

# Black-Box Optimization Benchmarking: Results for the BayEDA<sub>cG</sub> Algorithm on the Noiseless Function Testbed

Marcus Gallagher  
School of Information Technology and Electrical Engineering  
The University of Queensland  
Brisbane, 4072. Australia  
marcusg@itee.uq.edu.au

## 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

This paper presents experimental results for the BayEDA<sub>cG</sub> continuous optimization algorithm on the BBOB noise free benchmark problem suite as part of the GECCO'09 Workshop.

## 2. ALGORITHM PRESENTATION

BayEDA<sub>cG</sub> is an Estimation of Distribution Algorithm that uses Bayesian Inference to learn a posterior distribution over model parameters for the probability density estimation model used [2]. In this algorithm, the distribution is a product of univariate Gaussian distributions and inference is performed over mean and variance parameters for each dimension in the search space. The algorithm is described for a one-dimensional problem in Table 1. Given the factorized probability model, the extension to the multidimensional case is straightforward.

The following description of the algorithm is from [2]:

For a univariate Gaussian (Normal) model distribution, Bayesian inference can readily be carried out: the resulting expressions given here are drawn from Gelman et al. [3]. We consider the simplest case of a noninformative (flat) prior for the model parameters, expressing no preference for any

Table 1: Algorithm: BayEDA<sub>cG</sub>.

---

```

Given: population size  $M$ , selection parameter  $\tau$ 
BEGIN (set  $t = 0$ ) Generate  $M$  individ. uniformly in  $S$ 
REPEAT for  $t = 1, 2, \dots$  until stopping criterion is met
    Select  $M_{sel} = \text{Round}(M \cdot \tau)$  individ. via truncation
    Calculate sample mean  $\bar{x}$  and variance  $s^2$  of  $D$ 
    Sample  $M$  individuals from  $p_t(\mathbf{x}|D, \theta)$ :
    FOR  $i=1:M$ 
        Draw sample variance  $\tilde{\sigma}^2 \sim \text{Inv} - \chi^2(M_{sel} - 1, s^2)$ 
        Draw sample mean  $\tilde{\mu} \sim N(\bar{x}, \tilde{\sigma}^2/(M_{sel}))$ 
        Draw new individual  $\mathbf{x}_i \sim N(\tilde{\mu}, \tilde{\sigma}^2)$ 
    ENDFOR
ENDREPEAT
END

```

---

particular values for the model parameters before observing any data. In this case, inference depends only on the data (selected individuals). The standard noninformative prior is uniform on  $(\mu, \log \sigma^2)$  or

$$p(\mu, \sigma^2) \propto (\sigma^2)^{-1}$$

The joint posterior can be factorised as

$$p(\mu, \sigma^2|D) = p(\mu|\sigma^2, D)p(\sigma^2|D)$$

In this case, the marginal density for  $\sigma$  is

$$\sigma^2|D \sim \text{Inv} - \chi^2(M_{sel} - 1, s^2) \quad (1)$$

where  $s^2$  is the sample variance of the data. The conditional density for  $\mu$  is

$$\mu|D, \sigma^2 \sim N(\bar{x}, \sigma^2/M_{sel}) \quad (2)$$

where  $\bar{x}$  is the sample mean of the data  $D$ .

The predictive distribution for  $\tilde{x}$  given the data,  $\mu$  and  $\sigma$  is

$$\tilde{x}|D, \mu, \sigma^2 \sim N(\mu, \sigma^2) \quad (3)$$

In the BayEDA<sub>cG</sub> algorithm, sampling from the posterior predictive distribution  $p(\tilde{x}|D)$  can be easily carried out in a three-step process. Firstly, a sample  $\tilde{\sigma}^2$  is drawn from (1), then this sample is used to draw a sample  $\tilde{\mu}$  from (2) and finally both samples are used to draw a sample  $\tilde{x}$  from (3). The process is repeated  $M$  times to produce the population for use in the next generation.

