

BBO-Benchmarking of Pure Random Search for Noiseless Function Testbed

An example BBOB 2009 Workshop Paper *

Anne Auger[†]
TAO Team - INRIA Saclay
LRI, Bat 490, Univ. Paris-Sud
91405 Orsay Cedex, France
anne.auger@inria.fr

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

The BBOBies

ABSTRACT

As an example, we benchmark the Pure-Random-Search algorithm on the noise-free BBOB 2009 testbed. Each candidate solution is sampled uniformly in $[-5, 5]^D$, where D denotes the search space dimension. The maximum number of function evaluations is chosen as 10^5 times the search space dimension.

Keywords

Benchmarking, Pure Random Search, Monte-Carlo, Black-box optimization, Evolutionary computation

1. INTRODUCTION

The pure random search, first proposed by Brooks in 1958 [1] is the most simple stochastic search algorithm that consists in sampling each search point independently in the search domain and keeping the best solution found.

2. METHODS

We have used a uniform sampling in $[-5, 5]^D$, where D denotes the dimension of the search space. The experiments according to [3] on the benchmark functions given in [2, 4] have been conducted using both a C-code and Matlab code. The algorithm implementation in Matlab is given in Figure 1. A maximum of $10^5 \times D$ function evaluations has been used. The simulations for 2; 5; 10 and 20 D were done with the C-code and took 2 hours and a half. The 40 D experiments were done at the same time using the Matlab code and took 17 hours.

No parameter tuning was done and the crafting effort CrE [3] is computed to zero.

*Camera-ready paper due April 17th.

[†]Dr. Auger insisted his name be first.

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'09, July 8–12, 2009, Montréal Québec, Canada.
Copyright 2009 ACM 978-1-60558-505-5/09/07 ...\$5.00.

3. RESULTS AND DISCUSSION

Results from experiments according to [3] on the benchmark functions given in [2, 4] are presented in Figures 3 and 4 and in Table 1.

Since we use a uniform sampling in the search domain, we obtain as a by-product of the results an estimate of the volume of the sublevel sets: the sublevel sets of a function $f : \mathbb{R}^D \rightarrow \mathbb{R}$ are defined as $S_c = \{x \in \mathbb{R}^D | f(x) \leq c\}$ for c spanning \mathbb{R} . If S_c is a subset of $[-5, 5]^D$, the hitting time T_c (assuming infinite horizon) of the sublevel set S_c is distributed according to a geometric random variable of parameter $p_c = Vol(S_c)/Vol([-5, 5]^D)$. The expected running time $ERT(\Delta f)$ estimates the expected value of $T_{\Delta f}$ (see Figure 2), that equals $1/p_c$ since $T_{\Delta f}$ is a geometric random variable. And thus $ERT(\Delta f)$ gives the ratio between $Vol([-5, 5]^D)$ and $Vol(S_c)$.

4. CPU TIMING EXPERIMENT

For the timing experiment the Pure Random Search was run with a maximum of $10^5 \times D$ function evaluations and restarted until 30 seconds has passed (according to Figure 2 in [3]). The experiments have been conducted with an Intel Core 2 Duo 2.53 GHz under Mac OS X Version 10.5.6 using the C-code provided. The time per function evaluation was 2.0; 2.3; 2.8; 4.2; 6.9 times 10^{-7} seconds in dimensions 2; 3; 5; 10; 20; 40 respectively.

5. CONCLUSION

We have presented the results of the Pure Random Search, a non-adaptive algorithm, that does not use information gathered during search for guiding its next steps. Those results provide a baseline comparison that every adaptive algorithm should outperform.

Acknowledgments

The first two authors would like to acknowledge the great and hard work of the BBOB team with special kudos to Steffen Finck and (of course!) to the natural leader of the team Nikolaus Hansen.

6. REFERENCES

- [1] S. H. Brooks. A discussion of random methods for seeking maxima. *Operations Research*, 6:244–251, 1958.

