Testing the Impact of Parameter Tuning on a Variant of IPOP-CMA-ES with a Bounded Maximum Population Size on the Noiseless BBOB Testbed

Tianjun Liao
IRIDIA, CoDE, Université Libre de Bruxelles (ULB), Brussels, Belgium
tliao@ulb.ac.be

Thomas Stützle
IRIDIA, CoDE, Université Libre de Bruxelles (ULB), Brussels, Belgium
stuetzle@ulb.ac.be

ABSTRACT

In this paper, we experimentally explore the influence tuned parameter settings have on an IPOP-CMA-ES variant that uses a maximum bound on the population size. We followed our earlier work, where we exposed seven parameters that control parameters of IPOP-CMA-ES, and tune them by applying `irace`, an automatic algorithm configuration tool. A comparison of the tuned to the default settings on the BBOB benchmark shows that for difficult problems such as multi-modal functions with weak global structure, the tuned parameter settings can result in significant improvements over the default settings.

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. INTRODUCTION

Assigning appropriate values for the parameters of optimization algorithms is an important task [4]. Over the recent years, evidence has been given that the performance of many algorithms can be improved by using automatic algorithm configuration and tuning tools [1–3, 5, 10, 11, 17]. Many successful studies involve configuring discrete optimization algorithms [2, 5, 10], but also the tuning of continuous optimization algorithms has received some attention [3,11–13,16,19,20]. Even if automatic algorithm configuration tools sometimes do not result in very strong performance improvements, they have the advantage of reducing typically the time necessary for tuning parameter values and they help in reducing the bias in algorithm comparisons.

In this paper, we tune the parameters of a IPOP-CMA-ES variant for which we have presented computational results in another to BBOB 2013 [14]. That variant, to which we refer as IP-10DDr, uses a bound on the maximum population size to be used in IPOP-CMA-ES and, once this bound is reached, it restarts the scheme for the variation of the population size at the initial default value used by IPOP-CMA-ES. Here, we experimentally explore the influence of the tuned parameter settings in the performance of IP-10DDr. We found that for the difficult problems such as multi-modal functions with weak global structure, the tuned parameter settings can result in significant improvements over the default ones.

2. EXPERIMENTAL PROCEDURE

For tuning IP-10DDr, we considered seven parameters related to the above mentioned default settings, following previous work we have presented in [13]. The seven default settings of IP-10DDr are as follows. The initial population size is $\lambda = 4 + \lfloor 3\ln(D) \rfloor$, where D is the number of dimensions of the function to be optimized. The number of selected search points in the parent population is $\mu = \lfloor 0.5\lambda \rfloor$. The initial step-size is $\sigma^{(0)} = 0.5(B - A)$, where $[A, B]^D$ is the initial search interval. At each restart, the population size is multiplied by a factor of two. Restarts occur if the stopping criterion is met. The three parameters are *stopTolFunHist*$(= 10^{-20})$, *stopTolFun*$(= 10^{-12})$ and *stopTolX*$(= 10^{-12})$; they refer to refer to the range of the improvement of the best objective function values in the last $10 + \lceil 30D/\lambda \rceil$ generations, all function values of the recent generation, and the standard deviation of the normal distribution in all coordinates, respectively.

The tuned parameters are given in Table 1. The first four parameters are actually used in a formula to compute some internal parameters of IP-10DDr and the remaining three are used to define the termination of CMA-ES. The first three columns of Table 1 give the parameters we use, the formula where they are used, the range that we considered for tuning.

As tuner we use `irace` [15], a publicly available implementation of the automatic configuration method Iterated F-Race [2]. The budget of each run of `irace` is set to 5 000 runs of IP-10DDr. We consider a separation between tun-

Table 1: Parameters that have been considered for tuning. Given are the continuous range we considered for tuning. The last two columns are the parameter settings obtained in [16] (tany) and for the tuning for the final solution quality at $100 \times D$ function evaluations (texp) respectively.

Para (tuning)	Internal parameter	Range	Tuned tany	Tuned texp
a	Init pop size: $\lambda_0 = 4 + \lfloor a \ln(D) \rfloor$	$[1, 10]$	3.676	2.675
b	Parent size: $\mu = \lfloor \lambda / b \rfloor$	$[1, 5]$	1.750	1.351
c	Init step size: $\sigma_0 = c \cdot (B - A)$	$(0, 1)$	0.325	0.102
d	IPOP factor: $ipop = d$	$[1, 4]$	1.840	2.88
e	$stopTolFun = 10^e$	$[-20, -6]$	-9.653	-8.607
f	$stopTolFunHist = 10^f$	$[-20, -6]$	-10.000	-14.77
g	$stopTolX = 10^g$	$[-20, -6]$	-9.528	-9.529

ing and test sets. The training instances are the same as those used in [16], which are a subset of the functions in the SOCO benchmark [9], with dimension $D \in [5, 40]$. The performance measure used for tuning is the error of the objective function value obtained by the tuned algorithm after $100 \times D$ function evaluations. It should be highlighted that we focus here on the first 100D function evaluations as the same results are re-used in the expensive optimization scenario at BBOB 2013 and we wanted to avoid a re-tuning. (In fact, the exploration of the impact different tuning setups have on the performance of the tuned IP-10DDr we leave for future work.) From this tuning we obtain a configuration of IP-10DDr we call henceforth `texp`. We also use the parameter settings we obtained in another paper [16], where IPOP-CMA-ES was tuned to improve its anytime behavior; the algorithm using these parameter settings is labeled as `tany` (Table 1). The default parameter settings are labeled `def` in this paper.

3. RESULTS

Results from experiments according to [7] on the benchmark functions given in [6, 8] are presented in Figures 1, 2 and 3 and in Tables 2 and 3. 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 [7, 18]. **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.

In the experiments, we clearly observe that performance may vary with the parameter settings on at least some of the functions. On some functions, `texp` is slower in reaching optimal threshold that either `tany` or `def`, the most noteworthy example being f_5 ($D = 2, 3, 40$). However, `texp` performs better than `tany` or `def` on several multi-modal functions with weak global structure. For example, it reaches the optimal threshold in f_{22} ($D = 20$) and f_{24} ($D = 5, 20$) where `tany` and `def` cannot do so. `tany` reaches optimal threshold in f_{23} ($D = 40$) and f_{24} ($D = 40$) where `texp` and `def` cannot do so. We also clearly observe that `texp` uses fewer function evaluations to reach optimal threshold in f_{21} ($D = 10, 20, 40$), f_{22} ($D = 2, 3, 5, 10$), f_{24} ($D = 2, 3$) than `tany` and `def`. Only on f_{23} ($D = 2, 3, 5$), `def` obviously uses fewer function evaluations to reach optimal threshold than `tany` and `texp`. The overview of the overall results in Figure 2 confirms the observation above: `texp` performs better than `tany` or `def` especially on the weakly structured multi-modal functions.

