

Benchmarking sep-CMA-ES on the BBOB-2009 Function Testbed

Raymond Ros

► **To cite this version:**

Raymond Ros. Benchmarking sep-CMA-ES on the BBOB-2009 Function Testbed. GECCO, Jul 2009, Montréal, Canada. inria-00377087

HAL Id: inria-00377087

<https://hal.inria.fr/inria-00377087>

Submitted on 20 Apr 2009

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.

Benchmarking sep-CMA-ES on the BBOB-2009 Function Testbed

Raymond Ros
Univ. Paris-Sud, LRI
UMR 8623 / INRIA Saclay, projet TAO
F-91405 Orsay, France
raymond.ros@lri.fr

ABSTRACT

A partly time and space linear CMA-ES is benchmarked on the BBOB-2009 noiseless function testbed. This algorithm with a multistart strategy with increasing population size solves 17 functions out of 24 in 20-D.

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, Evolutionary computation, Covariance matrix adaptation, Evolution strategy

1. INTRODUCTION

The sep-CMA-ES algorithm introduced in [7] is a variant of the covariance matrix adaptation evolution strategy (CMA-ES) [5] that is linear in time and space. This property combined with a faster learning rate makes sep-CMA-ES appropriate for separable function and larger dimensions. A mixed strategy of using sep-CMA-ES and CMA-ES is proposed here and benchmarked on a noiseless function testbed.

2. ALGORITHM PRESENTATION

In its design, the sep-CMA-ES differs from the CMA-ES by two aspects: first, the covariance matrix is constrained to be diagonal at each of its update, second, the learning rate is increased by a factor of $\frac{n+3/2}{3}$, where n is the dimension of the search space¹. These modifications result

¹Please note that the factor for the learning rate is smaller than the one in [7].

in an algorithm that trades model complexity with a time and space scaling that is better than the original CMA-ES. The $(\mu/\mu_w, \lambda)$ -sep-CMA-ES has been shown to outperform $(\mu/\mu_w, \lambda)$ -CMA-ES on separable functions.

We propose here what would be the best of two worlds: to use sep-CMA-ES for the first few iterations and then switch to CMA-ES. At the time of the switch, all parameters are retained except for the learning rate that is decreased back to its default value. This implies the diagonal covariance matrix acquired using sep-CMA-ES is directly used by CMA-ES. This mixed strategy is therefore expected to be faster than CMA-ES on separable functions. Ongoing work has also shown that for some test functions the first iterations using sep-CMA-ES would not disadvantage the latter use of CMA-ES in any way. In other terms, the cost of initially using sep-CMA-ES would not induce a penalty in the cost of solving the function with CMA-ES afterwards. The author admits some functions could induce such a penalty.

As for the multistart strategy, we use the increasing population size IPOPOP-CMA-ES [1]. Though this approach has shown its limits [6], independent restart may improve the probability of the algorithm reaching a given target function value.

3. EXPERIMENTAL PROCEDURE

The Matlab implementation of the CMA-ES (version 3.23 beta) is used². We use the $(\mu/\mu_w, \lambda)$ -IPOPOP-CMA-ES variant with an initial default population size $\lambda = 4 + \lfloor 3 \ln(n) \rfloor$ increasing twice at each restart. Except the learning rate, all other algorithm parameters are set to their default values. The covariance matrix is constrained to be diagonal only for the first $1 + 100n/\sqrt{\lambda}$ iterations of the *first start*. A maximum of 8 independent restarts is conducted. Restarts occur after $100 + 300n\sqrt{n/\lambda}$ iterations or if any of the default stopping criterion is met. The initial stepsize has been set to 2 and the starting point has been chosen uniformly in $[-4, 4]^n$. The maximum number of function evaluations was set to 10^4 times the dimension. No parameter tuning was done, the CrE [3] is computed to zero.

4. RESULTS AND DISCUSSION

Results from experiments according to [3] on the benchmark functions given in [2, 4] are presented in Figures 1 and 2 and in Table 1. The algorithm solves 17 out of the 24 functions in 20-D. The algorithm performs well on uni-

²Latest version available here: <http://www.lri.fr/~hansen/cmaesintro.html>

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.

