Advanced Control

March 14, 2013

École Centrale Paris, Châtenay-Malabry, France

slides partly inspired by E. Zitzler: "Bio-inspired Optimization and Design", lecture@ETH

Anne Auger
INRIA Saclay – Ile-de-France

Dimo Brockhoff
INRIA Lille – Nord Europe

Course Overview

Date		Topic
Fri, 11.1.2013	DB	Introduction to Control, Examples of Advanced Control, Introduction to Fuzzy Logic
Fri, 18.1.2013	DB	Fuzzy Logic (cont'd), Introduction to Artificial Neural Networks
Fri, 25.1.2013	~~AA~~ DB	Bio-inspired Optimization, discrete search spaces
Fri, 1.2.2013	AA	The Traveling Salesperson Problem
Fri, 22.2.2013	AA	Continuous Optimization I
Fri, 1.3.2013	AA	Continuous Optimization II
Fr, 8.3.2013	DB	Controlling a Pole Cart
Do, 14.3.2013	DB	Advanced Optimization: multiobjective optimization, constraints, ...
Tue, 19.3.2013		written exam (paper and computer)

next Tuesday exam at **! 9h45-13h00 !**

Remark to last lecture

All information also available at

`http://researchers.lille.inria.fr/~brockhof/advancedcontrol/`

(exercise sheets, lecture slides, additional information, links, ...)

(More) Advanced Concepts of Optimization

(Evolutionary) Multiobjective Optimization + MCDM
Constrained Optimization
Possible Thesis Projects

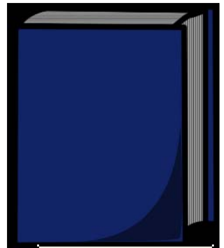
The Single-Objective Knapsack Problem

weight = 750g profit = 5	weight = 1500g profit = 8	weight = 300g profit = 7	weight = 1000g profit = 3

Fabien1309

Denae Bedard

Simon A. Eugster



PUBLIC DOMAIN

?

Joadl

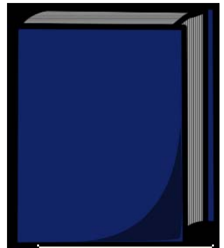
Single-objective Goal:

choose a subset that

- maximizes overall profit
- w.r.t a weight limit (constraint)

The Multiobjective Knapsack Problem

weight = 750g profit = 5	weight = 1500g profit = 8	weight = 300g profit = 7	weight = 1000g profit = 3



?

Single-objective Goal:

choose a subset that

- maximizes overall profit
- w.r.t a weight limit (constraint)

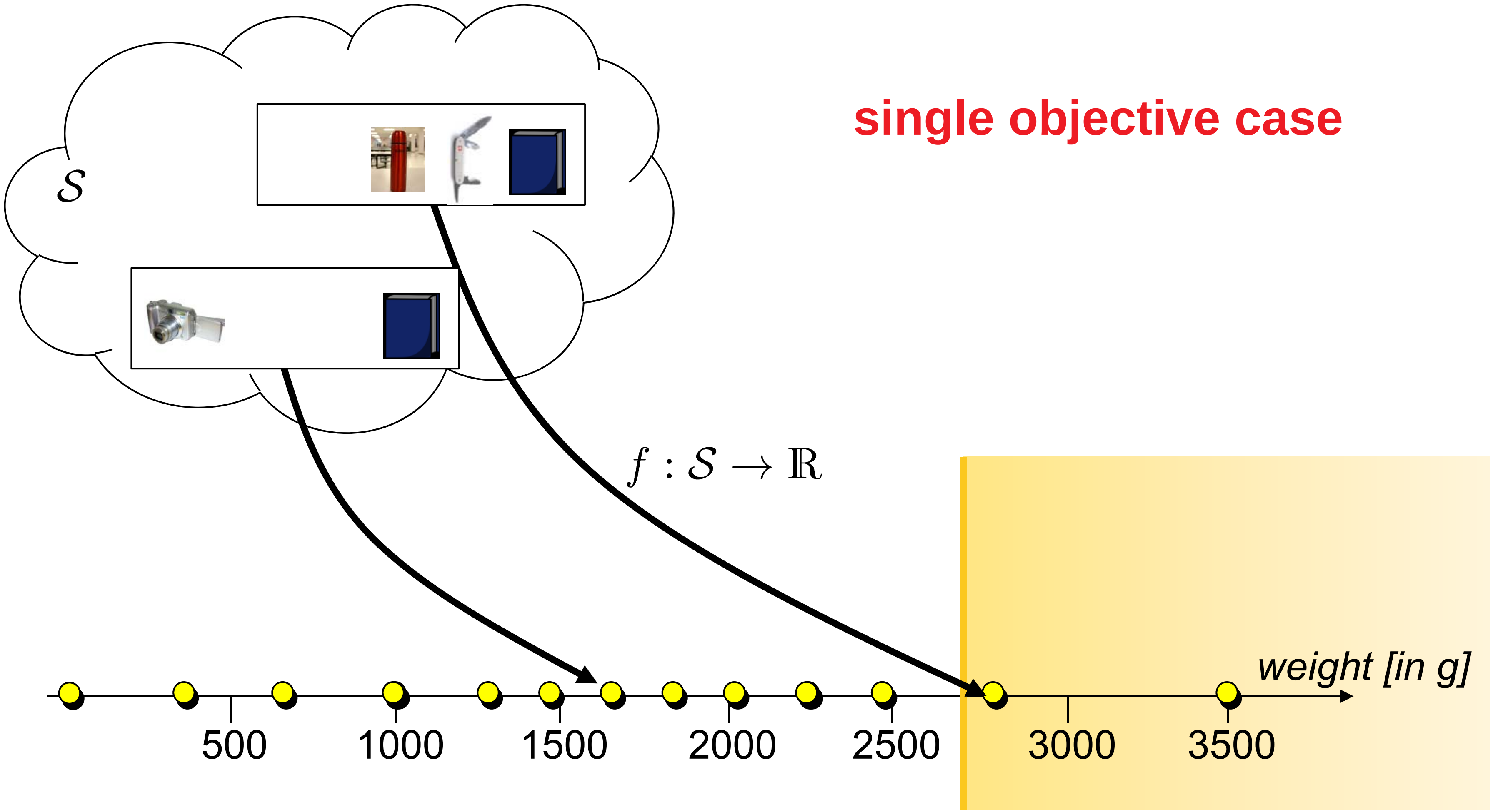
Multiobjective Goal:

Choose a subset that

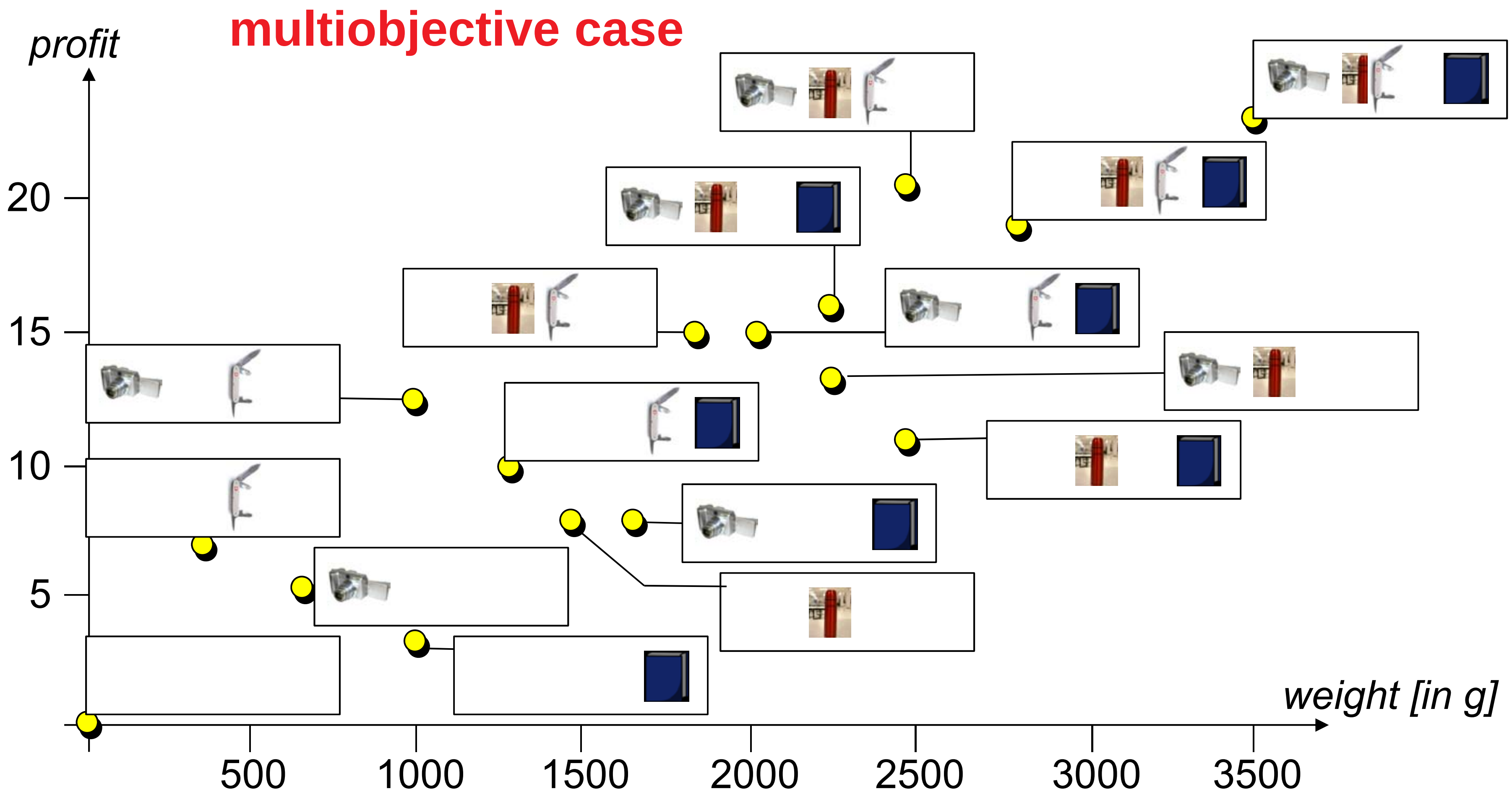
- maximizes overall profit
- minimizes overall weight

Knapsack problem: all solutions plotted

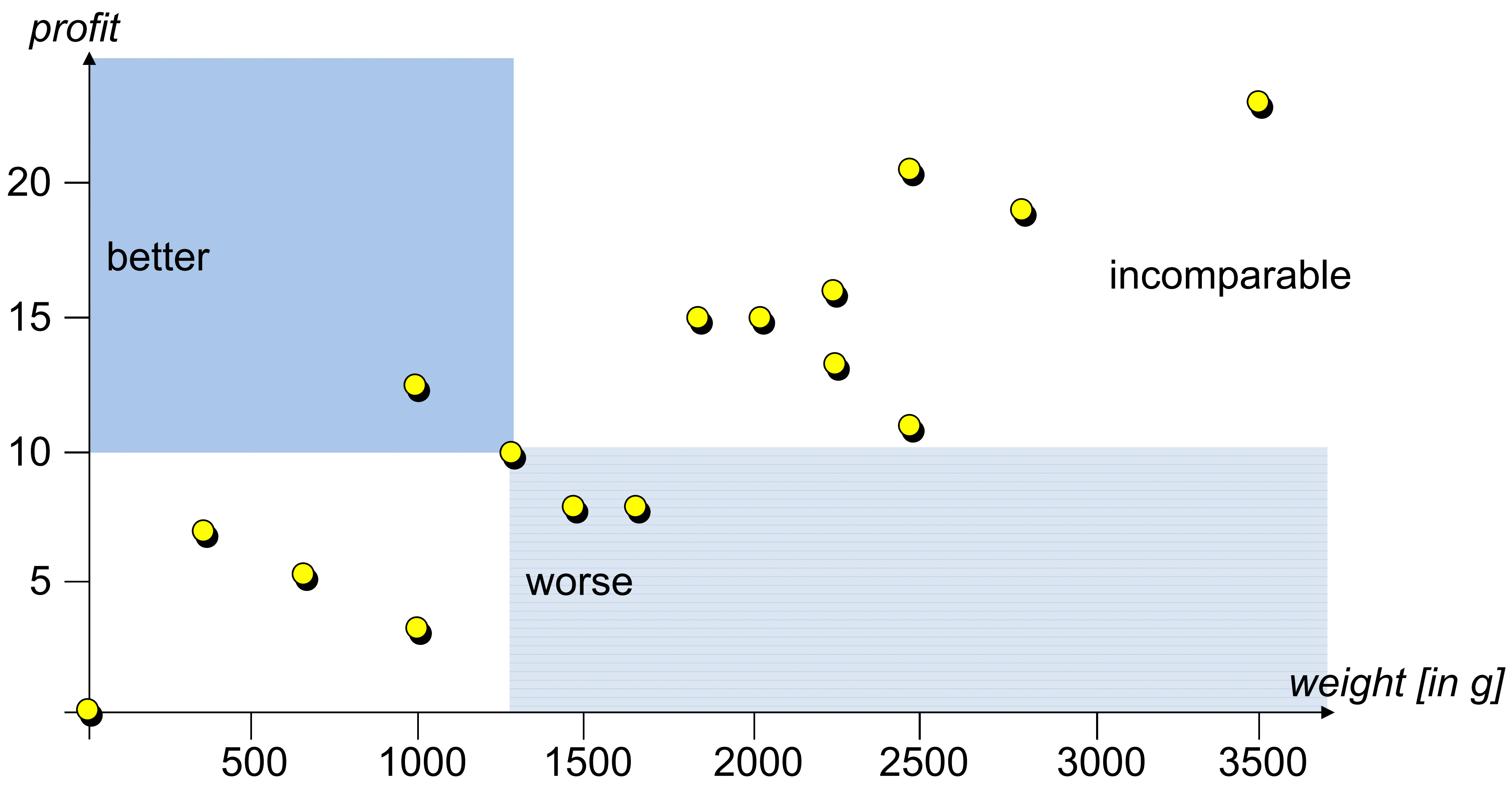
single objective case



Knapsack problem: all solutions plotted

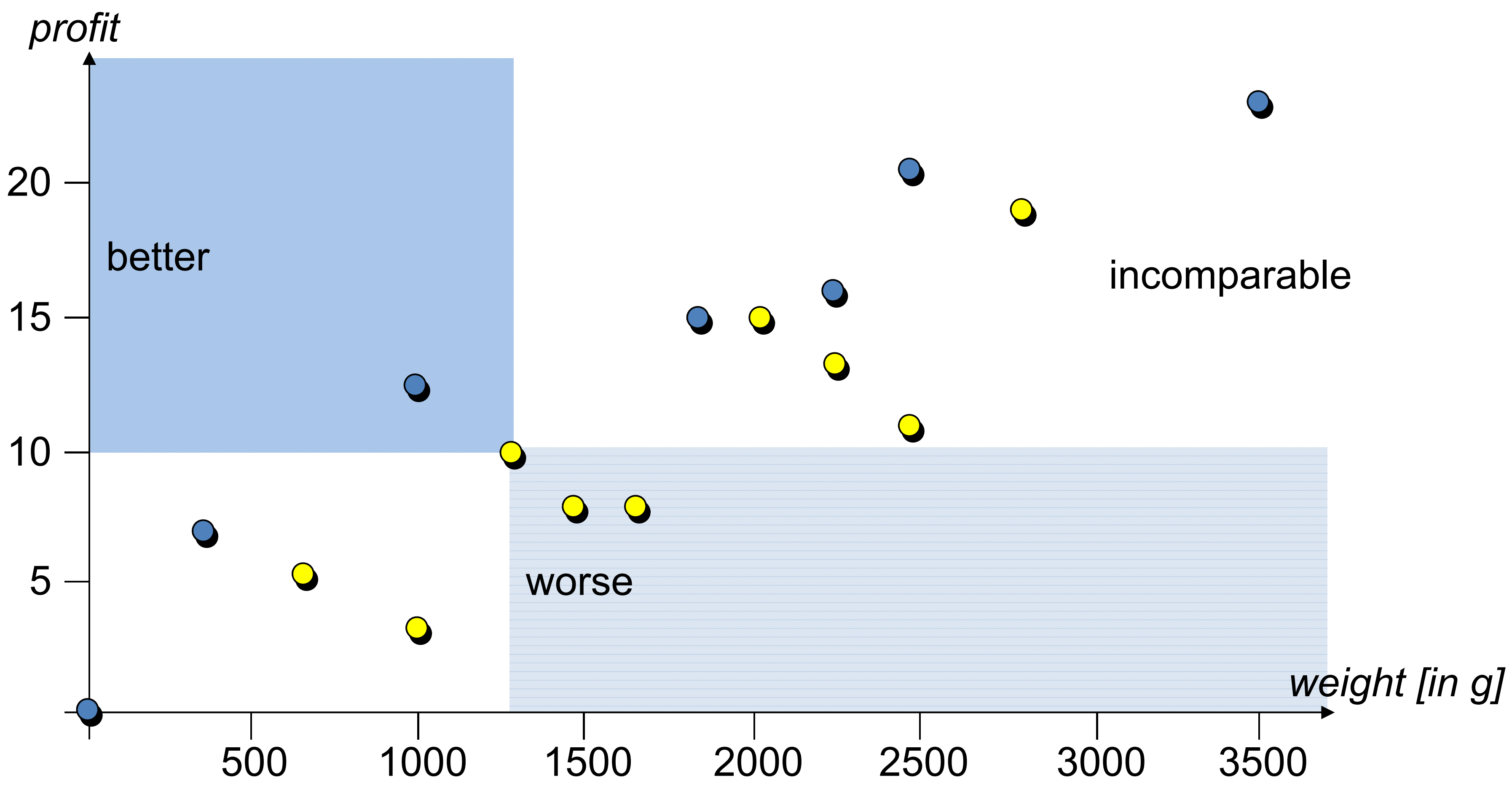


Knapsack problem: all solutions plotted

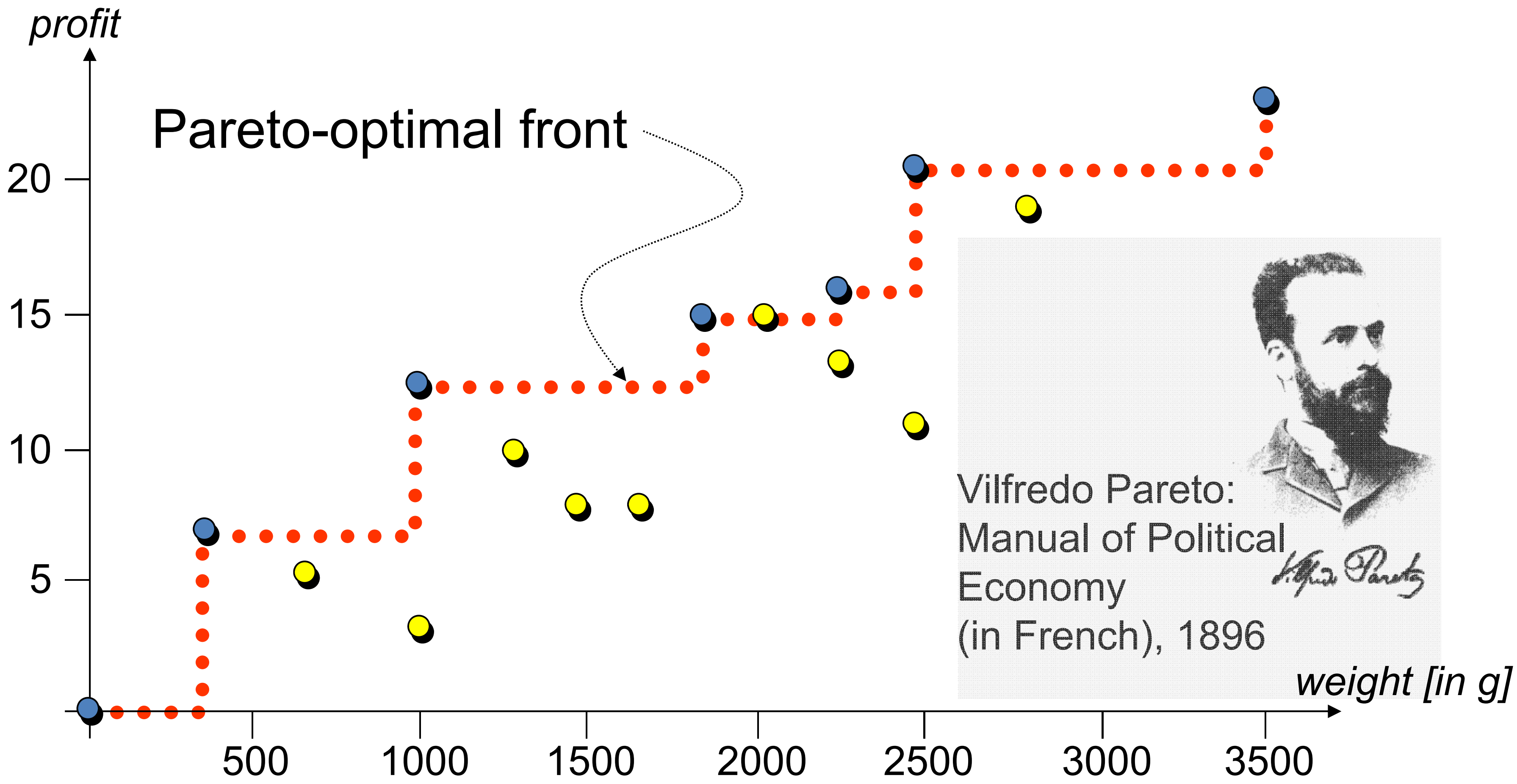


Principles of Multiple Criteria Decision Analysis

Observations: ❶ there is no single optimal solution, but
❷ some solutions (blue) are better than others (yellow)

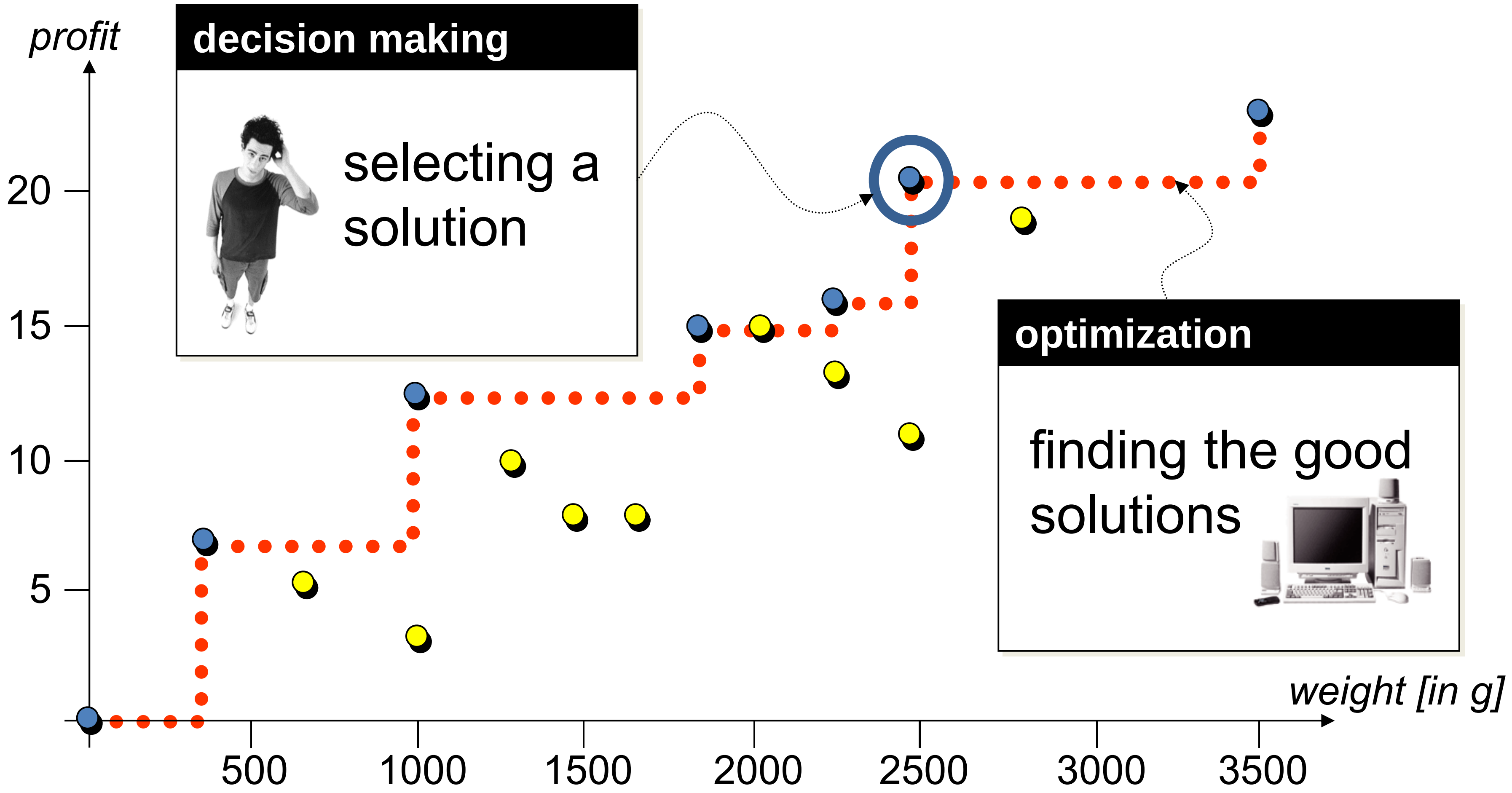


Observations: ❶ there is no single optimal solution, but
❷ some solutions (●) are better than others (●)



Principles of Multiple Criteria Decision Analysis

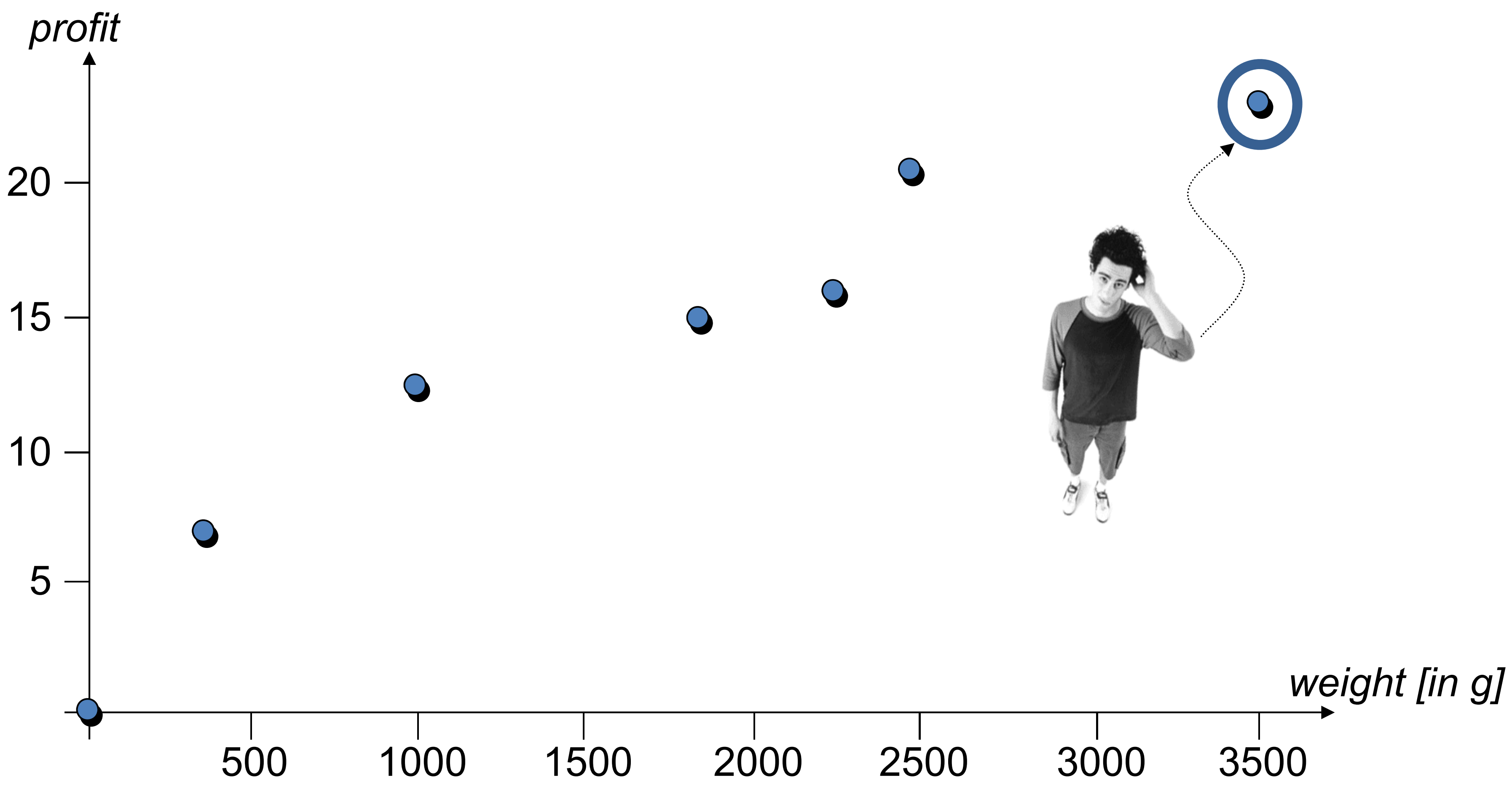
Observations: ❶ there is no single optimal solution, but
❷ some solutions (●) are better than others (●)



Decision Making: Selecting a Solution

Possible Approach:

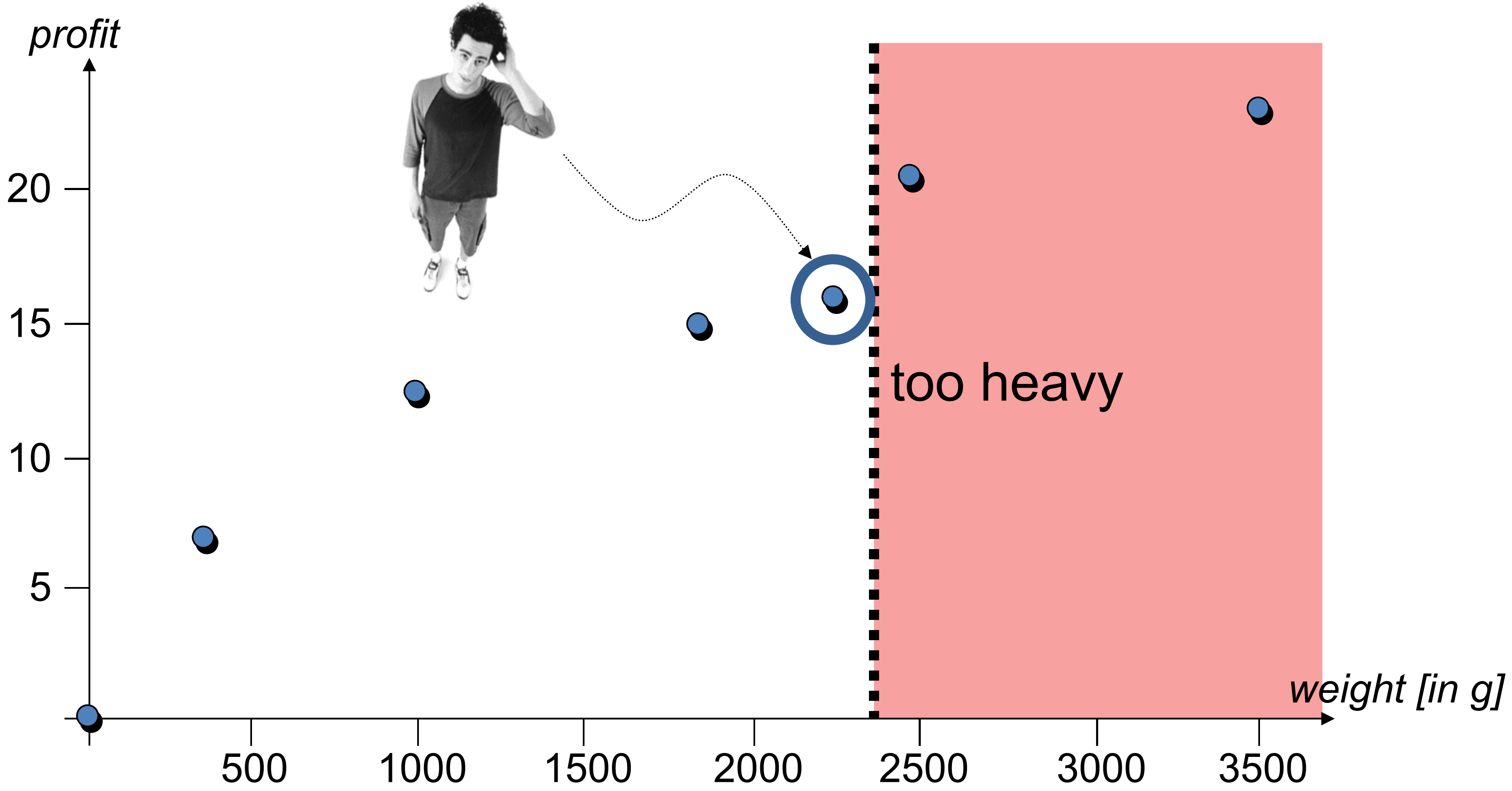
- profit more important than weight (ranking)



Decision Making: Selecting a Solution

Possible Approach:

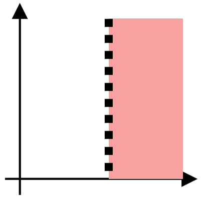
- profit more important than weight (ranking)
- weight must not exceed 2400 (constraint)



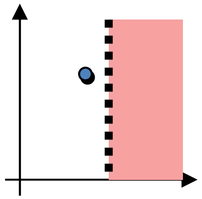
When to Make the Decision

Before Optimization:

rank objectives,
define constraints,…

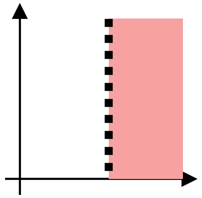


search for one
(blue) solution



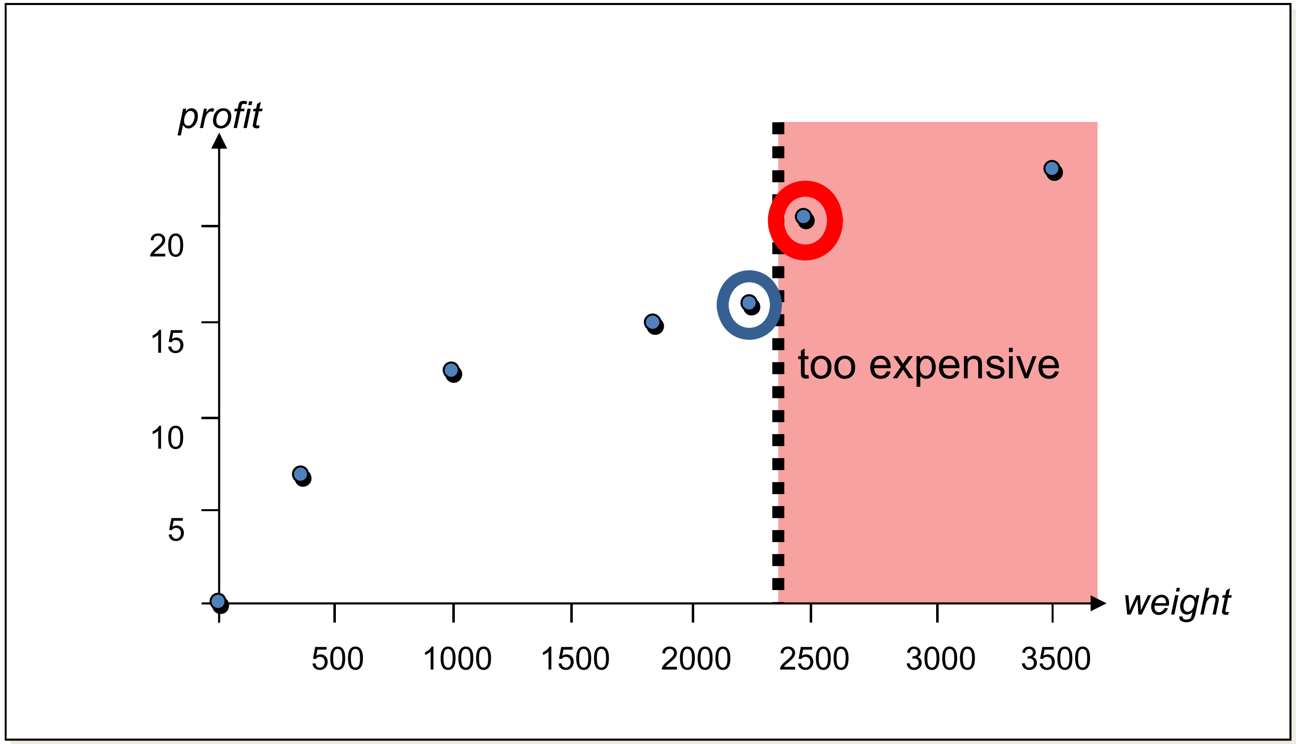
When to Make the Decision

Before Optimization:



rank objectives,
define constraints,…

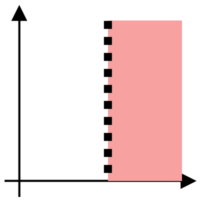
search for one
(blue) solution



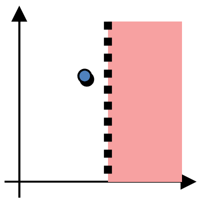
When to Make the Decision

Before Optimization:

rank objectives,
define constraints,…

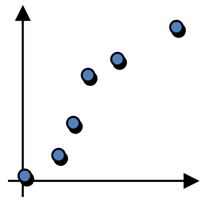


search for one
(blue) solution

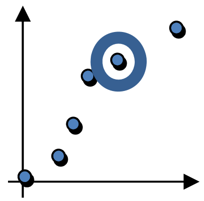


After Optimization:

search for a set of
(blue) solutions



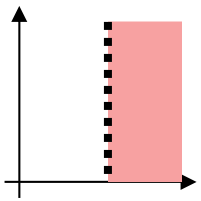
select one solution
considering
constraints, etc.



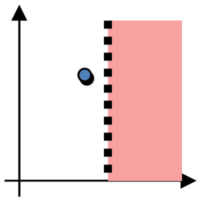
When to Make the Decision

Before Optimization:

rank objectives,
define constraints,…

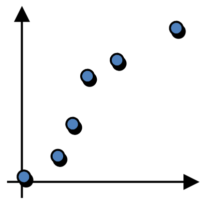


search for one
(blue) solution

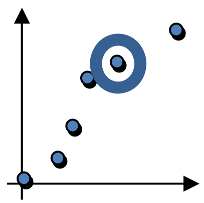


After Optimization:

search for a set of
(blue) solutions



select one solution
considering
constraints, etc.



Focus: learning about a problem

- trade-off surface
- interactions among criteria
- structural information

Multiple Criteria Decision Making (MCDM)

Definition: MCDM

MCDM can be defined as the study of methods and procedures by which concerns about multiple conflicting criteria can be formally incorporated into the management planning process

International Society on Multiple Criteria Decision Making

model

$$\min_x \, [\mu_1(x), \mu_2(x), \ldots, \mu_n(x)]^T$$

$s.t.$

$g(x) \leq 0$

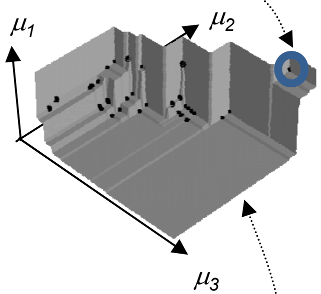
$h(x) = 0$

$x_l \leq x \leq x_u$

decision making

(exact) optimization

trade-off surface

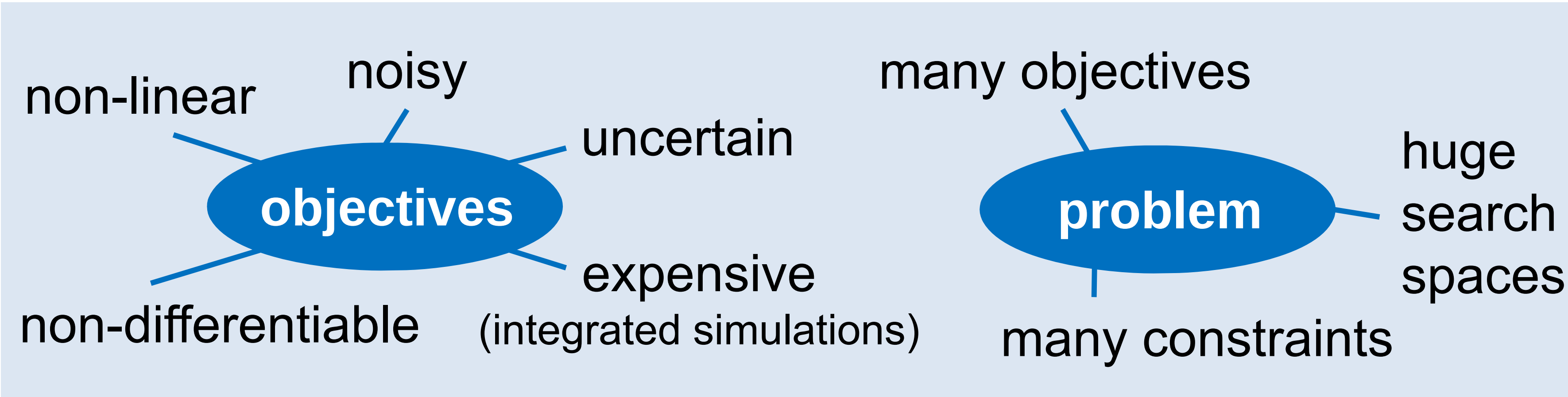


Multiple Criteria Decision Making (MCDM)

Definition: MCDM

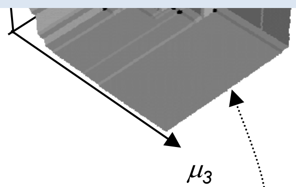
MCDM can be defined as the study of methods and procedures by which concerns about multiple conflicting criteria can be formally incorporated into the management planning process

International Society on Multiple Criteria Decision Making



$g(x) \leq 0$
$h(x) = 0$
$x_l \leq x \leq x_u$

(exact) optimization



Multiple Criteria Decision Making (MCDM)

Definition: MCDM

MCDM can be defined as the study of methods and procedures by which concerns about multiple conflicting criteria can be formally incorporated into the management planning process

International Society on Multiple Criteria Decision Making

non-linear

noisy

many objectives

uncerta

objectives

expens

non-differentiable

(integrated si

$g(x) \leq 0$
$h(x) = 0$
$x_l \leq x \leq x_u$

(exact) optim

Black box optimization

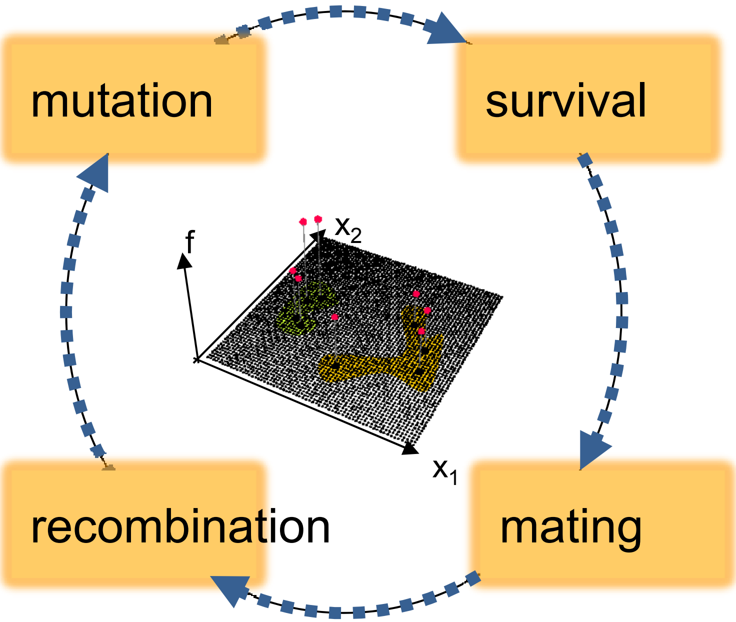
$x \in X \rightarrow f \rightarrow (f_1(x), \ldots, f_k(x))$

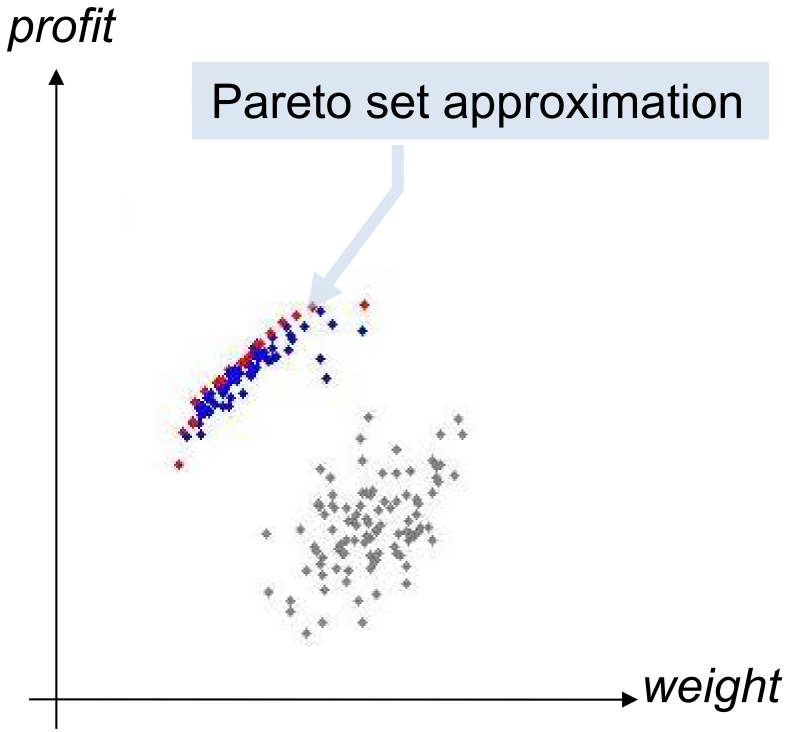
only mild assumptions

Evolutionary Multiobjective Optimization (EMO)

Definition: EMO

EMO = **evolutionary algorithms** / randomized search algorithms
- applied to multiple criteria decision making (in general)
- used to approximate the Pareto-optimal set (mainly)

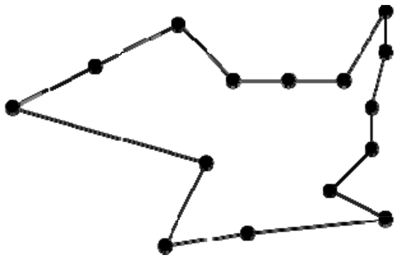




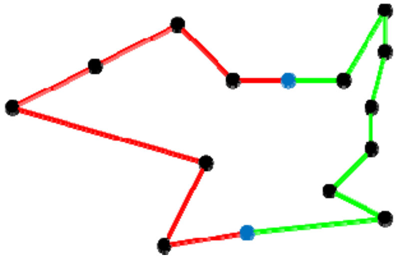
Multiobjectivization

Some problems are easier to solve in a multiobjective scenario

example: TSP
[Knowles et al. 2001]



$\pi \in S_n \rightarrow f(\pi)$



$\pi \in S_n \rightarrow (f_1(\pi, a, b), f_2(\pi, a, b))$

Multiobjectivization

- by addition of new “helper objectives” [Jensen 2004]
  - job-shop scheduling [Jensen 2004], frame structural design [Greiner et al. 2007], theoretical (runtime) analyses [Brockhoff et al. 2009]
- by decomposition of the single objective
  - TSP [Knowles et al. 2001], minimum spanning trees [Neumann and Wegener 2006], protein structure prediction [Handl et al. 2008a], theoretical (runtime) analyses [Handl et al. 2008b]

Overview EMO part

The Big Picture

Basic Principles of Multiobjective Optimization

- algorithm design principles and concepts
- performance assessment

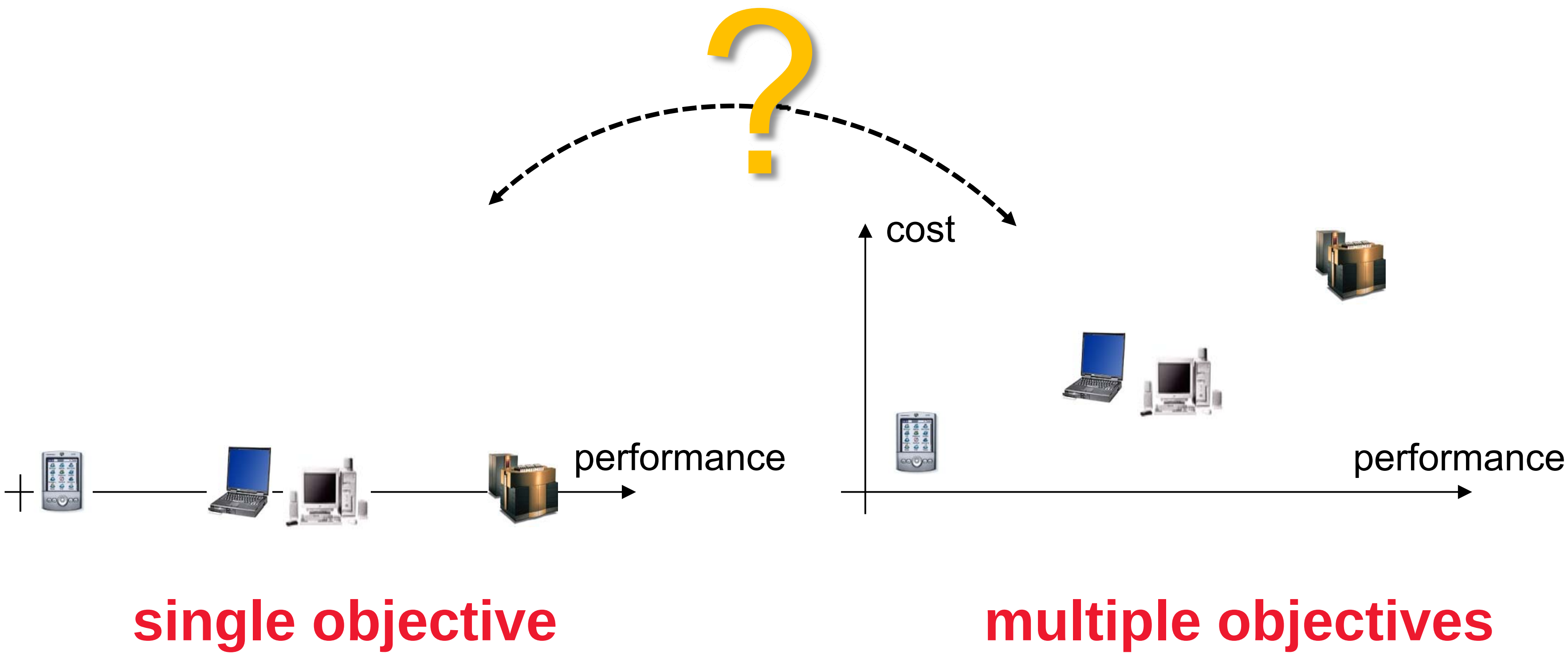
Selected Advanced Concepts

- indicator-based EMO
- preference articulation

A Few Examples From Practice

Starting Point

What makes evolutionary multiobjective optimization different from single-objective optimization?



A General (Multiobjective) Optimization Problem

A multiobjective optimization problem: $(X, Z, \mathbf{f}, \mathbf{g}, \leq)$

X	search / parameter / decision space
$Z = \mathbb{R}^n$	objective space
$\mathbf{f} = (f_1, \ldots, f_n)$	vector-valued objective function with $f_i : X \mapsto \mathbb{R}$
$\mathbf{g} = (g_1, \ldots, g_m)$	vector-valued constraint function with $g_i : X \mapsto \mathbb{R}$
$\leq \subseteq Z \times Z$	binary relation on objective space

Goal: find decision vector(s) $\mathbf{a} \in X$ such that

❶ for all $1 \leq i \leq m : g_i(\mathbf{a}) \leq 0$ and

❷ for all $\mathbf{b} \in X : \mathbf{f}(\mathbf{b}) \leq \mathbf{f}(\mathbf{a}) \Rightarrow \mathbf{f}(\mathbf{a}) \leq \mathbf{f}(\mathbf{b})$

A Single-Objective Optimization Problem

decision space — objective space — objective function

total order

$$(X, Z, f: X \rightarrow Z, rel \subseteq Z \times Z)$$

A Single-Objective Optimization Problem

decision space — objective space — objective function

total order

$$(X, Z, f: X \rightarrow Z, rel \subseteq Z \times Z)$$

$$(X, prefrel)$$

total preorder where

$a \text{ prefrel } b \Leftrightarrow f(a) \text{ rel } f(b)$

A Single-Objective Optimization Problem

Example: ONEMAX Problem

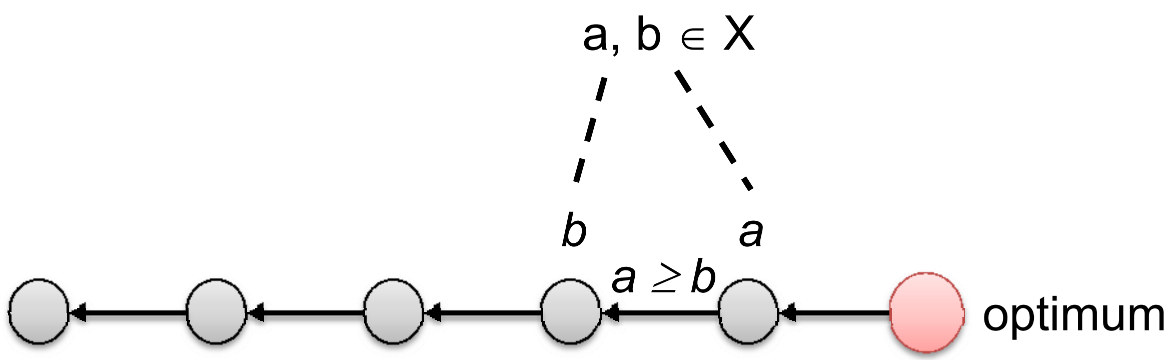
$(X, Z, f: X \rightarrow Z, rel \subseteq Z \times Z)$

$(X, prefrel)$

$(\{0,1\}^n, \{0,1, 2, ..., n\}, f_{OM}, \geq)$ where $f_{OM}(a) = \sum_i a_i$

Simple Graphical Representation

Example: $\geq$ (total order)



totally ordered

≤ is a total order if

1) *a* ≤ *b* and *b* ≤ *a* then a = b (antisymmetry),
2) a ≤ b and b ≤ c then a ≤ c (transitivity), and
3) a ≤ b or b ≤ a (totality).

