6th GECCO Workshop on Blackbox Optimization Benchmarking (BBOB): Wrap-Up

The BBOBies
https://github.com/numbbo/coco
Numerical Blackbox Optimization

Optimize $f: \Omega \subset \mathbb{R}^n \mapsto \mathbb{R}^k$

$x \in \mathbb{R}^n \quad f(x) \in \mathbb{R}^k$

derivatives not available or not useful
Need: Benchmarking

• understanding of algorithms
• algorithm selection
• putting algorithms to a standardized test
  • simplify judgement
  • simplify comparison
  • regression test under algorithm changes
that's where COCO and BBOB come into play

Comparing Continuous Optimizers Platform

https://github.com/numbbo/coco
Available Data Sets in COCO before and after 2016

Before 2016

- **bbob**: 140+ algo data sets
- **bbob-noisy**: 40+ algo data sets
- **bbob-biobj**

In 2016

- **bbob**: 4 data sets
- **bbob-noisy**: 4 data sets
- **bbob-biobj**: 15 data sets

**BBOB Submissions (#data sets)**

- # bbo - algdatasets
- # bbo-noisy algdatasets
- # bbo-biobj algdatasets
extension of COCO to multi-objective optimization
bbob-biobj Testbed (new in 2016)

- 55 functions, combining bbob functions
- 6 dimensions (2..40D)
- no normalization
- ideal/nadir known
- but Pareto set/front not (only refsets)
Bi-objective Performance Assessment

algorithm quality =

\[
\text{normalized* hypervolume (HV) of all non-dominated solutions} \\
\text{if a point dominates nadir} \\
\text{closest normalized* negative distance to region of interest } [0,1]^2 \\
\text{if no point dominates nadir}
\]

* such that ideal=[0,0] and nadir=[1,1]
Again, as in last session's wrap-up:

- results are relative to a reference set, given as the best Pareto front approximation known (since exact Pareto set not known)
  - note: improved reference sets compared to workshop papers
- actual absolute hypervolume targets used are
  \[ HV(\text{refset}) - \text{targetprecision} \]
  with 51 fixed targetprecisions between 1 and \(10^{-5}\) (same for all functions, dimensions, and instances) in the displays
- all 10 instances are displayed
let's dig into the data...
all results for the bbob-biobj suites
Bi-objective Performance Assessment

The ECDFs are actually influenced by

1. the number and set of instances and by
2. the reference set and the reference hypervolume values

...so let's have a brief look behind the scenes of BBOB
the influence of the instances
BBOB-2016: Instances

- on the bbo-biobj test suite, experiments were run on 10 instances
- but all plots were based on the first 5 instances only
  - practical reason 1: we did not have enough data to produce good hypervolume reference values for all instances
  - practical reason 2: setting allows to investigate potential overfitting ("split between training and test")
BBOB-2016: Instances 1-5 (2-D)
BBOB-2016: Instances 6-10 (2-D)

Proportion of function+target pairs

log10 of (# f-evals / dimension)
BBOB-2016: Instances 1-10 (2-D)
now 20-D
BBOB-2016: Instances 1-5

Proportion of function+target pairs vs. log10 of (# f-evals / dimension)

- bbob-biobj-f1-f55, 20-D
- 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5 instances

Best 2016
- HMO-CMA-ES
- UP-MO-CMA-ES
- RM-MEDA
- DEMO
- SMS-EMOA-DE
- GA-MULTIOBJ(NSG)
- SMS-EMOA-PM
- MO-DIRECT-hv(HV)
- RS-4
- RS-5
- MO-DIRECT-hv(Ra)
- MO-DIRECT-hv(ND)
- MAT-DIRECT
- MAT-SMS
- RS-100
BBOB-2016: Instances 6-10

bbob-biobj + f1-f55, 20-D
5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5

Proportion of function+target pairs

log10 of (# f-evals / dimension)
BBOB-2016: Instances 1-10

Proportion of function+target pairs vs. log10 of (# f-evals / dimension)

Methods:
- HMO-CMA-ES
- UP-MO-CMA-ES
- RM-MEDA
- DEMO
- SMS-EMOA-DE
- GA-MULTIOBJ(NSG)
- SMS-EMOA-PM
- MO-DIRECT-hv(HV-4)
- MO-DIRECT-hv(HV-5)
- MO-DIRECT-hv(Rar)
- MO-DIRECT-hv(ND)
- MAT-DIRECT
- MAT-SMS
- RS-100
Influence of the Instance Set

• is relatively small
• sometimes, last 5 instances harder, sometimes first 5 (depending on dimension)
• no indication of overfitting to the first 5 instances
2 the influence of the reference set
BBOB-2016: Reference Set Before

Proportion of function+target pairs

log10 of (# f-evals / dimension)
BBOB-2016: Reference Set After

Proportion of function+target pairs vs. log10 of (\# f-evals / dimension) for various algorithms.
quick check: first 5 instances
Influence of the Reference Set

• impact by the workshop algorithms the largest
• mainly on second five instances
  • which means the provided and displayed reference sets were okay
• continue with the current best in the future
  • updated reference hypervolume values will be provided in one of the next releases (this summer for sure)
• investigations on the single functions show that for some, we still do not have a good enough reference set yet
The Future of COCO

• bi-objective data will be made available online in the next days

• towards more realistic problems
  • large-scale test suite soon ready for release
  • constraints potentially ready in 2017
  • "almost real-world" problems

• online visualization of data
Your Participation it Welcome...

• ...always 😊

• benchmark your own algorithm and submit next year
• report bugs, issues, and feature requests
  • https://github.com/numbbo/coco
• contribute to the code base on github
  • issue tracker has special flag easy

• or even join us in Paris
  • as an engineer (funding for 1 year available)
  • or as postdoc, PhD student, or intern