Example paper: Black-Box Optimization Benchmarking Template for the Comparison of Two Algorithms on the Noiseless 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. The **expected running time (ERT)**, used in the figures and table, depends on a given target function value, $f_\mathrm{t} = f_\mathrm{opt} + \Delta f$, 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 [**?**, **?**]. **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.

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

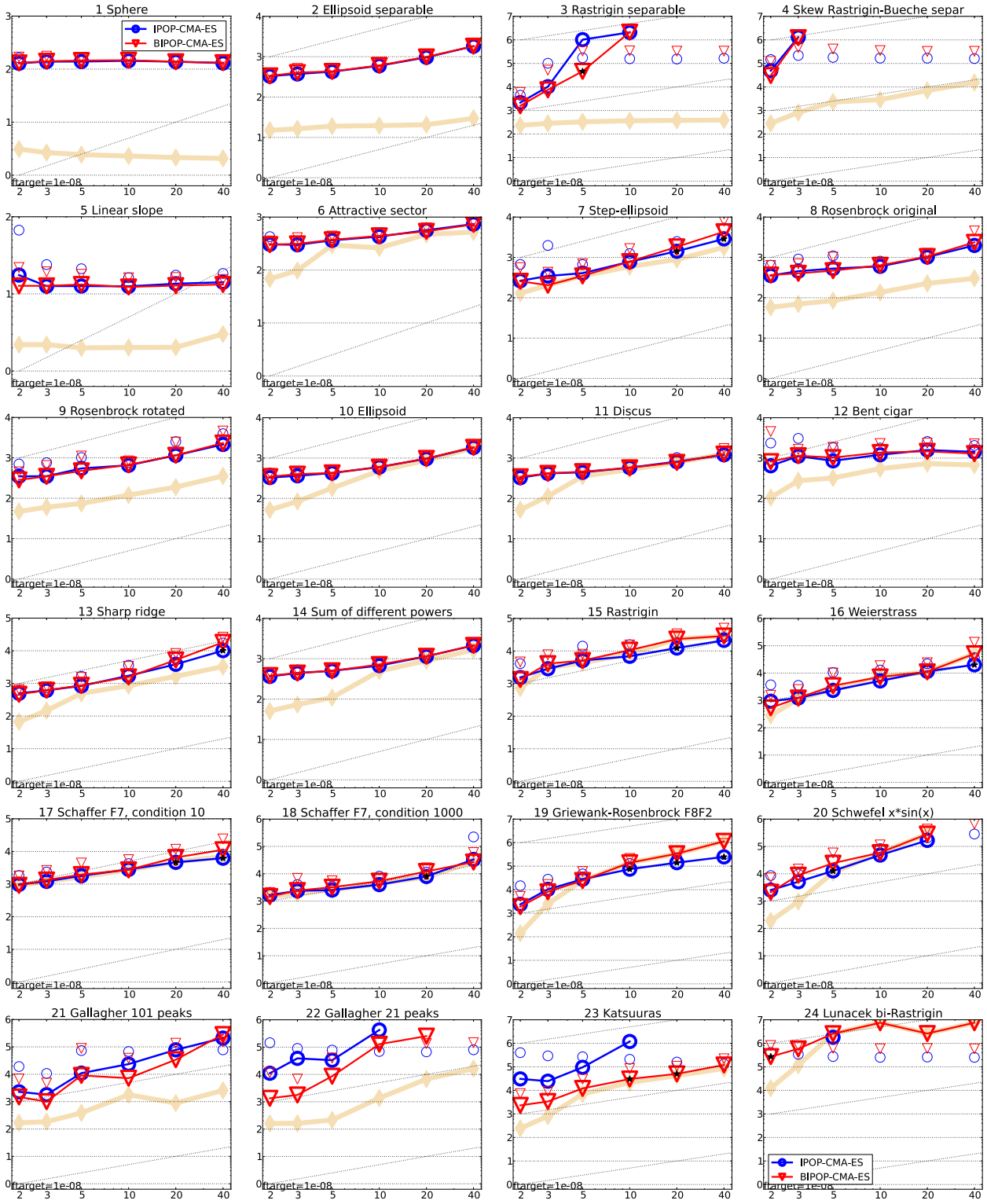

Figure 1: Expected running time (ERT in number of f-evaluations) divided by dimension for target function value 10^{-8} as $\log_{10}$ values versus dimension. Different symbols correspond to different algorithms given in the legend of f_1 and f_{24}. Light symbols give the maximum number of function 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: ○: IPOP-CMA-ES, ▽: BIPOP-CMA-ES.

