



**HAL**  
open science

# Black-Box Optimization Benchmarking the IPOP-CMA-ES on the Noisy Testbed

Raymond Ros

► **To cite this version:**

Raymond Ros. Black-Box Optimization Benchmarking the IPOP-CMA-ES on the Noisy Testbed. Genetic and Evolutionary Computation Conference 2010, Jul 2010, Portland, OR, United States. inria-00473778

**HAL Id: inria-00473778**

**<https://inria.hal.science/inria-00473778>**

Submitted on 16 Apr 2010

**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.

# Black-Box Optimization Benchmarking the IPOP-CMA-ES on the Noisy Testbed

## Comparison to the BIPOP-CMA-ES

Raymond Ros  
TAO Team-Project – INRIA Saclay  
LRI, Bat 490, Univ. Paris-Sud  
F-91405 Orsay Cedex, France  
raymond.ros@inria.fr

### ABSTRACT

We benchmark the IPOP-CMA-ES on the noisy testbed of the BBOB 2010 workshop. The performances of the IPOP-CMA-ES are compared to those of the BIPOP-CMA-ES. Both algorithms are shown to perform comparably on the BBOB noisy testbed.

### 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, Evolution strategies

## 1. ALGORITHM PRESENTATION

The algorithm Covariance Matrix Adaptation-Evolution Strategy (CMA-ES) [9] is a stochastic search method based on a population. We choose to apply the  $(\mu/\mu_w, \lambda)$ -CMA-ES [3, 7, 8] in this paper. The Increasing POPulation-size (IPOP) restart policy was proposed for the CMA-ES in [1]. The resulting IPOP-CMA-ES algorithm uses a population doubling in size at each restarts.

We compare the performances of the IPOP-CMA-ES to those of the BIPOP-CMA-ES [4] which was proposed to the BBOB 2009 workshop. The BIPOP-CMA-ES distributes the allocated budget—number of function evaluations—between a doubling population size and a small population size policy. The BIPOP-CMA-ES showed good performances on the function testbeds of the BBOB 2009 workshop.

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'10, July 7–11, 2010, Portland, Oregon, USA.

Copyright 2010 ACM 978-1-4503-0073-5/10/07 ...\$10.00.

## 2. EXPERIMENTAL PROCEDURE

The IPOP-CMA-ES was tested using the same experimental set-up as that of the BIPOP-CMA-ES [4] tested for the BBOB 2009 workshop. In particular parameter  $c_1$  and  $c_\mu$ , terms of the learning rate  $c_{cov}$  are set to one fifth of the values that used for the noiseless testbed.

The only difference with the BIPOP-CMA-ES is that all the budget in terms of number of function evaluations is allocated to the doubling population size policy.

The crafting effort for IPOP-CMA-ES [5] computes to CrE = 0.

## 3. RESULTS

Results of the CPU timing experiment are given in the paper benchmarking IPOP-CMA-ES on the noiseless testbed.

Results from experiments according to [5] on the benchmark functions given in [2, 6] are presented in Figures 1, 2, 3 and 4 and in Tables 1 and 2. The **expected running time (ERT)**, used in the figures and tables, depends on a given target function value,  $f_t = f_{opt} + \Delta f_t$ , 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, 10]. **Statistical significance** is tested with the rank-sum test for a given target  $\Delta f_t$  ( $10^{-8}$  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 performances of both IPOP-CMA-ES and BIPOP-CMA-ES are pretty close with IPOP-CMA-ES being slightly faster overall but not significantly. The exception is function  $f_{117}$  (Ellipsoid function with uniform noise model) in 20-D where the IPOP-CMA-ES is significantly faster but only by a factor smaller than two.

## 4. REFERENCES

- [1] A. Auger and N. Hansen. A restart CMA evolution strategy with increasing population size. In *Proceedings of the IEEE Congress on Evolutionary Computation (CEC 2005)*, pages 1769–1776. IEEE Press, 2005.
- [2] S. Finck, N. Hansen, R. Ros, and A. Auger. Real-parameter black-box optimization benchmarking