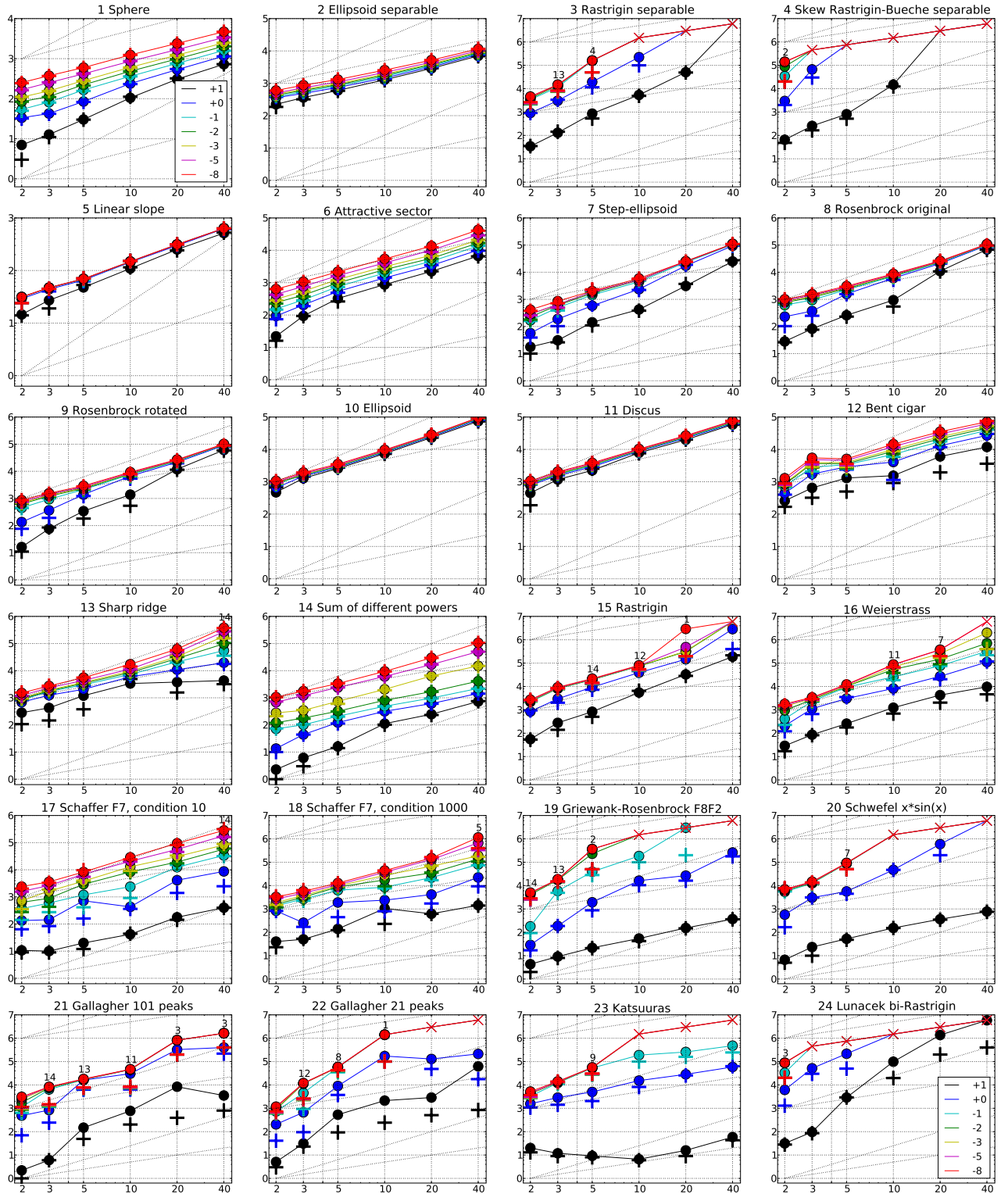


Figure 1: Expected Running Time (ERT, ●) to reach $f_{\text{opt}} + \Delta f$ and median number of function evaluations of successful trials (+), shown for $\Delta f = 10, 1, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-5}, 10^{-8}$ (the exponent is given in the legend of f_1 and f_{24}) versus dimension in log-log 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 $\#FEs(-\infty)$. Numbers above ERT-symbols indicate the number of successful trials. Annotated numbers on the ordinate are decimal logarithms. Additional grid lines show linear and quadratic scaling.