The algorithm is summarized in Table 1. Note that for implementation purposes, a random draw  $y$  from an inverse- $\chi^2$  distribution can be obtained by firstly drawing a sample

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.

$z$  from the  $\chi^2$  distribution and applying  $y = s^2/z$ . The  $\chi^2$  distribution is also a special case of the gamma distribution (see [3] for details).

### 3. EXPERIMENTAL PROCEDURE

The BayEDA<sub>cG</sub> algorithm was run on the current set of BBOB noiseless benchmark functions (see other document for results on noisy functions). No parameter tuning was attempted with respect to the functions. The population size ( $M$ ) was set to 10 times the dimensionality of the problem. The selection threshold  $\tau$  (for truncation selection) was set (rather arbitrarily) to 0.8. The total number of function evaluations was set to 2000 times the problem dimensionality (making the total number of generations for each run of the algorithm equal to 200).

The *crafting effort* for this set of experiments is zero in this case.

### 4. RESULTS

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

### 5. CPU TIMING EXPERIMENT

For the timing experiment the BayEDA<sub>cG</sub> algorithm was run as suggested on  $f_8$  until at least 30 seconds has passed. The experiments in this paper were conducted on an Intel Pentium 4 quad core 2.4Ghz processor running Linux 2.6.24-23 SMP and Matlab R2008a. The results were (in seconds per function evaluation) 2.9;3.2;3.8;5.3 and  $8.3 \times 10^{-4}$  for dimension 2;3;5;10 and 20D and  $1.4 \times 10^{-3}$  for 40D.

### 6. CONCLUSION

The results show a wide variety of performance across the different test functions. Some functions seem reasonably well-solved for a range of precision and dimensionality values while others show only modest performance. This is for a reasonably small number of function evaluations. Nevertheless, it will be interesting to see comparative analysis at the workshop.

### 7. REFERENCES

- [1] 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.
- [2] M. Gallagher, I. Wood, J. Keith, and G. Sofronov. Posterior predictive optimization with thresholding: Bayesian inference in EDAs. In *Proc. Congress on Evolutionary Computation (CEC'07)*, pages 127–133, 2007.
- [3] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, second edition, 2004.
- [4] N. Hansen, A. Auger, S. Finck, and R. Ros. Real-parameter black-box optimization benchmarking 2009: Experimental setup. Technical Report RR-6828, INRIA, 2009.
- [5] 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.