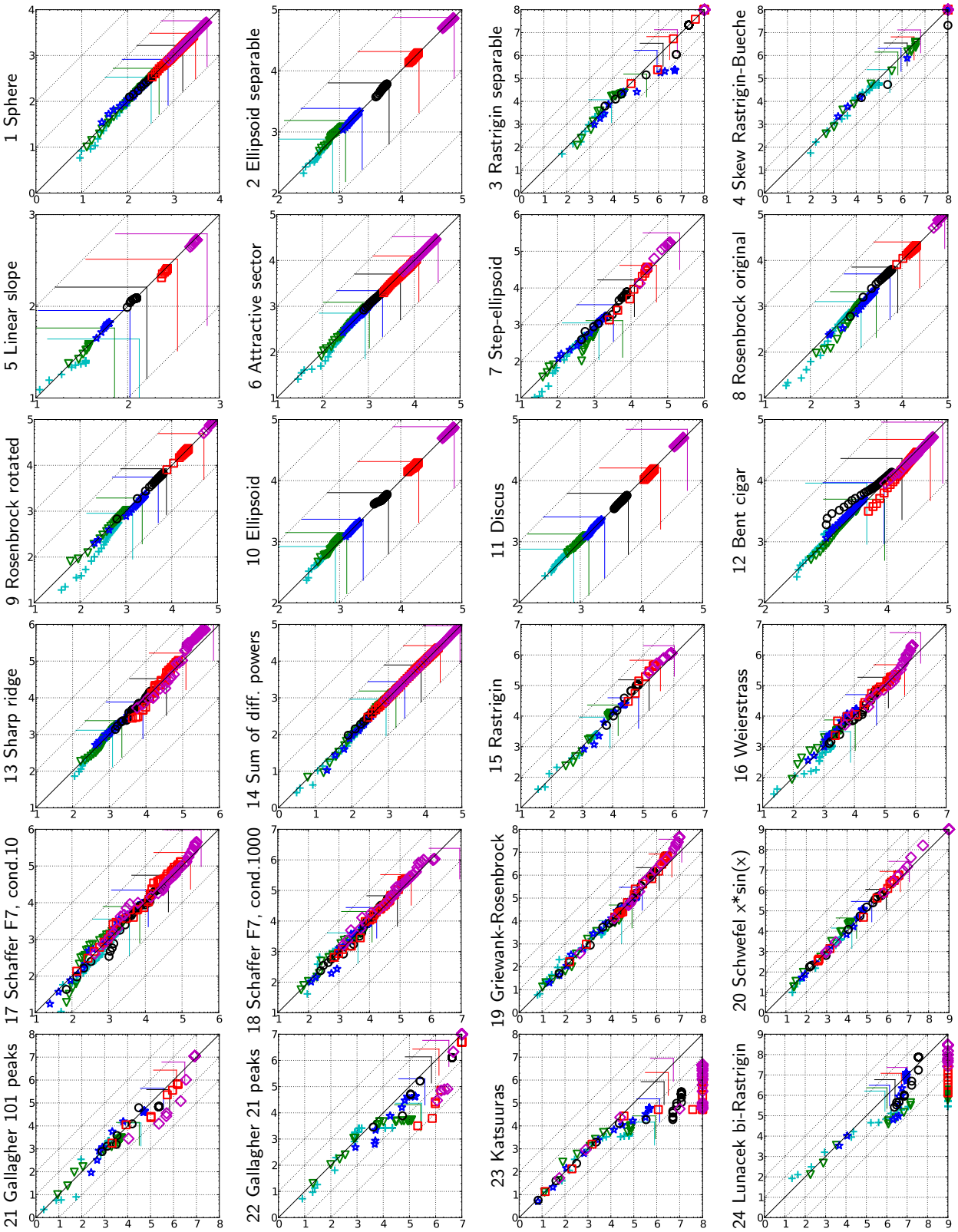

Figure 2: Expected running time (ERT in $\log_{10}$ **of number of function evaluations) of IPOP-CMA-ES** (x**-axis) versus BIPOP-CMA-ES** (y**-axis) for 46 target values** $\Delta f \in [10^{-8}, 10]$ **in each dimension on functions** f_1–f_{24}**. Markers on the upper or right edge indicate that the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

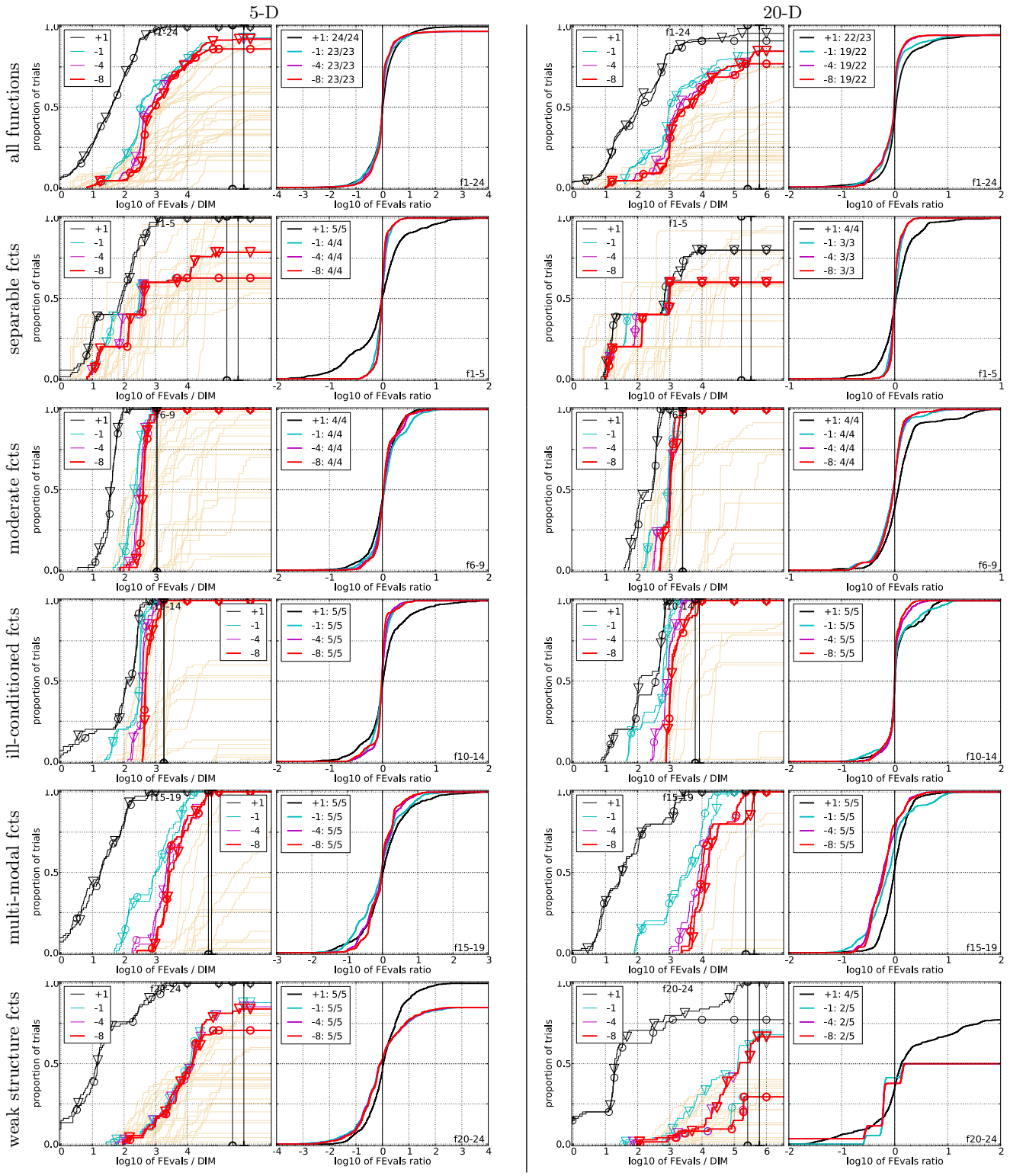

Figure 3: Empirical cumulative distributions (ECDF) of run lengths and speed-up ratios in 5-D (left) and 20-D (right). Left sub-columns: ECDF of the number of function evaluations divided by dimension D (FEvals/D) to reach a target value $f_{\mathrm{opt}} + \Delta f$ with $\Delta f = 10^k$, where $k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

5-D

Δf	1e+1	1e-1	1e-3	1e-5	1e-7	#succ
f_1	11	12	12	12	12	15/15
1: IPO	2.5(2)	14(3)	27(3)	39(3)	51(3)	15/15
2: BIP	3.2(2)	15(4)	27(5)	40(4)	53(6)	15/15
f_2	83	88	90	92	94	15/15
1: IPO	14(4)	18(2)	19(2)	21(2)	22(2)	15/15
2: BIP	13(4)	18(2)	20(2)	21(2)	22(2)	15/15
f_3	716	1637	1646	1650	1654	15/15
1: IPO	2.2(4)	3130(3775)	3113(3800)	3106(3831)	3099(3779)	2/15
2: BIP	1.4(1)	139(107)	139(107)	139(107)	140(107)	14/15
f_4	809	1688	1817	1886	1903	15/15
1: IPO	2.0(3)	∞	∞	∞	∞*8.5e5*	0/15
2: BIP	2.7(3)	∞	∞	∞	∞*1.8e6*	0/15