f_1 in 5-D, N=15, mFE=738					f_1 in 20-D, N=15, mFE=2594					f_2 in 5-D, N=15, mFE=1554					f_2 in 20-D, N=15, mFE=5750						
Δf	#	ERT	10%	90%	RT _{succ}	#	ERT	10%	90%	RT _{succ}	#	ERT	10%	90%	RT _{succ}	#	ERT	10%	90%	RT _{succ}	
10	15	3.0e1	2.3e1	3.7e1	3.0e1	15	3.1e2	2.9e2	3.2e2	3.1e2	10	6.0e2	5.5e2	6.5e2	6.0e2	15	2.9e3	2.8e3	3.0e3	2.9e3	
1	15	8.5e1	7.8e1	9.2e1	8.5e1	15	5.3e2	5.1e2	5.5e2	5.3e2	1	15	7.4e2	7.0e2	7.8e2	7.4e2	15	3.2e3	3.1e3	3.3e3	3.2e3
1e-1	15	1.7e2	1.5e2	1.8e2	1.7e2	15	7.6e2	7.4e2	7.8e2	7.6e2	1e-1	15	8.3e2	7.9e2	8.6e2	8.3e2	15	3.5e3	3.4e3	3.6e3	3.5e3
1e-3	15	2.8e2	2.7e2	3.0e2	2.8e2	15	1.2e3	1.2e3	1.3e3	1.2e3	1e-3	15	9.7e2	9.4e2	1.0e3	9.7e2	15	4.0e3	3.9e3	4.1e3	4.0e3
1e-5	15	4.2e2	4.0e2	4.3e2	4.2e2	15	1.7e3	1.6e3	1.7e3	1.7e3	1e-5	15	1.1e3	1.1e3	1.1e3	1.1e3	15	4.5e3	4.4e3	4.6e3	4.5e3
1e-8	15	6.0e2	5.8e2	6.2e2	6.0e2	15	2.4e3	2.4e3	2.4e3	2.4e3	1e-8	15	1.3e3	1.3e3	1.3e3	1.3e3	15	5.2e3	5.1e3	5.3e3	5.2e3
f_3 in 5-D, N=15, mFE=50422					f_3 in 20-D, N=15, mFE=200352					f_4 in 5-D, N=15, mFE=50430					f_4 in 20-D, N=15, mFE=200376						
Δf	#	ERT	10%	90%	RT _{succ}	#	ERT	10%	90%	RT _{succ}	#	ERT	10%	90%	RT _{succ}	#	ERT	10%	90%	RT _{succ}	
10	15	8.6e2	5.9e2	1.1e3	8.6e2	15	5.1e4	4.2e4	6.0e4	5.1e4	10	15	8.1e2	5.7e2	1.1e3	8.1e2	0	14e+0	13e+0	16e+0	1.1e5
1	13	1.9e4	1.3e4	2.7e4	1.5e4	0	60e-1	40e-1	70e-1	1.1e5	1	0	37e-1	30e-1	40e-1	1.8e4					
1e-1	4	1.6e5	9.6e4	3.4e5	3.7e4						1e-1										
1e-3	4	1.6e5	9.8e4	3.5e5	3.8e4						1e-3										
1e-5	4	1.6e5	9.9e4	3.5e5	3.8e4						1e-5										
1e-8	4	1.6e5	1.0e5	3.4e5	3.9e4						1e-8										
f_5 in 5-D, N=15, mFE=106					f_5 in 20-D, N=15, mFE=410					f_6 in 5-D, N=15, mFE=2650					f_6 in 20-D, N=15, mFE=24422						