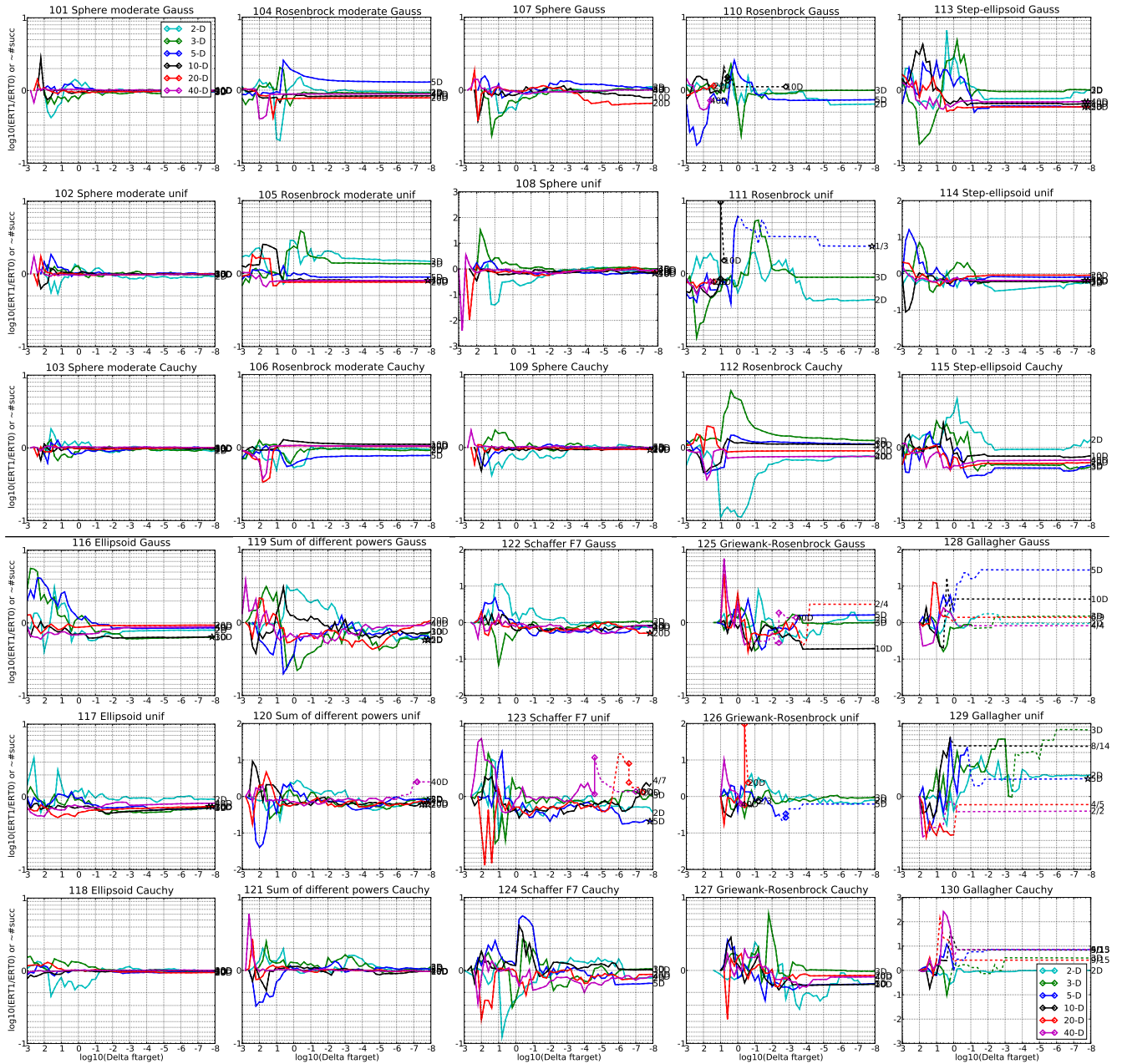


Figure 1: Ratio of the expected running times (ERT) of IPOP-CMA divided by BIPOP-CMA versus  $\log_{10}(\Delta f)$  for  $f_{101}-f_{130}$  in 2, 3, 5, 10, 20, 40-D. Ratios  $< 10^0$  indicate an advantage of IPOP-CMA, smaller values are always better. The line gets dashed when for any algorithm the ERT exceeds thrice the median of the trial-wise overall number of  $f$ -evaluations for the same algorithm on this function. Symbols indicate the best achieved  $\Delta f$ -value of one algorithm (ERT gets undefined to the right). The dashed line continues as the fraction of successful trials of the other algorithm, where 0 means 0% and the y-axis limits mean 100%, values below zero for IPOP-CMA. The line ends when no algorithm reaches  $\Delta f$  anymore. The number of successful trials is given, only if it was in  $\{1 \dots 9\}$  for IPOP-CMA (1st number) and non-zero for BIPOP-CMA (2nd number). Results are statistically significant with  $p = 0.05$  for one star and  $p = 10^{-\#\star}$  otherwise, with Bonferroni correction within each figure.