f_5	10	10	10	10	10	15/15
1: IPO	4.6(2)	6.3(3)	6.3(3)	6.3(3)	6.3(3)	15/15
2: BIP	4.5(2)	6.6(2)	6.6(2)	6.6(2)	6.6(2)	15/15
f_6	114	281	580	1038	1332	15/15
1: IPO	2.5(0.9)	2.2(0.4)	1.7(0.2)	1.3(0.2)	1.2(0.1)	15/15
2: BIP	2.3(1)	2.2(0.6)	1.7(0.2)	1.3(0.3)	1.3(0.2)	15/15
f_7	24	1171	1572	1572	1597	15/15
1: IPO	4.4(3)	1.2(0.9)	1.2(0.6)	1.2(0.6)	1.2(0.6)	15/15
2: BIP	5.0(5)	1(1)	1(0.9)	1(0.9)	1(0.9)	15/15
f_8	73	336	391	410	422	15/15
1: IPO	3.5(2)	5.3(4)	5.6(4)	5.8(3)	6.1(3)	15/15
2: BIP	3.2(1)	4.5(2)	4.8(2)	5.1(2)	5.4(2)	15/15
f_9	35	214	300	335	369	15/15
1: IPO	6.0(2)	8.7(6)	7.5(5)	7.3(4)	7.2(4)	15/15
2: BIP	5.8(2)	7.2(2)	6.4(2)	6.3(1)	6.2(1)	15/15
f_{10}	349	574	626	829	880	15/15
1: IPO	3.6(1.0)	2.7(0.3)	2.8(0.3)	2.3(0.2)	2.3(0.2)	15/15
2: BIP	3.5(0.8)	2.7(0.4)	2.8(0.2)	2.3(0.2)	2.4(0.1)	15/15
f_{11}	143	763	1177	1467	1673	15/15
1: IPO	8.6(2)	2.1(0.2)	1.6(0.1)	1.4(0.1)	1.3(0.1)	15/15
2: BIP	8.4(3)	2.2(0.3)	1.6(0.2)	1.4(0.1)	1.3(0.1)	15/15
f_{12}	108	371	461	1303	1494	15/15
1: IPO	9.4(7)	6.2(4)	6.3(4)	2.8(2)	2.8(2)	15/15
2: BIP	11(12)	7.4(6)	7.7(5)	3.3(2)	3.3(2)	15/15
f_{13}	132	250	1310	1752	2255	15/15
1: IPO	3.1(1)	5.3(3)	1.4(0.6)	1.6(0.3)	1.6(0.6)	15/15
2: BIP	3.9(3)	5.9(3)	1.6(0.3)	1.5(0.2)	1.7(0.8)	15/15
f_{14}	10	58	139	251	476	15/15
1: IPO	2.2(3)	3.8(1)	4.7(1)	5.4(0.7)	4.4(0.5)	15/15
