T-79.300 Postgraduate Course in Theoretical Computer Science

Ruggedness and Neutrality

Pekka Isto

Introduction

Fitness landscapes have been proposed as a theoretical construct for investigating and explaining a range of phenomena from biology, physics and combinatorial optimization. A simple definition of a fitness landscape can be given as a triple *(V, E, f)*, where *V* is a set of vertices, *E* is a set of undirected edges between the vertices of *V*, and $f: V \rightarrow \mathbb{R}$ is an arbitrary function. The vertices of *V* are called configurations and two configurations are called neighbors, if there is an edge in *E* between the corresponding vertices of *V*. The function *f* is called the fitness function, and the values of the function on a particular vertex $x \in V$ is called the fitness of the vertex. In combinatorial optimization, landscape properties may prove to be useful in explaining why some problems are difficult and why some heuristics or exact algorithms perform better than others for a given problem. The Traveling Salesman Problem (TSP) is a prototypical problem in combinatorial optimization and one that has been a subject of extensive research. It is known that small modifications to the basic TSP problem can cause significant differences in the performance of alternative solution techniques [1]. The hypothesis is that the differences can be explained by the properties of the corresponding landscapes, in this particular case, by correlation between the neighboring configurations.

Ruggedness and neutrality are two related but independent characterizations of a landscape. Ruggedness is informally defined as the opposite of smoothness [2]. More formal definitions have been based on the correlation structure of the landscape and on the sizes and distribution of the local minima of the fitness function *f*. Neutrality describes the degree to which neighboring configurations of the landscape have the same fitness value.

Ruggedness and neutrality are just two concepts in a wider pursuit for a theory of landscapes. Weinberger states in his paper as the goal to derive a mathematical model that allows derivation of statistical properties of the landscape with "moderate" amount of data about the landscape [2]. His intention is to ultimately develop a theory that would classify different types of landscapes. When fully developed, the theory would provide pragmatic benefit of guiding the selection of optimization technique for each problem by its landscape type. From theoretical point of view, the landscape theory could complement the traditional complexity theory.

This paper is organized as follows. Next section defines ruggedness as autocorrelation of the landscape. The following sections present measures of local optima and barriers of the landscapes. Then, a brief section on neurality follows. The final section presents some concluding remarks.

In order to ease the reading of the full developments in the original papers, the definitions, lemmas and colloraries in this paper retain the numbering of the original papers. The original numbering and the bibliographical reference are used as a cross-reference within this paper. Definitions etc. are presented verbatim as much as possible.

Ruggedness as Autocorrelation

The characterization of the ruggedness of a landscape by its autocorrelation appears to originate in the work of Weinberger [2]. He considers the fitnesses of the configurations as random variables with a joint

Gaussian distribution and obtains an exponentially decaying random walk autocorrelation function for such landscapes. Stadler derives similar result in a wider context of elementary landscapes [4]. The definition of an elementary landscape relies on the definition of the graph Laplacian. Let $\Gamma=(V,E)$ be a directed graph and $x \in V$ a vertex of Γ. Furthermore, let $\mathbf{A}$ be the adjacency matrix of Γ so that

$$\mathbf{A} \overset{def}{=} \begin{cases} 1 & \text{if } \{x, y\} \in E \\ 0 & \text{otherwise} \end{cases}$$

Definition [4]. Let $\mathbf{D}$ be the diagonal matrix of vertex degrees, i.e., $\mathbf{D}_{xx}$ is the number of edges incident into x, and let $\mathbf{A}$ be the adjacency matrix of Γ. Then the matrix

$$-\boldsymbol{\Delta} = \mathbf{D} - \mathbf{A}$$

is called the Laplacian of Γ. For D-regular graph, i.e., graphs for which all vertices have degree D, this becomes $\boldsymbol{\Delta} = \mathbf{A} - D\mathbf{I}$.

Let $\{\varphi_i\}$ denote a complete ortonormal set of eigenvectors of the graph Laplacian $-\boldsymbol{\Delta}$. Then the following expansion is called the Fourier expansion of the landscape:

$$f(x) = \sum_{i=1}^{|V|} a_i \varphi_i(x) .$$

With these preliminaries, the definition of an elementary landscape can be presented.

Definition [4]. A landscape $f: V \to \mathbb{R}$ is elementary if there are constants f^* and λ such that

$$\boldsymbol{\Delta} f + \lambda (f - f^* \mathbf{1}) = 0 .$$

where $\mathbf{1}$ is a vector with all entries equal to 1.

Additional definitions are required to present further characterization of elementary landscapes.

Definition [4]. For each landscape $f: V \to \mathbb{R}$ we define

$$\bar{f} \overset{def}{=} \frac{1}{|V|} \sum_{x \in V} f(x) \qquad \sigma_f^2 \overset{def}{=} \frac{1}{|V|} \sum_{x \in V} \left[f(x) - \bar{f} \right]^2 = \overline{f^2} - \bar{f}^2$$

A landscape with $\sigma_f^2 = 0$ is called flat.

It should be noted that a landscape is flat if and only if f is constant. Clearly, $\bar{f}$ is the mean of the landscape, and σ_f^2 can be interpreted as the variance of the landscape. However, both are functionals of f and are not "statistical" in the sense of being probabilistic quantities.

The following lemma will give a further characterization of elementarity and later provide a bridge between elementarity and correlational properties in the sketch of proof for theorem 1[4].