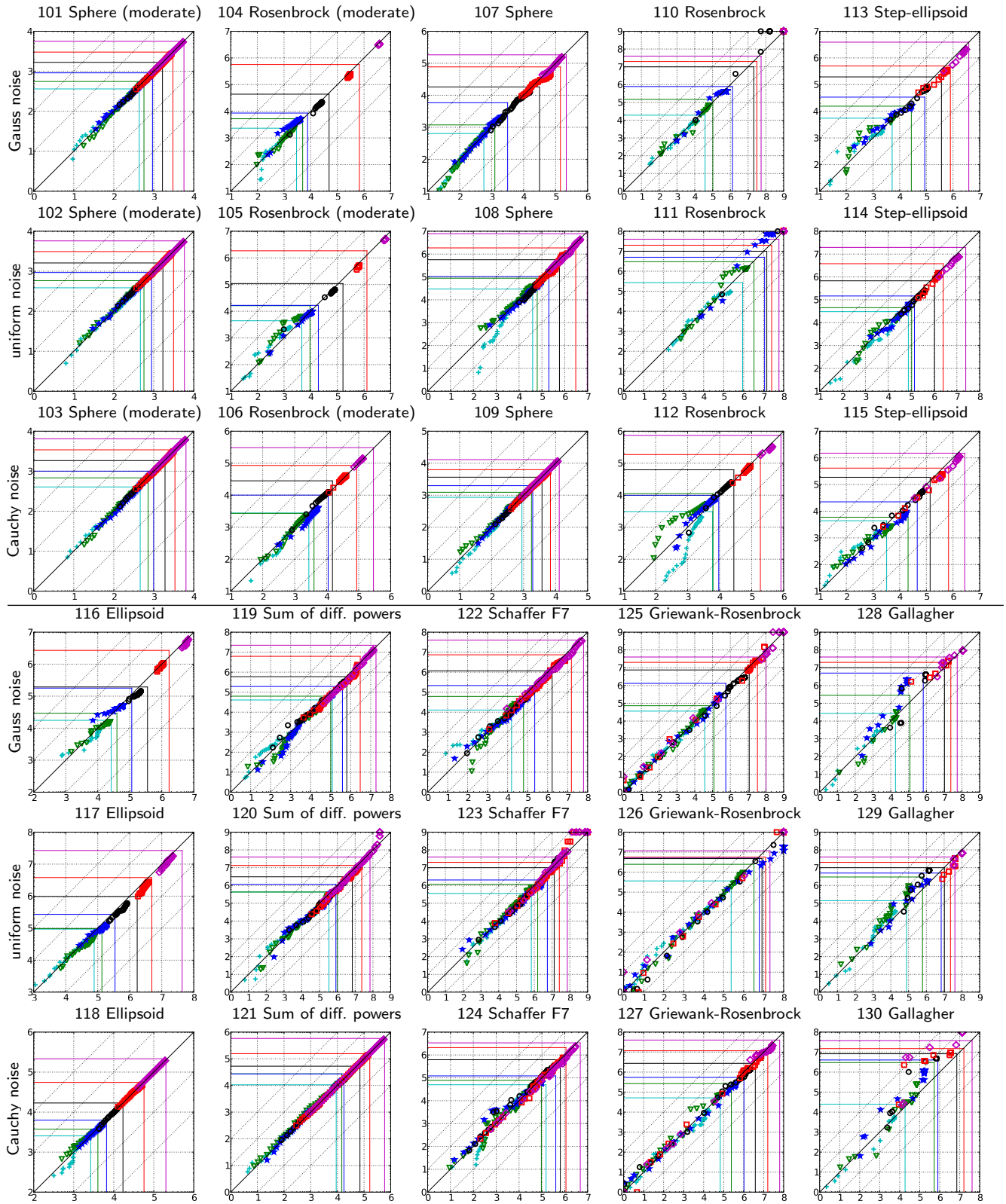
