Black-box expensive optimization: Learn to optimize

Sébastien Verel

Laboratoire d'Informatique, Signal et Image de la Côte d'opale (LISIC)
Université du Littoral Côte d'Opale, Calais, France
http://www-lisic.univ-littoral.fr/~verel/

August, 29th/30th, 2019

AI : Machine Learning, Optimization, perception, etc.

Learning :

Minimize an error function

$\{M_\theta\}$: models to learn on data
Search $\theta^\star = \arg\min_\theta Error(M_\theta, data)$

According to the model dimension, variables, error function, etc., huge number of optimization algorithms

Optimization :

Learn a design algorithm

$\{A_\theta\}$: search algorithms for problems (X, f)
Learn $A_{\theta^\star}$ such that $x = A_{\theta^\star}(X, f)$ is a good solution

According to the class of algorithms, search spaces, functions, etc., huge number of learning algorithms

Black-box (expensive) optimization

$x \longrightarrow$ $\longrightarrow f(x)$

No direct analytic definition of the objective function f

Objective fonction :

- can be irregular, non continuous, non differentiable, etc.
- given by a computation or an (expensive) simulation

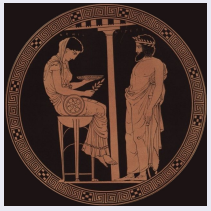
Few examples from the OSMOSE team at LISIC :

- Nuclear power plant (Valentin Drouet),
- Logistic simulation (Brahim Aboutaib),
- Mobility simulation (Florian Leprêtre),
- Plant's biology, plant growth (Amaury Dubois),
- Cellular automata, grammatical inference (Bilal El Ghadyry),

Real-world black-box expensive optimization

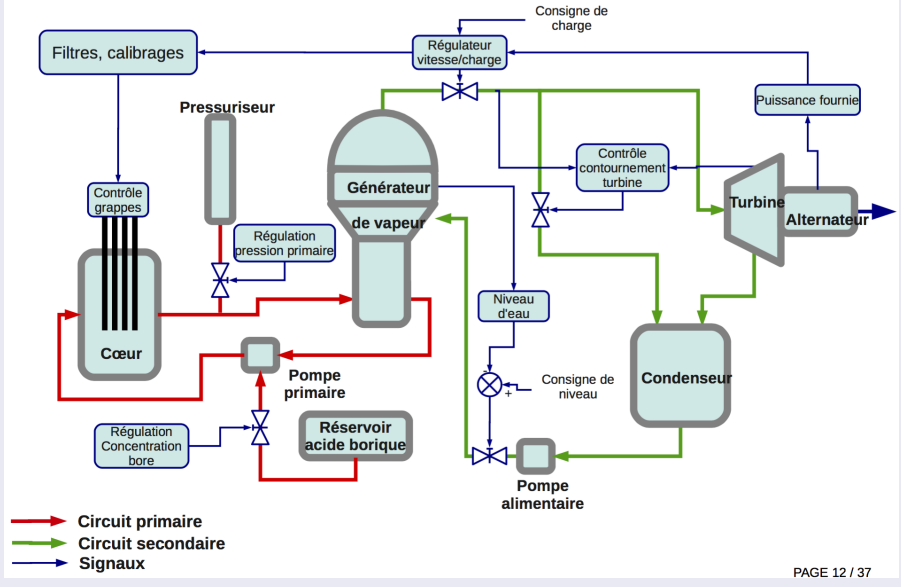
PhD of Mathieu Muniglia, 2014-2017, Valentin Drouet, 2017-2020,
Jean-Michel Do, Jean-Charles Le Pallec, CEA-Saclay, Paris

$x \longrightarrow$



$\longrightarrow f(x)$

$(73, .., 8) \rightarrow$



Multi-physic simulator

$\rightarrow (\Delta_z P, V_e)$

Expensive multiobjective optimization :

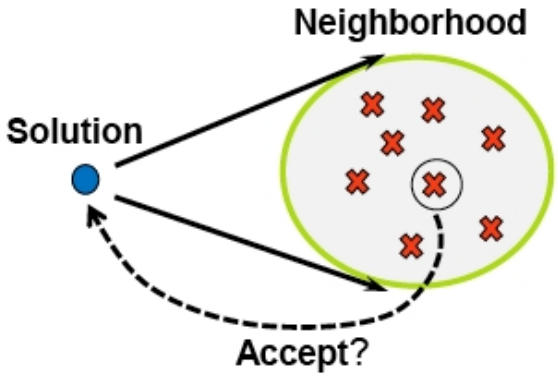
parallel computing, and surrogate model.

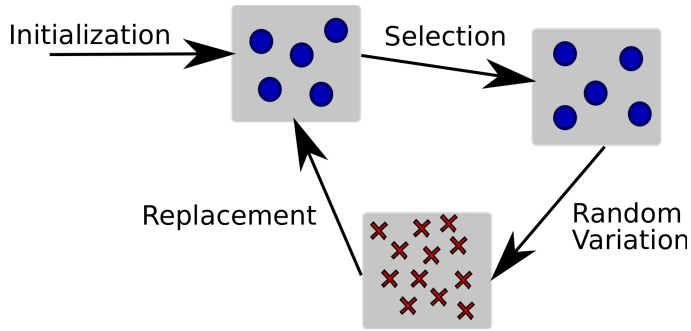
Search algorithms

Principle

Enumeration of the search space

- A lot of ways to enumerate the search space
- Main topics : local search, evolutionary algorithms, metaheuristics, etc.





Main challenges :

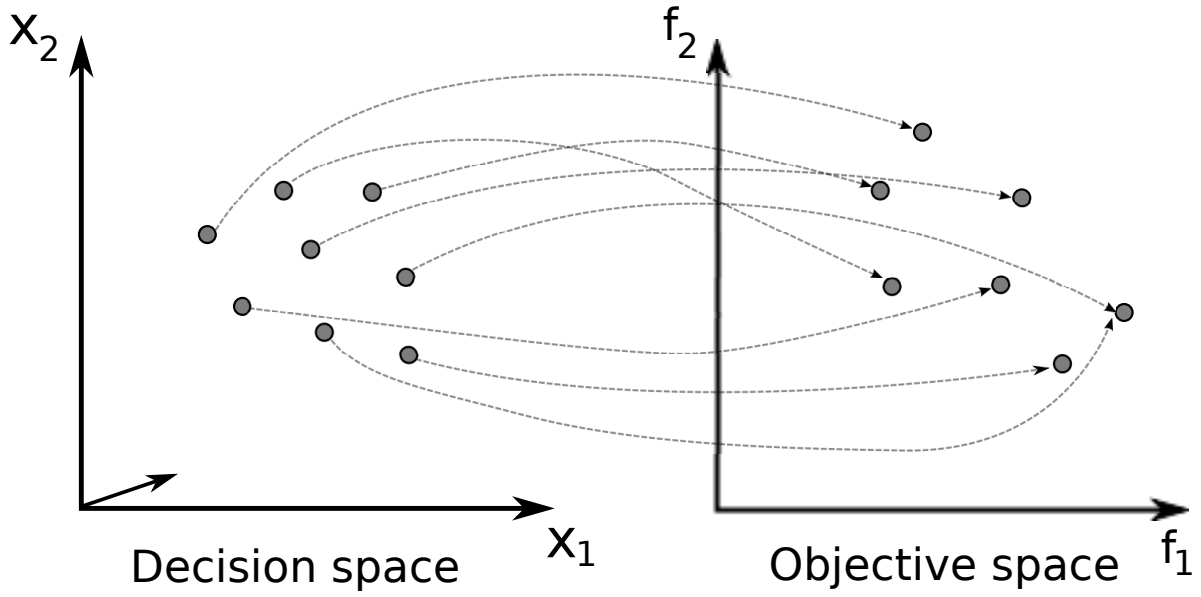
Understand search dynamics,

Parameters, design components

Multiobjective optimization

Multiobjective optimization problem

- $\mathcal{X}$: set of feasible solutions in the decision space
- $M \geqslant 2$ objective functions $f = (f_1, f_2, \ldots, f_M)$ (to maximize)
- $\mathcal{Z} = f(\mathcal{X}) \subseteq \mathbb{R}^M$: set of feasible outcome vectors in the objective space

