

Mirrored Variants of the (1,4)-CMA-ES Compared on the Noisy BBOB-2010 Testbed

[Black-Box Optimization Benchmarking Workshop]

Anne Auger, Dimo Brockhoff, and Nikolaus Hansen
Projet TAO, INRIA Saclay—Ile-de-France
LRI, Bât 490, Univ. Paris-Sud
91405 Orsay Cedex, France
firstname.lastname@inria.fr

ABSTRACT

Derandomization by means of mirrored samples has been recently introduced to enhance the performances of $(1, \lambda)$ - and $(1 + 2)$ -Evolution-Strategies (ESs) with the aim of designing fast stochastic local search algorithms. In this paper, we investigate the impact of mirrored samples for noisy optimization. Since elitist selection is detrimental for noisy optimization, we investigate non-elitist ESs only here. We compare on the BBOB-2010 noisy benchmark testbed two variants of the (1,4)-CMA-ES where mirrored samples are implemented with the baseline (1,4)-CMA-ES. Each algorithm implements a restart mechanism. A total budget of $10^4 D$ function evaluations per trial has been used, where D is the dimension of the search space.

The comparison shows that using mirroring within the (1,4)-CMA-ES improves the performance in the noisy BBOB-2010 scenario: the $(1,4_m)$ -CMA-ES with mirrored mutations improves significantly over the (1,4)-CMA-ES by 13–60% on 6 functions whereas no function with decreased performance can be reported. The $(1,4_m^s)$ -CMA-ES, employing in addition to the mirroring a sequential selection, further improves the results over the $(1,4_m)$ -CMA-ES by additional 20–62%, depending on the function. Compared to the BBOB-2009 benchmarking, the $(1,4_m^s)$ -CMA-ES improves over the function-wise best algorithm on 7 functions with Cauchy noise type by 12–68% (in both 5D and 20D).

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

Evolution Strategies (ESs) are robust stochastic search algorithms for black-box optimization where the function to be minimized, f , maps the continuous search space \mathbb{R}^D into \mathbb{R} . ESs evolve a population of candidate solutions that are created by sampling λ independent random vectors following a multivariate normal distribution. Recently, a new derandomization technique replacing the independent sampling of new solutions (or offspring) by mirrored samples has been introduced to enhance the performances of ESs [1]. With *mirrored sampling*, a single sample \mathcal{N} of a multivariate normal distribution is used for two offspring of the same iteration. Denoting X the current solution, the two offspring will equal $X + \mathcal{N}$ and $X - \mathcal{N}$ respectively. The resulting offspring are thus symmetric or *mirrored* with respect to X and are thus *negatively correlated*. Mirrored samples have been implemented in the Covariance-Matrix-Adaptation Evolution-Strategy (CMA-ES), an ES whose characteristic is to adapt the full covariance matrix of the multivariate normal search distribution [7]. Another new concept called sequential selection was introduced together with mirrored samples [1]. *Sequential selection*, consists in performing sequential evaluations of the offspring and breaking the evaluation loop as soon as an offspring is better than the current solution X and thus saving the remaining fitness evaluations.

In this paper, we assess quantitatively the improvement that can be brought by mirrored samples and by mirrored samples coupled with sequential selection. We compare on the BBOB-2010 noisy testbed the (1,4)-CMA-ES with two variants: first the $(1,4_m)$ -CMA-ES where mirrored samples are used, and second the $(1,4_m^s)$ -CMA-ES that, in addition to the mirrored samples, uses sequential selection. The algorithms and the CPU timing experiments are described in a complementing paper in the same proceedings [2].

2. RESULTS

2.1 Comparing (1,4)- and $(1,4_m)$ -CMA-ES

Results from experiments comparing (1,4)-CMA-ES and $(1,4_m)$ -CMA-ES according to [5] on the benchmark functions given in [4, 6] are presented in Figures 1 and 2 and in Table 1. The **expected running time (ERT)**, used in the figures

and table, depends on a given target function value, $f_t = f_{\text{opt}} + \Delta f_t$, 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 [5, 8]. **Statistical significance** is tested with the rank-sum test for a given target Δf_t using, for each trial, either the number of needed function evaluations to reach Δf_t (inverted and multiplied by -1), or, if the target was not reached, the best Δf -value achieved, measured only up to the smallest number of overall function evaluations for any unsuccessful trial under consideration.

