Black-Box Optimization Benchmarking Template for Noiseless Function Testbed

Draft version [*]

Forename Name

ABSTRACT

to be written

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. CPU TIMING

In order to evaluate the CPU timing of the algorithm, we have run the MY-ALGORITHM-NAME on the function f_8 with restarts for at least 30 seconds and until a maximum budget equal to $400(D+2)$ is reached. The code was run on a Mac Intel(R) Core(TM) i5-2400S CPU @ 2.50GHz with 1 processor and 4 cores. The time per function evaluation for dimensions 2, 3, 5, 10, 20, 40 equals $x.x$, $x.x$, $x.x$, xx, xxx, and xxx milliseconds respectively.

2. RESULTS

Results of PSO DE modified from experiments according to [**?**] on the benchmark functions given in [**?**, **?**] are presented in Figures 1, 2, 3, and 4 and in Tables 1.

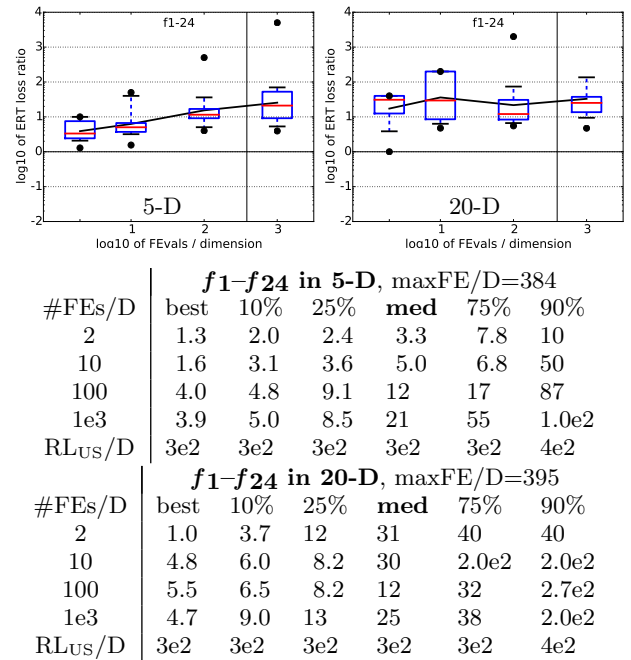|		$f1$–$f24$ in **5-D**, maxFE/D=384				
#FEs/D	best	10%	25%	**med**	75%	90%
2	1.3	2.0	2.4	3.3	7.8	10
10	1.6	3.1	3.6	5.0	6.8	50
100	4.0	4.8	9.1	12	17	87
1e3	3.9	5.0	8.5	21	55	1.0e2
RL$_{US}$/D	3e2	3e2	3e2	3e2	3e2	4e2
		$f1$–$f24$ in **20-D**, maxFE/D=395				
#FEs/D	best	10%	25%	**med**	75%	90%
2	1.0	3.7	12	31	40	40
10	4.8	6.0	8.2	30	2.0e2	2.0e2
100	5.5	6.5	8.2	12	32	2.7e2
1e3	4.7	9.0	13	25	38	2.0e2
RL$_{US}$/D	3e2	3e2	3e2	3e2	3e2	4e2

