



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_μ , 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, $f_t = f_{\text{opt}} + \Delta f_t$, 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, 10]. **Statistical significance** is tested with the rank-sum test for a given target Δf_t (10^{-8} in Figure 1) 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 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

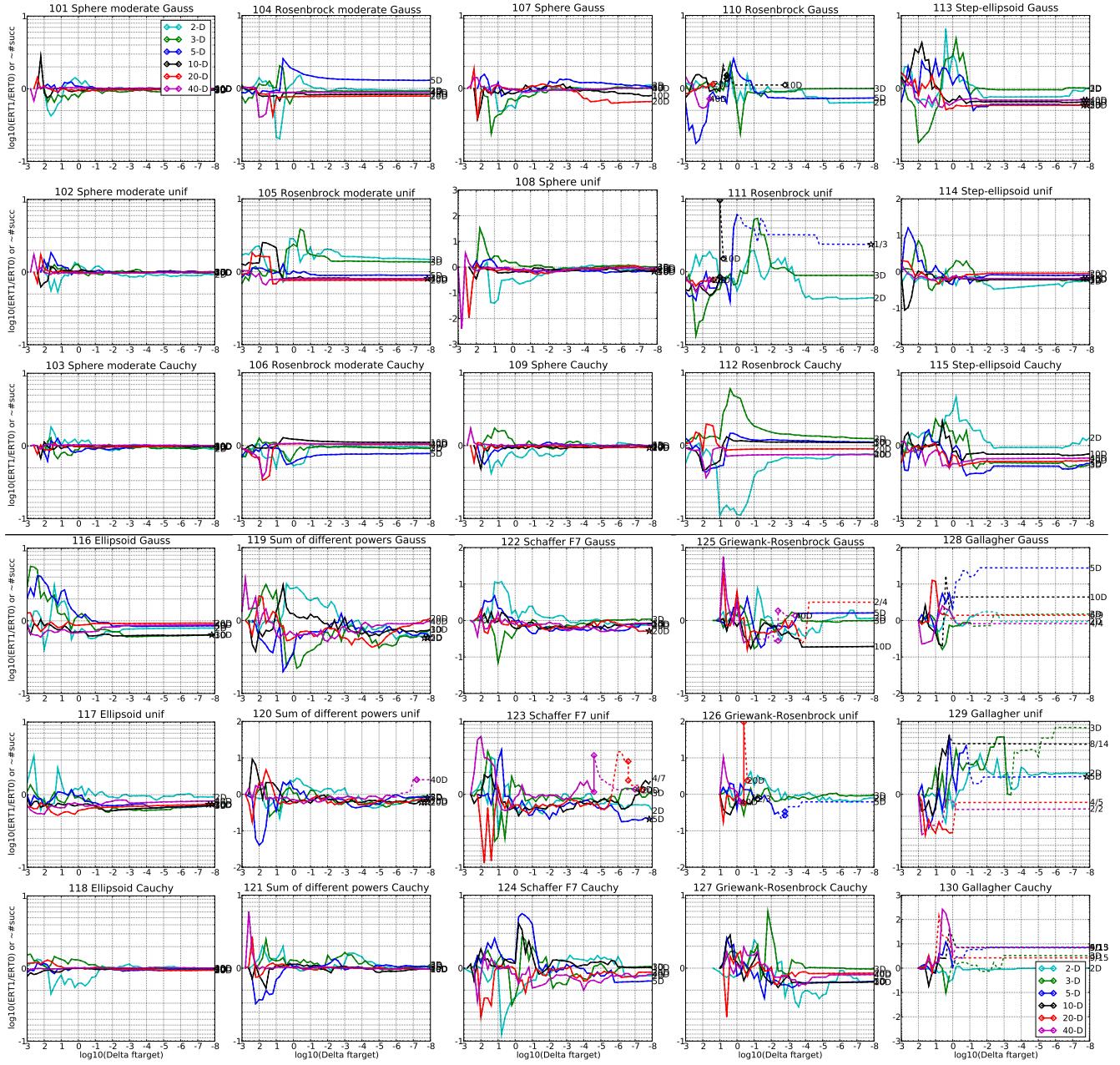


Figure 1: Ratio of the expected running times (ERT) of IPOPCMA divided by BIPOPCMA versus $\log_{10}(\Delta f)$ for $f_{101}\dots f_{130}$ in 2, 3, 5, 10, 20, 40-D. Ratios $< 10^0$ indicate an advantage of IPOPCMA, smaller values are always better. The line gets dashed when for any algorithm the ERT 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 Δf -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 IPOPCMA. The line ends when no algorithm reaches Δf anymore. The number of successful trials is given, only if it was in $\{1\dots 9\}$ for IPOPCMA (1st number) and non-zero for BIPOPCMA (2nd number). Results are statistically significant with $p = 0.05$ for one star and $p = 10^{-\#\star}$ otherwise, with Bonferroni correction within each figure.