2: BIP	1.1(1.0)	3.7(0.9)	4.6(0.7)	5.4(0.5)	4.5(0.3)	15/15
f_{15}	511	19369	20073	20769	21359	14/15
1: IPO	2.3(3)	1.2(1)	1.2(1)	1.2(1)	1.2(1)	15/15
2: BIP	1.6(2)	1.2(0.7)	1.2(0.7)	1.2(0.7)	1.2(0.7)	15/15
f_{16}	120	2662	10449	11644	12095	15/15
1: IPO	2.5(2)	1.7(1)	0.96(0.9)	0.94(0.7)	0.95(0.7)	15/15
2: BIP	3.0(3)	2.6(1)	1.3(2)	1.4(2)	1.4(2)	15/15
f_{17}	5.2	899	3669	6351	7934	15/15
1: IPO	4.8(6)	0.97(2)	0.77(1)	0.81(0.7)	1.0(0.4)	15/15
2: BIP	3.4(3)	1(2)	1(0.7)	1(0.5)	1.2(0.5)	15/15
f_{18}	103	3968	9280	10905	12469	15/15
1: IPO	1.2(0.5)	0.87(1)	1.0(0.4)	1.0(0.3)	0.99(0.1)	15/15
2: BIP	1(0.7)	1(1)	1(0.3)	1.2(0.7)	1.3(0.6)	15/15
f_{19}	1	242	1.2e5	1.2e5	1.2e5	15/15
1: IPO	21(25)	125(94)	1.1(0.7)	1.1(0.7)	1.1(0.7)	15/15
2: BIP	20(16)	161(175)	1(0.7)	1(0.7)	1(0.7)	15/15
f_{20}	16	38111	54470	54861	55313	14/15
1: IPO	3.9(2)	1.4(0.9)	1.1(0.7)	1.1(0.7)	1.1(0.7)	15/15
2: BIP	3.3(3)	2.8(1)	2.1(0.8)	2.2(0.8)	2.2(0.8)	15/15
f_{21}	41	1674	1705	1729	1757	14/15
1: IPO	6.3(17)	30(34)	31(38)	31(41)	31(40)	14/15
2: BIP	2.3(2)	24(35)	25(36)	25(36)	25(36)	15/15
f_{22}	71	938	1008	1040	1068	14/15
1: IPO	12(26)	166(222)	161(211)	158(215)	155(207)	11/15
2: BIP	6.9(11)	45(94)	42(88)	41(85)	40(83)	15/15
f_{23}	3.0	14249	31654	33030	34256	15/15
1: IPO	2.2(2)	33(50)	15(22)	14(20)	14(20)	11/15
2: BIP	1.7(2)	3.7(4)	1.8(2)	1.8(2)	1.8(2)	15/15
