Sum-of-Ranks Bandit for Adaptive Strategy Selection Compared with the Probability Matching-based one within Differential Evolution on the Noiseless Testbed

Draft version *

Álvaro Fialho[1], Marc Schoenauer[1,2], Michèle Sebag[1,2]

[1]Microsoft Research–INRIA Joint Centre
Parc Orsay Université
91893 Orsay Cedex, France

[2]Project-Team TAO, INRIA Saclay - Île-de-France
LRI, Bât. 490, Université Paris-Sud
91405 Orsay Cedex, France

FirstName.LastName@inria.fr

ABSTRACT

The decision of which of the several existent strategies should be applied for the offspring generation is critical for the performance of the Differential Evolution algorithm, besides being problem-dependent. In this paper, we use the BBOB noiseless benchmarking suite to better empirically validate the Sum-of-Ranks Bandit Adaptive Strategy Selection, a comparison-based technique used to automatically select between the available strategies while solving the problem, recently proposed in [1], referred to as *SR-R*. It is compared with another recently proposed approach for adaptive strategy selection, the PM-AdapSS-DE [3].

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 [4] on the benchmark functions given in [2, 5] are presented in Figures 1, 2 and 3 and in Table 1. The **expected running time (ERT)**, used in the figures and table, depends on a given target function value, $f_t = f_{opt} + \Delta f$, 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 [4, 6]. **Statistical significance** is tested with the rank-sum test for a given target Δf_t (10^{-8} 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 25th.

2. REFERENCES

[1] A. Fialho, M. Schoenauer, and M. Sebag. Toward comparison-based adaptive operator selection. In J. B. et al., editor, *GECCO'10: Proceedings of the 12th Annual Conference on Genetic and Evolutionary Computation*. ACM Press, July 2010. to appear.

[2] S. Finck, N. Hansen, R. Ros, and A. Auger. Real-parameter black-box optimization benchmarking 2009: Presentation of the noiseless functions. Technical Report 2009/20, Research Center PPE, 2009. Updated February 2010.

[3] W. Gong, A. Fialho, and Z. Cai. Adaptive strategy selection in differential evolution. In J. B. et al., editor, *GECCO'10: Proceedings of the 12th Annual Conference on Genetic and Evolutionary Computation*. ACM Press, July 2010. to appear.

[4] N. Hansen, A. Auger, S. Finck, and R. Ros. Real-parameter black-box optimization benchmarking 2010: Experimental setup. Technical Report RR-7215, INRIA, 2010.

[5] N. Hansen, S. Finck, R. Ros, and A. Auger. Real-parameter black-box optimization benchmarking 2009: Noiseless functions definitions. Technical Report RR-6829, INRIA, 2009. Updated February 2010.

[6] K. Price. Differential evolution vs. the functions of the second ICEO. In *Proceedings of the IEEE International Congress on Evolutionary Computation*, pages 153–157, 1997.

