Black-Box Optimization Benchmarking Comparison of Two Algorithms on the Noiseless Testbed

An Example BBOB 2010 Workshop Paper*

The BBOBies

ABSTRACT

This example paper shows results from the BBOB experimental procedure when comparing two algorithms. Two templates for comparing two algorithms are available: one for the noiseless and one for the noisy BBOB testbed. In this example, results on the noiseless testbed are shown, comparing NEWUOA with BIPOP-CMA-ES.

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

This is an example paper comparing the performance of NEWUOA [6] to BIPOP-CMA-ES [2].

2. PARAMETER TUNING

The parameter settings of NEWUOA and BIPOP-CMA-ES are described in [6] and [2]. Both algorithm have a crafting effort [3] equal to zero.

3. RESULTS

Results from experiments according to [3] on the benchmark functions given in [1, 4] are presented in Figures 1, 2 and 3 and in Table 1. The **expected running time (ERT)**, used in the figures and table, depends on a given target function value, $f_t = f_{opt} + \Delta f$, 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 [3, 5]. **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.

NEWUAO outperforms BIPOP-CMA-ES on f_1 by a factor of about 30 and on the Linear and the Rosenbrock function by a factor of about three. On the other unimodal functions the picture is comparatively mixed, presumably due to local deformations in the function topographies: besides f_1, all function deviate significantly from a quadratic form. The most surprising results can be observed on the multimodal functions f_{21} and f_{22}, where NEWUAO consistenly outperforms the BIPOP-CMA-ES, for larger dimension and the more difficult target values even by a factor between 10 and 100. The applied independent restarts of NEWUOA appear to be more effective than the large population size of BIPOP-CMA-ES, which is in turn more helpful on the remaining multi-modal functions.

4. CPU TIMING EXPERIMENTS

For the timing experiments, both algorithms were run on f_8 and restarted until at least 30 seconds (according to [3]. The experiments for NEWUOA has been conducted on a Intel Core 2 6700 processor (2.66 GHz) on Linux 2.6.24.7. The results were 8.1 ; 11 ; 21 ; 58 ; 170 ; 620 and 2500 $\times 10^{-6}$ seconds per function evaluations for NEWUOA in dimensions 2 ; 3 ; 5 ; 10 ; 20 ; 40 and 80 respectively. The experiments for BIPOP-CMA-ES has been conducted on a Intel Core 2 6700 processor (2.66 GHz) on Linux 2.6.24.7 using Matlab R2008a. The results were 6.2 ; 5.8 ; 5.6 ; 5.7 ; 5.8 ; 5.9 and 6.3 $\times 10^{-4}$ seconds per function evaluation for BIPOP-CMA-ES in dimensions 2 ; 3 ; 5 ; 10 ; 20 ; 40 and 80 respectively.

5. REFERENCES

[1] S. Finck, N. Hansen, R. Ros, and A. Auger. Real-parameter black-box optimization benchmarking 2009: Presentation of the noiseless functions. Technical

*Submission deadline: March 25th.

