Theoretical Issues of Evolutionary Multiobjective Optimization: Selected Research Topics and Open Problems

Dimo Brockhoff

September 16, 2011, 5th SPO Symposium, TU Dortmund
study of CS (Dipl. inform.) in Dortmund, Germany

Dr. sc. ETH at ETH Zurich, Switzerland

postdoc at INRIA Saclay – Ile-de-France

postdoc at Ecole Polytechnique

chargé de recherche (CR2) INRIA Lille Nord-Europe
Contributions in EMO

Theory

- Many-Objective Optimization and Objective Reduction
  - ECJ '09
  - chapter '07
  - PPSN '06
  - IEEE-CEC '07
  - MOPGP '08
  - OR '06

- Hypervolume-Based Search
  - PhD thesis '09

- Hypervolume Sampling
  - ACM-GECCO '09a

- Algorithms
  - Weighted Hypervolume
    - EMO '07

- Applications
  - Hazmat Routing
    - CTW '11
  - Wireless Sensor Networks
    - MCDM '08
  - Radar Waveforms

- Set-Based EAs
  - EMO '09

- Algorithms
  - Hypervolume-Based Search
  - Optimal μ-distributions
    - ACM-FOGA '09
    - ACM-GECCO '09b
    - PPSN '10b
    - TCS '11
    - SEAL '10

- Runtime analyses
  - IEEE-TEC '09
  - ACM-GECCO '07
  - PPSN '08

© Dimo Brockhoff, LIX, Ecole Polytechnique

“Theoretical Issues of EMO” @ TU Dortmund, September 16, 2011
Most problems are multiobjective in nature...

$\min_{x \in X} f(x) = (f_1(x), \ldots, f_k(x)) \in \mathbb{R}^k$
Blackbox Optimization

Most problems are multiobjective in nature...

\[
\min_{x \in X} f(x) = (f_1(x), \ldots, f_k(x)) \in \mathbb{R}^k
\]
Blackbox Optimization

Most problems are multiobjective in nature...

\[
\min_{x \in X} f(x) = (f_1(x), \ldots, f_k(x)) \in \mathbb{R}^k
\]
Blackbox Optimization

Most problems are multiobjective in nature...

\[
\min_{x \in X} f(x) = (f_1(x), \ldots, f_k(x)) \in \mathbb{R}^k
\]

Issues:

- non-linear
- noisy
- objectives
- non-differentiable
- expensive (e.g. simulations)
- many objectives
- many constraints
- huge search spaces

Pareto Front

Cost

power consumption

© Dimo Brockhoff, LIX, Ecole Polytechnique

“Theoretical Issues of EMO” @ TU Dortmund, September 16, 2011
**Blackbox Optimization**

Most problems are multiobjective in nature...

\[ \min_{x \in X} f(x) = (f_1(x), \ldots, f_k(x)) \in \mathbb{R}^k \]

**Blackbox optimization**

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

Features:
- function f used as an *oracle*
- only mild locality assumptions

---

© Dimo Brockhoff, LIX, Ecole Polytechnique

“Theoretical Issues of EMO” @ TU Dortmund, September 16, 2011
Blackbox Optimization

Most problems are multiobjective in nature...

\[ \min_{x \in X} f(x) = (f_1(x), \ldots, f_k(x)) \in \mathbb{R}^k \]

Features:
- function f used as an oracle
- only mild locality assumptions

Evolutionary Multiobjective Optimization (EMO)

EMO = randomized search heuristics optimizing on solution sets

“sampling” the Pareto front to inform decision maker
Main Purpose of My Talk

- Talk about some of my work
- A subjective list of “hot topics” in the theory of EMO
- Share interesting open questions and ideas

Why?
- build foundation for later discussions this week
- have content for possible collaborations/thesis topics

the GECCO deadline is soon ;-)}
Overview

Benchmarking
“on how to compare sets of solutions”

Indicator-based Search and Preference Articulation
“on how to optimize and steer the search in many-objective problems”

Objective Reduction and Multiobjectivization
“on when to reduce and when to increase the number of objectives”
Once Upon a Time...