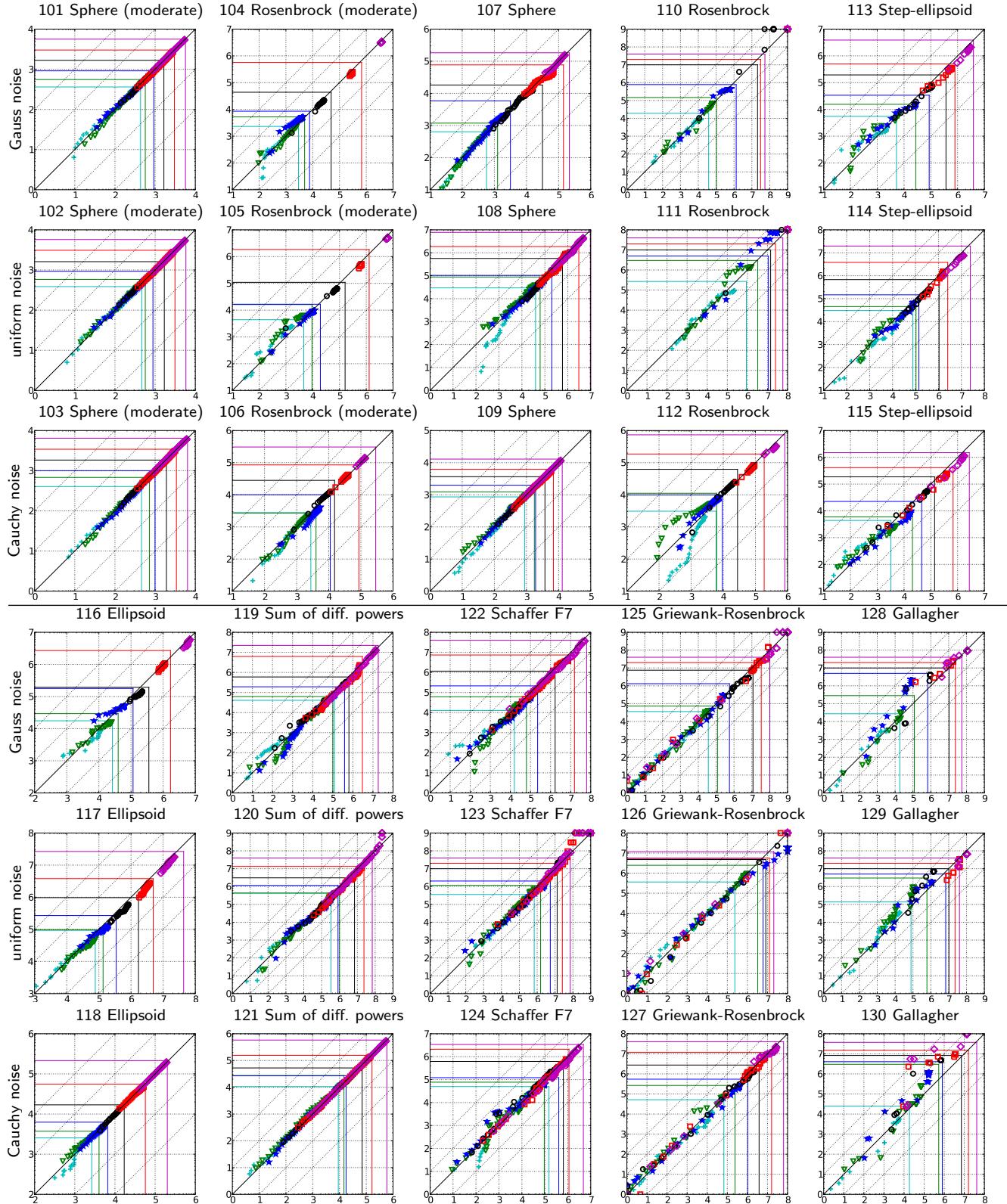


Figure 2: Expected running time (ERT in \log_{10} of number of function evaluations) of IPOP-CMA versus BIPOP-CMA for 46 target values $\Delta f \in [10^{-8}, 10]$ in each dimension for functions f_{101} – f_{130} . Markers on the upper or right edge indicate that the target value was never reached by IPOP-CMA or BIPOP-CMA respectively. Markers represent dimension: 2: $\textcolor{cyan}{+}$, 3: $\textcolor{green}{\triangledown}$, 5: $\textcolor{blue}{\star}$, 10: $\textcolor{black}{\circ}$, 20: $\textcolor{red}{\square}$, 40: $\textcolor{magenta}{\diamond}$.

