

# On the Impact of Active Covariance Matrix Adaptation in the CMA-ES With Mirrored Mutations and Small Initial Population Size on the Noiseless BBOB Testbed

Dimo Brockhoff  
INRIA Lille - Nord Europe  
Dolphin team  
59650 Villeneuve d'Ascq  
France  
dimo.brockhoff@inria.fr

Anne Auger  
Projet TAO, INRIA  
Saclay—Ile-de-France  
LRI, Bât 490, Univ. Paris-Sud  
91405 Orsay Cedex, France  
anne.auger@inria.fr

Nikolaus Hansen  
Projet TAO, INRIA  
Saclay—Ile-de-France  
LRI, Bât 490, Univ. Paris-Sud  
91405 Orsay Cedex, France  
nikolaus.hansen@inria.fr

## ABSTRACT

Mirrored mutations as well as active covariance matrix adaptation are two techniques that have been introduced into the well-known CMA-ES algorithm for numerical optimization. Here, we investigate the impact of active covariance matrix adaptation in the IPOP-CMA-ES with mirrored mutation and a small initial population size. Active covariance matrix adaptation improves the performance on 8 of the 24 benchmark functions of the noiseless BBOB test bed. The effect is the largest on the ill-conditioned functions with the largest improvement on the discus function where the expected runtime is more than halved. On the other hand, no statistically significant adverse effects can be observed.

## Categories and Subject Descriptors

G.1.6 [Numerical Analysis]: Optimization—*global optimization, unconstrained optimization*; F.2.1 [Analysis of Algorithms and Problem Complexity]: Numerical Algorithms and Problems

## General Terms

Algorithms

## Keywords

Benchmarking, Black-box optimization

## 1. INTRODUCTION

The covariance matrix adaptation evolution strategy (CMA-ES) is considered as one of the standard stochastic optimization algorithms for continuous domain. In recent years, several algorithmic improvements of the original version have been made. One is the idea of increasing the population size (by a factor of two) after each restart of the algorithm

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

GECCO'12 Companion, July 7–11, 2012, Philadelphia, PA, USA.  
Copyright 2012 ACM 978-1-4503-1178-6/12/07 ...\$10.00.

(IPOP-CMA-ES, [2]). Two other improvements, are the active covariance matrix update [9] and mirrored mutations [1]. In an accompanying paper [3], it is shown that both active covariance matrix adaptation and mirrored mutations can be easily combined in the standard IPOP-CMA-ES. Here we pursue the question how strong is the influence of the active covariance matrix adaptation on the overall performance. To this end, we compare the IPOP-CMA-ES with mirrored mutations and a small initial population size with the algorithm enhanced by active covariance matrix adaptation on the noiseless BBOB test bed. We find that the additional active covariance matrix adaptation improves the algorithm significantly on about one third of all functions while no detrimental effect can be observed.

## 2. ALGORITHM PRESENTATION

The baseline algorithm for our comparison is the IPOP-CMA-ES with mirrored mutations and an initial population size of  $\lfloor (4 + \lfloor 3 \log(D) \rfloor) / 2 \rfloor$ , i.e., a population size that is half as large as in the original IPOP-CMA-ES where  $D$  is the problem dimension. The reason is that in initial experiments, mirrored mutations showed a larger impact when the population size is small [4]. We denote this baseline algorithm as  $\text{CMA}_{\text{mh}}^1$ . On the other hand, we have the algorithm  $\text{CMA}_{\text{mah}}$  with additional active covariance matrix adaptation which is implemented according to [8].

Both algorithms are run for a maximum of  $2 \cdot 10^5 \cdot D$  function evaluations and restarted up to 10 times. Besides the reduced initial population size, all parameters are set as recommended in the original CMA-ES. The MATLAB source code is version 3.54.beta.mirrors of the CMA-ES and can be found at <http://canadafrance.gforge.inria.fr/mirroring/>.