The $(1,4_m)$ -CMA-ES with mirrored mutations clearly outperforms the $(1,4)$ -CMA-ES. There is no function where the $(1,4_m)$ -CMA-ES shows worse results than the $(1,4)$ -CMA-ES, but on all 9 functions, that can be solved by the two algorithms in 20D, the $(1,4_m)$ -CMA-ES shows lower expected running times, of which 6 differences are statistically significant: the $(1,4_m)$ -CMA-ES is about 20% faster on f_{101} and on f_{102} , 13% faster on f_{103} , 25% better on f_{109} , 60% faster on f_{118} , and about 60% faster on f_{121} . On four functions, the $(1,4_m)$ -CMA-ES shows the same or even better results than the best algorithm of the BBOB-2009 benchmarking on these functions [3] (f_{103} , f_{106} , f_{109} , and f_{130}) with the most remarkable improvement of 41% on the Gallagher function with Cauchy noise (f_{130}) where also the $(1,4)$ -CMA-ES improves over the function-wise best algorithm of BBOB-2009 by 37%. Note that all functions where the BBOB-2009 function-wise best algorithm is beaten comprise an underlying Cauchy noise.

2.2 Comparing $(1,4_m)$ - and $(1,4_m^s)$ -CMA-ES

The results of this comparison can be found in Fig. 3 and 4 and in Table 2 and show an even better performance if mirroring is combined with the sequential selection: the $(1,4_m^s)$ -CMA-ES is in 20D on all 9 functions that are solved better than the $(1,4_m)$ -CMA-ES where the differences are for 8 functions statistically significant: about 20% improvement on f_{101} and f_{102} , 27% on f_{103} , 38% on f_{106} , 43% on f_{109} , 62% on f_{112} , 50% on f_{118} , and 56% on f_{121} . These improvements also result in much better expected running times than the best algorithm of the BBOB-2009 benchmarking for those functions [3], in particular on f_{103} (37% better in 20D), f_{106} and f_{118} ($\geq 33\%$ better in 5D and 20D), f_{121} ($\geq 44\%$ better in 5D and 20D), f_{109} and f_{112} (30% better in 5D and 50% better in 20D), and f_{130} (about 68% better in 5D and 20D), see Table 2.

2.3 Comparing $(1,4)$ - and $(1,4_m^s)$ -CMA-ES

Due to space limitations, we refrain from showing the plots and tables of this comparison which are not too informative—keeping in mind that the $(1,4_m^s)$ -CMA-ES already outperforms the $(1,4_m)$ -CMA-ES. Here, the $(1,4_m^s)$ -CMA-ES, in 20D, improves over the $(1,4)$ -CMA-ES on all 9 functions that are solved of which 8 show high statistical significance (with a p -value of at least 0.01) with a large improvement factor ranging from about 35% improvement on $f_{101-103}$, to improvement factors of 2.4 (f_{106} and f_{109}), 3.3 (f_{118}), and more than 4 (f_{121} and f_{112}).

3. CONCLUSIONS

The idea behind derandomization by means of mirroring introduced in [1] is to use only one random sample from a

multivariate normal distribution to create two (negatively correlated or *mirrored*) offspring. Thereby, one offspring is generated by adding a random sample to the parent solution and a second offspring then equals the solution which is symmetric to the first one with respect to the parent (by adding the negative sample to the parent). Here, this concept of mirroring has been integrated within two variants of a simple $(1,4)$ -CMA-ES (of which the $(1,4_m^s)$ -CMA-ES uses sequential selection [1] in addition and the $(1,4_m)$ -CMA-ES does not). The three algorithms are then compared on the noisy BBOB-2010 testbed.

