T-79.300 Postgraduate Course in Theoretical Computer Science:

Ruggedness and Neutrality

Pekka Isto

Contents

- Motivations
- Ruggedness and autocorrelation
- Ruggedness and local optima, basins
- Barriers and depths, barrier tree
- Neutrality

Motivations

- Landscape properties seem to play an important role in combinatorial optimization, e.g. TSP operations.
- One would like to gather “moderate” amount of data about the landscape and infer properties of the landscape and select best possible optimization strategy (Weinberger, 1990).
- Discovery of connections between landscape and complexity theory (Weinberger, 1990).

Ruggedness

- Intuitively, ruggedness is the opposite of "smoothness" (Reidys & Stadler, to appear).
- Ruggedness can be quantified by the correlation of between adjacent configurations.
- The number and distribution of local minima can provide another characterization of ruggedness.
- Correlation length conjecture bridges the above two approaches.
- Nodal domain theorem.

Random Walk on the Landscape

- Weinberger (1990) proposes to consider fitnesses of configurations as random variables and to obtain their statistical properties.
- He uses random walk on the landscape to approximate autocovariance function.
- He recognizes that landscapes with exponentially decaying autocovariance function are of particular interest and proposes the use of "Fourier-like decomposition" to study the landscape structure.

Correlation Functions of Landscapes

- The following will present two correlation functions on landscapes:
  - Random walk correlation function for “time series” on vertices.
  - Autocorrelation function for partitions of vertex pairs.

The Expected Autocorrelation

The expected autocorrelation of the time series $\{f(x_0), f(x_1), \ldots\}$ along the walk $\{x_0, x_1, \ldots\}$ on Γ is defined as

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

where expectations are take over all "times" t and initial conditions x_0.

The Expected Autocorrelation

Noting that initial conditions are uniformly distributed allows a simplification:

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

Mean and Variance of Landscape

Definition (Stadler, 1996). For each landscape $f\!: V \rightarrow$ we define:

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

These are functionals of f. The former is the mean of the landscape. The latter is interpreted as the variance of the landscape. $\sigma^2{}_f = 0$ if and only if f is constant, i.e. for *flat* landscapes.

Functions on Random Walk

Lemma 1 (Stadler, 1996). Let Γ be a regular graph and let $f\colon V \to \mathbb{R}$ 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 (Stadler, 1996). Let Γ be a regular graph and let $f\colon V \times V \to \mathbb{R}$. Then

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

Autocorrelation Function on Regular Graph

Corollary 1 (Stadler, 1996). Let $f : V \to \mathbb{R}$ be a non-flat landscape on a D-regular graph Γ with adjacency matrix $\mathbf{A}$. Then by application of lemma 1 with $F=f$ and $F=f^2$ and lemma 2 with $F(x,y)=f(x)f(y)$:

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

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

Autocorrelation Function

(Reidys & Stadler, to appear) relaxes the requirement that Γ is regular.

Noting that autocorrelation is invariant under the transformation $f \to f - \langle f \rangle$ and

$$\tilde{f} = f - \langle f \rangle, \qquad \langle \tilde{f} \rangle = 0$$

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

Autocorrelation Function

Eq (4.1) (Reidys & Stadler, to appear). Let **T** be a transition matrix of a reversible Markov process on V with stationary distribution $_0$. The expected autocorrelation function along a **T**-random walk on V is

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

Correlation Functions of Elementary Landscapes

Theorem 1 (Stadler, 1996). 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 $r(s) = \varrho^s$.