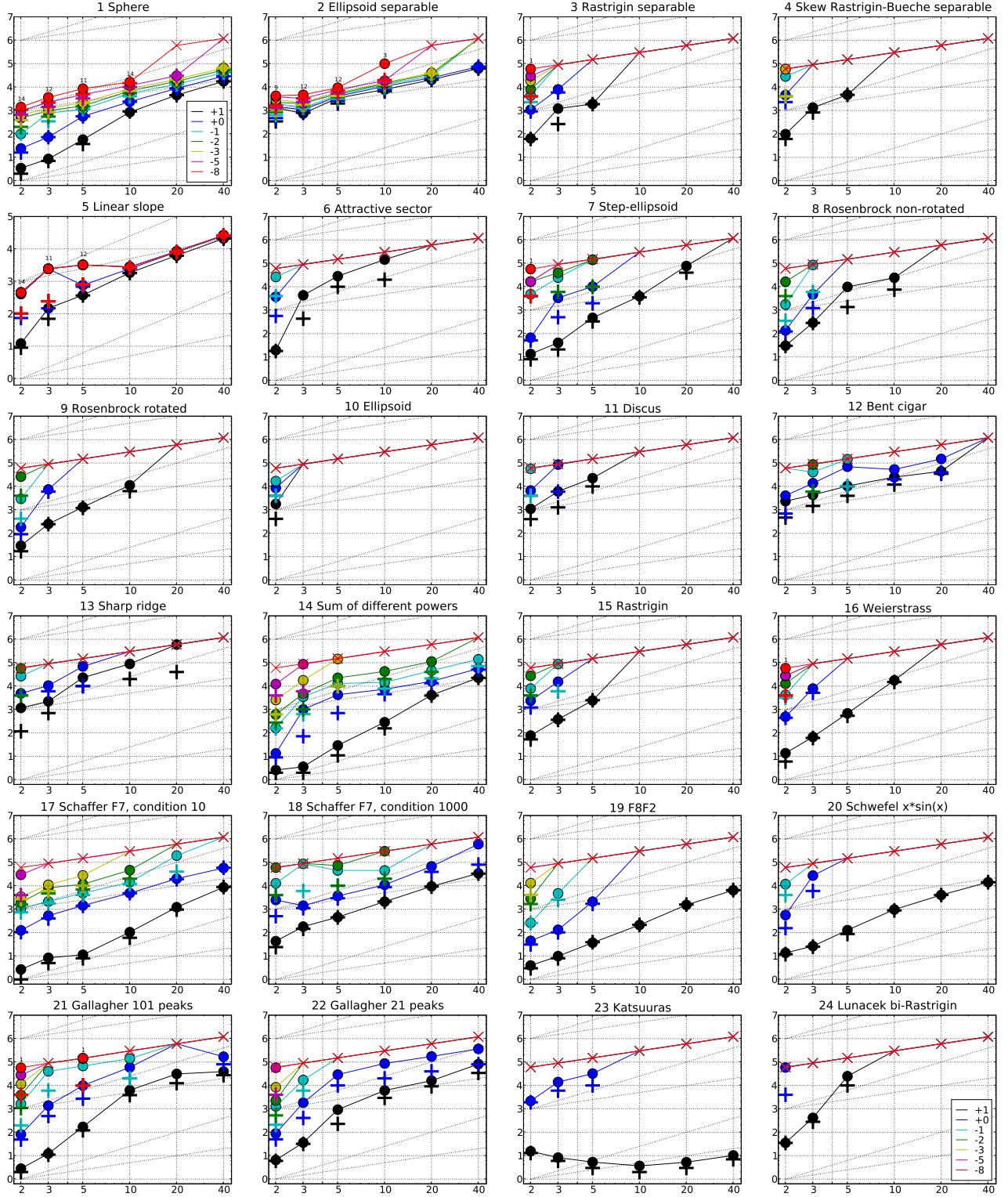
