

Fitness Landscapes and Problem Hardness in Evolutionary Computation



Leonardo Vanneschi

Dipartimento di Informatica, Sistemistica
e Comunicazione (DISCo)

University of Milano-Bicocca, Italy

vanneschi@disco.unimib.it

<http://personal.disco.unimib.it/Vanneschi>



Sébastien Verel

Laboratoire I3S

University of Nice-Sophia Antipolis /CNRS, France

verel@i3s.unice.fr

<http://www.i3s.unice.fr/~verel>

Dishes of the day

• Part I - Fitness Landscapes (Sébastien Verel)



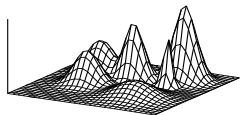
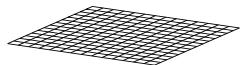
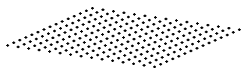
- Definition of Fitness Landscape
- Types of Fitness Landscapes
 - Multimodal
 - Rugged
 - Neutral
- Measures to quantify multimodality and ruggedness
- Measures to study neutrality

• Part II - Problem Difficulty in GP (Leonardo Vanneschi)



- Binding between Fitness Landscapes and Problem Difficulty
- Measures of difficulty applied to GP

Fitness landscape

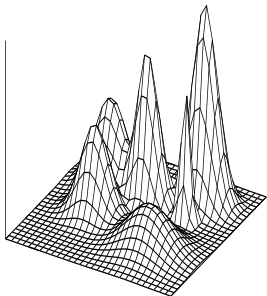


from biological science (Wright 1930 [30]) :
Modelisation of species evolution

Used to model dynamical systems :

- statistical physics,
- molecular evolution,
- ecology, etc

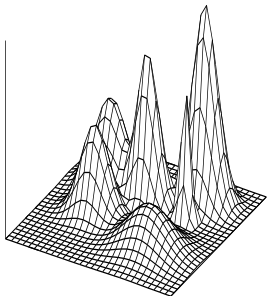
In combinatorial optimization



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

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

Definition of neighborhood



$\mathcal{V} : \mathcal{S} \rightarrow 2^{\mathcal{S}}$: neighborhood function

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

$$\mathcal{V}(x) = \{y \in \mathcal{S} \mid y = op(x)\}$$

$$\mathcal{V}(x) = \{y \in \mathcal{S} \mid d(y, x) \leq 1\}$$

Goal of the fitness landscapes study

- the "geometry" of fitness landscape corresponds to the dynamic of a local search algorithm
- the geometry is linked to the problem hardness :
 - probability or time to have a fitness level for a given local heuristic

Study of the geometry of the landscape allows to study the difficulty.

Two main geometries :

- multimodal / ruggedness
- neutral

Multimodal Fitness landscapes

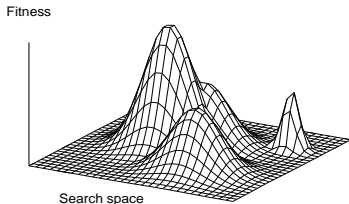
Local optima :

no neighbor solution with higher fitness value

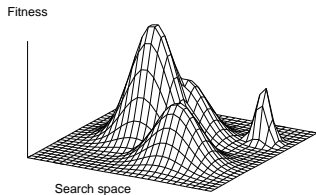
Adaptive walk : (s_0, s_1, \dots) where $s_{i+1} \in \mathcal{V}(s_i)$ and $f(s_i) < f(s_{i+1})$

Attraction basin of s_{opt} :

set of solutions of adaptive walks to s_{opt}



Multimodal Fitness landscapes

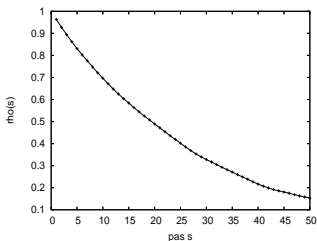
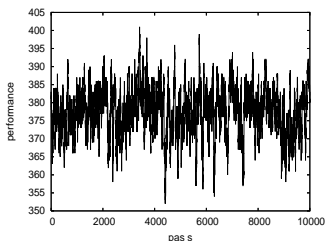


Optimisation difficulty :
number and size of attractive
basins (Garnier *et al* [8])

The idea :

- if the size of attractive basin of global optima is small
- the problem is difficult to optimize

Rugged/smooth fitness landscapes



Autocorrelation of $(f(s_1), f(s_2), \dots)$ along a random walk (s_1, s_2, \dots) (Weinberger 1990 [29]) :

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

autocorrelation length $\tau = \frac{1}{\rho(1)}$

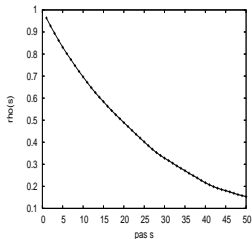
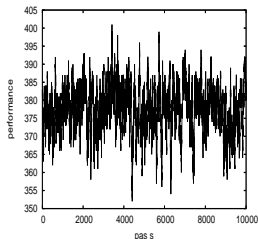
- small τ : rugged landscape
- long τ : smooth landscape

Results on rugged fitness landscapes (Stadler 96 [23])

Problem	parameter	$\rho(1)$
symmetric TSP	n number of towns	$1 - \frac{4}{n}$
anti-symmetric TSP	n number of towns	$1 - \frac{4}{n-1}$
Graph Coloring Problem	n number of nodes α number of colors	$1 - \frac{2\alpha}{(\alpha-1)n}$
NK landscapes	N number of proteins K number of epistasis links	$1 - \frac{K+1}{N}$

Ruggedness decreases with the size of those problems :
 small variation has less effect on the fitness values

Multimodality, ruggedness, epistasis ?



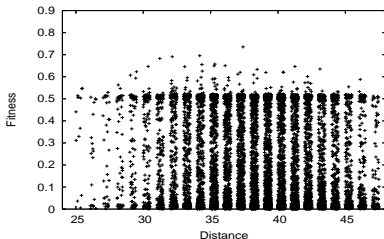
- multimodality/ruggedness :
 conjecture (Stadler 92 [24], Garcia 97 [7]) :
 on average, 1 local optimum per
 sphere of rayon τ
 - epistasis/ruggedness :
 NK fitness landscapes (Kauffman [14])

$$\tau = \frac{-1}{\ln(1 - \frac{K+1}{N})}$$
 et
$$d = \frac{N \log_2(K+1)}{2(K+1)}$$
 - But some counterexamples...

$$\sum_{i=1}^N \exp(i)x_i$$
 et
$$\prod_{i=1}^N x_i$$
- ... open question ?

Fitness Distance correlation (FDC) (Jones 95 [13])

Correlation between distance to global optimum and fitness



Experimental scale :

- $\rho < -0.15$, easy optimization
- $\rho > 0.15$, hard optimization
- $-0.15 < \rho < 0.15$, undecided zone

Fitness landscape and crossover ?

- Crossover of complementary strings (Culberson 94 [6]) : comparison with mutation-landscape
- Generalisation of graph theory :
 - Hypergraph of Gitchoff [9]
 - P-structure of Stadler [25]
- crossover with a random solution
- Space of pair of solutions (Jones, Defoin [19] : study of linear GP crossover)

Maybe the schemata theorem or the study of distances between pairs give better results?...

Synthesis

- Metaphor from the biology
- Study of multimodal fitness landscapes :
→ optimization algorithms (SA, Tabu Search, Island Model...)

Goal of study of fitness landscapes :

- Links with problem hardness : to make better choices (mapping, fitness function, operators), design of algorithms, etc.
- Proof of convergence (with speed of convergence)
- Autoadaptation of parameters of research

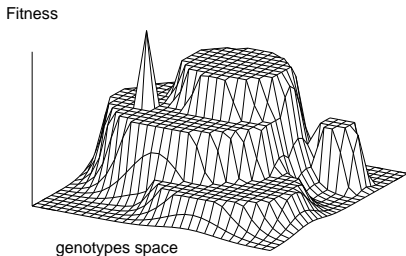
Limits :

- 1 operator = 1 landscape ?
- Crossover ? and link with population ?
- Edges : usefull information ?

Neutral Fitness Landscapes

Neutral theory (Kimura \approx 1960 [15])
Theory of mutation and random drift

A considerable number of mutations have no effects on fitness values



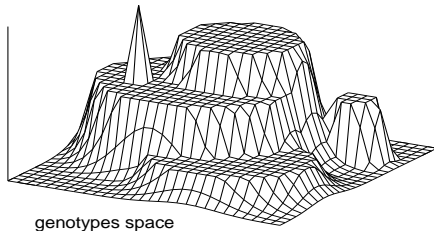
- plateaus
- neutral degree
- neutral networks
[Schuster 1994 [22],
RNA folding]

Neutral fitness landscapes

Combinatorial optimization

- Redundant problem (symetries, ...) (Goldberg 87 [10])
- Problem “not well” defined or dynamic environment (Torres 04 [12])

Fitness



Applicative problems :

- Robot controller
- Circuit design
- genetic programming
- Protein Folding
- learning problems

What optimizers do with neutrality?

