



# On the Impact of a Small Initial Population Size in the IPOP Active CMA-ES with Mirrored Mutations on the Noiseless BBOB Testbed

Dimo Brockhoff, Anne Auger, Nikolaus Hansen

## ► To cite this version:

Dimo Brockhoff, Anne Auger, Nikolaus Hansen. On the Impact of a Small Initial Population Size in the IPOP Active CMA-ES with Mirrored Mutations on the Noiseless BBOB Testbed. GECCO Companion '12, Jul 2012, Philadelphia, PA, United States. pp.285-290, 10.1145/2330784.2330825 . hal-00746122

**HAL Id: hal-00746122**

**<https://hal.inria.fr/hal-00746122>**

Submitted on 27 Oct 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# On the Impact of a Small Initial Population Size in the IPOP Active CMA-ES with Mirrored Mutations on the Noiseless BBOB Testbed

Dimo Brockhoff  
INRIA Lille - Nord Europe  
Dolphin team  
59650 Villeneuve d'Ascq  
France  
dimobrockhoff@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

Active Covariance Matrix Adaptation and Mirrored Mutations have been independently proposed as improved variants of the well-known optimization algorithm Covariance Matrix Adaptation Evolution Strategy (CMA-ES) for numerical optimization. This paper investigates the impact of the algorithm's population size when both active covariance matrix adaptation and mirrored mutation are used in the CMA-ES. To this end, we compare the CMA-ES with standard population size  $\lambda$ , i.e.,  $\lambda = 4 + \lfloor 3 \log(D) \rfloor$  with a version with half this population size where  $D$  is the problem dimension.

## 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 IPOP-CMA-ES [2] has the special feature of increasing the population size of the CMA-ES algorithm at each restart. Together with a standard population size of  $\lambda^s = 4 + \lfloor 3 \log(D) \rfloor$  where  $D$  is the problem dimension, the IPOP-CMA-ES is a (nearly) parameterless algorithm that automatically restarts CMA-ES with increased population size if the given size is not sufficient to solve the problem at hand.

More recently, an active covariance matrix adaptation update has been proposed for CMA-ES [9] and mirrored mutations with pairwise selection and selective mirroring have been suggested for evolution strategies with weighted recombination [1]. While the former one allows for negative weights in the covariance matrix update for bad mutations, the latter mirrors the bad mutations and evaluates them again before to proceed.

The combination of both approaches into the IPOP-CMA-ES with active covariance matrix adaptation and mirrored mutations, denoted by  $\text{CMA}_{\text{ma}}$ , has been introduced and tested empirically in an accompanying paper [3]. Here, we test how a different starting population size influences the performance of this algorithm. A previous study showed that in the  $(1, \lambda)$ -ES, the largest effect of mirrored mutations is observed for small population sizes, i.e.  $\lambda = 2$  and  $\lambda = 4$  [4]. Hence, we could conjecture that in the IPOP-CMA-ES with mirrored mutations, a positive effect on the performance can be observed if the initial population size is chosen smaller than the standard size of  $\lambda^s$ . To test this hypothesis, we run the  $\text{CMA}_{\text{ma}}$  with an initial population size of  $\lambda^s$  and compare it with the  $\text{CMA}_{\text{mah}}$  that employs an initial population size of  $\lfloor \lambda^s/2 \rfloor$  on the noiseless BBOB test bed [6].

The algorithms are described in more detail in Sec. 2. Section 3 gives the mandatory results of the BBOB timing experiments while Sec. 4 presents the general results of the comparison. Section 5 concludes the paper.

## 2. TESTED CMA-ES VARIANTS

We tested two variants of the IPOP-CMA-ES with active covariance matrix adaptation and mirrored mutations: the  $\text{CMA}_{\text{ma}}$  with standard initial population size  $\lambda^s = 4 + \lfloor 3 \log(D) \rfloor$  and the  $\text{CMA}_{\text{mah}}$  with reduced initial population size  $\lfloor \lambda^s/2 \rfloor$ . Both implementations can be downloaded from <http://canadafrance.gforge.inria.fr/mirroring/> in the used version 3.54.beta.mirrors. Besides the difference in the initial population size, the number of restarts is increased to 10 for the  $\text{CMA}_{\text{mah}}$  instead of 9 for the  $\text{CMA}_{\text{ma}}$  to allow the final restarts of both algorithms to operate with the same (range of) population size. All other parameters are equal for the two algorithms and, besides  $2 \cdot 10^5 \cdot D$  as the maximal number of function evaluations, chosen according to the standard recommendations for the CMA-ES. For

more details of the algorithm, see also the accompanying paper [3].