Using mirroring within the $(1,4_m)$ -CMA-ES improves significantly over the $(1,4)$ -CMA-ES by 13–60% on 6 functions whereas no function with decreased performance can be reported. Employing, in addition, sequential selection within the $(1,4_m^s)$ -CMA-ES further improves the results over the $(1,4_m)$ -CMA-ES by 20–62% on 8 functions in 20D. Important to note is that the $(1,4_m^s)$ -CMA-ES beats the expected running time of the function-wise best algorithm of the BBOB-2009 benchmarking by 30–68% on 7 functions which are all functions comprising Cauchy noise.

Acknowledgments

This work receives support by the French national research agency (ANR) within the SYSCOMM project ANR-08-SYSC-017 and within the COSINUS project ANR-08-COSI-007-12.

4. REFERENCES

- [1] A. Auger, D. Brockhoff, and N. Hansen. Mirrored sampling and sequential selection for evolution strategies. Rapport de Recherche RR-7249, INRIA Saclay—Île-de-France, April 2010.
- [2] A. Auger, D. Brockhoff, and N. Hansen. Mirrored variants of the $(1,4)$ -CMA-ES compared on the noiseless BBOB-2010 testbed. In *GECCO (Companion)*, 2010.
- [3] A. Auger, S. Finck, N. Hansen, and R. Ros. BBOB 2009: Comparison tables of all algorithms on all noisy functions. Technical Report RT-0384, INRIA, 04 2010.
- [4] S. Finck, N. Hansen, R. Ros, and A. Auger. Real-parameter black-box optimization benchmarking 2010: Presentation of the noisy functions. Technical Report 2009/21, Research Center PPE, 2010.
- [5] N. Hansen, A. Auger, S. Finck, and R. Ros. Real-parameter black-box optimization benchmarking 2010: Experimental setup. Technical Report RR-7215, INRIA, 2010.
- [6] N. Hansen, S. Finck, R. Ros, and A. Auger. Real-parameter black-box optimization benchmarking 2009: Noisy functions definitions. Technical Report RR-6869, INRIA, 2009. Updated February 2010.
- [7] N. Hansen, S. D. Müller, and P. Koumoutsakos. Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation. *Evolutionary Computation*, 11(1):1–18, 2003.
- [8] K. Price. Differential evolution vs. the functions of the second ICEO. In *Proceedings of the IEEE International Congress on Evolutionary Computation*, pages 153–157, 1997.