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

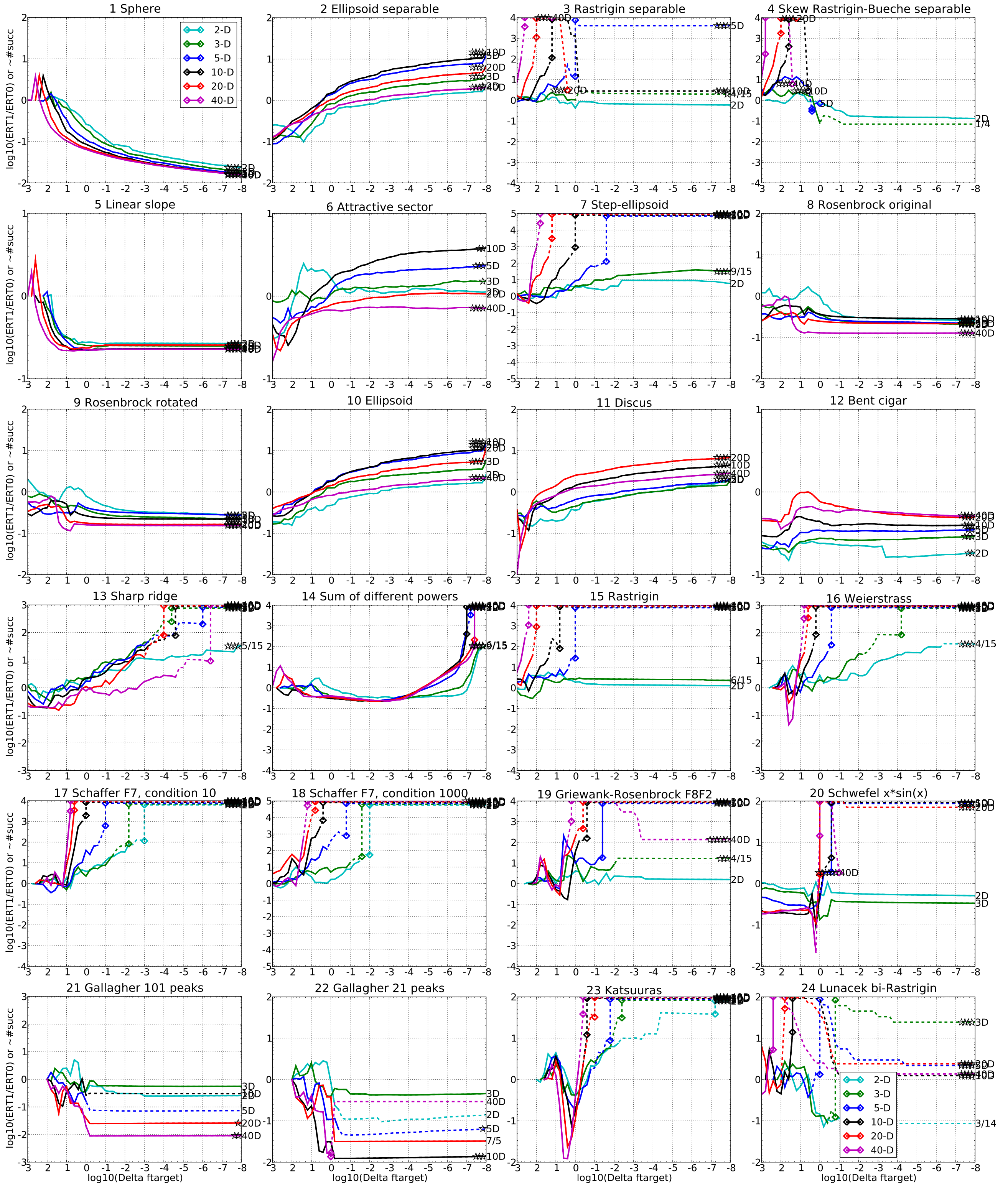

Figure 1: ERT ratio of NEWUOA divided by BIPOP-CMA versus $\log_{10}(\Delta f)$ for f_1–f_{24} in 2, 3, 5, 10, 20, 40-D. Ratios $< 10^0$ indicate an advantage of NEWUOA, smaller values are always better. The line gets dashed when for any algorithm the ERT exceeds thrice the median of the trial-wise overall 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 NEWUOA. The line ends when no algorithm reaches Δf anymore. The number of successful trials is given, only if it was in $\{1 \ldots 9\}$ for NEWUOA (1st number) and non-zero for BIPOP-CMA (2nd number). Results are significant with $p = 0.05$ for one star and $p = 10^{-\#\star}$ otherwise, with Bonferroni correction within each figure.