### 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 CMA<sub>ma</sub> and CMA<sub>mah</sub> 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  $2 \cdot 10^5 D$  function evaluations until 30 seconds have passed. The per-function-evaluation-runtimes were 16; 16; 11; 6.5; 4.2; 4.6; and 7.2 times  $10^{-4}$  seconds for the CMA<sub>ma</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 first observation is the fact that both algorithm variants behave quite similar with only a few cases where the differences are statistically significant. The two main exceptions are the sphere function ( $f_1$ ) for which the variant with smaller initial population size is about 25% faster in all dimensions and all difficult targets and the discus function ( $f_{11}$ ) where the variant with standard population size is about 20% faster in 5D and 10% faster in 20D (see Table 1). Figure 1 reveals a few more statistically significant differences for the target value  $10^{-8}$ : while the algorithm with standard population size is faster for several lower dimensions ( $f_2$  in 2D and 3D,  $f_{10}$  in 2D, 3D, and 10D,  $f_{13}$  in 2D, 3D, and 5D,  $f_{17}$  in 3D and 10D,  $f_{18}$  in 3D and 20D) as well as on  $f_6$  in 20D and 40D and for  $f_{14}$  in all dimensions but 20, the algorithm with reduced initial population size is sometimes faster for larger dimensions ( $f_5$  in 20D and 40D,  $f_8$  and  $f_{12}$  in 40D). Furthermore, one can observe that, in 20D, unsuccessful runs occur for eight of the 24 functions and the functions  $f_3$ ,  $f_4$ , and  $f_{19}$ – $f_{24}$  cannot be solved by both algorithms in any of the 15 runs. When compared to the best algorithm of the BBOB'2009 exercise, both algorithms significantly improve the performance on  $f_{10}$  (faster by a factor of 1.4),  $f_{14}$  (factor of  $\geq 1.5$ ), and  $f_{11}$  and  $f_{15}$  (factor of  $> 2$ , all results in 40D) which is mainly due to the active covariance matrix adaptation [8].

### 5. CONCLUSIONS

When investigating the impact of the initial population size in the IPOP-CMA-ES with active covariance matrix adaptation and mirrored mutation, no general recommendation towards one of the two algorithms CMA<sub>ma</sub> and CMA<sub>mah</sub> can be made. While a lower population size is generally helpful on the sphere function and less effective on the discus function, the positive effect of the lower population size is often more pronounced for larger dimensions with the exception of the attractive sector function where the opposite is the case. As a general conclusion, we remark that the change of the initial population size has overall comparatively small effects.

