Selected Research Topics in Stochastic Blackbox Optimization

Dimo Brockhoff

KanGAL, IIT Kanpur, December 8, 2010

Dimo Brockhoff

2000-2005

study of CS (Dipl. inform.) in
Dortmund, Germany

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

2005-2009

Dr. sc. ETH at
ETH Zurich, Switzerland

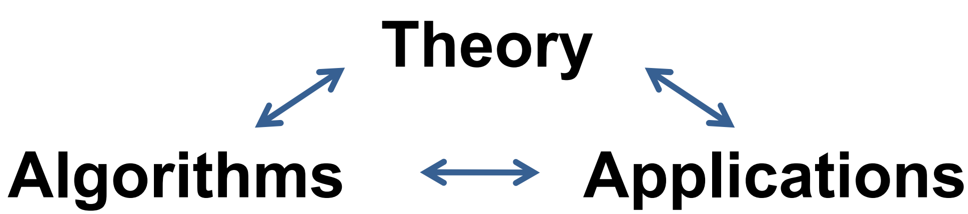
2009- 2010

postdoc at
INRIA Saclay---Ile-de-France

since November 2010

postdoc at
Ecole Polytechnique

Main Interests

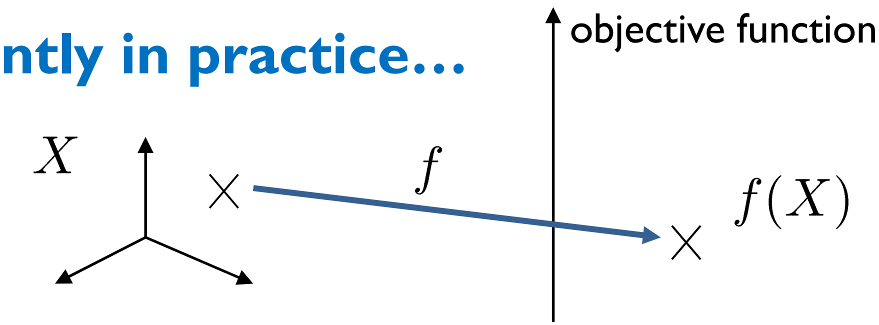


Blackbox Optimization: The Big Picture

Optimization problems occur frequently in practice...

$$\min_{x \in X} f(x) = (f_1(x), \ldots, f_k(x))$$

where $x \in X \mapsto f(x) \in \mathbb{R}^k$



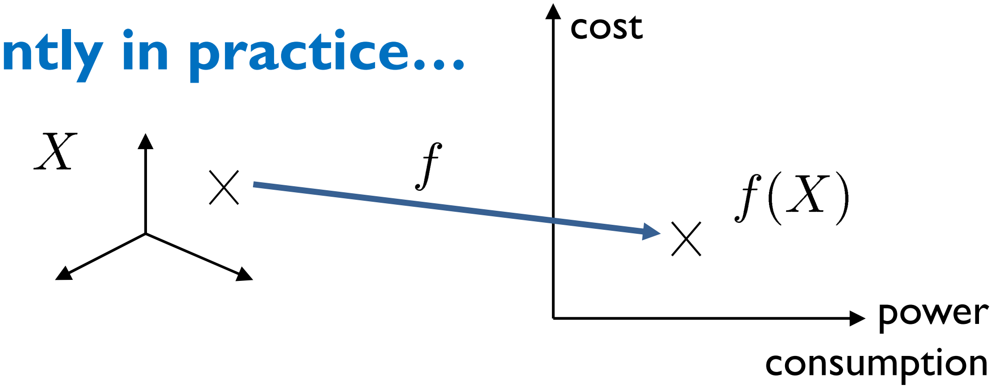
single-objective: total order

Blackbox Optimization: The Big Picture

Optimization problems occur frequently in practice…

$$\min_{x \in X} f(x) = (f_1(x), \ldots, f_k(x))$$

$$\text{where } x \in X \mapsto f(x) \in \mathbb{R}^k$$



multiple objectives:

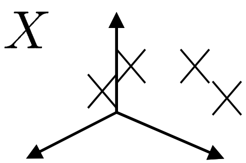
dominance relation (partial order)

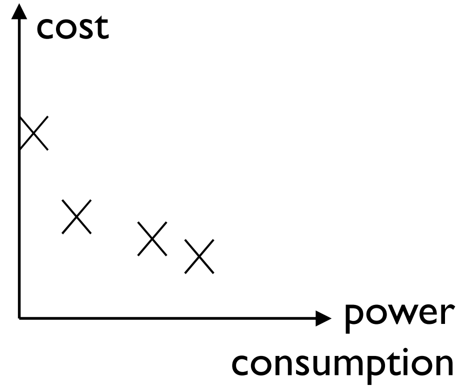
Blackbox Optimization: The Big Picture

Optimization problems occur frequently in practice...

$$\min_{x \in X} f(x) = (f_1(x), \ldots, f_k(x))$$

where $x \in X \mapsto f(x) \in \mathbb{R}^k$





multiple objectives:

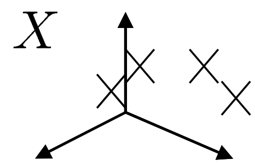
dominance relation (partial order)

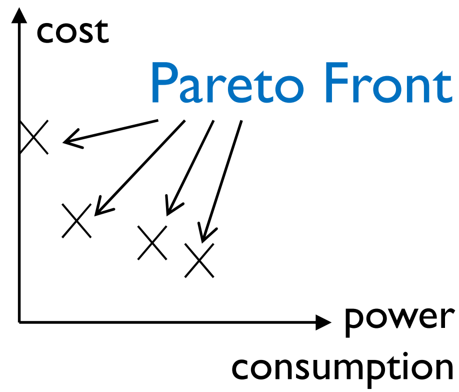
Blackbox Optimization: The Big Picture

Optimization problems occur frequently in practice...

$$\min_{x \in X} f(x) = (f_1(x), \ldots, f_k(x))$$

$$\text{where } x \in X \mapsto f(x) \in \mathbb{R}^k$$





multiple objectives:

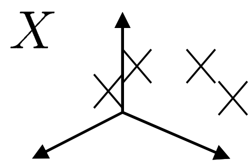
dominance relation (partial order)

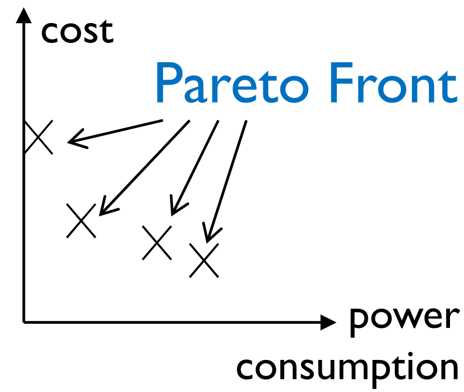
Blackbox Optimization: The Big Picture

Optimization problems occur frequently in practice...

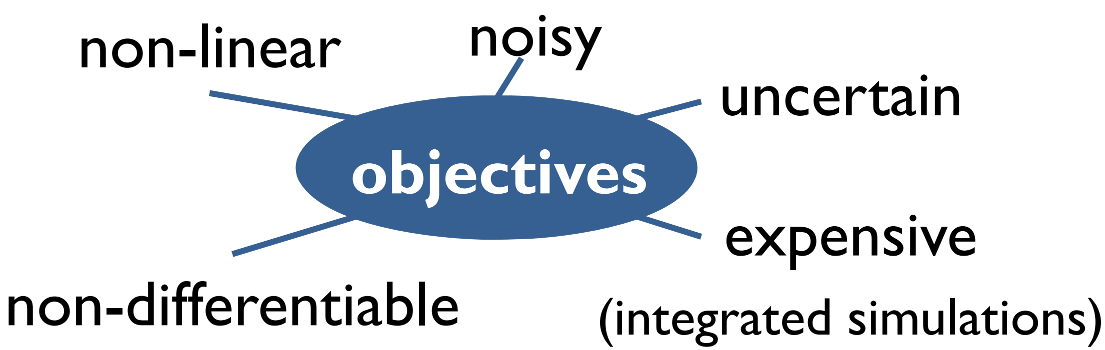
$$\min_{x \in X} f(x) = (f_1(x), \ldots, f_k(x))$$

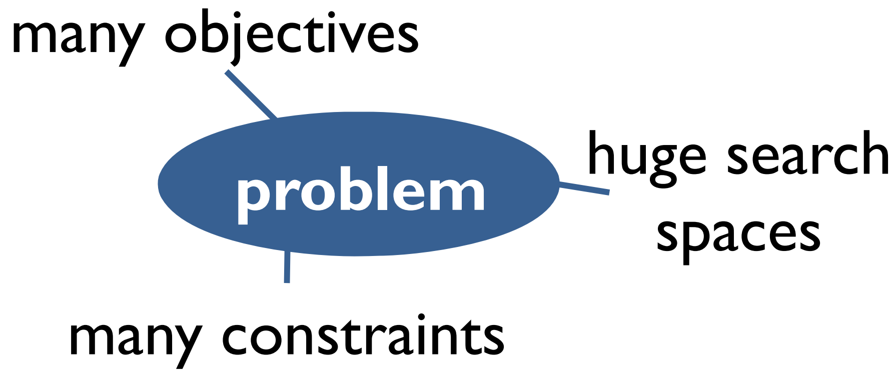
$$\text{where } x \in X \mapsto f(x) \in \mathbb{R}^k$$





Issues:



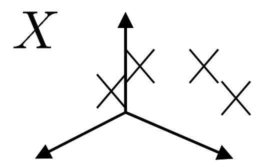


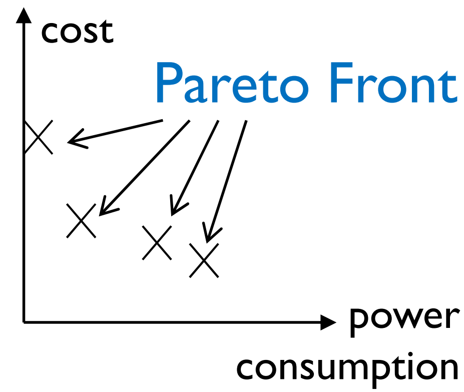
Blackbox Optimization: The Big Picture

Optimization problems occur frequently in practice...

$$\min_{x \in X} f(x) = (f_1(x), \ldots, f_k(x))$$

$$\text{where } x \in X \mapsto f(x) \in \mathbb{R}^k$$

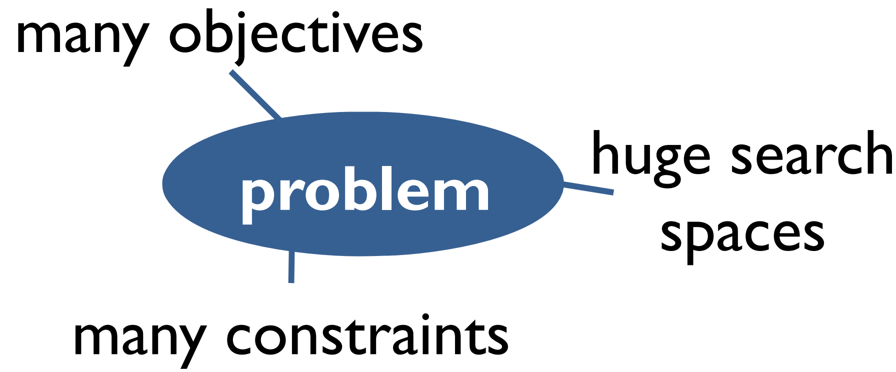




Black box optimization

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

$\longrightarrow$ no assumptions

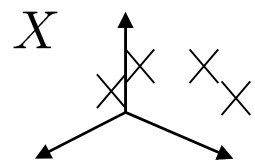


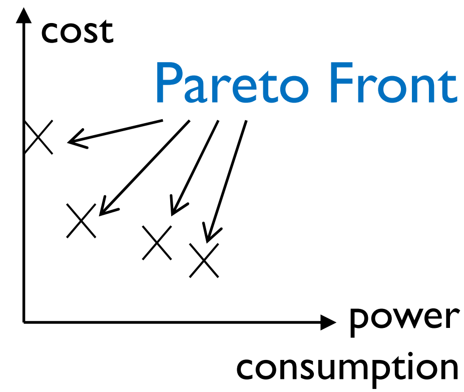
Blackbox Optimization: The Big Picture

Optimization problems occur frequently in practice...

$$\min_{x \in X} f(x) = (f_1(x), \ldots, f_k(x))$$

where $x \in X \mapsto f(x) \in \mathbb{R}^k$





Black box optimization

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

$\longrightarrow$ no assumptions

many objectives

problem

huge search spaces

many constraints

Evolutionary Algorithms (EAs)

Evolutionary algorithms =

randomized search algorithms

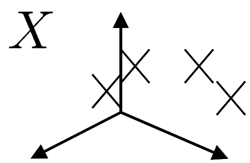
optimizing on solution **sets**

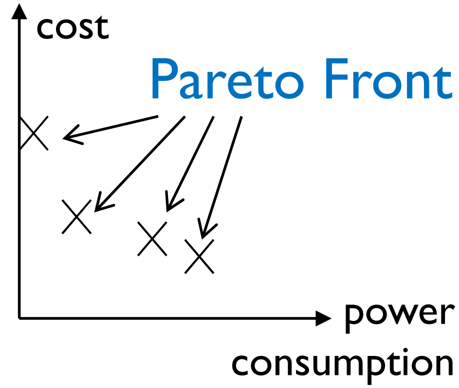
Blackbox Optimization: The Big Picture

Optimization problems occur frequently in practice...

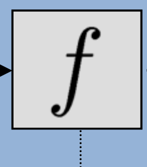
$$\min_{x \in X} f(x) = (f_1(x), \ldots, f_k(x))$$

where $x \in X \mapsto f(x) \in \mathbb{R}^k$



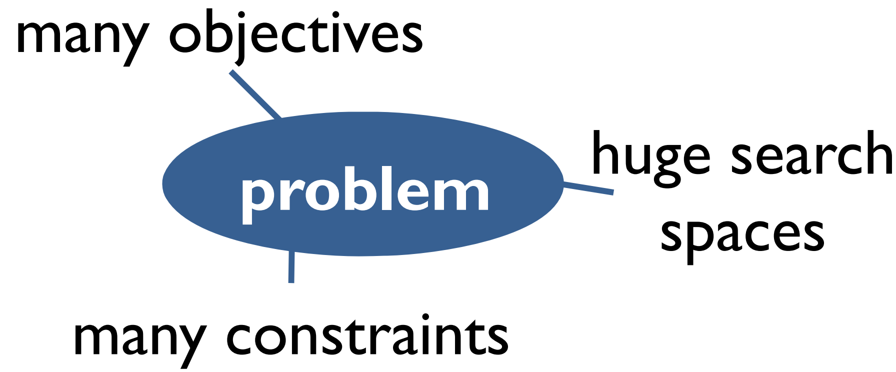


Black box optimization



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

$\longrightarrow$ no assumptions



Evolutionary Algorithms (EAs)

Evolutionary algorithms =

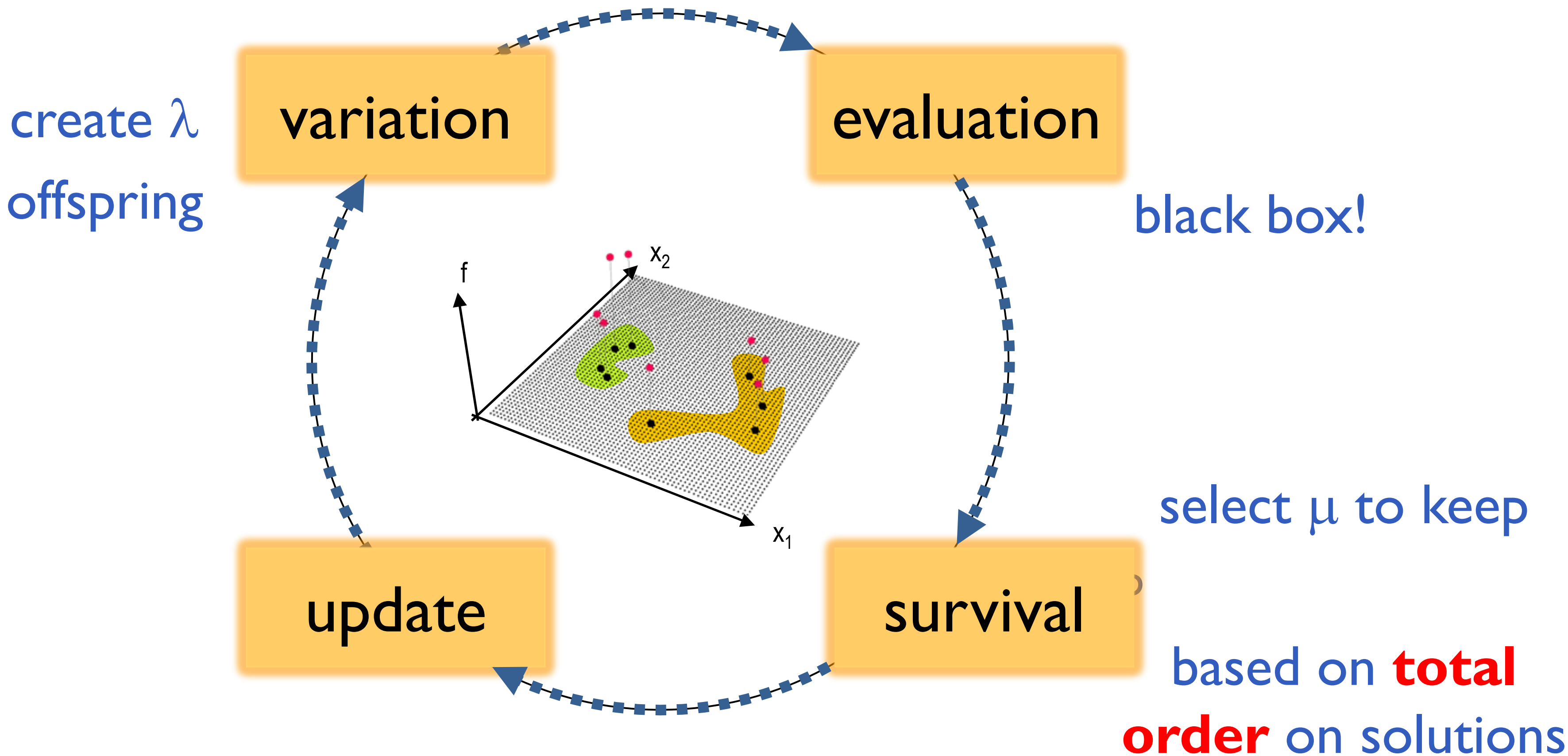
randomized search algorithms

optimizing on solution **sets**

- robust
- multimodal problems
- "sampling" the (Pareto-) optimal solutions to **inform** DM

General Principle of Evolutionary Algorithms

How these algorithms work…



Overview

Objective Reduction

Hypervolume-based search

Mirroring and Sequential Selection

Overview

Objective Reduction

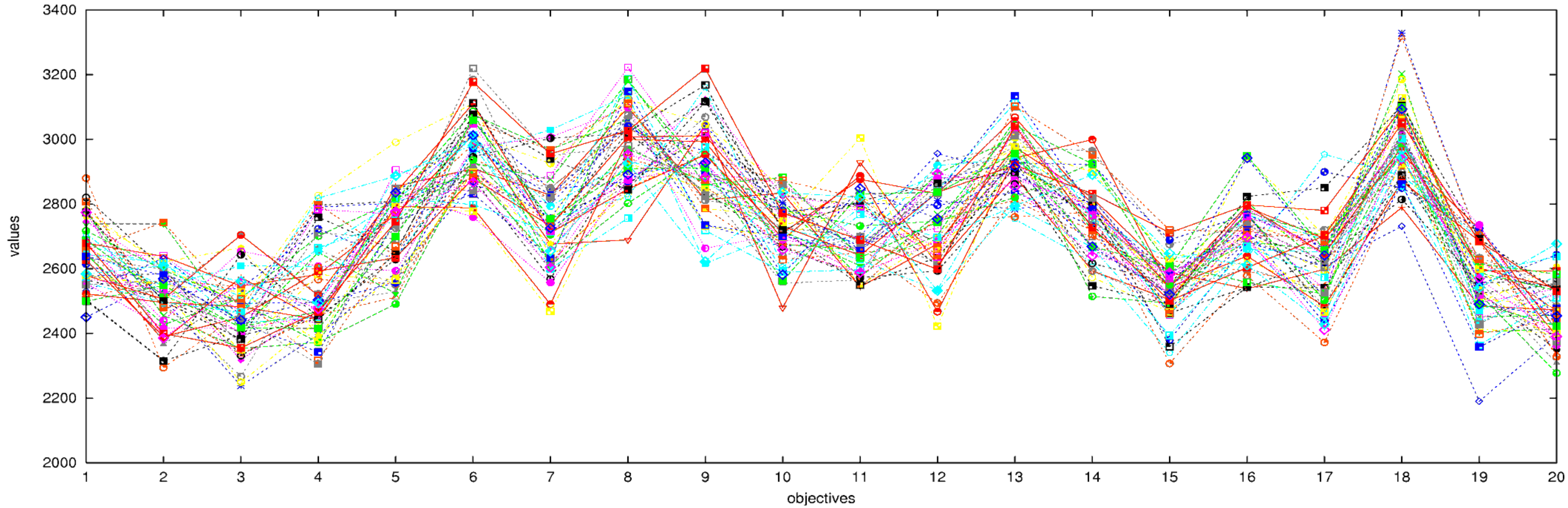
Hypervolume-based search

Mirroring and Sequential Selection

} multiobjective

} single-objective

Objective Reduction: Motivation



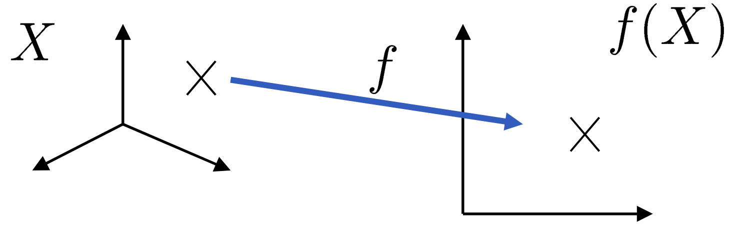
- Which objectives are the most important ones?
- What is the relationship between the objectives?
- Are all objectives necessary or can objectives be omitted?
- Are additional objectives always bad?
- Can efficient methods be developed?
- How can user preferences be incorporated into the search?

