Black-Box Optimization Benchmarking Template for Noiseless Function Testbed

March 24, 2015

1 Results

Results from experiments according to [**?**] on the benchmark functions given in [**?**, **?**] are presented in Figures **??**, **??**, **??**, **??**, **??**, and **??** and Tables **??** and **??**. 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} 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 if available.

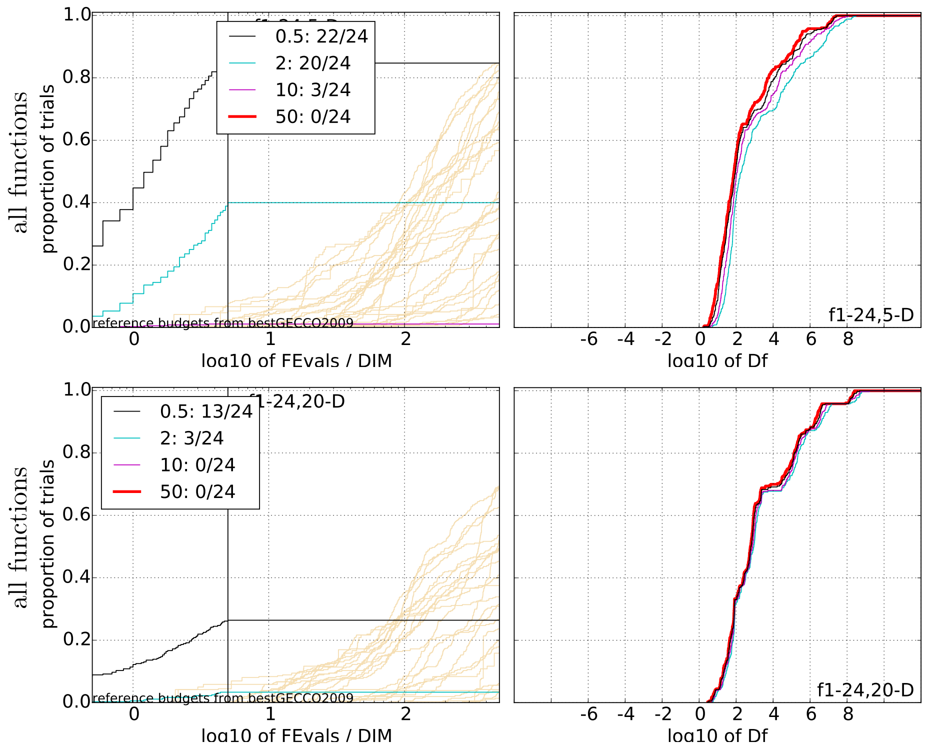


Figure 1: Empirical cumulative distribution functions (ECDF), plotting the fraction of trials with an outcome not larger than the respective value on the x-axis. Left subplots: ECDF of number of function evaluations (FEvals) divided by search space dimension D, to fall below $f_{\mathrm{opt}} + \Delta f$ 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. The top row shows results for 5-D and the bottom row for 20-D.

