Example paper: Black-Box Optimization Benchmarking Template for Noiseless Function 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 of NEWUOA from experiments according to [?] on the benchmark functions given in [?, ?] are presented in Figures 1, 2 and 3 and in Tables 1 and 2.

*Submission deadline: March 28th.

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

Table 2: ERT **loss ratio compared to the respective best result from BBOB-2009 for budgets given in the first column (see also Figure 3). The last row $\mathrm{RL_{US}}$/D gives the number of function evaluations in unsuccessful runs divided by dimension. Shown are the smallest, 10%-ile, 25%-ile, 50%-ile, 75%-ile and 90%-ile value (smaller values are better). The ERT Loss ratio equals to one for the respective best algorithm from BBOB-2009. Typical median values are between ten and hundred.**

	***f*1–*f*24 in 5-D**, maxFE/D=102505					
#FEs/D	best	10%	25%	**med**	75%	90%
2	1.4	1.6	2.1	3.2	5.3	10
10	1.2	1.3	1.9	3.4	8.0	50
100	1.0	2.2	3.6	9.4	24	72
1e3	1.0	1.2	3.1	11	40	88
1e4	1.0	1.2	3.1	9.1	52	1.6e2
1e5	1.0	1.2	3.1	8.8	2.7e2	4.2e2
$\mathrm{RL_{US}}$/D	1e5	1e5	1e5	1e5	1e5	1e5
	***f*1–*f*24 in 20-D**, maxFE/D=17757					
#FEs/D	best	10%	25%	**med**	75%	90%
2	1.0	5.4	12	33	40	40
10	1.0	1.8	2.6	3.9	8.2	2.0e2
100	2.4	2.7	3.9	12	40	1.2e3
1e3	3.0	4.7	11	27	88	2.3e2
1e4	4.3	8.6	21	72	3.5e2	1.0e3
1e5	5.7	9.1	16	3.1e2	1.5e3	6.7e3
$\mathrm{RL_{US}}$/D	1e4	1e4	1e4	1e4	1e4	1e4

