

# On the Effect of Mirroring in the IPOP Active CMA-ES 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 and active covariance matrix adaptation are two recent ideas to improve the well-known covariance matrix adaptation evolution strategy (CMA-ES)—a state-of-the-art algorithm for numerical optimization. It turns out that both mechanisms can be implemented simultaneously. In this paper, we investigate the impact of mirrored mutations on the so-called IPOP active CMA-ES. We find that additional mirrored mutations improve the IPOP active CMA-ES statistically significantly, but by only a small margin, on several functions while never a statistically significant performance decline can be observed. Furthermore, experiments on different function instances with some algorithm parameters and stopping criteria changed reveal essentially the same results.

## 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 a state-of-the-art algorithm for numerical optimization. If after each restart, its population size is increased by a factor of two, we talk about the IPOP-CMA-ES which showed good and competitive results on a variety of test problems [2, 12] as well as in practice. A recent variant

of the standard CMA-ES is the *active CMA-ES* [10] where also “bad” steps are taken into account to shrink the covariance matrix of the search distribution along axes with inferior performance. When combined with the IPOP-CMA-ES, a substantial performance gain can be observed on the BBOB test bed [9]. Another recent algorithm development for evolution strategies is the idea of *mirrored mutations* [3, 1]. The main idea here is to use dependent, mirrored mutations instead of the standard independent ones with the additional trick to only evaluate the mirrors of solutions with bad function values (selective mirroring). This variant has shown improved performance over the standard CMA-ES in particular on the sphere function. Both ideas can, in principle, be easily combined and in this paper we investigate the impact of mirrored mutations on the IPOP active CMA-ES.

It turns out that on the one hand, the additional mirrored mutations have a positive effect on the performance of the IPOP active CMA-ES and on the other hand that small changes in the algorithm parameters and stopping criteria do not essentially change the results on the BBOB test functions.

## 2. ALGORITHM PRESENTATION

### 2.1 The IPOP-CMA-ES

The standard CMA-ES samples in each iteration  $\lambda$  normally distributed solutions  $x_i = \bar{m}_t + \sigma_t N(0, C_t)$  in the search space  $\mathbb{R}^D$ , given the current mean  $\bar{m}_t$ , step size  $\sigma_t$ , and covariance matrix  $C_t$  at time  $t$ . After evaluating the  $\lambda$  new search points on the objective function, the best  $\mu$  solutions  $x_{1:\lambda}, \dots, x_{\mu:\lambda}$  are recombined via weighted recombination to the new mean  $\bar{m}_{t+1} = \sum_{i=1}^{\mu} w_i x_{i:\lambda}$  and step size and covariance matrix are adapted via the so-called cumulative evolution path(s) [8]. The so-called IPOP-CMA-ES where IPOP stands for increasing population size, restarts the algorithm with a doubled population size when convergence is detected by some stopping criteria [2]. With an initial population size of  $\lambda^s = 4 + \lceil 3 \log(D) \rceil$ , the algorithm is quasi parameter-free.

### 2.2 The IPOP Active CMA-ES

A CMA-ES variant with negative or active covariance matrix adaptation allows negative weights in the update of the covariance matrix taking into account also the worst solutions. The main idea behind this update is “to use information about unsuccessful offspring candidate solutions in order to actively reduce variances of the mutation distri-

bution in unpromising directions of the search space” [10]. The additional active covariance adaptation has been shown to improve the performance of the IPOP-CMA-ES on the BBOB test bed [9].