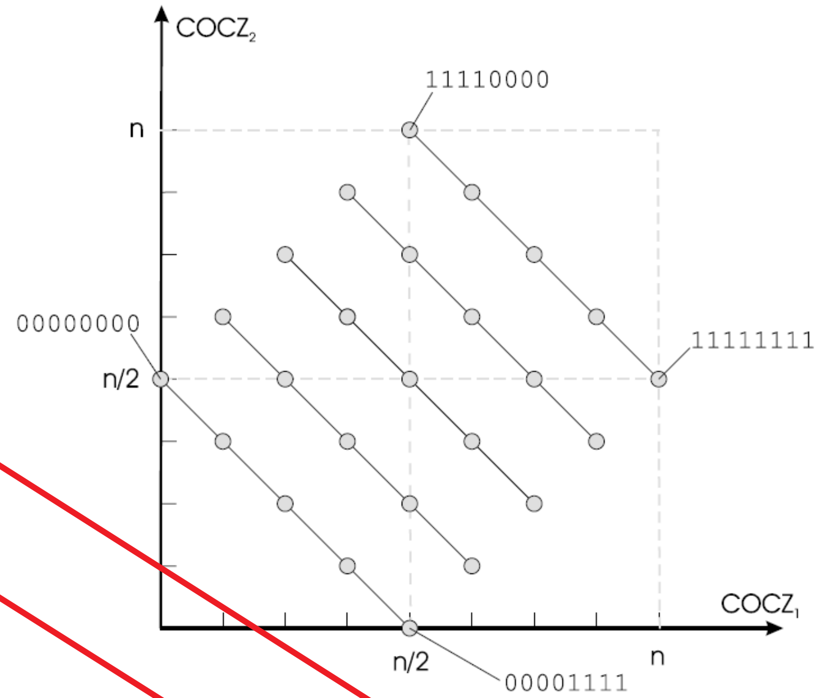
see, e.g., wikipedia

A Multiobjective Optimization Problem

Example: Counting Ones Counting Zeros Problem (COCZ)

(X, Z, f: X → Z, rel ⊆ Z × Z)

(X, prefrel)

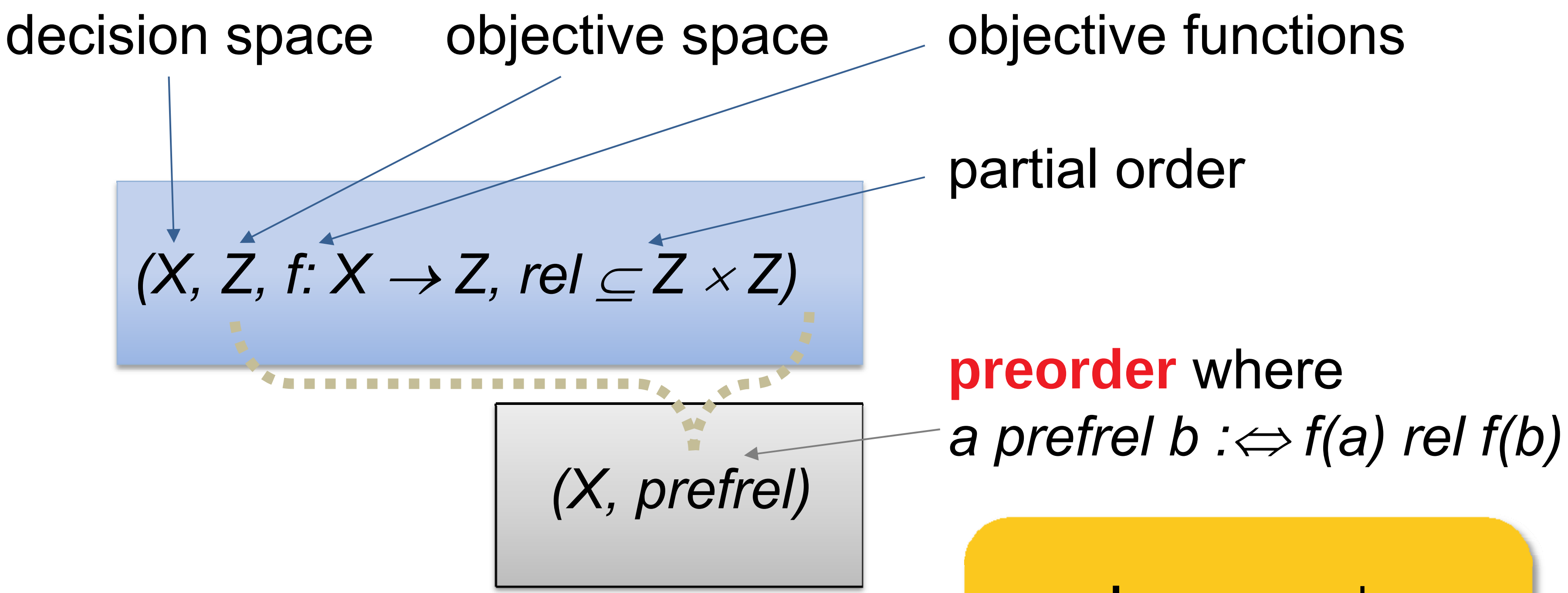


copyright: M. Laumanns, L. Thiele, and E. Zitzler, 2003

$(\{0,1\}^n$ [n even], $\{0,1,2,\ldots,n\} \times \{0,1,2,\ldots,n\}$, (f_{OM}, f_{ZM}), **?**)

$$f_{\mathrm{OM}}(a) = \sum_{i=1}^{n} a_i \qquad f_{\mathrm{OM}}(a) = \sum_{i=1}^{n/2} a_i + \sum_{i=n/2+1}^{n} (1 - a_i)$$

Preference Relations



preorder on search space
induced by **partial order**
on objective space

Parenthesis: Relations

≤ is a **preorder** if

1) a ≤ *a* (reflexivity) for all *a* in P and

2) a ≤ b and b ≤ c then a ≤ c (transitivity)

≤ is a **partial order** if

1) a ≤ *a* (reflexivity) for all *a* in P,

2) a ≤ *b* and *b* ≤ *a* then *a* = *b* (antisymmetry), and

3) a ≤ *b* and *b* ≤ *c* then *a* ≤ *c* (transitivity).

see, e.g., wikipedia

Preference Relations

decision space   objective space   objective functions

partial order

$(X, Z, f: X \rightarrow Z, rel \subseteq Z \times Z)$

$(X, prefrel)$

preorder where

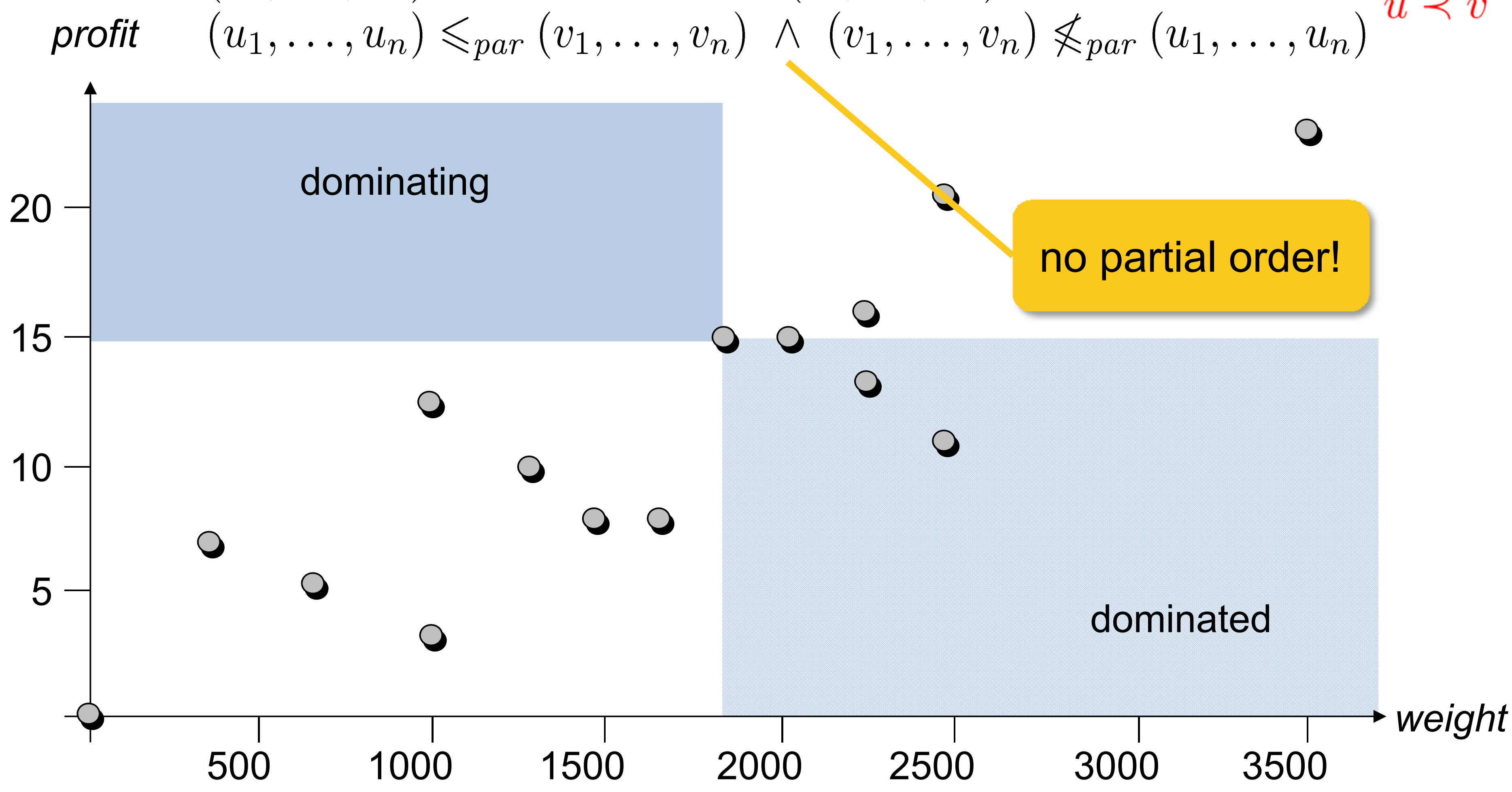
$a \; prefrel \; b :\Leftrightarrow f(a) \; rel \; f(b)$

preorder on search space **induced** by **partial order** on objective space

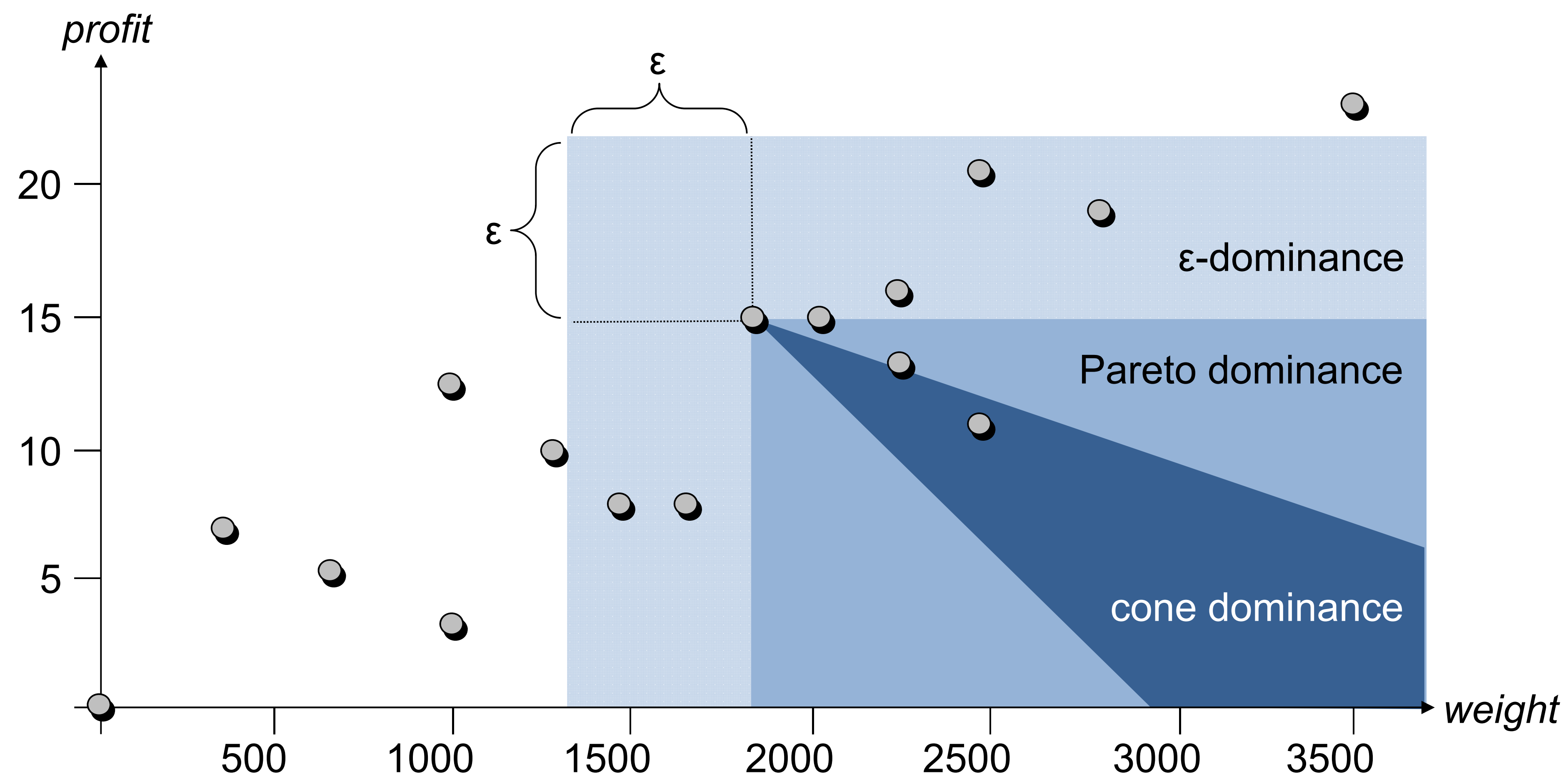
Pareto Dominance

$(u_1, \ldots, u_n)$ *weakly Pareto dominates* $(v_1, \ldots, v_n)$:
$$(u_1, \ldots, u_n) \leqslant_{par} (v_1, \ldots, v_n) :\Leftrightarrow \forall 1 \leq i \leq n : u_i \leq v_i$$
$u \preceq v$

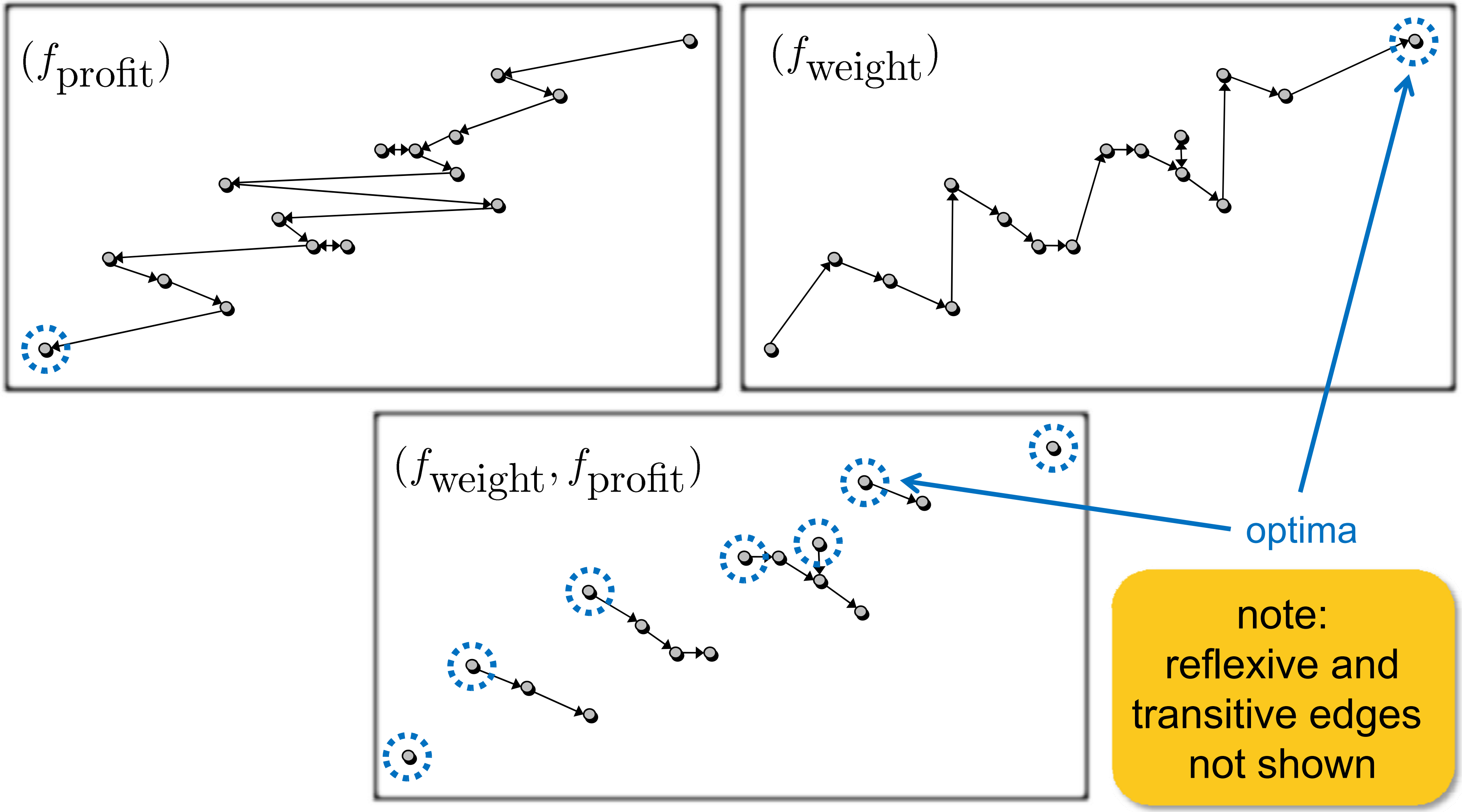
$(u_1, \ldots, u_n)$ *Pareto dominates* $(v_1, \ldots, v_n)$:
$$(u_1, \ldots, u_n) \leqslant_{par} (v_1, \ldots, v_n) \;\wedge\; (v_1, \ldots, v_n) \not\leqslant_{par} (u_1, \ldots, u_n)$$
$u \prec v$



Different Notions of Dominance



Visualizing Preference Relations



weak Pareto dominance: $(X, \preccurlyeq)$

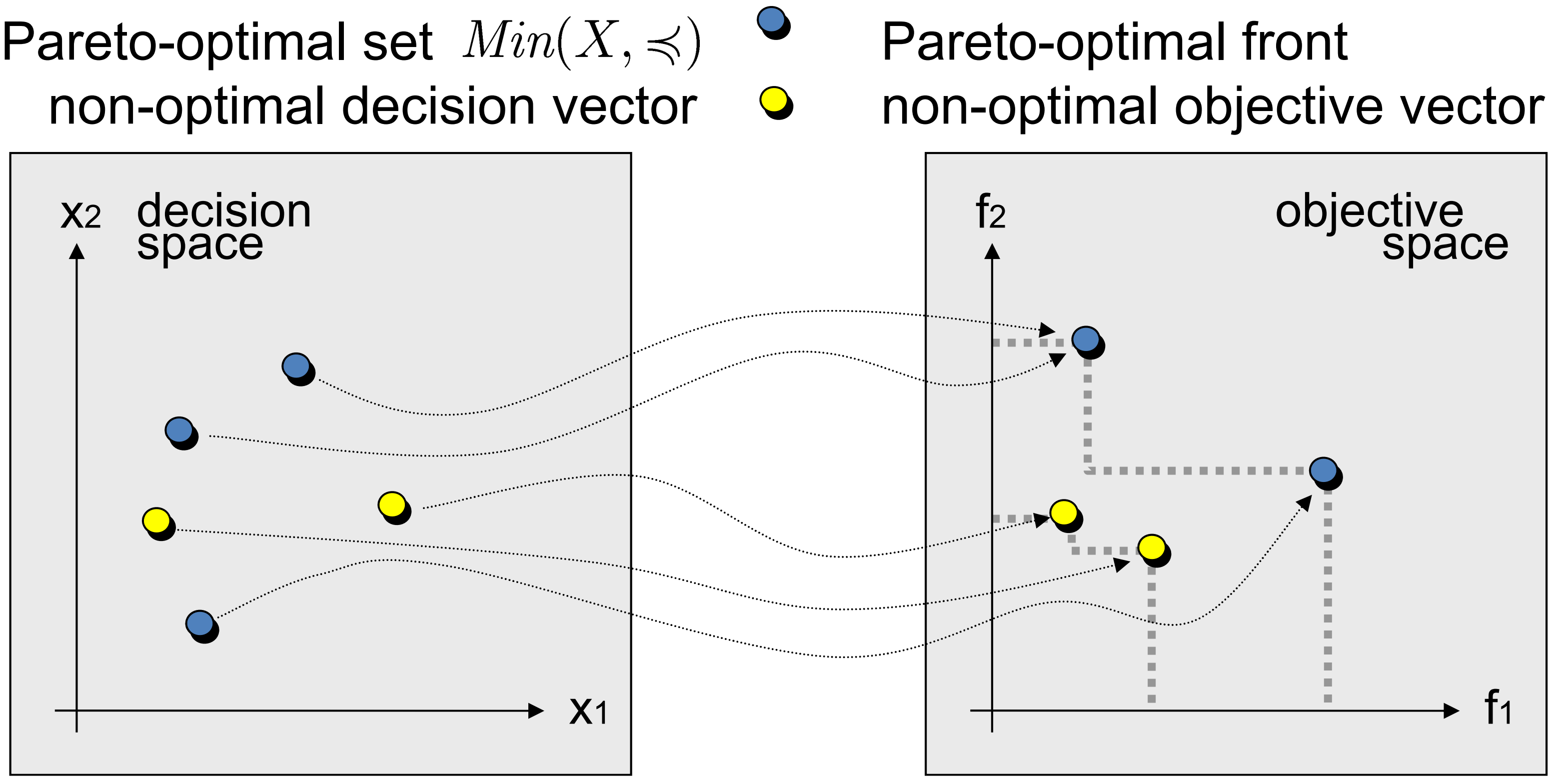
The Pareto-optimal Set

The *minimal set* of a preordered set $(Y, \leqq)$ is defined as

$$Min(Y, \leqq) := \{a \in Y \mid \forall b \in Y : b \leqq a \Rightarrow a \leqq b\}$$

Pareto-optimal set $Min(X, \preccurlyeq)$ — Pareto-optimal front

non-optimal decision vector — non-optimal objective vector

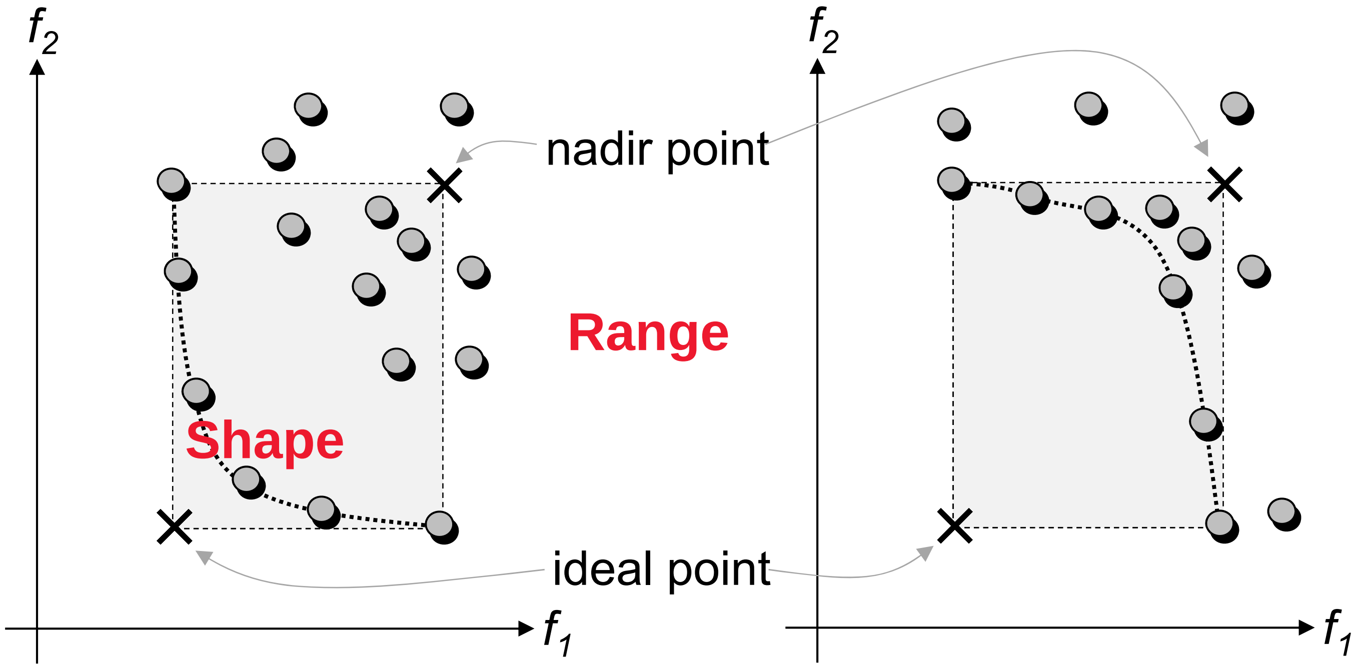


Remark: Properties of the Pareto Set

Computational complexity:

multiobjective variants can become NP- and #P-complete

Size: Pareto set can be exponential in the input length (e.g. shortest path [Serafini 1986], MSP [Camerini et al. 1984])



Approaches To Multiobjective Optimization

A multiobjective problem is as such underspecified

…because not any Pareto-optimum is equally suited!

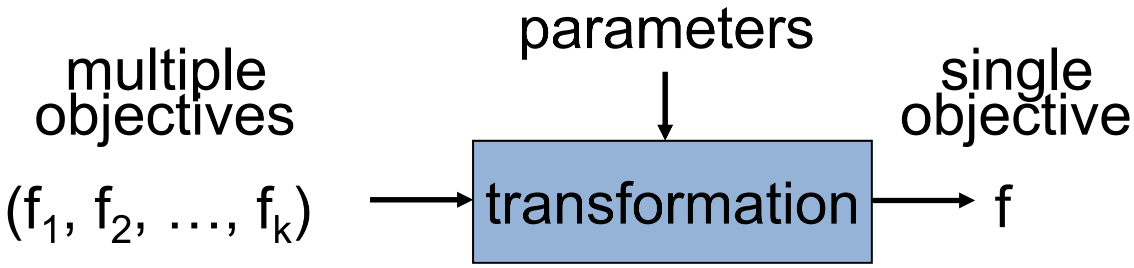
Additional preferences are needed to tackle the problem:

Solution-Oriented Problem Transformation:
Induce a total order on the decision space, e.g., by aggregation.

Set-Oriented Problem Transformation:
First transform problem into a set problem and then define an objective function on sets.

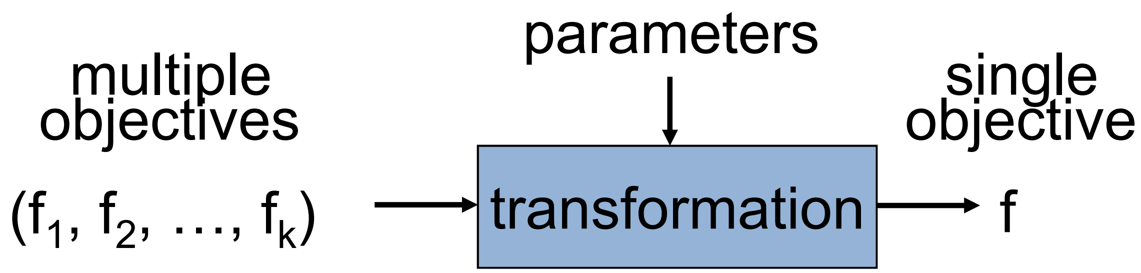
Preferences are needed in any case, but the latter are weaker!

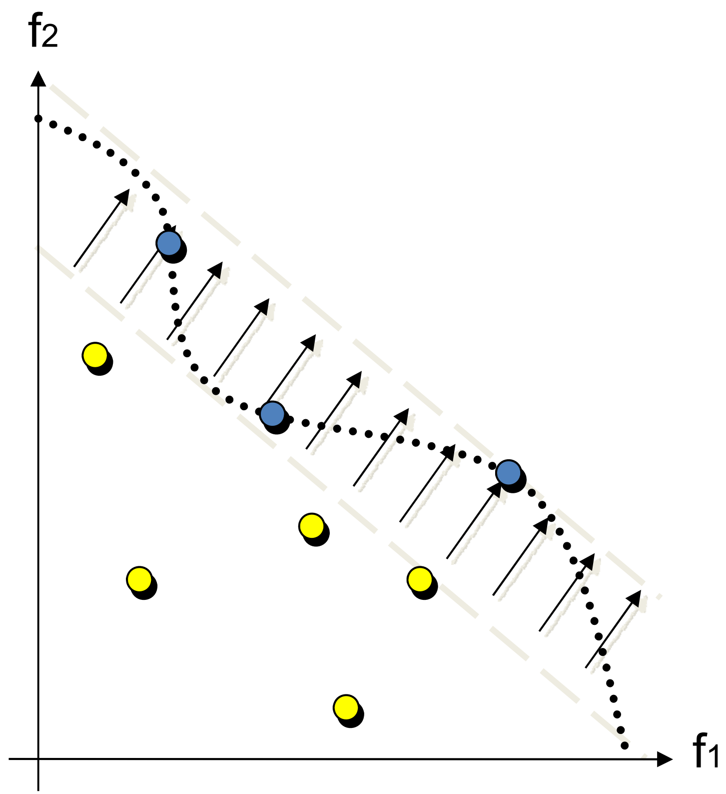
Solution-Oriented Problem Transformations



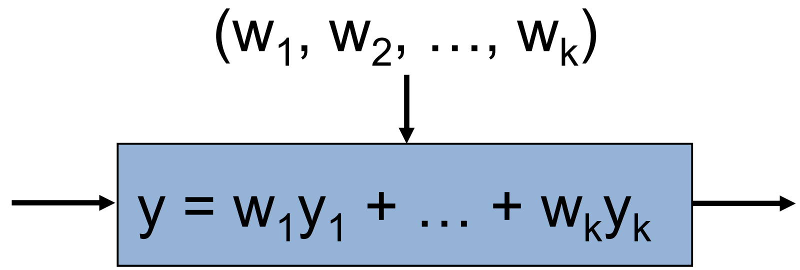
A *scalarizing function* s is a function $s : Z \mapsto \mathbb{R}$ that maps each objective vector $(u_1, \ldots, u_n) \in Z$ to a real value $s(u_1, \ldots, u_n) \in \mathbb{R}$.

