Quelques questions en optimisation stochastique

Outline

1. Introduction
2. A study of local optimas' basins and local optima networks
   - NK Landscapes
   - Landscapes as Networks
   - Statistics of Maxima Networks
   - Structure of the Associated Basins
3. Discussion

Optimisation problem

Definition

$\mathcal{P}$=($\mathcal{S}$, f)

- $\mathcal{S}$: set of potential solutions, search space
- $f : \mathcal{S} \rightarrow \mathbb{R}$: objective function to maximize (or minimize)

Goal : find the set $\mathcal{S}_{opt} \in \mathcal{F} \subset \mathcal{S}$ ($\mathcal{F}$ feasable solutions) such that

$$\forall s_{opt} \in \mathcal{S}_{opt}, \forall s \in \mathcal{F}, f(s) \leq f(s_{opt})$$

Optimisation

- Methodes exactes :
  - Branch and bound
  - Prog. Dynamique
  - prog. par contraintes
  - A^*
  - algorithmes de gradients
- methodes approchees :
  - Heuristiques : specifique à un problème
  - Meta-heuristiques

Metaheuristics classification

Stochatic algorithms with unique solution :

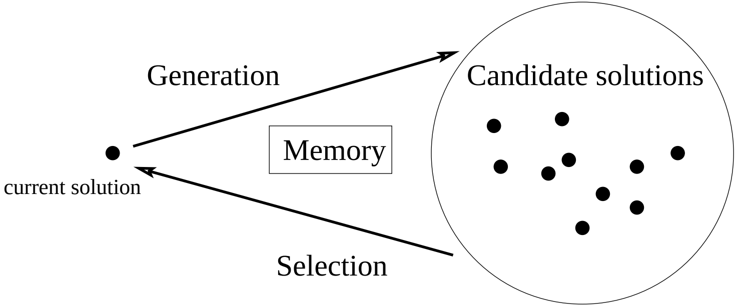
- (Random search),
- Ascent Algorithms : Hill-Climber (HC), Stochatic Hill-Climber
- Simulated Annealing (Kirkpatrick *et al* 1983)
- Tabu search (Glover 1986)

Stochatic algorithms with population of solutions :

- Evolutionary Algorithm : the oecumenic algorithm
- Ant optimization (Bonabeau 1999) : route on graph
- memetic algorithms : combined local search with population
- Particule Swarm Optimisation (PSO)

Stochatic algorithms with unique solution

- $\mathcal{S}$ set of solution (search space)
- $f : \mathcal{S} \rightarrow \mathbb{R}$ objective function
- $\mathcal{V}(s)$ set of neighbor's solution of s



Metaheuristics
Random search / Hill Climbing

Algorithm 1 Random walk

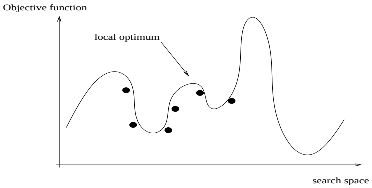
Choose randomly initial solution $s \in \mathcal{S}$
repeat
  Choose $s^{'} \in \mathcal{V}(s)$ randomly
  $s \leftarrow s^{'}$
until ...

Algorithm 2 Hill-climbing

Choose randomly initial solution $s \in \mathcal{S}$
repeat
  Choose $s^{'} \in \mathcal{V}(s)$ such as $f(s^{'})$ is maximal
  $s \leftarrow s^{'}$
until s local optimum

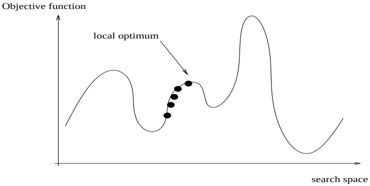
Metaheuristics
Random search / Hill Climbing

Random walk



maximal exploration , diversification

Hill-climbing



maximal exploitation , intensification

Main issue : exploration / exploitation tradeoff

escape from local optima, etc.

$\Longrightarrow$ simulated annealing, tabu search

Simulated annealing

Algorithm 3 Simulated annealing

Choisir solution initiale $s \in \mathcal{S}$ et temperature initiale T
repeat
  choisir aléatoirement $s^{'} \in \mathcal{V}(s)$, $\Delta = f(s^{'}) - f(s)$
  if $\Delta > 0$ **then**
    $s \leftarrow s^{'}$
  else
    u nombre aléatoire de $[0, 1]$
    if $u < e^{\frac{\Delta}{T}}$ **then**
      $s \leftarrow s^{'}$
    end if
  end if
  update temperature T
until Critère d'arrêt vérifié

Métaheuristiques
Recherche tabou

Glover 1986.