Three possibilities :

- Decrease the neutrality
- Use a specific metaheuristic
- Increase the neutrality with redundant genotype/phenotype mapping

Decreasing the neutrality (minLA : E. Rodriguez, PPSN05 [20])

Redundant encoding is a drawback, lack of information
 graph $G = (V, E)$: labeled each nodes

$$LA(G, \varphi) = \sum_{(u,v) \in E} |\varphi(u) - \varphi(v)| \in \mathbb{N}$$

“LA represents a potential drawback because different linear arrangements can result in the same total edge length. This incomplete information can prevents the search process from finding better solution.”

$$\phi(G, \varphi) = LA(G, \varphi) + I_{norm}(G, \varphi)$$

with $I_{norm}(G, \varphi) \in [0, 1]$

- I_{norm} is higher when the differents could be optimized
- I_{norm} makes different between equivalent labelling

Use a specific metaheuristic

Neutrality of the problem can not be changed

- Netcrawler (L. Barnett [2])

step \leftarrow 0

Choose initial solution $s \in \mathcal{S}$

repeat

 choose $s' \in \mathcal{V}(s)$ randomly

if $f(s) \leq f(s')$ **then**

$s \leftarrow s'$

end if

 step \leftarrow step + 1

until stepMax \leq step

Good results on ϵ -correlated landscapes :

low probability to find a better solution

high probability to find a solution with same fitness

Use a specific metaheuristic

Neutrality of the problem can not be changed

- Extrema selection (Stewart 2001 [26]) :

“It is sufficient to recognise that the neutrality of a fitness function may be a significant issue when evolving solutions. With this in mind, the remainder a novel modification to the standard GA which is specifically designed to take in advantage of Neutral Network”

When the solutions are in the same plateaus (at 90% from best solution)

→ selection according to the distance from the centroide of the population

Increase the neutrality of the landscape with a redundant coding

Escape from local optima

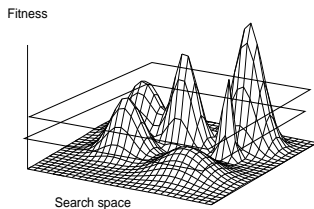
- crossing a Barrier (Nimwegen et Crutchfield 99 [18]) :
 - fitness barrier : decreases the fitness by croosing a valley
 - Entropy barrier : lack of information on a plateaus
- en Cartesian GP (Vassilev *et al* 00 [27]) :

“(...) the role of landscape neutrality for adaptive evolution is to provide a path for crossing landscape regions with poor fitness.”
- Duality (Collard, Clergue 00 [5]) : add one bit and use a specific operator
 $f(x_0) = f(x)$, $f(x_1) = f(\bar{x})$ et $op(x_1) = \bar{x}_0$

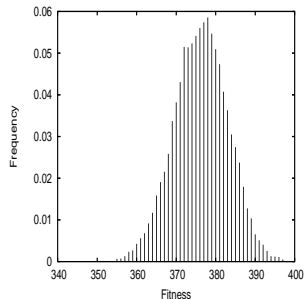
What do we do ?

- In our knowledge, there is no definitive answer about neutrality / problem hardness
 - Certainly, it is depend on the nature of neutrality of the fitness landscape
- ⇒ A better description of the geometry of neutral fitness landscapes is needed

Neutral sets



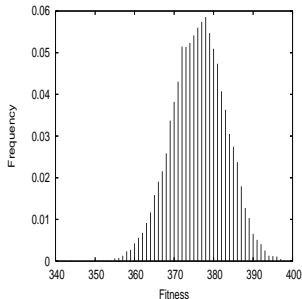
Set of solution with fitness value



Density of state (D.O.S.)

- Introduce in physic (Rosé 1996 [21])
- Optimization (Belaidouni, Hao 00 [4])

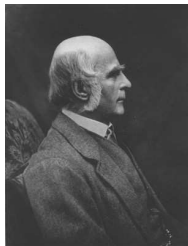
Neutral sets



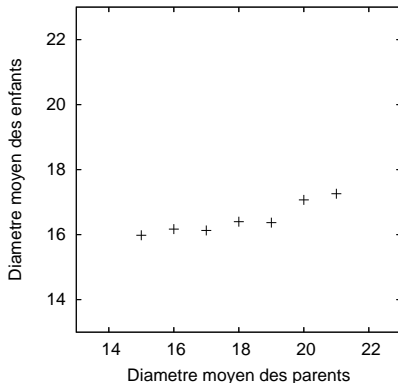
Density of state (D.O.S.)

- Performance of random search
- the tail of the distribution is an indicator of difficulty :
 - the faster the decay, the harder the problem
- But do not care about the neighborhood relation

Make a roundabout with Sir F. Galton

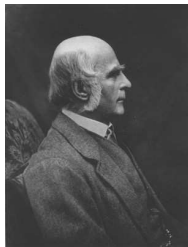


(1822 - 1911)

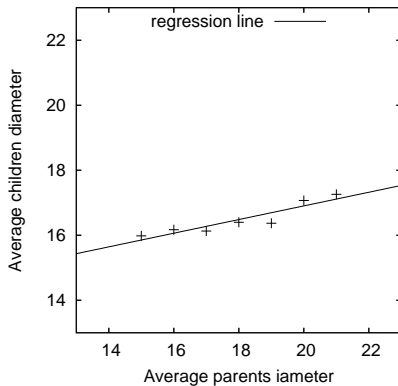


Original data of sweet peas sizes (1877)

Make a roundabout with Sir F. Galton

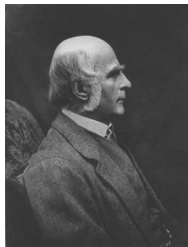


(1822 - 1911)

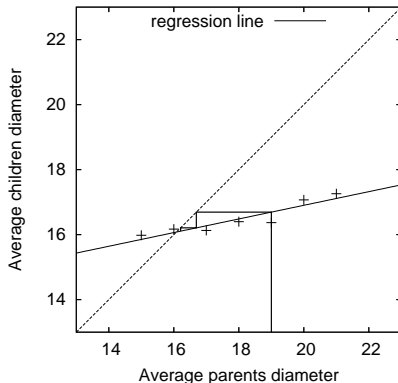


Original data of sweet peas sizes (1877)

Make a roundabout with Sir F. Galton



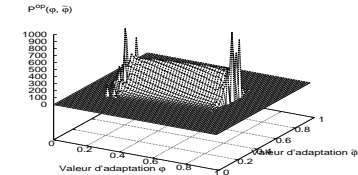
(1822 - 1911)



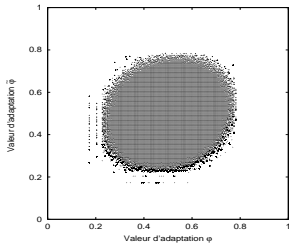
Original data of sweet peas sizes (1877)

Fitness Cloud

Combinatorial optimization



$$X(s) = f(s)$$
$$Y(s) = f(op(s))$$

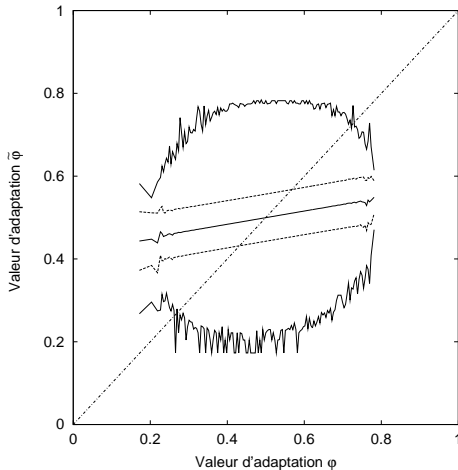


Fitness Cloud de op :

conditional probability density
function of Y given X

Fitness cloud

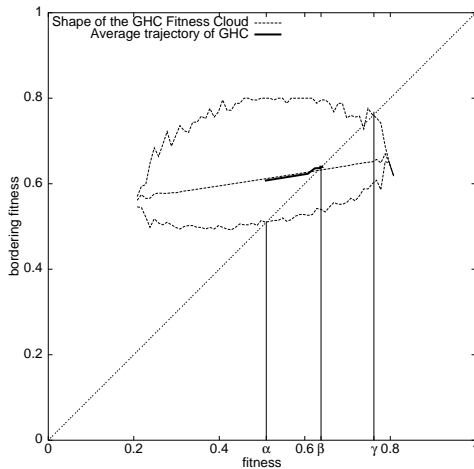
Measure of evolvability



- Probability of finding better solutions
- Average fitness of better neighbor solutions
- Average and standart deviation of fitnesses

Fitness cloud

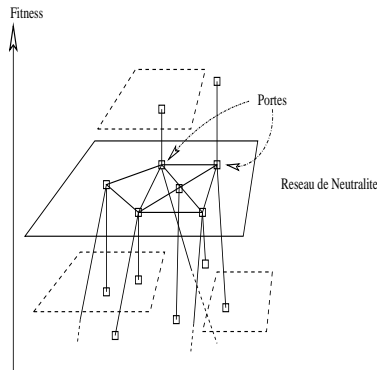
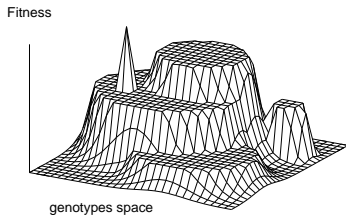
Prediction of evolution (CEC 2003)



Neutral fitness landscapes

- Neutral sets : set of solution with same fitness \rightarrow no structure
- Fitness cloud : neighborhood relation between neutral sets
- Introduction of neighborhood structure on neutral sets \rightarrow Neutral Networks

Neutral networks (Schuster 1994 [22])



Definitions

- A *test of neutrality* is a predicate

$$isNeutral : S \times S \rightarrow \{true, false\}$$

For example, $isNeutral(s_1, s_2)$ is *true* if :

- $f(s_1) = f(s_2)$.
 - $|f(s_1) - f(s_2)| \leq 1/M$ with M is the search population size.
 - $|f(s_1) - f(s_2)|$ is under the evaluation error.
- The *neutral neighborhood* of s is the set of neighbors which have the same fitness $f(s)$

$$\mathcal{V}_{neut}(s) = \{s' \in \mathcal{V}(s) \mid isNeutral(s, s')\}$$

- The *neutral degree* of a solution is the number of its neutral neighbors

$$nDeg(s) = \#(\mathcal{V}_{neut}(s) - \{s\}).$$

Definitions

- A *neutral walk* : $W_{neut} = (s_0, s_1, \dots, s_m)$
for all $i \in [0, m - 1]$, $s_{i+1} \in \mathcal{V}(s_i)$
for all $(i, j) \in [0, m]^2$, $isNeutral(s_i, s_j)$ is true.
- A *Neutral Network* : graph $G = (V, E)$
 $V \subset \mathcal{S}$: for all s and s' from V , there is a neutral walk W_{neut}
belonging to V from s to s' ,
Two vertices are connected by an edge of E if they are neutral
neighbors.