Lemma 3 [4]. A non-flat landscape on a connected graph Γ is elementary if and only if

$$f(x) = \bar{f} + \varphi(x) \quad \forall x \in V ,$$

where φ is an eigenfunction of Δ with eigenvalue $\lambda>0$.

Several definitions and lemmas are needed for the derivation of the autocorrelation functions for landscapes. First, for the "random walk" autocorrelation on *V*, the expected autocorrelation of a time series $\{f(x_0), f(x_1), \ldots\}$ along the walk $\{x_0, x_1, \ldots\}$ is defined as

$$r(s) \stackrel{def}{=} \frac{\left\langle f(x_t) f(x_{t+s}) \right\rangle_{x_0,t} - \left\langle f(x_t) \right\rangle_{x_0,t} \left\langle f(x_{t+s}) \right\rangle_{x_0,t}}{\sqrt{\left(\left\langle f(x_t)^2 \right\rangle_{x_0,t} - \left\langle f(x_t) \right\rangle_{x_0,t}^2 \right) \left(\left\langle f(x_{t+s})^2 \right\rangle_{x_0,t} - \left\langle f(x_{t+s}) \right\rangle_{x_0,t}^2 \right)}}$$

where the expectation is take over the all "times" in the random walk and all initial conditions x_0.

Noting that the initial conditions are uniformly distributed allows the simplification

$$r(s) \stackrel{def}{=} \frac{\left\langle f(x_t) f(x_{t+s}) \right\rangle_{x_0,t} - \left\langle f(x_t) \right\rangle_{x_0,t}^2}{\left\langle f(x_t)^2 \right\rangle_{x_0,t} - \left\langle f(x_t) \right\rangle_{x_0,t}^2}$$

This simplification takes advantage of the fact that since the expectation is taken over all initial conditions $\left\langle f(x_{t+s}) \right\rangle_{x_0,t} = \left\langle f(x_t) \right\rangle_{x_0,t}$.

Two lemmas are necessary for the proof of the following corollary. Let $\mathbf{T} \stackrel{def}{=} \mathbf{D}^{-1}\mathbf{A}$.

Lemma 1 [4]. Let Γ be a regular graph and let $F: V \to \mathbb{R}$ be arbitrary function. Let $\{x_t\}$ be a simple random walk on Γ. Then

$$\left\langle F(x_t) \right\rangle_{x_0,t} = \overline{F}$$

Lemma 2 [4]. Let Γ be a regular graph, and let $F: V \times V \to \mathbb{R}$. Then

$$\left\langle F(x_{t+s}, x_t) \right\rangle_{x_0,t} = \frac{1}{|V|} \sum_{x,y \in V} F(x,y) \quad [\mathbf{T}]^s_{xy}$$

Let $\langle \mathbf{a}, \mathbf{b} \rangle$ signify the dot product of **a** and **b**. Finally, the autocorrelation function on a regular graph can be presented as a corollary.

Corollary 1 [4]. Let $f: V \to \mathbb{R}$ be a non-flat landscape on a *D*-regular graph Γ with adjacency matrix **A**. Then

$$r(s) = \frac{\frac{1}{|V|} \langle f, \mathbf{T}^s f \rangle - \bar{f}^2}{\bar{f}^2 - \bar{f}^2}$$

where $\mathbf{T} = (1/D)\mathbf{A}$.

Sketch of proof. The result follows by application of lemma 1 [4] with $F=f$ and $F=f^2$ and lemma 2 [4] with $F(x,y)=f(x)f(y)$ and substitution into the definition of $r(s)$.❑

Note that in the dot product of the above corollary *f* is to be taken as a vector of fitness values over the configurations. This becomes more clear as the requirement of regularity is relaxed in the following.

Noting that autocorrelation is invariant under the transformation $f \to f - \langle f \rangle$ and defining $\tilde{f} = f - \langle f \rangle$ with $\langle \tilde{f} \rangle = 0$, the autocorrelation can be written

$$r(s) = \frac{\langle f(x_t) f(x_{t+s}) \rangle - \langle f(x_t) \rangle^2}{\langle f(x_t)^2 \rangle - \langle f(x_t) \rangle^2} = \frac{\langle \tilde{f}(x_t) \tilde{f}(x_{t+s}) \rangle}{\langle \tilde{f}(x_t)^2 \rangle}$$

In order to present the following result further definition of notation is needed. Let

$$\langle f, g \rangle_p = \sum_x p(x) f(x) g^*(x)$$

where * denotes complex conjugation. All the functions in the theory are real, but the complex conjugation is retained to stay faithful to the original notation.

Eq (4.1) [2]. Let **T** be a transition matrix of a reversible Markov process on *V* with stationary distribution φ_0. The expected autocorrelation function along a **T**-random walk on *V* is

$$r(s) = \frac{\langle \tilde{f}(x) \tilde{f}(y) \rangle}{\langle \tilde{f}(x)^2 \rangle} = \frac{\sum_{x,y \in V} \tilde{f}(x) (\mathbf{T}^s)_{xy} \tilde{f}^*(y) \varphi_0(y)}{\sum_{x \in V} |\tilde{f}^2(x)| \varphi_0(x)} = \frac{\langle \tilde{f}, \mathbf{T}^s \tilde{f} \rangle_{\varphi_0}}{\langle \tilde{f}, \tilde{f} \rangle_{\varphi_0}}$$

The above equation calculates the autocorrelation as the ratio between average value of the "paths" of length *s* for all *y* and *x* biased with the stationary distribution φ_0 and the expected value of the squared fitness. Comparing to the equation of corollary 1 [4] above makes it obvious that the essential difference is that both the stationary distribution and transition probabilities are taken as uniform in the case of a regular graph.