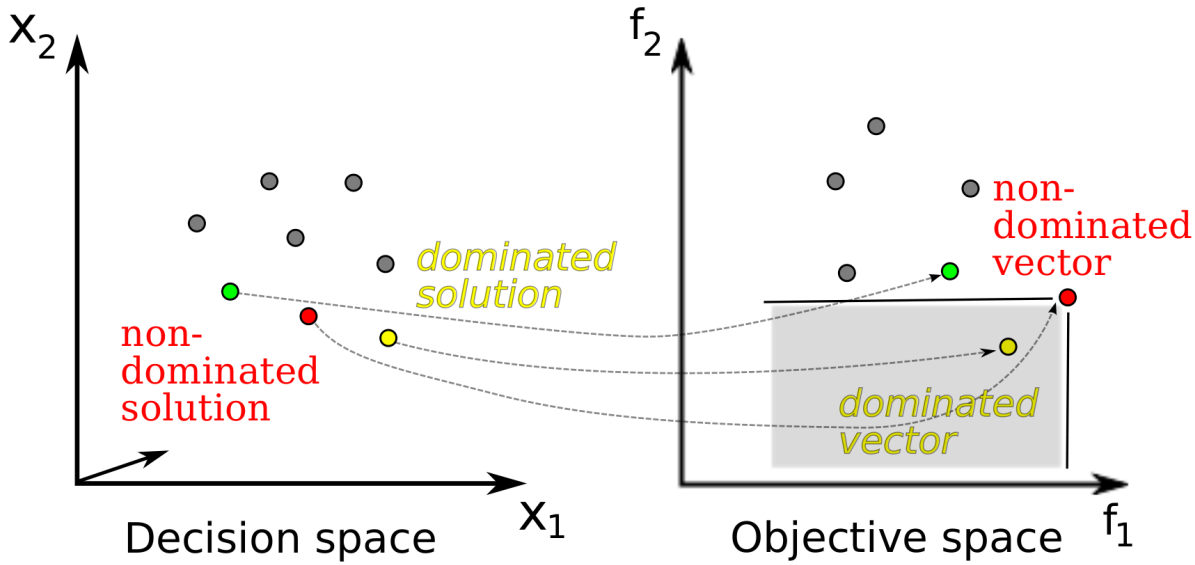


Definition

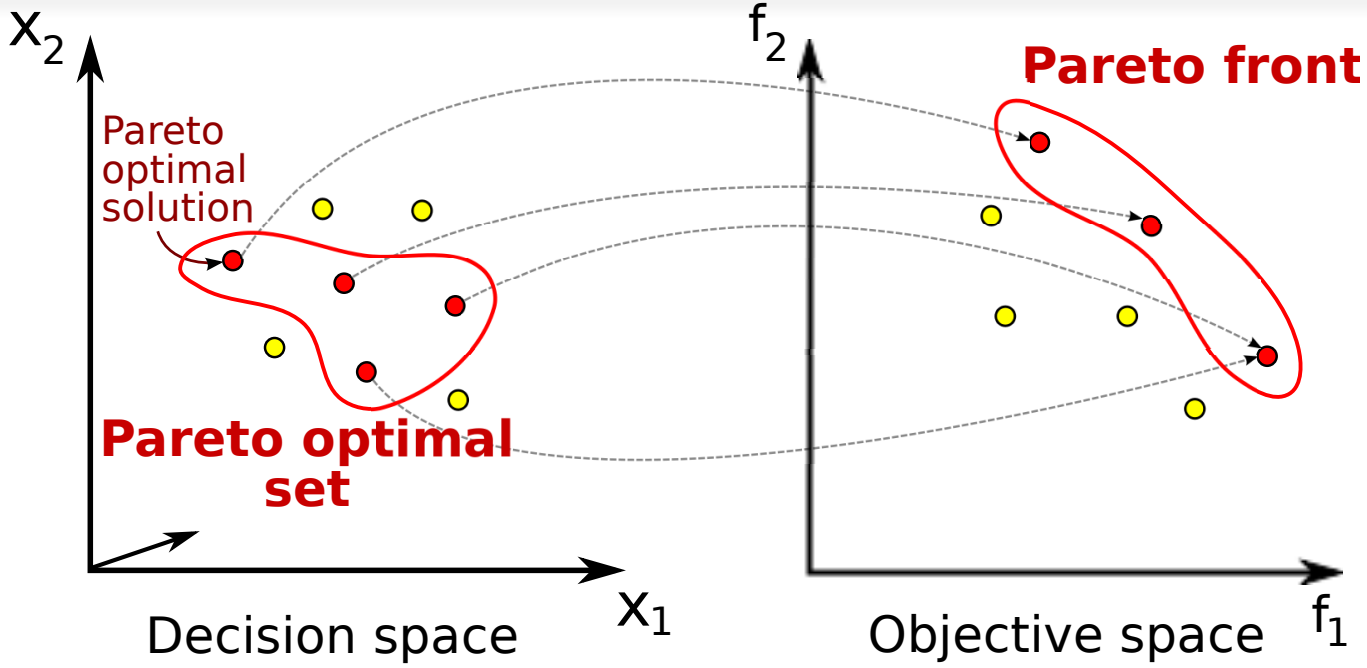
Pareto dominance relation

A solution $x \in \mathcal{X}$ dominates a solution $x' \in \mathcal{X}$ ($x' \prec x$) iff

- $\forall i \in \{1, 2, \ldots, M\}$, $f_i(x') \leqslant f_i(x)$
- $\exists j \in \{1, 2, \ldots, M\}$ such that $f_j(x') < f_j(x)$



Pareto set, Pareto front



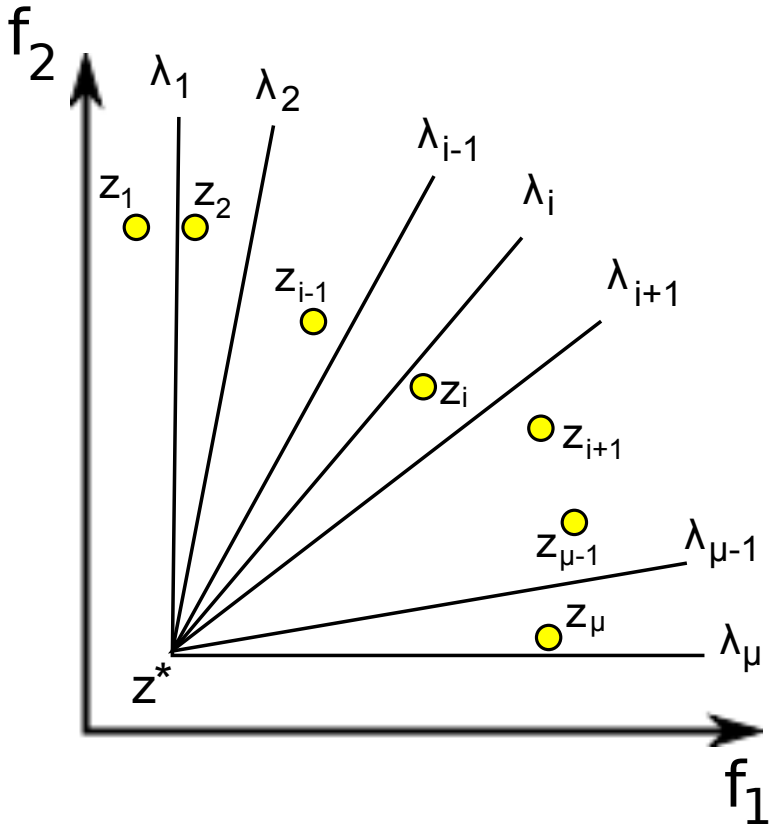
Goal

Find the Pareto Optimal Set,
or a good approximation of the Pareto Optimal Set

NPP problem

discrete optimization problem,
$\approx$ 30min of computation per value, 24h total, 10^3 to 3.10^3 cores.

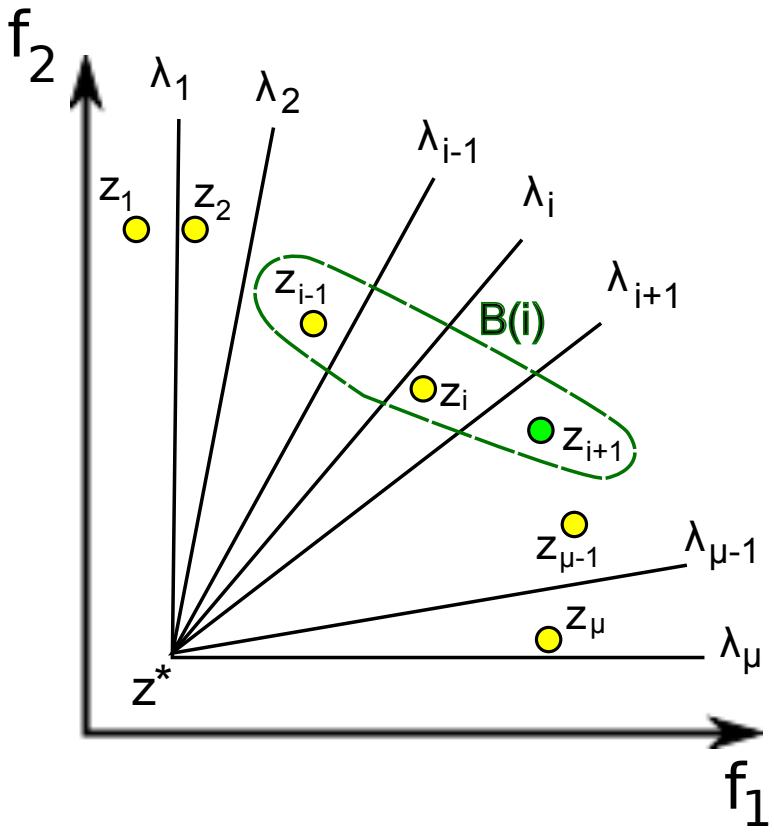
Decomposition based approaches, MOEA/D



Population at iteration t

- Minimization problem
- One solution x_i for each sub pb. i
- Representation of solutions in objective space : $z_i = g(x_i|\lambda_i, z_i^\star)$
- Same reference point for all sub-pb. $z^\star = z_1^\star = \ldots = z_\mu^\star$
- Scalar function g : Weighted Tchebycheff
- Neighborhood size $\sharp B(i) = T = 3$