... multiobjective EAs were mainly compared visually:

ZDT6 benchmark problem: IBEA, SPEA2, NSGA-II
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

<table>
<thead>
<tr>
<th>Indicator</th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypervolume indicator</td>
<td>6.3431</td>
<td>7.1924</td>
</tr>
<tr>
<td>$\epsilon$-indicator</td>
<td>1.2090</td>
<td>0.12722</td>
</tr>
<tr>
<td>$R_2$ indicator</td>
<td>0.2434</td>
<td>0.1643</td>
</tr>
<tr>
<td>$R_3$ indicator</td>
<td>0.6454</td>
<td>0.3475</td>
</tr>
</tbody>
</table>

see e.g. [Zitzler et al. 2003]
Don’t use an arbitrary quality indicator, but a meaningful one...

<table>
<thead>
<tr>
<th>Indicator</th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generational distance</td>
<td>3.46396</td>
<td>2.37411</td>
</tr>
<tr>
<td>Spacing (Schott)</td>
<td>0.26476</td>
<td>0.19989</td>
</tr>
<tr>
<td>Max Pareto front error</td>
<td>3.35489</td>
<td>3.31314</td>
</tr>
<tr>
<td>Extent</td>
<td>3.56039</td>
<td>3.57319</td>
</tr>
</tbody>
</table>

\[ A \preceq_I B \iff I(A) \leq I(B) \]
Refinements

\[ \text{ref} \ \preceq \ \text{refines} \ \text{a preference relation} \ \preceq \ \text{iff} \]

\[ A \preceq B \land B \not\preceq A \Rightarrow A \preceq B \land B \not\preceq A \]

(\text{better } \Rightarrow \text{ better})

\[ \Rightarrow \ \text{fulfills requirement} \]

…sought are total refinements!

(such as the hypervolume indicator)
but still...

- difficult to interpret absolute numbers
- better: relative values: how far from the optimum (as in single-obj. opt.)

Question:
- what is the optimum?
Optimal $\mu$-Distributions

When the goal is to maximize the hypervolume…

- this yields sets with only Pareto-optimal solutions
  
  [Fleischer 2003]

- those sets, if unrestricted in size, cover the entire Pareto front

- many hypervolume-based EMO algorithms have a population size $\mu$!

Optimal $\mu$-Distribution:

A set of $\mu$ solutions that maximizes a certain (unary) indicator $I$ among all sets of $\mu$ solutions is called optimal $\mu$-distribution for $I$. 
Optimal μ-Distributions

Questions:
- how are optimal μ-distributions characterized?
  - understand the bias of the indicator (influence on DM)
  - 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 [Auger et al. 2009]

Results for 2 objectives only… (except [Auger et al. 2010])

\[ f : x \in D \rightarrow f(x) \]

hypervolume indicator:

\[ I_H((x_1, \ldots, x_\mu)) := \sum_{i=1}^{\mu} (x_{i+1} - x_i)(f(x_0) - f(x_i)) \]
Proposition 1. (Necessary condition for optimal $\mu$-distributions) If $f$ is continuous, differentiable and $(x_1^\mu, \ldots, x_\mu^\mu)$ denote the $x$-coordinates of a set of $\mu$ points maximizing the hypervolume indicator, then for all $x_{\min} < x_i^\mu < x_{\max}$

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

(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 $x_i^\mu$ or $x_i^\mu$ is at the boundary of the domain
**Interpretation of Necessary Condition**

**Example:** equal distances (only) on linear fronts

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

\[ \alpha (x_{i+1}^\mu - x_i^\mu) = f(x_i^\mu) - f(x_{i-1}^\mu) = \alpha (x_i^\mu - x_{i-1}^\mu) \]

generalization of results in [Emmerich et al. 2005, Beume et al. 2007]

exact optimal \( \mu \)-distribution for linear fronts and any choice of reference point

[Brockhoff 2010]
A Density Result: When $\mu$ Goes to Infinity

**Observation:**

general front shapes too difficult to investigate for finite $\mu$

**Question:**

can we characterize optimal $\mu$-distributions with respect to a density

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

[Auger et al. 2009]
Result and Interpretation

The resulting density is

\[ \delta(x) = \frac{\sqrt{-f'(x)}}{\int_0^{x_{\text{max}}} \sqrt{-f'(x)} \, dx} \]

How can we interpret this?