Theorem 1: Sketch of Proof

- Express *f* as decomposition $f = \underset{i}{\ } a_i \varphi_i$
- Substitute into *r(s)* noting orthogonality and normalization constraints.
- Define normalized amplitudes $A_i \overset{def}{=} \dfrac{|a_i|^2}{\underset{j \neq i}{\ } |a_j|^2}$
- Simplify: $r(t) = \underset{i \neq 0}{\ } A_i \left(1 - \lambda_i / D\right)^s$
- r(t) is exponential iff all nonzero A_i belong to single eigenvalue λ_k of -Δ.
- This is the case only when $f = \left(a_0 / \sqrt{|V|}\right)\mathbf{1} + \varphi$
- For some eigenvector of -Δ, thus complying with elementarity (lemma 3 of (Stadler, 1996))

Correlation Length

For elementary landscape a single parameter determines the correlation function:

$$\rho \overset{def}{=} r(1) = \left(1 - \lambda_k / D\right)$$

Correlation length is defined

$$l = \begin{cases} 0, & \text{if } \varsigma = 0 \\ \dfrac{-1}{\ln|\rho|} & \text{if } \varsigma \neq 0 \end{cases}$$

Relaxing the Regularity

Again, the assumption of D-regularity can be relaxed to obtain for transition matrix **T**:

$$r(t) = \sum_{\lambda \neq 1} B_T(\lambda)\lambda^t$$

$$l = \sum_{t=0}^{\infty} r(t)$$

Autocorrelation Functions on Partitions

Similar results have been derived for relations on $V \times V$.

Definition 4.1 (Reidys & Stadler, to appear). Given a relation R on $V \times V$, the autocorrelation function of f w.r.t R is:

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

If the partition of $V \times V$ is “nice”, the correlation function has “useful” algebraic properties.

Table 1(Stadler, 1996). Compare antisymmetric TSP with transpositions and inversions. Any other peculiarities?

Ruggedness and Local Optima

Local minima are configurations

$$\hat{x} \in V : f(\hat{x}) \leq f(y) \quad \text{for all } y \in N(\hat{x})$$

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 system size n.

Correlation Length Conjecture

Let *N(r)* denote number of vertices in a neighborhood with radius *r*.

(Stadler, P. F. & Schnabl W., 1992):

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

That is one local optimum on a patch with radius of correlation length. Said to be supported by numerical data (3-opt!).

Growth of M_f

Estimates of M_f are available for some models.

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

Table 3 (Stadler, 1995). A’s are similar, but larger than expected from correlation length conjecture (compare to last line).

Basins

Assume that a "search space" E can be split into partition E_i, $i=1,..,M_f$ of subspaces which are the attraction basins of the local maxima m_i. Furthermore, let there be an algorithm which is able to find from any $x \in E$ the corresponding local maximum. Steepest ascent is a suitable candidate, but may be ambiguous, if several neighbors have same fitness value. Stringent definition of basin is an open question.

Sampling for the Basins

The problem is to estimate the probability of hitting every basin with a random sample of size m.

Proposition 4.1 (Garnier,& Kallel, 2000). If we denote by $\alpha_j = |E_j|/|E|$ 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.

Sampling for the Global Optimum

Corollary 4.3 (Garnier,& Kallel, 2000). Assuming α_j´s are jointly uniformly distributed over a simplex of $\;^{N}$ and $M_f >> 1$ and $m{=}aM_f^2$ then

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

meaning that O($M_f^{2)}$ points provide a finite chance to find the the basin of global optima.

Estimating the Basin Distribution

Let β_j be the number of minima detected with j points. Let H^γ signify that α_j´s can be described as $(Z_1/T_N, \ldots Z_N/T_N)$ with Z_j following the distribution

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

and

$$T_N = \sum_{i=1}^{N} Z_i$$

Estimating the Basin Distribution

Proposition 5.1 (Garnier,& Kallel, 2000) Under H^{γ} the expected values $\beta_{j,\gamma} = \mathrm{E}(\beta_j)$ of the β_j's can be computed in the asymptotic framework N>>1, $m=aM_f^2$ as

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

This results can be used to compute estimate for the number of local minima from observed β_j's.

Fitness Barriers and Saddle Points

The local minima are separated by saddle points and fitness barriers. The fitness barrier separating local minima $\hat{x}, \hat{y}$

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

$\hat{z} \in X$ satisfying above minmax condition is called a saddle point of the landscape. Barriers can be represented as a three with minima at leaves and saddle points at internal nodes.