4. CPU TIMING EXPERIMENT

The `texp` and `tany` were run on f_8 until at least 30 seconds have passed. These experiment were conducted with Intel Xeon E5410 (2.33 GHz) on Linux (kernel 2.6.9 - 78.0.22). The results of `texp` were 1.6E−05, 2.4E−05, 6.2E−06, 9.7E−06, 1.5E−05 and 5.8E−05 seconds per function evaluation in dimensions 2, 3, 5, 10, 20, and 40, respectively. The results of `tany` were 1.8E−05, 2.4E−05, 1.7E−05, 7.5E−06, 1.3E−05 and 4.6E−05 seconds per function evaluation in dimensions 2, 3, 5, 10, 20, and 40, respectively.

5. CONCLUSIONS

In this paper we have examined the influence of tuning on the performance of IP-10DDr. When compared to default settings, further performance improvements could be observed on few difficult functions. These results suggest that it would be interesting to further explore the impact different tuning setups have. In fact, here we used a setting where we considered only very short runs for tuning and also a tuning target, where we try to obtain for a fixed number of function evaluations as good as possible solutions. Changing the tuning setup to minimize the number of function evaluations to target, the point of view actually followed in the BBOB benchmark analysis, we would expect further improved performance.

6. ACKNOWLEDGMENTS

The authors would like to thank the great and hard work of the BBOB team. This work was supported by the Meta-X project funded by the Scientific Research Directorate of the French Community of Belgium. Thomas Stützle acknowledges support from the Belgian F.R.S.-FNRS, of which he is a Research Associate. Tianjun Liao acknowledges a fellowship from the China Scholarship Council.

7. REFERENCES

[1] B. Adenso-Diaz and M. Laguna. Fine-tuning of algorithms using fractional experimental designs and local search. *Operations Research*, 54:99–114, 2006.

[2] P. Balaprakash, M. Birattari, and T. Stützle. Improvement strategies for the F-Race algorithm: sampling design and iterative refinement. In Bartz-Beielstein et al., editors, *Proceedings of the 4th International Conference on Hybrid Metaheuristics*, volume 4771 of *LNCS*, pages 108–122. Springer, Berlin, Germany, 2007.

[3] T. Bartz-Beielstein. *Experimental Research in Evolutionary Computation: the New Experimentalism.* Springer, Berlin, Germany, 2006.

[4] M. Birattari. *Tuning Metaheuristics: A Machine Learning Perspective.* Springer, Berlin, Germany, 1st ed. 2005. 2nd printing edition, 2009.

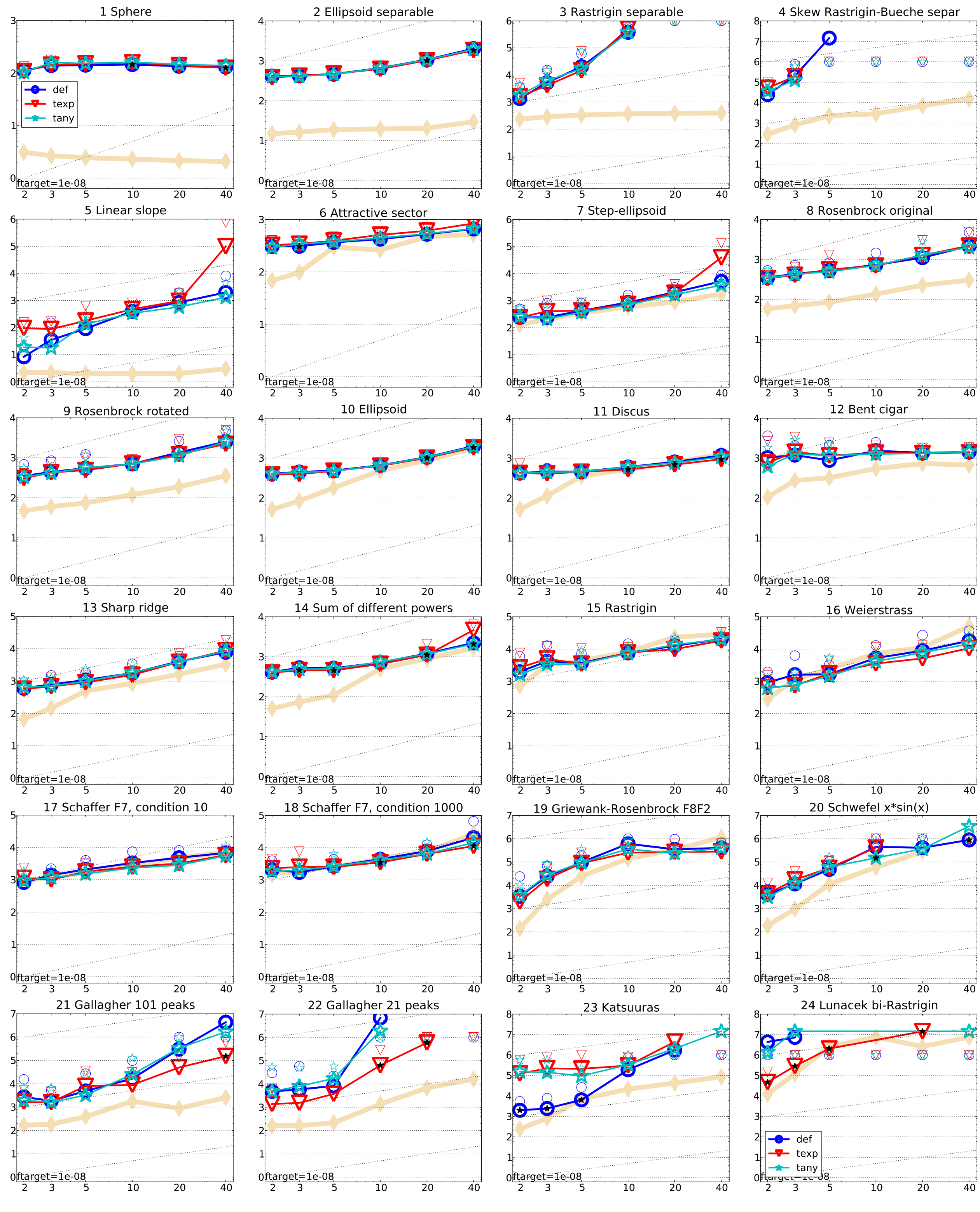

Figure 1: Expected running time (ERT in number of f-evaluations) divided by dimension for target function value 10^{-8} as $\log_{10}$ values versus dimension. Different symbols correspond to different algorithms given in the legend of f_1 and f_{24}. Light symbols give the maximum number of function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