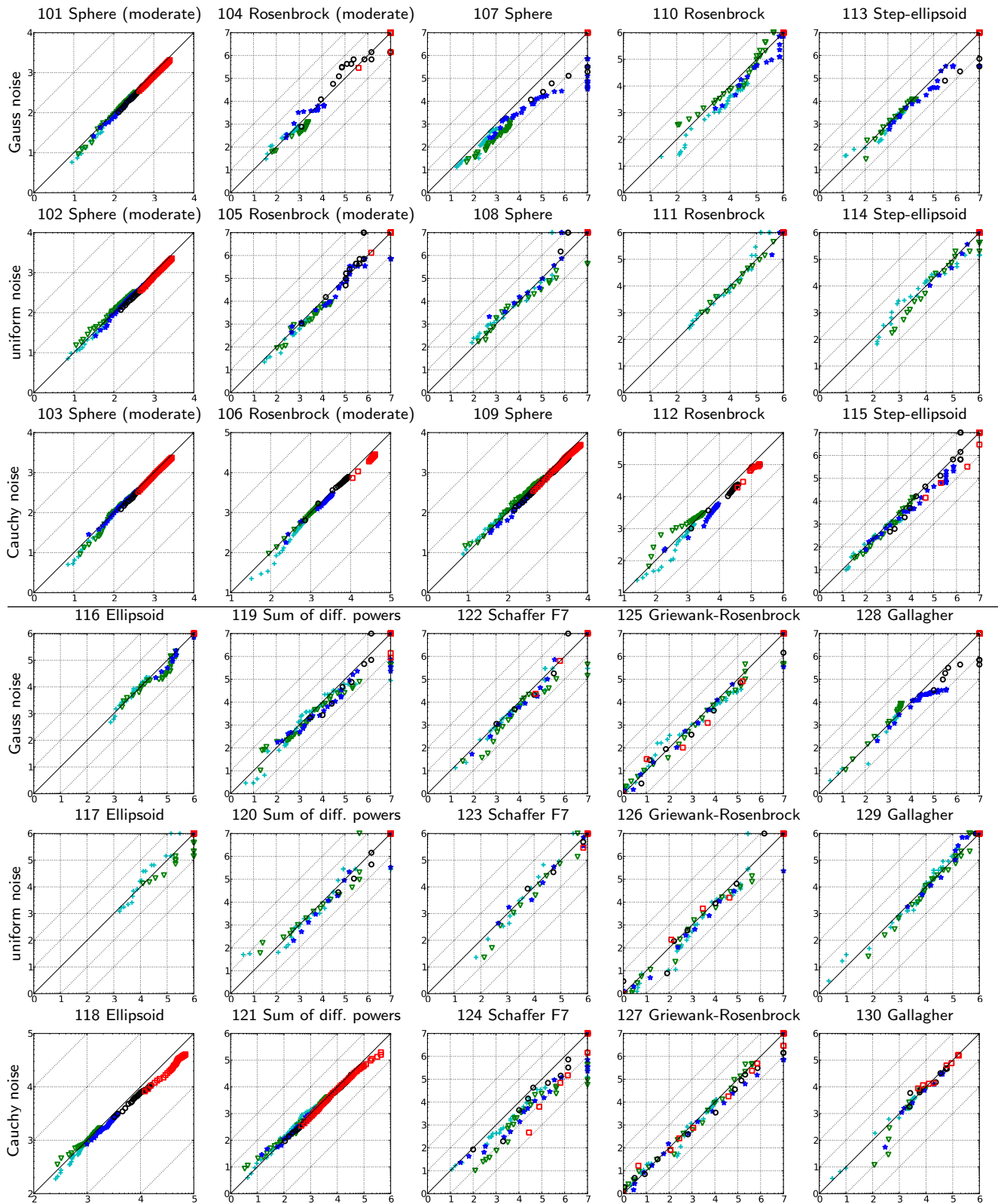


Figure 1: Expected running time (ERT in log10 of number of function evaluations) of $(1,4_m)$ -CMA-ES versus $(1,4)$ -CMA-ES for 46 target values $\Delta f \in [10^{-8}, 10]$ in each dimension for functions f_{101} - f_{130} . Markers on the upper or right edge indicate that the target value was never reached by $(1,4_m)$ -CMA-ES or $(1,4)$ -CMA-ES respectively. Markers represent dimension: 2: +, 3: ∇, 5: *, 10: o, 20: □.