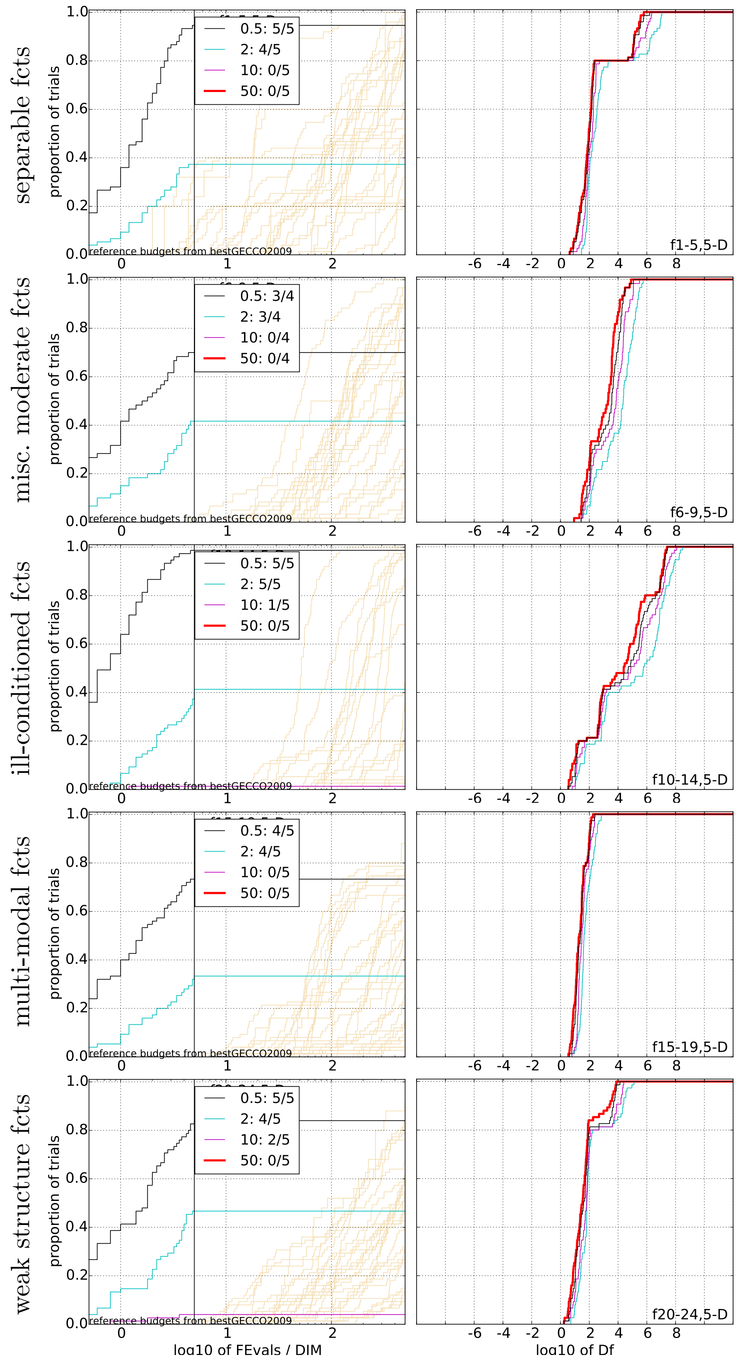


Figure 2: Subgroups of functions 5-D. See caption of Figure ?? for details.