→ Learning about the problem

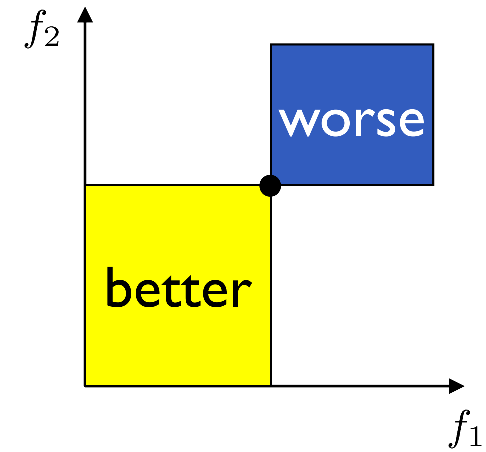
Recall: Multiobjective Optimization

- w.l.o.g. $\min_{x \in X} f(x) = (f_1(x), \ldots, f_k(x))$

  where $x \in X \mapsto f(x) \in \mathbb{R}^k$



- weak Pareto dominance relation wrt set $\mathcal{F} = \{f_1, \ldots, f_k\}$ of objectives: $x \preceq_{\mathcal{F}} y \Leftrightarrow \forall f_i \in \mathcal{F} : f_i(x) \leq f_i(y)$



- incomparable/comparable/indifferent
- $x^* \in X$ Pareto-optimal: $\nexists x \in X : x \preceq_{\mathcal{F}} x^* \wedge x^* \not\preceq_{\mathcal{F}} x$

Goal (without decision maker):

- find or approximate set of Pareto-optimal solutions (Pareto set)
- in practice: as close as possible & as diverse as possible

Set Problem View and Refinements

set problem: generalize Pareto dominance on sets

$$A \preceq_{\mathcal{F}} B \Leftrightarrow \forall b \in B : \exists a \in A : a \preceq_{\mathcal{F}} b$$

Sought: total refinement

i.e. a total order on sets that is compliant with dominance

Definition 2.4: Given a set Ψ. Then the preference relation $\preccurlyeq_{\text{ref}}$ refines $\preccurlyeq$ if for all $A, B \in \Psi$ we have

$$(A \preccurlyeq B) \wedge (B \not\preccurlyeq A) \Rightarrow (A \preccurlyeq_{\text{ref}} B) \wedge (B \not\preccurlyeq_{\text{ref}} A).$$

from [ztb2010a in IEEETEC'10]

Many-Objective Optimization

Main Problem

weak Pareto dominance gives no search direction with many objectives

Needed:

“more total” order

One Idea:

Reduce the number of objectives automatically

$\rightarrow$ omitting objectives results in a refinement!

Automated Objective Reduction

Related Work:

- MCDM approaches: [gl1977a, agre1997a, mali2006a, mt2007a, mt2008a, mali2008a]
  - for linear objectives only
- PCA-based: Deb and Saxena [ds2006a, sd2007a, sd2008b]
  - no control over dominance relation ("what happens?")

MOSS: The Minimum Objective Subset Problem

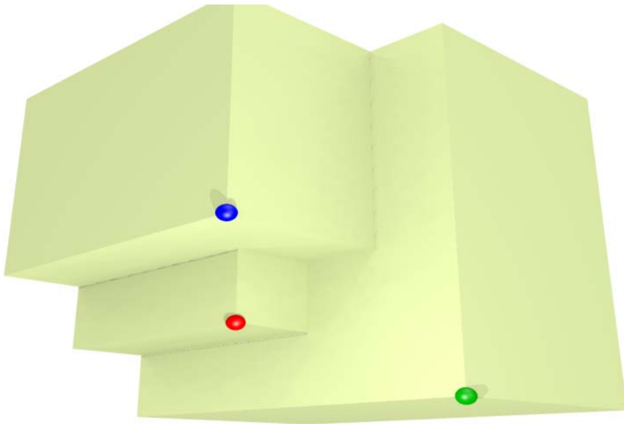
Given a set A of solutions with relations $\preceq_{f_i} \subseteq A \times A$,

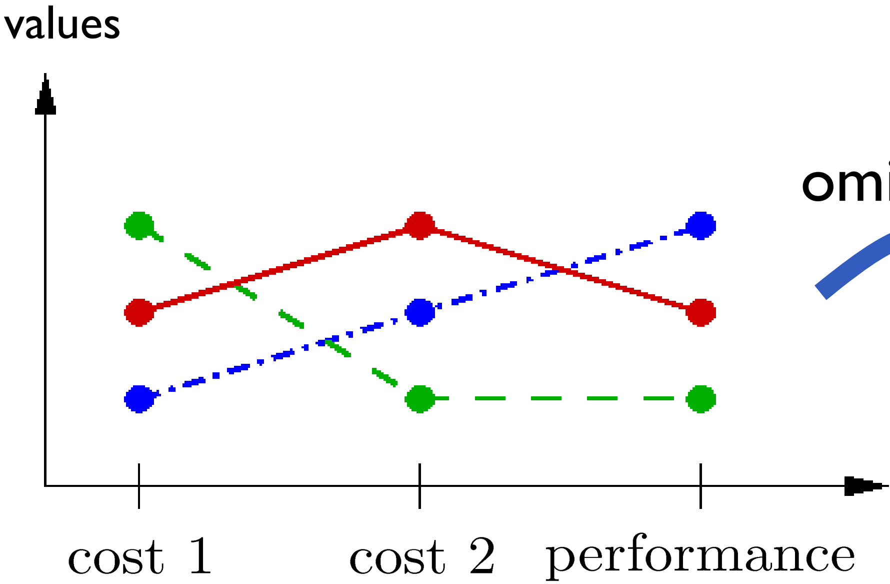
Find minimum objective set $\mathcal{F}' \subseteq \mathcal{F}$ preserving the relation ($\preceq_{\mathcal{F}'} = \preceq_{\mathcal{F}}$)

MOSS is NP-complete

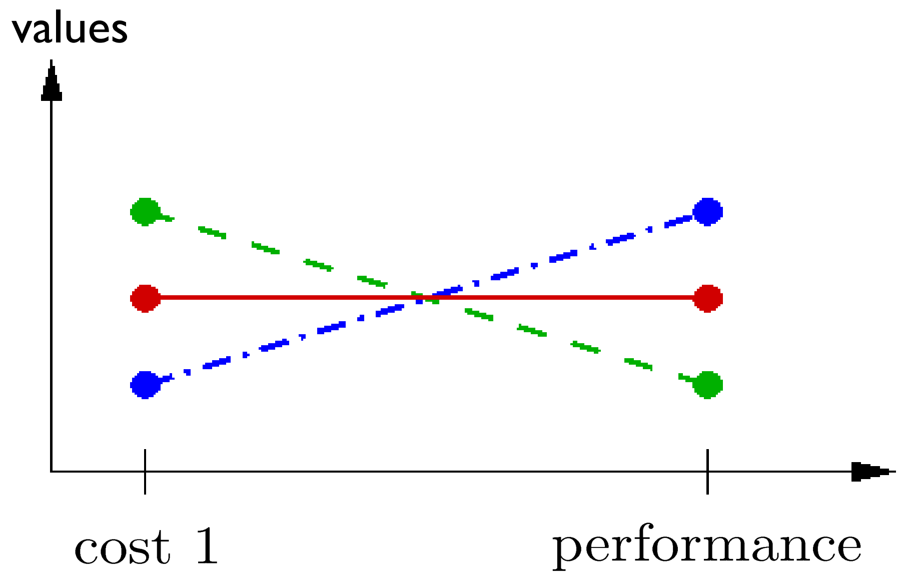
- Reduction from SETCOVER
- As a result, consideration of objective sets of fixed size is not sufficient

An Example





omit cost 2



still the same relations

Algorithms for the MOSS Problem [bz2007d]

Exact algorithm

- Correctness proof
- Runtime: $O(|A|^2 \cdot k \cdot 2^k)$
- Worst case: $\Omega(|A|^2 \cdot 2^{k/3})$

$S := \emptyset$
for each pair $\mathbf{x}, \mathbf{y} \in A$ of solutions **do**
  $S_x := \{\{i\} \mid i \in \{1,\dots,k\} \wedge \mathbf{x} \preceq_i \mathbf{y} \wedge \mathbf{y} \not\preceq_i \mathbf{x}\}$
  $S_y := \{\{i\} \mid i \in \{1,\dots,k\} \wedge \mathbf{y} \preceq_i \mathbf{x} \wedge \mathbf{x} \not\preceq_i \mathbf{y}\}$
  $S_{xy} := S_x \sqcup S_y$ where
    $S_1 \sqcup S_2 := \{s_1 \cup s_2 \mid s_1 \in S_1 \wedge s_2 \in S_2$
      $\wedge (\nexists p_1 \in S_1, p_2 \in S_2 : p_1 \cup p_2 \subset s_1 \cup s_2)\}$
  if $S_{xy} = \emptyset$ **then** $S_{xy} := \{1,\dots,k\}$
  $S := S \sqcup S_{xy}$
end for
Output a smallest set $s_{\min}$ in S

Simple greedy heuristic

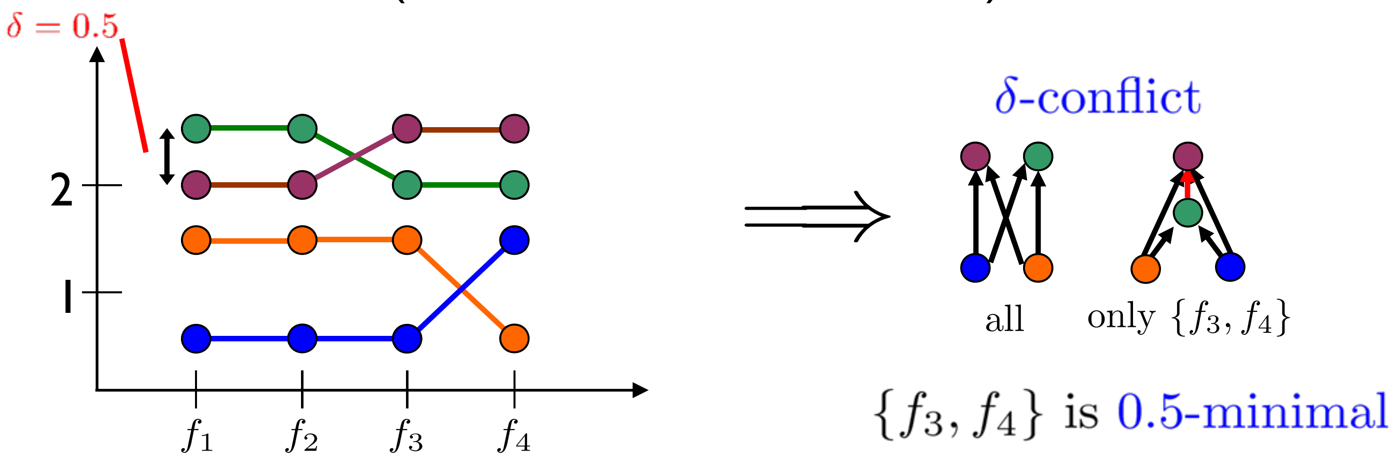
- Correctness proof
- Runtime: $O(k \cdot |A|^2)$
- Best possible approximation ratio of $\Theta(\log |A|)$ [feig1998a]

$E := \preceq_{\mathcal{F}}^C$ where $\preceq_{\mathcal{F}}^C := (A \times A) \setminus \preceq_{\mathcal{F}}$
$I := \emptyset$
while $E \neq \emptyset$ **do**
  choose an $i \in (\{1,\dots,k\} \setminus I)$
    such that $|\preceq_i^C \cap E|$ is maximal
  $E := E \setminus \preceq_i^C$
  $I := I \cup \{i\}$
end while

Generalizations

❶ Exactly conserving dominance structure sometimes too strict

$\rightarrow$ δ-error versions (based on ε-dominance) [bz2006d]

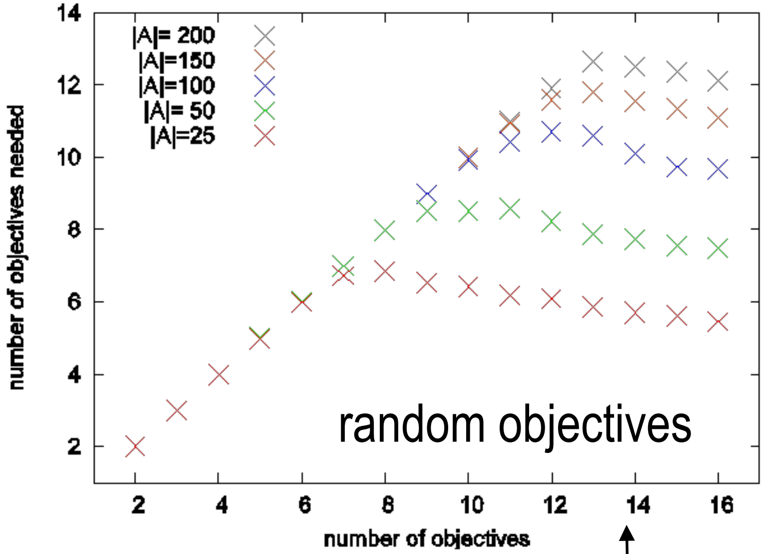


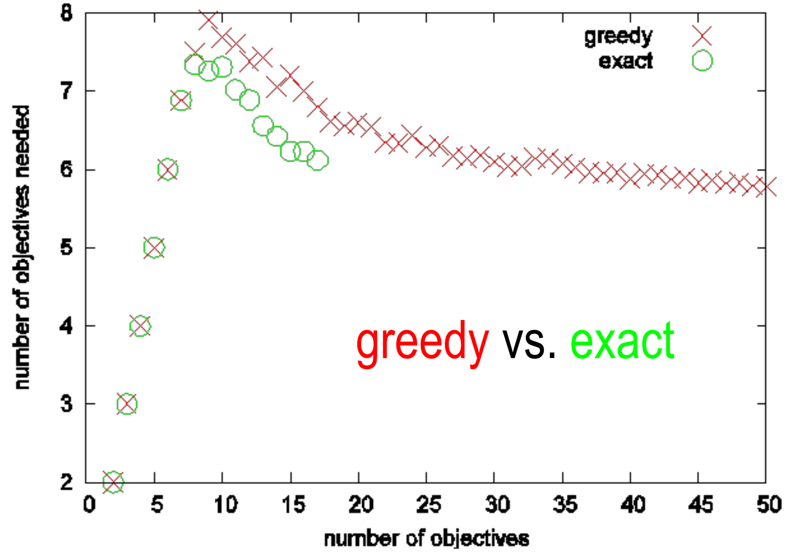
❷ Omitting Objectives during search might yield bad objective values

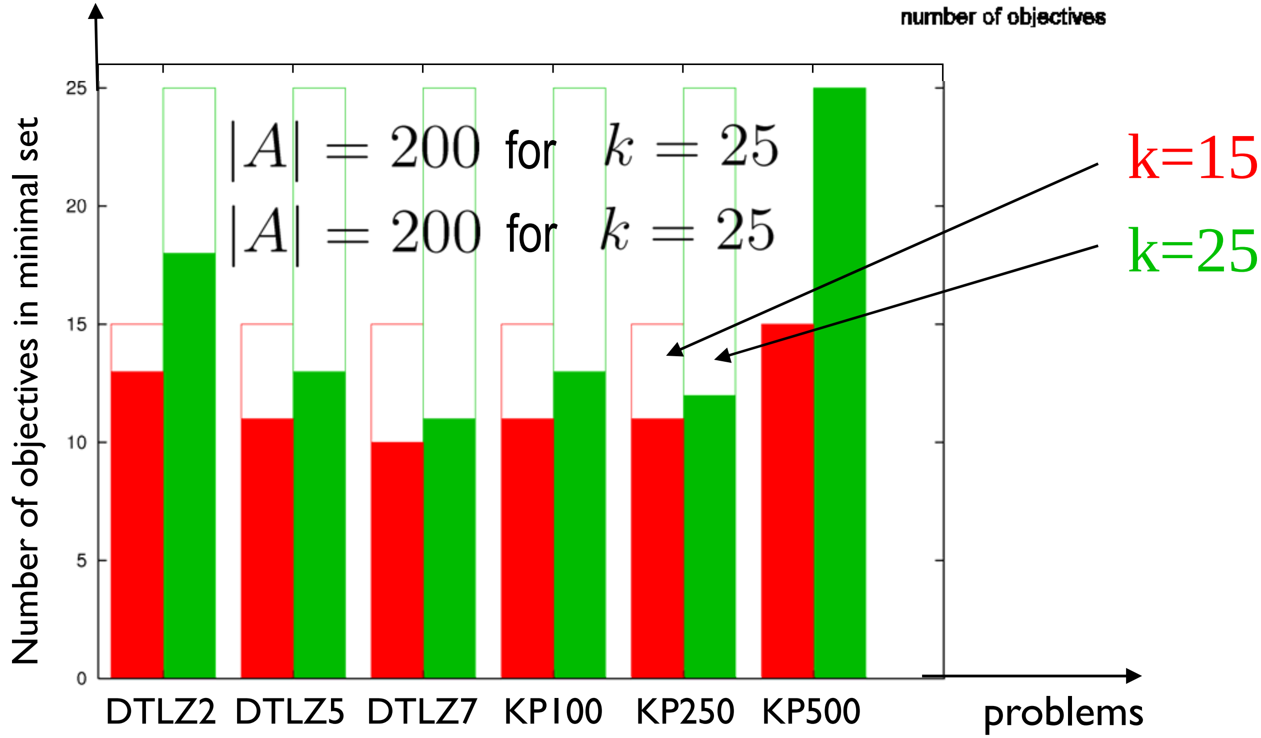
$\rightarrow$ Aggregation of objectives [bz2010a]

$\rightarrow$ Interestingly: also a refinement for weighted sum

Objective Reduction Results (Excerpts)







Objective Reduction: Open Problems

Online objective reduction

- fast algorithms
- objective aggregation
- objective decomposition
- one idea: using multi-armed bandits
- …