### 6. 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.

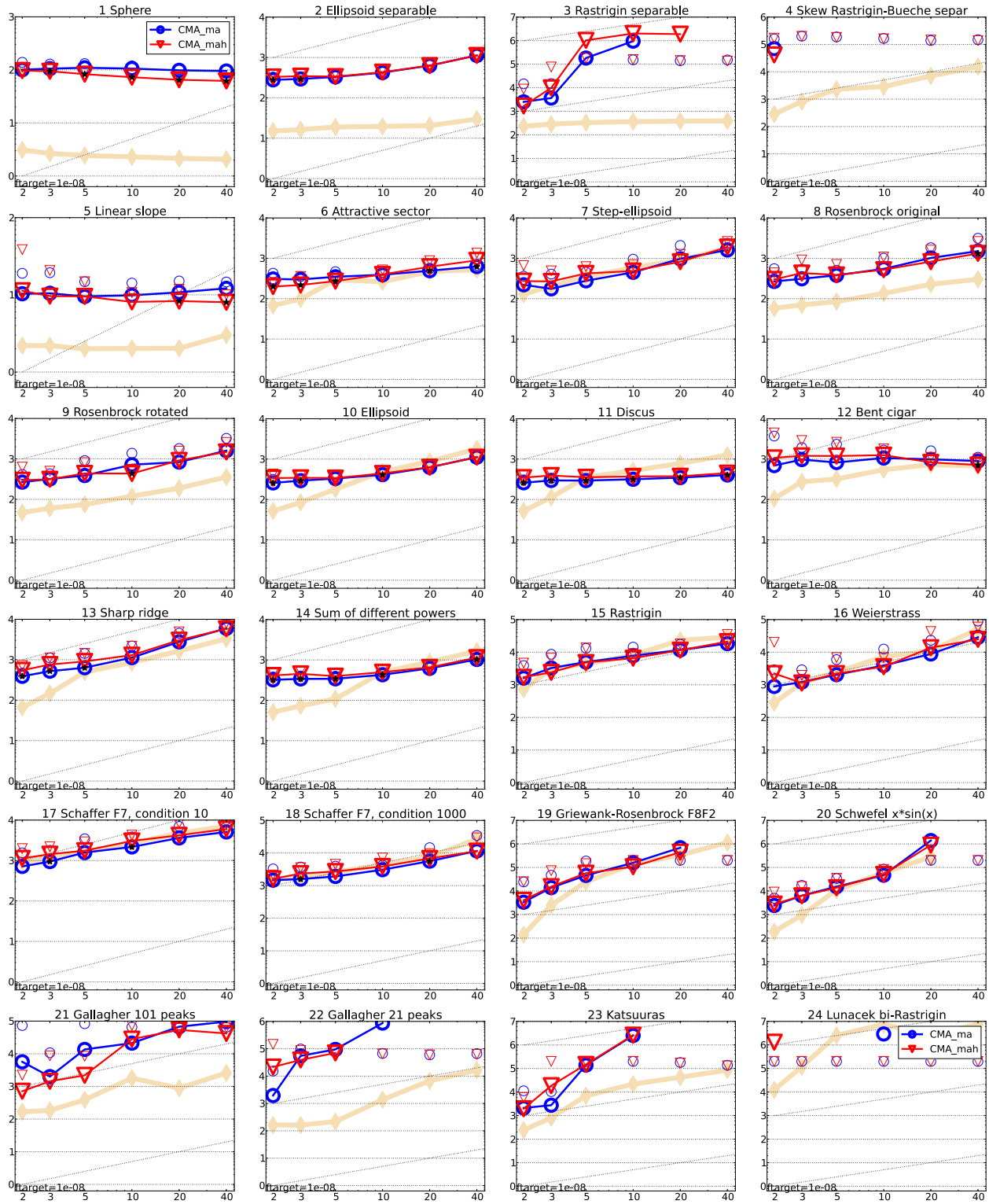


Figure 1: Expected running time (ERT in number of  $f$ -evaluations) divided by dimension for target function value  $10^{-8}$  as  $\log_{10}$  values versus dimension. Different symbols correspond to different algorithms given in the legend of  $f_1$  and  $f_{24}$ . Light symbols give the maximum 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<sub>ma</sub>,  $\nabla$ :CMA<sub>mah</sub>.

