Landscape analysis for explainable optimization

Arnaud Liefooghe
Univ. Littoral Côte d'Opale – LISIC
Calais, France

arnaud.liefooghe@univ-littoral.fr

Sébastien Verel
Univ. Littoral Côte d'Opale – LISIC
Calais, France

verel@univ-littoral.fr

WCCI/CEC conference, Yokohama, June 31th, 2024.

Sources

Final version can be found :

`https://www-lisic.univ-littoral.fr/~verel/`

Program for today

1. The Basics of Fitness Landscapes
2. Geometries of Fitness Landscapes
3. Local Optima Network
4. Multi-objective Fitness Landscapes

1. The Basics of Fitness Landscapes

Outline

Single-objective optimization

- Search space : set of candidate solutions

$$X$$

- Objective fonction : quality criteria (or non-quality)

$$f : X \to \mathbb{R}$$

X discrete : combinatorial optimization
$X \subset \mathbb{R}^n$: numerical optimization

Solve an optimization problem (maximization)

$$X^\star = \mathrm{argmax}_X \ f$$

or find an approximation of $X^\star$.

Context : black-box optimization

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

No information on the objective function definition f

Objective fonction :

- can be irregular, non continuous, non differentiable ...
- given by a computation or a simulation

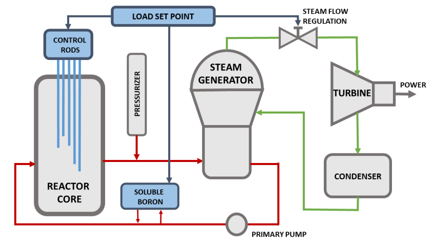
Real-world black-box optimization : an example

PhD of M. Muniglia / V. Drouet / B. Gasse, Saclay Nuclear Research Centre (CEA), Paris

$x \longrightarrow$

$\longrightarrow f(x)$

$(73, \ldots, 8) \longrightarrow$



$\longrightarrow \Delta_z P$

Multi-physic simulator

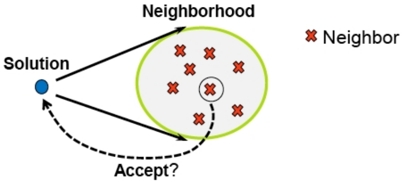
Search algorithms

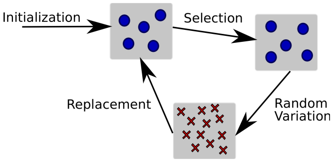
Principle

(implicite) **enumeration of a subset of the search space**

- Many ways to enumerate the search space
  - Exact methods : $A^\star$, Branch&Bound . . .
  - Random sampling : Monte Carlo, approximation with guarantee, bayesian optimization, . . .

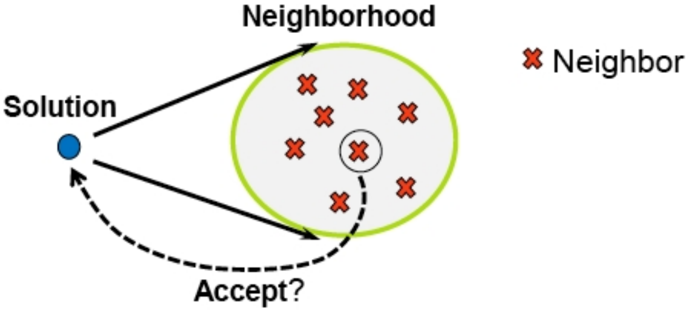
Local search / Evolutionary algorithms





Stochastic algorithms with a single solution (Local Search)

- X set of candidate solutions (the search space)
- $f : X \to \mathbb{R}$ objective function
- $\mathcal{N}(x)$ set of neighboring solutions from x



So, we need a tool to study this...

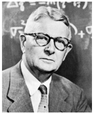
Motivations on fitness landscape analysis

For the search to be efficient, the sequence of local optimization problems must be related to the global problem

Main motivation : "Why using local search"

- Study the search space from the point of view of local search
  $\Rightarrow$ Fitness Landscape Analysis
- To understand and design effective local search algorithms

Fitness landscape : original plots from S. Wright [Wri32]



S. Wright. "The roles of mutation, inbreeding, crossbreeding, and selection in evolution.", 1932.

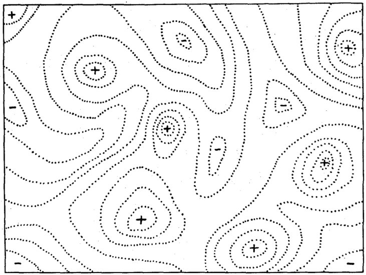


FIGURE 2.—Diagrammatic representation of the field of gene combinations in two dimensions instead of many thousands. Dotted lines represent contours with respect to adaptiveness.

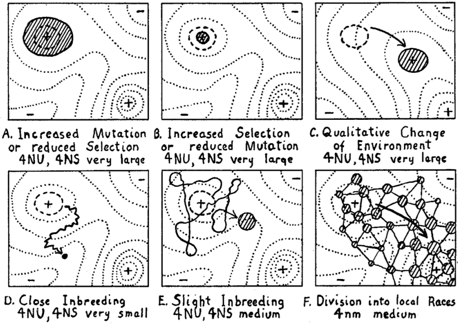
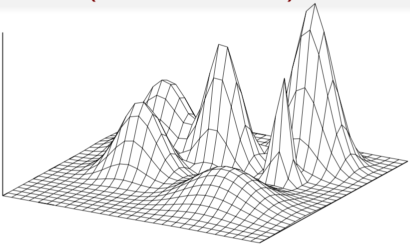


FIGURE 4.—Field of gene combinations occupied by a population within the general field of possible combinations. Type of history under specified conditions indicated by relation to initial field (heavy broken contour) and arrow.

source : Encyclopaedia Britannica Online

Fitness landscapes in (evolutionary) biology



- Metaphorical uphill struggle across a "fitness landscape"
  - mountain **peaks** represent high "fitness" (ability to survive/reproduce)
  - **valleys** represent low fitness
- Evolution proceeds :
  population of organisms
  performs an "**adaptive walk**"

be careful : "2 dimensions instead of many thousands"

Fitness landscapes as Complex System tool

Dynamical system

Predict, and understand the evolutionary paths

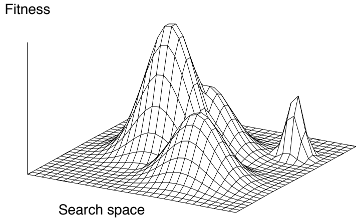
$$X \longrightarrow X$$

- Quasispecies equation : mean field analysis

  x_t
- Stochastic process : Markov chain

  $\Pr(x_{t+1} \mid x_t)$
- Individual scale : network analysis

Fitness landscape for combinatorial optimization [Sta02]



Definition

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

- **search space** :

$$X$$

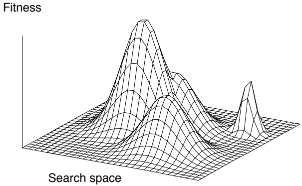
- **neighborhood relation** :

$$\mathcal{N} : X \to 2^X$$

- **objective function** :

$$f : X \to \mathbb{R}$$

What is a neighborhood ?



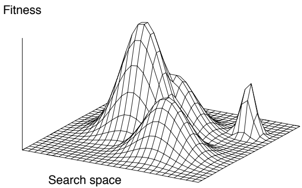
Neighborhood function :

$$\mathcal{N} : X \to 2^X$$

Set of "**neighbor**" solutions associated to each solution

$$\mathcal{N}(x) = \{y \in X \mid y = op(x)\}$$

What is a neighborhood ?



Neighborhood function :

$$\mathcal{N} : X \to 2^X$$

Set of "**neighbor**" solutions associated to each solution

$$\mathcal{N}(x) = \{y \in X \mid y = op(x)\}$$

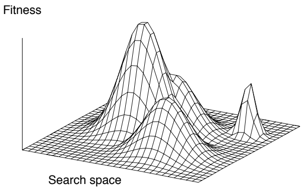
or

$$\mathcal{N}(x) = \{y \in X \mid \Pr(y = op(x)) > 0\}$$

or

$$\mathcal{N}(x) = \{y \in X \mid \Pr(y = op(x)) > \varepsilon\}$$

What is a neighborhood ?



Ordre

Neighborhood function :

$$\mathcal{N} : X \to 2^X$$

Set of "**neighbor**" solutions associated to each solution

$$\mathcal{N}(x) = \{y \in X \mid y = op(x)\}$$

or

$$\mathcal{N}(x) = \{y \in X \mid \Pr(y = op(x)) > 0\}$$

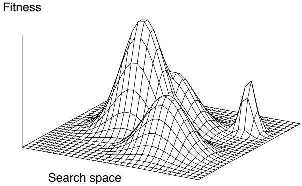
or

$$\mathcal{N}(x) = \{y \in X \mid \Pr(y = op(x)) > \varepsilon\}$$

or

$$\mathcal{N}(x) = \{y \in X \mid \text{distance}(x, y) = 1\}$$

What is a neighborhood ?



Neighborhood function :

$$\mathcal{N} : X \to 2^X$$

Set of "**neighbor**" solutions associated to each solution

Important !

Neighborhood must be based on the operator(s) used by the algorithm

Neighborhood $\Leftrightarrow$ Operator

Ordre

$$\mathcal{N}(x) = \{y \in X \mid y = op(x)\}$$

or

$$\mathcal{N}(x) = \{y \in X \mid \Pr(y = op(x)) > 0\}$$

or

$$\mathcal{N}(x) = \{y \in X \mid \Pr(y = op(x)) > \varepsilon\}$$

or

$$\mathcal{N}(x) = \{y \in X \mid \text{distance}(x, y) = 1\}$$

Typical example : bit strings

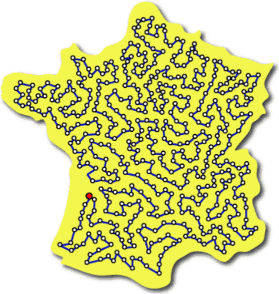
Search space : $X = \{0, 1\}^N$

$$\mathcal{N}(x) = \{y \in X \mid d_{Hamming}(x, y) = 1\}$$

Example :
$\mathcal{N}(01101) = \{11101, 00101, 01001, 01111, 01100\}$

Typical example : permutations

Traveling Salesman Problem :
find the shortest tour which cross one time every town



Search space : $X = \{\sigma \mid \sigma \text{ permutations }\}$

$$\mathcal{N}(x) = \{y \in X \mid y = op_{2opt}(x)\}$$

cf. exchange, insertion, etc.

More than 1 operator. . . ?

What can we do with 2 operators (ex : memetic algorithm) ?

$\mathcal{N}_1(x) = \{y \in X \mid y = op_1(x)\}$ $\mathcal{N}_2(x) = \{y \in X \mid y = op_2(x)\}$

More than 1 operator...?

What can we do with 2 operators (ex : memetic algorithm)?

$\mathcal{N}_1(x) = \{y \in X \mid y = op_1(x)\}$ $\mathcal{N}_2(x) = \{y \in X \mid y = op_2(x)\}$

Severals possibilities according to the goal :

- Study 2 landscapes : $(X, \mathcal{N}_1, f)$ and $(X, \mathcal{N}_2, f)$
- Study the landscape of "union" : $(X, \mathcal{N}, f)$

$$\mathcal{N} = \mathcal{N}_1 \cup \mathcal{N}_2 = \{y \in X \mid y = op_1(x) \text{ or } y = op_2(x)\}$$

- Study the landscape of "composition" : $(X, \mathcal{N}, f)$

$$\mathcal{N} = \{y \in X \mid y = op \circ op'(x) \text{ with } op, op' \in \{id, op_1, op_2\}\}$$

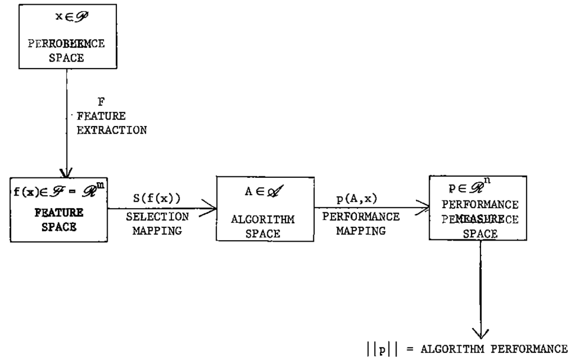
Main goals

- Engineering goal :
  How to analyze fitness landscape ?
  Predict performance, select algorithm / configuration, etc.

- Scientific goal :
  Why there is this search dynamic on the problem ?
  What are the properties of Fitness Landscape ?
  Understand relation between properties, and search dynamic

Rice's framework for algorithm selection [Ric76]

Algorithm selection



Rice, J. R. (1976). The algorithm selection problem. Advances in computers, 15, 65-118.

Fitness landscape analysis

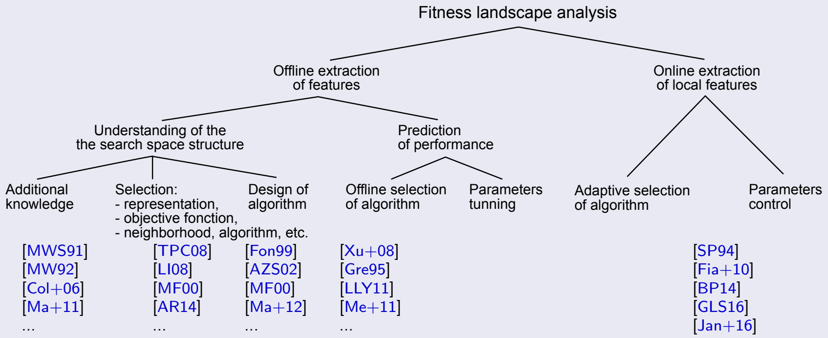
Algebraic approach, grey-box :

$$\Delta f = \lambda.(f - \bar{f})$$

Statistical approach, black-box :

Problems $\rightsquigarrow$ Features $\rightsquigarrow$ Algorithm $\rightsquigarrow$ Performances

Goals



J. J. Grefenstette, in FOGA 3, 1995.[Gre95]

"Predictive Models Using Fitness Distributions of Genetic Operators"

"An important goal of the theory of genetic algorithms is to build **predictive models** of how well genetic algorithms are expected to perform, given a representation, a **fitness landscape**, and a set of genetic operators. (...)"

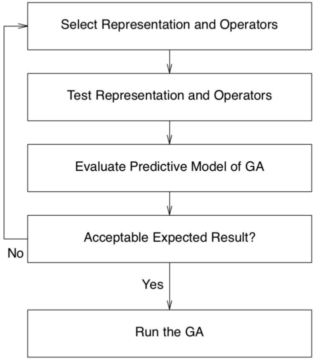
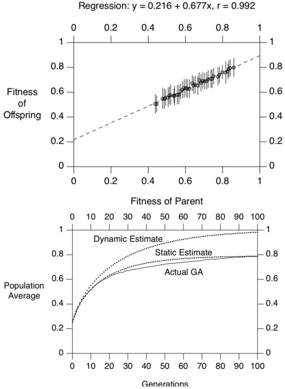


Figure 1: Predicting GA Performance

$FD_{op}(F_p) = Prob\,(F_c = \text{fitness of offspring} \mid \text{parents have mean fitness } F_p)$



Typical use cases of fitness landscapes analysis

1. **Comparing the difficulty of two landscapes** :
   - one problem, different encodings : $(X_1, \mathcal{N}_1, f_1)$ vs. $(X_2, \mathcal{N}_2, f_2)$ different representations, variation operators, objectives ...

     Which landscape is easier to solve ?
2. **Choosing one algorithm** :
   - analyzing the global geometry of the landscape

     Which algorithm shall I use ?
3. **Tuning the algorithm's parameters** :
   - *off-line* analysis of the fitness landscape structure

     What is the best mutation operator ? the size of the population ? the number of restarts ? ...
4. **Controlling the algorithm's parameters at runtime** :
   - *on-line* analysis of structure of fitness landscape

     What is the optimal mutation operator according to the current estimation of the structure ?

Beyong the use cases of fitness landscapes analysis : Why

1. **Comparing the difficulty of two landscapes** :
   - one problem, different encodings : $(X_1, \mathcal{N}_1, f_1)$ vs. $(X_2, \mathcal{N}_2, f_2)$ different representations, variation operators, objectives ...

   Which landscape is easier to solve ?

2. **Choosing one algorithm** :
   - analyzing the global geometry of the landscape

   Which algorithm shall I use ?

3. **Tuning the algorithm's parameters** :
   - *off-line* analysis of the fitness landscape structure

   What is the best mutation operator ? the size of the population ? the number of restarts ? ...

4. **Controlling the algorithm's parameters at runtime** :
   - *on-line* analysis of structure of fitness landscape

   What is the optimal mutation operator according to the current estimation of the structure ?

Short summary for this part

Studying the structure of the fitness landscape
allows to **understand**/**explain** the difficulty,
and to design better optimization algorithms

The fitness landscape is a **graph** $(X, \mathcal{N}, f)$:

- nodes are solutions and have a value (the fitness)
- edges are defined by the neighborhood relation

pictured as a real landscape

So next, what are the properties (features), how have been designed, what are their meanings ?

References I

John J Grefenstette.
Predictive models using fitness distributions of genetic operators.
In *Foundations of Genetic Algorithms*, volume 3, pages 139–161. Elsevier, 1995.

John R. Rice.
The algorithm selection problem.
Advances in Computers, 15 :65–118, 1976.

P. F. Stadler.
Fitness landscapes.
In M. Lässig and Valleriani, editors, *Biological Evolution and Statistical Physics*, volume 585 of *Lecture Notes Physics*, pages 187–207, Heidelberg, 2002. Springer-Verlag.

References II

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.

2. Geometries of Fitness Landscapes

Outline

Metrics, features of fitness landscape

Main idea

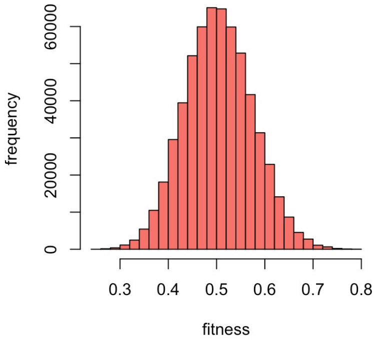
The "shape" of the neighborhood (local description)
is related to
the dynamics of the local search, and its performance

Main questions

- How to design relevant metrics ?
- What are the meaning of the metrics (benefits, and caveats) ?
- How to estimate the metrics ?

In the following, a comprehensive methodology of fitness analysis

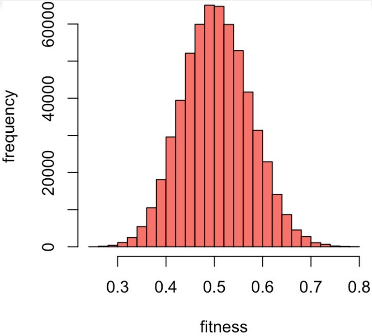
Fitness distribution : Density of states



density of fitness values across the search space

- Introduced in physics : Rosé 1996 [REA96]
- In optimization : Belaidouni, Hao 00 [BH00]

Fitness distribution : Density of states



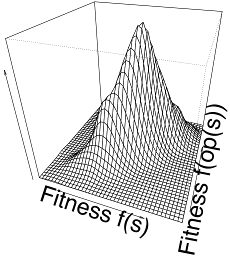
Interpretations :

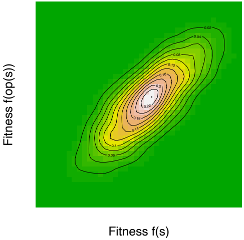
- Performance of random search
- The faster the decay, the harder the problem
- Not so far from a normal distribution (in practice, and theory)

Features : Average, sd, kurtosis, . . .

Estimation : Sample of random solutions (size $\approx 10^3$)

Fitness cloud [Verel et al. 2003]





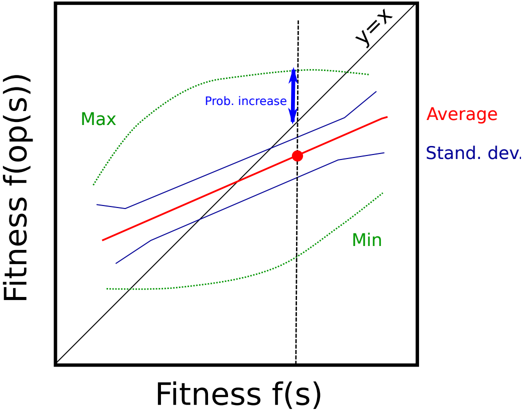
- $(X, \mathcal{F}, \Pr)$: probability space
- $op : X \to X$ stochastic operator of the local search

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

Fitness Cloud of *op*

Conditional probability density function of Y given X

Fitness cloud : a measure of evolvability



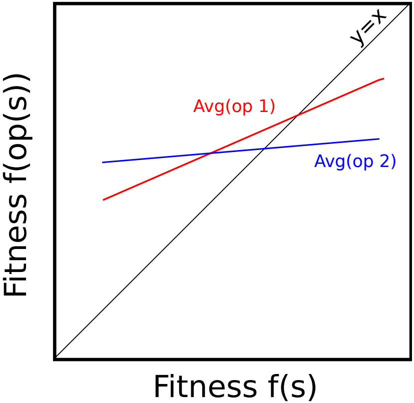
Evolvability

Ability to evolve : fitness in the neighborhood vs fitness of current solution

- Probability of finding better solutions
- Average fitness of better neighbors
- Average and standard dev. of fitness-values

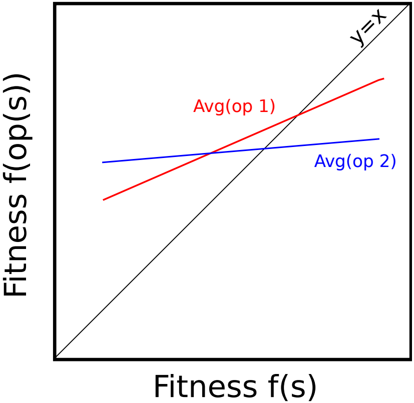
Fitness cloud : comparing difficulty

Average of evolvability



- Operator 1 ? ? Operator 2

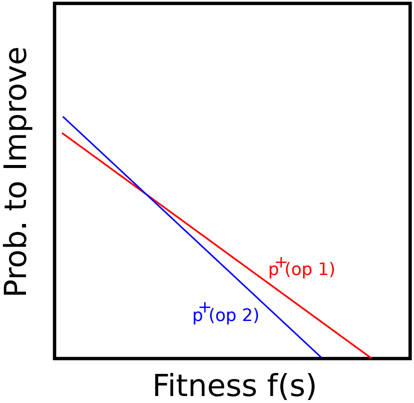
Fitness cloud : comparing difficulty
Average of evolvability



- Operator 1 > Operator 2
- Because Average 1 more correlated with fitness
- Linked to autocorrelation
- Average is often a line :
  - See works on Elementary Landscapes (Stadler, D. Wihtley, F. Chicano and others)
  - See the idea of Negative Slope Coefficient (NSC)

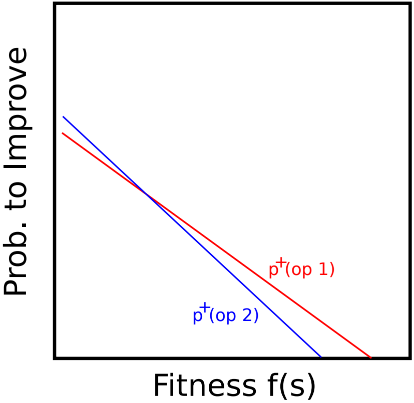
Fitness cloud : comparing difficulty

Probability to improve



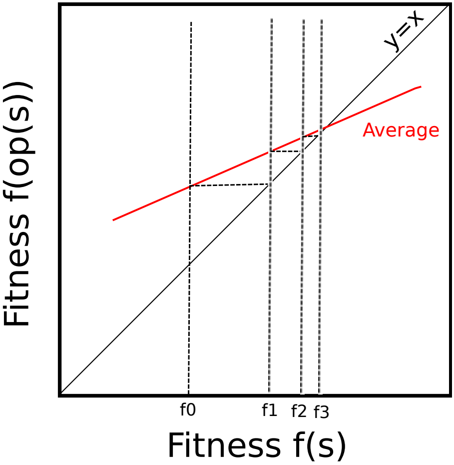
- Operator 1 ? ? Operator 2

Fitness cloud : comparing difficulty
Probability to improve



- Operator 1 > Operator 2
- Prob. to improve of Op 1 is often higher than Prob. to improve of Op 2
- Probability to improve is often a line
- See also works on fitness-probability cloud (G. Lu, J. Li, X. Yao [LLY11])
- See theory of EA and fitness level technics

Fitness cloud : estimating the convergence point



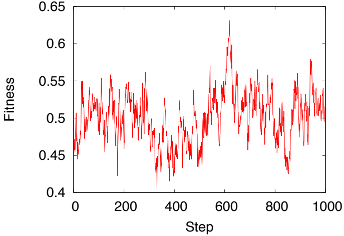
- Approximation (only approximation) of the fitness value after few steps of local operator
- Indication on the quality of the operator
- See fitness level technic

Random walk tools

Fitness cloud Estimator :

Random solutions, and one random neighbor

ex. sample size $\approx 2 \times 10^3$ (at least $2n\log(n)$ to sample all dim.)

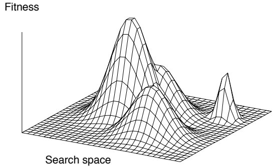


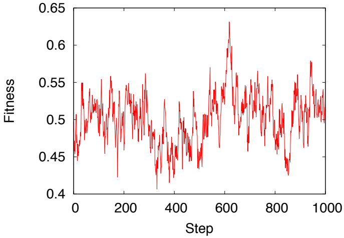
- Random walk :

  $(x_1, x_2, \ldots)$ where $x_{i+1} \in \mathcal{N}(x_i)$ and equiprobability on $\mathcal{N}(x_i)$

ex. sample size $\approx 10^3$ (at least $n\log(n)$ to sample all dim.)

Random walk to estimate ruggedness





Gives useful information on the profile of fitness landscape, and on local properties (neighborhood)

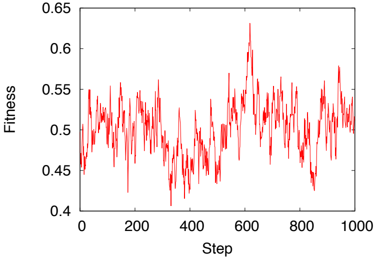
Interpretation :

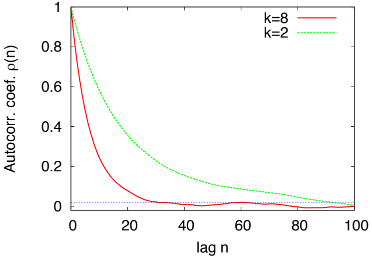
- if the profile of fitness is irregular,
- then the "information" between neighbors is low

Feature :

- Study the fitness profile as a signal

Rugged/smooth fitness landscapes





Autocorrelation function of the time series of fitness-values [Wei90] :

$$\rho(n) = \frac{\mathbb{E}[(f(x_i) - \bar{f})(f(x_{i+n}) - \bar{f})]}{\mathrm{Var}(f(x_i))}$$

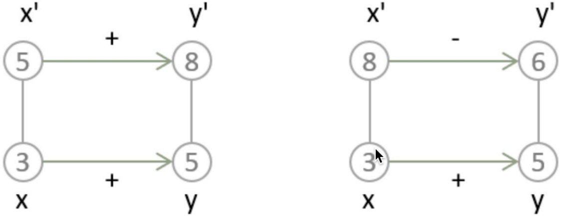
Autocorrelation length

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

"How many random steps such that correlation becomes insignificant"

Other correlation metrics are possible *e.g.* Kendall, entropy (see [])

Rugged/smooth fitness landscapes : sign epistasis



Degree of epistasis :

Ratio of "negative" square (*i.e.* Kendall correlation coeff.)

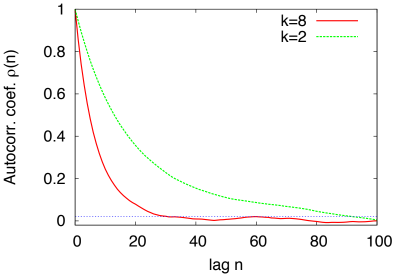
References :

Biology : Poelwijk *et al.* [PKWT07]

EA : Basseur *et al.* [BG15]

Estimator : sample size ≈ 2.10^3

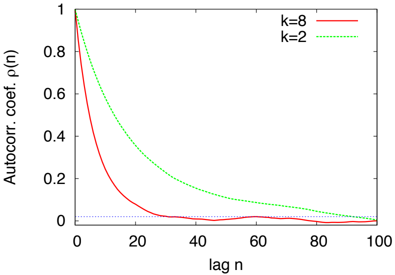
"Easy / Difficult" landscapes



Question

Which landscape is "easier" ? Green or red one ?

"Easy / Difficult" landscapes



Question

Which landscape is "easier" ? Green or red one ?

- small τ : rugged landscape, more difficult landscape
- long τ : smooth landscape, easier landscape

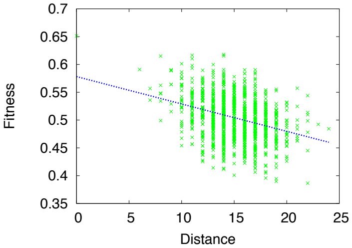
Theoretical results on autocorrelation (Stadler 96 [Sta96])

Ruggedness decreases with the size of those problems

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	$1 - \frac{K+1}{N}$
random max-k-SAT	n number of variables k variables per clause	$1 - \frac{k}{n(1-2^{-k})}$

Fitness distance correlation (FDC) (Jones 95 [Jon95])

Correlation between fitness and distance to global optimum



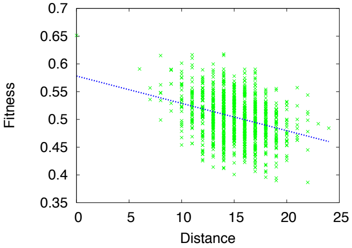
"easy" "hard"

Classification based on experimental studies

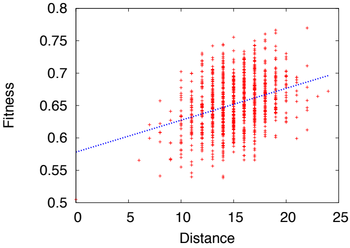
- $\rho < -0.15$: easy optimization
- $\rho > 0.15$: hard optimization
- $-0.15 < \rho < 0.15$: undecided zone

Fitness distance correlation (FDC) (Jones 95 [Jon95])

Correlation between fitness and distance to global optimum



"easy"

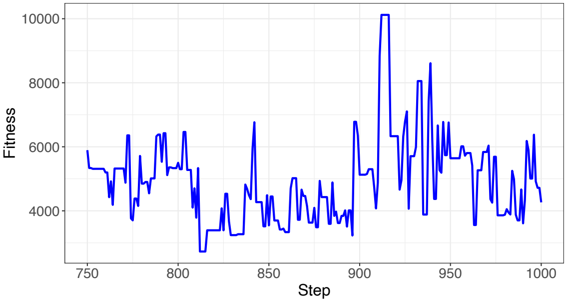


"hard"

- Important concept to understand search difficulty
- Not useful in "practice"
  (difficult to estimate, global opt. unknown)

Random walks on real world problems

Random walk on the problem of
"nuclear power plant design" [MVLPD17]



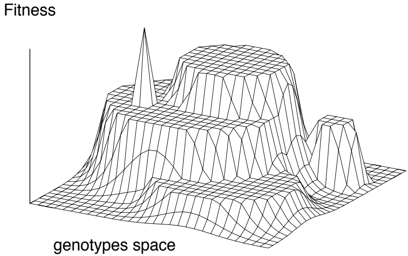
- Move/Mutation without fitness change (here ≈ 30%)
- Low impact of variable modification, "flat" shape

Neutral fitness landscapes

Neutral theory (Kimura ≈ 1960 [Kim83])

Theory of mutation and random drift

Many mutations have no effects on fitness-values



- plateaus
- neutral degree
- neutral networks [Schuster 1994 [SFSH94], RNA folding]

Neutral degree

Neutral neighborhood

Set of neighbors which have the same fitness value

$$\mathcal{N}_{neutral}(x) = \{x' \in \mathcal{N}(x) \mid f(x') = f(x)\}$$

Nota : $f(x') = f(x)$ can be replaced by $|f(x') - f(x)| \leq \varepsilon$.

Neutral degree

Number of neutral neighbors : $\sharp\mathcal{N}_{neutral}(x)$

Neutral rate

Relative number of neutral neighbors : $\frac{\sharp\mathcal{N}_{neutral}(x)}{\sharp\mathcal{N}(x)}$

Estimation of the neutral rate with random walk

- The neutral rate can be estimated with a random walk :

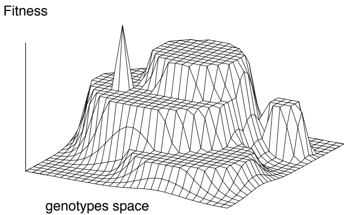
$$(x_1, x_2, \ldots, x_\ell) \text{ where } x_{t+1} \in \mathcal{N}(x_t)$$

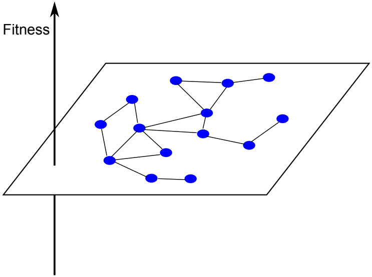
Neutral rate estimation [LDV+17]

$$\frac{\sharp\{(x_t, x_{t+1}) \ : \ f(x_t) = f(x_{t+1}), t \in \{1, \ell - 1\}\}}{\ell - 1}$$

Nota : With single random walk, fitness distribution, autocorrelation of fitness, probability of improvement, neutral rate can be estimated

Neutral networks (Schuster 1994 [SFSH94])



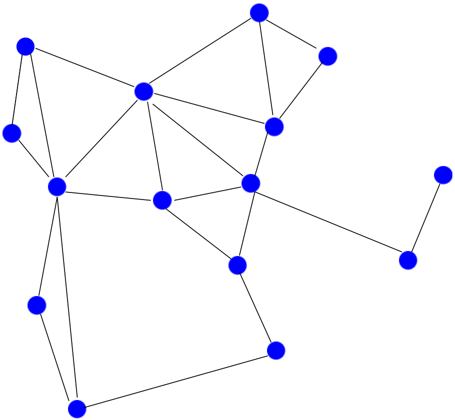


Basic definition of Neutral Network

Graph where :

- Node = solution with the same fitness-value
- Edge = neighborhood relation

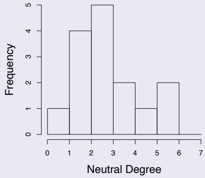
Features of neutral networks



1. **Size**

   avg, distribution . . .

2. **Neutral degree distribution**



3. **Autocorrelation of the neutral degree**
   - neutral random walk
   - autocorr. of degrees

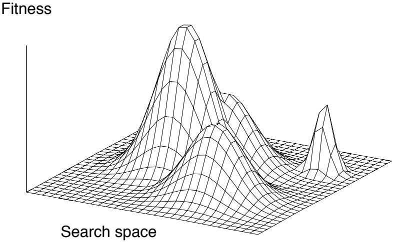
4. **Evolvability metrics,**

Multimodal fitness landscapes

Local optima $x^\star$

no neighboring solution with strictly better fitness value (maximization)

$$\forall x \in \mathcal{N}(x^\star), \quad f(x) \leqslant f(x^\star)$$



nota : If $\mathcal{N}$ is modified (distance, op), the local optima are modified

Typical example : bit strings

Search space : $X = \{0,1\}^N$

$$\mathcal{N}(x) = \{y \in X \mid d_{Hamming}(x,y) = 1\}$$

Example :
$x = 01101$ and $f_1(x) = f_2(x) = f_3(x) = 5$

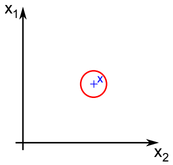
	11101	00101	01001	01111	01100
f_1	4	2	3	0	3
f_2	2	3	6	2	3
f_3	1	5	2	2	4

Question

Is x is a local maximum for f_1, f_2, and/or f_3 ?

Not so typical example : continuous optimization

Still an open question...



Search space : $X = [0,1]^d$

$$\mathcal{N}_\alpha(x) = \{y \in X \mid \|y - x\| \leqslant \alpha\}$$
with $\alpha > 0$

Classical definition of local optimum

x is local maximum iff

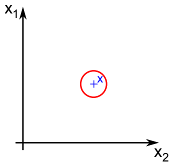
$\exists \varepsilon > 0, \forall y$ such that $\|y - x\| \leqslant \varepsilon,\ f(y) \leqslant f(x)$

Questions

Local search definition with $\mathcal{N}_\alpha \Rightarrow$ classical definition ?
Classical definition $\Rightarrow$ local search definition with $\mathcal{N}_\alpha$?

Not so typical example : continuous optimization

Still an open question...



Search space : $X = [0,1]^d$

$$\mathcal{N}_\alpha(x) = \{y \in X \mid \|y - x\| \leqslant \alpha\}$$

with $\alpha > 0$

Classical definition of local optimum

x is local maximum iff

$$\exists \varepsilon > 0, \forall y \text{ such that } \|y - x\| \leqslant \varepsilon,\ f(y) \leqslant f(x)$$

Questions

Local search definition with $\mathcal{N}_\alpha \Rightarrow$ classical definition ?

Classical definition $\Rightarrow$ local search definition with $\mathcal{N}_\alpha$?

Still some works to do...

Sampling local optima by adaptive walks

Adaptive walk

$$(x_1, x_2, \ldots, x_\ell) \text{ such that } x_{i+1} \in \mathcal{N}(x_i) \text{ and } f(x_i) < f(x_{i+1})$$

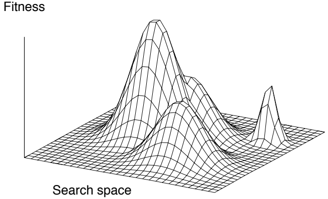
Hill-Climbing algorithm (first-improvement)

Choose initial solution $x \in X$
repeat
  choose $x' \in \{y \in \mathcal{N}(x) \mid f(y) > f(x)\}$
  if $f(x) < f(x')$ **then**
    $x \leftarrow x'$
  end if
until x is a Local Optimum

Basin of attraction of $x^\star$

$$\{x \in X \mid HillClimbing(x) = x^\star\}.$$

Multimodality and problem difficulty



The core idea :

- if the size of the basin of attraction of the global optimum is "small",
- then, the "time" to find the global optimum is "long"

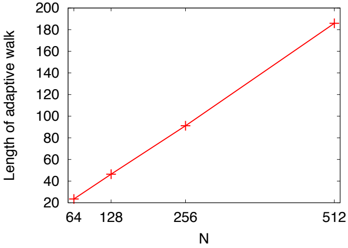
Optimization difficulty :
Number and size of the basins of attraction (Garnier *et al.* [GK02])

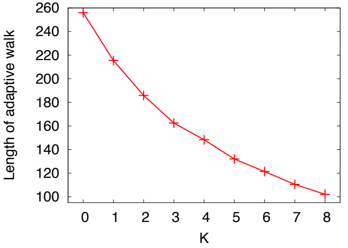
Feature to estimate the basins size :

- **Length of adaptive walks**

cost : sample size $\times \ell \times |\mathcal{N}|$

Multimodality and problem difficulty





ex. nk-landscapes with $n = 512$

The core idea :

- if the size of the basin of attraction of the global optimum is "small",
- then, the "time" to find the global optimum is "long"

Optimization difficulty :

Number and size of the basins of attraction (Garnier *et al.* [GK02])

Feature to estimate the basins size :

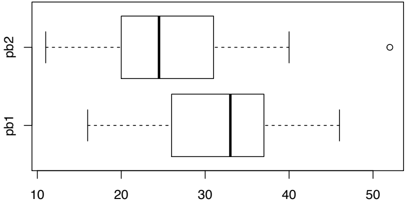
- **Length of adaptive walks**

cost : sample size $\times \ell \times |\mathcal{N}|$

Example

2 instances of the same problem :

same problem dimension, same neighborhood operator



Adaptive walks length distribution

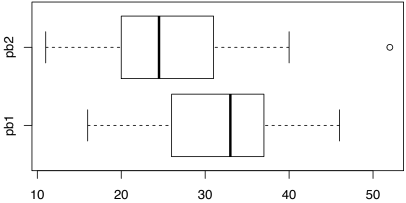
Question

Which one seems to be easier ?

Example

2 instances of the same problem :

same problem dimension, same neighborhood operator



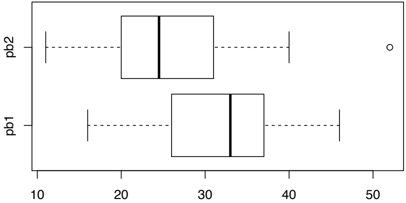
Adaptive walks length distribution

Question

Which one seems to be easier ? problem 2

Example

2 instances of the same problem :
same problem dimension, same neighborhood operator



Adaptive walks length distribution

Question

Which one seems to be easier ? problem 2

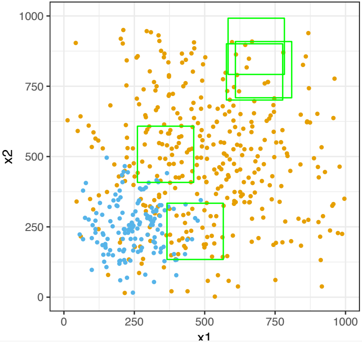
Indeed, basic hypothesis (but only hypothesis) :
$\sharp X = 2^d$, $\sharp \text{Basin} = 2^{\alpha\ell}$
Avg. number of local opt. : $\log(\sharp X / \sharp \text{Basin}) = (\alpha\ell - d) \log 2$

Practice : the Squares Problem

a program design problem ?

Squares Problem (SP)

Find the position of 5 squares in order to maximize inside squares the number of brown points without blue points



Candidate solutions

$$X = ([0,1000] \times [0,1000])^5$$

	x_1	x_2
1	577	701
2	609	709
3	366	134
4	261	408
5	583	792

Fitness function

$$f(x) = \text{number of brown points} - \text{number of blue points inside squares}$$

Practice : computing the autocorrelation function

Source code `exo02.R` :

- `mutation_create` :
  Create a mutation operator,
  modify each square according to rate p,
  a new random value from $[(x-r, y-r), (x+r, y+r)]$.
- `main` :
  Code to obtain autocorrelation function

Questions

- Define the function `random_walk` to compute the fitness values during a random walk
- Execute line by line the `main` function to compute a sample of fitness value collected during a random walk
- Compare the first autocorrelation coefficient of the SP problems 1 and 2

Source code in R : ex01.R

Source code : `https://www-lisic.univ-littoral.fr/~verel/`

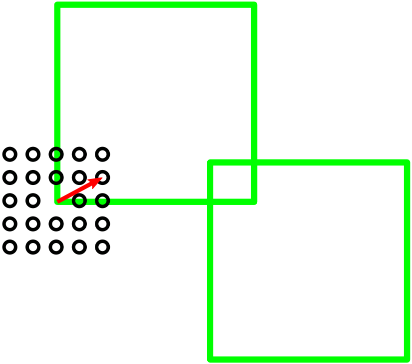
Different functions are already defined :

- `main` : example to execute the following functions
- `draw` and `draw_solution` :
  draw a problem and the squares of a solution
- `fitness_create` :
  create a fitness function from a data frame of points
- `pb1_create` and `pb2_create` :
  create two particular SP problems
- `init` :
  create a random solution with *n* squares
- `hc_ngh` :
  hill-climbing local search based on neighborhood

Neighborhood

Questions

- Execute line by line the `main` function
- Define the `neighborhood_create` which creates a neighborhood : a neighbor move one square



Adaptive walks to compare problem difficulty

Pre-defined functions :

- `adaptive_length` :
  run the hill-climber and compute a data frame with the length of adaptive walks
- `main_adaptive_length_analysis` :
  Compute the adaptive length of two different SP problems

Questions

- Execute line by line the `main_adaptive_length_analysis` function to compute a sample of adaptive walk lengths
- Compare the lengths of adaptive walks for the two SP problems
- Which one is more multimodal ?

Practice : computing the neutral rate

Source code `exo03.R` :

- `main` :
  Code to compute the neutral rates

Questions

- Define the function `neutral_rate` to compute the neutral rate estimated with a random walk
- Execute the `main` function to compute the neutral rate
- Compare the neutrality of the SP problems 1 and 2

Practice : Performance *vs.* fitness landscape features

Explain the performance of ILS with fitness landscape features ?

- 20 random SP problems have been generated : `pb_xx.csv`
- The performance of Iterated Local Search has been computed in `perf_ils_xx.csv` (30 runs)
- Goal : regression of ILS performance with fitness landscape features

Practice : Performance *vs.* fitness landscape features

Source code `exo04.R` :

- `fitness_landscape_features` :
  Compute the basic fitness landscape features
- `random_walk_samplings` :
  Random walk sampling on each problem (save into file)
- `fitness_landscape_analysis` :
  Compute the features for each problems
- `ils_performance` :
  Add the performance of ILS into the data frame
- `main` :
  Execute the previous functions

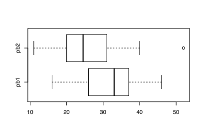
Practice : Performance *vs.* fitness landscape features

Questions

- What are the features computed by the function `fitness_landscape_features` ?
- Execute the `random_walk_samplings` function to compute the random walk samples
- Compute the correlation plots between features and ILS performance (use ggpairs)
- Compute the linear regression of performance with fitness landscape features

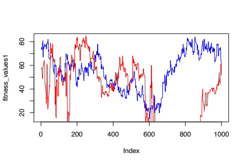
Practice : example of results

Adaptive walk lengths



pb1 is "easier" than pb2

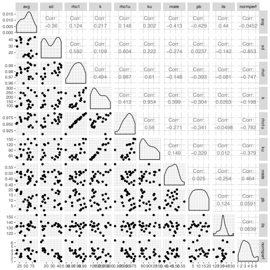
Random walks



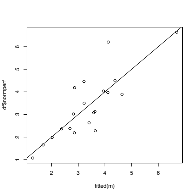
pb1 : $\rho(1) = 0.9856$, $nr = 0.513$

pb2 : $\rho(1) = 0.9872$, $nr = 0.498$

Correlation between features



ILS perf. prediction (lin. mod.)