Now, combining above statements, the connection between elementarity and autocorrelation can be stated with the following lemma.

Theorem 1[4]. Let *f* be a non-flat landscape on a D-regular graph Γ and let *r(s)* be the "random walk" correlation function of *f*. Then *f* is elementary if and only if *r(s)* is an exponential function, i.e., iff $r(s) = \rho^s$.

Sketch of proof. Express (the vector) *f* as Fourier expansion $f = \sum_i a_i \varphi_i$ and substitute into the expression of *r(s)*. Noting ortogonality and normalization of φ_i will lead to significant simplification of the expression. The remaining amplitudes a_i can be replaced with normalized amplitudes $A_i \overset{def}{=} \frac{|a_i|^2}{\sum_{j \neq 0} |a_j|^2}$ yielding

$$r(s) = \sum_{i \neq 0} A_i (1 - \lambda_i / D)^s$$

Now, *r(t)* remain exponential if and only if all nonzero A_i belong to single eigenvalue λ_k of $-\Delta$. This is the case if and only if *f* is of the form $f = (a_0 / \sqrt{|V|}) \mathbf{1} + \varphi$ where φ is an eigenvector of $-\mathbf{\Delta}$. Applying lemma 3 [4] above with completes the proof. ❑

The correlation of an elementary landscape can be described with a single value *l* called correlation length:

$$\rho \overset{def}{=} r(1) = (1 - \lambda_k / D) \qquad l \overset{def}{=} \begin{cases} 0, & \text{if } \rho = 0 \\ \dfrac{-1}{\ln|\rho|} & \text{if } \rho \neq 0 \end{cases}$$

Again, the assumption of D-regularity can be relaxed to obtain autocorrelation and correlation length with transition matrix **T [2].** Expanding f w.r.t of the eigenvectors of **T,** it can be shown that

$$r(t) = \sum_{\lambda \neq 1} B_T(\lambda)\lambda^t \qquad\qquad l = \sum_{t=0}^{\infty} r(t)$$

where $B_{\mathbf{T}}(\lambda)$ are the amplitudes of f w.r.t. the eigenspaces of **T.**

The above autocorrelation functions are defined for time series over vertices of graph Γ. The following defines autocorrelation function for relations on $V \times V$.

Definition 4.1[2]. Given a relation R on $V \times V$, the autocorrelation function of f w.r.t R is:

$$\rho(R) = \frac{|V|^2}{|R|} \quad \frac{\sum_{(x,y)\in R}(f(x)-\bar{f})(f(y)-\bar{f})}{\sum_{x,y\in V}(f(x)-\bar{f})(f(y)-\bar{f})}$$

The autocorrelation of a relation can be seen as the ratio of scaled covariances in the relation and in the whole graph. The autocorrelation function may have useful algebraic properties, if the partition R has enough regularity. For details, see [4].

Table 1 presents properties of some known elementary landscapes. The problems are Not-All-Equal-Satisfiability (NAES), weight partition (WP), graph coloring problem (GCP), p-spin, traveling salesman problem (TSP), graph matching problem (GMP) and graph bipartioning problem (GBP). There are two version of TSP. The cost matrix **W** of TSP can be uniquely decomposed into symmetric component $\mathbf{W}^{\sigma} = (\mathbf{W} + \mathbf{W}^{\mathbf{T}})/2$ and antisymmetric component $\mathbf{W}^{\alpha} = (\mathbf{W} - \mathbf{W}^{\mathbf{T}})/2$. TSP properties with both types of cost matrix are presented into the table. Furthermore, both inversions (*2-opt*) and transpositions are considered as operations for TSP producing total of four different versions of TSP in the table.

The table gives the graph type (Γ) capturing the neighborhood structure of the problem, number of neighbors each configuration has (D), the eigenvalue λ of the Laplacian present in the Fourier expansion of the landscape, correlation between neighboring configurations (σ) and correlation length (l). Graphs are Hamming graph of size n with alphabet size of α (Q_{α}^{n}), graph with the symmetric group S_n as the vertex set V and either transpositions (T) or inversions (I) as the move set defining the edge set of the graph, and Johnson graph. Johnson graph *J(n,k)* is a graph with subsets of exactly k elements as the vertex set of the graph and edges between vertices that have exactly *k-1* elements in common.

Unfortunately, interpretation of the results in the table does not give much information about the properties of the problems. The only distinction seems to be that while for most of the problems the correlation length is essentially a linear function of the problem size, it is a function of inverse of the logarithm of the problem size for the antisymmetric TSP with inversions. It means that for this particular TSP version, the correlation length becomes shorter as the problems size is increased.

