

Benchmarking the (1+1)-ES with One-Fifth Success rule on the BBOB-2009 Noisy Testbed

Anne Auger

► **To cite this version:**

Anne Auger. Benchmarking the (1+1)-ES with One-Fifth Success rule on the BBOB-2009 Noisy Testbed. ACM-GECCO Genetic and Evolutionary Computation Conference, Jul 2009, Montreal, Canada. 2009. <inria-00430516>

HAL Id: inria-00430516

<https://hal.inria.fr/inria-00430516>

Submitted on 8 Nov 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 the (1+1)-ES with One-Fifth Success rule on the BBOB-2009 Noisy Testbed

Anne Auger
TAO Team, INRIA Saclay Ile-de-France
LRI, Bat 490 Univ. Paris-Sud
91405 Orsay Cedex France
Anne.Auger@inria.fr

ABSTRACT

The (1+1)-ES with one-fifth success rule is one of the first and simplest stochastic algorithm proposed for optimization on a continuous search space in a black-box scenario. In this paper, we benchmark an independent-restart (1+1)-ES with one-fifth success rule on the BBOB-2009 noisy testbed. The maximum number of function evaluations used equals 10^6 times the dimension of the search space. The algorithm could only solve 3 functions with moderate noise in 5-D and 2 functions 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

1. INTRODUCTION

The (1+1)-ES with one-fifth success rule is one of the earliest and simplest adaptive stochastic search algorithm [8, 7, 3]. This paper complements [1] where an independent-restart implementation of the (1+1)-ES with one-fifth success rule is benchmarked on the BBOB-2009 noise-free testbed. Indeed, we test exactly the same algorithm, using the same settings on the BBOB-2009 noisy testbed. For the description of the algorithm and the settings we refer to [1].

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.

2. RESULTS AND DISCUSSION

Results from experiments according to [5] on the benchmark functions given in [4, 6] are presented in Figures 1 and 2 and in Tables 1 and 2.

We observe that globally the algorithm performs poorly. In 5-D, only f_{101} , f_{102} , f_{103} are solved and in 20-D only f_{101} and f_{102} are solved. The functions solved belong to the class of functions with moderate noise.

Acknowledgments

The author would like to acknowledge Nikolaus Hansen, Steffen Finck and Raymond Ros for their great and hard work on the BBOB project.

3. REFERENCES

- [1] A. Auger. Benchmarking the (1+1) Evolution Strategy with One-Fifth Success Rule on the BBOB-2009 Function Testbed. In *Workshop Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2009)*. ACM Press, 2009.
- [2] A. Auger and N. Hansen. Benchmarking the (1+1)-CMA-ES on the BBOB-2009 Function Testbed. In *Workshop Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2009)*. ACM Press, 2009.
- [3] L. Devroye. The compund random search. In *International Symposium on Systems Engineering and Analysis*, pages 195–110. Purdue University, 1972.
- [4] S. Finck, N. Hansen, R. Ros, and A. Auger. Real-parameter black-box optimization benchmarking 2009: Presentation of the noisy functions. Technical Report 2009/21, Research Center PPE, 2009.
- [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, S. Finck, R. Ros, and A. Auger. Real-parameter black-box optimization benchmarking 2009: Noisy functions definitions. Technical Report RR-6869, INRIA, 2009.
- [7] I. Rechenberg. *Evolutionstrategie*. Friedrich Frommann Verlag (Günther Holzboog KG), Stuttgart, 1973.
- [8] M. Schumer and K. Steiglitz. Adaptive step size random search. *Automatic Control, IEEE Transactions on*, 13:270–276, 1968.

