

Example paper: Black-Box Optimization Benchmarking Template for the Comparison of Two Algorithms on the Noisy Testbed

Draft version *

BBOBies

ABSTRACT

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. RESULTS

Results from experiments according to [?] on the benchmark functions given in [?, ?] 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, $f_t = f_{\text{opt}} + \Delta f_t$, 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 [?, ?]. **Statistical significance** is tested with the rank-sum test for a given target Δf_t (10^{-8} as 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.

*Submission deadline: March 28th.

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'12, July 7–11, 2012, Philadelphia, USA.

Copyright 2012 ACM 978-1-4503-0073-5/10/07 ...\$10.00.

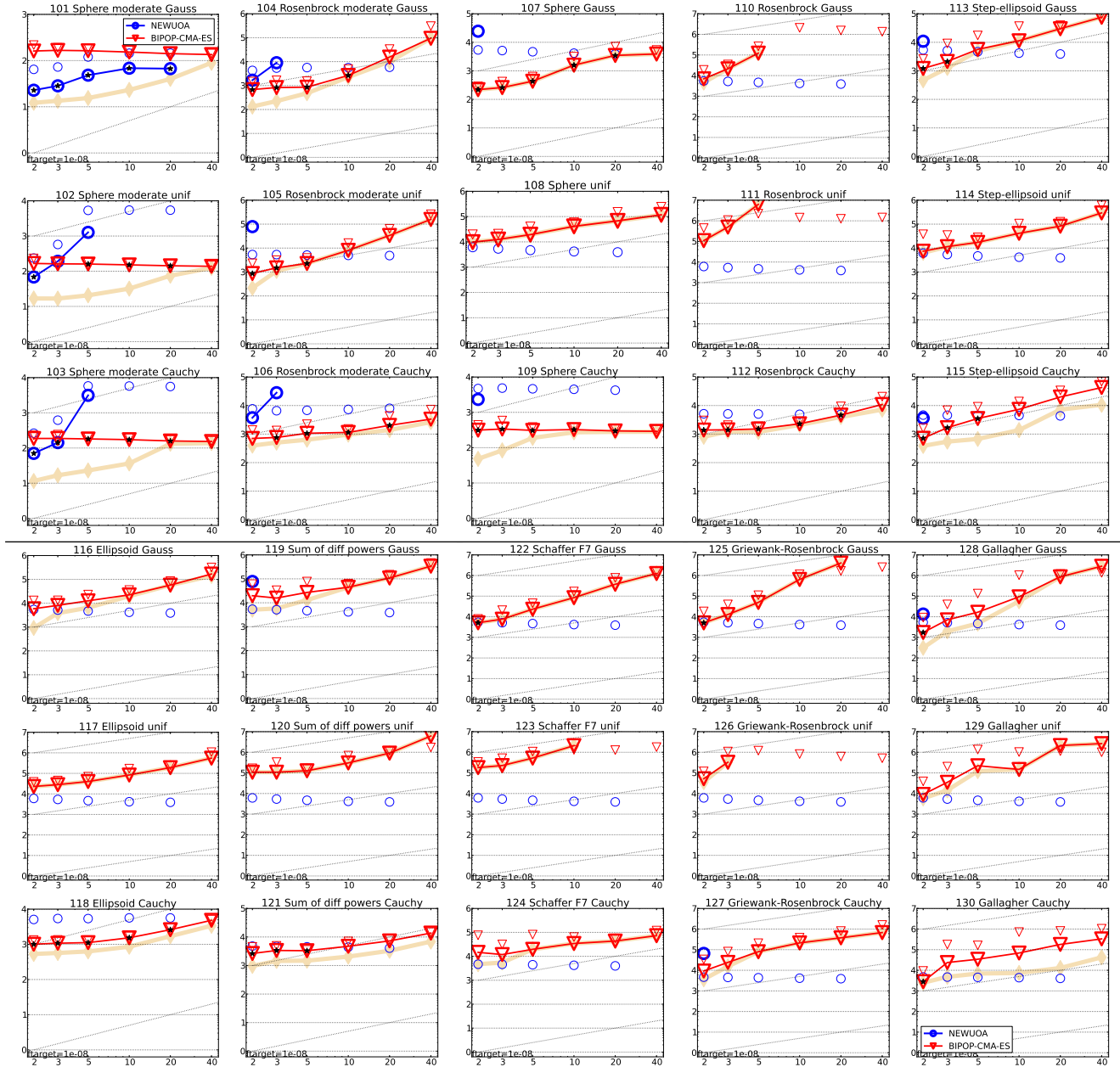


Figure 1: Expected running time (ERT in number of f -evaluations) divided by dimension for target function value 10^{-8} as \log_{10} values versus dimension. Different symbols correspond to different algorithms given in the legend of f_{101} and f_{130} . Light symbols give the maximum number of function 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 : NEWUOA, ∇ : BIPOP-CMA-ES.

