Population based Fitness Landscapes

Dagstuhl Seminar 10361: Theory of Evolutionary Algorithms

Sébastien Verel[1,2]

http://www.i3s.unice.fr/~verel
[1]DOLPHIN team - INRIA Lille-Nord Europe (France)

[2]I3S Laboratory - University of Nice-Sophia Antipolis / CNRS (France)

September 9, 2010

Origin of the fitness landscapes model

- **Biology** : S. Wright (1932), Mid-1920
  provide imagery for his theory of speciation
  - 1 point = 1 individual : *shifting balance theory* (subpopulation, drift and selection)
  - 1 point = 1 population : *population genetics* (avg gene/fitness)

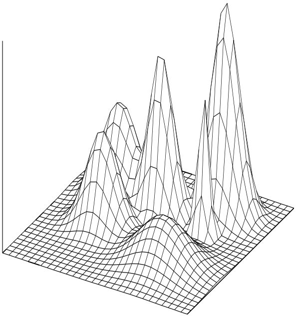
Origin of the fitness landscapes model

- **Biology** : S. Wright (1932), Mid-1920
  provide imagery for his theory of speciation
  - 1 point = 1 individual : *shifting balance theory*
    (subpopulation, drift and selection)
  - 1 point = 1 population : *population genetics* (avg gene/fitness)
- **Physics, chimistry** : 1970-80, Energy function to minimize
  - Evolution of biological macromolecule (Eigen, 1971)
  - Spin glass model of evolution (Sherrington Kirkpatrick, 1975, Binder et al, 1986, Mézard et al 1987)
  - NK-landscapes (Kauffman, Weinberger 1989)

Origin of the fitness landscapes model

- **Biology** : S. Wright (1932), Mid-1920
  provide imagery for his theory of speciation
  - 1 point = 1 individual : *shifting balance theory* (subpopulation, drift and selection)
  - 1 point = 1 population : *population genetics* (avg gene/fitness)
- **Physics, chimistry** : 1970-80, Energy function to minimize
  - Evolution of biological macromolecule (Eigen, 1971)
  - Spin glass model of evolution (Sherrington Kirkpatrick, 1975, Binder et al, 1986, Mézard et al 1987)
  - NK-landscapes (Kauffman, Weinberger 1989)
- **Combinatorial optimization** : Late 1980, 1990
  - RNA folding (Fontana, Schuster, Schnabl, Bornberg-Bauer, Stadler, Huynen, Reidys, etc. 1987...)
  - Spectral analysis, Fourrier decomposition (Stadler from 1992)
  - Genetic algorithms (Manderick 91, Whitley, Mitchel Forest Holland 92, Pruggel-Bennett Shapiro 94, Jones 95, Hordijk 96, Bornholdt 97, Barnett Harvey 98, Reeves, Greffenstette, Vassilev Miller Fogarty, Merz 99, Naudts Kallel 00, etc.)

Fitness landscape definition



Fitness landscape $(\mathcal{S}, \mathcal{N}, f)$

- $\mathcal{S}$: set of admissible solutions,
- $\mathcal{N} : \mathcal{S} \rightarrow 2^{\mathcal{S}}$: neighborhood function,
- $f : \mathcal{S} \rightarrow \mathbb{R}$: fitness function.

$\forall x \in \mathcal{S}$,

$\mathcal{N}(x) = \{y \in \mathcal{S} \mid \mathbb{P}(y = op(x)) > 0\}$

or

$\mathcal{N}(x) = \{y \in \mathcal{S} \mid \mathbb{P}(y = op(x)) > \epsilon\}$

or

$\mathcal{N}(x) = \{y \in \mathcal{S} \mid d(y, x) \leq k\}$

Goal of the fitness landscapes study

- "Geometry" (features) of fitness landscape
  $\Rightarrow$ dynamics of a local search algorithm
- Geometry is linked to the complexity of search algorithms :
  - "If there are a lot of local optima, the probability to find the global optimum is lower."
  - "If the fitness landscape is flat, discovering better solutions is rare."
- Geometry guides the design of the search algorithm :
  - "According to the probability to increase the fitness, the sample size of the neighborhood is not the same."

Study of the fitness landscape features
allows to study
the performance of a search algorithm
(and design efficient one...)

Goal and usefulness of the fitness landscapes study

1. To compare two search spaces :
   - One problem with 2 (or more) possible codings : $(S_1, \mathcal{N}_1, f_1)$ and $(S_2, \mathcal{N}_2, f_2)$ different coding, mutation operator, fitness function, etc.

