note that at x_{min}, $S_{eq} = n_{eq}(x_{min})f(x_{min})$. By Equations (1) and (4) we have

$$\frac{1}{d}(S_{eq} + S_{gr}) = f(x_{min}) + \frac{k}{d}(\bar{f} - f(x_{min}))$$

$$S_{eq} + S_{gr} = df(x_{min}) + k(\bar{f} - f(x_{min}))$$

$$n_{eq}(x_{min})f(x_{min}) + S_{gr} = df(x_{min}) + k(\bar{f} - f(x_{min}))$$

$$n_{eq}(x_{min}) = d + \frac{k(\bar{f} - f(x_{min}))}{f(x_{min})} - \frac{S_{gr}}{f(x_{min})}$$

Now consider the difference of neutral moves among a and b.

$$n_{eq}(a) - n_{eq}(b) = \left[d + \frac{k(\bar{f} - f(a))}{f(a)} - \frac{S_{gr}^a}{f(a)} \right] -$$
$$\left[d + \frac{k(\bar{f} - f(b))}{f(b)} - \frac{S_{gr}^b}{f(b)} \right]$$
$$= \frac{S_{gr}^b - S_{gr}^a}{f(a)}$$

Where S_{gr}^a and S_{gr}^b denote the sum over all neighbors with strictly greater objective evaluation for a and b respectively. Since we assumed that $f(x)$ for all $x \in X$ is positive and that $n_{eq}(a) - n_{eq}(b)$ is non-negative, we have

$$S_{gr}^b \geq S_{gr}^a$$

and this proves the result. $\square$

A similar result holds for all local maxima on plateaus.

4.2 Implications for search

An interesting consequence of the expected value version of the wave equation is that a local search algorithm on an elementary landscape can always compute the expected value of a random solution in its neighborhood without expanding a single solution in this neighborhood.

Furthermore, we can expand a *partial* neighborhood and predict the expected value of the remaining neighborhood. Suppose $M \subseteq N(x)$ is a subset of the neighborhood of a point x. Then the expected value of a point r uniformly drawn from the *remaining* neighbors in $N(x)$ can be written as

$$\mathbb{E}[f(r)] = \frac{1}{d - |M|} \left(\sum_{z \in N(x)} f(z) - \sum_{z \in M} f(z) \right)$$

$$= \frac{1}{d - |M|} \left(d \left(f(x) + \frac{k}{d}(\bar{f} - f(x)) \right) - \sum_{z \in M} f(z) \right)$$

by Equations (2) and (1). Let the expected *improvement* be

$$D(r; x) = \mathbb{E}[f(r)] - f(x)$$
$$= \frac{1}{d - |M|} \left(d \left(f(x) + \frac{k}{d}(\bar{f} - f(x)) \right) - \sum_{z \in M} f(z) \right)$$
$$- f(x)$$

This is the expected difference in value if we move (or mutate) at uniform random to an element in the remaining neighbor set $N(x) - M$.

Now suppose x_1 and x_2 are two solutions in X. Furthermore, suppose M_1 and M_2 are two partial expansions
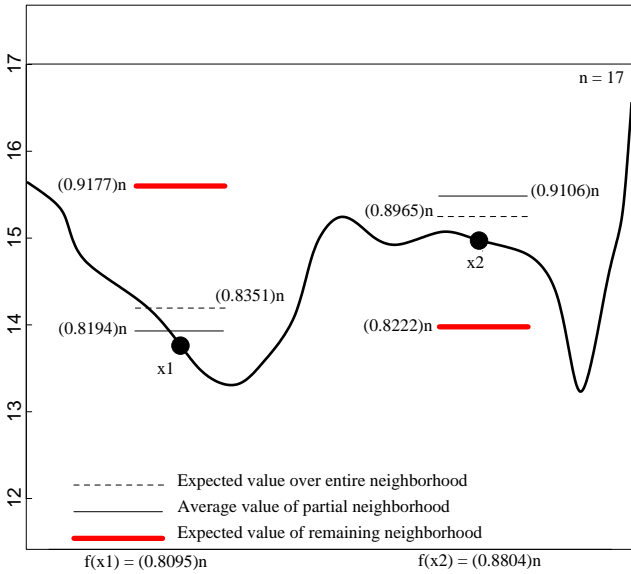
Figure 5: **A stylized sketch of partial and whole neighborhood relationships for two solutions x_1 and x_2. The average evaluation of the partial neighborhood of x_1 is worse than x_2; the expected evaluation of remaining neighbors of x_1 are *better* than those of x_2. Numerical data were generated using two random solutions from (normalized) 17-city TSP `gr17.tsp` instance.**

of the neighborhoods of x_1 and x_2. If r_1 and r_2 are randomly drawn elements from the *remaining* neighborhoods $N(x_1) - M_1$ and $N(x_2) - M_2$, then we can compute the expected values of r_1 and r_2 above and the expected improvement $D(r_1; x_1)$ and $D(r_2; x_2)$ for each.