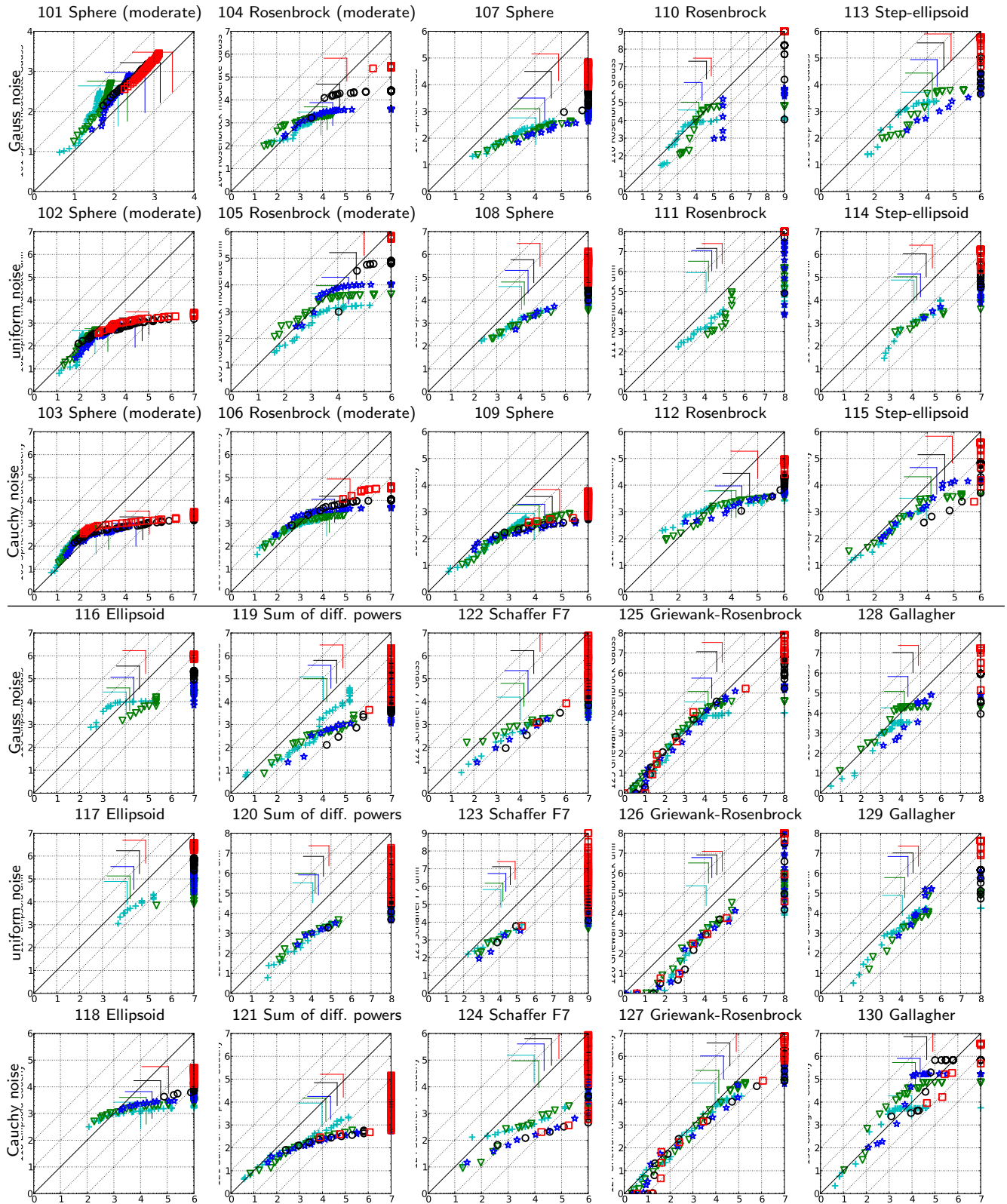


Figure 2: Expected running time (ERT in \log_{10} of number of function evaluations) of NEWUOA (x -axis) versus BIPOP-CMA-ES (y -axis) for 46 target values $\Delta f \in [10^{-8}, 10]$ in each dimension on functions f_{101} – f_{130} . Markers on the upper or right edge indicate that the target value was never reached. Markers represent dimension: 2:+, 3:∇, 5:*, 10:○, 20:□, 40:◇.