$R^2 = 0.69$

Short summary

- Geometries :
  Multimodality, ruggedness, neutrality

- Metrics/features based on the neighborhood :
  probability to improve, fitness distribution, sign, etc.
- Covariance of the metrics across search space :
  autocorrelation, pearson/spearman/kendall correlation, entropy, etc.

- Estimation of metrics/features :
  random sampling, random walk, adaptive walk, etc.
  sample size, length, number : use sampling methodology

References I

Khulood Alyahya and Jonathan E Rowe.
Simple random sampling estimation of the number of local optima.
In International Conference on Parallel Problem Solving from Nature, pages 932–941. Springer, 2016.

Matthieu Basseur and Adrien Goëffon.
Climbing combinatorial fitness landscapes.
Applied Soft Computing, 30 :688–704, 2015.

References II

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.

Josselin Garnier and Leila Kallel.
Efficiency of local search with multiple local optima.
SIAM Journal on Discrete Mathematics, 15(1) :122–141, 2002.

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

References III

M. Kimura.
The Neutral Theory of Molecular Evolution.
Cambridge University Press, Cambridge, UK, 1983.

Arnaud Liefooghe, Bilel Derbel, Sébastien Verel, Hernán Aguirre, and Kiyoshi Tanaka.
Towards Landscape-Aware Automatic Algorithm Configuration : Preliminary Experiments on Neutral and Rugged Landscapes, pages 215–232.
Springer International Publishing, Cham, 2017.

Guanzhou Lu, Jinlong Li, and Xin Yao.
Fitness-probability cloud and a measure of problem hardness for evolutionary algorithms.
In European Conference on Evolutionary Computation in Combinatorial Optimization, pages 108–117. Springer, 2011.

References IV

Mathieu Muniglia, Sébastien Verel, Jean-Charles Le Pallec, and Jean-Michel Do.
A Fitness Landscape View on the Tuning of an Asynchronous Master-Worker EA for Nuclear Reactor Design.
In
International Conference on Artificial Evolution (Evolution Artificielle
pages 30–46, Paris, France, October 2017.

Frank J Poelwijk, Daniel J Kiviet, Daniel M Weinreich, and Sander J Tans.
Empirical fitness landscapes reveal accessible evolutionary paths.
Nature, 445(7126) :383–386, 2007.

References V

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.

References VI

E. D. Weinberger.
Correlated and uncorrelatated fitness landscapes and how to tell the difference.
In Biological Cybernetics, pages 63 :325–336, 1990.

3. Local Optima Network

Outline

1. ~~The Basics of Fitness Landscapes~~
2. ~~Geometries of Fitness Landscapes~~
3. Local Optima Network
   - Features from the network, algorithm design and performance
   - Performance prediction and algorithm portfolio
4. Multi-objective Fitness Landscapes

Joint initial work with

- Gabriela Ochoa, University of StirlingUK
- Marco Tomassini, University of Lausanne, Switzerland
- Fabio Daolio, University of Stirling, UK

Key idea : complex system tools

Principle of variable aggregation

A model for dynamical systems with two scales (time/space)

- Split the state space according to the different scales
- Study the system at the large scale

Key idea : complex system tools

Principle of variable aggregation

A model for dynamical systems with two scales (time/space)

- Split the state space according to the different scales
- Study the system at the large scale

$$X \xrightarrow{op} X$$

Variable aggregation for fitness landscape

- At solutions level (small scale) :
  - Stochastic local search operator
  - Exponential number of solutions
  - Exponential size of the stochastic matrix of the process (Markov chain)
- Projection on a relevant space :
  - Reduce the size of state space
  - Potentially loose some information
  - Relevant information remains when $p(op(x)) \approx op'(p(x))$

Key idea : complex system tools

Principle of variable aggregation

A model for dynamical systems with two scales (time/space)

- Split the state space according to the different scales
- Study the system at the large scale

$$\begin{array}{ccc} X & \xrightarrow{op} & X \\ p \downarrow & & \downarrow p \\ E & \xrightarrow{op'} & E \end{array}$$

Variable aggregation for fitness landscape

- At solutions level (small scale) :
  - Stochastic local search operator
  - Exponential number of solutions
  - Exponential size of the stochastic matrix of the process (Markov chain)
- Projection on a relevant space :
  - Reduce the size of state space
  - Potentially loose some information
  - Relevant information remains when

$$p(op(x)) \approx op'(p(x))$$

Key idea : complex system tools

Complex network

Bring the tools from complex networks analysis to study the structure of combinatorial fitness landscapes

Methodology

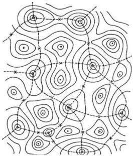
- **Design a network** that represents the landscape
  - Nodes : local optima
  - Edges : a notion of adjacency between local optima
- **Extract features** :
  - "complex" network analysis
- **Use the network features** :
  - search algorithm design, difficulty ...

J. P. K. Doye, The network topology of a potential energy landscape : a static scale-free network., Phys. Rev. Lett., 88 :238701, 2002. [Doy02]

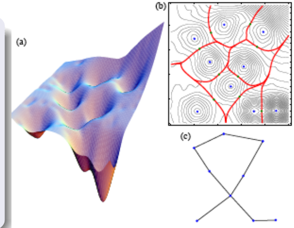
Energy surface and inherent networks

Inherent network

- **Nodes** : energy minima
- **Edges** : two nodes are connected if the energy barrier separating them is sufficiently low (transition state)



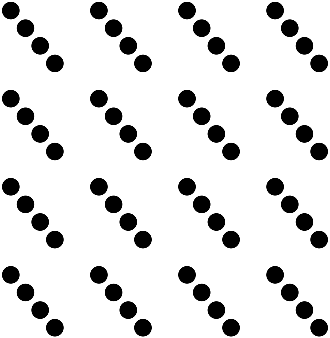
(a) Energy surface
(b) Contours plot : partition of states space into basins of attraction
(c) Landscape as a network



F. H Stillinger, T. A Weber. Packing structures and transitions in liquids and solids. Science, 225.4666 , p. 983-9, 1984. [SW84]
J. P. K. Doye, The network topology of a potential energy landscape : a static scale-free network. Phys. Rev. Lett., 88 :238701, 2002. [Doy02]

Basins of attraction in combinatorial optimization

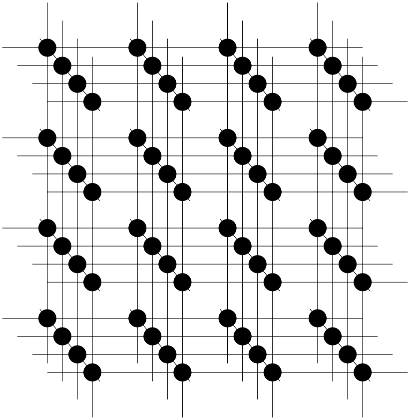
Example of a small NK landscape with $N = 6$ and $K = 2$



- Bit strings of length $N = 6$
- $2^6 = 64$ solutions
- one point = one solution

Basins of attraction in combinatorial optimization

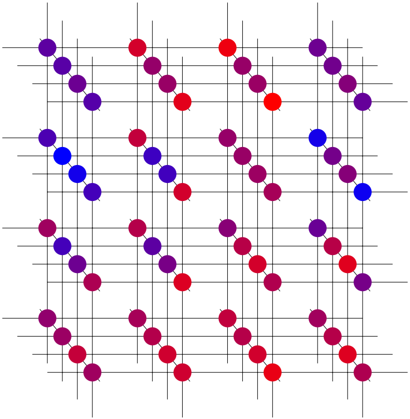
Example of a small NK landscape with $N = 6$ and $K = 2$



- Bit strings of length $N = 6$
- Neighborhood size = 6
- Line between points = solutions are neighbors
- Hamming distances between solutions are preserved (except for at the border of the cube)

Basins of attraction in combinatorial optimization

Example of small NK landscape with $N = 6$ and $K = 2$

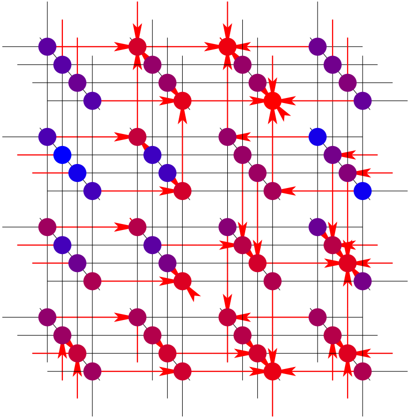


The color represents the fitness-values

- high fitness
- low fitness

Basins of attraction in combinatorial optimization

Example of small NK landscape with $N = 6$ and $K = 2$



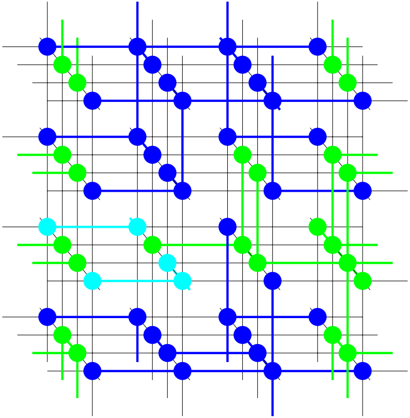
- Color represent fitness value
  - high fitness
  - low fitness
- → point towards the solution with highest fitness in the neighborhood

Exercise :

Why not making a Hill-Climbing walk on it ?

Basins of attraction in combinatorial optimization

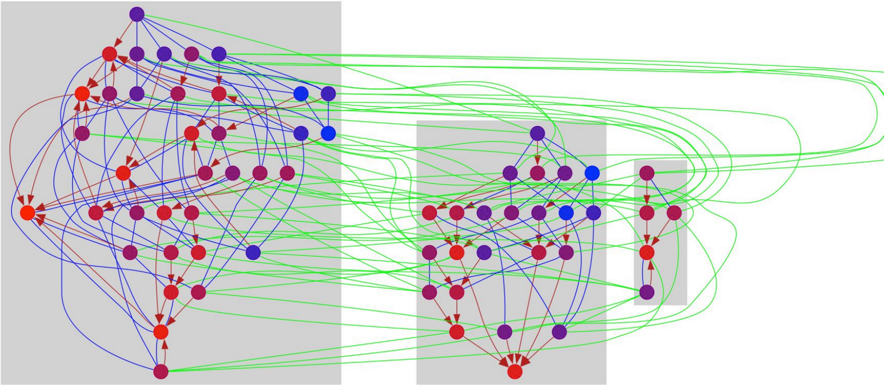
Example of small NK landscape with $N = 6$ and $K = 2$



- Each color corresponds to one basin of attraction
- Basins of attraction are interlinked and overlapped
- Basins have no "interior"

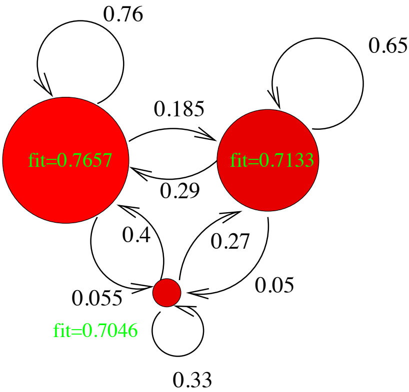
Basins of attraction in combinatorial optimization

Example of small NK landscape with $N = 6$ and $K = 2$



- Basins of attraction are interlinked and overlapped !
- Most neighbors of a given solution are outside its basin

Local optima network



- **Nodes** :
  local optima
- **Edges** :
  transition probabilities

Local optima network

Definition : Local Optima Network (LON)

Oriented weighted graph (V, E, w)

- Nodes V : set of local optima $\{LO_1, \ldots, LO_n\}$
- Edges E : notion of connectivity between local optima

Local optima network

Definition : Local Optima Network (LON)

Oriented weighted graph (V, E, w)

- Nodes V : set of local optima $\{LO_1, \ldots, LO_n\}$
- Edges E : notion of connectivity between local optima

2 possible definitions for edges

- **Basin-transition edges** :
  transition between random solutions from basin b_i to basin b_j
  ([OTVD08], [VOT08], [TVO08], [VOT10])
- **Escape edges** :
  transition from Local Optimum i to basin b_j
  (EA 2011, GECCO 2012, PPSN 2012, EA 2013 [DVOT13])

Basin-transition edges : random transition between basins

Edges

e_{ij} between LO_i and LO_j if $\exists\ x_i \in b_i$ and $x_j \in b_j : x_j \in \mathcal{N}(x_i)$

Prob. from solution x to solution x'

$$p(x \rightarrow x') = \Pr(x' = op(x))$$

Prob. from solution s to basin b_j

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

Weights : Transition prob. from basin b_i to basin b_j

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

LON with escape edges

Definition : Local Optima Network (LON)

Orienter weighted graph (V, E, w)

- Notes V : set of local optima $\{LO_1, \ldots, LO_n\}$
- Edges E : notion of connectivity between local optima

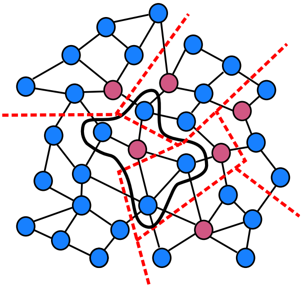
Escape edges

Edge e_{ij} between LO_i and LO_j
if $\exists x$: $distance(LO_i, x) \leqslant D$ and $x \in b_j$

Weights

$$w_{ij} = \sharp\{x \in X \mid d(LO_i, x) \leqslant D,\, x \in b_j\}$$

can be normalized by the number of solutions at distance D



LON with escape edges

Definition : Local Optima Network (LON)

Orienter weighted graph (V, E, w)

- Notes V : set of local optima $\{LO_1, \dots, LO_n\}$
- Edges E : notion of connectivity between local optima

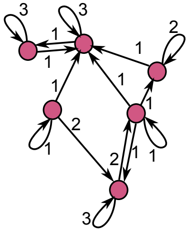
Escape edges

Edge e_{ij} between LO_i and LO_j
if $\exists x$: $distance(LO_i, x) \leqslant D$ and $x \in b_j$

Weights

$$w_{ij} = \sharp\{x \in X \mid d(LO_i, x) \leqslant D,\, x \in b_j\}$$

can be normalized by the number of solutions at distance D



Methodology

- Design, and understand LON metrics
  on tunable enumerable problem instances
  nk-landscapes, qap, ubqp, flow-shop
- Understand, and predict algorithm performances
  on enumerable instances
- Define sampling techniques for large size instance
- Understand, and predict algorithm performances
  on large instances

NK-landscapes

[Kauffman 1993] [Kau93]

$$x \in \{0,1\}^n \qquad f(x) = \frac{1}{n} \sum_{i=1}^{n} f_i(x_j, x_{i_1}, \ldots, x_{i_k})$$

Two parameters

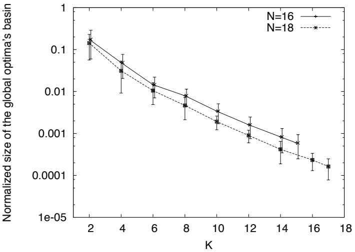
- Problem size n
- Non-linearity k < n
  (multi-modality, epistatic interactions)
  - $k = 0$: linear problem, one single maxima
  - $k = n - 1$: random problem, number of local optima $\frac{2^N}{N+1}$

note : similar results for QAP and flowshop

Basins of attraction features

- **Basin of attraction** :
  - Size :
    average, distribution . . .
  - Fitness of local optima :
    average, distribution, correlation . . .

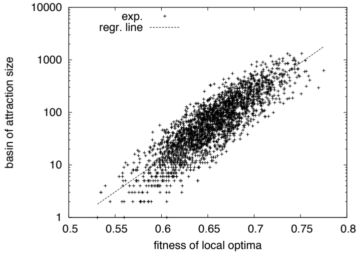
Global optimum basin size *vs.* non-linearity degree k



Size of the global maximum basin as a function of non-linearity degree k

- Basin size of maximum decreases exponentially with non-linearity degree
- $\Rightarrow$ Difficulty of (best-improvement) hill-climber from a random solution

Fitness of local optima *vs.* basin size

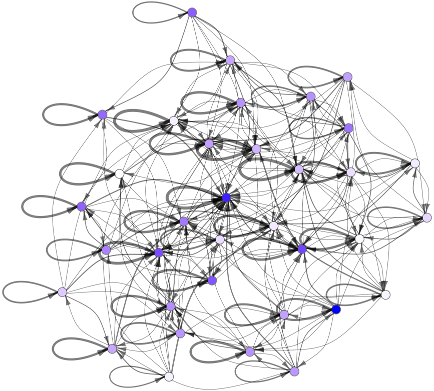


Correlation fitness of local optima *vs.* their corresponding basins sizes

The highest, the largest !

- On average, the global optimum is easier to find than one given other local optimum
- . . . but more difficult to find, as the number of local optima increases exponentially with k

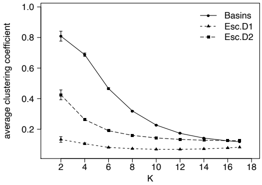
Features form the local optima network

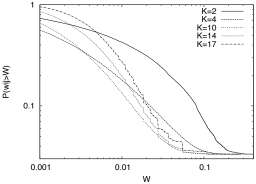


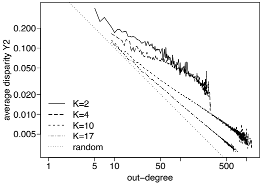
- nv : ♯vertices
- lv : avg path length
  $d_{ij} = 1/w_{ij}$
- lo : path length to best
- fnn : fitness corr.
  $(f(x), f(y))$ with $(x, y) \in E$
- wii : self loops
- wcc : weighted clust. coef.
- $zout$: out degree
- $y2$: disparity
- knn : degree corr.
  $(deg(x), deg(y))$ with $(x, y) \in E$

Structure of the local optima network

- NK-landscapes (small instances) :
  most of features are correlated with k
  relevance of the LON definition

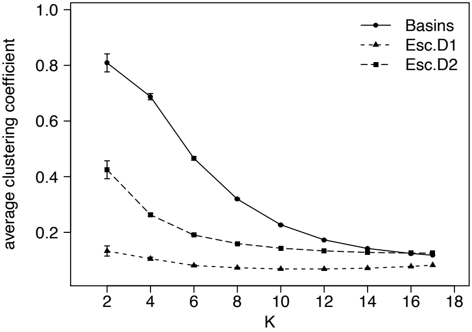






- LON is **not a random** network (NK, QAP, FSSP) :

Example : clustering coefficient for NK-landscapes



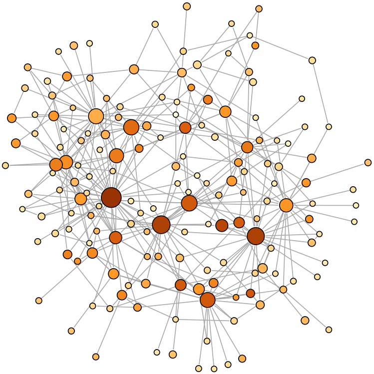
- Network highly clustered
- Clustering coefficient decreases with the degree of non-linearity k

LON to compare instance difficulty

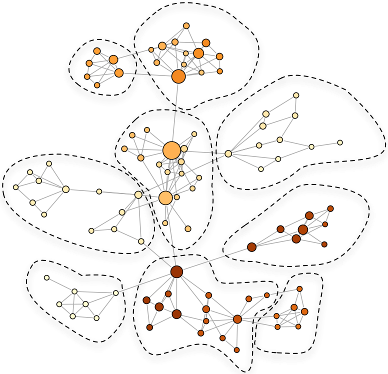
Local Optima Network for the Quadratic Assignment Problem (QAP) [DTVO11]

→ Community detection, Funnel, Fractal dimension

Random instance



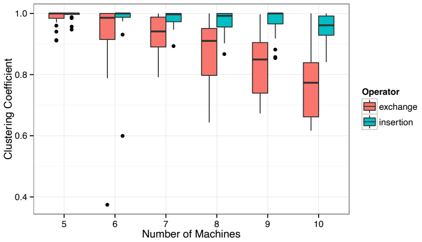
Real-like instance

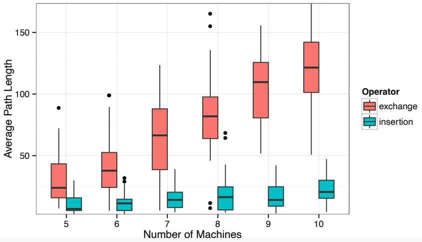


the structure of the LON is related to **problem difficulty**

Configuration : LON to compare algorithm components (1)

comparaison of **operators** for the Flowshop Scheduling Problem





Configuration : LON to compare algorithm components (2)

comparaison of the hill-climbing's **pivot rule** for NK-landscapes :
First *vs.* Best improvement HC

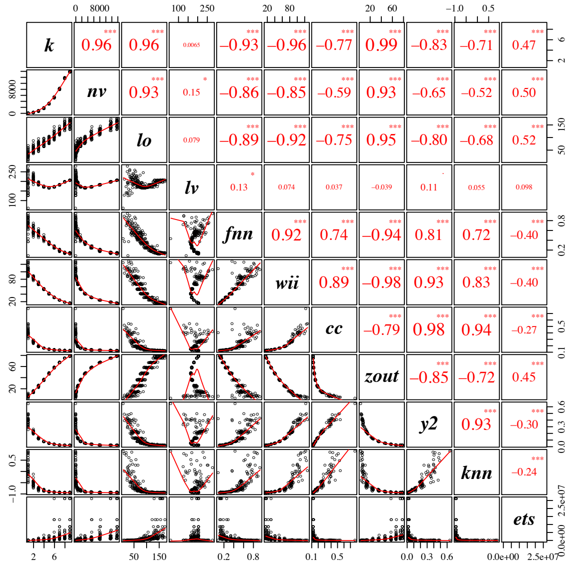
K	$\bar{n}_e / \bar{n}_v^2$		$\bar{Y}$		$\bar{d}$		$\bar{d}_{best}$	
	b-LON	f-LON	b-LON	f-LON	b-LON	f-LON	b-LON	f-LON
2	0.81	0.96	0.326	0.110	56	39	16	12
4	0.60	0.92	0.137	0.033	126	127	35	32
6	0.32	0.79	0.084	0.016	170	215	60	70
8	0.17	0.65	0.062	0.011	194	282	83	118
10	0.09	0.53	0.050	0.009	206	340	112	183
12	0.05	0.44	0.043	0.008	207	380	143	271

Information given by the local optima network

Advanced questions

- Can we explain the performance from LON features ?
- Can we predict the performance from LON features ?
- Can we select the relevant algorithm from LON features ?

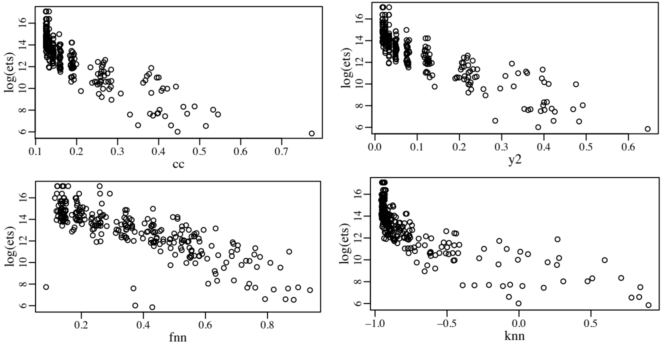
Correlation matrix (small size problem instances)



LON features *vs.* performance : simple correlation

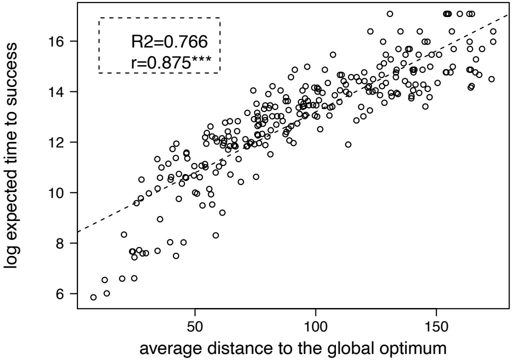
Algorithm : Iterated Local Search on NK-landscapes with $N = 18$
Performance : $ert = \mathbb{E}(T_s) + \left(\frac{1-p_s}{p_s}\right) T_{max}$

n_v	$\bar{d}_{best}$	$\bar{d}$	fnn	w_{ii}	$\bar{C}^w$	zout	$\bar{Y}$	knn
0.885	0.915	0.006	−0.830	−0.883	−0.875	0.885	−0.883	−0.850



ILS performance vs LON metrics

NK-landscapes [DVOT12]



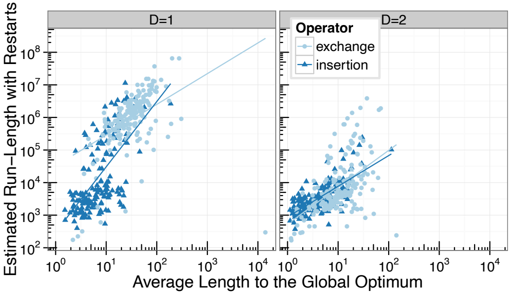
Expected running time

vs.

Average shortest path to the global optimum

ILS performance vs LON metrics

Flow-Shop Scheduling Problem [EA'13]



Expected running time

vs.

Average shortest path to the global optimum

LON features *vs.* performance : multi-linear regression

1. Multiple **linear** regression on all possible predictors :

$$\log(ert) = \beta_0 + \beta_1 k + \beta_2 \log(nv) + \beta_2 lo + \cdots + \beta_{10} knn + \varepsilon$$

2. Step-wise **backward elimination** of each predictor in turn

Predictor	$\widehat{\beta_i}$	Std. Error	p-value
(Intercept)	10.3838	0.58512	$9.24 \cdot 10^{-47}$
lo	0.0439	0.00434	$1.67 \cdot 10^{-20}$
zout	−0.0306	0.00831	$2.81 \cdot 10^{-04}$
y2	−7.2831	1.63038	$1.18 \cdot 10^{-05}$
knn	−0.7457	0.40501	$6.67 \cdot 10^{-02}$

Multiple R^2 : 0.8494, Adjusted R^2 : 0.8471

LON features *vs.* performance : multi-linear regression

for the **Flowshop Scheduling Problem** using exhaustive selection

♯P	$\log(N_V)$	CC^w	F_{nn}	k_{nn}	r	$\log(L_{opt})$	$\log(L_V)$	w_{ii}	Y_2	k_{out}	C_p	$adjR^2$
1						2.13					265.54	0.574
2		−5.18				1.43					64.06	0.675
3						1.481	0.895			−0.042	16.48	0.700
4		−2.079				1.473	0.540			−0.032	8.75	0.704
5		−2.388			−1.633	1.470	0.528			−0.030	5.97	0.706

Explicability using feature importance in an interpretable model

Sampling methodology for large-size instances

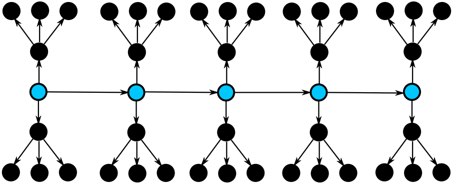
Two mains techniques (Thomson *et al.* [TOVV20]) :

- Random walk on local optima network
- Adaptive walk lon local optima network

Sampling methodology for large-size instances

From the sampling of large-size complex network :

- Random walk on the network
- Breadth-First-Search



Set of estimated LON features for large-size instances

	LON metrics
fit	Average fitness of local optima in the network
wii	Average weight of self-loops
zout	Average outdegree
y_2	Average disparity for outgoing edges
knn	Weighted assortativity
wcc	Weighted clustering coefficient
fnn	Fitness-fitness correlation on the network
	Metrics from the sampling procedure
lhc	Average length of hill-climbing to local optima
mlhc	Maximum length of hill-climbing to local optima
nhc	Number of hill-climbing paths to local optima

Performance prediction based on estimated features

- Optimization scenario using off-the-shelf metaheuristics :
  TS, SA, EA, ILS on 450 instances for NK and QAP
- Performance measures :
  average fitness / average rank
- Regression model :
  multi-linear model / random forest
- Set of features :
  - *basic* : 1st autocorr. coeff. of fitness (rw of length 10^3)
    Avg. fitness of local optima (10^3 hc)
    Avg. length to reach local optima (10^3 hc)
  - *lon* : see previous
  - *all* : *basic* and *lon* features
- Quality measure of regression :
  R^2 on cross-validation (repeated random sub-sampling)

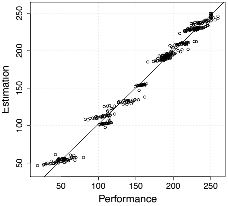
R^2 on cross-validation for NK-landscapes and QAP

Sampling parameters : length $\ell = 100$, sampled edge $m = 30$, deep $d = 2$

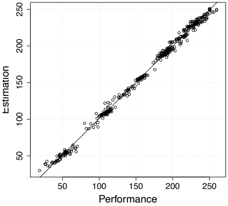
			NK					QAP				
Mod.	Feat.	Perf.	TS	SA	EA	ILS	avg	TS	SA	EA	ILS	avg
lm	basic	fit	0.8573	0.8739	0.8763	0.8874	0.8737	-38.42	-42.83	-41.63	-39.06	-40.48
lm	lon	fit	0.8996	0.9015	0.9061	0.8954	0.9007	**0.9995**	**1.0000**	**1.0000**	**0.9997**	0.9998
lm	all	fit	**0.9356**	**0.9455**	**0.9442**	**0.9501**	0.9439	**0.9996**	**0.9997**	**0.9999**	**0.9997**	0.9997
lm	basic	rank	0.8591	0.9147	0.6571	0.6401	0.7678	0.2123	0.8324	-0.0123	0.4517	0.3710
lm	lon	rank	0.9517	**0.9332**	**0.7783**	**0.7166**	0.8449	**0.7893**	**0.9673**	**0.8794**	**0.9015**	0.8844
lm	all	rank	**0.9534**	**0.9355**	**0.7809**	**0.7177**	0.8469	0.6199	0.9340	0.8577	**0.9029**	0.8286
rf	basic	fit	**0.9043**	0.9104	0.9074	**0.8871**	0.9023	0.8811	0.8820	0.8806	0.8801	0.8809
rf	lon	fit	0.8323	0.8767	0.8567	0.8116	0.8443	0.9009	0.9025	0.9027	0.9019	0.9020
rf	all	fit	0.8886	**0.9334**	**0.9196**	**0.8778**	0.9048	**0.9431**	**0.9445**	**0.9437**	**0.9429**	0.9436
rf	basic	rank	**0.9513**	0.9433	0.7729	**0.8075**	0.8687	**0.9375**	**0.9653**	0.8710	0.9569	0.9327
rf	lon	rank	0.9198	0.9291	0.7979	0.7798	0.8566	0.9308	0.9630	**0.8820**	0.9601	0.9340
rf	all	rank	**0.9554**	**0.9465**	**0.8153**	0.8151	0.8831	**0.9381**	**0.9668**	**0.8779**	**0.9643**	0.9368

Observed vs. estimated performance

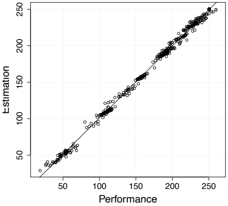
- On the 32 possibles cases (Mod. × Feat. × Algo.), the best set of features : *all* 27 times, *lon* 12 times, *basic* 6 times
- With linear model : *basic* set is never the one of the best set,
  lon features are more linearly correlated with performance
- Random forest model obtains higher regression quality :
  basic can be one of the best set (2 times)
  Nevertheless, 7/8 cases, *all* features are the best one



basic, $R^2 = 0.9327$



lon, $R^2 = 0.9601$



all, $R^2 = 0.9643$

Portfolio scenario

- Portfolio of 4 metaheuristics : TS, SA, EA, ILS
- Classification task : selection of one of the best metaheuristic
- Models : logit, random forest, svm
- Quality of classification :
  error rate (algo. is not one of the best) on cross-validation

		Avg. error rate				
Probl.	Feat.	logit	rf	svm	cst	rnd
NK	basic	0.0379	0.0278	**0.0158**		
	lon	**0.0203**	**0.0249**	**0.0168**	0.4711	0.6749
	all	0.0244	**0.0269**	**0.0165**		
QAP	basic	**0.0142**	0.0107	0.0771		
	lon	**0.0156**	**0.0086**	**0.0456**	0.4222	0.6706
	all	**0.0161**	0.0106	**0.0431**		

Conclusions and perspectives

- The structure of the local optima network . . .
  . . .can explain problem difficulty
- LON-features can be used for performance prediction
- The sampling methodology gives relevant estimation of LON features for performance prediction and algorithm portfolio

Perspectives

- Reducing the cost and improving the efficiency of the sampling
- Other (real-world, black-box) problems and algorithms
- Understanding the link between the problem definition
  and the LON structure
- Studying the LON as a fitness landscape at a large scale

In brief

Features :

input of machine learning models

Indeed, **explainability** starts with features

→ Meaningful features for meaningful analysis

How ?

- Define the neighborhood relation,
  but also search space, and fitness function
- Use/define meaningful local properties,
- Estimation of properties using sampling techniques

→ Insights about the dynamics of the optimization algorithm

References I

J. P. K. Doye.
The network topology of a potential energy landscape : a static scale-free network.
Phys. Rev. Lett., 88 :238701, 2002.

Fabio Daolio, Marco Tomassini, Sébastien Verel, and Gabriela Ochoa.
Communities of Minima in Local Optima Networks of Combinatorial Spaces.
Physica A : Statistical Mechanics and its Applications, 390(9) :1684 – 1694, July 2011.

References II

Fabio Daolio, Sébastien Verel, Gabriela Ochoa, and Marco Tomassini.
Local optima networks and the performance of iterated local search.
In
Proceedings of the fourteenth international conference on Genetic an
pages 369–376, Philadelphia, United States, July 2012. ACM.

Fabio Daolio, Sébastien Verel, Gabriela Ochoa, and Marco Tomassini.
Local Optima Networks of the Permutation Flow-Shop Problem.
In Springer, editor,
International Conference on Artificial Evolution (EA 2013),

References III

Lecture Notes in Computer Science, pages –, Bordeaux, France, October 2013.

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

Gabriela Ochoa, Marco Tomassini, Sébastien Verel, and Christian Darabos.
A Study of NK Landscapes' Basins and Local Optima Networks.
In Proceedings of the 10th annual conference on Genetic and evolutionary computation Genetic And Evolutionary Computation Conference, pages 555–562, Atlanta États-Unis d'Amérique, 07 2008. ACM New York, NY, USA.
best paper nomination.

References IV

Frank H Stillinger and Thomas A Weber.
Packing structures and transitions in liquids and solids.
Science(Washington, DC), 225(4666) :983–9, 1984.

SL Thomson, G Ochoa, S Verel, and N Veerapen.
Inferring future landscapes : Sampling the local optima level.
Evolutionary computation, page 1, 2020.

Marco Tomassini, Sébastien Verel, and Gabriela Ochoa.
Complex-network analysis of combinatorial spaces : The NK landscape case.
Physical Review E : Statistical, Nonlinear, and Soft Matter Physics, 78(6) :066114, 12 2008.
89.75.Hc ; 89.75.Fb ; 75.10.Nr.

References V

Sébastien Verel, Gabriela Ochoa, and Marco Tomassini.
The Connectivity of NK Landscapes' Basins : A Network Analysis.
In Proceedings of the Eleventh International Conference on the Simulation and Synthesis of Living Systems Artificial Life XI, pages 648–655, Winchester France, 08 2008. MIT Press, Cambridge, MA.
tea team.

Sébastien Verel, Gabriela Ochoa, and Marco Tomassini.
Local Optima Networks of NK Landscapes with Neutrality.
IEEE Transactions on Evolutionary Computation, volume 14(6) :783 – 797, November 2010.

– Tutorial on Landscape Analysis for Explainable Optimization –

4. Multi-objective Landscapes

Arnaud Liefooghe & Sébastien Verel

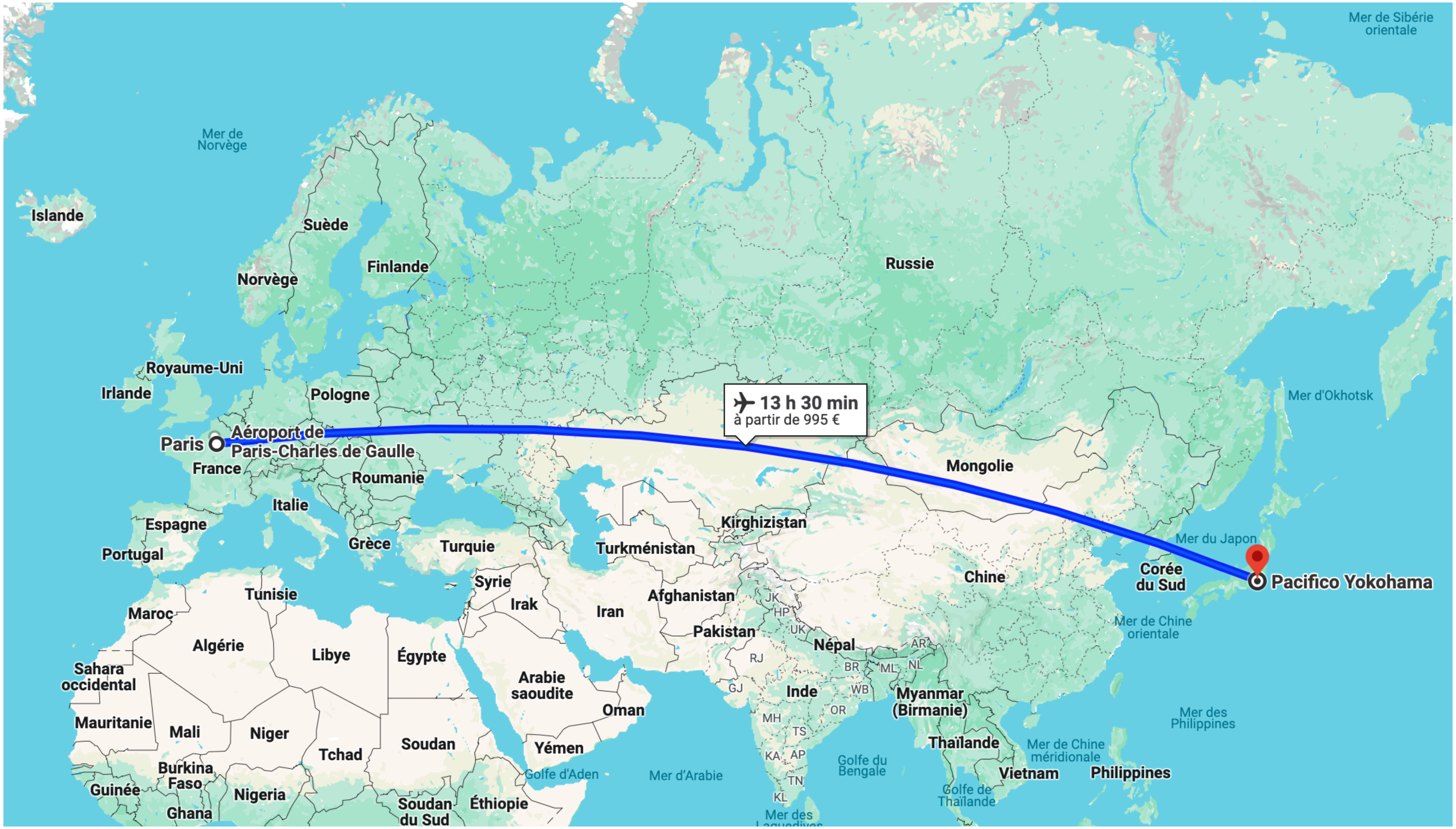
Université du Littoral Côte d'Opale – LISIC
arnaud.liefooghe@univ-littoral.fr, sebastien.verel@univ-littoral.fr

This work has benefited from the support of the National Research Agency under France 2030, MAIA Project ANR-22-EXES-0009