- Concept de mémoire : interdiction de mouvement
- Critère d'aspiration : lever exception d'interdiction

Métaheuristiques

Recherche tabou

Algorithm 4 Recherche tabou

Choisir une solution initiale $s \in \mathcal{S}$
Initialiser la liste tabou M
repeat
  Choisir $s' \in V(s)$ telle que :
   $f(s')$ meilleure solution de $V(s)$ ET critère d'aspiration
   OU
   $f(s')$ meilleure solution de $V(s)$ non taboue
   $s \leftarrow s^{'}$
  Mettre à jour la liste tabou M
until s optimum local

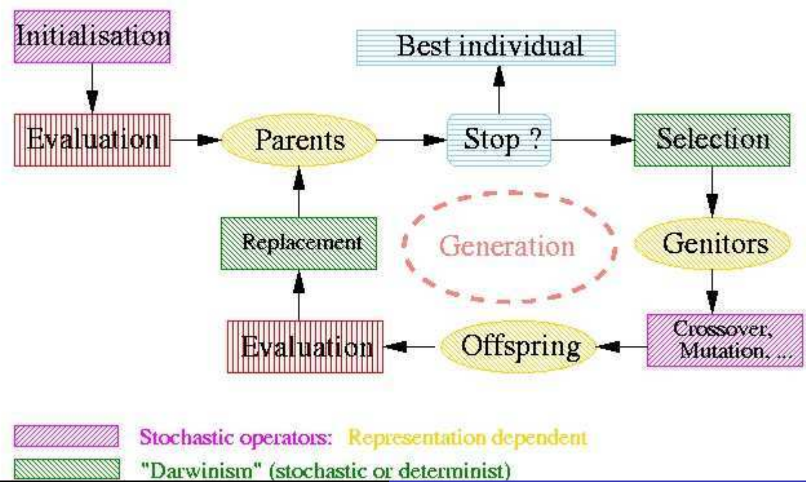
Metaheuristics

Population based metaheuristics

Principles

- At each iteration,
  the algorithme select the best solutions,
  stochastic variation of the solutions to compose the new population
- the algorithm stop according to the convergence or computation time

Generic Evolutionary Algorithm



Classification of evolutionary algorithms

Classification according to the search space :

- Evolution strategies (Schweffel 1970) : real optimization
- Genetic algorithms (Holland 1975 and even before) : binary strings
- Genetic programming (Koza 92) : "program" where solution are computed by a machine

Memetic algorithms (hybrid algorithm)

Combined :

- efficient local search (local operator)
- population search based (global operator)

$\Rightarrow$ initialisation of the population use local search
$\Rightarrow$ "mutation" operator is replaced by local search

Avantages :

- at least, the same performance as efficient (specific) local search !
- Use the global operator on population,
- better compromize between exploration of search space and exploitation of "good" solution

Algorithms framework

- EO, paradisEO (c++) :
  evolutionary algorithm, local search, multiobjective optimisationn parallelisation method
- openBeagle (c++) :
  less used than paradisEO
- ECJ (java) :
  nearly the same asparadisEO, but neither local search, neither parallelisation
- matlab, scilab :
  numerical optimisation (CMA-ES)
- Frontend GUIDE : graphic user interface to develop an EA with ECJ or EO

What kind of problems could be solve with metaheuristics ?

- very large search space (combinatorial optimisation) :
  NP-complet problem
- "complex" objective function :
  irregular, non differentiable, non continue...
- objective function given by a computation or an experience :
  "black box" optimisation

Recent work : Quadratic Assignment Problem (QAP)

Definition of the problem

- Problème NP-Complet.
- Deux matrices de dimension $n \times n$, $A = (a_{ij})$ et $B = (b_{ij})$.
- Permutation π de $[1, 2, \ldots, n]$ qui minimise la quantité :

$$z(\pi) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} b_{\pi(i)\pi(j)}$$

Recent work : Quadratic Assignment Problem (QAP)

Une application

Application : terminaux aéroport

Dans un aéroport, on désire placer n portes d'embarquement sur n emplacements dont on connait les distances (matrice A). On connait également les flux des passagers entre chaque porte d'embarquement (matrice B).

Optimisation

On cherche ainsi à affecter les portes d'embarquements aux avions pour minimiser la marche de l'ensemble des passagers.

Problèmes de logistique, etc.

Recent work : Quadratic Assignment Problem (QAP)

Instances de Taillard

QAPLIB

Les instances de Taillard ont été générés uniformément et comptent parmi les instances les plus difficiles du QAP.

Difficultés

L'optimalité des solutions pour les instances pour $n <= 25$ a été prouvée par des méthodes exactes de type *branch and bound*. Les instances de cardinalité plus grande relèvent encore du défi.

Utilisation des métaheuristiques

Les métaheuristiques sont ainsi pour l'instant les meilleurs algorithmes pour les grandes instances du QAP.