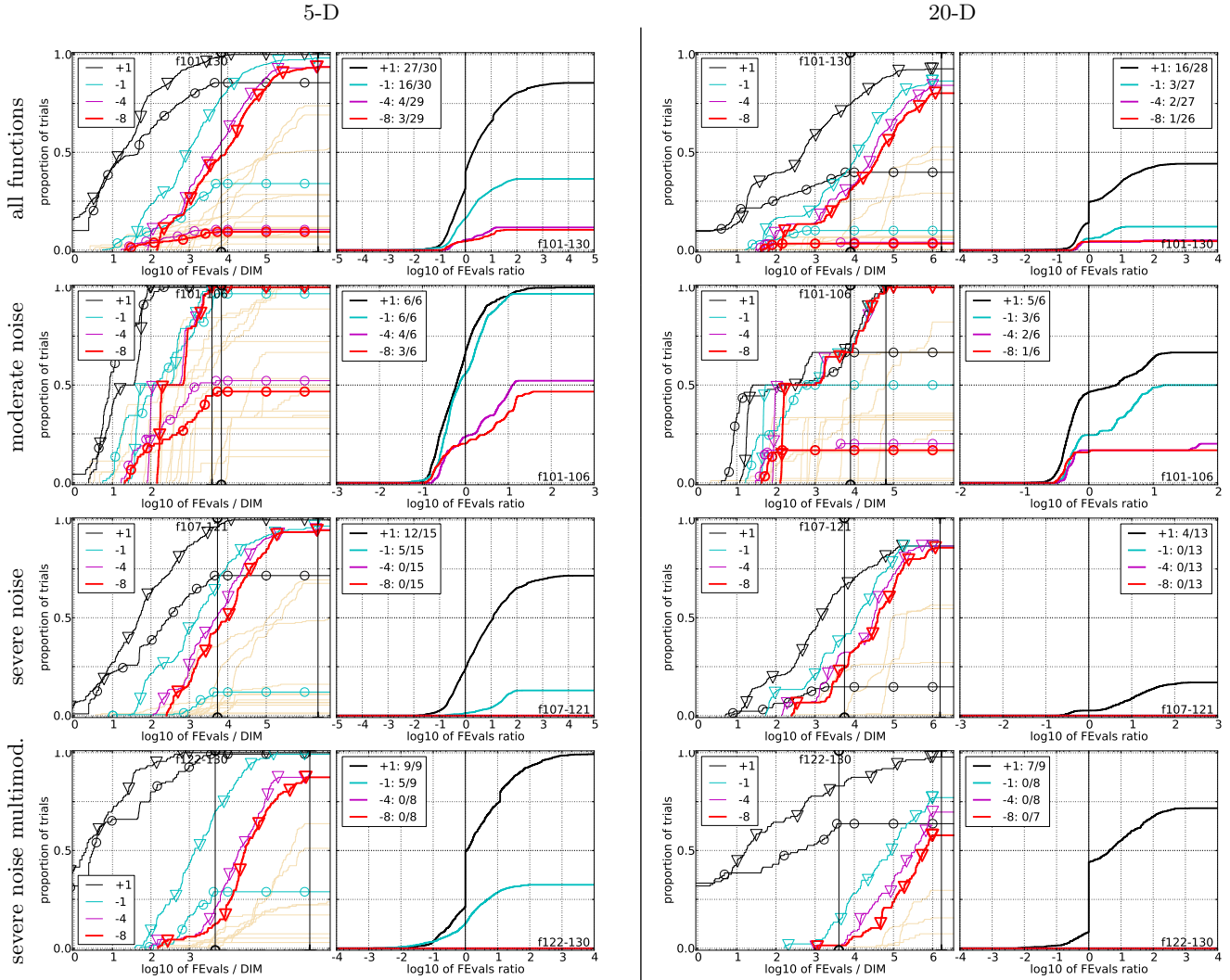


Figure 3: Empirical cumulative distributions (ECDF) of run lengths and speed-up ratios in 5-D (left) and 20-D (right). Left sub-columns: ECDF of the number of function evaluations divided by dimension D (FEvals/ D) to reach a target value $f_{opt} + \Delta f$ with $\Delta f = 10^k$, where $k \in \{1,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