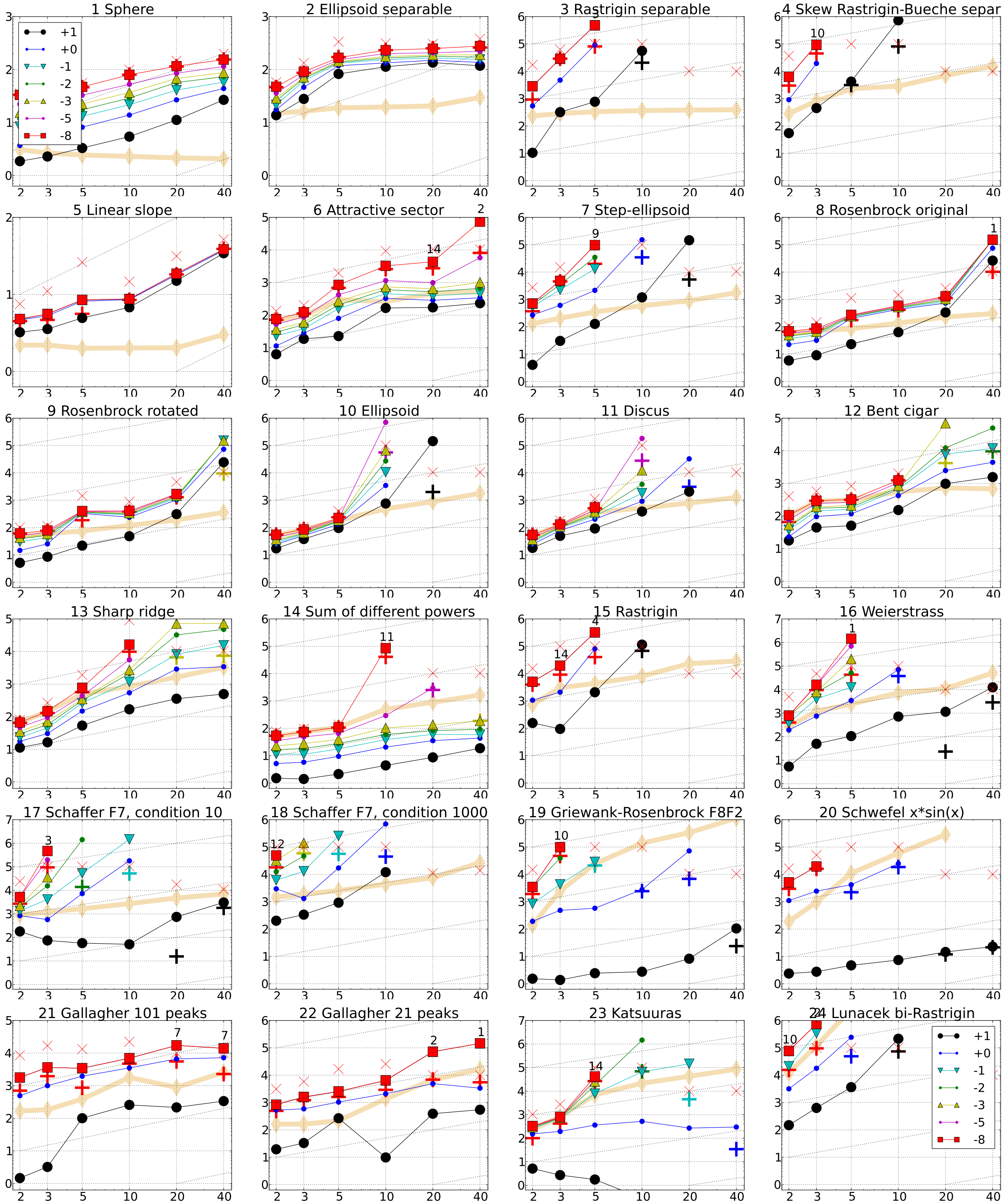

Figure 1: Expected number of f-evaluations (ERT, with lines, see legend) to reach $f_{\mathrm{opt}} + \Delta f$, median number of f-evaluations to reach the most difficult target that was reached at least once (+) and maximum number of f-evaluations in any trial (×), all divided by dimension and plotted as $\log_{10}$ values versus dimension. Shown are $\Delta f = 10^{\{1,0,-1,-2,-3,-5,-8\}}$. Numbers above ERT-symbols indicate the number of successful trials. The light thick line with diamonds indicates the respective best result from BBOB-2009 for $\Delta f = 10^{-8}$. Horizontal lines mean linear scaling, slanted grid lines depict quadratic scaling.