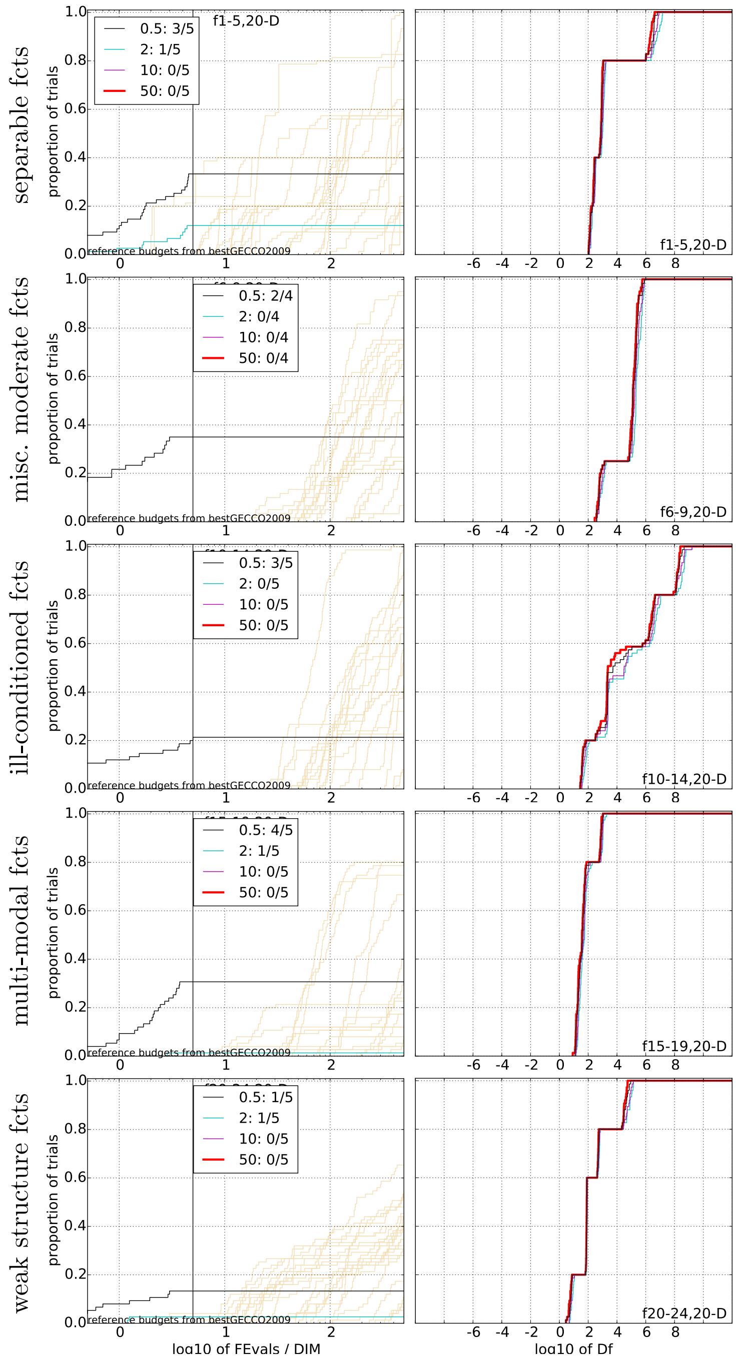
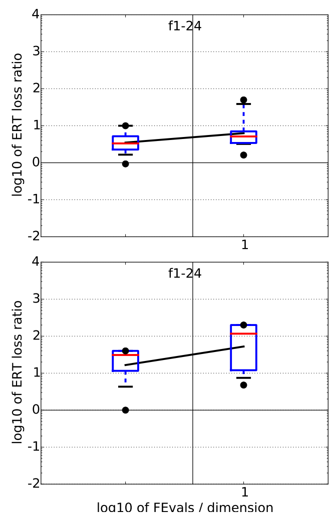


Figure 3: Subgroups of functions 20-D. See caption of Figure **??** for details.



#FEs/D	f_1–f_{24} in 5-D, maxFE/D=5					
	best	10%	25%	**med**	75%	90%
2	0.93	1.5	2.3	3.3	5.7	10
10	1.6	3.1	3.4	5.1	7.1	50
$\mathrm{RL_{US}}$/D	5	5	5	5	5	5

#FEs/D	f_1–f_{24} in 20-D, maxFE/D=5					
	best	10%	25%	**med**	75%	90%
2	1.0	3.5	11	31	40	40
10	4.8	7.4	12	1.2e2	2.0e2	2.0e2
$\mathrm{RL_{US}}$/D	5	5	5	5	5	5

Figure 4: ERT loss ratio. Left: plotted versus given budget FEvals = #FEs in log-log display. Box-Whisker plot shows 25-75%-ile (box) with median, 10-90%-ile (caps), and minimum and maximum ERT loss ratio (points). The black line is the geometric mean. The vertical line gives the maximal number of function evaluations. Right: tabulated ERT loss ratios in 5-D (top table) and 20-D (bottom table). maxFE/D gives the maximum number of function evaluations divided by the dimension. $\mathrm{RL_{US}}$/D gives the median number of function evaluations for unsuccessful trials.