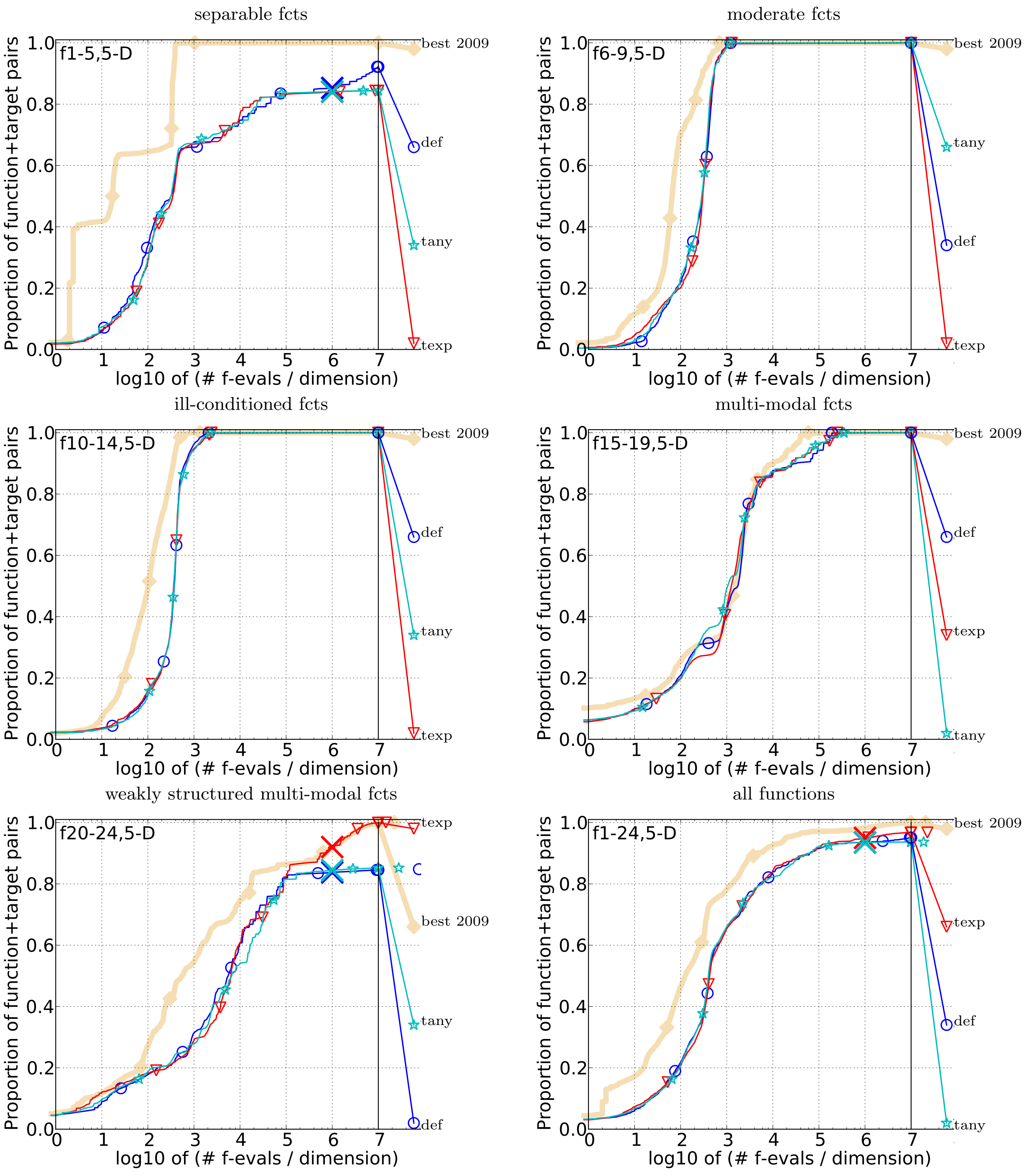

Figure 2: Bootstrapped empirical cumulative distribution of the number of objective function evaluations divided by dimension (FEvals/D) for 50 targets in $10^{[-8..2]}$ for all functions and subgroups in 5-D. The "best 2009" line corresponds to the best ERT observed during BBOB 2009 for each single target.