Figure 3: ERT loss ratio versus the budget in number of f-evaluations divided by dimension. For each given budget FEvals, the target value f_t is computed as the best target f-value reached within the budget by the given algorithm. Shown is then the ERT to reach f_t for the given algorithm or the budget, if the GECCO-BBOB-2009 best algorithm reached a better target within the budget, divided by the best ERT seen in GECCO-BBOB-2009 to reach f_t. 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 a single trial in this function subset. See also Figure 4 for results on each function subgroup.
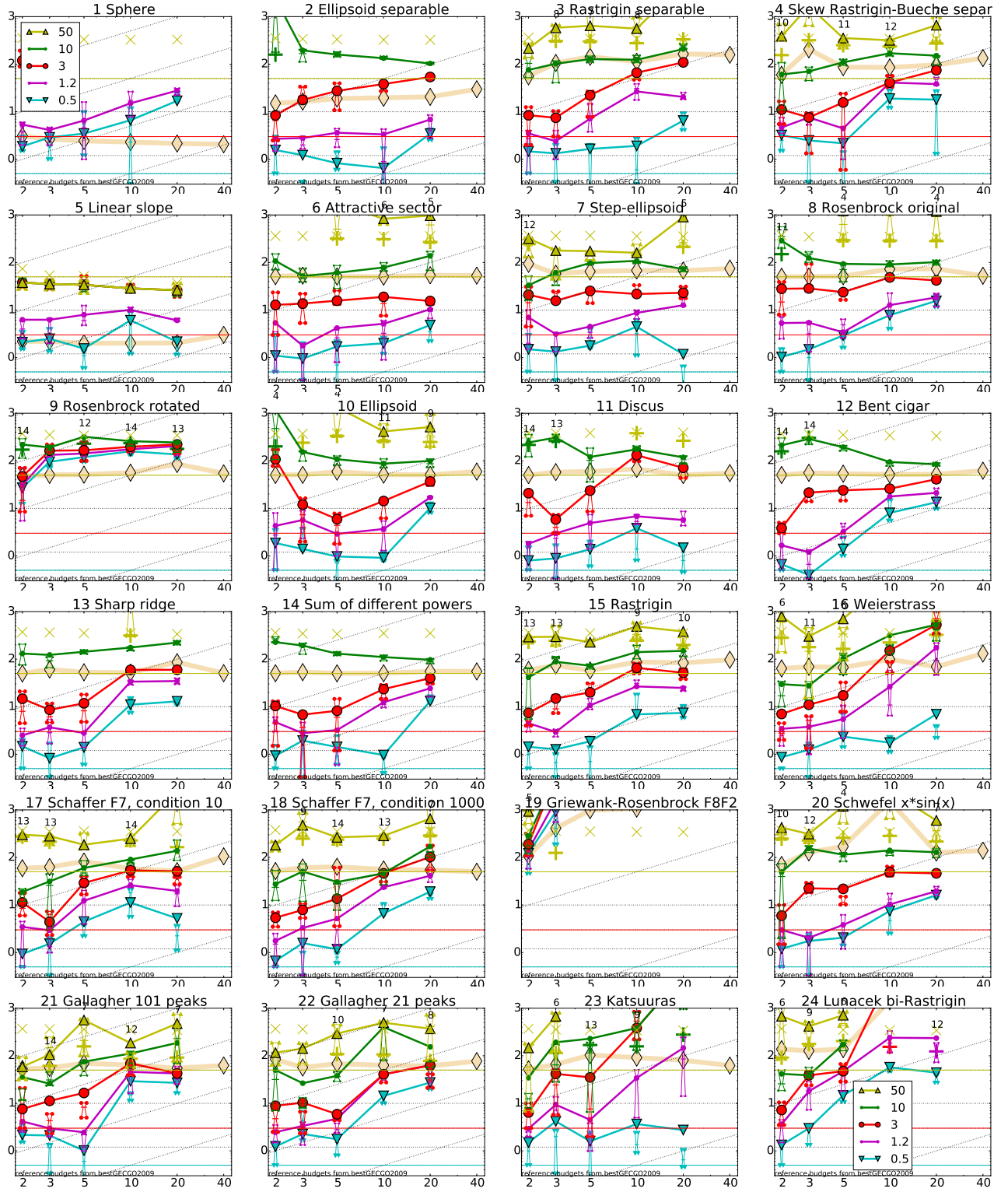