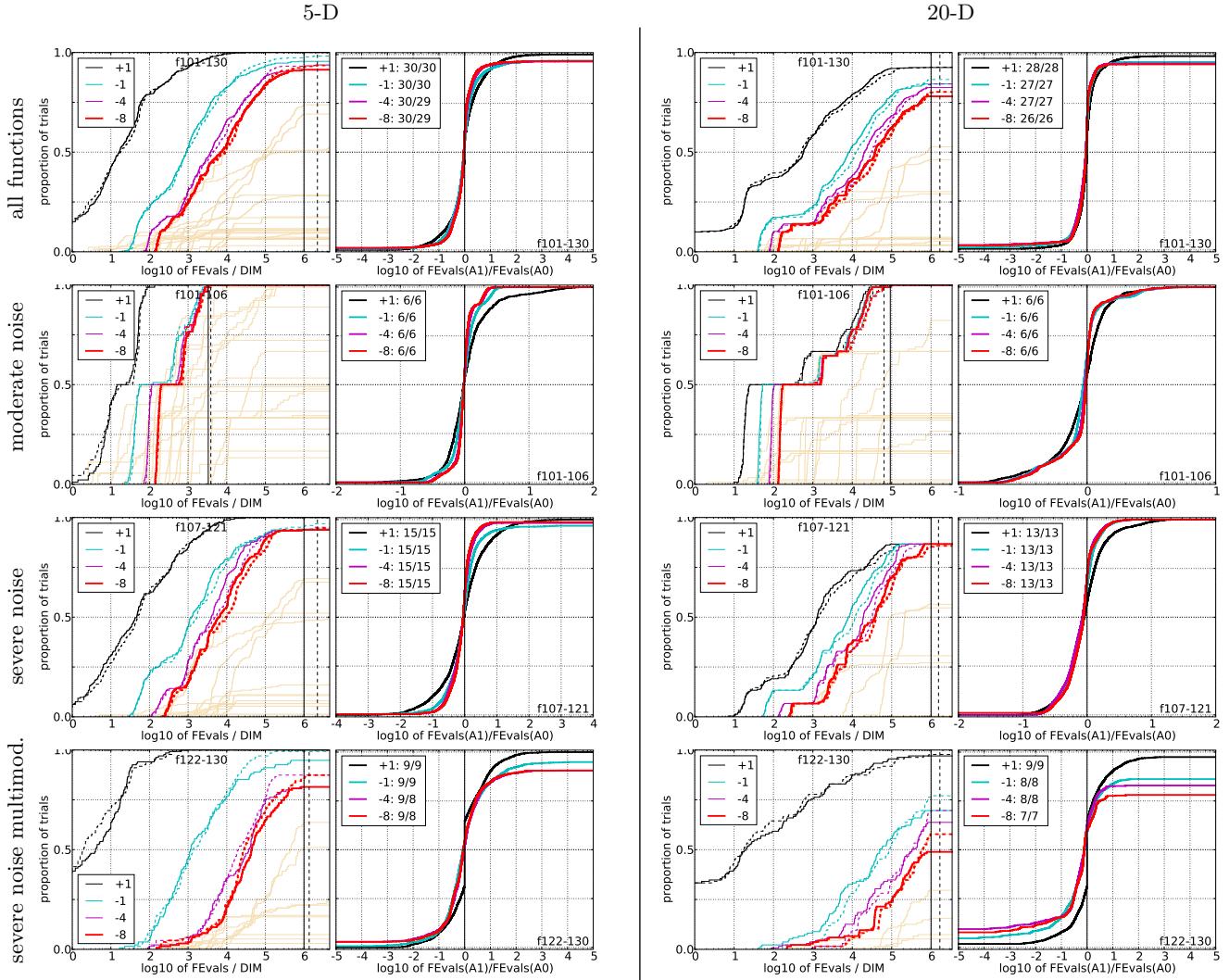


Figure 3: Empirical cumulative distributions (ECDF) of run lengths and speed-up ratios in 5-D (left) and 20-D (right). Left sub-columns: ECDF of the number of necessary function evaluations divided by dimension D (FEvals/D) to reached a target value $f_{\text{opt}} + \Delta f$ with $\Delta f = 10^k$, 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).