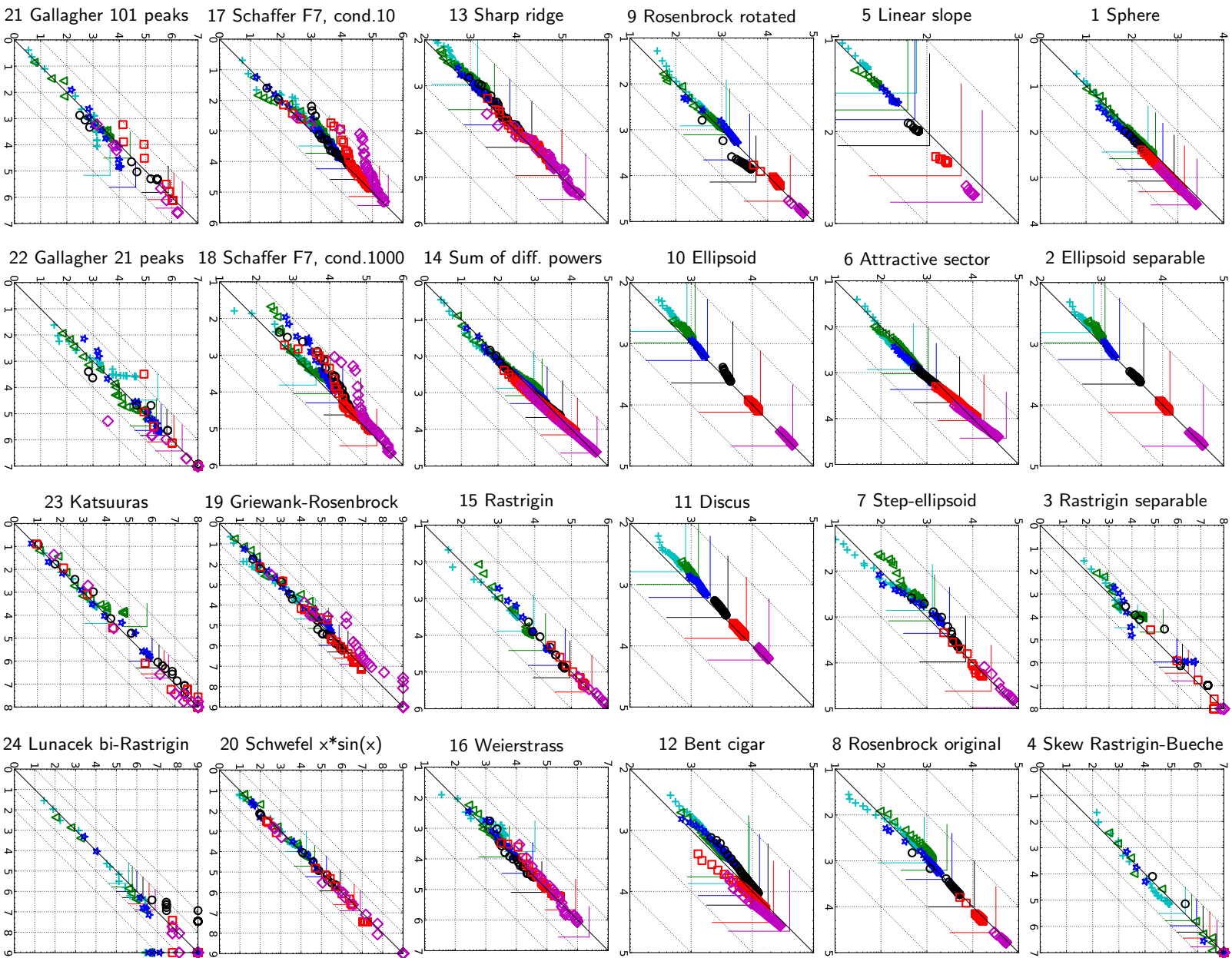


Figure 2: Expected running time (ERT in  $\log_{10}$  of number of function evaluations) of CMAma ( $x$ -axis) versus CMAmah ( $y$ -axis) for 46 target values  $\Delta f \in [10^{-8}, 10]$  in each dimension on functions  $f_1-f_{24}$ . Markers on the upper or right edge indicate that the target value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