Decision Making

- what are the most important objectives?
- what can we learn more about the objective relations?
- incorporation of decision space
- …

Overview

Objective Reduction

Hypervolume-based search

Mirroring and Sequential Selection

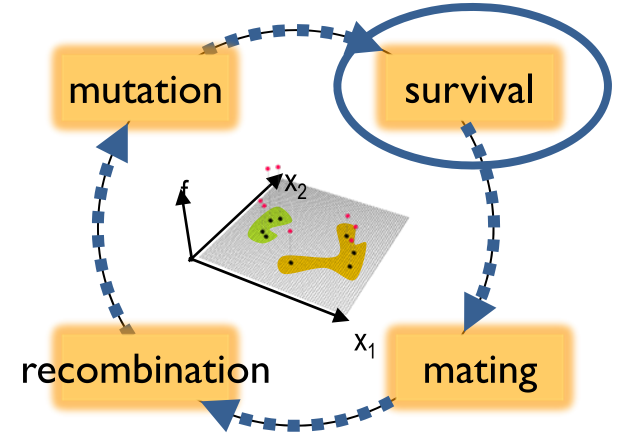
} multiobjective

} single-objective

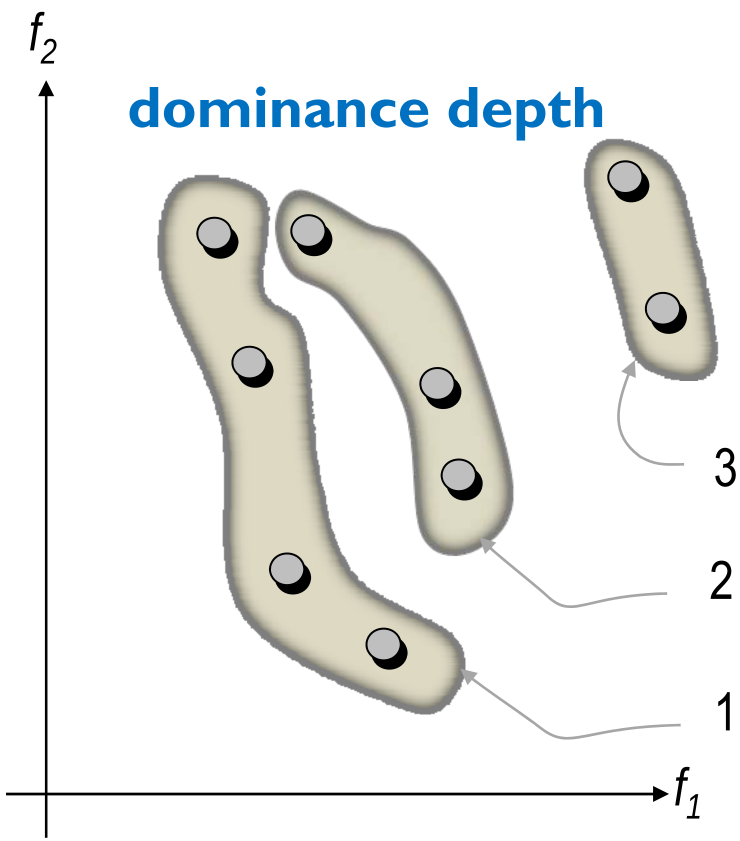
A "Classical" Algorithm: NSGA-II

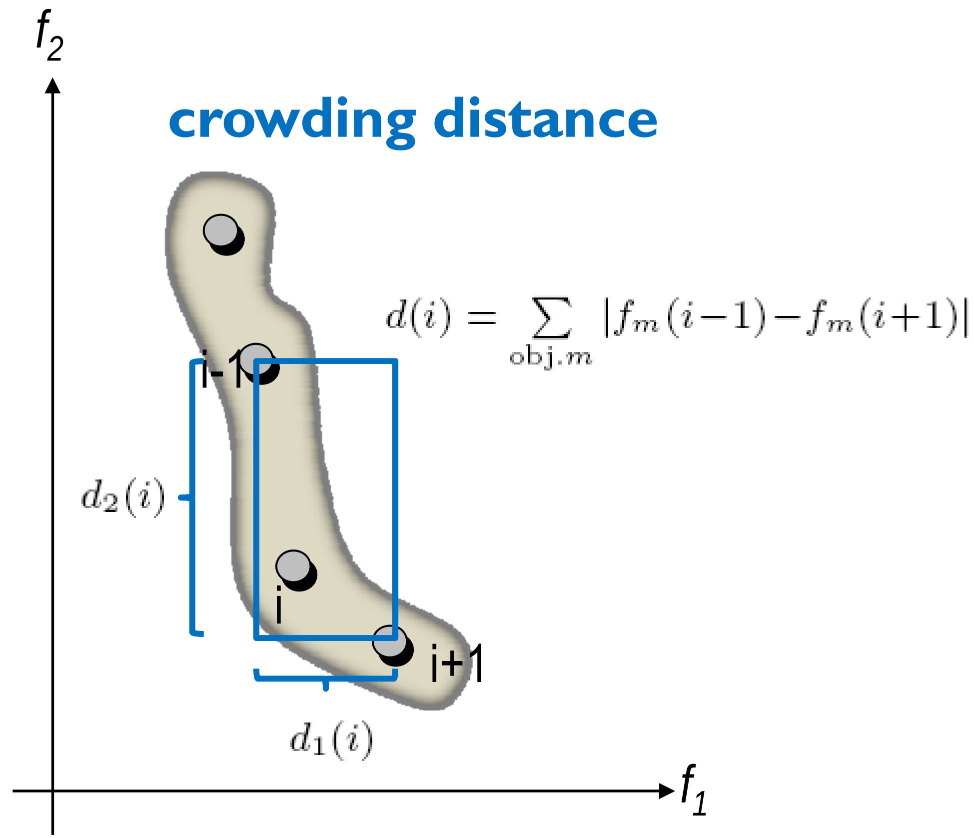
Selection:

NSGA-II [dapm2002a]



- 1[st] criterion: Pareto dominance
- 2[nd] criterion: crowding distance
- Optimizing crowding distance introduces cycles!



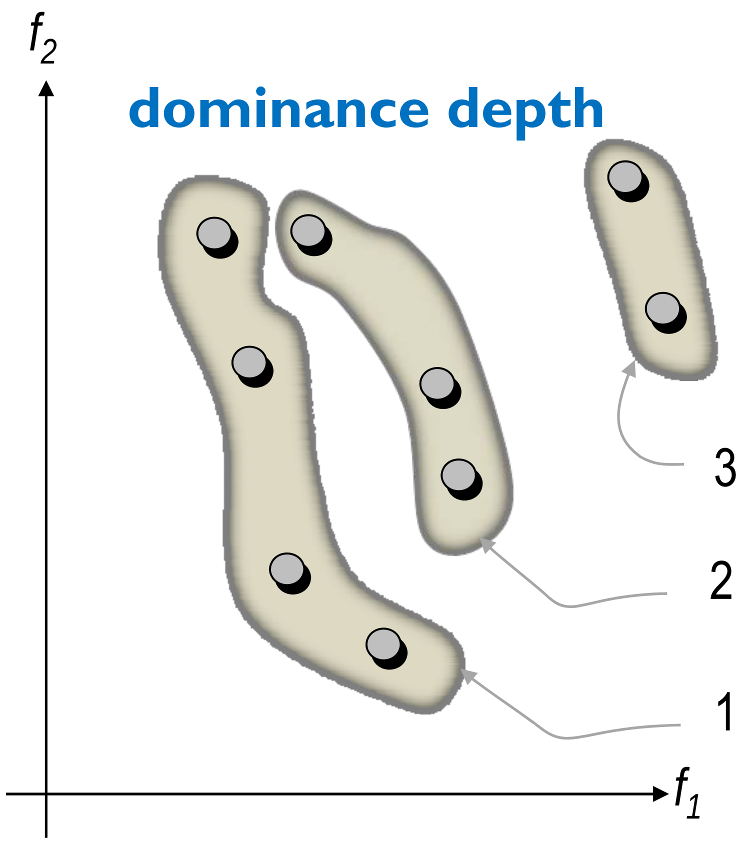


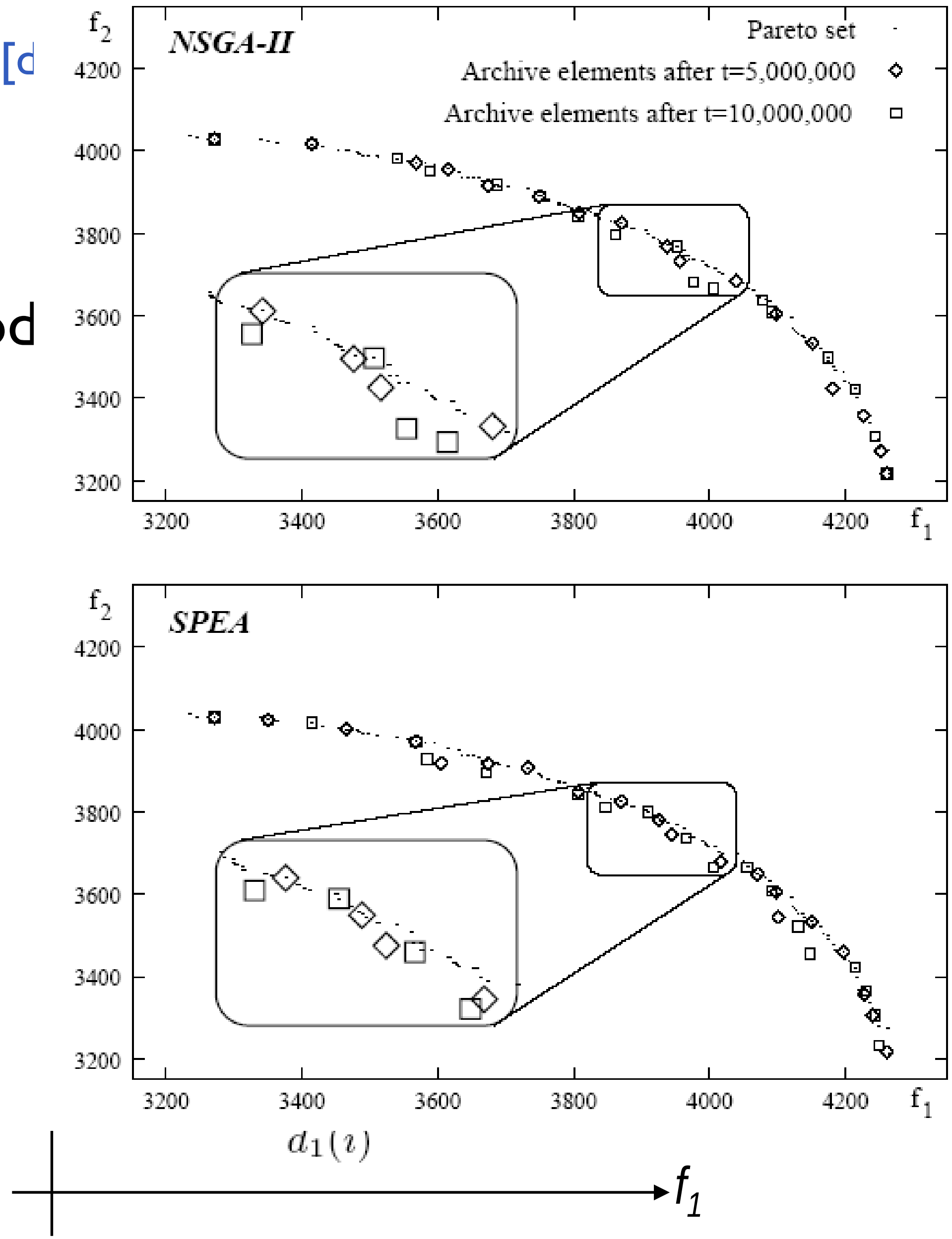
A "Classical" Algorithm: NSGA-II

Selection:

NSGA-II [d

- 1^{st} criterion: Pareto dominance
- 2^{nd} criterion: crowding distance
- Optimizing crowding distance introd





Hypervolume-Based Evolutionary Algorithms

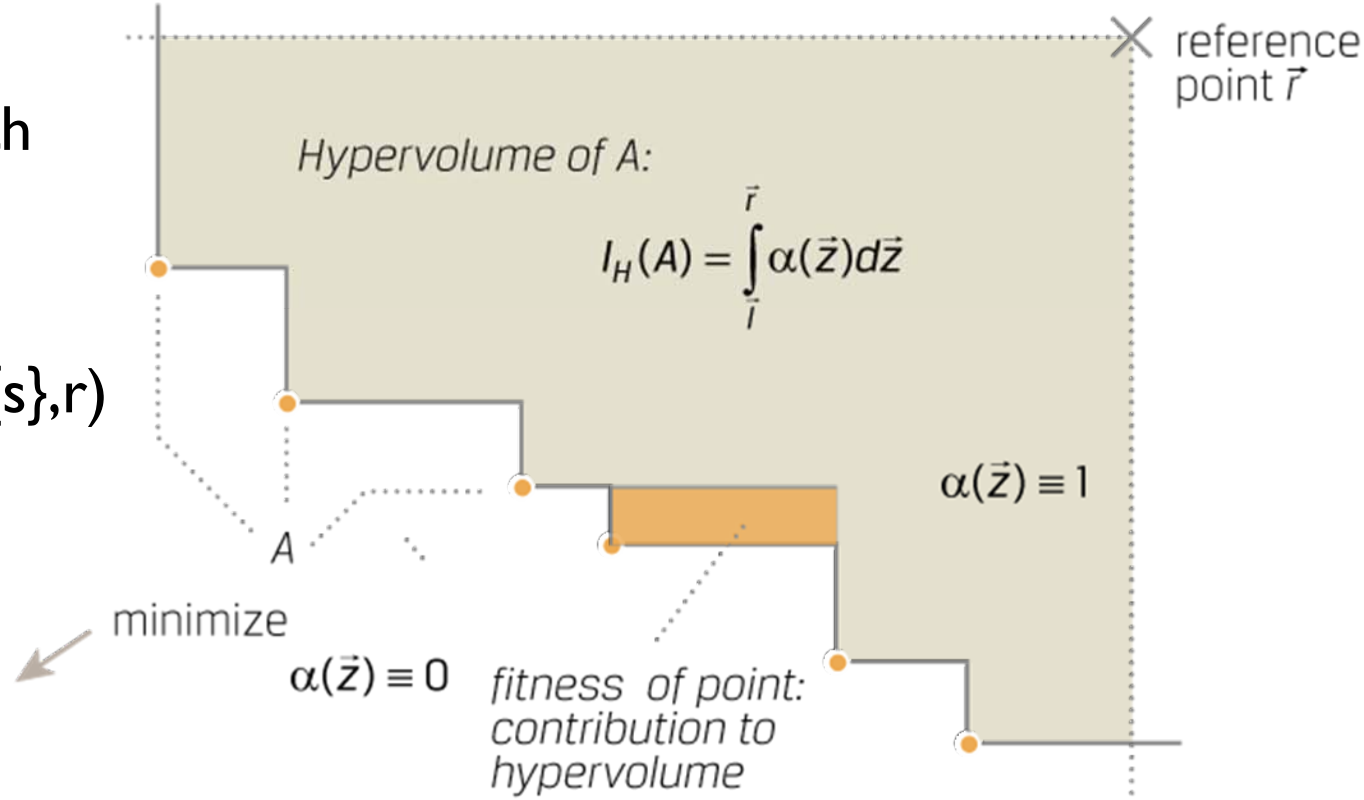
State-of-the-art algorithms (SMS-EMOA, MO-CMA-ES, HypE, …)

use hypervolume indicator as 2^{nd} selection criterion: no cycles!

refinement!

Main idea

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



Optimal μ-Distributions

When the goal is to maximize the hypervolume…

refinement!

- this yields sets with only Pareto-optimal solutions [flei2003a @ EMO'03]
- those sets, if unrestricted in size, cover the entire Pareto front
- many hypervolume-based EMO algorithms have a population size μ!

Optimal μ-Distribution:

A set of μ solutions that maximizes the hypervolume indicator among all sets of μ solutions is called optimal μ-distribution.

Optimal μ-Distributions

Questions:

- how are optimal μ-distributions characterized?
  - understand the bias of the indicator (influence on DM)
  - how can it be changed?
- what is their indicator value?
  - helpful for performance assessment (target values)
- what is the influence of the indicator's parameters on optimal μ-distributions?
  - guidelines for practical usage
- do algorithms converge to optimal μ-distributions?

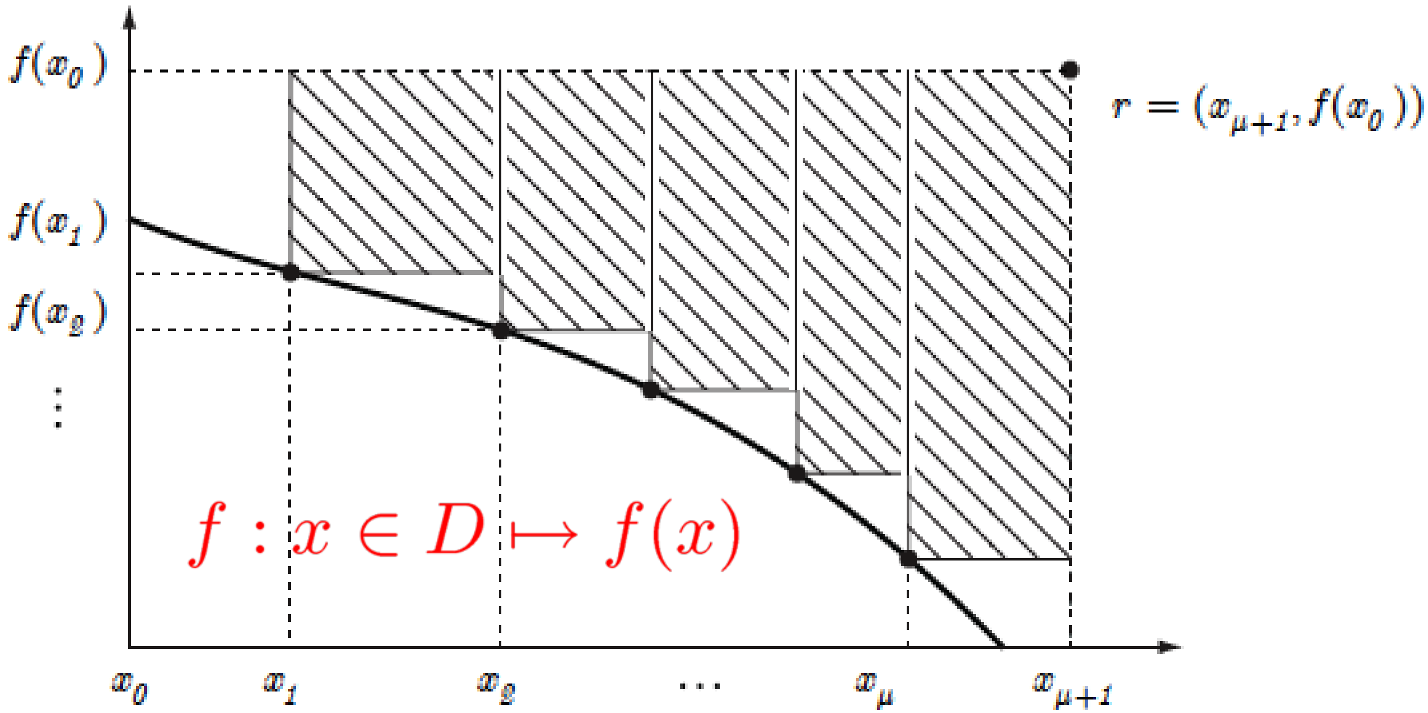
Optimal μ-Distributions

Questions:

- how are optimal μ-distributions characterized?
  - understand the bias of the indicator (influence on DM)
  - how can it be changed?
- what is their indicator value?
  - helpful for performance assessment (target values)
- what is the influence of the indicator's parameters on optimal μ-distributions?
  - guidelines for practical usage
- do algorithms converge to optimal μ-distributions?