5-D

Δf	1e+1	1e+0	1e-1	1e-3	1e-5	1e-7	#succ
f₁₀₁	11	37	44	62	69	75	15/15
0: BIP	3.2	3.1	4.6	6.1	8.0	10	15/15
1: IPO	3.3	3.4	4.7	6.0	7.8	9.3	15/15
f₁₀₂	11	35	50	72	86	99	15/15
0: BIP	2.7	3.0	4.0	5.1	6.3	7.2	15/15
1: IPO	3.4	3.1	4.1	5.1	6.5	7.3	15/15
f₁₀₃	11	28	30	31	35	115	15/15
0: BIP	3.5	4.7	7.4	13	17	6.9	15/15
1: IPO	3.6	4.0	6.6	12	17	7.1	15/15
f₁₀₄	173	773	1287	1768	2040	2284	15/15
0: BIP	1.4	1.9	2.0	2.0	1.9	1.8	15/15
1: IPO	1.4	3.4	2.9	2.7	2.5	2.4	15/15
f₁₀₅	167	1436	5174	10388	10824	11202	15/15
0: BIP	1.7	3.7	1.7	1	1	1	15/15
1: IPO	1.6	3.8	1.6	0.90	0.90	0.90	15/15
f₁₀₆	86	529	1050	2666	2887	3087	15/15
0: BIP	3.6	4.3	3.2	1.6	1.7	1.7	15/15
1: IPO	3.3	2.5	2.2	1.2	1.3	1.3	15/15
f₁₀₇	40	228	453	940	1376	1850	15/15
0: BIP	1.7	1	1	1	1	1	15/15
1: IPO	2.1	0.98	1.1	1.3	1.2	1.1	15/15
f₁₀₈	87	5144	14469	30935	58628	80667	15/15
0: BIP	6.1	1.0	1	1	1	1	15/15
1: IPO	9.1	0.80	0.67	0.77	0.62 \downarrow	0.69	15/15
f₁₀₉	11	57	216	572	873	946	15/15
0: BIP	3.5	2.2	1.1	1.1	1.1	1.5	15/15
1: IPO	2.9	2.2	1.2	1.0	1.1	1.5	15/15
f₁₁₀	949	33625	1.20e5	5.93e5	6.03e5	6.11e5	15/15
0: BIP	1	4.8	3.7	1	1	1	15/15
1: IPO	0.73	8.3	3.4	0.72	0.73	0.74	15/15
f₁₁₁	6856	6.12e5	8.83e6	2.30e7	3.10e7	3.13e7	3/15
0: BIP	1	2.5	1	1	1	1	3/15
1: IPO	0.78	15	3.9	3.2	2.4	2.4	1/15
f₁₁₂	107	1684	3421	4502	5132	5596	15/15
0: BIP	4.0	1	1.2	1.3	1.3	1.3	15/15
1: IPO	2.1	1.4	1.4	1.5	1.5	1.5	15/15
f₁₁₃	133	1883	8081	24128	24128	24402	15/15
0: BIP	1.5	1.3	1.7	1.1	1.1	1.1	15/15
1: IPO	3.7	1.4	1.4	0.67	0.67	0.67	15/15
f₁₁₄	767	14720	56311	83272	83272	84949	15/15
0: BIP	2.2	1	1	1	1	1	15/15
1: IPO	3.2	0.45	0.48	0.79	0.79	0.80	15/15
f₁₁₅	64	485	1829	2550	2550	2970	15/15
0: BIP	1.5	2.6	6.5	5.9	5.9	5.7	15/15
1: IPO	1.7	2.4	2.7	3.1	3.1	2.7	15/15
