6th GECCO Workshop on Blackbox Optimization Benchmarking (BBOB): Session II Wrap-up

The BBOBies

https://github.com/numbbo/coco

slides based on previous ones by A. Auger, N. Hansen, and D. Brockhoff

Need: Benchmarking

- understanding of algorithms
- algorithm selection
- putting algorithms to a standardized test
  - simplify judgement
  - simplify comparison
  - regression test under algorithm changes

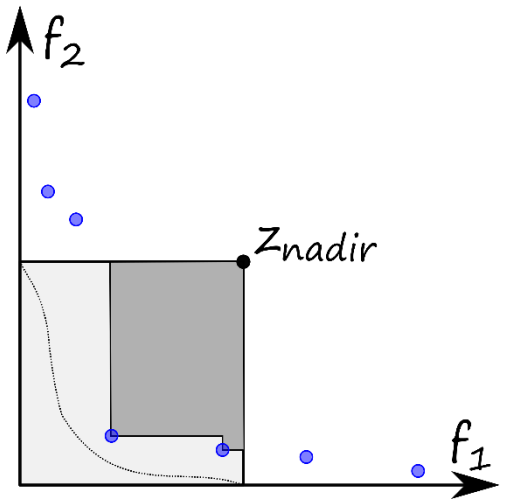
Kind of everybody has to do it (and it is tedious):

- choosing (and implementing) problems, performance measures, visualization, stat. tests, ...
- running a set of algorithms

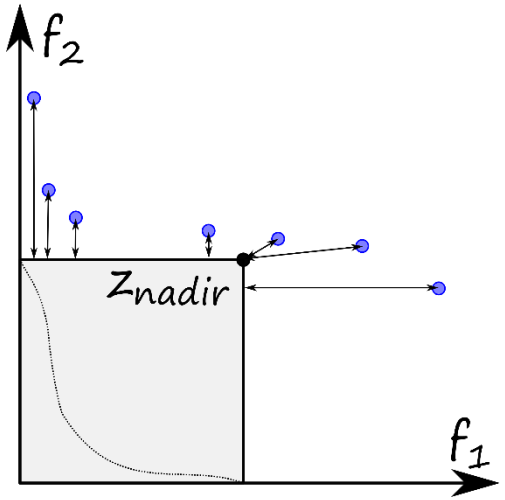
Bi-objective Performance Assessment

algorithm quality =

normalized* hypervolume (HV)
of all non-dominated solutions
if a point dominates nadir



closest normalized* negative distance
to region of interest $[0,1]^2$
if no point dominates nadir



* such that ideal=[0,0] and nadir=[1,1]

Bi-objective Performance Assessment

We measure runtimes to reach (HV indicator) targets:

- relative to a reference set, given as the best Pareto front approximation known (since exact Pareto set not known)
  - for the workshop: `before_workshop` values
  - from now on: updated `current_best` values incl. all non-dominated points found by the 15 workshop algos: will be available soon and hopefully fixed for some time
- actual absolute hypervolume targets used are

HV(refset) – targetprecision

~~51~~ **51** with ~~58~~ fixed targetprecisions between 1 and ~~10^{-4}~~ **10^{-5}** (same for all functions, dimensions, and instances) in the displays

all 10 instances !

Bi-objective Performance Assessment

Session 2

All algorithms of this session + best of BBOB-2016 as reference

- MO-DIRECT-hv(HV-Rank)
- MO-DIRECT-hv(ND)
- MO-DIRECT-hv(Rank)

- MAT-DIRECT
- MAT-SMS

- UP-MO-CMA-ES

data of all session II algorithms

Bi-objective Performance Assessment

Session 2

For better comparison, we also provide some baselines:

- random search within [-100, 100] (RS-100)
- random search within [-5, 5] (RS-5)
- random search within [-4, 4] (RS-4)
- NSGA-II (MATLAB gamultiobj implementation, with restarts)
- SMS-EMOA (MATLAB implementation of T. Wagner)
  - SMS-EMOA-PM: with polynomial mutation and SBX
  - SMS-EMOA-DE: with differential evolution
- RM-MEDA (implementation by Q. Zhang et al.)

data of session II algorithms plus baselines

Bi-objective Performance Assessment

Session 2 - Conclusions

- MAT-DIRECT and MAT-SMS the worst algorithms
  - confirming the results from the single-objective MATSuMoTo library
- MO-DIRECT variants quite good for small dimensions and small budgets
  - variant MO-DIRECT-hv(HV-Rank) clearly the best one
- Impact of initialization
  - search domain of initial point important (closer to 0 → better)
- Impact of dimension n
  - UP-MO-CMA-ES not much affected (always relative to refset!)
  - the other algos are becoming (relatively) worse with larger n

General Advice

- be careful when looking at data!

Science is a way of trying not to fool yourself.

The first principle is that you must not fool yourself, and you are the easiest person to fool. So you have to be very careful about that. After you've not fooled yourself, it's easy not to fool other[scientist]s. You just have to be honest in a conventional way after that. -- *Richard P. Feynman*

- e.g. ECDFs can look different although they come from the same data because we bootstrap runtimes of a simulated restarted algorithm