## 3. TIMING EXPERIMENTS

In order to see the dependency of the algorithms on the problem dimension, the requested BBOB'2012 timing experiment has been performed for the two algorithms  $\text{CMA}_{\text{mh}}$  and  $\text{CMA}_{\text{mah}}$  on an Intel Core2 Duo T9600 laptop with 2.80GHz, 4.0GB of RAM, and MATLAB R2008b on Windows Vista SP2. The algorithms have been restarted for up to  $2 \cdot 10^5 N$  function evaluations until 30 seconds have

<sup>1</sup>The index 'm' stands for mirrored mutations and the index 'h' for half the standard population size.

passed. The per-function-evaluation-runtimes were 22; 19; 12; 8.3; 5.8; 5.8 and 12 times  $10^{-4}$  seconds for the CMA<sub>mh</sub> and 21; 19; 11; 8.3; 6.1; 5.7 and 11 times  $10^{-4}$  seconds for the CMA<sub>mah</sub> in 2, 3, 5, 10, 20, 40, and 80 dimensions respectively.

## 4. RESULTS

Results from experiments according to [6] on the benchmark functions given in [5, 7] are presented in Figures 1, 2 and 3 and in Tables 1. The **expected running time (ERT)**, used in the figures and table, depends on a given target function value,  $f_t = f_{\text{opt}} + \Delta f$ , and is computed over all relevant trials as the number of function evaluations executed during each trial while the best function value did not reach  $f_t$ , summed over all trials and divided by the number of trials that actually reached  $f_t$  [6, 10]. **Statistical significance** is tested with the rank-sum test for a given target  $\Delta f_t$  ( $10^{-8}$  as in Figure 1) using, for each trial, either the number of needed function evaluations to reach  $\Delta f_t$  (inverted and multiplied by  $-1$ ), or, if the target was not reached, the best  $\Delta f$ -value achieved, measured only up to the smallest number of overall function evaluations for any unsuccessful trial under consideration.

The main observation is that the IPOP-CMA-ES variant with additional covariance matrix adaptation is never statistically worse than the one without active covariance matrix adaptation. The strongest positive effect of the additional covariance matrix adaptation can be observed for the ill-conditioned functions (Fig. 3) where for all test functions in this class the CMA<sub>mah</sub> statistically significantly outperforms the CMA<sub>mh</sub> for a target of  $10^{-8}$  (Fig. 1). The factor of improvement in terms of ERT for these functions lies between 1.3 (for the ellipsoid function) and 2.24 (for the discuss function) in 40D, see Table 1. Also on the separable ellipsoid, the CMA<sub>mah</sub> statistically significantly outperforms the CMA<sub>mh</sub> for all tested dimensions. Also on other functions, statistically significant improvements can be observed, mainly found for larger dimensions: for 20D and 40D on  $f_6$ , for 10D, 20D, and 40D on  $f_7$ , for 20D on  $f_8$ , for 20D and 40D on  $f_{12}$ , and for 40D on  $f_{18}$ . Overall, there are only two functions where the CMA<sub>mh</sub> is (slightly) better in 20D for difficult targets than the CMA<sub>mah</sub>. Furthermore, within the given budget, for  $f_3$ ,  $f_4$ ,  $f_{19}$  and for all weakly structured functions, unsuccessful runs occur for both algorithms with a slight advantage for the CMA<sub>mah</sub> (with 6 successful runs more than the CMA<sub>mh</sub>).

In summary, active covariance matrix adaptation regularly improves the performance of IPOP-CMA-ES with mirrored mutations, also when the initial population size is smaller than the standard population size.