It is clearly more promising to continue expanding the neighborhood corresponding to the better of the two expected improvements (i.e. smallest, if we are minimizing): this is more likely to yield a better change in value. This idea is illustrated in Figure 5 where the partial neighborhood that results in an average value *worse* than the expected value of the entire neighborhood yields a better expected value over the remaining neighborhood.

On an elementary landscape, neighborhoods can be considered rather localized. From the *component* point of view, neighbors are very similar in structure to the current incumbent solution. Neighbors also have similar objective function values (on average). The neighborhood does not support any mechanism that allows sampling in more distant parts of the search space. This makes local search on these landscapes fundamentally myopic. It should be pointed out this is not true of all forms of local search. The neighborhoods used by Gray- and binary-coded representation of parameter optimization problems do a better job of systematically globally sampling in distant parts of the search space. These neighbors are used by many forms of genetic algorithms. Pattern Search is also better at systematically globally sampling in distance parts of the search space.

Stadler [13] showed that a landscape (with a symmetric neighborhood operator) is elementary if and only the time series generated by a random walk on the landscape using transitions defined by the neighborhood operator is an AR(1) process: an observation that was later generalized by Dimova et al. [7]. This observation, along with Grover's results about local extrema lying above or below the mean solution value, imply that elementary landscapes describe a class of relatively smooth and structured problems.

When compared to other conventional combinatorial problems, the TSP will be comparatively well-behaved. As Equation (1) predicts, candidate solutions in the TSP will have a constant relationship with the average neighborhood value (with respect to the mean). This is not necessarily a pervasive feature in commonly studied combinatorial problems.

Consider the mean-centered ratio between the evaluation of an arbitrary solution x and its average neighborhood value.

$$R(x) = \frac{f(x) - \bar{f}}{(1/d)\sum_{y \in N(x)} f(y) - \bar{f}} \tag{5}$$

Suppose x lies on an elementary landscape. Equations (1) and (2) give us

$$R_{elem}(x) = \frac{f(x) - \bar{f}}{f(x) + \frac{k}{d}(\bar{f} - f(x)) - \bar{f}}$$
$$= \frac{d}{d - k}$$

Thus the ratio is constant for landscapes that obey Equation (1).

To contrast this relationship empirically with other combinatorial problems, we selected an instance of the permutation flowshop scheduling problem (FSP) under the shift neighborhood [15], and the asymmetric (ATSP) under the exchange neighborhood. On each landscape, we generated 100 random solutions and calculated the value of $R(x)$, Equation (5), for each solution x. The values are plotted in Figure 6 for `car3.fsp`: a 12×5 flowshop problem instance originally published by Carlier [4] and contained in the online OR-LIBRARY[2]. the `pr152.tsp` and `ali535.tsp` (symmetric) TSP instances and the `ft53.atsp` ATSP instance from TSPLIB. The ratio is constant for the TSP problems, but varies dramatically for the ATSP and FSP problems.

We can therefore assume that search algorithms that perform competitively on TSP are likely to be exploiting the partial decomposability of the landscape. On the one hand, this implies that such an algorithm may easily be adaptable to other well known elementary problems. On the other hand, search algorithms that are benchmarked on elementary problem classes may not generalize well to composite, non-elementary landscapes.

5. CONCLUSION

The elementary property of certain landscapes introduces several interesting constraints on the behavior of local search algorithms. Landscapes with this property tend to be relatively smooth when contrasted to other combinatorial optimization problems with well-studied local move operators.