- bias only depends on slope of \( f \) \textit{in contrast to} [Deb et al. 2005, Zitzler and Thiele 1998]
- density highest where slope = 45° \textit{compliant to} [Beume et al. 2007]
- experimental results for finite and small \( \mu \) support the result
Implications for Benchmarking

- now we can transform multiobjective benchmarking into a single-objective problem (where we sometimes know the optimum)
- we can use exactly the same methodology than for single-objective benchmarking:
  - horizontal view (i.e., fixing target values instead of runtime)
  - ERT
  - performance plots a la BBOB

Observation:

we are not as advanced in EMO as in single-objective optimization
Open Questions

Optimal \( \mu \)-distributions
- uniqueness proofs
- other test problems & other indicators
- >2D
- efficient calculation/approximation
- ‘numbers’ for practical usage (on web page?)

Linear convergence speed
- what’s the problem in current algorithms?
- how to achieve it?

Others
- “good” test functions
- multiobjective BBOB
- effective restarts in EMO
Indicator-based Search and Preference Articulation

“on how to optimize and steer the search in many-objective problems”
Assume, we have chosen a total refinement and therefore an optimization goal
- how to achieve it as fast as possible?

**Example: hypervolume indicator**
- SMS-EMOA (changing the reference point might be bad?!)
- Even with fixed reference point, greedy selection might be bad
- HypE (?!)
- Something else?
- Isn’t the variation operator even more important?

**Needed:**
- better understanding of what’s happening in search
- (first) examples of runtime analyses/convergence speed
Idea of Hypervolume-Based Selection

Main Idea (SMS-EMOA, MO-CMA-ES, HypE, …)
use hypervolume indicator to guide the search: refinement!

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

But: can result in cycles [Judt et al. 2011]
is expensive [Bringmann and Friedrich 2008]
and can result in arbitrarily bad sets compared to the optimal one [Bringmann and Friedrich 2009]
**A Simple Algorithm: SIBEA**

Properties:

- No worsenings of $I_H$
- Duplicated solutions removed first
- Selection similar to SMS-EMOA [Emmerich et al. 2005] and MO-CMA-ES [Igel et al. 2007]

$(\mu+1)$SIBEA

generate initial population $P \subseteq \{0, 1\}^n$ at random

repeat:

1. mutate randomly selected $x \in P$ to $x'$ by flipping each bit of $x$ with probability $1/n$

   \[ P' = P \cup \{x'\} \]

2. for all solutions $x \in P$, determine the hypervolume loss

   \[ d(x) = I_H(P') - I_H(P' \setminus \{x\}) \]

3. choose a $z \in P$ with smallest loss $d(z)$

   \[ P = P' \setminus \{z\} \]
Theorem [Brockhoff et al. 2008]: If $\mu \geq n+1$, the $(\mu+1)$SIBEA optimizes LOTZ in $O(\mu n^2)$ generations.

Sketch of Proof:
- 2k mutations increase $I_H$ (prob. $\frac{1}{\mu} n \left(1 - \frac{1}{n}\right)^{n-1} \geq \frac{1}{e\mu n}$)
- Total increase $\geq \max\{X_{\max}, Y_{\max}\} \geq \sqrt{X_{\max} \cdot Y_{\max}} \geq \sqrt{I_H}$
- Exp. increase for 1 mutation $\geq \sqrt{I_H} / 2k$; with Markov: i.e., in 8k good mutations $\sqrt{I_H}$ w.h.p.
- Exp. runtime for increase by $\sqrt{I_H}$ is $O\left(\frac{\mu n}{2k} \cdot 8k\right) = O(\mu n)$
- By induction, $O(n)$ such increases sufficient to reach front, then $O(\mu n)$ time enough to find all other $n$ points
A More Involved Selection Scheme: HypE

**Idea** [Bader and Zitzler 2011]

Solution quality = expected loss, when removing the point and \((\text{randomly})\) \(k-1\) others

Comparison HypE/standard:

<table>
<thead>
<tr>
<th></th>
<th>opt.</th>
<th>dist</th>
<th>better</th>
</tr>
</thead>
<tbody>
<tr>
<td>new</td>
<td>59.7%</td>
<td>0.00109</td>
<td>30.2%</td>
</tr>
<tr>
<td>standard</td>
<td>44.5%</td>
<td>0.00261</td>
<td>3.2%</td>
</tr>
</tbody>
</table>

**Question:**
can we show the improvement also theoretically?
Articulating User Preferences

What if user wants something else than finding the optimal $\mu$-distribution for the hypervolume indicator? E.g.
- (p)reference points
- stressing extremes
- simulate classical scalarizing function approaches

Idea:
[Zitzler et al. 2007]
Articulating User Preferences

What if user wants something else than finding the optimal $\mu$-distribution for the hypervolume indicator? E.g.