   Which one is more promising to solve ?
2. To choose the algorithm :
   - analysis of global geometry of the landscape

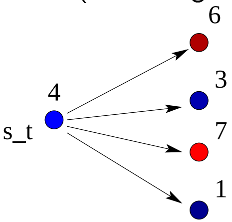
   Which algorithm can I use ?
3. To tune the parameters :
   - *off-line* analysis of structure of fitness landscape

   Which is the best mutation operator ? the size of the population ? etc.
4. To control the parameters during the run :
   - *on-line* analysis of structure of fitness landscape

   Which is the optimal mutation operator according to the estimation of structure ?

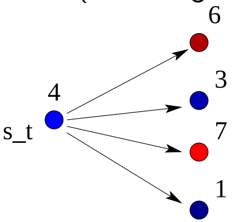
Black-box and Before putting a particular heuristic

FL = (Sol., Neighbors, Fitness)

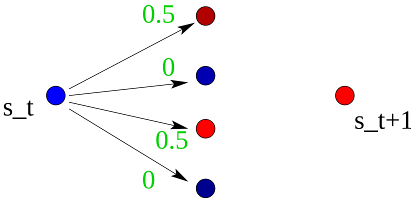


Black-box and Before putting a particular heuristic

FL = (Sol., Neighbors, Fitness)



Put prob. from your heuristic :
Markov Chain = (Sol., M)



- Sample the neighborhood to have information on local features of the search space
- From local information : deduce some global features like general shape of search space, basins structure, etc.

Limitations, drawbacks

- Problem difficulty measures :
  - Measures based on a sample can be misleading (maximum entropy hypothesis)
  - A predictive measure solves a decision problem (between P and NP, worth case)

  ⇒ Compare the features of search spaces
  before studying "the" "difficulty"
- Gap between Solution and Population :
  - most works on the fitness landscapes are based on mutation of one solution
- Crossover operator :
  - few FL analysis as mutation fitness landscapes analysis

Solution *vs.* Population : Markov chain

$$P \overset{W}{\longrightarrow} P$$

- P : population of solutions
- W : operator, one iteration of the EA : selection, mutation, crossover and replacement.

- First time hitting : difficult to compute

Solution *vs.* Population : Markov chain

$$
\begin{array}{ccc}
P & \xrightarrow{W} & P \\
p \downarrow & & \downarrow p \\
X & \xrightarrow[T]{} & X
\end{array}
$$

where "$p \circ W = T \circ p$" and T has some good properties (drift condition, etc.)

- P : population of solutions
- W : operator, one iteration of the EA : selection, mutation, crossover and replacement.

- First time hitting : difficult to compute
- Use (find) a better space with a projection

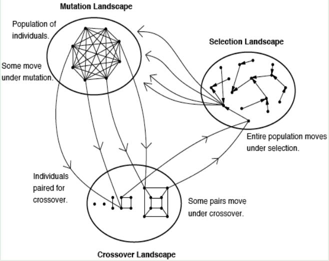
$\Rightarrow$ Fitness landscapes analysis should also work on population

Solution *vs.* Population : T. Jones (1995)

$\text{FL} = (R, \Phi, f, \mathcal{F}, \leq_{\mathcal{F}})$

- R : representation space (search space)
- $\mathcal{M}(R)$: multiset on R
- $\Phi : \mathcal{M}(R) \rightarrow \mathcal{M}(R)$ stochatic operator on *population* (neighborhood)
- $(\mathcal{F}, \leq_{\mathcal{F}})$ partiel order set (objectif space)
- $f : \mathcal{M}(R) \rightarrow (\mathcal{F}, \leq_{\mathcal{F}})$: fitness function (but not necessarily $\mathcal{F} = \mathbb{R}$)

"One operator = one landscape"



- Crossover HC stops when $max(f_i', f_j') < max(f_i, f_j)$
- Difficult links with pop.
- No crossover FL analysis

Previous works on crossover landscapes

- **Operator correlation** (Manderick 91) :
  correlation of fitness between parents and children.

  No direct links between fitness of parents

- **Couple of solution** (T. Jones 95) : see previous slide
- **Crossover landscapes for the oneMax** (Horn Reeves 96) :
  crossover with the complementary bit string.

  Isomorphism with mutation FL

- **Hypergraph** (Gritchov) **P-Structure** (Stadler Wagner 96) :
  crossover with random solution.