*A fitness landscape is neutral
if there are many solutions with high neutral degree.*

Measures on neutral fitness landscapes

To introduce measure of neutrality,
we will use three possible families of neutral fitness :

- based on NK fitness landscapes :
 \mathcal{S} : bit strings of length N ,
$$f(x) = \frac{1}{N} \sum_{i=1}^N f_i(x_i, x_{j_1}, \dots, x_{j_K})$$
- corresponding to three possible ways to introduce redundancy in additive fitness functions
- two parameters :
one for non-linearity (epistasis K), one for neutrality
- the measures could be analyse with the parameters and supposed difficulty

Neutral NK fitness landscapes

NK (Kauffman 1993)

$$f(s) = \frac{1}{N} \quad (\quad 0.02 \quad + \quad 0.31 \quad + \quad 0.91 \quad + \dots + \quad 0.20 \quad)$$

NK_q (Newmann *et al* 1998) [17] : q values for the terms

$$f(s) = \frac{1}{N(q-1)} \quad (\quad 1 \quad + \quad 3 \quad + \quad 3 \quad + \dots + \quad 0 \quad)$$

NK_M (Lobo 2004 [16]) : M different values

$$f(s) = \frac{1}{N.M} E[M. \quad (\quad 0.02 \quad + \quad 0.31 \quad + \quad 0.91 \quad + \dots + \quad 0.20 \quad)]$$

NK_p (Barnett 1998 [1]) : prob. p to have 0

$$f(s) = \frac{1}{N} \quad (\quad 0.02 \quad + \quad 0.31 \quad + \quad 0 \quad + \dots + \quad 0.20 \quad)$$

Intra network Measures

Classical measures of graph to describe NN :

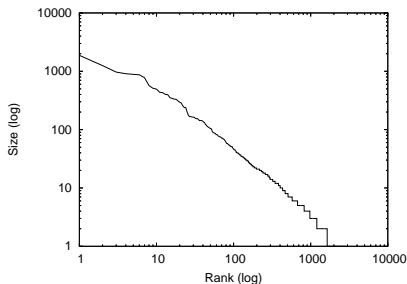
- 1 the *size* : number of nodes of NN,
- 2 neutral degree distribution :
measure of the quantity of “neutrality”
- 3 Autocorrelation of neutral degree during neutral walk (Bastolla 03 [3]) :
comparaison with random graph,
measure of the correlation structure of *NN*

Intra network Measures

Size

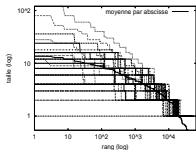
Classical measures of graph to describe NN :

- 1 The size : number of nodes of NN,
rank-size of NN in log-log :

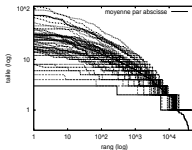


Frequency of apparition of a word in a text (Ziff law 1960)

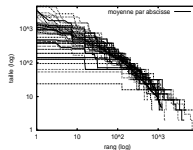
Size of neutral networks



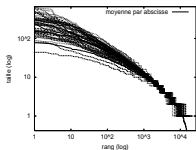
$K = 1 \quad q = 10$



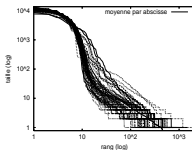
$K = 1 \quad M = 160$



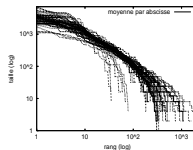
$K = 1 \quad p = 0.8$



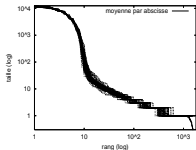
$K = 2 \quad q = 4$



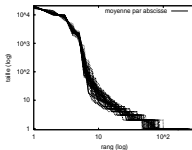
$K = 2 \quad M = 32$



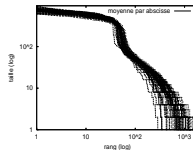
$K = 2 \quad p = 0.9$



$K = 5 \quad q = 2$



$K = 5 \quad M = 16$



$K = 5 \quad p = 0.95$

Size of neutral networks

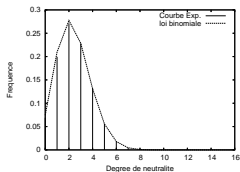
- When epistasis and neutrality increase, the random variation of distribution of size decreases
- Those neutral fitness landscapes are dominated by few and large neutral networks :
maybe, could be prove with percolation theory...
→ important information to design search algorithms
- The Ziff law is approaching only when epistasis and neutrality are low

Intra network Measures

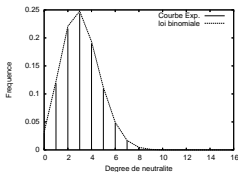
Classical measures of graph to describe NN :

- 1 the *size* : number of nodes of NN,
- 2 neutral degree distribution :
measure of the quantity of “neutrality”
- 3 Autocorrelation of neutral degree during random neutral walk
(Bastolla 03 [3]) :
comparaison with random graph,
measure of the correlation structure of *NN*

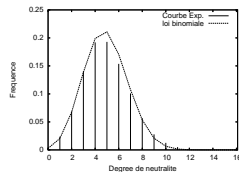
Distribution of neutral degrees ($N = 16, K = 2$)



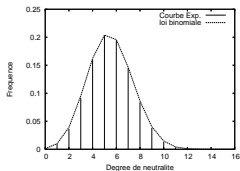
$q = 4$



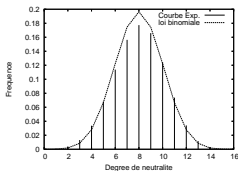
$M = 48$



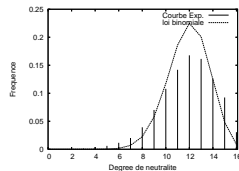
$p = 0.8$



$q = 2$



$M = 16$



$p = 0.95$

Experimental distribution (impulse), binomial distribution (line).

Distribution of neutral degrees

- Barnett (98) gives the probability of neutral mutation for NK_p -landscapes :

$$p_{neutr} = p^2 \left(1 - \frac{K}{N-1} (1 - p^2)^{N-1}\right) \approx p^2 e^{-K(1-p^2)}$$

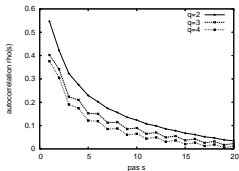
- For NK_q and NK_M landscapes the distribution is nearly a binomial distribution

Intra network Measures

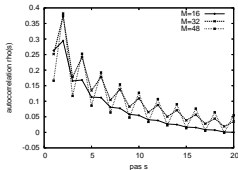
Classical measures of graph to describe NN :

- 1 the *size* : number of nodes of NN,
- 2 neutral degree distribution :
measure of the quantity of “neutrality”
- 3 Autocorrelation of neutral degree during neutral walk (Bastolla 03 [3]) :
comparaison with random graph,
measure of the correlation structure of *NN*

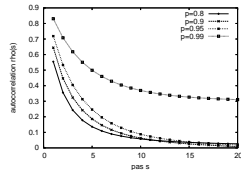
Autocorrelation of neutral degrees



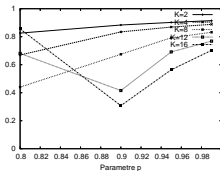
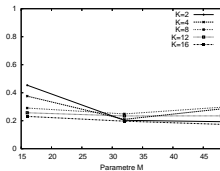
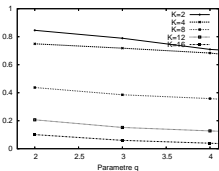
NK_q land.



NK_M land.



NK_p land.

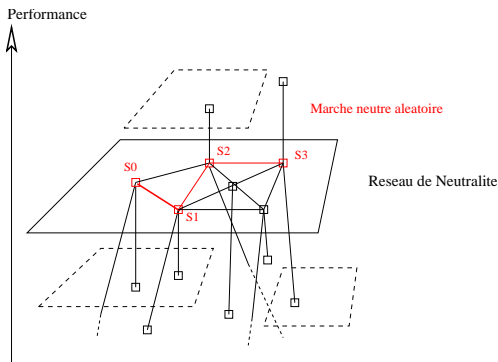


Autocorrelation coefficient of order 1

Autocorrelation of neutral degrees

- Neutral networks are not random graph
- NK_M landscapes (with M fitness values) is more closely to random graph
- epistasis parameter (K) has more influence on the structure of neutral networks than neutrality parameter
- \rightarrow important to design search algorithm

Inter networks measures



- 1 rate of innovation (Huynen 96 [11]) :
The number of new accessible structures (fitness) per mutation
- 2 Autocorrelation of evolvability [28] :
autocorrelation of the sequence $(evol(s_0), evol(s_1), \dots)$.

Rate of innovation on neutral NK landscapes

The number of new accessible structures (fitness) per mutation

- No information on neutral NK fitness landscapes :
 - No link with the parameters
 - difficult to estimate for the whole fitness landscape

Autocorrelation of maximal evolvability (Verel 06 [28])

Evolvability : ability to evolve

Definition

Autocorrelation of evolvability is the autocorrelation function of the serie $(evol(s_0), evol(s_1), evol(s_2), \dots)$ where (s_0, s_1, s_2, \dots) is neutral random walk on a neutral network and $evol$ is a measure of evolvability of a solution.