Δf	f1 in 5-D, N=15, mFE=500000					f1 in 20-D, N=15, mFE=2000000					f2 in 5-D, N=15, mFE=500000					f2 in 20-D, N=15, mFE=2000000						
	#	ERT	10%	90%	RT _s	#	ERT	10%	90%	RT _s	#	ERT	10%	90%	RT _s	#	ERT	10%	90%	RT _s		
10	15	8.7e1	6.2e1	1.0e2	8.7e1	0	<i>33e+0</i>	<i>29e+0</i>	<i>40e+0</i>	1.1e6	10	0	<i>26e+1</i>	<i>62e+0</i>	<i>55e+1</i>	2.5e5	0	<i>21e+4</i>	<i>16e+4</i>	<i>28e+4</i>	1.0e6	
1	15	2.5e4	1.8e4	3.3e4	2.5e4	1	
1e-1	2	3.5e6	3.3e6	3.7e6	4.2e5	1e-1	
1e-3	0	<i>19e-2</i>	<i>86e-3</i>	<i>42e-2</i>	1.8e5	1e-3	
1e-5	1e-5
1e-8	1e-8

Δf	f3 in 5-D, N=15, mFE=500000					f3 in 20-D, N=15, mFE=2000000					f4 in 5-D, N=15, mFE=500000					f4 in 20-D, N=15, mFE=2000000					
	#	ERT	10%	90%	RT _s	#	ERT	10%	90%	RT _s	#	ERT	10%	90%	RT _s	#	ERT	10%	90%	RT _s	
10	4	1.6e6	1.4e6	1.8e6	4.4e5	0	<i>30e+1</i>	<i>26e+1</i>	<i>34e+1</i>	1.3e6	10	1	7.4e6	7.4e6	7.5e6	5.0e5	0	<i>39e+1</i>	<i>34e+1</i>	<i>43e+1</i>	1.0e6
1	0	<i>13e+0</i>	<i>63e-1</i>	<i>16e+0</i>	2.8e5	1	0	<i>17e+0</i>	<i>10e+0</i>	<i>21e+0</i>	4.0e5

Δf	f5 in 5-D, N=15, mFE=500000					f5 in 20-D, N=15, mFE=2000000					f6 in 5-D, N=15, mFE=500000					f6 in 20-D, N=15, mFE=2000000					
	#	ERT	10%	90%	RT _s	#	ERT	10%	90%	RT _s	#	ERT	10%	90%	RT _s	#	ERT	10%	90%	RT _s	
10	15	4.9e4	3.5e4	6.1e4	4.9e4	0	<i>13e+1</i>	<i>13e+1</i>	<i>14e+1</i>	7.9e5	10	15	2.5e4	1.8e4	3.4e4	2.5e4	0	<i>69e+2</i>	<i>37e+1</i>	<i>93e+3</i>	7.1e5
1	0	<i>54e-1</i>	<i>42e-1</i>	<i>71e-1</i>	3.2e5	1	1	7.3e6	7.0e6	7.5e6	5.0e5

Δf	f7 in 5-D, N=15, mFE=500000					f7 in 20-D, N=15, mFE=2000000					f8 in 5-D, N=15, mFE=500000					f8 in 20-D, N=15, mFE=2000000					
	#	ERT	10%	90%	RT _s	#	ERT	10%	90%	RT _s	#	ERT	10%	90%	RT _s	#	ERT	10%	90%	RT _s	
10	15	7.9e2	3.4e2	1.1e3	7.9e2	0	<i>14e+1</i>	<i>11e+1</i>	<i>18e+1</i>	7.1e5	10	3	2.2e6	2.0e6	2.4e6	5.0e5	0	<i>14e+3</i>	<i>91e+2</i>	<i>17e+3</i>	5.0e5
1	9	5.0e5	3.8e5	5.7e5	3.0e5	1	0	<i>16e+0</i>	<i>86e-1</i>	<i>26e+0</i>	1.8e5

Table 1: Shown are, for a given target difference to the optimal function value Δf : the number of successful trials (#); the expected running time to surpass $f_{\text{opt}} + \Delta f$ (ERT, see Figure 3); the 10%-tile and 90%-tile of the bootstrap distribution of ERT; the total number of function evaluations in unsuccessful trials divided either by the number of successful trials or by 1, if none was successful (RT_{US}). If $f_{\text{opt}} + \Delta f$ was never reached, figures in *italics* denote the best achieved Δf -value of the median trial and the 10% and 90%-tile trial. Furthermore, N denotes the number of trials, and mFE denotes the maximum of number of function evaluations executed in one trial. See Figure 3 for the names of functions.