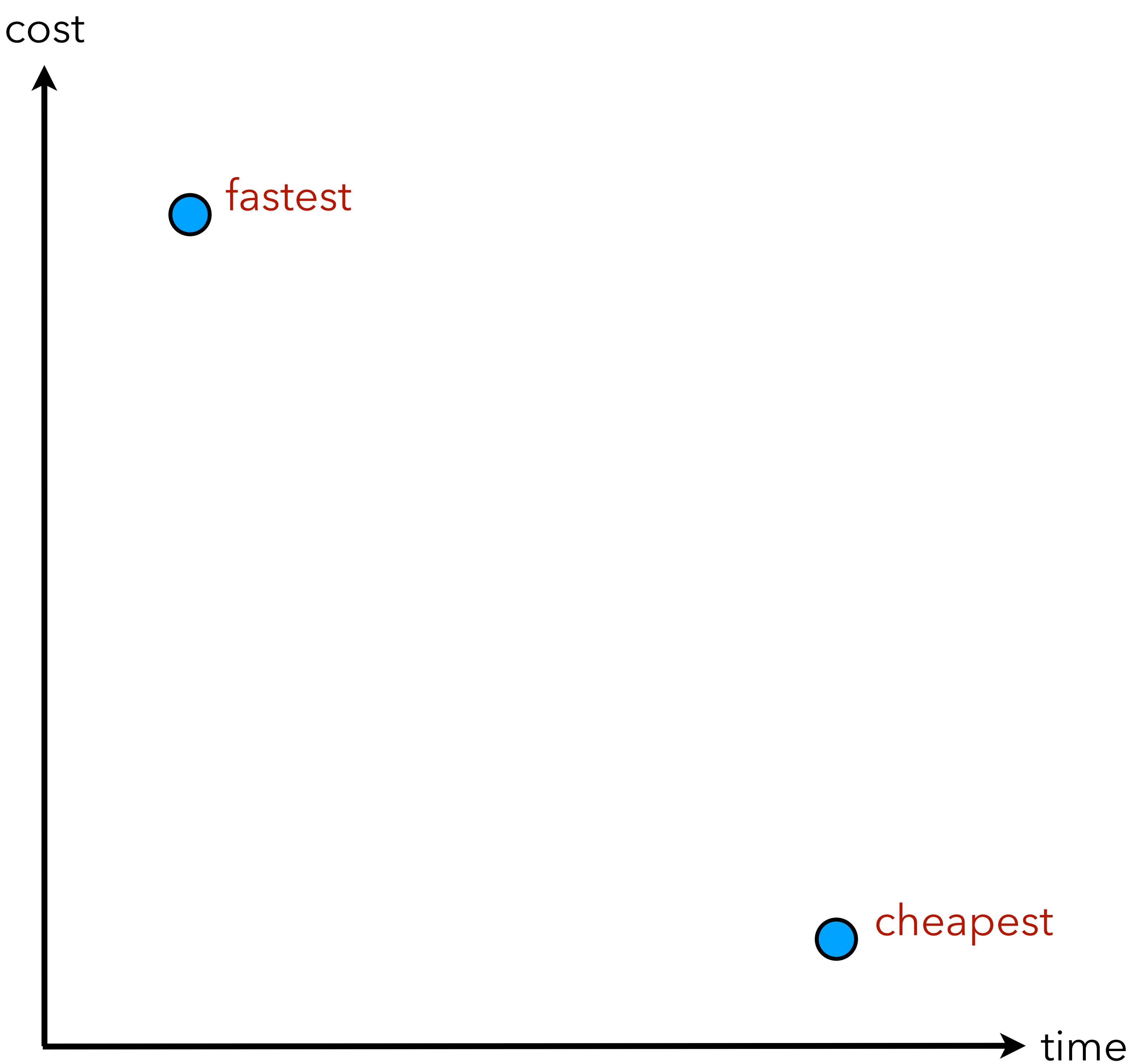
Contents

Contents

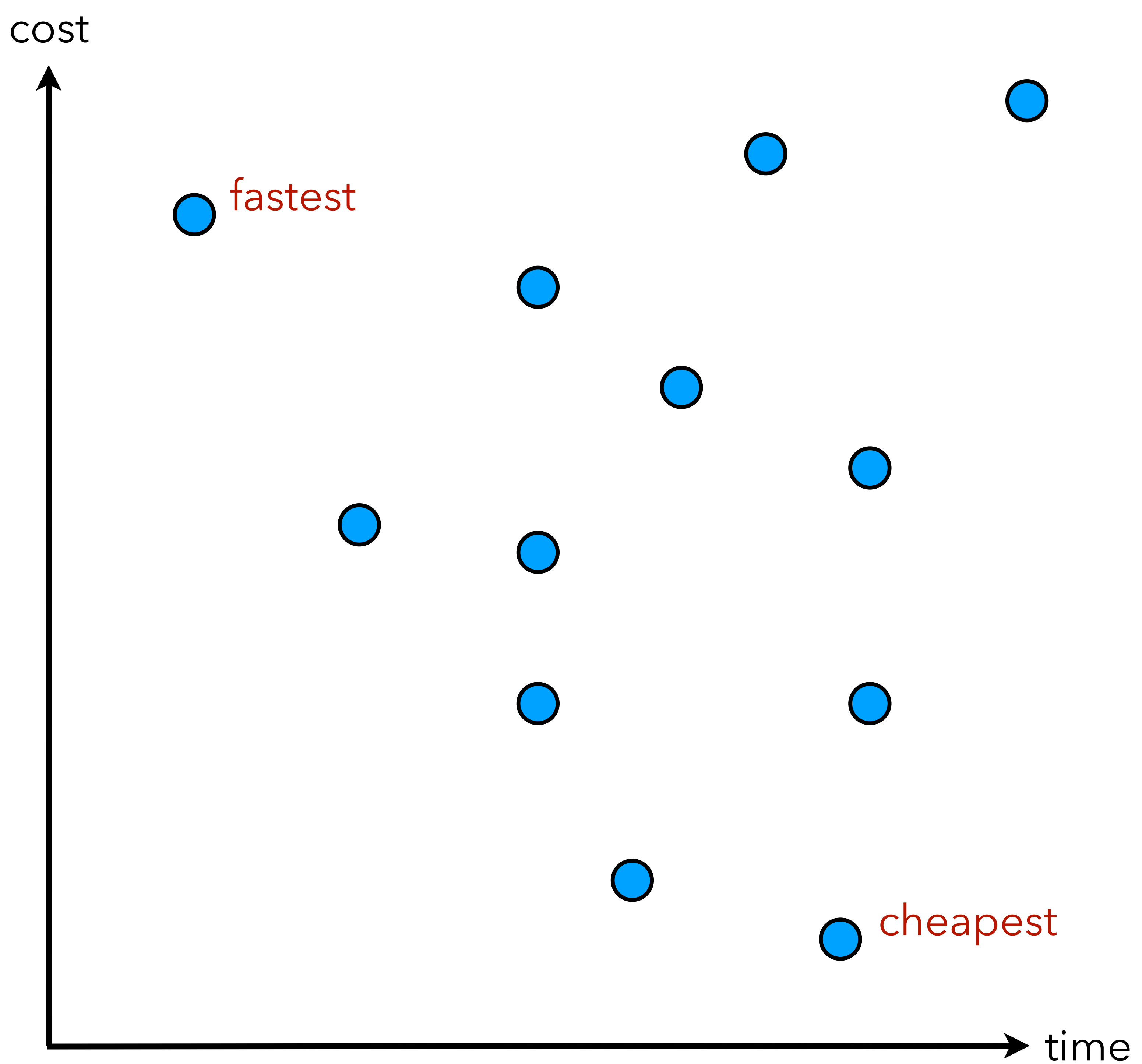
Example: Shortest Path



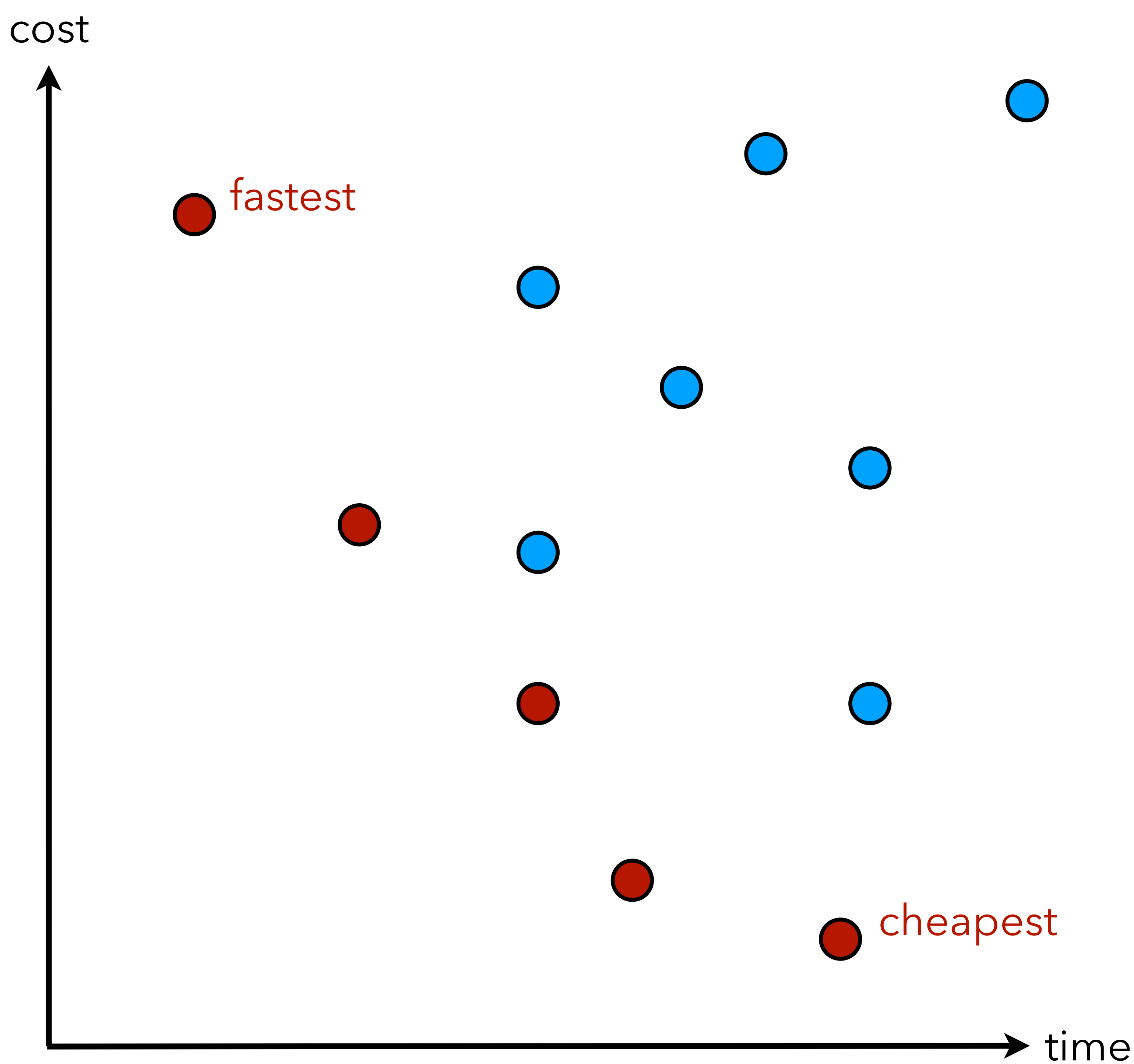
Multi-objective Shortest Path



Multi-objective Shortest Path

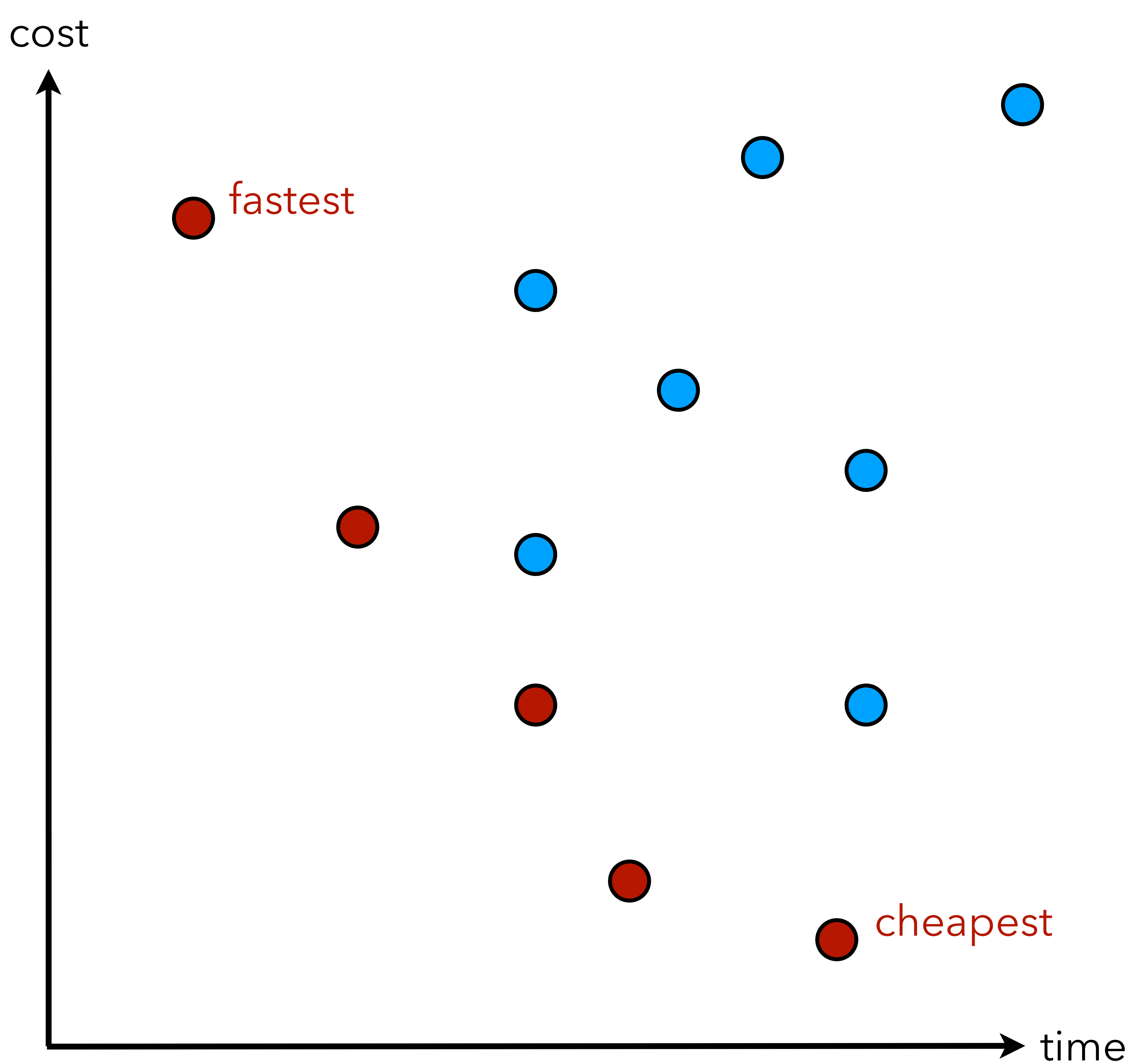


Multi-objective Shortest Path



some paths are better

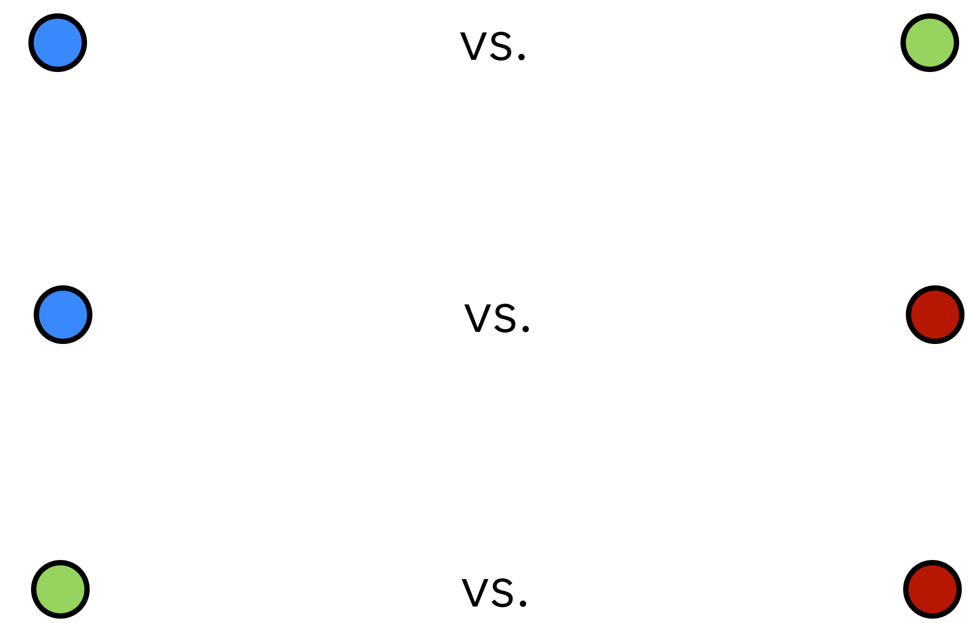
Multi-objective Shortest Path



some paths
are better

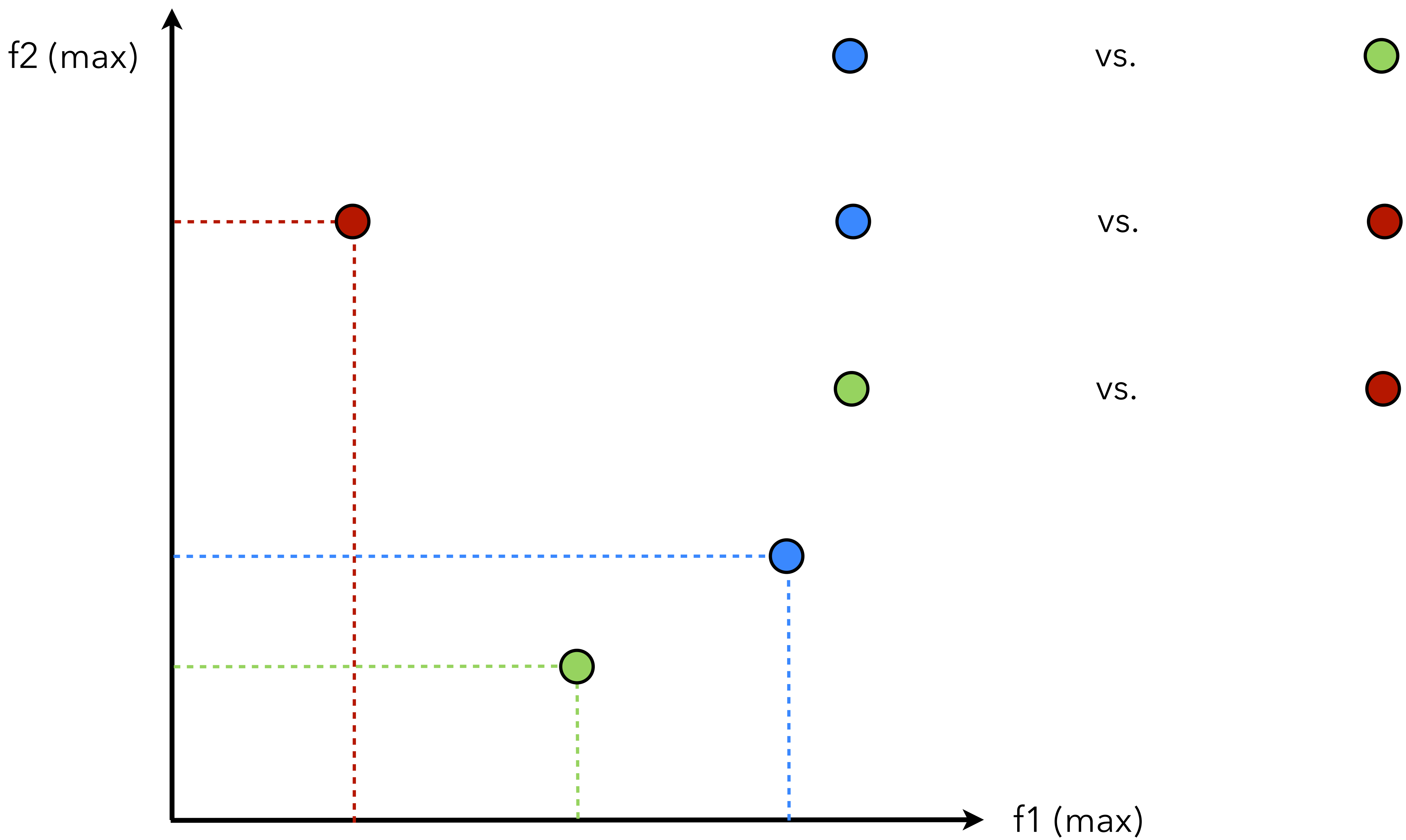
which path
is optimal?

One vs. Multiple Objectives

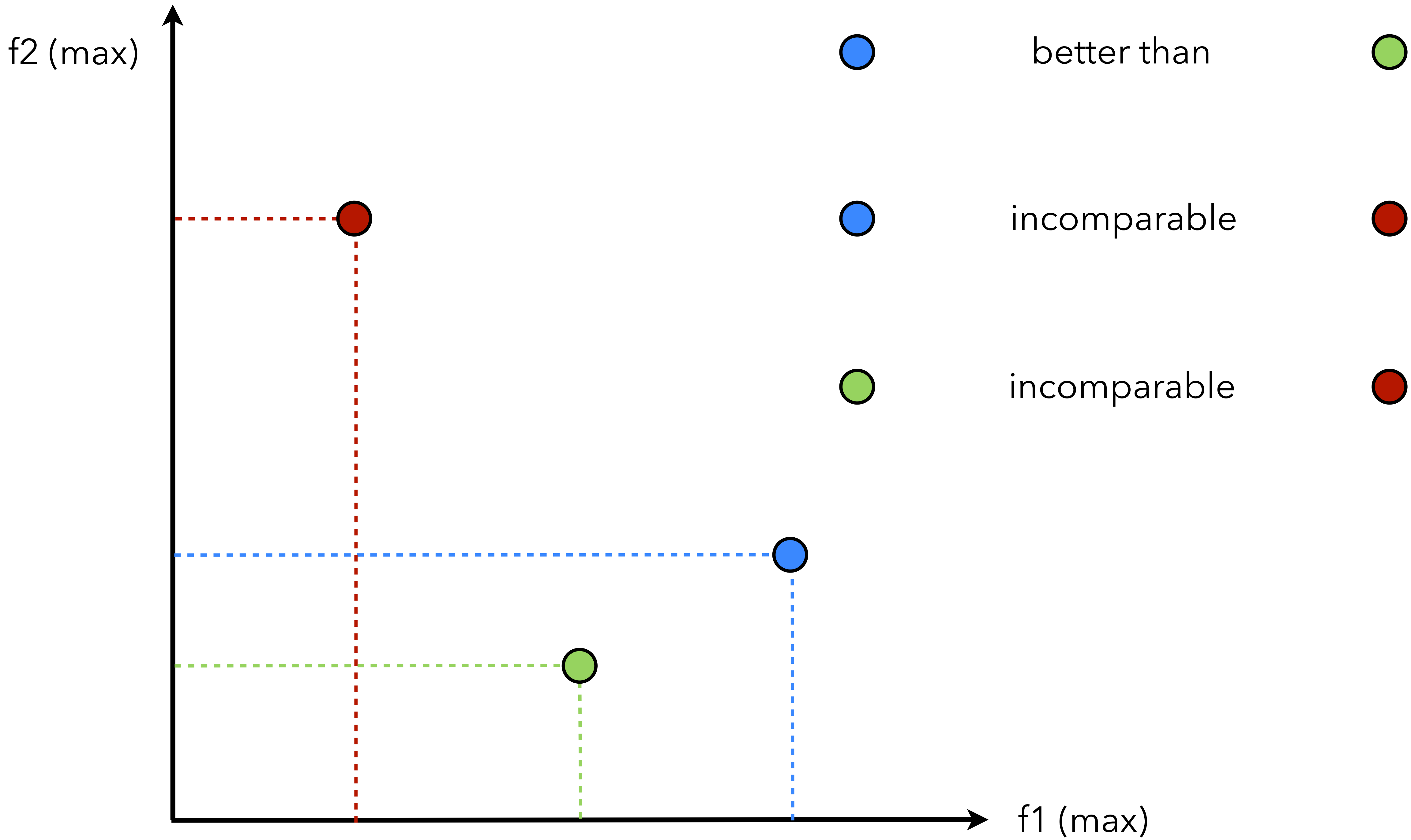


One vs. Multiple Objectives

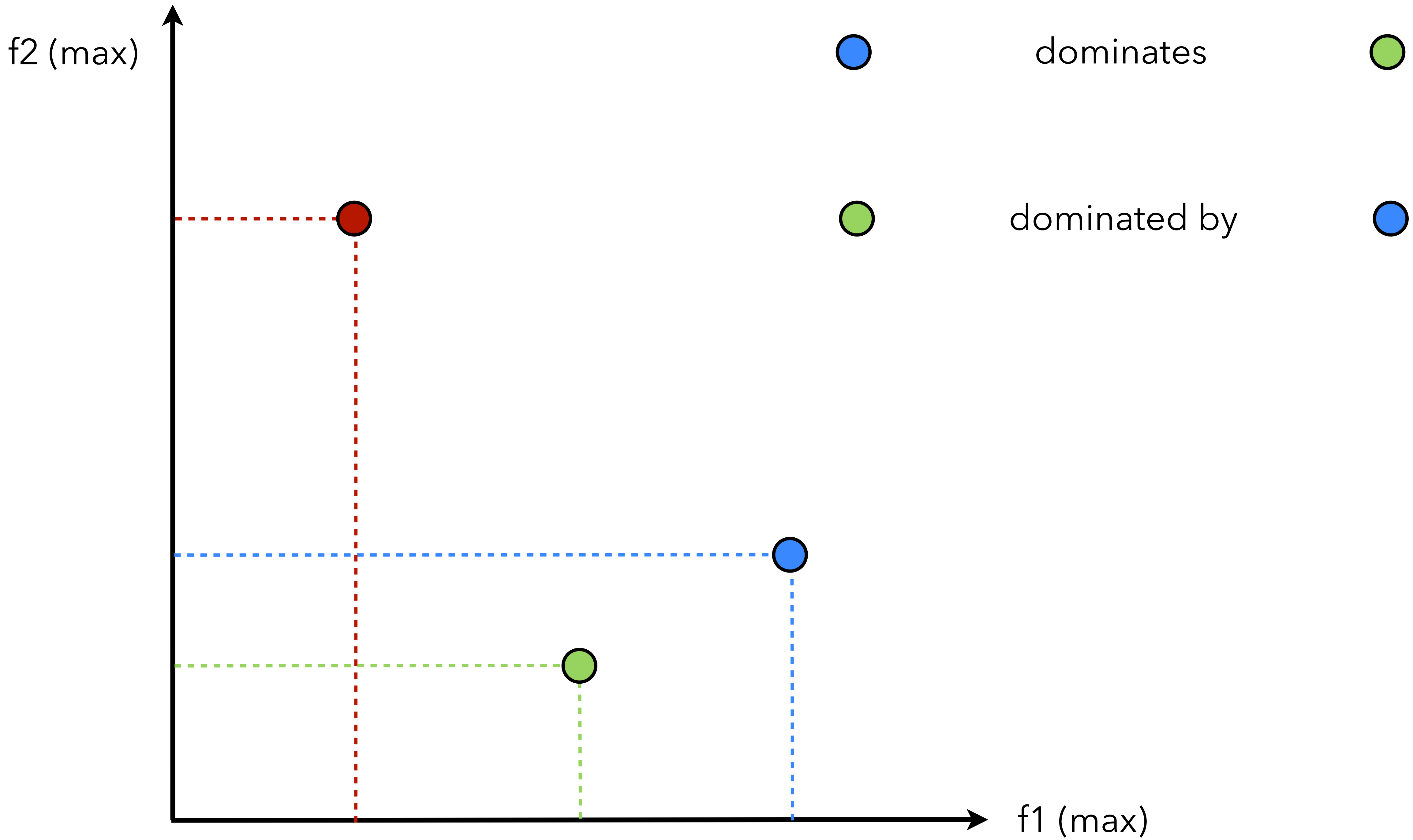
One vs. Multiple Objectives



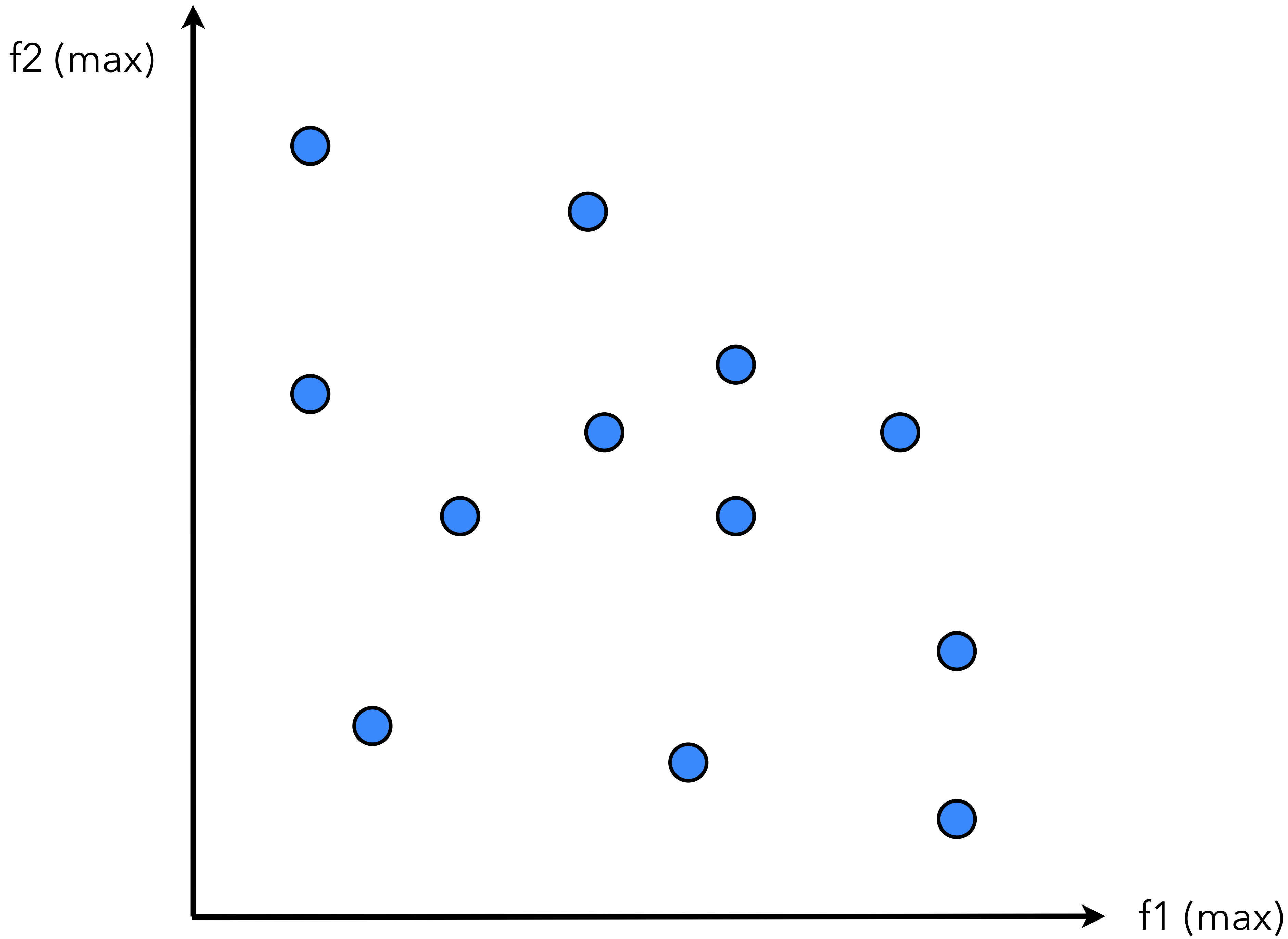
One vs. Multiple Objectives



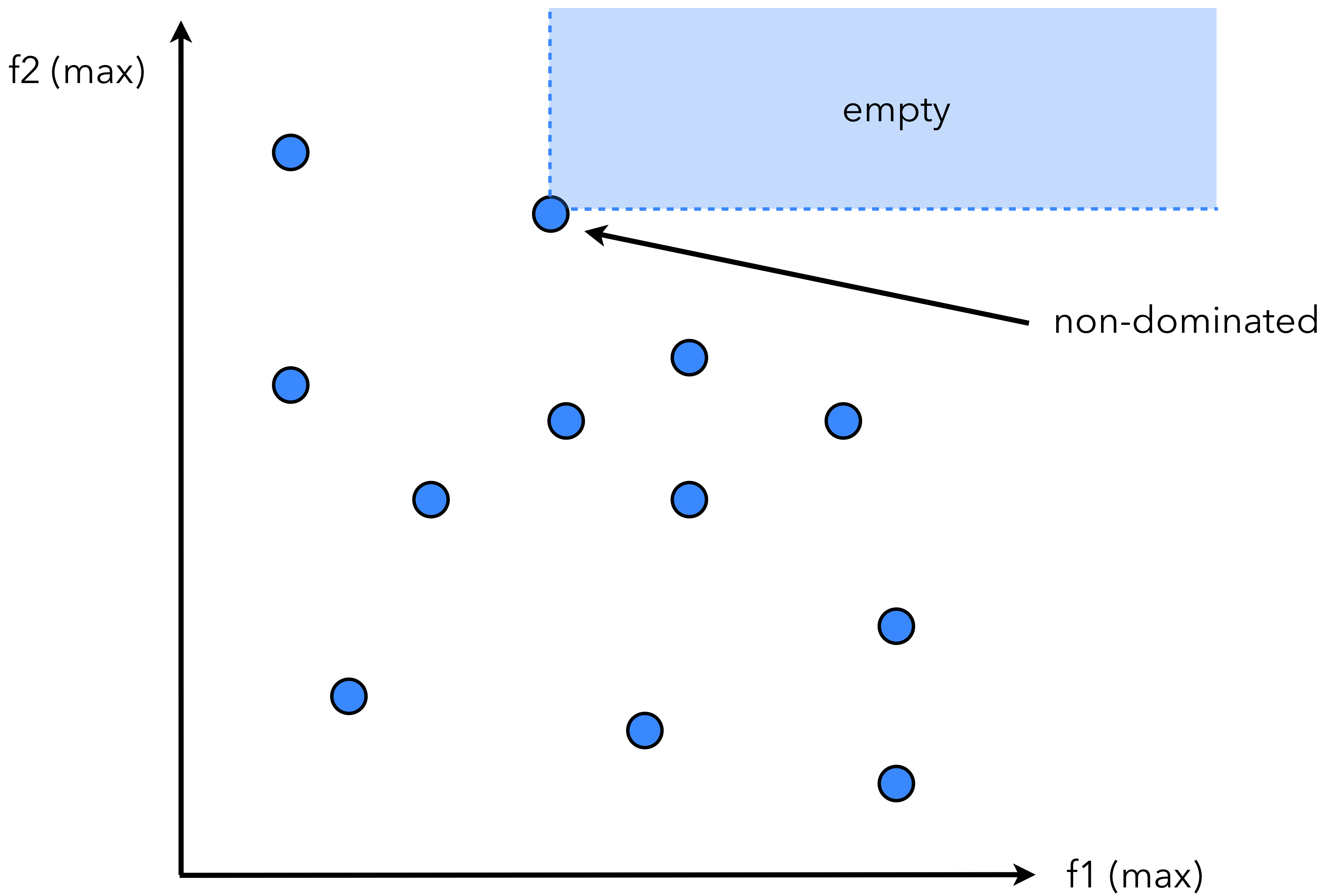
Pareto Dominance



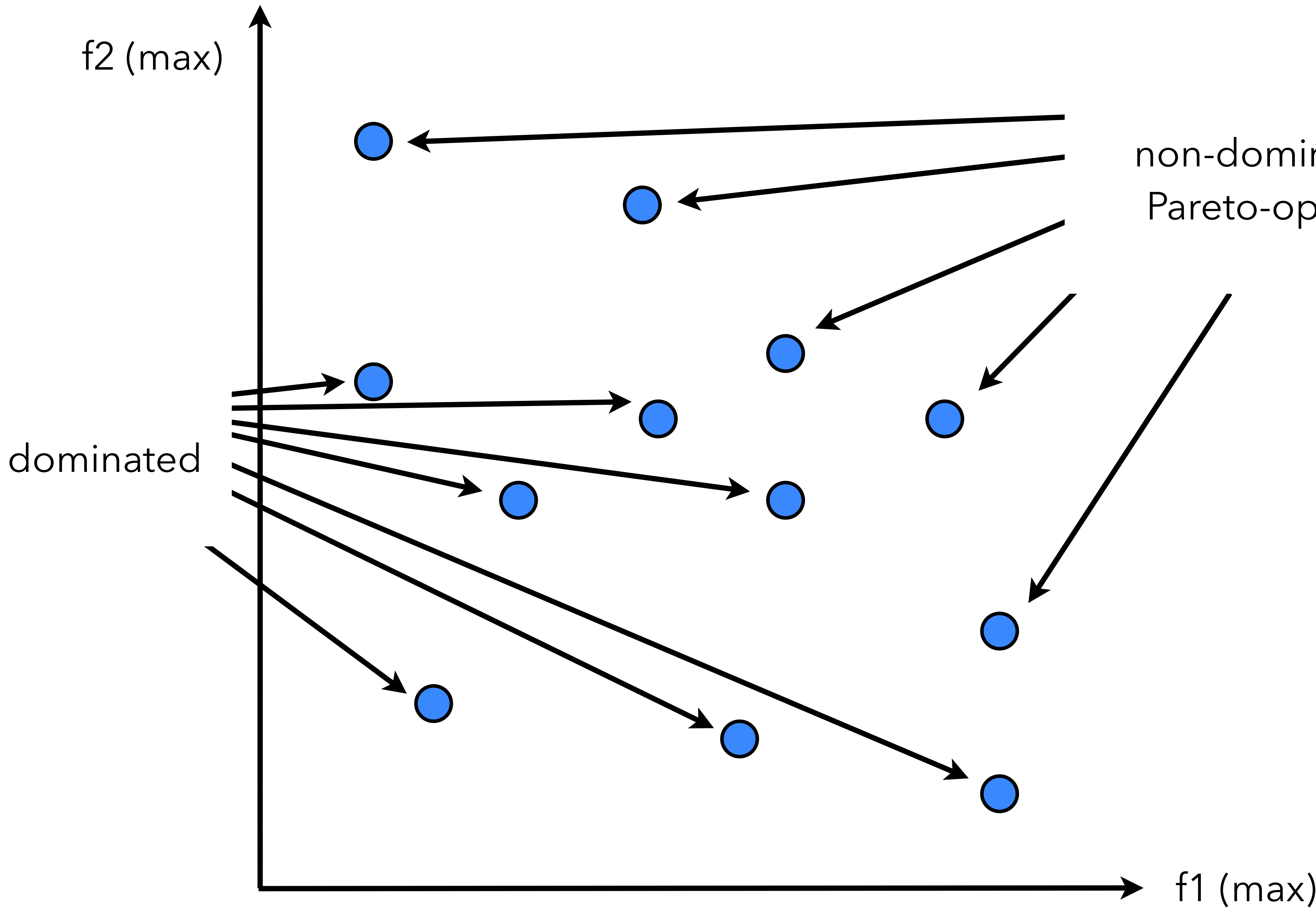
Pareto Dominance



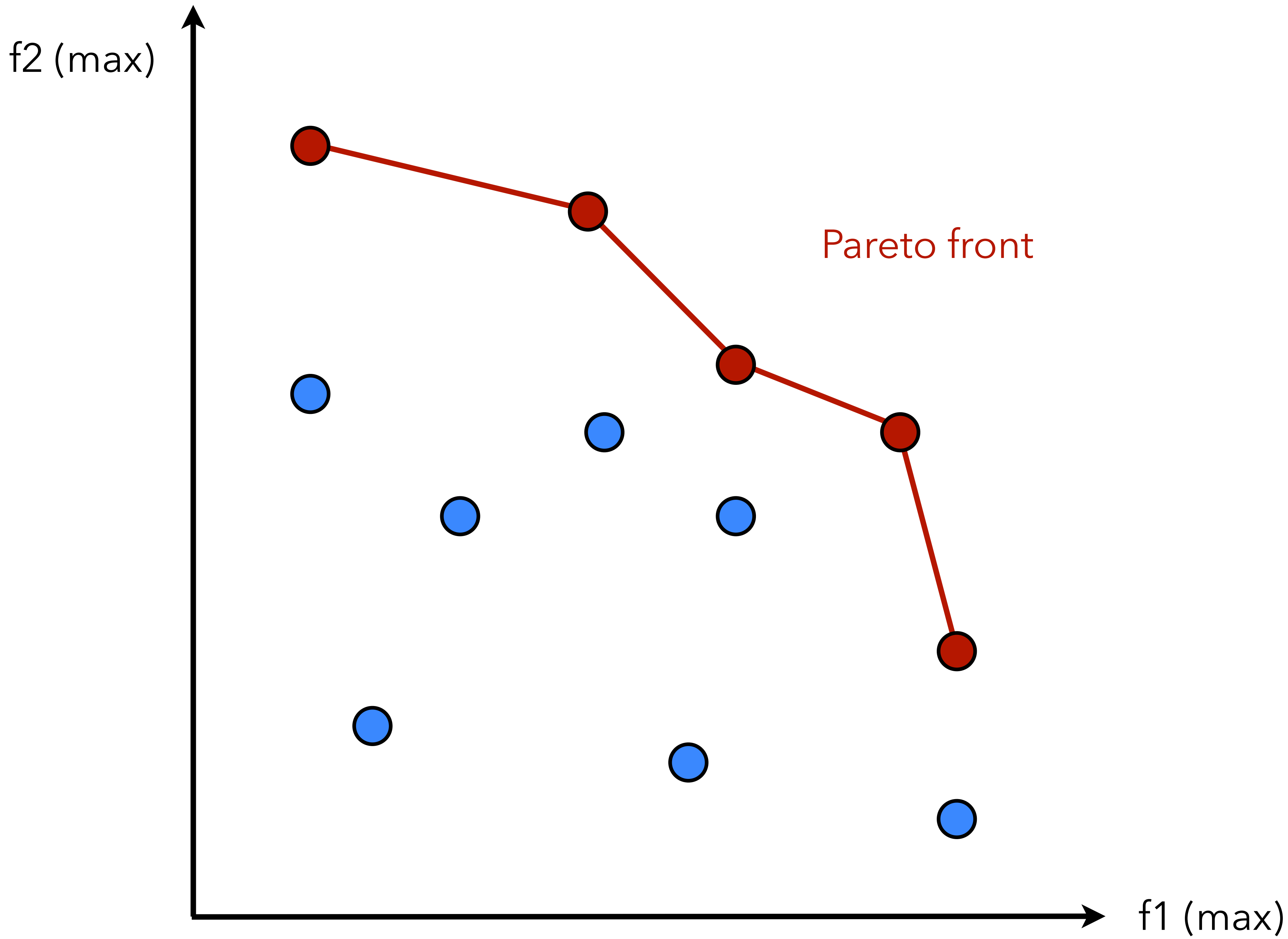
Pareto Dominance



Pareto Dominance



Pareto Front



Challenges

- **Variables** many, heterogeneous, intricate structure
- **Objectives** multiple/many, heterogeneous, conflicting, black-box (expensive)
- **Complexity** deciding if a solution is Pareto optimal is difficult for many problems
- **Intractability** number of Pareto optimal solutions often grows exponentially

How about a Pareto set approximation?

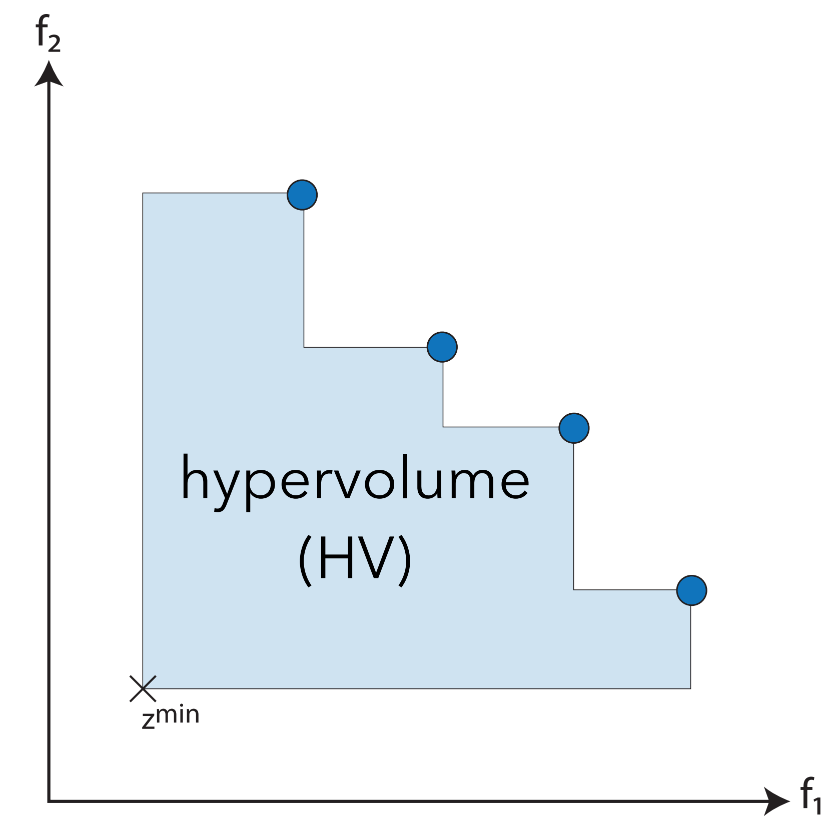
Pareto Set Approximation

Rule of thumb

- closeness to the (exact) Pareto front
- well-distributed solutions in the objective space

Quality indicators

- scalar value that reflects approximation quality
  e.g. HV, EPS, IGD, R-metrics



Local vs. Global Search

local search	global search
multi-objective hill-climber	multi-objective (1+1)–EA
PLS	G-SEMO
[Paquete et al. 2004]	[Laumanns et al. 2004]

$A \leftarrow \{x_0\}$
repeat
  select $x \in A$ at random
  for all x' s.t. $||x - x'||_1 = 1$ **do**
    $A \leftarrow$ non-dominated solutions from $A \cup \{x'\}$
  end for
until stop

$A \leftarrow \{x_0\}$
repeat
  select $x \in A$ at random
  $x' \leftarrow x$
  flip each bit x'_i with a rate $\frac{1}{n}$
  $A \leftarrow$ non-dominated solutions from $A \cup \{x'\}$
until stop

Evolutionary Search

(1) Dominance-based selection e.g. NSGA-II, G-SEMO

- search process guided by a dominance relation

(2) Indicator-based selection e.g. IBEA, SMS-EMOA

- search process guided by a quality indicator

(3) Decomposition-based selection e.g. MOEA/D

- multiple aggregations of the objectives

Evolutionary Search

(1) Dominance-based selection e.g. NSGA-II, G-SEMO

- search process guided by a dominance relation

(2) Indicator-based selection e.g. IBEA, SMS-EMOA

- search process guided by a quality indicator

(3) Decomposition-based selection e.g. MOEA/D

- multiple aggregations of the objectives

+ many other **parameters** … **which algorithm should I use?**

Contents

Sébastien Verel

Fabio Daolio

Bilel Derbel

Hernán Aguirre

Kiyoshi Tanaka

Multi-objective Optimization

Foundations of **MO Landscapes**

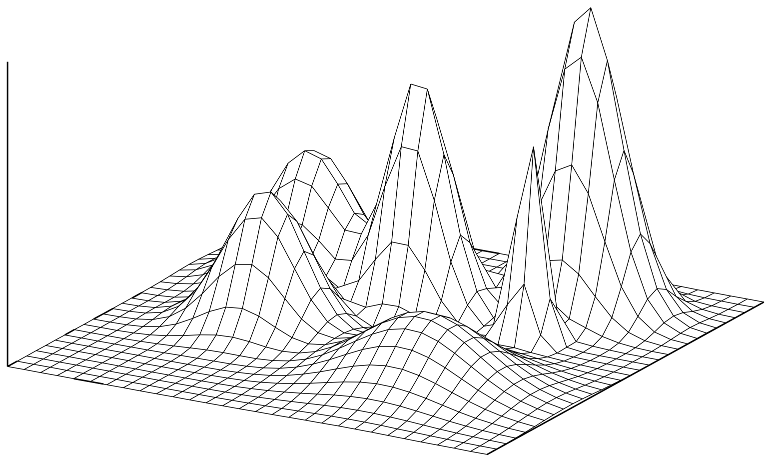
Set- and **Indicator**-based **Landscapes**

A Glimpse on related **Research Directions**

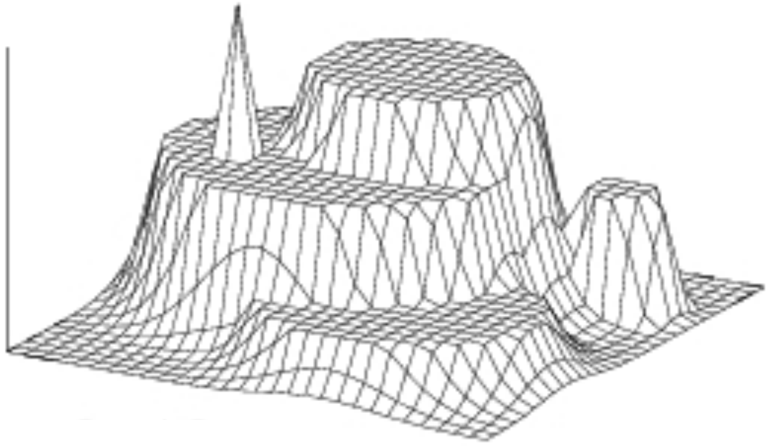
Multi-objective Landscape

- Triplet (X, N, f) such that:
  - X is a variable space
  - N : X → 2^X is a neighborhood relation
  - f : X → Z is a (black-box) objective function vector
- Features to portray multi-objective landscapes
  - Capture what makes a problem hard, a search efficient
  - Performance prediction, algorithm selection

Analyze Objectives Independently?