Aggregation-Based Approaches





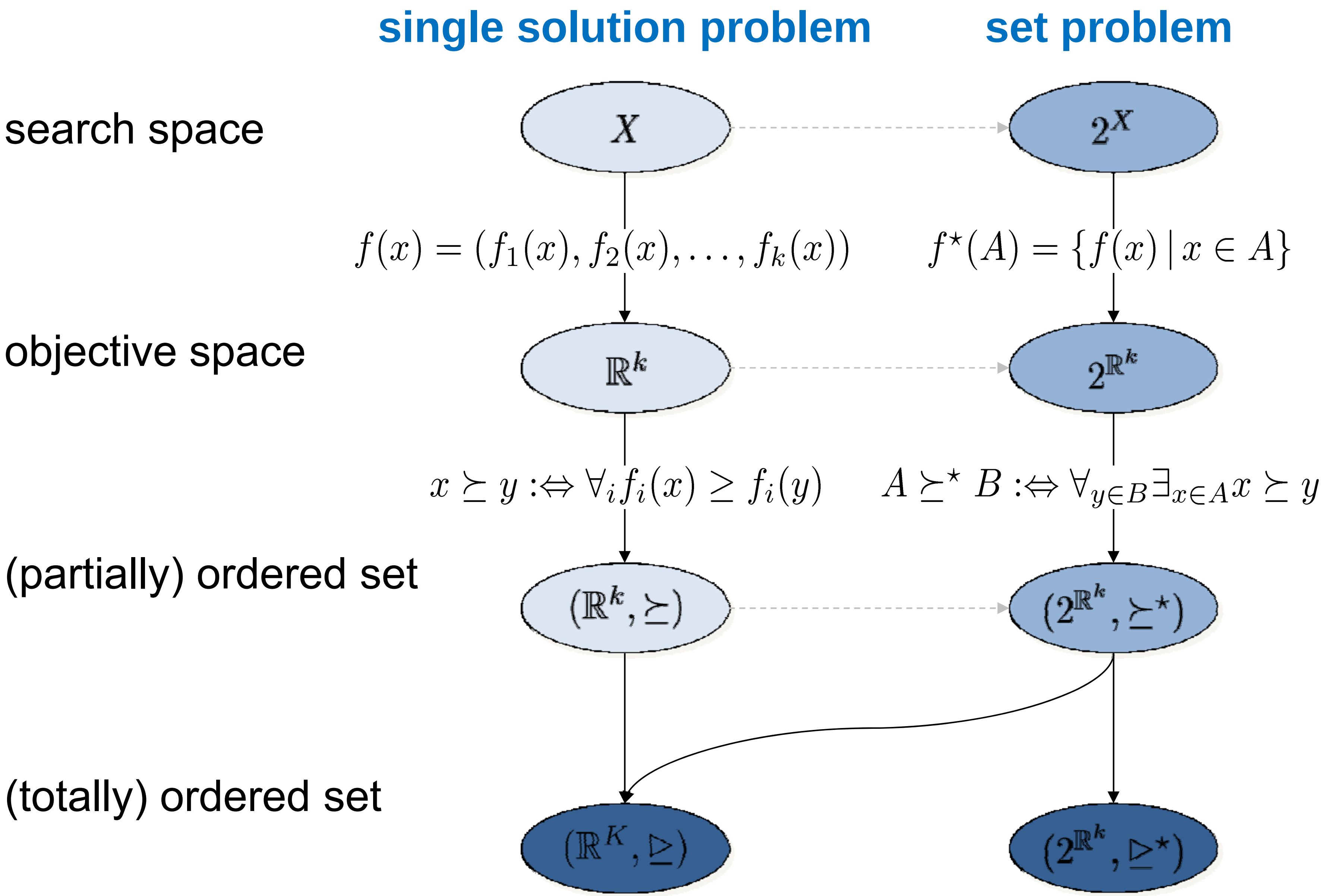
Example: weighting approach



Other example: Tchebycheff

$$y = \max w_i(u_i - z_i)$$

Problem Transformations and Set Problems



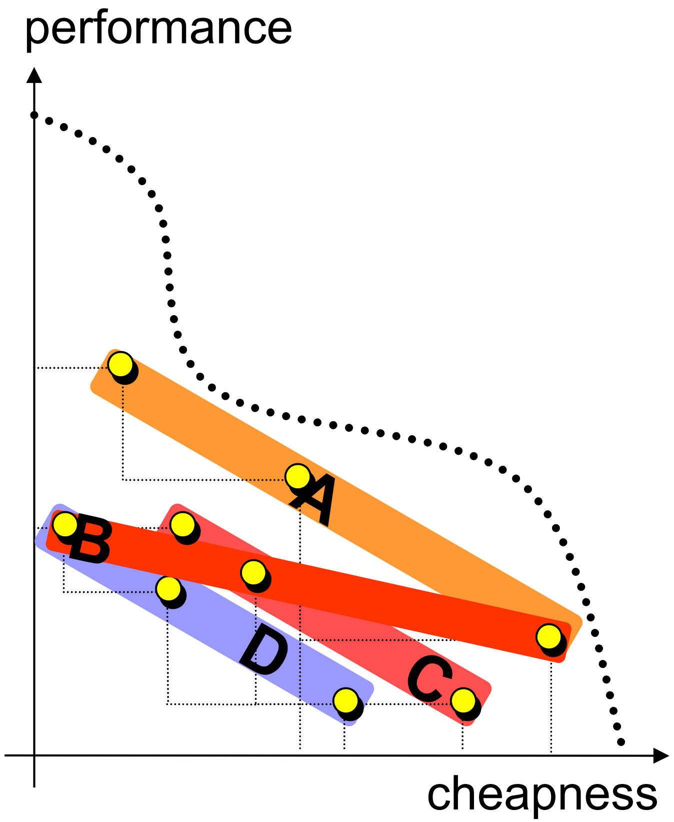
Set-Oriented Problem Transformations

For a multiobjective optimization problem $(X, Z, \mathbf{f}, \mathbf{g}, \leqslant)$, the associated *set problem* is given by $(\Psi, \Omega, F, \mathbf{G}, \preccurlyeq)$ where

- $\Psi = 2^X$ is the space of decision vector sets, i.e., the powerset of X,
- $\Omega = 2^Z$ is the space of objective vector sets, i.e., the powerset of Z,
- F is the extension of $\mathbf{f}$ to sets, i.e., $F(A) := \{\mathbf{f}(\mathbf{a}) \ : \ \mathbf{a} \in A\}$ for $A \in \Psi$,
- $\mathbf{G} = (G_1, \ldots, G_m)$ is the extension of $\mathbf{g}$ to sets, i.e., $G_i(A) := \max\{g_i(\mathbf{a}) \ : \ \mathbf{a} \in A\}$ for $1 \leq i \leq m$ and $A \in \Psi$,
- $\preccurlyeq$ extends $\leqslant$ to sets where $A \preccurlyeq B :\Leftrightarrow \forall \mathbf{b} \in B \ \exists \mathbf{a} \in A : \mathbf{a} \leqslant \mathbf{b}$.

Pareto Set Approximations

Pareto set approximation (algorithm outcome) =
set of (usually incomparable) solutions



A weakly dominates B
= not worse in all objectives and sets not equal

C dominates D
= better in at least one objective

A strictly dominates C
= better in all objectives

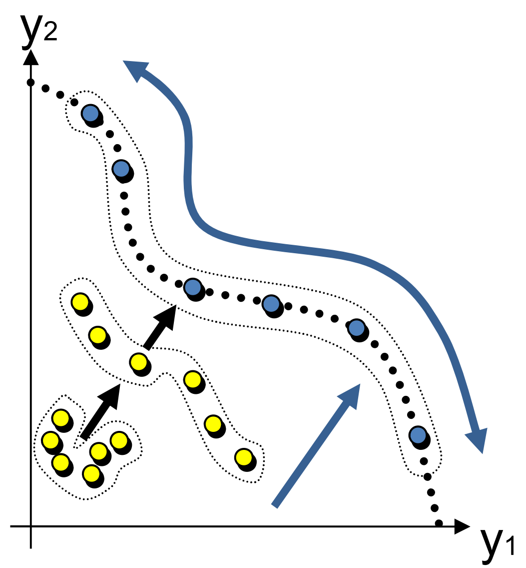
B is **incomparable** to C
= neither set weakly better

What Is the Optimization Goal (Total Order)?

- Find all Pareto-optimal solutions?
  - Impossible in continuous search spaces
  - How should the decision maker handle 10000 solutions?
- Find a representative subset of the Pareto set?
  - Many problems are NP-hard
  - What does representative actually mean?
- Find a good approximation of the Pareto set?
  - What is a good approximation?
  - How to formalize intuitive understanding:

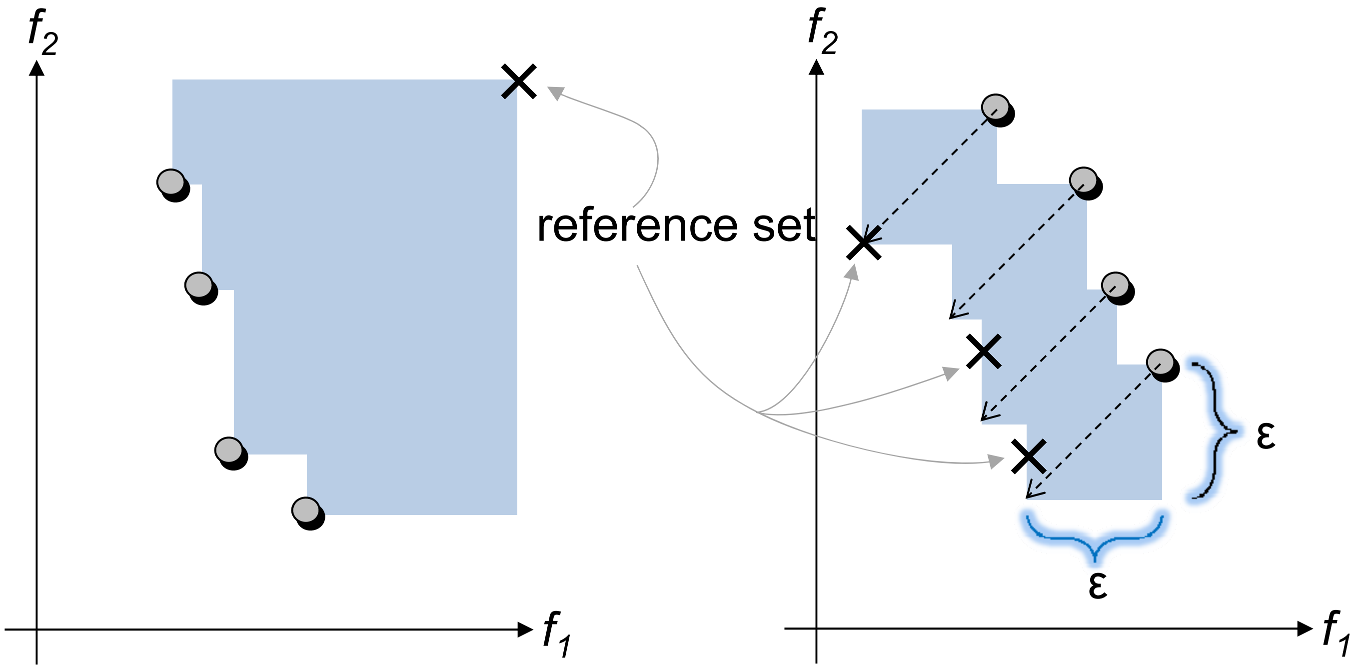
    ❶ close to the Pareto front

    ❷ well distributed



Quality of Pareto Set Approximations

A (unary) *quality indicator* I is a function $I : \Psi \mapsto \mathbb{R}$ that assigns a Pareto set approximation a real value.



hypervolume indicator

epsilon indicator

General Remarks on Problem Transformations

Idea:

Transform a preorder into a total preorder

Methods:

- Define single-objective function based on the multiple criteria *(shown on the previous slides)*
- Define any total preorder using a relation *(not discussed before)*

Question:

Is any total preorder ok resp. are there any requirements concerning the resulting preference relation?

$\Rightarrow$ Underlying dominance relation *rel* should be reflected

Refinements and Weak Refinements

❶ $\overset{\text{ref}}{\preceq}$ **refines** a preference relation $\preceq$ iff

$$A \preceq B \wedge B \not\preceq A \Rightarrow A \overset{\text{ref}}{\preceq} B \wedge B \overset{\text{ref}}{\not\preceq} A \qquad \textit{(better} \Rightarrow \textit{better)}$$

$\Rightarrow$ fulfills requirement

❷ $\overset{\text{ref}}{\preceq}$ **weakly refines** a preference relation $\preceq$ iff

$$A \preceq B \wedge B \not\preceq A \Rightarrow A \overset{\text{ref}}{\preceq} B \qquad \textit{(better} \Rightarrow \textit{weakly better)}$$

$\Rightarrow$ does not fulfill requirement, but $\overset{\text{ref}}{\preceq}$ does not contradict $\preceq$

…sought are total refinements…

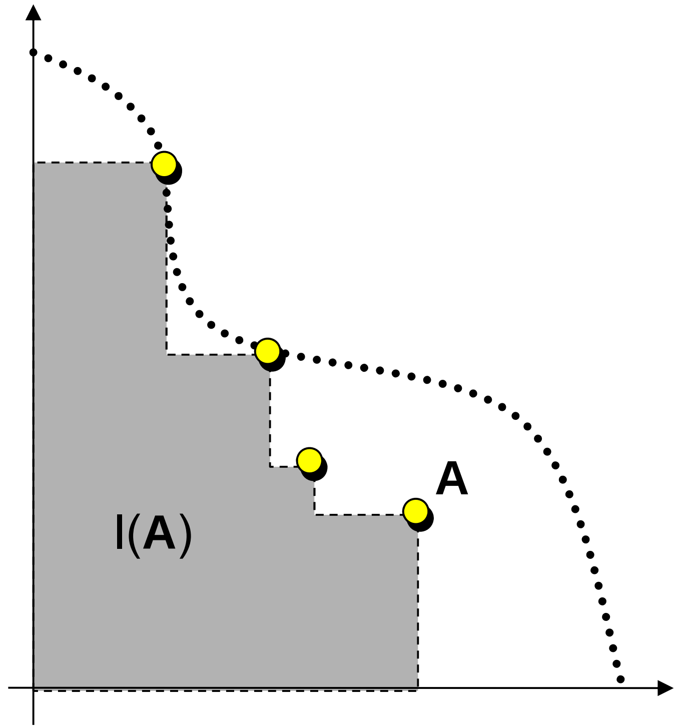
Example: Refinements Using Indicators

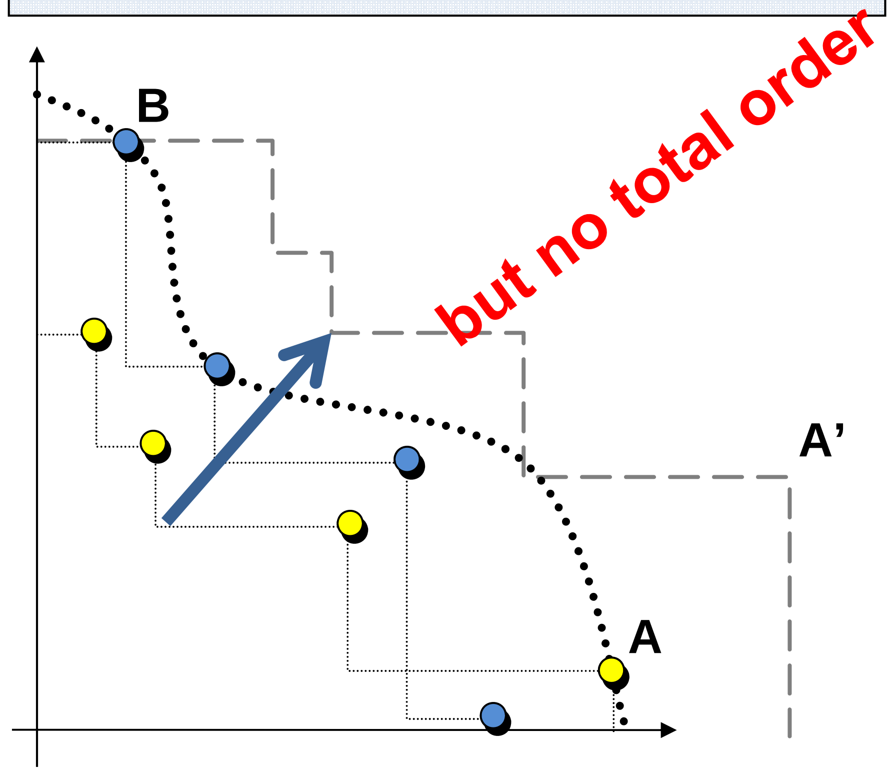
$$A \overset{\text{ref}}{\preccurlyeq} B :\Leftrightarrow I(A) \geq I(B)$$

$$A \overset{\text{ref}}{\preccurlyeq} B :\Leftrightarrow I(A,B) \leq I(B,A)$$

I(A) = volume of the weakly dominated area in objective space

I(A,B) = how much needs A to be moved to weakly dominate B





unary hypervolume indicator

binary epsilon indicator

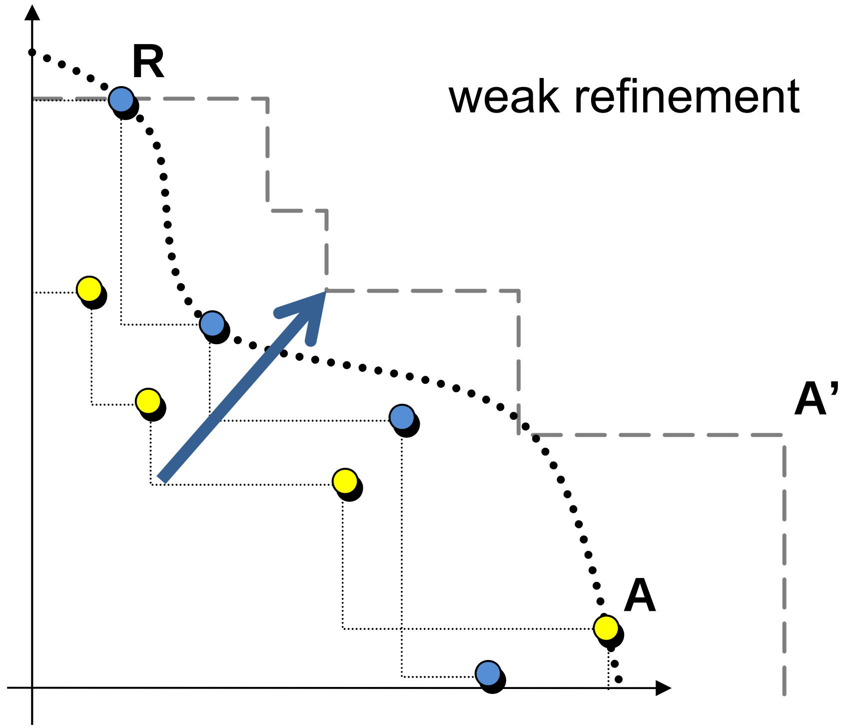
Example: Weak Refinement / No Refinement

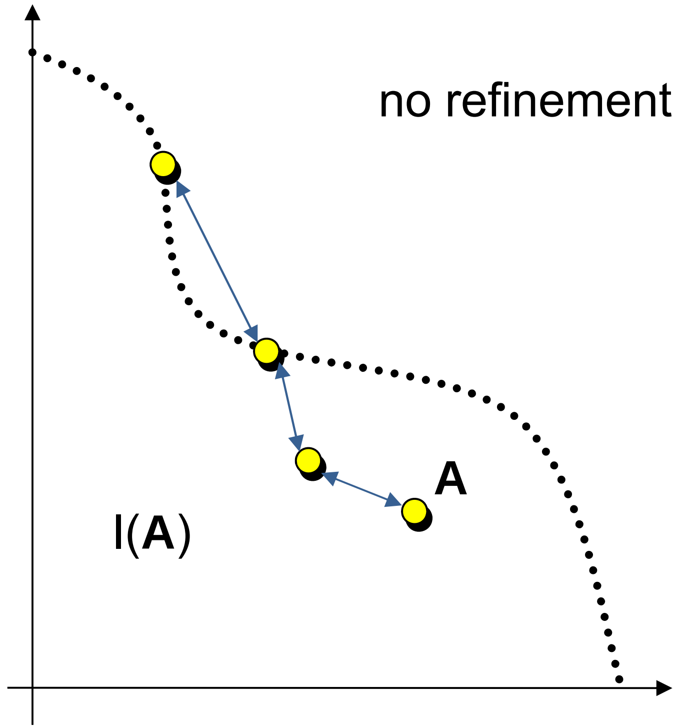
$$A \overset{\text{ref}}{\preccurlyeq} B :\Leftrightarrow I(A,R) \leq I(B,R)$$

$$A \overset{\text{ref}}{\preccurlyeq} B :\Leftrightarrow I(A) \leq I(B)$$

I(A,R) = how much needs A to be moved to weakly dominate R

I(A) = variance of pairwise distances





unary epsilon indicator

unary diversity indicator

Overview

The Big Picture

Basic Principles of Multiobjective Optimization

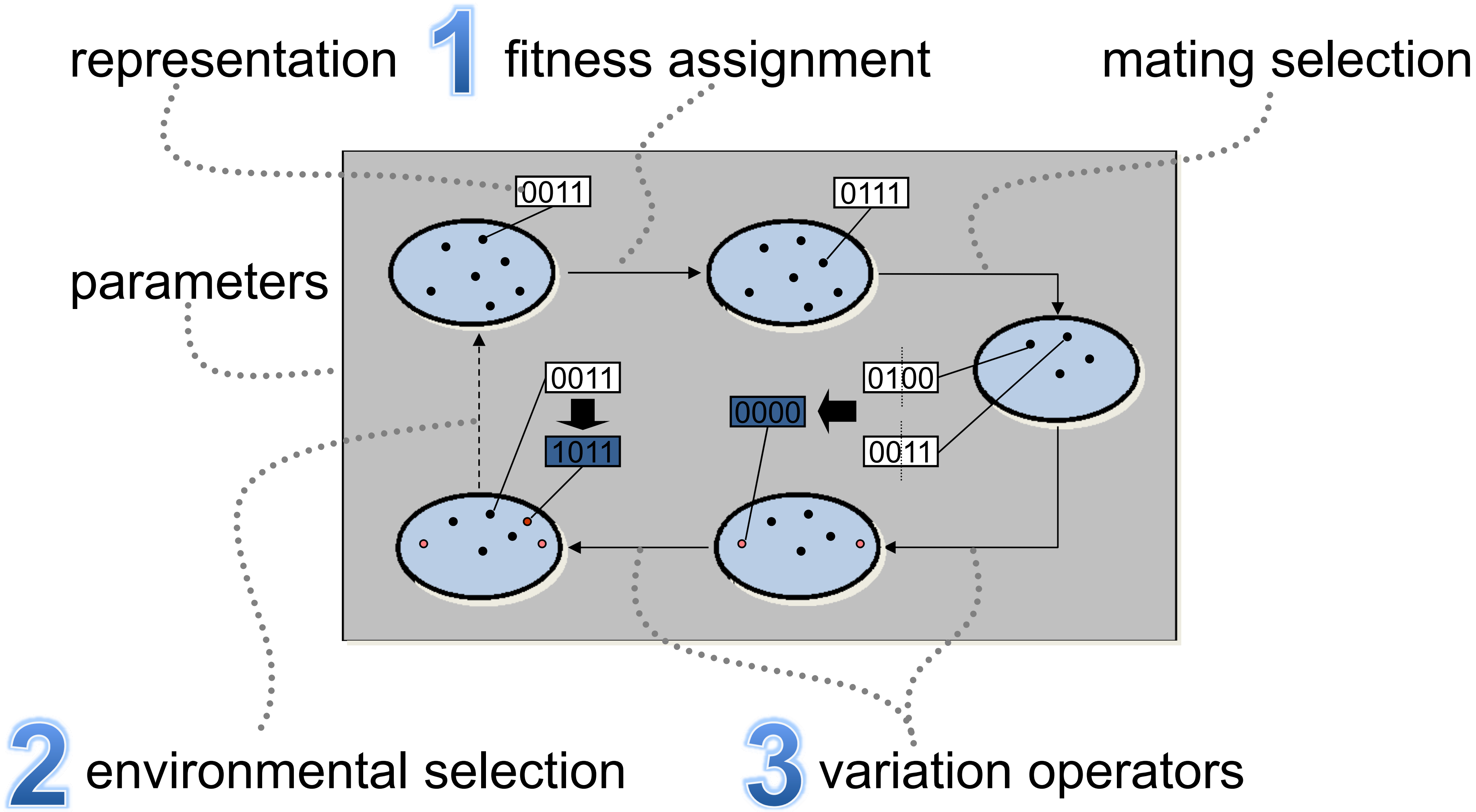
- algorithm design principles and concepts
- performance assessment

Selected Advanced Concepts

- indicator-based EMO
- preference articulation

A Few Examples From Practice

Algorithm Design: Particular Aspects

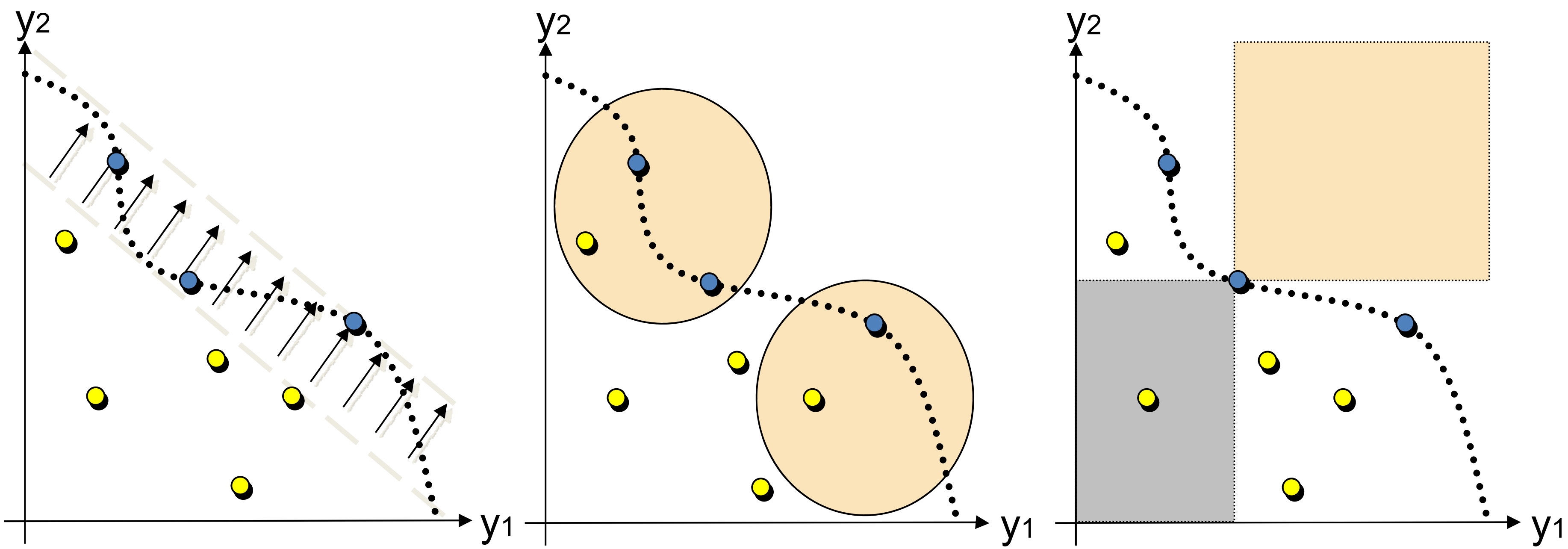


Fitness Assignment: Principal Approaches

aggregation-based
weighted sum

criterion-based
VEGA

dominance-based
SPEA2



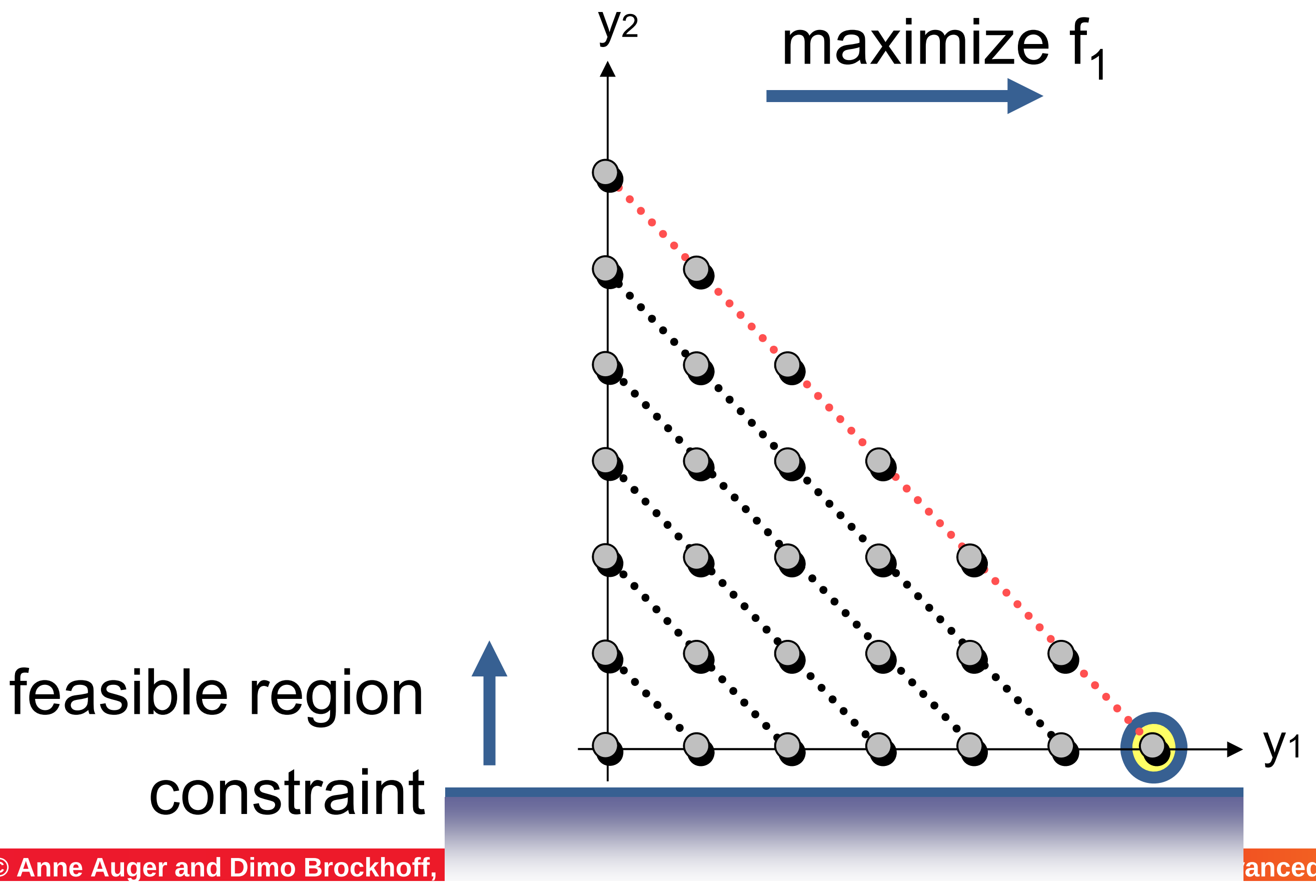
parameter-oriented
scaling-dependent

set-oriented
scaling-independent

Aggregation-Based: Multistart Constraint Method

Underlying concept:

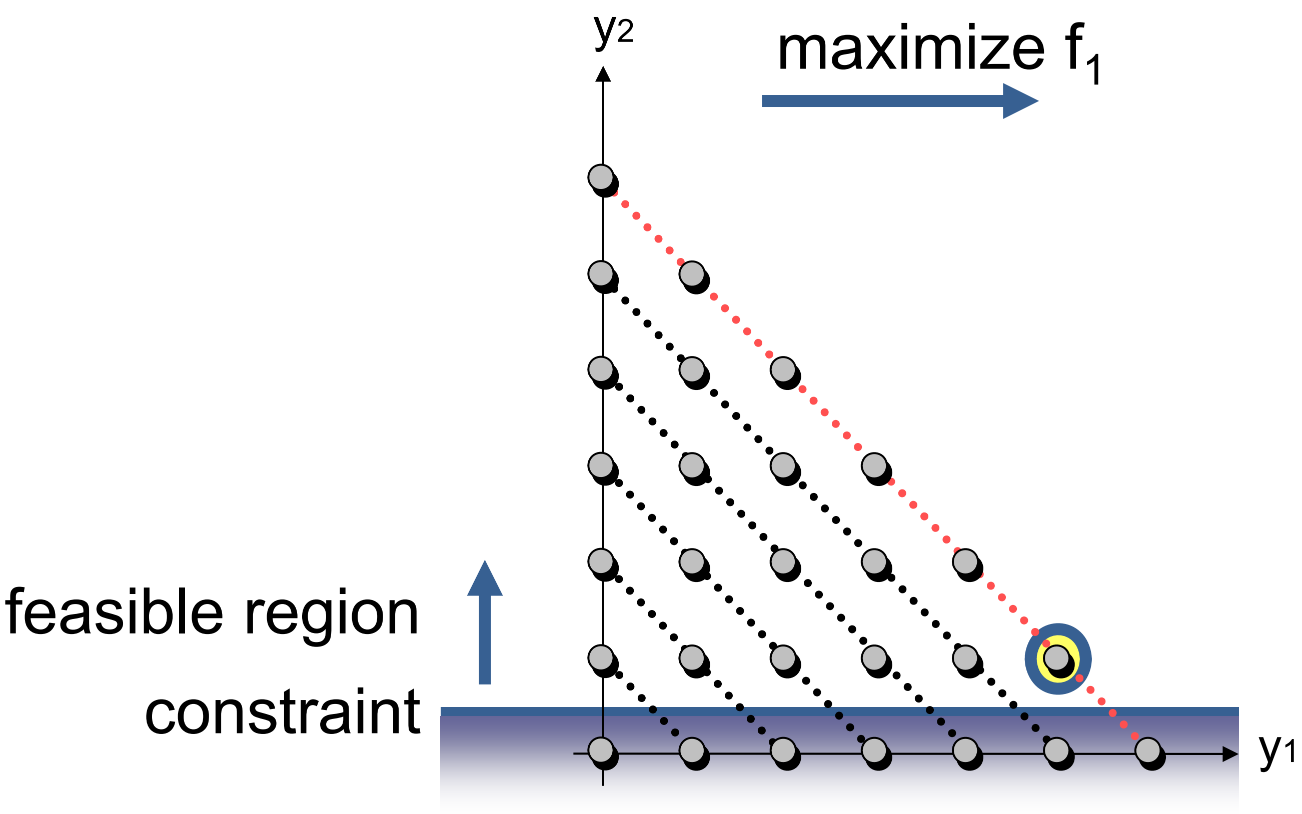
- Convert all objectives except of one into constraints
- Adaptively vary constraints



Aggregation-Based: Multistart Constraint Method

Underlying concept:

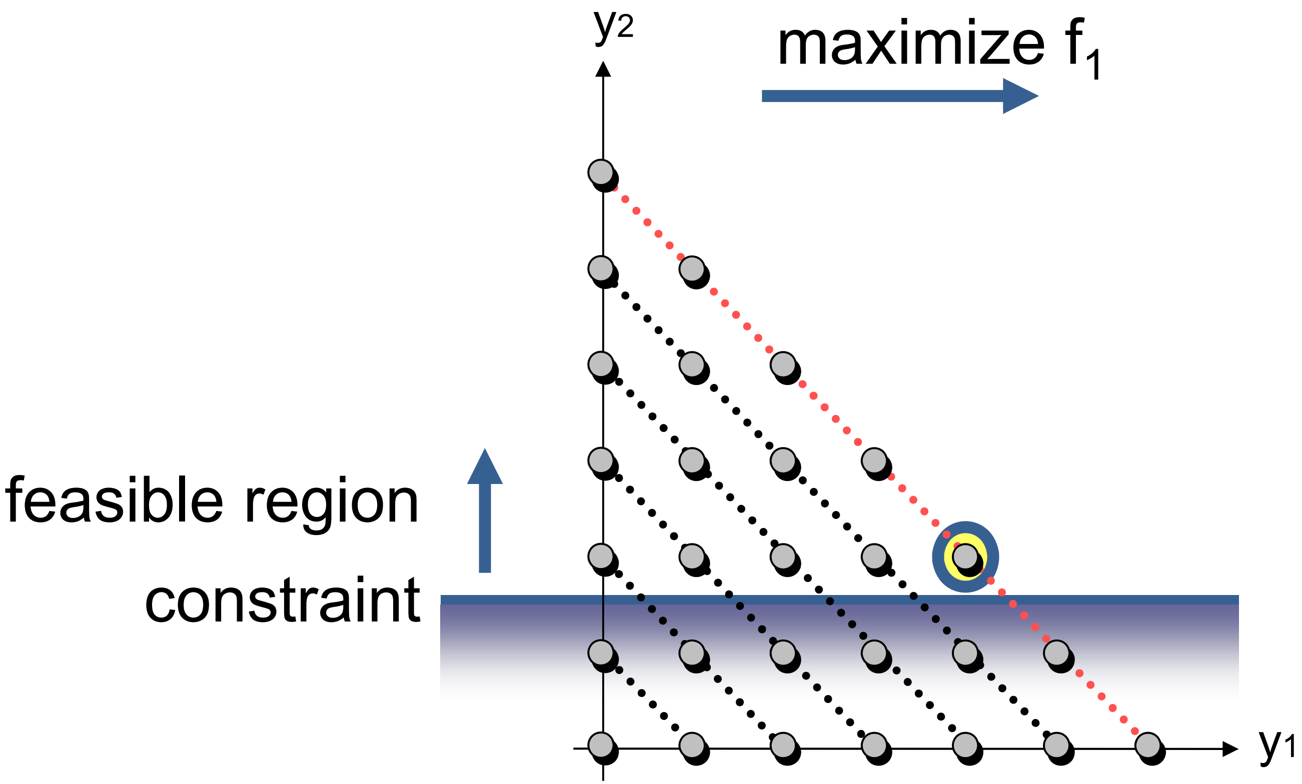
- Convert all objectives except of one into constraints
- Adaptively vary constraints



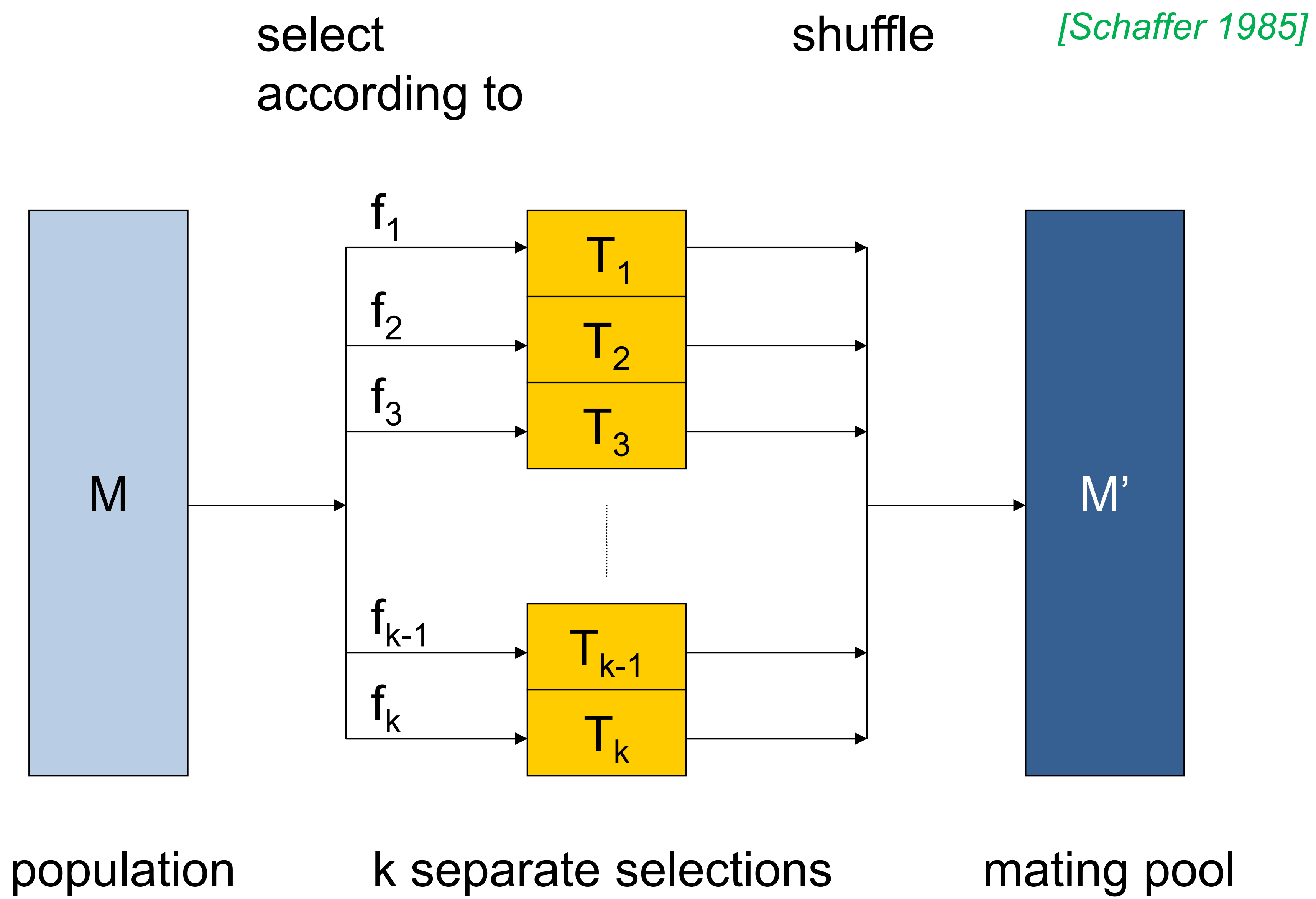
Aggregation-Based: Multistart Constraint Method

Underlying concept:

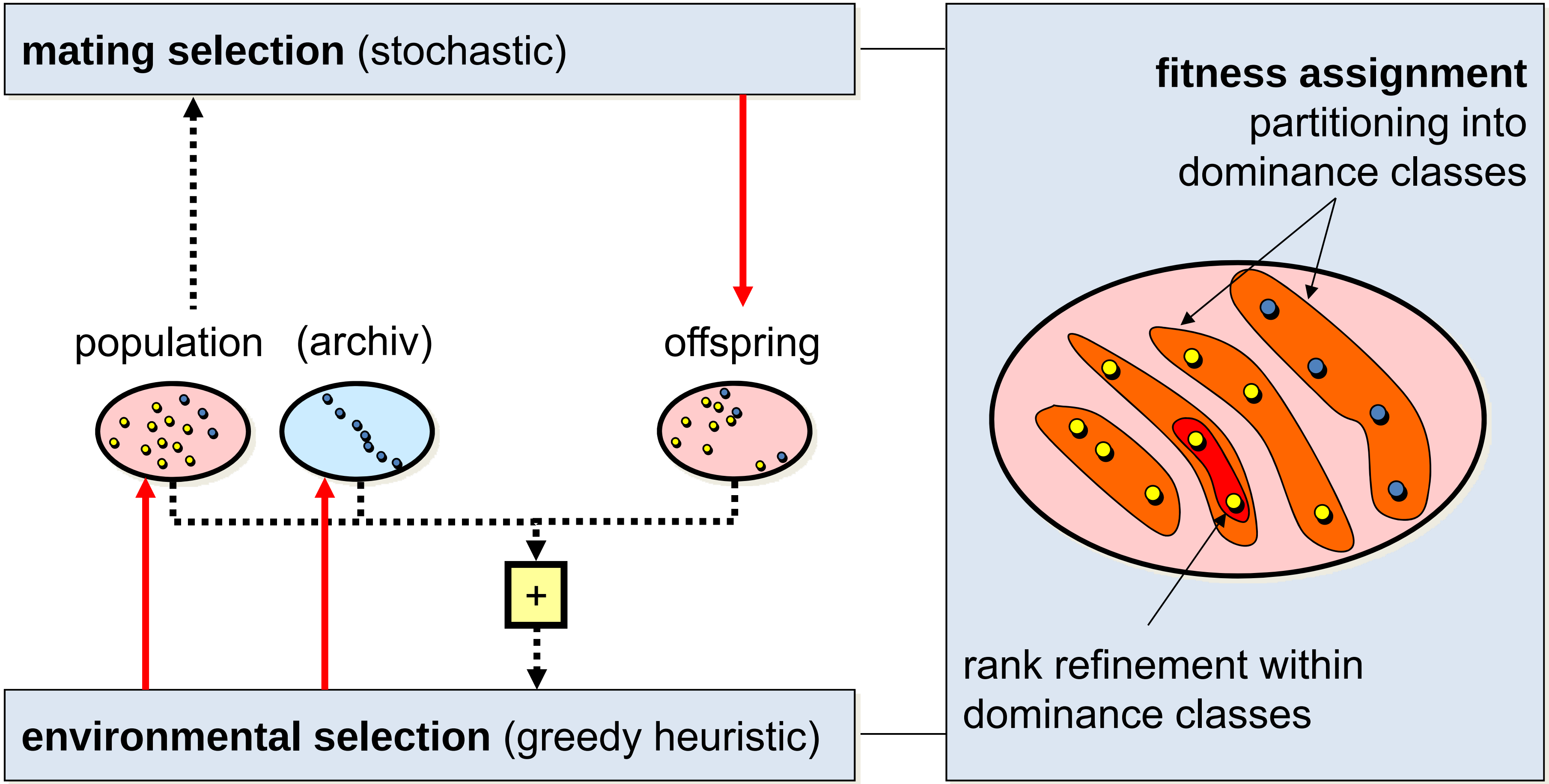
- Convert all objectives except of one into constraints
- Adaptively vary constraints



Criterion-Based Selection: VEGA



General Scheme of Dominance-Based EMO



Note: good in terms of set quality = good in terms of search?

Ranking of the Population Using Dominance

... goes back to a proposal by David Goldberg in 1989.

... is based on pairwise comparisons of the individuals only.

- **dominance rank:** by how many individuals is an individual dominated?
  MOGA, NPGA
- **dominance count:** how many individuals does an individual dominate?
  SPEA, SPEA2
- **dominance depth:** at which front is an individual located?
  NSGA, NSGA-II

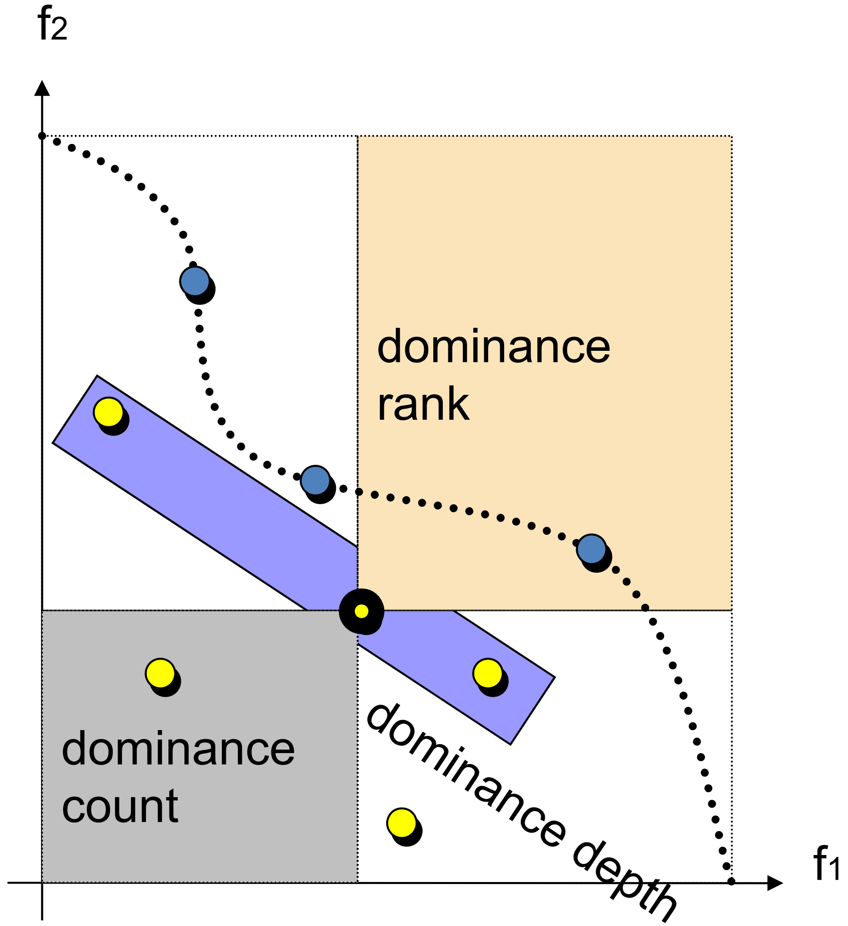
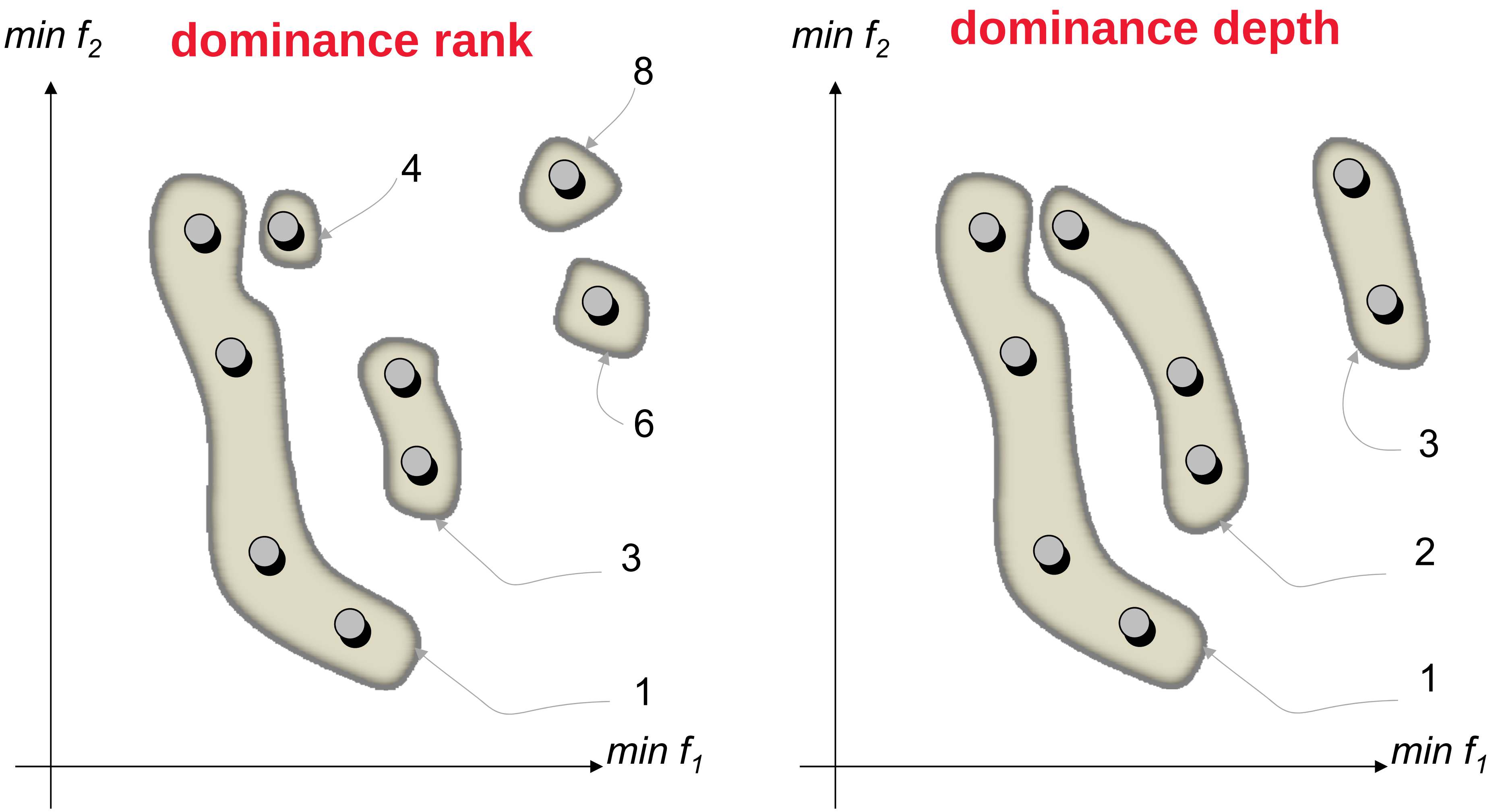


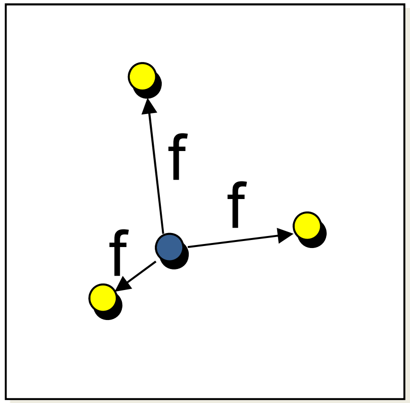
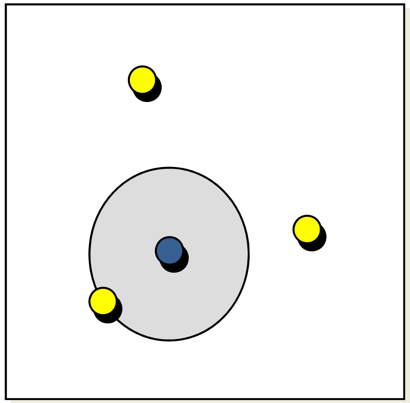
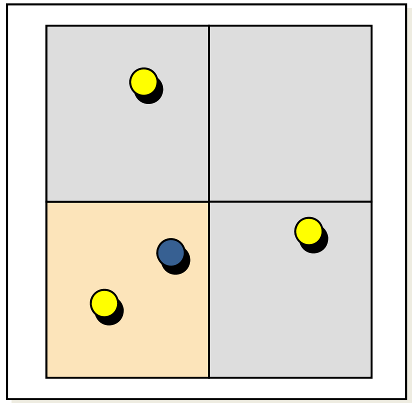
Illustration of Dominance-based Partitioning



Refinement of Dominance Rankings

Goal: rank incomparable solutions within a dominance class

❶ Density information (good for search, but usually no refinements)

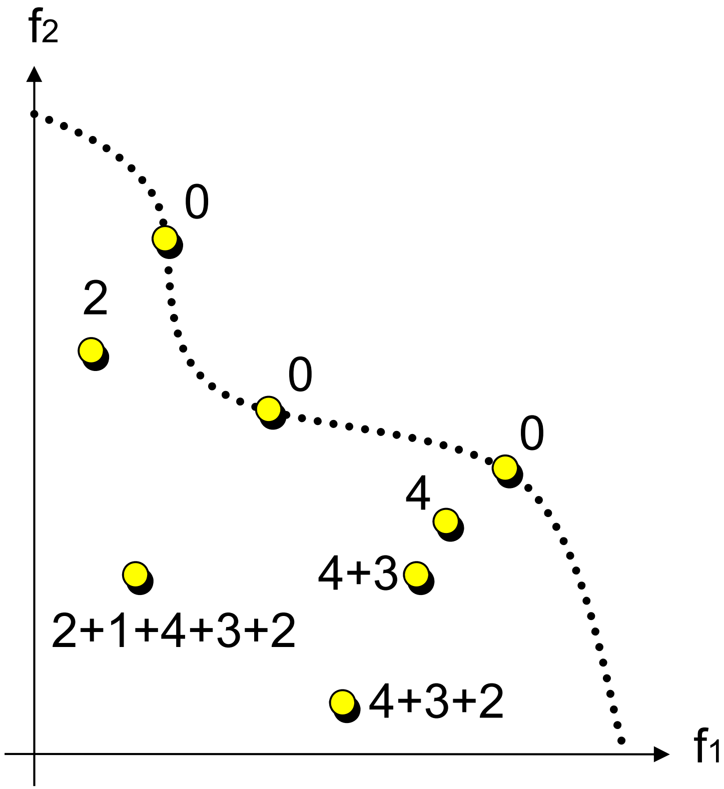
Kernel method	**k-th nearest neighbor**	**Histogram method**
density = function of the distances	density = function of distance to k-th neighbor	density = number of elements within box
		

❷ Quality indicator (good for set quality): soon...

Example: SPEA2 Dominance Ranking

Basic idea: the less dominated, the fitter...

Principle: first assign each solution a weight (strength), then add up weights of dominating solutions



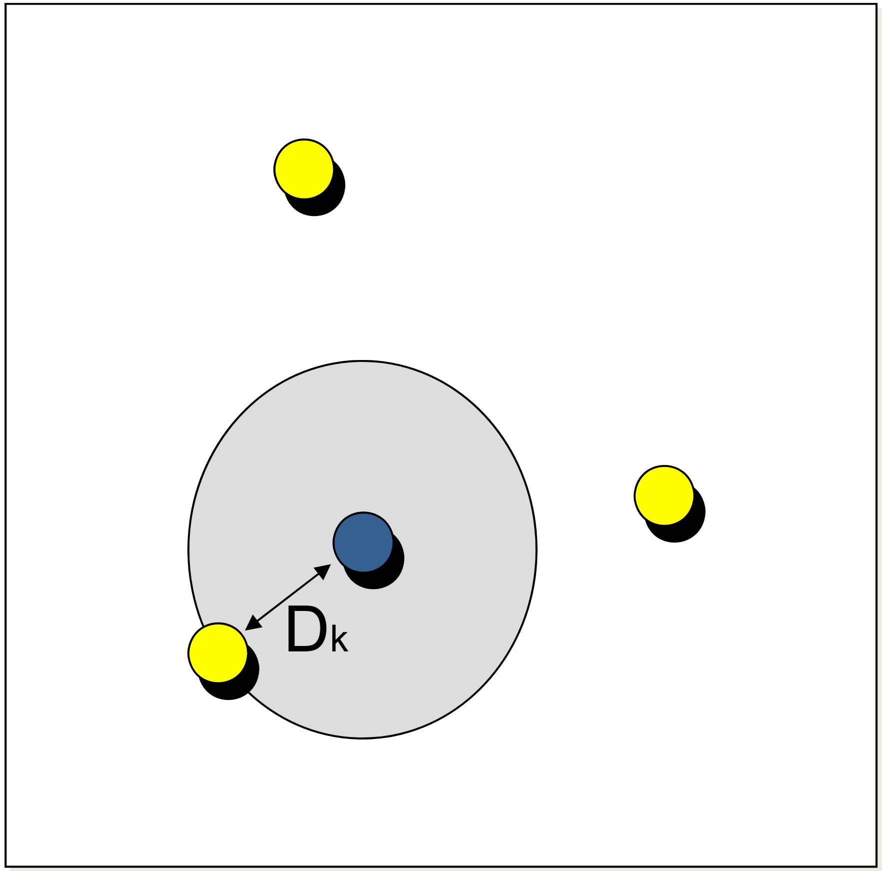
- S (strength) = #dominated solutions
- R (raw fitness) = $\sum$ strengths of dominators

Example: SPEA2 Diversity Preservation

Density Estimation

k-th nearest neighbor method:

- Fitness = R + 1 / (2 + D_k)
  (1 / (2 + D_k) < 1)
- D_k = distance to the k-th nearest individual
- Usually used: k = 2



Example: NSGA-II Diversity Preservation

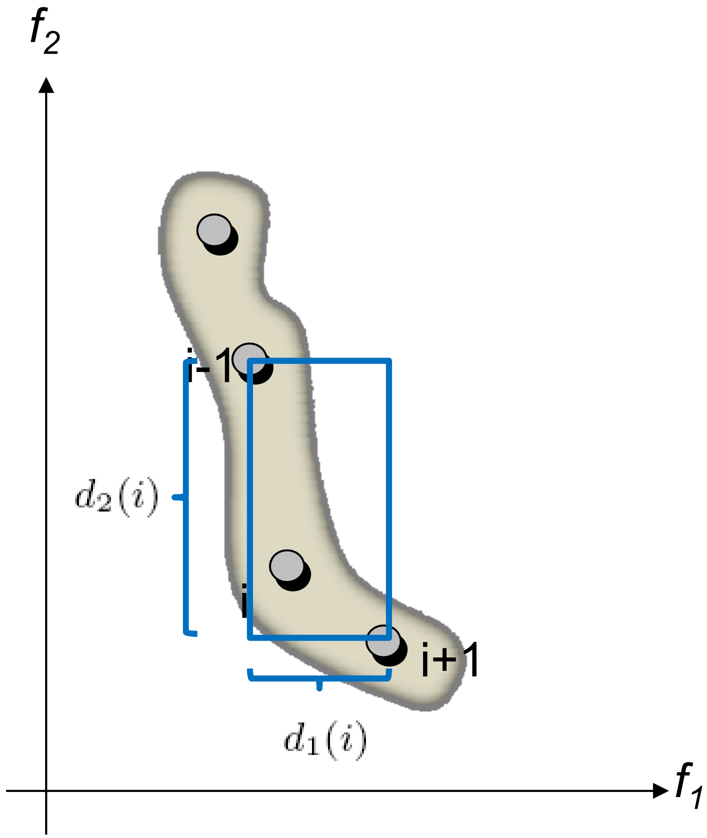
Density Estimation

crowding distance:

- sort solutions wrt. each objective

- crowding distance to neighbors:

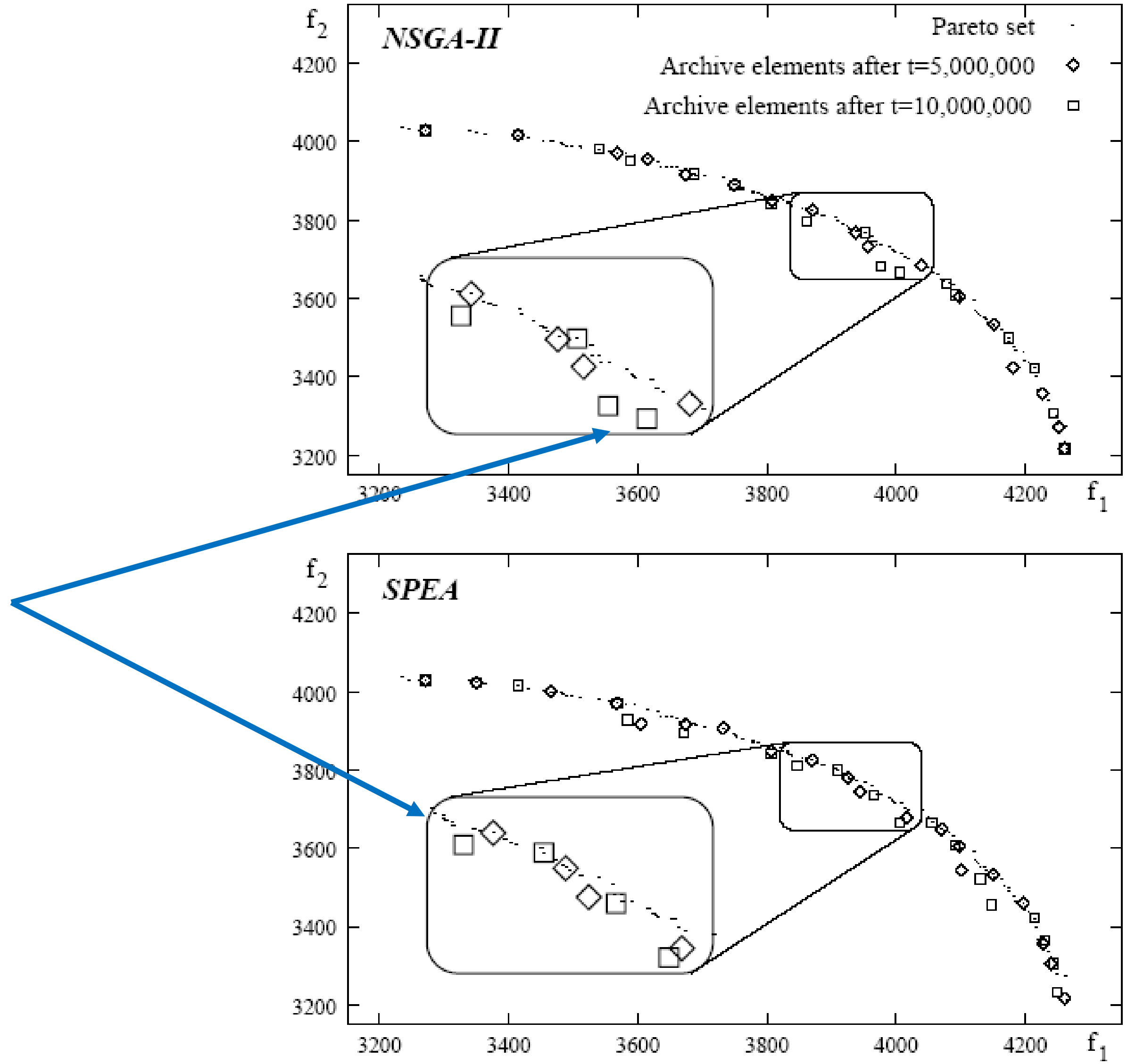
$$d(i) - \sum_{\text{obj. } m} |f_m(i-1) - f_m(i+1)|$$



SPEA2 and NSGA-II: Cycles in Optimization

Selection in SPEA2 and NSGA-II can result in
deteriorative cycles

non-dominated solutions already found can be lost



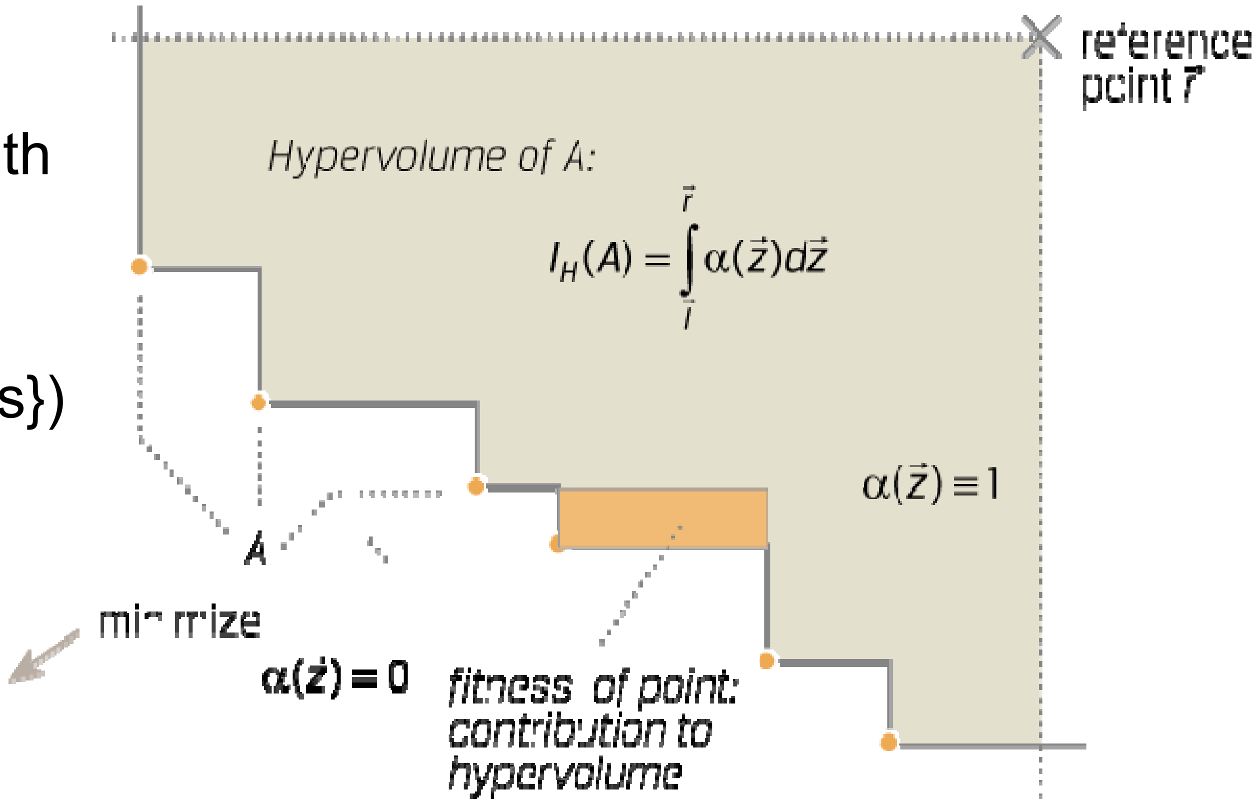
Hypervolume-Based Selection

Latest Approach (SMS-EMOA, MO-CMA-ES, HypE, …)

use (hypervolume) indicator to guide the search: refinement!

Main idea

Delete solutions with the smallest hypervolume loss $d(s) = I_H(P) - I_H(P / \{s\})$ iteratively



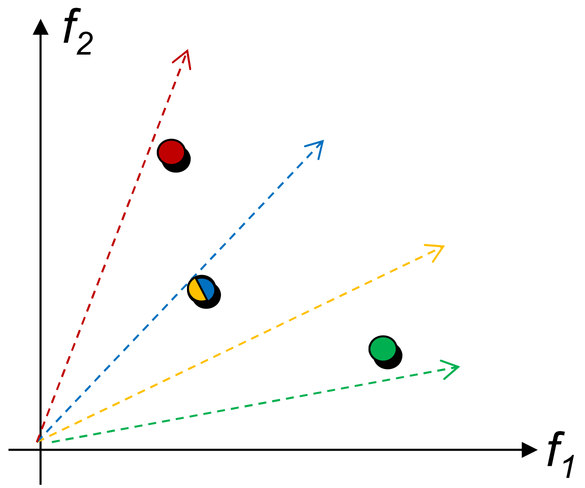
But: can also result in cycles [Judt et al. 2011] and is expensive [Bringmann and Friedrich 2009]

Decomposition-Based Selection: MOEA/D

MOEA/D: Multiobjective Evolutionary Algorithm Based on Decomposition [Zhang and Li 2007]

Ideas:

- Optimize N scalarizing functions in parallel
- Use only best solutions of "neighbored scalarizing function" for mating
- keep the best solutions for each scalarizing function
- use external archive for non-dominated solutions



Overview

The Big Picture

Basic Principles of Multiobjective Optimization

- algorithm design principles and concepts
- performance assessment

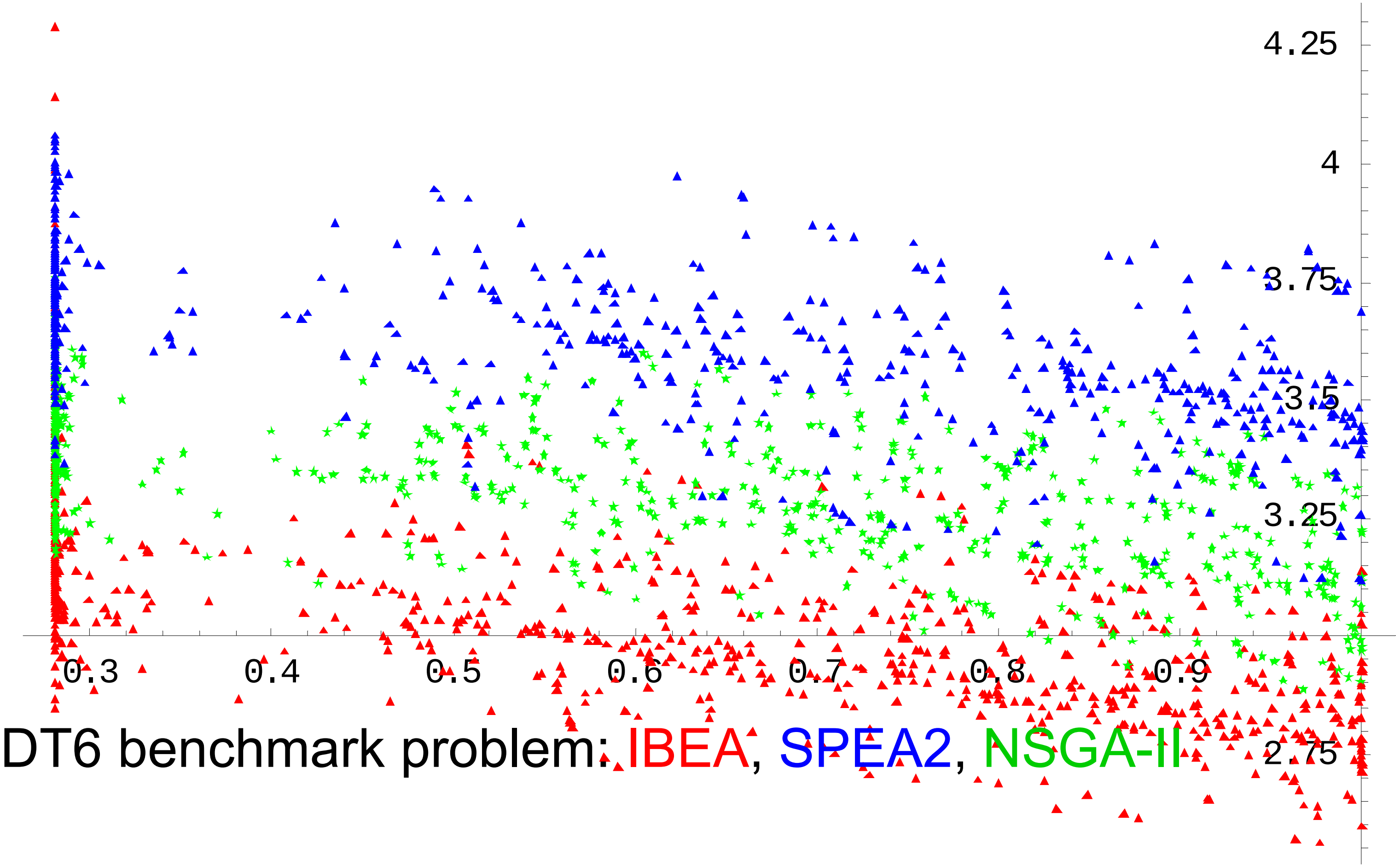
Selected Advanced Concepts

- indicator-based EMO
- preference articulation

A Few Examples From Practice

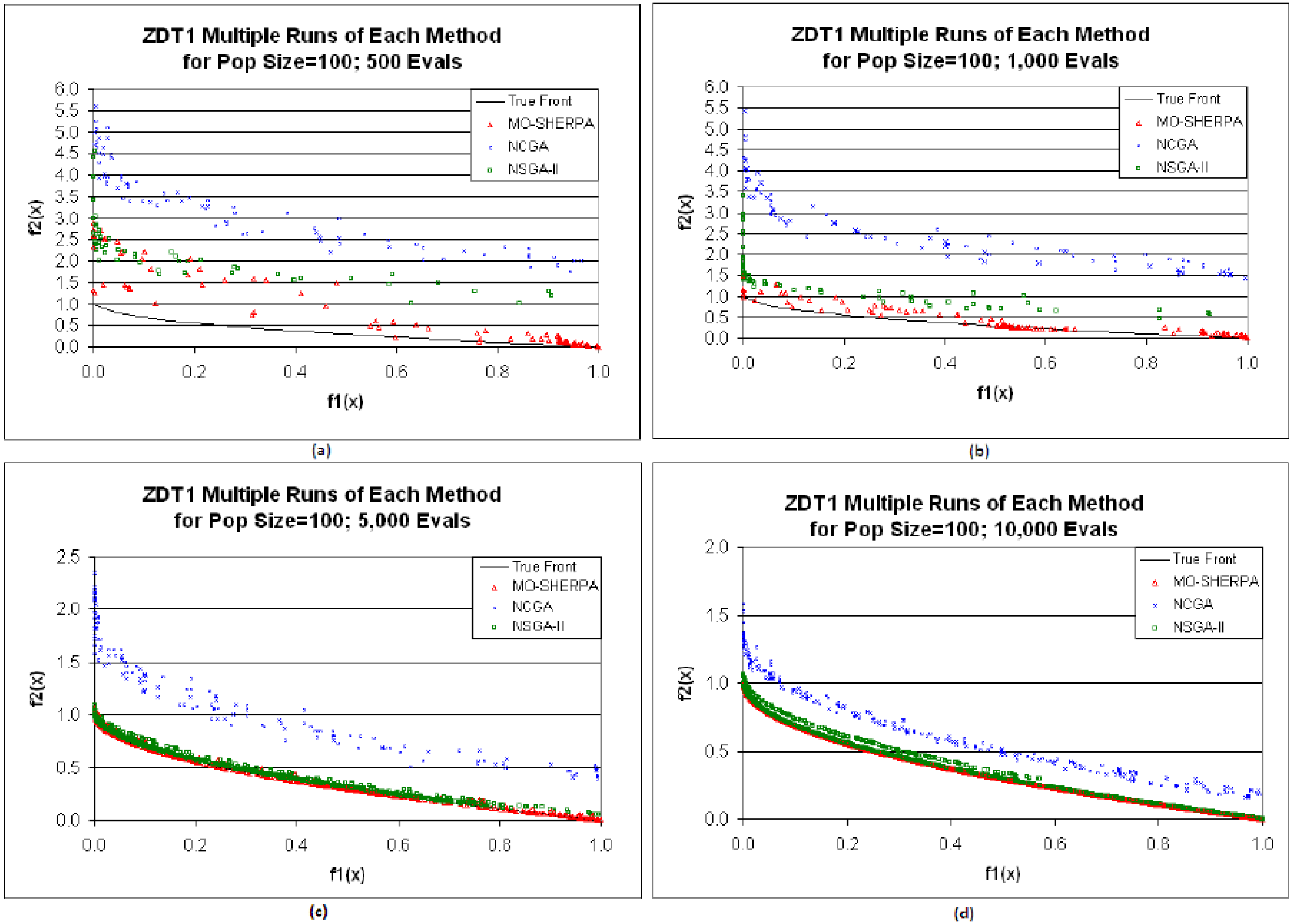
Once Upon a Time...

... multiobjective EAs were mainly compared visually:



ZDT6 benchmark problem: IBEA, SPEA2, NSGA-II

...And Even Today!



[found in a paper from 2009]

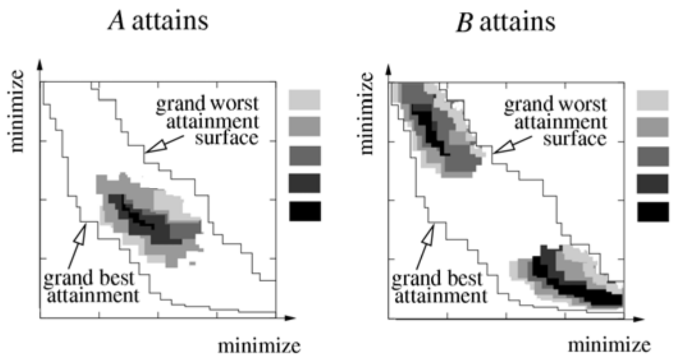
Two Approaches for Empirical Studies

Attainment function approach:

- Applies statistical tests directly to the samples of approximation sets
- Gives detailed information about how and where performance differences occur

Quality indicator approach:

- First, reduces each approximation set to a single value of quality
- Applies statistical tests to the samples of quality values

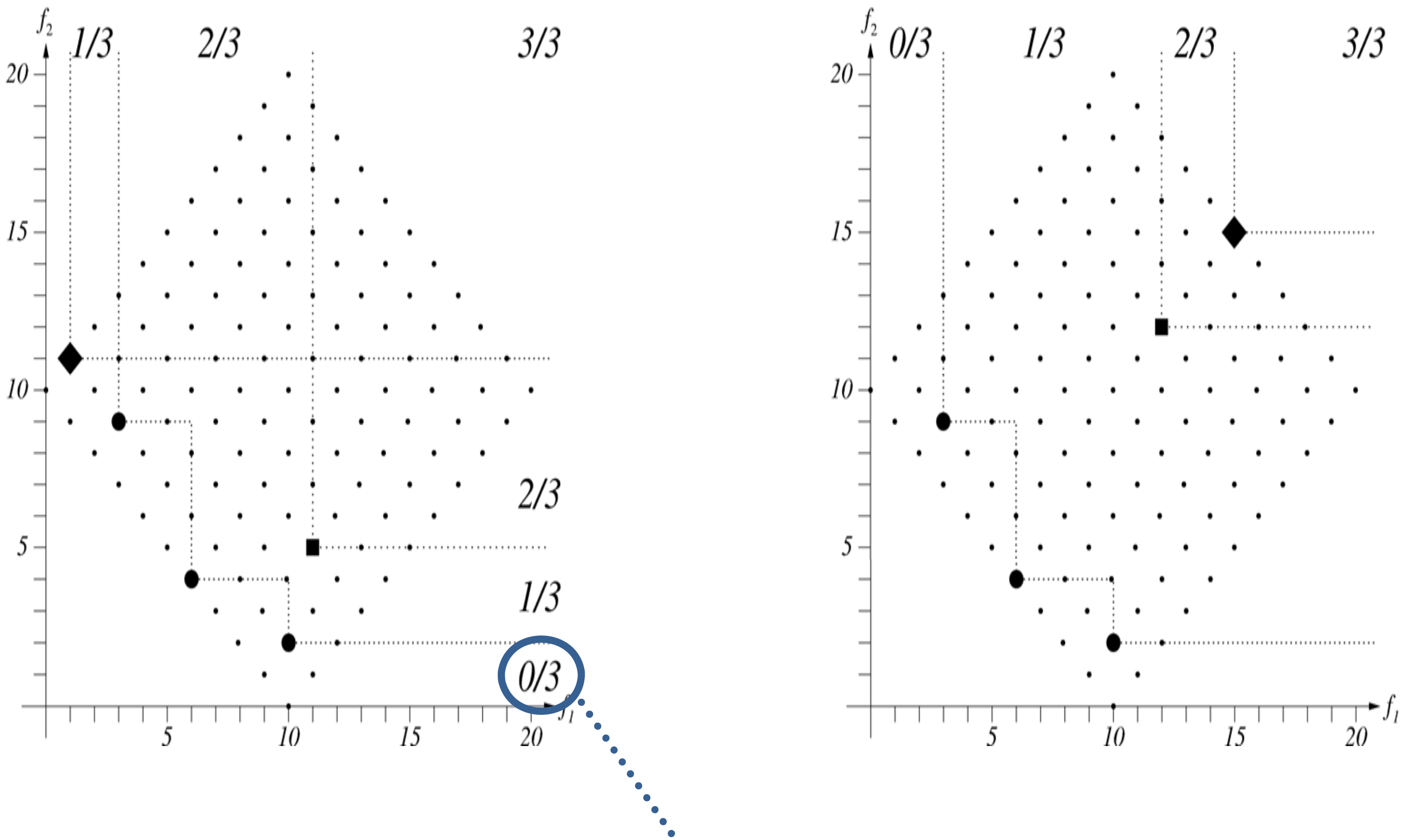


Indicator	A	B
Hypervolume indicator	6.3431	7.1924
ϵ-indicator	1.2090	0.12722
R_2 indicator	0.2434	0.1643
R_3 indicator	0.6454	0.3475

see e.g. [Zitzler et al. 2003]

Empirical Attainment Functions

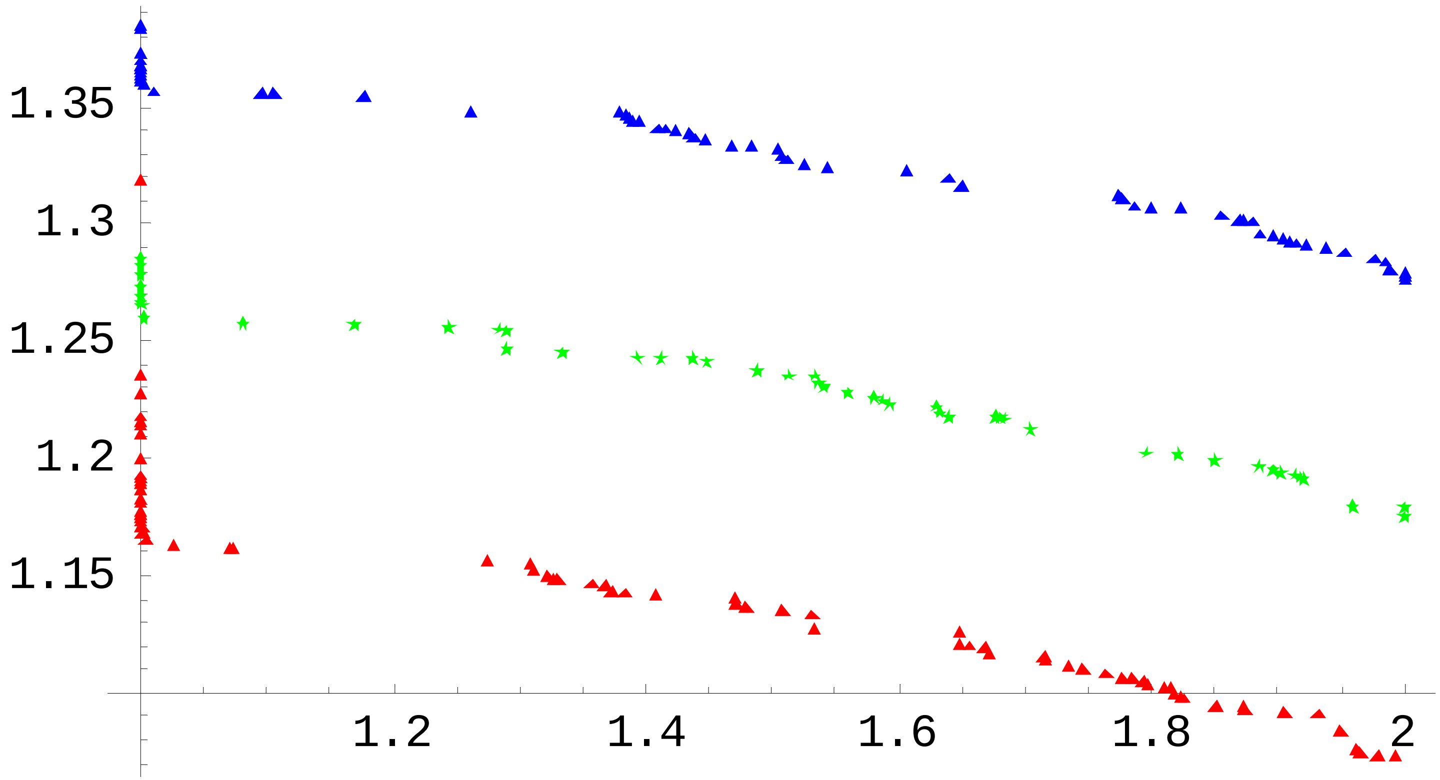
three runs of two multiobjective optimizers



frequency of attaining regions

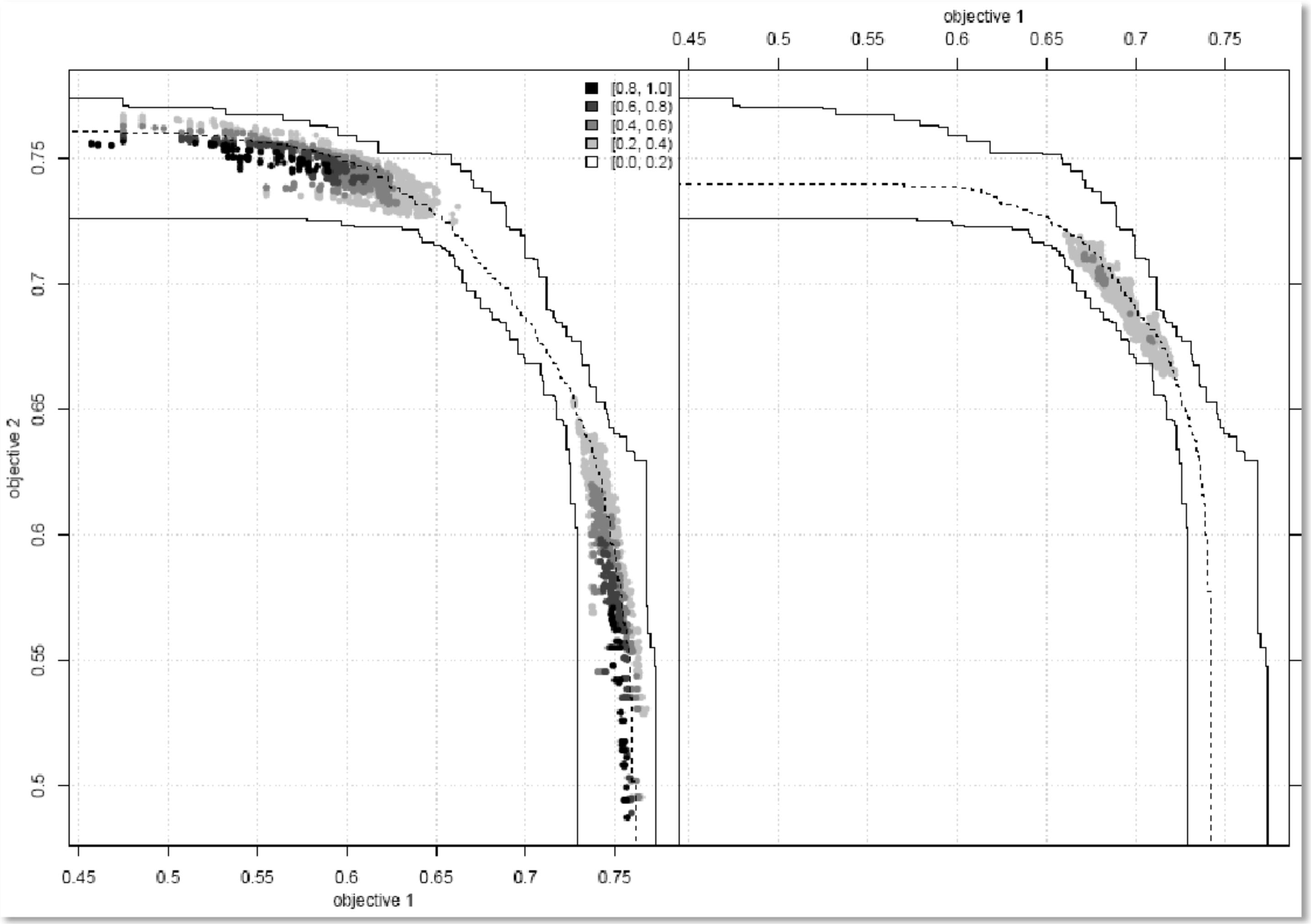
Attainment Plots

50% attainment surface for IBEA, SPEA2, NSGA2 (ZDT6)



latest implementation online at
http://eden.dei.uc.pt/~cmfonsec/software.html
see [Fonseca et al. 2011]

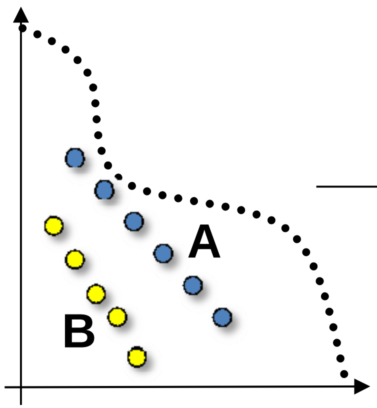
Attainment Plots



latest implementation online at
http://eden.dei.uc.pt/~cmfonsec/software.html
see [Fonseca et al. 2011]

