4. Fitness landscape analysis for multi-objective optimization problems

Fitness landscape analysis for understanding and designing local search heuristics

Sébastien Verel

LISIC - Université du Littoral Côte d'Opale, Calais, France
http://www-lisic.univ-littoral.fr/~verel/

The 51st CREST Open Workshop
Tutorial on Landscape Analysis
University College London

27th, February, 2017

Outline of this part

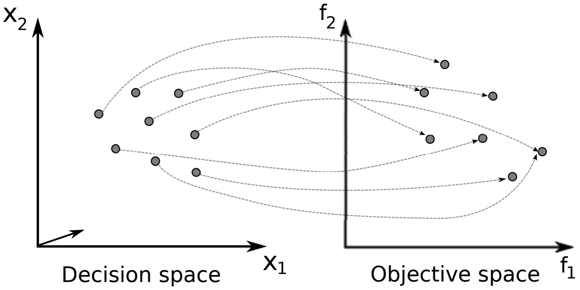
- Basis of fitness landscape *(Done)*
- Geometries of fitness landscapes *(Done)*
- Local optima network : *(Done)*
  - Definition inspired by complex systems science *(Done)*
  - Features of the network, design and performance *(Done)*
  - Performance prediction and portfolio *(Done)*
- **Fitness landscape analysis**
  for multi-objective optimization problems
  - Brief introduction on MO and MO algorithms
  - Fitness landscapes features
  - Performance prediction based on MO features
  - Set-based fitness landscape

Multiobjective optimization

Multiobjective optimization problem

- Search space : *decision space* X
- Objective function : *objective space* $f : X \to \mathbb{R}^m$

component f_i : objective, criterion.

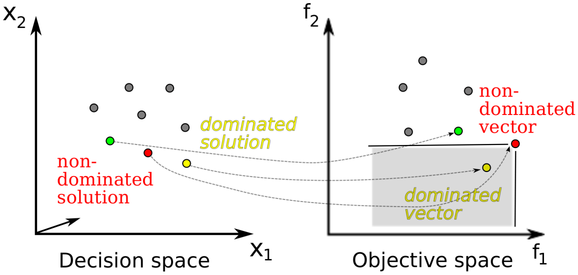


Definition

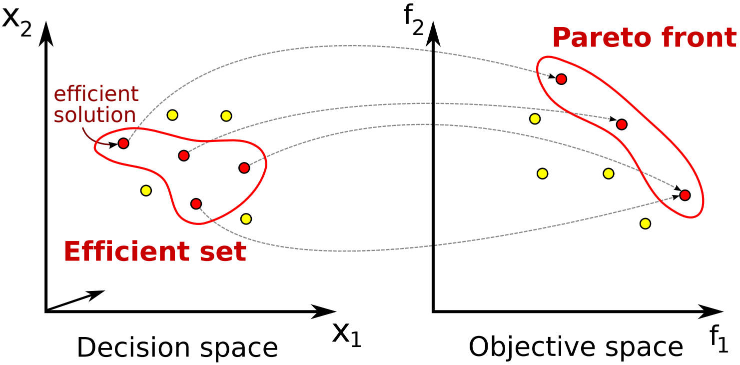
Pareto dominance relation

A solution $x \in X$ dominates a solution $x' \in X$ $(x' \prec x)$ 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 of multi-objective optimization

Find the whole **Pareto Optimal Set**,

or a good approximation of the Pareto Optimal Set

Why multiobjective optimization ?

Position of multiobjective optimization : Decision making

- No *a priori* on the importance of the different objectives
- *a posteriori* selection by a decision marker :
  Selection of one Pareto optimal solution after a deep study of the possible solutions.

Multi-objective, many-objective optimization

Approximative definition

- Multi-objective : 2, 3 or 4 objectives
- Many-objective : 4, 5 and more objectives

Number of Pareto optimal solutions

Suppose that :

- Probability to improve : p for all objective,
- Objective are independent.

Probability to be non-dominated for m objectives is :

Multi-objective, many-objective optimization

Approximative definition

- Multi-objective : 2, 3 or 4 objectives
- Many-objective : 4, 5 and more objectives

Number of Pareto optimal solutions

Suppose that :

- Probability to improve : p for all objective,
- Objective are independent.

Probability to be non-dominated for m objectives is : $1 - p^m$

Multi-objective, many-objective optimization

Approximative definition

- Multi-objective : 2, 3 or 4 objectives
- Many-objective : 4, 5 and more objectives

Number of Pareto optimal solutions

Suppose that :

- Probability to improve : p for all objective,
- Objective are independent.

Probability to be non-dominated for m objectives is : $1 - p^m$

Intuitive goals

Convergence toward the front, and diversity of the solutions.
Many-objective : convergence "easy", diversity "hard"

Note : objective correlation is also important.

Performance assessment

Quality of the approximation of the Pareto front [FKTZ05]

- Goal : indicator function related to the quality of the set.
- No universal indicator

Indicator functions

- Hypervolume indicator
- Epsilon indicator
- Atteinment function :

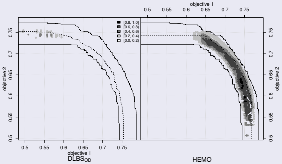
Performance assessment

Quality of the approximation of the Pareto front [FKTZ05]

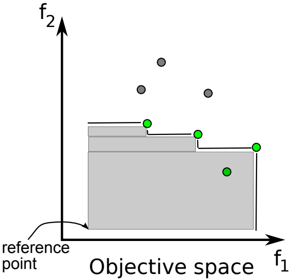
- Goal : indicator function related to the quality of the set.
- No universal indicator

Indicator functions

- Hypervolume indicator
- Epsilon indicator
- Atteinment function :
  Probability to reach a point in objective space with algo.



Hypervolume ($\mathtt{I}_H$)



Volume cover
by the set.