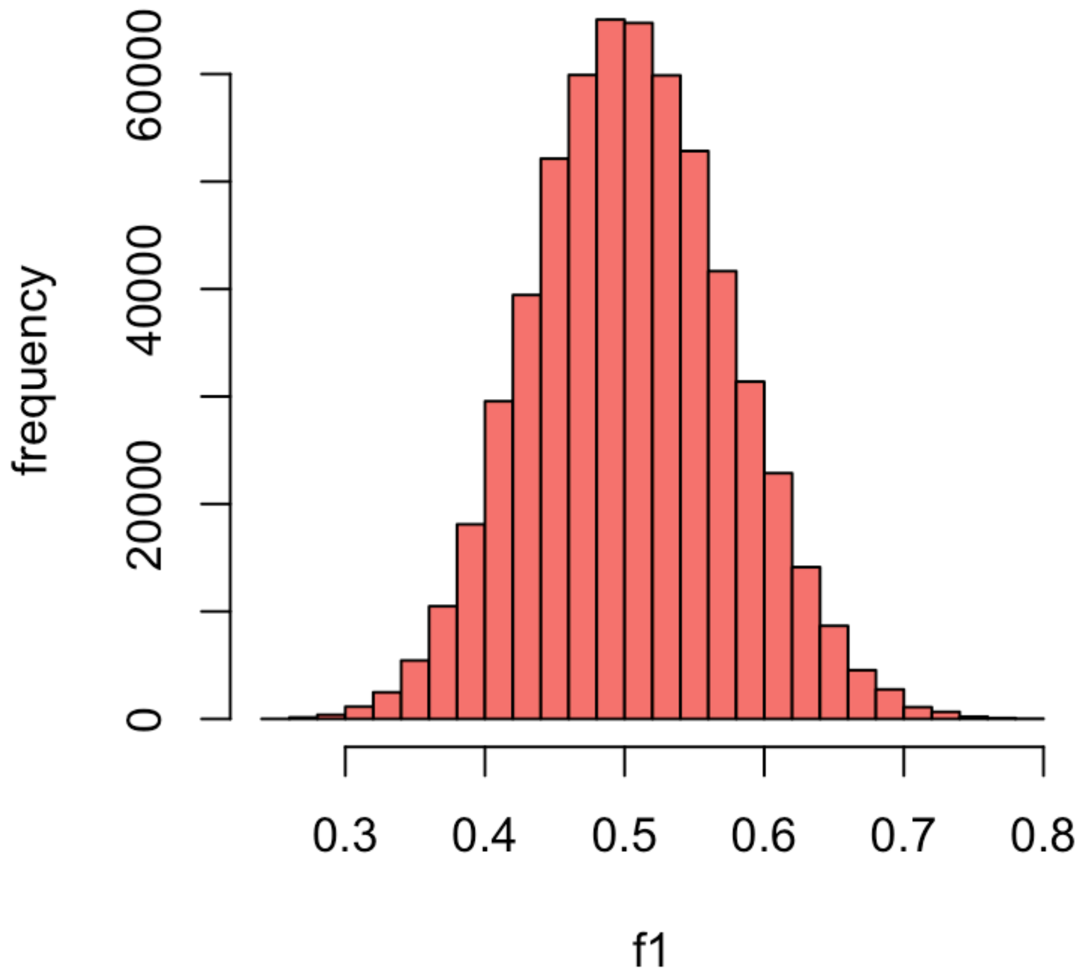


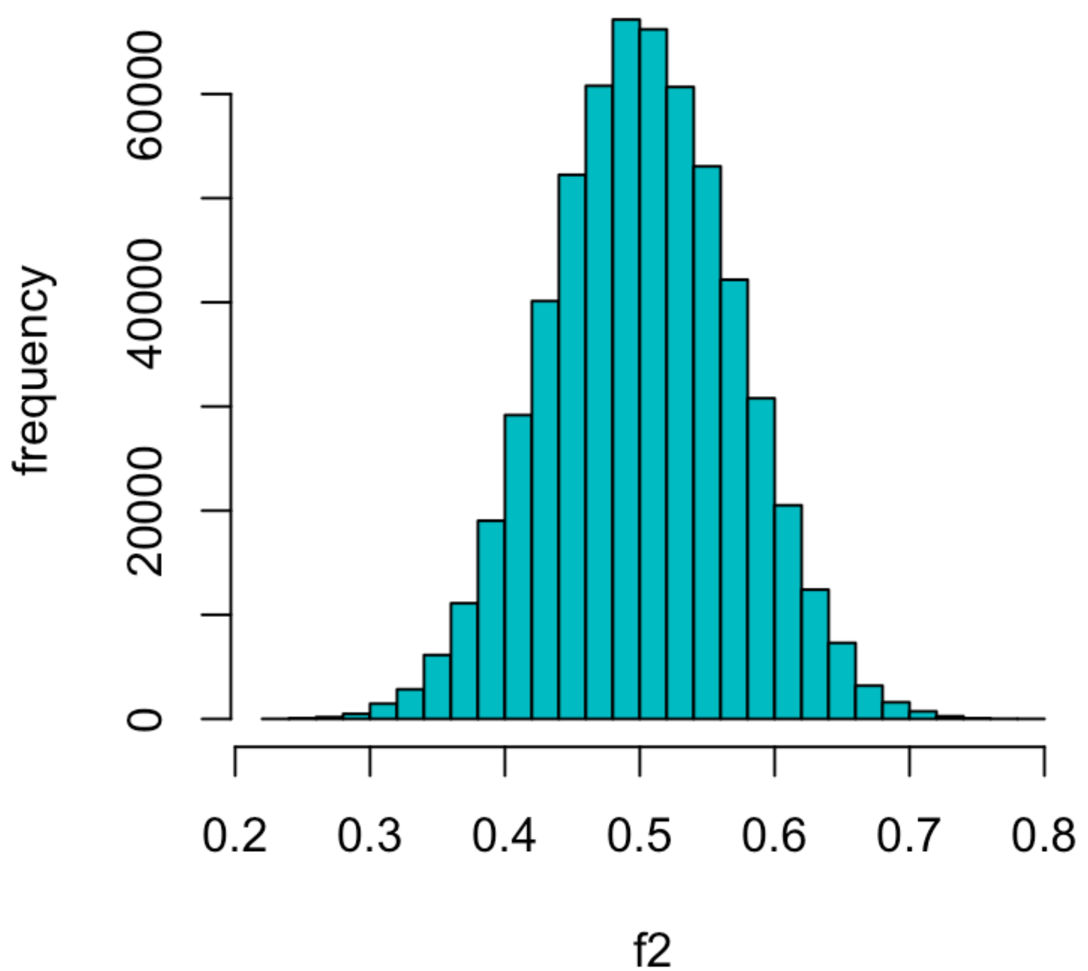
f1



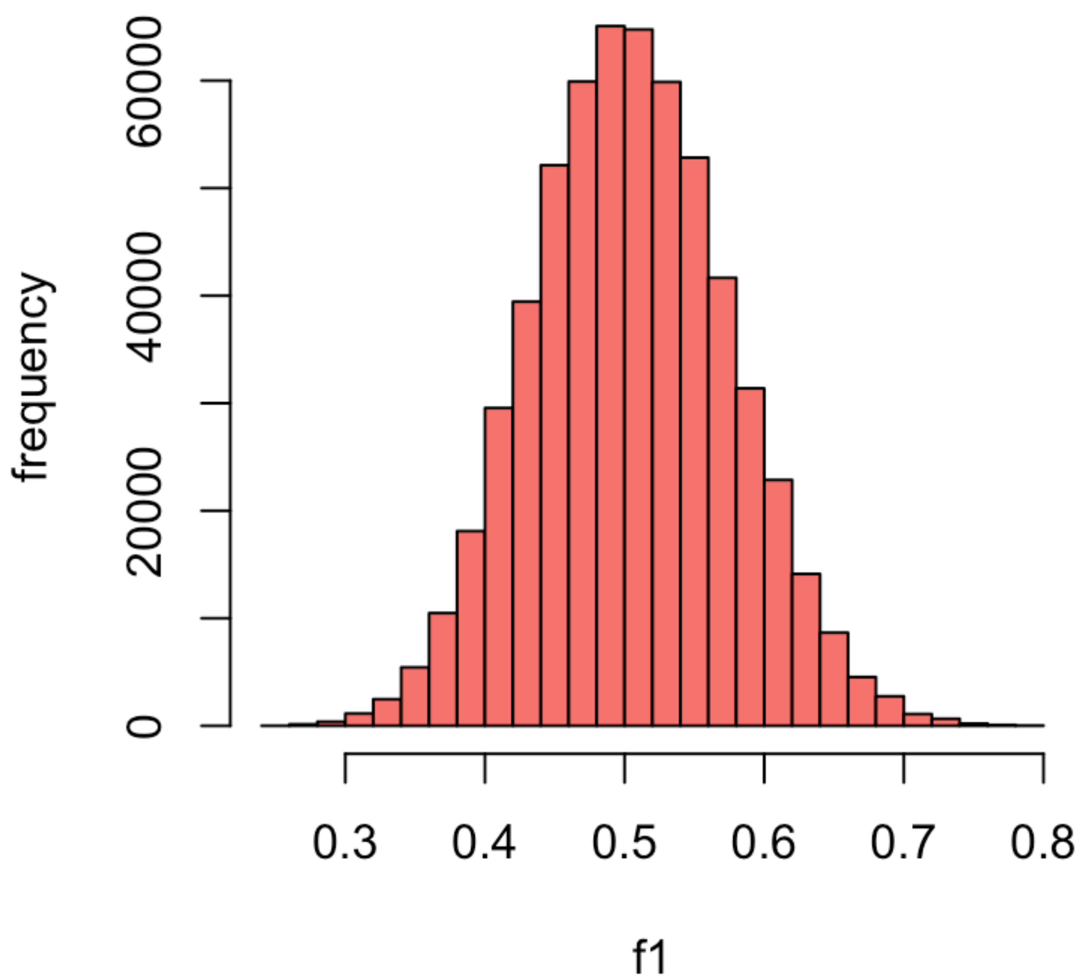
f2

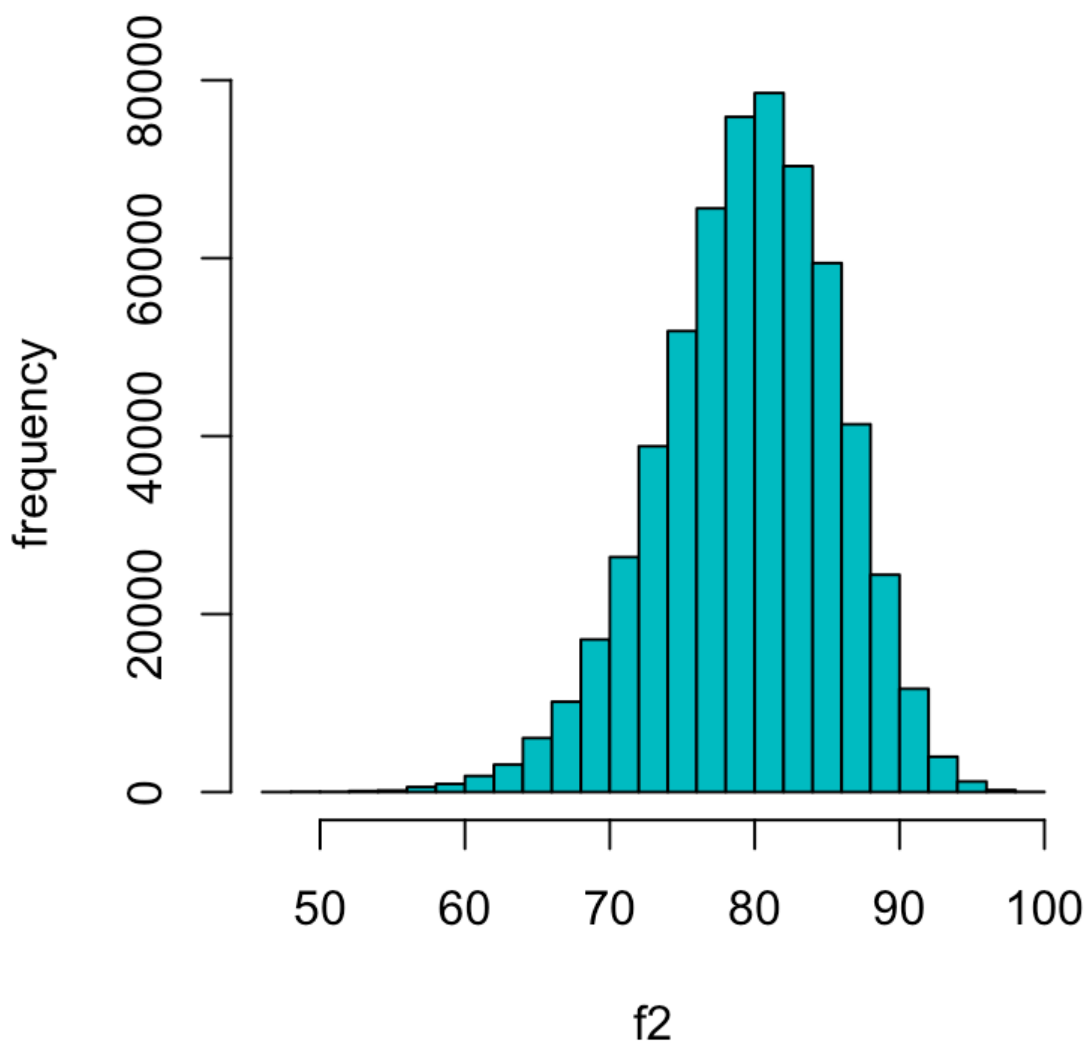
Distribution of Objective Values



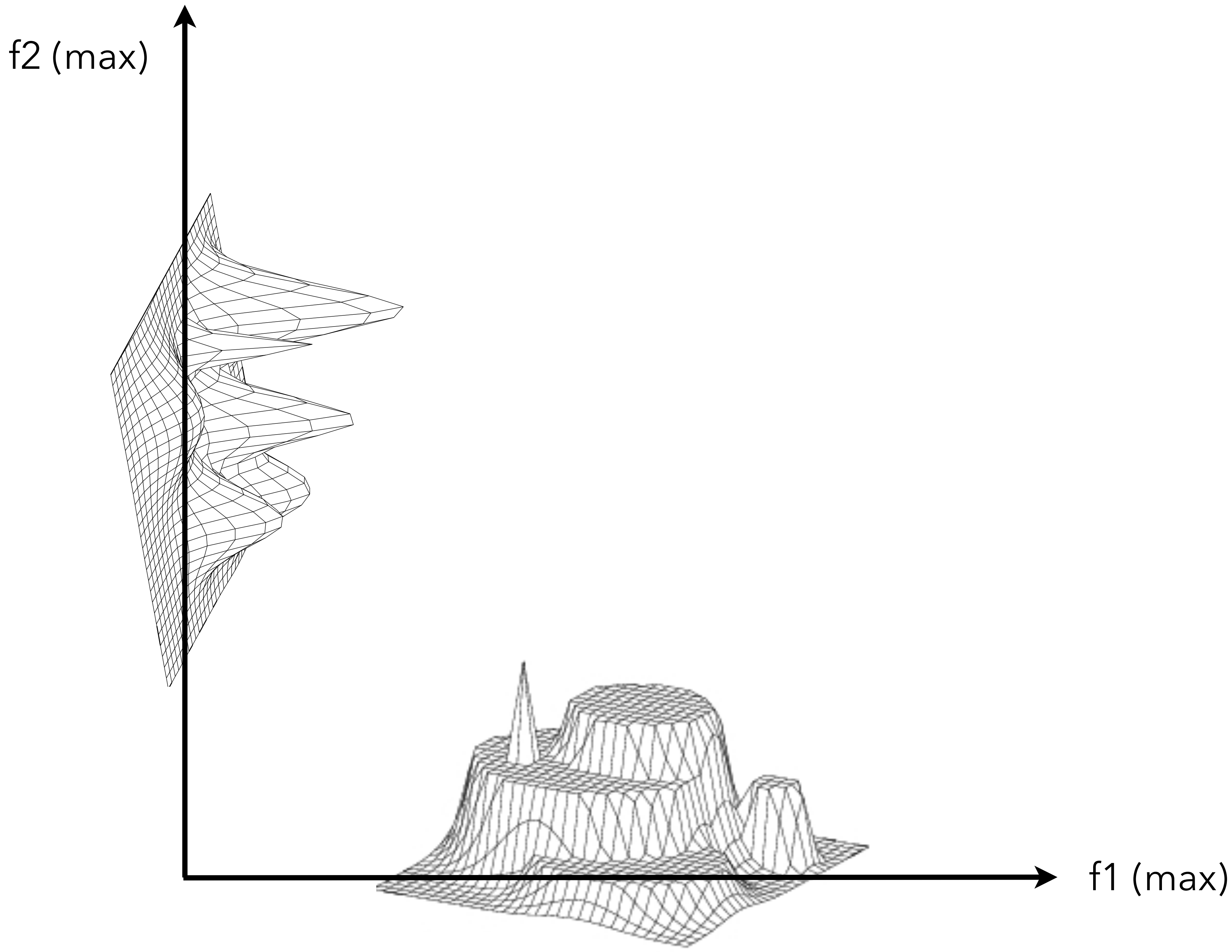


Distribution of Objective Values

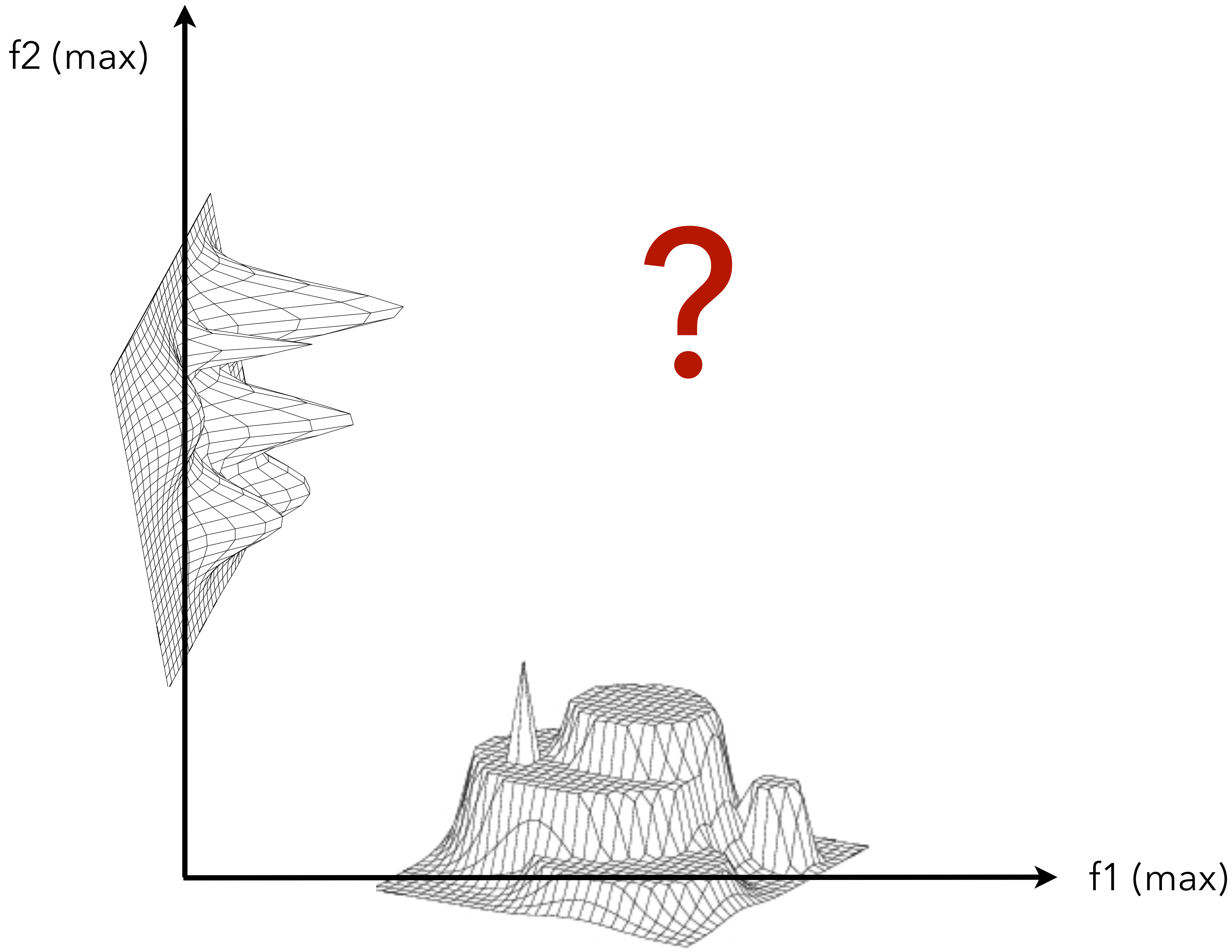




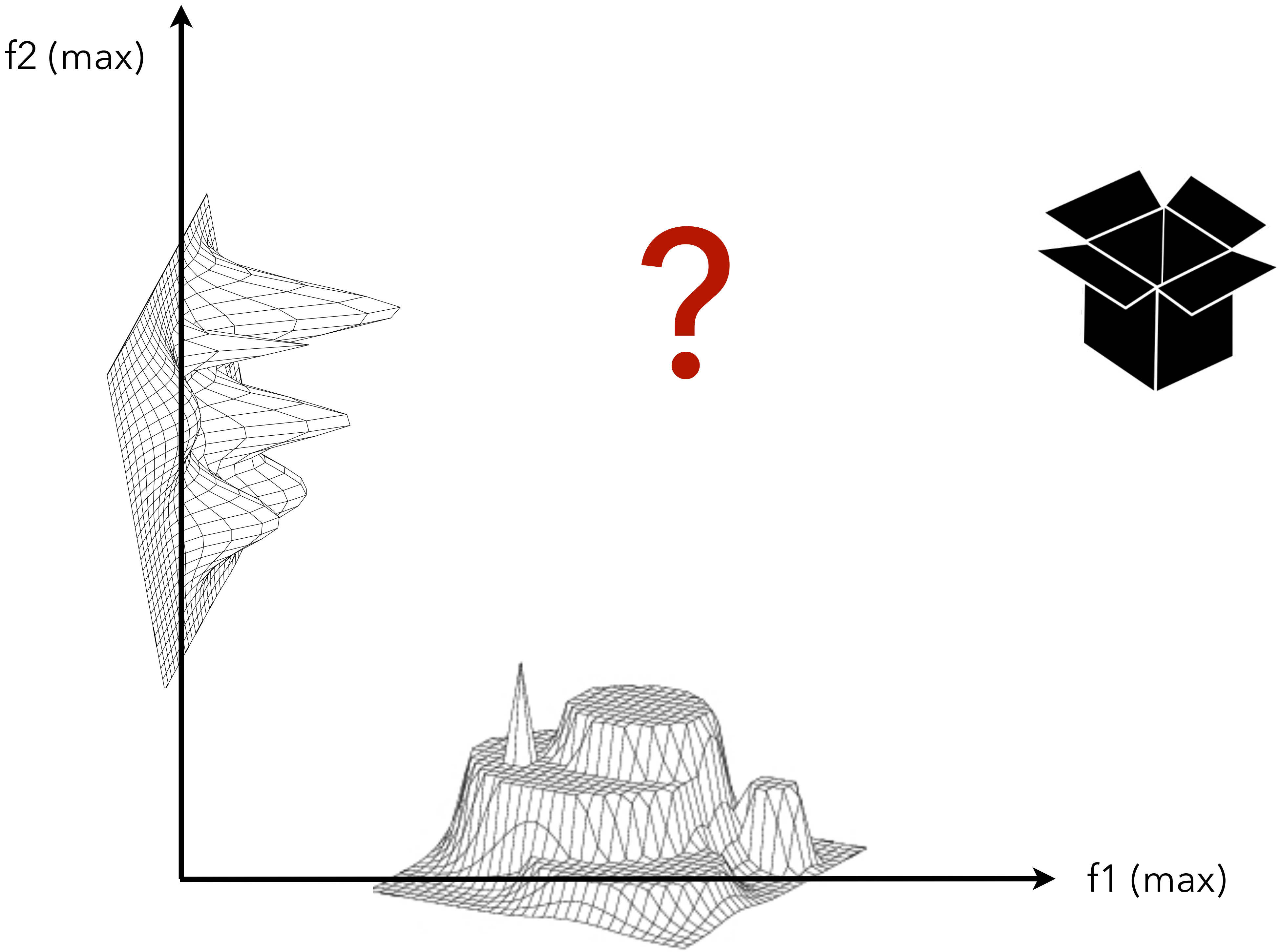
Objectives Interaction



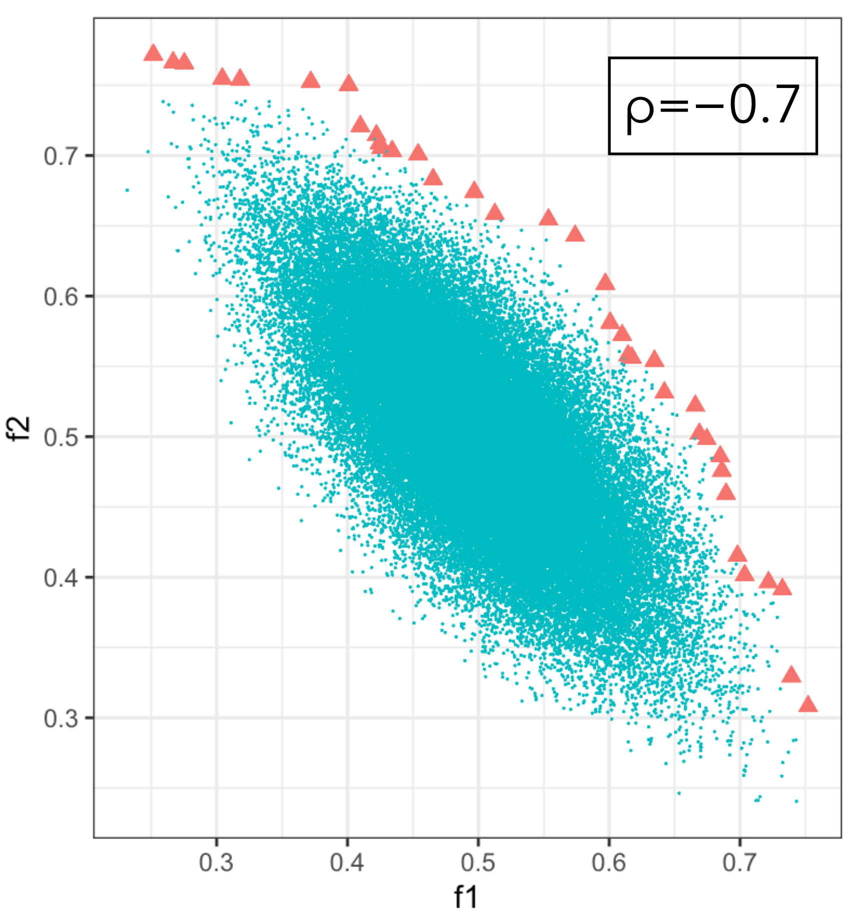
Objectives Interaction



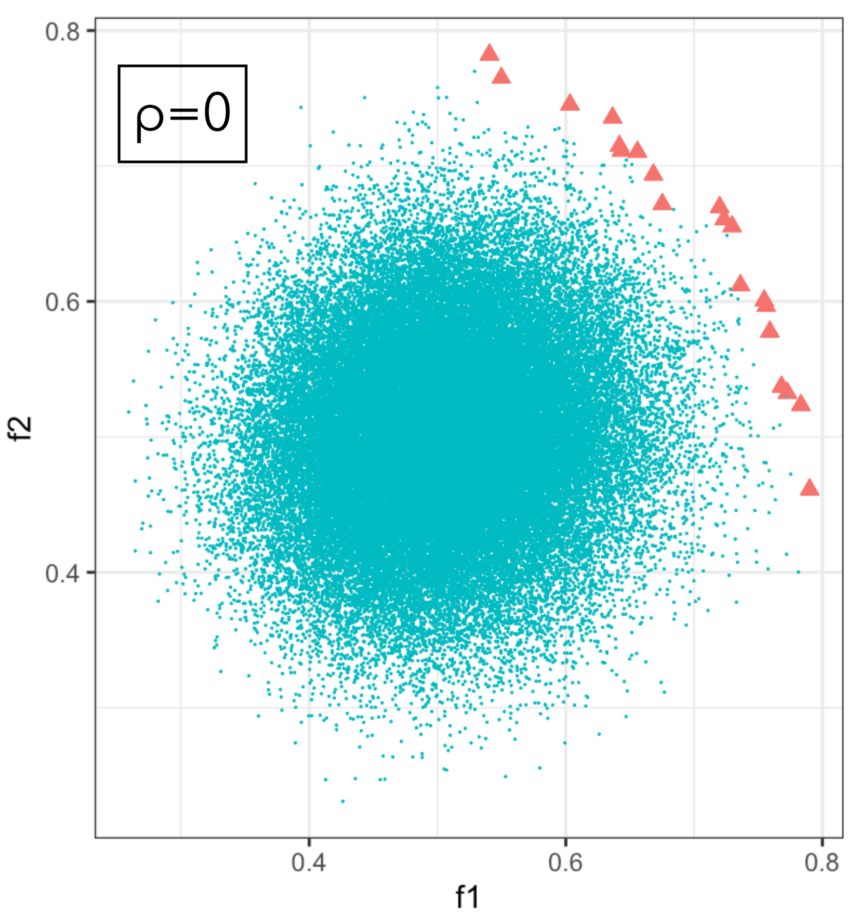
Objectives Interaction



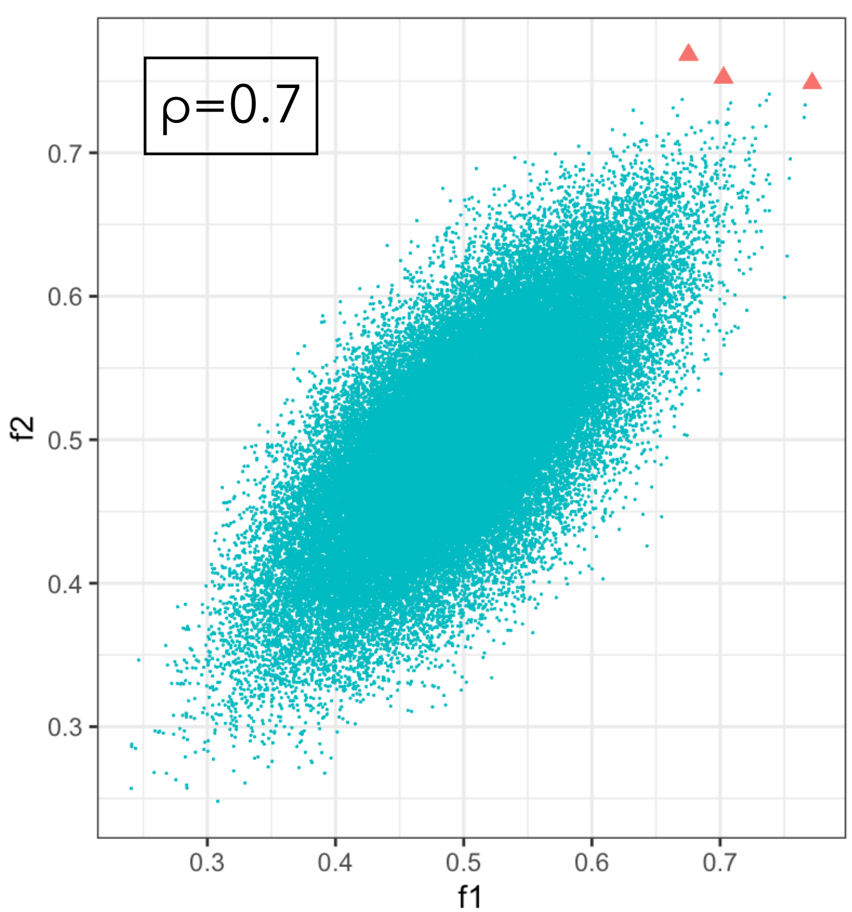
Objective Correlation



conflicting
objectives



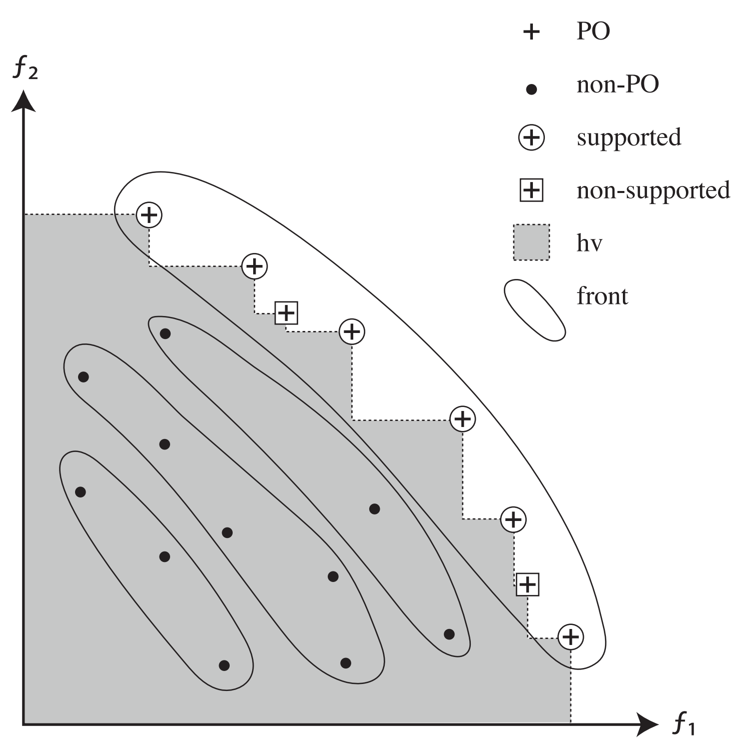
independent
objectives



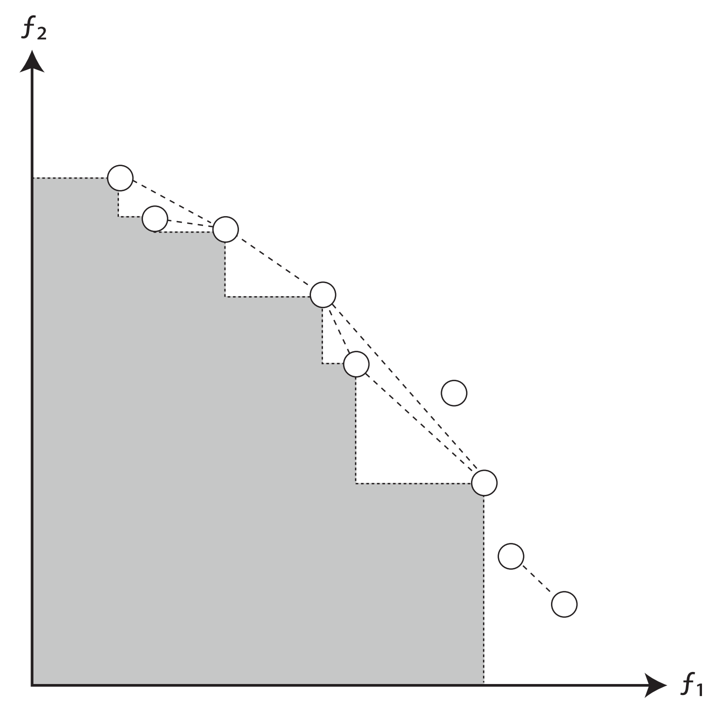
correlated
objectives

enumeration

"Global" Features



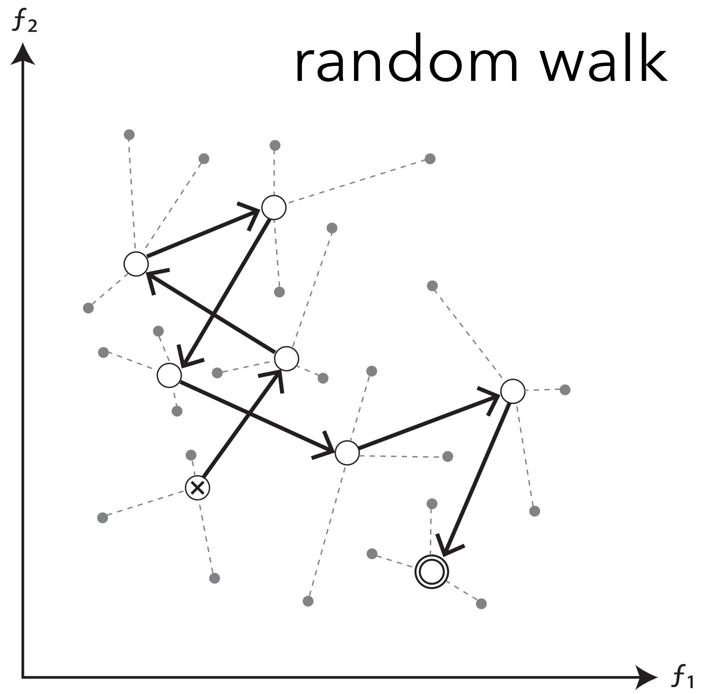
features from **solution space** and **Pareto set**

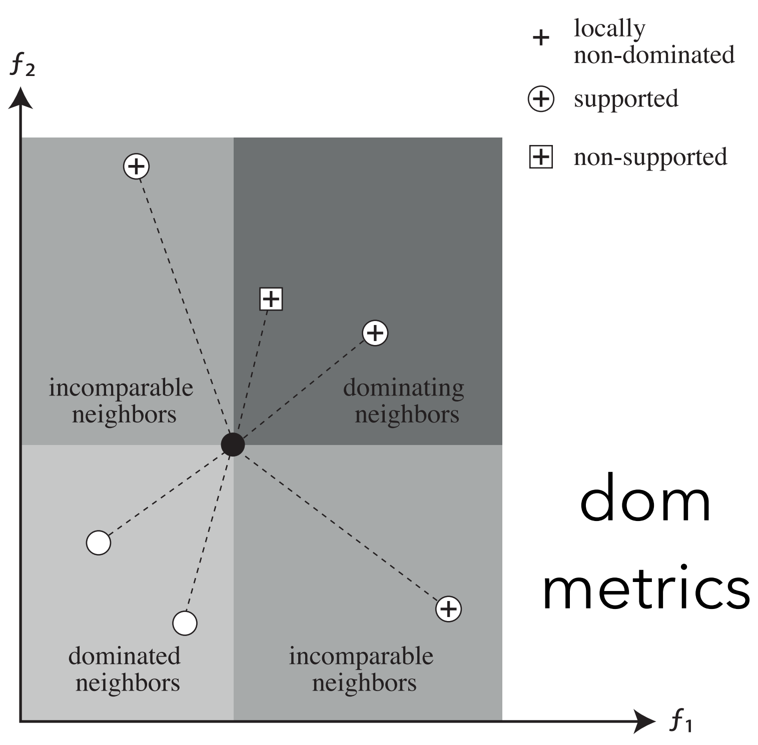


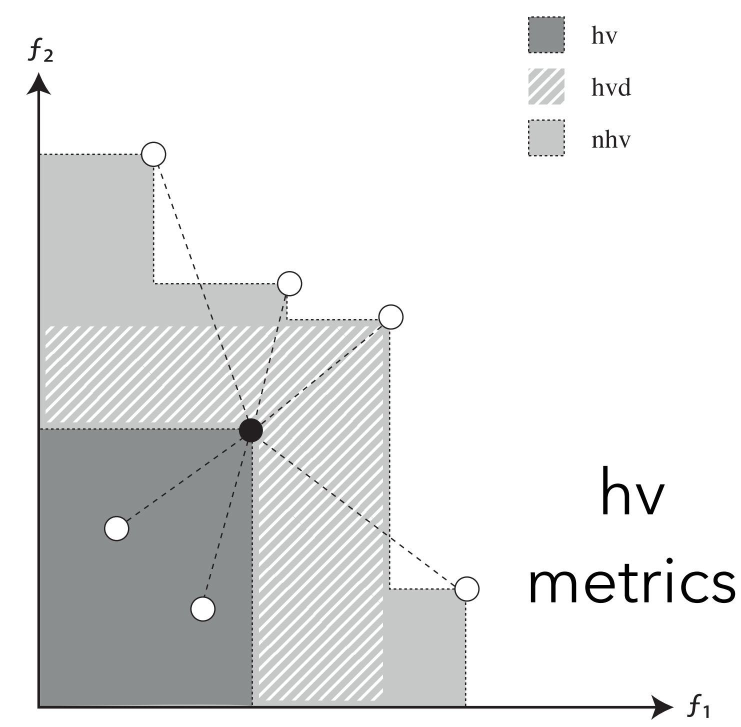
features from **Pareto graph** (connectedness) sampling

Local Features

1. Sampling
   - walk $(x_0, x_1, \ldots, x_\ell)$ s.t. $x_t \in N(x_{t-1})$
2. Measures
   - autocorrelation (ruggedness)
   - average

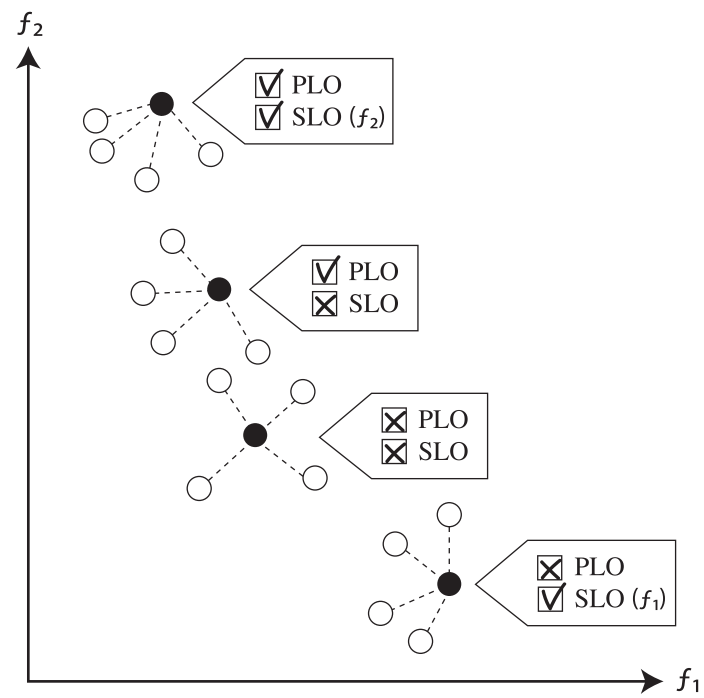






enumeration

Multimodality



multimodality /
local optimality

sampling

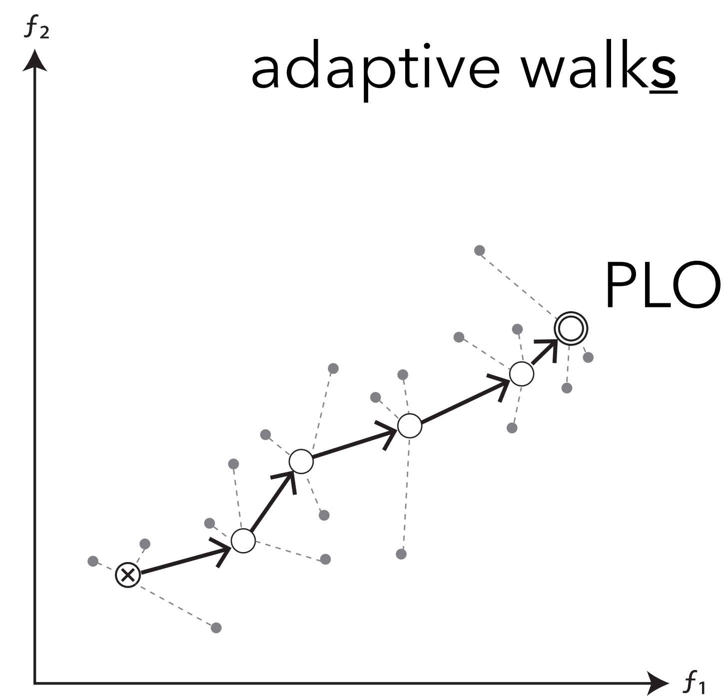
Local Features

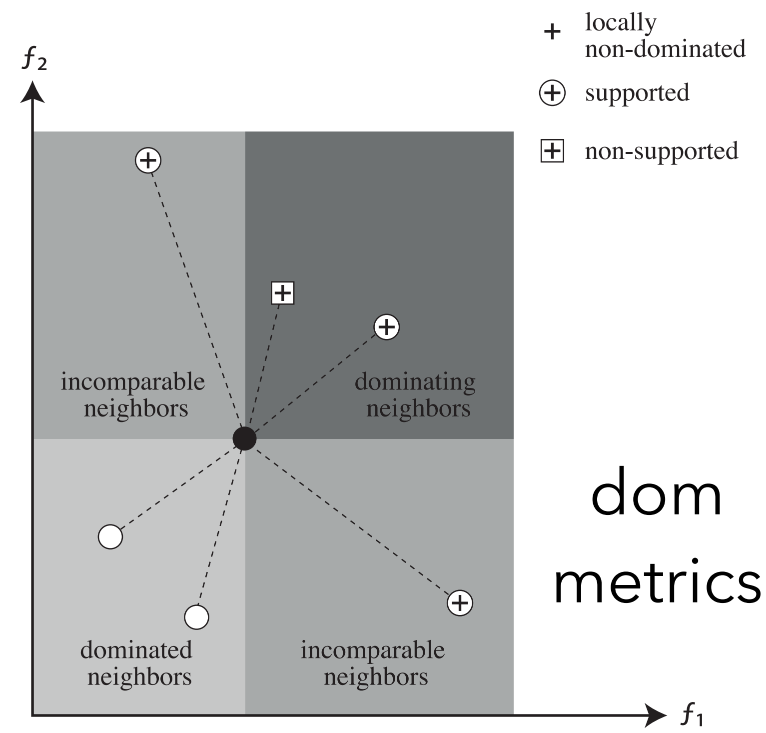
1. Sampling

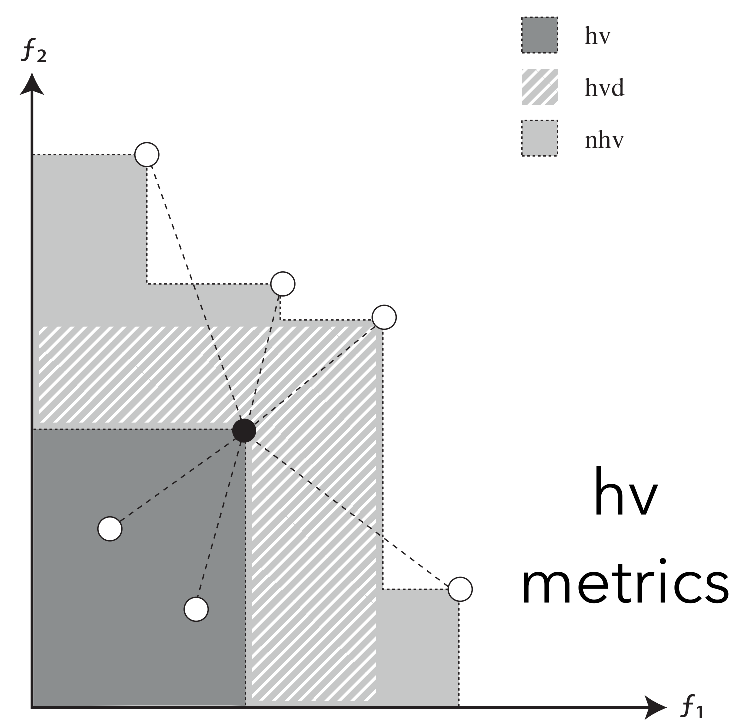
- walk $(x_0, x_1, \ldots, x_\ell)$
  s.t. $x_t \in N(x_{t-1})$
  and x_t dom x_t