5-D

20-D

Δf	1e+1	1e-1	1e-3	1e-5	1e-7	#succ	Δf	1e+1	1e-1	1e-3	1e-5	1e-7	#succ
f₁₀₁	11	44	62	69	75	15/15	f₁₀₁	59	571	700	739	783	15/15
1: IPO	2.5(1)	2.1(0.9)*2	2.6(2)*2	3.0(2)*3	3.1(2)*4	15/15	1: IPO	3.1(0.8)*3	0.90(0.2)*4	1.1(0.2)*4	1.5(0.3)*2	1.6(0.4)*2	15/15
2: BIP	3.2(2)	4.6(0.9)	6.1(0.5)	8.0(0.4)	10(0.7)	15/15	2: BIP	6.1(1)	1.6(0.1)	2.1(0.1)	2.7(0.1)	3.3(0.2)	15/15
f₁₀₂	11	50	72	86	99	15/15	f₁₀₂	231	579	921	1157	1407	15/15
1: IPO	6.3(11)	7.0(9)	20(19)	33(32)	41(57)	15/15	1: IPO	2.9(4)	6.3(7)	45(49)	∞	∞	0/15
2: BIP	2.7(2)	4.0(0.6)	5.1(0.5)	6.3(0.5)	7.2(0.7)	15/15	2: BIP	1.6(0.3)	1.6(0.2)	1.6(0.1)*4	1.8(0.1)*4	1.8(0.1)*4	15/15
f₁₀₃	11	30	31	35	115	15/15	f₁₀₃	65	629	1313	1893	2464	14/15
1: IPO	2.4(1.0)	5.7(9)	60(66)	178(172)	136(175)	12/15	1: IPO	2.3(0.9)*4	5.9(6)	1231(1388)	∞	∞	0/15
2: BIP	3.5(4)	7.4(1)	13(1)	17(2)*2	6.9(0.9)*2	15/15	2: BIP	5.5(1)	1.5(0.1)	1.2(0.1)*4	1.2(0.1)*4	1.2(0.1)*4	15/15
f₁₀₄	173	1287	1768	2040	2284	15/15	f₁₀₄	23690	1.7e5	1.8e5	1.9e5	2.0e5	15/15
1: IPO	1.2(2)	6.0(8)	24(24)	∞	∞	0/15	1: IPO	68(74)	∞	∞	∞	∞	0/15
2: BIP	1.4(0.3)	2.0(0.3)	2.0(0.2)*2	1.9(0.2)*4	1.8(0.2)*4	15/15	2: BIP	10(7)	1.7(1)	1.6(1)	1.6(1.0)	1.6(0.9)	15/15
f₁₀₅	167	5174	10388	10824	11202	15/15	f₁₀₅	1.9e5	6.3e5	6.5e5	6.6e5	6.7e5	15/15
1: IPO	1.7(2)	3.3(4)	∞	∞	∞	0/15	1: IPO	∞	∞	∞	∞	∞	0/15
2: BIP	1.7(0.4)	1.7(0.9)	1(0.4)*4	1(0.4)*4	1(0.4)*4	15/15	2: BIP	2.7(2)	1(0.6)	1(0.6)	1(0.6)	1(0.6)	15/15
f₁₀₆	92	1050	2666	2887	3087	15/15	f₁₀₆	11480	23746	25470	26492	27360	15/15
1: IPO	0.93(0.7)*4	5.0(6)	59(56)	∞	∞	0/15	1: IPO	7.0(5)	∞	∞	∞	∞	0/15
2: BIP	3.3(0.9)	3.2(3)	1.6(1)*4	1.7(1)*4	1.7(1)*4	15/15	2: BIP	1.0(0.3)*3	1.4(1)*4	1.5(1)*4	1.5(1)*4	1.5(1)*4	15/15
f₁₀₇	40	453	940	1376	1850	15/15	f₁₀₇	8571	16226	27357	52486	65052	15/15
1: IPO	60(50)	∞	∞	∞	∞	0/15	1: IPO	∞	∞	∞	∞	∞	0/15
2: BIP	1.7(2)*	1(0.5)*4	1(0.3)*4	1(0.2)*4	1(0.2)*4	15/15	2: BIP	1(0.4)*4	1(0.6)*4	1(0.4)*4	1(0.8)*4	1(0.8)*4	15/15