Stadler examines the nearest neighbor correlations (σ) of TSP versions and compares those with the known facts about the performance of simulated annealing in solving TSP [4]. A TSP with symmetric cost matrix **W** has obviously only symmetric component $\mathbf{W}^{\sigma}$ in the matrix, while TSP with asymmetric cost matrix has both symmetric component $\mathbf{W}^{\sigma}$ and antisymmetric component $\mathbf{W}^{\alpha}$ in the cost matrix. The components of the cost matrix yield components to the landscape, so that a TSP instance has an elementary landscape, if it has either symmetric or antisymmetric cost matrix. Asymmetric TSP has a composite landscape with symmetric and antisymmetric components. It is know that inversions are an effective move set for symmetric TSP but a very inefficient move set for asymmetric TSP. This observation can be understood in terms of the correlations so that inversions with symmetric cost matrix yield a single smooth elementary landscape, but the asymmetric TSP has additional antisymmetric component with nearest neighbor

correlation tending towards zero. In Stadler's words, the symmetric TSP has as smooth landscape as possible while asymmetric TSP has extremely rugged landscape. This explanation is somewhat contradicted by the observation that *3-opt* inversions have smaller correlation length than *2-opt* inversions, but still *3-opt* inversions lead to better solutions [1]. Stadler and Schnabl use additional neighborhood properties to explain away this contradictory observation, but it remains as an indication that correlation has only limited explanatory power.

Problem	Γ	*D*	λ	σ	*l*
NAES	Q_2^n	*n*	*4*	$1-\frac{4}{n}$	$\frac{1}{4}n-\frac{1}{2}-\frac{1}{3}\frac{1}{n}+O\left(\frac{1}{n}\right)$
WP	Q_2^n	*n*	*4*	$1-\frac{4}{n}$	$\frac{1}{4}n-\frac{1}{2}-\frac{1}{3}\frac{1}{n}+O\left(\frac{1}{n}\right)$
p-spin	Q_2^n	*n*	*2p*	$1-\frac{2p}{n}$	$\frac{1}{2p}n-\frac{1}{2}-\frac{p}{6}\frac{1}{n}+O\left(\frac{1}{n}\right)$
GCP	Q_α^n	*(α-1)n*	*2α*	$1-\frac{2\alpha}{(\alpha-1)n}$	$\frac{\alpha-1}{2\alpha}n-\frac{1}{2}+\frac{\alpha}{6(\alpha-1)}\frac{1}{n}+O\left(\frac{1}{n}\right)$
Symmetric TSP	$\Gamma(S_n,T)$	*n(n-1)/2*	*2(n-1)*	$1-\frac{4}{n}$	$\frac{1}{4}n-\frac{1}{2}-\frac{1}{3}\frac{1}{n}+O\left(\frac{1}{n}\right)$
	$\Gamma(S_n,I)$	*n(n-1)/2*	*n*	$1-\frac{2}{n-1}$	$\frac{1}{2}n-1-\frac{1}{6}\frac{1}{n}+O\left(\frac{1}{n}\right)$
Antisymmetric TSP	$\Gamma(S_n,T)$	*n(n-1)/2*	*2n*	$1-\frac{4}{n-1}$	$\frac{1}{4}n-\frac{3}{4}-\frac{1}{3}\frac{1}{n}+O\left(\frac{1}{n}\right)$
	$\Gamma(S_n,I)$	*n(n-1)/2*	*n(n+1)/2*	$-\frac{2}{n-1}$	$\frac{1}{\ln n}-\ln 2\frac{1}{\ln^2 n}+O(\ln^3 n)$
GMP	$\Gamma(S_n,T)$	*n(n-1)/2*	*2(n-1)*	$1-\frac{4}{n}$	$\frac{1}{4}n-\frac{1}{2}-\frac{1}{3}\frac{1}{n}+O\left(\frac{1}{n}\right)$
GBP	$J(n,n/2)$	$n^2/4$	*2(n-1)*	$1-\frac{8}{n}+\frac{8}{n^2}$	$\frac{1}{8}n-\frac{3}{8}-\frac{13}{24}\frac{1}{n}+O\left(\frac{1}{n}\right)$

Table 1: Summary of known elementary landscapes [4]. See text for details.

Ruggedness and Local optima

Above presentation used the correlation between configurations as a measure of the ruggedness of the landscape. Saddle-points and local minima play a crucial role in classical optimization theory. Alternative measure of ruggedness can be based on the number and distribution of the local minima in the landscape. Let *N(x)* be the set of neighbors of $x \in V$, that is $N(x)=\{y\in V:\{x,y\}\in E\}$. Local minima are configurations $\hat{x}\in V: f(\hat{x})\le f(y)$ for all $y\in N(\hat{x})$. Local maxima are defined by replacing *f* with *-f*.

A measure of ruggedness can be based on the growth of the number of local optima M_f: Landscape is rugged, if M_f scales exponentially with a measure of the system size *n* [2]. Unfortunately, there is no general method for computing M_f except by exhaustive numeration or random sampling. Part of the motivation of the correlation length as a measure of ruggedness is provided by a conjecture that suggests an association between correlation length and number of local minima in the landscape. Let $\mathbf{N}(r)$ denote number of vertices in a neighborhood with radius *r* around an arbitrary vertex of the graph $\Gamma=(V,E)$. The Correlation Length Conjecture [1] states that there are *O(1)* local optima in a "patch" with radius of correlation length *l*:

$$\Psi = \mathrm{Prob}\{\text{local optimum}\} \approx \frac{\mathbf{N}(l)}{|V|}$$

Estimates of the growth of M_f have been derived for some models and compiled together by Stadler. Let

$$A \stackrel{def}{=} \lim_{n\to\infty} \frac{1}{n} \ln E(M_f) \qquad\qquad \xi \stackrel{def}{=} l/n$$

Now, A gives the exponent of the expected asymptotic growth of the number of local minima as the system size in increased without a bound and ξ is the correlation length normalized with the system size n. Table 2 presents values of A for spin models and Nk models together with the prediction given by Correlation Length Conjecture. The block model is such a variation of the Nk-model that the fitness function is computed as a sum of fitnesses of blocks of the sequence.