Figure 1: Pure Random Search in Matlab. At each iteration (*iter*), 200 points are sampled and stored in a matrix of size $D \times 200$ so as to reduce loops and function calls within Matlab.

```

function MY_OPTIMIZER(FUN, DIM, ftarget, maxfunevals)
% MY_OPTIMIZER(FUN, DIM, ftarget, maxfunevals)
% samples new points uniformly randomly in [-5,5]^DIM
% and evaluates them on FUN until ftarget of maxfunevals
% is reached, or until 1e8 * DIM fevals are conducted.
% Relies on FUN to keep track of the best point.

maxfunevals = min(1e8 * DIM, maxfunevals);
popsize = min(maxfunevals, 200);
for iter = 1:ceil(maxfunevals/popsize)
    feval(FUN, 10 * rand(DIM, popsize) - 5);
    if feval(FUN, 'fbest') < ftarget % task achieved
        break;
    end
    % if useful, modify more options here for next start
end

```

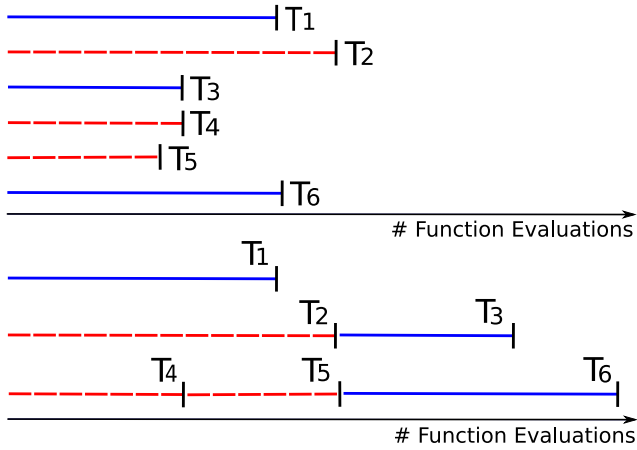


Figure 2: Illustration that $ERT(\Delta f)$ estimates the expected hitting time of an algorithm restarted until success (assuming infinite horizon): among 6 runs of the same algorithm A, the 1st, 3rd and 6th are successful while the 2nd, 4th and 5th are unsuccessful and thus T_1 , $T_2 + T_3$ and $T_4 + T_5 + T_6$ are 3 instances of the algorithm restart-A (i.e., algorithm A restarted until success). Thus an estimate of the expected hitting time of restart-A is $(T_1 + (T_2 + T_3 + T_4) + (T_4 + T_5 + T_6))/3$, i.e., total number of function evaluations divided by number of successes of algorithm A, i.e., $ERT(\Delta f)$. In the case where algorithm A is the pure random search, the picture is simpler because unsuccessful runs always reach the maximum number of evaluations and thus the 2nd, 4th and 5th runs have the same length. T_1 , $T_2 + T_3$ and $T_4 + T_5 + T_6$ represent then 3 instances of the pure random search that would be run with infinite horizon until a success is reached and $ERT(\Delta f)$ estimates thus the expected hitting time of the pure random search with infinite horizon.

- [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.
- [3] N. Hansen, A. Auger, S. Finck, and R. Ros. Real-parameter black-box optimization benchmarking 2009: Experimental setup. Technical Report RR-6828, INRIA, 2009.
- [4] 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.