Properties

- Compliant with the (weak) Pareto dominance relation
  $\rightarrow A \prec B \Rightarrow \mathtt{I}_H(A) \leqslant \mathtt{I}_H(B)$
- A single parameter :
  the reference point
- Minimal solution-set maximizing $\mathtt{I}_H$
  $\rightarrow$ subset of the Pareto optimal set

$$\arg\max_{\sigma \in \Sigma} \ \mathtt{I}_H(\sigma)$$

Epsilon indicator (I_{ε})

Definition

Smallest coefficient ε to translate the set A to "cover" each point of the set B

$\forall z_b \in B,\ \exists z_a \in A$ such that $z_b \prec (1+\varepsilon)\ .\ z_a$

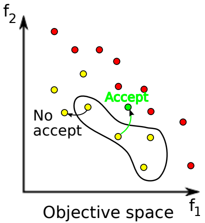
Properties

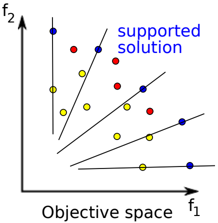
- Compliant with the (weak) Pareto dominance relation (using Pareto front)
  $\rightarrow A \prec B \Rightarrow \mathrm{I}_{\varepsilon}(A) \leqslant \mathrm{I}_{\varepsilon}(B)$
- A parameter : the reference set

Main types of MO local search algorithms

Three main classes :

- Pareto-based approaches : directly or indirectly focus the search on the Pareto dominance relation.
  Pareto Local Search (PLS), Global SEMO, NSGA-II, etc.
- Indicator approaches : Progressively improvement the indicator function : IBEA, SMS-MOEA, etc.
- Scalar approaches : multiple scalarized aggregations of the objective functions : MOEA/D, etc.





Pareto-based approaches

Simple dominance-based EMO with unbounded archive

local search :
Pareto Local Search (PLS)

Pick a random solution $x_0 \in X$
$A \leftarrow \{x_0\}$
repeat
  Select a non-visited $x \in A$
  Create $N(x)$ by flipping each bit of x in turns
  Flag x as visited
  $A \leftarrow$ non-dom. from $A \cup N(x)$
until all-visited $\vee$ maxeval

[Paquete *et al.* 2004][PCS04]

global search :
Global-Simple EMO (G-SEMO)

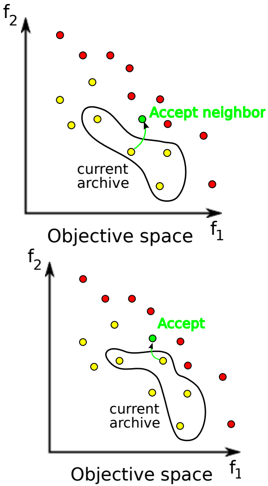
Pick a random solution $x_0 \in X$
$A \leftarrow \{x_0\}$
repeat
  Select $x \in A$ at random
  Create x' by flipping each bit of x with a rate $1/N$

  $A \leftarrow$ non-dom. from $A \cup \{x'\}$
until maxeval

[Laumanns *et al.* 2004][LTZ04]

A Pareto-based approach : Pareto Local Search

- Archive solutions using **Dominance relation**
- Iteratively improve this archive by exploring the neighborhood



SMS-MOEA : $\mathcal{S}$ metric selection- EMOA

[Beume *et al.* 2007][BNE07]

$P \leftarrow$ initialization()
repeat
  $q \leftarrow$ Generate(P)
  $P \leftarrow$ Reduce($P \cup \{q\}$)
until maxeval

Reduction

- Remove the worst solution according to non-dominated sorting, and $\mathcal{S}$ metric

Algorithm 2. Reduce(Q)

1: $\{\mathcal{R}_1, \ldots, \mathcal{R}_v\} \leftarrow$ fast-nondominated-sort(Q) /* all v fronts of Q */
2: $r \leftarrow \text{argmin}_{s \in \mathcal{R}_v}[\Delta_{\mathcal{S}}(s, \mathcal{R}_v)]$ /* $s \in \mathcal{R}_v$ with lowest $\Delta_{\mathcal{S}}(s, \mathcal{R}_v)$ */
3: **return** $(Q \backslash \{r\})$ /* eliminate detected element */

IBEA : Indicator-Based Evolutionary algorithm

[Zitzler *et al.* 2004][ZK04]

$P \leftarrow$ initialization()
repeat
  $P^{'} \leftarrow$ selection(P)
  $Q \leftarrow$ random_variation($P^{'}$)
  Evaluation of Q
  $P \leftarrow$ replacement(P, Q)
until maxeval

Fitness assignment

- Pairwise comparison of solutions in a population w.r.t. indicator I
- Fitness value : "loss in quality" in the population P if x was removed

$$f(x) = \sum_{x^{'} \in P \setminus \{x\}} (-e^{-I(x^{'},x)/\kappa})$$

- Often the ε-indicator is used

Original MOEA/D [ZL07] (minimization)

/ μ sub-problems defined by μ directions */*
$(\lambda^1, \ldots, \lambda^\mu) \leftarrow$ initialization_direction()
Initialize $\forall i = 1..\mu$ $B(i)$ the neighboring sub-problems of sub-problem i
/ one solution for each sub-problem */*
$(x^1, \ldots, x^\mu) \leftarrow$ initialization_solution()
repeat
  for $i = 1..\mu$ **do**
    Select x and x' randomly in $\{x_j \ : \ j \in B(i)\}$
    $y \leftarrow$ mutation_crossover(x, x')
    for $j \in B(i)$ **do**
      if $g(y|\lambda_j, z_j^\star) < g(x_j|\lambda_j, z_j^\star)$ **then**
        $x_j \leftarrow y$
      end if
    end for
  end for
until max_eval