## 5. REFERENCES

- [1] A. Auger, D. Brockhoff, and N. Hansen. Mirrored Sampling in Evolution Strategies With Weighted Recombination. In *Genetic and Evolutionary Computation Conference (GECCO 2011)*, pages 861–868. ACM, 2011.
- [2] A. Auger and N. Hansen. A Restart CMA Evolution Strategy With Increasing Population Size. In *Congress on Evolutionary Computation (CEC 2005)*, volume 2, pages 1769–1776. IEEE Press, 2005.
- [3] D. Brockhoff, A. Auger, and N. Hansen. On the Effect of Mirroring in the IPOP Active CMA-ES on the Noiseless BBOB Testbed. In *GECCO (Companion) workshop on Black-Box Optimization Benchmarking (BBOB'2012)*. ACM, 2012.
- [4] D. Brockhoff, A. Auger, N. Hansen, D. V. Arnold, and T. Hohm. Mirrored Sampling and Sequential Selection for Evolution Strategies. In *Conference on Parallel Problem Solving from Nature (PPSN XI)*, pages 11–21. Springer, 2010.
- [5] S. Finck, N. Hansen, R. Ros, and A. Auger. Real-parameter black-box optimization benchmarking 2009: Presentation of the noiseless functions. Technical Report 2009/20, Research Center PPE, 2009. Updated February 2010.
- [6] N. Hansen, A. Auger, S. Finck, and R. Ros. Real-parameter black-box optimization benchmarking 2012: Experimental setup. Technical report, INRIA, 2012.
- [7] N. Hansen, S. Finck, R. Ros, and A. Auger. Real-parameter black-box optimization benchmarking 2009: Noiseless functions definitions. Technical Report RR-6829, INRIA, 2009. Updated February 2010.
- [8] N. Hansen and R. Ros. Benchmarking a weighted negative covariance matrix update on the BBOB-2010 noiseless testbed. In *Genetic and Evolutionary Computation Conference (GECCO 2010)*, pages 1673–1680, New York, NY, USA, 2010. ACM.
- [9] G. Jastrebski and D. Arnold. Improving evolution strategies through active covariance matrix adaptation. In *IEEE Congress on Evolutionary Computation (CEC 2006)*, pages 2814–2821, 2006.
- [10] K. Price. Differential evolution vs. the functions of the second. In *Proceedings of the IEEE International Congress on Evolutionary Computation (ICEO)*, pages 153–157, 1997.