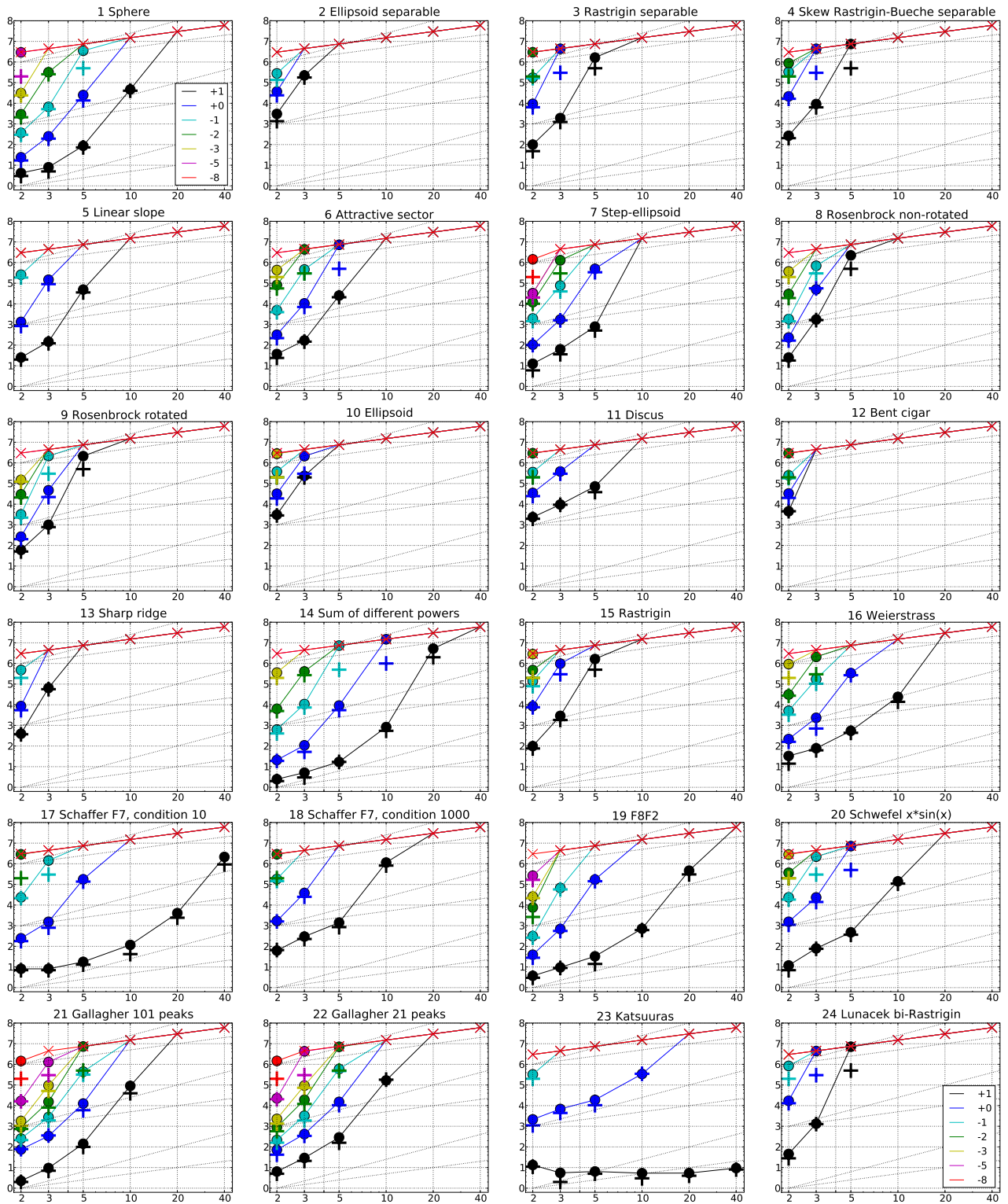


Figure 3: Expected Running Time (ERT, ●) and number of function evaluations of the median trial (+) to reach $f_{\text{opt}} + \Delta f$, shown for $\Delta f = 10, 1, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-5}, 10^{-8}$ (the exponent is given in the legend of f_1 and f_{24}) versus dimension in log-log presentation. The ERT(Δf) equals to $\#FEs(\Delta f)$ divided by the number of successful trials, where a trial is successful if $f_{\text{opt}} + \Delta f$ was surpassed during the trial. The $\#FEs(\Delta f)$ are the total number of function evaluations while $f_{\text{opt}} + \Delta f$ was not surpassed during the trial from all respective trials (successful and unsuccessful), and f_{opt} denotes the optimal function value. Crosses (×) indicate the total number of function evaluations $\#FEs(-\infty)$. Annotated numbers on the ordinate are decimal logarithms. Additional grid lines show linear and quadratic scaling.