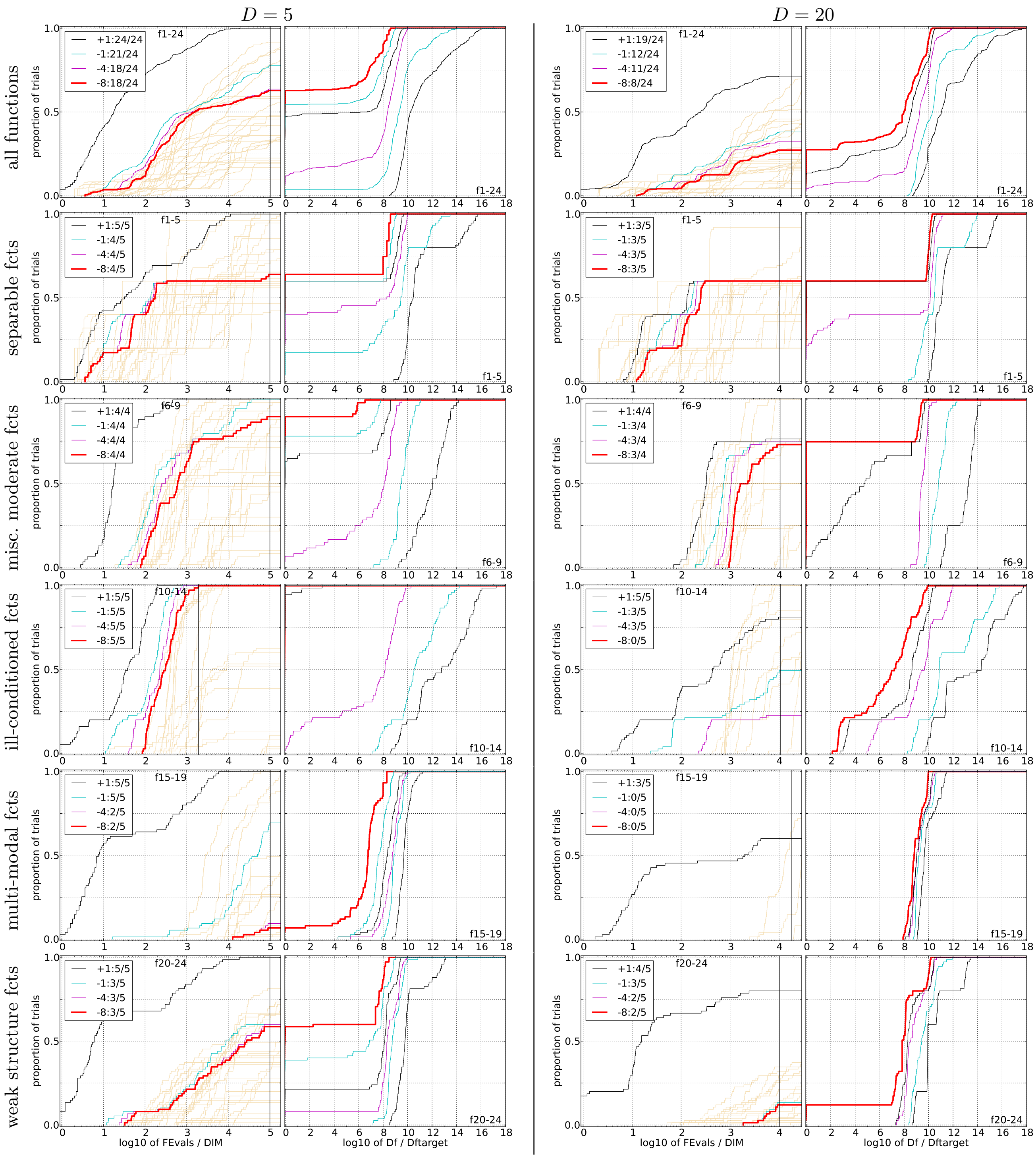

Figure 2: Empirical cumulative distribution functions (ECDFs), plotting the fraction of trials with an outcome not larger than the respective value on the x-axis. Left subplots: ECDF of number of function evaluations (FEvals) divided by search space dimension D, to fall below $f_{\mathrm{opt}} + \Delta f$ with $\Delta f = 10^k$, where k is the first value in the legend. Right subplots: ECDF of the best achieved Δf divided by 10^{-8} for running times of $D, 10\,D, 100\,D, \ldots$ function evaluations (from right to left cycling black-cyan-magenta). The thick red line represents the most difficult target value $f_{\mathrm{opt}} + 10^{-8}$. Legends indicate the number of functions that were solved in at least one trial. Light brown lines in the background show ECDFs for $\Delta f = 10^{-8}$ of all algorithms benchmarked during BBOB-2009.

5-D

Δf	1e+1	1e+0	1e-1	1e-3	1e-5	1e-7	#succ
f1	11	12	12	12	12	12	15/15
	1.5(0.9)	3.3(2)	5.4(1)	9.2(2)	13(1)	17(2)	15/15
f2	83	87	88	90	92	94	15/15
	5.0(3)	6.8(2)	7.4(2)	7.9(2)	8.3(2)	8.6(2)	15/15
f3	716	1622	1637	1646	1650	1654	15/15
	5.4(7)	282(272)	1464(1498)	1456(1519)	1452(1486)	1449(1483)	3/15
f4	809	1633	1688	1817	1886	1903	15/15
	26(23)	∞	∞	∞	∞	∞ *5.0e5*	0/15
f5	10	10	10	10	10	10	15/15
	2.5(1)	4.1(4)	4.2(5)	4.2(5)	4.2(5)	4.2(5)	15/15
f6	114	214	281	580	1038	1332	15/15
	1(0.5)	1.9(2)	2.8(3)	2.3(1)	2.0(0.9)	2.6(1)	15/15
f7	24	324	1171	1572	1572	1597	15/15
	27(46)	33(36)	56(57)	307(352)	307(338)	302(333)	9/15
f8	73	273	336	391	410	422	15/15
	1.6(1)	3.7(4)	3.3(4)	3.1(3)	3.1(3)	3.2(3)	15/15
f9	35	127	214	300	335	369	15/15
	3.1(4)	13(24)	8.2(14)	6.2(10)	5.8(9)	5.4(8)	15/15
f10	349	500	574	626	829	880	15/15
	1.4(0.9)	1.3(0.7)	1.4(0.7)	1.5(0.7)	1.2(0.7)	1.2(0.6)	15/15
f11	143	202	763	1177	1467	1673	15/15
	3.2(3)	5.0(3)	1.7(0.5)	1.5(0.7)	1.5(0.7)	1.6(0.7)	15/15
f12	108	268	371	461	1303	1494	15/15
	2.3(1)	2.2(2)	2.2(1)	2.3(1)	1(0.6)	1(0.6)	15/15
f13	132	195	250	1310	1752	2255	15/15
	2.0(3)	3.8(4)	5.3(3)	1.3(0.6)	1.2(0.7)	1.3(0.9)	15/15
f14	10	41	58	139	251	476	15/15
	1.1(1)	1.2(0.6)	1.5(0.5)	1.4(0.3)	1.3(0.3)	1(0.1)	15/15
f15	511	9310	19369	20073	20769	21359	14/15
	20(24)	43(56)	83(86)	80(91)	77(86)	75(82)	4/15
f16	120	612	2662	10449	11644	12095	15/15
	4.4(9)	28(45)	23(22)	95(98)	302(342)	597(641)	1/15
f17	5.2	215	899	3669	6351	7934	15/15
	55(190)	170(159)	295(278)	∞	∞	∞ *5.0e5*	0/15
f18	103	378	3968	9280	10905	12469	15/15
	45(43)	228(303)	321(332)	∞	∞	∞ *5.0e5*	0/15
f19	1	1	242	1.2e5	1.2e5	1.2e5	15/15
	12(6)	2885(5138)	590(575)	∞	∞	∞ *5.0e5*	0/15
f20	16	851	38111	54470	54861	55313	14/15
	1.5(1)	25(28)	∞	∞	∞	∞ *5.0e5*	0/15
f21	41	1157	1674	1705	1729	1757	14/15
	12(27)	8.4(8)	10(18)	10(17)	10(17)	10(17)	15/15
f22	71	386	938	1008	1040	1068	14/15
	19(33)	13(25)	13(10)	13(10)	12(9)	12(9)	15/15
f23	3.0	518	14249	31654	33030	34256	15/15
	2.9(3)	3.5(6)	2.7(3)	4.0(4)	4.6(4)	5.6(5)	14/15
f24	1622	2.2e5	6.4e6	9.6e6	1.3e7	1.3e7	3/15
	11(11)	5.6(7)	∞	∞	∞	∞ *5.0e5*	0/15