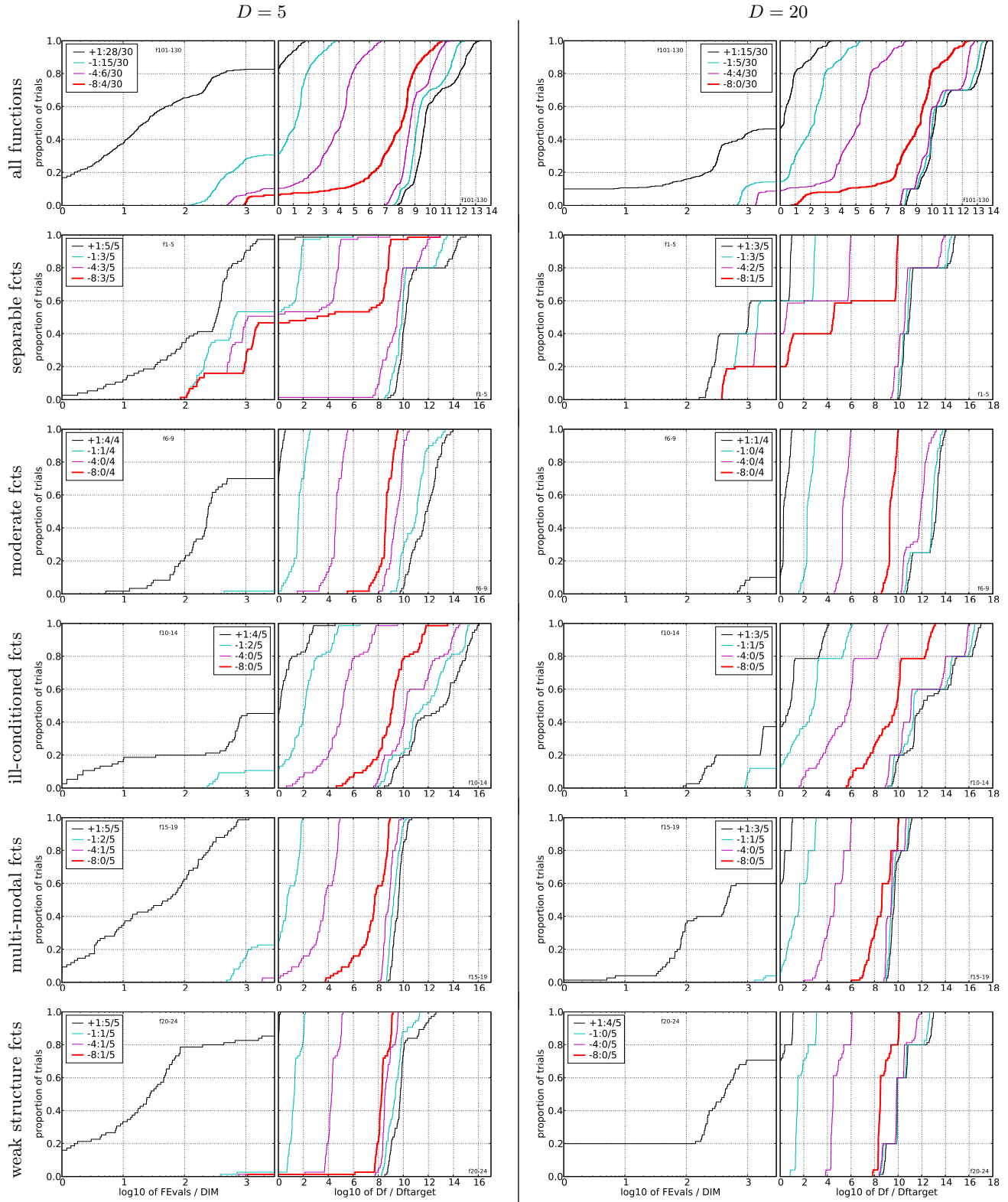