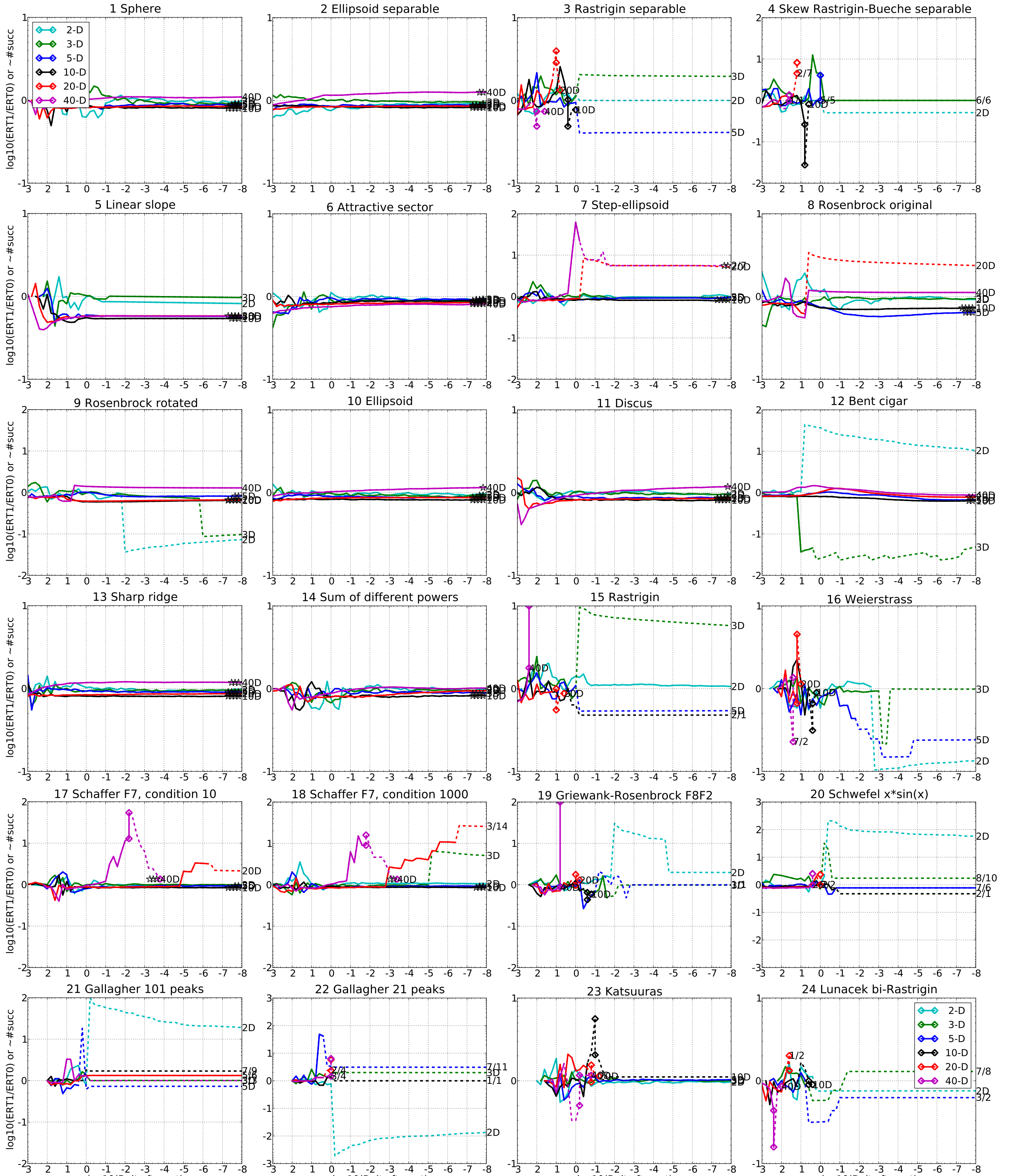

Figure 1: ERT ratio of ALG1-acronym divided by ALG0-acronym versus $\log_{10}(\Delta f)$ for f_1–f_{24} in 2, 3, 5, 10, 20, 40-D. Ratios $< 10^0$ indicate an advantage of ALG1-acronym, smaller values are always better. The line gets dashed when for any algorithm the ERT exceeds thrice the median of the trial-wise overall number of f-evaluations for the same algorithm on this function. Symbols indicate the best achieved Δf-value of one algorithm (ERT gets undefined to the right). The dashed line continues as the fraction of successful trials of the other algorithm, where 0 means 0% and the y-axis limits mean 100%, values below zero for ALG1-acronym. The line ends when no algorithm reaches Δf anymore. The number of successful trials is given, only if it was in $\{1 \dots 9\}$ for ALG1-acronym (1st number) and non-zero for ALG0-acronym (2nd number). Results are significant with $p = 0.05$ for one star and $p = 10^{-\#\star}$ otherwise, with Bonferroni correction within each figure.