f₁₀₈	87	14469	30935	58628	80667	15/15	f₁₀₈	58063	2.0e5	4.5e5	6.3e5	9.0e5	15/15
1: IPO	77(89)	∞	∞	∞	∞	0/15	1: IPO	∞	∞	∞	∞	∞	0/15
2: BIP	6.1(10)*2	1(0.8)*4	1(0.6)	1(0.4)	1(0.3)	15/15	2: BIP	1(0.5)*4	1(0.5)	1(0.5)	1(0.5)	1(0.4)	15/15
f₁₀₉	11	216	572	873	946	15/15	f₁₀₉	333	1138	2287	3583	4952	15/15
1: IPO	4.8(10)	83(105)	∞	∞	∞	0/15	1: IPO	17(23)	∞	∞	∞	∞	0/15
2: BIP	3.5(2)	1.1(0.3)*4	1.1(0.2)*4	1.1(0.3)*4	1.5(0.3)*4	15/15	2: BIP	1.2(0.3)	1.1(0.2)*4	1.1(0.1)*4	1.1(0.1)*4	1.0(0.1)*4	15/15
f₁₁₀	949	1.2e5	5.9e5	6.0e5	6.1e5	15/15	f₁₁₀	∞	∞	∞	∞	∞	0
1: IPO	118(124)	∞	∞	∞	∞	0/15	1: IPO	∞	∞	∞	∞	∞	0/15
2: BIP	1(1)*4	3.7(4)	1(0.7)	1(0.7)	1(0.6)	15/15	2: BIP	∞	∞	∞	∞	∞	0/15
f₁₁₁	6856	8.8e6	2.3e7	3.1e7	3.1e7	3/15	f₁₁₁	∞	∞	∞	∞	∞	0
1: IPO	∞	∞	∞	∞	∞	0/15	1: IPO	∞	∞	∞	∞	∞	0/15
2: BIP	1(1.0)*3	1(1)	1(0.9)	1(0.9)	1(1)	3/15	2: BIP	∞	∞	∞	∞	∞	0/15
f₁₁₂	107	3421	4502	5132	5596	15/15	f₁₁₂	25552	69621	73557	76137	78238	15/15
1: IPO	1.9(3)	105(122)	∞	∞	∞	0/15	1: IPO	∞	∞	∞	∞	∞	0/15
2: BIP	4.0(2)	1.2(0.2)*4	1.3(0.2)*4	1.3(0.2)*4	1.3(0.2)*4	15/15	2: BIP	1(0.3)*4	1.1(0.8)*4	1.2(0.7)*4	1.2(0.7)*4	1.2(0.7)*4	15/15
f₁₁₃	133	8081	24128	24128	24402	15/15	f₁₁₃	50123	5.6e5	5.9e5	5.9e5	5.9e5	15/15
1: IPO	13(16)	∞	∞	∞	∞	0/15	1: IPO	∞	∞	∞	∞	∞	0/15
2: BIP	1.5(1.0)*	1.7(2)*4	1.1(1)	1.1(1)	1.1(1)	15/15	2: BIP	1(1.0)*4	1(0.4)	1(0.4)	1(0.4)	1(0.4)	15/15
f₁₁₄	767	56311	83272	83272	84949	15/15	f₁₁₄	2.1e5	1.4e6	1.6e6	1.6e6	1.6e6	15/15
1: IPO	43(47)	∞	∞	∞	∞	0/15	1: IPO	∞	∞	∞	∞	∞	0/15
2: BIP	2.2(2)*3	1(0.7)	1(0.7)	1(0.7)	1(0.7)	15/15	2: BIP	1(0.4)	1(0.5)	1(0.5)	1(0.5)	1(0.5)	15/15
f₁₁₅	64	1829	2550	2550	2970	15/15	f₁₁₅	2405	91749	1.3e5	1.3e5	1.3e5	15/15
1: IPO	2.9(3)	42(44)	∞	∞	∞	0/15	1: IPO	236(283)	∞	∞	∞	∞	0/15
2: BIP	1.5(0.8)	6.5(7)*	5.9(6)*	5.9(6)*	5.7(5)*	15/15	2: BIP	1(1)*4	3.9(2)	3.0(1)	3.0(1)	3.0(2)	15/15