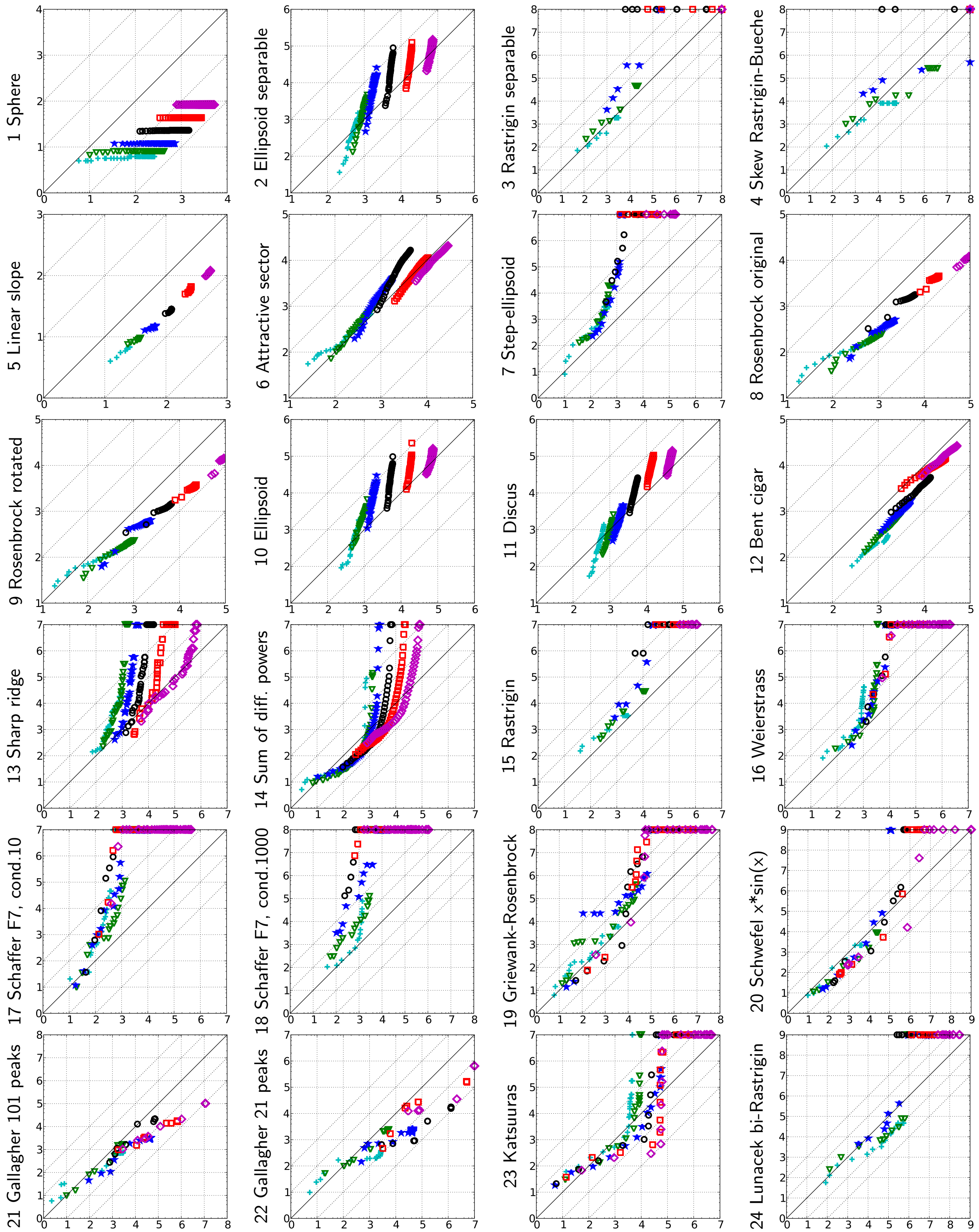

Figure 2: Expected running time (ERT in log10 of number of function evaluations) of NEWUOA versus BIPOP-CMA for 46 target values $\Delta f \in [10^{-8}, 10]$ in each dimension for functions f_1–f_{24}. Markers on the upper or right egde indicate that the target value was never reached by NEWUOA or BIPOP-CMA respectively. Markers represent dimension: 2:+, 3:▽, 5:★, 10:○, 20:□, 40:◇.