### 2.3 The IPOP-CMA-ES With Mirrored Mutations

The idea of mirrored mutations has been introduced recently into weighted recombination evolution strategies [1]. The main idea is thereby the following. Instead of sampling all  $\lambda$  new solutions in iteration  $t$  of the algorithm independently, only  $\lambda_{\text{iid}}$  samples are drawn independently and identically distributed while the remaining  $\lambda_{\text{m}}$  solutions are *mirrors* of already sampled solutions. The mirror of a solution  $x_i = \bar{m}_t + \vec{d}$  (with  $\vec{d}$  distributed according to  $\sigma_t N(0, C_t)$ ) is the solution in the opposite direction of the mean  $\bar{m}_t$ , i.e., the solution  $x'_i = \bar{m}_t - \vec{d}$ . *Pairwise selection* is furthermore used to prevent a bias in the step size: at most one solution of each mirrored/unmirrored solution pair is used within the weighted recombination and to update the algorithm’s evolution path(s). Finally, with the idea of *selective mirroring*, only the worst of the  $\lambda_{\text{iid}}$  solutions are mirrored—improving the convergence rate of the algorithm significantly [1]. Here, we use selective mirroring with  $\lambda_{\text{m}} = \lfloor 0.5 + 0.159\lambda_{\text{iid}} \rfloor$  (where pairwise selection becomes superfluous), and recombination weights as in [1].

### 2.4 Combining Active Covariance Matrix Adaptation and Mirrored Mutations

The implementations of active covariance matrix adaptation and mirrored mutations are, in principle, independent of each other and can be turned on and off individually. The additional active covariance matrix adaptation is thereby performed after the standard covariance matrix adaptation where the mirrored mutations are taken into account. Combining both approaches is of conceptual advantage even on the linear and the sphere function, where the updates tend to cancel each other out: if we denote a bad solution as  $x = \bar{m}_t + \sigma_t N$  and its mirrored counterpart as  $x' = \bar{m}_t - \sigma_t N$ , then  $w^- N N^T$  with  $w^- < 0$  is added to the covariance matrix within active covariance matrix adaptation and  $w^+ (-N)(-N)^T = w^+ N N^T$  with  $w^+ > 0$  is added to the covariance matrix during the default update, if  $-N$  turns out to be a successful direction. On the linear and the sphere function the mirrors of bad solutions tend to be successful thereby stabilizing the covariance matrix in its original shape. This is desirable in both cases.

### 2.5 Implementation Details

In the following, we compare three IPOP-CMA-ES variants: active CMA-ES as benchmarked for BBOB’2010 [9], denoted by IPOP-ACTCMA-ES, our current implementation of active CMA-ES (version 3.54.beta.mirrors), denoted as CMA<sub>a</sub>, and the CMA<sub>a</sub> with additional mirrored mutations which we denote as CMA<sub>ma</sub> in the following. The first two algorithms use different parameters and stopping criteria. The CMA<sub>a</sub> uses

- cumulation parameter  $c_c = (4 + \mu_{\text{eff}}/D)/(D + 4 + 2\mu_{\text{eff}}/D)$  instead of  $c_c = 4/(D + 4)$  for the IPOP-ACTCMA-ES,
- a maximum number of  $1000(D+5)^2/\sqrt{\lambda}$  iterations per restart instead of  $100 + 50(D + 3)^2/\sqrt{\lambda}$ ,

- TolX=2e-11 instead of TolX=2e-12
- TolHistFun=1e-13 instead of TolHistFun=1e-12
- TolUpSigma=inf (i.e., not implemented) instead of TolUpSigma=1e20, see [5]. This difference is likely to have a noticeable impact on some multimodal functions like  $f_{19}$ .

Additionally, in CMA<sub>ma</sub>, the step-size damping is reduced by  $(1 - \min(1, (\lambda_{\text{mirr}}/(0.159\lambda) - 1)^2))/2 \in [0, 1/2]$ . Since we used the original BBOB’2010 data for the IPOP-ACTCMA-ES, also the function instances differ. While for the IPOP-ACTCMA-ES, the instances 1, . . . , 15 are used, the new results for the CMA<sub>a</sub> and the CMA<sub>ma</sub>, are obtained for the BBOB’2012 instances 1, . . . , 5, 21, . . . , 30.

