



HAL
open science

Black-Box Optimization Benchmarking of NEWUOA compared to BIPOP-CMA-ES

Raymond Ros, Nikolaus Hansen

► **To cite this version:**

Raymond Ros, Nikolaus Hansen. Black-Box Optimization Benchmarking of NEWUOA compared to BIPOP-CMA-ES. Genetic and Evolutionary Computation Conference 2010, Jul 2010, Portland, OR, United States. inria-00473779

HAL Id: inria-00473779

<https://inria.hal.science/inria-00473779>

Submitted on 16 Apr 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Black-Box Optimization Benchmarking of NEWUOA compared to BIPOP-CMA-ES

On the BBOB Noiseless Testbed

Nikolaus Hansen
Microsoft Research–INRIA Joint Centre
28 rue Jean Rostand
91893 Orsay Cedex, France
Nikolaus.Hansen@inria.fr

Raymond Ros
TAO Team-Project – INRIA Saclay
LRI, Bat 490, Univ. Paris-Sud
F-91405 Orsay Cedex, France
raymond.ros@inria.fr

ABSTRACT

In this paper, the performances of the NEW Unconstrained Optimization Algorithm (NEWUOA) on some noiseless functions are compared to those of the BI-POPulation Covariance Matrix Adaptation-Evolution Strategy (BIPOP-CMA-ES). The two algorithms were benchmarked on the BBOB 2009 noiseless function testbed. The comparison shows that NEWUOA outperforms BIPOP-CMA-ES on some functions like the Sphere or the Rosenbrock functions. Also the independent restart procedure used for NEWUOA allows it to perform better than BIPOP-CMA-ES on the Gallagher functions. Nevertheless, BIPOP-CMA-ES is faster and has a better success probability than NEWUOA in reaching target function values smaller than one on all other functions.

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, Evolution strategy, Derivative-free optimization

1. INTRODUCTION

In the context of Black-Box Optimization (BBO), there is the community of Derivative-Free Optimization (DFO) which has proposed the NEW Unconstrained Optimization Algorithm (NEWUOA) [8]. NEWUOA is a trust-region method. NEWUOA computes a quadratic interpolation of

the objective function in the current trust region and performs a truncated conjugate gradient minimization of the surrogate model in the trust region.

From the community of evolutionary computation, the Covariance Matrix Adaptation-Evolution Strategy (CMA-ES) is a state-of-the-art stochastic population-based search method. A BI-POPulation (BIPOP) CMA-ES was introduced in [4] and makes use of a multi-start strategy and two populations with different sizes.

Comparisons of NEWUOA and CMA-ES on a small set of essentially unimodal functions were done in [1, 2]. NEWUOA considerably outperforms CMA-ES on well-conditioned problems and on convex problems with moderate condition number. It performs particularly well on separable convex problems. On non-convex (unimodal) problems with a moderate condition number of 10^4 and on non-separable problems with a condition number of 10^6 , the performance of NEWUOA and CMA-ES align. With even larger condition numbers CMA-ES becomes somewhat advantageous.

In this paper, we compare NEWUOA to BIPOP-CMA-ES based on the experimental data obtained for the Black-Box Optimization Benchmarking (BBOB) workshop that was held at the Genetic and Evolutionary Computation Conference 2009.

For more details on the algorithms, their parameter tuning, we refer to [10] for NEWUOA and [4] for BIPOP-CMA-ES.

2. EXPERIMENTAL PROCEDURE

We used the data obtained in [10] for NEWUOA using $2n + 1$ points for interpolating the quadratic model, where n is the dimension of the search space, and the data from [4] for BIPOP-CMA-ES. For benchmarking NEWUOA on the BBOB 2009 noiseless function testbed, an independent multi-start procedure had been implemented as advised in [5]. The BIPOP-CMA-ES includes a restart procedure but adds a population size management policy.

We use the BBOB 2010 post-processing software to compare the performances of NEWUOA and BIPOP-CMA-ES. This can be done without any modifications of the data from BBOB 2009. In any case, the differences in the experimental set-up of BBOB 2009 and BBOB 2010 are minimal. The differences reside in the function instances considered (1 to 5 versus 1 to 15 resp.) and their repetition (3 times versus 1 time resp.).