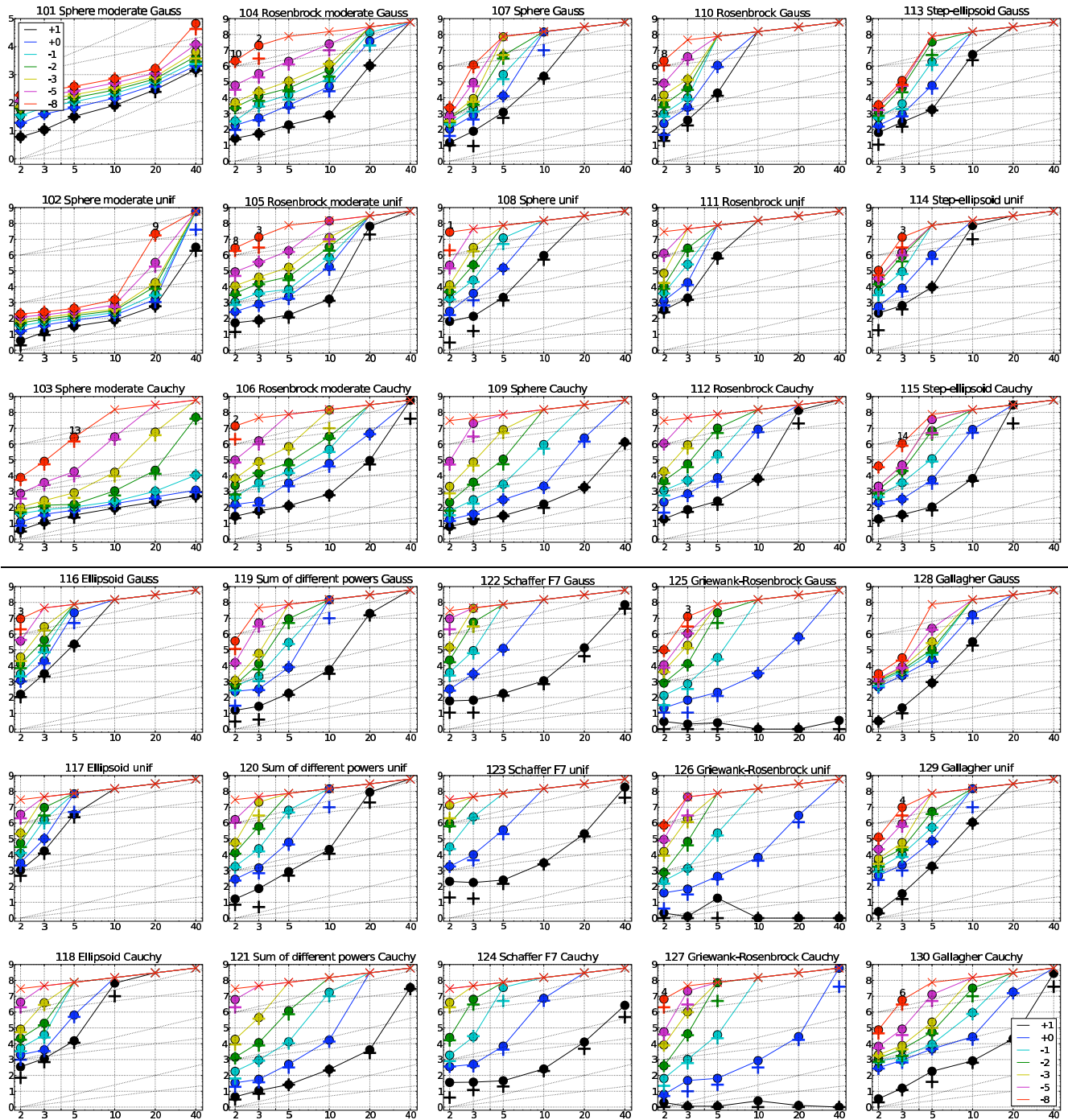


Figure 1: Expected Running Time (ERT, ●) to reach $f_{\text{opt}} + \Delta f$ and median number of function evaluations of successful trials (+), shown for $\Delta f = 10, 1, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-5}, 10^{-8}$ (the exponent is given in the legend of f_{101} and f_{130}) versus dimension in log-log presentation. The $ERT(\Delta f)$ 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 (x) 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_{101} in 5-D, N=15, mFE=438					f_{101} in 20-D, N=15, mFE=2971					f_{102} in 5-D, N=15, mFE=886					f_{102} in 20-D, N=15, mFE=2.00e7						
Δf	#	ERT	10%	90%	RT _{succ}	#	ERT	10%	90%	RT _{succ}	#	ERT	10%	90%	RT _{succ}	#	ERT	10%	90%	RT _{succ}	
10	15	3.2e1	2.7e1	3.8e1	3.2e1	15	2.9e2	2.4e2	3.5e2	2.9e2	15	3.2e1	2.7e1	3.7e1	3.2e1	15	6.1e2	5.0e2	7.2e2	6.1e2	
1	15	6.7e1	5.9e1	7.4e1	6.7e1	15	4.3e2	3.7e2	4.9e2	4.3e2	1	15	7.6e1	6.9e1	8.2e1	7.6e1	15	1.6e3	1.2e3	2.0e3	1.6e3
le-1	15	1.1e2	9.9e1	1.2e2	1.1e2	15	5.7e2	5.2e2	6.3e2	5.7e2	le-1	15	1.1e2	1.0e2	1.2e2	1.1e2	15	3.2e3	2.4e3	4.0e3	3.2e3
le-3	15	1.9e2	1.8e2	2.0e2	1.9e2	15	8.4e2	7.9e2	9.0e2	8.4e2	le-3	15	2.0e2	1.9e2	2.0e2	2.0e2	15	1.8e4	1.4e4	2.3e4	1.8e4
le-5	15	2.6e2	2.5e2	2.7e2	2.6e2	15	1.1e3	1.1e3	1.2e3	1.1e3	le-5	15	2.8e2	2.7e2	2.9e2	2.8e2	15	3.4e5	2.4e5	4.4e5	3.4e5
le-8	15	3.8e2	3.7e2	3.9e2	3.8e2	15	1.6e3	1.5e3	1.8e3	1.6e3	le-8	15	4.3e2	4.0e2	4.8e2	4.3e2	9	2.3e7	1.7e7	3.3e7	1.3e7