Notations for 2-Objective Case [abbz2009a]

Results for 2 objectives only… (except [abb2010a])



hypervolume indicator: $$I_H((x_1, \ldots, x_\mu)) := \sum_{i=1}^{\mu} (x_{i+1} - x_i)(f(x_0) - f(x_i))$$

A Necessary Condition [abbz2009a]

Proposition 1. **(Necessary condition for optimal μ-distributions)** *If f is continuous, differentiable and $(x_1^\mu, \ldots, x_\mu^\mu)$ denote the x-coordinates of a set of μ points maximizing the hypervolume indicator, then for all $x_{min} < x_i^\mu < x_{max}$*

$$f'(x_i^\mu)\left(x_{i+1}^\mu - x_i^\mu\right) = f(x_i^\mu) - f(x_{i-1}^\mu),\ i = 1 \ldots \mu \qquad (3)$$

where f' denotes the derivative of f, $f(x_0^\mu) = r_2$ and $x_{\mu+1}^\mu = r_1$.

Proof idea:

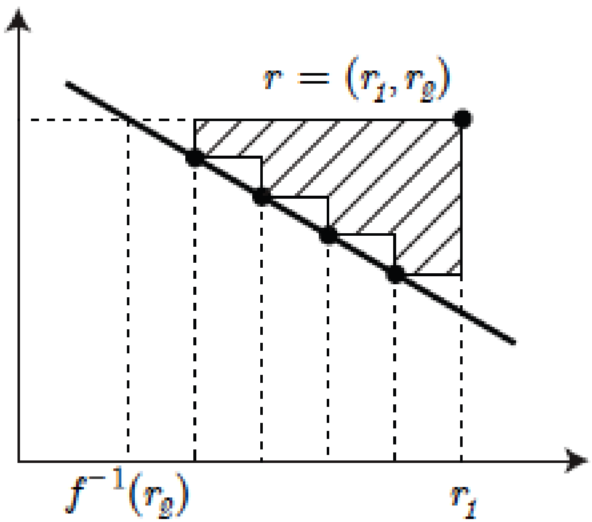
I_H max $\Rightarrow$ derivative is 0 at each x_i^μ or x_i^μ is at the boundary of the domain

Interpretation of Necessary Condition

Example: equal distances (only) on linear fronts

$$f : x \in [x_{min}, x_{max}] \mapsto \alpha x + \beta$$

$$\alpha \left(x^\mu_{i+1} - x^\mu_i\right) = f(x^\mu_i) - f(x^\mu_{i-1}) = \alpha(x^\mu_i - x^\mu_{i-1})$$

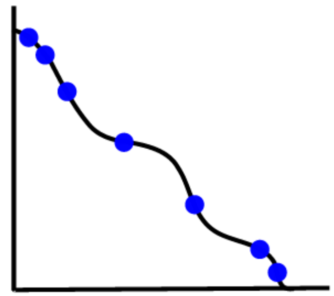


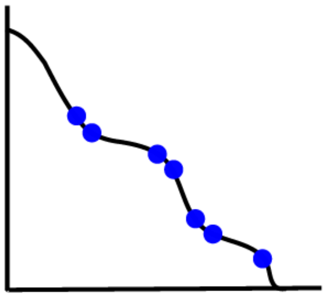
generalization of results in [ebn2005a,bne2007a]

Previous Belief About the Hypervolume

"Belief" about Bias:

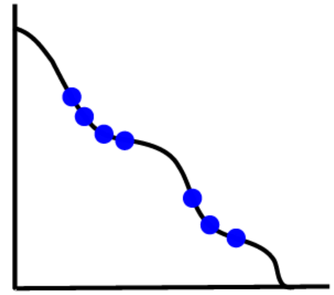
"biased towards the boundary solutions" [dmm2005a]





focuses on knee points; points less dense on extremes [bne2007a]

"convex regions might be preferred to concave regions" [zt1998b]



A Density Result: When μ Goes to Infinity

Observation:

general front shapes too difficult to investigate for finite μ

Question:

can we characterize optimal μ-distributions with respect to a density

$$\delta(x) = \lim_{\substack{\mu\to\infty \\ h\to 0}} \left(\frac{1}{\mu h} \sum_{i=1}^{\mu} \mathbf{1}_{[x,x+h[}(x_i^{\mu}) \right) ?$$

[abbz2009a]

Result and Interpretation

The resulting density is

$$\delta(x) = \frac{\sqrt{-f'(x)}}{\int_0^{x_{max}} \sqrt{-f'(x)}dx}$$

How can we interpret this?

- bias only depends on slope of f *in contrast to* [dmm2005a,zt1998b]
- density highest where slope = 45° *compliant to* [bne2007a]
- experimental results for finite and small μ support the result

Conclusion:

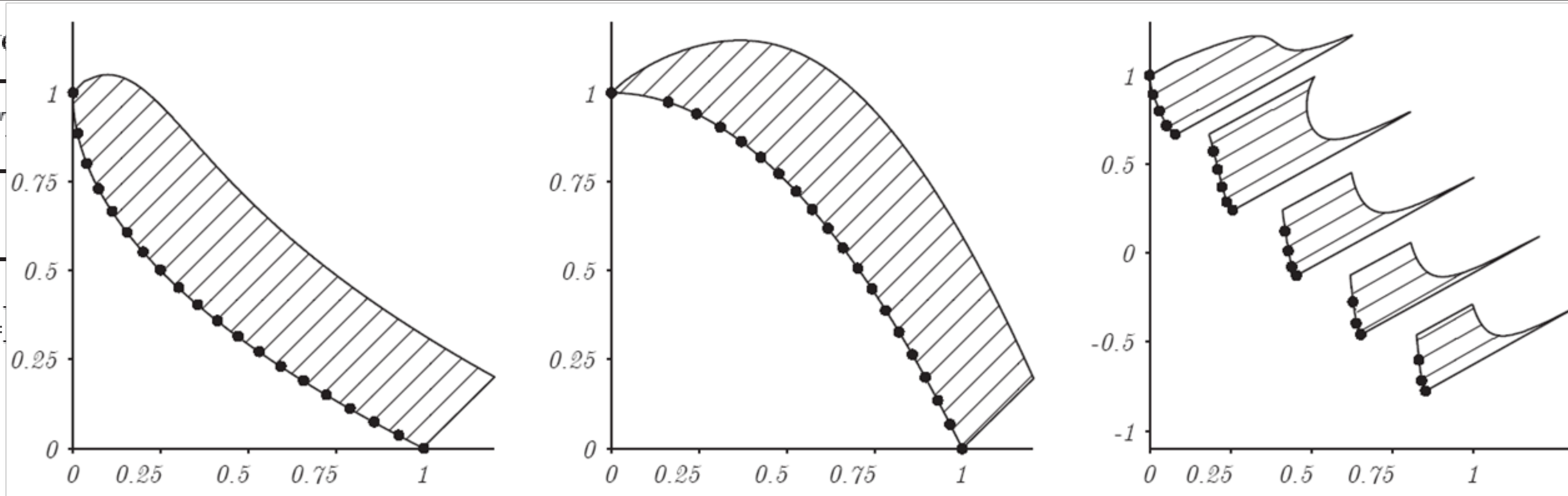
only theoretical results make it possible to understand the bias

How to Use the Result in Performance Assess.

Problem	front description	density
bi-objective sphere	$f(x) = \left((b-a) - x^{1/\alpha}\right)^{\alpha}$	$\delta(x) = C \cdot \sqrt{(b-a-x^{1/\alpha})^{\alpha-1} \cdot x^{\frac{\alpha-a}{\alpha}}}$
ZDT1, ZDT4 [24]	$f(x) = 1 - \sqrt{x}$	$\delta(x) = \frac{3}{4x^{1/4}}$
ZDT2 [24]	$f(x) = 1 - x^2$ for $x \in [0,1]$	$\delta(x) = \frac{3}{2}\sqrt{x}$
ZDT3* [24]	$f(x) = 1 - \sqrt{x} - x \cdot \sin(10\pi x)$ for all $x \in F$ where $F = [0, 0.0830015349] \cup]0.1822287280, 0.2577623634] \cup]0.4093136748, 0.4538821041] \cup]0.6183967944, 0.6525117038] \cup]0.8233317983, 0.8518328654]$	$\delta(x) = 1.5609 \cdot \sqrt{\frac{1}{2\sqrt{x}} + \sin(10\pi x) + 10\pi x \cos(10\pi x)}$
ZDT6 [24]	$f(x) = 1 - x^2$ for $x \in [\frac{\arctan(9\pi)}{6\pi}, 1] \approx [0.08146, 1]$	$\delta(x) = C \cdot \sqrt{x}$ with $C = \frac{3}{2}\left(1 - \frac{\arctan(9\pi)}{6\pi}^{3/2}\right)^{-1} \approx 1.53570$
DTLZ1 [8]	$f(x) = \frac{1}{2} - x$	$\delta(x) = 1$
DTLZ2, DTLZ3, DTLZ4 [8]	$f(x) = \sqrt{1 - x^2}$	$\delta(x) = 1.1803 \cdot \sqrt{\frac{x}{\sqrt{1-x^2}}}$
DTLZ7* [8]	$f(x) = 4 - x(1 + \sin(3\pi x))$ for all $x \in F$ where $F = [0, 0.2514118361] \cup]0.6316265307, 0.8594008566] \cup]1.3596178368, 1.5148392681] \cup]2.0518383519, 2.1164268079]$	$\delta(x) = 0.6566 \cdot \sqrt{1 + \sin(3\pi x) + 3\pi x \cos(3\pi x)}$

How to Use the Result in Performance Assess.

Problem	front description	density
bi-objective		
ZDT1, ZD		
ZDT2 [24]		
ZDT3* [24]		cos(10πx)
		1041] ∪
ZDT6 [24]	$f(x) = 1 - x$ for $x \in [\frac{\arctan(9\pi)}{6\pi}, 1] \approx [0.08146, 1]$	$\delta(x) = C \cdot \sqrt{x}$ with $C = \frac{3}{2}\left(1 - \frac{\arctan(9\pi)}{6\pi}^{3/2}\right)^{-1} \approx 1.53570$
DTLZ1 [8]		
DTLZ2, DT [8]		
DTLZ7* [8]		3πx)
		2681] ∪
	[2.0518383519, 2.1164268079]	



(a) ZDT1 and ZDT4 (b) ZDT2 (c) ZDT3

numerical optimization of $I_H(A,r,\mu)$ possible!

then: plot difference to optimal hypervolume

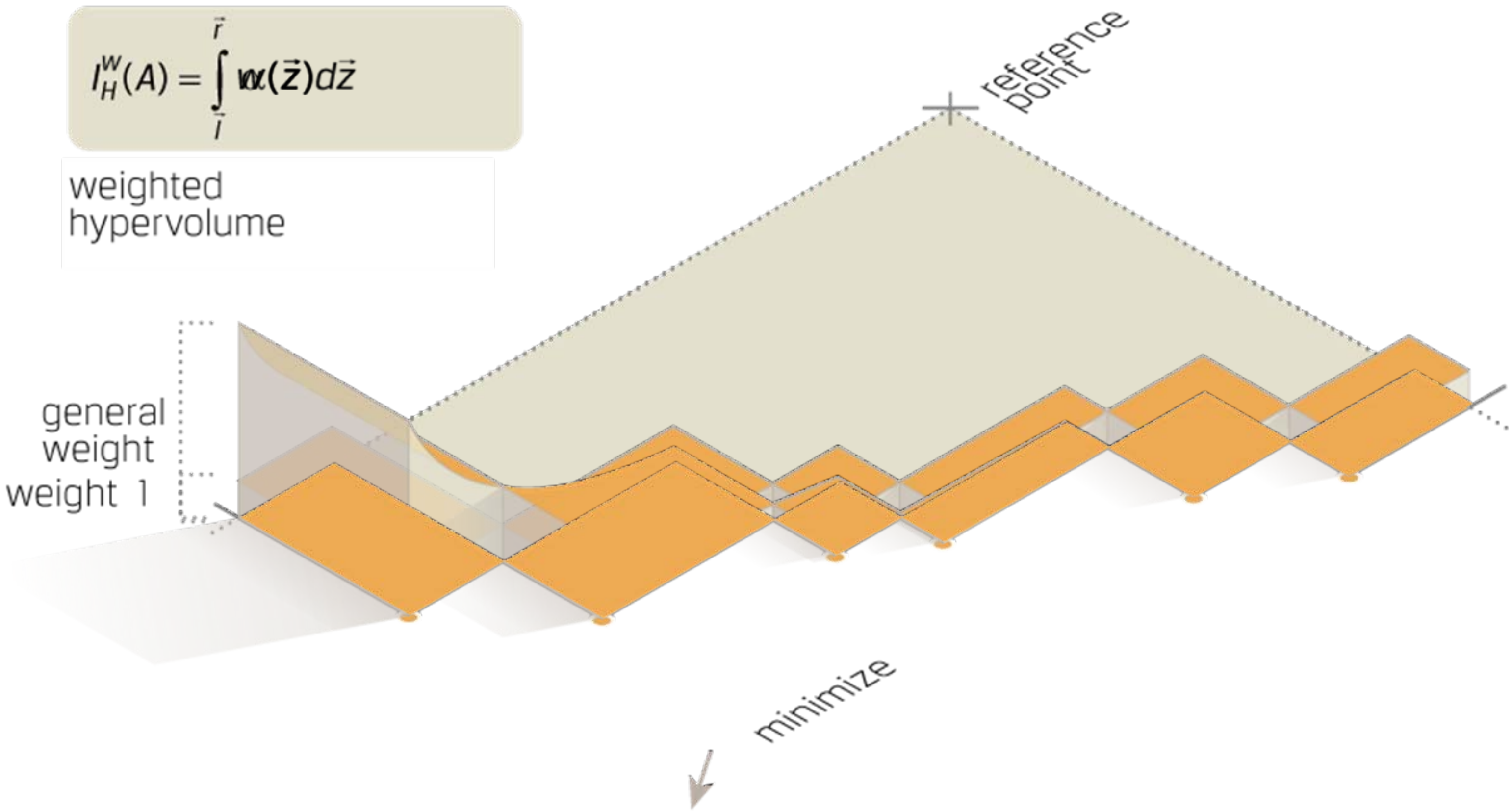
→ should decrease log-linear!

How to Change the Bias?

Goal: Incorporate user preferences into search (interactive optimization)

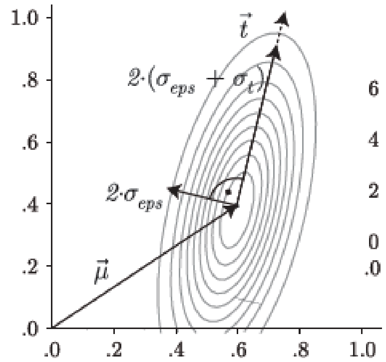
- (p)reference points, stressing extremes
- simulate classical scalarizing function approaches
- while keeping the refinement property

Idea: [zbt2007a]

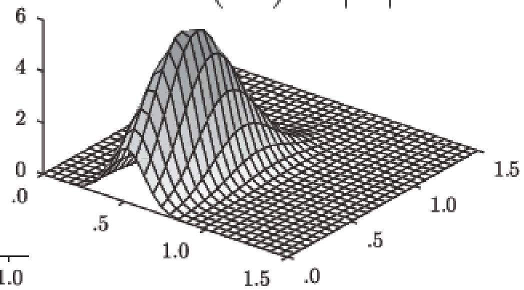


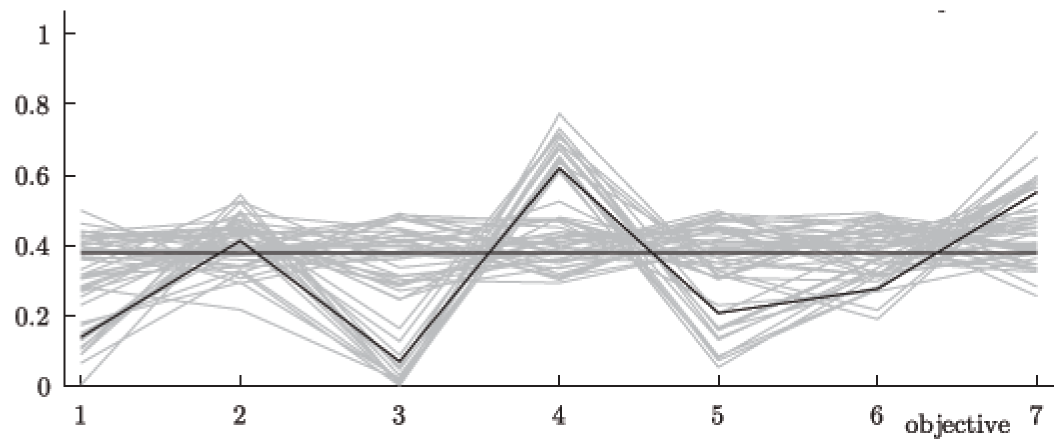
Examples of Weight Functions

preference point

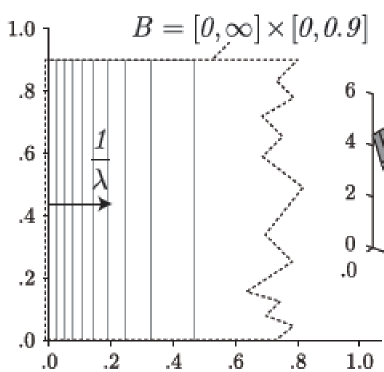


$$w(\vec{z}) = \frac{1}{(2\pi)^{k/2}|C|^{1/2}} e^{-\frac{1}{2}}$$

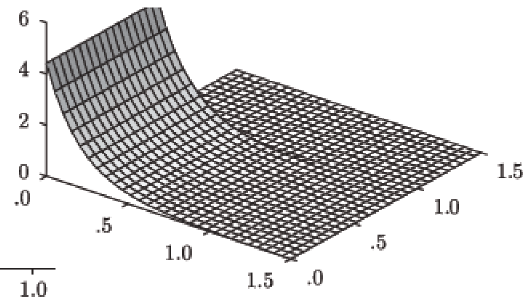


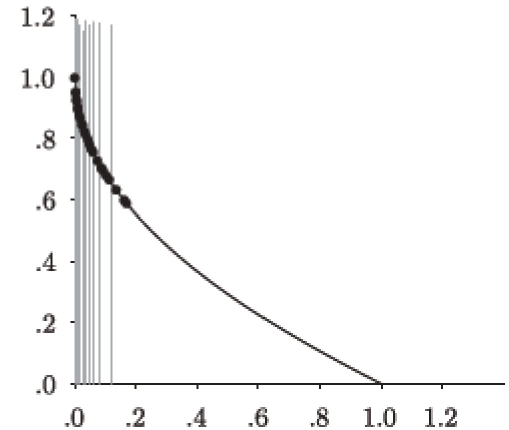


stressing one objective

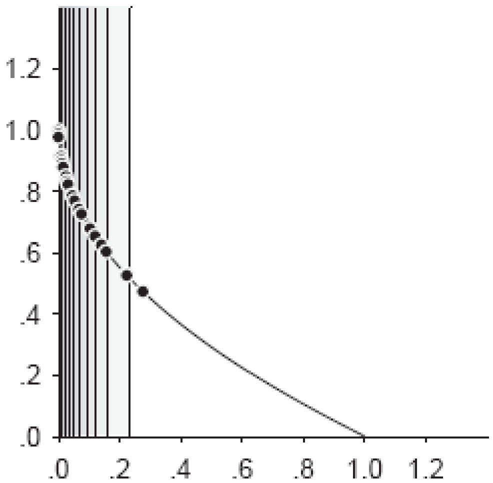


$$w(z_1,\cdots,z_k) = \begin{cases} \left(\prod_{i\neq s}(b_i^u - b_i^l)\right)^{-1} \lambda e^{-\lambda(z_s - b_s^l)} & \vec{z}\in B \\ 0 & \vec{z}\notin B \end{cases}$$