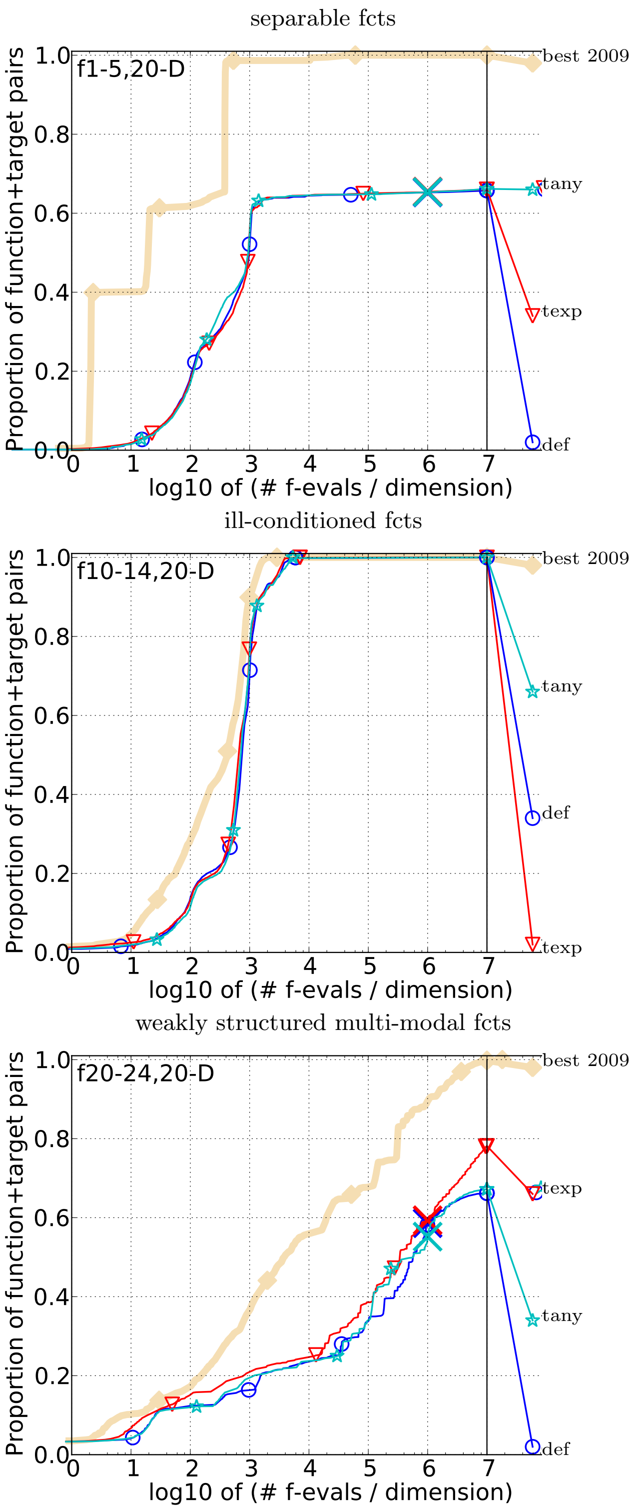

Figure 3: Bootstrapped empirical cumulative distribution of the number of objective function evaluations divided by dimension (FEvals/D) for 50 targets in $10^{[-8..2]}$ for all functions and subgroups in 20-D. The "best 2009" line corresponds to the best ERT observed during BBOB 2009 for each single target.

Δf_{opt}	1e1	1e0	1e-1	1e-3	1e-5	1e-7	#succ
f1	11	12	12	12	12	12	15/15
def	3.0(1)	7.5(1)	**13**(3)	**27**(3)	**40**(4)	**53**(6)	15/15
texp	**2.7**(2)	**6.7**(3)	13(6)	29(4)	44(4)	55(4)	15/15
tany	2.7(2)	7.4(2)	16(4)	30(3)	44(3)	56(3)	15/15
f2	83	87	88	90	92	94	15/15
def	**13**(4)	17(5)	19(4)	21(3)	23(3)	24(2)	15/15
texp	15(4)	18(4)	20(2)	22(2)	23(2)	24(2)	15/15
tany	15(3)	**17**(3)	**18**(2)	**20**(1)	**22**(1.0)	**23**(1)	15/15
f3	716	1622	1637	1646	1650	1654	15/15
def	0.99(1)	15(21)	60(67)	60(67)	61(67)	61(67)	15/15
texp	1.1(1)	10(12)	**42**(35)	**43**(35)	**43**(34)	**44**(34)	15/15
tany	**0.94**(1)	**7.2**(8)	51(41)	51(41)	51(40)	52(40)	15/15
f4	809	1633	1688	1817	1886	1903	15/15
def	2.6(3)	**2721**(2484)	**4.3e4**(5e4)	**4.0e4**(4e4)	**3.8e4**(4e4)	**3.8e4**(4e4)	1/15
texp	**1.6**(2)	5050(4796)	∞	∞	∞	∞ *5e6*	0/15
tany	2.0(2)	3009(3498)	∞	∞	∞	∞ *5e6*	0/15
f5	10	10	10	10	10	10	15/15
def	**6.2**(3)	**23**(18)	**30**(25)	**41**(33)	**45**(43)	**46**(43)	15/15
texp	12(11)	36(18)	49(24)	67(36)	76(50)	88(71)	15/15
tany	8.6(7)	32(24)	40(27)	54(31)	62(42)	65(53)	15/15
f6	114	214	281	580	1038	1332	15/15
def	1.6(0.5)	**1.8**(0.4)	**2.1**(0.3)	**1.6**(0.2)	**1.2**(0.1)	**1.2**(0.1)	15/15
texp	1.7(1.0)	2.0(0.6)	2.2(0.4)	1.7(0.3)	1.3(0.2)	1.3(0.2)	15/15
tany	**1.5**(0.7)	1.9(0.5)	2.1(0.4)	1.6(0.2)	1.3(0.1)	1.3(0.1)	15/15
f7	24	324	1171	1572	1572	1597	15/15
def	5.3(2)	**1.1**(0.4)	**1.2**(0.9)	1.2(0.7)	1.2(0.7)	1.3(0.8)	15/15
texp	**2.7**(2)	2.1(2)	1.4(0.7)	1.2(0.6)	1.2(0.6)	1.3(0.6)	15/15
tany	3.5(2)	1.7(2)	1.3(1)	**1.1**(0.9)	**1.1**(0.9)	**1.1**(0.9)	15/15
f8	73	273	336	391	410	422	15/15
def	3.2(1.0)	4.3(2)	5.1(1)	5.5(1)	5.8(1)	6.1(1)	15/15
texp	**2.6**(0.7)	4.7(2)	5.6(2)	5.9(2)	6.1(2)	6.4(2)	15/15
tany	2.7(0.7)	**3.7**(1)	**4.6**(1)	**5.0**(1)	**5.3**(1)	**5.6**(1)	15/15
f9	35	127	214	300	335	369	15/15
def	6.1(2)	10(10)	8.6(6)	7.6(5)	7.4(4)	7.2(4)	15/15
texp	**3.0**(2)$^{\star 2}$	**8.1**(4)	**7.4**(2)	**6.6**(2)	**6.5**(1)	**6.4**(1)	15/15
tany	5.2(2)	11(13)	8.7(8)	7.6(6)	7.5(6)	7.4(5)	15/15
f10	349	500	574	626	829	880	15/15
def	3.4(0.9)	**3.1**(0.6)	3.1(0.3)	3.2(0.3)	2.6(0.2)	2.7(0.2)	15/15
texp	**3.4**(1)	3.1(0.7)	3.1(0.4)	3.1(0.3)	2.6(0.2)	2.6(0.2)	15/15
tany	3.4(0.9)	3.1(0.6)	**2.9**(0.4)	**3.0**(0.3)	**2.5**(0.3)	**2.6**(0.2)	15/15
f11	143	202	763	1177	1467	1673	15/15
def	7.4(4)	7.1(0.7)	2.1(0.2)	1.6(0.1)	1.4(0.1)	1.3(0.1)	15/15
texp	**6.9**(3)	**6.7**(1)	**2.0**(0.3)	**1.5**(0.1)	**1.3**(0.1)	**1.3**(0.1)	15/15
tany	8.1(1)	7.1(0.7)	2.1(0.2)	1.5(0.1)	1.4(0.1)	1.3(0.1)	15/15
f12	108	268	371	461	1303	1494	15/15
def	7.9(5)	**5.9**(3)	**6.0**(3)	**6.3**(4)	**2.8**(2)	**2.8**(2)	15/15
texp	11(10)	8.7(9)	8.4(8)	8.5(8)	3.7(3)	3.7(3)	15/15
tany	**6.2**(4)	7.1(5)	7.5(7)	8.2(5)	3.7(2)	3.7(2)	15/15