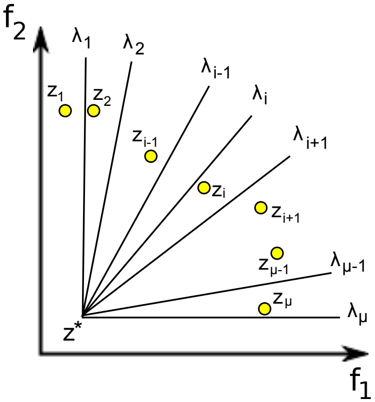
$B(i)$ is the set of the T closest neighboring sub-problems of sub-problem i
$g(\ |\lambda_i, z_i^\star)$: scalar function of sub-pb. i with λ_i direction, and $z_i^\star$ reference point

MOEA/D steady-state variant

Another MOEA/D (minimization)

/* μ *sub-problems defined by* μ *directions* */
$(\lambda^1, \ldots, \lambda^\mu) \leftarrow$ initialization_direction()
Initialize $\forall i = 1..\mu$ $B(i)$ the neighboring sub-problems of sub-problem i
/* *one solution for each sub-problem* */
$(x^1, \ldots, x^\mu) \leftarrow$ initialization_solution()
repeat
  Select i at random $\in 1..\mu$
  Select x randomly in $\{x_j \ : \ j \in B(i)\}$
  $y \leftarrow$ mutation_crossover(x_i, x)
  for n_r sub-problems $\in B(i)$ **do**
    if $g(y|\lambda_j, z_j^\star) < g(x_j|\lambda_j, z_j^\star)$ **then**
      $x_j \leftarrow y$
    end if
  end for
until max_eval

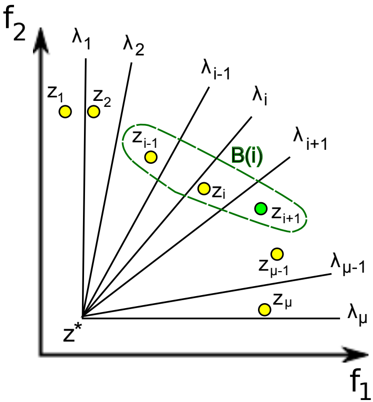
Representation of steady-state MOEA/D



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 $\sharp B(i) = T = 3$

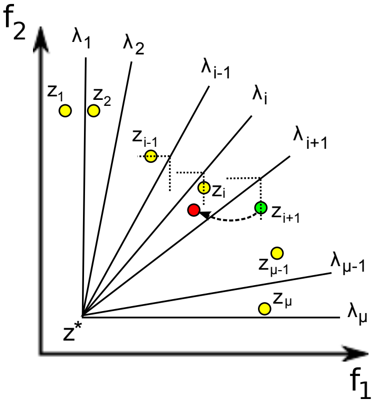
Representation of steady-state MOEA/D



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 $\sharp B(i) = T = 3$

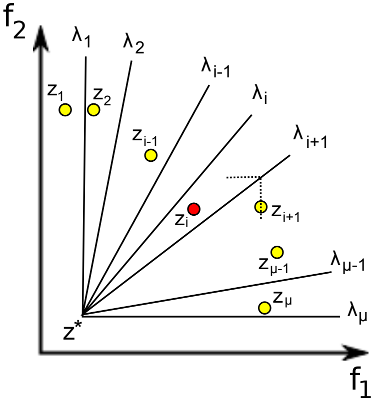
Representation of steady-state MOEA/D



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 $\sharp B(i) = T = 3$

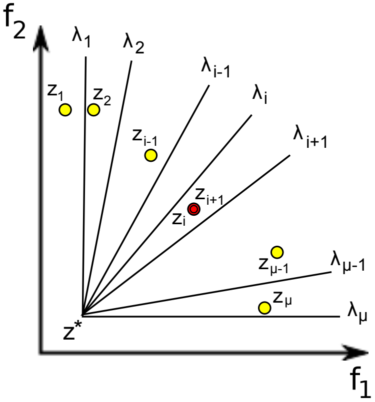
Representation of steady-state MOEA/D



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 $\sharp B(i) = T = 3$

Representation of steady-state 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 $\sharp B(i) = T = 3$

Join work with

Arnaud Liefooghe, Fabio Daolio, Hernan Aguirre, Kiyoshi Tanaka, Clarisse Daehnens, Laetitia Jourdan, Manuel López-Ibáñez, Mathieu Basseur, Adrien Goëffon

Slides are shared with those co-authors, special thanks to Arnaud and Fabio

Fitness landscape for Multi-Objective optimization

Context

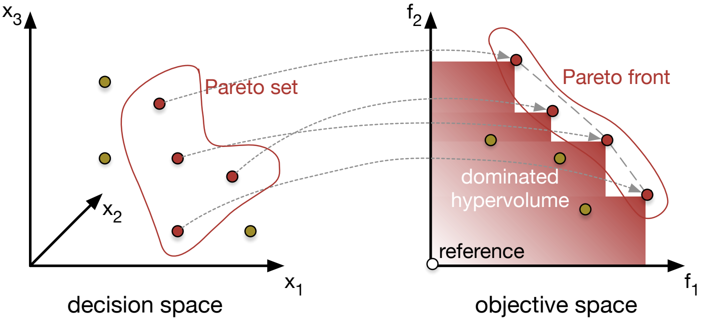
- MO problems are hard :
  underlying single-objective functions, join functions, etc.
- Learning the problem structure :
  to understand and improve algorithms
- Questions :
  - What are the relevant problem features ?
  - Replace one-dimensional by d-dimensional objective function ?