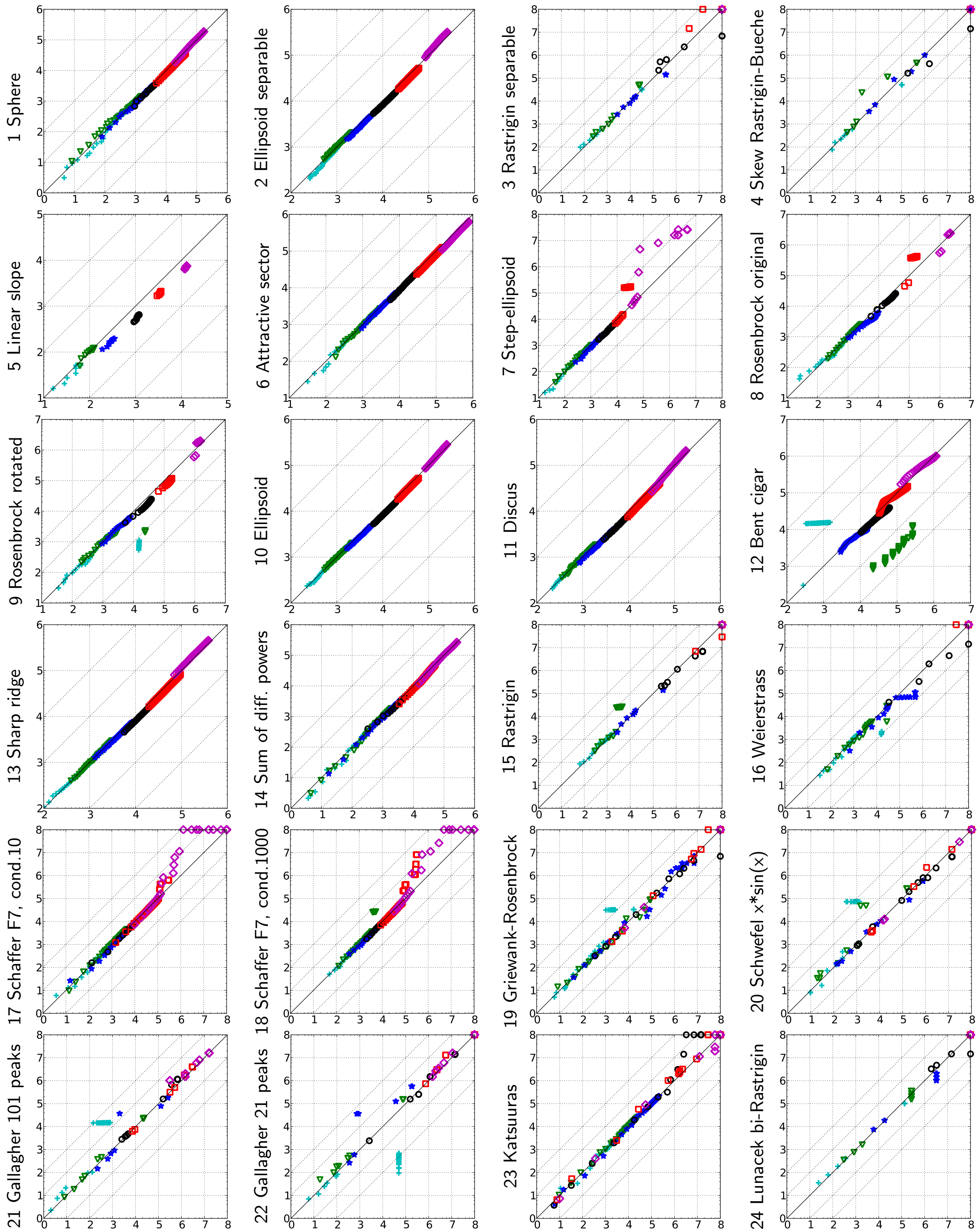

Figure 2: Expected running time (ERT in log10 of number of function evaluations) of ALG1-acronym versus ALG0-acronym for 46 target values $\Delta f \in [10^{-8}, 10]$ in each dimension for functions f_1–f_{24}. Markers on the upper or right egde indicate that the target value was never reached by ALG1-acronym or ALG0-acronym respectively. Markers represent dimension: 2:+, 3:▽, 5:⋆, 10:○, 20:□, 40:◇.