Figure 1: Expected number of f-evaluations (ERT, lines) to reach $f_{\mathrm{opt}} + \Delta f$; median number of f-evaluations (+) to reach the most difficult target that was reached not always but at least once; maximum number of f-evaluations in any trial ($\times$); interquartile range with median (notched boxes) of simulated runlengths to reach $f_{\mathrm{opt}} + \Delta f$; all values are divided by dimension and plotted as $\log_{10}$ values versus dimension. Shown is the ERT for targets just not reached by the artificial GECCO-BBOB-2009 best algorithm within the given budget $k \times \mathrm{DIM}$, where k is shown in the legend. Numbers above ERT-symbols indicate the number of trials reaching the respective target. The light thick line with diamonds indicates the respective best result from BBOB-2009 for the most difficult target. Slanted grid lines indicate a scaling with $\mathcal{O}(\mathrm{DIM})$ compared to $\mathcal{O}(1)$ when using the respective 2009 best algorithm.
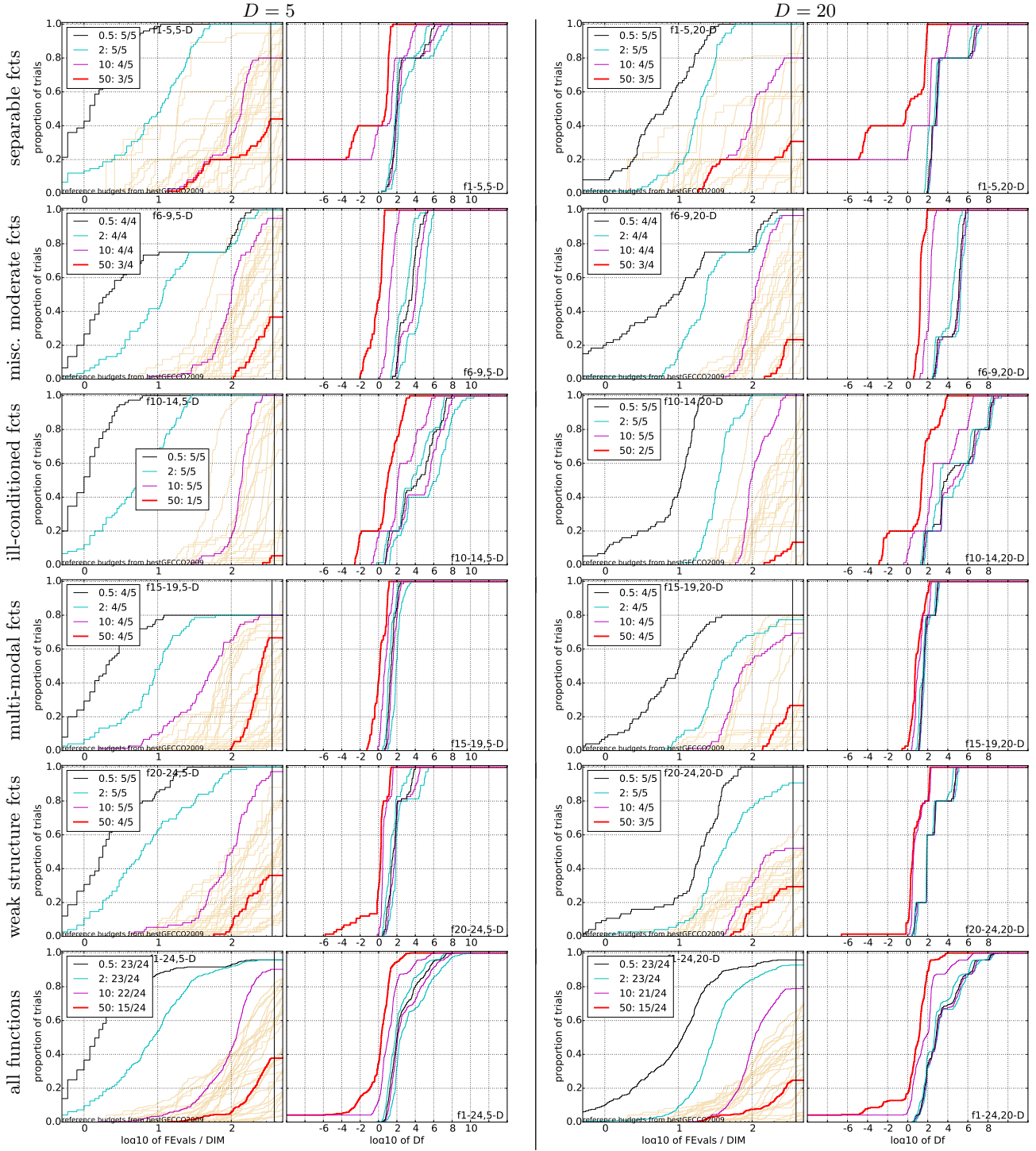
Figure 2: Empirical cumulative distribution functions (ECDF), plotting the fraction of trials with an outcome not larger than the respective value on the x-axis. Left subplots: ECDF of number of function evaluations (FEvals) divided by search space dimension D, to fall below $f_{\mathrm{opt}} + \Delta f$ where Δf is the target just not reached by the GECCO-BBOB-2009 best algorithm within a budget of $k \times \mathrm{DIM}$ evaluations, where k is the first value in the legend. Legends indicate for each target the number of functions that were solved in at least one trial within the displayed budget. Right subplots: ECDF of the best achieved Δf for running times of $0.5D, 1.2D, 3D, 10D, 100D, 1000D, \ldots$ function evaluations (from right to left cycling cyan-magenta-black...) and final Δf-value (red), where Δf and **Df** denote the difference to the optimal function value. Light brown lines in the background show ECDFs for the most difficult target of all algorithms benchmarked during BBOB-2009.