Δf	#	ERT	10%	90%	RT _{succ}	#	ERT	10%	90%	RT _{succ}	#	ERT	10%	90%	RT _{succ}	#	ERT	10%	90%	RT _{succ}	
10	15	4.8e1	4.2e1	5.4e1	4.8e1	15	2.5e2	2.3e2	2.7e2	2.5e2	10	15	3.2e2	2.6e2	3.9e2	3.2e2	15	2.2e3	2.1e3	2.3e3	2.2e3
1	15	6.6e1	6.0e1	7.2e1	6.6e1	15	3.0e2	2.8e2	3.2e2	3.0e2	1	15	5.3e2	4.7e2	6.1e2	5.3e2	15	3.4e3	3.2e3	3.5e3	3.4e3
1e-1	15	6.8e1	6.2e1	7.4e1	6.8e1	15	3.1e2	2.9e2	3.3e2	3.1e2	1e-1	15	7.5e2	6.8e2	8.3e2	7.5e2	15	4.6e3	4.4e3	4.7e3	4.6e3
1e-3	15	6.9e1	6.4e1	7.5e1	6.9e1	15	3.1e2	2.9e2	3.3e2	3.1e2	1e-3	15	1.3e3	1.2e3	1.3e3	1.3e3	15	7.0e3	6.8e3	7.2e3	7.0e3
1e-5	15	6.9e1	6.4e1	7.5e1	6.9e1	15	3.1e2	2.9e2	3.3e2	3.1e2	1e-5	15	1.7e3	1.6e3	1.7e3	1.7e3	15	1.0e4	9.3e3	1.1e4	1.0e4
1e-8	15	6.9e1	6.4e1	7.5e1	6.9e1	15	3.1e2	2.9e2	3.3e2	3.1e2	1e-8	15	2.2e3	2.1e3	2.2e3	2.2e3	15	1.4e4	1.3e4	1.5e4	1.4e4
f_7 in 5-D, N=15, mFE=3288					f_7 in 20-D, N=15, mFE=46966					f_8 in 5-D, N=15, mFE=5788					f_8 in 20-D, N=15, mFE=28922						
Δf	#	ERT	10%	90%	RT _{succ}	#	ERT	10%	90%	RT _{succ}	#	ERT	10%	90%	RT _{succ}	#	ERT	10%	90%	RT _{succ}	
10	15	1.4e2	1.2e2	1.7e2	1.4e2	15	3.1e3	2.6e3	3.6e3	3.1e3	10	15	2.6e2	2.1e2	3.1e2	2.6e2	15	1.1e4	1.1e4	1.1e4	1.1e4
1	15	5.8e2	5.0e2	6.6e2	5.8e2	15	1.7e4	1.5e4	2.0e4	1.7e4	1	15	1.6e3	1.3e3	1.9e3	1.6e3	15	2.1e4	2.1e4	2.1e4	2.1e4
1e-1	15	1.4e3	1.2e3	1.7e3	1.4e3	15	2.3e4	2.0e4	2.6e4	2.3e4	1e-1	15	2.3e3	2.1e3	2.5e3	2.3e3	15	2.3e4	2.2e4	2.3e4	2.3e4
1e-3	15	1.9e3	1.6e3	2.1e3	1.9e3	15	2.5e4	2.2e4	2.8e4	2.5e4	1e-3	15	2.7e3	2.5e3	2.9e3	2.7e3	15	2.4e4	2.4e4	2.5e4	2.4e4
1e-5	15	1.9e3	1.6e3	2.1e3	1.9e3	15	2.5e4	2.2e4	2.8e4	2.5e4	1e-5	15	2.8e3	2.6e3	3.1e3	2.8e3	15	2.5e4	2.5e4	2.6e4	2.5e4