As seen from the data in the table, all the models exceed the growth rate A predicted by Correlation Length Conjecture. A futher interesting detail is that while the empirical correlation function is the same for both spin glass models, they have different values of A.

Model	A
2-spin SK	0.1992
2-spin chain	0.2415[1]
Nk, k=3	0.3466
Block Model	0.3466
Correlation Length Conjecture	0.1308

Table 2: Estimated values of A for various landscapes with $\zeta=l/4$ on Boolean hypercubes [5].

Basins

Assuming a local search algorithm that can find from any configuration $x \in V$ an unique local optima $\hat{x}_i$, the landscape can be split into a set of attraction basins $\mathbf{B}(\hat{x}_i)$ of the local minima. The steepest descent algorithm is a possible candidate for defining the basins, although it is well known that it may lead to ambiguous basins, if several neighbors have the same smallest fitness value. Stringent definition of basin is still an open question and not addressed in this paper. However, the distribution of the basin sizes is crucial for the performance of many optimization techniques, such as simulated annealing, and therefore the estimation and characterization of this distribution is an important issue.

Garnier and Kallel have presented a method for estimating the distribution of attraction basins given a random sample of points on the landscape [6]. In derivation of the method, they present a number of interesting results. The first of their results presented here describes the probability of having at least one sample in every basin of the landscape including the global optimum with a sample of size m.

Proposition 4.1 [6]. If we denote by $\alpha_j = \mathbf{B}(\hat{x}_j)/|V|$ the normalized size of the j-th attraction basin, then

$$p(m) = \sum_{k=0}^{M_f} (-1)^k \sum_{1 \le j_1 \le \cdots \le j_k \le M_f} (1 - \alpha_{j_1} - \cdots - \alpha_{j_k})^m$$

gives the probability of having at least one point of the random sample in each basin.

---

[1] Two values 0.2415 and 0.2410 are provided for this quantity in the original paper [5], both which are different from the given symbolic expression $4/\pi$ of the value for A. The table has been modified to use $A=\ln(4/\pi)\approx 0.2415$ which appears to be the intended value.

Of course, obtaining the actual values of α_j is a difficult task. Assuming that the basin sizes are "random", that is uniformly distributed, the following corollary states that $O(M_f^2)$ points provide a finite chance to find the basin of global optima.

Corollary 4.3 [6]. Assuming α_j´s are jointly uniformly distributed over a simplex of $\mathbb{R}^N$ and $M_f >> 1$ and $m=aM_f^2$ then

$$p(m) \overset{m\to\infty}{\to} \exp(-1/a)$$

Assuming that the distribution of basin sizes follow some suitable parameterized family of distributions, it is possible to estimate the number and distribution of basins of the landscape from the basins observed with the sample. Let β_j be the number of minima detected with j points. For example, a sample of six points detecting 3 local optima so that two optima were detected with one point and one optima with four points would yield $\beta_1 = 2, \beta_2 = 0, \beta_3 = 0$, and $\beta_4 = 1$. Let H' signify that α_j´s can be described as $\left(Z_1 / T_{M_f}, \cdots, Z_{M_f} / T_{M_f}\right)$ with Z_j following the distribution

$$p_\gamma(z) = \frac{\gamma^\gamma}{\Gamma(\gamma)} z^{\gamma-1} e^{-\gamma z}$$

and $T_{M_f} = \sum_{i=1}^{M_f} Z_j$.

Proposition 5.1 [6]. Under H' the expected values $\beta_{j,\gamma} = \mathrm{E}(\beta_j)$ of the β_j's can be computed in the asymptotic framework N>>1, $m=aM_f$ as

$$\beta_{j,\gamma} = M_f \frac{\Gamma(j+\gamma)}{j!\Gamma(\gamma)} \frac{a^j \gamma^\gamma}{(a+\gamma)^{j+\gamma}} + o(M_f).$$

Given an observed set of β_j's, its distribution is compared with X^2 test to the above distribution with each value of γ to obtain the best approximation of the distribution. Once γ is determined, letting $r=m/M_f$ and noting $M_f = \sum_{j=1}^{\infty} \beta_{j,\gamma}$, the number of local optima M_f can be estimated from the unique solution r to the equation

$$\frac{\sum_{j=1}^{\infty} \beta_{j,\gamma}}{m} = \frac{1 - \left(1 + \frac{r}{\gamma}\right)^{-\gamma}}{r}.$$

Another description of the landscape is given by the saddle points and fitness barriers that separate the local minima. The fitness barrier separating local minima $\hat{x}$ and $\hat{y}$ is

$$f[\hat{x}, \hat{y}] = \min\{\max[f(z) \mid z \in \mathbf{p}] \mid \mathbf{p} : \text{path from } \hat{x} \text{ to } \hat{y}\}$$

A point $\hat{z}$ satisfying the above minmax condition is called a saddle point of the landscape. The saddle points and local minima can be represented as a tree with minima at leaves and saddle points at internal nodes, see Figure 1. An algorithm for constructing a barrier tree is presented in [7]. Realistic barrier trees for a number of problems are presented in the Figure 2.