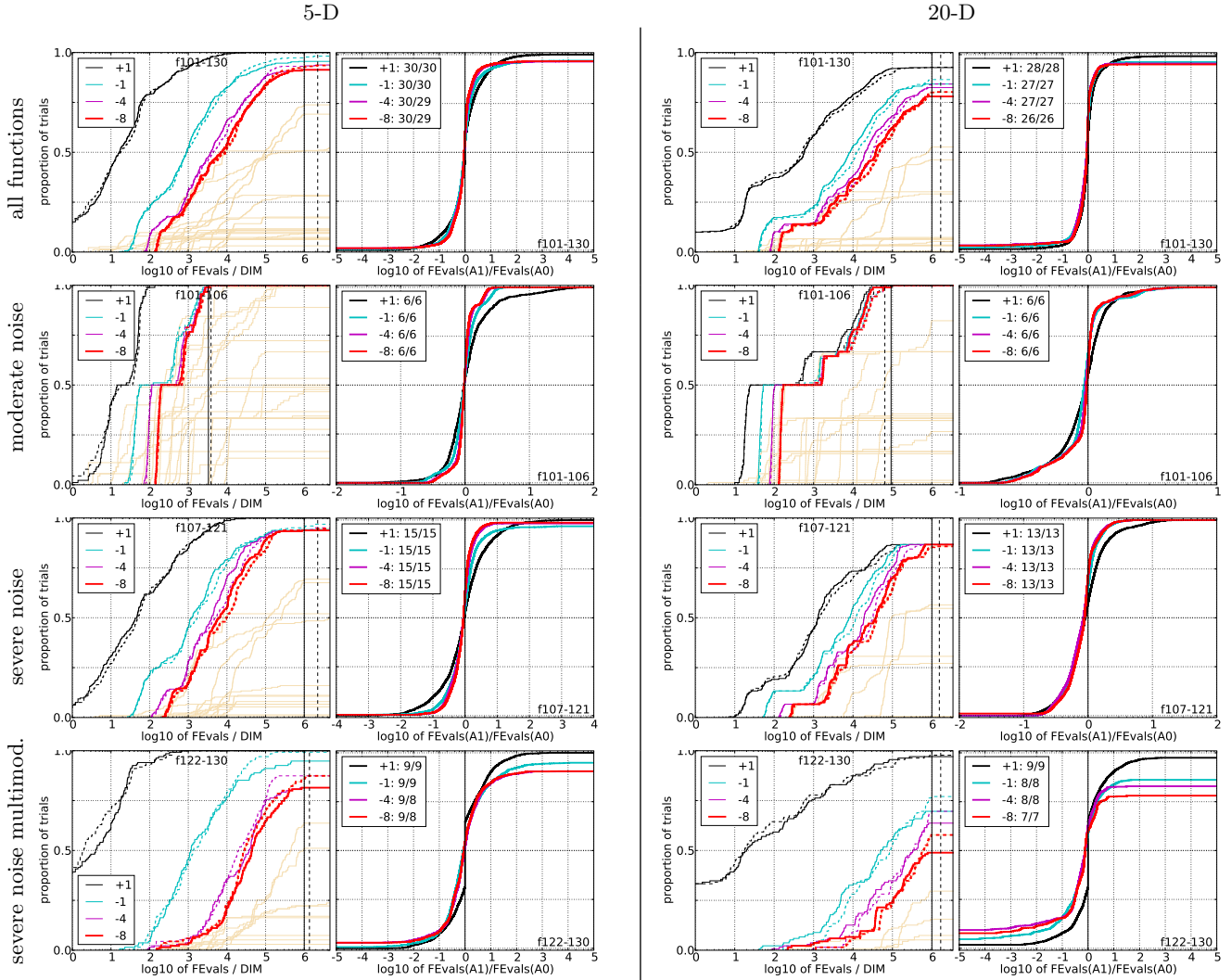