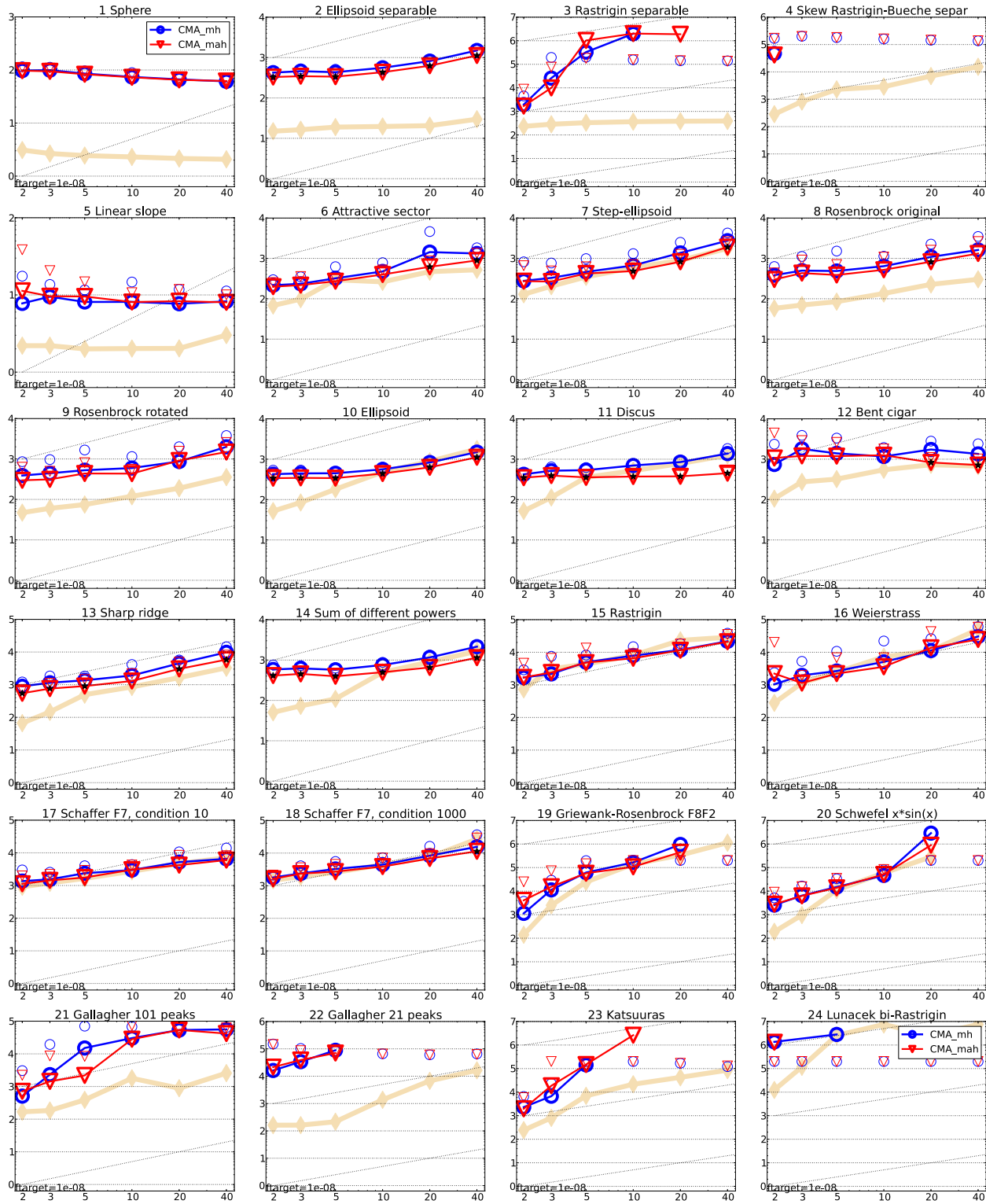


Figure 1: Expected running time (ERT in number of  $f$ -evaluations) divided by dimension for target function value  $10^{-8}$  as  $\log_{10}$  values versus dimension. Different symbols correspond to different algorithms given in the legend of  $f_1$  and  $f_{24}$ . Light symbols give the maximum number of function evaluations from the longest trial divided by dimension. Horizontal lines give linear scaling, slanted dotted lines give quadratic scaling. Black stars indicate statistically better result compared to all other algorithms with  $p < 0.01$  and Bonferroni correction number of dimensions (six). Legend:  $\circ$ :CMA\_mh,  $\nabla$ :CMA\_mah.