Problem features from the Pareto set

$\#po$ proportion of solutions that belong to the Pareto set

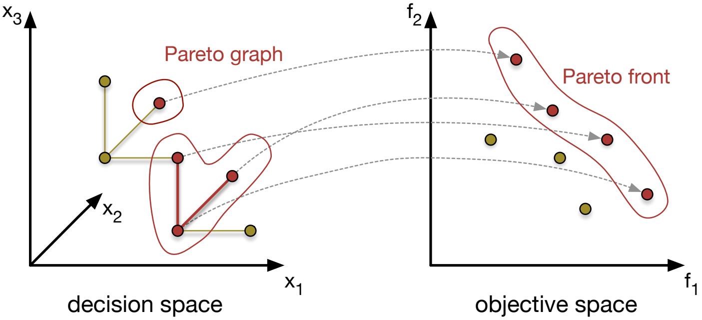
$\#supp$ proportion of supported solutions in the Pareto set

hv hypervolume-value of the Pareto front $\left(z^{\text{ref}} = (0, \ldots, 0)\right)$



Problem features from the Pareto set topology

dist_avg average distance between solutions from the Pareto set
dist_max maximum distance between solutions from the Pareto set
#cc relative number of connected components in the Pareto set
#lcc proportional size of the largest connected component



Problem features related to multi-modality and ruggedness

> Features sampled along adaptive walks :

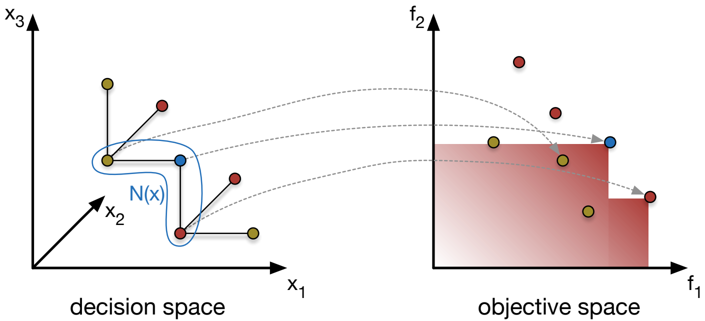
#lnd_avg_aws proportion of locally non-dominated solutions

ladapt length of a Pareto-based adaptive walk

> Features sampled along random walks :

hv_r1_rws autocorrelation coeff. of solution hypervolume

nhv_r1_rws autocorrelation coeff. of neighborhood hypervolume



Summary of problem features

	BENCHMARK **parameters**	
n	number of (binary) variables	
k_n	proportional number of variable interactions (epistatic links) : k/n	
m	number of objectives	
ρ	correlation between the objective values	
	FEATURES FROM **enumeration**	
#po	proportion of Pareto optimal (PO) solutions	*knowles2003*
#supp	proportion of supported solutions in the Pareto set	*knowles2003*
hv	hypervolume-value of the (exact) Pareto front	*aguirre2007*
#plo	proportion of Pareto local optimal (PLO) solutions	*paquete2007*
#slo_avg	average proportion of local optimal solutions per objective	
podist_avg	average Hamming distance between Pareto optimal solutions	*liefooghe2013*
podist_max	maximal Hamming distance between Pareto optimal solutions (diameter of the Pareto set)	*knowles2003*
po_ent	entropy of binary variables from Pareto optimal solutions	*knowles2003*
fdc	fitness-distance correlation in the Pareto set (Hamming dist. in solution space *vs.* Manhattan dist. in objective space)	*knowles2003*
#cc	proportion of connected components in the Pareto graph	*paquete2009*
#sing	proportion of isolated Pareto optimal solutions (singletons) in the Pareto graph	*paquete2009*
#lcc	proportional size of the largest connected component in the Pareto graph	*verel2011a*
lcc_dist	average Hamming distance between solutions from the largest connected component	
lcc_hv	proportion of hypervolume covered by the largest connected component	
#fronts	proportion of non-dominated fronts	*aguirre2007*
front_ent	entropy of the non-dominated front's size distribution	

Summary of problem features (cont'd)

	FEATURES FROM RANDOM WALK **sampling**	
hv_avg_rws	average (single) solution's hypervolume-value	
hv_r1_rws	first autocorrelation coefficient of (single) solution's hypervolume-values	*liefooghe2013*
hvd_avg_rws	average (single) solution's hypervolume difference-value	
hvd_r1_rws	first autocorrelation coefficient of (single) solution's hypervolume difference-values	*liefooghe2013*
nhv_avg_rws	average neighborhood's hypervolume-value	
nhv_r1_rws	first autocorrelation coefficient of neighborhood's hypervolume-value	
#lnd_avg_rws	average proportion of locally non-dominated solutions in the neighborhood	
#lnd_r1_rws	first autocorrelation coefficient of the proportion of locally non-dominated solutions in the neighborhood	
#lsupp_avg_rws	average proportion of supported locally non-dominated solutions in the neighborhood	
#lsupp_r1_rws	first autocorrelation coefficient of the proportion of supported locally non-dominated solutions in the neighborhood	
#inf_avg_rws	average proportion of neighbors dominated by the current solution	
#inf_r1_rws	first autocorrelation coefficient of the proportion of neighbors dominated by the current solution	
#sup_avg_rws	average proportion of neighbors dominating the current solution	
#sup_r1_rws	first autocorrelation coefficient of the proportion of neighbors dominating the current solution	
#inc_avg_rws	average proportion of neighbors incomparable to the current solution	
#inc_r1_rws	first autocorrelation coefficient of the proportion of neighbors incomparable to the current solution	
f_cor_rws	estimated correlation between the objective values	
	FEATURES FROM ADAPTIVE WALK **sampling**	
hv_avg_aws	average (single) solution's hypervolume-value	
hvd_avg_aws	average (single) solution's hypervolume difference-value	
nhv_avg_aws	average neighborhood's hypervolume-value	
#lnd_avg_aws	average proportion of locally non-dominated solutions in the neighborhood	
#lsupp_avg_aws	average proportion of supported locally non-dominated solutions in the neighborhood	
#inf_avg_aws	average proportion of neighbors dominated by the current solution	
#sup_avg_aws	average proportion of neighbors dominating the current solution	
#inc_avg_aws	average proportion of neighbors incomparable to the current solution	
length_aws	average length of Pareto-based adaptive walks	*verel2011a*