f₁₁₆	5730	14472	22311	26868	30329	31661	15/15
0: BIP	1.2	2.0	1.9	2.1	2.0	2.0	15/15
1: IPO	3.1	2.3	1.9	1.8	1.7	1.7	15/15
f₁₁₇	26686	76052	1.10e5	1.37e5	1.73e5	1.92e5	15/15
0: BIP	1	1	1	1	1	1	15/15
1: IPO	1.1	0.95	0.77	0.73	0.67	0.69	15/15
f₁₁₈	429	1217	1555	1998	2430	2913	15/15
0: BIP	3.2	2.0	1.9	2.1	2.0	1.8	15/15
1: IPO	3.2	2.0	1.9	2.0	1.9	1.7	15/15
f₁₁₉	12	657	1136	10372	35296	49747	15/15
0: BIP	1.9	1	1	1	1.5	2.3	15/15
1: IPO	1.1	0.35	0.70	0.83	1.0	1.4	15/15
f₁₂₀	16	2900	18698	72438	3.33e5	5.48e5	15/15
0: BIP	17	1.1	1	1	1	1	15/15
1: IPO	6.0	1.6	0.68	0.69	0.55	0.83	15/15
f₁₂₁	8.6	111	273	1583	3870	6195	15/15
0: BIP	2.7	1.1	1	1.1	2.0	2.2	15/15
1: IPO	1.9	1.1	1.0	1.1	2.1	2.3	15/15
f₁₂₂	10	1727	9190	30087	53743	1.11e5	15/15
0: BIP	2.2	1	1	1	1	1	15/15
1: IPO	4.8	0.94	0.44	0.56	0.68	0.67 \downarrow	15/15
f₁₂₃	11	16066	81505	3.36e5	6.71e5	2.22e6	15/15
0: BIP	8.1	1	1	1	1	1	15/15
1: IPO	23	0.62	0.52	0.74	0.65	0.45	15/15
f₁₂₄	10	202	1040	20478	45337	95200	15/15
0: BIP	1.5	1.1	1	1.1	1.2	1	15/15
1: IPO	2.8	1.3	4.0	1.2	0.93	0.65	15/15
f₁₂₅	1	1	1	2.39e5	2.43e5	2.46e5	15/15
0: BIP	1.1	17	3443	1	1	1	15/15
1: IPO	1	27	2599	0.78	1.3	1.3	15/15
f₁₂₆	1	1	1	∞	∞	∞	0
0: BIP	1	160	13292	∞	∞	∞	0/15
1: IPO	1	10254	1.21e7	1.88e7	1.89e7	2/15	0/15
f₁₂₇	1	1	1	3.42e5	3.89e5	3.95e5	15/15
0: BIP	1	19	2136	1	1	1	15/15
1: IPO	1	15	1542	0.58	0.64	0.65	15/15
f₁₂₈	111	4248	7808	12447	17217	21162	15/15
0: BIP	2.2	6.9	10	6.6	4.8	3.9	15/15
1: IPO	1.0	14	166	183	132	108	10/15
f₁₂₉	64	10710	59443	2.85e5	5.11e5	5.80e5	15/15
0: BIP	12	7.1	9.2	3.9	2.2	1.9	13/15
1: IPO	8.5	13	18	6.7	3.8	3.3	11/15
f₁₃₀	55	812	3034	32823	33889	34528	10/15
0: BIP	1.9	57	55	5.1	5.0	5.0	15/15
1: IPO	1.2	59	321	37	36	35	12/15