  Random walk, autocorrelation, algebric approach,
  decomposition into a superpostion of elementary landscapes

Population based fitness landscapes

Population based Fitness Landscape (Σ, N, Φ_p)

- Σ is the set of multiset of μ admissible solutions,
$$\Sigma = \{\sigma \subset \mathcal{S} \ : \ |\sigma| = \mu\}, \text{ so } |\Sigma| = |\mathcal{S}|^\mu / \mu!$$
- $N : \Sigma \longrightarrow 2^\Sigma$ a neighborhood structure is a function that assigns to every $\sigma \in \Sigma$ a set of neighbors $N(\sigma)$.
- $\Phi_p : \Sigma \longrightarrow \mathbb{R}$ is the population based fitness function : $\forall \sigma = \{s_1, \ldots, s_\mu\} \in \Sigma$
$$\Phi_p(\sigma) = \left(\sum_{i=1}^{\mu} f(s_i)^p \right)^{1/p}$$

Solution based Fitness Landscape $(\mathcal{S}, \mathcal{N}, f)$:

- $\mathcal{S}$: set of admissible solutions,
- $\mathcal{N} : \mathcal{S} \rightarrow 2^{\mathcal{S}}$: neighborhood function,
- $f : \mathcal{S} \rightarrow \mathbb{R}$: fitness function.

Fitness function for population

Fitness function

$$\forall \sigma = \{s_1, \dots, s_\mu\} \in \Sigma$$

$$\Phi_p(\sigma) = \left(\sum_{i=1}^{\mu} f(s_i)^p \right)^{1/p}$$

Li and Fang, "On the Entropic Regularization Method for Solving Min-Max Problems with Applications", Mathematical Methods of Operations Research (1997) 46 :119-130.

p=1 → p=oo

sum → max

- $\Phi_1(\sigma) = \sum_{i=1}^{\mu} f(s_i)$
  - Information of the most interesting solutions is lost
  - Differentiable
- Φ_p : Differentiable in f_i
- $\lim_{p \to \infty} \Phi_p(\sigma) = \max\{f(s_1), \dots, f(s_\mu)\}$
  - Nearly all solutions are "silent"
  - Non differentiable
  - Solve $\Phi_\infty \Rightarrow$ solve f

Neighborhood N

$\sigma^{'}$ neighbor of $\sigma = \{s_1, \ldots, s_\mu\}$

- Mutation of one solution :
  $\sigma^{'} = \{s_1, \ldots, s_{i-1}, s_i^{'}, s_{i+1}, \ldots, s_\mu\}$ with $s_i^{'} \in \mathcal{N}(s_i)$
- Mutation of several solutions with rate r :
  $\sigma^{'} = \{s_1, \ldots, s_{i_1}^{'}, \ldots, s_{i_2}^{'}, \ldots, s_{i_3}^{'}, \ldots, s_\mu\}$ with $s_{i_k}^{'} \in \mathcal{N}(s_{i_k})$
- Crossover between 2 solutions :
  $\sigma^{'} = \{s_1, \ldots, s_i^{'}, \ldots, s_j^{'}, \ldots, s_\mu\}$ with $(s_i^{'}, s_j^{'}) = \mathcal{X}(s_i, s_j)$
- Crossover between several solutions
- Apply the selection operator :
  $N(\sigma) = \{\sigma^{'} \in \Sigma \mid \mathbb{P}(\sigma^{'} = select(\sigma)) > 0\}$

Neighborhood : Mutation of one random solution

$\sigma = \{s_1, \dots, s_\mu\}$, and $\sigma^{'} \in N(\sigma)$ such that :
$\exists i \in [1, \mu]$, $\sigma^{'} = \{s_1, \dots, s_{i-1}, s_i^{'}, s_{i+1}, \dots, s_\mu\}$ with $s_i^{'} \in \mathcal{N}(s_i)$

Local optima

σ is loc. opt. for (Σ, N, Φ_p)
$\Leftrightarrow \forall i \in [1, \mu]$, s_i is loc. opt. for $(\mathcal{S}, \mathcal{N}, f)$

Evolvability

if $f(s_i^{'}) = (1 + \delta) f(s_i)$
then $\Phi_p(\sigma^{'}) = \left(\Phi_p(\sigma^{'})^p + \sum_{k=1}^{p} \binom{p}{k} \delta^k f(s_i)^p \right)^{1/p}$