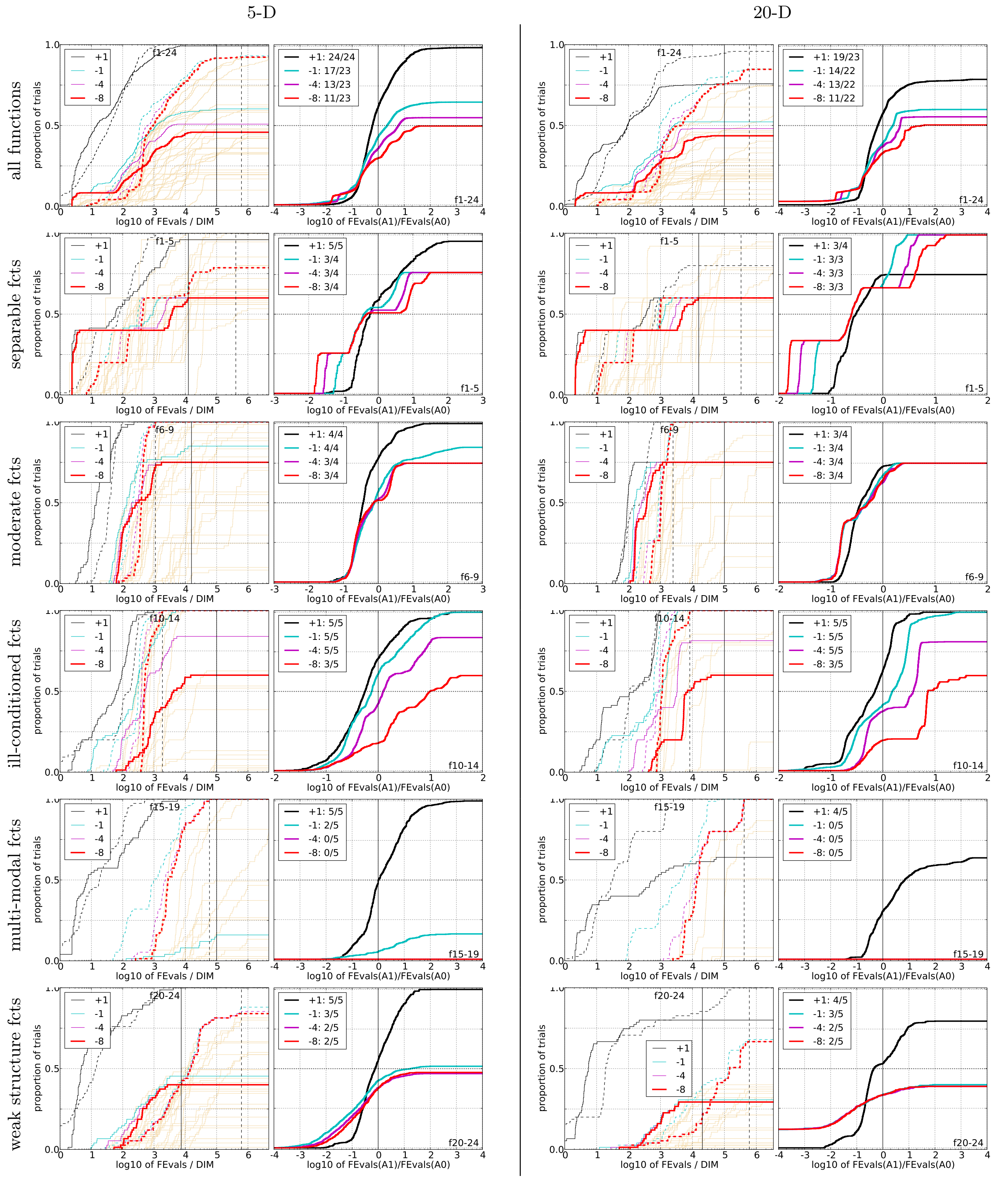

Figure 3: Empirical cumulative distributions (ECDF) of run lengths and speed-up ratios in 5-D (left) and 20-D (right). Left sub-columns: ECDF of the number of function evaluations divided by dimension D (FEvals/D) to reach a target value $f_{\rm opt} + \Delta f$ with $\Delta f = 10^k$, where $k \in \{1, -1, -4, -8\}$ is given by the first value in the legend, for NEWUOA (solid) and BIPOP-CMA (dashed). Light beige lines show the ECDF of FEvals for target value $\Delta f = 10^{-8}$ of algorithms benchmarked during BBOB-2009. Right sub-columns: ECDF of FEval ratios of NEWUOA 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 (NEWUOA first).