The parameter settings of NEWUOA and BIPOP-CMA-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

GECCO '10, July 7–11, 2010, Portland, Oregon, USA.

Copyright 2010 ACM 978-1-4503-0073-5/10/07 ...\$10.00.

ES are described in [10] and [4]. For both algorithms, the crafting effort [6] is equal to CrE= 0.

3. CPU TIMING EXPERIMENTS

For the timing experiments, both algorithms were run on f_8 and restarted until at least 30 seconds (according to [6]). The experiments for NEWUOA has been conducted on a Intel Core 2 6700 processor (2.66 GHz) on Linux 2.6.24.7. The results were 8.1 ; 11 ; 21 ; 58 ; 170 ; 620 and 2500×10^{-6} seconds per function evaluations for NEWUOA in dimensions 2 ; 3 ; 5 ; 10 ; 20 ; 40 and 80 respectively. The results show a dependency between the time per function evaluations and the dimension of the search space.

The experiments for BIPOP-CMA-ES has been conducted on a Intel Core 2 6700 processor (2.66 GHz) on Linux 2.6.24.7 using Matlab R2008a. The results were 6.2 ; 5.8 ; 5.6 ; 5.7 ; 5.8 ; 5.9 and 6.3×10^{-4} seconds per function evaluation for BIPOP-CMA-ES in dimensions 2 ; 3 ; 5 ; 10 ; 20 ; 40 and 80 respectively.

4. RESULTS

Results from experiments according to [6] on the benchmark functions given in [3, 7] are presented in Figures 1, 2, 3 and 4 and in Tables 1 and 2. The **expected running time (ERT)**, used in the figures and table, depends on a given target function value, $f_t = f_{\text{opt}} + \Delta f$, 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 [6, 9]. **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.

NEWUOA outperforms BIPOP-CMA-ES on f_1 by a factor of about 50 and on the Linear Slope and the Rosenbrock function by a factor of about three. On the other unimodal functions the picture is comparatively mixed, presumably due to local deformations in the function topographies: besides f_1 , all functions deviate significantly from a quadratic form. The most surprising results can be observed on the multi-modal functions f_{21} and f_{22} , where NEWUOA consistently outperforms the BIPOP-CMA-ES, for larger dimension and the more difficult target values even by a factor between 10 and 100. The applied independent restarts of NEWUOA appear to be more effective than the large population size of BIPOP-CMA-ES, which is in turn more helpful on the remaining multi-modal functions. On the most difficult multi-modal functions, the performance is not comparable, as BIPOP-CMA-ES were allowed to execute more function evaluations than NEWUOA. Overall, NEWUOA considerably outperforms BIPOP-CMA-ES on about seven functions, while BIPOP-CMA-ES considerably outperforms NEWMAN on about eleven functions.

In conclusion, NEWUOA and BIPOP-CMA-ES are two quite complementary algorithms in their performance. On most problems, one of them considerably outperforms the other. This makes both of them good candidates to be used in an ensemble of black-box search algorithms.

5. REFERENCES

- [1] A. Auger, N. Hansen, J. M. Perez Zerpa, R. Ros, and M. Schoenauer. Empirical comparisons of several derivative free optimization algorithms. In *Acte du 9ieme colloque national en calcul des structures*, May 2009.
- [2] A. Auger, N. Hansen, J. M. P. Zerpa, R. Ros, and M. Schoenauer. Experimental comparisons of derivative free optimization algorithms. In J. Vahrenhold, editor, *SEA*, volume 5526 of *Lecture Notes in Computer Science*, pages 3–15. Springer, 2009.
- [3] 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.
- [4] N. Hansen. Benchmarking a BI-population CMA-ES on the BBOB-2009 function testbed. In Rothlauf [11], pages 2389–2396.
- [5] N. Hansen, A. Auger, S. Finck, and R. Ros. Real-parameter black-box optimization benchmarking 2009: Experimental setup. Technical Report RR-6828, INRIA, 2009.
- [6] N. Hansen, A. Auger, S. Finck, and R. Ros. Real-parameter black-box optimization benchmarking 2010: Experimental setup. Technical Report RR-7215, INRIA, 2010.
- [7] 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.
- [8] M. J. D. Powell. The NEWUOA software for unconstrained optimization without derivatives. *Large Scale Nonlinear Optimization*, pages 255–297, 2006.
- [9] 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.
- [10] R. Ros. Benchmarking the NEWUOA on the BBOB-2009 function testbed. In Rothlauf [11], pages 2421–2428.
- [11] F. Rothlauf, editor. *Genetic and Evolutionary Computation Conference, GECCO 2009, Proceedings, Montreal, Québec, Canada, July 8-12, 2009, Companion Material*. ACM, 2009.