Recent work : Quadratic Assignment Problem (QAP)

Results (The Van Luong, PhD INRIA futur Lille sept. 08)

Results

- Implementation of best know algorithm :
  confirm the best know solution found, but not the average performances of the algorithm
- Parallelisation of the algorithm (proActive) :
  distribution of the local search,
  with 17 computers : speed up 11 times.
- Study of the parameters :
  find the main parameters of the algorithm

Recent work : Quadratic Assignment Problem (QAP)

Results (The Van Luong, PhD INRIA futur Lille sept. 08)

Perspectives

- Study the influence of the distribution method on simple model of QAP
- Proposed an algorithm with the best parameters and distribution method

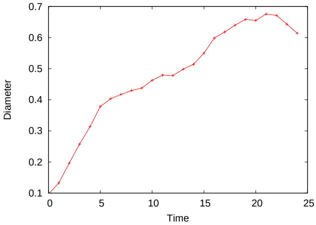
Recent work : pupil diameter

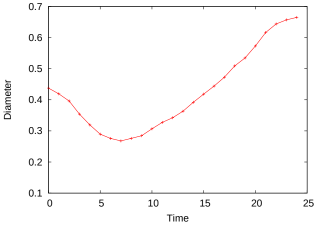
Definition : Cognitive Load (or mental workload) is the property emerging from the interaction between the requirements of the task, their circumstances and the skills, behaviors, and perceptions of the user

- Working Memory (Sweller, 1988) :
  Problem solving, reasoning, language
- Attentional capacity :
  Divided vs selective attention (Dual-Task Paradigm)
  - Filter Theory of selective attention (Broadbent, 1958)
  - Single channel theory (Welford, 1967, 1980)
  - Resource Theory (Kahneman, 1973 ; Wickens, 1984)
- Serial vs Parallel allocation of attention during Dual-Task ?

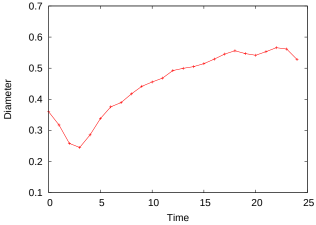
$\Longrightarrow$ pupil diameter is a measure of cognitive load

Recent work : pupil diameter





How to explain the curve of dual task with the single tasks ?



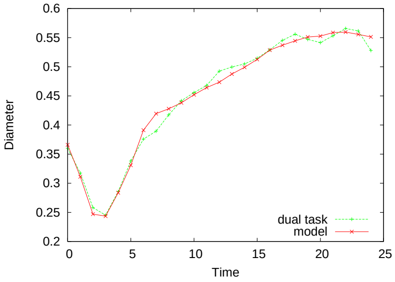
Model

$$F(t) = \begin{cases} K_1(\alpha_1 f_{DM}(t) + (1-\alpha_1) f_{WS}(\omega_1 t)) & \text{if} \quad \forall t < T \\ K_2(\alpha_2 f_{DM}(t) + (1-\alpha_2) f_{WS}(\omega_1 T + \omega_2 (t - T))), & \text{if} \quad \forall t \geq T \end{cases}$$

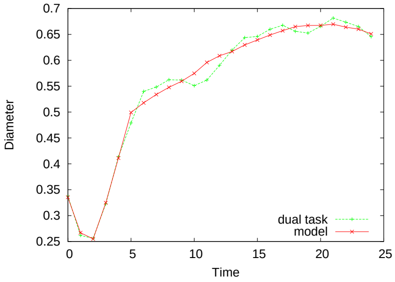
$$\omega_2 = \frac{T_{max} - \omega_1 T}{T_{max} - T}$$

Minimisation of the least square

Recent work : pupil diameter



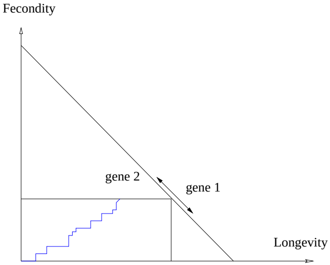
5 digits



9 digits

Recent work : insects
E. Wajnberg, P. Coquillard

Stochatic model which describe the life of a parasitoide.

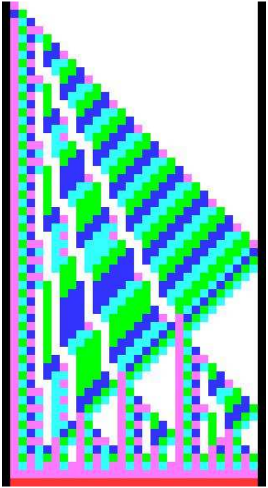


- gene 1 : tradeoff between fecondity and life time duration
- gene 2 : size of the range
- gene 3 : beta for the bayesian estimation
- gene 4 : first tradeoff according to the instability environment

Recent work : insects

- use of CMA-ES instead of genetic algorithm (binary coding for real parameter)
- performances of better, to continue...