NK-landscapes [Kauffman 1993] [Kau93]

general-purpose family of multi-modal pseudo-boolean optimization functions
superposition of n Walsh functions of order k+1

$$\begin{array}{lll} \max & f(x) = \frac{1}{n}\sum_{j=1}^{n} c_j(x_j, x_{j_1}, \dots, x_{j_k}) & \\ \text{s.t.} & x_j \in \{0,1\} & , \quad j \in \{1,\dots,n\} \end{array}$$

> problem size n (decision space dimension)

> problem non-linearity $k <$ n
(multi-modality, epistatic interactions)

$x_1x_2x_4$	c_1
000	0.9
001	0.6
010	0.1
011	0.2
...	...

$x_1x_2x_3$	c_2
000	0.4
001	0.8
010	0.3
011	0.2
...	...

$x_2x_3x_4$	c_3
000	0.2
...	...
101	0.9
110	0.1
111	0.5

$x_1x_2x_4$	c_4
000	0.1
001	0.2
010	0.8
011	0.0
...	...

MNK-landscapes [Aguirre and Tanaka 2007][AT07]

$$\begin{array}{lll} \max & f_i(x) = \frac{1}{n}\sum_{j=1}^{n} c_j^i(x_j, x_{j_1}, \ldots, x_{j_k}) & , \quad i \in \{1, \ldots, m\} \\ \text{s.t.} & x_j \in \{0,1\} & , \quad j \in \{1, \ldots, n\} \end{array}$$

- m **independent** objective functions
- different variable interactions for each objective function

$x_1x_2x_4$	c_1^1
000	0.9
001	0.6
010	0.1
011	0.2
...	...

$x_1x_2x_3$	c_2^1
000	0.4
001	0.8
010	0.3
011	0.2
...	...

$x_2x_3x_4$	c_3^1
000	0.2
...	...
101	0.9
110	0.1
111	0.5

$x_1x_2x_4$	c_4^1
000	0.1
001	0.2
010	0.8
011	0.0
...	...

$x_1x_2x_4$	c_1^2
000	0.1
001	0.2
010	0.7
011	0.3
...	...

$x_1x_2x_3$	c_2^2
000	0.2
001	0.6
010	0.4
011	0.2
...	...

$x_2x_3x_4$	c_3^2
000	0.6
...	...
101	0.1
110	0.3
111	0.9

$x_1x_2x_4$	c_4^2
000	0.9
001	0.2
010	0.6
011	0.1
...	...

ρMNK-landscapes [Verel et al. 2011] [VLJD11]

$$\begin{array}{lll} \max & f_i(x) = \frac{1}{n}\sum_{j=1}^{n} c_j^i(x_j, x_{j_1}, \ldots, x_{j_k}) & , \quad i \in \{1, \ldots, m\} \\ \text{s.t.} & x_j \in \{0, 1\} & , \quad j \in \{1, \ldots, n\} \end{array}$$

- c_j^i's follow a **multivariate uniform law** of correlation ρ
- same variable interactions for each objective function

$x_1x_2x_4$	c_1^1
000	0.9
001	0.6
010	0.1
011	0.2
...	...

$x_1x_2x_3$	c_2^1
000	0.4
001	*0.8*
010	0.3
011	0.2
...	...

$x_2x_3x_4$	c_3^1
000	0.2
...	...
101	0.9
110	0.1
111	0.5

$x_1x_2x_4$	c_4^1
000	0.1
001	0.2
010	0.8
011	0.0
...	...

$x_1x_2x_4$	c_1^2
000	0.8
001	0.7
010	0.2
011	0.3
...	...

$x_1x_2x_3$	c_2^2
000	0.6
001	*0.7*
010	0.1
011	0.2
...	...

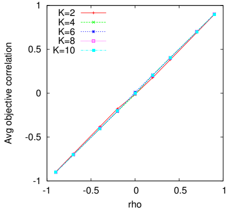
$x_2x_3x_4$	c_3^2
000	0.5
...	...
101	0.6
110	0.2
111	0.3

$x_1x_2x_4$	c_4^2
000	0.4
001	0.3
010	0.7
011	0.1
...	...

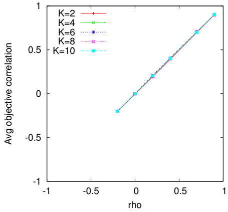
ρMNK-landscapes

$$
\begin{array}{lll}
\max & f_i(x) = \frac{1}{n} \sum_{j=1}^{n} c_j^i(x_j, x_{j_1}, \ldots, x_{j_k}) & , \quad i \in \{1, \ldots, m\} \\
\text{s.t.} & x_j \in \{0, 1\} & , \quad j \in \{1, \ldots, n\}
\end{array}
$$

m = 2



m = 5

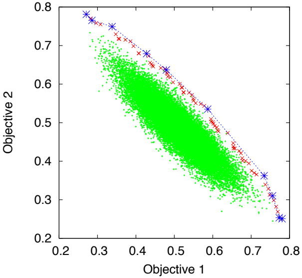


Benchmark parameters :

- problem size n (decision space dimension)
- problem non-linearity $k <$ n (multi-modality, epistatic interactions)
- number of objective functions m (objective space dimension)
- objective correlation $\rho > -\frac{1}{m-1}$