Figure 1: Expected Running Time (ERT, ●) to reach  $f_{\text{opt}} + \Delta f$  and median number of function evaluations of successful trials (+), shown for  $\Delta f = 10, 1, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-5}, 10^{-8}$  (the exponent is given in the legend of  $f_1$  and  $f_{24}$ ) versus dimension in log-log presentation. The ERT( $\Delta f$ ) equals to  $\text{\#Fes}(\Delta f)$  divided by the number of successful trials, where a trial is successful if  $f_{\text{opt}} + \Delta f$  was surpassed during the trial. The  $\text{\#Fes}(\Delta f)$  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  $f_{\text{opt}}$  denotes the optimal function value. Crosses (x) indicate the total number of function evaluations  $\text{\#Fes}(-\infty)$ . 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=10000						$f_1$ in 20-D, N=15, mFE=40000						$f_2$ in 5-D, N=15, mFE=10000						$f_2$ in 20-D, N=15, mFE=40000									
$\Delta f$	#	ERT	10%	90%	RT <sub>succ</sub>	#	ERT	10%	90%	RT <sub>succ</sub>	$\Delta f$	#	ERT	10%	90%	RT <sub>succ</sub>	#	ERT	10%	90%	RT <sub>succ</sub>	#	ERT	10%	90%	RT <sub>succ</sub>	
10	15	5.7e1	3.9e1	7.6e1	5.7e1	15	4.6e3	4.4e3	4.7e3	4.6e3	10	13	3.4e3	2.2e3	4.6e3	3.1e3	15	2.1e4	2.0e4	2.1e4	2.1e4	15	2.1e4	2.0e4	2.1e4	2.1e4	
1	15	5.6e2	5.3e2	6.0e2	5.6e2	15	8.8e3	8.6e3	9.0e3	8.8e3	1	13	4.1e3	3.0e3	5.2e3	3.7e3	15	2.5e4	2.5e4	2.5e4	2.5e4	15	2.5e4	2.5e4	2.5e4	2.5e4	
1e-1	15	1.1e3	1.1e3	1.2e3	1.1e3	15	1.3e4	1.3e4	1.3e4	1.3e4	1e-1	13	4.6e3	3.6e3	5.7e3	4.1e3	15	2.9e4	2.9e4	3.0e4	2.9e4	15	2.9e4	2.9e4	3.0e4	2.9e4	
1e-3	15	2.1e3	2.0e3	2.2e3	2.1e3	15	2.1e4	2.1e4	2.2e4	2.1e4	1e-3	13	5.7e3	4.9e3	6.6e3	5.1e3	14	4.1e4	4.0e4	4.1e4	3.8e4	14	4.1e4	4.0e4	4.1e4	3.8e4	
1e-5	13	4.8e3	3.8e3	5.8e3	4.3e3	15	3.0e4	3.0e4	3.1e4	3.0e4	1e-5	12	7.7e3	6.9e3	8.5e3	6.0e3	0	34e-5	23e-5	42e-5	3.5e4	0	34e-5	23e-5	42e-5	3.5e4	
1e-8	11	8.5e3	7.6e3	9.6e3	6.8e3	0	61e-9	39e-9	11e-8	3.5e4	1e-8	12	9.4e3	8.9e3	1.0e4	7.4e3											
$f_3$ in 5-D, N=15, mFE=10000						$f_3$ in 20-D, N=15, mFE=40000						$f_4$ in 5-D, N=15, mFE=10000						$f_4$ in 20-D, N=15, mFE=40000									
$\Delta f$	#	ERT	10%	90%	RT <sub>succ</sub>	#	ERT	10%	90%	RT <sub>succ</sub>	$\Delta f$	#	ERT	10%	90%	RT <sub>succ</sub>	#	ERT	10%	90%	RT <sub>succ</sub>	#	ERT	10%	90%	RT <sub>succ</sub>	
10	15	1.9e3	1.7e3	2.1e3	1.9e3	0	73e+0	60e+0	92e+0	3.5e4	10	15	4.7e3	4.1e3	5.2e3	4.7e3	0	69e+0	60e+0	80e+0	3.5e4	10	15	4.7e3	4.1e3	5.2e3	4.7e3
1	0	29e-1	19e-1	45e-1	7.9e3						1	0	69e-1	51e-1	94e-1	7.1e3											
1e-1											1e-1																
1e-3											1e-3																
1e-5											1e-5																
1e-8											1e-8																