Δf_{opt}	1e1	1e0	1e-1	1e-3	1e-5	1e-7	#succ
f13	132	195	250	1310	1752	2255	15/15
def	4.7(4)	5.8(4)	**5.7**(3)	**1.5**(0.7)	1.9(0.8)	1.9(0.8)	15/15
texp	**3.7**(3)	**5.4**(3)	6.5(2)	1.6(0.3)	1.7(0.3)	**1.7**(0.4)	15/15
tany	3.9(3)	5.5(3)	5.9(2)	1.6(0.4)	**1.7**(0.2)	1.7(0.3)	15/15
f14	10	41	58	139	251	476	15/15
def	1.2(1)	2.4(1)	3.7(1.0)	4.7(0.9)	5.7(0.4)	4.6(0.4)	15/15
texp	**0.93**(1)	**1.9**(1)	**3.0**(1)	**3.8**(0.6)	**4.9**(0.7)	**4.0**(0.3)	15/15
tany	0.94(1)	2.6(1)	3.9(2)	4.5(1)	5.5(0.7)	4.2(0.4)	15/15
f15	511	9310	19369	20073	20769	21359	14/15
def	**1.5**(0.6)	1.1(0.9)	0.85(0.4)	0.85(0.4)	0.85(0.4)	0.86(0.4)	15/15
texp	2.1(2)	0.94(0.9)	**0.76**(0.8)	**0.78**(0.7)	**0.79**(0.7)	**0.80**(0.7)	15/15
tany	1.8(2)	**0.76**(0.6)	0.80(0.5)	0.80(0.6)	0.80(0.6)	0.81(0.5)	15/15
f16	120	612	2662	10449	11644	12095	15/15
def	2.0(2)	2.2(3)	1.1(1)	**0.58**(0.5)	0.67(0.3)	0.67(0.3)	15/15
texp	**1.1**(1)	**1.5**(2)	1.2(1)	0.70(0.3)	0.68(0.3)	0.70(0.3)	15/15
tany	1.9(1)	2.1(3)	**0.90**(0.6)	0.62(0.7)	**0.59**(0.6)	**0.59**(0.6)	15/15
f17	5.2	215	899	3669	6351	7934	15/15
def	**3.2**(3)	**0.79**(0.2)	**0.77**(0.2)	0.75(0.5)	1.1(0.4)	1.3(0.4)	15/15
texp	47(7)	3.0(7)	2.8(6)	1.2(2)	1.1(0.7)	1.1(0.6)	15/15
tany	3.6(6)	1.1(0.6)	1.1(2)	**0.71**(0.5)	**0.83**(0.7)	**0.93**(0.6)	15/15
f18	103	378	3968	9280	10905	12469	15/15
def	1.1(0.4)	**1.8**(0.3)	**0.56**(0.5)	1.0(0.3)	0.97(0.3)	1.0(0.1)	15/15
texp	1.3(0.8)	2.8(5)	0.81(0.6)	**0.88**(0.4)	**0.90**(0.4)	**0.98**(0.4)	15/15
tany	**0.95**(0.4)	1.8(0.7)	1.2(1)	1.0(0.7)	1.1(0.7)	1.0(0.6)	15/15
f19	1	1	242	1.2e5	1.2e5	1.2e5	15/15
def	13(14)	**1466**(1310)	519(658)	4.0(3)	4.0(3)	4.0(3)	15/15
texp	**3.5**(2)	1812(2224)	746(669)	**3.7**(4)	**3.7**(4)	**3.7**(4)	15/15
tany	13(16)	2343(3986)	**391**(388)	3.9(5)	3.9(5)	3.9(5)	15/15
f20	16	851	38111	54470	54861	55313	14/15
def	3.4(1)	**12**(11)	6.2(5)	**4.4**(4)	**4.4**(4)	**4.4**(4)	15/15
texp	**1.6**(0.6)	17(19)	7.1(6)	5.0(4)	5.0(4)	5.0(4)	15/15
tany	3.0(2)	12(10)	**5.0**(5)	5.5(7)	5.5(7)	5.4(7)	15/15
f21	41	1157	1674	1705	1729	1757	14/15
def	1.6(1)	11(12)	14(18)	14(18)	14(17)	14(17)	15/15
texp	**1.1**(0.6)	13(16)	23(24)	23(24)	23(24)	23(23)	15/15
tany	1.6(1)	**3.8**(5)	**9.3**(8)	**9.3**(8)	**9.2**(8)	**9.2**(8)	15/15
f22	71	386	938	1008	1040	1068	14/15
def	**7.4**(11)	42(53)	43(59)	40(55)	39(53)	39(52)	15/15
texp	22(9)	**13**(24)	**18**(20)	**18**(19)	**17**(19)	**17**(18)	15/15
tany	12(9)	29(51)	91(93)	85(87)	82(84)	80(82)	15/15
f23	3.0	518	14249	31654	33030	34256	15/15
def	2.1(2)	7.2(5)	**1.7**(1)	**0.96**(0.8)	**0.94**(0.8)	**0.93**(0.8)	15/15
texp	3.0(3)	**7.0**(7)	6.4(8)	3.8(4)	3.6(3)	3.5(3)	14/15
tany	**1.7**(1)	7.5(6)	5.2(5)	2.7(2)	2.7(2)	2.6(2)	15/15
f24	1622	2.2e5	6.4e6	9.6e6	1.3e7	1.3e7	3/15
def	2.0(2)	104(122)	∞	∞	∞	∞ *5e6*	0/15
texp	**1.4**(1)	**0.76**(1)$^{\star 3}$	**0.50**(0.5)$^{\star 4}$	**1.1**(1)$^{\star 4}$	**0.80**(0.8)$^{\star 4}$	**0.80**(0.8)$^{\star 4}$	6/15
tany	2.2(2)	20(18)	∞	∞	∞	∞ *5e6*	0/15

Table 2: Expected running time (ERT in number of function evaluations) divided by the respective best ERT measured during BBOB-2009 (given in the respective first row) for different Δf values in dimension 5. The central 80% range divided by two is given in braces. 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 $f_{\mathrm{opt}} + 10^{-8}$. Best results are printed in bold.