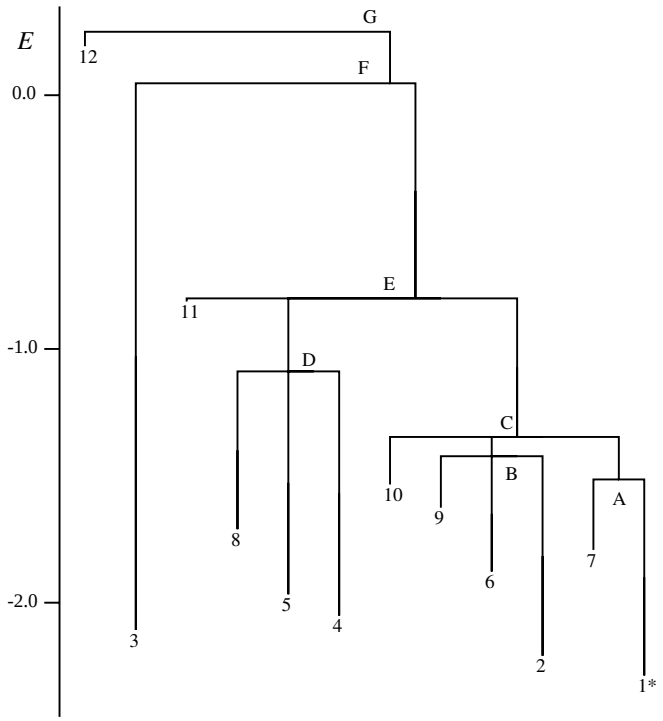

Figure 1: An example of a barrier tree of a landscape [7]. The numbers 1-12 label local minima and letters A-G label the saddle points. The global optimum 1 is marked with an asterisk. The data comes from a Gaussian Random Energy Model with system size n=7.

Barrier trees provide a nice visualization of the saddle point structure of the landscape. The information can be further compacted as follows. Let

$$B(\hat{x}) = \min\{f[\hat{x}, \hat{y}] - f(\hat{x}) \mid \hat{y} : f(\hat{y}) < f(\hat{x})\}$$

be the barrier enclosing local minima $\hat{x}$ [2]. Barrier is the height of the lowest saddle point giving access to a more favorable minimum. Let Ω_f be the set of global minima of the landscape. Now depth D and difficulty ψ are defined

$$D = \max\{B(s) \mid s \notin \Omega_f\} \qquad \Psi = \max\left\{\frac{B(s)}{f(s) - f(x_{\min})} \mid s \notin \Omega_f, x_{\min} \in \Omega_f\right\}$$

Depth gives the highest barrier of the landscape and difficulty gives the most “expensive” barrier of the landscape in terms of relative improvement in fitness gained by climbing over the barrier. These parameters play a role in the theory of simulated annealing. However, Kern conjectures that computing the depth of a problem is as expensive as solving the optimization problem [8].

As a concluding remark of this section, it should be noted that there seems to be a correlation between the fitness or energy of the local minima and the size of its basin of attraction [7]. At the moment, this is an empirical result with no convincing theoretical explanation.