2. Measures

- length ℓ
- average







	GLOBAL FEATURES FROM **full enumeration** (16)	
#po	proportion of Pareto optimal (PO) solutions	$\Theta(\lvert X \rvert)$
#supp	proportion of supported solutions in the Pareto set	
hv	hypervolume-value of the (exact) Pareto front	
#plo	proportion of Pareto local optimal (PLO) solutions	
#slo_avg	average proportion of single-objective local optimal solutions per objective	
podist_avg	average Hamming distance between Pareto optimal solutions	
podist_max	maximal Hamming distance between Pareto optimal solutions (diameter of the Pareto set)	
po_ent	entropy of binary variables from Pareto optimal solutions	
fdc	fitness-distance correlation in the Pareto set (Hamming dist. in solution space vs. Manhattan dist. in objective space)	
#cc	proportion of connected components in the Pareto graph	
#sing	proportion of isolated Pareto optimal solutions (singletons) in the Pareto graph	
#lcc	proportional size of the largest connected component in the Pareto graph	
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	
front_ent	entropy of the non-dominated front's size distribution	
	LOCAL FEATURES FROM RANDOM WALK **sampling** (17)	
hv_avg_rws	average (single) solution's hypervolume-value	$\Theta(\ell_{rws} \cdot \#neig)$
hv_r1_rws	first autocorrelation coefficient of (single) solution's hypervolume-values	
hvd_avg_rws	average (single) solution's hypervolume difference-value	
hvd_r1_rws	first autocorrelation coefficient of (single) solution's hypervolume difference-values	
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	$\Theta(\ell_{rws})$
	LOCAL FEATURES FROM ADAPTIVE WALK **sampling** (9)	
hv_avg_aws	average (single) solution's hypervolume-value	$\Theta(n_{aws} \cdot \ell_{aws} \cdot \#neig)$
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	$\Theta(n_{aws} \cdot e_{aws})$

http://mocobench.st.net

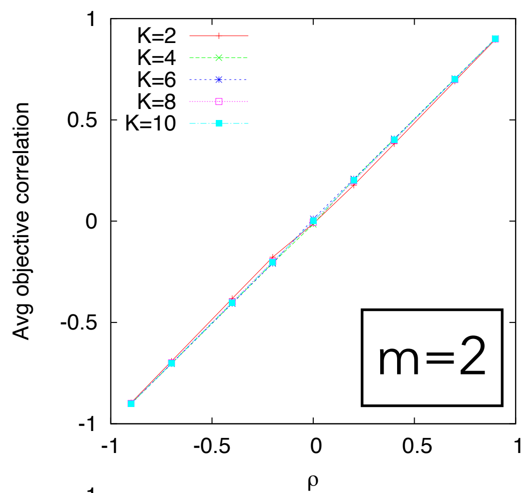
e.g. pmnk Landscapes

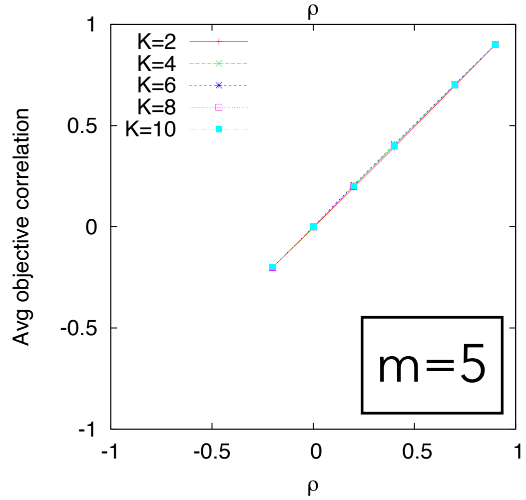
[Verel et al. 2013]

$$\max f_i(x) = \frac{1}{n} \sum_{j=1}^{n} c_j^i(x_j, x_{j_1}, \ldots, x_{j_k}) \qquad i \in \{1, \ldots, m\}$$

$$\textbf{s.t. } x_j \in \{0,1\} \qquad j \in \{1, \ldots, n\}$$

- number of variables n
- variable interactions k < n
- number of objectives m
- objective correlation ρ > −1 / (m−1)





http://mocobench.st.net

e.g. pmnk Landscapes

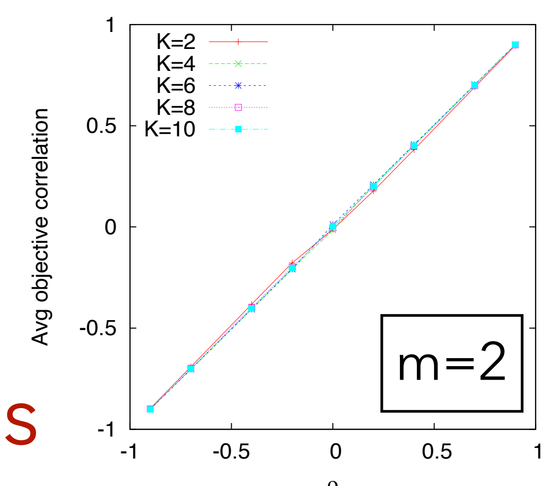
[Verel et al. 2013]

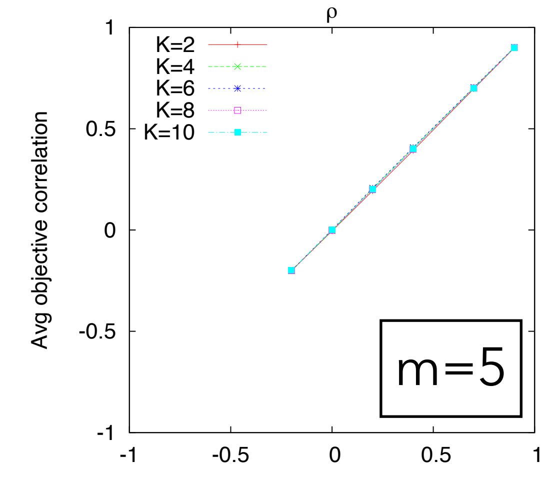
$$\max f_i(x) = \frac{1}{n} \sum_{j=1}^{n} c_j^i (x_j, x_{j_1}, \ldots, x_{j_k}) \qquad i \in \{1, \ldots, m\}$$

$$\textbf{s.t. } x_j \in \{0,1\} \qquad j \in \{1, \ldots, n\}$$

- number of variables n
- variable interactions k < n
- number of objectives m
- objective correlation ρ > −1 / (m−1)

unknown for black-box problems





small landscapes

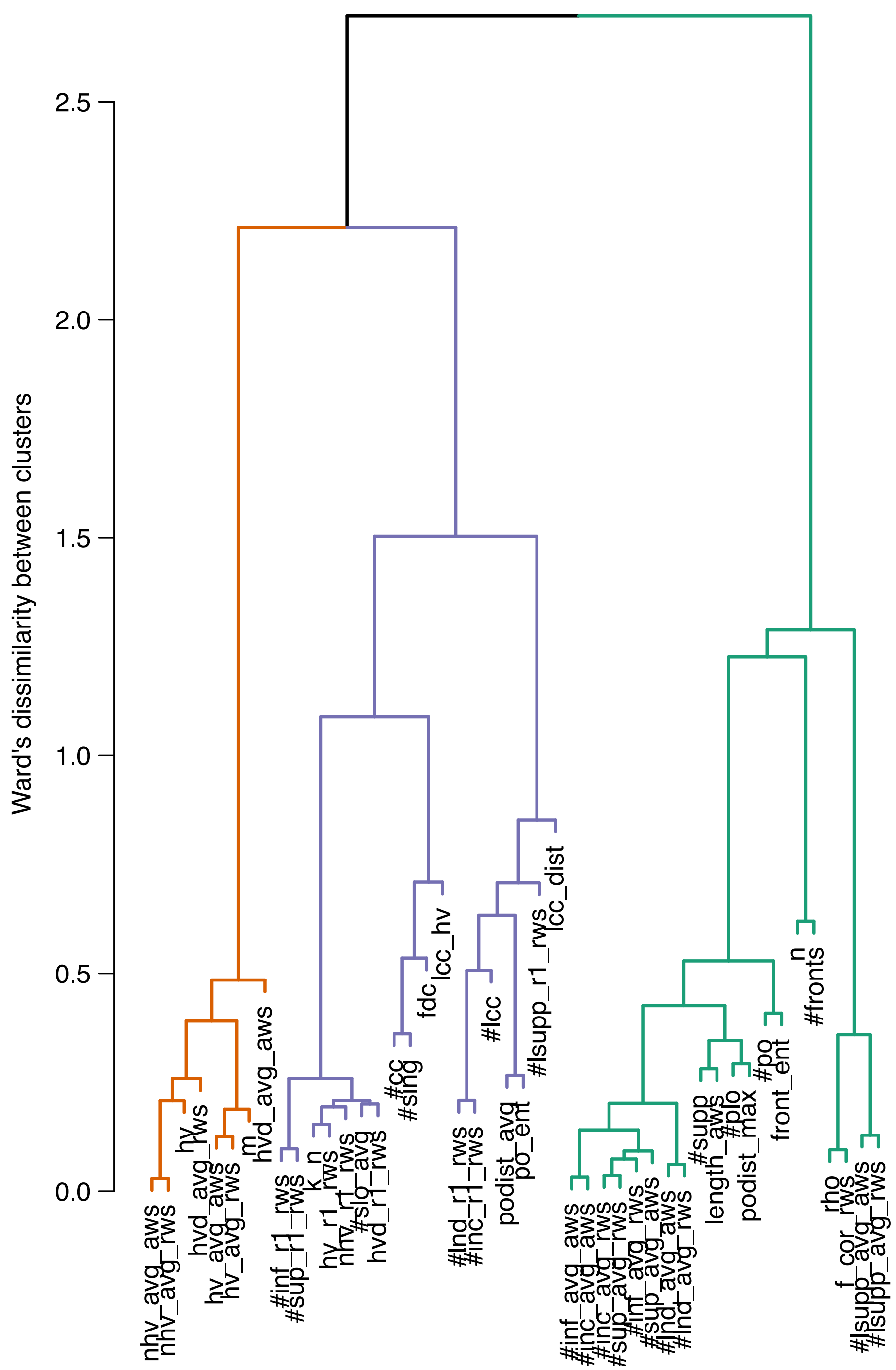
Experimental Setup

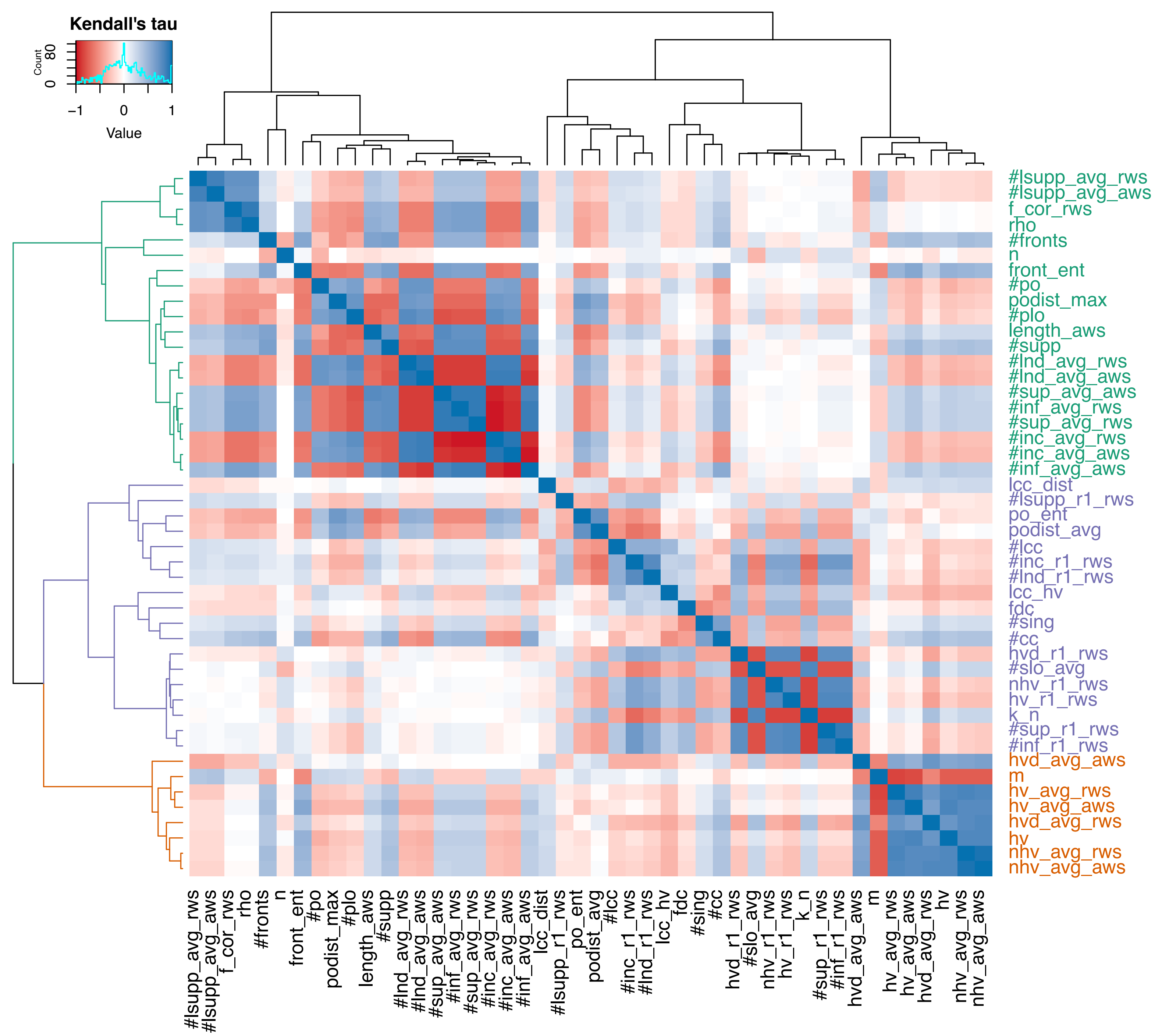
60 480 instances (factorial design, x30 per setting)

- number of variables $n \in \{10, 11, 12, 13, 14, 15, 16\}$
- variable interactions $k \in \{0, 1, 2, 3, 4, 5, 6, 7, 8\}$
- number of objectives $m \in \{2, 3, 4, 5\}$
- objective correlation $\rho > -1/(m-1)$
  $\rho \in \{-0.8, -0.6, -0.4, -0.2, 0, 0.2, 0.4, 0.6, 0.8, 1\}$

small landscapes

Pairwise Feature Correlation





large landscapes

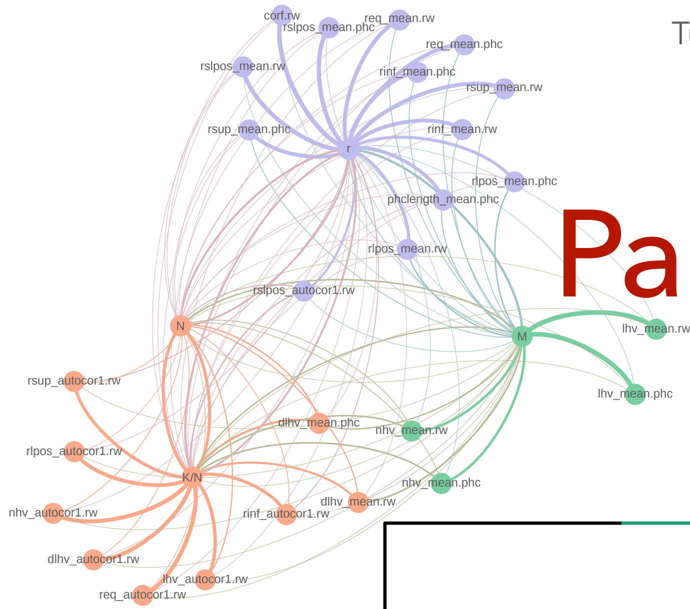
Experimental Setup

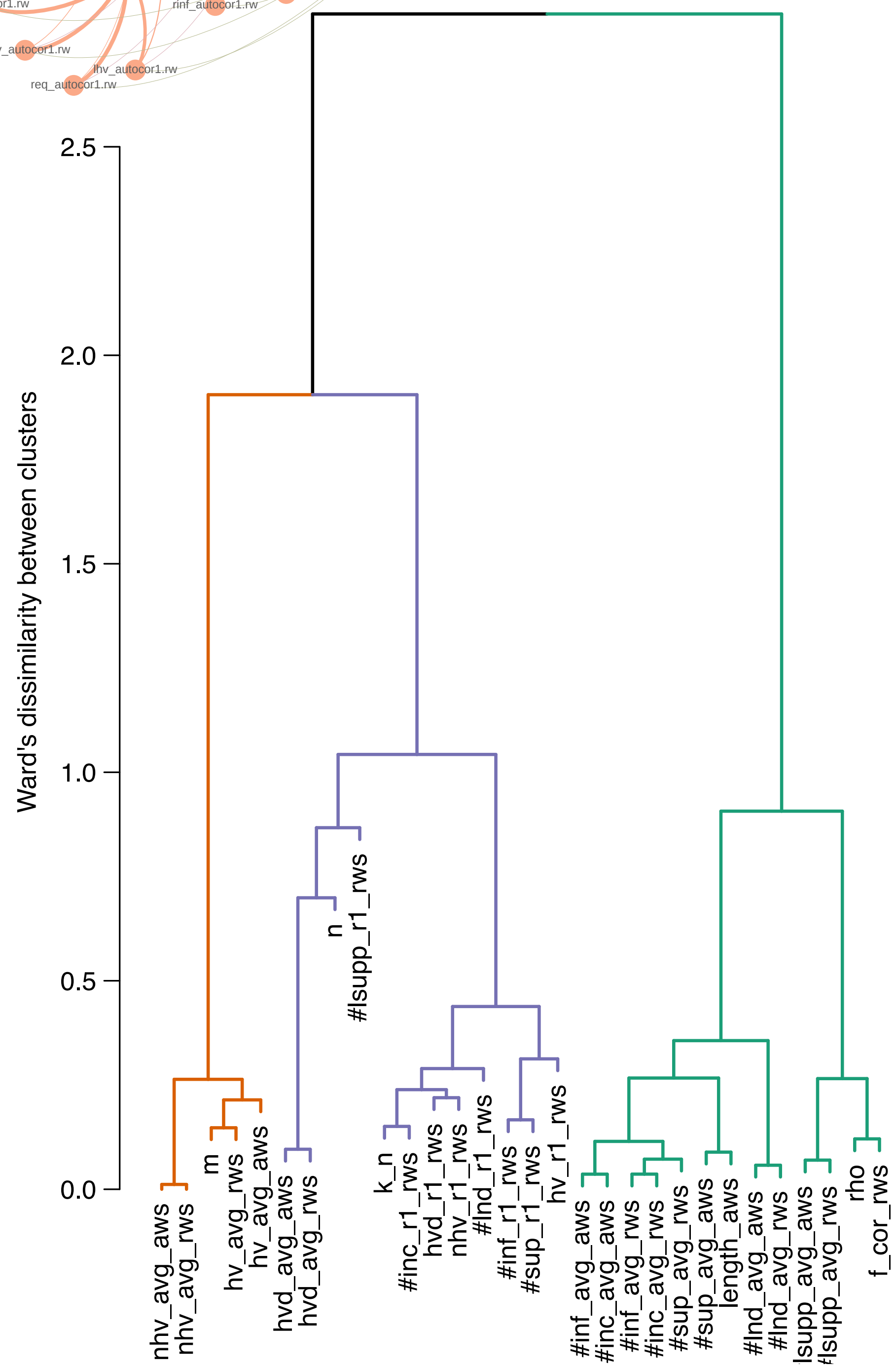
1 000 landscapes (design of experiments)

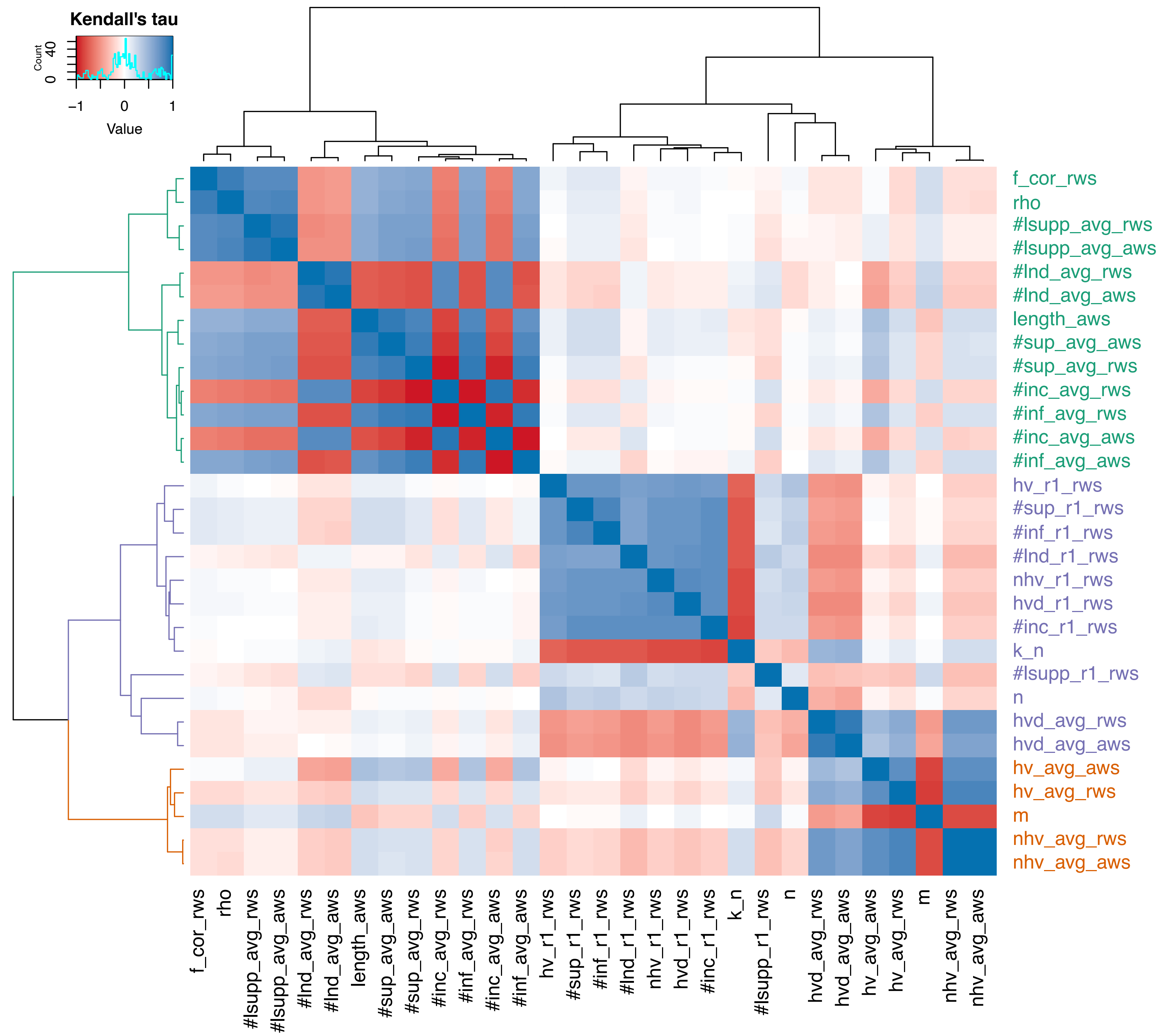
- number of variables $n \in [\![64, 256]\!]$
- variable interactions $k \in [\![0, 8]\!]$
- number of objectives $m \in [\![2, 5]\!]$
- objective correlation $\rho \in [-1/(m-1), 1]$

large landscapes

Pairwise Feature Correlation







large landscapes

Experimental Setup

Algorithms

- **Evolutionary search** (**G-SEMO**) vs. **Local search** (iterated PLS = **I-PLS**)

Performance

- 30 independent runs per instance, fixed budget of 100 000 evaluations
- (Expected) epsilon approximation ratio to best non-dominated set

Statistics

- Regression = extremely randomized trees (RF variant)

large landscapes

Prediction Accuracy

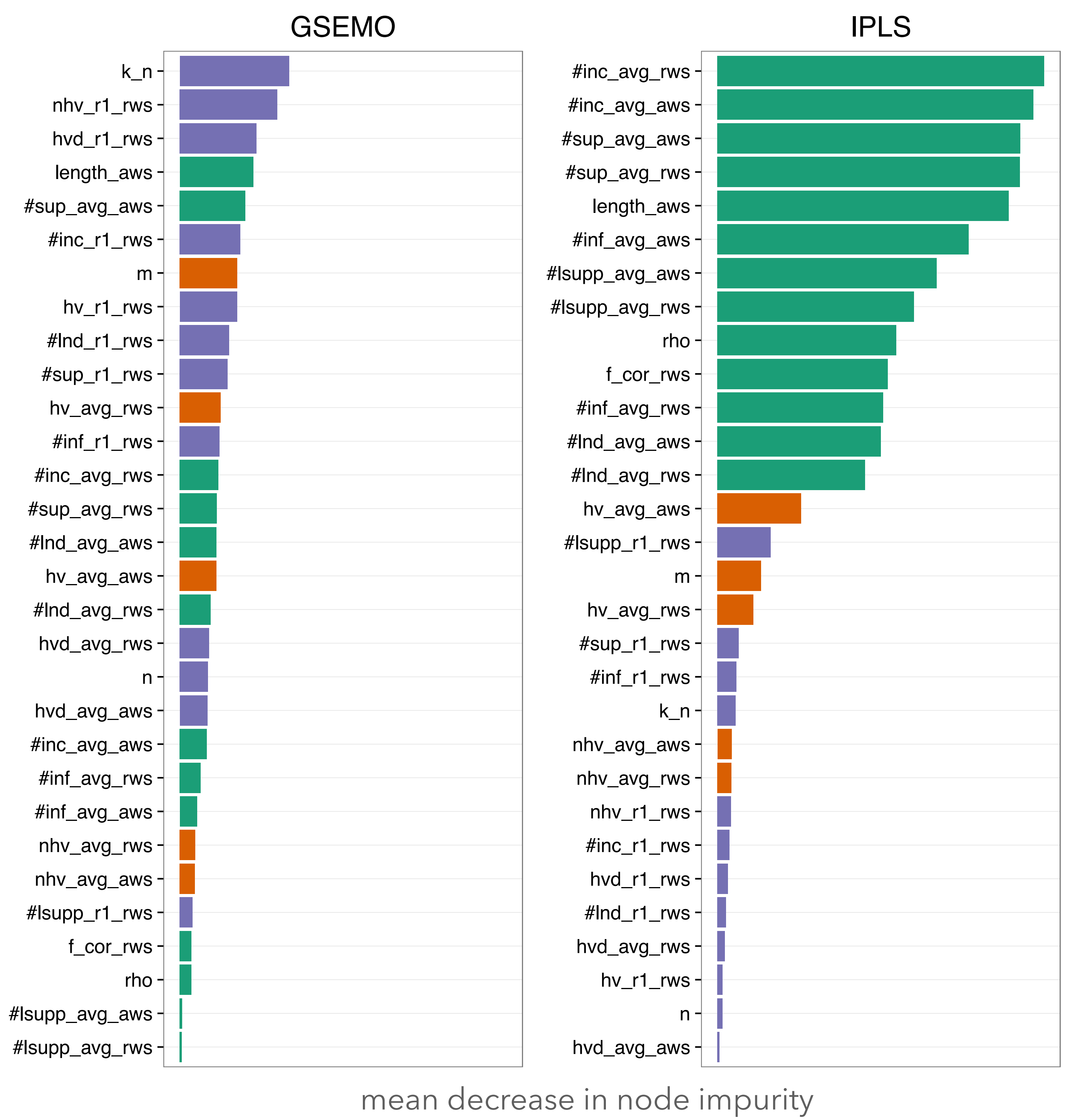
algo.	set of features	MAE		MSE		R^2		adjusted R^2		rank
		avg	std	avg	std	avg	std	avg	std	
G-SEMO	all features	**0.003049**	0.000285	**0.000017**	0.000004	0.891227	0.024584	0.843934	0.035273	1
	local features	**0.003152**	0.000295	**0.000018**	0.000004	0.883909	0.026863	0.838126	0.037457	1
	local features (random walk)	**0.003220**	0.000314	0.000019	0.000004	0.878212	0.028956	0.849287	0.035833	1.5
	local features (adaptive walk)	0.003525	0.000329	0.000023	0.000006	0.854199	0.032339	0.834089	0.036799	5
	$\{\rho, \mathtt{m}, \mathtt{n}, \mathtt{k\_n}\}$	**0.003084**	0.000270	**0.000017**	0.000003	0.892947	0.020658	0.888440	0.021528	1
	$\{\mathtt{m}, \mathtt{n}\}$	0.010813	0.000830	0.000206	0.000030	-0.303336	0.188046	-0.330209	0.191923	6
I-PLS	all features	**0.004290**	0.000430	**0.000034**	0.000008	0.886568	0.026980	0.837249	0.038710	1
	local features	**0.004359**	0.000423	**0.000035**	0.000008	0.883323	0.027274	0.837309	0.038030	1
	local features (random walk)	**0.004449**	0.000394	**0.000036**	0.000008	0.879936	0.026335	0.851421	0.032589	1
	local features (adaptive walk)	0.004663	0.000403	0.000039	0.000008	0.871011	0.025903	0.853219	0.029476	3.5
	$\{\rho, \mathtt{m}, \mathtt{n}, \mathtt{k\_n}\}$	**0.004353**	0.000320	**0.000033**	0.000006	0.889872	0.024505	0.885235	0.025537	1
	$\{\mathtt{m}, \mathtt{n}\}$	0.016959	0.001473	0.000472	0.000077	-0.568495	0.228629	-0.600836	0.233343	6

random subsampling cross-validation
(50 iterations, 90/10 split)

error < 1% $R^2 > 0.8$

large landscapes

Importance of Features



large landscapes

Experimental Setup

Algorithms

- **NSGA-II** vs. **IBEA** vs. **MOEA/D** (default setting, population size = 100)

Performance

- 20 independent runs per instance, 1 000 000 evaluations
- (Expected) hypervolume relative deviation (hvrd)

Statistics

- Classification = extremely randomized trees, decision tree

large landscapes

Automated Algorithm Selection

Algorithm portfolio = {NSGA-II, IBEA, MOEA/D}

Model (classif, RF) = {algo} ~ (n, k/n, m, ρ, {features})

set of features	error rate of best average performance mean	std	rank	error rate of best statistical rank mean	std	rank
all features	**0.122222**	0.031033	1	**0.012727**	0.014110	1
local features	**0.123030**	0.030521	1	**0.013737**	0.014103	1
local features (random walk)	**0.118788**	0.029187	1	**0.013333**	0.012149	1
local features (adaptive walk)	**0.130303**	0.029308	1	**0.015354**	0.014026	1
{ρ, m, n, k_n}	**0.125859**	0.028875	1	**0.014141**	0.013382	1
{m, n}	0.413333	0.045533	6	0.197374	0.043778	6

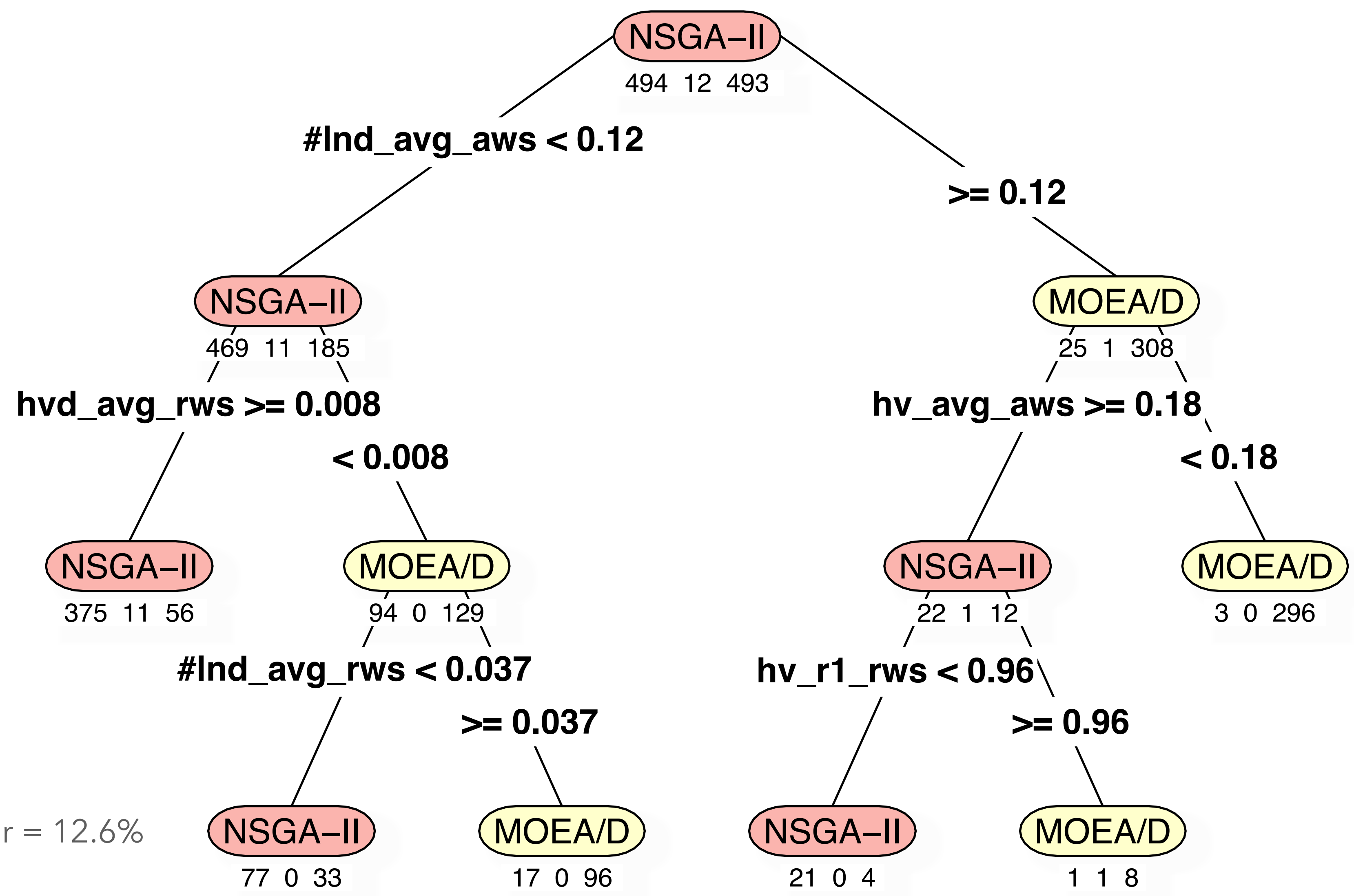
random subsampling cross-validation
(50 iterations, 90/10 split)

avg-best > 85% stat-best > 98%

large landscapes

Automated Algorithm Selection

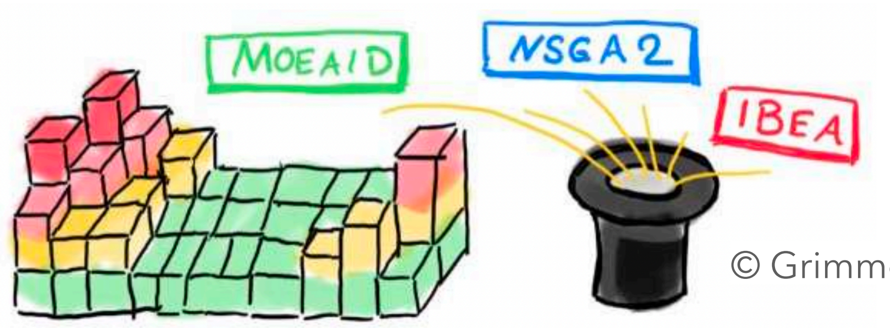
Model (classif, decision tree) = {algo} ~ (n, k/n, m, ρ, {features})



classification error = 12.6%

large mQAP

Automated Algorithm Selection



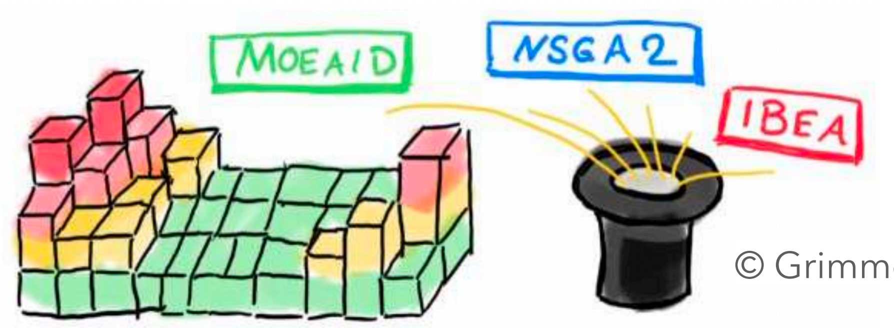
© Grimme

Algorithm portfolio = {NSGA-II, IBEA, MOEA/D}

Model (classif, RF) = {algo} ~ (n, m, ρ, type, {features})

large mQAP

Automated Algorithm Selection



© Grimme

Algorithm portfolio = {NSGA-II, IBEA, MOEA/D}

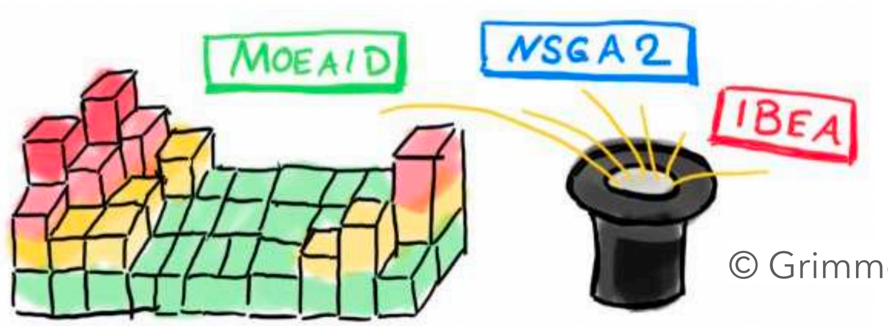
Model (classif, RF) = {algo} ~ (n, m, ρ, type, {features})

subset of features	**classification error**	**error predicting statistical best**
{n, m}	.1962	.0332
{type, n, m, ρ}	.1197	.0072
{⋆_rws, n, m}	.1114	**.0062**
{⋆_aws, n, m}	.1125	.0065
{⋆_rws, length_aws, n, m}	**.1089**	**.0056**
{⋆_rws, ⋆_aws, n, m}	**.1077**	**.0063**
{⋆_rws, ⋆_aws, type, n, m, ρ}	**.1078**	**.0063**
random classifier	.6667	.3810
dummy classifier (MOEA/D)	.4200	.1040

random subsampling cross-validation
(100 repetitions, 80/20% split)

large mQAP

Automated Algorithm Selection



© Grimme

Algorithm portfolio = {NSGA-II, IBEA, MOEA/D}

Model (classif, RF) = {algo} ~ (n, m, ρ, type, {features})

subset of features	classification error	error predicting statistical best
{n, m}	.1962	.0332
{type, n, m, ρ}	.1197	.0072
{⋆_rws, n, m}	.1114	**.0062**
{⋆_aws, n, m}	.1125	.0065
{⋆_rws, length_aws, n, m}	**.1089**	**.0056**
{⋆_rws, ⋆_aws, n, m}	**.1077**	**.0063**
{⋆_rws, ⋆_aws, type, n, m, ρ}	**.1078**	**.0063**
random classifier	.6667	.3810
dummy classifier (MOEA/D)	.4200	.1040

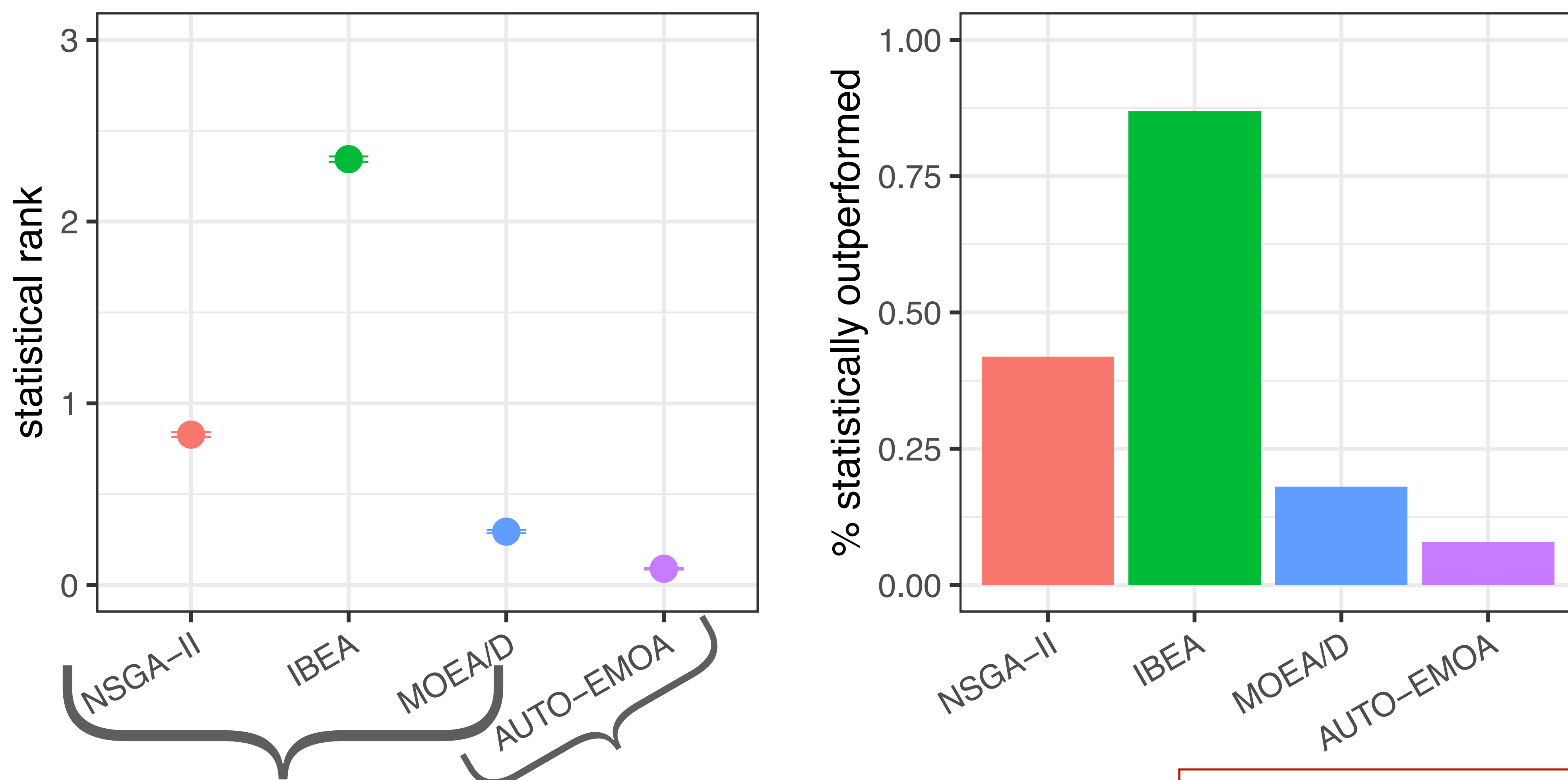
random subsampling cross-validation
(100 repetitions, 80/20% split)

avg-best > 89% stat-best > 99%

large mQAP

Automated Algorithm Selection

(low-cost) features extracted from search budget



100% · 10^6 eval. for search

~5% · 10^6 eval. for features
~95% · 10^6 eval. for search

hv-dev < 3.5%
stat-best > 92%

Contents

Multi-objective Optimization

Foundations of MO Landscapes

Set- and Indicator-based Search

A Glimpse on related Research Directions

Manuel López-Ibánez

Luís Paquete

Sébastien Verel

Multi-objective Optimization

Foundations of **MO Landscapes**

Set- and **Indicator-** based Search

A Glimpse on related **Research Directions**

Sets and Indicators

- Multi-objective optimization is a set problem [Zitzler et al. 2010]
  - Seeking the best set of solutions e.g. $\arg\max_{X' \subseteq X} hv(X')$
  - (Evolutionary) multi-objective algorithms are (local) search heuristics performing on sets
- How to compare sets?
  - Same as for performance evaluation (benchmarking)
  - Set preference relation e.g. set dominance, quality indicator
- How does the set preference relation impacts search difficulty?