[5] M. Birattari, T. Stützle, L. Paquete, and K. Varrentrapp. A racing algorithm for configuring metaheuristics. In *Proceedings of the Genetic and Evolutionary Computation Conference, GECCO'02*, pages 11–18, San Francisco, CA, USA, 2002. Morgan Kaufmann.

[6] 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.

[7] N. Hansen, A. Auger, S. Finck, and R. Ros. Real-parameter black-box optimization benchmarking 2012: Experimental setup. Technical report, INRIA, 2012.

[8] 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.

[9] F. Herrera, M. Lozano, and D. Molina. Test suite for the special issue of soft computing on scalability of evolutionary algorithms and other metaheuristics for large scale continuous optimization problems, 2010. URL: `http://sci2s.ugr.es/eamhco/`.

[10] F. Hutter, H. Hoos, K. Leyton-Brown, and T. Stützle. ParamILS: an automatic algorithm configuration framework. *Journal of Artificial Intelligence Research*, 36(1):267–306, 2009.

[11] F. Hutter, H. H. Hoos, K. Leyton-Brown, and K. P. Murphy. An experimental investigation of model-based parameter optimisation: Spo and beyond. In *Proceedings of the Genetic and Evolutionary Computation Conference, GECCO'09*, pages 271–278, New York, NY, USA, 2009. ACM.

Δf_{opt}	1e1	1e0	1e-1	1e-3	1e-5	1e-7	#succ
f1	43	43	43	43	43	43	15/15
def	7.5(0.6)	14(0.7)	20(2)	**32**(3)	**45**(2)	**57**(3)	15/15
texp	**6.6**(1)	**13**(1)	**20**(2)	33(2)	47(2)	59(4)	15/15
tany	7.7(0.9)	15(0.9)	22(1)	36(1)	49(3)	61(4)	15/15

Δf_{opt}	1e1	1e0	1e-1	1e-3	1e-5	1e-7	#succ
f2	385	386	387	390	391	393	15/15
def	35(4)	42(4)	48(4)	51(2)	52(2)	53(2)	15/15
texp	**34**(5)	**41**(5)	**45**(5)	**49**(2)	**50**(2)	**51**(2)*	15/15
tany	35(4)	42(5)	46(4)	50(3)	52(2)	53(2)	15/15

Δf_{opt}	1e1	1e0	1e-1	1e-3	1e-5	1e-7	#succ
f3	5066	7626	7635	7643	7646	7651	15/15
def	15(16)	∞	∞	∞	∞	∞ *2e7*	0/15
texp	**8.5**(8)	3.9e4(4e4)	∞	∞	∞	∞ *2e7*	0/15
tany	10(7)	**3.8e4**(4e4)	∞	∞	∞	∞ *2e7*	0/15

Δf_{opt}	1e1	1e0	1e-1	1e-3	1e-5	1e-7	#succ
f4	4722	7628	7666	7700	7758	1.4e5	9/15
def	**2.9e4**(3e4)	∞	∞	∞	∞	∞ *2e7*	0/15
texp	∞	∞	∞	∞	∞	∞ *2e7*	0/15
tany	6.1e4(7e4)	∞	∞	∞	∞	∞ *2e7*	0/15

Δf_{opt}	1e1	1e0	1e-1	1e-3	1e-5	1e-7	#succ
f5	41	41	41	41	41	41	15/15
def	**51**(35)	**111**(93)	165(106)	233(151)	309(245)	403(293)	15/15
texp	92(122)	177(150)	231(205)	309(284)	357(306)	436(324)	15/15
tany	69(89)	134(146)	**159**(157)	**202**(173)	**235**(202)	**274**(212)	15/15

Δf_{opt}	1e1	1e0	1e-1	1e-3	1e-5	1e-7	#succ
f6	1296	2343	3413	5220	6728	8409	15/15
def	**1.3**(0.2)	**1.1**(0.2)	**1.1**(0.2)	**1.1**(0.1)	**1.1**(0.1)	**1.1**(0.1)	15/15
texp	1.4(0.2)	1.2(0.2)	1.2(0.2)	1.2(0.2)	1.3(0.2)	1.3(0.2)	15/15
tany	1.4(0.4)	1.2(0.2)	1.1(0.2)	1.1(0.2)	1.2(0.1)	1.2(0.1)	15/15

Δf_{opt}	1e1	1e0	1e-1	1e-3	1e-5	1e-7	#succ
f7	1351	4274	9503	16524	16524	16969	15/15
def	1.7(2)	6.2(3)	3.9(2)	2.4(1.0)	2.4(1.0)	2.4(1.0)	15/15
texp	1.6(2)	6.5(4)	3.7(3)	2.3(2)	2.3(2)	2.2(2)	15/15
tany	**1.0**(0.3)	**4.6**(2)	**3.2**(0.8)	**2.0**(0.5)	**2.0**(0.5)	**1.9**(0.5)	15/15

Δf_{opt}	1e1	1e0	1e-1	1e-3	1e-5	1e-7	#succ
f8	2039	3871	4040	4219	4371	4484	15/15
def	**4.1**(0.8)	**4.3**(0.3)	**4.6**(0.4)	**4.8**(0.3)	**4.8**(0.3)	**4.9**(0.3)	15/15
texp	4.4(0.9)	5.0(0.5)	5.4(0.4)	5.5(0.4)	5.5(0.4)	5.6(0.4)	15/15
tany	4.6(1)	5.1(0.6)	5.5(0.6)	5.6(0.6)	5.6(0.6)	5.7(0.5)	15/15

Δf_{opt}	1e1	1e0	1e-1	1e-3	1e-5	1e-7	#succ
f9	1716	3102	3277	3455	3594	3727	15/15
def	5.0(1)	6.9(5)	7.2(5)	7.3(4)	7.3(4)	7.2(4)	15/15
texp	**4.8**(1)	6.0(0.6)	6.3(0.6)	6.5(0.6)	6.5(0.5)	6.5(0.5)	15/15
tany	5.0(1)	**5.3**(0.7)	**5.7**(0.7)	**6.0**(0.6)	**6.0**(0.6)	**6.0**(0.6)	15/15