f_{103} in 5-D, N=15, mFE=5.00e6					f_{103} in 20-D, N=15, mFE=2.00e7					f_{104} in 5-D, N=15, mFE=5.00e6					f_{104} in 20-D, N=15, mFE=2.00e7						
Δf	#	ERT	10%	90%	RT _{succ}	#	ERT	10%	90%	RT _{succ}	#	ERT	10%	90%	RT _{succ}	#	ERT	10%	90%	RT _{succ}	
10	15	3.3e1	2.5e1	4.1e1	3.3e1	15	2.2e2	2.1e2	2.3e2	2.2e2	10	15	1.9e2	1.4e2	2.5e2	1.9e2	15	1.1e6	8.0e5	1.3e6	1.1e6
1	15	6.8e1	6.1e1	7.5e1	6.8e1	15	4.6e2	3.9e2	5.3e2	4.6e2	1	15	3.3e3	2.3e3	4.4e3	3.3e3	6	3.8e7	2.5e7	6.5e7	1.5e7
le-1	15	1.1e2	9.9e1	1.1e2	1.1e2	15	1.0e3	8.5e2	1.3e3	1.0e3	le-1	15	1.5e4	1.1e4	2.0e4	1.5e4	2	1.3e8	6.6e7	>3e8	1.2e7
le-3	15	8.4e2	5.2e2	1.2e3	8.4e2	14	5.7e6	4.0e6	7.5e6	5.5e6	le-3	15	1.1e5	7.7e4	1.5e5	1.1e5	0	12e-1	97e-3	33e-1	7.9e6
le-5	15	1.8e4	9.3e3	2.9e4	1.8e4	0	42e-5	32e-5	74e-5	7.9e6	le-5	14	2.0e6	1.5e6	2.6e6	2.0e6
le-8	13	2.6e6	1.9e6	3.3e6	2.4e6	le-8	0	25e-8	15e-9	73e-7	1.8e6

f_{105} in 5-D, N=15, mFE=5.00e6					f_{105} in 20-D, N=15, mFE=2.00e7					f_{106} in 5-D, N=15, mFE=5.00e6					f_{106} in 20-D, N=15, mFE=2.00e7						
Δf	#	ERT	10%	90%	RT _{succ}	#	ERT	10%	90%	RT _{succ}	#	ERT	10%	90%	RT _{succ}	#	ERT	10%	90%	RT _{succ}	
10	15	1.7e2	1.2e2	2.2e2	1.7e2	4	6.6e7	4.0e7	1.4e8	1.5e7	10	15	1.3e2	1.1e2	1.4e2	1.3e2	15	9.2e4	4.6e4	1.4e5	9.2e4
1	15	2.4e3	1.9e3	2.9e3	2.4e3	0	13e+0	37e-1	15e+0	1.1e7	1	15	3.5e3	2.4e3	4.6e3	3.5e3	15	4.6e6	3.5e6	5.8e6	4.6e6
le-1	15	6.9e3	5.0e3	8.9e3	6.9e3	le-1	15	1.8e4	1.3e4	2.4e4	1.8e4	0	42e-2	13e-2	62e-2	1.0e7
le-3	15	1.7e5	1.2e5	2.2e5	1.7e5	le-3	15	6.9e5	4.6e5	9.4e5	6.9e5
le-5	14	1.9e6	1.4e6	2.4e6	1.8e6	le-5	0	10e-5	35e-6	28e-5	3.2e6
le-8	0	42e-8	13e-8	38e-7	2.8e6	le-8