$f_5$ in 5-D, N=15, mFE=10000						$f_5$ in 20-D, N=15, mFE=12400						$f_6$ in 5-D, N=15, mFE=10000						$f_6$ in 20-D, N=15, mFE=40000									
$\Delta f$	#	ERT	10%	90%	RT <sub>succ</sub>	#	ERT	10%	90%	RT <sub>succ</sub>	$\Delta f$	#	ERT	10%	90%	RT <sub>succ</sub>	#	ERT	10%	90%	RT <sub>succ</sub>	#	ERT	10%	90%	RT <sub>succ</sub>	
10	15	3.8e2	3.2e2	4.3e2	3.8e2	15	6.2e3	6.1e3	6.4e3	6.2e3	10	4	2.8e4	2.3e4	3.3e4	5.5e3	0	60e+0	27e+0	90e+0	4.0e4	10	4	2.8e4	2.3e4	3.3e4	5.5e3
1	15	7.6e2	6.8e2	8.4e2	7.6e2	15	8.0e3	7.7e3	8.3e3	8.0e3	1	0	13e+0	33e-1	28e+0	6.3e2											
1e-1	12	3.2e3	1.7e3	4.8e3	3.0e3	15	8.3e3	8.0e3	8.7e3	8.3e3	1e-1																
1e-3	12	3.2e3	1.7e3	4.8e3	3.0e3	15	8.4e3	8.0e3	8.8e3	8.4e3	1e-3																
1e-5	12	3.2e3	1.7e3	4.8e3	3.0e3	15	8.4e3	8.0e3	8.8e3	8.4e3	1e-5																
1e-8	12	3.2e3	1.7e3	4.8e3	3.0e3	15	8.4e3	8.0e3	8.7e3	8.4e3	1e-8																
$f_7$ in 5-D, N=15, mFE=10000						$f_7$ in 20-D, N=15, mFE=40000						$f_8$ in 5-D, N=15, mFE=10000						$f_8$ in 20-D, N=15, mFE=40000									
$\Delta f$	#	ERT	10%	90%	RT <sub>succ</sub>	#	ERT	10%	90%	RT <sub>succ</sub>	$\Delta f$	#	ERT	10%	90%	RT <sub>succ</sub>	#	ERT	10%	90%	RT <sub>succ</sub>	#	ERT	10%	90%	RT <sub>succ</sub>	
10	15	4.7e2	3.0e2	6.8e2	4.7e2	6	7.7e4	6.7e4	8.7e4	2.9e4	10	8	9.9e3	7.6e3	1.2e4	5.6e3	0	48e+0	18e+0	81e+0	4.0e4	10	8	9.9e3	7.6e3	1.2e4	5.6e3
1	8	1.0e4	7.4e3	1.3e4	4.6e3	0	11e+0	48e-1	21e+0	2.5e4	1	0	45e-1	30e-1	28e+0	2.8e3											
1e-1	1	1.4e5	1.3e5	1.5e5	1.0e4						1e-1																
1e-3	0	73e-2	18e-2	64e-1	2.0e3						1e-3																
1e-5											1e-5																
1e-8											1e-8																
$f_9$ in 5-D, N=15, mFE=10000						$f_9$ in 20-D, N=15, mFE=40000						$f_{10}$ in 5-D, N=15, mFE=10000						$f_{10}$ in 20-D, N=15, mFE=40000									
$\Delta f$	#	ERT	10%	90%	RT <sub>succ</sub>	#	ERT	10%	90%	RT <sub>succ</sub>	$\Delta f$	#	ERT	10%	90%	RT <sub>succ</sub>	#	ERT	10%	90%	RT <sub>succ</sub>	#	ERT	10%	90%	RT <sub>succ</sub>	
10	15	1.3e3	1.1e3	1.5e3	1.3e3	0	18e+0	18e+0	19e+0	4.0e4	10	0	28e+2	30e+1	57e+2	1.4e3	0	42e+3	21e+3	11e+4	3.5e4	10	0	28e+2	30e+1	57e+2	1.4e3
1	0	39e-1	30e-1	47e-1	3.5e3						1																
1e-1											1e-1																
1e-3											1e-3																
1e-5											1e-5																
1e-8											1e-8																
$f_{11}$ in 5-D, N=15, mFE=10000						$f_{11}$ in 20-D, N=15, mFE=40000						$f_{12}$ in 5-D, N=15, mFE=10000						$f_{12}$ in 20-D, N=15, mFE=40000									
$\Delta f$	#	ERT	10%	90%	RT <sub>succ</sub>	#	ERT	10%	90%	RT <sub>succ</sub>	$\Delta f$	#	ERT	10%	90%	RT <sub>succ</sub>	#	ERT	10%	90%	RT <sub>succ</sub>	#	ERT	10%	90%	RT <sub>succ</sub>	
10	5	2.3e4	1.9e4	2.6e4	8.5e3	0	14e+1	95e+0	15e+1	7.1e3	10	9	1.0e4	8.8e3	1.2e4	5.2e3	12	4.4e4	4.3e4	4.5e4	3.5e4	12	4.4e4	4.3e4	4.5e4	3.5e4	