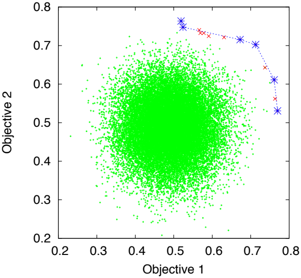
`http://mocobench.sf.net`

Some intuitions on objective correlation ρ



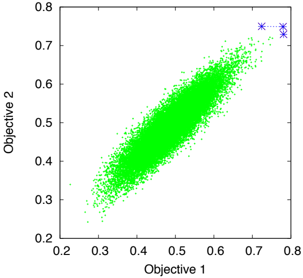
conflicting objectives

$\rho = -0.9$



independent objectives

$\rho = 0.0$



correlated objectives

$\rho = 0.9$

$\boxed{m = 2 \qquad n = 18 \qquad k = 4}$

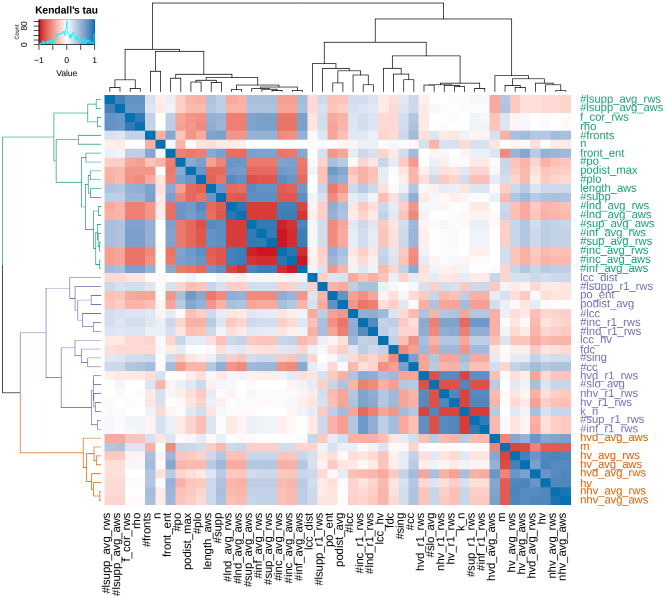
Experimental setup for enumerable instances

Small-size ρ*MNK-landscapes*, factorial design, 30 instances/unit

- problem size $n \in \{10, 11, 12, 13, 14, 15, 16\}$
- problem non-linearity $k \in \{0, 1, 2, 3, 4, 5, 6, 7, 8\}$
- number of objectives $m \in \{2, 3, 4, 5\}$
- obj. corr. $\rho \in \{-0.8, -0.6, -0.4, -0.2, 0, 0.2, 0.4, 0.6, 0.8, 1\}$, $\rho > \frac{-1}{m-1}$

60 480 problem instances overall

Pairwise feature association (enumerable instances)



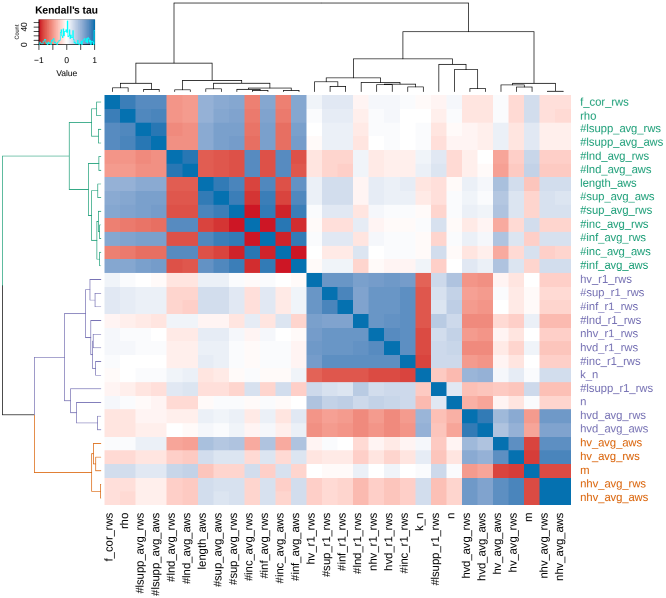
Experimental setup for large-size instances

Large-size ρ*MNK-landscapes*, constrained random LHS DOE

- problem size $n \in \{64, 256\}$
- problem non-linearity $k \in \{0, 8\}$
- number of objectives $m \in \{2, 5\}$
- objective correlation $\rho \in [-1, 1]$, $\rho > \frac{-1}{m-1}$

1 000 problem instances overall

Pairwise feature association (large instances)



Experimental setup for large-size instances

Large-size *ρMNK-landscapes*, constrained random LHS DOE

- problem size $n \in \{64, 256\}$
- problem non-linearity $k \in \{0, 8\}$
- number of objectives $m \in \{2, 5\}$
- objective correlation $\rho \in [-1, 1]$, $\rho > \frac{-1}{m-1}$

1 000 problem instances overall

GSEMO and IPLS *algorithms*

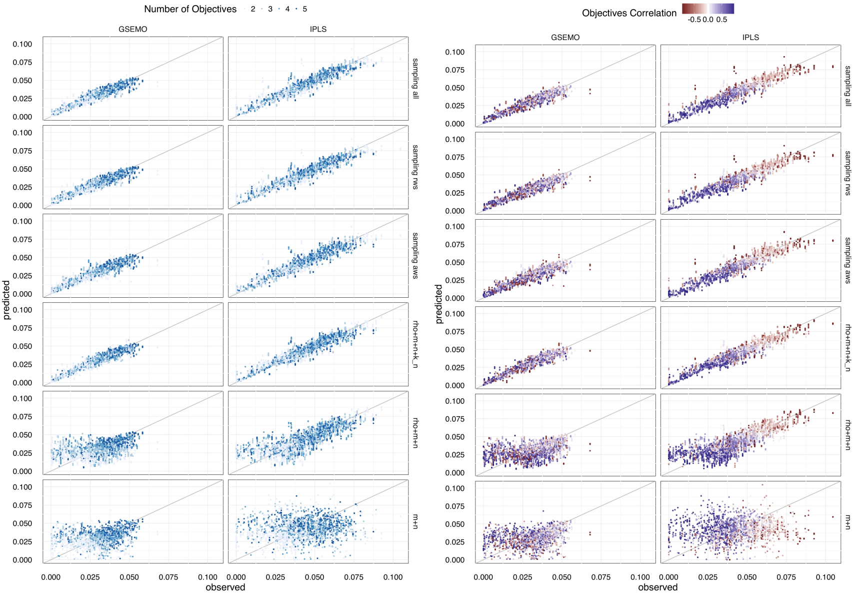
- 30 independent runs per instance
- Fixed budget of 100 000 evaluation calls
- epsilon approximation ratio to best-found non-dominated set