All algorithms are restarted maximally nine times and stopped after at most  $2 \cdot 10^5 \cdot D$  function evaluations. The implementations of the CMA<sub>a</sub> and the CMA<sub>ma</sub> 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 CMA-ES variant with active covariance matrix adaptation without mirrored mutations (CMA<sub>a</sub>) and the same algorithm with mirrored mutations (CMA<sub>ma</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 25, 18, 13, 7.9, 5.5, 5.5, and 7.4 times  $10^{-4}$  seconds for the CMA<sub>a</sub> and 16, 16, 11, 6.5, 4.2, 4.6, and 7.2 times  $10^{-4}$  seconds for the CMA<sub>ma</sub> in 2, 3, 5, 10, 20, 40, and 80 dimensions respectively. The used implementation is version 3.54.beta.mirrors and can be downloaded from <http://canadafrance.gforge.inria.fr/mirroring/>. In comparison, [9] correspondingly states 2.0, 1.7, 1.4, 1.2, 1.1, 1.4, and 3.6 times  $10^{-4}$  seconds per function evaluation on an Intel Core 2 6700 processor (2.66 GHz) with Linux 2.6.28-18 and Matlab R2008a for the 2010 implementation of the IPOP-ACTCMA-ES.

## 4. RESULTS

Results from experiments according to [6] on the benchmark functions given in [4, 7] are presented in Figures 1, 2 and 3 and in Tables 1 and 2. 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, 11]. **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.

### *Impact of Mirrored Mutations.*

Mirrored mutations show a small but consistent improvement on seven unimodal functions, where on each function the effect is significant for at least one dimensionality. The effect is however small enough that it might be mainly explained by the setting of the step-size damping parameter  $d_\sigma$  that depends on the number of mirrored mutations. Further effects of mirrored mutations are neither systematic nor significant and likely to have happened by chance. In particular, mirrored mutations seem not to be harmful on any of the 24 test functions.

### *Impact of Parameter Setting and Instance Numbers.*

We find a statistically significant difference between CMA<sub>a</sub> and IPOP-ACTCMA-ES on  $f_{13}$  in 2D, on  $f_{16}$  in 40D, on  $f_{17}$  in 10D, and on  $f_{19}$  in 5D. Only the last two effects are quantitatively relevant and only in the last case, IPOP-ACTCMA-ES outperforms CMA<sub>a</sub>. Furthermore is on  $f_{19}$  and  $f_{20}$  the success rate in 20-D significantly higher with IPOP-ACTCMA-ES. No other statistically significant differences are observed (the test results not shown). The improvement of IPOP-ACTCMA-ES on  $f_{19}$ , where IPOP-ACTCMA-ES seems to be consistently better than CMA<sub>a</sub> and on  $f_{20}$  (and possibly on  $f_{22}$ ) is likely to be caused by the To1UpSigma termination. The other parameter differences seem to have minor or no effect.

## 5. SUMMARY

Two main statements have been supported by empirical results on the noiseless BBOB testbed. First, additional mirrored mutations improve the performance of the IPOP active CMA-ES on several functions by a small margin while no detrimental effects can be observed. Second, changes of some parameters and stopping criteria in IPOP active CMA-ES do only have a comparatively small effect on the results and are only observable when restarts are necessary to solve the function.

## 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, 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.
- [4] 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.
- [5] N. Hansen. Benchmarking a BI-population CMA-ES on the BBOB-2009 function testbed. In F. Rothlauf, editor, *GECCO (Companion)*, pages 2389–2396. ACM, 2009.
- [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 A. Ostermeier. Completely Derandomized Self-Adaptation in Evolution Strategies. *Evolutionary Computation*, 9(2):159–195, 2001.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] R. Ros. Black-box optimization benchmarking the IPOP-CMA-ES on the noiseless testbed: comparison to the BIPOP-CMA-ES. In *Genetic and Evolutionary Computation Conference (GECCO 2010)*, pages 1503–1510, New York, NY, USA, 2010. ACM.