Firing squad problem (J.Myhill 1957)



- one firing man = one cell
- line of length n of firing squad
- Initial configuration :
  - the left cell in "general" state
  - right cells in "quiescent" state
- Exchange information between neighbors firing men = local transition function
- Goal : find a transition function such that
  $\forall n$, the final configuration is a line of cells in "firing" state.

Previous works

Transition function which solve FSP in optimal time ($2n-2$ iterations) :

- E. Goto [1962] with thousand of states
- Waksman [1966] with 16 states
- Balzer [1967] with 8 states
- Mazoyer [1987] with 6 states

No transition function with 4 states (Balzer)

Open Problem :

Does it exist transition which solve FSP with 5 states ?

Firing Squad Problem
Search space

$\mathcal{S} = \{\ \delta \mid (\mathcal{Q}, \delta) IfAC,\ \sharp\mathcal{Q} = 5\}$, but rules with "firing" state are not use

Neighborhood of s : solutions with one rule changed

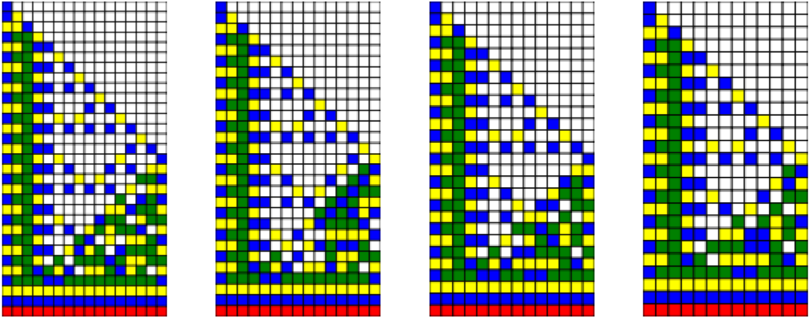
- Whole search space :
  - number of rules : $4^3 - 1$ "central" rules, $4^2 - 1$ rules for left and right bounds
    So, 93 rules
  - size of search space : 5 possible states for each rule
    $5^{93} \approx 10^{66}$
  - size of neighborhood : 4 other possible states for each rule
    $93 \times 4 = 372$

Recent work : Firing squad problem
Results

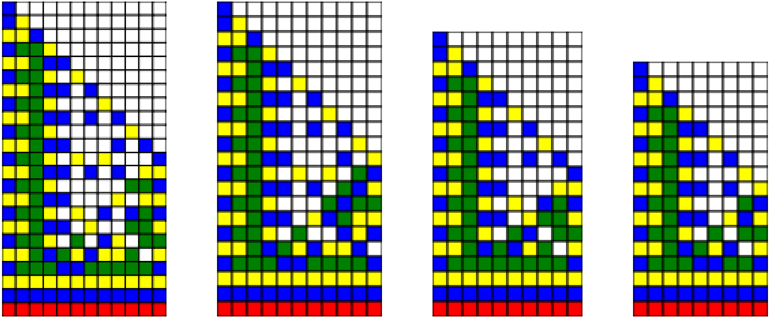
synchronisation de lignes de tailles 2 jusqu'à n

Method	avg	best
Hill-climbing	$5.27_{1.36}$	12
Tabu	$7.1_{2.19}$	13
SA	$4.23_{0.43}$	5
EA	5.31	8
backtrack		12
signaux		restriction of search space