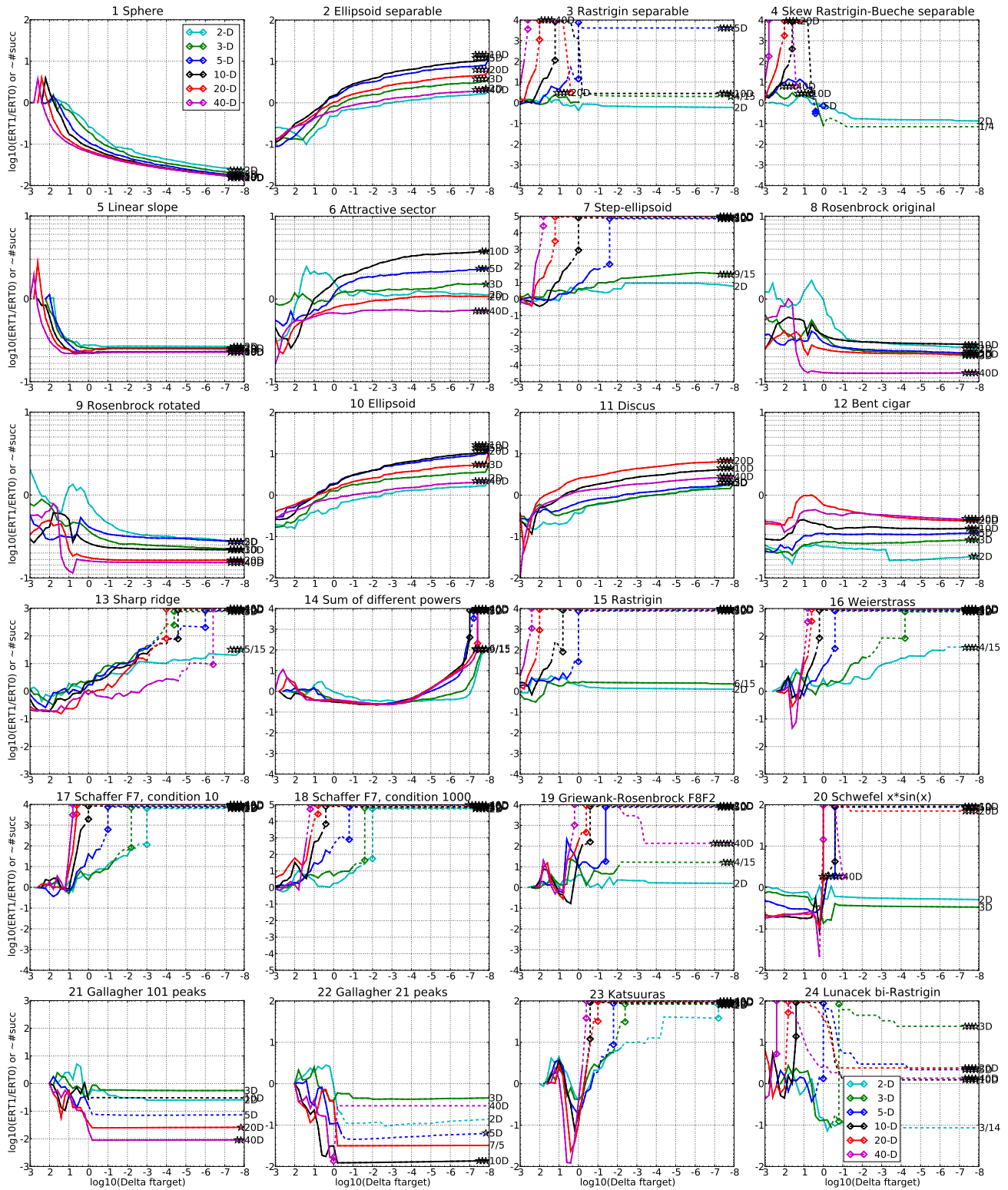


Figure 1: ERT ratio of NEWUOA divided by BIPOP-CMA-ES versus $\log_{10}(\Delta f)$ for f_1 - f_{24} in 2, 3, 5, 10, 20, 40-D. Ratios $< 10^0$ indicate an advantage of NEWUOA, smaller values are always better. The line gets dashed when for any algorithm the ERT exceeds thrice the median of the trial-wise 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 NEWUOA. 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 NEWUOA (1st number) and non-zero for BIPOP-CMA-ES (2nd number). Results are significant with $p = 0.05$ for one star and $p = 10^{-\#\ast}$ otherwise, with Bonferroni correction within each figure.