Figure 4.1 (Reidys & Stadler, to appear)

Figure 3 (Ferreira, Fontanari, & Stadler, 2000). Barrier trees for Low Autocorrelated Binary String Problem, Mean Field approximation to LABSP, +/- 4-spin glass model and Random Energy Model.

Barriers and Depth

Let's define the height of lowest saddle point giving access to better minimum:

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

Now, depth and difficulty are defined over the set of global minima Ω_{f}

$$D = \max\{B(s) \mid s \notin \Omega_f\}$$

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

Depth and Difficulty

Depth and Difficulty summarize the energy barriers and play a role in the theory of Simulated Annealing. However,

Conjecture 1 (Kern,1993). Computing the depth function is at least as hard as solving an optimization problem.

Conjecture 2 (Kern,1993). Computing the depth function is at most as hard as solving the optimization problem.

Proofs?

Neutrality

Let Γ be a graph with vertex set V and edge-set E. The number of neutral neighbors of $x \in V$ is

$$\nu(x) = \underset{y \in N(x)}{} \delta\big(f(x), f(y)\big)$$

which can be studied as a landscape on Γ. Trivial neutrality can be obtained by embedding a combinatorial optimization problem in a state space that is too large, e.g. Graph Matching Problem in (S_n, T).

Neutrality in ARL's

Let $$y, y', y'' \in N(x)$$

$$\Phi \subseteq M \text{ of non-zero constants} \in R$$

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

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

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

$\mathit{\Xi}$ is the expected variance of the family ν_x across given landscape.

Neutrality in ARL's

Theorem 4.6 (Reidys & Stadler, to appear). Let F be an ARL with coefficients c_i, satisfying

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

Neutrality in ARL's

Theorem 4.6 (Cont.) then

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

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

$$\Xi = \frac{1}{|V|}\left[\sum_y \mathrm{V}(\nu_y) - \frac{1}{|V|}\sum_{y,y'} \mathrm{Cov}(\nu_y, \nu_{y'})\right] + \frac{1}{|V|}\mathrm{E}\left[\nu_x\right]^2 - \left(\frac{1}{|V|}\mathrm{E}\left[\nu_x\right]\right)^2$$

Neutrality in p-Spin model

Corollary 4.7 (Reidys & Stadler, to appear). For a p-spin model with coefficients c_i satisfying (4.6):

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

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

$$\Xi = 0$$

For short range spin glasses:

$$\mu_0 = 1 - \frac{z}{n^{p-1}}, \text{ for } z > 0 \text{ determined by connectivity.}$$

Giving

$$\lim_{n\to\infty} \mathrm{E}[\nu/\mathrm{n}] = e^{-z} \qquad\qquad \lim_{n\to\infty} \mathrm{V}[\nu/\mathrm{n}] = 0$$

Neutrality and Ruggedness

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 ($\lambda=2p$).

Ruggedness and neutrality are independent features of (random) landscape.

Some Impressions

- The correlation theory hinges on the Correlation Length Conjecture. One should take a serious look at that.
- Characterization of the distribution of the local minima seems an interesting new direction.
- Neutrality has specific applications but it's relevance to combinatorial *optimization* is not obvious.

References

Ferreira, F. F. , Fontanari, J. F., & Stadler, P. F. , 2000 Landscape Statistics of the Low Autocorrelated Binary String Problem J. Phys. A: Math. Gen., 33, 8635-8647.

Garnier, J. & Kallel, L., 2000, Efficiency of Local Search with Multiple Local Optima, SIAM J. Discr. Math., Vol 15, No 1, 122-141.

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

Reidys, C. M. & Stadler P. F., to appear, Combinatorial Landscapes.

Stadler P. F., 1995, 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, 77-163.

Stadler, P. F., 1996 , Landscapes and Their Correlation Functions, J. Math. Chem., 20, 1-45.

Stadler, P. F. & Schnabl W., 1992, The Landscape of the Traveling Salesman Problem, Physics Letters A, 161, 337-344.

Weinberger, E. D., 1990, Correlated and Uncorrelated Fitness Landscapes and How to Tell the Difference, Biological Cybernetics, vol. 63, no. 5, 325-336.