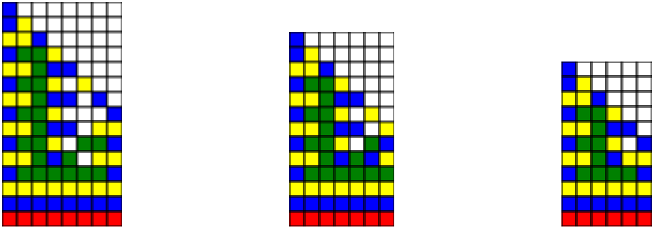
Recent work : Firing squad problem



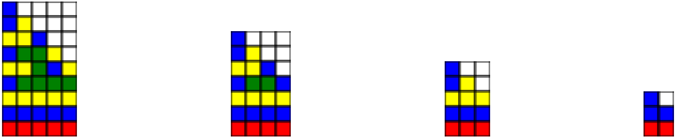
Recent work : Firing squad problem



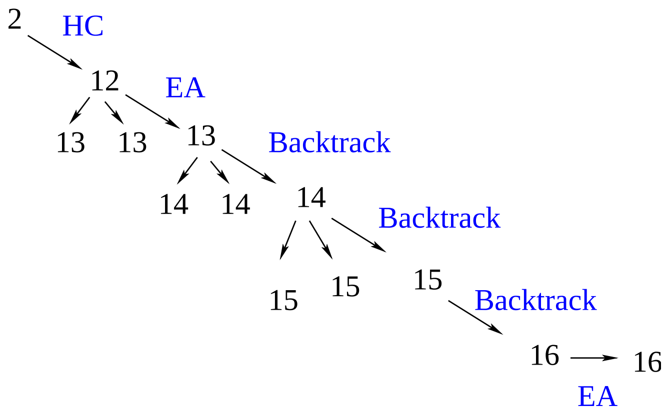
Recent work : Firing squad problem



Recent work : Firing squad problem



Recent work : Firing squad problem



Conclusions provisoires

- echecs :
  - J'ai essaye en boite noire, ca ne marche pas...
  - J'ai essaye sur un problème facile, c'était ridiculement lent, compare à ...
- Contextes recommandés :
  - Problèmes non résolus fonctions chahutées, contraintes chahutées
  - Plusieurs optima critères implicites, multi-critères
  - Problèmes (très) mal posés validation de lutilisateur
  - A coupler avec des méthodes locales avec mesure

Choix crucial : la représentation et les opérateurs de variation

Main questions

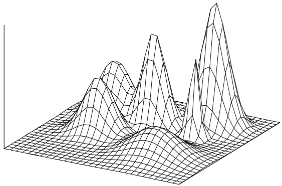
- the choice of the representation and local search operators :
  to have the more regular neighborhood
- Exploration / exploitation tradeoff :
  use the local information and explore new part of the search space
- design of the search parameters :
  link to other problem

Origin of the work

- J. P. K. Doye. The network topology of a potential energy landscape : a static scale-free network. *Phys. Rev. Lett.*, 88 :238701, 2002.
- J. P. K. Doye and C. P. Massen. Characterizing the network topology of the energy landscapes of atomic clusters. *J. Chem. Phys.*, 122 :084105, 2005.

and previous work with Marco Tomassini

Fitness landscape



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

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

NK Landscapes

An idea of S. Kauffman

- A string s of N "spins" or "genes" represented by binary variables $s_i \in \{0, 1\}$
- A real stochastic function Φ defined on $\{0, 1\}^N$:

$$\Phi : \{0, 1\}^N \longrightarrow \mathbb{R}_+$$

- the integer variable K ($0 \leq K < N$) determines how many other spin values in the string influence a given spin s_i
- The value of Φ is the average contribution of all the spins :

$$\Phi(s) = \frac{1}{N} \sum_{i=1}^{N} \phi_i(s_i, s_{i_1}, \ldots, s_{i_K})$$

NK Landscapes ctd.

- For K going from 0 to $N-1$ NK landscapes can be tuned from "smooth" to "rugged" (easy to difficult respectively)
- For $K=0$ there are no correlations, Φ is an additive function, and there is a single maximum
- For $K=N-1$ the landscape is completely random and the expected number of local optima is $2^N/(N+1)$
- Intermediate values of K interpolate between these two extreme cases and have a variable degree of epistasis (i.e. gene interaction)

NK Landscapes ctd.

- The K variables that influence a given spin can be chosen randomly or sequentially in the lexical neighborhood of s_i. Here we use the adjacent neighborhood model
- The search space S for the NK landscapes is the N-dimensional hypercube with $|S| = 2^N$
- The neighborhood V of a given configuration s_i depends on the way the space S is traversed. Here we use the customary single bit-flip operator in which the neighborhood V of a configuration s_i is the set of configurations that are at Hamming distance 1 from s_i
- The real stochastic function Φ previously defined is the fitness function to be maximized

Landscapes as Networks

The Configuration Space Graph G

Let's introduce the graph $G(S, E)$ where S is the set of vertices $\equiv$ possible configurations (the search space), and E is the set of edges. Each edge represents a 1-bit move in the space (Boolean N-dimensional hypercube for the NK landscapes)

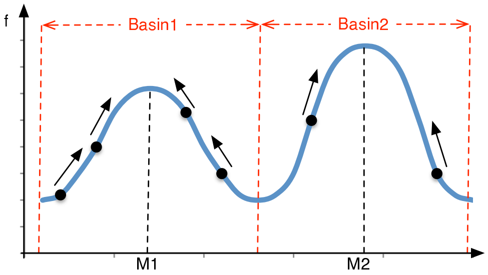
The Optima (or Inherent) Graph M

$M(V^*, \Gamma)$ is the graph whose vertices $V^* \subseteq S$ are the local maxima of the configuration space and whose edges Γ are possible transitions among maxima of V^*.
Note that in general M need not be a subgraph of G, i.e. $\Gamma \not\subset E$