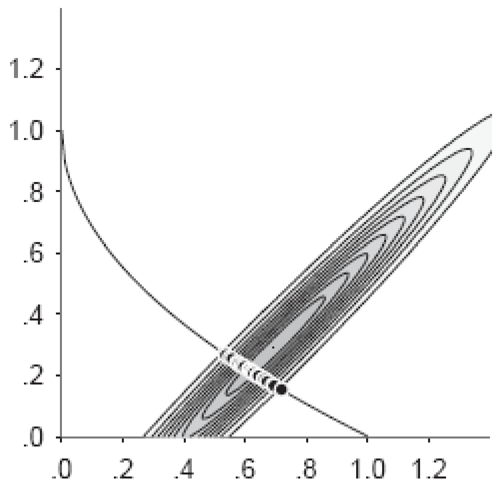




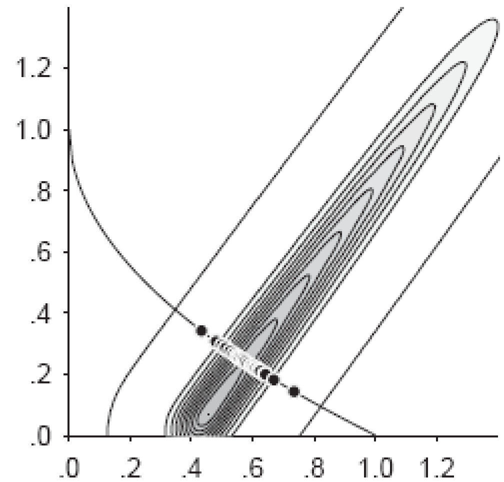
Results in 2D



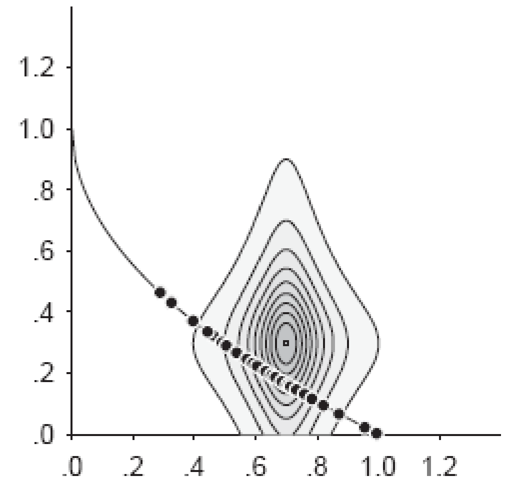
stress f_1



preference point

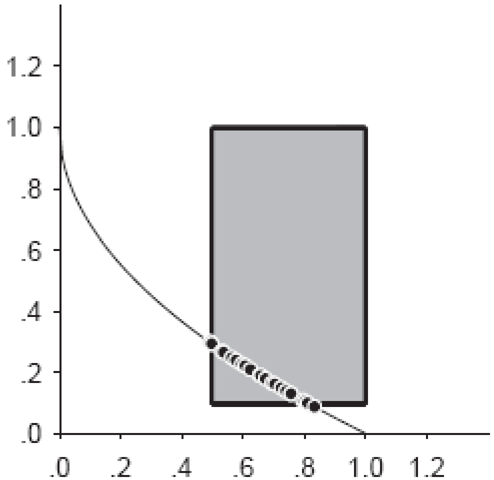


Tchebycheff

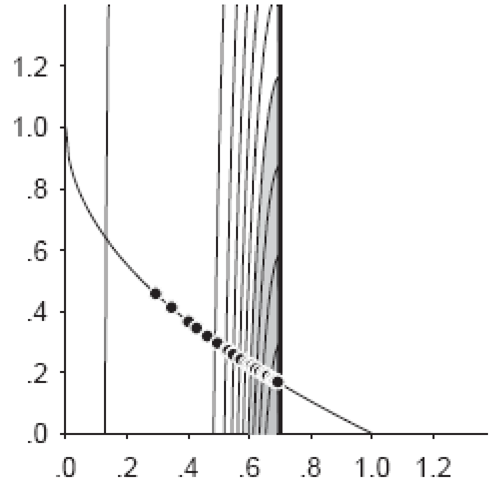


desirability function

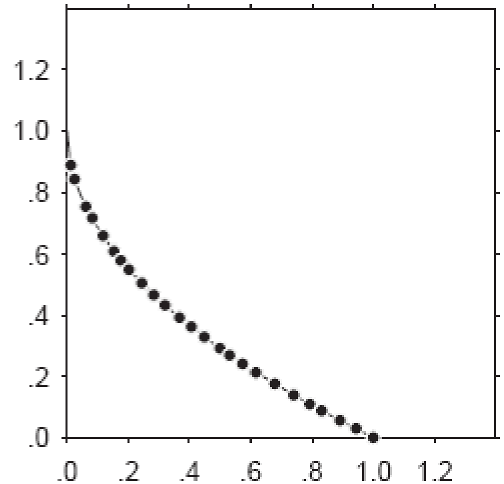
uniform



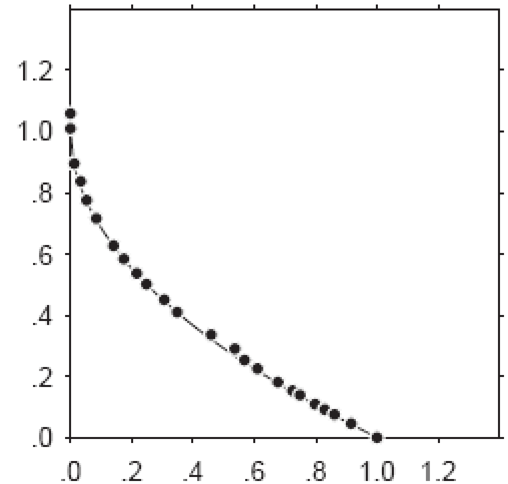
ε-constraint



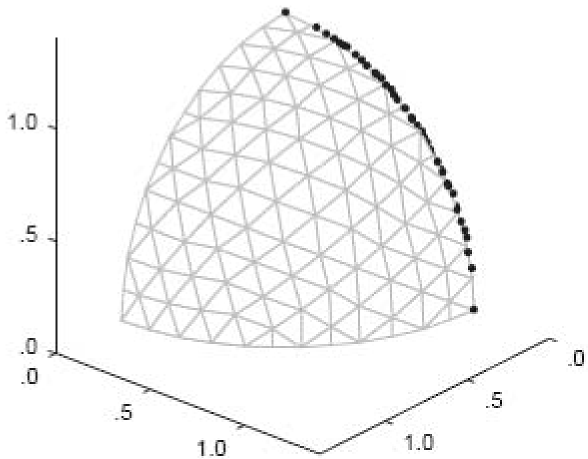
SPEA2



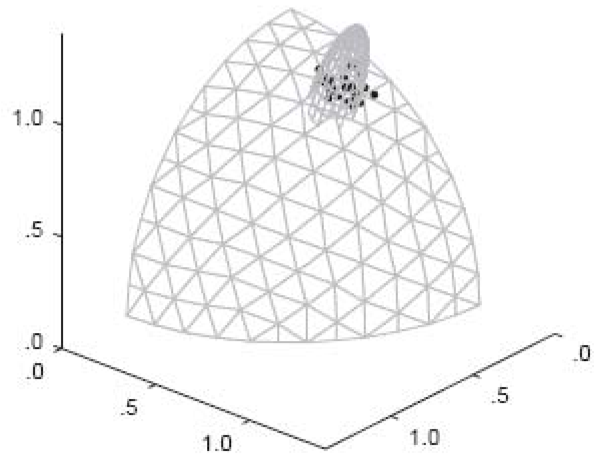
NSGA-II



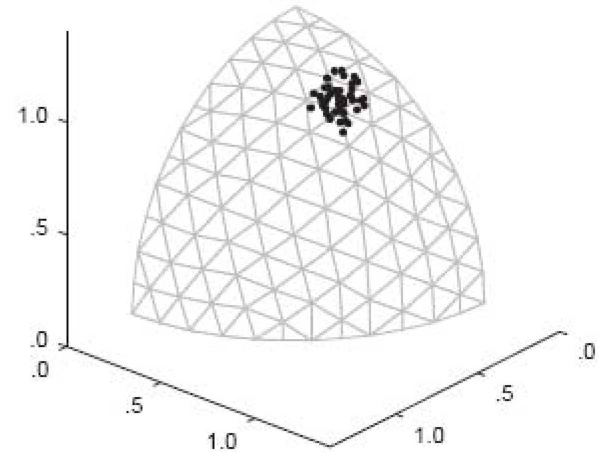
Results in 3D



stress f_1

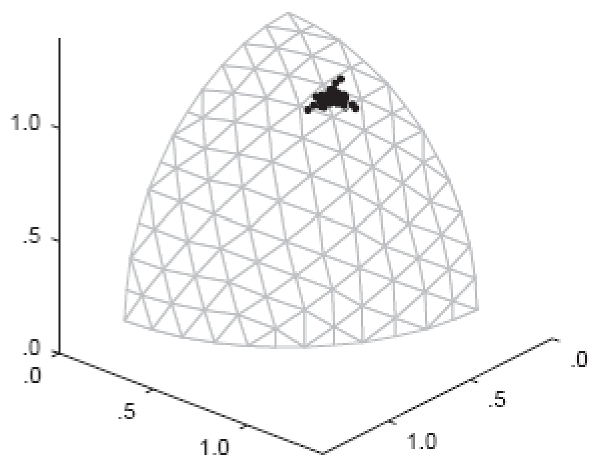


preference point

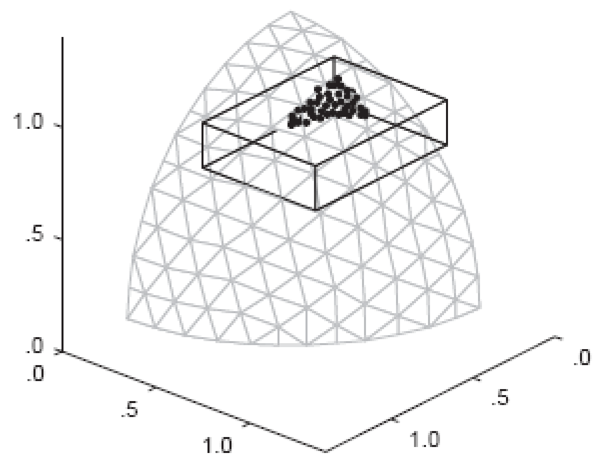


Tchebycheff

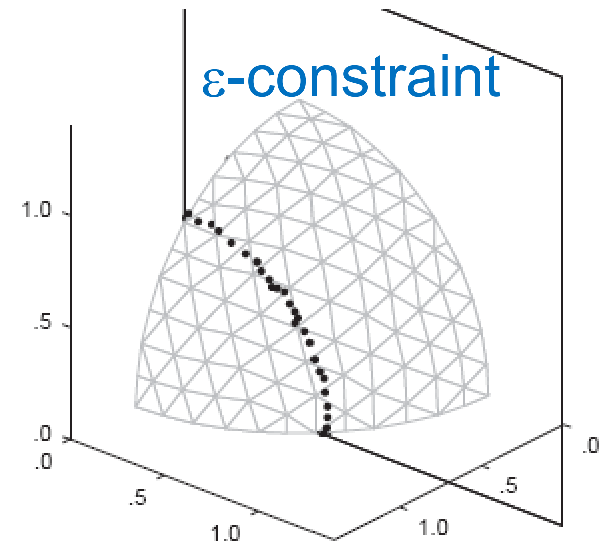
desirability function



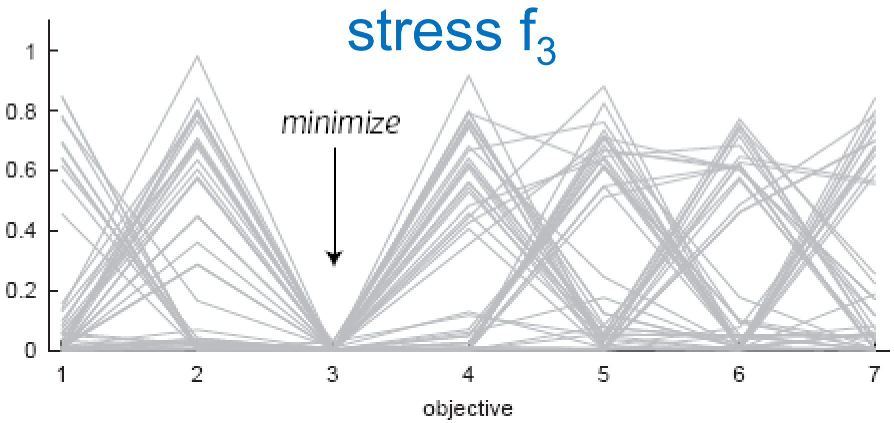
uniform

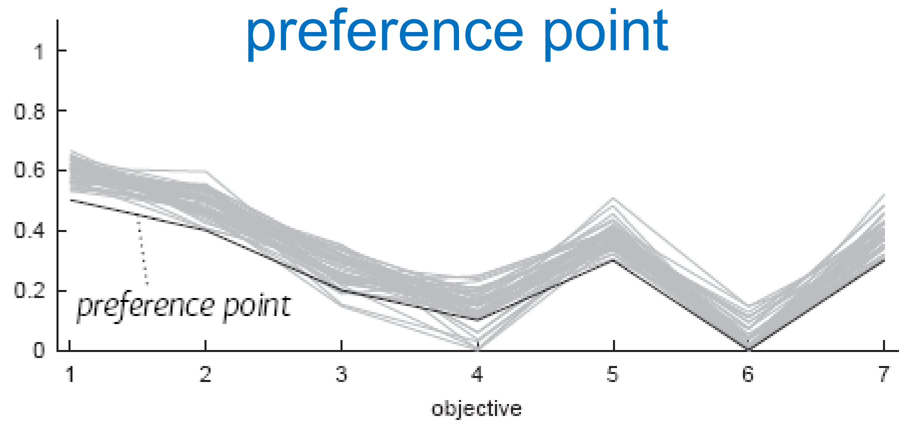


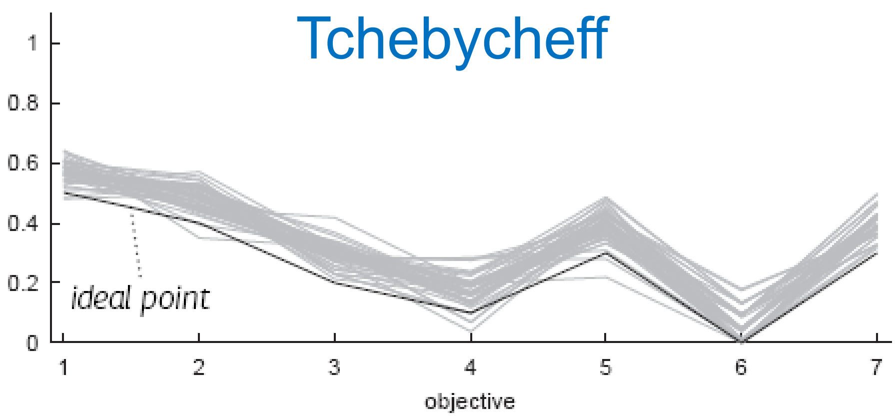
ε-constraint

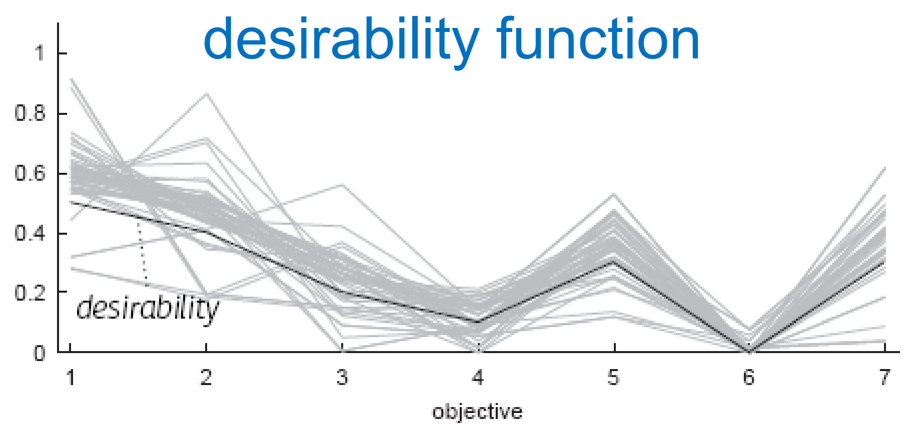


Results in 7D



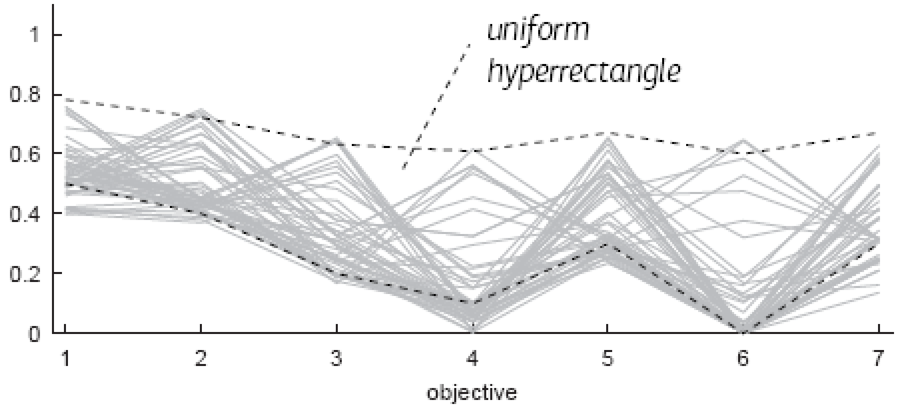




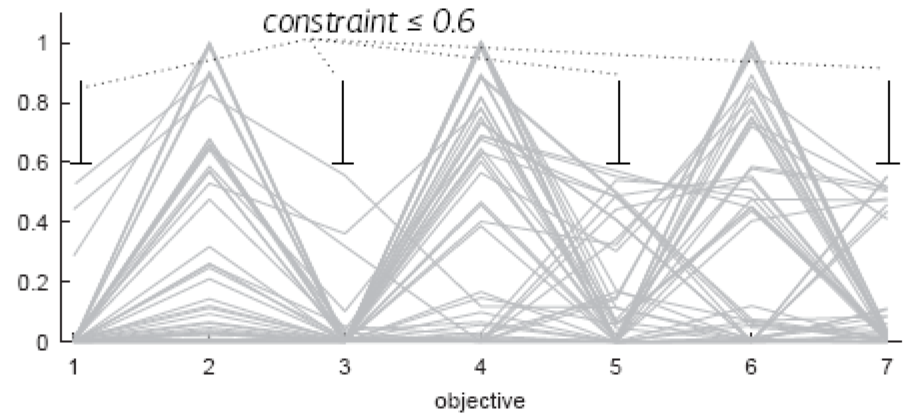


Results in 7D

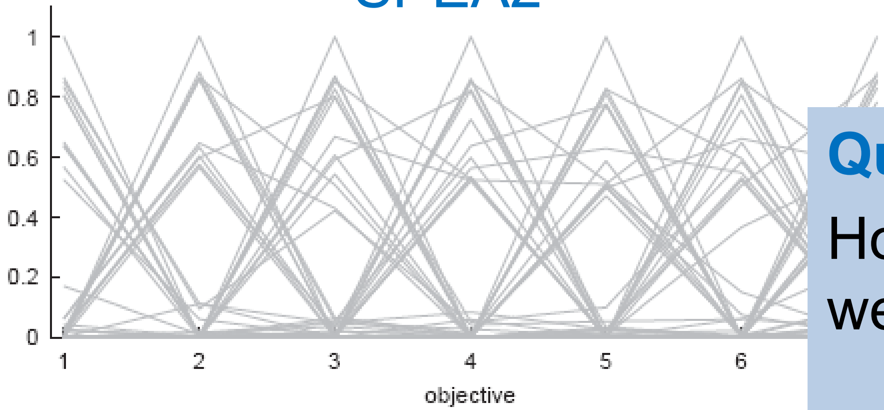
uniform



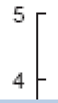
ε-constraint



SPEA2



NSGA-II



Question:

How do optimal μ-distributions for the weighted hypervolume indicator look like?