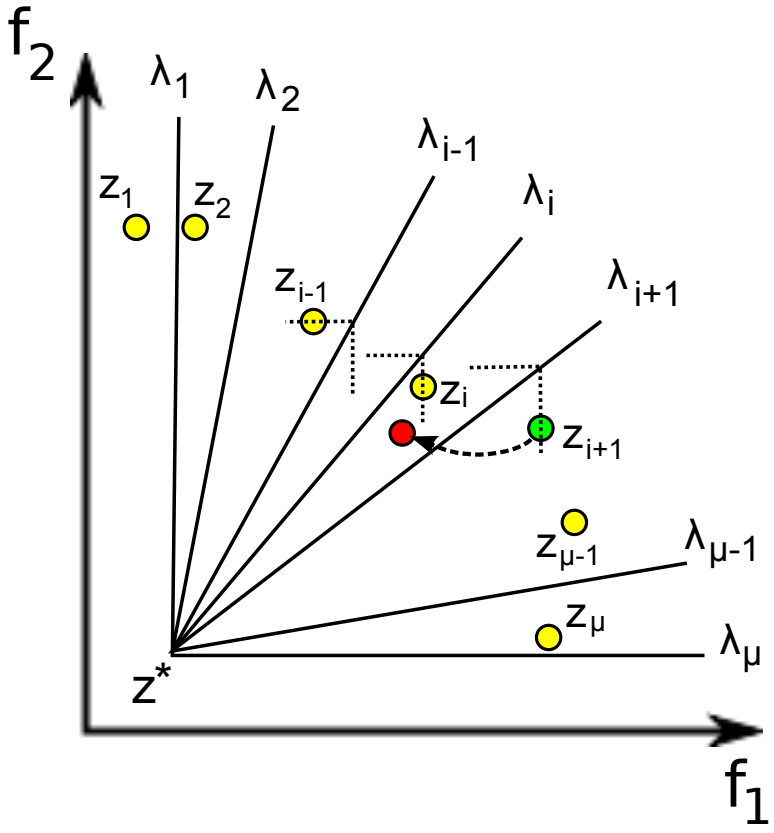
Decomposition based approaches, MOEA/D



From the neigh. $B(i)$ of sub-pb. i, x_{i+1} is selected

- Minimization problem
- One solution x_i for each sub pb. i
- Representation of solutions in objective space : $z_i = g(x_i|\lambda_i, z_i^\star)$
- Same reference point for all sub-pb. $z^\star = z_1^\star = \ldots = z_\mu^\star$
- Scalar function g : Weighted Tchebycheff
- Neighborhood size $\sharp B(i) = T = 3$

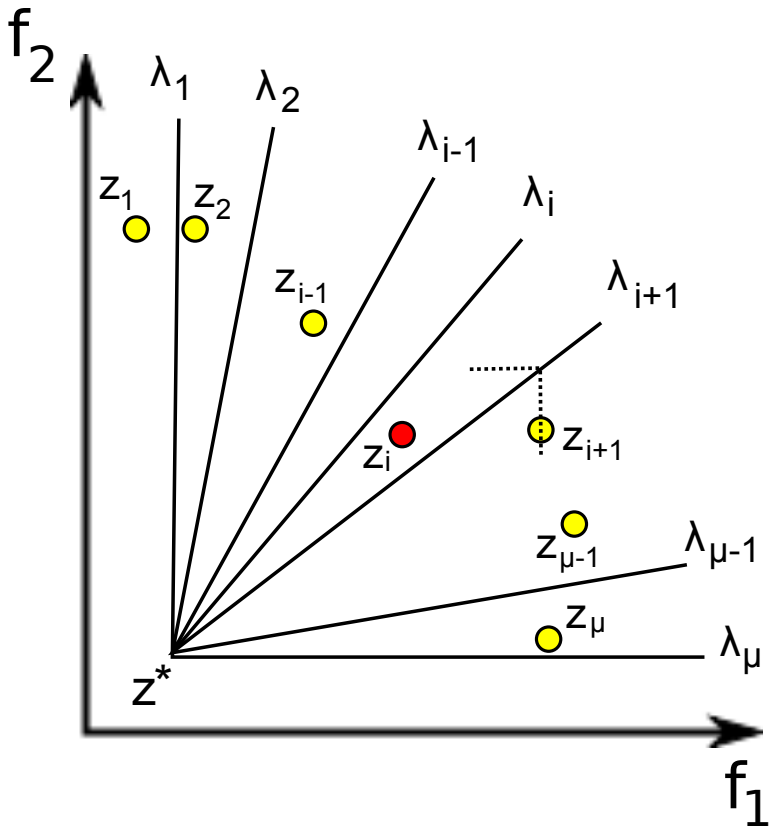
Decomposition based approaches, MOEA/D



The mutated solution y is created

- Minimization problem
- One solution x_i for each sub pb. i
- Representation of solutions in objective space : $z_i = g(x_i|\lambda_i, z_i^\star)$
- Same reference point for all sub-pb. $z^\star = z_1^\star = \ldots = z_\mu^\star$
- Scalar function g : Weighted Tchebycheff
- Neighborhood size $\sharp B(i) = T = 3$

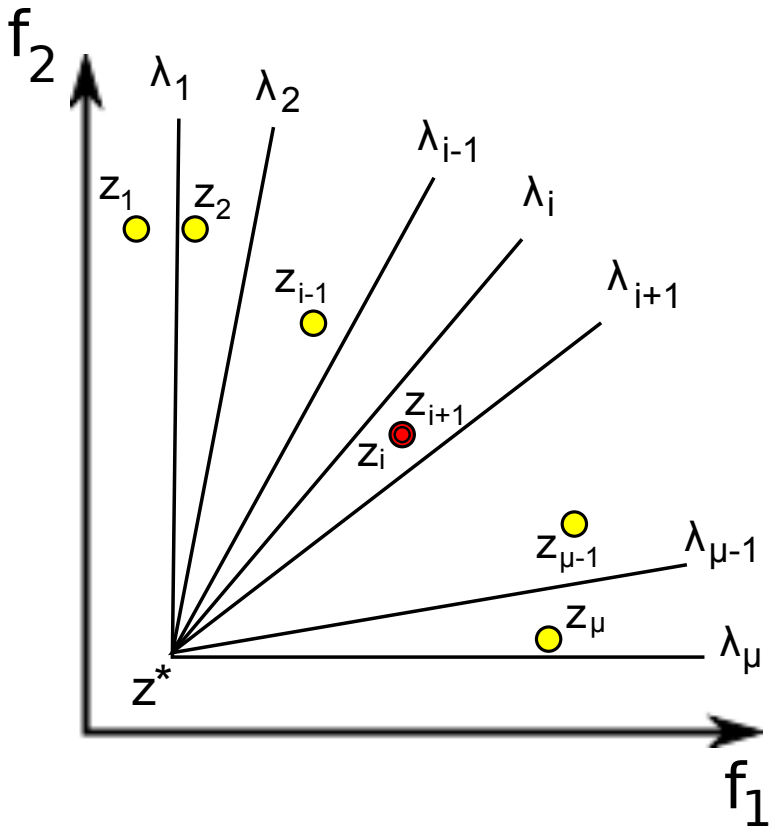
Decomposition based approaches, MOEA/D



According to scalar fonction,
y is worst than x_{i-1},
y is better than x_i and replaces it.

- Minimization problem
- One solution x_i for each sub pb. i
- Representation of solutions in objective space : $z_i = g(x_i|\lambda_i, z_i^\star)$
- Same reference point for all sub-pb. $z^\star = z_1^\star = \ldots = z_\mu^\star$
- Scalar function g :
  Weighted Tchebycheff
- Neighborhood size $\sharp B(i) = T = 3$

Decomposition based approaches, MOEA/D

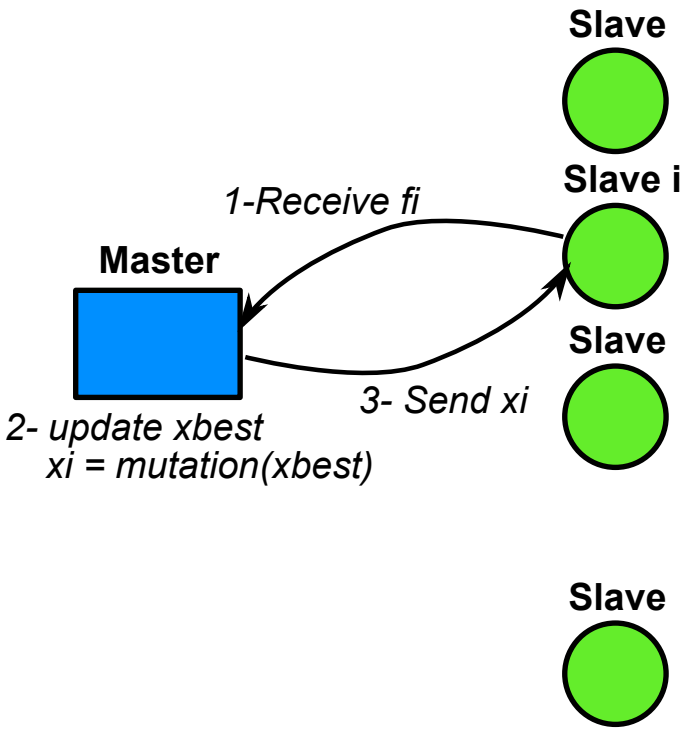


According to scalar fonction,
y is also better than x_{i+1}
and replaces it for the next iteration.