Basins of Attraction

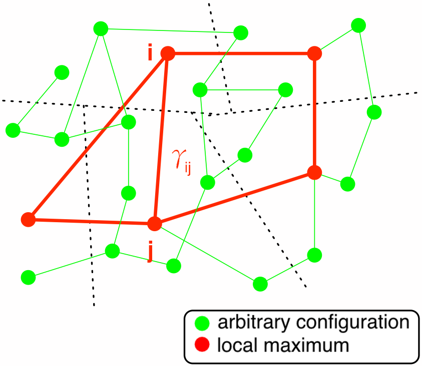
Basin of Attraction

The basin of attraction b_i of a local optimum i is the set of configurations $s \in S$ such that a best improvement local search starting at $s \neq i$ always ends in i



How Are The Edges Γ of M To Be Determined ?

Let's call b_i and b_j the attraction basins of maxima i and j. There is an edge γ_{ij} between two local maxima i and j if there is at least a pair of direct neighbors s_i and s_j such that $s_i \in b_i$ and $s_j \in b_j$



Statistical Characterization of the Optima Graphs M for the NK Landscapes

We shall use the following statistics :

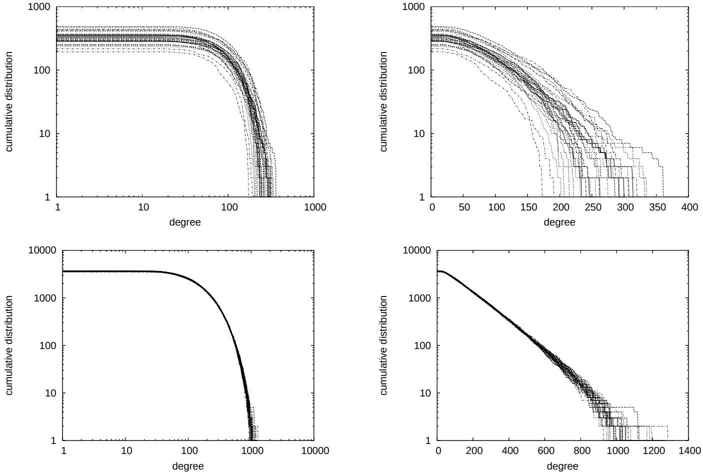
- mean degree
- degree distribution $p(d)$ (ddf) and cumulated degree distribution $P(d < D) = \sum_{i=1}^{D} p(d_i)$
- clustering coefficient $C = \frac{1}{N} \sum_{i=1}^{N} C_i$, where $C_i = 2E_i / k_i(k_i - 1)$, with E_i is the number of edges in the neighborhood of node i
- mean path length $\bar{L}$

Statistical Characterization of the Optima Graphs

Average values over 30 randomly generated landscapes for $N = 18$. $\bar{n}_v$ and $\bar{n}_e$ represent the number of vertexes and edges, $\bar{C}$, the mean clustering coefficient, whilst C_r is the clustering coefficient of a random graph with the same number of vertexes and mean degree, $\bar{z}$ represent the mean degree, $\bar{l}$ the mean path length

$N = 18$						
K	$\bar{n}_v$	$\bar{n}_e$	$\bar{C}$	C_r	$\bar{z}$	$\bar{l}$
2	50_{25}	478_{342}	$0.62_{0.106}$	$0.414_{0.1697}$	$17.08_{4.930}$	$1.66_{0.210}$
4	330_{72}	$17,576_{4898}$	$0.61_{0.044}$	$0.332_{0.0573}$	$105.39_{8.106}$	$1.67_{0.058}$
6	994_{73}	$93,043_{8588}$	$0.51_{0.016}$	$0.189_{0.0115}$	$187.07_{4.650}$	$1.82_{0.012}$
8	$2,093_{70}$	$214,844_{6793}$	$0.41_{0.007}$	$0.098_{0.0038}$	$205.29_{2.615}$	$1.92_{0.006}$
10	$3,619_{61}$	$348,761_{5275}$	$0.33_{0.004}$	$0.053_{0.0011}$	$192.76_{1.150}$	$2.05_{0.009}$
12	$5,657_{59}$	$476,614_{3416}$	$0.27_{0.002}$	$0.030_{0.0005}$	$168.50_{1.003}$	$2.29_{0.012}$
14	$8,352_{60}$	$594,902_{2459}$	$0.23_{0.001}$	$0.017_{0.0002}$	$142.46_{0.652}$	$2.56_{0.007}$
16	$11,797_{63}$	$707,326_{2296}$	$0.21_{0.001}$	$0.010_{0.0001}$	$119.92_{0.368}$	$2.72_{0.003}$
17	$13,795_{77}$	$762,197_{2299}$	$0.20_{0.001}$	$0.008_{0.0001}$	$110.51_{0.377}$	$2.79_{0.005}$

Cumulative Degree Distribution Functions



Cumulative degree distributions for $N = 18$ and $K = 4$ (top), $K = 10$ (bottom). All 30 curves are plotted. Left : Log-log plots. Right : Lin-log plots.