Prediction accuracy (large instances)

cross validation with repeated subsampling, 50 iterations, 90/10 split

feature set	MAE	MSE	R^2	rank
		GSEMO		
all	**0.003049**	**0.000017**	**0.891227**	1
sampling all	0.003152	**0.000018**	**0.883909**	1.3
sampling rws	0.003220	0.000019	0.878212	2
sampling aws	0.003525	0.000023	0.854199	3
ρ+m+n+k/n	**0.003084**	**0.000017**	**0.892947**	1
ρ+m+n	0.009062	0.000148	0.065258	4
m+n	0.010813	0.000206	-0.303336	5
		IPLS		
all	**0.004290**	**0.000034**	**0.886568**	1
sampling all	**0.004359**	**0.000035**	**0.883323**	1
sampling rws	**0.004449**	0.000036	**0.879936**	1.3
sampling aws	0.004663	0.000039	0.871011	2
ρ+m+n+k/n	**0.004353**	**0.000033**	**0.889872**	1
ρ+m+n	0.008415	0.000119	0.600965	3
m+n	0.016959	0.000472	-0.568495	4

Predicted vs observed values (out-of-folds)



Portfolio accuracy

cross validation with repeated subsampling, 50 iterations, 90/10 split

Portfolio : { GSEMO, IPLS }

feature set	error rate	rank
all	**0.0128**	1
sampling all	**0.0138**	1
sampling rws	**0.0150**	1
sampling aws	**0.0144**	1
ρ+m+n+k/n	**0.0134**	1
ρ+m+n	0.0824	2
m+n	0.1328	3
const=GSEMO	0.0880	
const=IPLS	0.7250	

Practice : Bi-objective squares problem

Squares Problem (SP)

Find the position of 5 squares
in order to maximize inside squares the number of brown points
and the number of blue points

Fitness function

$f(x) =$ (number of brown points, number of blue points)

Fitness landscape analysis

Execute line by line the `main` function of the code `ex05.R` :
random walk and Pareto adaptive walk

Set-based multiobjective fitness landscapes [VLD11]

The key idea

EMO algorithms are local search algorithms performing on sets

Set-based multiobjective fitness landscape

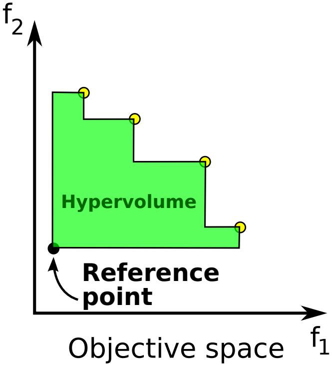
$$(\Sigma, \mathtt{N}, \mathtt{I})$$

- **Set-domain search space** (Σ)
  $\Sigma \subset 2^X$ is a set of feasible solution-sets
  (where X is the set of feasible solutions)
- **Set-domain neighborhood relation** $(\mathtt{N})$
  $\mathtt{N} : \Sigma \to 2^\Sigma$ is a neighborhood relation between solution-sets
- **Set-domain fitness function** $(\mathtt{I})$
  $\mathtt{I} : \Sigma \to \mathbb{R}$ is a unary quality indicator,
  i.e. a fitness function measuring the quality of solution-sets

Set-domain fitness function

Quality indicator

$\mathrm{I}(\sigma) \in \mathbb{R}$: quality of a set of solutions σ



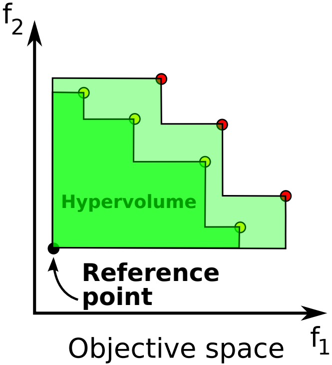
Hypervolume (I_H)

- Compliant with the (weak) Pareto dominance relation :
  $$\sigma \prec \sigma' \Rightarrow \mathrm{I}_H(\sigma) \leqslant \mathrm{I}_H(\sigma')$$
- A single parameter :
  the reference point
- Minimal set maximizing I_H :
  subset of the Pareto optimal set

Set-domain fitness function

Quality indicator

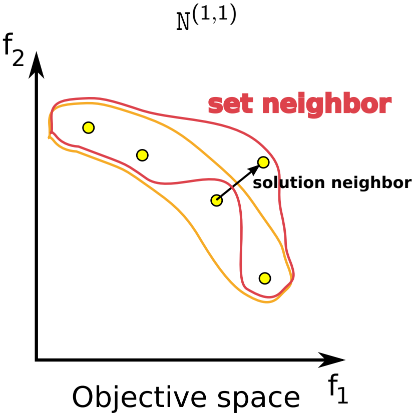
$\mathrm{I}(\sigma) \in \mathbb{R}$: quality of a set of solutions σ

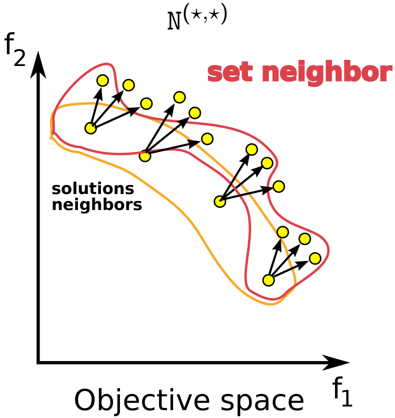


Hypervolume (I_H)