1	0	15e+0	20e-1	43e+0	2.0e3						1	2	6.9e4	6.3e4	7.5e4	7.1e3	4	1.5e5	1.5e5	1.5e5	4.0e4	4	1.5e5	1.5e5	1.5e5	4.0e4	
1e-1											1e-1	1	1.4e5	1.4e5	1.5e5	1.0e4	0	21e-1	71e-2	15e+1	3.5e4	0	21e-1	71e-2	15e+1	3.5e4	
1e-3											1e-3	0	71e-1	10e-2	25e+0	6.3e3											
1e-5											1e-5																
1e-8											1e-8																
$f_{13}$ in 5-D, N=15, mFE=10000						$f_{13}$ in 20-D, N=15, mFE=40000						$f_{14}$ in 5-D, N=15, mFE=10000						$f_{14}$ in 20-D, N=15, mFE=40000									
$\Delta f$	#	ERT	10%	90%	RT <sub>succ</sub>	#	ERT	10%	90%	RT <sub>succ</sub>	$\Delta f$	#	ERT	10%	90%	RT <sub>succ</sub>	#	ERT	10%	90%	RT <sub>succ</sub>	#	ERT	10%	90%	RT <sub>succ</sub>	
10	5	2.3e4	2.0e4	2.6e4	7.3e3	1	5.9e5	5.8e5	6.0e5	4.0e4	10	15	2.9e1	1.6e1	4.4e1	2.9e1	15	4.1e3	3.7e3	4.5e3	4.1e3	15	4.1e3	3.7e3	4.5e3	4.1e3	
1	2	6.9e4	6.3e4	7.5e4	6.9e3	0	49e+0	15e+0	11e+1	3.5e4	1	7	4.3e3	2.5e3	6.1e3	2.3e3	15	1.6e4	1.4e4	1.8e4	1.6e4	15	1.6e4	1.4e4	1.8e4	1.6e4	
1e-1	0	19e+0	75e-2	87e+0	8.9e3						1e-1	1	1.3e4	1.0e4	1.7e4	3.9e3	9	4.6e4	4.1e4	5.1e4	3.5e4	9	4.6e4	4.1e4	5.1e4	3.5e4	
1e-3											1e-3	1	1.4e5	1.4e5	1.5e5	1.0e4	0	15e-3	59e-4	27e-2	3.5e4	0	15e-3	59e-4	27e-2	3.5e4	
1e-5											1e-5	0	12e-2	13e-4	20e-1	3.5e3											
1e-8											1e-8																
$f_{15}$ in 5-D, N=15, mFE=10000						$f_{15}$ in 20-D, N=15, mFE=40000						$f_{16}$ in 5-D, N=15, mFE=10000						$f_{16}$ in 20-D, N=15, mFE=40000									
$\Delta f$	#	ERT	10%	90%	RT <sub>succ</sub>	#	ERT	10%	90%	RT <sub>succ</sub>	$\Delta f$	#	ERT	10%	90%	RT <sub>succ</sub>	#	ERT	10%	90%	RT <sub>succ</sub>	#	ERT	10%	90%	RT <sub>succ</sub>	
10	15	2.4e3	2.0e3	2.8e3	2.4e3	0	93e+0	83e+0	11e+1	2.8e4	10	15	6.9e2	4.9e2	8.8e2	6.9e2	0	21e+0	17e+0	24e+0	2.0e4	10	15	6.9e2	4.9e2	8.8e2	6.9e2
1	0	61e-1	48e-1	78e-1	7.1e3						1	0	35e-1	25e-1	63e-1	3.5e3											
1e-1											1e-1																
1e-3											1e-3																
1e-5											1e-5																
1e-8											1e-8																
$f_{17}$ in 5-D, N=15, mFE=10000						$f_{17}$ in 20-D, N=15, mFE=40000						$f_{18}$ in 5-D, N=15, mFE=10000						$f_{18}$ in 20-D, N=15, mFE=40000									
$\Delta f$	#	ERT	10%	90%	RT <sub>succ</sub>	#	ERT	10%																			



**Figure 2: Empirical cumulative distribution functions (ECDFs), plotting the fraction of trials versus running time (left) or  $\Delta f$ .** 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  $\Delta f$  divided by  $10^k$  (upper left lines in continuation of the left subplot), and best achieved  $\Delta f$  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  $\Delta f$  and Df denote the difference to the optimal function value.