Measure of evolvability :

- Probability to have fitter solution in the neighborhood
- maximum evolvability : the fitness of best solution in the neighborhood

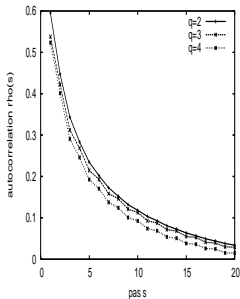
Autocorrelation of maximal evolvability (Verel 06 [28])

- if correlation is high, then the neutral networks are not randomly distributed over the fitness landscapes.
The problem is easier to optimize than...
- if the correlation is low, the neutral networks are randomly distributed over the fitness landscapes

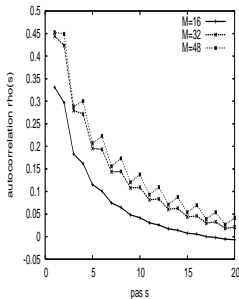
→ and this information could be introduced into a search algorithm

Autocorrelation of maximal evolvability (Verel 06 [28])

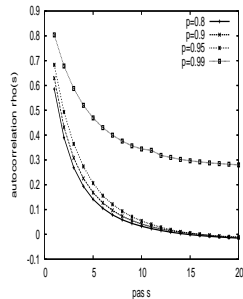
maximal evolvability autocorrelation function for $N = 16$, $K = 2$



NK_q

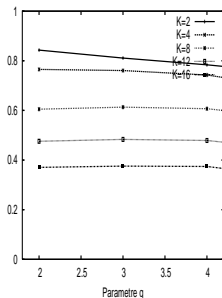


NK_M

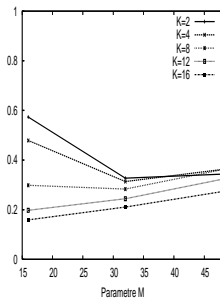


NK_p

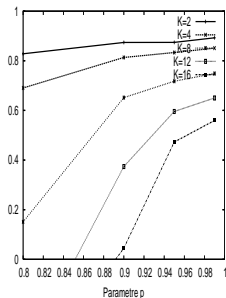
Autocorrelation of maximal evolvability



NK_q



NK_M



NK_p

autocorrelation coefficient of order 1 for $N = 64$

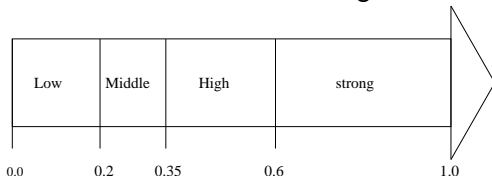
Autocorrelation of maximal evolvability

- Neutral networks are not randomly distributed
 - NK_M landscapes (with M fitness values) is more "random"
 - epistasis parameter (K) has more important than neutrality parameter
- take care to design search algorithm !

Synthesis

Measures

- neutral degrees distribution :
“How neutral is the fitness landscape ?”
- Autocorrelation of neutral degrees : network “structure”

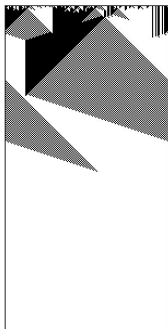


- rank-size of NN in log-log :
well adapted representation (complex systems, percolation)
- rate of innovation :
low information for combinatorial optimization
- Autocorrelation of maximal evolvability :
information on the links between NN

Presentation

majority or *density* task : two-state CA

Does the initial state contain more 0s than 1s?



$\rho_0 < 0.5$



$\rho_0 > 0.5$

- Difficult problem : coordination among the automata
- Paradigm of the phenomenon of *emergence* in complex systems.

Definition

Finite CA of size N ($N = 149$), radius $r = 3$ (2^{128} rules)

ρ_0 be the fraction of 1s in the *Initial Configuration* (IC).

- If $\rho_0 > 1/2$ then the CA must relax to $(1)^N$
- If $\rho_0 < 1/2$ then the CA must relax to $(0)^N$

after $M = 2N$ time steps

Standard performance :

fraction of correct classifications over $n = 10^4$ randomly chosen ICs.

Binomial distribution sampled : each bit is independently drawn with probability $1/2$ of being 0

Best Rules Known

- No CA can perform the task perfectly [Land 95]
- Finding a good rule is a hard Problem

Best rules know :

GKL (1978), By hand, 0.815 005F005F005F005F005FFF5F005FFF5F	Das (1996), By hand, 0.823 009F038F001FBF1F002FFB5F001FFF1F
Davis (1996), By hand, 0.818 070007FF0F000FFF0F0007FF0F310FFF	ABK (1996), Gen. Prog, 0.824 050055050500550555FF55FF55FF55FF
Coe1 (1998), coevol GA, 0.851 011430D7110F395705B4FF17F13DF957	Coe2 (1998), coevol GA, 0.860 1451305C0050CE5F1711FF5F0F53CF5F

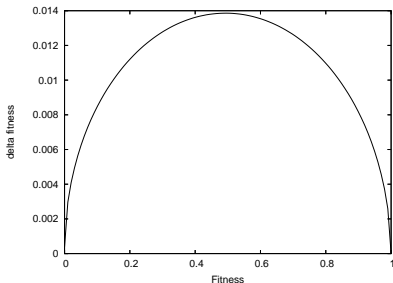
- No investigations of the difficulty of this fitness landscape
- ⇒ **Goal** : To statistically quantify the degree of difficulty

Neutrality in Majority Problem

Standard performance : error of evaluation due to random variation of samples of ICs.

ICs are chosen independently,

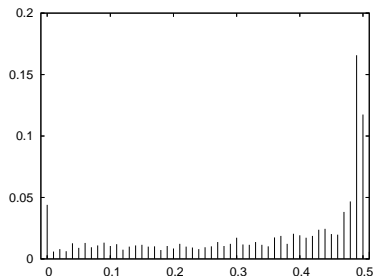
fitness value f follows a normal law $\mathcal{N}(f, \sqrt{\frac{f(1-f)}{n}})$



isNeutral(s, s') is true
if t-test accepts the hypothesis of
equality of $f(s)$ and $f(s')$ with
95% of confidence.

Density Of States

Sampling	Random	Metropolis-Hastings
Fitness = 0	3979	176
Total	4000	4000



Metropolis-Hastings sampling :

- a considerable number of solutions sampled with a fitness ≈ 0.5
- No solution with a fitness value superior to 0.55

Study of Neutral Networks

Study of two important large neutral networks :

- $NN_{0.5}$: fitness around 0.5
Automata that solve the problem on only half of ICs,
5 neutral walks.
- $NN_{0.76}$: fitness around 0.765
Solutions near a CA found by Mitchell (GA),
19 neutral walks.

Neutral walks :

- Same starting point on each NN
- Strictly increasing the Hamming distance from the starting solution,
- Stops when there is no neutral step that increases distance.

Diameter

Average length of neutral walks (max 128) :

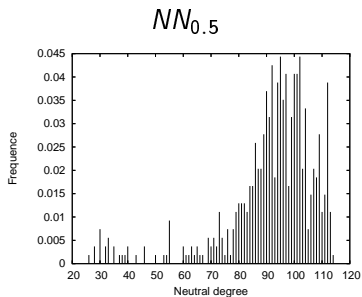
$NN_{0.5}$	108.2
$NN_{0.76}$	33.1

Result on diameter :

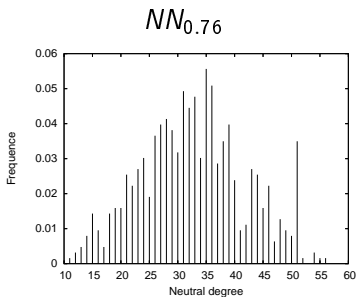
Diameter of $NN_{0.5} >$ Diameter of $NN_{0.76}$.

Neutral Degree Distribution

Distribution of neutral degree collected along all neutral walks.



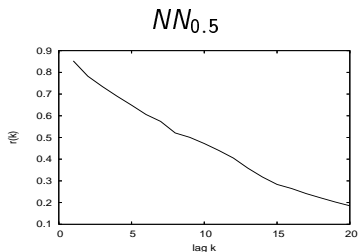
$$Avg_{Stddev} = 91.6_{16.6}$$



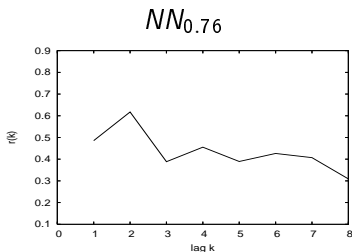
$$Avg_{Stddev} = 32.7_{9.2}$$

- $NN_{0.76}$: close to normal,
- $NN_{0.5}$: skewed and approximately bimodal

Autocorrelation of Neutral Degree



$$\rho(1) = 0.85$$



$$\rho(1) = 0.49$$

- Correlation is not null
- Correlation for $NN_{0.5} >$ Correlation for $NN_{0.76}$
- Graphs of Neutral Networks are not random graphs
- Variation of neutral degree is smooth on NN

⇒ important consequence on metaheuristic design

Definition of Olympus Landscape

Two symmetries that do not change performance :

0/1 symmetry and *right/left* symmetry.

Symmetries of *blok* which maximize the number of joint bits

$$\begin{aligned} GKL' &= GKL, & Das' &= Das, & Davis' &= S_{01}(Davis), \\ ABK' &= S_{01}(ABK), & Coe1' &= Coe1 & Coe2' &= S_{rl}(Coe2). \end{aligned}$$

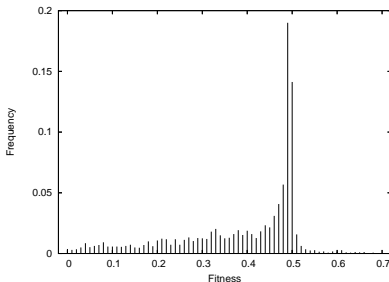
Olympus Landscape, subspace of dimension 77 :

```
000*0*0* 0*****1** 0***00** **0**1** 000***** 0*0**1** ***** 0*0**1*1
0*0***** *****1** 111111** **0**111 ***** 0**1*1*1 1111**1 0*01*111
```


Density Of States of Olympus subspace

Null fitness value :

Sampling	Random	M-H
Whole search space	99.9%	4.4%
Olympus	28.6%	0.3%



Metropolis-Hastings sampling :

- No solution with a fitness value superior to 0.68
- Advantageous to concentrate the search in the Olympus landscape.