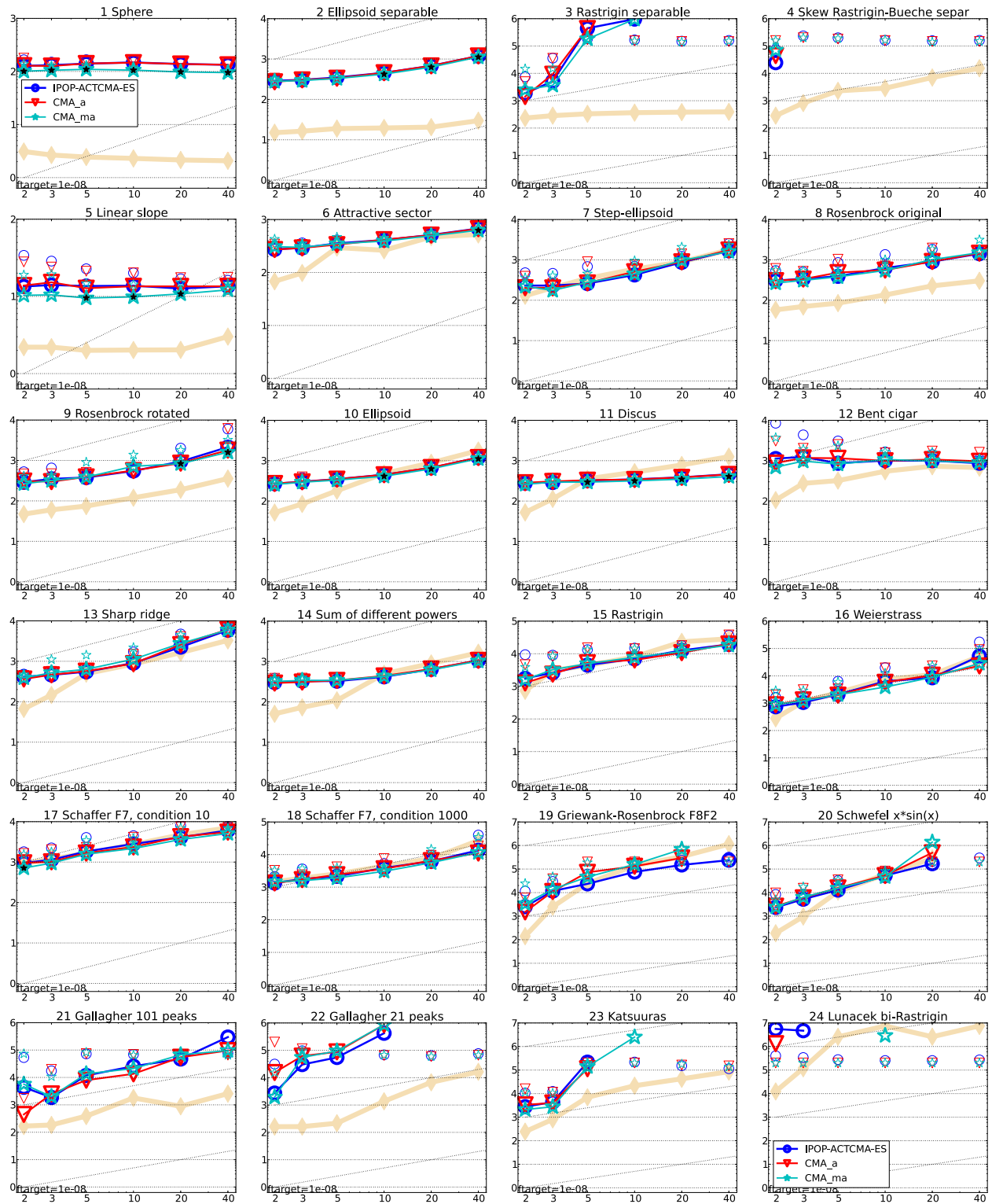


Figure 1: Expected running time (ERT in number of  $f$ -evaluations) divided by dimension for target function value  $10^{-8}$  as  $\log_{10}$  values versus dimension. Different symbols correspond to different algorithms given in the legend of  $f_1$  and  $f_{24}$ . Light symbols give the maximum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