1e-8	15	2.0e3	1.8e3	2.3e3	2.0e3	15	2.5e4	2.2e4	2.9e4	2.5e4	1e-8	15	3.1e3	2.9e3	3.3e3	3.1e3	15	2.6e4	2.6e4	2.7e4	2.6e4
f_9 in 5-D, N=15, mFE=4914					f_9 in 20-D, N=15, mFE=55924					f_{10} in 5-D, N=15, mFE=4042					f_{10} in 20-D, N=15, mFE=30386						
Δf	#	ERT	10%	90%	RT _{succ}	#	ERT	10%	90%	RT _{succ}	#	ERT	10%	90%	RT _{succ}	#	ERT	10%	90%	RT _{succ}	
10	15	3.4e2	1.9e2	4.9e2	3.4e2	15	1.2e4	1.1e4	1.2e4	1.2e4	10	15	2.6e3	2.5e3	2.7e3	2.6e3	15	2.3e4	2.2e4	2.4e4	2.3e4
1	15	1.4e3	1.2e3	1.6e3	1.4e3	15	2.2e4	1.9e4	2.4e4	2.2e4	1	15	2.8e3	2.7e3	2.9e3	2.8e3	15	2.5e4	2.4e4	2.6e4	2.5e4
1e-1	15	2.1e3	2.0e3	2.3e3	2.1e3	15	2.4e4	2.1e4	2.6e4	2.4e4	1e-1	15	2.9e3	2.8e3	3.0e3	2.9e3	15	2.6e4	2.6e4	2.7e4	2.6e4
1e-3	15	2.5e3	2.3e3	2.7e3	2.5e3	15	2.5e4	2.3e4	2.7e4	2.5e4	1e-3	15	3.1e3	3.0e3	3.2e3	3.1e3	15	2.7e4	2.7e4	2.8e4	2.7e4
1e-5	15	2.7e3	2.5e3	2.8e3	2.7e3	15	2.6e4	2.4e4	2.8e4	2.6e4	1e-5	15	3.3e3	3.2e3	3.3e3	3.3e3	15	2.8e4	2.7e4	2.8e4	2.8e4
1e-8	15	2.9e3	2.8e3	3.1e3	2.9e3	15	2.7e4	2.5e4	2.9e4	2.7e4	1e-8	15	3.5e3	3.4e3	3.6e3	3.5e3	15	2.9e4	2.8e4	2.9e4	2.9e4
f_{11} in 5-D, N=15, mFE=4226					f_{11} in 20-D, N=15, mFE=31418					f_{12} in 5-D, N=15, mFE=9042					f_{12} in 20-D, N=15, mFE=58382						
Δf	#	ERT	10%	90%	RT _{succ}	#	ERT	10%	90%	RT _{succ}	#	ERT	10%	90%	RT _{succ}	#	ERT	10%	90%	RT _{succ}	
10	15	2.2e3	1.9e3	2.5e3	2.2e3	15	2.0e4	1.9e4	2.0e4	2.0e4	10	15	1.3e3	9.4e2	1.7e3	1.3e3	15	6.1e3	4.0e3	8.2e3	6.1e3
1	15	2.9e3	2.8e3	3.0e3	2.9e3	15	2.2e4	2.2e4	2.3e4	2.2e4	1	15	2.9e3	2.6e3	3.2e3	2.9e3	15	1.3e4	9.9e3	1.6e4	1.3e4
1e-1	15	3.2e3	3.1e3	3.3e3	3.2e3	15	2.3e4	2.3e4	2.4e4	2.3e4	1e-1	15	3.4e3	2.9e3	3.8e3	3.4e3	15	1.8e4	1.4e4	2.1e4	1.8e4