Discussion of Graph Statistics

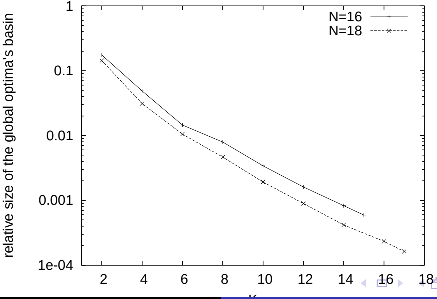
1. From the values of $\bar{C}$ and $\bar{L}$ one can conclude that the Inherent graphs of NK landscapes are of the small-world type. This is not surprising given that the base hypercube G is itself a small-world with a diameter $log\ N$
2. The degree distribution functions are best fitted by exponentials for $N = 18$ and even faster-decaying functions for $N = 16$
3. There seems to be no qualitative difference in the graph ddf for easy landscapes (low K) and rugged ones (high K). The shape of the curves for a given N seems to be essentially the same independent of K
4. This is in contrast to many chemical-physics energy landscapes for which scale-free ddfs have been observed which causes a so-called "funnel" effect (starting anywhere in the space there are many paths that bring the system into the global minimum).

Structure of the Basins of Attraction

The structure of the basins of attraction plays an essential role in search algorithms.

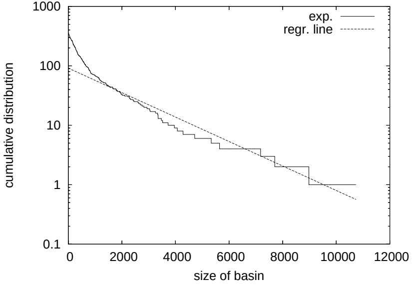
The main results are the following :

- The size of the basin corresponding to the global maximum shrinks exponentially with increasing K (average values over 30 landscapes) :



Structure of the Basins of Attraction ctd.

- The distribution of the number of basins of a given size is essentially exponential

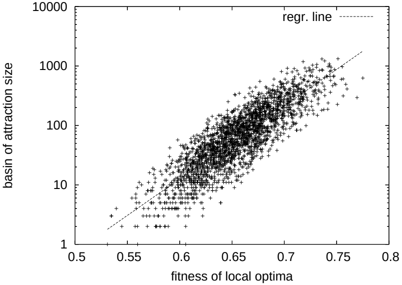


An example with $N = 18$ and $K = 4$.

This result contradicts the often used assumption (for mathematical convenience) that the distribution is uniform

Structure of the Basins of Attraction ctd.

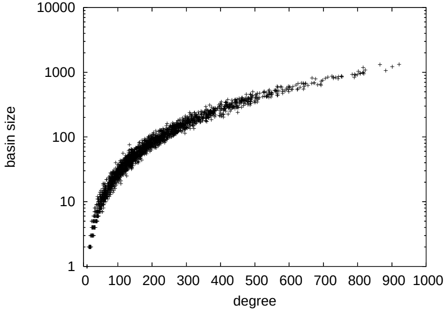
- Basins are broader for lower values of K. This confirms the notion that the landscape becomes harder and more rugged with increasing K
- There is a strong positive correlation between the basin's size and the fitness of the corresponding local optimum :



An example with $N = 18$ and $K = 8$.

Structure of the Basins of Attraction ctd.

- There is a clear positive correlation between the size of the local optima basins and their degree in the maxima graph :



A representative instance with $N = 18$ and $K = 8$.

This shows that the simplified view provided by the maxima graph is a useful one

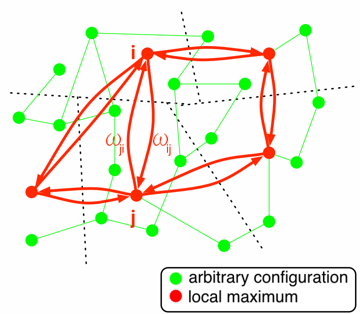
Remarks

- Inspired by a previous model for energy landscapes, we have proposed a new synthetic graph representation of combinatorial fitness landscapes which has proved useful
- We have found no clear relationship between landscape search difficulty and optima network degree distribution, at least for the class of synthetic *NK* landscapes
- We have found new results for the structure of the attraction basins and confirmed already known results starting from the graph view of the landscape
- Among these, the distribution of the basins' sizes, the correlation between local maximum fitness and basin size, and the correlation between maximum degree in the graph and basin size are particularly remarkable