Table 1: Shown are, for functions f_{101} - f_{120} and 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 1); the 10%-tile and 90%-tile of the bootstrap distribution of ERT; the average number of function evaluations in successful trials or, if none was successful, as last entry the median number of function evaluations to reach the best function value (RT_{succ}). 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 1 for the names of functions.

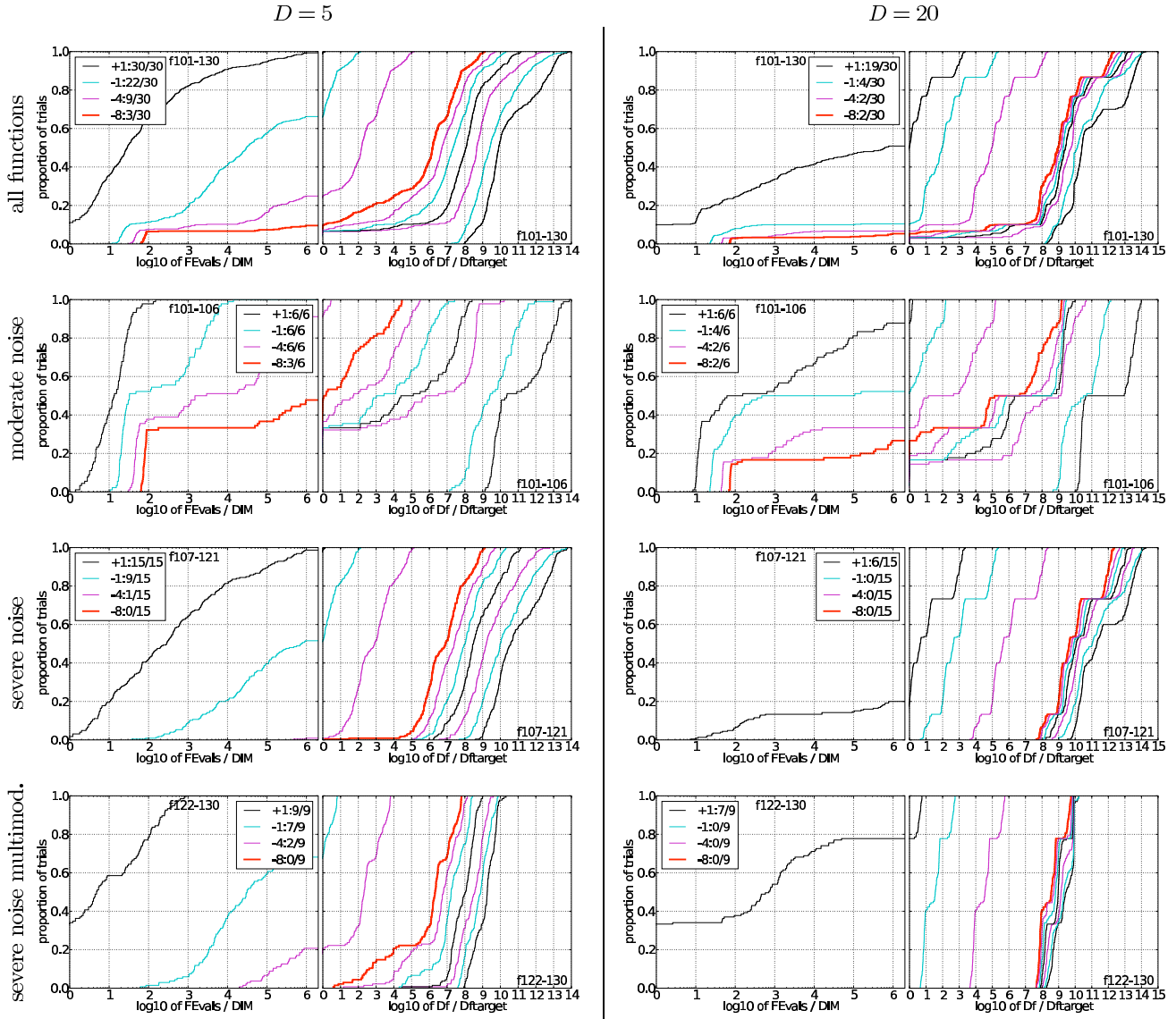


Figure 2: Empirical cumulative distribution functions (ECDFs), plotting the fraction of trials versus running time (left subplots) or versus Δf (right subplots). The thick red line represents the best achieved results. Left subplots: ECDF of the running time (number of function evaluations), divided 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: moderate noise functions; third row: severe noise functions; fourth row: severe noise and highly-multimodal functions. 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.