- Minimization problem
- One solution x_i for each sub pb. i
- Representation of solutions in objective space : $z_i = g(x_i|\lambda_i, z_i^\star)$
- Same reference point for all sub-pb. $z^\star = z_1^\star = \ldots = z_\mu^\star$
- Scalar function g :
  Weighted Tchebycheff
- Neighborhood size $\sharp B(i) = T = 3$

Asynchronous distributed $(1 + \lambda)$-Evolution Strategy

Master-slaves architecture



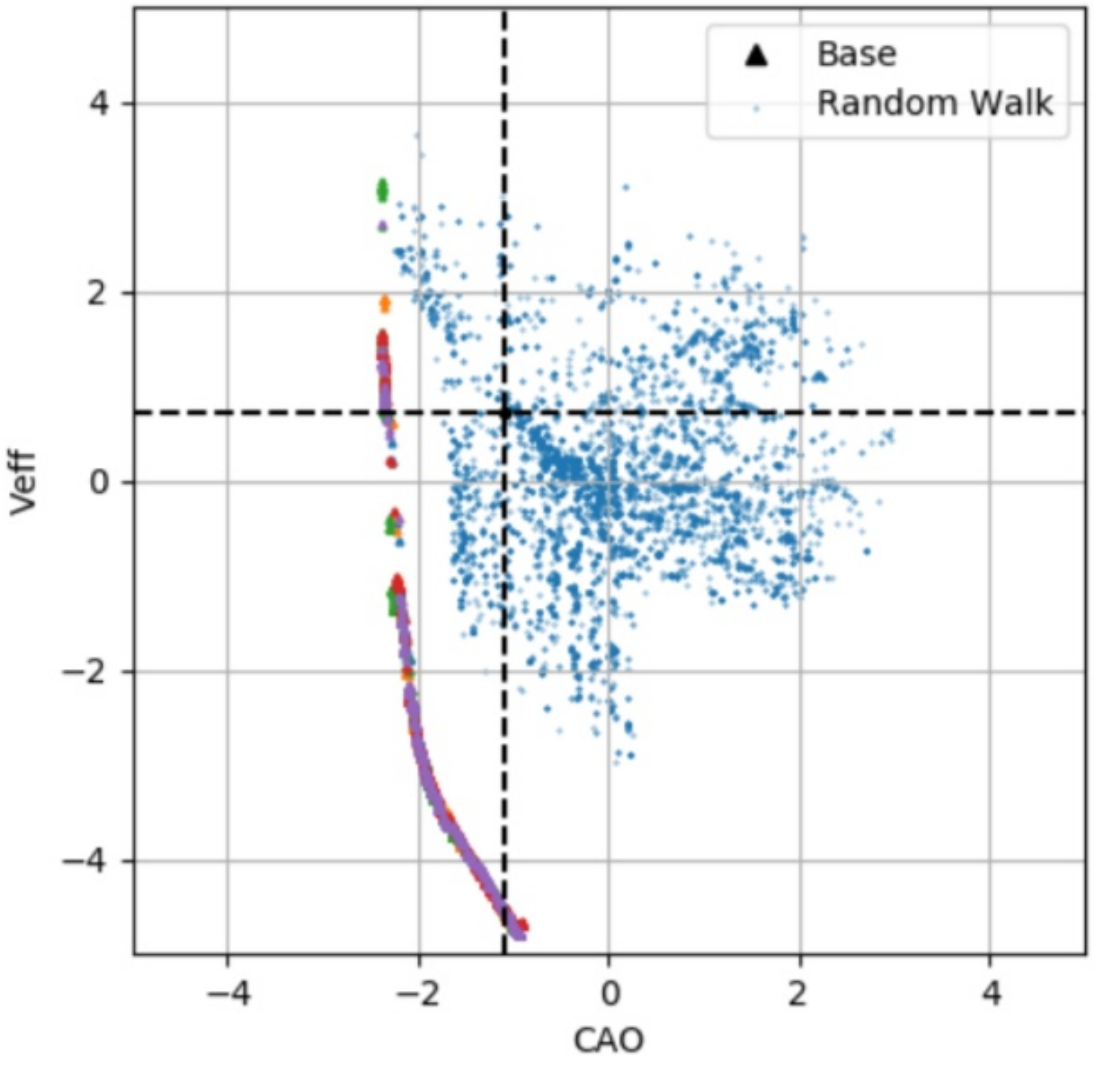
Algorithm on Master

$\{x_1, \ldots, x_\lambda\} \leftarrow$ **Initialization**()
for $i = 1..\lambda$ **do**
  Send (Non-blocking) x_i to slave S_i
end for
$x_{best} \leftarrow \emptyset$, and $f_{best} \leftarrow \infty$
repeat
  if there is a pending mess. from S_i **then**
    Receive fitness f_i of x_i from S_i
    if $f_i \leqslant f_{best}$ **then**
      $x_{best} \leftarrow x_i$, and $f_{best} \leftarrow f_i$
    end if
    $x_i \leftarrow \texttt{mutation}(x_{best})$
    Send (Non-blocking) x_i to slave S_i
  end if
until time limit

Parameters of mutation :

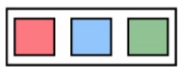
fitness landscape analysis (random walk, and var. importance)

Résultat

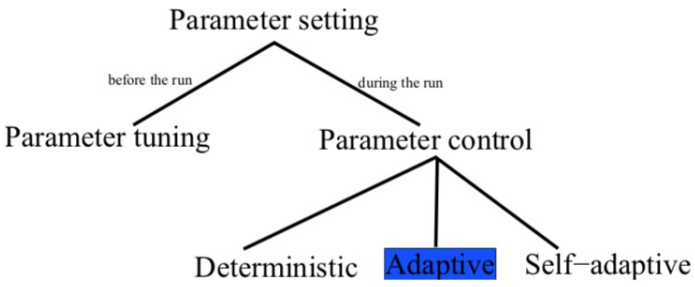


Adaptive distributed optimization algorithms

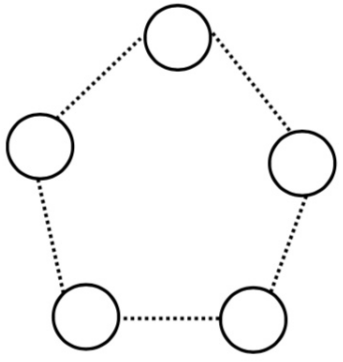
B. Aboutaib (2018-), C. Jankee (2014-2018), C. Fonlupt (LISIC), B. Derbel (univ. Lille)



A portfolio of metaheuristics



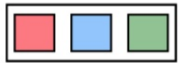
Parameter setting [Eiben et al., 1999]



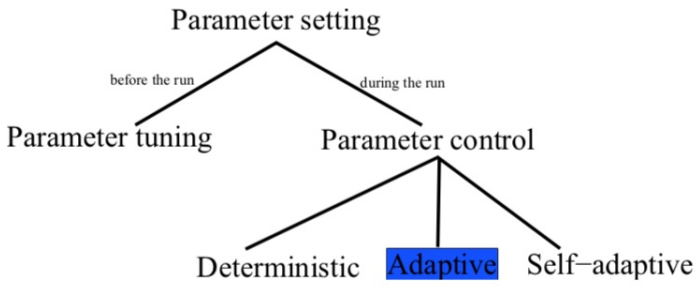
Distributed algorithm

Adaptive distributed optimization algorithms

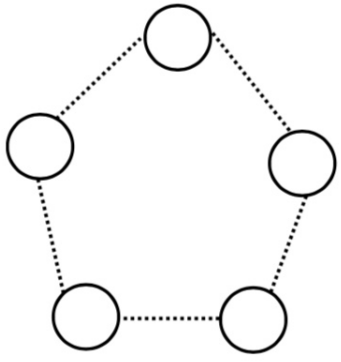
B. Aboutaib (2018-), C. Jankee (2014-2018), C. Fonlupt (LISIC), B. Derbel (univ. Lille)