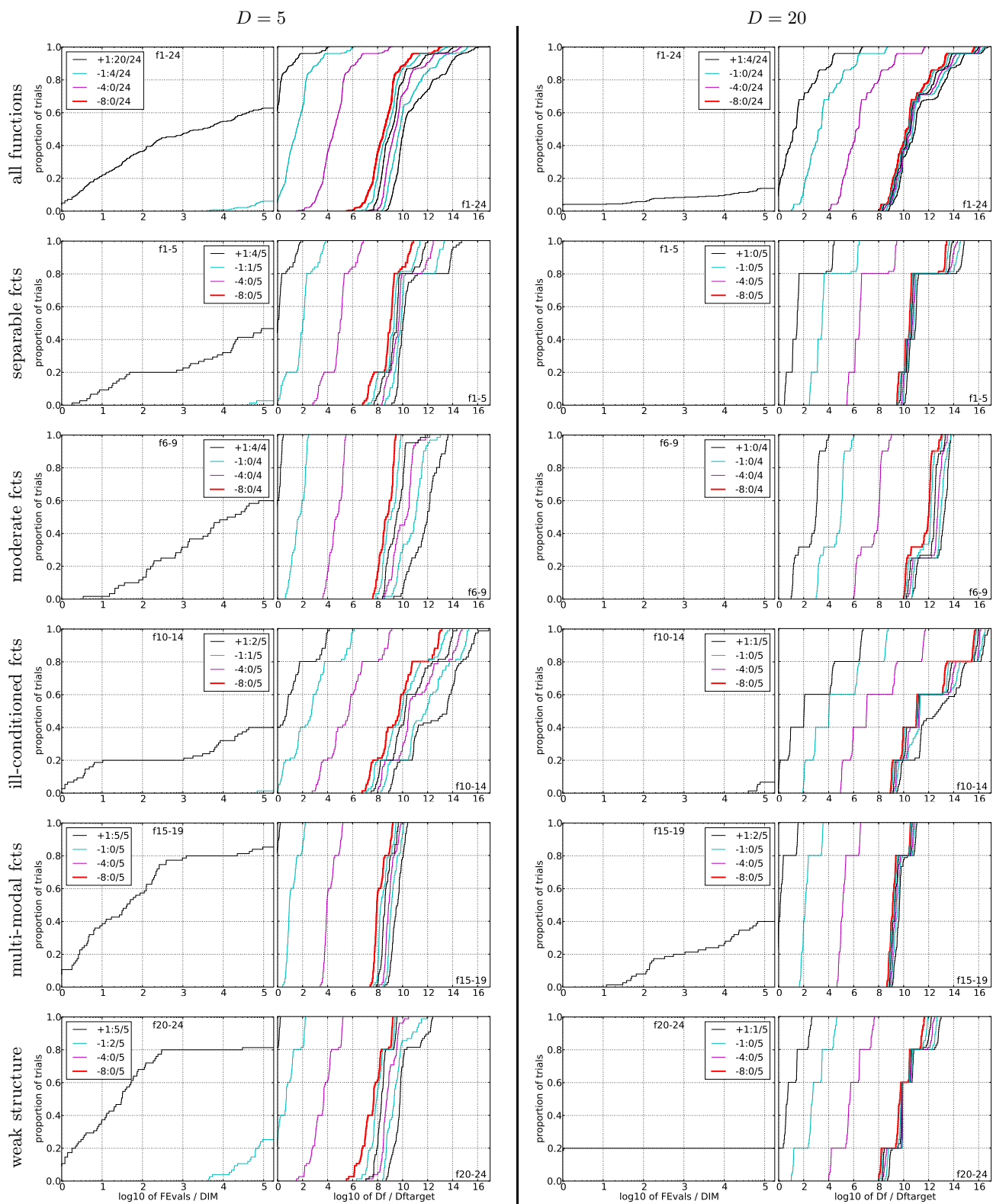


Figure 4: Empirical cumulative distribution functions (ECDFs), plotting the fraction of trials versus running time (left) or Δf . Left subplots: ECDF of the running time (number of function evaluations), divided by search space dimension D , to fall below $f_{\text{opt}} + \Delta f$ with $\Delta f = 10^k$, where k is the first value in the legend. Right subplots: ECDF of the best achieved Δf divided by 10^k (upper left lines in continuation of the left subplot), and best achieved Δf divided by 10^{-8} for running times of $D, 10D, 100D \dots$ function evaluations (from right to left cycling black-cyan-magenta). Top row: all results from all functions; second row: separable functions; third row: misc. moderate functions; fourth row: ill-conditioned functions; fifth row: multi-modal functions with adequate structure; last row: multi-modal functions with weak structure. The legends indicate the number of functions that were solved in at least one trial. FEvals denotes number of function evaluations, D and DIM denote search space dimension, and Δf and Df denote the difference to the optimal function value.