5-D

Δf	1e+1	1e+0	1e-1	1e-3	1e-5	1e-7	#succ
f_1	11	12	12	12	12	12	15/15
0: CMA	3.2	9	15	27	40	53	15/15
1: NEW	1.1	**1**$^{\star 3}$	**1**$^{\star 3}$	**1**$^{\star 3}$	**1**$^{\star 3}$	**1**$^{\star 3}$	15/15
f_2	83	87	88	90	92	94	15/15
0: CMA	13	16	**18***	**20**$^{\star 2}$	**21**$^{\star 3}$	**22**$^{\star 3}$	15/15
1: NEW	**5.7**$^{\star 2}$	22	45	85	130	170	15/15
f_3	720	1600	1600	1600	1700	1700	15/15
0: CMA	1.4	**16**$^{\star 3}$	**140**$^{\star 2}$	**140**$^{\star 2}$	**140**$^{\star 2}$	**140**$^{\star 2}$	14/15
1: NEW	6.1	230	∞	∞	∞	∞ *2.5e4*	0/15
f_4	810	1600	1700	1800	1900	1900	15/15
0: CMA	**2.7**$^{\star 3}$	∞	∞	∞	∞	∞ *1.8e6*	0/15
1: NEW	27	300	∞	∞	∞	∞ *3.4e4*	0/15
f_5	10	10	10	10	10	10	15/15
0: CMA	4.5	6.5	6.6	6.6	6.6	6.6	15/15
1: NEW	**1.3**$^{\star 3}$	**1.5**$^{\star 3}$	**1.5**$^{\star 3}$	**1.5**$^{\star 3}$	**1.5**$^{\star 3}$	**1.5**$^{\star 3}$	15/15
f_6	110	210	280	580	1000	1300	15/15
0: CMA	2.3	2.1	2.2	**1.7***	**1.3**$^{\star 2}$	**1.3**$^{\star 2}$	15/15
1: NEW	1.7	2.4	3.6	3.3	2.7	2.9	15/15
f_7	24	320	1200	1600	1600	1600	15/15
0: CMA	5	1.5	**1**$^{\star 3}$	**1**$^{\star 3}$	**1**$^{\star 3}$	**1**$^{\star 3}$	15/15
1: NEW	9.9	13	60	∞	∞	∞ *2.9e4*	0/15
f_8	73	270	340	390	410	420	15/15
0: CMA	3.2	3.7	4.5	4.8	5.1	5.4	15/15
1: NEW	**1**$^{\star 2}$	**1.1**$^{\star 2}$	**1.2**$^{\star 3}$	**1.2**$^{\star 3}$	**1.2**$^{\star 3}$	**1.2**$^{\star 3}$	15/15
f_9	35	130	210	300	340	370	15/15
0: CMA	5.8	8.7	7.2	6.4	6.3	6.2	15/15
1: NEW	**1.8**$^{\star 3}$	3.6	**2.5**$^{\star 2}$	**1.9**$^{\star 2}$	**1.9**$^{\star 3}$	**1.7**$^{\star 3}$	15/15
f_{10}	350	500	570	630	830	880	15/15
0: CMA	3.5	2.9	2.7	**2.8**$^{\star 3}$	**2.3**$^{\star 3}$	**2.4**$^{\star 3}$	15/15
1: NEW	3.1	5.5	8.1	14	16	21	15/15
f_{11}	140	200	760	1200	1500	1700	15/15
0: CMA	8.4	7.2	2.2	1.6	**1.4**$^{\star 3}$	**1.3**$^{\star 3}$	15/15
1: NEW	**3.5**$^{\star 3}$	**4.7***	1.8	1.8	2	2.2	15/15
f_{12}	110	270	370	460	1300	1500	15/15
0: CMA	11	7.4	7.4	7.7	3.3	3.3	15/15
1: NEW	3.5	**2.6***	**2.5***	**2.6**$^{\star 2}$	**1.1**$^{\star 2}$	**1.1***	15/15
f_{13}	130	190	250	1300	1800	2300	15/15
0: CMA	3.9	5.4	5.9	**1.6**$^{\star 3}$	**1.5**$^{\star 3}$	**1.7**$^{\star 3}$	15/15
1: NEW	3.1	9.3	35	54	330	∞ *4.0e4*	0/15
f_{14}	9.8	41	58	140	250	480	15/15
0: CMA	1.1	2.8	3.7	4.6	5.4	**4.5**$^{\star 3}$	15/15
1: NEW	1.7	**1**$^{\star 3}$	**1**$^{\star 3}$	**1.2**$^{\star 3}$	5.5	2.5e3	0/15
f_{15}	510	9300	1.9e4	2.0e4	2.1e4	2.1e4	14/15
0: CMA	1.6	**1.5**$^{\star 3}$	**1.2**$^{\star 2}$	**1.2**$^{\star 2}$	**1.2**$^{\star 2}$	**1.2**$^{\star 2}$	15/15
1: NEW	5.8	41	∞	∞	∞	∞ *2.5e4*	0/15
f_{16}	120	610	2700	1.0e4	1.2e4	1.2e4	15/15
0: CMA	3	**3.6**$^{\star 2}$	**2.6**$^{\star 3}$	**1.3**$^{\star 3}$	**1.4**$^{\star 3}$	**1.4**$^{\star 3}$	15/15
1: NEW	2.1	29	∞	∞	∞	∞ *3.6e4*	0/15
f_{17}	5.2	210	900	3700	6400	7900	15/15
0: CMA	3.4	**1***	**1**$^{\star 3}$	**1**$^{\star 3}$	**1**$^{\star 3}$	**1.2**$^{\star 3}$	15/15
1: NEW	2.3	40	620	∞	∞	∞ *3.4e4*	0/15
f_{18}	100	380	4000	9300	1.1e4	1.2e4	15/15
0: CMA	1	**3.4**$^{\star 3}$	**1**$^{\star 3}$	**1**$^{\star 3}$	**1.2**$^{\star 3}$	**1.3**$^{\star 3}$	15/15
1: NEW	31	1.4e3	∞	∞	∞	∞ *9.2e4*	0/15
f_{19}	1	1	240	1.2e5	1.2e5	1.2e5	15/15
0: CMA	20	2.8e3	160	**1**$^{\star 3}$	**1**$^{\star 3}$	**1**$^{\star 3}$	15/15
1: NEW	14	2.7e4	1.4e3	∞	∞	∞ *5.0e5*	0/15
f_{20}	16	850	3.8e4	5.4e4	5.5e4	5.5e4	14/15
0: CMA	3.3	8.2	2.8	2.1	2.2	2.2	15/15
1: NEW	**1**$^{\star 2}$	3.3	∞	∞	∞	∞ *3.2e4*	0/15
f_{21}	41	1200	1700	1700	1700	1800	14/15
0: CMA	2.3	14	24	25	25	25	15/15
1: NEW	1.1	2.2	1.8	1.8	1.8	1.9	15/15
f_{22}	71	390	940	1000	1000	1100	14/15
0: CMA	6.9	20	45	42	41	40	15/15
1: NEW	2.1	2.1	2	2.1	2.3	2.4	15/15
f_{23}	3	520	1.4e4	3.2e4	3.3e4	3.4e4	15/15
0: CMA	**1.7**$^{\star 2}$	13	3.7	1.8	1.8	1.8	15/15
1: NEW	6.2	2.4	7.1	∞	∞	∞ *3.3e4*	0/15
f_{24}	1600	2.2e5	6.4e6	9.6e6	1.3e7	1.3e7	3/15
0: CMA	2.1	1.6	1	1	1	1	3/15
1: NEW	2.9	2.1	∞	∞	∞	∞ *3.0e4*	0/15