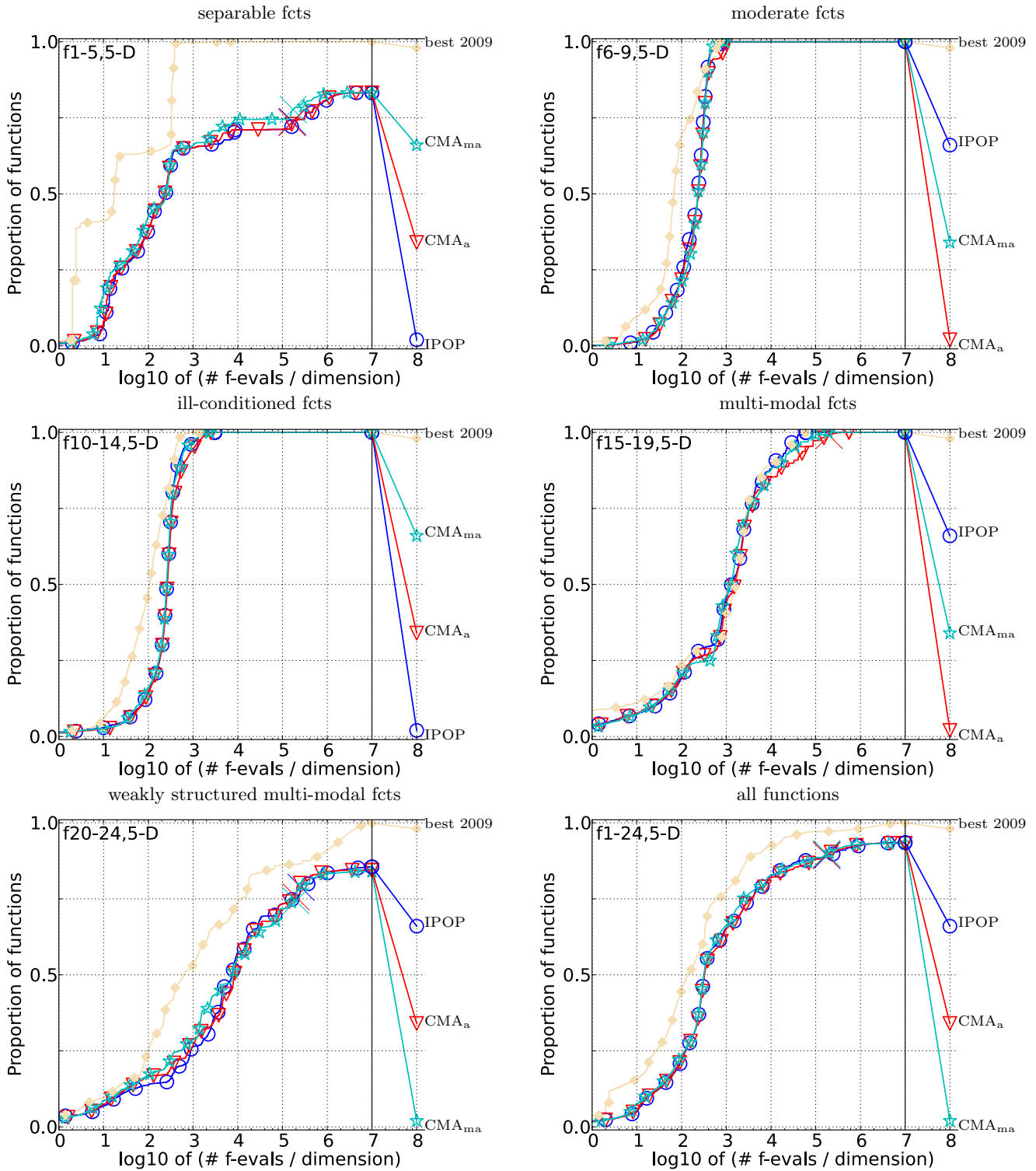


Figure 2: Bootstrapped empirical cumulative distribution of the number of objective function evaluations divided by dimension (FEvals/D) for 50 targets in  $10^{[-8..2]}$  for all functions and subgroups in 5-D. The “best 2009” line corresponds to the best ERT observed during BBOB 2009 for each single target.