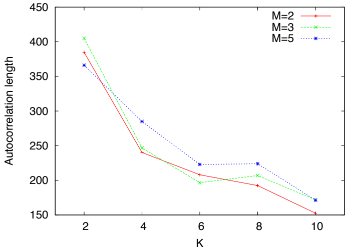
- Compliant with the (weak) Pareto dominance relation :
  $\sigma \prec \sigma' \Rightarrow \mathrm{I}_H(\sigma) \leqslant \mathrm{I}_H(\sigma')$
- A single parameter :
  the reference point
- Minimal set maximizing I_H :
  subset of the Pareto optimal set

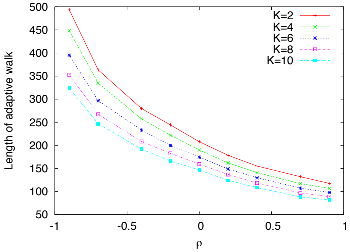
Set-domain neighborhood relations [BGLV13]





Ruggedness and multimodality





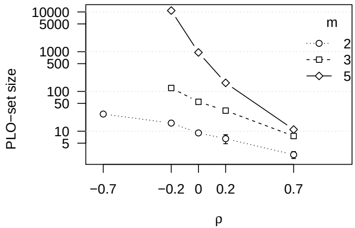
Autocorrelation length (τ)

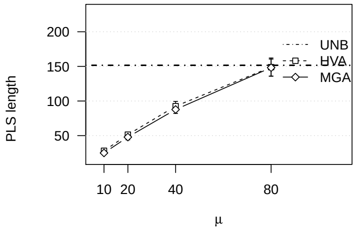
- non-linearity degree K
  $\rightarrow$ τ decreases wrt. K
- number of objectives M
  $\rightarrow$ no influence
- objective correlation ρ
  $\rightarrow$ no influence

Adaptive walk length ($\mathcal{L}$)

- non-linearity degree K
  $\rightarrow$ $\mathcal{L}$ decreases wrt. K
- number of objectives M
  $\rightarrow$ no clear trend
- objective correlation ρ
  $\rightarrow$ $\mathcal{L}$ decreases wrt. ρ ! !

Pareto Local Optima set *vs.* archive size





Pareto set-size

for unbounded solution-set (archive)

- objective correlation ρ
  $\rightarrow$ decreases wrt. ρ
- number of objectives M
  $\rightarrow$ increases wrt. M

Pareto Local search Length

for bounded solution-set (archive)

- Increases with archive size

Fitness landscapes analysis for MO optimization

Summary

- A number of features are shown to be related to problem difficulty
- Better knowledge on problem difficulty for MO optimization
- Performance can be predicted using such relevant features (for some algo/problems)
- And now,
  - toward a better design ?
  - toward a algorithm portfolio ?
  - toward more theoretical works ?
  - design low complexity features ?...

References I

H. E. Aguirre and Kiyoshi Tanaka.
Working principles, behavior, and performance of MOEAs on MNK-landscapes.
European Journal of Operational Research, 181(3) :1670–1690, 2007.

Matthieu Basseur, Adrien Goëffon, Arnaud Liefooghe, and Sébastien Verel.
On set-based local search for multiobjective combinatorial optimization.
In *Genetic and Evolutionary Computation Conference (GECCO 2013)*, pages 471–478, Amsterdam, Netherlands, 2013.

References II

Nicola Beume, Boris Naujoks, and Michael Emmerich.
Sms-emoa : Multiobjective selection based on dominated hypervolume.
European Journal of Operational Research, 181(3) :1653–1669, 2007.

Carlos M Fonseca, Joshua D Knowles, Lothar Thiele, and Eckart Zitzler.
A tutorial on the performance assessment of stochastic multiobjective optimizers.
In *Third International Conference on Evolutionary Multi-Criterion Optimization (EMO 2005)*, volume 216, page 240, 2005.

References III

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

Marco Laumanns, Lothar Thiele, and Eckart Zitzler.
Running time analysis of evolutionary algorithms on a simplified multiobjective knapsack problem.
Nat Comput, 3(1) :37–51, 2004.

L. Paquete, M. Chiarandini, and T. Stützle.
Pareto local optimum sets in the biobjective traveling salesman problem : An experimental study.
In *Metaheuristics for Multiobjective Optimisation*, volume 535 of *Lecture Notes in Economics and Mathematical Systems*, chapter 7, pages 177–199. Springer, 2004.

References IV

Sébastien Verel, Arnaud Liefooghe, and Clarisse Dhaenens.
Set-based Multiobjective Fitness Landscapes : A Preliminary Study.
In *Genetic And Evolutionary Computation Conference*, pages 769–776, Dublin, Ireland, June 2011. ACM.

Sébastien Verel, Arnaud Liefooghe, Laetitia Jourdan, and Clarisse Dhaenens.
Analyzing the Effect of Objective Correlation on the Efficient Set of MNK-Landscapes.
In C.A. Coello Coello, editor, *Learning and Intelligent OptimizatioN Conference (LION 5)*, volume 6683/2011 of *Lecture Notes in Computer Science (LNCS)*, pages 116–130, Rome, Italy, January 2011. Springer.

References V

Eckart Zitzler and Simon Künzli.
Indicator-based selection in multiobjective search.
In *Parallel Problem Solving from Nature-PPSN VIII*, pages 832–842. Springer, 2004.

Qingfu Zhang and Hui Li.
Moea/d : A multiobjective evolutionary algorithm based on decomposition.
Evolutionary Computation, IEEE Transactions on, 11(6) :712–731, 2007.