5-D

#FEs/D	0.5	1.2	3	10	50	#succ
f_1	2.5e+1:4.8	1.6e+1:7.6	1.0e-8:12	1.0e-8:12	1.0e-8:12	15/15
	3.6(3)	4.3(4)	∞	∞	∞1600	0/15
f_2	1.6e+6:2.9	4.0e+5:11	4.0e+4:15	6.3e+2:58	1.0e-8:95	15/15
	1.4(1)	1.6(2)	9.0(8)	14(4)	∞1600	0/15
f_3	1.6e+2:4.1	1.0e+2:15	6.3e+1:23	2.5e+1:73	1.0e+1:716	15/15
	2.0(2)	2.4(2)	4.7(5)	8.8(2)	4.5(2)	7/15
f_4	2.5e+2:2.6	1.6e+2:10	1.0e+2:19	4.0e+1:65	1.6e+1:434	15/15
	4.2(4)	2.3(4)	4.2(4)	8.4(2)	4.1(0.8)	11/15
f_5	6.3e+1:4.0	4.0e+1:10	1.0e-8:10	1.0e-8:10	1.0e-8:10	15/15
	1.9(2)	4.0(2)	17(6)	17(7)	17(6)	15/15
f_6	1.0e+5:3.0	2.5e+4:8.4	1.0e+2:16	2.5e+1:54	2.5e-1:254	15/15
	2.8(2)	2.5(2)	4.9(2)	5.6(4)	33(18)	3/15
f_7	1.6e+2:4.2	1.0e+2:6.2	2.5e+1:20	4.0e+0:54	1.0e+0:324	15/15
	2.1(2)	3.6(2)	6.3(2)	9.0(3)	2.7(0.7)	15/15
f_8	1.0e+4:4.6	6.3e+3:6.8	1.0e+3:18	6.3e+1:54	1.6e+0:258	15/15
	3.2(3)	2.5(2)	6.7(3)	8.6(5)	24(17)	4/15
f_9	2.5e+1:20	1.6e+1:26	1.0e+1:35	4.0e+0:62	1.6e-2:256	15/15
	30(5)	28(14)	24(6)	26(21)	∞1700	0/15
f_{10}	2.5e+6:2.9	6.3e+5:7.0	2.5e+5:17	6.3e+3:54	2.5e+1:297	15/15
	1.7(3)	2.1(4)	1.8(2)	10(4)	23(19)	4/15
f_{11}	1.0e+6:3.0	6.3e+4:6.2	6.3e+2:16	6.3e+1:74	6.3e-1:298	15/15
	2.3(1)	4.0(4)	7.4(8)	8.1(6)	∞1900	0/15
f_{12}	4.0e+7:3.6	1.6e+7:7.6	4.0e+6:19	1.6e+4:52	1.0e+0:268	15/15
	1.9(1)	2.1(1)	6.2(2)	19(7)	∞1700	0/15
f_{13}	1.0e+3:2.8	6.3e+2:8.4	4.0e+1:52	6.3e+1:52	6.3e-2:264	15/15
	2.5(4)	1.7(2)	3.5(4)	14(5)	∞1700	0/15
f_{14}	1.6e+1:3.0	1.0e+1:10	6.3e+0:15	2.5e-1:53	1.0e-5:251	15/15
	2.4(2)	1.6(2)	2.7(2)	12(1)	∞1700	0/15
f_{15}	1.6e+2:3.0	1.0e+2:13	6.3e+1:24	4.0e+1:55	1.6e+1:289	5/5
	3.1(3)	4.1(3)	4.1(3)	6.6(3)	3.9(0.7)	15/15
f_{16}	4.0e+1:4.8	2.5e+1:16	1.6e+1:46	1.0e+1:120	4.0e+0:334	15/15
	2.4(4)	1.7(2)	1.9(2)	4.3(4)	11(14)	6/15
f_{17}	1.0e+1:5.2	6.3e+0:26	4.0e+0:57	2.5e+0:110	6.3e-1:412	15/15
	4.3(4)	2.4(2)	2.6(3)	2.7(1)	2.2(0.6)	15/15
f_{18}	6.3e+1:3.4	4.0e+1:7.2	2.5e+1:20	1.6e+1:58	4.0e-1:318	15/15
	1.7(0.7)	3.6(7)	3.4(5)	2.6(3)	4.2(0.6)	14/15
f_{19}	1.6e-1:172	1.0e-1:242	6.3e-2:675	4.0e-2:3078	2.5e-2:4946	15/15
	∞	∞	∞	∞	∞1700	0/15
f_{20}	6.3e+3:5.1	4.0e+3:8.4	4.0e+1:15	2.5e+0:69	1.0e+0:851	15/15
	2.0(2)	2.3(3)	7.2(5)	8.3(4)	7.0(3)	4/15
f_{21}	4.0e+1:3.9	2.5e+1:11	1.6e+1:31	6.3e+0:73	1.6e+0:347	5/5
	1.3(1)	1.2(1)	2.7(3)	5.2(5)	8.1(16)	7/15
f_{22}	6.3e+1:3.6	4.0e+1:15	2.5e+1:32	1.0e+1:71	1.6e+0:341	5/5
	2.4(1)	1.6(1.0)	0.92(0.7)	2.7(1)	4.3(5)	10/15
f_{23}	1.0e+1:3.0	6.3e+0:9.0	4.0e+0:33	2.5e+0:84	1.6e+0:518	15/15
	2.7(2)	2.5(3)	5.3(5)	14(9)	∞1700	0/15
f_{24}	6.3e+1:15	4.0e+1:37	4.0e+1:37	2.5e+1:118	1.6e+1:692	15/15
	5.0(2)	6.5(4)	6.5(3)	7.3(3)	5.2(4)	6/15