20-D

Δf	1e+1	1e+0	1e-1	1e-3	1e-5	1e-7	#succ
f1	43	43	43	43	43	43	15/15
	5.2(2)	12(5)	19(8)	32(8)	40(6)	49(9)	15/15
f2	385	386	387	390	391	393	15/15
	7.0(0.8)	7.8(1)	8.6(1)	10(1)	11(0.9)	12(1)	15/15
f3	5066	7626	7635	7643	7646	7651	15/15
	∞	∞	∞	∞	∞	∞ *2.0e5*	0/15
f4	4722	7628	7666	7700	7758	1.4e5	9/15
	∞	∞	∞	∞	∞	∞ *2.0e5*	0/15
f5	41	41	41	41	41	41	15/15
	7.4(1)	8.8(2)	9.2(2)	9.2(2)	9.2(2)	9.2(2)	15/15
f6	1296	2343	3413	5220	6728	8409	15/15
	2.7(1)	2.4(0.9)	2.3(0.9)	2.5(0.7)	3.0(0.5)	5.4(4)	14/15
f7	1351	4274	9503	16524	16524	16969	15/15
	2160(2304)	∞	∞	∞	∞	∞ *2.0e5*	0/15
f8	2039	3871	4040	4219	4371	4484	15/15
	3.3(1)	3.8(3)	4.1(3)	4.9(3)	5.3(3)	5.6(3)	15/15
f9	1716	3102	3277	3455	3594	3727	15/15
	3.6(1)	6.6(6)	7.2(6)	8.4(5)	8.8(5)	8.9(4)	15/15
f10	7413	8661	10735	14920	17073	17476	15/15
	390(447)	∞	∞	∞	∞	∞ *2.0e5*	0/15
f11	1002	2228	6278	9762	12285	14831	15/15
	41(48)	292(328)	∞	∞	∞	∞ *2.1e5*	0/15
f12	1042	1938	2740	4140	12407	13827	15/15
	19(27)	26(25)	57(59)	338(360)	∞	∞ *2.0e5*	0/15
f13	652	2021	2751	18749	24455	30201	15/15
	11(14)	29(33)	60(66)	77(86)	∞	∞ *2.0e5*	0/15
f14	75	239	304	932	1648	15661	15/15
	2.3(1)	3.0(2)	3.9(1)	2.9(0.6)	36(31)	∞ *2.1e5*	0/15
f15	30378	1.5e5	3.1e5	3.2e5	4.5e5	4.6e5	15/15
	∞	∞	∞	∞	∞	∞ *2.0e5*	0/15
f16	1384	27265	77015	1.9e5	2.0e5	2.2e5	15/15
	17(21)	∞	∞	∞	∞	∞ *2.0e5*	0/15
f17	63	1030	4005	30677	56288	80472	15/15
	237(585)	∞	∞	∞	∞	∞ *2.0e5*	0/15
f18	621	3972	19561	67569	1.3e5	1.5e5	15/15
	∞	∞	∞	∞	∞	∞ *2.0e5*	0/15
f19	1	1	3.4e5	6.2e6	6.7e6	6.7e6	15/15
	165(149)	1.4e6(2e6)	∞	∞	∞	∞ *2.0e5*	0/15
f20	82	46150	3.1e6	5.5e6	5.6e6	5.6e6	14/15
	3.5(2)	∞	∞	∞	∞	∞ *2.0e5*	0/15
f21	561	6541	14103	14643	15567	17589	15/15
	7.7(12)	20(18)	24(25)	23(24)	22(22)	20(20)	7/15
f22	467	5580	23491	24948	26847	1.3e5	12/15
	17(35)	18(19)	61(59)	58(60)	54(55)	11(12)	2/15
f23	3.2	1614	67457	4.9e5	8.1e5	8.4e5	15/15
	2.1(4)	3.3(5)	43(50)	∞	∞	∞ *2.0e5*	0/15
f24	1.3e6	7.5e6	5.2e7	5.2e7	5.2e7	5.2e7	3/15
	∞	∞	∞	∞	∞	∞ *2.0e5*	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 $f_{\mathrm{opt}} + 10^{-8}$.