f_{24}	1622	6.4e6	9.6e6	1.3e7	1.3e7	3/15
1: IPO	2.9(3)	1.4(2)	0.94(1)	0.70(0.8)	0.70(0.9)	2/15
2: BIP	2.1(2)	1(1.0)	1(1)	1(1)	1(1)	3/15

20-D

Δf	1e+1	1e-1	1e-3	1e-5	1e-7	#succ
f_1	43	43	43	43	43	15/15
1: IPO	8.0(1)	20(2)	33(1)	46(2)	58(2)	15/15
2: BIP	7.9(2)	20(2)	33(4)	45(3)	57(3)	15/15
f_2	385	387	390	391	393	15/15
1: IPO	35(4)	43(3)	45(3)	47(2)	48(2)	15/15
2: BIP	35(7)	44(4)	47(2)	48(2)	50(2)	15/15
f_3	5066	7635	7643	7646	7651	15/15
1: IPO	13(6)	∞	∞	∞	∞*2.9e6*	0/15
2: BIP	12(7)	∞	∞	∞	∞*5.7e6*	0/15
f_4	4722	7666	7700	7758	1.4e5	9/15
1: IPO	∞	∞	∞	∞	∞*2.8e6*	0/15
2: BIP	∞	∞	∞	∞	∞*5.5e6*	0/15
f_5	41	41	41	41	41	15/15
1: IPO	5.8(1)	6.7(1.0)	6.7(1.0)	6.7(1.0)	6.7(1.0)	15/15
2: BIP	5.1(0.8)	6.3(1)	6.3(1)	6.3(1)	6.3(1)	15/15
f_6	1296	3413	5220	6728	8409	15/15
1: IPO	1.7(0.3)	1.2(0.1)	1.2(0.1)	1.2(0.1)	1.2(0.1)	15/15
2: BIP	1.5(0.4)	1.2(0.2)	1.1(0.2)	1.2(0.1)	1.2(0.1)	15/15
f_7	1351	9503	16524	16524	16969	15/15
1: IPO	1.9(2)	2.7(2)	**1.7**(1.0)*	**1.7**(1.0)*	**1.6**(0.9)*	15/15
2: BIP	1(0.5)	3.5(0.6)	2.2(0.3)	2.2(0.3)	2.1(0.3)	15/15
f_8	2039	4040	4219	4371	4484	15/15
1: IPO	3.7(0.9)	4.2(0.5)	4.4(0.4)	4.4(0.4)	4.5(0.4)	15/15
2: BIP	4.0(1)	4.3(0.6)	4.5(0.6)	4.6(0.6)	4.6(0.6)	15/15
f_9	1716	3277	3455	3594	3727	15/15
1: IPO	4.6(0.8)	6.0(0.4)	6.1(0.4)	6.1(0.4)	6.1(0.3)	15/15