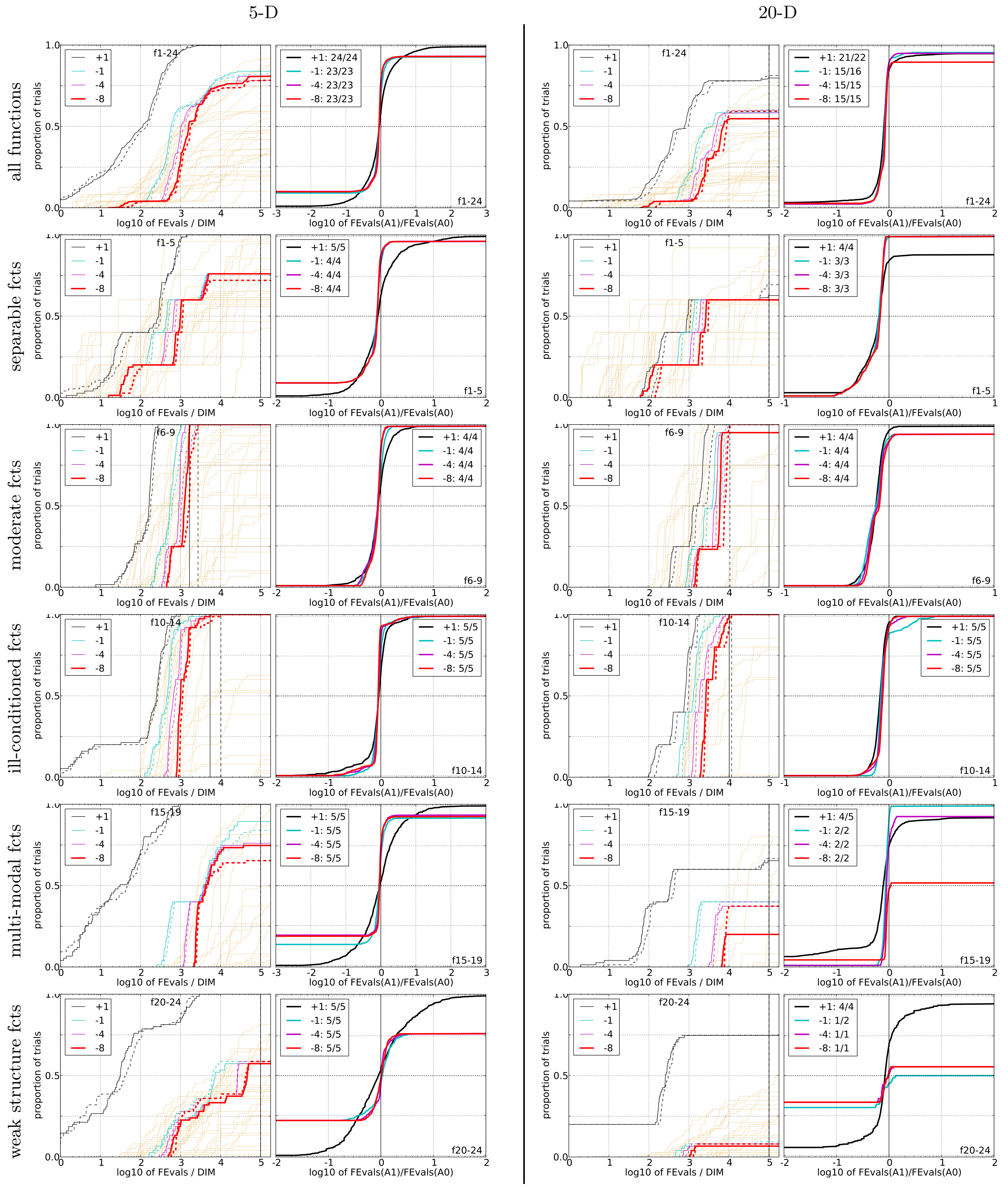

Figure 3: Empirical cumulative distributions (ECDF) of run lengths and speed-up ratios in 5-D (left) and 20-D (right). Left sub-columns: ECDF of the number of function evaluations divided by dimension D (FEvals/D) to reach a target value $f_{\mathrm{opt}} + \Delta f$ with $\Delta f = 10^k$, where $k \in \{1, -1, -4, -8\}$ is given by the first value in the legend, for ALG1-acronym (solid) and ALG0-acronym (dashed). Light beige lines show the ECDF of FEvals for target value $\Delta f = 10^{-8}$ of algorithms benchmarked during BBOB-2009. Right sub-columns: ECDF of FEval ratios of ALG1-acronym divided by ALG0-acronym, 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 (ALG1-acronym first).