f_{121} in 5-D, N=15, mFE=5.00e6					f_{121} in 20-D, N=15, mFE=2.00e7					f_{122} in 5-D, N=15, mFE=5.00e6					f_{122} in 20-D, N=15, mFE=2.00e7							
Δf	#	ERT	10%	90%	RT _{succ}	#	ERT	10%	90%	RT _{succ}	Δf	#	ERT	10%	90%	RT _{succ}	#	ERT	10%	90%	RT _{succ}	
10	15	2.6e1	2.1e1	3.2e1	2.6e1	15	4.1e3	3.0e3	5.3e3	4.1e3	10	15	1.8e2	1.3e2	2.2e2	1.8e2	15	1.4e5	6.1e4	2.2e5	1.4e5	
1	15	5.0e2	3.6e2	6.5e2	5.0e2	0	<i>10e-1</i>	<i>12e-1</i>	<i>21e-1</i>	7.9e6	1	15	1.2e5	9.2e4	1.6e5	1.2e5	0	<i>57e-1</i>	<i>39e-1</i>	<i>62e-1</i>	7.9e6	
1e-1	15	1.3e4	1.0e4	1.7e4	1.3e4	1e-1	0	<i>24e-2</i>	<i>17e-2</i>	<i>40e-2</i>	1.6e6	
1e-3	0	<i>51e-4</i>	<i>28e-4</i>	<i>80e-4</i>	2.5e6	1e-3	
1e-5	1e-5
1e-8	1e-8
f_{123} in 5-D, N=15, mFE=5.00e6					f_{123} in 20-D, N=15, mFE=2.00e7					f_{124} in 5-D, N=15, mFE=5.00e6					f_{124} in 20-D, N=15, mFE=2.00e7							
10	15	2.5e2	1.6e2	3.4e2	2.5e2	15	2.1e5	1.4e5	2.9e5	2.1e5	10	15	4.6e1	1.8e1	7.5e1	4.6e1	15	1.3e4	8.4e3	1.7e4	1.3e4	
1	15	3.7e5	2.6e5	4.9e5	3.7e5	0	<i>64e-1</i>	<i>59e-1</i>	<i>68e-1</i>	7.9e6	1	15	7.0e3	4.5e3	9.8e3	7.0e3	0	<i>36e-1</i>	<i>32e-1</i>	<i>43e-1</i>	7.1e6	
1e-1	0	<i>59e-2</i>	<i>45e-2</i>	<i>63e-2</i>	2.2e6	1e-1	2	3.6e7	1.9e7	>7e7	5.0e6	
1e-3	1e-3	0	<i>12e-2</i>	<i>92e-3</i>	<i>15e-2</i>	2.0e6	
1e-5	1e-5
1e-8	1e-8
f_{125} in 5-D, N=15, mFE=5.00e6					f_{125} in 20-D, N=15, mFE=2.00e7					f_{126} in 5-D, N=15, mFE=5.00e6					f_{126} in 20-D, N=15, mFE=2.00e7							
10	15	2.5e0	1.1e0	4.0e0	2.5e0	15	1.0e0	1.0e0	1.0e0	1.0e0	10	15	1.8e1	2.9e0	3.4e1	1.8e1	15	1.0e0	1.0e0	1.0e0	1.0e0	
1	15	2.1e2	1.3e2	3.0e2	2.1e2	15	6.5e5	5.0e5	8.2e5	6.5e5	1	15	4.3e2	2.7e2	6.1e2	4.3e2	15	3.1e6	2.1e6	4.1e6	3.1e6	
1e-1	15	3.3e4	2.3e4	4.3e4	3.3e4	0	<i>75e-2</i>	<i>66e-2</i>	<i>77e-2</i>	1.3e7	1e-1	15	2.4e5	1.7e5	3.0e5	2.4e5	0	<i>85e-2</i>	<i>77e-2</i>	<i>90e-2</i>	8.9e6	