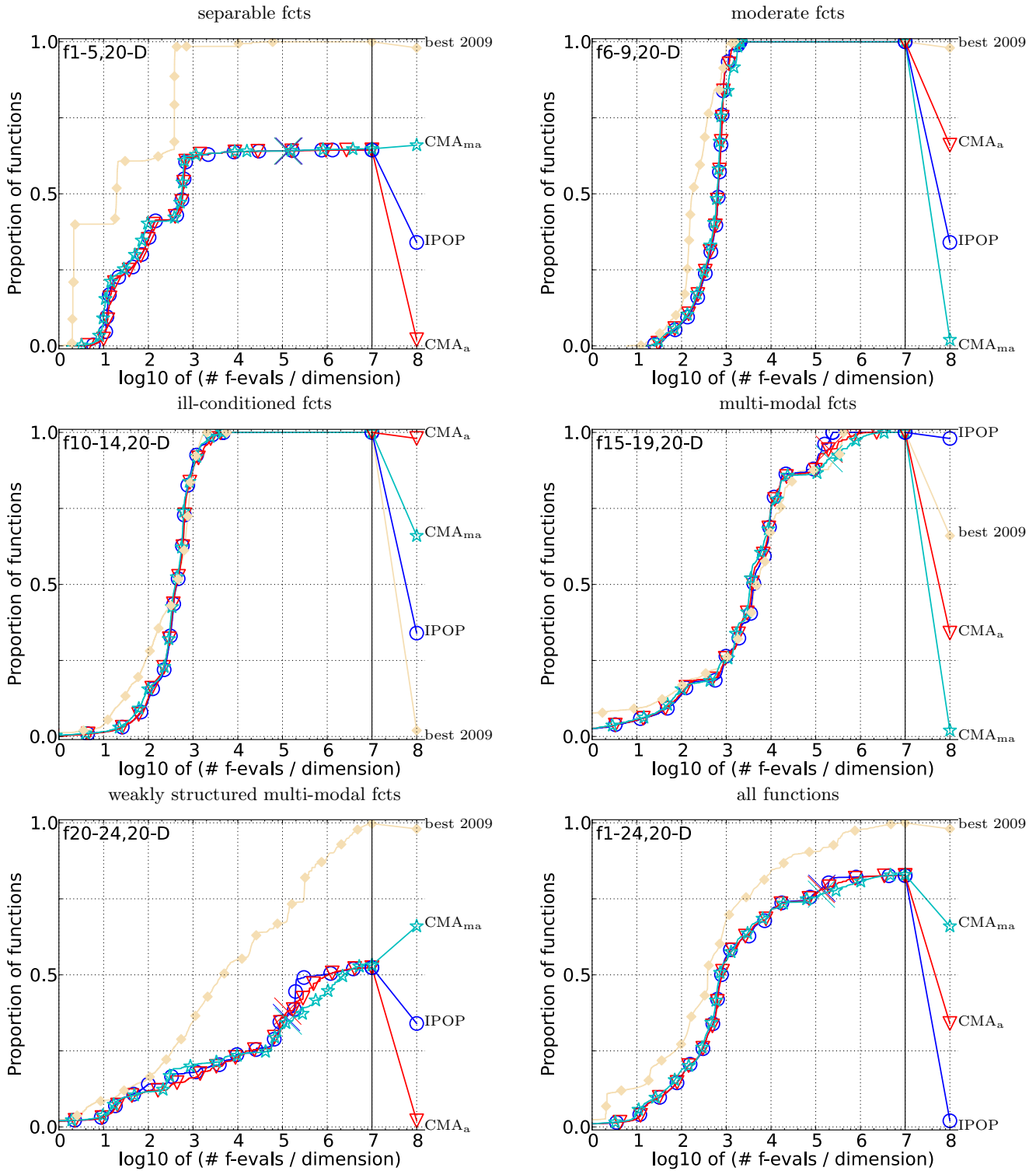


Figure 3: Bootstrapped empirical cumulative distribution of the number of objective function evaluations divided by dimension (FEvals/D) for 50 targets in  $10^{[-8..2]}$  for all functions and subgroups in 20-D. The “best 2009” line corresponds to the best ERT observed during BBOB 2009 for each single target.