Quality Indicator Approach

Goal: compare two Pareto set approximations A and B



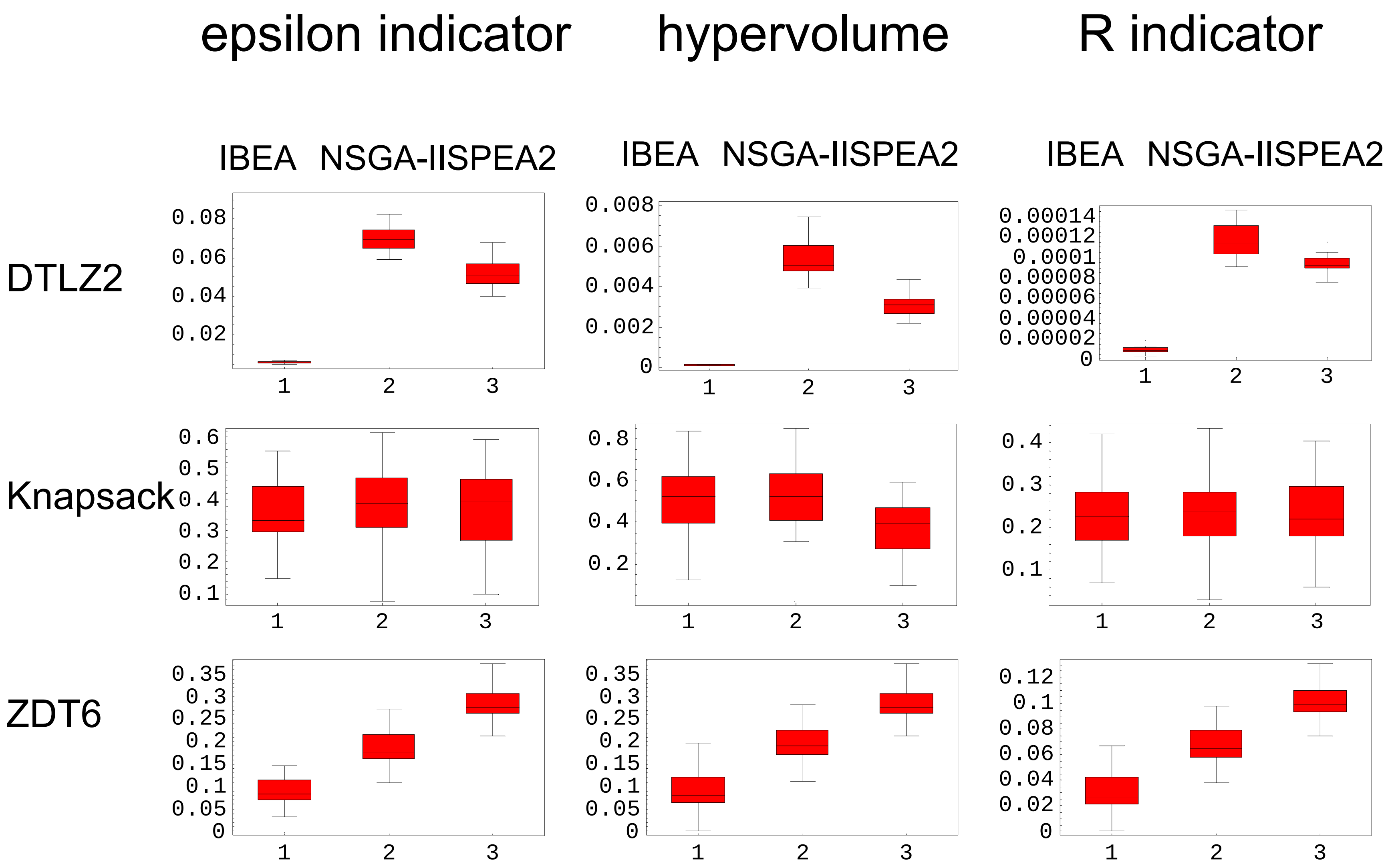
	A	B
hypervolume	*432.34*	*420.13*
distance	*0.3308*	*0.4532*
diversity	*0.3637*	*0.3463*
spread	*0.3622*	*0.3601*
cardinality	*6*	*5*

→ "**A** better"

Comparison method C = quality measure(s) + Boolean function

A, B —(quality measure / **reduction**)→ $\mathbb{R}^n$ —(Boolean function / **interpretation**)→ statement

Example: Box Plots



Statistical Assessment (Kruskal Test)

ZDT6
Epsilon

is better than →

	IBEA	NSGA2	SPEA2
IBEA		~0 ☺	~0 ☺
NSGA2	1		~0 ☺
SPEA2	1	1	

Overall p-value = 6.22079e-17.
Null hypothesis rejected (alpha 0.05)

DTLZ2
R

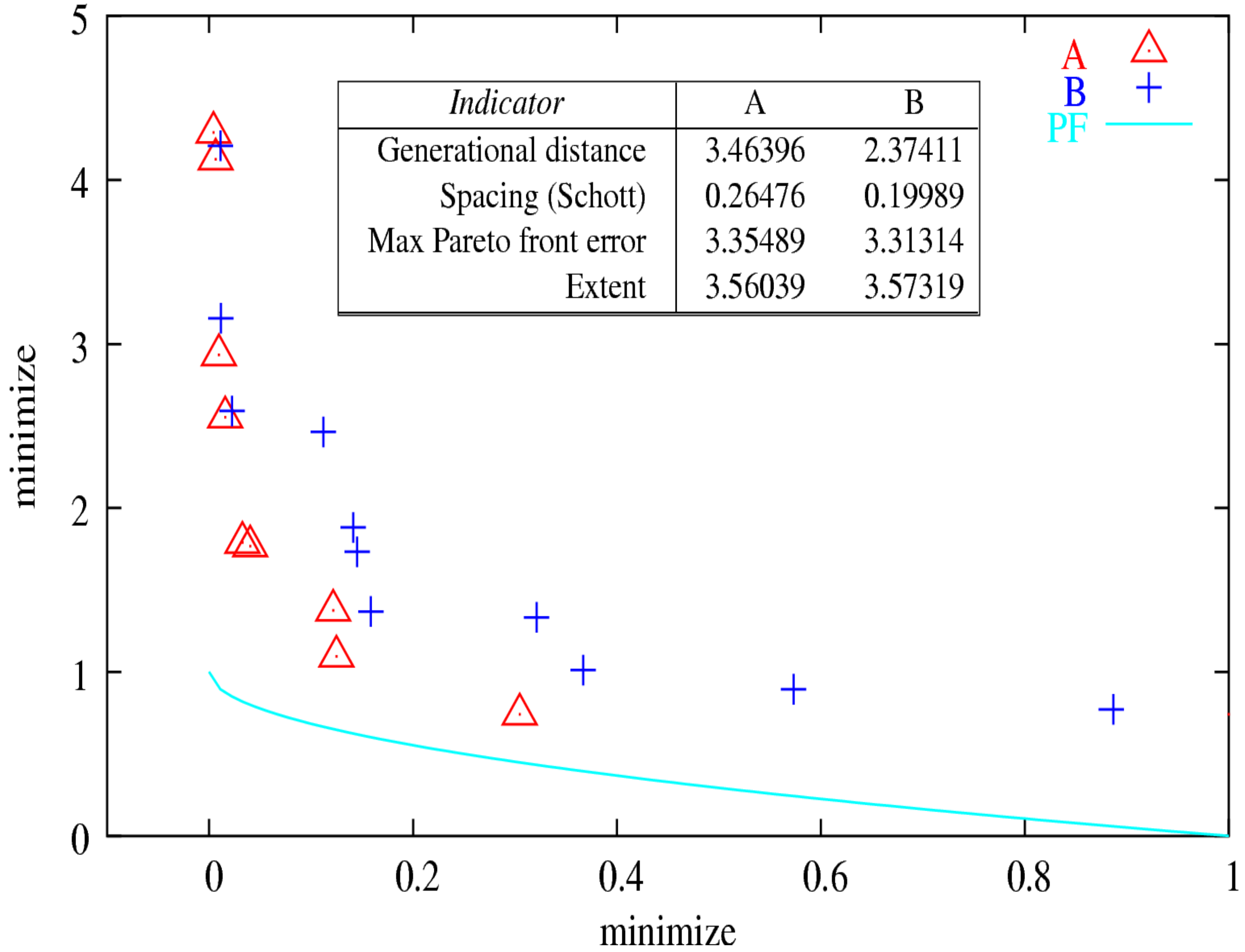
is better than →

	IBEA	NSGA2	SPEA2
IBEA		~0 ☺	~0 ☺
NSGA2	1		1
SPEA2	1	~0 ☺	

Overall p-value = 7.86834e-17.
Null hypothesis rejected (alpha 0.05)

Knapsack/Hypervolume: H_0 = No significance of any differences

Problems With Non-Compliant Indicators



Indicator	A	B
Generational distance	3.46396	2.37411
Spacing (Schott)	0.26476	0.19989
Max Pareto front error	3.35489	3.31314
Extent	3.56039	3.57319

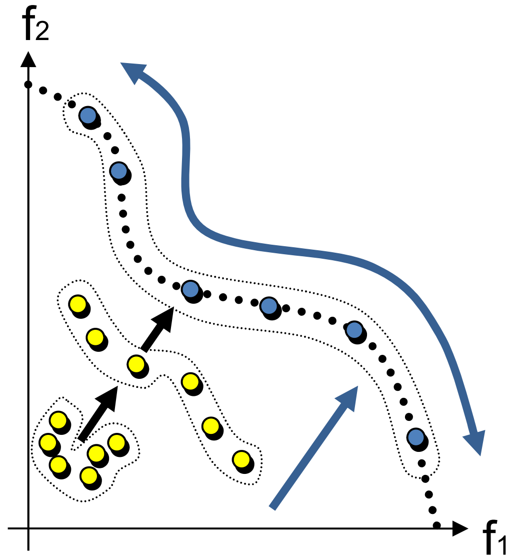
What Are Good Set Quality Measures?

There are **three aspects** [Zitzler et al. 2000]

Comparing different optimization techniques experimentally always involves the notion of performance. In the case of multiobjective optimization, the definition of quality is substantially more complex than for single-objective optimization problems, because the optimization goal itself consists of multiple objectives:

- The distance of the resulting nondominated set to the Pareto-optimal front should be minimized.
- A good (in most cases uniform) distribution of the solutions found is desirable. The assessment of this criterion might be based on a certain distance metric.
- The extent of the obtained nondominated front should be maximized, i.e., for each objective, a wide range of values should be covered by the nondominated solutions.

In the literature, some attempts can be found to formalize the above definition (or parts



Wrong! [Zitzler et al. 2003]

An infinite number of unary set measures is needed to detect in general whether A is better than B

Overview

The Big Picture

Basic Principles of Multiobjective Optimization

- algorithm design principles and concepts
- performance assessment

Selected Advanced Concepts

- indicator-based EMO
- preference articulation

A Few Examples From Practice

Indicator-Based EMO: Optimization Goal

When the goal is to maximize a unary indicator…

- we have a single-objective set problem to solve
- but what is the optimum?
- important: population size μ plays a role!

Multiobjective Problem —Indicator→ Single-objective Problem

Optimal μ-Distribution:

A set of μ solutions that maximizes a certain unary indicator I among all sets of μ solutions is called optimal μ-distribution for I. [Auger et al. 2009a]

Optimal μ-Distributions for the Hypervolume

Hypervolume indicator refines dominance relation

$\Longrightarrow$ most results on optimal μ-distributions for hypervolume

Optimal μ-Distributions (example results)

[Auger et al. 2009a]:

- contain equally spaced points iff front is linear
- density of points $\propto \sqrt{-f'(x)}$ with f' the slope of the front

[Friedrich et al. 2011]:

optimal μ-distributions for the hypervolume correspond to ε-approximations of the front

OPT	$1 + \frac{\log(\min\{A/a, B/b\})}{n}$
HYP	$1 + \frac{\sqrt{A/a} + \sqrt{B/b}}{n-4}$
logHYP	$1 + \frac{\sqrt{\log(A/a)\,\log(B/b)}}{n-2}$

! (probably) does not hold for > 2 objectives

Overview

The Big Picture

Basic Principles of Multiobjective Optimization

- algorithm design principles and concepts
- performance assessment

Selected Advanced Concepts

- indicator-based EMO
- preference articulation

A Few Examples From Practice

Articulating User Preferences During Search

What we thought: EMO is preference-less

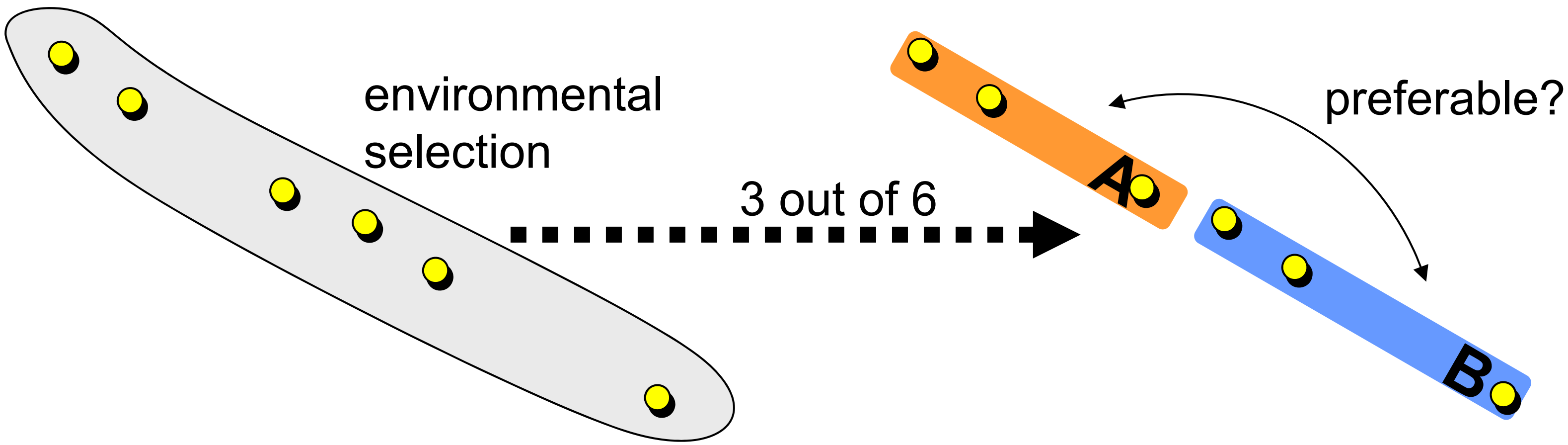
given by the DM.

Search before decision making: Optimization is performed without any preference information given. The result of the search process is a set of (ideally Pareto-optimal) candidate solutions from which the final choice is made by the DM.

Decision making during search: The DM can articulate preferences during

[Zitzler 1999]

What we learnt: EMO just uses weaker preference information

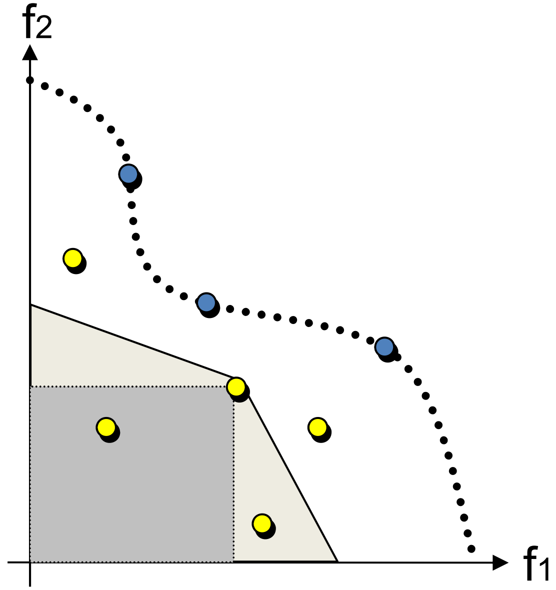


Incorporation of Preferences *During* Search

Nevertheless...

- the more (known) preferences incorporated the better
- in particular if search space is too large

[Branke 2008], [Rachmawati and Srinivasan 2006], [Coello Coello 2000]



❶ Refine/modify dominance relation, e.g.:

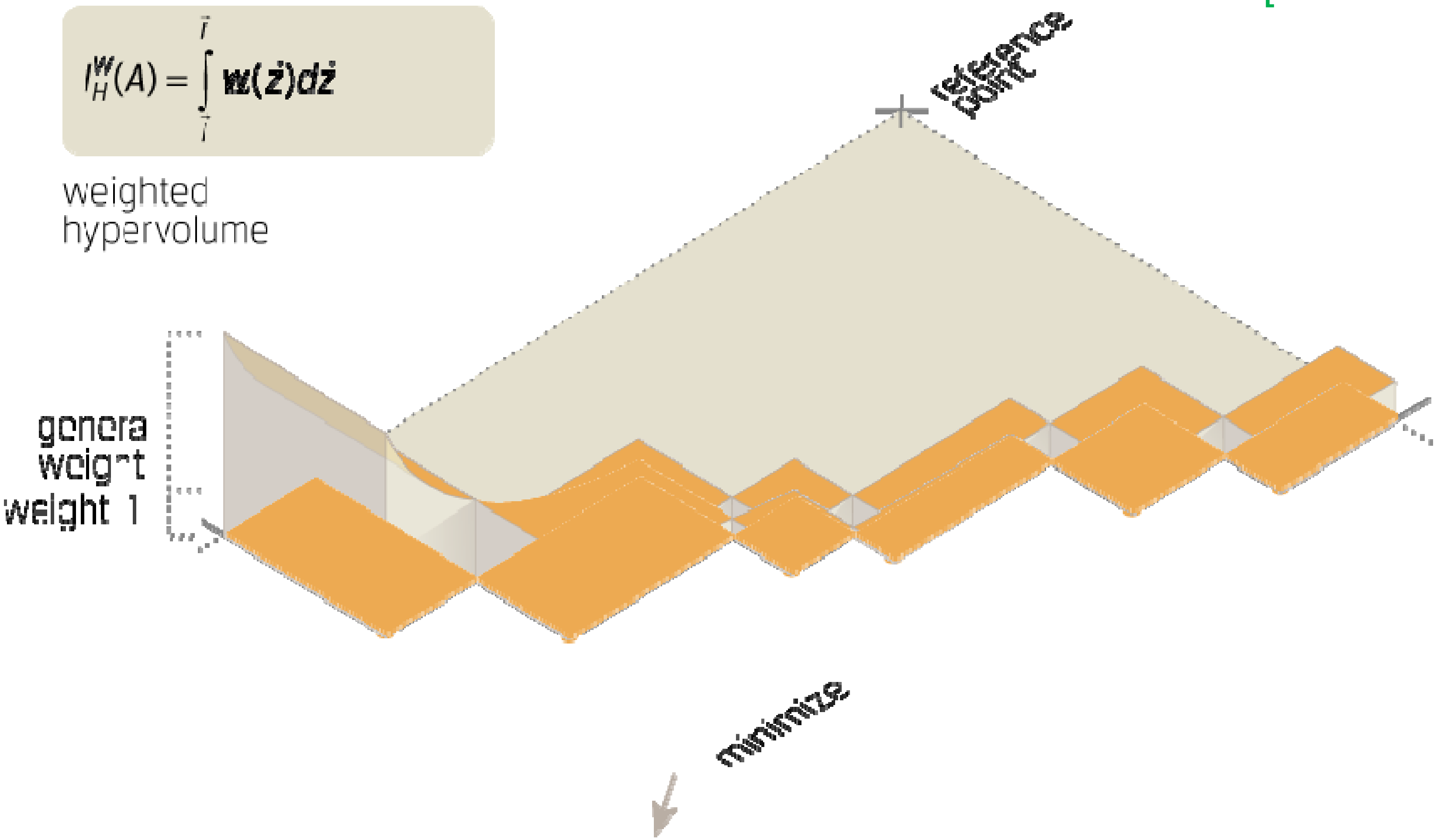
- using goals, priorities, constraints [Fonseca and Fleming 1998a,b]
- using different types of cones [Branke and Deb 2004]

❷ Use quality indicators, e.g.:

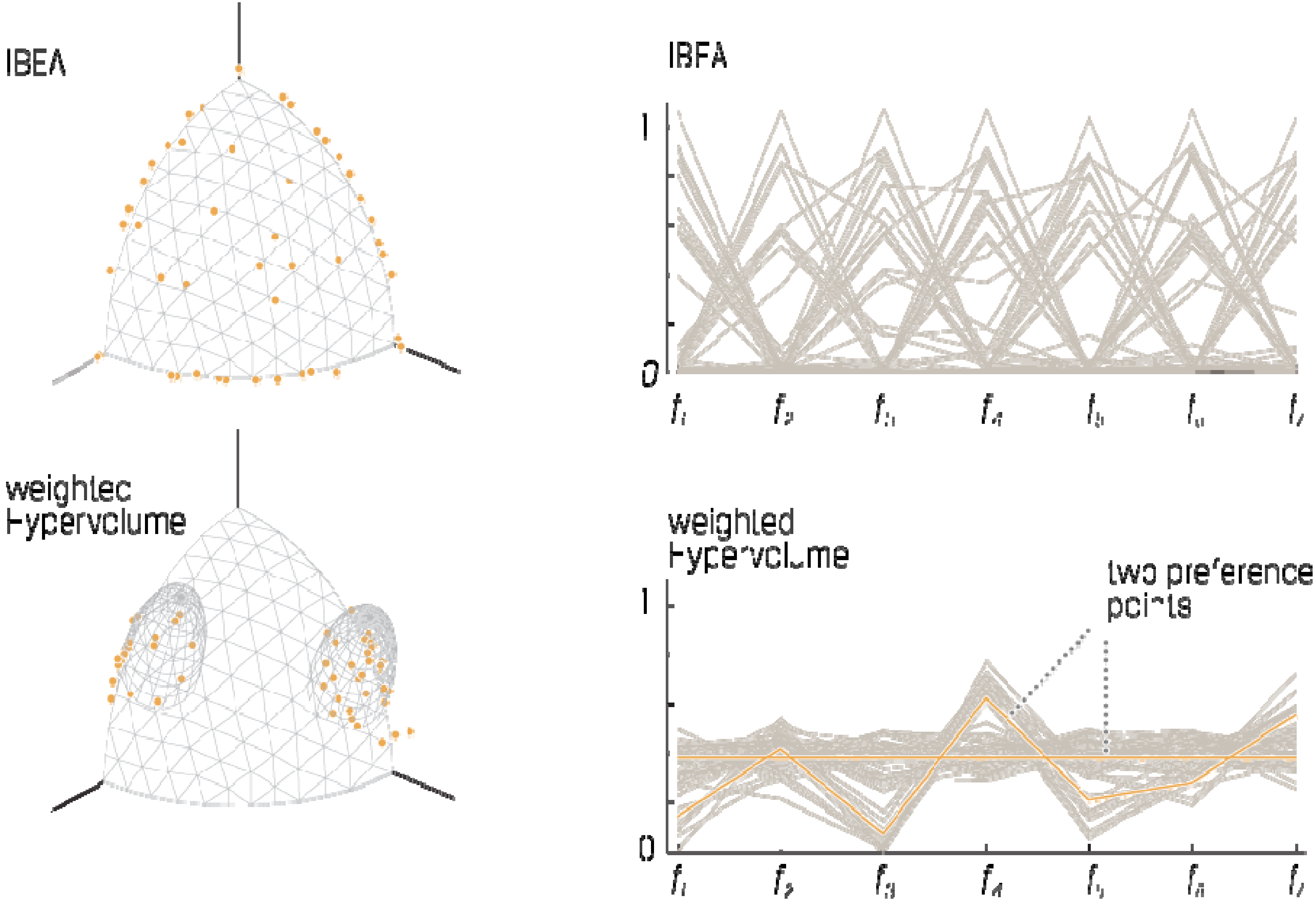
- based on reference points and directions [Deb and Sundar 2006, Deb and Kumar 2007]
- based on binary quality indicators [Zitzler and Künzli 2004]
- based on the hypervolume indicator (now) [Zitzler et al. 2007]

Example: Weighted Hypervolume Indicator

[Zitzler et al. 2007]



Weighted Hypervolume in Practice



[Auger et al. 2009b]

Overview

The Big Picture

Basic Principles of Multiobjective Optimization

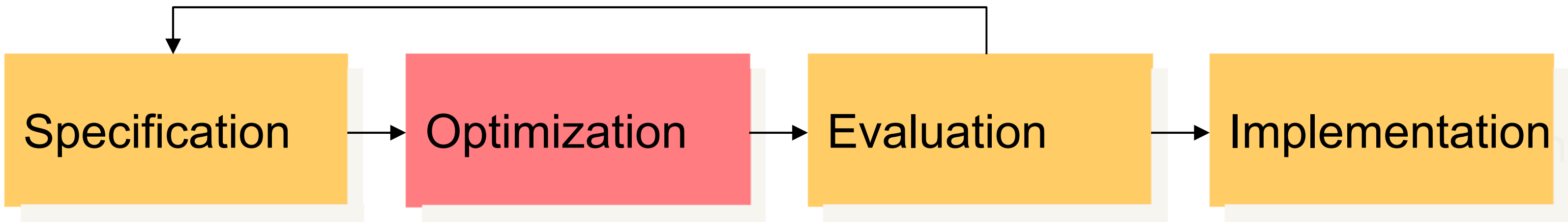
- algorithm design principles and concepts
- performance assessment

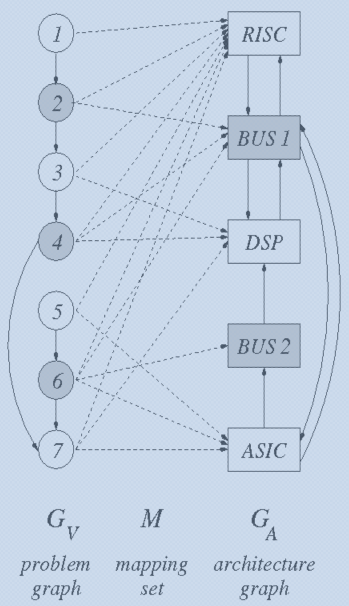
Selected Advanced Concepts

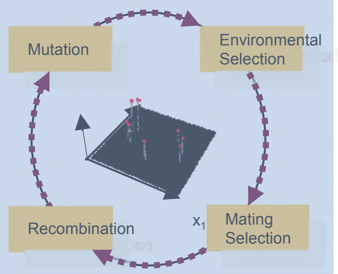
- indicator-based EMO
- preference articulation

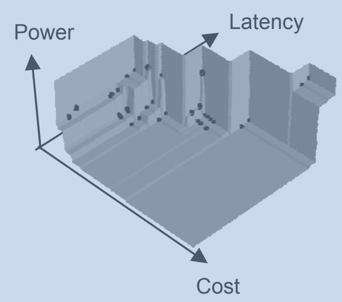
A Few Examples From Practice

Application: Design Space Exploration





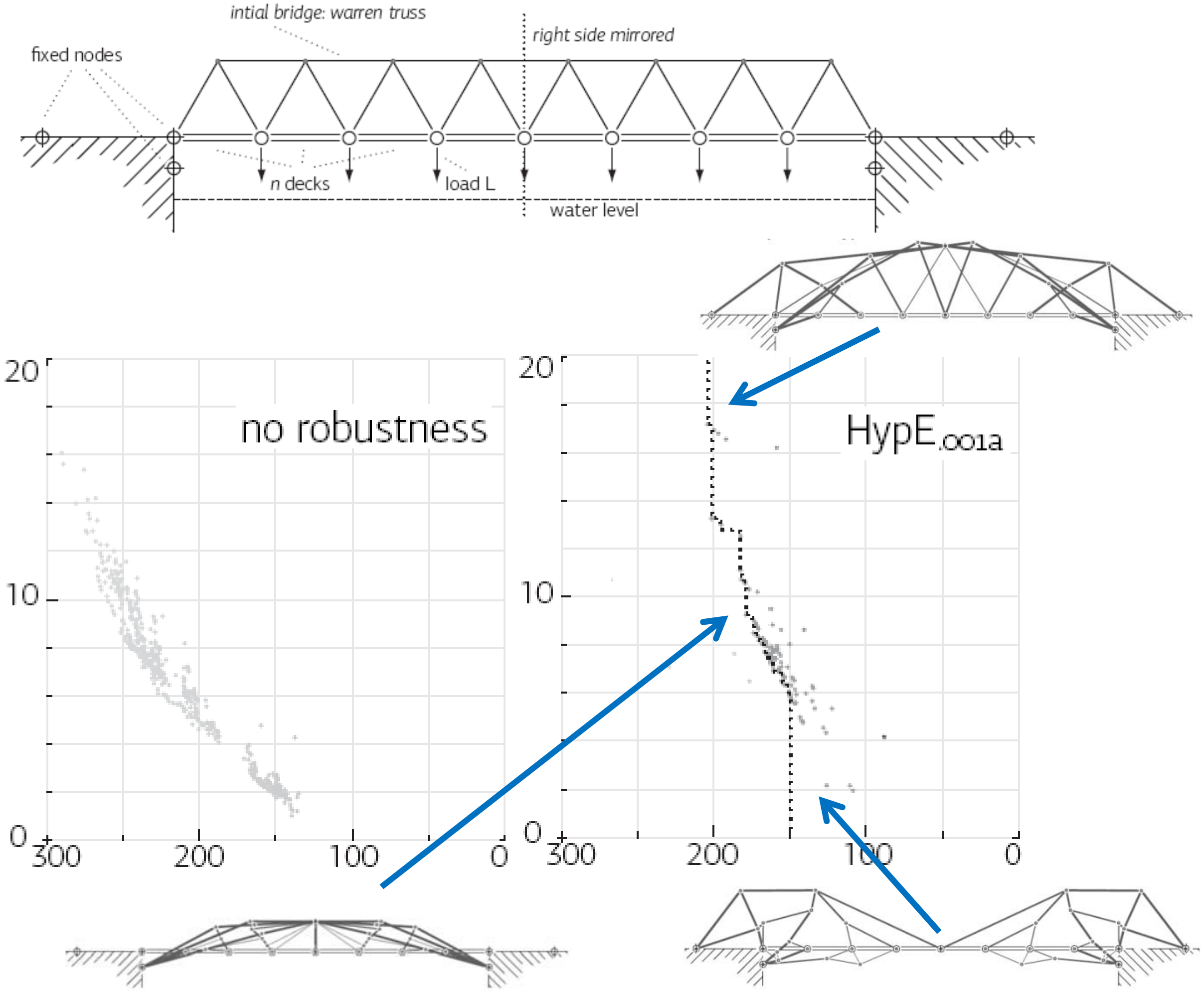




Application: Design Space Exploration

Truss Bridge Design

[Bader 2010]