20-D

#FEs/D	0.5	1.2	3	10	50	#succ
f_1	6.3e+1:24	4.0e+1:42	1.0e-8:43	1.0e-8:43	1.0e-8:43	15/15
	14(6)	13(3)	∞	∞	∞6500	0/15
f_2	4.0e+6:29	2.5e+6:42	1.0e+5:65	1.0e+4:207	1.0e-8:412	15/15
	2.4(4)	3.3(3)	17(3)	10(2)	∞6400	0/15
f_3	6.3e+2:33	4.0e+2:44	1.6e+2:109	1.0e+2:255	2.5e+1:3277	15/15
	3.9(3)	9.4(2)	20(5)	16(4)	31(29)	1/15
f_4	6.3e+2:22	4.0e+2:91	2.5e+2:250	1.6e+2:332	6.3e+1:1927	15/15
	17(9)	8.4(3)	5.9(2)	9.1(2)	6.9(7)	7/15
f_5	2.5e+2:19	1.6e+2:34	1.0e-8:41	1.0e-8:41	1.0e-8:41	15/15
	2.3(1)	3.5(0.9)	13(2)	13(4)	13(3)	15/15
f_6	2.5e+5:16	6.3e+4:43	1.6e+4:62	1.6e+2:353	1.6e+1:1078	15/15
	6.0(9)	4.8(3)	5.0(2)	7.8(3)	18(28)	5/15
f_7	1.0e+3:11	4.0e+2:39	2.5e+2:74	6.3e+1:319	1.6e+1:1351	15/15
	2.2(2)	6.5(4)	6.3(1)	4.6(1)	13(7)	5/15
f_8	4.0e+4:19	2.5e+4:35	4.0e+3:67	2.5e+2:231	1.6e+1:1470	15/15
	17(6)	10(6)	13(2)	8.8(2)	16(17)	4/15
f_9	1.0e+2:357	6.3e+1:560	4.0e+1:684	2.5e+1:756	1.0e+1:1716	15/15
	7.6(3)	7.3(4)	6.5(6)	6.4(4)	∞6700	0/15
f_{10}	1.6e+6:15	1.0e+6:27	4.0e+5:70	6.3e+4:231	4.0e+3:1015	15/15
	14(4)	12(5)	10(4)	8.6(5)	10(8)	9/15
f_{11}	4.0e+4:11	2.5e+3:27	1.6e+2:313	1.0e+2:481	1.0e+1:1002	15/15
	2.6(2)	4.2(3)	4.6(0.9)	4.9(2)	116(47)	1/15
f_{12}	1.0e+8:23	6.3e+7:39	2.5e+7:76	4.0e+6:209	1.0e+1:1042	15/15
	12(4)	11(5)	11(3)	8.1(1)	∞6600	0/15
f_{13}	1.6e+3:28	1.0e+3:64	6.3e+2:79	4.0e+1:211	2.5e+0:1724	15/15
	9.4(6)	11(3)	15(1)	21(3)	∞6700	0/15
f_{14}	2.5e+1:15	1.6e+1:42	1.0e+1:75	1.6e+0:219	6.3e-1:1106	15/15
	18(6)	11(5)	11(4)	8.9(3)	∞6900	0/15
f_{15}	6.3e+2:15	4.0e+2:67	2.5e+2:292	1.6e+2:846	1.0e+2:1671	15/15
	10(3)	7.4(2)	3.5(2)	3.5(2)	4.6(5)	10/15
f_{16}	4.0e+1:26	2.5e+1:127	1.6e+1:540	1.6e+1:540	1.0e+1:1384	15/15
	5.3(7)	28(38)	20(12)	20(19)	75(79)	1/15
f_{17}	1.6e+1:11	1.0e+1:63	6.3e+0:305	4.0e+0:468	1.0e+0:1030	15/15
	10(12)	6.3(6)	3.3(2)	5.9(3)	47(57)	2/15
f_{18}	4.0e+1:116	2.5e+1:252	1.6e+1:430	1.0e+1:621	4.0e+0:1090	15/15
	3.3(2)	3.3(0.9)	4.8(10)	5.5(4)	12(9)	7/15
f_{19}	1.6e-1:2.5e5	1.0e-1:3.4e5	6.3e-2:3.4e5	4.0e-2:3.4e5	2.5e-2:3.4e5	3/15
	∞	∞	∞	∞	∞6800	0/15
f_{20}	1.6e+4:38	1.0e+4:42	2.5e+2:62	2.5e+0:250	1.6e+0:2536	15/15
	8.4(5)	9.3(4)	15(3)	10(1)	4.7(2)	7/15
f_{21}	6.3e+1:36	4.0e+1:77	4.0e+1:77	1.6e+1:456	4.0e+0:1094	15/15
	15(8)	11(5)	11(5)	8.2(19)	8.7(12)	7/15
f_{22}	6.3e+1:45	4.0e+1:68	4.0e+1:68	1.6e+1:231	1.6e+0:1219	15/15
	12(3)	19(27)	19(50)	13(2)	6.1(4)	8/15
f_{23}	6.3e+0:29	4.0e+0:118	2.5e+0:306	2.5e+0:306	1.6e+0:1614	15/15
	1.9(3)	25(17)	329(233)	329(383)	∞6800	0/15
f_{24}	2.5e+2:208	1.6e+2:918	1.0e+2:6628	6.3e+1:9885	4.0e+1:31629	15/15
	4.2(2)	5.1(5)	∞	∞	∞6800	0/15