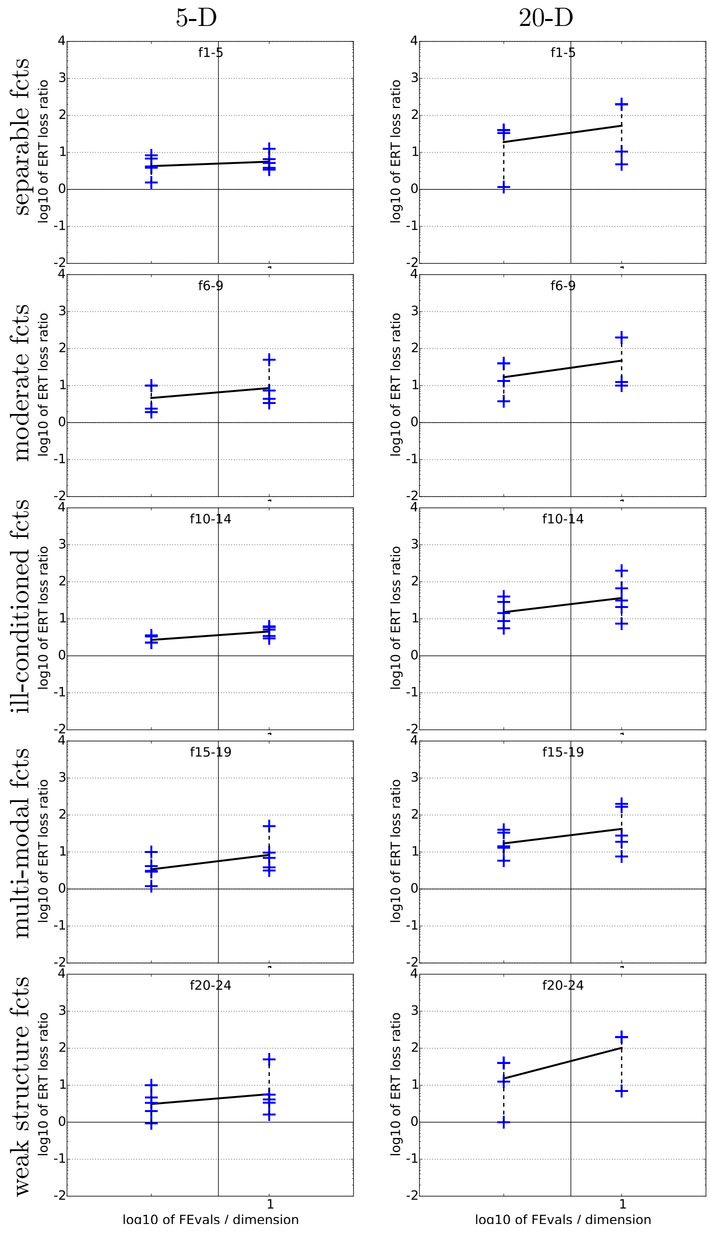


Figure 5: ERT loss ratio versus given budget FEvals divided by dimension in log-log display. Crosses give the single values on the indicated functions, the line is the geometric mean. The vertical line gives the maximal number of function evaluations in the respective function subgroup.