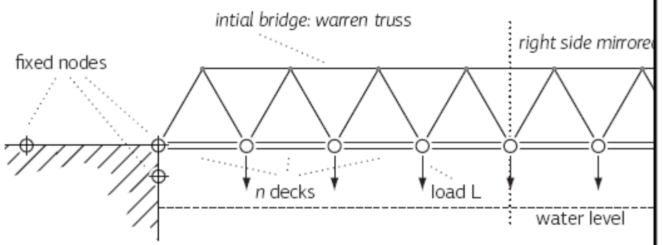


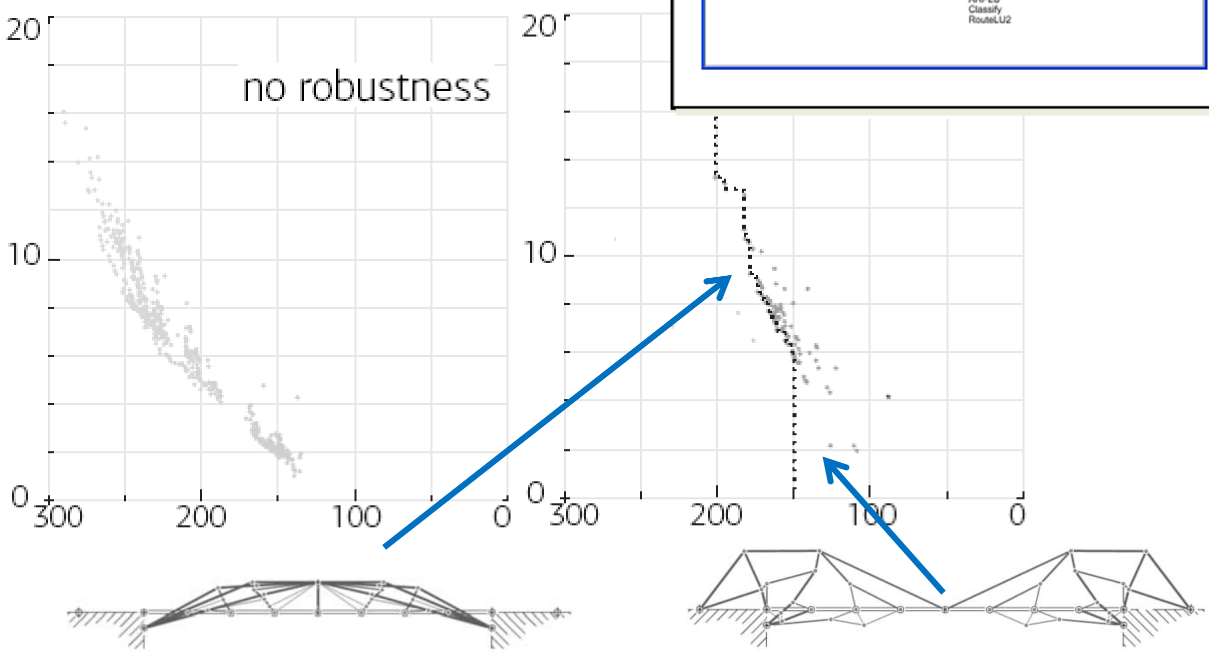
Implementation

Application: Design Space Exploration

Truss Bridge Design

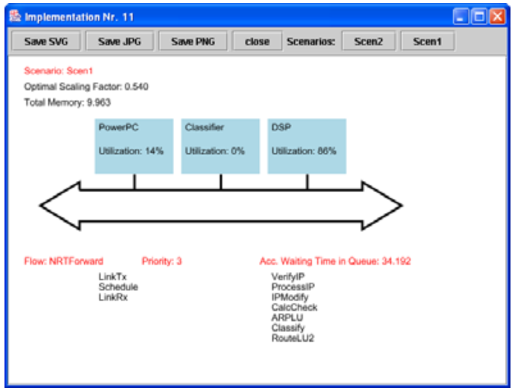
[Bader 2010]

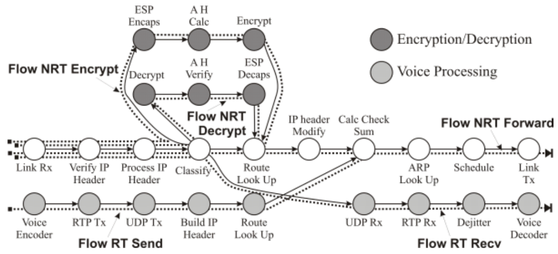


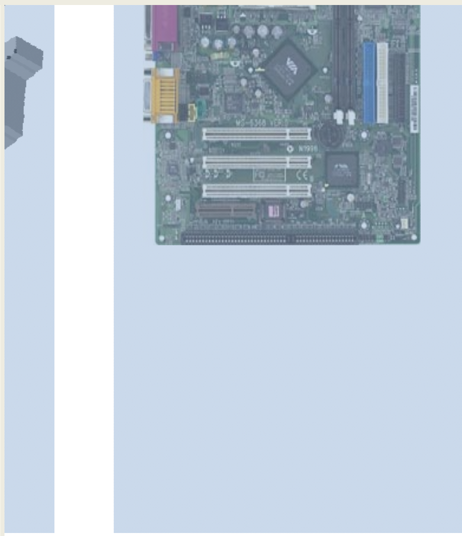


Network Processor Design

[Thiele et al. 2002]



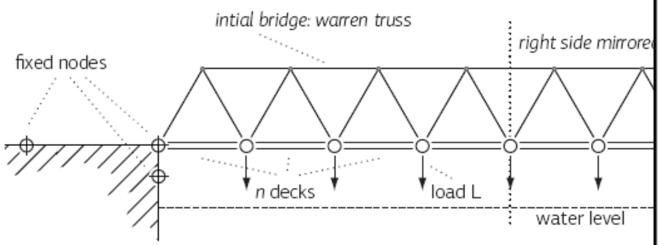




Application: Design Space Exploration

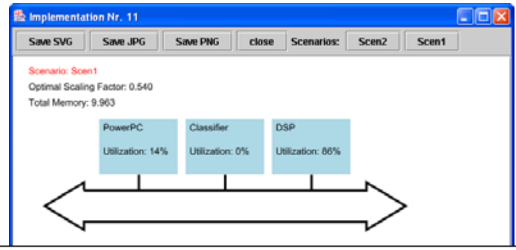
Truss Bridge Design

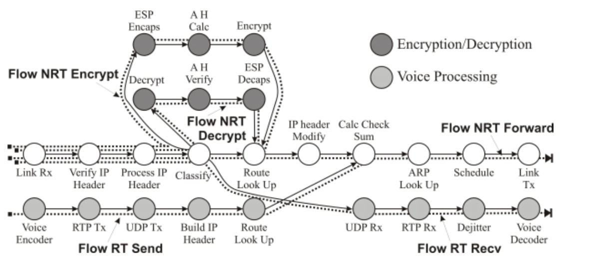
[Bader 2010]

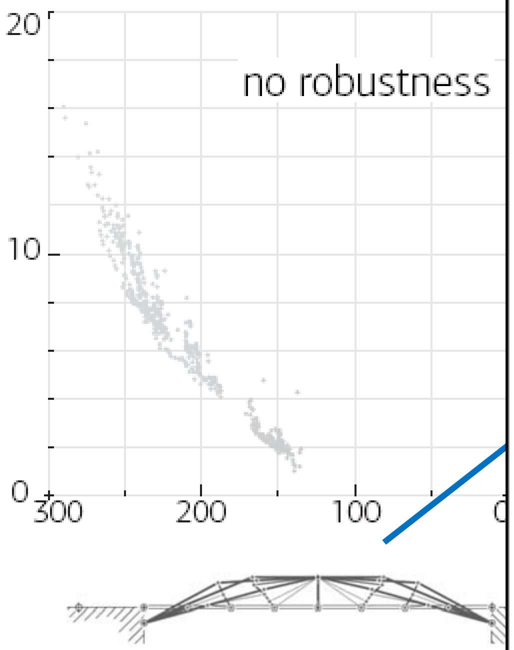


Network Processor Design

[Thiele et al. 2002]

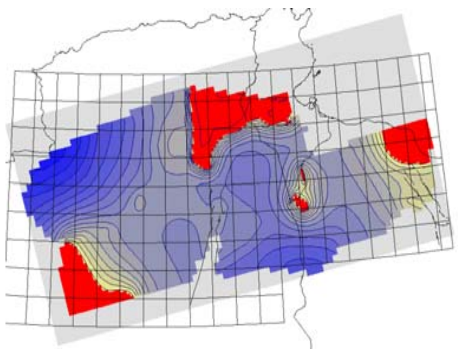


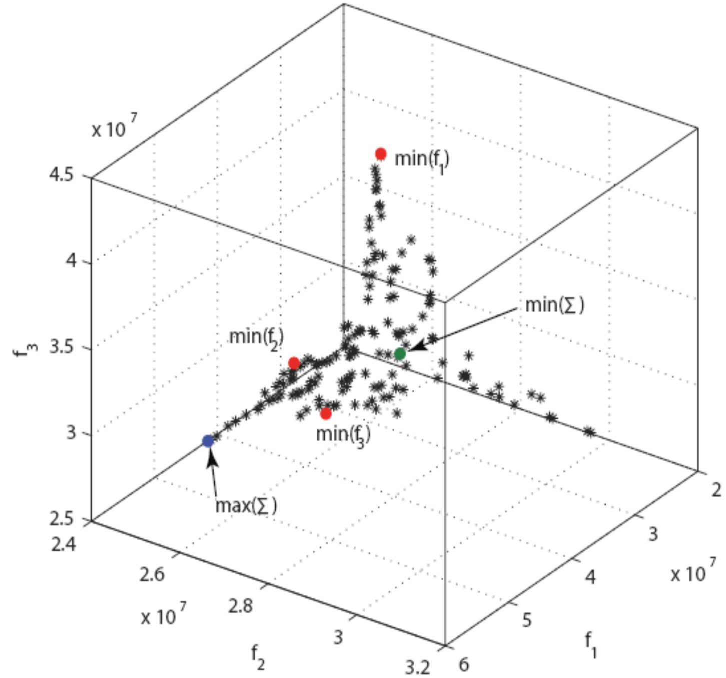




Water resource management

[Siegfried et al. 2009]



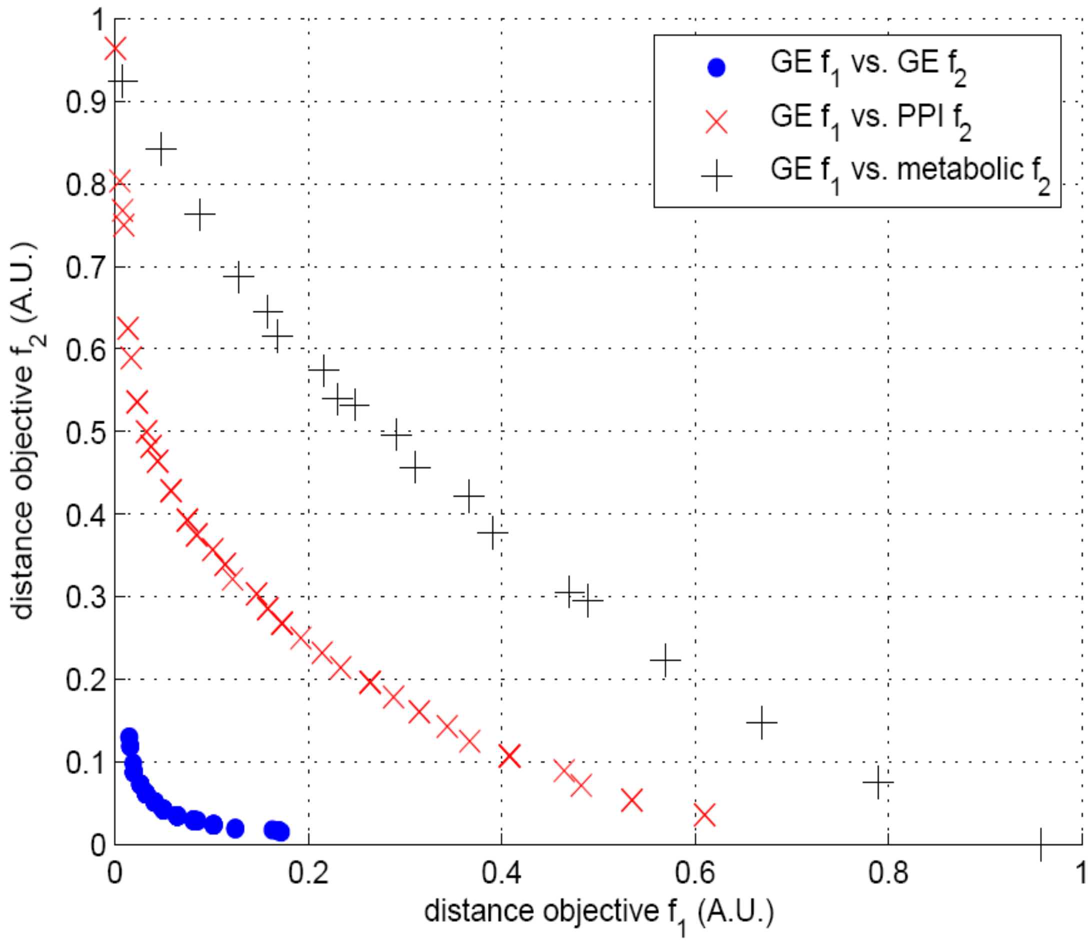


Application: Trade-Off Analysis

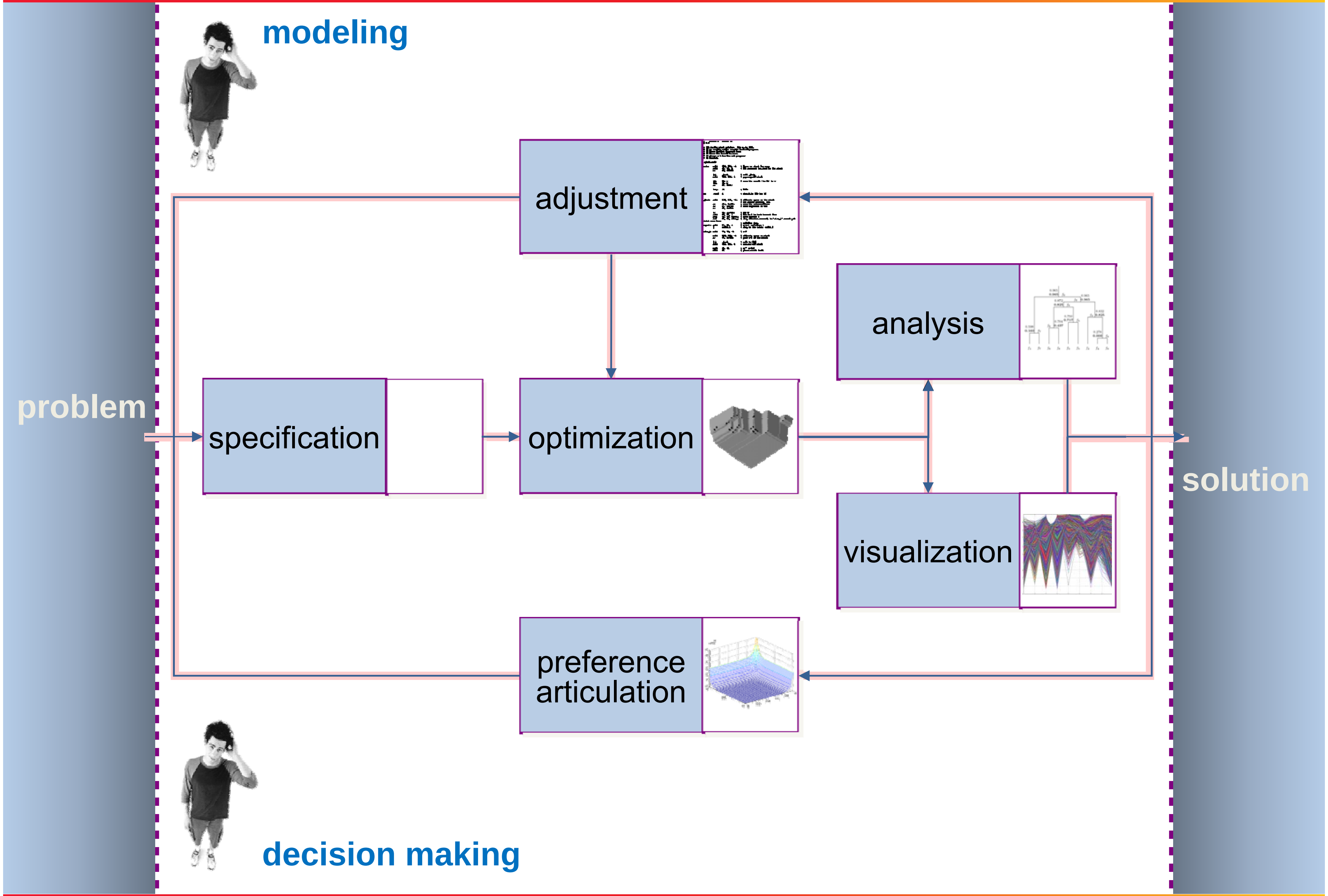
Module identification from biological data [Calonder et al. 2006]

Find group of genes wrt different data types:

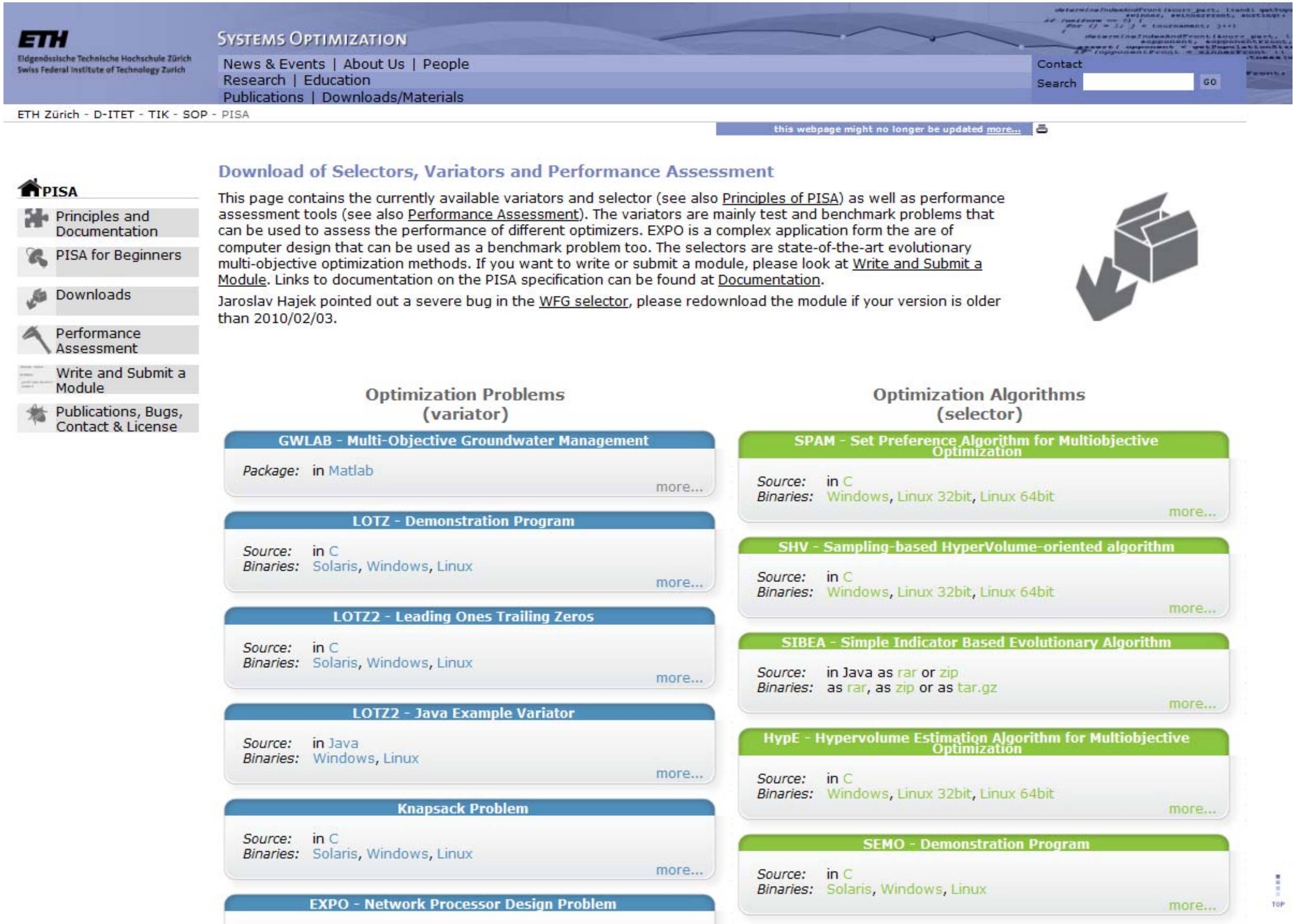
- similarity of gene expression profiles
- overlap of protein interaction partners
- metabolic pathway map distances



Conclusions: EMO as Interactive Decision Support



PISA: http://www.tik.ee.ethz.ch/pisa/



Exercise:
Runtime Analysis of a Simple EMO Algorithm

Constrained Optimization

Constrained Optimization

- up-to-now only unconstrained problems considered
- but constraints are frequent in practice
  - most combinatorial optimization problems have constraints (think about knapsack, scheduling, ...)
  - also continuous problems might have constraints (remember the initial Ariane launcher problem?)

$$\min f(x)$$
$$\text{s.t.}$$
$$g_i(x) \leq 0 \text{ for all } 1 \leq i \leq m$$
$$h_j(x) = 0 \text{ for all } \quad 1 \leq j \leq p$$

Main approaches:

- straightforward rejection sampling
- penalty functions
- special representations and operators
- multiobjective formulation

The Simplest Approach: Rejection Sampling

until candidate solution is feasible: resample the search space
according to current probability distribution

- no information about infeasible region is used
- not applicable if feasible region is (too) small
- other approaches used (much) more frequently

Penalty Functions

- transform constrained problem into unconstrained one: incorporate (the amount of) constraint violations into the objective function
- already proposed in the 1940s by Richard Courant
- general example:

constants

$$\phi(x) = f(x) + \sum_{i=0}^{m} \gamma_i \cdot G_i + \sum_{i=0}^{p} \eta_i \cdot H_i$$

"old" function

unconstrained function

function of $g_i(x)$

function of $h_i(x)$

Remarks:

- death penalty = rejection sampling
- difficult to find appropriate constants and G_i and H_i functions

Special Representations And Operators

we have seen already examples:

- TSP: permutations
- single-objective knapsack problem: no details

- either operators directly produce a feasible solution (like for TSP)
- or after the operator is used, a repair strategy transforms any infeasible solution into a feasible one
  - e.g. in knapsack problem: greedily reduce the number of chosen items according to $\text{profit}_i/\text{weight}_i$ until constraint fulfilled

Constraints and Multiobjective Optimization

Idea 1:

use a multiobjective algorithms to solve a problem where all constraints are objective functions

Idea 2:

solve a bi-objective problem where the second objective is the sum of all constraint violations

Possible problem:

the algorithm might only search in the infeasible domain

Box Constraints With CMA-ES

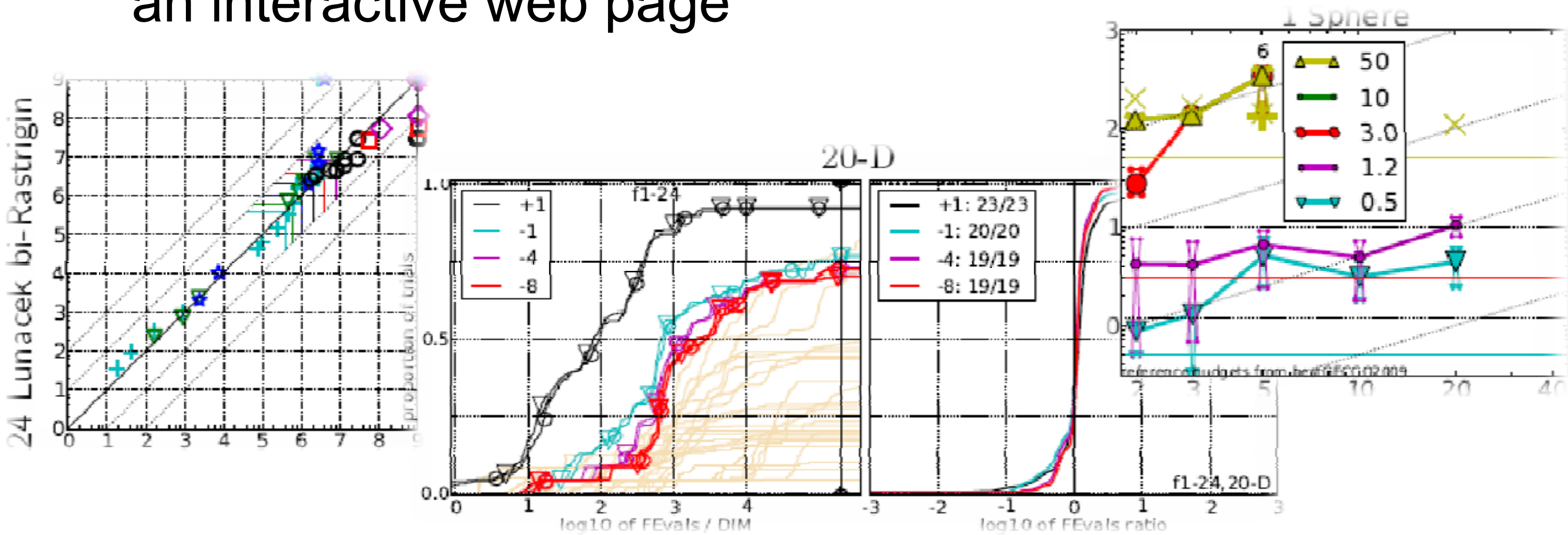
Main idea of standard implementation:

- using standard covariance matrix update with unchanged differences as if the new points are feasible
- projection of infeasible points onto boundary variable-by-variable to compute a feasible function value
- no penalty

Possible Thesis Projects

INRIA Saclay – Ile-de-France with Anne

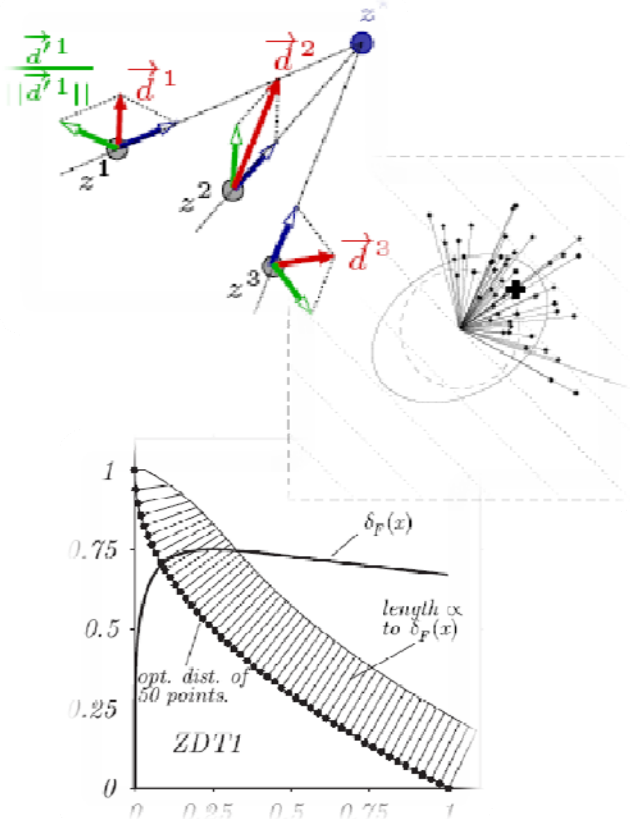
- more related to continuous single-objective optimization
- both theoretical and practical projects
- several possible options with respect to benchmarking
  - constrained optimization
  - large-scale optimization
  - expensive optimization
  - more concrete: visualizing algorithm comparison results on an interactive web page



INRIA Lille – Nord Europe with Dimo

More related to multiobjective optimization (but probably also possible in Saclay)

- scalarization/aggregation-based EMO (also parallelization possible)
- variation operators for set-based EMO
- stopping criteria and restarts for EMO
- optimal μ-distributions (both theoretical and numerical)



for questions, just ask us:
firstname.lastname@inria.fr

Good luck for the exam!