In this paper we have explained the elementary property in terms of neighborhood sampling and observed some of its consequences on real world instances. The constraints imposed by the wave equation precludes the existence of certain plateau structures, and forces certain relationships between local optima with the same evaluation. The ele-

---

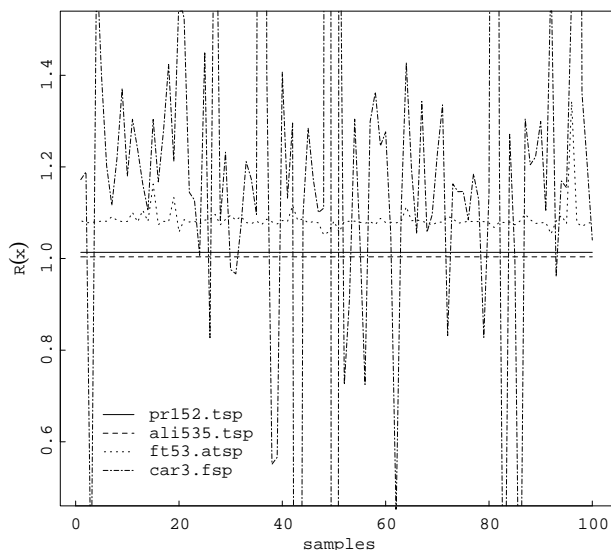[2]`http://people.brunel.ac.uk/~mastjjb/jeb/info.html`

Figure 6: Ratio $R(x)$ of mean-centered objective and mean-centered average of neighborhood objective for 100 random solutions each on two elementary TSP instances, an ATSP instance, and a flowshop scheduling instance.

mentary property also allows us to make predictions about the value of partial or full neighborhoods during search.

The Traveling Salesman Problem is widely used as a benchmark to testing new heuristic search algorithms. But, there are many properties of TSP that make it unlike many other combinatorial optimization problems. The wave equation and the associated characterization by constituent solution components presented in this paper make it clear that the TSP is in some sense partially decomposable. This allows for the partial update of the cost function. It also means that near optimal solutions can be found by methods that exploit the partial decomposability of the problem. Any method that heuristically exploits information about the individual components of the cost matrix may be a risk of over specialization. The results may not generalize well beyond the TSP, or at least beyond the family of elementary landscapes.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] L. Barbulescu, A. E. Howe, L. D. Whitley, and M. Roberts. Understanding algorithm performance on an oversubscribed scheduling application. *Journal of Artificial Intelligence Research*, 27:577–615, Dec 2006.

[2] J. W. Barnes, B. Dimova, and S. P. Dokov. The theory of elementary landscapes. *Applied Mathematics Letters*, 16(3):337–343, April 2003.

[3] L. Barnett. Netcrawling – optimal evolutionary search with neutral networks. In *Proceedings of the 2001 Congress of Evolutionary Computation (CEC2001)*, pages 30–37, Seoul, Korea, May 2001.

[4] J. Carlier. Ordonnancements a contraintes disjonctives. *R.A.I.R.O. Recherche operationelle/Operations Research*, 12:333–351, 1978.

[5] B. Codenotti and L. Margara. Local properties of some NP-complete problems. Technical Report TR 92-021, International Computer Science Institute, Berkeley, CA, 1992.

[6] B. Colletti and J. W. Barnes. Linearity in the traveling salesman problem. *Applied Mathematics Letters*, 13(3):27–32, April 2000.

[7] B. Dimova, J. Wesley Barnes, and E. Popova. Arbitrary elementary landscapes & ar(1) processes. *Appl. Math. Lett.*, 18(3):287–292, 2005.

[8] J. Frank, P. Cheeseman, and J. Stutz. When gravity fails: Local search topology. *Journal of Artificial Intelligence Research*, 7:249–281, 1997.

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

[10] C. M. Reidys and P. F. Stadler. Neutrality in fitness landscapes. *Applied Mathematics and Computation*, 117:321–350, 2001.

[11] A. Solomon, J. W. Barnes, S. P. Dokov, and R. Acevedo. Weakly symmetric graphs, elementary landscapes, and the tsp. *Appl. Math. Lett.*, 16(3):401–407, 2003.

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

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

[14] P. F. Stadler and G. P. Wagner. Algebraic theory of recombination spaces. *Evolutionary Computation*, 5(3):241–275, 1997.

[15] É. Taillard. Some efficient heuristic methods for the flow shop sequencing problem. *European Journal of Operations Research*, 47:65–74, 1990.

[16] S. Wright. The roles of mutation, inbreeding, crossbreeding, and selection in evolution. In *Proceedings of the Sixth Congress of Genetics*, volume 1, 1932.