Neutral Degree : Sampling Method

- Solutions < 0.5 : randomly chosen in Olympus.
- Solutions > 0.5 : from 2 runs of a GA during 10^3 generations.

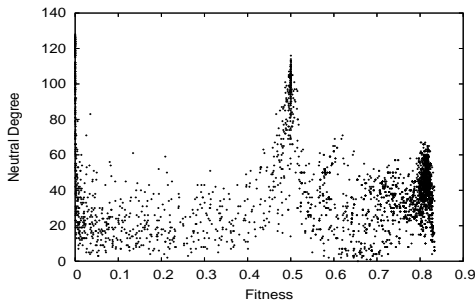
AG used :

- Based on GA defined by Mitchell
- Operators : restricted to Olympus subspace
- Selection : tournament selection taking into account the neutrality.

Justified the usefulness of Olympus and neutrality :

- Discover a lot of solutions between 0.80 and 0.835
- Over 50 runs, average performances $0.832_{0.006}$
(coevolutionary AG of Page 02 : $0.80_{0.02}$).

Neutral Degree



- Two large neutral networks at fitnesses 0 and 0.5 :
neutral degree > 70 .
- Over fitness 0.5 :
average of neutral degree if 37.6.

Main Results on Fitness landscape

- Considerable number of CAs of performance 0 or 0.5
- High neutrality of the landscape
- Neutral networks studies are not random graphs
- Fitness landscape of Majority Problem is very hard!
- Defined the *Olympus* landscape :
exploiting similarities between the six best rules.
- Less solutions with performance 0
- Easy to find solutions over 0.80 with a simple GA.
- Over performance 0.5 : neutrality of landscape is still high

Synthesis

Fitness landscape is a representation of

- notion of neighborhood
- fitness of solutions

Goal :

- local description : fitness between neighbor solutions
Ruggedness, local optima, fitness cloud, neutral networks...
- and to deduce global results :
Difficulty!
to decide a good choice of the representation, operator and fitness function

Open questions

- How to dynamically change the parameters and/or operators of the algorithm with the local description of fitness landscape ?
- Can fitness landscape describe the dynamic of a population of solutions ?
- Links between neutrality and fitness difficulty ?
- Links between neutralities and fitness difficulty ?
-



L. Barnett.

Ruggedness and neutrality - the NKp family of fitness landscapes.

In C. Adami, R. K. Belew, H. Kitano, and C. Taylor, editors, *ALIFE VI, Proceedings of the Sixth International Conference on Artificial Life*, pages 18–27. ALIFE, The MIT Press, 1998.



Lionel Barnett.

Netcrawling - optimal evolutionary search with neutral networks.

In *Proceedings of the 2001 Congress on Evolutionary Computation CEC2001*, pages 30–37, COEX, World Trade Center, 159 Samseong-dong, Gangnam-gu, Seoul, Korea, 27-30 2001. IEEE Press.



U. Bastolla, M. Porto, H. E. Roman, and M. Vendruscolo.
Statistical properties of neutral evolution.

Journal Molecular Evolution, 57(S) :103–119, August 2003.



Meriema Belaidouni and Jin-Kao Hao.

An analysis of the configuration space of the maximal constraint satisfaction problem.

In *PPSN VI : Proceedings of the 6th International Conference on Parallel Problem Solving from Nature*, pages 49–58, London, UK, 2000. Springer-Verlag.



P. Collard, M. Clergue, and M. Defoin Platel.

Synthetic neutrality for artificial evolution.

In *Artificial Evolution : Fourth European Conference AE'99*, pages 254–265. Springer-Verlag, 2000.

Selected papers in *Lecture Notes in Computer Sciences 1829*.



J. C. Culberson.

Mutation-crossover isomorphisms and the construction of discrimination function.

Evolutionary Computation, 2 :279–311, 1994.



Ricardo Garcia-Pelayo and Peter F. Stadler.

Correlation length, isotropy, and meta-stable states.

Physica D, 107 :240–254, 1997.

Santa Fe Institute Preprint 96-05-034.



Josselin Garnier and Leila Kallel.

Efficiency of local search with multiple local optima.

SIAM Journal on Discrete Mathematics, 15(1) :122–141, 2002.



P. Gitchoff and G. Wagner.

Recombination induced hypergraphs : A new approach to mutation-recombination isomorphism, 1996.



David E. Goldberg and Philip Segrest.

Finite markov chain analysis of genetic algorithms.

In *ICGA*, pages 1–8, 1987.



M. Huynen.

Exploring phenotype space through neutral evolution.
Journal Molecular Evolution, 43 :165–169, 1996.



E. Izquierdo-Torres.

The role of nearly neutral mutations in the evolution of dynamical neural networks.
In J. Pollack and al, editors, *Ninth International Conference of the Simulation and Synthesis of Living Systems (Alife 9)*, pages 322–327. MIT Press, 2004.







T. Jones.

Evolutionary Algorithms, Fitness Landscapes and Search.
PhD thesis, University of New Mexico, Albuquerque, 1995.



S. A. Kauffman.

The Origins of Order.
Oxford University Press, New York, 1993.

-  M. Kimura.
The Neutral Theory of Molecular Evolution.
Cambridge University Press, Cambridge, UK, 1983.
-  J. Lobo, J. H. Miller, and W. Fontana.
Neutrality in technology landscape, 2004.
-  M. Newman and R. Engelhardt.
Effect of neutral selection on the evolution of molecular species.
In *Proc. R. Soc. London B.*, volume 256, pages 1333–1338, 1998.
-  Erik Van Nimwegen, James P. Crutchfield, and Martijn Huynen.
Metastable evolutionary dynamics : Crossing fitness barriers or escaping via neutral paths?
Technical Report 99-07-041, SanteFe institute, 1999.



M. Defoin Platel.

Homologie en Programmation Génétique - Application à la résolution d'un problème inverse.

PhD thesis, Université de Nice Sophia Antipolis, France, 2004.



Eduardo Rodriguez-Tello, Jin-Kao Hao, and Jose Torres-Jimenez.

A new evaluation function for the minla problem.

In *Proceedings of the MIC 2005*, pages 796–801, Vienna Austria, 2005.



Helge Rosé, Werner Ebeling, and Torsten Asselmeyer.

The density of states - a measure of the difficulty of optimisation problems.

In *Parallel Problem Solving from Nature*, pages 208–217, 1996.



P. Schuster, W. Fontana, P. F. Stadler, and I. L. Hofacker.

From sequences to shapes and back : a case study in RNA secondary structures.

In *Proc. R. Soc. London B.*, volume 255, pages 279–284, 1994.



Peter F. Stadler.

Landscapes and their correlation functions.

J. Math. Chem., 20 :1–45, 1996.



Peter F. Stadler and W. Schnabl.

The landscape of the traveling salesman problem.

Phys. Letters, A(161) :337–344, 1992.



Peter F. Stadler and Gunter P. Wagner.

Algebraic theory of recombination spaces.

Evolutionary Computation, 5(3) :241–275, 1997.



Terry Stewart.

Extrema selection : Accelerated evolution on neutral networks.

In *Proceedings of the 2001 Congress on Evolutionary Computation CEC2001*, pages 25–29, COEX, World Trade Center, 159 Samseong-dong, Gangnam-gu, Seoul, Korea, 27-30 May 2001. IEEE Press.



Vesselin K. Vassilev and Julian F. Miller.

The advantages of landscape neutrality in digital circuit evolution.

In *ICES*, pages 252–263, 2000.



Sebastien Verel, Philippe Collard, and Manuel Clergue.

Measuring the evolvability landscape to study neutrality.

In M. Keijzer and et al., editors, *Poster at Genetic and Evolutionary Computation – GECCO-2006*, pages 613–614, Seattle, 8-12 July 2006. ACM Press.



E. D. Weinberger.

Correlated and uncorrelated fitness landscapes and how to tell the difference.

In *Biological Cybernetics*, pages 63 :325–336, 1990.



S. Wright.

The roles of mutation, inbreeding, crossbreeding, and selection in evolution.

In *Proceedings of the Sixth International Congress of Genetics 1*, pages 356–366, 1932.

PART II

Problem Hardness in Genetic Programming

Leonardo Vanneschi

Dipartimento di Informatica, Sistemistica
e Comunicazione (DISCo)

University of Milano-Bicocca, Italy

vanneschi@disco.unimib.it

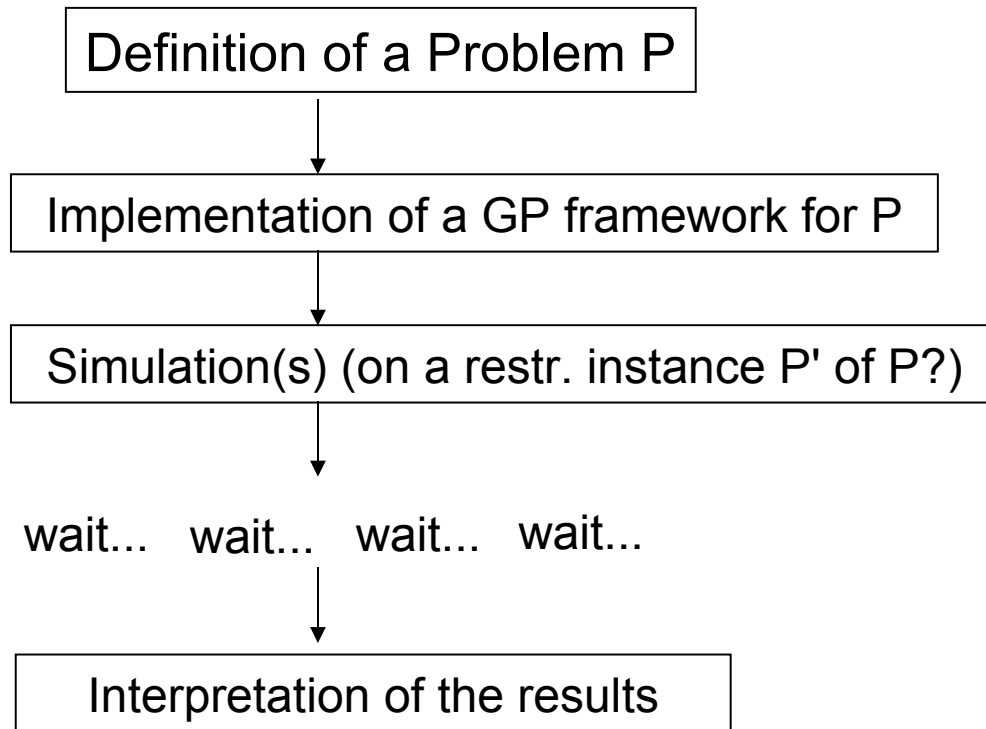
<http://personal.disco.unimib.it/Vanneschi>



Why predicting the difficulty of a problem is important?