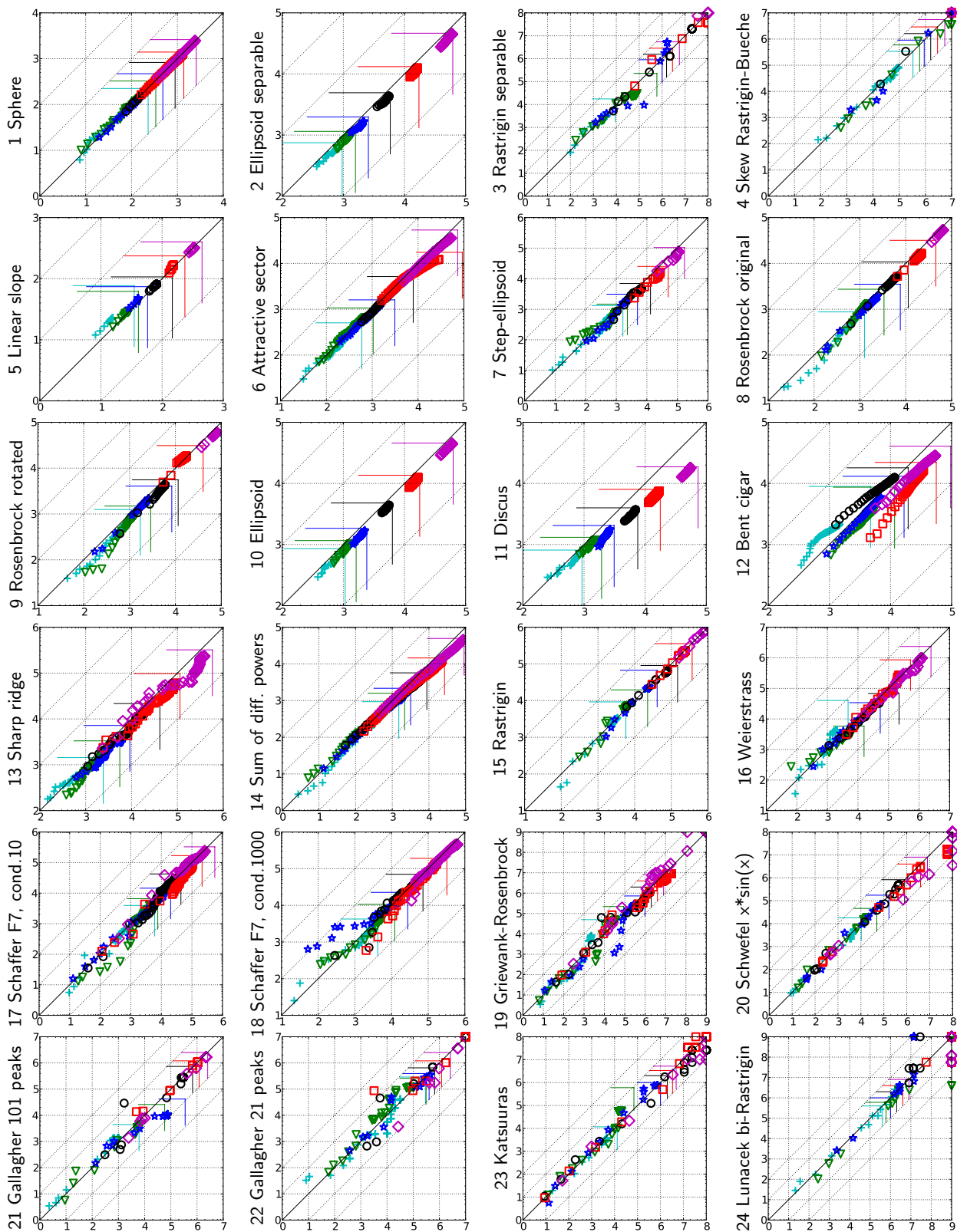


Figure 2: Expected running time (ERT in  $\log_{10}$  of number of function evaluations) of CMA\_mh ( $x$ -axis) versus CMA\_mah ( $y$ -axis) for 46 target values  $\Delta f \in [10^{-8}, 10]$  in each dimension on functions  $f_1$ - $f_{24}$ . Markers on the upper or right edge indicate that the target value was never reached. Markers represent dimension: 2:+, 3:∇, 5:\*, 10:○, 20:□, 40:◇.