20-D

Δf	1e+1	1e+0	1e-1	1e-3	1e-5	1e-7	#succ
f_1	43	43	43	43	43	43	15/15
0: CMA	7.9	14	20	33	45	57	15/15
1: NEW	**1**$^{\star 3}$	**1**$^{\star 3}$	**1**$^{\star 3}$	**1**$^{\star 3}$	**1**$^{\star 3}$	**1**$^{\star 3}$	15/15
f_2	380	390	390	390	390	390	15/15
0: CMA	35	40	**44**$^{\star 2}$	**47**$^{\star 3}$	**48**$^{\star 3}$	**50**$^{\star 3}$	15/15
1: NEW	**18**$^{\star 3}$	42	71	130	170	220	15/15
f_3	5100	7600	7600	7600	7600	7700	15/15
0: CMA	**12**$^{\star 3}$	∞	∞	∞	∞	∞ *5.7e6*	0/15
1: NEW	∞	∞	∞	∞	∞	∞ *1.3e5*	0/15
f_4	4700	7600	7700	7700	7800	1.4e5	9/15
0: CMA	∞	∞	∞	∞	∞	∞ *5.5e6*	0/15
1: NEW	∞	∞	∞	∞	∞	∞ *2.2e5*	0/15
f_5	41	41	41	41	41	41	15/15
0: CMA	5.1	6.2	6.3	6.3	6.3	6.3	15/15
1: NEW	**1.2**$^{\star 3}$	**1.5**$^{\star 3}$	**1.6**$^{\star 3}$	**1.6**$^{\star 3}$	**1.6**$^{\star 3}$	**1.6**$^{\star 3}$	15/15
f_6	1300	2300	3400	5200	6700	8400	15/15
0: CMA	1.5	1.3	1.2	1.1	1.2	1.2	15/15
1: NEW	**1**$^{\star 2}$	1	1	1.1	1.3	1.3	15/15
f_7	1400	4300	9500	1.7e4	1.7e4	1.7e4	15/15
0: CMA	**1**$^{\star 3}$	**4.9**$^{\star 3}$	**3.5**$^{\star 3}$	**2.2**$^{\star 3}$	**2.2**$^{\star 3}$	**2.1**$^{\star 3}$	15/15
1: NEW	∞	∞	∞	∞	∞	∞ *4.8e5*	0/15
f_8	2000	3900	4000	4200	4400	4500	15/15
0: CMA	4	4	4.3	4.5	4.6	4.6	15/15
1: NEW	**1**$^{\star 3}$	**1**$^{\star 3}$	**1**$^{\star 3}$	**1**$^{\star 3}$	**1**$^{\star 3}$	**1**$^{\star 3}$	15/15
f_9	1700	3100	3300	3500	3600	3700	15/15
0: CMA	4.7	5.7	6	6.1	6.1	6.1	15/15
1: NEW	**1**$^{\star 3}$	**1**$^{\star 3}$	**1**$^{\star 3}$	**1**$^{\star 3}$	**1**$^{\star 3}$	**1**$^{\star 3}$	15/15
f_{10}	7400	8700	1.1e4	1.5e4	1.7e4	1.7e4	15/15
0: CMA	1.9	**1.8**$^{\star 2}$	**1.6**$^{\star 3}$	**1.2**$^{\star 3}$	**1.1**$^{\star 3}$	**1.1**$^{\star 3}$	15/15
1: NEW	1.7	2.6	3.3	4	4.7	5.8	15/15
f_{11}	1000	2200	6300	9800	1.2e4	1.5e4	15/15
0: CMA	**10**$^{\star 3}$	**5.1**$^{\star 3}$	**1.9**$^{\star 3}$	**1.4**$^{\star 3}$	**1.2**$^{\star 3}$	**1**$^{\star 3}$	15/15
1: NEW	15	13	5.8	6.1	6.6	6.5	15/15
f_{12}	1000	1900	2700	4100	1.2e4	1.4e4	15/15
0: CMA	3	4	4.5	4.5	1.9	2	15/15
1: NEW	3	3	3	2.5	**1**$^{\star 2}$	**1**$^{\star 3}$	15/15