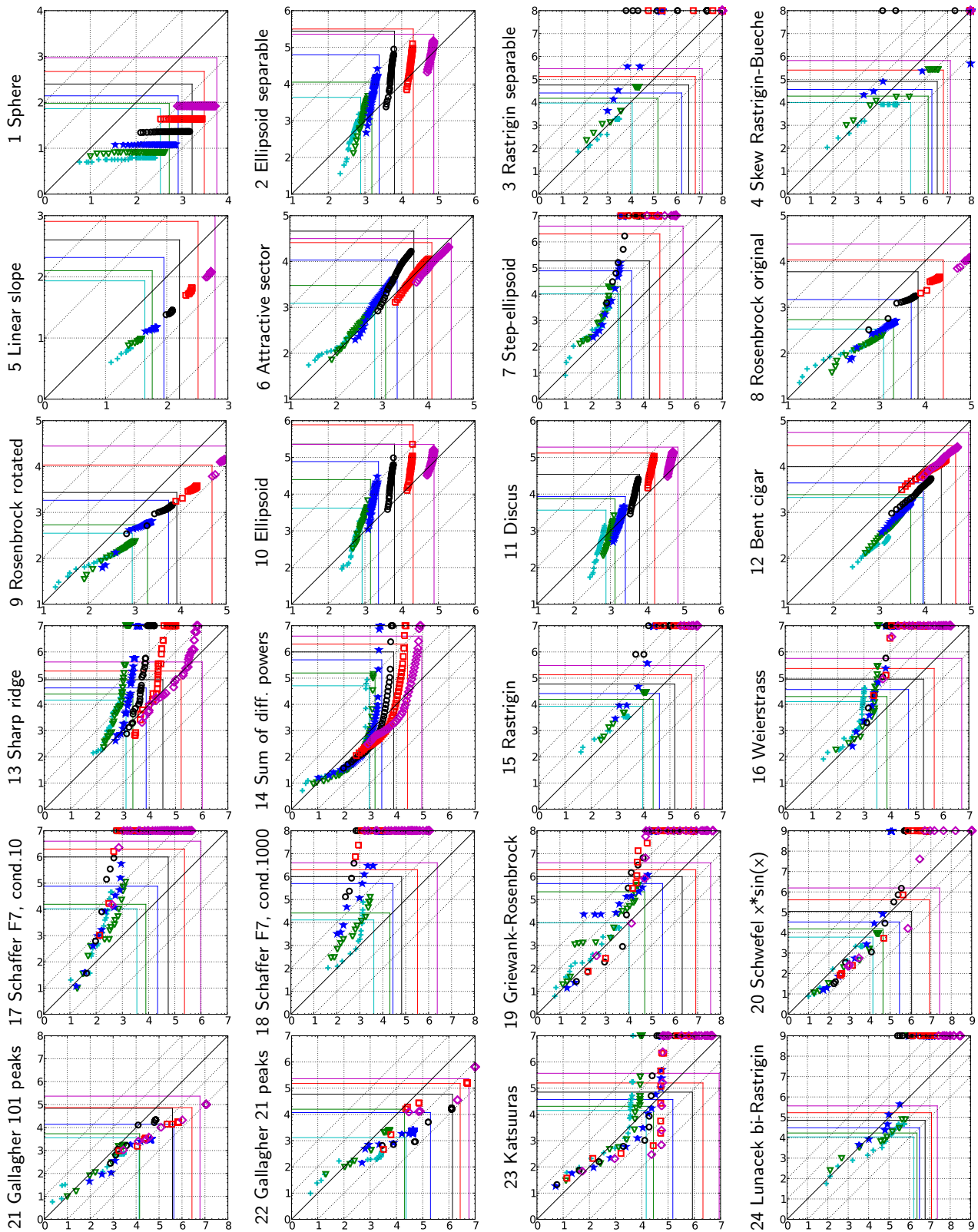


Figure 2: Expected running time (ERT in \log_{10} of number of function evaluations) of NEWUOA versus BIPOP-CMA-ES for 46 target values $\Delta f \in [10^{-8}, 10]$ in each dimension for functions f_1-f_{24} . Markers on the upper or right edge indicate that the target value was never reached by NEWUOA or BIPOP-CMA-ES respectively. Markers represent dimension: 2:+, 3:∇, 5:*, 10:○, 20:□, 40:◇. The colored lines indicate maximum number of function evaluations