20-D

Δf	1e+1	1e+0	1e-1	1e-3	1e-5	1e-7	#succ
f₁₀₁	59	361	513	700	739	783	15/15
0: BIP	6.1	1.8	1.8	2.1	2.7	3.3	15/15
1: IPO	6.0	1.7	1.7	2.0	2.6	3.2	15/15
f₁₀₂	231	399	579	921	1157	1407	15/15
0: BIP	1.6	1.6	1.6	1.6	1.8	1.8	15/15
1: IPO	1.6	1.6	1.6	1.7	1.7	1.8	15/15
f₁₀₃	65	417	629	1313	1893	2464	14/15
0: BIP	5.5	1.6	1.5	1.2	1.2	1.2	15/15
1: IPO	5.5	1.5	1.4	1.2	1.2	1.2	15/15
f₁₀₄	23690	85656	1.71e5	1.82e5	1.89e5	1.96e5	15/15
0: BIP	10	3.2	1.7	1.6	1.6	1.6	15/15
1: IPO	7.5	2.5	1.3	1.3	1.3	1.2	15/15
f₁₀₅	1.92e5	6.11e5	6.32e5	6.49e5	6.60e5	6.70e5	15/15
0: BIP	2.7	1	1	1	1	1	15/15
1: IPO	1.9	0.76	0.77	0.77	0.76	0.76	15/15
f₁₀₆	11480	21668	23746	25470	26492	27360	15/15
0: BIP	1.0	1.3	1.4	1.5	1.5	1.5	15/15
1: IPO	1.0	1.4	1.5	1.5	1.5	1.5	15/15
f₁₀₇	8571	13582	16226	27357	52486	65052	15/15
0: BIP	1	1	1	1	1	1	15/15
1: IPO	1.1	0.95	1.1	0.96	0.68	0.65	15/15
f₁₀₈	58663	97228	2.03e5	4.46e5	6.30e5	8.98e5	15/15
0: BIP	1	1	1	1	1	1	15/15
1: IPO	0.72	0.87	0.66	0.77	0.94	1.0	15/15
f₁₀₉	333	632	1138	2287	3583	4952	15/15
0: BIP	1.2	1.2	1.1	1.1	1.1	1.0	15/15
1: IPO	1.1	1.2	1.1	1.1	1.0	1.0	15/15
f₁₁₀	∞	∞	∞	∞	∞	∞	0
0: BIP	∞	∞	∞	∞	∞	∞	0/15
1: IPO	∞	∞	∞	∞	∞	∞	0/15
f₁₁₁	∞	∞	∞	∞	∞	∞	0
0: BIP	∞	∞	∞	∞	∞	∞	0/15
1: IPO	∞	∞	∞	∞	∞	∞	0/15
f₁₁₂	25552	64124	69621	73557	76137	78238	15/15
0: BIP	1	1.1	1.1	1.2	1.2	1.2	15/15
1: IPO	0.95	0.94	1.0	1.1	1.1	1.1	15/15
f₁₁₃	50123	3.64e5	5.60e5	5.88e5	5.88e5	5.91e5	15/15
0: BIP	1	1	1	1	1	1	15/15
1: IPO	1.0	0.53	0.58 \downarrow	0.59 \downarrow	0.59 \downarrow	0.59 \downarrow	15/15
f₁₁₄	1.20e5	1.12e6	1.45e6	1.57e6	1.57e6	1.58e6	15/15
0: BIP	1	0.59 \downarrow	0.68	0.84	0.91	0.92	15/15
f₁₁₅	2405	30268	91749	1.27e5	1.27e5	1.29e5	15/15
0: BIP	1	6.5	3.9	3.0	3.0	3.0	15/15
1: IPO	1.1	4.8	2.2	1.8	1.8	1.9	15/15
f₁₁₆	4.98e5	6.94e5	8.93e5	1.03e6	1.08e6	1.12e6	15/15
0: BIP	1.4	1.2	1.1	1	1	1	15/15
1: IPO	1.2	1.1	1.00	0.92	0.93	0.93	15/15
f₁₁₇	1.79e6	2.46e6	2.60e6	2.91e6	3.24e6	3.62e6	15/15
0: BIP	1	1	1	1	1	1	15/15
1: IPO							