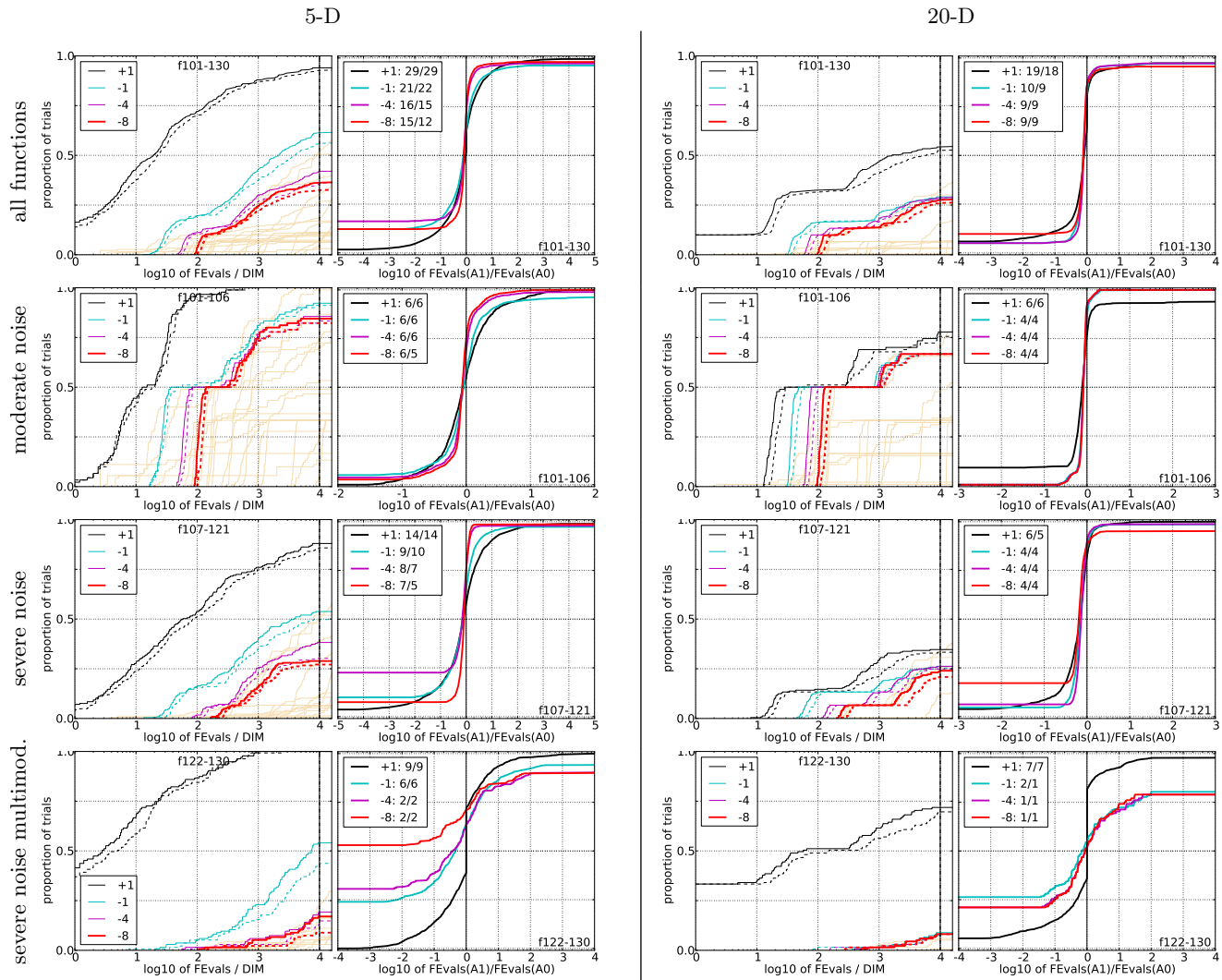


Figure 2: Empirical cumulative distributions (ECDF) of run lengths and speed-up ratios in 5-D (left) and 20-D (right). Left sub-columns: ECDF of the number of necessary function evaluations divided by dimension D (FEvals/D) to reached a target value $f_{\text{opt}} + \Delta f$ with $\Delta f = 10^k$, where $k \in \{1, -1, -4, -8\}$ is given by the first value in the legend, for $(1,4_m)$ -CMA-ES (solid) and $(1,4)$ -CMA-ES (dashed). 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 $(1,4_m)$ -CMA-ES divided by $(1,4)$ -CMA-ES, 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 ($(1,4_m)$ -CMA-ES first).