Is GP the good technique to solve my problem?



Is there a better way?



Define some measures to quantify the ability of GP to solve a problem from its high level specification!

Not obvious:

- GP is stochastic
- GP *works well* on P' ... but how does it work on P ?



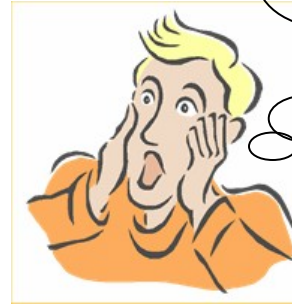
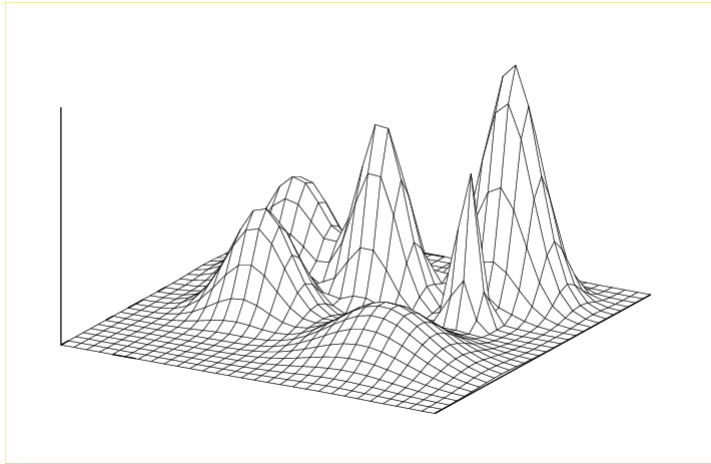
The first step: J. R. Koza, 1992

Proposed "measure" of problem hardness: number of individuals that have to be sampled by GP before finding a solution with a given probability p (usually $p = 0.99$).

Remarks

- It can't be calculated without executing (many times!) GP
- It can be used to confirm the results of another hardness measure

Fitness Landscapes in GP



- Huge search spaces!
- Multidimensionality of neighborhoods!

Very complex neighborhood structures (genotypes = trees, strings of dynamic size, graphs, ...)

Impossible to draw a Fitness Landscape also for simple problems!

We look for measures able to catch some interesting properties of Fitness Landscapes

Autocorrelation. Kinnear, 1994

Proposed measure of problem hardness for GP: [autocorrelation function](#) (Weinberg in 1990 and Manderick in 1991 had studied the same measure for GAs).

Basically no clear relationship between autocorrelation values and problem hardness was observed for GP

Fitness Landscapes in GP are very complex, but...

"Why Ants are Hard?" Langdon, Poli, 1998

Enumeration of a small fraction of the total search space and random sampling characterise it as rugged with many multiple plateaus split by deep valleys and many local and global optima. This suggests it is difficult for hill climbing algorithms.

Many other similar studies in "Foundations of Genetic Programming", **Langdon, Poli, 2002.**

This book also contains an important first step towards the study of problem hardness using the results obtained for the Schema Theorem.

Relationship between Neutrality and Evolvability

T. Yu, J. Miller 2001

Neutrality is particularly interesting in GP since functional redundancy and introns naturally foster neutrality

G1: `nor (and x1 x2) (nor x1 x2)`

G2: `nor (nand (nand x1 x2) (or x1 x2)) (nor x1 x2)`

G3: `nor (and x1 x2) (nor (nor x1 x2) (nand x1 x2))`

different programs (genotypes), same functional behavior (phenotype)

Implicit neutrality

Yu and Miller introduce *explicit neutrality* and a way to measure it for Cartesian GP.

Neutrality Measured with Hamming distance

Let G be an individual in the population at a certain time step.

Let $G1$ be an individual obtained by mutating G .

If G and $G1$ have the same fitness (the mutation is neutral), Yu and Miller accept $G1$ as a legal offspring (and thus allow him to take part in the evolution) only if G and $G1$ have a smaller Hamming distance than a given constant k .

Changing this constant k (Hamming distance threshold) allows us to control the *amount of allowable neutral mutations*, i.e. the *amount of neutrality*.

Yu and Miller Results (Synthesis)

Larger amount of neutrality allow GP to generate fitter individuals

(results criticized by Collins, 2005)

Relationship between Code Growth and Problem Difficulty: Gustafson, Ekárt, et al., 2004

They used two different types of symbolic regression increased instance difficulty.

Results

Increased difficulty induces higher selection pressure and less genetic diversity, which both contribute toward an increased rate of code growth

Discussion

Problem difficulty is

- bound to neutrality
- bound to code growth
- bound to tree-shapes (Daida et al., 2001)
- ...

but....

we still miss mathematical measures
of problem hardness

Hardness Measures for GP

- Fitness-Distance Correlation (*fdc*)
- Negative Slope Coefficient (*nsc*)

Collaborators:

- Marco Tomassini (University of Lausanne, Switzerland)
- Philippe Collard (University of Nice-Sophia Antipolis, France)
- Manuel Clergue (University of Nice-Sophia Antipolis, France)
- Sébastien Verel (University of Nice-Sophia Antipolis, France)

Fitness Distance Correlation (*fdc*) [T. Jones, 1995]

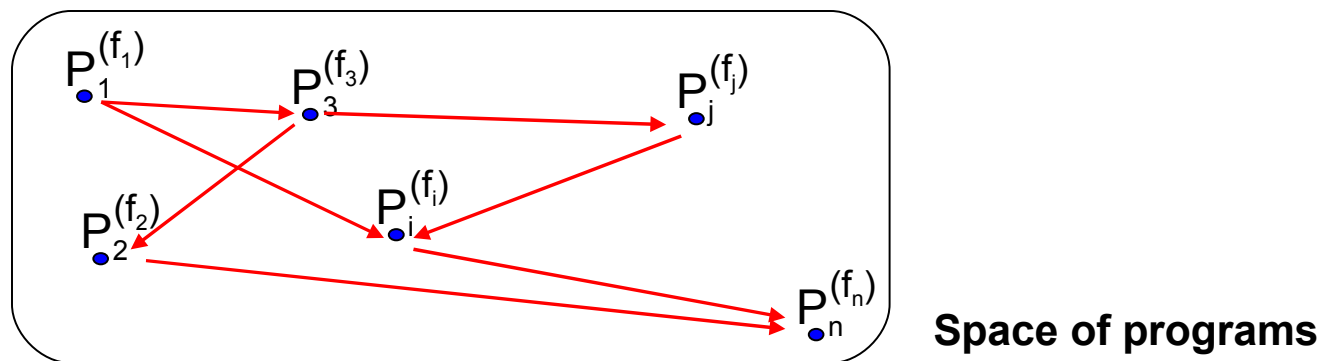
Given a sample of n individuals, let's suppose to know:

- the set $F = \{f_1, f_2, \dots, f_n\}$ of the individual fitnesses
- the genotype of the global optimum (individual with the best fitness)
- a measure to express the genotypic distance between individuals

Let $D = \{d_1, d_2, \dots, d_n\}$ be the n distances to the global optimum, then

fdc is the correlation between sets F and D

Main idea



- Notion of **distance**.
- Relationship between **fitness** and **distance to the goal**.

First use of the *fdc* in GP: Nicolaev and Slavov, 1998

They used the *fdc* to choose a mutation operator among a set of given ones

In 2005 we tried to use the *fdc* for GP much more in the same way Jones intended to use it for GAs.

***fdc* as tool for problem hardness [T. Jones, 1995]**

For GAs, problems can be classified in three classes:

- **Misleading** ($fdc \geq 0.15$) in which fitness increases with distance.
- **Difficult** ($-0.15 < fdc < 0.15$) in which there is no correlation between fitness and distance.
- **Straightforward** ($fdc \leq -0.15$) in which fitness increases as the global optimum approaches.

To (experimentally) verify if the same property is also valid for GP:

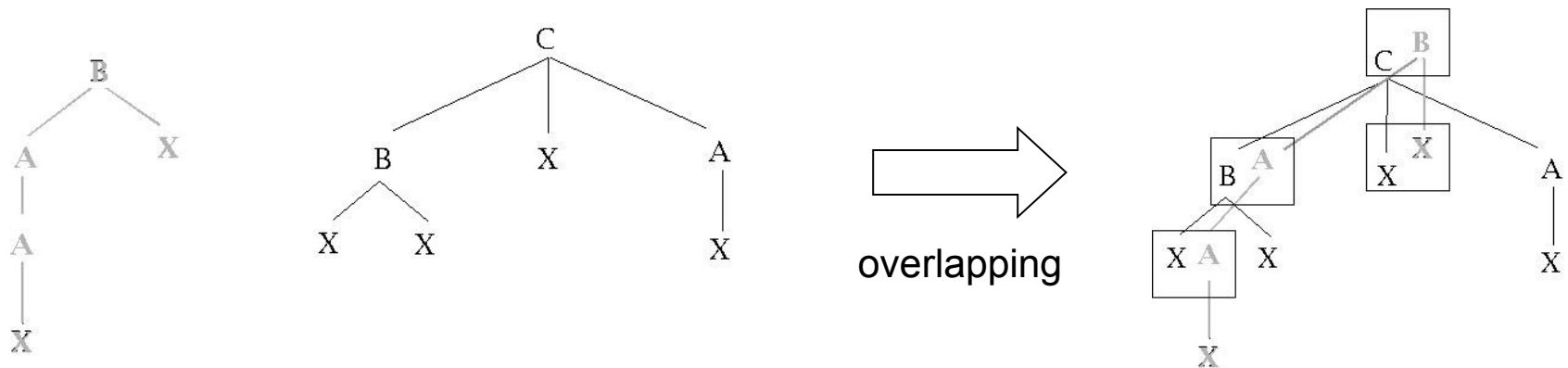
First step: to choose a distance between genotypes (trees!)