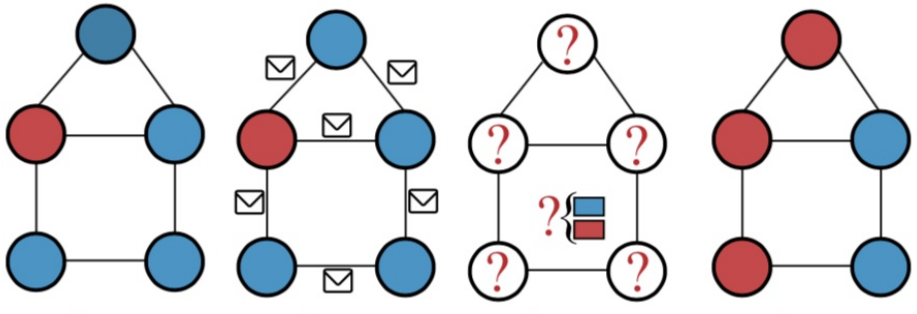
A portfolio of metaheuristics



Parameter setting [Eiben et al., 1999]



Distributed algorithm



repeat
- Migration of solutions
- Reward computation
- Metaheuristic selection
- Apply metaheuristic

until *Stopping condition is satisfied*

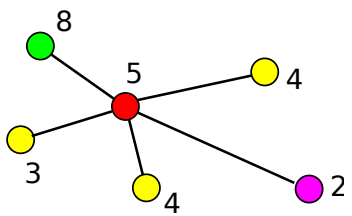
Main challenges

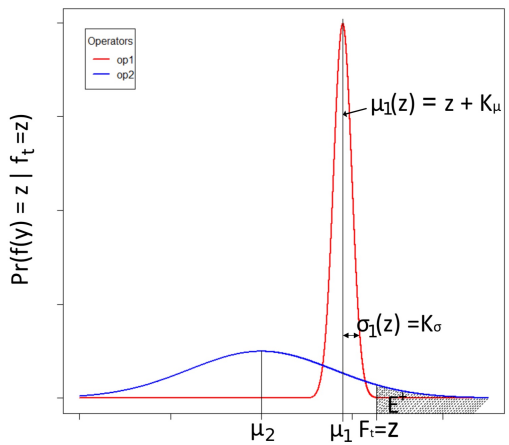
- How to select an algorithm ?

  Design reinforcement learning methods for distributed computing (ϵ-greedy, adapt. pursuit, UCB, ...)

- How to compute a reward ?

  Aggregation function of local rewards (mean, max, etc.) for a global selection



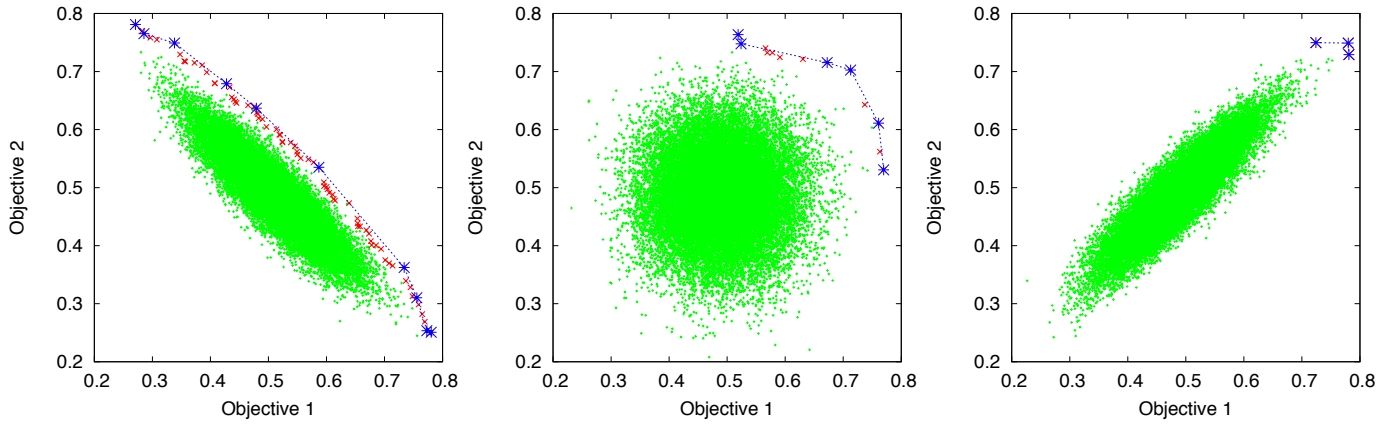


Methodology : Use designed benchmark functions with designed properties and experimental analysis

Features to learn : Mult.-Obj. fitness landscape

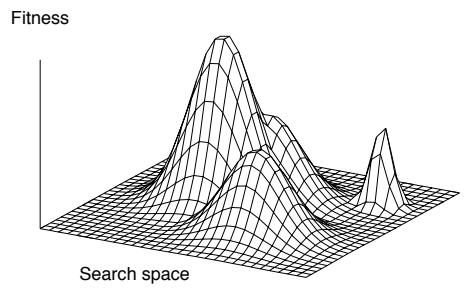
K. Tanaka, H. Aguirre (Univ. Shinshu), A. Liefooghe, B. Derbel (univ. Lille), 2010 - 2018

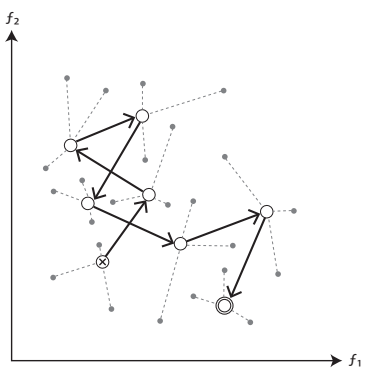
Fitness landscape : $(\mathcal{X}, f, \mathcal{N})$, Search space, obj. func., neighborhood relation

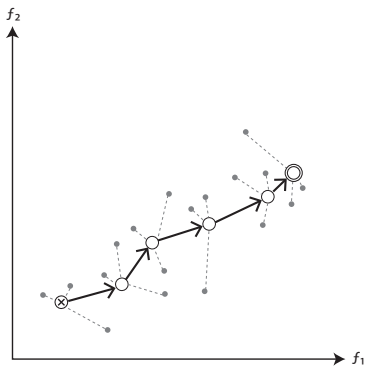


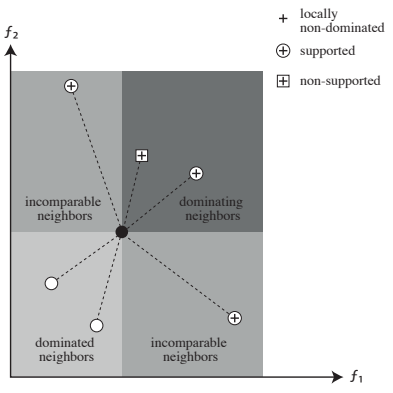
conflicting objectives *independent objectives* *correlated objectives*

Fitness landscape analysis :