f_{13}	650	2000	2800	1.9e4	2.4e4	3.0e4	15/15
0: CMA	4.3	2.7	5.1	**1.5**$^{\star 2}$	**2.3**$^{\star 3}$	**3**$^{\star 3}$	15/15
1: NEW	**1***	3	9.3	19	∞	∞ *1.8e5*	0/15
f_{14}	75	240	300	930	1600	1.6e4	15/15
0: CMA	3.9	2.9	3.7	4.1	**6.2**$^{\star 3}$	**1.2**$^{\star 3}$	15/15
1: NEW	**1.5**$^{\star 3}$	**1**$^{\star 3}$	**1**$^{\star 3}$	**1**$^{\star 3}$	9.1	43	0/15
f_{15}	3.0e4	1.5e5	3.1e5	3.2e5	4.5e5	4.6e5	15/15
0: CMA	**1**$^{\star 3}$	**2**$^{\star 3}$	**1.4**$^{\star 3}$	**1.4**$^{\star 3}$	**1**$^{\star 3}$	**1**$^{\star 3}$	15/15
1: NEW	∞	∞	∞	∞	∞	∞ *1.3e5*	0/15
f_{16}	1400	2.7e4	7.7e4	1.9e5	2.0e5	2.2e5	15/15
0: CMA	**1.7**$^{\star 2}$	**1**$^{\star 3}$	**1.2**$^{\star 3}$	**1**$^{\star 3}$	**1**$^{\star 3}$	**1**$^{\star 3}$	15/15
1: NEW	16	∞	∞	∞	∞	∞ *2.3e5*	0/15
f_{17}	63	1000	4000	3.1e4	5.6e4	8.0e4	15/15
0: CMA	2.2	**1**$^{\star 3}$	**1**$^{\star 3}$	**1.2**$^{\star 3}$	**1.3**$^{\star 3}$	**1.4**$^{\star 3}$	15/15
1: NEW	16	∞	∞	∞	∞	∞ *1.5e6*	0/15
f_{18}	620	4000	2.0e4	6.8e4	1.3e5	1.5e5	15/15
0: CMA	**1**$^{\star 3}$	**2.4**$^{\star 3}$	**1.2**$^{\star 3}$	**1.1**$^{\star 3}$	**1.7**$^{\star 3}$	**1.6**$^{\star 3}$	15/15
1: NEW	1.2e4	∞	∞	∞	∞	∞ *1.6e6*	0/15
f_{19}	1	1	3.4e5	6.2e6	6.7e6	6.7e6	15/15
0: CMA	170	**2.4e4***	**1.2**$^{\star 3}$	**1**$^{\star 3}$	**1**$^{\star 3}$	**1**$^{\star 3}$	15/15
1: NEW	**76**$^{\star 2}$	4.3e6	∞	∞	∞	∞ *2.0e6*	0/15
f_{20}	82	4.6e4	3.1e6	5.5e6	5.6e6	5.6e6	14/15
0: CMA	4.3	9.2	1	1	1	1	14/15
1: NEW	**1**$^{\star 3}$	15	∞	∞	∞	∞ *3.8e5*	0/15
f_{21}	560	6500	1.4e4	1.5e4	1.6e4	1.8e4	15/15
0: CMA	3.2	55	48	46	43	39	13/15
1: NEW	1.7	2.2	1.2	1.2	1.1	1	15/15
f_{22}	470	5600	2.3e4	2.5e4	2.7e4	1.3e5	12/15
0: CMA	6.8	13	210	200	190	37	5/15
1: NEW	**1***	4.9	6.8	6.4	6	1.2	7/15
f_{23}	3.2	1600	6.7e4	4.9e5	8.1e5	8.4e5	15/15
0: CMA	4.3	32	**1**$^{\star 3}$	**2**$^{\star 2}$	**1.2**$^{\star 2}$	**1.2**$^{\star 2}$	15/15
1: NEW	12	**3.5**$^{\star 3}$	32	∞	∞	∞ *1.5e5*	0/15
f_{24}	1.3e6	7.5e6	5.2e7	5.2e7	5.2e7	5.2e7	3/15
0: CMA	**1**$^{\star 2}$	**1**$^{\star 2}$	**1**$^{\star 2}$	**1**$^{\star 2}$	**1**$^{\star 2}$	**1**$^{\star 2}$	3/15
1: NEW	∞	∞	∞	∞	∞	∞ *1.7e5*	0/15