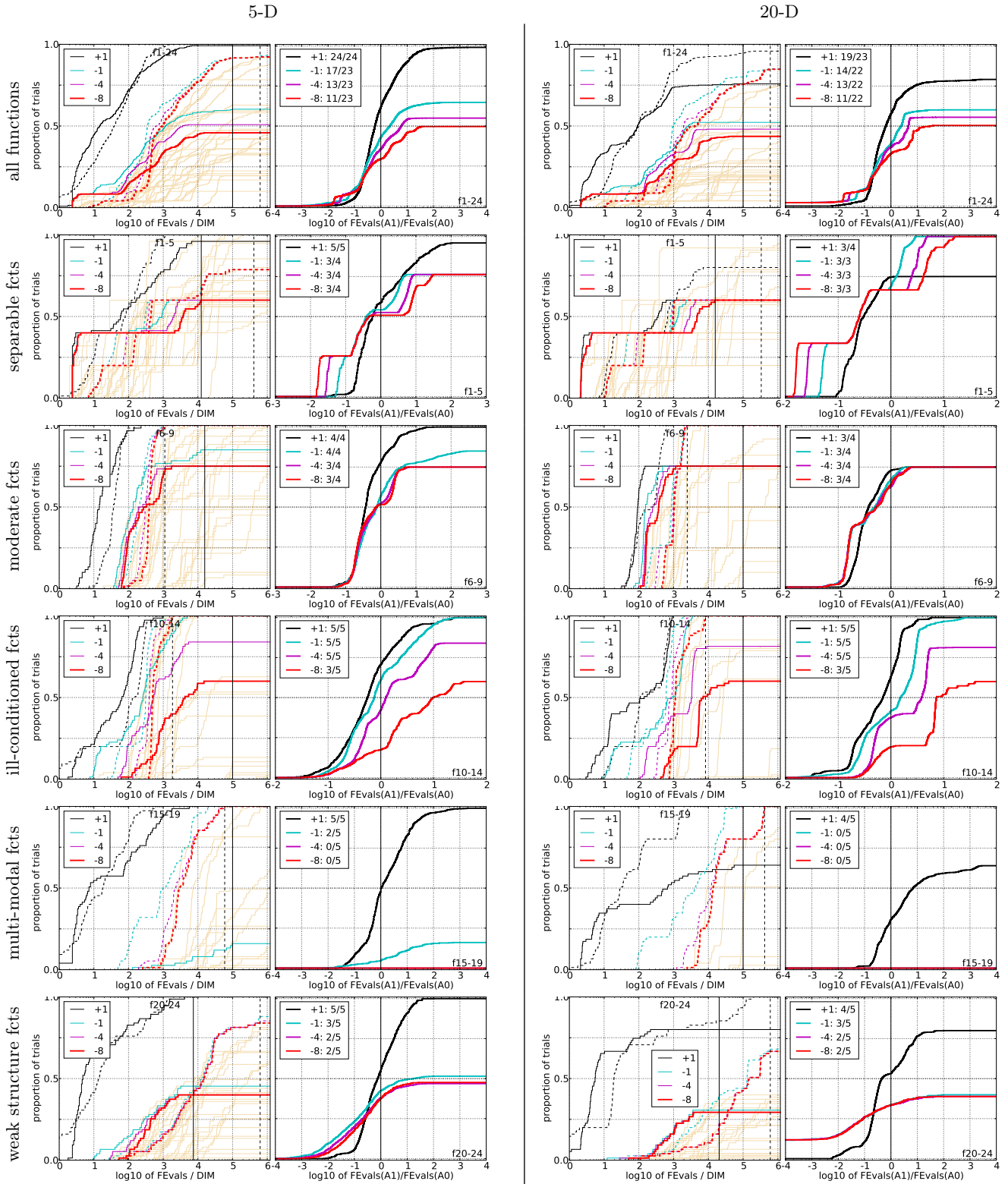


Figure 3: Empirical cumulative distributions (ECDF) of run lengths and speed-up ratios in 5-D (left) and 20-D (right). Left sub-columns: ECDF of the number of function evaluations divided by dimension D (FEvals/ D) to reach a target value $f_{\text{opt}} + \Delta f$ with $\Delta f = 10^k$, where $k \in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