2: BIP	4.7(2)	6.0(1)	6.1(1)	6.1(1.0)	6.1(0.9)	15/15
f_{10}	7413	10735	14920	17073	17476	15/15
1: IPO	1.8(0.2)	1.5(0.1)	1.2(0.1)	1.1(0.0)	1.1(0.0)	15/15
2: BIP	1.9(0.2)	1.6(0.1)	1.2(0.0)	1.1(0.0)	1.1(0.0)	15/15
f_{11}	1002	6278	9762	12285	14831	15/15
1: IPO	11(2)	2.1(0.3)	1.4(0.2)	1.2(0.2)	1.1(0.1)	15/15
2: BIP	10(0.5)	1.9(0.1)	1.4(0.0)	1.2(0.0)	1.0(0.0)	15/15
f_{12}	1042	2740	4140	12407	13827	15/15
1: IPO	4.8(4)	5.5(5)	5.1(3)	2.1(1.0)	2.2(0.9)	15/15
2: BIP	3.0(2)	4.5(3)	4.5(2)	1.9(0.7)	2.0(0.7)	15/15
f_{13}	652	2751	18749	24455	30201	15/15
1: IPO	6.5(5)	6.2(5)	1.4(0.8)	1.7(0.8)	2.3(1)	15/15
2: BIP	4.3(6)	5.1(6)	1.5(0.8)	2.3(2)	3.0(2)	15/15
f_{14}	75	304	932	1648	15661	15/15
1: IPO	3.7(2)	3.6(0.5)	3.9(0.5)	6.0(0.6)	1.2(0.1)	15/15
2: BIP	3.9(1)	3.7(0.4)	4.1(0.3)	6.2(0.5)	1.2(0.1)	15/15
f_{15}	30378	3.1e5	3.2e5	4.5e5	4.6e5	15/15
1: IPO	1.1(0.7)	**0.69**(0.4)*	**0.70**(0.4)*	**0.52**(0.3)$^{\star\downarrow}$	**0.53**(0.3)$^{\star\downarrow}$	15/15
2: BIP	1(0.4)	1.4(0.5)	1.4(0.5)	1(0.3)	1(0.3)	15/15
f_{16}	1384	77015	1.9e5	2.0e5	2.2e5	15/15
1: IPO	1.7(0.8)	0.92(0.6)	0.84(0.4)	1.1(0.7)	1.0(0.6)	15/15
2: BIP	1.7(0.4)	1.2(0.7)	1(0.7)	1(0.7)	1(0.7)	15/15
f_{17}	63	4005	30677	56288	80472	15/15
1: IPO	2.1(1)	1.2(2)	0.76(0.6)	0.99(0.3)	1.0(0.7)	15/15
2: BIP	2.2(2)	1(1)	1.2(1)	1.3(0.6)	1.4(0.7)	15/15
f_{18}	621	19561	67569	1.3e5	1.5e5	15/15