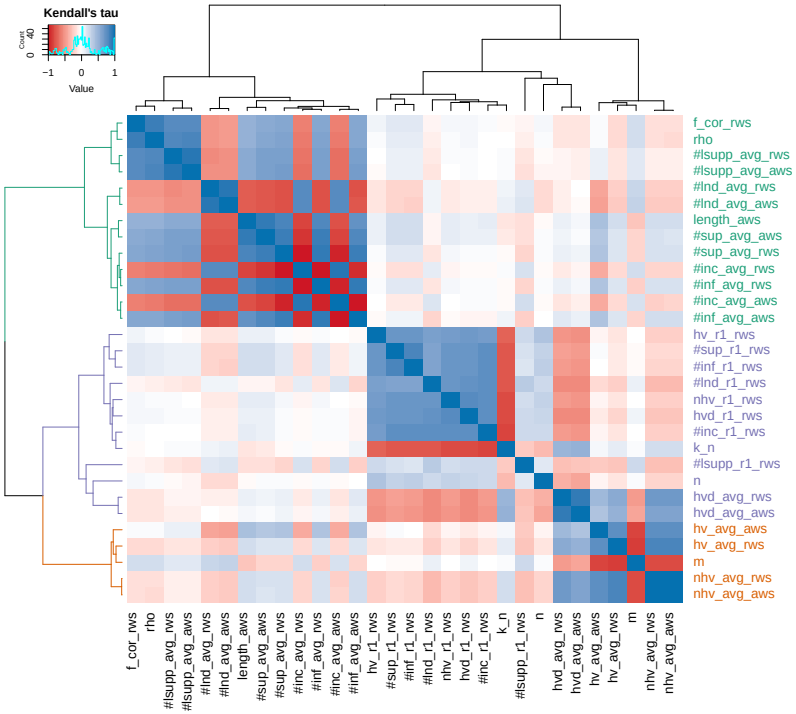






Features to learn : Mult.-Obj. fitness landscape

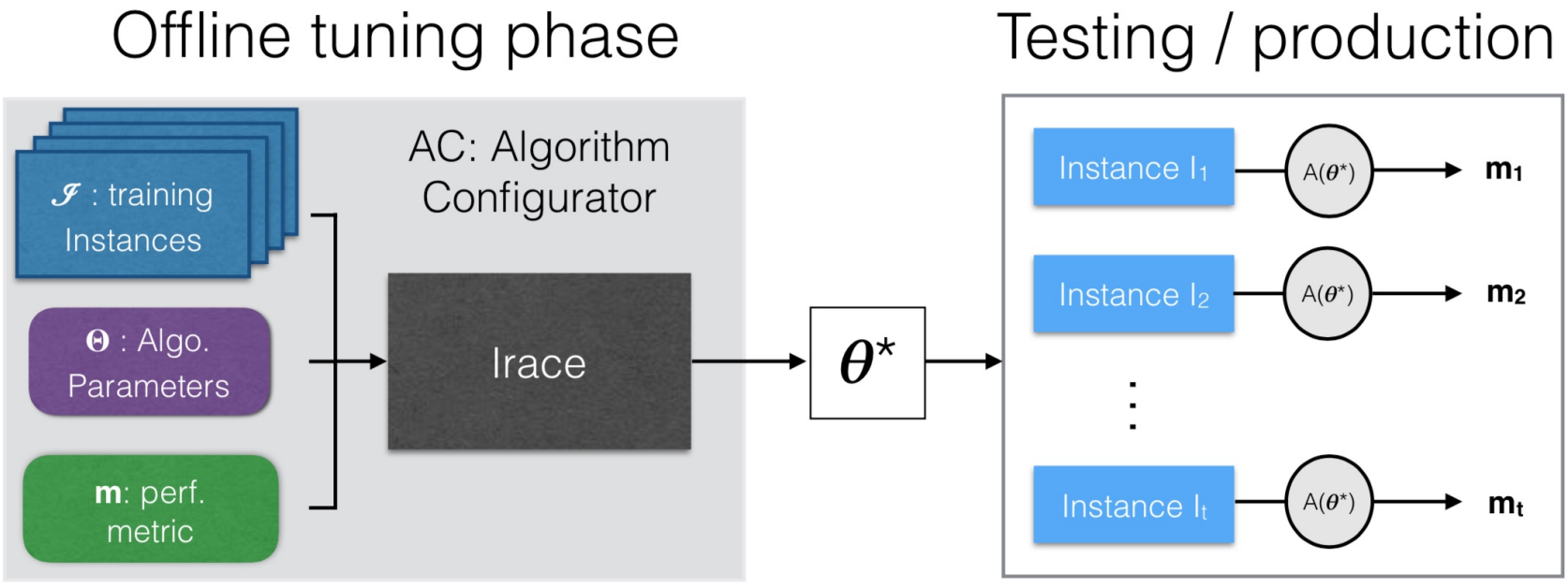
Performance prediction of MO algorithms
based on fitness landscape features
(regression model : random forest, etc.)



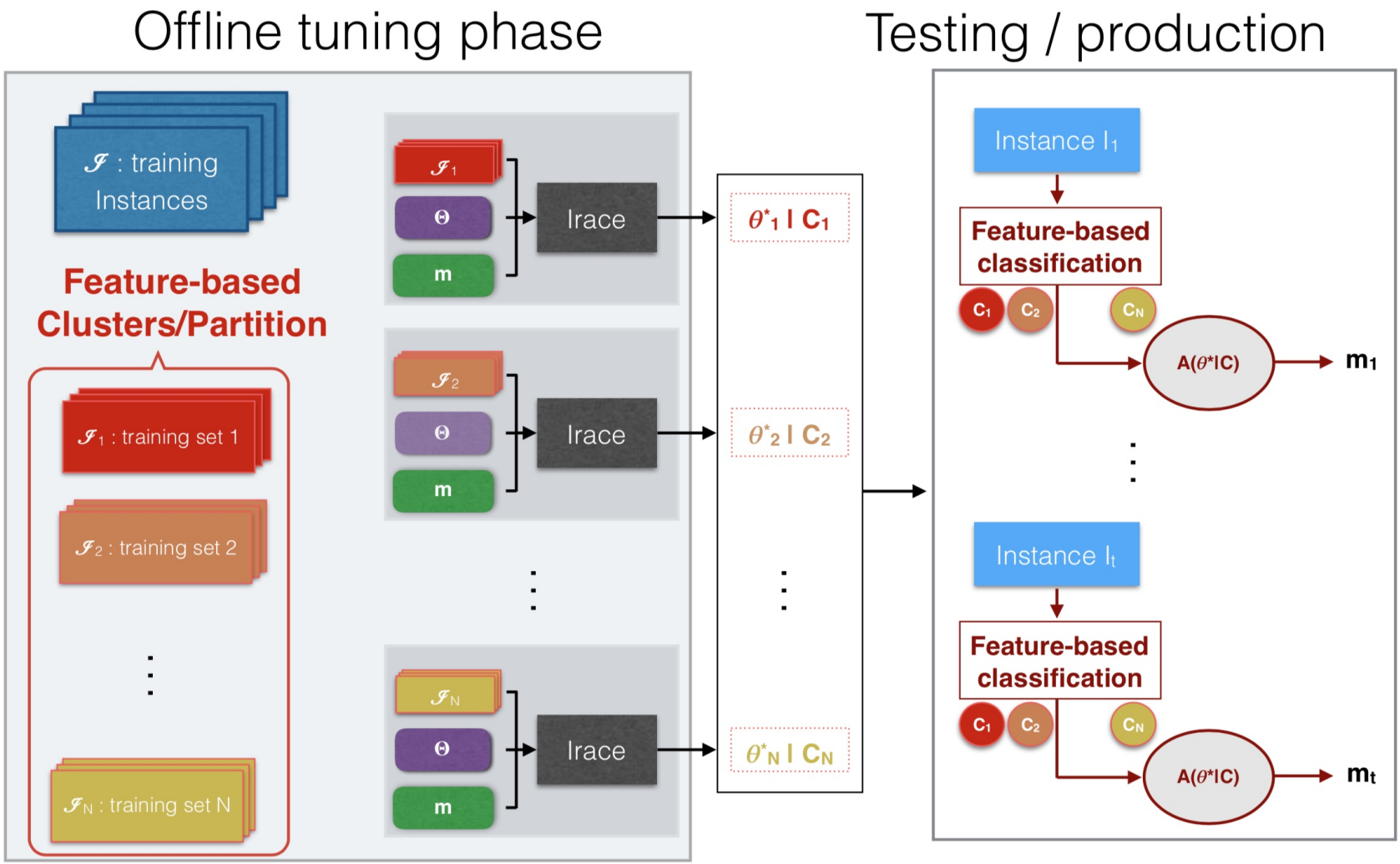
Perf. prediction (cross-val.)

feature set	MAE	MSE	R^2	rank
	GSEMO			
all	**0.007781**	**0.000118**	**0.951609**	1
enumeration	0.008411	0.000142	0.943046	2
sampling all	0.009113	0.000161	0.932975	3
sampling rws	0.009284	0.000167	0.930728	4
sampling aws	0.010241	0.000195	0.917563	5
$\{r, m, n, k/n\}$	0.010609	0.000215	0.911350	6
$\{r, m, n\}$	0.026974	0.001123	0.518505	7
$\{m, n\}$	0.032150	0.001545	0.340715	8