Δf_{opt}	1e1	1e0	1e-1	1e-3	1e-5	1e-7	#succ
f10	7413	8661	10735	14920	17073	17476	15/15
def	1.8(0.2)	1.9(0.2)	1.7(0.1)	1.3(0.0)	1.2(0.0)	1.2(0.0)	15/15
texp	**1.7**(0.2)	**1.7**(0.2)	**1.6**(0.1)	**1.2**(0.1)	**1.1**(0.1)	**1.1**(0.0)	15/15
tany	1.8(0.3)	1.8(0.2)	1.6(0.1)	1.3(0.0)	1.2(0.0)	1.2(0.0)	15/15

Δf_{opt}	1e1	1e0	1e-1	1e-3	1e-5	1e-7	#succ
f11	1002	2228	6278	9762	12285	14831	15/15
def	10(0.6)	5.1(0.2)	2.0(0.1)	1.4(0.0)	1.2(0.0)	1.1(0.0)	15/15
texp	**8.4**(0.6)$^{\star 3}$	**4.3**(0.3)$^{\star 3}$	**1.7**(0.1)$^{\star 4}$	**1.2**(0.1)$^{\star 4}$	**1.0**(0.1)$^{\star 4}$	**0.90**(0.0)$^{\star 4}_{\downarrow 4}$	15/15
tany	9.4(0.4)	4.8(0.2)	1.9(0.1)	1.3(0.0)	1.1(0.0)	0.99(0.0)	15/15

Δf_{opt}	1e1	1e0	1e-1	1e-3	1e-5	1e-7	#succ
f12	1042	1938	2740	4140	12407	13827	15/15
def	2.6(2)	3.6(3)	4.1(3)	4.1(2)	1.8(0.6)	1.9(0.5)	15/15
texp	**2.3**(0.2)	3.0(3)	**3.4**(3)	**3.8**(1)	**1.7**(0.5)	1.8(0.5)	15/15
tany	2.4(0.2)	**2.7**(2)	3.7(2)	3.9(1)	1.7(0.5)	**1.8**(0.4)	15/15

Δf_{opt}	1e1	1e0	1e-1	1e-3	1e-5	1e-7	#succ
f13	652	2021	2751	18749	24455	30201	15/15
def	**2.4**(0.3)	**4.3**(5)	**6.4**(4)	**1.7**(1)	2.2(0.9)	2.3(0.9)	15/15
texp	3.8(4)	4.3(4)	6.9(6)	1.7(0.7)	**2.0**(0.8)	**2.2**(0.3)	15/15
tany	4.4(6)	4.8(3)	6.7(6)	1.7(1)	2.0(1)	2.5(1)	15/15

Δf_{opt}	1e1	1e0	1e-1	1e-3	1e-5	1e-7	#succ
f14	75	239	304	932	1648	15661	15/15
def	3.8(0.7)	2.7(0.5)	3.4(0.4)	4.2(0.4)	6.3(0.4)	1.2(0.1)	15/15
texp	**1.6**(0.8)$^{\star 3}$	**2.3**(0.3)	**3.2**(0.4)	**4.1**(0.4)	**5.8**(0.4)*	**1.1**(0.1)$^{\star 2}$	15/15
tany	3.6(1)	3.0(0.5)	3.8(0.3)	4.5(0.5)	6.4(0.5)	1.2(0.1)	15/15

Δf_{opt}	1e1	1e0	1e-1	1e-3	1e-5	1e-7	#succ
f15	30378	1.5e5	3.1e5	3.2e5	4.5e5	4.6e5	15/15
def	1.1(0.7)	0.99(0.5)	0.70(0.2)	0.71(0.2)	0.53(0.2)$_{\downarrow 2}$	0.54(0.2)$_{\downarrow 2}$	15/15
texp	**0.69**(0.4)	**0.75**(0.4)	**0.52**(0.2)	**0.54**(0.2)	**0.41**(0.2)$_{\downarrow 4}$	**0.42**(0.2)$_{\downarrow 4}$	15/15
tany	0.88(0.5)	1.0(0.3)	0.77(0.2)	0.78(0.2)	0.58(0.1)$_{\downarrow 3}$	0.59(0.1)$_{\downarrow 3}$	15/15

Δf_{opt}	1e1	1e0	1e-1	1e-3	1e-5	1e-7	#succ
f16	1384	27265	77015	1.9e5	2.0e5	2.2e5	15/15
def	1.2(0.3)	0.27(0.3)	0.42(0.3)$_{\downarrow 2}$	0.78(0.6)	0.85(0.6)	0.79(0.5)	15/15
texp	**0.86**(0.3)	**0.16**(0.2)	**0.37**(0.2)$_{\downarrow 3}$	**0.34**(0.2)$_{\downarrow 3}$	**0.47**(0.5)	**0.45**(0.4)	15/15
tany	1.3(0.4)	0.29(0.3)	0.46(0.1)$_{\downarrow 2}$	0.69(0.5)	0.73(0.5)	0.70(0.4)	15/15

Δf_{opt}	1e1	1e0	1e-1	1e-3	1e-5	1e-7	#succ
f17	63	1030	4005	30677	56288	80472	15/15
def	1.9(0.8)	**0.78**(0.1)$_{\downarrow}$	**0.67**(0.1)	0.93(0.4)	1.1(0.4)	1.1(0.5)	15/15
texp	**1.2**(0.9)	0.96(0.2)	0.99(2)	0.82(0.5)	0.82(0.3)	0.73(0.2)$_{\downarrow 2}$	15/15
tany	1.7(0.8)	0.96(0.2)	0.75(0.2)	**0.64**(0.5)$_{\downarrow}$	**0.67**(0.2)$_{\downarrow}$	**0.68**(0.2)$_{\downarrow 2}$	15/15

Δf_{opt}	1e1	1e0	1e-1	1e-3	1e-5	1e-7	#succ
f18	621	3972	19561	67569	1.3e5	1.5e5	15/15
def	0.87(0.2)	0.46(0.2)	0.66(0.4)	1.2(0.8)	0.96(0.4)	1.0(0.4)	15/15
texp	0.97(0.2)	1.3(2)	0.85(0.9)	**0.84**(0.0)	**0.80**(0.4)	**0.80**(0.4)	15/15
tany	**0.85**(0.2)	**0.45**(0.1)	**0.48**(0.4)	0.92(0.3)	0.87(0.2)	0.86(0.2)	15/15