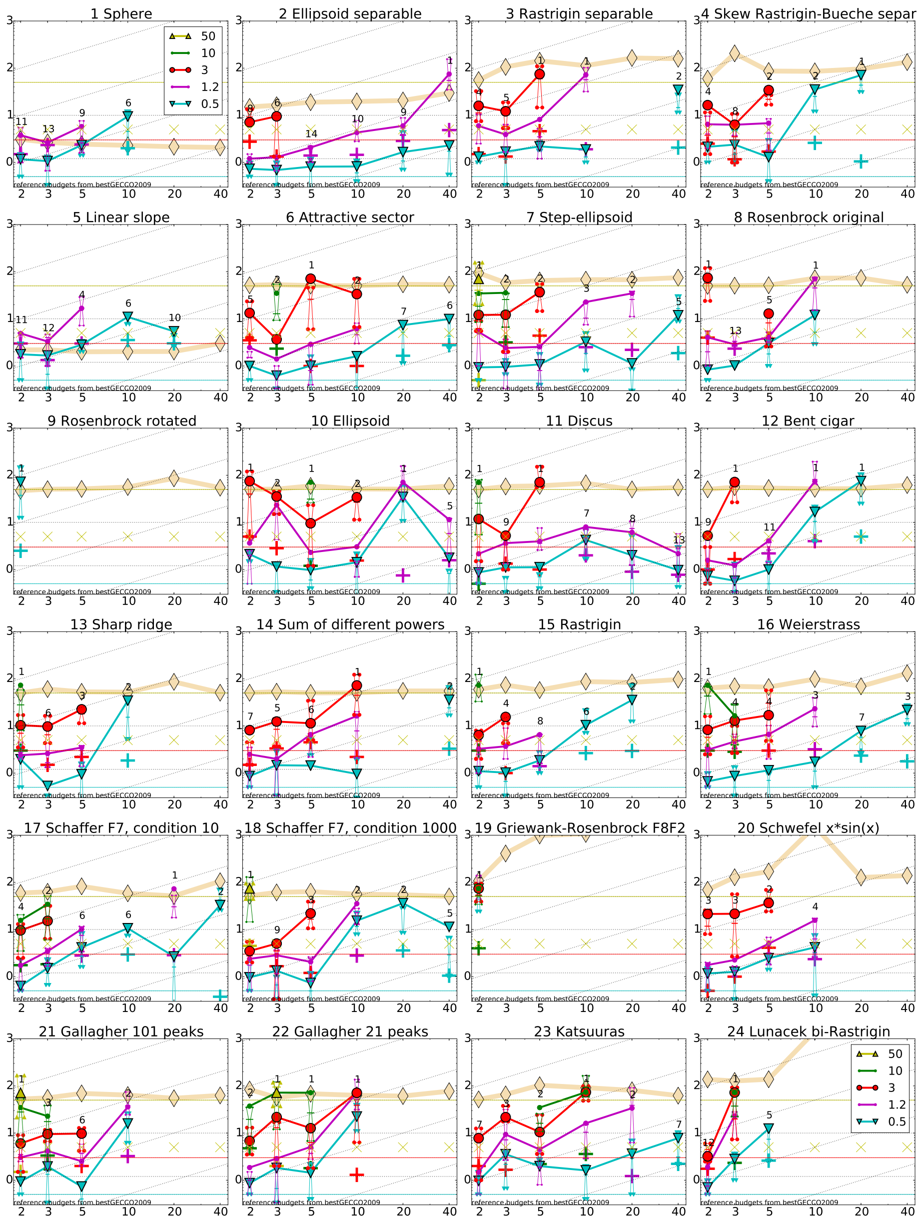


Figure 6: Expected number of f-evaluations (ERT, lines) to reach $f_{\mathrm{opt}} + \Delta f$; median number of f-evaluations (+) to reach the most difficult target that was reached not always but at least once; maximum number of f-evaluations in any trial (×); interquartile range with median (notched 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.