Our approach

- To chose a distance between genotypes to calculate fdc
- To define some genetic operators *consistent* with this distance
- To test fdc on a set of functions

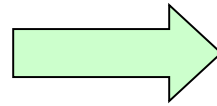
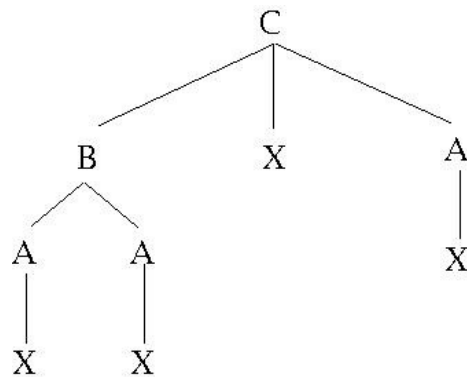
Structural Distance (Intuition)

[Ekàrt-Németh 2002]

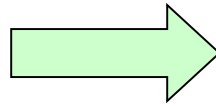
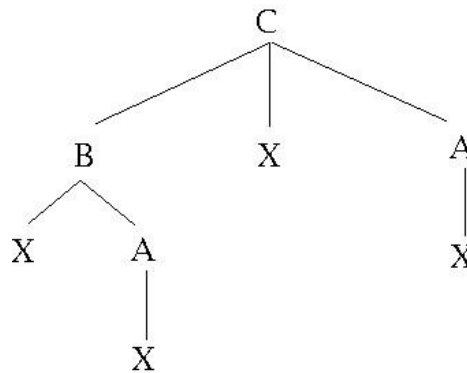
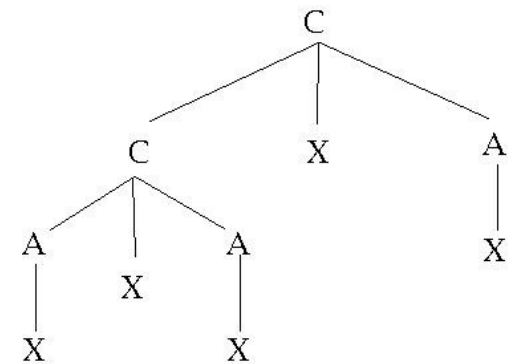


- We assign a *weight* to each node
- We calculate the difference of the weights of nodes at corresponding positions
- The distance is the *weighted sum* of these differences

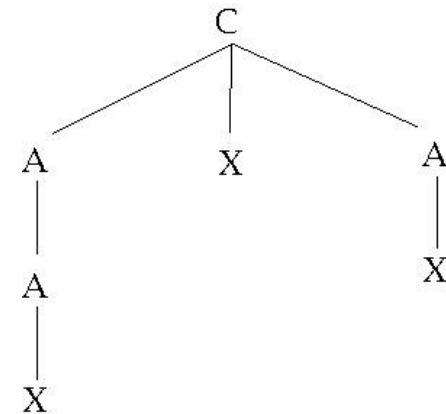
Operators of Structural Mutation



Inflate Mutation



Deflate Mutation



GP based only on these operators:

Structural *M*utation Genetic *P*rogramming (*SMGP*).

Property (Distance/Operators Consistency)

Let:

- $F = \{A, B, C, \dots\}$ $T = \{X\}$
- s.t. $s \in \{F \cup T\} : c(s) = \text{arity}(s) + 1$
- T_1 et T_2 two trees composed by symbols $\{F \cup T\}$
- $k = 1, z = 1$

If

$$\text{dist}(T_1, T_2) = D$$

then

T_2 can be obtained from T_1 with a sequence of $D/2$ operations of structural mutation

Summary of *fdc* results

Fdc correctly measures the difficulty of:

- Unimodal and Multimodal Trap Functions (Deb, Goldberg)
- Royal Trees (Punch)
- Max Problem (Gathercole)

Are we happy ?

***fdc* drawbacks**

- Existence of counterexamples

Ridged Royal Trees

(inspired by the counterexample for GAs of [Quick *et al.*, 1998])

- Not a predictive measure

Optima must be known "a priori"

(this drawback makes *fdc* "almost" unusable in practical cases)

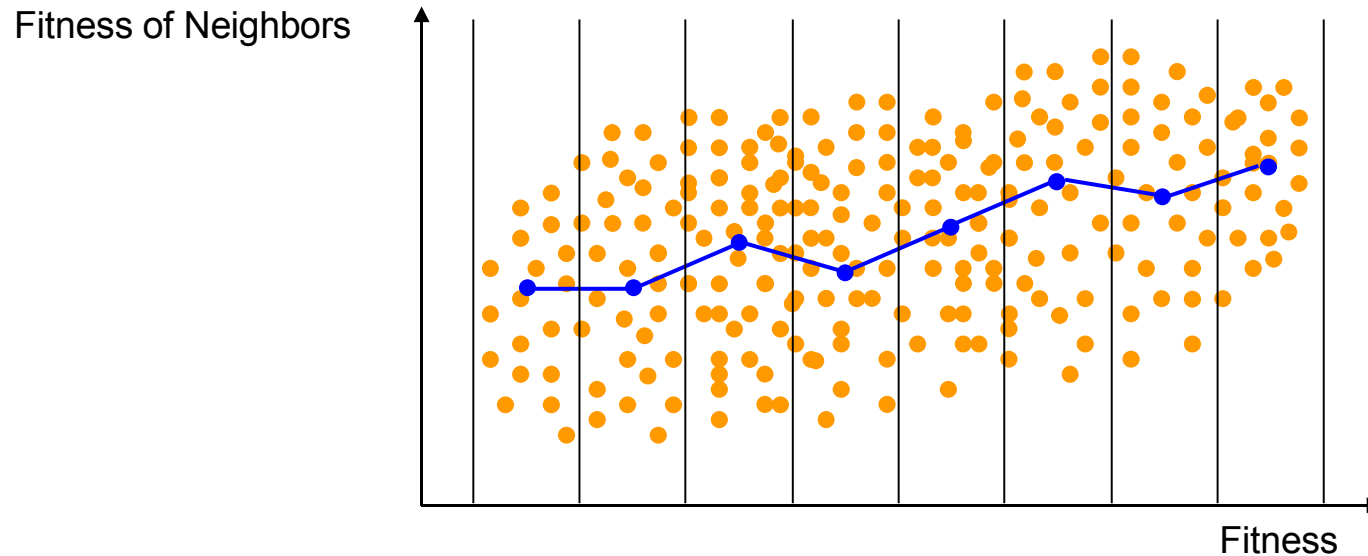
A new measure is needed to quantify the difficulty of "*real*" problems.

The measure we have proposed is based on the concept of ***fitness clouds***.

Measure of Problem Hardness Based on Fitness Clouds

Negative Slope Coefficient (nsc)

- A fitness cloud is partitioned into n bins
- For each bin, a point is calculated, such that its abscissa is the average of the abscissas and its ordinate is the average of the ordinates.
- All these points are joined by segments $\{S_1, S_2, \dots, S_{n-1}\}$



Negative Slope Coefficient Definition

$$nsc = \sum_{i=1}^{n-1} p_i$$

where, $i \in [1, n-1]$

$$p_i = \min \{0, \text{slope}(S_i)\}$$

Hypothesis:

- $nsc = 0$ the problem is **easy**
- $nsc < 0$ the problem is **difficult** and the magnitude of nsc quantifies the difficulty

Idea:

If $nsc < 0$ then there is at least one area of the fitness landscape where evolvability is bad.

Sampling the search space and the neighborhoods

Main idea:

Evolvability makes sense if it is calculated on "good" individuals ("bad" ones are probably discarded by selection).

Sampling the search space:

Importance sampling (Metropolis-Hastings technique)

Sampling the neighborhoods:

selection (tournament selection of size 10).

Summary of *nsc* results

- Good hardness indicator for:
 - Trap Functions
 - Royal Trees
 - Binomial-3 Problem [Daida *et al.*, 2001]
 - Even Parity Problem [Koza, 1992]
 - Artificial Ant on the Santa Fe Trail [Koza, 1992]
- Many ways of calculating the *nsc* have been used:
 - Number of neighbors for each sampled individual
 - Number of mutations to generate neighbors
 - Different types of mutations to generate neighbors
 - Different techniques to partition the fitness clouds into bins
- *nsc* is predictive it can be used on *any* problem
- *nsc* has not been normalized yet into a given range (classification of different problems by their difficulty)
- *nsc* lacks formal/theoretical justification

A first step towards a theoretical justification of *nsc*

R. Poli and L. Vanneschi. Fitness-proportional negative slope coefficient as a hardness measure for genetic algorithms. In *Proceedings of the Genetic and Evolutionary Computation Conference, GECCO 2007*. ACM Press, 2007. Nominated for the *Best Paper Award* for the Genetic Algorithms track.

Presentation:

Monday 9 July

at 10:40

Room: Roberts G06

What about Crossover?