[abbz2009c,abbz2011a]

A New Idea of How to Articulate Preferences

Idea: [abbz2009c]

- compute theoretical result for weighted case

$$\delta(x) = \frac{\sqrt{-f'(x)w(x,f(x))}}{\int_0^{x_{max}} \sqrt{-f'(x)w(x,f(x))}dx}$$

- use „inverse“:
  - define a desired density
  - compute the corresponding weight
  - optimize with hypervolume-based algorithm

Problems:

- theoretical result for weight on front only
- front in practice not known
- efficient calculation of the hypervolume

A New Idea of How to Articulate Preferences

Idea:

- compute theoretical result for weighted case

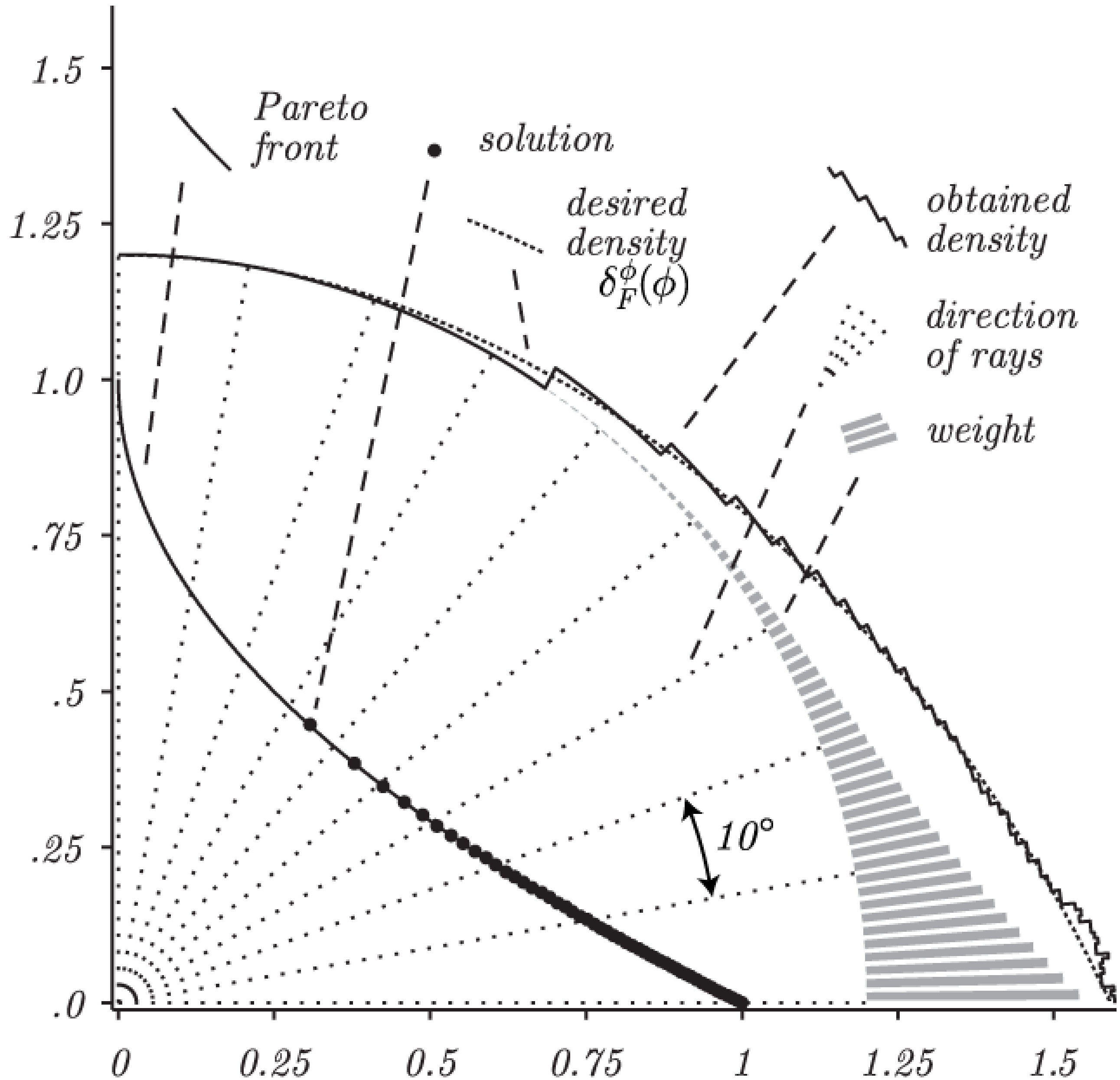
$$\delta(x) = \frac{\sqrt{-f'(x)w(x, f(x))}}{\int_0^{x_{max}} \sqrt{-f'(x)w(x, f(x))}dx}$$

- use „inverse“:
  - define a desired density
  - compute the corresponding weight
  - optimize with hypervolume-based algorithm

Problems:

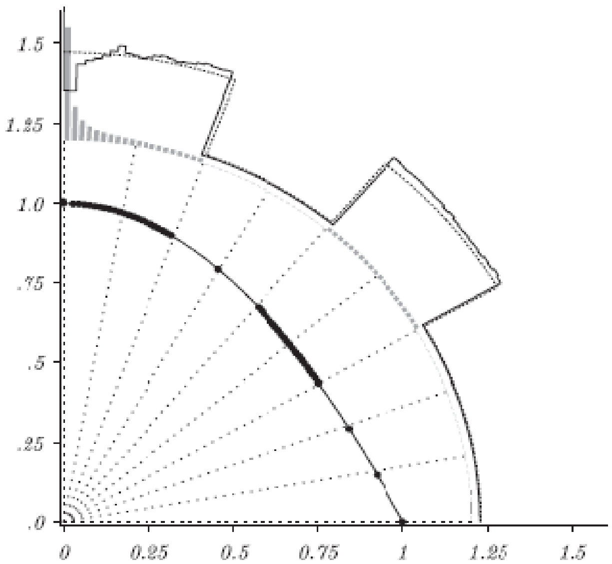
- theoretical result for weight on front only (extend with const. w)
- front in practice not known (assume expected front)
- efficient calculation of the hypervolume (dynamic programming)
- define density as function of angle ϕ instead of x

Results I

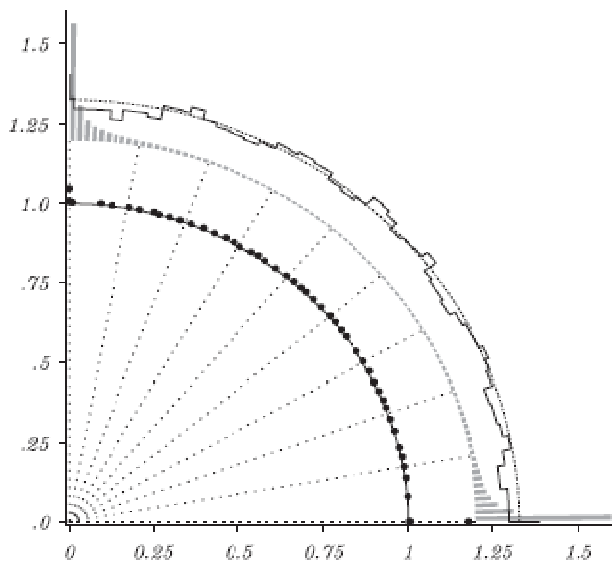


Pareto front
solution
desired density $\delta_F^\phi(\phi)$
obtained density
direction of rays
weight
10°
1.5
1.25
1.0
.75
.5
.25
.0
0
0.25
0.5
0.75
1
1.25
1.5

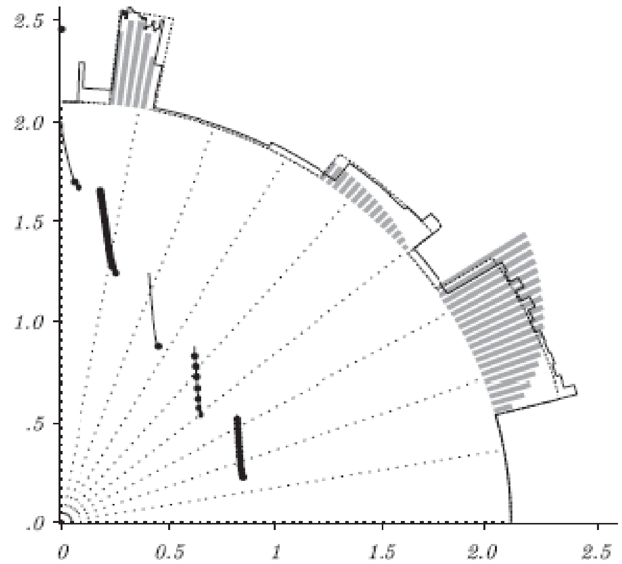
Results II



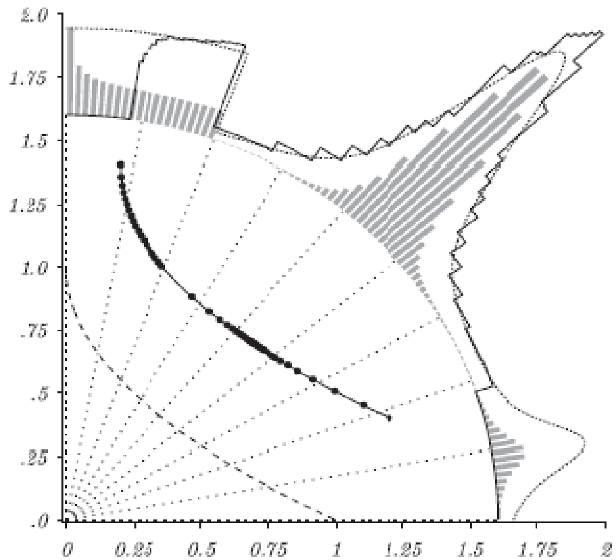
1.5
1.25
1.0
.75
.5
.25
.0
0
0.25
0.5
0.75
1
1.25
1.5



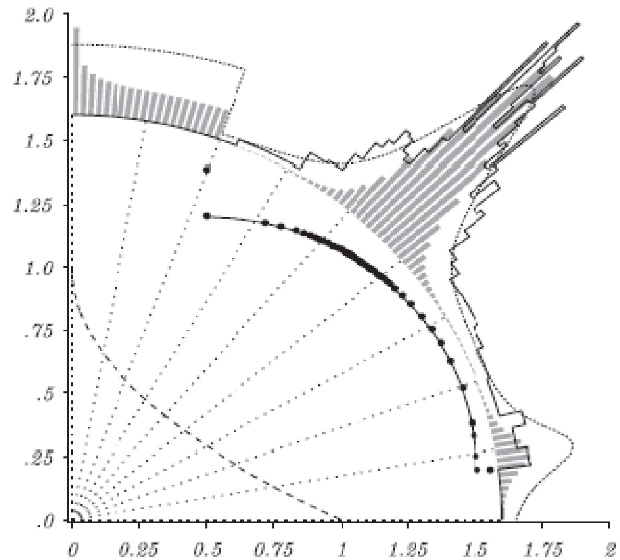
1.5
1.25
1.0
.75
.5
.25
.0
0
0.25
0.5
0.75
1
1.25
1.5



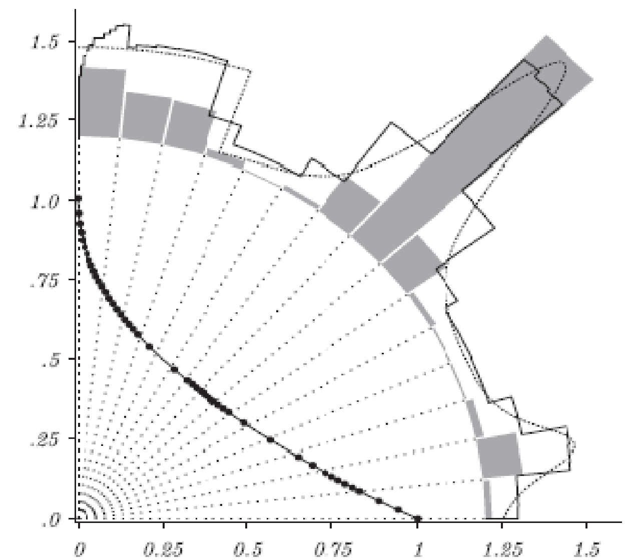
2.5
2.0
1.5
1.0
.5
.0
0
0.5
1
1.5
2.0
2.5



2.0
1.75
1.5
1.25
1.0
.75
.5
.25
.0
0
0.25
0.5
0.75
1
1.25
1.5
1.75
2



2.0
1.75
1.5
1.25
1.0
.75
.5
.25
.0
0
0.25
0.5
0.75
1
1.25
1.5
1.75
2



1.5
1.25
1.0
.75
.5
.25
.0
0
0.25
0.5
0.75
1
1.25
1.5

Hypervolume: Open Questions

Optimal μ-distributions

- uniqueness
- more objectives
- other indicators
- exact results
- faster algorithms to compute them
- convergence (greedy approach, HypE)
- linear convergence

Articulating User Preferences

- changing preferences over time
- simulating other classical approaches (from AI?)
- interactive

Overview

Objective Reduction

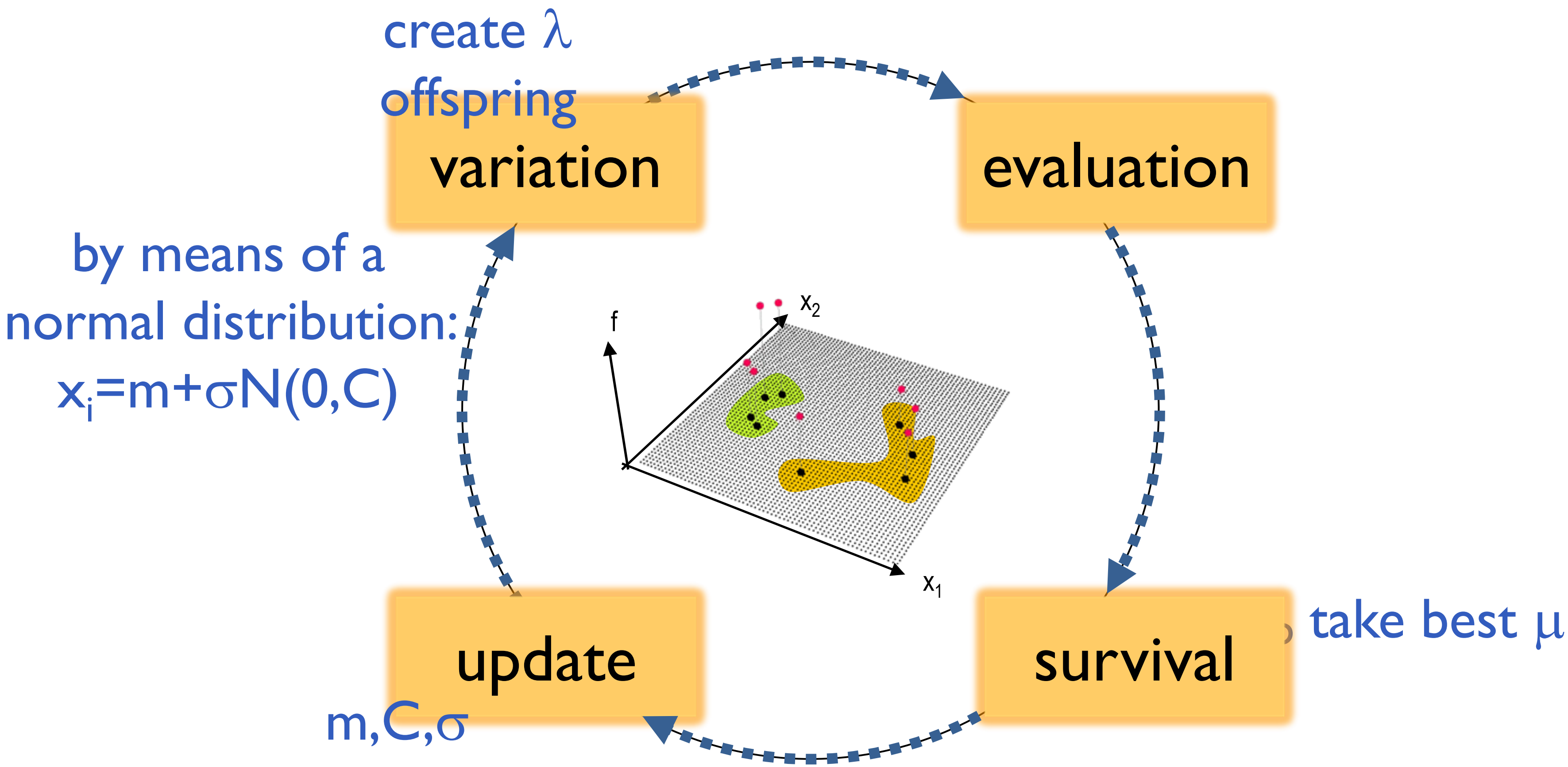
Hypervolume-based search

} multiobjective

Mirroring and Sequential Selection } single-objective

The "best" single-objective blackbox algorithm:

- Covariance Matrix Adaptation Evolution Strategy and variances
- continuous optimization



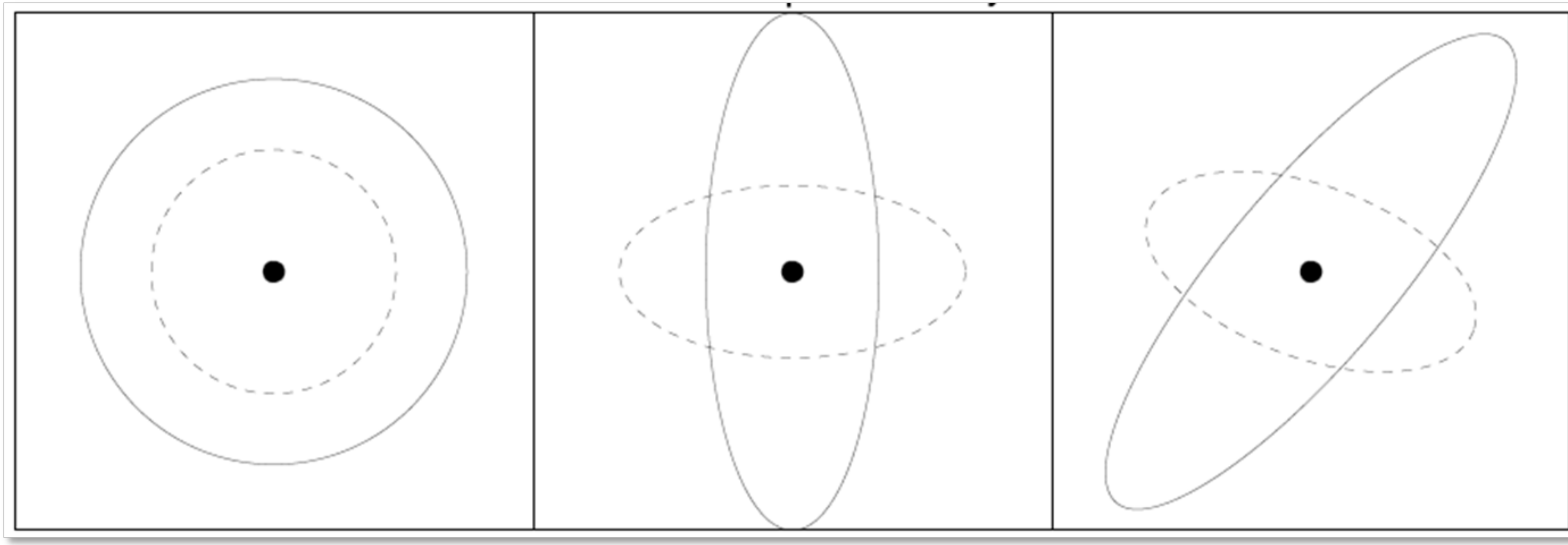
The CMA-ES: Equations

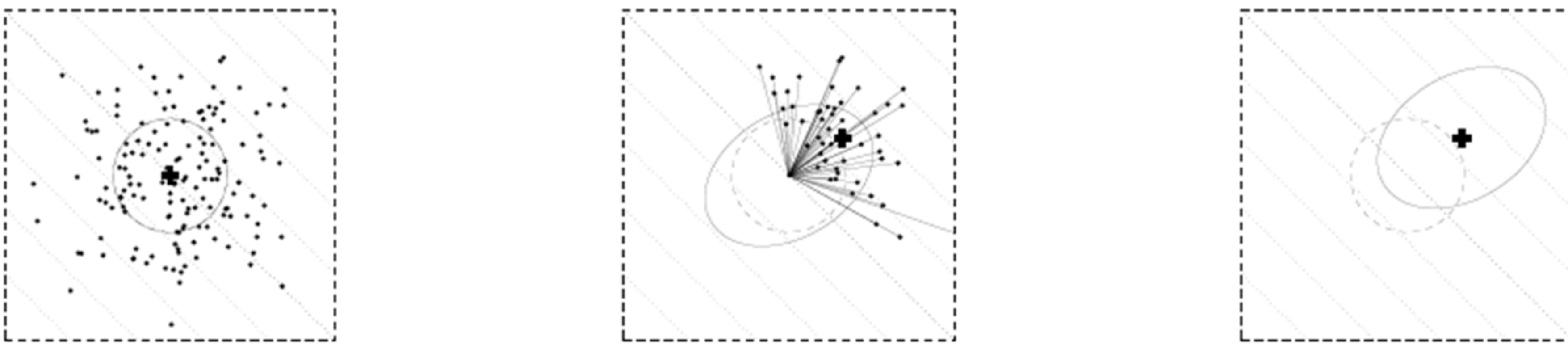
Input: $m \in \mathbb{R}^n$, $\sigma \in \mathbb{R}_+$, λ
Initialize: $\mathbf{C} = \mathbf{I}$, and $p_c = \mathbf{0}$, $p_\sigma = \mathbf{0}$,
Set: $c_c \approx 4/n$, $c_\sigma \approx 4/n$, $c_1 \approx 2/n^2$, $c_\mu \approx \mu_w/n^2$, $c_1 + c_\mu \leq 1$, $d_\sigma \approx 1 + \sqrt{\frac{\mu_w}{n}}$,
and $w_{i=1\ldots\lambda}$ such that $\mu_w = \frac{1}{\sum_{i=1}^{\mu} {w_i}^2} \approx 0.3\,\lambda$
While not terminate