1e-3	15	3.4e3	3.4e3	3.5e3	3.4e3	15	2.5e4	2.4e4	2.5e4	2.5e4	1e-3	15	3.9e3	3.4e3	4.6e3	3.9e3	15	2.5e4	2.4e4	2.8e4	2.5e4
1e-5	15	3.6e3	3.5e3	3.7e3	3.6e3	15	2.6e4	2.5e4	2.6e4	2.6e4	1e-5	15	4.5e3	3.8e3	5.2e3	4.5e3	15	3.0e4	2.7e4	3.3e4	3.0e4
1e-8	15	3.8e3	3.8e3	3.9e3	3.8e3	15	2.7e4	2.6e4	2.7e4	2.7e4	1e-8	15	5.1e3	4.3e3	5.9e3	5.1e3	15	3.5e4	3.2e4	3.8e4	3.5e4
f_{13} in 5-D, N=15, mFE=6308					f_{13} in 20-D, N=15, mFE=105488					f_{14} in 5-D, N=15, mFE=3634					f_{14} in 20-D, N=15, mFE=31850						
Δf	#	ERT	10%	90%	RT _{succ}	#	ERT	10%	90%	RT _{succ}	#	ERT	10%	90%	RT _{succ}	#	ERT	10%	90%	RT _{succ}	
10	15	1.2e3	8.5e2	1.6e3	1.2e3	15	3.8e3	2.6e3	5.0e3	3.8e3	10	15	1.6e1	1.2e1	2.0e1	1.6e1	15	2.5e2	2.3e2	2.7e2	2.5e2
1	15	2.1e3	1.7e3	2.4e3	2.1e3	15	1.1e4	8.4e3	1.3e4	1.1e4	1	15	1.2e2	1.1e2	1.4e2	1.2e2	15	6.1e2	5.8e2	6.5e2	6.1e2
1e-1	15	3.0e3	2.8e3	3.2e3	3.0e3	15	2.1e4	1.7e4	2.5e4	2.1e4	1e-1	15	2.1e2	1.9e2	2.3e2	2.1e2	15	9.8e2	9.4e2	1.0e3	9.8e2
1e-3	15	3.7e3	3.6e3	3.8e3	3.7e3	15	3.3e4	2.9e4	3.6e4	3.3e4	1e-3	15	7.4e2	6.3e2	8.5e2	7.4e2	15	6.4e3	6.1e3	6.7e3	6.4e3
1e-5	15	4.4e3	4.3e3	4.5e3	4.4e3	15	4.7e4	4.4e4	5.0e4	4.7e4	1e-5	15	2.4e3	2.3e3	2.5e3	2.4e3	15	1.7e4	1.6e4	1.7e4	1.7e4
1e-8	15	5.5e3	5.4e3	5.7e3	5.5e3	15	6.2e4	5.7e4	6.8e4	6.2e4	1e-8	15	3.3e3	3.3e3	3.4e3	3.3e3	15	3.0e4	2.9e4	3.0e4	3.0e4
f_{15} in 5-D, N=15, mFE=50430					f_{15} in 20-D, N=15, mFE=200340					f_{16} in 5-D, N=15, mFE=32578					f_{16} in 20-D, N=15, mFE=200340						
Δf	#	ERT	10%	90%	RT _{succ}	#	ERT	10%	90%	RT _{succ}	#	ERT	10%	90%	RT _{succ}	#	ERT	10%	90%	RT _{succ}	
10	15	8.3e2	5.1e2	1.2e3	8.3e2	15	3.3e4	2.8e4	3.8e4	3.3e4	10	15	2.6e2	2.0e2	3.1e2	2.6e2	15	4.3e3	2.3e3	6.3e3	4.3e3
1	15</																				