Research Questions

Is it harder for a multi-objective local search to find a good approximation set for few or for many objectives?

Is it harder for a multi-objective local search to find a good approximation set with few or with many solutions?

Is it harder for a multi-objective local search to find a good approximation set in terms of dominance or indicator?

Set-based Multi-objective Landscape

- The search space $\Sigma \subset 2^X$ is a collection of sets

  e.g. sets of mutually non-dominated solutions with a cardinality bound μ

- The neighborhood $N : \Sigma \rightarrow 2^\Sigma$ is a relation between sets

  e.g. two sets are neighbors if they differ by one (neighboring) solution

- The set preference relation $\preccurlyeq$ is a pre-order between sets

$$A \preccurlyeq B \wedge \neg(B \preccurlyeq A) \iff A \prec B$$

Set Preference Relations

[Zitzler et al. 2010]

- (weak) Set dominance relation

$$A \preccurlyeq_{dom} B \iff \forall b \in B, \exists a \in A \textbf{ s.t. } a \preccurlyeq_{dom} b$$

- Quality indicators

$$A \preccurlyeq_{dom} B \implies I_{eps}(A) \leqslant I_{eps}(B)$$

$$A \prec_{dom} B \implies I_{hv}(A) > I_{hv}(B)$$

- Indicator-based preference relations

$$A \preccurlyeq_{eps} B \iff I_{eps}(A) \leqslant I_{eps}(B)$$

$$A \preccurlyeq_{hv} B \iff I_{hv}(A) \geqslant I_{hv}(B)$$

Set-based Local Optimality

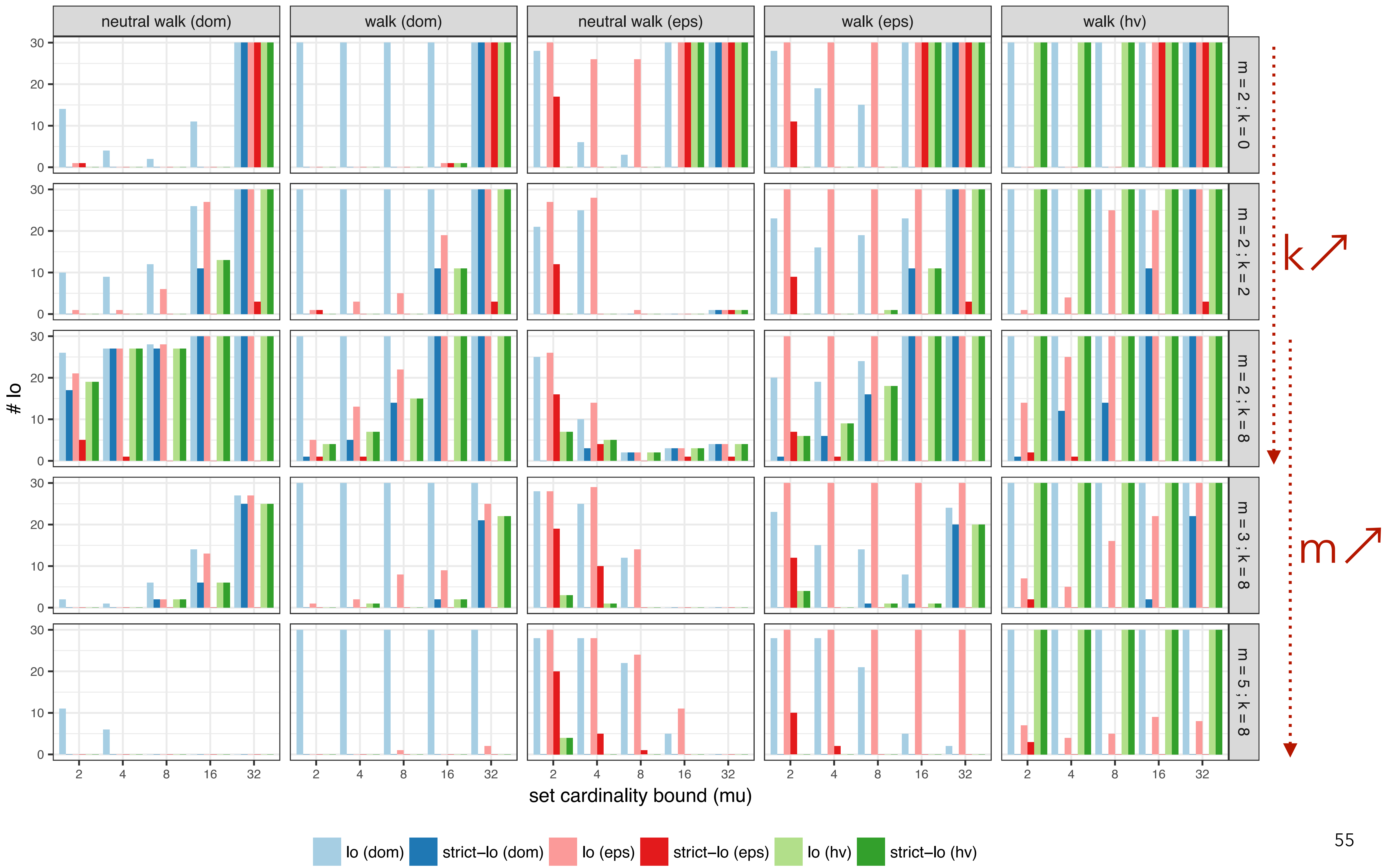
- A set A ∈ Σ is a **local optimal set** (LO-set) iff

$$\forall B \in N(A) \backslash A \; , \; \neg(B \preccurlyeq A)$$

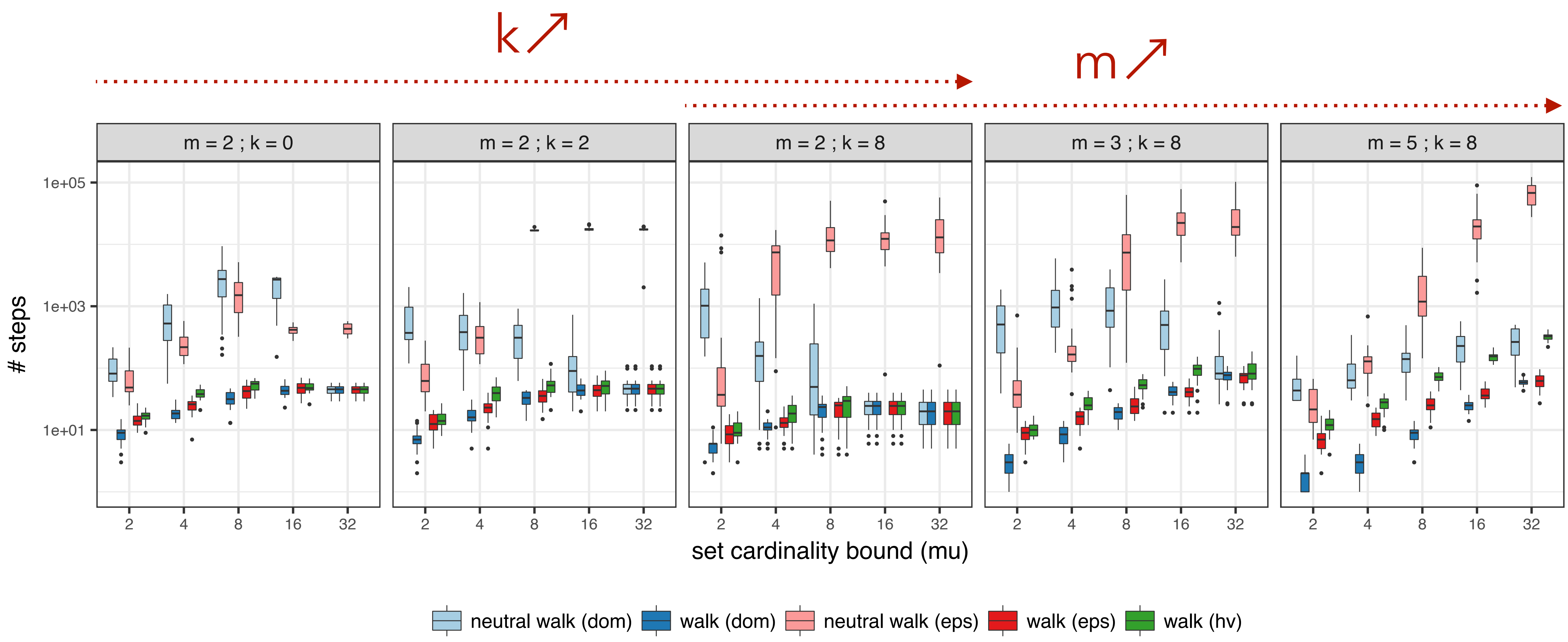
- A set A ∈ Σ is a **strict local optimal set** (sLO-set) iff

$$\forall B \in N(A) \backslash A \; , \; A \prec B$$

Number of (s)LO-sets (Adaptive Walks)



Length of Adaptive Walks



Research Questions

Is it harder for a multi-objective local search to find a good approximation set for **few** or for **many objectives**?

Is it harder for a multi-objective local search to find a good approximation set with **few** or with **many solutions**?

Is it harder for a multi-objective local search to find a good approximation set in terms of **dominance** or **indicator**?

Research Questions

Is it harder for a multi-objective local search to find a good approximation set for **few** or for **many objectives**?

- set-based landscapes with **fewer objectives** are <u>more</u> **multimodal**

Is it harder for a multi-objective local search to find a good approximation set with **few** or with **many solutions**?

- set-based landscapes with **fewer solutions** are <u>more</u> **multimodal**

Is it harder for a multi-objective local search to find a good approximation set in terms of **dominance** or **indicator**?

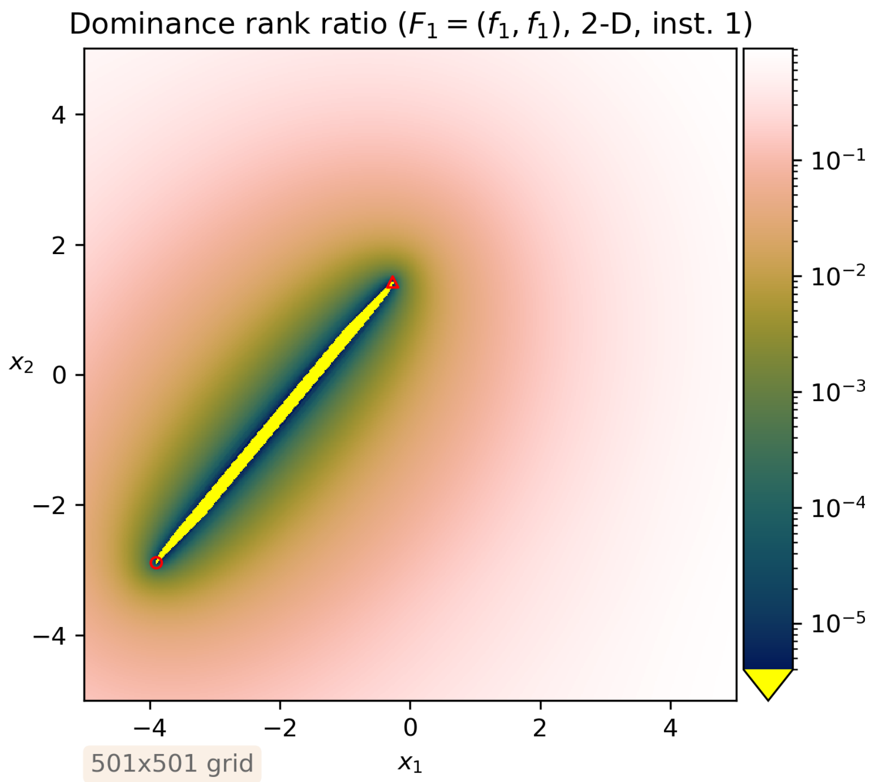
- set-based landscapes under **dominance** are <u>more</u> **multimodal**
  … but they are <u>more</u> "strictly" multimodal under indicators

Contents

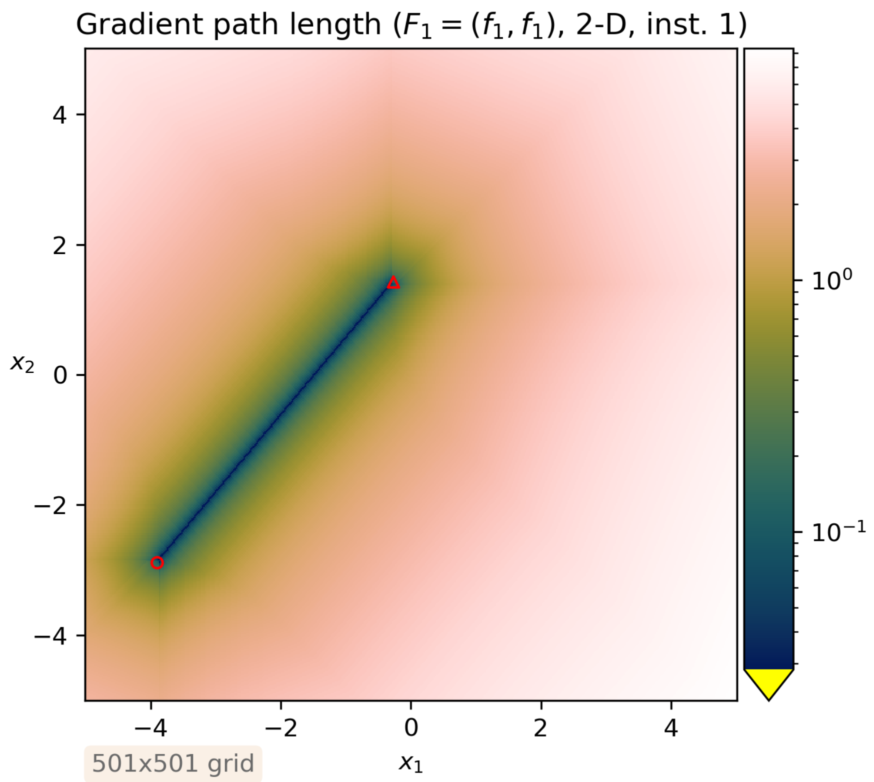
Visualizing Multi-objective Landscapes

Visualizing Multi-objective Landscapes

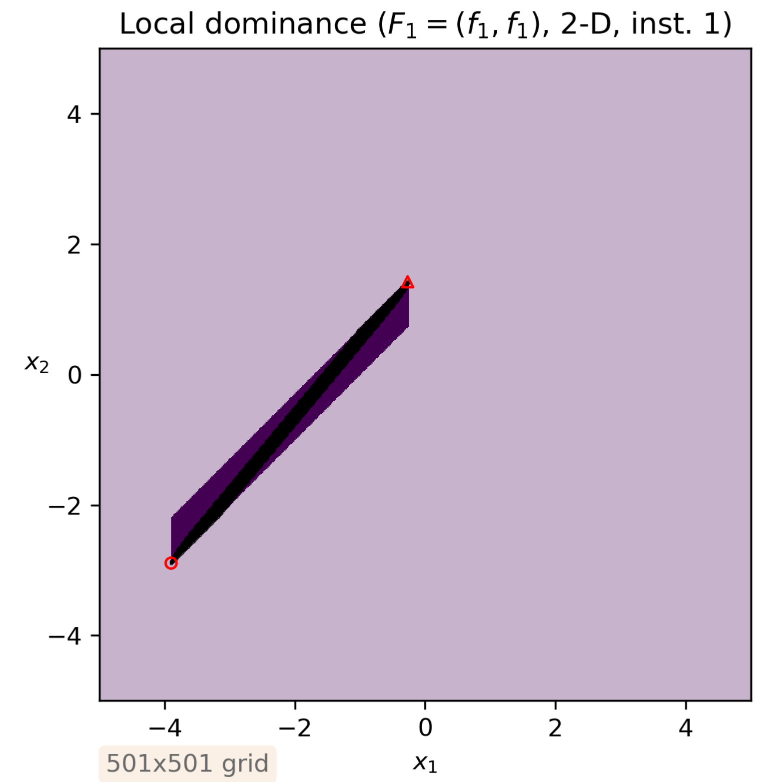
dominance ratio
[Fonseca 1995]



gradient path length
[Kerschke, Grimme 2017]



local dominance
[Fieldsend et al. 2019]

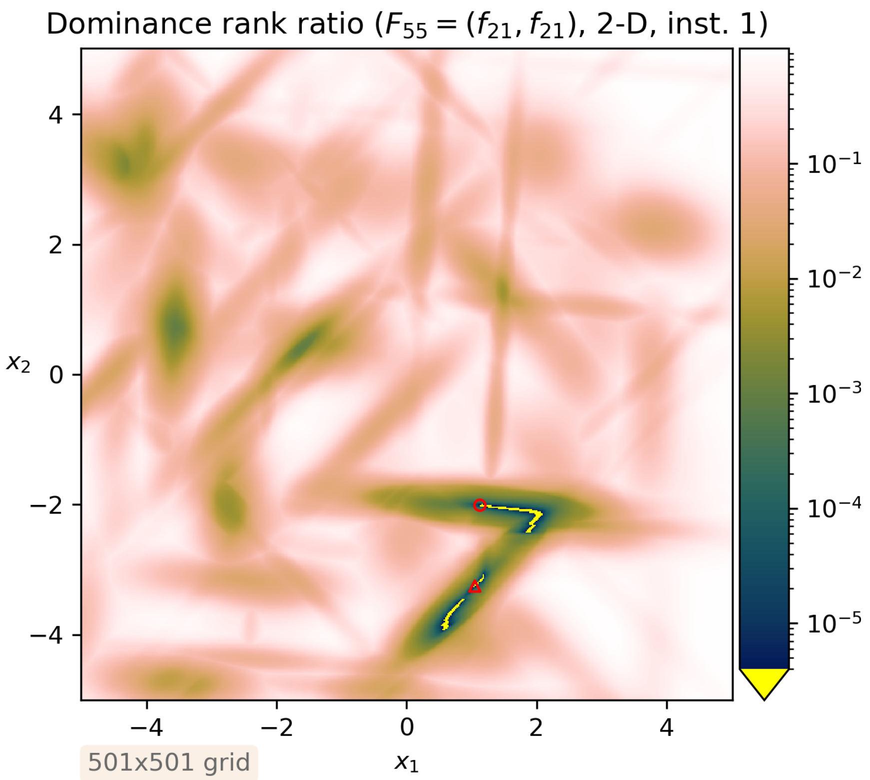


F1: Sphere/Sphere
Dimension 2

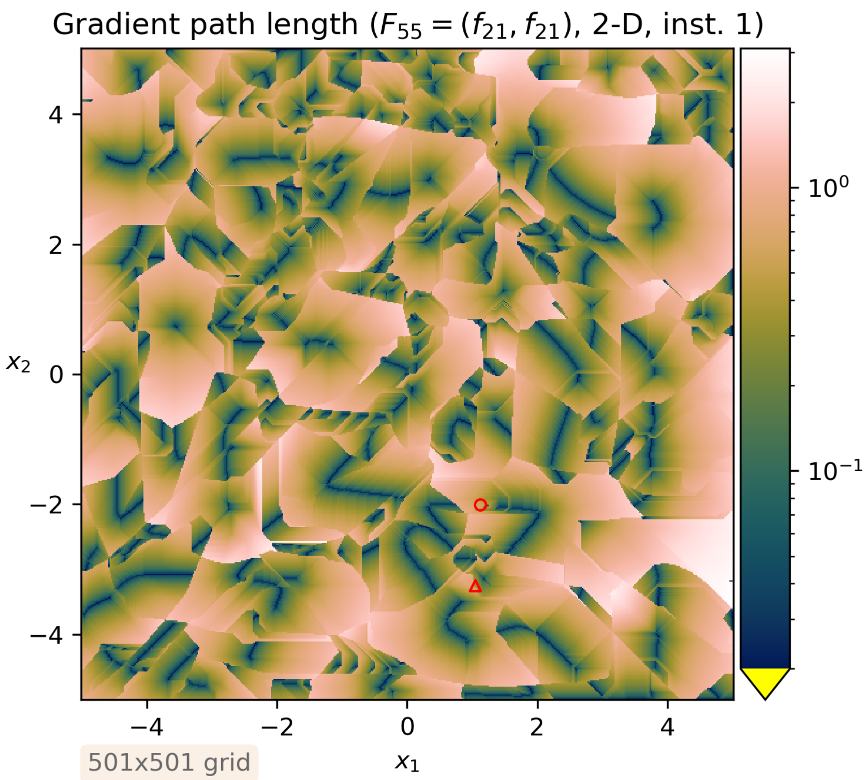
https://numbbo.github.io/bbob-biobj/

Visualizing Multi-objective Landscapes

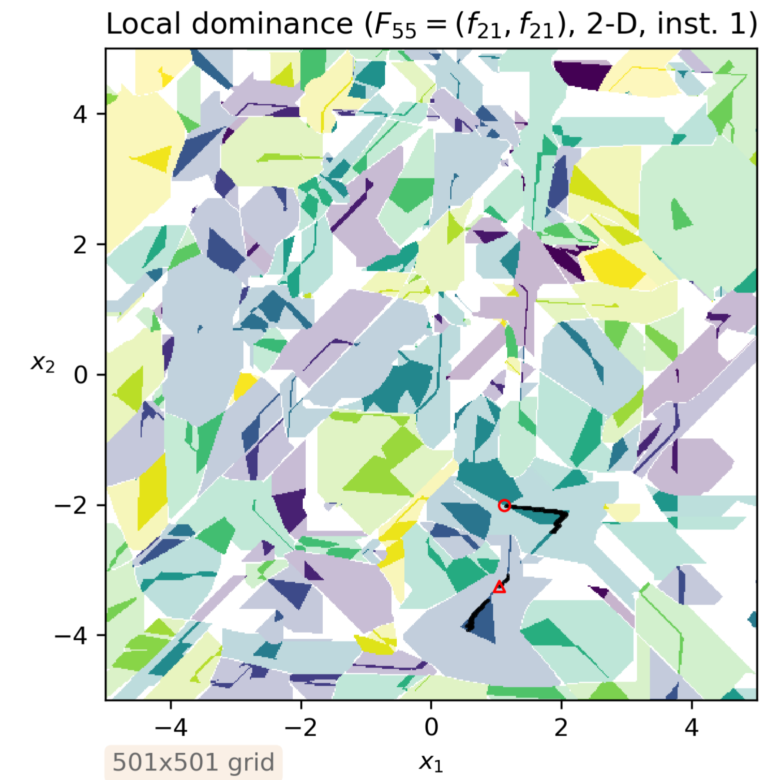
dominance ratio
[Fonseca 1995]



gradient path length
[Kerschke, Grimme 2017]



local dominance
[Fieldsend et al. 2019]



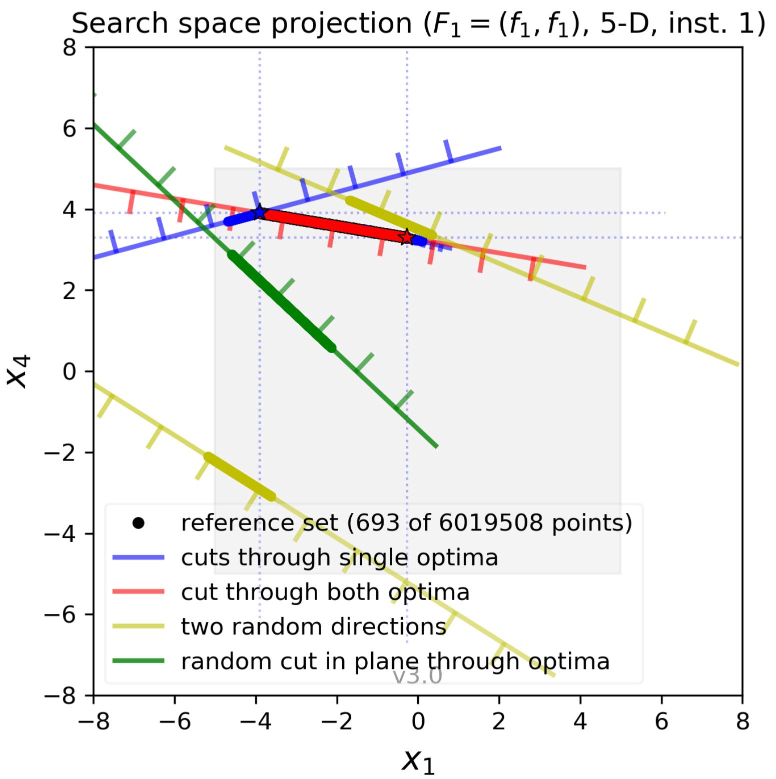
F55: Gallagher 101 peaks/Gallagher 101 peaks
Dimension 2

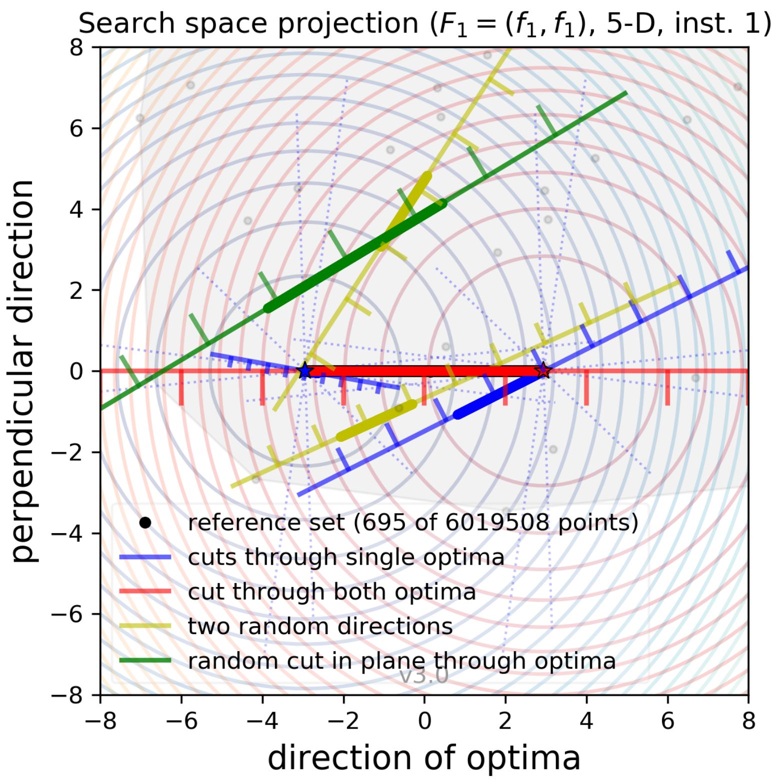
https://numbbo.github.io/bbob-biobj/

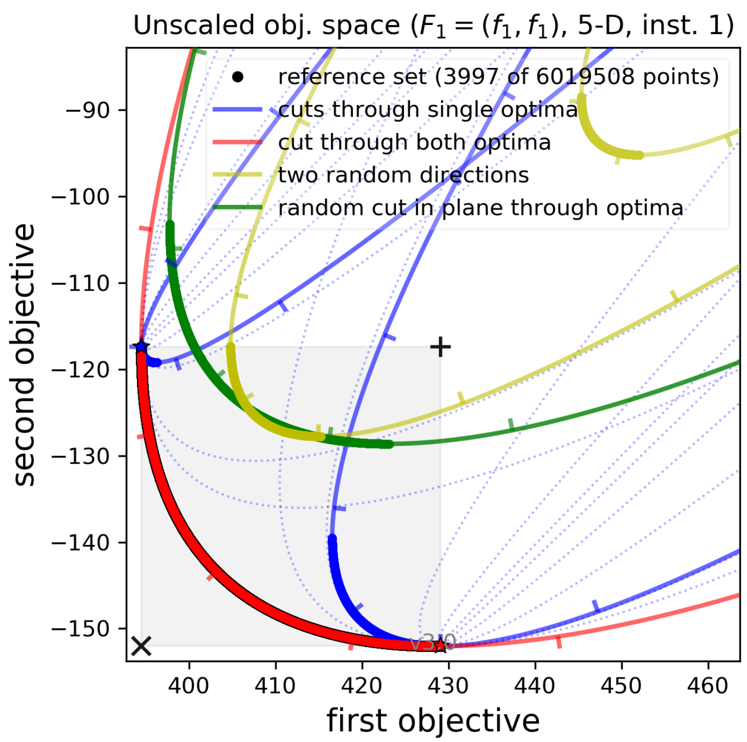
Visualizing Multi-objective Landscapes

line cuts

[Brockhoff et al. 2022]







F1: Sphere/Sphere

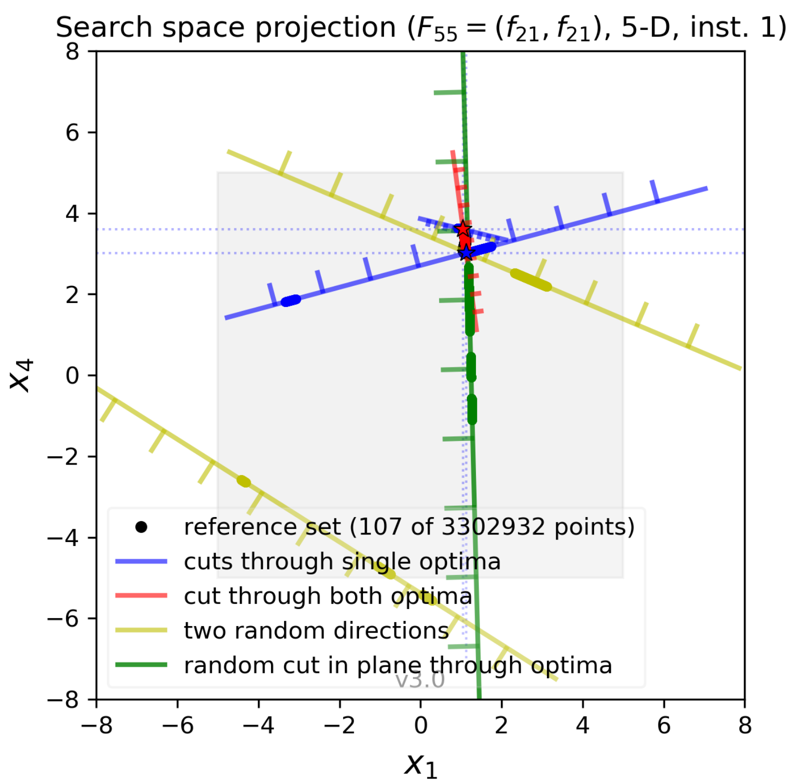
Dimension 5

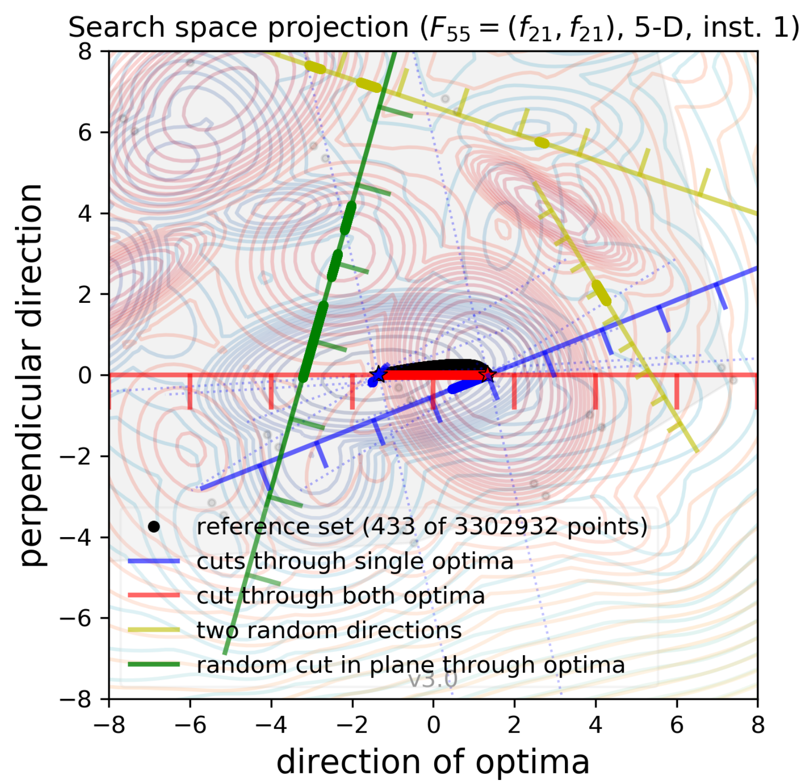
https://numbbo.github.io/bbob-biobj/

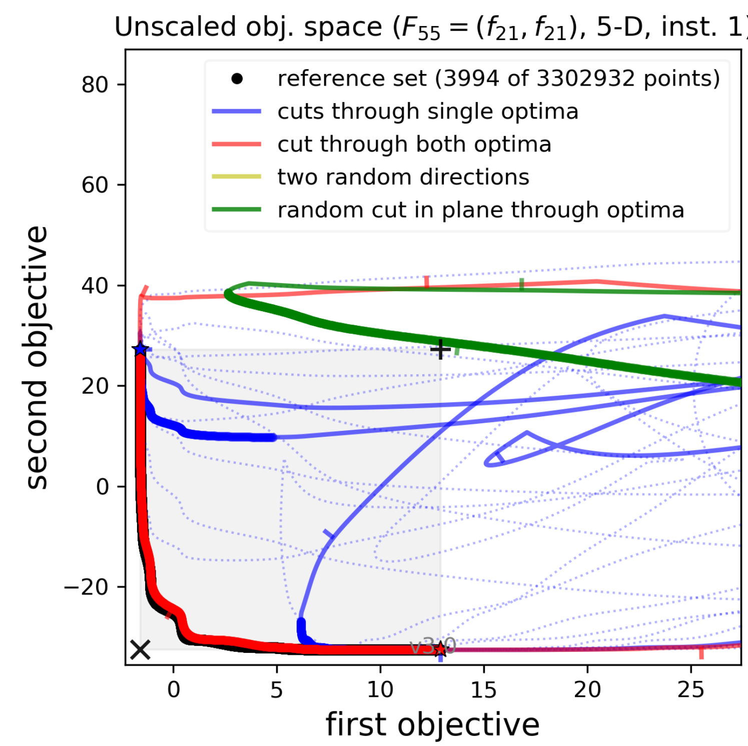
Visualizing Multi-objective Landscapes

line cuts

[Brockhoff et al. 2022]







F1: Sphere/Sphere

Dimension 5

https://numbbo.github.io/bbob-biobj/

COCO / bbob-biobj

https://numbbo.github.io/bbob-biobj/

HOME COCO CODE DATA ARCHIVE POSTPROCESSED DATA COCO HOME

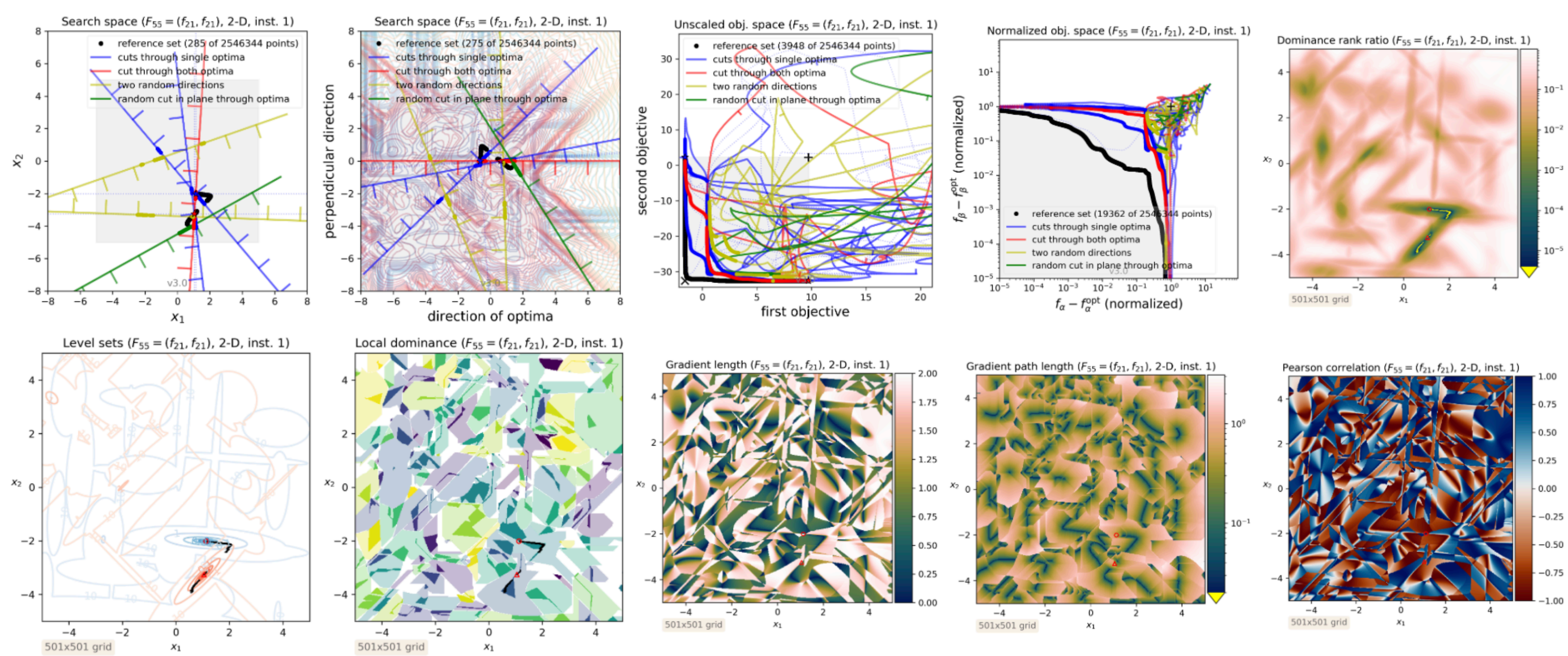
- Home
- Function definitions
- Visualizations
- Gradient angle plots
- Postprocessed data

Visualizations of problem landscapes

Plots

Show plots in – 5 + columns (click on **Dimension/Function/Instance/Visualization type** below to show all plots for the chosen category)

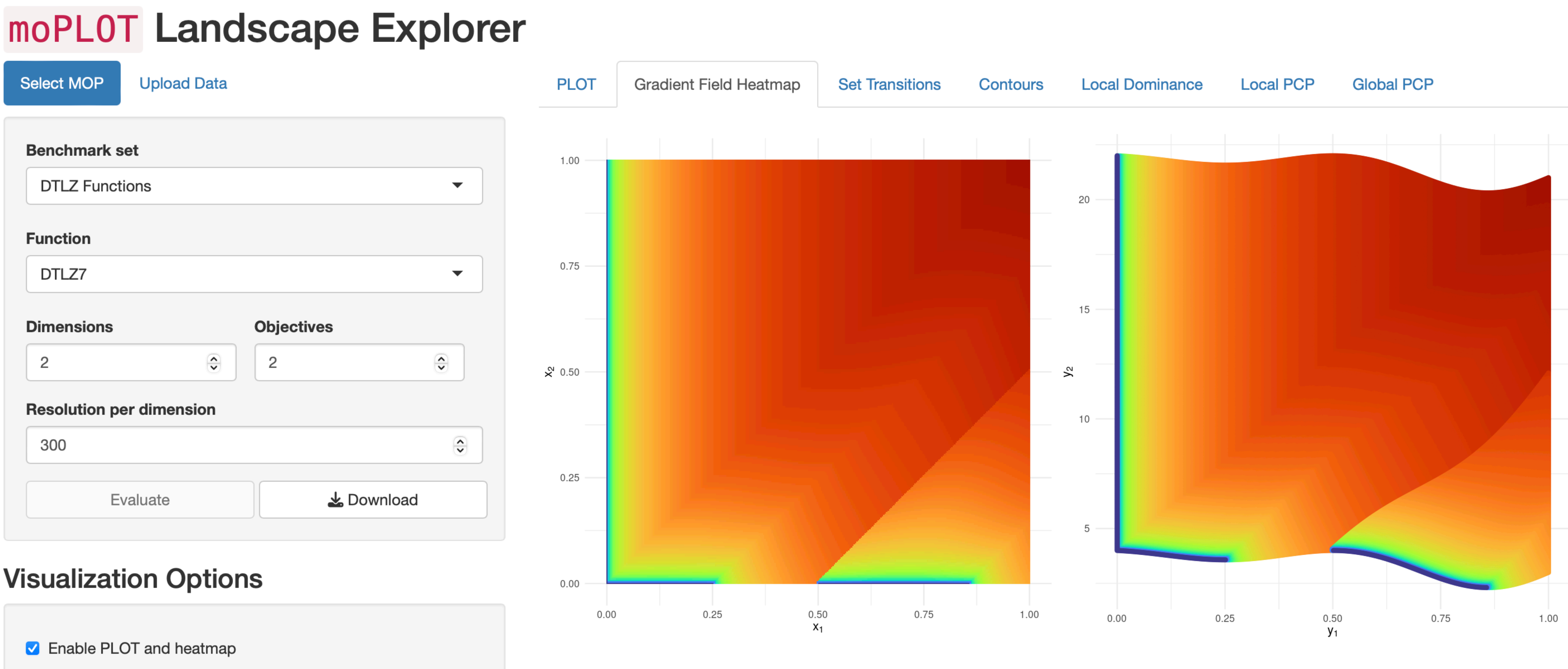
Dimension	Function	Instance	Visualization type
– 2 +	– 55 +	– 1 +	< Pareto set approximation >



Objectives = 2

moPLOT

https://schaepermeier.shinyapps.io/moPLOT/



Dimension = 2, 3 Objectives = 2, 3

(Compressed) PLOS-net


Manuel López-Ibánez


Hernán Aguirre

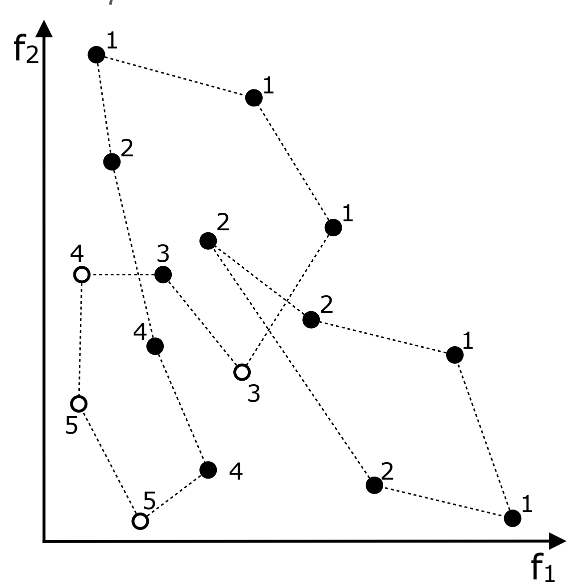

Kiyoshi Tanaka


Sébastien Verel

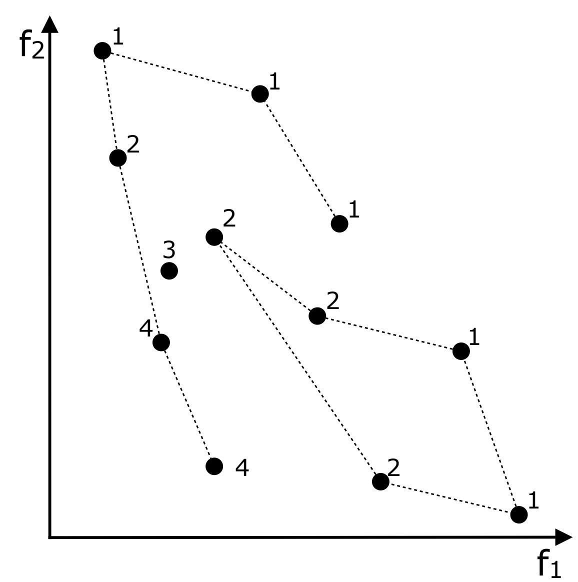

Bilel Derbel


Gabriela Ochoa

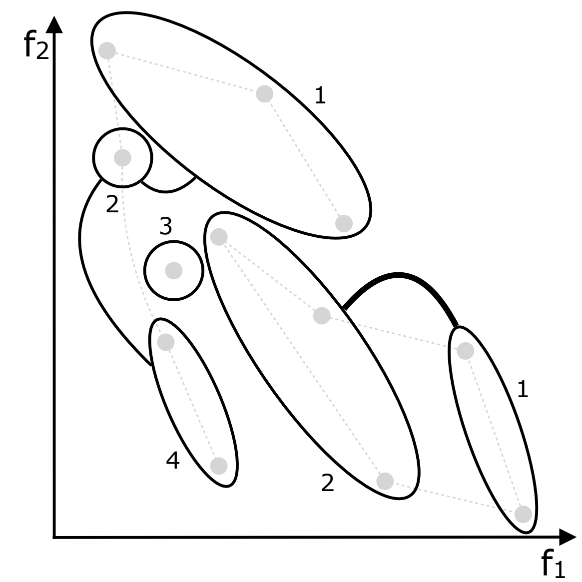
rank w.r.t. non-dominated sorting



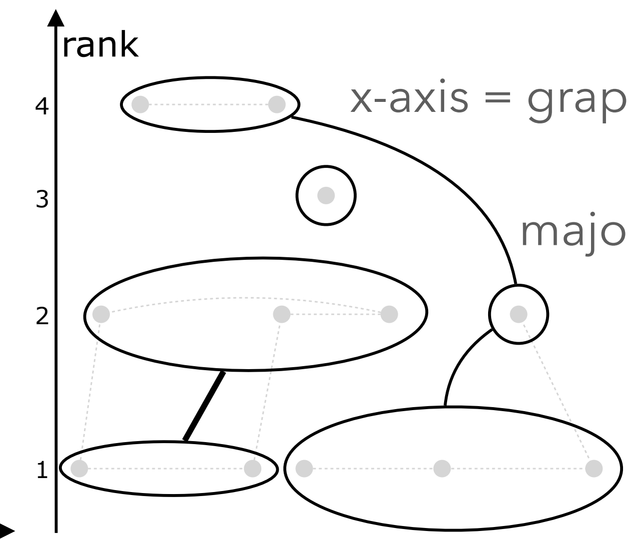
Full landscape



PLOS-net
[PPSN'18]