#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
	2.4(3)	3.8(5)	∞	∞	∞*25*	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(0.7)	0.93(0.7)	∞	∞	∞*25*	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
	2.6(3)	2.8(2)	16(37)	∞	∞*25*	0/15
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
	2.5(2)	3.5(3)	9.1(7)	∞	∞*25*	0/15
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
	3.5(4)	8.5(10)	∞	∞	∞*25*	0/15
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
	1.7(2)	1.7(3)	22(53)	∞	∞*25*	0/15
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
	1.3(2)	2.1(2)	9.1(8)	∞	∞*25*	0/15
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
	3.3(3)	3.0(2)	3.6(3)	∞	∞*25*	0/15
f_9	*2.5e+1*:20	*1.6e+1*:26	*1.0e+1*:35	*4.0e+0*:62	*1.6e-2*:256	15/15
	∞	∞	∞	∞	∞*25*	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(1)	1.6(0.5)	2.8(3)	6.6(8)	∞*25*	0/15
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
	1.9(2)	3.2(3)	22(31)	∞	∞*25*	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
	1.4(0.7)	2.7(4)	∞	∞	∞*25*	0/15
f_{13}	*1.0e+3*:2.8	*6.3e+2*:8.4	*4.0e+2*:17	*6.3e+1*:52	*6.3e-2*:264	15/15
	1.7(2)	2.1(4)	6.7(6)	∞	∞*25*	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(8)	3.4(4)	3.7(4)	∞	∞*25*	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(6)	2.5(2)	∞	∞	∞*25*	0/15
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
	1.2(1)	2.1(2)	1.8(2)	∞	∞*25*	0/15
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
	4.0(4)	2.0(2)	∞	∞	∞*25*	0/15
f_{18}	*6.3e+1*:3.4	*4.0e+1*:7.2	*2.5e+1*:20	*1.6e+1*:58	*1.6e+0*:318	15/15
	1.1(0.9)	1.4(3)	5.5(9)	∞	∞*25*	0/15
f_{19}	*1.6e-1*:172	*1.0e-1*:242	*6.3e-2*:675	*4.0e-2*:3078	*2.5e-2*:4946	15/15
	∞	∞	∞	∞	∞*25*	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
	2.4(3)	3.0(2)	12(14)	∞	∞*25*	0/15
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
	0.93(0.5)	1.2(2)	1.6(1)	∞	∞*25*	0/15
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
	2.0(1)	1.7(2)	2.0(1)	5.1(5)	∞*25*	0/15
f_{23}	*1.0e+1*:3.0	*6.3e+0*:9.0	*4.0e+0*:33	*2.5e+0*:84	*1.0e+0*:518	15/15
	3.3(4)	2.5(4)	1.6(2)	2.1(1)	∞*25*	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
	4.3(4)	∞	∞	∞	∞*25*	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 ↓ symbol, with Bonferroni correction by the number of functions.
Results of PUT MY BBOB DATA PATH in 5-D.