$x_i = m + \sigma\, y_i, \quad y_i \sim \mathcal{N}_i(\mathbf{0}, \mathbf{C})\,, \quad$ for $i = 1, \ldots, \lambda$	sampling
$m \leftarrow \sum_{i=1}^{\mu} w_i\, x_{i:\lambda} = m + \sigma\, y_w \quad$ where $y_w = \sum_{i=1}^{\mu} w_i\, y_{i:\lambda}$	update mean
$p_c \leftarrow (1 - c_c)\, p_c + \mathbb{1}_{\{\lVert p_\sigma \rVert < 1.5\sqrt{n}\}} \sqrt{1 - (1 - c_c)^2}\, \sqrt{\mu_w}\, y_w$	cumulation for $\mathbf{C}$
$p_\sigma \leftarrow (1 - c_\sigma)\, p_\sigma + \sqrt{1 - (1 - c_\sigma)^2}\, \sqrt{\mu_w}\, \mathbf{C}^{-\frac{1}{2}} y_w$	cumulation for σ
$\mathbf{C} \leftarrow (1 - c_1 - c_\mu)\, \mathbf{C} + c_1\, p_c {p_c}^{\mathrm{T}} + c_\mu \sum_{i=1}^{\mu} w_i\, y_{i:\lambda} y_{i:\lambda}^{\mathrm{T}}$	update $\mathbf{C}$
$\sigma \leftarrow \sigma \times \exp\left(\frac{c_\sigma}{d_\sigma}\left(\frac{\lVert p_\sigma \rVert}{\mathsf{E}\lVert \mathcal{N}(\mathbf{0}, \mathbf{I}) \rVert} - 1\right)\right)$	update of σ

Not covered on this slide: termination, restarts, useful output, boundaries and encoding

The CMA-ES: Ideas





$x_i = m + \sigma y_i, \quad y_i \sim \mathcal{N}(\mathbf{0}, C)$

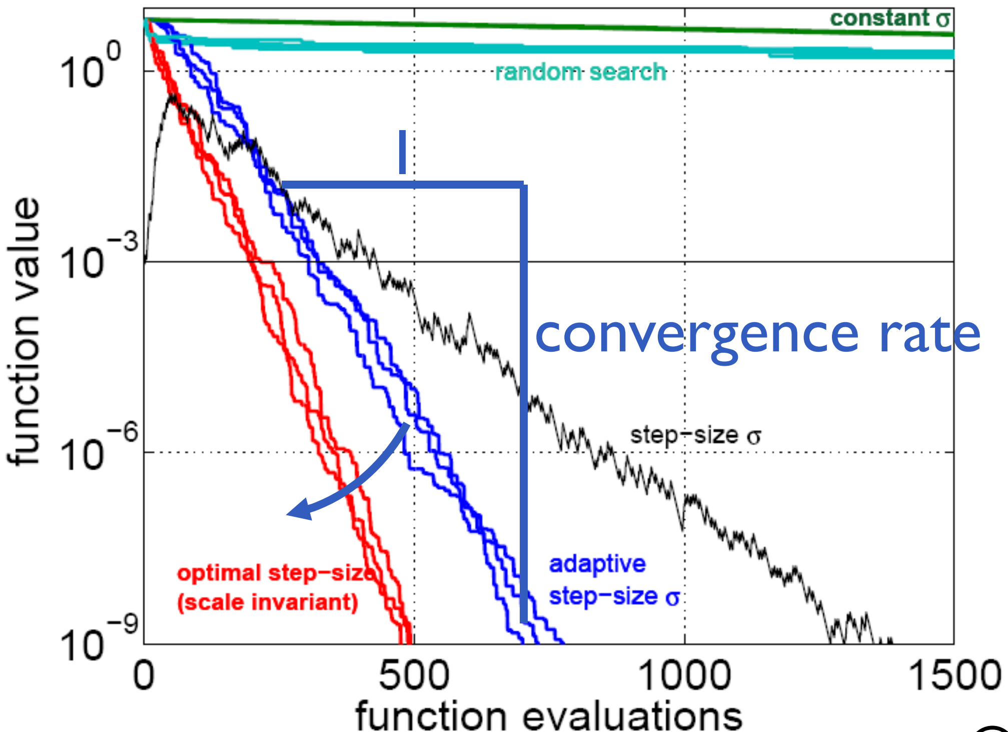
$C_\mu = \frac{1}{\mu} \sum y_{i:\lambda} y_{i:\lambda}^{T}$

$C \leftarrow (1-1) \times C + 1 \times C_\mu$

$m_{\text{new}} \leftarrow m + \frac{1}{\mu} \sum y_{i:\lambda}$

Two Independent Ideas to Make Local (CMA-)ES Faster

- local: only a few children (λ small) [baha2010a, abh2011a]
- derandomized mutations
- stopping generations whenever better than parent



$$f(\boldsymbol{x}) = \sum_{i=1}^{n} x_i^2$$

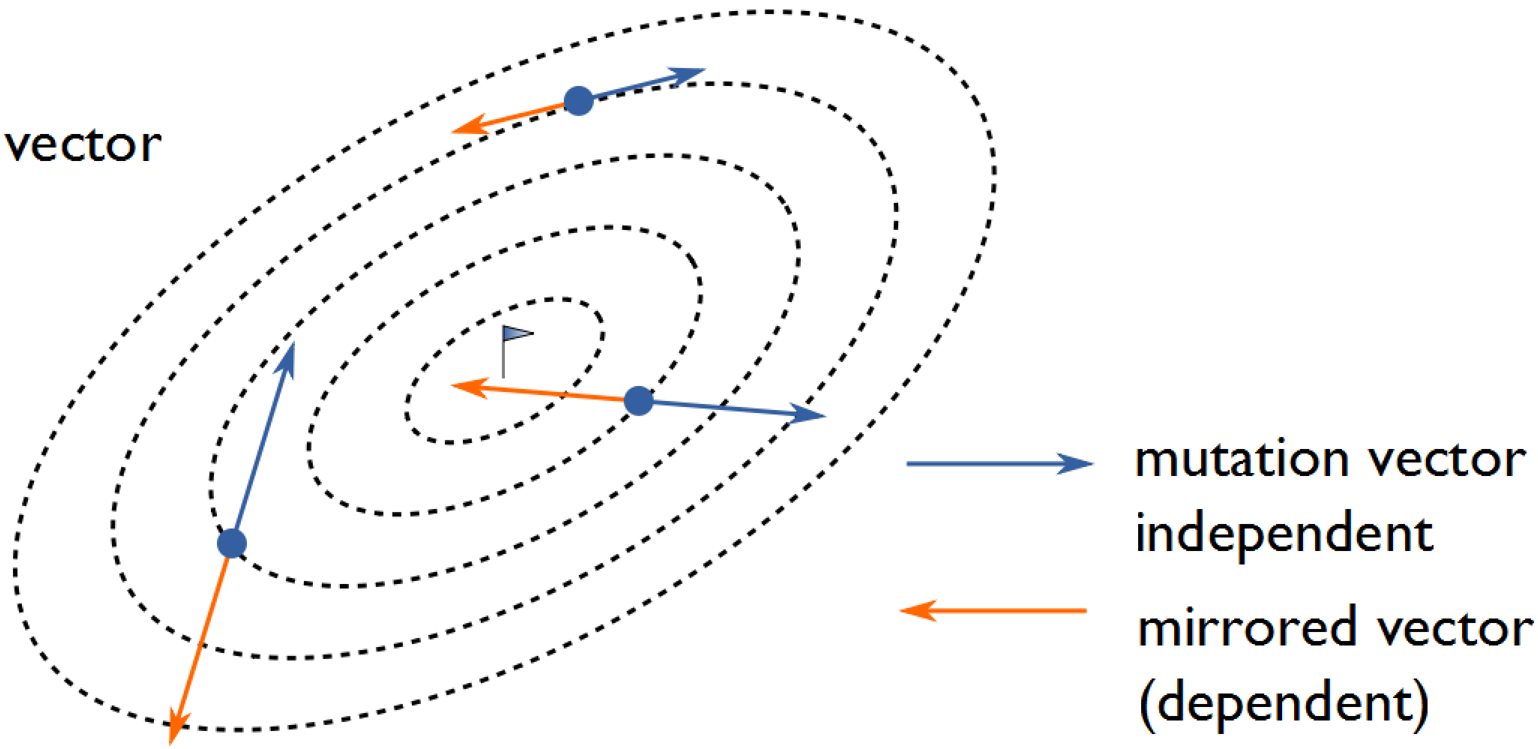
in $[-0.2, 0.8]^n$
for $n = 10$

© N. Hansen & A. Auger

Mirrored Mutations

Idea

use one random vector
to generate two
offspring



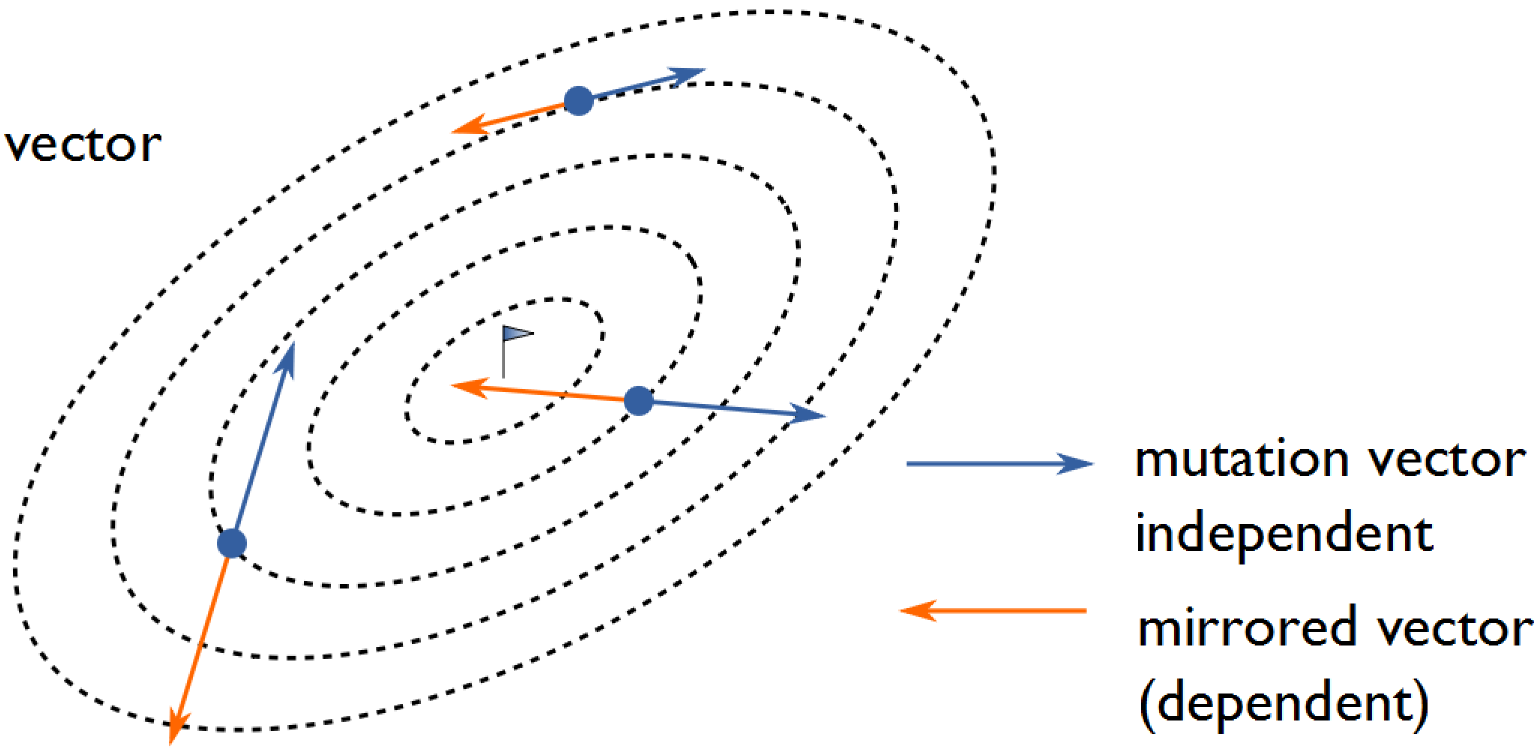
Mirrored Mutations

Idea

use one random vector to generate two offspring

Reasoning

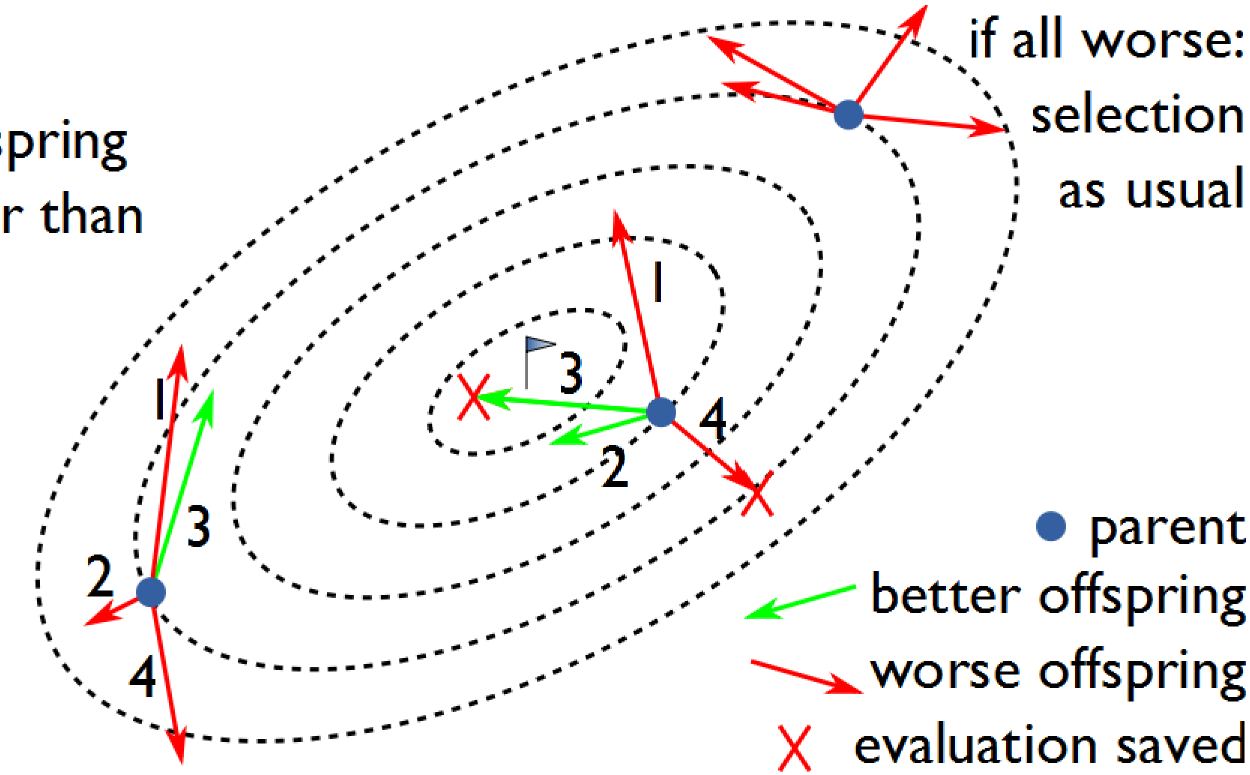
often "good" and "bad" in opposite directions



Sequential Selection

Idea

stop generation of new offspring as soon as a solution, better than the parent, is found



Sequential Selection

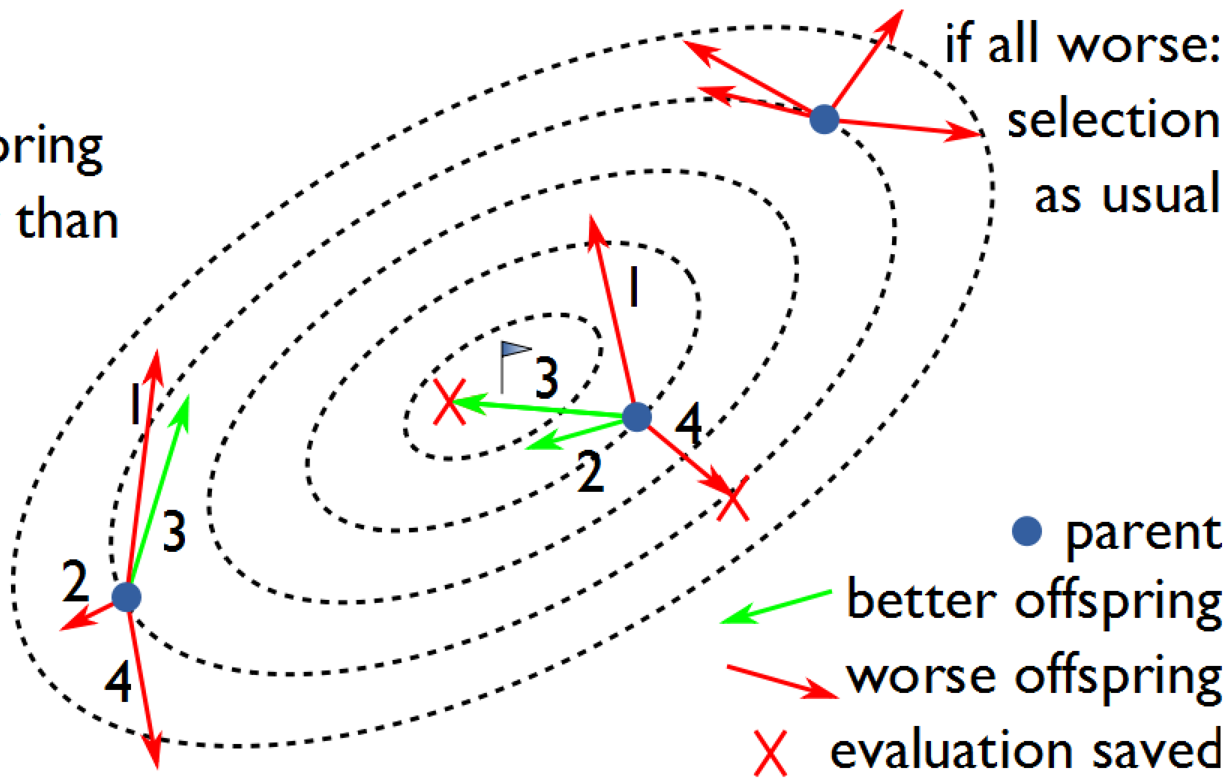
Idea

stop generation of new offspring as soon as a solution, better than the parent, is found

Reasoning

if sublevel sets convex one better is enough

in particular with mirroring



Theoretical Results on Convergence Rates

- for several variants of scale-invariant ($\sigma_t = \sigma|X_t|$) Evolution Strategies