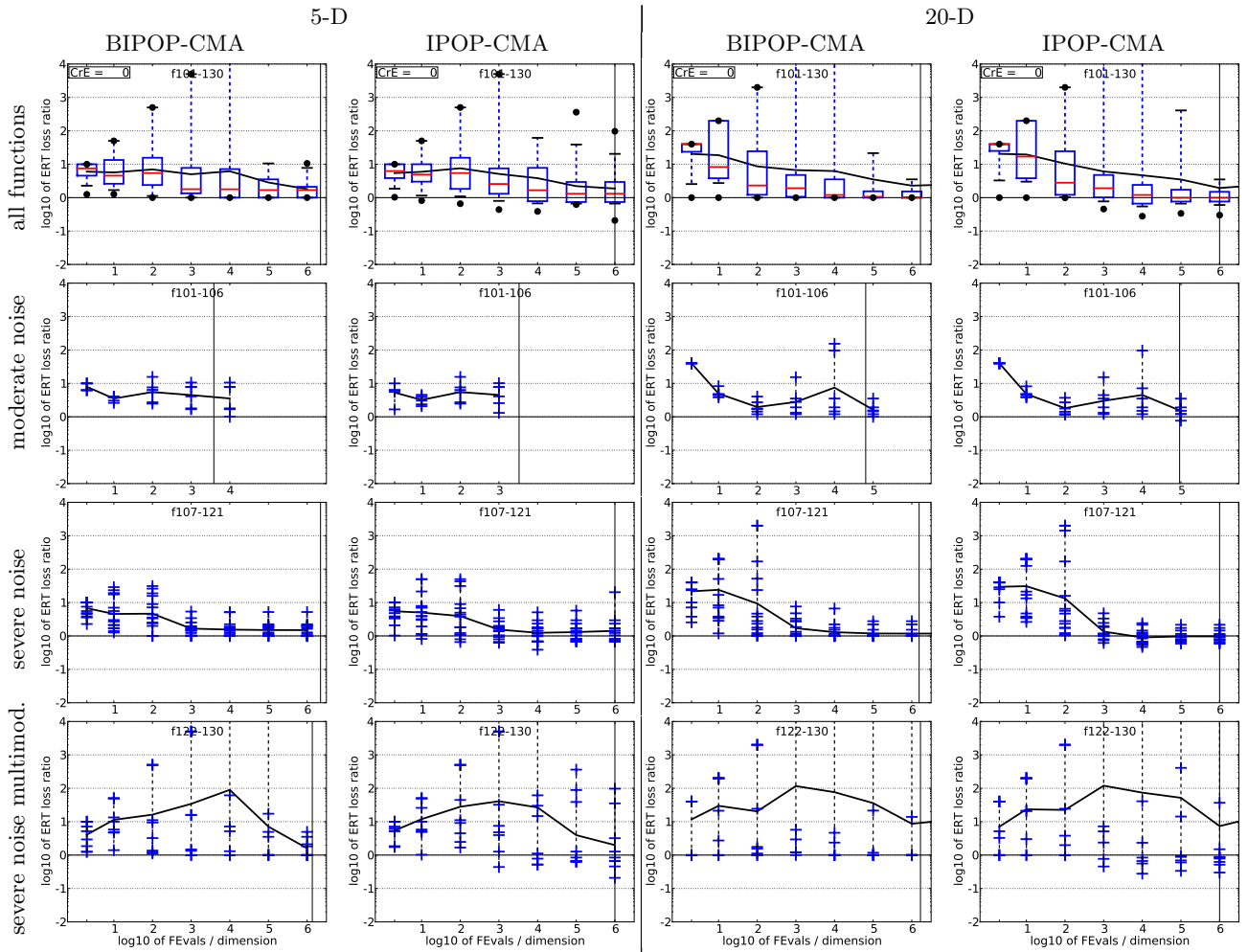