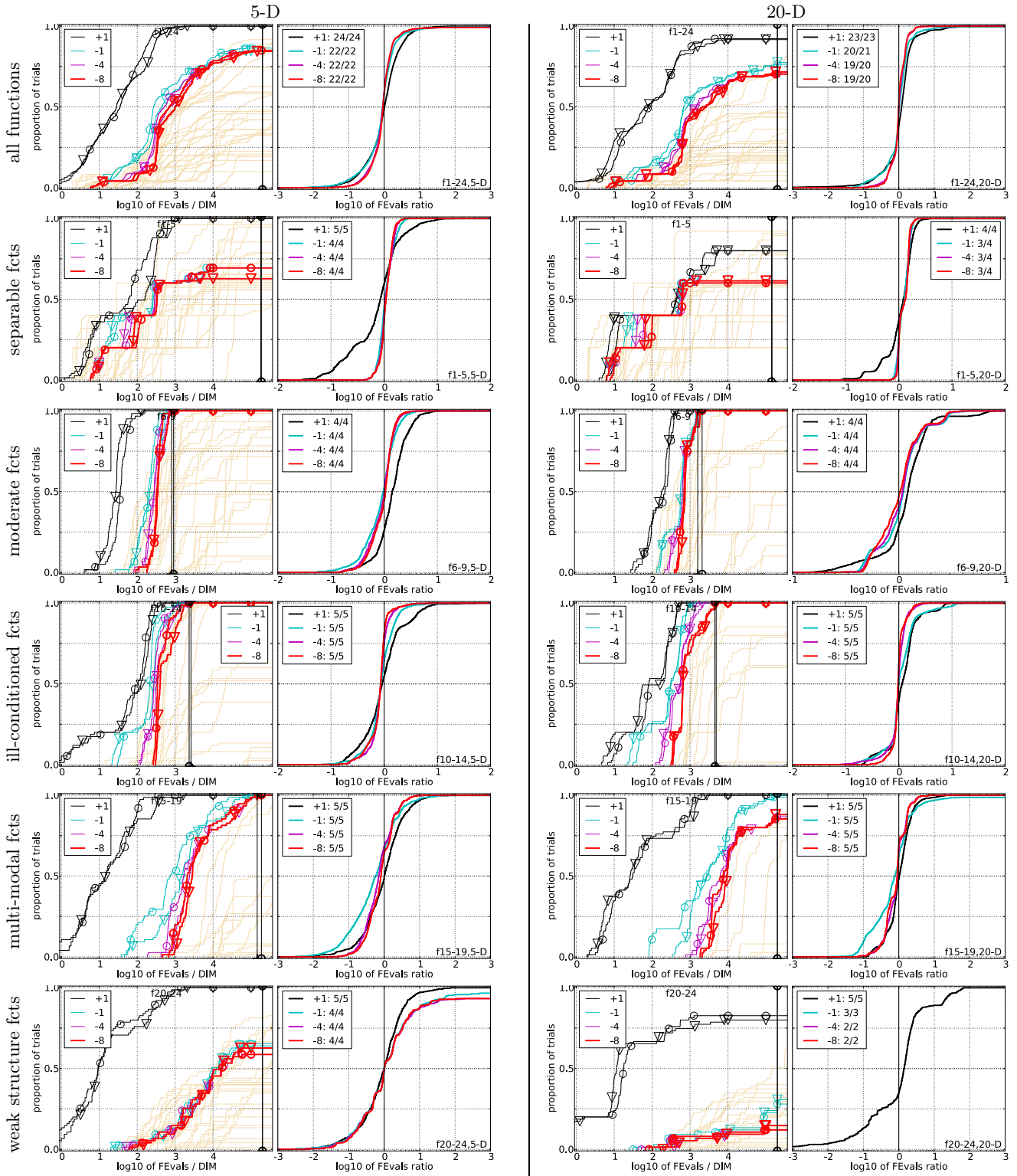


Figure 3: Empirical cumulative distributions (ECDF) of run lengths and speed-up ratios in 5-D (left) and 20-D (right). Left sub-columns: ECDF of the number of function evaluations divided by dimension  $D$  (FEvals/ $D$ ) to reach a target value  $f_{\text{opt}} + \Delta f$  with  $\Delta f = 10^k$ , where  $k \in \{1, -1, -4, -8\}$  is given by the first value in the legend, for  $\text{CMA}_{\text{ma}}$  ( $\circ$ ) and  $\text{CMA}_{\text{mah}}$  ( $\nabla$ ). 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  $\text{CMA}_{\text{ma}}$  divided by  $\text{CMA}_{\text{mah}}$ , all trial pairs for each function. Right 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 ( $\text{CMA}_{\text{ma}}$  first).