5-D

Δf	1e+1	1e+0	1e-1	1e-3	1e-5	1e-7	#succ
f_1	11	12	12	12	12	12	15/15
0: wen	7.4	41	80	150	220	300	15/15
1: F-S	6.3	37	**69***	**130**^{*2}	200	**260**^{*2}	15/15
f_2	83	87	88	90	92	94	15/15
0: wen	20	25	29	38	47	55	15/15
1: F-S	18	**21**^{*3}	**25**^{*2}	**33**^{*3}	**42**^{*3}	**49**^{*3}	15/15
f_3	720	1600	1600	1600	1700	1700	15/15
0: wen	3.9	11	220	210	210	210	9/15
1: F-S	3.7	10	87	88	88	88	12/15
f_4	810	1600	1700	1800	1900	1900	15/15
0: wen	4.8	620	∞	∞	∞	∞ *5.0e5*	0/15
1: F-S	4.4	620	∞	∞	∞	∞ *5.0e5*	0/15
f_5	10	10	10	10	10	10	15/15
0: wen	19	32	34	34	34	34	15/15
1: F-S	12	**19***	**20***	**20***	**20***	**20***	15/15
f_6	110	210	280	580	1000	1300	15/15
0: wen	8.2	8.2	9.4	7.3	5.8	5.8	15/15
1: F-S	6.9	7.9	8.5	6.9	**5.4***	**5.4**^{*2}	15/15
f_7	24	320	1200	1600	1600	1600	15/15
0: wen	11	2.5	1.1	1.4	1.4	1.5	15/15
1: F-S	10	2.2	1	1.3	1.3	1.3	15/15
f_8	73	270	340	390	410	420	15/15
0: wen	14	12	16	20	21	22	15/15
1: F-S	13	8.7	**10***	**11**^{*2}	**13**^{*2}	**14**^{*2}	15/15
f_9	35	130	210	300	340	370	15/15
0: wen	28	20	20	20	20	20	15/15
1: F-S	25	20	17	15	16	17	15/15
f_{10}	350	500	570	630	830	880	15/15
0: wen	4.8	4.3	4.5	5.6	5.2	6	15/15
1: F-S	4.4	3.9	4	**4.9**^{*2}	**4.6**^{*3}	**5.2**^{*3}	15/15
f_{11}	140	200	760	1200	1500	1700	15/15
0: wen	6	6.4	2.4	2.3	2.5	2.7	15/15
1: F-S	5.3	5.9	2.1	**2**^{*2}	**2.1**^{*2}	**2.3**^{*2}	15/15
f_{12}	110	270	370	460	1300	1500	15/15
0: wen	27	14	14	18	9.3	10	15/15
1: F-S	23	14	14	15	6.8	6.8	15/15
f_{13}	130	190	250	1300	1800	2300	15/15
0: wen	10	11	12	3.4	3.5	3.4	15/15
1: F-S	9.6	10	11	**3.1***	**3.2**^{*3}	**3.1**^{*3}	15/15
f_{14}	9.8	41	58	140	250	480	15/15
0: wen	1.8	10	18	15	12	9	15/15
1: F-S	1.4	9.2	15	14	**11***	**8.1**^{*3}	15/15
f_{15}	510	9300	1.9e4	2.0e4	2.1e4	2.1e4	14/15
0: wen	5.3	1.9	14	13	13	13	10/15
1: F-S	4	2	7.4	7.2	7	6.9	12/15
f_{16}	120	610	2700	1.0e4	1.2e4	1.2e4	15/15
0: wen	5.6	43	41	27	41	39	8/15
1: F-S	2.7	34	25	6.8	9.7	9.5	13/15
f_{17}	5.2	210	900	3700	6400	7900	15/15
0: wen	3	4.3	2.5	1.5	1.4	1.5	15/15
1: F-S	5.1	3.4	2.2	**1.3**^{*2}	**1.3***	**1.4***	15/15
f_{18}	100	380	4000	9300	1.1e4	1.2e4	15/15
0: wen	3.5	4.2	0.78	0.72	0.92	1	15/15
1: F-S	3.8	3.9	0.68	0.65	0.85	**0.96***	15/15
f_{19}	1	1	240	1.2e5	1.2e5	1.2e5	15/15
0: wen	41	3.2e3	1.6e3	60	60	59	1/15
1: F-S	37	3.0e3	1.1e3	60	59	59	1/15
f_{20}	16	850	3.8e4	5.4e4	5.5e4	5.5e4	14/15
0: wen	8.5	8.9	20	14	14	14	6/15
1: F-S	9	9	15	11	11	11	7/15
f_{21}	41	1200	1700	1700	1700	1800	14/15
0: wen	5.4	110	150	150	150	140	10/15
1: F-S	3.6	68	110	110	110	110	11/15
f_{22}	71	390	940	1000	1000	1100	14/15
0: wen	5	96	200	180	180	170	11/15
1: F-S	3.7	330	610	570	550	540	7/15
f_{23}	3	520	1.4e4	3.2e4	3.3e4	3.4e4	15/15
0: wen	2.1	9.7	2.1	3.1	4.8	5.9	15/15
1: F-S	1.4	8.8	2.3	3.2	4.9	6.1	15/15
f_{24}	1600	2.2e5	6.4e6	9.6e6	1.3e7	1.3e7	3/15
0: wen	3.7	15	0.52	0.34	0.26	0.26	2/15
1: F-S	4.7	4.8	0.32	0.21	0.16	0.16	3/15