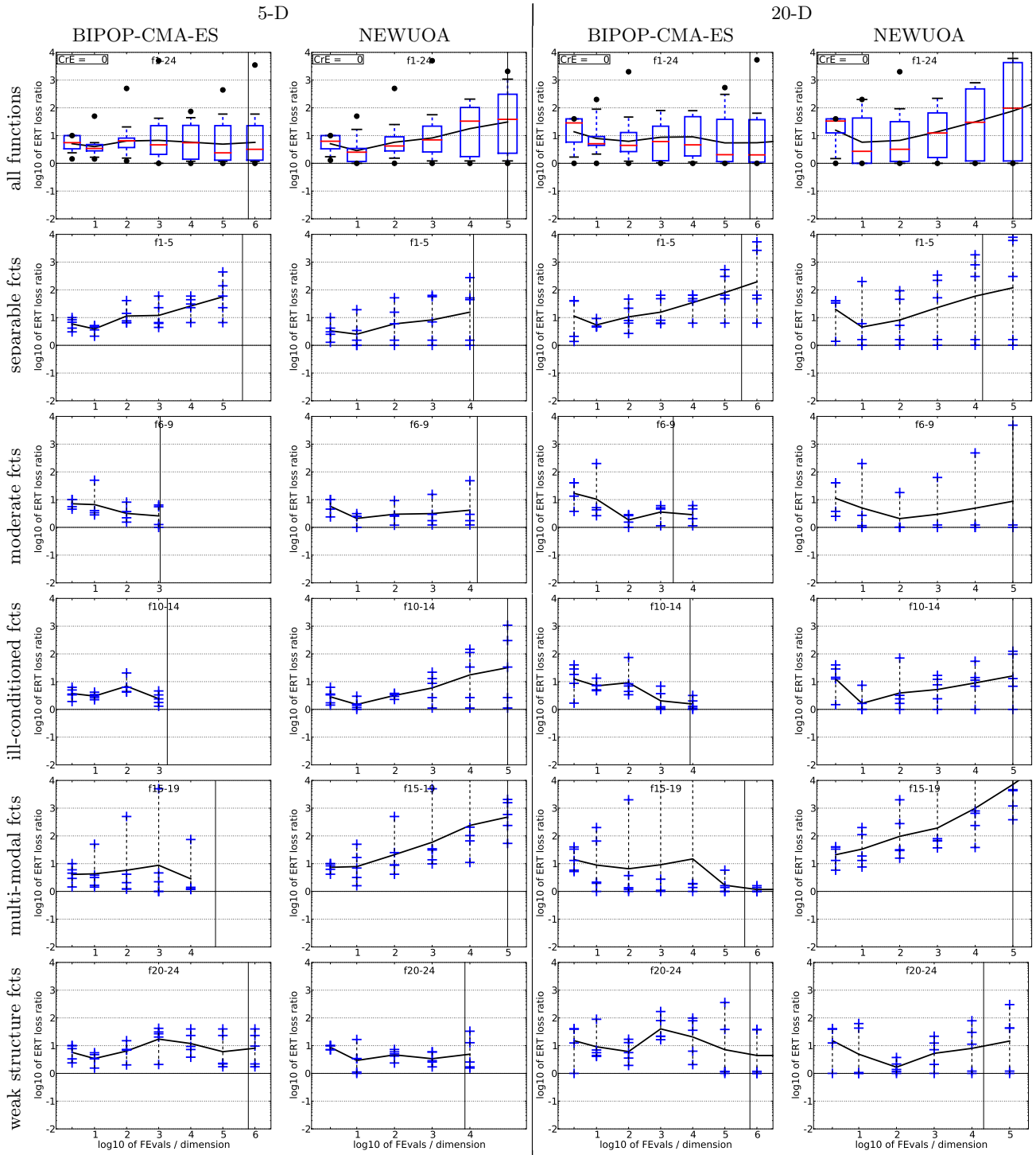


Figure 4: ERT loss ratio versus given budget FEvals. The target value f_t for ERT is the smallest (best) recorded function value such that $ERT(f_t) \leq FEvals$ for the presented algorithm. Shown is FEvals divided by the respective best $ERT(f_t)$ from BBOB-2009 for functions f_1-f_{24} in 5-D and 20-D. Each ERT is multiplied by $\exp(CrE)$ correcting for the parameter crafting effort. 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 this function subset.

Table 2: ERT loss ratio (see Figure 4) compared to the respective best result from BBOB-2009 for budgets given in the first column. The last row 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). ERT Loss ratio is equal to zero if the algorithm considered outperformed all algorithms from BBOB-2009.

BIPOP-CMA-ES							NEWUOA						
#FEs/D	$f1-f24$ in 5-D, maxFE/D=622854						#FEs/D	$f1-f24$ in 5-D, maxFE/D=100000					
	best	10%	25%	med	75%	90%		best	10%	25%	med	75%	90%
2	1.4	2.3	3.3	5.3	9.2	10	2	1.3	1.7	3.2	6.2	10	10