Learning/tuning parameters according to features

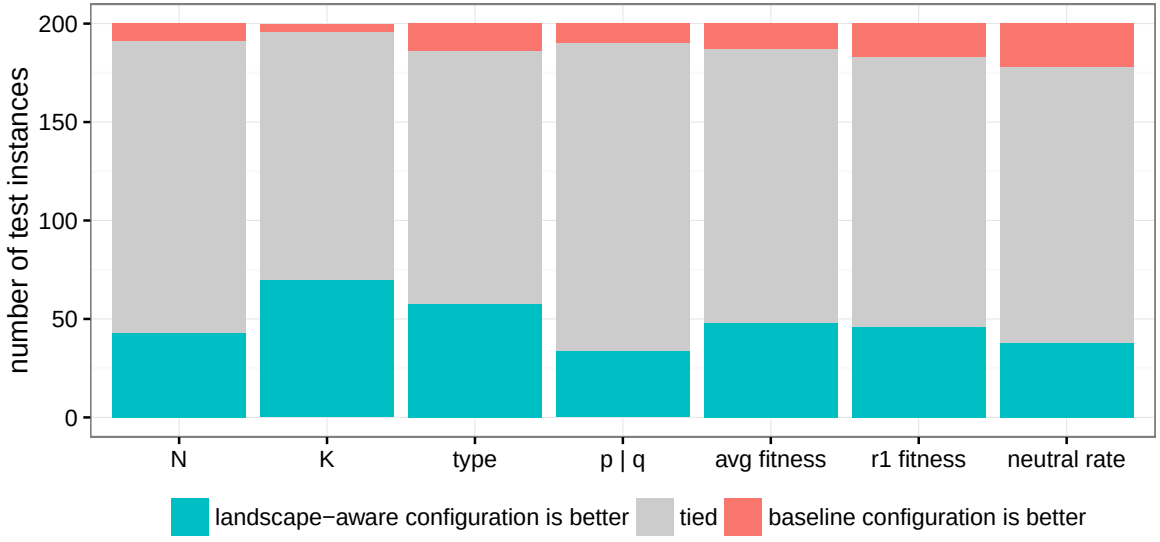


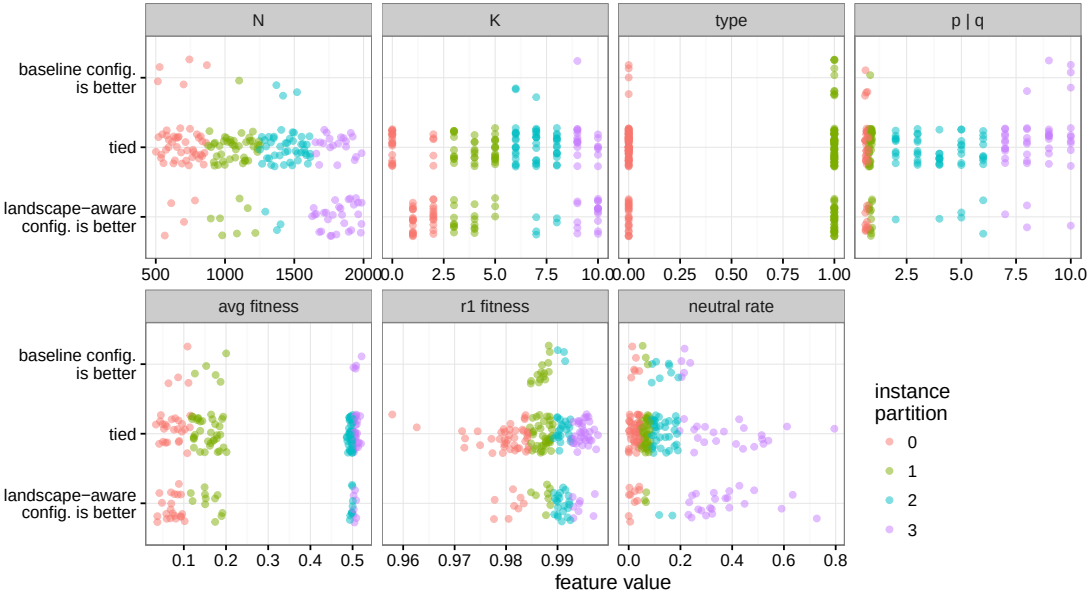
Learning/tuning parameters according to features



Learning/tuning parameters according to features

Overview of some results





Surrogate-assisted opt. of pseudo-boolean problems

Florain Lepretre, Virginie Marion, Cyril Fonlupt,
K. Tanaka, H. Aguirre (Univ. Shinshu), A. Liefooghe, B. Derbel (univ. Lille)

Goal : Replace/learn the (expensive) objective function with a (sheep) meta-model during the optimization process

Continuous optimization : huge number of works

Gaussian Process/krigging (EGO), polynomial chaos, NN, etc.

Combinatorial optimization : few number of works.

$$f : \{0,1\}^n \longrightarrow \mathbb{R}$$

Surrogate-Assisted Optimization

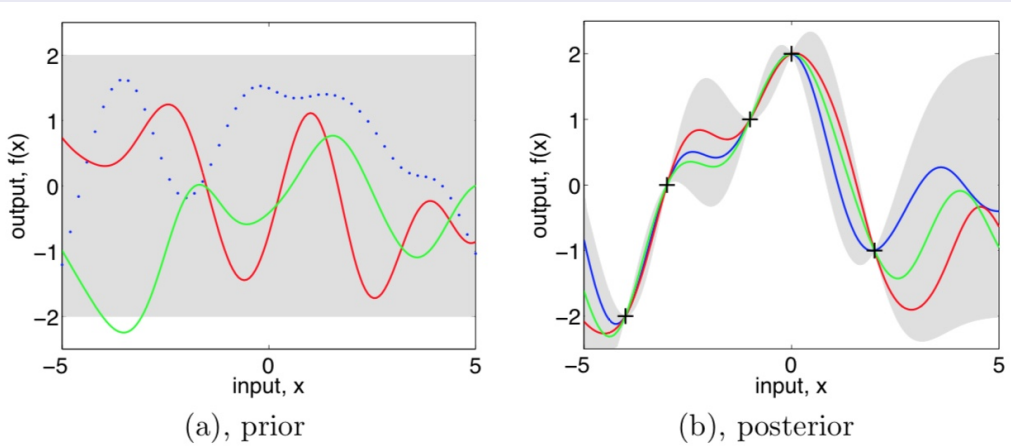
$X \leftarrow$ initial sample
repeat
  $M \leftarrow$ Build model from X
  $x^\star \leftarrow$ Optimize *w.r.t.* an acquisition function
  $X \leftarrow X \cup \{x^\star\}$
until time limit

Example : Efficient Global Optimizer

- Model : Gaussian Process (GP)
- Acquisition function : Expected Improvement

GP : Random variables which have joint Gaussian distribution.

mean : $m(y(x)) = \mu$
covariance : $cov(y(x), y(x')) = exp(-\theta d(x, x')^p)$



(a), prior (b), posterior

from : Rasmussen, Williams, GP for ML, MIT Press, 2006.

A Hamming distance can be used for pseudo-boolean functions.

Surrogate model for pseudo-boolean functions

Walsh functions

$$\forall x \in \{0,1\}^n, \quad \varphi_k(x) = (-1)^{\sum_{j=0}^{n-1} k_j x_j}$$

Normal, and orthogonal basis

$$\frac{1}{2^n} \sum_{x \in \{0,1\}^n} \varphi_i(x)\varphi_j(x) = \delta_{ij}$$

Any function can be written as :

$$f(x) = \sum_{k=0}^{2^n-1} w_k . \varphi_k(x)$$

Example, with order 2 :

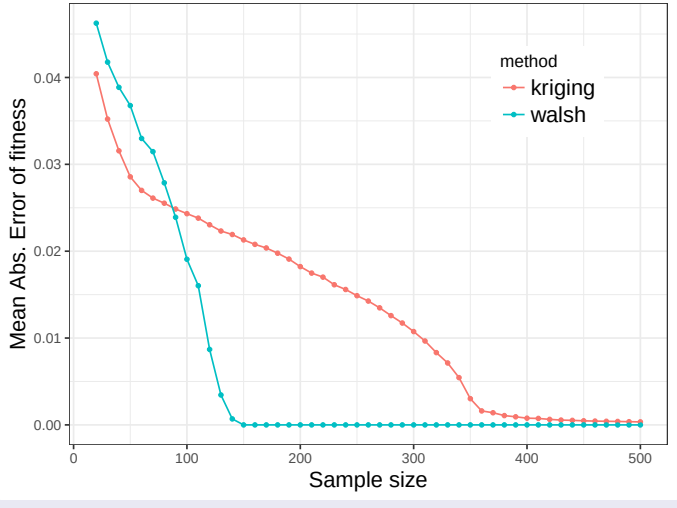
$$f(x) = w_0 + \sum_{i=1}^{n} w_i . (-1)^{x_i} + \sum_{i<j=1}^{n} w_{ij} . (-1)^{x_i + x_j}$$

Walsh Surrogate-assisted Optimizer (WSaO)

Surrogate model : $W(x) = \sum_{k \ : \ \mathrm{ord}(\varphi_k) \leqslant d} \widehat{w}_k . \varphi_k(x)$

Estimation of the coefficients :

linear regression with sparse techniques, LARS, or LASSO



Walsh Surrogate-assisted Optimizer (WSaO)

$X \leftarrow$ initial sample
repeat
  $W \leftarrow$ Build Walsh model from X
  $x^\star \leftarrow$ Optimize W
  $X \leftarrow X \cup \{x^\star\}$
until time limit

Optimizer

Efficient optimization algorithm (ILS) using the additive property :

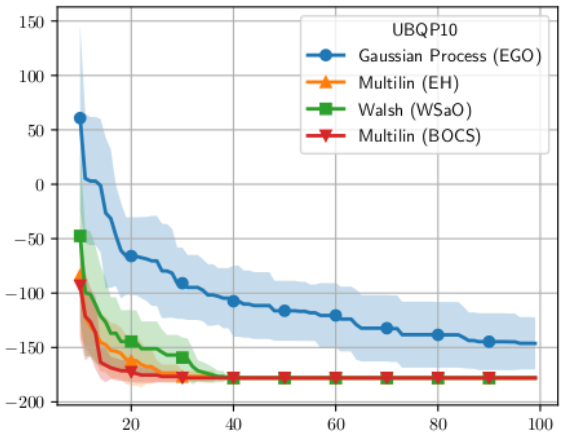
$$\delta_i(x) = W(x \bigoplus i) - W(x) = -2 \sum_{k \supset i} w_k \varphi_k(x)$$
$$\delta_{ij}(x) = \delta_i(x \bigoplus j) - \delta_i(x) = 4 \sum_{k \supset i \& k \supset j} w_k \varphi_k(x)$$

Find best improving move in $O(d)$ at each step

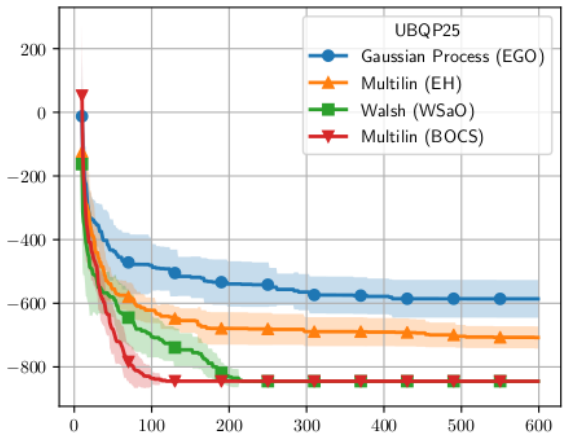
Walsh Surrogate-assisted Optimizer (WSaO)

UBQP (min.)