CPLOS-net
objective-space layout



x-axis = graph layout (stress majorization)

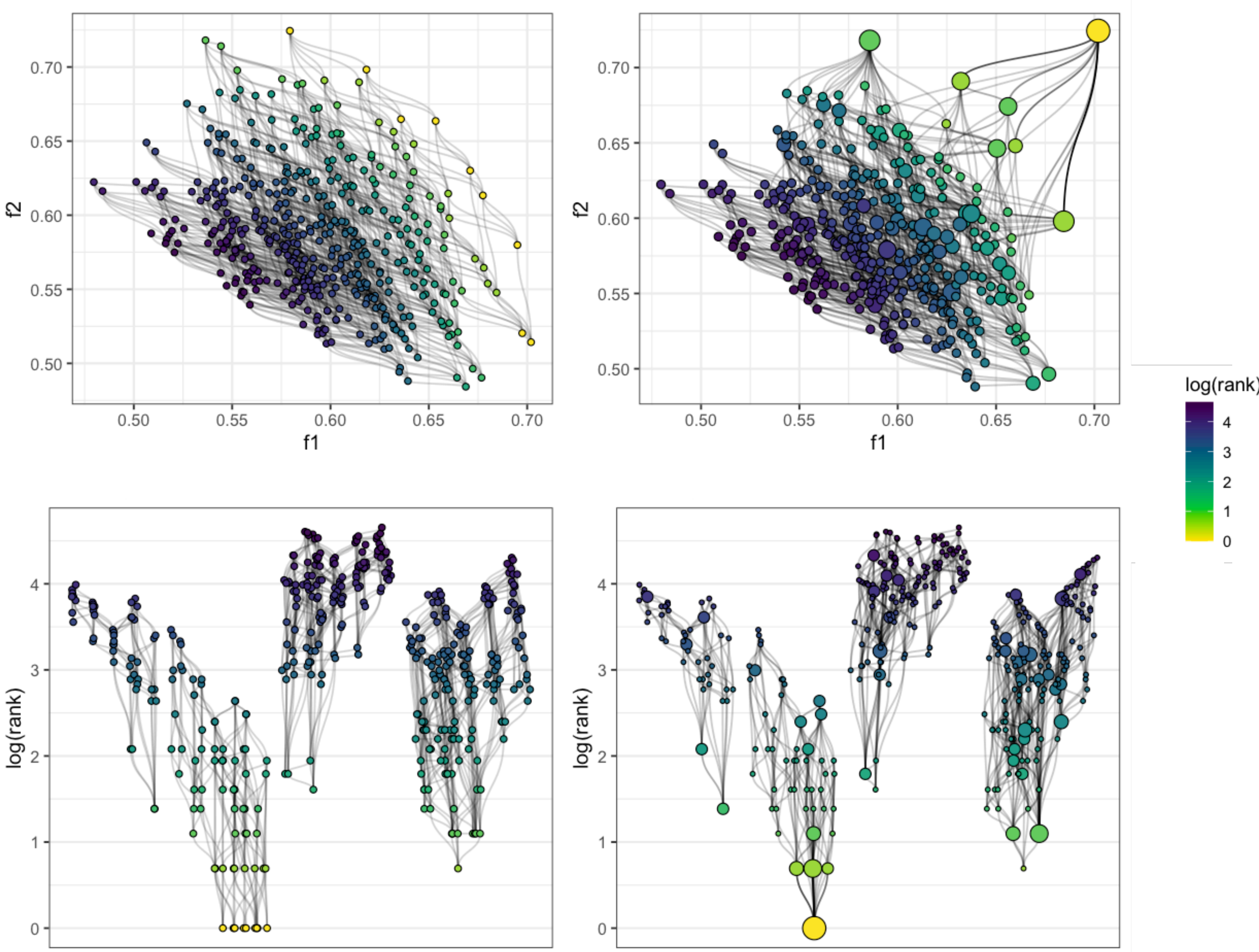
CPLOS-net
rank layout

[GECCO'23]

display for 2 (/3) objectives

scale to any number of objectives
invariant to **some** objective transformations

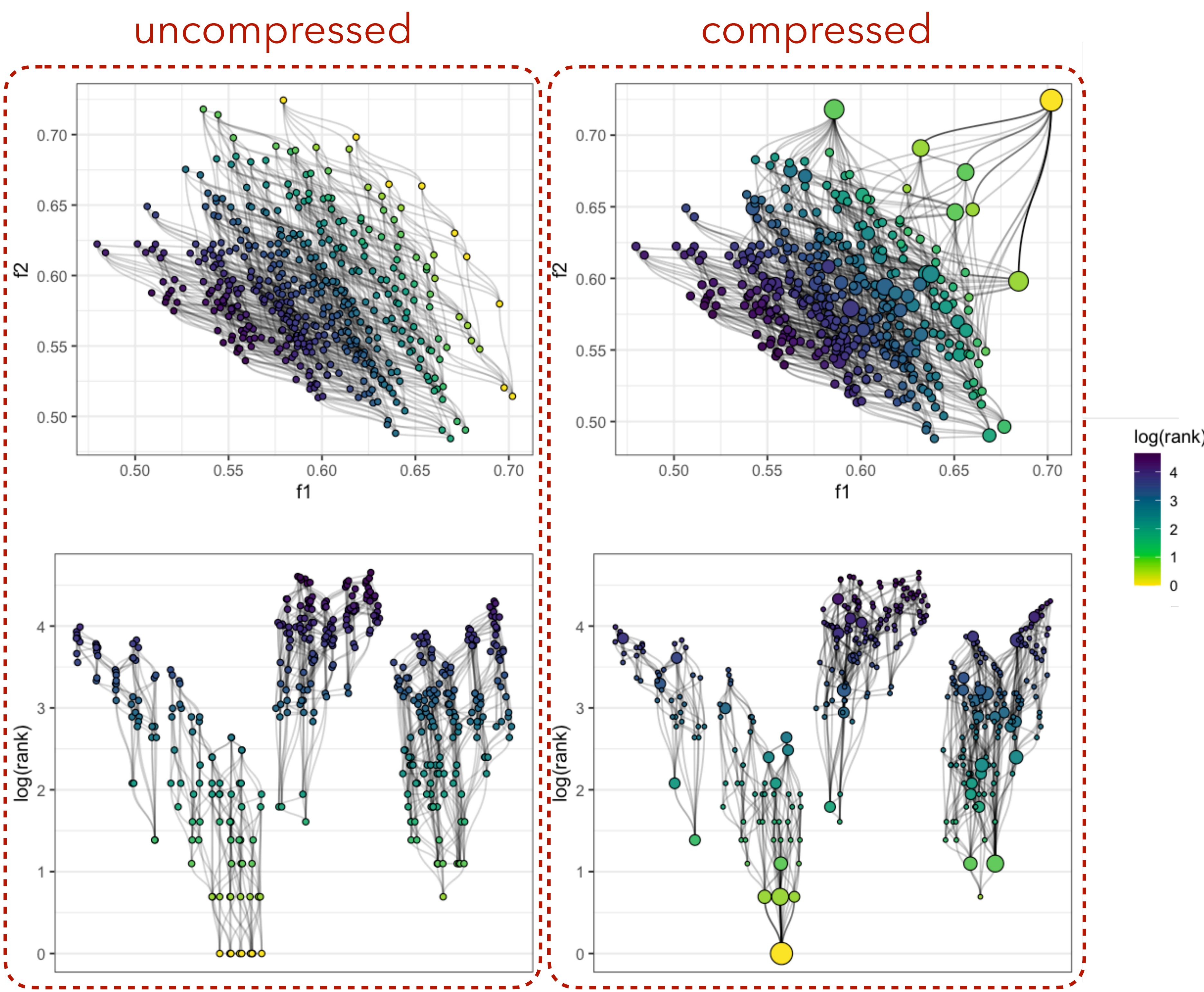
Uncompressed vs. Compressed PLOS-net



ρmnk-landscape

- $\rho = 0.4$
- $m = 2$
- $n = 16$
- $k = 1$

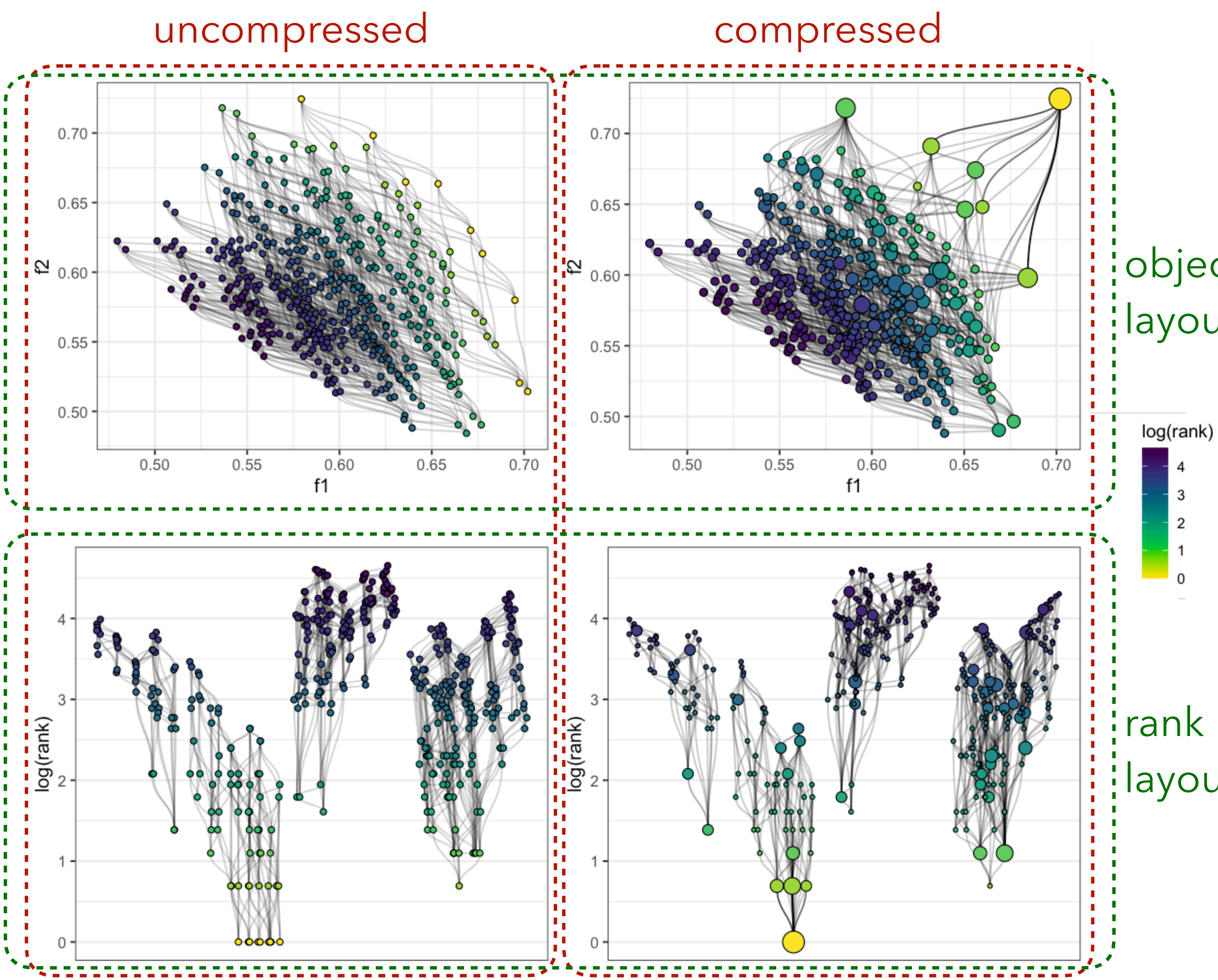
Uncompressed vs. Compressed PLOS-net



ρmnk-landscape

- $\rho = 0.4$
- $m = 2$
- $n = 16$
- $k = 1$

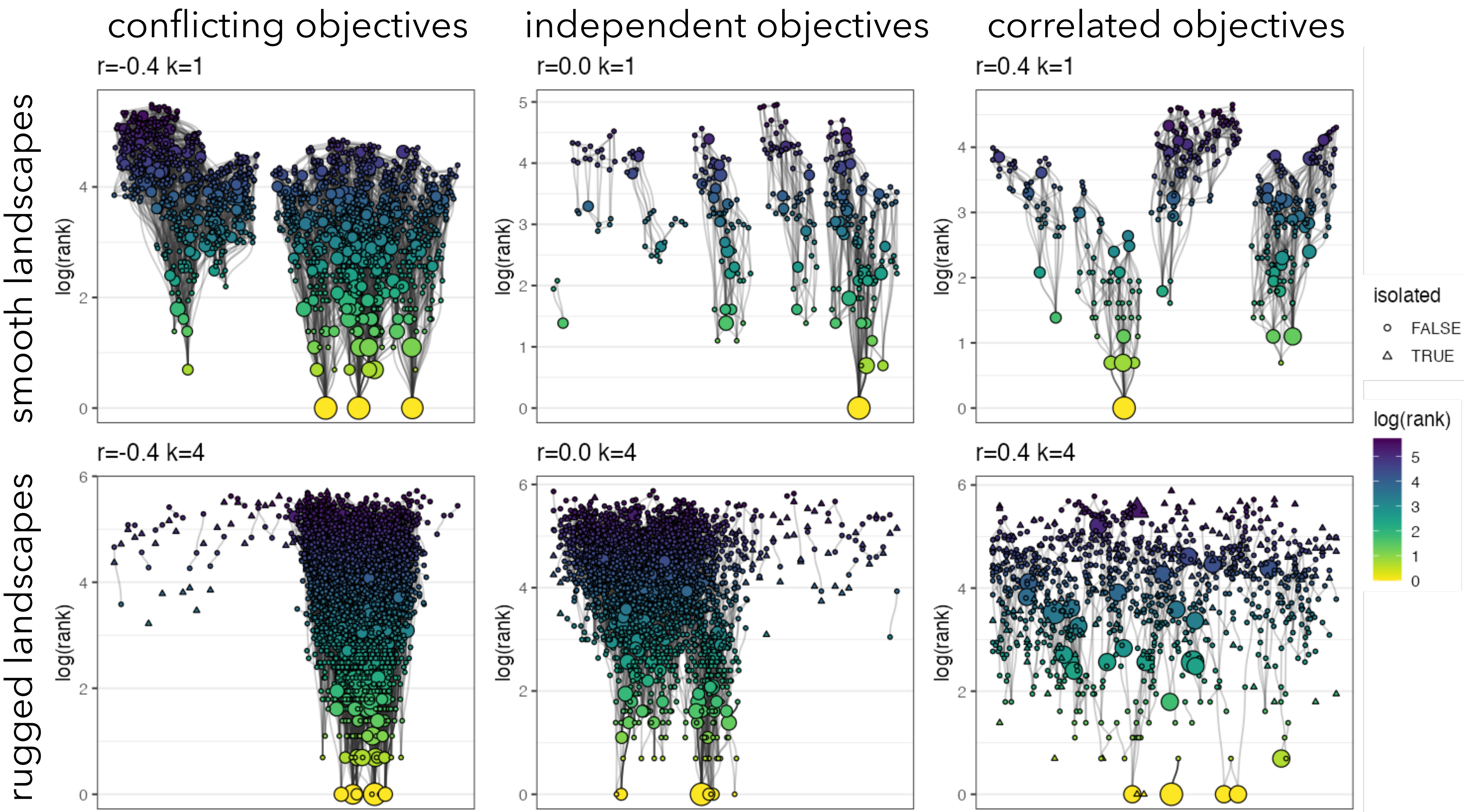
Uncompressed vs. Compressed PLOS-net



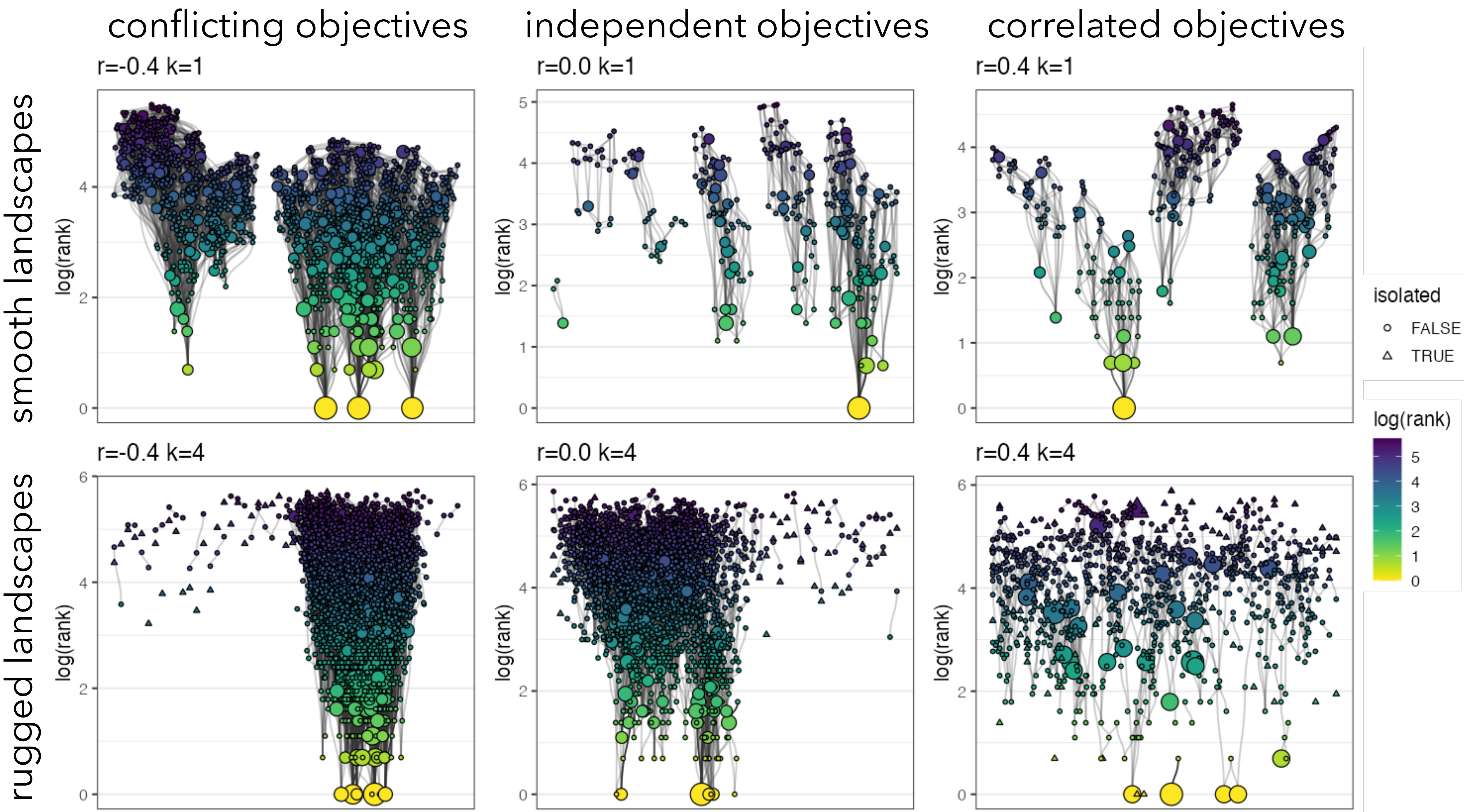
pmnk-landscape

- $\rho = 0.4$
- $m = 2$
- $n = 16$
- $k = 1$

C-PLOS-net visualization for 2 objectives



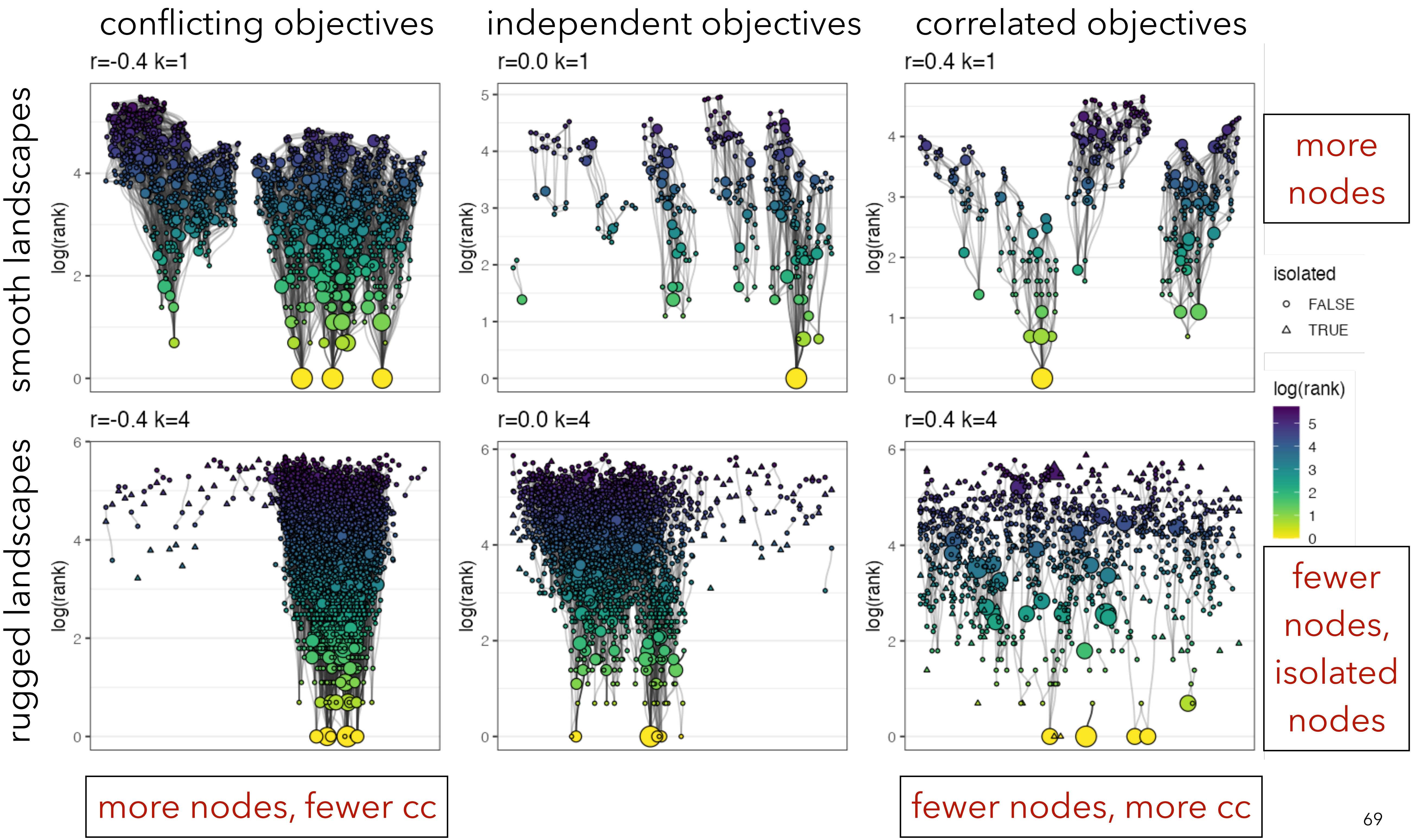
C-PLOS-net visualization for 2 objectives



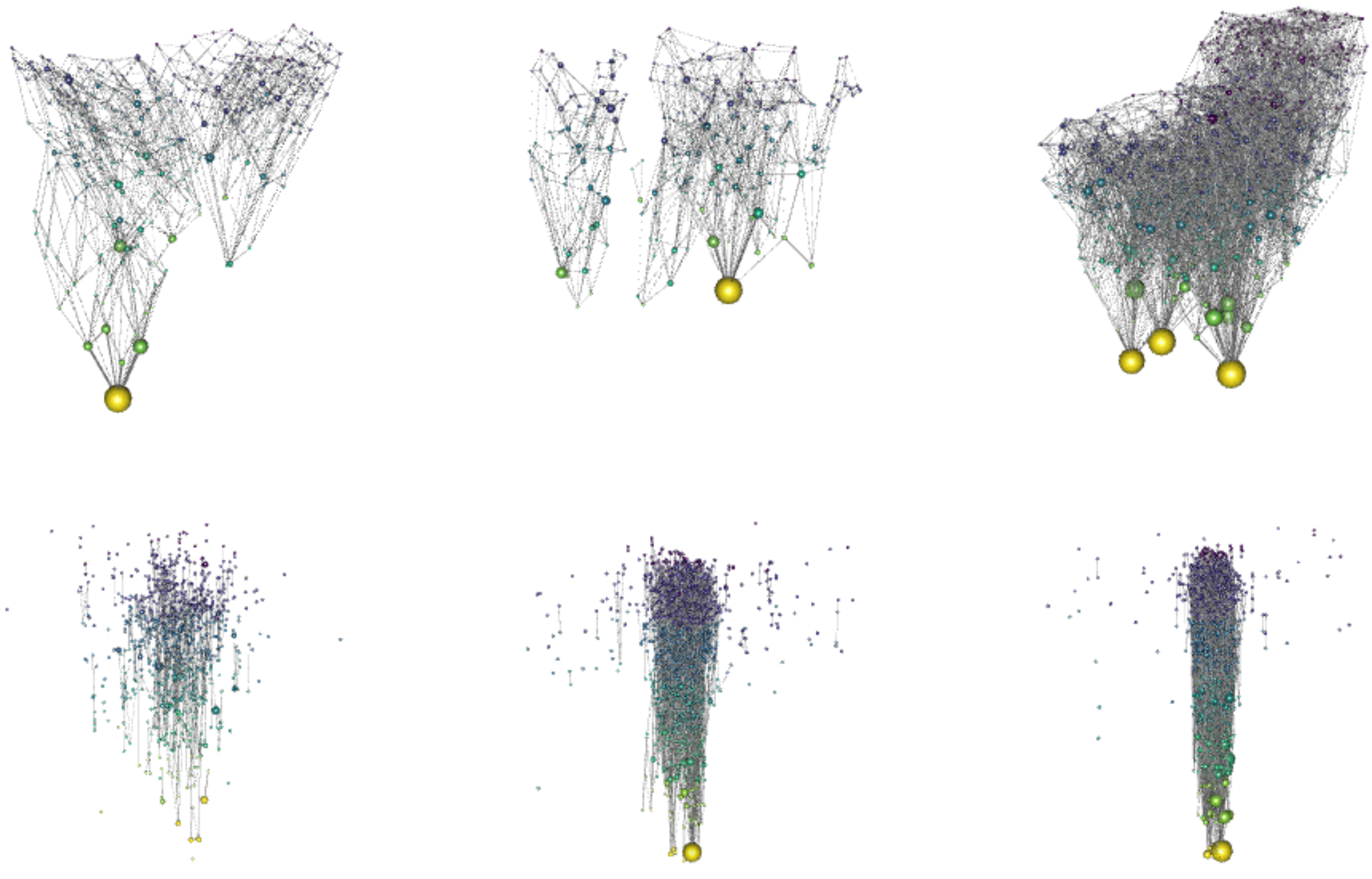
more nodes, fewer cc

fewer nodes, more cc

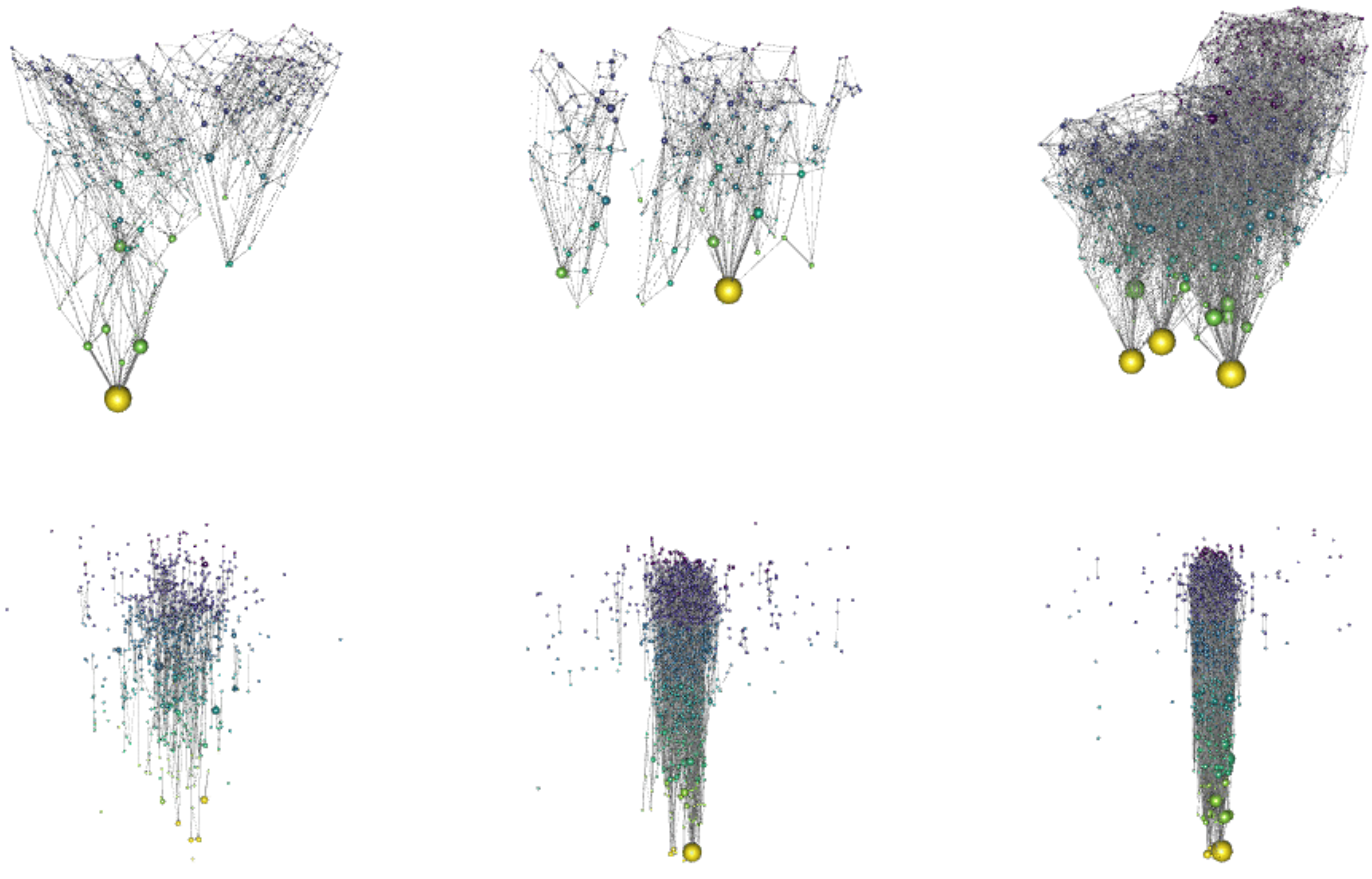
C-PLOS-net visualization for 2 objectives



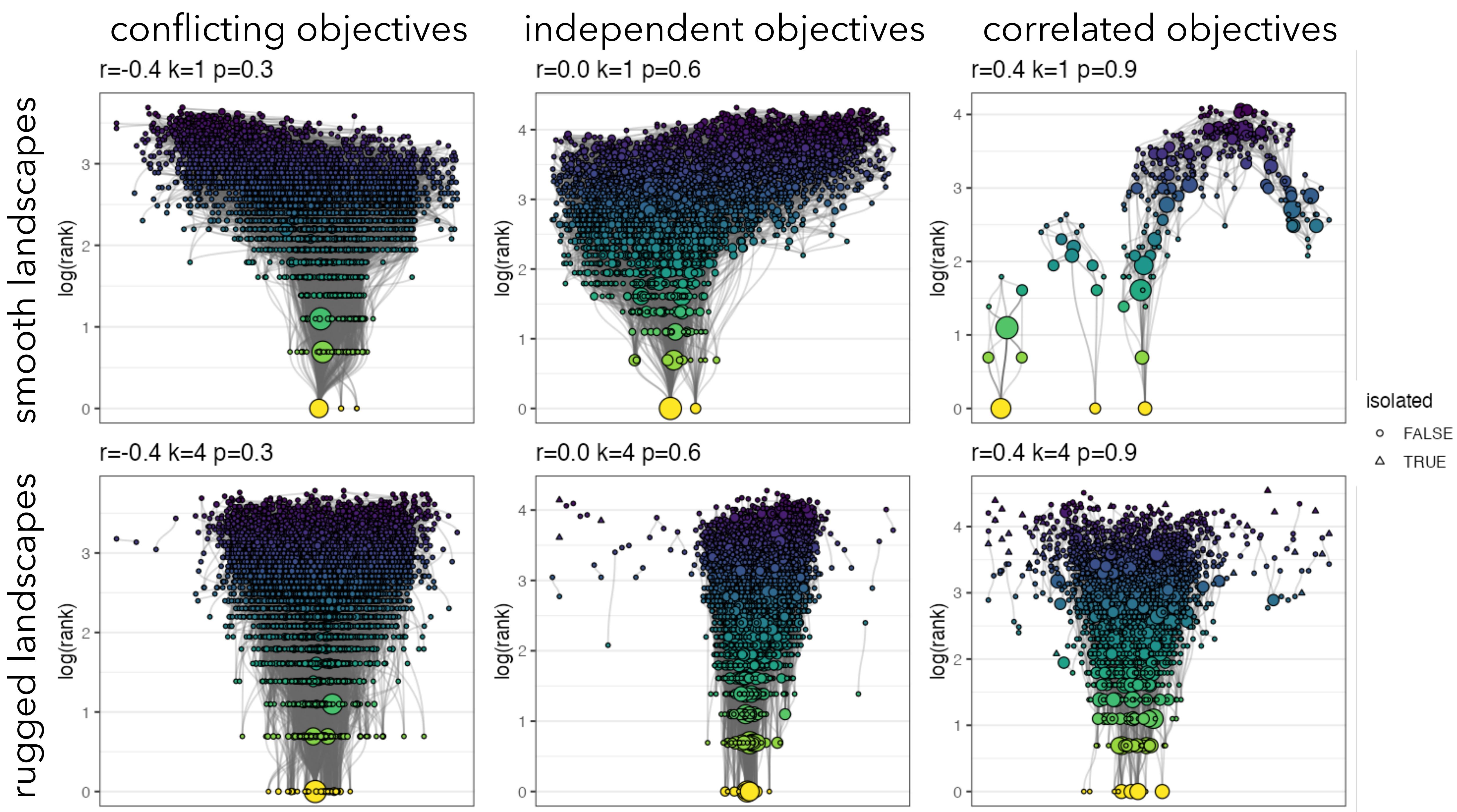
C-PLOS-net visualization for 2 objectives



C-PLOS-net visualization for 2 objectives



C-PLOS-net visualization for 3 objectives



pruning : p ∈ {0.3, 0.6, 0.9} for ρ ∈ {−0.4.0.0, 0.4}

Network metrics

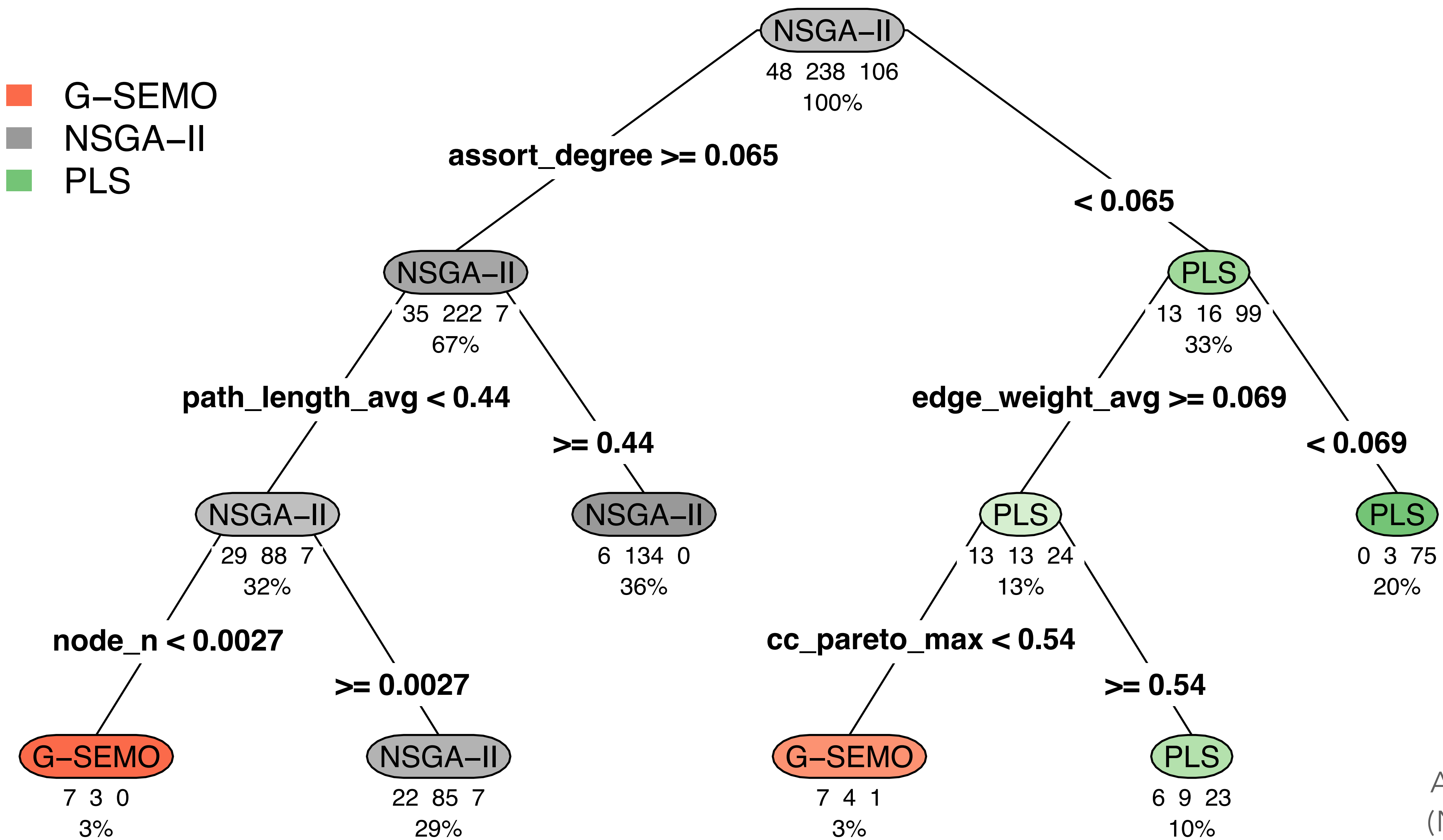
PLOS-nets metrics (24 + 10)

- **Support** and **complement** our visual **intuitions**
- Adapted from **complex networks**
- **Meaningful** for search

	metric	description
uncompressed and compressed networks	node_n	proportion of nodes
	node_pareto_n	proportion of Pareto nodes (nodes with rank 1)
	node_adj_pareto_n	proportion of nodes adjacent to a Pareto node
	node_rank_worst	maximum (worst) node rank
	degree_avg	average degree of nodes
	rank_degree_cor	node rank-vs-degree correlation
	isolated_n	proportion of isolated nodes
	pareto_isolated_n	proportion of Pareto nodes that are isolated
	isolated_rank_avg	average rank of isolated nodes
	edge_density	density of edges
	assort_degree	assortativity by degree
	cc_n	proportion of connected components (cc)
	cc_max	size of largest cc
	cc_avg	average size of cc
	cc_max_pareto	size of largest cc that contains a Pareto node
	cc_pareto_max	(average) size of cc with most Pareto nodes
	cc_pareto_avg	average number of Pareto nodes per cc
	cc_rank_avg_avg	mean of average rank per cc
	cc_rank_best_avg	mean of best rank per cc
	path_length_avg	average path length
	path_length_max	longest path length (diameter)
	path_pareto_exist	number of nodes connected to a Pareto node
	path_pareto_avg	avg. nb. of Pareto nodes a node is connected to
	path_length_pareto_avg	avg. (existing) path length to a Pareto node
compressed networks	node_width_avg	average node width
	node_cmpr	compression rate over nodes
	strength_avg	average node strength
	strength_pareto	sum of strengths of Pareto nodes
	rank_strength_cor	node rank-vs-strength correlation
	edge_weight_avg	average edge weight
	edge_cmpr	compression rate over edges
	dist_avg	average distance
	dist_max	longest distance
	dist_pareto_avg	avg. dist. to Pareto nodes (existing paths)

Interpretable algorithm prediction

Classification Model (CART tree) = {algo} ~ {metrics}

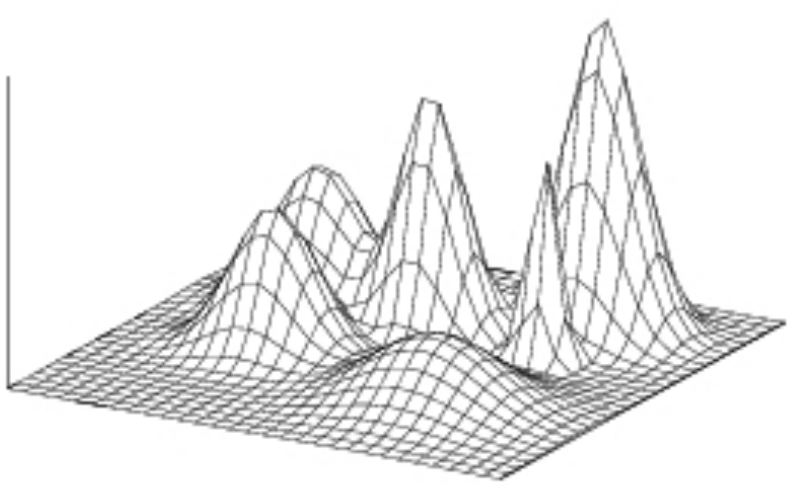


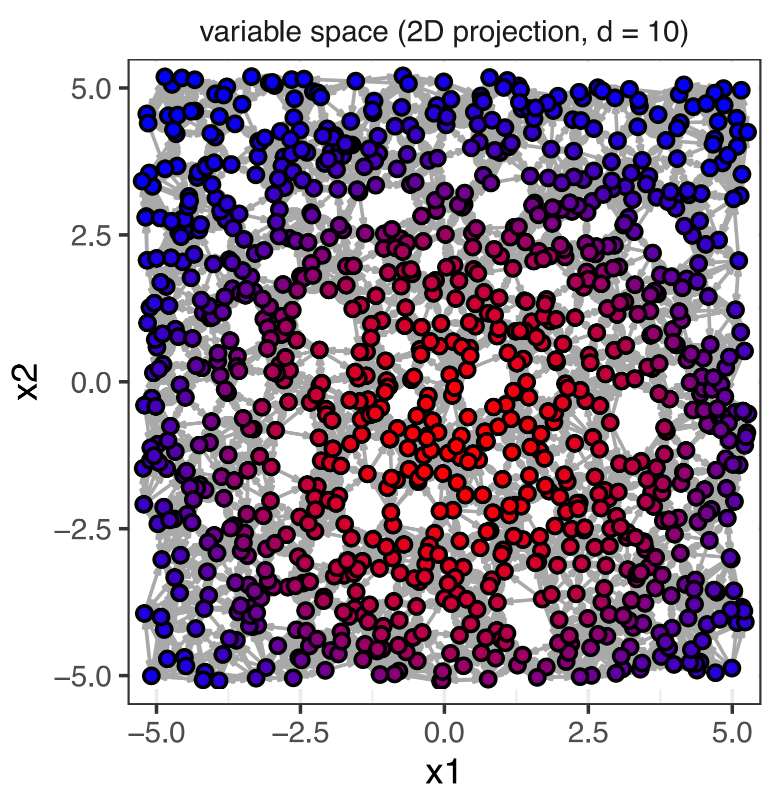
Accuracy = 84%
(NSGA-II = 61%)

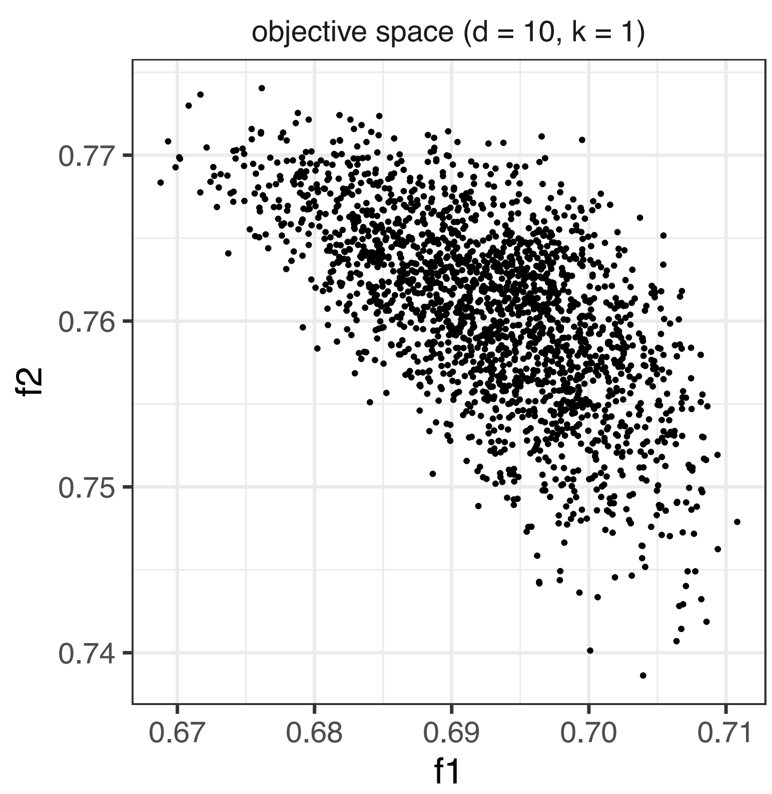
Landscape Features for Continuous MO Optimization