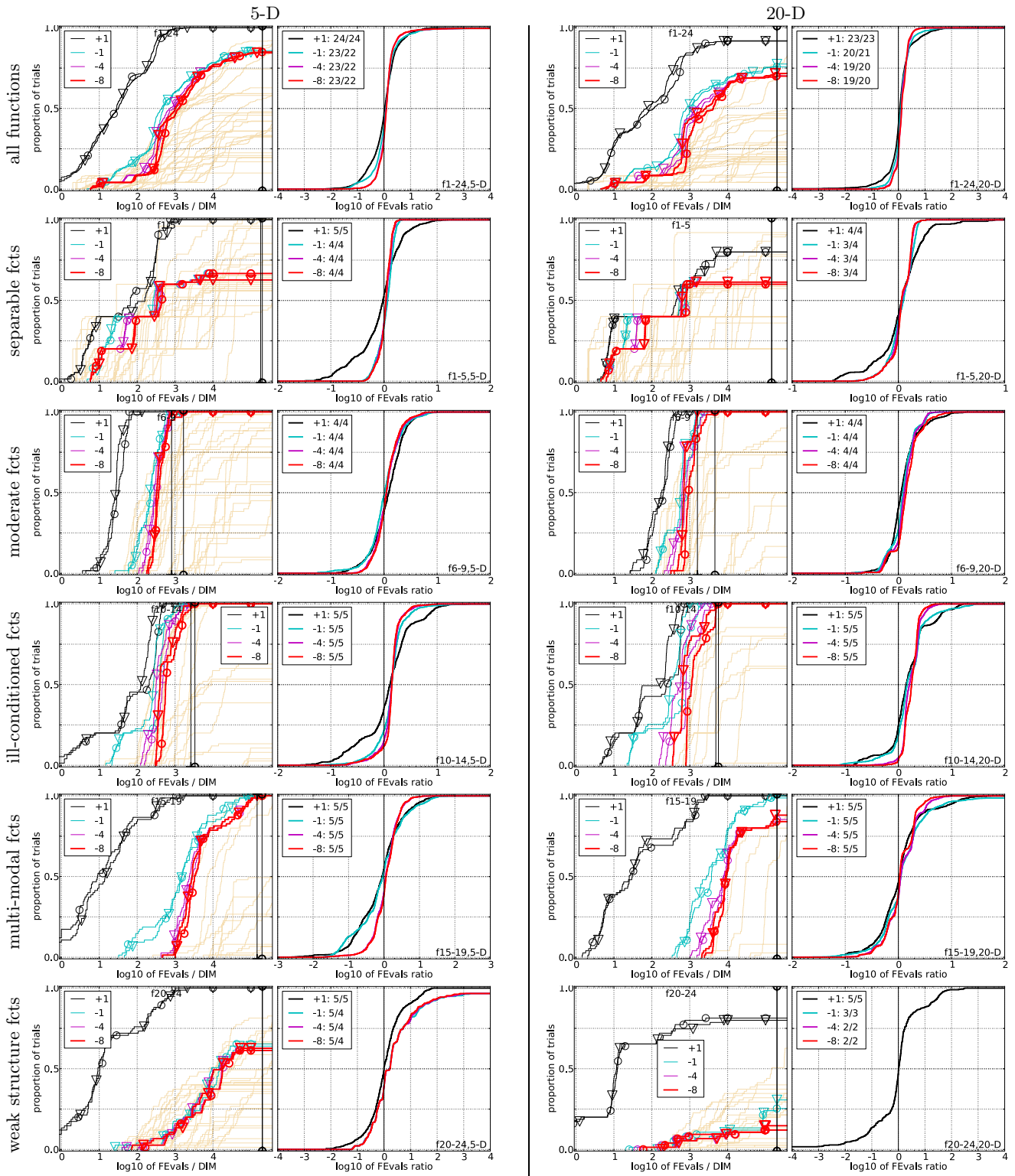


Figure 3: Empirical cumulative distributions (ECDF) of run lengths and speed-up ratios in 5-D (left) and 20-D (right). Left sub-columns: ECDF of the number of function evaluations divided by dimension  $D$  (FEvals/ $D$ ) to reach a target value  $f_{\text{opt}} + \Delta f$  with  $\Delta f = 10^k$ , where  $k \in \{1, -1, -4, -8\}$  is given by the first value in the legend, for CMA\_mh ( $\circ$ ) and CMA\_mah ( $\nabla$ ). Light beige lines show the ECDF of FEvals for target value  $\Delta f = 10^{-8}$  of all algorithms benchmarked during BBOB-2009. Right sub-columns: ECDF of FEval ratios of CMA\_mh divided by CMA\_mah, all trial pairs for each function. Pairs where both trials failed are disregarded, pairs where one trial failed are visible in the limits being  $> 0$  or  $< 1$ . The legends indicate the number of functions that were solved in at least one trial (CMA\_mh first).