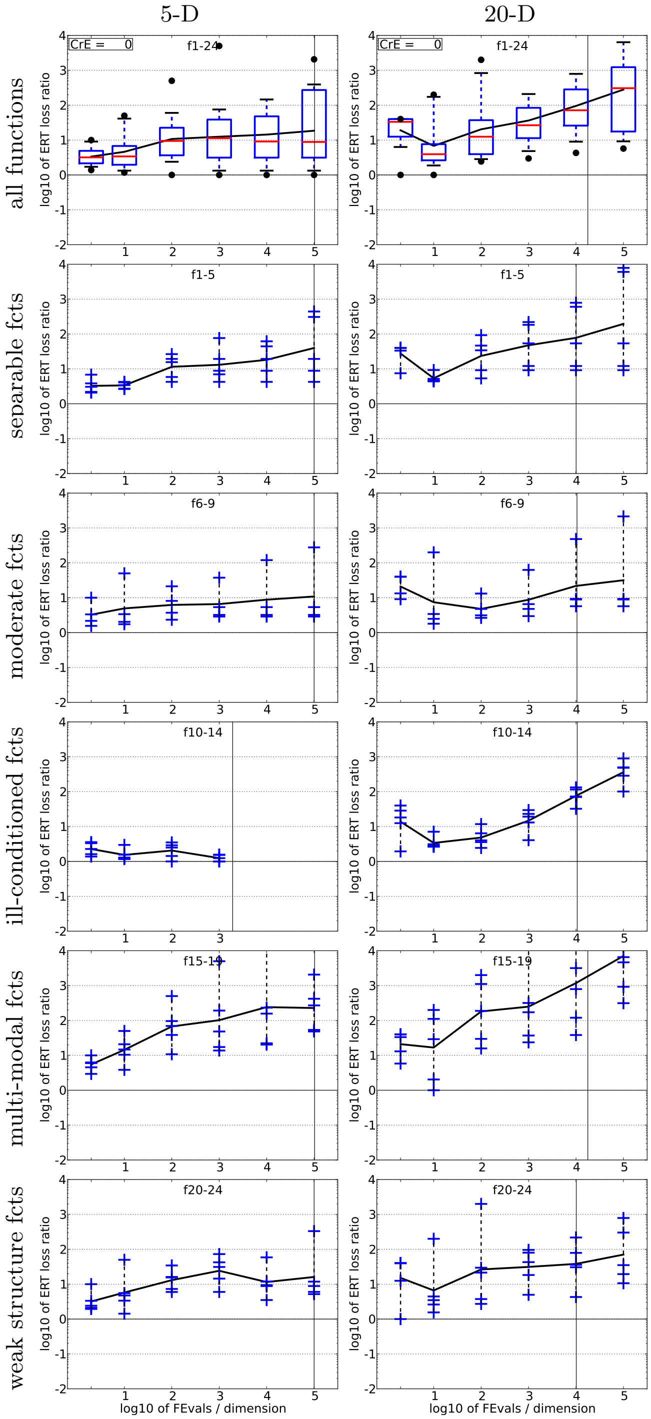

Figure 3: ERT loss ratio vs. a given budget FEvals. Each cross (+) represents a single function. The target value f_t used for a given FEvals is the smallest (best) recorded function value such that $\mathrm{ERT}(f_t) \leq$ FEvals for the presented algorithm. Shown is FEvals divided by the respective best $\mathrm{ERT}(f_t)$ from BBOB-2009 for functions f_1–f_{24} in 5-D and 20-D. 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.