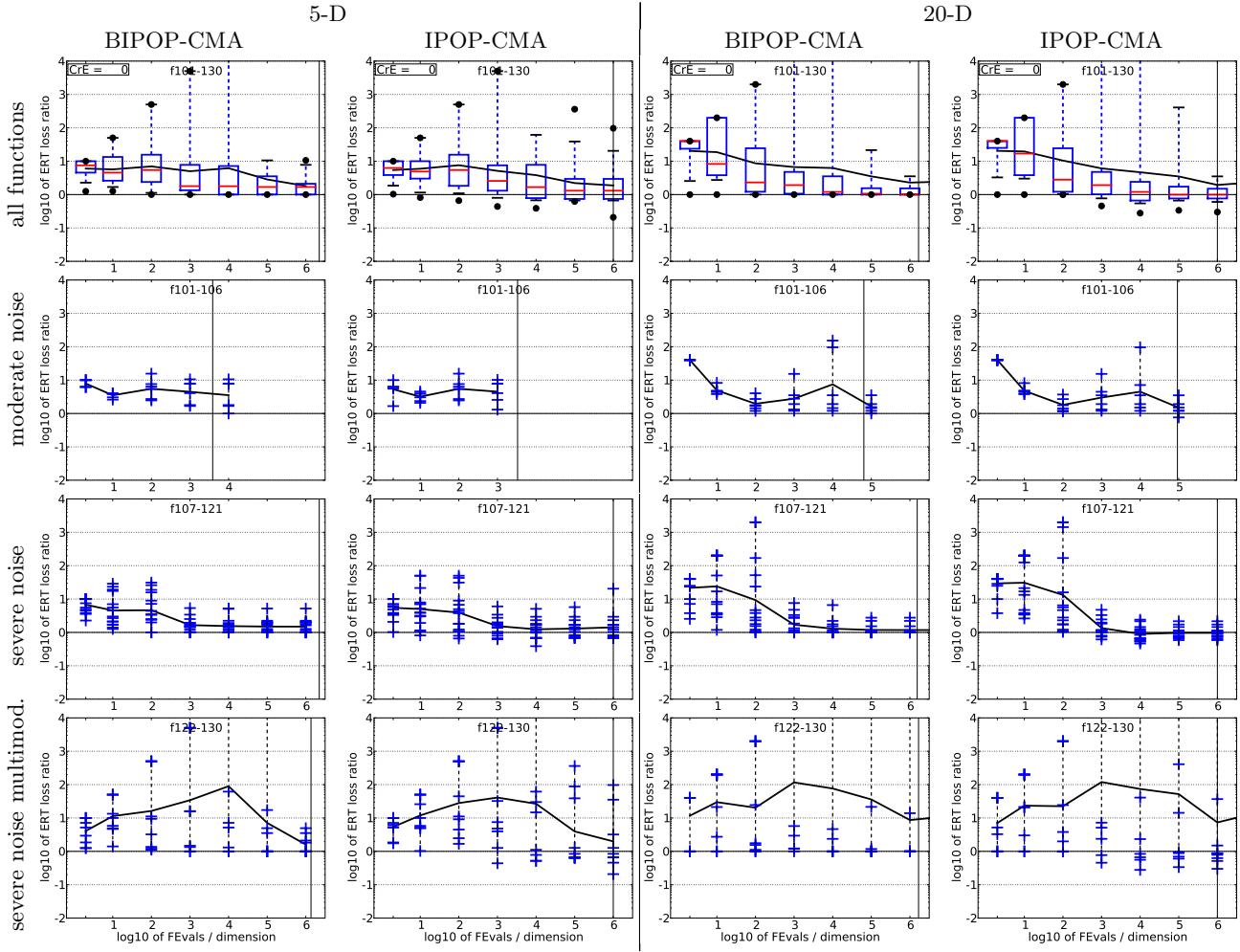


Figure 4: ERT loss ratio versus given budget FEvals. The target value f_t for ERT is the smallest (best) recorded function value such that $\text{ERT}(f_t) \leq \text{FEvals}$ for the presented algorithm. Shown is FEvals divided by the respective best $\text{ERT}(f_t)$ from BBOB-2009 for functions $f_{101}-f_{130}$ in 5-D and 20-D. Each ERT is multiplied by $\exp(\text{CrE})$ 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. Larrañaga, 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.