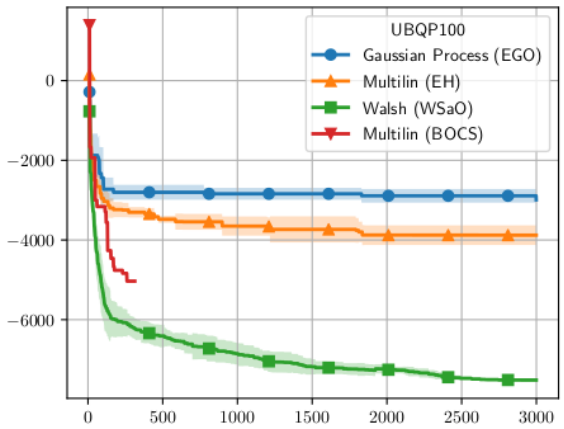
$n = 10$



$n = 25$



$n = 100$

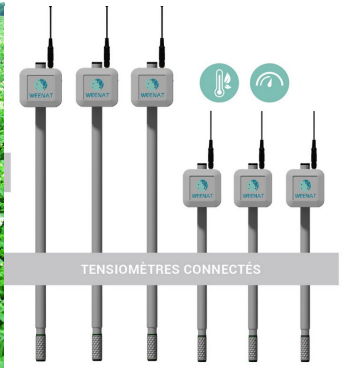


- Krigging : distance in high dimension
- BOCS : bayesian estimation of multilinear polynome, basis, optimization method (SA)

Will be applied to bus-stop position problem for Calais city

Agronomy : plant irrigation

Amaury Dubois, 2018-2021, Jérôme Leroy, Weenat WEENAT, F. Teytaud, Eric Ramat

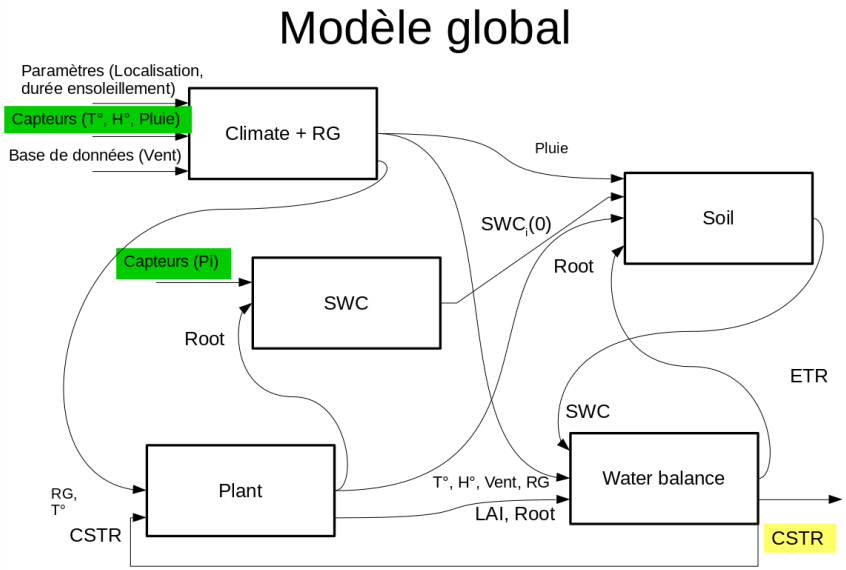


Goal :

Setting of
plant and sol parameters
for hydric stress prediction

Key points :

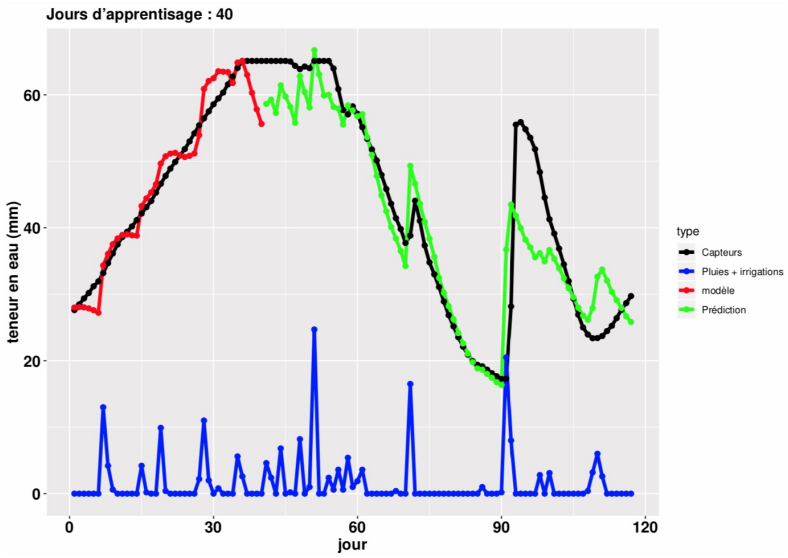
- Algorithm design for multimodal optimization problems (QRDS)
- Combine offline learning (optimization techniques), and online learning during saison (ensemble learning techniques)

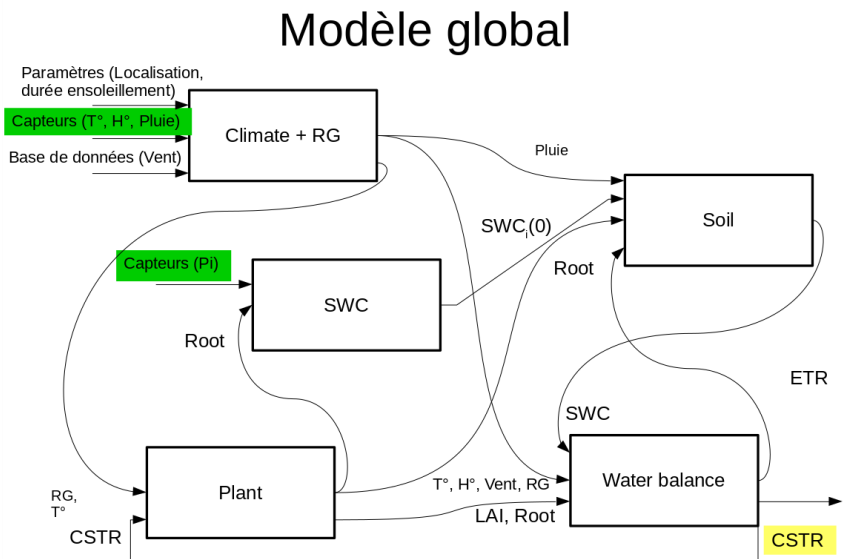


Numerical simulator

Agronomy : plant irrigation

Amaury Dubois, 2018-2021, Jérôme Leroy, Weenat WEENAT, F. Teytaud, Eric Ramat





Goal :

Setting of
plant and sol parameters
for hydric stress prediction

Key points :

- Algorithm design for multimodal optimization problems (QRDS)
- Combine offline learning (optimization techniques), and online learning during saison (ensemble learning techniques)

Grammatical Inference

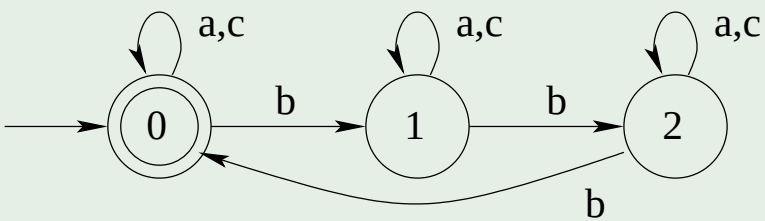
Bilal El Ghadyry (2019-2022), Faissal Ouardi (Univ. M5, Rabat)

Goal

Learning a Determinist Finite Automata from examples of words

Example

$\{abbab, bababaa, bbb, bbbabbb\} \in L$
$\{bb, ba, bbbb, abbabbb\} \notin L$



Applications : biologie (génétique), physique, informatique, etc.

- Combinatorial optimization, genetic programing
- Active learning
- Representation

Conclusions

- Combinatorial optimization : black-box, expensive
- Multiobjective optimization
- Fitness landscapes for performance prediction
- Adaptive algorithm
- Surrogate model for combinatorial optimization