Mutation of one random solution on oneMax problem

Average fitness in the neighborhood

$p = 1$ (sum)

$$E[\Phi_1(\sigma^{'})] = (1 - \frac{2}{\mu n})\Phi_1(\sigma) + 1$$

(like oneMax of size μn)

$p = \infty$ (max)

Let be $M = \Phi_\infty(\sigma) = \max\{f(s_1), \ldots, f(s_\mu)\}$, $k = \sharp\{i : f(s_i) = M\}$
if $k = 1$ then

$$E[\Phi_\infty(\sigma^{'})] = (1 - \tfrac{2}{\mu n})\Phi_\infty(\sigma) + \tfrac{1}{\mu}$$

if $k > 1$ then

$$E[\Phi_\infty(\sigma^{'})] = (1 - \tfrac{k}{\mu n})\Phi_\infty(\sigma) + \tfrac{k}{\mu}$$

NK fitness landscapes : ruggedness and epistasis

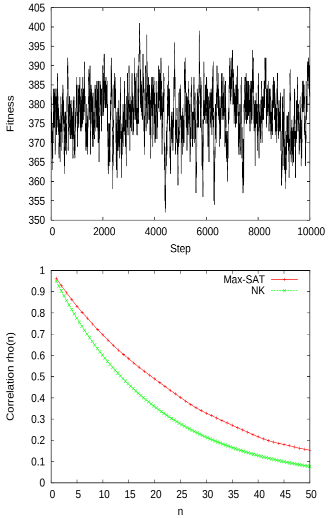
NK-landscapes : Model of problems

N size of the bit-strings
K from 0 to $N-1$, NK landscapes can be tuned from smooth to rugged :

- $K = 0$ no correlations, f is an additive function, and there is a single maximum
- $K = N-1$ landscape completely random, the expected number of local optima is $\frac{2^N}{N+1}$
- Intermediate values of K interpolate between these two extreme cases and have a variable degree of epistasis (i.e. gene interaction)

- $N = 128$, $K \in \{2, 4, \ldots, N-2, N-1\}$
- 30 random instances for each case

Rugged/smooth FL by autocorrelation functions



Autocorrelation of time series of fitnesses $(f(s_1), f(s_2), \ldots)$ along a random walk $(s_1, s_2, \ldots)$ (Weiberger 90) :

$$\rho(n) = \frac{E[(f(s_i) - \bar{f})(f(s_{i+n}) - \bar{f})]}{var(f(s_i))}$$

autocorrelation length

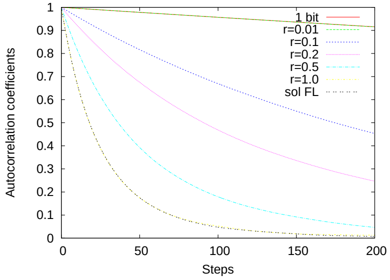
$\tau = \frac{1}{\log(\rho(1))}$

- small τ : rugged landscape
- long τ : smooth landscape

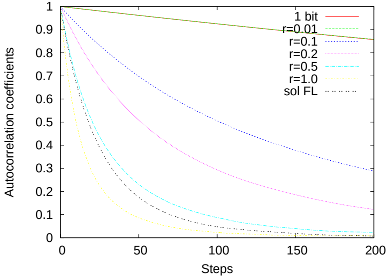
1 bit-flip on several solutions with rate r ($\mu = 100$)

Autocorrelation functions

$p = 2$



$p = 30$

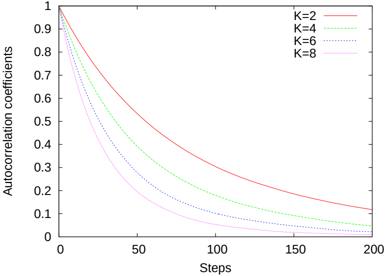


- Ruggedness increases with mutation rate
- When p small, $r = 1.0$ gives same ruggedness as solution based FL
- When p large, the ruggedness of $r = 1$ is larger than solution based FL