Continuous MO Landscape Features

- Discretize space + 'standard' measures from landscape analysis
  - Budget = *n* solutions, from random latin hypercube design
  - Neighbors = *d* closest solutions (euclidean dist. in var. space)







Benchmark

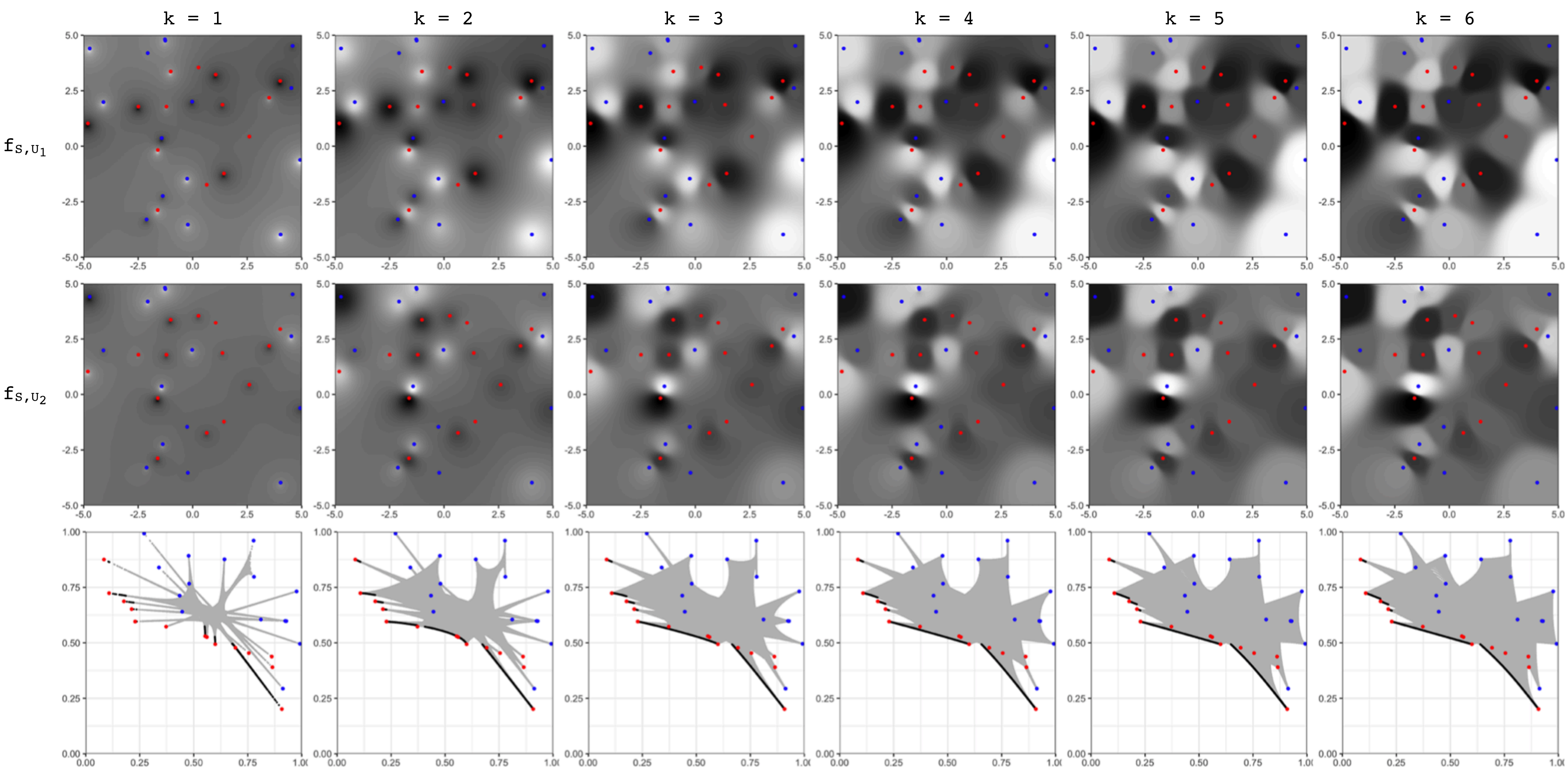



Sébastien Verel · Benjamin Lacroix · Ciprian Zăvoianu · John McCall

Interpolated Continuous MOPs

d	number of variables
k	power of interpolation
seed_n	proportional number of seeds $\lvert S \rvert$
nd_seed_n	proportion of non-dominated seeds $\lvert S_{nd} \rvert$
dom_seed_n	proportion of dominated seeds $\lvert S_d \rvert$

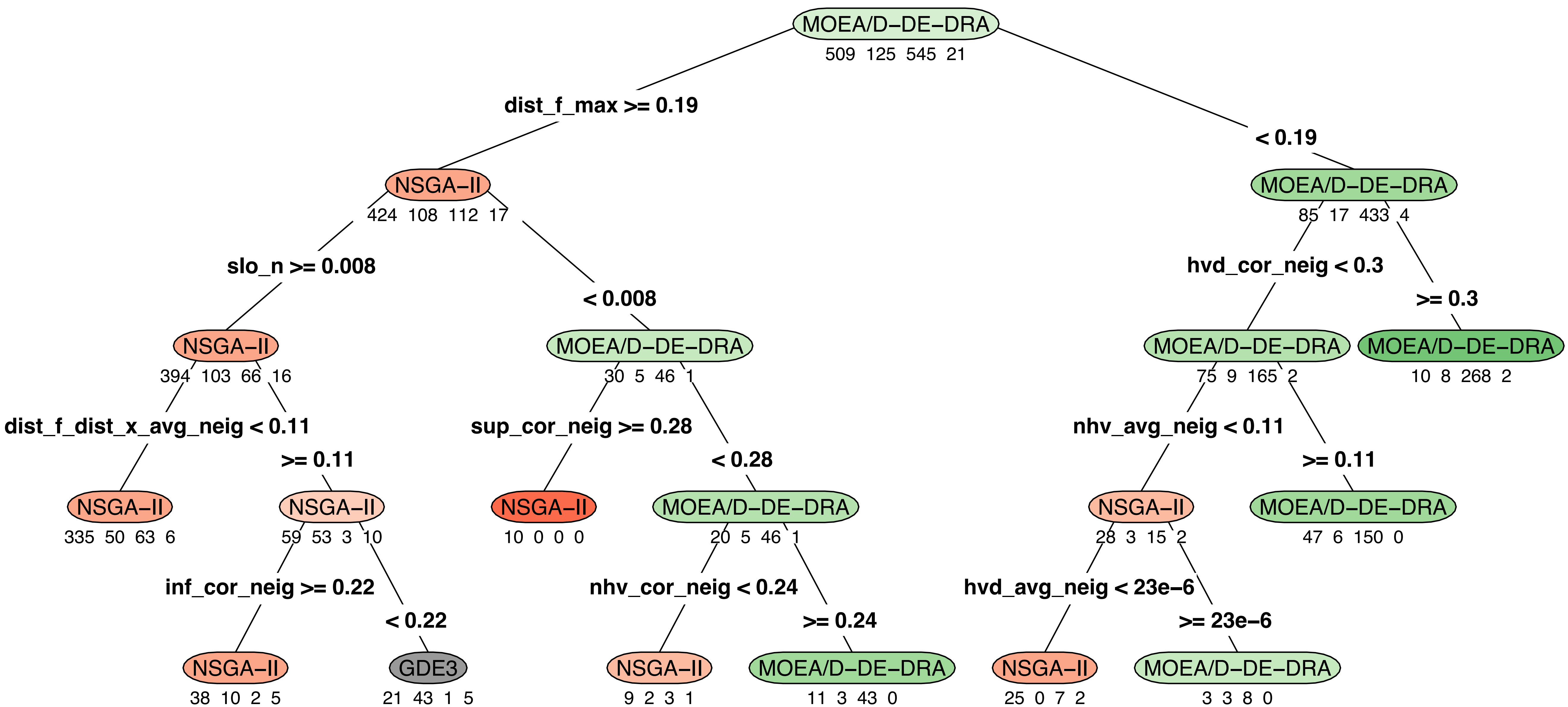
$$f_{S,U_i}(x) = \begin{cases} \dfrac{\sum_{j=1}^{N} \frac{u_{i,j}}{e(x,s_j)^k}}{\sum_{j=1}^{N} \frac{1}{e(x,s_j)^k}} & \text{if } e\left(x, s_j\right) \neq 0 \text{ for all } j \\ u_{i,j} & \text{if } e\left(x, s_j\right) = 0 \text{ for some } j \end{cases}$$



Importance of Features

(budget = 50000)

Model (classification, decision tree) = {algo} ~ {features}

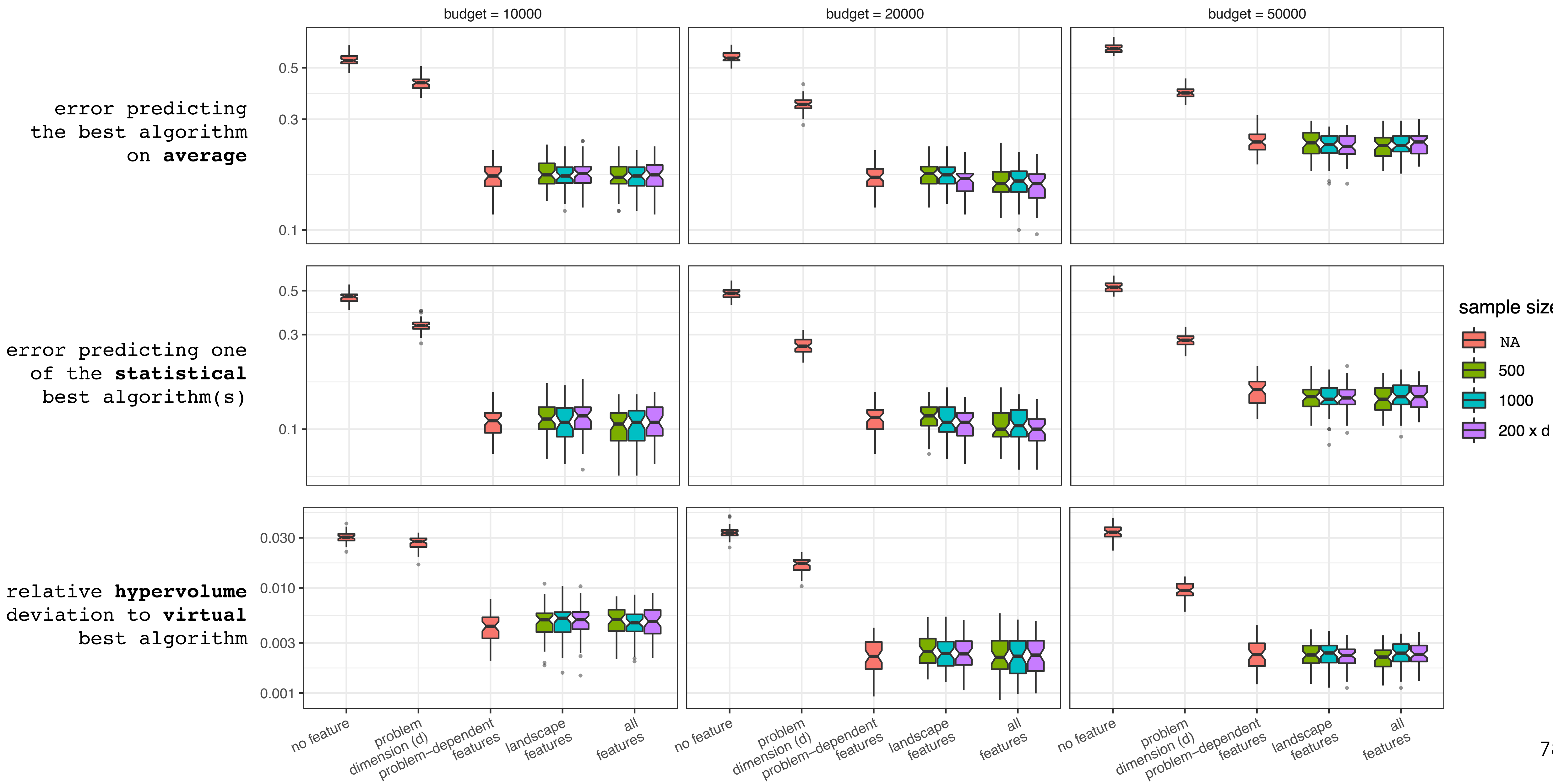


classification error = 22.58%

Prediction Accuracy

Model (classification, random forest) = {algo} ~ {features}

random subsampling cross-validation (x 50, 80/20% split)



Continuous MO Landscapes

Towards Constructing a Suite of Multi-objective Optimization Problems with Diverse Landscapes

Andrejaana Andova[1,2](✉), Tobias Benecke[3], Harald Ludwig[4], and Tea Tušar[1,2]

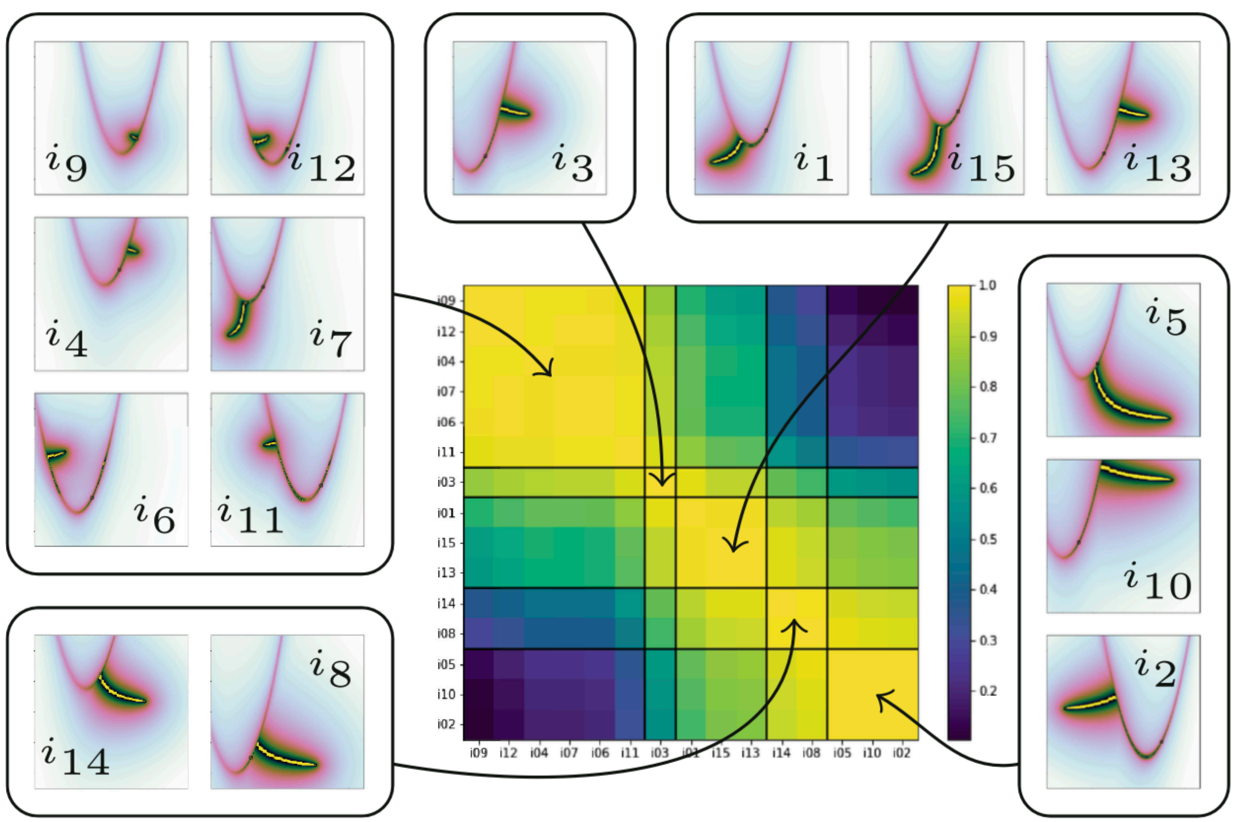
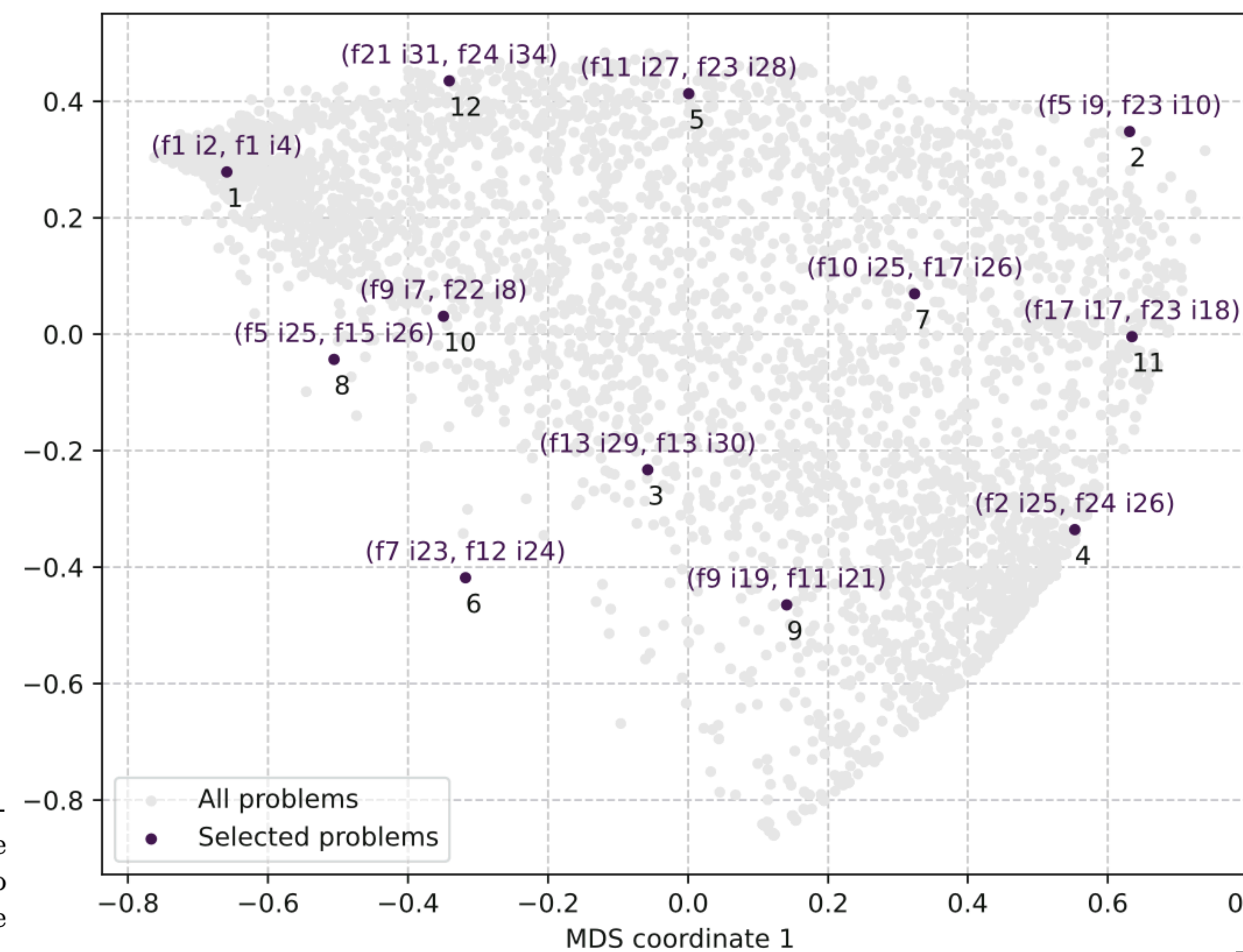


Fig. 1. Problem similarity heatmap for the 15 instances of the 2-D function combination (f_1, f_8) (function F_4 in the `bbob-biobj` suite) and their corresponding landscape visualizations using dominance ranking ratio [3,7] show that the size of the Pareto set (yellow curves) is proportional to the similarity among instance. The instances are numbered as in the `bbob-biobj` suite. (Color figure online)

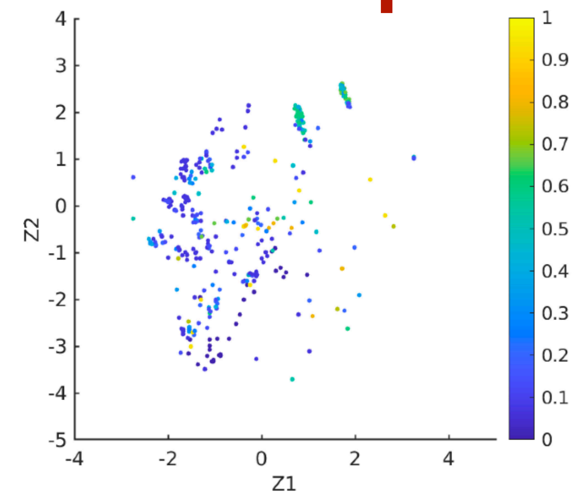


Constrained Continuous MO Landscapes

IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, VOL. 27, NO. 5, OCTOBER 2023 1427

An Instance Space Analysis of Constrained Multiobjective Optimization Problems

Hanan Alsouly, Michael Kirley, and Mario Andrés Muñoz



Features Used to Characterize the Multiobjectives Landscape of CMOP

Type	Feature	Description	Source	Focus
Global	upo_n	Proportion of unconstrained PO solutions	[27]	Set-Cardinality
	uhv	Hypervolume-value of the $\widetilde{UPF}$	[28]	Set-Distribution
	corr_obj	correlation between objective values	[29]	evolvability
	mean_f	Average of unconstrained ranks	[12]	y-distribution
	std_f	Standard deviation of unconstrained ranks	[5]	y-distribution
	max_f	Maximum of unconstrained ranks	[5]	y-distribution
	skew_f	Skewness of unconstrained ranks	[5]	y-distribution
	kurt_f	Kurtosis of unconstrained ranks	[5]	y-distribution
	kurt_avg	Average of objectives kurtosis	[5]	y-distribution
	kurt_min	Minimum of objectives kurtosis	[5]	y-distribution
	kurt_max	Maximum of objectives kurtosis	[5]	y-distribution
	kurt_rnge	Range of objectives kurtosis	[5]	y-distribution
	skew_avg	Average of objectives skewness	[5]	y-distribution
	skew_min	Minimum of objectives skewness	[5]	y-distribution
	skew_max	Maximum of objectives skewness	[5]	y-distribution
	skew_rnge	Range of objectives skewness	[5]	y-distribution
	f_mdl_r2	Adjusted coefficient of determination of a linear regression model for varibles and unconstrained ranks	[5]	variable scaling
	f_range_coeff	Difference between maximum and minimum of the absolute value of the linear model coefficients	[5]	variable scaling
Random Walk	dist_f_avg_rws	Average distance from neighbours in the objective space	[12]	evolvability
	dist_f_r1_rws	First autocorrelation coefficient of dist_f_avg_rws	[12]	ruggedness
	dist_f_dist_x_avg_rws	Ratio of dist_f_avg_rws to dist_x_avg_rws	[12]	evolvability
	dist_f_dist_x_avg_r1	First autocorrelation coefficient of dist_f_dist_x_avg_rws	[12]	ruggedness
	nuhv_avg_rws	Average unconstrained hypervolume-value of neighborhood's solutions	[29]	evolvability
	nuhv_r1_rws	First autocorrelation coefficient of nuhv_avg_rws	[29]	ruggedness

Features Used to Characterize the Multiobjectives Landscape of CMOP

Type	Feature	Description	Source	Focus
Global	upo_n	Proportion of unconstrained PO solutions	[27]	Set-Cardinality
	uhv	Hypervolume-value of the $\widetilde{UPF}$	[28]	Set-Distribution
	corr_obj	correlation between objective values	[29]	evolvability
	mean_f	Average of unconstrained ranks	[12]	y-distribution
	std_f	Standard deviation of unconstrained ranks	[5]	y-distribution
	max_f	Maximum of unconstrained ranks	[5]	y-distribution
	skew_f	Skewness of unconstrained ranks	[5]	y-distribution
	kurt_f	Kurtosis of unconstrained ranks	[5]	y-distribution
	kurt_avg	Average of objectives kurtosis	[5]	y-distribution
	kurt_min	Minimum of objectives kurtosis	[5]	y-distribution
	kurt_max	Maximum of objectives kurtosis	[5]	y-distribution
	kurt_rnge	Range of objectives kurtosis	[5]	y-distribution
	skew_avg	Average of objectives skewness	[5]	y-distribution
	skew_min	Minimum of objectives skewness	[5]	y-distribution
	skew_max	Maximum of objectives skewness	[5]	y-distribution
	skew_rnge	Range of objectives skewness	[5]	y-distribution
	f_mdl_r2	Adjusted coefficient of determination of a linear regression model for varibles and unconstrained ranks	[5]	variable scaling
	f_range_coeff	Difference between maximum and minimum of the absolute value of the linear model coefficients	[5]	variable scaling
Random Walk	dist_f_avg_rws	Average distance from neighbours in the objective space	[12]	evolvability
	dist_f_r1_rws	First autocorrelation coefficient of dist_f_avg_rws	[12]	ruggedness
	dist_f_dist_x_avg_rws	Ratio of dist_f_avg_rws to dist_x_avg_rws	[12]	evolvability
	dist_f_dist_x_avg_r1	First autocorrelation coefficient of dist_f_dist_x_avg_rws	[12]	ruggedness
	nuhv_avg_rws	Average unconstrained hypervolume-value of neighborhood's solutions	[29]	evolvability
	nuhv_r1_rws	First autocorrelation coefficient of nuhv_avg_rws	[29]	ruggedness

Features Used to Characterize the Violation Landscape of CMOP. The Proposed Features Marked as New, While the (*) Indicates That the Feature Has Been Modified to Characterize CMOP

Type	Feature	Description	Source	Focus
Global	min_cv	Minimum of constraints violations	[5] *	y-distribution
	skew_cv	Skewness of constraints violations	[5] *	y-distribution
	kurt_cv	Kurtosis of constraints violations	[5] *	y-distribution
	cv_mdl_r2	Adjusted coefficient of determination of a linear regression model for varibles and violations	[5] *	variable scaling
	cv_range_coeff	Difference between maximum and minimum of the absolute value of the linear model coefficients	[5] *	variable scaling
	dist_c_corr	Violation-distance correlation	[30] *	deception
Random Walk	dist_c_avg_rws	Average distance from neighbours in the constraints space	[12] *	evolvability
	dist_c_r1_rws	first autocorrelation coefficient of dist_c_avg_rws	[12] *	ruggedness
	dist_c_dist_x_avg_rws	Ratio of dist_c_avg_rws to dist_x_avg_rws	[12] *	evolvability
	dist_c_dist_x_r1_rws	First autocorrelation coefficient of dist_c_dist_x_avg_rws	[12] *	ruggedness
	ncv_avg_rws	Average single solution's violation-value	New	evolvability
	ncv_r1_rws	first autocorrelation coefficient of ncv_avg_rws	New	ruggedness
	nncv_avg_rws	Average neighborhood's violation-value	New	evolvability
	nncv_r1_rws	first autocorrelation coefficient of nncv_avg_rws	New	ruggedness
	bncv_avg_rws	Average violation-value of neighborhood's non-dominated solutions	New	evolvability
	bncv_r1_rws	first autocorrelation coefficient of bncv_avg_rws	New	ruggedness

TABLE III

Features Used to Characterize the Violation Landscape of CMOP. The Proposed Features Marked as New, While the (*) Indicates That the Feature Has Been Modified to Characterize CMOP

Type	Feature	Description	Source	Focus
Global	min_cv	Minimum of constraints violations	[5] *	y-distribution
	skew_cv	Skewness of constraints violations	[5] *	y-distribution
	kurt_cv	Kurtosis of constraints violations	[5] *	y-distribution
	cv_mdl_r2	Adjusted coefficient of determination of a linear regression model for varibles and violations	[5] *	variable scaling
	cv_range_coeff	Difference between maximum and minimum of the absolute value of the linear model coefficients	[5] *	variable scaling
	dist_c_corr	Violation-distance correlation	[30] *	deception
Random Walk	dist_c_avg_rws	Average distance from neighbours in the constraints space	[12] *	evolvability
	dist_c_r1_rws	first autocorrelation coefficient of dist_c_avg_rws	[12] *	ruggedness
	dist_c_dist_x_avg_rws	Ratio of dist_c_avg_rws to dist_x_avg_rws	[12] *	evolvability
	dist_c_dist_x_r1_rws	First autocorrelation coefficient of dist_c_dist_x_avg_rws	[12] *	ruggedness
	ncv_avg_rws	Average single solution's violation-value	New	evolvability
	ncv_r1_rws	first autocorrelation coefficient of ncv_avg_rws	New	ruggedness
	nncv_avg_rws	Average neighborhood's violation-value	New	evolvability
	nncv_r1_rws	first autocorrelation coefficient of nncv_avg_rws	New	ruggedness
	bncv_avg_rws	Average violation-value of neighborhood's non-dominated solutions	New	evolvability
	bncv_r1_rws	first autocorrelation coefficient of bncv_avg_rws	New	ruggedness

Constrained Continuous MO Landscapes

Information Sciences 607 (2022) 244–262

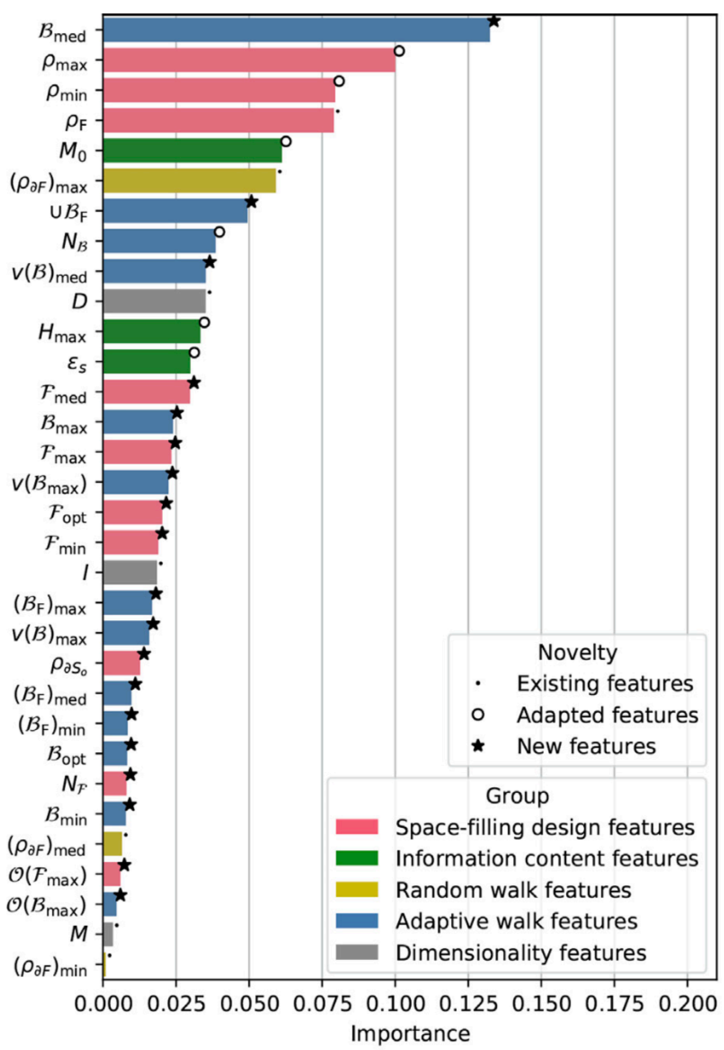
Contents lists available at ScienceDirect

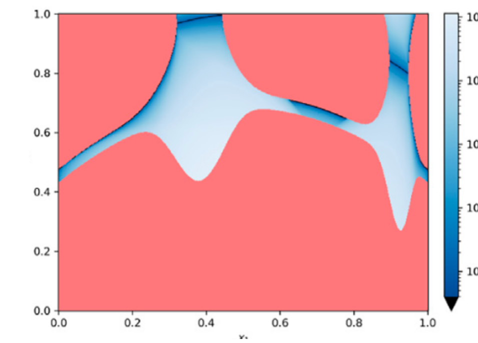
Information Sciences

journal homepage: www.elsevier.com/locate/ins

Characterization of constrained continuous multiobjective optimization problems: A feature space perspective

Aljoša Vodopija [a,b,*], Tea Tušar [a,b], Bogdan Filipič [a,b]





The proposed ELA features to characterize CMOPs categorized into four groups: space-filling design, information content, random walk, and adaptive walk. "New" indicates that the corresponding feature is proposed in this paper.

Space-filling design features		
$N_{\mathscr{F}}$	Number of feasible components	New
$\mathscr{F}_{\text{min}}$	Smallest feasible component	New
$\mathscr{F}_{\text{med}}$	Median feasible component	New
$\mathscr{F}_{\text{max}}$	Largest feasible component	New
$\mathcal{O}(\mathscr{F}_{\text{max}})$	Proportion of Pareto-optimal solutions in $\mathscr{F}_{\text{max}}$	New
$\mathscr{F}_{\text{opt}}$	Size of the "optimal" feasible component	New
ρ_F	Feasibility ratio	[19]
ρ_{min}	Minimum correlation	[18][a]
ρ_{max}	Maximum correlation	[18][a]
$\rho_{\partial S_o}$	Proportion of boundary Pareto-optimal solutions	New
Information content features		
H_{max}	Maximum information content	[27][b]
ε_s	Settling sensitivity	[27][b]
M_0	Initial partial information	[27][b]
Random walk features		
$(\rho_{\partial F})_{\text{min}}$	Minimum ratio of feasible boundary crossings	[18,19]
$(\rho_{\partial F})_{\text{med}}$	Median ratio of feasible boundary crossings	[18,19]
$(\rho_{\partial F})_{\text{max}}$	Maximum ratio of feasible boundary crossings	[18,19]
Adaptive walk features		
$N_{\mathscr{B}}$	Number of basins	[28][b]
$\mathscr{B}_{\text{min}}$	Smallest basin	New
$\mathscr{B}_{\text{med}}$	Median basin	New
$\mathscr{B}_{\text{max}}$	Largest basin	New
$(\mathscr{B}_F)_{\text{min}}$	Smallest feasible basin	New
$(\mathscr{B}_F)_{\text{med}}$	Median feasible basin	New
$(\mathscr{B}_F)_{\text{max}}$	Largest feasible basin	New
$\cup\mathscr{B}_F$	Proportion of feasible basins	New
$v(\mathscr{B})_{\text{med}}$	Median constraint violation over all basins	New
$v(\mathscr{B})_{\text{max}}$	Maximum constraint violation of all basins	New
$v(\mathscr{B}_{\text{max}})$	Constraint violation of $\mathscr{B}_{\text{max}}$	New
$\mathcal{O}(\mathscr{B}_{\text{max}})$	Proportion of Pareto-optimal solutions in $\mathscr{B}_{\text{max}}$	New
$\mathscr{B}_{\text{opt}}$	Size of the "optimal" basin	New

Contents

Conclusions

- Many (E)MO algorithms, few recommendations w.r.t. target problem
  - where are the key differences in behavior among them?
- Multi-objective landscapes (interpretable features, visualization)
- Multi-objective optimization is a set problem
  - solution-level features capture information about the neighboring set
  - set-level features are insightful, but also challenging

Related Issues

- Multi-objectivization
- Multi-objective landscape features from decomposition
- Many-objective landscapes (they tend to get easier in some respects)
- …

References (1)

- Aguirre, Tanaka. **Working principles, behavior, and performance of MOEAs on MNK-landscapes**. European Journal of Operational Research, vol. 181, no. 3, pp. 1670–1690, 2007
- Alsouly, Kirley, Muñoz. **An instance space analysis of constrained multiobjective optimization problems**. IEEE Transactions on Evolutionary Computation, vol. 27, no. 5, pp. 1427-1439, 2023
- Andova, Benecke, Ludwig, Tusar. **Towards constructing a suite of multi-objective optimization problems with diverse landscapes.** EvoApplications 2023, LNCS vol. 13989, pp. 442-457, Brno, Czech Republic, 2023
- Brockhoff, Auger, Hansen, Tusar. **Using well-understood single-objective functions in multiobjective black-box optimization test suites**. Evolutionary Computation, vol. 30, no. 2, pp. 165-193, 2022
- Brockhoff, Friedrich, Hebbinghaus, Klein, Neumann, Zitzler. **Do additional objectives make a problem harder?** GECCO 2007, pp. 765–772, London, UK, 2007
- Cosson, Derbel, Liefooghe, Aguirre, Tanaka, Zhang. **Decomposition-based multi-objective landscape features and automated algorithm selection**. EvoCOP 2021, LNCS vol. 12692, pp. 34-50 Spain, 2021
- Cosson, Santana, Derbel, Liefooghe. **Multi-objective NK landscapes with heterogeneous objectives**. GECCO 2022, pp.502-510, Boston, MA, USA, 2022
- Daolio, Liefooghe, Verel, Aguirre, Tanaka. **Problem features versus algorithm performance on rugged multiobjective combinatorial fitness landscapes**. Evolutionary Computation, vol. 25, no. 4, pp. 555–585, 2017
- Ehrgott. **Multicriteria optimization**. Springer (2nd ed.), 2005
- Fieldsend, Alyahya. **Visualising the landscape of multi-objective problems using local optima networks**. GECCO 2019, pp. 1421–1429, Prague, Czech Republic, 2019
- Fonseca. **Multiobjective genetic algorithms with application to control engineering problems**. PhD thesis, University of Sheffield, 1995
- Garrett, Dasgupta. **Multiobjective landscape analysis and the generalized assignment problem**. LION 2, LNCS vol. 5313, pp. 110–124, Trento, Italy, 2007

References (2)

- Garrett, Dasgupta. **Plateau connection structure and multiobjective metaheuristic performance**. CEC 2009, pp. 1281–1288, 2009
- Grimme, Kerschke, Trautmann. **Multimodality in multi-objective optimization – More boon than bane?** EMO 2019, LNCS vol. 11411, pp. 126–138, Lansing, MI, USA, 2019
- Handl, Lovell, Knowles. **Multiobjectivization by decomposition of scalar cost functions**. PPSN 2008, LNCS vol. 5199, pp. 31-40, Dortmund, Germany, 2008
- Jensen. **Helper-objectives: Using multi-objective evolutionary algorithms for single-objective optimisation**. Journal of Mathematical Modelling and Algorithms vol. 3, no. 4, pp. 323–347, 2004
- Kauffman. **The origins of order**. Oxford University Press, 1993
- Kerschke, Grimme. **An expedition to multimodal multi objective optimization landscapes**. EMO 2017, LNCS vol. 10173 pp. 329–343, Münster, Germany, 2017
- Knowles, Corne. **Instance generators and test suites for the multiobjective quadratic assignment problem**. EMO 2003, LNCS vol. 2632, pp. 295–310, Faro, Portugal, 2003
- Knowles, Corne. **Quantifying the effects of objective space dimension in evolutionary multiobjective optimization**. EMO 2007, LNCS vol. 4403, pp. 757–771, Matsushima, Japan, 2007
- Knowles, Watson, Corne. **Reducing local optima in single-objective problems by multi-objectivization**. EMO 2001, LNCS Vol.1993, pp. 269–283, 2001
- Liefooghe, Daolio, Verel, Derbel, Aguirre, Tanaka. **Landscape-aware performance prediction for evolutionary multi-objective optimization**. IEEE Transactions on Evolutionary Computation, vol. 24, no. 6, pp. 1063-1077, 2020
- Liefooghe, Derbel, Verel, López-Ibáñez, Aguirre, Tanaka. **On Pareto local optimal solutions networks**. PPSN 2018, LNCS vol. 11102, pp. 232–244, Coimbra, Portugal, 2018
- Liefooghe, López-Ibáñez. **Many-objective (combinatorial) optimization is easy**. GECCO 2023, pp.704-712, Lisbon, Portugal, 2023

References (3)

- Liefooghe, López-Ibáñez, Paquete, Verel. **Dominance, epsilon, and hypervolume local optimal sets in multi-objective optimization, and how to tell the difference**. GECCO 2018, pp. 324-331, Kyoto, Japan, 2018
- Liefooghe, Ochoa, Verel, Derbel. **Pareto local optimal solutions networks with compression, enhanced visualization and expressiveness**. GECCO 2023, pp.713-721, Lisbon, Portugal, 2023
- Liefooghe, Verel, Derbel, Aguirre, Tanaka. **Dominance, indicator and decomposition based search for multi-objective QAP: landscape analysis and automated algorithm selection**. PPSN 2020, LNCS vol. 12269, pp. 33-47, Leiden, Netherlands, 2020
- Liefooghe, Verel, Lacroix, Zăvoianu, McCall. **Landscape features and automated algorithm selection for multi-objective interpolated continuous optimisation problems**. GECCO 2021, pp. 421-429, Lille, France, 2021
- Paquete, Schiavinotto, Stützle. **On local optima in multiobjective combinatorial optimization problems**. Annals of Operations Research, vol. 156, no. 1, pp. 83–97, 2007
- Paquete, Stützle. **Clusters of non-dominated solutions in multiobjective combinatorial optimization: An experimental analysis**. Multiobjective Programming and Goal Programming: Theoretical Results and Practical Applications, LNEMS vol. 618, pp. 69–77, 2009
- Schäpermeier, Grimme, Kerschke. **One PLOT to show them all: Visualization of efficient sets in multi-objective landscapes**. PPSN 2020, LNCS vol. 12269, pp. 154–167, Leiden, Netherlands, 2020
- Tanaka, Takadama, Sato. **Impacts of single-objective landscapes on multi-objective optimization**. CEC 2022, pp. 1-8, Padua, Italy, 2022
- Verel, Liefooghe, Jourdan, Dhaenens. **On the structure of multiobjective combinatorial search space: MNK-landscapes with correlated objectives**. European Journal of Operational Research, vol. 227, no. 2, pp. 331–342, 2013
- Vodopija, Tusar, Filipic. **Characterization of constrained continuous multiobjective optimization problems: A feature space perspective**. Information Science, vol. 607, pp. 244-262, 2022
- Zitzler, Thiele, Bader. **On set-based multiobjective optimization**. IEEE Transactions on Evolutionary Computation, vol. 14, no. 1, pp. 58–79, 2010

non-exhaustive list… any important reference missing? please let us know!

– Tutorial on Landscape Analysis for Explainable Optimization –

General Conclusions

Arnaud Liefooghe & Sébastien Verel

Université du Littoral Côte d'Opale – LISIC
arnaud.liefooghe@univ-littoral.fr, sebastien.verel@univ-littoral.fr

This work has benefited from the support of the National Research Agency under France 2030, MAIA Project ANR-22-EXES-0009

General Conclusions

- Landscape analysis: valuable tool for understanding / explaining problem difficulty and algorithm performance / behavior
- Bridge the gap between theory and practice

- Combinatorial vs. continuous landscapes, mixed landscapes
- Fitness vs. violation landscape for constraint-handling
- Landscape-aware automated algorithm selection and configuration
- Key issues in benchmarking: heterogeneous problems, algorithm complementarity, multiple performance measures, anytime…
- Fitness landscape for real-world applications (e.g. in ML / DL)

General Conclusions

- Automated perf. prediction, algorithm selection, configuration
  - Computationally intensive, repeated from scratch for each scenario
  - What have we learn from this?
- How about the knowledge acquired by EC researchers to make optimization more explainable?
  - … and EC algorithms more reliable?
- A prerequisite is interpretable landscape tools
  - can be complemented by XAI/XML
- Few (interpretable) features vs. many features
  - Unexplainable features: artifacts or unexpected discovery?
- Towards explainable landscape analysis (XLA) © Katherine ;)

Further Reading

algorithms

MDPI

Article

A Survey of Advances in Landscape Analysis for Optimisation

Katherine Mary Malan

Department of Decision Sciences, University of South Africa, Pretoria 0002, South Africa; malankm@unisa.ac.za

Abstract: Fitness landscapes were proposed in 1932 as an abstract notion for understanding biological evolution and were later used to explain evolutionary algorithm behaviour. The last ten years has seen the field of fitness landscape analysis develop from a largely theoretical idea in evolutionary computation to a practical tool applied in optimisation in general and more recently in machine learning. With this widened scope, new types of landscapes have emerged such as multiobjective landscapes, violation landscapes, dynamic and coupled landscapes and error landscapes. This survey is a follow-up from a 2013 survey on fitness landscapes and includes an additional 11 landscape analysis techniques. The paper also includes a survey on the applications of landscape analysis for understanding complex problems and explaining algorithm behaviour, as well as algorithm performance prediction and automated algorithm configuration and selection. The extensive use of landscape analysis in a broad range of areas highlights the wide applicability of the techniques and the paper discusses some opportunities for further research in this growing field.

Keywords: fitness landscape; landscape analysis; violation landscape; error landscape; automated algorithm selection

Landscape Analysis of Optimisation Problems and Algorithms

Katherine Malan[1] & Gabriela Ochoa[2]
[1]University of South Africa, Pretoria, South Africa
[2]University of Stirling, Stirling, UK
malankm@unisa.ac.za, gabriela.ochoa@stir.ac.uk

http://gecco-2023.sigevo.org/

GECCO '23 Companion, July 15 - 19, 2023, Lisbon, Portugal

ACM ISBN 979-8-4007-0120-7/23/07
https://doi.org/10.1145/3583133.3595051

GECCO