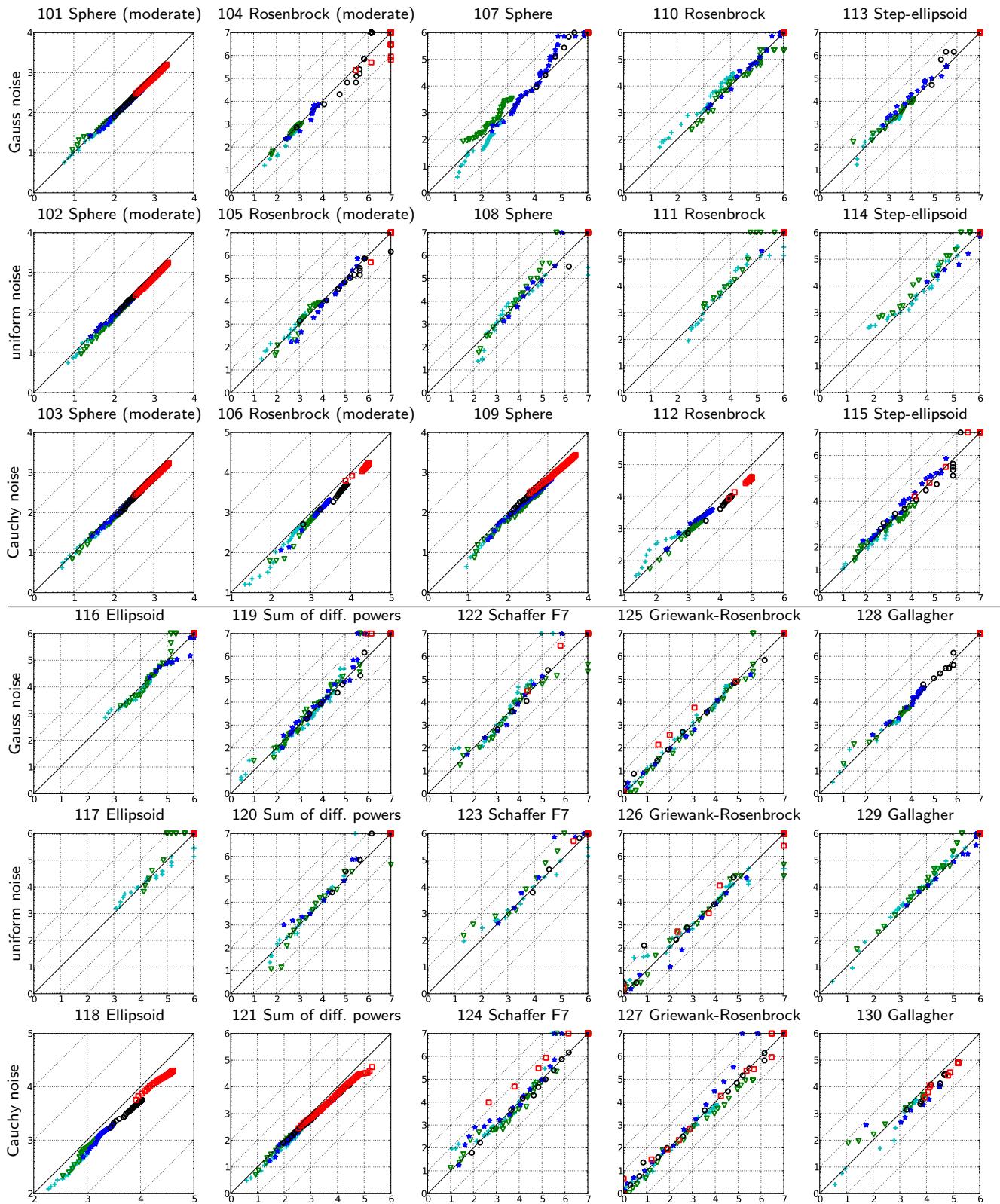


Figure 3: Expected running time (ERT in log10 of number of function evaluations) of $(1,4_m^s)$ -CMA-ES versus $(1,4_m)$ -CMA-ES for 46 target values $\Delta f \in [10^{-8}, 10]$ in each dimension $f_{101}-f_{130}$. Markers on the upper or right edge indicate that the target value was never reached by $(1,4_m^s)$ -CMA-ES or $(1,4_m)$ -CMA-ES respectively. Markers represent dimension: 2: +, 3: ∇, 5: *, 10: ○, 20: □.