1 bit-flip on several solutions with rate r

Influence of epistasis K

$p = 2$ and $r = 0.5$

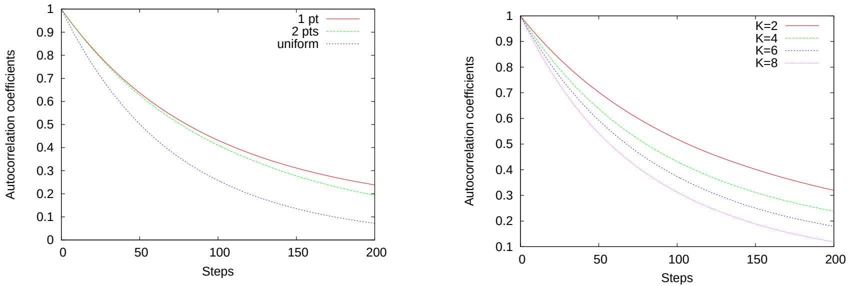


- Ruggedness increases with epistasis K
- Same result as solution based FL

1 crossover between 2 solutions

Autocorrelation functions ($p = 2$)

$K = 4$ 1-point crossover



- Uniform more rugged than 2-point more rugged than 1-point
- Ruggedness for crossover increases with epistasis
- But performances of uniform $\geq$ 2-point $\approx$ 1-point (diversity rather BB)

A problem with different perf. according to crossover...

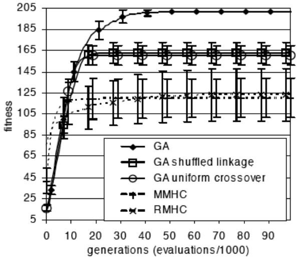
R. Watson and T. Jansen, "A building-block royal road where crossover is provably essential", GECCO 07.

Build-Block Royal Road

Royal Road where sub-blocks (of size k) have several local optima :
$F(G) = \sum_{i=1}^{B} f(g_i)$ where $f(g_i) = \sum_{j=1}^{T_i} c(g_i, t_{ij})$, and

$$c(g_i, t_{ij}) = \begin{cases} w_{ij} & \text{if } d(g_i, t_{ij}) = 0 \\ (1 + d(g_i, t_{ij}))^{-1} & \text{otherwise} \end{cases}$$

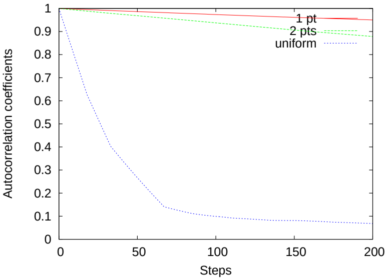
Theorem 1. The multi-deme EA with g demes, one receiving deme, with deme size μ, crossover probability p_c, and epoch length e, finds the unique global optimum of F within $O(\mu e g N^2 \log N)$ function evaluations with probability $1-O(1/N)$ if $g \geq c \log\sqrt{N}$ and $\xi \leq p_c \leq 1-\xi$ hold for some constants $\xi > 0$ and $c > 1$ sufficiently large.



1 crossover between 2 solutions on BB RR

Autocorrelation functions

$p = 2$



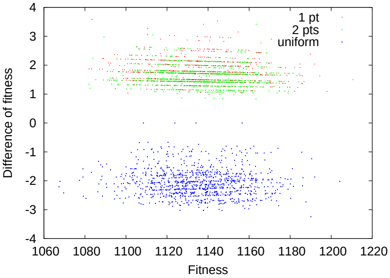
- 1-point and 2-point are highly correlated
- Uniform : rugged operator
- Expected result !
- Correspond to performances of different algorithms

1 crossover between 2 solutions on BB RR

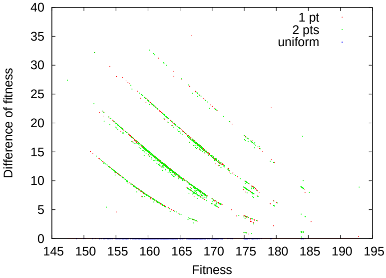
Fitness cloud (parent fitness vs. children fitness)

- Large neighborhood such as GP problems
- From a population σ of local optima (for f), compute the difference of fitness Φ_p between the best neighbor from $\mu(\mu-1)/2$ random neighbors and the fitness $\Phi_p(\sigma)$ the population

$p = 2$



$p = 30$



Conclusion, Discussion

- This definition of population based Fitness Landscapes could work

Some possible studies

- Relation between diversity and evolvability (in Φ_p)
- Use standart tools of fitness landscapes analysis : autocorrelation, length of adaptive walk, FDC, Local Optima Network, etc.
- Properties on sol. FL report on pop FL ? (elementary landscapes, FDC, evolvability, etc.)
- Online control of parameters : Bandit (DMAB) with Φ_p
- Why not optimize directly Φ_p with tabu, SA, etc.