#FEs/D	0.5	1.2	3	10	50	#succ
$\mathbf{f_{1}}$	*6.3e+1*:24	*4.0e+1*:42	*1.0e-8*:43	*1.0e-8*:43	*1.0e-8*:43	15/15
	∞	∞	∞	∞	∞ *100*	0/15
$\mathbf{f_{2}}$	*4.0e+6*:29	*2.5e+6*:42	*1.0e+5*:65	*1.0e+4*:207	*1.0e-8*:412	15/15
	1.2(1.0)	2.8(2)	∞	∞	∞ *100*	0/15
$\mathbf{f_{3}}$	*6.3e+2*:33	*4.0e+2*:44	*1.6e+2*:109	*1.0e+2*:255	*2.5e+1*:3277	15/15
	∞	∞	∞	∞	∞ *100*	0/15
$\mathbf{f_{4}}$	*6.3e+2*:22	*4.0e+2*:91	*2.5e+2*:250	*1.6e+2*:332	*6.3e+1*:1927	15/15
	66(76)	∞	∞	∞	∞ *100*	0/15
$\mathbf{f_{5}}$	*2.5e+2*:19	*1.6e+2*:34	*1.0e-8*:41	*1.0e-8*:41	*1.0e-8*:41	15/15
	5.7(5)	∞	∞	∞	∞ *100*	0/15
$\mathbf{f_{6}}$	*2.5e+5*:16	*6.3e+4*:43	*1.6e+4*:62	*1.6e+2*:353	*1.6e+1*:1078	15/15
	9.2(13)	∞	∞	∞	∞ *100*	0/15
$\mathbf{f_{7}}$	*1.0e+3*:11	*4.0e+2*:39	*2.5e+2*:74	*6.3e+1*:319	*1.0e+1*:1351	15/15
	2.1(5)	18(17)	∞	∞	∞ *100*	0/15
$\mathbf{f_{8}}$	*4.0e+4*:19	*2.5e+4*:35	*4.0e+3*:67	*2.5e+2*:231	*1.6e+1*:1470	15/15
	∞	∞	∞	∞	∞ *100*	0/15
$\mathbf{f_{9}}$	*1.0e+2*:357	*6.3e+1*:560	*4.0e+1*:684	*2.5e+1*:756	*1.0e+1*:1716	15/15
	∞	∞	∞	∞	∞ *100*	0/15
$\mathbf{f_{10}}$	*1.6e+6*:15	*1.0e+6*:27	*4.0e+5*:70	*6.3e+4*:231	*4.0e+3*:1015	15/15
	46(118)	52(49)	∞	∞	∞ *100*	0/15
$\mathbf{f_{11}}$	*4.0e+4*:11	*2.5e+3*:27	*1.6e+2*:313	*1.0e+2*:481	*1.0e+1*:1002	15/15
	3.6(1)	4.6(5)	∞	∞	∞ *100*	0/15
$\mathbf{f_{12}}$	*1.0e+8*:23	*6.3e+7*:39	*2.5e+7*:76	*4.0e+6*:209	*1.0e+1*:1042	15/15
	65(100)	∞	∞	∞	∞ *100*	0/15
$\mathbf{f_{13}}$	*1.6e+3*:28	*1.0e+3*:64	*6.3e+2*:79	*4.0e+1*:211	*2.5e+0*:1724	15/15
	∞	∞	∞	∞	∞ *100*	0/15
$\mathbf{f_{14}}$	*2.5e+1*:15	*1.6e+1*:42	*1.0e+1*:75	*1.6e+0*:219	*6.3e-4*:1106	15/15
	∞	∞	∞	∞	∞ *100*	0/15
$\mathbf{f_{15}}$	*6.3e+2*:15	*4.0e+2*:67	*2.5e+2*:292	*1.6e+2*:846	*1.0e+2*:1671	15/15
	46(61)	∞	∞	∞	∞ *100*	0/15
$\mathbf{f_{16}}$	*4.0e+1*:26	*2.5e+1*:127	*1.6e+1*:540	*1.6e+1*:540	*1.0e+1*:1384	15/15
	5.9(8)	∞	∞	∞	∞ *100*	0/15
$\mathbf{f_{17}}$	*1.6e+1*:11	*1.0e+1*:63	*6.3e+0*:305	*4.0e+0*:468	*1.0e+0*:1030	15/15
	5.0(2)	23(35)	∞	∞	∞ *100*	0/15
$\mathbf{f_{18}}$	*4.0e+1*:116	*2.5e+1*:252	*1.6e+1*:430	*1.0e+1*:621	*4.0e+0*:1090	15/15
	6.2(5)	∞	∞	∞	∞ *100*	0/15
$\mathbf{f_{19}}$	*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
	∞	∞	∞	∞	∞ *100*	0/15
$\mathbf{f_{20}}$	*1.6e+4*:38	*1.0e+4*:42	*2.5e+2*:62	*2.5e+0*:250	*1.6e+0*:2536	15/15
	∞	∞	∞	∞	∞ *100*	0/15
$\mathbf{f_{21}}$	*6.3e+1*:36	*4.0e+1*:77	*4.0e+1*:77	*1.6e+1*:456	*4.0e+0*:1094	15/15
	∞	∞	∞	∞	∞ *100*	0/15
$\mathbf{f_{22}}$	*6.3e+1*:45	*4.0e+1*:68	*4.0e+1*:68	*1.6e+1*:231	*6.3e+0*:1219	15/15
	∞	∞	∞	∞	∞ *100*	0/15
$\mathbf{f_{23}}$	*6.3e+0*:29	*4.0e+0*:118	*2.5e+0*:306	*2.5e+0*:306	*1.0e+0*:1614	15/15
	2.5(4)	5.7(5)	∞	∞	∞ *100*	0/15
$\mathbf{f_{24}}$	*2.5e+2*:208	*1.6e+2*:918	*1.0e+2*:6628	*6.3e+1*:9885	*4.0e+1*:31629	15/15
	∞	∞	∞	∞	∞ *100*	0/15

Table 2: 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 ↓ symbol, with Bonferroni correction by the number of functions.
Results of PUT MY BBOB DATA PATH in 20-D.