Table 1: Expected running time (ERT in number of function evaluations) divided by the best ERT measured during BBOB-2009 (given in the respective first row) for different Δf values for functions f_1–f_{24}. The median number of conducted function evaluations is additionally given in *italics*, if $\mathrm{ERT}(10^{-7}) = \infty$. #succ is the number of trials that reached the final target $f_{\mathrm{opt}} + 10^{-8}$. 0: CMA is BIPOP-CMA and 1: NEW is NEWUOA. Bold entries are statistically significantly better compared to the other algorithm, with $p = 0.05$ or $p = 10^{-k}$ where $k > 1$ is the number following the $\star$ symbol, with Bonferroni correction of 48.

Report 2009/20, Research Center PPE, 2009. Updated February 2010.

[2] N. Hansen. Benchmarking a bi-population CMA-ES on the BBOB-2009 function testbed. In Rothlauf [7], pages 2389–2396.

[3] N. Hansen, A. Auger, S. Finck, and R. Ros. Real-parameter black-box optimization benchmarking 2010: Experimental setup. Technical Report RR-7215, INRIA, 2010.

[4] 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.

[5] 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.

[6] R. Ros. Benchmarking the NEWUOA on the BBOB-2009 function testbed. In Rothlauf [7], pages 2421–2428.

[7] F. Rothlauf, editor. *Genetic and Evolutionary Computation Conference, GECCO 2009, Proceedings, Montreal, Québec, Canada, July 8-12, 2009, Companion Material*. ACM, 2009.