10	1.4	1.6	2.7	3.4	4.6	10	10	1.0	1.0	1.1	2.5	3.4	17
100	1.2	1.5	3.0	6.4	7.9	23	100	1.0	1.5	2.7	4.2	8.9	28
1e3	1.0	1.0	1.9	4.6	22	44	1e3	1.0	1.2	2.2	6.5	19	57
1e4	1.0	1.2	1.4	5.1	23	46	1e4	1.0	1.2	1.7	23	84	2.1e2
1e5	1.0	1.2	1.3	2.3	15	68	1e5	1.0	1.2	2.0	36	3.1e2	1.1e3
1e6	1.0	1.2	1.3	2.8	16	68	RL_{US}/D	4e3	5e3	6e3	7e3	1e4	1e5
RL_{US}/D	3e5	3e5	4e5	4e5	6e5	6e5							

BIPOP-CMA-ES							NEWUOA						
#FEs/D	$f1-f24$ in 20-D, maxFE/D=605134						#FEs/D	$f1-f24$ in 20-D, maxFE/D=100000					
	best	10%	25%	med	75%	90%		best	10%	25%	med	75%	90%
2	1.0	1.7	5.5	23	40	40	2	1.0	1.5	9.2	33	40	40
10	1.0	2.1	4.4	5.0	8.3	1.0e2	10	1.0	1.0	1.0	2.2	31	2.0e2
100	1.0	1.2	2.3	4.1	11	49	100	1.0	1.0	1.1	2.8	31	1.1e2
1e3	1.0	1.0	1.2	6.1	22	89	1e3	1.0	1.0	1.4	12	64	2.3e2
1e4	1.0	1.1	1.6	3.9	44	81	1e4	1.0	1.0	1.2	22	3.9e2	9.0e2
1e5	1.0	1.0	1.1	2.0	22	3.1e2	1e5	1.0	1.0	1.2	71	2.7e3	6.2e3
1e6	1.0	1.0	1.1	1.8	22	3.2e2	1e6	1.0	1.0	1.2	3.7e2	1.7e4	4.6e4
1e7	1.0	1.0	1.1	1.8	22	2.7e3	RL_{US}/D	5e3	6e3	8e3	1e4	7e4	1e5
RL_{US}/D	1e5	1e5	3e5	3e5	3e5	5e5							