20-D

Δf	1e+1	1e+0	1e-1	1e-3	1e-5	1e-7	#succ
f_1	43	43	43	43	43	43	15/15
0: wen	110	210	310	510	710	910	15/15
1: F-S	**90**^{*2}	**170**^{*3}	**260**^{*3}	**430**^{*3}	**600**^{*3}	**770**^{*3}	15/15
f_2	380	390	390	390	390	390	15/15
0: wen	57	68	80	100	120	140	15/15
1: F-S	**48**^{*3}	**57**^{*3}	**67**^{*3}	**86**^{*3}	**110**^{*3}	**120**^{*3}	15/15
f_3	5100	7600	7600	7600	7600	7700	15/15
0: wen	730	∞	∞	∞	∞	∞ *2.0e6*	0/15
1: F-S	2.9e3	∞	∞	∞	∞	∞ *2.0e6*	0/15
f_4	4700	7600	7700	7700	7800	1.4e5	9/15
0: wen	∞	∞	∞	∞	∞	∞ *2.0e6*	0/15
1: F-S	∞	∞	∞	∞	∞	∞ *2.0e6*	0/15
f_5	41	41	41	41	41	41	15/15
0: wen	72	87	88	88	88	88	15/15
1: F-S	**41**^{*3}	**50**^{*3}	**51**^{*3}	**51**^{*3}	**51**^{*3}	**51**^{*3}	15/15
f_6	1300	2300	3400	5200	6700	8400	15/15
0: wen	23	18	16	16	16	16	15/15
1: F-S	**18**^{*3}	**15**^{*3}	**13**^{*3}	**13**^{*3}	**14**^{*3}	**14**^{*3}	15/15
f_7	1400	4300	9500	1.7e4	1.7e4	1.7e4	15/15
0: wen	6.3	3.7	2.3	1.8	1.8	1.9	15/15
1: F-S	**5.2**^{*2}	**3.3***	17	10	10	10	14/15
f_8	2000	3900	4000	4200	4400	4500	15/15
0: wen	35	33	35	37	38	39	15/15
1: F-S	**22**^{*3}	99	97	96	94	94	13/15
f_9	1700	3100	3300	3500	3600	3700	15/15
0: wen	38	40	43	45	46	47	15/15
1: F-S	**26**^{*3}	**25**^{*3}	**27**^{*3}	**29**^{*3}	**30**^{*3}	**31**^{*3}	15/15
f_{10}	7400	8700	1.1e4	1.5e4	1.7e4	1.7e4	15/15
0: wen	2.9	3	2.8	2.6	2.8	3.2	15/15
1: F-S	**2.4**^{*3}	**2.5**^{*3}	**2.4**^{*3}	**2.2**^{*3}	**2.4**^{*3}	**2.8**^{*3}	15/15
f_{11}	1000	2200	6300	9800	1.2e4	1.5e4	15/15
0: wen	9.6	6.2	2.9	2.8	2.9	3	15/15
1: F-S	**7.5**^{*2}	**5.1**^{*3}	**2.4**^{*3}	**2.3**^{*3}	**2.4**^{*3}	**2.5**^{*3}	15/15
f_{12}	1000	1900	2700	4100	1.2e4	1.4e4	15/15
0: wen	31	19	17	22	12	13	15/15
1: F-S	**27**^{*2}	20	21	21	9.6	**10***	15/15
f_{13}	650	2000	2800	1.9e4	2.4e4	3.0e4	15/15
0: wen	30	14	13	2.8	2.9	2.9	15/15