Modeling/Studying GP Crossover

- Schema Theorem [R. Poli and coworkers]
- Geometric Crossover [R. Poli and A. Moraglio]
- Homologous Crossover [M. Defoin-Platel, P. Collard et al.]
- Crossover (pseudo-)distance [S. Gustafson and L. Vanneschi]

Crossover Distance

Collaborator:

Steven Gustafson (GE Global Research, Niskayuna, NY,
USA)

The (Basic) Idea

We don't have to count *how many crossovers* it takes to transform a tree T_1 into another tree T_2 , but *how probable it is* to obtain T_2 by applying crossover to T_1 (in just one step!).

[S. Gustafson, L. Vanneschi, *Operator based distance for Genetic Programming: Subtree Crossover Distance*, EUROGP 2005]

Subtree Crossover Distance (SCD)

between two trees T_1 and T_2

=

Probability of:

- Selecting a subtree S_{T_1} from T_1 , and
- Finding a subtree S_{T_2} in the population P

Such that:

Replacing S_{T_1} with S_{T_2} in T_1 we get T_2

Terminology

- SCD is a *probability!*
- SCD between two trees T_1 and T_2 is a function of T_1 , T_2 and the population (P) in which T_1 and T_2 are!

Thus

SCD is NOT a distance (metric) !!

We need a *similarity / dissimilarity measure (for subtree crossover)*, not necessarily an (Euclidean) distance metric.

The term *pseudo-distance* would be more appropriate.

SCD Definition

```
func SCD( $T_1, T_2, P$ ) {  
     $S = \text{diff}(T_1, T_2)$   
     $res = 1$   
     $\forall (s_{T_1}^i, s_{T_2}^i) \in S:$   
         $ps1 = \text{probSelecting}(s_{T_1}^i, T_1)$   
         $ps2 = \text{probCreating}(s_{T_2}^i, P)$   
         $res = res * (1 - ps1 * ps2)$   
    return( $res$ )  
}
```

the complexity is
"reasonable"!

The operator $\text{diff}(T_1, T_2)$ returns the set

$$S = \{(s_{T_1}^1, s_{T_2}^1), (s_{T_1}^2, s_{T_2}^2), \dots, (s_{T_1}^n, s_{T_2}^n)\}$$

such that:

if we replace $s_{T_1}^i$ with $s_{T_2}^i$ ($i \in [1, n]$) in T_2 we obtain T_1

Summary of Crossover-Distance results

SCD appropriate for:

- Measuring the FDC *dynamically* (during evolution)
- Fitness Sharing

Our hypothesis: SCD appropriately models subtree crossover

- SCD diversity behave differently than ED diversity (slightly increasing and larger than zero for successful runs, approximately zero for unsuccessful runs)

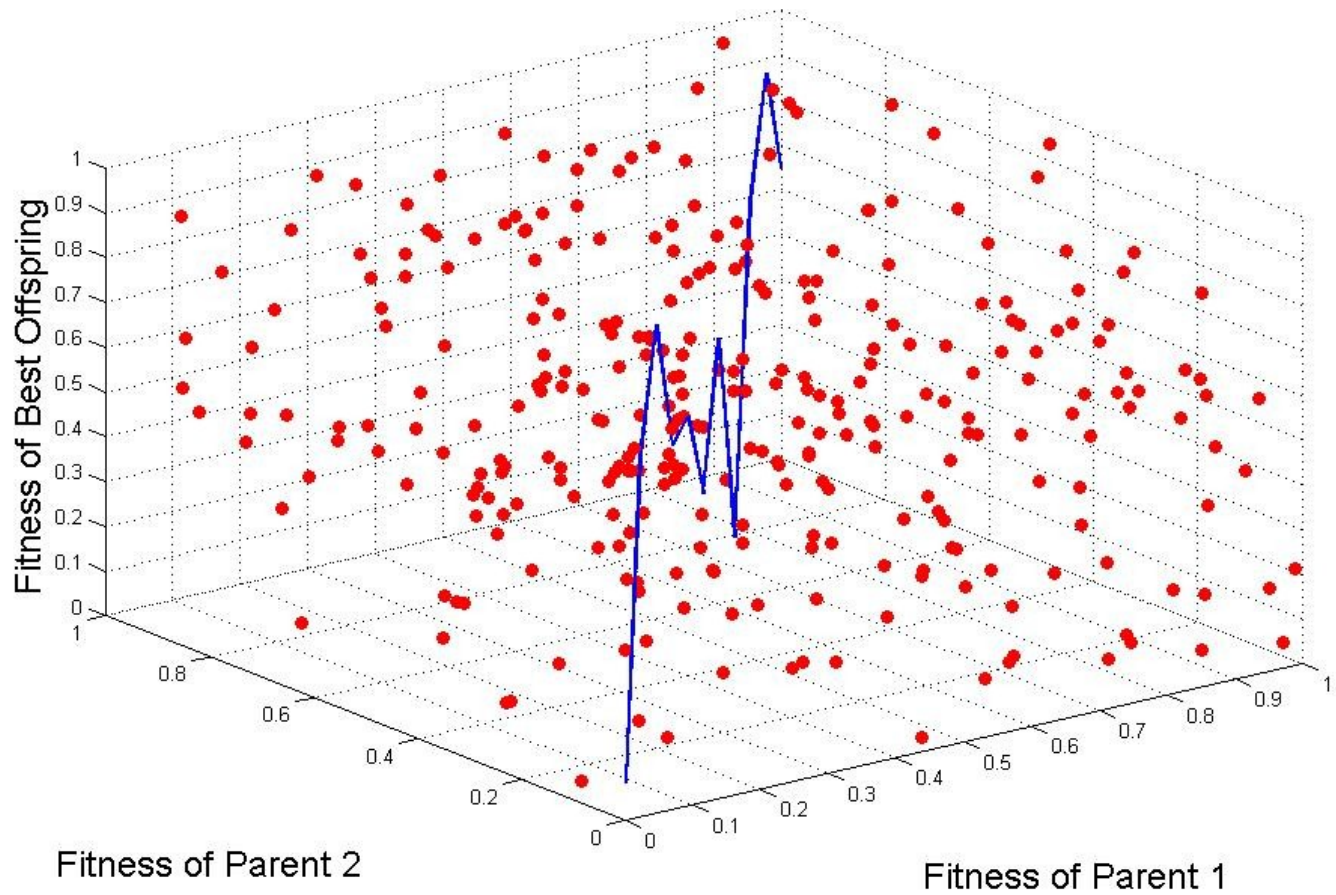
Can it be used to predict the behavior of GP runs?

Last Discussion:

Can we define an NSC "with" crossover?

A possible idea

- Generate two samples of individuals S_1 and S_2
- Repeat
 - Take one individual i_1 from S_1 , one individual i_2 from S_2
 - Perform the crossover between i_1 and i_2 , let j_1 and j_2 be the offspring
 - Let $j = \text{best}(j_1, j_2)$
 - Plot the triple (i_1, i_2, j) on a 3D plane
 - Eliminate i_1 from S_1 and i_2 from S_2
- Until S_1 and S_2 are empty



Bibliography

- J. M. Daida, R. Bertram, S. Stanhope, J. Khoo, S. Chaudhary, O. Chaudhary
What makes a problem GP-hard? Analysis of a tunably difficult problem in genetic programming.
Genetic Programming and Evolvable Machines, 2:165–191, 2001.
- J. M. Daida, H. Li, R. Tang, A. M. Hilss
What makes a problem GP-hard? Validating a hypothesis of structural causes.
In R. Poli et al. editors *Genetic and Evolutionary Computation – GECCO-2003*,
volume 2724 of *LNCS*, pages 1665–1677. Springer-Verlag, Berlin.
- K. E. Kinnear
Fitness landscapes and difficulty in genetic programming.
In *Proceedings of the First IEEE Conference on Evolutionary Computing*,
pages 142–147. IEEE Press, Piscataway, NY.
- J. R. Koza
Genetic Programming
The MIT Press, Cambridge, Massachusetts, 1992

- W. B. Langdon, R. Poli
Foundations of Genetic Programming
Springer, 2002
- N. I. Nikolaev, V. Slavov.
Concepts of inductive genetic programming.
In W. B. Langdon et al. editors,
Genetic Programming, Proceedings of EuroGP'1998,
volume 1391 of *LNCS*, pages 49–59. Springer-Verlag, 1998
- M. Tomassini, L. Vanneschi, P. Collard, M. Clergue
A study of fitness-distance correlation as a difficulty measure in genetic programming
Evolutionary Computation, 13(2): 213-239, 2005
- L. Vanneschi, M. Tomassini, P. Collard, S. Verel
Negative Slope Coefficient. A measure to characterize genetic programming fitness
landscapes
In P. Collet et al. editors, *Genetic Programming, 9th European Conference,
EuroGP 2006*, pages 178-189, Lecture Notes in Computer Science. 2006
- L. Vanneschi, S. Gustafson, G. Mauri
Using subtree crossover distance to investigate genetic programming dynamics
In P. Collet et al. editors, *Genetic Programming, 9th European Conference,
EuroGP 2006*, pages 238-249, Lecture Notes in Computer Science. 2006

- T. Yu, J. Miller
Neutrality and the evolvability of boolean function landscape.
In J. Miller et al., editor, *Proceedings of the Fourth European Conference on Genetic Programming (EuroGP-2001)*, volume 2038 of LNCS, pages 204–217, Lake Como, Italy, 2001. Springer, Berlin, Heidelberg, New York. Lecture notes in Computer Science vol. 2038.
- W. B. Langdon and R. Poli,
Why Ants are Hard.
In *Genetic Programming 1998: Proceedings of the Third Annual Conference*
Morgan Kaufmann, J. R. Koza et al. editors, pages = 193-201, 1998
- S. Gustafson, A. Ekárt, E. K. Burke, G. Kendall
Problem Difficulty and Code Growth in Genetic Programming
Genetic Programming and Evolvable Machines, 2004, Volume: 5, Issue: 3, p. 271-290