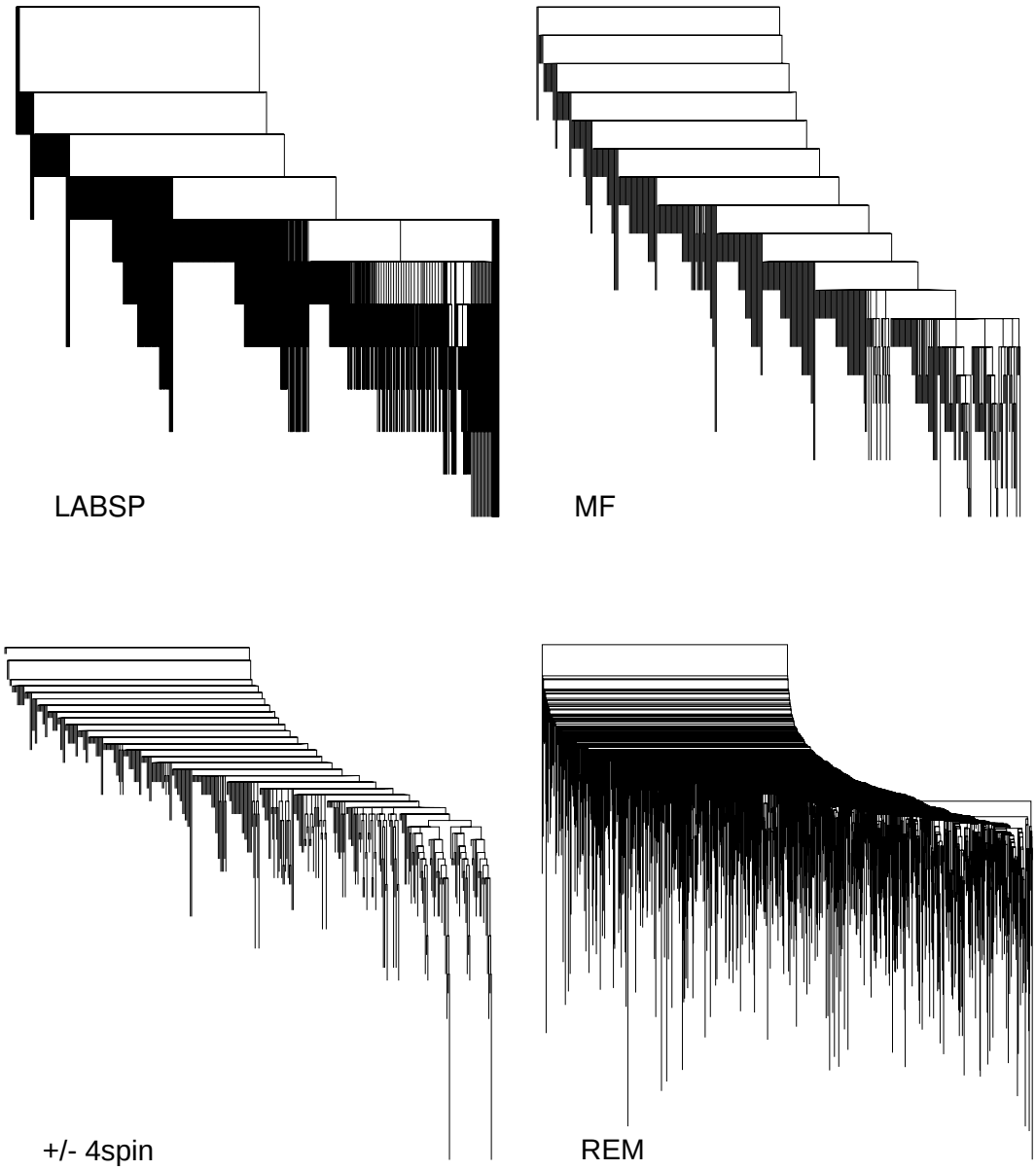

Figure 2: Typical barrier trees for Low Autocorrelated Binary String Problem, Mean Field approximation, 4-spin model, and Gaussian Random Energy Model [7]. REM has system size n=14, other models have n=16.

Neutrality

Two configurations $x, y \in V$ are considered neutral, if they have the same fitness value: $f(x) = f(y)$ [9]. If a "substantial fraction" of the configurations of a landscape are neutral, the landscape as a whole is referred to as neutral [9]. Neutrality plays an important role for example in RNA folding and sequential dynamical systems. It may also help characterize the behavior of combinatorial optimization algorithms, but relevant

results similar to the role of correlation in TSP or depth and difficulty in simulated annealing are still to be presented.

A measure of neutrality of a configuration can be defined as the number of neutral neighbors. Let Γ be a graph with vertex set V and edge-set E. The number of neutral neighbors of $x \in V$ is $v(x) = \sum_{y \in N(x)} \delta(f(x), f(y))$. Trivial neutrality can be obtained by embedding a combinatorial optimization problem in a state space that is too large. As an example consider that if the landscape of graph matching problem is presented as $\Gamma(S_n, T)$, the transpositions of form $\tau_{2k-1,2k}$ leave the fitness value unchanged as they merely exchange the endpoints of a single edge.

More interesting results can be derived by studying neutrality in additive random landscapes. Let $\mathbf{A}_j$ be a σ–field.

Definition 1 [9]. Let V and M be finite sets and $\Theta = (\mathbf{v}_j)_{j \in M}$ a family of maps $\mathbf{v}_j : V \to \mathbb{R}$. Further, let c_j, $j \in M$ be independent, real valued random variables (over the respective probability spaces $\Omega_j = (\mathbb{R}, \mathbf{A}_j, \mu_j)$) and

$$\Omega_V = \left\{ f : V \to R \mid f(x) = \sum_{j=1}^{M} c_j \mathbf{v}_j(x) \right\}.$$

An additive random landscape (arl), $\mathbf{F}(V, \Theta, (c_j))$, is the probability space

$$\left(\Omega_V, \otimes_j \mathbf{A}_j, \otimes_j \mu_j\right)$$

The number of neutral neighbors $v(x)$ can be studied as an another landscape on Γ. The following will present statistics for the neutrality of an additive random landscape, when the coefficients c_j vanish with a non-zero probability. First, some definitions are needed. Let $y, y', y'' \in N(x)$. Additionally, for any set $\Phi \subseteq M$:

$$c_x(y) = \left| \left\{ j \in \Phi \mid \mathbf{v}_j(x) \neq \mathbf{v}_j(y) \right\} \right|$$

$$w_x(y', y'') = \left| \left\{ j \in \Phi \mid \mathbf{v}_j(x) \neq \mathbf{v}_j(y') \wedge \mathbf{v}_j(x) \neq \mathbf{v}_j(y'') \right\} \right|$$

$$\Xi = E\left[\frac{1}{|V|} \sum_x \left(v_x - \frac{1}{|V|} \sum_{x'} v_{x'} \right)^2 \right]$$

Ξ is the expected variance of the family ν_x across given landscape.