Theorem 4. *For a* $(1, 2^{\mathrm{s}}_{\mathrm{m}})$*-ES with scale-invariant step-size* ($\sigma_k = \sigma\|\boldsymbol{X}_k\| > 0$) *on the sphere function* $g(\|\boldsymbol{x}\|)$*, for* $g \in \mathcal{M}$*, linear convergence holds and*

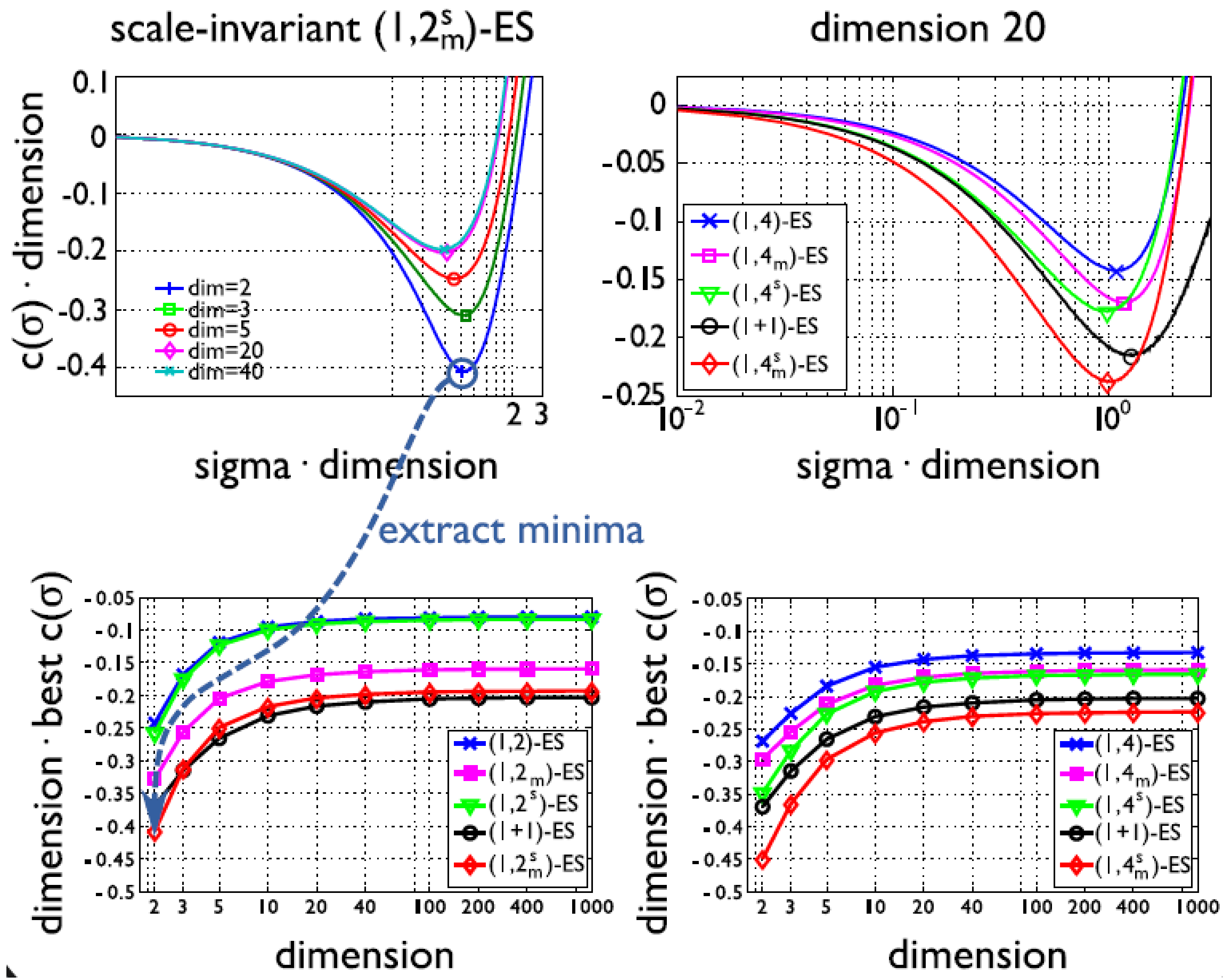
$$\frac{1}{T_k} \ln \frac{\|\boldsymbol{X}_k\|}{\|\boldsymbol{X}_0\|} \xrightarrow[k \to \infty]{} \frac{1}{2} \frac{1}{2 - p_s(\sigma)} \times E\left[\ln\left(1 - 2\sigma|[\boldsymbol{\mathcal{N}}]_1| + \sigma^2\|\boldsymbol{\mathcal{N}}\|^2\right)\right] \; a.s.$$

where T_k *is the random variable for the number of function evaluations until iteration* k*,* $\boldsymbol{\mathcal{N}}$ *is a random vector following a multivariate normal distribution, and* $p_s(\sigma) = \Pr(2[\boldsymbol{\mathcal{N}}]_1 + \sigma\|\boldsymbol{\mathcal{N}}\|^2 < 0)$ *is the probability that the first offspring is successful.*

exemplary

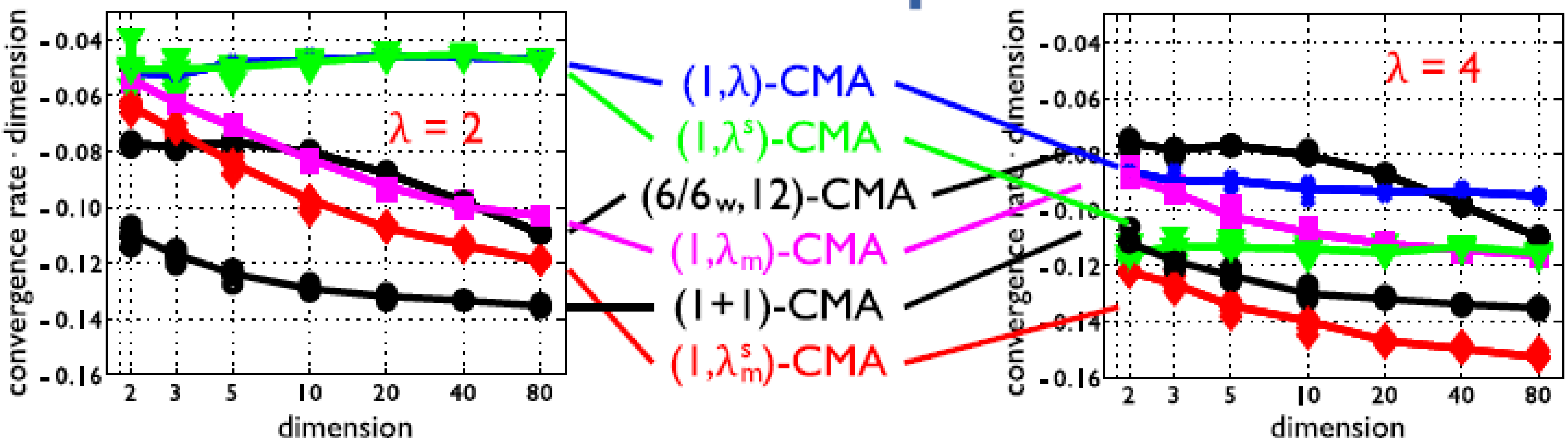
- can be estimated via Monte Carlo Sampling

Results: Estimated Convergence Rates



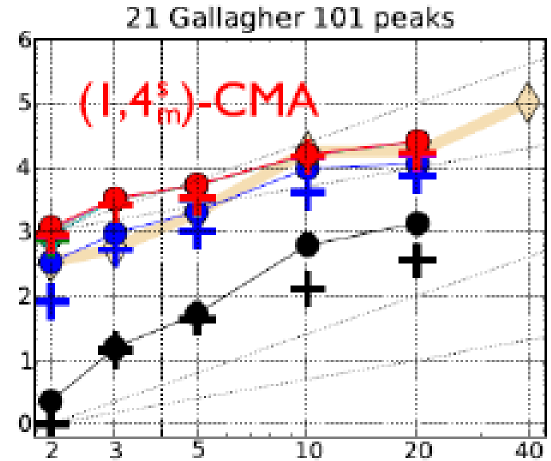
Implementation in CMA-ES

results on sphere



results on BBOB'2010

- $(1,4^s_m)$-CMA-ES turned out to be fastest
    local non-elitist strategy tested
- 3rd best of BBOB'2009/10 on Gallagher with
    101 peaks (3x faster than (1+1)-CMA-ES)
- even more competitive on noisy functions



Mirroring and Sequ. Selection: Open Questions

- how to implement in μ/μ_w-CMA-ES without bias in step-size?
- further mirroring (more dependencies)
- does it make sense in multiobjective CMA-ES?
- …

Conclusions

❶ Objective Reduction

- idea, algorithms
- omission and aggregation of objectives

❷ Hypervolume-based Search

- weighted hypervolume indicator
- optimal μ-distributions
- a new way to articulate user preferences in 2D

❸ Mirroring and Sequential Selection

- idea, results
- log-linear convergence

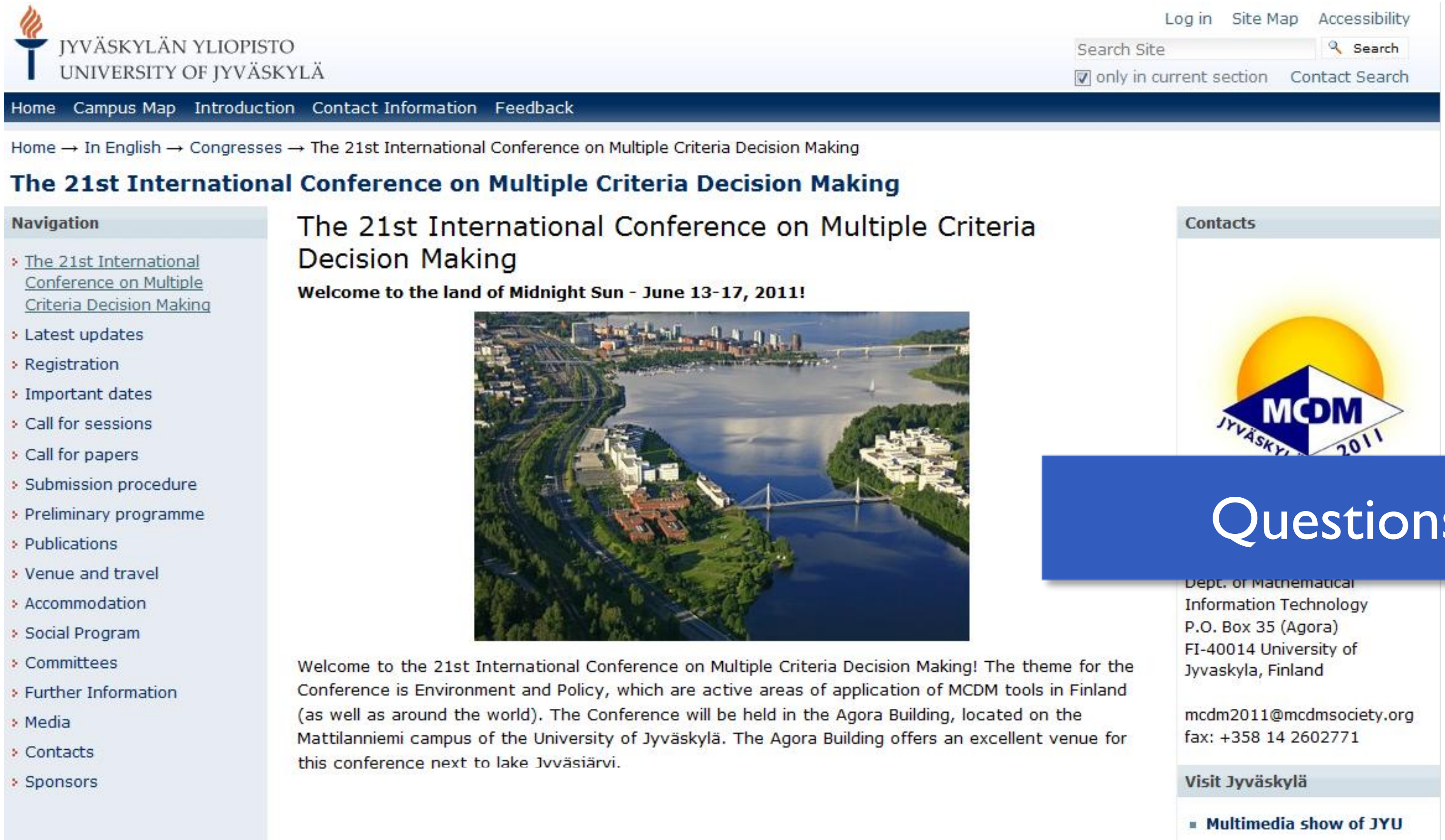
Announcement

EMO session @ MCDM'2011 in Jyväskylä, Finland

organizers: Dimo Brockhoff and Kalyanmoy Deb

tentative deadline: **Jan. 31, 2011** (full papers & abstracts)

http://emoatmcdm.gforge.inria.fr



Questions?

References I

[abb2010a] A. Auger, J. Bader, and D. Brockhoff. Theoretically Investigating Optimal μ-Distributions for the Hypervolume Indicator: First Results For Three Objectives. In R. Schaefer et al., editors, Conference on Parallel Problem Solving from Nature (PPSN XI), volume 6238 of LNCS, pages 586–596. Springer, 2010

[abbz2009a] A. Auger, J. Bader, D. Brockhoff, and E. Zitzler. Theory of the Hypervolume Indicator: Optimal μ-Distributions and the Choice of the Reference Point. In Foundations of Genetic Algorithms (FOGA 2009), pages 87–102, New York, NY, USA, 2009. ACM

[abbz2009c] A. Auger, J. Bader, D. Brockhoff, and E. Zitzler. Investigating and Exploiting the Bias of the Weighted Hypervolume to Articulate User Preferences. In G. Raidl et al., editors, Genetic and Evolutionary Computation Conference (GECCO 2009), pages 563–570, New York, NY, USA, 2009. ACM

[abbz2011a] A. Auger, J. Bader, D. Brockhoff, and E. Zitzler. Hypervolume-based Multiobjective Optimization: Theoretical Foundations and Practical Implications. Theoretical Computer Science, 2010. to appear

[abh2011a] A. Auger, D. Brockhoff, and N. Hansen. Analyzing the Impact of Mirrored Sampling and Sequential Selection in Elitist Evolution Strategies. In Foundations of Genetic Algorithms (FOGA 2011). ACM, 2011. to appear

[agre1997a] P. J. Agrell. On Redundancy in Multi Criteria Decision Making. European Journal of Operational Research, 98(3):571–586, 1997

[baha2010a] D. Brockhoff, A. Auger, N. Hansen, D. V. Arnold, and T. Hohm. Mirrored Sampling and Sequential Selection for Evolution Strategies. In R. Schaefer et al., editors, Conference on Parallel Problem Solving from Nature (PPSN XI), volume 6238 of LNCS, pages 11–21. Springer, 2010

[bne2007a] N. Beume, B. Naujoks, and M. Emmerich. SMS-EMOA: Multiobjective Selection Based on Dominated Hypervolume. European Journal of Operational Research, 181(3):1653–1669, 2007

[bz2006d] D. Brockhoff and E. Zitzler. Are All Objectives Necessary? On Dimensionality Reduction in Evolutionary Multiobjective Optimization. In T. P. Runarsson et al., editors, Conference on Parallel Problem Solving from Nature (PPSN IX), volume 4193 of LNCS, pages 533–542, Berlin, Germany, 2006. Springer

[bz2007d] D. Brockhoff and E. Zitzler. Dimensionality Reduction in Multiobjective Optimization: The Minimum Objective Subset Problem. In K. H. Waldmann and U. M. Stocker, editors, Operations Research Proceedings 2006, pages 423–429. Springer, 2007

[bz2010a] D. Brockhoff and E. Zitzler. Automated Aggregation and Omission of Objectives to Handle Many-Objective Problems. In Conference on Multiple Objective and Goal Programming (MOPGP 2008), Lecture Notes in Economics and Mathematical Systems, pages 81–102. Springer, 2010

[dapm2002a] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation, 6(2):182–197, 2002

[dmm2005a] K. Deb, M. Mohan, and S. Mishra. Evaluating the ε-Domination Based Multi-Objective Evolutionary Algorithm for a Quick Computation of Pareto-Optimal Solutions. Evolutionary Computation, 13(4):501–525, Winter 2005

[ds2006a] K. Deb and D. K. Saxena. Searching For Pareto-Optimal Solutions Through Dimensionality Reduction for Certain Large-Dimensional Multi-Objective Optimization Problems. In Congress on Evolutionary Computation (CEC 2006), pages 3352–3360. IEEE Press, 2006

[ebn2005a] M. Emmerich, N. Beume, and B. Naujoks. An EMO Algorithm Using the Hypervolume Measure as Selection Criterion. In Conference on Evolutionary Multi-Criterion Optimization (EMO 2005), volume 3410 of LNCS, pages 62–76. Springer, 2005

[feig1998a] U. Feige. A Threshold of ln n for Approximating Set Cover. J. ACM, 45(4):634–652, 1998

References II

[flei2003a] M. Fleischer. The Measure of Pareto Optima. Applications to Multi-Objective Metaheuristics. In C. M. Fonseca et al., editors, Conference on Evolutionary Multi-Criterion Optimization (EMO 2003), volume 2632 of LNCS, pages 519–533, Faro, Portugal, 2003. Springer

[gl1977a] T. Gal and H. Leberling. Redundant Objective Functions in Linear Vector Maximum Problems and Their Determination. European Journal of Operational Research, 1(3):176–184, 1977

[ho1996a] N. Hansen and A. Ostermeier. Adapting arbitrary normal mutation distributions in evolution strategies: the covariance matrix adaptation. In Congress on Evolutionary Computation (CEC 1996), pages 312–317, Piscataway, NJ, USA, 1996. IEEE

[ho2001a] N. Hansen and A. Ostermeier. Completely Derandomized Self-Adaptation in Evolution Strategies. Evolutionary Computation, 9(2):159–195, 2001

[mali2006a] A. B. Malinowska. Nonessential Objective Functions in Linear Multiobjective Optimization Problems. Control and Cybernetics, 35(4):873–880, 2006

[mali2008a] A. B. Malinowska. Weakly and Properly Nonessential Objectives in Multiobjective Optimization Problems. Operations Research Letters, 36:647–650, 2008

[mt2007a] A. B. Malinowska and D. F. M. Torres. Nonessential Functionals in Multiobjective Optimal Control Problems. Proceedings of the Estonian Academy of Sciences: Physics and Mathematics, 56(4):336–346, 2007

[mt2008a] A. B. Malinowska and D. F. M. Torres. Computational Approach to Essential and Nonessential Objective Functions in Linear Multicriteria Optimization. Journal of Optimization Theory and Applications, 139(2):577–590, 2008

[sd2007a] D. K. Saxena and K. Deb. Non-linear Dimensionality Reduction Procedures for Certain Large-Dimensional Multi-objective Optimization Problems: Employing Correntropy and a Novel Maximum Variance Unfolding. In Conference on Evolutionary Multi-Criterion Optimization (EMO 2007), volume 4403 of LNCS, pages 772–787. Springer, 2007

[sd2008b] D. K. Saxena and K. Deb. Dimensionality Reduction of Objectives and Constraints in Multi-objective Optimization Problems: A System Design Perspective. In Congress on Evolutionary Computation (CEC 2008), pages 3204–3211. IEEE Press, 2008

[zbt2007a] E. Zitzler, D. Brockhoff, and L. Thiele. The Hypervolume Indicator Revisited: On the Design of Pareto-compliant Indicators Via Weighted Integration. In S. Obayashi et al., editors, Conference on Evolutionary Multi-Criterion Optimization (EMO 2007), volume 4403 of LNCS, pages 862–876, Berlin, 2007. Springer

[zt1998b] E. Zitzler and L. Thiele. Multiobjective Optimization Using Evolutionary Algorithms - A Comparative Case Study. In Conference on Parallel Problem Solving from Nature (PPSN V), volume 1498 of LNCS, pages 292–301, Amsterdam, 1998

[ztb2010a] E. Zitzler, L. Thiele, and J. Bader. On Set-Based Multiobjective Optimization. IEEE Transactions on Evolutionary Computation, 14(1):58–79, 2010