Table 1: ERT in number of function evaluations divided by the best ERT measured during BBOB-2009 given in the respective first row and the half inter-80%ile in brackets for different Δf values. #succ is the number of trials that reached the final target $f_{\text{opt}} + 10^{-8}$. 1:IPO is NEWUOA and 2:BIP is BIPOP-CMA-ES. Bold entries are statistically significantly better compared to the other algorithm, with $p = 0.05$ or $p = 10^{-k}$ where $k \in \{2, 3, 4, \dots\}$ is the number following the \star symbol, with Bonferroni correction of 60. A \downarrow indicates the same tested against the best BBOB-2009.

5-D

20-D

Δf	1e+1	1e-1	1e-3	1e-5	1e-7	#succ	Δf	1e+1	1e-1	1e-3	1e-5	1e-7	#succ
f₁₁₆	5730	22311	26868	30329	31661	15/15	f₁₁₆	5.0e5	8.9e5	1.0e6	1.1e6	1.1e6	15/15
1: IPO	∞	∞	∞	∞	∞	∞	0/15	1: IPO	∞	∞	∞	∞	0/15
2: BIP	1.2(1)*⁴	1.9(2)	2.1(2)	2.0(2)	2.0(2)	15/15	2: BIP	1.4(0.9)	1.1(0.5)	1(0.4)	1(0.4)	1(0.4)	15/15
f₁₁₇	26686	1.1e5	1.4e5	1.7e5	1.9e5	15/15	f₁₁₇	1.8e6	2.6e6	2.9e6	3.2e6	3.6e6	15/15
1: IPO	∞	∞	∞	∞	∞	∞	0/15	1: IPO	∞	∞	∞	∞	0/15
2: BIP	1(0.7)	1(0.7)	1(0.6)	1(0.6)	1(0.5)	15/15	2: BIP	1(0.5)	1(0.2)	1(0.2)	1(0.2)	1(0.2)	15/15
f₁₁₈	429	1555	1998	2430	2913	15/15	f₁₁₈	6908	17514	26342	30062	32659	15/15
1: IPO	4.3(6)	116(122)	∞	∞	∞	∞	0/15	1: IPO	∞	∞	∞	∞	0/15
2: BIP	3.2(1)	1.9(0.7)*⁴	2.1(0.4)*⁴	2.0(0.4)*⁴	1.8(0.3)*⁴	15/15	2: BIP	1.9(0.4)*⁴	1.6(0.2)*⁴	1.5(0.1)*⁴	1.6(0.1)*⁴	1.6(0.1)*⁴	15/15