1e-3	0	<i>13e-3</i>	<i>71e-4</i>	<i>23e-3</i>	2.0e6	1e-3	0	<i>27e-3</i>	<i>15e-3</i>	<i>41e-3</i>	2.0e6	
1e-5	1e-5
1e-8	1e-8
f_{127} in 5-D, N=15, mFE=5.00e6					f_{127} in 20-D, N=15, mFE=2.00e7					f_{128} in 5-D, N=15, mFE=5.00e6					f_{128} in 20-D, N=15, mFE=2.00e7							
10	15	1.1e0	1.0e0	1.3e0	1.1e0	15	1.3e0	1.0e0	1.5e0	1.3e0	10	15	8.4e2	6.4e2	1.0e3	8.4e2	0	<i>34e+0</i>	<i>24e+0</i>	<i>48e+0</i>	7.1e6	
1	15	6.5e1	3.1e1	1.0e2	6.5e1	15	2.8e4	1.8e4	3.8e4	2.8e4	1	15	2.5e4	1.9e4	3.0e4	2.5e4	
1e-1	15	3.6e4	2.5e4	4.9e4	3.6e4	0	<i>62e-2</i>	<i>50e-2</i>	<i>66e-2</i>	1.1e7	1e-1	15	6.5e4	4.9e4	8.3e4	6.5e4	
1e-3	0	<i>24e-3</i>	<i>15e-3</i>	<i>30e-3</i>	2.5e6	1e-3	15	3.2e5	2.2e5	4.3e5	3.2e5	
1e-5	1e-5	14	2.3e6	1.7e6	3.0e6	2.1e6	
1e-8	1e-8	0	<i>17e-7</i>	<i>46e-9</i>	<i>68e-7</i>	2.5e6	
f_{129} in 5-D, N=15, mFE=5.00e6					f_{129} in 20-D, N=15, mFE=2.00e7					f_{130} in 5-D, N=15, mFE=5.00e6					f_{130} in 20-D, N=15, mFE=2.00e7							
10	15	1.9e3	1.5e3	2.3e3	1.9e3	0	<i>48e+0</i>	<i>34e+0</i>	<i>56e+0</i>	1.0e7	10	15	1.8e2	1.1e2	2.5e2	1.8e2	15	2.1e4	1.4e4	2.7e4	2.1e4	
1	15	7.3e4	5.2e4	9.2e4	7.3e4	1	15	4.7e3	3.4e3	6.0e3	4.7e3	10	1.8e7	1.2e7	2.8e7	9.2e6	
1e-1	15	5.3e5	4.3e5	6.4e5	5.3e5	1e-1	15	9.3e3	7.3e3	1.1e4	9.3e3	0	<i>74e-2</i>	<i>46e-2</i>	<i>16e-1</i>	7.1e6	
1e-3	0	<i>60e-4</i>	<i>24e-4</i>	<i>18e-3</i>	2.2e6	1e-3	15	2.3e5	1.4e5	3.3e5	2.3e5	
1e-5	1e-5	5	1.3e7	8.6e6	2.3e7	4.3e6	
1e-8	1e-8	0	<i>57e-6</i>	<i>31e-7</i>	<i>13e-5</i>	2.5e6	

Table 2: Shown are, for functions f_{121} - f_{130} and 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 1); the 10%-tile and 90%-tile of the bootstrap distribution of ERT; the average number of function evaluations in successful trials or, if none was successful, as last entry the median number of function evaluations to reach the best function value (RT_{succ}). 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 1 for the names of functions.