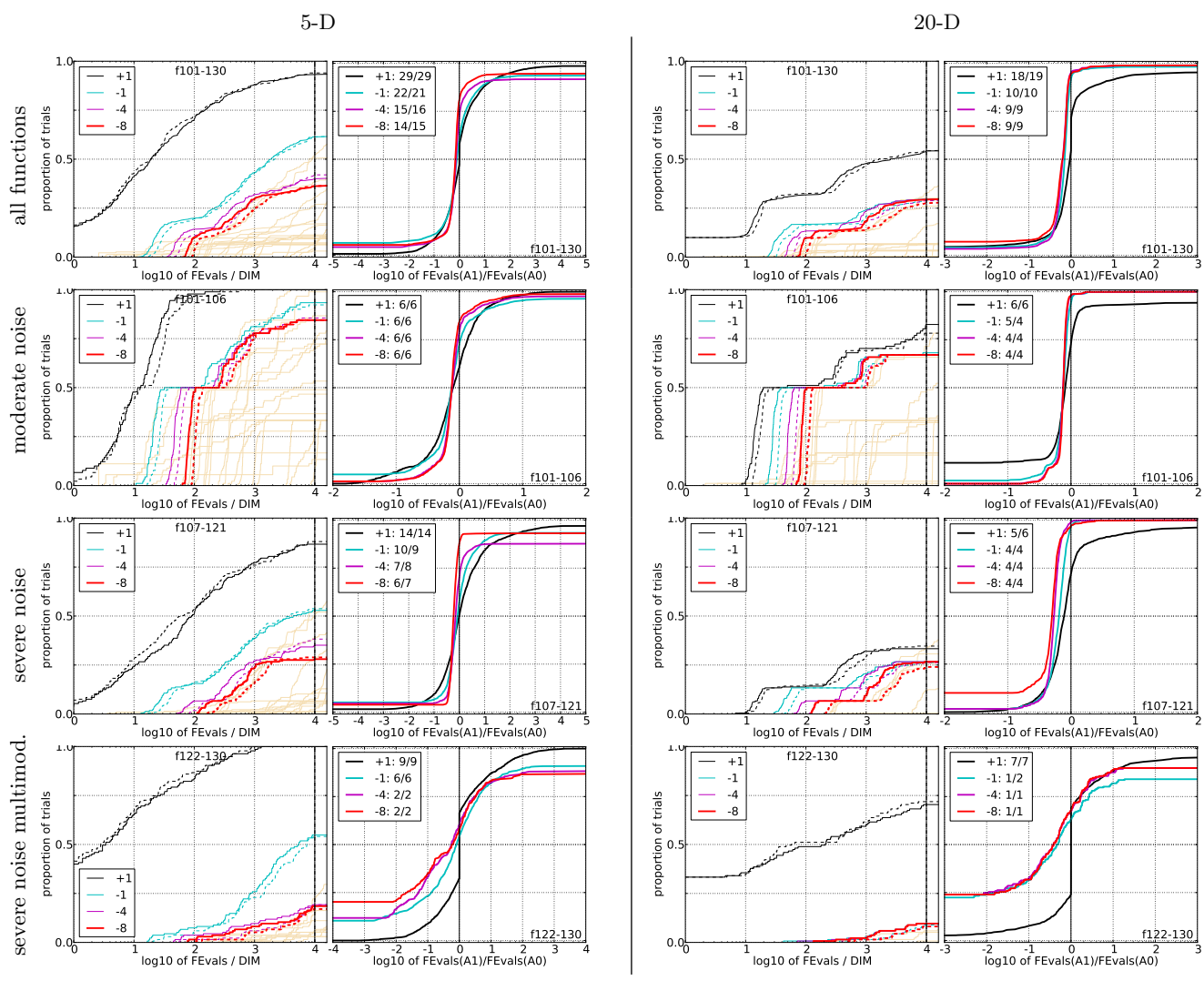


Figure 4: Empirical cumulative distributions (ECDF) of run lengths and speed-up ratios in 5-D (left) and 20-D (right) as in Fig.2 but now for $(1,4_m^s)$ -CMA-ES (solid) and $(1,4_m)$ -CMA-ES (dashed) and ratios of $(1,4_m^s)$ -CMA-ES divided by $(1,4_m)$ -CMA-ES respectively.