f₁₁₉	12	1136	10372	35296	49747	15/15	f₁₁₉	2771	35930	4.1e5	1.4e6	1.9e6	15/15
1: IPO	26(40)	∞	∞	∞	∞	∞	0/15	1: IPO	398(435)	∞	∞	∞	0/15
2: BIP	1.9(3)	1(2)*⁴	1(0.6)*⁴	1.5(0.8)	2.3(1)	15/15	2: BIP	1.6(1)*⁴	1(1)*⁴	1(0.5)	1.3(0.3)	1.1(0.2)	15/15
f₁₂₀	16	18698	72438	3.3e5	5.5e5	15/15	f₁₂₀	36040	2.8e5	1.6e6	6.7e6	1.4e7	13/15
1: IPO	130(196)	∞	∞	∞	∞	∞	0/15	1: IPO	∞	∞	∞	∞	0/15
2: BIP	17(16)*	1(0.6)*⁴	1(0.8)	1(0.5)	1(0.4)	15/15	2: BIP	1(0.6)*⁴	1(0.6)	1(0.6)	1(0.4)	1(0.4)	13/15
f₁₂₁	8.6	273	1583	3870	6195	15/15	f₁₂₁	249	1426	9304	34434	57404	15/15
1: IPO	4.8(13)	76(75)	∞	∞	∞	∞	0/15	1: IPO	31(63)	∞	∞	∞	0/15
2: BIP	2.7(3)	1(0.2)*²	1.1(0.5)*⁴	2.0(0.2)*⁴	2.2(0.2)*⁴	15/15	2: BIP	1.2(0.5)*²	1.2(0.3)*⁴	1.1(0.2)*⁴	1.3(0.1)*⁴	1.9(0.1)	15/15
f₁₂₂	10	9190	30087	53743	1.1e5	15/15	f₁₂₂	692	1.4e5	7.9e5	2.0e6	5.8e6	15/15
1: IPO	14(23)	∞	∞	∞	∞	∞	0/15	1: IPO	82(115)	∞	∞	∞	0/15
2: BIP	2.2(2)	1(0.8)*⁴	1(0.5)	1(0.6)	1(0.6)	15/15	2: BIP	1.8(2)*	1(0.7)	1(0.7)	1(0.5)	1(0.8)	15/15
f₁₂₃	11	81505	3.4e5	6.7e5	2.2e6	15/15	f₁₂₃	1063	1.5e6	5.3e6	2.7e7	1.6e8	0
1: IPO	65(85)	∞	∞	∞	∞	∞	0/15	1: IPO	174(196)	∞	∞	∞	0/15
2: BIP	8.1(11)	1(0.6)	1(0.6)	1(0.6)	1(0.9)	15/15	2: BIP	5.7(4)*²	1(0.7)	1(0.6)	1(0.8)	1(1)	0/15
f₁₂₄	10	1040	20478	45337	95200	15/15	f₁₂₄	192	40840	1.3e5	3.9e5	8.0e5	15/15
1: IPO	3.0(1)	∞	∞	∞	∞	∞	0/15	1: IPO	91(187)	∞	∞	∞	0/15