Table 1: Expected running time (ERT in number of function evaluations) divided by the best ERT measured during BBOB-2009. The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear in the second row of each cell, the best ERT (preceded by the target Δf-value in *italics*) in the first. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Bold entries are statistically significantly better (according to the rank-sum test) compared to the best algorithm in BBOB-2009, with $p = 0.05$ or $p = 10^{-k}$ when the number $k > 1$ is following the $\downarrow$ symbol, with Bonferroni correction by the number of functions.
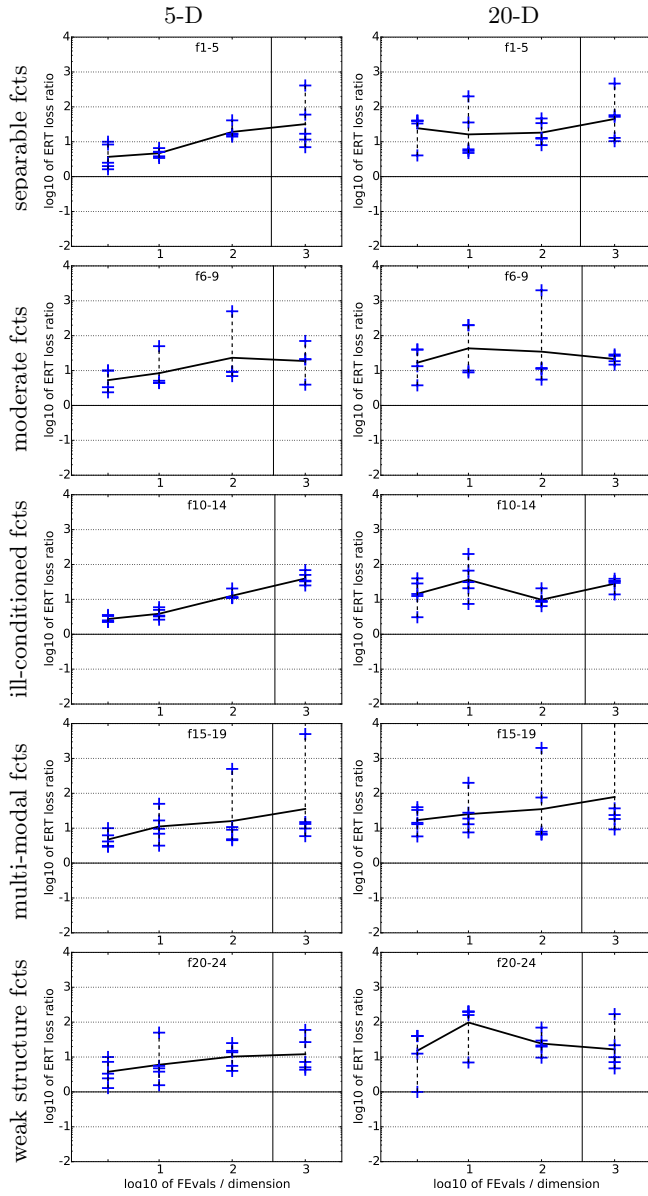
Figure 4: ERT loss ratios (see Figure 3 for details). Each cross (+) represents a single function, the line is the geometric mean.