Lemma 3 [9]. Let M be a finite index set and c_j, $j \in M$, be independent real valued random variables such that

$$\mu(c_j = \xi) = \begin{cases} \mu_0 > 0 & \text{if } \xi = 0 \\ 0 & \text{otherwise} \end{cases}$$

Then we have for any set $\Phi \subseteq M$ of non-zero constants $0 \neq \eta_j \in \mathbb{R}$

$$\mu\left\{ \sum_{j \in \Phi} c_j \eta_j = 0 \right\} > 0 \Rightarrow \forall j \in \Phi, c_j = 0 .$$

Theorem 1 [9]. For any arl whose coefficients c_i fulfill lemma 3 [9] the following assertions hold

$$\mathrm{E}[\nu_x] = \sum_{y \in N(x)} \mu_{\mu_0}^{c_x(y)}$$

$$\mathrm{V}(\nu_x) = \sum_{y', y''} \mu_0^{c_x(y') + c_x(y'')} \left[\mu_0^{-w_x(y', y'')} - 1 \right]$$

$$\Xi = \frac{1}{|V|}\left[\sum_{y} \mathrm{V}(v_y) - \frac{1}{|V|}\sum_{y,y'} \mathrm{Cov}(v_y, v_{y'})\right] + \frac{1}{|V|}\sum_{\mathrm{y}} \mathrm{E}[v_x]^2 - \left(\frac{1}{|V|}\sum_{\mathrm{y}} \mathrm{E}[v_x]\right)^2$$

where $\frac{1}{|V|}\sum_{y,y'} Cov(v_y, v_{y'}) \geq 0$. In particular, if E[v_y] is independent of y, we have

$$\Xi = \frac{1}{|V|}\left[\sum_{y} \mathrm{V}(v_y) - \frac{1}{|V|}\sum_{y,y'} \mathrm{Cov}(v_y, v_{y'})\right]$$

In the particular case of p-spin glass, a more insightful result can be derived.

Proposition 3 [9]. For a p-spin model, where the coefficients c_i fulfil lemma 3 [9] the following assertions hold:

$$\mathrm{E}[v] = n\mu_0^{\binom{n-1}{p-1}}$$

$$\mathrm{V}[v] = n(n-1)\mu_0^{2\binom{n-1}{p-1}}\left[\mu_0^{-\binom{n-1}{p-1}} - 1\right] + n\mu_0^{\binom{n-1}{p-1}}\left[1 - \mu_0^{\binom{n-1}{p-1}}\right]$$

$$\Xi = 0$$

The interpretation of the result yields the observation that the neutrality of p-spin can be tuned to any desired value with parameter μ_0. On the other hand, p can be used to prescribe any desired degree of ruggedness ($\lambda=2p$). Thus, ruggedness and neutrality are independent features of a (random) landscape.

Conclusions

This seminar paper presented several definitions of ruggedness of a landscape and possible connections between them. Additionally, the neutrality of a landscape was defined and a more specific expression was presented for the p-spin landscape. The use of correlation as an explanation for the differing efficiency of combinatorial optimization algorithms hinges on the Correlation Length Conjecture. No rigorous direct results linking the correlation and algorithm performance are available. Furthermore, it seems that the contradictory evidence against the conjecture has not been given high emphasis.

Characterization of the distribution of the local minima seems an interesting new direction. The Garnier and Kallel method for estimating the number and sizes of attraction basins has obvious uses in heuristic algorithms, but due to its approximative nature, its use in theoretical investigations is limited. As noted by Garnier and Kallel, the distribution of the basin sizes does not fully characterize the landscape. The observation that there seems to be a correlation between the size of the basin and the fitness value of the optimum should be examined in more detail.

Neutrality has specific applications but its relevance to combinatorial optimization is not obvious: there are no results that would connect neutrality to the efficiency of any algorithm. Such results cannot be excluded, however.

Bibliography

[1] P. F. Stadler, and W. Schnabl, The Landscape of the Traveling Salesman Problem, Physics Letters A, Vol. 161, 1992, 337-344. http://www.tbi.univie.ac.at/papers/Abstracts/92-01-003.ps.gz

[2] C. M. Reidys and P. F. Stadler, Combinatorial Landscapes, to appear. http://www.santafe.edu/sfi/publications/Working-Papers/01-03-014.pdf

[3] E. Weinberger, Correlated and Uncorrelated Fitness Landscapes and How to Tell the Difference, Biological Cybernetics, Vol. 63, No. 5, September 1990, 325-336.

[4] P. F. Stadler, Landscapes and Their Correlation Functions, J. Math. Chem, 20, 1996, 1-45. http://www.tbi.univie.ac.at/papers/Abstracts/95-07-07.ps.gz

[5] P. F. Stadler, Towards a Theory of Landscapes, in Complex Systems and Binary Networks, R. Lopez-Pena, R. Capovilla, R. Garcia-Pelayo, H. Waelbroeck, and F. Zertuche, eds., Berlin, New York, Springer Verlag, 1995, 77-163. http://www.santafe.edu/sfi/publications/Working-Papers/95-03-030.ps

[6] J. Garnier, & L. Kallel, Efficiency of Local Search with Multiple Local Optima, SIAM J. Discr. Math., Vol 15, No 1, 2000, 122-141. http://www.cmap.polytechnique.fr/users/www.garnier/publi/bassin.ps.gz

[7] F. F. Ferreira, J. F. Fontanari, & P. F. Stadler, Landscape Statistics of the Low Autocorrelated Binary String Problem J. Phys. A: Math. Gen., 33, 2000, 8635-8647. http://www.tbi.univie.ac.at/papers/Abstracts/00-07-033.pdf

[8] W. Kern, On the Depth of Combinatorial Optimization Problems, Discr. Appl. Math., 43, 1993, 115-129.

[9] C. M. Reidys and P. F. Stadler, Neutrality in Fitness Landscapes, Appl. Math. & Comput., 117, 2001, 321-350. http://www.tbi.univie.ac.at/papers/Abstracts/98-10-089.ps.gz