Work in Progress

The only modification in the new model (to appear in the proceedings of Alife 08 in August) concerns the meaning of the edges of the optima network

The new notion of an edge better reflects the workings of stochastic local search algorithms by attributing a weight w_{ij} to it proportional to the probability of inter-basin transitions in both directions (i.e. in general $w_{ij} \neq w_{ji}$) :



Weigthed basins' graph

Definition : Edge weight.
$p(s \rightarrow s^{'})$ as the probability to pass from s to $s^{'}$ with the local operation.

$$p(s \rightarrow b_j) = \sum_{s^{'} \in b_j} p(s \rightarrow s^{'})$$

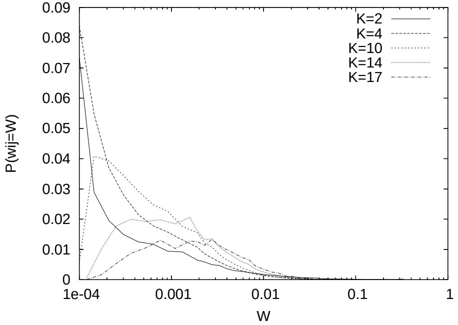
$$p(b_i \rightarrow b_j) = \frac{1}{\sharp b_i} \sum_{s \in b_i} p(s \rightarrow b_j)$$

Weigthed basins' graph

K	$\bar{n}_v$	$\bar{n}_e$	$\bar{C}^w$	$\bar{Y}$	$\bar{d}$
		$N = 18$			
2	50_{25}	1579_{1854}	$0.95_{0.0291}$	$0.307_{0.0630}$	73_{15}
4	330_{72}	26266_{7056}	$0.92_{0.0137}$	$0.127_{0.0081}$	174_{9}
6	994_{73}	146441_{18685}	$0.78_{0.0155}$	$0.076_{0.0044}$	237_{5}
8	$2,093_{70}$	354009_{18722}	$0.64_{0.0097}$	$0.056_{0.0012}$	273_{2}
10	$3,619_{61}$	620521_{20318}	$0.52_{0.0071}$	$0.044_{0.0007}$	292_{1}
12	$5,657_{59}$	899742_{14011}	$0.43_{0.0037}$	$0.038_{0.0003}$	297_{1}
14	$8,352_{60}$	1163640_{11935}	$0.36_{0.0023}$	$0.034_{0.0002}$	293_{1}
16	$11,797_{63}$	1406870_{6622}	$0.32_{0.0012}$	$0.032_{0.0001}$	283_{1}
17	$13,795_{77}$	1524730_{4818}	$0.30_{0.0009}$	$0.032_{0.0001}$	277_{1}

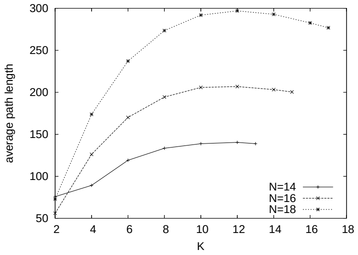
Weigthed basins' graph

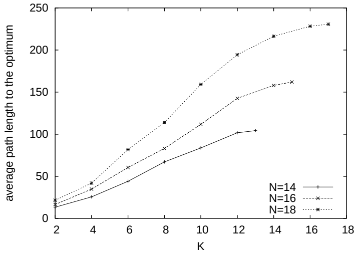
Probability distribution of the network weights w_{ij} for outgoing edges with $j \neq i$ in logscale on x-axis. Averages of 30 instances for each N and K are reported.



more than 99% of the solutions are on the boundary of the basin.

Weigthed basins' graph





Weigthed basins' graph : return to Doye

- Doye found inherent networks were of the scale-free type with the global minimum being often the most connected node : "funnel" landscape
- not "funnel" landscape for the NK-landscapes

Some conclusions

- First problem was MAX-SAT :
  too small number of basins
  use the NK-landscape which is a generalisation of MAX-SAT
- the methods of the physician (complex systems) help to understand the structure of search :
  the distribution of size of basin should be the same for other problem
- use the results to estimate the number and size of basins :
  number of runs to find the global optimum
  restart strategy possible

Discussion

- main principles in the design of metaheuristics are known :
  - probability to explore new solutions
  - use memory
  - use a population
  - combined different methods
- metaheuristics are very flexible and can be at the crossroads of different optimisation methods
- easy parallelisation
- lack of automatisation of the design of the "ingredients"

$\longrightarrow$ study the structure of the problem to design automatisation of search process (learning)

The last conclusion before your questions

- Domaine il y a encore beaucoup de choses à explorer
- méthodes de plus en plus utiliser dans l'industrie à mesure que les bibliothèques (et la connaissance) se diffusent

Your questions