Δf_{opt}	1e1	1e0	1e-1	1e-3	1e-5	1e-7	#succ
f19	1	1	3.4e5	6.2e6	6.7e6	6.7e6	15/15
def	178(144)	5.2e5(2e5)	4.0(3)	0.84(0.6)	1.0(0.9)	1.0(0.9)	15/15
texp	**3.3**(2)	4.5e4(5e4)	**2.0**(2)	**0.67**(0.6)	0.80(0.8)	0.80(0.8)	15/15
tany	130(116)	**3.6e4**(3e4)	3.7(4)	0.69(0.4)$_{\downarrow}$	**0.70**(0.4)	**0.70**(0.4)	15/15

Δf_{opt}	1e1	1e0	1e-1	1e-3	1e-5	1e-7	#succ
f20	82	46150	3.1e6	5.5e6	5.6e6	5.6e6	14/15
def	5.4(1)	**5.6**(3)	0.91(0.6)	1.5(1)	1.5(1)	1.5(1)	15/15
texp	**2.6**(0.9)$^{\star 3}$	13(7)	48(51)	∞	∞	∞ *2e7*	0/15
tany	5.5(0.9)	8.3(4)	**0.81**(0.4)$_{\downarrow 2}$	**1.3**(2)	**1.3**(2)	**1.3**(2)	14/15

Δf_{opt}	1e1	1e0	1e-1	1e-3	1e-5	1e-7	#succ
f21	561	6541	14103	14643	15567	17589	15/15
def	6.3(11)	147(189)	430(665)	414(641)	390(603)	345(533)	14/15
texp	2.8(4)	**47**(68)	**70**(94)	**67**(91)	**64**(85)	**56**(76)	15/15
tany	**2.5**(5)	51(70)	479(735)	461(718)	434(675)	384(598)	12/15

Δf_{opt}	1e1	1e0	1e-1	1e-3	1e-5	1e-7	#succ
f22	467	5580	23491	24948	26847	1.3e5	12/15
def	96(168)	1871(3584)	∞	∞	∞	∞ *2e7*	0/15
texp	**62**(42)	**532**(830)	**502**(586)$^{\star 2}$	**473**(512)$^{\star 2}$	**440**(476)$^{\star 2}$	**88**(94)$^{\star 2}$	12/15
tany	83(23)	1944(3584)	∞	∞	∞	∞ *2e7*	0/15

Δf_{opt}	1e1	1e0	1e-1	1e-3	1e-5	1e-7	#succ
f23	3.2	1614	67457	4.9e5	8.1e5	8.4e5	15/15
def	1.7(2)	35(39)	5.3(5)	71(66)	**43**(40)	**41**(38)	7/15
texp	**1.3**(1)	28(36)	28(45)	182(190)	110(117)	106(114)	3/15
tany	1.7(1)	**20**(23)	**3.0**(6)	**55**(62)	51(56)	49(55)	6/15

Δf_{opt}	1e1	1e0	1e-1	1e-3	1e-5	1e-7	#succ
f24	1.3e6	7.5e6	5.2e7	5.2e7	5.2e7	5.2e7	3/15
def	∞	∞	∞	∞	∞	∞ *2e7*	0/15
texp	**1.2**(1)*	**0.68**(0.7)$^{\star 4}$	**5.5**(6)$^{\star 4}$	**5.5**(6)$^{\star 4}$	**5.5**(6)$^{\star 4}$	**5.5**(6)$^{\star 4}$	1/15
tany	28(33)	∞	∞	∞	∞	∞ *2e7*	0/15

Table 3: Expected running time (ERT in number of function evaluations) divided by the respective best ERT measured during BBOB-2009 (given in the respective first row) for different Δf values in dimension 20. The central 80% range divided by two is given in braces. 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 $f_{\mathrm{opt}} + 10^{-8}$. Best results are printed in bold.

[12] T. Liao, M. A. Montes de Oca, and T. Stützle. Tuning parameters across mixed dimensional instances: A performance scalability study of Sep-G-CMA-ES. In *Proceedings of the Workshop on Scaling Behaviours of Landscapes, Parameters and Algorithms of the Genetic and Evolutionary Computation Conference, GECCO'11*, pages 703–706, New York, NY, USA, 2011. ACM.

[13] T. Liao, M. A. Montes de Oca, and T. Stützle. Computational results for an automatically tuned CMA-ES with increasing population size on the CEC'05 benchmark set. *Soft Computing - A Fusion of Foundations, Methodologies and Applications*, 2013. In press.

[14] T. Liao and T. Stützle. Bounding the population size of IPOP-CMA-ES on the noiseless BBOB testbed. In *Proceedings of Genetic and Evolutionary Computation Conference*, GECCO Companion'13, 2013. submitted.

[15] M. López-Ibáñez, J. Dubois-Lacoste, T. Stützle, and M. Birattari. The `irace` package, iterated race for automatic algorithm configuration. Technical Report TR/IRIDIA/2011-004, IRIDIA, Université Libre de Bruxelles, Belgium, 2011.

[16] M. López-Ibáñez, T. Liao, and T. Stützle. On the anytime behavior of IPOP-CMA-ES. LNCS 7491, pages 357–366. Springer, Heidelberg, Germany, 2012.

[17] V. Nannen and A. E. Eiben. Relevance estimation and value calibration of evolutionary algorithm parameters. In *Proceedings of the 20th International Joint Conference on Artifical Intelligence*, pages 975–980, San Francisco, CA, USA, 2007. Morgan Kaufmann.

[18] 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.

[19] S. Smit and A. Eiben. Comparing parameter tuning methods for evolutionary algorithms. In *IEEE Congress on Evolutionary Computation (CEC'09)*, pages 399–406, Piscataway, NJ, 2009. IEEE Press.

[20] Z. Yuan, M. Montes de Oca, M. Birattari, and T. Stützle. Modern Continuous Optimization Algorithms for Tuning Real and Integer Algorithm Parameters. In M. Dorigo et al., editors, *Proceedings of the Seventh International Conference on Swarm Intelligence, ANTS 2010*, volume 6234 of *LNCS*, pages 204–215. Springer, Berlin, Germany, 2010.