2: BIP	1.5(2)	1(0.3)*⁴	1.1(0.7)*⁴	1.2(1.0)	1(0.5)	15/15	2: BIP	1.1(0.5)	1(1.0)*⁴	1(0.9)	1(0.8)	1(0.4)	15/15
f₁₂₅	1	1	2.4e5	2.4e5	2.5e5	15/15	f₁₂₅	1	1	2.5e7	8.0e7	8.1e7	4/15
1: IPO	3.9(6)	6088(8822)	∞	∞	∞	∞	0/15	1: IPO	1	∞	∞	∞	0/15
2: BIP	1.1	3443(2609)	1(0.7)	1(0.7)	1(0.7)	15/15	2: BIP	1	9.8e6(7e6)	1(0.9)	1(1)	1(1)	4/15
f₁₂₆	1	1	∞	∞	∞	0	f₁₂₆	1	1	∞	∞	∞	0
1: IPO	1.2	3.5e5(4e5)	∞	∞	∞	0/15	1: IPO	4.2	∞	∞	∞	∞	0/15
2: BIP	1	13292(10642)*³	∞	∞	∞	0/15	2: BIP	1	∞	∞	∞	∞	0/15
f₁₂₇	1	1	3.4e5	3.9e5	4.0e5	15/15	f₁₂₇	1	1	4.4e6	7.3e6	7.4e6	15/15
1: IPO	2.5(6)	7248(10060)	∞	∞	∞	∞	0/15	1: IPO	3.7	∞	∞	∞	0/15
2: BIP	1	2136(1530)	1(1.0)	1(0.8)	1(0.8)	15/15	2: BIP	1	9.0e5(1e6)	1(0.6)	1(0.7)	1(0.7)	15/15
f₁₂₈	111	7808	12447	17217	21162	15/15	f₁₂₈	1.4e5	1.7e7	1.7e7	1.7e7	1.7e7	9/15
1: IPO	12(22)	43(51)	∞	∞	∞	∞	0/15	1: IPO	∞	∞	∞	∞	0/15
2: BIP	2.2(2)	10(17)	6.6(11)	4.8(8)	3.9(6)	15/15	2: BIP	1(2)	1(1)	1(1)	1(1)	1(1)	9/15
f₁₂₉	64	59443	2.8e5	5.1e5	5.8e5	15/15	f₁₂₉	7.8e6	4.2e7	4.2e7	4.2e7	4.2e7	5/15
1: IPO	124(124)	∞	∞	∞	∞	∞	0/15	1: IPO	∞	∞	∞	∞	0/15
2: BIP	12(15)*³	9.2(2)	3.9(12)	2.2(7)	1.9(6)	13/15	2: BIP	1(1)	1(1)	1(1)	1(1)	1(1)	5/15
f₁₃₀	55	3034	32823	33889	34528	10/15	f₁₃₀	4904	2.5e5	2.5e5	2.6e5	2.6e5	7/15
1: IPO	2.3(4)	10(11)	∞	∞	∞	∞	0/15	1: IPO	9.1(9)	∞	∞	∞	0/15
2: BIP	1.9(1)	55(101)	5.1(9)	5.0(9)	5.0(9)	15/15	2: BIP	1.9(4)*	14(28)	14(27)	14(27)	14(27)	15/15

Table 2: Relative ERT in number of f -evaluations, see Table 1 for details.