Evaluation Report on Twelve Experiments

Anne Anonymous

***Abstract*—This is the evaluation report on twelve experiments, namely P-DCN, holtschulte2013_ga100, holtschulte2013_hill, hutter2013_CMAES, liao2013_IPOP, liao2013_IPOP-500, liao2013_IPOP-tany, liao2013_IPOP-texp, pal2013_DE, pal2013_HMLSL, pal2013_fmincon, and pal2013_simplex on 144 benchmark instances. This report has been generated with the version 0.8.4 of the Evaluator Component of the Optimization Benchmarking Tool Suite.**

I. Performance Comparisons

A. Estimated Cumulative Distribution Function

We analyze the estimated cumulative distribution function (*ECDF*) [1], [2], [3] of F over $\log_{10} FEs$. The $\mathrm{ECDF}(FEs, F \leq \text{1.E-8})$ represents the fraction of runs which reach a value of F less than or equal to 1.E-8 for a given ellapsed runtime measured in *FEs*. The *ECDF* is always computed over the runs of an experiment for a given benchmark instance. If runs for multiple instances are available, we aggregate the results by computing their arithmetic mean. The x-axis does not represent the values of *FEs* directly, but instead $\log_{10} FEs$. The *ECDF* is always between 0 and 1 — and the higher it is, the better.

The corresponding plot is illustrated in Figure 1.

B. Estimated Cumulative Distribution Function

We analyze the estimated cumulative distribution function (*ECDF*) [1], [2], [3] of F over $\log_{10} FEs$. The $\mathrm{ECDF}(FEs, F \leq \text{1.E-8})$ represents the fraction of runs which reach a value of F less than or equal to 1.E-8 for a given ellapsed runtime measured in *FEs*. The *ECDF* is always computed over the runs of an experiment for a given benchmark instance. If runs for multiple instances are available, we aggregate the results by computing their arithmetic mean. The x-axis does not represent the values of *FEs* directly, but instead $\log_{10} FEs$. The *ECDF* is always between 0 and 1 — and the higher it is, the better. The instance run sets belonging to instances with the same value of the feature *dim* grouped together.

The corresponding plots are illustrated in Figure 2.

C. Estimated Cumulative Distribution Function

We analyze the estimated cumulative distribution function (*ECDF*) [1], [2], [3] of F over $\log_{10} FEs$. The $\mathrm{ECDF}(FEs, F \leq \text{1.E-5})$ represents the fraction of runs which reach a value of F less than or equal to 1.E-5 for a given ellapsed runtime measured in *FEs*. The *ECDF* is always computed over the runs of an experiment for a given benchmark instance. If runs for multiple instances are available, we aggregate the results by computing their arithmetic mean. The x-axis does not represent the values of *FEs* directly, but instead $\log_{10} FEs$. The *ECDF* is always between 0 and 1 — and the higher it is, the better. The instance run sets belonging to instances with the same value of the feature *cond* grouped together.

The corresponding plots are illustrated in Figure 3.

D. Median of Medians

We analyze the median of medians (*med med*) computed based on $\log_{10} F$ over $\log_{10}\left(\frac{FEs}{dim^2}\right)$. The $\mathrm{med\ med}(FEs, \log_{10} F)$ represents the median of the $\log_{10} F$ for a given ellapsed runtime measured in *FEs*. The median is always computed over the runs of an experiment for a given benchmark instance. If runs for multiple instances are available, we aggregate these medians by computing their median. The x-axis does not represent the values of *FEs* directly, but instead $\log_{10}\left(\frac{FEs}{dim^2}\right)$. The instance run sets belonging to instances with the same value of the feature *sep* grouped together.

The corresponding plots are illustrated in Figure 4.

References

[1] H. H. Hoos and T. Stützle, "Evaluating las vegas algorithms — pitfalls and remedies," in *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence (UAI'98)*, G. F. Cooper and S. Moral, Eds. Madison, WI, USA: San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., Jul. 24–26, 1998, pp. 238–245. [Online]. Available: http://www.intellektik.informatik.tu-darmstadt.de/TR/1998/98-02.ps.Z

[2] D. A. D. Tompkins and H. H. Hoos, "Ubcsat: An implementation and experimentation environment for sls algorithms for sat and max-sat," in *Revised Selected Papers from the Seventh International Conference on Theory and Applications of Satisfiability Testing (SAT'04)*, ser. Lecture Notes in Computer Science (LNCS), H. H. Hoos and D. G. Mitchell, Eds., vol. 3542. Vancouver, BC, Canada: Berlin, Germany: Springer-Verlag GmbH, May 10–13, 2004, pp. 306–320. [Online]. Available: http://ubcsat.dtompkins.com/downloads/sat04proc-ubcsat.pdf?attredirects=0

[3] N. Hansen, A. Auger, S. Finck, and R. Ros, "Real-parameter black-box optimization benchmarking: Experimental setup," Orsay, France: Université Paris Sud, Institut National de Recherche en Informatique et en Automatique (INRIA) Futurs, Équipe TAO, Tech. Rep., Mar. 24, 2012. [Online]. Available: http://coco.lri.fr/BBOB-downloads/download11.05/bbobdocexperiment.pdf