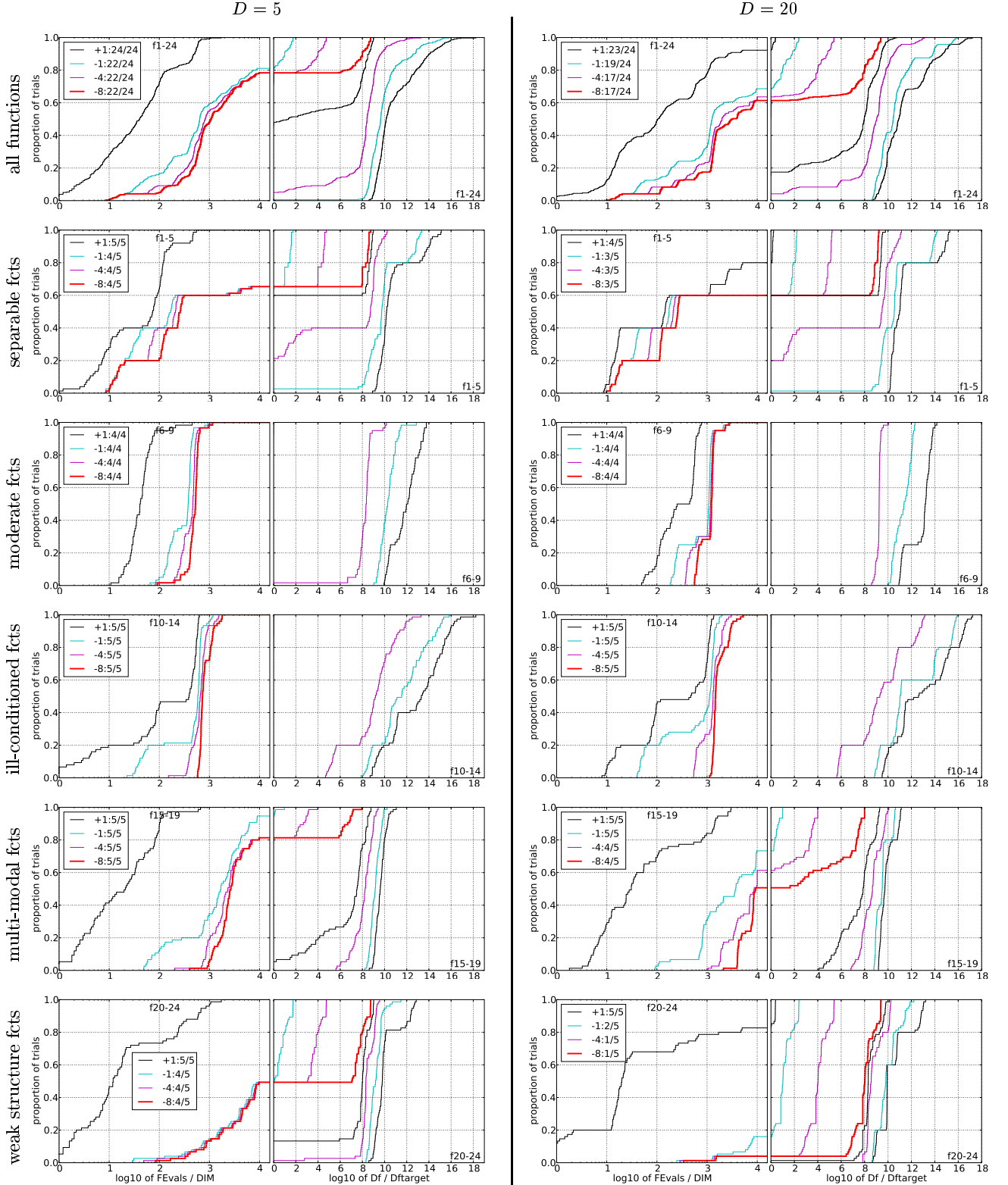


Figure 2: Empirical cumulative distribution functions (ECDFs), plotting the fraction of trials versus running time (left subplots) or versus Δf (right subplots). The thick red line represents the best achieved results. Left subplots: ECDF of the running time (number of function evaluations), divided 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.

modal separable functions as expected. Its performances on multimodal functions, even separable ones such as f_3 and f_4 , are limited though. Whereas for multimodal functions increasing the maximum number of function evaluations is likely to improve the performances of the algorithm, this should not be the case for f_{24} . For the timing experiment, the proposed algorithm was run on f_8 and restarted until at least 30 seconds have passed (according to Figure 2 in [3]). The experiments were conducted with an Intel Core 2 6700 processor (2.66GHz) with Matlab R2008a on Linux 2.6.24.7. The results were 15, 13, 11, 9.7, 9.9, and 13×10^{-5} seconds per function evaluations in dimension 2, 3, 5, 10, 20, and 40 respectively.

Acknowledgments

The first author would like to acknowledge the support, help, and work of the BBOB team with particular kudos to Anne Auger, Steffen Finck and Nikolaus Hansen.

5. REFERENCES

- [1] A. Auger and N. Hansen. A restart CMA evolution strategy with increasing population size. In *Proceedings of the IEEE Congress on Evolutionary Computation (CEC 2005)*, pages 1769–1776. IEEE Press, 2005.
- [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.
- [5] N. Hansen and A. Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evolutionary computation*, 9(2):159–195, 2001.
- [6] M. Lunacek, D. Whitley, and A. Sutton. The impact of global structure on search. In G. Rudolph, T. Jansen, S. M. Lucas, C. Poloni, and N. Beume, editors, *PPSN*, volume 5199 of *Lecture Notes in Computer Science*, pages 498–507. Springer, 2008.
- [7] R. Ros and N. Hansen. A simple modification in CMA-ES achieving linear time and space complexity. In G. Rudolph, T. Jansen, S. M. Lucas, C. Poloni, and N. Beume, editors, *Parallel Problem Solving from Nature - PPSN X, 10th International Conference Dortmund, Germany, September 13-17, 2008, Proceedings*, volume 5199 of *Lecture Notes in Computer Science*, pages 296–305. Springer, 2008.