1: F-S	**25**^{*3}	**12**^{*3}	**11**^{*3}	**2.4**^{*3}	**2.5**^{*3}	**2.5**^{*3}	15/15
f_{14}	75	240	300	930	1600	1.6e4	15/15
0: wen	46	37	47	27	22	3	15/15
1: F-S	**34**^{*2}	**29**^{*3}	**37**^{*3}	**22**^{*3}	**19**^{*2}	2.8	15/15
f_{15}	3.0e4	1.5e5	3.1e5	3.2e5	4.5e5	4.6e5	15/15
0: wen	230	∞	∞	∞	∞	∞ *2.0e6*	0/15
1: F-S	230	∞	∞	∞	∞	∞ *2.0e6*	0/15
f_{16}	1400	2.7e4	7.7e4	1.9e5	2.0e5	2.2e5	15/15
0: wen	2.1e4	∞	∞	∞	∞	∞ *2.0e6*	0/15
1: F-S	∞	∞	∞	∞	∞	∞ *2.0e6*	0/15
f_{17}	63	1000	4000	3.1e4	5.6e4	8.0e4	15/15
0: wen	23	14	7.5	2.1	2.1	3.6	14/15
1: F-S	19	13	**6.5**^{*2}	**1.9**^{*2}	4.4	7.9	12/15
f_{18}	620	4000	2.0e4	6.8e4	1.3e5	1.5e5	15/15
0: wen	14	5.5	2	1.2	2.1	2.1	14/15
1: F-S	**11**^{*2}	**4.8**^{*3}	**1.7**^{*2}	3.2	8.6	56	3/15
f_{19}	1	1	3.4e5	6.2e6	6.7e6	6.7e6	15/15
0: wen	1.9e3	1.4e7	∞	∞	∞	∞ *2.0e6*	0/15
1: F-S	**1.3e3***	1.4e7	∞	∞	∞	∞ *2.0e6*	0/15
f_{20}	82	4.6e4	3.1e6	5.5e6	5.6e6	5.6e6	14/15
0: wen	51	300	∞	∞	∞	∞ *2.0e6*	0/15
1: F-S	**42**^{*2}	300	∞	∞	∞	∞ *2.0e6*	0/15
f_{21}	560	6500	1.4e4	1.5e4	1.6e4	1.8e4	15/15
0: wen	13	210	210	210	190	170	6/15
1: F-S	11	270	280	270	260	230	5/15
f_{22}	470	5600	2.3e4	2.5e4	2.7e4	1.3e5	12/15
0: wen	1.6e3	990	∞	∞	∞	∞ *2.0e6*	0/15
1: F-S	1.6e3	2.3e3	∞	∞	∞	∞ *2.0e6*	0/15
f_{23}	3.2	1600	6.7e4	4.9e5	8.1e5	8.4e5	15/15
0: wen	2.4	940	430	∞	∞	∞ *2.0e6*	0/15
1: F-S	2	1.2e3	∞	∞	∞	∞ *2.0e6*	0/15
f_{24}	1.3e6	7.5e6	5.2e7	5.2e7	5.2e7	5.2e7	3/15
0: wen	∞	∞	∞	∞	∞	∞ *2.0e6*	0/15
1: F-S	∞	∞	∞	∞	∞	∞ *2.0e6*	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 $f_{\mathrm{opt}} + 10^{-8}$. 0: wen is ALG0-acronym and 1: F-S is ALG1-acronym. Bold entries are statistically significantly better compared to the other algorithm, with $p = 0.05$ or $p = 10^{-k}$ where $k > 1$ is the number following the $\star$ symbol, with Bonferroni correction of 48.