- (p)reference points
- stressing extremes
- simulate classical scalarizing function approaches

Idea:

[Zitzler et al. 2007]
Examples of Weight Functions

preference point

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

stressing one objective

\[ w(z_1, \ldots, z_k) = \begin{cases} \left( \prod_{i \neq s} (b_i^u - b_i^l) \right)^{-1} \lambda e^{-\lambda(z_i - b_i^l)} & \text{if } \vec{z} \in B \\ 0 & \text{if } \vec{z} \notin B \end{cases} \]

Question:

Does this work also interactively?
Some Experimental Results

Preliminary results shows yes:

interaction every 100 iterations:
choose alternatively leftmost/rightmost point

Observation:
Very difficult to assess those interactive methods in a decent way
Open Questions

HypE
- why is HypE better than normal HYP-based selection?
- and when? (Is there an example where it’s provably better?)
- by how much (convergence speed?)
- greedy vs. oneShot: advantages and disadvantages
- a more advanced scheme than assuming uniform deletion?

SMS-EMOA: does algo becomes faster if HYP worsenings are not allowed (eg. by keeping old population if new one is worse)?

Convergence to optimal $\mu$-distribution
- do other algorithms converge to optimal $\mu$-distribution for other indicators?

Others
- more runtime analyses of indicator-based EMO
  - weighted hypervolume $\rightarrow$ reduced pop size of SEMO?
- preferences: how to evaluate/compare algos objectively?
Objective Reduction and Multiobjectivization

“on when to reduce and when to increase the number of objectives”
Statements are contradictory: some studies say that...

- **Problems may become harder**
  - [Fonseca and Fleming 1995], [Deb 2001], [Coello et al. 2002], and others:
    - conflicts between objectives
    - Pareto front size
    - # incomparable solutions
  - [Winkler 1985]:
    - theoretical work for random objectives

- **Problems may become easier**
  - [Knowles et al. 2001]:
    - multiobjectivization
  - [Jensen 2004]:
    - helper-objectives
  - [Scharnow et al. 2002], [Neumann and Wegener 2006]:
    - theoretical investigations
    - 2D faster than 1D
    - decomposition
Adding Objectives: Runtime Analysis

\[ \text{PLATEAU}_1(x) := \begin{cases} 
|x|_0 & : x \not\in \{1^i0^{n-i}, 1 \leq i \leq n\} \\
n + 1 & : x \in \{1^i0^{n-i}, 1 \leq i < n\} \\
n + 2 & : x = 1^n.
\]
Conclusions When Adding Objectives

Additional objectives can:
- turn a region with direction into a plateau of incomparable solutions
- add direction to a plateau of indifferent solutions

Contrary, removing objectives can do the opposite
- and therefore might also reduce the optimization time
- interesting: removing objectives results in a refinement!

Several works on automated objective reduction
- for reducing the runtime of hypervolume-based methods in many-objective optimization
- for giving insights into the problem for the decision maker
Open Questions

- faster aggregation heuristics
- what happens exactly when aggregating objectives?
  - which orders can be generated by e.g. a weighted sum?
- test problems with changing conflict
- GUI for decision support (incl. innovization?)
- online reduction:
  - when to delete, when to add objectives? (MAB problem)
- more examples of multiobjectivization:
  - both with runtime analysis + experimental
Conclusions

- Three aspects of Theory in EMO
  - benchmarking
  - indicator-based search and preference articulation
  - objective reduction and multiobjectivization

- Many open questions
- Lots of ideas for future work

...let’s do it 😊
French Summer School in Evolutionary Algorithms

June 12-15, 2012
Quiberon (Bretagne)

organizers: D. Brockhoff, L. Jourdan, A. Liefooghe, S. Verel
References


References


References