Figure 2: Expected running time (ERT in  $\log_{10}$  of number of function evaluations) of IPOP-CMA versus BIPOP-CMA for 46 target values  $\Delta f \in [10^{-8}, 10]$  in each dimension for functions  $f_{101}-f_{130}$ . Markers on the upper or right edge indicate that the target value was never reached by IPOP-CMA or BIPOP-CMA respectively. 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 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 IPOP-CMA (solid) and BIPOP-CMA (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 IPOP-CMA divided by BIPOP-CMA, 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 (IPOP-CMA first).





**Figure 4: ERT loss ratio versus given budget FEvals. The target value  $f_t$  for ERT is the smallest (best) recorded function value such that  $ERT(f_t) \leq FEvals$  for the presented algorithm. Shown is FEvals divided by the respective best  $ERT(f_t)$  from BBOB-2009 for functions  $f_{101}$ – $f_{130}$  in 5-D and 20-D. Each ERT is multiplied by  $\exp(CrE)$  correcting for the parameter crafting effort. Line: geometric mean. Box-Whisker error bar: 25-75%-ile with median (box), 10-90%-ile (caps), and minimum and maximum ERT loss ratio (points). The vertical line gives the maximal number of function evaluations in this function subset.**

- 2010: Presentation of the noisy functions. Technical Report 2009/21, Research Center PPE, 2010.
- [3] N. Hansen. The CMA evolution strategy: a comparing review. In J. Lozano, P. Larranaga, I. Inza, and E. Bengoetxea, editors, *Towards a new evolutionary computation. Advances on estimation of distribution algorithms*, pages 75–102. Springer, 2006.
- [4] N. Hansen. Benchmarking a bi-population CMA-ES on the BBOB-2009 noisy testbed. In F. Rothlauf, editor, *GECCO (Companion)*, pages 2397–2402. ACM, 2009.
- [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 and S. Kern. Evaluating the CMA evolution strategy on multimodal test functions. In X. Yao et al., editors, *Parallel Problem Solving from Nature - PPSN VIII, LNCS 3242*, pages 282–291. Springer, 2004.
- [8] N. Hansen, A. Niederberger, L. Guzzella, and P. Koumoutsakos. A method for handling uncertainty in evolutionary optimization with an application to feedback control of combustion. *IEEE Transactions on Evolutionary Computation*, 13(1):180–197, 2009.
- [9] N. Hansen and A. Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 9(2):159–195, 2001.
- [10] 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.

**Table 2:** ERT loss ratio (see Figure 4) compared to the respective best result from BBOB-2009 for budgets given in the first column. The last row  $RL_{US}/D$  gives the number of function evaluations in unsuccessful runs divided by dimension. Shown are the smallest, 10%-ile, 25%-ile, 50%-ile, 75%-ile and 90%-ile value (smaller values are better). ERT Loss ratio is equal to zero if the algorithm considered outperformed all algorithms from BBOB-2009.

BIPOP-CMA							IPOP-CMA						
#FEs/D	$f_{101}-f_{130}$ in 5-D, maxFE/D=2.23e6						#FEs/D	$f_{101}-f_{130}$ in 5-D, maxFE/D=1.00e6					
	best	10%	25%	med	75%	90%		best	10%	25%	med	75%	90%
2	1.3	2.1	4.5	7.3	10	10	2	1.0	1.8	3.8	6.2	10	10
10	1.3	1.5	2.6	4.3	13	39	10	0.81	1.1	3.0	4.7	10	50
100	1.0	1.1	2.4	4.9	16	2.7e2	100	0.66	1.1	1.8	5.2	16	2.8e2
1e3	1.0	1.0	1.3	1.8	7.8	2.5e3	1e3	0.44	0.73	1.3	2.3	7.5	2.5e3
1e4	1.0	1.0	1.0	1.7	7.2	2.5e4	1e4	0.39	0.60	0.78	1.5	7.8	46
1e5	1.0	1.0	1.0	1.7	3.5	9.2	1e5	0.62	0.66	0.74	1.3	2.9	25
1e6	1.0	1.0	1.0	1.7	2.1	6.5	1e6	0.21	0.66	0.74	1.3	2.9	15
$RL_{US}/D$	8e5	8e5	9e5	1e6	1e6	1e6	$RL_{US}/D$	4e5	4e5	5e5	8e5	1e6	1e6

BIPOP-CMA							IPOP-CMA						
#FEs/D	$f_{101}-f_{130}$ in 20-D, maxFE/D=1.68e6						#FEs/D	$f_{101}-f_{130}$ in 20-D, maxFE/D=1.00e6					
	best	10%	25%	med	75%	90%		best	10%	25%	med	75%	90%
2	1.0	1.8	24	40	40	40	2	1.0	2.1	25	40	40	40
10	1.0	2.0	3.8	8.3	2.0e2	2.0e2	10	1.0	2.8	3.8	15	2.0e2	2.0e2
100	1.0	1.0	1.2	2.0	24	2.0e3	100	0.99	1.0	1.2	2.8	24	2.0e3
1e3	1.0	1.0	1.1	1.9	4.8	2.0e4	1e3	0.46	0.76	1.0	1.9	4.8	2.0e4
1e4	1.0	1.0	1.0	1.2	3.5	1.0e5	1e4	0.28	0.54	0.66	1.2	2.4	1.0e5
1e5	1.0	1.0	1.0	1.0	1.5	12	1e5	0.34	0.63	0.76	0.96	1.7	2.1e2
1e6	1.0	1.0	1.0	1.0	1.5	3.1	1e6	0.30	0.60	0.76	0.96	1.5	2.8
1e7	1.0	1.0	1.0	1.0	1.5	3.1	1e7	0.52	0.63	0.77	1.1	1.7	3.7
$RL_{US}/D$	4e5	5e5	1e6	1e6	1e6	1e6	$RL_{US}/D$	1e5	2e5	6e5	9e5	1e6	1e6