1: IPO	1.1(0.3)	1.1(0.6)	0.97(0.7)	**1.0**(0.4)*	**1.1**(0.4)$^{\star 2}$	15/15
2: BIP	1.0(0.4)	1.2(0.9)	1.1(0.6)	1.7(0.7)	1.6(0.6)	15/15
f_{19}	1	3.4e5	6.2e6	6.7e6	6.7e6	15/15
1: IPO	161(86)	0.71(0.5)	**0.38**(0.1)$^{\star 3\downarrow 4}$	**0.41**(0.2)$^{\star 3\downarrow 3}$	**0.41**(0.2)$^{\star 3\downarrow 3}$	15/15
2: BIP	169(74)	1.2(0.6)	1(0.3)	1(0.3)	1(0.3)	15/15
f_{20}	82	3.1e6	5.5e6	5.6e6	5.6e6	14/15
1: IPO	4.6(1)	0.65(0.3)	0.57(0.2)	0.58(0.2)	0.58(0.2)	15/15
2: BIP	4.3(1)	1(0.5)	1(0.3)	1(0.3)	1(0.3)	14/15
f_{21}	561	14103	14643	15567	17589	15/15
1: IPO	3.7(5)	110(142)	106(135)	100(127)	88(113)	7/15
2: BIP	3.2(6)	48(87)	46(90)	43(79)	39(77)	13/15
f_{22}	467	23491	24948	26847	1.3e5	12/15
1: IPO	445(1384)	∞	∞	∞	∞*1.3e6*	0/15
2: BIP	6.8(13)	215(274)	202(241)	188(231)	37(43)	5/15
f_{23}	3.2	67457	4.9e5	8.1e5	8.4e5	15/15
1: IPO	4.3(3)	∞	∞	∞	∞*2.5e6*	0/15
2: BIP	4.3(5)	**1**(0.8)$^{\star 3}$	**2.0**(1)$^{\star 3}$	**1.2**(0.9)$^{\star 3}$	**1.2**(0.9)$^{\star 3}$	15/15
f_{24}	1.3e6	5.2e7	5.2e7	5.2e7	5.2e7	3/15
1: IPO	∞	∞	∞	∞	∞*5.1e6*	0/15
2: BIP	**1**(0.9)$^{\star 3}$	1(1)	1(1)	1(1.0)	1(1)	3/15

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 IPOP-CMA-ES 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 48. A $\downarrow$ indicates the same tested against the best BBOB-2009.**